

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

UNIVERSITE L'ARBI BEN MHIDI-OU M EL BOUAGHI
FACULTE DES SCIENCES EXACTE ET SCIENCE DE NATURE ET DE LA
VIE



DEPARTEMENT DES SCIENCES DE LA NATURE ET DE
LA VIE

Cours de Biologie Moléculaire et Génie génétique

**Destiné aux étudiants de 3^{ème} année Licence Microbiologie
2017-2018**

Réalisé par Dr KAROUCHE SAIDA

Préface :

La biologie moléculaire est une discipline scientifique au croisement de la génétique, de la biochimie et de la physique, dont l'objet est la compréhension des mécanismes de fonctionnement de la cellule au niveau moléculaire.

La biologie moléculaire est apparue au XX^e siècle, à la suite de l'élaboration des lois de la génétique, la découverte des chromosomes et l'identification de l'ADN comme support chimique de l'information génétique.

Après la découverte de la structure en double hélice de l'ADN en 1953 par James Watson (1928-), Francis Crick (1916-2004), Maurice Wilkins (1916-2004) et Rosalind Franklin (1920-1958), la biologie moléculaire a connu d'importants développements pour devenir un outil incontournable de la biologie moderne à partir des années 1970.

Depuis la fin des années 1950 et le début des années 1960, les biologistes moléculaires ont appris à caractériser, isoler et manipuler les composants moléculaires des cellules et des organismes. Ces composants incluent l'ADN, support de l'information génétique, l'ARN, et les protéines, molécules structurelles et enzymatiques les plus importantes des cellules.

Une des techniques les plus élémentaires en biologie moléculaire pour étudier le rôle des protéines est le clonage d'expressions. Dans cette technique, l'ADN codant la protéine qui nous intéresse est cloné en utilisant la réaction en chaîne par polymérase (PCR) et/ou des enzymes de restriction dans un plasmide (qu'on appelle vecteur d'expression). Ce plasmide peut avoir des éléments de séquences promotrices spéciales pour diriger la production de la protéine en question et peut aussi avoir des marqueurs de résistance antibiotique pour aider à suivre le plasmide.

Ce polycopié est le support de cours de la matière biologie moléculaire et génie génétique ; unité d'enseignement fondamentale 2 (UEF 3.1.2) destiné aux étudiants de troisième année LMD Licence microbiologie avec un volume horaire semestriel de 16 semaines.

Le polycopié est divisé en deux parties : la première partie concerne la biologie moléculaire, permettra aux étudiants de comprendre la structure et l'organisation du génome avec toute sa complexité de transcription, traduction, régulation transcriptionnelle et traductionnelle. Ainsi que les techniques de bases (PCR, séquençage, électrophorèse, transfert sur membrane et marquage). Tandis que la seconde partie présente également l'ensemble des techniques de manipulation d'acides nucléiques (ADN, ARN), transfert de gènes appelées aussi techniques de génie génétique.

Chapitre I : l'expression de l'information génétique

Partie 1 : La transcription

Objectifs :

- Différencier la transcription chez les Procaryotes et les Eucaryotes
- Connaître les grandes étapes de la transcription qui permettent de passer de l'ADN à l'ARNm
- Connaître, pour chaque étape, les différentes protéines associées.

Introduction :

La fonction capitale de l'ADN d'une cellule vivante est de contenir l'information génétique nécessaire au fonctionnement cellulaire. Ce fonctionnement est assuré, dans les cellules eucaryotes d'un organisme humain, par environ 30 000 protéines distinctes.

Le mécanisme général de formation des ARN qui seront nécessaires à la synthèse protéique nécessite : • ARN messager (ARNm) ; ARN de transfert (ARNt) ; ARN constitutifs des ribosomes (ARNr). Des études ont montré que le mécanisme de copie décrit chez les bactéries est valable aussi chez les eucaryotes à quelques nuances près. Cependant une différence fondamentale existe : l'absence de noyau chez les bactéries et sa présence dans la cellule eucaryote.

I) Définition de la transcription :

La transcription est le processus de copie du matériel génétique (ADN) en ARN.

Chez les **procaryotes** une **seule ARN polymérase-ADN dépendante** effectue la transcription pour tous les types d'ARN, tandis que chez les **eucaryotes**, **trois** ARN polymérases (ARNpol) interviennent : l'ARNpol I pour les ARN ribosomiques transcrits dans le nucléole (28S, 18S et 5,8S), l'ARNpol II pour les ARNm, et l'ARNpol III pour les petits ARN (ARNt, ARNr 5S, ARNsn). Pour certains virus à ARN enfin, l'ARN est transcrit par une ARNpol-ARN dépendante appelée aussi réplisase.

a) Les particularités des ARN polymérases sont :

- Qu'elles n'ont pas besoin d'amorces pour fonctionner.

- La polymérisation se fait dans le sens 5' 3', donc lecture du brin d'ADN dans le sens 3' → 5'.
- L'ARN ne reste pas apparié au brin d'ADN pendant la synthèse.
- Les ARN polymérases n'ont pas d'activité de relecture (activité exonucléasique 3' → 5').
- La vitesse de synthèse est inférieure à celles des ADN polymérases (environ 50 nucléotides/seconde).

On peut d'emblée se poser une question : quel brin sera traduit ? Si l'on considère le brin d'ADN 3' → 5' antisens (transcrit), l'ARN lui sera complémentaire et antiparallèle. Si l'on considère l'autre brin d'ADN en 5' → 3' (non transcrit, informatif), le brin d'ARN écrit en 5' 3' lui sera identique, à l'exception toutefois de T remplacé par U.

Notons toutefois que lors d'une transcription, l'un des brins d'ADN est transcrit et l'autre pas, mais ce ne sera pas toujours le même brin. Dans une autre transcription, les rôles pourront fort bien être inversés figure 1.

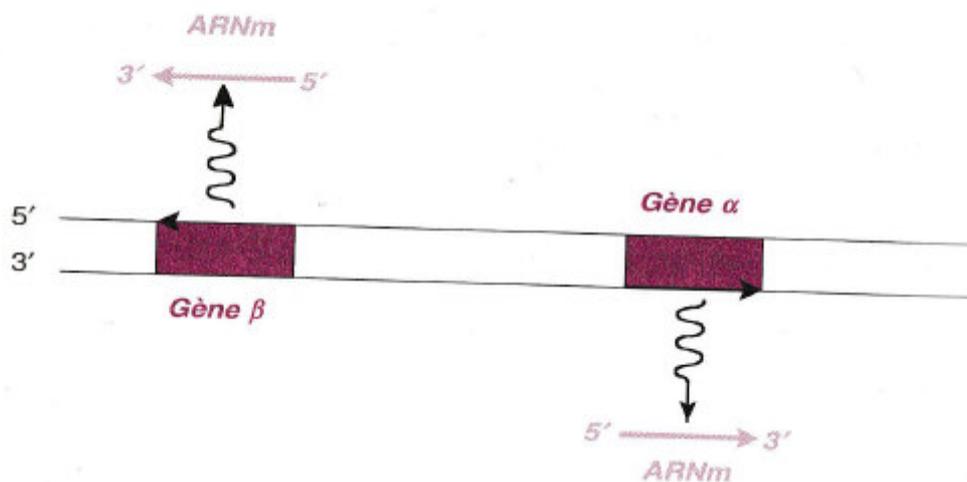


Figure 1 : Lecture des brins d'ADN lors de la transcription. Pour les gènes α et β , ce n'est pas le même brin d'ADN qui est lu, mais la lecture se fait toujours dans le sens 3' → 5', et la synthèse de l'ARNm de 5' vers 3'. (Beaumont S, 2007)

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 (figure 2).

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

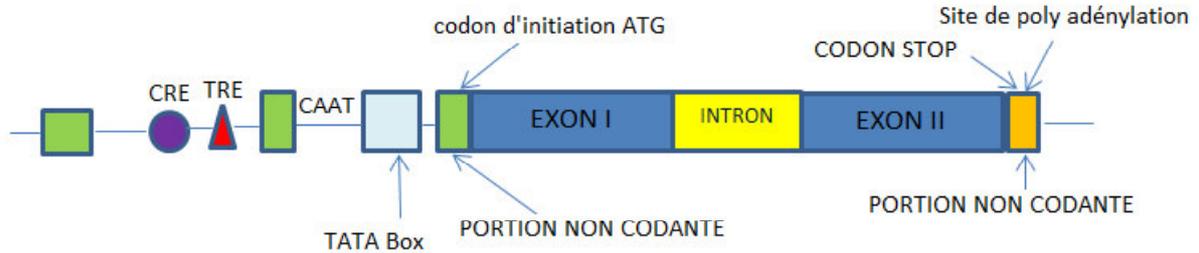


Figure 2: Structure général d'un gène chez les eucaryotes. (CRE [Cyclic AMP Responsive Element], TRE [12-O-Tetradacanoylphorbol 13-acetate Responsive Element]: des éléments de réponse).

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 x10⁶ pb), elles sont qualifiés d'éléments cis.

II) Transcription de l'ADN procaryote

La transcription s'effectue chez les bactéries grâce à une enzyme, l'ARN polymérase. C'est une enzyme de grande de taille (PM 450 000 Da) contenant 5 sous-unités ; chacune codée par un gène différent (tableau 1). Elle comprend la sous-unité σ et deux sous-unité α et un exemplaire de chacune des sous-unités β , β' (figure 3), 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 $\alpha_2 \beta \beta' \omega - \sigma$

Tableau 1: Gènes codant les sous-unités de l'ARN polymérase

Gène	Sous unité
<i>rpoA</i>	Deux sous unité α (36 500Da)
<i>rpoB</i>	Une sous unité β (151000 Da)
<i>rpoC</i>	Une sous unité β' (155000 Da)
<i>rpoD</i>	Une sous unité σ (70 Da)

L'ARN polymérase est une protéine ADN dépendante, **multimérique** possédant les sous-unités α , β , β' et σ . Elle est présente sous deux formes le **core-enzyme** ($\alpha_2\beta\beta'$) et l'**holoenzyme** ($\alpha_2\beta\beta'\sigma$). Les ARN polymérases ne nécessitent pas d'amorce et ne possèdent pas d'activité exonucléasique et donc de correction d'erreur, le taux d'erreur est ainsi plus important que pour les ADN-polymérases, mais ce taux est supporté.

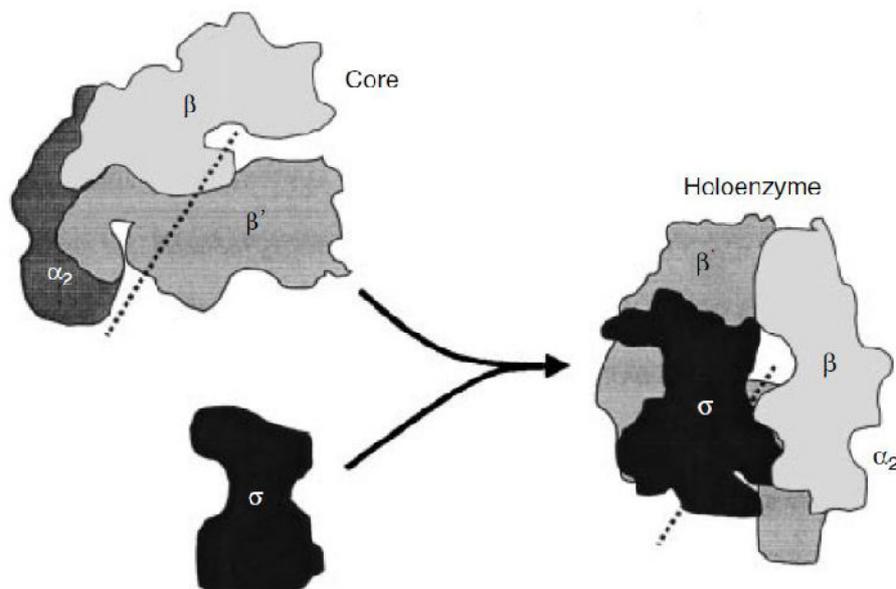
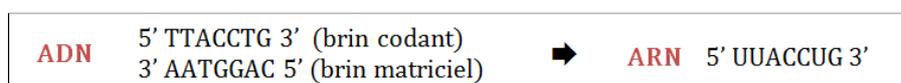


Figure 3 : Structure de l'ARN polymérase des bactéries.

Les nucléotides triphosphates sont additionnés à l'extrémité 3' de la chaîne en cours de synthèse par complémentarité de la matrice d'ADN. L'hydrolyse de la liaison anhydride fournit l'énergie pour la synthèse de la liaison phosphodiester.

La molécule d'ADN est composée d'un **brin matriciel** (ou **brin anti-codant**) dirigé par définition de 3' vers 5' et servant comme son nom l'indique de matrice à l'ARN-polymérase, ainsi que d'un **brin codant** dirigé par définition de 5' vers 3' et ayant une séquence identique à l'ARN transcrit mise à part le fait que la thymine est changée par l'uridine. L'ARN messager est lui monocaténaire et dirigé tout comme le brin codant de 5' vers 3'.



Chez E-coli, **une seule ARN-polymérase catalyse la synthèse de tous les ARN** de la cellule.

La transcription chez les procaryotes se fait au niveau du cytoplasme, elle est polycistronique car plusieurs gènes peuvent être transcrits par la même polymérase (exemple : L'opéron lactose). 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 (figure 4).

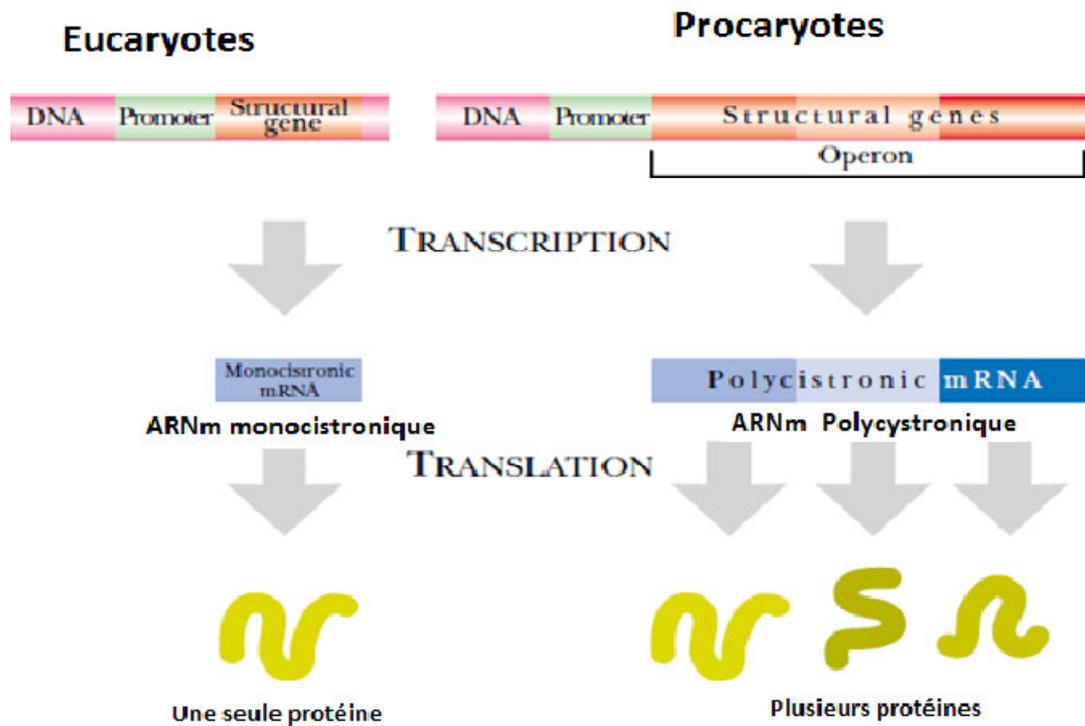


Figure 4: Synthèse monocistronique (eucaryotes) et polycistronique (procaryotes) de l'ARNm (Clark, 2005).

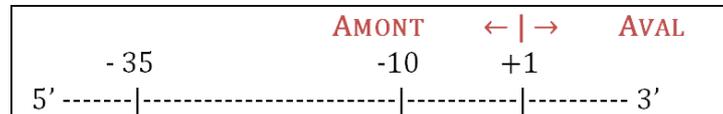
II.1 Étapes de la transcription chez les procaryotes

La transcription est divisée en 3 étapes : **l'initiation, l'élongation et la terminaison.**

Le promoteur est une séquence d'une centaine de nucléotides située dans la région régulatrice et désignant le début de la transcription. Il est situé en **amont** du **site d'initiation** et porte des séquences reconnues par l'ARN-polymérase et déterminant le sens de la transcription.

Le promoteur est constitué de courtes séquences conservées d'une unité de transcription à l'autre et appelées **séquences consensus** :

- En **-10** du site d'initiation on trouve la **TATA box** ou **boîte de Pribnow** : «TATAAT »
- En **-35** du site d'initiation on trouve : « TTGACA »



L'affinité de l'ARN-polymérase pour l'ADN dépend de la forme de l'enzyme : le core-enzyme a une affinité faible et non spécifique, l'holoenzyme a une affinité très forte et spécifique pour le promoteur. On peut faire la remarque que la sous-unité sigma σ à l'état libre ne se fixe pas sur l'ADN. La sous-unité β' étant basique et l'ADN étant acide, ce sera elle qui facilitera l'interaction du complexe avec le promoteur.

La **sous-unité sigma σ** permet donc une reconnaissance spécifique du promoteur par l'ARN-polymérase et diminue l'affinité de l'enzyme pour les régions non promotrices. Il agit de manière cyclique, en effet après l'initiation faite, le facteur sigma se détache pour être recyclé et réutilisé pour d'autres initiations de gènes.

L'ARN-polymérase entraîne la **dénaturation des deux brins d'ADN** sur 14 paires de nucléotides, on parle de complexe ouvert qui augmente encore l'affinité de l'enzyme pour la double hélice.

II.1.A Initiation :

fixation de L'ARN polymérase sur les sites promoteurs

Chez toutes les bactéries, le démarrage de la transcription est conditionner par la reconnaissance de séquences spécifiques de l'ADN appelées sites promoteurs. L'analyse de nombreuses séquences d'ADN situées en amont du site de démarrage de la transcription, et d'autres expériences comportant l'étude de l'ADN après sa liaison à l'ARN polymérase (détermination des séquences d'ADN protégées par l'ARN polymérase) ont permis de définir des séquences « type » ou séquences consensus de liaison de l'ARN polymérase. Ces séquences constituent le site promoteur (Figure 5).

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.



Figure 5 : Représentation du site promoteur des Procaryotes. Les chiffres indiquent la position des séquences par rapport à l'extrémité 3' de la TATA box prise comme référence. En gras apparaissent les nucléotides les plus conservés. (Beaumont S, 2007)

L'initiation de la transcription se fait au niveau du promoteur. L'ARN polymérase réalise la polymérisation de ribonucléotides face à un bras d'ADN de 5 vers 3. Cette enzyme grâce à son facteur σ ; est capable d'assurer la reconnaissance de séquences spécifiques situées l'égerment en amont du point d'initiation de la transcription (+1).

La sous-unité σ se lie au site promoteur. Les séquences -10 et -35 sont reconnues par la sous-unité sigma σ de l'ARN polymérase. La sous-unité σ rentre en contact avec les deux séquences, mais semble également rentrer en contact avec une zone adjacente située sur l'autre brin.

Les points de contacts sont situés des mêmes cotés de la double hélice, ce sont essentiellement des groupements chimiques donneurs ou accepteurs de liaisons hydrogènes situées dans le grand sillon. L'attachement se fait en deux temps :

- **D'abord à la région -35 ou une liaison relativement faible s'établit,**
- **puis à la région -10 ou une liaison très forte se produit. Cette deuxième liaison se traduit par une ouverture de la double hélice sur 15 paires de bases environ.**

L'initiation va durer le temps que la polymérase associe 7 à 8 ribonucleotides sous formes d'un polymère hybride au brin matrice.

Comme le core-enzyme présente une très forte affinité pour les hétéroduplex (brin d'ADN et brin d'ARN) le facteur σ se décroche et c'est le core-enzyme qui va seul continuer la transcription.

L'interaction de la **sous-unité β** est inhibée par la **rifampicine** et la **steptolydigne**. Ces substances permettent donc d'inhiber le développement des bactéries en empêchant la synthèse des protéines qui leur sont nécessaires lors de leur croissance. Une mutation dans la sous unité β induit l'apparition de souches bactériennes résistantes à la rifampicine.

Le déroulement des premières étapes de la transcription est donc :

1. liaison non spécifique de l'holoenzyme.
2. formation d'un complexe fermé au niveau du promoteur
3. formation du complexe ouvert (déroulement sur 14 nucléotides)
4. Mise en place du premier nucléotide (très souvent A ou G)
5. Allongement de 7 à 8 nucléotides
6. Détachement du facteur sigma, après la transcription des premiers nucléotides.

Des points de contact supplémentaires entre l'ARN polymérase (les autres sous-unités) et l'ADN se forment, et la transcription va pouvoir commencer. La sous-unité σ se séparera alors de l'enzyme (dès que la chaîne d'ARN aura quelques nucléotides) pour aller se lier à une autre molécule d'ARN polymérase.

Sans (ou avec) la sous-unité σ , l'ARN polymérase a une affinité faible pour l'ADN ; elle « glisse » donc sur l'ADN jusqu'à ce que la sous-unité σ reconnaisse le site promoteur, ce qui facilite la reconnaissance du site promoteur.

La sous-unité σ se trouve d'un côté de la double hélice. C'est cette orientation qui va permettre la copie du brin matrice orienté 3' --> 5'. C'est ce qui explique que l'ARN polymérase ne copie qu'un seul brin. Il y a donc reconnaissance de la TATA box dans le sens 3' --> 5', pour que l'ARN polymérase polymérise l'ARN dans le sens 5' --> 3'.

C'est aussi l'orientation du site promoteur qui explique que dans une molécule d'ADN bactérien, l'une ou l'autre des chaînes sera copiée (Figure 6).

Des séquences d'ADN situées plus loin, entre —50 et —150 nucléotides, sont aussi très importantes pour que l'activité du site promoteur soit maximale (elles influencent la vitesse de l'initiation).

1. Les premiers nucléotides de la chaîne d'ARN formée, ne sont pratiquement jamais ceux que l'on retrouve dans l'ARN traduit.
2. Des nucléases enlèvent quelques nucléotides ; parfois un clivage plus éloigné se produit. Ce qui sera important, en fait, c'est le codon d'initiation pour la traduction, et les séquences d'ARN permettant l'interaction avec le ribosome.

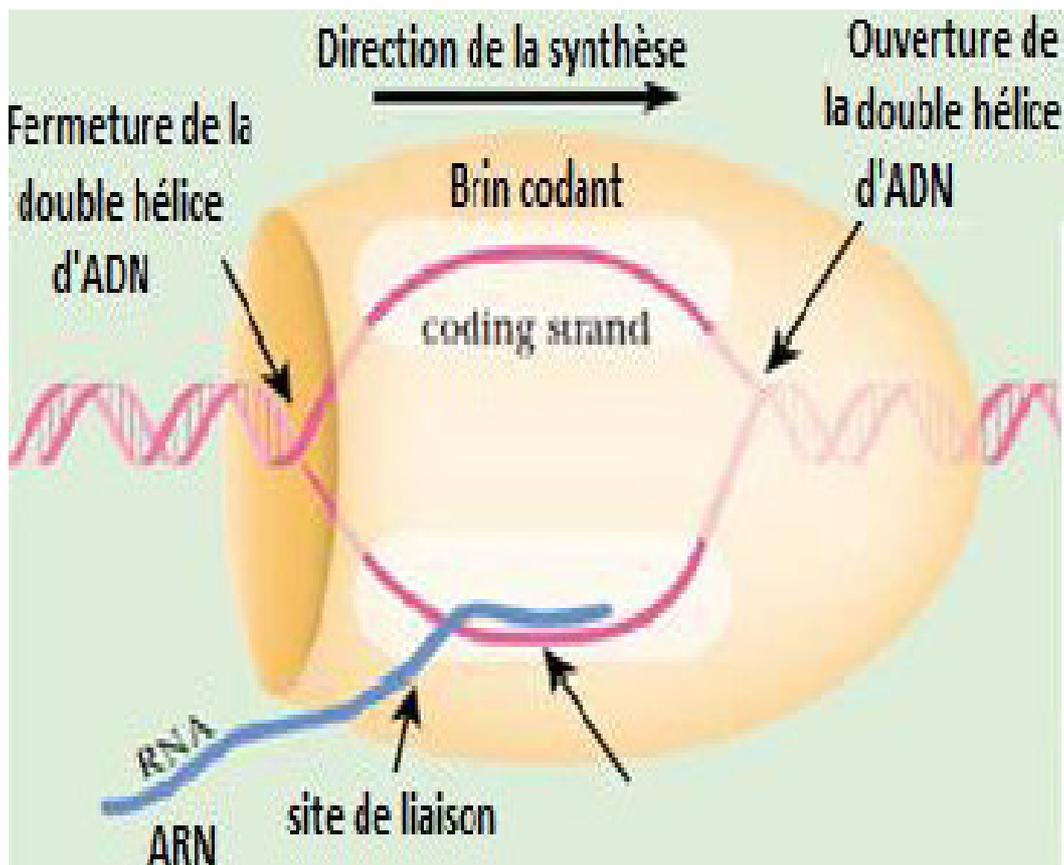


Figure 7: Elongation et direction de la synthèse.

II.1.C Terminaison

La terminaison de la chaîne est assurée par des séquences particulières appelées **signaux de terminaison**. Un signal de terminaison est précédé par des sites dits de pause où la polymérase sera ralentie. Ces sites sont, soit des séquences **riches en G-C**, difficiles à ouvrir, soit des structures à caractère **palindromique**, c'est-à-dire présentant une symétrie.

On en distingue deux types :

- Signaux ayant une structure d'ADN particulière comportant un centre de symétrie d'ordre 2.

Ces séquences particulières, lorsqu'elles seront transcrites en ARNm, permettront à celui-ci de se boucler sur lui même puisque ses séquences seront complémentaires et pourront donc s'apparier grâce à des liaisons hydrogènes.

Quand cette structure en épingle à cheveux se forme (Figure 8), elle correspond à une dissociation du complexe ADN-ARN sur environ 20 nucléotides. Il reste alors une séquence poly U liée à l'ADN. C'est un hybride très instable qui pourra se dissocier immédiatement de l'ADN (donc tout l'ARN est dissocié de l'ADN).

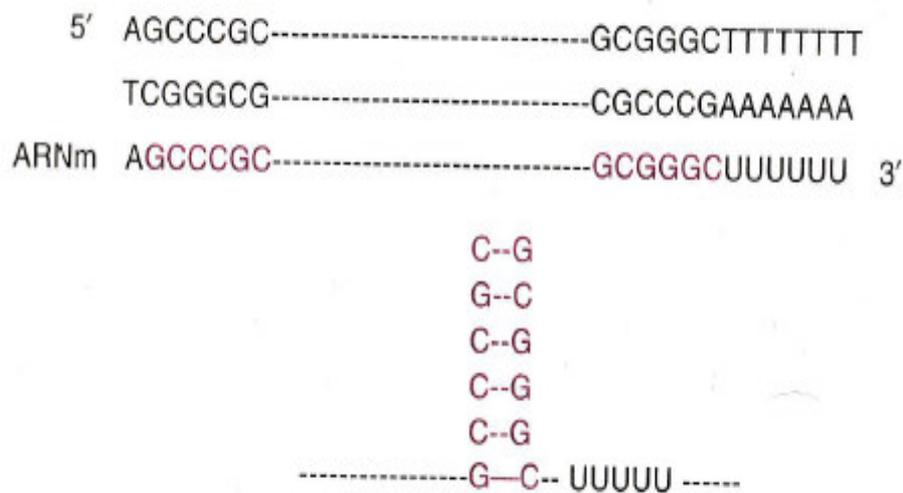


Figure 8 : Terminaison de transcription chez les Procaryotes.

En haut les deux brins d'ADN, dont l'un est transcrit en ARNm qui contient des nucléotides capables de former la structure en épingle à cheveu. (Beaumont S, 2007)

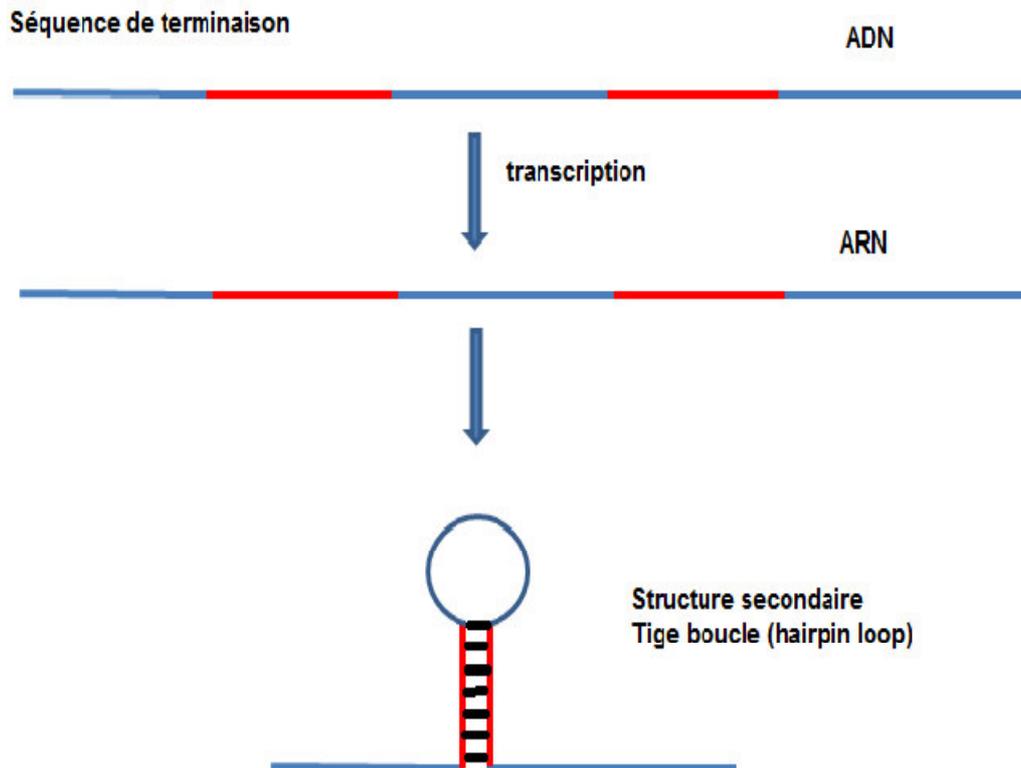


Figure 9: Formation du signal de terminaison de la transcription chez les bactéries (Aouf, 2016)

Ces structures secondaires suivies d'une série d'uridines sont des terminateurs efficaces de la transcription (figure 10). 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**".

- Certains signaux de terminaison ont une structure en épingle à cheveu insuffisante pour entraîner une dissociation immédiate de l'ARN. Ils ont besoin de la liaison d'une **protéine p (Rho)**. La protéine p se fixe, dans une réaction impliquant l'hydrolyse d'ATP, sur l'ARN, alors que l'ARN polymérase est dans un site de pause ; l'ARN entoure la protéine p sur 80 nucléotides ; elle interagit aussi avec les protéines associées et détache l'ARN de sa liaison à l'enzyme (l'ATP sert de donneur d'énergie dans cette réaction) (figure 11).

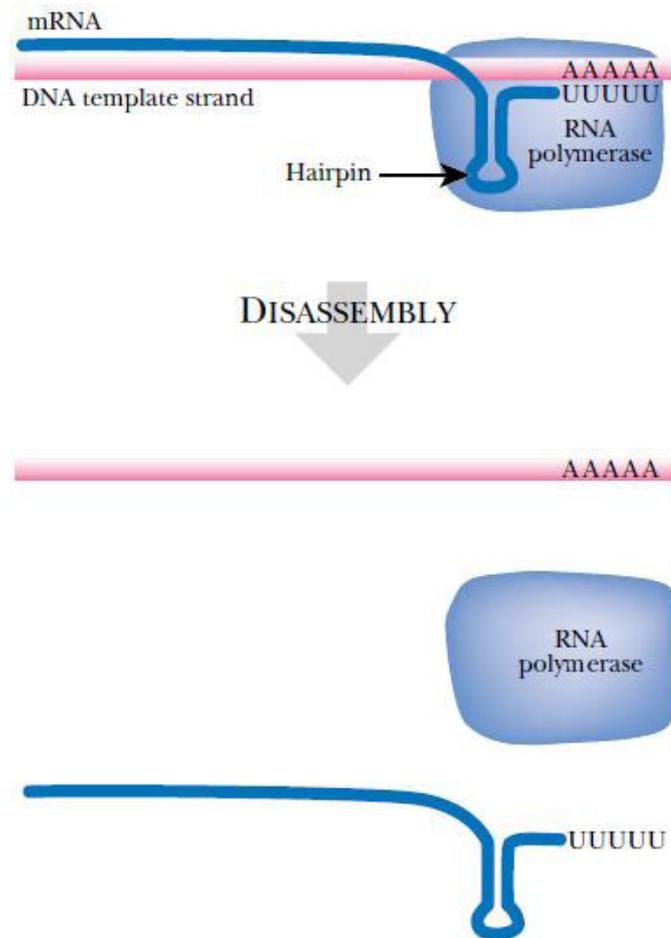


Figure 10: Terminaison de la transcription par des terminateurs intrinsèques (Clark, 2005).

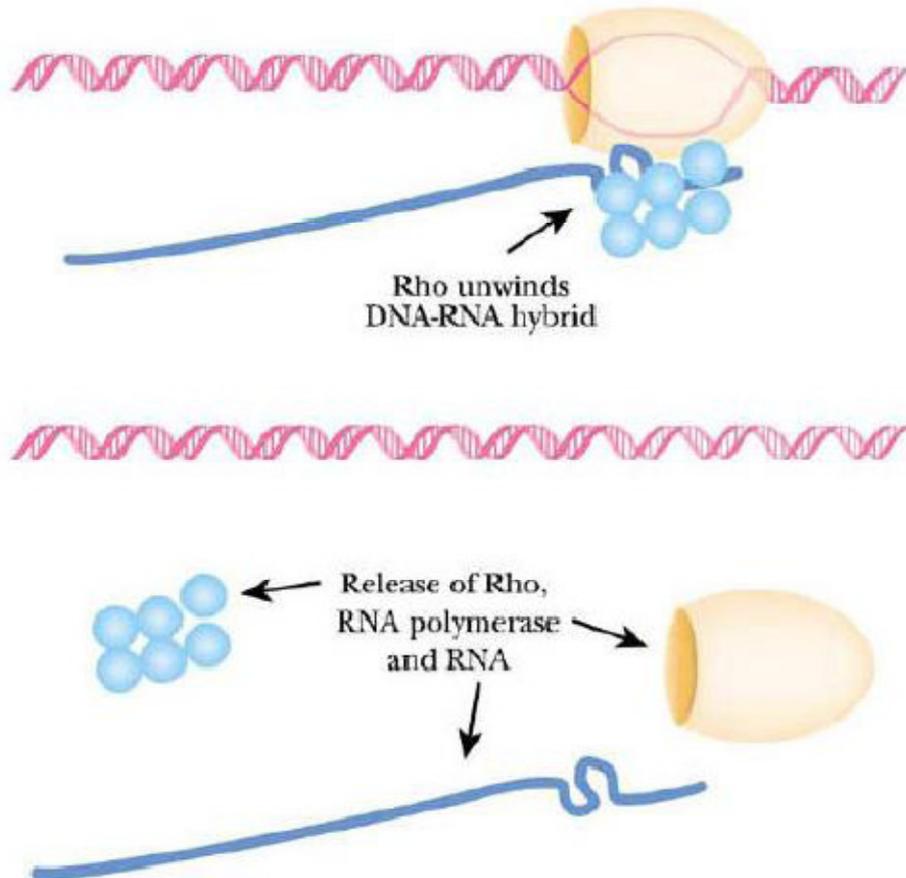


Figure 11: Terminaison de la transcription chez les bactéries par le facteur *Rho* (Clark, 2005).

Bibliographie :

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 180-193p

Clark D. (2005). Molecular Biology: Understanding the Genetic Revolution. Southern Illinois University. USA Elsevier Academic Press.784p.

Klug S. W., M. R. Cummings, C. A. Spencer & M. A. Palladino. (2013). Essentials of genetics. 8th Ed. Pearson Education. New Jersey. USA. 608p.

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.

Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003).Molecular Cell Biology. 6th Ed. 937p.

Madigan M., J. Martinko. (2007). Biologie des microorganismes. 11e Ed.Pearson Education-Paris-France. 1047p.

WatsonJ., T. Baker, S. Bell, A. Gann, M. Levine, R. Losick. (2009). Biologie Moléculaire du gène. 6eEd. Pearson Education- Paris-France. 688p.

Aouf Abdelhakim, (2016). Cours de Biologie moléculaire et génie génétique. Université Ferhat-Abbas Setif 1, 133p.

III) Transcription de l'ADN eucaryote

Chez les eucaryotes le mécanisme de base de la transcription est identique à ce qui a été décrit pour les procaryotes ; cependant **la structure du promoteur** est différente et les transcrits primaire obtenus sont toujours **monocistronique**.

Trois points essentiels distinguent la transcription des eucaryotes et des procaryotes :

- chez les eucaryotes, l'ADN est contenu dans le noyau et la transcription s'effectue dans celui-ci alors que la traduction en protéines s'effectue dans le cytoplasme. Chez les procaryotes, ces deux phénomènes sont associés et la traduction commence avant même que la transcription soit fini. De plus, l'ADN des Eucaryotes est présent sous forme d'une structure appelée la chromatine. La transcription de l'ADN en ARN implique une structure décondensée de la chromatine, c'est-à-dire une chromatine où il existe peu d'histone H1, où la structure est en collier de perles avec des zones sans nucléosomes.

- Chez les Eucaryotes, il existe trois ARN polymérases distinctes (alors qu'une seule fait «tout le travail» chez les bactéries)

- ARN polymérase I : ARNr (sauf l'ARN 5S)
- ARN polymérase III : ARNt, ARN 5S et petits ARN nucléaires
- ARN polymérase II : ARNm.

Ces polymérases ne fonctionnent correctement (en particulier pour la fixation sur des sites promoteurs) qu'en présence de très nombreuses protéines supplémentaires appelées **facteurs de transcription**.

- chez les eucaryotes, la transcription d'un gène en ARN m, r, t ne produit pas immédiatement une molécule fonctionnelle. Dans le noyau c'est un précurseur (ou transcrit primaire) qui est produit dans la grande majorité des cas, il subira plusieurs modifications avant son entre dans le cytoplasme.

III.1 Transcription par l'ARN polymérase I :

La synthèse de l'ARN ribosomal s'effectue dans un compartiment spécialisé du noyau, le nucléole. Cette synthèse représente plus de la moitié de l'ARN total cellulaire avec près de 80 % de tous les ARN.

Elle s'effectue sous forme d'un précurseur de grande taille (en moyenne 45S) qui sera converti en ARNr 18S, 5,8S et 28S. Ces ARN, avec l'ARN 5S transcrit sous contrôle d'une autre ARN polymérase (ARN polymérase III), vont former les constituants non protéiques du ribosome.

Cette synthèse est sous contrôle d'une ARN polymérase spécialisée, l'ARN polymérase I.

III.1.A Fonctionnement de la transcription

Le site promoteur comporte deux séquences situées vers -140 (site distal) et vers -35 (site proximal). Le premier site est le plus important pour la transcription, mais une transcription très active ne peut avoir lieu sans le deuxième site (Figure 1).

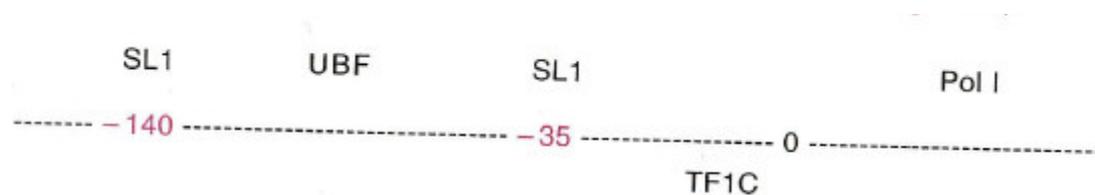


Figure 1 : Structure du site d'initiation de transcription pour l'ARN pol I. (Beaumont , 2007)

La séquence de ces deux sites est spécifique de l'espèce, ce qui correspond à une fixation de facteurs de transcription qui ne reconnaissent que le site promoteur de l'espèce dont ils proviennent.

Deux protéines se lient au promoteur :

- SL1 ou TFI D ou TFI B (fixation sur les deux sites)
- UBF (fixation entre les deux sites).

Seule la protéine SL1 a la spécificité d'espèce.

Remarque :

Un autre facteur protéique TFI C semble nécessaire à ce démarrage de la transcription. La structure ressemble à la sous-unité σ de la polymérase *d'E. coli*.

Enfin l'ARN polymérase I est nécessaire. Elle est composée de 65 sous-unités chez les vertébrés. La plus grande ressemble à β *d'E. coli*.

La terminaison est assurée par une structure composée de deux parties :

- une première où se fixe une protéine qui a été identifiée aussi bien chez un eucaryote inférieur, la Levure *S. cerevisiae* (Reb-1p), que chez les Mammifères (TTF-1)
- une seconde qui est indispensable aussi à la libération du transcrit d'ARN et qui est une séquence riche en paire A-T.

La protéine va bloquer la polymérase en situation de « pause » et cette dernière va reculer ce qui va créer un hybride ARN-ADN instable et dissocier l'ARN de sa liaison avec l'ADN et avec l'ARN polymérase.

III.1.B Maturation du précurseur 45 S et assemblage :

Quand la copie de l'ADN en pré-ARNr est faite, le pré-ARNr est incorporé dans un grand complexe ribonucléoprotéique qui clive ce pré-ARNr. Ce complexe contient la ribonucléoprotéine U3 : découpage d'abord dans la partie 5' puis du côté 3', puis entre les deux. Ce n'est pas une élimination d'introns sous forme d'épissage, mais une action de type nucléasique. Après transcription, le pré-ARNr ou ARN 45 S (environ 14 000 bases) subit aussi des modifications post-transcriptionnelles : méthylation, réduction, formation de pseudo-uridine, etc. comme chez les Procaryotes.

L'assemblage des ARNr obtenus après maturation avec les protéines constitutives des deux sous-unités ribosomales 40S et 60S (85 protéines différentes) et avec l'ARN 5S (qui s'associe à la grande sous-unité 60S) se fait dans le nucléole. Ce n'est que lorsque l'assemblage est terminé que les sous-unités vont quitter le nucléole pour aller vers le cytoplasme.

III.2 Transcription par l'ARN polymérase III :

L'ARN polymérase III copie :

- l'ARN 5S constitutif des ribosomes
- les ARN de transfert
- certains ARN nucléaires appelés petits ARN nucléaires ou ARNsn.

III.2.A Le site promoteur pour l'ARN 5S et la fixation de l'ARN polymérase III :

On sait qu'il est situé à l'intérieur du gène et qu'il se lie à un facteur de transcription TFIII A.

La liaison de ce facteur à l'ADN est une structure dite doigt de zinc Cys 2-His 2 (deux Cys et deux His en interaction avec un ion zinc II) qui se lie au grand sillon de l'ADN.

La partie C-terminale de TFIII A se lie à TFIII B et C. L'ensemble va ensuite interagir avec l'ARN polymérase III. Le complexe recouvre alors la totalité du gène 5S.

C'est en définitive l'interaction ARN pol III-TFIII B qui permet à celle-ci de commencer la transcription. Il est à noter que cette transcription se fait sans dissociation du complexe TFIII A-ADN.

III.2.B La transcription des ARNt :

Le site promoteur pour l'ARNt est constitué de deux séquences (+ 10 à + 20) et (+ 50 à + 60) qui vont fixer le facteur de transcription TFIII C.

L'interaction se fait en plusieurs étapes :

- TFIII C se lie à l'ADN
- après fixation, TFIII B se fixe à TFIII C
- le complexe stable formé permet alors la liaison de la polymérase III.

III.2.C La transcription et la maturation du transcrit primaire :

- 5S, ARNsn : pas de maturation.
- L'ARN de transfert subit un processus d'épissage : élimination d'une séquence « inutile » (intron) et « recollage » des séquences « utiles », avant de passer dans le cytoplasme. Ce processus d'élimination de l'intron est assez différent de celui utilisé pour l'ARNm. Ce n'est qu'après l'élimination de l'intron que l'ARNt passe dans le cytoplasme.

III.3 Transcription par l'ARN polymérase II :

III.3.1 Transcription d'un ARN messager :

III.3.1.1. L'initiation de la transcription :

Tout l'ADN n'est pas transcrit, seules les régions correspondant à des gènes le sont. Et encore, cette expression peut être régulée selon le stade de développement, le type cellulaire, l'environnement,... Dès lors, un acteur doit intervenir pour déterminer à quel endroit une région d'ADN doit commencer à être transcrite : c'est le rôle du promoteur.

