



الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة وهران للعلوم والتكنولوجيا محمد بوضياف
كلية علوم الطبيعة والحياة
قسم الوراثة الجزيئية التطبيقية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
Université des Sciences et de la Technologie d'Oran "Mohamed BOUDIAF"
Faculté des Sciences de la Nature et de la Vie
Département de Génétique Moléculaire Appliquée

Polycopié de cours

Biologie Moléculaire et Génie Génétique

Cours destinés aux étudiants de 3^{ème} année Licence Génétique

Elaboré par :

Dr. BOUSHABA Nadjat

Maître de Conférences B

Année universitaire : 2021/2022

Avant-propos

Les notes de cours ont pour fonction de transmettre des connaissances, de proposer une démarche d'apprentissage, de suggérer une méthodologie, de susciter la participation active des étudiants et de guider l'apprentissage. Ce polycopié est un support de cours de la matière de Biologie Moléculaire et Génie Génétique. Il est destiné comme support pédagogique aux étudiants de la troisième année Licence, semestre 5, spécialité "Génétique", année universitaire 2021-2022 suivant le canevas ministériel en complément des notions générales de Génétique Moléculaire et Biochimie déjà acquises par les étudiants en deuxième année de Licence en Biologie du système LMD. La matière de Biologie Moléculaire et Génie Génétique fournira, à travers des exemples concrets, un aperçu des développements des techniques de Biologie Moléculaire et de ses applications dans les sciences de la vie.

Les cours serviront d'introduction à la réalisation d'exercices par les étudiants sous forme de travaux dirigés qui illustreront l'éventail des approches basées sur la Biologie Moléculaire et Génie Génétique. L'objectif de ce polycopié de cours est d'initier les étudiants au langage et aux concepts de ce domaine. A l'issue de cette matière, les étudiants auront une connaissance des différentes méthodologies existantes sur lesquelles elles sont fondées. Ils seront capables de choisir et mettre en œuvre ces méthodologies en fonction des informations dont ils disposent (données de génétiques, type de matériel biologique ; animal, végétal, micro-organisme, ...).

Cet enseignement permettra aux étudiants de faire le lien entre les différents enseignements de biologie et leurs applications en biotechnologie et recherche fondamentale et appliquée.

Résumé

Envisager de façon cohérente les multiples possibilités d'usage d'un même enseignement amène à réfléchir et conceptualiser une approche qui permet de constituer des ressources pédagogiques pouvant être automatiquement mises en forme pour les supports visés : que ce soit un polycopié rédigé qui sera imprimé ou visualisé dynamiquement sur une plateforme. Le but de ce polycopié est l'élaboration conceptuelle d'un cours de Biologie Moléculaire et Génie Génétique. Ce polycopié est divisé en deux parties présentant : les principes élémentaires et fondamentaux de la Biologie Moléculaire, les propriétés physiques et chimiques des acides nucléiques, l'expression et la transmission des informations génétiques ; les principales méthodes expérimentales d'étude. La deuxième partie traite des outils du Génie Génétique et les applications possibles notamment en santé humaine (production de protéines thérapeutiques, élimination de séquences virales persistantes), en agriculture biotechnologique (mise au point de nouvelles générations de plantes génétiquement modifiées) ou pour la mise au point d'outils destinés à explorer la fonction d'un gène. Le polycopié se termine par des références bibliographiques.

Mots clés : Acides nucléiques - Biologie Moléculaire - Génie Génétique.

Abréviations

ADN	Acide D ésoxyribo N ucléique
ADNc	Acide D ésoxyribo N ucléique C omplémentaire
ADNg	Acide D ésoxyribo N ucléique g énomique
ARN	Acide R ibo N ucléique
ARNm	Acide R ibo N ucléique m essenger
ARNr	Acide R ibo N ucléique r ibosomique
ARNt	Acide R ibo N ucléique de t ransfert
BET	B romure d' E thidium
dNTP	d ésoxy N ucléotides T ri P hosphates
DO	D ensité O ptique
EDTA	Ethylène D iamine T étra- A cétique
Kb	K ilobase
M	M olaire
ml	M illitre
mn	M inute
NaCl	Sodium C hlorure
nm	N anomètre
OGM	O rganisme G énétiquement M odifié
pb	p aire de b ases
pH	p otentiel H ydrogène
PCR	<i>"Polymerase Chain Reaction"</i>
RFLP	<i>"Restriction Fragments Length Polymorphism"</i>
SNP	<i>"Simple Nucleotide Polymorphism"</i>
Taq	<i>"Termus aquaticus"</i>
Tm	<i>Melting Temperature</i>
°C	D egré C elsius
µg	M icrogramme
µl	M icrolitre
%	P ourcentage

Liste des tableaux

Tableau 1 : Nomenclature des nucléosides	08
Tableau 2 : Exemples de nomenclature des enzymes de restriction	28
Tableau 3 : Différence entre une banque d'ADNg /ADNc	63
Tableau 4 : Différents vecteurs et leurs hôtes	77

Liste des figures

Figure 01 : Différence entre ribose et désoxyribose	06
Figure 02 : Acides phosphoriques	06
Figure 03 : Bases puriques	07
Figure 04 : Bases pyrimidiques	07
Figure 05 : Structure d'un nucléotide	08
Figure 06 : Structure secondaire de la double hélice	09
Figure 07 : Formes alternatives de la double hélice	10
Figure 08 : Différents niveaux d'organisation de la chromatine	11
Figure 09 : Mise en évidence de la séparation des deux brins d'un ADN et mesure de la température de fusion T_m	12
Figure 10 : Erreurs de réplication	16
Figure 11 : Dépuration	16
Figure 12 : Fourche de réplication de l'ADN	24
Figure 13 : Modèle de réparation des mésappariements chez <i>E. coli</i>	26
Figure 14 : Système restriction/modification <i>EcoRI</i>	27
Figure 15 : Carte de restriction et test de digestion dans un gel d'agarose	31
Figure 16 : Principe de la technique d'hybridation sur gel "blot"	32
Figure 17 : Estimation de la taille des fragments d'ADN dans un gel d'agarose	32
Figure 18 : Représentation schématique d'un échange génétique	33
Figure 19 : Transposons et rétrotransposons	36
Figure 20 : Structure d'un transposon d'ADN	36
Figure 21 : Conformation d'une molécule d'ARN	39
Figure 22 : Etape de la transcription	40
Figure 23 : Structure de l'ARN de transfert (ARNt)	43
Figure 24 : Etape de la traduction d'un ARNm	44

Figure 25 : Fragments d'ADN séparés par électrophorèse dans un gel d'agarose	50
Figure 26 : Spectre d'absorption caractéristique d'ADN pur	51
Figure 27 : Procédé général du génotypage d'un marqueur SNP sur une puce Illumina en utilisant la technologie <i>Infinium assay</i>	52
Figure 28 : Amplification génique par PCR	53
Figure 29 : Méthode de séquençage selon Maxam et Gilbert	55
Figure 30 : Méthode de séquençage de Sanger	56
Figure 31 : Polymorphisme de longueur de fragments de restriction	57
Figure 32 : Synthèse d'ADNc double brin à partir d'ARNm	59
Figure 33 : Principe de construction d'une banque plasmidique d'ADN génomique total	60
Figure 34 : Principe du Clonage de l'ADNc	62
Figure 35 : Composants basiques d'un vecteur plasmidique chez <i>E. coli</i>	64
Figure 36 : Plasmide pBR322	65
Figure 37 : Plasmide pUC18	65
Figure 38 : Fixation d'un phage	66
Figure 39 : Différents cycles de répllication des phages	67
Figure 40 : Formes linéaires et circulaires de l'ADN λ	68
Figure 41 : Clonage dans le phage	69
Figure 42 : Cosmide typique pour cloner de longs fragments d'ADN	71
Figure 43 : Structure du vecteur PAC	72
Figure 44 : Vecteur pBeloBAC11	73
Figure 45 : Vecteur YIp5	73
Figure 46 : Plasmides de levure	74
Figure 47 : Vecteur YAC	75
Figure 48 : Insertion d'ADN étranger dans un vecteur YAC	76
Figure 49 : Gale de collet	80
Figure 50 : Représentation simplifiée des principales régions du plasmide <i>Ti</i> d' <i>Agrobacterium tumefaciens</i>	81
Figure 51 : Un fragment d'ADN étranger inséré dans un plasmide en utilisant des enzymes de restriction et l'ADN ligase	83
Figure 52 : Insertion d'un insert dans un vecteur par des homopolymères	84
Figure 53 : Absorption d'ADN plasmidique par une cellule bactérienne compétente	86
Figure 54 : Transformation bactérienne par un plasmide	89

Figure 55 : Sélection des clones recombinants par complémentation	90
Figure 56 : Principe de l'intégration ciblée d'un gène en un site donné du génome	93
Figure 57 : Marquage de l'ADN par <i>nick-translation</i>	94
Figure 58 : Marquage de l'ADN par amorçage au hasard	94
Figure 59 : Méthode de détection des clones recombinants par hybridation moléculaire des colonies	95
Figure 60 : Utilisation d'un anticorps purifié pour détecter des protéines dans des colonies recombinantes	96
Figure 61 : Trois principales méthodes de transgénèse	97
Figure 62 : Principales étapes de l'obtention de souris transgéniques par micro-injection pronucléaire	98
Figure 63 : Transgénèse utilisant les cellules <i>ES</i>	100
Figure 64 : Clonage reproductif du mouton "Dolly"	102
Figure 65 : Remplacement de l'ADN-T par un gène d'intérêt permet d'envisager une technique de transgénèse	104
Figure 66 : Transformation des cellules végétales par <i>Agrobacterium tumefaciens</i> recombinants	105
Figure 67 : Principe général de la recombinaison homologue	110
Figure 68 : Mutagenèse dirigée en utilisant la PCR	111

Table des matières

Avant-propos	i
Résumé	ii
Abréviations.....	iii
Liste des tableaux	iv
Liste des figures	v
Liste des matières	viii
Introduction	01
A. Biologie Moléculaire	06
Chapitre I. Support de l'information génétique, l'ADN	06
1. Structure et dynamique de l'ADN.....	06
1.1 Structure de base	06
1.2 Structure secondaire.....	09
1.3 Formes alternatives de la double hélice	09
1.4 Interactions avec les protéines	10
1.5 Manipulations topologiques.....	11
1.6 Dénaturation-renaturation	12
1.7 Implications biologiques.....	13
2. Structure et organisation du génome dans le noyau	13
Chapitre II. Mutations, mutagenèse et détection.....	15
1. Mutations géniques	15
1.1 Définition	15
1.2 Intérêt des mutations	15
1.3 Réarrangements génétiques des mutations	15
1.4 Mutations naturelles	16
1.5 Mutations induites.....	17
1.6 Agents mutagènes	17
1.7 Effets des mutations.....	18
1.8 Expression des mutations.....	18
1.9 Réversions et suppressions	19
2. Mutagenèse physique, chimique et biologique et techniques de modification du matériel génétique	19
2.1 Mutagenèse physique	19
2.2 Mutagenèse chimique.....	20
2.3 Mutagenèse biologique	21
2.4 Techniques de modification du matériel génétique	21
3. Détection et mise en évidence des mutations (diagnostic génotypique).....	22
Chapitre III. Transmission et maintien de l'information génétique	23
1. Réplication de l'ADN et sa régulation	23
2. Réparation de l'ADN et détection du pouvoir mutagène.....	25
3. Systèmes de restriction-modification	26
4. Cartes de restriction.....	30
5. Intérêt et analyse du polymorphisme de restriction	31
Chapitre IV. Mécanismes moléculaires de la recombinaison.....	33
1. Recombinaison homologue	33

2. Cartographie	33
3. Analyse et construction génétique	34
4. Recombinaison à un site spécifique	34
5. Eléments génétiques mobiles (transposons et rétrotransposons)	35
6. Utilisation des transposons	37
6.1 Marquage	37
6.2 Mutagenèse	37
6.3 Clonage	37
6.4 Mobilisation du matériel génétique	38
6.5 Carte génétique	38
Chapitre V. Expression de l'information génétique et son contrôle	39
1. Structure de l'ARN	39
2. Transcription et maturation de l'ARN	40
3. Traduction et maturation des protéines	42
4. Régulation du fonctionnement et de l'expression des gènes	45
4.1 Structure chromatinienne des gènes actifs	45
4.2 Modification de la structure primaire de l'ADN	46
4.3 Régulation transcriptionnelle et post-transcriptionnelle	46
4.4 Régulation traductionnelle et post-traductionnelle	47
5. Voies de régulation des gènes par les signaux extracellulaires	48
Chapitre VI. Méthodologie en Biologie Moléculaire	49
1. Méthodes de caractérisation et analyse de l'ADN	49
1.1 Extraction et purification de l'ADN	49
1.2 Fragmentation	49
1.3 Séparation analytique	50
1.4 Visualisation	50
1.5 Quantification	51
1.6 Hybridation et microarrays	52
1.7 Amplification (la PCR et ses applications)	53
1.8 Séquençage	54
1.9 Restriction et analyse des polymorphismes	56
1.10 Interaction avec les protéines	57
B. Génie Génétique	58
Chapitre I. Sources et préparation de l'ADN à cloner	58
1. ADN génomique	58
2. ADN complémentaire	59
3. ADN synthétique	59
4. Notion de banques d'ADN génomique et complémentaire	59
4.1 Banque d'ADN génomique	59
4.2 Banque ADNc	61
Chapitre II. Vecteurs de clonage	64
1. Vecteurs bactériens	64
1.1 Plasmides	64
1.2 Phages	66
1.3 Cosmides	70
1.4 PAC	72
1.5 BAC	72

2. Vecteurs de clonage dans la levure	73
2.1 Vecteurs intégratifs	73
2.2 Vecteurs autonomes dérivés du chromosome ou du plasmide 2 μ m	74
2.3 Chromosomes artificiels	74
3. Vecteurs eucaryotes (cellules animales)	77
3.1 Plasmides non répliatifs	77
3.2 Plasmides répliatifs	77
3.3 Virus.....	78
3.4 Chromosomes artificiels	79
4. Vecteurs eucaryotes (cellules végétales) : Plasmide Ti et ADN-T	80
Chapitre III. Procédés de ligation du vecteur et de l'ADN à cloner	82
1. Enzymes ligases	82
2. Procédés de ligation	82
Chapitre IV. Transfert de l'ADN dans les cellules	85
1. Transfert direct	85
1.1 Biolistique	85
1.2 Micro-injection	85
2. Transformation/transfection	85
2.1 Méthodes chimiques : au chlorure de calcium (bactéries).....	85
2.2 Co-précipitation de l'ADN et du phosphate de calcium.....	86
2.3 DAEA-dextran	86
2.4 Polycation-DMSO (cellules eucaryotes)	86
2.5 Fusion des protoplastes	87
2.6 Lipofection.....	87
2.7 Peptides.....	87
2.8 Electroporation.....	88
2.9 Transduction virale (encapsulation <i>in vitro</i>).....	88
Chapitre V. Sélection des transformants recombinants	89
1. Sélection par complémentation	89
2. Sélection par marqueurs dominants de résistance à des agents toxiques.....	91
3. Sélection par des marqueurs métaboliques	91
Chapitre VI. Criblage pour la détection des clones d'intérêt.....	92
1. Complémentation génétique.....	92
2. Recombinaison.....	92
3. Hybridation des acides nucléiques	93
4. Détection des produits d'expression	96
Chapitre VII. Notions de transgénèse animale et végétale	97
1. Transgénèse animale	97
2. Transgénèse végétale	104
Chapitre VIII. Mutagenèse aléatoire et dirigée	109
1. Notion et principe.....	109
1.1 Mutagenèse aléatoire	109
1.2 Mutagenèse dirigée par PCR	110
Références bibliographiques	112

1. Introduction

Le support de cours est un élément indispensable de l'enseignement, qui soutient et illustre le discours de l'enseignant pendant le cours magistral. Il n'est compréhensible qu'avec la narration qu'il accompagne. Les nouveaux supports multimédia de la connaissance permettent aux enseignants universitaires de reconsidérer dans de nouvelles conditions leur politique de structuration pédagogique et de diffusion du savoir. Un enseignant dispose aujourd'hui de nombreuses solutions pour présenter et diffuser le cours dont il a la responsabilité : il peut rédiger un polycopié qui sera imprimé ou projeter le cours sur une plateforme. Cependant, chaque support requiert un mode spécifique d'écriture et de mise en forme du contenu. Le but du polycopié est de fournir une approche et un outil permettant de constituer des ressources pédagogiques pouvant être automatiquement mises en forme pour les différents supports. C'est dans ce contexte que s'inscrit ce polycopié pédagogique proposé aux étudiants de troisième année de la licence LMD qui s'articulera autour de deux parties :

La première partie de Biologie Moléculaire nécessite la reconnaissance des structures de l'acide phosphorique, du ribose et du désoxyribose, des bases puriques et pyrimidiques. Sur des schémas qui seront présentés, l'étudiant doit savoir reconnaître les liaisons impliquées dans la complémentarité des bases ainsi que les éléments structuraux essentiels d'une double hélice de l'ADN (Acide DéoxyriboNucléique), de la chromatine de l'ARN (Acide RiboNucléique) ainsi que la fonction et de la régulation de l'expression des gènes. L'étudiant doit savoir comment l'information génétique d'un organisme vivant est exprimée et régulée aux niveaux de la transcription, la traduction et les modifications post-traductionnelles.

La première partie de Biologie Moléculaire est structurés en six (06) chapitres :

Le premier chapitre définit la Biologie Moléculaire et la structure primaire et secondaire de l'ADN ainsi que les manipulations topologiques.

Le deuxième chapitre décrit les différentes mutations (naturelles, induites) et les agents mutagènes et leurs effets sur l'ADN.

Le troisième chapitre décrit le mécanisme permettant le maintien de l'information génétique au cours du temps dans chaque cellule de l'organisme (réplication de l'ADN, réparation) et la construction une carte génétique (position des sites de restriction sur un fragment d'ADN digéré).

Le quatrième chapitre décrit les mécanismes moléculaires de la recombinaison homologue où les séquences de nucléotides sont échangées entre des molécules d'ADN

identiques (homologues) ou similaires (homéologues) ainsi que les transposons, ou éléments génétiques transposables, participent à l'évolution des génomes, à l'augmentation de leur taille et à la diversité biologique.

Le cinquième chapitre définit les différents types d'ARN et décrit les grandes étapes du passage de l'ADN aux protéines (transcription, maturation, traduction et régulation).

Le sixième chapitre traite les méthodes de caractérisation de l'ADN (extraction et purification, quantification, fragmentation et séparation analytique) ainsi que les techniques de Biologie Moléculaire (*Southern-blot* et PCR).

La deuxième partie de Génie Génétique présente aussi comment les outils du Génie Génétique sont mis à profit pour isoler et caractériser des gènes, les modifier et les transférer entre espèces. Cet enseignement permettra aux étudiants de faire le lien entre les différents enseignements de biologie et leurs applications en biotechnologie et recherche fondamentale et appliquée. Cette partie sera présentée en huit (08) chapitres :

Le premier chapitre définit le terme de Génie Génétique et décrit les différents types d'ADN à cloner (ADNg, ADNc et ADN synthétique) et les notions de banque d'ADN génomique et la banque d'ADN complémentaire et le chapitre VII qui décrit les méthodes de criblage pour la détection des clones d'intérêt.

Le deuxième chapitre décrit les vecteurs bactériens de clonage (plasmides et leurs propriétés générales les vecteurs de clonage bactériens (bactériophages), les vecteurs dans la levure et les vecteurs (cellules animales et végétales).

Le troisième chapitre définit les enzymes de restrictions utilisés en Génie Génétique (origine, nomenclature, mécanisme d'action et la carte de restriction) les procédés de ligation d'un vecteur recombinant dans une cellule.

Le quatrième chapitre décrit le transfert de l'ADN dans les cellules végétales en utilisant des méthodes chimiques, physiques et biologiques.

Le cinquième chapitre décrit la sélection des transformants recombinants par différentes méthodes (complémentation, marqueurs de résistance à des agents toxiques et marqueurs métaboliques).

Le sixième chapitre décrit le criblage par hybridation de l'ADN des clones transformés avec une sonde spécifique afin d'identifier un clone spécifique (parmi une banque génomique ou banque ADNc) et par méthodes immunologiques.

Le septième chapitre décrit les différentes techniques de transgénèse animale et végétale et leurs avantages et leurs inconvénients

Le huitième chapitre décrit deux méthodes possibles pour introduire des mutations dans un génome d'une façon précise et volontaire (mutagenèse dirigée) ou aléatoire (la mutagenèse aléatoire) afin d'obtenir un grand nombre de mutants, pouvant être par la suite sélectionnés.

Enfin, ce polycopié est clôturé par une bibliographie qui englobe les ressources mentionnées.

2. Compétences acquises

Ces connaissances permettront aux étudiants d'acquérir les compétences théoriques nécessaires pour appréhender les concepts utiles à la pratique expérimentale de la Biologie Moléculaire et du Génie Génétique.

À la fin de ces enseignements, l'étudiant doit :

- Avoir une connaissance générale du vocabulaire génétique ;
- Comprendre et savoir utiliser la nomenclature et la structure des acides nucléiques ;
- Maîtriser les bases générales sur les concepts de génome et de gène ;
- Connaître les bases des différentes techniques de manipulation des acides nucléiques ;
- Apprendre comment l'information génétique portée par l'ADN est transmise, réparée et exprimée afin de répondre aux besoins de la cellule ;
- Comprendre les principales techniques de bases rencontrées dans les laboratoires de Biologie Moléculaire telles que l'extraction de l'ADN génomique à partir de cellules eucaryotes, la PCR (*Polymerase Chain Reaction*) ;
- Avoir acquis la notion que des méthodes simples permettent d'extraire et de purifier de l'ADN génomique à partir de cellules eucaryotes.
- Aborder les premières notions sur les processus d'évolution des génomes (mutations géniques et chromosomiques, délétions et duplications, polymorphismes nucléotidiques, éléments transposables...) ;
- Expliquer les méthodologies principales du Génie Génétique ;
- Choisir, en fonction du problème de génie génétique posé, parmi différentes stratégies utilisées pour isoler un gène, le modifier et l'introduire dans d'autres organismes ;
- Proposer une démarche expérimentale intégrée permettant de répondre à des problèmes concrets de génie génétique dans les domaines microbien, animal et végétal ;

- Savoir reconnaître/représenter le mode d'action des principales enzymes impliquées dans la manipulation de l'ADN incluant les endonucléases de restriction, ligases, ADN polymérase ;
- Il doit avoir acquis la notion des problèmes éthiques soulevés par la constitution d'une banque d'ADN génomique humain ;
- Savoir expliquer les grands principes de l'étude de l'ADN génomique en maîtrisant la notion de clonage ;
- Et enfin, être capable d'interpréter des résultats expérimentaux obtenus par les techniques les plus courantes de Biologie Moléculaire appliquées à la recherche biomédicale ou à l'exploration de matériel génétique par les laboratoires de biologie médicale.

3. Méthodes pédagogiques

3.1 Enseignements en présentiel

La matière comprend deux séances de cours (1h30) et une séance de travaux dirigés (1h30) par semaine.

- Semestre : 5
- Unité d'enseignement Fondamentale 2 (UEF 3.1.2) : Biologie Moléculaire
- Matière 1 : Biologie Moléculaire et Génie Génétique
- Crédits : 06
- Coefficient : 03

Les présentations sont de type Powerpoint avec des phrases courtes, diagrammes, schémas, tableaux,...) et l'utilisation d'un marqueur sur tableau blanc. L'objectif est de développer l'aptitude à raisonner sur des problèmes de Biologie Moléculaire et Génie Génétique et d'apprendre à appliquer les concepts aux données expérimentales. L'étudiant doit prendre des notes lors des explications de l'enseignant.

3.2 Utilisation d'une plateforme

- Dépôts de documents et diapositives du cours sur une plateforme (<https://elearning.univ-usto.dz/course/view.php?id=2244>)
- Outils de communication : forum, messagerie.

4. Mode d'évaluation

La note finale de la matière se compose des notes suivantes :

- 40% Contrôles Continus (minimum 02)

- 60% Examen semestriel écrit de type QCM et un exercice (1h30) nécessitant de mobiliser les notions vues en cours et en travaux dirigés et la note moyenne de 3 contrôles continus. Ils évalueront l'acquisition des connaissances enseignées aussi bien en cours qu'en travaux. Le téléphone portable et les notes de cours ne sont pas permis pour consultation durant les examens et les contrôles contenus.

A. Biologie Moléculaire

Chapitre I.
Support de l'information génétique, l'ADN

A. Biologie Moléculaire

Chapitre I. Support de l'information génétique, l'ADN

1. Structure primaire et dynamique de l'ADN

1.1 Structure de base

L'ADN (Acide DéoxyriboNucléique) est le support matériel de l'information génétique des êtres vivants. C'est un polymère de déoxyribonucléiques reliés entre eux par des liaisons phosphodiester (Pierce, 2012).

- Ribose, désoxyribose

Le ribose est un pentose. Le désoxyribose, composant des acides déoxyribonucléiques (ADN) est dérivé du ribose par une réduction de la fonction alcool secondaire du carbone n°2 (Figure 1).

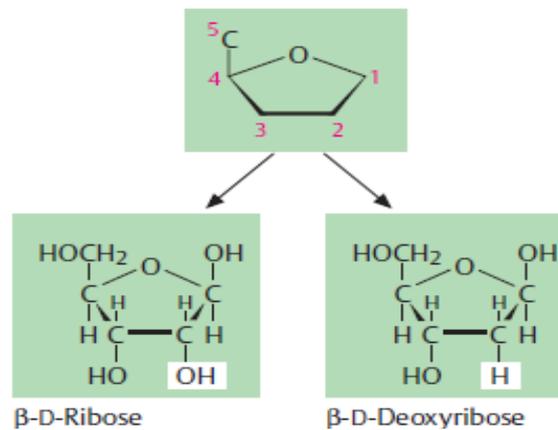


Figure 1 : Différence entre ribose et désoxyribose (Passarge, 2007).

- Acide phosphorique (H_3PO_4)

Des esters de phosphate peuvent se former entre un phosphate et un groupement hydroxyle libre (alcool, énol, phénol...) (Figure 2).

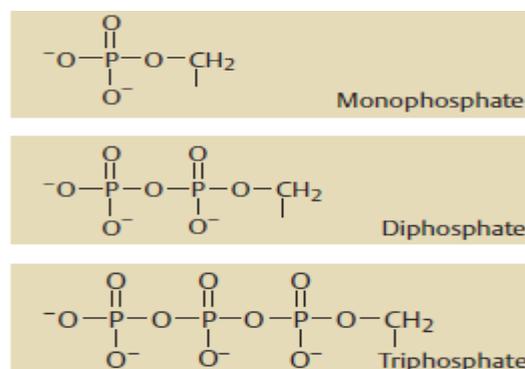


Figure 2 : Acides phosphoriques (Passarge, 2007).

- Bases azotéesa) *bases puriques*

Les bases azotées sont des molécules aromatiques dont le noyau est soit une purine (bases puriques), soit une pyrimidine (bases pyrimidiques).

- Les bases puriques sont au nombre de 2 : l'adénine et la guanine.
- Les purines ont un double noyau aromatique comportant à gauche un cycle hexagonal de 4 carbones et 2 azotes et à droite un cycle pentagonal de 3 carbones (dont 2 communs avec le précédent) et 2 azotes.
- L'adénine est constituée d'un noyau purine dont le carbone 6 est substitué par une fonction amine. Elle est la seule des bases nucléiques dont la formule ne contient pas d'atome d'oxygène.
- La guanine est constituée d'un noyau purine dont le carbone 2 est substitué par une fonction amine et le carbone 6 par une fonction cétone (Figure 3).

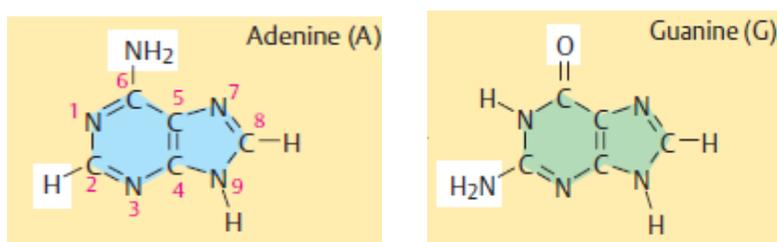


Figure 3 : Bases puriques (Passarge, 2007).

b) *Bases pyrimidiques*

Les bases pyrimidiques sont au nombre de 3 : la cytosine, la thymine et l'uracile.

- Les pyrimidines ont un noyau aromatique hexagonal de 4 carbones et 2 azotes.
- La cytosine est constituée d'un noyau pyrimidine dont le carbone 4 est substitué par une fonction amine et le carbone 2 par une fonction cétone.
- La thymine est aussi constituée d'un noyau pyrimidine dont les carbones 2 et 4 portent des fonctions cétone, mais dont le carbone 5 est substitué par un méthyl.
- L'uracile est constitué d'un noyau pyrimidine dont les carbones 2 et 4 portent des fonctions cétone (Figure 4).

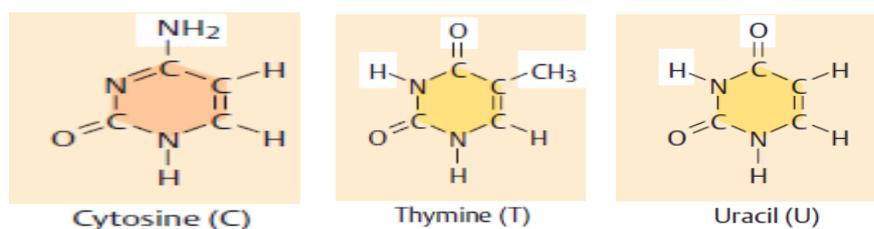


Figure 4 : Bases pyrimidiques (Passarge, 2007).

- Nucléosides, nucléotides

Les sucres (ribose ou désoxyribose) se lient aux bases azotées par des liaisons impliquant un des azotes de la base (azote n°1 des pyrimidines ou azote n°9 des purines) et le carbone n°1 de l'ose. Ce sont des liaisons N-osidiques.

La liaison d'un nucléoside avec un phosphate se fait par une estérification de la fonction alcool primaire (carbone n°5') du sucre et une des trois fonctions acides du phosphate. L'ester obtenu est un nucléotide. Un nucléotide est donc formé d'une base azotée, liée par une liaison osidique avec un sucre, lui-même lié par une liaison ester avec un phosphate (Figure 5).

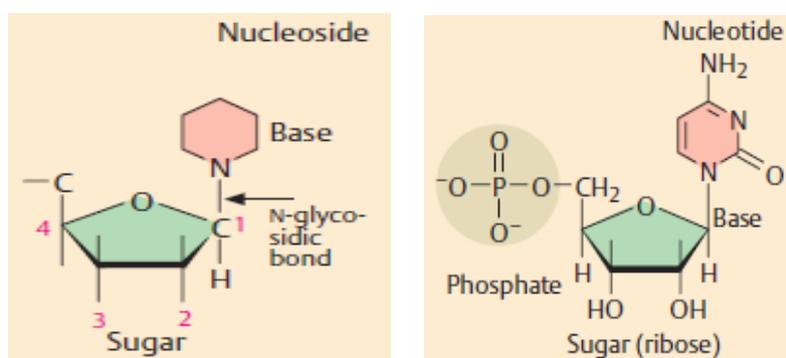


Figure 5 : Structure d'un nucléotide (Passarge, 2007).

Les noms des nucléosides ont comme suffixe :

"Osine" pour les nucléosides puriques

"Idine" pour les nucléosides pyrimidiques (Tableau 1).

Tableau 1 : Nomenclature des nucléosides.

	Bases	Nucléosides	Nucléotides 5'-mono, di, tri phosphate
Purines	Adénine (A)	d-adénosine	d- AMP/ADP/ATP
	Guanine (G)	d-guanosine	d- GMP/GDP/GTP
Pyrimidines	Cytosine (C)	d-cytidine	d- CMP/CDP/CTP
	Thymine (T)	d-thymidine	d- TMP/TDP/TTP

1.2 Structures secondaires

En 1953, Watson et Crick démontrent par diffraction aux rayons X que l'ADN est structuré en hélice double brin. La structure secondaire de l'ADN est telle que les deux brins sont enroulés l'un autour de l'autre. Chacun des deux brins est orienté (5'→3') dans le sens opposé à celui de l'autre brin (3'→5'). Les deux brins ont une orientation antiparallèle et établissent entre eux des liaisons de type hydrogène. Les bases azotées sont tournées vers l'intérieur de la double hélice de façon à ce que chacune s'hybride avec une base de l'autre brin (A avec T, C avec G) (Figure 6). On dit que les bases successives de chacun des brins sont complémentaires. La double hélice a un "pas" de 3,4 nm c'est à dire qu'il y a environ 10 paires de nucléotides pour chaque tour d'hélice. L'appariement suit la loi de Chargaff : $A+G/C+T = 1$. D'autre part, le rapport $(A+T)/(G+C)$ est spécifique de chaque espèce.

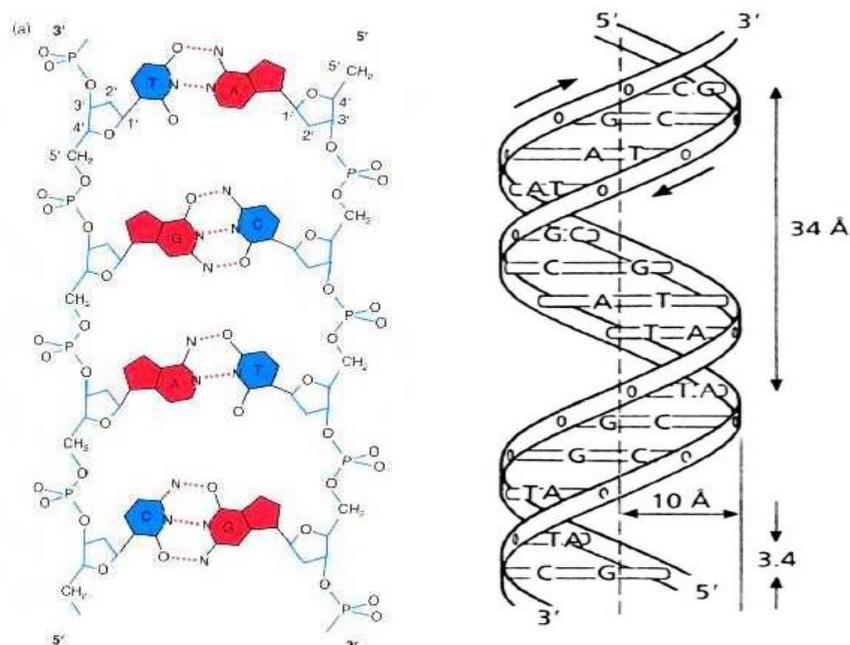


Figure 6 : Structure secondaire de la double hélice (Watson et Crick, 1953).

1.3 Formes alternatives de la double hélice

L'ADN double brin peut avoir trois structures secondaires : A, B et Z. Si on change les conditions d'humidité ou de salinité, on peut obtenir des structures différentes de la forme B

- La structure en échelle se transforme en hélice quand on la fait tourner à droite et on obtient la forme B de l'ADN.
- Les barreaux de l'échelle se rapprochent à 0,34 nm. Comme l'hélice se répète approximativement toutes les 10 bases, son pas (tour complet de l'hélice) est de 3,4 nm.

- La forme B de l'ADN est la conformation majeure de l'ADN en solution.
- Une forme alternative à la double hélice B est l'ADN A. Cette forme apparaît quand la quantité d'eau disponible pour hydrater la double hélice est insuffisante.
 - Dans la double hélice A, le pas est de 2,46 nm et un tour complet de l'hélice demande 11 paires de bases.
 - Bien que les fibres d'ADN, relativement déshydratées, peuvent adopter la conformation A en conditions physiologiques, il n'est pas clair si l'ADN adopte cette conformation *in vivo*. Cependant, les hybrides ADN-ARN double hélice ont probablement une conformation A (Figure 7).
 - Il existe une troisième forme de l'ADN, qui présente une double hélice tournée vers la gauche. Elle s'appelle l'ADN Z. Dans cette molécule d'ADN, les liaisons N-glycosyl des résidus Guanine sont tournées de 180° par rapport à leur conformation dans l'ADN B.

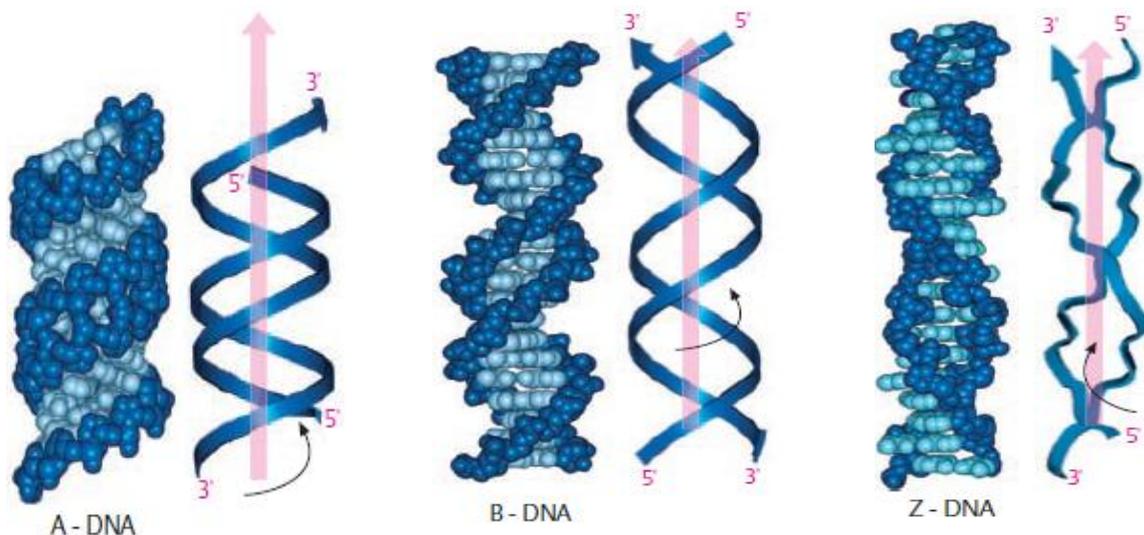


Figure 7 : Formes alternatives de la double hélice (Passarge, 2007).

