

République Algérienne Démocratique et Populaire
Ministère De L'enseignement Supérieur et de la Recherche Scientifique

Université Echahid Hamma Lakhdar d'El-Oued

Faculté des Sciences de la Nature et de la Vie

Département de Biologie Cellulaire et Moléculaire



Ingénierie moléculaire

Cours de 1^{ère} Année Master Biochimie Appliquée

Préparé par : Dr. MEDIL A Ifriqya

MCA à l'Université d'El-Oued



2020/2021

TABLE DES MATIERES

Chapitre I. Génétique de la levure, Animaux, et plantes transgéniques	1
1. Généralités	1
2. Les levures	1
3. Les animaux transgéniques	3
4. Les plantes transgéniques	4
Chapitre II. Protéines recombinantes en recherche fondamentale	5
1. Les vecteurs d'expression et les cellules hôtes	5
2. Expression en système procaryote: <i>Escherichia coli</i>	6
2.1. Stratégie générale	7
2.2. Avantages et défauts du système	8
2.3. Construction des vecteurs d'expression	9
a-Origin de répllication	10
b-Les promoteurs	11
c- Séquences responsables de la traduction	14
d-Séquences responsables de l'arrêt de la transcription	14
e-Étiquette de détection/purification (TAG)	14
2.4. Choix de la localisation subcellulaire de la protéine exprimée	17
3. Production chez les levures	19
3.1. <i>Saccharomyces cerevisiae</i>	19
3.1.1. Avantages et défauts du système	19
3.1.2. Vecteurs d'expression	20
3.1.3. Système de sécrétion	22
3.1.4. Modifications post-traductionnelles	23
3.2. <i>Pichia pastoris</i>	24
3.2.1. Avantages et défaut de systeme	24
3.2.2. Étapes d'expression d'une protéine chez <i>Pichia pastoris</i>	25
3.2.3. Vecteurs d'expression	26
3.2.4. Mode d'intégration du gène d'intérêt :	26
3.2.5. Système de sécrétion	28
4. Production en cellules d'insectes	29
4.1. Avantages et défauts de système	29

4.2. Cycle de multiplication d'un baculovirus	29
4.3. Utilisation des baculovirus comme vecteurs d'expression de gènes étrangers	31
4.4. Cellules d'insecte ayant intégré le gène d'intérêt dans leur génome	32
5. Production en cellules de mammifère	33
5.1. Expression transitoire	34
5.2. Pools stables	36
6. Préparation de protéines recombinantes à partir des animaux transgéniques	37
6.1. Avantages et inconvénients du système	37
6.2. Systèmes de production de protéines recombinantes par les animaux transgéniques.	37
6.3. Les techniques de transgénèse	40
A. Micro-injection dans le pronucléus	40
B. Transfection de cellule embryonnaire souches (ES)	42
7. La production de protéines recombinante dans les plantes	43
7.1. Les différents systèmes végétaux de production de protéines recombinantes	43
7.2. Méthodes de transformation	44
7.2.1. à l'aide d' <i>Agrobacterium tumefaciens</i>	45
7.2.2. à l'aide de la biolistique	46
7.2.3. à l'aide de l'électroporation	46
7.3. Les avantages du système plante	47
7.4. Les limitations du système plante et les stratégies de contournement développées	48
A. La glycosylation chez les plantes	48
B. Adressage des protéines vers la voie sécrétoire	48
C. L'activité protéolytique et la stabilité de protéines recombinantes	49
D. L'optimisation des codons	49
E. Le « gene silencing » et les supresseurs	49
8. Bilan	51
Chapitre III. Modulation de l'expression génique ou de l'activité protéique	52
1. Le Knock-out ou l'invalidation d'un gène par recombinaison homologue	52
1.1. Généralités	52
1.2. Principe de la technology Knock Out	53

1.3. La recombinaison homologue : un mécanisme à l'origine de la technologie Knockout	53
1.4. La sélection des événements de recombinaison.	54
1.5. Les différentes constructions utilisées	54
1.6. Exemple de l'invalidation du gène Bmp7	56
2. Le système Cre/Lox	60
2.1. Fonctionnement de système	60
2.2. Exemples d'application	62
2.2.1. Mutations «propres»	62
2.2.2. Le système Cre/loxP et la mutagenèse conditionnelle	63
3. Les oligonucléotides triples hélices	66
3.1. Fonctionnalité de système	66
3.2. Avantage et limitation des Applications potentielles des oligonucléotides triple hélice en thérapeutique	67
4. Stratégie ribozyme.	68
4.1. Classification des ribozymes	68
4.2. Domaines d'application	69
5. L'approche anti-sens	70
5.1. Les mécanismes d'action des oligonucléotides antisens	70
5.1.1. Mécanisme impliquant la ribonucléase	70
5.1.2. Inhibition par blocage physique de l'ARN messenger	71
5.1.3. Modification du métabolisme des ARN messagers	71
5.1.4. Inhibition de l'épissage des ARN pré-messagers	71
6. Modifications chimiques des oligonucleotides (les phosphoramidates, les morpholinos, PNA, LNA)	73
7. L'interférence d'ARN (RNAi)	74
7.1. Quelques grandes étapes de leur découverte	75
7.2. Le mécanisme de l'inhibition traductionnelle	76
7.3. Les applications de l'ARNi	79
Chapitre IV. Analyse du génome, du transcriptome	81
1. Détection des mutations	81
1.1. Méthodes basées sur la séquence	81
1.2. Méthodes basées sur l'hybridation	84

1.3. Méthodes basées sur la PCR	88
1.4. Méthodes basées sur la conformation de l'ADN	89
1.5. Sensibilité d' hétéroduplex à une coupure	90
2. Analyse du polymorphisme	94
2.1. Etablissement du profil des protéines	94
2.2. <i>Analyse des fragments de restriction RFLP</i>	94
2.3. Amplified Fragment Length Polymorphism AFLP	95
2.4. ADN polymorphe amplifié au hasard (Random Amplified Polymorphic DNA-RAPD)	95
2.5. Les microsatellites ou SSR	98
3. Les techniques d'analyse du transcriptome	100
3.1. Puces à ADN	100
3.2. Technique SAGE (<i>serial analysis of gene expression</i>)	102
Références bibliographiques	104

I. Génétique de la levure, Animaux, et plantes transgéniques

1. Généralités

On qualifie de transgénèse la transformation intentionnellement opérée dans le patrimoine génétique « naturel » (ou encore « sauvage ») d'une espèce pour créer des populations dites transgéniques constituées d'organismes génétiquement modifiés (O.G.M.) ayant acquis une nouvelle caractéristique liée à l'expression du transgène. Cette technique permet des progrès considérables dans la connaissance du fonctionnement des gènes et des mécanismes qui gouvernent les fonctions biologiques. Ses applications potentielles pour produire des médicaments, des vaccins, des protéines alimentaires ou en amélioration génétique sont très vastes, mais encore limitées en raison de son coût et de la difficulté de sa mise en œuvre.

2. Les levures

D'une façon générale, des microorganismes génétiquement modifiés sont élaborés en vue de deux types d'utilisations :

a) Production d'une molécule purifiée. Le microorganisme est utilisé comme bioréacteur et ne se retrouve pas dans la formulation finale du produit. La plupart des microorganismes OGM répertoriés entrent dans cette catégorie. Les molécules ainsi purifiées sont principalement de deux types :

- Molécule d'ADN pour la création subséquente d'un autre organisme transgénique (mycète, plante, animal, virus...).
- Autre molécule d'intérêt, en fonction de la typologie des finalités :
 - production de connaissances;
 - usage médico-pharmaceutique;
 - produit à valeur ajoutée;
 - amélioration des modes de production;
 - protection de l'environnement;
 - production d'autres bioproduits.

b) Exploitation in situ.

Les propriétés intrinsèques du microorganisme sont altérées en vue d'une exploitation in situ à des fins diverses, en fonction de la même typologie des finalités :

- production de connaissances;
- usage médico-pharmaceutique;
- produit à valeur ajoutée;
- amélioration des modes de production;
- protection de l'environnement.

Les levures sont des organismes unicellulaires qui se comportent à bien des égards comme des bactéries. Bien que la levure soit capable de reproduction sexuée, les cellules se multiplient généralement de façon végétative par bourgeonnement avec un temps de dédoublement de quelques heures sur les milieux de croissance usuels. Son cycle cellulaire comprend des stades haploïde et diploïde. Les levures croissent sous forme de cellules isolées en culture liquide et forment des colonies sur milieu solide, tout comme les bactéries. Le bagage génétique des levures contient environ cinq fois plus de bases que le génome bactérien, réparties en plusieurs chromosomes.

Deux types d'événements générant la variation génétique sont connus chez les levures. Tout d'abord, des mutations occasionnent des changements dans la séquence d'ADN. En second lieu, le processus de recombinaison entre deux séquences d'ADN homologues engendre de nouvelles combinaisons génétiques. Les mutations peuvent être simples et ne concerner que le changement d'une base dans l'ADN, se limitant à la modification d'une seule protéine qui se traduira par des effets phénotypiques correspondants. Des mutations plus complexes peuvent également intervenir à une plus grande échelle par délétion, insertion, duplication ou réarrangement de l'ADN, entraînant des variations phénotypiques plus vastes qui peuvent aller jusqu'à l'apparition ou à la disparition de voies métaboliques entières. La propriété de mobilité de certains éléments d'ADN, les transposons, ainsi que la perte ou l'acquisition de plasmides peuvent aussi engendrer des changements phénotypiques d'importance variable.

Le traitement de l'ADN avec des agents chimiques (agents désaminant, oxydant, alkylant ou intercalant, analogues de bases) ou physiques (rayons ultraviolets ou radiations ionisantes) peut de son côté induire la mutagenèse avec une efficacité accrue et certains de ces agents ne vont agir que sur un type spécifique de bases.

Afin d'obtenir un mutant, des techniques d'enrichissement sélectif doivent être utilisées. Lorsqu'une mutation survient en présence d'une pression de sélection, les mutants les plus performants dans les conditions utilisées seront favorisés, vont se multiplier plus activement et envahir la culture.

L'avènement de la technologie de l'ADN recombinant a toutefois permis d'envisager l'altération des génomes à des sites particuliers de façon à créer des mutations spécifiques. Cette mutagenèse dite dirigée, qui implique une recombinaison entre séquences d'ADN homologues, suppose la caractérisation de la région d'ADN ciblée pour la mutagenèse ainsi que l'utilisation de molécules recombinantes afin d'inactiver spécifiquement un gène ou de le remplacer par une allèle modifiée.

3. Les animaux transgéniques

La transgénèse animale a été réalisée avec succès pour la première fois il y a 17 ans. De nombreuses utilisations de cette technique existent pour la recherche fondamentale. Elles consistent à ajouter, à inactiver ou à remplacer spécifiquement des gènes dans les génomes des animaux. Ces expériences apportent une moisson d'informations incomparables sur le fonctionnement du génome et sur les mécanismes de régulation des fonctions biologiques. De nombreux modèles animaux sont également obtenus pour l'étude de maladies humaines. La production de protéines recombinantes dans le lait d'animaux transgéniques est en passe de devenir une réalité industrielle. Le transfert de certains organes (cœur, rein, poumon...) et cellules (pancréas, foie) de l'animal transgéniques à l'espèce humaine est un objectif qui ne paraît plus inaccessible. Les applications de la transgénèse pour l'amélioration des productions animales sont encore à peu près inexistantes. Elles se cantonnent essentiellement à l'obtention de modèles pour des études de gènes et de fonctions biologiques particulières. La difficulté et le coût de la transgénèse chez les animaux domestiques sont une des causes essentielles de la lenteur des applications dans ce domaine. Toutefois, le transfert de gène dans des cellules fœtales cultivées suivi de leur transfert dans des ovocytes énucléés devrait contribuer grandement à améliorer cette situation.

4. Les plantes transgéniques

Des plantes transgéniques sont « naturellement » produites par l'action d'une bactérie (*Agrobacterium tumefaciens*) qui contient un plasmide porteur de gènes capables d'induire des tumeurs dites « galle du collet » et de s'intégrer dans le génome des plantes. En remplaçant les gènes capables de provoquer la tumeur par le gène que l'on désire introduire dans une plante, on transforme le plasmide bactérien en un vecteur permettant le transfert de gènes dans les espèces végétales qui peuvent être infectées par *Agrobacterium*. Des plantes transgéniques peuvent être obtenues par infection de protoplastes (cellules isolées dépourvues de paroi cellulaire) suivie de régénération à partir de ces protoplastes. Cette technique longue est limitée aux plantes capables de régénérer à partir de protoplastes. Il est aussi possible de découper de petits disques dans les feuilles et de les infecter par *Agrobacterium* contenant le plasmide recombiné porteur du gène à introduire. Placés sur des milieux adaptés, ces petits disques régénèrent, des racines poussent et une plante transgénique est obtenue. Cette technique est particulièrement bien adaptée à la production de tomates, pétunias, tabacs transgéniques, espèces qui, d'une part, sont facilement infectées par *Agrobacterium* et, d'autre part, produisent facilement des plantes à partir d'explants de feuilles. Les succès obtenus avec les dicotylédones ne furent pas, dans un premier temps, reproduits avec les monocotylédones, et en particulier avec les principales céréales. Cela conduisit à utiliser des techniques d'introduction directe d'ADN (en particulier par biolistique) dans les protoplastes ou encore dans la lignée germinale des céréales. La technique biolistique consiste à envoyer sur les cellules cibles des projectiles constitués de billes de tungstène ou d'or, recouvertes des gènes à introduire.

La transgénése est un moyen essentiel pour étudier le rôle des gènes dans l'expression des fonctions biologiques ainsi que leur fonctionnement. Elle permet également d'envisager des applications biotechnologiques diverses. De nouveaux outils, qu'il est encore nécessaire de perfectionner, permettent de mieux utiliser la transgénése pour des études fondamentales et de développer diverses applications dans les domaines agronomique et industriel.

II. Protéines recombinantes en recherche fondamentale

Le génie génétique, né dans les années 70, a ensuite connu un essor de grande ampleur en raison des applications variées dont il est à l'origine. En effet, de nombreuses molécules, dont la complexité de structure ne permettait pas la synthèse chimique et qui devaient auparavant être extraites directement de sources biologiques, peuvent désormais être produites en intégrant à une cellule hôte un matériel génétique hétérologue. 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. La protéine, synthétisée par une cellule différente de sa cellule d'origine, est ainsi dite hétérologue ou recombinante. C'est ainsi par exemple, que l'insuline, utilisée dans le traitement du diabète, provenait de pancréas de porcs ou de bœufs jusqu'en 1982, date à partir de laquelle cette protéine a pu être produite à l'échelle industrielle à partir de cultures de bactéries : ce fut le premier cas de commercialisation d'une protéine recombinante. De même, en 1985, l'hormone de croissance humaine, qui était jusqu'alors extraite d'hypophyses prélevées sur des cadavres, fut également mise sur le marché sous forme recombinante. En 1987, le rtPA (recombinant tissue plasminogen activator) est la première protéine produite par culture de cellules CHO (Chinese hamster ovary) à être commercialisée. Ce type de technologie trouve aujourd'hui un champ d'applications très vaste, principalement dans l'industrie pharmaceutique (médicaments, vaccins, et probablement bientôt thérapie génique), mais aussi dans le domaine plus controversé actuellement de l'agriculture (amélioration des plantes, résistance aux maladies, aux insectes, etc ...).

1. Les vecteurs d'expression et les cellules hôtes

Un système recombiné résulte de l'association entre un hôte, qui peut être un micro-organisme ou un organisme multicellulaire, et un vecteur d'expression (plasmide par exemple) contenant un gène codant pour une protéine d'intérêt.

Dans le but de produire en grande quantité une protéine de structure correcte, le vecteur idéal doit avoir plusieurs qualités particulières : il doit, premièrement, pouvoir pénétrer en de nombreux exemplaires dans la cellule hôte ou se répliquer afin que le gène de la protéine hétérologue y soit présent en de multiples copies, deuxièmement, contenir un promoteur "fort" pour que ces copies soient exprimées préférentiellement par rapport aux

autres gènes de la cellule hôte, et troisièmement, être stable afin de persister dans la cellules le plus longtemps possible.

Les qualités de l'hôte idéal sont encore plus diverses. Il doit, bien entendu, exécuter correctement les instructions portées par le gène étranger mais aussi pouvoir, si nécessaire, réaliser des modifications post-traductionnelles permettant d'obtenir une protéine ressemblant au mieux à la protéine naturelle. De plus, afin de permettre une production de masse en bioréacteur, les cellules hôtes doivent être résistantes, peu exigeantes en éléments nutritifs et pouvoir se multiplier rapidement et à densité élevée. S'il s'agit de plantes ou d'animaux entiers, le coût de leur culture ou élevage doit être l'imité.

Une multitude de combinaisons de cellules hôtes, de vecteurs, d'étiquettes et de conditions de culture sont disponibles pour exprimer et purifier une protéine recombinante. Tout projet d'expression de protéine recombinante commence par le choix de l'hôte, qui conditionne le vecteur à utiliser, les équipements et les milieux de culture. La définition de la stratégie de production hôte/vecteur est aussi essentielle, puisqu'elle conditionne les étapes de purification nécessaires à l'obtention de la protéine d'intérêt à la pureté voulue. Il existe plusieurs systèmes d'expression : procaryotes (*Escherichia coli* et *Bacillus subtilis* par exemple), eucaryotes (tels que les levures, les cellules d'insectes et de mammifères).

Le choix de système hôte sera en fonction de la protéine à produire et de son utilisation. Le taux d'expression, qui est l'un des éléments de décision, varie d'un système à l'autre pour une même protéine, et d'une protéine à l'autre pour un même système. Aussi et surtout, le temps et les moyens à mettre en oeuvre pour les opérations en aval : extraction, purification, contrôle de la qualité. Le choix peut être évident dans certains cas, par exemple, si la protéine nécessite des glycosylations, (les systèmes bactériens, ne permettant pas cette modification post-traductionnelle), seront éliminés d'office.

2. Expression en système procaryote: *Escherichia coli*

La bactérie *Escherichia coli* (*E. coli*) est un hôte de choix car elle a une vitesse de croissance remarquablement rapide, avec un temps de génération de 20 à 30 minutes dans les conditions optimales de culture. *E. coli* présente aussi l'avantage de pouvoir croître à des densités de cellules très élevées (10^{10} à 10^{13} bactéries/ml de milieu de culture liquide) comparativement aux cellules eucaryotes (10^6 à 10^8 cellules/ml). Avec un équipement de base (Erlenmeyer de culture et incubateur agitant), la biomasse produite est de l'ordre de 5-

10 g/1 de culture, sachant qu'elle peut atteindre 150 g/1 en bioréacteur. D'autre part, les milieux pour la culture d' *E. coli* peuvent être préparés à moindre coût par rapport à ceux utilisés en système eucaryote. Enfin, cette bactérie a servi à produire les premières protéines recombinantes, comme l'insuline humaine, en 1982, ce qui a entraîné le développement de nombreux vecteurs d'expression et de souches génétiquement modifiées. *E. coli* est ainsi devenue un véritable cheval de bataille pour l'expression de protéines recombinantes procaryotes ou eucaryotes. Dans ce dernier cas, il faut veiller à l'absence de codons rares pour *E. coli* dans la séquence.

2.1. Stratégie générale

D'une façon synthétique, l'expression d'un ADN recombinant s'appuie sur la succession d'opérations suivantes, chacune générant ses propres difficultés :

- tout d'abord, il s'agit d'isoler un gène d'intérêt, codant pour une protéine dont la synthèse artificielle en quantité est nécessaire : (médicament, additif, ...);
- le gène d'intérêt est alors inséré dans un vecteur (en général un plasmide), qui doit être capable de se répliquer et de répliquer le gène étranger qui lui est rattaché ;
- la molécule d'ADN recombinant obtenue (vecteur + gène d'intérêt) est ensuite introduite dans une cellule hôte, qui exécute les instructions qui lui sont fournies par le gène d'intérêt, et réalise les modifications post-traductionnelles dans l'objectif de fabriquer la protéine recherchée. L'hôte peut être une bactérie (*E. coli*), une levure, une cellule de mammifère ou encore une plante ou un animal transgéniques ;
- il s'agit alors de passer par une phase de production, généralement en fermenteur (en ce qui concerne les micro-organismes et cellules de mammifères), afin de produire les volumes de protéine nécessaires ;
- la protéine doit enfin être récupérée puis purifiée, cette étape faisant appel à des techniques plus classiques (séparation, extraction, purification) mais pouvant se révéler longue et onéreuse ;
- il s'ensuit des essais biochimiques et immunologiques (contrôle qualité) permettant de contrôler la pureté de la protéine obtenue, son activité, etc.

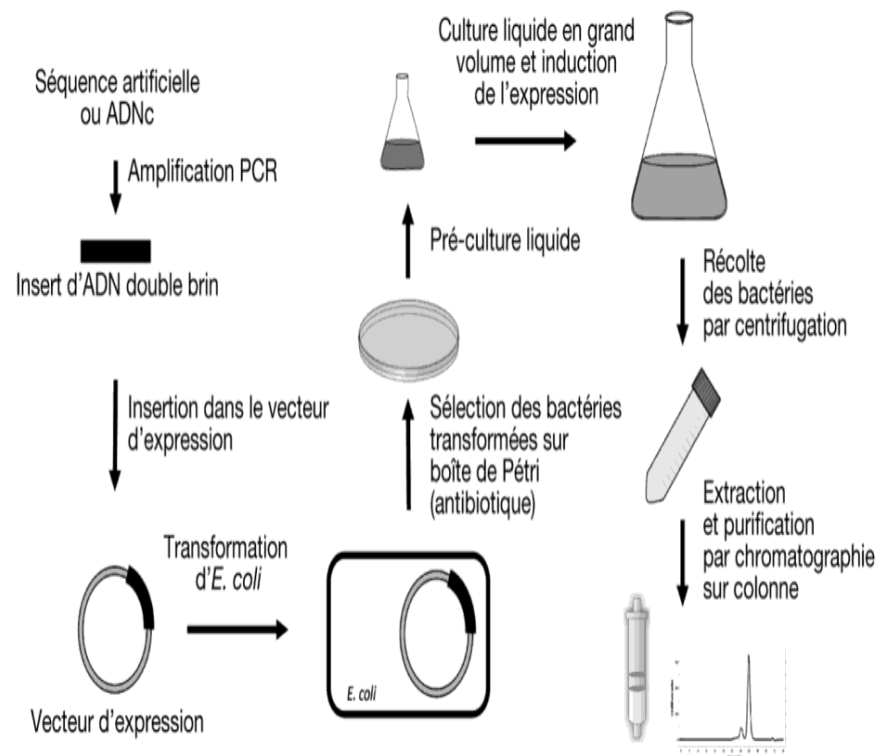


Figure 1. Stratégie générale de production des protéines recombinantes chez les procaryotes

2.2. Avantages et défauts du système

Avantages

Ses avantages incontestables sont :

- une excellente caractérisation génétique et physiologique,
- un temps de génération court et une bonne adaptation à la culture en masse.
- Nombreux vecteurs d'expression disponibles, facilité de transfection dans les cellules hôtes
- Bons rendements de production des protéines (*E. coli* possède une grande capacité à accumuler des protéines étrangères, puisqu'elles peuvent représenter plus de 20% des protéines cellulaires totales. Le taux d'expression de la protéine hétérologue peut atteindre plusieurs grammes par litre.
- Corps d'inclusion cytoplasmique : facilité de purification.

Inconvénients

- Absence de modifications post-transductionnelles.
- Influence du vecteur sur le taux de croissance cellulaire, la consommation d'énergie
- activité biologique et l'immunologie peuvent différer de la protéine native.
- Corps d'inclusion : protéines insolubles ou mal repliée.
- Sécrétion dans l'espace périplasmique : rendement moins important.
- Présence d'endotoxine bactérienne

2.3. Construction des vecteurs d'expression

L'utilisation d'*E. coli* pour la production d'une protéine d'intérêt repose essentiellement sur l'utilisation d'un plasmide contenant un gène codant cette protéine. Ce plasmide est introduit dans les bactéries pour permettre une surexpression de la protéine correspondante. Ce système peut être largement adapté ou modulé par rapport à la protéine que l'on souhaite exprimer qui peut s'avérer insoluble, dégradée, inactive ou toxique pour l'hôte. De ce point de vue, le succès dépend du choix du vecteur d'expression, qui joue un rôle central pour surmonter ces éventuelles difficultés. Un vecteur typique pour l'expression de protéines chez *E. coli* comporte différents éléments dont certains sont optionnels.

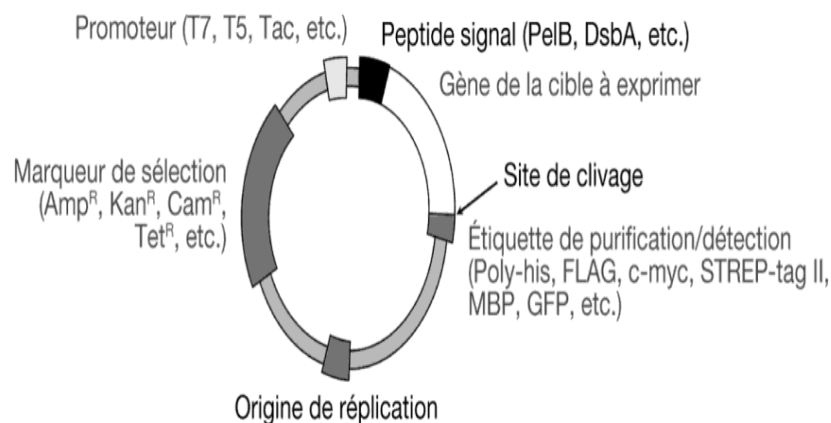


Figure 2. Organisation générale d'un vecteur d'expression procaryote

Les vecteurs d'expression dans *E. coli* sont généralement construits autour de plusieurs éléments essentiels. La figure 2 présente la conformation de base d'un vecteur d'expression. Chaque type de vecteur a sa conformation type mais des points communs existent.

Il faut que chaque vecteur contienne :

- Une origine de réplication pour contrôler le nombre de copies du vecteur d'expression.
- Un promoteur, permet l'initiation de la transcription. Il est placé 10 à 100 nucléotides avant le site de fixation du ribosome.
- Le site de clonage multiple (MCS) est l'élément indispensable pour intégrer le gène à exprimer au sein du vecteur d'expression. Sa configuration peut être une succession de sites de restriction mais également un site de clonage unique spécifique de kits commerciaux.
- La séquence Shine-Dalgarno (SD) permet l'initiation de la traduction.
- Un marqueur de sélection peut également être présent sur le vecteur. Il s'agit en général d'un gène de résistance à un antibiotique. Le milieu de culture sera alors complété de l'antibiotique adéquat et permettra ainsi la sélection des bactéries transformées.
- Etiquette ou tag généralement insérer dans le polylinker qui permettra une protéine de fusion facilement détectable et purifiable, souvent par chromatographie d'affinité.
- Signal de terminaison transcription-terminateur

a-Origine de réplication

Le gène doit être répliqué dans la bactérie, autrement, on va le perdre très rapidement. La méthode générale consiste à l'insérer dans un plasmide qui porte une origine de réplication. L'utilisation de plasmides pour exprimer un gène dans *E. coli* est brevetée, ceci a amené l'utilisation d'une autre origine de réplication, celle du chromosome bactérien.

Il existe plusieurs origines de réplication (*ori*) disponibles. Celle de bluescript (500-700 copies par chromosomes), celle de pBR322 (15-20 copies) et celle de P15A (10-12 copies).

Si le produit du gène est toxique on utilisera une origine de réplication qui donne peu de copies, au contraire si le gène n'est pas toxique et s'il est bien régulé, on utilisera une origine à fort nombre de copies.

b-Les promoteurs

Un promoteur est une séquence d'ADN que l'on trouve en amont du codon correspondant au premier acide aminé de la protéine à exprimer. Il permet à l'ARN polymérase de la cellule hôte de s'arrimer à l'ADN, pour initier la production des ARN messagers à traduire en protéine. Deux domaines en amont du site d'initiation de la transcription sont importants dans les promoteurs procaryotes. Le domaine à -10 (la boîte, 5' T-A-T-A-A 3') et un domaine à -35 (5' T-T-G-A-C-A 3'). Ces deux domaines sont en contact avec la RNA polymérase lors de l'initiation de la transcription.

La séquence promotrice est donc indispensable. Elle a un impact direct sur le taux de transcrits et par conséquent sur le rendement de production de la protéine.

Deux caractéristiques des promoteurs sont importantes pour l'expression des protéines, la force du promoteur et son inductibilité.

- *La force du promoteur,*

Les promoteurs bactériens sont relativement faibles du moins pour les besoins de production. La séquence promotrice a un impact direct sur le taux de transcrits et par conséquent sur le rendement de production de la protéine.

- *L'inductibilité, c'est à dire son contrôle :*

Le plus souvent on désire utiliser un promoteur inductible par exemple lorsque la protéine est très toxique pour la bactérie. Dans ce cas, la protéine est produite en début de la culture et inhibe la croissance cellulaire. En utilisant un promoteur inductible, on cherchera à inhiber la production au début de la phase de croissance des bactéries puis à l'approche de la phase stationnaire, on induira la production.

Il existe plusieurs systèmes d'induction (tableau 1), le système d'induction le plus utilisé est l'induction par l'isopropyl β -D thiogalactoside (IPTG est un réactif en biologie moléculaire. Ce composé est utilisé comme un équivalent de l'allolactose, un métabolite du lactose qui entraîne la transcription de l'opéron lactose. Contrairement à l'allolactose, l'atome de soufre de l'IPTG crée un pont chimique qui n'est pas hydrolysable par la cellule, ce qui rend sa concentration constante. D'une manière générale, les promoteurs que l'on trouve dans les vecteurs d'expression sont directement ou indirectement sous le contrôle d'une séquence opératrice lactose, sur laquelle vient se lier le répresseur Lac I pour

empêcher la transcription. L'ajout au milieu de culture d'un analogue du lactose non hydrolysable (l'IPTG, pour isopropyl p-D-l thiogalactopyranoside) va lever cette inhibition en se liant à LacI, ce qui permet de déclencher la transcription au moment optimal pour l'expression.

Tableau 1. Les différents systèmes d'induction d'expression.

Promoteur	Induction	Caracteristiques
LacUV5	lactose et IPTG	titrable, fuites importantes
Trp	tryptophane	titrable, fuites
T7	IPTG	nombreux vecteurs disponible, fuites
Ara	arabinose	titrable, répression
IpL	temperature	non titrable

On peut distinguer plusieurs types de promoteurs utilisables chez *E. coli* :

- Les promoteurs natifs d'*E. coli* : ils sont peu utiles pour l'expression recombinante, car ils ne sont pas suffisamment forts et peuvent laisser fuiter une expression basale.

-Les promoteurs hybrides : Le remplacement des séquences du promoteur du gène *lacZ* en amont de la position -20 par les séquences du promoteur de l'opéron tryptophane a permis de faire un promoteur hybride : Ptac. Cette construction donne une augmentation de la transcription de 10 fois. Bien qu'ils combinent leur force de transcription à l'avantage de la régulation du promoteur lac, les promoteurs hybrides ont été supplantés ces dernières années par ceux d'origine virale.

-Les promoteurs d'origine virale (T5 et T7).

Promoteur T7 d'origine virale, le promoteur de la protéine 10 du phage T7, est aujourd'hui probablement le plus utilisé pour l'expression recombinante, car la protéine peut représenter jusqu'à 50 % du total des protéines des cellules d'*E. coli*. Le promoteur T7 n'est pas reconnu par l'ARN polymérase d'*E. coli*. La transcription utilisant l'ARN polymérase T7 est très efficace. En effet,

- La polymérase utilise la plupart des nucléotides triphosphate de la cellule ce qui inhibe la transcription des gènes de l'hôte. Elle est 8 fois plus rapide que les polymérases bactériennes.

- L'ARN polymérase T7 est très spécifique, elle ne reconnaît pas les promoteurs de la bactérie, si bien qu'elle ne transcrit que le gène d'intérêt.
- L'ARN polymérase T7 n'est pas sensible aux inhibiteurs des ARN polymérases bactériennes tels que la rifampicine. Ainsi, l'ajout de rifampicine dans le milieu bloque toutes les polymérases sauf celle de T7, seul le plasmide est transcrit, tous les ribonucléotides sont utilisés pour la transcription du gène d'intérêt.

On le retrouve dans les vecteurs de la série pET (Novagen). Le gène à exprimer est cloné juste derrière le promoteur T7 reconnu très spécifiquement par l'ARN polymérase du phage T7. Dans ce système, il est nécessaire d'utiliser des souches d'*E. coli* modifiées (telles que BL21(DE3)) par l'intégration dans le génome bactérien du gène de ARN polymérase T7. Ce gène est sous le contrôle du promoteur *lacUV5* dérivé de *lac*, ce qui rend le système inductible par l'IPTG. L'expression basale peut être réprimée de façon plus stringente par une version mutée de Lac I (Lac I^Q) ainsi que par la co-expression du lysozyme T7, fourni par le plasmide pLysS ou pLysE dans les souches BL21 (DE3) pLyS ou E, qui en se fixant sur l'ARN polymérase T7 l'inhibe. Ainsi, avant induction de l'expression de la protéine d'intérêt, si un peu d'ARN polymérase T7 est produite, elle sera inhibée par le lysozyme T7. Après induction par l'IPTG, la quantité de la polymérase va dépasser ce que le lysozyme peut inhiber. On aboutit donc à un système qui réprime très fortement l'expression avant induction.