III.3.1.1.a. Site promoteur :

Il faut bien distinguer, étant donné le niveau d'expression variable des gènes, deux zones distinctes :

- le site promoteur proximal
- les sites de régulation.

Le promoteur correspond à une région non transcrite de l'ADN, généralement juste en amont du début de la région transcrite, dont la séquence permet le recrutement de l'ARNpol II. Certaines séquences du promoteur (surnommées "boîtes") ont une importance particulière dans ce processus, essentiellement parce que ces séquences sont reconnues spécifiquement par différentes protéines appartenant au complexe d'initiation:

- la "**boîte TATA**" riche en thymine et adénine, la plus importante, est située vers **-25 à -30** nucléotide du site de démarrage de la transcription (noté +1)
- des éléments proximaux :
 - la "**boîte CAAT**" (facultative), contenant de la cytosine, est située vers **-120 à -80** nucléotides du site de démarrage de la transcription.
 - la "**boîte GC**" (facultative également), riche en guanine et cytosine, peut être présente entre la boîte CAAT et la boîte TATA.

Signalons que si ces séquences sont souvent bien conservées, une variabilité non négligeable peut également être observée. Ainsi, il existe des promoteurs sans "boîte TATA".

III.3.1.1.b. Le complexe d'initiation :

Contrairement à ce qui se passe chez les procaryotes, l'ARN pol II des eucaryotes ne reconnaît pas seul le promoteur proximal. Elle effectue ce travail en compagnie de nombreux co-facteurs protéiques qui se recrutent les uns les autres et qui forment avec elle un complexe d'initiation. Ces facteurs sont notés TFIIA, TFIIB,... pour **Transcription Factor for RNA polymerase II**. Ils correspondent aux facteurs généraux de la transcription car ils s'assemblent sur tous les promoteurs utilisés par l'ARN pol II. La séquence d'assemblage du complexe d'initiation se fait dans un ordre précis : TFIID puis TFIIA, B, ARN pol II se fixe ensuite suivie par TFIIF, E, H, pour produire finalement un complexe fonctionnel capable d'initier la transcription (figure 2).

Le recrutement de la polymérase II est très compliqué. Jusqu'à récemment, on considérait l'accrochage séquentiel suivant pour les gènes : TFII D, puis TFII A qui augmente la liaison de TFII D à TATA, puis TFII B qui reconnaît TBP (TATA box Binding Protein) et permet le recrutement de l'ARN pol II avec II F, II E, faiblement associé, II H et II B.

Il est capital de comprendre que la transcription effectuée en présence de ces différents facteurs est très faible. Le rôle de différentes protéines régulatrices qui s'associent à des séquences situées en amont ou en aval, voire à l'intérieur du gène, sera essentiellement d'augmenter la stabilité du complexe d'initiation de la transcription : liaison à TBP mais surtout aux TAF (TBP Associated Factors), liaison aux SRB. La fixation de différents facteurs de régulation va augmenter de 100 à 1 000 le taux de transcription.

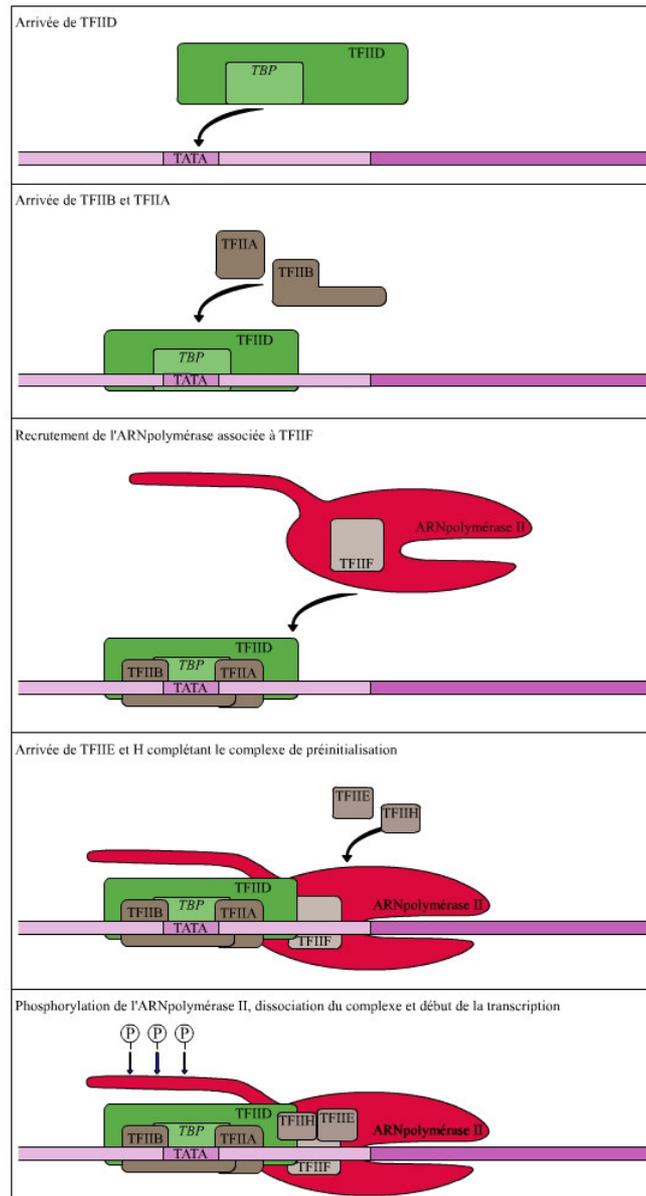


Figure 2 : Mise en place du complexe d'initiation

La liaison du complexe de transcription au promoteur proximal provoque l'ouverture et le déroulement des deux brins de son ADN, tout en indiquant le brin qui va être transcrit.

III.3.1.1.c. Intervention de facteurs spécifiques de la transcription :

Le complexe d'initiation composé de l'ARN pol II et des différents TFII est suffisant pour obtenir une activité transcriptionnelle *in vitro* mais à très faible taux. L'augmentation de cette activité basale (ou sa répression) est sous la dépendance de facteurs spécifiques qui vont interagir avec le complexe d'initiation. Ces protéines activatrices ou inhibitrices se lient à des promoteurs distaux, séquences spécifiques de l'ADN, appelées **enhancer** lorsqu'il recrute des

cofacteurs **activateurs**, ou **silencer** lorsqu'ils recrutent des cofacteurs **inhibiteurs**. Ces promoteurs distaux peuvent être situés à des milliers de nucléotides du promoteur proximal et agissent sur le promoteur proximal par le jeu de courbures de l'ADN (voir figure 3).

Mais d'autres participants existent encore. On trouve ainsi des promoteurs proximaux alternatifs et des promoteurs distaux, situés à des milliers de nucléotides du promoteur proximal. Malgré la distance qui sépare les promoteurs proximaux des promoteurs distaux, ces derniers agissent sur le promoteur proximal comme activateurs (enhancers; voir figure 3) ou répresseurs (silencers) par le jeu de courbures de l'ADN, des facteurs de transcription et du médiateur qui maintient liés tous ces acteurs.

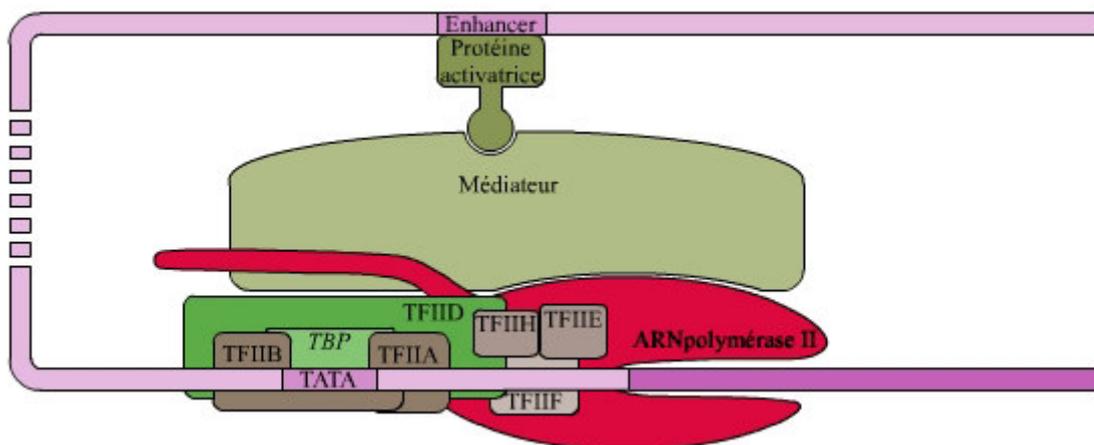


Figure 3 : La régulation spécifique de la transcription

Les enhancers (activateurs) sont des séquences situées parfois à plusieurs milliers de nucléotides d'un promoteur qui activent ce dernier par un jeu d'interactions protéiques qui stabilisent le complexe d'initiation. Ces interactions sont rendues possibles par la courbure de l'ADN qui rapproche des éléments situés à de grandes distances les uns de autres. Cette activation est spécifique du gène et utilise une multitude de facteurs de transcriptions spécifiques (les protéines activatrices) qui agissent généralement sous forme dimérique). Une répression spécifique peut agir selon le même type de modalité.

III.3.1.2. L'élongation :

La TBP est la première protéine qui reconnaît une séquence spécifique de l'ADN initiatrice de la transcription (la boîte TATA). Le facteur TFIIB semble impliqué dans la sélection précise du site d'initiation (nucléotide à partir duquel se déroule la transcription). Le

facteur TFIIFH comporte plusieurs activités enzymatiques dont une activité **hélicase** permettant l'ouverture de la double hélice d'ADN au niveau du promoteur, et une activité **kinase** responsable de la phosphorylation de la queue C-terminale de l'ARN polymérase II (domaine CTD, pour C-Terminal Domain) est un des événements « clef » du passage à l'élongation. Cette phosphorylation entraîne une modification de la structure tridimensionnelle de l'ARN polymérase entraînant la dissociation du complexe d'initiation et le début de la transcription.

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 TFIIFH. La régulation du niveau de phosphorylation du CTD de Pol II contrôle aussi les étapes ultérieures (élongation, maturation).

TFII = Transcription Factor for RNA polymerase II (Facteurs généraux de la transcription pour l'ARN polymérase II).

TBP = TATA box-Binding Protein (Protéine de liaison à la boîte TATA).

L'ARN pol II est équipé de facteurs protéiques d'élongation qui facilitent sa progression au travers d'une chromatine dont ils relâchent la structure. Un ARN pré-messager complémentaire du brin matrice de l'ADN (brin antisens), donc identique au brin codant de l'ADN (brin sens) aux riboses et uraciles près, commence à être synthétisé selon la direction 5'-3' (voir figure 4).

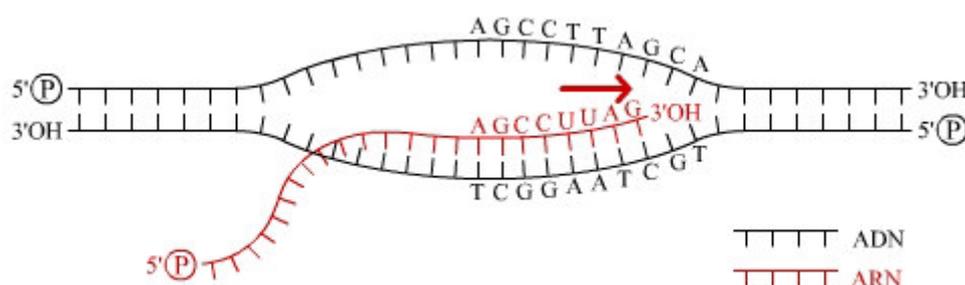


Figure 4 : La phase d'élongation de la transcription

L'ARN polymérase lit le brin antisens qui est complémentaire du brin codant (ou brin sens) depuis son extrémité 3' vers son extrémité 5'. Le brin d'ARN néosynthétisé est donc identique au brin codant d'ADN, aux uraciles et riboses près.

III.3.1.3. La terminaison :

L'ARN pol II est également équipée de facteurs protéiques de terminaison, et reconnaît ainsi un ou plusieurs signaux de terminaison portés par le brin progressivement parcouru et qui annoncent la fin de la transcription sur le brin **d'ADN matrice (TTATTT)**. Elle arrête bientôt son travail de transcription et libère l'ARN prém qu'elle vient d'assembler.

III.4 Inhibiteurs de transcription :

Il existe certains antibiotiques qui se comportent comme des inhibiteurs de la transcription, en agissant à des niveaux différents :

- *La rifampicine*, qui inhibe l'étape d'initiation de la transcription, en se plaçant dans le site actif de l'ARN polymérase.
- *L'actinomycine D*, s'intercale entre les plans des bases de l'ADN, et l'empêche alors de servir de matrice à la fabrication de l'ARN.
- La toxine de l'amanite phalloïde (champignon vénéneux), *l'alpha amanitine* inhibe l'ARN polymérase II des Eucaryotes en s'y fixant avec une très grande affinité.

Bibliographie :

Barmak Modrek & Christopher Lee, (2002). A genomic view of alternative splicing. Nature Genetics 30 : 13 - 19

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 180-193p

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.

Housset C. et A. Raisonier. (2006). Biologie Moléculaire. Biochimie PCEM1 Université Paris-VI. 204p.

PERRIN Pascale. (2010), Le contrôle de l'expression génétique FLBI399. Université Montpellier 2-53p.

BELKADI Bouchra. (2010). Cours de Biologie Moléculaire, L'expression génétique. Université Mohamed V - Agdal • Faculté des Sciences B.P. 1014 - Rabat – MAROC- 45p.

Carl Herrmann. (2006), la régulation de la transcription. Institut de Biologie du Développement Marseille Luminy. Université de la Méditerranée 56p.

Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003). Molecular Cell Biology. 6th Ed. 937p.

Madigan M., J. Martinko. (2007). Biologie des microorganismes. 11e Ed. Pearson Education- Paris-France. 1047p.

Watson J., T. Baker, S. Bell, A. Gann, M. Levine, R. Losick. (2009). Biologie Moléculaire du gène. 6e Ed. Pearson Education- Paris-France. 688p.

Partie II : la traduction

Objectifs

- Connaître les caractéristiques du code génétique
- Connaître les éléments nécessaires à la traduction
- Connaître les étapes de la traduction
- Savoir établir le bilan énergétique de la traduction

Introduction

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.

1. Définition de la traduction

La traduction est un processus permettant la synthèse d'une chaîne polypeptidique (protéine) à partir d'un brin d'ARN messager par respect du code génétique. Elle s'effectue dans le cytoplasme de la cellule.

1.A. Le code génétique

Le code génétique est un code qui permet la conversion d'une séquence de nucléotides (ADN puis ARN) en séquence d'acides aminés (protéines). Le code implique les bases A, C, T et G ainsi que les 20 acides aminés. Le code génétique possède différentes caractéristiques :

- **Les codons sont des triplets de nucléotides** et ils codent pour un acide aminé.
- **La séquence du gène et la séquence de la protéine codée sont colinéaires**, c'est-à-dire que la longueur du gène et la longueur de la structure primaire de la protéine finale sont proportionnelles.
- **Le code génétique est universel**. En effet chaque acide aminé dispose d'un ou plusieurs codons et ceci au niveau d'une multitude d'organismes vivants procaryote et eucaryote. Cependant, depuis les années 1980, cette universalité a été remise en cause par quelques exceptions, découvertes d'abord dans l'ADN mitochondrial. Il y a en effet dans cet AND quelques codons qui varient.

Depuis, d'autres exceptions ont été découvertes chez des levures, des protozoaires. . .

- **Le code génétique est redondant (ou dégénéré)**. Plusieurs codons codent pour un même acide-aminé : on trouve 64 codons et 20 acides aminés. Souvent se sont les deux premiers nucléotides du codon qui définissent l'acide aminé, la redondance est donc due au troisième nucléotide du codon.
- **Le code génétique est non-chevauchant**. Les nucléotides d'un codon ne participe qu'au code d'un seul acide aminé, ainsi le prochain acide-aminé sera codé par le prochain codon présent sur l'ARNm. On parle du **cadre de lecture** (ou ORF Open

reading frame). 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 eucaryotes) 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, si non le cadre de lecture sera décalé (différente protéine, aucune protéine s'il y a introduction d'un ou plusieurs codons stop).

- **Le code possède un système de ponctuation.** Le codon d'initiation est le codon AUG (GUG dans la mitochondrie) et les codons de terminaison sont les codons UAA, UAG et UGA. Le codon UGA n'est pas présent au niveau de la mitochondrie.

Remarque :

Le cadre de lecture est défini par le codon d'initiation, ainsi le véritable codon de terminaison sera le premier codon qui sera dans le cadre de lecture imposé par ce codon d'initiation.

Parmi les 64 acides aminés, 3 sont des codons de terminaison ou codon stop les 61 restants sont des codons codant.

1.B. Les acteurs de la traduction

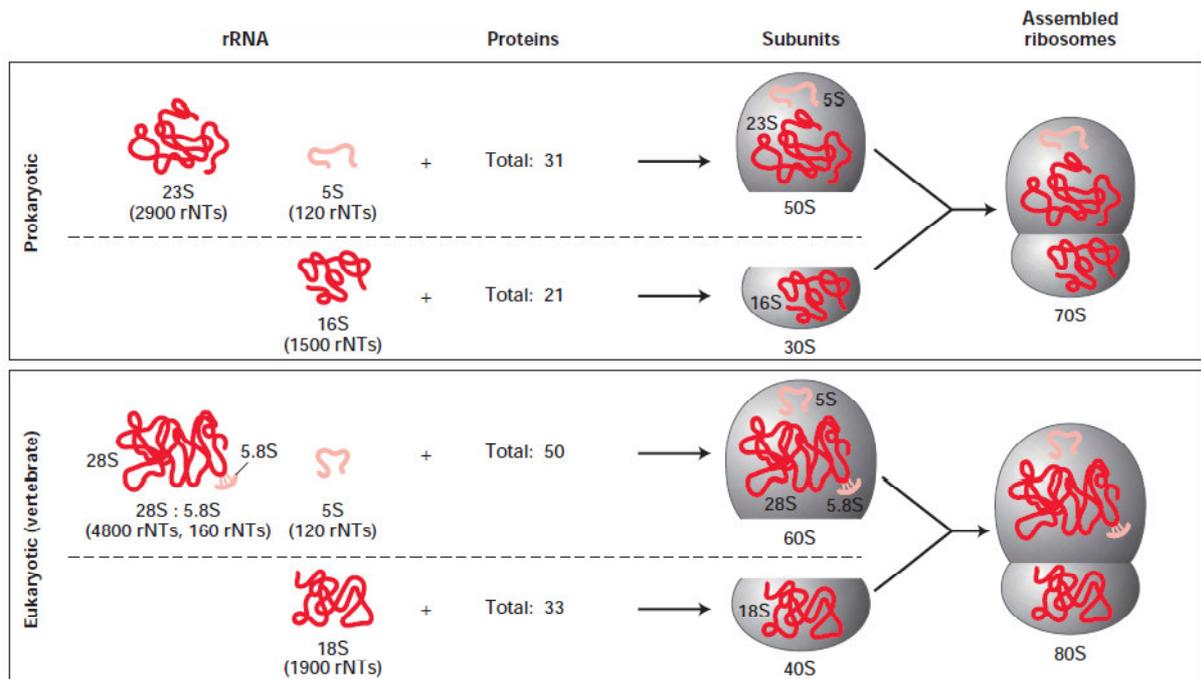
Les acteurs de la traduction sont l'ARN messager (ARNm), les ARN de transfert (ARNt), les ribosomes, les acides aminés, les aminoacyl ARNt synthétases, le Mg^{2+} , le GTP et l'ATP.

1.B.1 Les ribosomes :

Sont constitués d'ARN ribosomiques (ARNr) et de protéines et sont structurés sous forme de deux sous-unités que ce soit chez les procaryotes ou chez les eucaryotes. Leur taille est définie en unité Svedberg (unité de coefficient de sédimentation) (figure 1).

- Les ribosomes procaryotes (**70S**) sont constitués d'une petite sous-unité **30S** et d'une grande sous-unité **50S**.
 - La sous-unité 30S est constituée d'un **ARNr 16S** (1541 nucléotides) et de **21 protéines**.
 - La sous-unité 50S est constituée des **ARNr 23S** (2904 nucléotides) et **5S** (120 nucléotides) ainsi que de **32 protéines**.

-



- **Figure 1**: Structure générale des ribosomes procaryotes et eucaryotes (Aouf, 2016).
- Les ribosomes eucaryotes (**80S**) sont constitués d'une petite sous-unité **40S** et d'une grande sous-unité **60S**.
 - La sous-unité 40S est constituée d'un **ARNr 18S** et de **33 protéines**.
 - La sous-unité 60S est constituée des **ARNr 28S, 5,8S et 5S** ainsi que de **49 protéines**.

1.B.1.A Topographie schématique du ribosome bactérien

Le ribosome bactérien comporte des sites spécifiques :

- **Site A** : (= site Acide-aminé ou Accepteur) fixation des acides aminés.
- **Site P** : (= site Peptidique ou Donneur) fixation de f-Met.
- **Site E** : (= site Exit) sortie de l'ARN de transfert.

Chez les **eucaryotes** le premier acide aminé est la **méthionine** et non pas la **f-Met** présent chez les **procaryotes**.

1.B.1.B Topographie schématique du ribosome chez les eucaryote

Le ribosome chez les eucaryotes contient les trois sites de fixation pour les ARNt (figure 2) :

Le site A qui fixe le complexe aminoacyl-ARNt

Le site P qui fixe le complexe peptidyl-ARNt

Le site E pour exit qui fixe l'ARNt libre

Au niveau des ribosomes associés au réticulum endoplasmique les protéines en voie de synthèse pénètrent dans les vésicules du réticulum directement après le site E.

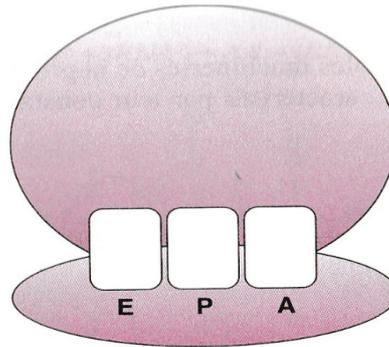


Figure 2 : les trois sites de fixation des ARNt dans le ribosome

1.B.2 Les ARNt

1.B.2.a Structure des ARNt

Les ARNt ont une structure secondaire en forme de **trèfle à 3 feuilles** et une structure tertiaire en forme de L à l'envers (voir figure 3). Lors du mécanisme de traduction il y a un appariement antiparallèle entre l'ARNm et l'ARNt : **reconnaissance codon-anticodon** au niveau de la boucle de l'anticodon.

Les ARNt possèdent également un bras de l'acide aminé qui le fixe en **3' (CCA)** sur le ribose, il s'agit d'une liaison covalente : liaison ester riche en énergie. Les acides aminés ne vont ainsi par arriver libre sur le ribosome mais associés à leurs ARNt respectifs. On trouve 40 à 60 ARNt différents par cellule, il existe donc plusieurs ARNt différents pour un acide aminé.

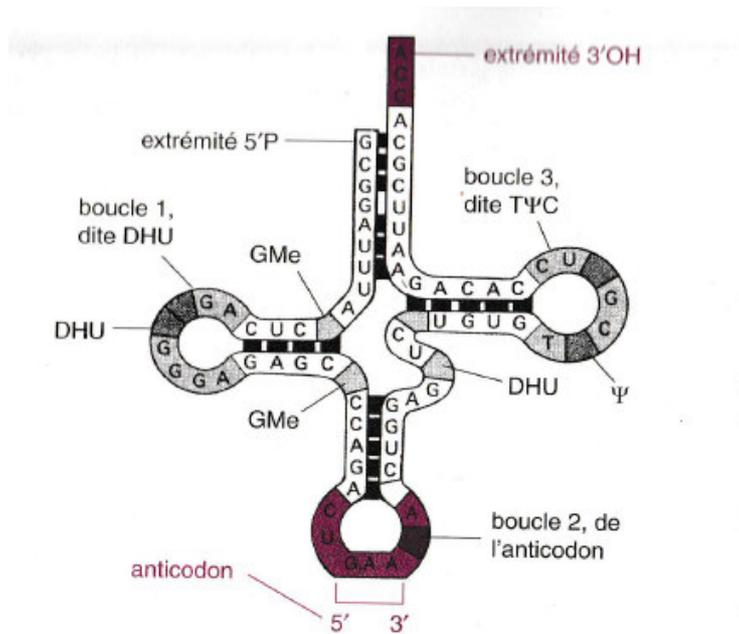
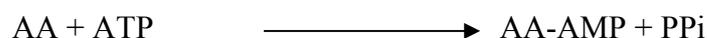


Figure 3 : Structures secondaire et tertiaire d'une molécule d'ARNt. La structure secondaire plane, en « feuille de trèfle », est formée d'une succession de tiges double brin stabilisées par des paires GC, et de boucles dont la longueur n'exède pas 10 bases (quelques bases modifiées ont été localisées). Cette forme plane s'organise dans l'espace pour donner une structure tertiaire compacte, telle que les boucles 1 et 3 se replient l'une sur l'autre et que la molécule prend l'aspect d'un L majuscule massif. Les deux parties fonctionnellement importantes de celle-ci : anticodon et le site d'accrochage de l'acide aminé en 3' OH, sont situées aux deux extrémités opposées.

1.B.2.b Chargement de l'acide aminé sur l'ARNt :

La formation du complexe **aminoacyl-ARNt (aa-ARNt)** nécessite une **Aminoacyl-ARNt-synthétase** spécifique de l'acide aminé, qui doit ainsi reconnaître toutes les formes de codon de cet acide aminé. Le chargement correct de l'ARNt est un élément important dans la traduction.

L'acide aminé (aa) est tout d'abord activé et cette activation nécessite de l'énergie sous forme d'**ATP** pour permettre la formation d'aa-AMP (liaison anhydride d'acide riche en énergie) (figure 4).



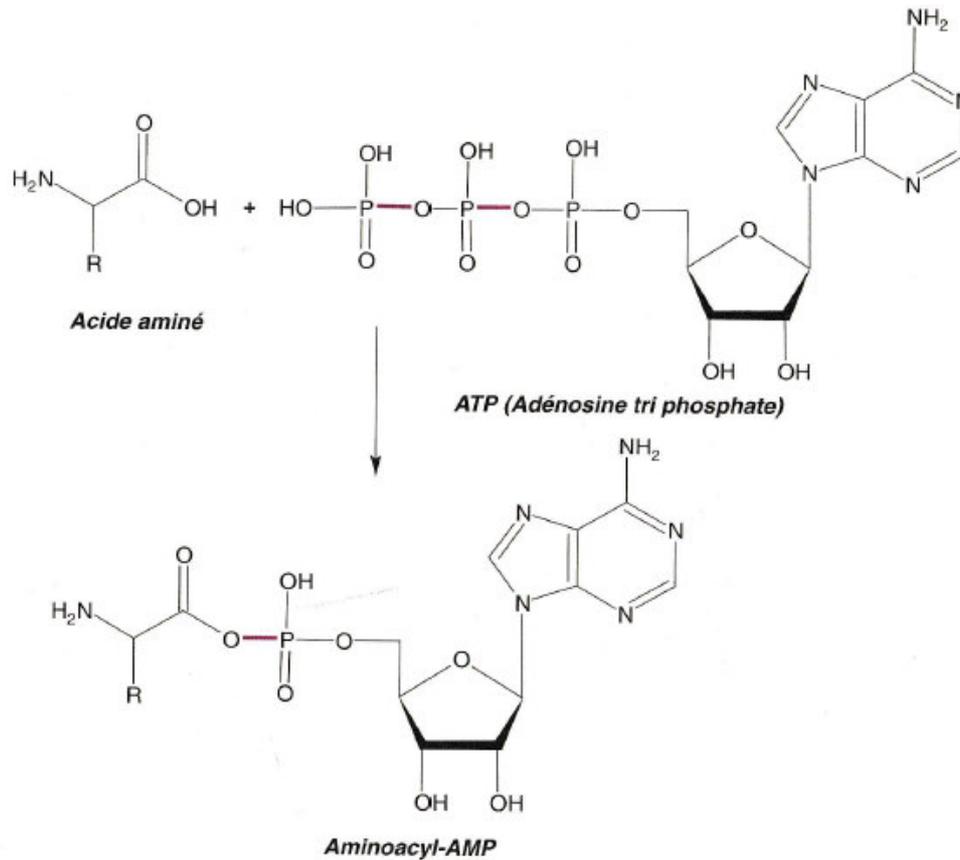


Figure 4 : formation de l' aminoacyl-AMP (en rose sont représentées les liaisons énergétiques)

La liaison formée entre l'ARNt et l'acide aminé est une liaison covalente de type carboxy-ester. Les Aminoacyl-ARNt-synthétase sont au nombre de 20 dans la cellule, autant qu'il y a d'acides aminés qui rentrent en compte dans la traduction. L'acide aminé complexé peut ainsi s'associer à la chaîne.

1.B.3 L'ARNm

C'est l'ARNm qui contient l'enchaînement des codons, et indique donc l'ordre dans lequel les différents AA devront être assemblés. Il sera lu dans le sens 5'- 3'. 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).

L'ARNm contient un ORF qui délimite le début et la fin de la séquence de l'ARNm qui va être traduit en peptide.

2. Les différentes étapes de la traduction procaryote

2.1 Initiation

Un ribosome reconnaît le début de la séquence codante, il utilise des signaux d'adressage en amont entre -8 et -13 du codon initiateur (AUG) qui correspond à la **séquence de Shine-Dalgarno** ou **RBS** (AGGAGG). Il y a appariement antiparallèle de bases entre l'ARNm et la petite sous-unité (30S) du ribosome, dû à une complémentarité de séquences entre l'ARNm et l'ARNr 16S (figure 5).

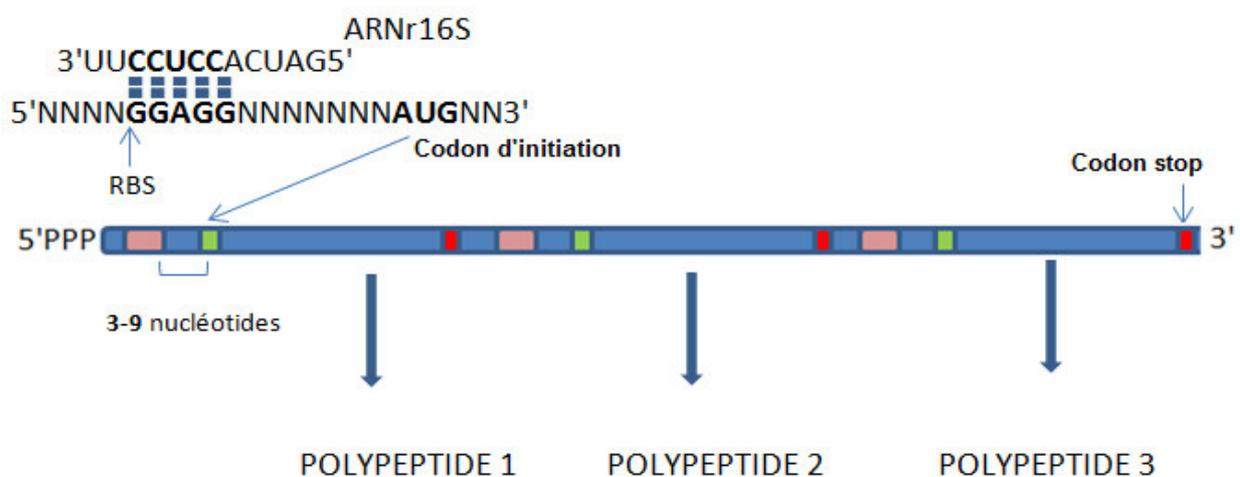


Figure 5: Structure d'un ARNm procaryote polycistronique. Chaque site de liaison au ribosome est indiqué RBS (Ribosome Binding Site). (Aouf, 2016)

Les bactéries nécessitent un acide aminé particulier pour l'initiation ; cet acide aminé est la méthionine et elle nécessite une **formylation** sur l'extrémité NH_2 (ajout d'un formyl) pour former la **f-Met**, c'est un phénomène **pré-traductionnel**. La particularité de conformation de son ARNt lui permet d'être placé directement dans le **site P** et non pas dans le site A (figure 6).

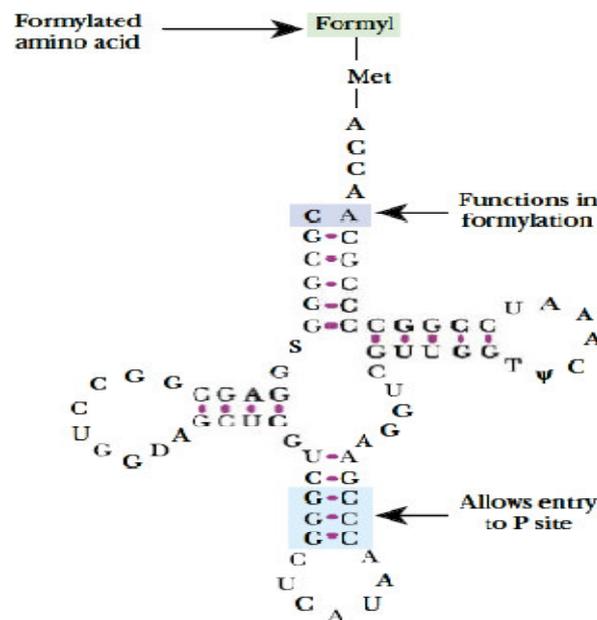


Figure 6: ARNt initiateur procaryote chargé par N-formyl-Méthionine. Les bases surlignées en bleu claire permettent la rentrée dans le site P du ribosome (Lodish *et al.*, 2003).

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 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. 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.

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 occupant le site P, et un site A libre.

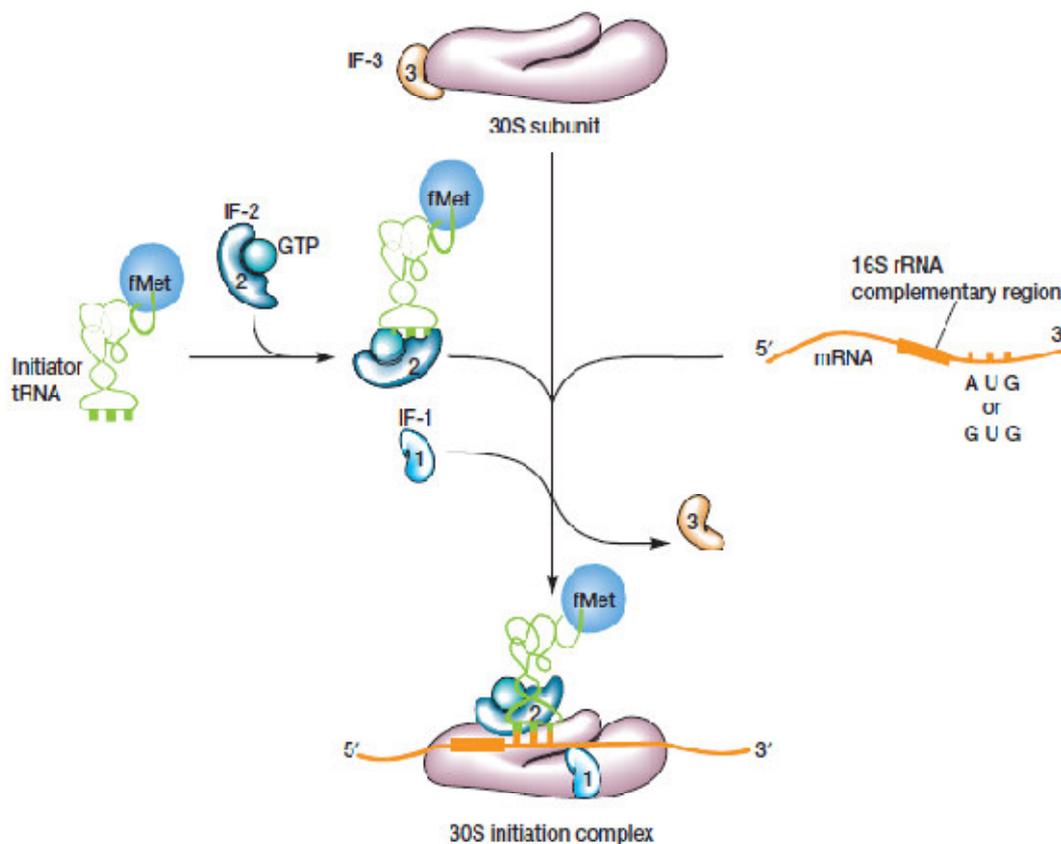


Figure 7-a: Le processus de l'initiation de la traduction chez les procaryotes (Prescott, 2002).

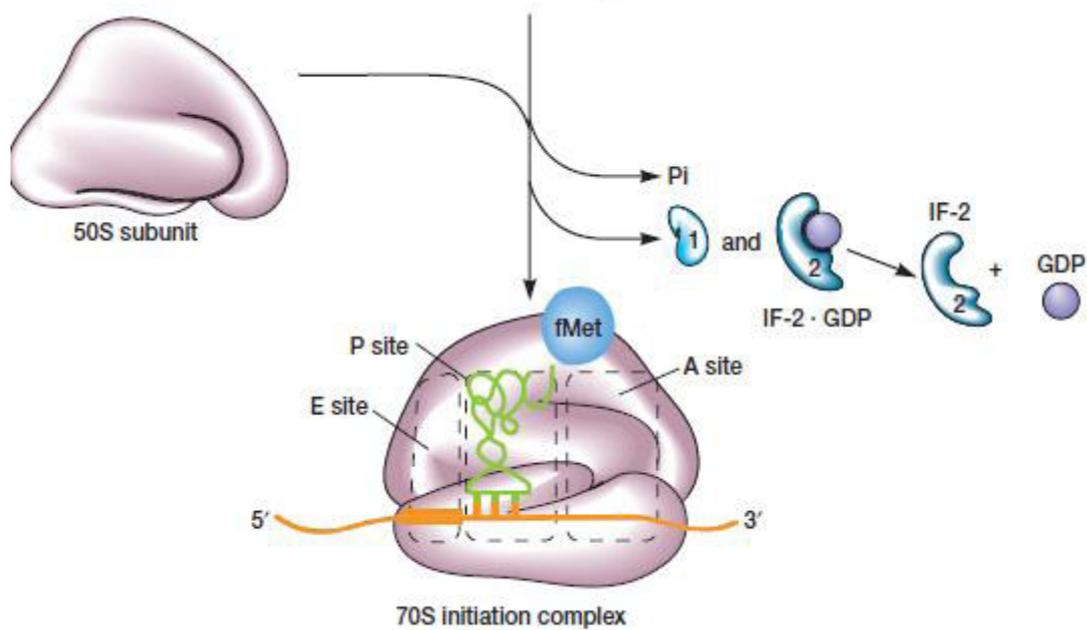


Figure 7-b: Le processus de l'initiation de la traduction chez les procaryotes (Prescott, 2002).

2.2 Elongation

L'élongation correspond à une synthèse protéique par ajout d'acides aminés à l'extrémité C-Terminale de la chaîne peptidique naissante, réaction catalysée par l'activité **peptidyl-transférase** de la grande SU des ribosomes. La lecture de l'ARNm par le ribosome se fait de 5' vers 3'. Il y a formation d'une liaison particulière appelée **liaison peptidique**, les deux fonctions formant la liaison étant portées par le carbone- α de deux acides aminés différents. Cette réaction entraîne l'élimination d'une molécule d'eau.

Deux protéines (facteurs) auxiliaires contrôlent l'élongation, ces 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**.

Pour chaque liaison peptidique formée on peut caractériser 3 étapes : **la réaction de couplage, la formation de la liaison peptidique et la translocation.**

2.2.a Réaction de couplage

L'étape de couplage correspond au transfert de l'acide aminé complexé à l'ARNt sur la chaîne protéique en voie d'élongation. On peut faire la remarque que durant la traduction c'est l'extrémité N-terminal (fonction amine) qui sort en premier du ribosome. Ainsi c'est l'extrémité C-terminale (fonction carboxyle COOH) du premier acide-aminé qui permettra la formation de la liaison peptidique avec la fonction amine NH₂ du deuxième acide-aminé. Ainsi le deuxième complexe aa-ARNt arrive dans le site A, la f-Met étant positionnée dans le site P.

2.2.b Formation de la liaison peptidique et libération du premier ARNt

La liaison riche en énergie qui lie le premier ARNt avec la f-Met se rompt amenant l'énergie pour permettre la formation de la liaison peptidique, ceci étant uniquement possible à ce moment là car la fonction COOH était jusqu'alors engagée dans la liaison à l'ARNt. L'ARNt est alors expulsé vers le cytoplasme où il sera recyclé, en même temps que la formation de la liaison peptidique, réaction catalysée par l'activité peptidyl-transférase de la grande sous-unité (ARNr 23S chez les procaryotes et l'ARNr 28S chez les eucaryotes).

2.2.c Translocation

La lecture de l'ARNm se fait de 5' vers 3', de ce fait le ribosome se déplace de 3 nucléotides (ou d'un codon) dans cette direction de telle sorte que l'ARNt portant les deux premiers nucléotides se retrouve dans le site P, il y a translocation. L'ARNt portant le troisième peut ensuite prendre place dans le site A à nouveau libre.

Ce cycle de trois étapes va donc se reproduire autant de fois qu'il est nécessaire jusqu'au codon stop (voir figure 8).

2.2.d Bilan énergétique au cours de l'élongation

L'énergie nécessaire à l'élongation est fournie sous la forme d'ATP et de GTP.

L'ATP apporte deux liaisons riches en énergie, et est consommé lors de l'activation de l'acide aminé, c'est-à-dire lors de la formation d'aa-AMP qui sera ensuite complexé par l'aminoacyl-ARNt-synthétase à l'ARNt correspondant. Il y aura ainsi une liaison riche en énergie entre l'acide aminé et son ARNt. L'énergie utilisée pour la formation de la liaison peptidique est l'énergie récupérée par la rupture de la liaison riche en énergie du complexe aa-ARNt

précédent et non pas de la liaison du complexe s'ajoutant à la chaîne peptidique en cours d'élongation.

- Le premier **GTP** est utilisé lors du positionnement du complexe aa-ARNt dans la loge A du ribosome. En effet l'acide aminé-ARNt forme un complexe avec EF-Tu, qui lui permet de se fixer sur le ribosome (site A). Après fixation du complexe, le GTP est hydrolysé en GDP et le facteur EF-Tu-GDP se détache.
- Le deuxième **GTP** est utilisé lors de la translocation, également par un facteur d'élongation ayant une activité GTPasique.

2.3 Terminaison

La terminaison de la traduction se fait au niveau d'un des 3 codons stop **UAA**, **UAG** et **UGA** qui ne code pour aucun acide aminé qui viendra se placer dans le site A, il aura alors rupture de la liaison entre le dernier acide aminé et le dernier ARNt. Trois facteurs de libération RF (*Releasing Factor* : 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.

La fixation des RF au codon non sens au site A transforme la peptidyl-transférase en une hydrolase, qui coupe la chaîne peptidique de l'ARNt auquel elle est fixée.

Hydrolyse de GTP est nécessaire pour la dissociation des facteurs RFs, des sous-unités ribosomales et du nouveau peptide (voir figure 9).

La terminaison fait intervenir, tout comme l'initiation, l'hydrolyse d'une molécule de **GTP**.

Remarque

La traduction bactérienne peut être inhibée par des antibiotiques tels que les aminosides qui inhibent la petite sous-unité ou les macrolides qui agissent au niveau de la grande sous-unité.

3. Les spécificités de la traduction eucaryote

Le ribosome est de taille différente et composé d'ARN ribosomiques différents bien que la structure générale et l'activité soit comparable.

Près de l'extrémité 5 (à quelque centaine de nucléotides), on trouve le codon initiateur est également **AUG** et c'est généralement le 1^{er} AUG présent sur l'ARNm. C'est lui qui porte le signal de début de traduction.

Vers 1980 **Kozak a met en évidence une séquence consensus encadrant le triple initiateur appelée séquence de Kozak (GCCGCC(A/G)CCAUGG)**. Une mutation sur la base A ou G (-3) ou en +4 (G) affecte fortement la traduction, voir l'empêche complètement.

Chez les eucaryotes le premier acide aminé est la méthionine qui sera le plus souvent enlevée juste après la synthèse de la chaîne peptidique (voir figure 10).

