

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la
Recherche Scientifique

Université Ferhat Abbas-Sétif1



الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي و البحث العلمي

جامعة فرحات عباس-سطيف1

Cours de Biologie Moléculaire et Génie Génétique

Dr. Abdelhakim Aouf

Destiné aux étudiants de 3^e année Licence de Microbiologie

2015-2016

*Life begins with cells and cell begins with the secret of
life:
"DNA"*

★

*DNA begins with four nucleotides and nucleotide begins
with five atoms:
"H, C, O, N, P"*

★

*Atoms begins with electrons, protons, neutrons and
other substructures*

★

So, Try to understand Life

| <h1>Sommaire</h1> | | P A G E |
|--------------------|--|----------------------------|
| Première partie | <h2>Biologie Moléculaire</h2> | |
| Chapitre I | Support de l'information génétique | 1 |
| 1 | Introduction | 1 |
| 2 | ADN comme matériel génétique | 1 |
| 3 | Composition chimique de l'ADN | 2 |
| 4 | Structure de l'ADN | 4 |
| 5 | Interactions entre C-G et A-T | 5 |
| 5-1 | Liaison hydrogène | 5 |
| 5-2 | Stabilisation de la double hélice de l'ADN | 5 |
| 6 | Formes de l'ADN | 5 |
| 7 | Propriétés physico-chimiques de l'ADN | 7 |
| 7-1 | Température de fusion | 7 |
| 7-2 | Facteurs influençant la T _m | 8 |
| 7-3 | Absorption de la lumière ultraviolette | 8 |
| Chapitre II | Expression de l'information génétique | 9 |
| 1 | Transcription | 9 |
| 1-1 | Gène | 10 |
| 1-2 | ARN polymérase | 10 |
| 1-3 | Étapes de la transcription chez les procaryotes | 12 |
| 1-3-1 | Initiation | 12 |
| 1-3-1-1 | Séquences consensus | 13 |
| 1-3-1-2 | Fixation de l'ARN polymérase | 13 |
| 1-3-1-3 | Éléments trans | 13 |
| 1-3-2 | Elongation | 13 |
| 1-3-3 | Terminaison | 14 |
| 1-4 | Transcription chez les eucaryotes | 16 |
| 1-4-1 | Étapes de la transcription chez les eucaryotes | 17 |
| 1-4-1-1 | Initiation | 17 |
| 1-4-1-2 | Elongation | 19 |
| 1-4-1-3 | Terminaison de la transcription | 20 |
| 1-5 | Maturation des ARNm | 20 |
| 1-5-1 | Ribozymes | 21 |
| 1-5-2 | Splicéosome | 22 |
| 2 | Traduction | 24 |
| 2-1 | ARN messenger et le code génétique | 24 |
| 2-1-1 | Code génétique | 25 |
| 2-1-2 | Reconnaissance, l'activation et le chargement des ARNt | 25 |

| | | |
|-----------------------------|---|-----------|
| 2-2 | Ribosomes | 27 |
| 2-3 | Rôles des ARN dans la traduction | 28 |
| 3 | Étapes de la traduction chez les procaryotes | 29 |
| 3-1 | Initiation de la traduction | 29 |
| 3-2 | Élongation | 31 |
| 3-3 | Terminaison de la traduction | 34 |
| 4 | Traduction chez les eucaryotes | 34 |
| 4-1 | Initiation de la traduction chez les eucaryotes | 34 |
| 4-2 | Élongation | 37 |
| 4-3 | Terminaison de la traduction | 37 |
| Chapitre III | Régulation de l'expression génétique | 38 |
| 1 | Introduction | 38 |
| 2 | Régulation de l'activité enzymatique | 39 |
| 2-1 | Retroinhibition (Feed Back or end-product inhibition) | 40 |
| 2-2 | Modification covalente des enzymes | 42 |
| 3 | Protéines se liants à l'ADN | 44 |
| 3-1 | Structure des protéines se liant à l'ADN | 44 |
| 4 | Régulation transcriptionnelle | 45 |
| 4-1 | Contrôle négatif de la transcription "Répression et induction" | 46 |
| 4-1-1 | Régulation de l'expression de l'opéron <i>lacZYA</i> | 46 |
| 4-1-2 | Régulation de l'expression de l'opéron <i>argCBH</i> | 49 |
| 4-2 | Contrôle positif de la transcription | 50 |
| 4-2-1 | Régulation de l'expression de l'opéron <i>malEFG</i> | 50 |
| 5 | Régulon | 50 |
| 6 | Régulation positive et négative de l'opéron <i>araBAD</i> | 50 |
| 7 | Contrôle de l'expression génétique par atténuation | 51 |
| 8 | Régulation par des facteurs sigma (σ) alternatifs | 53 |
| 9 | Détection du quorum (quorum sensing: QS) | 53 |
| 10 | Contrôle de l'expression génétique par un système de régulation à deux composants | 55 |
| 11 | ARN de régulation et les riboswitch | 56 |
| 11-1 | Riboswitchs | 57 |
| 2^e partie | Génie génétique | |
| | Glossaire | 58 |
| | Préambule | 61 |
| CHAPITRE I | Enzymes de restriction et de modification des acides nucléiques | 62 |
| 1 | Enzymes de restriction | 62 |
| 1-1 | Nomenclature des enzymes de restriction | 62 |
| 1-2 | Classification des enzymes de restriction | 65 |
| 1-3 | Enzymes de restriction utilisées en génie génétique | 65 |
| 1-4 | Utilisation des endonucléases de restriction pour établir la mappe de restriction (restriction mapping) | 66 |

| | | |
|---------------------|---|-----------|
| 1-5 | Isoschizomères et famille d'enzymes | 67 |
| 1-4 | Méthylation de l'ADN | 68 |
| 1-4-1 | Dam méthylase | 68 |
| 1-4-2 | Dcm méthylase | 68 |
| 2 | Enzymes de modification des acides nucléiques | 69 |
| 2-1 | Ribonucléases A (RNase): EC 3.1.27.5 | 70 |
| 2-2 | Désoxyribonucléase I (DNase I): EC 3.1.21.1 | 70 |
| 2-3 | Nucléase S1 : EC 3.1.30.1 | 71 |
| 2-4 | Exonucléase de phage λ : EC 3.1.11.3 | 72 |
| 2-5 | Exonucléase III: EC 3.1.11.2 | 72 |
| 2-6 | Ribonucléase T1: EC 3.1.27.3 | 72 |
| 2-7 | Ribonucléase T2: EC 3.1.27.1 | 73 |
| 2-8 | Nucléase Bal31 | 73 |
| 2-9 | ADN ligase EC 6.5.1.1 | 74 |
| CHAPITRE II | Hôtes de clonage | 76 |
| 1 | L'hôte idéal | 76 |
| 2 | Méthodes d'introduction de l'ADN à cloner dans la cellule hôte | 77 |
| 2-1 | Électroporation | 77 |
| 2-2 | Un canon à Particules ou Biolistics | 77 |
| 2-3 | Microinjection | 77 |
| CHAPITRE III | Vecteurs de clonage | 79 |
| 1 | Vecteurs bactériens | 79 |
| 1-1 | Plasmides | 79 |
| 1-1-1 | Plasmide pBR322 (Vecteur de première génération) | 80 |
| 1-1-1-1 | Caractéristiques du plasmide pBR322 | 81 |
| 1-1-2 | Vecteurs de seconde génération | 81 |
| 1-2 | Bactériophage λ | 83 |
| 1-2-1 | Principales étapes de la construction d'un vecteur phagique | 85 |
| 1-3 | Bactériophage M13 | 86 |
| 1-4 | Cosmides | 87 |
| 1-5 | Phagmides ou phasmides | 89 |
| 1-6 | Chromosomes artificiels bactériens (Bacterial Artificial Chromosomes) BAC | 89 |
| 2 | Chromosomes artificiels des levures (Yeast Artificial Chromosomes) YAC | 90 |
| 3 | Chromosomes artificiels dérivé de P1 "PAC" | 90 |
| 4 | Plasmide Ti | 91 |
| 5 | Vecteur spécifiques | 92 |
| 5-1 | Vecteurs navettes et d'expression | 92 |
| 6 | Régulation de la transcription des vecteurs d'expression | 93 |
| 7 | Traduction de gènes clonés | 93 |
| 8 | Expression de gènes de mammifères chez les bactéries | 93 |
| 9 | Production d'insuline par génie génétique | 94 |
| Chapitre IV | Techniques de biologie moléculaire | 96 |

| | | |
|-----------------------------|---|------------|
| 1 | Extraction et purification des acides nucléiques | 96 |
| 1-1 | Source d'ADN à cloner | 96 |
| 1-2 | Préparation de l'ADN à partir du sang total | 97 |
| 1-2-1 | Protocole expérimental | 98 |
| 1-2-2 | Dosage des acides nucléiques | 98 |
| 2 | Electrophorèse | 99 |
| 2-1 | Introduction | 99 |
| 2-2 | Définition | 99 |
| 2-3 | Détermination de la taille des fragments | 100 |
| 3 | Hybridation des acides nucléiques | 101 |
| 3-1 | Formules utilisées pour calculer la Th | 102 |
| 3-2 | Facteurs influençant l'hybridation | 102 |
| 3-3 | Types d'hybridation | 102 |
| 3-3-1 | Hybridation d'ADN sur support solide : SOUTHERN BLOT | 102 |
| 3-3-2 | Hybridation in situ sur chromosome | 103 |
| 3-3-3 | Hybridation sur colonie de bactéries, sur plage de lyse | 103 |
| 4 | Réaction de Polymérisation en Chaîne "PCR" | 104 |
| 4-1 | Historique | 104 |
| 4-2 | Étapes de d'amplification de l'ADN par PCR | 104 |
| 4-3 | Thermocycleur | 105 |
| 4-4 | Applications et la sensibilité de la PCR | 106 |
| 5 | Séquençage | 107 |
| 5-1 | Méthode de Sanger | 107 |
| 5-2 | Marquage des produits d'extension | 108 |
| 5-3 | Séquençage d'un produit de PCR | 109 |
| 6 | ADN C " ADN complémentaire" | 109 |
| 6-1 | Caractéristiques de l'ARNm eucaryote | 109 |
| 6-2 | Synthèse de l'ADNc | 109 |
| 6-3 | Insertion de l'ADNc dans un vecteur approprié | 112 |
| 3^e Partie | Introduction à la bioinformatique | 113 |
| 1 | Introduction | 113 |
| 2 | Historique | 113 |
| 3 | Banques de données biologiques | 113 |
| 4 | Principes de bases de l'alignement des séquences | 114 |
| 4-1 | Alignement et détermination du score | 115 |
| 4-2 | Alignement global | 116 |
| 4-2-1 | Algorithme de Needleman et Wunsch | 115 |
| 4-3 | Alignement local | 117 |
| 4-4 | BLAST (acronym de: Basic Local Alignment Search Tool) | 117 |
| 4-5 | Alignement de séquences multiples | 119 |
| 4-5-1 | ClustalW / ClustalX "Cluster alignment" | 119 |
| 4-5-1-1 | Étapes de l'alignement multiples de type clustal | 119 |
| 6 | Arbres phylogénétiques | 119 |

| | | |
|------------|--|------------|
| 6-1 | Méthode des distances (UPGMA: Unweighted Pair Group Method with Arithmetic Mean) | 119 |
| 6-2 | Méthode de parcimonie (Maximum Parsimony) | 120 |
| 6-3 | Méthode de maximum de vraisemblance (Maximum Likelihood- ML) | 120 |
| 7 | Recherche de motifs et de domaines | 121 |
| 8 | Prédiction et modélisation moléculaire | 121 |
| 8-1 | Prédiction de structures secondaires | 121 |
| 8-2 | Modélisation moléculaire | 122 |
| | Références bibliographiques | 124 |

Première partie:

Biologie Moléculaire

Chapitre I

Support de l'information génétique

Objectifs

- Expériences ayant conduit à la découverte de l'ADN comme matériel génétique.
- Composition chimique de l'ADN.
- Structure de l'ADN.
- Caractérisation et analyses de l'ADN.

1-Introduction

La variété des organismes vivants est extraordinaire, on estime qu'il ya de nos jours plus de 10 millions d'espèces. Chaque espèce est différente des autres. Cette différence est due au contenu en information génétique de chaque espèce. Les généticiens ont envisagé que des molécules spécifiques étaient porteuses de cette information génétique. Les êtres vivants sont subdivisés en deux groupes: les eucaryotes et les procaryotes, se différencient par la présence ou non d'un noyau.

2-ADN comme matériel génétique

En 1869: Fredrich Micher utilisa une protéase (pepsine: découverte en 1836) pour séparer le noyau du cytoplasme. Il a observé une petite tache de poudre grise qui s'était détachée d'un liquide claire jaunâtre (cytoplasme). Il l'appela nucléine.

En 1889: Le biologiste Altman a précisé que la nucléine c'est un acide nucléique.

En 1928 : L'expérience de Griffith a montrée que la souche S tuée ne tue pas la souris, mais lorsqu'elle est mélangée avec la souche R vivante, la souris injectée meurt (fig. 1). Comme conclusion la souche R est transformée en souche S, et le facteur transformant est l'ADN.

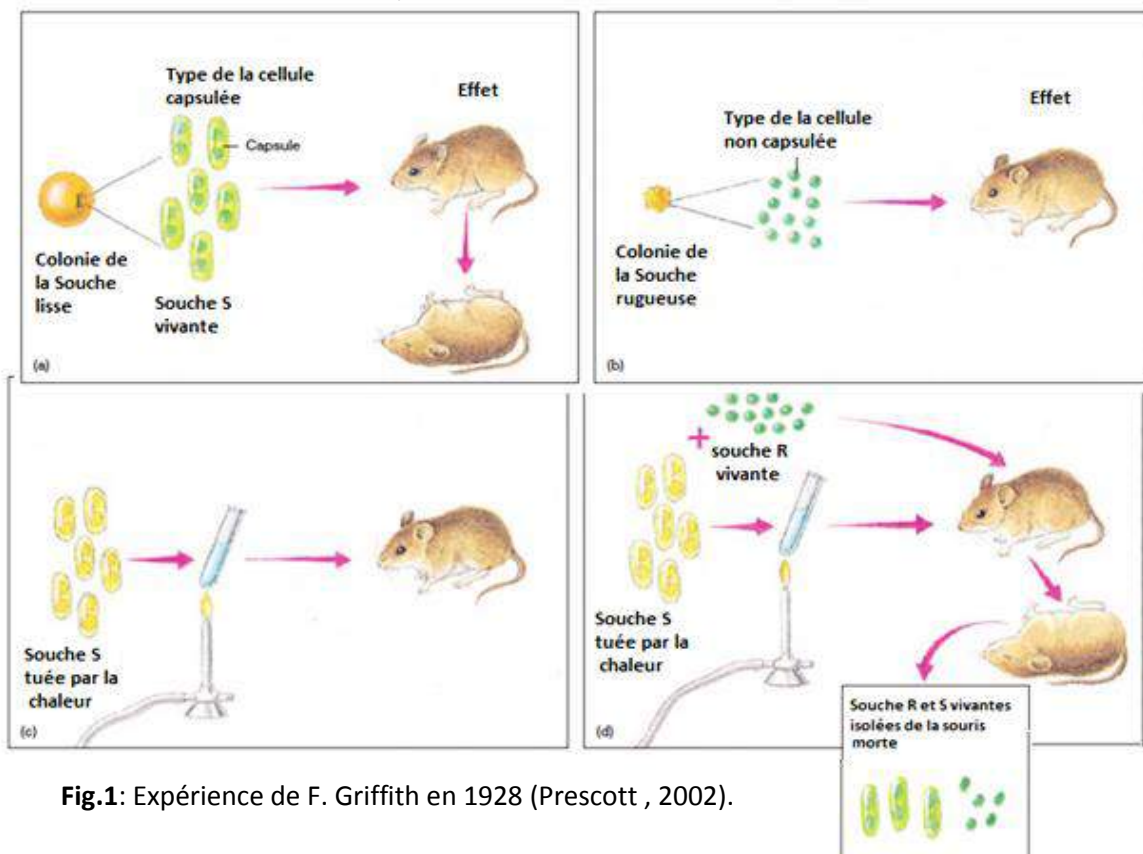


Fig.1: Expérience de F. Griffith en 1928 (Prescott , 2002).

En 1931: J. Hammerling, il travaille sur l'*Acetobularian* qui est une algue unicellulaire, verte ayant l'aspect d'une petite ombrelle (tige + chapeau). Il existe plusieurs formes selon le type d'algues.

- Il a mis les noyaux d'une souche de type A dans une souche de type B (différent chapeaux) dont les chapeaux furent coupés.
- Il a observé que des chapeaux de type A se développent sur le type B.

Les travaux de Hammerling ont démontré pour la première fois dans l'histoire que le noyau est le seul responsable de la production des êtres vivants (eucaryotes).

En 1944: Avery O.T., Macleod C. M. et Mc Carty M. (Rockefeller institute - New York) ont repris les expériences effectués en 1928 par Fred Griffith en utilisant les extraits de différentes biomolécules de la cellule S traités ou non par des enzymes spécifiques. Les résultats obtenus montrent qu'il ya transformation d'un nombre de souches R en souche S lorsque l'ADN purifié est utilisé (selon la fréquence de transformation). Les mêmes résultats ont été obtenus en dégradant les protéines et les ARN par des enzymes spécifiques. En conclusion, l'ADN est le facteur transformant et le support de l'information génétique

En 1952: Expérience de Hershey et Chase. L'utilisation de l'ADN et des protéines phagiques marqués, montre que seulement l'ADN qui est injecté dans les cellules bactériennes, entraînant la production des phages. C'est donc l'ADN qui est porteur de l'information génétique du virus.

En 1946: Lederberg et Tatum ont recombinié génétiquement des bactéries de souches auxotrophes incapables de pousser sur un milieu dépourvu de facteurs de croissance).

En 1950: Chargaff a fait l'équivalence en bases dans l'ADN des animaux, végétaux, bactéries et phages. En conclusion il a trouvé que :

- Le rapport $A + G / T + C \approx 1$ (purines / pyrimidines ≈ 1) chez tous les organismes testés.
- Le rapport AT / GC est variable selon les espèces.

En 1952-1953: - Franklin a obtenu une excellente photographie d'ADN par diffraction des rayons X.

- La même année Watson et Crick utilisant les données de Chargaff et l'image obtenue par Franklin ont établi le modèle de la double hélice (la forme B).

3-Composition chimique de l'ADN

Les travaux de Kornberg révélèrent que les précurseurs de l'ADN étaient des nucléotides riches en énergie (dNTP, dCTP, dTTP, dGTP). Chaque nucléotide est composé d'acide phosphoriques (3), base azotée (A, T, C, G) et un pentose(2'-désoxy- β -D-Ribose). La liaison formée entre deux phosphates donne un anhydride avec une liaison très riche en énergie, et la liaison entre le groupement acide du phosphate et l'hydroxyle du sucre donne une liaison ester. La liaison formée entre le sucre et la base est une liaison osidique de type β -N-glycosidique (fig. 2, 3 et 4).

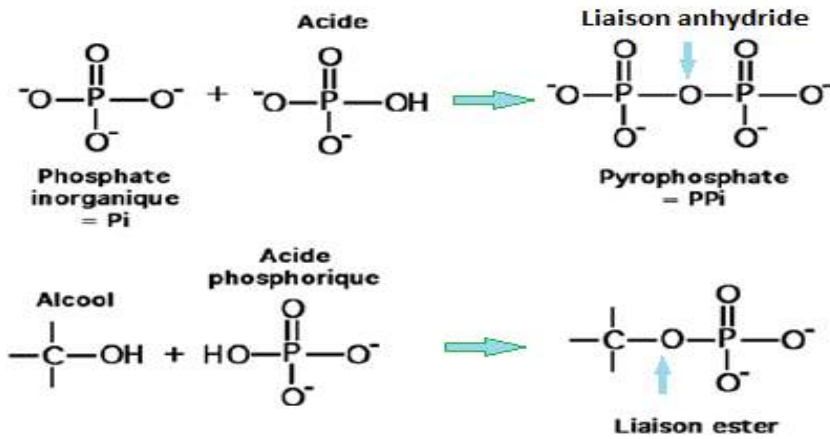


Fig.2: Formes du phosphate et les liaisons formées.

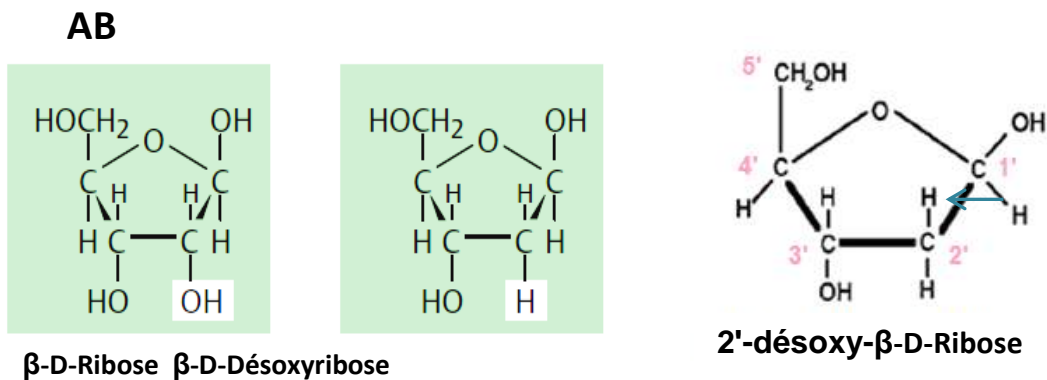


Fig.3:A- Les pentoses qui rentrent dans la composition des acides nucléiques (ADN,ARN).B-Numérotation d'un sucre, le prime (') est ajouté dans la numérotation pour la distinguer à celle des bases.

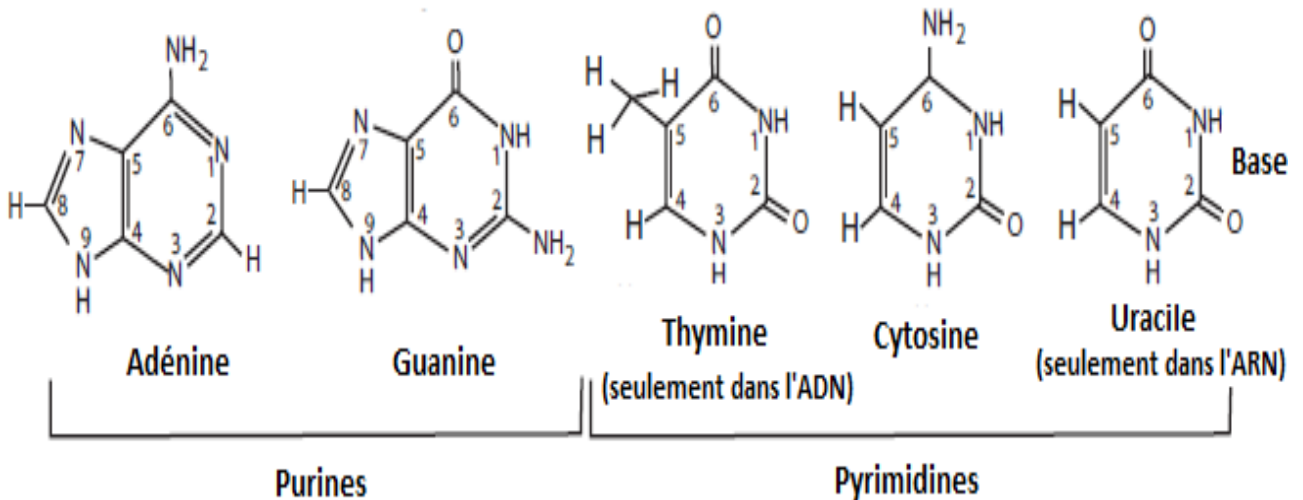


Fig.4:Les cinq bases azotées de l'ADN et l'ARN (Purines [A et G], Pyrimidines [C, T et U])

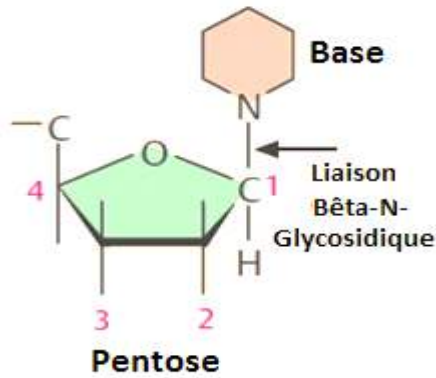


Fig.5 : Nucléoside. Le sucre se lie à la base azotée par une liaison impliquant un des azotes (l'azote n°1 des pyrimidines ou azote n°9 des purines) et le carbone 1' de l'ose (carbone réducteur ou fonction semi-acétalique). C'est une liaison N-osidique.

4-Structure de l'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 (La forme B a un diamètre de 2,47 nm) (fig. 6).

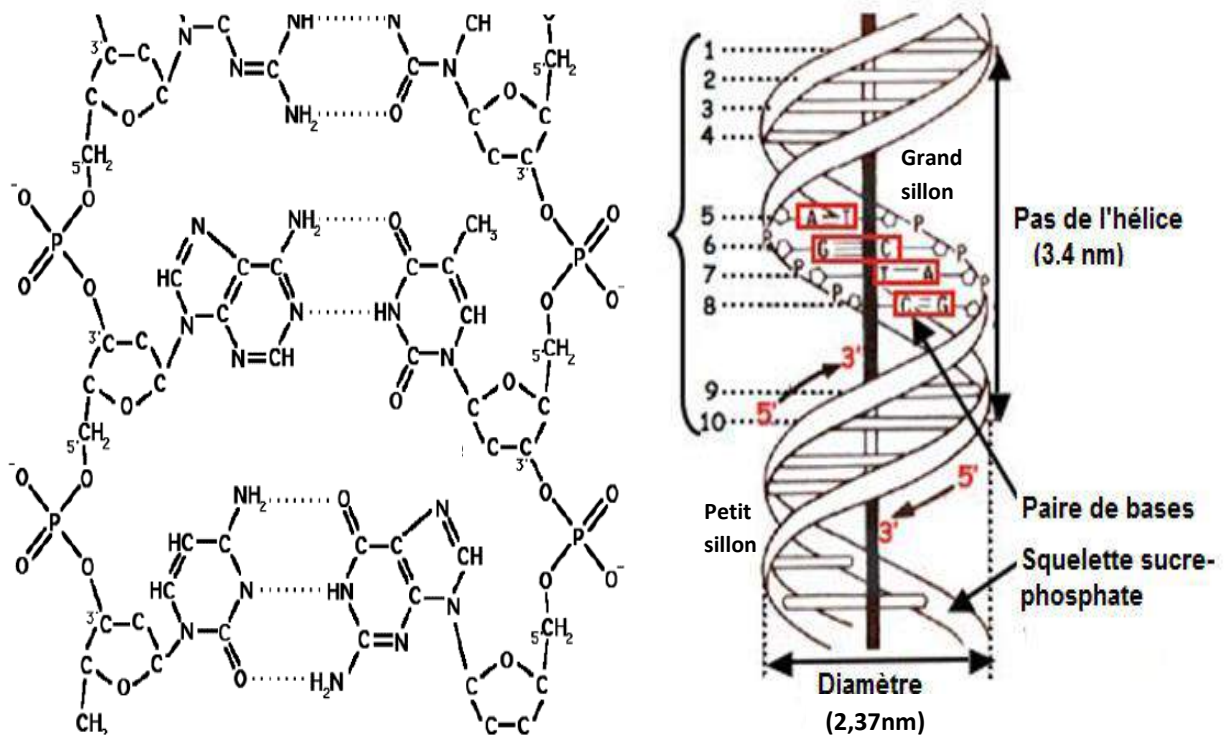


Fig.6: Structure détaillée de la double hélice de l'ADN (forme B). Un tour d'hélice recouvre environ 10.5 pb (34 Å). Le squelette de chaque brin est formé de sucre et de phosphate. Les bases azotées sont projetées vers l'intérieur de la molécule mais restent accessibles au solvant par les deux sillons (petit et grand sillon).

5-Interactions entre C-G et A-T

Deux types de paires de bases, souvent appelées paires de bases complémentaires prédominants dans la plupart des ADN A-T et C-G. Le paire de base A-T est associé par deux liaisons hydrogène et le G-C par trois.

5-1-Liaison hydrogène

Les liaisons hydrogène entre les bases complémentaires sont l'une des caractéristiques fondamentales de la double hélice, car elles assurent la stabilité thermodynamique de l'hélice et la spécificité de l'appariement. Une liaison hydrogène est formée entre un atome ou groupement donneur d'hydrogène chargé positivement et un autre accepteur chargé négativement (fig.7). Dans l'ADN, les liaisons hydrogène stabilisent les paires de base de la double hélice. La complémentarité entre les bases repose sur des raisons stériques.

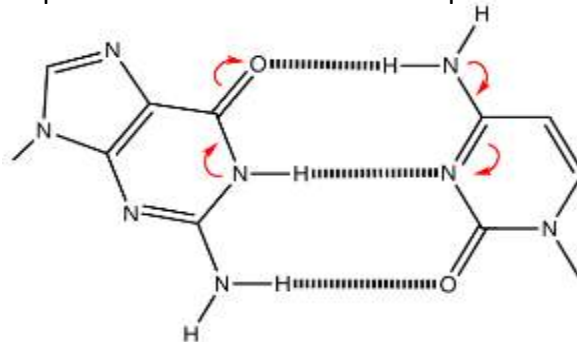


Fig.7: Mise en évidence des groupements impliqués dans les liaisons H

L'appariement de base entre deux purines, deux pyrimidines ou des bases non complémentaires A-C ou G-T est défavorisé, parce que les liaisons hydrogènes appropriées ne peuvent se former sous peine de rompre la géométrie de l'hélice. Une conséquence de cette règle d'appariement des bases est que la séquence d'un brin définit la séquence des bases de l'autre brin.

5-2-Stabilisation de la double hélice de l'ADN

La stabilisation de la double hélice s'effectue par des liaisons chimiques faibles et non covalentes à la fois internes et externes:

- 1- Liaisons hydrogène entre les bases.
- 2- Interactions hydrophobes et électroniques entre les bases
- 3- Interactions des groupements sucre et phosphate avec le milieu aqueux (interactions hydrophiles).

6-Formes de l'ADN

Dans l'espace les deux chaînes présentent une configuration hélicoïdale. Elles s'enroulent autour d'un axe imaginaire pour constituer une double hélice à rotation droite (Ex. Forme A et B) ou bien à rotation gauche (Ex. Forme Z). Il existe plusieurs structures hélicoïdale de l'ADN jusqu'à maintenant présent, six formes ont été décrites (A → E et Z), mais la plupart d'entre elles ont été trouvées dans des conditions expérimentales contrôlées.

Forme B

La forme B de l'ADN est la forme biologique la plus importante, elle correspond à la forme décrite par Watson et Crick en 1953. Cette forme est caractérisée par un pas (tour) d'hélice de 10.5 pb (34 Å) et un diamètre de 2,37nm (fig. 8). La forme B de l'ADN est caractérisée aussi par la présence de deux sillons, le petit sillon et le grand sillon (fig. 9).

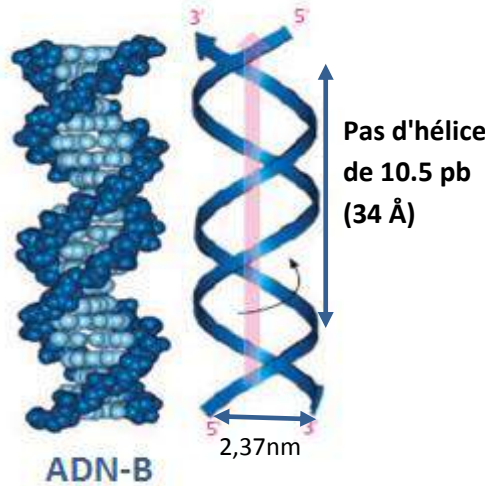


Fig.8: ADN forme B. Toutes les bases ont la même conformation par rapport à l'ose et sont à l'intérieur de l'hélice. Les groupements phosphate et les désoxyriboses sont à l'extérieur (Passarge, 2007).

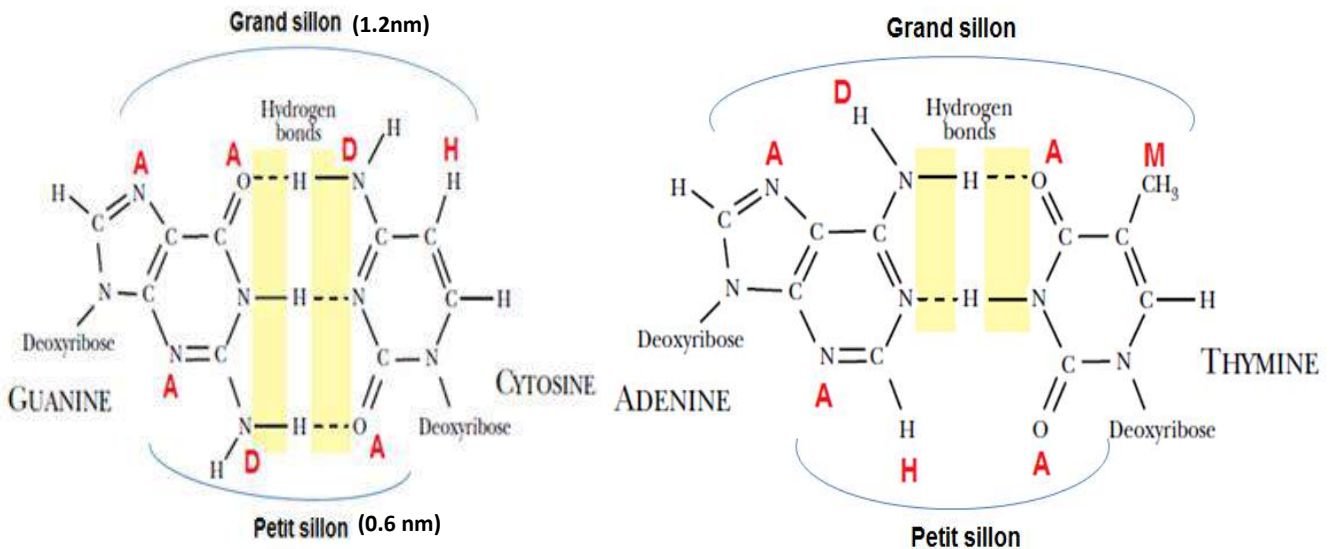


Fig.9: Les groupements chimiques des bases exposés au solvant (projetés vers l'extérieur) dans la double hélice. A: atomes accepteurs d'hydrogène, D: atomes donneurs d'hydrogène, H: hydrogène non polaire, M: groupements méthyles.

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

Les différentes formes de l'ADN se distinguent par:

- 1- Nombre de paires de bases dans chaque tour d'hélice.
- 2- Pas de l'hélice (longueur)

- 3- Diamètre hélicoïdal de la molécule
- 4- Sens de la double hélice (droit, gauche).

Forme Z

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

- Le squelette présente une conformation en zigzag favorisée par un taux de GC élevé (moins lisse que l'ADN-B).
- Il ya un seul sillon, qui ressemble au petit sillon de l'ADN-B
- Les phosphates sont plus proches les uns des autres que dans l'ADN-B.
- L'ADN-Z ne peut pas former de nucléosomes.
- La transformation de l'ADN-B en ADN-Z se fait lors de la transcription des gènes (au site d'initiation de la transcription).

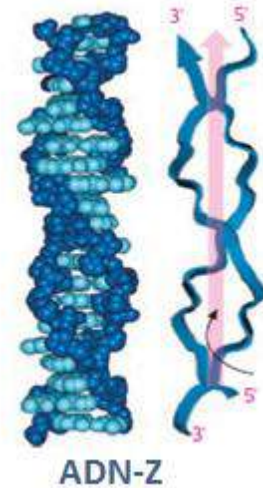


Fig. 10: Forme Z de DNA
(Passarge, 2007)

7- Propriétés physico-chimiques de l'ADN

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

7-1- Température de fusion

Si une solution d'ADN est chauffée, à une certaine température, les liaisons hydrogène qui assurent la cohésion des 2 brins appariés se rompent. On parle de fusion de l'ADN caractérisée par la température de fusion (T_m : melting temperature) (fig. 11).

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

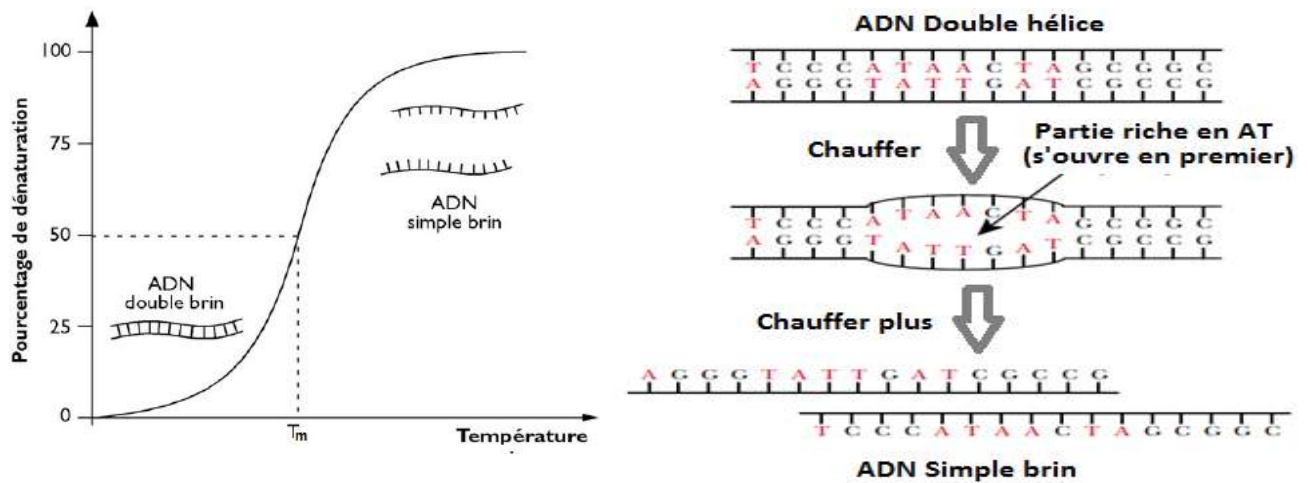


Fig.11: La séparation des brins d'ADN en fonction de la température. T_m est la température pour parvenir à 50% de séparation des deux brins. Il ya une augmentation de l'absorption à 260 nm. Les régions riches en A et T s'ouvrent plus rapidement à celles riches en C et G (Clark, 2005).

7-2- Facteurs influençant la T_m

La température de fusion est influencée par plusieurs facteurs :

- C + G % : la T_m augmente avec l'augmentation du% CG (fig.11). Pour les petites séquences comme les amorces (17 à 31) $T_m = 4(C+G) + 2(A+T)$
- Les cations (mono ou divalent).
- Les polyamines.
- Les protéines.

7-3- Absorption de la lumière ultraviolette

La propriété d'absorption des purines et pyrimidines dans l'UV à 260nm (fig. 13), et les protéines à 280nm permet de doser les acides nucléiques (C: concentration), et aussi bien d'estimer la contamination par les protéines lors de la purification des acides nucléiques.

$$C = A_{260} * DF * 100 \quad (\text{unité: } \mu\text{g}/\mu\text{l} \quad A: \text{Absorbance, DF: facteur de dilution})$$

$$P_{(\text{pureté})} = A_{260} / A_{280} \quad (\text{Une solution d'ADN est considérée pure si } 1.7 \leq P \leq 2)$$

Biologie Moléculaire

Chapitre II

Expression de l'information génétique

Objectifs

- Rôle des éléments nécessaires (éléments *cis*, *trans*, protéines, enzymes...etc.) pour transcrire un gène.
- Etapes de la transcription et les facteurs impliqués.
- Différences entre la transcription chez les procaryotes et les eucaryotes.
- Maturation des ARNm chez les eucaryotes.
- Cadres de lecture ouvert (ORF) et Code génétique
- Eléments nécessaires (ARNs, protéines, enzymes) pour traduire un ARNm en polypeptide.
- Etapes de la traduction et les facteurs impliqués.
- Différences entre la traduction chez les procaryotes et les eucaryotes.

1-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, le métabolisme cellulaire, et le développement des tissus et les organes. 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 (transcription, traduction). On appelle transcription la synthèse de brin d'ARN dont la séquence est dictée par celle de la molécule d'ADN correspondante.

Les trois principaux types d'ARN connus sont:

- ARN messager (ARN_m) : Spécifie les codons
- ARN de transfert (ARN_t): Spécifie les anticodons
- ARN ribosomique (ARN_r): Constituant de ribosomes et impliqué dans la synthèse des protéines.

Chez les procaryotes ces ARN sont synthétisées par une seule ARN polymérase. Chez les eucaryotes elles sont synthétisées par trois types de polymérases, ARN pol I, ARN pol II, et ARN pol III (tab. 1). La réaction nécessite un brin d'ADN qui sert de modèle, les nucléosides trois phosphate (ATP, GTP, CTP, UTP) et Mg⁺⁺.

Tab.1: Les ARN polymérases chez les cellules eucaryotes

| Enzyme | Location | Produit |
|--------------------|-------------------------------|-----------------------|
| ARN polymérase I | Nucléole | rRNA (5.8S, 18S, 28S) |
| ARN polymérase II | Chromatine, matrice nucléaire | ARNm |
| ARN polymérase III | Chromatine, matrice nucléaire | ARNt, 5S, ARNr |

1-1- Gène

Le gène est l'unité fonctionnelle de l'information génétique. Il est constitué d'un ensemble de nucléotides qui contient toutes les informations nécessaires pour transcrire un ARNm (fig. 1).

ARNm: Une succession de nucléotides sous forme de long filament à un seul brin. C'est un vecteur du message dictant la séquence des protéines entre chaque gène et la protéine correspondante.

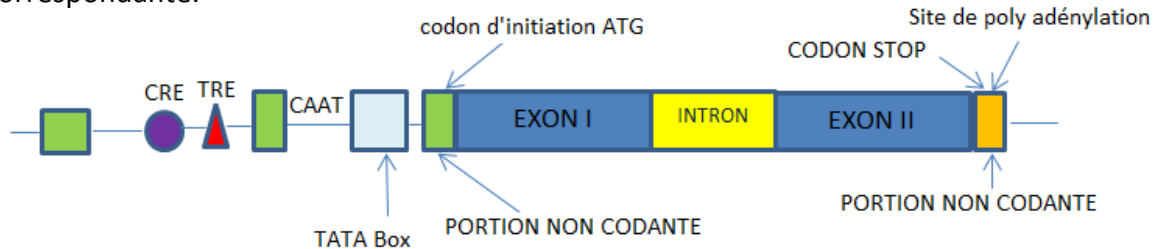


Fig.1: Structure général d'un gène chez les eucaryotes. (CRE [Cyclic AMP Responsive Element], TRE [12-O-Tetradecanoylphorbol 13-acetate Responsive Element]: des éléments de réponse).

Promoteur: Site sur l'ADN d'une centaine de bases auquel l'ARN polymérase (ainsi que les facteurs d'initiation requis) se fixe pour commencer la transcription. Chez *E. coli* il ya environ 2000 sites promoteurs (4.8×10^6 pb), elle sont qualifiés d'éléments *cis* (fig. 2, 3).

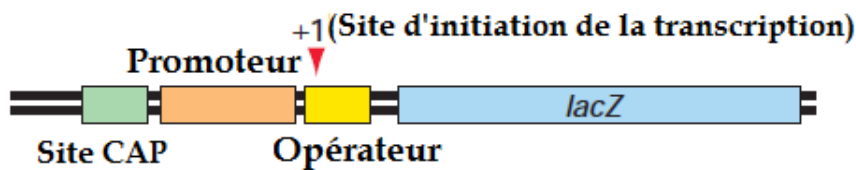


Fig.2: Structure partielle de l'opéron lactose d'*E. coli* (Lodish *et al.*, 2003).

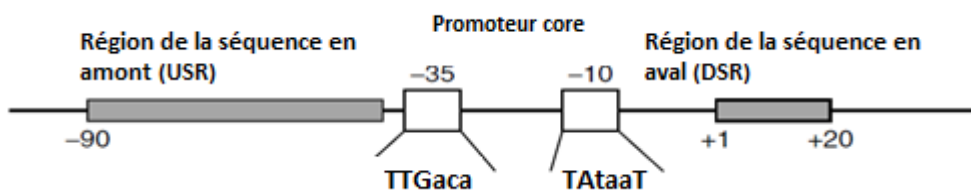


Fig.3: Structure du promoteur bactérien.

1-2-ARN polymérase

L'ARN polymérase assume de multiples fonctions dans le processus de la transcription:

- Elle cherche les promoteurs
- Elle déroule le fragment à transcrire pour produire une matrice simple brin
- Elle sélectionne les NTPs corrects et catalyse la formation des liaisons phosphodiester.
- Elle détecte les signaux de terminaison.

- Elle interagit avec les protéines de régulation de l'expression génétique (activateurs, répresseur).

Les ARN polymérases réalisent essentiellement les mêmes réactions dans toutes les cellules procaryotes et eucaryotes. Les bactéries possèdent une seule ARN polymérase (l'enzyme minimale) capable à elle seule de synthétiser de l'ARN, elle est composée de quatre sous unités, chacune codée par un gène différent (tab. 2). Elle comprend la sous-unité σ et deux sous-unités α et un exemplaire de chacune des sous-unités β , β' (fig. 4), et une sous-unité ω (n'est pas essentielle à la transcription mais elle a un rôle dans la stabilisation du complexe). La forme active de l'enzyme contient les sous-unités $\alpha_2 \beta \beta' \omega - \sigma$ (PM \approx 500000 Da). Parmi ces sous unités les polypeptides β , β' sont à l'origine de l'activité catalytique et constituent le site actif de la transcription, ce site ressemble à celui de l'ADN polymérase par le fait que sous sa forme active, il contient deux ions métalliques. La sous unité σ aide à trouver un site promoteur ou doit commencer la transcription.

Cette enzyme est très proche des polymérases eucaryotes. Plus particulièrement, les deux grandes sous-unités β , β' sont homologues des deux grandes sous unités de pol II (RPB1, EPB2). Les sous-unités α sont homologues de RPB3 et RPB11 et la sous-unité ω est homologue de RPB6.

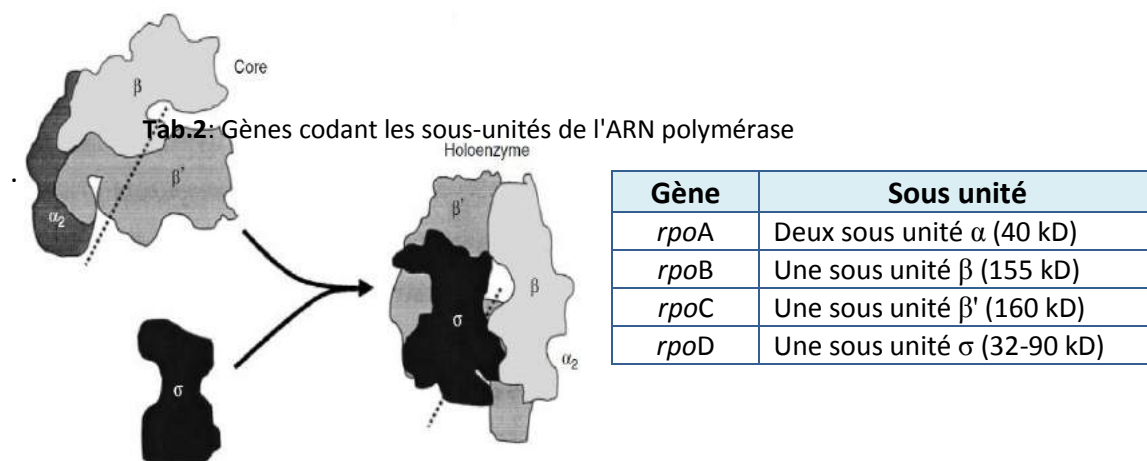


Fig.4:Structure de l'ARN polymérase des bactéries.

La transcription chez les procaryotes ce fait au niveau du cytoplasme, elle est polycistronique car plusieurs gènes peuvent être transcrit par la même polymérase (Ex. L'opéron lactose). Par contre chez les eucaryotes, celle-ci se fait au niveau du noyau, et une polymérase transcrit un seul gène, on dit qu'elle est monocistronique (fig. 5).

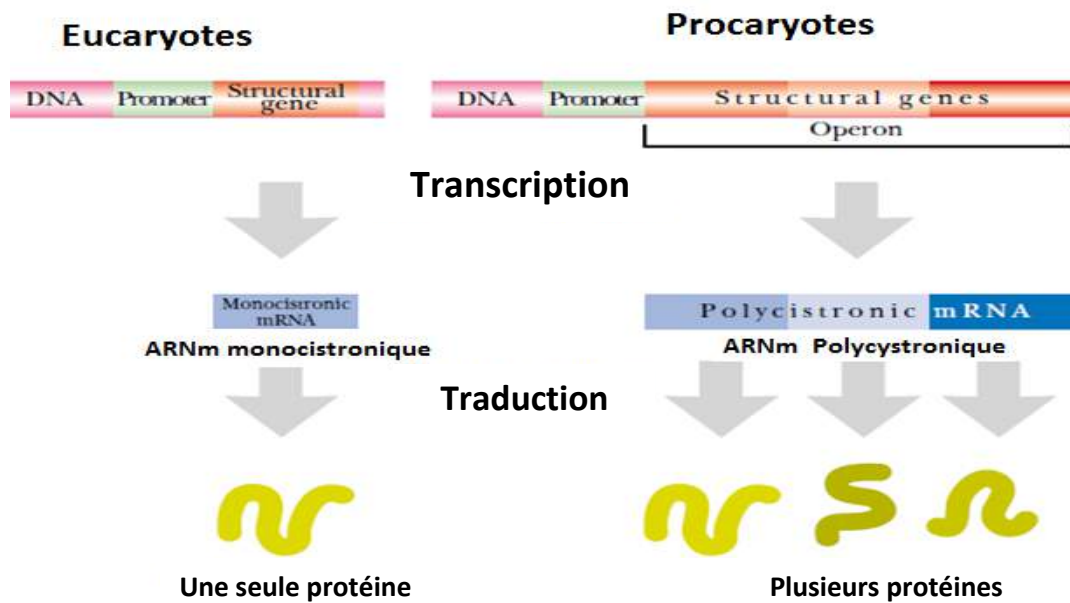


Fig.5: Synthèse monocistronique (eucaryotes) et polycistronique (procaryotes) de l'ARNm (Clark, 2005).

1-3-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.

1-3-1- Initiation

Cette étape initiale est appelée "fixation sur le brin matrice". Chez les bactéries, cette fixation est établie quand la sous unité σ (élément *trans*) de l'ARN polymérase reconnaît le promoteur (éléments *cis*) localisé dans la région 5' (en amont) du point de transcription initiale d'un gène (fig. 6).

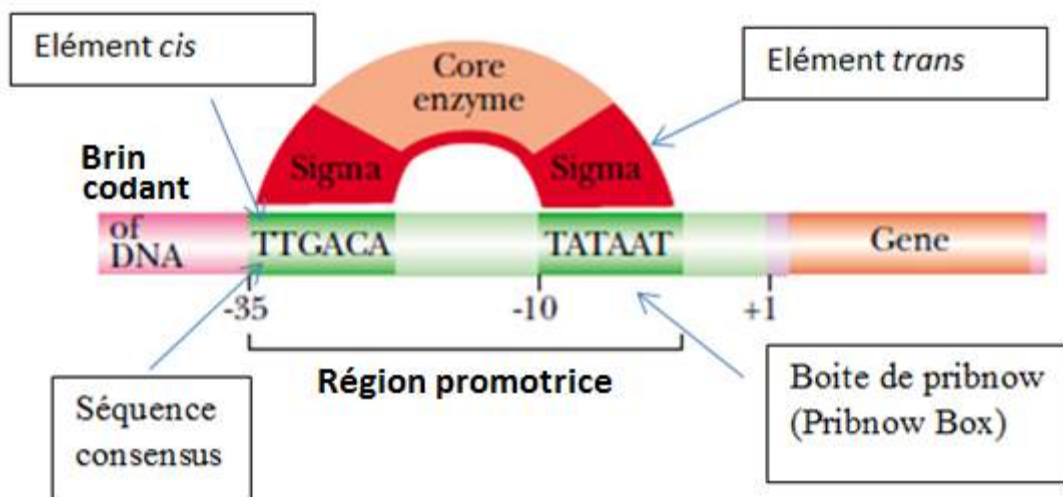


Fig.6: Interaction entre l'ARN polymérase et le promoteur bactérien (Clark, 2005).

1-3-1-1-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:

- **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 coté 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.

