

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة جيلالي بونعاما خميس مليانة
Université Djilali Bounaama de khemis Miliana
كلية علوم الطبيعة و الحياة وعلوم الأرض
Faculté des Sciences de la Nature et de la vie et science de la Terre



Polycopié de Cours de Biologie moléculaire

Département : Biologie

Cycle : 3ème année Licence

Présenté par :

Dr. SAADI Wiam

Année Universitaire

2022/2023

Table des matières



Objectifs	4
I - Introduction générale	5
1. Définition	5
2. Bref historique.....	5
3. Les Acides Nucléiques	6
3.1. L'Acide DésoxyriboNucléique (ADN).....	7
3.2. L'Acide RiboNucléique (ARN)	8
II - Notion de gène et transmission de l'information génétique	9
1. Introduction	9
2. Réplication de l'ADN	9
2.1. La réplication chez les procaryotes (<i>E.coli</i>)	10
2.2. La réplication chez les eucaryotes	16
III - Structure et fonction des gènes	18
1. Structure des gènes.....	18
1.1. Structure des gènes procaryotes.....	18
1.2. Structure des gènes eucaryotes.....	18
2. Les grandes étapes de l'utilisation de l'information génétique.....	19
2.1. La Transcription.....	19
2.2. La transcription chez les procaryotes	22
2.3. La transcription chez les eucaryotes	27
2.4. La traduction chez les procaryotes.....	31
2.5. Les spécificités de la traduction eucaryote	36
IV - Régulation de l'expression des gènes	37
1. L'expression d'un gène	37
2. La régulation génétique	37
3. Principes de la régulation génique.....	37
4. Rappel: organisation des gènes bactériens	38
5. Les différents niveaux de régulation génétique chez les bactéries	38
6. Opéron inductible et opéron répressible.....	38
6.1. Régulation de l'opéron lactose chez <i>E. coli</i>	39
6.2. Régulation de l'opéron Tryptophane chez <i>E. coli</i>	41

V - Prédiction de gènes	44
1. Annotation des génomes	44
2. Annotation structurale	44
2.1. Les méthodes <i>Ab initio</i>	45
2.2. Les méthodes <i>par homologie de séquence</i>	47
VI - Comparaison de séquences protéiques	48
1. L'annotation fonctionnelle	48
1.1. <i>Annotation fonctionnelle expérimentale</i>	48
1.2. <i>L'annotation fonctionnelle in silico</i>	48
VII - Les mutation et les mécanismes de réparation de l'ADN	53
1. Introduction	53
2. Les mutations et les lésions de l'ADN	53
3. Les mutations ponctuelles.....	54
3.1. <i>Mutation « silencieuse »</i>	54
3.2. <i>Mutation « faux-sens »</i>	54
3.3. <i>Mutation « non-sens » (stop)</i>	54
4. Mutations pendant la réplication de l'ADN	55
4.1. <i>Mutation par substitution</i>	55
4.2. <i>Mutations par délétion</i>	56
4.3. <i>Mutations par insertion</i>	56
5. Altération de l'ADN survenant en dehors de la réplication	57
5.1. <i>Les lésions spontanées</i>	57
5.2. <i>Mutation induites par des agents mutagènes</i>	58
6. Correction des erreurs d'appariement lors de la réplication.....	63
6.1. <i>Correction immédiate : Fonction d'édition de l'ADN polymérase</i>	63
6.2. <i>Le système MMR (Methyl mediated -Mismatch Repair system)</i>	63
7. Correction des autres mutations de l'ADN	64
7.1. <i>Réparation par réversion directe (DRR)</i>	64
7.2. <i>Réparation par excision</i>	65
8. Réparation des lésions non réparées par les systèmes précédents.....	67
8.1. <i>Réparation par recombinaison homologue</i>	67
8.2. <i>Le système SOS</i>	68
VIII - Transferts des gènes	69
1. Les transferts horizontaux d'information génétique chez les bactéries.....	69
1.1. <i>La Transformation</i>	70
1.2. <i>La transduction</i>	73
1.3. <i>La conjugaison</i>	75
IX - Les références bibliographiques	79

Objectifs



Ce cours est divisé en trois grandes parties :

- La première partie permettra aux étudiants de comprendre la structure et l'organisation du génome avec toute sa complexité (Réplication, transcription, traduction, régulation transcriptionnelle et traductionnelle).
- La deuxième partie du cours est consacrée à l'annotation des génomes (outils et base de données). Elle vise à initier les étudiants à la bioinformatique.
- La dernière partie est dédiée à l'étude de gènes et d'ADN.

Introduction générale

Définition	5
Bref historique	5
Les Acides Nucléiques	6

1. Définition

La biologie moléculaire est un domaine qui étudie la composition, la structure et les interactions de molécules cellulaires telles que les acides nucléiques et les protéines qui effectuent les processus biologiques essentiels pour les fonctions et la maintenance de la cellule.

2. Bref historique

En 1865: la notion de gène est apparue après la découverte des lois de l'hérédité, établies par « *Gregor Mendel* ».

En 1869: « *Johann Friedrich Miescher* » isolait l'ADN et les protéines associées, la " *nucléine* ", à partir des noyaux cellulaires.

En 1889: le terme *nucléine* fut remplacé par l'appellation « *acide nucléique* » sur proposition de « *Richard Altmann* ».

En 1909: le mot *Gène* fut proposé par le biologiste danois « *W.Johannsen* ».

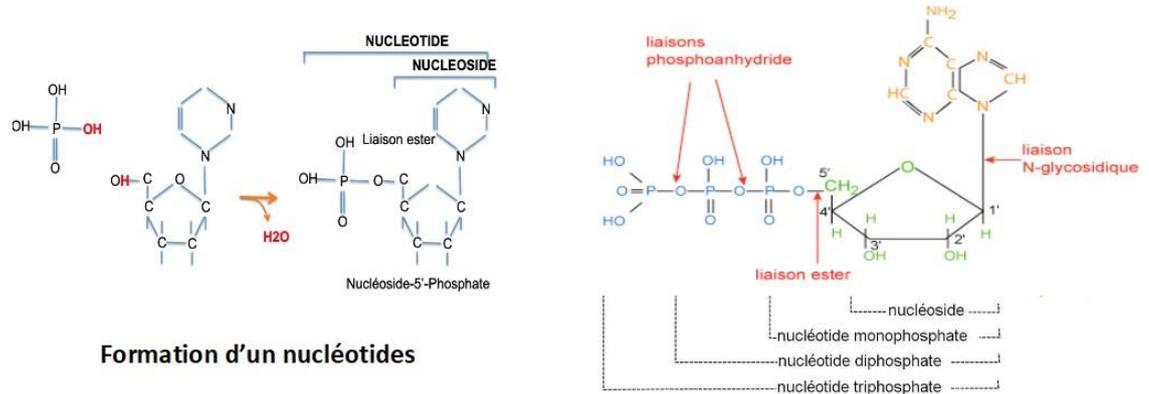
En 1910: « *Morgan* » avait expérimentalement établi que:

- Les caractères phénotypiques sont déterminés par les gènes.
- Les gènes sont localisés sur les chromosomes.
- Chaque chromosome contient un groupe déterminé de gènes.

Deux expériences fondamentales ont conduit à la découverte de l'ADN comme matériel génétique :

1. *En 1928* « *Griffith* » a mis en évidence un *principe transformant* de souche R en souche S (virulence liée à polysaccharide surface qui donne aspect smooth).
2. *En 1944* « *Avery, MacLeod et MacCarthy* » prouvent que l'agent en cause est l'ADN.

- Identique pour les nucléotides de l'ADN et de l'ARN



Formation d'un nucléotides

Base + sucre + phosphate = NUCLEOTIDE

Figure 1

3.1. L'Acide DésoxyriboNucléique (ADN)

Les acides désoxyribonucléiques sont de très grandes molécules composées de deux chaînes polynucléotidiques enroulées l'une autour de l'autre pour former une double hélice régulière (Figure 2).

L'ADN forme des pelotes microscopiques, situées :

- Dans le noyau chez les eucaryotes
- Directement dans le cytoplasme de la cellule chez les procaryotes.
- A l'intérieur de la capsid chez les Virus à ADN

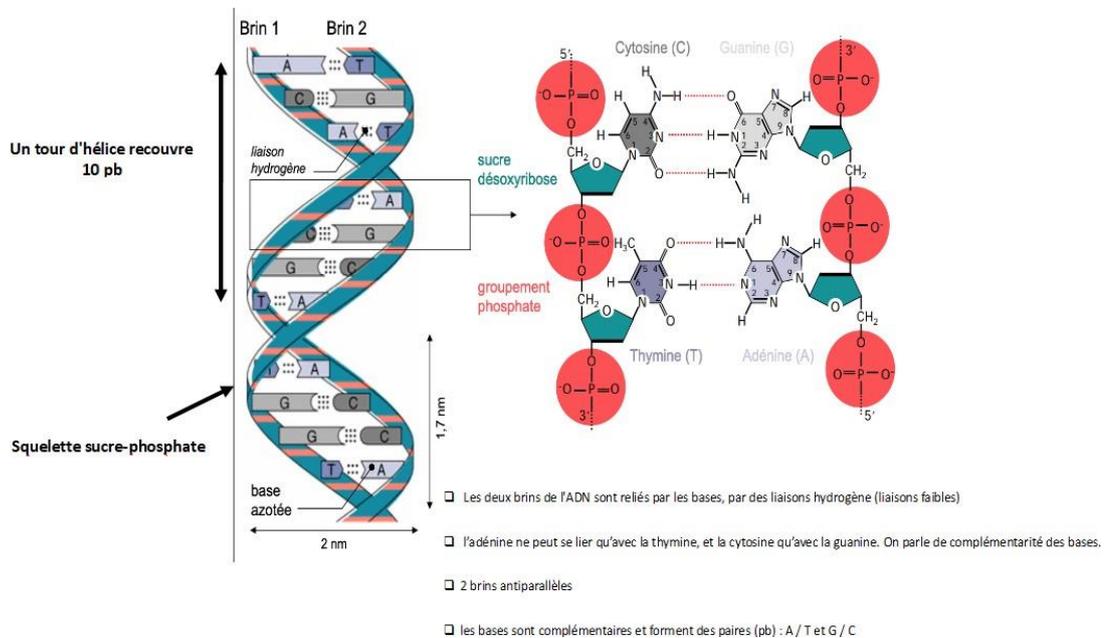


Figure 2 : Structure de l'ADN

3.2. L'Acide RiboNucléique (ARN)

- C'est un polymère linéaire de ribonucléotides liés par des liaisons phosphodiester.
- On les trouve dans le noyau et le cytoplasme.
- Un ARN est monocaténaire, orienté 5'P→3'OH mais sa chaîne peut se replier pour former une structure secondaire stable (épingle à cheveu) en formant des liaisons hydrogène entre les bases (Figure 3).

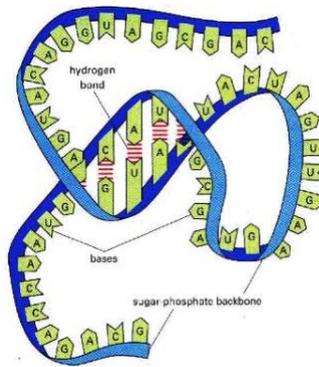


Figure 3 : Structure de l'ARN

- L'ARN est un produit de la transcription de l'ADN par l'ARN polymérase ADN dépendante.
- Il existe de nombreuses familles d'ARN (ARNr, ARNm, ARNt, ARNnc..), dont chacune possède une structure ou une fonction particulière, chacune assure une fonction particulière.

Notion de gène et transmission de l'information génétique



Introduction

9

Réplication de l'ADN

9

1. Introduction

Au cours de la division cellulaire, quand une cellule-mère donne deux cellules-filles, il est essentiel que l'ADN présent dans les cellules-filles soit la copie identique de l'ADN présent dans la cellule-mère. La transmission des caractères génétiques des parents vers leur descendance nécessite la réplication de l'ADN. La réplication assure donc le transfert de l'information génétique d'une cellule à ses descendantes.



Définition

- La *réplication* correspond à un doublement des molécules d'ADN par synthèse, qui reproduit exactement le génome d'une cellule au cours du cycle cellulaire afin de préparer la division de cette cellule.
- Le *cycle cellulaire* se définit, quant à lui, comme un processus cyclique qui permet à une cellule de se diviser.

2. Réplication de l'ADN

Dès leur publication originale sur la structure de l'ADN, « *Watson et Crick* » avaient proposé un modèle semi-conservatif de réplication. Les expériences de « *Taylor* » en 1957 et de « *Meselson et Stahl* » en 1958 ont contribué à valider cette hypothèse.



2.1. La réplication chez les procaryotes (*E.coli*)

2.1.1. Éléments nécessaires pour la réplication

- Un brin d'ADN comme brin matrice (ou parental) : chacun des deux bras sert de matrice pour synthétiser le brin antiparallèle.
- Les dNTP (dATP, dCTP, dTTP et dGTP) rentrent dans la réaction sous forme triphosphate, et sont intégrés dans le brin sous forme monophosphate.
- Des enzymes qui permettent de catalyser cette réaction (séparation des brins, incorporation des nucléotides.....).
- Les ions Mg^{2+} stabilisent les dNTP en les protégeant d'une hydrolyse enzymatique et sont également importants pour l'ADN polymérase.

2.1.2. Synthèse de l'ADN et protéines impliquées

Proteine	Role	Taille (Kda)	Molécule par cellule
Hélicase	Déroule la double hélice	300	20
Primase	Synthétise les amorces d'ARN	60	50
Protéine SSB	Stabilise les régions monocaténares	74	300
ADN gyrase	Réduit la tension due a la torsion	400	250
ADN polymérase III	Synthétise l'ADN	900	20
ADN polymérase I	Élimine l'amorce et comble les vides	103	300
ADN ligase	Reunit les extrémités des segments d'ADN répare l'ADN	74	300

Tableau 1 : Enzymes d'*E.coli* utilisées pour la réplication de l'ADN

a) Déroulement de la double hélice

- Assuré par l'hélicase (*DNA β*), enzyme permettant l'ouverture du duplex d'ADN par coupure des liaisons hydrogène.
- Le déplacement le long de l'ADN parental nécessite une énergie (hydrolyse des molécules d'ATP en ADP et P_i).
- Après l'action hélicase, des protéines se positionnent : les protéines SSB (*Single-strand DNA-binding*), qui se lient à des simples brins, pour éviter la renaturation spontanée de l'ADN (Figure 4).

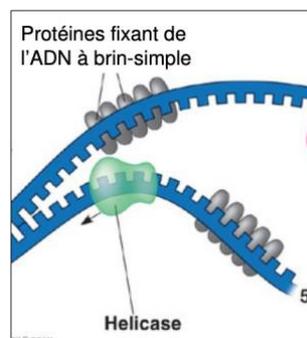


Figure 4

b) Modification du surenroulement de l'ADN

- La gyrase (une topoisomérase de type 2) agit 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 donc supprimer ces surenroulements positifs et introduire des surenroulements négatifs (passage de la forme super enroulée à une forme relâchée).
- Cette action est simultanée aux hélicases et permet une bonne accessibilité aux enzymes de la réplication à la molécule d'ADN (Figure 5).

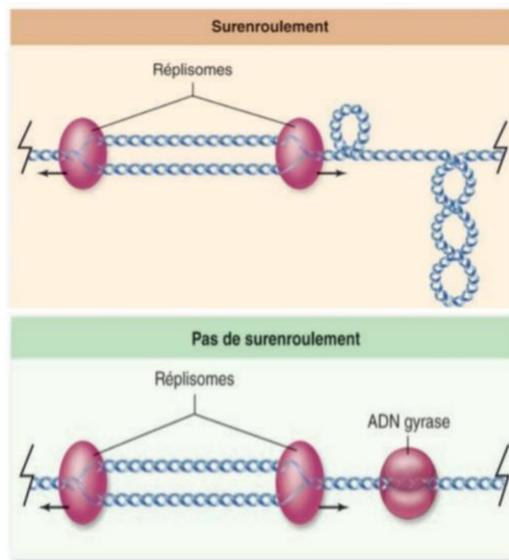


Figure5

c) Formation des nouveaux brins

- La polymérisation est assurée par l'ADN polymérase qui additionne les nucléotides (désoxyribonucléotides triphosphates ; dNTP) à une séquence préexistante.
- 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 (Figure 6).
- L'énergie nécessaire pour former cette liaison phosphodiester est fournie par la libération sous forme de pyrophosphate (ppi) des deux phosphates terminaux du nucléotide ajouté.

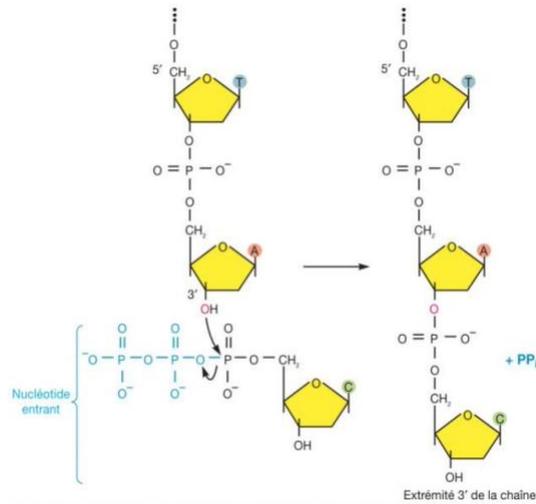


Figure 6

- Chez les procaryotes, il existe plusieurs ADN polymérases (I, II et III) (Tableau 2). Ces enzymes ont une activité polymérisique et peuvent être dotées d'activité exonucléasique. Chez *E. coli*, l'ADN polymérase III est celle qui est indispensable à la réplication.

	ADN polymérase I	ADN polymérase II	ADN polymérase III
Structure	Monomérique	> 4 sous unités	> 10 sous unités (core + clamp + protéines associées)
Rôle	Élimine les amorces lors de la réplication + réparation	Réparation de l'ADN	Réplication de l'ADN génomique
Polymérisation 5' → 3'	Oui	Oui	Oui (sous-unité α)
Exonucléase 3' → 5'	Oui	Oui	Oui (sous-unité ε)
Exonucléase 5' → 3'	Oui	Non	Non
Vitesse de polymérisation	16-20 bases / sec	5-10 bases / sec	250-1000 bases / sec

Tableau 2

- L'activité de l'ADN polymérase nécessite une initiation : démarrage par une amorce denucléotides. Pour la réplication, l'amorce est un ARN synthétisée par polymérisation par une primase (ARN polymérase) ADN-dépendante (Figure 7).

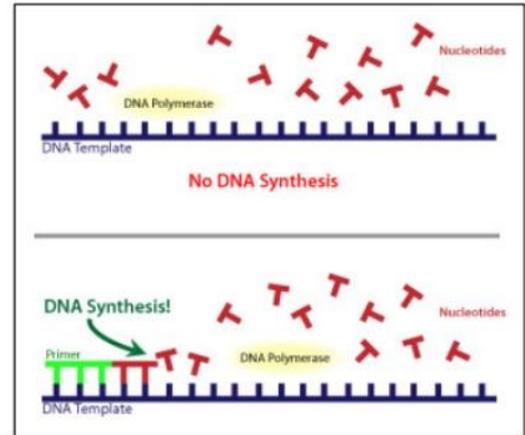
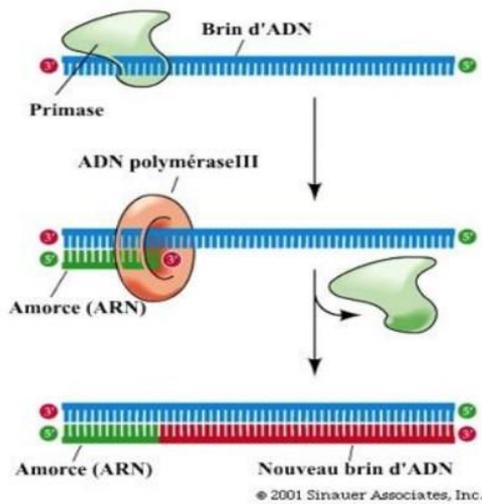


Figure 7

2.1.3. Mécanismes de réplication chez les procaryotes (E. COLI)

Le processus de réplication comprend 3 étapes principales :

1. Initiation
2. Élongation
3. Terminaison

a) Initiation

- La réplication démarre à partir d'une ouverture au niveau de la double hélice d'ADN. Il s'agit de l'origine de la réplication « *ori C* » (Figure 8).

- La synthèse de l'ADN à partir d'une origine de réplication est bidirectionnelle : les ADN polymérases vont en sens inverse à partir d'une origine commune, ce qui crée 2 fourches de réplication (Figure 8).

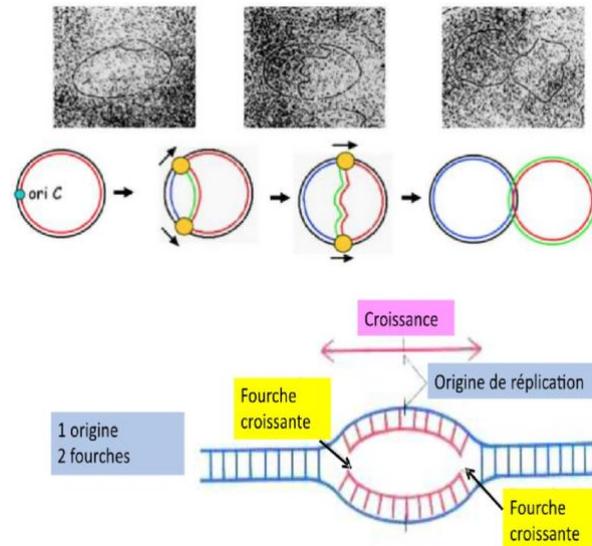


Figure 8 : origine de réplication

- Un premier complexe est formé, dans lequel 10 à 20 monomères de *DnaA* sont fixés à la séquence *oriC* (Figure 9).
- La formation de ce complexe permet l'ouverture de la double hélice faisant apparaître les ébauches des 2 fourches de réplication et l'entrée de l'hélicase (*DnaB*).
- Aidée par la protéine *DnaC* (chargeur d'hélicase), l'hélicase va se fixer sur l'origine de réplication et ouvrir la double chaîne dans les deux directions, formant un complexe de préinitiation.
- En présence de *SSB*, protéine fixant l'ADN simple brin qui stabilise les régions dénaturées, de la gyrase, de la primase (qui synthétise une amorce dont l'extrémité 3' OH est libre et sera reconnue par l'ADN pol) et de la polymérase (Pol III), la réplication du chromosome peut commencer (Figure 9).