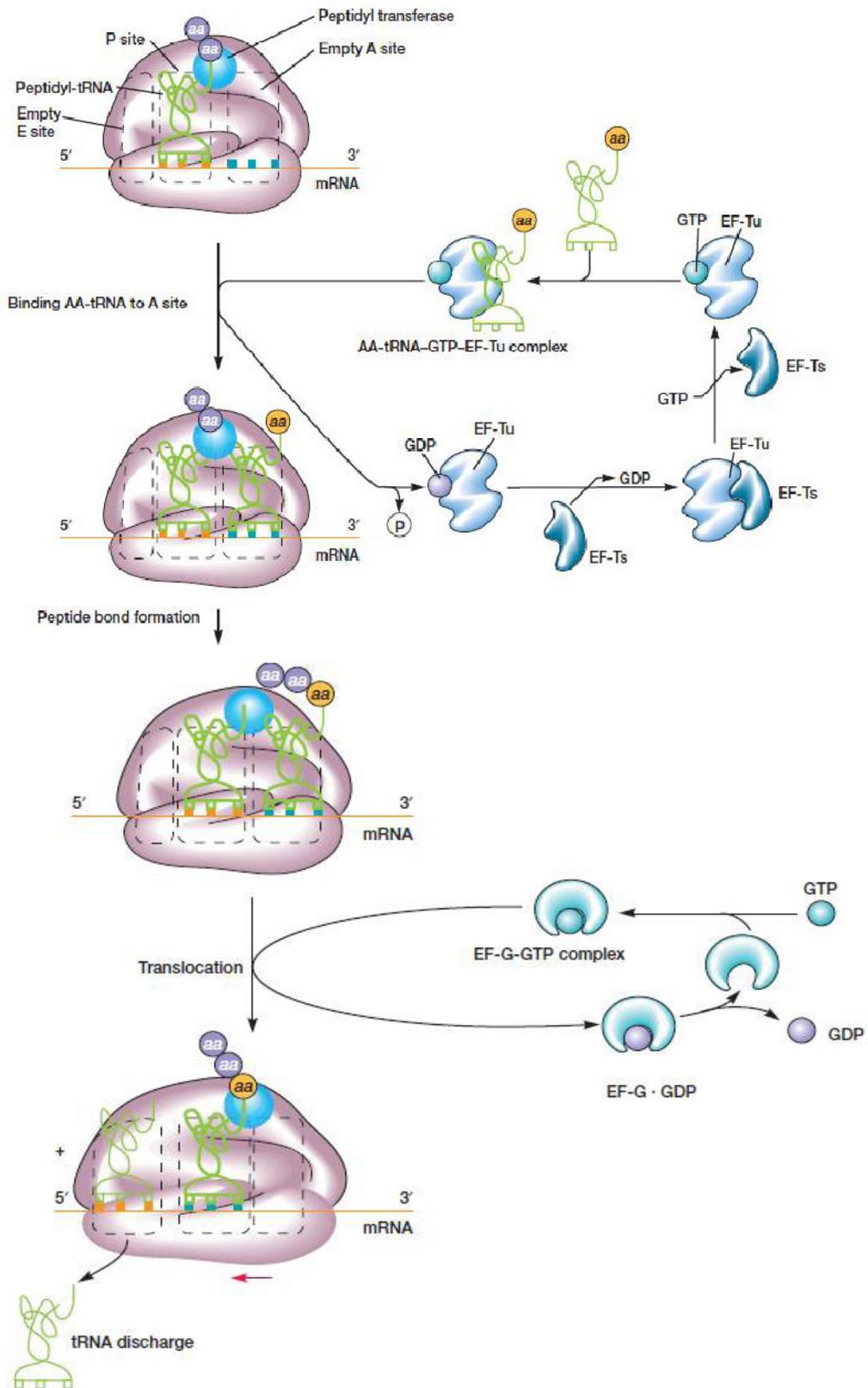


Figure 8: Etape d'élongation de la synthèse protéique chez les procaryotes. (Aouf, 2016)

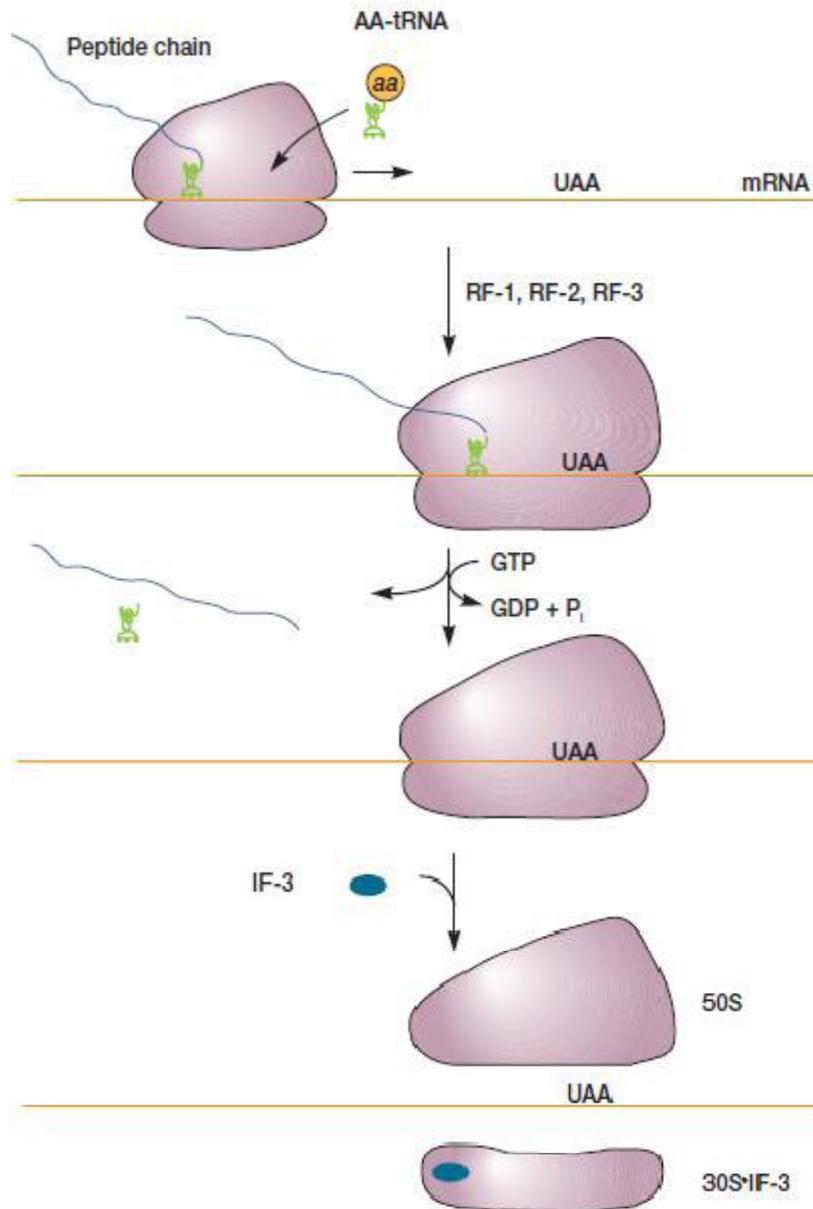


Figure 9: Étape de terminaison de la traduction chez les procaryotes (Prescott, 2002).

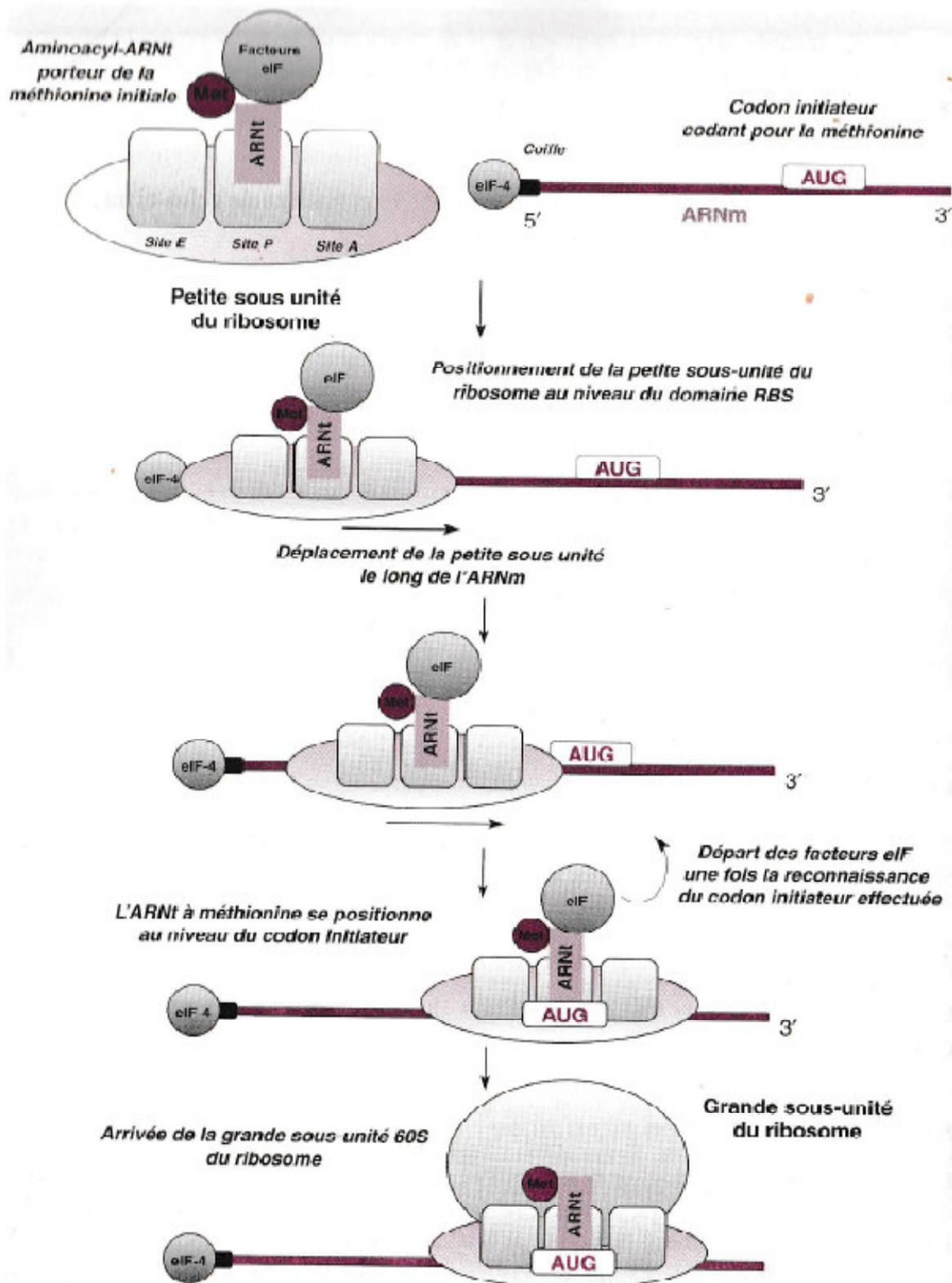


Figure 10 : principales étapes de l'initiation de la traduction avec le positionnement du ribosome sur l'ARNm chez les eucaryotes (l'hydrolyse de la molécule GTP en GDP n'a pas été placé sur le schéma).

4. Les antibiotiques inhibiteurs de la traduction chez les procaryotes :

- Streptomycine : qui inhibe la fixation de l'ARNt initiateur (N-formyl methionine)
- Gentamycine et la néomycine qui se fixe sur l'ARNr 16S des ribosomes et empêche ainsi l'appariement codon-anticodon
- Chloramphénicol qui inhibe l'activité peptidyl-transférase
- Acide fusidique ; utilisé principalement comme antibiotique ophtalmique, se lie à l'un des facteurs d'élongation et empêche ainsi la fixation des aminoacyl-ARNt

Bibliographie :

Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003). Molecular Cell Biology. 6th Ed. 937p.

Clark D. (2005). Molecular Biology: Understanding the Genetic Revolution. Southern Illinois University. USA Elsevier Academic Press. 784p.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 180-193p

J.-C. Callen, « Biologie cellulaire », Dunod, 2005

Prescott M. L., J. P. Harley, et D. A. Klein. (2002). Microbiology. 5th Ed. The McGraw-Hill Companies. 1139p.

Madigan M., J. Martinko. (2007). Biologie des microorganismes. 11e Ed. Pearson Education-Paris-France. 1047p.

Klug S. W., M. R. Cummings, C. A. Spencer & M. A. Palladino. (2013). Essentials of genetics. 8th Ed. Pearson Education. New Jersey. USA. 608p.

Filali-Maltouf Abdelkarim, Cours Biologie Moléculaire ; 2ème Partie, L'expression de l'information génétique et les signaux impliqués. Faculté des sciences. Université Mohamad V Rabat 50p.

Aouf Abdelhakim, (2016). Cours de Biologie moléculaire et génie génétique. Université Ferhat-Abbas Setif 1, 133p.

Chapitre 2 : régulation de l'expression génétique

I : régulation transcriptionnelle

Objectifs :

- Comprendre que la transcription peut être régulée à des niveaux différents
- Connaitre ces niveaux de régulation
- Connaitre les mécanismes précis de ces régulations

Introduction

Les cellules d'un même organisme multicellulaire sont dotées d'un patrimoine génétique identique, cependant elles diffèrent les unes des autres par leur transcriptome, leur protéome et leur métabolome qui dépendent de l'expression génique. Ainsi, le niveau d'information diffère en fonction du type cellulaire, du niveau de différenciation, du stade de développement et des différents stimuli perçus par la cellule. Ces variations de l'expression génique permettent à la cellule d'assurer des fonctions très diverses, notamment en ce qui concerne les métabolismes secondaires. Cette expression est soumise à une régulation temporelle et spatiale fine et complexe. L'acide désoxyribonucléique (ADN) est le support de l'information génétique, il contient des séquences codantes qui sont transcrites en acides ribonucléiques messagers (ARNm) par des ARN polymérases ADN dépendantes, avant d'être traduits en protéines. Ces étapes intermédiaires entre le gène et la protéine constituent une opportunité de choix entre les différents gènes à transcrire et constituent donc des étapes clés dans la régulation génique.

La modification transcriptionnelle des ARN est une étape importante du contrôle de l'expression des gènes, on considère généralement que les différentes modifications sont les éléments de régulation transcriptionnelle de l'expression des gènes. Ces différentes modifications peuvent influencer sur différentes caractéristiques de l'ARN, telle que sa stabilité, sa capacité à être traduit ou bien même modifier la séquence à traduire.

1. Définition des modifications transcriptionnelles :

Représentent l'ensemble des modifications qu'un ARN subit après avoir été transcrit. On parle également de **maturation de l'ARN**. Les modifications les plus connues sont l'ajout d'une coiffe, la polyadénylation ou encore l'épissage, mais il en existe bien d'autres.

La maturation des transcrits primaires a lieu dans le noyau de la cellule.

2. Modifications de l'ARN après transcription chez les procaryotes :

Elles diffèrent selon la nature de l'ARN :

- L'ARN_m ne subit pas de modification
- Les ARN_t et ARN_r subissent une maturation qui correspond à une coupure à partir de précurseurs :

1. Pour les ARN_t, le précurseur contient plusieurs séquences d'ARN_t et, dans certains cas, des séquences d'ARN_r.

2. Pour les ARN_r, la maturation implique aussi le clivage et l'élagage d'un précurseur contenant tous les ARN_r et des ARN_t.

Enfin, les bases de l'ARN peuvent subir de nombreuses modifications :

- méthylation de l'uracile la transformant en thymine, formation de 5-méthylcytosine, de méthyladénine, diméthyladénine.
- Réduction de l'uracile en dihydro-uracile
- transformation de l'uridine en pseudo-uridine.

Ces modifications touchent essentiellement les ARN_t et ARN_r et jouent un rôle très important pour la fonction de ces molécules.

Le transcrit primaire n'est pas utilisé tel quel pour la synthèse protéique (la traduction). Il doit subir des modifications qui répondent à plusieurs impératifs (augmentation de la demi-vie, modification de la séquence). Toutes ces modifications sont réalisées au fur et à mesure de la progression de la synthèse du préARN_m dans le nucléoplasme. Il existe trois grands types de modifications transcriptionnelles, catalysées chacune par des enzymes de nature protéique ou ribonucléique.

3. régulation transcriptionnelle chez les eucaryotes :

3.1 L'addition d'une coiffe en 5'(ou capping) :

Elle a lieu dès le début de la transcription avant que la chaîne ne compte plus de 30 nucléotides. Elle consiste en l'**ajout** d'un nucléotide à **guanine sur l'extrémité 5'** de l'ARN suivi de sa **méthylation sur l'azote 7 de la base** (provenant d'un GTP hydrolysé en GMP), ainsi que de la **méthylation en 2' du ribose des un ou deux premiers nucléotides** du transcrit primaire. La particularité de cet ajout consiste dans le type de liaison mis en jeu.

Cette coiffe a plusieurs rôles :

- Cet ensemble sert de coiffe protectrice à l'extrémité 5' de l'ARNm. Il en résulte que l'extrémité 5' de l'ARNm n'est pas porteuse des trois acides phosphoriques libres habituels, mais d'un GMP, ce qui limite la réactivité de cette extrémité et sa reconnaissance par les exonucléase (protection contre la dégradation).
- Elle est également nécessaire à l'exportation de l'ARNm vers le cytoplasme
- et à la liaison de ce dernier avec la petite sous-unité du ribosome lors de l'étape d'initiation de la traduction.
- Rendre l'ARN accessible au processus d'excision – épissage.

La coiffe est ajoutée grâce à un complexe protéique appelé « **Cap-Binding-Complex** » qui possédant une activité triphosphatase, une activité guanylyl-transférase et une activité méthyl-transférase.

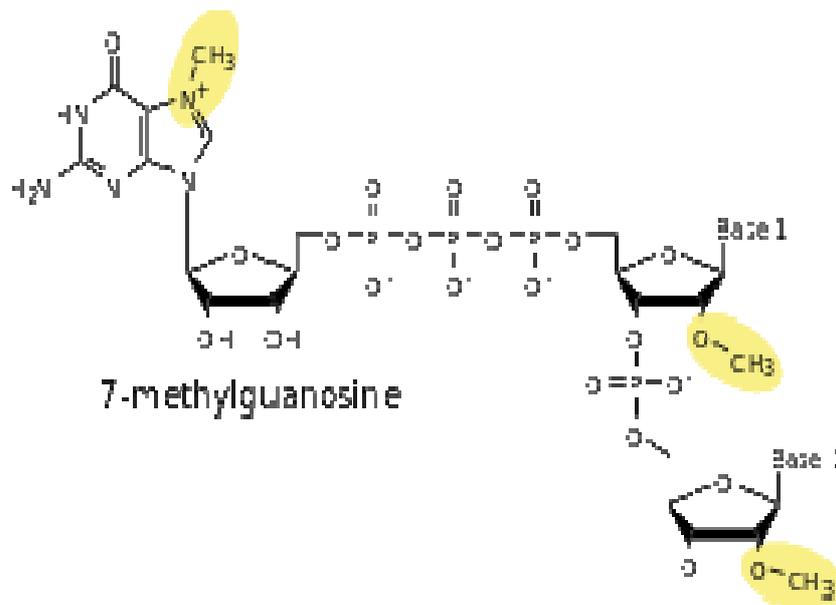


Figure 1 : Structure de la coiffe d'un ARN

La coiffe se caractérise par l'ajout d'un nucléotide à guanine à l'extrémité 5' du brin d'ARN. Ce nucléotide à guanine est relié à la chaîne nucléotidique par une liaison inhabituelle : au lieu d'être reliés par une liaison ester-phosphorique entre le groupement OH porté par le carbone 3' du ribose du nucléotide à guanine et l'acide phosphorique alpha du premier nucléotide de l'ARN natif, les deux nucléotides sont reliés par une liaison anhydride d'acide

entre les acides phosphoriques des deux nucléotides. Par la suite, le cycle imidazole de la guanine terminale est méthylé sur son azote 7. Par ailleurs, le ribose du premier nucléotide de l'ARN natif est méthylé sur l'oxygène porté par le carbone 2'. Ce peut également être le cas du nucléotide suivant.

3.2 L'addition d'une queue poly A en 3'

La poly-adénylation correspond à l'ajout de jusqu'à 200 adénines à l'extrémité 3' du transcrit primaire et ceci sans matrice par la **poly-A-polymérase (PAP)** qui utilise des ATP comme donneurs d'adénine.

Le mécanisme implique la prise en charge de deux séquences de l'ARN dont l'une, **AAUAAA** (séquence consensus retrouvée à la fin de tous les gènes), appelée signal de polyadénylation, est située approximativement à **15 nucléotides** en amont d'un site de clivage endonucléasique, qui servira ensuite de site d'addition de poly-A alors que l'autre, une séquence riche en **U** ou en **G/U** est située en aval du site de coupure. Ces deux séquences sont reconnues chez les mammifères par **deux complexes protéiques** interagissant l'un avec l'autre : **CPSF** (Cleavage and Poly-adenylation Specify Factor), reconnaissant AAUAAA et **CstF** (Clivage Stimulation Factor) reconnaissant la séquence riche en G/U. À cet ensemble vient s'ajouter la poly A-polymérase et deux autres protéines probablement responsables du clivage endonucléasique CFI et CFII.

Remarque : Certains ARNm, en particulier ceux codant pour les séquences des histones, ne possèdent pas de queue poly A.

La queue poly A aurait plusieurs rôles :

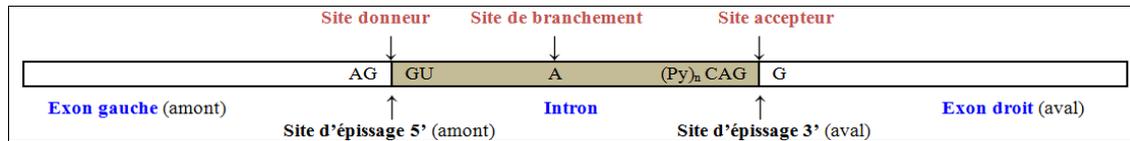
- Protection de l'ARNm lors de la traduction
- Aide au passage de l'ARNm du noyau vers le cytoplasme

3.3 L'excision-épissage

Chez les eucaryotes, les gènes sont morcelés : constitués d'une alternance d'**exons** (parties codantes du gène) et d'**introns** (parties non codantes, bornées par des séquences de bases spécifiques : 5'GU et 3'AG), ils sont d'abord intégralement copiés dans l'ARNprém, puis subissent une opération d'excision des introns (ainsi que celle parfois de petits morceaux d'exons) suivi d'un épissage (**splicing**), c'est à dire la réunion bout à bout des exons restants

qui constituent l'ARNm. Ce remaniement se déroule au fur et à mesure de la progression de la transcription.

Il faut en effet que cet épissage soit parfait, car une erreur d'un seul nucléotide aboutirait à un changement du cadre de lecture, et donc à une protéine totalement différente.



Les jonctions d'épissage sont reconnues par les **snRNP** (ou **snurps** pour *Small-Nuclear-Ribonucleo-protein-Particules*). Les snRNP correspondent à l'association de snRNA (snRNA **U1, U2, U4, U5, U6**) et de protéines et l'ensemble des snRNPs s'appelle le **spliceosome ou complexe d'épissage**. La taille de ces snRNA varie de 100 à 300 nucléotides chez la plupart des eucaryotes.

Les snRNP jouent trois rôles dans l'épissage :

- Reconnaittent 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 séquences de début de l'intron (**GTAAGT** dans l'ADN, **GUAAGU** dans l'ARN) et de fin de l'intron (**AG**) sont invariantes. De plus, il existe toujours une **adénine** située près de l'extrémité 3' de l'intron située à environ -20 à -30 nucléotides. Ces séquences sont reconnues dans l'ARN prémessager par différentes RNPsn au sein du complexe d'épissage. En fonction de leur rôle dans le processus d'épissage, les séquences de début d'intron, de fin d'intron et celles où se trouve l'adénine sont respectivement appelées : site donneur, site accepteur et point de branchement. La fixation des RNPsn s'effectue grâce à des appariements entre bases de l'ARN messager et de l'ARN de la ribonucléoprotéine.

Le snRNP U1 reconnaît le site donneur et le snRNP U2 reconnaît le site de branchement et le site accepteur.

L'excision des introns et l'épissage des exons se fait en plusieurs étapes :

- Le **snRNP U1** permet la reconnaissance du site donneur d'épissage et entraîne la rupture de la liaison phosphodiester entre le premier exon et l'intron.
- Cette rupture de la liaison **phosphodiester** entraîne la formation d'un **lasso**, qui n'est autre que l'extrémité 5' de l'intron. Ce lasso forme une liaison avec le site de branchement, lui-même situé sur le même intron qui se replie ainsi sur lui-même.
- U2 se lie au point de branchement, puis les protéines U4 et U6 s'associent à U2 ; le **site de branchement** est reconnu par le **snRNP U2** et permet la liaison par l'intermédiaire d'une adénosine.
- U5 se fixe près du site accepteur.
- Le **snRNP U2** permet également la reconnaissance du **site accepteur** d'épissage. Suite à cette reconnaissance il y a rupture de la liaison phosphodiester au niveau de l'extrémité 3' de l'intron.
- Le groupement 3'OH du premier exon peut ainsi réagir avec l'extrémité 5'phosphate du deuxième exon pour former une liaison phosphodiester et permettre la libération de l'intron qui sera dégradé par des **ribonucléases** (figure 2).

Remarque :

- Il existe certains ARN qui possèdent des introns auto-catalytiques qui ne nécessitent ainsi aucune protéine, l'activité enzymatique est portée par l'ARN lui-même, on parle de **ribozyme**. Elle se trouve dans quelques gènes des organites eucaryotes et de procaryotes (Archaea) et ARNr nucléaire, 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. Ce ribozyme est donc une endoribonucléase spécifique d'une séquence donnée et réalise une réaction analogue à celle du splicéosome.

- Au cours de cette réaction (épissage) que le nombre de liaisons phosphodiester reste le même, ce qui explique que ce mécanisme ne nécessite pas l'apport d'ATP.

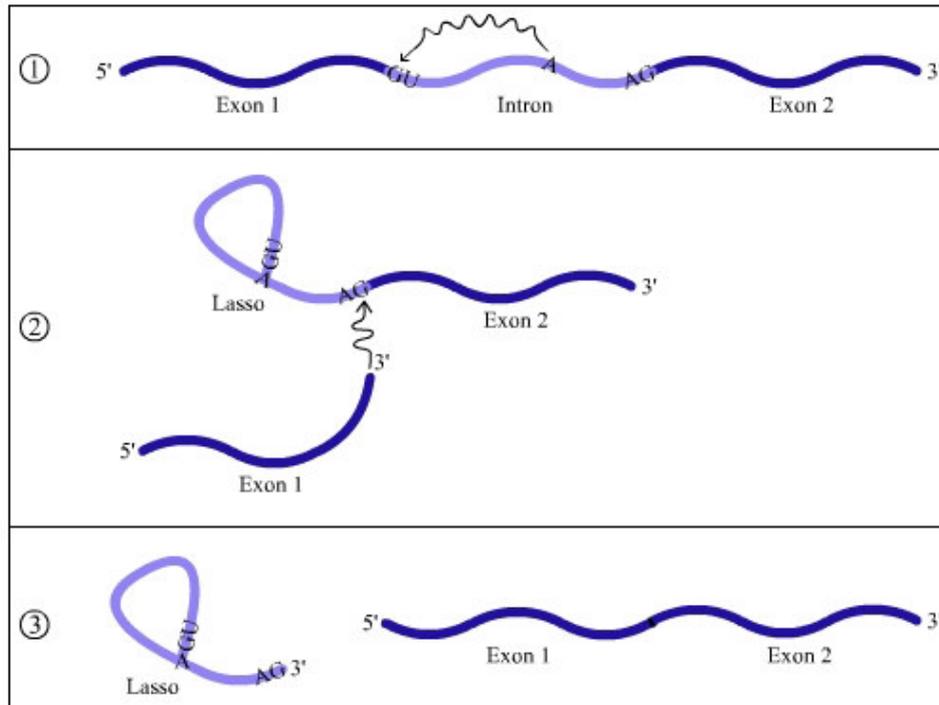


Figure 2 : Excision de l'intron par formation d'un lasso

1^{re} étape : clivage du côté 5' de l'intron par l'attaque de l'hydroxyle OH situé en 2' de L'adénine du branchement sur le groupe phosphodiester (cette attaque est rendue possible par des phénomènes de repliement dans la chaîne). IL se forme alors une Liaison 2'-5' entre L'adénine et la guanine à l'extrémité de l'intron qui crée un Lasso.

2^e étape : le 3'OH de La guanine en fin de l'exon 1 attaque le groupe phosphodiester qui relie la guanine de L'exon 2 à l'intron du côté 3', ce qui provoque le clivage de cette Liaison.

3^e étape : il se produit alors une soudure des deux exons, et libération du Lasso qui sera dégradé par des nucléases cellulaires.

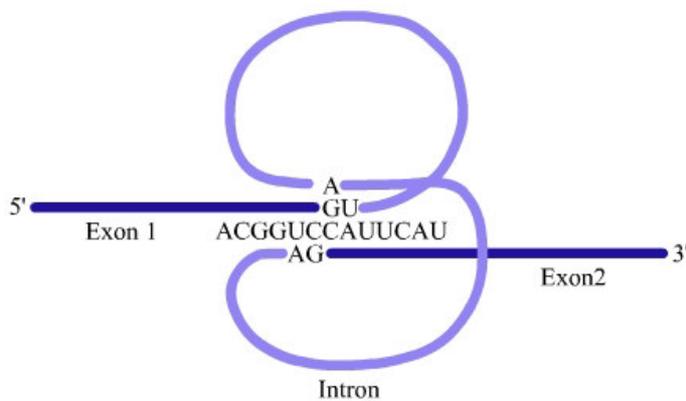
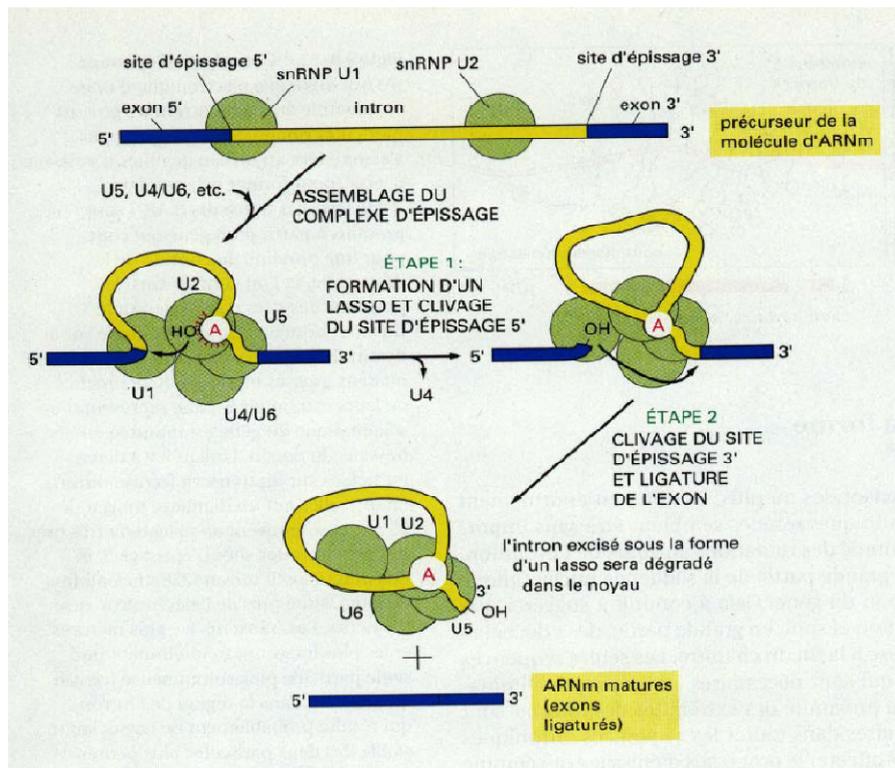


Figure 3 : Excision de l'intron par formation d'un lasso : le splicéosome

3.4 L'épissage alternatif

A partir d'un transcrit primaire on peut avoir deux ou plus ARNm matures qui seront à l'origine de la formation des protéines-isoformes. Ceci est possible grâce à l'épissage alternatif qui consiste en **l'élimination de certains exons**. En effet certains exons sont constants au niveau des différents ARNm matures et d'autres sont variables et spécifiques du tissu dans lequel se trouve la protéine isoforme.

Exemple :

Au niveau du gène de la tropomyosine on met en évidence 12 exons au total dont 7 sont constants et 5 sont alternatifs (voir figure 4).

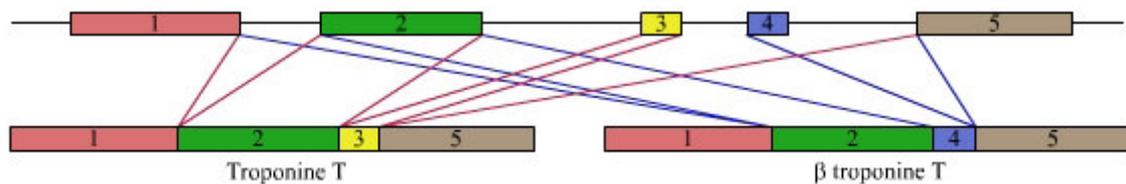


Figure 4 : Exemple d'épissage alternatif

A partir du même préARNm peuvent être obtenus deux ARNm matures différents codant pour deux isoformes de la troponine, une molécule impliquée dans la structure des myofibrilles.

4. Structure des facteurs protéiques de la transcription

La régulation de la transcription des gènes est assurée en partie par l'intervention de facteurs protéiques agissant en « trans ». Certains de ces facteurs sont capables de se lier directement à l'ADN et d'autres interviennent dans des complexes protéiques sans se lier à l'ADN. L'association de l'ensemble de ces protéines en complexes de haut poids moléculaire permet de réguler la transcription du gène en fonction des signaux cellulaires.

Ces protéines nucléaires sont capables de reconnaître et de se lier avec des régions promotrices de gènes cibles. Deux types de facteurs de transcription participent à la régulation de l'expression des gènes chez les eucaryotes : les facteurs généraux de la transcription qui interagissent avec la machinerie basale de la transcription au niveau du promoteur central et

les facteurs de transcriptions spécifiques interagissant avec les régions distales des promoteurs.

Les facteurs de transcription possèdent généralement deux domaines caractéristiques : un site de liaison direct à une séquence spécifique d'ADN et un domaine d'activation ou de répression qui permet au facteur de transcription d'interagir avec des cofacteurs pour contrôler (activer ou réprimer) l'activation de la machinerie basale.

4.1. Structure des activateurs

Il faut distinguer dans ces facteurs le domaine de liaison à l'ADN (noté DBD pour DNA Binding Domain), et un ou plusieurs domaines d'activation.

4.1.1 Protéines à Leu Zipper

Il s'agit de deux hélices α formant ainsi un dimère, reliées entre elles par des liaisons impliquant des AA hydrophobes tels que la leucine, ce qui crée des liaisons faibles susceptibles d'être assez facilement rompues puis recréées, ce qui évoque le système d'une fermeture éclair (Figure 5). En amont de cette région « Leu Zipper », on trouve une zone riche en AA basiques (chargés positivement) qui vont interagir avec les charges négatives de l'ADN, la encore par des liaisons faibles. C'est principalement par cette région que la liaison à l'ADN se fera.

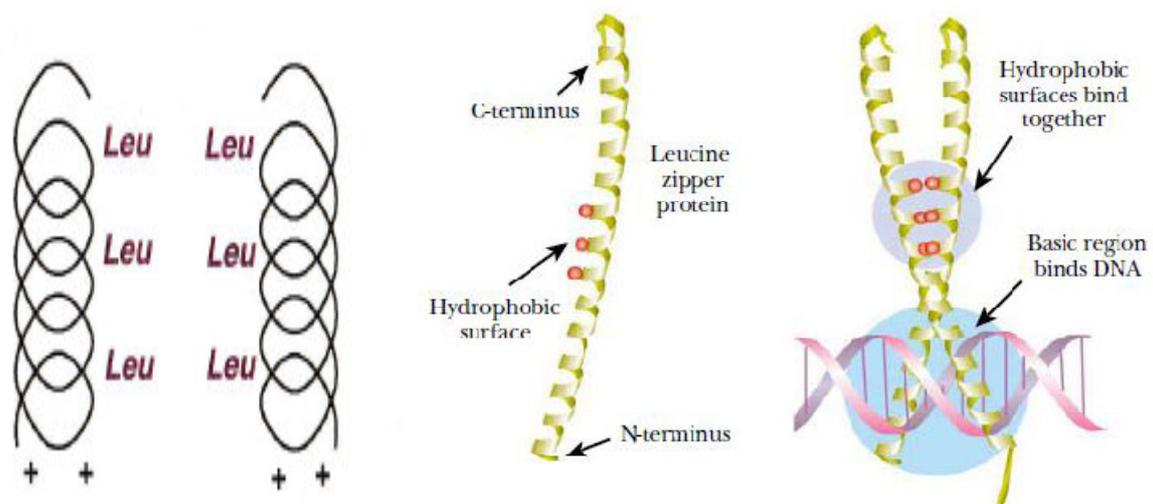


Figure 5: structure des Leu-Zipper (Aouf, 2016)

4.1.2 Protéines à doigts de zinc

Il s'agit d'une chaîne peptidique repliée sur elle-même en boucle sur une trentaine d'AA, stabilisé par un ion Zn^{+2} situé à la base du doigt (figure 6). Cet ion échange des liaisons de coordination avec 4 AA eux aussi situés à la base du doigt (cystéines ou histidine). Ceci stabilise le reste du doigt qui entre en interaction avec l'ADN par l'intermédiaire d'AA hydrophobes chargés positivement.

L'utilisation du zinc dans les structures biologiques est assez rare, et on pense d'ailleurs que c'est pour cela qu'il est utilisé ici ; en effet, ne faisant pas partie de chaînes d'oxydoréduction « classiques », cet ion zinc ne risque pas d'interagir avec l'ADN et d'y provoquer éventuellement des mutations.

On peut noter que certaines protéines possèdent plusieurs de ces structures en doigt de zinc, comme les hormones stéroïdiennes qui en comptent deux.

Certains doigts de zinc sont formés avec quatre cystéines, d'autres avec un mélange de cystéine et d'histidine (2 à 2).

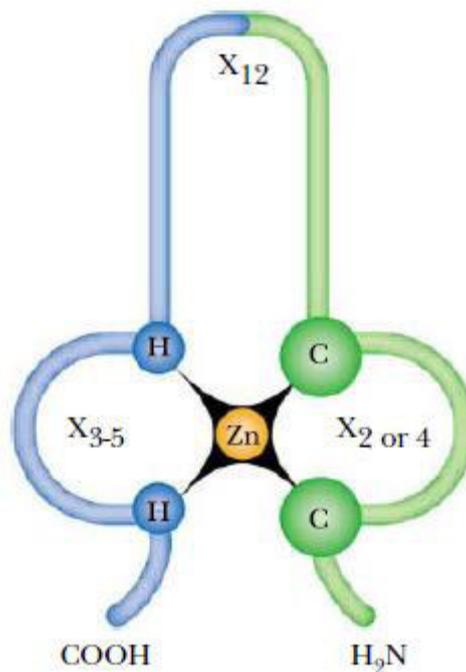


Figure 6: structure des doigts de zinc (Aouf, 2016)

Les protéines présentant des motifs en doigts de zinc permettent d'activer la transcription en se liant à des régions riches en GC.

4.1.3 Protéines à homeodomains

Ces protéines ont été tout d'abord rencontrées comme facteurs de régulation de gènes

impliqués dans le développement embryonnaire, puis dans d'autres facteurs.

Leur structure est du type hélice tour — hélice (hélice α — tour β — hélice α), dont une des chaînes interagit grâce à ses charges positives avec les charges négatives du grand sillon de l'ADN. Il semble que la structure contienne au total trois hélices α et deux coudes β et que ce soit l'hélice n° 3 qui interagisse principalement avec le grand sillon de l'ADN par des charges positives.

Les protéines contenant des motifs HD reconnaissent généralement des structures d'ADN riches en bases AT.

Ces facteurs protéiques sont codés par des gènes dits homéobox, et sont donc impliqués dans le développement, mais également dans la forme du corps chez les vertébrés.

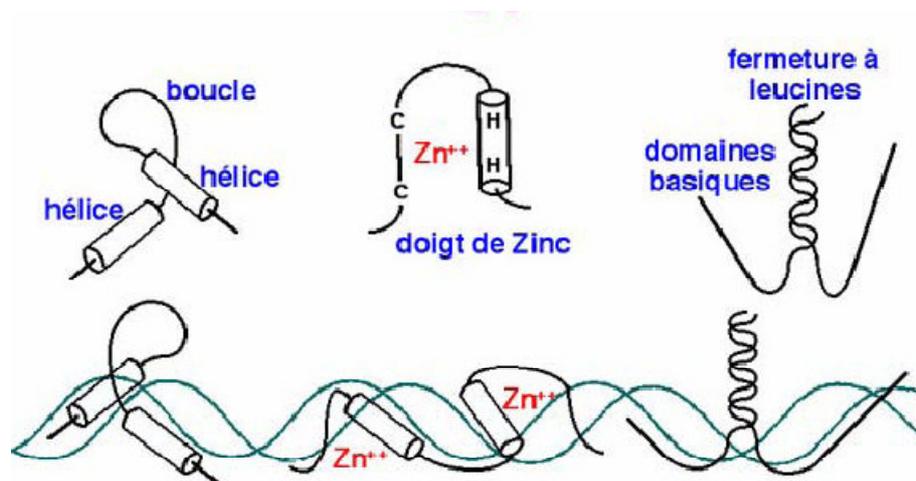


Figure 7 : Représentation schématique de domaines de liaison à l'ADN couramment rencontrés.

Bibliographie :

PERRIN Pascale. (2010), Le contrôle de l'expression génétique FLBI399. Université Montpellier 2-53p.

Cyril Esnault (2007), Étude des interactions entre le Médiateur et les facteurs généraux de la transcription par l'ARN polymérase II. Thèse pour obtenir le grade de DOCTEUR de l'université PARIS XI ORSAY 8-35p.

Carl Herrmann. (2006), la régulation de la transcription. Institut de Biologie du Développement Marseille Luminy. Université de la Méditerranée 56p.

HOMMEL Marie, (2007). La régulation transcriptionnelle de l'expression génique dans le fruit de tomate : caractérisation fonctionnelle de promoteurs fruit-spécifiques et d'un cofacteur de la transcription de type MBF1.thèse Présentée pour obtenir le titre de docteur de l'institut national polytechnique de Toulouse 16-25p.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 180-193p

Fainsod A, (1986) Bogarad LD, Ruusala T, Lubin M, Crothers DM, Ruddle FH. The homeo domain of a murine protein binds 5' to its own homeo box. Proc. Natl. Acad. Sci. USA., 83: 9532-6.

Mitchell PJ and Tjian R. (1989) Transcriptional regulation in mammalian cells by sequencespecific DNA binding proteins. Science, 245: 371-8.

Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003).Molecular Cell Biology. 6th Ed. 937p.

Aouf Abdelhakim, (2016). Cours de Biologie moléculaire et génie génétique. Université Ferhat-Abbas Setif 1, 133p.

II : La régulation de l'opéron lactose (régulation de la synthèse des protéines procaryotiques)

Objectifs :

Comprendre la régulation génétique chez les bactéries et son rôle.

Savoir la notion de l'opéron.

Connaître le principe du fonctionnement d'un opéron catabolique.

Introduction

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

1. Qu'est-ce-qu'un opéron ?

Le contrôle concerté de l'expression des gènes est essentiel pour le maintien équilibré de la croissance cellulaire. Ce contrôle permet à la cellule d'ajuster ses synthèses aux conditions environnementales. Chez la bactérie *Escherichia coli*, les gènes impliqués dans un processus métabolique sont souvent groupés sur le chromosome et organisés en unité de transcription appelée opéron. Le contrôle coordonné de l'expression de ces gènes est possible grâce à des protéines régulatrices. Elles régulent le taux de transcription en se liant à l'ADN au niveau de séquences spécifiques. Dans ce cas, l'ARNm est polycistronique, c'est-à-dire qu'il contient l'information nécessaire à la synthèse des différentes protéines.

2. Le métabolisme du lactose :

La bactérie trouve sa source de carbone dans le catabolisme des sucres. Si le glucose est la source de carbone "préférée", le lactose (qui est un β -galactoside) peut également être consommé par la bactérie et métabolisé en galactose et glucose. Les enzymes nécessaires à l'utilisation du lactose ne seront synthétisées qu'en présence de ce substrat. La lactose perméase permet l'entrée dans la cellule du lactose avec un flux de protons et la β -galactosidase qui hydrolyse le lactose en hexoses qui suivent d'autres voies métaboliques.