Une séquence consensus comparable à celle de la région -10 a été identifiée dans la plupart des gènes eucaryotes étudiés. Elle est appelée boîte TATA (TATA Box).

1-3-1-2-Fixation de l'ARN polymérase

Le degré de la fixation de l'ARN polymérase aux différents promoteurs varie considérablement (promoteur fort, promoteur faible), en provoquant la variation de l'expression génétique. Généralement ce phénomène est attribué à la variation des séquences dans les promoteurs.

1-3-1-3-Éléments *trans*

Chez les bactéries, il existe plusieurs formes alternatives de l'ARN polymérase, chaque forme contient une sous-unité σ particulière (σ^{32} , σ^{54} , σ^{70} , σ^S et σ^E). La sous-unité σ^{70} est la plus grande forme (70 kDa). Ces éléments *trans* reconnaissent les séquences consensus de différents promoteurs et permettent une spécificité de l'initiation de la transcription. Cela permet de contrôler leur expression.

Une fois qu'elle a reconnu le promoteur, et qu'elle s'y fixe, l'ARN polymérase catalyse l'**initiation**, c'est-à-dire l'insertion du premier NTP du site d'initiation du brin d'ADN matrice (aucune amorce n'est nécessaire : initiation de la synthèse *de novo*). Ce processus se poursuit dans la direction 5'→3', créant un duplexe ADN-ARN temporaire de 8 pb.

1-3-2-Elongation

Une fois le duplexe temporaire est formé, la sous-unité σ se dissocie de l'holoenzyme et l'élongation de la chaîne se poursuit sous la direction de la partie centrale de l'enzyme (fig.7). Chez *E. coli* ce processus progresse à la vitesse d'environ 50 nucléotides/seconde à 37°C. Par la suite la polymérase parcourt le gène entier jusqu'à rencontrer une séquence dite de terminaison.

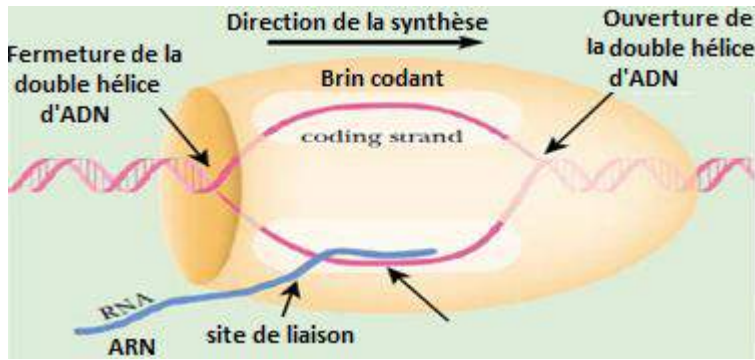


Fig.7: Elongation et direction de la synthèse.

1-3-3-Terminaison

La terminaison commence dès l'enzyme rencontre un signal de terminaison d'une séquence de 40 pb. Un aspect intéressant de la terminaison chez les bactéries est que la séquence de terminaison ci-dessus est en fait transcrite en ARN. Cette dernière forme une structure en tige boucle par appariement de bases du même brin. L'ADN au niveau de la séquence de terminaison est inversé au centre duquel une zone non répétée est insérée (fig.8).

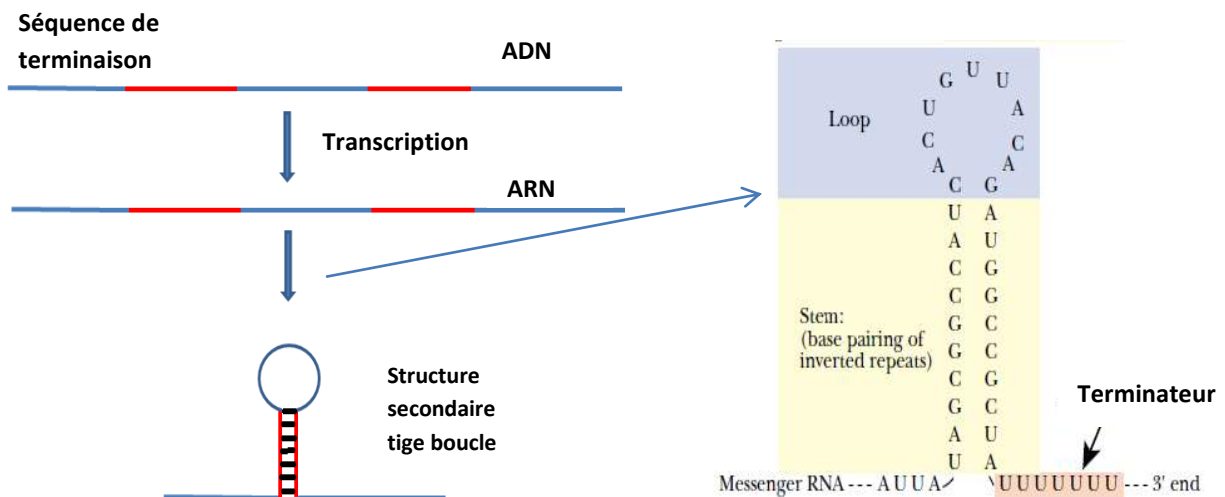


Fig. 8: Formation du signal de terminaison de la transcription chez les bactéries

Ces structures secondaires suivies d'une série d'uridines sont des terminateurs efficaces de la transcription (fig. 8, 9). Des séquences riches en GC suivie par autres riche en AT sont aussi des sites spécifiques de terminaison. Ces régions qui ne nécessitent pas de facteurs additionnels sont nommées "terminateurs intrinsèques".

D'autres types de terminateurs nécessitent des facteurs protéiques spécifiques. Chez *E. coli* la protéine *Rho* se fixe fortement à l'ARN (mais pas ni à l'enzyme ni à l'ADN), elle parcourt alors la chaîne jusqu'au complexe ARN polymérase-ADN. Dès que l'ARN polymérase s'arrête à un site de terminaison *Rho* dépendant. Le facteur *Rho* favorise la libération de l'ARN et de l'enzyme positionnée sur l'ADN, terminant ainsi la transcription (fig. 10). Il ya aussi d'autres protéines comme *Rho* sont aussi impliquées dans la terminaison de la transcription.

Dans tous les cas, les séquences impliquées dans la terminaison sont localisées au niveau de l'ARN, cependant l'ARN étant transcrit depuis l'ADN, la terminaison est au final toujours déterminée par des séquences nucléotidiques spécifiques sur l'ADN.

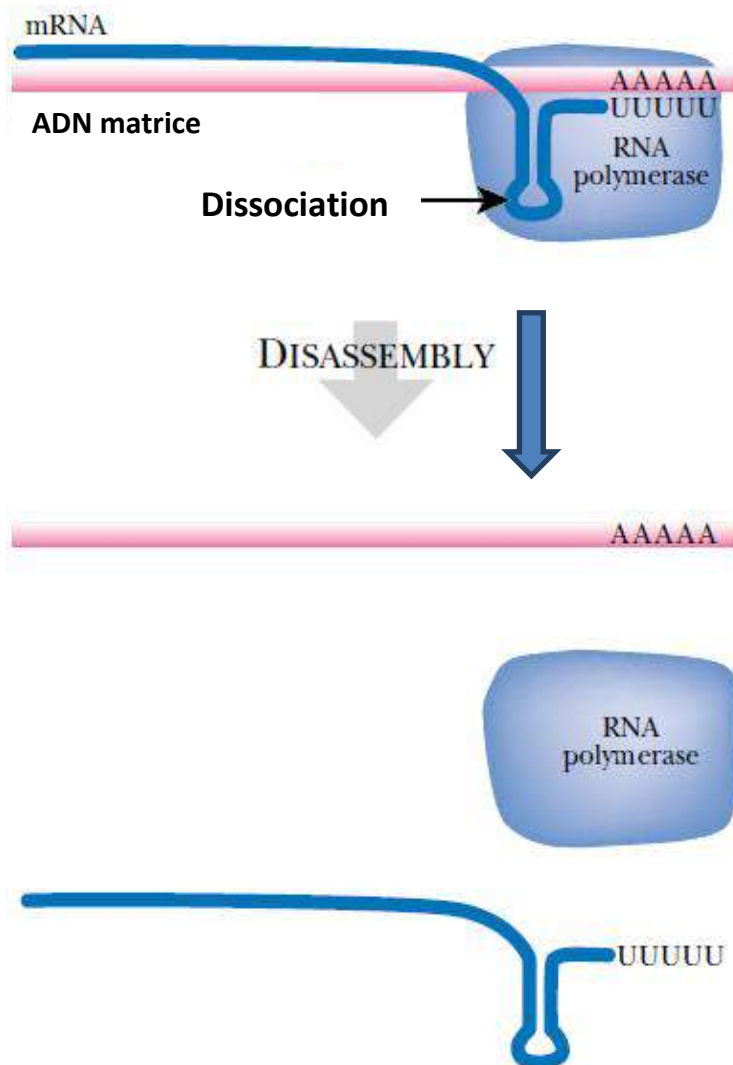


Fig.9: Terminaison de la transcription par des terminateurs intrinsèques (Clark, 2005).

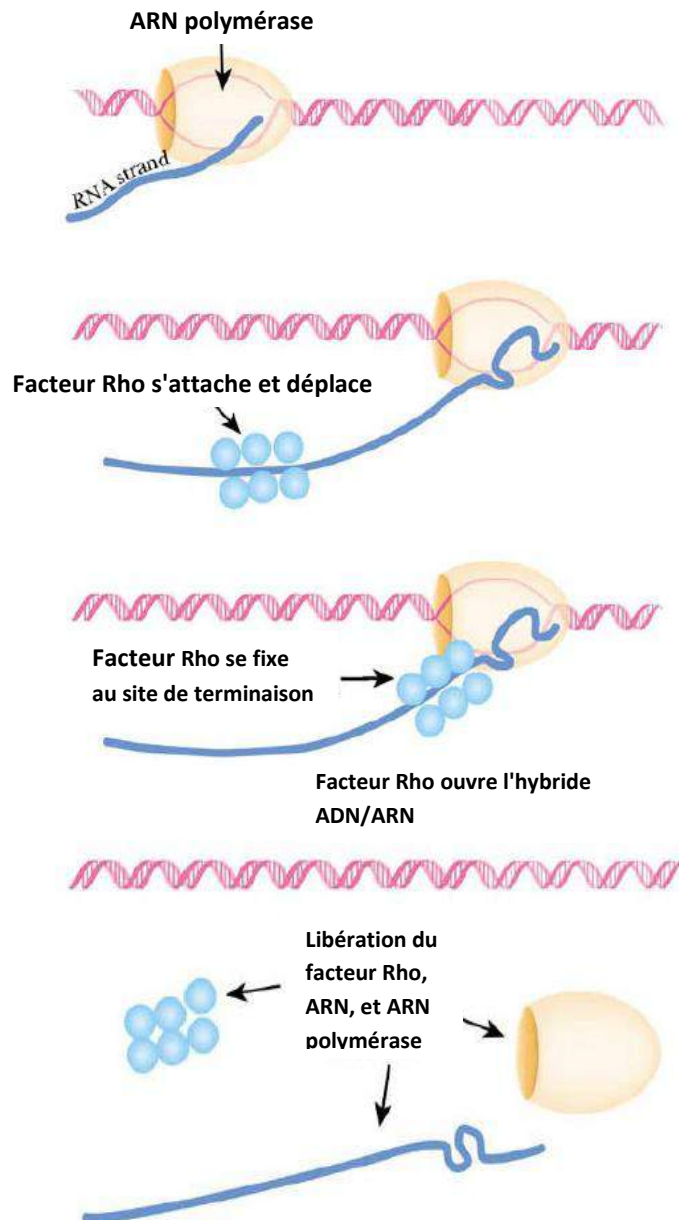


Fig.10: Terminaison de la transcription chez les bactéries par le facteur *Rho*(Clark, 2005).

1-4- Transcription chez les eucaryotes

La transcription chez les eucaryotes est prise en charge par des ARN polymérase très apparentées à celles présentes chez les procaryotes. Mais c'est un processus beaucoup plus complexe et il existe plusieurs différences notables:

- La transcription chez les eucaryotes est prise en charge par 3 ARN polymérases (Pol I, II, et III) très apparentées à celles présentes chez les procaryotes.
- La transcription au contraire de la traduction se fait au niveau du noyau.
- Elle nécessite plusieurs facteurs d'initiation (facteurs généraux de transcription (FGT: TFII A, TFII B, TFII D (TBP, TAF), TFII E, TFIIH).

- Remodelage de la chromatine par des protéines régulatrices et des enzymes de modification de la chromatine ----> : changement de conformation lors de l'ouverture de la double hélice de l'ADN (ADN inclus dans les nucléosomes).
 - En plus des éléments *cis* (ex: TATA Box) il existe en plus du promoteur des autres séquences ou unités de contrôle (fig. 11):
 - Activateurs : Enhancers
 - Extincteurs: Silencers.
- Ces séquences peuvent être localisées dans la région régulatrice en 5', en amont du point d'initiation, mais aussi parfois à l'intérieur du gène ou même dans la région 3' en aval de la séquence codante (fig. 12).

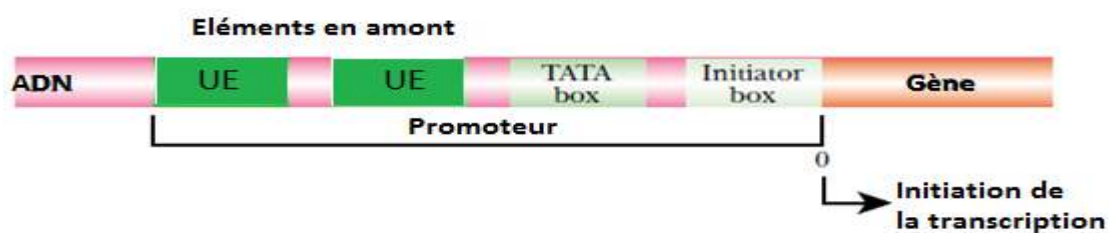


Fig.11: Les composants d'un promoteur eucaryote.

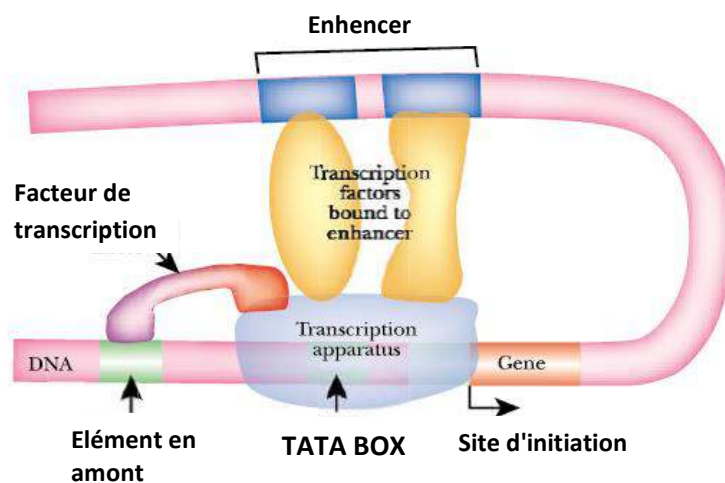


Fig.12: Contrôle de la transcription par un enhancer localisé en aval (downstream) du site de l'initiation de la transcription. L'ADN forme une loop pour faciliter le contact de l'enhancer avec le complexe de l'initiation de la transcription par l'intermédiaire des facteurs de transcription (Clark ; 2005).

1-4-1- Étapes de la transcription chez les eucaryotes

Comme chez les procaryotes, la transcription se fait en 3 étapes :initiation, élongation et terminaison.

1-4-1-1- Initiation

Chez les eucaryotes, plusieurs facteurs d'initiation sont nécessaires pour une initiation efficace, alors que les procaryotes besoin d'un seul (σ).

L'ARN polymérase II est régulée à la fois par les éléments agissant en *cis* (dans le gène lui-même) et par certains nombres de facteurs agissant en *trans* qui se fixent à ces éléments d'ADN. Au moins trois éléments d'ADN agissant en *cis* régulent l'initiation de la transcription par l'ARN polymérase II. Les facteurs généraux de la transcription (agissant en *trans*) aident la polymérase à se fixer au promoteur et à séparer les brins d'ADN. Ils aident aussi la polymérase à se détacher du promoteur et à s'engager dans la phase d'élongation.

a- Complexe de préinitiation

La reconnaissance du promoteur et la mise en place du complexe de pré initiation chez la plupart des pol II débute au niveau du TATA box (fig. 13), ce dernier est reconnu par TFIID, plus précisément par la sous unité TBP (TATA box Binding Protein). Les bases A et T sont préférées parce qu'elles sont plus aisément déformables pour permettre l'élargissement initial du petit sillon. D'autres sous unités de ce complexe sont appelées TAF (TBP associated factors) reconnaissent d'autres éléments du promoteur, tels Inr, DPE et DCE, bien que la liaison la plus forte soit celle entre TBP et TATA. Le complexe TBP-TATA attire sur le promoteur d'autres facteurs généraux de transcription [TFIIA → liaison du Pol II au promoteur (upstream), TFIIB → liaison du Pol II au promoteur (downstream), TFIIF → accompagne Pol II dès qu'elle se fixe sur le promoteur, TFIIIE → essentiel pour l'élongation et la libération du Pol II du promoteur, et TFIIH → il a un rôle dans la phosphorylation du CTD du Pol II et il a aussi un rôle dans l'élongation] et la polymérase elle-même. La double hélice à ce niveau se sépare par l'intermédiaire du TFIIH en hydrolysant la liaison anhydride de l'ATP (comme l'hélicase dans la réplication).

b- Libération de la polymérase du promoteur

Cette étape d'initiation abortive (car le complexe est instable, donc elle se répète un certain nombre de fois avant que se forme un hybride ADN:ARN stable incluant un transcrit de 11 nucléotides) implique deux étapes :

- 1- Hydrolyse de l'ATP
- 2- Phosphorylation de la polymérase

La grande sous unité de pol II possède sur l'extrémité C-terminale une série de séquence répétée de 7 acides aminés (heptapeptide: Tyr-Ser-Pro-Thr-Ser-Pro-Ser) appelée domaine carboxy terminal (CTD) ou la queue. Le nombre de répétitions dépend des espèces: 27 fois chez les levures, 45 fois chez la mouche *Drosophila*, et 52 fois chez l'homme. Chaque motif répété contient des sites de phosphorylation par des kinases spécifiques, notamment une qui est une sous unité de TFIIH. La régulation du niveau de phosphorylation du CTD de Pol II contrôle aussi les étapes ultérieures (élongation, maturation).

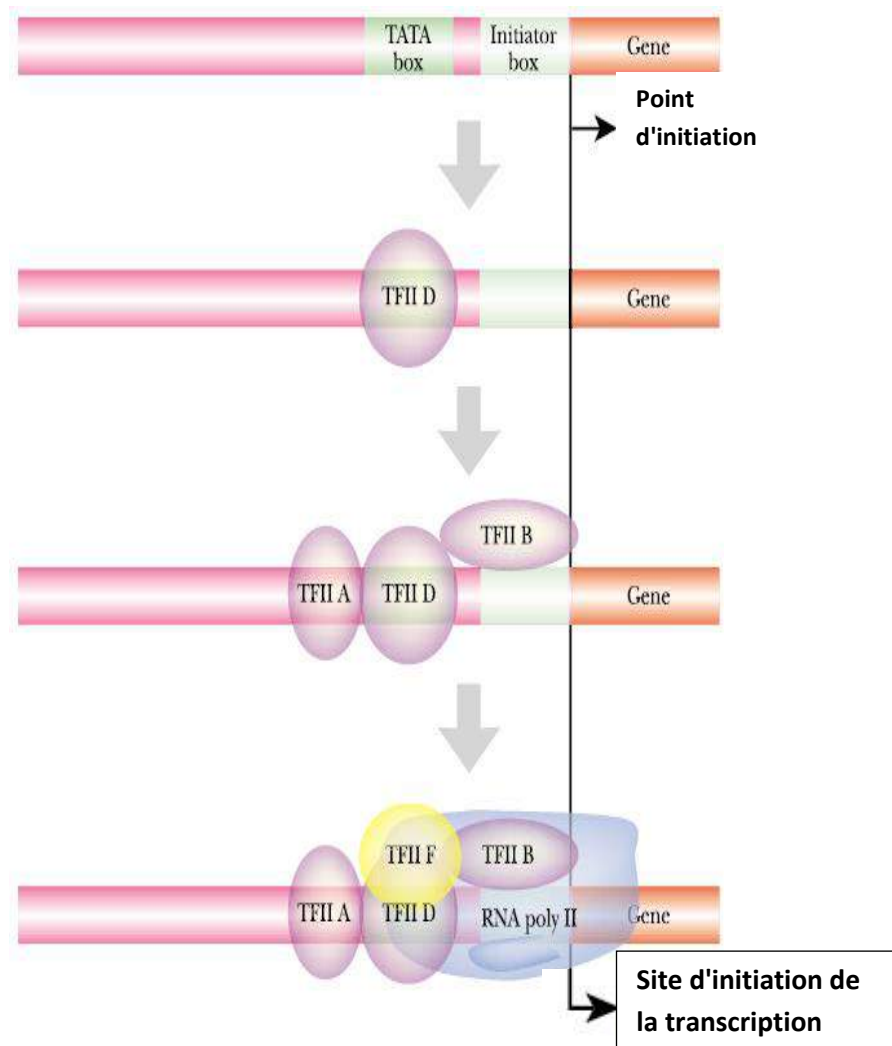


Fig.13: Liaison de l'ARN pol II au promoteur (Clark ; 2005).

1-4-1-2- Elongation

Dés qu'il y'a formation de l'hybride ADN/ARN stable l'élongation du transcrit d'ARN progresse. A partir de ce moment on peut dire que la transcription devient un processus de polymérisation de l'ARN hautement efficace.

Dans l'étape de l'élongation la Pol II se débarrasse de la plupart de ses facteurs d'initiation, comme les FGT et le médiateur (requis pour atteindre des niveaux élevés de transcription *in vivo*: car l'ADN est empaqueté sous forme de chromatine). De nouveaux facteurs appelés facteurs d'élongation vont les remplacer.

Différents protéines sont supposées stimuler l'élongation comme la kinase P-TEFb (phosphoryle des résidus Ser de CTD, et active le facteur d'élongation hSPT5), TAT-SF1, P-TEFb (stimule l'élongation par trois voies distinctes). Il existe aussi des facteurs d'élongation de type ELL (suppriment les arrêts temporaires de l'enzyme). Un autre facteur important dans l'élongation et n'a aucun rôle dans l'initiation est le TFIIIS, ce facteur joue de rôle important le

premier est de réduire les pause de la polymérase et le deuxième est de corriger les erreurs (mésappariements) par l'activité RNase.

Comme l'initiation de la transcription, l'élongation se déroule en présence d'histones, ce qui fait appel à des facteurs facilitant la transcription en présence de chromatine. Le facteur FACT (facilitates Chromatin Transcription) rend la transcription sur type de matrice chromatine beaucoup plus efficace. C'est un hétérodimère de deux protéines, la première Spt16 se lie sur le module H2A/H2B et la deuxième SSRP1 se lie sur le module H3/H4 pour démanteler les nucléosome.

Durant l'élongation la polymérase est associée à un nouvel ensemble de facteurs protéiques requis pour différents types de modification de l'ARN avant d'être exporté du noyau pour être traduit. Ces modifications incluent la coiffe, l'épissage et la polyadénylation.

1-4-1-3-Terminaison

Lorsque la transcrit progresse jusqu'à la portion d'ADN qui signal la terminaison, le complexe devient "instable". Le clamp s'ouvre, l'ADN et l'ARN étant tous deux libérés de l'enzyme lorsque la transcription est terminée.

1-5- Maturation des ARNm

La maturation des ARNm prématurés se fait au niveau du noyau et entraîne des changements dans leur structure (fig. 14 et fig.15):

- Addition d'une coiffe (7mG) en 5' et d'une queue (poly A) en 3' pour la plupart des transcrits destinés à devenir des ARNm.
- Autres modifications ont lieu à l'intérieur des séquences nucléotidiques.
- Pré-ARNm: les transcrits initiaux se trouvent uniquement dans le noyau. Elles sont désignées par l'appellation collective d'ARN nucléaires hétérogènes (ARNnh). 25% d'ARNnh sont convertis en ARNm par excision des introns et assemblage des exons. Concept de gènes discontinus et d'épissage (splicing) chez les eucaryotes.

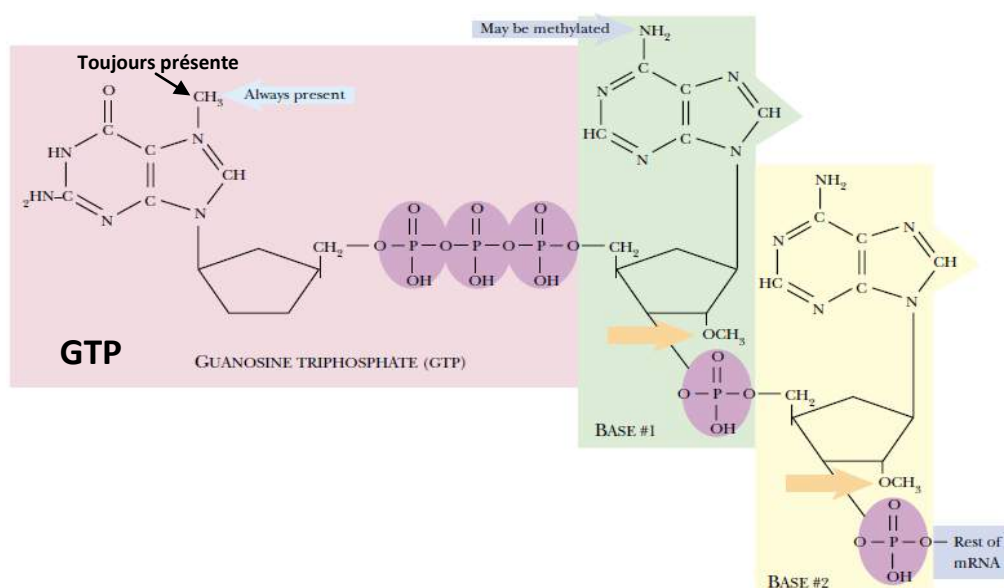


Fig.14: Addition de la coiffe: 7m-Guanosine.

La transcription et la maturation des ARNm sont couplées, elles sont coordonnées par le domaine carboxyle terminal (CTD) de la pol II. Nous avons vu que le CTD est constitué d'une répétition d'une séquence unique et conservée de sept acide aminés (YSPTSPS). Soit S₂, soit S₅, soit les deux sont phosphorylés dans les diverses répétitions. L'état de phosphorylation du CTD est contrôlé par un grand nombre de kinases et de phosphatases. Cela rend le CTD capable de fixer beaucoup de protéines et enzymes ayant des rôles variés dans la transcription et la maturation de l'ARN. Elles comprennent:

- 1- Les enzymes de la formation de la coiffe, qui méthyle la guanine 5' du ARN pm (pré-mRNA) immédiatement après que la transcription commence.
- 2- Les composants de la machinerie d'épissage qui initient l'excision de chaque intron au fur et à mesure de la synthèse. Il ya plusieurs classes des introns (GA-AG, AT-AC, Groupe I, II, et III, Twintrons...etc.), la plus fréquente est GT-AG (ou GU-AG dans le transcrit d'ARN).
- 3- Une endonucléase qui clive le transcrit au niveau du site d'addition du poly(A), créant un groupe 3'-OH libre qui est la cible de l'adénylation 3'.

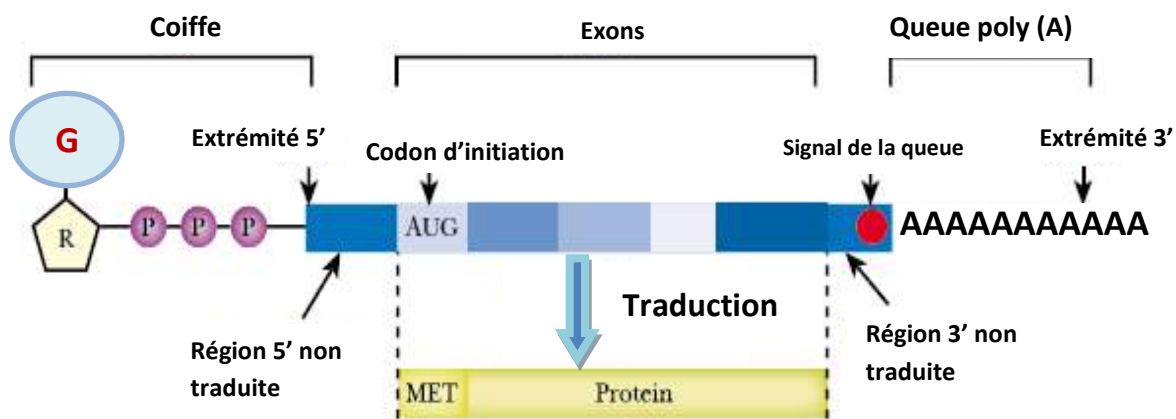


Fig.15: ARNm eucaryote mature prêt à être traduit. Coiffe ajoutée (CAP ADDED), La queue polyA ajoutée (TAIL ADDED), introns éliminés (INTRONS REMOVED).

1-5-1- Ribozymes

Certains types d'ARN fonctionnent dans la cellule comme enzymes et comme ARN, ces ARN sont appelés ribozymes.

Les ribozymes sont rares, elles interviennent dans l'épissage des introns du groupe II et I, Elle se trouve dans quelques gènes des organites eucaryotes et de procaryotes (*Archaea*) (II et I) et ARNr nucléaire (I), et elles fonctionnent comme les enzymes protéiques, en ce sens qu'ils possèdent un site actif qui lie le substrat et catalyse la formation d'un produit. La plupart des ribozymes sont des introns auto-épissant (self splicing introns) chez les eucaryotes. La taille d'un ribozyme caractérisée chez le protozoaire *Tetrahymena* est de 413 ribonucléotides qui en présence de guanosine, se sépare d'un ARN précurseur plus long. Le ribozyme joint deux exons adjacents pour former l'ARNm mature (fig. 16). Ce ribozyme est donc une endoribonucléase spécifique d'une séquence donnée et réalise une réaction analogue à celle du splicéosome. *In vitro* il ne nécessite pas d'autres facteurs, mais *in vivo* nécessite des facteurs protéiques pour déclencher le processus de l'épissage.

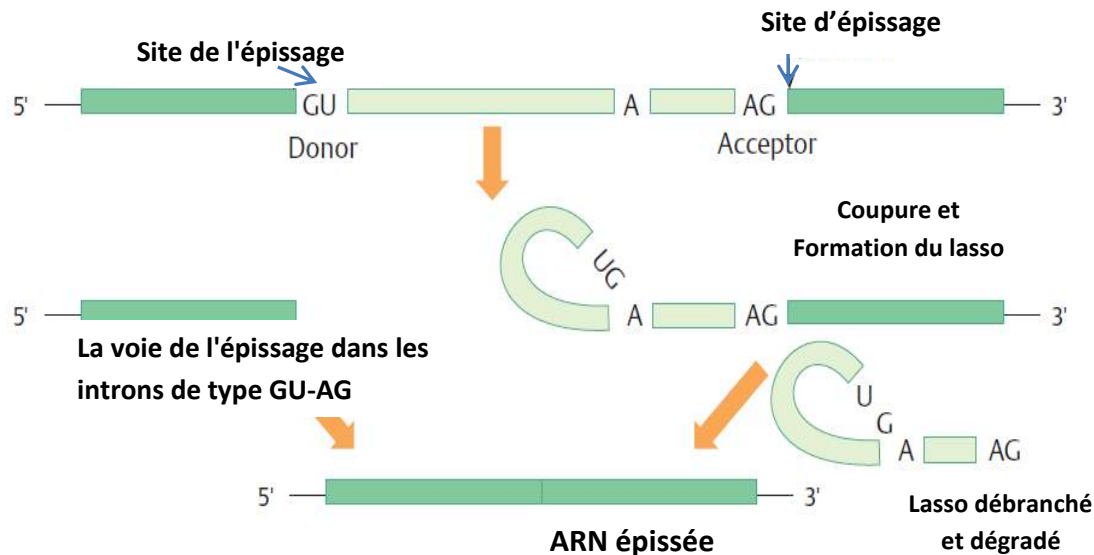


Fig.16: Mécanisme d'épissage (splicing) des introns GU-AG (Intron auto-épissant du groupe II). La présence d'une adénine très réactionnelle dans l'intron initiant l'épissage et conduisant à la formation d'un produit en forme de lasso (lariat), dans lequel le résidu adénylate est lié à trois nucléotides par des liaisons phosphodiester (sert de boîte de branchement)(Passarge, 2007).

1-5-2-Splicéosome

Le splicéosome est un complexe macromoléculaire de grande taille (\approx la taille d'un ribosome) et environ 150 protéines et 5 ARN sont impliqués dans son activité, sans compter des facteurs protéiques ne faisant pas partie du splicéosome lui-même. Le splicéosome enlève les introns et effectue la jointure des exons pour former un ARNm mature. Les réactions catalysées consomment plusieurs molécules d'ATP, et elles sont réalisées par les ARNs plutôt que les protéines, et même la localisation des introns se fait par les snRNA (Ex: U1, U2, U2AF). La taille de ces snRNA varie de 100 à 300 nucléotides chez la plupart des eucaryotes. Ces petits ARN sont souvent liés à des protéines et le complexe formé est appelé : snRNP (petites ribonucléoprotéines nucléaires).

Les snRNP jouent trois rôles dans l'épissage

- Reconnait le site d'épissage 5', le site de branchement et le site d'épissage en 3'.
- Rapprochent ces sites de manière adéquate.
- Catalysent des réactions de clivage et de ligation (ou participent à la catalyse) par des interactions ARN-ARN, ARN-protéine et protéine-protéine

Les introns sont repérés par l'intermédiaire des séquences consensus au niveau de l'ARNpm, les séquences les plus conservées sont les sites d'épissage 5'-GU et AG-3' et A au site de branchement. Soit pour les introns auto épissant (II) ou les splicéosomes, l'intron est enlevé par deux réactions successives de trans-estérification.

La première est déclenchée par le 2'-OH du A agissant comme un nucléophile pour attaquer le groupe phosphorylé du G. A la suite de cette réaction, la liaison phosphodiester entre le sucre et le phosphate à la jonction intron-exon 5' est rompue. L'extrémité 5' de l'intron

libérée est jointe de A du site de branchement (fig. 17). Ainsi, en plus des liaisons 5' et 3' du squelette, une troisième liaison phosphodiester est formée entre le 2'-OH de A avec le P-5' de G.

Dans la deuxième réaction le groupement 3'-OH de l'exon 5' libéré change de rôle et devient nucléophile qui attaque le groupe phosphoryle au niveau du site d'épissage 3' en joignant les deux exons et libérant l'intron sous forme de lasso (l'intron libéré est linéaire pour le groupe I).

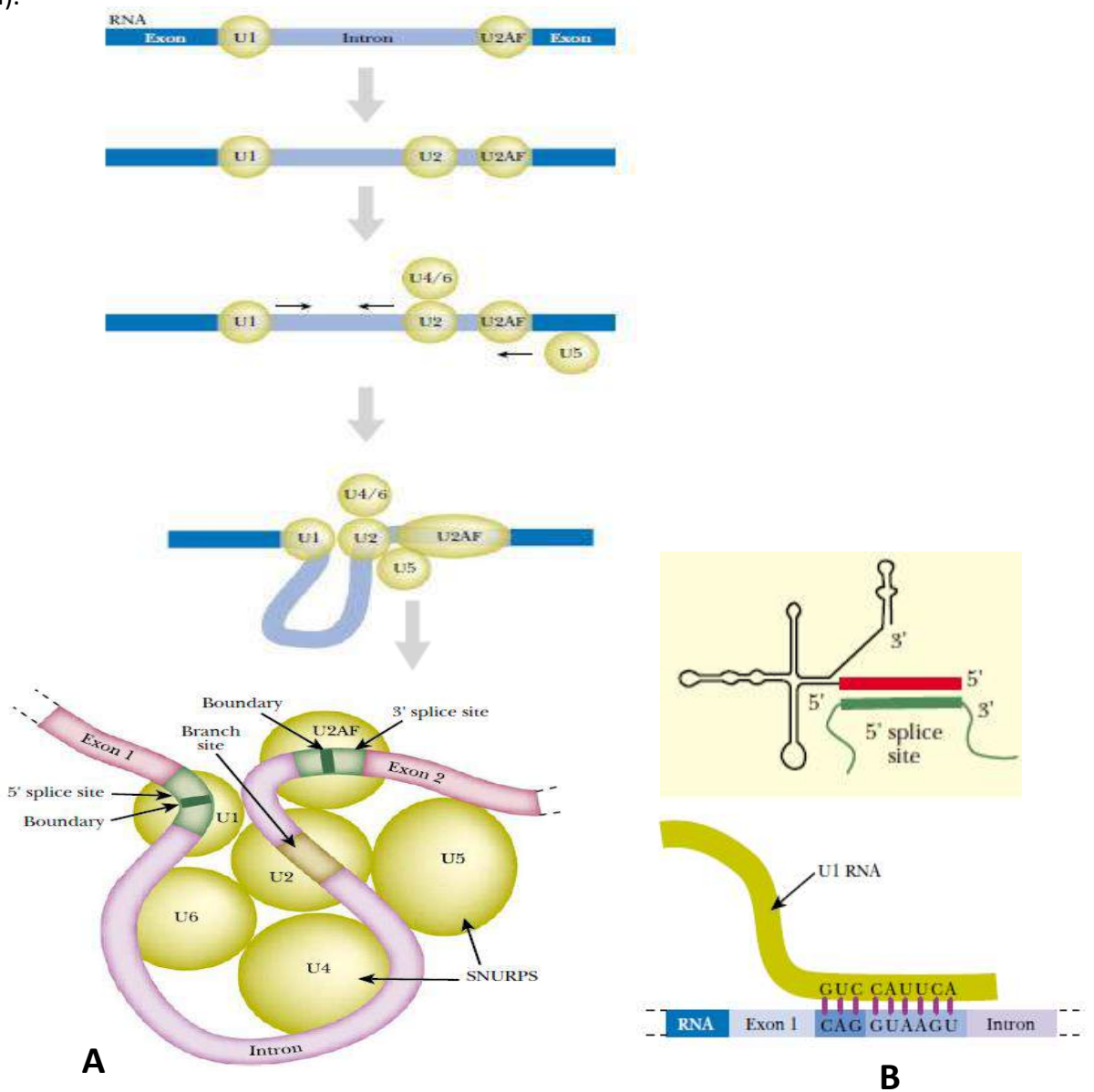


Fig.17: **A-** Les étapes d'assemblage du splicéosome, il ya six SNURPs (or snRNP: small nuclear ribonucleoproteins) numérotés U1, U2, U4, U5 et U6, le U3 est un snRNA (se trouve dans le nucléole), U1 s'apparie sur le site d'épissage en 5', U2 s'apparie avec le site de branchement, et U2AF avec le site d'épissage en 3'. **B-** Appariement spécifique d'U1 au site d'épissage en 5' (Clark, 2005).

2 -Traduction

Dans la réplication et la transcription le transfert de l'information génétique nécessite une matrice d'ADN pour produire un acide nucléique (ADN, ARN), mais dans la traduction la matrice est l'ARN et le produit final est une polypeptide ou protéine.

La traduction est parmi les événements les plus conservés chez tous les organismes, et les plus coûteuses en énergie. Dans des bactéries en croissance rapide, jusqu'à 80% de l'énergie cellulaire et 50% du poids sec de la cellule sont consacrés à la synthèse des protéines. En effet, la synthèse d'une protéine nécessite environ 100 protéines et ARN.

Quatre composants principaux sont impliqués dans la traduction:

- 1- **ARNm**: Cette macromolécule fournit l'information qui doit être interprétée par la machinerie de la traduction et constitue la matrice pour la traduction. La région de l'ARNm codant la protéine est constituée d'une série ordonnée d'unités de trois nucléotides appelées codons (spécifient l'ordre des acides aminés).
- 2- **ARNt**: Ils fournissent l'interface entre les acides aminés ajoutés à la chaîne peptidique en croissance et les codons dans l'ARNm.
- 3- **Aminoacyl-ARNt synthétase**: Ces enzymes couples des acides aminés avec des ARNt spécifiques qui reconnaissent le codon approprié.
- 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.

2-1- ARN messenger et le code génétique

On a vu en détail la structure d'un transcrit ou ARNm dans le chapitre de la transcription. La question qui se pose, est-ce que toute la séquence de l'ARNm est décodée?

Dans tout ARNm, la région (ou les régions) codant un polypeptide est composée d'une suite de codons adjacents et non chevauchants (au moins 100 codons) appelée cadre de lecture ouvert ou ORF (**O**pen **R**eading **F**rame). Chaque ORF spécifie un seul polypeptide. La traduction commence à l'extrémité 5' de l'ORF et procède par le codon d'initiation (AUG: *N*-formyle méthionine chez les bactéries, et méthionine chez les *Eukarya* et *Archaea*) puis continue codon par codon jusqu'à le codon stop (UAA, UAG, UGA) à l'extrémité 3'. La traduction doit commencer à un point de départ précis (fig. 18), si non le cadre de lecture sera décalé (différente protéine, aucune protéine s'il y a introduction d'un ou plusieurs codons stop).

Il y a des logiciels permettant la recherche des ORF à partir d'une banque d'ADN. L'analyse correcte des ORF permet d'identifier plus de 90% des gènes des bactéries et les levures. Ces logiciels permettent aussi de rechercher les promoteurs, les séquences de Shine Dalgarno et les différents codons (voir TD bioinformatique). La recherche de ces éléments est très importante en génomique et en génie génétique. La génomique a montré que l'utilisation des codons varie selon les organismes. Par exemple, chez les polypeptides d'*E. coli*, seule 1 isoleucine sur 20 est codé par le codon AUA, les 19 autres le sont par AUU et AUC.

I- 5'----AUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUCCUGA----3'
 MK A I F V L K G W W R T S STOP

II- 5'—AUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUCCUGA--3'
 STOP K Q F S Y STOP K V G G A L P

III- 5'—AUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUCCUGA---3'
 E S N F R T E R L V A H F L

Fig.18: Cadre de lectures possibles dans un ARNm. Le décalage du cadre de lecture par 1 ou 2 nucléotides a donné deux séquences différentes. La détermination du codon d'initiation détermine la localisation de tous les codons suivants. Un ORF contient un seul codon stop.

2-1-1- Code génétique

Le code génétique établit une correspondance entre la matrice d'acides nucléiques et le polypeptide produit (Un acide aminé correspond à un codon). Le code génétique est écrit en ARNm plutôt qu'avec l'ADN. L'aspect le plus intéressant du code génétique est que certains acides aminés sont codés par plusieurs codons apparentés mais différents. Parmi un total de 64 codons, 4 sont dédiés à l'initiation et à l'arrêt de la traduction (tab. 1).

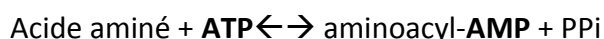
2-1-2- Reconnaissance, activation et chargement des ARNt

Les ARNt agissent comme adaptateurs entre les codons et les acides aminés qu'ils spécifient (fig. 19). Il ya plusieurs types d'ARNt mais chacun reconnaît un codon particulier, ou quelques codons particuliers, dans l'ARNm. La réaction spécifique entre l'acide aminé et l'ARNt est catalysée par l'aminoacyl-ARNt synthétase. Il ya deux classes de ces enzymes:

Classe I: Attache l'acide aminé à l'extrémité 2' de l'ARNt et sont généralement monomériques (Glu, Gln, Arg, Cys, Met, Val, Ile, Leu, Tyr, Trp).

Classe II: Attache l'acide aminé à l'hydroxyle 3' de l'ARNt et sont typiquement dimériques ou tétramériques (Gly, Ala, Pro, Ser, Thr, His, Asp, Asn, Lys, Phe).

La réaction est catalysée en deux étapes, premièrement il ya activation de l'acide aminé par l'ATP:



L'acide aminé est attaché à l'AMP *via* une liaison acyle riche en énergie dans lequel le groupe carbonyle de l'acide aminé est lié au groupe phosphoryle de l'AMP. L'intermédiaire formé par adénylylation (transfert d'AMP) reste lié solidement à l'enzyme jusqu'à ce qu'il y ait une rencontre avec l'ARNt appropriée, l'acide aminé est transféré sur l'extrémité 3' de l'ARNt sur le groupe hydroxyle 2' ou 3' avec libération de l'AMP.

La formation de l'aminoacyl-ARNt est très précise. L'anticodon du fait de son caractère unique est une région essentielle dans la reconnaissance par l'enzyme. Un petit nombre de nucléotides sont aussi impliqués dans cette reconnaissance.

Tab.1: Correspondance entre les triplets de bases de l'ARNm et les acides aminés et les non sens (codon stop).

| Nucleotide base | | | | | |
|-----------------|-----------------------|-------------------|-----------------------|--------------------|-------|
| First | Second | | | | Third |
| | Uracil (U) | Cytosine (C) | Adenine (A) | Guanine (G) | |
| Uracil (U) | F Phenylalanine (Phe) | S Serine (Ser) | Y Tyrosine (Tyr) | C Cysteine (Cys) | U |
| | F Phenylalanine (Phe) | S Serine (Ser) | Y Tyrosine (Tyr) | C Cysteine (Cys) | C |
| | L Leucine (Leu) | S Serine (Ser) | Stop Codon | Stop Codon | A |
| | L Leucine (Leu) | S Serine (Ser) | Stop Codon | W Tryptophan (Trp) | G |
| Cytosine (C) | L Leucine (Leu) | P Proline (Pro) | H Histidine (His) | R Arginine (Arg) | U |
| | L Leucine (Leu) | P Proline (Pro) | H Histidine (His) | R Arginine (Arg) | C |
| | L Leucine (Leu) | P Proline (Pro) | Q Glutamine (Gln) | R Arginine (Arg) | A |
| | L Leucine (Leu) | P Proline (Pro) | Q Glutamine (Gln) | R Arginine (Arg) | G |
| Adenine (A) | I Isoleucine (Ile) | T Threonine (Thr) | N Asparagine (Asn) | S Serine (Ser) | U |
| | I Isoleucine (Ile) | T Threonine (Thr) | N Asparagine (Asn) | S Serine (Ser) | C |
| | I Isoleucine (Ile) | T Threonine (Thr) | K Lysine (Lys) | R Arginine (Arg) | A |
| | Start (Methionine) | T Threonine (Thr) | K Lysine (Lys) | R Arginine (Arg) | G |
| Guanine (G) | V Valine (Val) | A Alanine (Ala) | D Aspartic acid (Asp) | G Glycine (Gly) | U |
| | V Valine (Val) | A Alanine (Ala) | D Aspartic acid (Asp) | G Glycine (Gly) | C |
| | V Valine (Val) | A Alanine (Ala) | E Glutamic acid (Glu) | G Glycine (Gly) | A |
| | V Valine (Val) | A Alanine (Ala) | E Glutamic acid (Glu) | G Glycine (Gly) | G |

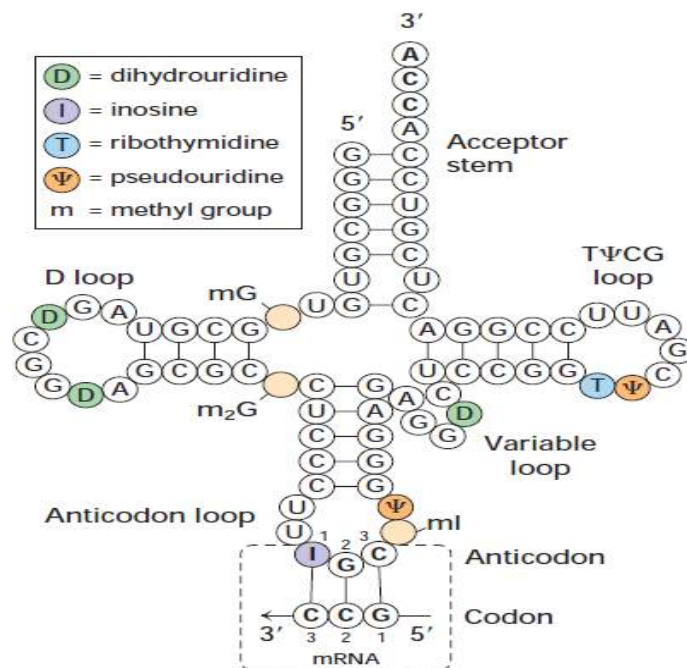


Fig.19: Structure de l'ARNt déchargée (sans acide aminé). Cette structure dite en feuille de trèfle. L'acide aminé se fixe au niveau du ribosome du A terminal de l'extrémité acceptrice. Nucléotides modifiés: ψ , I, T, et D. m: méthyle (Lodish *et al.*, 2003).

2-2- Ribosomes

Une cellule en phase active de croissance possède des milliers de ribosomes. Ce sont les sites de synthèse protéique. Chaque ribosome est composé de deux sous unités 30S (unité Svedberg: unité de coefficient de sédimentation) et 50S, l'ensemble donne des ribosomes 70S. La machinerie de polymérisation des acides aminés est composée d'au moins 3 ARN et 50 protéines différentes. Les ARNm procaryotes possèdent un site de liaison au ribosome (RBS) appelé séquence de Shine Dalgarno (S/D). Cet élément typiquement localisé 3 à 9 nucléotides en amont du codon d'initiation, il s'apparie avec la séquence complémentaire localisée sur l'ARNr 16S (fig.20).

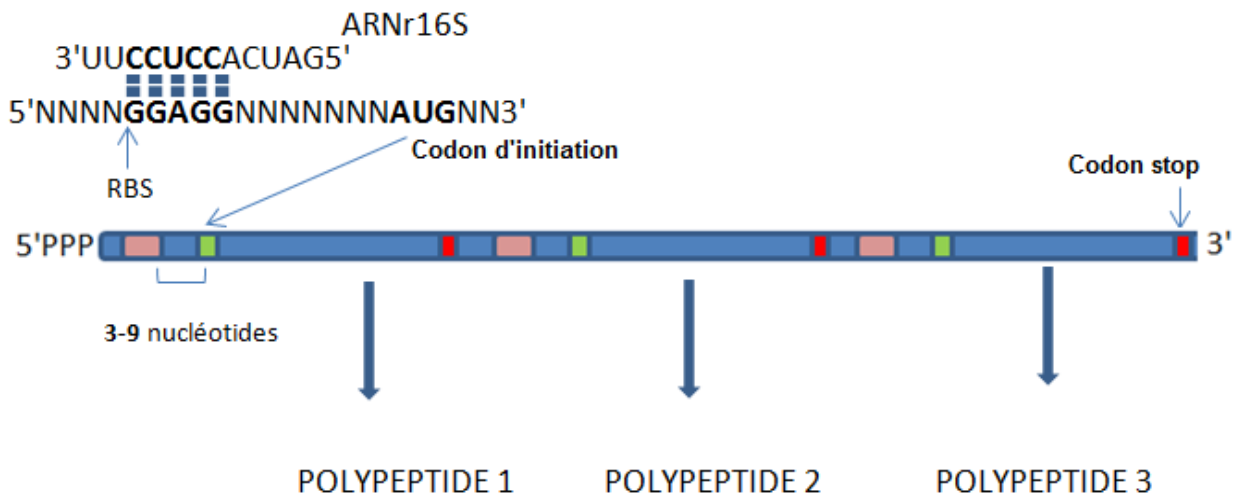


Fig.20: Structure d'un ARNm procaryote polycistronique. Chaque site de liaison au ribosome est indiqué RBS (Ribosome Binding Site).

La grande sous-unité contient le centre peptidyl-transférase qui est responsable de la formation des liaisons peptidiques. La petite sous-unité contient le centre de décodage dans lequel les ARNt chargés lisent ou décodent les codons de l'ARNm.

Les sous-unités ribosomiques sont composées d'un ou plusieurs ARNr et de nombreuses protéines (fig. 21).