1.4 Interactions avec les protéines

L'ADN est associé à des protéines appelées histones. Il existe huit molécules : 2H2A, 2H2B, 2H3 et 2H4 appelées octamère. Cette structure supérieure du matériel génétique constitue la chromatine. Un segment d'ADN de 147 pb et un octamère d'histone constituent un nucléosome (Luger et *al.*, 1997). Cette structure est appelée fibre chromatinienne. Dans la chromatine, deux zones peuvent être distinguées à savoir l'hétérochromatine (régions de forte condensation non transcrite, regroupées près du nucléole et de la membrane nucléaire) et

l'euchromatine (régions de plus faible concentration, déroulées et transcrites) représentant la majeure partie du génome. Elle contient la plupart des gènes structuraux. Les fibres compactées de 30 nm vont ensuite former des boucles d'environ 300 nm de long, qui s'ancrent dans une matrice protéique (sorte de "squelette" interne du chromosome, constitué de protéines acides). Enfin, cette structure en boucles successives va former une hélice d'environ 700 nm de diamètre, dont chaque tour est constitué d'environ 60 boucles. Ce niveau représente le taux de compaction maximal de l'ADN dans un chromosome (Thiry et *al.*, 2016) (Figure 8).

Les chromosomes sont de longs fragments d'ADN organisés à différentes échelles et stockés à l'intérieur du noyau des cellules eucaryotes. Les outils de la biologie moderne permettent de sonder et de déterminer les mécanismes moléculaires qui gouvernent leur architecture.

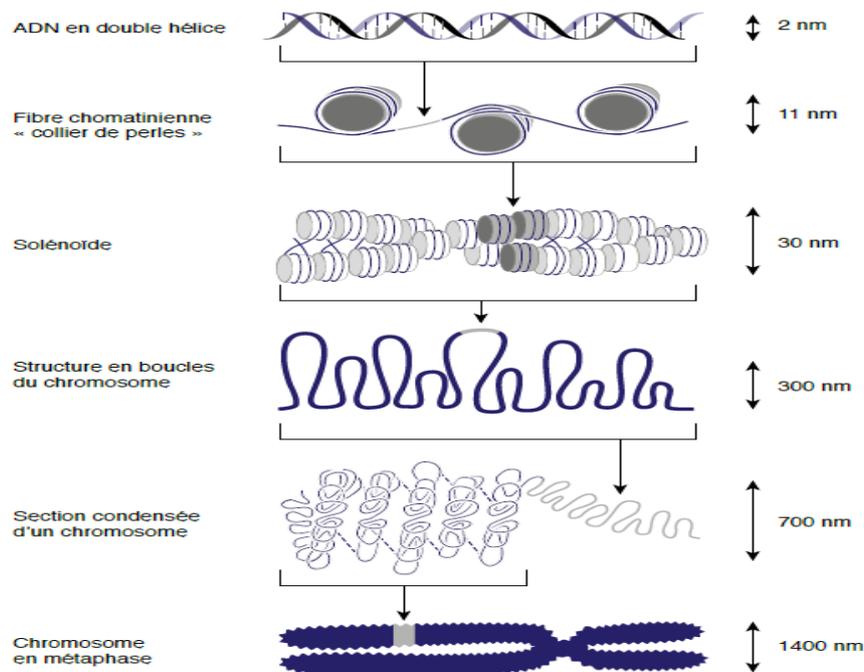


Figure 8 : Différents niveaux d'organisation de la chromatine (Thiry et *al.*, 2016).

1.5 Manipulations topologiques

Dans le cadre de l'ADN bactérien qui est circulaire, il peut exister sous trois formes topologiques (superenroulée, relâchée, linéaire) objectivées par plusieurs techniques telle l'ultracentrifugation, la microscopie électronique ou tout simplement l'électrophorèse en gel d'agarose (technique d'usage courant). La centrifugation sur gradient de densité en présence d'un agent intercalant BET (Bromure d'ETHidium) repose sur l'affinité différente du BET

pour les différentes formes topologique : l'insertion du BET diminue la densité moyenne de la molécule d'ADN. Dans une molécule circulaire, la création de supertours positifs rapprochera les plateaux des paires de bases et rendra progressivement la molécule réfractaire à l'insertion de nouvelles molécules de BET. Le bromure d'éthidium est de plus en plus souvent remplacé par des molécules moins dangereuses comme le bleu de méthylène, bleu de nile, l'azure-A ou le *SyberSafe* (Tagu et al., 2018).

1.6 Dénaturation-renaturation

1.6.1 Dénaturation

Lorsque l'on dénature le double brin d'ADN, les liaisons hydrogènes entre les deux brins sont rompues, les deux chaînes se déroulent et les deux brins se séparent. Cependant, aucune liaison covalente n'est détruite. Cette dénaturation peut être réalisée *in vitro* en soumettant l'ADN à tout agent chimique ou physique capable de déstabiliser les liaisons hydrogène, comme le *pH*, la température, certains solvants, des concentrations ioniques élevées, des agents alcalins. L'appariement G≡C est plus stable à la chaleur que l'appariement A=T, car il possède une liaison hydrogène supplémentaire (Klug et al., 2006).

L'ADN riche en G≡C requiert donc une température plus élevée pour être totalement dénaturé. Lorsque l'absorption à 260 nm (DO_{260}) est mesurée en fonction de la température, on obtient un profil de fusion de la molécule d'ADN. Le point médian de ce profil, ou courbe, est appelée température de fusion (T_m pour *melting temperature*) et correspond à la température à laquelle 50 % de l'ADN est dénaturé (Figure 9). Lorsque le plateau de la courbe est atteint (densité optique maximale), la dénaturation est complète et la solution ne contient que de l'ADN simple brin (Klug et al., 2006).

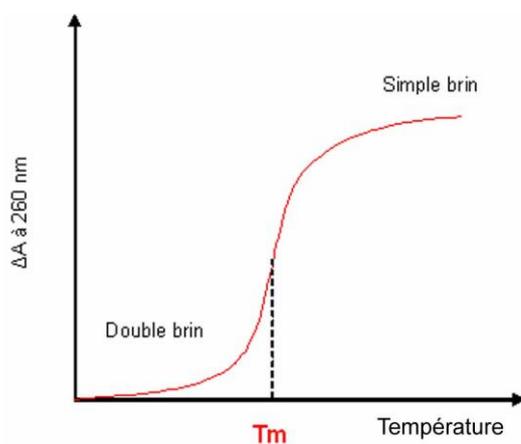


Figure 9 : Mise en évidence de la séparation des deux brins d'un ADN et mesure de la température de fusion T_m (Kaplan et Delpech, 2007).

La T_m peut être déterminée expérimentalement mais est plus généralement calculée à partir de la formule simple : $T_m = (4 \times [G + C]) + (2 \times [A + T])$ °C.

Dans laquelle $[G + C]$ est le nombre de nucléotides G et C dans la séquence d'amorces, et $[A + T]$ est le nombre de nucléotides A et T (Brown, 2010).

Exemple :

- *E. coli* $G + C = 50\%$, $T_m = 65^\circ\text{C}$ - *P. aeruginosa* $G + C = 68\%$, $T_m = 76^\circ\text{C}$.

La température de fusion est influencée par plusieurs facteurs :

- La force ionique du milieu : plus la concentration en ion augmente, plus la T_m diminue (en milieu NaCl > 1M)

- La composition en bases : plus le pourcentage de bases GC est grand, plus la T_m est élevée, plus le nombre de liaisons hydrogène est grand, plus la molécule est stable (dénaturation difficile) (Kaplan et Delpech, 22007).

1.6.2 Renaturation

Le processus de dénaturation est réversible. Si la molécule d'ADN ayant été dénaturé par la chaleur est doucement refroidie, les brins complémentaires se réassocient pour reformer la molécule d'ADN double-brin (Klug et *al.*, 2006). Pour que la renaturation puisse se faire, deux conditions doivent être satisfaites :

- La concentration saline doit être suffisamment importante pour que soient éliminés les répulsions électrostatiques des groupements phosphates des deux brins, ce qui correspond en NaCl de 0,15 à 0,50 M ;

- La température optimale pour la renaturation est de 20 à 25 °C inférieure à la valeur de T_m (Freifelder, 1990).

1.7 Implications biologiques

Tous les ADN naturels existent sous forme superenroulée négativement dans les cellules. Ce superenroulement a des implications biologiques. La création de supers tours demandent de l'énergie : les supertours constituent une réserve d'énergie qui peut faciliter les transitions structurales (la séparation des brins lors de la réplication, transcription, recombinaison requiert de l'énergie par 10 pb).

2. Structure et organisation du génome dans le noyau

Le génome est l'ensemble du matériel génétique d'un organisme contenu dans chacune de ses cellules sous la forme de chromosomes. Dans le monde du vivant, les espèces ont des

génomés parfois très différents. En effet, les molécules d'ADN des espèces peuvent varier par leur :

- Taille : de quelques milliers de paires de bases à plusieurs milliards ;
- Nombre : une molécule d'ADN chez les procaryotes, plusieurs chez les eucaryotes (par exemple 46 molécules chez l'Homme) ;
- Forme : circulaire (chez la plupart des procaryotes) ou linéaire (chez la majorité des eucaryotes) ;
- Localisation : chez les eucaryotes, l'ADN est par exemple contenu dans un noyau, alors que l'ADN des procaryotes est dans le cytoplasme ;
- Séquence : certaines caractéristiques de chaque espèce, d'autres caractéristiques de chaque individu dans une même espèce.

Les génomes sont constitués de régions particulières appelées "gènes", séparées par des régions appelées séquences intergéniques. Les séquences intergéniques sont des régions non-codantes de l'ADN, situées entre les gènes. Chez les procaryotes, il y a très peu de ces régions, la majorité du génome correspondant à des gènes.

La plus grande partie de l'ADN nucléaire est non codante : les 25 000 gènes chez l'homme couvrent un peu plus de 1% de l'ADN pour leur part codante, soit 33% en incluant les séquences géniques non codantes.

Les séquences répétées à fonction connue comprennent les familles de gènes (et leurs pseudogènes, qui n'ont pas de fonction) et les superfamilles de gènes. Ces gènes dérivent d'un gène ancestral unique ayant été dupliqué (une ou plusieurs fois), ces répliques ayant divergé (plus ou moins) au cours de l'évolution.

Chapitre II.
Mutations, mutagenèse et détection

Chapitre II. Mutations, mutagenèse et détection

1. Mutations géniques

1.1 Définition

Une mutation est définie comme un changement héréditaire de l'information génétique. Les mutations peuvent être un facteur d'évolution, elles sont l'origine de la variation génétique. Les mutations qui affectent un seul gène, appelées *mutations géniques*. Par contre, de nombreuses mutations sont nuisibles provoquant chez l'homme et l'animal diverses anomalies et maladies héréditaires. Chez les organismes pluricellulaires, on peut distinguer deux catégories générales de mutation :

- Les mutations somatiques se produisent dans les tissus somatiques. Quand une cellule somatique contenant une mutation se divise par mitose, la mutation est transmise aux cellules filles, et il se forme une population de cellules mutantes génétiquement identiques – un clone.

- Les mutations germinales apparaissent dans des cellules qui qui se différencieront en gamètes. Une mutation germinale peut être transmise aux générations ultérieures, produisant des individus porteurs de la mutation dans toutes leurs cellules somatiques et germinales (Pierce, 2012).

1.2 Intérêt des mutations

Les mutations peuvent être au détriment de l'individu (maladie génétique) ou se faire à son avantage (c'est le principe de l'évolution des êtres vivants). Elle peut toucher des segments plus ou moins importants de l'ADN. A l'échelle d'une maladie, la connaissance du spectre des mutations permet de mieux comprendre la physiopathologie moléculaire (perte totale ou partielle de fonction, gain de fonction) et d'expliquer le mode de transmission récessif ou dominant. Néanmoins, le problème central à l'heure actuelle est d'analyser et de comprendre les relations entre génotype et phénotype (Hanna et *al.*, 2005).

1.3 Réarrangements génétiques des mutations

Ces mutations correspondent à des échanges mutuels de position entre segments d'ADN à l'intérieur ou à l'extérieur d'un gène. Un exemple simple : les inversions dans lesquelles une portion de la séquence d'ADN est excisée puis réinsérée à la même position, mais dans une orientation opposée. Les mutations de grande ampleur, parce qu'elles impliquent des altérations majeures des séquences des gènes, ont toujours un effet sérieux sur la protéine codée et sont fréquemment associées à un phénotype mutant (Winter et *al.*, 1999).

1.4 Mutations naturelles

La réplication de l'ADN est imparfaite. De temps en temps, les ADN polymérase insèrent un mauvais nucléotide dans le brin en cours de synthèse. La plupart du temps, les ADN polymérase corrigent ces erreurs grâce à leur activité correctrice 3' → 5' exonucléase. Mais parfois ces nucléotides peuvent persister après la réplication (Figure 10). Si ces erreurs ne sont pas détectées et réparées par les mécanismes de correction, elles conduiront à des mutations (Klug et al., 2006).

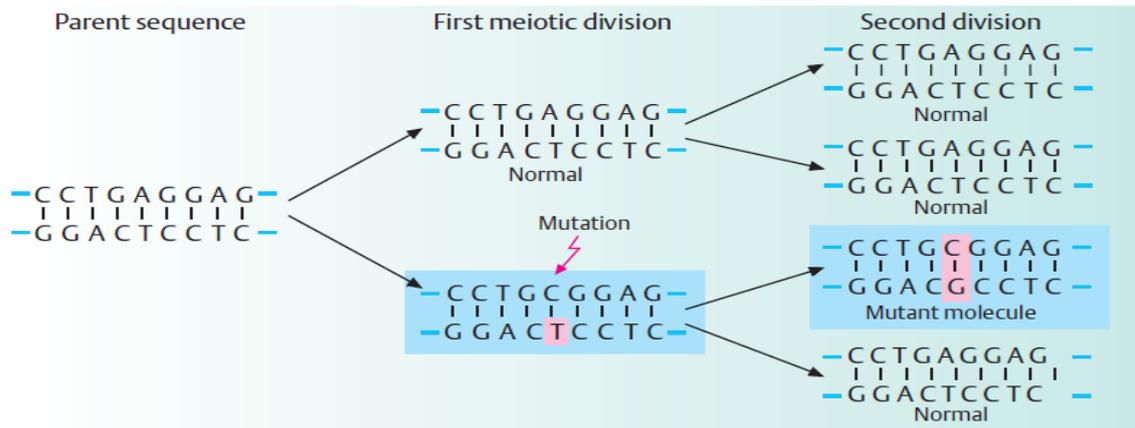


Figure 10 : Erreurs de réplication (Passarge, 2007).

- Dépurination : implique la perte d'une base azotée dans une double hélice intacte. Ce phénomène affecte plus souvent les bases puriques, aussi bien l'adénine que la guanine (Figure 11). Ces bases peuvent être perdues si la liaison glycosidique entre la base et le désoxyribose est rompue. Il en résulte la création d'un site apurinique sur l'un des brins de l'ADN. Au cours de la réplication, les sites apuriniques ainsi produits ne peuvent plus spécifier la base complémentaire de la purine d'origine (Klug et al., 2006).

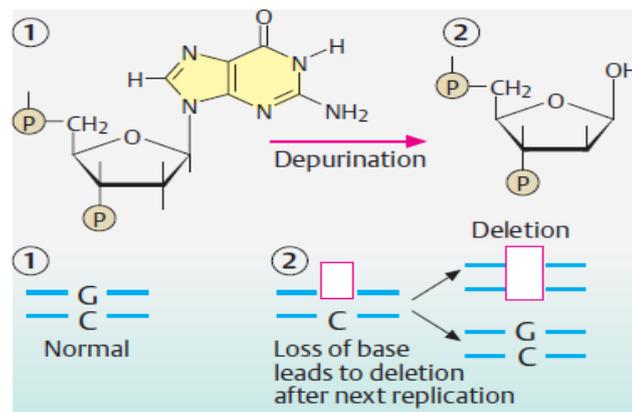


Figure 11 : Dépurination (Passarge, 2007).

- Tautomérisation : les bases de l'ADN sont l'objet d'altérations structurales spontanées appelées tautomérisations. Chaque base peut exister sous deux formes : ainsi la guanine peut revêtir la forme "enol". Ces deux formes sont appelées tautomères ou isomères structuraux. Les différentes formes de tautomères ont des propriétés d'appariements différentes. Si durant la réplication, la guanine (G) est sous la forme "enol", l'ADN polymérase pourra positionner, en face, une thymine (T) à la place d'une cytosine (C) parce que les modalités d'appariement ont changé (et ce n'est pas une erreur de la polymérase). Le résultat est une transition du dinucléotide G-C en A-T (Klug et *al.*, 2006).

- Désamination : le groupe amino d'une cytosine ou d'une adénine est converti en groupe cétone. La cytosine est alors convertie en uracile et l'adénine en hypoxanthine. Le principal effet de ces conversions est d'altérer la spécificité de l'appariement de ces deux bases durant la réplication. A titre d'exemple, la cytosine s'apparie normalement à la guanine. A cause de la conversion en uracile, qui s'apparie avec l'adénine, la paire de bases originale G≡C est convertie en paire A=U, puis, après réplication suivante, en paire A=T. Si c'est l'adénine qui est désaminée, la paire originale A=T est convertie en paires G≡C parce que l'hypoxanthine s'apparie naturellement avec la cytosine (Klug et *al.*, 2006).

1.5 Mutations induites

Les mutations induites sont dues à l'influence de facteurs externes. Elles peuvent être dues à des agents mutagènes naturels ou artificiels. Ainsi, les rayonnements d'origine minérale et cosmique, ou encore les rayonnements ultraviolets produits par le soleil sont des émissions d'énergie auxquelles sont exposés la plupart des organismes (Dupuy et Nougier, 2005). Ces mutations induites peuvent entraîner la rupture d'un chromosome et son réarrangement incorrect. L'action est liée souvent à un choc d'un quantum d'énergie sur un atome ou un groupes d'atomes, ou bien à la formation de substances mutagènes (de peroxydes par exemple) (Lewin, 1999).

1.6 Agents mutagènes

Un agent mutagène est un agent naturel ou résultant de l'activité humaine (certains polluants industriels et les rayons X utilisés en médecine), physique ou chimique, qui peut altérer la structure de l'ADN dans laquelle il détermine des lésions variées telles que ruptures de liaisons diester, modifications chimiques de certaines bases, ponts de covalence établis

entre bases de différents niveaux. Les mutations sont les conséquences plus ou moins directes de ces lésions.

1.7 Effets des mutations

Les mutations qui auront un effet peuvent être réparties en deux catégories :

- Une *perte de fonction* est le résultat normal d'une mutation qui réduit ou abolit l'activité d'une protéine.

- Les mutations avec *gain de fonction* sont beaucoup plus rares. Ces mutations sont celles qui donnent une activité anormale à une protéine. Beaucoup de mutation avec gain de fonction sont situées dans les régions de régulation plutôt que dans les régions codantes, ce qui entraîne nombre de conséquences (Brown, 2004).

1.8 Expression des mutations

Dans la mutation génique, l'allèle d'un gène est changé en un autre allèle. Puisqu'un tel changement se produit à l'intérieur d'un seul gène et est localisé en un locus unique du chromosome, on appelle parfois les mutations géniques des mutations ponctuelles.

Il en existe deux principales catégories : des mutations ponctuelles qui affectent un seul nucléotide sur un seul gène :

1.8.1 Substitutions de nucléotides : Transition-transversion

Transition : une purine est remplacée par une purine différente ou une pyrimidine est remplacée par une pyrimidine différente. Exemple: A → G, C → T.

Transversion : une purine est remplacée par une pyrimidine ou une pyrimidine est remplacée par une purine. Exemple: A → C, A → T.

Les substitutions d'une seule paire de base sont plusieurs effets possibles, qui sont des conséquences directes de deux aspects du code génétique : la dégénérescence du code et l'existence de codons de terminaison de la traduction (Griffiths et *al.*, 2006) :

- Mutations synonymes : la mutation change un codon spécifiant un acide aminé en un autre codon du même acide aminé. Les mutations synonymes sont également appelées mutations *silencieuses*. Ceci n'a pas d'effet sur la fonction codante du génome : le gène muté s'exprime exactement par la même protéine que le gène non muté (Brown, 2004).

- Mutation faux sens : le codon d'un acide aminé est remplacé par le codon d'un autre acide aminé. Les mutations faux-sens sont également appelées mutations *non synonymes*.

- Mutation non-sens : le codon d'un acide aminé est remplacé par un codon de terminaison de la traduction (codon stop). Il en résulte une protéine tronquée (Brown, 2004).

1.8.2 Mutation par décalage du cadre de lecture

Les mutations responsables d'un décalage du cadre de lecture (insertions ou suppressions d'un nucléotide ou plusieurs) suppriment toute ressemblance entre la séquence d'acides aminés située en aval du site mutant lors de la traduction, et la séquence originelle d'acides aminés. Pour cette raison, les mutations par décalage du cadre de lecture conduisent en général à une perte complète de la structure et de la fonction normales de la protéine (Griffiths et *al.*, 2006).

1.9 Réversions et suppressions

Une mutation de réversion serait typiquement une sorte de mutation bénéfique ou du moins, dans le cas où la mutation d'origine était une mutation néfaste. Le nucléotide muté redevient normal. *Exemple* : ATC → AAC (mutation primaire) → ATC (réversion).

Une mutation suppressive est une seconde mutation qui atténue ou annule les effets phénotypiques d'une mutation existante dans un processus de sauvetage synthétique défini. La suppression génétique rétablit donc le phénotype observé avant la mutation d'origine. Les mutations suppressives sont utiles pour identifier de nouveaux sites génétiques qui affectent un processus biologique d'intérêt.

2. Mutagenèse physique, chimique et biologique et techniques de modification du matériel génétique

2.1 Mutagenèse physique

2.1.1 Radiations ionisantes

Les rayons X, les rayons gamma et les rayons cosmiques ont des longueurs d'onde plus faibles et sont donc plus énergétiques que la lumière ultraviolette (UV). La conséquence est qu'ils peuvent pénétrer profondément dans les tissus et causer une ionisation des molécules sur leur passage. Ces réactions peuvent directement ou indirectement affecter le matériel génétique en altérant les bases puriques ou pyrimidiques de l'ADN, ce qui génère des mutations ponctuelles. Les radiations ionisantes peuvent aussi rompre les liaisons

phosphodiester, altérant l'intégrité des chromosomes et produisant ainsi des anomalies chromosomiques comme des délétions, des translocations ou des cassures (Klug et *al.*, 2006).

2.1.2 Rayonnement ultraviolet

Toute énergie sur la terre est constituée d'ondes électromagnétiques de différentes longueurs d'onde. Les rayonnements de longueur d'onde inférieure à celle de la lumière visible sont plus énergétiques et ont un effet délétère sur les molécules organiques. Leurs longueurs d'onde sont absorbées préférentiellement par des bases de l'ADN et par les acides aminés aromatiques des protéines. Cela aboutit à la création de liaisons covalentes sur deux pyrimidines adjacentes appartenant au même brin d'ADN : formation de dimère de pyrimidine et plus particulièrement des dimères de thymine. Les dimères C-C et C-T peuvent également être formés mais sont plus rares. Ces dimères réent des distorsions dans la conformation de la double hélice, ce qui inhibe la réplication normale et introduit des erreurs dans la séquence d'ADN (Klug et *al.*, 2006).

2.2 Mutagenèse chimique

Un certain nombre d'agents présents dans l'environnement provoquent des lésions dans l'ADN. Il peut s'agir d'agents chimiques ou physiques (radiations). Tout agent présent dans l'environnement et qui augmente significativement le taux de mutation par rapport au taux de mutation spontanée est un mutagène (Pierce, 2012).

2.2.1 Analogues des bases

Les analogues de bases sont des mutagènes chimiques qui ressemblent suffisamment aux bases azotées normales pour parfois être incorporées à la place de celle-ci lors de la réplication au niveau du brin fils. Ces analogues ont des propriétés d'appariement différentes, ce qui entraîne des mutations après réplication. Le 5-bromo-uracile (5-BU) est un analogue de la thymine, dans lequel un atome de brome (Br) remplace le groupement méthyle (CH₃) sur l'atome de carbone 5. Normalement, le 5-bromo-uracile s'apparie avec l'adénine, comme le fait la thymine, mais la présence du bromo modifie la distribution des électrons et favorise la formation spontanée d'une forme ionisée qui s'apparie avec la guanine, provoquant l'apparition d'une transition (T-A → 5BU-A → 5BU-G → C-G) (Pierce, 2012).

2.2.2 Agents intercalants

La proflavine, l'acridine orange, le bromure d'éthidium et les dioxines sont des agents intercalants, qui produisent des mutations en s'insérant entre les bases empilées de l'ADN. La structure tridimensionnelle de la double hélice est déformée et cette distorsion provoque des insertions et des délétions d'un nucléotide à la réplication. Ces insertions et délétions décalent fréquemment le cadre de lecture, et l'effet mutagène des agents intercalants est donc sévère (Pierce, 2012).

2.2.3 Agents alkylants

Les agents alkylants sont des composés chimiques qui servent de donneurs de groupement alkyle, comme le groupement méthyle (CH₃) ou éthyle (CH₃-CH₂) dans des réactions d'alkylation des bases de l'ADN. Par exemple, l'éthyl – méthanesulfonate (EMS) donne un groupement éthyle à la guanine, ce qui produit la O⁶ – éthyl-guanine qui s'apparie avec la thymine. Donc l'EMS induit des transitions C≡G → A=T. Le gaz moutarde est aussi un agent alkylant (Pierce, 2012).

2.3 Mutagenèse biologique

Différents virus sont capables d'interférer sur le génome des cellules hôtes qu'ils ont infectées, cela peut entraîner :

- Une sur-expression ou sous-expression de certains gènes régulateurs de la division cellulaire,
- La production d'une cellule tumorale.

2.4 Techniques de modification du matériel génétique

Les nouvelles techniques de génie génétique regroupent des procédés très divers. Actuellement, un type de technique a accéléré et multiplié la capacité d'intervention dans les mécanismes du vivant, c'est la méthode d'édition génomique qui se réfère à diverses méthodes de biologie moléculaire qui sont utilisées afin de couper dans le matériel héréditaire et copier/coller de nouvelles séquences d'ADN à l'endroit de la coupure. Celle-ci est basée sur l'utilisation de ciseaux moléculaires, appelés nucléases programmables, qui coupent dans le génome et permettent, au travers du système de réparation de l'ADN des cellules, de muter, modifier et insérer des gènes existants ou entièrement synthétiques dans une multitude d'organismes végétaux et animaux et de passer outre la barrière des espèces. Par rapport aux méthodes conventionnelles, ces méthodes permettent une augmentation du nombre de

changements fait au génome en une seule opération (*multiplexing*) et sont applicables à un nombre plus important d'organismes vivants.

3. Détection et mise en évidence des mutations (diagnostic génotypique)

L'analyse de l'ADN génomique de n'importe quelle cellule nucléée permet d'effectuer un diagnostic moléculaire des maladies génétiques constitutionnelles monofactorielles récessives ou dominantes dont le gène est connu ou localisé. Il présente l'intérêt d'être véritablement étiologique. Il permet donc de valider le diagnostic chez un individu symptomatique et d'effectuer un conseil génétique dans la famille, par la détection des sujets transmetteurs et par le diagnostic prénatal dans les grossesses à risque. Le diagnostic génotypique a aussi d'autres indications : prédispositions, maladies du génome (cancers) et de l'épigénome, identification des clones cellulaires des individus (Kaplan et Delpech, 2007).

Chapitre III.

Transmission et maintien de l'information génétique

Chapitre III. Transmission et maintien de l'information génétique

1. Réplication de l'ADN et sa régulation

Tant chez les procaryotes que chez les eucaryotes, la réplication de l'ADN est un préalable à la division cellulaire. Cette phase de réplication de l'ADN est appelée *phase S* (de synthèse). Les deux molécules filles d'ADN qui se sont formées au cours de la réplication deviennent des chromosomes à part entière dans les cellules filles (Griffiths et *al.* 2001).

L'enzyme qui catalyse la polymérisation des nucléotides s'appelle l'*ADN polymérase*. Cette enzyme ajoute des désoxyribonucléotides dATP, dGTP, dCTP et dTTP. L'ADN polymérase intervient au niveau de la fourche de réplication, la zone au niveau de laquelle se déroule et expose ses simples-brins pour qu'ils servent de matrices. La nouvelle synthèse ne peut avoir lieu de façon continue que sur l'un des brins matrices. Le nouveau brin ainsi formé est appelé *brin précoce*. La synthèse sur l'autre matrice se fait également à partir de l'extrémité 3' mais elle procède par courts fragments et se déroule dans le mauvais sens (car pour ce brin, le sens 5' → 3' va en s'éloignant de la fourche de réplication). Pour ce brin, l'ADN polymérase s'éloigne de la matrice exposée et doit attendre que la fourche de réplication expose davantage d'ADN avant de pouvoir synthétiser un autre court fragment d'ADN. Ces fragments, appelés fragments d'*Okazaki* sont ensuite reliés les uns les autres par une enzyme appelée *ADN ligase*. Le nouveau brin ainsi formé est le *brin retardé*.

Chez *Escherichia coli* l'*ADN polymérase III (pol III)* catalyse l'addition des nucléotides en 3' au niveau de la fourche de réplication. Toutefois, les polymérases de ce type ne peuvent ajouter des nucléotides qu'à des chaînes nucléotidiques préexistantes. C'est pourquoi au début de la réplication des brins précoce et retardé, de courts fragments d'ARN sont synthétisés pour servir de points de départ, d'amorces à la polymérisation. Pour le brin précoce, une seule amorce initiale est nécessaire car après le premier amorçage, l'addition continue peut utiliser le brin d'ADN en cours de synthèse comme amorce. En revanche, sur le brin retardé, chaque fragment d'Okazaki a besoin de sa propre amorce. Les amorces sont synthétisées par un ensemble de protéines appelé un *primosome*, dont un des composants principaux est une enzyme, la *primase*, un type d'ARN polymérase.

Le retrait des amorces d'ARN et le remplissage par de l'ADN, des vides laissés par ce retrait est assuré par une autre ADN polymérase, *pol I* ensuite une ligase relie l'extrémité 3' de l'ADN, comblant les vides à l'extrémité 5' du fragment d'Okazaki situé en aval.

Le déplacement de la fourche de réplication est assuré par une enzyme, l'*hélicase*, qui rompt les liaisons hydrogènes établies entre les bases appariées et déroule la double hélice devant l'ADN polymérase en cours de progression. Des protéines capables de se lier à l'ADN simple-brin empêchent les simples-brins d'ADN ainsi créés de se réassocier. Lorsque l'ADN se déroule, il a tendance à former des supertours ; La double hélice est ramenée à son état de départ (sans supertour) par l'action d'une autre enzyme, la *gyrase*, qui est un type de *topoisomérase* (Figure 12) (Griffiths et al., 2001).

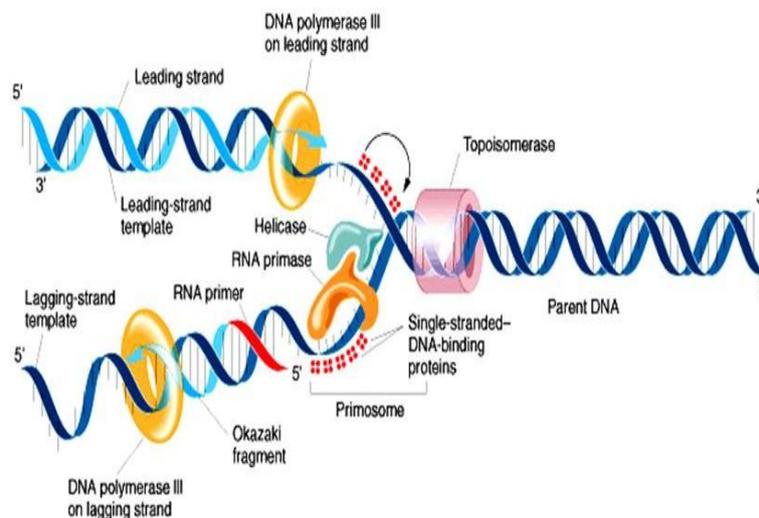


Figure 12 : Fourche de réplication de l'ADN (Griffiths et al., 2001).

La nécessité de coordonner réplication et division cellulaire implique la régulation de ces deux processus. La réplication du chromosome d'*E. coli* est contrôlée au moment du démarrage. Par ailleurs, la division cellulaire est bloquée lorsque la réplication a été perturbée. Le modèle du réplicon, proposé par Jacob et Brenner en 1963 pour expliquer la régulation de la réplication, reste souvent le modèle de choix. Il supposait l'existence d'un "réplicateur", mettant en route la réplication par fixation sur une cible lorsqu'il atteint une concentration cellulaire suffisante. Chez *E. coli*, ce rôle est joué par la protéine *DnaA*, dont la synthèse est contrôlée et qui est, de plus, en état d'équilibre entre deux formes, l'une active, l'autre inactive. La cible de *DnaA*, *oriC*, est elle-même soumise à un deuxième niveau de régulation. Après l'initiation, l'origine de réplication est hémi-méthylée, puisque la méthylation du brin néosynthétisé n'est pas immédiate, et, dans cet état hémi-méthylé, *oriC* est séquestrée sous forme de complexe avec des protéines membranaires. Ce processus évite une réutilisation immédiate des séquences *oriC* (Michel et Baldacci, 1998).

2. Réparation de l'ADN et détection du pouvoir mutagène

La réparation de l'ADN est la correction d'un mésappariement entre les nucléotides (les nucléotides situés sur les deux brins ne sont plus complémentaires (A-T ou C-G), c'est-à-dire d'une absence de complémentarité entre les nucléotides des deux brins d'un fragment d'ADN. L'un des deux brins est identifié en tant que matrice et c'est l'autre brin qui sera réparé en cas de mésappariement. Plusieurs mécanismes de réparation permettent respectivement de remplacer une base, un nucléotide ou un fragment plus ou moins long d'un brin de l'ADN.

2.1 Excision de base

Une base modifiée est d'abord excisée par une ADN glycosylase qui ouvre la double hélice, puis hydrolyse la liaison N-glycosidique de la base endommagée et laisse un nucléotide apurinique ou apyrimidique (dépourvu de base) (Pierce, 2012). La réparation sera faite par une des ADN polymérases de réparation : l'ADN polymérase β qui hydrolyse le nucléotide apurinique, puis le remplace par le nucléotide attendu sur l'extrémité 3'OH libre et la brèche sera refermée par l'ADN ligase qui reconstitue la liaison phosphodiester en 3' du nucléotide ajouté.

2.2 Mésappariements

Chez *E. coli*, le système enzymatique de réparation est multi-protéique. Il est capable de reconnaître un mésappariement et une séquence GATC située dans l'environnement du mésappariement (jusqu'à 1000 pb), afin de vérifier le brin méthylé. Ce complexe multienzymatique est codé par les gènes *MUT* et constitué des enzymes *MUT*. Le complexe de réparation rapproche les bases mal appariées de la séquence GATC et identifie le brin neuf. Il a une activité endonucléasique : il peut cliver les liaisons phosphodiester à l'intérieur d'une chaîne. Ce complexe excise le fragment d'ADN simple brin qui contient le mésappariement (donc sur le brin non méthylé) et des exonucléases éliminent le nucléotide du brin néoformé depuis GATC jusqu'à un peu à l'endroit du défaut d'appariement. Cela entraîne la formation d'une lacune qui doit être comblée. L'ADN polymérase III se positionne alors en 3'OH et synthétise l'ADN complémentaire et antiparallèle. La dernière liaison phosphodiester est effectuée par l'ADN ligase (Griffiths et *al.*, 2006) (Figure 13).

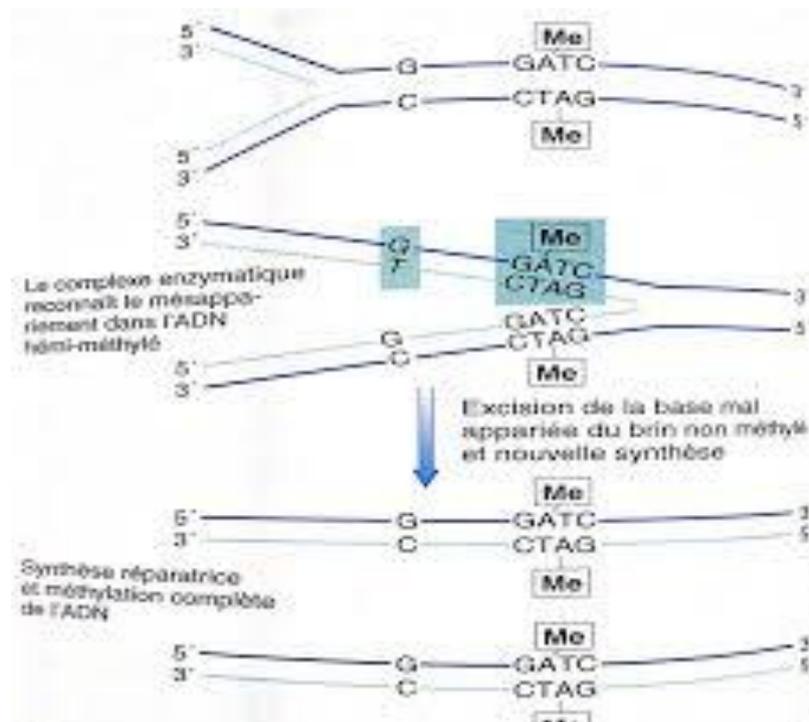


Figure 13 : Modèle de réparation des mésappariements chez *E. coli* (Griffiths et al., 2006).