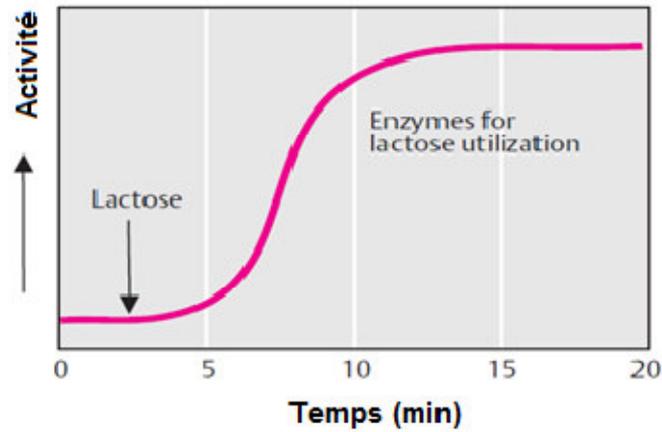


Figure 1: Cinétique de l'induction par le lactose.

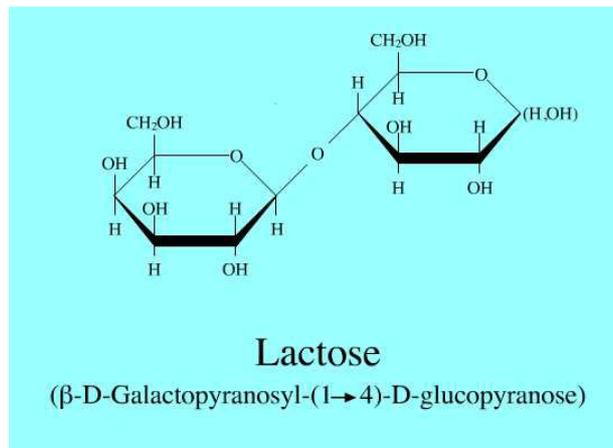


Figure 2 : Lactose

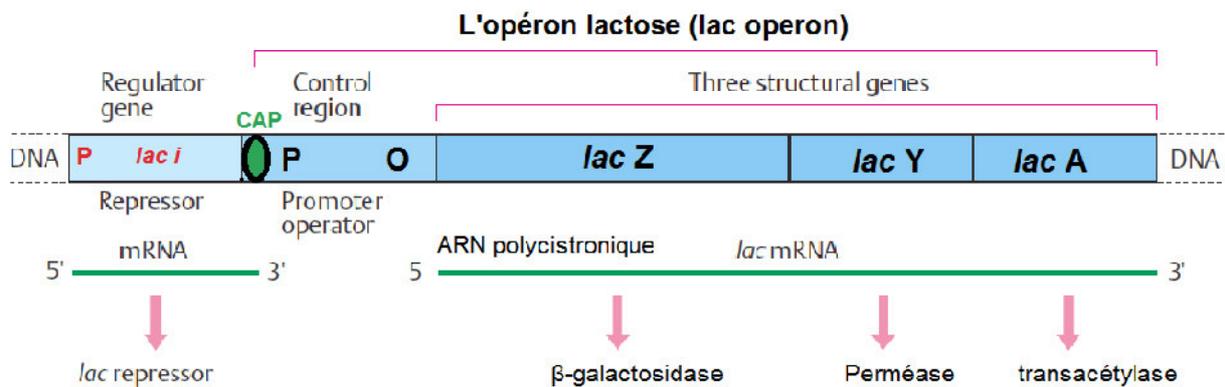


Figure 3: Organisation de l'opéron lactose d'*E. coli* (Passarge, 2007).

Dans l'opéron lactose, on trouve les trois gènes indispensables à la dégradation du lactose. Ils codent :

- La β -galactosidase (gène *lacZ*). Elle hydrolyse la liaison β 1-4 osidique des β -galactosides.
- La lactose perméase (gène *lacY*). Cette protéine membranaire permet l'entrée du lactose dans la cellule.
- La thiogalactoside transacétylase (gène *lacA*). Son rôle n'est pas bien connu. Elle acétyle les β -galactosides non métabolisables qui peuvent alors être éliminés hors de la cellule par diffusion à travers la membrane plasmique.

Ces trois gènes de structure sont précédés par une région responsable de la régulation de leur expression. Cette région régulatrice comprend **le promoteur et l'opérateur**. On trouve également en amont de l'opéron lactose, le gène régulateur (*lacI*) qui code une protéine régulatrice. Celle-ci agit en inhibant l'expression des gènes de l'opéron lactose par transactivation en se liant spécifiquement sur l'ADN au niveau de l'opérateur. L'expression de ce répresseur est constitutive, c'est à dire qu'il est exprimé quelque soient les conditions de croissance de la bactérie. Par contre, son affinité pour l'opérateur sera modifiée en présence de lactose.

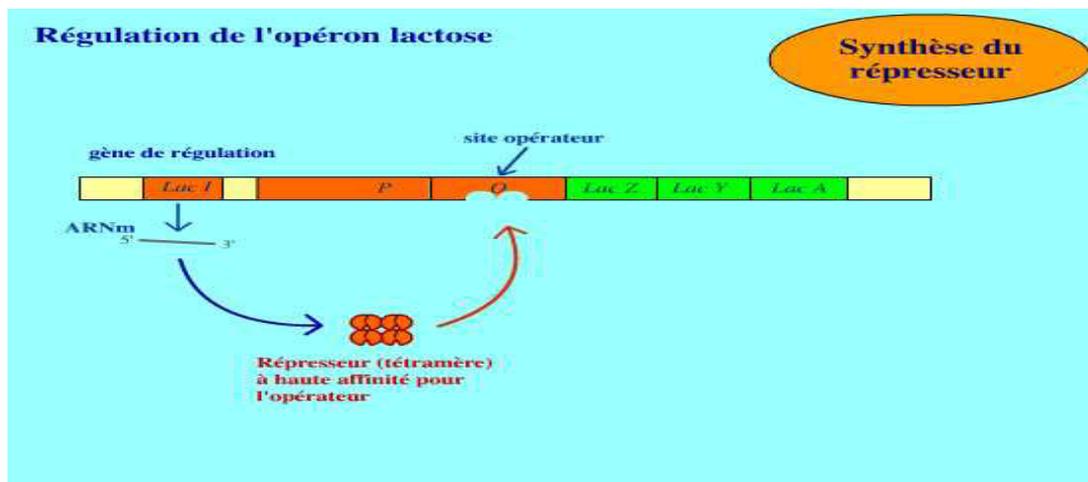


Figure 4 : synthèse du répresseur

3. Régulation négative de la transcription :

En absence de lactose, le répresseur est sous sa forme active. Il va se lier spécifiquement au niveau de l'opérateur de l'opéron lactose bloquant l'accès de l'ARN polymérase au site

d'initiation de la transcription. Ainsi, il y a régulation négative de la transcription des gènes de l'opéron lactose. Les enzymes nécessaires au métabolisme du lactose ne sont pas synthétisées car inutiles en absence de lactose.

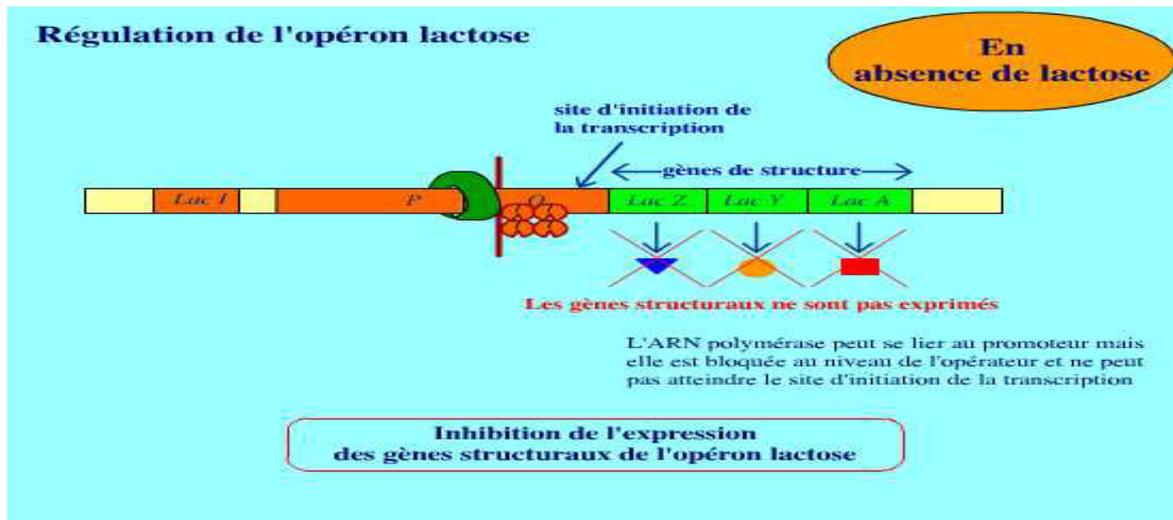


Figure 5 : inhibition de l'expression des gènes structuraux de l'opéron lactose

4. Levée de l'inhibition de la transcription :

En présence de lactose, c'est l'allolactose, un isomère du lactose, qui va jouer le rôle d'inducteur en se liant au répresseur pour l'inactiver. Cette liaison entraîne un changement conformationnel du répresseur qui perd alors son affinité pour l'opérateur. Le site opérateur étant libéré, l'ARN polymérase peut atteindre le site d'initiation de la transcription et synthétiser l'ARN polycistronique. La production des enzymes nécessaires au métabolisme du lactose est donc dépendante de la présence du substrat.

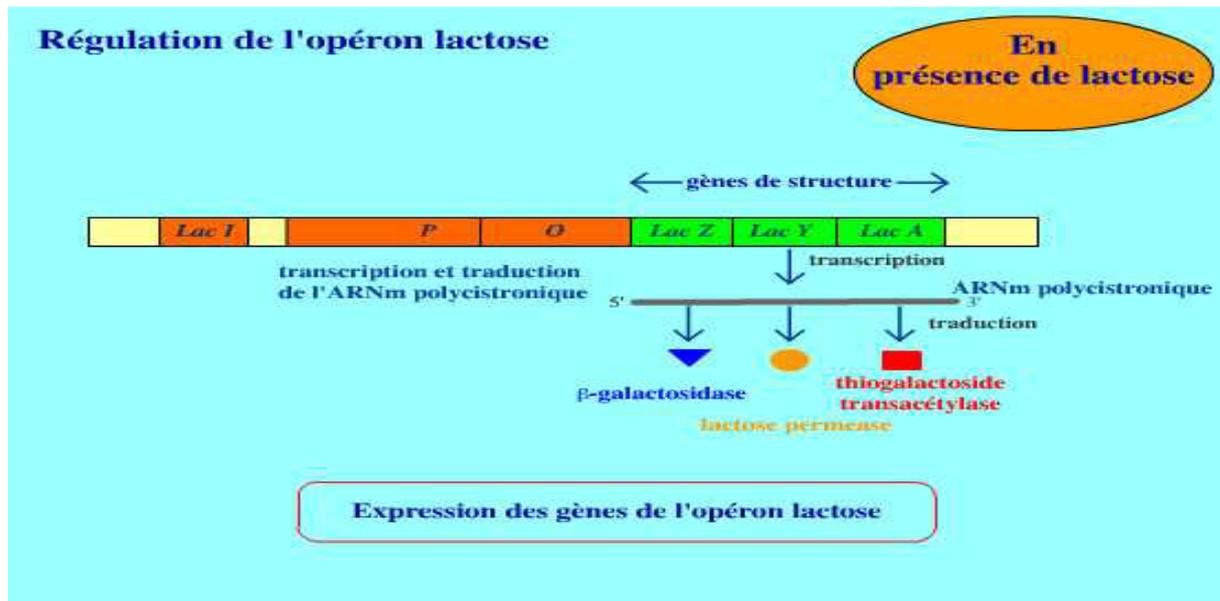


Figure 6 : expression des gènes de l'opéron lactose

5. Répression catabolique, régulation positive

Lorsque les bactéries, par chance, disposent en même temps de glucose et de lactose, la situation se complique !

Le répresseur de l'opéron lactose est inactivé par l'allolactose, le site opératoire de l'opéron est donc libre et les gènes pourraient être transcrits. Or, tant que du glucose est présent, la bactérie va le métaboliser préférentiellement : elle n'a donc pas besoin des enzymes nécessaires au métabolisme du lactose. Ceci implique l'existence d'un autre mécanisme de régulation que l'on appelle la répression catabolique. Ce n'est que lorsque la concentration en glucose diminue que le métabolisme du lactose devient nécessaire. Un signal de carence alimentaire est alors déclenché sous forme d'une augmentation du taux d'AMPc. Cet AMPc forme un complexe avec la protéine CAP (pour Catabolite gene Activator Protein, ou CRP pour cAMP Receptor Protein). Ce complexe se lie à l'ADN en amont du site de fixation de l'ARN polymérase. L'interaction du complexe CAP-AMPc va agir comme un inducteur et augmenter l'affinité de l'ARN polymérase pour le promoteur de l'opéron. Cette régulation positive peut permettre d'augmenter d'un facteur 50 la transcription de l'opéron lactose. On comprend alors qu'en présence de glucose, il n'y a pas de complexes CAP-AMPc disponibles : le niveau de transcription de l'opéron lactose est donc très faible.

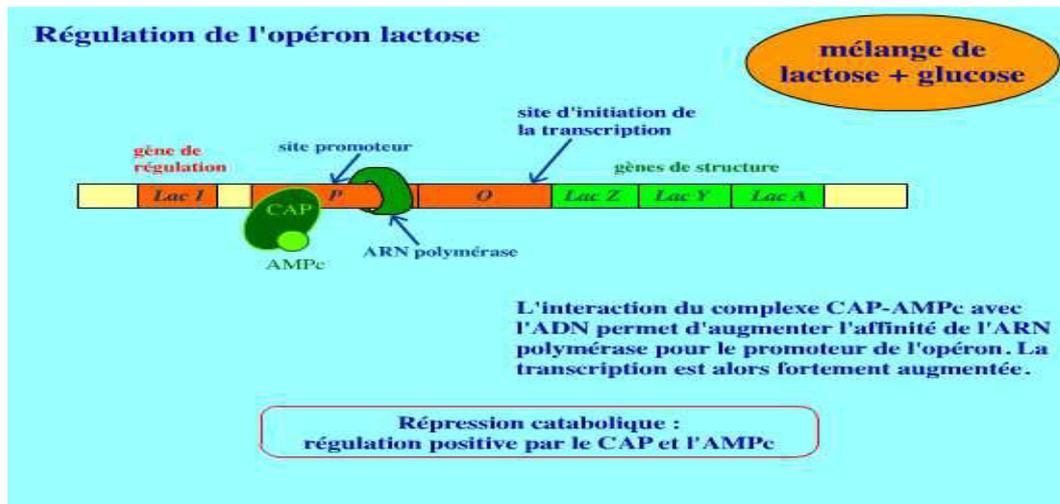


Figure 7 : régulation positive par le CAP et l'AMPc

Conclusion

L'induction de l'opéron lactose nécessite deux conditions, il faut que le lactose soit présent et que le glucose soit absent. La transcription de l'opéron lactose est donc sous le contrôle de deux protéines régulatrices :

- le répresseur lacI qui se fixe au niveau de l'opérateur en absence de lactose et bloque l'ARN polymérase. C'est une régulation négative,
- la protéine CAP (ou CRP) qui, sous forme d'un complexe avec l'AMPc, se lie à l'ADN et permet d'augmenter l'affinité de l'ARN polymérase pour le promoteur. C'est une régulation positive.

Allolactose : β -D-galactopyranosyl-(1-6)- β -D-glucopyranose, isomère du lactose

AMPc : Adénosine 5' monophosphate cyclique, c'est un second messager intracellulaire qui est formé par l'adénylate cyclase.

ARNm polycistronique : chez les procaryotes, les ARN peuvent contenir l'information nécessaire pour former plusieurs protéines différentes.

β -galactosidase : enzyme qui permet d'hydrolyser le lactose en galactose et en glucose (elle est codée par le gène *lacZ*).

CAP (ou CRP) : « Catabolite gene Activator Protein », ou CRP pour « cAMP Receptor Protein ». Cette protéine contrôle l'initiation de la transcription des gènes du catabolisme qui va permettre d'utiliser d'autres molécules nutritives quand le glucose est absent.

CAP-AMPc : complexe formé après l'interaction de la protéine CAP (ou CRP) avec l'AMPc.

Inducteur : molécule signal qui en se liant à une protéine régulatrice induit une augmentation de l'expression d'un gène donné.

lac A : gène de l'opéron lactose qui code pour la thiogalactoside transacétylase.

lac I : gène en amont de l'opéron lactose qui code pour le répresseur.

lac Y : gène de l'opéron lactose qui code pour la lactose perméase.

lac Z : gène de l'opéron lactose qui code pour la β -galactosidase.

Lactose : β -D-galactopyranosyl-(1-4)- β -D-glucopyranose.

Lactose perméase : protéine membranaire qui permet l'entrée du lactose dans la bactérie (elle est codée par le gène *lacY*).

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

Opéron : unité d'expression génétique qui comprend un ou plusieurs gènes et des séquences régulatrices (promoteur et opérateur) qui régule leur transcription.

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

Protéines régulatrices : ces protéines interviennent dans le contrôle de l'expression des gènes en se fixant sur des régions particulières de l'ADN. Elles peuvent ainsi activer ou réprimer de façon spécifique la transcription des gènes. L'action de ces protéines est réversible. On parle également de facteurs de transcription.

Répresseur : protéine qui se lie sur une séquence régulatrice ou sur l'opérateur d'un gène, bloquant sa transcription.

Thiogalactoside transacétylase : cette protéine est codée par le gène *lacA*, son rôle physiologique n'est pas bien connu. Elle acétyle les galactosides non métabolisés par la β -galactosidase, cette acétylation permet ainsi la diffusion de ces sucres à travers la membrane plasmique et leur élimination hors de la cellule.

Gène de structure : code une protéine structurale, une enzyme ou une protéine régulatrice

Gène de régulation: code une protéine impliquée dans la régulation d'expression d'autre gène.

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

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

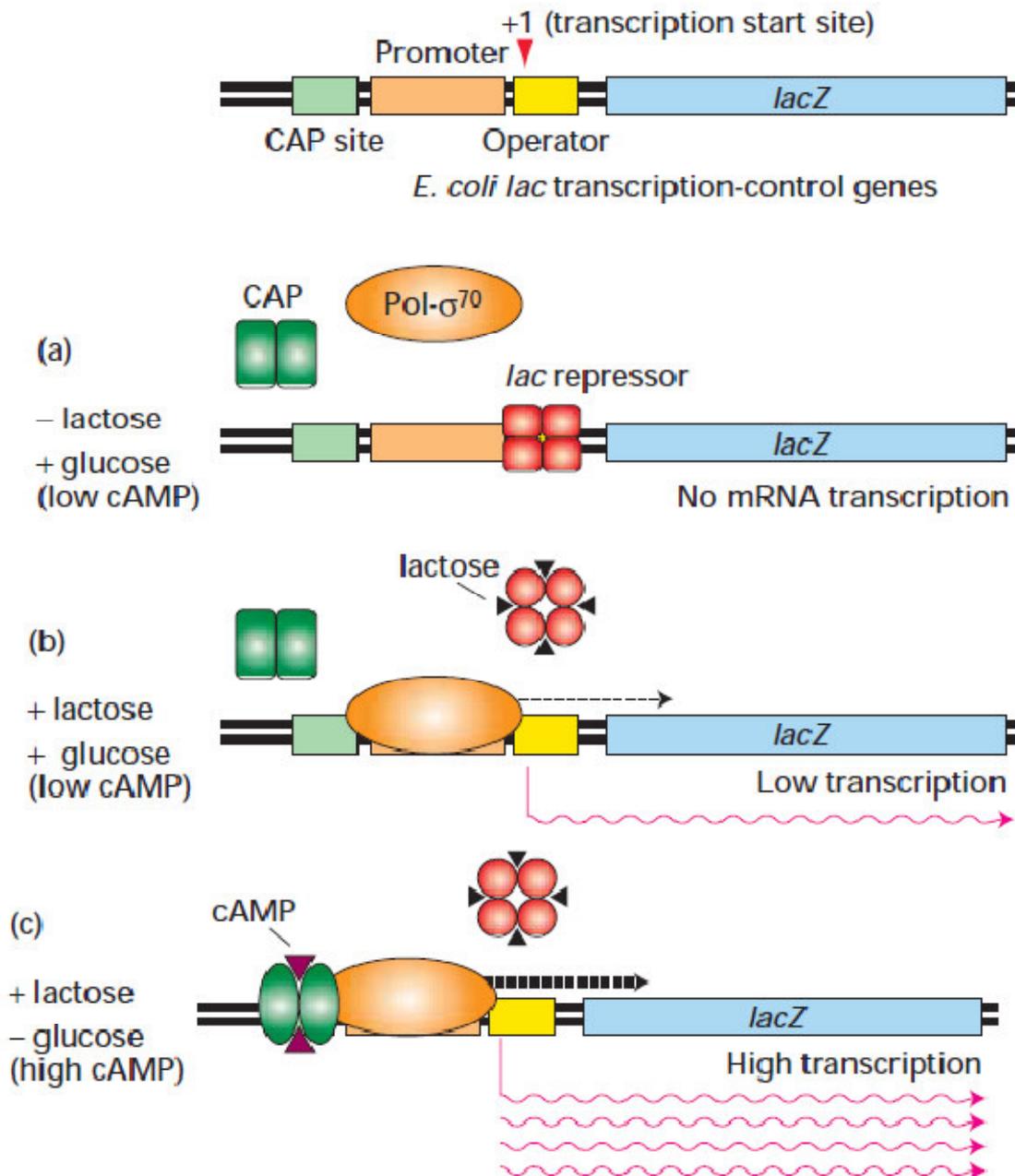


Figure 8: Expression des gènes *lac* en présence/absence du glucose et le lactose. Quand le répresseur Lac est lié sur l'opérateur, il exclut la polymérase, que CAP soit présente ou non.

CAP recrute la polymérase sur le promoteur *lac* où elle s'isomérise spontanément en complexe ouvert. L'allolactose (inducteur) résulte d'une transglycosylation du lactose (Lodish et al., 2003).

Bibliographie :

Passarge E. (2007). Color Atlas of Genetics. 3rd Ed. Flexibook. Thieme Stuttgart, NY. USA. 486p.

Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003). Molecular Cell Biology. 6th Ed. 937p.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 180-193p

Xie et al, (2003), Microbiol and Molecular Biolreviews, Vol. 67 p.303–342

Isabelle Borde, Dominique Boucher. L'opéron lactose, Biologie moléculaire. Université de Sorbonne. https://rnbio.upmc.fr/bio-mol_operon-lactose_sommaire

<http://www.cours-de-biochimie.fr/operons.php>

Medraoui Leila. (2015) Régulation génétique chez les bactéries: opérons cataboliques & anaboliques. Faculté des Sciences Rabat. Laboratoire de Microbiologie et Biologie Moléculaire. Université Mohamed V Agdal 88p.

C. CHEVALET, Florence CORPET M. GILLOIS A. MICALI. (1983) Modélisation dynamique de systèmes génétiques de régulation. L'induction de l'opéron lactose d'Escherichia coli : élaboration d'un modèle. *Genet. Sel. Evol.* 15 : (1), 1-30

Houghton Mifflin Harcourt. (2016) The Lactose Operon—A Look at Regulation

<https://www.cliffsnotes.com/study-guides/biology/biochemistry-ii/rna-and-transcription/the-lactose-operon-a-look-at-regulation>

Joël LUNARDI. (2012) Régulation de l'expression du message génétique. Université Joseph Fourier – Grenoble 1. UE1 : Biochimie – Biologie moléculaire 29 p.

III : régulation traductionnelle

Objectifs :

- Comprendre que la traduction peut être régulée à des niveaux différents
- Connaitre ces niveaux de régulation
- Connaitre les mécanismes précis de ces régulations

Introduction

L'ARNm est une copie simple brin linéaire de l'ADN, qui comprend la région codante une protéine, encadrée de régions non codantes. Il est synthétisé sous forme de précurseur dans le noyau de la cellule lors d'un processus appelé transcription. Il subit alors plusieurs étapes de maturation, ses deux extrémités sont modifiées, certaines régions non codantes appelées introns peuvent être excisées lors d'un processus appelé épissage. L'ARNm mûré est exporté dans le cytoplasme où il est traduit en protéine par un ribosome. L'information portée par l'ARNm est constituée d'une série de codons, des triplets de nucléotides consécutifs qui codent chacun pour un acide aminé de la protéine correspondante. L'enchaînement de ces codons constitue le gène proprement dit ou cistron. La correspondance entre codons et acides aminés constitue ce qu'on appelle le code génétique.

La transcription des ARNm et leur traduction sont des processus qui sont l'objet de contrôles cellulaires importants et permettent à la cellule de réguler l'expression des différentes protéines dont elle a besoin pour son métabolisme.

Chez les eucaryotes, un ARNm correspond en général à un seul gène et code une seule protéine (une seule phase ouverte de lecture). Ce sont des ARN monocistroniques. Chez les bactéries, les ARNm peuvent en revanche coder plusieurs protéines (plusieurs phases ouvertes de lectures successives), on les qualifie alors d'ARN polycistroniques. Chez ces organismes qui ne possèdent pas de noyau, la transcription des ARN messagers et leur traduction en protéine sont le plus souvent couplées.

On distingue **trois principales régions fonctionnelles** dans un ARNm : la région 5' non traduite (5'-UTR), le ou les cistrons codants et enfin la région 3' non traduite (3'-UTR). Les deux régions non traduites ou régions UTR (de l'anglais untranslated regions) contiennent souvent des signaux d'expression ou de maturation de l'ARN (figure 1).



Figure 1 : Structure d'un ARN messager eucaryote mature

Les ARNm maturés sont ensuite traduits en protéines par **les ribosomes** dans le cytoplasme. La petite sous-unité du ribosome (40S chez les eucaryotes) se fixe d'abord dans **la région amont** de l'ARNm et glisse jusqu'au codon de démarrage. Cette étape nécessite l'intervention d'un ensemble de protéines spécifiques appelées **facteurs d'initiation**. Il recrute alors le premier ARN de transfert et la grande sous-unité du ribosome (60S chez les eucaryotes) s'associe à l'ensemble. Le ribosome ainsi assemblé démarre la traduction.

À la fin de cette phase d'initiation, **les facteurs d'initiation** quittent le ribosome assemblé qui va allonger la protéine codée par l'ARNm, au cours de cycles d'élongations successives. À chaque cycle, le ribosome lit un codon et y associe l'ARNt de codon complémentaire, qui porte l'acide aminé correspondant. Cette étape nécessite l'action **de facteurs d'élongation**. Une fois l'ensemble du cistron lu par le ribosome, celui-ci parvient sur le codon d'arrêt qui le termine. Le ribosome libère la protéine terminée, puis le dernier ARNt et enfin l'ARNm. Ces étapes nécessitent l'intervention de **facteurs de terminaison**.

1) L'ARNm dans le cytoplasme :

Dans le cytoplasme des cellules eucaryotes, les ARNm existent sous forme de complexes associés à des protéines, sous forme **pseudo-circulaire**. La coiffe en 5' est en effet reconnue par le **facteur de démarrage de la traduction (eIF4E)**. La queue poly(A) est recouverte par une protéine spécifique, la **PABP** (poly(A) binding protein). Ces protéines, eIF4E et PABP interagissent toutes les deux avec une même protéine organisatrice, le facteur de démarrage **eIF4G**. Ceci conduit à une circularisation de l'ARNm au travers de ce complexe protéique, ses deux extrémités 5' et 3' étant ainsi rapprochées. Ce complexe contient aussi d'autres facteurs de démarrage, comme **eIF4A** qui est une **ARN hélicase**.

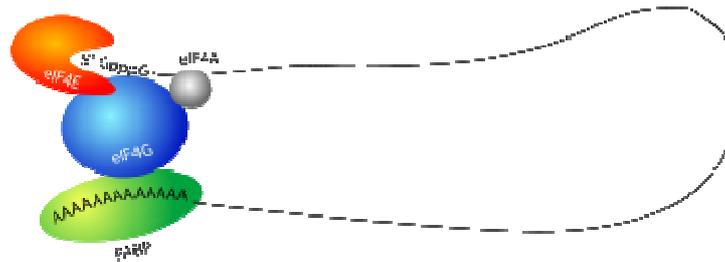


Figure 2 : Structure en pseudo-cercle de l'ARNm, via l'interaction de la PABP, liée au poly(A), eIF4E, lié à la coiffe, et eIF4G

L'un des effets de la formation de ces complexes circulaires est de permettre le rapprochement de la 5'-UTR et de la 3'-UTR de l'ARNm. Ces deux régions peuvent contenir des signaux d'expression qui vont ainsi pouvoir jouer conjointement sur le recrutement du ribosome. C'est aussi un moyen pour la cellule de vérifier que l'ARNm est complet et contient bien simultanément une coiffe et une queue poly(A) avant de démarrer la traduction.

2) Recrutement du ribosome :

Le recrutement du ribosome sur l'ARNm est une étape clé de la synthèse des protéines. L'efficacité de cette étape détermine en particulier le taux de synthèse final de la ou des protéines codées par l'ARNm. C'est à ce niveau que s'exercent la plupart **des régulations traductionnelles** de l'expression du génome.

2.a) Démarrage de la traduction eucaryote (initiation):

Chez les eucaryotes, les ARNm matures existent sous forme de complexes pseudo-circulaires dans le cytoplasme, associant les facteurs de démarrage eIF4E, eIF4G et eIF4A, ainsi que la PABP. C'est le facteur eIF4G qui joue le rôle de protéine organisatrice dans ce complexe car il est lié simultanément à la coiffe via eIF4E et à la queue poly(A) via la PABP. Lorsque ces deux interactions sont réalisées, eIF4G peut effectuer le recrutement de la petite sous-unité (**40S**) du ribosome.

Cette dernière est elle-même associée à plusieurs facteurs au sein d'un complexe de pré-initiation 43S. Celui-ci contient, outre la sous-unité ribosomique 40S, l'ARNt de démarrage aminoacylé par la méthionine associé au facteur eIF2 complexé au GTP, le facteur eIF3 qui maintient la sous-unité 40S dissociée de la grande sous-unité, ainsi que eIF1A et eIF5. Au sein de ce complexe de pré-initiation, eIF3 interagit avec eIF4G au sein du complexe ARNm.

Ceci permet le recrutement du ribosome au niveau de la coiffe de l'ARNm. Le complexe d'initiation complet ainsi formé est appelé complexe 48S. Il contient donc :

- l'ARNm coiffé et polyadénylé
- la sous-unité 40S du ribosome
- l'ARNt de démarrage au site P de la précédente ; cet ARNt est aminoacylé par la méthionine et associé à eIF2:GTP
- les facteurs de démarrage eIF1, eIF1A, eIF2, eIF3, eIF4A, eIF4B, eIF4E, eIF4G et eIF5.

Ce complexe 48S, recruté sur la coiffe de l'ARNm, va alors avancer le long de l'ARNm jusqu'à atteindre le premier triplet **AUG** qui va être reconnu comme **codon d'initiation**. Il se forme alors une interaction entre l'anticodon de l'ARNt et ce codon AUG. Ceci permet de caler le ribosome sur le cadre de lecture du gène. Ce « balayage » de l'ARNm par le ribosome est facilité par l'activité hélicase d'eIF4A qui, avec eIF4B, va ouvrir les structures secondaires éventuelles présentes dans le 5'-UTR. Une fois le codon de démarrage reconnu et l'interaction codon-anticodon formée, il y a hydrolyse du GTP par eIF2, **dissociation des facteurs de démarrage** et recrutement de la sous-unité **60S**. Le ribosome assemblé peut alors commencer la synthèse de la protéine (voir figure 3).

2.b) Élongation :

Pendant ce processus, un ou plusieurs ARN de transfert sont associés à l'ARNm au niveau de trois sites présents dans le ribosome :

- **le site A** (pour aminoacyl-ARNt) : il accueille les ARNt portant l'acide aminé qui va être ajouté à la chaîne. C'est au niveau du site A que s'effectue le décodage de l'information génétique, grâce à la reconnaissance de l'interaction entre le codon sur l'ARNm et l'anticodon sur l'ARNt
- **le site P** (pour peptidyl-ARNt) : il accueille l'ARNt qui porte la chaîne peptidique déjà synthétisée. Ce dernier est toujours apparié à l'ARNm par son anticodon
- **le site E** (pour exit) ou site de sortie : il accueille l'ARNt libre après transfert de la chaîne peptidique sur l'ARNt suivant, juste avant qu'il ne quitte le ribosome.

Ces trois sites sont occupés séquentiellement par les ARNt au fur et à mesure de la progression du ribosome sur l'ARNm. À tout instant du processus de traduction, au plus deux de ces trois sites sont occupés simultanément : soit le site P et le site A, soit le site E et le site P.

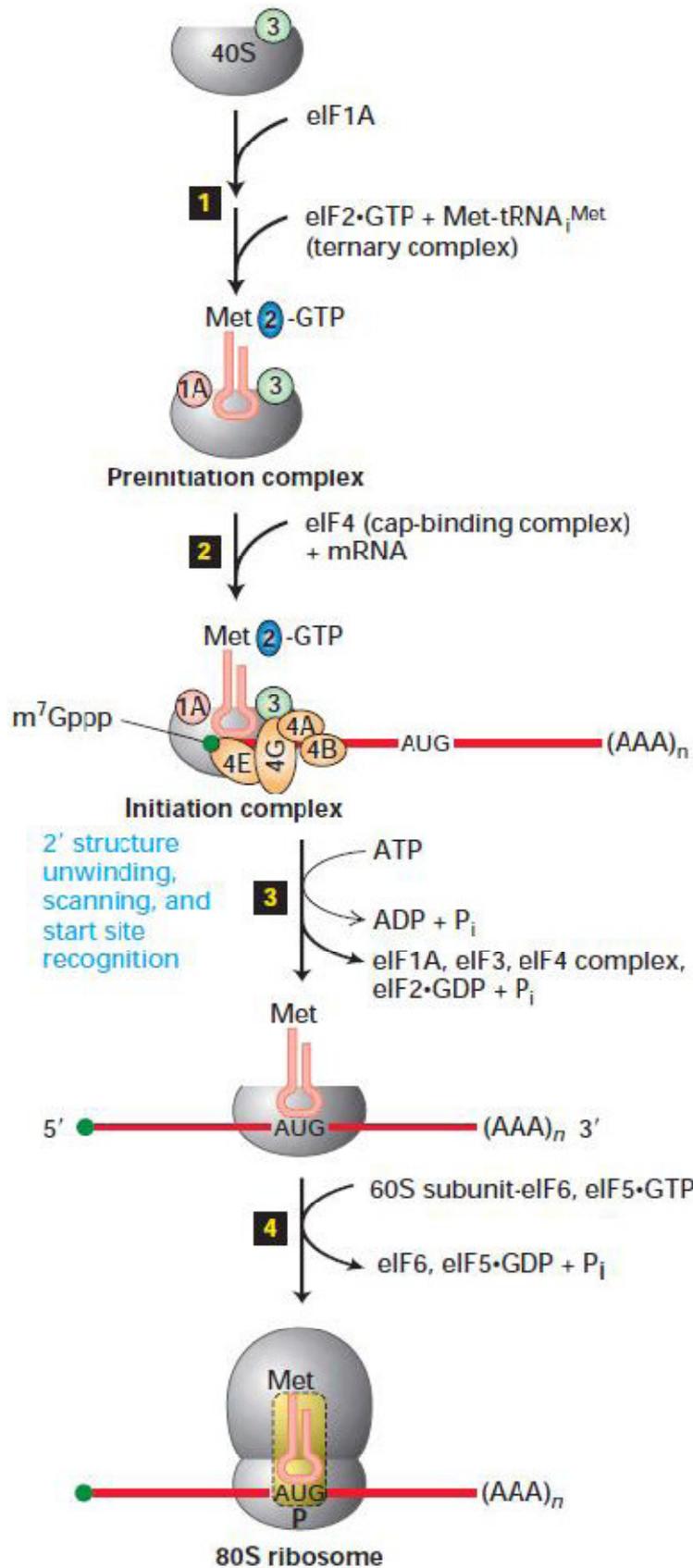


Figure 3: Formation du complexe d'initiation chez les eucaryotes (Lodish *et al.*, 2003).

Le début du cycle de synthèse démarre avec le ribosome ayant un seul ARNt fixé au site P par l'interaction codon-anticodon. Celui-ci porte estérifié à son extrémité 3' le début de la chaîne peptidique synthétisée. Le site A est initialement vacant. Ceux-ci sont associés à un facteur d'élongation (**eEF1** chez les eucaryotes) qui est une GTPase. Le ribosome teste si l'appariement codon-anticodon est correct. C'est la petite sous-unité du ribosome qui effectue ce contrôle au niveau du site de décodage. Si l'appariement codon-anticodon n'est pas correct, l'ARNt est rejeté et le processus d'essai-erreur se répète. Lorsqu'un ARNt correct s'apparie enfin au codon, le facteur d'élongation hydrolyse une molécule de GTP, ce qui provoque un changement de conformation du complexe ribosomique et la dissociation du facteur d'élongation. Il y a alors formation de la liaison peptidique entre l'acide aminé porté par l'ARNt au site A et celui porté par l'ARNt au site P. Cette réaction se produit au niveau du site catalytique de la grande sous-unité du ribosome et conduit au transfert de la chaîne peptidique sur l'ARNt lié au site A. Le ribosome recrute alors un facteur de translocation (**eEF2** chez les eucaryotes), lui aussi lié à une molécule de GTP. Ce facteur utilise l'énergie d'hydrolyse de ce GTP pour permettre la progression du ribosome sur l'ARNm. L'ARNt nu qui était au site P se retrouve décalé au site E de sortie et l'ARNt qui était au site A vient au site P. Le site A redevient ainsi vacant. L'ARNt nu du site E diffuse en dehors du ribosome et on revient dans l'état initial pour un nouveau cycle d'élongation (voir figure 4).

2.c) Terminaison de la traduction :

est le processus qui permet la libération de la protéine terminée une fois tout le gène lu par le ribosome, ainsi que le recyclage des différents acteurs impliqués, à savoir les sous-unités du ribosome, l'ARNm et le dernier ARNt. Cette étape fait intervenir des protéines spécifiques, appelées **facteurs de terminaison** de la traduction (RF ou eRF, pour *release factors*). La terminaison est déclenchée par l'arrivée d'un **codon d'arrêt** dans le site A du ribosome. À ce stade, la protéine complète est encore accrochée par une liaison **ester** au dernier ARNt situé au site P. Il n'existe normalement pas d'ARNt portant un anticodon complémentaire du codon d'arrêt et pouvant se lier à l'ARNm au site A. À la place, un premier facteur de terminaison (eRF1 chez les eucaryotes) se lie au ribosome et interagit avec le codon d'arrêt. Ceci déclenche l'hydrolyse de la liaison ester entre la protéine et le dernier ARNt. La protéine ainsi libérée quitte le ribosome. Un second facteur (RF3 ou eRF3) qui est une **GTPase** se lie alors au ribosome, l'hydrolyse du GTP permettant alors le départ des deux facteurs. L'ARNt et

l'ARNm sont finalement libérés par l'action de RRF (ribosome release factor) et de facteurs d'élongation et d'initiation (EF-G).

2.d) Contrôle qualité des ARNm :

Les cellules vivantes ont développé différents mécanismes de contrôle qualité des ARNm. Cela permet de vérifier que l'ARNm est intact et qu'il code bien pour une protéine complète. Ceci évite la production de protéines anormales qui en s'accumulant pourraient finir par tuer la cellule. Ces mécanismes sont différents chez les eucaryotes.

Chez les eucaryotes, le principal mécanisme de contrôle qualité est la vérification de la présence de la coiffe et de la queue poly(A). Sans ces éléments, l'ARNm n'est pas exporté du noyau et ne peut être traduit car il n'y a pas formation du pseudo-cercle via l'interaction eIF4E-eIF4G-PABP. Ainsi, par exemple, si un ARNm est endommagé et subit une coupure chimique ou enzymatique, il ne sera pas traduit, car l'extrémité 5' qui porte la coiffe et l'extrémité 3' qui porte la queue poly(A) seront séparées. Les eucaryotes disposent d'un autre mécanisme de contrôle qualité « dégradation de l'ARNm par l'intermédiaire du codon non-sens ». Ce mécanisme conduit à la dégradation des ARNm qui contiennent un **codon d'arrêt** (dit également codon non-sens) **prématuré**, résultant par exemple d'une mutation dans le génome. C'est le **ribosome**, lors de la première traduction de l'ARNm, qui semble être le médiateur de ce contrôle.

2.e) Dégradation des ARN messagers :

La dégradation des ARN messagers est effectuée par un complexe protéique appelé **exosome**, présent dans toutes les cellules eucaryotes. La dégradation est en général précédée par le raccourcissement de la queue poly(A) et par le retrait de la coiffe par une enzyme spécifique (decapping enzyme). Le dégradosome agit ensuite de manière exonucléolytique, c'est-à-dire qu'il hydrolyse progressivement l'ARN à partir d'une de ses extrémités, en l'occurrence à partir de **l'extrémité 3'**.

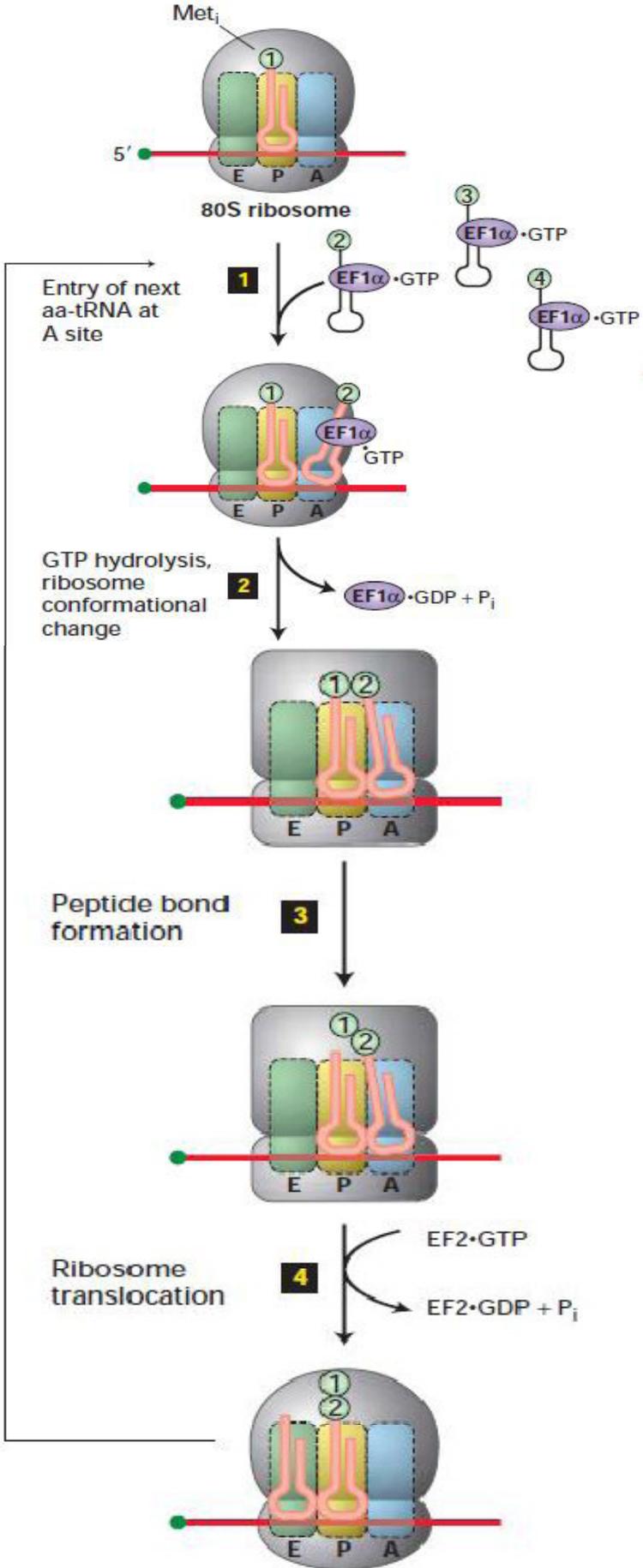


Figure 4: Élongation de la traduction chez les eucaryotes (Lodish *et al.*, 2003)

3) Modifications post-traductionnelles :

Les chaînes polypeptidiques nouvellement synthétisées par les ribosomes peuvent subir des modifications **post-traductionnelles** ; par exemple de lier par covalence des groupes fonctionnels tels que phosphate (phosphorylation), acétate (acétylation) ou méthyle (méthylation), voire des oses, des oligosaccharides ou des polysaccharides (glycosylation), ou encore des lipides (prénylation).

Bibliographie :

Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003).Molecular Cell Biology. 6th Ed. 937p.

Filali-Maltouf Abdelkarim, Cours Biologie Moléculaire ; 2ème Partie, L'expression de l'information génétique et les signaux impliqués. Faculté des sciences. Université Mohamad V Rabat 50p.

BELKADI Bouchra. (2012). Elément 2: Biologie Moléculaire. Partie 2: Expression génétique Traduction. Université Mohamed V - Agdal• Faculté des Sciences B.P. 1014 - Rabat – MAROC 33p.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 180-193p

Pascale PERRIN - FLBI399. Cours : le contrôle de l'expression génétique. Université Montpellier 2 53p.