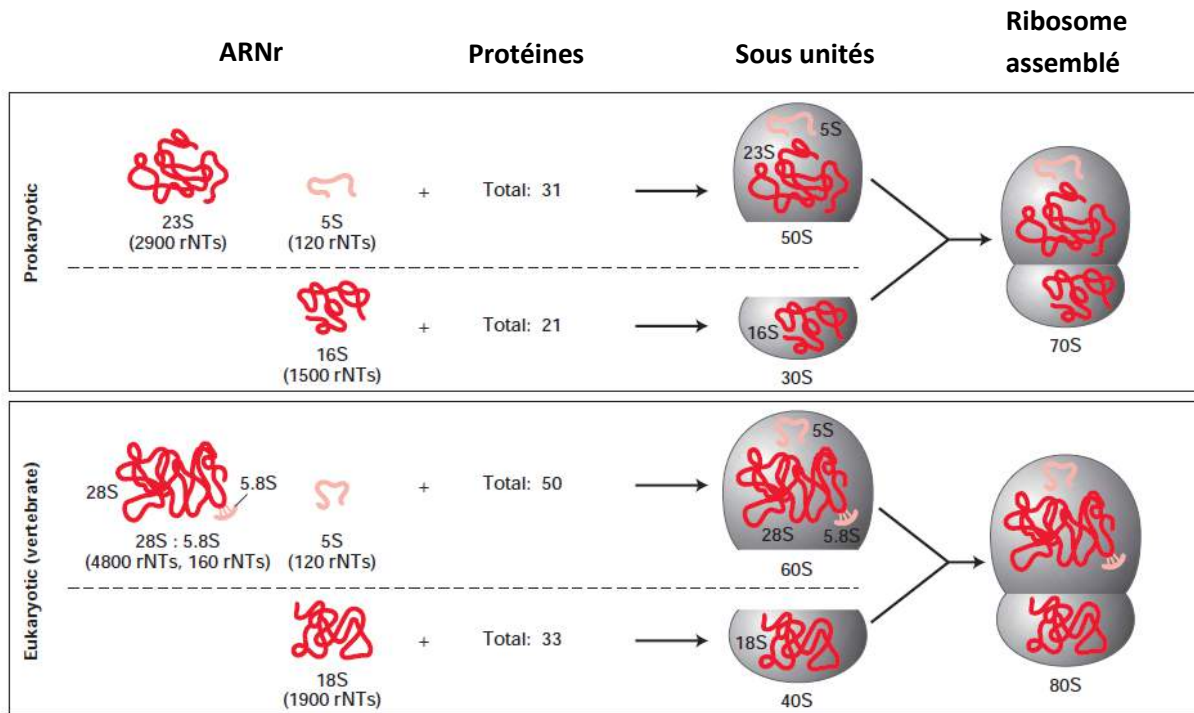


Fig.21: Structure générale des ribosomes procaryotes et eucaryotes. La figure montre la composition en ARNr et en protéines des différents sous unités, de même que la longueur des ARNr et le nombre de protéines ribosomiques.

2-3- Rôles des ARN dans la traduction

En plus du rôle de l'ARNm qui décode l'information génétique portée par l'ADN, l'ARNt joue le rôle d'adaptateur entre les codons et les acides aminés qu'ils spécifient. Et 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) (fig. 22).

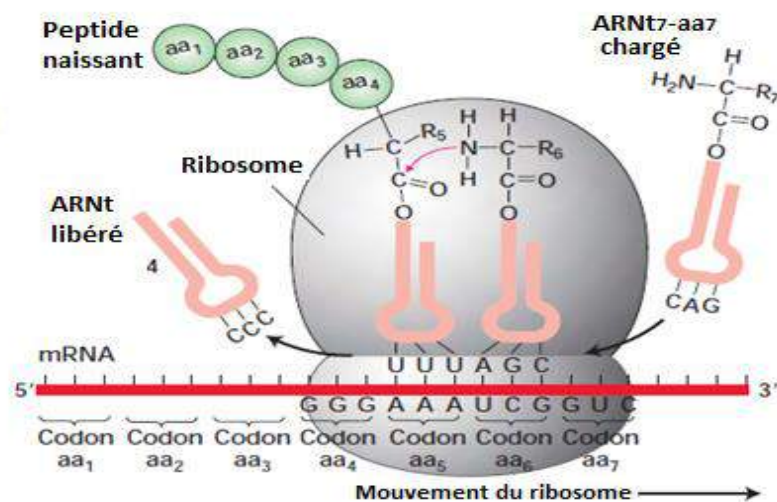


Fig.22: Les rôles des ARN dans la traduction(Clark, 2005).

3-Étapes de la traduction chez les procaryotes

La synthèse protéique par la voie ribosomale est un processus complexe. C'est aussi un processus continu mais il est possible d'y distinguer trois étapes: l'initiation, l'élongation et la terminaison. Au delà des aminoacyl-ARNt synthétases, des ARNt, les ribosomes, et les ARNm cette synthèse nécessite plusieurs protéines appelées : facteurs d'initiation, d'élongation et de terminaison, tandis que l'énergie est fournie par la molécule de GTP (fig.23, 24, 25, 26, 27 et 28)

Les structures différentes dans les ARNm procaryotes et eucaryotes impliquent des voies différentes pour ces événements.

3-1-Initiation de la traduction

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

- 1- Le ribosome doit être recruté sur l'ARNm
- 2- Un ARNt initiateur chargé par N-formyl-Met doit s'insérer dans le site P du ribosome
- 3- Le ribosome doit être positionné de manière précise sur le codon d'initiation.

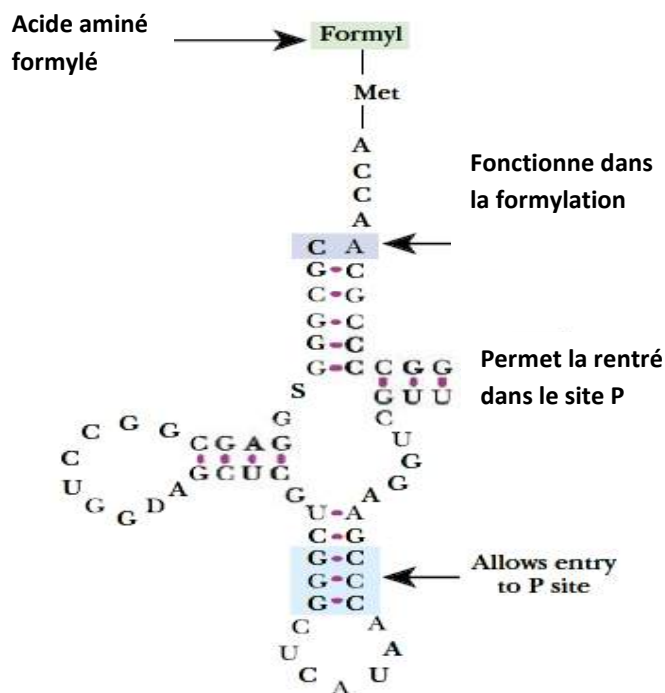


Fig.23: ARNt initiateur procaryote chargé par N-formyl-Méthionine. Les bases surlignées en bleu claire permettent la rentrée dans le site P du ribosome (Lodish *et al.*, 2003).

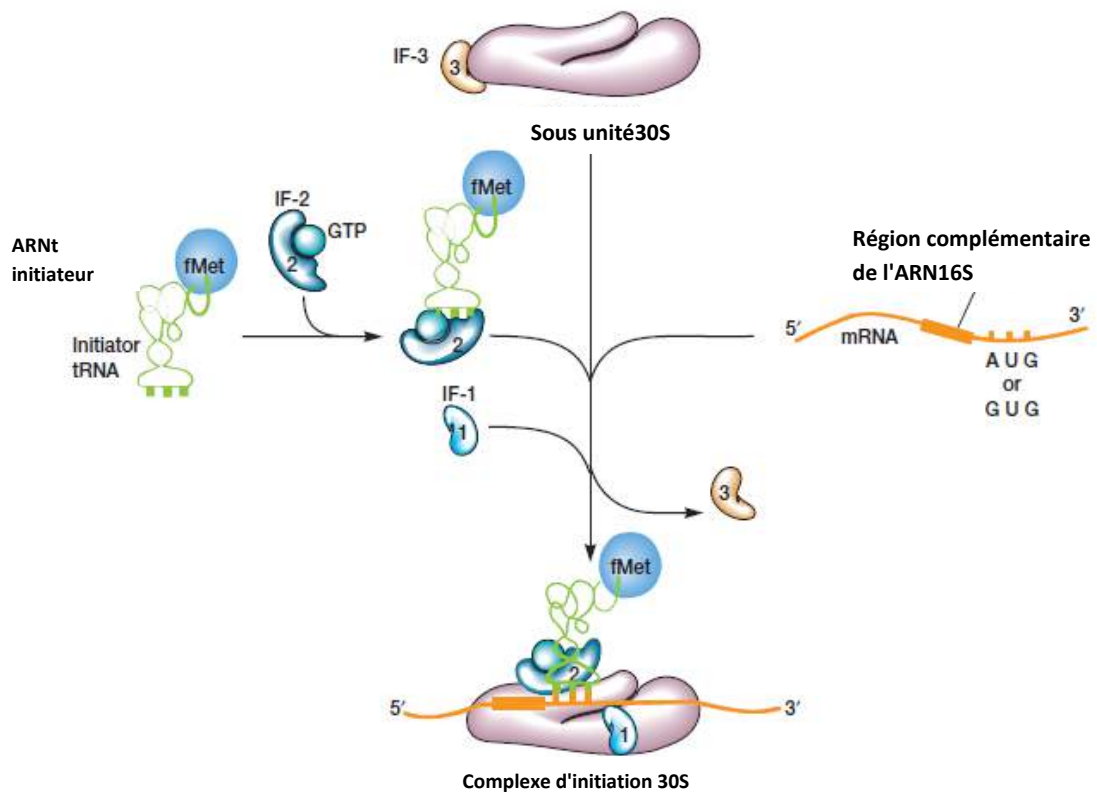


Fig.24: Le processus de l'initiation de la traduction chez les procaryotes. Le *N*-formylmethionyl- ARNt initiateur se lie d'abord à la sous-unité libre 30S. Ensuite, l'ARNm se lie à la sous-unité 30S et se positionne correctement par des interactions à la fois avec l'extrémité 3' de l'ARNr 16S et avec l'anticodon fmét-ARNt. Finalement la sous-unité 50S s'attache au complexe Sous-unité 30S –ARNm.

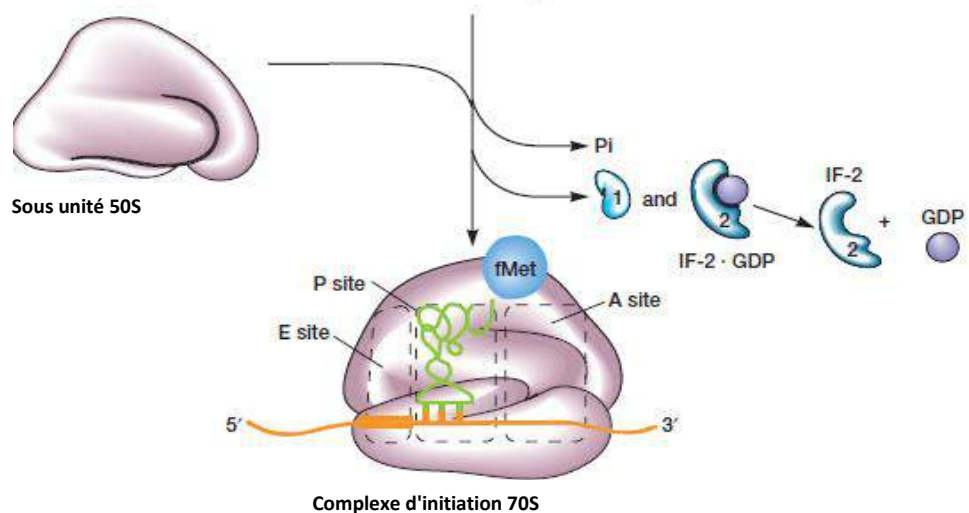


Fig.25: Formation du complexe d'initiation 30S puis 70S. IF: initiation factor. A: Aminoacyl or acceptor site, P: peptidyl site, E: Exit site (Prescott, 2002).

Chez les procaryotes trois facteurs d'initiation dirigent l'assemblage d'un complexe d'initiation.

1-Facteur d'initiation 3 : IF-3 se lie à la petite sous-unité et l'empêche de s'associer avec la grande sous-unité libre, et favorise la liaison correcte de l'ARNm.

2-Facteur d'initiation 2 : IF-2 c'est une protéine liant et hydrolysant la GTP (GTPase), il lie le GTP au fMet-ARNt initiateur pour faciliter sa liaison à la sous unité 30S et empêche la fixation de d'autre ARNt chargé.

3-Facteur d'initiation 1: IF-1 c'est un facteur semble nécessaire à la libération du IF-2 et du GDP. Il empêche la fixation d'ARNt sur la partie de la petite sous-unité qui fera partie du site A. IF-1 peut aussi faciliter l'assemblage des deux sous-unités.

Chacun des facteurs d'initiation se lie sur, ou à proximité, d'un des trois sites de liaison de l'ARNt sur la sous-unité 30S. En accord avec son rôle de bloquer de l'accès d'ARNt chargé au site A, IF-1 se lie directement à la zone de la petite sous-unité qui fera partie du site A. IF-2 se lie à IF-1 et s'étend jusqu'au site P où il prend contacte avec le fMet-ARNt_i^{fMET}. Enfin, IF-3 occupe la zone de la sous-unité qui fera partie du site E. Ainsi, des trois sites de liaison d'ARNt potentiels sur la petite sous unité, seul le site P est disponible pour lier un ARNt en présence des facteurs d'initiation. L'appariement de bases entre le fMet-ARNt et l'ARNm permet de positionner le codon d'initiation au niveau du site P.

Quand le codon d'initiation et le fMet-ARNt s'apparient, la sous-unité 30S subit un changement de conformation. Cette modification entraîne la libération de IF-3, en absence de cet facteur la grande sous-unité peut se lier à la petite. Cette liaison stimule l'activité GTPase d'IF-2-GTP provoquant l'hydrolyse du GTP et la libération d'IF-2-GDP à cause de la réduction de l'affinité au ribosome et fMet-ARNt_i^{fMet}.

Le résultat de l'initiation est la formation d'un ribosome complet (70S) un niveau du site d'initiation de l'ARNm, avec le fMet-ARNt_i^{fMet} occupant le site P, et un site A libre.

Les aminoglycosides comme la streptomycine sont des antibiotiques largement utilisés contre les infections bactériennes comme la tuberculose. Ces antibiotiques inhibent la formation du complexe d'initiation 70S en bloquant l'assemblage des deux sous unités.

3-2-Élongation

Une fois le ribosome 70S assemblé, avec l'ARNt initiateur chargé au site P, la synthèse de polypeptide peut commencer. Trois événements clés doivent se produire pour l'addition correcte des acides aminés.

- 1- La liaison de l'aa-ARNt au niveau du site A en concordance avec le codon qui est exposé.
- 2- La réaction de transpeptidation.
- 3- La translocation vers le site P.

Après la translocation (libération du site A) le ribosome est prêt pour un autre cycle de polymérisation. Deux protéines auxiliaires (facteurs d'élongation) contrôlent ces événements. Les deux facteurs utilisent la GTP comme source d'énergie pour accroître la vitesse et la précision du fonctionnement du ribosome. L'aminocyl-ARNt est ramené au site A par le facteur d'élongation EF-Tu, ce facteur est lié à une GTP. L'activité GTPase est activée juste après l'établissement d'un appariement codon-anticodon correcte, et le EF-Tu-GDP sera libéré et converti en EF-Tu-GTP de nouveau par l'intermédiaire d'un deuxième facteur d'élongation appelé EF-Ts.

Après la réaction peptidyl-transférase, l'ARNt du site P est désacylé est l'acide aminé se lie à l'ARNt chargé du site A. Au cours de cette réaction, le groupe α -aminé de l'acide aminé au site A réalise une attaque nucléophile du groupe α -carboxyle de l'acide aminé lié à l'ARNt du site P (Position C-terminale). L'ARNr 23 S est le composant majeur de la peptidyl transférase, il n'ya pas des protéines dans la région du site actif. Une adénine spécifique semble participer à la catalyse de la formation de la liaison peptidique.

Le chloramphénicol c'est un antibiotique inhibant la synthèse des protéines chez les bactéries. Il empêche le positionnement correcte de l'Amino-acyl-ARNt dans le site A pour la réaction peptidyl transférase dans la grande sous-unité.

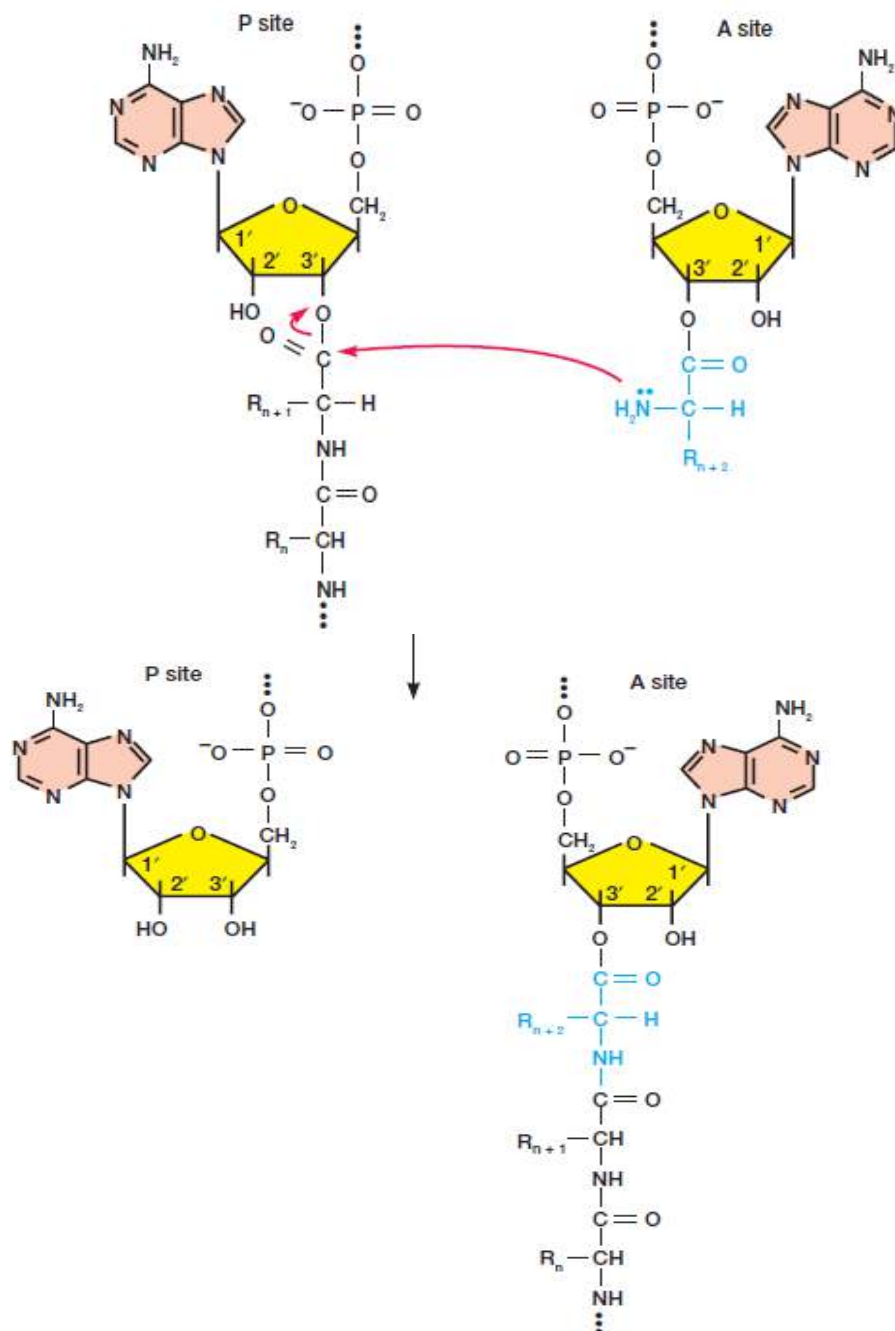


Fig.26: La réaction de transpeptidation (Clark, 2005).

Pour qu'une nouvelle élongation puisse se produire l'ARNt du site P doit se déplacer vers le site E et l'ARNt du site A doit se déplacer au site P. En même temps l'ARNm doit se déplacer de trois nucléotides pour présenter le codon suivant. Cette phase finale d'élongation nécessite l'EF-G ou translocase et l'hydrolyse de GTP.

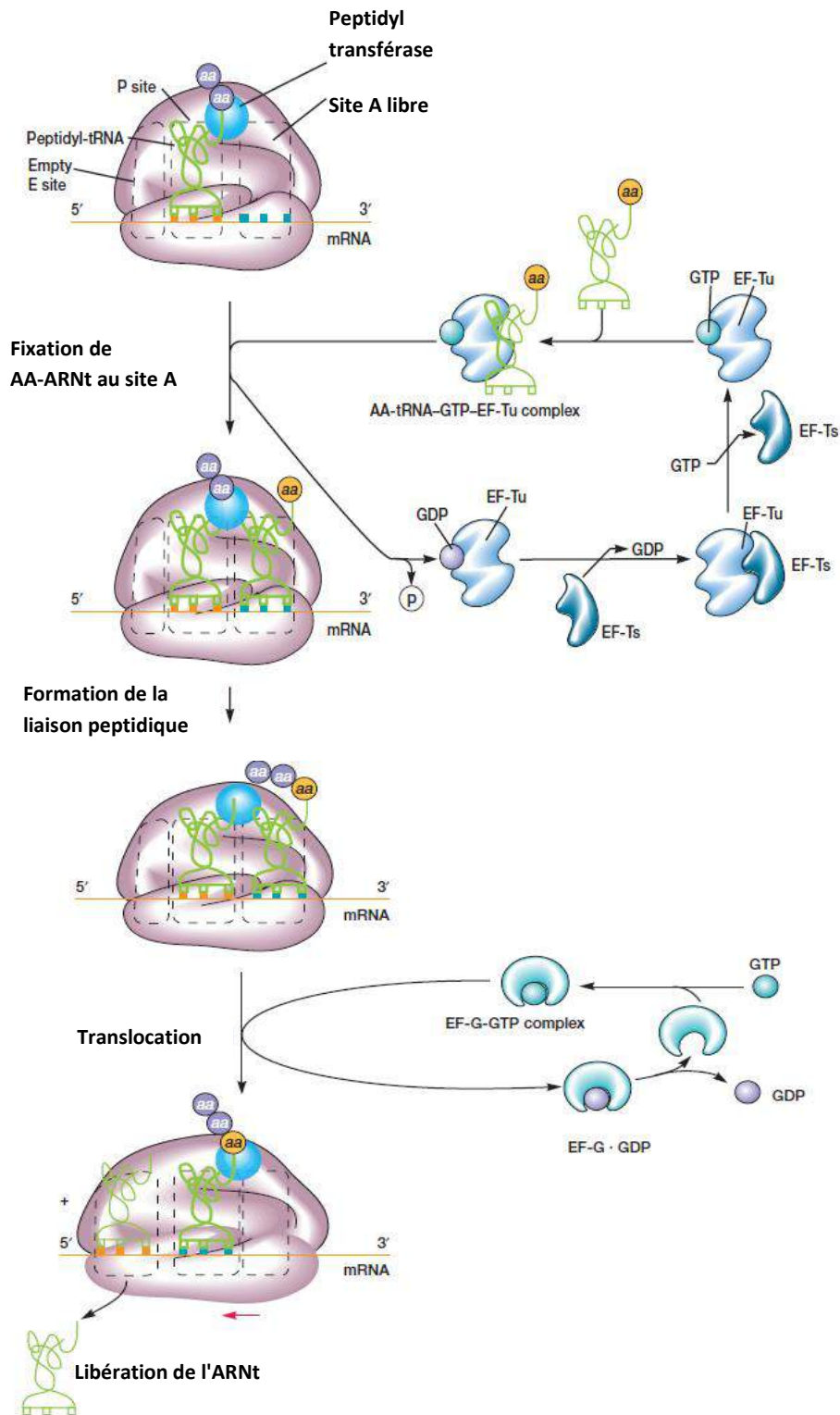


Fig.27: Etape d'élongation de la synthèse protéique.

3-3- Terminaison de la traduction

La terminaison de la traduction se fait quand un codon stop (non-sens: UAA, UAG, et UGA) est atteint, les ARNt ne pouvant plus s'y fixer. Trois facteurs de libération RF (RF-1, RF-2, et RF-3) aident le ribosome à reconnaître ces séquences et coupent le polypeptide attaché à l'ARNt terminal, libérant le produit fini. Par la suite les sous-unités se dissocient, elles peuvent alors former de nouveaux complexes d'initiation et recommencer le processus.

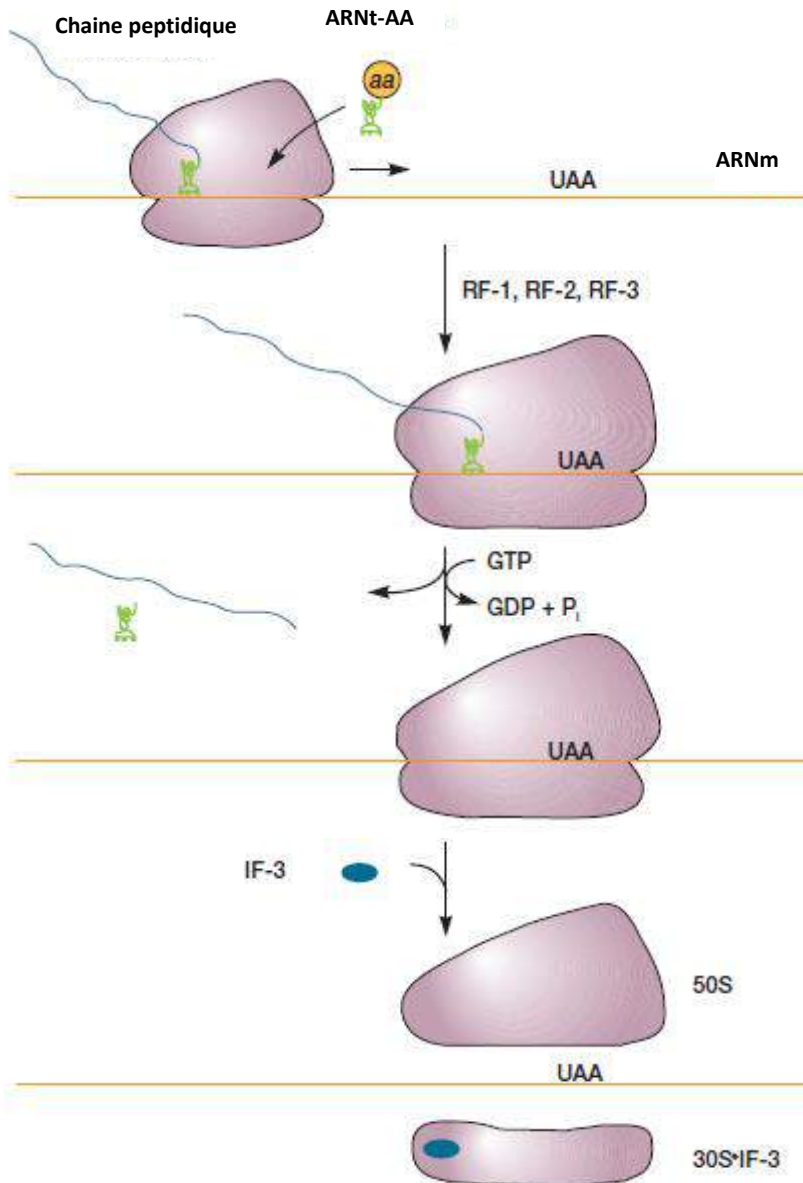


Fig.28: Étape de terminaison de la traduction (Prescott, 2002).

4- Traduction chez les eucaryotes

Comme chez les procaryotes on distingue trois étapes dans la synthèse protéique chez les eucaryotes (initiation, élongation et terminaison).

4-1- Initiation de la traduction chez les eucaryotes

L'initiation est similaire à celle chez les procaryotes par plusieurs aspects. Les deux utilisent un codon d'initiation et un ARNt initiateur, et les deux font appel à des facteurs d'initiation pour former un complexe avec la petite sous-unité ribosomique qui lie l'ARNm avant l'ajout de la grande sous-unité. Néanmoins, les eucaryotes utilisent une méthode complètement différente pour reconnaître l'ARNm et le codon d'initiation.

Chez les eucaryotes, la petite sous unité est déjà associée avec un ARNt initiateur. Au contraire des procaryotes la liaison de l'ARNt initiateur (chargé par une méthionine non modifiée) à la sous unité 40S précède toujours l'association avec l'ARNm. La reconnaissance de l'ARNm fait appel à plusieurs facteurs auxiliaires. 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). Enfin, la grande sous unité du ribosome est recruté Après l'appariement de bases de l'ARNt initiateur avec le codon d'initiation. A la fin d'un cycle de traduction, le ribosome eucaryote se dissocie en ses sous-unités constitutives (fig. 29, 30 et 31).

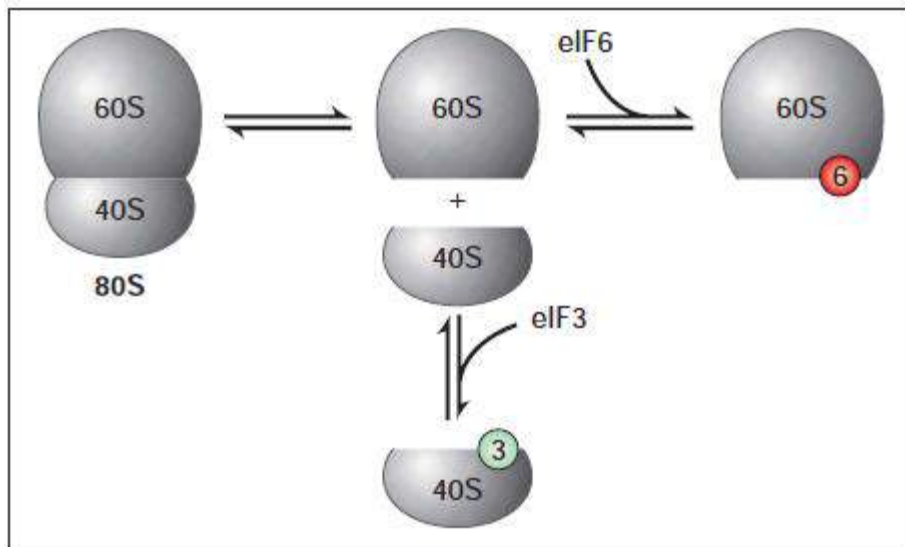


Fig.29: Dissociation du ribosome procaryote (Lodish *et al.*, 2003).

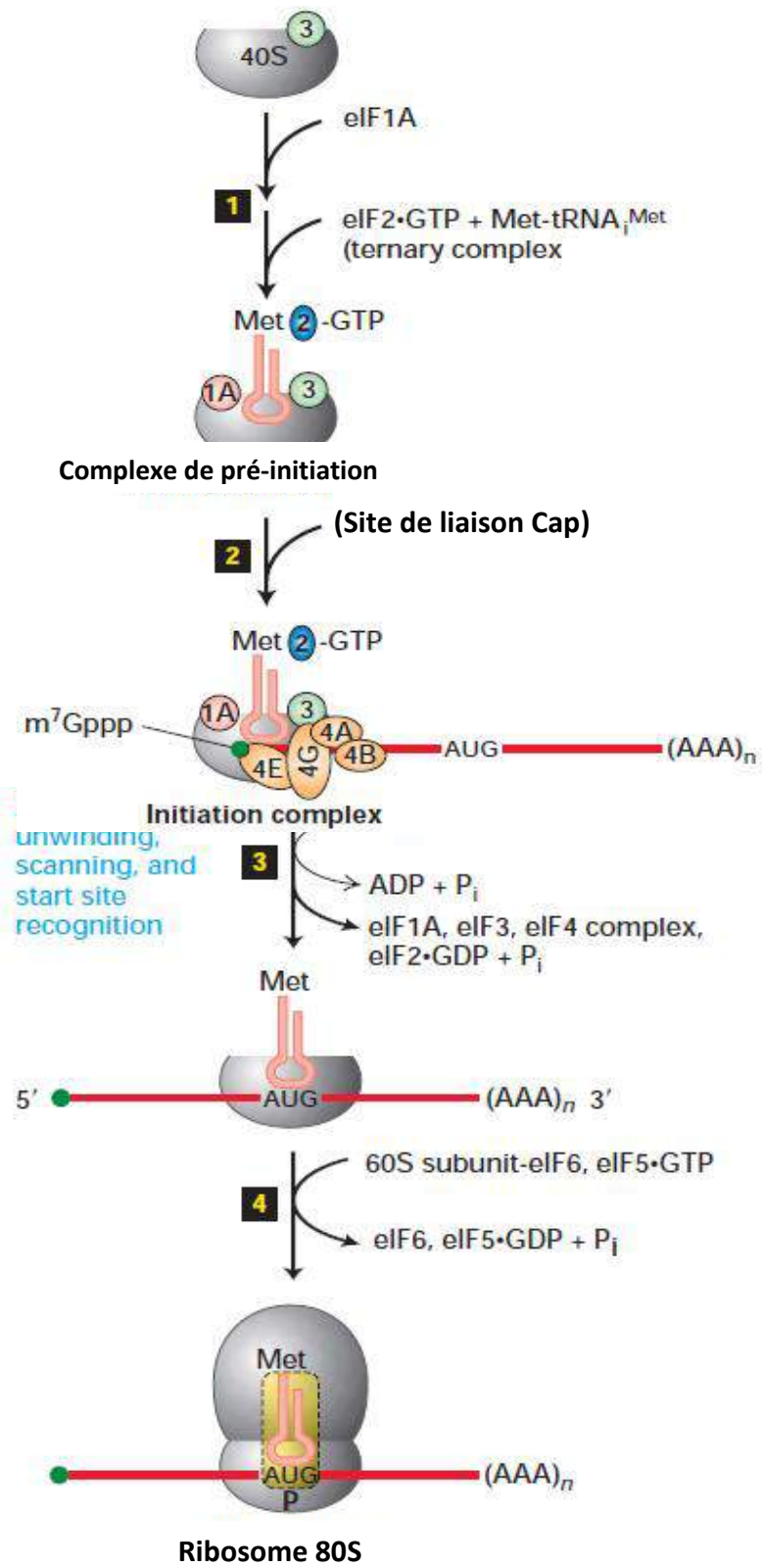


Fig.30: Formation du complexe d'initiation chez les eucaryotes(Lodish *et al.*, 2003).

4-2- Élongation

Une fois le ribosome assemblé, avec l'ARNt initiateur chargé au site P, la synthèse du polypeptide peut commencer (fig. 16).

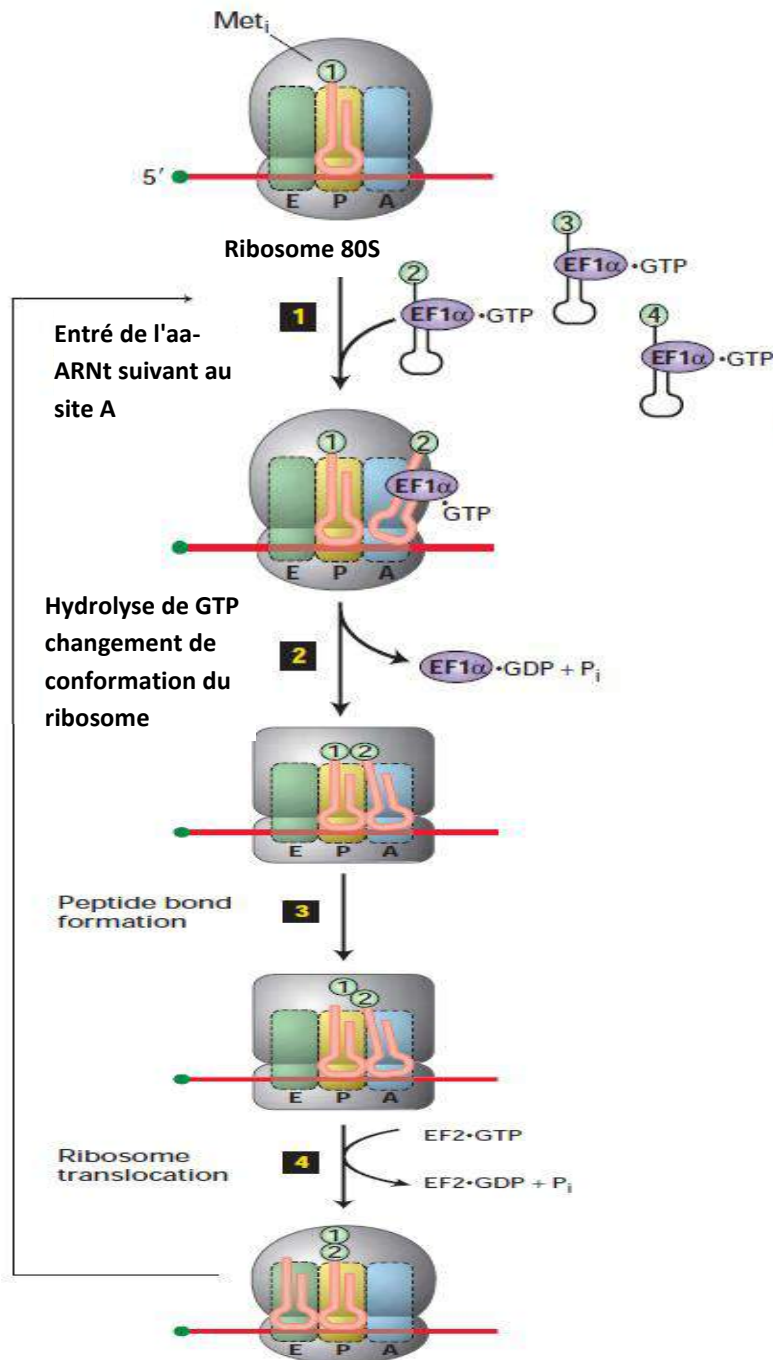


Fig.31:Élongation de la traduction chez les eucaryotes(Lodish *et al.*, 2003)

4-3- 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.

Biologie Moléculaire

Chapitre III

Régulation de l'expression génétique

Objectifs

- Différents niveaux de la régulation
- Différents mécanismes de la régulation de l'expression génétique
- Importance de la régulation de l'expression génétique

1- Introduction

La régulation de l'expression génétique des cellules procaryotes et eucaryotes c'est un sujet extrêmement compliqué et la compréhension de la nature de cette régulation est complexe. Les cellules microbiennes ont divers mécanismes par lesquels régulent leurs activités en réponse au changement de l'environnement externe, afin d'optimiser ses fonctions principales (nutrition, croissance, résistance...etc.). Ex1: l'augmentation de la température ou épuisement des nutriments pousse des souches de *Bacillus* sp. à la sporulation, Ex2: la présence du lactose dans le milieu pousse des souches d'*E. coli* à produire une β -galactosidase et une perméase afin de l'utiliser comme source de carbone et d'énergie, en absence bien sûr de sources plus facilement assimilables comme le glucose).

Dans les chapitres précédents nous avons vu en détails les différentes étapes de l'expression génétique chez les cellules procaryotes et eucaryotes: "Comment un gène se transcrit en ARNm et ce dernier se traduit à un polypeptide". Dans ce chapitre on va traiter la régulation de l'expression génétique aux différents niveaux.

La régulation de l'expression génétique c'est un processus d'intérêt majeur pour toutes les cellules, elle permet d'économiser l'énergie et les ressources. Il y a deux modes majeurs de régulation dans les cellules:

- 1- Régulation de l'activité des enzymes présentes (post-traductionnelle).
- 2- Régulation de la production des enzymes (transcriptionnelle ou post-transcriptionnelle).

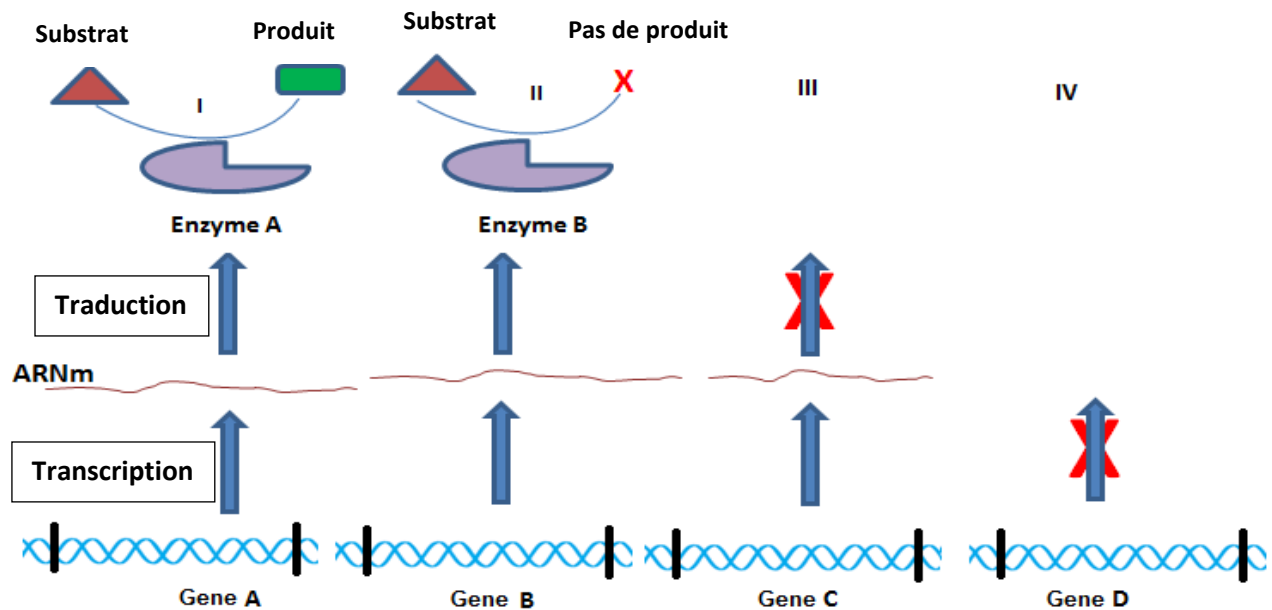


Fig.1: Les différents niveaux de la régulation de l'expression génétique [I: Absence de contrôle, II: Contrôle de l'activité enzymatique (post traductionnel). III: Contrôle traductionnel (pas de synthèse de l'enzyme). IV: Contrôle transcriptionnelle (pas de synthèse d'ARNm).

2- Régulation de l'activité enzymatique:

Les enzymes sont généralement des protéines (il ya des ARN catalytiques: ribozymes) dont la fonction d'augmenter la vitesse de catalyse des réactions chimiques. Chaque enzyme a un **E.C.** (Enzyme Commission number) spécifique : **E.C.** 1^{er} chiffre. 2^{eme}chiffre. 3^{eme} chiffre . 4^{eme} chiffre.

- Premier chiffre indique la classe : Ex. **1:** Oxydoréductases, **2:** Transférases, **3:** Hydrolases, **4:** Lyases, **5:** Isomérases, **6:** Ligases.
- Deuxième chiffre indique la sous-classe: Ex. Les enzymes de restriction sont de la sous classe des estérases [1].
- Troisième chiffre indique la sous sous-classe : Ex. les enzymes de restriction sont de la sous sous classe **21** (coupe avant le phosphate 5': le dernier nucléotide garde son 5' phosphate)
- Quatrième chiffre indique le type de coenzyme: les enzymes de restriction de type II sont de la sous sous sous classe **4** (utilise Mg^{++} comme cofacteur) Ex. EC3.1.21.4

Il y'adeux types d'enzymes:

a-Enzymes constitutives : La synthèse et l'activité de ces enzymes ne subit pas de contrôle, car elles sont nécessaires en quantités égales quelle que soit les conditions de croissance.

b- Enzymes inductibles (adaptatives): Ces enzymes sont produites uniquement en présence du substrat (exemple: la bêta-galactosidase).

La régulation de l'activité enzymatique s'effectue par deux moyens principaux :

- 1- Le contrôle de la quantité d'enzyme : Des systèmes complexes modulent la quantité d'enzyme en réponse à des conditions changeantes (Exemple le cas des enzymes inductibles).

2- Le contrôle de l'activité enzymatique: L'activité enzymatique peut être directement modifiée par des modifications covalentes et conformationnelles, ou par des modifications non-covalentes en modulant l'affinité des enzymes au substrat par des molécules ou sous-unités protéiques régulatrices.

2-1- Retro-inhibition (Feed Back or end-product inhibition)

Avant de parler sur la rétro-inhibition on doit savoir la notion des termes suivant:

Enzyme allostérique: Une enzyme qui possède deux sites de fixation, le site actif (sur lequel se fixe le substrat) et le site allostérique ou modulateur (sur lequel se fixe l'effecteur).

Allostérie: C'est un mécanisme d'inhibition enzymatique par un effecteur allostérique, lorsque l'effecteur se fixe de façon non-covalente au niveau du site allostérique, la conformation spatiale de l'enzyme est modifiée et le substrat ne plus se fixer au site actif.

La rétro-inhibition c'est un mécanisme majeur du contrôle de l'activité enzymatique. Elle assure essentiellement la régulation des voies de biosynthèse complètes. Comme la synthèse des acides aminés et les nucléotides (exemple: synthèse des acides aminés aromatiques).

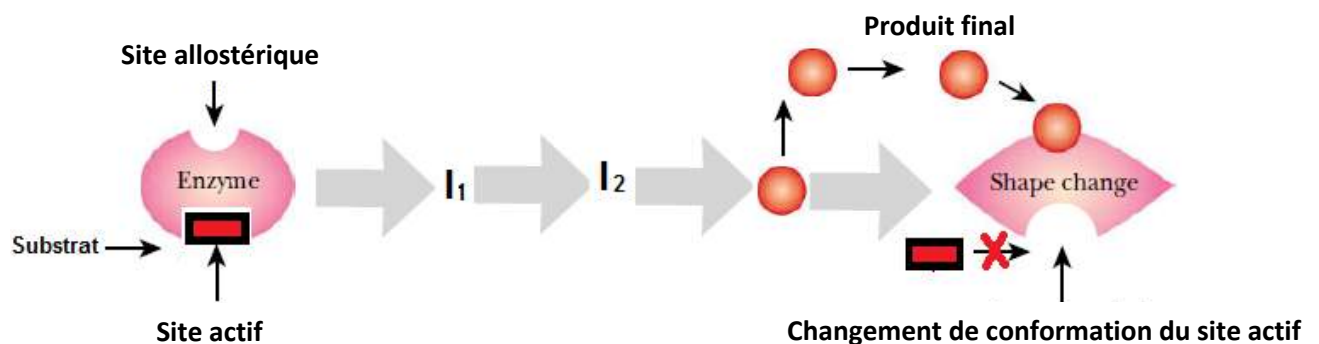


Fig. 2: Rétro-inhibition d'une activité enzymatique. A l'absence de l'effecteur le substrat est transformé par l'enzyme en intermédiaires I1, I2 et puis en produit final. La fixation non covalente de l'effecteur (produit final: end product) sur le site allostérique change la conformation de l'enzyme et la rend inactive (incapable de fixer le substrat).

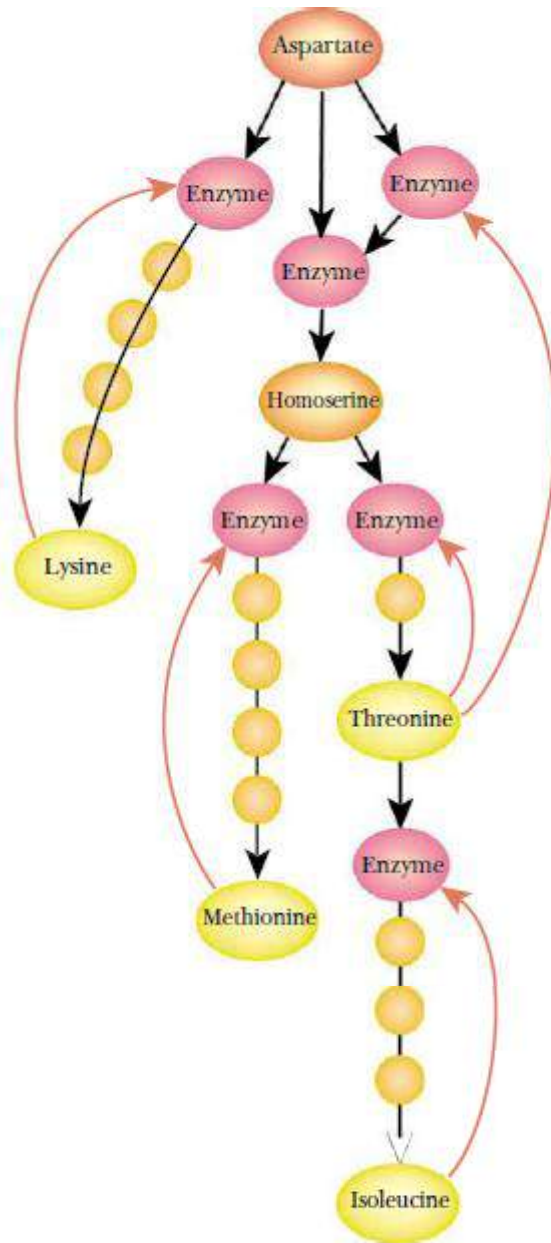


Fig.3: Biosynthèse des acides aminés dérivés de l'aspartate et les boucles de rétro-inhibition (les flèches en rouge indiquent les boucles de rétro-inhibition) Clark, 2005).

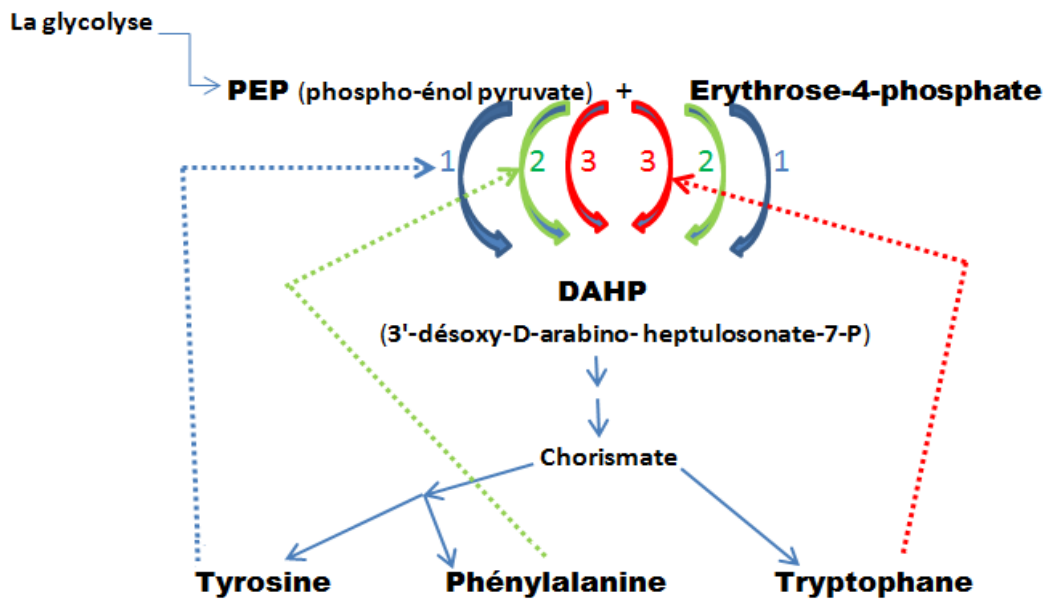


Fig.4: Les trois enzymes impliquées dans la synthèse de DAHP (précurseur des acides aminés aromatiques) chez *E. coli*, chaque enzyme est inhibée par un produit différent (**DAHP synthétases** : 1: Tyrosine, 2: Phénylalanine, 3: Tryptophane).

Les isoenzymes: Elles sont des enzymes catalysant la même réaction, mais soumises à des contrôles différents (Exemple: les DAHP synthétases [1, 2 et 3]).

Question: Comment pouvez vous surproduire la phénylalanine par les outils de biologie moléculaire et génie génétique?

2-2- Modification covalente des enzymes

Le contrôle de l'activité enzymatique par modification covalente est rare chez les bactéries, mais extrêmement communs chez les eucaryotes. Les mécanismes usuels de modification covalente comprennent l'ajout de l'AMP, ADP et le Pi. La phosphorylation par les kinases et la déphosphorylation par les phosphatases sont très utilisées pour ce type de contrôle, surtout chez les cellules animales (fig. 5).

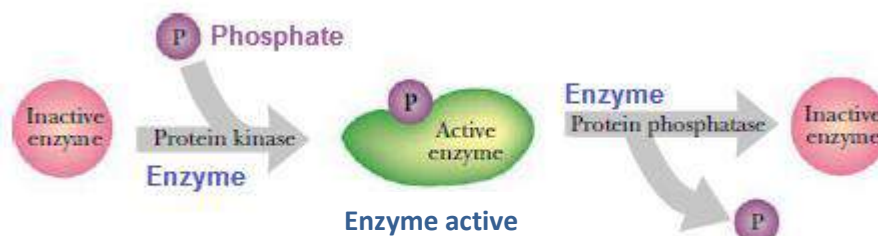


Fig. 5: La phosphorylation et la déphosphorylation contrôlent l'activité enzymatique (Clark, 2005).