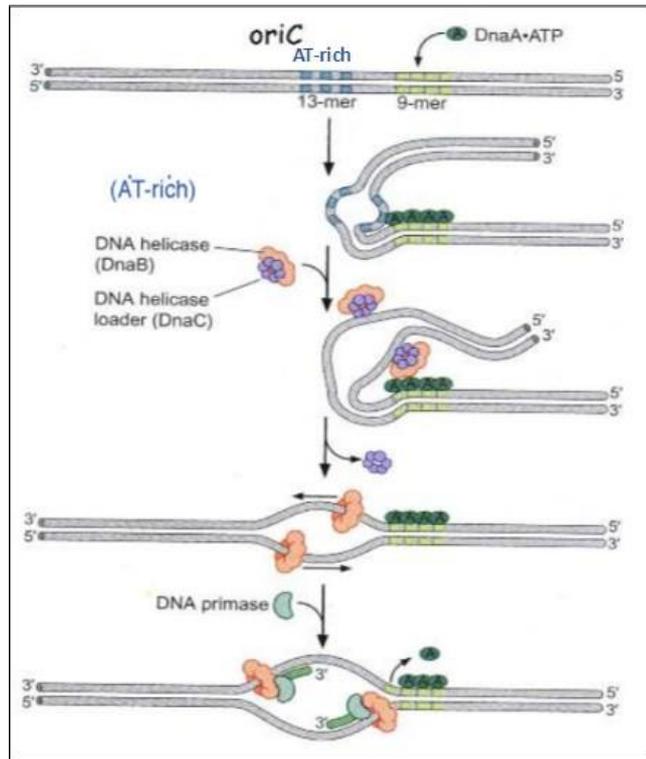


Figure 9 : L'initiation de la réplication

b) L'élongation

- La réplication est continue pour un brin, il est dit brin précoce ou avancé et discontinue pour l'autre brin, il est dit brin retardé. En effet, la synthèse d'un brin d'ADN n'est possible que dans le sens 5' → 3'. De ce fait les nucléotides peuvent se lier, de façon continue au niveau de l'extrémité libre 3' de l'un des 2 brins parentales.
- Par contre, pour l'autre brin parental les nucléotides ne peuvent pas se lier à l'extrémité 5' et la synthèse se fait par addition successive de petits fragments d'ADN, appelés fragments d'OKAZAKI.

Au cours de l'élongation :

1. La synthèse se fait par l'ADN pol III jusqu'à rencontrer l'amorce suivante.
2. L'ADN pol I remplace ensuite l'amorce d'ARN en continuant la polymérisation du brin complémentaire.
3. L'ADN ligase rattache les fragments d'Okazaki : elle est capable de catalyser la formation la liaison phosphodiester entre le carbone 3'-OH et le phosphate-5' de deux nucléotides voisins sur un brin d'ADN (Figure10).

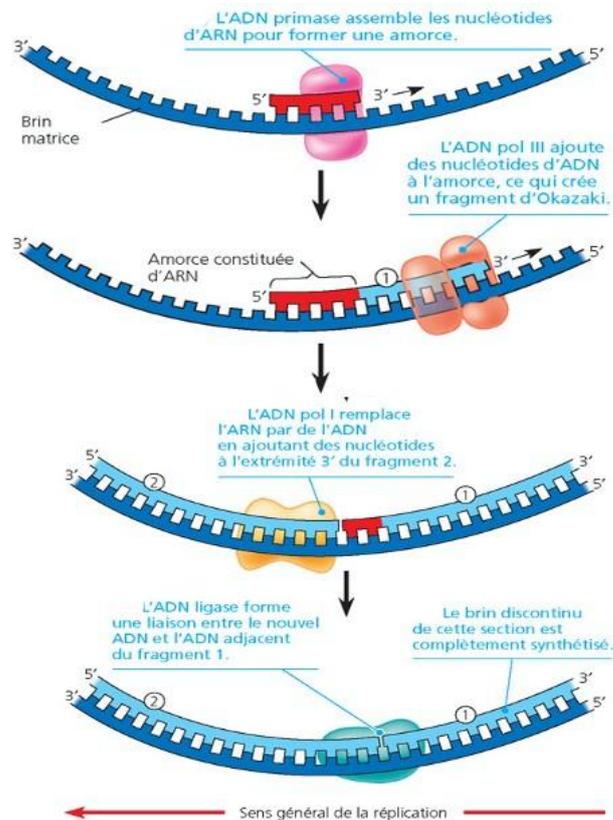


Figure 10 : Les étapes franchies dans la synthèse de chaque fragment d'ADN sur la chaîne tardive

c) La terminaison

- La terminaison a lieu lors de la rencontre des deux fourches. Elle se fait au niveau d'une séquence *TER* située à l'opposé de l'origine de réplication reconnue par la protéine *Tus*. Le complexe *Ter-Tus* bloque les fourches, mettant fin à la réplication (Figure 11).

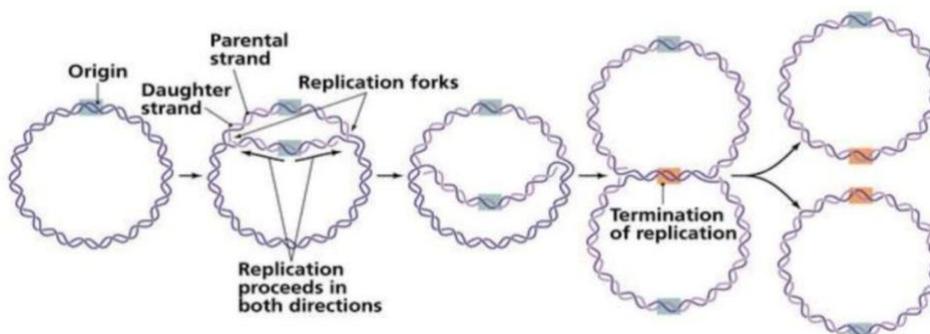


Figure 11

2.2. La réplication chez les eucaryotes

- La réplication se fait en de nombreux points d'initiation (plusieurs origines).
- La présence de nucléosomes : Pendant la phase S du cycle cellulaire, de nouveaux nucléosomes sont synthétisés pour compléter les nucléosomes parentaux dans les cellules filles.

- Elle fait intervenir un nombre d'ADN polymérases plus important que chez les procaryotes (9 ADN polymérases chez les eucaryotes).
- La présence des télomères : la fin de la réplication est marquée par l'addition, par des enzymes spécifiques, les télomérases, de séquences répétées qui forment les télomères (sont indispensables pour préserver l'intégrité du matériel génétique au cours du cycle cellulaire).
- La vitesse de réplication chez les eucaryotes est plus faible.

Structure et fonction des gènes

Structure des gènes

18

Les grandes étapes de l'utilisation de l'information génétique

19

1. Structure des gènes

1.1. Structure des gènes procaryotes

- Chez les procaryotes, les gènes sont constitués d'une seule pièce, et sont fréquemment organisés en opérons.

Définition : Les opérons

- Un opéron est un ensemble de gènes (chacun entouré d'un codon start et d'un codon stop) ayant vocation à fonctionner de manière coordonnée de façon à produire les protéines répondant à une voie physiologique bien intégrée, ainsi qu'à apporter les mécanismes de régulation de cette voie (Figure 12).

- Les protéines produites sont souvent impliquées dans un même processus métabolique, ou peuvent constituer les différentes sous-unités d'une protéine multimérique qui présentera la fonction biologique requise.

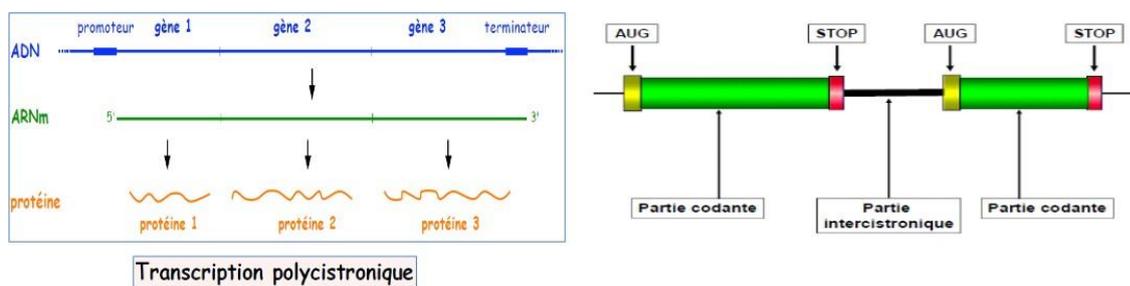


Figure 12

1.2. Structure des gènes eucaryotes

- Selon le type d'ARN, on trouve chez les eucaryotes plusieurs types de gènes. Chacun de ces gènes à une organisation différente. Les gènes à l'origine de la synthèse protéique présentent une structure

morcelée : ils sont constitués de régions non codantes parfois très longues, les introns, qui alternent avec des portions effectivement traduites en protéines, les exons (Figure 13).

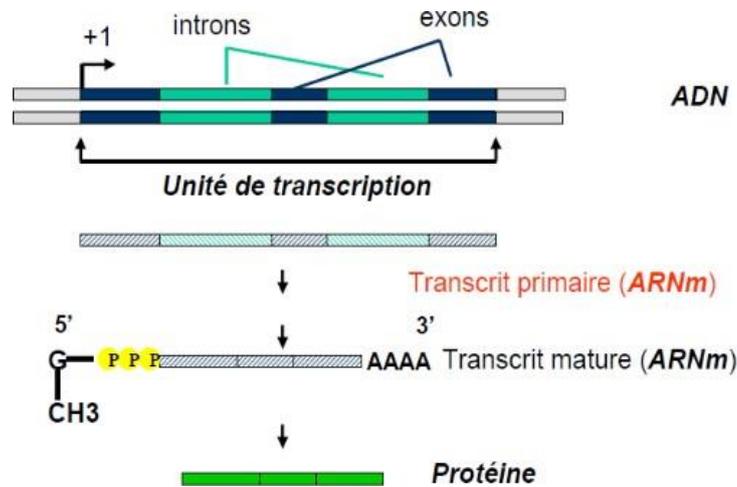


Figure 13

2. Les grandes étapes de l'utilisation de l'information génétique

- L'information génétique est stockée dans l'ADN. Cependant, l'expression de cette information génétique nécessite le passage de l'ADN à l'ARN, puis aux protéines.
- Elle se réalise en deux étapes (Figure 14) :
 1. La transcription, qui permet d'obtenir de l'ARNm.
 2. La traduction, qui permet d'obtenir les protéines.



Figure 14

2.1. La Transcription

- La transcription est le premier processus de régulation utilisé par les cellules, tissus et organismes pour faciliter et contrôler les programmes complexes de l'expression génétique.
- On appelle transcription la synthèse de brin d'ARN dont la séquence est dictée par celle de la molécule d'ADN correspondante.
- Contrairement à la réplication qui intéresse la totalité du génome le programme de transcription n'est pas fixe : seules, de petites portions du génome sont transcrites à une époque donnée de la vie de la cellule et ces portions varient en fonction du développement, de l'environnement etc...

- La transcription commence donc en un point précis de l'ADN pour se terminer en un point également précis, l'espace entre les deux constitue une unité de transcription.

Remarque

- La transcription chez les procaryotes se fait au niveau du cytoplasme, elle est *polycistronique* car plusieurs gènes peuvent être transcrits par la même polymérase (exemple : L'opéron lactose) (Figure 15).

- Chez les eucaryotes, celle-ci se fait au niveau du noyau, et une polymérase transcrit un seul gène, on dit qu'elle est *monocistronique* (Figure 15).

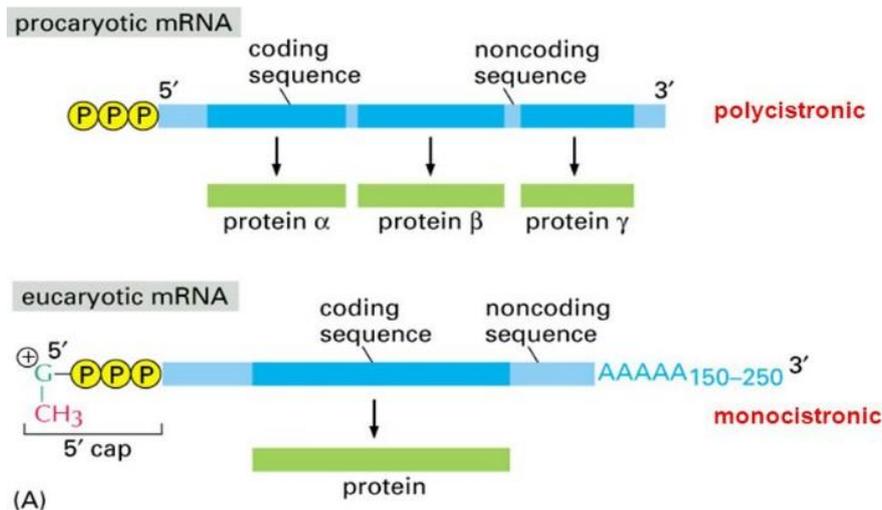


Figure 15

2.1.1. Caractéristiques générales

- Comme pour la réplication, la chaîne d'ARN est toujours synthétisée dans le sens 5' → 3'.
- Chaque nouveau nucléotide est ajouté à l'extrémité 3' de la chaîne en cours de synthèse de façon complémentaire selon les règles d'appariement des bases (Figure 16).
- La chaîne d'ARN synthétisée est :
 1. Identique à un brin codant ou brin complémentaire.
 2. Complémentaire de l'autre brin d'ADN, le brin matrice.

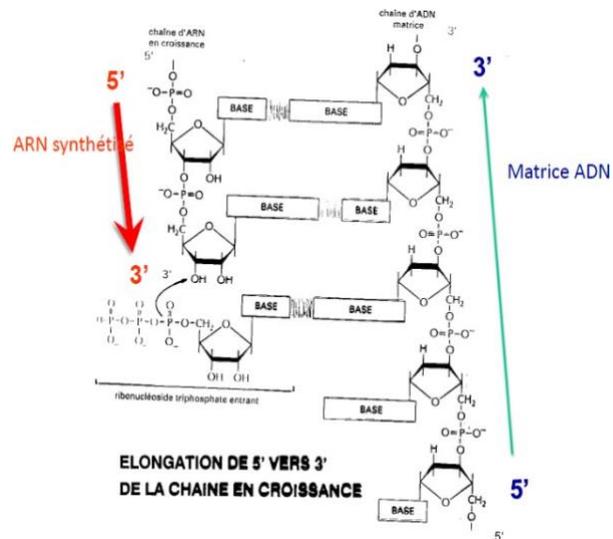


Figure 16

2.1.2. Les Mécanismes de la Transcription

- Ne concerne qu'une portion de l'ADN :
 1. Démarre à un Promoteur
 2. S'arrête à un terminateur (pas chez les eucaryotes)
- Elle se déroule en trois étapes :
 1. Initiation
 2. Élongation
 3. Terminaison
- Enzyme responsable : *L'ARN polymérase*.

2.2. La transcription chez les procaryotes

2.2.1. Structure de l'ARN polymérase

- La transcription s'effectue chez les bactéries grâce à une enzyme, l'ARN polymérase.
- L'ARN polymérase bactérienne ou holoenzyme (500kDa) est une enzyme multimérique composée de 5 sous-unités $\alpha 2\beta\beta'\sigma$ (Figure 17).
- Elle se charge de la synthèse d'ARNt, ARNr ou ARNm indifféremment.

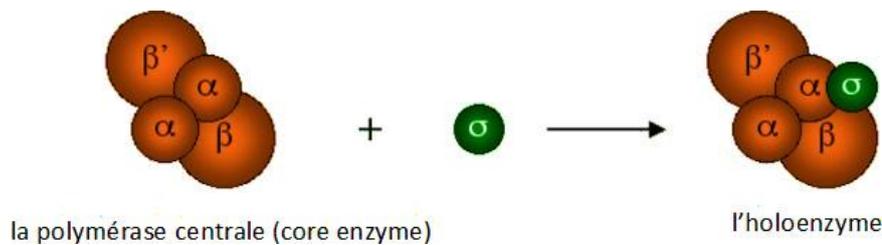


Figure 17 : Structure de l'ARN polymérase

2.2.2. Les particularités des ARN polymérases

- Elles n'ont pas besoin d'amorces pour fonctionner.
- La polymérisation se fait dans le sens $5' \rightarrow 3'$, donc la lecture du brin d'ADN se fait dans le sens $3' \rightarrow 5'$.
- L'ARN ne reste pas apparié au brin d'ADN pendant la synthèse.
- Les ARN polymérases n'ont pas d'activité de relecture (activité exonucléasique $3' \rightarrow 5'$) le taux d'erreur est ainsi plus important que pour les ADN-polymérases, mais ce taux est supporté.
- Elle assume de multiples fonctions dans le processus de la transcription :
 1. Elle cherche les promoteurs.
 2. Elle déroule le fragment à transcrire pour produire une matrice simple brin.
 3. Elle sélectionne les NTPs corrects et catalyse la formation des liaisons phosphodiester.
 4. Elle détecte les signaux de terminaison.

2.2.3. Étapes de la transcription chez les procaryotes

- La synthèse de l'ARN, comme presque toutes les réactions biologiques de polymérisation, comprend trois étapes : l'initiation, l'élongation et la terminaison.

a) Initiation

- Cette étape initiale est appelée "fixation sur le brin matrice".

- Chez les bactéries, cette fixation est établie quand la sous unité σ de l'ARN polymérase reconnaît le promoteur localisé dans la région 5' (en amont) du point de transcription initiale d'un gène (Figure 18).

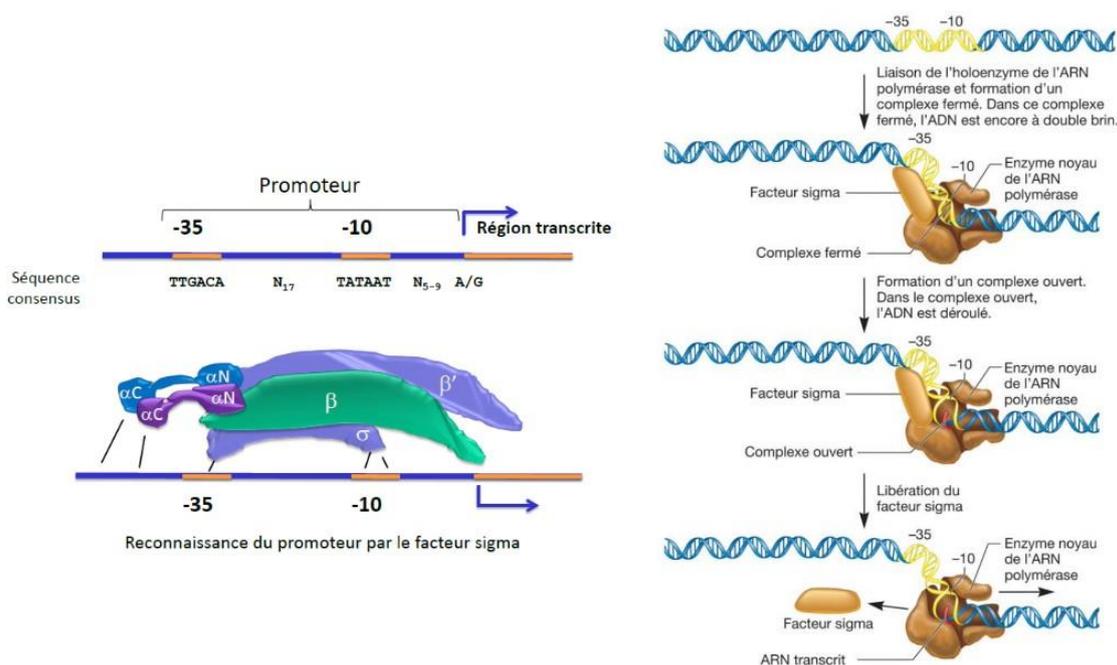


Figure 18 : Interaction entre l'ARN polymérase et le promoteur bactérien

i Séquences consensus

- Ces séquences présentent des homologies avec celles des autres gènes du même organisme ou d'un ou plusieurs gènes d'organismes apparentés. Leur conservation au cours de l'évolution atteste la nature critique de leur rôle dans les processus biologiques.
- Dans les promoteurs bactériens deux séquences ont été détectées (Figure 19) :
- **TATAAT** : Localisée 10 nucléotides en amont (upstream) du site d'initiation de la transcription (région -10, ou Pribnow Box).
- **TTGACA** : Localisée 35 nucléotides en amont du site d'initiation de la transcription (région -35).
- Ces séquences sont appelées "éléments agissant en cis" (sur le même côté de la molécule d'ADN dans le gène lui-même). Contrairement aux "facteurs agissant en trans" qui sont des molécules qui se fixent à ces éléments d'ADN.

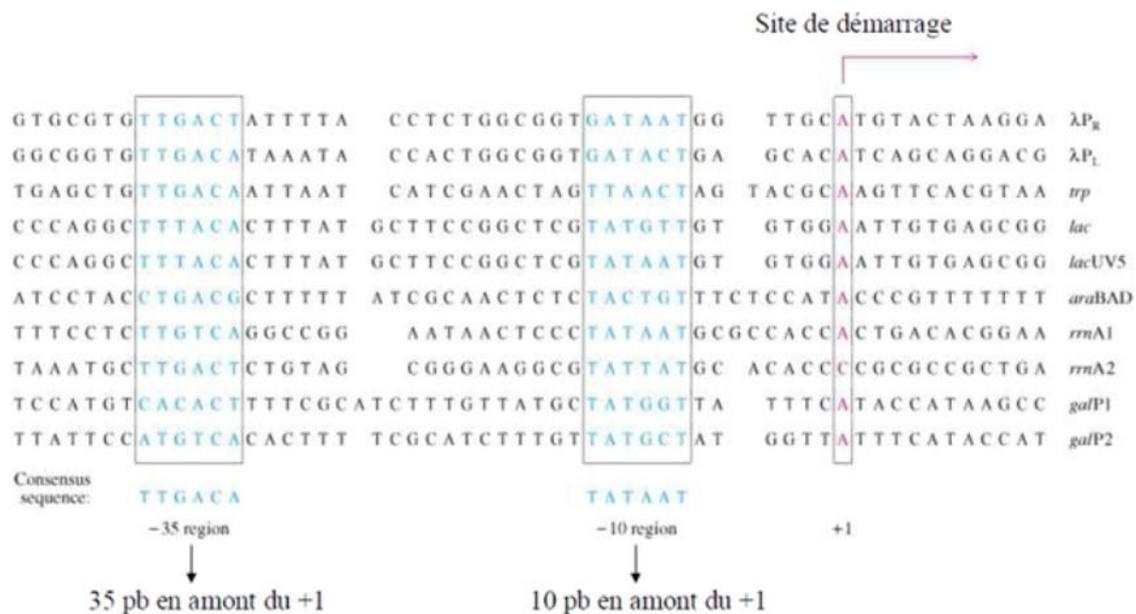


Figure 19

b) Élongation de la chaîne d'ARN

- La boucle de transcription se déplace dans le sens 3'→5' du brin matriciel et la chaîne d'ARNm s'allonge dans le sens 5'→3' (Figure 20).
- L'élongation de la molécule d'ARNm se fait par l'appariement des bases complémentaires et par l'addition successive de nucléotides 5' triphosphates.
- L'ADN lu se rembobine immédiatement après la lecture.

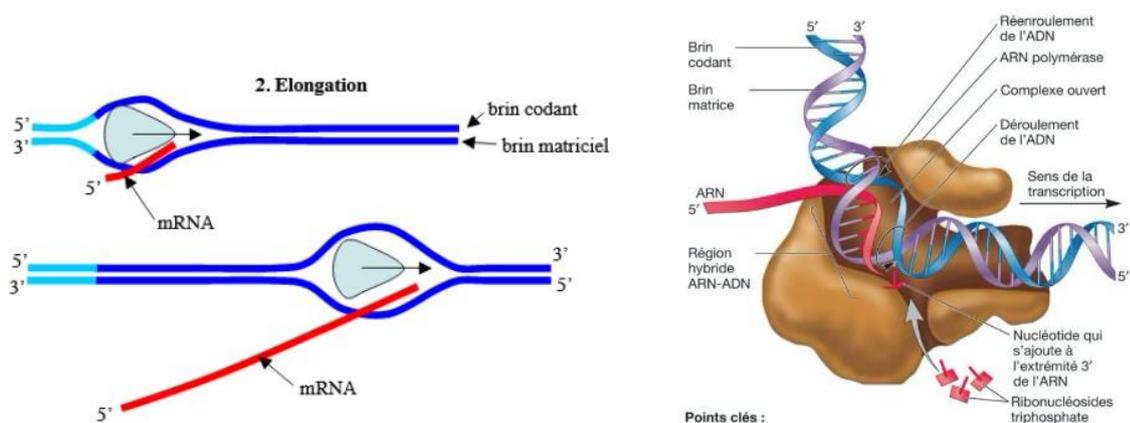


Figure 20 : L'élongation de la chaîne d'ARN

c) La Terminaison

- Définie comme étant le processus conduisant à la dissociation du complexe ARN polymérase de l'extrémité 3' du gène.