Avner et Heard (2001). Nat. Rev. Genet. 2:59-67

Nedjma Ameziane, Marc Bogard, Jérôme Lamoril. (2005) Principes de biologie moléculaire en biologie clinique »

Joël LUNARDI. (2012) Régulation de l'expression du message génétique, UE1 : Biochimie – Biologie moléculaire. Université Joseph Fourier de Grenoble 29p.

Chapitre 3 : techniques de base de biologie moléculaire :

Préparation des acides nucléiques (extraction et purification)

I) définition :

L'**extractation de l'ADN** est une technique permettant d'isoler l'ADN de cellules ou de tissus. L'ADN ainsi extrait peut ensuite être utilisé pour des recherches de biologie moléculaire, telles que le **séquençage**, la **PCR** ou le **clonage**. Il existe différents protocoles pour extraire l'ADN, qui suivent approximativement le même schéma de principe :

- Lyse des cellules
- Elimination des protéines
- Elimination des autres acides nucléiques (ARN,...)
- Concentration de l'ADN par précipitation à l'alcool

Différentes variantes sont employées, suivant que l'on cherche à extraire de l'**ADN génomique** (issu du ou des chromosomes des cellules analysées) ou de l'**ADN plasmidique** (provenant de plasmides portés le plus souvent par des cellules bactériennes comme *Escherichia coli*). Il existe aujourd'hui des kits commerciaux permettant de réaliser rapidement ces extractions à l'aide de réactifs prêts à l'emploi.

1) Préparation d'ADN génomique :

On commence en général par une **lyse des cellules** ou des tissus, consistant éventuellement en un broyage, suivi d'une extraction par des **détergents**, qui vont disperser les bicouches lipidiques des membranes et dénaturer les protéines, et en particulier celles qui sont associées à l'ADN dans la chromatine. La solution obtenue est en général très visqueuse, car l'ADN ainsi libéré forme de très longs filaments qui s'opposent aux écoulements hydrodynamiques.

1.1. Préparation d'extraits bruts acellulaires :

◇Lyse des cellules : pour lyser les cellules on peut utiliser soit un système mécanique soit un système chimique suivant le type de cellules concerné.

1.1.A Lyse mécanique:

Pour la lyse mécanique, on utilise de préférence des méthodes ou systèmes qui ne dénaturent pas l'ADN, qui ne «cassent» pas les molécules d'ADN. Il faut donc que la méthode utilisée ne génère pas trop de forces de cisaillement: on utilise **le choc thermique** (cycles de congélation/décongélation) ou **le choc osmotique**.

On délaisse les méthodes d'agitation; on délaisse aussi l'emploi des ultra-sons.

Ces méthodes ne sont pas adaptées à l'extraction d'ADN à partir des cellules procaryotes. En effet, les traitements mécaniques utilisés pour casser des cellules procaryotes doivent être beaucoup plus draconiens que ceux utilisés pour les cellules eucaryotes (presse de French, broyeur à bille, ...): les bactéries, cellules de petite taille peuvent se glisser entre les faisceaux de forces de cisaillement et sont donc parfois très difficiles à lyser mécaniquement.

Les traitements mécaniques employés pour les procaryotes sont donc trop dénaturant pour l'ADN. La lyse mécanique est préférentiellement réservée aux cellules eucaryotes.

1.1.B Lyse chimique:

Pour les cellules procaryotes on préférera une lyse chimique.

* Fragilisation enzymatique des parois de cellules bactériennes, végétales et fongiques

La paroi est un obstacle majeur à la lyse cellulaire. Pour rendre efficace la désorganisation de la membrane plasmique, des hydrolases spécifiques peuvent être employées

Organisme	Bactérie	Champignon	Végétaux
Type de paroi	Peptidoglycane	Chitine	Cellulose, Hémicellulose, Pectines
Enzyme	Lysozyme	Chitinase	Cellulase, Hémicellulase, Pectinase

Ce traitement crée des brèches dans la paroi. Avec la perte de protection assurée par la paroi contre la forte pression osmotique intracellulaire, la cellule gonfle (afflux d'eau vers l'intérieur de la cellule) jusqu'à rupture de la membrane plasmique.

* On réalisera obligatoirement une lyse chimique pour extraire sélectivement les plasmides des bactéries. Les conditions de cette lyse chimique permettent de récupérer dans le lysat uniquement les plasmides sans l'ADN génomique, on parle alors de lysat clair. Pour arriver à cette extraction sélective la lyse chimique est obligatoire, la lyse mécanique ne permet d'obtenir qu'un lysat brut c'est-à-dire qui contient à la fois ADN génomique et plasmides.

* **Désorganisation des membranes par des détergents et solubilisation des lipides membranaires :**

Sodium Dodécyl Sulfate (SDS) appelé aussi laurylsulfate de sodium, sarcosyl **sont des détergents qui vont solubiliser les lipides membranaires sous forme de micelles**. Cela permet de créer des pores membranaires suffisamment larges pour libérer le contenu du cytoplasme hors des cellules. Suivant leur force (chargés ou pas), les détergents vont aussi plus ou moins dénaturer les protéines membranaires. La dénaturation des protéines de la membrane plasmique contribue également à la lyse de la cellule.

L'étape suivante est **la déprotéinisation** de la solution qui se fait par une extraction au moyen de **solvants organiques**, en général du **phénol** additionné de plus ou moins de **chloroforme**. Les protéines dénaturées forment un précipité à l'interface phénol-eau, tandis que l'ADN reste en solution dans la phase aqueuse qui est récupérée par décantation ou par centrifugation.

L'ADN est ensuite **précipité** par addition **d'éthanol ou d'isopropanol** dans la phase aqueuse, collecté par centrifugation et dissout dans du tampon. Pour éliminer les traces de phénol et d'autres contaminants, on peut enfin pratiquer une dialyse ou une étape de purification par chromatographie préparative.

1.1.C Elimination des protéines :

Cette étape sera très importante pour l'élimination des histones eucaryotes. On peut procéder de deux façons différentes.

- **Déprotéinisation par hydrolyse enzymatique:**

Elle est réalisée en faisant agir une endoprotéase non spécifique comme **la protéinase K**, active jusqu'à 65°C c'est donc une enzyme très stable. Cette digestion est souvent conduite en présence d'un détergent dénaturant comme le SDS. Le SDS, en plus de solubiliser les lipides et contribuer à la lyse de la cellule, facilite l'action de la protéinase K car il déploie la chaîne protéique.

- **Précipitation des protéines en utilisant un agent chaotropique:**

On parle aussi de déprotéinisation par défécation. Un agent chaotropique est un sel (donc des ions) qui modifie la solubilité des molécules (protéines ou acides nucléiques) et qui peut provoquer leur précipitation. Certains sont dénaturants d'autres pas.

L'agent chaotropique peut agir de plusieurs façons:

- Neutraliser certaines charges ioniques requises en surface pour le maintien de la solubilité.
- Interférer dans les interactions que les protéines établissent avec l'eau et modifier la solubilité des protéines. L'agent chaotrope entre en compétition avec la protéine pour établir des interactions avec les molécules d'eau disponibles dans la solution. L'eau en se liant à l'agent chaotrope prive la protéine des molécules d'eau qui l'hydrataient (la solvataient), la protéine sort alors de la solution et précipite.
- pour certains, il peut aussi dénaturer les protéines, on le qualifie alors de dénaturant précipitant. Par exemple pour les protéines, il provoque la rupture des liaisons hydrogènes qui maintiennent leur structure tertiaire entraînant ainsi le démasquage de leurs régions hydrophobes. Les régions hydrophobes ont tendance à s'agréger et les protéines précipitent (défécation).

On peut citer comme agents chaotropiques:

- Des agents chaotropiques qui ne dénaturent pas les protéines: **le chlorure de sodium** (NaCl) à forte concentration et **le sulfate d'ammonium**, qui lui est largement utilisé dans la purification des protéines
- des agents chaotropiques dénaturants pour les protéines, **le perchlorate de sodium** (NaClO₄), **le thiocyanate de guanidine** (TCG), **l'iodure de sodium** (NaI) et **le chlorure de lithium**.

Les agents chaotropiques les plus utilisés sont **le thiocyanate de guanidine** (GTC) et **l'iodure de sodium** (NaI). En effet, ces agents chaotropiques par leur action dénaturante sur les protéines assurent deux fonctions à la fois:

- disloquer les membranes des tissus en agissant sur les protéines membranaires et conduire donc à la lyse des cellules.
- et éliminer les protéines en provoquant leur précipitation. Après centrifugation, il reste peu de protéines dans le surnageant qui contient, lui, tous les acides nucléiques.

1.1.D Réactifs divers ajoutés au tampon d'extraction :

Lorsqu'on utilise les agents chaotropiques pour éliminer les protéines par précipitation, on ajoute quelque fois dans le tampon d'extraction **des thiols** pour empêcher la **reformation de ponts disulfures des protéines** qui restent ainsi à l'état dénaturé.

Une forte concentration saline (NaCl 0,15 M) du milieu d'extraction empêche la séparation des deux brins de l'ADN en formant un écran protecteur de contre-ions autour de la double hélice: les ions Na^+ s'associent avec les groupements phosphate PO_4^- , les neutralisent et diminuent les forces de répulsion qui s'exercent entre les deux brins de même charge électrique, les ions Na^+ stabilisent ainsi la structure en double hélice. Le citrate de sodium et l'acétate de sodium aussi utilisés dans les tampons d'extraction jouent le même rôle = stabiliser l'ADN natif c'est-à-dire sous sa forme bicaténaire.

Le tampon contient souvent une substance chélatrice telle que l'EDTA qui séquestre par complexation les ions divalents ou le citrate qui possède 3 groupements carboxyle qui s'ils sont chargés COO^- piègent les ions positifs. Ces deux réactifs sont utilisés pour piéger notamment le magnésium, cofacteur des DNases et RNases afin de mieux préserver les acides nucléiques par inhibition des nucléases.

1.1.E Elimination des ARN lors de l'extraction de l'ADN :

Les extraits acellulaires bruts contiennent les deux types d'acides nucléiques: ADN et ARN.

On peut hydrolyser les ARN de deux façon, par:

- Hydrolyse chimique avec NaOH :

. A pH alcalin l'ARN est hydrolysé et pas l'ADN qui est protégé de la lyse alcaline par l'absence de groupement hydroxyle en 2' sur le désoxyribose

- Mais on préfère souvent opérer par **digestion enzymatique**. Pour diminuer la concentration en ARN par digestion à la RNase, on utilise une RNase «DNase free», c'est-à-dire dépourvue d'activité DNase; pour cela les DNases contaminant éventuellement les préparations de RNase du commerce sont dénaturées par un chauffage (par exemple 5 min à 100°C) auquel résiste la RNase, enzyme particulièrement thermostable. La RNase est souvent apportée dès le début de l'extraction-purification grâce à sa grande stabilité.

1.2. Purification par extraction phénol-chloroforme :

1.2.A Principe de la méthode d'extraction :

Les méthodes utilisent la solubilité différentielle des molécules (acides nucléiques / contaminant comme les protéines et les lipides) entre deux phases non miscibles.

- En pratique : On mélange vigoureusement la solution d'acides nucléiques en phase aqueuse à une phase non miscible hydrophobe

. Après centrifugation, on récupère la phase aqueuse contenant les acides nucléiques à la pipette (phase aqueuse = phase supérieure).

• **Deux extractions successives** qui se distinguent selon la phase non miscible:

* **L'extraction phénolique**: est utilisée pour débarrasser les acides nucléiques des protéines car le phénol est un déprotéinisant puissant. Les protéines précipitent, elles sédimentent au fond de la phase aqueuse mais et elles restent à l'interface c'est-à-dire qu'elles restent à la surface de la phase phénolique qui est une phase hydrophobe. Les débris membranaire lipidiques vont aller dans la phase phénol hydrophobe.

Le phénol doit être très pur et saturé en tampon (pH 8 pour extraire l'ADN). L'inconvénient du phénol réside dans son caractère très corrosif, c'est donc un produit à manipuler avec précaution.

* **L'extraction au chloroforme**: complète toujours l'extraction précédente pour éliminer toutes traces de phénol aqueux. Toute trace de phénol doit être éliminée pour permettre l'action ultérieure d'enzyme (de restriction par exemple) sur l'acide nucléique extrait. Le chloroforme est additionné d'alcool isoamylique (AIA = 3-méthyl-1-butanol = $(\text{CH}_3)_2\text{CHCH}_2\text{CH}_2\text{OH}$) qui est un agent anti-mousse stabilisant la séparation des phases (agent déstabilisant de l'émulsion).

Lors de la préparation d'un extrait ADN, l'élimination de l'ARN peut être effectué à la fin ou avant l'étape phénolique. S'il est effectué après l'étape phénolique, ce traitement doit être suivi d'une deuxième déprotéinisation afin d'éliminer la RNase résiduelle.

L'extraction des ARN nécessite de prendre des précautions particulières: port de gants, matériels et réactifs à usage unique traités sans ribonucléases. Ces précautions sont mises en place pour remédier à la présence de ribonucléases sur la peau et à l'existence de tissus naturellement enrichis en ribonucléases (pancréas par exemple). L'extraction des ARN est donc plus délicate que celle de l'ADN car les ARN sont très sensibles à l'action des RNases qui, de leur côté, sont des enzymes très stables et qui donc, même si elles sont présentes en petite quantité, réaliseront la dégradation des ARN.

→L'extraction débute par une lyse cellulaire ou tissulaire en présence d'agent chaotrope, isothiocyanate de guanidine ou de chlorure de lithium, l'agent chaotrope est utilisé pour dénaturer les RNases endogènes. On y ajoute souvent un inhibiteur de ribonucléases comme le β -mercaptoéthanol ($\text{HOCH}_2\text{CH}_2\text{SH}$) qui est un agent réducteur et qui va rompre les ponts disulfures dénaturant ainsi les ribonucléases. (Il est communément employé pour réduire les Ponts disulfures présents dans les protéines et peut jouer un rôle d'antioxydant biologique).

→La lyse est ensuite suivie par une extraction en phénol acide. Le phénol utilisé est un Phénol acide c'est-à-dire qu'il a été mis en solution et équilibré avec un tampon de pH acide (pH 5). Dans ces conditions les protéines histones, riches en acides aminés basiques (portant dans leur radical un groupement NH₂), vont avoir à pH acide une forte charge positive (NH₃⁺), elles vont alors s'associer plus fortement à l'ADN génomique chargé négativement et vont l'entraîner avec elles lors de leur précipitation.

Ainsi l'ADN et les protéines qui ont précipité restent à la surface de la phase phénolique et sont ainsi rassemblés à l'interface. La phase aqueuse (supérieure) contiendra en solution les ARN débarrassés de l'ADN. La phase phénolique (inférieure) contiendra les lipides.

D'autres techniques rapides ont été élaborées sans l'utilisation du phénol acide chloroforme: purification sur mini-colonne contenant une membrane de gel de silice présentant une affinité pour les ARN avec élimination des ADN par l'emploi d'une DNase.

2) Préparation d'ADN plasmidique :

La préparation d'ADN plasmidique à partir de bactéries est l'une des techniques les plus courantes de la **biologie moléculaire**. Le principe de l'extraction est connu sous le nom de **lyse alcaline**. Cette méthode permet de préparer sélectivement l'ADN du plasmide contenu dans les bactéries, tout en éliminant l'ADN du chromosome bactérien.

Le principe de cette méthode consiste à effectuer la lyse des cellules au moyen d'un **détergent** (dodécyl sulfate de sodium) en présence de **soude**, à **pH 13**. À ce pH très alcalin, l'ADN est **dénaturé**, c'est-à-dire que les deux brins de la double-hélice sont séparés. On neutralise ensuite rapidement la solution, ce qui provoque la renaturation brutale (réappariement des brins du duplex d'ADN). L'ADN chromosomique, très long (~10⁶ paires de base), ne parvient pas à se réapparier complètement et forme des enchevêtrements insolubles. L'ADN plasmidique, court (~10³ paires de base), parvient à se réassocier complètement et reste en solution. On sépare alors les espèces par centrifugation. Les protéines précipitées, sont également éliminées avec le détergent et l'ADN chromosomique. L'ADN plasmique, resté en solution, est alors concentré par précipitation à **l'alcool**.

Différentes variantes ou améliorations de cette méthode existent notamment dans un grand nombre de kits commerciaux. Ces derniers contiennent souvent des petites colonnes de résine chromatographique échangeuse d'ion, permettant d'améliorer la pureté de l'ADN obtenu. Pour

éliminer les **ARN**, on ajoute fréquemment **des ribonucléases** qui vont hydrolyser sélectivement ces acides nucléiques, en laissant l'ADN intact.

3) Contrôle de l'extraction :

- Par **électrophorèse sur gel d'agarose**, on peut analyser la présence et la taille des acides nucléiques contenus dans la préparation. Ceux-ci sont révélés par **le bromure d'éthidium**, un colorant dont la fluorescence augmente très sensiblement quand il interagit avec l'ADN. Dans le cas d'une préparation d'ADN plasmidique, on observe sous UV une ou plusieurs bandes colorées discrètes dans le gel, correspondant à l'ADN du plasmide. Dans le cas d'ADN chromosomique, on observe un continuum de bandes, correspondant aux fragments d'ADN issus de la coupure statistique du chromosome résultant des traitements mécaniques lors de l'extraction de l'ADN.
- Par mesure de l'absorbance dans l'UV au moyen **d'un spectrophotomètre** : 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 **50µg/ml**
 - Une solution de DNA simple brin ou RNA à **25µg/ml**

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 (pureté) = A260 / A280** Une solution d'ADN est considérée pure si : **1.7 ≤ P ≤ 2**.

Bibliographie :

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 147-150p

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES. UE2. Ellipses 10-23p.

Philippe Lepoivre, Jean Kummert, Dominique Collnet, Olivier Duterme, Christine Anceau. (1994). Techniques moléculaires de détection et d'identification des agents Phytopathogènes, Cahiers Agricultures, 3 : 217-25

Passarge E. (2007). Color Atlas of Genetics. 3rd Ed. Flexibook. Thieme Stuttgart, NY. USA. 486p.

Streips N. U. & R. Yasbin. (2002). Modern Microbial Genetics. 2nd Ed. Wiley-Liss, Inc. New York. USA; 603p.

Medraoui Leila. (2014). Cours Biologie Moléculaire ; Techniques de séparation et d'analyse des acides nucléiques: ADN, ARN & PLASMIDES. Faculté des sciences. Université Mohamad V Rabat 44p.

Séparation des acides nucléiques : (électrophorèse sur gel d'agarose) :

1. définition :

L'**électrophorèse sur gel d'agarose** est une méthode utilisée en biochimie et en biologie moléculaire pour séparer l'ADN, l'ARN ou des protéines en fonction de leur poids moléculaire.

La technique de l'électrophorèse sur gel d'agarose est basée sur la séparation des acides nucléiques chargés négativement sous l'effet d'un champ électrique. Cette séparation s'effectue à travers la matrice du gel d'agarose : les molécules de plus petites tailles se déplacent plus rapidement et migreront plus loin que les molécules de tailles supérieures

2. Principe général:

L'électrophorèse est basée sur la séparation des acides nucléiques chargés négativement à pH 7~8, vers l'anode (+) sous l'effet d'un champ électrique. Cette séparation s'effectue à travers la matrice du gel en fonction de la taille des molécules.

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.

3. Applications de l'électrophorèse :

- Séparation de fragments ADN digérés
- Estimation du poids moléculaire de fragment d'ADN après une digestion par des enzymes de restriction
- Analyse d'ADN après une amplification par PCR

Les avantages de l'électrophorèse sur gel d'agarose sont les suivants:

- préparation aisée, rapide, et peu coûteuse des gels d'agarose
- pas de dénaturation des échantillons
- l'agarose plus ferme et moins toxique que le gel de polyacrylamide
- les échantillons peuvent être récupérés en vue d'analyses supplémentaires

4. Facteurs affectant la migration :

Les facteurs les plus importants sont :

- **la longueur de la molécule d'ADN** : la séparation se fait en fonction du poids moléculaire et donc de la taille de l'ADN.
- **la concentration du gel** : L'augmentation de la concentration d'agarose dans un gel réduit la vitesse de migration et permet la séparation de fragment d'ADN de plus petite taille (voir tableau 1).
- **le voltage** : plus le voltage est important, plus la vitesse de migration augmente. Toutefois le voltage est limité en intensité (5V/cm): un fort voltage induit une augmentation de température ce qui peut faire fondre le gel.
- **La conformation de l'ADN** : l'ADN plasmidique, non digéré par une enzyme de restriction, migre à différentes vitesses (du plus lent au plus rapide) : ADN circulaire, ADN linéaire et ADN superenroulé.

Tableau 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		

5. Gel d'agarose :

On utilise généralement un gel **1 % m/V** (1 g d'agarose pour 100 ml de volume final) en électrophorèse. Plus on veut un gel discriminant, plus on augmentera le pourcentage d'agarose.

5.a. Agarose :

L'**agarose** est un polymère à base d'agar purifié. Différentes puretés d'agarose sont disponibles. En général, de l'agarose de grande pureté à la solidification lente est utilisé lorsque l'ADN doit être extrait du gel après migration.

L'agarose est utilisé à des concentrations de **0,5% à 2%** (poids/volume) et permet de séparer des molécules de très grande taille, principalement de l'ADN ou de l'ARN.

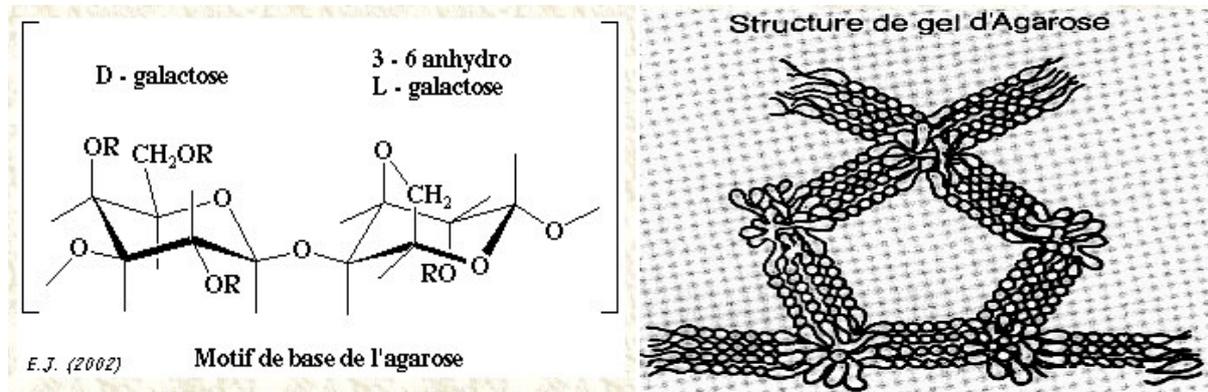


Figure 1 : structure de gel d'agarose

5.b Tampons :

Il existe un nombre très varié de tampons. Les plus souvent utilisés sont le Tris/Acétate/EDTA (TAE), le Tris/Borate/EDTA (TBE) et le sodium borate (SB).

Le TAE possède le plus faible pouvoir tampon mais produit une meilleure séparation pour les fragments d'ADN de grande taille.

Le SB est relativement nouveau et inefficace pour la séparation de fragments d'ADN d'une taille supérieure à 5000 paires de bases (5 kb). Cependant sa faible conductivité permet l'utilisation d'un plus fort voltage (jusqu'à 35V/cm), ceci réduisant considérablement le temps de migration.

Des fragments d'ADN avec seulement quelques paires de bases de différences sont séparés en utilisant un gel d'agarose à 3 % et avec un tampon SB de très faible conductivité (1 mM lithium borate).

6. Révélation de l'ADN :

La méthode de révélation la plus utilisée est la révélation au **bromure d'éthidium** ou **BET**. Le bromure d'éthidium est un agent d'intercalation couramment utilisé comme marqueur d'acide nucléique dans les laboratoires de biologie moléculaire. Lorsqu'il est exposé

à des rayonnements ultraviolets, il devient fluorescent avec une couleur rouge-orangée, 20 fois plus intense lorsqu'il est lié à l'ADN.

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 (figure 2).

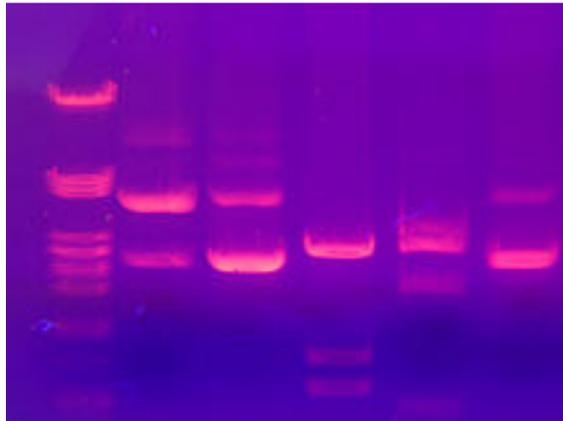


Figure 2 : Gel d'électrophorèse sous UV

6.1 Limites de résolution :

L'électrophorèse sur gel d'agarose permet théoriquement la séparation de fragments d'ADN d'une taille allant de 50 paires de bases à plusieurs millions. Cependant, elle est généralement utilisée pour la séparation de fragment d'une taille allant de 100 pb à 20 kpb. La durée moyenne de migration est d'une heure.

Les petits fragments d'acides nucléiques sont mieux séparés par électrophorèse sur gel de polyacrylamide. Les fragments de tailles importantes sont plus difficiles à séparer. En général, l'utilisation de gel d'agarose à forte concentration (3 à 4 %) est alors nécessaire pour des fragments inférieurs à 150 pb, car elle permet une meilleure séparation et résolution des différentes bandes en fonction de leur différence de taille. Le principal désavantage est le temps de migration, qui peut aller jusqu'à plusieurs jours. Pour pallier ces problèmes, il est avantageux d'effectuer une électrophorèse en champ pulsé.

6.2 Analyse du gel :

Après la migration d'électrophorèse, le gel est éclairé sous ultraviolet afin d'observer les bandes d'ADN fluorescente.

Les bandes peuvent être alors découpées et séparées du gel, puis dissoutes afin de récupérer l'ADN purifié.

6.2.a Détermination de la taille des fragments :

L'estimation de la taille des fragments est faite grâce à la comparaison avec l'échelle de marqueur de taille moléculaire (DNA-ladder) utilisée simultanément dans un autre puits lors de la migration.

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

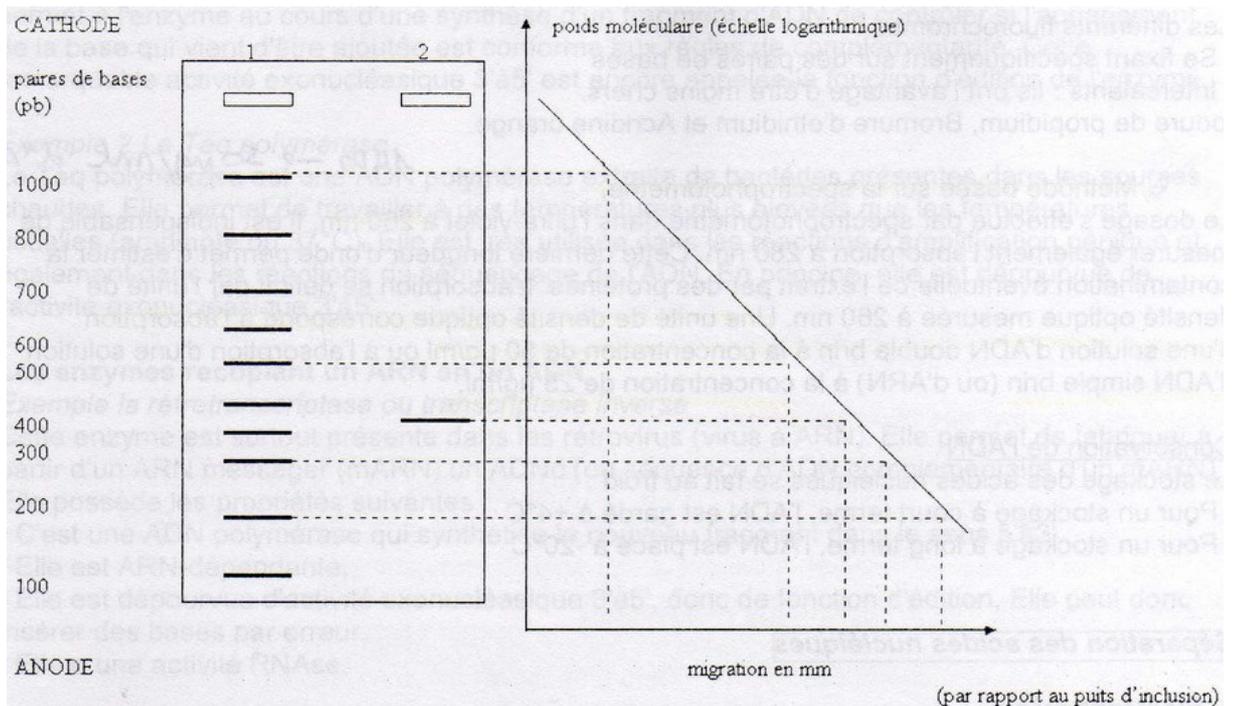


Figure 3 : Distance de migration en fonction du logarithme de la taille des fragments d'ADN.

Légendes de la figure :

Figure de gauche

- PISTE 1: Marqueurs de taille (en paires de bases).
- PISTE 2: Echantillon d'ADN dont la taille est à déterminer.

Figure de droite

- Courbe d'étalonnage en abscisses, les distances par rapport au puits d'inclusion et en ordonnées les poids moléculaires (échelle logarithmique).

Exemples de marqueurs de taille :

- Phage λ coupé par BstEII (0.12 \longrightarrow 8.45)
- Phage λ coupé par HindIII (0.13 \longrightarrow 23.13)
- pBR322 coupé par BstNI (0.12 \longrightarrow 1.86)

- 100 pb Ladder (100 \longrightarrow 1000)

7. L'électrophorèse en champ pulsé :

Pour séparer des fragments de taille supérieure à 20 kb.

Le principe de cette électrophorèse consiste à **changer l'orientation et/ou la polarité du champ électrique alternativement au cours du temps**. A chaque modification du champ, la molécule d'ADN doit se réorienter parallèlement au nouveau champ. Le temps nécessaire à la réorientation est proportionnel à la longueur de la molécule. Lorsque le champ est rétabli dans son sens initial, la molécule doit une nouvelle fois se réorienter. Ces temps de réorientation provoquent un retardement de la migration nette qui est proportionnel à la taille de la molécule. Le support de migration est un gel d'agarose à 1% et la taille des fragments séparés est de l'ordre de 50 kb à quelque mégabases.

Il n'est pas possible d'utiliser les méthodes classiques pour analyser des échantillons d'ADN (longueur environ 50 kb). Aussi les cellules dont on souhaite analyser l'ADN sont incluses dans un bloc d'agarose et la digestion par les enzymes de restriction est effectuée in situ. Puis on utilise le principe de l'électrophorèse en champ pulsé.

Bibliographie :

Medraoui Leila. (2014). Cours Biologie Moléculaire ; Techniques de séparation et d'analyse des acides nucléiques: ADN, ARN & PLASMIDES. Faculté des sciences. Université Mohamad V Rabat 44p.

Brody JR, Calhoun ES, Gallmeier E, Creavalle TD, Kern SE (2004). Ultra-fast high-resolution agarose electrophoresis of DNA and RNA using low-molarity conductive media. *Biotechniques*. 37:598-602.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 147-150p

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES. UE2. Ellipses 10-23p.

Sambrook J, Russel DW (2001). *Molecular Cloning: A Laboratory Manual* 3rd Ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Gel electrophoresis of DNA. In: Sambrook, J., Fritsch, E.F. and Maniatis, T. (Eds.) *Molecular Cloning: a Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, chapter 6.

Westermeier, R. (1997). *Electrophoresis in Practice: a Guide to Methods and Applications of DNA and Protein Separation*, VCH, Weinheim.

Fanny Demay – BTS BioAnalyses & Contrôles.

Détection, caractérisation et identification des acides nucléiques

1. Marquage et suivi des acides nucléiques

Le marquage est principalement employé dans toutes les techniques qui utilisent une sonde (hybridation, northern,). On distingue le marquage dit « chaud » utilisant des isotopes radioactifs, et les marquages « froids » qui utilisent des molécules aux propriétés fluorescentes, luminescentes. Ces dernières sont de plus en plus employées car plus pratiques.

1.1. Marquage radioactif :

On distingue plusieurs méthodes selon la localisation du marquage (extrémités ou interne à la molécule) et selon la nature de la séquence marquée (simple ou double brin).

Le Phosphore 32 est le radioisotope le plus utilisé. Incorporé dans la sonde enzymatiquement au moyen d'un ou plusieurs nucléotides tri P radiomarqués. Il existe des sondes mono ou double brins. Soufre 35, H 3 utilisés plutôt pour le séquençage et hybridation in situ. On peut aussi réaliser un marquage en 5' : avec la T4 polynucléotide kinase. La radioactivité est aussi utile pour le marquage des oligonucléotides de synthèse. Deux méthodes de marquage d'ADN peuvent être utilisées : **Random Printing ou Nick Translation**

1.1.1 Les sondes doubles brins :

1.1.1.A Marquage par amorçage au hasard (Random Printing) :

Très employé dans les laboratoires pour par exemple les Southern et Northern Blot, il permet d'obtenir des activités spécifiques plus élevées, nécessaires pour détecter un gène sur quelques mg d'ADN génomique.

Dans cette technique de marquage, les deux brins d'ADN de la sonde sont préalablement séparés par chauffage suivi d'un refroidissement brutal. Puis, on ajoute un mélange d'oligonucléotides (hexanucléotides) de synthèse correspondant à toutes les combinaisons mathématiquement possibles (soit $4^6 = 4096$ nucléotides). Ces oligonucléotides vont s'hybrider avec la sonde pour une partie d'entre eux. Ces oligonucléotides fixés vont servir d'amorces à l'ADN polymérase I qui va reconstituer l'intégrité des deux fragments en présence de désoxynucléosides triphosphates marqués au ^{32}P . Des ADN polymérases par exemple dérivée du **phage T7** sont actuellement les plus utilisées (figure 1).

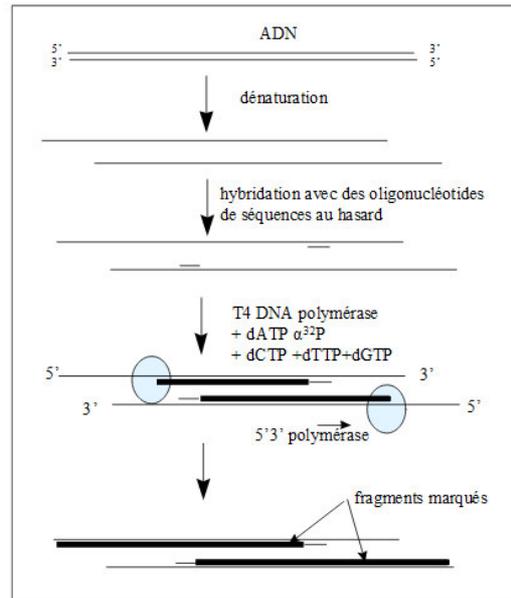


Figure 1 : marquage par amorçage au hasard

L'ADN peut être marqué à son extrémité 5' à l'aide d'une kinase, par exemple, la T4 polynucléotide kinase extraite d'E. coli infecté par le bactériophage T4. En présence d'ATP avec du ^{32}P en position g ou $[\text{}^{32}\text{P}]\text{g-ATP}$, il est possible d'échanger le groupement 5'-phosphate présent sur le fragment d'ADN avec le phosphate radioactif en position g sur l'ATP. Cette méthode est générale. Elle est plus efficace si les extrémités 5' sont préalablement déphosphorylées, par exemple par l'action d'une phosphatase alcaline.

1.1.1.B Marquage par translation de coupure (Nick Translation) :

utilise 2 enzymes :

***DNase I** dans des conditions ménagées pour générer quelques coupures simples brin dans le fragment d'intérêt.

***DNA pol. I** pour dégrader l'ADN dans sens 5'-3' au niveau de ces coupures et repolymériser en présence d'un nucléotide chaud.

Les outils enzymatiques utilisés en biologie moléculaire montrent que l'ADN double brin traité par la Dnase I était clivé au hasard. La réparation des coupures réalisées par la Dnase I nécessite l'action de l'ADN polymérase I en présence de désoxynucléosides triphosphates marqués au phosphore radioactif (^{32}P). Les désoxynucléosides triphosphates utilisés sont

marqués en position alpha au ^{32}P . Cette technique est appelée "technique de nick translation". Cette technique est actuellement en retrait par rapport à la technique suivante (figure 2).

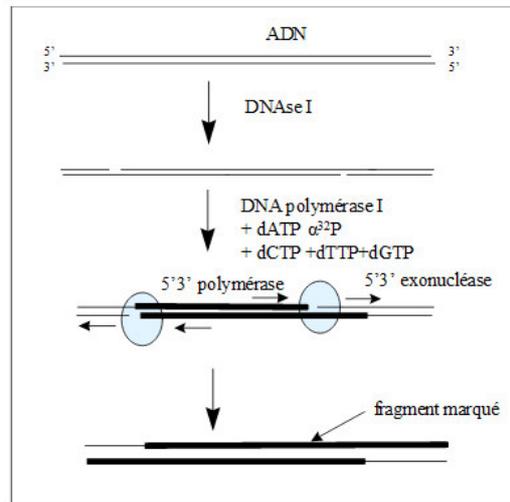


Figure 2 : technique de nick translation

1.1.2 Les sondes simples brin :

Pour obtenir une sonde spécifique d'un seul brin on utilisera plutôt une sonde ARN synthétisée in vitro.

Dans certains cas on a besoin que la sonde soit marquée uniquement à une extrémité. On peut marquer soit l'extrémité 5' soit l'extrémité 3'.

1.1.2.A Marquage en 5' des ARN :

La première base des ARN est généralement une adénine. Si on effectue une transcription in vitro en présence de $\gamma\text{ATP } 32\text{P}$ (ou $\beta\text{ATP } 32\text{P}$, le seul A marqué sera celui qui est en 5'. Les autres adénines incorporées dans l'ARN néosynthétisé ne seront pas marquées, en effet seul le phosphate en α restera dans le polymère (figure 3).

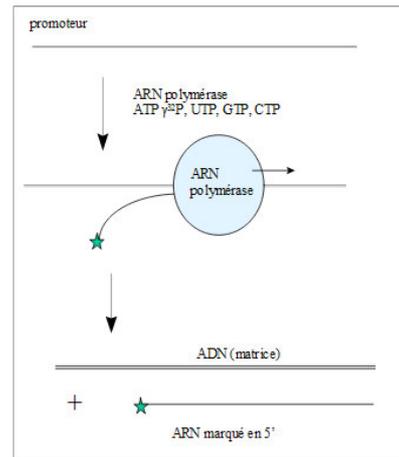


Figure 3 : ARN marqué en 5'

1.1.2.B Marquage en 5' des acides nucléiques (ADN ou ARN) :

A partir d'une extrémité 5'OH on peut ajouter un phosphate marqué à l'aide d'une kinase et de $\gamma^{32}\text{P}$ ATP.

Il faudra au préalable obtenir l'acide nucléique avec un 5'OH.

Pour les oligonucléotides synthétisés chimiquement, la synthèse s'effectue de 3' vers 5' et l'extrémité 5' est OH. On n'a donc aucun traitement à faire.

Pour les ADN obtenus par digestion par une endonucléase de restriction, on a un 5' phosphate. Ce phosphate peut être éliminé à l'aide d'une phosphatase on obtient une extrémité 5'OH (on détruit ensuite la phosphatase (chauffage, phénol par exemple), sur le produit de la réaction avant de faire agir la kinase.

Pour les ARN messagers eucaryotes qui ont un cap en 5', il faut détruire ce cap, par action d'une pyrophosphatase et d'une phosphatase

Cette méthode est généralement employée pour les oligonucléotides pour lesquels les marquages internes ne sont pas possibles du fait de leurs petites tailles.

1.1.2.C Marquage en 3' :

Dans ce cas il va falloir ajouter un ou plusieurs nucléotides à l'extrémité de la sonde. Pour les ARN on utilisera une RNA ligase en présence d'un nucléotide monophosphate et d'ATP pour fournir de l'énergie. Le nucléotide sera ajouté en 3' OH. Pour qu'il n'y ait pas plusieurs

nucléotides ajoutés, on utilisera un nucléotide ayant son extrémité 3' bloquée par exemple par un phosphate.

Pour les ADN, on peut utiliser une deoxynucleotidyl terminal transférase avec un $\alpha^{32}\text{P}$ nucléotide. Il y a alors ajout de plusieurs nucléotides en 3'.

Si on veut n'en rajouter qu'un, il faut utiliser un didéoxynucléotide marqué à la place d'un déoxynucléotide, la polymérisation sera ainsi bloquée après l'ajout du premier nucléotide.

Une autre méthode consiste à utiliser l'activité terminale transférase des polymérases. La plupart des polymérases présente en effet une activité terminale transférase en ayant comme matrice un ADN double brin. Cette limitation dans la matrice fait qu'un seul nucléotide est rajouté en 3' même si on utilise un déoxynucléotide. C'est une adénine qui est préférentiellement ajoutée, on fera donc le marquage avec de l'ATP $\alpha^{32}\text{P}$. On utilisera une polymérase dépourvue d'activité 3'5' exonucléase dans ce cas pour ne pas retirer le nucléotide ajouté en 3' (figure 5).

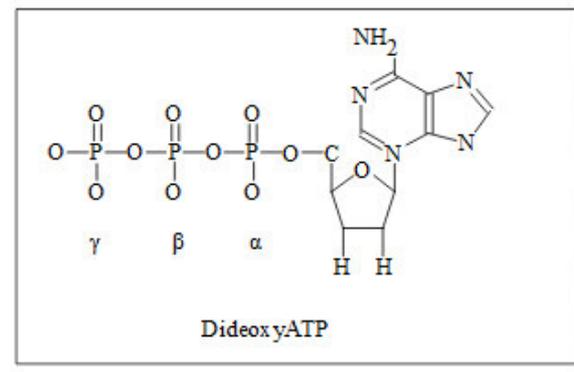


Figure 4 : Dideoxy-ATP

Les sondes radioactives présentent de nombreux **inconvénients** :

- nécessité de se protéger du rayonnement émis, un maniement des sondes inconfortables.
- Décroissance rapide du P^{32} , d'où un besoin de marquer les sondes fréquemment.

Pour pallier à ces inconvénients, on peut utiliser les sondes « froides ».

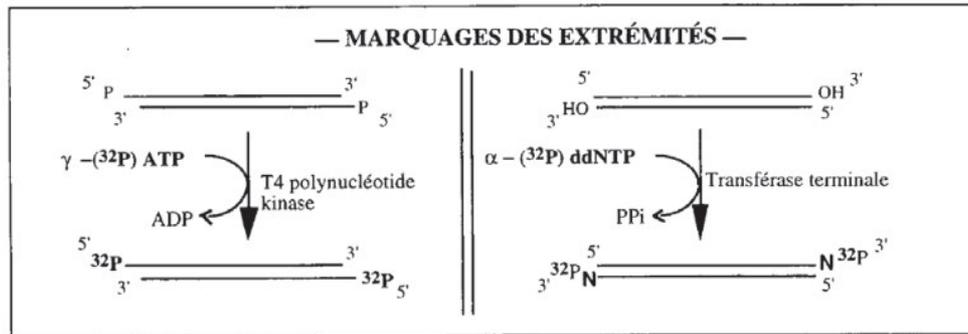
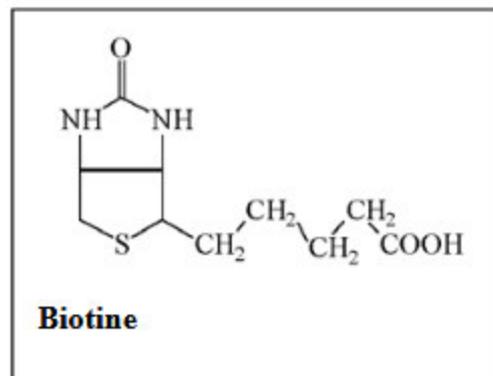
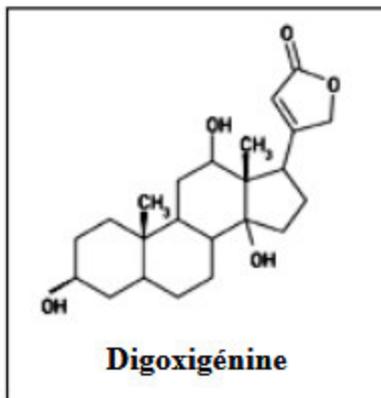


Figure 5 : marquage des extrémités

1.2 Marquages froids :

On peut utiliser des marquages non radioactifs. Dans ce dernier cas, on parle de sonde froide. Les deux types de marquage froid les plus souvent utilisés sont les marquages à la biotine et à la digoxigénine. Ces deux composés sont fixés sur un nucléotide, lui-même incorporé à la sonde.