Plusieurs exemples de modification covalente sont connus chez les bactéries, l'exemple le plus connu est la modulation de l'activité de la glutamine synthétase par l'AMP. Cette enzyme est composée de 12 sous-unités identiques, elle a un grand rôle dans l'assimilation de l'ammoniac (source d'azote) par la cellule. L'activité de cette enzyme est contrôlée par deux mécanismes, le premier est la rétro-inhibition (par neuf composés [acides aminés et des

intermédiaires du métabolisme des nucléotides]. Le deuxième est par modification covalente "adénylylation" [addition des AMP par P_{II}] (contrôlé par la glutamine et l' α -céto-glutarate). Lorsque la concentration de la glutamine diminue dans la cellule, P_{II} subit une modification covalente (uridylylation par l'UTP), cette modification entraîne la désadénylylation du GS et la synthèse De la glutamine se reprend de nouveau.

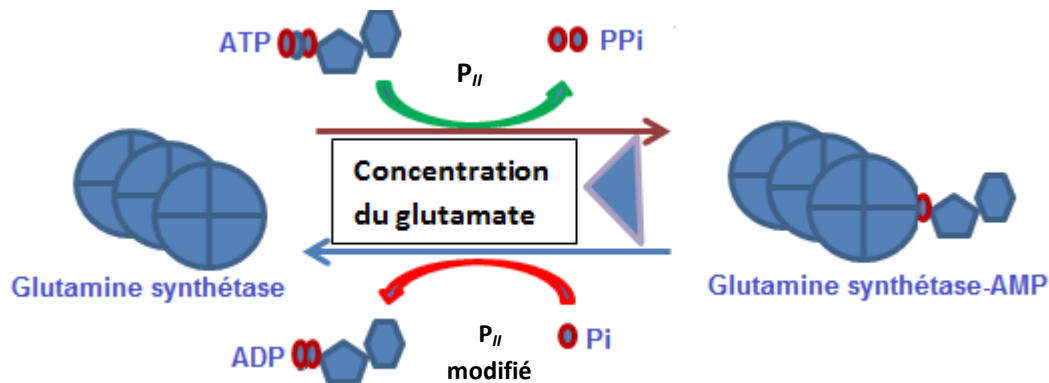


Fig.6: Contrôle de l'activité de la glutamine synthétase par adénylylation. Jusqu'à 12 groupes d'AMP peuvent être ajoutés. Plus qu'il ya addition des AMP (par P_{II}) plus qu'il ya réduction de l'activité enzymatique et vice-versa.

Exemples de modifications post-traductionnelles

Les principales modifications post-traductionnelles sont :

- Phosphorylation (Ser, Thr, ou Tyr: important dans la transduction de signaux à travers les membranes.
- L'élimination du formyl-méthionine ou méthionine par une enzyme spécifique (Met-aminopeptidase : MAP)se fait chez plus de 50% des protéines des deux types de cellules. Le N-formyl de la première méthionine est enlevé chez *E. coli* par une deformylase.
- Clivage de chaine polypeptidique (Ex. L'insuline)
- Glycosylation
- Addition des lipides
- Formation des ponts disulfures ou pont covalents (élastine, collagène).
- Acétylation (\approx 100 protéines cytoplasmique ont un N-terminal acétylé), la réaction est catalysée par N- α - acétyl transférase (exemple: acétylation des histones).
- Hydroxylation (hydroxylation des prolines et lysine)
- Biotinylation:Addition de la biotine [vit B8] (coenzyme de quelques enzymes de carboxylation [transfert de CO_2]).
- Élimination des séquences surnuméraires au niveau protéique, ce processus est appelé l'épissage protéique et le peptide éliminé est nommé intéine.

Les intéines qui assurent l'auto-épissage ont l'activité d'une protéase hautement spécifique. Comme par exemple, la maturation de la sous-unité A de l'ADN gyrase (Gyr A).

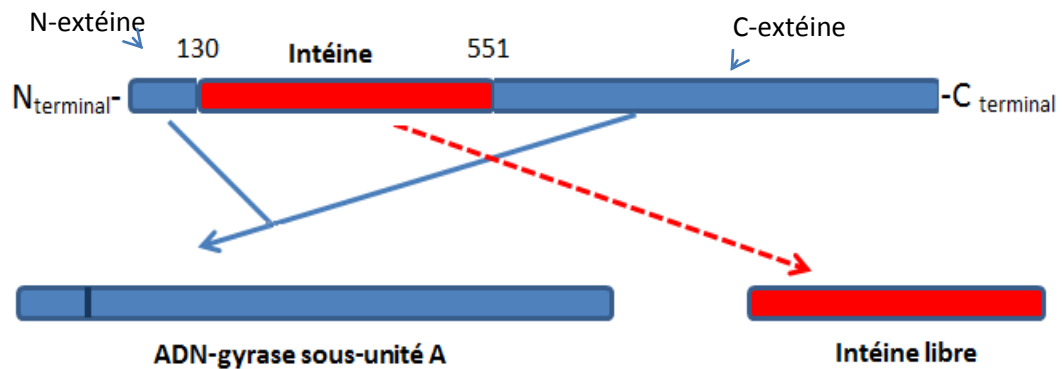


Fig.7: Epissage protéique: la protéine synthétisée par *Mycobacterium liprae* compte 1273 aa. L'auto-épissage produit une intéine (130-550) libre et la sous-unité A de l'ADN gyrase.

3- Protéines se liants à l'ADN

Les interactions protéines/acides nucléiques ont un rôle central dans la réplication, la transcription et la traduction, ainsi pour la régulation des ces processus. La plupart des protéines se liants à l'ADN présentent une spécificité de séquence, cette spécificité est le résultat de l'interaction avec plusieurs nucléotides (complémentarité moléculaires), ces séquences sont souvent des répétitions inversées (Ex. L'opérateur de l'opéron lactose). Et les protéines sont classiquement des homodimères (Ex. Répresseur de l'opéron lactose chez *E.coli*).

3-1-Structure des protéines se liant à l'ADN

Il existe trois motifs :

- 1- Le motif hélice-tour-hélice (Helix-Turn-Helix motif):** Ce motif est souvent observé chez les bactéries, bactériophages, et généralement chez les procaryotes et les eucaryotes

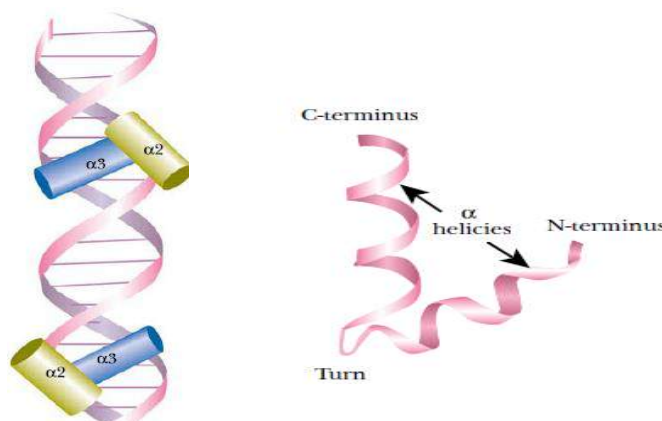


Fig.8: Le motif hélice-tour-hélice, et sa liaison à l'ADN.

2- Le motif doit de Zinc (Zink zipper)

Ce motif est fréquemment trouvé chez les eucaryotes (facteurs de transcription).

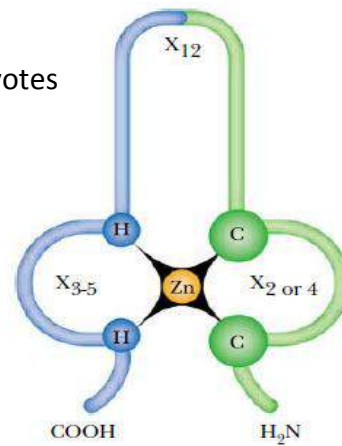


Fig. 9: Le motif doit de zinc

3- Le motif fermeture à leucine

Un motif fermeture à leucine ou glissière à leucine, (leucine zipper), est un motif structural tridimensionnel courant des protéines.

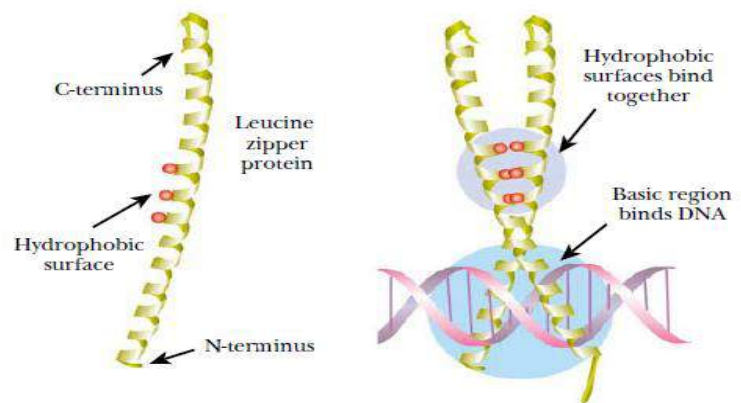


Fig. 10: Le motif « la fermeture à leucine » (glissière à leucine), lié à un segment d'ADN bicaténaire.

4- Régulation transcriptionnelle

L'expression d'un gène peut être régulée à plusieurs étapes, la plus commune étant celle de l'initiation de la transcription. Il y a deux raisons qui peuvent justifier cela.

- Le coût en énergie et en ressources pour la transcription
- La régulation lors de l'initiation est plus facilement réalisable.

Cette régulation est souvent contrôlée par des signaux extracellulaires, dans le cas des bactéries, ces signaux sont donnés par les molécules présentes dans le milieu. Ils sont ensuite transmis aux gènes par des protéines régulatrices. Ces protéines régulatrices contrôlent l'accès de l'ARN polymérase aux promoteurs de l'ADN bactériens (contrôle du taux de la transcription). Donc ces protéines aident les cellules à évaluer leur environnement. Les gènes codant ces protéines sont appelés les gènes régulateurs.

Il y a deux types de protéines régulatrices :

- 1- Les activateurs (régulateurs positifs: augmente la transcription d'un gène)
- 2- Les répresseurs (régulateurs négatifs: diminue ou bloque sa transcription)

4-1- Contrôle négatif de la transcription "Répression et induction"

Répression: Activité de synthèse d'une enzyme empêchée lorsque le produit de la réaction est présent en excès.

Induction: Production d'une enzyme seulement lorsque son substrat est présent.

Chez les bactéries, un opéron c'est un ensemble de gènes transcrit en un seul ARNm polycistronique sous la direction d'un seul promoteur (le mot cistron désigne la plus petite unité génétique servant à une seule fonction). Les gènes d'un opéron donné sont soumis à un contrôle coordonné (système de contrôle remarquablement économique).

L'exemple le plus étudié de ce type de régulation est le contrôle de l'opéron lactose d'*E.coli*. Historiquement, les expériences portées sur l'analyse génétique de mutants bactériens a permis de découvrir tous les éléments régulateurs du métabolisme du lactose (disaccharide présent dans le lait des mammifères), surtout les recherches de François Jacob et Jacques Monod à la fin des années 50.

4-1-1- Régulation de l'expression de l'opéron *lacZYA*

Lorsque *E. coli* est cultivé dans un milieu de culture où le lactose est la seule source de carbone, la concentration cellulaire de trois protéines augmente (fig. 11).

- **β -galactosidase:** enzyme hydrolysant le lactose en glucose et galactose.
- **Perméase:** Protéine de la membrane interne permettant l'absorption du lactose.
- **Transacétylase:** son rôle exact dans le métabolisme du lactose reste encore obscure.

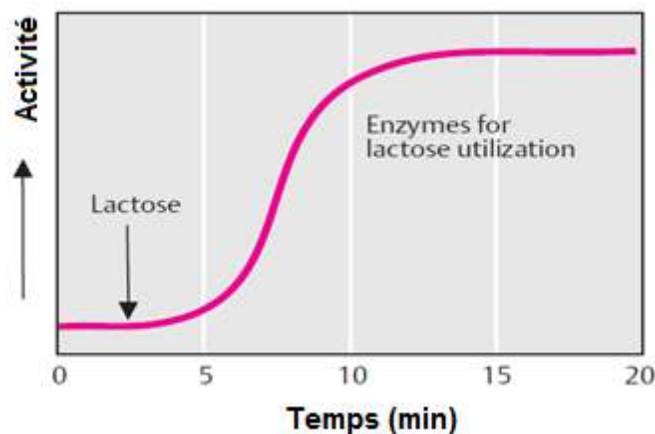


Fig.11: Induction par le lactose.

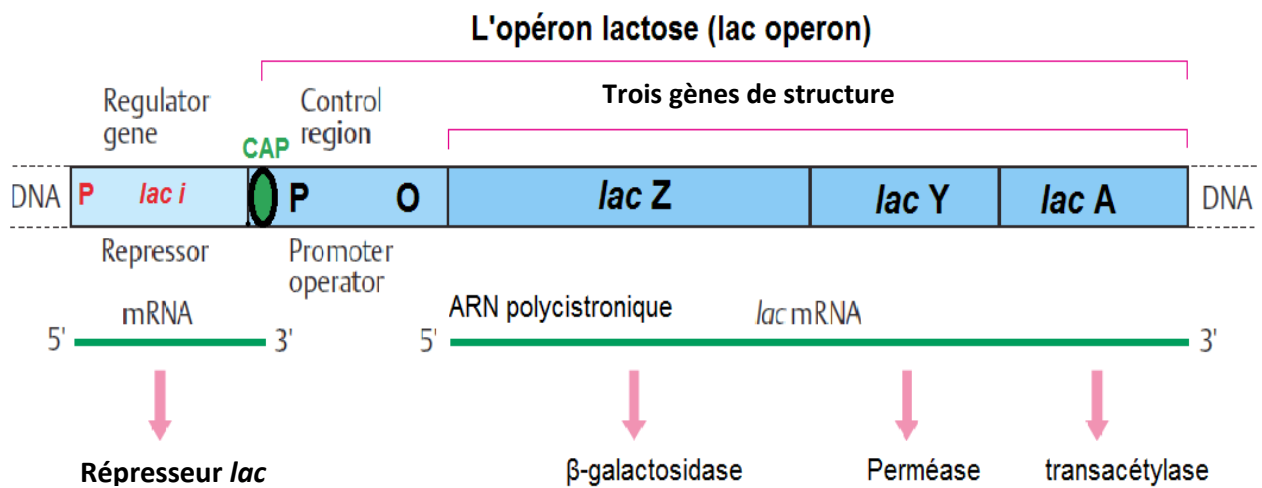


Fig.12: Organisation de l'opéron lactose d'*E. coli*(Passarge, 2007).

Les gènes de l'opéron *lac* ont un fort niveau d'expression seulement en présence du lactose, et en absence du glucose (source de carbone et d'énergie préférée). Deux protéines régulatrices sont impliquées:

- Un activateur appelé **CAP (Catabolite Activator Protein)**, connue aussi par le nom de CRP (**cAMP receptor protein**). Le gène codant CAP au contraire du ce codant le répresseur est localisé sur un autre endroit sur le chromosome. Il se fixe sur le site CAP.
- Le répresseur *lac* (codé par le gène *lac I*), il se fixe sur l'opérateur.

Chacune de ces protéines régulatrices répond à un signal de l'environnement et le communique aux gènes *lac*. Le CAP transmet le signal de glucose, alors que le répresseur transmet le signal du lactose.

Le blocage de l'expression de l'opéron lactose en présence du glucose ou autre source d'énergie et de carbone facilement assimilable que le lactose est appelé la répression catabolique. L'une des conséquences de cette répression est la croissance diauxique ou triauxique (selon le nombre).

L'opéron *lac* ne possède pas un promoteur fort (faible séquences consensus surtout la région -35) pour cette raison la fixation de l'ARN polymérase doivent être stimulée par une protéine d'activation spécifique (contrôle positif), appelé protéine réceptrice d'AMPc (CRP) ou protéine activatrice du catabolisme (CAP).

En absence du glucose, les niveaux d'AMPc (synthétisée à partir de l'ATP par l'adénylate cyclase) augmentent, aboutissant à la fixation de la CRP à l'AMPc. Le complexe CRP-AMPc se fixe sur le site CAP. L'ADN subit alors une inclinaison de 90°, qui stimule la fixation des éléments *trans* de l'ARN polymérase aux éléments *cis* du promoteur pour augmenter le taux de la transcription de 50 fois.

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. L'ARN polymérase activée par le complexe CRP-AMPc peut initier la synthèse de l'ARN avec un taux de transcription élevé.

La protéine régulatrice CAP ou CRP avec l'AMPc c'est un exemple typique de la régulation globale de l'expression génétique (contrôle positif et négatif). Ce qui permet à la cellule d'économiser l'énergie et les ressources au maximum.

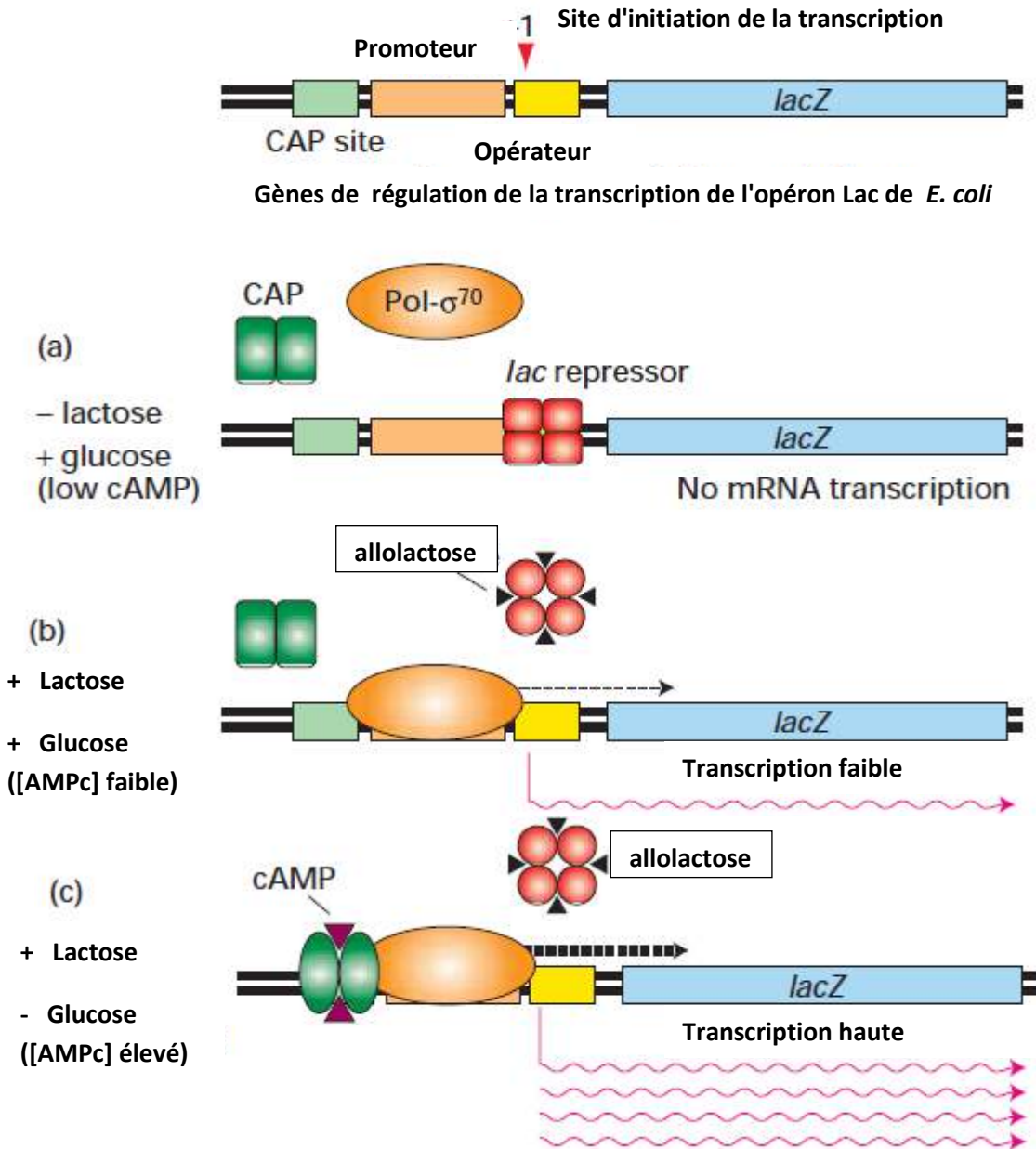


Fig.13: Expression des gènes *lac* en présence/absence du glucose et le lactose. Quand le répresseur Lac est lié sur l'opérateur, il exclut la polymérase, que CAP soit présente ou non. CAP recrute la polymérase sur le promoteur *lac* où elle s'isomérisse spontanément en complexe ouvert. L'allolactose (inducteur) résulte d'une transglycosylation du lactose (Lodish et *al.*, 2003).

4-1-2- Régulation de l'expression de l'opéron *argCBH*

Un autre exemple de la régulation de l'expression génétique par les mécanismes de répression et d'induction est le contrôle de la production de l'arginine.

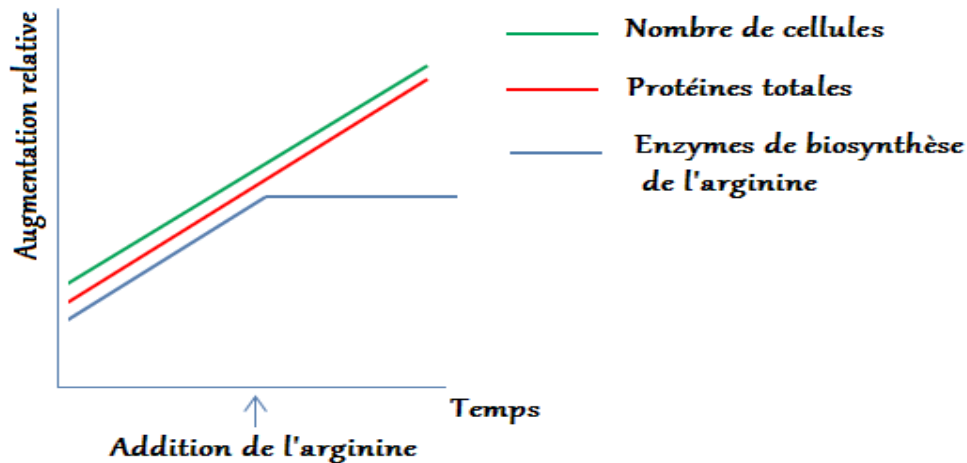


Fig.14: Effet de l'addition de l'arginine dans un milieu de culture dépourvu de l'arginine et ensemencé par *E. coli*.

L'addition de l'arginine dans le milieu de culture a inhibé (bloqué) la production des enzymes de la biosynthèse de l'arginine. Aucun effet observé sur la croissance et la concentration des protéines totales.

La transcription de l'opéron a lieu lorsque le répresseur ne peut pas se lier à l'opérateur. Après la fixation d'une petite molécule appelée corépresseur dans ce cas là c'est l'arginine sur le répresseur, ce dernier se fixe sur l'opérateur et bloque la transcription de l'opéron *argCBH* (fig. 15).

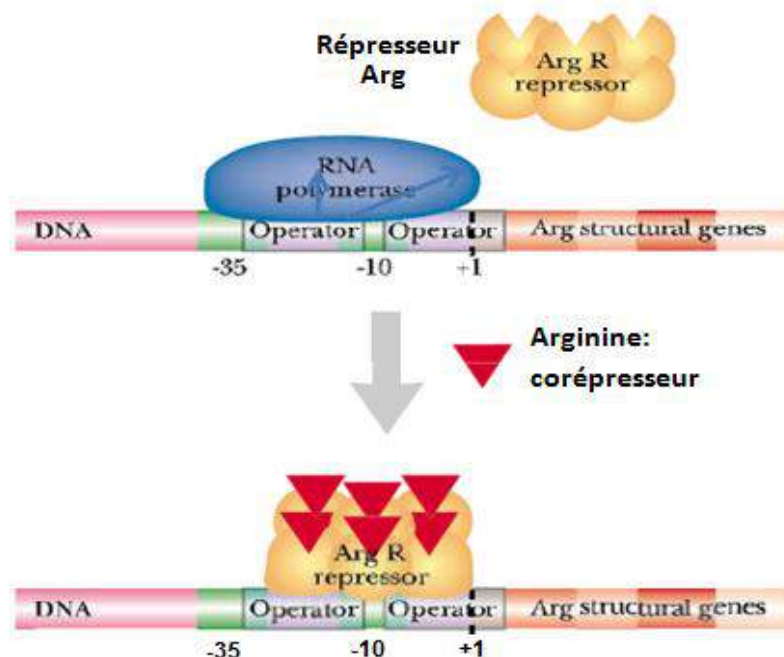


Fig.15: Mécanisme de contrôle de l'expression de l'opéron *argCBH*(Clark, 2005).

4-2- Contrôle positif de la transcription

Dans ce type de contrôle, une protéine régulatrice active la fixation de l'ARN polymérase d'où le terme positif. Un excellent exemple de ce type de contrôle est l'opéron *malEFG* (l'opéron du catabolisme du maltose) chez *E. coli*.

4-2-1- Régulation de l'expression de l'opéron *malEFG*

En absence d'inducteur, ni la protéine activatrice, ni l'ARN polymérase ne peuvent se lier à l'ADN. La liaison d'une molécule inductrice dans ce cas là c'est le maltose à la protéine activatrice provoque un changement de sa conformation (protéine allostérique), à son tour, se lie au site de fixation de l'activateur. Ceci permet à l'ARN polymérase de se fixer sur le promoteur et commencer la transcription (fig.16).

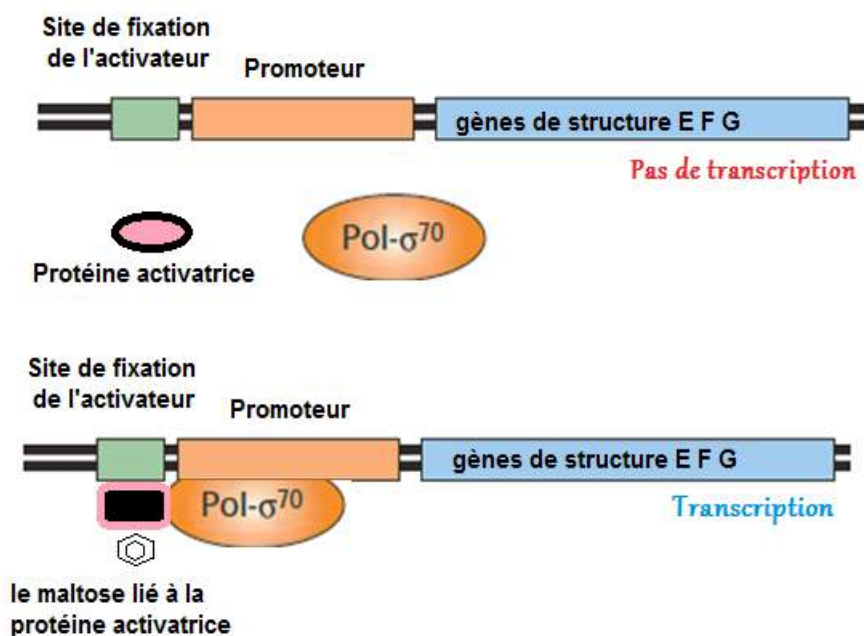


Fig.16: Mécanisme de contrôle positif de l'opéron *malEFG* chez *E. coli*(Lodish *et al.*,2003).

5- Régulon

Dans les exemples précédents (opéron d'arginine, opéron maltose), dans les figures on a montré un seul opéron mais en réalité le catabolisme du maltose et la synthèse de l'arginine nécessite plus qu'un opéron. La régulation de ces opérons se fait par un seul activateur (*malEFG*) ou un seul répresseur (*argCBH*). Alors on dit un régulon (ensemble d'opéron), lorsqu'une seule protéine régulatrice contrôle plus d'un opéron.

6-Régulation positive et négative de l'opéron *araBAD*

Le catabolisme de l'arabinose est contrôlée par une protéine de régulation AraC qui joue le rôle de répresseur en absence du sucre et d'activateur à sa présence (lié). La fixation de l'arabinose sur la protéine AraC provoque un changement de sa conformation, et se lie à l'ADN au site 1 et 2 ce qui active la fixation de l'ARN polymérase sur le promoteur et la transcription

des gènes codants les enzymes [isomérase(B), kinase (A), et épimérase (D)] du catabolisme de l'arabinose en ribulose puis rébulose 5-phosphate et ce dernier en Xylulose 5-phosphate. A l'absence de l'arabinose le répresseur AraC occupe les sites 2 et 3 et forme une structure secondaire (loop) en inactivant le gène (fig. 17).

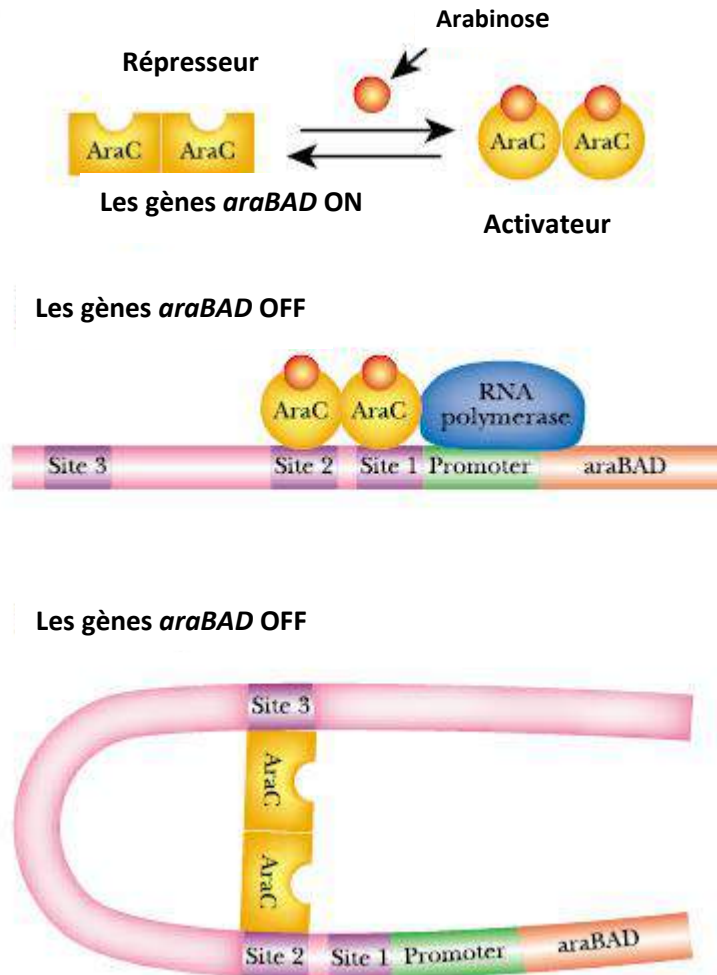


Fig.17: Mécanisme de contrôle du catabolisme de l'arabinose par la protéine régulatrice AraC (activateur, répresseur)(Clark, 2005).

7- Contrôle de l'expression génétique par atténuation

L'atténuation dépend du fait que la transcription et la traduction sont étroitement couplées. Chez *E. coli* les opérons de biosynthèse des acides aminés (Ex. Tryptophane) sont contrôlés par ce mécanisme. L'opéron Trp est composé de cinq gènes adjacents codant les enzymes qui synthétisent le tryptophane. Ces gènes sont contrôlés par un répresseur et ne s'expriment qu'à l'épuisement de tryptophane (l'absence du tryptophane qui arrête la répression).

Le phénomène d'atténuation contrôle la décision de transcrire un ARNm entier. Si les taux de tryptophane sont élevés, la plupart des transcrits terminent leur synthèse de façon

prématurée (incomplète). Au contraire, si les taux sont insuffisants, la polymérase transcrit les gènes en entier.

Une séquence leader de 161 nucléotides localisés en amont du premier codon du gène *trpE*, elle contient 4 régions (1, 2, 3, et 4) (fig. 18). Cette séquence contient aussi deux codons pour le tryptophane, alors s'il ya manque de tryptophane la synthèse du peptide leader sera incomplet à cause du manque des ARNt chargés par le Trp.

L'atténuation (l'arrêt de la transcription) dépend de la capacité de l'ARN à former des structures secondaires entre des régions complémentaires (2=3 ou 3=4) (fig. 19).

Si le tryptophane est abondant, le ribosome traduira la séquence leader jusqu'au codon stop, le reste de l'ARNm leader forme une structure secondaire entre les régions 3 et 4 suivi par une séquence riche en uracile qui provoque finalement la terminaison.

Si le tryptophane est absent ou en faible quantité, la synthèse du peptide leader sera incomplet. Le blocage du ribosome permet dans ce cas là la formation d'une autre structure secondaire (entre les régions 2 et 3), cette structure n'est pas un signal de terminaison de la transcription et empêche la formation du signal de terminaison. Ceci permet à l'ARN polymérase de transcrire tous les gènes de structure.

La régulation par atténuation est observée aussi pour au moins six opérons qui codent pour la biosynthèse des acides aminés par exemple: Thr, His, et le Phe. Les peptides leader des opérons de synthèses des ces acides aminés contiennent dans leur séquences plusieurs monomères de l'acide aminé concerné (8 Thr, 7 His, 7 Phe). L'ensemble des ces opérons ne présentent pas la même combinaison d'éléments de contrôle de la régulation que celle de l'opéron *trp*. (Exemple: pour l'opéron *his*, l'atténuation est le seul mécanisme de contrôle [pas de régulation Opér-Répr]).

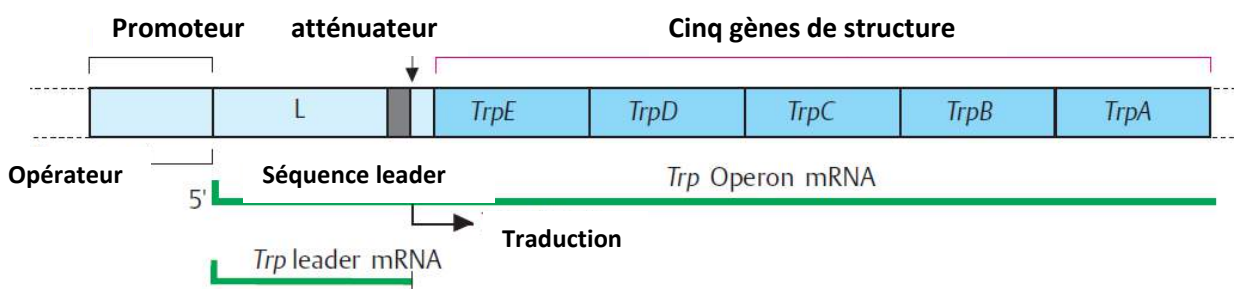


Fig.18: L'opéron tryptophane chez *E. coli*(Lodish *et al.*, 2003).

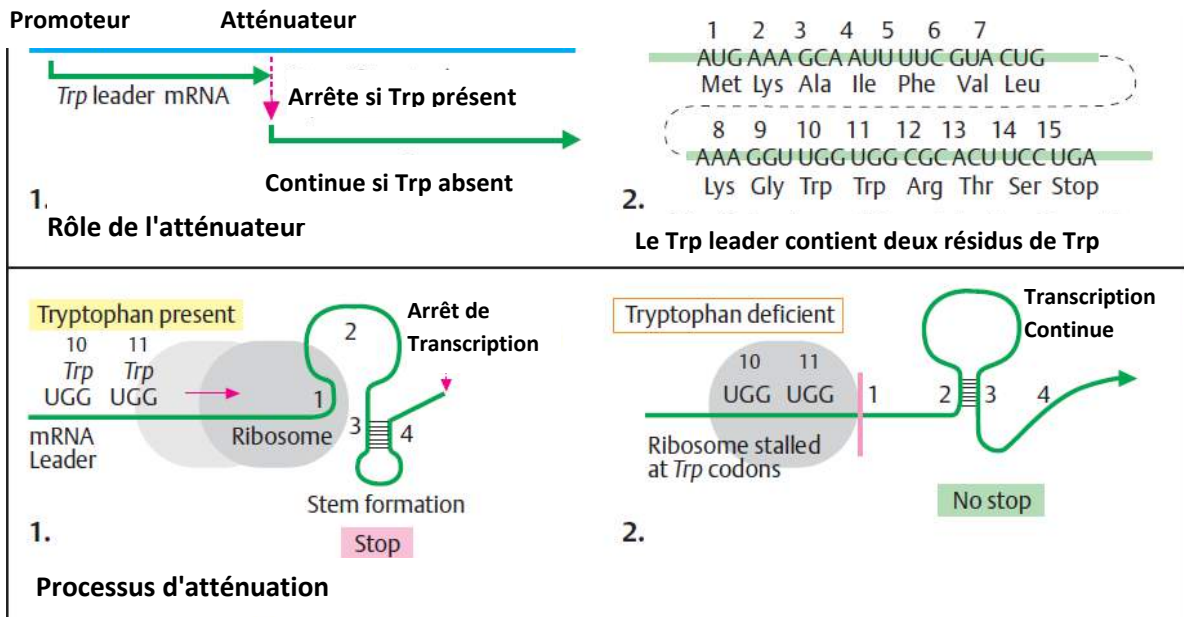


Fig.19: Processus de l'atténuation de l'opéron tryptophane d'*E. coli* (Passarge et al.,2007).

8- Régulation par des facteurs sigma (σ) alternatifs

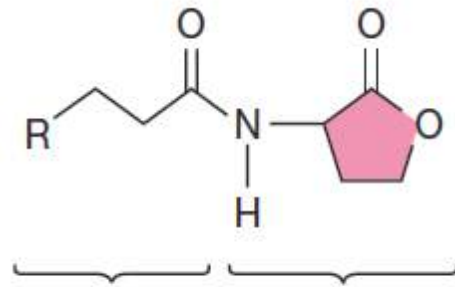
Comme on a vu dans le chapitre de la transcription, le facteur sigma (élément *trans*) est responsable de la reconnaissance des éléments *cis* (séquences consensus) des promoteurs lors de l'initiation de la transcription. Plusieurs bactéries produisent des facteurs sigma alternatifs.

La réponse au choc thermique est un exemple chez *E. coli* où l'expression des gènes est considérablement modifiée par l'intervention de différents facteurs sigma. La transcription des ces gènes met en jeu une ARN polymérase utilisant un facteur alternatif σ^{32} codé par le gène *rpoH*, c'est une protéine mineur moins abondante que le facteur σ^{70} . Le facteur σ^{32} possède des séquences consensus spécifiques sur les promoteurs (TNTCNCTTGAA). Si *E. coli* est soumise à une augmentation de température, la synthèse de 17 protéines du choc thermique est déclenché. La synthèse des ces protéines est transitoire entre 37°C et 42°C et sont les seules protéines synthétisées par la bactérie à des températures extrêmes (+ 50 °C).

Il existe d'autres exemples de ce type de régulation comme la réponse au choc froid, la différenciation des bactéries sporulées en réponse aux conditions défavorables, etc.

9- Détection du quorum (Quorum Sensing: QS)

La détection du quorum est un mode de signalisation très répandu chez les bactéries à Gram négative. Ce type de communication a été observé pour la première fois chez *Vibrio fischeri*, une bactérie bioluminescente. Il repose sur la production de petites molécules appelées "autoinducteurs". Lorsqu'un seuil de concentration est atteint, ces molécules interagissent avec un régulateur transcriptionnel, permettant l'expression spécifique d'un groupe de gènes. L'un des autoinducteurs le plus étudiés est le N-acylhomosérine lactone (AHL).



La chaîne Acyle Acyle homosérine lactone

Fig.20: Structure de l'autoinducteur N-acylhomosérine lactone (AHL).

Une cellule capable de pratiquer la détection de quorum transcrit et traduit les gènes codant l'enzyme: acylhomosérine lactone synthétase à des niveaux de base. Quand le nombre de cellules de la même espèce atteint un certain niveau, la concentration en AHL (fig. 21) augmente suffisamment pour se fixer à la protéine activatrice (inducteur). Cette protéine active la transcription des gènes codant les protéines spécifiques du quorum.

Parmi les exemples les mieux étudiés sur la détection de quorum est la formation du biofilm et la régulation de la production des facteurs de virulence chez *Pseudomonas aeruginosa*. Ce qui accroît la pathogénécité de ce microbe et empêche la pénétration des antibiotiques.

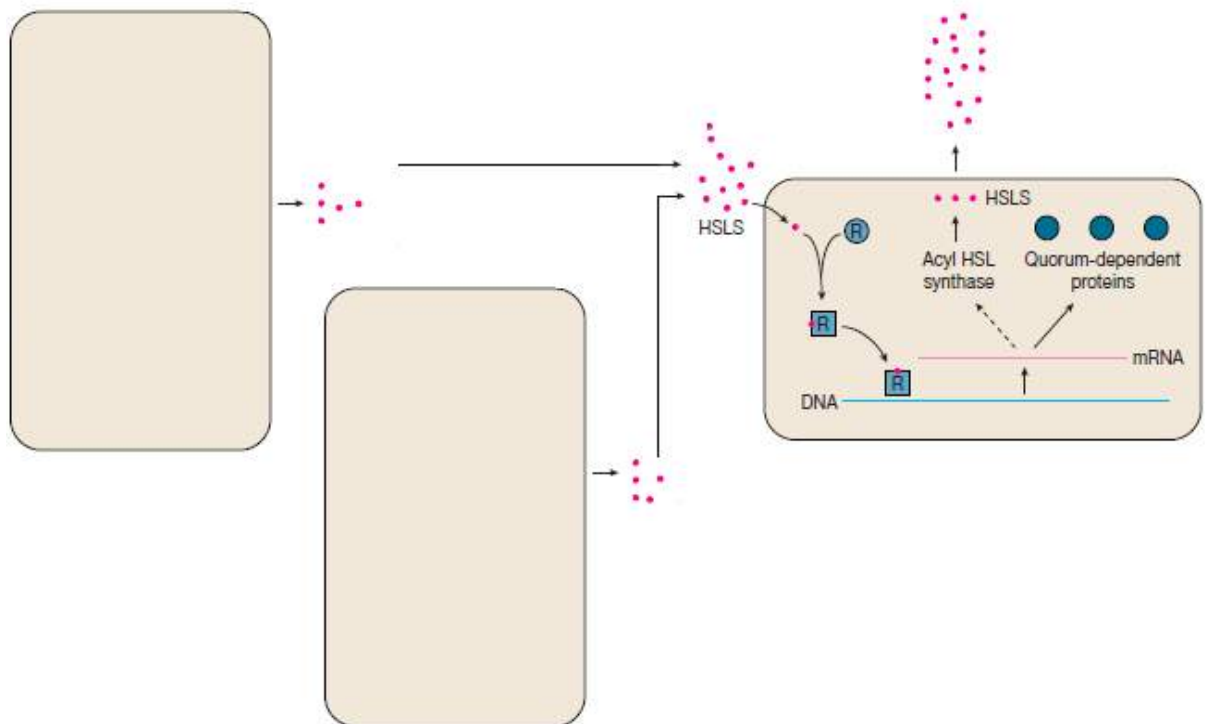


Fig.21: Détection de quorum (quorum sensing).Lorsqu'un seuil de concentration est atteint, ces molécules interagissent avec un régulateur transcriptionnel, permettant l'expression spécifique d'un groupe de gènes. R: protéine activatrice (inducteur) (Prescott, 2002).

10-Contrôle de l'expression génétique par un système de régulation à 2 composants

Les composants principaux du système comptent une kinase de détection dans la membrane cellulaire capable d'autophosphorylation en réponse à un signal environnemental. Le groupe phosphorylé est ensuite transféré à un autre composant principal du système, un régulateur de réponse. Ce dernier c'est un répresseur il se fixe alors sur l'opérateur en bloquant la transcription. Une phosphatase peut amorcer un nouvel cycle.

Exemples:

- L'utilisation NH_4^+ par *E. coli*.
- Régulation des porines (pression osmotique)
- La sporulation de *Bacillus sp.* (température élevée, épuisement de nutriments, agents chimiques toxiques etc.).

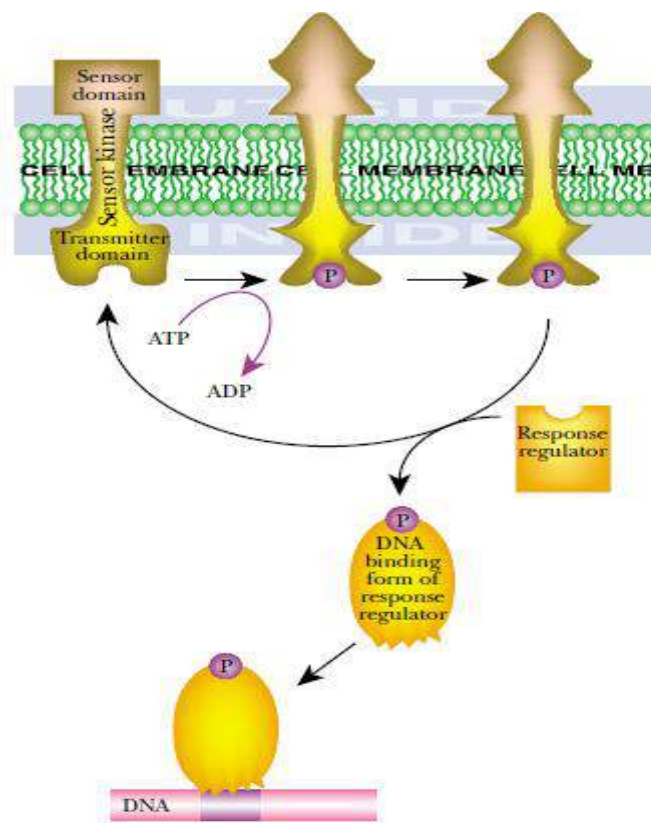


Fig.22: Contrôle de l'expression génétique par un système de régulation à 2 composants (kinase, répresseur) (Clark, 2005).

11- ARN de régulation et les riboswitch

Dans les exemples précédents sur la régulation de l'expression des gènes, les protéines régulatrices jouent le rôle clés, alors que dans cette partie c'est les ARN. Ce type de régulation est connu chez les cellules eucaryotes et procaryotes.

Des expériences sur *E. coli* montrent qu'il ya des petites ARN (pARN: 40 à 400 nucléotides) participent à la régulation des fonctions physiologiques de cette bactérie en se

fixant à d'autres ARN et même à des petites molécules dans certains cas. Ces ARN de régulation ne codent ni protéines, ni ARNr, ni ARNt. Il ya trois mécanismes d'action de ces molécules:

- Les pARN rentrent dans la formation d'une ribonucléo-protéine (particule de reconnaissance de signal) pour l'identification des protéines neosynthétisées qui doivent être excrétées.
- Elles s'apparient à l'ARNm (exemple au niveau de la séquence S/D), en bloquant sa traduction et accélérant sa dégradation par des ribonucléases spécifiques. Dans ce cas-là sont appelés les ARN antisens.

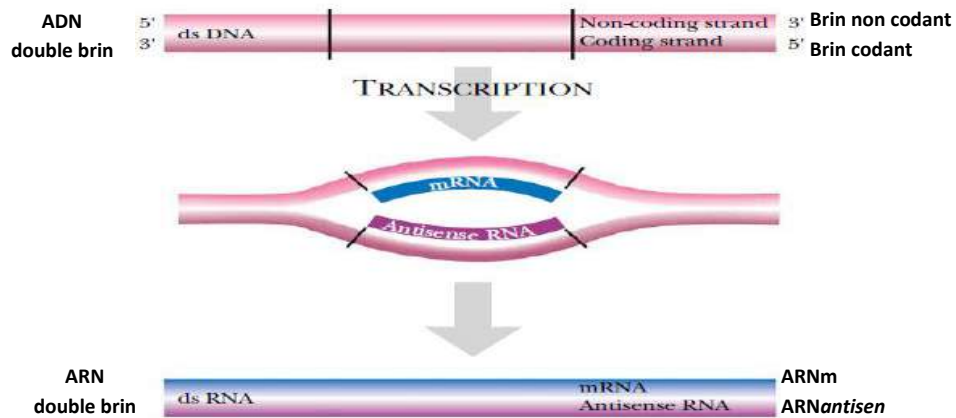


Fig.23: Formation d'hybride ARNm/ARNantisens.

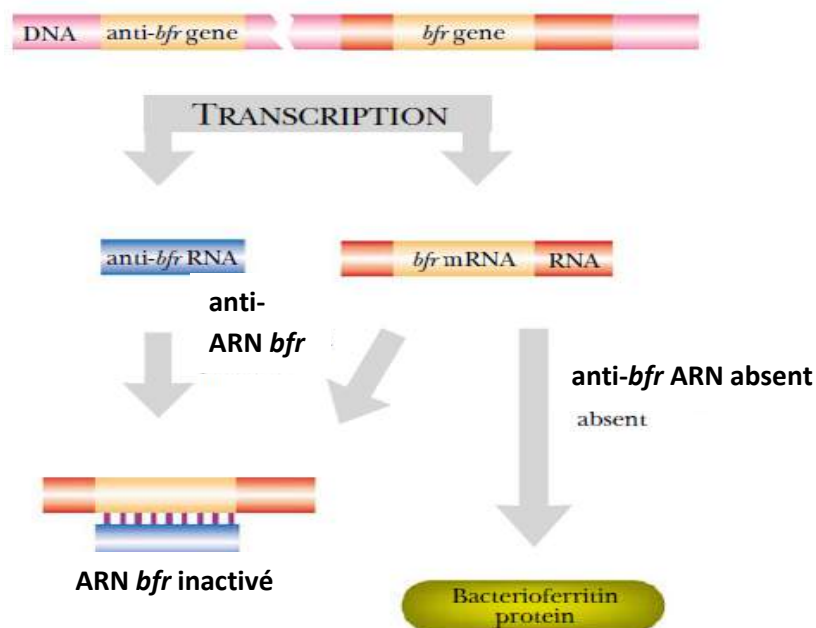


Fig.24: Régulation de la synthèse du bactérioferritine par l'ARN antisens, le chromosome bactérien contient les gènes pour les deux *bfr* ARNm et *anti-bfr* ARN. En présence d'une quantité suffisante de fer, les deux molécules d'ARN sont transcrites et l'ARNm de *bfr* ne sera pas traduit à cause de la formation d'hybride entre *bfr* ARNm et *anti-bfr* ARN. S'il ya carence en fer le gène *anti-bfr* ne sera pas transcrit et la bactérioferritine est produit par la traduction du *bfr* ARNm (Clark, 2005).

11-1- Riboswitchs

Les riboswitchs sont trouvés que chez quelques bactéries, quelques plantes et quelques champignons. Ils sont une forme particulière de pARN, ce sont des ARNm qui contiennent des séquences en amont de la séquence codante et peuvent se lier à des petites molécules. Exemples de biosynthèses contrôlées par ce mécanisme: la synthèse des vitamines: thiamine, riboflavine, B12, quelques acides aminés et les bases puriques : adénine et guanine.

La molécule synthétisée par la protéine codée par l'ARNm du riboswitch, se lit à des endroits spécifiques du riboswitch et forme une structure secondaire. Cette dernière peut bloquer la transcription ou la traduction (fig. 25).