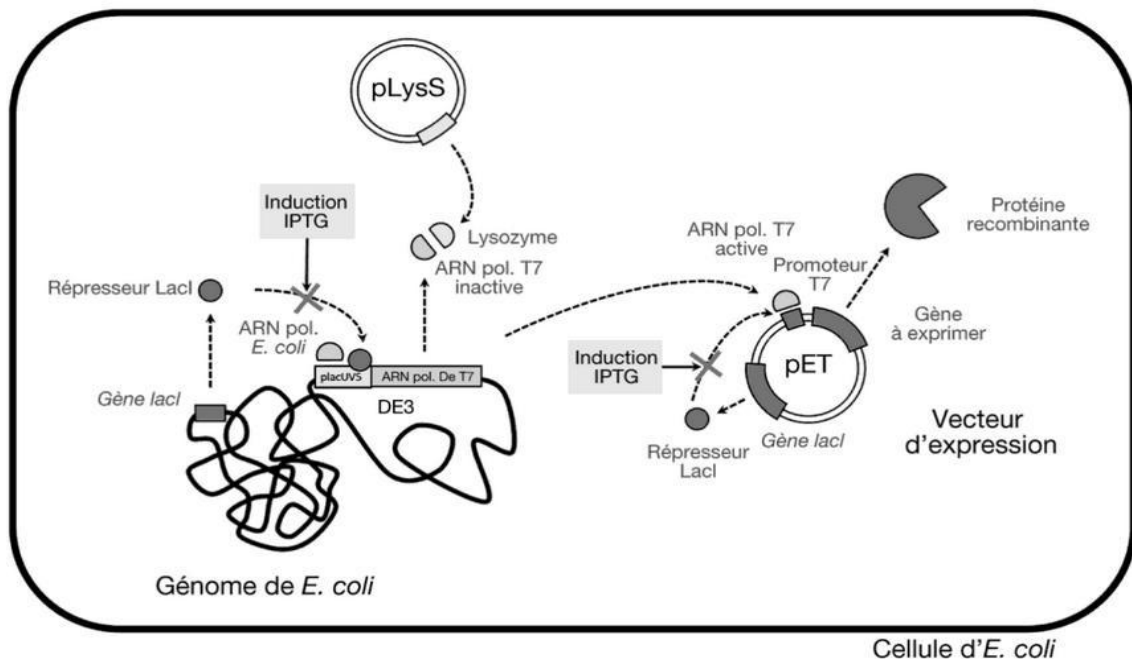


Figure 3. Le système d'expression pET/ BL21(DE3).

c- Séquences responsables de la traduction

Chez les procaryotes, l'initiation s'effectue par reconnaissance d'une séquence particulière (RBS, ribosome binding site). L'efficacité de la traduction est dépendante de la position de la séquence Shine-Dalgarno (SD) par rapport à la position du codon d'initiation. La séquence Shine-Dalgarno 5'-TAAGGAGG-3', est positionnée entre 4 et 44 nucléotides en amont du codon d'initiation. Elle permet l'initiation de la traduction. La composition de la séquence est complémentaire de l'extrémité 3' de l'ARN ribosomal 16s.

d-Séquences responsables de l'arrêt de la transcription

On peut dans certains cas ajouter en 3' du gène des séquences responsables de l'arrêt de la transcription. Au niveau d'un signal de terminaison, la polymérase libère la matrice et l'ARN monocaténaire néosynthétisé.

e-Étiquette de détection/purification (TAG)

Les protéines recombinantes sont fréquemment produites sous forme de protéines de fusion, afin de faciliter la détection de la protéine d'intérêt au cours du processus de production mais également pour faciliter sa purification. Une séquence de taille variable (tableau 2) peut être ainsi ajoutée à la séquence de la protéine d'intérêt, étiquette peptidique pour les séquences de quelques acides aminés, ou protéine de fusion pour les étiquettes de taille plus importante (jusqu'à 60 kDa parfois).

Les étiquettes peptidiques de petite taille peuvent être choisies pour essayer de limiter l'impact de l'étiquette sur la conformation de la protéine d'intérêt. La disponibilité de vecteurs d'expression permettant de cloner la protéine avec une étiquette peptidique en N-terminal ou en C-terminal permet de tester l'influence de l'étiquette sur la production et sur la fonction. Les étiquettes peptidiques les plus couramment utilisées sont les étiquettes FLAG, poly-His, c-Myc, poly-Arg, Strep II tag. La disponibilité d'anticorps reconnaissant ces peptides permet de tracer la cinétique de production de la protéine d'intérêt par la cellule hôte mais également de disposer d'une étape de purification très efficace par chromatographie d'affinité. Par exemple, les protéines portant une séquence FLAG sont purifiées sur des résines d'agarose couplées à un anticorps spécifique de cette étiquette. Les protéines portant une étiquette poly-His peuvent être purifiées par chromatographie d'affinité IMAC (Immobilized Métal ion Affinity Chromatography) sur des résines couplées à des ions métalliques divalents (nickel, cobalt, cuivre, zinc).

Les ions métalliques présentent des différences en matière de spécificité et d'adsorption. Le nickel, fréquemment utilisé, présente une forte capacité d'adsorption et

donc des rendements de purification élevés alors que le cobalt présente une plus grande spécificité de fixation, et donc une plus grande pureté de la protéine obtenue. De très nombreux types de résines commerciales facilitant la purification des protéines recombinantes sont disponibles aujourd'hui sous de nombreux formats : résines en viac, colonnes, « spin » colonnes, billes magnétiques ; ces deux derniers formats étant plutôt adaptés à des petites quantités de protéines.

Certaines protéines ne sont pas produites efficacement par la bactérie car elles sont instables ou peu solubles dans le compartiment cytoplasmique bactérien, notamment si elles sont de faible poids moléculaire (inférieur à 10 kDa) ou si elles sont membranaires. L'ajout d'une protéine de fusion à ces protéines améliore leur solubilité, limite leur protéolyse et permet ainsi une meilleure production. Les protéines de fusion les plus fréquemment utilisées sont la MBP [Maltose-Binding Protein], la GST (Glutathione S Transferase), la NusA (N-utilization substance protein-A), la Trx (Thioredoxin), l'ubiquitin et la SUMO. Les raisons de l'impact de ces protéines sur la solubilité ne sont pas toujours très claires. Nus-A, MBP et Trx présentent la meilleure capacité d'augmentation de la solubilité. La protéine Nus-A est un facteur de transcription bactérien qui diminue la vitesse de transcription, limitant ainsi l'agrégation de protéines mal renaturées. Dans le cas de la protéine MBP, sa présence améliore la solubilité des protéines de fusion par une activité intrinsèque de protéine chaperonne qui facilite le repliement de la protéine. Trx est une enzyme d'oxydoréduction bactérienne qui limite la formation de corps d'inclusion, en limitant la formation de ponts disulfures inappropriés. La protéine MBP présente un intérêt pour la purification des protéines de fusion par rapport aux deux autres, puisqu'elle peut être purifiée par chromatographie d'affinité. À l'inverse, l'ajout d'une étiquette peptidique supplémentaire est requis pour les protéines Nus-A et Trx. Pour optimiser la production et la purification des protéines, une combinaison d'étiquettes peut être utilisée (TAP ou Tandem Affinity Purification) permettant d'allier la fonction des protéines de fusion à l'efficacité des étiquettes peptidiques pour faciliter la purification.

De par la taille de ces protéines de fusion, leur impact sur l'immunogénicité, la fonction et la structure de la protéine d'intérêt est difficilement prévisible. Pour limiter cet effet, des séquences de clivage enzymatique sont souvent ajoutées afin d'éliminer la protéine de fusion sur le produit final, après la production et la purification. C'est également le cas pour éliminer les étiquettes peptidiques dans le cas des protéines produites pour l'industrie pharmaceutique par exemple.

Plusieurs enzymes sont utilisées pour cliver les étiquettes : entéropeptidase, facteur Xa, thrombine, protéase TEV (Tobacco Etch Virus) et protéase PreScission. Le choix entre les différentes protéases est basé sur la spécificité du clivage, le coût, le nombre d'acides aminés conservés après clivage et la facilité de séparation de l'enzyme et de la protéine d'intérêt après clivage. La sensibilité de ces enzymes à la composition du tampon est également un critère de choix, puisque par exemple l'entéropeptidase et la thrombine sont incompatibles avec les solutions contenant des agents réducteurs, et le facteur Xa ne fonctionne pas en présence d'agent chélatant. La température et le pH d'activité des enzymes peuvent également être des facteurs discriminants puisque certaines protéines y sont sensibles. La PreScission présente par exemple une activité de clivage maximale à 4 °C.

Tableau 2. Les différent TAG utilisée pour la production de protéine recombinante chez *E. coli*

Étiquette	Nombre d'acides aminés (taille en kDa)	Ligand ou matrice	Utilisation
Poly-His	2-10 (0,84)	Ion métallique divalent	Purification et détection
FLAG	8 (1)	Anticorps anti-FLAG	Purification et détection
c-Myc	11 (1,2)	Anticorps anti-c-Myc	Purification et détection
Strep-tag II	8 (1,1)	Strep-Tactin	Purification et détection
MBP	396 (42)	Résine couplée à l'amylose	Purification, détection, augmentation de la solubilité et de l'expression
GST	211 (26)	Agarose-glutathion	Purification, détection, augmentation de la solubilité et de l'expression
Nus-A	495 (54,8)	Pas de matrice dédiée	Augmentation de l'expression et de la solubilité
Trx	109 (11,7)	Pas de matrice dédiée	Augmentation de la solubilité
SUMO	100 (12)	Pas de matrice dédiée	Augmentation de la solubilité et de l'expression

2.4. Choix de la localisation subcellulaire de la protéine exprimée

Corps d'inclusion

Lors d'une expression cytoplasmique, la localisation finale des protéines recombinantes est le cytoplasme. Une expression trop forte ou l'expression de protéines ne parvenant pas à atteindre leur conformation native (ce qui est souvent le cas des protéines humaines nécessitant des ponts disulfures et/ou des glycosylations) entraîne une agrégation dans le cytoplasme et la formation de corps d'inclusions. On estime que 70% des protéines hétérologues produites se retrouvent dans les corps d'inclusions. En dépit de l'insolubilité des protéines résultantes, ces corps d'inclusions présentent des avantages non négligeables, telles que la stabilité et la résistance aux protéases, et une purification simplifiée, une simple centrifugation après lyse cellulaire permettant d'obtenir la protéine recombinante pure à près de 90% (Ventura et al., 2006). Mais les corps d'inclusions doivent être traités afin de les rendre solubles, généralement au moyen d'agents chaotropiques induisant la dénaturation des protéines recombinantes. Les étapes permettant de recouvrer leur activité sont très délicates et nécessitent de nombreuses mises aux points souvent très spécifiques de la protéine produite.

Soluble dans le cytoplasme

Certaines protéines exprimées dans le cytoplasme restent solubles et la renaturation n'est alors pas nécessaire. Pour augmenter les taux d'expression il est possible d'intervenir sur des paramètres extérieurs comme la température. En effet une baisse de la température lors de la production permet de diminuer le taux de synthèse protéique, évitant ainsi non seulement une accumulation dans le cytoplasme, mais limitant également les interactions hydrophobes qui initient des agrégats et aboutissent à la formation des corps d'inclusion (Fahnert et al., 2004).

La co-expression de protéines chaperonnes permet également d'accroître le taux de protéines correctement repliées. Cependant, cette technique, en sollicitant « la machinerie de synthèse » de l'hôte, peut également diminuer le rendement de production (Rinas et al., 2007).

Expression dans le périplasme (4% des protéines bactériennes)

Il faut pour cela l'ajout d'une séquence signal procaryote (phoA, OmpA) en 5' de la protéine exprimée. L'avantage principal est que l'environnement est oxydant, et permet

donc la formation de ponts disulfure. Cependant, les mécanismes de translocation sont mal connus, et les rendements sont faibles.

Expression extracellulaire

Les protéines sécrétées passent par le périplasme avant d'être libérées dans le milieu de culture. Cette voie confère les avantages de l'expression périplasmique décrits ci-dessus (milieu oxydant, protéines solubles et fonctionnelles). Mais des avantages supplémentaires sont apportés par cette expression :

- très peu de protéines endogènes sont excrétées, ce qui permet d'effectuer une purification plus simple et d'obtenir une bonne qualité ;
- une faible activité protéase propre au milieu extracellulaire permet d'éviter toute dégradation excessive ;
- l'excrétion a lieu en continu, évitant l'accumulation des protéines recombinantes surexprimées et la formation des corps d'inclusion ;
- et enfin la production de protéines dont l'expression cytoplasmique ou périplasmique serait nocive ou toxique pour la cellule se trouve ainsi délocalisée.

La présence de la membrane externe chez *E. coli* empêche la libération de la forme mature de la protéine dans le milieu extra-cellulaire. Il n'y a pas de phénomène d'excrétion. Il faut afin de libérer les protéines, altérer la perméabilité de la membrane externe. Chez *E. coli*, la protéine sécrétée se retrouve soit au niveau de la membrane externe soit dans l'espace périplasmique. Ces dernières peuvent alors diffuser à travers la membrane externe modifiée. Ainsi, des méthodes physiques ont été utilisées pour réaliser la libération des protéines dans l'espace extracellulaire telles le choc osmotique ou le traitement thermique. Toutefois ces méthodes sont difficiles à reproduire à grandes échelles, de plus elles s'accompagnent souvent de produits indésirables tels que les endotoxines qui doivent être ultérieurement éliminées. Toutefois beaucoup de micro-organismes excrètent la protéine dans le milieu de culture comme *Bacillus subtilis*.

3. Production chez les levures

3.1. *Saccharomyces cerevisiae*

S. cerevisiae est également très utilisée comme organisme modèle en biologie cellulaire et en génétique. En 1996, ce fut le premier eucaryote dont le génome a été séquencé. Son génome de 16 chromosomes est composé de 13 millions de paires de bases et de 6 275 gènes. On estime que l'homme partage 23 % de son génome avec cette levure. L'ensemble des données génomiques concernant *S. cerevisiae* sont rassemblées sur Saccharomyces Genome Database.

Cette levure fait partie des organismes eucaryotes les plus connus et par conséquent le premier organisme eucaryote à avoir été utilisé pour l'expression hétérologue de gènes (Boer et al., 2007). *S. cerevisiae* a permis la production du premier vaccin efficace contre l'hépatite B chez l'homme par expression hétérologue (McAleer et al., 1984).

3.1.1. Avantages et défauts du système

La levure présente deux avantages principales (en plus d'autres présentés dans le tableau 3) : d'une part, on dispose de vecteurs qui peuvent se maintenir dans les cellules comme des plasmides bactériens, et d'autre part on peut facilement intégrer un gène par recombinaison dans le chromosome.

De plus on dispose de plusieurs méthodes pour faire rentrer un ADN dans la levure. On peut utiliser soit une transformation chimique avec le l'acétate de lithium, soit une transformation avec du CaCl_2 sur des sphérobastes (les sphérobastes sont obtenus par une lyse de la paroi des levures par une lyticase), soit par électroporation.

Tableau 3. Les avantages et les défauts du système.

Avantages	Inconvénients
- Petit génome eucaryote facilement manipulable et bien caractérisé.	-N-glycosylation : immunogène pour l'homme.
- Absence d'endotoxine.	- Mauvais repliement de la protéine produite.
- Pas toxique.	- Fermentation peu couteuse.
- Bons rendements (quelques grammes par litre de culture).	- Hypoglycosylation.
- Modifications post traductionnelles simples.	- Mal sécrétion des protéines sauf petits polypeptides, ou sécrétion possible mais avec rendement diminué.
- Capables de synthétiser des protéines complexes.	

3.1.2. Vecteurs d'expression

La production de protéines recombinantes chez *S. cerevisiae* peut être réalisée à l'aide de trois types de vecteurs:

- les plasmides d'intégration (YIp),
- les plasmides épisomiques (YEp)
- et les plasmides centromériques (YCp).

Les YEps sont des plasmides basés sur l'origine de réplication endogène de 2 μ maintenue à un nombre élevé de copies à l'intérieur de la cellule (5 à 30 copies), ce qui permet une expression génétique robuste, mais peut imposer une charge considérable aux cellules, entraînant une instabilité accrue du plasmide. Les YCP sont des plasmides basés sur une combinaison de séquence à réplication autonome (ARS) et de séquence centromérique de levure (CEN) maintenues dans un faible nombre de copies (1 à 2 copies), qui sont plus stables, mais des niveaux d'expression géniques plus faibles limitent leur utilisation. Ainsi, l'intégration d'une cassette d'expression dans un locus cible sur un chromosome de levure natif est bénéfique, car elle permet l'élimination de la pression sélective après la construction de la souche recombinante.

- **Marqueurs de sélection**

De nombreuses souches de laboratoire sont auxotrophes pour des acides aminés ou des nucléotides, c'est à dire qu'elles ne peuvent pas pousser en l'absence du composé dans le milieu. On peut donc utiliser la complémentation des enzymes entrant dans les voies de biosynthèse. Les plus utilisés sont les gènes URA3 ou LEU2 responsables de la synthèse de l'uracile ou de la leucine. Mais il existe d'autres systèmes de complémentation. L'inconvénient de ce système est qu'il faut utiliser des milieux biochimiquement définis or dans ces milieux, les levures poussent moins bien et de ce fait, la production est moins bonne. De plus les milieux définis sont plus chers à fabriquer qu'un milieu non défini comportant par exemple de l'extrait de levure et de la peptone comme un milieu de culture. Aussi, on préfère souvent utiliser des marqueurs de sélection qui donnent la résistance à des drogues, la tunicamycine (TUN) ou l'hygromycine (Hgm), mais il y a toujours un risque d'apparition de résistance spontanée dans la souche utilisée et donc de perte du plasmide.

- **Promoteurs**

Pour exprimer un gène chez la levure, deux types de promoteurs existent: les promoteurs constitutifs et les promoteurs inductibles. Des systèmes basés sur des promoteurs inductibles ont cependant été utilisés en recherche chez la levure. Les promoteurs constitutifs offrent une simplicité et des niveaux d'expression relativement constants, tandis que les promoteurs inductibles sont couramment utilisés lorsque la séparation de la croissance et de la production est souhaitée, ce qui peut éventuellement empêcher une sélection involontaire de non recombinants à croissance plus rapide.

Les systèmes GAL1, GAL7, GAL10, CUP1, ADH2 ou MET25 (sont des promoteurs inductibles naturellement existants chez *S. cerevisiae*).

Exp :

-La régulation des gènes utilisés pour la métabolisation du galactose est très utilisée pour réguler la production de protéines recombinantes. Ce promoteur est en effet très fort : en présence de galactose, la production de la kinase GAL1 est augmentée de 1000 fois et représente 0.8% des protéines totales de la levure.

- Un autre gène inductible très utilisé est celui codant pour la phosphatase acide (PHO5). Cette phosphatase acide a pour rôle de déphosphoryler les protéines du milieu extérieur et ainsi de procurer une source de phosphate à la cellule. Lorsque le phosphate inorganique dans le milieu est faible, le gène est transcrit, la phosphatase acide est sécrétée, elle déphosphoryle des protéines, le phosphate libéré est incorporé dans la levure. A l'inverse,

quand il y a du phosphate inorganique dans le milieu, la phosphatase devient inutile, le gène la codant n'est pas transcrit. On peut utiliser ce système en employant un milieu riche en phosphate inorganique pour éviter la transcription et un milieu dépourvu pour l'augmenter.

Un éventail de promoteurs constitutifs associés à des niveaux de transcription variés (ou force de promoteurs) est communément utilisé chez la levure. Ils sont souvent issus de la voie de la glycolyse de *S. cerevisiae*: on y trouve les promoteurs des gènes de la phosphoglycérate kinase (PGK1), du pyruvate decarboxylase (PDC1), de la triosephosphate isomérase (TPI1), de l'alcool déshydrogénase I (ADH1), ou encore de la glyceraldéhyde-3-phosphate déshydrogénase (TDH3 ou GPD). Le promoteur du facteur d'élongation TEF1 est également largement utilisé et connu pour conférer un niveau de transcription élevé au gène qui se trouve sous son contrôle.

Tableau 4. Différents vecteurs d'expression chez *S. cerevisiae*

Vecteur	Promoteur	Marqueur de sélection (<i>E. coli</i> / <i>S. cerevisiae</i>)	Vecteur de base	Elément pour l'expression	Elément pour réplication	Elément pour intégration
pAS404	ADH1	Ampicilline/TRP1	pBR322	---	---	---
pACT2AD	ADH1	Ampicilline	---	Termineur ADH	2μ	---
pYC2/NTC	GAL 1	Ampicilline/URA3	pBR322	Epitope V5, poly- histidine, termineur CYC1	---	CEN6/ARS4
pYES_{2,1}	GAL 1	Ampicilline/URA3	pUC	Epitope V5, poly- histidine, termineur CYC1	2μ	---

3.1.3. Système de sécrétion

Les peptides signaux hétérologues fonctionnent pour la plupart dans la levure, il n'est donc pas nécessaire d'en changer lorsqu'on veut exprimer une protéine habituellement sécrétée. Dans le cas de l'expression d'une protéine cytoplasmique, il suffit de faire une fusion avec un peptide signal tel que le "facteur α". Ce peptide signal est constitué d'une préséquence de 19 acides aminés suivi d'une prorégion de 67 acides aminés. La proséquence possède des sites de glycosylation et un site de clivage pour une endopeptidase permettant de cliver la proséquence.

Toutefois, il est difficile de produire des protéines sécrétées dans *S. cerevisiae*.

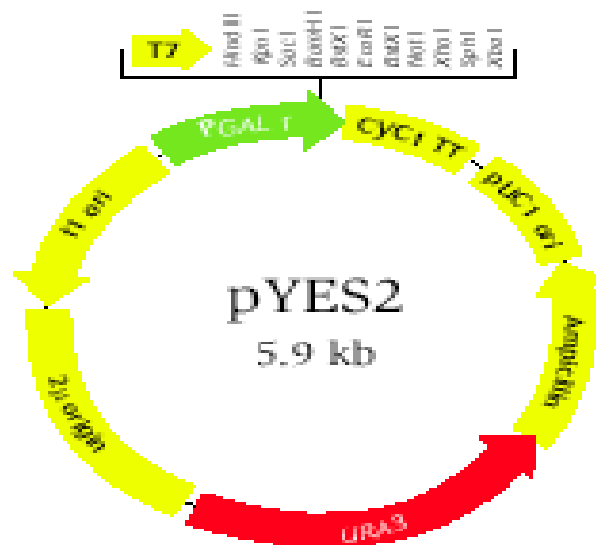
- En étudiant la glycosylation des protéines, il est apparu qu'il existe une étape limitante dans cette maturation quelque part entre le réticulum endoplasmique et l'appareil de Golgi.

Une première technique consiste à changer d'espèce et utiliser *Pichia pastoris*. Cette dernière permet l'expression de protéines membranaires qui sont difficiles à obtenir avec *S. cerevisiae*.

- La sécrétion ne s'effectue que pendant la phase de croissance exponentielle, pour une protéine sécrétée, il est donc préférable d'utiliser un promoteur constitutif.

3.1.4. Modifications post-traductionnelles

La levure *S. cerevisiae* permet l'expression hétérologue de gènes en assurant des modifications post-traductionnelles comme la glycosylation. Cette modification est notamment nécessaire pour l'obtention d'une activité enzymatique, la localisation cellulaire, la stabilité de la protéine ou encore l'interaction entre protéines. L'importante activité de glycosylation de *S. cerevisiae* pose cependant deux problèmes. Tout d'abord, le phénomène d'hyper-glycosylation (ajout d'une centaine de résidus mannose à la place d'une dizaine) des sites de glycosylation modifie la protéine, diminue la quantité produite ainsi que les capacités de sécrétion. De plus, la glycosylation n'est pas identique à celle réalisée pour les protéines de mammifères, les sites de liaison et les acides aminés visés sont différents. Ce phénomène peut provoquer de mauvaises conformations et des pertes d'activité. Face aux difficultés rencontrées, l'utilisation d'autres levures s'est avérée nécessaire.



3.2. *Pichia pastoris*

Depuis les 20 dernières années, *Pichia pastoris* a remplacé dans de nombreux domaines *S. cerevisiae*. Son utilisation permet d'atteindre des niveaux d'expression de protéines recombinantes (par exemple l'hirudine) de trois à trente fois supérieurs à ceux obtenus chez *S. cerevisiae*.

Pichia pastoris est une levure méthylotrophe, c'est à dire qui peut utiliser le méthanol comme source unique de carbone. L'alcool oxydase (AOX), codée par deux genes AOX1 et AOX2, est l'enzyme clé qui lui permet d'oxyder le méthanol (CH₃OH) en formaldéhyde (HCOH). Cette réaction entraîne la formation d'une molécule de peroxyde d'hydrogène toxique pour la cellule, c'est pourquoi elle a lieu dans un compartiment cellulaire appelé le peroxysome. Cette organelle peut représenter jusqu'à 80 % du volume cellulaire total lorsque les cellules de *Pichia pastoris* croissent en utilisant le méthanol comme source de carbone. À cet endroit, le peroxyde d'hydrogène sera décomposé en oxygène et eau par la seconde enzyme de cette voie métabolique, la catalase. Une partie du formaldéhyde quitte le peroxysome pour être transformé en énergie à la suite de plusieurs réactions impliquant deux autres enzymes. L'autre partie sera convertie en dihydroxyacétone par l'enzyme DHAS («dihydroxyacétone synthase») pour la production de constituants cellulaire. Suite à des cultures en bioréacteur utilisant le méthanol comme source de carbone, les chercheurs ont observé que l'enzyme AOX pouvait représenter jusqu'à 30 % des protéines solubles totales. L'étude de la voie métabolique du méthanol a démontré que cette enzyme a une très faible affinité pour l'oxygène, la cellule compense donc cette faiblesse en produisant de très grandes quantités de l'enzyme.

À partir de cette observation, il fut proposé d'utiliser cette levure. La stratégie à employer pour la production de protéines hétérologues est donc de cloner le gène d'intérêt sous le contrôle du promoteur du gène AOX1. L'activation ou l'inactivation des gènes AOX1 et AOX2 permet d'obtenir trois phénotypes au niveau de l'utilisation du méthanol, soit Mut⁺, Mut^s ou Mut⁻ (utilisation du méthanol).

3.2.1. Avantages et défaut de système

Avantages

Système levure le plus adapté à la production de grande quantité de protéine recombinantes

- Même avantage que *Saccharomyces cerevisiae* en terme de facilité, de coût de production

- Biomasse plus importante : en fermenteur, on peut obtenir des densités cellulaires 10 fois plus importantes qu'avec *Saccharomyces cerevisiae*
- Rendement en protéines recombinantes plus importants qu'avec *Saccharomyces cerevisiae*.
- Sécrétion très efficace des protéines dans le milieu de culture (Purification facilitée) .
- Glycosylation plus semblable à la glycosylation des mammifères:

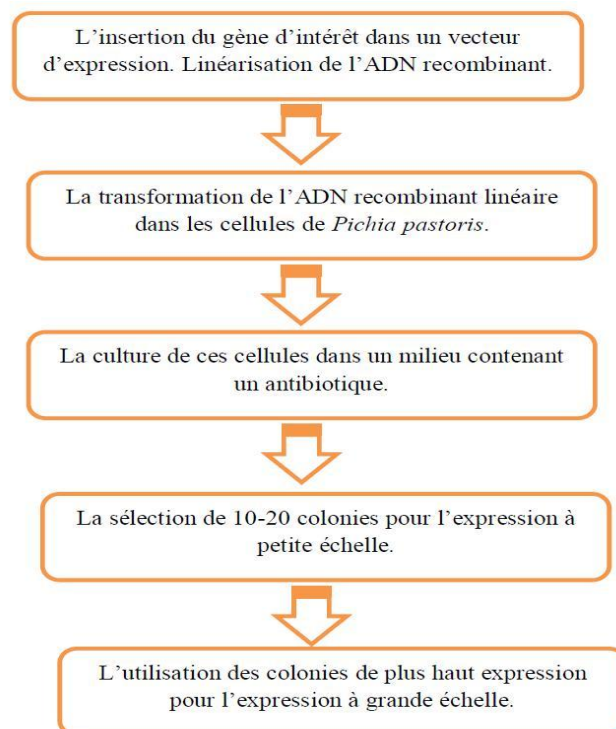
-les deux levures font des N-glycosylation type mannose, la longueur de la chaîne d'oligosaccharide est plus courte chez *Pichia* (8-14 mannose au lieu de 50-150 mannose)

-chez *Saccharomyces*, oligosaccharides ont des liaisons α 1-3 Glycane qui sont responsables du pouvoir hyper-antigénique des protéines recombinantes. *Pichia* non

Inconvénients

- Culture en fermenteur souvent requise pour atteindre un haut taux de production.
- Vecteurs disponibles limités.
- Modifications post-traductionnelles parfois inadéquates.
- Difficulté de sécrétion de certaines protéines.
- Possibilité de dégradation de la protéine d'intérêt par les protéases du milieu.

II.3.2.2. Étapes d'expression d'une protéine chez *Pichia pastoris*



3.2.3. Vecteurs d'expression

La première étape de la métabolisation du méthanol est une oxydation par le produit du gène AOX1 (le produit d'un deuxième gène, AOX2 est aussi impliqué, mais AOX1 est responsable de la majeure part de l'oxydation). L'expression d'AOX1 est régulée au niveau transcriptionnel.

Le vecteur typique contient un gène de résistance à un antibiotique et une origine de réplication pour la propagation dans *E. coli*. Il contient, en outre, le promoteur d'AOX1 et la partie terminale (3') d'AOX1. Ces deux parties sont interrompues par un site de clonage multiple suivi au besoin d'une séquence codant pour un peptide signal pour insérer le gène d'intérêt. Comme moyen de sélection, on utilise soit un gène codant pour une histidinol dehydrogenase permettant la synthèse d'histidine (HIS4) soit un gène bactérien procurant la résistance, soit un gène de résistance à la zéocine (Figure 4).

Ce plasmide est introduit dans *P. Pastoris* soit par électroporation soit par transformation de sphéroblastes comme pour *S. cerevisiae*.

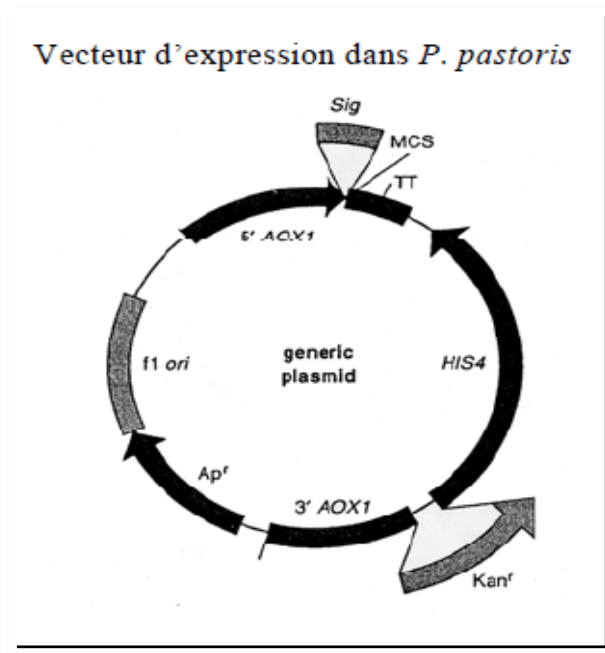


Figure 4. Vecteur d'expression chez *P. pastoris*

3.2.4. Mode d'intégration du gène d'intérêt

Les vecteurs utilisés pour l'expression des protéines recombinantes sont des YIP c'est-à-dire qu'ils ne possèdent pas d'origine de réplication chez la levure et donc ne peuvent se propager que par leur intégration dans le génome de l'hôte.

L'insertion de la cassette d'expression se fait par recombinaison homologue entre l'ADN plasmidique et les régions qui lui sont homologues dans le génome de *Pichia*

pastoris, cette recombinaison dirigée par le gène AOX1 montre une grande stabilité surtout lorsque le gène d'intérêt est présent en plusieurs exemplaires.

Il y a plusieurs possibilités d'intégration dans le chromosome :

- Le remplacement de gène ou insertion oméga. Dans ce cas il y a deux crossing over entre les promoteurs de l'AOX1 et la partie 3' de l'AOX1, les recombinants ont le phénotype Mut^s. (Mut^s signifie que la levure n'est plus capable d'oxyder le méthanol et donc plus capable d'utiliser cet alcool comme seule source de carbone). Dans ce cas, le plasmide est coupé en deux endroits avant la transformation (Figure 5).
- L'insertion de gène s'effectue par simple crossing-over entre le locus AOX1 du chromosome et une des séquences d'AOX1 présente sur le vecteur (région 5' ou région 3'). On a dans ce cas insertion d'un plasmide en amont ou en aval d'AOX1 qui peut demeurer fonctionnel. Le phénotype de tels transformant est Mut⁺ (Figure 6). Plusieurs copies peuvent ainsi s'intégrer en tandem. De telles insertions multiples peuvent être sélectionnées en utilisant la résistance au G418 ou à la zéomycine. Chaque gène entraîne une faible résistance, en augmentant la dose d'antibiotique, on sélectionne les recombinants qui ont le plus de copies.

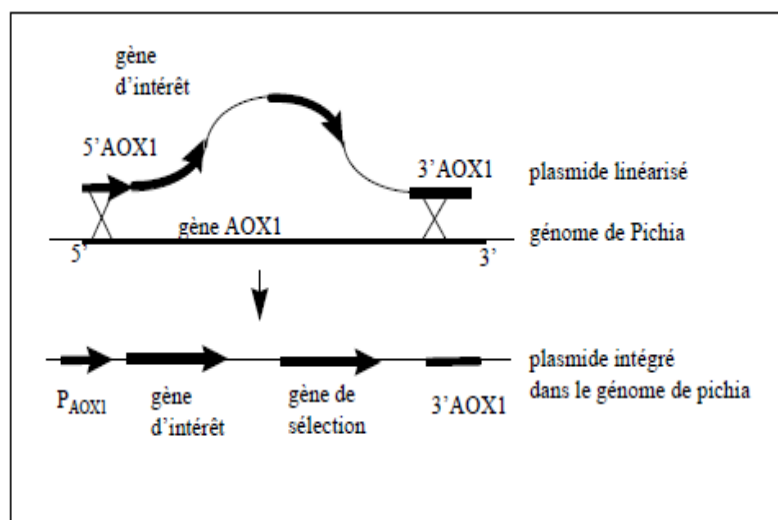


Figure 5. L'insertion de gène par deux crossing over (l'insertion oméga).

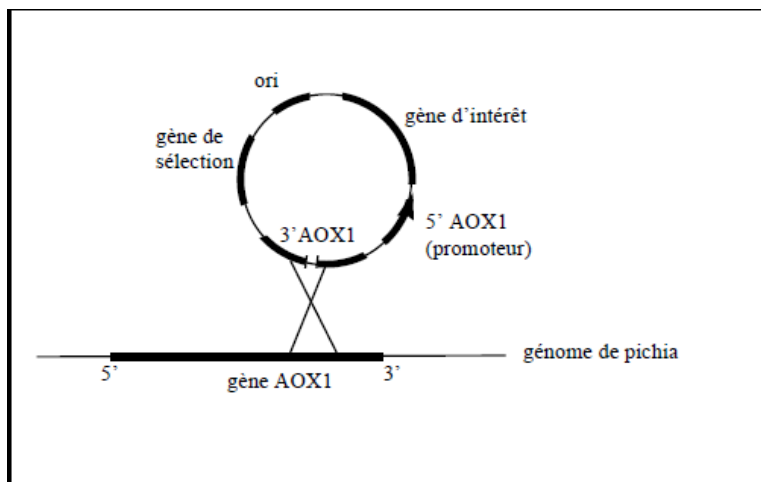


Figure 6. L'insertion de gène par un seul crossing over.