2.3 Excision de nucléotides

Dans la réparation par excision de nucléotides, un segment d'ADN simple brin contenant le(s) nucléotide(s) lésé(s) est excisé et remplacé par un nouvel ADN. Le système ressemble donc à l'excision de base sauf qu'il n'est pas précédé par le retrait sélectif d'une base, et que le segment de polynucléotide retiré est plus long. L'exemple le mieux étudié est le système patch court de *E. coli*, ainsi appelé parce que le morceau de polynucléotide qui est excisé et ensuite remplacé est relativement court, habituellement d'une longueur de 12 nucléotides (Brown, 2004).

3. Systèmes de restriction-modification

Les endonucléases de restriction reconnaissent spécifiquement des séquences dans l'ADN entrant et clivent l'ADN en fragments (Figure 14a), soit à des sites spécifiques ou plus aléatoire. La méthylation peut réduire la sensibilité des ADN au clivage par des endonucléases de restriction et l'efficacité de la transformation de l'ADN (Figure 14b). Quelques sites sinon tous sont résistants au clivage par les endonucléases, lorsqu'ils

proviennent de souches possédant les méthylase Dcm (méthyle les résidus cytosine) ou Dam (méthyle les résidus adénine).

En outre, les systèmes de restriction-modification sont utiles à d'autres fins médicales: par exemple, pour typer des souches de bactéries pathogènes telles que "*Haemophilus influenzae* non typable (*NTHi*)" et pour suivre les changements dans le microbiome intestinal lors de changements de régime alimentaire ou de régimes de traitement de la maladie (Loenen, 2019).

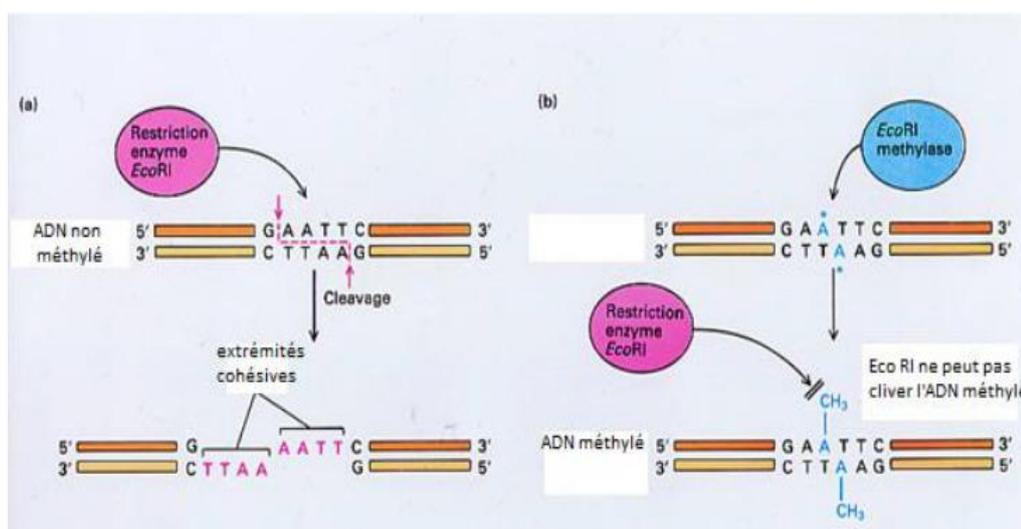


Figure 14 : Système restriction/modification *EcoRI*. (a) restriction, (b) modification (Primrose et Twyman, 2006).

➤ Origine des enzymes de restriction

Les enzymes de restriction, découvertes à partir de 1973, sont des protéines synthétisées par des bactéries pour se protéger des infections de virus (bactériophages). Les bactéries peuvent être parasitées par des virus à ADN, elles produisent des enzymes de restriction qui sont capables de cliver les ADN étrangers quel que soit leur origine. Pour éviter une autodestruction de leur propre ADN, elles se protègent contre leurs propres enzymes de restriction par une modification des sites de restriction correspondants.

➤ Nomenclature des enzymes de restriction

Les règles de nomenclature ont été proposées pour la première fois par Smith et Nathans (1973). Leur nom comporte plusieurs lettres (3 ou 4). La première lettre de dénomination de

l'enzyme est écrite en majuscule, elle correspond au genre de la bactérie d'où a été extraite l'enzyme. La seconde lettre et la troisième lettre (en minuscule) correspondent à l'espèce de la bactérie d'où l'enzyme est extraite (Tableau 2). On peut avoir une quatrième lettre écrite en majuscule correspondant à la souche bactérienne. Enfin pour terminer, un chiffre romain indique l'ordre de caractérisation de ces enzymes (Primrose et Twyman, 2006).

Exemple : EcoRI

- Genre : *Escherichia*

- Espèce : *coli*

- Souche : R

- Ordre de découverte : 1

Tableau 2 : Exemples de nomenclature des enzymes de restriction (Primrose et Twyman, 2006).

Enzyme	Origine de l'Enzyme	Séquence reconnue
<i>SmaI</i>	<i>Serratia marcescens</i> , 1ère enzyme	CCCGGG
<i>HaeIII</i>	<i>Hemophilus aegyptius</i> , 3ème enzyme	GGCC
<i>HindII</i>	<i>Hemophilus influenzae</i> , souche d, 2ème enzyme	GTPyPuAC
<i>HindIII</i>	<i>Hemophilus influenzae</i> , souche d, 3ème enzyme	AAGCTT
<i>BamHI</i>	<i>Bacillus amyloliquefaciens</i> , souche H, 1ère enzyme	GGATCC

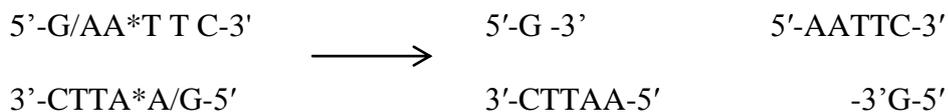
➤ Séquences cibles et modes de clivage

La plupart des endonucléases de restriction de type II, mais pas toutes, reconnaissent et clivent l'ADN dans des séquences particulières de quatre à huit nucléotides. Ces séquences ont un double axe de symétrie de rotation, sont similaires mais inversées : appelés palindromes. Ces enzymes reconnaissent et clivent des séquences différentes. De telles séquences sont souvent appelées palindromes en raison de leur similitude avec des mots qui lisent la même chose vers l'arrière que vers l'avant. Par exemple, les enzymes de restriction et de modification R. *EcoRI* et M. *EcoRI* reconnaissent la séquence (Primrose et Twyman, 2006) :



Type de fragments à extrémités franches ou bouts francs.

La position à laquelle l'enzyme de restriction coupe est habituellement indiquée par le symbole «/» et les nucléotides méthylés par l'enzyme de modification sont habituellement marqués d'un astérisque (*). Pour *EcoRI*, ceux-ci seraient représentés ainsi :



Type de fragments à extrémités cohésives ou bouts collants.

➤ Types d'enzymes de restriction

Il existe au moins quatre types différents de systèmes de restriction/modification selon leurs modes et sites d'action. Cependant un seul système est largement exploité dans les manipulations génétiques *in vitro* ; c'est l'enzyme de type II pour la précision du clivage, et son indépendance vis-à-vis des cofacteurs énergétiques.

- Enzyme de type I : une fois la séquence reconnue, l'enzyme se déplace sur l'ADN, s'arrête de manière aléatoire 1000 à 5 000 paires de bases plus loin et libère quelques dizaines de nucléotides ;

- Enzyme de type II : deux enzymes différentes qui reconnaissent à la fois la même séquence cible qui est symétrique. Les deux enzymes clivent et/ou modifient ou non la séquence de reconnaissance (ne nécessite pas d'ATP, et coupe au niveau de courtes séquences palindromiques de 4 à 8 pb).

- Enzyme de type III : une enzyme avec deux sous-unités différentes, l'une pour la reconnaissance et la modification et une pour le clivage. Reconnaît et méthyle la même séquence, mais clive 24-26 pb plus loin (nécessite de l'ATP).

- Enzyme de type IV : deux enzymes différentes, mais la séquence de reconnaissance est asymétrique. Le clivage se produit sur un côté de la séquence de reconnaissance jusqu'à 20 pb plus loin (Primrose et *al.*, 2004).

➤ Autres types d'enzymes de restriction

- Enzymes isoschizomères :

Les *isoschizomères* étaient des endonucléases de restriction provenant de différentes espèces bactériennes qui reconnaissaient la même séquence. Cela ne signifiait pas nécessairement le même site de clivage (par exemple, *SmaI* clive CCC / GGG et *XmaI* C/CCGGG. Ces enzymes ont ensuite été nommées *néoschizomères* (Loenen, 2019).

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

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

DNase I : endonucléase active sur les simples et doubles brins d'ADN, coupant préférentiellement après un nucléotide pyrimidique. L'ADN coupé reste phosphorylé en 5'. Cet enzyme est utilisé en combinaison avec l'ADN polymérase I dans les marquages de l'ADN par *nick-translation* (Kamoun, 1997).

Nucléase SI : cette enzyme dégrade spécifiquement les acidesnucléiques simple brin (Kaplan et Delpech, 2007).

- RNase A : cette enzyme très résistante (se maintient après 1 heure à 90°C) hydrolyse spécifiquement les ARN simple brin après une pyrimidine. Elle permet également de détecter les mismatches dans les hybrides ADN-ARN (Primrose, 2006).

- RNase H : digère l'ARN dans un complexe ARN-ADN. On s'en sert pour éliminer l'ARN après avoir fabriqué un premier brin d'ADNc à l'aide de la reverse transcriptase (Primrose, 2006).

4. Cartes de restriction

Une carte de restriction d'un fragment d'ADN cloné indique le nombre de sites de restriction présents, leur ordre et les distances qui les séparent sur le fragment. Les cartes de restriction de différents ADN clonés sont généralement suffisamment différentes pour servir de carte d'identité à chaque clone. Les unités des cartes de restriction sont exprimées en paires de bases (pb) ou de plus grandes longueurs, en paires de kilobases (kb). La taille des fragments pourra être définie par électrophorèse (Klug et *al.*, 2006).

L'enzyme *Eco*RI utilisée pour la digestion coupe à un seul endroit pour un ADN linéaire, il y aura 2 fragments 6kb et 8 kb, l'enzyme *Hind*III coupe à un autre endroit pour donner 2 fragments 7 kb et 10 kb. La somme des tailles des fragments obtenus doit donner la taille initiale de la molécule d'ADN. Dans l'illustration ci-dessous, dans le puits 1, l'ADN

digéré par *EcoRI* seule est déposé, dans le puits 2, l'ADN digéré par *HindIII* seule et dans le puits 3, l'ADN digéré par *EcoRI* et *HindIII*.

La détection des différents fragments est réalisée à l'aide d'une sonde marquée du gène étudié (Figure 15).

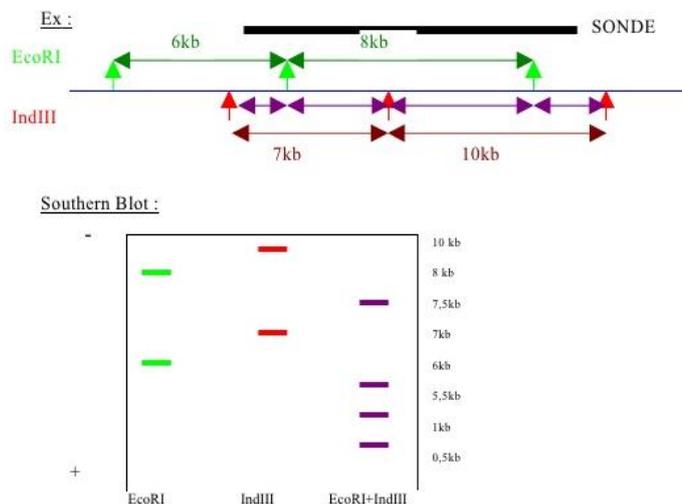


Figure 15 : Carte de restriction et test de digestion dans un gel d'agarose.

5. Intérêt et analyse du polymorphisme de restriction

Le polymorphisme qui se rencontre au niveau des sites d'enzymes de restriction et une sonde radioactive est appelé polymorphisme des longueurs des fragments de restriction ou RFLP (*Restriction Fragment Length Polymorphism*) (Botstein et al., 1985), qui est défini par la variation individuelle d'une séquence d'ADN révélée par des modifications de la longueur des fragments de restriction. Le profil obtenu avec une enzyme de restriction donnée montre les différences individuelles sur un autoradiogramme.

5.1 Technique de Southern-blot

Le Southern-blot est une technique mise au point par E.M. Southern (1975) pour rechercher spécifiquement des fragments d'ADN transféré sur un filtre par leur hybridation à des séquences complémentaires radioactives (sondes). Les étapes sont les suivantes :

- L'ADN génomique est digéré par des enzymes de restriction, les fragments obtenus sont séparés selon leur taille par électrophorèse sur gel d'agarose ;
- Les fragments sont dénaturés par un traitement alcalin, neutralisés ensuite transférés sur une membrane de nylon par capillarité grâce à un montage pour faire passer les fragments d'ADN grâce à une montée de tampon imprégnant le gel puis une membrane de nylon où l'ADN va se fixer par des liaisons stable après exposition aux U.V ou l'action de la chaleur ;

➤ La membrane de nylon avec l'ADN fixé est alors mise à incuber dans un sac contenant une solution d'une sonde radioactive complémentaire du fragment d'ADN qu'on recherche, à une température assez basse pour que l'hybride se forme mais assez élevée pour que cet hybride soit parfaitement complémentaire ;

➤ La membrane des molécules de la sonde qui ne sont pas fixées à leur DNA complémentaire est lavée, puis mise en présence d'un film radiographique pour que la sonde radioactive fixée sur les fragments d'ADN impressionne le film. Les fragments reconnus par la sonde forment des hybrides stables à une température précise Une ou plusieurs bandes sont détectées sur un autoradiogramme (Figure 16).

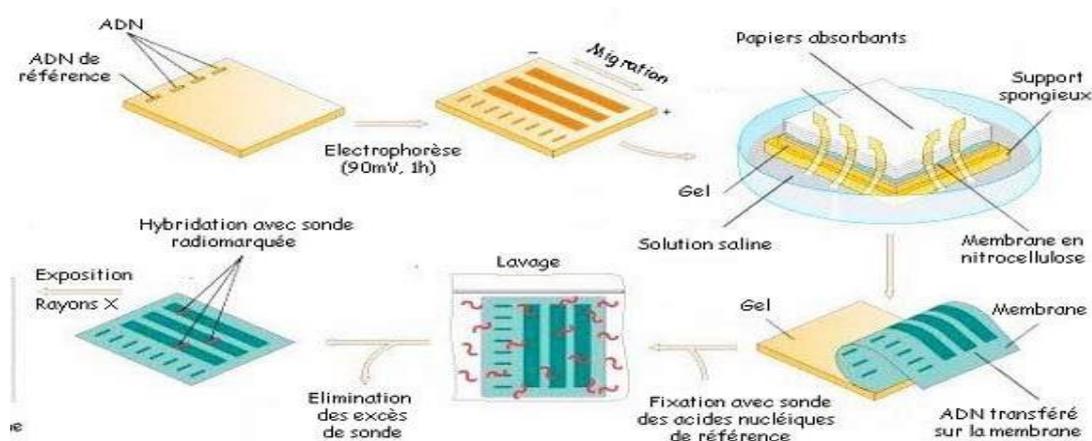


Figure 16 : Principe de la technique d'hybridation sur gel "blot" (Griffiths, 1996).

Une mesure plus précise de la taille des fragments est obtenue en utilisant les mobilités des fragments du marqueur de taille pour construire une courbe d'étalonnage; les tailles des fragments inconnus peuvent alors être déterminées à partir des distances auxquelles ils ont migré (Figure 17).

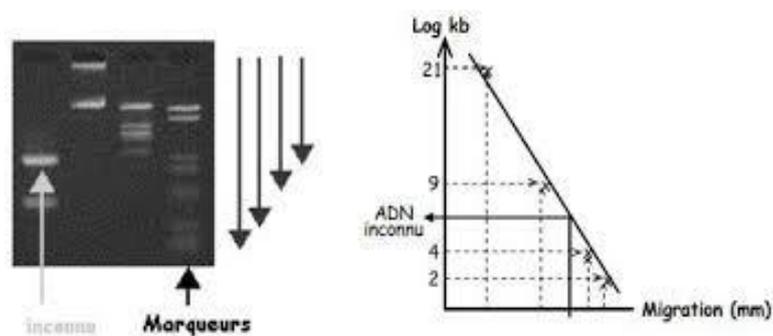


Figure 17 : Estimation de la taille des fragments d'ADN dans un gel d'agarose.

(<http://www.univ-oeb.dz/fsesnv/wp-content/uploads/2020/04/E-SAE8-3-2020.pdf>)

Chapitre IV.
Mécanismes moléculaires de la recombinaison

Chapitre IV. Mécanismes moléculaires de la recombinaison

1. Recombinaison homologue

La recombinaison homologue est un type de recombinaison génétique qui inclut une variété de phénomène permettant le réarrangement de *loci* génétiques. Dans un premier temps, deux chromosomes possédant les génotypes $a^+ b^-$ et $a^- b^+$ s'apparient (Figure 18). Puis une coupure a lieu au hasard de manière symétrique sur les deux chromosomes. Ensuite, les quatre fragments résultant de ces coupures sont recollés pour former deux chromosomes recombinants dont les génotypes sont $a^+ b^+$ (type sauvage) et $a^- b^-$ (double mutant) (Freifelder et *al.*, 1990).

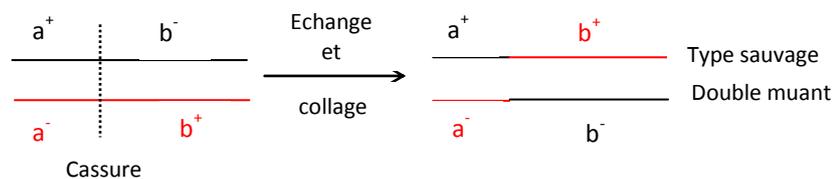


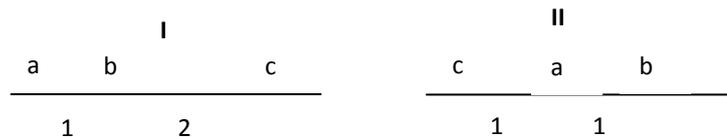
Figure 18 : Représentation schématique d'un échange génétique (Freifelder et *al.*, 1990).

2. Cartographie

Dans la cartographie génétique, la distance sur le chromosome entre deux *loci* recombinants (ou deux mutations) détermine la fréquence de recombinaison. Tant que les deux *loci* ne sont pas trop près l'un de l'autre et que les coupures se font au hasard, la fréquence de recombinaison est proportionnelle à la distance. Ainsi, dans le croisement suivant, les génotypes des chromosomes sont $a^+ b^- c^-$ et $a^- b^+ c^+$, les gènes sont placés par ordre alphabétique et équidistants (Freifelder et *al.*, 1990) :

$$\begin{array}{ccc} a^+ & b^- & c^- \\ \hline & x & \\ \hline a^- & b^+ & c^+ \end{array}$$

Il apparaît deux fois plus de recombinants $a^+ c^+$ que de recombinants $a^+ b^+$ du fait que le locus *a* est deux fois plus éloigné du locus *c* que du locus *b*. Etant donné que la fréquence de recombinaison est proportionnelle à la distance, elle peut être utilisée pour déterminer l'ordre des gènes sur le chromosome. Ceci peut être montré à travers un exemple simple. Considérons trois gènes *a*, *b* et *c* dont l'ordre sur le chromosome est inconnu :



Les deux arrangements possibles de trois gènes dans lesquels a et b sont distants de 1 cM et b et c de 2 cM (Freifelder et *al.*, 1990).

En utilisant la notation $p \times q = m$, pour indiquer une fréquence de recombinaison de $m\%$ entre les gènes p et q , nous supposons que $a \times b = 1\%$ et que $b \times c = 2\%$ de recombinaison. La détermination de la fréquence de recombinaison entre a et c devrait permettre de trancher entre les deux possibilités. Supposons que ce résultat soit de 1% , alors seul l'arrangement II est possible. L'ordre $c a b$ de ces trois gènes et leurs distances relatives constituent une carte génétique. N'importe quel gène peut être cartographié par cette méthode. Dans ce cas considérons l'addition d'un quatrième gène, d , aux trois précédents. Si $d \times b = 0,5\%$, d doit être localisé à 0,5 unité soit à gauche, soit à droite de b . Si $a \times d = 1,5\%$, alors d est clairement à droite de b et l'ordre des gènes devient $c a b d$. Si $a \times d = 0,5\%$, l'ordre des gènes serait $c a d b$ (Freifelder et *al.*, 1990).

3. Analyse et construction génétique

Une séquence d'ADN destinée au transfert dans une cellule, comprenant un gène d'intérêt, les séquences promotrices et régulatrices indispensables à son expression et à sa régulation dans la cellule receveuse et un gène marqueur. La première étape, après identification de l'organisme donneur, consiste à intégrer le gène d'intérêt dans une construction génétique associant parfois un gène marqueur. Ce marqueur permet de sélectionner aisément les cellules qui ont intégré le gène d'intérêt. La deuxième étape consiste à transférer la construction génétique dans une cellule végétale.

4. Recombinaison à un site spécifique

La recombinaison spécifique diffère de la recombinaison homologue par différents aspects différents. En termes de manipulation des gènes, les différences les plus importantes entre ces processus concernent la disponibilité de la recombinase ainsi que la taille et la spécificité de la séquence cible. Les systèmes de recombinaison spécifique sont des systèmes

très spécialisés qui diffèrent selon les organismes. La recombinaison homologue a lieu entre des séquences d'ADN présentant de longues régions d'homologie sans être spécifique d'une séquence particulière, tandis que la recombinaison à spécificité de site s'effectue à des sites de reconnaissance petits et spécifiques. Cela implique que les sites-cibles pour la recombinaison spécifique peuvent aisément être introduits de façon discrète dans les transgènes, mais la recombinaison n'aura lieu dans une cellule hétérologue que si l'on fournit également une source de recombinase (Primrose et *al.*, 2004).

5. Eléments génétiques mobiles (transposons et rétrotransposons)

3.1 Transposons

Dans les années 40, Barbara McClintock découvre un segment d'ADN responsable de l'instabilité chromosomique du maïs (Mc Clintock, 1948). Depuis, des éléments transposables ont pu être mis en évidence chez tous les organismes vivants où ils ont été cherchés. Chez l'homme, 40 à 45 % du génome est constitué d'éléments transposables, ces séquences d'ADN étant moyennement répétées et dispersées (Lander *et al.*, 2001), puis aux alentours de 1980 chez la drosophile, chez la levure et chez *E. coli*. Ils sont des séquences d'ADN moyennement répétées et dispersées dans le génome dont la taille varie de quelques centaines à quelques milliers de pb, leur nombre est rarement plus élevé que 100 par génome (Solignac et *al.*, 1995).

Une propriété caractéristique de la transposition est que le segment transféré est flanqué par une paire de courtes séquences répétées de même sens. Les différents types d'éléments transposables connus chez les eucaryotes et les procaryotes sont répartis approximativement en trois catégories en fonction du mécanisme de transposition :

- Les transposons d'ADN qui sont transposés de façon répllicative, le transposon d'origine restant en place et une nouvelle copie apparaissant ailleurs dans le génome ;
- Les transposons d'ADN qui se transposent de façon conservative, le transposon d'origine se déplaçant vers un site nouveau par un système de couper-coller (Figure 19) ;
- Les rétroéléments qui se transposent tous par un ARN intermédiaire (Brown, 2004).

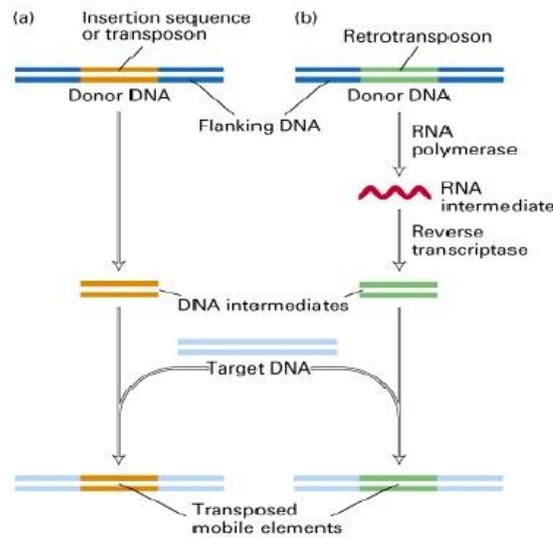


Figure 19 : Transposons et rétrotransposons (Lodish et *al.*, 2003).

Les transposons d'ADN, également connus sous le nom d'éléments transposables de classe 2, sont flanqués aux deux extrémités de répétitions inversées terminales (Figure 20). Les répétitions inversées sont des compléments les unes des autres (la répétition à une extrémité est une image miroir de, et composée de nucléotides complémentaires à, la répétition à l'extrémité opposée) (Pray, 2008).

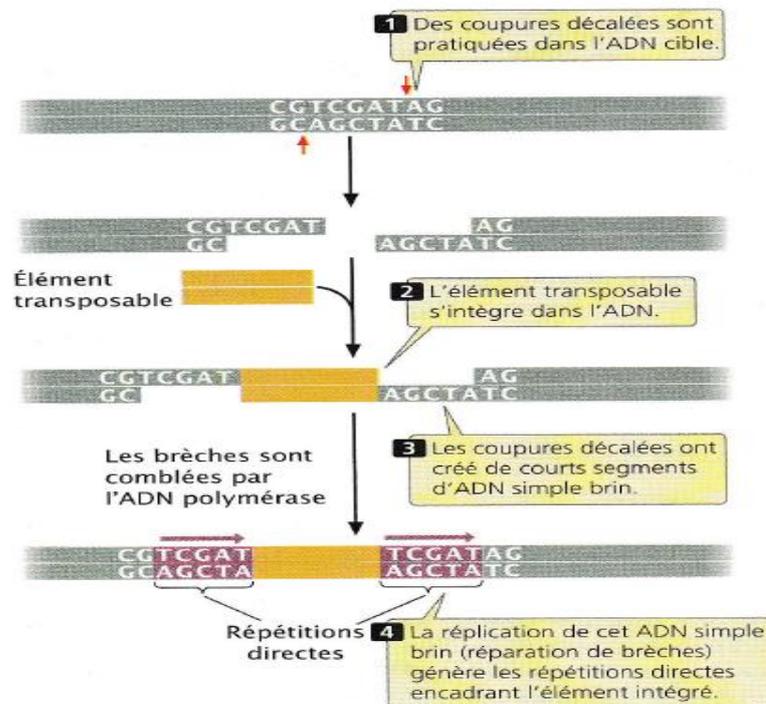


Figure 20 : Structure d'un transposon d'ADN (Pierce, 2012).

3.2 Rétrotransposons

Les éléments transposables (ET) dont la transposition dépend d'une transcription suivie d'une transcription inverse sont appelées rétrotransposons. Leur séquence code pour plusieurs protéines dont une transcriptase inverse. Il existe deux grandes classes de rétrotransposons :

Classe I : ou retrotransposons de type viral, leur structure et leur priorités transcriptionnelles ressemblent à celles des séquences retréovirales intégrées dans le génome cellulaire sous la forme de provirus endogène, sont bornées par de longues répétitions terminales présentant la même orientation (*Long Terminal Repeats, LTR*). Cette class est rencontrée chez les plantes (tabac), les champignons (la levure), les invertébrés (la drosophile) et de nombreux mammifères.

Classe II : de type non viral, ne possédant pas de répétitions terminales et leurs extrémités 3' possède une trace riche en A + T. Aucun type de rétrotransposon n'a encore été trouvé chez les procaryotes (Solignac et *al.*, 1995).

6. Utilisation des transposons

6.1 Marquage

Le marquage des transposons permet d'inclure plusieurs mutants simultanément lors des rondes de criblage dans les souris. Si les transposons n'étaient pas uniques, il serait impossible de différencier les mutants les uns par rapport aux autres et les infections de souris auraient lieu avec un seul mutant à la fois.

6.2 Mutagenèse

Les éléments transposables peuvent altérer à la fois l'organisation et l'expression des gènes, à des fréquences qui excèdent celles des mutations spontanées. De nombreux réarrangements chromosomiques peuvent être imputés à des éléments transposables : insertions, excisions, délétions, inversions, duplications et translocations. Une nouvelle insertion dans la région codante d'un gène conduit fréquemment à l'inactivation de ce gène. Des insertions dans la région régulatrice peuvent avoir des modifications dans l'expression du gène. Une insertion dans un intron peut fortement perturber l'épissage de l'ARNm (Solignac et *al.*, 1995).

6.3 Clonage

Les gènes dans lesquels un élément transposable a été inséré peuvent être clonés en utilisant des sondes d'hybridation spécifiques de l'élément transposable adjacent. Trois

familles d'éléments transposables, *AciDs*, *Spm* (En) et *Mu*, ont été utilisées avec succès pour cloner des gènes de maïs (Chandler, 1994).

6.4 Mobilisation de matériel génétique

Les transposons mobilisables sont des éléments transposables qui codent les fonctions nécessaires à leur mobilisation, mais, à la différence des transposons conjugatifs, ils ne codent pas l'ensemble de protéines nécessaires à leur transfert par conjugaison. Ils portent cependant leur propre site *oriT* (origine de transfert ou site *RSA*), une protéine de mobilisation *TnpZ*, et leur transfert d'une cellule à une autre dépend de la machinerie de conjugaison d'éléments présents dans leur hôte (plasmide ou transposon conjugatif).

6.5 Carte génétique

Une carte génétique sert à décrire l'ordre relatif de marqueurs génétiques au sein d'un groupe de liaison. La distance entre les marqueurs est exprimée en unités de recombinaison. L'unité standard de distance génétique est le centimorgan (*cM*) qui exprime le pourcentage de descendants pour lesquels il y a recombinaison entre deux marqueurs. Le nom de l'unité fait référence à Thomas Hunt Morgan ; le *cM* est l'unité arbitraire correspondant à une fréquence de recombinaison de 0,01 qui est égal à 1000 kb (Gibson et Muse, 2004).

Dans de nombreux cas, les *loci* étudiés sont des gènes, dont les schémas d'héritage sont suivis en surveillant les phénotypes de la progéniture produite après un entre parents avec des caractéristiques contrastées (par exemple, grands et courts pour les plants de pois étudiés par Mendel). Les modèles d'hérédité révèlent l'étendue du lien génétique entre les gènes présents sur le même chromosome, permettant de déduire les positions relatives de ces gènes et de construire une *carte génétique*. Plus récemment, des techniques ont été conçues pour la cartographie génétique de séquences d'ADN qui ne sont pas des gènes mais qui présentent toujours une variabilité dans la population humaine. Les plus importants de ces marqueurs ADN sont : Les marqueurs RFLP, microsatellites, SNP... (Brown, 2010).

Chapitre V.

Expression de l'information génétique et son contrôle

Chapitre V. Expression de l'information génétique et son contrôle

1. Structure de l'ARN

Les acides nucléiques simples brins (ARN ou Acide RiboNucléique) comme l'ADN, est un polymère de nucléotides joints par des liaisons phosphodiester. Il y a toutefois des différences de structure importante entre l'ADN et l'ARN. Au lieu du désoxyribose présent dans les nucléotides de l'ADN, les nucléotides de l'ARN contiennent un sucre ribose. Une autre différence importante est que la thymine, une des deux pyrimidines présentes dans l'ADN, est remplacée par l'uracile dans l'ARN (Pierce, 2012).

L'ARN existe habituellement sous la forme d'une molécule simple brin (Figure 21). Une conséquence de la nature monocaténaire de l'ARN est que de courtes régions complémentaires internes peuvent s'apparier et former des structures secondaires.

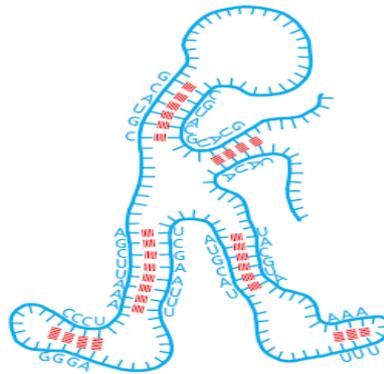


Figure 21 : Conformation d'une molécule d'ARN (Alberts et *al.*, 2017).

1 Classes d'ARN

Les molécules d'ARN accomplissent une variété de tâches dans la cellule. L'ARN ribosomal (*ARNr*) associé à des protéines constitue les sous-unités du ribosome, le siège de l'assemblage des protéines. L'ARN messager (*ARNm*) transporte de l'ADN au ribosome les instructions codées par la synthèse de chaînes polypeptidiques. L'ARN de transfert (*ARNt*) fait le lien entre la séquence nucléotidique codante de l'ARNm et la séquence d'acide aminé d'une chaîne polypeptidique (Pierce, 2012).

D'autres classes de molécules d'ARN sont présentes dans le noyau des cellules eucaryotes. Les petits ARN nucléaires (*ARNpn*) se combinent avec de petites sous-unités protéiques pour former de petits ribonucléoprotéines nucléaires (*RNPpn*). Les micro-ARN

(*ARNmi*) et les petits ARN interférants (*ARNpi*) peuvent aider, par un processus, à déclencher la dégradation de l'ARNm ou inhiber sa traduction en protéine (Pierce, 2012).

2. Transcription et maturation de l'ARN

2.1 Transcription

A partir de la séquence d'un gène, la synthèse d'une protéine comporte obligatoirement différentes étapes : (1) la transcription du gène en pré-ARN messenger (Figure 22) ; (2) la maturation du pré-ARN messenger, qui comprend l'épissage des introns, l'addition de la coiffe et la polyadénylation ; (3) le transport de l'ARN messenger mature depuis le noyau jusqu'au cytoplasme, dans lequel sa concentration est contrôlée par des éléments permettant sa stabilisation ou sa dégradation ; et simultanément (4) sa traduction plus ou moins efficace. L'expression d'un gène dans une cellule résulte donc de la régulation coordonnée de chacune de ces étapes. En particulier, lors du transfert de l'ARNm depuis le noyau vers sa destination cytoplasmique, l'ARNm interagit avec des protéines régulatrices qui vont gouverner son devenir (Pallier, 2001).

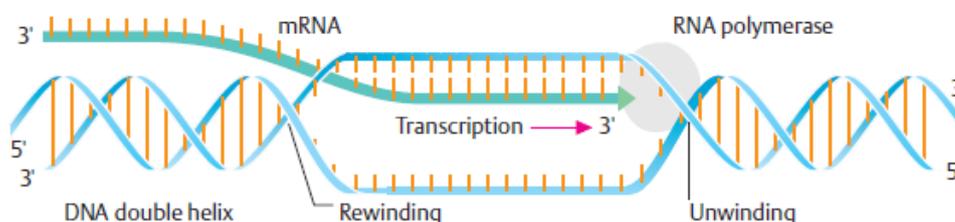


Figure 22 : Etape de la transcription (Passarge, 2007).

Les trois étapes de la transcription de l'ADN, en utilisant comme exemple le système de la bactérie *E. coli* chez les procaryotes, sont :

Initiation : une séquence d'ADN à laquelle se fixe une ARN polymérase pour initier la transcription est appelée un promoteur. Un promoteur fait partie de la région régulatrice adjacente à la région codante d'un gène.

Elongation : l'ARN polymérase avance le long de l'ADN, maintenant ouverte une bulle de transcription pour exposer le brin matrice, et catalyse l'élongation en 3' du brin d'ARN

encours de synthèse. La polymérase compare les ribonucléotides triphosphate libres avec la base exposée qui suit dans l'ADN matrice et, lorsqu'il y a complémentarité.

Terminaison : lorsque l'ARN polymérase reconnaît des séquences nucléotidiques spécifiques dans l'ADN qui agissent comme des signaux de terminaison de la synthèse de la chaîne, le brin d'ARN et la polymérase sont libérés de la matrice d'ADN. Les séquences de terminaison sont constituées d'environ 40 pb et se terminent par une séquence riche en GC suivie d'une succession de six A, ou plus, sur le brin matrice. Les séquences GC sont arrangées de telle sorte que le transcrit est capable de former avec lui-même des liaisons complémentaires dans cette région. On appelle la région d'ARN double-brin résultante, une *épingle à cheveux*. Elle se poursuit et se termine par une succession de U qui correspond aux résidus A sur la matrice d'ADN. L'épingle à cheveux et la succession de résidus U semblent servir de signal pour la libération de l'ARN polymérase et la terminaison de la transcription (Griffiths et al., 2001).

2.2 Maturation de l'ARN

L'ARN messenger sert de support pour la synthèse des protéines ; il apporte l'information génétique de l'ADN à un ribosome et aide l'assemblage bien ordonné des acides aminés. Chez les bactéries, l'ARNm est le produit direct de la transcription, mais, chez les eucaryotes, le transcrit primaire est un pré-ARNm qui est ensuite modifié pour produire un ARNm mature (Pierce, 2012).

- *Addition d'une coiffe 5'* : au cours de leur maturation, une coiffe, constituée d'une guanine méthylée en position N7, est ajoutée à l'extrémité 5' de la plupart des pré-ARNm par une liaison 5'-5' triphosphate. Cette coiffe participe aux réactions d'épissage et de polyadénylation nucléaires, au transport nucléocytoplasmique des ARNm, à l'initiation de leur traduction et à leur protection contre la dégradation par les exoribonucléases dans le sens 5' 3' (Lewis et Izaurralde, 1997).