- Deux mécanismes :
 1. Terminaison "rho-indépendante"
 2. Terminaison "rho-dépendante"

i Terminaison rho-indépendante (intrinsèque)

- Le site spécifique de terminaison est constitué de 3 segments caractéristiques : deux séquences répétées inversées particulièrement riches en G et C, séparées par un court segment.
- Cette région palindromique est terminée par un segment de bases répétées.
- Une série de 6 à 8 bases A sur le brin matrice codant pour un poly-U (région de faible énergie)
- C'est la transcription de cette séquence par l'ARN polymérase qui conduit à la terminaison de la transcription car elle aboutit à la formation d'une structure en épingle à cheveux suivie d'une succession de 6 résidus U (Figure 21).
- La structure en épingle à cheveux de l'ARN est formée par l'appariement de séquences répétées inversées (qui forment la tige) séparées par une courte séquence (qui forme la boucle).
- La séquence poly-U entraîne la dissociation de l'enzyme de la matrice pendant qu'elle est arrêtée par l'épingle à cheveux (Figure 21).

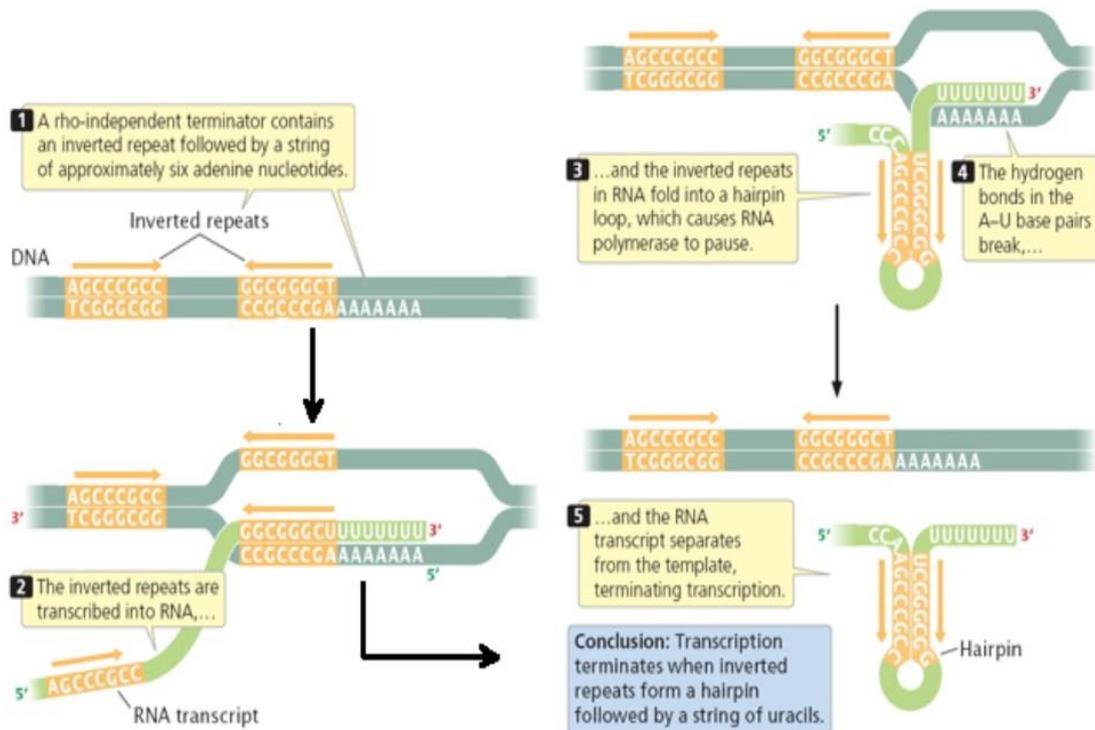


Figure 21

ii La Terminaison Rho-dépendante

- Une protéine homo-hexamérique, l'hélicase Rho, reconnaît et fixe une région du transcrit, appelé rut (rho utilization site), puis transloque de 5' vers 3', rejoint l'ARN polymérase arrêtée (ou déstabilisée) au

niveau d'un site de pause et dissocie le complexe d'élongation (Figure 22).

- La progression de Rho le long du transcrit (nécessite hydrolyse ATP).

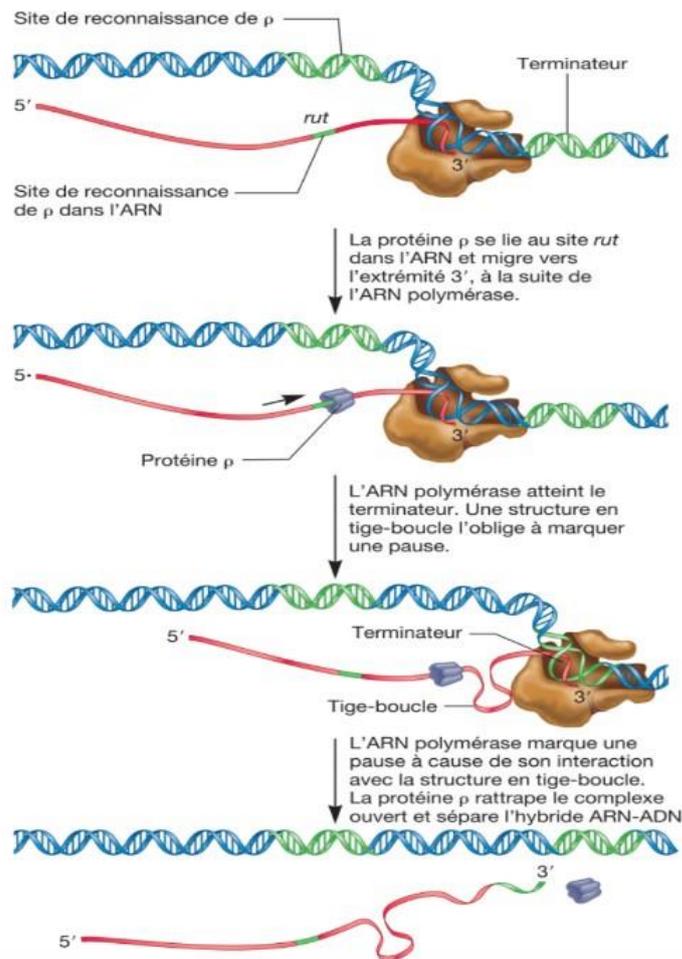


Figure 22

d) Visualisation de la transcription

- Le processus de la transcription peut être visualiser par microscopie électronique (première fois en 1970) (Figure 23).

- Les molécules d'ADN : filaments.
- Les molécules d'ARN : Branches.

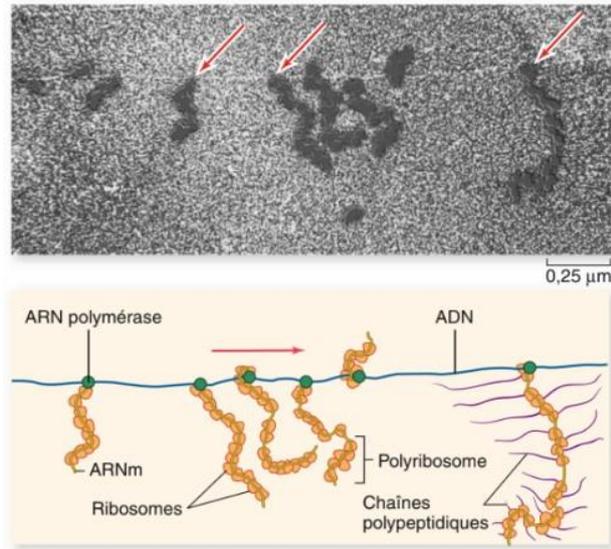


Figure 23 : Visualisation de la transcription

2.3. La transcription chez les eucaryotes

- C'est un processus beaucoup plus complexe et il existe plusieurs différences notables : La transcription au contraire de la traduction se fait au niveau du noyau (Figure 24).

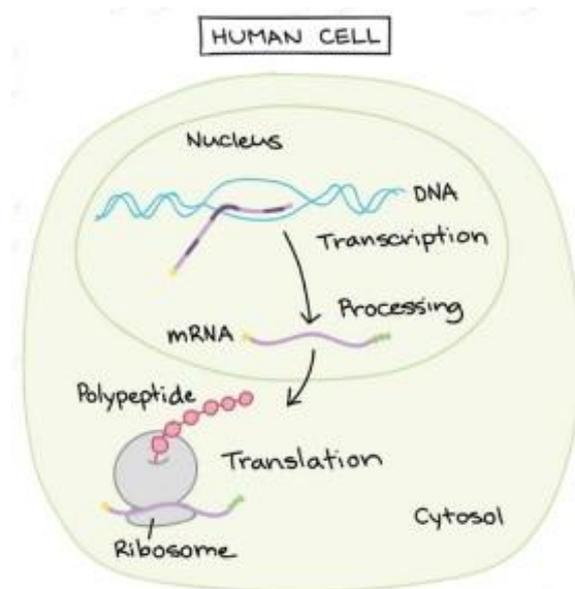


Figure 24

- La TATA box (= boîte de Hogness) 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 (Figure 25).

Eukaryotic promoter site:

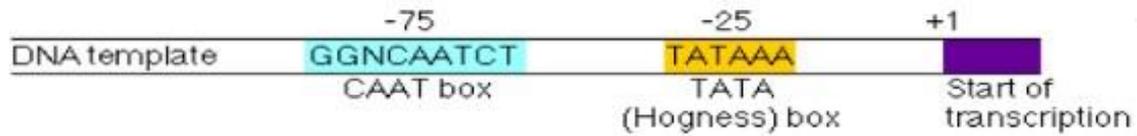


Figure 25

- La transcription chez les eucaryotes est prise en charge par 3 ARN polymérase (Figure 26)

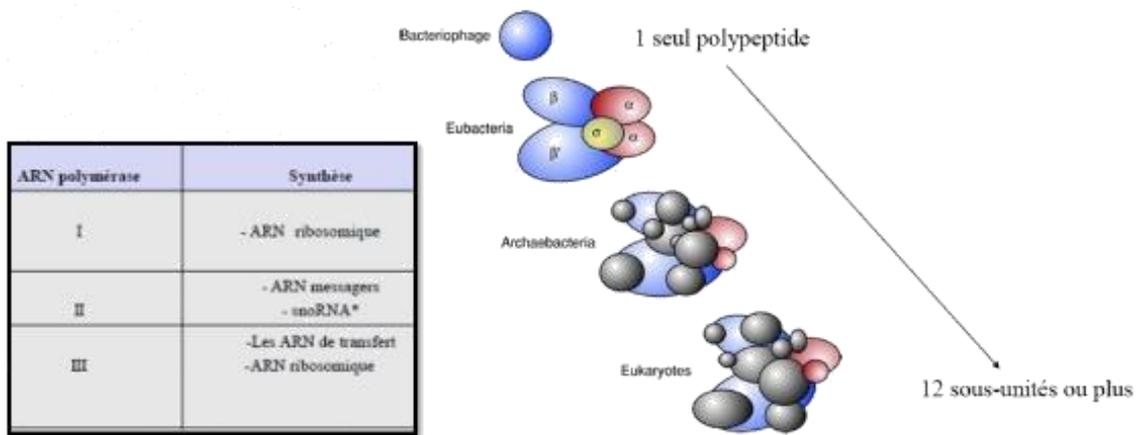


Figure 26

- Elle nécessite plusieurs facteurs d'initiation (facteurs généraux de transcription (FGT : TFII A, TFII B, TFII D (TBP, TAF), TFII E, TFIIH).

2.3.1. L'initiation de la transcription

- Certains de ces facteurs ont un rôle particulier :

- Le facteur TFII D : Sa sous-unité TBP (protéine de liaison à la TATA box = TATA binding protein) 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 (Figure 27).

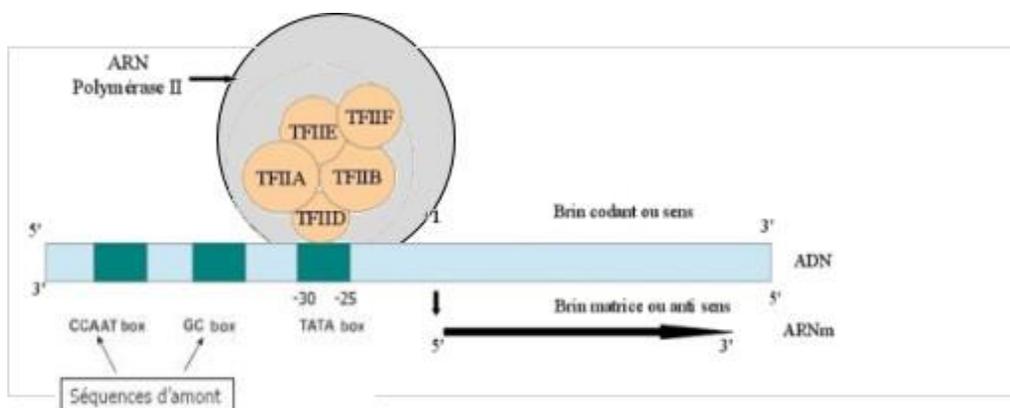


Figure 27 : Le complexe d'initiation eucaryote

- Les promoteurs peuvent être stimulés par des enhancers (séquences activatrices) qui sont des séquences définies par le fait qu'elles :

- Augmentent considérablement le taux de transcription du gène auquel elles sont associées.
- Peuvent être localisées en 5' ou en 3' ou dans un intron du gène.
- Des séquences identiques mais à effet opposé ont été mises en évidence : elles sont appelées Silencers ou séquences extinctrices (Figure 28).

Rôle des séquences cis-régulatrices

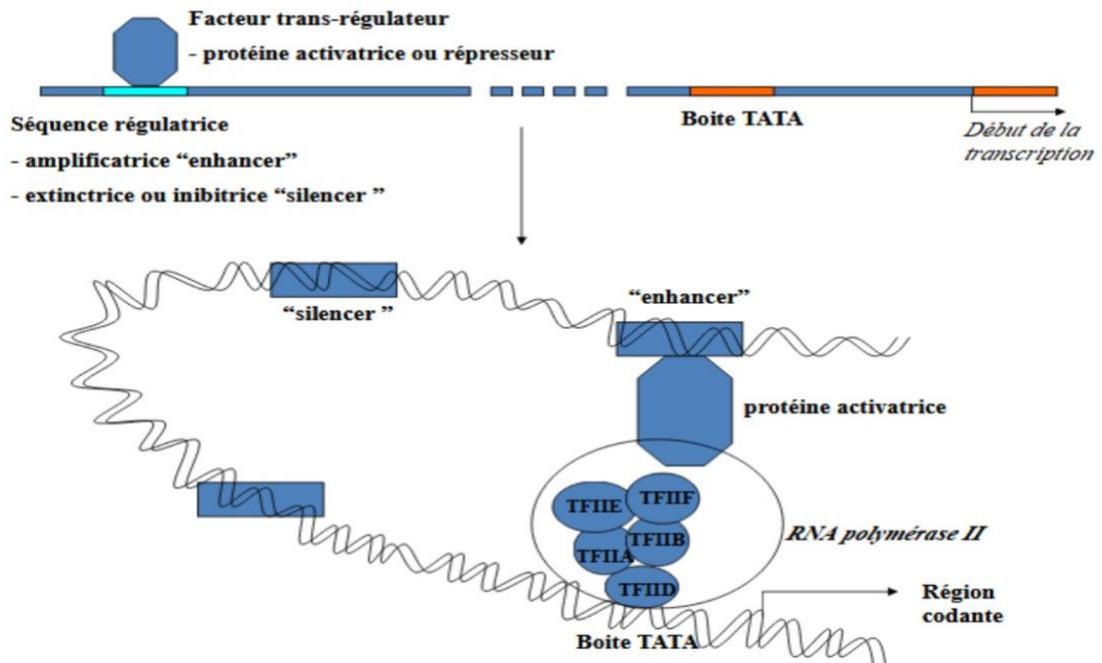


Figure 28

2.3.2. L'élongation de la transcription

- Dans l'étape de l'élongation la Pol II se débarrasse de la plupart de ses facteurs d'initiation de nouveaux facteurs appelés facteurs d'élongation vont les remplacer (des facteurs protéiques d'élongation), qui permettent le maintien le plus longtemps possible de l'ARN polymérase II sur l'ADN matrice).

- Les topoisomérases permettent l'élimination des surenroulements positifs créés par l'avancée de l'ARN pol II.

2.3.3. La terminaison de la transcription

- L'ARN pol II est également équipée de facteurs protéiques de terminaison, et reconnaît ainsi un ou plusieurs signaux de terminaison portés par le brin progressivement parcouru et qui annoncent la fin de la transcription sur le brin d'ADN matrice (TTATTT).

- Elle arrête bientôt son travail de transcription et libère Pré-ARNm qu'elle vient d'assembler.

2.3.4. Maturation du pré-ARNm en ARNm mature propre à la traduction

a) Modification des extrémités

Le transcrit reçoit une coiffe (coiffe de méthylguanosine) à l'extrémité 5' et une queue poly (A) à l'extrémité 3' (polyadénylation) (Figure 29).

1- Rôle de la coiffe de méthylguanosine :

- Transport de l'ARNm dans le cytoplasme
- Protection contre les nucléases
- Rôle important dans l'initiation de la traduction

2- Rôle dans la queue poly (A) :

- Stabilisation de l'ARNm (évite la dégradation)

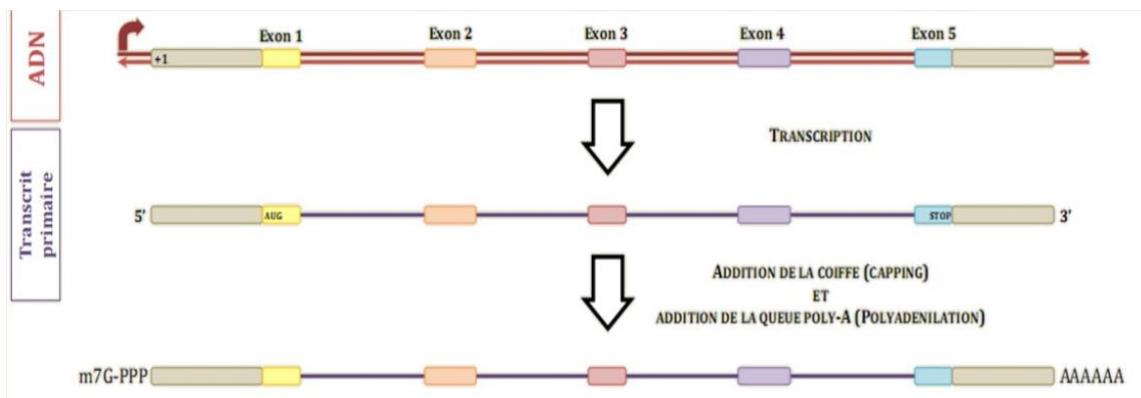


Figure 29

b) Épissage (élimination de séquences non-codantes)

- Séquences codantes (conservées) = exons
- Séquences non codantes (non conservées) = introns
- Rupture de l'ARN aux bordures 5' et 3' de chaque intron = site d'épissage
- Ligature des exons situés de part et d'autre des zones excisées.

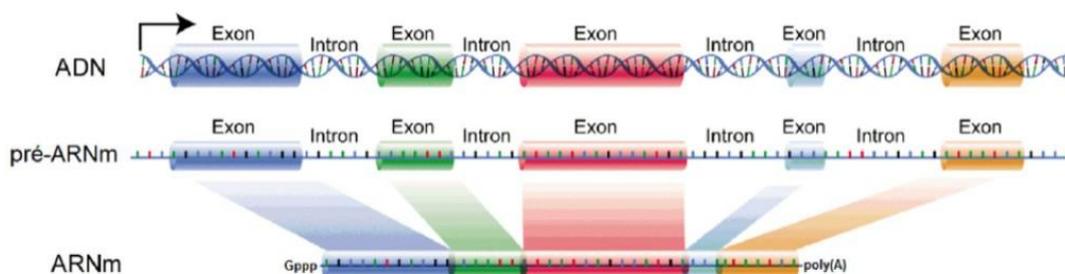


Figure 30

2.4. La traduction chez les procaryotes

La traduction correspond à la synthèse des protéines = conversion de séquence nucléotidique de l'ARN en une chaîne d'acides aminés qui constituent un peptide puis une protéine fonctionnelle.

Passage d'un code à 4 lettres (nucléotides) à un code à 20 lettres (acides aminés) = traducteur = Code Génétique

2.4.1. Code génétique universel

- Le code génétique est le système de correspondance permettant au message génétique (acides nucléiques) d'être traduit en protéine par une cellule.
- Il y a 64 combinaisons codon-acide aminé possibles pour 20 acides aminés seulement (Figure 31).
- Chez les procaryotes, le codon d'initiation est AUG mais peut parfois être remplacé par le codon GUG et plus rarement UUG.
- La redondance minimise les effets néfastes des mutations. Par exemple, il arrive qu'un changement dans l'ADN entraîne la mise en place d'un autre codon mais que ce dernier code pour le même acide aminé que le codon d'origine. La mutation est alors dite « silencieuse ».

		Second base				Third base (3' end)
		U	C	A	G	
First base (5' end)	U	UUU Phe UUC UUA Leu UUG	UCU UCC Ser UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG	U C A G
	C	CUU CUC Leu CUA CUG	CCU CCC Pro CCA CCG	CAU His CAC CAA Gln CAG	CGU CGC Arg CGA CGG	U C A G
	A	AUU AUC Ile AUA AUG Met or start	ACU ACC Thr ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
	G	GUU GUC Val GUA GUG	GCU GCC Ala GCA GCG	GAU Asp GAC GAA Glu GAG	GGU GGC Gly GGA GGG	U C A G

Figure 31

2.4.2. Les principaux composants qui sont impliqués dans la traduction

1- ARNm : Fournit l'information qui doit être interprétée par la machinerie de la traduction et constitue la matrice pour la traduction

2- *L'ARNt* : joue le rôle d'adaptateur entre les codons et les acides aminés qu'ils spécifient.

3- *L'ARNr* : en plus de son rôle structural des ribosomes en association avec des protéines, il a un rôle dans la fixation de l'ARNm (ARNr16S chez les procaryotes) et la formation des liaisons peptidique (activité peptidyl transférase de l'ARNr 23S chez les procaryotes).

4- *Ribosomes* : Le ribosome coordonne la reconnaissance de l'ARNm par chaque ARNt et catalyse la formation de la liaison peptidique entre la chaîne polypeptidique en croissance et l'acide aminé chargé sur l'ARNt sélectionné (Figure 32).

- Site A (pour aminoacyl-ARNt) qui est le site d'entrée de l'ARNt chargé à l'AA.
- Site P (pour peptidyl-ARNt) qui est le site où se fait le transfert de l'AA sur la chaîne polypeptidique naissante.
- Site E = site de sortie des ARNt (exit).