3.2.5. Système de sécrétion

La sécrétion d'une protéine dans le milieu extracellulaire implique une dépense supplémentaire d'énergie qui a comme effet une baisse du rendement de production de la protéine d'intérêt. Toutefois, la purification d'une protéine sécrétée chez *Pichia pastoris* est beaucoup plus simple, rapide et engendre moins de perte que la purification d'une protéine intracellulaire. Il peut donc être avantageux de construire un système permettant la sécrétion de la protéine d'intérêt dans le milieu extracellulaire.

Chez *Pichia pastoris*, la protéine d'intérêt contrôlée par le promoteur du gène AOX1 et liée à un facteur de sécrétion peut représenter plus de 80 % des protéines totales extracellulaires. Le peptide signal α de *S. cerevisiae* est efficace pour la sécrétion de protéines de petites tailles chez *P. pastoris*.

4. Production en cellules d'insectes

Les cellules d'insecte du système baculovirus a été montré à une excellente expression, c'est un système capable de fournir de grandes quantités de protéines recombinantes biologiquement actives. Il prévoit la construction d'un nucléopolyédrovirus recombinant contenant le gène étranger d'intérêt sous le contrôle du promoteur polyédrine fin même, qui est exceptionnellement forte. Lors de l'infection par le baculovirus recombinant les cellules expriment souvent des niveaux élevés de même de la protéine recombinante avec toutes les modifications post-traductionnelles nécessaires.

4.1. Avantages et défauts de système

Avantages

Ce système de production en infectant soit des insectes soit des cellules d'insectes présente plusieurs avantages :

- Fonctionnalité des protéines recombinantes : Les insectes sont des eucaryotes et ils sont capables d'effectuer tous les repliements des protéines avec les bons ponts disulfures et les bonnes oligomérisations, ce qui n'est pas le cas de bactéries ou des levures.
- Modifications post-traductionnelles : La plupart des modifications post-traductionnelles sont effectuées en baculovirus aux sites où on les trouve dans les protéines natives. Toutefois, ce système produit une très grande quantité de protéine et on observe parfois une saturation du système.
- Haut niveau d'expression : Le record de l'expression mentionnée avec baculovirus est de 1 gramme de protéine recombinante pour 10^9 cellules.

Inconvénients

- Coûts élevé de la production.
- Les cellules d'insectes possèdent les équipements enzymatiques leur permettant d'effectuer les modifications post-traductionnelles semblables à celles des Vertébrés. Toutefois, ce système produit une très grande quantité de protéine et on observe parfois une saturation du système.
- Systèmes de fermentations spéciaux (22°C, CO₂).

4.2. Cycle de multiplication d'un baculovirus

Dans la nature les virus sont présents à l'intérieur de polyèdres, structures cristallines facilement observables en microscopie photonique, qui forment de véritables “

corps d'inclusion ». La matrice protéique de ces corps d'inclusion est constituée par la polyédrine qui est une protéine de 29 kDa codée par le génome viral. Le virus, protégé des facteurs de l'environnement par ce corps d'inclusion, est ainsi disséminé à la mort de l'insecte. Après ingestion par un autre insecte, les polyèdres sont dissous dans le suc intestinal, libérant les particules virales qui pourront pénétrer les cellules de l'épithélium intestinal par fusion membranaire ou endocytose. Après passage de l'épithélium intestinal, les virus infectent les tissus de l'hôte. Dans les cellules, la nucléocapside est amenée jusqu'au noyau où elle libère son ADN (figure 7). La transcription des gènes viraux précoces peut alors commencer. La réplication de l'ADN commence environ 7 heures après l'infection et l'expression des gènes précoces est alors réprimée au bénéfice des gènes tardifs. Après environ 10 heures d'infection, des particules virales bourgeonnent à la surface des cellules. Plus tard, en fin d'infection, le gène codant la polyédrine est exprimé en très grande quantité grâce à la présence en amont d'un promoteur fort. La polyédrine est essentielle à la formation des corps d'inclusions permettant la dissémination des virus dans la nature. Cependant, en culture cellulaire *in vitro*, l'expression du gène de la polyédrine n'est pas indispensable pour obtenir un cycle de multiplication.

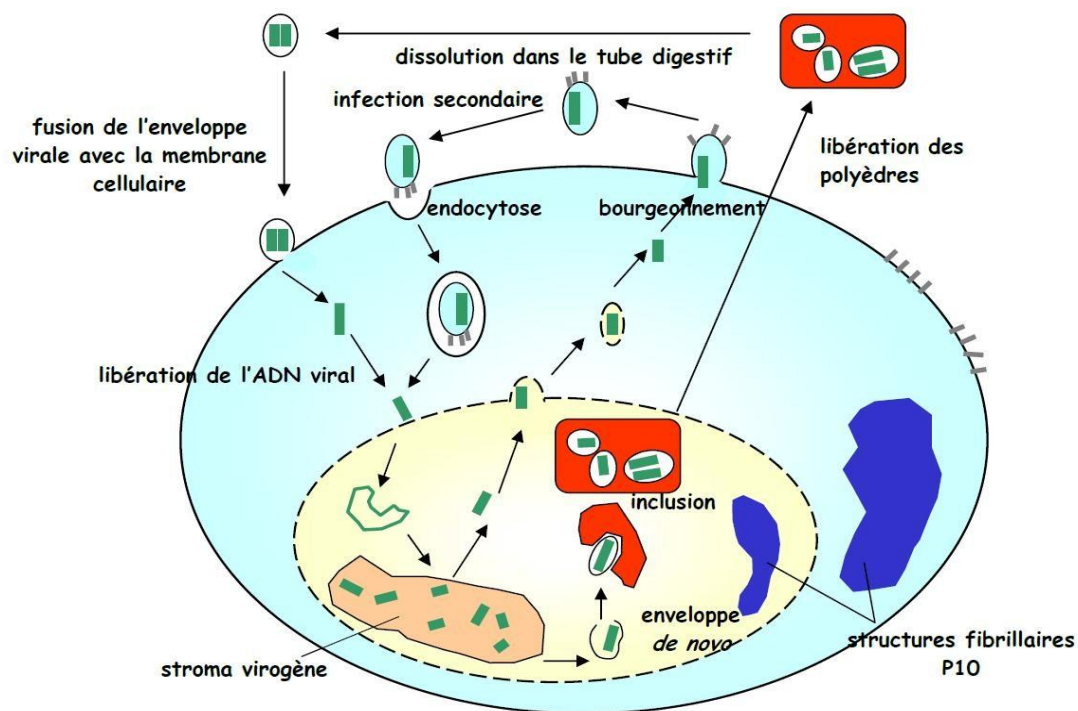


Figure 7. Cycle infectieux du baculovirus

4.3. Utilisation des baculovirus comme vecteurs d'expression de gènes étrangers

Le principe de l'utilisation des baculovirus comme vecteurs de gènes étrangers consiste à remplacer la région du génome du baculovirus qui code pour la polyédrine par un gène qui code une protéine "étrangère" d'intérêt. Ainsi, un gène étranger cloné à la place de la polyédrine sera à son tour fortement exprimé.

La taille du génome du baculovirus ne permet pas le clonage direct du gène étranger. Ce clonage se fait par l'intermédiaire d'un plasmide de transfert (par exemple : pBlue Bac III) qui contient le promoteur de la polyédrine. Le plasmide de transfert recombinant et l'ADN du baculovirus sauvage linéarisé sont co-transfectés dans les cellules d'insecte. Le gène d'intérêt X et le gène *lacZ* sur le plasmide de transfert sont encadrés de séquences de recombinaison identiques à celles situées sur l'ADN du baculovirus sauvage. Il se produit alors une recombinaison homologue entre le plasmide recombinant et le génome du baculovirus ce qui permet la production de baculovirus recombinants. Ainsi les cellules infectées par les baculovirus recombinants ne présentent plus de cristaux de polyédrine contrairement aux cellules infectées par les baculovirus sauvages. De plus, le gène *lacZ* étant transféré en même temps que le gène d'intérêt, il est possible de détecter les cellules infectées par le baculovirus recombinant : elles se colorent en bleu en présence de X-Gal. Il est alors possible de sélectionner les virus recombinants en prélevant les foyers des cellules colorées en bleu. Les baculovirus recombinants purifiés sont alors utilisés pour infecter des cellules afin de produire en grande quantité la protéine d'intérêt.

Un grand nombre de lignées de cellules d'insecte est aujourd'hui disponible pour étudier la multiplication des baculovirus et produire des protéines étrangères. Dans le cas du baculovirus d'*Autographa californica*, se sont essentiellement les cellules Sf21 (ou Sf9) qui proviennent de tissu ovarien de larves du lepidoptère *Spodoptera frugiperda*.

Principe :

- Clonage du gène codant la protéine d'intérêt dans un vecteur de « transfert »
- Co-tranfection du vecteur de transfert recombinant avec de l'ADN baculoviral linéarisé dans des cellules d'insectes de type Sf9. Il y a recombinaison homologue entre l'ADN viral et le vecteur de transfert le gène codant la protéine d'intérêt est inséré sur l'ADN viral
- Récolte du stock viral recombinant
- Infection de cellules Sf9 et expression de la protéine d'intérêt.

4.4. Cellules d'insecte ayant intégré le gène d'intérêt dans leur génome

Toutes les lignées cellulaires (la lignée Sf9, Sf21) ont la propriété intéressante de pouvoir intégrer de l'ADN dans leurs chromosomes. Si on transfecte une cellule avec un plasmide on a tout d'abord expression transitoire à partir du plasmide puis dans un deuxième temps établissement d'une culture cellulaire exprimant la protéine d'une manière stable puisque le plasmide s'est intégré dans le génome. Des centaines de copies du plasmide se retrouvent dans le génome.

Le promoteur : pour obtenir ces lignées stables on a besoin d'un promoteur reconnu par la machinerie transcriptionnelle de la cellule et d'un agent de sélection. On connaît un très grand nombre de promoteurs chez les insectes. On utilise soit un promoteur constitutif soit un promoteur inductible. Dans ce dernier cas, on peut utiliser le promoteur de la metallothionéine qui est inductible par le sulfate de cuivre. Une autre alternative est d'utiliser un promoteur viral précoce. Ainsi le promoteur OpIE2 du baculovirus d'*Orgyia pseudotsugata* est utilisé car il est fonctionnel dans un grand nombre de cellules d'insecte différentes.

La sélection : le facteur limitant est la transfection, si une cellule incorpore de l'ADN, elle incorpore un grand nombre de molécules présentes dans le milieu. Une première technique consiste à faire une cotransfection avec un plasmide comportant un gène de résistance. Les cellules ayant incorporé des plasmides dans leur génome seront hygromycine résistante, les autres seront sensibles et seront éliminées. Une deuxième technique consiste à intégrer le gène de sélection au plasmide. On augmente la pression de sélection pour sélectionner les cellules qui sont les plus résistantes, ces cellules comportent plus de plasmide, ainsi plus de copies du gène d'intérêt et donc produiront plus de protéines. En augmentant graduellement la sélection, on peut espérer sélectionner des lignées produisant un grand nombre de copies par recombinaison.

5. Production en cellules de mammifère

Les cellules mammifères sont dans la plupart des cas le meilleur système pour produire des protéines humaines, voir plus globalement mammifères, car elles permettent d'obtenir des protéines très proches de la forme native (modifications post-traductionnelles, repliement...). En effet, les cellules de mammifère, étant des eucaryotes plus complexes, sont en mesure d'effectuer des modifications post-traductionnelles complètes sur les protéines produites et les sécréter efficacement. Les rendements de production sont cependant souvent plus faibles que ceux obtenus à partir des autres systèmes, et nécessitent par conséquent beaucoup plus de ressources afin d'obtenir des quantités de protéines purifiées convenables.

Par contre, les cellules de mammifère sont caractérisées par un rythme de croissance plutôt lent (temps de génération d'environ 20 heures comparativement à 20 minutes pour les bactéries), une consommation rapide des nutriments du milieu, une production élevée en métabolites toxiques (lactate, ammonium, etc.) contribuant à altérer la culture et la qualité de la production protéique, et elles entraînent des coûts de culture plus élevés. En bref, ce système assure une bonne qualité du produit, mais le rendement protéique total est bas comparativement aux autres systèmes. (Altamirano *et al*, 2013, Berlec et Strukelj, 2013). Les nombreux développements et améliorations constantes de ce système d'expression le rendent toutefois toujours plus intéressant pour l'industrie biopharmaceutique (Pham *et al*, 2006). En effet, les avancées des méthodes de culture, des bioprocédés et de l'ingénierie métabolique des cellules de mammifères contribuent à augmenter les rendements protéiques.

On peut distinguer deux grands types d'expression :

- Expression transitoire: le vecteur d'expression transfecté reste présent transitoirement dans les cellules: à chaque division cellulaire, les cellules filles perdent le plasmide et donc le niveau d'expression diminue progressivement.
- Transfection stable: le vecteur d'expression transfecté reste présent continuellement dans les cellules transfectées. Il est transmis fidèlement à la descendance et en général s'intègre dans le génome. Les cellules sont donc génétiquement modifiées. Idéalement, le niveau d'expression à partir du plasmide est constant dans le temps.

5.1. Expression transitoire

Vecteurs d'expression et Véhicules de transfection

À l'échelle moléculaire, l'ADNc codant pour la protéine d'intérêt est inséré dans un plasmide de quelques kilobases qu'on désigne «vecteur d'expression». Ce dernier possède habituellement une origine de réplication, des cassettes pour effectuer sélection en bactérie et/ou en cellules eucaryotes, un promoteur fort placé en amont du gène d'intérêt à transcrire, des séquences *enhancers* et des éléments impliqués au niveau post-transcriptionnel (introns, signal de polyadénylation, 5'«UTR» comprenant la séquence de Kozak, 3'«UTR», etc.). Ces derniers augmentent la stabilité, facilitent le transport des ARNm produits, et augmentent le taux d'initiation de la traduction, ce qui a un effet positif sur le niveau de production. Bien qu'une multitude de combinaisons d'éléments soient possibles dans la conception d'un vecteur d'expression, il n'y a pas de vecteur universel qui soit optimal pour tous les procédés de production. Il faut donc s'assurer de la qualité du vecteur qu'on utilise selon les conditions de la situation.

Une fois conçu, le vecteur d'expression est par la suite transfecté dans un système d'expression hôte, ce dernier permettant la transcription de l'ADNc sur le vecteur, la traduction des transcrits générés et l'expression de la protéine d'intérêt.

La transfection d'ADN nu dans des cellules est grandement inefficace. Pour remédier à ce problème, il existe plusieurs agents de transfection utilisés en biotechnologie pour assurer le transport des transgènes à l'intérieur de la cellule et du noyau, notamment les vecteurs viraux et les vecteurs non-viraux.

- Les vecteurs viraux sont des particules virales modifiées par génie génétique afin d'empêcher toute réplication virale dans les cellules hôtes. Ils profitent toutefois des millions d'années d'évolution qui ont permis aux virus d'acquérir des processus très efficaces d'entrée et de livraison d'acides nucléiques dans les cellules. Parmi les virus utilisés, on retrouve entre autres les adénovirus, les baculovirus, les lentivirus, etc. Les principaux inconvénients à l'utilisation des vecteurs viraux sont la probabilité de les voir redevenir compétents pour la réplication, la complexité et la durée des manipulations qui leur sont associées, et les coûts associés à leur production. Une fois traversés les membranes cytoplasmique et nucléaire, les vecteurs d'expression portés par les vecteurs viraux décapsidés sont pris en charge par la machinerie transcriptionnelle de la cellule.

- Les méthodes de transfection non virales utilisent des vésicules naturelles ou synthétiques pour compacter, protéger et livrer les acides nucléiques dans la cellule. Par opposition aux vecteurs viraux, celles-ci sont moins coûteuses, plus simples à manipuler et offrent une meilleure stabilité dans la cellule. L'efficacité du transfert d'ADN est toutefois plus limitée. Les méthodes comprennent entre autres le traitement au phosphate de calcium (CaPi), l'électroporation, la lipofection et le transfert via la liaison de l'ADN à des polymères peptidiques. Après leur pénétration au travers de la membrane cytoplasmique par endocytose, les véhicules non viraux se libèrent du transport endocytaire et sont transportés au noyau, où ils relâchent les vecteurs d'expression. L'agent de transfection majoritairement utilisé dans les cellules de mammifère est un polymère polycationique : le polyéthylèneimine (PEI).

Les marqueurs de sélection

Un des principes de base des techniques de biologie moléculaire est d'utiliser un marqueur biologique pour identifier ou sélectionner les cellules contenant des molécules d'ADN recombinant. Le gène de sélection doit lui-même être en aval d'un promoteur fonctionnant dans la cellule transfectée.

Une première technique consiste à compléter une déficience de la cellule. Exemple de la thymidine kinase. Cette enzyme est codée par le gène *tk* qui a été isolé chez le virus de l'herpès simplex (HSV). Elle est impliquée dans une des voies de synthèse de la thymidine monophosphate à partir de la thymidine et donc de la thymidine triphosphate, nucléotide indispensable pour la synthèse de l'ADN et donc pour la réplication et la viabilité de la cellule.

Une deuxième méthode consiste à utiliser des enzymes résistants à une drogue comme par exemple :

- Une dihydrofolate réductase (DHFR) résistante au méthotrexate qui peut alors être ajouté au milieu de culture.
- Une hygromycine phosphotransférase (HPH) inactive l'hygromycine qui inhibe aussi la synthèse protéique.
- Une aminoglycosyl phosphotransférase (APH) donne la résistance au G418, proche de la néomycine. Le G418 bloque la synthèse protéique et est détruit par l'APH. C'est le marqueur le plus utilisé car le gène bactérien de résistance à la kanamycine procure la résistance au G418 chez les eucaryotes. Dans ce cas, le même gène marqueur peut être

utilisé pour les étapes de construction dans *E. coli* comme pour les étapes d'expression dans les cellules eucaryotes.

Une autre méthode consiste à exprimer une protéine qui colore les cellules. On utilise le plus souvent la GFP (green fluorescent protein) qui colore les cellules en vert.

Les promoteurs

Les promoteurs et enhancers consistent en des séquences qui interagissent avec les protéines impliquées dans la transcription. La combinaison de ces séquences détermine l'efficacité avec laquelle un gène donné est transcrit dans une cellule donnée. On distingue grossièrement deux types de séquence dans un promoteur, la "TATA box" et les éléments en amont. La "TATA box" est localisée 25 à 30 bp en amont du site d'initiation de la transcription. Ce site est directement impliqué dans le positionnement de l'ARN polymérase II qui commencera la synthèse au bon site.

5.2. Pools stables

La méthode la plus répandue pour la production de protéines recombinantes dans les cellules de mammifères consiste à intégrer dans le génome, de manière stable, le transgène codant pour la protéine d'intérêt. Ceci est possible grâce à l'action de nucléases qui linéarisent le vecteur d'expression contenant le transgène, puis de recombinaisons qui l'insèrent aléatoirement dans le génome. Le vecteur d'expression portant le transgène d'intérêt doit contenir un marqueur de sélection, c'est-à-dire un gène de résistance à un antibiotique (puromycine, G418, etc.) ou à un inhibiteur enzymatique (méthionine sulfoximine, méthotrexate, etc.) afin de sélectionner les cellules dans lesquelles ces événements se sont produits. La population de cellules générée, appelée pool stable, peut donc répliquer le transgène et exprimer la protéine d'intérêt sur de nombreuses générations cellulaires. La population d'un pool stable est cependant très hétérogène en termes d'expression de la protéine d'intérêt. En effet, certaines cellules, lors de la transfection, auront intégré très peu de copies du transgène dans leur génome alors que d'autres en auront intégré plusieurs. Une expression élevée est plus probable chez les cellules qui ont intégré plusieurs copies du transgène. En effet, les cellules qui survivent à des concentrations croissantes de l'agent de sélection sont celles où des événements de duplication et de recombinaison du gène conférant la résistance sont survenus.

6. Préparation de protéines recombinantes à partir des animaux transgéniques

L'utilisation des protéines comme médicament a commencé avec l'administration d'insuline de porc aux diabétiques. D'autres protéines également obtenues par extraction ont suivi. Le passage aux protéines recombinantes a eu lieu au début des années 1980 avec la préparation d'insuline humaine par des bactéries. Ce succès a été suivi d'autres mais il est vite apparu que les bactéries étaient incapables de procéder à certaines des modifications post-traductionnelles qui sont indispensables pour que nombre de protéines soient biologiquement actives. C'est le cas en particulier des glycosylations. Il a donc fallu avoir recours à des cellules animales et même à celles de mammifères pour obtenir certaines protéines. Ce sont des cellules en culture qui ont été tout d'abord sollicitées. Dès 1987, il a pu être montré que les animaux transgéniques, et plus précisément leur lait, pouvaient être une source industrielle de protéines thérapeutiques humaines biologiquement actives. Quelques entreprises dédiées à cette nouvelle branche de l'industrie pharmaceutique ont alors été créées et, en 2006, la première protéine thérapeutique extraite du lait de chèvres transgéniques a reçu l'autorisation de mise sur le marché par l'EMA (*European Agency for the Evaluation of Medicinal Products*).

6.1. Avantages et inconvénients du système

Avantages

- Modifications post-traductionnelles très proche de la protéine native.
- Sécrétion de la protéine recombinante dans les fluides biologiques (lait, sérum et les urines).
- Purification rapide.
- Production de grosses protéines recombinantes complexes (protéines humaines, anticorps etc...)

Inconvénients

- Coût élevé de la production.
- Difficulté d'obtenir des animaux transgéniques.
- Problèmes bioéthiques.

6.2. Systèmes de production de protéines recombinantes par les animaux transgéniques.

Les avantages comparatifs des animaux sont résumés dans le *tableau 5*. Le sang peut être une source de protéines recombinantes mais seulement pour celles qui sont stables et

n'exercent pas d'actions biologiques délétères chez les animaux producteurs. Les larves de drosophile se présentent comme un système peu coûteux mais produisant des protéines imparfaitement glycosylées car provenant de cellules d'insectes. Plusieurs animaux ont été et sont encore sollicités à cet effet. La souris est utilisée pour améliorer les vecteurs d'expression des constructions de gènes et les valider. Le lapin, le porc, la chèvre, le mouton et la vache sont effectivement exploités dans le but de produire industriellement des protéines dans leur lait. Chez ces différents mammifères, les quantités de protéines recombinantes sont de l'ordre de quelques centaines de milligrammes à quelques grammes par litre de lait. Ceci place d'emblée le coût brut de production des protéines dans le lait à un niveau 5 à 100 fois plus faible que celui actuellement en vigueur pour les cellules CHO. Le lait est le système animal de production le plus avancé techniquement et probablement le plus performant. La glande mammaire synthétise en effet des quantités considérables de protéines qui sont sécrétées en dehors du corps de la femelle réduisant ainsi leurs effets intempestifs potentiels chez l'animal. Chacune de ces espèces présente ses avantages et ses inconvénients.

Le lapin est de plus en plus apprécié car la transgénèse par micro-injection de gène est relativement aisée chez cette espèce. Une durée de 7 mois entre la micro-injection de gène dans les embryons est suffisante pour obtenir les premiers laits contenant la protéine recombinante.

De plus, le lapin se reproduit rapidement et les lignées d'intérêt peuvent être conservées sous forme d'embryons ou de semence congelés offrant ainsi une grande souplesse d'exploitation. Le lapin peut, aisément et à des coûts réduits, être élevé dans des animaleries sans germe.

Tableau 5. Comparaison des avantages et des limites des différents systèmes de production de protéines recombinantes par les animaux transgéniques.

Systèmes de production	Sang	Lait	Blanc d'œuf	Plasma séminal	Urine	Glande séricigène	Larve de drosophile
Niveau de production	+++++	+++++	+++++	++	++	++	++
Investissement	+++	+++	+++	+	+	+++	+++
Coût de production	++++	++++	++++	++	+	+++++	++++
Souplesse	+++++	+++++	+++++	++	+	+++++	++++
Conservation des lignées	+++++	+++++	+++++	+++++	+++++	+++++	+++++
Délais pour la première production	+++	++ (+)	+++	++	+	++++	++++
Augmentation d'échelle	++++	++++	++++	++++	+	++++	+++
Collecte	+++++	++++	+++++	+++	+++	+++++	+++++
Effet sur les animaux	++	+++	+++ (+)	+++ (+)	+++ (+)	++++	++++
Modifications post-traductionnelles	+++++	++++	+++ (+)	+++ (+)	+++ (+)	++ (+)	++ (+)
Glycosylation	++++(+)	++++	+++	+++ (+)	+++ (+)	++	++
Purification	++	+++	+++	++ (+)	++ (+)	+++	++ (+)
Contamination par des pathogènes	++	+++	+++	+++	++	++++	++++
Dissémination dans l'environnement	+++++	+++++	+++++	+++++	+++++	+++++	+++++
Propriété intellectuelle	++++	+++	+++	+++	+++	+++	+++
Considérations bioéthiques	+++	+++	+++	+++	+++	+++++	+++++

Le lapin est par ailleurs naturellement résistant aux maladies à prions et il ne transmet pas de maladies majeures à l'homme. Chaque lapine peut fournir 15 litres de lait par an par une traite mécanique qui ne pose pas de problème particulier.

Un système de production de protéines qui était attendu est le blanc d'œuf de poule. Ce système possède en effet la plupart des avantages du lait. Ce procédé ne s'est imposé que très récemment car l'obtention de poules transgéniques s'est heurtée pendant 15 ans à des problèmes techniques difficilement maîtrisables.

6.3. Les techniques de transgénèse

La transgénèse chez les animaux est plus laborieuse que chez les plantes et ceci est dû essentiellement aux techniques de reproduction des deux catégories d'organismes vivants.

A. Micro-injection dans le pronucléus

- Injecter ADN dans le noyau d'un œuf fertilisé.
- Implanter l'œuf dans une femelle porteuse.
- Cribler les animaux pour la présence d'ADN étrangers.
- Établir par croisements les nouvelles lignées transgéniques (Figure 8 et 9).

Inconvénients

- Faible efficacité (<5%) pour 1000 œufs injectés on obtiendra 30-50 animaux transgéniques.
- 66% des œufs survivent à l'injection.
- 25% des œufs se développent.
- 25% des descendants sont transgéniques.
- Dans beaucoup de cas ce transgène ne s'exprime pas.
- Le transgène s'intègre au hasard.
- Expression très variable.
- Pas de gène de sélection requis.

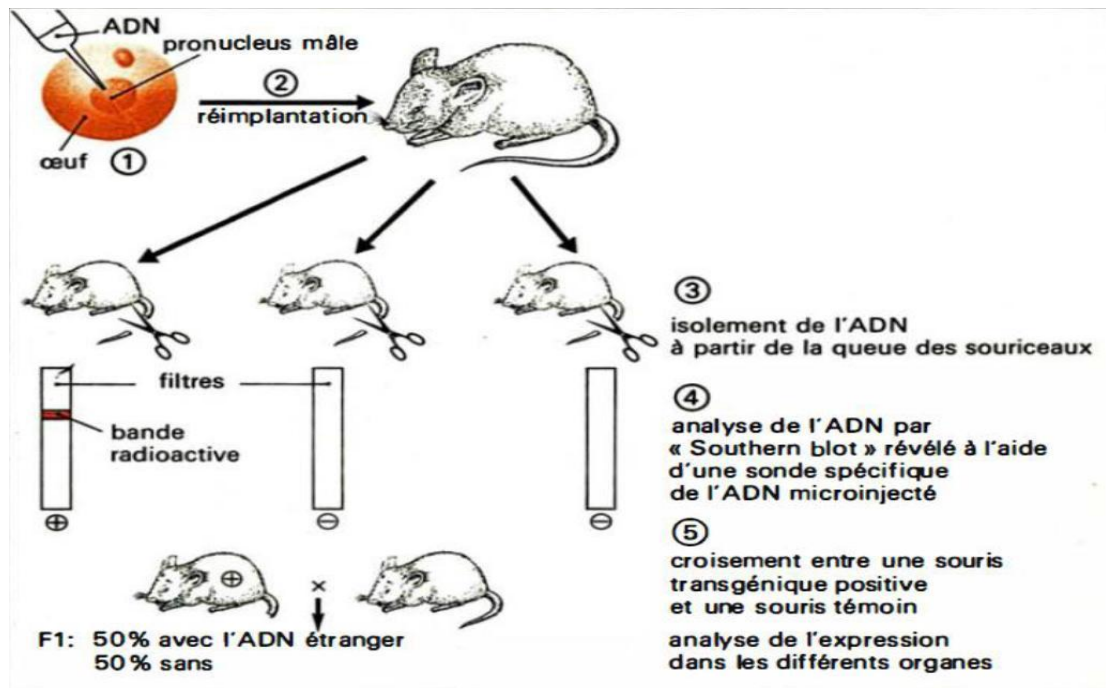


Figure 8. La micro-injection de l'ADN étranger dans le pronucléus.

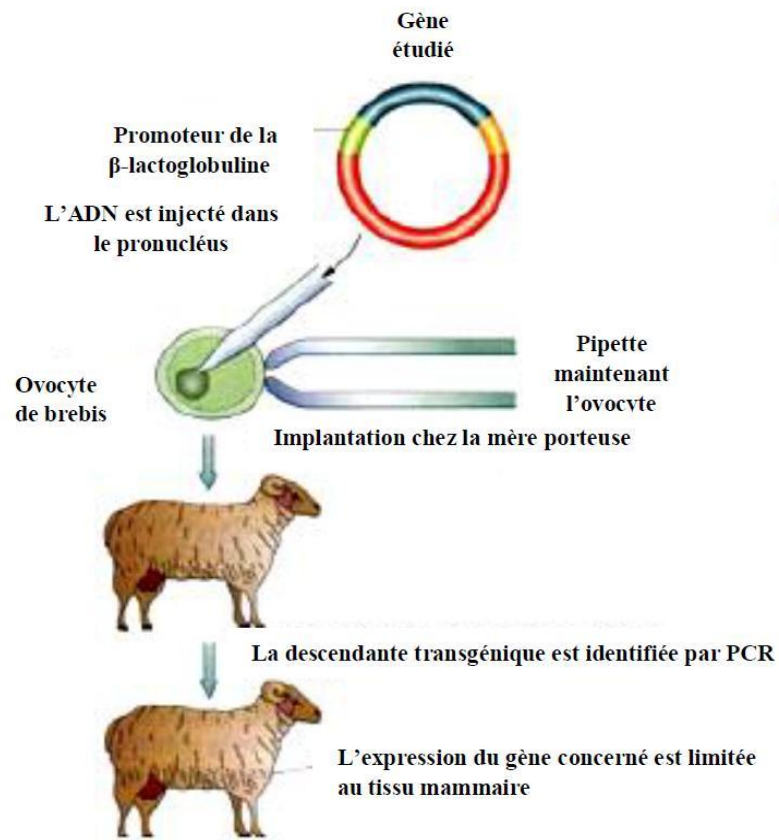


Figure 9. Microinjection chez les brebis.

B. Transfection de cellule embryonnaire souches (ES)

✚ Capacité à participer à tous les tissus.

✚ Il est possible de les cultiver en continu.

- Introduction de l'ADN par transfection ou infection virale.
- Sélection par marqueurs usuels (G418 par exemple).
- Propice à la recombinaison homologue.

- Cellules ES recombinantes introduites dans les blastocytes ----->embryon chimérique.

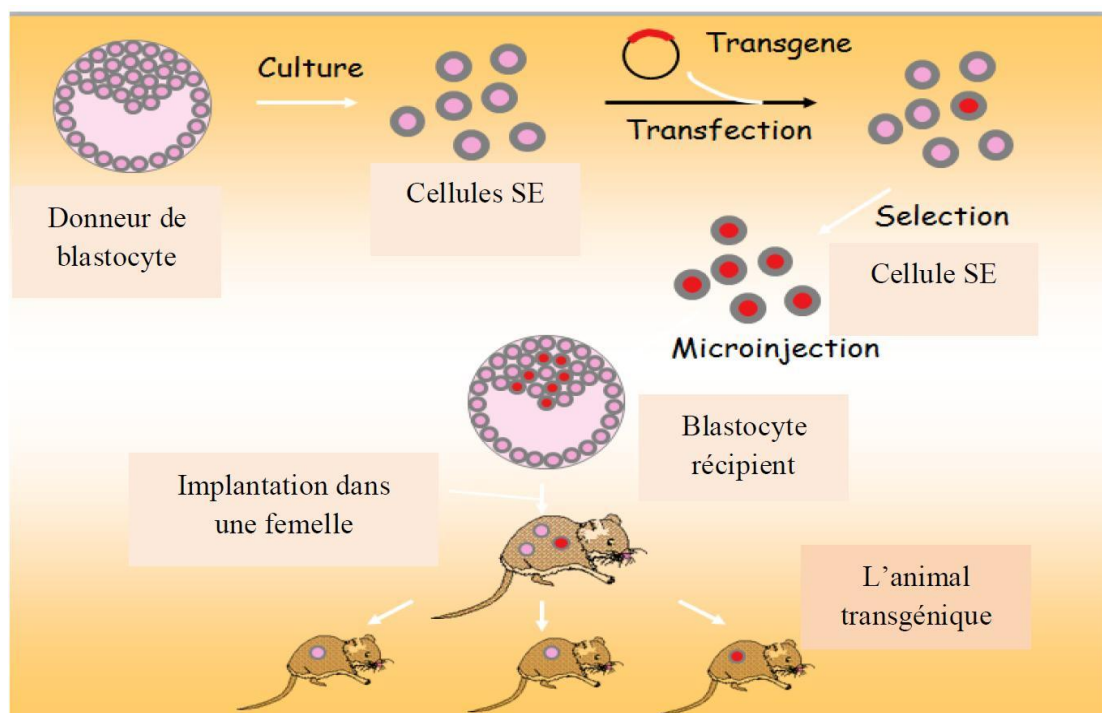


Figure 10. La transfection dans des cellules embryonnaires souches.