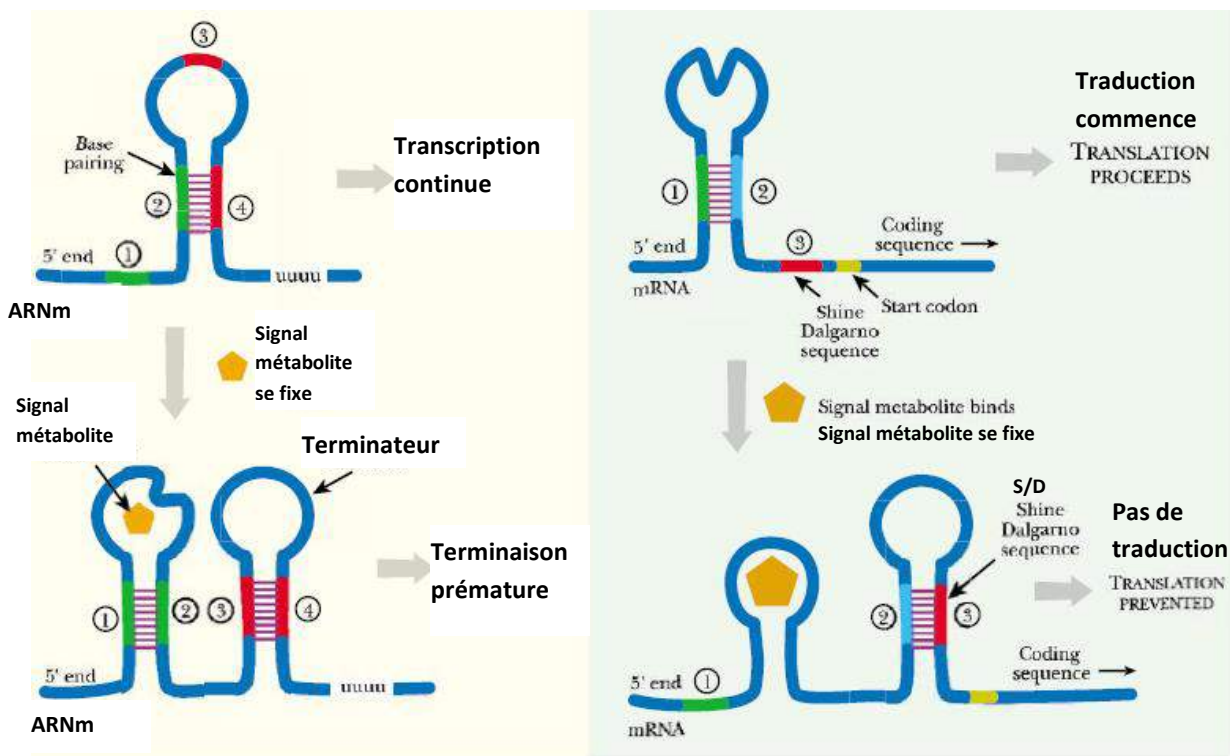


Fig.25: Régulation de l'expression génétique par le mécanisme des riboswitchs (Clark, 2005).

Deuxième partie:

Génie Génétique

Génie Génétique

Glossaire

Adaptateur: Courte séquence nucléotidique capable de réaliser le pontage entre deux fragments d'ADN terminés par des séquences non complémentaires.

ADN: Acide désoxyribonucléique. C'est le support du matériel de l'hérédité à l'exception de certains virus à ARN.

ADNc ou complémentaire: C'est une copie d'ARN obtenu par transcription reverse.

Cellule hôte: cellule hébergeant un matériel génétique étranger apporté par un vecteur.

Clonage: Manipulation génétique permet d'isoler des fragments d'acides nucléiques de leur organismes d'origine et de les propager dans un organisme hôte qui peut être le même que l'organisme original ou un organisme différent. La capacité de cloner l'ADN ouvre la voie à de nombreuses applications (isolement, amplification, séquençage, expression, etc.).

Clonage moléculaire: Isolement et incorporation d'un fragment d'ADN dans un vecteur pouvant le répliquer.

Criblage: Opération d'identification et de tri de clones recombinants.

Génie génétique:

- Ensemble de techniques permettant de modifier le patrimoine héréditaire d'une cellule par la manipulation de gène *in vitro*
- La capacité à manipuler de l'ADN et à l'insérer dans un organisme qui l'exprimera.
- L'utilisation de techniques *in vitro* pour l'isolement, la manipulation, la recombinaison, et l'expression des gènes, ainsi le développement d'OGM.
- La formation de nouvelles combinaisons de matériel héréditaire par l'insertion de molécules d'acides nucléiques dans un vecteur (un virus, un plasmide bactérien, ou tout autre vecteur) pour les introduire dans un organisme hôte dans lequel elles ne sont pas présentes naturellement, mais où elles deviennent capables de propagation stable.

Ligaturer : Action de former une liaison phosphodiester entre deux polynucléotides.

Mutagenèse dirigé: Introduction d'une mutation précise.

Séquence palindromique: Séquence d'ADN pouvant se lire de la même façon dans les deux sens par rapport à un point central soit sur le même brin ou sur les deux brins.

Vecteur de clonage: Élément génétique capable de se répliquer et permettant le transfert dans une cellule réceptrice d'un fragment d'ADN étranger préalablement inséré.

Préambule

Le génie génétique est principalement basé sur le clonage moléculaire qui permet la recombinaison d'un fragment double brin d'ADN avec un vecteur et son introduction dans un hôte approprié.

Cette technologie a d'abord concerné les organismes les plus simples qui ont souvent des génomes de petite taille comme les bactéries (Ex. *E. coli*, *B. subtilis*, *Agrobacterium tumefaciens*...etc.), et les levures (*S. cerevisiae*). Ce n'est que plus tard que la transgénèse stable des cellules végétales et animales.

Actuellement les OGM (végétaux, animaux transgéniques) sont nombreux et certains commencent à parvenir dans nos assiettes.

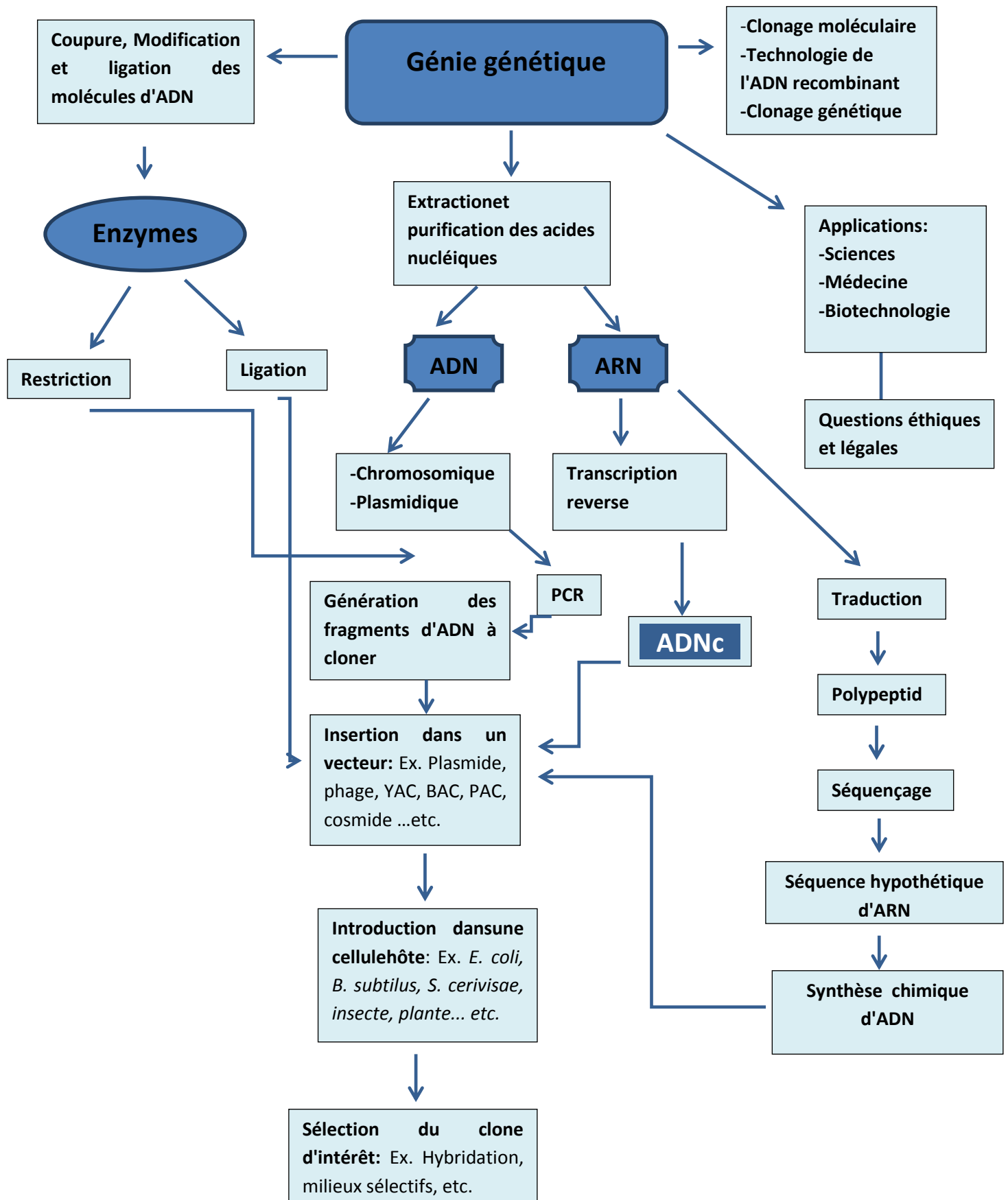
Le premier apport de cette technologie à la médecine fut de rendre des protéines thérapeutiques telle: insuline, HGH (human growth hormone), disponibles en quantités suffisantes. La médecine a retiré bien d'autres bénéfices de la technologie de l'ADN recombinant, vaccins (anti hépatite, rage, ...etc.), sondes (diagnostique, identification ...etc.), agents thérapeutiques (acides nucléiques anti sens, thérapie génique, ...etc.).

La technologie de l'ADN recombinant a touché d'autres produits d'intérêt agro-alimentaire et industriel, enzymes (présure, protéases, ...etc.), acides aminés (glutamate, succinate, ...etc.), vitamines (B2, B12, C ...etc.).

Dans les deux dernières décennies plusieurs milliers d'entreprises de biotechnologie ont été créés dans le monde surtout en USA et le Japon.

Les différentes étapes successives à entreprendre en génie génétique sont les suivants:

- 1- Isolement d'une séquence d'acide nucléique, idéalement un gène.
- 2- L'insertion de ce gène dans un vecteur (Ex. plasmide ou phage).
- 3- L'incorporation de ce vecteur dans un hôte (Ex. *E. coli*).
- 4- L'hôte transformé peut assurer la propagation du vecteur à chaque cycle de réplication.



Carte de conception de génie génétique.

Génie génétique

Chapitre I

Enzymes de restriction et de modification de l'ADN

1- Enzymes de restriction

Les enzymes de restriction sont des endonucléases spécifiques, leur utilisation permet d'obtenir des fragments d'ADN aux extrémités bien caractérisées, dont certains peuvent contenir des gènes.

Il existe trois types d'enzymes de restriction isolées des bactéries.

- 1- **Type I** : Dont l'action nécessite la présence de Mg^{++} , d'ATP comme cofacteur, et de S-adénosyle-méthionine. Leur site de coupure est éloigné de leur site de reconnaissance (jusqu'à plusieurs milliers de nucléotides, plus loin dans certain cas).
- 2- **Type II** : Pratiquement les seules utilisées en génie génétique, reconnaissent l'ADN à des sites particuliers et coupe coupes dans ces sites ou à proximité immédiate d'eux. Ce sont des nucléases qui coupent à l'intérieur, donc ce sont des endonucléases (Type II restriction endonucléases).
- 3- **Type III** : Ressemblent à celles du type I pour la séparation des sites de reconnaissance et de coupure, mais qui s'apparentent à celles du type II par leur mode d'action.

Tab.1: Nombre des sites de coupure de quelques enzymes de restriction de type II sur le génome des virus SV40, λ , et l'adénovirus 2.

| Origine microbienne | Enzyme | Site de coupure | Nombre de sites de reconnaissance par génome | | |
|-------------------------------------|---------------|-----------------------|--|-----------|--------------|
| | | | SV40 | λ | Adénovirus 2 |
| <i>E. coli</i> KY13 | <i>EcoRI</i> | G [↓] AA*TTC | 1 | 5 | 5 |
| <i>Haemophilus influenzae</i> Rd | <i>HindII</i> | *A [↓] AGCTT | 6 | 6 | 11 |
| <i>Haemophilus parainfluenza</i> | <i>HpaII</i> | C [↓] C*GG | 1 | >50 | >50 |
| <i>Bacillus amyloliquefaciens</i> H | <i>BamHI</i> | G [↓] GA*TCC | ? | ? | ? |
| <i>Haemophilus aegypticus</i> | <i>HaeIII</i> | GG [↓] CC | 18 | >50 | >50 |

* indique la méthylation.

1-1- Nomenclature des enzymes de restriction

Les trois premières lettres de chacune d'entre elle concernent

- La première lettre du nom de genre, par exemple E (*Escherichia*)
- Les deux premières lettres du nom de l'espèce, par exemple co (*coli*).
- La quatrième concerne la souche bactérienne d'où est extraite l'enzyme en question.
- Le chiffre romain: indique l'ordre de la caractérisation de l'enzyme chez la même souche. *HindIII* signifie que c'est la troisième (III) Enzyme de restriction isolée et caractérisée de la souche bactérienne *Haemophilus influenzae* Rd.

Tab. 2: Exemples des enzymes de restriction de type II.

| Enzyme | L'organisme source | Séquence de reconnaissance |
|---------------|-----------------------------------|--------------------------------|
| <i>HpaII</i> | <i>Haemophilus parainfluenzae</i> | C/CGG GGC/C |
| <i>MboI</i> | <i>Moraxella bovis</i> | /GATC CTAG/ |
| <i>NdeI</i> | <i>Neisseria denitrificans</i> | /GATC CTAG/ |
| <i>EcoRI</i> | <i>Escherichia coli</i> RY13 | G/AATTC CTTAA/G |
| <i>EcoRII</i> | <i>Escherichia coli</i> RY13 | /CC(A or T)GG GG(T or A)CC/ |
| <i>EcoRV</i> | <i>Escherichia coli</i> J62/pGL74 | GAT/ATC CTA/TAG |
| <i>BamHI</i> | <i>Bacillus amyloliquefaciens</i> | G/GATCC CCTAG/G |
| <i>SauI</i> | <i>Staphylococcus aureus</i> | CC/TNAGG GGANT/CC |
| <i>BglI</i> | <i>Bacillus globigii</i> | GCCNNNN/NGGC CGGN/NNNNCCG |
| <i>NotI</i> | <i>Nocardia otitidis-caviarum</i> | GC/GGCCGC CGCCGG/CG |
| <i>DraI</i> | <i>Deinococcus radiophilus</i> | RG/GNCCY YCCNG/GR |

N: n'importe quelle base **R:** n'importe quelle purine **Y:** n'importe quelle pyrimidine /: Position de coupure.

La plupart des enzymes de restriction de type II ont un site de reconnaissance de 4 ou 6 paires de bases, et un certain nombre d'enzymes de restriction reconnaissent des séquences à 5 ou à 7, 8 ... paires de bases.

Les enzymes à site de reconnaissance tétra-nucléotidique ou hexa-nucléotidique présentent en générale un axe de symétrie rotatoire palindromique (fig. 1), au centre de leur séquence.

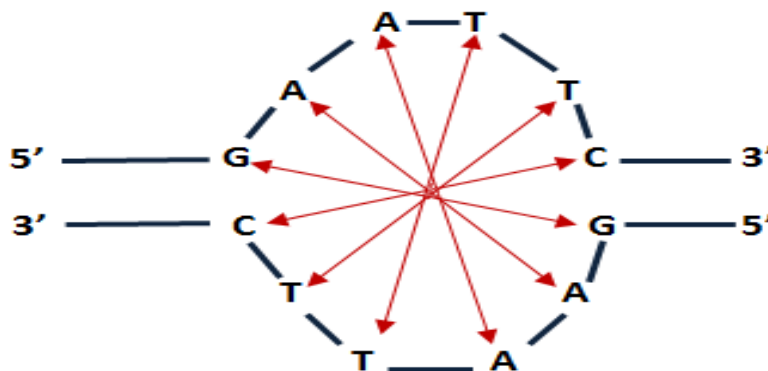


Fig.1. Axe de symétrie rotatoire du site de coupure *EcoRI*.

Le site de coupure, compris dans le site de reconnaissance est représenté par une flèche verticale (tab.1), et les produits de la digestion d'un génome par les enzymes de restriction sont appelés les fragments de restriction. Selon la position de la coupure (dans l'axe ou loin de l'axe) on distingue deux types de coupure:

- Coupure **franche** : Dans l'axe \longrightarrow Extrémités franches
- Coupure **cohésive** (collant) : N'est pas dans l'axe. \longrightarrow Extrémités cohésives

Dès 1953, on avait observé que lorsque des molécules d'ADN provenant d'une souche d'*E.coli* sont introduites dans une autre souche, elles ne sont que très rarement exprimées. En fait, l'ADN étranger est la plupart du temps rapidement coupé en petits fragments par une enzyme dite de restriction. Pour qu'il puisse se répliquer dans une nouvelle souche bactérienne, l'ADN introduit doit être protégé par une enzyme de modification qui ajoute des groupements méthyles à l'ADN.

En accord avec ces expériences sur *E. coli*, Hamilton Smith à l'Université Johns Hopkins s'aperçoit fortuitement que la bactérie *Haemophilus influenzae* dégrade rapidement un ADN étranger alors que son propre ADN reste intact. Cette découverte l'a amené à purifier *HindII*, la première enzyme de restriction coupant à un site spécifique et responsable de la dégradation de l'ADN étranger.

Donc ces enzymes participent à un mécanisme de défense des bactéries vis-à-vis des virus "système de restriction méthylation" (fig. 2). Les endonucléases de restriction coupent l'ADN étranger à des sites spécifiques. Et afin de protéger l'ADN bactérien de l'hydrolyse par les enzymes, une méthylase, codé par le gène de méthylation, va méthyler l'ADN au niveau des sites de coupure pour qu'ils ne soient plus reconnus par l'enzyme de restriction.

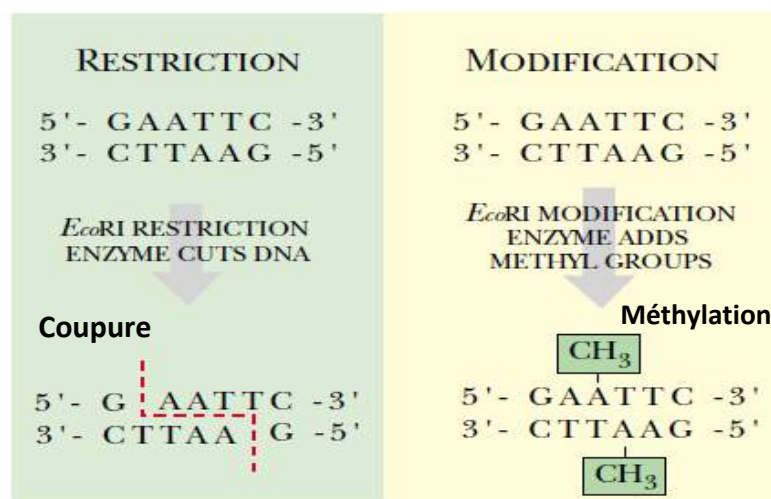


Fig.2: Systèmes de restriction et de méthylation.

1-2- Classification des enzymes de restriction

Les enzymes sont souvent des protéines présentant des propriétés de catalyse spécifique d'une réaction chimique. Elles sont réparties en 6 classes :

E.C. 1 Oxydoréductases: catalysent les réactions redox.

E.C. 2 Transférases: transfèrent d'un groupement fonctionnel.

E.C. 3 Hydrolases: catalysent l'hydrolyse de différentes liaisons.

E.C. 4 Lyases: Coupe des liaisons variées sans réaction d'hydrolyse ou de redox

E.C. 5 Isomérases: isomérisation au niveau de la même molécule

E.C. 6 Ligases: joinsdeux molécules par une liaison covalente.

1-3-Enzymes de restriction utilisées en génie génétique

Les enzymes de restriction de type II sont pratiquement les plus utilisées en génie génétique, à cause de leurs propriétés de coupure et de reconnaissance. Le E.C. de ces enzymes est le suivant: **E.C.3.1.21.4.**

Nom Commun: Les endonucléases de restriction de type II

Réaction: Coupure à l'intérieur de la séquence d'ADN db au niveau des sites de reconnaissance et de coupure spécifiques, générant des fragments d'ADN double brin se terminant par 5'-phosphate.


3 (classe): Hydrolase

1 (sous classe): Agissent sur les liaisons esters (**Estérases**)

21 (sous-sous classe): Leurs produits gardent leurs phosphates **5'**initial

4 : Nécessite la présence de **Mg⁺⁺** dans le milieu.

Réaction générale

ADN (Substrat) ADN digéré 
(H₂O, énergie, Enzyme, Mg⁺⁺)

La réaction générale catalysée par les enzymes de restriction (E.C.3.1.21.n) implique la présence dans le milieu des facteurs suivants.

- **Enzyme** (E.C.3.1.21.n)
- **Substrat** ADN double brin non digéré
- **Cofacteur** : En général, seul le **Mg⁺⁺** est indispensable. Quelques enzymes de restriction font appel à d'autres cofacteurs. Les enzymes de méthylation ont pour coenzymes le S-adénosyl-Méthionine.
- **Facteurs physico-chimiques:** Ces facteurs contrôlent aussi ces réactions, **pH**, potentiel d'oxydoréduction, **forces ioniques**. L'énergie libre dégagée par l'hydrolyse est suffisamment importante pour que la réaction ne soit pratiquement pas réversible dans les conditions habituelles.
- Différents tampons sont utilisés pour les enzymes de restriction permettant l'optimisation des réactions de multiples enzymes avec le même tampon (pH:7- 7.9, PotOxyRed: dithiothreitol 1mM, Force ionique : NaCl ou CH₃COOK de 0 à 100mM ; varient d'un tampon à l'autre).

Exemples d'enzymes de restriction de type II

1- **EcoRI** : E.C. 3.1.21.4: C'est la deuxième enzyme caractérisée et identifiée chez *E. coli* souche R, le site de reconnaissance et de coupure est formé de six paires de bases (nucléotides).

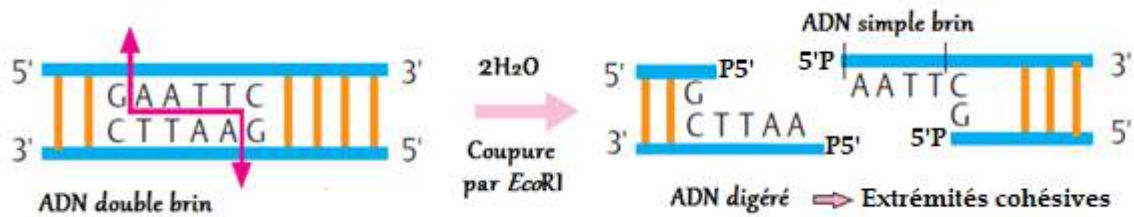


Fig.3: Coupure de l'ADN double brin par *EcoRI*, générant des extrémités cohésives.

2- **HaeIII**: C'est la troisième enzyme de restriction caractérisée et identifiée chez la bactérie *Haemophilus aegypticus*.

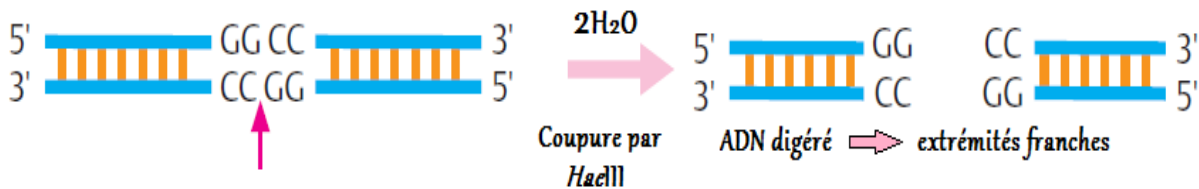


Fig.4: Coupure de l'ADN double brin par *HaeIII*, générant des extrémités franches.

1-4 Utilisation des endonucléases de restriction pour établir la mappe de restriction

La figure au dessous montre la digestion d'un fragment d'ADN de 5 kb par deux enzymes de restriction (*BamHI* et *EcoRI*).

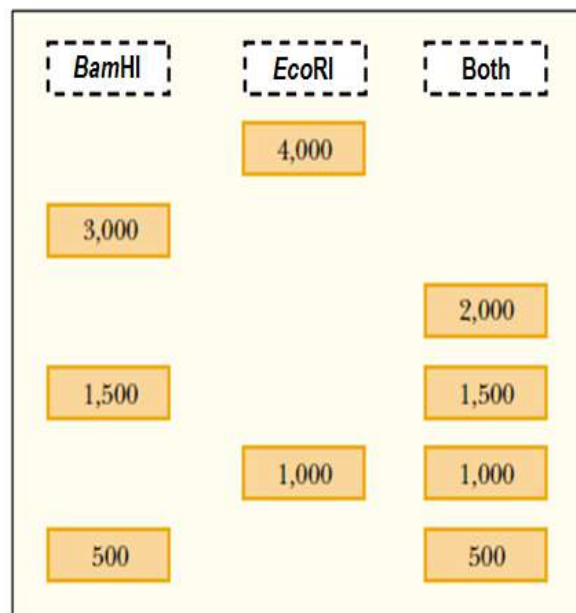


Fig.5: Gel d'agarose des produits de digestion d'un fragment de 5kb par deux enzymes de restriction.

L'analyse des résultats de la digestion (fig. 5) nous permet de déterminer les différentes possibilités de la localisation des sites de restriction pour chaque enzyme (fig. 6 et 7).

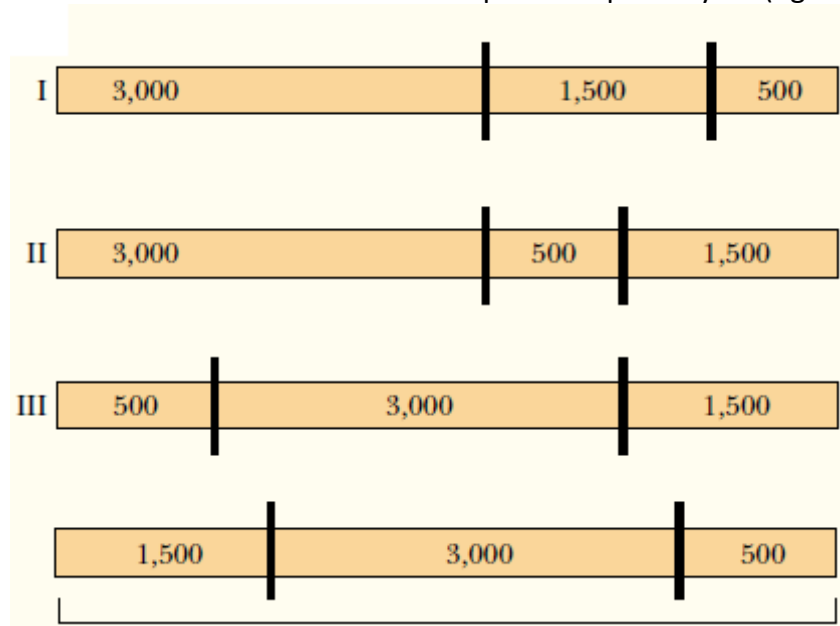


Fig.6: Les différentes possibilités de la localisation des sites de coupure par *Bam*HI.

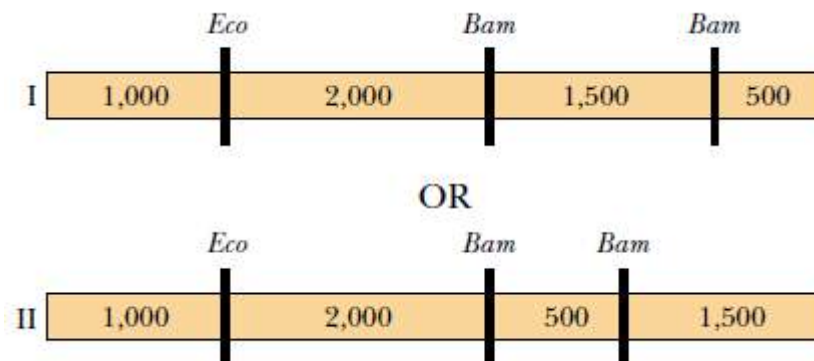


Fig. 7: Les différentes possibilités de localisation des sites de restriction des deux enzymes (*Eco*RI, et *Bam*HI).

1-5- Isoschizomères et famille d'enzymes:

En général, des enzymes de restriction différentes reconnaissent des séquences différentes. Cependant, plusieurs enzymes isolées d'organismes différents reconnaissent des séquences identiques. On les appelle isoschizomères. Ceux-ci ne coupent pas toujours la séquence reconnue au même endroit.

| | | |
|------------|----------------|--------------------------------|
| Exemples : | <i>Asp</i> 718 | 5' g gatcc 3' 3' cctag g 5' |
| | <i>Kpn</i> I | 5' ggatc c 3' 3' c ctagg 5' |
| | <i>Bam</i> HI | 5' g gatcc 3' 3' cctag g 5' |

Une famille d'enzymes est caractérisée par le fait que tous les membres de cette famille produisent les mêmes extrémités, même si elles ont des séquences de reconnaissance différentes.

Exemples:

- **gatc** : *Sau* 3AI, *Bgl* I, *Bam* HI, *Bcl* II, *Xho* II.
- **ctag** : *Mae* I, *Spe* I, *Nhe* I, *Avr* I, *Xba* I.



1-6-Méthylation de l'ADN

La plupart des souches de *E. coli* contiennent deux enzymes qui méthylent l'ADN : la Dam méthylase et la Dcm méthylase.

1-6-1-Dam méthylase

La Dam méthylase méthyle l'adénine dans la séquence 5'-GA*TC-3'. Les sites de reconnaissance de plusieurs enzymes de restriction telles que *PvuI*, *Bam*HI, *Bcl*I, *Bgl*II, *Xho*II, *Mbo*I et *Sau*3AI contiennent cette séquence de même qu'une certaine proportion des sites reconnus par *Cl*I (1 site sur 4), *Taq*I (1 site sur 16), *Mbo*II (1 site sur 16) et *Hph*I (1 site sur 16). Certaines enzymes ne coupent pas si l'adénine est méthylée. L'enzyme *Mbo*I ne coupe pas la séquence GATC si elle est méthylée tandis que l'enzyme *Sau*3AI reconnaît la même séquence que *Mbo*I et n'est pas affectée par la méthylation *dam*.

1-6-2-Dcm méthylase

Cette méthylase ajoute un groupement méthyle à la cytosine interne des séquences 5'-CC*AGG-3' et 5'-CC*TGG-3'. Une enzyme affectée par la *dcm* méthylation est *Eco*RII. On peut contourner le problème en utilisant *Bst*NI qui reconnaît la même séquence qu'*Eco*RII, mais qui ne la coupe pas au même endroit. Si *Bst*NI ne convient pas, on peut préparer l'ADN à partir de souches d'*E. coli* qui sont *dcm*-.

E. coli K contient au moins 3 systèmes de restriction différents dépendant de la méthylation qui reconnaissent leurs séquences cibles seulement lorsqu'elles sont méthylées : *mrr* (6-méthyladénine), *mcrA* (m5CG où m5C est la 5-méthylcytosine) et *mcrB* (Pum5C). Des fragments d'ADN méthylés à un de ces sites sont clonés inefficacement dans des souches sauvages de *E. coli*. Par exemple, comme les CG de l'ADN de mammifère sont largement méthylés *in vivo*, l'ADN humain est restreint par *mcrA*.

2- Enzymes de modification des acides nucléiques

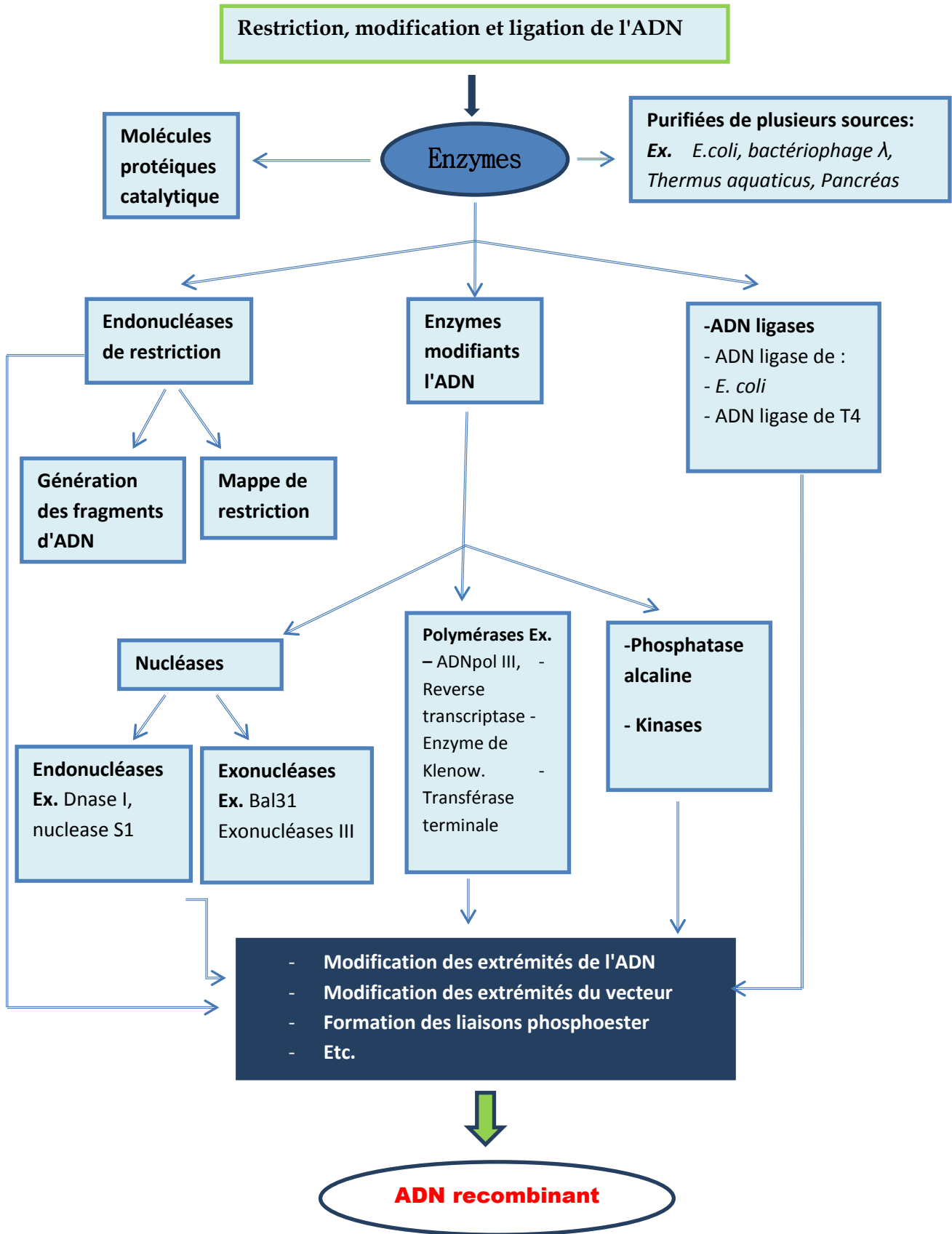


Fig.8: Carte de conception des enzymes utilisées en génie génétique

2-1- Ribonucléases A (RNase): EC3.1.27.5

Cette hydrolase agit comme une endonucléase, préférentiellement après les nucléotides à pyrimidines (U, C). En hydrolysant la liaison entre le phosphate et le carbone 5' du nucléotide suivant. Elle hydrolyse les ARN simple brin (inactive sur l'ARN double brin) jusqu'à un mélange d'oligonucléotides se terminant tous par un nucléotide à pyrimidine estérifié par un phosphate en 3'.

Il existe des inhibiteurs de l'RNase, Ex. Le SDS (Sodium Dodecyl Sulphate), des protéines extraites du placenta (ces inhibiteurs sont utilisés souvent pour protéger les ARN lors de l'extraction ou les réactions enzymatiques).

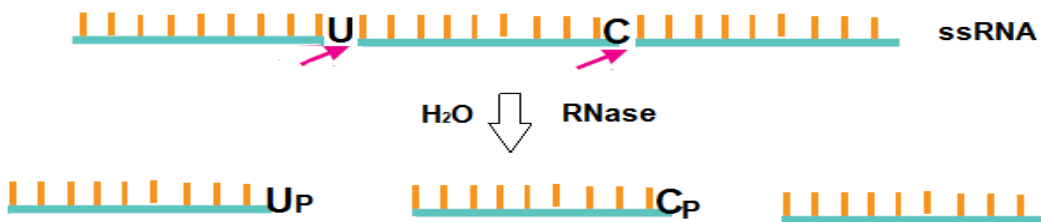


Fig.9: Digestion de l'ARN simple brin par l'RNase.

2-2- Désoxyribonucléase I (DNase I): EC 3.1.21.1

La DNase agit sur les liaisons adjacentes aux nucléosides pyrimidiques (C, T) des ADN simple ou double brin. L'activité de cette enzyme est dépendante de la présence du Mg^{++} ou Mn^{++}

- Mg^{++} : Hydrolyse au hasard
- Mn^{++} : Hydrolyse dépendante de la séquence

Elle est utilisée pour l'analyse des sites de liaison protéines –ADN par la technique de foot-printing (empreinte de pas). Car la protéine liée à l'ADN le protège de l'hydrolyse par cette enzyme (fig. 10). Cette enzyme est extraite du pancréas de bœuf.

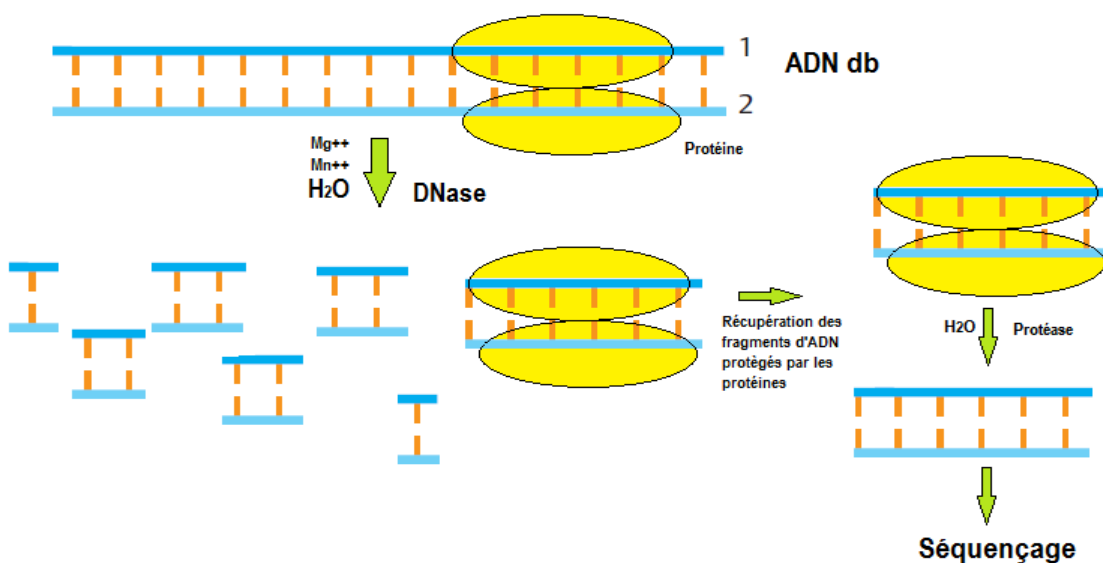


Fig.10: Détection des sites de liaison des protéines sur l'ADN par la technique de "foot printing".

2-3- Nucléase S1 : EC 3.1.30.1

La nucléase S1 est une métalloprotéine (PM: 32000 Da) produite par le champignon *Aspergillus oryzae*, c'est une nucléase spécifique des ADN (ou ARN) simple brin, bien qu'à des concentrations élevées elle agit sur les hybrides. Elle agit en milieu acide (pH 4-4.3) en présence d'ion Zinc (Zn^{++}). L'activité de cette enzyme diminue avec l'augmentation du pH (diminution de 50% à pH 4.9), et aussi avec la diminution de la concentration du phosphate libre. Cette métallo-enzyme est inhibée par les agents chélateurs (Ex. EDTA: Ethylene Diamine Tetra-acetic Acid). Elle est utilisée en plusieurs applications:

Pour faire des bouts francs aux extrémités des fragments d'ADN double brin.

- Pour hydrolyser les fragments d'ADN simple brin au niveau des brèches (même qu'il manque une seule liaison).
- Pour ouvrir les épingle-à-cheveux formés lors de la synthèse de l'ADNc.
- Pour isoler des hybrides ADN/ADN, ADN/ARN, ou ARN/ARN. C'est un outil de choix pour mesurer le taux d'hybridation.

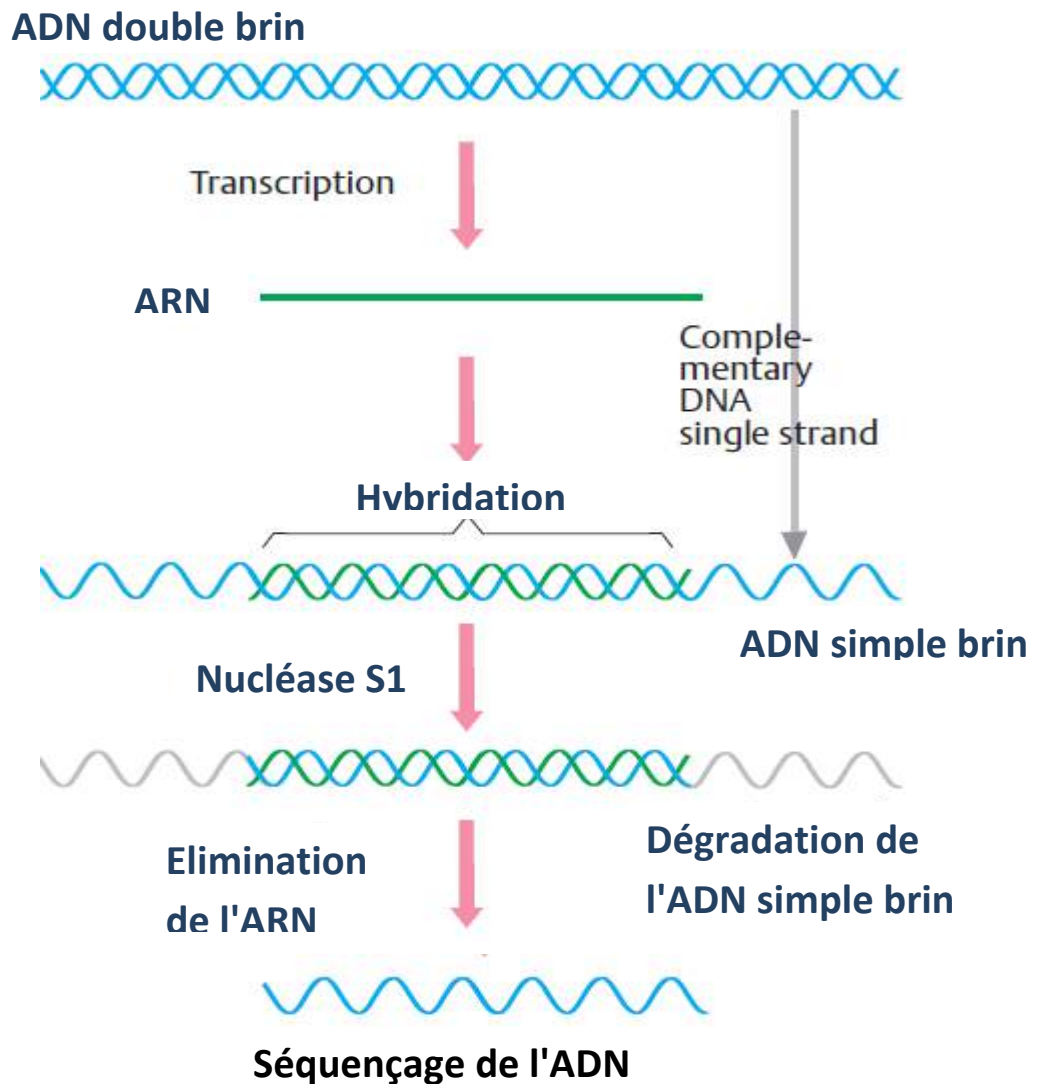


Fig.11: L'utilisation du nucléase S1 pour la détermination du site d'initiation de la transcription (Passarge, 2007).

2-4- Exonucléase de phage λ : EC 3.1.11.3

L'exonucléase de phage λ , elle est aussi rencontrée chez le T4 ou T7, ainsi que chez les animaux (DNase IV) hydrolyse de préférence les extrémités 5'-phosphate des DNA double brin en continuant vers le coté 3' (5' \rightarrow 3' exonucléase). Elle produit des nucléosides 5'-phosphate (fig. 12). Cette enzyme est produite par les cellules bactériennes infectées par les phages.

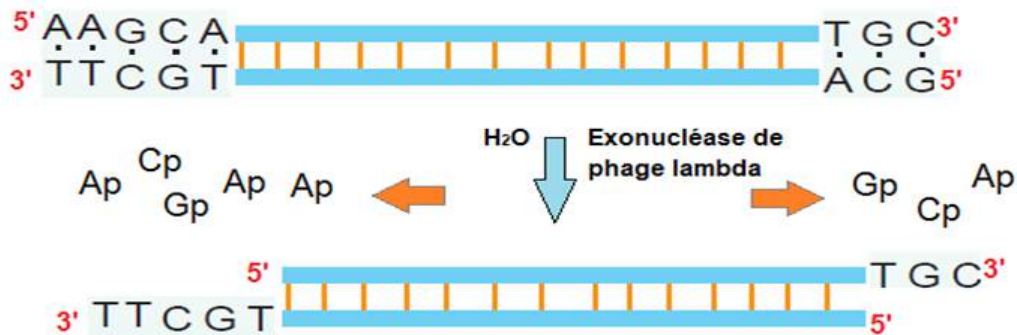


Fig. 12: Hydrolyse des extrémités 5' de l'ADN db par l'exonucléase de phage λ .

I-2-5 Exonucléase III: EC 3.1.11.2

Cette enzyme produite par *E. coli*, et aussi par *Haemophilus influenza*, au contraire de l'exonucléase de phage λ , hydrolyse de préférence les extrémités 3' des ADN doubles brin en remontant vers le coté 5' (3' \rightarrow 5' exonucléase). Elle produit des nucléotides 5'-phosphate (fig.13). Cette enzyme peut hydrolyser aussi un radical phosphorylé lié à la fonction 3'-OH du dernier nucléotide (action phospho-mono-estérase).

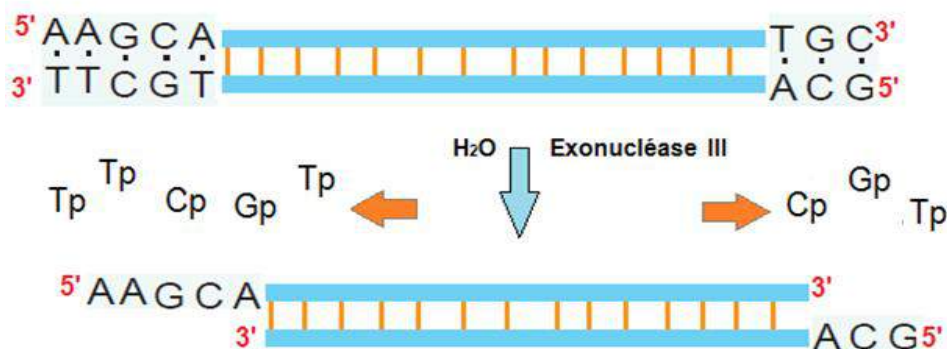


Fig.13: L'hydrolyse des extrémités 3' de l'ADN db par l'exonucléase III.

2-6-Ribonucléase T1: EC 3.1.27.3

Cette nucléase est produite par *Aspergillus oryzae*, elle a un PM de 11000 Da, et hydrolyse spécifiquement les ARN sb en rompant les liaisons 3'-5' phosphoester en aval des GMP, de telle manière que le produit soit une guanosine 3'-phosphate ou un oligonucléotide se

terminant par une guanosine 3'-phosphate (fig. 14). Cette enzyme est utilisée surtout dans le séquençage des ARN.

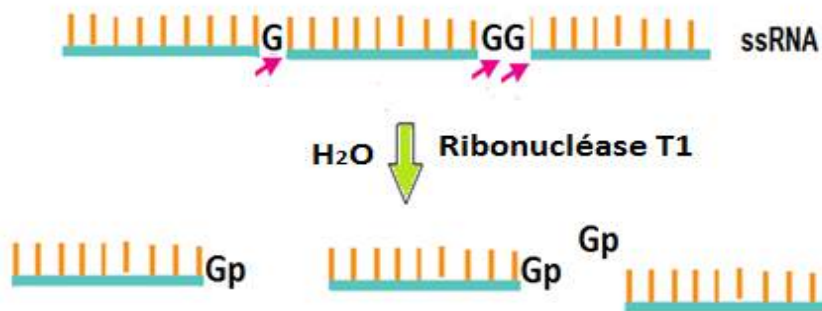


Fig.14: L'hydrolyse de l'ARN sb par la ribonucléase de T1.

2-7- Ribonucléase T2

Cette nucléase comme celle T1 est produite par *Aspergillus oryzae*, elle a un PM de 36000 Da, et hydrolyse spécifiquement les ARN sb en rompant les liaisons 3'-5' phosphoester en aval de l'adénosine (AMP) (fig. 15). Elle est utilisée aussi dans le séquençage de l'ARN.

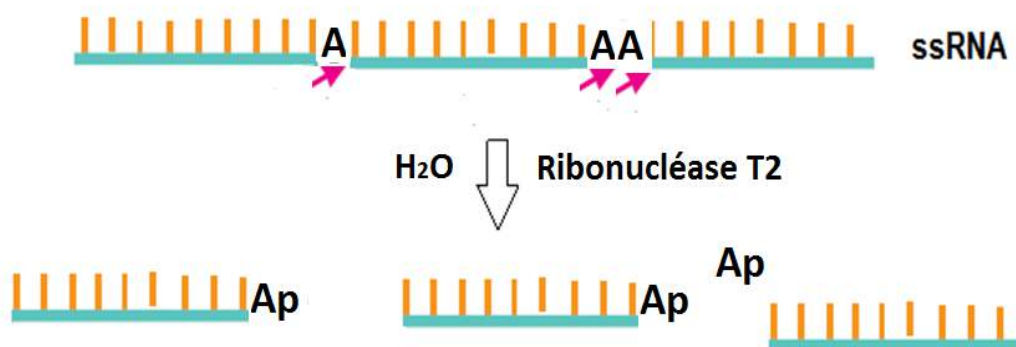


Fig.15: Hydrolyse de l'ARN sb par la ribonucléase T2.

2-8-Nucléase Bal31

Cette enzyme multifonctionnelle est produite par *Alteromonas espejana*. Elle possède à la fois:

- Activité **exonucléase** hydrolysant simultanément les deux extrémités 3' et 5' de l'ADN db.
- Activité **endodésoxyribonucléase** hautement spécifique des ADN sb.

La Bal31 agit séquentiellement, comme exonucléase dans les sens 3' → 5' puis par élimination endonucléotidique des simples brins sortants (fig. 16).

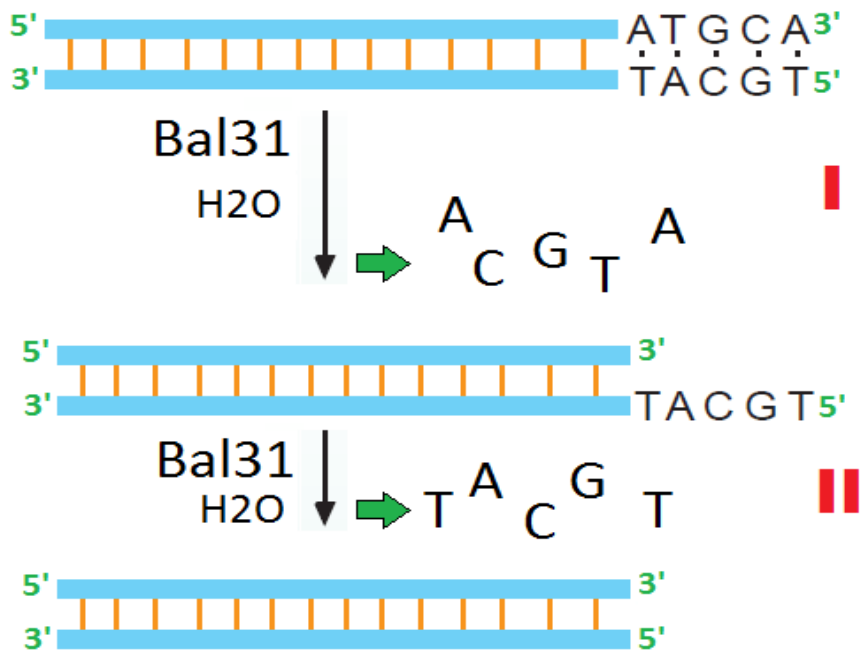


Fig.16: Activité exonucléase sur l'extrémité 3' puis endonucléase sur le fragment simple brin sortant du BAL31 sur l'ADN.

2-9- ADN ligase

La ligation entre deux molécules d'ADN par une liaison covalente entre l'extrémité 3'-OH d'un brin et celle 5'-phosphate de l'autre brin est le résultat de l'action d'une ligase. Cette réaction nécessite la consommation d'une molécule d'ATP.

L'ADN-ligase d'*E. coli* et celle du phage T4 sont les plus utilisées en génie génétique. La première utilise le NAD comme cofacteur et la deuxième utilise l'ATP et le Mg^{++} .