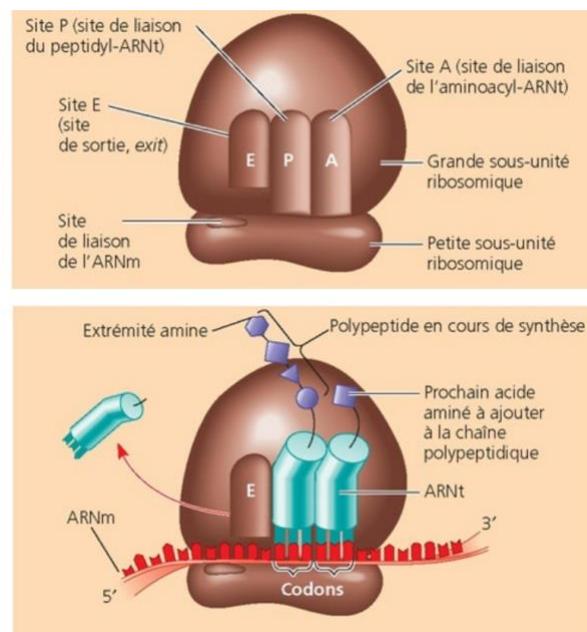


Figure 32

2.4.3. L'initiation de la traduction

Pour que la traduction soit initiée avec succès, trois événements doivent se produire :

A- La petite sous-unité 30S reconnaît le site de liaison à l'ARNm : la séquence de *Shine Dalgarno* (Figure33).

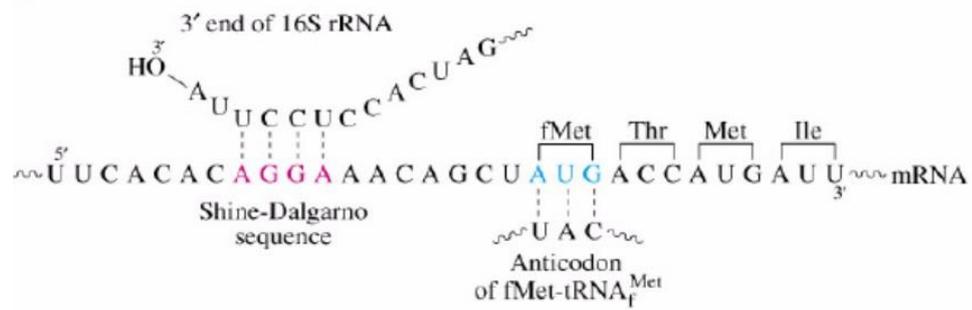


Figure 33

B- Ensuite l'ARNt chargé à la méthionine : le N-Formyl-méthionyl ARNt se fixe sur l'AUG, au niveau du site P du ribosome (Figure 34).

C- La grosse sous-unité se positionne sur la petite pour former le complexe de traduction (Figure 34).

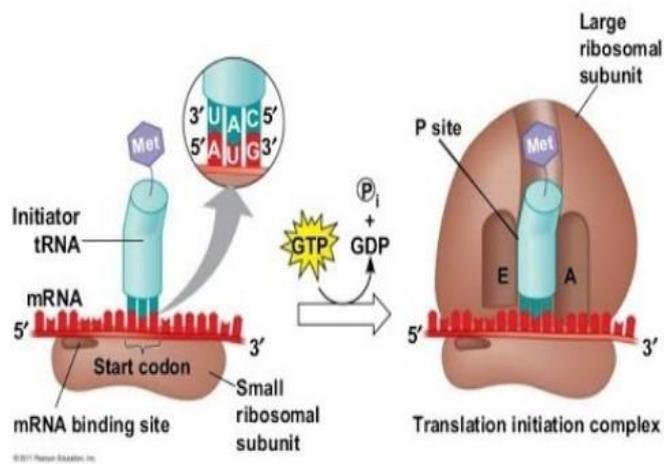


Figure 34

Les facteurs impliqués dans l'initiation de la traduction chez les procaryotes sont résumés dans la figure 35.

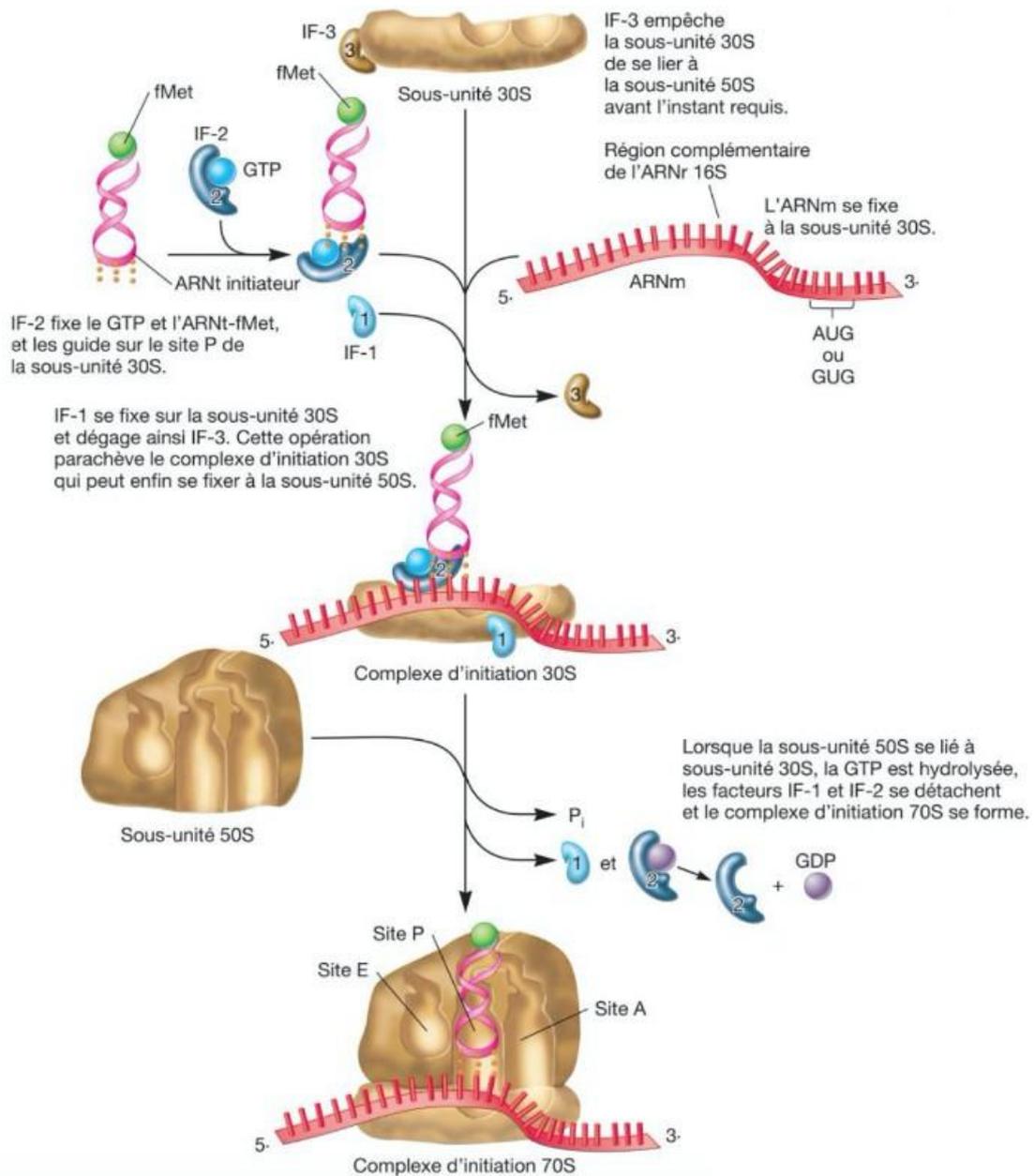


Figure 35

2.4.4. Élongation

Le ribosome se déplace de codon à codon (Figure 36).

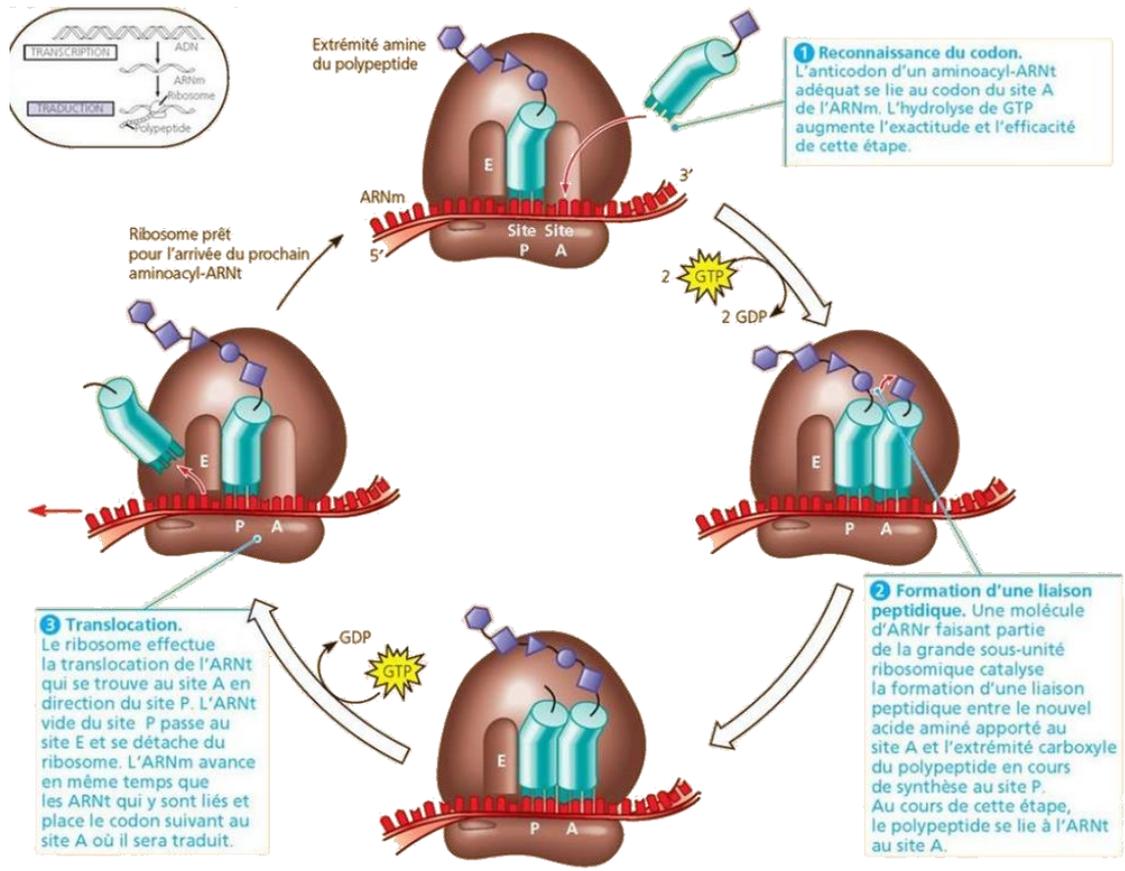


Figure 36

2.4.5. La terminaison

- À la rencontre d'un codon stop Il n'existe pas d'ARNt portant d'anticodon correspondant (Figure 37).
- Un facteur de terminaison se fixe sur le site A à la place d'un ARNt, ce qui provoque la libération de la protéine, et la dissociation du ribosome (Figure 37).
- Les facteurs de terminaison :
 1. RF1 reconnaît les codons UAG et UAA au site A
 2. RF2 reconnaît les codons UGA et UAA au site A
 3. RF3 éjecte RF1 et RF2 du site A en hydrolysant 1 GTP

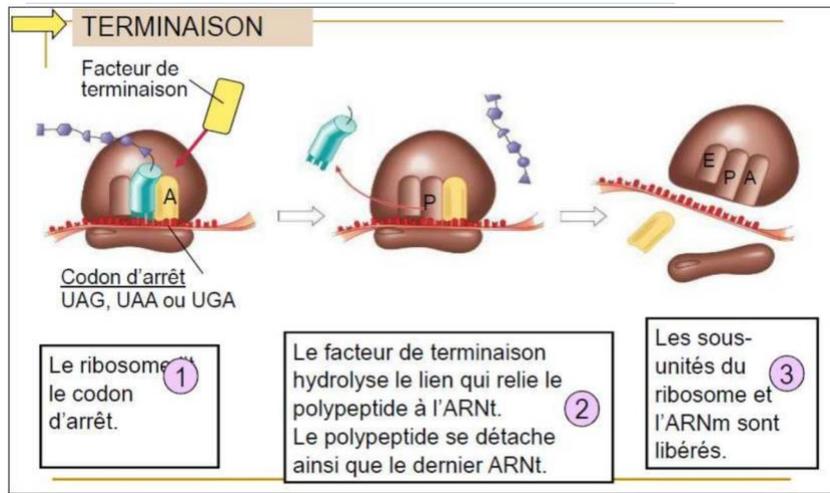


Figure 37

2.5. Les spécificités de la traduction eucaryote

2.5.1. Initiation de la traduction

- Le ribosome est de taille différente et composé d'ARN ribosomiques différents bien que la structure générale et l'activité soit comparable.
- Le ribosome recruté au niveau de la coiffe 5' de l'ARNm scanne l'ARNm dans le sens 5' - 3' jusqu'à atteindre la séquence de Kozak 5'CCAUGG3', cette séquence permet de connaître le codon d'initiation (AUG).
- Chez les eucaryotes le premier acide aminé est la méthionine et non pas la f-Met présent chez les procaryotes.
- Les facteurs d'initiation sont du type eIF (pour eukaryotic Initiation Factor), d'eIF1 à eIF6.

2.5.2. Élongation

Une fois le ribosome assemblé, avec l'ARNt initiateur chargé au site P, la synthèse du polypeptide peut commencer.

2.5.3. Terminaison de la traduction

Tous comme l'initiation et l'élongation, la terminaison de la traduction est contrôlée par une série d'attachement-relâchement de facteurs dans un ordre précis.

Régulation de l'expression des gènes

IV

L'expression d'un gène	37
La régulation génétique	37
Principes de la régulation génique	37
Rappel: organisation des gènes bactériens	38
Les différents niveaux de régulation génétique chez les bactéries	38
Opéron inductible et opéron répressible	38

1. L'expression d'un gène

Définition

Elle désigne le processus biochimique par lequel l'information héréditaire stockée dans un gène est lue pour aboutir à la fabrication de molécules qui auront un rôle actif dans le fonctionnement cellulaire, comme les protéines ou les ARN.

2. La régulation génétique

Définition

Un moyen pour la cellule de développer des mécanismes qui lui permettent d'activer les gènes qui codent pour des protéines nécessaires et de les réprimer au moment où ils deviennent inutiles.

La régulation de l'expression des gènes passe avant tout par la régulation de la transcription et de l'expression des ARNm, qui aboutit à réguler la production et la nature des protéines.

Elle est assurée de plusieurs façons :

-*Régulation quantitative* :

- Modulation du niveau de transcription d'un gène
- Régulation de l'accumulation des ARNm

-*Régulation qualitative* :

- Structure des ARNm

3. Principes de la régulation génique

En 1960, « *Francois Jacob et Jacques Monod* » découvrent le premier système de régulation

génétique, appelé opéron. Ils ont étudié comment l'expression génétique change chez les procaryotes en fonction d'un changement dans l'environnement et ont proposé qu'un répresseur pouvait bloquer la synthèse d'ARN d'un ensemble spécifique de gènes, l'opéron lactose, à moins qu'un inducteur, le lactose, se lie au répresseur. Ils ont gagné *le prix Noble de médecine et de physiologie en 1965*.

4. Rappel : organisation des gènes bactériens

- 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

5. Les différents niveaux de régulation génétique chez les bactéries

1. *Le contrôle transcriptionnel* : accélérer ou ralentir la transcription d'un ARNm.
2. *Le contrôle post-transcriptionnel* : L'ARNm peut être décomposé avant la traduction.
3. *Le contrôle post-translationnel* : la protéine peut être décomposée ou modifiée chimiquement pour contrôler la vitesse à laquelle elle devient active.

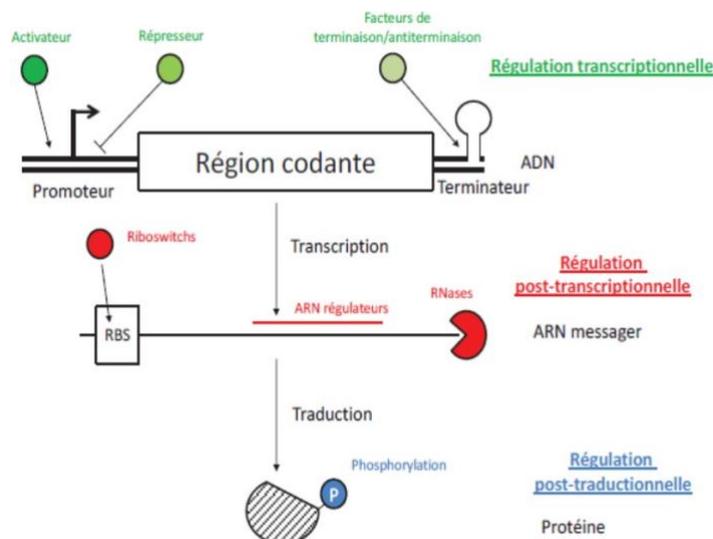


Figure 38

6. Opéron inductible et opéron répressible

Définition : Opérons inductibles

Codent pour des enzymes impliquées dans la voie catabolique (processus de dégradation) et sont

induits par le substrat de cette voie.

Exemple : Opéron Lactose

Synthèse des enzymes nécessaires au métabolisme (catabolisme) du lactose.

Définition : Opérons répressibles

Codent pour des enzymes impliquées dans l'anabolisme (processus de biosynthèse) et l'expression est contrôlée par le produit final de la chaîne

Exemple : Opéron Tryptophane

Synthèse des enzymes impliquées dans la biosynthèse (anabolisme) du tryptophane.

6.1. Régulation de l'opéron lactose chez *E. coli*

E.coli utilise le glucose s'il est disponible, mais peut métaboliser d'autres sucres si le glucose est absent.

Les enzymes exigées pour métaboliser le lactose sont synthétisées si le glucose est épuisé et le lactose est disponible.

6.1.1. Organisation de l'opéron lactose d'*E.coli*

L'opéron comprend :

1. Les gènes de structure
2. Un ou plusieurs gènes régulateurs codants des protéines régulatrices : répresseurs ou activateurs.
3. Des éléments de contrôle présents dans la séquence d'ADN : promoteur, opérateur et terminateur.

Rappel

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

- Un Promoteur est une région d'ADN en amont du site d'initiation de la transcription sur laquelle l'ARN polymérase peut se lier.

- Un Termineur est une région d'ADN qui marque la fin de la transcription

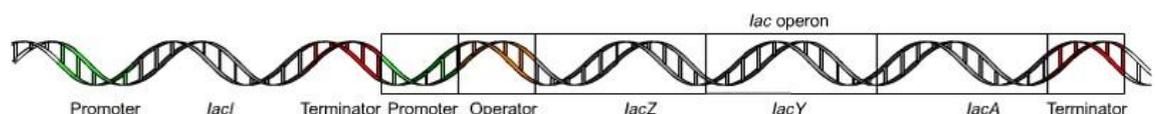


Figure 39 : Structure de l'opéron lac

- Le gène *lacZ* code l'enzyme β -galactosidase, qui hydrolyse le lactose (en galactose et glucose)

- Le gène *lacY* code la galactosidase perméase, une protéine de transport pour le lactose.
- Le gène *lacA* code la thiogalactoside acétyltransférase.
- Le gène *lacI* (gène adjacent n'appartenant pas à l'opéron) code un répresseur qui bloque la transcription de l'opéron lactose.

6.1.2. En présence de lactose et absence de glucose

En absence du glucose et la présence du lactose, l'allolactose (β -D-galactopyranosyl-(1-6)- β -glucopyranose), une molécule inductrice synthétisée à partir du lactose (β , 1-4) par un réaction de transglycosylation se fixe sur le répresseur et l'inactive de sorte qu'il ne puisse plus se fixer à la séquence de l'opérateur t permettant donc à l'ARN polymérase de progresser et d'entamer la transcription et par la suite la production des enzymes chargées de métabolisme du lactose (Figure 40). Donc le lactose agit comme inducteur des enzymes chargées de son métabolisme

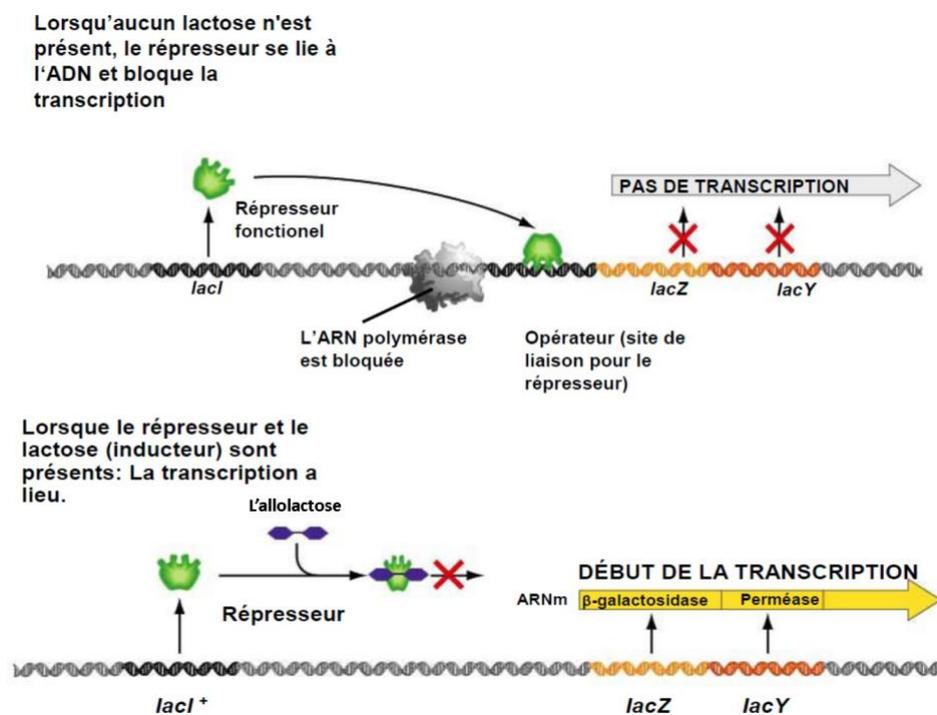


Figure 40

6.1.3. En présence de glucose et de lactose

Un système supplémentaire de contrôle s'ajoute au système répresseur-opérateur. Ce système existe car les cellules possèdent des enzymes spécifiques qui favorisent l'absorption du glucose et son métabolisme. Si le lactose et le glucose sont présents simultanément, la synthèse de la β -galactosidase n'est pas induite tant que le glucose n'a pas été épuisé. Ainsi, la cellule économise sa machinerie métabolique.

La présence du glucose va entraîner la diminution de concentration de l'AMPc donc pas de fixation de l'AMPc à la protéine CAP et par voie de conséquence pas de fixation de l'ARN polymérase donc pas d'expression de l'opéron lactose (Figure 41).

Après l'épuisement du glucose, la bactérie se trouve privée d'une source d'énergie, elle accumule l'AMPc qui va se lier à la protéine CAP formant le complexe CAP-AMPc qui va se fixer sur le promoteur permettant ainsi la transcription de l'ARN polymérase et la production des enzymes chargées (Figure 41).