7. La production de protéines recombinante dans les plantes

Les plantes répondent à plusieurs des critères recherchés dans les nouvelles usines cellulaires: elles ont une machinerie cellulaire complexe et sophistiquée ; elles ne sont pas porteuses des agents pathogènes couramment associés aux infections humaines ; leur contenu cellulaire, hormis la protéine recombinante, est connu de l'homme par son utilisation dans l'alimentation, en cosmétique et en pharmacie.

Les plantes peuvent être propagées à l'infini, leur production à grande échelle ou en conditions contrôlées fait partie du patrimoine culturel de toutes les nations et elles représentent la biomasse la moins onéreuse à produire à ce jour par unité de volume. Certaines d'entre elles peuvent être considérées comme des sources renouvelables de molécules, sans impact négatif mesurable sur l'environnement.

7.1. Les différents systèmes végétaux de production de protéines recombinantes

Dans la plupart des systèmes végétaux utilisés pour la production de protéines recombinantes à grande échelle, les cellules végétales transformées ne sont pas multipliées *in vitro* dans des bioréacteurs, mais cultivées dans des conditions qui permettent la régénération de plantes mûres. Ce sont ces plantes mûres qui constituent l'« usine de production » et, à ce titre, les usines végétales se distinguent des autres systèmes de production.

Les systèmes d'expression utilisés chez les végétaux ont été adaptés aux impératifs anatomiques et physiologiques de cette grande classe d'organismes. A l'intérieur de cette grande classe, pour chacune des espèces visées, deux groupes de systèmes végétaux ont été développés pour bénéficier des sources de biomasse les plus accessibles et les plus communes, c'est-à-dire le feuillage et la graine. Chez le tabac, la luzerne et quelques autres espèces, le feuillage abondant est la cible de l'expression du transgène. Chez le maïs, le colza, le carthame, le soja et le riz, les vecteurs d'expression stimulent la production et l'accumulation des protéines recombinantes dans la graine. Chacune de ces stratégies a ses avantages et ses inconvénients, et aucune d'entre elles ne semble convenir à l'expression de toutes les protéines visées. Les feuilles ont un métabolisme actif et complexe qui offre beaucoup de possibilités, mais elles ont aussi une activité protéasique importante qui limite l'accumulation de certaines protéines. Les graines présentent l'avantage d'avoir un contenu en eau moins élevé et offrent donc un milieu d'accumulation plus stable. Les grains de maïs transgéniques sont utilisés par exemple pour la production de lipase gastrique. Aussi les protéines recombinantes peuvent être stockées 2 à 3 ans sans que la protéine recombinante

ne s'altère. En revanche, elles ne sont pas adaptées à la synthèse de certaines protéines complexes, et la nécessité d'atteindre la floraison peut représenter un danger accru de dispersion du transgène.

D'autres compartiments cellulaires ont été ciblés pour la production de protéines chez les plantes. L'expression peut se faire également dans un fruit comme la banane ou la tomate, permettant la production par exemple de vaccins oraux.

L'expression des protéines recombinantes dans les chloroplastes offre quelques avantages, notamment des rendements très élevés. Le transgène est introduit dans les organites par bombardement de particules et est intégré directement dans le génome du chloroplaste par recombinaison homologue. Un chloroplaste possède son propre génome, sous forme d'ADN circulaire (200Kb), et chaque chloroplaste contient environ 100 copies identiques du génome. Comme une cellule contient une centaine de chloroplastes, il en résulte qu'elle peut renfermer environ 10 000 copies du transgène. Cette transformation stable des chloroplastes permet donc l'accumulation de grandes quantités de protéines. Bénéficiant d'une faible activité protéolytique, une protéine exprimée dans cette organelle peut représenter plus de 20% des protéines totales de la feuille.

Dans certains cas, quand la concentration en protéines recombinantes exprimées dans le chloroplaste est trop élevée, des corps d'inclusion contenant les protéines peuvent se former, simplifiant ainsi la purification. Mais les chloroplastes n'ont pas la capacité de glycosyler les protéines, et semblent uniquement adaptés à la production de molécules simples. A noter qu'une voie alternative permettant l'adressage de glycoprotéines dans le chloroplaste (comme compartiment de stockage) par la voie sécrétoire (RE, Golgi) a été décrite chez *Arabidopsis thaliana*.

Dans tout système de production hétérologue, la molécule recombinante doit être extraite et purifiée à partir de l'ensemble des protéines endogènes de l'organisme. Pour chaque système végétal faisant l'objet de développement commercial, l'enjeu de la récupération est au centre même de la rentabilisation du procédé. La purification d'une protéine recombinante compte en effet pour plus de 80 % de ses coûts de production.

7.2. Méthodes de transformation

Différentes techniques sont employées afin de transférer un gène d'intérêt dans le patrimoine génétique d'une plante : vecteurs bactériens, biolistique, électroporation du protoplaste, la lipotransfection et injection de cellules embryonnaires totipotentes. Chaque méthode a ses avantages et inconvénients, et le choix d'une des deux méthodes dépend

d'une combinaison de facteurs, notamment l'espèce hôte choisie, et l'expérience personnelle.

7.2.1. à l'aide d'*Agrobacterium tumefaciens*

Agrobacterium tumefaciens est une bactérie en forme de bâtonnet, de la famille des Rhizobiacées. Elle se développe dans le sol. Elle est attirée par des composés phénoliques dégagés par les plantes dicotylédones lorsqu'elles sont blessées. Au niveau de cette blessure, *Agrobacterium* est capable de se fixer sur les cellules du végétal. À la suite de ce contact, ces cellules végétales se multiplient de manière importante, donnant naissance à une formation tumorale. Elle est en général située au niveau du collet, d'où le nom de cette formation : la galle du collet (crown gall).

La bactérie *Agrobacterium tumefaciens* apporte une méthode simple pour la plupart des espèces Dicotylédones et est généralement utilisée en agriculture moléculaire avec le tabac, la luzerne, le pois, la tomate et la pomme de terre.

Cette bactérie renferme un plasmide, appelé plasmide Ti, qui s'intègre en partie dans l'ADN de la cellule de la plante au moment de l'infection. Cette bactérie, capable d'insérer du nouvel ADN dans la cellule de la plante hôte est un manipulateur génétique naturel. Les biotechnologues modifient le plasmide Ti pour le doter du gène qu'ils souhaitent intégrer dans la cellule de la plante (figure). Le plasmide est ensuite réintroduit dans la cellule bactérienne. Lorsque cette bactérie infecte une cellule de plante, le nouveau gène s'intègre dans le génome de la cellule de la plante. La cellule modifiée de la plante peut ensuite se transformer en une plante complète, qui contiendra le nouvel ADN dans toutes ses cellules.

D'autres méthodes ont été mises au point pour introduire des gènes dans les types de plantes qui, généralement, ne sont pas infectées par l'agrobactérie. Ces méthodes sont décrites ci-après.

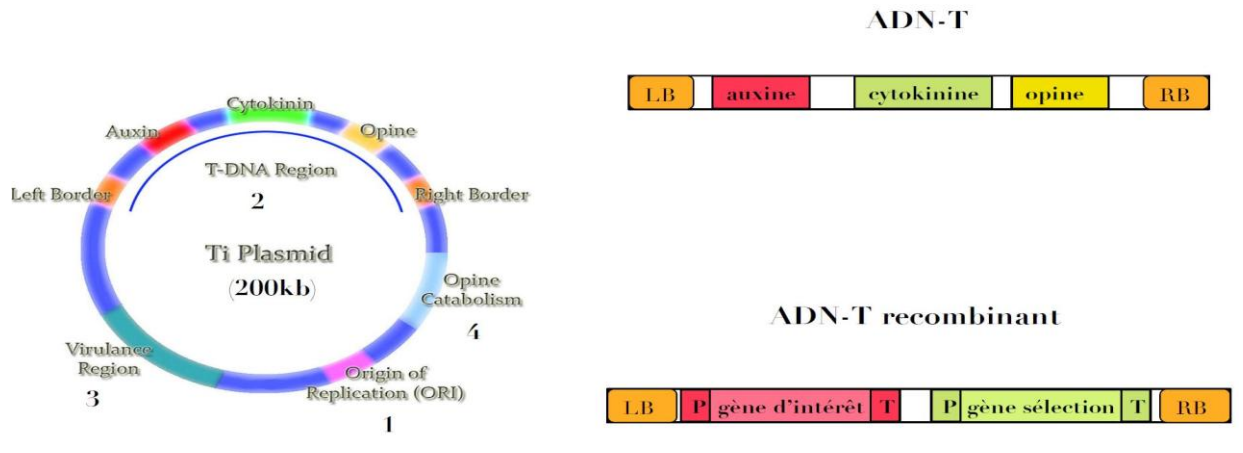


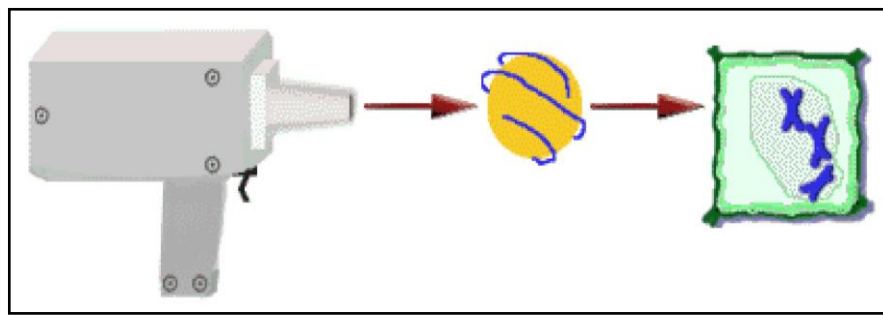
Figure 11. Structure de plasmide Ti, ADN-T et ADN-T recombinant

7.2.2. à l'aide de la biolistique

On appelle parfois " **fusil génétique** " le **Système biolistique d'introduction des particules**.

Le fusil propulse de minuscules particules d'or ou de tungsten recouvertes d'ADN dans le tissu de la plante. Les particules pénètrent dans la cellule à travers la paroi cellulaire rigide de nombreuses cellules de la plante présentes dans le tissu, et y libèrent l'ADN contenant les gènes recherchés.

Les cellules qui intègrent le nouvel ADN dans leur génome sont sélectionnées et donnent naissance à des plantes transgéniques adultes.³



7.2.3. à l'aide de l'électroporation

Normalement, l'incubation de l'ADN dans une solution comprenant des cellules de plantes ne suffit pas pour que celles-ci absorbent l'ADN, parce que chaque cellule de plante est entourée d'une membrane cellulaire et d'une paroi cellulaire, qui font obstacle à la pénétration de l'ADN. Toutefois, lorsqu'on **applique de courtes décharges électriques de**

forte intensité aux protoplastes de la cellule de la plante (cellules de plante dont on a retiré la paroi, mais qui possèdent encore leur membrane), de petits pores se forment dans la membrane cellulaire.

Ces pores sont suffisamment grands pour permettre à l'ADN de la solution de pénétrer dans la cellule. Après les décharges, les pores se referment, piégeant le nouvel ADN dans la cellule.

Un petit **nombre des cellules de la solution non seulement absorberont l'ADN, mais l'intégreront dans leur génome**. Ces cellules sont sélectionnées et donnent naissance à des plantes transgéniques exprimant le gène recherché.

7.3. Les avantages du système plante

- L'utilisation de plantes transgéniques pourrait permettre de palier à cette forte demande en protéines d'intérêt notamment en anticorps. En effet, même si les rendements sont souvent encore limités dans les systèmes d'expression végétaux, la capacité de production dans les plantes transgéniques est quasiment illimitée puisqu'elle dépend exclusivement des surfaces mises en culture. Ainsi un "bioréacteur" végétal permettra d'obtenir jusqu'à 20 kg de protéines recombinantes par hectare qu'il s'agisse de tabac, de maïs, de soja ou de luzerne. Enfin, le principal avantage des plantes transgéniques par rapport aux systèmes traditionnels de culture de cellules de mammifères en fermenteur, est que la production d'anticorps recombinants serait jusqu'à 500 fois moins coûteuse.

- Les cellules végétales étant des cellules eucaryotes (comme les cellules humaines), elles disposent d'un système permettant dans de nombreux cas de produire des protéines complexes ayant des propriétés thérapeutiques équivalentes aux protéines humaines. Elles sont en effet capables d'effectuer toutes les modifications post-traductionnelles requises pour obtenir une molécule bioactive. Ceci est illustré par leur capacité à produire des protéines de mammifères complexes comme des protéines sanguines et plasmatiques, des antigènes, des facteurs de croissance, des hormones, des cytokines, des enzymes et des anticorps. La production d'immunoglobulines dans les cellules végétales est un bon exemple : ces dernières sont capables de synthétiser correctement, et d'assembler les chaînes polypeptidiques lourdes et légères constituant un anticorps.

- Il n'existe pas, en l'état actuel des connaissances, de pathogènes végétaux capables d'infecter l'homme et l'animal, éliminant ainsi le risque par exemple d'infection ou de contamination virale par les protéines produites par les plantes, à la différence des

protéines produites par les cellules de mammifères ou d'animaux transgéniques. La sécurité biologique est donc assurée.

- Le niveau actuel des biotechnologies végétales permet de cibler de façon spécifique les tissus dans lesquels s'exprimera la protéine d'intérêt. En particulier, dans le cas du maïs, la protéine peut être ciblée dans les grains, permettant un stockage efficace et une bonne stabilité de la protéine d'intérêt.
- Enfin, la capacité de régénération et la pérennité du végétal sont des qualités qui entrent en compte pour la production de protéines recombinantes. La luzerne, par exemple, est un végétal bien adapté à la production de protéines recombinantes, en raison de la facilité de son bouturage et de l'homogénéité et de la stabilité de la descendance, le point négatif étant une possibilité accrue de persistance dans l'environnement. Ce problème de dissémination peut être solutionné par la culture des plantes en serre S2 mais sur des superficies restreintes.

7.4. Les limitations du système plante et les stratégies de contournement développées

A. La glycosylation chez les plantes

Un des inconvénients majeurs de la production de protéines recombinantes par des plantes transgéniques reste la glycosylation végétale.

En effet, si la production de protéines par les plantes est relativement aisée, leur utilisation en thérapie humaine est limitée car elles diffèrent des protéines natives humaines par leurs motifs de glycosylation. Ceux-ci les rendent fortement allergéniques pour l'Homme, et modifient leur devenir et leur biodistribution dans l'organisme. Les plantes n'ont pas le même équipement enzymatique que les cellules de mammifères pour assurer les modifications post-traductionnelles telles que les glycosylations sur les protéines.

B. Adressage des protéines vers la voie sécrétoire

Chez les plantes et les animaux, le réticulum endoplasmique permet l'entrée des protéines dans la voie sécrétoire, ce qui assure l'assemblage correct des protéines. Beaucoup de protéines recombinantes produites dans les plantes sont sécrétées dans l'espace intercellulaire ou l'apoplaste. Cet adressage dépend seulement de la présence d'un peptide signal en partie N-terminale de la protéine, clivé pendant l'insertion traductionnelle du peptide en formation dans le RE.

Il a été observé que les protéines recombinantes adressées vers la voie sécrétoire, peuvent être excrétées dans les rhizosecrétions des racines, dans des quantités plus importantes que

dans les tissus racinaires. Cette technologie, développée par l'entreprise Phytomedics, permet de supprimer l'étape d'extraction de la protéine des tissus et de simplifier le processus de purification. Récemment, des immunoglobulines ont été produites dans leur forme active dans les rhizosécrétions de tabac transgénique. Les protéines recombinantes GFP et la phosphatase alcaline placentaire humaine ont également été exprimées dans les fluides issus de la guttation de feuilles de tabac transgéniques.

C. L'activité protéolytique et la stabilité de protéines recombinantes

Les protéases présentes dans les différents compartiments de la plante peuvent altérer la stabilité des protéines recombinantes pendant leur extraction des tissus. Les protéases vacuolaires, actives à pH acide, sont très dommageables pour l'intégrité des protéines sécrétées dans ce compartiment. Comme décrit ci-dessus, des stratégies d'adressage vers les compartiments RE ou vers les chloroplastes, ayant une plus faible activité protéolytique sont développées. Récemment, il a été mis en évidence qu'il était possible de bloquer l'activité des protéases *in planta* grâce à des inhibiteurs de protéase recombinant, permettant d'augmenter la production de protéines dans les organes cibles, sans compromettre le développement de la plante hôte. Par exemple, l'expression de l'oryzacystatine I, un inhibiteur de protéinase à cystéine de riz, dans le cytoplasme de cellules foliaires de tabac transgénique, a montré une augmentation de 40% de la quantité de protéines solubles dans les feuilles.

D. L'optimisation des codons

L'optimisation des codons du gène d'intérêt peut aussi permettre d'augmenter le niveau d'expression de la protéine dans la plante de 5 à 100 fois. En effet, le code génétique étant redondant, certains codons sont utilisés préférentiellement par certains organismes, notamment par les plantes.

E. Le « gene silencing » et les supresseurs

Un autre problème se pose dans le système plante. Au cours de l'évolution, les plantes ont développé un mécanisme de défense permettant de se protéger contre les infections virales. Lorsque les ARN viraux sont présents en quantité trop importante dans une cellule, un système de régulation spécifique appelé « gene silencing » post transcriptionnel (PTGS) se met en place. Or la surexpression de gènes chez les plantes transgéniques s'apparente à la présence surnuméraire de certains ARN. En effet, ce mécanisme se met en place en présence d'ARN viral dans la plante, mais aussi en présence

d'ARN étranger, notamment sous contrôle d'un promoteur fort. Dans les plantes transformées, l'inactivation épigénétique post-transcriptionnelle correspond à la dégradation spécifique des ARN messagers (ARNm) codés par le transgène par la production massive d'ARNi doublebrin (ARN interférence) de 20-30 paires de bases complémentaires. L'ARNdb est dégradé par une nucléase de type Rnase III en petites molécules d'ARN de 21 à 23 nucléotides de long qui vont servir de guide à un complexe multi-enzymatique pour dégrader tous les ARN messagers homologues. Le transgène est alors éteint. Cette extinction posttranscriptionnelle peut être associée à un phénomène de « gene silencing » transcriptionnel (TGS) correspondant à la méthylation du gène en question, bloquant sa transcription.

Plusieurs stratégies peuvent être adoptées pour éviter ce phénomène, en exprimant par exemple simultanément au gène d'intérêt un suppresseur de silencing. Il existe en effet des suppresseurs viraux de gene silencing. Par exemple, le HC-Pro (Helper Component Proteinase) est capable de bloquer la production d'ARNi. Précédemment, un autre gène a été mis en évidence, le gène *rgs-CaM* chez le tabac. Il code pour une protéine liant le calcium capable d'inhiber le « gene silencing » et interagit avec HC-Pro dans le mécanisme de suppression du PTGS. Un des meilleurs suppresseurs caractérisé est la protéine p19 provenant du Tomato Bushy Stunt Virus (TBSV). Ce suppresseur, coexprimé avec une protéine recombinante a permis d'augmenter de 50 fois le niveau d'expression de la protéine d'intérêt.

8. Bilan

Le choix des outils pour l'expression hétérologue dépend de nombreux critères qu'il faut définir en fonction des besoins et des moyens à disposition. Le tableau 5 montre les caractéristiques générales à retenir pour chaque hôte.

Tableau 5. Principaux critères de choix d'un organisme d'expression

Caractéristiques	<i>Escherichia coli</i>	Champignons Filamenteux	Levure	Cellules d'insecte	Cellules de mammifères	Cellules de plantes
Temps de croissance	Plusieurs heures à plusieurs jours	Plusieurs jours à une semaine	Plusieurs jours à une semaine	Plusieurs jours à une semaine	Plusieurs semaines	Plusieurs mois
Coût du milieu de culture	faible à moyen	faible à moyen	faible à moyen	élevé	élevé	moyen à élevé
Taux d'expression	faible à élevé	faible à élevé	faible à élevé	faible à élevé	faible à élevé	faible
Evaluation de la méthode	facile	difficile	facile	difficile	difficile	moyen
Capacités de sécrétion	oui dans le périplasme	oui dans le milieu	oui dans le milieu	oui dans le milieu	oui dans le milieu	oui dans le milieu
Modifications post-traductionnelles						
Repliement de protéines	Repliement nécessaire	Repliement pouvant être nécessaire	Repliement pouvant être nécessaire	Repliment assuré	Repliement assuré	Repliement assuré
N-glycosylation	Non	Oui, similaire à ceux des mammifères, pas d'acide sialique, incompatible avec certaines protéines humaines	Oui, élevée en mannose, pas d'acide sialique, incompatible avec certaines protéines humaines	Oui glycosylation complexe, pas d'acide sialique, incompatible avec certaines protéines humaines	Oui glycosylation complexe, incompatible avec certaines protéines humaines	Oui glycosylation complexe, pas d'acide sialique, incompatible avec certaines protéines humaines
O-glycosylation, O-phosphorylation, O-acetylation, O-acylation	Non	Oui	Oui	Oui	Oui	Oui

III. Modulation de l'expression génique ou de l'activité protéique

La production des animaux modèles porteurs de mutations génétiques, similaires à celles retrouvées dans des pathologies humaines d'origine génétique, ainsi que la mise au point de thérapies géniques efficaces constituent deux voies de recherche d'enjeu majeur. Des études ont d'abord été réalisées grâce à l'étude de mutants, naturels ou provoqués par l'utilisation de produits chimiques ou ionisants. C'est ainsi que, par une stratégie de mutagenèse particulièrement importante. Néanmoins, ces techniques de mutagenèse aléatoire posent un gros problème : les chercheurs n'ont aucun contrôle sur les mutations qui vont avoir lieu... Ceci impose un fastidieux travail de criblage pour rechercher les mutants. De plus, les mutations obtenues ont souvent pour conséquence une diminution de l'expression du gène ou une diminution de l'efficacité de la protéine codée. Il est ainsi difficile de comprendre si l'effet observé est dû à l'absence de la protéine, à une expression incorrecte de la protéine, à la présence d'une protéine modifiée, etc.

Dans ce chapitre on va discuter quelques techniques permettant de modifier les génomes de manière fine, contrôlée et avec une grande efficacité.

1. Le Knock-out ou l'invalidation d'un gène par recombinaison homologue

1.1. Généralités

Le knock-out d'un gène (ou K.O.) signifie la perte physique de la séquence (ou d'une partie de la séquence) de fondamental d'échange de séquence d'ADN homologue au niveau chromosomique, la recombinaison homologue. Un organisme dans lequel un seul gène de choix ou d'intérêt soit inactivé par knock-out d'une manière qui laisse tous les autres gènes inchangés et la meilleure façon de délimiter la fonction d'un gène. Ces organismes porteurs de tels gènes sont connus sous le nom d'organismes Knockout ou knock-out simples, ils sont utilisés pour attribuer une fonction à un gène spécifique ayant une fonction inconnue.

Cette technique est à l'opposé de Gene knock-in. La suppression simultanée de deux gènes dans un organisme est connue sous le nom de double knockout (DKO).

1.2. Principe de la technology Knock Out

- Le knockout est accompli par une combinaison de techniques, commençant par la préparation d'un vecteur (plasmide, des chromosomes artificiels bactériens BAC ou une autre construction d'ADN);
- Souvent, le but est de créer un animal transgénique qui a le gène modifié. Pour cela, les cellules souches embryonnaires (cellules ES) sont génétiquement transformées et insérées dans des embryons précoces;
- Les animaux résultants avec le changement génétique dans les cellules de la lignée germinale peuvent alors souvent transmettre le knock-out du gène aux générations futures.

1.3. La recombinaison homologue : un mécanisme à l'origine de la technologie Knockout

Ce mécanisme moléculaire permet l'échange d'information génétique entre une séquence chromosomique endogène, la cible à modifier, et une séquence exogène portant une région homologue à la cible. La séquence exogène est choisie et introduite dans la cellule par l'expérimentateur. Comparativement, les événements de recombinaison homologue sont plus rares dans les cellules mammifères que dans les levures, et l'intégration aléatoire de séquences exogènes dans le génome des cellules mammifères est beaucoup plus fréquente que l'événement de recombinaison homologue.

La recombinaison ce gène conduisant à l'absence de l'ARN messager et donc de la protéine. Le KO est donc transmissible à la descendance. Attention cependant : chez les eucaryotes, le KO d'un gène signifie que la séquence des deux allèles est affectée, la descendance respecte les lois de Mendel (figure 1). Le KO revient à réaliser de la mutagenèse dirigée (par délétion). Ceci est rendu possible par l'existence d'un mécanisme homologue est un échange de fragments d'ADN entre deux molécules (d'ADN) au niveau des séquences nucléotidiques homologues. C'est ce qu'il se passe lors de crossing over à la méiose entre les chromatides des paires de chromosomes.

Les facteurs qui favorisent la recombinaison homologue :

- La taille de la région homologue (de 4 à 9 kb, on multiplie par 20 la RH)
- Le fort pourcentage d'homologie
- L'ADN sous forme linéaire (par rapport à circulaire)

- Le mode de transfection

Il semble en revanche que ni le nombre de copies présentes dans la cellule, ni la localisation chromosomique du gène à invalider, n'influe sur la RH.

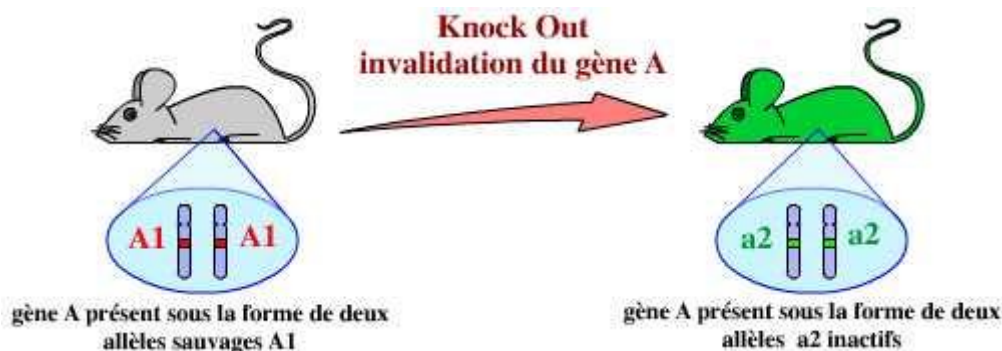


Figure 1. Principe du knock-out

1.4. La sélection des événements de recombinaison.

En tout état de cause, il est indispensable de sélectionner les événements de recombinaison. Les différents types de sélection utilisés chez les eucaryotes supérieurs reposent sur l'utilisation de gènes de résistance à des drogues telles que la néomycine (ou G418), la bléomycine, l'hygromycine, la puromycine. D'autres marqueurs de sélection sont également utilisés :

Le gène codant pour la DHFR (dihydrofolate réductase) résiste au méthotrexate.

Le gène HPRT : Hypoxanthine phosphoribosyl-transferase : Ce gène permet une double sélection, positive et négative. Les cellules possédant ce gène poussent en milieu HAT (hypoxanthine, aminoptérine, thymidine) mais sont sensibles à la 6-thioguanine, inversement l'absence de ce gène rend les cellules sensibles au milieu HAT et résistantes à la 6-thioguanine.

Le gène HSV-TK : la thymidine kinase du virus herpès simplex. Cette enzyme est capable de phosphoryler le gancyclovir qui devient ainsi toxique pour les cellules.

1.5. Les différentes constructions utilisées

a. L'interruption d'un exon

La figure 2 montre un exemple de construction simple aboutissant, après recombinaison homologue, à la perte de l'expression d'un gène. Le gène de résistance à la néomycine a été introduit dans la séquence de l'exon 3 d'un gène quelconque. Le

gène de la thymidine kinase du virus de l'herpès est introduit en aval de la séquence du gène. Les événements de recombinaison homologue sont sélectionnés par la double sélection : positive par la néomycine et négative par la thymidine kinase. Les cellules recombinées au niveau du gène sont néomycine résistantes et gancyclovir résistantes, alors que les intégrations au hasard seront néomycine résistantes et gancyclovir sensibles.

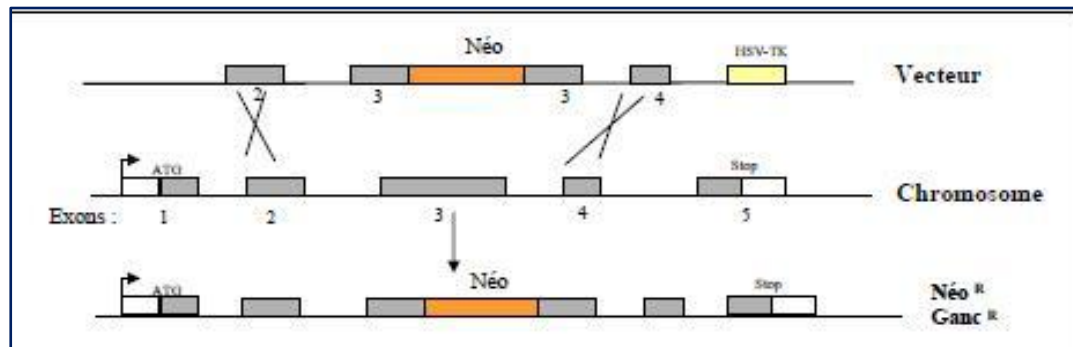


Figure 2. L'interruption d'un exon

b. Creation de mutation ponctuelle

Si maintenant le gène de résistance à la néomycine se situe dans un intron et que l'on a créé une mutation dans l'exon3 du gène, la recombinaison homologue de cette construction aboutira à l'introduction de la mutation au niveau du génome (figure 3). La présence de la cassette de résistance dans un intron n'empêche pas la traduction ni l'épissage, les modifications qu'entraîne cette construction sont théoriquement dues uniquement à la mutation. Il a cependant été montré que dans certains cas la présence d'une cassette de sélection dans un intron peut avoir un effet sur le niveau de transcription du gène. Ce qui a conduit les chercheurs à imaginer des stratégies pour éliminer le gène de sélection afin d'aboutir à des mutations « propres ».

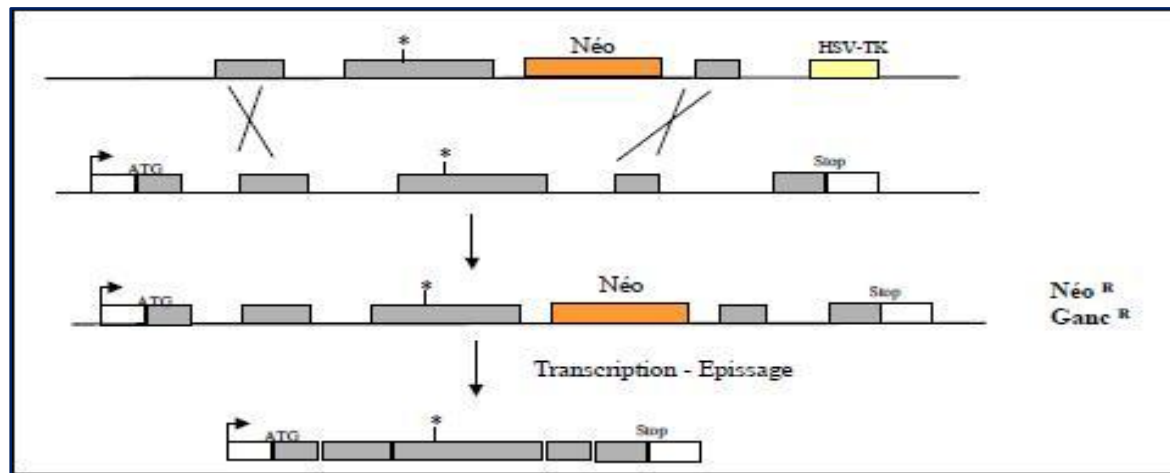


Figure 3: Creation de mutation ponctuelle

c. Mutation propre par double remplacement

Dans cet exemple on se base sur la propriété du gène HPRT permettant une double sélection (figure 4). Le premier événement de recombinaison apporte gène *hprt* dans l'intron (cellules HAT résistantes et 6-thioguaninesensibles). La seconde RH amène la mutation et élimine le gène *hprt*, les cellules deviennent HAT^S et 6-TG^R.

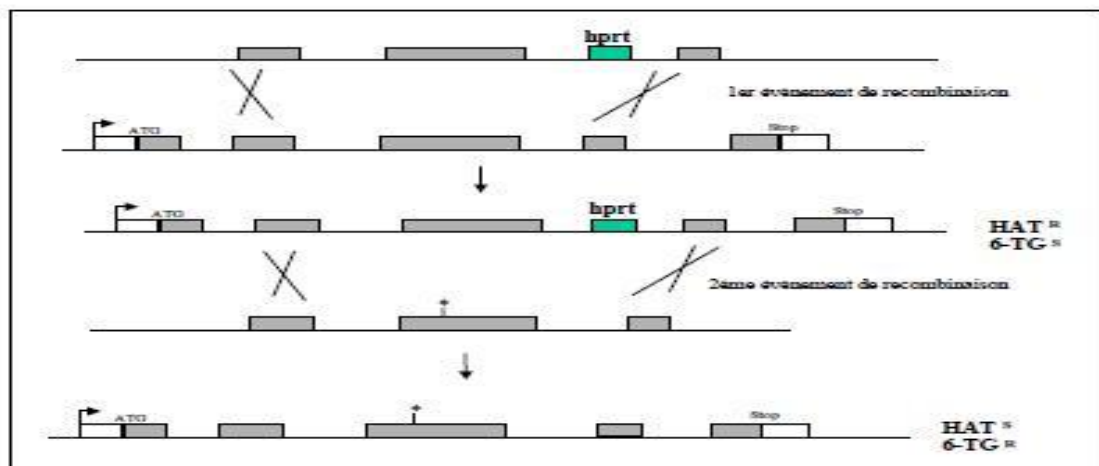


Figure 4. Mutation propre par double remplacement

1.6. Exemple de l'invalidation du gène Bmp7

Dans ce cas, le gène ciblé est Bmp7. (A) Les cellules souches embryonnaire (ES) d'un blastocyste de souris sont cultivées. (B) Les gènes Bmp7 clones sont coupés avec une enzyme de restriction, et un gène de résistance à la néomycine est inséré

(figure 5). La construction contient les gènes Bmp7 mutants sont électroporés dans des cellules ES. Dans certains des tell, la recombinaison homologue échange un gène de type sauvage par la copie mutante. Ces cellules sont sélectionnées par leur résistance à la néomycine. (C) Les cellules ES hétérozygotes sélectionnées sont insérées dans la masse cellulaire interne d'un embryon de type sauvage, et le blastocyste est implanté dans l'utérus d'une souris. La souris résultante est une chimère composée de tissus avec des gènes hétérozygotes Bmp7 et de tissus avec le gène Bmp7 de type sauvage. L'accouplement de la souris chimérique à une souris de type sauvage produit une progéniture hétérozygote Bmp7 si les cellules ES ont contribué à la lignée germinale. Les résultats de cette experimentation de Dudley et al.(1995) est représenté dans la figure 6. Ces souris hétérozygotes peuvent être reproduites ensemble, et environ 25% de leur descendance devrait être homozygote pour le mutant Bmp7.