- *Addition d'une queue poly (A)* : une deuxième modification de l'ARNm eucaryotique est l'addition de 50 à 250 nucléotides adénine à son extrémité 3'. Ce processus joue un rôle critique dans l'expression des gènes. En effet, la queue poly (A) d'un ARNm mature est essentielle pour ses fonctions, notamment la stabilité, la translocation vers le cytoplasme et la traduction et durée de vie de l'ARNm (Wahle et Kuhn, 1997).

- *Excision-épissage de l'ARNm* : une autre modification très importante du pré-ARNm eucaryotique est l'élimination des introns par excision suivie de d'épissage des exons (réunion bout à bout des exons). Le clivage des introns se fait au niveau des séquences *consensus* se trouvant de part et d'autre des introns. La première séquence est dite *donneur d'épissage GT* (côté 5') l'autre *accepteur d'épissage AG*. De plus, une troisième séquence intronique est impliquée dans l'épissage, c'est le *site de branchement* se trouve à environ 40 nucléotides avant le nucléotide AG. Le site de branchement accroche un bout de l'intron formant un lasso (Pierce, 2012).

3. Traduction et maturation des protéines

La biosynthèse d'une protéine comporte deux étapes principales : la première correspond à la synthèse de la chaîne polypeptidique par décodage de l'ARNm ; c'est l'étape de traduction ; la seconde consiste à maturer la chaîne polypeptidique néosynthétisée qui peut subir diverses modifications avant de devenir une protéine fonctionnelle (Leroux et Tossier-Klopp, 2000).

3.1 Traduction

Les caractéristiques fondamentales de la traduction sont les suivantes : la traduction procède de 5' en 3' sur l'ARNm, et la protéine croît de l'extrémité aminée vers l'extrémité carboxylée. Il est utile de se rappeler que, dans la lecture d'une phrase, le sens importe également. Le début de la "phrase" - l'extrémité 5' de l'ARNm et l'extrémité aminée du polypeptide est représentée, par convention, à gauche. La fin extrémité -3' de l'ARNm et l'extrémité carboxylée du polypeptide -l'est à droite. Le message est donc lu de gauche à droite.

Le mécanisme de la traduction d'une séquence de codons en une chaîne polypeptidique est complexe et implique un grand nombre d'étapes répétitives. La traduction s'accomplit dans le cytoplasme sur des petites particules appelées ribosomes. De concert avec les ARNt, les ribosomes constituent la machinerie qui convertit les séquences de nucléotides des ARNm en séquences d'acides aminés dans les protéines. Cependant, ce processus met également en jeu une variété d'intervenants (Leroux et Tossier-Klopp, 2000).

Code génétique

Dès 1964, le code génétique était totalement connu. Chaque codon est constitué de trois nucléotides adjacents dans un brin d'ARNm. Soixante et un codons spécifient les 20 acides aminés et chacun d'entre eux n'en spécifie qu'un seul. Le triplet ATG (ou son équivalent AUG dans l'ARN) exerce deux fonctions. D'une part, il code pour l'acide aminé méthionine, d'autre part, il indique le début d'une séquence codante pour une protéine - le codon "départ". Trois triplets TAG (UAG), TAA (UAA) et TGA (UGA) - ne spécifient aucun acide aminé mais signalent plutôt la fin d'une séquence codante (codon "arrêt" ou codon "stop").

Le code génétique est dit dégénéré parce que le même acide aminé peut être spécifié par plus d'un codon mais, point très important, il n'est nullement ambigu puisqu'un codon donné ne spécifie qu'un seul acide aminé (Leroux et Tossier-Klopp, 2000).

Mécanisme de traduction

La traduction de l'ARNm fait intervenir un grand nombre de macromolécules dont des ARN de transfert (ARNt) des enzymes et des ribosomes. Les ARNt sont de petits ARN de 74 à 94 nucléotides ayant une structure en forme de trèfle du fait de l'appariement de certaines bases (Harry, 2008). Les ribosomes, chez les eucaryotes, comportent 2 sous-unités : 60S et 40S. La sous-unité 60S est constituée de 3 ARNr (5, 5,8 et 28S) et de 49 protéines.

La sous-unité 40S est formée de l'assemblage d'un seul ARNr (18S) et de 33 protéines. Les ribosomes ont deux sites pour lier les ARNt : le site A (site acide aminé) où viendra l'ARNt porteur de l'acide aminé et le site P (site peptidique) pour l'ARNt porteur de la chaîne peptidique en cours d'élongation (Figure 23) (Leroux et Tossier-Klopp, 2000).

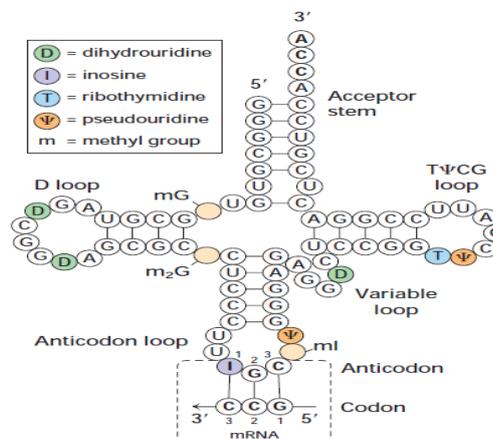


Figure 23 : Structure de l'ARN de transfert (ARNt) (Lodish et *al.*, 2003).

La traduction se déroule en trois étapes : l'initiation, l'élongation et la terminaison.

Phase d'initiation. Elle débute par la reconnaissance et la fixation de l'extrémité 5' (chapeau ou coiffe) de l'ARNm et un complexe "ARNt.Met initiateur + facteurs d'initiation eIF + sous-unité ribosomique 40S". L'ensemble va ensuite migrer dans le sens 5' vers 3' jusqu'au codon d'initiation AUG. Les nucléotides entourant le triplet initiateur sont très comparables d'un ARNm à l'autre. Lorsque, dans la partie 5' de l'ARNm, deux triplets AUG sont présents, la comparaison des nucléotides entourant ces codons peut aider à prévoir lequel aura le plus de chance d'être initiateur. Une fois le site d'initiation atteint, les facteurs d'initiation sont libérés et la sous-unité ribosomique 60S s'associe à la 40S pour donner un ribosome fonctionnel (Leroux et Tosser-Klopp, 2000).

Phase d'élongation. L'acide aminé suivant dans la chaîne polypeptidique est amené au site ribosomique de traduction par un ARNt dont l'anticodon reconnaît par appariement de bases le deuxième codon de l'ARNm. La première liaison peptidique est alors formée entre la méthionine et le deuxième acide aminé : la chaîne peptidique est commencée. Au fur et à mesure que les codons sont traduits, les acides aminés sont ajoutés à la chaîne naissante. Ce processus se répète jusqu'à ce que tous les codons de la séquence codante aient été traduits. Cette étape d'élongation fait intervenir des facteurs d'élongation dont *eEF-1* et *eEF-2* respectivement nécessaires à la fixation de l'amino acylARNt sur le site A et à la translocation du peptidyl-ARNt du site A vers le site P. Cette translocation se traduit par un glissement par triplet du ribosome sur l'ARNm (Figure 24).

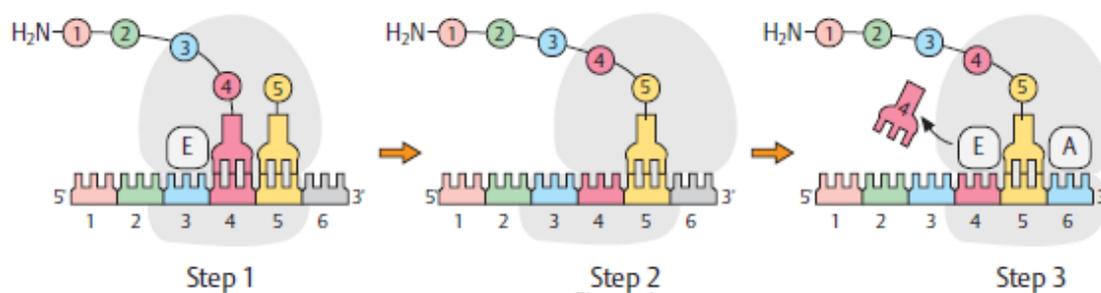


Figure 24 : Etape de la traduction d'un ARNm (Passarge, 2007).

Enfin, lorsque la machinerie de traduction atteint le signal de fin de traduction (l'un des trois codons stop UAA, UAG ou UGA), la chaîne polypeptidique complète est libérée après l'hydrolyse, par une peptidyl-transférase, de la liaison ester unissant le dernier ARNt au dernier acide aminé de la chaîne. A cette dernière étape de la traduction, la fixation du facteur protéique de terminaison sur le codon stop provoque la libération du dernier ARNt et de la chaîne peptidique. Le ribosome se dissocie en deux sous-unités qui pourront recommencer une nouvelle lecture d'ARNm (Leroux et Tosser-Klopp, 2000).

3.2 Maturation des protéines

b) Modifications post-traductionnelles

La traduction n'est pas l'aboutissement de la voie d'expression du génome. Le polypeptide qui émerge du ribosome est inactif, il doit subir au moins la première des quatre réactions de maturation post-traductionnelle avant de pouvoir assurer son rôle fonctionnel dans la cellule :

- *Repliement des protéines*. Le polypeptide est inactif tant qu'il n'est pas replié pour présenter sa structure tertiaire correcte.

- *Coupure protéolytique*. Certaines protéines vont être coupées, au cours de leur maturation, par des enzymes appelées protéases. Ces coupures peuvent éliminer des segments à une extrémité du polypeptide ou aux deux, ce qui donne une forme raccourcie de la protéine, ou elles peuvent couper le polypeptide en plusieurs morceaux qui peuvent ou non être actifs.

- *Modifications chimiques*. Des acides aminés particuliers du polypeptide peuvent être modifiés par la fixation de nouveaux groupements chimiques.

- *Epissage de l'intéine*. Les intéines sont des séquences interférentes dans certaines protéines, semblables, dans une certaine mesure, aux introns des ARNm. Elles doivent être éliminées et les extéines ligaturées pour que la protéine devienne active (Brown, 2004).

4. Régulation du fonctionnement et de l'expression des gènes

4.1 Structure chromatinienne des gènes actifs

Une digestion partielle par faibles quantités de *DNase I* permet de définir des zones d'accessibilité préférentielle à l'enzyme. Il existe deux niveaux de sensibilité, d'où les notions de *sites sensibles* et de *sites hypersensibles*. Cette sensibilité à la *DNase I* reflète des structures particulières, dites ouvertes, de la chromatine. Elle résulte, en grande partie : 1) de

modifications post-traductionnelle des histones (acétylation, méthylation, etc.) qui diminuent les forces d'interaction entre les histones du nucléosome et l'ADN ; 2) des *remodelages* de la chromatine qui, eux aussi, ont pour but de rendre l'ADN plus accessible au niveau de promoteurs ; 3) de la fixation de *facteurs transcriptionnels* sur l'ADN (Kaplan et Delpech, 2007)

4.2 Modification de la structure primaire de l'ADN

Certaines séquences *CG* de l'ADN peuvent être méthylées, la méthylation ne portant que sur les cytosines. Un ensemble de résultats expérimentaux ont montré que la méthylation des cytosines situées dans la région en 5' non transcrite des gènes était associée à une diminution de leur activité transcriptionnelle.

L'ADN des cellules cancéreuses, dont la transcription est très active, est à l'inverse largement hypométhylé. Le chromosome X inactivé est globalement hyperméthylé, ce qui peut être mis en relation avec l'inactivation de la plupart de ses gènes.

L'effet inhibiteur de la méthylation sur la transcription résulte d'au moins deux types de mécanismes possibles :

- La méthylation modifie la forme et les caractères physicochimiques de la molécule d'ADN, ce qui modifie la reconnaissance des séquences par les facteurs de transcription.

- L'autre mécanisme est lié à l'intervention de protéines *MBD* (*Methyl Binding Domain*), protéines qui contiennent un domaine de fixation à l'ADN méthylé. Leur fixation interdit l'accès des facteurs de transcription à l'ADN. Elles sont par ailleurs associées à d'autres molécules comme les histone-désacétylases qui ont un effet inhibiteur sur la transcription.

4.3 Régulation transcriptionnelle et post-transcriptionnelle

Les mécanismes de la régulation post-transcriptionnelle résultent essentiellement d'interactions entre les régions non traduites situées aux extrémités 5' et 3' de l'ARNm, et des protéines à effet en *trans* qui contrôlent spécifiquement : (1) la localisation cytoplasmique ; (2) la traduction et/ou (3) la stabilité de l'ARNm (Pallier, 2001).

- Au niveau transcriptionnel

L'expression des gènes est souvent contrôlée par des signaux extracellulaires. Dans le cas des bactéries, ces signaux sont donnés par des molécules présentes dans le milieu de

culture. Ils sont ensuite transmis aux gènes par des protéines régulatrices de deux types : les régulateurs positifs ou activateurs et les régulateurs négatifs ou répresseurs. Ces régulateurs sont des protéines de liaison à l'ADN qui reconnaissent de façon spécifique des sites situés dans ou près des gènes qu'ils contrôlent. Des activateurs augmentent la transcription d'un gène, ils agissent sur les promoteurs faibles : des répresseurs diminuent ou bloquent sa transcription en se liant sur un site d'ADN qui recouvre le promoteur, empêchant ainsi la liaison de l'ARN polymérase (Watson et *al.*, 2012).

- Au niveau post-transcriptionnel

Les mécanismes de la régulation post-transcriptionnelle résultent essentiellement d'interactions entre les régions non traduites situées aux extrémités 5' et 3' de l'ARNm, et des protéines à effet *in trans* qui contrôlent spécifiquement: (1) la localisation cytoplasmique ; (2) la traduction et/ou (3) la stabilité de l'ARNm. Chez les eucaryotes, les ARN messagers sont synthétisés par l'ARN polymérase. Ils sont sous forme de précurseurs, appelés pré-ARN messagers. Au cours de leur maturation, une coiffe, constituée d'une guanine méthylée en position N7, est ajoutée à l'extrémité 5' de la plupart des pré-ARNm par une liaison 5'-5' triphosphate (Varani, 1997). Cette coiffe participe aux réactions d'épissage et de polyadénylation nucléaires, au transport nucléocytoplasmique des ARNm, à l'initiation de leur traduction et à leur protection contre la dégradation par les exoribonucléases dans le sens 5' 3' [2]. En aval de la région 5' non traduite (5'UTR), le cadre ouvert de lecture (ORF) débute avec le codon d'initiation AUG et s'achève avec le codon de terminaison de la traduction. Les pré-ARNm se terminent au niveau de la région 3' non traduite (3'UTR) à laquelle est ajoutée une séquence poly(A) au cours de leur maturation (Wahle et Kuhn, 1997). Tous les transcrits synthétisés par l'ARN polymérase sont polyadénylés, à l'exception de ceux des histones.

4.4 Régulations traductionnelle et post-traductionnelle

- Au niveau traductionnel

La régulation peut faire intervenir un régulateur traductionnel exogène qui peut être soit une protéine, soit un autre ARN qui interfère avec la traduction par le ribosome, en réponse à un signal extérieur (concentration d'un métabolite, température, présence d'un facteur protéique...). Ce régulateur peut être un activateur ou un répresseur, comme dans le cas de la régulation de la transcription. La régulation de la traduction se passe dans le cytoplasme de la

cellule, où se déroule cette étape de l'expression des gènes, tandis que la régulation de la transcription a lieu dans le noyau (chez les cellules eucaryotes).

- Au niveau post-traductionnel

L'acétylation : elle se fait sur des résidus lysine par les histones acétyl-transférases, ce qui diminue les charges positives des lysines, ce qui aurait pour conséquences de réduire les contacts avec l'ADN chargé négativement (Hong et al., 1993).

La méthylation : elle peut concerner les deux types d'acides aminés basiques, la lysine et l'arginine par l'ajout d'un groupement méthyl. La protéine qui ajoute la méthylation est l'histone méthyl-transférase et celle qui l'enlève est l'histone déméthylase. Si la méthylation se fait au niveau des résidus arginine, il y a un effet d'activation de la transcription et si les résidus lysine sont méthylés, il y a une fonction double (signal répresseur donc chromatine fermée ou formation d'euchromatine, c'est l'ouverture de la chromatine) (Winter et al., 2008).

La phosphorylation : elle concerne les sérines de l'histone *H3*, mais aussi les thréonines. Si on observe au niveau des promoteurs de gènes des histones *H3* avec une sérine 10 phosphorylée, alors elle est corrélée à une activité transcriptionnelle de ces gènes et si la kinase est inhibée, la phosphorylation de ces résidus est empêchée, ce qui aboutit à une mauvaise condensation du chromosome (Winter et al., 2008).

5. Voies de régulation des gènes par les signaux extracellulaires

Une cellule peut recevoir différents types de signaux en provenance d'autres cellules qui activent sa prolifération. Ces signaux sont portés par les facteurs de croissance, comme le facteur de croissance épithélial *EGF* (*Epithelial Growth Factor*) qui active la prolifération des cellules épithéliales. L'*EGF* est une protéine extracellulaire qui représente un signal de prolifération. Elle se lie à son récepteur présent à la surface des cellules épithéliales, l'*EGFR* (récepteur de l'*EGF*). Cette liaison active l'*EGFR* qui déclenche dans la cellule une cascade d'activations en chaîne de plusieurs protéines signalisatrices. Ces protéines s'organisent en une voie de signalisation de la prolifération appelée voie mitogène. La signalisation parvient jusqu'au noyau de la cellule où se trouve le génome. La voie mitogène active alors l'expression des gènes de la prolifération cellulaire. Les protéines qui orchestrent la prolifération cellulaire sont synthétisées : la cellule se multiplie en réponse au signal de prolifération porté par l'*EGF*.

Chapitre VI.
Méthodologie en Biologie Moléculaire

Chapitre VI. Méthodologie en Biologie Moléculaire

1. Méthodes de caractérisation et analyse de l'ADN

1.1 Extraction et purification de l'ADN

L'extraction de l'ADN est une technique permettant d'isoler l'ADN de cellules ou de tissus. L'ADN ainsi extrait peut ensuite être utilisé pour des recherches de biologie moléculaire, telles que le séquençage, la PCR ou le clonage. Il existe différents protocoles pour extraire l'ADN à partir soit de tissus (organes, biopsies.....), cellules en culture, bulbe de poils ou de cheveux, muqueuse buccale, salive, le sperme, l'urine, pellicules (peau morte), larmes, sécrétions vaginales ou des lymphocytes d'une prise de sang total sur EDTA (anticoagulant).

La technique la plus simple à réaliser est celle utilisant du NaCl pour précipiter les protéines associées à l'ADN (Miller et *al.*, 1988). L'utilisation de cette technique écarte tout risque de contamination ou d'intoxication par des produits dangereux comme le cas de l'extraction au phénol. Le matériel le plus pratique permettant d'obtenir de l'ADN est le sang total, fraîchement recueilli ou décongelé. Le sang doit être mélangé à une solution hypotonique pour faire éclater les globules rouges. Les solutions de lyse des globules rouges sont nombreuses et variées. La lyse est en général réalisée à 4 °C pendant 20 à 30 minutes. Le culot des lymphocytes est lysé en présence d'un détergent (SDS) et l'action de la *protéinase K* qui a pour but de digérer les protéines qui lui étaient associées. Le lysat est centrifugé et, après élimination du surnageant, le culot cellulaire contenant les leucocytes est repris dans une solution saline (NaCl 5 M) et est traité par une solution de lyse des globules blancs et enfin, l'ADN est précipité en un nuage cotonneux blanchâtre (la méduse) par l'éthanol absolu froid. L'ADN est lavé à l'éthanol absolu avant de redissoudre dans un tampon.

L'ADN est le plus souvent conservé à + 4°C en tampon (Tris-HCl 10 mM pH 8, EDTA 1 mM). Pour une longue conservation, la congélation à -80°C est préférable (Kamoun, 1997).

1.2 Fragmentation

Grâce aux enzymes de restriction, de fractionner la molécule d'ADN pour obtenir des fragments dont la taille est compatible avec l'utilisation des techniques de laboratoire. D'autres types de fragmentation sont également utilisés (sonication, fragmentation mécanique, irradiation...);

Une manière de "fragmenter" l'ADN est de ne cibler dès le départ qu'une petite portion du génome : l'utilisation de la PCR entre dans ce cadre (Riquet et Pitel, 2000).

1.3 Séparation analytique

La molécule d'ADN étant fragmentée, il faut séparer les différents fragments pour les visualiser individuellement ou pour sélectionner ceux qui nous intéressent. L'une des techniques appartenant au quotidien du biologiste moléculaire est l'électrophorèse. Elle permet de séparer les fragments d'ADN en fonction de leur taille : la migration dans un champ électrique est effectuée sur différents types de gels - dans lesquels la mobilité est influencée par la densité du maillage moléculaire - avec différentes conditions de migration, qui peuvent affecter la mobilité des fragments. En fonction du pourcentage d'agarose dans le gel, des fragments de quelques dizaines à quelques milliers de paires de bases peuvent être discriminés, en champ électrique constant. Ces gels sont utilisés couramment pour séparer des fragments d'ADN avant Southern-blot (voir plus loin), vérifier la qualité ou la quantité d'un échantillon d'ADN, déterminer la taille d'un produit PCR, d'une insertion d'ADN dans un vecteur... Grâce à l'utilisation de champs électriques alternatifs, on peut aller jusqu'à la séparation de fragments de plus de 10 000 kb : c'est l'électrophorèse en champs pulsés (Riquet et Pitel, 2000).

1.4 Visualisation

Les fragments d'ADN sont vérifiés par électrophorèse dans un gel d'agarose (Figure 25). En utilisant les propriétés d'agent intercalant et de fluorescence aux UV du bromure d'éthidium (BET) (colorant qui se complexe à l'ADN en s'intercalant entre les bases (Kamoun, 1997). Un marqueur de taille (piste 1) est également déposé. Il sert de contrôle pour déterminer la taille des ADN déposés (Griffiths et *al.*, 2001).

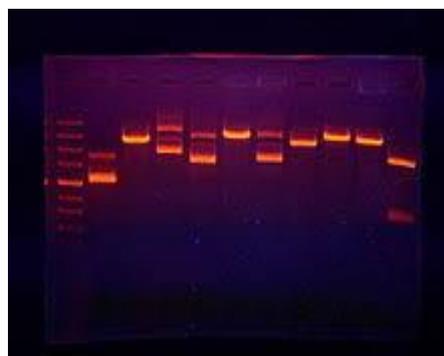


Figure 25 : Fragments d'ADN séparés par électrophorèse dans un gel d'agarose (Griffiths et *al.*, 2001).

1.5 Quantification

En biologie moléculaire, il est important de quantifier et d'analyser la pureté de l'ADN après sa purification. La méthode la plus répandue pour le dosage de l'ADN est la spectrophotométrie qui mesure l'absorbance (ou densité optique) des acides nucléiques à 260 nm (absorbent dans l'ultraviolet). Une unité d'absorbance à 260nm correspond à une concentration de 50 ng/μl d'ADN double-brin, ou 33 ng/μl d'ADN simple-brin (Sambrook et *al.*, 1989). Parallèlement leur pureté est déterminée en mesurant l'absorbance des protéines à 280 nm. Le ratio entre les valeurs à 260 nm et à 280 nm fournit une estimation du degré de pureté des acides nucléiques (Warburg et Christian, 1942). Des préparations pures d'ADN ont des valeurs à DO_{260}/DO_{280} de 1,8 et 2,0 respectivement. Le spectre d'absorption d'une solution pure d'acides nucléiques est représenté par une courbe caractéristique représenté par la Figure 26. Un changement de profil de cette courbe est un bon indicateur de l'impureté d'une solution y compris pour les solutions d'acides nucléiques.

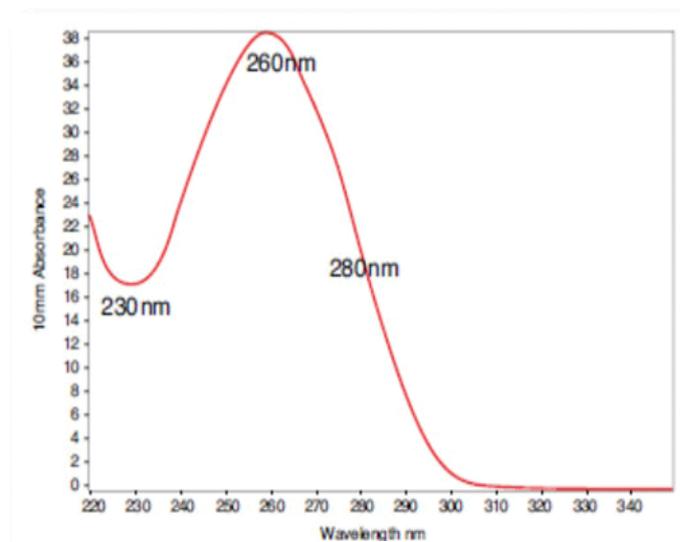


Figure 26 : Spectre d'absorption caractéristique d'ADN pur (Warburg et Christian, 1942).

La formule suivante présente l'estimation de la concentration de l'ADN :

$DO_{260} \times 50 \times \text{facteur de dilution}$ (Warburg et Christian, 1942).

- 1 unité $A_{260\text{nm}}$ = ~50 μg/ml d'ADN double brin
- 1 unité $A_{260\text{nm}}$ = ~33 μg/ml d'ADN simple brin
- 1 unité $A_{260\text{nm}}$ = ~40 μg/ml d'ARN simple brin

1.6 Hybridation et microarrays

Dans les microarrays la densité du dépôt est plus élevée (1000 ADN/cm^2) et il se fait sur une lame de verre. L'ADN de l'individu est amplifié puis fragmenté et chargé sur la puce (étape 1). Les fragments s'apparient spécifiquement à l'amorce correspondante (étape 2). Une fois appariés, il y a une réaction d'extension d'une base avec des nucléotides marqués (étape 3). Les bases insérées sont révélées par activation du fluorochrome. Le signal résultant est lu et analysé par le logiciel *Genome Studio*, qui renvoie automatiquement le génotype du marqueur SNP (*Simple Nucleotide Polymorphism*) pour chaque individu (étapes 4 et 5) (Figure 27) (Gunderson et al., 2005). Les utilisations sont multiples : analyse de l'expression de gènes, détection et caractérisation d'espèces, recherche de mutations...

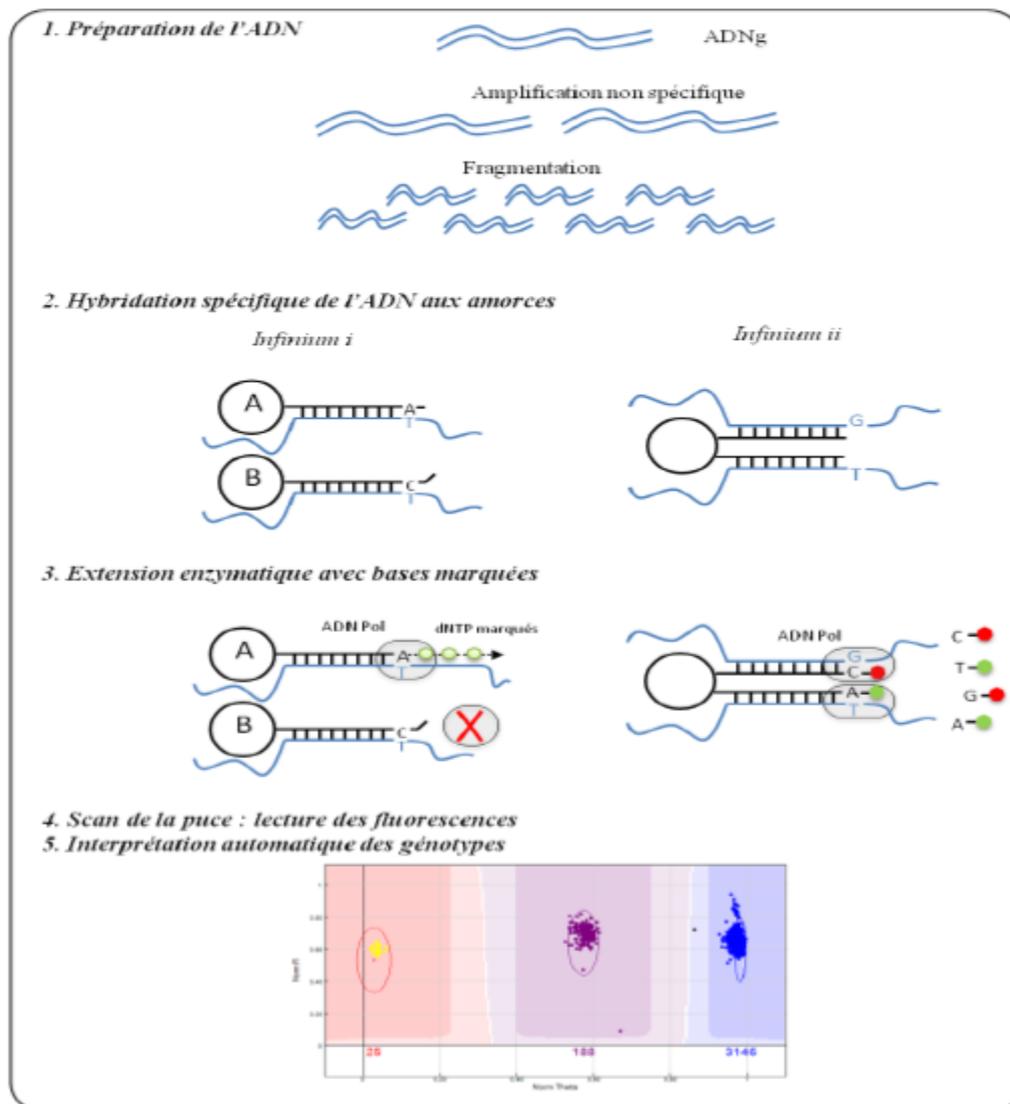


Figure 27 : Procédé général du génotypage d'un marqueur SNP sur une puce Illumina utilisant la technologie *Infinium assay* (Gunderson et al., 2005).

1.7 Amplification (PCR et ses applications)

La PCR (*Polymerase Chain Reaction*) ou en français Amplification en chaîne par polymérase a été découverte par Kary B. Mullis en 1985 et a obtenu le prix Nobel de chimie en 1993. L'utilisation de la technique de PCR (Mullis et *al.*, 1986) permet d'amplifier spécifiquement une région de l'ADN double brin de quelques centaines de paires de bases en des millions de copies, grâce à un appareil qui contrôle les changements de températures à différentes étapes de la réaction : température de dénaturation, d'hybridation et d'extension.

L'ADN doit d'abord être séparé en simples brins (dénaturation à 95°C), un couple d'amorces (oligonucléotides synthétiques d'environ 20 bases, complémentaires des deux extrémités de la région à amplifier) est ajouté qui va s'hybrider à 50-60°C environ avec la séquence complémentaire sur chacun des brins de l'ADN, les amorces sont choisies de façon à encadrer la séquence d'ADN à amplifier et les quatre désoxyribonucléotides (dNTP) qui serviront de substrats. Le tout est soumis à l'activité d'une ADN polymérase (*Taq polymerase*) qui synthétise à 72°C un brin complémentaire à partir du 3'OH de l'amorce hybridée. Le nombre de brins obtenus à la fin du premier cycle est le double du nombre de brins initialement présents. L'opération est répétée 35 fois, ce qui aboutit à des millions de copies qui représente une quantité suffisante pour étudier le fragment de DNA amplifié (Figure 28).

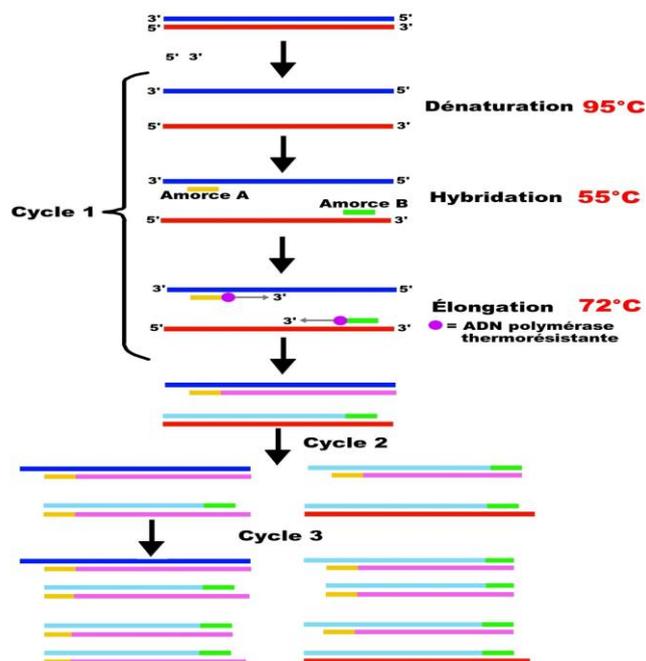


Figure 28 : Amplification génique par PCR.

(<https://microbiologiemedicale.fr/biologie-moleculaire-amplification-genique-pcr/>)

Applications de la PCR

La PCR a de multiples applications :

- Déterminer la présence d'une mutation dans le cas d'une maladie génétique, notamment en dépistage prénatal
- Détecter spécifiquement une bactérie lors d'une infection, en utilisant des amorces spécifiques d'un gène bactérien donne
- Détecter des gènes ou d'Organismes Génétiquement Modifiés (OGM)
- Recherche d'une pathologie en : parasitologie, bactériologie, virologie
- Détection du polymorphisme génétique
- Séquençage
- Recherche de paternité.

1.8 Séquençage

Le séquençage de l'ADN constitue une méthode dont le but est de déterminer la succession linéaire des bases A, C, G et T prenant part à la structure de l'ADN. La lecture de cette séquence permet d'étudier l'information biologique contenue par celle-ci. Les deux premières techniques de séquençage de l'ADN : celle de Maxam-Gilbert (1973) et celle de Sanger et collaborateurs (1977) ont été décrites en 1977.

a) Méthode chimique de Maxam-Gilbert

Il s'agit d'une méthode chimique de séquençage. Les réactifs clivent spécifiquement après chacune des bases A, C, G, [A + G], [C + T]. Cette technique est basée sur la propriété de certains agents chimiques, l'hydrazine, le diméthyl sulfate (DMS) et l'acide formique, de modifier les bases de l'ADN. Dans un second temps, la pipéridine est ajoutée et "casse" les brins d'ADN au niveau des bases modifiées.

Les agents chimiques sont utilisés dans des conditions telles qu'ils n'agissent qu'avec un faible pourcentage des bases de l'ADN étudié. Le DMS agit au niveau des bases "G". L'acide formique agit au niveau des bases "A + G". L'hydrazine agit au niveau des bases "C + T" (en milieu alcalin, l'hydrazine agit uniquement au niveau des "C"). L'ADN à séquencer est marqué à une extrémité. Le plus souvent, il s'agit d'un marqueur radioactif.

Le produit de séquence est déposé sur un gel d'acrylamide, puis la séquence lue après autoradiographie (Figure 29). L'ADN étudié peut être simple ou double brin. Cette technique permettait d'analyser des fragments allant jusqu'à 500 pb.

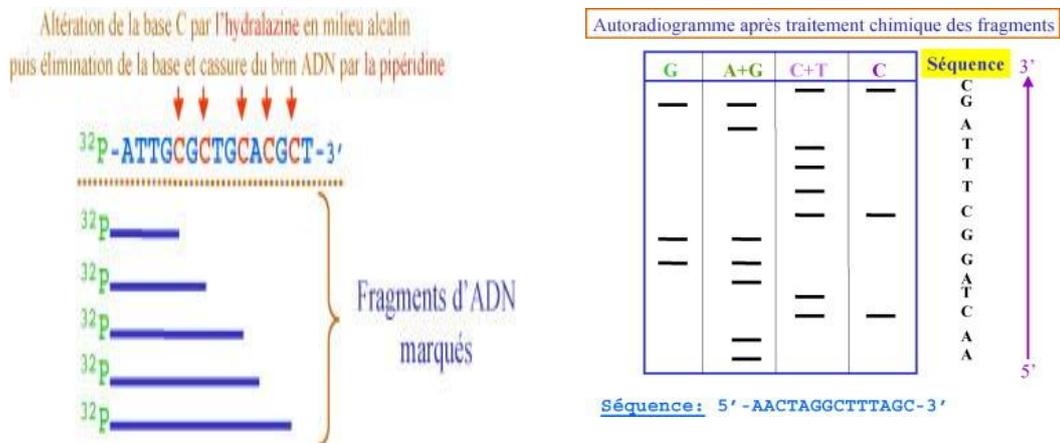


Figure 29 : Méthode de séquençage selon Maxam et Gilbert (Lamoril et *al.*, 2008)

b) Méthode enzymatique de Sanger

Dans un premier temps, il est nécessaire d'amplifier l'ADN cible par PCR, puis de le dénaturer afin d'obtenir un ADN simple brin. À l'aide d'une amorce spécifique et complémentaire du brin étudié (sens ou antisens), identique ou différente de celle utilisée pour la PCR, une ADN polymérase effectue alors la synthèse de l'ADN complémentaire à partir de cette amorce. De l'extrémité 5' vers l'extrémité 3', cette enzyme ajoute les désoxyribonucléotides-triphosphates (dNTP) complémentaires et de manière aléatoire et inconstante des didésoxyribonucléotides triphosphates (ddNTP), par exemple un ddGTP sera parfois ajouté à la place d'un dGTP. La réaction se faisant dans un seul tube, les ddNTP (ddATP, ddGTP, ddCTP et ddTTP) sont marqués à l'aide de fluorophores différents pour chaque ddNTP (fluorophores "quatre couleurs").