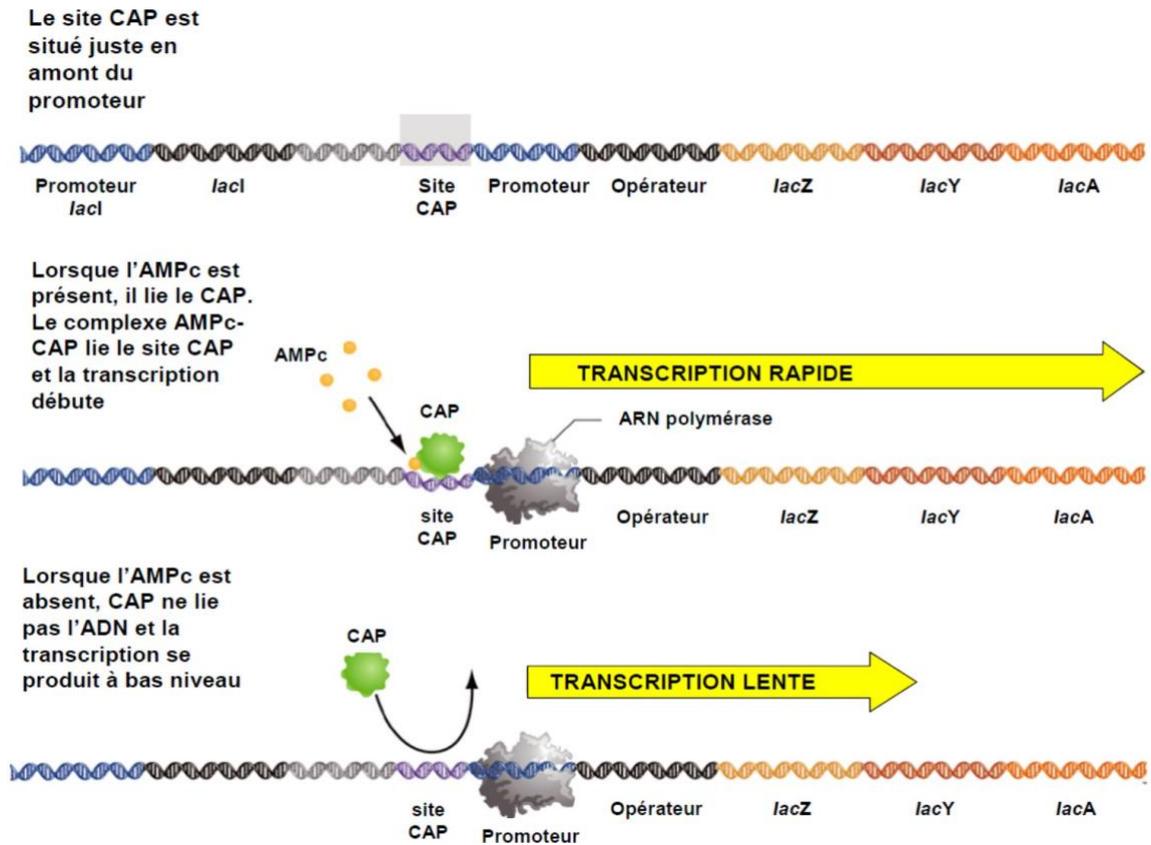


Figure 41

6.2. Régulation de l'opéron Tryptophane chez E. coli

Le tryptophane est un acide aminé qui est :

1. Produit à partir de l'acide chorismique.
2. Nécessaire à la synthèse des protéines.
3. Peu fréquent dans les protéines.
4. Régulation à différents niveaux : Activation/répression de la transcription. Atténuation de la transcription.

6.2.1. Organisation de l'opéron Trp d'E. coli

L'opéron tryptophane est constitué des éléments suivants (Figure 42) :

1. Cinq gènes de structure : trpA, trpB, trpC, trpD, trpE qui sont des gènes qui permettent de transformer le chorismate en tryptophane.

2. Un gène régulateur *trpR* codant pour un apo-répresseur.
3. Des éléments de contrôle : représentés par le promoteur et le l'opérateur.

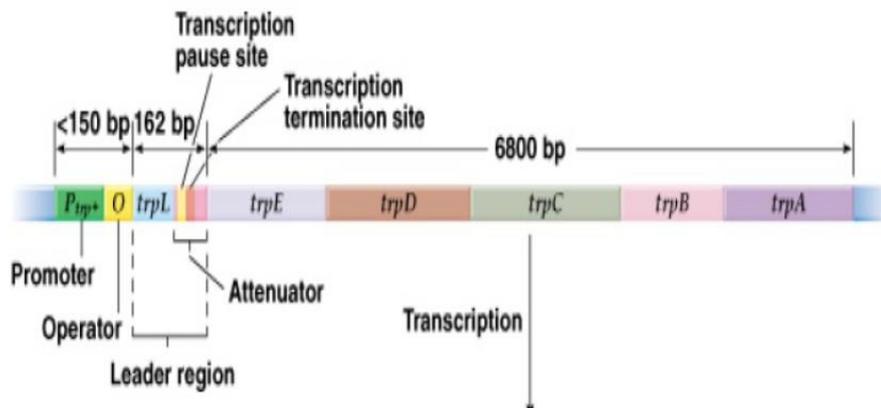


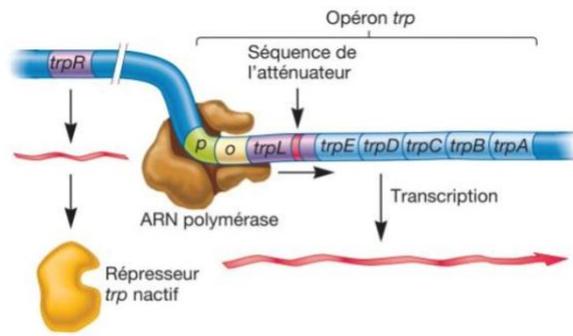
Figure 42

6.2.2. En absence de tryptophane

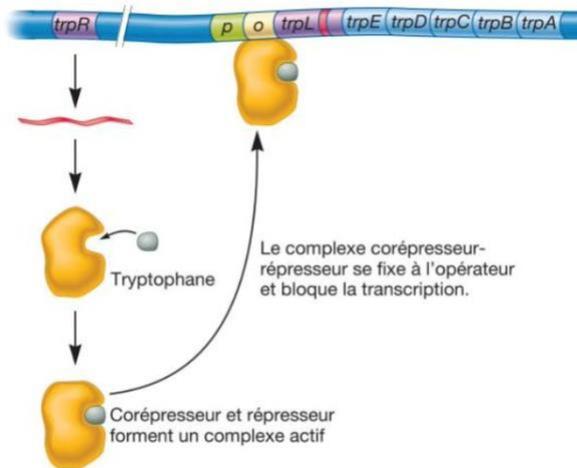
La synthèse de l'ARNm de l'opéron est contrôlée par un répresseur qui bloque la transcription lorsqu'il est lié par le tryptophane (co-répresseur). En absence de tryptophane l'ARN polymérase se fixe sur le promoteur permettant la transcription des gènes de structure ainsi leur traduction aboutissant à la production de tryptophane.

6.2.3. En présence de tryptophane

Le tryptophane se lie au répresseur modifiant ainsi sa conformation ce qui lui permet de se fixer sur l'opérateur et d'empêcher la fixation de l'ARN polymérase et donc la transcription (Figure 43).



(a) Faibles teneurs en tryptophane, l'opéron *trp* entier est transcrit



(b) Teneurs élevées en tryptophane, il y a répression

Figure 43

Prédiction de gènes

V

Annotation des génomes

44

Annotation structurale

44

1. Annotation des génomes

- L'annotation du génome est le processus d'identification des éléments fonctionnels le long de la séquence d'un génome, lui donnant ainsi un sens.
- Elle est nécessaire, car le séquençage de l'ADN produit des séquences dont la fonction est inconnue.
- Elle fournit les informations nécessaires pour la compréhension du fonctionnement de la cellule et des relations entre les gènes.



Définition

L'annotation du génome consiste à prédire et localiser l'ensemble des séquences codantes (gènes) du génome et à déterminer et identifier leur structure (annotation syntaxique), leur fonction (annotation fonctionnelle) ainsi que les relations entre les entités biologiques relatives au génome (annotation relationnelle). L'information résultante enrichit les bases de données biologiques.

On distingue deux grands types d'annotation génomique :

1- *L'annotation structurale (Prédiction de gènes)* : elle est utilisée pour définir des régions d'intérêt sur la séquence comme les bornes de début et de fin des gènes, ou des motifs d'intérêt dans le fonctionnement de la cellule (i.e., le motif RBS, les sites de régulation.).

2- *L'annotation fonctionnelle* : elle recherche à prédire les fonctions biologiques des gènes prédits.

2. Annotation structurale

Informations utilisées

Deux groupes d'information :

1- *Les informations intrinsèques (Méthodes ab-initio)*

Les informations contenues dans la séquence nucléique considérée.

2- *Les informations extrinsèques (Méthodes par homologie de séquence)*

Les informations obtenues par comparaison de la séquence nucléique d'intérêt avec des séquences

déjà connues.

2.1. Les méthodes Ab initio

Les approches ab initio ont pour objectif de prédire l'ensemble des gènes présents dans une séquence nucléique sans autre connaissance extérieure. Pour cela, elles tirent parti des signaux présents dans la séquence et des biais de composition des séquences codantes.

Le premier signal qui peut être exploité provient simplement des bornes des gènes en particulier pour les organismes procaryotes.

2.1.1. Les signaux liés à la structure du gène

- La reconnaissance des gènes est facilitée par l'identification de zones particulières (Figure 44-45).
- L'annotation des génomes de procaryotes est relativement plus aisée que celle des génomes eucaryotes pour les raisons suivantes :
 - 1- Les génomes procaryotes sont plus petits que les génomes eucaryotes et ont surtout une densité de codage bien plus importante, de l'ordre de 80-90 %, tandis qu'elle peut aller de 70% chez la levure à quelques pourcentages chez l'humain.
 - 2- Les gènes procaryotes sont fréquemment organisés en opéron, c'est-à-dire qu'une seule unité de transcription peut contenir plusieurs séquences codantes.
 - 3- Les gènes procaryotes ne sont pas morcelés contrairement à ceux des eucaryotes.

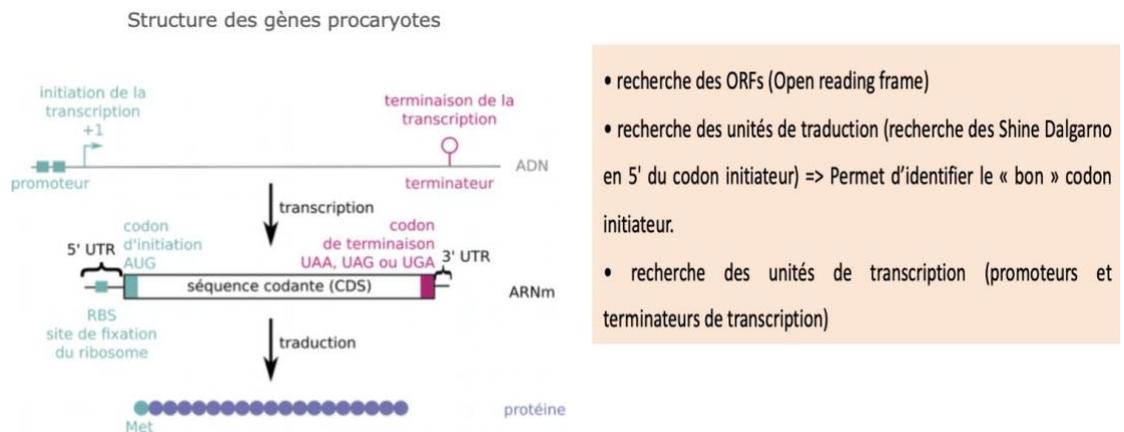
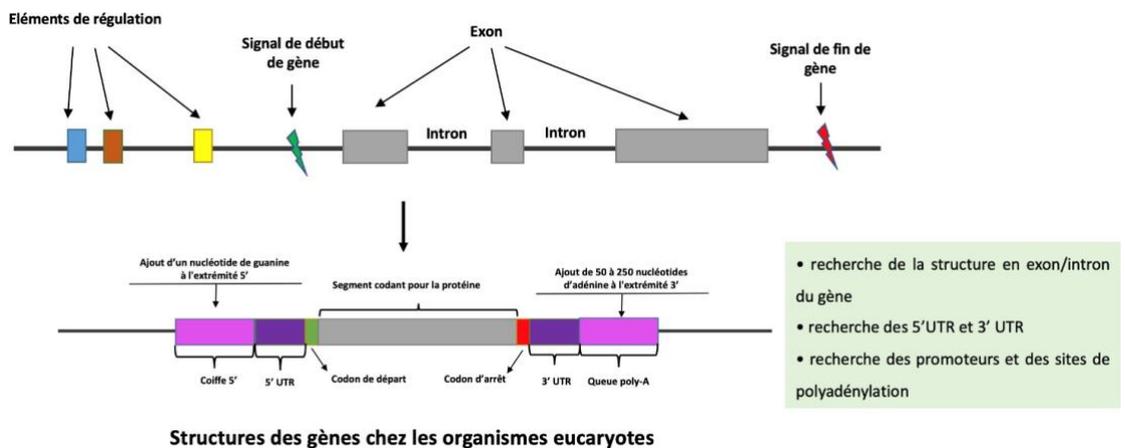


Figure 44

Chez les génomes eucaryotes, l'annotation est nettement plus compliquée pour les raisons suivantes :

- 1- Les génomes eucaryotes ont une faible densité de codage. Il y a donc de larges régions génomiques sans séquence codante ;
- 2- Les gènes eucaryotes sont morcelés ; ils subissent des modifications de la séquence nucléotidique (l'épissage) du pré-ARN messager.
- 3- Enfin, l'épissage peut être alternatif : différents profils d'épissage existent pour un même pré-ARN

messager et par conséquent un gène peut produire différentes CDS.



Structures des gènes chez les organismes eucaryotes
Figure 45

Exemple : Détection des ORF (Open Reading Frame)

- La phase ouverte ou « Open Reading Frame (ORF) » est la région de l'ADN qui sépare deux codons STOP. Dans celle-ci, une séquence codante (CDS, CoDing Sequence, région traduite en protéine) commence par un codon START (ATG+ autres), se termine par le codon STOP (TAA, TAG, TGA) et est précédée d'un site de liaison aux ribosomes (RBS) (Figure 46).

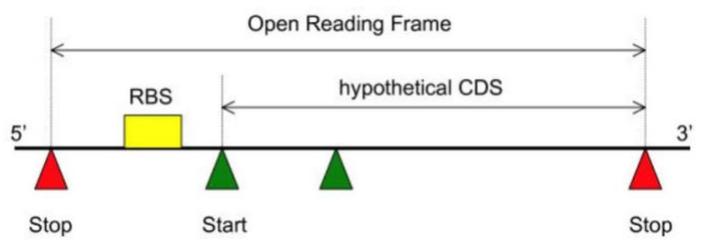


Figure 46

- La lecture (traduction en protéine) peut se faire sur les deux brins d'ADN complémentaires et selon trois cadres de lecture possibles pour chaque brin :

- La recherche des régions codantes doit donc en pratique être effectuée sur six séquences virtuelles différentes.
- Un cadre de lecture est l'une des six possibilités pour traduire une séquence génomique double brin donnée en acides aminés (Figure 47).
- Plusieurs logiciels donnent une représentation graphique des cadres de lecture ouverts.

Le logiciel ORF finder disponible sur le site du NCBI (<http://www.ncbi.nlm.nih.gov/>) recherche pour une séquence donnée tous les ORFs possibles sur les six phases de lecture.

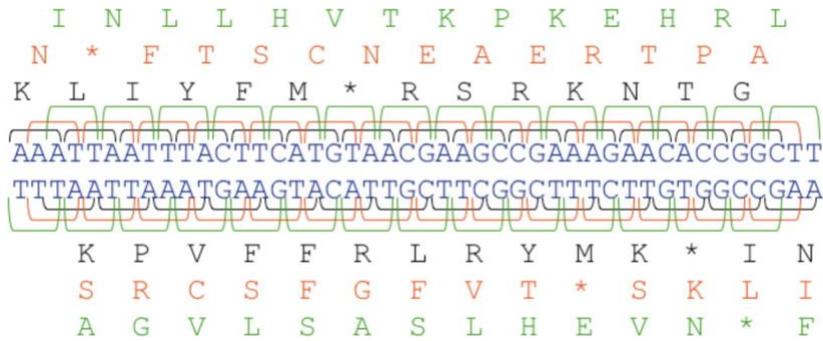


Figure 47

Remarque

- Un cadre de lecture qui n'encadre pas un gène devrait avoir de courts ORFs dû à la présence d'une grande quantité de codons STOP.
- Tous les ORF ne sont pas des gènes -> validation avec comparaison de séquences ou statistiques.
- Comment choisir le bon codon START ? Pas toujours le codon le plus éloigné du codon STOP.

On va valider en cherchant des séquences connues pour être situées en début de gènes chez l'espèce étudié.

- Il est possible d'avoir plus d'un gène sur un ARNm.
- Il arrive qu'il y ait des gènes qui se chevauchent codant deux protéines différentes => Ces gènes sont trouvés dans des cadres de lecture différents.

2.1.2. Les biais de composition de la séquence codante

L'usage du code génétique est la fréquence d'utilisation des codons synonymes. Ces derniers ne sont pas employés à la même fréquence soit entre espèces soit au sein d'une espèce. Il existe des préférences de certains codons par rapport à d'autres, c'est le biais d'usage du code (BUC).

Exemple

Chez *E. coli* K12, l'emploi des codons pour l'arginine : sur les 6 codons, deux seulement (CGC et CGT) totalisent plus de 78% des codons utilisés, tandis que le rassemblement des 3 codons AGA, AGG et CGA forme moins de 12% des codons employés.

2.2. Les méthodes par homologie de séquence

Les méthodes par homologie de séquences utilisent comme première et principale source d'information la similarité avec des séquences connues et déjà annotées => On compare la séquence des ORF avec les séquences d'autres espèces qui sont disponibles dans les banques de données.

L'hypothèse de ces méthodes est la suivante : deux séquences significativement similaires ont généralement une fonction identique ou proche.

Comparaison de séquences protéiques

VI

1. L'annotation fonctionnelle

- L'annotation fonctionnelle des gènes correspond au processus d'attribution d'un rôle biologique à chacun de ces gènes. En effet, ils vont produire des protéines disposant d'une ou plusieurs fonctionnalités pour l'organisme.
- Ces fonctions sont habituellement classées en trois catégories : moléculaires, cellulaires, phénotypiques

1.1. Annotation fonctionnelle expérimentale

- L'annotation fonctionnelle peut être basée sur des méthodes expérimentales.

Les validations expérimentales peuvent démontrer la fonction d'un gène avec une grande confiance.

- Assigner des fonctions par des méthodes expérimentales (biochimie, biologie moléculaire...).

Exemple

- Un gène peut être cloné dans un vecteur d'expression pour produire la protéine correspondante dans un autre organisme facilement manipulable en laboratoire.
- Cette expression hétérologue permet ensuite de purifier la protéine et de tester, par exemple, son activité biochimique dans le cas d'une enzyme.
- Ces méthodes expérimentales ne sont pas envisageables pour caractériser la fonction de tous les gènes d'un organisme car elles demandent beaucoup de temps de conception et de réalisation.

1.2. L'annotation fonctionnelle in silico

- Aujourd'hui, le plus souvent, l'annotation fonctionnelle est réalisée *in silico*.
- Elle est généralement basée sur la recherche de *similarités de séquence* avec des gènes de fonctions connues, répertoriés dans les bases de données.

- Classiquement, la fonction de la protéine possédant le plus fort degré de similarité est alors assignée à la protéine candidate. Cette prédiction est basée sur l'hypothèse que des gènes homologues possèdent des fonctions similaires, héritées de leur dernier ancêtre commun.

- Pour mesurer la similarité entre deux séquences on doit :

1- *Aligner les séquences => Alignement = Comparaison=* Écrire une séquence au-dessous de l'autre pour pouvoir les comparer.

2- *Calculer le score de cet alignement*

- Ces 2 opérations sont effectuées par des méthodes de bio-informatique

👉 Exemple

- La ressemblance entre les séquences est illustrée par les rectangles gris qui encadrent l'alignement (Figure 48).

- *Figure 48 (1)* : l'alignement d'une protéine humaine (Homo sapiens) avec une protéine de la souris (Mus musculus).

- *Figure 48 (2)* : l'alignement de cette même protéine humaine avec une espèce évolutivement moins proche de l'Homme que la souris, un poisson Danio rerio.

09Y5Y	MEKRLLLGLRYQEVYTRHPAATAQLETAVRGFSYLLAGRFADSHSELVYSASNLLVLL	60	PEX16_HUMAN	09Y5Y	MEKRLLLGLRYQEVYTRHPAATAQLETAVRGFSYLLAGRFADSHSELVYSASNLLVLL	60	PEX16_HUMAN
091XC9	MEKRLLLSRLRYQEVYTRHPAATAQLETAVRGFSYLLAGRFSDSHSELVYSASNLLVLL	60	PEX16_MOUSE	040RH7	MEKRLTRVFERVQEVYTRSPAASHLESYVALSYLTAGRFSDSHSELVYSASNLLVLL	60	PEX16_DANRE
09Y5Y	NDGILRKLKRLKPLVSLDQKLLTWLSVLECEVFMEMGAAMVGEVGRMLVIALQLAK	120	PEX16_HUMAN	09Y5Y	NDGILRKELRKKEPLVSLDQKLLTWLSVLECEVFMEMGAAMVGEVGRMLVIALQLAK	120	PEX16_HUMAN
091XC9	NDGILRKLKRLKPLVSLDQKLLTWLSVLECEVFMEMGAAMVGEVGRMLVIALQLAK	120	PEX16_MOUSE	040RH7	NDGILRKNLSRATLPMISDQKLLTWLSVLECEVFMEMGAAMVGEVGRMLVIALQLAK	120	PEX16_DANRE
09Y5Y	AVLRMLLLWFKAGLQTSPPVPLDRETAQPPDGDHSFGNHEOSYVGRKSNRVVRLQN	180	PEX16_HUMAN	09Y5Y	AVLRMLLLWFKAGLQTSPPVPLDRETAQPPDGDHSPGNHESYVGRKSNRVVRLQN	180	PEX16_HUMAN
091XC9	AVLRMLLLWFKAGLQTSPPVPLDRETAQPPDGDHSPGSHHESYVGRKSNRVVRLQN	180	PEX16_MOUSE	040RH7	AVLRMLLLWFKAGLQTSPPVPLDRETAQPPDGDHSPGNHEOSYVGRKSNRVVRLQN	180	PEX16_DANRE
09Y5Y	TPSLHSRHWGAPQDREIRKQDQHEELSATPTPLGQETIAEFLYIARPLLHLSLGLWGQ	240	PEX16_HUMAN	09Y5Y	TPSLHSRHWGAPQDREIRKQDQHEELSATPTPLGQETIAEFLYIARPLLHLSLGLWGQ	240	PEX16_HUMAN
091XC9	SPSLHSRHWGAPQDREIRKQDQHEELSTPTPLGQETIAEFLYIARPLLHLSLGLWGQ	240	PEX16_MOUSE	040RH7	APSLHSRHWGAPQDREIRKQDQHEELSATPTPLGQETIAEFLYIARPLLHLSLGLWGQ	239	PEX16_DANRE
09Y5Y	RSWPKLLEGVVQVTSLSLSDRKLTRERRELRRAITLLLYLRSPPYDFSEARIL	300	PEX16_HUMAN	09Y5Y	RSWPKMLLAQVVDVTSLSLSDRKLTRERRELRRAITLLLYLRSPPYDFSEARIL	300	PEX16_HUMAN
091XC9	RSWPKLLEGVVQVTSLSLSDRKLTRERRELRRAITLLLYLRSPPYDFSEARIL	300	PEX16_MOUSE	040RH7	RSWPKMLLAQVVDVTSLSLSDRKLTRERRELRRAITLLLYLRSPPYDFSEARIL	299	PEX16_DANRE
09Y5Y	FLLQLLADHPGVGLVTRPLMDYLRTPWQKIYFYWG	336	PEX16_HUMAN	09Y5Y	FLLQLLADHPGVGLVTRPLMDYLRTPWQKIYFYWG	336	PEX16_HUMAN
091XC9	FLLQLLADHPGVGLVTRPLMDYLRTPWQKIYFYWG	336	PEX16_MOUSE	040RH7	FLLRFLADHPGVGLVTRPLMDYLRTPWQKIYFYWG	335	PEX16_DANRE

1

2

Figure 48

1.2.1. Types d'alignements

Alignement global

- Prise en compte de la totalité de la séquence (Figure 49).