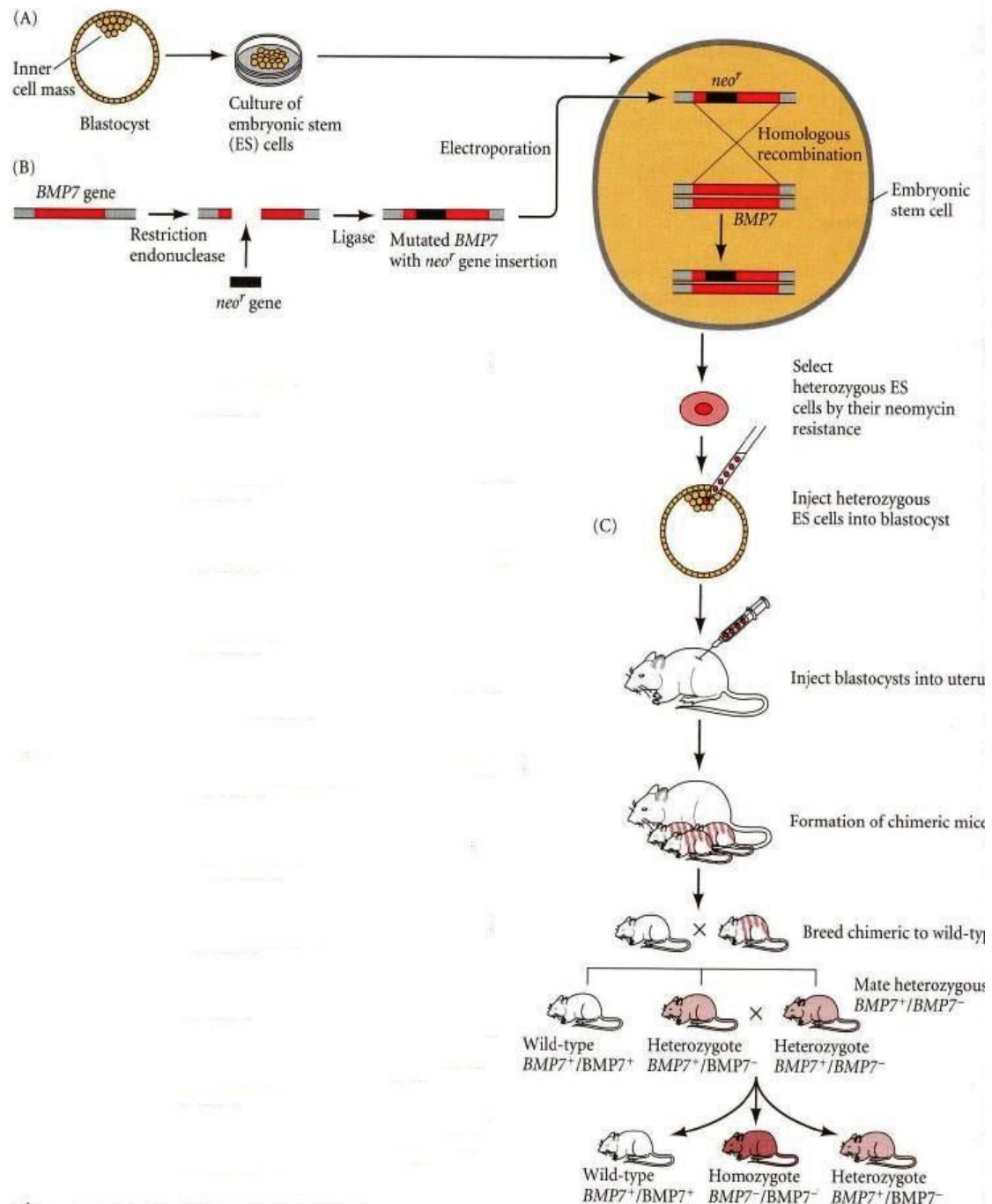


Figure 5. Technique de ciblage génétique par knock out.

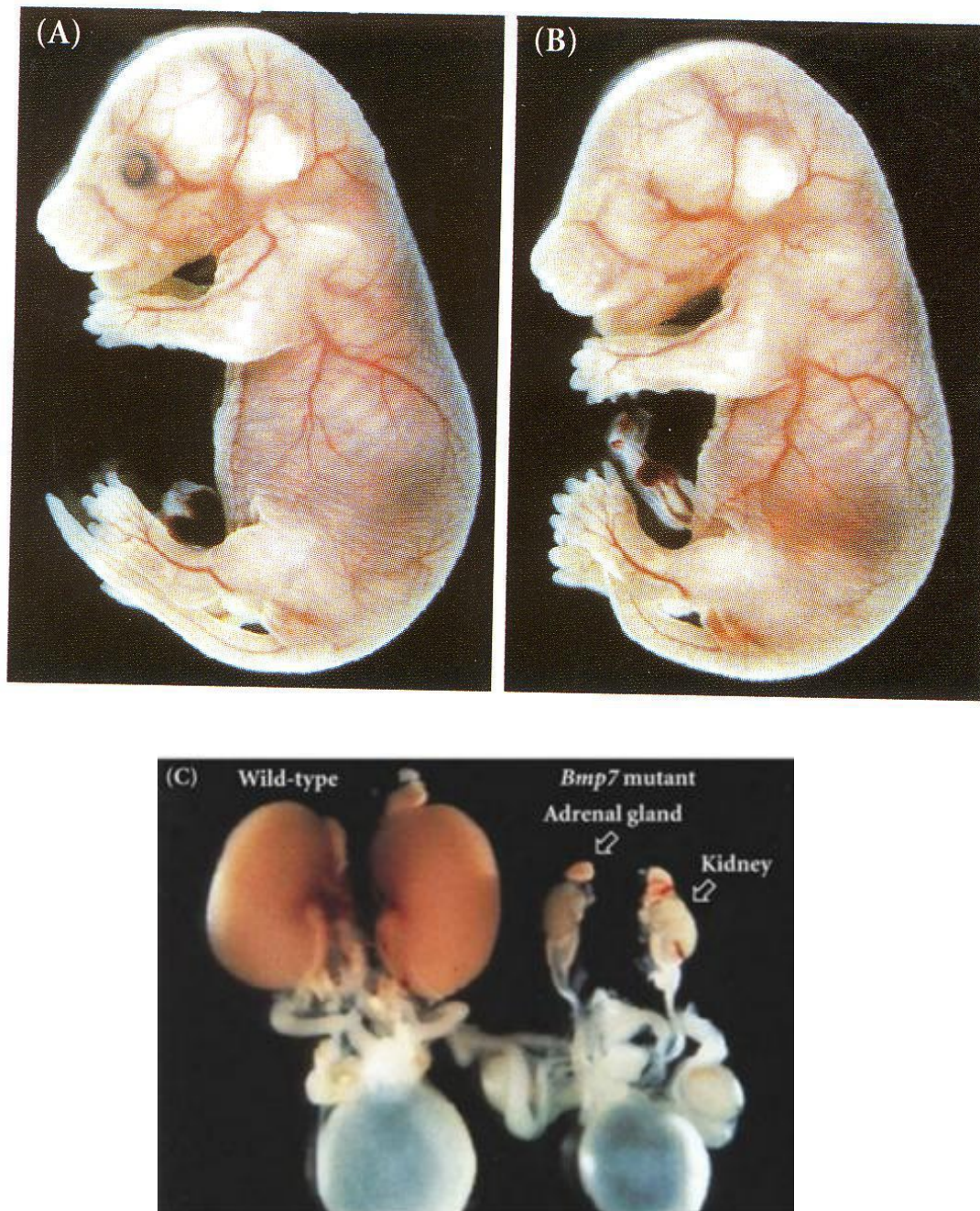


Figure 6. Analyse morphologique de souris knock-out Bmp7 (Dudley et al., 1995)

(A) Souris de type sauvage et (B) homozygote Bmp7-déficiente au jour 17 de leur gestation de 21 jours. La souris déficiente en Bmp7 n'a pas d'yeux. (C) Les reins de ces souris au jour 19 de la gestation. Le rein de la souris déficiente en Bmp7 (à droite) est sévèrement atrophié. Des coupes microscopiques révèlent la mort des cellules qui auraient autrement formé les néphrons.

2. Le système Cre/loxP

Les modèles de souris génétiquement modifiés sont généralement préférés pour étudier les maladies humaines en raison des similitudes génétiques et physiopathologiques entre les souris et les humains. En particulier, le système Cre-loxP est largement utilisé comme outil expérimental intégral pour générer des souris modèles. Ce système a permis aux chercheurs d'étudier des gènes d'intérêt dans un tissu / cellule (contrôle spatial) et / ou un temps (contrôle temporel) de manière spécifique. Diverses lignes de souris Cre-driver spécifiques aux tissus ont été générées à ce jour, et de nouvelles lignes Cre sont toujours en cours de développement.

2.1. Fonctionnement de système

La Cre est une recombinase (de la famille des intégrases) du bactériophage P1. C'est une protéine de 38 KDa qui catalyse la recombinaison entre deux site de reconnaissance, les sites loxP. loxP est une séquence d'ADN de 34 pb comprenant aux extrémités 13 nucléotides palindromiques séparées par une séquence centrale asymétrique de 8 pb (figure 7). (il est donc impossible de trouver cette séquence dans un génome eucaryote). La recombinase Cre reconnaît spécifiquement un site loxP au sein duquel elle clive l'ADN puis catalyse la réunion de cet ADN avec une séquence d'ADN clivé à un autre site loxP. Les deux sites peuvent être très éloignés l'un de l'autre et être quand même recombinaisonnés par la Cre. Cette recombinase fonctionne dans les cellules eucaryotes génétiquement modifiées comportant deux sites loxP (figure 8).

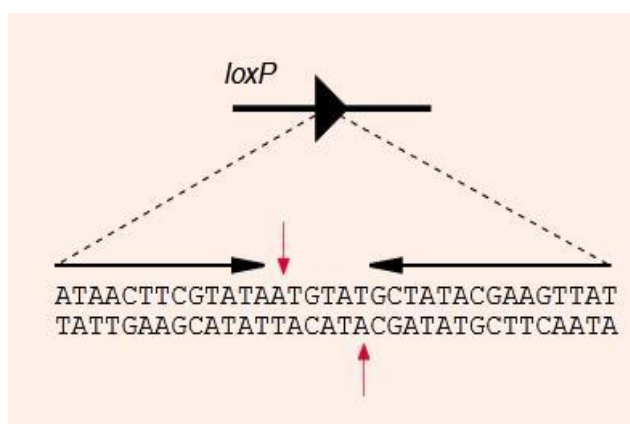


Figure 7. Structure de site Lox P

Plusieurs types d'événements de recombinaison peuvent être produits suivant que les deux sites LoxP sont portés par la même molécule d'ADN (recombinaison en cis) ou par deux molécules d'ADN différentes (recombinaison en trans) et suivant l'orientation respective des deux sites loxP (l'orientation d'un site loxP est donnée par la séquence de 8 pb non palindromique).

Recombinaison en cis

Si les deux sites loxP sont dans la même orientation, la région d'ADN située entre ces sites est délétée lors de la recombinaison. Ce type de configuration est utilisé pour créer des mutations «propres» (élimination de la cassette de sélection, voir figure 9), des mutations conditionnelles (voir figure 10) et des délétions. Si les deux sites loxP sont en orientation inverse, la recombinaison conduit à l'inversion de la région comprise entre les sites.

Recombinaison en trans

Dans le cas où un site loxP est intégré dans le génome et l'autre site est porté par un plasmide circulaire, il peut y avoir insertion des séquences portées par le plasmide au niveau du site loxP intégré. Néanmoins, l'insertion étant défavorisée par rapport à la délétion (c'est-à-dire la réaction inverse). Dans le cas où les sites loxP sont tous les deux intégrés dans le génome, la recombinaison en trans induit des remaniements chromosomiques : délétions, duplications ou translocations. Ces événements de recombinaison sont rares et doivent être sélectionnés pour être observés. Pour ce faire, il est possible d'utiliser des cassettes de sélection tronquées et non fonctionnelles hp-loxP et loxP-rt. Après recombinaison entre les sites loxP, et seulement dans ce cas, une cassette hp-loxP-rt fonctionnelle (le site loxP restant est situé dans un intron) est reconstituée, permettant de sélectionner le remaniement chromosomique désiré. Par ailleurs, l'orientation relative des sites loxP par rapport à l'axe centrotélomérique des chromosomes est importante. En effet, dans le cas d'une mauvaise orientation relative, l'événement de recombinaison aboutirait à la formation de chromosomes acentriques ou dicentriques qui, du fait de leur grande instabilité, seraient éliminés de la cellule.

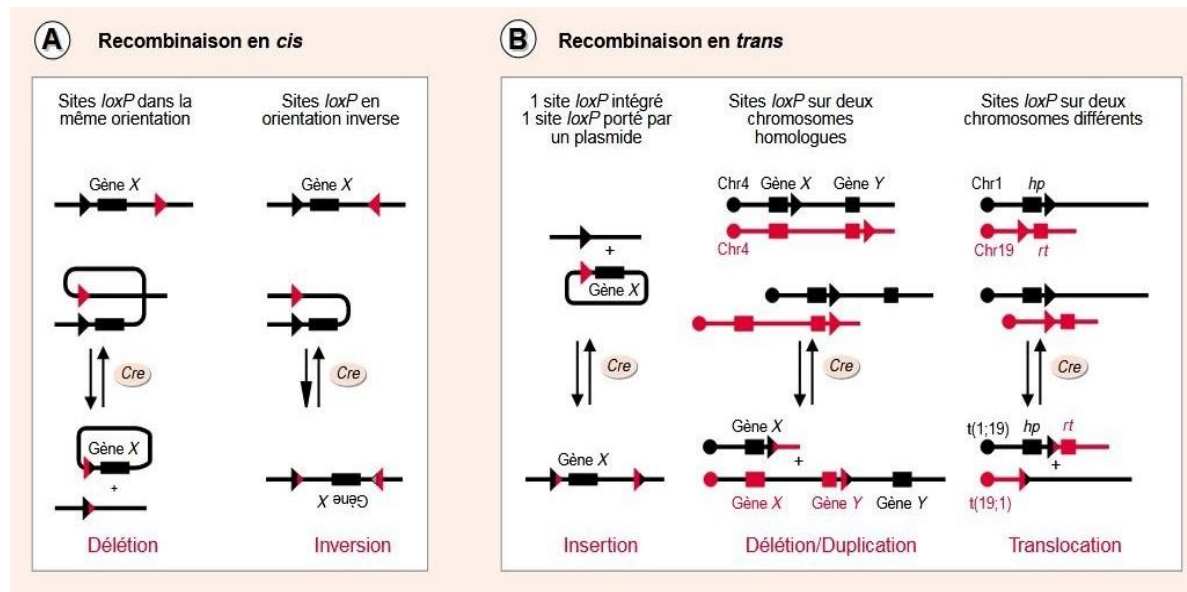


Figure 8. Le système Cre/loxP et ses applications

2.2. Exemples d'application

2.2.1. Mutations «propres»

La persistance, dans l'allèle modifié, d'une cassette de sélection avec son propre promoteur et des séquences activatrices fortes peut avoir des effets sur le locus cible et les locus avoisinants. S'agissant de la création de mutations subtiles (mutations ponctuelles, petites délétions et insertions) il est préférable d'éliminer les cassettes de sélection. Les deux stratégies les plus utilisées pour créer ce type de modification sont schématisées sur la figure 9. À gauche : la stratégie Knoch out de double remplacement. Cette approche nécessite l'utilisation d'une lignée de cellules ES *hprt*⁻, mutée pour le gène *hprt*. La première étape (A) consiste en l'introduction dans le gène cible d'une cassette d'expression pour le gène *hprt*. La sélection des cellules recombinantes (*hprt*⁺) s'effectue en présence de HAT. Dans un deuxième temps (B), ces cellules sont transfectées avec un deuxième vecteur de remplacement présentant une mutation subtile et dépourvu de cassette de sélection. L'événement de recombinaison homologue, qui aboutit à la perte de la cassette d'expression *hprt*, est sélectionné en présence de 6-TG. L'utilisation d'autres vecteurs de remplacement portant des modifications différentes permet de créer rapidement plusieurs allèles pour le même gène cible. À droite : utilisation du système Cre/loxP. Dans un premier temps (C), le gène cible est modifié à l'aide d'un vecteur de ciblage comprenant une

mutation subtile et une cassette de sélection «floxée», c'est-à-dire entourée de 2 sites loxP dans la même orientation. Dans un deuxième temps (D), l'expression transitoire de la recombinase Cre dans les cellules recombinantes induit la délétion de la cassette de sélection. En dehors de la modification subtile souhaitée, seul un site loxP de 34 pb persiste dans l'allèle modifié final. La position de ce site loxP est choisie de manière à ne pas interférer avec l'expression du gène cible (en général dans un intron).

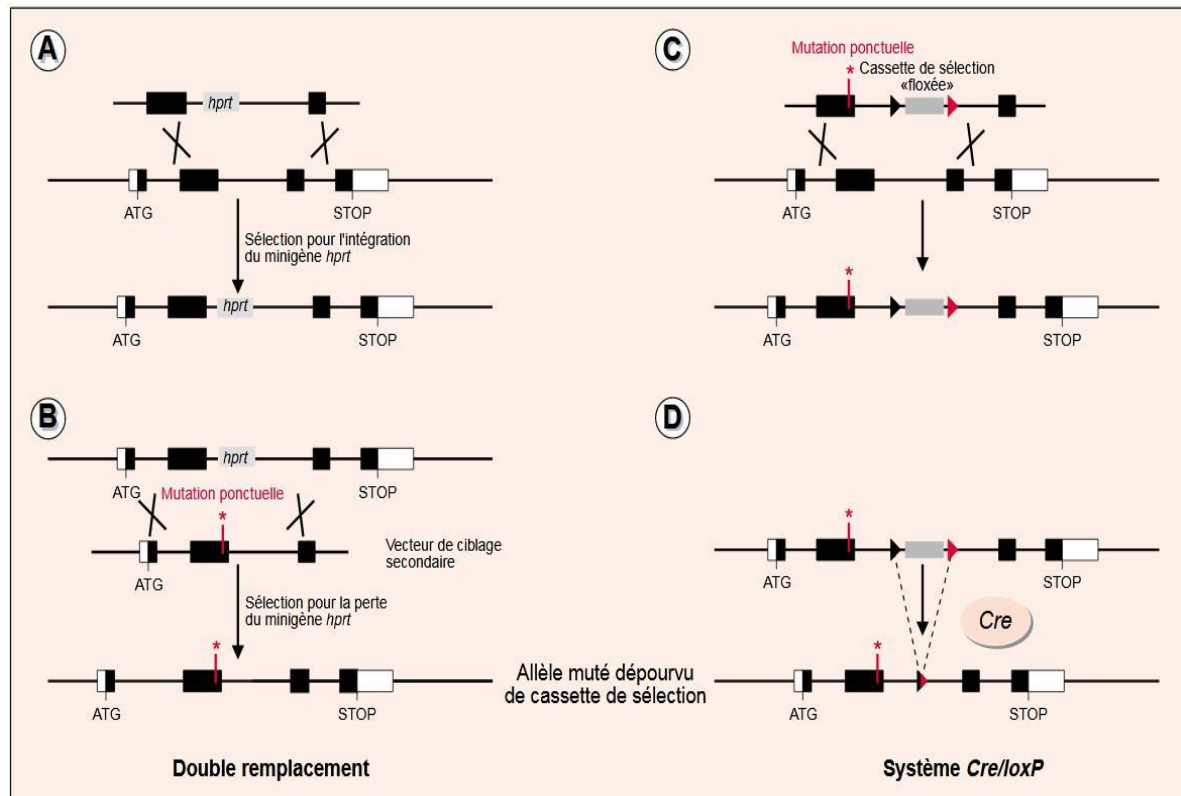


Figure 9. Mutations «propres» par double remplacement et par l'utilisation de système Cre/loxP

2.2.2. Le système Cre/loxP et la mutagenèse conditionnelle

Nous avons indiqué l'intérêt considérable pour l'analyse fonctionnelle du génome de la création programmée de souris porteuses de modifications génétiques variées (mutations nulles, mutations discrètes, remaniements chromosomiques...). Cependant, ces situations, dans lesquelles les souris mutantes portent la mutation dans toutes leurs cellules, ont aussi leurs limites notamment pour deux types de raisons : (1) dans le cas où la mutation entraîne à l'état homozygote une létalité embryonnaire (un cas extrême est celui où la mutation entraîne la mort cellulaire), il devient

impossible d'étudier la fonction éventuelle du gène, au-delà du moment où les embryons meurent et donc pendant toute la vie adulte ;

(2) un gène peut avoir un profil d'expression très large et son invalidation entraîner un phénotype complexe qui affecte de multiples tissus. Il serait alors utile, pour simplifier l'analyse, de créer des souris porteuses de la mutation seulement dans l'un ou l'autre de ces tissus.

Une stratégie repose sur les propriétés remarquables du système Cre/loxP peuvent être utilisée. Il s'agit, dans un premier temps, de créer des souris porteuses d'un allèle dans lequel deux sites loxP encadrent une partie essentielle du gène d'intérêt sans pour autant perturber son fonctionnement, en les plaçant par exemple dans les introns (on parle alors d'allèle «floxé»). S'agissant d'un gène dont les mutations nulles sont létales à l'état homozygote, on vérifiera que les souris homozygotes pour l'allèle floxé sont parfaitement viables. On les croise alors avec une souris transgénique exprimant la recombinaise dans un type cellulaire particulier grâce à l'emploi de séquences régulatrices appropriées dans le transgène (la recombinaise entraîne l'excision des séquences situées entre les sites loxP et donc induit une mutation nulle dans le type cellulaire dans lequel le transgène s'exprime). Cette stratégie est d'une très grande puissance, car elle permet non seulement de contourner le problème de la létalité embryonnaire qui se produit lorsque toutes les cellules de l'embryon portent la mutation, mais encore d'examiner l'effet de cette mutation, en principe dans n'importe quel tissu, pourvu que l'on dispose d'une lignée de souris transgéniques exprimant la protéine Cre dans le tissu en question.

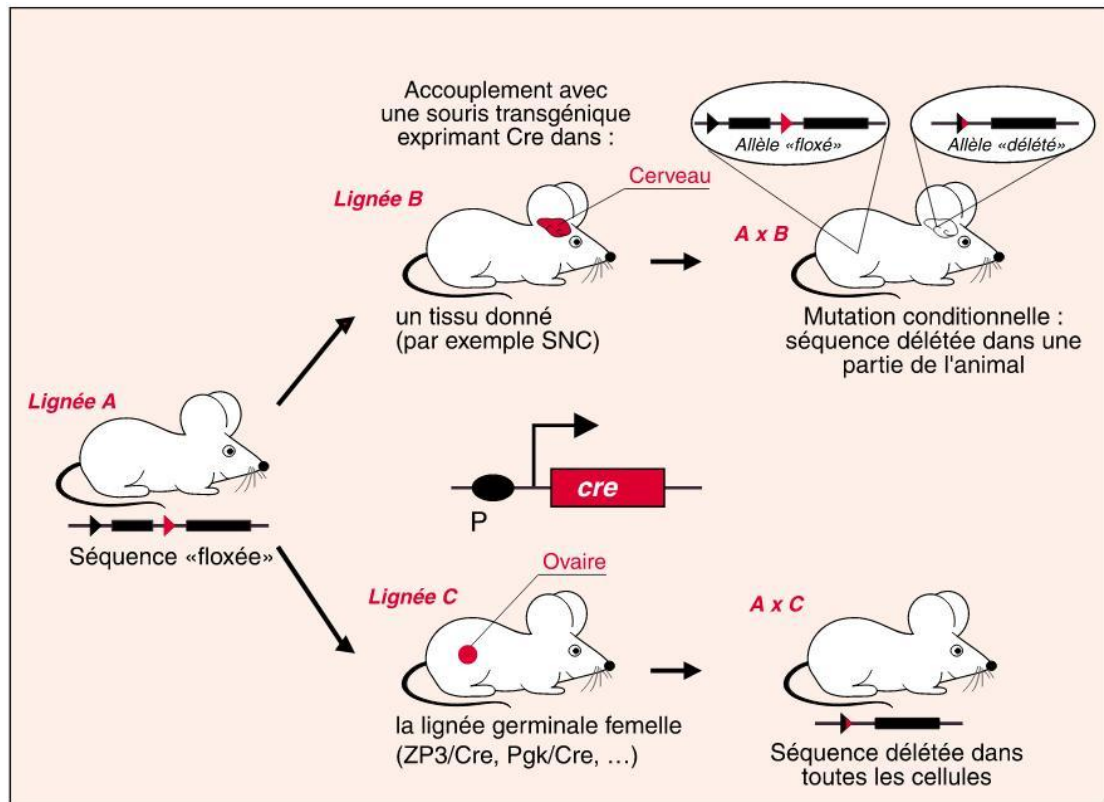


Figure 10. Le système Cre/loxP *in vivo*.

Un raffinement supplémentaire consiste à contrôler l'induction de la mutation, non seulement dans l'espace comme nous venons de le voir, mais également dans le temps. Pour ce faire, la protéine Cre est exprimée sous la forme d'une protéine de fusion avec le domaine de fixation au ligand d'un récepteur des stéroïdes (LBD); cette protéine de fusion ne présente pas d'activité Cre (figure 11-A). En revanche, en présence d'un ligand approprié, un changement de conformation se produit qui permet la restauration de l'activité Cre. Ainsi chez les souris transgéniques portant les deux allèles «floxés» du gène d'intérêt, et le transgène exprimant la protéine de fusion LBD/Cre sous la dépendance d'un promoteur spécifique d'un tissu donné, l'induction de la mutation nulle dans ce tissu peut être obtenue par l'injection, au moment voulu, du ligand approprié (figure 11-B).

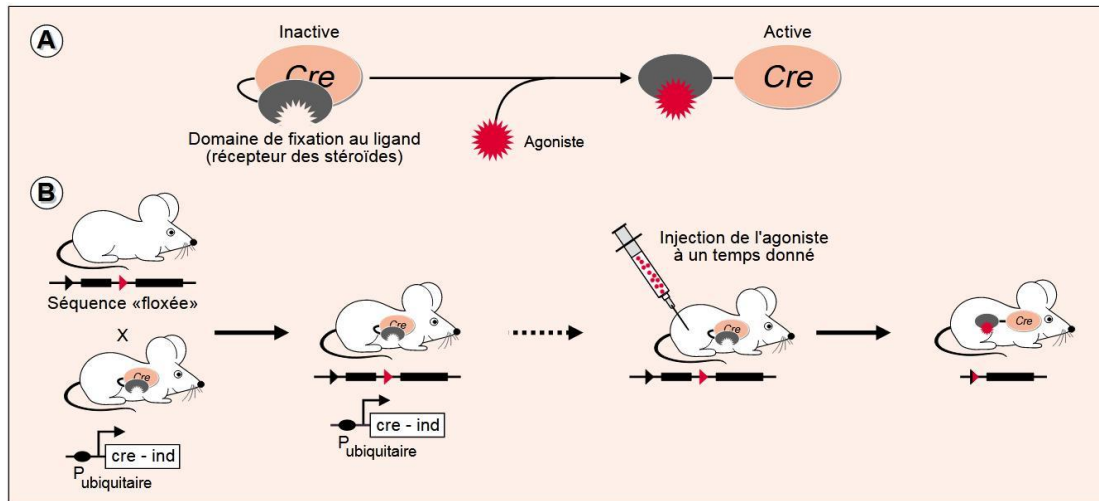


Figure 11. Système Cre/loxP et mutations inductibles.

3. Les oligonucléotides triples hélices

3.1. Fonctionnalité de système

L'idée est d'utiliser des oligonucléotides capables de former des triples hélices, nommées ainsi Triplex-forming oligonucleotides (TFOs), (figure 12) pour bloquer le site d'initiation de la transcription du double brin d'ADN, les triplex n'étant pas assez stables pour bloquer l'élongation de la transcription une fois celle-ci amorcée [187]. L'oligonucléotide 3ème brin se lie dans le grand sillon de la double hélice d'ADN par des liaisons hydrogènes de type Hoogsten ou Hoogsten inverse. Ce type de liaison ne dénature pas les liaisons de Watson-Crick, mais elles ne peuvent se produire qu'au niveau des séquences homopuriques (figure 13).

Deux modes d'action possibles :

- empêcher la fixation des facteurs de transcription
- empêcher l'ouverture de l'hélice



Figure 12. Représentation de la formation d'une triple hélice entre l'ADN et un oligonucléotide.



Figure 13. Séquences de formation de triplex dans le gène c-MYC humain.

3.2. Avantage et limitation des Applications potentielles des oligonucleotides triple hélice en thérapeutique

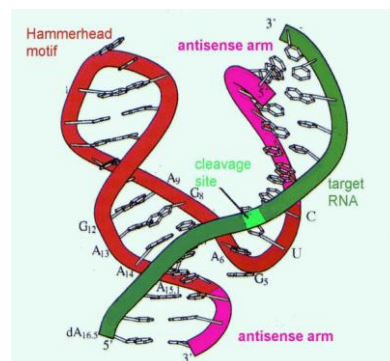
La capacité de cibler des gènes spécifiques pour moduler leur structure et / ou leur fonction dans le génome a des implications de grande envergure en biologie, biotechnologie et médecine. Les TFO représentent des molécules quasi idéales à cet effet en raison de leur capacité à se lier à l'ADN duplex avec une forte affinité et une spécificité élevée, à des sites spécifiques dans un génome. Les oligonucléotides ont été conçus pour cibler les acides nucléiques ainsi que les protéines dans une variété d'applications. D'autres caractéristiques positives générales de cette approche sont: la synthèse facile des réactifs; la disponibilité d'une variété de modifications chimiques (aux bases, au squelette sucre-phosphate et aux extrémités 5' et 3') pour améliorer l'absorption cellulaire, la liaison à la cible, la spécificité et la stabilité; et la disponibilité de systèmes de livraison efficaces.

Cependant, les obstacles à cette approche ont limité le potentiel de la technologie triplex en tant que méthode cohérente et reproductible de régulation transcriptionnelle. Plusieurs limitations à cette technologie comprennent la délivrance et l'absorption de TFO dans les cellules, la stabilité de TFO a niveau intracellulaire, le manque d'affinité et de spécificité de liaison du site cible a cause des concentrations intracellulaires en sel et du pH, le déplacement de l'AND par les activités métaboliques de l'ADN (par exemple, transcription, réplication et réparation), et la structure de la chromatine qui peut présenter une barrière à l'accessibilité du site cible. Il est également possible que des activités alternatives du TFO dans la cellule excluent son but recherché dans la liaison de sa cible duplex pour inhiber l'expression génique. Il a été démontré que les TFO agissent comme des oligonucléotides «leurres», qui lient les facteurs de transcription, de sorte qu'ils ne sont pas disponibles pour lier leurs séquences consensus duplex pour l'activation de la transcription [29]. Des effets aptamères des TFO ont également été observés dans lesquels les oligonucléotides

peuvent se lier aux protéines et inhiber leur activité. Des efforts ont été faits pour surmonter ces limitations, en grande partie par la modification chimique.

4. Stratégie ribozyme

Avant les années 1980, on croyait que seules les protéines étaient capables de catalyser des réactions biologiques dans les cellules vivantes, jusqu'à ce que Dr. Thomas Cech (1986) et son équipe découvrent que l'épissage des ARNm avait toujours lieu même en absence de toute protéine. Ceci a conduit à la découverte du premier ARN ayant une activité catalytique, l'auto-épissage des *introns du groupe I*. Cette découverte a bouleversé les théories de la biologie les plus importantes disant que l'ARN ne jouait qu'un rôle de messenger entre l'ADN qui porte l'information génétique et les protéines qui sont les principaux acteurs des cellules vivantes. Ainsi, diverses autres réactions chimiques ont été réétudiées, ce qui a engendré l'établissement d'une nouvelle classe d'ARN nommée «ribozymes».



4.1. Classification des ribozymes

Initialement, les ARN catalytiques ont été classifiés en deux grandes familles selon leur taille et leur mécanisme de réaction. Ainsi, la famille des grands ARN catalytiques est constituée de l'intron du groupe I, de l'intron du groupe II et de la ribonucléase P (RNase P). La grandeur de ces molécules est généralement de plusieurs centaines de nucléotides. Les grands ARN catalytiques génèrent lors de leur réaction de transestérification des produits ayant des extrémités 5'-phosphate et 3'-hydroxyle. Les ribozymes hammerhead, hairpin, delta et VS sont quant à eux les membres de la deuxième famille, soit les petits ARN catalytiques. La grandeur des petits ARN catalytiques comme leur nom l'indique varie de 35 nt à environ 155 nt. Ces ribozymes génèrent des produits aux extrémités 5'-hydroxyle et 2'-3'-phosphocyclique.

4.2. Domaines d'application

Malgré un mécanisme de réaction différent, tous les ribozymes sont capables d'induire la coupure d'une autre molécule d'ARN. Cette découverte a permis d'émettre l'hypothèse que ces ARN pourraient être utilisés comme outils moléculaires pour cibler des ARNm ou des ARN viraux.

Une grande partie de l'excitation dans le domaine des ribozymes provient de l'application potentielle des ribozymes dans l'inhibition spécifique de l'expression génique (figure 14). Bien que simple en théorie, l'aspect pratique de cette application pose des défis tels que la sélection du site d'ARNm, la livraison et la localisation cellulaire des ribozymes.