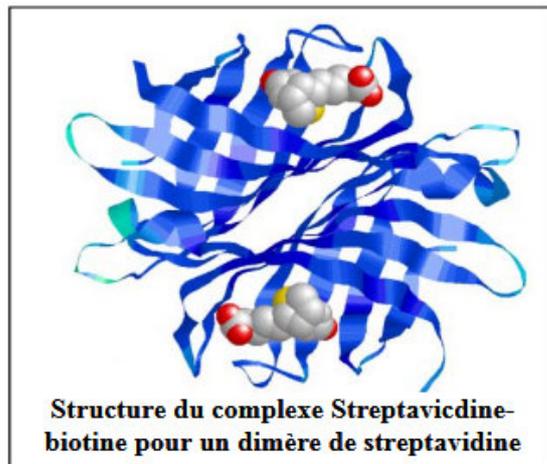
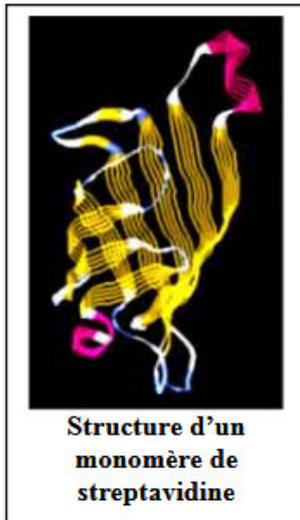


La biotine est la vitamine H. C'est un cofacteur de réactions enzymatiques impliquées dans l'incorporation ou le transfert de CO₂ (si on mange trop de blanc d'œuf, on inhibe l'incorporation de CO₂ dans le foie).

1.2.1 La détection des sondes froides

La biotine est détectée avec de la streptavidine. L'avidine est une protéine du blanc d'œuf, elle comporte 4 sites de liaison pour la biotine. Elle est généralement produite par génie génétique dans *Streptomyces avidinii*, sous une forme non glycosylée et prend alors le nom de streptavidine.

La digoxigénine se révèle à l'aide d'un anticorps.



La streptavidine ou les anticorps n'ont pas d'activité enzymatique propre, on utilise des protéines de fusion pour les détecter. L'anticorps ou la streptavidine sont couplés avec une protéine ayant une activité enzymatique facilement détectable comme la phosphatase alcaline ou la peroxydase.

La peroxydase du raifort (radis noir) (HRP, horseradish peroxidase) est une protéine de 40 kDa. Elle catalyse la réaction : $H_2O_2 + \text{donneur d'électron} \rightarrow \text{produit coloré} + H_2O$ on utilise un chromogène comme donneur d'électrons

- soit du tétrachlorure de diamino-3-3'-benzidine (DAB) qui donne un précipité brun
- soit de l' amino-3-ethyl-9-carbazole (AEC) qui donne un précipité rouge.

La phosphatase alcaline

C'est une protéine de 40 kDa isolé d'intestin de veau. Plusieurs chromogènes peuvent être utilisés: new fushine, fast red, BCIP/NBT...

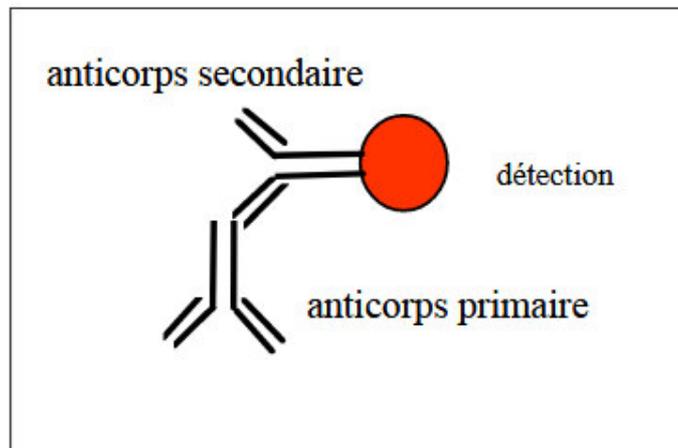
1.2.2 L'amplification du signal :

Pour augmenter la sensibilité des sondes froides on cherche à amplifier le signal détecté. Il existe de nombreuses variantes pour atteindre ce but.

1.2.2.A Méthode en deux couches :

Elle permet une simple détection de l'anticorps primaire sans amplification du signal, sauf si l'anticorps secondaire est polyclonal et reconnaît donc plusieurs épitopes. L'anticorps secondaire est conjugué avec l'enzyme ou le marqueur de fluorescence.

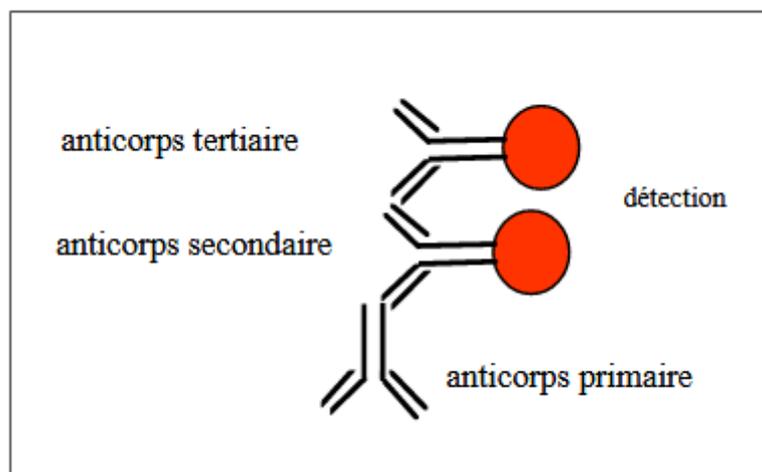
L'anticorps secondaire peut être remplacé par la protéine A.



1.2.2.B Méthodes en trois couches :

On utilise un anticorps primaire qui reconnaît l'antigène, un anticorps secondaire qui reconnaît l'anticorps primaire et un anticorps tertiaire qui reconnaît l'anticorps secondaire.

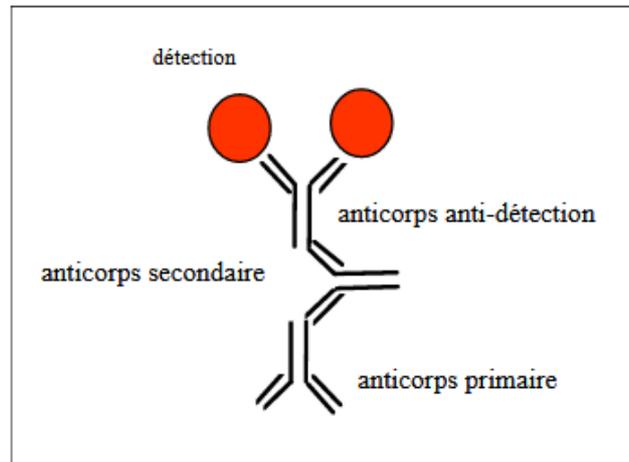
Les anticorps secondaires et tertiaires sont conjugués au même système de détection.



1.2.2.C Méthode utilisant un complexe entre l'enzyme de détection et un anticorps :

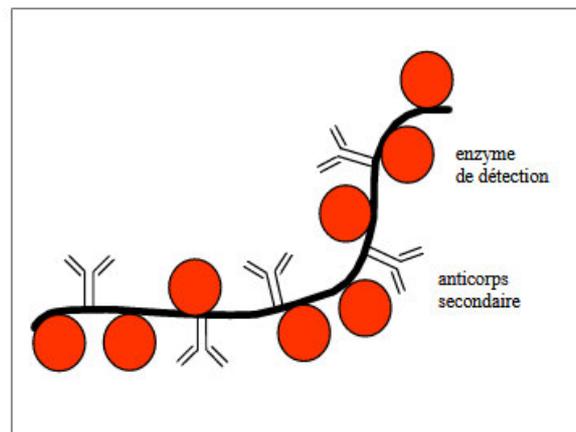
Cette méthode est parfois appelée méthode des anticorps non marqués, ou suivant l'enzyme de détection PAP (peroxydase antiperoxydase) ou APAAP (alkalin phosphatase anti alcaline).

phosphatase). L'anticorps utilisé contre l'enzyme de détection doit être de même nature (obtenu dans la même espèce animale) que l'anticorps primaire.



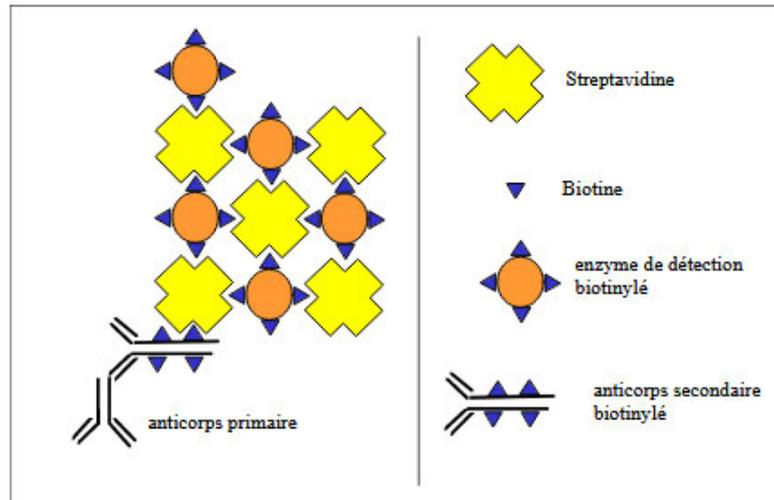
1.2.2.D Méthode EPOS (Enhanced Polymer One Step Staining) et Envision (Polymer two step staining) (DAKO) :

Les anticorps primaires (EPOS) ou secondaires (Envision) sont fixés sur un polymère de dextran avec la protéine servant de révélation (peroxydase, phosphatase alcaline.)



1.2.2.E Méthodes utilisant les complexes avidine-biotine. (AB Complex) :

L'anticorps secondaire est biotinylé comme la protéine de révélation. En présence d'avidine qui a quatre sites de liaisons à la biotine, on obtient un empilement de l'enzyme de détection.



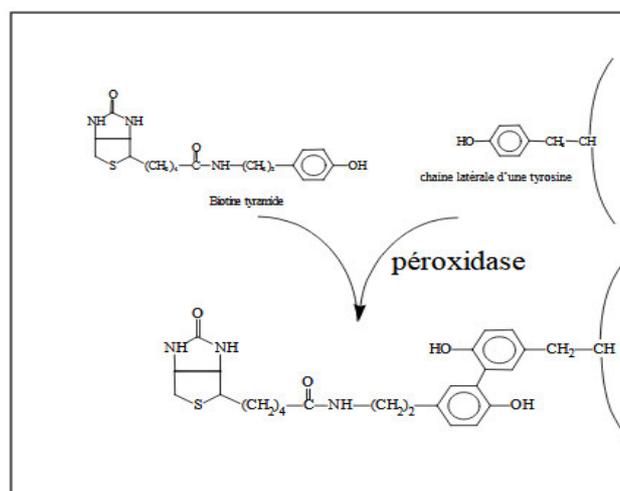
1.2.2.F CSA (Catalysed Signal Amplification) :

Utilisation de la biotinylation-tyramide. Les tyramides sont convertis par la peroxydase en un intermédiaire qui est réactif avec les résidus aromatiques riches en électrons tel que la tyrosine, présente dans les protéines.

La streptavidine-péroxydase reconnaît la biotine sur la sonde. La peroxydase active la

biotinylation-tyramide qui se fixe sur la protéine (ou sur une protéine proche) et, ainsi, on a une nouvelle biotine fixée à proximité de la sonde. Cette biotine peut alors être reconnue par la streptavidine-péroxydase.

Cette technique peut être utilisée pour les sondes biotinylées mais aussi pour toutes les sondes qui sont reconnues par un anticorps. Dans ce dernier cas on utilisera un anticorps secondaire marqué à la biotine.



Bibliographie :

Vincent Ecochard, Didier Fournier, Laurence Nieto, Laurent Paquereau. (2011). Techniques et Stratégies en Biologie Moléculaire. Première partie. Université Paul Sabatier, Toulouse, 201p.

Clémence Allain. (2006) Sondes fluorescentes pour l'ADN : marquages covalent et non-covalent. Université Pierre et Marie Curie - Paris VI, 260p.

D. Tagu, C. Moussard, (2003). Principes des techniques de biologie moléculaire, , 2e éd., INRA-Quæ (ISBN 2-7380-1067-9).

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 150-157p.

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES. UE2. Ellipses 47-51p.

Medraoui Leila. (2015). Cours Biologie Moléculaire ; Initiation aux techniques usuelles de biologie moléculaire. Faculté des sciences. Université Mohamad V Rabat 132p.

Strachan Tom, Read Andrew. (2004). Génétique Moléculaire Humaine. Hybridation des acides nucléiques : principes et applications. 4^e édition ; Médecine Sciences publications. Lavoisier 16p.

Hybridation moléculaire :

Introduction :

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 (figure 1).

I. Définition :

L'hybridation moléculaire désigne l'association qui peut avoir lieu entre deux acides nucléiques simples brins de séquences complémentaires et qui conduit à la formation d'un double brin ou duplex. Cette association s'effectue par l'établissement de liaisons hydrogènes spécifiques : deux liaisons entre l'adénine (A) et la thymine (T) (ou l'uracile U) et trois entre la cytosine (C) et la guanine (G). La formation et la stabilité des duplex dépendent de nombreux facteurs en plus de la composition en bases : longueur des duplex, complexité de la séquence.

L'hybridation est à la base de nombreuses techniques de biologie moléculaire impliquant la mise en présence d'au moins deux brins simples d'acides nucléiques dans des conditions physico-chimiques précises. Le brin, dont on connaît au moins une partie de la séquence, est une sonde, l'autre brin, celui que l'on souhaite caractériser constitue la cible. L'un des deux brin est marqué par couplage chimique avec une molécule pouvant générer un signal.

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.

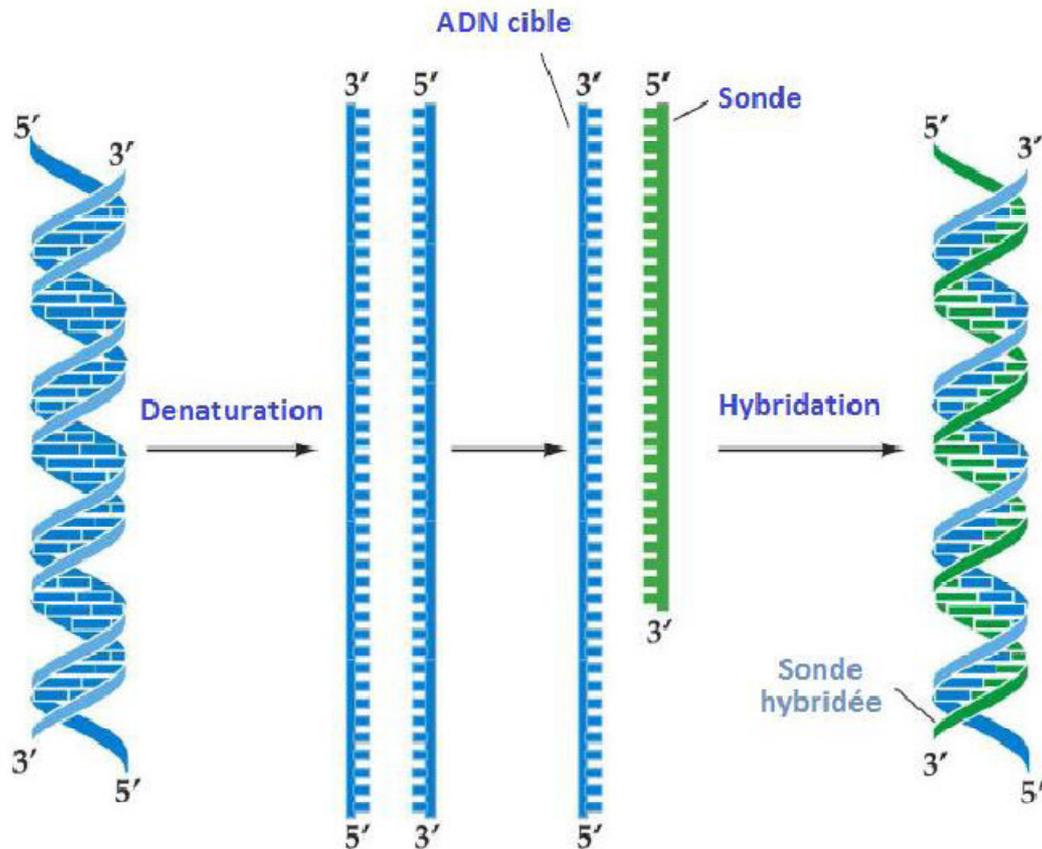


Figure 1: Hybridation des acides nucléiques (Passarge, 2007).

I.1 Les sondes nucléotidiques :

A. Définition :

Une sonde nucléotidique est un segment de nucléotides qui permet de rechercher de manière spécifique un fragment d'acide nucléique que l'on désire étudier. Cette réaction sonde-fragment correspond à une réaction d'hybridation moléculaire.

B. Caractéristiques générales :

Une sonde nucléotidique peut être soit une séquence d'ADN ou d'ARN, mais obligatoirement monobrin. Sa taille est très variable: oligonucléotides de 20-30 nucléotides ou à l'opposé de plusieurs centaines de nucléotides. La sonde est complémentaire et antiparallèle du fragment recherché. Dans un mélange complexe où s'effectue l'hybridation moléculaire, la sonde doit être facilement repérable grâce à un marquage avec un radioisotope (marquage chaud), mais il existe également des sondes appelées sondes froides sans marquage par un radioisotope.

C. Obtention d'une sonde :

Il existe plusieurs possibilités pour obtenir une sonde nucléotidique:

- Une sonde oligonucléotidique peut être fabriquée par synthèse chimique, si la séquence de l'ADN à repérer est connue. Si elle est inconnue, on peut étudier la protéine correspondante et remonter grâce au code génétique à la séquence d'ADN. Dans ce dernier cas, le travail est particulièrement laborieux (nombre de codons élevé pour un même acide aminé).
- Une sonde peut être un ADNc. Une partie seulement de l'ADNc est utilisée (après action d'enzymes de restriction et clonage des fragments obtenus).
- Une sonde peut être théoriquement de l'ARNm.

D. L'hybridation moléculaire avec la sonde :

L'hybridation moléculaire sonde-fragment d'ADN à repérer nécessite des conditions physico-chimiques parfaites (tampon, pH, température, etc..). Ces conditions sont appelées la stringence. Plusieurs facteurs peuvent également intervenir comme la longueur de la sonde et la complémentarité sonde-fragment avec possibilité de mauvais appariements.

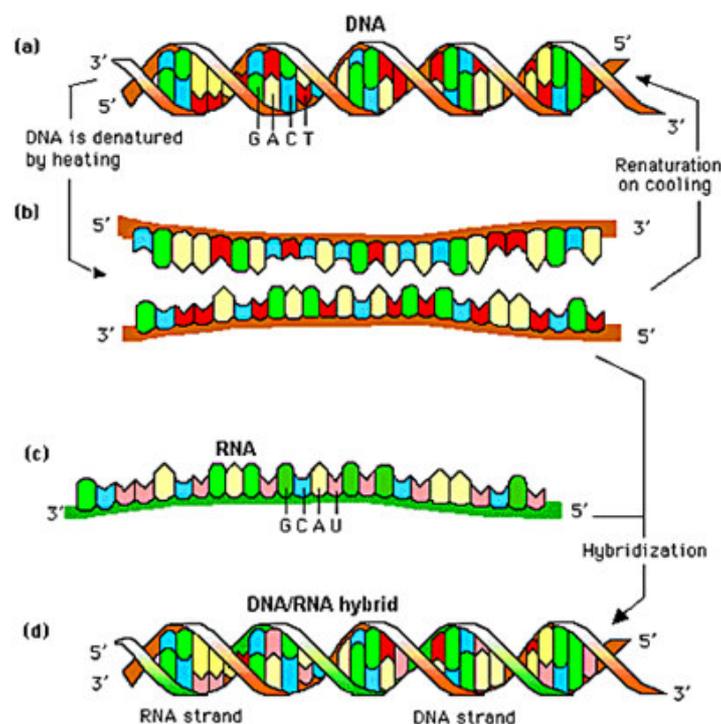


Figure 2 : hybridation ADN/ARN

II. 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.).

II.1 Hybridation d'ADN sur support solide :

II.1.1 Southern Blot :

Cette méthode a été initialement décrite par E.M. SOUTHERN en 1975. Elle consiste à détecter spécifiquement des fragments d'ADN transférés sur filtre par leur hybridation à des séquences complémentaires marquées par un radioisotope. Les étapes successives de cette technique sont les suivantes:

1-Extraction de l'ADN génomique.

Cette extraction s'effectue à partir des leucocytes circulants par exemple obtenus à partir de sang total.

2-Digestion par des enzymes de restriction différentes du même ADN génomique.

L'ADN génomique est digéré par des enzymes de restriction différentes: dans le tube 1, on réalisera une digestion par l'enzyme 1; dans le tube 2, une digestion par l'enzyme 2; etc....). On peut bien entendu réaliser des digestions par deux enzymes dans un même tube. Dans ces conditions, on obtient un très grand nombre de fragments, mais seuls quelques fragments correspondent à une partie ou à la totalité du gène étudié.

3-Séparation électrophorétique des fragments d'ADN par électrophorèse dans un gel d'agarose.

Après séparation électrophorétique en gel d'agarose des fragments d'ADN bicaténaires obtenus par digestion enzymatique, on réalise une dénaturation des fragments par un traitement alcalin du gel d'électrophorèse. Ce traitement transforme les fragments d'ADN double brin (ou bicaténaires) en fragments d'ADN monobrin (ou monocaténaires).

4-Transfert des fragments monocaténaires du gel d'agarose à un support souple (membrane de nitrocellulose ou de nylon par exemple).

Le transfert des fragments monocaténaires du gel d'agarose à un support type nylon s'effectue par simple capillarité.

5-Fixation des fragments monocaténares d'ADN sur le support souple et hybridation dans des conditions optimales de stringence avec une sonde complémentaire marquée à un radioisotope.

Les fragments monocaténares d'ADN transférés sur un support solide souple (nylon) sont mis en présence d'une sonde qui va s'hybrider dans des conditions physico-chimiques bien définies. La sonde s'apparie avec les fragments d'ADN monocaténaire selon les règles de complémentarité. De plus, elle est marquée avec un radioisotope (soit à son extrémité 5', soit à l'intérieur de la chaîne polynucléotidique).

6-Lavages et révélation (dans ce cas par autoradiographie).

Après de nombreux lavages, le support solide est mis en contact avec un film photographique pendant plusieurs jours. Le film est ensuite révélé. Les bandes d'ADN monocaténares hybridées avec la sonde radioactive sont visibles sous forme de bandes noires sur un fond blanc. La position de ces bandes par rapport à des témoins de poids moléculaire permet de déterminer la taille de ces fragments (voir figure 2).

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

II.1.2 Northern Blot :

Le principe est le même que pour le Southern Blot mais ici ce sont les **ARN** qui sont étudiés. Donc plus besoin de digérer par enzyme de restriction.

La visualisation d'un ARN par une sonde permet de :

- apprécier sa distribution dans les tissus, étudier son abondance relative
- déterminer sa taille
- détecter les intermédiaires de maturation et les différentes formes d'épissage de l'ARN.

Nous citerons parmi les applications de la technique de SOUTHERN:

Ex1 : Carte de restriction de l'ADN

Les enzymes de restriction coupent l'ADN au niveau de séquences parfaitement définies. Il est donc possible d'établir une véritable carte d'un gène donné par la technique de Southern. Cette carte porte le nom de carte de restriction.

En pratique, l'ADN à étudier est réparti en différentes fractions. Chaque fraction est traitée par une enzyme de restriction ou un couple d'enzymes de restriction. Ainsi dans l'illustration ci-dessous. Dans le puits 1, on dépose l'ADN digéré par Eco RI seule; dans le puits 2, on dépose l'ADN digéré par Hind III seule et dans le puits 3, l'ADN digéré par Eco RI et Hind III. La détection des différents fragments est réalisée à l'aide d'une sonde marquée du gène étudié (voir illustration).

Illustration des fragments après migration

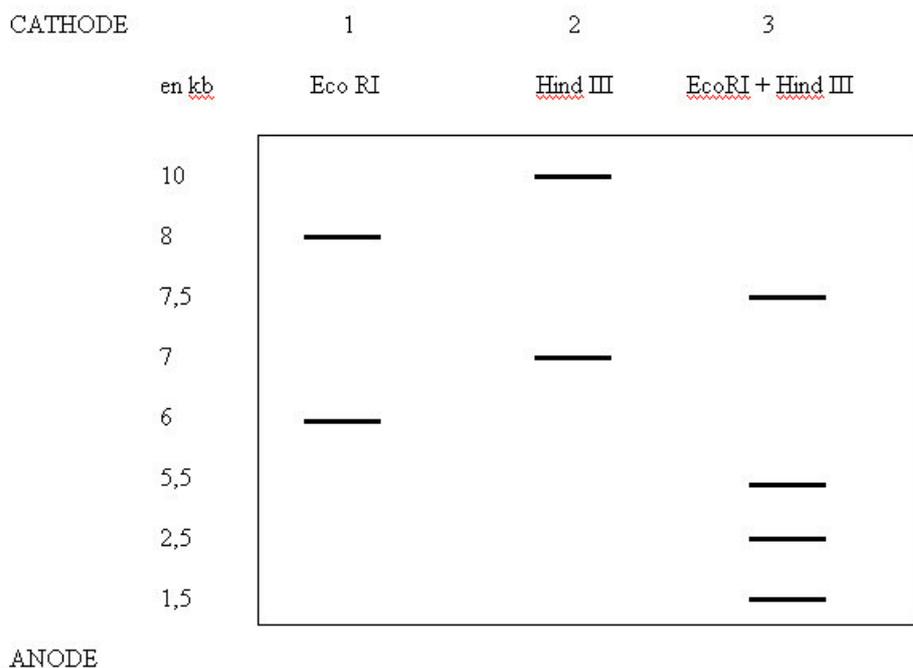
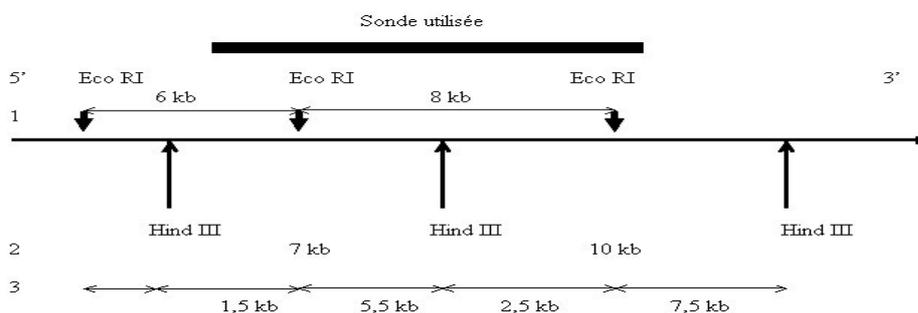


Illustration des positions sur l'ADN



Légende :

- (1): Position des coupures par Eco RI.
- (2): Position des coupures par Hind III.
- (3): Position des coupures par Hind III + Eco RI.

Ex2 : Si on dispose de sondes spécifiques du gène à explorer. Mise en évidence des délétions dans un gène ou une zone juxta-génique. L'étendue de la délétion peut être estimée par la détermination de la taille des fragments d'ADN.

Ex3 : Comparaison de l'ADN de différents individus avec mise en évidence des polymorphismes de restriction.

En présence de mutation(s) ponctuelle(s) dans des sites de restriction, des variations dans la taille de certains fragments seront observées et ceci par exemple à l'intérieur d'une même population. Ces petites différences entre des individus dans une même population sont appelées de polymorphisme de taille des fragments de restriction (ou RFLP pour « restriction fragment length polymorphism »).

Ex4 : Mise en évidence des recombinaisons entre des gènes.

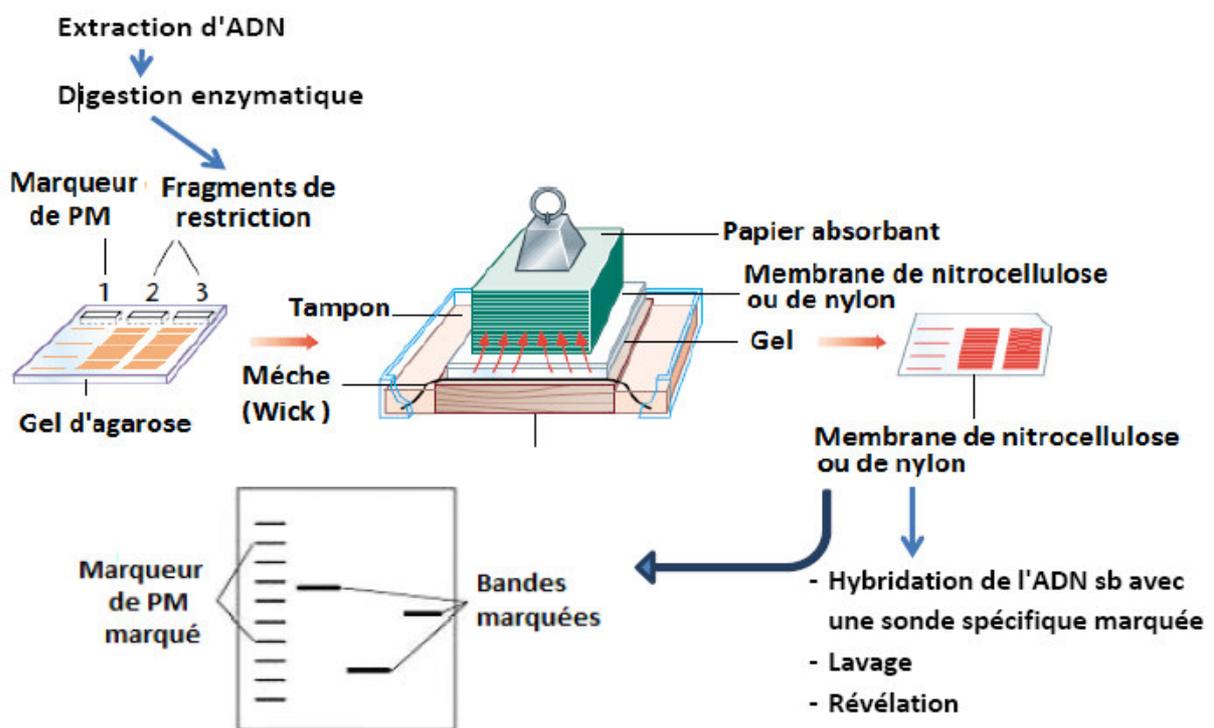


Figure 2: Southern Blot

II.1.3 Hybridation in situ (HIS) :

L'**hybridation in situ** est utilisée pour repérer une séquence donnée d'ARN ou d'ADN à l'intérieur d'une cellule. Elle peut être réalisée sur :

- des coupes de tissus
- des chromosomes métaphasiques (localisation des gènes sur les chromosomes eucaryotes)
- des colonies bactériennes (permet de repérer les colonies recombinantes qui ont intégré le vecteur contenant la séquence d'intérêt lors d'une expérience de clonage de gène)

L'hybridation in situ est aussi appelée **technique FISH** = Fluorescent In Situ Hybridation

L'HIS est un outil incomparable pour étudier l'expression des gènes. Elle est très proche, dans son principe, des Southern et des Northern blots et repose, comme eux, sur l'hybridation d'une sonde d'acide nucléique (ADN ou ARN) marquée avec une séquence complémentaire d'acides nucléiques que l'on cherche à identifier et à localiser.

II.1.4 Hybridation sur colonie, criblage d'une banque de plasmide :

Le principe est le même qu'en Southern ou en northern mais les fragments d'ADN sont séparés par clonage dans des bactéries. L'ADN d'une espèce peut être sous forme fragmenté dans un plasmide, dans une bactérie, si les fragments font 3000 bp en moyenne, pour l'homme il représente 1/1000 000 du génome. L'ensemble de ces bactéries s'appelle une banque. Les bactéries sont étalées sur une boîte de pétri, chaque bactérie donne une colonie, un clone. Si chacune d'entre elle porte un plasmide contenant un fragment d'ADN humain, on peut trouver celle qui porte un fragment complémentaire à une séquence donnée par hybridation (voir figure 3).

- Culture pure sur boîte de Petri (colonies ou plage de lyse).

Chaque colonie contient un plasmide qui porte un morceau de génome. Toutes les bactéries d'une colonie ont le même plasmide et constituent un clone. Deux colonies différentes portent deux plasmides différents, avec chacun un morceau de génome différent.

- On transfère quelques bactéries de chaque clone sur nitrocellulose.
- On lyse les bactéries présentes sur le filtre et on dénature l'ADN par la soude. On fixe l'ADN sur la membrane, et on sature le filtre avec de l'ADN.

- On effectue une hybridation avec une sonde marquée, on lave et on révèle l'hybridation par autoradiographie.

Cette technique permet de visualiser une séquence complémentaire à la sonde, présente dans le clone, dans la colonie bactérienne, dans le plasmide.

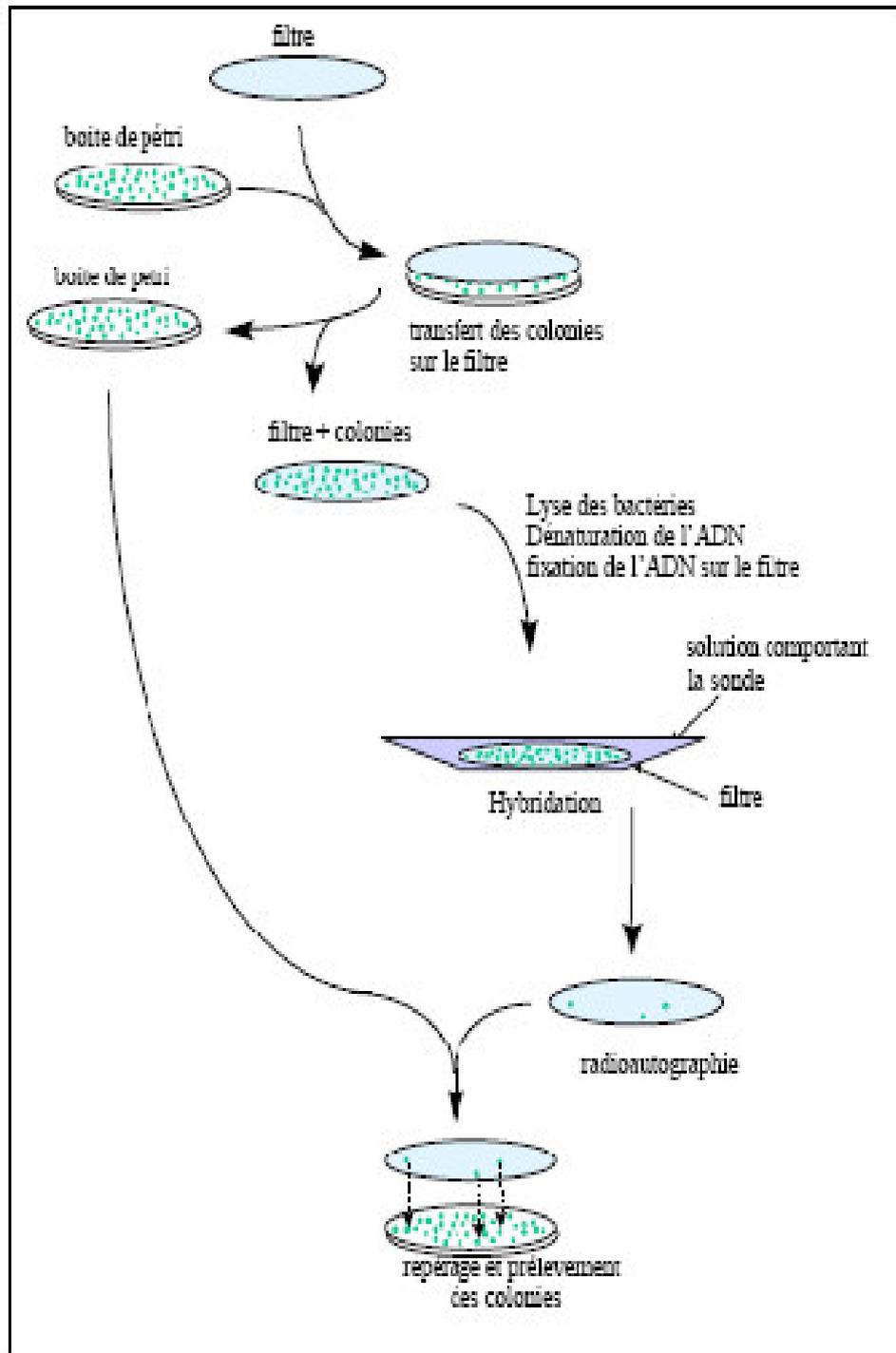


Figure 3 : Hybridation sur colonie

Bibliographie :

Vincent Ecochard, Didier Fournier, Laurence Nieto, Laurent Paquereau. (2011). Techniques et Stratégies en Biologie Moléculaire. Première partie. Université Paul Sabatier, Toulouse, 201p.

Passarge E. (2007). Color Atlas of Genetics. 3rd Ed. Flexibook. Thieme Stuttgart, NY. USA. 486p.

Clémence Allain. (2006) Sondes fluorescentes pour l'ADN : marquages covalent et non-covalent. Université Pierre et Marie Curie - Paris VI, 260p.

D. Tagu, C. Moussard, (2003). Principes des techniques de biologie moléculaire, , 2e éd., INRA-Quæ (ISBN 2-7380-1067-9).

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 157-158p.

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES. UE2. Ellipses 58-65p.

Medraoui Leila. (2015). Cours Biologie Moléculaire ; Initiation aux techniques usuelles de biologie moléculaire. Faculté des sciences. Université Mohamad V Rabat 132p.

Strachan Tom, Read Andrew. (2004). Génétique Moléculaire Humaine. Hybridation des acides nucléiques : principes et applications. 4^e édition ; Médecine Sciences publications. Lavoisier 16p.

Amplification in vitro des acides nucléiques PCR et RT-PCR :

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.

2. Principe :

La réaction PCR (Polymerase Chain Reaction) permet d'amplifier *in vitro* une région spécifique d'un acide nucléique donné afin d'en obtenir une quantité suffisante pour le détecter et l'étudier.

Pour se faire, une série de réactions permettant la réplication d'une matrice d'ADN double brin est répétée en boucle. Ainsi, au cours de la réaction PCR, les produits obtenus à la fin de chaque cycle servent de matrice pour le cycle suivant, l'amplification est donc exponentielle.

Pour avoir réplication d'un ADN double brin, il faut agir en trois étapes.

- 1- Il faut dénaturer l'ADN pour obtenir des matrices simple brin : **dénaturation**.
- 2- Borner et amorcer la réplication de la séquence à amplifier à l'aide d'oligonucléotides amorces spécifiques : **hybridation**.
- 3- Réaliser la réaction de polymérisation du brin complémentaire. A la fin de chaque cycle, les produits sont sous forme d'ADN double brin : **polymérisation**.

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. Cet appareil permet d'exposer les tubes à des températures choisies et pour des durées déterminées par l'expérimentateur. La réaction PCR est extrêmement rapide et ne dure que quelques heures (2 à 3 heures pour une PCR de 30 cycles).

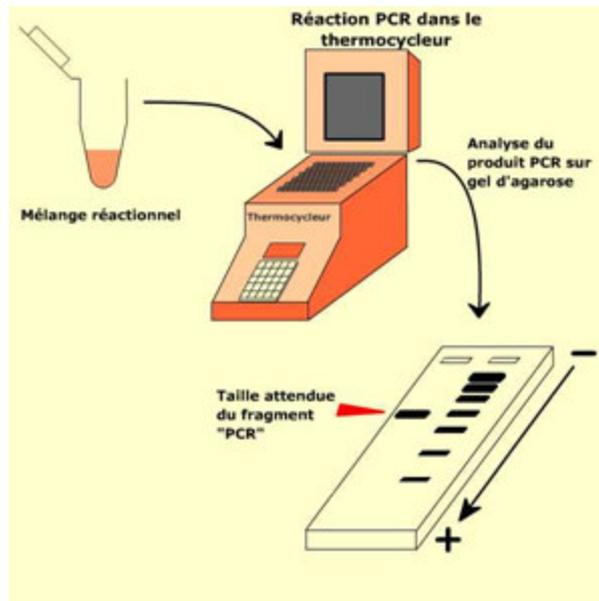


Figure 1: Réaction d'amplification d'ADN dans le thermocycleur et visualisation du produit de PCR par électrophorèse.

3. Les amorces :

Dans la mise au point de la réaction PCR, le choix des amorces est crucial. Elles vont avoir un double rôle : en s'hybridant à l'ADN matrice, elles délimitent la région d'ADN à amplifier (étape 2 du cycle) et avec leur extrémité 3' OH libre servir d'amorce pour l'ADN polymérase (étape 3 du cycle).

Les oligonucléotides amorces s'hybrident aux extrémités de la séquence qui va être amplifiée, il faut donc connaître les séquences nucléotidiques des extrémités de la région ADN amplifiée. C'est en effet au niveau de ces extrémités que les oligonucléotides amorces vont s'hybrider. Pour faciliter le choix des séquences des deux amorces, des logiciels d'analyse de séquences permettent de vérifier les points suivants :

- des **T_m** comparables. Les oligonucléotides amorces doivent s'hybrider à l'ADN matrice dans les mêmes conditions de température
- des séquences qui ne soient pas complémentaires entre elles (en particulier dans la région 3')

- des séquences qui ne contiennent pas de séquences répétées inversées (pas de repliement intramoléculaire).

Il est alors possible de faire synthétiser les deux oligonucléotides (d'une vingtaine de bases). La synthèse chimique d'ADN permet d'obtenir des oligonucléotides dont l'extrémité 5' n'est pas phosphorylée (5'OH) contrairement aux ADN naturels. Ainsi, tous les fragments d'ADN amplifiés par PCR sont déphosphorylés à leurs extrémités 5'.

4. Les températures :

Les trois étapes, constituant un cycle de PCR, sont effectuées à des températures différentes permettant de contrôler l'activité enzymatique :

- l'étape de **dénaturation**, est réalisée à environ 95°C, pour une dissociation complète des deux brins d'ADN.
- l'étape d'**hybridation** se fait à une température qui sera définie selon la nature des amorces (cette température varie de 50 à 60°C). Cette température va déterminer la stabilité des hybrides une fois que l'appariement amorces / matrice est réalisé.
- l'étape de **polymérisation** est à environ 72°C, température de « travail » de l'ADN polymérase thermorésistante utilisée. Au cours de cette étape, les brins complémentaires d'ADN sont synthétisés à partir des extrémités 3'OH libre des amorces hybridées.

Les températures de dénaturation et de polymérisation sont fixes, seule la température d'hybridation devra être calculée pour chaque nouvelle PCR. Cette température d'hybridation dépend de la composition en bases des oligonucléotides amorces. Elle est légèrement inférieure (environ de 5°C) au T_m qui est la température de demi-dénaturation.

Pour calculer le T_m d'un oligonucléotide inférieur à 30 nucléotides, on utilise la relation suivante :

$$T_m = 2 (A + T) + 4 (G + C)$$

(Où A, T, G et C sont respectivement le nombre de chacune de ces bases dans l'oligonucléotide).

5. L'ADN matriciel :

En théorie une copie ADN de la séquence recherchée est suffisante pour avoir une amplification, il faut cependant tenir compte de la probabilité de « rencontre » des molécules d'ADN matrice avec les amorces. Dans la pratique plusieurs copies sont nécessaires pour avoir un résultat correct. Mais attention, la mauvaise qualité et/ou une quantité trop importante d'ADN matrice peut conduire à une amplification aspécifique, voire à une inhibition enzymatique. Il est possible d'expliquer cette inhibition par la présence de contaminants provenant de l'échantillon ou des réactifs utilisés dans les protocoles d'extraction d'ADN.

6. L'ADN polymérase :

Les premières ADN polymérases utilisées provenaient d'une bactérie thermophile (résistante à des températures très élevées), comme par exemple *Thermus aquaticus* (Taq polymérase). De nos jours, les enzymes utilisées sont dites recombinantes, ce qui simplifie considérablement leur obtention, et leurs propriétés ont été largement modifiées pour les rendre plus efficaces et plus fidèles.