Lorsqu'un ddNTP est incorporé à la place d'un dNTP, l'ADN polymérase ne peut plus continuer sa polymérisation. Il se forme donc une multitude de brins complémentaires inachevés, chacun terminé par un didésoxyribonucléotide fluorescent caractéristique de la base de ce dernier nucléotide. En séparant par électrophorèse ces fragments complémentaires on sépare dans le gel chacun des fragments, du plus petit au plus grand et on les détecte au passage par un faisceau laser qui excite la fluorescence et une cellule photoélectrique qui lit la lumière émise à chacune des longueurs d'onde des fluorescences caractéristiques des quatre

bases (Figure 30). L'ordinateur reçoit donc une série de mesure d'intensité lumineuse en forme de pics correspondant au passage de chacun des fragments : en interprétant la couleur de la fluorescence de chaque pic l'ordinateur écrit la séquence d'ADN

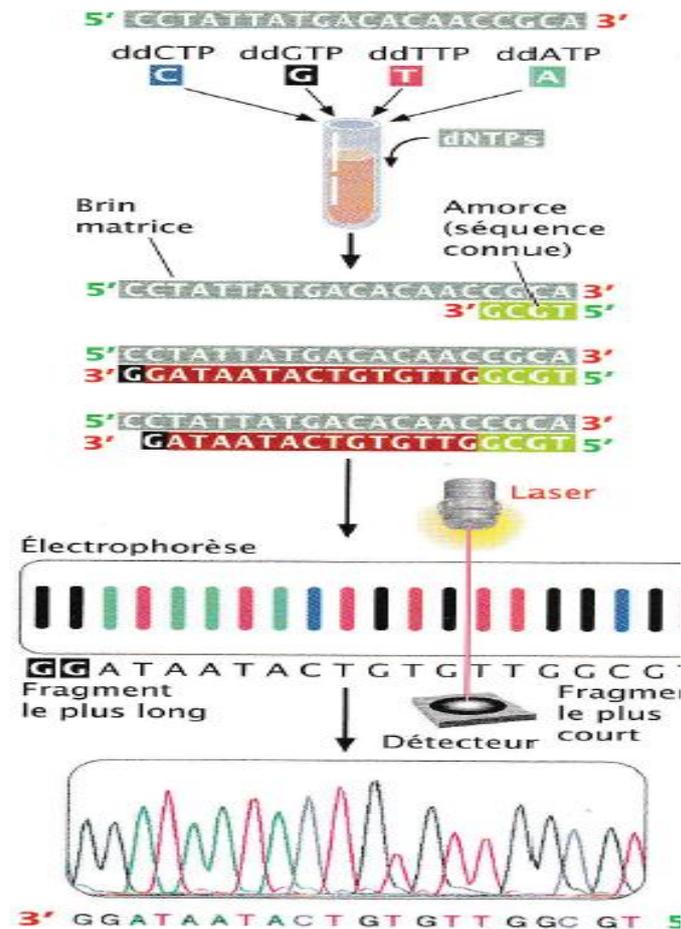


Figure 30 : Méthode de séquençage de Sanger (Pierce, 2012).

1.9 Restriction et analyse des polymorphismes

Le polymorphisme de restriction : variations individuelles d'une séquence de l'ADN révélée par des modifications de la longueur des fragments de restriction. Le profil électrophorétique obtenu avec une enzyme de restriction donnée montre des différences individuelles dans la taille des fragments de restriction (RFLP, *Restriction Fragment Length Polymorphism*) (Figure 31). Après une amplification de la région d'intérêt, le produit de PCR est soumis à la digestion par une enzyme de restriction (Lagziel et al., 2000).

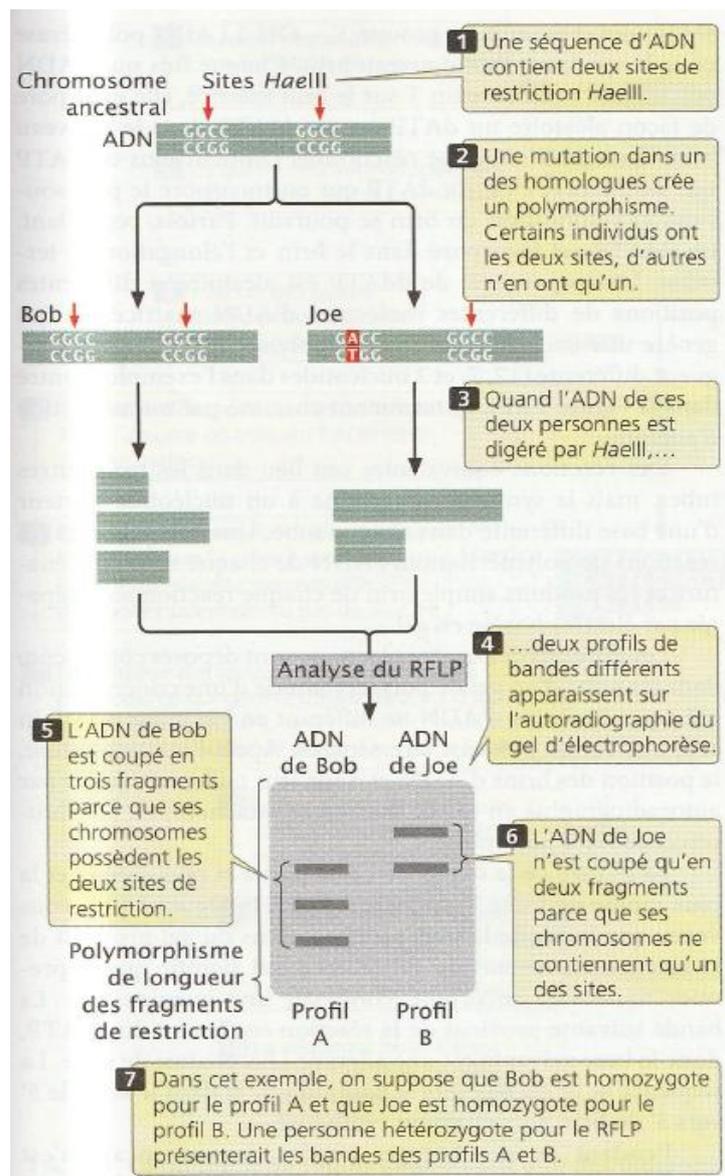


Figure 31 : Polymorphisme de longueur de fragments de restriction (Pierce, 2012).

1.10 Interaction avec les protéines

Les protéines régulatrices (facteurs trans-régulateurs) ont la capacité de se lier de façon spécifique à de courtes séquences de l'ADN et de réguler de façon positive ou négative la transcription d'un gène. Dans la plupart des cas la protéine s'insère dans le grand sillon de l'hélice et réalise une série de contacts moléculaires avec les paires de bases. La liaison avec l'ADN se fait par l'intermédiaire de plusieurs types interactions (liaison hydrogène, liaison ionique ou interaction hydrophobe).

B. Génie Génétique

Chapitre I.
Sources et préparation de l'ADN à cloner

B. Génie Génétique

Chapitre I. Sources et préparation de l'ADN à cloner

Le Génie Génétique est l'ensemble de techniques permettant de modifier artificiellement la composition génétique (génome) des cellules ou des organismes. Le terme Génie Génétique a été introduit au milieu des années 70. Ces différentes méthodes peuvent être regroupées en deux voies : la voie descendante, des acides nucléiques vers la protéine et la voie ascendante, des protéines vers les acides nucléiques. Le génie génétique utilise des outils cellulaires et moléculaires :

1. Outils cellulaires

L'isolement et l'amplification d'un gène impose son clonage et sa multiplication dans des cellules de procaryotes (bactéries) (*E. coli* se divise très rapidement environ 20mn) ou d'eucaryotes pouvant être mises en culture *in vitro*.

A partir de cellules données, il y a isolement d'un gène à cloner du génome de cette cellule et sa multiplication dans d'autres cellules appelées cellules hôtes.

Cet isolement et cette multiplication nécessitent l'utilisation d'outils moléculaires adaptés.

2. Outils moléculaires

Des enzymes de restriction permettent de travailler sur les acides nucléiques de l'ADN, de le modifier, de le couper, de recoller les morceaux etc...

Les vecteurs sont capable de transporter l'ADN d'intérêt (d'où le nom de vecteur) dans différents types d'hôtes, c'est un morceau d'ADN capable d'autoréplication. Ce sont les vecteurs de clonage.

1. ADN génomique

L'ADN génomique (ADNg) est obtenu à partir de lignées cellulaires, de tissus ou d'organismes entiers en fonction de la taille de l'organisme. Pour un organisme donné on part de n'importe quel lot de cellules, l'ADN génomique sera le même. A partir du sang (lymphocytes) environ 10 ml, une extraction d'ADN peut être réalisée par les techniques nécessaires sous forme de longs filaments correspondant à l'ADN de haut poids moléculaires. Ces longues chaînes d'ADN peuvent être coupées par une ou plusieurs enzymes de restrictions en fragments de tailles variables.

2. ADN complémentaire

L'ADN complémentaire (ADNc) est obtenu à partir d'ARNm, qui servira à l'obtention d'une banque ADNc. Cet ADN ne contient que les séquences codantes d'un gène (Figure 32).

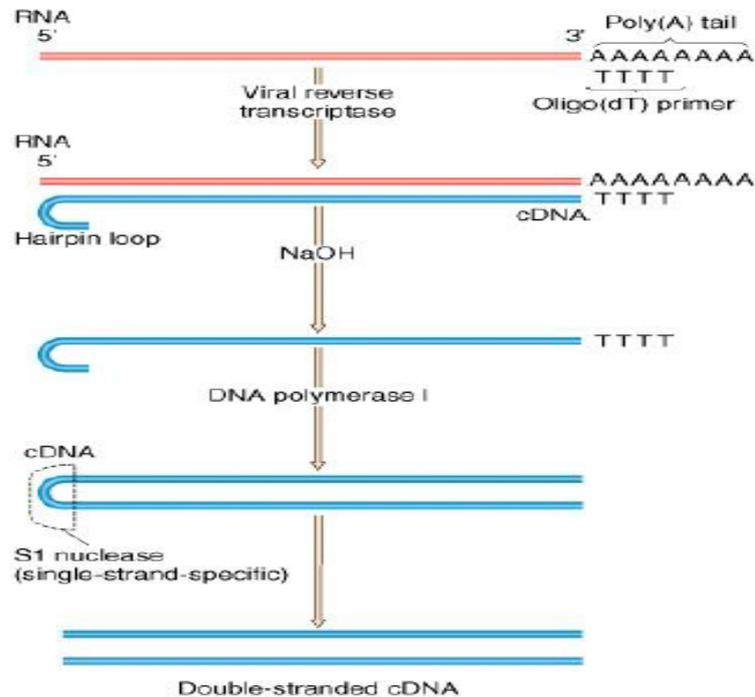


Figure 32 : Synthèse d'ADNc double brin à partir d'ARNm (Griffiths et *al.*, 2001).

3. ADN synthétique

Cet ADN est produit de façon chimique et synthétique pour des petits fragments de l'ordre de 50 à 100 bases par des automates. Ce type d'ADN est utilisé comme sonde ou amorces pour le criblage.

4. Notion de banques d'ADN génomique et complémentaire

Une banque d'ADN est un ensemble de larges fragments d'ADN d'un génome d'intérêt qui sont clonés dans un vecteur répliatif (exemple : plasmide) et introduit dans une cellule hôte facile à répliquer. On distingue deux types de banque ADN :

4.1 Banque d'ADN génomique

Une banque d'ADN génomique est un ensemble de vecteurs recombinants contenant chacun un fragment différent du génome de l'espèce étudiée. Le principe consiste à : (i) extraction de l'ADN génomique de l'espèce d'intérêt ; (ii) les enzymes de restriction sont

utilisées pour cliver l'ADN génomique en plusieurs fragments permettant de libérer des fragments de restriction de taille compatible avec le vecteur choisi ; (iii) incorporation des fragments d'ADN dans des vecteurs, par exemple dans de l'ADN de phage, et sont répliqués dans des bactéries (Figure 33). La banque obtenue est ainsi constituée de millions de clones recombinants (Tug et *al.*, 2018).

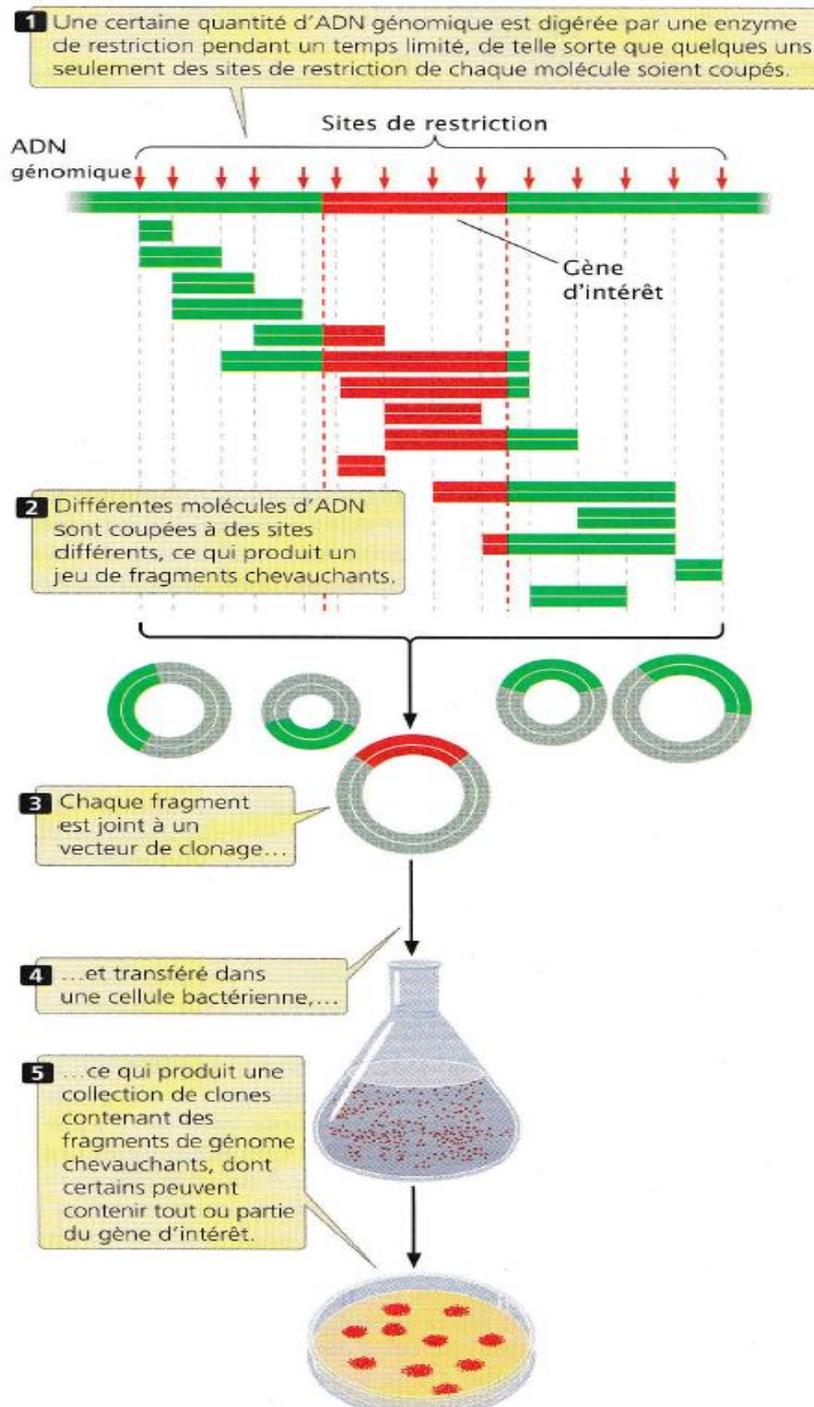


Figure 33 : Principe de construction d'une banque plasmidique d'ADN génomique total (Pierce, 2012).

Une banque génomique est caractérisée par :

- Le type de vecteur utilisé (plasmide, cosmide, phage, YAC...);
- La nature des inserts (ADNg, ADNc) et leur origine ;
- Le pourcentage de vecteurs possédant un insert ;
- La taille moyenne des inserts ;
- La taille de la banque (quantité de clones).

Avantages des banques génomiques

Les banques génomiques permettent de connaître:

- Les régions transcrites non traduites 5'UTR et 3' UTR ;
- Les séquences de régulation ;
- Les gènes silencieux transcrits et non traduits ;
- Les pseudogènes non transcrits et non traduits ;
- Les introns ;
- Les gènes chevauchants.

4.2 Banque ADN complémentaire (ADNc)

Les populations d'ARNm accumulées dans un tissu donné sont représentatives de ce tissu. Il est donc de première importance de pouvoir caractériser les gènes exprimés dans un organe déterminé. Mais la manipulation des ARNm est délicate (faible quantité, sensibilité aux nucléases) et il est nécessaire de transformer ces ARNm en ADN bicaténaires aisément manipulables et adaptés à des clonages dans des vecteurs bactériens. Une banque d'ADNc (ADN complémentaire) est donc représentative des populations d'ARNm présentes dans un tissu donné à un stade déterminé de son développement. Contrairement à la banque génomique, une banque d'ADNc sera spécifique d'un tissu. Une banque d'ADNc peut être considérée comme "une photographie instantanée" des populations d'ARNm représentées dans un organe.

La construction d'une banque d'ADNc nécessite de nombreuses étapes (Figure 34). Le principe repose sur : (i) l'extraction d'ARN et parfois la purification des ARNm polyadénylés de l'organe (par exemple par chromatographie d'affinité sur une colonne polyT) 1 ; (ii) la copie de ces ARNm en ADN complémentaires par l'action d'une transcriptase inverse ; (iii) l'élimination spécifique des ARNm par l'ARNase H ou la soude ; (iv) la synthèse du second brin d'ADN par une ADN polymérase ; (v) la ligature d'oligonucléotides (adaptateurs) pour créer des sites de restriction ; (vi) la ligature dans un vecteur de clonage (plasmide ou phage) ;

(vii) l'intégration des vecteurs recombinants dans une bactérie. Les vecteurs de clonage utilisés pour la construction de banques d'ADNc sont soit les plasmides, soit les phages. Les étapes de synthèse du second brin d'ADN et de ligature d'adaptateurs sont de plus en plus souvent réalisées à l'aide de la PCR. La synthèse du brin d'ADNc est amorcée généralement par fixation d'une courte séquence polyT sur l'extrémité polyA de l'ARNm. Une autre approche consiste à utiliser comme amorce pour la synthèse du brin d'ADNc par la transcriptase inverse un mélange d'hexanucléotides synthétiques correspondant à de nombreuses séquences différentes et s'hybridant au hasard sur l'ARNm. Cependant, comme par définition l'amorçage de la synthèse de l'ADNc se fait sur des parties internes des ARNm, les ADNc obtenus ne correspondent qu'à des fragments d'ARNm (Tagu et *al.*, 2018).

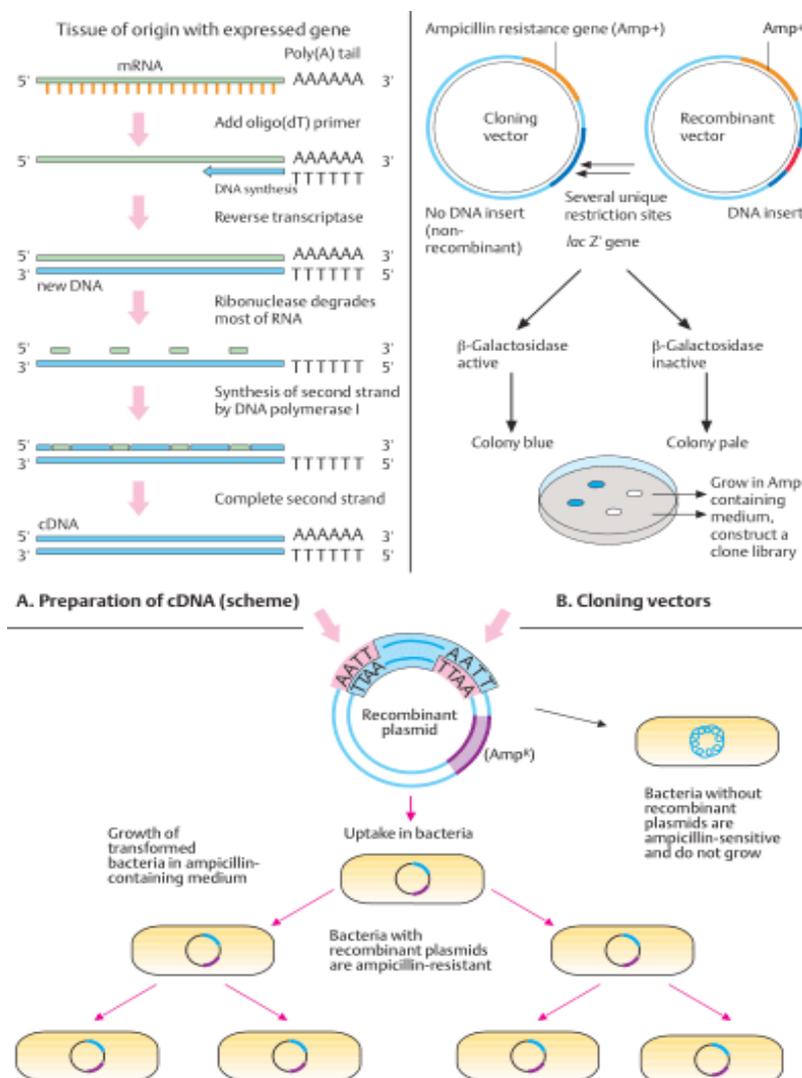


Figure 34 : Principe du Clonage de l'ADNc.

(<https://www.magazinescience.com/tag/adnc/?print=print-search>).

L'ADN complémentaire (ADNc) et l'ADN génomique sont deux types de molécules d'ADN utilisées dans les expériences de recherche en biologie moléculaire. L'ADNc et l'ADN génomique sont tous deux constitués de nucléotides d'ADN. L'ADNc est produit par la transcription inverse de l'ARN extrait du tissu. Les types d'ARN utilisés pour la transcription inverse peuvent être l'ARN total, le pré-ARN, l'ARN, l'ARN ribosomal, ou l'ARNt. Cependant, l'ADN génomique peut être directement isolé de la cellule. La différence principale entre l'ADNc et l'ADN génomique est que L'ADNc représente le transcriptome d'un organisme particulier alors que l'ADN génomique représente le génome (Tableau 3).

Tableau 3 : Différence entre une banque d'ADNg /ADNc (Tagu et *al.*, 2018).

	Banque d'ADNc	Banque d'ADNg
Origine de l'ADN	ADNc obtenus par reverse transcription d'ARNm, or la présence d'un messenger donné dépend du tissu étudié. Le tissu et le stade de développement est à spécifier.	ADNg de n'importe quelle cellule. Digestion partielle pour obtenir des clones chevauchants.
Type de vecteur de clonage utilisé / Taille des fragments d'ADN	Plasmides : insert < 5-10 kb Phages : insert < 20 kb	Cosmides : insert < 45 kb BAC : insert < 300 kb YAC : insert < 1000-2000 kb
Utilisation	Production de protéines recombinantes. Absence d'introns dans la séquence d'ADNc : identification des cadres de lecture ouverts	Caractérisation d'une séquence génomique particulière (connaître la structure d'un gène, son promoteur). Etablissement de cartes physiques d'un génome. Séquençage à grande échelle de portions ou de la totalité du génome

Chapitre II.

Vecteurs de clonage

Chapitre II. Vecteurs de clonage

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

1. Vecteurs bactériens

1.1 Plasmides

Les plasmides sont des molécules d'ADN accessoires qui existent dans les cellules bactériennes, ils se répliquent indépendamment du génome bactérien. Bien que la plupart des plasmides soient circulaires, il en résulte aussi des linéaires. Certains plasmides peuvent avoir jusqu'à 50 copies par cellule hôte, tandis que d'autres seulement 1 à 2 copies par cellule hôte. La taille des plasmides peut aller de quelques kilobases à quelques centaines de kilobases (Hartl et Jones, 2003). Les plasmides sont utilisés comme vecteur de clonage. Ils présentent les caractéristiques suivantes (Figure 35) :

- Une origine de réplication (*séquence Ori*).

Un gène de résistance à un antibiotique (exemple : ampicilline). C'est un marqueur de sélection qui permet de sélectionner les bactéries ayant incorporé le plasmide de celles qui ne l'ont pas incorporé (détection de la présence du plasmide dans les cellules hôtes). La présence de ce gène permettra à la bactérie porteuse de ce plasmide de ne pas être sensible à l'effet de cet antibiotique ;

- Un marqueur phénotypique, le gène *LacZ* qui permet de sélectionner les colonies n'ayant pas reçu le plasmide par la méthode de coloration bleue/blanche ;

- Une région avec des sites multiples et uniques pour des enzymes de restriction connues. Cette région est appelée *polylinker*. Son rôle est de permettre la linéarisation du plasmide et l'insertion d'un fragment d'ADN étranger.

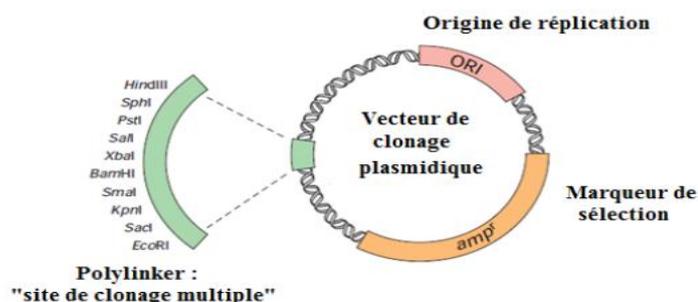


Figure 35 : Composants basiques d'un vecteur plasmidique chez *E. coli* (Lodish *et al.*, 2003).

Plasmides artificiels

Plasmide pBR322 (*plasmide Boliver Rodriguez*), le meilleur exemple de la première génération de vecteurs synthétiques. Ce plasmide a une taille de 4.3 kb. Il possède deux gènes de résistance aux antibiotiques (TET^R et AMP^R) et 40 sites de restriction dont 11 se trouvent dans le gène TET^R et deux autres sites se trouvent dans le promoteur de ce gène. Six sites uniques sont localisés dans le gène AMP^R des deux gènes de résistance aux antibiotiques. Une insertion réussie au niveau du gène de l'ampicilline se traduira par l'inactivation de ce gène qui ne sera plus capable de conférer la résistance à l'ampicilline à la cellule hôte (Primrose et *al.*, 2004) (Figure 36).

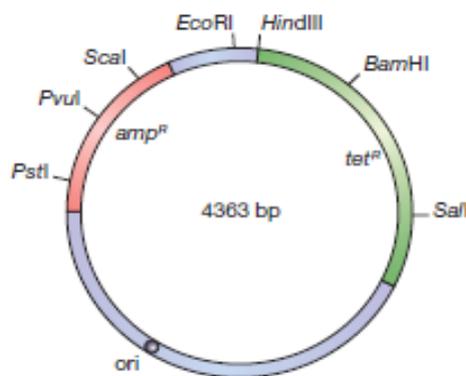


Figure 36 : Plasmide pBR322 (Brown, 2010).

Plasmide pUC18 (*plasmide University of California*) (Figure 37), présente en plus du gène AMP^R , une portion de l'opéron lactose dont le promoteur du gène *LacZ* d'*E. coli*. Dans cette région, un adaptateur ou un *polylinker*, comportant de nombreux sites de restriction, a été introduit (Harry, 2008).

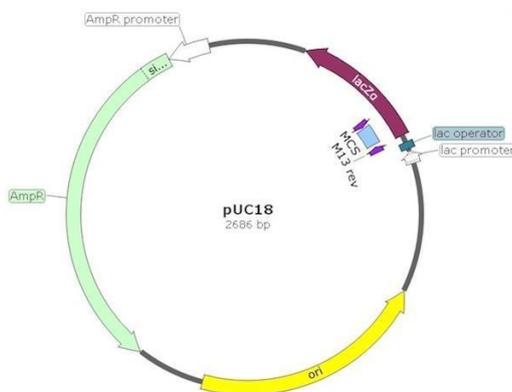


Figure 37 : Plasmide pUC18 (Lobet et *al.*, 1989).

- Avantages des plasmides
 - Petite taille du vecteur, permettant un travail expérimental aisé
 - Sélection des plasmides recombinants (sélection par les antibiotiques).
- Inconvénients des plasmides
 - Faible efficacité pour la transformation des bactéries (intégration du plasmide)
 - Impossibilité d'insérer de grands fragments d'ADN.

1.2 Phages

1.2.1 Propriétés générales des phages

Les bactéries sont sensibles à l'attaque de petits organismes appelés *bactériophages* ou simplement *phages*. Ce sont de petites particules, appartenant à la classe des virus, qui ne peuvent se multiplier qu'à l'intérieur d'une bactérie. Les molécules protéiques peuvent être organisées suivant trois modèles. Dans le modèle le plus commun, elles forment une enveloppe appelée *capside* ou *tête du phage*, à laquelle est généralement attachée une *queue* (Figure 38). L'acide nucléique des phages peut être de l'ADN double-brin (le plus fréquent), de l'AN simple-brin, de l'ARN simple-brin ou de l'ARN double-brin (le moins fréquent) (Freifelder, 1990).

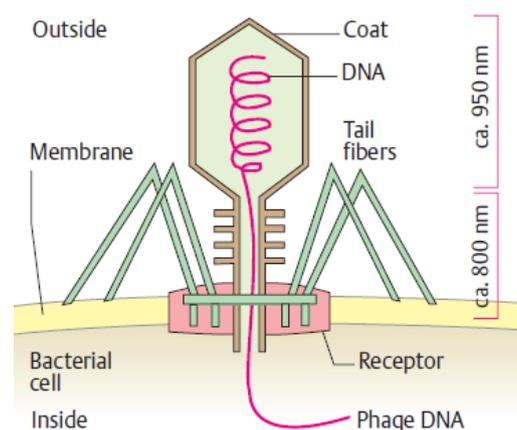


Figure 38 : Fixation d'un phage (Passarge, 2007).

Au cours de l'infection, un phage se fixe à une bactérie et injecte son matériel génétique dans le cytoplasme bactérien. De nombreux phages descendants sont libérés lorsque la paroi de la cellule bactérienne se rompt. Ce processus d'éclatement est appelé *lyse*.

A contrario, les *phages tempérés* s'intègrent dans le chromosome bactérien. Le phage inséré est alors répliqué en même temps que le chromosome bactérien. Dans ces conditions, le phage est appelé un *prophage* et l'hôte bactérien est dit *lysogénique* (Figure 39). Une bactérie *lysogénique* ou *lysogène* est résistante à toute infection ultérieure, car la présence du prophage lui assure une immunité. L'état lysogène peut être transmis génétiquement pendant de nombreuses générations bactériennes. Toutefois, un prophage peut à l'occasion s'exciser du chromosome bactérien et s'engager dans un cycle lytique (Griffiths et al., 2001).

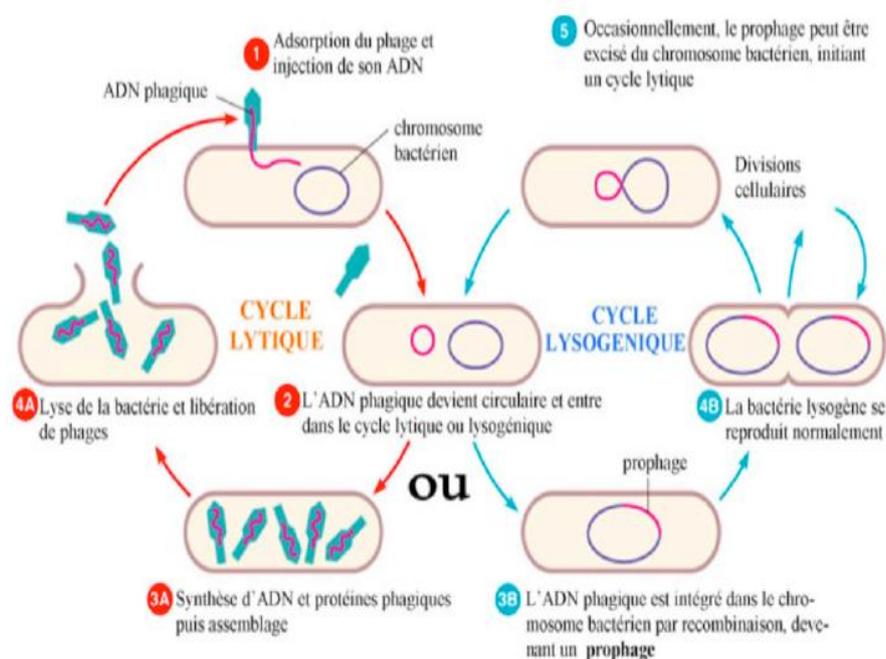


Figure 39 : Différents cycles de réplication des phages (Trotreau et al., 2018).

1.2.2 Types de phages

➤ *Phage lambda* (λ)

Le phage λ est un vecteur de clonage commode pour plusieurs raisons. Une tête de phage empaquète un chromosome de façon sélective un chromosome d'environ 50 kb de

long. A chaque extrémité de la molécule se trouve un court tronçon de 12 nucléotides dans lequel l'ADN est monocaténaire. Les deux brins simples sont complémentaires, et peuvent donc s'apparier l'un à l'autre pour former une molécule circulaire, complètement double brin (Figure 40). Les extrémités cohésives du phage λ sont appelées sites *cos* et jouent deux rôles distincts au cours du cycle d'infection. Premièrement, ils permettent de circulariser la molécule d'ADN linéaire injectée dans la cellule, ce qui est un préalable nécessaire à l'insertion dans le génome bactérien. Le deuxième rôle est assez différent, et entre en jeu après que le prophage s'est excisé du génome de l'hôte. A ce stade, un grand nombre de nouvelles molécules d'ADN e sont produites par le mécanisme de réplication en cercle roulant; dans lequel un brin d'ADN continu est "roulé" de la molécule matrice (Griffiths et *al.*, 2001).

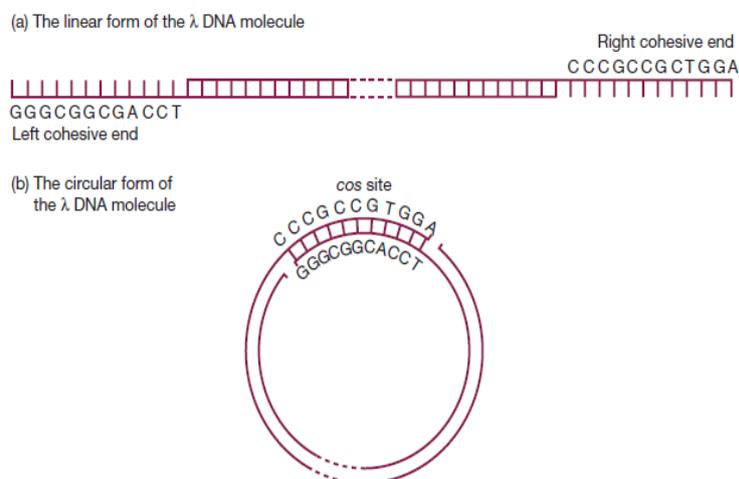


Figure 40 : Formes linéaires et circulaires de l'ADN λ . (a) forme linéaire, montrant les extrémités cohésives gauche et droite. (b) appariement de bases entre les extrémités cohésives donne la forme circulaire de la molécule (Brown, 2010).

La région centrale du génome du phage λ n'est pas nécessaire pour la réplication ou l'encapsulation des molécules d'ADN de λ dans *E. coli* et peut donc être excisée par des enzymes de restriction. Les "deux bras" sont réunis à l'ADN donneur digéré par les mêmes enzymes de restriction. Les molécules chimériques peuvent, soit être introduites dans des bactéries *E. coli* par la transformation, soit empaquetées dans des têtes de phage *in vitro*.

Dans les deux méthodes, les molécules recombinantes contenant des inserts de 10 à 15 kb seront les plus efficacement empaquetées dans les têtes de phage, car cette taille d'insert

correspond à celle de la partie centrale déléetée du génome phagique. La molécule a ainsi une taille totale de 50 kb. La présence d'une plaque de lyse phagique sur le tapis bactérien signale automatiquement la présence d'un phage recombinant contenant un insert (Figure 41).