Les ribozymes - sont en cours de développement pour traiter une variété de maladies allant des troubles métaboliques innés aux infections virales et aux maladies acquises telles que le cancer. Les ribozymes peuvent être utilisés à la fois pour réguler et pour réparer les gènes pathogènes. Dans certains cas, une délivrance exogène à court terme d'ARN stabilisé est souhaitable, mais de nombreux traitements nécessiteront une délivrance à médiation virale pour fournir une expression à long terme du catalyseur thérapeutique. Les applications actuelles de thérapie génique utilisent des variations sur les ribozymes naturels, mais la sélection *in vitro* a fourni de nouveaux catalyseurs d'ARN et d'ADN, et la recherche sur le trans-épissage et la RNase P a suggéré des moyens d'exploiter les ribozymes endogènes de la cellule à des fins thérapeutiques

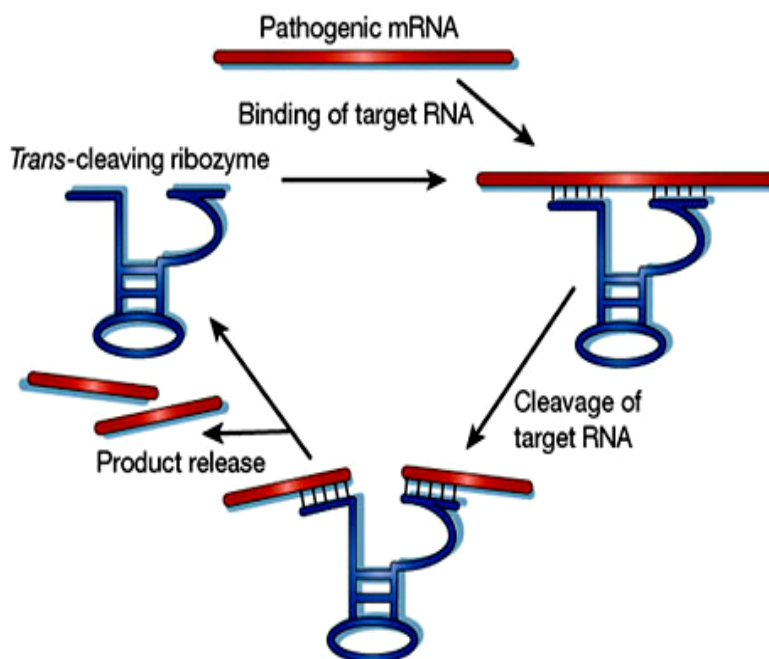


Figure 14. Mode d'action des Ribozymes

5. L 'approche anti-sens

Dans la stratégie antisens, un enchaînement de nucléotides (oligonucléotide) est utilisé pour reconnaître spécifiquement un ARN messager et inhiber la synthèse de la protéine correspondante. Des ARN antisens peuvent être produits directement à l'intérieur d'une cellule à l'aide d'une construction génétique dans laquelle tout ou partie d'un gène (ou d'un ADN) est placé, dans une orientation opposée à celle du gène lui-même, sous le contrôle d'un promoteur, inductible ou non, qui règle l'expression de l'ARN antisens.

5.1. Les mécanismes d'action des oligonucléotides antisens

5.1.1. Mécanisme impliquant la ribonucléase

La présence de la ribonucléase H (RNase H) est essentielle pour assurer l'efficacité d'oligonucléotides antisens dirigés contre la partie codante d'un ARN messager. La RNase H reconnaît les hybrides ADN /ARN et hydrolyse seulement les liaisons phosphodiester de la partie ARN. Cette activité enzymatique est présente dans toutes les cellules puisqu'elle est nécessaire pour éliminer les amorces d'ARN utilisées lors de la réplication de l'ADN. Des cellules en division rapide contiennent un taux élevé de RNase H. Il existe en fait plusieurs enzymes dans les cellules eucaryotes possèdent également une activité RNase H associée.

5.1.2. Inhibition par blocage physique de l'ARN messenger

La fixation d'un oligonucléotide entre la coiffe (CAP) et le codon d'initiation AUG, ou dans une région recouvrant ce codon d'initiation, empêche l'assemblage correct de la machinerie de traduction (protéines spécifiques de la coiffe, fixation de la sous-unité 40S du ribosome, glissement de cette sous-unité le long de la région 5' non codante, fixation de la sous-unité 60S . . .) (Figure 15).

5.1.3. Modification du métabolisme des ARN messagers

Les structures secondaires (et tertiaires) des ARN messagers jouent un rôle important dans leur métabolisme, en empêchant leur dégradation par des ribonucléases cellulaires. Si la Fixation d'un oligonucléotide modifie ces structures et rend certaines régions de l'ARN messenger accessibles aux ribonucléases, le temps de vie et la concentration stationnaire du message seront diminués. Ce mécanisme est difficile à distinguer d'un mécanisme impliquant la RNase H puisqu'une coupure au sein d'un ARN messenger se traduit, la plupart du temps, par une hydrolyse rapide des fragments.

La déstructuration de l'ARN messenger pourrait représenter une étape importante dans le mécanisme d'action des ARN antisens produits in situ par des constructions génétiques, en plus de l'effet éventuel des ribonucléases spécifiques de structures d'ARN en double hélice.

5.1.4. Inhibition de l'épissage des ARN pré-messagers

En e fixant au niveau de jonction exon-intron, un oligonucléotide peut empêcher le fonctionnement normal de la machinerie d'épissage (Figure 16). D'autres sites fonctionnels importants pour l'épissage, qui sont situés à l'intérieur des introns (par exemple, le site de formation de la structure en lasso), peuvent également être la cible d'oligonucléotides antisens. La fixation sur l'un des ARN nucléaires qui participe à la constitution du complexe nucléo-protéique assurant l'épissage peut entraîner une inhibition non spécifique de l'épissage.

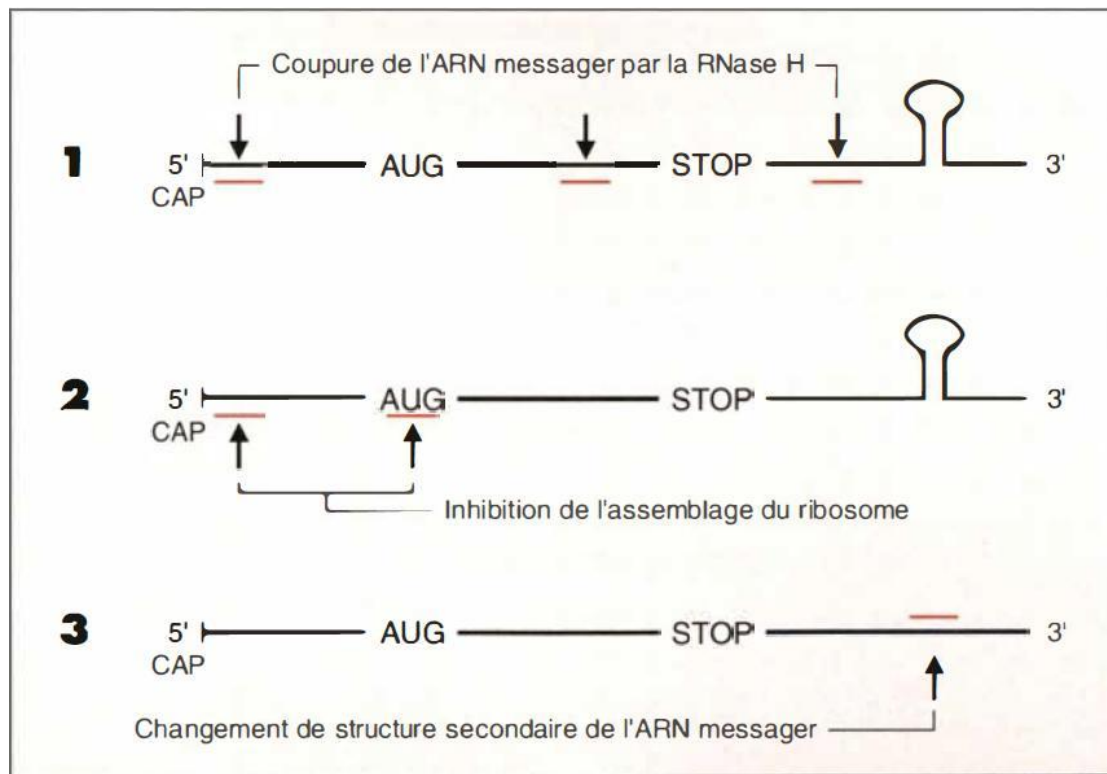


Figure 15. Mécanismes d'action des oligonucléotides antisens.

1. Mécanisme impliquant la RNase H ; 2. Inhibition par blocage physique de l'ARNm ; 3. Changement de la structure secondaire de l'ARNm.

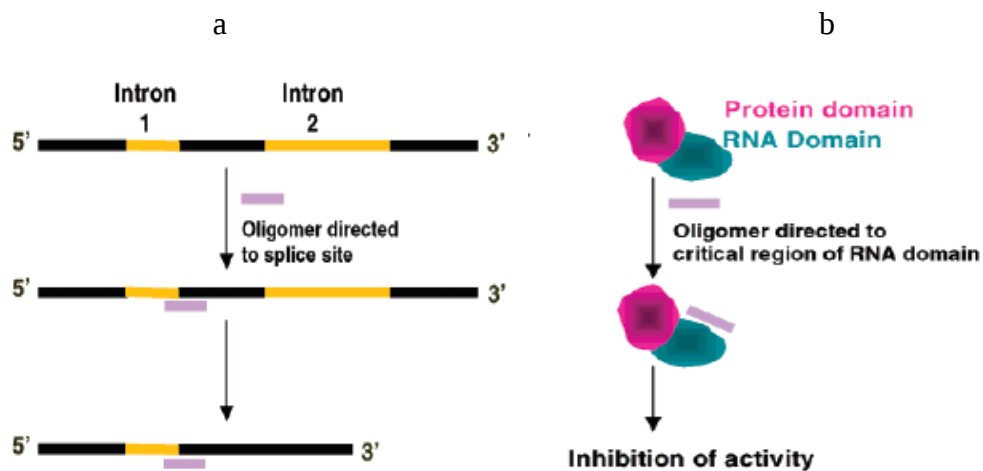


Figure 16. Inhibition de l'épissage par des oligonucléotides antisens.

a. inhibition de l'épissage alternatif ; b. inhibition de l'activité des ribonucléoprotéides.

6. Modifications chimiques des oligonucleotides (les phosphoramidates, les morpholinos, PNA, LNA)

Pour des applications thérapeutiques *in vivo*, il est nécessaire que les oligonucléotides employés soient stables vis à vis de la dégradation par des nucléases, et qu'ils présentent une bonne affinité pour la cible choisie ainsi qu'une bonne biodisponibilité (pénétration cellulaire). L'utilisation d'oligonucléotides (triple hélice et antisens) non modifiés est pour cela limitée, en revanche l'introduction d'acides nucléiques portant une ou plusieurs modification(s) permet de faire face à ces problèmes. Ces dernières peuvent être introduites à plusieurs niveaux au sein de la séquence oligonucléotidique à savoir au niveau du lien phosphodiester (Figure 17), du sucre (Figure 17), et de la base azotée. Elles peuvent être appliquées aux désoxyribo ou aux ribonucléotides, mais nous ne ferons pas de distinction particulière entre les modifications appliquées à l'une ou l'autre famille d'acides nucléiques.

Des modifications du squelette phosphodiester naturel ont été conçues pour améliorer l'affinité et la stabilité de liaison des oligonucleotides. Parmi ceux-ci, il y a des modifications chimiques qui aboutissent à des squelettes neutres ou cationiques pour réduire la répulsion électrostatique entre le squelette phosphodiester chargé négativement de l'oligonucleotide avec celui du d'ADN ou l'ARN cible. Un exemple de telles modifications : les morpholino qui sont caractérisés par le squelette dépourvu de charge négative, groupements morphine et liaisons phosphodiaminates ce qui contribue à la stabilisation des brins.

D'autres modifications du squelette oligonucléotidique qui permettent d'améliorer l'activité des oligonucleotides comprennent PNA (**peptide nucleic acids**) et LNA (**locked nucleic acids**). Les PNA sont des analogues d'acides nucléiques non ioniques dans lesquels le squelette sucre-phosphate est remplacé par une structure polyamide à base de N-aminoéthyl glycine. Les PNA se lient à l'ADN simple brin via l'appariement de bases Watson Crick et peuvent également former des triples hélices via l'appariement de bases Hoogsteen avec le duplex ADN / PNA.

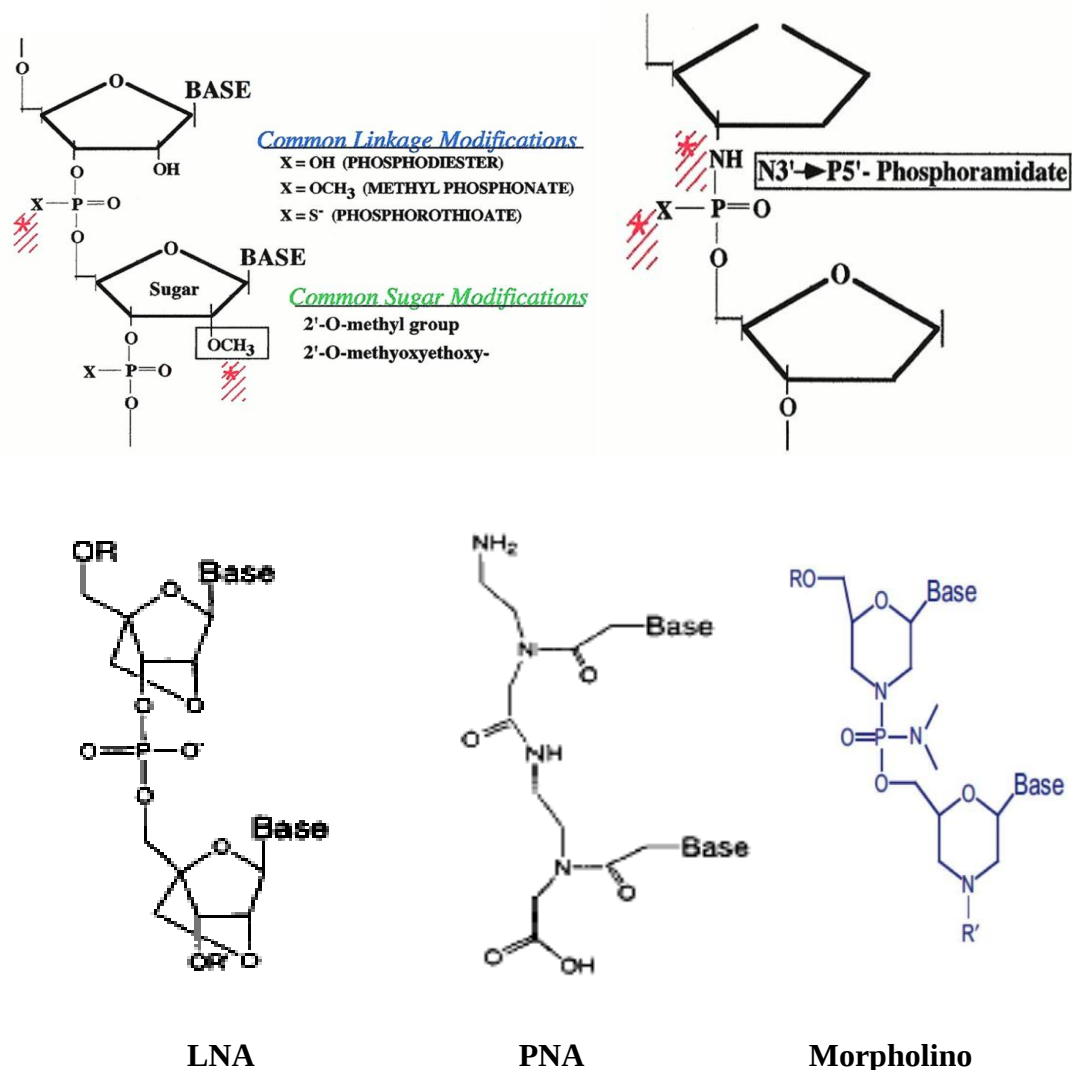


Figure 17. Exemples des modifications des oligonucleotides

7. L'interférence d'ARN (RNAi)

Longtemps, les ARN ont paru être de simples intermédiaires entre ADN et protéines : ARN codants, correspondant aux ARN messagers (ARNm) transmettant le message du gène dans le cytoplasme, ARN non-codants, correspondant aux ARN ribosomiques et ARN de transfert œuvrant au déchiffrement et à la traduction de l'ARNm en protéine, et ARN catalytiques participant à l'épissage des introns.

Mais tout un nouveau monde de petits ARN non codant (ARNnc) a été découvert depuis la fin des années 90 (voir figure ci-dessous). Il comprend notamment deux nouvelles classes de petits ARN : les micro-ARN (ARNmi) et les petits ARN interférents (ARNsi : *small interfering RNA*) qui remplissent de

nombreuses fonctions, en particulier celle de l'inhibition post-transcriptionnelle de l'expression des gènes.

7.1. Quelques grandes étapes de leur découverte

La première manifestation du phénomène d'ARN interférence a été observée en 1990 par Jorgensen et son équipe. Ces biologistes en physiologie végétale voulaient intensifier la couleur des pétales de pétunias. Par transgénèse, ils ont introduit dans les plantes des copies supplémentaires du gène de la chalcone synthétase. Ils eurent la surprise de constater qu'au lieu d'obtenir la couleur pourpre attendue, certaines plantes (plus de 40%) exprimaient des fleurs blanches : non seulement le transgène ne s'était pas exprimé, mais il y avait inhibition de l'expression du gène endogène. Ils observèrent également que ce phénomène était réversible puisque les fleurs retrouvaient leur pigmentation lors de la perte du transgène. A cette époque, ce phénomène fut qualifié de co-suppression.

En 1995, Guo et Kemphues, de l'université Cornell à New York, étudiaient, chez Un ver rond nématode, *Caenorhabditis elegans* (*C. elegans*), le gène *par-1* indispensable à l'établissement de la polarité antéro-postérieure lors de la première division de la cellule-œuf. Au cours de l'étude de la fonction de ce gène, ils ont tenté d'éteindre son expression en injectant, dans la gonade, l'ARN complémentaire du messager de ce gène, c'est-à-dire l'ARN-antisens. Il était attendu par les auteurs que cette injection conduise, dans les embryons produits par l'animal injecté, à la production d'un long ARN double brin résultant de l'appariement de l'ARNm avec l'ARN-antisens injecté. Un tel appariement devait empêcher l'ARNm de *par-1* d'être traduit.

En 1998, Andrew Fire et Craig Mello découvrent que l'introduction d'ARN double brin dans les cellules de nématode, *C. elegans*, permet de réduire spécifiquement l'expression de protéines en se liant à leur ARN messager.

En parallèle, le groupe de G. Hannon a montré la présence dans des extraits de drosophile d'une activité nucléase spécifique de séquence qui était induite par l'introduction d'ARNdb. De plus ils ont détecté la présence dans cette nucléase d'un petit ARN, ce qui les a conduit à proposer qu'il servait de guide pour la reconnaissance de l'ARN messager ciblé. Peu de temps après, le même groupe a décrit le rôle d'une enzyme de la famille de la RNaseIII dans le processus de

fractionnement des ARNdb en molécules d'une vingtaine de nucléotides qui ont été appelés petits ARN interférant. Cette enzyme a été appelée dicer (l'éminceuse).

7.2. Le mécanisme de l'inhibition traductionnelle

Les petits ARNmi et ARNsi ne sont pas codés par la même partie du génome. Les premiers (ARNmi) proviennent de gènes qui leur sont propres et dont ils sont les uniques produits. Les seconds (ARNsi) dérivent de deux sources, exogène ou endogène. La source exogène correspond soit à l'apport extérieur d'un ARN double brin par injection expérimentale, par exemple dans le nématode *C. elegans*, soit à un apport en provenance de génomes à ARN double brin comme ceux de certains virus à ARN des plantes. La source endogène correspond à une production naturelle par utilisation des parties exoniques (codantes ou non codantes) d'un gène plus étendu.

Concernant les ARNmi, leur transcrit est d'abord, pour chacun, une molécule d'ARN simple brin d'un millier de nucléotides. Chacune de ces molécules présente, le long de sa séquence, des motifs nucléotidiques intrachâînes partiellement complémentaires, qui les conduisent à adopter une structure secondaire en forme d'une longue épingle à cheveux. Cette structure peut être divisée en trois parties : corps, tête et les jambes.

L'enzyme Drosha (ribonucléase de type III) y effectue une première série de coupures qui réduit cette structure en une courte molécule d'une centaine de nucléotides, toujours en épingle à cheveux (short hairpin RNA : shRNA), mais amputée de ses « jambes ». L'exportine-5 la transporte alors dans le cytoplasme. Là, une seconde ribonucléase de type III, appelée Dicer, achève le travail commencé par l'enzyme Drosha : par une deuxième série de coupures, elle élimine la « tête » de l'épingle à cheveux tandis que d'autres protéines qui lui sont associées (de type hélicase) dissocient le petit ARNdb (double brin) subsistant, d'une vingtaine de nucléotides, en deux ARN simple brin (Figure 18). L'enzyme Dicer (Il s'agit d'une ribonucléase reconnaissant spécifiquement tout ARN double brin, indépendamment de sa séquence, afin de produire de petits ARN de manière ATP dépendante) clive et sépare également les longs ARN double brin des ARNsi en morceaux d'environ 21 nucléotides. L'un des simples brins obtenus est le microARN (ARNmi) ou le petit ARN interférent (ARNsi). Il s'associe à un complexe protéique en formant un nouveau complexe appelé RISC (RNA Induced Silencing Complex). Le complexe RISC a une composition protéique variable selon les espèces, mais il est toujours doté

d'une activité hélicase ATP-dépendante et d'une activité RNase. C'est en cet équipage qu'il se porte sur sa cible: l'ARNm auquel il s'apparie. En cas d'appariement parfait (ARNsi), l'ARNm est détruit et il n'y a pas traduction. En cas d'appariement imparfait (ARNmi), l'ARNm est détruit et il n'y a pas traduction. Les RISC qui sont à l'origine de cette destruction restent ensuite parfaitement fonctionnels, ce qui leur permet d'opérer de nouveau sur d'autres ARNm de même spécificité. C'est cette réutilisation qui les rend particulièrement offensifs. Mais même lorsque l'appariement est imparfait (ARNmi), il y a tout de même barrage à la traduction de l'ARNm.

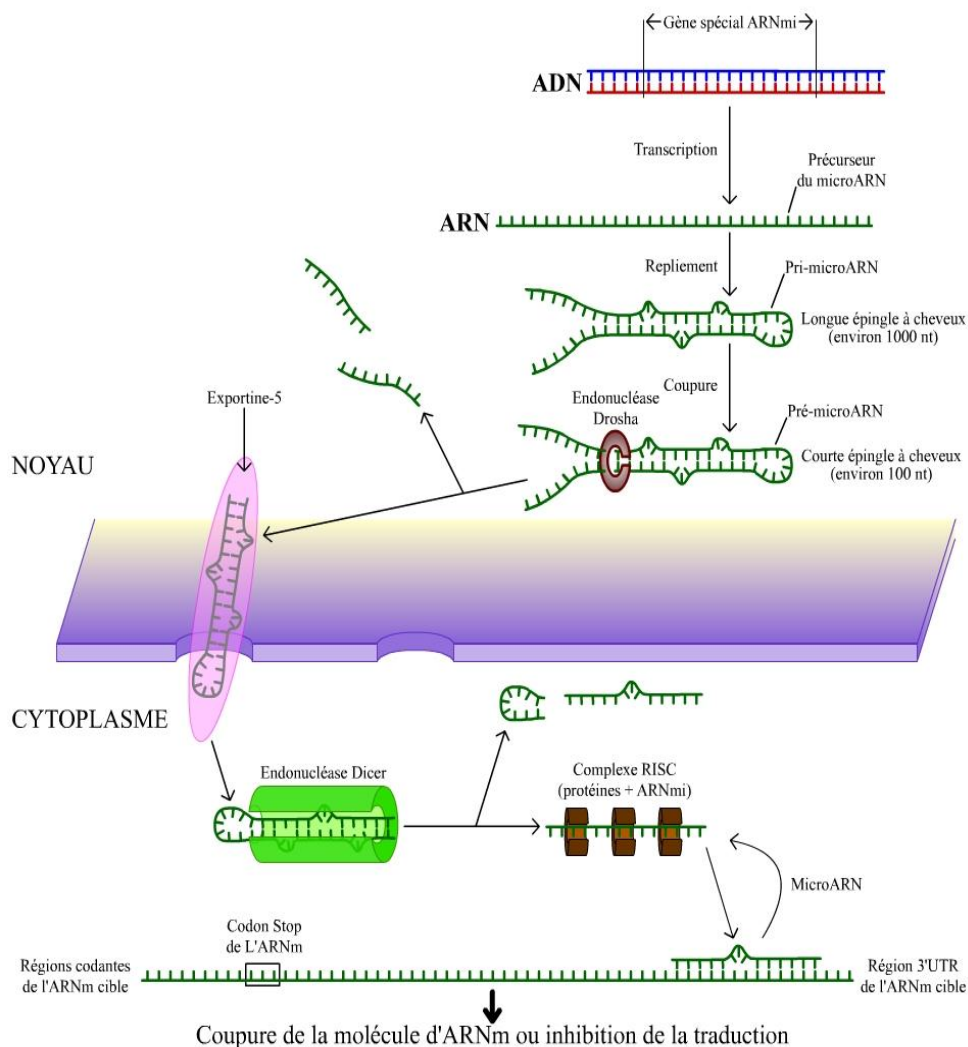


Figure 18. Mécanisme de production et d'action des ARNm

en cas des ARNsi, Lorsqu'un ARN double brin est présent dans le cytoplasme des cellules eucaryotes, il est pris en charge par une enzyme à activité RNase III, Dicer, qui le clive en petits ARN double brin, longs de 21 nucléotides, et présentant à chaque extrémité 3' deux nucléotides non appariés, tandis qu'ils sont phosphorylés en 5' (Figure 19). Une fois formés, les siRNA s'associent à un complexe protéique effecteur, le RISC. Les fonctions du RISC semblent nombreuses mais à ce jour, elles ne sont connues que dans les grandes lignes. Tout d'abord, ce complexe enzymatique déroule le petit ARN, de façon dépendante de l'ATP, pour dissocier les deux brins. La plus faible stabilité en 5' de l'antisens permettrait le déroulement sélectif et l'incorporation de ce brin dans le RISC. Le complexe activé ne conserve que le brin antisens, qui lui sert de guide pour reconnaître l'ARNm cible. Le complexe RISC activé s'apparie à l'ARNm par l'intermédiaire des siRNA, qui ont une correspondance de séquence parfaite avec leur cible. Cela déclenche alors la dégradation de l'ARN messenger. Lorsque l'ARN messenger est clivé par le complexe RISC, les fragments générés, longs d'une vingtaine de nucléotides, sont ultérieurement dégradés par des RNases cellulaires, alors que le complexe endonucléolytique RISC est recyclé pour aller cliver d'autres ARNm.

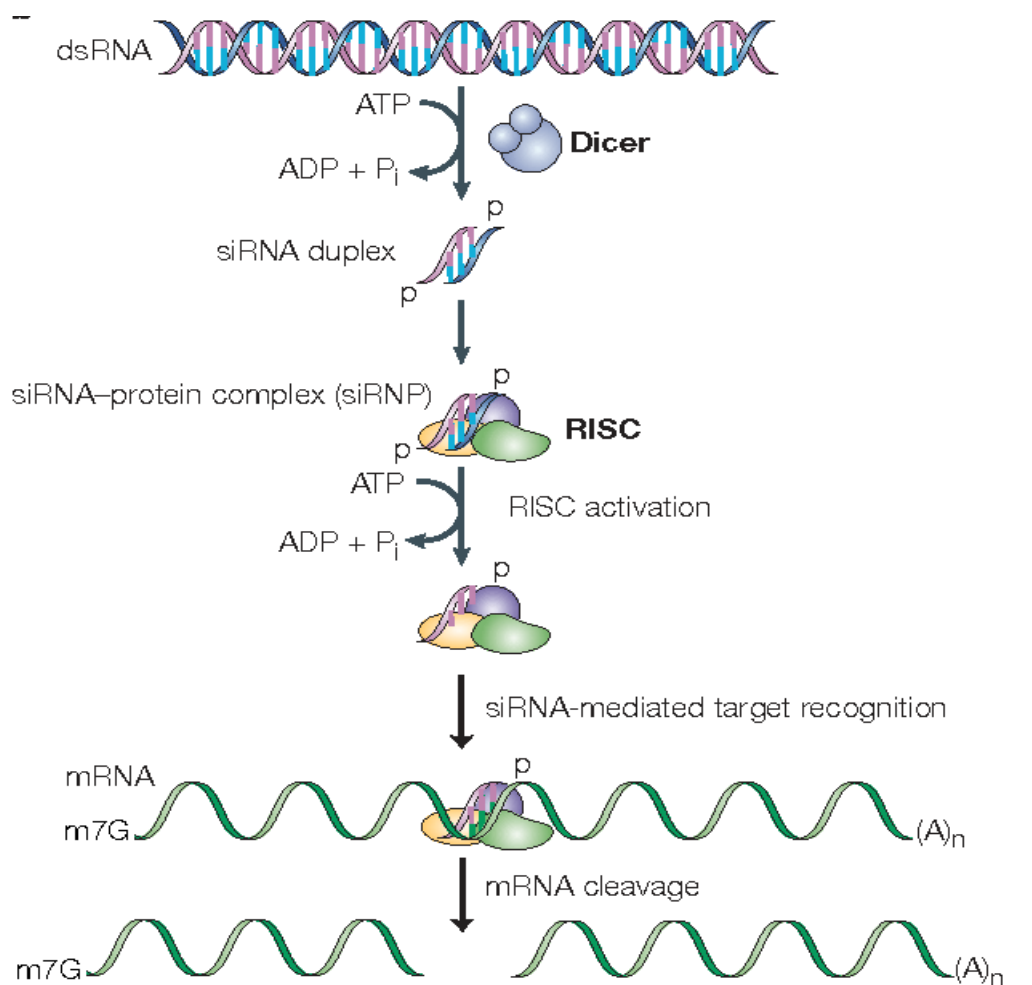


Figure 19. Mécanisme de production et d'action des siRNA.

7.3. Les applications de l'ARNi

L'ARNi représente une méthode de premier ordre pour étudier la fonction des gènes, pour valider de potentielles cibles thérapeutiques ou même pour développer de nouvelles thérapeutiques, par exemple dans les domaines de l'oncologie ou de la virologie. Compte tenu du caractère naturel de ce mécanisme, il se révèle très efficace pour inhiber l'expression des gènes, bien plus que les outils précédemment développés, tels que les stratégies antigènes (triplex) ou antisens. Malgré tout, le problème de la délivrance reste un obstacle majeur à l'utilisation courante de l'ARNi. Pour des applications thérapeutiques chez l'Homme, des problèmes de stabilité, de biodistribution, de ciblage, de pénétration cellulaire, ou d'induction d'éventuels effets secondaires, notamment d'ordre immunologique sont principalement rencontrés.

L'intérêt suscité par le mécanisme d'ARNi et les efforts de recherche scientifique ont d'ores et déjà permis de développer des thérapeutiques aujourd'hui évaluées sur l'Homme dans des essais cliniques.

IV. Analyse du génome et du transcriptome

Tout chercheur, aussi bien dans le domaine fondamental qu'appliqué, se voit un jour rencontrer la notion de mutation. On entend par mutation toute modification qui touche le matériel génétique ou génome d'une cellule, contenant le message héréditaire.

Elles peuvent entraîner la perte totale d'une information et la suppression de la protéine codée par le gène muté, ou alors une altération de la structure de cette protéine. Une protéine modifiée représente une molécule non fonctionnelle et ce dysfonctionnement est d'une importance capitale dans le diagnostic moléculaire, pouvant être à l'origine de la plupart des pathologies.

Il existe également des mutations n'affectant pas radicalement la structure ou la fonction de la protéine concernée, mais cependant variées et très répandues au sein des populations. Dès lors on peut parler de polymorphisme des individus, et il s'agit souvent de variations d'un seul nucléotide connues sous le nom de SNP ou SNiPs (Single Nucleotide Polymorphism). Quoique non essentielles pour le fonctionnement de la cellule, ces variations présentent un grand intérêt dans la médecine légale et sont utilisées pour l'établissement de la paternité ou de l'identité.

Nous nous intéressons dans ce chapitre aux méthodes utilisées pour la détection des mutations et/ou celles utilisées pour l'étude de polymorphisme

1. Détection des mutations

La détection de mutations dans l'ADN représente une étape essentielle de la biologie moléculaire, pour la recherche fondamentale comme pour les applications médicales.

1.1. Méthodes basées sur la séquence

Le séquençage des acides nucléiques est sans aucun doute le domaine analytique ayant bénéficié de l'évolution la plus importante au cours des vingt dernières années. Cette méthode permet d'identifier la mutation d'une manière sûre. Toutefois elle est lourde et génère plus d'information qu'il n'en faut. Aussi elle n'est souvent utilisée que pour la mise en évidence pour la première fois d'une mutation.

Deux méthodes principales ont d'abord été utilisées : la méthode chimique de Maxam et Gilbert, aujourd'hui pratiquement abandonnée, et surtout la méthode enzymatique de Sanger aux di-désoxynucléotides qui s'est généralisée en s'adaptant à l'évolution des outils de la génétique moléculaire et des moyens techniques.

Toutefois, le séquençage du génome humain avec la technologie Sanger a été un immense effort qui a pris plus de dix ans et a coûté environ trois milliards de dollars. Il est donc évident que cette méthode n'est pas adaptée au séquençage de grands génomes, et c'est pour cette raison que de nouvelles technologies ont été développées.

Des méthodes dites « séquençage à haut débit » ou « 2^{ème} génération » ont révolutionné la génomique en permettant de séquencer un génome humain en quelques jours pour moins de 1000 dollars. Ainsi, les technologies de troisième génération, encore plus récentes, qui, depuis quelques années, constituent une nouvelle révolution.