La quantité de l'ADN ligase nécessaire dans chaque réaction et l'activité de cette enzyme dépendent de certains nombre de facteurs:

- Nature des fragments d'ADN à ligaturer (extrémités franches ou cohésives).
- Longueur d'extrémités cohésives.
- Stabilité des liaisons hydrogène.
- La température d'incubation.
- La concentration des fragments.

Exemple:

- La vitesse de ligation des fragments de restriction de pBR322, de même longueur et tous à extrémités cohésives, varie selon la séquence de ces extrémités.

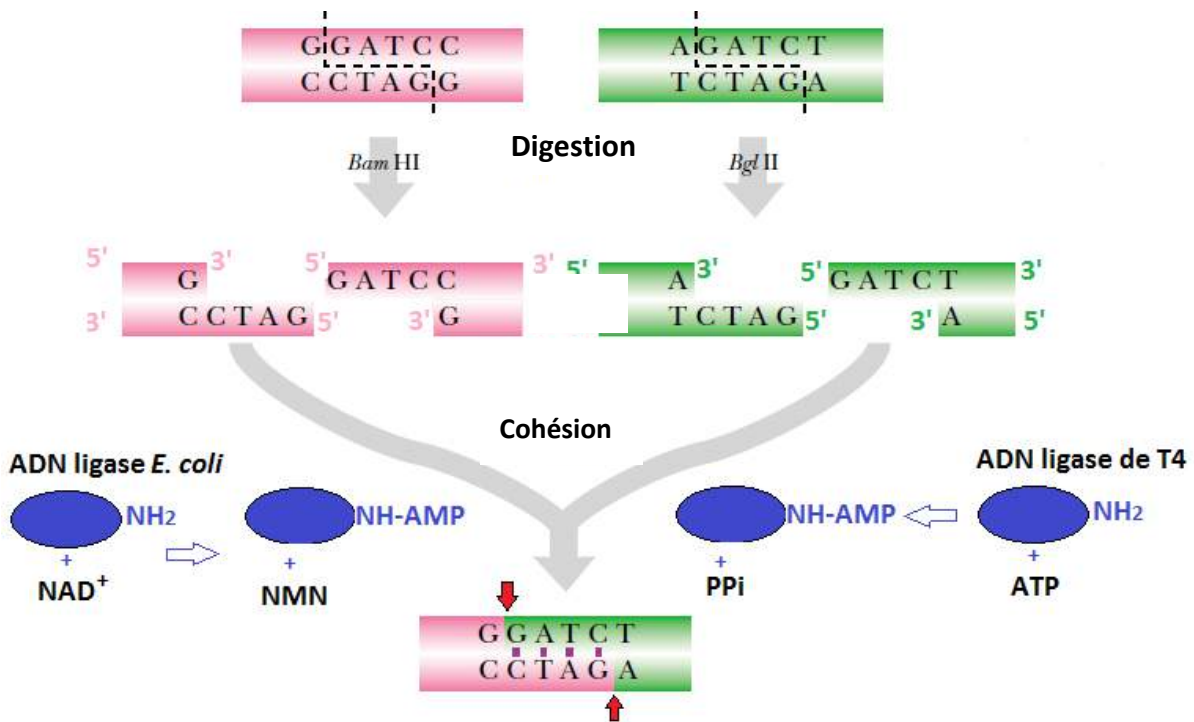


Fig.17: L'une des modalités de l'utilisation des ligases en génie génétique (ligation des extrémités cohésives des enzymes de restriction). L'enzyme ligature les extrémités 3'-OH et 5'-phosphate. L'ADN ligase est adénylylée par NAD (*E. coli*) ou par l'ATP (phage T4). Après l'adénylylation, l'extrémité 5' phosphate favorise la formation de liaison phosphoester par attaque nucléophilique ($^-\ddot{O}H$).

Génie génétique

Chapitre II

Hôtes de clonage

En biotechnologie, le génie génétique est utilisé à des fins commerciales comme pour la production de nouveaux vaccins, des grandes quantités de protéines valorisables, ou l'introduction de gènes spécifiques dans un organisme animal ou végétale. Dans chaque cas, le choix de l'hôte est essentiel puisqu'il nous orientera vers un type de vecteur adapté.



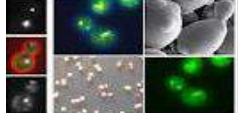
1- L'hôte idéal

Pour obtenir de grande quantité d'ADN cloné l'hôte idéal doit

- Se développer rapidement dans un milieu de culture peu onéreux.
- Être non pathogène.
- Être capable d'incorporer l'ADN.
- Être stable en culture
- Possède des enzymes appropriées pour la réplication du vecteur.

Les hôtes répondant à ces critères sont des microorganismes eucaryotes ou procaryotes dont les génomes sont bien connus car entièrement séquencés, génétiquement manipulables.

Tab.1: Hôtes de clonage moléculaire.

| <i>Escherichia coli</i> | <i>Bacillus subtilis</i> | <i>Saccharomyces cerevisiae</i> |
|---|--|--|
| Procaryotes | | Eucaryotes |
| Bacille à Gram –  | Bacilles à Gram + Spore +  |  |
| Avantages | | |
| <ul style="list-style-type: none"> - Génétiquement très bien connue. - Nombreuses souches disponibles - Procaryote le plus connu | <ul style="list-style-type: none"> - Facilement transformable - Non pathogène - Protéines secrétés naturellement - Formation d'endospores facilitant les cultures. | <ul style="list-style-type: none"> - Génétiquement très bien connue - Non pathogènes - Assure la maturation des ARNm et des protéines - Facile à cultiver. |
| Inconvénients | | |
| <ul style="list-style-type: none"> - Potentiellement pathogènes -Périplasma piégeant les | <ul style="list-style-type: none"> - Génétiquement instable - Génétique moins connue | <ul style="list-style-type: none"> - Plasmides instables -Pas de réplication pour la |

| | | |
|-----------|--------------------|------------------------------------|
| protéines | qu' <i>E. coli</i> | plupart des plasmides procaryotes. |
|-----------|--------------------|------------------------------------|

Il faut noter que *E. coli* est l'organisme le plus utilisé en clonage moléculaire. Malgré que cette bactérie soit comptée parmi la flore normale de l'intestin de l'homme et les animaux, il est aussi un pathogène potentiel (surtout les souches sauvages). Et aussi la synthèse d'endotoxines (LPS) susceptible de contaminer les produits finis, cela constitue un problème potentiel notamment pour les produits pharmaceutiques administrés par voie intraveineuse.

Aussi le problème, que *E. coli* retiens des protéines extracellulaires dans son espace périplasmique, ce qui peut rendre l'isolement et la purification des protéines recombinantes difficiles et coûteuse.

Avec *Bacillus subtilis* l'inconvénient majeur reste la difficulté de maintenir la répllication plasmidique dans les sous cultures, ce qui engendre souvent la perte de l'ADN cloné.

Des vecteurs plasmidique et des YAC (Yeast Artificial Chromosome) ont été développés pour le clonage dans la levure *Saccharomyces cerevisiae*. L'avantage que présente cet hôte est qu'il possède les ARN et les systèmes post traductionnels complexes nécessaire à la synthèse de produits de gènes d'organismes supérieurs. Les processus post traductionnels peuvent être à l'origine de problème de clonage.

La culture de cellules de mammifères présente un coût élevé et des difficultés de production à grande échelle. En plus le niveau d'expression des gènes clonés est souvent faible (aussi pour les insectes, plantes, etc.).

On dit dans les cas des bactéries transformation, le processus d'intégration de l'ADN étranger, mais pour les eucaryotes on dit la transfection. Car, la transformation des cellules de mammifère désigne habituellement la conversion en cellules malignes (tumoraux, cancéreuses). L'exemple le plus connu de l'application de cette caractéristique en biotechnologie est la production des anticorps monoclonaux.

2- Méthodes d'introduction de l'ADN à cloner dans la cellule hôte

Il y'a plusieurs méthodes sont largement utilisées pour introduire l'ADN dans les cellules hôtes.

N.B. Chez les bactéries on peut transférer l'ADN à cloner par trois méthodes: la transformation, la transduction et la conjugaison (voir cours génétique microbienne).

2-1- Électroporation

Cette technique implique l'exposition de l'hôte à des décharges électriques afin d'ouvrir les pores (temporairement) dans la membrane par lesquels, l'ADN cloné, ajouté dans le milieu, peut pénétrer sans lyse des cellules (Fig. 1).

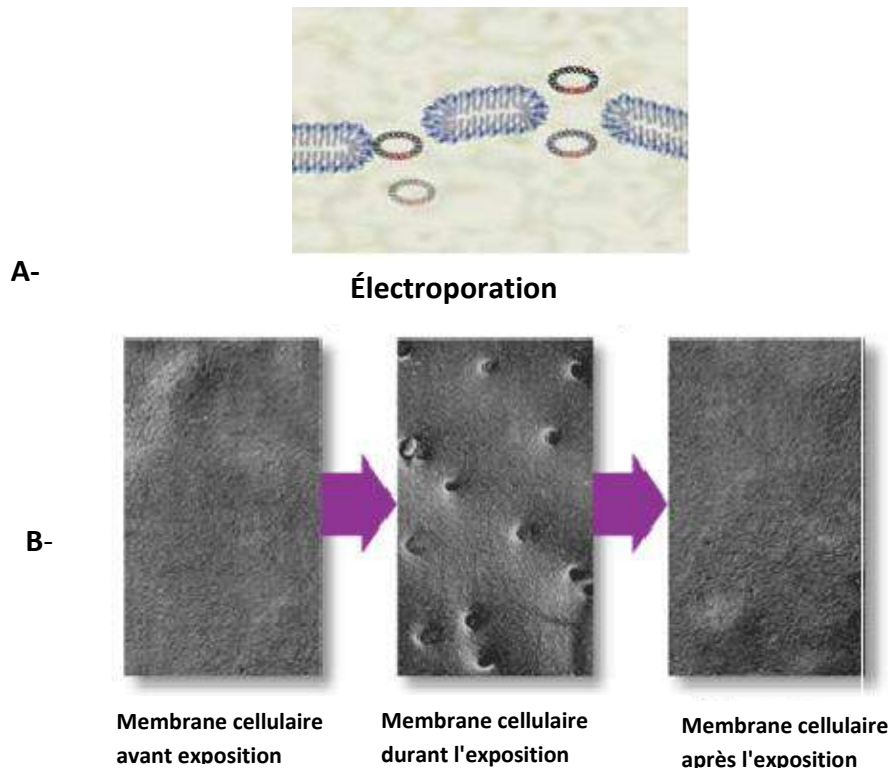


Fig.1: A : Représentation graphique des plasmides en passant par les pores aqueux dans la membrane plasmique. B : Phénomène d'électroporation.

2-2- Un canon à Particules

La transfection des cellules cibles se fait par des billes métalliques (généralement de tungstène) recouvertes d'acides nucléiques, en perçant paroi et membranes plasmiques sans provoquer de lyse cellulaire. Cette technique a été utilisée sur des levures, des algues, des cellules de plantes et même des mitochondries et chloroplastes. De plus contrairement à l'électroporation, cette technique peut être utilisée pour introduire de l'ADN dans des tissus intacts comme des graines de plantes.

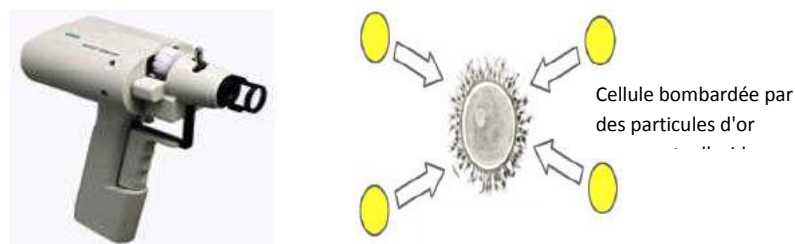
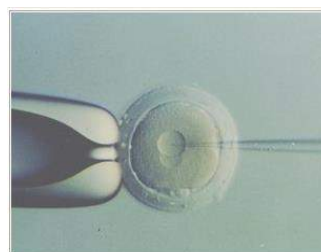


Fig.2:Canon à acides nucléique (*biolistics*) pour la transfection d'eucaryotes.

2-3-Microinjection

Dans les cellules animales, l'ADN peut être injecté dans le noyau par microinjection.

Fig.3:Microinjection des gènes dans le pronucléus mâle d'un ovocyte de lapin



Génie génétique

Chapitre III

Vecteurs de clonage

Un vecteur de clonage est un petit élément génétique à réplication autonome, utilisé pour produire de multitude de copie du gène d'intérêt. Ces vecteurs de clonage sont spécialement conçus pour permettre l'intégration de la portion d'ADN exogène dans un site spécifique sans que cela affecte sa propre réplication. Il existe plusieurs types de vecteurs pouvant introduire des molécules d'ADN recombinées dans les bactéries, les levures, les cellules végétales ou les cellules de mammifères et humaines.

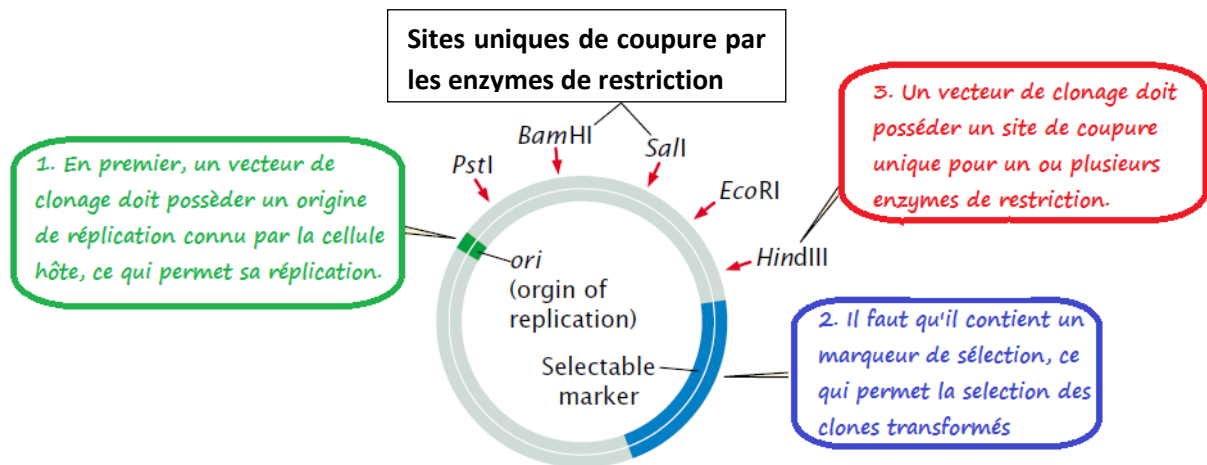


Fig.1: Les trois caractéristiques d'un vecteur de clonage idéal.

1- Vecteurs bactériens

La bactérie *E. coli* est la bactérie la plus utilisée en clonage moléculaire, l'ADN recombinant peut être introduit sous forme de plasmides, Phages, phagmides, cosmides, BAC etc.

1-1- Plasmides

Les plasmides sont des petits éléments génétiques extra-chromosomiques doués de la réplication autonome, ce sont typiquement des molécules d'ADN double brin, circulaires, leur taille varie de 1 kb à deux ou trois centaines de kb (< 5% de la taille du chromosome bactérien). Ils contiennent des gènes codants souvent pour des protéines qui donnent un ou des avantage(s) à la cellule hôte. Comme par exemple :

- Résistance aux antibiotiques
- Résistance aux métaux lourds
- Dégradation de composés aromatiques

- Production de toxines
- Production d'antibiotiques
- Induction des tumeurs chez les plantes

L'amplification de l'ADN plasmidique par réplication varie entre les différents plasmides

- Plasmides à petit nombre de copies.
- Plasmides à grand nombre de copies.

Il faut noter que des mutations dans les plasmides peuvent affecter le nombre de copies du plasmide par cellule, surtout celles qui touchent au répresseur. En absence de ce répresseur, le nombre de copies augmente jusqu'à ce que la machinerie cellulaire responsable de la réplication soit saturée. On obtient la même chose par l'utilisation des antibiotiques interférant avec la synthèse des protéines. Par exemple le chloramphénicol (inhibe la synthèse du répresseur), le nombre peut atteindre plusieurs milliers.

Les plasmides naturels possèdent dans leur génomes un locus nommé "*par*" qui fonctionne à la façon de centromère des chromosomes eucaryotes, ce locus contrôle la répartition précise des plasmides parmi les cellules filles lors de la division cellulaire. Pour réduire la taille des plasmides qui servent de vecteurs de clonage, le locus "*par*" est très souvent éliminé. C'est notamment le cas du plasmide pBR322 et leurs dérivés.

Les plasmides sont de bons vecteurs de clonage chez les bactéries parce qu'ils se multiplient en nombre de copies important et qu'ils sont aisément purifiables. Les marqueurs de sélection (Ex. gènes de résistance aux antibiotiques) permettent l'identification des bactéries recombinantes (transformées) qu'ils les portent.

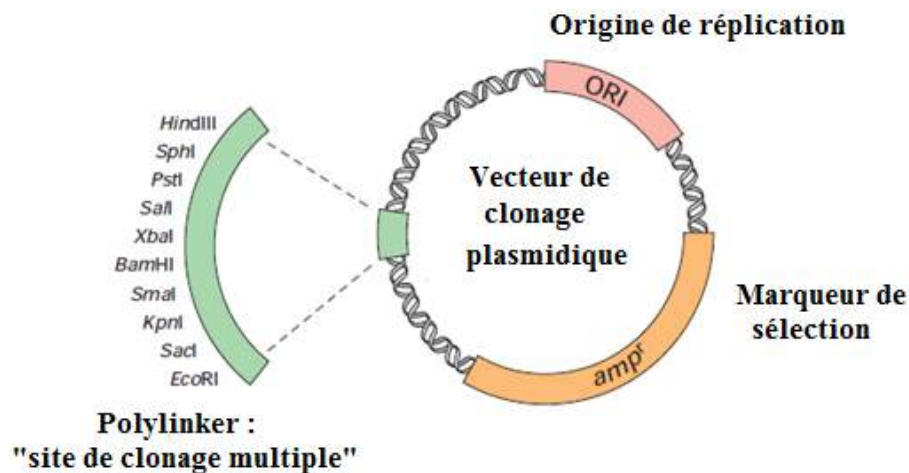


Fig.2: Composants basiques d'un vecteur plasmidique qui peut se répliqué chez *E. coli*(Lodish et al.,2003).

1-1-1-Plasmide pBR322

Le pBR322 appartient à une série de vecteurs de clonage de première génération, partiellement construit par génie génétique. Qui fut construit sur la base d'une chimère rassemblant les éléments intéressants des plasmides naturels et en augmentant le nombre de sites uniques de coupure par les enzymes de restriction.

De nouvelle génération de plasmides plus puissants sans cesse croissantes ont été développés depuis pBR322 et ces dérivés. C'est le cas de la famille pUC appelés (les vecteurs de clonage de seconde génération).

Les vecteurs de clonage de seconde génération ce sont des petits plasmides d'environ 2700pb. Le plasmide pUC19 contient le gène de résistance à l'ampicilline de pBR322, mais en plus il possède une partie du gène *lacZ* dans lequel a été introduit un site multiple de clonage (Polylinker), contenant toute une série de sites de coupure unique.

Le fait que ce polylinker soit inséré dans le gène *lacZ* qui intervient dans le catabolisme du lactose permet de révéler facilement l'intégration d'un insert par « inactivation insertionnelle ».

L'utilisation d'un inducteur coloré comme le X-gal (5-bromo-4chloro-3-indonyl-β-D-galactoside), l'hydrolyse de ce dernier en dibromo-5,5-dichloro-4, 4-indigo (de couleur bleu), indique la production de β-galactosidase.

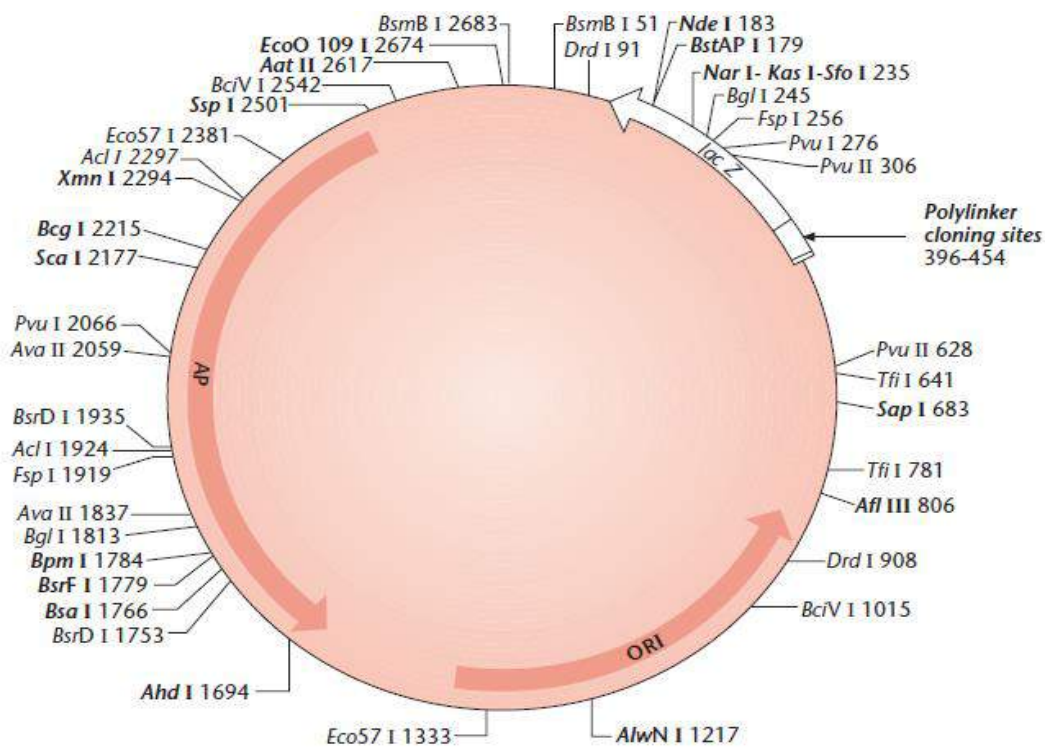


Fig.4: La carte génétique de certain vecteur de type pUC dérivés de pBR322, le site de clonage multiple (polylinker) est introduit dans le gène *lacZ*, sans interrompre la fonction du gène (Primose et al.,2002).

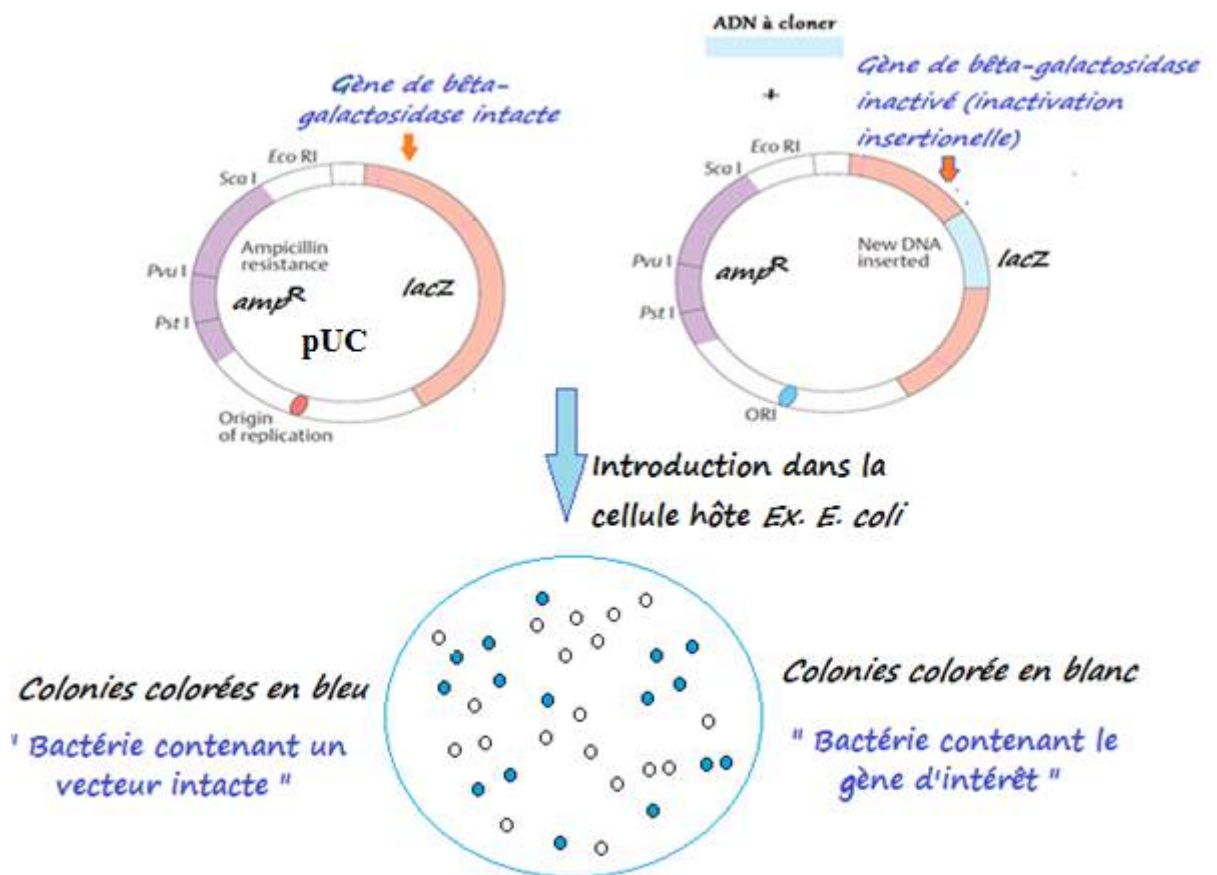


Fig.5: Insertion d'un fragment d'ADN (< 10 kb) dans un vecteur de deuxième génération (type pUC). L'inactivation insertionnelle du gène *lacZ* permet de révéler facilement les colonies contenant le gène d'intérêt (sélection positive: colonies blanche). Le milieu de sélection est supplémenté par l'ampicilline.

La sélection des clones contenant le gène d'intérêt par le test x-gal. Cette stratégie a également été largement employée dans le développement de vecteurs de clonage de type viraux.

Dans certains vecteurs le polylinker est introduit dans un gène qui lorsqu'il est exprimé, est létal pour la cellule. Donc seules les cellules qui contiennent un plasmide dans lequel le gène létal est inactivé sont capables de se développer.

La nécessité de vecteur pour le clonage de grands fragments d'ADN et d'autres objectifs pour répondre à des besoins précis. Une collection de vecteurs très spécialisés s'est constituée.

1-2-Bactériophage λ

Le bactériophage λ a été découvert par E.M. Lederberg en 1950. C'est un virus d'*E. coli*, l'ADN de ce phage est une molécule linéaire d'ADN double brin de 48 kb. A chaque extrémité 5' se trouve une région monocaténaire de 12 nucléotides, l'une complémentaire de l'autre et leur association donne une structure circulaire à l'ADN dans la cellule hôte. L'association de ces extrémités cohésive naturelles forme le site *cos* (fig.) [*cos*: Des éléments important pour la réplication et l'encapsidation de bactériophage λ].

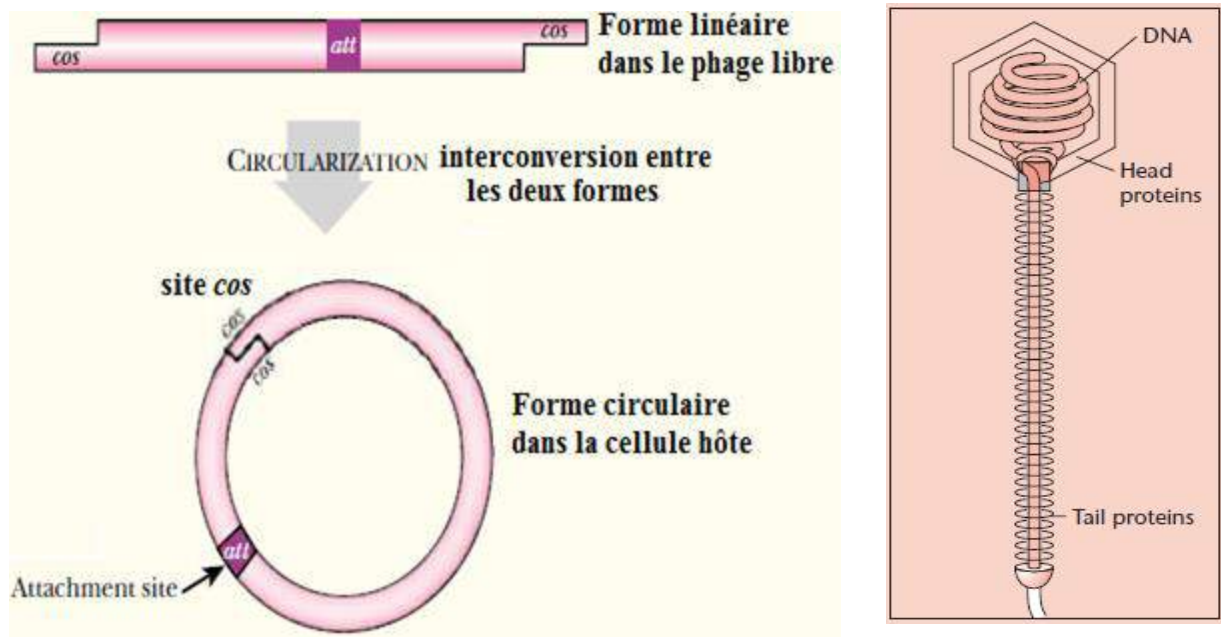


Fig.5: Différentes formes du génome du bactériophage λ , et sa structure (Primose *et al.*, 2002).

Le bactériophage λ a donné naissance aux premiers vecteurs de types phagiques car:

- Ça biologie est bien connue
- La quantité d'ADN qu'il peut intégrer est plus importante que celle véhiculée par les vecteurs plasmidiques. Il est possible d'insérer jusqu'au 22kb, après élimination de la partie indispensable au cycle de vie du phage (fig. 6)
- La possibilité d'empaqueter *in vitro* l'ADN phagique nu recombinant dans les têtes de phage.
- Il a une capacité d'infection (transfection) de l'hôte très rapide.
- Le nombre de copies par cellule étant considérable.
- Le rendement de cette transfection est très supérieur à ce qui est obtenu lors de la transformation de la bactérie par les plasmides.

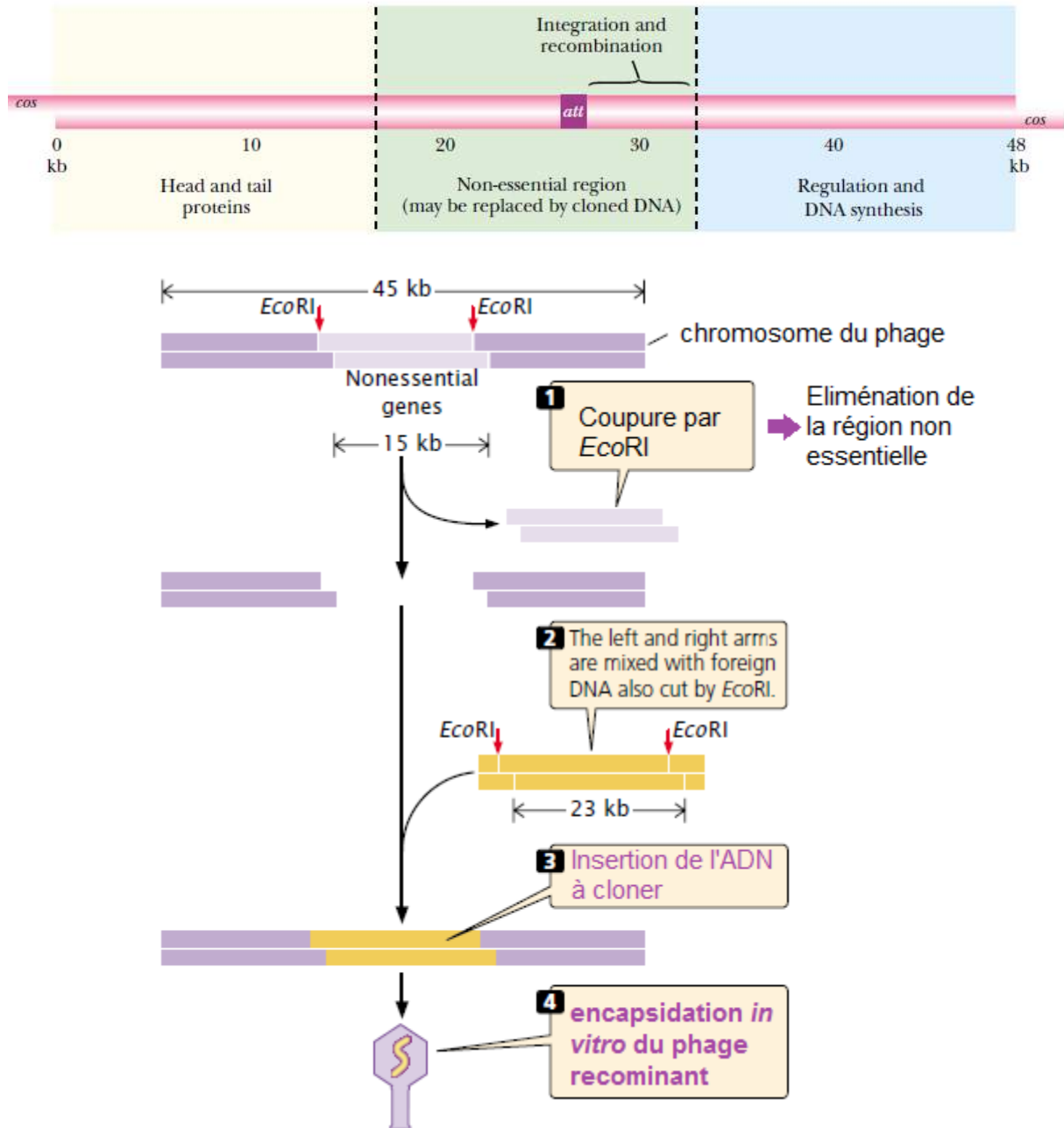


Fig.6: L'élimination de la partie indispensable (non essentielle) au cycle de vie du bactériophage λ augmente la capacité d'insertion de grand fragment (22kb) (Passarge, 2007).

1-1-3-1-Principales étapes de la construction d'un vecteur phagique

- 1- Produire une grande quantité de phage, la purifier, puis en extraire son ADN génomique, qui sera digéré par une enzyme de restriction.
- 2- Hybrider les deux bras du phage avec le fragment d'ADN à cloner (ce dernier doit avoir une taille adéquate) puis souder par l'ADN ligase.
- 3- Procéder à l'encapsidation *in vitro* de l'ADN recombinant en ajoutant les protéines phagiques de tête et de la queue. Ces derniers ils vont s'auto assembler pour former les nouveaux virions recombinants infectieux.

- 4- Infecter des bactéries (cellules hôtes) et les étaler sur boîte de Pétri, chaque plage de lyse correspond à un phage recombinant qui peut être récupéré.
- 5- Vérifier la présence d'un insert dans l'ADN recombinant par toute procédure appropriée (hybridation ADN-ADN, séquençage, test bleu-blanc etc.)

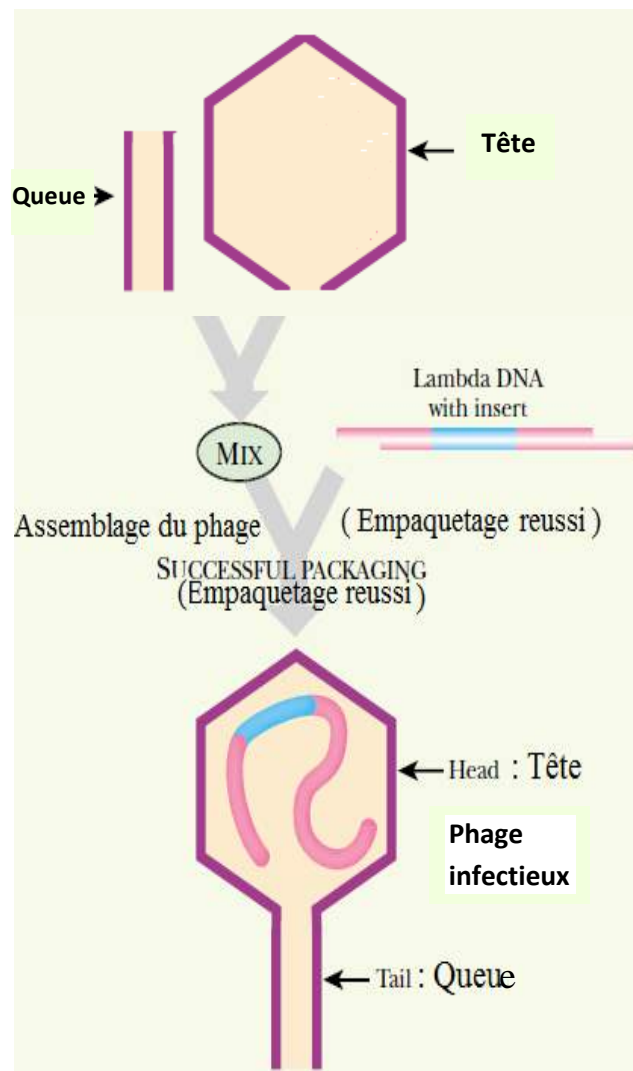


Fig.7: Encapsidation *in vitro* (Passarge, 2007).

1-3-Bactériophage M13

Comme le bactériophage λ , le M13 est un petit virus d'*E. coli* (6.407 kb). L'ADN de ce phage est sous forme simple brin circulaire positif. Mais lors de la réplication un ADN négatif complémentaire est synthétisé, c'est la forme répliquative (replicative form) ou la forme intermédiaire. La forme répliquative présente toute les qualités des vecteurs plasmidiques. En plus cette forme sert de matrice à la synthèse d'ADN simple brin qui peut être encapsidé et par conséquent facilement isolable.

Le M13 à des dérivés contenant une partie du gène *lacZ* (peptide α), avec un polylinker (MCS: multiple cloning site) de 13 sites uniques de restriction (54pb). Ce qui permet d'insérer des fragments d'ADN dans ces sites, et la sélection des colonies blanches sur des boites contenant l'analogue X-gal.

Ce vecteur et ces dérivés peuvent être utilisés :

- Pour séquencer des fragments d'ADN même de séquences inconnues par la technique de Sanger.
- Pour le clonage des fragments d'ADN de taille jusqu'à six fois plus grand que l'ADN viral.
- Pour la transfection des cellules compétentes d'*E. coli* par les deux formes (MC, BC [RF]).
- Dans la mutagenèse dirigée (Fig. 8).

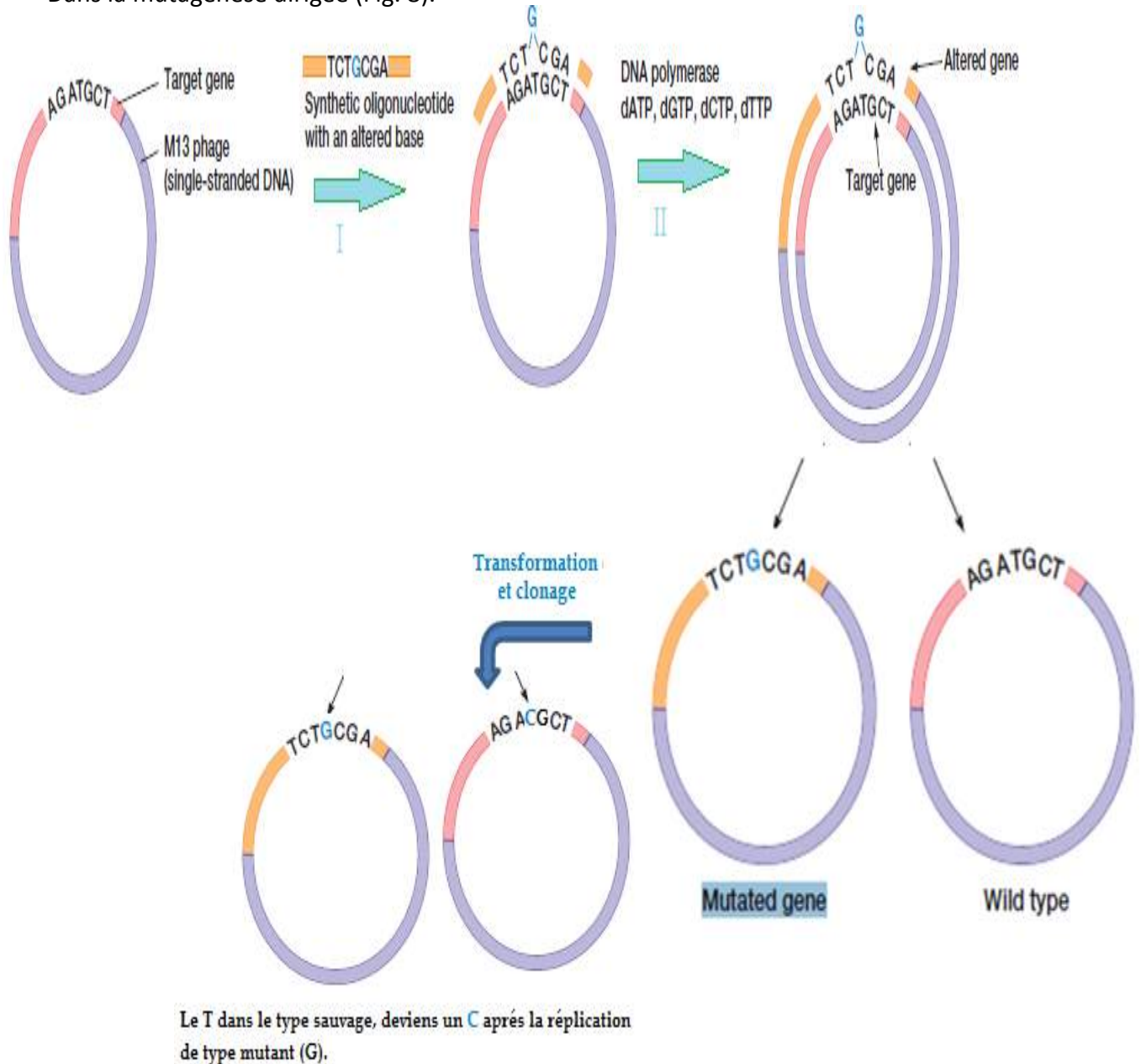


Fig.8: Utilisation du vecteur M13 dans la mutagenèse dirigée.

1-4-Cosmides

Les cosmides sont des vecteurs artificiels ($\approx 5\text{kb}$) constitués d'un plasmide classique auquel ont été ajoutées les séquences *cos* du phage λ . Ces vecteurs rassemblent à la fois les propriétés intéressantes des plasmides comme :

- L'origine de réplication
- Gène de résistance à un antibiotique

Et celles du bactériophage

- Encapsidation *in vitro* de grand fragment d'ADN.

N.B. Il faut insérer de **32 à 47 kb** d'ADN étranger dans un vecteur de type cosmide pour qu'il soit encapsidé.

Après l'encapsidation *in vitro*, la particule virale formée peut infecter une cellule hôte appropriée. L'ADN du cosmide recombinant injecté se circularise à l'intérieur de la cellule comme un ADN de phage. Mais il se réplique comme un plasmide, sans exprimer les fonctions du phage.

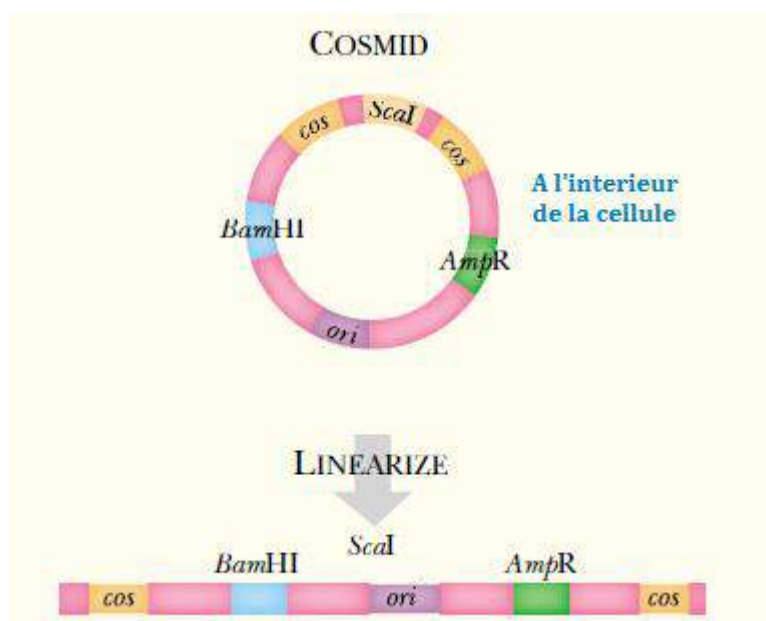


Fig.9: Les deux formes (linéaire et circulaire) d'un cosmide.

Les cellules transformées sont sélectionnées sur un milieu contenant de l'antibiotique (ampicilline). Comme conclusion les cosmides permettent :

- L'encapsidation d'un plasmide modifié (recombinant) dans un virion.
- D'obtenir des rendements d'intégration bien supérieurs à ceux que donne une transformation bactérienne par un plasmide.
- Le clonage d'un fragment d'ADN plus grand (de 32 à 47 kb) que celui véhiculé par un plasmide ($\approx 10\text{ kb}$) ou par le bactériophage λ ($\approx 22\text{ kb}$).
- Nécessite moins de clones pour créer une banque génomique.
- Les cosmides sont plus stables que les plasmides.

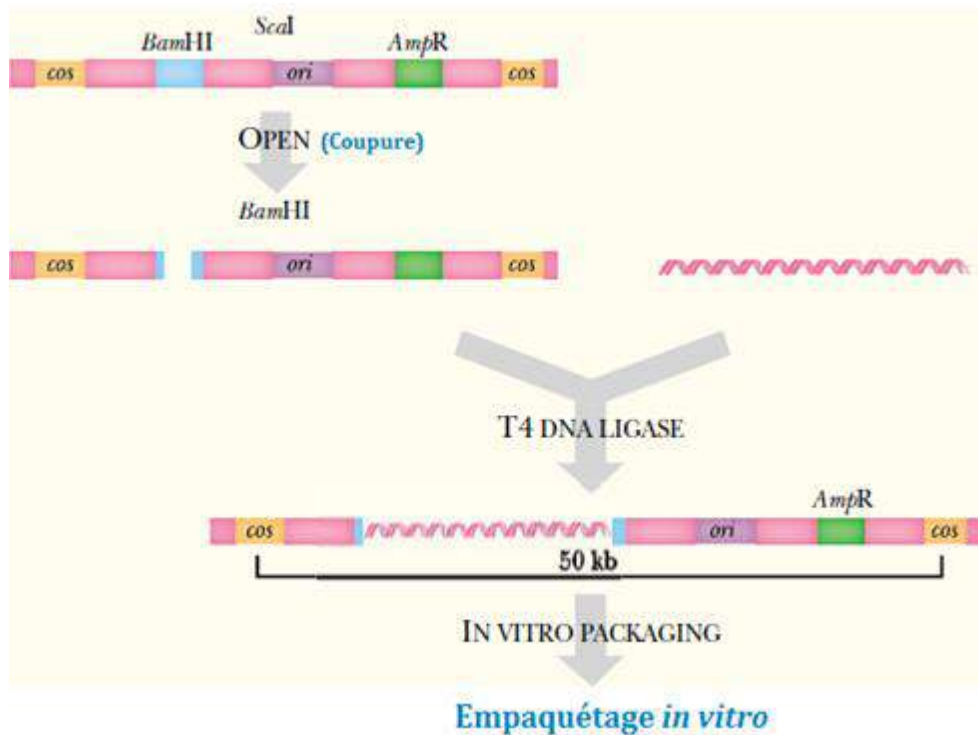


Fig.10: L'empaquetage *in vitro* d'un cosmide recombinant.

1-5-Phagemides ou phasmides

Sont des vecteurs qui combinent des éléments d'origine plasmidiques et phagiques. Le phagmide le plus utilisé est **pBluescriptII KS**, c'est un dérivé du plasmide pUC19, il contient un polylinker, interrompu par deux promoteurs (T3 et T7) se lus en sens opposés, il contient aussi un promoteur *lac* inducible avec une partie du gène *lacZ* (blanc-bleu sélection). Une origine de réplique dérivée de M13 (phage filamenteux). Un *ori ColEI* pour permettre la réplique du phage comme un plasmide. Ils sont utilisés pour cloner de grand fragment d'ADN et la manipulation des gènes.

1-6- Chromosomes artificiels bactériens "BAC"

Les BACs (Bacterial Artificial Chromosomes) sont construits à partir du plasmide F (99.2 kb) à cause de ces propriétés intéressantes (Ex. Phénotype Hfr: mobilisation du ou souvent d'une partie du chromosome bactérien d'une cellule donatrice à une cellule réceptrice, cela après l'intégration au chromosome). Il contient qu'une petite fraction du plasmide F, et comme les vecteurs plasmidiques, il possède un polylinker, gène de résistance à un antibiotique comme marqueur de sélection (*cat*: résistance au chloramphénicol). Une origine de réplique et *rep E* qui sont nécessaires à la réplique. Des régulateurs de la réplique (*sop B*, *sop A*) pour maintenir un faible nombre de copies, sa taille est de 6.5 kb.

L'hôte typique d'un BAC est une souche mutante de *E. coli*, cette souche ne dispose pas des systèmes de modification et de restriction de la souche sauvage, pour empêcher la destruction du BAC. Ainsi, elle a perdu les capacités de recombinaison normales, cela interdit la recombinaison et les réarrangements de l'ADN cloné du BAC avec le chromosome de l'hôte. Le BAC a la capacité d'insérer jusqu'au **300 kb**.

2-Chromosomes artificiels des levures "YAC"

Les YACs (Yeast Artificial Chromosomes) doivent avoir :

- Une origine de réplication
- Des télomères pour la réplication de l'ADN aux extrémités du chromosome
- Un centromère (ségrégation lors de la mitose).
- Site de clonage multiple (MCS : multiple cloning site or polylinker)
- Marqueur de sélection.

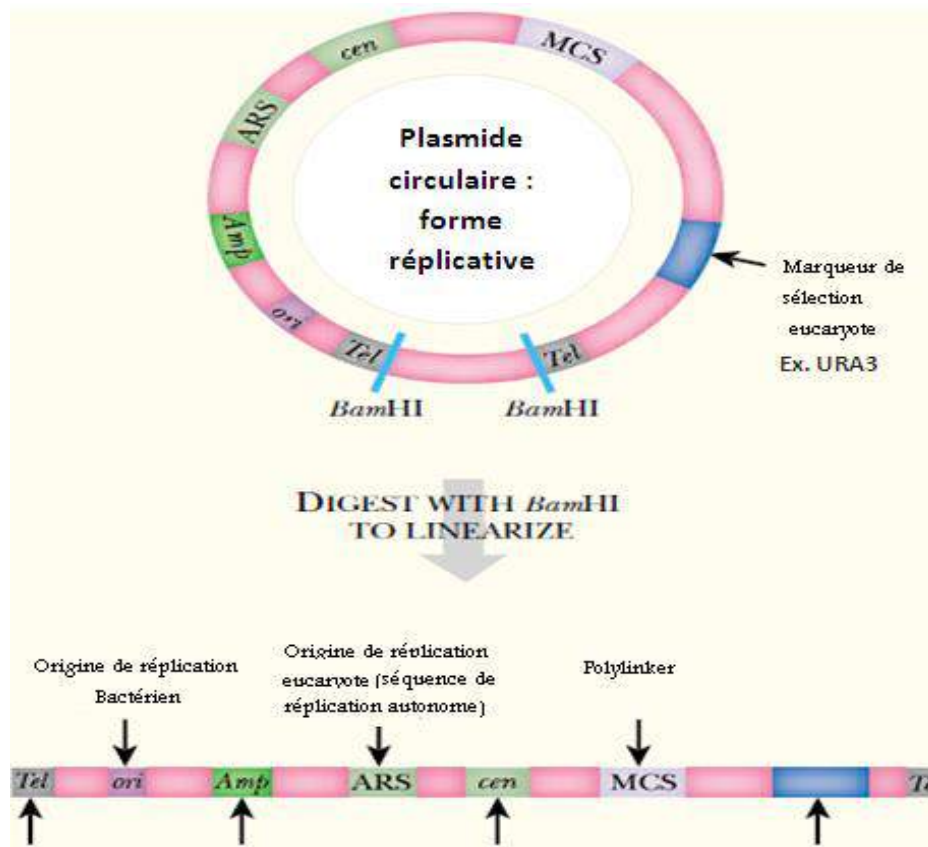


Fig.11: Diagramme d'un YAC contenant des éléments essentiels pour la réplication chez les bactéries (*Ori*, gène de résistance (*Amp*), forme circulaire), et chez les levures (forme linéaire obtenue après digestion par *Bam*HI) (Clark, 2005).