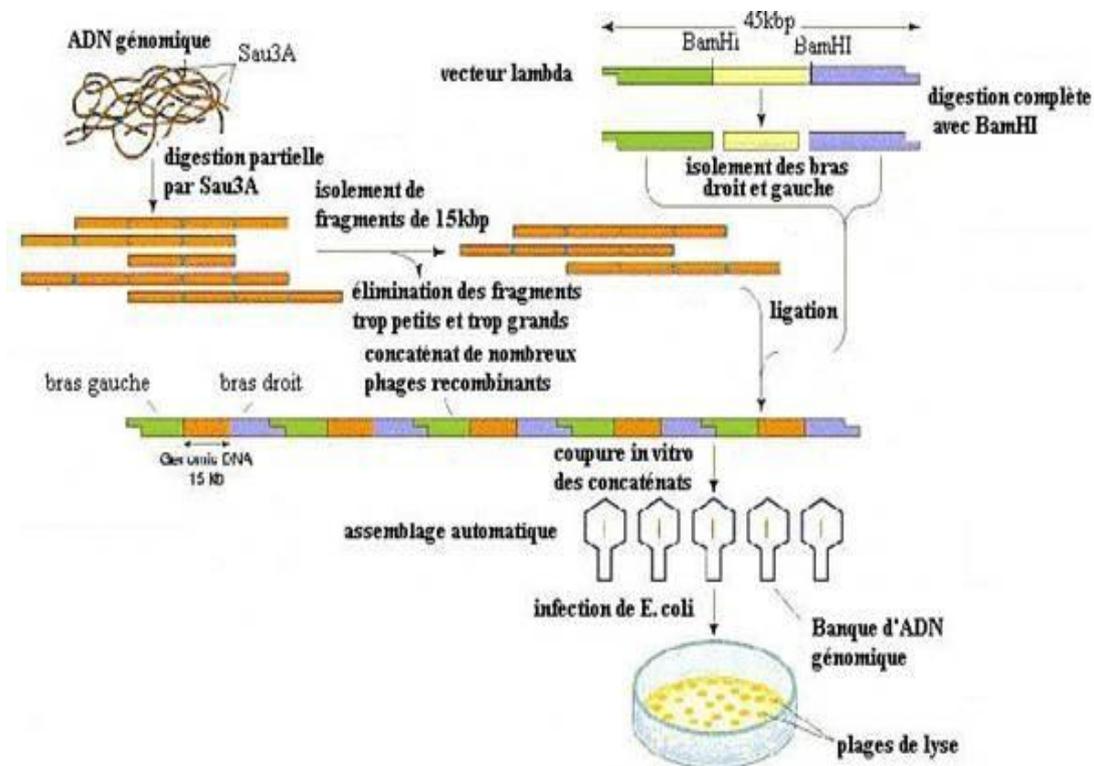


Figure 41 : Clonage dans le phage (Griffiths et *al.*, 2001).

➤ Phage λ gt10 et gt11

λ gt10 et λ gt11 étaient les vecteurs standard pour le clonage d'ADNc jusqu'à environ 1990. Les deux λ gt10 et λ gt11 sont des vecteurs d'insertion, et ils peuvent accepter respectivement environ 7,6 kb et 7,2 kb d'ADN étranger. Dans chaque cas, l'ADN étranger est introduit sur un site de clonage *EcoRI* unique. λ gt10 est utilisé pour créer des banques qui sont criblées par hybridation. Le site *EcoRI* interrompt le gène *cI* du phage, permettant une sélection sur la base de la morphologie de la plaque. λ gt11 contient un gène *lacZ* d'*E. coli* dirigé, les séquences d'ADNc clonées dans ce vecteur peuvent être exprimées sous forme de protéines de fusion de la β -galactosidase, et peuvent être détectées par criblage immunologique ou criblage avec d'autres ligands (Primrose et Tawyman, 2006).

➤ Phage M13

Les particules phagiques sont constitués d'un ADN simple brin circulaire bases noté brin (+). Lors de la réplication, un ADN négatif complémentaire est synthétisé. M13 est un petit bactériophage d'*E. coli* de 6 407 pb. Ses dérivés (M13mp18, M13mp19, ...) portent une partie du gène *lacZ* (peptide α) avec un *polyinker* de 13 sites de restriction (54 pb), ce qui permet d'insérer des fragments d'ADN dans des sites uniques et de sélectionner les colonies blanches sur des boîtes contenant l'analogue *X-gal* (Crépin, 1987).

➤ Phagemide

Les phagemides sont des molécules hybrides entre un plasmide et un phage. Ce sont des molécules bicaténaires, circulaires, qui peuvent être obtenues sous forme monocaténaire dans certaines conditions. Ils possèdent une origine de réplication, au moins un gène de résistance à un antibiotique, un site de polyclonage et une séquence provenant du phage M13 contenant l'origine de réplication qui permet d'obtenir la forme monocaténaire. En général, des promoteurs reconnus par des ARN polymérase ont été introduits en amont et/en aval du site de polyclonage, afin de pouvoir produire des ARN par transcription *in vitro*. Des fragments d'ADN de l'ordre de 4 kb peuvent être introduits dans ces vecteurs, mais il est parfois possible d'intégrer des inserts de l'ordre de 10 kb (Tug et al., 2018).

• Avantages des phages

- La taille des fragments d'ADN insérables est supérieure à celle des plasmides 40 kb
- La transformation des bactéries (incorporation des phages) est plus efficace que pour les plasmides.

• Inconvénients des phages

- Nombre de sites de restriction restreints dans le génome des phages.
- Obligation d'empaqueter l'ADN.

1.3 Cosmides

Les cosmides sont des vecteurs hybrides des phages λ et des plasmides. Leur ADN peut se répliquer dans la cellule comme celui d'un plasmide et être empaqueté comme celui d'un phage. Toutefois, les cosmides peuvent transporter des inserts d'ADN environ trois fois plus longs que ceux transportés par λ lui-même (jusqu'à 45 kb). Le secret de cette construction réside dans le fait que la majeure partie de la structure du phage λ a été délétée

mais que des séquences de signalisation responsables de l'empaquetage de l'acide nucléique dans la tête du phage (sites *cos*) sont toujours présentes (Figure 42).

Cette structure modifiée permet de remplir les têtes des phages avec quasiment n'importe quel ADN donneur. L'ADN cosmique peut être intégré dans des particules phagiques à l'aide d'un système *in vitro* (Griffiths et al., 2001).

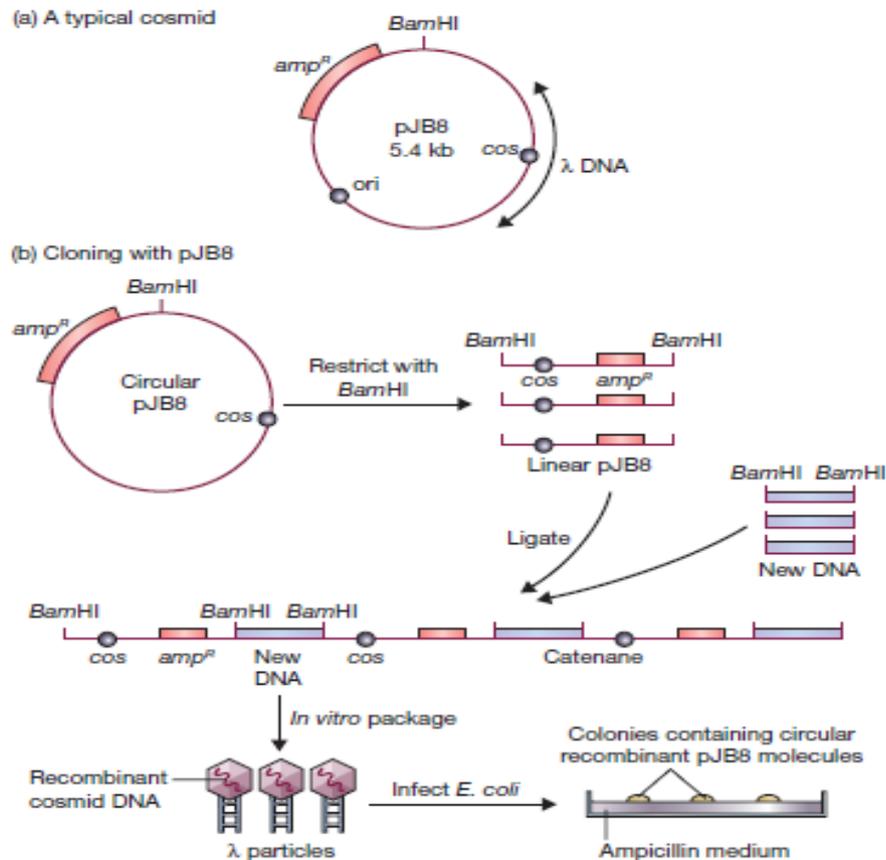


Figure 42 : Cosmide typique pour cloner de longs fragments d'ADN (Brown, 2010).

- Avantages des cosmides

- La taille des fragments insérables peut atteindre 50 kb
- Leur incorporation dans les bactéries (transformation) est plus efficace que pour les plasmides.

- Inconvénients des cosmides

- Obligation d'empaqueter l'ADN.

1.4 PAC

Les vecteurs PAC (*P1-derived Artificial Chromosome*) (Sternberg, 1990) sont des vecteurs de clonage similaires au vecteur λ obtenus à partir d'une version de λ dans laquelle une partie du génome naturel du phage a été délétée. Les capacités d'insertion dépendent de la taille du fragment de génome de λ délété et du nombre de bases comprises dans les espaces intermédiaires. Le génome de P1 est plus grand que celui de λ et la particule encapsidable plus longue, si bien qu'un vecteur P1 peut cloner des fragments d'ADN d'environ 25 kb en routine (Brown, 2004). Les éléments essentiels du vecteur sont contenus dans une séquence d'ADN de 17 kb :

- *ori* : origine de répllication plasmidique
- *kan^R* : gène de résistance à un antibiotique, la kanamycine
- *sacB* : gène de saccharose synthase; crible de sélection des bactéries recombinantes; site de clonage (*Bam*H1). La dégradation du saccharose produit un composé hautement toxique pour les bactéries. Sur un milieu contenant du saccharose, seules les bactéries dont le gène *sac B* est modifié se développeront (Hartl et Jones, 2003) (Figure 43).

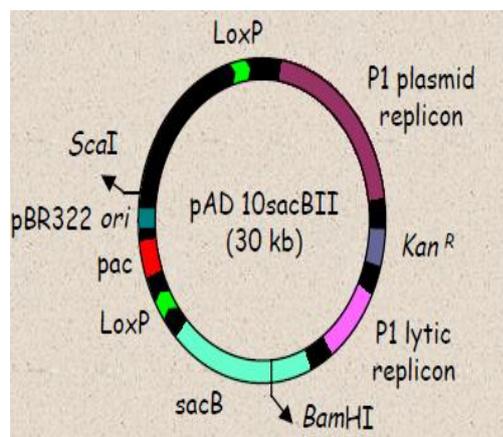


Figure 43 : Structure du vecteur PAC.

(<https://docplayer.fr/9227546-Clonage-dans-un-plasmide-puc.html>)

1.5 BAC

Les vecteurs BAC (*Bacterial Artificial Chromosome*) (Shizuya *et al.*, 1992) ont pour origine le *plasmide F* qui existe naturellement chez *E. coli*. Contrairement aux plasmides utilisés aux débuts des recherches, le plasmide F est relativement plus grand et les vecteurs en dérivant ont la capacité d'insérer des fragments d'ADN de 300 kb ou même plus. Le vecteur pBeloBAC11 Il contient des gènes qui contrôlent la répllication unidirectionnelle du facteur F

(*oriS* et *repE*) et son nombre de copies (*parA* et *parB*). La sélection des clones est réalisée avec le gène de résistance au chloramphénicol (*CMR*). Le clonage, souvent au site *HindIII*, interrompt le fragment *LacZ* (Rogel- Gaillard, 2000) (Figure 44).

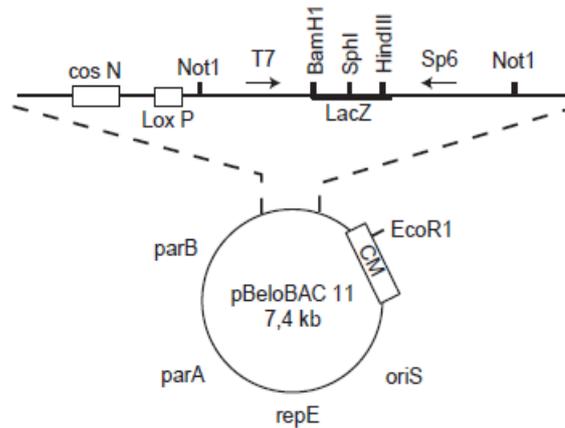


Figure 44 : Vecteur pBeloBAC11 (Rogel-Gaillard, 2000).

2. Vecteurs de clonage dans la levure

2.1 Vecteurs intégratifs

Les vecteurs intégratifs ou YIps (*Yeast Integrative plasmids*) ont la capacité d'insérer leur matériel génétique dans le génome de la cellule hôte (Baum et al., 2006). Ils sont essentiellement des vecteurs bactériens pBR322 contenant un gène de résistance à l'ampicilline (amp^R), un gène de résistance à la tétracycline (tet^R), un gène de levure (*URA3*) et l'origine de répllication spécifique d'*E. coli* (*ori*) (Brown, 2004).

Un exemple est le vecteur YIp5, qui est le pBR322 avec un gène *URA3* inséré. Un YIp ne peut pas se répliquer en tant que plasmide car il ne contient aucune partie du plasmide de 2 μ m et dépend à la place pour sa survie de l'intégration dans l'ADN chromosomique de levure, ils dépendent pour leur survie de leur intégration dans l'ADN chromosomique de levure par recombinaison homologue (Figure 45).

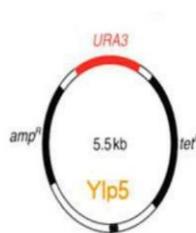


Figure 45 : Vecteur YIp5 (Brown, 2010).

2.2 Vecteurs autonomes dérivés du chromosome ou du plasmide 2 μ m

Le plasmide 2 μ m est un plasmide trouvé dans la plupart des souches de levure *Saccharomyces cerevisiae* (Figure 46a) à environ 60-100 copies par cellule. Le plasmide possède la nouvelle capacité d'amplification répliquative induite par la recombinaison site-spécifique (Frank et al., 1992). Le plasmide 2 μ m, comme on l'appelle, est l'un des rares plasmides trouvés dans les cellules eucaryotes. Les vecteurs dérivés du plasmide 2 μ m sont appelés plasmides épisomiques de levure YEps (*Yeast Episomal plasmids*). Certains YEps contiennent l'intégralité du plasmide 2 μ m, d'autres incluent uniquement l'origine de répllication. Un exemple de ce dernier type est YEp13 (Figure 46b).

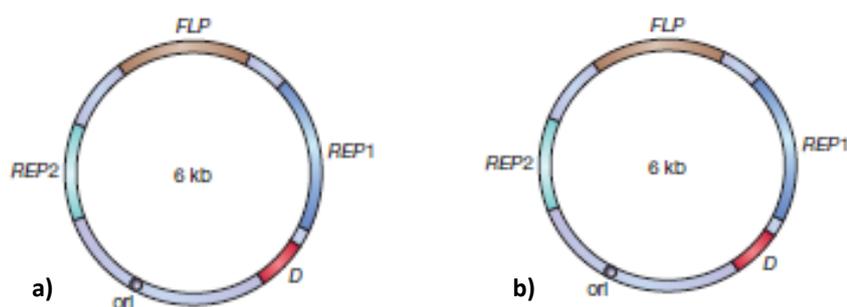


Figure 46 : Plasmides de levure. a) plasmide 2 μ m, b) plasmide YEp13.

REP1 et REP2 sont impliqués dans la répllication du plasmide et FLP code pour une protéine qui peut convertir la forme A du plasmide (représentée ici) en la forme B dans laquelle l'ordre des gènes a été réorganisé par recombinaison intramolécule. La fonction de D n'est pas connue avec précision (Brown, 2010).

2.3 Chromosomes artificiels

Les vecteurs YACs (*Yeast Artificial Chromosomes*) (Burke et al., 1987) sont dérivés des chromosomes de levures *Sacharomyces cerviciae*. Le génome de la levure *Saccharomyces cerevisiae* est constitué de 16 chromosomes de taille comprise entre 250 et 2000 kb. Chez la levure, trois régions chromosomiques sont importantes pour sa répllication. Une séquence ARS (*Autonomous Replicating Sequence*), une séquence centromérique *CEN* et les séquences télomériques *TEL* leur permettant de se comporter comme un chromosome une fois introduit sous forme linéarisée dans la levure (Billault et al., 1996). Plusieurs gènes de sélection (*URA3* et *TRP1*) codent pour des enzymes permettant de sélectionner les levures ayant incorporé un vecteur viable. Un site de clonage unique situé dans le gène (*SUP4*), permet l'introduction d'ADN étranger ainsi que la sélection des vecteurs recombinants. En effet, l'insertion d'ADN

interrompt le gène *SUP4* (Figure 47), ce qui permet la détection des colonies ayant incorporé un vecteur recombinant par apparition d'une coloration rose (Rogel-Gaillard, 2000).

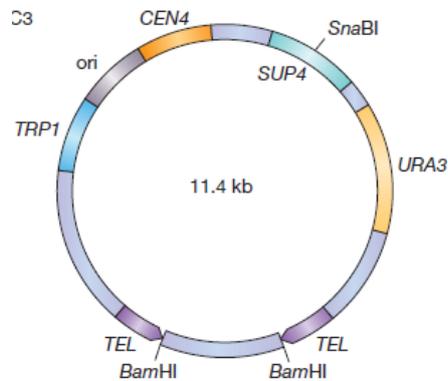


Figure 47 : Vecteur YAC (Brown, 2010).

Insertion d'un ADN exogène

L'insertion d'un ADN étranger est obtenue après coupure par *Bam*HI et *Eco*RI et qui permet d'obtenir les deux "bras" du YAC, chaque bras contenant un gène de sélection. Le système qui a été le plus couramment utilisé associe le vecteur de clonage *pYAC4* à la souche de levure *Saccharomyces cerevisiae* *AB1380*. Les bras gauche et droit contiennent des gènes qui synthétisent, respectivement, le tryptophane (*TRP1*) et l'uracile (*URA3*). L'insertion d'un fragment d'ADN génomique entre les deux bras du vecteur entraîne la coupure du gène de levure *SUP4* qui intervient dans la synthèse d'adénine. Ainsi, les clones recombinants sont repérés par leur couleur rouge sombre en comparaison des clones non recombinants qui sont blancs. De très grands fragments d'ADN exogène (de 150 à 1000 kb) peuvent être introduits dans les YACs.

Le pseudo-chromosome de levure est capable après introduction dans ce microorganisme de se répliquer (grâce à *ARS*), de stabiliser ses extrémités (grâce aux séquences *TEL*) et de se répartir entre les cellules filles lors de la division cellulaire (grâce à la séquence *CEN*) (Figure 48).

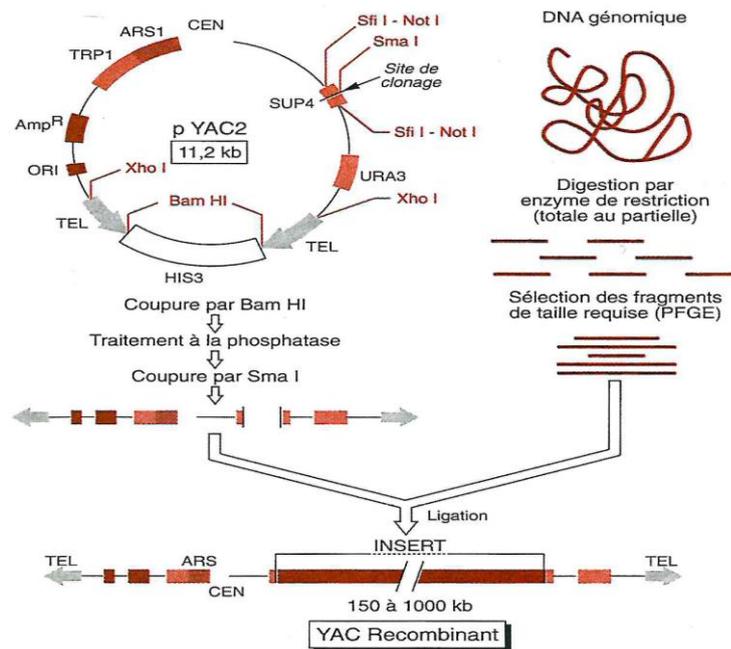


Figure 48 : Insertion d'ADN étranger dans un vecteur YAC (Strachan et Read, 1998).

• Avantages des YACs

L'avantage des YAC est que, contrairement aux autres vecteurs plasmidiques, leur stabilité augmente avec la taille de l'insert. Il n'y a donc pas de limite de taille pour les YAC et ils constituent un outil essentiel pour tous les projets de séquençage de génomes (Primrose et *al.*, 2004).

- Le vecteur YAC est un petit génome entièrement séquencé ;
- Croissance dans un milieu de culture simple ;
- Intégration des fragments de grande taille (jusqu'à 400 kb) ;
- Modification post-traductionnelle.

• Inconvénients des YACs

Le principal inconvénient des YAC, à l'heure actuelle, est leur instabilité mais cet inconvénient momentané ne met pas la technique en cause. Les YAC sont un outil de base dans le programme de séquençage du génome humain, mais également dans l'identification de gènes et dans la compréhension de certains aspects de leur régulation (Kochko, 2000).

De très grands fragments d'ADN exogène (de 150 à 1 000 kb) peuvent être introduits dans les YACs. Actuellement, le plus grand insert clone a une taille de 2 900 kb. Ces vecteurs sont donc de bons outils pour l'analyse des génomes complexes mais sont sujets à des instabilités. D'autres vecteurs permettent également le clonage de grands fragments d'ADN (Tableau 4): les *Bacterial Artificial Chromosome* (BAC) ou les *Pl-derived Artificial Chromosome* (PAC). Ces vecteurs sont capables, tout comme les YACs, d'incorporer de grands fragments d'ADN. Les BACs se comportent comme des plasmides de très grande taille alors que les PAC se manipulent comme des bactériophages de type 1. Les BACs sont de plus en plus utilisés : par rapport aux YACs, le clonage de l'ADN y est plus aisé et efficace, et les clones sont plus stables dans *E. coli* que dans la Levure. Cependant, la taille des fragments intégrés ne dépasse pas 300 kb.

Tableau 4 : Différents vecteurs et leurs hôtes (Tagu et *al.*, 2018).

Vecteurs	Forme du vecteur	Hôte	Taille de l'insert	Usage majeur
Plasmides	ADN double brin circulaire	<i>E. coli</i>	0,1 à 10 kb	Banque d'ADNc
Bactériophages	Virus, ADN double brin linéaire	<i>E. coli</i>	10 à 20 kb	Banque d'ADN génomique et ADNc
Cosmides	ADN double brin circulaire	<i>E. coli</i>	45 kb	Banque d'ADN génomique
Phagemides	Hybride plasmide/phage	<i>E. coli</i>	4 à 10 kb	Banque d'ADNc
PACs	P1, chromosome artificiel	<i>E. coli</i>	130 à 150 kb	Banque d'ADN génomique/ ADNc
YACs	Levure, chromosome artificiel	Levure	200 à 1000kb	Banque d'ADN génomique
BACs	Bactérie, chromosome artificiel	<i>E. coli</i>	100 à 150 kb	Banque d'ADN génomique/ADNc

3. Vecteurs eucaryotes (cellules animales)

3.1 Plasmides non répliatifs

Ce sont des vecteurs qui sont délétés ou mutés dans des gènes essentiels afin de réduire l'expression des protéines virales. Ils transduisent efficacement les cellules. Ils sont également utilisés pour des stratégies de vaccination. Leur intérêt majeur réside dans l'absence de toute

séquence virale codante, ce qui leur confère par ailleurs une capacité de clonage accrue : 150 kb.

3.2 Plasmides réplicatifs

Les plasmides réplicatifs (*YRps*) sont capables de se multiplier en tant que plasmides indépendants car ils portent une séquence d'ADN chromosomique qui comprend une origine de réplication. Les origines de réplication sont connues pour être situées très près de plusieurs gènes de levure, dont un ou deux qui peuvent être utilisés comme marqueurs sélectionnables. Le plasmide *YRp7* est un exemple de plasmide réplicatif. Il est composé de *pBR322* et du gène de levure *TRP1*. Ce gène, impliqué dans la biosynthèse du tryptophane, est situé à côté d'une origine chromosomique de réplication. Le fragment d'ADN de levure présent dans *YRp7* contient à la fois *TRP1* et l'origine (Brown, 2010).

3.3 Virus

La méthode des vecteurs viraux consiste tout simplement à utiliser les virus. Cette méthode a été découverte par Paul Berg en 1970. En effet, il l'a utilisée pour introduire des fragments d'ADN dans des cellules mammifères. Aujourd'hui, la méthode a été reprise pour la thérapie génique. Les virus sont des particules microscopiques qui sont composés soit d'ADN soit d'ARN et sont englobés dans une capsid de protéines (structure entourant l'ensemble du génome). Les virus offrent de nombreux avantages pour le clonage et les applications des gènes clonés qui en découlent. Comme l'efficacité d'infection des virus est élevée, le gène cloné peut être introduit dans une cellule à une fréquence élevée. Ils servent à cloner des fragments de tailles définies. Les virus, pour pouvoir se reproduire dans l'organisme, détournent l'activité de la cellule dans laquelle ils se logent : c'est la transduction. Les cellules infectées par le virus sont alors transduites. Différents types de virus existent pour cela :

- Adénovirus

Les adénovirus sont des virus à ADN possédant un génome linéaire bicaténaire d'environ 36 kb. Ils ont été largement utilisés comme vecteurs de transfert et d'expression des gènes, en raison de leurs nombreuses caractéristiques favorables, comme leur stabilité, leur grande taille d'insert accepté, un spectre d'hôte large comprenant des cellules qui ne se divisent pas et de leur minimum d'effets cytotoxiques (Primrose et *al.*, 2004).

- Rétrovirus

Les rétrovirus sont des virus à ARN, ce sont les vecteurs les plus couramment utilisés pour la thérapie génique. Bien qu'ils s'insèrent à des positions aléatoires, les intégrants résultants sont très stables, ce qui signifie que les effets thérapeutiques du gène cloné persisteront pendant un certain temps (Brown, 2010).

3.4 Chromosomes artificiels

De petits chromosomes artificiels de mammifères (MAC ou *Mammalien Artificial Chromosomes*) et d'humains (HAC ou *Human Artificial Chromosomes*) capables de se répliquer, ont été construits [4-91]. Ils sont transmis de cellule mère à cellule fille et permettent l'expression des gènes qu'ils contiennent (Kochko, 2000). Depuis leur première description à la fin des années 1990, les chromosomes artificiels humains (HAC) porteurs d'un kinétochore fonctionnel ont été considérés comme un système prometteur pour la délivrance et l'expression de gènes avec le potentiel de surmonter plusieurs problèmes causés par l'utilisation de systèmes de transfert de gènes à base virale. Les HACs ont la capacité à transporter des *loci* génomiques entiers avec tous les éléments de régulation leur permet d'imiter fidèlement le modèle normal d'expression des gènes naturels. De plus, non seulement des gènes uniques, mais également des groupes de gènes codant pour des voies complexes peuvent être transportés sur un seul HAC (Kouprina et al., 2018).

La principale utilité que l'on prête à ce nouveau type de chromosome artificiel HAC est la thérapie génique. Cette technique consistera à placer sur ces chromosomes la forme active d'un ou de plusieurs gènes déficients chez un malade. Après introduction dans des cellules en culture de ce même malade, ces cellules (transformées) seront réimplantées dans l'organisme du malade. Ceci permettra de ne plus utiliser des vecteurs viraux (Kochko, 2000).

Les Chromosomes Artificiels de Mammifères (MAC) peuvent être construits à partir d'ADN centromère humain cloné et de répétitions de télomères, prouvant le principe selon lequel les chromosomes peuvent se former à partir de molécules d'ADN nues transfectées dans des cellules humaines. Les MAC étaient stables sur le plan mitotique, avec un faible nombre de copies et des anticorps liés associés à des centromères actifs (Schindelbauer, 1999).

4. Vecteurs eucaryotes (cellules végétales) : plasmide Ti et ADNT

Les seuls vecteurs utilisés de façon routinière pour produire des plantes transgéniques sont dérivés de la bactérie du sol *Agrobacterium tumefaciens*. Cette bactérie provoque une maladie appelée *gale du collet* (ou tumeur de *crown-gall*) au cours de laquelle la plante infectée produit des excroissances incontrôlées (tumeurs ou gales), généralement à la base (collet) de la plante (Griffiths et al., 2001). La gale du collet survient lorsqu'une blessure sur la tige permet aux bactéries *A. tumefaciens* d'envahir la plante. Après l'infection, les bactéries provoquent une prolifération cancéreuse du tissu de la tige dans la région de la couronne (Figure 49).

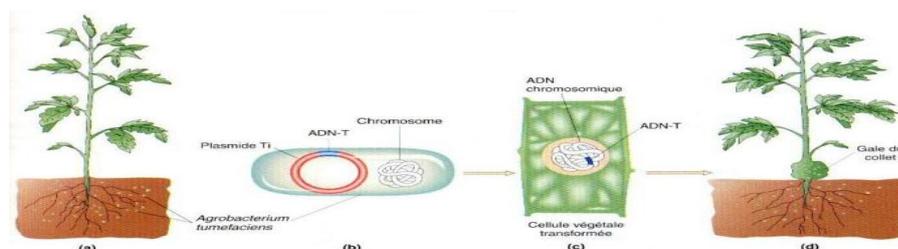


Figure 49 : Gale du collet (Griffith et al., 2001).

Le responsable de la formation de tumeurs est un grand ADN plasmidique circulaire supérieur à 200 kb, le plasmide *Ti* (inducteur de tumeurs : *tumor-inducing*) (Figure 50). Lorsque la bactérie infecte une cellule végétale, une partie du plasmide *Ti*, une région appelée *ADN-T* est transférée et insérée, plus au moins au hasard dans le génome de la plante hôte. Les fonctions nécessaires à ce transfert se trouvent dans le plasmide *Ti*, à l'extérieur de l'ADN-T.

L'ADN-T a une taille comprise entre 15 et 30 kb, selon la souche. Il est maintenu sous une forme stable dans la cellule végétale et est transmis aux cellules filles en tant que partie intégrante des chromosomes. Mais la caractéristique la plus remarquable du plasmide *Ti* est que l'ADN-T contient environ huit gènes qui sont exprimés dans la cellule végétale et sont responsables des propriétés cancéreuses des cellules transformées. Ces gènes dirigent également la synthèse de composés inhabituels, appelés opines, que les bactéries utilisent comme nutriments. *A. tumefaciens* conçoit génétiquement la cellule végétale à ses propres fins (Brown, 2010). Lorsque le plasmide *Ti* a intégré les éléments souhaités du gène, ce vecteur intermédiaire peut à son tour être inséré dans le plasmide *Ti*, formant un plasmide intégratif qui peut être introduit par transformation dans une cellule végétale. Le plasmide *Ti* qui recevra le vecteur intermédiaire est d'abord atténué. C'est-à-dire qu'il est délété de toute région située à droite de son ADN-T, qui comprend les gènes responsables de la formation des tumeurs. La seule région située à gauche qui lui est laissée (*L* pour *Left* : gauche) de son

ADN-T, qui servira de site de *crossing-over* pour l'incorporation du vecteur intermédiaire (Griffiths et *al.*, 2001).

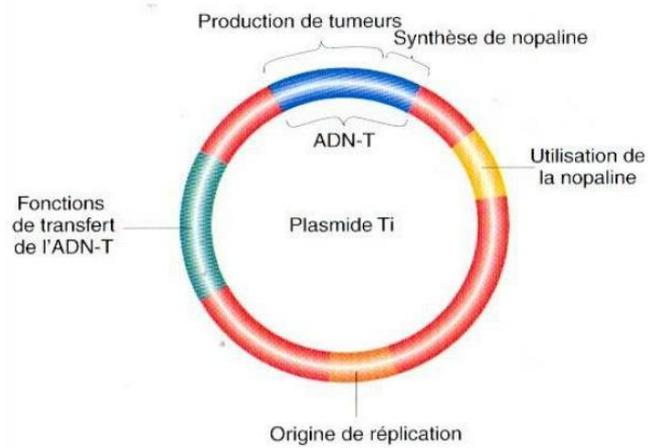


Figure 50 : Représentation simplifiée des principales régions du plasmide *Ti* d'*Agrobacterium tumefaciens* (Griffiths et *al.*, 2001).

Chapitre III.

Procédés de ligation du vecteur et de l'ADN à cloner

Chapitre III. Procédés de ligation du vecteur et de l'ADN à cloner

1. Enzymes ligases

La ligation entre molécules d'ADN est le résultat de l'action catalytique particulières appelées ligases, action qui consiste à réunir par une ligation covalente l'extrémité 3'-hydroxyle d'un brin d'ADN avec celle 5'-phosphate d'un autre brin ; la ligation se réalise entre les deux extrémités d'un même brin sur une matrice, le brin complémentaire des deux extrémités à relier. Une telle modalité de ligation intervient in vivo lors du processus de réparation de l'ADN ou au niveau des discontinuités lors de la constitution de la fourche de réplication. Deux types d'ADN ligases ont été isolés et caractérisés, et sont utilisés de façon courante en génie génétique (Lucotte, 1983) :

- L'ADN ligase *E. coli* ne peut agir que si les deux ADN sont associés par des extrémités cohésives. Elle utilise comme cofacteur le NAD^+ (Kaplan et Delpech, 2007).

- L'ADN Ligase T4, purifiée à partir d'une souche d'*Escherichia coli* infestée par le phage T4 à partir. Elles peuvent effectuer des ligatures sur des fragments d'ADN avec bouts francs (Primrose et al., 2004).

2. Procédés de ligation

2.1 Digestion par la même enzyme de restriction

Le plasmide et l'ADN étranger (insert) sont digérés par le même type d'enzyme de restriction et intégration de l'insert dans le vecteur en utilisant l'ADN ligase d'*Escherichia coli*. Dans cet exemple *EcoRI* G/AATTC (Figure 51) :

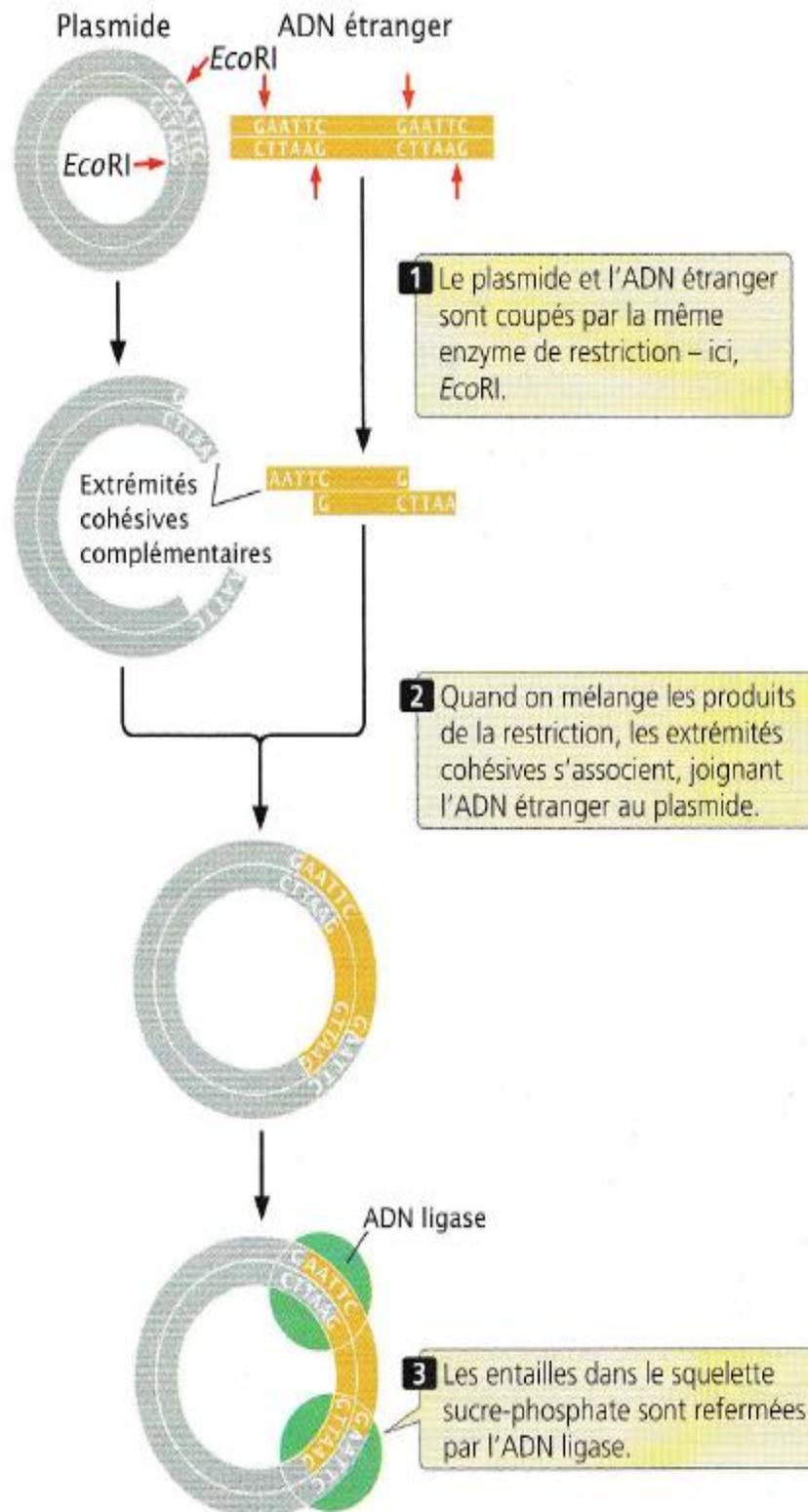


Figure 51 : Un fragment d'ADN étranger inséré dans un plasmide en utilisant des enzymes de restriction et l'ADN ligase (Pierce, 2012).

2.2 Addition d'extrémités cohésives homopolymériques

Une méthode très générale pour joindre des molécules d'ADN consiste à utiliser l'association de séquences homopolymériques complémentaires. Si on ajoute des séquences oligo (dA) aux extrémités 3' d'une population de molécules d'ADN et des séquences oligo (dT) aux extrémités 3' d'une autre population, les deux types de molécules peuvent s'associer pour former des dimères circulaires mixtes.

La désoxyribonucléotidyl transférase terminale, une enzyme purifiée à partir du thymus de veau, permet de synthétiser les extensions homopolymériques (Primrose et *al.*, 2004). (Figure 52).