7. Le tampon :

Le tampon utilisé pour la réaction PCR sert à maintenir stable le pH du milieu réactionnel au niveau optimal pour la Taq polymérase (TrisHCl à pH basique 8,5 à 9). Il contient des cations bivalents Mg^{2+} , cofacteurs indispensables pour la réaction de polymérisation avec la Taq polymérase. La présence dans le milieu réactionnel des cations bivalents Mg^{2+} et de cations monovalents (K^+ ou NH_4^+) vont neutraliser les charges négatives des groupements phosphates au niveau de l'ADN et ainsi stabiliser les hybrides ADN/ADN. En pratique, la concentration en sel doit cependant rester compatible avec l'activité de l'ADN polymérase.

8. Aspects quantitatifs de la réaction PCR :

Le nombre de cycle est défini par l'expérimentateur selon le degré d'amplification souhaité, dans la plupart des cas, ce nombre est d'environ 30. En théorie (et seulement en théorie!) il y a un facteur d'amplification de l'ordre de 2^{30} soit un facteur d'un milliard. En pratique les conditions expérimentales évoluent au fur et à mesure de l'avancée des cycles, le rendement

de la réaction de polymérisation évolue de même, on peut espérer obtenir avec 30 à 40 cycles, un facteur d'amplification de l'ordre de 10^5 à 10^6 .

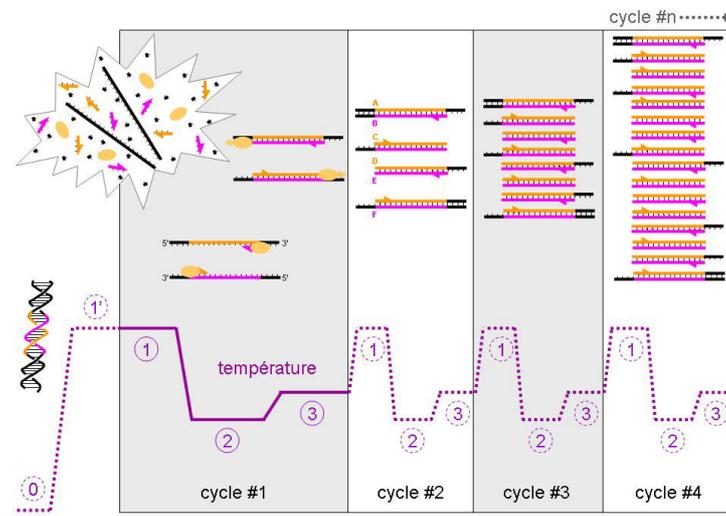


Figure 2 : Evolution de la température et des différents types de brins d'ADN au cours des 4 premiers cycles de la PCR

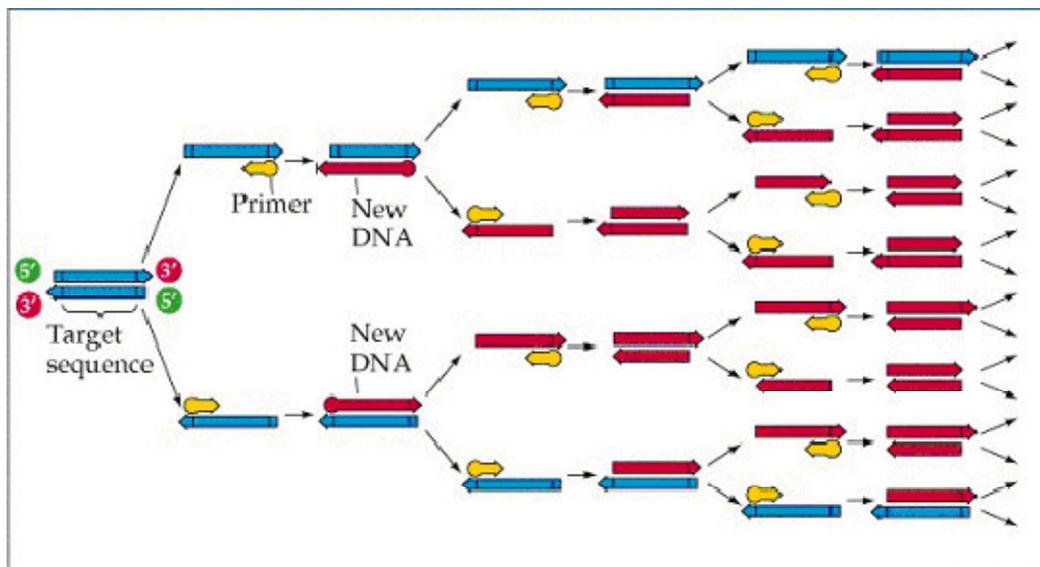


Figure 3: Amplification d'ADN par PCR. Le produit d'un cycle sert de matrice pour le cycle qui suit.

9. RT-PCR :

La RT-PCR se déroule en deux phases. Une première phase correspond à la copie d'ARN messager en ADN complémentaire (ADNc) et une seconde phase correspond à une réaction PCR classique sur l'ADNc synthétisé.

Dans la première phase, l'ARN messager à étudier est repéré en utilisant une sonde oligonucléotidique spécifique (amorce 1 qui s'hybride à l'extrémité 3' du seul mARN auquel on s'intéresse), puis la transcriptase inverse (ou rétrotranscriptase) permet la synthèse du brin complémentaire (sous une forme de ADNc simple brin), une seconde amorce oligonucléotidique spécifique (amorce 2) permettra la synthèse du second brin par extension.

L'ADN complémentaire synthétisé servira ensuite de matrice pour une réaction PCR classique.

La technique RT-PCR a permis de montrer que la transcription de tous les gènes s'effectuait dans tous les tissus et ceci même pour les gènes qui présentent une très grande spécificité tissulaire. On parle dans ces conditions de transcription illégitime. Il est évident qu'avant les techniques d'amplification génique, la sensibilité des méthodes classiques n'avait pas permis de mettre en évidence un tel phénomène.

10. Utilisations des produits PCR :

Les utilisations des produits PCR sont très variées, nous citerons quelques exemples:

-Mise en évidence de mutations ponctuelles par hybridation des produits PCR avec des sondes oligonucléotidiques :

Les produits PCR peuvent après transformation en monobrin être hybridés avec des sondes oligonucléotidiques fixées sur un support solide. Ces sondes correspondent à des séquences normales et pathologiques (présence de mutations ponctuelles par exemple) pour un gène donné.

-Analyse de restriction :

Le produit PCR est soumis à une digestion enzymatique par une enzyme de restriction. Si une mutation ponctuelle modifie le site de restriction initialement présent, la taille des fragments d'ADN obtenus après digestion sera modifiée et décelable après électrophorèse des fragments d'ADN (sur gel d'agarose ou gel de polyacrylamide).

-Introduction du produit PCR dans un vecteur: clonage du produit PCR.

-Séquençage direct du produit PCR

Bibliographie :

Vincent Ecochard, Didier Fournier, Laurence Nieto, Laurent Paquereau. (2011). Techniques et Stratégies en Biologie Moléculaire. Première partie. Université Paul Sabatier, Toulouse, 201p.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 155-157p.

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES. UE2. Ellipses 70-74p.

D. Tagu, C. Moussard, (2003). Principes des techniques de biologie moléculaire, , 2e éd., INRA-Quæ (ISBN 2-7380-1067-9).

Medraoui Leila. (2015). Cours Biologie Moléculaire ; Initiation aux techniques usuelles de biologie moléculaire. Faculté des sciences. Université Mohamad V Rabat 132p.

Strachan Tom, Read Andrew. (2004). Génétique Moléculaire Humaine. Hybridation des acides nucléiques : principes et applications. 4 e édition ; Médecine Sciences publications. Lavoisier 16p.

Le séquençage de l'ADN :

Introduction :

La séquence d'ADN contient l'information nécessaire aux êtres vivants pour survivre et se reproduire. Déterminer cette séquence est donc utile aussi bien pour les recherches visant à savoir comment vivent les organismes que pour des sujets appliqués. En médecine, elle peut être utilisée pour identifier, diagnostiquer et potentiellement trouver des traitements à des maladies génétiques et à la virologie. En biologie, l'étude des séquences d'ADN est devenue un outil important pour la classification des espèces.

1. Historique :

Les premières techniques de séquençage ont été développées en parallèle au milieu des années 1970. Les méthodes de **Sanger** (Grande-Bretagne) et **Gilbert** (Etats-Unis) ont toutes deux été récompensées d'un prix Nobel de chimie en 1980.

Le premier organisme a été séquencé en 1977. Il s'agissait du virus bactériophage ϕ X174, possédant un ADN simple brin ne nécessitant donc pas l'étape de dénaturation utilisé dans les méthodes de Sanger et Maxam et Gilbert.

Depuis une trentaine d'année, l'amélioration de la technique de production des amorces, de l'amplification des brins et la généralisation des traceurs fluorescents ont permis d'améliorer considérablement les techniques de séquençage. L'apparition des **séquenceurs automatiques** a notamment permis l'automatisation de ces technologies et ont contribué à la finalisation du génotypage humain en 2003.

1.1 Définition de séquençage :

Le **séquençage de l'ADN** est une technique d'analyse de l'ADN, consiste à déterminer l'ordre d'enchaînement des nucléotides pour un fragment d'ADN donné. Actuellement la technique de séquençage repose majoritairement sur la méthode enzymatique de Sanger.

1.2 Les différents acteurs du séquençage :

- l'ADN provient des organismes dont on souhaite séquencer le génome. L'ADN, le plus souvent sous la forme double-brin, est dénaturé afin de séparer les deux brins. Celui qui sera séquencé s'appelle le brin « matrice »

- les **nucléotides**, ou plutôt désoxynucléotides, sont les bases de l'ADN (A, C, G ou T). Ils sont attachés les uns aux autres grâce à des liaisons chimiques nécessitant la présence d'un groupement OH particulier (sur le carbone 3' du ribose)
- les **didésoxynucléotides** sont des nucléotides privés du groupement OH en position 3'. Leur incorporation dans une chaîne d'ADN interrompt définitivement la synthèse de l'ADN (voir figure 1).
- une **amorce** est un brin d'ADN très court (une vingtaine de nucléotide), qui peut s'hybrider à une séquence complémentaire spécifique
- l'**ADN polymérase** est une enzyme qui a pour rôle de copier l'ADN, en synthétisant un brin complémentaire au brin matrice. L'ADN polymérase ne fonctionne qu'en ajoutant des nucléotides à **une amorce** déjà présente, en fonction de la succession de nucléotides du brin matrice (un A est toujours positionné en face d'un T et un C toujours en face d'un G).

1.3 Méthode de Sanger :

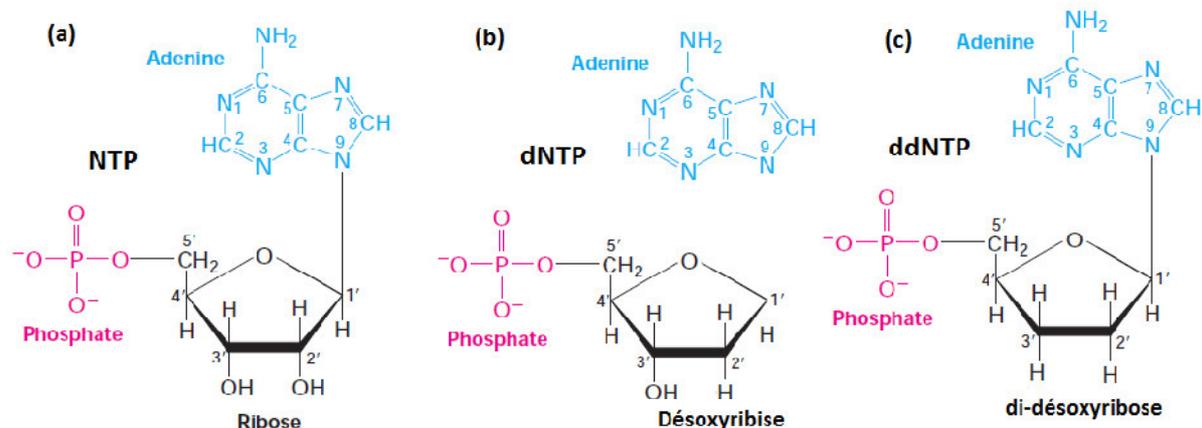


Figure 1 : les différentes formes des nucléotides.

Dans le ddATP, le **groupement 3'-OH** est remplacé par un **hydrogène**. Cette modification empêche la poursuite de la synthèse de l'ADN qui continue normalement sur le 3-OH

1.4 Principe du séquençage par la méthode de Sanger :

Le principe de cette méthode consiste à initier la polymérisation de l'ADN à l'aide d'un petit oligonucléotide (amorce) complémentaire à une partie du fragment d'ADN à séquencer. L'élongation de l'amorce est réalisée par une ADN polymérase I dépourvue d'activité exonucléase 5'→3' et maintenue par des ADN polymérases thermostables, celles qui sont

utilisées pour la PCR. On utilise quatre tubes séparés contenant les quatre désoxyribonucléotides (dATP, dCTP, dGTP, dTTP), ainsi qu'une faible concentration de l'un des quatre didésoxyribonucléotides (ddATP, ddCTP, ddGTP ou ddTTP).

Ces didésoxyribonucléotides agissent comme des « poisons » terminateurs de chaîne : une fois incorporés dans le nouveau brin synthétisé, ils empêchent la poursuite de l'élongation. Cette terminaison se fait spécifiquement au niveau des nucléotides correspondant au didésoxyribonucléotide incorporé dans la réaction. Pour le séquençage complet d'un même fragment d'ADN, on répète cette réaction quatre fois en parallèle, avec les quatre didésoxyribonucléotides différents.

Par exemple, dans la réaction où on a ajouté du ddGTP, la synthèse s'arrête au niveau des G. Le mélange réactionnel contenant, à la fois du dGTP et un peu de ddGTP, la terminaison se fait de manière statistique suivant que l'ADN polymérase utilise l'un ou l'autre de ces nucléotides. Il en résulte un mélange de fragments d'ADN de tailles croissantes, qui se terminent tous au niveau d'un des G dans la séquence. Ces fragments sont ensuite séparés par électrophorèse sur un gel de polyacrylamide, ce qui permet ainsi de repérer la position des G dans la séquence (voir figure 2).

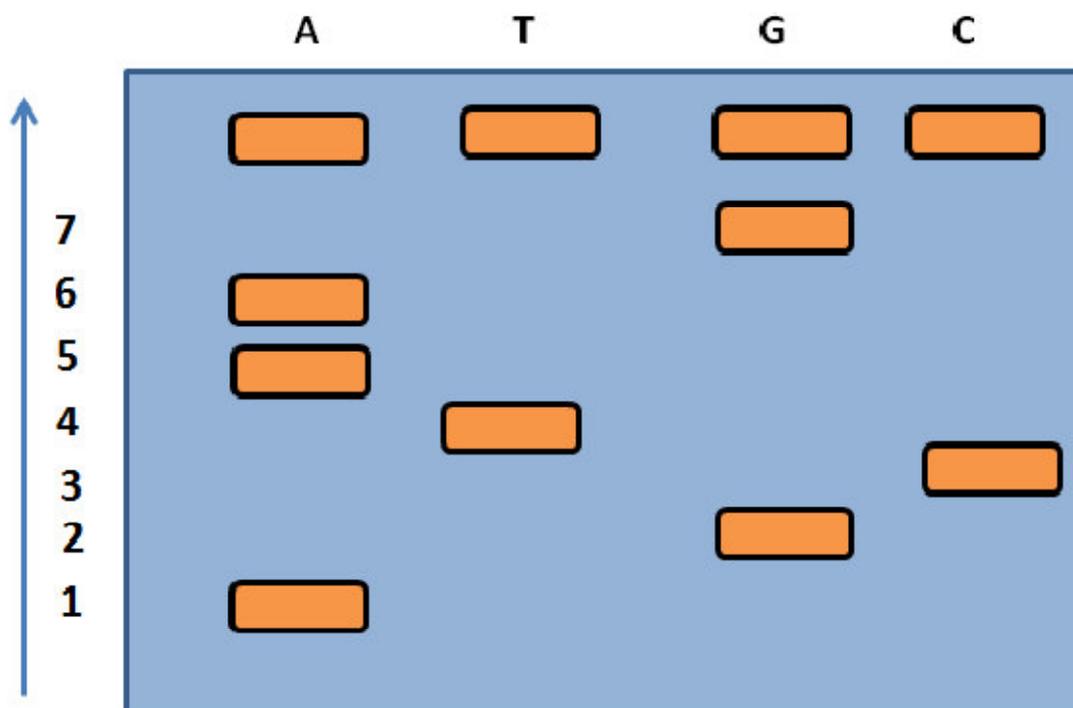


Figure 2: Séquençage de l'ADN par la méthode de Sanger

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

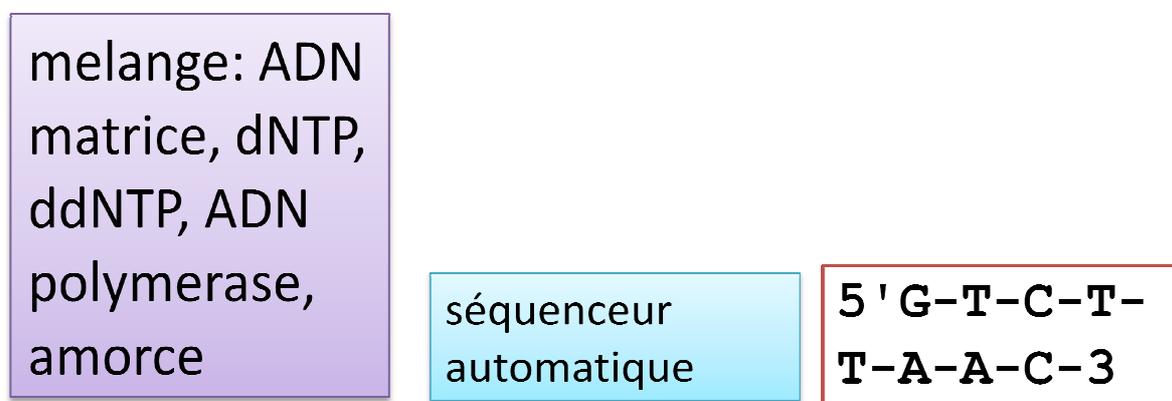
1- Synthèse du brin complémentaire d'un brin matrice à l'aide d'une polymérase à partir d'une amorce, en présence des 4 dNTPs et d'un ddNTP qui joue le rôle de terminateur de chaîne (absence de 3'-OH). Quatre réactions en parallèle sont nécessaires, chacune avec l'un des terminateurs présents en faible quantité.

2- Séparation par électrophorèse des fragments synthétisés, le nombre de ces fragments dépend du nombre de résidus nucléotidiques. Donc la position des terminateurs dans la séquence (figure 2).

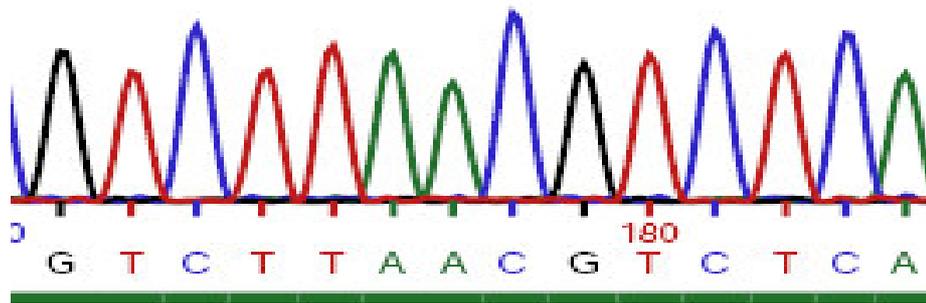
La détection des fragments ainsi synthétisés se fait en incorporant un traceur dans l'ADN synthétisé. Initialement ce traceur était radioactif; aujourd'hui, on utilise des traceurs fluorescents, attachés soit à l'oligonucléotide, soit au didésoxyribonucléotide.

1.5 Automatisation du séquençage :

Des séquenceurs automatiques sont capables de réaliser les réactions de séquence puis de les lire.



- Les didésoxynucléotides sont marqués avec des fluorescences différentes, cela permet de faire une seule réaction et une seule lecture. Le séquenceur détecte la fluorescence et repère la taille des fragments d'ADN.



- On obtient des courbes de fluorescences qui sont interprétées en terme de nucléotides

Bibliographie :

Vincent Ecochard, Didier Fournier, Laurence Nieto, Laurent Paquereau. (2011). Techniques et Stratégies en Biologie Moléculaire. Première partie. Université Paul Sabatier, Toulouse, 201p.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 159-165p.

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES. UE2. Ellipses 51-57p.

D. Tagu, C. Moussard, (2003). Principes des techniques de biologie moléculaire, , 2e éd., INRA-Quæ (ISBN 2-7380-1067-9).

Medraoui Leila. (2015). Cours Biologie Moléculaire ; Initiation aux techniques usuelles de biologie moléculaire. Faculté des sciences. Université Mohamad V Rabat 132p.

Strachan Tom, Read Andrew. (2004). Génétique Moléculaire Humaine. Hybridation des acides nucléiques : principes et applications. 4 e édition ; Médecine Sciences publications. Lavoisier 16p.

PARTIE 2 : génie génétique :

Chapitre 1 : clonage in vivo :

Introduction :

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.

I. Clonage :

I.1 Un clone correspond à un grand nombre de molécules ou de cellules identiques provenant d'un seul ancêtre (cellule ou molécule). L'opération qui consiste à obtenir ce grand nombre de cellules ou de molécules s'appelle *le clonage*.

I.1.1 Le clonage nucléique :

est une technique de biologie moléculaire qui consiste en l'insertion d'un fragment d'ADN dans un **vecteur**, ce vecteur étant propagé dans une **cellule hôte**. La culture de cette cellule et la purification ultérieure du vecteur permettent de produire des quantités quasiment illimitées du fragment d'ADN cloné que l'on désire étudier.

Cette technique de biologie moléculaire peut être utilisée pour un clonage partiel, ne portant que sur un fragment de matériel génétique (ADN), mais aussi pour le clonage d'un gène entier permettant la production de la protéine recombinante correspondante.

I.1.2 Les vecteurs :

Les fragments obtenus sont dépourvus de moyens de se répliquer (origine de répllication) et d'extrémités protégées. Il est donc nécessaire de les inclure dans une structure qui assurera leur propagation et leur protection. Ces structures sont appelées des vecteurs.

Un vecteur de clonage est un petit élément génétique à répllication autonome, utilisé pour produire de multitude de copie du gène d'intérêt.

Vecteur natif lorsqu'il ne contient que son ADN propre.

Vecteur recombiné lorsqu'il a intégré un fragment d'ADN étranger, quel qu'il soit.

Une origine de répllication est adaptée au mode de répllication de la cellule dans laquelle se multiplie le vecteur. Une origine de répllication de type eucaryote ne fonctionnera pas dans une bactérie et réciproquement. Lorsque l'on veut passer d'un type de cellule à l'autre il faut des vecteurs qui possèdent les deux types d'origine de répllication.

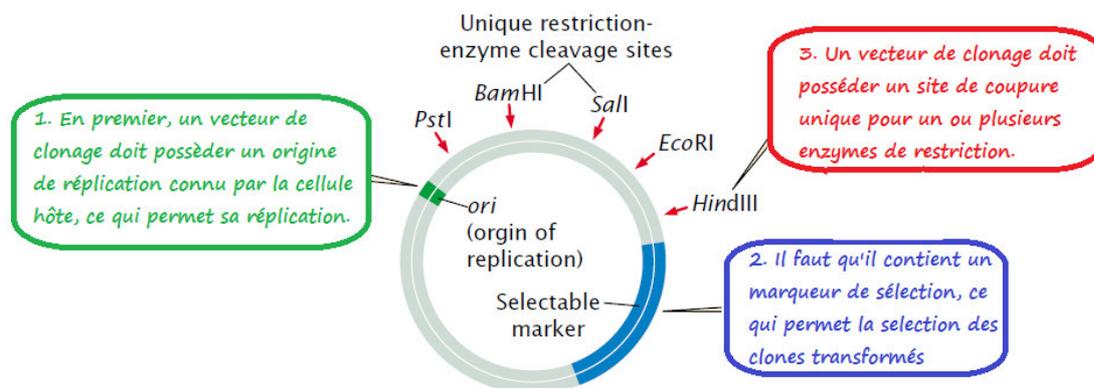


Figure 1: Les trois caractéristiques d'un vecteur de clonage idéal. (Aouf, 2016)

Pour protéger des extrémités d'ADN deux moyens existent.

- créer à chaque extrémité un bout collant compatible avec l'autre et religuer les deux extrémités l'une avec l'autre. On obtient un cercle d'ADN sans extrémités libres. C'est ce qui est réalisé dans les plasmides, chez le bactériophage lambda (chez les cosmides, BAC (chromosome artificiel bactérien) d'E. coli et PAC de P1 (chromosome artificiel du bactériophage P1).

On peut artificiellement coller aux extrémités de l'ADN linéaire des télomères d'origine chromosomique qui assureront la protection des bouts, c'est la voie suivie pour la construction des chromosomes artificiels.

I.1.2.A ADN recombinant :

Il s'agit d'un ADN hybride obtenu au laboratoire par la combinaison de deux ADN appartenant à deux espèces différentes.

I.1.2.B Les banques de clones :

La construction d'une banque de clones est une étape préalable nécessaire à la cartographie physique. L'étude d'un génome débute toujours par le découpage de l'ADN (digestion par des enzymes de restriction) en fragments qui sont insérés dans des constructions moléculaires appelées vecteurs de clonage. Ces derniers sont ensuite intégrés dans des cellules qui assureront le maintien et la réplication du fragment d'ADN. Un ensemble de cellules portant des fragments d'ADN de l'espèce étudiée constitue une banque de clones (un fragment par clone).

L'objectif de la banque est donc de conserver une collection de fragments d'ADN :

- bien identifiés,
- facilement manipulables,
- que l'on peut produire en grande quantité (en vue de leur étude).

On pourra notamment alors procéder au séquençage de chacun de ces fragments.

II. Les éléments nécessaires au clonage :**II.1 Les vecteurs de clonage: plasmides ; cosmides ; phages ; BAC ; PAC ; YAC :****Généralités :**

On appelle vecteur l'ADN dans lequel on insère le fragment d'ADN à étudier. L'ADN inséré est appelé insert ou ADN étranger ou ADN exogène. Cette séquence nucléotidique est capable de s'auto-répliquer.

Les vecteurs sont donc des petits ADN dans lesquels on insère un fragment d'ADN que l'on veut étudier. Ces petits ADN sont généralement des plasmides ou des bactériophages. Il est nécessaire d'introduire ces bactériophages ou plasmides dans les bactéries pour réaliser une multiplication de ceux-ci.

II.1.A Les plasmides :

Les plasmides sont des petits fragments d'ADN circulaire présents dans la cellule bactérienne et indépendants du génome bactérien.

II.1.A.1 Propriétés générales des plasmides :

Les plasmides présentent les caractéristiques suivantes:

- Leur ADN est bicaténaire, circulaire avec un nombre de nucléotides inférieur à 10 kb (1 kilobase = kb = 1000 nucléotides).

- Le nombre de plasmides dans une cellule bactérienne peut être considérable (plusieurs centaines).
- Les plasmides portent normalement des gènes qui leur confèrent un avantage sélectif, par exemple une résistance à un antibiotique, résistance aux métaux lourds, dégradation de composés aromatiques.
- Leur réplication est indépendante de celle du génome bactérien.
- 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.

II.1.A.2 Préparation des plasmides :

Les bactéries contenant les plasmides sont mises en culture dans un grand volume (jusqu'à 2 litres de milieu de culture). Quand la concentration bactérienne atteint une valeur suffisante, on traite par un antibiotique (comme le chloramphénicol) qui bloque la réplication du chromosome bactérien mais pas celle du plasmide. On récupère les bactéries par centrifugation. Puis, on perméabilise la paroi bactérienne par un traitement doux permettant d'obtenir un passage en solution des plasmides qui sont plus légers que l'ADN bactérien. La purification des plasmides peut être ensuite réalisée par chromatographie liquide à haute performance ou ultra-centrifugation en gradient de densité.

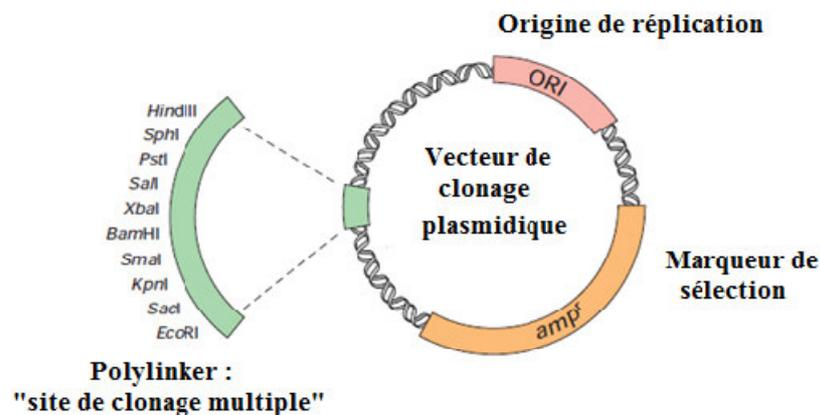


Figure 2: Composants basiques d'un vecteur plasmidique qui peut se répliquer chez *E. coli* (Lodish *et al.*, 2003).

II.1.A.3 Préparation des plasmides pour la recombinaison, constitution de l'hybride et transformation bactérienne :

L'insertion d'une séquence d'ADN bicaténaire dans un plasmide nécessite un traitement préalable par une enzyme de restriction. On choisit une enzyme qui a un site unique dans la séquence plasmidique. Après action de l'enzyme de restriction, le plasmide circulaire est linéarisé. Pour maintenir la linéarisation, il est indispensable de traiter le plasmide linéaire par une phosphatase alcaline qui enlève les groupements phosphates en 5'. Le plasmide linéarisé et traité par une phosphatase alcaline est ajouté au fragment d'ADN à insérer en présence d'une ligase. L'ensemble de ces étapes produit **un plasmide recombinant**.

Les bactéries peuvent être placées dans certaines conditions physico-chimiques pour permettre aux plasmides d'être incorporés à l'intérieur de celles-ci. Les bactéries traitées et prêtes à recevoir l'ADN étranger sont appelées **bactéries compétentes**. L'incorporation des plasmides dans les bactéries est appelée transformation bactérienne (figure 3).

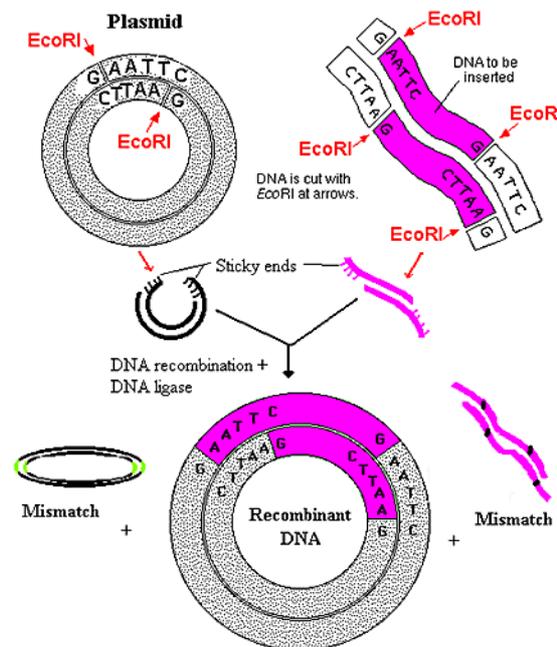


Figure 3 : insertion d'ADN dans le plasmide

II.1.A.4 Sélection des bactéries transformées par des plasmides :

Le taux de transformation des bactéries est généralement très faible. Il est fondamental de disposer de méthodes permettant d'obtenir des bactéries transformées par les plasmides et seulement celles-ci. Pour cela, l'artifice consiste à utiliser *une résistance à un antibiotique* (par exemple: l'ampicilline). La transformation d'une bactérie sensible à un antibiotique par des plasmides portant un gène de résistance au même antibiotique aboutit à l'apparition d'une résistance uniquement pour les bactéries ayant incorporé les plasmides.

Cette technique permet de séparer les bactéries comportant des plasmides et les bactéries sans plasmides car ces dernières sont tuées. Cependant, *elle ne permet pas de distinguer les bactéries avec plasmides intacts des bactéries avec plasmides recombinants*. On utilise pour cela une résistance à un second antibiotique, par exemple: une tétracycline. L'insertion du fragment d'ADN dans le plasmide doit inactiver le gène de résistance au deuxième antibiotique. Les bactéries transformées par les plasmides recombinants sont donc résistantes au premier antibiotique (ampicilline) et sensibles au deuxième antibiotique (tétracycline). Les bactéries non transformées sont sensibles aux deux antibiotiques.

II.1.A.5 Un exemple de plasmide: le puc19

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).

Un plasmide pUC est un plasmide pBR dans lequel on a remplacé le gène de résistance à la tétracycline (qui sert à repérer les plasmides recombinés) par un gène bactérien lacZ qui code la (galactosidase (revoir l'opéron lactose). Cette enzyme a pour propriété d'hydrolyser le lactose en glucose et galactose. Un produit non toxique pour la cellule bactérienne est également substrat de la galactosidase : le Xgal (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.

Un plasmide classique est le pUC19. La séquence complète est connue (2686 nucléotides). Il comporte les particularités suivantes:

- *Un gène de résistance à l'ampicilline (antibiotique)*. La présence de ce gène permettra à la bactérie porteuse de ce plasmide de ne pas être sensible à l'effet de cet antibiotique.
- *Le gène lac Z qui code pour la β -galactosidase dans l'opéron lactose.*
- *Enfin, une région avec des sites multiples et uniques pour des enzymes de restriction connues*. Cette région est appelée " *polylinker* ". Son rôle est de permettre l'insertion du fragment d'ADN étranger (figure 4).

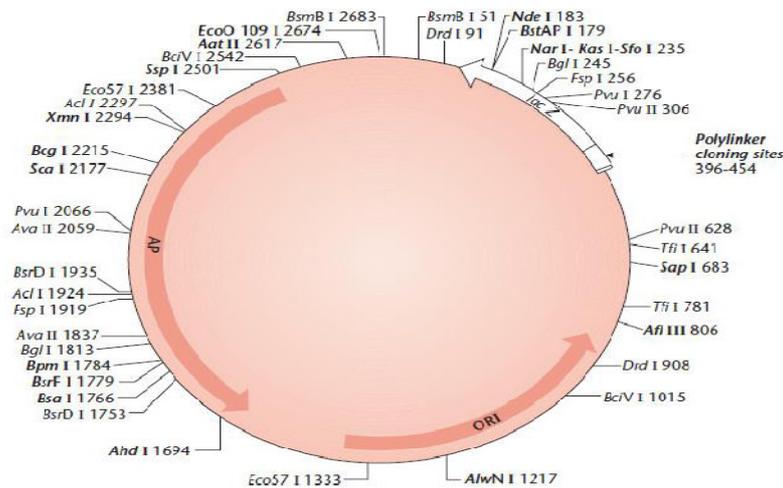


Figure 4: La carte génétique de certains vecteurs de type pUC dérivés de pBR322, le site de clonage multiple (polylinker) est introduit dans le gène *lacZ*, sans interrompre la fonction du gène (Primose et al., 2002).

Ce type de plasmide est intéressant car il montre un système de sélection des bactéries ayant été transformées par les plasmides recombinants. Un système enzymatique appartenant à l'opéron lactose peut être utilisé à la place d'un second antibiotique. L'insertion du fragment d'ADN dans le plasmide pUC 19 aboutit à l'inactivation du gène qui code pour la β -galactosidase. Pour vérifier la présence ou l'absence de l'activité enzymatique β -galactosidase, on utilise un galactoside dont la couleur passe de l'incolore au bleu quand il est clivé par la β -galactosidase, ce composé s'appelle *X-gal*. Pour pouvoir métaboliser le *X-gal*, la cellule doit être exposée à un inducteur, cet inducteur est le *IPTG* (*isopropylthio- β -D-galactoside*).

En culture et en présence d'*IPTG* et de *X-gal*, les bactéries résistantes à l'ampicilline et transformées par les plasmides recombinants se présentent sous forme de colonies blanchâtres car elles ont perdu la capacité de clivage de l'équivalent coloré du lactose (le *X-gal*) par la β -galactosidase. Par contre, les bactéries résistantes à l'ampicilline et non transformées par les plasmides recombinants se présentent sous forme de colonies bleues. La sélection visuelle des bactéries transformées par les plasmides recombinants est donc possible (figure 5).



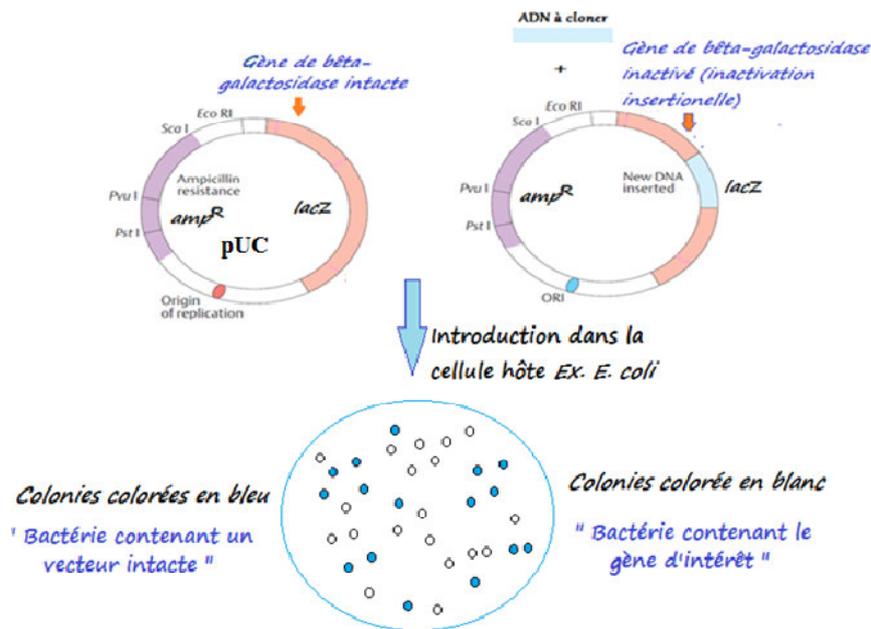


Figure 5 : Sélection des bactéries transformées portant un plasmide pUC recombiné (Aouf, 2016)

II.1.A.6 Avantages et désavantages des plasmides :

Avantages:

- Petite taille du vecteur, permettant un travail expérimental aisé.
- Sélection des plasmides recombinants (sélection par les antibiotiques).

Désavantages:

- Faible efficacité pour la transformation des bactéries (pénétration de plasmides).
- Impossibilité d'insérer des larges fragments d'ADN (La capacité maximum de ce plasmide est de 10kb d'ADN étranger).

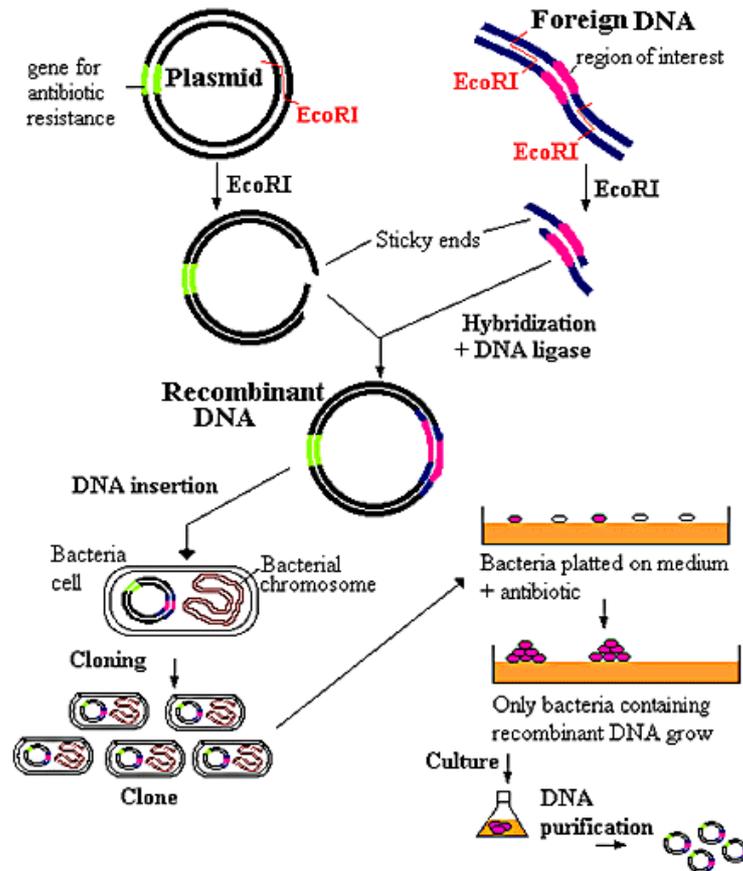


Figure 6 : les vecteurs de type plasmides

II.2 Les phages :

II.2.A Propriétés générales des phages :

Les phages sont des virus qui infectent les bactéries. Deux phages sont très utilisés comme vecteurs: le phage lambda et le phage M13 (mais maintenant les phages dérivés de ces deux phages). Les séquences de ces phages utilisés comme vecteurs sont connues (40 à 50 kb d'ADN double brin). Les extrémités de cet ADN sont simples brins sur une longueur réduite, et complémentaires l'une de l'autre et surtout formant *des extrémités cohésives*. Ces séquences sont appelées *séquences cos* (*pour cohésives*).

II.2.A.1 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* (figure 7) [*cos*: Des éléments important pour la réplication et l'encapsidation de bactériophage λ].

Deux parties sont individualisées dans le phage lambda:

- La tête du virus: elle renferme l'ADN viral.
- La queue du virus: elle renferme des protéines. Elle permet la fixation du virus sur la cellule-hôte bactérienne.

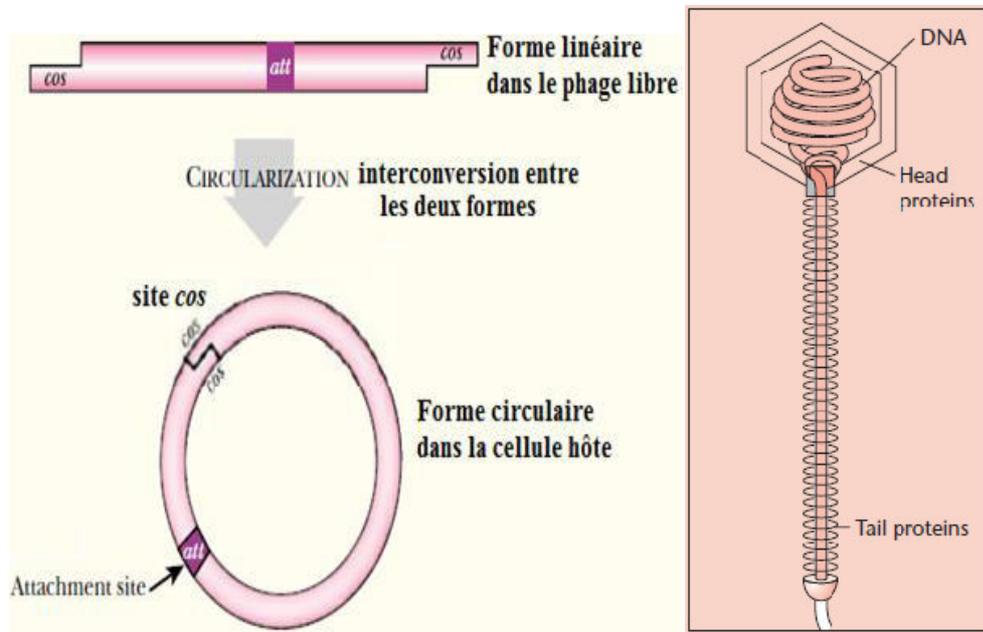


Figure 7: Différentes formes du génome du bactériophage λ , et sa structure (Primose *et al.*, 2002).

Le virus lambda se multiplie selon deux modes possibles: soit par *multiplication lytique*, soit par *multiplication lysogénique*.

- La multiplication lytique :

Le virus s'adsorbe spécifiquement à la surface des bactéries. Il injecte son ADN à la cellule-hôte. L'ADN viral se circularise dans la bactérie. Puis, une phase complexe de réplication débute et aboutit à la constitution de nombreuses particules virales à l'intérieur du cytoplasme bactérien. La lyse bactérienne provoque la libération des particules virales.

- La multiplication lysogénique :

Comme dans la multiplication lytique, le virus s'adsorbe et pénètre dans la bactérie. L'ADN viral s'intègre à l'ADN bactérien et sera répliqué en même que lui.