Les YACs ont une taille d'environ 10kb, mais ils peuvent recevoir de longs fragments d'ADN à cloner (de 200 jusqu'à 2000 kb). Malgré que les YAC puissent porter de plus grands fragments (inserts) que les BAC, les problèmes de recombinaison et de réarrangement de l'ADN cloné sont plus importants avec la levure qu'avec *E. coli*. Ce qui rend les BAC plus utilisés que les YAC en clonage génomique.

3-Chromosomes artificiels dérivé de P1 "PAC"

Ils sont construits par la combinaison des éléments du système P1, et du facteur F, les vecteurs PAC permettent de manipuler des inserts de 100 à 300 kb.

4-Plasmide Ti (Tumor inducing Plasmide)

La bactérie *Agrobacterium tumefaciens* renferme de grand plasmides (140 à 235 kb), mais les cellules de plante transformées intègrent seulement un petit fragment spécifique du plasmide d'une taille d'environ 23kb appelé l'ADN-t, ce fragment spécifie le type d'opine synthétisé dans le tissu végétal. Des souches d'*A.tumefaciens* porteuse du plasmide Ti sont maintenant disponibles pour produire des plantes transgénique. Exemple: l'obtention des plantes résistantes aux insectes (BT), aux herbicides, etc.

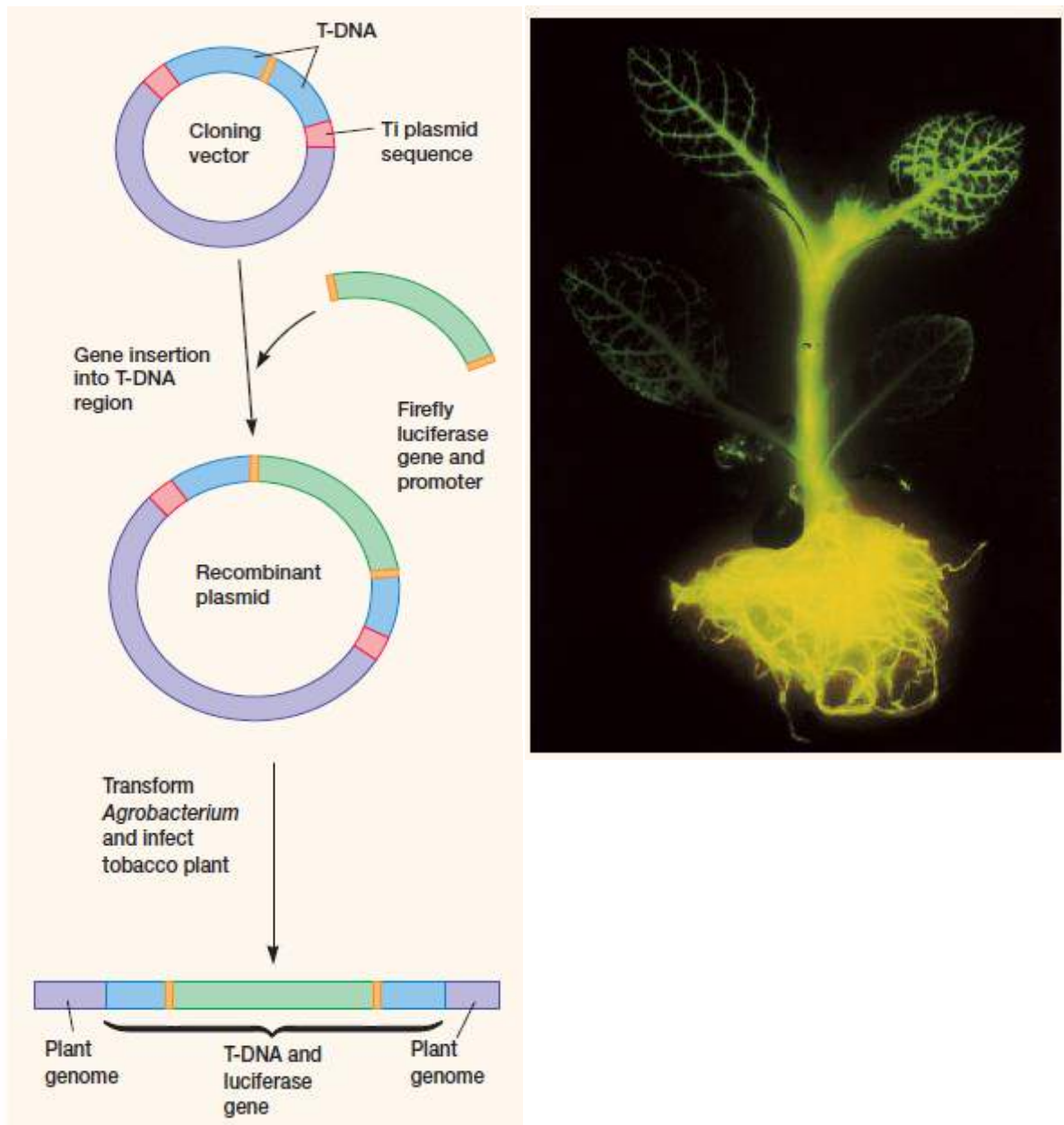


Fig.12: L'utilisation du vecteur Ti pour transformer la plante *Nicotiana tabacum* par le gène d'une luciférase. Le gène est introduit dans la région de l'ADN-T. Cette transformation rend les bactéries bioluminescentes (Prescott, 2002).

5-Vecteurs spécifiques

Pour répondre aux besoins de la biotechnologie, surtout si c'est le but du clonage est d'obtenir un niveau d'expression élevé d'un gène dans un hôte approprié, d'autres vecteurs spécialisés ont été produits.

5-1- Vecteurs navettes et d'expression

Les vecteurs navettes peuvent transporter un ADN cloné entre deux organismes différents et se répliquer de manière stable dans chacun d'eux.

Exemple: Un vecteur se réplique chez *E. coli* et dans une levure (fig. 14), ou bien autre bactérie et même une cellule de mammifère.

Il faut noter que les organismes possèdent des systèmes de régulation complexe qui sont des obstacles à l'expression des gènes étrangers. Pour cette raison les vecteurs d'expression ont été élaborés.

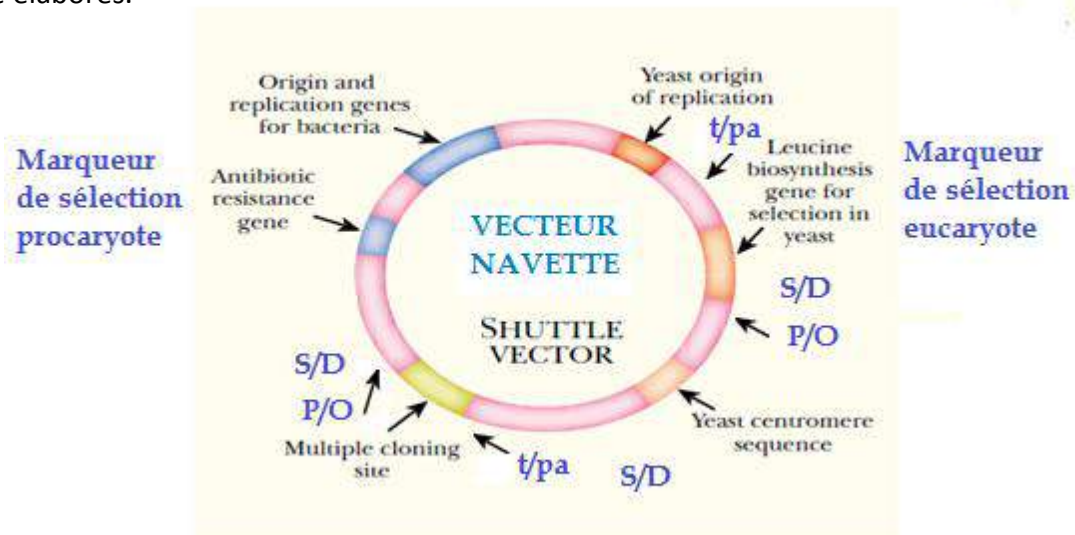


Fig. 13: Caractéristique d'un vecteur navette d'expression (shuttle expression vector). t/pa: signaux de terminaison de la transcription/ poly adénylation. T1, T2: terminateurs de la transcription. P/O : promoteur –Opérateur, S/D: séquence de Shine Dalgarno (Clark, 2005).

Un vecteur d'expression permet de cloner un gène, mais contient aussi les séquences de régulation nécessaire à son expression. Dans la plupart des cas on utilise avec le promoteur l'opérateur responsable sur sa contrôle. Comme sa on peut régler l'expression par l'ajout de l'inducteur ou du répresseur.

Exemples: I- Promoteur *lac* avec opérateur *lac*. II- Promoteur *trp* avec opérateur *lac*.

On peut aussi contrôler l'expression des vecteurs par des éléments de bactériophage. Par exemple, le bactériophage T7, lorsqu'il infecte *E. coli*, il code sa propre ARN polymérase qui reconnaît uniquement les promoteurs de T7 et inhibe la transcription de l'hôte (gènes de l'hôte reste silencieux). Donc l'utilisation de gènes clonés sous contrôle d'un promoteur T7, limite la transcription uniquement aux gènes clonés. Cela n'est possible qu'en introduisant dans le plasmide un gène codant l'ARN polymérase T7 placé sous contrôle d'un système de régulation simple comme le système *lac*.

6- Régulation de la transcription des vecteurs d'expression

- L'élément clés du vecteur d'expression est le mécanisme de contrôle transcriptionnel.
- Une forte expression suppose la production de grandes quantités d'ARNm.
- La région -10 → -35 est une zone particulièrement importante du promoteur.
- Le promoteur normalement associé au gène cloné risque d'être faiblement efficace dans un nouvelle hôte.

Exemple: Les promoteurs des eucaryotes ou même des autres procaryotes fonctionnent peu ou pas du tout chez *E. coli*.

Pour régler ce problème, des promoteurs d'*E. coli* ont été utilisés par la construction des vecteurs d'expression dont les promoteurs des opérons : *lac*, *trp*, *tac*, *trc* (hybrides synthétiques des promoteurs *lac* et *trp*). Et P_L lambda (P_λ). Ce sont des promoteurs forts qui peuvent être spécifiquement régulés. Le promoteur doit être fonctionnel et correctement situé afin de permettre la transcription des gènes clonés.

Exemple: Pour la production d'une protéine toxique pour la cellule hôte, il faut pouvoir déclencher simultanément l'expression des gènes dans toutes les copies en activant un mécanisme de régulation après une phase avancée de la croissance (pour produire suffisamment de biomasse).

7-Traduction des gènes clonés

Pour traduire un gène cloné dans un hôte procaryote (Ex. *E. coli*), il faut qu'il y'a tout les éléments nécessaires pour la traduction:

- Site de liaison au ribosome (S/D) : Shine Dalgarno sequence
- Codon d'initiation sur l'ARNm à coté de S/D.
- Les gènes eucaryotes n'ont pas le site S/D, qui doit être introduit dans un vecteur si un haut niveau d'expression est attendu.
- L'usage des codons peut être un obstacle, car certains codons sont plus utilisés que d'autres. Aussi, l'usage des codons diffère entre les organismes et donc la concentration des ARNt est différente.

Pour résoudre ce problème, la mutagenèse dirigée et l'ADN synthétique peuvent être utilisés pour changer des codons sélectionnés dans le gène, pour le rendre plus proche du mode d'usage des codons de l'hôte.

8-Expression de gène de mammifères chez les bactéries

Pour pouvoir cloner et exprimer des gènes de mammifères dans des cellules hôtes bactériennes, c'est l'un des grands défis de génie génétique, à cause de plusieurs facteurs et obstacles:

- La détection du bon clone
- La présence des introns c'est un obstacle.
- Les protéines synthétisées dans un hôte procaryote peuvent être dégradées par des protéases intracellulaires avant de pouvoir les isoler.

- Certaines protéines issue des d'eucaryotes sont toxiques chez un hôte procaryote engendrant ainsi la mort de l'hôte avant la synthèse de quantités suffisantes du produit.

Parfois les protéines d'intérêt forme des corps d'inclusion dans l'hôte facilement purifiable du fait de leur taille, mais très difficile à solubiliser (repliement incorrecte des protéines). Car une protéine pour fonctionner, elle doit prendre une conformation spéciale et ce placer à un endroit précis dans la cellule. Plusieurs protéines prennent spontanément une conformation active lors de leur synthèse, en revanche d'autre nécessite l'assistance des chaperons moléculaires (protéines impliquées dans la réconformation des protéines mâle repliées). Pour régler le problème repliement incorrecte et l'excrétion des protéines. Il faut utiliser un hôte qui surproduit desprotéines chaperons qui aident au repliement correcte. Les problèmes de repliement ne peuvent le plus souvent se résoudre que si la protéine du gène clonée est une protéine de fusion.

Pour résoudre le problème de présence des introns il faut utiliser l'ARNm pour synthétiser l'ADN à cloner (ADNc). Par exemple lors de la synthèse du facteur de coagulation sanguin (Facteur VIII) dans *E. coli*. Le gène est récupéré par la technique de l'ADNc, puis l'ADNc est amplifier par PCR et insérer à côté d'un promoteur fort du phage T7.

9-Production d'insuline

La production des protéines humaines est le secteur des biotechnologies le plus rentable.

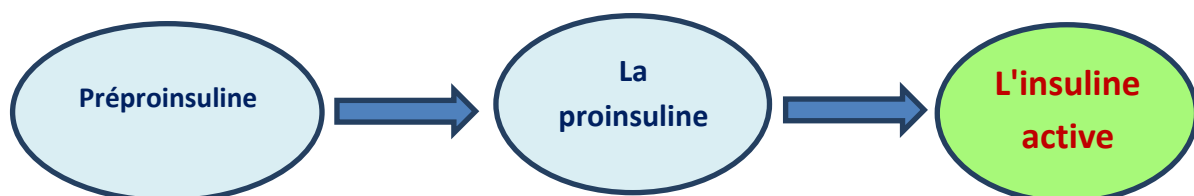
Question: Pourquoi l'utilisation des cultures microbienne est préférée?

Dans les cellules de mammifères la quantité de protéine de grand intérêt pharmaceutique comme l'insuline est très faible. Ce qui rend leur extraction et purification extrêmement couteuse (onéreuse). De plus, même si ces protéines peuvent être produites par des cultures cellulaires; c'est une vois beaucoup plus couteuse et difficile que la production par une culture microbienne à fort rendement. En conséquence l'industrie de la biotechnologie à mis au point des microorganismes génétiquement modifiés pour produire ces protéines.

La forme active de l'insuline est composée de deux polypeptides (A et B) reliés par des ponts disulfures, ces parties sont codées par des parties séparées du gène de l'insuline.

Préproinsuline: Polypeptide contenant:

- Une séquence signale impliquée dans l'excrétion de la protéine.
- Les polypeptides A et B de l'insuline active.
- Un polypeptide de connexion absent dans l'insuline active



Pour produire l'insuline utilisant *E. coli* comme cellule hôte. Le gène entier a pu être synthétisée, car cette hormone est une protéine relativement petite, cela permet d'éviter l'obstacle des codons utilisés par des espèces différents (ARNt différents). Le gène obtenu est

inséré en aval d'un promoteur fort d'*E. colide* telle sorte que le fragment d'insuline soit synthétisé comme partie d'une protéine de fusion.

Pour séparer l'insuline de la protéine de fusion, le gène codant l'insuline est séparé à celui codant la protéine de fusion par le codon ATG (MET), car l'insuline ne contient pas de méthionine. Cela permet de séparer les deux protéines utilisant le bromure de cyanogène, sans endommager la proinsuline.

L'insuline est obtenu a partir de la proinsuline par la formation des ponts disulfure. La proinsuline se replie naturellement en amenant face à face les résiduels cystéiniques impliqués dans la formation de ces ponts. Puis l'élimination du peptide de connexion par deux protéases (la trypsine et la carboxy peptidase B) n'ayant aucun effet sur le produit final.

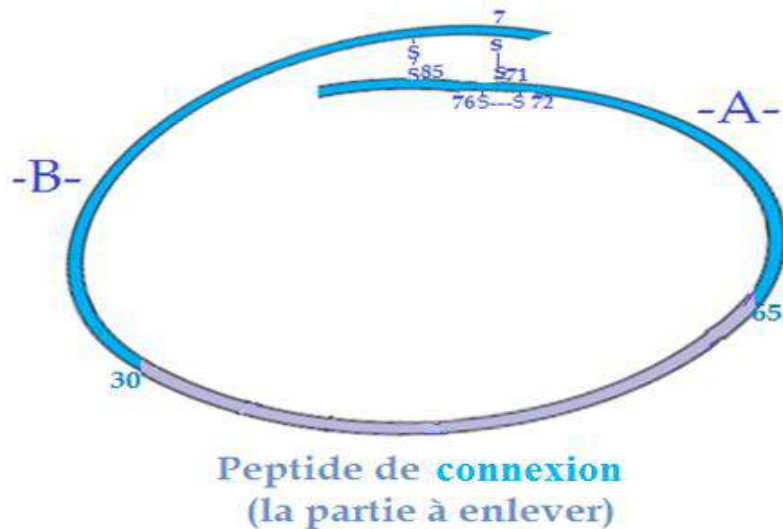


Fig.14: La proinsuline.

N.B. On peut utiliser *Saccharomyces cervisiae* pour produire cette hormone.

Génie Génétique

Chapitre IV

Techniques de Biologie moléculaire

L'étude des processus génétiques nécessite un grand nombre de techniques expérimentales. Cela consiste à utiliser des méthodes pour l'extraction, la purification, la séparation, restriction, l'amplification, le séquençage ...etc. des acides nucléiques.

Cette partie du polycopié est divisée sur six sujets. Le premier est consacré à l'extraction et la purification de l'ADN jusqu'au calcul de la concentration et la pureté. Le deuxième sur la visualisation de l'ADN sur gel d'agarose et la détermination du PM. Le troisième sur l'hybridation du principe jusqu'au la détection des clones spécifiques. Le quatrième sur l'amplification de l'ADN par PCR. Le cinquième sur le séquençage de l'ADN et en fin la synthèse de l'ADNc.

1- Extraction et purification des acides nucléiques

Toute étude de génétique moléculaire implique la disposition d'échantillon d'acides nucléiques. Les techniques d'extraction d'acides nucléiques sont relativement simples. Il convient simplement d'éviter toute destruction enzymatique ou mécanique. En effet les acides nucléiques qui sont stables dans la cellule intacte, deviennent très vulnérables à la digestion par les nucléases endogènes une fois la cellule lysée.

1-1- Source d'ADN à cloner

Un fragment d'ADN spécifique (Ex. gène) est récupéré par différents façon (PCR, RT-PCR, extraction...etc.) et de différentes sources (cellules nucléés, virus). La mappe au dessous montre les différents façons d'obtenir l'ADN, afin de l'utiliser dans le clonage moléculaire (fig.1).

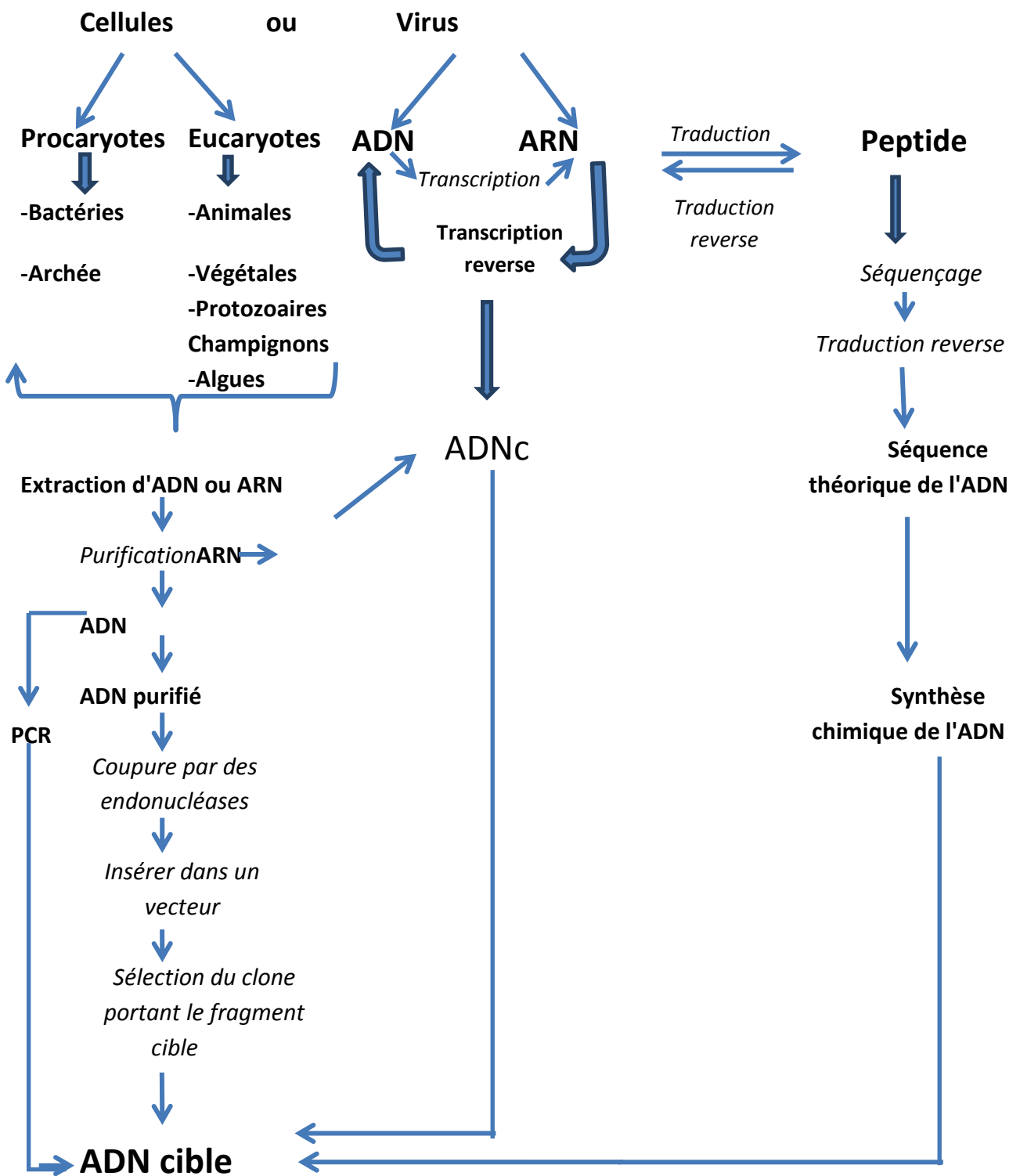


Fig.1: Carte de conception: Sources d'ADN

1-2- Préparation de l'ADN à partir du sang total

Les globules blancs (lymphocytes, macrophages...etc.) représentent la source majeure d'ADN en médecine. A partir de 10 à 30 ml de sang en présence d'un anticoagulant (ex. EDTA), il est possible de récupérer une quantité d'ADN ($\approx 100\mu\text{g}$) suffisante, pour des études qualitatives et quantitatives.

1-2-1- Protocole expérimental

- 1- Prélèvement de 10 à 30 ml de sang dans un tube contenant un anticoagulant (ex. EDTA).
- 2- Éclatement des globules rouges par une solution hypotonique
- 3- Récupération des globules blancs par centrifugation.
- 4- Préparation du lysat cellulaire en utilisant la solution de lyse (détergent comme le SDS ou sarcosyl + Protéinase K). Cette étape permet la libération de l'ADN nucléaire dans le milieu et la digestion des protéines qui lui étaient associées.
- 5- Extraction phénolique : cette étape permet de récupérer l'ADN dans la phase aqueuse après la centrifugation, car le phénol est un déprotéinisant puissant dans lequel les acides nucléiques ne sont pas solubles.
- 6-Extraction au chloroforme ou à l'éther : cette étape complète toujours une extraction phénolique, car elle permet d'éliminer les traces de phénol qui aurait pu être importé par la phase aqueuse (le phénol peut inhiber les enzymes utilisées dans l'amplification, la restriction...etc.).
- 7- Précipitation : cette étape permet la précipitation de l'ADN pour faciliter sa récupération. Il y'a deux types largement utilisés :
 - a-Précipitation à l'alcool éthylique : Elle doit être effectuée à haute force ionique (2.5 volumes d'éthanol 95° contre un volume d'échantillon). Pour les faibles concentrations, il faut prolonger le temps de précipitation ($> 10\text{h}$), comme il est possible d'accélérer la précipitation par le froid (-20°C à -70°C). L'ADN est récupéré par centrifugation.
 - b-Précipitation à l'isopropanol : Le principe est le même que la précipitation éthanolique. Ce type de précipitation se fait volume à volume. Mais deux caractéristiques majeures la différencient de la précipitation éthanolique.
 - Le sel n'est pas nécessaire
 - Les très petits fragments de DNA même à hautes concentrations ne sont pas précipités, ce qui permet de les éliminer.
- 8- Lavage : l'utilisation de l'éthanol 70° permet d'éliminer les sels, et les traces de l'isopropanol. L'ADN est récupéré par centrifugation.
- 9- Séchage: cette étape permet l'élimination de l'éthanol (s'évapore).
- 10- Resuspendre dans 10mM tris EDTA, l'eau purifiée de nucléases...etc.

1-2-2- Dosage des acides nucléiques

Les pyrimidines et les purines absorbent fortement les UV à 260nm. Une unité de densité optique à 260 nm correspond à :

- Une solution de DNA double brin à $50\mu\text{g/ml}$
- Une solution de DNA simple brin ou RNA à $25\mu\text{g/ml}$

$$C = A_{260} * DF * 100$$

Ces valeurs s'appliquent à des acides nucléiques parfaitement purs et en solution homogène. Pour vérifier la pureté de l'ADN il faut calculer :

$$P \text{ (pureté)} = A_{260} / A_{280} \text{ Une solution d'ADN est considérée pure si : } 1.7 \leq P \leq 2$$

2- Electrophorèse

2-1- Introduction

Les plus grands progrès des dernières décennies en biologie moléculaire ont été obtenus par l'analyse et la manipulation des macromolécules, l'ADN en particulier. Bien que la séquence des acides nucléiques détermine ses fonctions biologiques, la longueur de ces polymères constitue leur propriété physique la plus utile.

Parmi les techniques les plus utilisées pour la séparation des acides nucléiques selon leur taille est l'électrophorèse.

2-2-Définition

C'est une technique de bioséparation permettant la séparation des biomolécules selon leur charge et leur taille (ADN, ARN, protéines). Le terme électrophorèse vient de : électroce qui signifie énergie électrique et de phoresis (photo en grec) qui veut dire porter, avoir en soi.

Les acides nucléiques en solution portent une charge négative à cause de l'ionisation de leurs groupements phosphate (macromolécules polyanioniques uniformément chargées). Ils se déplacent donc vers l'électrode (+). Toutes les séquences d'acides nucléiques ont un rapport charge/masse à peu près identique, quelle que soit leur longueur, parce que tous les nucléotides sont à peu près de même masse (masse molaire moléculaire moyenne des nucléotides 330 g/mol).

Suivant les théories de l'électrophorèse la mobilité μ dans un champ électrique au sein d'un gel doué de pouvoir de filtration est:

$$\text{Log } \mu = \text{Log } \mu_0 - Kr \cdot C$$

Log μ_0 : La mobilité de la molécule en milieu liquide

C: Concentration du gel

Kr: Coefficient de retardement dû au gel

(Lui même est en fonction de la masse moléculaire de la molécule).

Deux sortes de matrice de gel sont utilisées: l'agarose et le polyacrylamide. Le gel agit à la manière d'un tamis moléculaire au travers duquel les molécules d'ADN en mouvement doivent passer; les molécules de grande taille ont plus de difficulté pour passer à travers les mailles (pores) créés par le réseau des microfibrilles du gel et vont donc migrer plus lentement que les petites molécules. La vitesse de migration de l'ADN est influencée par deux paramètres :

- La masse moléculaire (nombre de pb)
- La concentration du gel (tab. 1).

Il faut noter que la nature et la concentration du support de l'électrophorèse sont en fonction de la taille des fragments à séparer.

Tab.1: Taille des fragments à séparer selon la concentration du gel.

| % d'agarose | Taille de fragments à séparer en kb | % d'acrylamide | Taille de fragments à séparer en pb |
|-------------|-------------------------------------|----------------|-------------------------------------|
| 0.5 | 1 à 30 | 4 | 200 à 800 |
| 0.7 | 0.8 à 12 | 5 | 80 à 200 |
| 1 | 0.5 à 10 | 8 | 40 à 100 |
| 1.2 | 0.4 à 7 | 11 | 10 à 50 |
| 1.5 | 0.2 à 3 | | |

Dans toutes les électrophorèses, des marqueurs de taille (ou de poids moléculaire) sont déposés et migrent parallèlement au DNA étudié. Une fois l'électrophorèse terminée, les molécules fluorophores, comme le bromure d'éthidium, qui se fixe à l'ADN en s'intercalant entre les bases. Les bandes d'ADN (regroupent toute les molécules d'ADN de taille identique) sont visualisées sous UV (fig. 2).

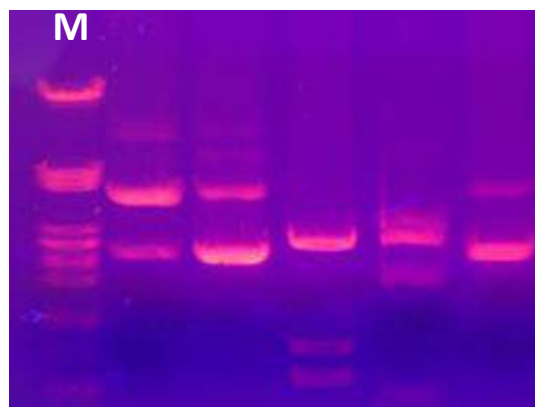


Fig.2: Visualisation des bandes d'ADN sous UV. **M:** marqueur de taille.

2-3- Détermination de la taille des fragments

La mobilité relative est le rapport entre la distance de migration d'une bande et la distance de migration du front de migration. La droite $\log(\text{nombre de kb}) = f(\text{mobilité relative})$, permet de déterminer la taille d'une molécule d'ADN inconnue (fig. 3).

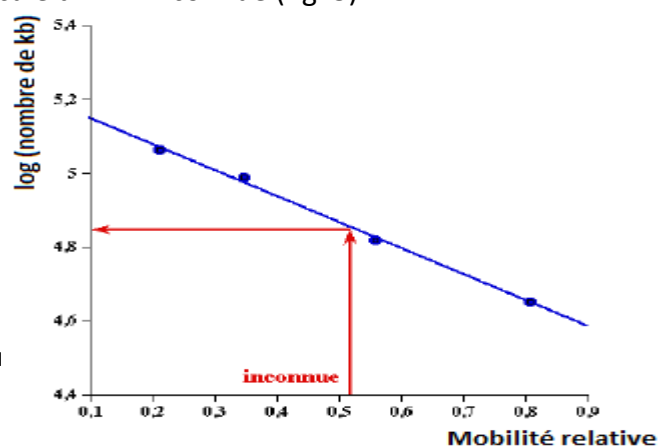


Fig.3: Distance de migration en fonction du logarithme de la taille des fragments d'ADN.

Exemples de marqueurs de taille

- Phage λ coupé par *Bst*EII (0.12→8.45)
- Phage λ coupé par *Hind*III (0.13→23.13)
- pBR322 coupé par *Bst*NI (0.12 → 1.86)
- 100 pb Ladder (100→1000)

Les topoisomères d'ADN (ont la même taille, mais dont les degrés de liaison sont différents) migrent différemment (fig. 4).

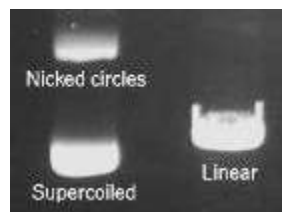


Fig.4: Migration des topoisomères d'ADN (circulaire relâché; linéaire, et surenroulé) sur gel d'agarose.

3- Hybridation des acides nucléiques

L'hybridation est une propriété fondamentale des acides nucléiques qui repose sur les règles de complémentarité. Il est possible d'apparier des brins d'ADN ou ARN avec des oligonucléotides qui reconnaissent spécifiquement des séquences sur les brins d'ADN de manière antiparallèle et complémentaire. Ces oligonucléotides sont appelés sondes nucléiques (fig. 5). Pour le cas de deux brins complémentaires de la même molécule d'ADN (brins homologues) on parle de renaturation (hybridation à 100%). Chaque 1% de non homologie diminue la température d'hybridation (T_h) de 1°C.

L'hybridation moléculaire est utilisée surtout dans la détection de l'homologie entre les molécules d'ADN de sources différentes. La complémentarité dépendra de la spécificité et de la sensibilité.

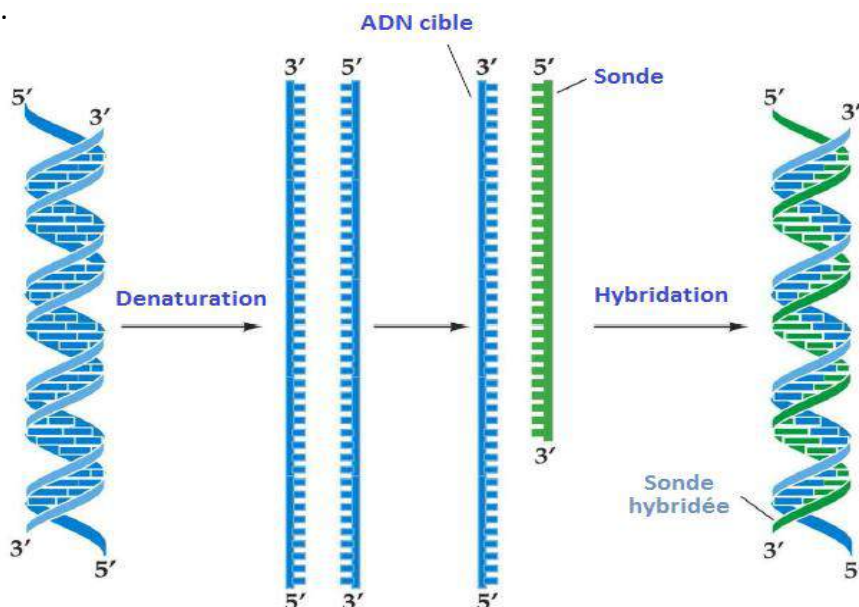


Fig.5: Hybridation des acides nucléiques (Passarge, 2007).

La Th est estimée en utilisant plusieurs formules, cela dépend de la:

- Longueur du fragment (nombres de nucléotides).
- Composition du milieu
- Mis-appariement
- Le contenu en C+G

3-1-Formules utilisées pour calculer la Th

Formule de Wallace: $T_m = 4(C+G) + 2(A+T)$; $T_h = T_m - 5$

La formule de Wallace permet de déduire la Th des petits oligonucléotides (Ex. Les amorces).

Pour les oligonucléotides avec $N < 100$ nucléotides on utilise la formule suivante en tenant compte la concentration du sel, les mésappariements et la concentration de formamide:

$T_m = 16.6\log[Na^+] + 0.41\%(C+G) + 81.5 - (675/N) - \%mismatch - 0.65\% \text{ de formamide}$

Dans les conditions réactionnelles standards la T_m pour les longs séquences (>100):

$T_m = 69.3 + 0.41\%(C+G)$

La formule globale:

$T_m = 81.5 + 16.6\log[\text{sel}] + 0.41\%(C+G) - \% \text{ mis-appariement} - (500/N) - 0.65\%(\text{formamide}).$

3-2 Facteurs influençant l'hybridation

- La concentration de l'ADN et le temps d'hybridation
- La température d'hybridation
- La force ionique
- La complexité des séquences

3-3-Types d'hybridation

L'objectif de l'hybridation est la détection de la présence d'un acide nucléique d'une séquence donnée par l'utilisation d'un fragment d'ADN complémentaire = sonde. Cela peut avoir lieu en solution ou sur support solide (immobilisation de la cible sur une membrane[nitrocellulose, nylon], sur verre, colonies bactériennes, chromosomes, plage de lyse ...etc.).

3-3-1-Hybridation d'ADN sur support solide : Southern blot

La procédure de ce type d'hybridation est résumée en sept étapes (fig. 6)

- 1- Extraction de l'ADN
- 2- Digestion enzymatique (par les enzymes de restriction)
- 3- Electrophorèse du produit de la digestion
- 4- Transfert sur membrane
- 5- Hybridation avec une sonde spécifique marquée.
- 6- Lavages

7- Révélation

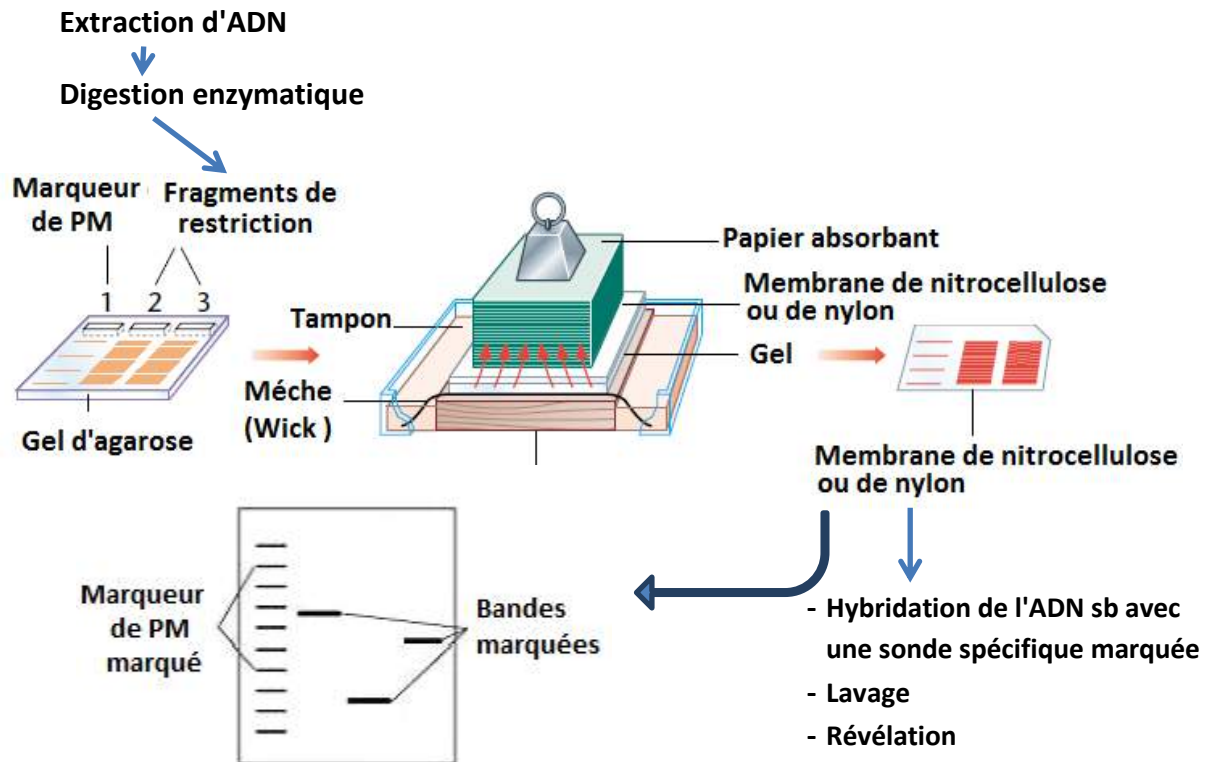


Fig.6: Southern Blot.

On peut aussi révéler la présence des séquences spécifiques dans les ARN (Northern Blot) et des protéines (Western Blot) utilisant le même principe.

3-3-2-Hybridation *in situ* sur chromosome

Cette technique permet de déterminer sur quel chromosome et dans quelle région se trouve le gène d'intérêt. Les chromosomes fixés sur lame sont hybridés avec une sonde marquée (H^3). L'examen microscopique révèle les signaux positifs sous forme de grains noirs disposés sur le chromosome. Comme exemple d'application est la détection des gènes viraux intégrés dans les chromosomes. Examine

3-3-3-Hybridation sur colonie de bactéries, sur plage de lyse

Ce type d'hybridation permet la détection parmi un grand nombre de bactéries ou de phages recombinants (plage de lyse) celle ou celui qui contient le fragment d'ADN cible. La réalisation de cette technique passe par les étapes suivantes:

- Culture pure sur boîte de Petri (colonies ou plage de lyse)
- Transfert sur membrane de nylon ou de nitrocellulose
- Lyse alcaline des bactéries et dénaturation de l'ADN
- Fixation par la chaleur ou les UV
- Hybridation avec une sonde marquée
- Lavage
- Autoradiographie

- Localisation des clones d'intérêt.

4- Réaction de Polymérisation en Chaîne "PCR"

La synthèse et le séquençage d'ADN ont permis l'émergence d'une méthode d'amplification de l'ADN appelée PCR (**P**olymerase **C**hain **R**eaction), à partir d'un gène (ou fragment) spécifique, de grand quantité d'ADN sont obtenues *in vitro*. L'ADN polymérase copiant jusqu'à un milliard de fois cette région cible « quantité suffisante pour être révélée ».

4-1-Historique

Cette méthode de Biologie Moléculaire a été mise au point en 1985 par Kary Mullis, qui obtint pour ces travaux le prix Nobel de Chimie en 1993.

Aujourd'hui, ce procédé révolutionnaire couplé à l'utilisation d'une ADN polymérase thermorésistante permet d'obtenir, sans clonage, une amplification considérable d'un fragment donné d'ADN.

4-2- Etapes de PCR :

La dénaturation de la double hélice nécessite une haute température (autour de 95°C), donc l'enzyme utilisée doit être résistante aux hautes températures ou bien elle devait être ajoutée lors de chaque cycle. Ce problème a été résolu en utilisant une ADN polymérase thermiquement stable d'une bactérie thermophile : *Thermus aquaticus*, isolée d'une source chaude. Cette enzyme nommée ADN Taq polymérase est stable à 95°C (tab. 2) et n'est pas affectée par l'étape de dénaturation. Avec l'utilisation du Taq polymérase, l'ADN étant copié à 72°C et non plus 37°C, son utilisation augmente la spécificité de la PCR. Car aux hauts températures l'hybridation non spécifique d'amorces à l'ADN **non** cible est rare. Les produits de l'ADN Taq polymérase sont donc plus homogènes que ceux obtenus avec l'ADN polymérase III d'*E. coli*. En résumé il y'a quatre étapes :

- 1- **Dénaturation** (autour de 95°C) : Sert à dénaturer l'ADN pour obtenir des matrices simple brin.
- 2- **Hybridation** (autour de T_h): Lors du refroidissement du mélange, les amorces étant en excès, la plupart des brins d'ADN cibles se fixent à celles et non entre eux.
- 3- **Extension** (72°C) : L'ADN polymérase « prolonge les amorces : côté 3'-OH libre » en utilisant les brins cibles comme matrice.
- 4- Après une période appropriée d'incubation, le mélange est chauffé de nouveau pour séparer les brins. Il est alors refroidi pour que les amorces s'hybrident aux régions complémentaires de l'ADN nouvellement synthétisé. Le processus complet est alors répété « cycle ».

Tab.2: Les différences entre ADN Taq Polymérase et l'ADN polymérase III.

| | ADN Taq polymérase | ADN polymérase III |
|----------------------------|--------------------|--------------------|
| Stabilité thermique à 95°C | +++ | - |
| Activité exonucléase | - | + (5' → 3') |

L'ADN polymérase de l'hyper thermophile *Pyrococcus furiosus* (croissance optimale à 100°C, la bactérie a été isolée d'une source hydrothermale), appeler Pfu polymérase. C'est la meilleure polymérase thermostable pour la correction des erreurs de réplication

(Fidélité élevée) grâce à son activité auto-correctrice. Cette enzyme est très utilisée lorsqu'une haute précision est requise. De plus elle est plus stable que la Taq polymérase.

4-3 Thermocycleur

Les trois étapes, constituant un cycle de PCR, sont effectuées à des températures différentes permettant de contrôler l'activité enzymatique. Pour effectuer ces transitions de températures, les microtubes contenant le mélange réactionnel sont placés dans un appareil programmable : un thermocycleur (fig. 7). Cet appareil permet d'exposer les tubes à des températures choisies et pour des durées déterminées par l'expérimentateur (effectuer automatiquement les cycles de chauffage) (fig. 8).

Le secret de la PCR est que le produit d'un cycle d'extension sert de matrice pour le suivant (fig. 8), ainsi la technique PCR est remarquable par le fait que l'ADN cible se double à chaque cycle. En pratique 20 à 40 cycles sont habituellement réalisés, ils génèrent 10^6 à 10^9 fois la séquence cible. « Elle repose sur la succession de plusieurs cycles » (fig.9).

La progression géométrique de raison 2 (tab. 3) permet d'établir une relation pour calculer le nombre des copies totale et cibles.

Tab. 3: Progression géométrique de raison 2 des cycles de PCR et nombre de copies.

| Cycle | 1 | 2 | 3 | 4 | 5 | 6 | -----n | n : nombre de cycle |
|----------------------|---|---|---|----|----|----|--------|---------------------|
| Copies totale (N) | 2 | 4 | 8 | 16 | 32 | 64 | ----- | $N=2^n$ |
| Copies parasites | 2 | 4 | 6 | 8 | 10 | 12 | ----- | $2n$ |
| Copies cibles | 0 | 0 | 2 | 8 | 22 | 52 | ----- | $2^n - 2n$ |

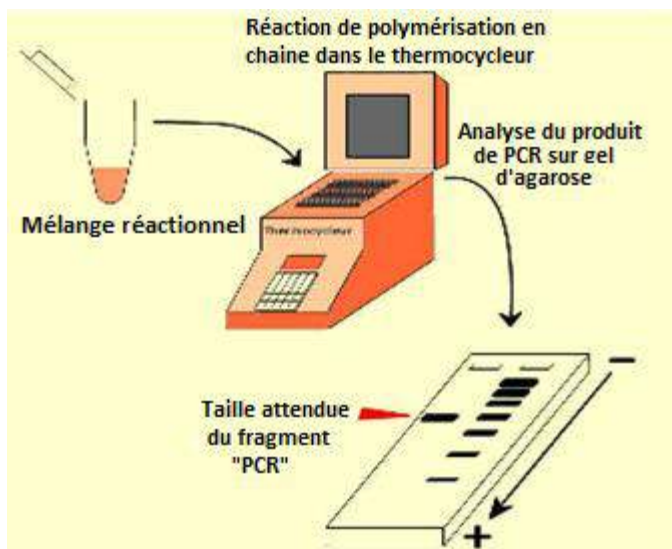


Fig.7: Réaction d'amplification d'ADN dans le thermocycleur et visualisation du produit de PCR par électrophorèse.

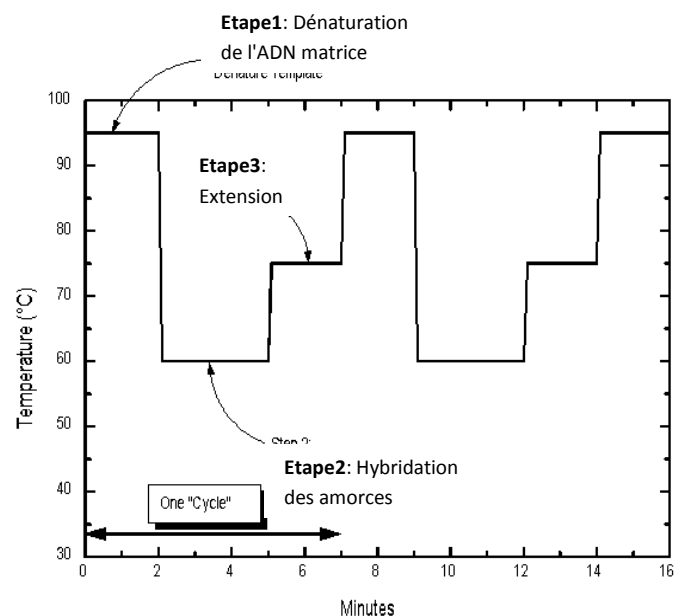


Fig.8: Profile de cycles de la température de la PCR

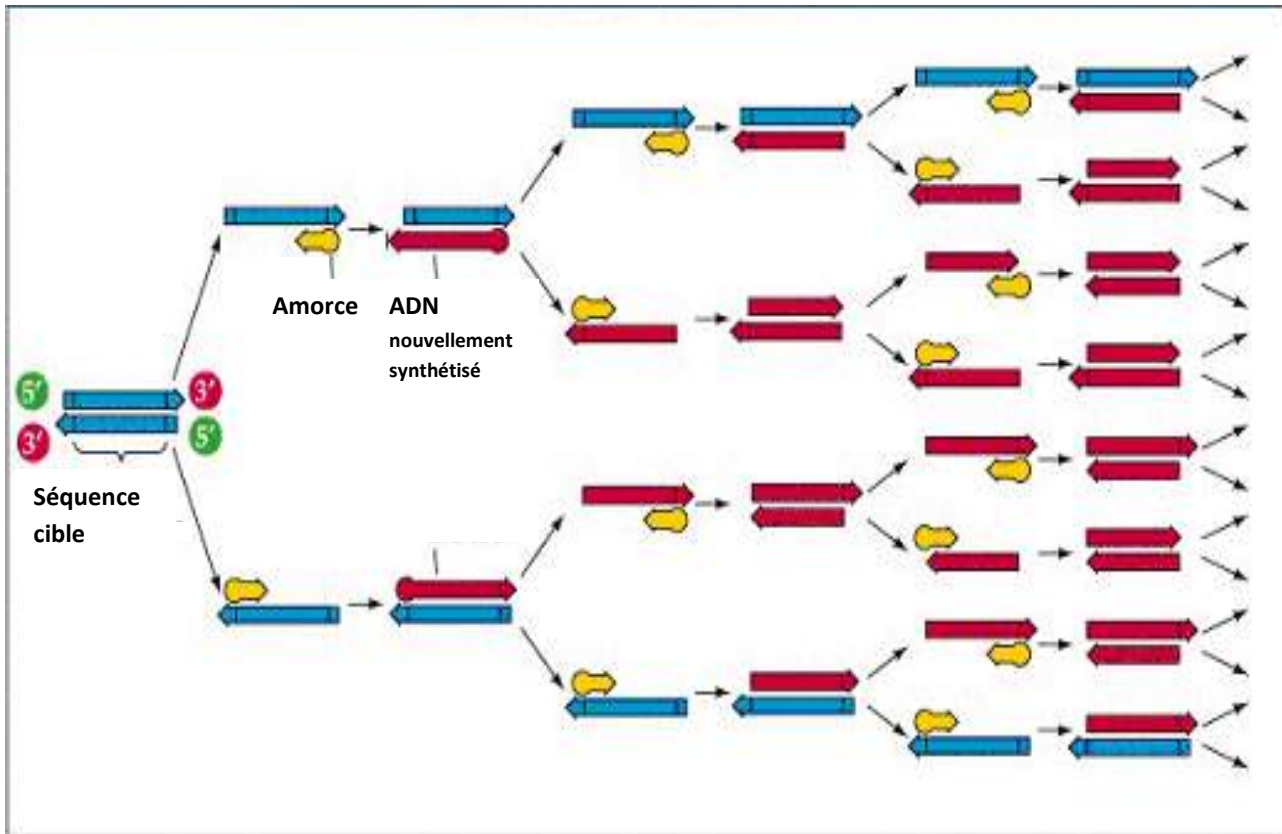


Fig.9: Amplification d'ADN par PCR. Le produit d'un cycle sert de matrice pour le cycle qui suit.

Le désigne des amorces (primers) repose sur les critères suivants:

- Taille entre 17 à 31 nucléotides
- Contenu GC : $\approx 50\%$
- T_h proches entre les deux amorces (F et R).
- Absence de répétition d'un même nucléotide
- Absence de complémentarité entre les deux amorces (pas de risque de formation de dimères)
- Faible risque de formation des structures secondaires (Ex. Epingle à cheveux).

Pour calculer le T_h d'une amorce on utilise la formule suivante:

$$T_h = 4(C+G) + 2(A+T) - 5$$

4-4-Applications et sensibilité de la PCR

La PCR est un outil puissant, facile d'utilisation, hautement efficace, extrêmement sensible et spécifique, lors de chaque cycle, la quantité d'ADN étant multiplié par deux (augmentation exponentielle).