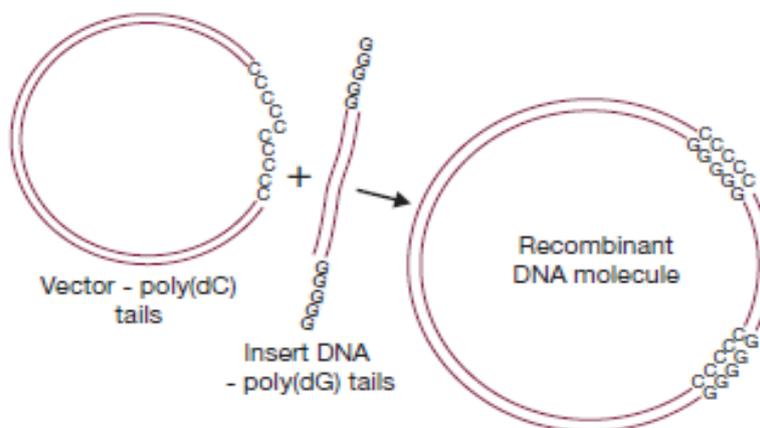


Figure 52 : Insertion d'un insert dans un vecteur par des homopolymères (Brown, 2010).

2.3 Addition de linkers

Des *linkers*, polymères synthétiques contenant des sites de restriction définis pour les enzymes, peuvent être fixés aux deux extrémités doubles brins de l'ADNc. Par exemple un *linker BamHI* (un décanucléotide contenant la séquence CCTAGG) peut être fixé avec la ligase de T4 aux extrémités franches d'un ADNc. Après coupure des molécules recombinées avec *BamHI*. Ces molécules recombinées seront reconstruites lors de la réplication dans la bactérie. Cette stratégie, qui conserve l'intégrité d l'ADNc, est de plus en plus utilisée pour fabriquer des banques ADNc (Crépin, 1987).

Chapitre IV.
Transfert de l'ADN dans les cellules

Chapitre IV. Transfert de l'ADN dans les cellules

1. Transfert direct

1.1 Biolistique

Ce procédé consiste en l'utilisation d'un pistolet modifié pour projeter de petites particules de métal (1 à 4 μm) dans des cellules de plantes à une vitesse suffisante pour traverser la paroi cellulaire. Un appareil utilisant une décharge électrique s'est révélé particulièrement utile au développement de méthodes de transfert de gènes non spécifiques d'une variété particulière dans le cas des céréales et des légumineuses les plus récalcitrantes (Primrose et *al.*, 2004).

1.2 Micro-injection

La micro-injection de gènes dans le noyau des cellules est la méthode la plus efficace. Elle peut conduire à l'obtention de 20% de cellules dans lesquelles le gène étranger est intégré. Elle nécessite un appareillage particulier et n'est justifiée que pour obtenir des clones stables de cellules pour lesquelles la transfection est difficile (Houdebine, 2003).

2. Transformation/transfection

2.1 Méthodes chimiques : au chlorure de calcium (bactéries)

La transformation d'*E. coli* est une étape essentielle de nombreuses expériences de clonage. Les cellules d'*E. coli* et l'ADN plasmidique interagissent de façon productive en présence d'ions calcium et à basse température. Le protocole consiste à resuspendre des cellules en croissance exponentielle dans une solution glacée de chlorure de calcium (CaCl_2) 50 mM, à une concentration finale de 10^{10} cellules/ml, puis de les laisser à 0°C pendant 30 mn. L'ADN plasmidique (0,1 μg) est ensuite ajouté à un petit échantillon (0,2 ml) de ces cellules qui sont désormais compétentes pour la transformation et l'incubation dans la glace est poursuivie pendant 30 mn, puis suivie d'un choc thermique de 2 mn à 42°C . Les cellules sont transférées dans un milieu nutritif et incubées à température ambiante (de 30 mn à 1h) (Figure 53) pour permettre l'expression phénotypique des propriétés conférées par le plasmide, par exemple la résistance à antibiotique qui est souvent utilisé comme marqueur de sélection des cellules contenant un plasmide (Primrose et *al.*, 2004).

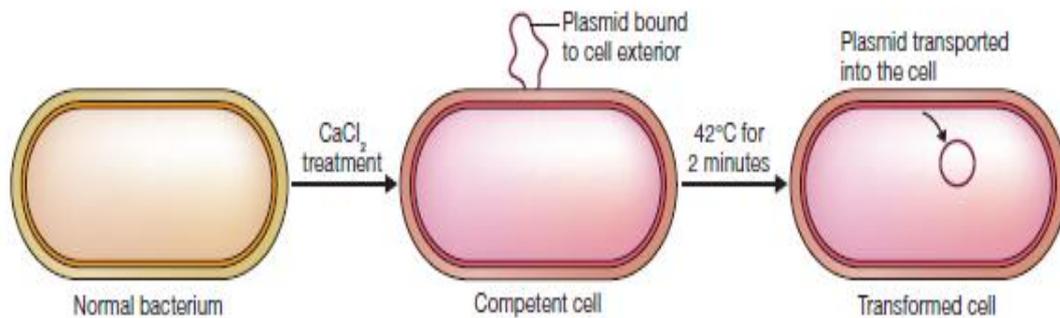


Figure 53 : Absorption d'ADN plasmidique par une cellule bactérienne compétente (Brown, 2010).

2.2 Co-précipitation de l'ADN et du phosphate de calcium

Le transfert réussi d'ADN dépendait de la formation d'un fin co-précipité ADN/phosphate de calcium, qui se dépose d'abord sur les cellules avant d'être internalisé. Ce précipité doit être préparé juste avant la transfection. Les petits grains de phosphate de calcium associés à l'ADN entrent par endocytose puis sont transportés vers le noyau, où certains ADN sont préservés et peuvent s'exprimer. Cette méthode ne s'applique qu'aux cellules qui poussent en monocouche, à l'exclusion de cellules en suspension ou formant des agrégats (Orrantia et Chang, 1990).

2.3 DEAE-dextran (Di-Ethyl-Amino Ethyl-dextran)

Le DEAE dextrane est un polycation qui facilite l'absorption de l'ADN sur la membrane plasmique et favorise son entrée à l'intérieur des cellules. Cependant le DEAE dextrane peut être toxique pour les cellules. L'optimisation des conditions de transfection est fonction de la sensibilité des cellules hôtes à ce polymère. Néanmoins, l'efficacité de transfection est élevée puisqu'on peut obtenir jusqu'à 25% de cellules transfectées (Houdebine, 2003).

2.4 Polycation-DMSO (cellules eucaryotes)

Le DMSO (diméthylsulfoxyde) est un solvant polaire organique, il augmente la perméabilité de la membrane et améliore l'adsorption et l'internalisation à la surface de la cellule.

2.5 Fusion des protoplastes

Les protoplastes sont des cellules dont les parois cellululosiques ont été retirées. Ils sont très utiles pour la manipulation génétique pour deux raisons: premièrement, plusieurs protocoles de transformation ont été développés qui fonctionnent spécifiquement avec des protoplastes; et deuxièmement, parce que dans certaines conditions, des protoplastes de types cellulaires similaires ou contrastés peuvent être fusionnés pour donner des hybrides somatiques, un processus connu sous le nom de fusion de protoplastes (Primrose et Twyman, 2006). Le polyéthylène glycol (PEG) est un polymère qui provoque des fusions entre membranes plasmiques. Il permet le rapprochement physique des molécules d'ADN et du plasmalemma, et des déstabilisations locales et réversibles de la membrane plasmique. Il en résulte la pénétration de molécules d'ADN dans le cytoplasme des protoplastes (Tagu et Moussard, 2003).

2.6 Lipofection

L'utilisation de liposomes comme système de transformation ou de transfection est appelée *lipofection*. C'est une technique utilisée pour injecter du matériel génétique dans une cellule au moyen de liposomes, qui sont des vésicules qui peuvent facilement se confondre avec la membrane cellulaire puisqu'elles sont toutes deux constituées d'une bicouche phospholipidique. La lipofection utilise généralement un lipide chargé positivement (cationique) pour former un agrégat avec le matériel génétique chargé négativement (anionique). Les complexes moléculaires obtenus, appelés *lipoplexes*, sont ensuite repris par les cellules. Les principaux avantages de la lipofection sont : son efficacité élevée, sa capacité à transférer tous les types d'acides nucléiques dans un large éventail de types de cellules, sa facilité d'utilisation, sa reproductibilité et sa faible toxicité (Primrose et Twyman, 2006).

2.7 Peptides

Les peptides ont été principalement utilisés comme agents auxiliaires pour d'autres systèmes de vectorisation, en particulier des polymères cationiques tels que la polylysine, auxquels ils peuvent apporter une fonction de ciblage cellulaire, une activité de type membranolytique ou encore un signal de localisation nucléaire. L'entrée se fait par endocytose, le composé doit pouvoir déstabiliser la membrane endosomale avant que l'ADN ne soit dégradé dans le lysosome. Une fois relargué dans le cytosol, l'ADN peut pénétrer dans le noyau au cours de la division cellulaire.

2.8 Electroporation

L'électroporation est une méthode simple pour introduire des gènes clonés dans des cellules très diverses. La technique consiste à utiliser des décharges électriques de haute tension qui induisent les membranes cytoplasmiques à fusionner. Les cellules soumises à un choc électrique captent de l'ADN exogène présent dans la suspension. Une partie de ces cellules sont transformées de façon stable et peuvent être sélectionnées si l'ADN transformant porte un marqueur approprié (Primrose et *al.*, 2004).

2.9 Transduction virale (encapsidation *in vitro*)

La transduction est le troisième mode de transfert génétique chez les bactéries. La transduction est un processus qui consiste en un transfert de matériel génétique (ADN bactérien), d'une bactérie donneuse à une bactérie receveuse, par l'intermédiaire d'un vecteur viral (un bactériophage). Le bactériophage infecte une première bactérie (bactérie donneuse) et y injecte son ADN viral à travers la paroi de la cellule. De nouveaux phages s'y développent, et certains intègrent une partie du génome bactérien dans leur capsid de phage. Lors de la libération des phages, ceux-ci vont infecter d'autres bactéries. Les virus comportant une partie d'ADN bactérien vont l'injecter dans une nouvelle bactérie (bactérie receveuse).

Chapitre V.
Sélection des transformants recombinants

Chapitre V. Sélection des transformants recombinants

Le rendement de la transformation bactérienne étant très faible (1 bactérie sur 10^5 à 10^7 par microgramme de plasmide recombinant), il convient de sélectionner les quelques bactéries recombinantes, parmi les milliards de bactéries qui n'ont rien incorporé. Cette sélection est obtenue par culture sur un milieu sélectif, la ou les propriétés apportées par le plasmide étant utilisées comme moyen de sélection. Il s'agit presque toujours d'une résistance à un antibiotique. Les bactéries ayant incorporé le plasmide recombinant deviennent, grâce au gène de résistance du plasmide, résistantes à l'antibiotique ; toutes les autres sont tuées. Cette technique ne permet pas de distinguer celles qui ont incorporé un plasmide non recombinant. Pour y parvenir, il faut utiliser, dans un second temps, un second marqueur de sélection (Kaplan et Delpech, 2007) (Figure 54).

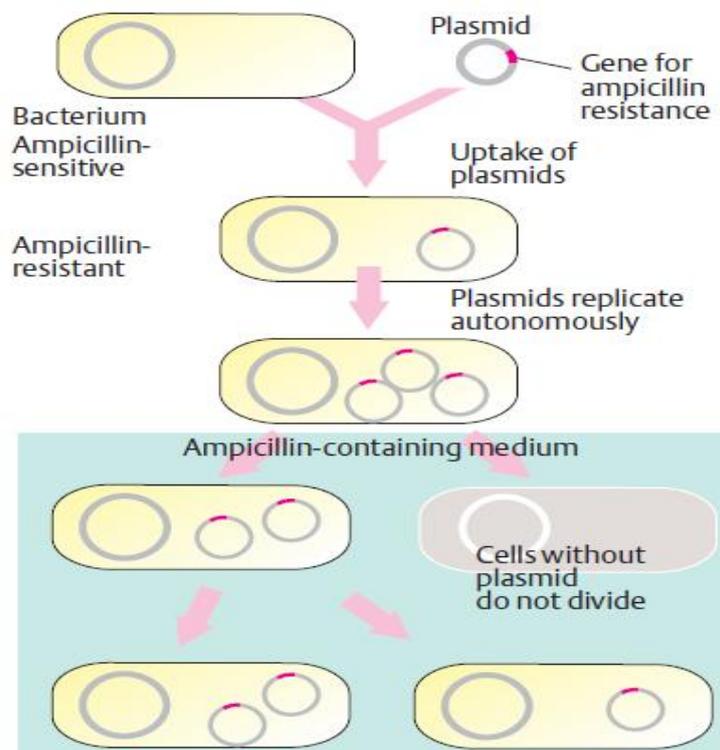


Figure 54 : Transformation bactérienne par un plasmide (Passarge, 2007).

1. Sélection par complémentation

Un système enzymatique appartenant à l'opéron lactose peut être utilisé à la place d'un second antibiotique. L'insertion du fragment d'ADN dans le plasmide *pUC* aboutit à l'inactivation du gène qui code pour la β -galactosidase. Pour vérifier la présence ou l'absence

de l'activité enzymatique β -galactosidase, on utilise un galactoside dont la couleur passe de l'incolore au bleu quand il est clivé par la β -galactosidase, ce composé chimique s'appelle *X-gal*. Pour pouvoir métaboliser le *X-gal*, la cellule doit être exposée à un inducteur, cet inducteur est l'IPTG (*isopropylthio-b-D-galactoside*).

Le $LacZ^+$ synthétise la β -galactosidase qui va dégrader le *X-gal*, les colonies auront une couleur bleue, ce qui indique que les bactéries ont reçu le vecteur non recombinant.

$LacZ^-$ pas de synthèse de β -galactosidase, pas de dégradation de *X-gal*, les bactéries auront une couleur blanche, ce qui indique que les bactéries ont reçu le vecteur recombinant.

En culture et en présence d'IPTG et de *X-gal*, les bactéries résistantes à l'ampicilline et transformées par les plasmides recombinants se présentent sous forme de colonies blanchâtres car elles ont perdu la capacité de clivage de l'équivalent coloré du lactose (*X-gal*) par la β -galactosidase. Par contre, les bactéries résistantes à l'ampicilline et non transformées par les plasmides recombinants se présentent sous forme de colonies bleues. La sélection visuelle des bactéries transformées par les plasmides recombinants est donc possible (Figure 55).

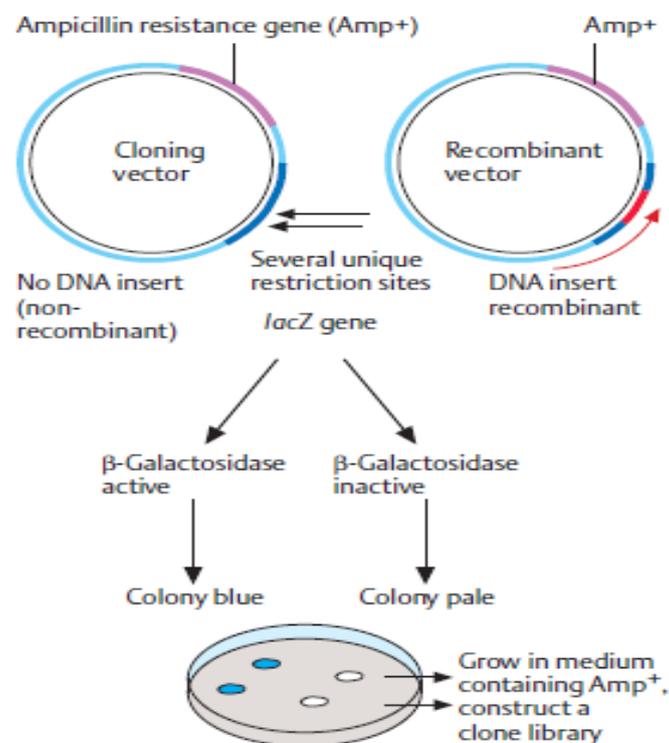


Figure 55 : Sélection des clones recombinants par complémentation (Passarge, 2007).

2. Sélection par marqueurs dominants de résistance à des agents toxiques

La détection de ces gènes marqueurs à sélection positive est subordonnée à l'utilisation d'agents toxiques herbicides ou médicaments. Ce sont les premiers gènes marqueurs à avoir été utilisés par les biotechnologies.

Cette résistance est conférée aux plantes par des gènes codant une forme tronquée de protéines endotoxines, fabriquées par certaines souches de *Bacillus thuringiensis* (bactéries vivant dans le sol ; d'où le nom de maïs *Bt*). Il existe de multiples toxines, actives sur différents types de larves d'insectes issues de cette souche bactérienne, reconnues depuis de longues années pour leurs propriétés insecticides utilisés en agriculture biologique ou conventionnelle. Plusieurs espèces ont été transformées avec un gène de résistance à un ravageur. Certaines sont commercialisées comme: le maïs, le coton, la tomate, la pomme de terre et d'autres sont à l'essai notamment le riz et l'aubergine.

3. Sélection par des marqueurs métaboliques

Dans cette approche, il est nécessaire d'utiliser une souche de levure mutée au niveau d'un gène codant pour une enzyme impliquée dans la synthèse d'un métabolite essentiel. Cette levure est maintenue en vie en lui fournissant ce métabolite dans le milieu de culture. Si elle est placée sur un milieu dépourvu de cette molécule, elle meurt. Mais elle survivra sur ce même milieu si elle a reçu par transformation une copie intacte de ce gène. Ainsi le gène *URA3* est fréquemment utilisé. Il code pour une enzyme catalysant la transformation d'orotidine-5'-phosphate en uridine monophosphate (*UMP*). Cette réaction fait partie d'une voie métabolique produisant uridine triphosphate (*UTP*) et cytidine triphosphate (*CTP*). Les mutants *ura3-* ne peuvent survivre qu'en présence d'uracile dans le milieu de culture. Si ces mutants acquièrent un fragment d'ADN comportant le gène *URA3* non muté, elles deviennent capables de croître sur un milieu dépourvu d'uracile.

Chapitre VI.
Criblage pour la détection des clones d'intérêt

Chapitre VI. Criblage pour la détection des clones d'intérêt

Le criblage d'une banque génomique ou d'ADNc consiste à sélectionner spécifiquement des clones bactériens contenant la séquence du gène recherché. Afin d'identifier les clones d'intérêt, une étape préalable de réplique de la banque est indispensable (Damier et Jacquot, 2003).

1. Complémentation génétique

Des clones spécifiques peuvent être détectés dans une banque bactérienne plasmidique ou phagique grâce à leur capacité à fournir une fonction absente à une lignée mutante de l'organisme donneur, qui joue le rôle de transformation. Cette procédure s'appelle la *complémentation fonctionnelle* ou le sauvetage de mutants. Le protocole est le suivant :

- Construire une banque bactérienne ou phagique contenant des inserts d'ADN donneur recombinant de type sauvage a^+ ;
- Utiliser la banque pour transformer des cellules de la lignée donneuse mutante a^- ;
- Sélectionner le phénotype a^+ parmi les cellules de l'organisme donneur ;
- Récupérer le gène a^+ à partir du clone phagique ou à partir des cellules donneuses transformées.

La raison de la réussite de cette technique de transformation tient dans le fait que le fragment transformant assure la complémentation fonctionnelle de la déficience due à l'allèle mutant chez le receveur.

2. Recombinaison

Le remplacement de gènes et leur intégration ciblée reposent sur un processus de recombinaison homologue. Une construction de gène peut être ciblée dans un site préalablement choisi du génome dans lequel a été introduite une séquence *LoxP*. La recombinase *Cre* permet au vecteur possédant lui-même une séquence *LoxP* de s'introduire dans le site *LoxP* du génome. Le ciblage par le système *Cre LoxP* permet d'introduire de multiples versions d'un gène dans le même site. Cela limite les effets de position sur le transgène à ceux intrinsèques au site choisi. Un site *LoxP* contenant 35 paires de bases est introduit dans le génome par recombinaison homologue, ou au hasard par microinjection. Le site retenu peut recevoir une construction de gène portant elle aussi un site *LoxP*.

La recombinaise *Cre* est apportée dans l'embryon en coinjectant un plasmide circulaire exprimant le gène *Cre* et la construction à intégrer. Le site *LoxP* peut avoir été bordé par un isolateur pour réduire les effets de position des gènes intégrés dans ce site (Figure 56).

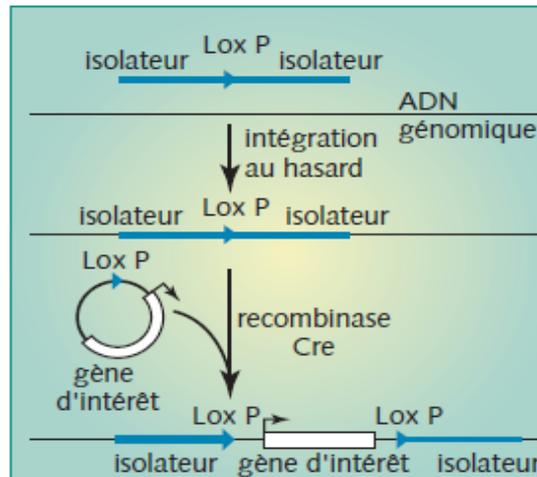


Figure 56 : Principe de l'intégration ciblée d'un gène en un site donné du génome (Houdebine, 2002).

3. Hybridation des acides nucléiques

3.1 Méthode de préparation d'une sonde radioactive

3.1.1 Marquage de l'ADN par *Nick-translation* (déplacement de cassure)

La technique de *nick translation* (déplacement de cassure) repose sur la réparation de cassures (*nicks*) simple brin dans l'ADN, formant des extrémités 3' OH et 5' P terminales exposées. Cette cassure peut être obtenue en utilisant une endonucléase appropriée, comme la désoxyribonucléase I du pancréas (*DNase I*). La cassure ainsi exposée va alors servir de point de départ pour l'introduction de nouveaux nucléotides en utilisant l'ADN polymérase d'*E. coli*, une enzyme composée de nombreuses sous-unités qui possède des activités d'ADN polymérase et d'exonucléase 5' → 3'. Alors que l'ADN polymérase ajoute de nouveaux nucléotides à l'extrémité 3'-hydroxyle de la cassure, les nucléotides sont supprimés à l'autre extrémité de la cassure par l'activité exonucléasique 5' → 3' de la même enzyme. La cassure est ainsi progressivement déplacée le long de l'ADN dans la direction 5' → 3'. Si la réaction est réalisée à une température relativement faible (environ 15 °C), elle ne se poursuit par au-delà du renouvellement complet de la séquence nucléotidique existante. Bien qu'il n'y ait pas de synthèse nette d'ADN à cette température, la réaction de synthèse permet l'incorporation de nucléotides marqués à la place de ceux qui existaient précédemment dans le nucléotide non marqué (Figure 57) (Strachan et Read, 2012).

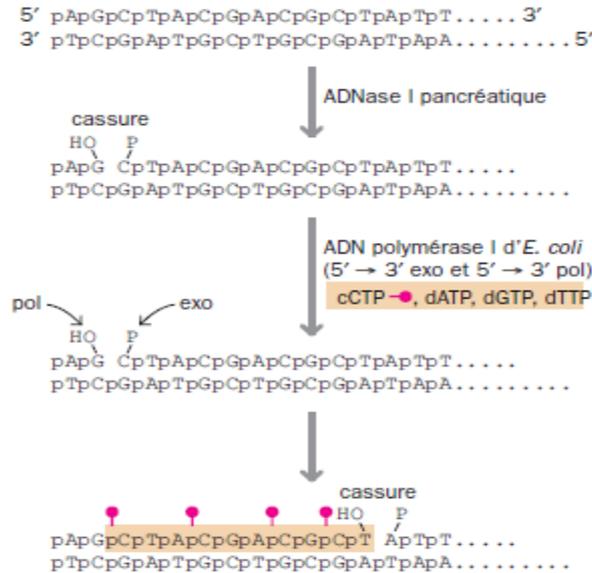


Figure 57 : Marquage de l'ADN par *nick-translation* (Strachan et Read, 2012).

3.1.2 Marquage par multi-amorçage au hasard (*Random priming*)

La technique de marquage de l'ADN à l'aide d'amorces aléatoires repose sur l'hybridation d'un mélange de nombreux hexanucléotides différentssynthésisés de façon presque aléatoire. qui se lient au hasard à des séquences complémentaires de l'ADN matrice dénaturé, et initient la synthèse de nouveaux brins d'ADN (Figure 58). La synthèse des nouveaux brins complémentaires est catalysée par la sous- unité de *Klenow* de l'ADN polymérase I d'*E. coli* (qui contient l'activité de polymérase 5' → 3' sans l'activité exonucléasique 5' → 3' associée) (Strachan et Read, 2012).

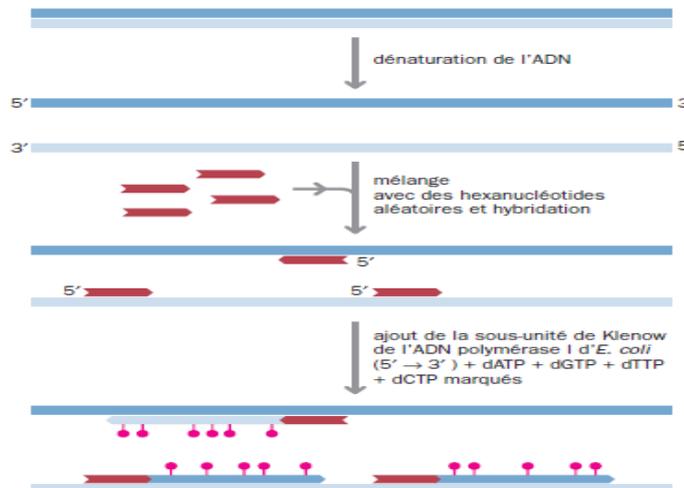


Figure 58 : Marquage de l'ADN par amorçage au hasard (Strachan et Read, 2012).

3.2 Hybridation moléculaire

Le criblage par hybridation de l'ADN des clones transformés avec une sonde spécifique constitue une méthode de choix. L'hybridation peut se faire *in situ* : on effectue des répliques de clones bactériens cultivés en boîte de Pétri sur des disques de nitrocellulose, après un temps de culture suffisant, les bactéries de ces répliques sont ensuite lysées par la soude qui, en même temps, dénature l'ADN, les molécules simple brin correspondantes se trouvent immobilisées à l'emplacement de chaque clone. Après hybridation, avec la sonde radioactive, l'autoradiographie révélera les clones positifs (Figure 59).

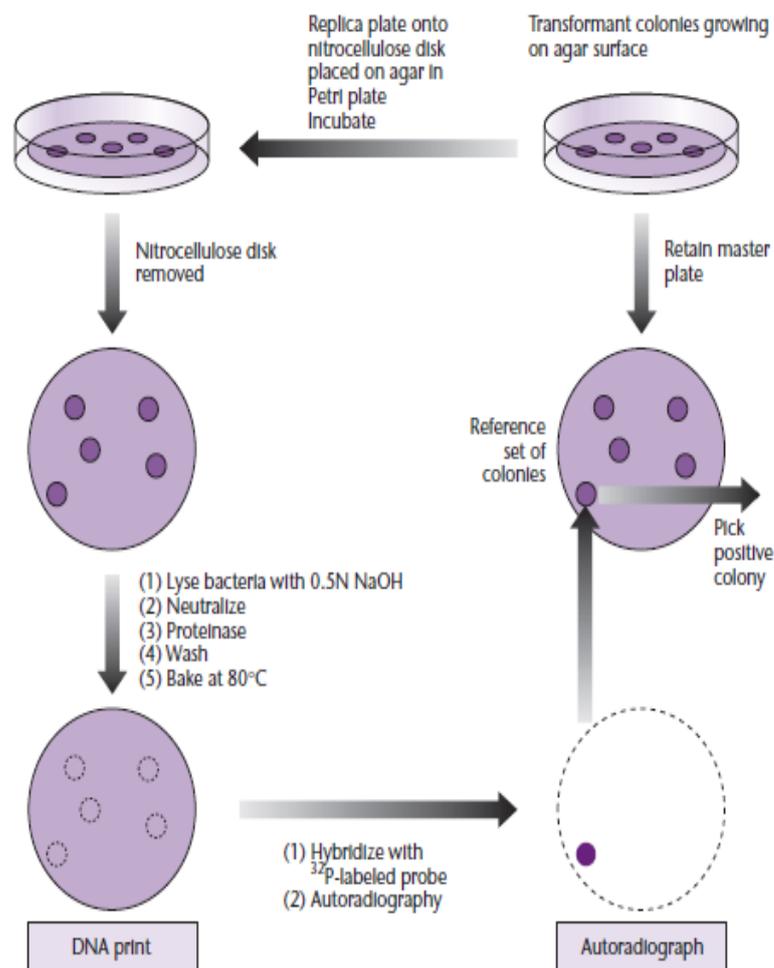


Figure 59 : Méthode de détection des clones recombinants par hybridation moléculaire des colonies (Primrose et Twyman, 2006).

4. Détection des produits d'expression

Il existe plusieurs versions de criblage immunologique, mais la méthode la plus utile est une contrepartie directe de l'hybridation de colonies par des sondes. Les colonies recombinantes sont transférées sur une membrane en polyvinyle ou en nitrocellulose, les cellules sont lysées et une solution contenant l'anticorps spécifique est ajoutée. Dans les méthodes originales, soit le l'anticorps lui-même a été marqué, ou la membrane a ensuite été lavée avec une solution de *protéine A* marquée, une protéine bactérienne qui se lie spécifiquement aux immunoglobulines dont les anticorps sont constitués (Figure 60). Dans les méthodes plus modernes, l'anticorps lié - l'anticorps primaire - est détecté en lavant la membrane avec un anticorps secondaire marqué, qui se lie spécifiquement à l'anticorps primaire. Plusieurs molécules d'anticorps secondaires peuvent se lier à une seule molécule d'anticorps primaire, augmentant la quantité de signal produit et permettant une détection plus claire de chaque colonie positive. Dans les trois méthodes, le marqueur peut être radioactif, auquel cas les colonies qui se lient au marqueur sont détectées par autoradiographie ou des marqueurs non radioactifs entraînant un signal fluorescent ou chimioluminescent peuvent être utilisés.

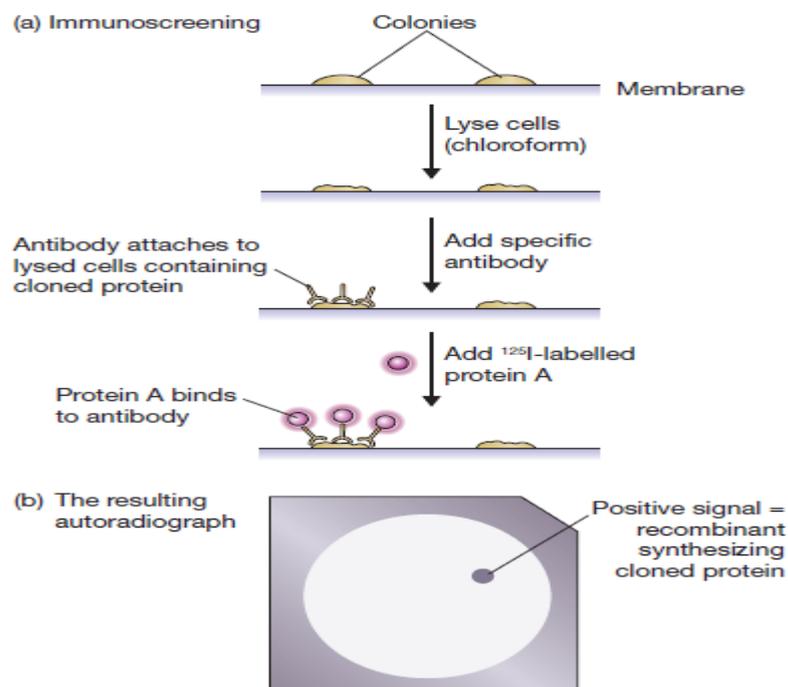


Figure 60 : Utilisation d'un anticorps purifié pour détecter des protéines dans des colonies recombinantes (Brown, 2010).

Chapitre VII.
Notions de transgénése animale et végétale

Chapitre VII. Notions de transgénèse animale et végétale

1. Transgénèse animale

Le terme de "transgénique" s'applique à des organismes vivants, qu'ils appartiennent au règne végétal ou animal, dans lesquels ont été transférés des gènes étrangers d'origine animale ou végétale à leur patrimoine héréditaire propre. Le gène étranger, appelé "transgène", se transmet à la descendance selon un mode mendélien (Kernbaum, 1998). La définition la plus récente des animaux ou végétaux transgéniques est suggérée par Beardmore (1997) : ce sont des "*organismes contenant des séquences intégrées d'ADN cloné (transgène), transférées en utilisant des techniques de génie génétique*". Cette définition est la plus générale et elle fait directement référence à la science à la base de cette révolution : le génie génétique.

1.1 Techniques de transgénèse

La transgénèse, au sens large, regroupe plusieurs techniques permettant différentes manipulations au niveau du génome : soit l'introduction d'un nouveau gène, soit l'inactivation temporaire ou définitive d'un gène déjà présent (Gordon, 1997 ; Houdebine, 1998) : La transgénèse permet ainsi l'étude de gènes particuliers ou encore la création de nouveaux modèles. Chez la souris, il existe trois méthodes de transgénèse très largement utilisées (Figure 61).

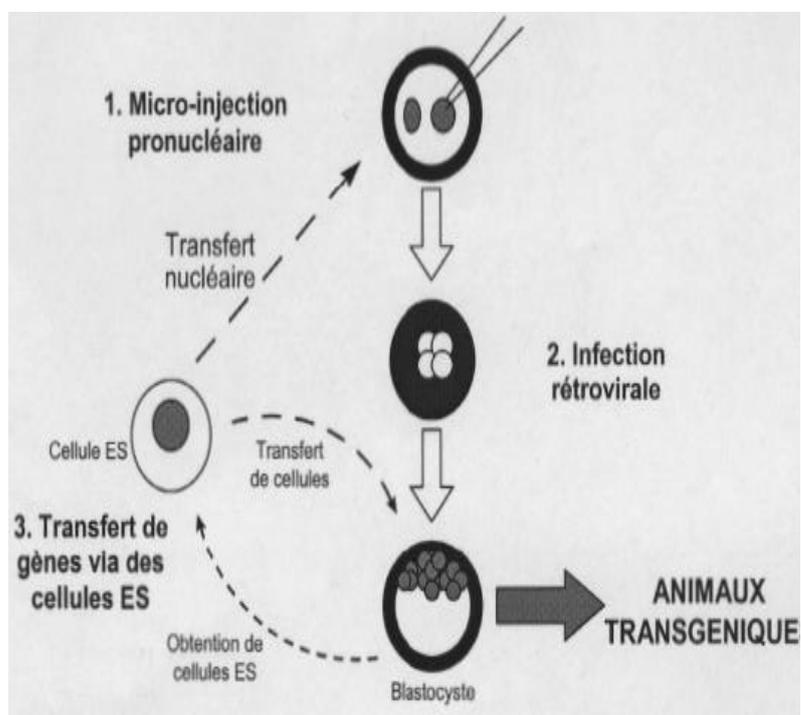


Figure 61 : Trois principales méthodes de transgénèse (Janne et Alhonen, 1996).

1.1.1 Micro-injection

Cette méthode est utilisée par l'équipe de Gordon lors de leur première expérience de transgénèse (Gordon et *al.*, 1980). C'est également la première méthode qui s'est révélée efficace chez les mammifères. Elle consiste à injecter des fragments d'ADN purifié (sous forme d'une solution) dans un des pronucléi (en général le pronucléus mâle car il est à la fois le plus grand et le plus proche de la surface) d'un œuf fertilisé au stade unicellulaire. Le zygote est réimplanté dans l'utérus d'une femelle pseudo-gestante où il poursuit son développement (Figure 62).

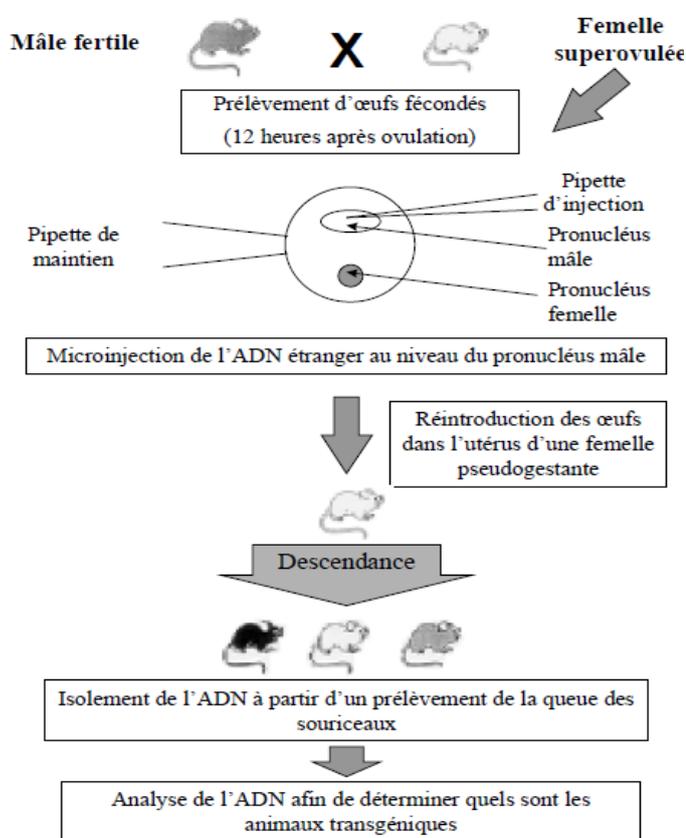


Figure 62 : Principales étapes de l'obtention de souris transgéniques par micro-injection pronucléaire (Jallat, 1991).