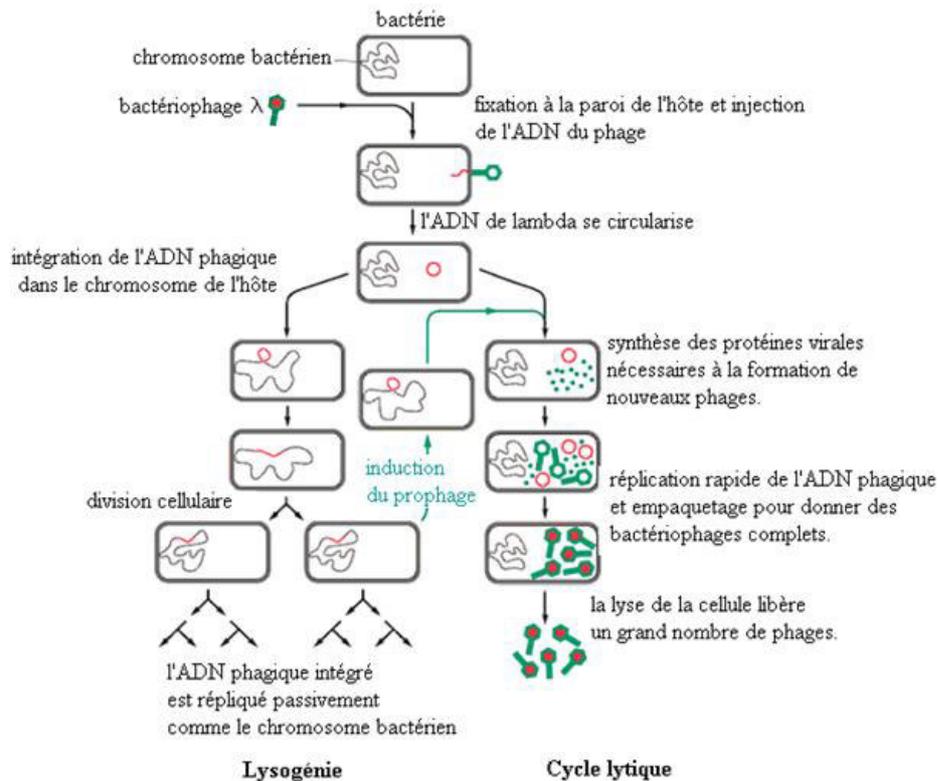


Figure 8: Cycles du bactériophage λ

Le génome du phage lambda peut être divisé en:

- Parties essentielles comprenant:
 - les gènes codant pour les protéines situées dans la tête et les protéines situées dans la queue du virus.
 - Les sites *cos*.
 - Le site de répllication de l'ADN viral.
- Parties non essentielles:
 - Ce sont les gènes impliqués dans la lysogénie (figure 9).

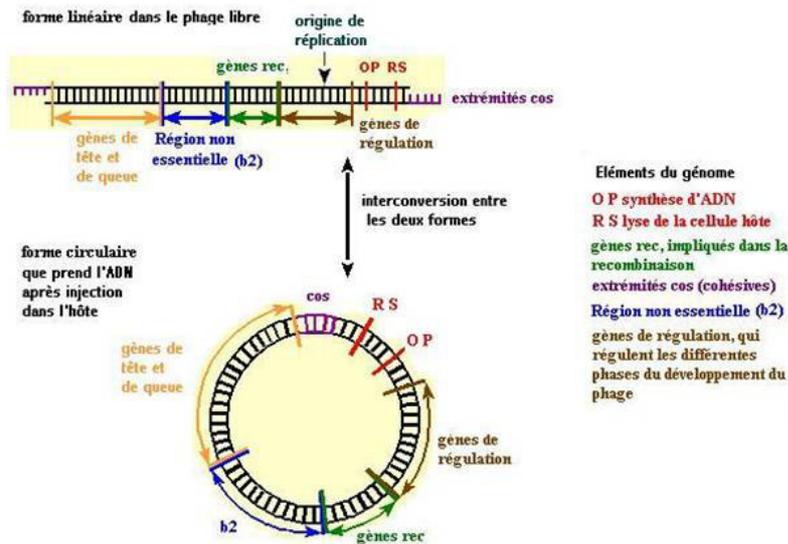


Figure 9 : Carte génétique du bactériophage λ

II.2.B λ Comme vecteur :

Ces propriétés du bactériophage ont permis de concevoir un nouveau vecteur à partir de λ capable de transporter plus d'ADN étranger que les plasmides pBR et pUC.

La région b1 puisqu'elle n'est pas essentielle peut être remplacée par un fragment d'ADN d'une taille équivalente. On a alors un vecteur λ de remplacement de taille voisine de celle du génome naturel du bactériophage. Ce phage accepte des fragments d'ADN de 9 kb à 23 kb.

II.2.B.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).

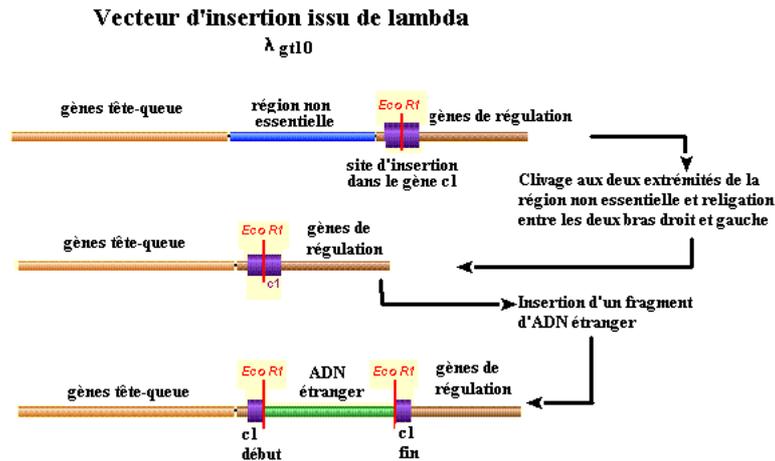


Figure 10: Le bactériophage λ utilisé comme vecteur d'insertion

II.2.C Avantages et désavantages des phages :

Avantages:

- La taille des fragments d'ADN insérables est supérieure à celle des plasmides.
- 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.

Désavantages:

Nombre de sites de restriction restreints dans le génome des phages.

Obligation d'empaqueter l'ADN.

Contraintes de taille pour l'ADN à insérer.

II.3 Les cosmides :

II.3.A Propriétés générales des cosmides :

Les cosmides (\approx 5kb) sont des vecteurs artificiels hybrides: phage lambda-plasmides. En fait, ils se comportent comme des plasmides avec des sites de restriction permettant l'insertion d'ADN étranger. Ils renferment également un gène de résistance aux antibiotiques (ampicilline). De plus, un site cos d'un virus lambda a été inclus dans leur ADN circulaire ce qui permettra au cosmide d'être empaqueté dans la tête d'un virus lambda (figure 11).

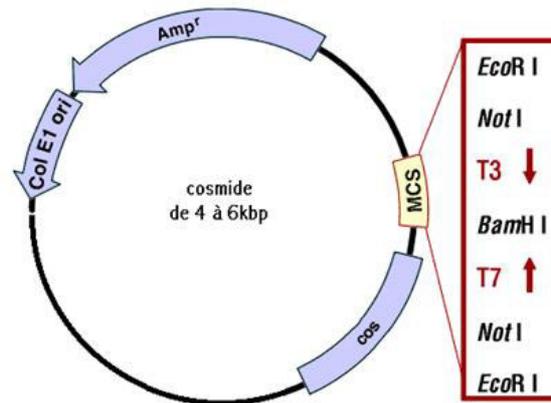


Figure 11: Schéma d'un cosmide (les éléments viennent des vecteurs précédents)

Les cosmides, après assemblage *in vitro* sont capables de se fixer sur la paroi de la cellule hôte et d'injecter leur ADN. Celui-ci grâce aux extrémités *cos* se circularise et se réplique comme un plasmide mais aucune des fonctions phagiques ne peut être exprimée donc les cellules survivent et forment des clones.

II.3.B Avantages et désavantages des cosmides :

Avantages:

- La taille des fragments insérables peut atteindre 50 kb.
- Leur incorporation dans les bactéries (transformation) est plus efficace que pour les plasmides. Les cellules transformées sont sélectionnées sur un milieu contenant de l'antibiotique (ampicilline).
- Les cosmides sont plus stables que les plasmides.

Désavantages:

- Obligation d'empaqueter l'ADN.

II.4 Autres types de vecteurs :

D'autres vecteurs existent avec possibilité de cloner des fragments d'ADN de taille élevée (au moins jusqu'à 500 kb).

II.4.1 Les BAC : (Bacterial Artificial Chromosome)

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 réplication et *rep E* qui sont nécessaires à la réplication. Des régulateurs de la réplication (*sop B*, *sop A*) pour maintenir un faible nombre de copies, sa taille est de 6.5 kb.

Bibliographie :

Vincent Ecochard, Didier Fournier, Laurence Nieto, Laurent Paquereau. (2011). Techniques et Stratégies en Biologie Moléculaire. Première partie. Université Paul Sabatier, Toulouse, 201p.

Clark D. (2005). Molecular Biology: Understanding the Genetic Revolution. Southern Illinois University. USA Elsevier Academic Press.784p.

Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003).Molecular Cell Biology. 6th Ed. 937p.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 151-155p.

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES.

UE2. Ellipses 21-46p.

D. Tagu, C. Moussard, (2003). Principes des techniques de biologie moléculaire, , 2e éd., INRA-Quæ (ISBN 2-7380-1067-9).

Passarge E. (2007).Color Atlas of Genetics. 3rdEd. Flexibook. Thieme Stuttgart, NY. USA. 486p.

Aouf Abdelhakim, (2016). Cours de Biologie moléculaire et génie génétique. Université Ferhat-Abbas Setif 1, 133p.

Primose S. B., R. M. Twyman and R.W. Old. (2002) Principles of Gene Manipulation. 6th Ed. Blackwell Scientific Publishing, Oxford. UK. 319p.

Strachan Tom, Read Andrew. (2004). Génétique Moléculaire Humaine. Hybridation des acides nucléiques : principes et applications. 4 e édition ; Médecine Sciences publications. Lavoisier 16p.

Chapitre 2 : Les enzymes de restriction

1. Généralités

Découvertes à partir de 1973. Les enzymes de restriction sont capables de reconnaître spécifiquement une courte séquence, de 4 à 10pb, et de cliver l'ADN au site reconnu. Ils permettent de fragmenter l'ADN en segments de taille réduite, ou de le couper à tel ou tel site désiré. Certains enzymes coupent le site en son milieu et produisent deux fragments dont les extrémités sont franches. Cependant, la plupart réalisent une coupure dissymétrique : on parle dans ce cas d'extrémités cohésives (chaque fragment possède une chaîne qui dépasse l'autre de quelques bases). Plusieurs centaines de ces enzymes ont été caractérisés : ils reconnaissent une grande variété de sites de coupure.

Les enzymes de restriction sont utilisés pour établir une carte de restriction de toute molécule d'ADN que l'on souhaite caractériser. Cela consiste à déterminer l'ordre des sites de restriction le long de cette molécule, qui vont produire, après "digestion enzymatique" de cette molécule, des fragments de tailles différentes dont la taille pourra être définie par électrophorèse.

Les enzymes de restriction appartiennent à la classe des endonucléases, c'est-à-dire des enzymes capables de cliver les liaisons phosphodiester entre deux nucléotides à l'intérieur d'un acide nucléique. Les endonucléases se différencient des exonucléases qui dégradent la molécule d'ADN à partir de l'une de ses extrémités (3' ou 5').

2. Séquences d'ADN reconnues par les enzymes de restriction

Les séquences de nucléotides reconnues par les enzymes de restriction sont habituellement des séquences dites palindromiques. Les séquences palindromiques sont des séquences où la succession des nucléotides lue dans le sens 5' à 3' (gauche-droite) pour le premier brin est identique à la séquence lue dans le sens droite-gauche pour le second brin (sens 5' à 3'). Ces séquences palindromiques sont le plus souvent constituées de 4, 5 ou 6 paires de bases. Il faut remarquer que dans les ADN, on rencontre statistiquement des séquences reconnues par des enzymes de restriction. Ainsi, la séquence **GATC** reconnue par l'enzyme **Mbo I** est présente avec une fréquence statistique de 1 / 256 paires de bases ($1/4^4$). En effet, la fréquence de coupure d'un ADN par une enzyme de restriction donnée peut être approchée statistiquement en considérant le nombre de nucléotides de la séquence spécifique reconnue par l'enzyme.

Ainsi, par exemple, dans la séquence de six nucléotides: **GGATCC** reconnue par l'enzyme **Bam HI**, on aura donc une fréquence de coupure statistique de $1 / 4^6$, soit 1 coupure tous les 4096 nucléotides.

3. Origine des enzymes de restriction

Les enzymes de restriction sont extraites de micro-organismes, le plus souvent des bactéries. Les bactéries peuvent être parasitées par des virus à ADN. Les bactéries fabriquent des enzymes de restriction qui sont capables de cliver les ADN étrangers. Pour éviter une auto-destruction de leur propre ADN, elles se protègent contre leurs propres enzymes de restriction par une modification des sites de restriction correspondants.

4. Nomenclature des enzymes de restriction

Les enzymes de restriction présentent une nomenclature bien précise. Leur nom comporte plusieurs lettres (3 ou 4). La première lettre de dénomination de l'enzyme est écrite en majuscule, elle correspond au **genre de la bactérie** d'où a été extraite l'enzyme. La seconde lettre et la troisième lettre (en minuscules) correspondent à **l'espèce de la bactérie** d'où l'enzyme est extraite. On peut avoir une quatrième lettre écrite en majuscule correspondant à **la souche bactérienne**. Enfin pour terminer, un chiffre romain indique **l'ordre** de caractérisation de ces enzymes.

Exemples:

Eco RI Extraite d'*Escherichia coli* RYB site reconnu: **G / AATTC**

Sma I Extraite de *Serratia marcescens* site reconnu: **CCC / GGG**

Pst I Extraite de *Providencia stuartii* site reconnu: **CTGCA / G**

4.1 Notion d'isoschizomères

Des enzymes de restriction différentes peuvent reconnaître des mêmes sites spécifiques, on les appelle isoschizomères. Les isoschizomères fournissent souvent après clivage enzymatique des fragments dont les extrémités sont différentes.

Exemple :

Soit la séquence suivante: GGTACC, cette séquence est coupée par l'enzyme Kpn I et l'enzyme Acc65 I:

Kpn I:

5'-G-G-T-A-C/C-3'

3'-C/C-A-T-G-G-5'

Acc65 I:

5'-G/G-T-A-C-C-3'

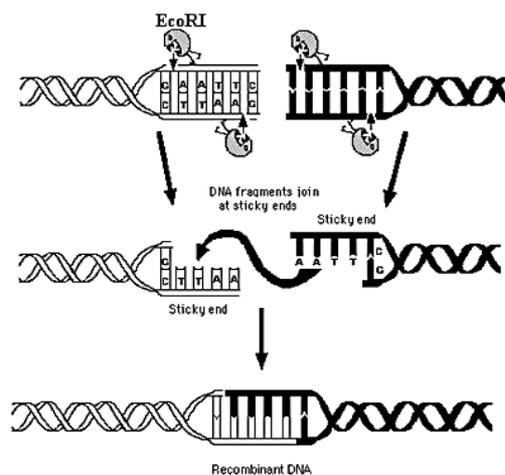
3'-C-C-A-T-G/G-5'

Ces enzymes sont des isoschizomères, elles reconnaissent en effet la même séquence nucléotidique GGTACC. On voit tout de suite que les extrémités des fragments obtenus diffèrent.

5. Types de coupures réalisées par les enzymes de restriction

Les enzymes de restriction peuvent donner deux types de coupures: la coupure à bouts francs et la coupure à bouts collants (figure 1).

- **La coupure à bouts francs** aboutit à une coupure au milieu de la séquence palindromique.
- **La coupure à bouts collants** (ou à extrémités adhésives) correspond à une coupure qui se fait de part et d'autre du centre de symétrie.



**Restriction Enzyme
Action of EcoRI**

Figure 1 : coupure à bouts francs et coupure à bouts collants réalisées par les enzymes de restriction.

Les sites de restriction sont repérés dans l'ADN par l'enzyme de restriction qui coupe l'ADN en principe autant que fois qu'il y a de sites de restriction. Ceci est valable pour une enzyme de restriction donnée, pour une autre enzyme, la coupure se fera en une position différente sur l'ADN. On voit tout de suite les possibilités considérables de ce type d'outils enzymatiques. L'ADN est découpé en fragments variables et ceci aussi bien l'ADN circulaire des bactéries ou des plasmides que l'ADN linéaire.

6. Les types d'enzymes de restriction

La plupart des enzymes de restriction utilisées au laboratoire présentent un site de reconnaissance (souvent palindromique) et un site de coupure identique ou proche du site de reconnaissance, ce sont **des enzymes de type II**. Certaines bactéries possèdent d'autres types d'enzymes de restriction.

Les enzymes de type I reconnaissent des séquences sur l'ADN sans aucune symétrie. Ces enzymes coupent non pas au niveau de la séquence de reconnaissance mais 1000 à 5000 paires de bases plus loin.

Les enzymes de type III présentent également un site de reconnaissance mais coupent une vingtaine de nucléotides plus loin.

Tableau 1: exemples d'enzymes de restriction de type II

Enzyme	Source	Séquence reconnue	Coupure	Extrémités (après coupure)
Eco RI	<i>Escherichia coli</i>	5'GAATTC 3'CTTAAG	5'---G AATTC---3' 3'---CTTAA G---5'	Cohésives
BamHI	<i>Bacillus amyloliquefaciens</i>	5'GGATCC 3'CCTAGG	5'---G GATCC---3' 3'---CCTAG G---5'	Cohésives
HindIII	<i>Haemophilus influenzae</i>	5'AAGCTT 3'TTCGAA	5'---A AGCTT---3' 3'---TTCGA A---5'	Cohésives
MstII	<i>Microcoleus species</i>	5'CCTNAGG 3'GGAMTCC	5'---CC TNAGG---3' 3'---GGAMT CC---5'	Cohésives
TaqI	<i>Thermus aquaticus</i>	5'TCGA 3'AGCT	5'---T CGA---3' 3'---AGC T---5'	Cohésives

NotI	<i>Nocardia otitidis</i>	5'GCGGCCGC 3'CGCCGGCG	5'--GC GGCCGC--3' 3'--CGCCGG CG--5'	Cohésives
HinfI	<i>Haemophilus influenzae</i>	5'GANTC 3'CTNAG		
AluI	<i>Arthrobacter luteus</i>	5'AGCT 3'TCGA	5'---AG CT---3' 3'---TC GA---5'	Franches
BgIII	<i>Bacillus globigii</i>	5'AGATCT 3'TCTAGA	5'---A GATCT---3' 3'---TCTAG A---5'	Cohésives
HaeIII	<i>Haemophilus aegyptius</i>	5'GGCC 3'CCGG	5'---GG CC---3' 3'---CC GG---5'	Franches
HhaI	<i>Haemophilus haemolyticus</i>	5'GCGC 3'CGCG	5'---GCG C---3' 3'---C GCG---5'	Cohésives
PstI	<i>Providencia stuartii</i>	5'CTGCAG 3'GACGTC	5'---CTGCA G---3' 3'---G ACGTC---5'	Cohésives
SmaI	<i>Serratia marcescens</i>	5'CCCGGG 3'GGGCCC	5'---CCC GGG---3' 3'---GGG CCC---5'	Franches

Le N et le M dans la séquence de MstII peuvent être remplacés par une des 4 bases et son complément. La séquence de restriction du MstII est un exemple de palindrome imparfait puisqu'il comporte un nombre impair de paires de bases.

7. Méthylation des sites de restriction et inactivation des enzymes de restriction

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 l'ADN bactérien présente des sites de restriction susceptibles d'être repérés par les enzymes de restriction que possède la bactérie. Pour éviter une auto-destruction, les enzymes de modification de l'ADN bactérien interviennent. Ces enzymes de modification sont des **méthylases** bactériennes (ou enzymes de méthylation). La méthylation de la cytosine (sur le carbone 5) ou de l'adénine (sur l'azote 6) appartenant à des sites de restriction aboutit à une inactivation de l'enzyme de restriction correspondante (figure 2). Cette méthylation peut se réaliser sur une base ou sur plusieurs bases appartenant au site de restriction. Les méthylases bactériennes sont très spécifiques.

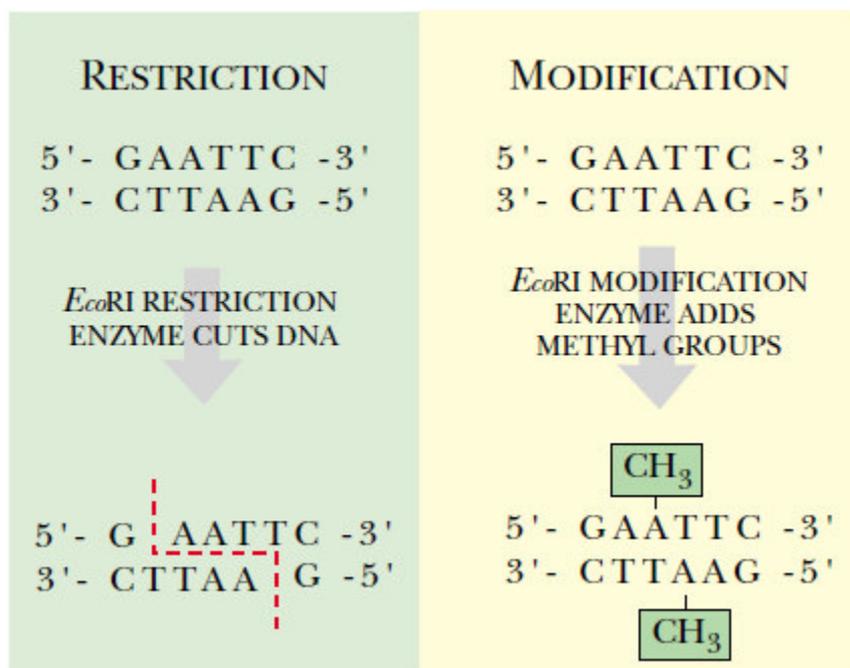


Figure 2: Systèmes de restriction et de méthylation (aouf, 2016)

8. Utilisations des enzymes de restriction

Les utilisations des enzymes de restriction sont très nombreuses en biologie moléculaire.

- Par exemple, elles permettent de fractionner l'ADN en multiples fragments susceptibles d'être séparés par les techniques d'électrophorèse.
- Les enzymes de restriction peuvent être utilisées pour préparer un fragment d'ADN d'un gène donné (insert) à être inséré dans un vecteur comme un plasmide.

- Les enzymes de restriction sont utilisées couramment pour rechercher des mutations dans le génome.

- Les enzymes de restriction sont utilisées pour rechercher dans l'ADN des cellules eucaryotes les méthylations de bases. Ces méthylations ont une signification complètement différente des méthylations de bases observées chez les procaryotes. Elles sont en relation directe avec des modifications de l'expression des gènes des eucaryotes. La méthylation provoque le verrouillage de l'expression de tel ou tel gène dans un tissu. Les méthylations dans le génome des eucaryotes concerne les cytosines impliquées dans les doublets dinucléotidiques CG. La recherche de ces doublets et de la présence ou non d'une méthylation sur les cytosines est réalisée à l'aide de deux enzymes de restriction, une enzyme insensible à la méthylation des cytosines et une autre enzyme sensible à la méthylation des cytosines.

9. Notion d'enzymes compatibles

Deux enzymes de restriction sont dites compatibles quand elles génèrent après digestion des fractions aux extrémités cohésives complémentaires. Ces fragments peuvent être facilement ligaturés.

Exemple: Les enzymes Bam HI (G / GATCC) et Mbo I (/ GATC):

Après action de Bam HI:

Premier fragment avant clivage par Bam HI:	Après clivage par Bam HI:
5'-G-G-A-T-C-C-3'	5'-G-3' (1)
3'-C-C-T-A-G-G-5'	3'-C-C-T-A-G-5'
	+
	5'-G-A-T-C-C-3' (2)
	3'-G-5'

Après action de Mbo I:

Second fragment avant clivage par Mbo I:	Après clivage par Mbo I
5'-A-G-A-T-C-A-G-C-3'	5'-A-3' (3)
3'-T-C-T-A-G-T-C-G-5'	3'-T-C-T-A-G-5'
	+
	5'-G-A-T-C-A-G-C-3' (4)
	3'-T-C-G-5'

Ligatures possibles : 1+4 et 2+3.

Bibliographie :

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 147-150p

Passarge E. (2007). Color Atlas of Genetics. 3rd Ed. Flexibook. Thieme Stuttgart, NY. USA. 486p.

Streips N. U. & R. Yasbin. (2002). Modern Microbial Genetics. 2nd Ed. Wiley-Liss, Inc. New York. USA; 603p.

Université de Pierre & Marie Curie, Genetique ; fragmentation de l'ADN :

http://www.edu.upmc.fr/sdv/masselot_05001/genes_et_genomes/fragmentation.html

Roberts RJ, Belfort M, Bestor T, Bhagwat AS, et al (2003). « SURVEY AND SUMMARY: A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes », Nucleic Acids Res. 31 : (7), 1805–12.

Smith HO, Nathans D, (1973) « Letter: A suggested nomenclature for bacterial host modification and restriction systems and their enzymes », J. Mol. Biol., 81 : (3), 419–23.

Le logiciel REBASE disponible sur le site : www.rebase.neb.com

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES. UE2. Ellipses 35-43p.

COULIBALY Fountouin Hamidou, N'NAN Oulo. Cours Magistral Génétique UE gène 2208, Master 1 LMD. 38p.

Aouf Abdelhakim, (2016). Cours de Biologie moléculaire et génie génétique. Université Ferhat-Abbas Setif 1, 133p.

Chapitre 3: Technologie de l' ADN recombinant

1. Définition de la protéine recombinante :

Une **protéine recombinante** est une protéine produite par une cellule dont le matériel génétique a été modifié par recombinaison génétique.

Un gène codant une protéine d'intérêt est introduit dans le génome de l'espèce productrice (bactéries, cellules mammifères en culture, animaux transgéniques). Les protéines recombinantes peuvent être purifiées et utilisées à des fins thérapeutiques, industrielles ou bien encore dans les activités de recherche.

Les protéines recombinantes sont ainsi qualifiées dans la mesure où elles sont produites par des cellules dont l'ADN a été modifié par recombinaison génétique. Au sens large, un système de production adapté à la fabrication d'une protéine recombinante donnée, est un processus biotechnologique qui s'appuie principalement sur :

- l'emploi d'un vecteur d'expression (en général un plasmide ou un virus -pour les vecteurs eucaryotes), jouant le rôle de transporteur génétique du gène d'intérêt codant pour la protéine recherchée
- l'utilisation d'une cellule hôte, chargée d'exécuter les instructions fournies par le gène d'intérêt qui lui est inséré, dans l'objectif de synthétiser la protéine recherchée
- une phase de production proprement dite permettant de fabriquer les volumes de protéines souhaités
- enfin, une séparation et extraction de la protéine du milieu de culture, suivie par une purification de celle-ci.

Au sens restreint, un système de production de protéines recombinantes est caractérisé par un couple constitué d'un vecteur d'expression et d'un hôte (cellule ou organisme).

2. ADNc et vecteur d'expression :

L'organisme donneur exprime la protéine d'intérêt. L'ARNm codant cette protéine est isolé de cet organisme. L'ARNm permet la synthèse de l'ADNc qui sera utilisé pour le clonage. En effet, l'ADN génomique eucaryote n'est pas utilisable dans un système bactérien car les bactéries ne possèdent pas la machinerie d'épissage des ARNm eucaryotes.

C'est pourquoi la technique RT-PCR est souvent utilisée pour l'obtention de l'ADNc codant le gène d'intérêt. À la suite de la transcription inverse des ARNm totaux en ADNc (simple brin) grâce à des amorces poly T spécifique des extrémités poly A des ARNm, l'ADNc correspondant à l'ARNm codant la protéine d'intérêt est amplifiée par PCR en utilisant des amorces spécifiques auxquels sont ajoutés des sites de restrictions qui seront ensuite utilisés pour le clonage.

3. Choix de l'hôte :

Une vaste gamme de systèmes de production de protéines recombinantes -c'est-à-dire, de couples vecteurs-hôtes- est aujourd'hui disponible, chacun d'eux présentant des avantages et des inconvénients. Les hôtes les plus utilisés à l'heure actuelle sont incontestablement la **bactérie *Escherichia coli*, la levure *Saccharomyces cerevisiae*** et les **cellules CHO extraites des ovaires de hamster**.

3.A *E. coli* : elle fut et reste encore le premier hôte utilisé pour la fabrication de protéines recombinantes.

- exemples de protéines recombinantes produites : hormone de croissance humaine, insuline, chymosine, interféron, interleukine-2...

3.A.1 principaux avantages :

- sa génétique est très bien connue. De nombreux vecteurs plasmidiques ont été construits et sont donc disponibles afin d'insérer et d'exprimer un gène étranger au sein de la bactérie.
- Elle est par ailleurs facile à utiliser, se prête très bien à la culture de masse en fermenteur. Enfin, les taux d'expression obtenus sont élevés, c'est-à-dire qu'elle permet de produire des quantités appréciables de protéines (jusqu'à plusieurs grammes par litre).

3.A.2 principaux inconvénients :

- sécrétant mal les protéines, il est souvent nécessaire de "casser" la bactérie afin de récupérer la protéine (ce qui induit des problèmes de purification, ou de solubilisation et de renaturation..., quelquefois au détriment des rendements).

- Autre inconvénient majeur : *E. coli* n'effectue pas les modifications post-traductionnelles des protéines (en particulier la glycosylation, la carboxylation, etc.), qui constituent souvent une condition principale d'activité de la protéine.
- Enfin *E. coli* étant une entérobactérie, il est donc nécessaire de s'assurer de l'absence d'endotoxines dans les protéines purifiées.

3.B *Saccharomyces cerevisiae* : il s'agit de la levure de boulanger, utilisée depuis des millénaires dans l'alimentation humaine.

- exemples de protéines recombinantes produites d'intérêts thérapeutiques : antigène de surface du virus de l'hépatite B, insuline, hirudine...

3.B.1 principaux avantages :

- son matériel génétique est simple et elle ne présente aucune toxicité. Bons vecteurs d'expression disponibles aujourd'hui.
- Les taux d'expression des protéines sont relativement bons (de l'ordre d'environ 100mg/L de culture).
- La levure est capable de fabriquer des protéines complexes et de réaliser des modifications post-traductionnelles (glycosylations simples, carboxylations, acylations...).

3.B.2 principaux inconvénients :

- les protéines synthétisées sont souvent obtenues à l'intérieur du cytoplasme et nécessitent de casser la cellule afin de les récupérer.
- La sécrétion est possible, mais en général au détriment des rendements : elle fonctionne bien pour des petits polypeptides tels que l'insuline, mais beaucoup moins bien pour de grandes protéines.

3.C Les cellules CHO (Chinese Hamster Ovary) :

- exemples de protéines recombinantes produites : antigène du virus de l'hépatite B, hormone de croissance humaine, cytokines, érythropoïétine, facteurs de coagulation...

3.C.1 principaux avantages :

- méthode déjà employée dans la production d'un vaccin contre l'hépatite B, les cellules CHO se prêtent très bien à la culture de masse en bioréacteur.
- Un avantage discriminant de ces cellules réside dans leur capacité à synthétiser des protéines complexes de poids moléculaire élevé. Certains vecteurs d'expression (BPV -Bovine Papilloma Virus-, SV40...) permettent d'introduire et de faire exprimer des gènes humains.

3.C.2 principaux inconvénients :

- rendements faibles (de l'ordre de 10 milligrammes par litre au maximum.
- En outre les cellules CHO s'avèrent fragiles et leur culture plus onéreuse.

D'autres systèmes de production, c'est-à-dire d'autres couples hôtes-vecteurs, se développent également aujourd'hui. Les hôtes sont principalement :

□ **D'autres bactéries :**

□ *Bacillus subtilis*, *Streptomyces*... possèdent des capacités de sécrétion supérieures à *E. coli*. Cependant leur génétique est moins connue, et le niveau de production de protéines est inférieur à celui obtenu avec *E. coli*.

□ d'autres **levures et champignons filamenteux** : *Pichia pastoris*, *Kluyveromyces lactis*, *Hanseluna polymorpha*, *Yarrowia lipolytica*, *Aspergillus niger*.

□ **des cellules d'insectes** (*Spodoptera frugiperda*), utilisant des vecteurs d'expression développés à partir du Baculovirus. Ces cellules sont capables de sécréter la protéine recombinante et d'effectuer les opérations post-traductionnelles. Par ailleurs il est possible également d'utiliser **des larves de vers à soie vivants**.

4. 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 des protéines 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.

4.1 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?

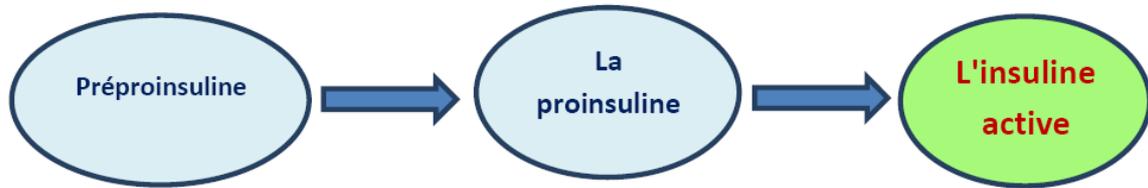
Dans les cellules de mammifères la quantité de protéine de grand intérêt pharmaceutique comme l'insuline est très faible. Ce qui rend leur extraction et purification extrêmement couteuse (onéreuse). De plus, même si ces protéines peuvent être produites par des cultures cellulaires; c'est une vois beaucoup 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. coli* de 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 obtenue à 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.

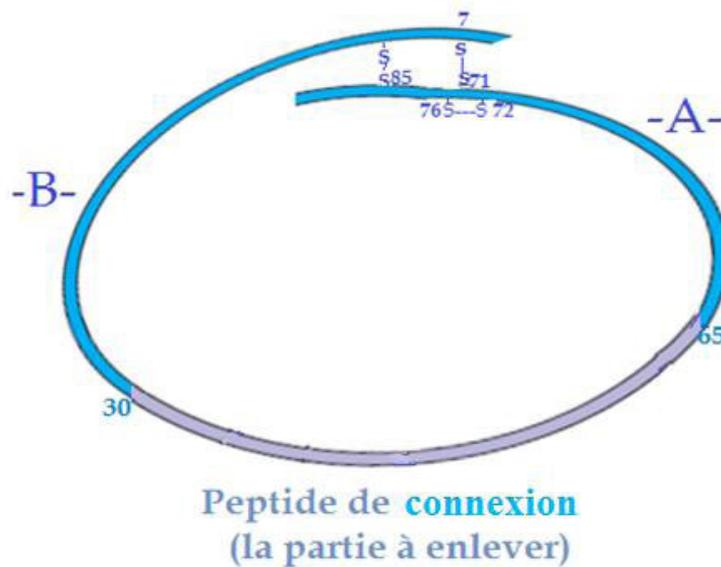


Figure 1: La proinsuline.

- On peut utiliser *Saccharomyces cerevisiae* pour produire cette hormone (Aouf, 2016).

Bibliographie :

Vincent Ecochard, Didier Fournier, Laurence Nieto, Laurent Paquereau. (2011). Techniques et Stratégies en Biologie Moléculaire. Première partie. Université Paul Sabatier, Toulouse, 201p.

Clark D. (2005). Molecular Biology: Understanding the Genetic Revolution. Southern Illinois University. USA Elsevier Academic Press.784p.

Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell. (2003).Molecular Cell Biology. 6th Ed. 937p.

Beaumont Simon, (2007). Biologie moléculaire PCEM1, Dunod, Paris 279-286p.

Perrin-Schmitt Fabienne, (2011). Biotechnologies et applications (génie génétique), PACES. UE2. Ellipses 21-46p.

D. Tagu, C. Moussard, (2003). Principes des techniques de biologie moléculaire, , 2e éd., INRA-Quæ (ISBN 2-7380-1067-9).

Passarge E. (2007).Color Atlas of Genetics. 3rdEd. Flexibook. Thieme Stuttgart, NY. USA. 486p.

Medraoui Leila. (2015). Cours Biologie Moléculaire ; Initiation aux techniques usuelles de biologie moléculaire. Faculté des sciences. Université Mohamad V Rabat 132p.

Primose S. B., R. M. Twyman and R.W. Old. (2002) Principles of Gene Manipulation. 6th Ed. Blackwell Scientific Publishing, Oxford. UK. 319p.

Strachan Tom, Read Andrew. (2004). Génétique Moléculaire Humaine. Hybridation des acides nucléiques : principes et applications. 4 e édition ; Médecine Sciences publications. Lavoisier 16p.

Aouf Abdelhakim, (2016). Cours de Biologie moléculaire et génie génétique. Université Ferhat-Abbas Setif 1, 133p.

SOMMAIRE

Chapitre I : l'expression de l'information génétique	1
Partie 1 : La transcription.....	1
I) Définition de la transcription :	1
II) Transcription de l'ADN procaryote	3
II.1 Étapes de la transcription chez les procaryotes.....	5
II.1.A Initiation :	6
II.1.B Elongation	9
II.1.C Terminaison.....	10
III) Transcription de l'ADN eucaryote	16
III.1 Transcription par l'ARN polymérase I.....	16
III.1.A Fonctionnement de la transcription	17
III.1.B Maturation du précurseur 45 S et assemblage.....	18
III.2 Transcription par l'ARN polymérase III.....	18
III.3 Transcription par l'ARN polymérase II	19
III.3.1 Transcription d'un ARN messager :	19
III.3.1.1. L'initiation de la transcription :	19
III.3.1.1.a. Site promoteur	19
III.3.1.1.b. Le complexe d'initiation.....	20
III.3.1.1.c. Intervention de facteurs spécifiques de la transcription.....	22
III.3.1.2. L'élongation :	23
III.3.1.3. La terminaison :	25
III.4 Inhibiteurs de transcription.....	25
Partie II : la traduction	26
1. Définition de la traduction.....	27
1.A. Le code génétique.....	27
1.B. Les acteurs de la traduction.....	28
1.B.1 Les ribosomes :	28
1.B.1.A Topographie schématique du ribosome bactérien.....	29
1.B.1.B Topographie schématique du ribosome chez les eucaryote.....	29
1.B.2 Les ARNt.....	30
1.B.3 L'ARNm	32
2. Les différentes étapes de la traduction procaryote	33
2.1 Initiation.....	33
2.2 Elongation.....	36

2.2.a Réaction de couplage	36
2.2.b Formation de la liaison peptidique et libération du premier ARNt	37
2.2.c Translocation	37
2.2.d Bilan énergétique au cours de l'élongation	37
2.3 Terminaison.....	38
3. Les spécificités de la traduction eucaryote.....	38
4. Les antibiotiques inhibiteurs de la traduction chez les procaryotes.....	43
Chapitre 2 : régulation de l'expression génétique.....	44
I : régulation transcriptionnelle.....	44
1. Définition des modifications transcriptionnelles	44
2. Modifications de l'ARN après transcription chez les procaryotes.....	45
3. régulation transcriptionnelle chez les eucaryotes :.....	45
3.1 L'addition d'une coiffe en 5'(ou capping) :.....	45
3.2 L'addition d'une queue poly A en 3'.....	47
3.3 L'excision-épissage	47
3.4 L'épissage alternatif.....	52
4. Structure des facteurs protéiques de la transcription.....	52
4.1. Structure des activateurs	53
4.1.1 Protéines à Leu Zipper	53
4.1.2 Protéines à doigts de zinc.....	53
4.1.3 Protéines à homeodomaines.....	54
II : La régulation de l'opéron lactose (régulation de la synthèse des protéines procaryotiques) .	57
1. Qu'est-ce-qu'un opéron ?	57
2. Le métabolisme du lactose :.....	57
3. Régulation négative de la transcription.....	59
4. Levée de l'inhibition de la transcription.....	60
5. Répression catabolique, régulation positive.....	61
III : régulation traductionnelle	66
1) L'ARNm dans le cytoplasme :.....	67
2) Recrutement du ribosome :.....	68
<i>2.a) Démarrage de la traduction eucaryote (initiation)</i>	<i>68</i>
2.b) Élongation :.....	69
2.c) Terminaison de la traduction :.....	72
2.d) Contrôle qualité des ARNm :.....	73
2.e) Dégradation des ARN messagers :	73

3) Modifications post-traductionnelles :	75
Chapitre 3 : techniques de base de biologie moléculaire :	76
Préparation des acides nucléiques (extraction et purification)	76
1) Préparation d'ADN génomique :	76
1.1. Préparation d'extraits bruts acellulaires	76
1.1.A Lyse mécanique	76
1.1.B Lyse chimique	77
1.1.C Elimination des protéines	78
1.1.D Réactifs divers ajoutés au tampon d'extraction.....	79
1.1.E Elimination des ARN lors de l'extraction de l'ADN	80
1.2. Purification par extraction phénol-chloroforme	80
1.2.A Principe de la méthode d'extraction.....	80
2) Préparation d'ADN plasmidique :	82
3) Contrôle de l'extraction :	83
Séparation des acides nucléiques : (électrophorèse sur gel d'agarose) :	85
1. définition :	85
2. Principe général:.....	85
3. Applications de l'électrophorèse :	85
4. Facteurs affectant la migration :	86
5. Gel d'agarose :	86
5.a. Agarose :	86
5.b Tampons :	87
6. Révélation de l'ADN :	87
6.1 Limites de résolution :	88
6.2 Analyse du gel :	88
6.2.a Détermination de la taille des fragments	89
7. L'électrophorèse en champ pulsé.....	90
Détection, caractérisation et identification des acides nucléique.....	92
1. Marquage et suivi des acides nucléiques.....	92
1.1. Marquage radioactif.....	92
1.1.1 Les sondes doubles brins.....	92
1.1.1.B Marquage par translation de coupure (Nick Translation).....	93
1.1.2 Les sondes simples brin	94
1.1.2.A Marquage en 5' des ARN	94
1.1.2.B Marquage en 5' des acides nucléiques (ADN ou ARN).....	95

1.1.2.C Marquage en 3'	95
1.2 Marquages froids	97
1.2.1 La détection des sondes froides	97
1.2.2 L'amplification du signal	98
1.2.2.A Méthode en deux couches	98
1.2.2.B Méthodes en trois couches	99
1.2.2.C Méthode utilisant un complexe entre l'enzyme de détection et un anticorps	99
1.2.2.E Méthodes utilisant les complexes avidine-biotine. (AB Complex)	100
1.2.2.F CSA (Catalysed Signal Amplification)	101
Hybridation moléculaire :	103
I. Définition	103
I.1 Les sondes nucléotidiques :	104
II. Types d'hybridation :	105
II.1 Hybridation d'ADN sur support solide :	106
II.1.1 Southern Blot :	106
II.1.2 Northern Blot :	107
II.1.3 Hybridation in situ (HIS) :	110
II.1.4 Hybridation sur colonie, criblage d'une banque de plasmide	110
Amplification in vitro des acides nucléiques PCR et RT-PCR :	113
1. Historique :	113
2. Principe :	113
3. Les amorces :	114
4. Les températures :	115
5. L'ADN matriciel	116
6. L'ADN polymérase	116
7. Le tampon	116
8. Aspects quantitatifs de la réaction PCR	116
9. RT-PCR :	117
10. Utilisations des produits PCR	118
Le séquençage de l'ADN :	120
1. Historique	120
1.1 Définition de séquençage :	120
1.2 Les différents acteurs du séquençage	120
1.3 Méthode de Sanger :	121
1.4 Principe du séquençage par la méthode de Sanger	121

PARTIE 2 : génie génétique :	125
Chapitre 1 : clonage in vivo :.....	125
Introduction :	125
I. Clonage :	125
I.1.1 Le clonage nucléique.....	125
I.1.2 Les vecteurs :	126
<i>I.1.2.A ADN recombinant.....</i>	<i>127</i>
<i>II. Les éléments nécessaires au clonage</i>	<i>127</i>
II.1 Les vecteurs de clonage: plasmides ; cosmides ; phages ; BAC ; PAC ; YAC :	127
II.1.A Les plasmides.....	127
<i>II.1.A.1 Propriétés générales des plasmides.....</i>	<i>127</i>
<i>II.1.A.2 Préparation des plasmides</i>	<i>128</i>
II.1.A.3 Préparation des plasmides pour la recombinaison, constitution de l'hybride et transformation bactérienne.....	129
II.1.A.4 Sélection des bactéries transformées par des plasmides :	129
II.1.A.5 Un exemple de plasmide: le puc19.....	130
<i>II.1.A.6 Avantages et désavantages des plasmides.....</i>	<i>132</i>
II.2 Les phages :	133
II.2.A Propriétés générales des phages	133
II.2.A.1 Bactériophage λ	133
II.2.B λ Comme vecteur	136
II.2.B.1 Principales étapes de la construction d'un vecteur phagique	136
II.2.C Avantages et désavantages des phages	137
II.3 Les cosmides :	137
II.3.A Propriétés générales des cosmides :	137
II.4 Autres types de vecteurs.....	138
3. Choix de l'hôte :	150