- 2 séquences, comparaison sur leur totalité, trouve le meilleur alignement sur toute leur longueur



Figure 49

Alignement local

- Recherche de la région de plus forte similarité entre deux séquences (Figure50).

- Les alignements locaux sont meilleurs pour des séquences avec un faible degrés de similarité ou de

tailles différentes.

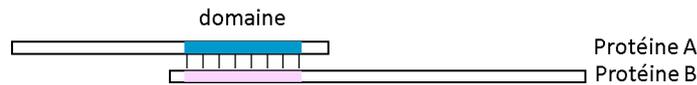


Figure 50

Alignement multiple

- L'alignement multiple qui consiste à aligner plus de deux séquences en même temps (Figure 51).
- L'alignement multiple permet d'étudier un groupe de protéines apparentées afin d'exhiber des motifs communs sensés jouer un rôle dans la fonction ou la structure des protéines d'une même famille.

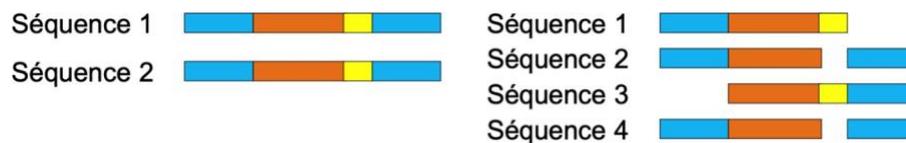


Figure 51

1.2.2. Évaluer un alignement d'AA

- Quand on calcule le score d'un alignement entre 2 séquences protéiques il faut qu'on tient compte des propriétés physico-chimiques des acide aminés.
- Les acides aminés composant une protéine peuvent avoir des propriétés physico-chimiques similaires.
- Une similitude au niveau de ces propriétés sera suffisante pour permettre la substitution d'un acide aminé en un autre sans perturber la fonction de la protéine (par exemple, échange de l'acide amine hydrophobe valine en leucine).

Exemple

- Arginine (R) et Lysine (K) (chargés positivement) ont des propriétés physico-chimiques similaires. Alors que arginine (R) (chargé positivement) et aspartate (D) (chargé négativement) ont des propriétés physico-chimiques différentes. Donc les substitutions R à D et R à K ne seront pas pénalisées de la même manière. Une substitution R à D sera pénalisée plus que R à K .
- Et les autres acides aminés ? Dans ce cas une matrice simple n'est pas utile.
- La table qui contient les valeurs de toutes les substitutions possibles s'appelle matrice de substitution (substitution matrix).

a) Matrice de substitution

Une matrice de substitution permet, pour chaque acide aminé, de connaître sa capacité à être substitué par chaque autre acide aminé, y compris lui-même.

Les matrices populaires sont :

PAM (POINT ACCEPTED MUTATIONS) Et BLOSUM (BLOCKS SUBSTITUTION MATRIX) (Figure 52).

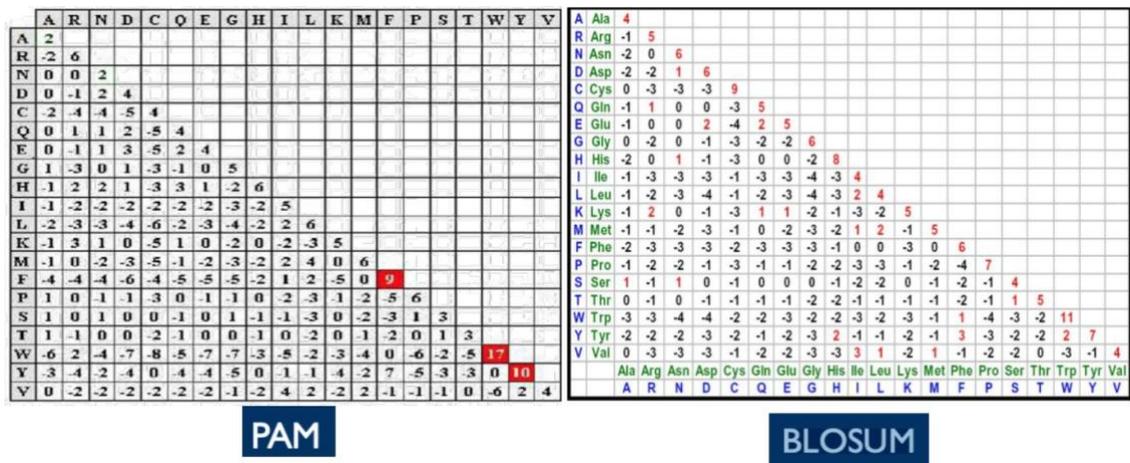


Figure 52

1.2.3. Outils bio-informatique pour la recherche de similarité et calcul du score d'alignement entre 2 séquences

La comparaison des séquences nucléiques se fait couramment à l'aide du logiciel d'alignement BLAST (Basic local alignment search tool). Ce logiciel compare la séquence requête hypothétique à toutes les séquences d'une banque de données.

Une des interfaces la plus souvent utilisée se trouve sur le site du NCBI : <http://www.ncbi.nlm.nih.gov/BLAST/>.

a) Principe de fonctionnement

BLAST recherche dans une base de données de séquence des segments qui sont localement homologues à une séquence-test fournie par l'utilisateur (query sequence).

BLAST utilise une matrice de similarité pour calculer des scores d'alignement. Il fournit un score pour chaque alignement et utilise ce score pour donner une évaluation statistique de la pertinence de cet alignement (probabilité qu'il soit dû au hasard)

i Quelques conseils

Méfiez-vous des résultats donnés par les logiciels

- La qualité des résultats est parfois diminuée au profit de la rapidité
- Certains problèmes admettent un ensemble infini de possibilités à ce n'est pas toujours la solution la meilleure qui est trouvée
- Certains logiciels ne font que de la prédiction
- Méfiez-vous des banques de données :

- Les données ne sont pas toujours fiables
- La mise à jour des données n'est pas systématiquement récente

Les mutations et les mécanismes de réparation de l'ADN

VII

Introduction	53
Les mutations et les lésions de l'ADN	53
Les mutations ponctuelles	54
Mutations pendant la réplication de l'ADN	55
Altération de l'ADN survenant en dehors de la réplication	57
Correction des erreurs d'appariement lors de la réplication	63
Correction des autres mutations de l'ADN	64
Réparation des lésions non réparées par les systèmes précédents	67

1. Introduction

- Les mutations génétiques sont le résultat d'une mauvaise séquence génétique qui s'est produite dans l'ADN de la cellule. Les gènes, qui produisent les protéines nécessaires au bon fonctionnement du corps, sont alors déréglés dans leur fonction.
- Les dommages à l'ADN intervenant spontanément et résultant d'agents de l'environnement existant en permanence, la plupart des organismes possèdent des capacités à réparer leur ADN. L'ADN est la seule macromolécule à être réparée par les cellules.
- La survie à court terme d'une cellule dépend de la stabilité génétique. Au niveau cellulaire, la cellule a donc à sa disposition des systèmes de prévention de l'altération de l'ADN. Sur 1000 modifications, seule une ne sera pas réparée, et entraînera alors une mutation permanente.

2. Les mutations et les lésions de l'ADN

Les conséquences d'une mutation seront différentes selon le type cellulaire qui est affecté :

Dans les cellules germinales : les mutations seront transmises à la descendance.

Dans les cellules somatiques : les mutations ne seront pas transmises à la descendance, elles entraînent des modifications au sein même de l'individu. On retrouvera ces mutations dans certaines cellules seulement (ex : lésions des cellules cancéreuses).

Au niveau moléculaire, les mutations peuvent être classées en trois catégories :

- Les mutations ponctuelles ou géniques affectant un ou plusieurs nucléotides (remplacement d'un ou plusieurs nucléotides, insertion et délétion). Un tel changement se produit à l'intérieur d'un seul gène et est localisé en un locus unique du chromosome.
- Des réarrangements chromosomiques modifiant l'organisation du génome (figure 53).
- Les mutations d'insertion due à l'intégration d'ADN étranger par transfert horizontal.

Remarque

Ces deux derniers types de mutations ne seront pas couverts dans ce chapitre.

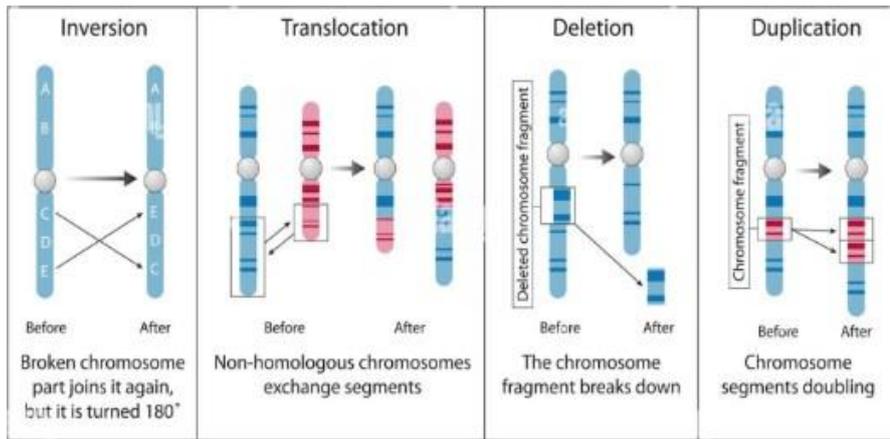


Figure 53

3. Les mutations ponctuelles

Ces mutations se regroupent en plusieurs catégories qui diffèrent par les conséquences sur la protéine codée par le gène muté.

3.1. Mutation « silencieuse »

Le codon muté code pour le même acide aminé que le codon sauvage (normal). La protéine est normale/Fonction normale.

3.2. Mutation « faux-sens »

Met en jeu l'altération d'une unique base ce qui modifie le codon de telle sorte que l'acide aminé soit lui aussi modifié donc modification de la protéine qui en résulte. De telles mutations ont généralement lieu dans l'une des deux premières bases du codon. Ce type de mutation peut provoquer une maladie si la fonction de la protéine est altérée.

3.3. Mutation « non-sens » (stop)

C'est une mutation ponctuelle qui change le codon d'un acide aminé en un codon stop. Cette mutation

provoque l'arrêt prématuré de la traduction et il en résulte une protéine plus courte. Ce type de mutation est souvent la cause de maladie car dans tous les cas, la fonction de la protéine est altérée.

4. Mutations pendant la réplication de l'ADN

Dans *le chapitre 2*, nous avons vu que l'ADN pouvait être répliqué, c'est-à-dire copié. Or, des erreurs peuvent survenir lors de cette copie : la séquence d'ADN est modifiée (Altération non corrigée par le système de correction d'édition de l'ADN polymérase), on a une mutation.

Ces mutations peuvent conduire soit à une délétion, soit à une substitution, soit à une insertion (Figure 54).

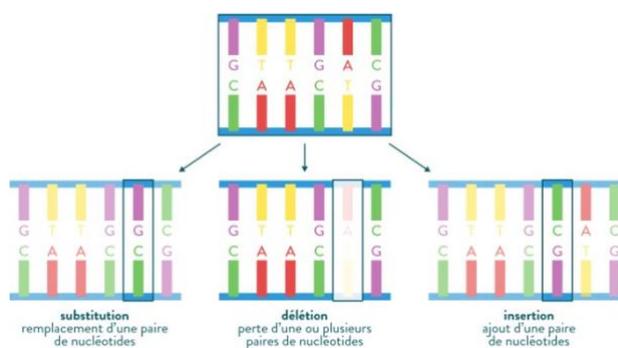


Figure 54

4.1. Mutation par substitution

Cela correspond à une mauvaise incorporation de nucléotide sur le brin fils. Une base est alors remplacée par une autre.

On distingue deux types de mutation par substitution (Figure 55) :

A- Substitution par transition : base pyrimidique substituée par une base pyrimidique ou base purique substituée par une base purique.

B- Substitution par transversion : base purique substituée par une base pyrimidique ou vice versa.

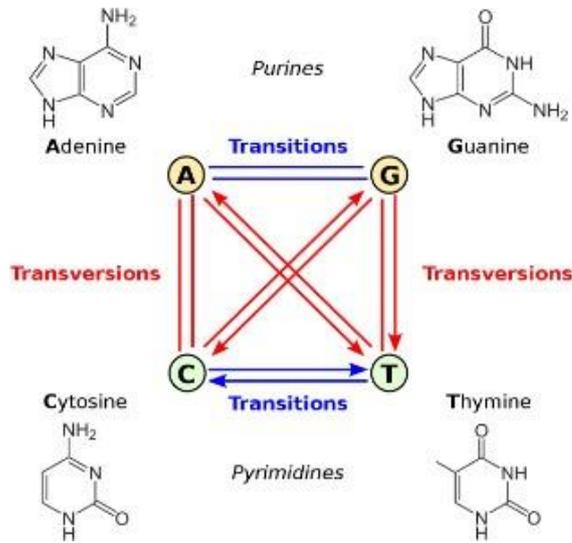


Figure 55

4.2. Mutations par délétion

- Il s'agit d'un oubli d'incorporation d'un nucléotide par l'ADN polymérase (Figure 56). Au bout de deux cycles successifs, la séquence est totalement différente par rapport aux séquences parentales.
- Lorsque le brin fils va servir d'ADN matrice, cela entraîne un décalage du cadre de lecture.

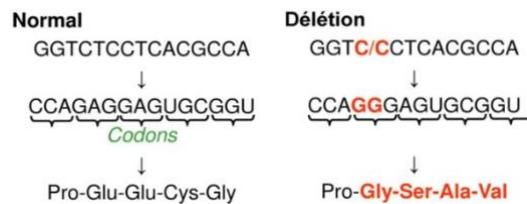


Figure 56

4.3. Mutations par insertion

Il s'agit d'une introduction d'un nucléotide en trop par l'ADN polymérase (Figure 57). Cela entraîne également un décalage de cadre de lecture.

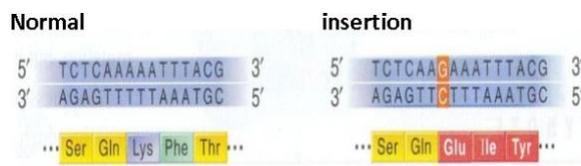


Figure 57

Remarque

Lorsque les délétions et insertions sont plus longues, elles peuvent supprimer l'expression d'une ou de plusieurs protéines, voire la fonction d'un gène entier (Ces délétions et insertions ne sont pas réversibles).

5. Altération de l'ADN survenant en dehors de la réplication

5.1. Les lésions spontanées

- Des altérations de la structure chimique des bases de l'ADN peuvent se produire spontanément.
- Les dommages plus classiques sont la désamination et la dépurination/dépyrimidination.

5.1.1. La désamination

- La désamination est une réaction chimique au cours de laquelle une molécule perd un groupement amine.
- Les enzymes catalysant cette réaction sont appelées désaminases.
- Elles donnent naissance à des bases modifiées (bases désaminées), qui établissent des liaisons particulières avec les autres bases et entraînent donc des mésappariements.
- La désamination la plus fréquente est celle de la cytosine en uracile, base normalement absente dans l'ADN. L'uracile est ensuite capable de s'apparier à une adénine (Figure 58).

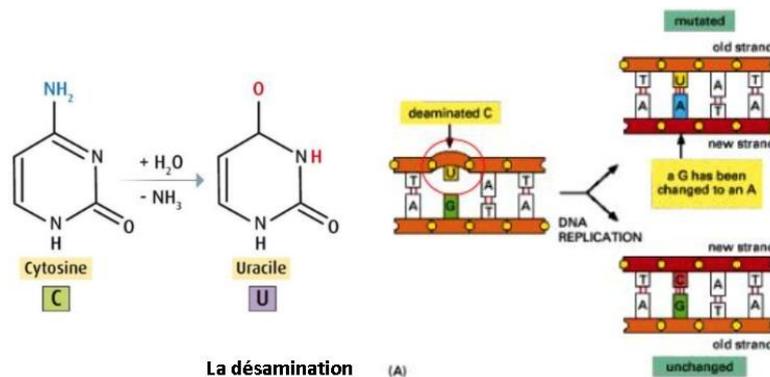


Figure 58

5.1.2. La dépurination/dépyrimidination

Correspond à l'interruption d'une liaison N-Glycosidique entre la base purique (A ou G) et le désoxyribose, donc perte d'un résidu guanine ou adénine de l'ADN et la création d'un site apurique (ou apurinique) (Figure 59).

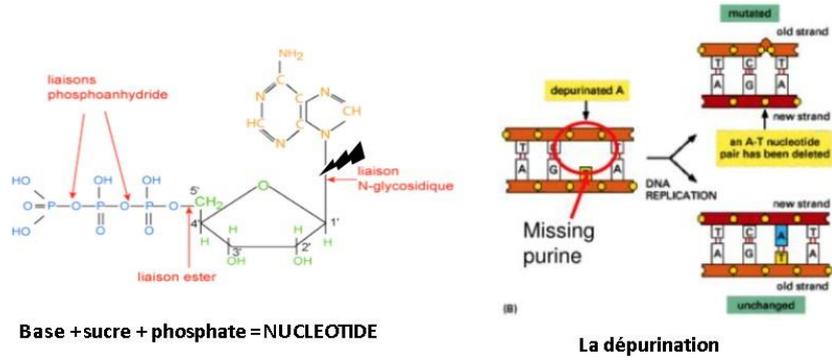


Figure 59

5.2. Mutations induites par des agents mutagènes

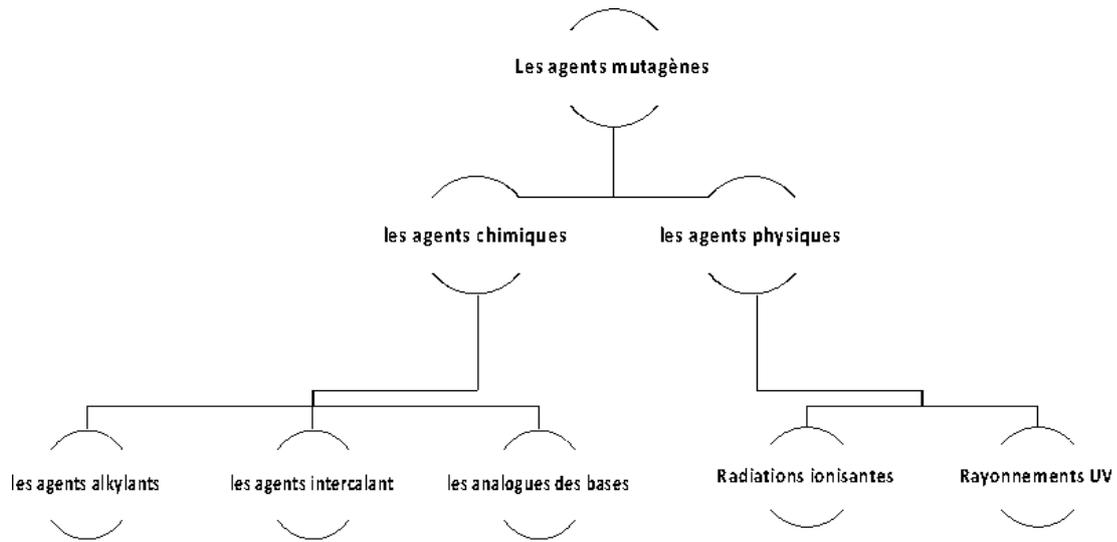


Figure 60

5.2.1. Les agents chimiques

a) Modification de bases par les agents alkylants

Ces agents altèrent les bases en ajoutant des groupes éthyle ou méthyle sur l'atome d'oxygène de la guanine ou de la thymine. Les bases modifiées ont des propriétés d'appariement différentes : par exemple, la 6-O-méthylguanosine s'apparie avec T (Figure 61).

☞ Exemple : Exemples d'agents alkylants

- Ethylméthane sulfate
- Nitrosoguanidine
- Diméthylsulfate.

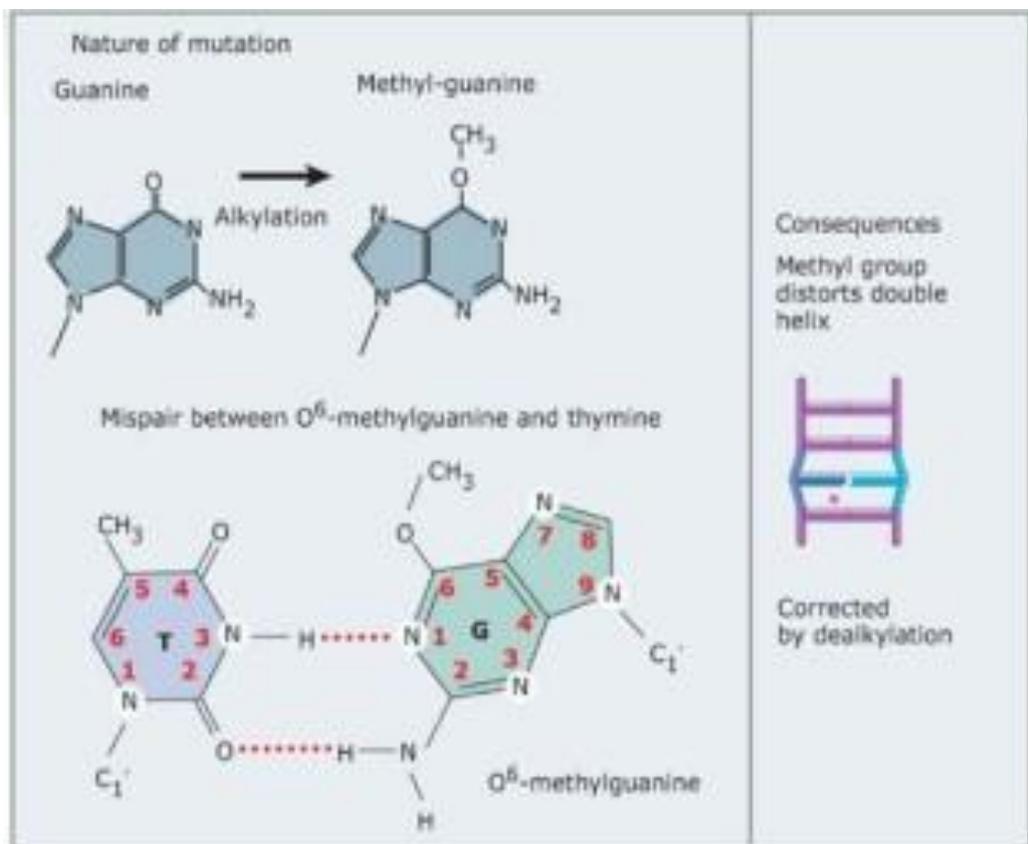


Figure 61

b) Insertions ou délétions de bases par les agents intercalants

Les agents intercalants sont des structures planes à 3 cycles s'apparentant à des paires de bases et qui vont ainsi s'intercaler entre les paires de bases de la double hélice. Elles induisent une distorsion causant à la réplication suivante une insertion de bases (fréquente) ou une délétion (Figure 62).