La méthode de séquençage de Sanger

Le recopiage d'un brin matrice par une ADN polymérase ADN dépendante est initiée par la fixation d'un oligonucléotide spécifique (amorce), complémentaire du brin matrice. Cette ADN polymérase va permettre l'élongation d'un nouveau brin complémentaire du brin matrice dans le sens 5' 3'. L'ADN polymérase permet l'incorporation de nucléotides (dNTP: déoxynucléotides) libres présents dans le milieu réactionnel par la formation d'un pont phosphodiester entre le 3'OH de la chaîne et le 5' phosphate du dNTP suivant. La réaction de Sanger repose sur l'incorporation alléatoire par cette ADN polymérase de didéoxynucléotides interrupteurs de chaîne (ddNTP) eux aussi présents dans le milieu réactionnel. Ces ddNTP diffèrent des dNTP par leur extrémité 3'. L'extrémité 3'OH des dNTP est remplacée par une extrémité 3'H. Cette modification empêche la formation de la liaison phosphodiester entre le ddNTP incorporé dans la chaîne et le nucléotide suivant. L'allongement de la chaîne est alors interrompu. Une migration électrophorétique du produit de cette réaction de séquence sur un gel très résolutif (polyacrylamide) va séparer tous les fragments présents en fonction de leur masse moléculaire (taille). Les plus petits fragments vont migrer plus rapidement que les grands. La grande résolution de ce gel permet de distinguer des fragments différents entre eux d'une paire de base (figure 1). L'identification du ddNTP présent à l'extrémité 3' de chaque fragment déterminera la séquence nucléotidique du brin matrice initial.

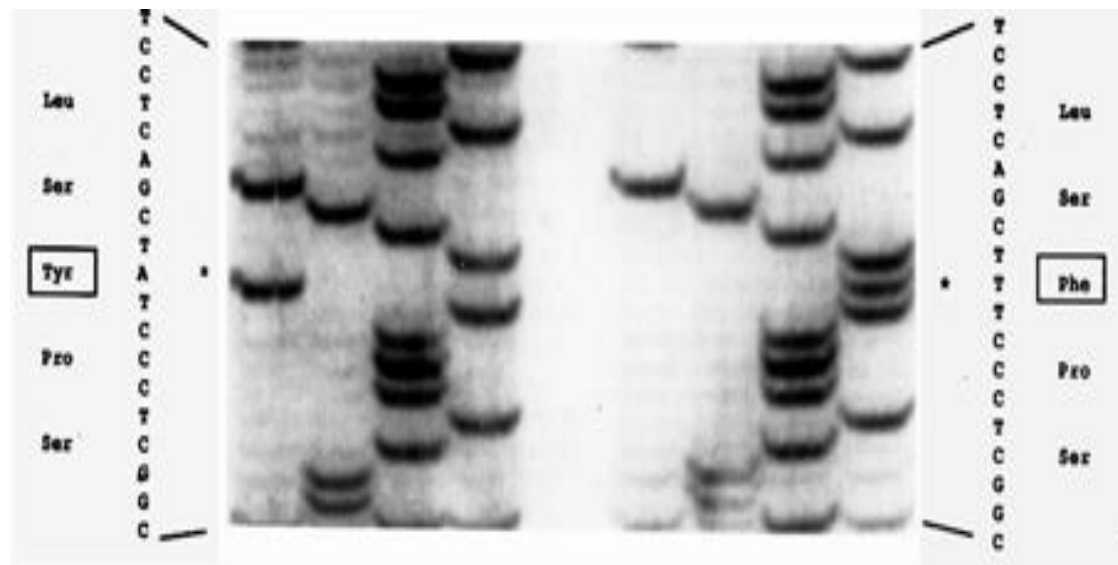


Figure 1. Exemple de mise en évidence d'une mutation ponctuelle par séquençage.

Pyrosequencing : Le séquençage haut-débit (2^e génération)

La méthode de séquence en utilisant le pyroséquençage peut être utilisée pour détecter des mutations. Le pyroséquençage est une technique qui permet d'effectuer un séquençage par quantification d'un signal lumineux émis lors de l'incorporation d'un nucléotide.

Pyroséquençage comprend le séquençage de l'ADN en synthétisant le brin complémentaire d'une base unique à la fois, tandis que la détermination de l'être nucléotidique spécifique incorporé au cours de la réaction de synthèse. La réaction se produit sur la matrice ADN simple brin immobilisé, où les quatre désoxyribonucléotides (dNTP) sont ajoutés de façon séquentielle et les dNTP non incorporées sont dégradés par voie enzymatique (dégradées par apyrase) avant l'addition de la prochaine dNTP à la réaction de synthèse. Si le nucléotide ajouté dans le milieu réactionnel correspond à celui attendu par la polymérase, il est incorporé dans le brin en cours de synthèse et libère un Pyrophosphate. Une ATP sulfurylase vient alors transformer ce Pyrophosphate en ATP qui est alors utilisé, couplé à une Luciférine, par une Luciférase. On a alors production d'Oxyluciférine et d'un signal lumineux. Le signal est détecté par un capteur CCD (Charge-Coupled Device) qui va le traduire sous forme d'un pic sur le *Pyrogramme* (figure 2). La hauteur de ce pic est fonction de l'intensité du signal lumineux, elle-même proportionnelle au nombre de nucléotides incorporés en même temps. L'ordre des dNTP qui produisent des signaux chimiluminescents détermine la séquence d'ADN de la matrice.

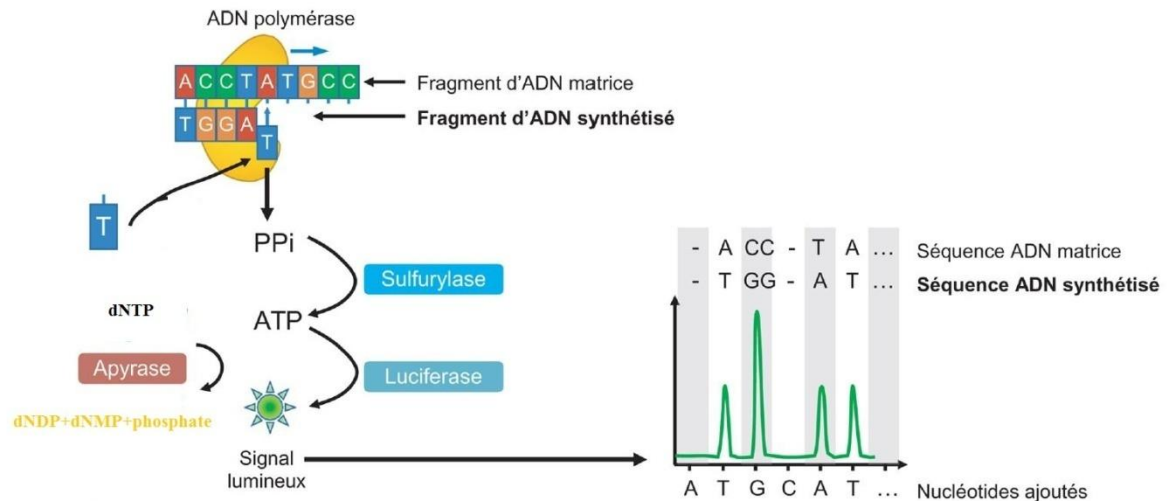


Figure 2. Séquençage de l'ADN à haut débit (exemple du pyroséquençage).

1.2. Méthodes basées sur l'hybridation

Southern Blot

L'hybridation de deux chaînes d'ADN est due à l'appariement de bases complémentaires A-T et G-C. Elle est donc diminuée lors de mésappariement. Le T_m d'un oligonucléotide de 10 à 20 bases sera donc abaissé s'il y a un mésappariement.

La méthode de Southern est particulièrement adaptée pour détecter les grandes altérations telles que les insertions, les délétions ou les réarrangements. Mais cette méthode peut aussi être utilisée pour détecter les mutations ponctuelles si on utilise un oligonucléotide comme sonde. 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

- Extraction de l'ADN génomique.
 - Digestion par des enzymes de restriction différentes du même ADN génomique.
- L'ADN génomique
- 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).

- Transfert des fragments monocaténaire du gel d'agarose à un support souple (feuille de nylon par exemple).Le transfert des fragments monocaténaire du gel d'agarose à un support type nylon s'effectue par simple capillarité.
- Fixation des fragments monocaténaire d'ADN sur le support souple et hybridation dans des conditions optimales de stringence avec une sonde complémentaire marquée à un radioisotope. La sonde s'apparie avec les fragments d'ADN monocaténaire selon les règles de complémentarité. De plus, elle est marquée avec un radioisotope.
- Lavages et révélation (dans ce cas par autoradiographie).Après de nombreux lavages, le support solide est mis en contact avec un film photographique. Le film est ensuite révéler. Les bandes d'ADN monocaténaire 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. La comparaison entre un profil d'un individu sauvage et cible permet de détecter la mutation (figure 3).

La figure 4 montre un exemple d'application de la technique southern pour la détection d'une mutation de délétion de la dystrophine. La piste 2 contient de l'ADN de Charles (enfant 1), dans lequel une bande est manquante. Cela correspond à l'exon 51 du gène de la dystrophine. L'ADN de sa mère se trouve dans la piste 1. Les pistes 3 et 4 contiennent respectivement des échantillons d'ADN de la grossesse précédente et de l'échantillon prénatal actuel.

Northern blot

Dans certains cas si l'altération se retrouve dans l'ARN, un northern peut être employé pour la détecter. 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 détecter les intermédiaires de maturation et les différentes formes d'épissage de l'ARN

Dot-blot

Dans une *dot blot*, on transfère les acides nucléiques (ADN ou ARN) depuis un milieu liquide directement sur une membrane, sans séparation préalable, donc sans électrophorèse préalable. Le transfert peut être réalisé par diffusion simple, par diffusion dans champs électrique (électrodifusion) ou par aspiration sous vide. La détection est similaire à celles utilisées dans Southern et northern blot.

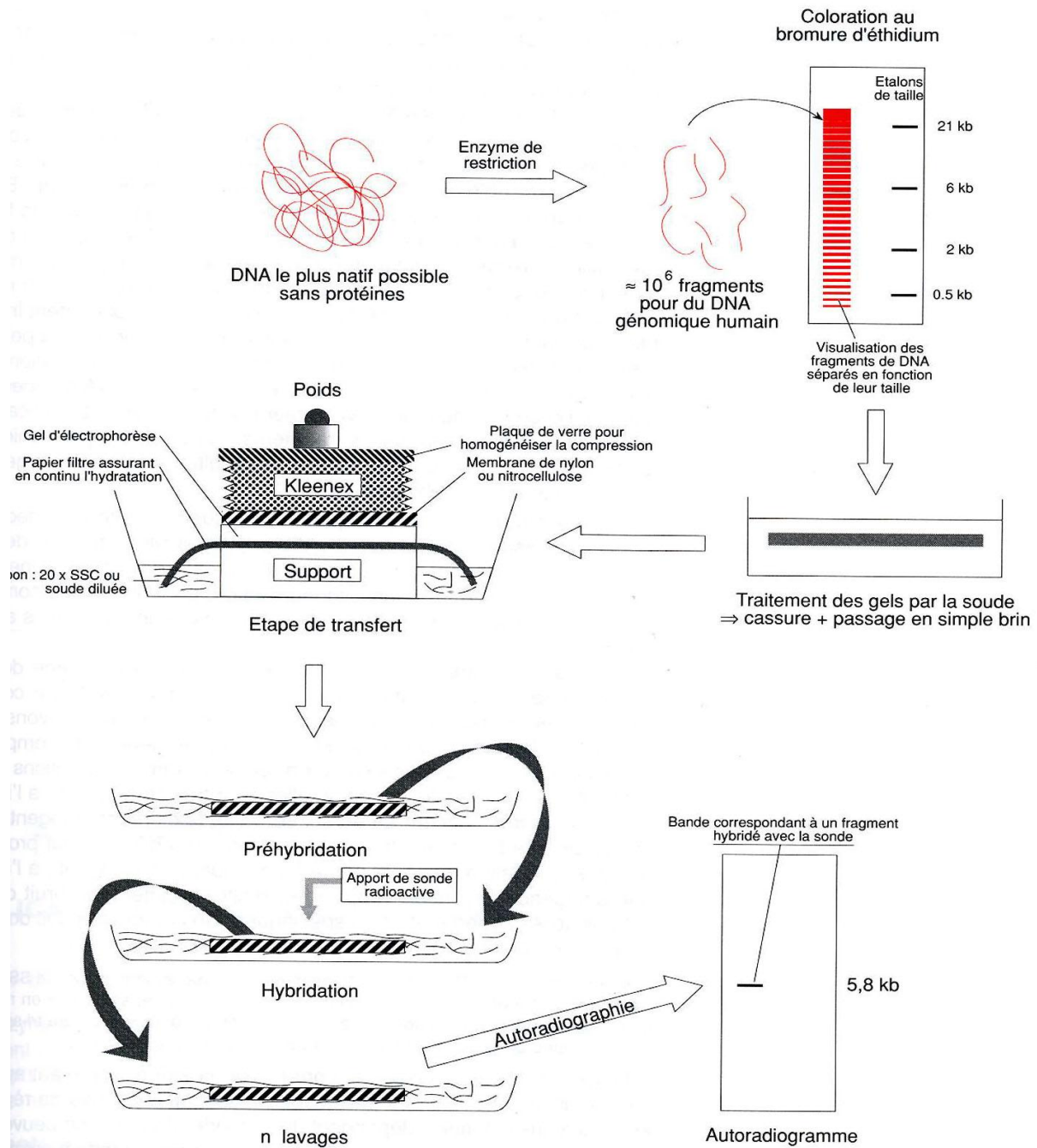


Figure 3. Principe de la technique Southern Blot.

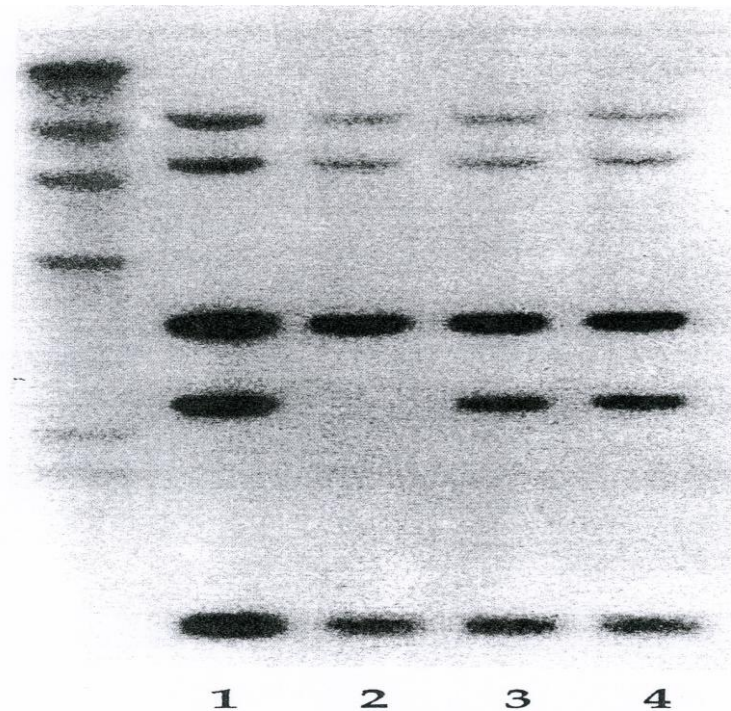


Figure 4. Identification de la délétion de la dystrophine par analyse de Southern.

DGGE = Denaturing Gradient Gel Electrophoresis

Cette méthode est basée sur la différence de mobilité électrophorétique entre deux molécules d'acides nucléiques qui diffèrent par une variation de séquence. La mobilité électrophorétique d'un fragment d'ADN double brin dans un gel de polyacrylamide est affectée par la séparation partielle de ses deux brins. Une molécule composée en partie de la double hélice classique et en partie de deux simples brins est fortement ralentie par rapport une molécule double brin ou complètement dénaturée. Dans des conditions dénaturantes (urée-formamide), le T_m du double brin d'ADN dépend de la séquence et la mobilité électrophorétique sera d'autant réduite qu'il atteindra le T_m tôt lors de la migration

Après amplification PCR des régions susceptibles de contenir une mutation, l'ADN de ces fragments migre dans un gel d'acrylamide où il rencontre des conditions de plus en plus dénaturantes (gradient en acrylamide). Cependant, si le gradient augmente de façon linéaire, la dénaturation de l'ADN, elle, ne se fait pas de façon progressive tout au long de la molécule, mais de domaine en domaine, qui s'ouvrent plus ou moins rapidement, en fonction de leur composition en bases AT et GC. La stabilité de chacun des domaines dépend de la structure primaire de la molécule. Deux molécules dont les séquences ne différeront que d'une paire de base pourront donc ne pas s'ouvrir au même moment et migreront alors différemment à travers un gradient précis.

DHPLC: Denaturing high-performance liquid chromatography

Cette méthode est proche de la DGGE mais ici la détection des hétéroduplex se fait par HPLC en utilisant une colonne de paire d'ions en phase reverse dans les conditions partiellement dénaturantes.

1.3. Méthodes basées sur la PCR**La multiplex PCR**

Les altérations du génome peuvent être détectées par PCR en utilisant plusieurs amorces, ce qui permet l'amplification simultanée à sélectivité dirigée de plusieurs régions.

Cette méthode a été employée pour la première fois par Chamberlain et coll. (1988) pour détecter les mutations de la dystrophine responsable de la maladie de Duchenne.

La multiplex PCR est une variante de la PCR dans laquelle deux séquences cibles ou plus peuvent être amplifiées en incluant plus d'une paire d'amorces dans la même réaction (figure 5). La PCR multiplex a le potentiel de générer des économies considérables de temps et d'efforts en laboratoire. Cette méthode a été appliquée avec succès dans de nombreux domaines des tests ADN, y compris l'analyse de délétion de gène, l'analyse de mutation et de polymorphisme, l'analyse quantitative. Dans le domaine des maladies infectieuses, la multiplex PCR s'est avérée un outil précieux pour l'identification des virus, bactéries et parasites.

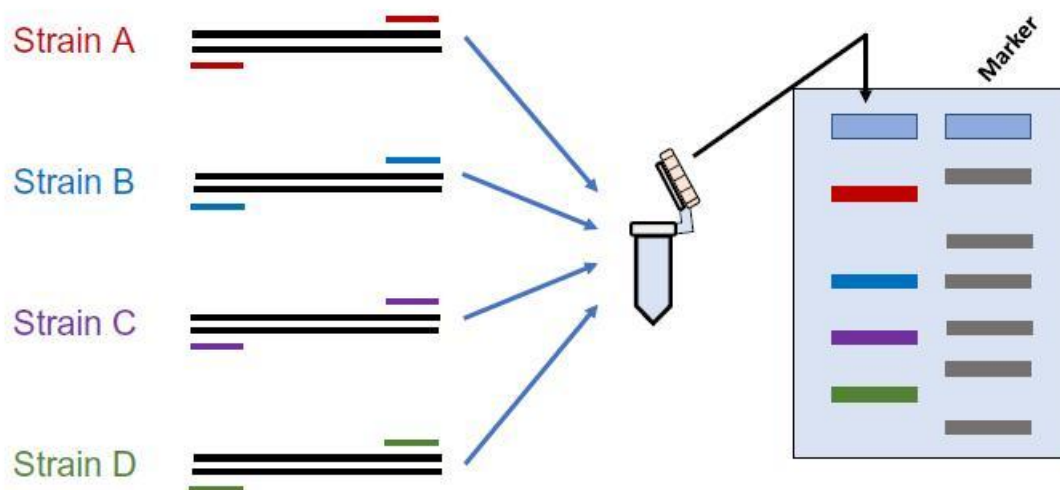


Figure 5. Principe de la multiplex PCR.

PCR oligonucléotide allèle spécifique, PASA (PCR Amplification of Specific Alleles)

Elle est basée sur un principe de misappariement d'une amorce en 3' qui empêche l'amplification du fragment. En PCR, la polymérase ne peut polymériser que si la dernière

base en 3' est stabilisée par des liaisons hydrogènes, si la dernière base est hybridée. Deux amplifications sont réalisées en parallèle. Une des amorces est commune aux deux amplifications. La deuxième existe en deux versions, une spécifique de la séquence normale et l'autre spécifique de la séquence mutante. Cette dernière est choisie de façon à ce que le dernier oligonucléotide en 3' corresponde à la mutation. L'amplification de la séquence mutée ne peut avoir lieu (Figure 6). Un témoin peut être ajouté, c'est un oligonucléotide qui hybride sur tous les allèles (connus) en amont de l'hybridation permettant de voir la mutation. On obtient donc deux bandes, une bande haute permettant de voir si la polymérisation a bien eu lieu et une bande basse permettant de voir la présence ou non de la mutation, suivant l'oligonucléotide qui a été utilisé.

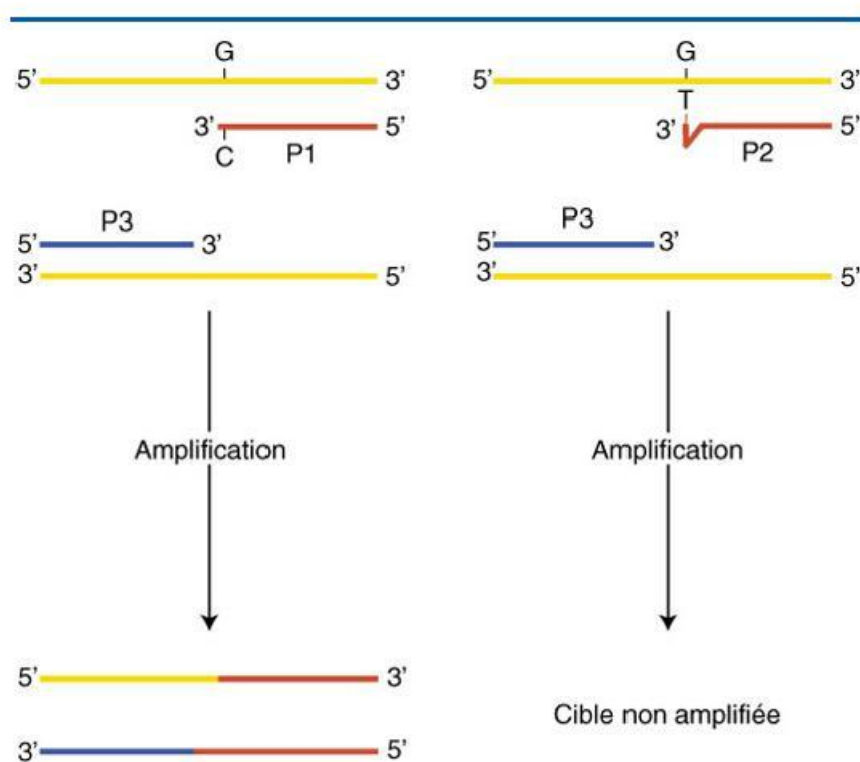


Figure 6. Principe de la *polymerase chain reaction* (PCR) spécifique d'allèles (P1, P2 et P3 : des amorces).

1.4. Méthodes basées sur la conformation de l'ADN

SSCP (Single Strand Conformation Polymorphism)

La technique SSCP est basée sur la capacité des simples brins d'ADN de pouvoir prendre dans des conditions NON dénaturantes différentes conformations 3D selon la séquence de départ. Deux brins ADN différents en structure primaire adopteront des structures spatiales différentes, ce qui se traduit en une différence de mobilité

électrophorétique (Figure 7). Après dénaturation thermique des fragments ADN amplifiés, on procède à un refroidissement brusque, pour que les molécules restent monocaténaires et adoptent les conformations les plus stables pour chacune d'entre elles. Les brins d'ADN forment des structures secondaires en tiges/boucles qui sont uniquement dépendantes de la séquence primaire, de plus il existe une structure tertiaire provenant de liaisons entre les sucres et les bases. Les profils de migration en gel d'acrylamide non dénaturant sont comparés en partant de l'idée qu'une différence de migration signifie une différence de séquence et par conséquent un polymorphisme ou une mutation.

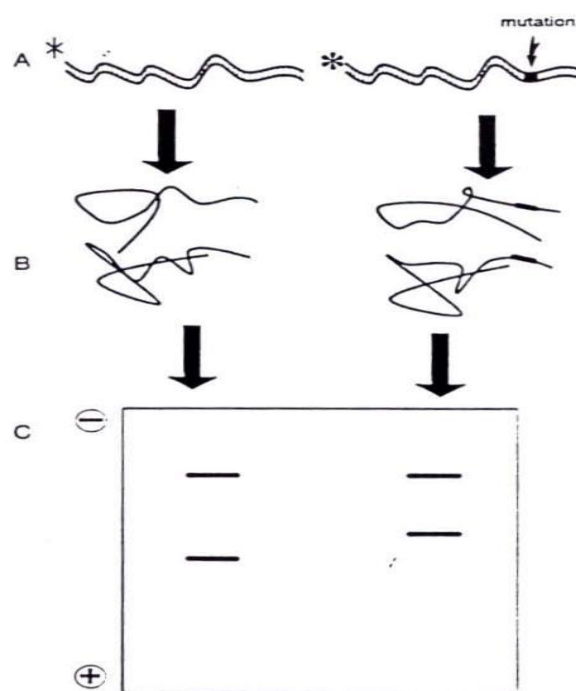


Figure 7. Principe de SSCP

ddF (Dideoxyfingerprinting)

Cette méthode est un hybride entre la séquence selon la méthode de séquençage à l'aide des dideoxynucléotides et la SSCP. Le fragment d'ADN où se trouve la mutation est amplifié puis une réaction de séquence est faite en employant un seul dideoxynucléotide et une amorce marquée en 5'. Le produit de la réaction est dénaturé puis chargé sur un gel d'acrylamide non dénaturant.

1.5. Sensibilité d' hétéroduplex à une coupure

Dans cette technique les deux fragments d'acide nucléique sont hybridés pour former un hétéroduplex (figure 8) et on détecte les mésappariements en les coupant avec des enzymes (nucléase) ou chimiquement.

Les produits de PCR sont dénaturés, ce qui conduit à la séparation des brins d'ADN, puis une renaturation lente conduit à la reliure des simples brins en une forme à deux brins. Les hétéroduplexes se forment créant ainsi un mésappariement (mismatch), lorsqu'un fragment simple brin avec une mutation complète se relie avec un fragment simple brin de type sauvage.

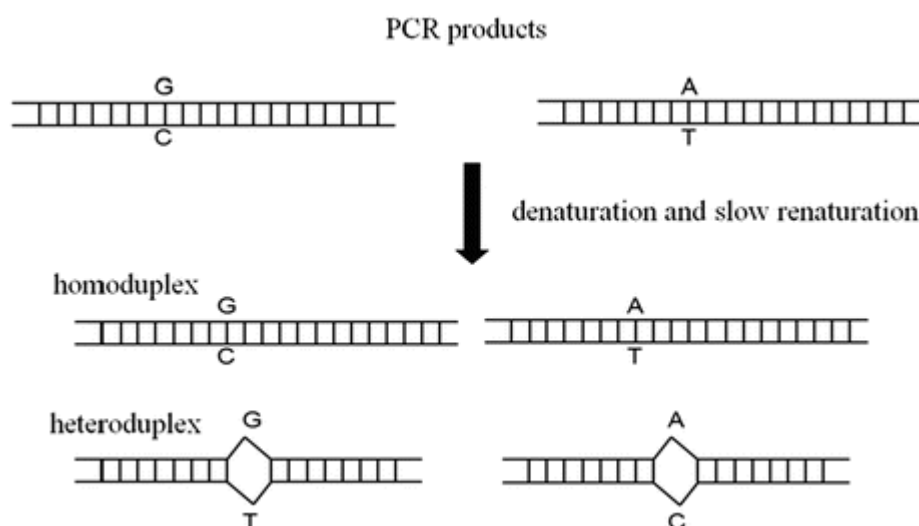


Figure 8. Formation d'hétéroduplex.

Utilisation des nucléase

Plusieurs nucléases peuvent être utilisées (Exemple : la nucléase S1 d'*Aspergillus oryzae*, la nucléase P1 de *Penicillium citrinum* ou la mungbean nucléase de *Vigna radiata* ou la nucléase CEL 1 du céleri). Ces nucléases sont spécifiques de l'ADN simple brin et ne digèrent pas l'ADN double brin.

Les produits de la PCR subissent une dénaturation et une renaturation lente. Les hétéroduplexes (fragments d'ADN présentant des mésappariements) n'apparaissent que dans les pools contenant de l'ADN avec une mutation dans le fragment en cours d'analyse. La nucléase (exemple La CEL I) est ensuite ajoutée à chaque échantillon afin de couper les hétéroduplex en position 3' du mésappariement. À la suite de la digestion, des fragments plus courts supplémentaires sont produits, ce qui indique la présence potentielle d'une mutation dans l'ADN du pool analysé (figure 9). La visualisation des produits sur un gel de polyacrylamide en utilisant un séquenceur LI-COR. L'astérisque bleu indique une amorce directe marquée avec IRDye 700; un astérisque vert indique une amorce inverse marquée avec

IRDye 800. I, un pool contenant de l'ADN sans mutation dans le gène amplifié; II, un pool contenant de l'ADN avec une mutation dans le gène amplifié.

La nucléase S1, la nucléase P1, la mungbean nucléase sont des protéines à Zinc et sont principalement actives à pH 5. Cette action à pH acide pose quelques problèmes pour les régions riches en AT en effet à ce pH on peut avoir une dénaturation partielle de l'ADN. Dans ce cas on peut utiliser une nucléase de plante comme la nucléase CEL 1 du céleri qui coupe l'ADN simple aux pH neutres.

Certaines nucléases comme la nucléase S1 digèrent très difficilement lorsque le mésappariement est limité à une seule base, mais au-dessus de deux bases la vitesse de réaction augmente énormément. Dans ce cas, la technique n'est donc valable pour détecter des délétions de petites tailles.

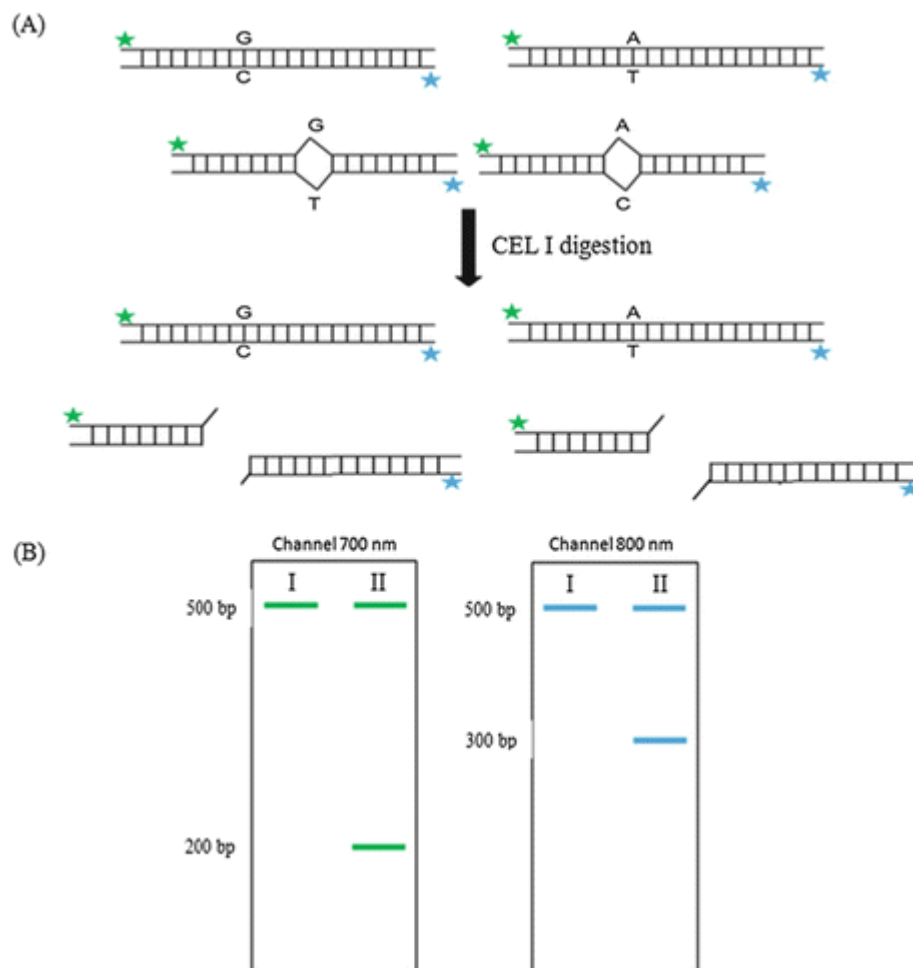


Figure 9. Un schéma simplifié de clivage de mismatch par la CEL I (A) et la visualisation des produits sur un gel de polyacrylamide(B).

Utilisation des RNase

Les RNase A, T1 ou T2 coupent les ARN sous forme simple brin et non lorsqu'ils sont hybridés. Lors d'une hybridation entre deux clones de deux individus différents, les mésappariements seront digérés et seules les régions hybridées resteront. La technique est donc équivalente à la digestion par la nucléase S1 mais ici on utilise une sonde ARN. Cependant les RNase ne digèrent pas tous les mésappariements, elles digèrent uniquement C-A, CC, C-T et U-T, c'est à dire 4 sur les 12 possibles. Pour augmenter ce nombre on peut faire deux ARNs marqués pour les deux espèces et dans ce cas on voit 60% des mésappariements.

Coupure par les méthodes chimiques (CCM= Chemical Cleavage of Mismatch)

Les méthodes chimiques ont été développées par les auteurs qui ont étudié les structures secondaires des ARN. Les produits chimiques présentent un certain nombre d'avantage sur les enzymes, ils agissent sur une gamme de pH et de force ionique beaucoup plus large, il n'y a pas de problème d'instabilité et ils ne nécessitent pas d'avoir un acide nucléique purifié.

De nombreux produits chimiques réagissent avec les bases des acides nucléiques, ils produisent des lésions qui peuvent être ensuite coupées par un traitement alcalin, c'est le principe de base de la méthode de séquence de Maxam et Gilbert. Parmi ces produits l'hydroxylamine (H) réagit avec les C mésappariés, le tétraoxyde d'osmium réagit avec les T mésappariés (OT). La combinaison des deux (HOT) permet de détecter les C et T.

2. Analyse du polymorphisme

2.1. Etablissement du profil des protéines

Les premières méthodes d'analyse du polymorphisme utilisaient les différences de migration des protéines d'un individu sur gel à l'état natif. Quelques protéines telles que les estérases étaient révélées par leurs activités.

Ces méthodes ont été très utilisées par les généticiens des populations mais le nombre de variants était assez faible et pouvait résulter d'un effet transcriptionnel.

2.2. Analyse des fragments de restriction RFLP

Les applications du RFLP en biologie moléculaire ont à la base deux notions fondamentales :

- **Les Enzymes de Restriction** – endonucléases qui reconnaissent des séquences spécifique de l'ADN double brin et scindent spécifiquement au niveau de ces sites ;
- **Le Polymorphisme** – différence génétique héréditaire au sein d'une population. Peut exister au niveau d'un seul nucléotide, cas des SNP (Single Nucleotide Polymorphism).