La connaissance de séquence de bordures entourant un gène d'intérêt est importante, les produits d'amplification obtenus par PCR peuvent être utilisés pour le clonage et le séquençage de ces gènes, ou pour des études comparatives ou phylogénétiques. Dans ces deux derniers cas, les amorces sont spécifiques à des régions conservées d'un gène commun d'une grande variété d'organismes. Par Exemple : le gène qui code pour l'ARNr 16S a des régions hautement conservées encadrant une région hautement variables. La PCR est utilisée aussi dans :

- La mutagénèse dirigée
- Les études paléontologiques (fossiles : momies, plantes, animaux, ...etc.)
- Le diagnostique microbiologique
- Le diagnostique moléculaire des maladies génétiques et infectieuses (surtout les cancers et les microorganismes difficilement cultivés Ex. *Bartonella*, virus ...etc.).

N.B. La présence d'un gène spécifique indique la présence vraie semblable de l'agent causative. Ainsi, la PCR évite de devoir cultiver l'organisme, une procédure souvent longue et onéreuse.

5- Séquençage

Le séquençage est une technique d'analyse de l'ADN, il permet de déterminer l'ordre des nucléotides. Actuellement la technique de séquençage repose majoritairement sur la méthode enzymatique de Sanger.

5-1- Méthode de Sanger

Cette méthode génère des fragments d'ADN terminés par l'une des quatre bases marquées au préalable, car l'élongation de la chaîne s'arrêtera après l'incorporation d'un didésoynucléotide (ddNTP) (fig. 10). C'est le principe sur lequel baser cette méthode.

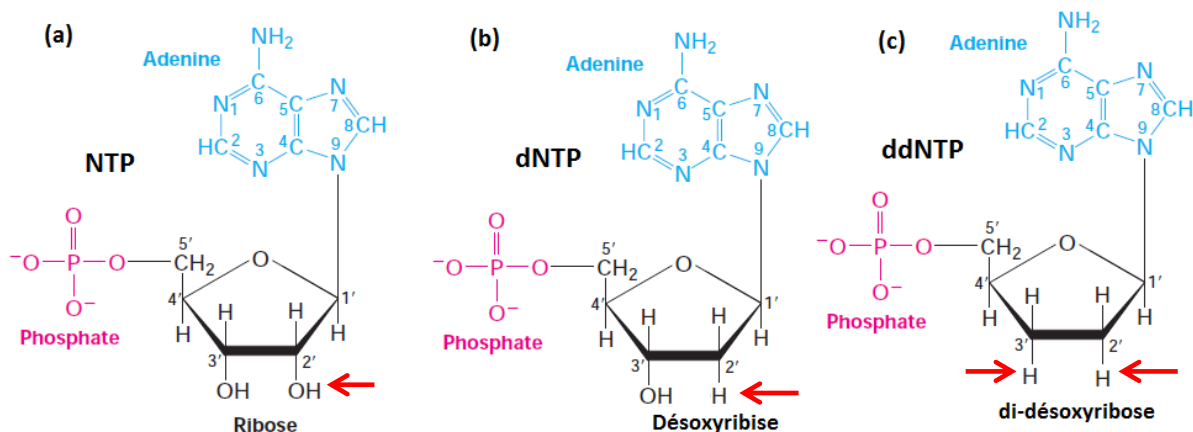


Fig.10: Les différentes formes des nucléotides.

La méthode de Sanger permet de déterminer une séquence inconnue par l'intermédiaire d'une copie d'ADN simple brin. Réalisé grâce à l'ADN polymérase et une amorce spécifique. Pour mieux comprendre le principe de séquençage, on utilise des ddNTP non marqués. Alors on utilise quatre tubes séparés contenant l'un des ddNTPs (ddATP, ddCTP, ddGTP, ddTTP). Chaque tube contient: ADN polymérase (Ex. Pfu Polymérase); amorce (F ou R), dNTPs (dATP, dTTP, dCTP, dGTP); eau purifiée de nucléases (WFN), $MgCl_2$ et l'ADN à séquencer.

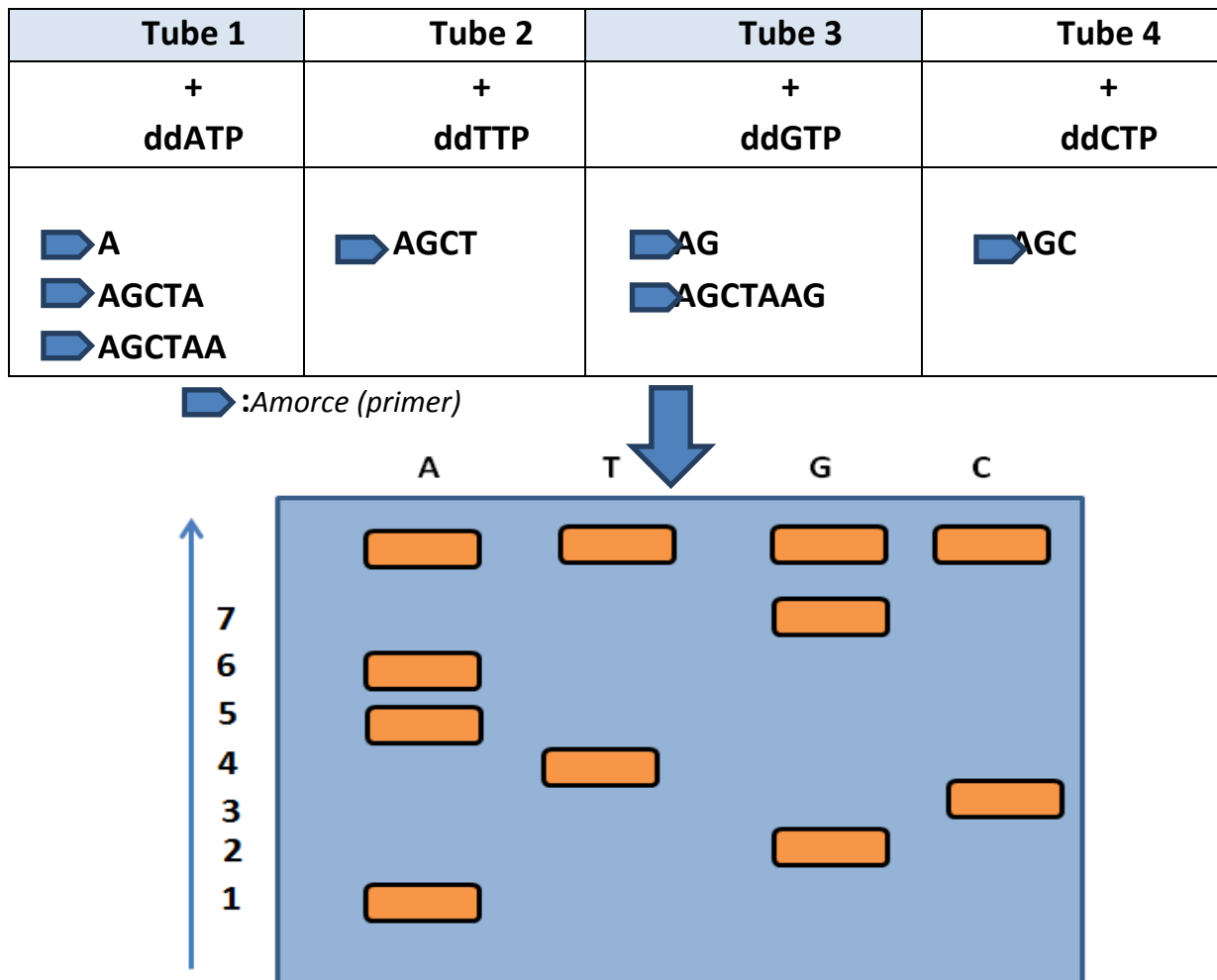


Fig.10: Séquençage de l'ADN par la méthode de Sanger

La détermination de l'enchaînement des nucléotides dans un fragment d'ADN simple brin par la méthode de Sanger comporte deux étapes:

- 1- Synthèse du brin complémentaire d'un brin matrice à l'aide d'une polymérase à partir d'une amorce, en présence des 4 dNTPs et d'un ddNTP qui joue le rôle de terminateur de chaîne (absence de 3'-OH). Quatre réactions en parallèle sont nécessaires, chacune avec l'un des terminateurs présents en faible quantité.
- 2- Séparation par électrophorèse (PAGE) des fragments synthétisés, le nombre de ces fragments dépend du nombre de résidus nucléotidiques. Donc la position des terminateurs dans la séquence (fig. 10).

5-2- Marquage des produits d'extension

Dans le séquençage automatique, des marqueurs fluorescents sont utilisés, un marqueur différent pour chaque ddNTP (bleu, rouge, vert, orange...etc.). La polymérisation se fait dans un seul tube contenant le mélange réactionnel et les quatre ddNTPs.

5-3- Séquençage d'un produit de PCR

Avant de séquencer un produit de PCR il faut le purifier en éliminant les composants suivants:

- Les dNTPs (car le rapport dNTP/ddNTP est important)
- Les amorces (on besoin d'une seule amorce F ou R pour amplifier un seul brin)
- Les sels (pour éviter l'inhibition de la polymérase lors du séquençage)
- Les produits de PCR non spécifiques (pour éviter les chevauchements de séquences: séquence artefacts).

Généralement on utilise des kits de purification permettant d'éliminer tous les composants non essentiels mêmes des produits non spécifiques < 100 pb par des colonnes de chromatographie (filtration sur gel). Mais les inconvénients de cette méthode qu'elle est relativement coûteuse et n'élimine pas les produit non-spécifique > 100pb.

S'il y'a des produits de PCR non spécifiques > 100 pb la purification se fait sur gel d'agarose (utilisant low melting agarose), seulement la bande spécifique qui est récupérée, puis purifiée et séquencée. Les inconvénients de la purification sur gel qu'elle est très longue, coûteuse et donne de faibles rendements et le risque des mutations lors de la révélation des bandes sous UV en utilisant un agent intercalent.

N.B. Il y'a des enzymes permettant d'éliminer les dNTPs (ex. phosphatase) et les amorces (Ex. nucléase agissant sur l'ADN simple brin). Mais le coût élevé limite leur utilisation.

6- ADN C " ADN complémentaire"

6-1- Caractéristiques de l'ARNm eucaryote

Chez les eucaryotes la région codante peut être interrompue par 1 ou plus d'introns. Donc si on utilise l'ARNm on peut réaliser les avantages suivants :

- 1- Les introns sont éliminés par épissage (splicing).
- 2- Les tissus spécifiques permettent d'avoir des quantités considérables d'ARNm (plus de copies par rapport au gène d'intérêt).

Il faut noter que l'ARNm chez les eucaryotes est caractérisé par la queue poly (A), donc on peut utiliser poly (T) (12 à 18) comme amorce pour des ARNm eucaryotes différents et aussi pour leur extraction (fie.11).

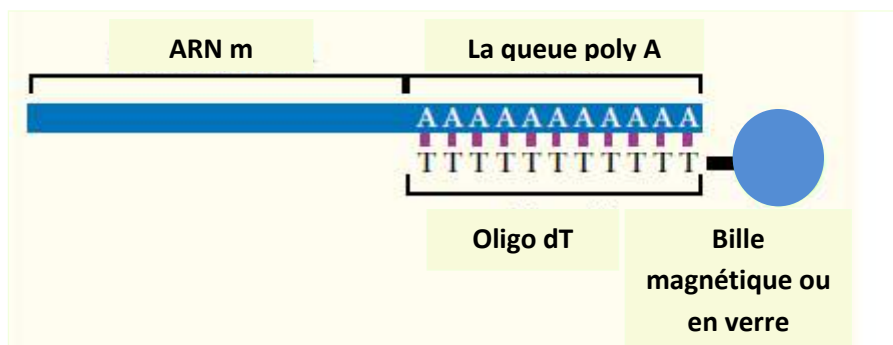


Fig.11: L'utilisation d'une queue poly (T) fixée sur un support pour l'extraction des ARNm eucaryotes.

6-2- Synthèse de l'ADNc

La synthèse d'ADNc est catalysée par des transcriptases inverses (RT). Ces enzymes sont des ADN polymérase ARN dépendantes, capables d'utiliser un brin d'ARN comme matrice pour catalyser la synthèse d'un brin d'ADN complémentaire. Cela correspond effectivement à l'«inverse» d'une réaction de transcription de l'ADN en ARN. Les transcriptases inverses sont issues de rétrovirus dont elles sont une des principales caractéristiques. Par exemples, la transcriptase inverse du Virus du Myéloblastome Aviaire (AMV), celle du Virus de la Leucémie Murine (Mo-MLV [Molony Murine Leukemia Virus] ou Mu-LV).

Comme toutes les ADN polymérase, les transcriptases inverses ne peuvent pas initier seules la synthèse d'un brin d'ADN. Elles ont besoin d'une amorce possédant une extrémité 3'-OH libre. Lorsque les ARN à amplifier sont polyadénylés en 3' (ARNm eucaryotes par exemple), l'amorce choisie peut être simplement une séquence polyT (12 à 18) constituée d'une succession de désoxythymidines (comme sur le schéma ci-dessous en rouge). Dans ce cas, tous les ARNm sont a priori copiés en ADNc.

La ribonucléase H ou RNase H (E.C.3.1.26.4) est une enzyme qui hydrolyse le brin d'ARN dans un double brin hybride ADN:ARN. Elle libère des extrémités 3'-OH et 5'-phosphate. La RNase H n'hydrolyse pas l'ARN simple ou double brin (fig.16).

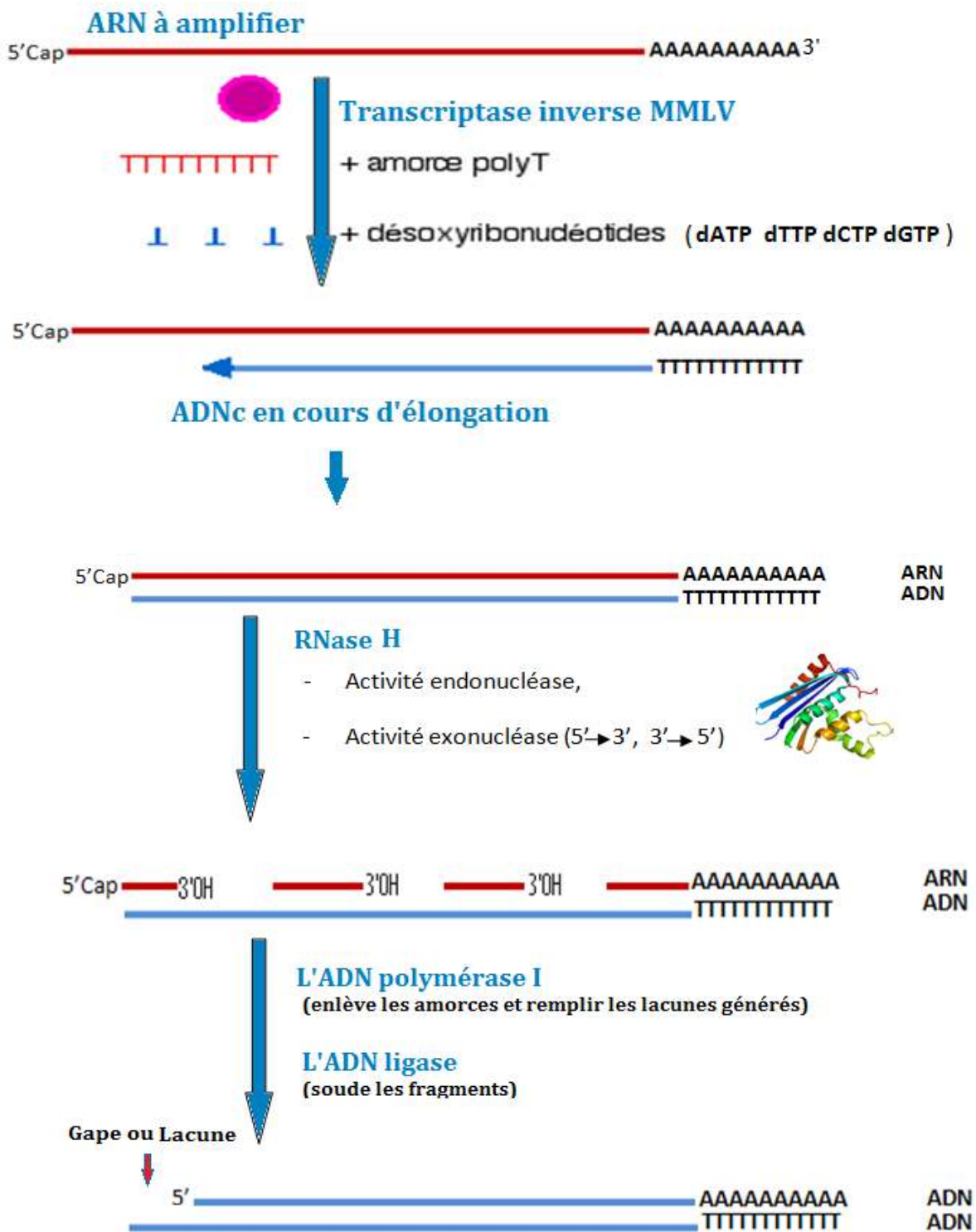


Fig.16: Les étapes de la synthèse d'ADNc. **N.B.** Il reste une GAP ou bien résidu 5' Cap-ARN.

6-3-Insertion de l'ADNc dans un vecteur approprié

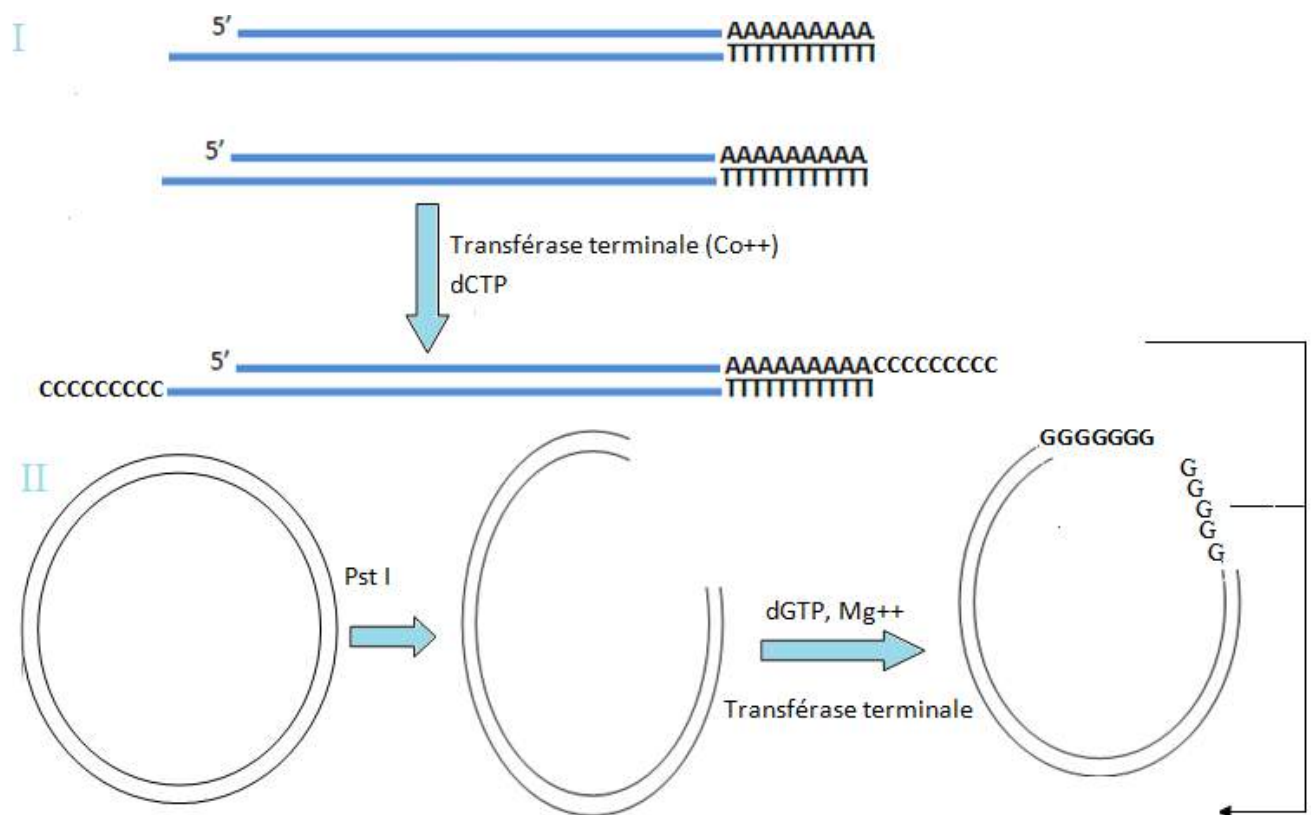
Il y'a deux méthodes d'insertion de l'ADNc dans les vecteurs:

1- Addition d'une queue homopolymérique.

2- Addition d'un joint (linker).

L'addition d'une queue homopolymérique se fait par une ADN polymérase inhabituelle trouvée chez des types de cellules eucaryotes appelées « pré-lymphocytes ». Cette enzyme est appelée transférase terminale, elle est capable d'ajouter un désoxynucléotide dans la partie 3'-OH d'un brin d'ADN (ne nécessite pas d'une amorce et d'une matrice).

À la présence de cations divalents (Mg^{++} , Co^{++}), l'enzyme catalyse l'addition de dNTPs (Purines : Mg^{++} ; Pyrimidines : Co^{++}) à l'extrémité 3' de l'ADN. Selon les conditions de réaction on ajoute trois à quelques milles nucléotides.



- Appariement entre les poly (G) et les poly (C).
- La lacune générée sera rempli par l'ADN Pol I.
- L'ADN ligase soude les fragments (formation de Liaisons phosphodiester)

Fig. 17: Insertion de l'ADNc dans un vecteur.

3^e partie:

Introduction à la Bioinformatique

Introduction à la bioinformatique

1- Introduction

La bioinformatique est une discipline relativement récente, son apparition, dans les années quatre vingt, coïncide avec la création des premières banques de données (Ex. GenBank, EMBL). Cette science concerne surtout l'étude et l'analyse "*in silico*" de l'information biologique contenue dans les séquences nucléotidiques et protéiques. A partir des années quatre vingt-dix, cette discipline devient indispensable avec l'accumulation des données de séquençage des protéines, des gènes d'identification comme ceux de l'ARN16S, des gènes de résistance, virulence et même de génomes complets.

Il est important pour le microbiologiste de comprendre les méthodes utilisées en bioinformatique, afin d'être en mesure d'appliquer avec succès les outils disponibles et interpréter correctement les résultats du laboratoire (Ex. L'identification, la caractérisation des mutations, les ORF, l'évolution, l'homologie...etc.). Parmi ces outils, celles appliqués à la comparaison des génomes bactériens comme l'alignement (Ex. BLAST), sont d'une importance cruciale.

Globalement, la bioinformatique propose des méthodes et des logiciels qui permettent:

- Le recueil, le stockage et la gestion des données biologiques et leur distribution à travers les réseaux.
- Les études phylogénétiques et l'évolution moléculaire des êtres vivants.
- Le développement des outils pour analyser les problèmes de biologie moléculaire.
- L'analyse, la comparaison et la prédiction de la structure des gènes.
- La modélisation et la prédiction de la structure et de la fonction des protéines.

2- Historique

1955-1965: Premiers langages informatiques, premier ordinateur commercial

1970: 1er programme pour la comparaison de séquences protéiques, Alignement optimal entre deux séquences (Needleman & Wunsch)

1971: PDB - Protein Data Bank (structures 3D macromolécules)

1978: Matrice de substitution (PAM) (Dayhoff *et al.*)

1981: Similarités de séquences dans les banques (Smith & Waterman)

1980/1986: Création de des banques de données: EMBL (1980), GenBank (1982), DDBJ (1986), SwissProt (1986)

1989: L'internet diffusé pour le publique

1990: BLAST – Simulation de séquence dans les banques.

2000: - Séquençage du 1er génome de plante, *Arabidopsis thaliana*

- Publication du "draft" de la première carte complète du génome Humain (3000 MB).

2001: Publication des travaux de séquençage du génome humain presque complet.

3- Banques de données biologiques

Les banques de données ou bases de données biologiques sont des bibliothèques contenant des données de séquences biologiques collectées grâce à des expériences scientifiques, largement diffusées par le réseau internet. Elles sont généralement reliées entre elles par des liens.

Tab.1: Exemples de quelques bases de données biologiques

| Base de données | Description | Site web (ou adresse internet) |
|---|---|---|
| Genbank | Banques de séquences nucléotidiques publiquement disponibles et leur traduction en protéines. | www.ncbi.nlm.nih.gov/enter (Acides nucléiques) |
| EMBL (European Molecular Biology Library). | | www.srs.ebi.ac.jp (Acides nucléiques) |
| NCBI (national center for biotechnology Information) | Portail américain de bioinformatique: BLAST, banques de séquences nucléotidiques et génomiques, PubMed. | www.ncbi.nlm.nih.gov (Portail) |
| Pubmed-Medline | Articles, publications en biologie et en médecine | www.ncbi.nlm.nih.gov (Bibliographie) |
| Google Scholar | Publication scientifiques avec google | http://scholar.google.com (Bibliographie) |
| Connectsciences | Articles, thèses | http://connectsciences.inist.fr (Bibliographie) |

4- Principes de bases de l'alignement des séquences

Durant les dernières décennies, des progrès énormes ont été réalisés dans trois domaines qui ont profondément affecté la classification des microorganismes. Il s'agit d'une part de l'étude détaillée de la structure cellulaire par microscopie électronique, d'autre part de la caractérisation physiologique et biochimique de nombreux microorganismes et enfin de la comparaison des séquences d'acides nucléiques et de protéines d'une grande variété de microorganismes. La diversité génétique est due à des mutations ponctuelles et à des insertions ou (et) délétions apparues au cours de l'évolution. L'alignement de séquences (globale, local ou multiple) de deux ou plusieurs microorganismes permet de déterminer leur degrés de similarité (deux organismes sont similaires si leur score de substitution est supérieur à 0), d'homologie pour montrer s'ils dérivent d'un ancêtre commun, et leur identité par estimation de la fraction de résidus identiques.

4-1- Alignement et détermination du score

L'alignement de séquences est une manière de représenter deux ou plusieurs séquences (nucléotides, acides aminés) les unes sous les autres, de manière à en faire ressortir les régions homologues ou similaires (les zones de concordance). Elle nous permet de révéler trois opérations d'édition: **1**) Substitution **2**) Délétion **3**) Insertion.

Les pénalités des brèches doivent être suffisamment coûteuses pour éviter les alignements sans signification biologique. La recherche de la similarité entre séquences nécessite la détermination d'un score de similarité.

Le score de l'alignement est la somme des scores élémentaires :

$$\text{Score} = \sum \text{scores élémentaires} - \sum \text{score pénalité.}$$

Exemple: Déterminer le score des séquences suivantes:

Séquence 1: **TTGACAAGCCTCG**
 ||| || |||
 Séquence 2: **TTCATGAGACATCG**

On a 8 appariements et six mésappariement (1 appariement "match": vaut +1 et 1 mésappariement "mismatch" vaut 0): **Score :8-0 = 8**

Mismatch Brèche "GAP"

Séquence 1: **ATGACTGGGCCACT**
 || ||||| |||
 Séquence 2: **ATCTGGGACAAC**

On a 12 appariements, 1 mésappariement, 2 brèches (1 brèche vaut -1):**Score:12-0-2 = 10**

4-2- Alignement global

Les alignements globaux (alignement des séquences sur la totalité de leur longueur en tenant compte de tous les résidus) sont plus souvent utilisés quand les séquences mises en jeu sont similaires et de tailles comparables. Une technique générale, appelée algorithme de Needleman-Wunsch et basée sur la programmation dynamique permet de réaliser des alignements globaux de manière optimale même pour les longues séquences.

4-2-1- Algorithme de Needleman et Wunsch

C'est la méthode la plus utilisée est connue qui réalise le meilleur alignement global entre deux séquences. Il permet d'évaluer le degré de similarité entre deux séquences connues (détermine le meilleur alignement évalué par un score). Pour cela il construit le chemin optimal allant du coin supérieur gauche au coin inférieur droit d'un tableau à deux dimensions. Pour déterminer la valeur d'une case on doit respecter les trois règles suivantes: Règle 1: chaque case va contenir un score; le score de l'alignement sera celui de la case en bas à droite. Règle 2: le score d'une case se déduit à partir de celui des cases au-dessus, à gauche ou en diagonale. Règle 3: un pas horizontal/vertical coûte 1 gap un pas diagonal coûte 1 position alignée (match ou mismatch). Ci-dessous un exemple de l'alignement de deux séquences:

Etape 1: Création d'un tableau indexé par les deux séquences.

| $j \backslash i$ | | A | C | G | G | C | T | A | T |
|------------------|-----|-----|----|----|----|-----|-----|-----|-----|
| | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 |
| A | -2 | 2 | 0 | -2 | -4 | -6 | -8 | -10 | -12 |
| C | -4 | 0 | 4 | 2 | 0 | -2 | -4 | -6 | -8 |
| T | -6 | -2 | 2 | 3 | 1 | -1 | 0 | -2 | -4 |
| G | -8 | -4 | 0 | 4 | 5 | 3 | 1 | -1 | -3 |
| T | -10 | -6 | -2 | 2 | 3 | 4 | 5 | 3 | 1 |
| A | -12 | -8 | -4 | 0 | 1 | 2 | 3 | 7 | 5 |
| G | -14 | -10 | -6 | -2 | -1 | 0 | 4 | 5 | 6 |

- Case Sim ($i; j$) : score entre les i premières bases de $U=ACGGCTAT$ et les j premières bases de $V=ACTGTAG$. Match = 2, Mismatch = -1, GAP (Insertion/Délétion) = -2.

- Cas de base (rouge): initialisation

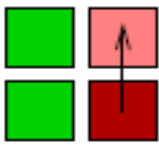
- Remplissage ligne par ligne en gardant en mémoire le mouvement qui donne le meilleur score

Etape 2: Recherche du chemin optimal dans la matrice.

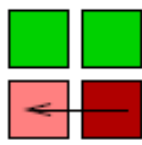
| $j \backslash i$ | | A | C | G | G | C | T | A | T |
|------------------|-----|-----|----|----|----|-----|-----|-----|-----|
| | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 |
| A | -2 | 2 | 0 | -2 | -4 | -6 | -8 | -10 | -12 |
| C | -4 | 0 | 4 | 2 | 0 | -2 | -4 | -6 | -8 |
| T | -6 | -2 | 2 | 3 | 1 | -1 | 0 | -2 | -4 |
| G | -8 | -4 | 0 | 4 | 5 | 3 | 1 | -1 | -3 |
| T | -10 | -6 | -2 | 2 | 3 | 4 | 5 | 3 | 1 |
| A | -12 | -8 | -4 | 0 | 1 | 2 | 3 | 7 | 5 |
| G | -14 | -10 | -6 | -2 | -1 | 0 | 4 | 5 | 6 |

Étape 3: Construction de l'alignement.

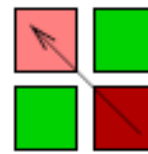
Sur le chemin des scores maximaux, on regarde quelle est l'opération correspondante.



Insertion



Déletion



Substitution

(match or mismatch).

Résultat:

```
ACGGCTAT
| | | |
ACTG-TAG
```

4-3- Alignement local:

Pour obtenir un alignement local optimal, une méthode a été développée par Smith et Waterman. Cette méthode permet l'alignement entre deux séquences portant sur des régions isolées et permettant de trouver des segments qui ont un haut degré de similitudes (score le plus élevé de la matrice). Cette propriété en fait un outil idéal, rapide et efficace, de recherche dans les bases de données en comparant une séquence inconnue avec les séquences de la banque. Elle permet de trouver des séquences homologues à notre séquence parmi les millions de séquences.

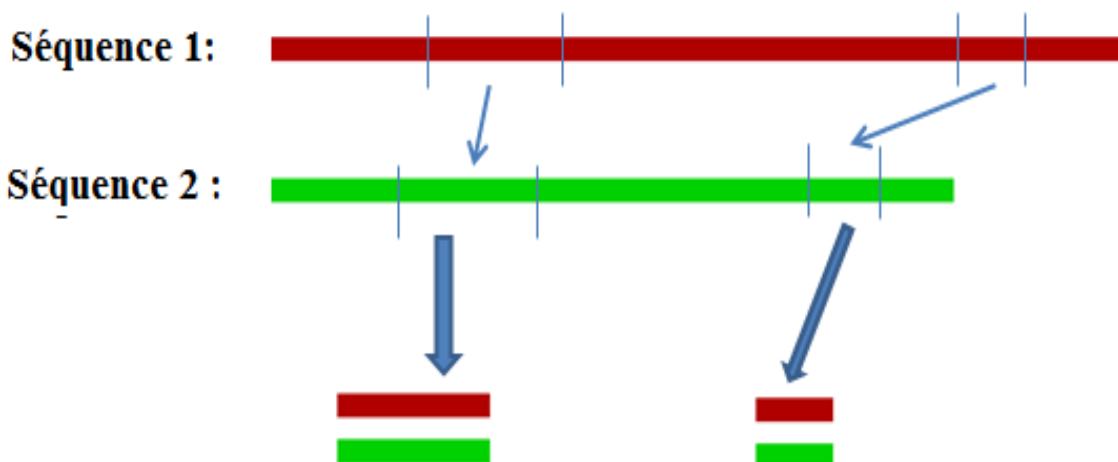


Fig.1: Alignement local.

4-4- BLAST

Le BLAST est acronyme de «Basic Local Alignment Search Tool » Cette méthode de recherche a été développée spécialement pour permettre de trouver les alignements locaux (régions similaires ou homologues entre deux ou plusieurs séquences de nucléotides ou d'acides aminés) statistiquement significatifs. Ce programme permet de retrouver rapidement dans des bases de données, les séquences

ayant des zones de similitude avec une séquence donnée (introduite par le chercheur). L'idée de base exploitée par l'algorithme est que les bons alignements doivent contenir quelque part des petites régions très riches en identité. Ces éléments, constituent les points d'ancrage à partir desquels l'alignement est étendu. Ce repérage initial permet de sélectionner rapidement les séquences de la banque potentiellement similaire à la séquence introduite. Le BLAST est utilisé pour trouver des relations fonctionnelles ou évolutives entre les séquences et peut aider à identifier les membres d'une même famille de gènes.

La signification statistique des alignements produits par BLAST est mesurée par E-value "Expected value". Cette valeur indique le nombre d'alignements différents ayant les mêmes degrés de similitude. Si $E = 10^{-2}$ cela signifie que l'alignement sur 100 sera trouvé par hasard. E-value est donné par la formule suivante: $e = k . m . n . e^{-\lambda.S}$

L'E-value dépend donc de la taille de la séquence (m), de la taille de la banque (n), et du score d'alignement (S), k et λ sont des paramètres caractérisant la banque de données. Un alignement est d'autant plus significatif que S est élevé et E faible.

Haemophilus influenzae strain 680 16S ribosomal RNA gene, complete sequence

Sequence ID: [ref|NR_044682.2|](#) Length: 1486 Number of Matches: 1

[▶ See 1 more title\(s\)](#)

Range 1: 432 to 856 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

| Score | Expect | Identities | Gaps | Strand |
|---------------|--------|---|-----------|-----------|
| 778 bits(421) | 0.0 | 425/425(100%) | 0/425(0%) | Plus/Plus |
| Query 1 | | GTTCCTTTCGGTATTGAGGAAGGTTGATGTGTTAATAGCACATCAAATTGACGTTAAATAC | | 60 |
| Sbjct 432 | | GTTCCTTTCGGTATTGAGGAAGGTTGATGTGTTAATAGCACATCAAATTGACGTTAAATAC | | 491 |
| Query 61 | | AGAAGAAGCACC GGCTAACTCCGTGCCAGCAGCCGGTAATACGGAGNGTGCAGCGTT | | 120 |
| Sbjct 492 | | AGAAGAAGCACC GGCTAACTCCGTGCCAGCAGCCGGTAATACGGAGNGTGCAGCGTT | | 551 |
| Query 121 | | AATCGGAATAACTGGGCGTAAAGGGCACGCAGGCGTTATTTAAGTGAGGTGTGAAAGCC | | 180 |
| Sbjct 552 | | AATCGGAATAACTGGGCGTAAAGGGCACGCAGGCGTTATTTAAGTGAGGTGTGAAAGCC | | 611 |
| Query 181 | | CCGGGCTTAACCTGGGNATTGCATTCAGACTGGGTAAGTAGAGTACTTTAGGGAGGGGT | | 240 |
| Sbjct 612 | | CCGGGCTTAACCTGGGNATTGCATTCAGACTGGGTAAGTAGAGTACTTTAGGGAGGGGT | | 671 |
| Query 241 | | AGAATTCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGAAGCGAAGGC | | 300 |
| Sbjct 672 | | AGAATTCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAATACCGAAGCGAAGGC | | 731 |
| Query 301 | | AGCCCTTGGGAATGTACTGACGCTCATGTGCGAAAGCGTGGGGAGCAAACAGGATTAGA | | 360 |
| Sbjct 732 | | AGCCCTTGGGAATGTACTGACGCTCATGTGCGAAAGCGTGGGGAGCAAACAGGATTAGA | | 791 |
| Query 361 | | TACCCTGGTAGTCCACGCTGTAAACGCTGTTCGATTTGGGGGTTGGGGTTAACTCTGGCA | | 420 |
| Sbjct 792 | | TACCCTGGTAGTCCACGCTGTAAACGCTGTTCGATTTGGGGGTTGGGGTTAACTCTGGCA | | 851 |
| Query 421 | 425 | CCCGT | | |
| Sbjct 852 | 856 | CCCGT | | |

Fig.2: Le résultat de la recherche par BLAST, montrant l'identification d'une souche de *Haemophilus influenzae* par l'alignement de la séquence d'amplification par PCR du gène codant l'ARN 16S à celle dans la banque de données.

4-5- Alignement de séquences multiples

Un alignement de séquences multiples (MSA: multiple sequence alignment) est un alignement de séquences de trois ou plusieurs séquences biologiques, généralement des protéines, l'ADN ou de l'ARN. Il permet de mettre en évidence les relations entre séquences que l'on ne peut visualiser en comparant les séquences deux à deux.

Ce type d'alignement est appliqué surtout pour :

- La caractérisation des régions (motifs, domaines) conservées des protéines.
- Prédiction des structures 2D ou 3D par comparaison avec des structures connues.
- Construction des arbres phylogénétiques des séquences homologues.

4-5-1- ClustalW / ClustalX "Cluster alignment"

Ce type de Clustal "Classique" est fondé sur l'utilisation d'un algorithme d'alignement progressif. Les séquences les plus similaires sont alignées en premier puis l'alignement progresse vers les séquences les plus distantes. C'est également un programme de construction d'arbres phylogénétiques.

4-5-1-1 Etapes de l'alignement multiples de type clustal

- 1) Alignement de toutes les séquences 2 à 2 et détermination des scores des alignements.
- 2) Construction d'une matrice de score (BLOSUM62) pour l'ensemble des séquences.
- 3) Construction d'un arbre guide à partir de la matrice traduisant les relations globales entre les séquences. Il exprime les relations entre les séquences dont la longueur des branches est égale (cladogramme). A distinguer du phylogramme dont la longueur des branches est proportionnelle au nombre de changements évolutifs (mutations).
- 4) Alignement progressif à partir de l'alignement des deux séquences les plus proches. Les séquences voisines sont alignées de proche en proche jusqu'à l'alignement multiple final.

5- Arbres phylogénétiques

La phylogénie moléculaire renseigne sur les changements survenus au niveau des séquences au cours de l'évolution(modifications graduelles au cours du temps).

Un arbre phylogénétique est un arbre schématique qui montre les relations de parentés entre des groupes d'êtres vivants (fig. 5; fig. 6). Il existe plusieurs techniques de construction des arbres phylogénétiques, plus ou moins rapides et plus ou moins fiables. Trois méthodes sont le plus souvent utilisées :

5-1- Méthode des distances (Unweighted Pair Group Method with Arithmetic Mean)

C'est une méthode basé sur les similarités entre paires de séquences, fondée sur l'alignement multiple et le calcul des distances entre ces séquences. La topologie de l'arbre est construite par la méthode Neighbor-Joining (NJ), c'est aussi une méthode de distance, qui consiste à calculer la longueur des branches, de telle manière à ce que les distances déduites de l'arbre soient les plus proches des distances mesurées. Elle a l'avantage d'être vraiment rapide. En général, elle est utilisée pour faire des arbres de plusieurs milliers de séquences. C'est la méthode utilisée par CLUSTALW.

5-2- Méthode de parcimonie (Maximum Parsimony)

Elle est fondée sur une méthode probabiliste qui recherche l'arbre le plus parcimonieux parmi tous les arbres possibles, c'est à dire celui qui nécessite le moins de changements pour expliquer les différences observées entre les séquences. C'est une méthode très lente par rapport à la méthode des distances et pas aussi précise que la méthode de maximum de vraisemblance (ML).

5-3- Méthode du maximum de vraisemblance (Maximum Likelihood-ML)

Cette méthode est souvent décrite comme étant la meilleure méthode et la plus fiable, c'est-à-dire la plus efficace pour trouver l'arbre le plus proche de la réalité. Elle repose sur un ou plusieurs caractères à étudier. Il s'agit d'une méthode probabiliste qui nécessite un modèle d'évolution. Le choix de ce modèle est crucial pour la qualité de l'arbre. L'inconvénient de cette méthode qu'elle est longue et nécessitant une grande puissance de calcul.

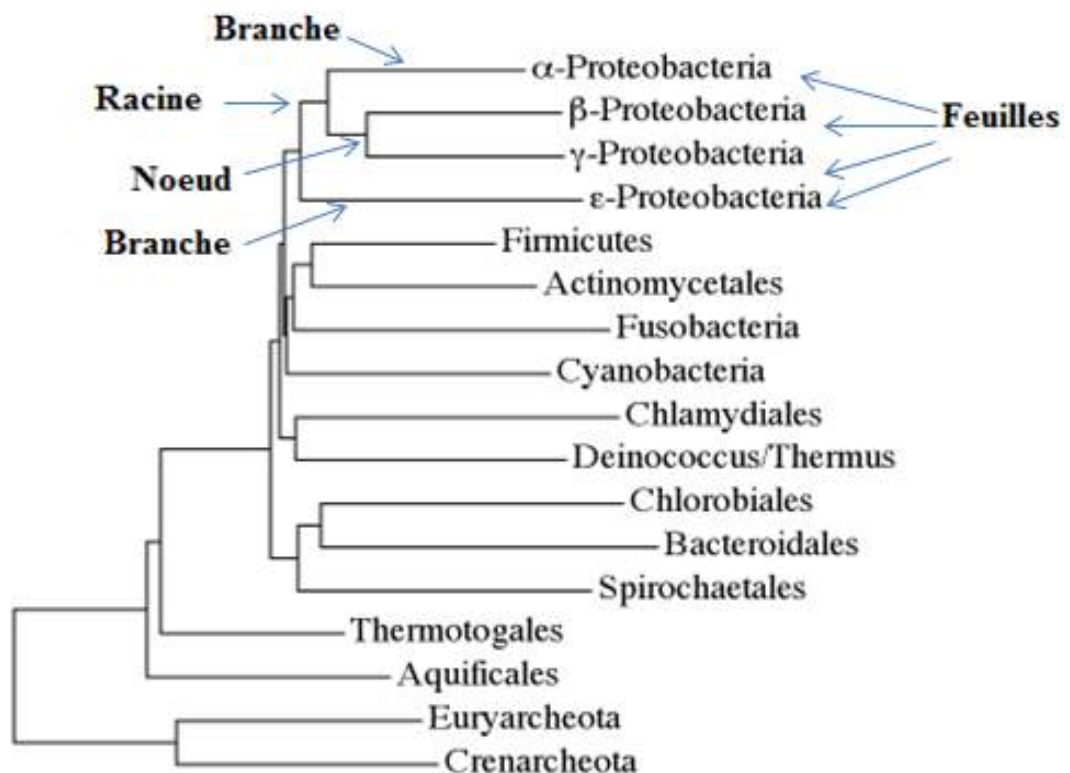


Fig.4: Arbre phylogénétique.

6- Recherche de motifs et de domaines

Les motifs sont de courts segments ou expressions. On sait que certains segments d'ADN ou de protéines sont déterminants dans l'analyse des séquences car ils correspondent à des sites précis d'activité biologique conservée au cours de l'évolution, comme par exemple les éléments de régulation des gènes ; la séquence de Shine Dalgarno, les ORF, les séquences consensus (séquence nucléique ou protéique reflétant une signature ou un trait commun entre différentes séquences), ou les signatures peptidiques. C'est pourquoi des bases spécialisées se sont naturellement constituées autour de ces séquences. Une séquence consensus est souvent établie, après alignement multiple de séquences.

L'utilisation des bases spécialisées comme les bases de motifs, est devenue un outil essentiel dans l'analyse des séquences pour tenter de déterminer la fonction de protéines inconnues ou savoir à quelle famille appartient une séquence non encore caractérisée.

En général, les motifs sont décrits avec une syntaxe spécifique. Les bases comme TFD ou IMD sont employées pour des séquences promotrices des gènes tandis que celles comme Prosite ou BLOCKS sont utilisées pour des protéines inconnues ou bien des séquences protéiques traduites à partir de cDNA ou de séquences génomiques.

Pour détecter une fonctionnalité sur une séquence, il suffit d'exécuter un programme qui s'appliquera à repérer la présence de certains motifs recensés dans ces bases et ainsi à prédire l'appartenance de la séquence testée à un groupe de séquences ayant une signature commune.

7- Prédiction et modélisation moléculaire

7-1- Prédiction de structures secondaires

Plusieurs paramètres physico-chimiques peuvent être déduits à partir de la séquence primaire de la protéine. Les principales caractéristiques physico-chimiques d'un acide aminé sont le volume, l'accessibilité au solvant, l'hydrophobicité, la taille (petit, gros), et l'aromaticité.

Le profil hydrophilique d'une séquence protéique permet d'obtenir quelques indices concernant la structure secondaire de la protéine correspondante. Par exemple, les acides aminés hydrophobiques tendent à se trouver à l'intérieur des protéines globulaires. Dans les protéines transmembranaires, ceux-ci ont tendance à se trouver au niveau de la membrane.

Les méthodes de prédiction de structures secondaires ont pour objectif principal de définir la localisation des éléments d'hélice α , de brin β , les coudes et les structures apériodiques. La structure secondaire est représentée par des valeurs d'angles phi (ϕ) et psi (ψ) particulières ($-57^\circ, -47^\circ$ pour les hélices par exemple) dont les possibilités constituent les degrés de liberté que peut prendre la chaîne principale.

La structure secondaire permet de faire des hypothèses structurales/fonctionnelles en l'absence de données expérimentales. Vouloir prédire la structure à partir de la séquence suppose l'hypothèse vérifiée que l'information nécessaire à l'acquisition de la structure 3D biologiquement active est contenue dans la séquence.

Différentes méthodes et approches ont été utilisées pour prédire la structure secondaire. Les méthodes statistiques, les méthodes de similarité est les méthodes neuronales. La méthode de chou et Fasman (1974-1978), la méthode de GarNIER Osguthorpe et Robson I (1978), et la méthode de plus proches voisins (Nearest Neighbor method).

7-2- Modélisation moléculaire

L'activité biologique des protéines et des ARN est intimement liée à la structure tridimensionnelle adoptée par ces biomolécules. Les molécules de par leurs dimensions sont invisibles à tout moyen d'investigation direct tel que la microscopie. C'est par l'analyse de données indirectes que les chercheurs peuvent reconstituer un modèle moléculaire, c'est-à-dire une construction intellectuelle présentant la meilleure adéquation avec les résultats expérimentaux. Ces données sont issues principalement d'analyses cristallographiques (étude des figures de diffraction des rayons X par un cristal), ou de Résonance Magnétique Nucléaire. Elles représentent les contraintes expérimentales exercées sur le modèle. Le modèle moléculaire obtenu ensuite est un ensemble de coordonnées atomiques dans l'espace. L'informatique intervient dans toutes les étapes conduisant de l'expérimentation au modèle, puis ensuite dans l'analyse du modèle par la visualisation moléculaire.

Plusieurs logiciels de visualisation créent un environnement tridimensionnel virtuel qui donne l'illusion de la profondeur. Ces logiciels permettent également la manipulation des structures et l'exportation sous forme de fichiers images. Elle est utilisée par exemple pour étudier les sites actifs d'une enzyme, mettre au point informatiquement une série d'inhibiteurs possibles pour cette enzyme, et ne synthétiser et ne tester que ceux qui semblent convenir. Cela permet de réduire les coûts de recherche et d'accélérer ces recherches. La visualisation de la structure tridimensionnelle d'acides nucléiques (ARN et ADN) fait également partie de la palette des outils bioinformatiques très utilisés. Parmi les logiciels les plus utilisés :

- [Rasmol: www.openrasmol.org] Visualisation, manipulation, pas de comparaison de structures, rapide, performant, simple et portable.
- [Chime: www.mdl.com/products/framework/chime] Version "Plug-in" de Rasmol commercialisé par [MDLI], téléchargement gratuit après inscription.
- [SPDBV: www.expasy.org/spdbv] Outil complet de modélisation moléculaire (modification de géométrie, mutations, construction...), consulter les tutoriels.

Références Bibliographiques

- Brown T. A. (2010). Gene Cloning and DNA Analysis: An Introduction. 6th Ed, Weley Blackwell. London, UK. 782p.
- Clark D. (2005). Molecular Biology: Understanding the Genetic Revolution. Southern Illinois University. USA Elsevier Academic Press. 784p.
- Gubta P. K. (2005). Microbiology, Cell Physiology and Biotechnology. 2nd Ed. National Offset Printers, Meerut, India. 311p.
- Hogg S. (2005). Essential Microbiology. British Library Cataloguing in Publication Data. John Wiley & Sons. UK. 468p.
- Housset C. et A. Raisonnier. (2006). Biologie Moléculaire. Biochimie PCEM1 Université Paris-VI. 204p.
- Klug S. W., M. R. Cummings, C. A. Spencer & M. A. Palladino. (2013). Essentials of genetics. 8th Ed. Pearson Education. New Jersey. USA. 608p.
- Lodish H., J. Darnell, D. Baltimore. (1989). La cellule: Biologie Moléculaire. Edition VIGOT. Bibliothèque nationale du Québec-Canada. 1189p.
- Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003). Molecular Cell Biology. 6th Ed. 937p.
- Madigan M., J. Martinko. (2007). Biologie des microorganismes. 11^e Ed. Pearson Education-Paris-France. 1047p.
- Misener S. and S. A. Krawetz. (2000). Bioinformatics Methods and protocols. Humana Press Inc, New Jersey. USA. 500p.
- Nicholl S. T. D. (2008). An introduction to Genetic Engineering. Third Edition. Cambridge University Press. UK. 292p.
- Okafor N. (2007). Modern Industrial Microbiology and Biotechnology. Science Publishers, Enfield, NH, USA. 530p.
- Passarge E. (2007). Color Atlas of Genetics. 3rd Ed. Flexibook. Thieme Stuttgart, NY. USA. 486p.
- Prescott M. L., J. P. Harley, et D. A. Klein. (2002). Microbiology. 5th Ed. The McGraw-Hill Companies. 1139p.
- Prescott M. L., J. P. Harley, et D. A. Klein. (2007). Microbiologie. 2^e Ed. Edition De Boek Université. Paris. 1137p.
- Primose S. B., R. M. Twyman and R.W. Old. (2002) Principles of Gene Manipulation. 6th Ed. Blackwell Scientific Publishing, Oxford. UK. 319p.
- Sambrook, J., and D. Russell. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory. New York. USA. 2344p.
- Singh B. D. (2007). Molecular Biology and genetic engineering. Kalyani Publishers. India. 311p.
- Smith E. J. (2005). Biotechnology. 3th Ed. Cambridge University Press. UK. 234p.

- Streips N. U. & R. Yasbin. (2002). Modern Microbial Genetics. 2nd Ed. Wiley-Liss, Inc. New York. USA; 603p.
- Turner P. C., A. G. Mc Lennan, A. D. Bates, M. R. H. White. (2002). Instant Notes in Molecular Biology. Bios scientific Publishers Limited. Oxford. UK. 386p.
- Ussery D. W., T. M. Wassenaar, and S. Borini. (2009). Computing for Comparative Microbial Genomics: Bioinformatics for Microbiologists. Computational Biology Series. Springer-Verlag London Limited. UK.270p.
- Watson J., T. Baker, S. Bell, A. Gann, M. Levine, R. Losick. (2009). Biologie Moléculaire du gène. 6^eEd. Pearson Education- Paris-France. 688p.