☞ Exemple : Exemples d'agents intercalants

- L'acridine orange
- La proflavine

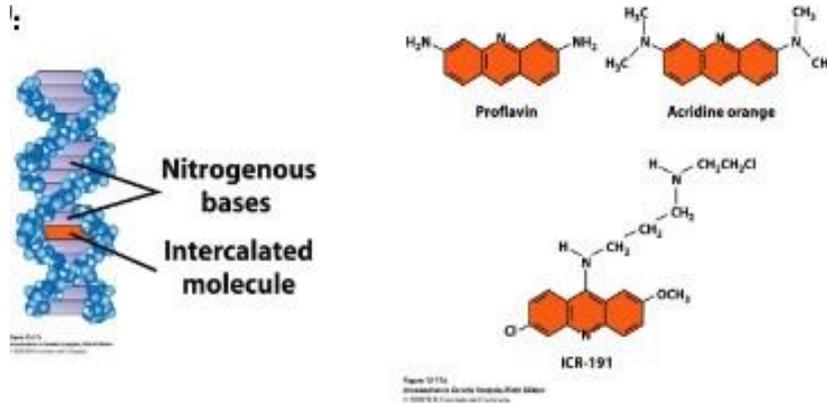


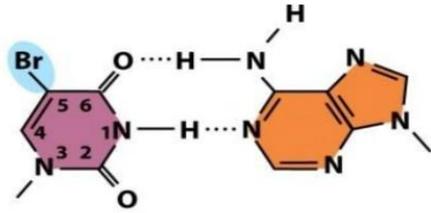
Figure 62

c) Incorporation d'analogues de bases dans la molécule d'ADN

Les analogues de bases sont des composés chimiques qui ressemblent suffisamment aux bases azotées normales de l'ADN pour être parfois incorporées à la place de celle-ci, lors de la réplication, au niveau du brin fils à la place des bases normales. Ces analogues ont des propriétés d'appariement différentes, ce qui entraîne des mutations après réplication (Figure 63).

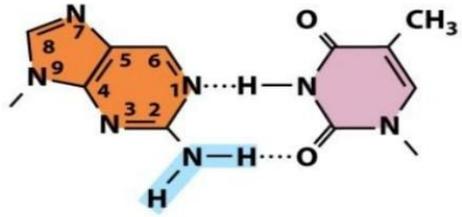
☞ Exemple

- Le 5-Bromo-uracile (5BU) : analogue de la thymine
- Le 2-Amino-purine (2AP) : analogue de l'adénine



Common keto form of 5-BU

Adenine



2-AP

Thymine

Figure 63

5.2.2. Les agents physiques

a) Lésions par radiations ionisantes

- Les radiations ionisantes peuvent aboutir à des cassures de brins, qui touchent soit un seul brin (cassure simple brin), soit les deux brins (cassure double brin) (Figure 64).
- Les radiations ionisantes capables d'induire ce type de lésions sont les rayons cosmiques, la radioactivité et les rayons X (utilisés à fin thérapeutique).
- Les mécanismes d'action de lésions par ces radiations sont multiples :

Action directe : cassure de brins puis création de sites AP.

Action indirecte : met en jeu la création de réactifs oxydatifs qui vont créer les lésions oxydantes.

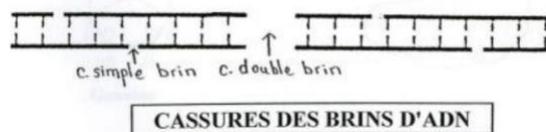


Figure 64

Définition : Les sites AP

Un site AP (site apurinique/apirimidique), ou site abasique, est un emplacement de l'ADN où la base (purine ou pyrimidine) est manquante. En termes de séquence génétique, c'est donc un site « vacant » (Figure 65).

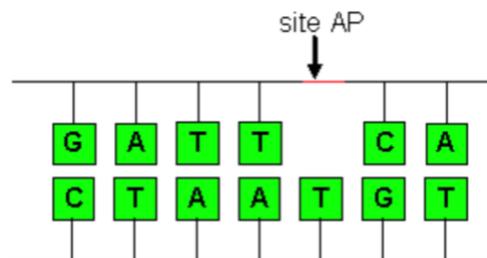


Figure 65

b) Lésions par la lumière UV

- Les rayonnements UV, peuvent également créer de liaisons covalentes sur deux pyrimidines adjacentes appartenant au même brin d'ADN : formation de dimère de pyrimidine (Figure 66).
- Ces liaisons covalentes s'établissent souvent entre deux thymines adjacentes ce qui aboutit à la création d'un dimère de thymines. Cela entraîne une distorsion de la molécule ADN ce qui bloque la réplication, s'il n'y a pas de réparation.
- Les UV solaires sont capables d'induire ces types de lésions, au niveau des cellules de la peau, et s'il n'y a pas de réparation, comme chez certaines personnes prédisposées, il y aura le développement de cancer de la peau : les mélanomes.

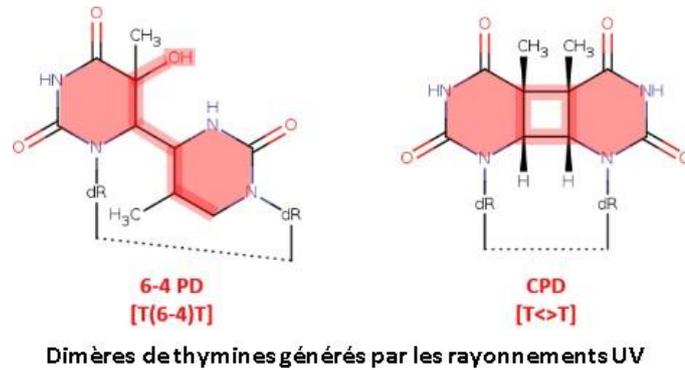


Figure 66

6. Correction des erreurs d'appariement lors de la réplication

6.1. Correction immédiate : Fonction d'édition de l'ADN polymérase

- La correction de la synthèse d'ADN lors de la réplication est effectuée par des protéines particulières que l'on nomme : nucléases (Figure 67).
- Elles sont associées au complexe enzymatique de l'ADN polymérase.

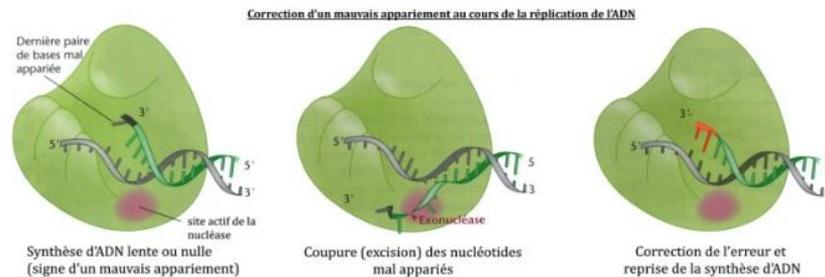


Figure 67

6.2. Le système MMR (Methyl mediated -Mismatch Repair system)

Les mésappariements de l'ADN susceptibles d'être réparés par le système MMR sont, dans la grande majorité des cas, la conséquence d'erreurs commises par l'ADN polymérase au cours du processus de réplication.

L'existence d'un processus MMR post-répliatif permet d'augmenter la fidélité de la réplication. Fait partie des voies de réparation par excision :

- Un seul brin est endommagé
- Retrait du dommage
- Remplissage de l'espace vide généré en utilisant le brin non endommagé comme matrice

- Ligation
- Le système de réparation MMR reconnaît et élimine les mésappariements mais aussi les erreurs d'insertion et de délétion
- Lorsque le système détecte un mésappariement, il détecte aussi le brin qui doit être corrigé.
- Pour pouvoir savoir quel est le brin avec le mésappariement, il se base sur le temps où il n'y a pas encore eu méthylation du brin fils. Le nucléotide du mésappariement du brin non méthylé est celui qui doit être corrigé.
- Ce système existe aussi bien chez les cellules eucaryotes et que chez les cellules procaryotes

7. Correction des autres mutations de l'ADN

7.1. Réparation par réversion directe (DRR)

La réparation par réversion directe (« Direct Reversal Repair », DRR) est associée à des altérations spécifiques où l'objectif général est d'inverser le mécanisme qui a conduit à l'altération de l'ADN.

Voici les deux exemples les plus courants :

- Mécanisme faisant intervenir la photolyase
- Mécanisme faisant intervenir les alkyltransférases

7.1.1. Mécanisme faisant intervenir la photolyase

- La photolyase permet de réparer la lésion induite par la lumière UV (les dimères de thymines ou photodimères).
- Elle est activée par la lumière visible.
- Quand elle est activée, elle se lie aux dimères de thymine pour le scinder afin de faire disparaître la liaison et revenir à deux thymines adjacentes (Figure 68).

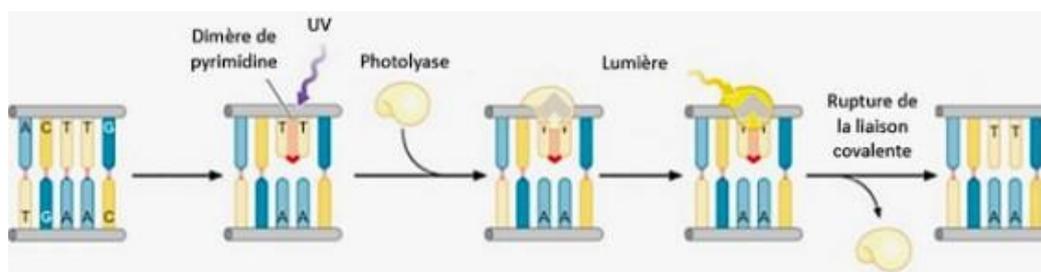


Figure 68

7.1.2. Mécanisme faisant intervenir les alkyltransférases

- Les alkyltransférases vont avoir pour rôle de réparer les liaisons induites par les agents alkylants.
- La formation de la O6 -méthylguanine est une mutation qui peut se transmettre.

- La réparation se fait grâce à l'enzyme O6-méthylguanine méthyltransférase.
- Elle se lie et transporte le groupement méthyle au niveau d'une cystéine interne à cette enzyme, où est localisé le site actif de l'enzyme. On a donc un retour à la guanine, et une dégradation du groupement méthyle (Figure 69).

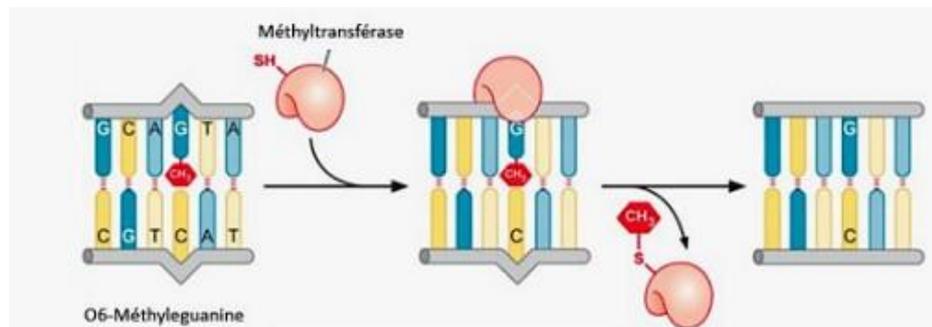


Figure 69

7.2. Réparation par excision

- C'est un mécanisme de réparation simple brin.
- On peut distinguer deux types de mécanismes en fonction de l'excision de la partie altérée (NER ou BER).
- C'est un mécanisme multi - étapes :
 - 1- Reconnaissance de la lésion.
 - 2- Excision de la partie altérée : soit sur une base (BER) soit sur plusieurs nucléotides (NER).
 - 3- Réparation par réplication pour combler la lacune (ADN POL et ligase).

7.2.1. Réparation par excision de base (BER)

Ce système de réparation est capable de réparer par exemple les désaminations et les dépurinations spontanées, il est communément utilisé pour éliminer par excision les bases endommagées.

Dans ce processus, une ADN glycosylase reconnaît la base modifiée et l'élimine par clivage de la liaison N-Glycosidique. Lorsque ces ADN-glycosylases agissent, elles aboutissent à la formation d'un site AP (Figure 70).

Il faut ensuite réparer le site AP. Cette réparation est faite avec deux autres enzymes. Tout d'abord l'AP-endonucléase qui a pour rôle de couper le squelette désoxyribose phosphate contenant le site AP, et donc d'enlever le sucre. Il y a donc clivage de la liaison phosphodiester, et retrait du site AP (Figure 70).

Une ADN polymérase remplit à nouveau l'espace libéré en utilisant la base complémentaire comme matrice. Enfin, une ADN ligase suture le brin réparé (Figure 70).

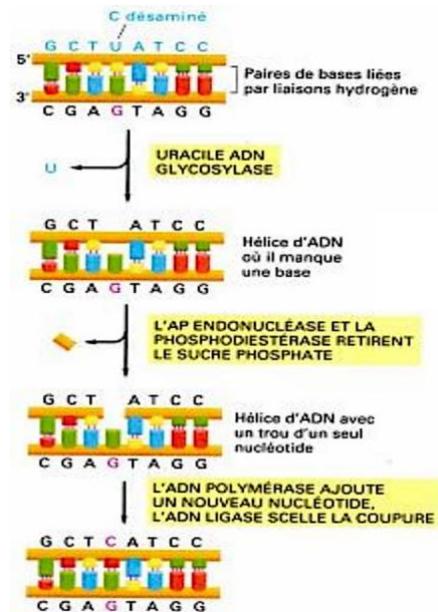


Figure 70

7.2.2. Réparation par excision de nucléotides (NER)

- Ce deuxième mécanisme est basé sur le même principe, mais il s'adresse à des lésions ou des modifications structurales importantes (ex : dimères de thymine ou pyrimidines).
- Chez *E. coli*, le NER est catalysé par un complexe multi-enzymatique : l'exonucléase ABC.
- Les constituants de ce complexe sont des protéines codées par *uvrA*, *uvrB* et *uvrC* (Figure 71).
- Ce complexe se positionne sur la lésion et reconnaît la modification. Après s'être fixé, il induit une distorsion de la double hélice.
- L'ADN hélicase permet de séparer les deux brins et de les dissocier, ce qui permet à l'exonucléase de cliver un fragment de 13 nt, contenant les lésions.
- L'ADN POL I se fixe sur l'extrémité 3'OH libre pour synthétiser l'ADN complémentaire et anti parallèle.
- La dernière liaison phosphodiester est réalisée par l'ADN ligase.

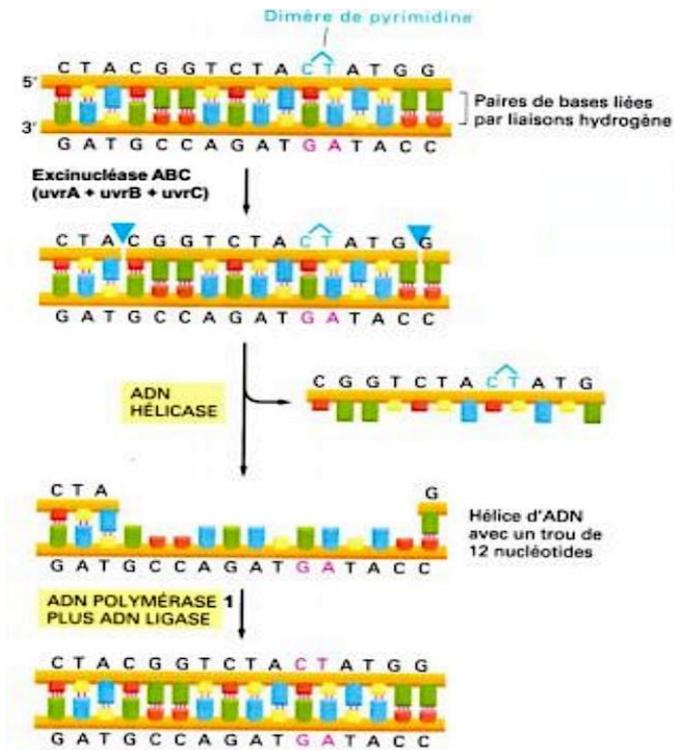


Figure 71

8. Réparation des lésions non réparées par les systèmes précédents

8.1. Réparation par recombinaison homologue

- C'est un système de réparation constitutif Il agit contre les lésions majeures de l'ADN, lorsque le système NER est débordé ou défectueux.
- S'il n'y a pas de réparation avant répllication, une lacune (post-répllicative) est laissée par l'ADN polymérase, lorsqu'elle rencontre une lésion. Cette lacune va être réparée par recombinaison homologue.

8.1.1. Mécanisme chez les procaryotes

- Lors de la répllication, une lacune est laissée sur le brin fils en face du dimère. Cette lacune est remplie par la séquence du brin parental opposé identique à ce brin fils, qui est prélevée par la protéine de recombinaison REC A (Figure 72).
- Cela va fournir la séquence correcte et créer une deuxième lacune sur le brin parental opposé. Puis le dimère est excisé par des enzymes d'excision. Les deux lacunes (à la place du dimère de thymines et sur le brin parental opposé) sont ensuite corrigées grâce à l'ADN polymérase qui synthétise la séquence complémentaire et antiparallèle de chaque brin parental.

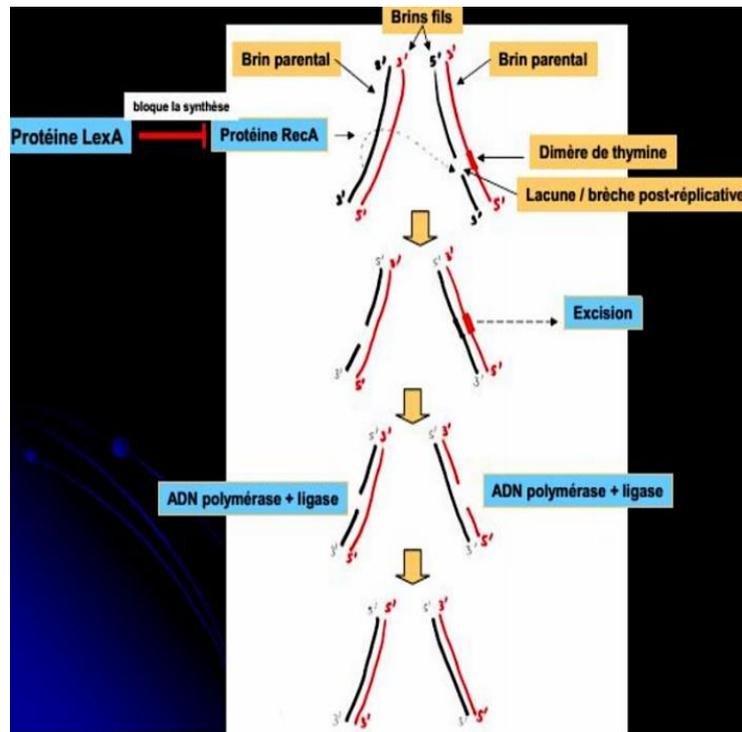


Figure 72

8.2. Le système SOS

- Si les dommages dans l'ADN sont trop importants : le système SOS se met en place. C'est le dernier système pour tenter de réparer les dommages de l'ADN.
- C'est un système de réparation par recombinaison homologe, sauf qu'il est inductible.
- Lorsque les systèmes de recombinaisons sont débordés, la réplication est stoppée. Le système SOS se met en place.
- Il active la deuxième fonction de la protéine REC A : son activité protéolytique (=dégradation de protéines). Cette activité aboutit à la dégradation de son propre répresseur LEX A. Il y a alors synthèse de protéines REC A et d'une vingtaine d'autres protéines issues des gènes SOS. Cela permet de stimuler le système de recombinaison homologe.

Transferts des gènes

VIII

Les transferts horizontaux d'information génétique chez les bactéries

69

1. Les transferts horizontaux d'information génétique chez les bactéries

- La transmission de gènes de parents à enfants est qualifiée de verticale.
- Des gènes peuvent aussi être échangés entre organismes de la même génération : ce sont des transferts horizontaux (aussi appelé transferts latéraux).



Définition

- Le transfert horizontal correspond à l'acquisition de matériel génétique à partir d'un donneur non ascendant (transferts d'information génétiques qui s'effectuent par une autre voie que celle de la reproduction).
- Chez les bactéries, le transfert horizontal joue un rôle prépondérant dans leur diversification et constitue un mécanisme d'évolution rapide.
- En effet, ce phénomène permet l'acquisition de gènes prêts-à-l'emploi en un seul événement à comparer aux mécanismes évolutifs nécessitant l'accumulation de mutations ponctuelles

L'ADN exogène peut être internalisé par trois mécanismes différents avant d'être recombiné au génome de la bactérie hôte (Figure 73) :

1. L'ADN peut provenir d'une bactérie morte (dont la paroi est ouverte) et être intégré par la bactérie receveuse : c'est la transformation bactérienne (c'est le mécanisme responsable des observations de Griffith).
2. Il peut également s'effectuer par l'intermédiaire d'un virus bactérien ou phage : c'est le transfert viral (la transduction).
3. Ou encore via un lien physique entre les deux bactéries : c'est la conjugaison bactérienne

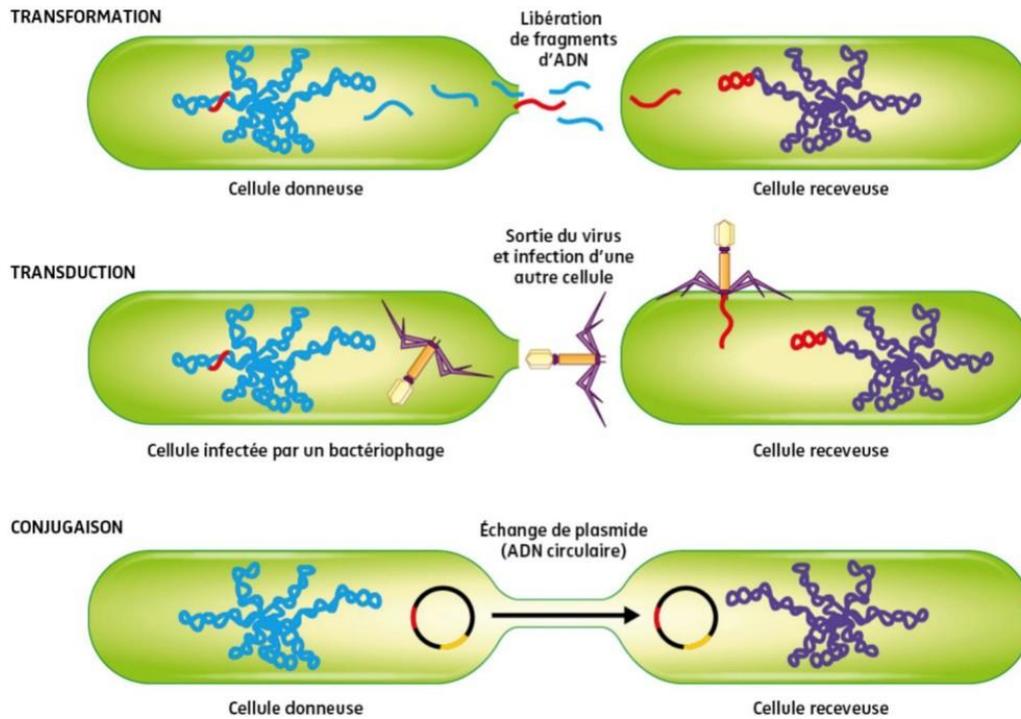


Figure 73

1.1. La Transformation

- C'est un processus permettant la capture et le transfert par diffusion d'ADN exogène nu par une bactérie et son incorporation au génome de celle-ci (Figure 74).
- La transformation est très répandue dans le monde bactérien aussi bien chez les Gram positif que négatif.
- La transformation requiert le développement d'un état physiologique particulier, génétiquement contrôlé : la compétence.