Les SNP ou des mutations peuvent causer l'apparition/disparition d'un site de restriction et donc une modification du profile de restriction. De ces notions émerge la notion de RFLP, qui définit concrètement certains fragments d'ADN génomique humain cloné qui, lorsqu'ils sont utilisés comme sondes sur de l'ADN génomique humain digéré par un enzyme de restriction, permettent la détection, pour un locus donné, des fragments dont la taille est différente selon les individus (correspondant aux différents allèles de ce locus). Ce polymorphisme peut être dû soit à la répartition des sites de restriction, soit à la présence de minisatellites diversement répétés. La taille des fragments reconnus dépend de la position des sites de coupures qui encadrent ou sont internes à la région reconnue par la sonde, ou du nombre de répétitions du minisatellite.

Le principe du RFLP est assez simple. Après amplification de la région d'intérêt par PCR, on réalise la digestion des fragments obtenus par l'enzyme (les enzymes) de restriction et on effectue une migration en gel d'agarose ou polyacrylamide, suivie de la visualisation au bromure d'éthidium sous UV. Cette méthode peut s'appliquer à tous les fragments d'ADN à la condition qu'ils ne soient pas trop grands, à condition que la digestion ne génère qu'un nombre limité de fragment. Chez les eucaryotes, les petits génomes sont représentés par les génomes mitochondriaux et chloroplastiques. On purifie donc les mitochondries ou les chloroplastes puis on isole leur ADN.

Les applications de la technique RFLP sont nombreuses et variées, parmi les plus utilisées étant la détermination de la paternité, analyse de l'identité en criminalistique, analyse de la transmission des maladies héréditaires ou établissement de cartes génétiques.

2.3. Amplified Fragment Length Polymorphism AFLP

Le principe de cette technique est schématisé dans la figure 2. L'ADN génomique est digéré par deux enzymes. Des adaptateurs sont ajoutés aux extrémités des sites de restriction. Ces adaptateurs permettent ensuite l'hybridation d'amorces spécifiques (la séquence de l'amorce est complémentaire à celle de l'adaptateur) pour une série d'amplifications PCR sélectives : on ajoute 1 à 3 bases d'ancrage en plus de la séquence de l'adaptateur, qui vont ne permettre l'hybridation des amorces que sur un sous-ensemble des fragments d'ADN présents dans le mélange. On diminue ainsi la complexité du mélange de fragments, de manière à obtenir un profil constitué de plusieurs dizaines de bandes distinctes. Le mélange obtenu est, par exemple, analysé sur un gel de polyacrylamide sur séquenceur automatique, grâce au marquage d'une des deux amorces en fluorescence. La présence ou l'absence de certaines bandes entre individus met en évidence autant de loci polymorphes (figure 10). On a donc des marqueurs multilocus, bialléliques, dominants ou codominants suivant la technique d'analyse utilisée.

2.4. ADN polymorphe amplifié au hasard (Random Amplified Polymorphic DNA-RAPD)

La technique RAPD est basée sur la méthode PCR. Leur principe repose sur l'amplification du DNA génomique est, cette fois-ci, réalisée à partir d'amorces de séquences aléatoires (10 bases, environ) et pouvant s'hybrider en plusieurs endroits du génome. Si deux amorces se fixent sur deux sites distants de moins de 1 kb et situés sur l'un et l'autre brin de l'ADN matrice, un fragment spécifique de ce locus sera amplifié. On obtient ainsi un profil multibandes (10 à 20 bandes) pour chaque individu. Le polymorphisme, qui peut être une mutation dans le site d'hybridation de l'amorce, une insertion ou une délétion, est visualisé par présence ou absence d'une même bande entre deux profils (figure 11).

Les problèmes de reproductibilité des profils entre laboratoires (utilisation de conditions PCR non stringentes) ont défavorisé leur utilisation.

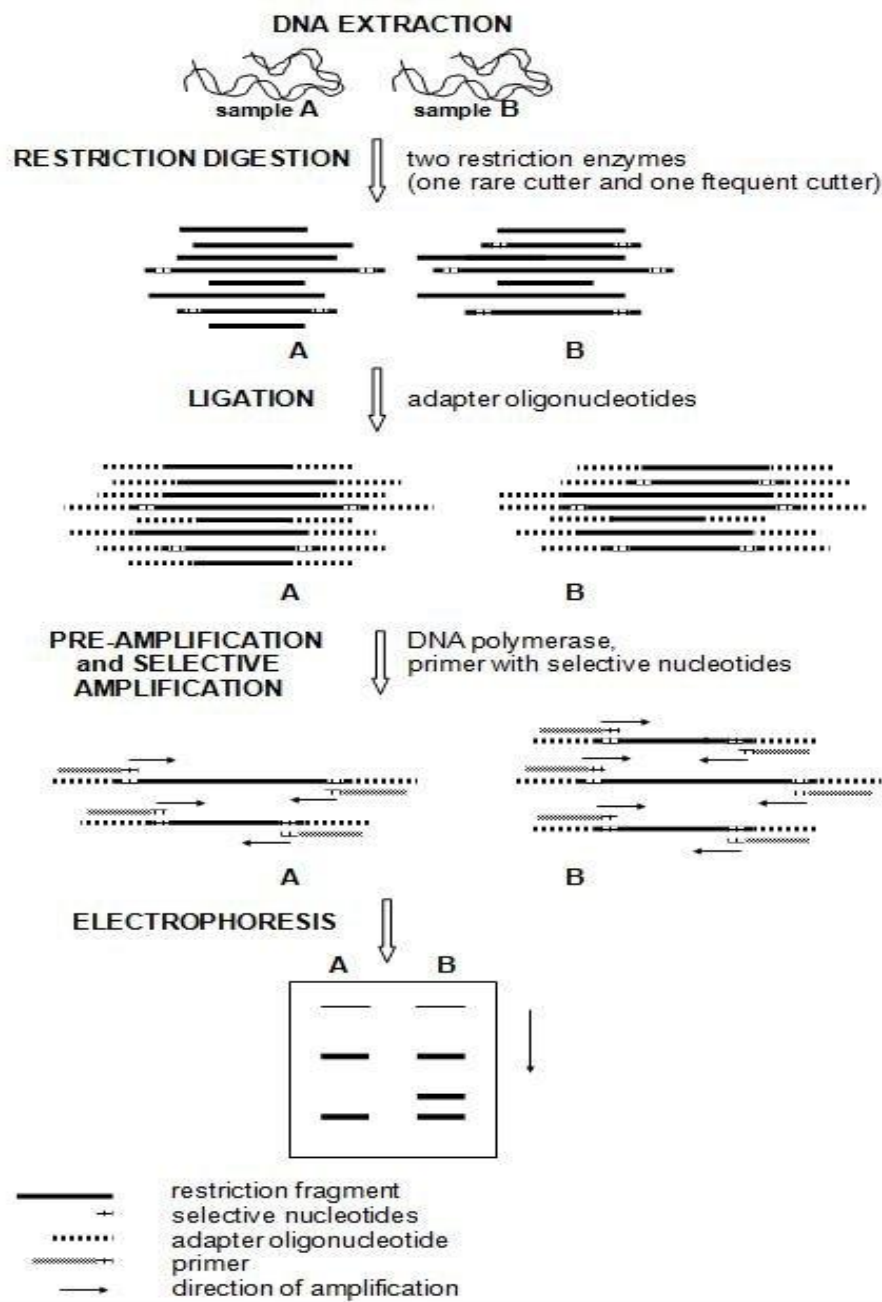
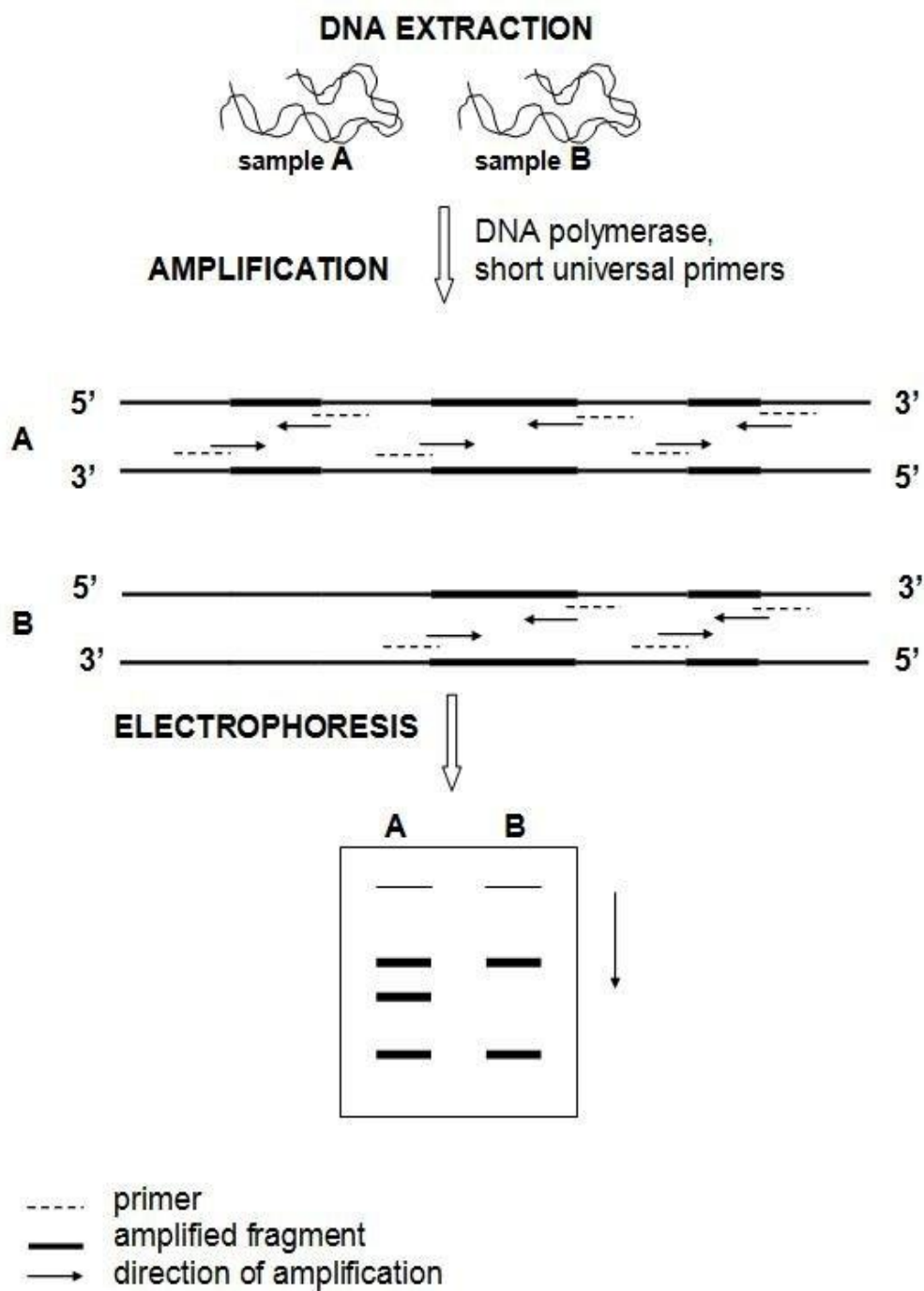


Figure 10. Principe de l'AFLP



a

Figure 11. Principe de la technique RAPD.

2.5. Les microsatellites ou SSR

Les différentes parties du génome sont plus ou moins variables, les parties codantes et les promoteurs présentent des séquences très conservées par contre, les parties non codantes sont plus variables. Mais il existe des éléments qui sont encore plus variables, ce sont les microsatellites (répétitions de 1-4 bases) et les minisatellites (ou VNTR : Variable Number Tandem Repeat- répétitions de 8-15 bases). Les plus étudiés sont les microsatellites.

Les microsatellites ou SSR sont constitués de séquences de di-, tri- ou tétra-nucléotides répétés en tandem (toujours dans le même sens, par opposition à répétitions inversées répétées). Ces éléments sont uniformément répartis en plusieurs exemplaires sur l'ensemble du génome d'une espèce et présentent un taux de polymorphisme élevé. Ce polymorphisme repose sur la variation du nombre d'unités de répétition constituant le microsatellite. Les clones d'ADN qui s'hybrident facilement contiennent des séquences répétées. Pour que ce DNA marqueur puisse être un repère non ambigu, il doit être entouré à droite et à gauche de séquences uniques. Ces deux séquences flanquant les éléments répétés permettent de définir une paire d'amorces qui sera utilisée pour l'amplification PCR. L'analyse des produits amplifiés s'effectue sur gel de polyacrylamide haute résolution (susceptible de séparer des fragments d'ADN de taille faible) (figure 12).

Un microsatellite (SSR) est donc défini par:

1. le motif répété qui le compose et

2. la paire d'amorces uniques qui l'encadrent et servent à l'identifier et à l'amplifier.

Si les SSR constituent de bons marqueurs moléculaires (reproductibles, codominants et aisés d'utilisation), leur caractérisation initiale est toutefois assez lourde. En effet, leur production doit passer d'abord par le clonage et le séquençage de ces éléments répétés.

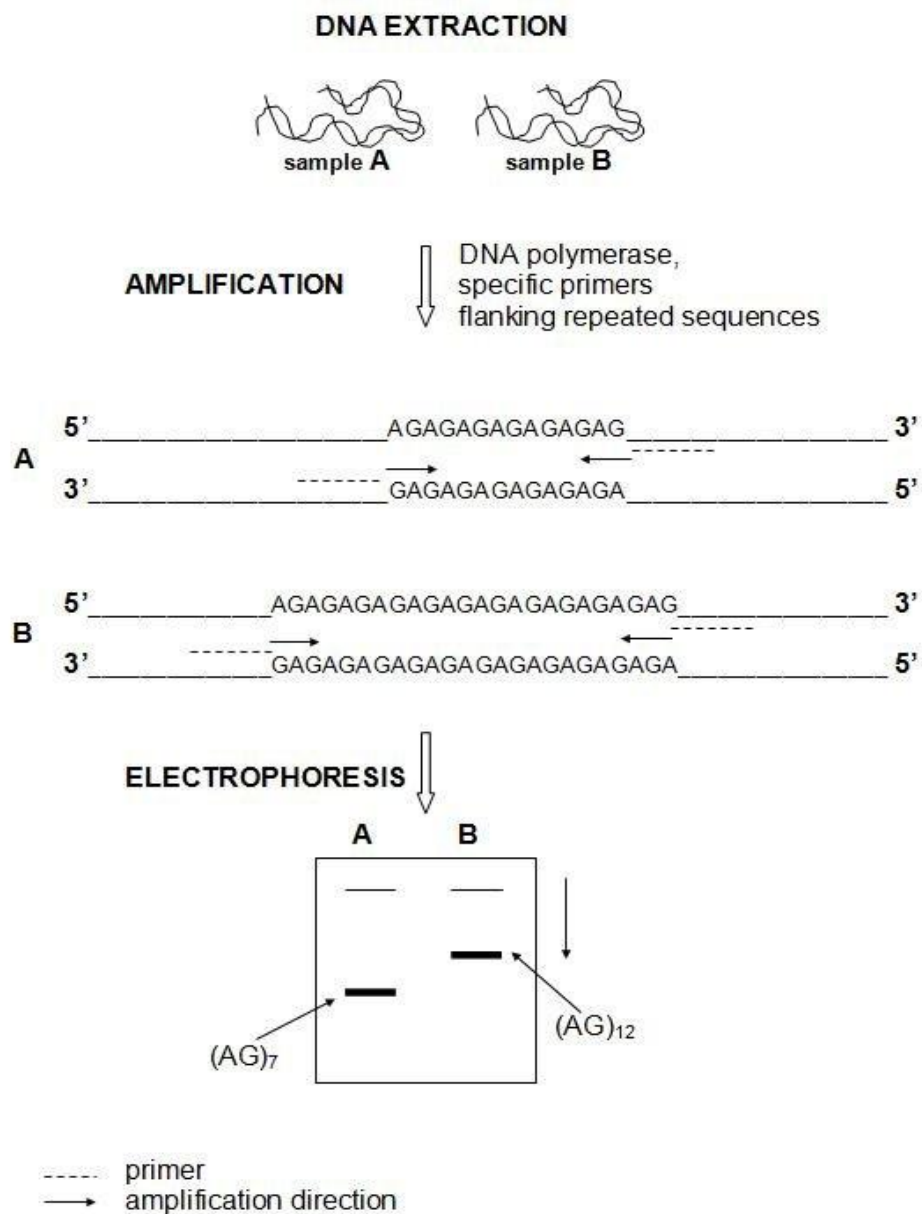


Figure 12. Principe de la technique SSR (L'exemple considère un microsatellite avec une unité de base di-nucléotidique (AG) n)

3. Les techniques d'analyse du transcriptome

L'analyse simultanée des niveaux d'expression de centaines ou de milliers de gènes dans un échantillon biologique donné peut être effectuée de différentes manières. Deux techniques majeures sont disponibles à l'heure actuelle : les puces à ADN, et la méthode SAGE.

L'étude du transcriptome global permet d'obtenir un profil (qualitatif et quantitatif) de l'expression des gènes d'un échantillon (cellules, tissus, biopsies) à un moment donné ou dans une condition physiologique ou pathologique déterminée. Ce profil constitue dès lors une caractéristique intrinsèque à l'échantillon rendant possible de nombreuses applications : découverte de nouvelles cibles thérapeutiques, recherche de biomarqueurs, étude de mécanismes d'action ou de voies de signalisation, connaissances fondamentales...

3.1. Puces à ADN

Cette technique consiste, à préparer une population de sondes ADNc fluorescentes ou radioactives par rétrotranscription des ARN de l'échantillon biologique, et à l'hybrider à une collection de fragments d'ADN déposés et fixés à un support solide selon une matrice à haute densité. Ce support est une lame de verre ou un filtre de nylon ou de plastique.

La quantification des niveaux d'ARNm d'un échantillon biologique nécessite de les marquer soit par un fluorochrome, soit par un atome radioactif. Cela est généralement réalisé par transcription inverse en présence de nucléotides couplés aux cyanines Cy3 ou Cy5 ou de nucléotides marqués au ^{33}P . Le ^{33}P est préféré au ^{32}P car il donne un signal beaucoup plus résolu, évitant des "débordements" d'un spot à son voisin lors de la révélation. Des imageurs β à haute résolution permettent de recueillir les valeurs quantitatives des signaux d'hybridation avec une excellente résolution spatiale et une grande fiabilité dans la mesure (les signaux enregistrés sont proportionnels à la quantité de radioactivité présente, de façon linéaire sur plusieurs log d'intensité).

Les techniques fluorescentes sont particulièrement adaptées à la comparaison d'échantillons deux à deux. En effet, il est possible de marquer une population d'ADNc à l'aide d'un fluorochrome vert (Cy5, par exemple) et l'autre population d'ADNc à l'aide d'un fluorochrome rouge (Cy3, par exemple). Les deux populations sont ensuite mélangées en proportions égales et hybridées sur le support porteur de la ressource d'ADN. La lecture des signaux de fluorescence pour chacune des longueurs d'onde caractéristiques des deux fluorochromes permet ensuite de mesurer le rapport d'intensités pour chaque gène étudié. Ce

rapport est normalisé à 1 pour tous les gènes dont le niveau d'expression est stable entre les deux conditions, et sera supérieur ou inférieur à 1 selon que le gène considéré est induit ou réprimé. Le résultat est généralement représenté en code couleur permettant de repérer immédiatement les gènes régulés (figure 13)

Les différentes puces à ADN

- Les filtres à haute densité (macroarrays) sont des plaques en nylon (12X8 cm) qui permettent de quantifier la présence de 2400 gènes par marquage radioactif d'un échantillon (ARNm totaux à étudier).
- Les lames de verre (microarrays) permettent de quantifier la présence de 10 000 gènes par marquage fluorescent avec deux conditions expérimentales par lame (voir infographie).
- Les puces à oligonucléotides (puce de 1,28cm x 1,28cm) contiennent 300 000 oligonucléotides (sondes) par lame et permettent de quantifier la présence d'ARNm grâce au marquage fluorescent de ces cibles issues d'une seule condition expérimentale.

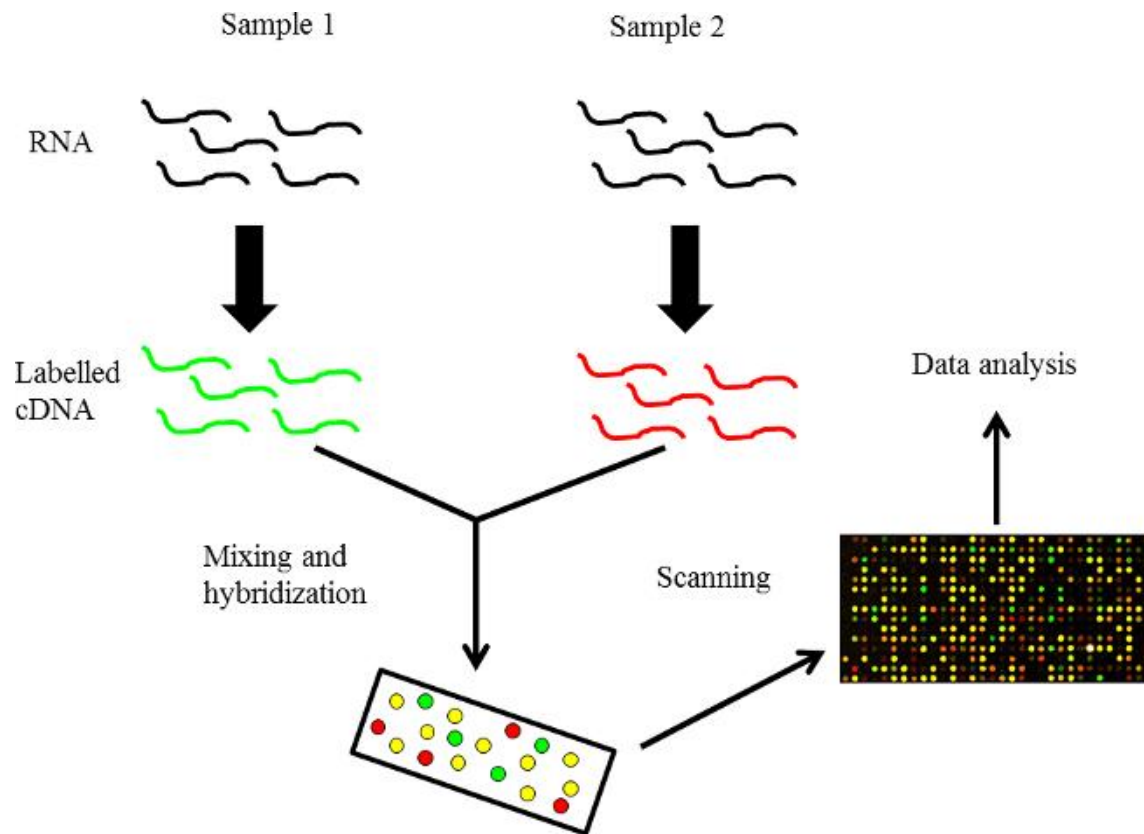


Figure 13. Principaux étapes de la technologie des puces à ADN.

3.2. Technique SAGE (*serial analysis of gene expression*)

Cette technique permet d'atteindre le même objectif que la méthode précédente par une approche expérimentale totalement différente.

Elle repose sur deux principes :

- une courte séquence nucléotidique de 10-14 pb (étiquette) contient suffisamment d'informations pour identifier un transcrit, à la condition que l'étiquette soit obtenue pour chaque transcrit à partir d'une position unique et reconnaissable ;
- les étiquettes de séquences peuvent être liées les unes aux autres pour former de grandes molécules qui peuvent être clonées et séquencées.

Comme l'indique la figure 14, la technique SAGE consiste donc à amplifier, par RT-PCR à grande échelle, des étiquettes de courtes séquences nucléotidiques à partir de positions précises. Ces étiquettes sont ensuite fusionnées et les polymères obtenus sont séquencés. La quantification du nombre de fois où une étiquette donnée est retrouvée dans les séquences mesure le niveau d'expression du transcrit correspondant. La limitation majeure de cette technique vraiment performante est le coût très élevé des opérations de séquençage. Pour la

comparaison du profil d'expression de deux échantillons d'ARN, la séquence d'environ 5 à 10 mégabases est nécessaire pour obtenir une représentation statistiquement correcte de la fréquence de chaque étiquette. Une autre limitation de la méthode est la quantité d'ARN nécessaire pour l'analyse : 2,5 à 5 µg d'ARNm, soit 250 à 500 µg d'ARN totaux.

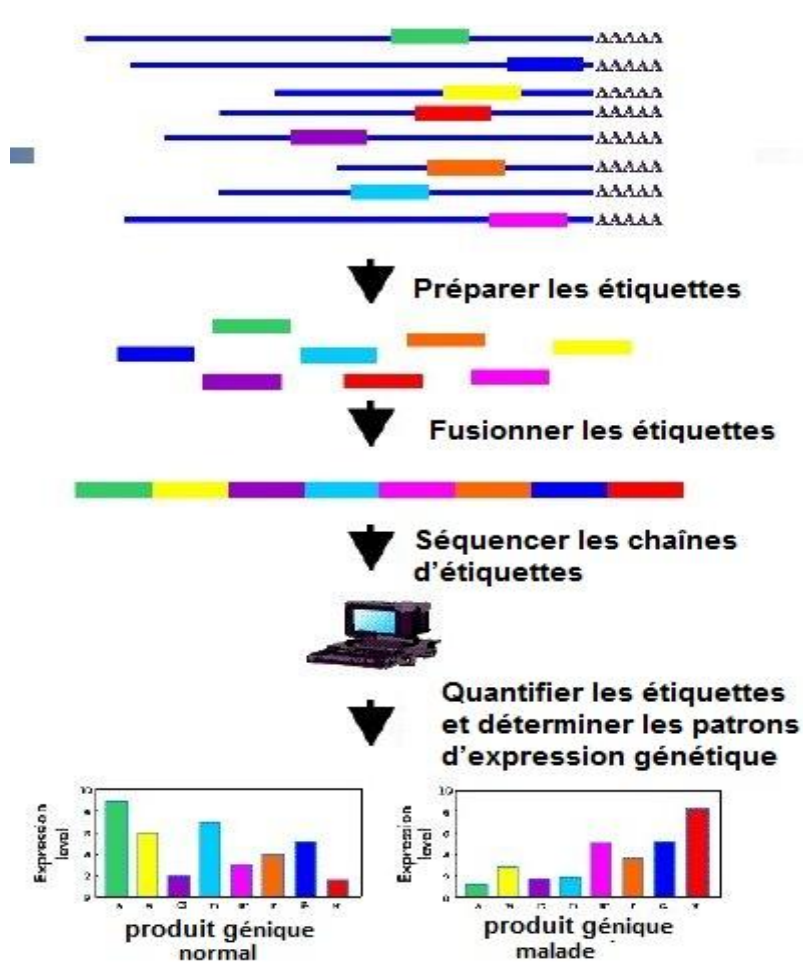


Figure 14. Principe de la technique SAGE.

Références bibliographiques

- Addamiano M. C. (2016). Synthèse et caractérisation de nucléotides et oligonucléotides modifiés pour l'obtention de structures capables de mimer l'activité enzymatique des protéases à sérine. Doctoral dissertation, Université de Toulouse, Université Toulouse III-Paul Sabatier).
- Anonyme. Cours de biochimie, Université d'Angers :
« [L'interférence ARN](http://biochimej.univ-angers.fr/Page2/COURS/3CoursdeBiochSTRUCT/3RNAi/1RNAi.htm) » <http://biochimej.univ-angers.fr/Page2/COURS/3CoursdeBiochSTRUCT/3RNAi/1RNAi.htm>
- Babinet C. (1995). Un nouveau pas dans l'utilisation du système Cre-LoxP chez les cellules souches embryonnaires de souris: la création de remaniements chromosomiques. *Mini-Synthèse*, 11 : 1154-7.
- Babinet C., & Cohen-Tannoudji M. (2000). Vingt ans d'interventions délibérées sur le génome de la souris: une révolution dans l'approche génétique de la biologie des mammifères. *médecine/sciences* ; 16 : 31-42
- Baldi, L., Hacker, D.L., Adam, M. et Wurm, F.M. (2007). Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *Biotechnol Lett.* 29(5): 677-84
- Baranwal D. K., Singh P., Singh R. K., & Shekhar S. Gene knockout technology and its application. *Biologix*, 55.
- Ben Naya R. (2013). Exploration des systèmes d'expression de protéines recombinantes pour la caractérisation d'un anticorps catalytique Thèse de doctorat université de technologie de Compiègne, pp :133.
- Berlec A. et Strukelj B. (2013). Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J Ind Microbiol Biotechnol.* 40(3-4): 257-74.
- Birch, J.R. et Racher, A.J. (2006). Antibody production. *Adv Drug Deliv Rev.* 58(5-6): 671-85
- Boudaud D. (2018). Interférence ARN : utilisation thérapeutique et vectorisation , thèse doctorat de pharmacie, Boudaud David, Université d'Angers.
- Creusat, G. (2009). L'ARN interférence, l'émergence d'une nouvelle stratégie thérapeutique. Doctoral dissertation, UHP-Université Henri Poincaré.

- Cummings, P. J., Ahmed, R., Durocher, J. A., Jessen, A., Vardi, T., Obom, K. M. Pyrosequencing for Microbial Identification and Characterization. *J. Vis. Exp.* (78), e50405, doi:10.3791/50405 (2013).
- Davis, M.E. (2002). Non-viral gene delivery systems. *Curr Opin Biotech.* 13(2): 128-31.
- Dudley A.T., Lyons K.M., and Robertson E.J. (1995). A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. *Genes Dev* 9, 2795-2807.
- Ecochard V, Fournier D, Nieto L, & Paquereau L. (2011). Techniques et Stratégies en Biologie Moléculaire Deuxième partie cours de master.
- Faye, L., Landry, N., Lerouge, P., Gomord, V., & Vézina, L. P. (2001). La production de protéines à usage biopharmaceutique dans les plantes. *Synthese*, 17 : 867-77
- Feige, J. J. (2003). Analyse du transcriptome: intérêt en endocrinologie. *Métabolismes Hormones Diabètes et Nutrition (VII)*, no 2, mars/avril , 2 : 67-71.
- Fiola K. (2005). Structure, mécanisme et applications du ribozyme delta. these de Doctorat. Université de Sherbrooke. pp : 253.
- Furelaud G. (2002). L'invalidation d'un gène : le knock-out <https://planet-vie.ens.fr/thematiques/cellules-et-molecules/l-invalidation-d-un-gene-le-knock-out>
- Grosse S., Thévenot G., Monsigny M. et Fajac I. (2006). Which mechanism for nuclear import of plasmid DNA complexed with polyethylenimine derivatives? *J Gene Med.* 8(7): 845-51
- Guo, S., & Kemphues, K. J. (1995). par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*, 81(4), 611-620.
- Hannig G et Makrides SC. (1998). Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends Biotechnol.* 16(2): 54-60.
- Hélène C., & Saison-Behmoaras E. (1994). La stratégie anti-sens: nouvelles approches thérapeutiques. *médecine/Synthèse* 1 994; JO: 253-73.
- Jain A., Wang G., & Vasquez K. M. (2008). DNA triple helices: biological consequences and therapeutic potential. *Biochimie*, 90(8) : 1117-1130.
- Kay MA., Glorioso JC et Naldini L (2001). Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nat Med.* 7(1): 33-40.
- Kayser, K., Lin, N., Allison, D., Donahue, L. et Caple, M. (2006). Cell line engineering methods for improving productivity. *BioProcess International Magazine*. Repéré à http://www.bioprocessintl.com/wp-content/uploads/bpicontent/0643ar01su_77541a.pdf

- Lewin A S, & Hauswirth W. W. (2001). Ribozyme gene therapy: applications for molecular medicine. *Trends in molecular medicine*, 7(5), 221-228.
- Mariati Ho S.C., Yap M.G. et Yang Y. (2010). Evaluating post-transcriptional regulatory elements for enhancing transient gene expression levels in CHO K1 and HEK293 cells. *Protein Expr Purif.* 69(1): 9-15
- Negura L, Carasevici E., HUM AM, & Artenie V. (2006). Méthodologie actuelle de la biologie moléculaire pour la détection des mutations. *Analele tiinifice ale Universit ii Alexandru Ioan Cuza Ia i (Serie nou), Sec iunea Genetic i Biologie Molecular*, 7, 13-20.
- Paris S., Burlacu A. et Durocher Y. (2008). Opposing roles of syndecan-1 and syndecan-2 in polyethyleneimine-mediated gene delivery. *J Biol Chem.* 282(12): 7697-704.
- Pasqualone A. (2013). Cultivar identification and varietal traceability in processed foods: A molecular approach. *Cultivars: Chemical Properties, Antioxidant Activities and Health Benefits*, 83-105.
- Pham P.L., Kamen A. et Durocher Y. (2006). Large-scale transfection of mammalian cells for the fast production of recombinant protein. *Mol Biotechnol.* 34(2): 225-37
- Romanos M. (1995). Advances in the use of *Pichia pastoris* for high-level gene expression. *Current Opinion in Biotechnology*, 6:527-533.
- Schubert S, & Kurreck J. (2004). Ribozyme-and deoxyribozyme-strategies for medical applications. *Current drug targets*, 5(8) : 667-681.
- Szurman-Zubrzycka M, Chmielewska B, Gajewska P, Szarejko I. (2017). Mutation Detection by Analysis of DNA Heteroduplexes in TILLING Populations of Diploid Species. In: Jankowicz-Cieslak J., Tai T., Kumlehn J., Till B. (eds) *Biotechnologies for Plant Mutation Breeding*. Springer, Cham. https://doi.org/10.1007/978-3-319-45021-6_18
- Vaish N. K., Kore A. R., & Eckstein F. (1998). Recent developments in the hammerhead ribozyme field. *Nucleic acids research*, 26(23) : 5237-5242.
- Wong V., Wong N., Tan H.K., Wang D. et Yap M. (2003). Enhancing Production of Recombinant Proteins from Mammalian Cells. *Molecular Engineering of Biological and Chemical Systems*. Repéré à <http://hdl.handle.net/1721.1/3782>
- Yang Z., L. Zhang Y, Zhang T et al.(2011). Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. *PLoS One*. 6:e22981.