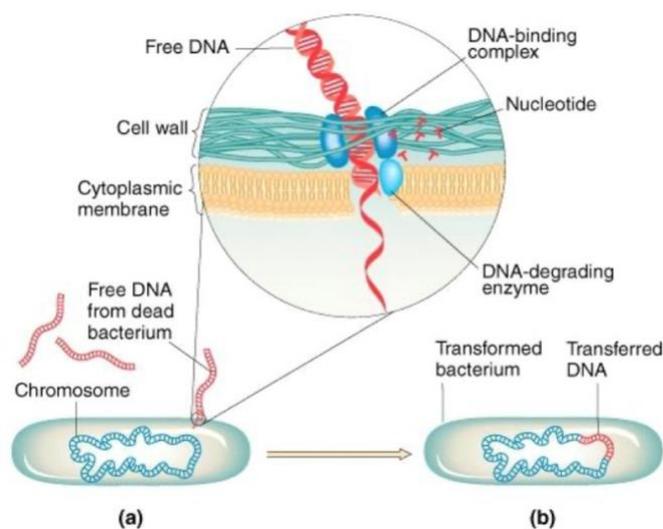


Figure 74

1.1.1. Notion de compétence

Une cellule qui est capable d'absorber de l'ADN et d'être transformée est dite compétente, et cette capacité est génétiquement déterminée. Il existe deux types de compétence :

A- La compétence naturelle

B- Compétence artificielle

a) La compétence naturelle

- La compétence chez la plupart des bactéries naturellement transformables est régulée, et des protéines particulières jouent un rôle dans la capture et le traitement de l'ADN.
- Ces protéines spécifiques de la compétence incluent une protéine membranaire fixant l'ADN, une autolysine, et différentes nucléases.
- Certaines espèces bactériennes sont naturellement compétentes (*Bacillus subtilis*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Haemophilus influenzae*, etc.); d'autres ne le sont pas mais peuvent le devenir expérimentalement (*Escherichia coli*), par exemple par un traitement au CaCl_2 qui rend la membrane cellulaire plus perméable au DNA.

i Étapes initiales du processus de transformation naturelle chez *S. pneumoniae*

1. La cellule peut capter un ADN double brin via un pseudo-pilus, formé par un ensemble de protéines codées par l'opéron *comG*.
2. Après capture, il interagit avec le récepteur membranaire ComEA avant d'être transféré à la nucléase EndA.
3. La nucléase EndA dégrade l'un des deux brins assurant ainsi l'internalisation d'un ADN simple brin par le pore transmembranaire ComEC.
4. Le brin d'ADN est enfin transporté de façon active via l'hydrolyse de molécules d'ATP dans le cytoplasme par ComFA
5. Une fois internalisé, DprA se lie à l'ADN, permettre la fixation et la polymérisation de RecA et la recombinaison homologue par recherche d'homologie et échange de brin.
6. De plus, des protéines SsbB peuvent se fixer à l'ADN simple brin afin de le protéger des nucléases cellulaires et de créer un réservoir d'ADN pour des recombinaisons homologues ultérieures (Figure 75).

Le type d'interaction du pseudo-pilus avec l'ADN reste encore inconnu!

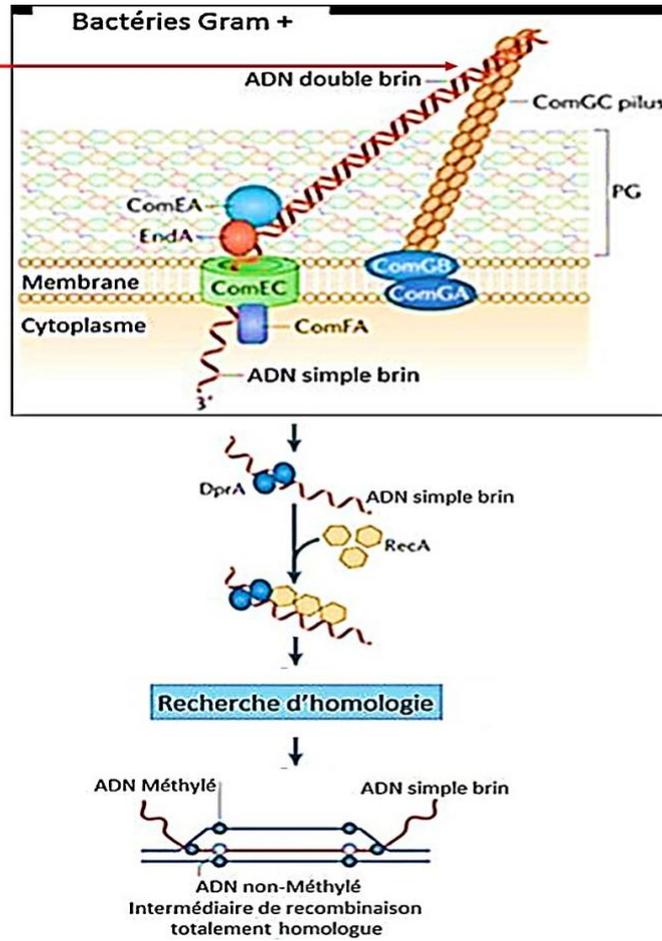


Figure 75

b) Compétence artificielle

- Un traitement des cellules non compétentes par du chlorure de calcium ou de rubidium ou par un choc thermique peut provoquer des altérations de l'enveloppe bactérienne augmentant ainsi leur capacité d'absorption d'ADN, les rendant ainsi compétentes (Figure 76).

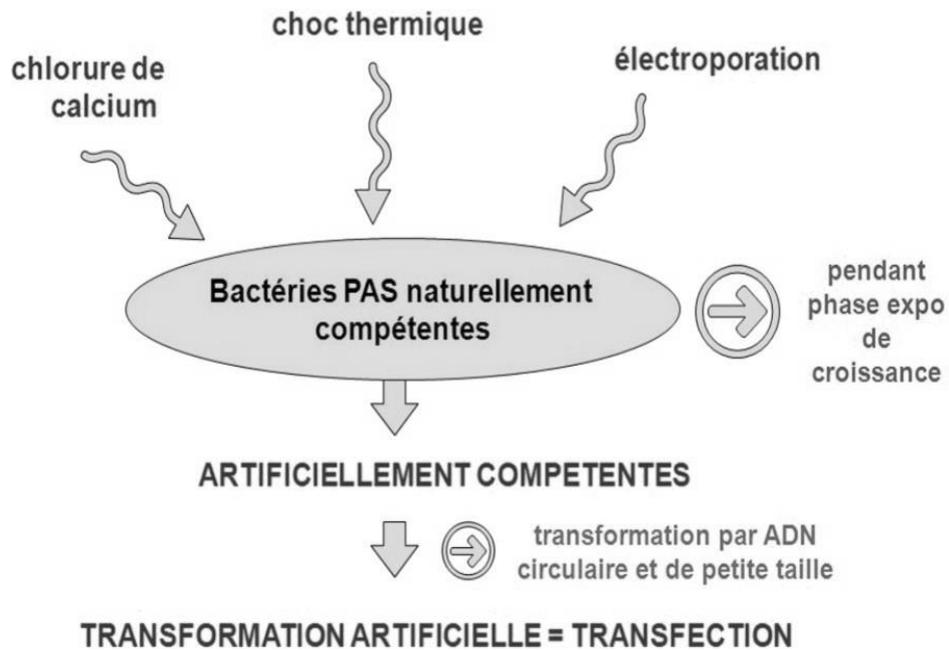


Figure 76

1.2. La transduction

- La transduction est un transfert génétique par l'intermédiaire d'un virus (bactériophage) qui est le vecteur de l'exogène (matériel génétique du donneur).

▼ Rappel : Les bactériophages

- Sont des particules infectieuses parasites strictes des bactéries ; ils sont constitués d'un génome (ADN mono- ou bicaténaire, ou ARN) protégé par une capsid protéinique.
- Il existe différents types morphologiques de capsid.
- Certains ont une action lytique, d'autres ont une action lysogénique.

1.2.1. Les différents types de transduction

Le transfert de gènes médié par les phages est connu pour se produire par trois mécanismes différents de transduction :

1. Spécialisé
2. Généralisé
3. Latérale

a) La transduction spécialisée

- Si une bactérie contenant du prophage est exposée à des facteurs de stress, tels que la lumière UV, des conditions de faibles nutriments ou des produits chimiques, le prophage peut spontanément s'extraire du génome hôte et entrer dans le cycle lytique.

- Ce processus, cependant, n'est pas parfait et le prophage peut parfois laisser des portions de son ADN derrière lui ou emporter un fragment adjacent du chromosome avec lui lorsqu'il se recircularise.
- Pour un phage donné et pour une espèce bactérienne donnée, ce sont toujours les mêmes gènes qui sont transduits.
- Le mécanisme est considéré comme spécialisé car il se limite au transfert de gènes bactériens immédiatement adjacents au prophage intégré.
- Dans le système *E. coli*/ phage lambda (λ), le prophage est inséré en un site spécifique connu (entre le gène gal et le gène bio) le prophage par erreur emporte soit le gène gal soit le gène bio (Figure 77)

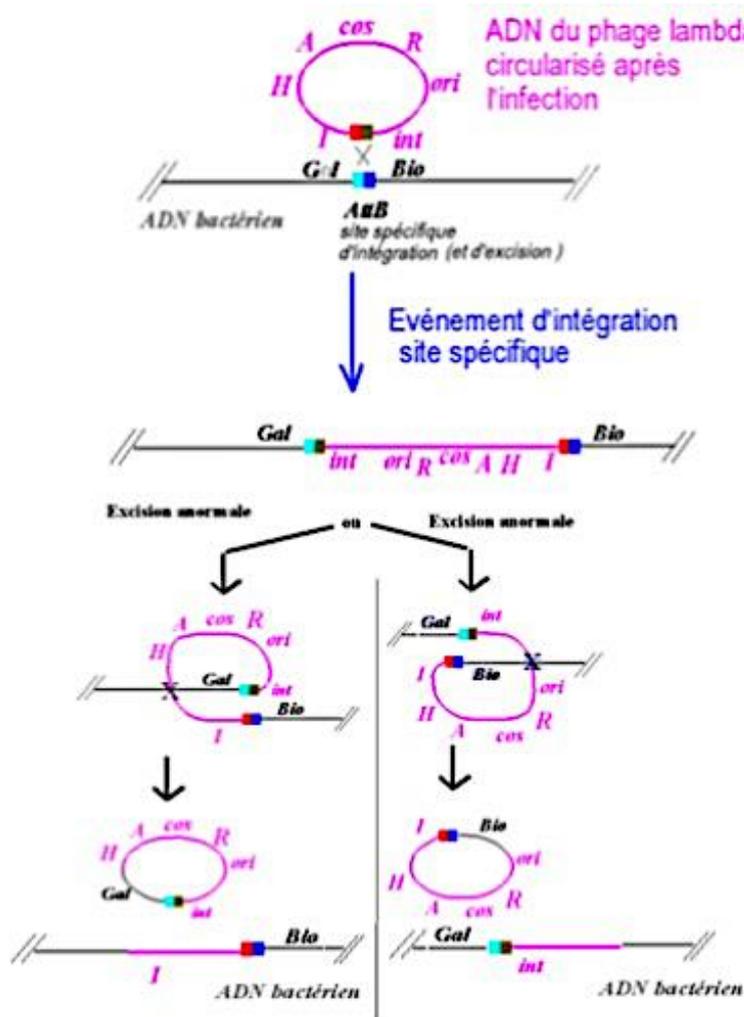


Figure 77

b) La transduction généralisée

- Au cours de l'infection, l'ADN de phage sert de matrice pour la synthèse de nouvelles molécules d'ADN viral. Il dirige également la synthèse de la capside protéique du phage.
- Pendant ce temps le chromosome bactérien est fragmenté par des enzymes virales, il arrive que, durant l'assemblage des phages, quelques fragments d'ADN bactériens soient enfermés par erreur à

l'intérieur de la capsidie protéique des bactériophages.

- Certaines particules virales ainsi formées contiennent alors de l'ADN bactérien plutôt que l'ADN phagique.
- Quand les particules virales libérées infectent par la suite une nouvelle population de bactéries, il y a à l'occasion transfert de gènes bactériens à des cellules receveuses.
- Le mécanisme est considéré comme généralisé car n'importe quel gène bactérien a théoriquement, la même probabilité d'être transduit (Figure 78).

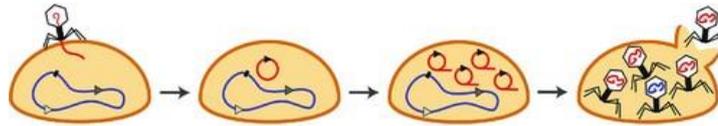


Figure 78

1.3. La conjugaison

- La conjugaison est processus par lequel un ADN est transféré d'une bactérie donneuse à une bactérie réceptrice. Il faut qu'il y ait un contact physique.
- La capacité d'une bactérie à donner son ADN dépend de la présence du plasmide F (facteur de fertilité).
- Le plasmide F est composé de 100Kb, est présent en une seule copie par bactérie et est extrêmement stable, donc jamais perdu pour la population, car il a développé des systèmes :
- De Partition : il va se répartir pareillement dans les deux cellules filles.
- De Killing : si le système de partition ne fonctionne pas et qu'une cellule fille ne porte pas ce plasmide, elle sera tuée.

1.3.1. La Découverte de la conjugaison



Joshua Lederberg



Edward L. Tatum

Figure 79

L'expérience de Joshua Lederberg et Edward L. Tatum en 1946 apporta la preuve initiale de la conjugaison bactérienne, le transfert de matériel génétique par contact direct entre les cellules.

Lederberg était persuadé que le transfert de matériel génétique chez les bactéries ne se faisait pas uniquement par multiplication cellulaire.

1- Il mit en place une expérience en mélangeant, en milieu liquide, deux mutants d'*E.coli*:

- Le premier étant déficient pour la synthèse de thréonine, leucine et thiamine,
- Le second déficient pour la synthèse de la biotine, de la phénylalanine et de la cystéine

2- Après quelques heures, il étala la culture sur un milieu minimum

3- 24 heures après, il a pu observer la croissance de certaines colonies, qui ne pouvaient être autre que Thr+, Leu+, Thiamine+, Biotine+, Phé+, et Cys+ (Figure 80).

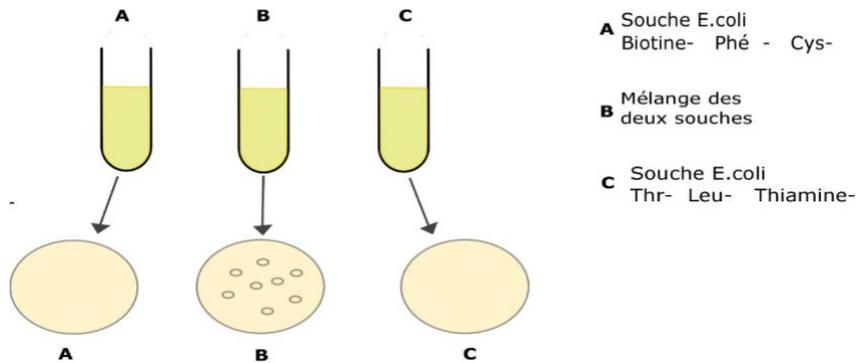


Figure 80

- Les deux mutant d'*E.coli* (boite A et C) sont déficients pour certains facteur de croissance, les empêchant ainsi de pousser sur un milieu minimum.
- Sur la boite B, on observe quelques colonies. On peut donc dire que ces conjuguant ont récupéré les gènes du type sauvage, sans déficience dans les facteurs de croissance et ce par recombinaison homologue.

1.3.2. Le croisement F+ X F-

- En 1952, William Hayes démontra que le transfert de gènes observé par Lederberg et Tatum's'effectuait dans un sens déterminé.
- En fait, il existe des souches définies comme donneuses (F+) et receveuse (F-), et le transfert de gènes n'est pas réciproque.
- Les souches F + contient un facteur F extra-chromosomique qui porte les gènes pour la formation des pili et pour le transfert de plasmide (Figure 81).

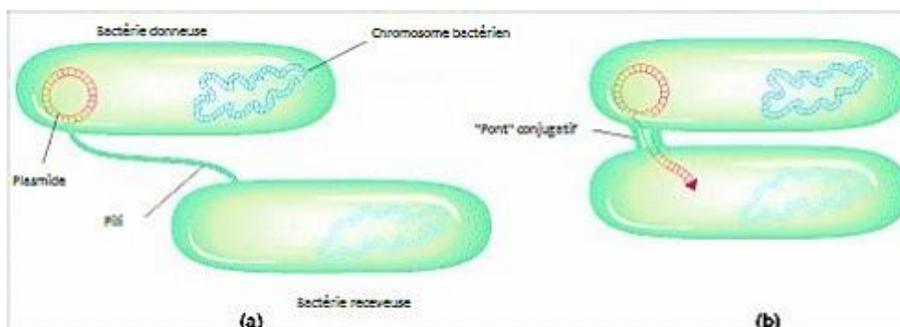


Figure 81

1.3.3. La conjugaison chez les Gram négatives

- Chez les bactéries à Gram négatif, le plasmide porte des gènes qui dirigent la synthèse de pili sexuels, prolongements de la surface de la cellule donneuse qui établissent un pont avec la cellule receveuse et contribuent à mettre les deux bactéries en contact direct.
- Au cours de la conjugaison le plasmide se réplique en même temps qu'une copie monocaténaire de son ADN est transférée à la cellule receveuse ou le brin complémentaire est synthétisé.
- On constate donc que les bactéries donneuses ayant des facteurs F (cellule F+) transfèrent ce plasmide à des cellules receveuses (cellule F-) qui deviennent alors F+.

1.3.4. La conjugaison chez les Gram positives

- La conjugaison plasmidique a aussi lieu chez les Gram positifs, en particulier dans les genres *Staphylococcus*, *Streptococcus*, *Streptomyces* et *Clostridium*.
- Dans certains cas (*Enterococcus*), il a été montré que le mécanisme est différent :
 - Le contact n'est pas réalisé par les pili comme chez les Gram négatifs.
 - Les receveurs relâchent dans le milieu externe des peptides diffusibles, les phéromones, qui induisent chez les donneurs (porteurs de plasmides conjugatifs) la synthèse d'une protéine de surface.
 - Cette protéine est une sorte d'adhésine : elle provoque la formation de grands agrégats entre donneurs et receveurs.
 - Ces agrégats sont le site des transferts génétiques dont les mécanismes demeurent encore inconnus.

1.3.5. Les bactérie Hfr

- Le plasmide F étant capable de s'intégrer sur le chromosome bactérien, une bactérie portant le facteur F sur son propre chromosome est appelée Hfr (haute fréquence de recombinaison) (Figure 82).

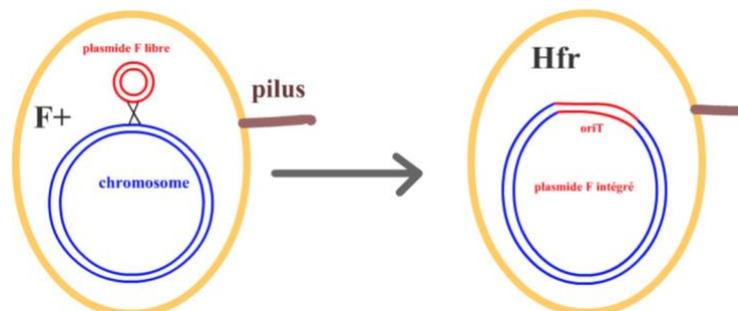


Figure 82

1.3.6. La conjugaison Hfr/F-

- Lors d'un croisement entre une bactérie Hfr et une bactérie F-, les différents gènes seront transférés l'un après l'autre et c'est uniquement lorsque l'ensemble du chromosome est transféré (environ 2 heures) que le facteur F est transféré lui aussi entièrement.
- Il est cependant très rare que l'ensemble du chromosome soit transmis, ainsi la bactérie receveuse ne reçoit pas le facteur F et reste donc F-.
- Elle a pu cependant acquérir, par recombinaison homologe, une partie des gènes de la bactérie donneuse (Figure 83).

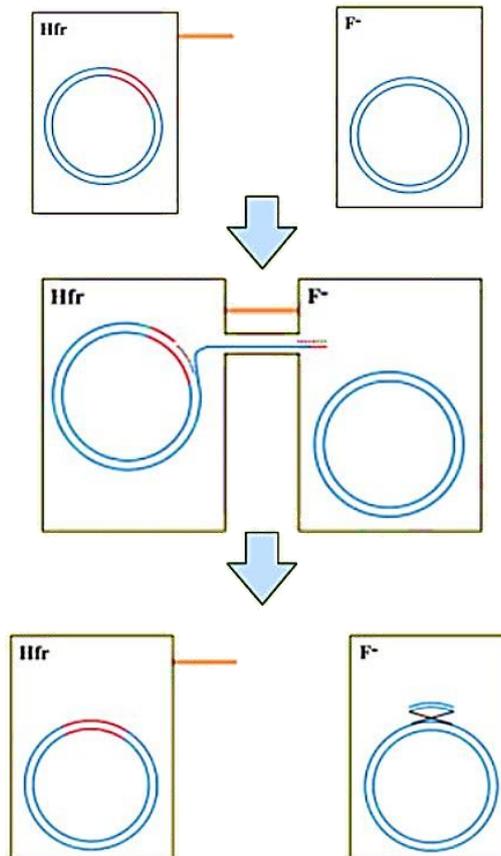


Figure 83

Les références bibliographiques

IX

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2014). *Molecular Biology of the Cell* (6th Ed.). Garland Science. 1464.
- Beaumont, S. (2007). *Biologie moléculaire PCEM1*. Dunod, Paris. 279-286.
- Bell, S. P., & Dutta, A. (2002). DNA replication in eukaryotic cells. *Annual Review of Biochemistry*, 71(1), 333-374.
- Brent, M. R., & Guigó, R. (2004). Recent advances in gene structure prediction. *Current Opinion in Structural Biology*, 14(3), 264-272.
- Clark, D. (2005). *Molecular Biology: Understanding the Genetic Revolution*. Elsevier Academic Press. 784.
- Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2), 113-127.
- Kunkel, T. A., & Erie, D. A. (2015). DNA Mismatch Repair. *Annual Review of Biochemistry*, 84, 649-680
- Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A., Krieger, M., Scott, M. P., Zipursky, S. L., & Darnell, J. (2003). *Molecular Cell Biology* (6th Ed.). 937.
- Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, 15, 423–437.
- O'Donnell, M., Langston, L., & Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harbor Perspectives in Biology*, 5(7), a010108.
- Passarge, E. (2007). *Color Atlas of Genetics* (3rd Ed.). Flexibook. Thieme Stuttgart, NY, USA. 486.
- Prescott, M. L., Harley, J. P., & Klein, D. A. (2002). *Microbiology* (5th Ed.). The McGraw–Hill Companies. 1139.
- Prescott, M. L., Willey, J. M., Sherwood, M. L., & Woolverton, J. C. (2018). *Microbiologie* (5ème édition). 1120.
- Raven, P. H., Losos, J. B., Mason, K. A., & Duncan, T. (2020). *Biologie*. 1480.
- Watson, J., Baker, T., Bell, S., Gann, A., Levine, M., & Losick, R. (2009). *Biologie Moléculaire du gène* (6e Ed.). Pearson Education- Paris-France. 688.
- Yandell, M., & Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics*, 13(5), 329-342.