Avantages et inconvénients

• Avantages

L'avantage principal de la micro-injection est son efficacité à produire des lignées transgéniques qui expriment la plupart du temps les gènes de manière prévisible (Jaenisch, 1988). Le travail de biologie moléculaire est assez simple (comparé aux autres techniques) : le

gène que l'on souhaite insérer dans le génome doit être identifié puis cloné afin d'avoir un nombre suffisant de copies à injecter (Gordon, 1997). En raison du faible temps de gestation de la souris et du peu de technicité de la phase de préparation de l'ADN, le temps de développement est réduit : entre 3 et 9 mois (Jaenisch, 1988).

- Inconvénients

Cependant cette technique ne permet que l'addition d'un gène (ni la délétion, ni la substitution ne sont possibles). L'intégration se faisant au hasard, il est difficile de prévoir les effets secondaires. Il est donc nécessaire de prendre en compte les effets du locus d'insertion sur le patron d'expression. Même si les gènes micro-injectés s'expriment de façon efficace, il se peut que le niveau d'expression soit sous la dépendance de différents facteurs comme des facteurs procaryotiques provenant des vecteurs utilisés qui peuvent inhiber l'expression du transgène (Gordon, 1997).

La dernière contrainte est la précision nécessaire à la technique de micro-injection, qui requiert de la dextérité de la part du technicien et donc un entraînement. Malgré ces quelques défauts, la technique de micro-injection pronucléaire est la plus efficace des techniques et également la plus utilisée pour obtenir des animaux transgéniques. Des fragments d'une longueur d'environ 50 kb peuvent être ainsi introduits dans le génome d'un œuf et ainsi ils seront exprimés à la fois dans les cellules somatiques et les cellules de la lignée germinale.

1.1.2 Transfert des cellules souches embryonnaires

Les cellules ES (*Embryonic Stem Cells*, cellules souches embryonnaires) sont des cellules embryonnaires précoces totipotentes (c'est-à-dire capables de donner naissance à tous les types de tissus différenciés). Elles sont obtenues en mettant en culture des cellules de la masse cellulaire interne provenant d'embryons au stade préimplantatoire (blastocyste). Il est possible d'injecter ces cellules dans la cavité d'un blastocyste receveur (Gassier et *al.*, 1987). Environ 80 % des embryons survivent à cette manipulation et jusqu'à 90 % des survivants sont chimériques. Les cellules ES présentent au moins deux caractéristiques intéressantes qui les rendent aptes à participer à la constitution d'individus chimériques. Elles possèdent un fort potentiel de différenciation et une constitution chromosomique normale. Avant de procéder à l'injection de cellules ES dans un blastocyste, il est important de vérifier ces deux caractéristiques principales, surtout si les cellules ont fait l'objet de manipulations génétiques ou de repiquages répétés. Les cellules ES cultivées *in vitro* sont transfectées avec l'ADN

exogène. Les clones recombinants homologues sont sélectionnés, puis injectés dans des blastocystes receveurs (Figure 63). Les cellules *ES* proviennent d'une lignée présentant un pelage agouti (gris) et les blastocystes d'une lignée albinos (blanche) ou non-agouti (noire). Les mâles chimériques (qui présentent des taches grises sur un pelage blanc ou noir) sont accouplés avec des femelles de la même lignée que celle qui a fourni les blastocystes. Les individus hétérozygotes ainsi obtenus sont gris car le caractère agouti est dominant, alors que les caractères albinos et non-agouti sont récessifs. Les hétérozygotes sont croisés entre eux afin d'observer les effets du gène modifié à l'état homozygote (Lemarchandel et Montagutelli, 1990).

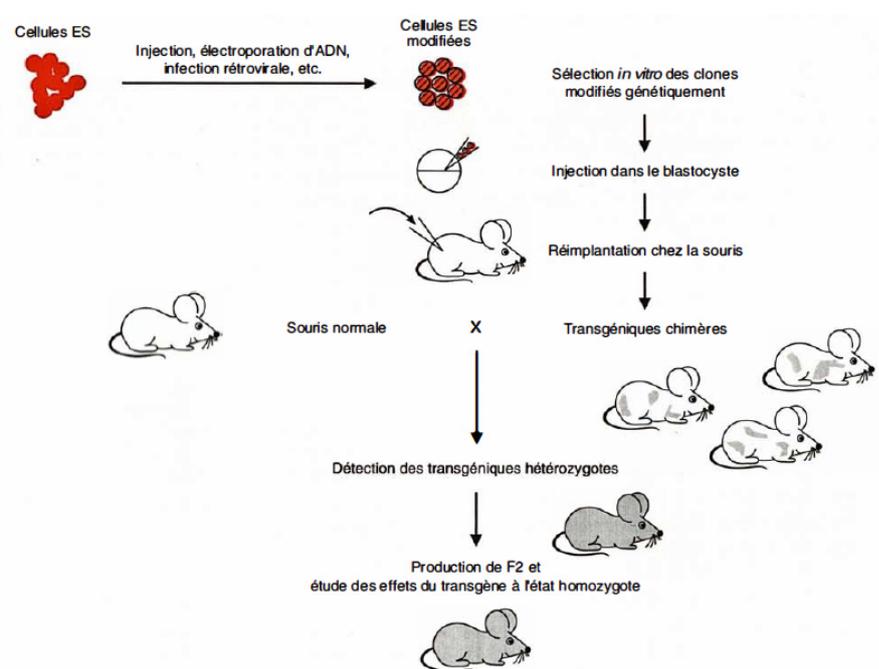


Figure 63 : Transgénèse utilisant les cellules *ES* (Lemarchandel et Montagutelli, 1990).

• Avantages

- Méthode permettant une transgénèse ciblée
- Pas de limitation de taille pour le transgène à intégrer.

• Inconvénients

- La période de temps entre la mise en culture de cellules souches, leur modification génétique, la sélection des clones à injecter, l'obtention de chimères fertiles et leur reproduction peut s'avérer très longue...

1.1.3 Infection rétrovirale

Cette technique repose sur une propriété particulière des rétrovirus : leur capacité à s'intégrer dans le génome des cellules qu'ils infectent. La découverte de Jaenish en 1976 a été l'étape déclenchant la possibilité de l'utilisation de cette technique pour obtenir des animaux transgéniques : il a été démontré que les rétrovirus pouvaient infecter des embryons à un stade précoce et que, suite à cette infection, il y avait insertion d'ADN sous forme de provirus dans le génome de ces embryons et que ce provirus était ensuite transmis à la descendance des souris l'ayant intégré (Jaenisch, 1976).

Des vecteurs viraux dans lesquels le transgène a été introduit par des techniques de recombinaison génétique, sont mis en présence de jeunes embryons de souris. Ceux-ci sont ensuite réimplantés dans l'utérus de souris femelles pseudo-gestantes.

1.1.4 Transfert de noyau

Cette procédure implique de retirer le noyau d'une cellule somatique et de l'insérer dans une cellule énucléée provenant d'un ovule non fécondé. Le noyau transplanté commence alors à se diviser et à se multiplier, comme dans une cellule normale, tout en gardant son identité génétique unique.

Au début de l'année 1997, une équipe écossaise a annoncé la naissance de Dolly au cours de l'année précédente, une agnelle clonée à partir d'une brebis adulte. La naissance de Dolly a été réalisée par un chercheur vétérinaire, le Dr Ian Wilmut et ses collègues du Roslin Institute. Pour créer Dolly, le groupe de Wilmut a utilisé le noyau d'une cellule mammaire "quiescente" provenant d'une brebis blanche de race Finn Dorset, c'est-à-dire une cellule qui avait arrêté de se diviser lorsqu'elle avait été privée au préalable de substances nutritives. Ensuite, le noyau a été implanté à travers la zone pellucide protectrice dans un ovocyte (un ovule non fécondé) énucléé prélevé sur une brebis de race Scottish *Blackface* (Wilmut et al., 1997). Une minuscule charge électrique l'a aidé à fusionner avec le cytoplasme de l'ovocyte. Après de nombreuses tentatives manquées, les chercheurs sont parvenus à obtenir une cellule d'ovule qui a commencé à se diviser normalement, et celle-ci a été implantée dans la mère porteuse Scottish *Blackface*. Après une période de gestation normale d'environ cinq mois, Dolly est née (Brown, 2000) (Figure 64).

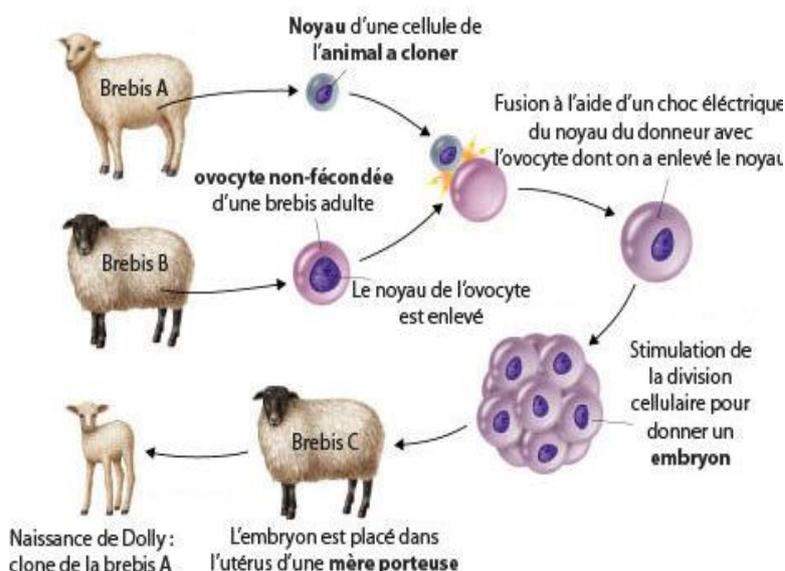


Figure 64 : Clonage reproductif du mouton "Dolly".

(https://www.pdfprof.com/PDF_Image.php?id=132792&t=24).

- Avantages de la transgénèse animale

- *Production de protéines recombinantes* : Des moutons et des brebis transgéniques ont été produits et qui secrétaient dans leur lait des protéines recombinantes humaines telles que la protéine de coagulation du sang (le facteur VIII) et une protéine plasmatique, l'antitrypsine (Winter et al., 1999).

- *Addition de composants nouveaux dans le lait destinés à la consommation humaine et animale* : Une opération plus volontaire peut consister à ajouter dans le lait des protéines qui ne s'y trouvent pas normalement, ces protéines pouvant ou non être naturellement présentes dans le lait selon les espèces animales. C'est le cas de la lactoferrine humaine qui peut être sécrétée dans le lait de vache grâce au transfert du gène correspondant (Houdebine, 1998).

- *Xénogreffe* : Le rat, le lapin et le porc sont des modèles d'étude pour définir les conditions permettant des xénogreffes d'organes d'animaux à des patients. Il est admis que le porc est le meilleur donneur potentiel d'organes. Des gènes ajoutés ou inhibés permettent aux organes de porcs, de ne plus être rejetés par le système du complément. D'autres

modifications génétiques sont nécessaires pour rendre les organes totalement compatibles avec les receveurs humains (Houdebine et Weill, 1999).

Malgré certains avantages que les animaux transgéniques apportent au niveau de la recherche, ils soulèvent des problèmes éthiques, surtout en ce qui concerne directement leur bien-être (Van Der Meer et *al.*, 1996).

- Inconvénients de la transgénèse animale

Les applications et les espoirs que font naître la transgénèse sont très nombreux. Néanmoins, l'utilisation de la transgénèse chez les végétaux et les animaux fait l'objet de vives réserves et controverses :

- L'utilisation de l'hormone de croissance humaine chez les animaux perturbe leur croissance et leur reproduction. Les moutons qui sur-expriment GH (hormone de croissance humaine) sont diabétiques et meurent avant l'âge d'un an ;

- Les risques liés à la consommation des produits OGM comme les protéines de la vache folle qui provoque des troubles chez l'homme. Les maïs transgéniques possèdent aussi des gènes de résistance aux ampicillines. Ces transgènes peuvent donc s'intégrer dans les bactéries du tube digestif humain et conférer des résistances bactériennes aux antibiotiques que nous utilisons pour nos traitements :

- Les risques liés à l'environnement : Ces risques sont liés à la dissémination des organismes transgéniques. Des bactéries et des souris transgéniques qui s'échappent d'un laboratoire de recherche ou d'un centre de production, les déchets des produits transgéniques déversés dans la nature constituent de grands dangers qui guettent sans cesse notre écosystème ;

- L'état de santé des animaux clonés ; certains présentent des maladies graves et meurent pendant la gestation ;

- Le syndrome de "gros animaux" ; développement fœtal anormal ;

- Le risque d'extinction d'autres races animales rustiques.

2. Transgénèse végétale

Des plantes peuvent-être régénérées assez facilement à partir d'une cellule somatique. La cellule végétale est donc apparue comme l'unité fondamentale dans le processus de la création d'une lignée de végétaux transgéniques. Sa propriété de totipotence lui confère, in vitro, dans des conditions contrôlées, la capacité de régénérer une plante entière.

2.1 Techniques de transgénèse

2.2.1 Transfert direct

C'est la réalisation d'une transformation par simple conjugaison entre deux types cellulaires, une cellule végétale et une cellule bactérienne du sol de type *Agrobacterium tumefaciens* en modifiant l'ADN-T du plasmide Ti. L'ADN-T est remplacé par un transgène, porteur d'un gène d'intérêt (GI) associé à un gène de sélection (GS) (Figure 65). Du fait des techniques utilisées, certaines séquences bactériennes sont encore présentes dans cette construction simple ; elles sont notées sur la figure (les constructions plus récentes ne possèdent plus ou presque plus de séquences bactériennes).

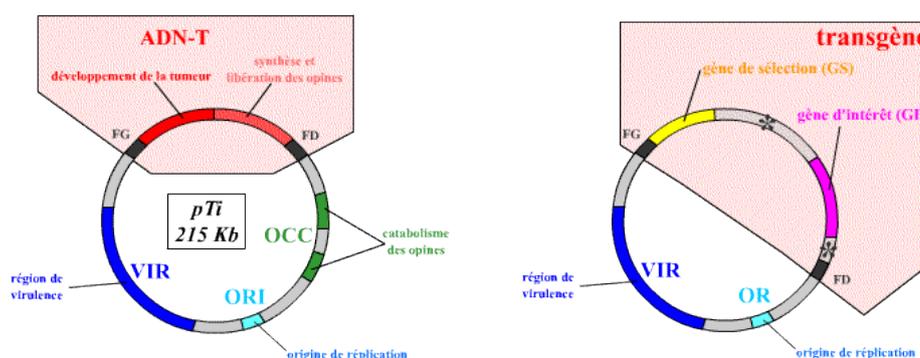


Figure 65 : Remplacement de l'ADN-T par un gène d'intérêt permet d'envisager une technique de transgénèse

(<https://planet-vie.ens.fr/thematiques/manipulations-en-laboratoire/la-transgenese-grace-a-agrobacterium-tumefaciens>).

Ce vecteur est en fait obtenu après recombinaison entre un plasmide pTi modifié et un plasmide portant la construction transgénique (Figure 66).

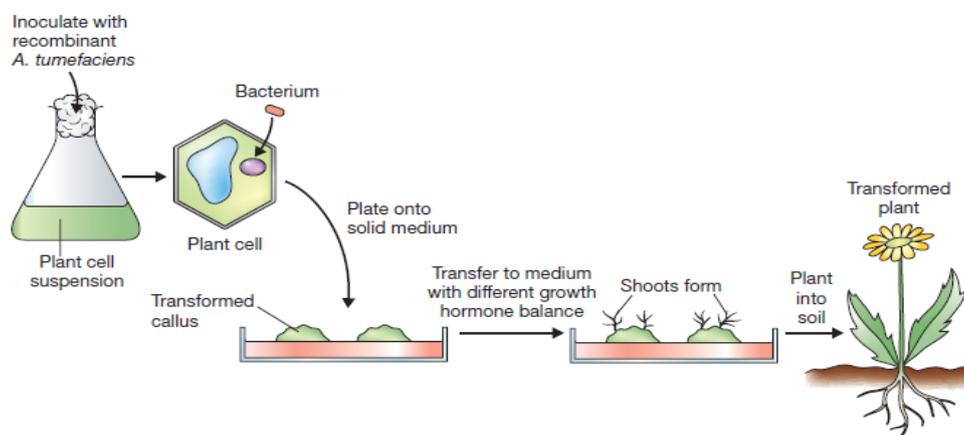


Figure 66 : Transformation des cellules végétales par *Agrobacterium tumefaciens* recombinants (Brown, 2010).

Le gène de sélection permet de repérer facilement les cellules ou amas de cellules qui ont intégré l'ADN transgénique à leur génome. Il s'agit en général d'un gène permettant la survie de ces cellules dans certaines conditions particulières, ou bien d'un gène aboutissant à la présence d'une molécule repérable facilement.

- Avantages de la transgénèse végétale

1. Application thérapeutique

- Production des composés à usages médicaux comme le collagène, l'albumine, l'hormone de croissance ou l'insuline ;

- Production des lipases à partir des plantes ou de graines (tabac-maïs) destinées à soigner les enfants atteints de mucoviscidose ;

- Production de l'hémoglobine dans des plantes pour obtenir des molécules indemnes d'agents infectieux ou pathogènes d'origine animale et pour s'affranchir des contraintes de compatibilité liées aux groupes sanguins, etc... ;

- Production de pomme de terre et de bananes contenant des vaccins. Cette méthode de production contribue aux opérations de vaccination à faibles coûts pour les populations des pays en voie de développement.

2. Application industrielle

- Modification de la qualité de la papeterie à partir d'essence forestière en modifiant le taux de lignine dans le but d'améliorer le rendement en pâte à papier tout en diminuant l'utilisation des produits chimiques d'extraction et de blanchiment ;

- Production des matières plastiques biodégradables ;

- Production des cotons de couleur pour l'industrie textile, dans le but de limiter l'utilisation des teintures, qui contribuent à la pollution ;

- Production de la soie et la laine modifiées à partir d'animaux.

3. Application agronomique

- Résistance à des insectes : Utilisation du gène de l'indoxine (*Bt*) de la bactérie *Bacillus thuringiensis*. En effet, les différentes souches de cette bactérie du sol recèlent plusieurs protéines insecticides ayant différents modes d'action, et affectant uniquement certains insectes tels que la larve (la pyrale) qui détruit le maïs ;

- Résistance à des herbicides : le glufosinate (Basta ou Liberty) et le glyphosate (Roundup) sont des herbicides totaux qui détruisent aussi bien les mauvaises herbes que les plantes cultivées. Les gènes de résistance à l'herbicide introduits dans une plante empêchent la matière active d'agir sur celle-ci, transformant l'herbicide total en herbicide sélectif sur cette plante. Ainsi l'herbicide détruit toutes les mauvaises herbes présentes tout en respectant totalement la plante cultivée ;

- Amélioration des qualités nutritionnelles ;

- Production du riz doré (*Golden Rice*) : introduction de quatre gènes pour la synthèse de la β -carotène (précurseur de la vitamine A) contre la cécité pour la population qui a une carence en vitamine A ;

- Tolérance au stress abiotique : salinité, acidité du sol, sécheresse, froid.

- Inconvénients de la transgénèse végétale

- Les risques toxicologiques : l'ajout d'un nouveau gène dans un organisme vivant peut provoquer l'action d'un gène inactif à l'état normal. Cette action peut être la production de toxines, ou l'augmentation de la production de toxine produite à l'état de traces naturellement. Cette surproduction de toxines peut être toxique pour l'homme. Mais les effets n'ont jamais été prouvés ;

- Le risque allergène : ces risques d'allergies alimentaires existent déjà en dehors des OGM. Mais la consommation d'OGM pourrait accentuer ces risques ;

- Le maïs transgénique contient un transgène qui peuvent s'intégrer dans les bactéries du tube digestif humain et confèrent la résistance aux antibiotiques que nous utilisons pour nos traitements ;

- La contamination entre variétés : dans l'espèce végétale, la reproduction s'effectue par croisements sexuels. C'est le pollen qui est le vecteur privilégié de cette dissémination, il est transporté par le vent ou les insectes comme les abeilles. Cette transmission peut se faire entre plantes de la même espèce ou en direction d'espèces sauvages (dites mauvaises herbes). Cette transmission peut entraîner l'apparition de mauvaises herbes résistantes aux herbicides ou aux insectes par exemple ;

- L'apparition d'insectes résistants aux plantes transgéniques : le principal insecte ravageur du maïs en France est la pyrale. A l'heure actuelle la protection des cultures de maïs est assurée par un traitement chimique mais la transgénèse a permis de créer de nouvelles variétés de maïs qui ont été transformées par des firmes privées pour produire dans leurs tissus la toxine (*Bt*) active contre la pyrale du maïs. On a déjà remarqué l'apparition d'insectes résistants à cette toxine dans certains pays (Malaisie, Japon, Hawaï). L'utilisation de ce maïs transgénique pourrait créer une nouvelle génération de pyrales résistantes à la toxine *Bt* ;

- L'éventuel impact sur les insectes utiles comme l'abeille : il est important de s'assurer que les plantes génétiquement modifiées (par insertion de gènes de résistance aux insectes ravageurs) ne soient pas toxiques vis à vis d'autres insectes non ciblés. Les insectes non ciblés sont ceux qui ne sont pas visés par la modification génétique qu'a subi la plante, c'est à dire ceux qui peuvent même être bénéfiques pour l'environnement, d'où leur

nom d'insectes "utiles". Ces insectes pouvant être les abeilles, des vers de terre ou des coccinelles par exemple ;

- Le riz doré est privé de la vitamine E ;

- Problème d'étiquetage : les animaux qui se nourrissent de l'alimentation modifiée génétiquement, peut-on étiqueter ces animaux ?

La transgénèse végétale présente des avantages mais malheureusement beaucoup d'inconvénients voire même des risques de santé pour l'homme.

Chapitre VIII.
Mutagenèse aléatoire et dirigée

Chapitre VIII. Mutagenèse aléatoire et dirigée

1. Notion et principe

La mutagenèse aléatoire ou dirigée est une approche utilisée par le génie génétique et la biologie pour comprendre la fonction des gènes ; elle consiste en l'introduction volontaire de mutations par l'action d'agents mutagènes chimiques ou physiques dans une séquence ADN afin de déduire des informations sur le rôle des gènes, à partir de l'analyse des effets de ces mutations. Il y a deux grands types de mutagenèse :

1.1 Mutagenèse aléatoire

Le segment d'ADN est amplifié *in vitro* en utilisant la technique de PCR. Le taux d'erreur d'incorporation des nucléotides est assez élevé pour que la probabilité d'obtention d'une copie présentant une mutation soit proche de 1. Ainsi, une population de molécules est obtenue dont un grand nombre diffère de l'allèle sauvage par le changement d'au moins une paire de nucléotides. Ces copies modifiées sont introduites par transfection dans des cellules de manière à ce que l'allèle sauvage du gène étudié soit remplacé par une des copies mutées *in vitro*. Les clones transformés qui présentent une différence phénotypique avec les cellules sauvages de référence sont recherchés. Ensuite, le gène muté est isolé à nouveau à partir des cellules du clone transformé considéré et le séquençage de son ADN peut déterminer la nature et la position de la modification qui y a été introduite. Cette technique permet d'identifier quelles sont les régions de la protéine étudiée qui jouent un rôle clé dans sa fonction (Rossignol et *al.*, 2000).

- Insertion ou délétion

On peut déléter un fragment de restriction du gène et ressouder sur lui-même le gène délété. Un gène marqueur peut aussi être inséré à la place de ce fragment (ou dans un site de restriction) ou un oligonucléotide synthétique qui interrompent le gène (mutagenèse insertionnelle).

- Inactivation d'un gène par recombinaison homologue (souris *knock out*)

Des études effectuées dans les années 1980, en particulier par les groupes de Smithies et de Capecchi, avaient démontré que les cellules de mammifères possèdent l'appareil enzymatique nécessaire à la recombinaison entre une séquence d'ADN exogène et la séquence homologue présente *in situ* dans les chromosomes, même s'il s'agit d'un événement

relativement rare en comparaison de l'intégration au hasard de ce même ADN (Smithies et *al.*, 1985). Le vecteur de ciblage comporte une cassette de sélection insérée dans un exon (les rectangles noir et blanc représentent respectivement un exon et un intron). Cette cassette de sélection est également entourée par des régions présentant des homologues avec le gène cible. La recombinaison avec ce dernier s'effectue au niveau de ces séquences homologues, ce qui conduit à la création d'un allèle nul et donc à l'inactivation du gène suite à l'insertion de la cassette de sélection. La cassette de sélection permettra ensuite de reconnaître les cellules où la recombinaison a eu lieu (Babinet et Cohen-Tannoudji, 2000) A ce jour, plusieurs centaines de gènes ont été invalidés de la sorte et les souris mutantes correspondantes créées. L'analyse des phénotypes engendrés par ces mutations a apporté de très nombreuses lumières sur la fonction des gènes concernés (Figure 67).

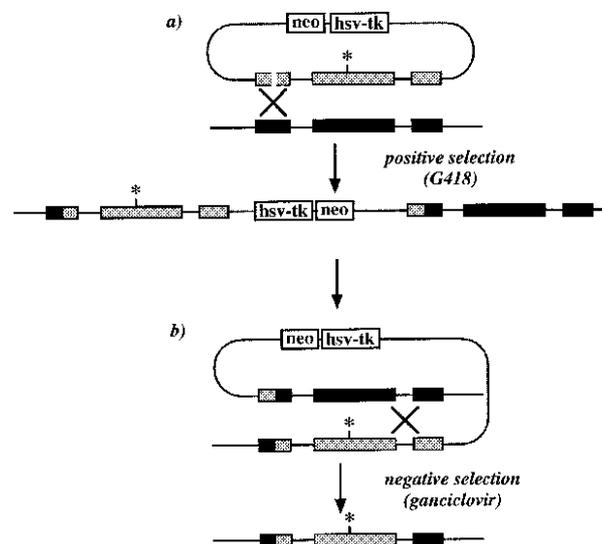


Figure 67 : Principe général de la recombinaison homologue (Cohen-Tannoudji et Babinet, 1998).

1.2 Mutagenèse dirigée par PCR

L'introduction d'une mutation par PCR permet de créer 100 de mutants. La première méthode nécessite l'utilisation de quatre amorces oligonucléotidiques différentes. Les deux premiers oligonucléotides (notés 1 et 2) sont complémentaires l'un de l'autre au niveau de la région 5'-P mutée. Les deux autres (notés 3 et 4) permettent d'amplifier la totalité de l'ADNc que l'on désire muter. Ils sont donc complémentaires des extrémités 3'-OH et 5'-P de l'ADNc présent dans le plasmide.

La seconde méthode est dite du "megaprimer". Elle nécessite l'utilisation de trois amorces oligonucléotidiques. La première porte la mutation que l'on désire introduire (notée 1) et est complémentaire de l'un des brins de l'ADN cible. Les deux autres (notées 3 et 4), comme précédemment, correspondent aux extrémités de l'ADNc.

Dans la première réaction de PCR, l'utilisation des amorces 1 et 4 permet l'obtention d'un fragment PCR portant la mutation et correspondant à une séquence tronquée de l'ADNc matrice. Ce fragment est purifié et va servir d'amorce (megaprimer) dans une deuxième réaction de PCR avec l'oligonucléotide 3. De cette façon, un fragment complet et muté sera généré (Figure 68).

L'un des inconvénients de la technique de mutagenèse par PCR, ce sont des erreurs de lecture de la *Taq* polymérase. Ce problème est aisément surmonté en utilisant un nombre réduit de cycles (< 30) (Damier et Jacquot, 2003).

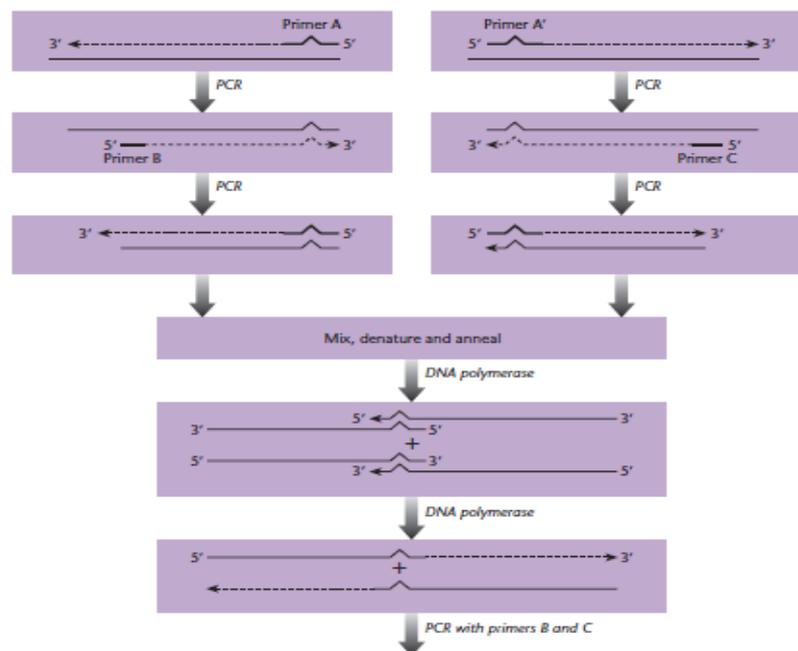


Figure 68 : Mutagenèse dirigée utilisant la PCR (Primrose et Twyman, 2006).

La moitié inférieure de la figure montre comment la mutation peut être introduite au milieu d'une molécule d'ADN. Les amorces sont représentées en traits gras. Les amorces A et A' sont complémentaires.

2. Intérêts de la mutagenèse

- Etude de la fonction des gènes (perte de fonction, gain de fonction, modifications ...)
- Produire de la diversité génétique, des nouveaux génotypes d'intérêt (amélioration des plantes, sélection animale, bactéries, etc...).

Références bibliographiques

Références bibliographiques

- Alberts B., Johnson A., Lewis J., Morgan D., Raff M., Roberts K. et Walter P. (2017). Biologie moléculaire de la cellule. 6ème Edition Lavoisier Médecine Sciences, Paris. 1463p.
- Babinet, C. et Cohen-Tannoudji, M. (2000). Vingt ans d'interventions délibérées sur le génome de la souris. *Médecine/sciences*, 16 (1) : 31-42.
- Baum C., Kustikova O., Modlich U., Li Z. et Fehse B. (2006). Mutagenesis and Oncogenesis by Chromosomal Insertion of Gene Transfer Vectors. *Hum. Gene Ther* 17 : 253-63.
- Beardmore J.A. (1997). Transgenics, autotransgenics, and allotransgenics, *Transgenics Research*, 6 : 107-108.
- Billault A., Susini L. et Soravito de Franceschi C. (1996). Les vecteurs YAC et BAC. Technoscope Biofutur, 153.
- Botstein D., White R.L., Skolnick M. et Davis R.W. (1985). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am J Hum Genet.*, 32 : 314-331.
- Burke D.T., Carle G.F. et Olson M.V. (1987). Cloning of large segment of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, 236 : 806-811.
- Brown T.A. (2004). Génomes. *Edition Médecine-Sciences Flammarion*, Paris. 572 p.
- Brown T.A. (2010). Gene Cloning and DNA Analysis: An Introduction. 6th Ed, *Weley-Blackwell*. University of Manchester, London, 376 p.
- Chandler V. (1994). Overview of a Ding Genes Using Transposon Tagging. *In* : The Maize Handbook-M. Freeling, V. Walbot, eds. *Springer-Verlag, New York, Inc.* pp. 647.
- Clancy S. et Brown W. (2008). Translation: DNA to mRNA to Protein. *Nature Education*, 1(1):101.
- Cohen S.N. (1975). The manipulation of genes. *Sci Am.* 233(1):25-33.
- Cohen-Tannoudji M. et Babinet C. (1998). Beyond 'knock-out' mice: new perspectives for the programmed modification of the mammalian genome. *Molecular Human Reproduction*, 4(10): 929-938
- Crépin M. (1987). Expression des gènes et génie génétique. *Edition Hermann*, Paris. 305 p.
- Damier L. et Jacquot J.P. (2003). *In* : Tagu D. et Moussard C. (2003). Principes des techniques de biologie moléculaire. 2^{ème} édition *INRA*, Paris. pp. 131.

- Dupuy Y. et Nougier P. (2005). Les micro-organismes - Du gène à la biosphère. Ellipses *Edition Marketing*, Paris. 256 p.
- Frank D. Scherson R.I. et Broach J.R. (1992). Direct simulation of yeast 2- μ m circle plasmid amplification. *Journal of Theoretical Biology*, 155, Issue 3 pp. 369-385.
- Freifelder D. (1990). Biologie moléculaire. *Edition Masson*, Paris. 338 p.
- Gibson G. et Muse S.V. (2004). Précis de génomique. 1^{ère} Edition *De Boeck*, Bruxelles. 347p.
- Gordon J.W., Scangos G.A., Plotkin, D.J., Barbosa, J.A., Ruddle, F.H. (1980). Genetic transformation of mouse embryos by micro-injection of purified DNA, *PNAS USA*, 77 : 7380-84.
- Gordon J.W. et Ruddle F.H. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei, *Science*, 214 : 1244-1246.
- Gordon J.W. (1997). Transgenic Technology and Laboratory Animal Science. *ILAR Journal*, 38 (1) 32-41.
- Griffiths A., Gelbart W.T., Miller J.H et Lewontin R. (2001). Analyse génétique moderne. 1^{ère} Edition *De Boeck*, Paris. 676 p.
- Griffiths Wessler Lewontin Gelbart et Miller. (2006). Introduction à l'analyse génétique. 4^{ème} édition *De Boeck*, Bruxelles. 782 p.
- Gunderson K.L., Steemers F.J., Lee G., Mendoza L.G. et Chee M.S. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genet.* 37:549-554.
- Hanna N., Parfait B., Vidaud D. et Vidaud M. (2005). Mécanismes et conséquences des mutations. *Médecine/Sciences*, 21 : 969-80.
- Harry M. (2008). Génétique moléculaire et évolutive. 2^{ème} Edition *Maloine*, Paris. 465 p.
- Hartl D. H et Jones E. W. (2003). Génétique, les grands principes. 3^{ème} Edition *Dunod*, Paris, 609 p.
- Hong L., Schroth G. P., Matthews H.R., Yau P., Bradbury E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA." *J Biol Chem.*, 268(1): 305-14.
- Houdebine L.M. et Weill B. (1999). The impact of transgenesis and cloning on cell and organ xenotransplantation to humans. ECS 9 Brussels. *Focus on Biotechnology*, 351-361.
- Houdebine L.M. (2002). Les apports de la transgénèse dans la recherche biomédicale. Vol (5) n° 6 : 240-247. *Act. Méd. Int. - Métabolismes - Hormones – Nutrition*.
- Houdebine L.M. (2003). In : Tagu D. et Moussard C. (2003). Principes des techniques de biologie moléculaire. 2^{ème} édition *INRA*, Paris. pp. 109.

- Jaenisch R. (1976). Germ line integration and Mendelian transmission of the exogenous Moloney Leukemia virus, *Proc. Nat. Acad. Sci. USA*, 73 (4) : 1260-1264.
- Jaenisch R. (1988). Transgenic Animals, *Science*, 240 :1468-1474.
- Jallat S. (1991). Les souris transgéniques. *Biofutur*, 45 : 3-11.
- Janne J. et Alhonen L. (1996). Transgenic animals: practical aspects, management, logistics. *In* : Proceedings of the Sixth Symposium of the Federation of European Laboratory Animal Science Associations, Harmonization of Laboratory Animal Husbandry, Basel, Switzerland, London : *The Royal Society of Medicine Press Limited*, 41-44.
- Kamoun P. (1997). Appareils et méthodes en biochimie et biologie moléculaire. Ed. *Flammarion Médecine-Sciences*, Paris. 418 p.
- Kaplan J.C. et Delpéch M. (2007). Biologie moléculaire et médecine. 3^{ème} édition Médecine-Sciences Flammarion, Paris. 815 p.
- Kernbaum S. (1998). Dictionnaire de Médecine, 6^{ème} Edition Médecine-Sciences, *Flammarion*, Paris. 1030 p.
- Klug W., Cummings M. et Spencer C. (2006). Génétique. 8^{ème} édition *Pearson*, Paris, 704 p.
- Kochko (2000). De l'utilisation des chromosomes artificiels de plantes comme outil pour la conservation et l'exploitation des ressources génétiques végétales. *Cahiers Agriculture*, 9 : 287-292.
- Kouprina N., Petrov N., Molina O., Liskovych M., Pesenti E., Ohzeki J.I., Masumoto H., Earnshaw W.C. et Vladimir Larionov V. (2018). Human Artificial Chromosome with Regulated Centromere: A Tool for Genome and Cancer Studies. *ACS Synth Biol.* 7(9): 1974-1989.
- Lagziel A., DeNise S., Hanotte O., Dhara S., Glazko V., Broadhead A., Davoli R., Russo V., et Soller M. (2000). Geographic and breed distribution of an *Msp* I PCR-RFLP in bovine growth hormone (bGH) gene. *Animal Genetics*, 31 : 210-213.
- Lander E.S., Linton L.M., Birren B. *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature*, 409 (6822): 860-921.
- Lemarchandel V. et Montagutelli X. (1990). La recombinaison homologue De nouvelles perspectives pour la transgénèse chez les mammifères. *médecine/sciences*, 6 : 18-29.
- Leroux C. et Tosser-Klopp G. (2000). 1 - Notions de base de génétique. La fonction du gène : les grandes étapes de l'utilisation de l'information génétique. *INRA Prod. Anim., numéro hors-série "Génétique moléculaire : principes et application aux populations animales"*, 21-28.
- Lewin B. (1999). Gènes VI. *Edition De Boeck*, Bruxelles. 179 p.

- Lewis J.D et Izaurralde E. (1997). The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem*, 247 : 461-9.
- Lodish H., Berk A., Matsudaira P., Kaiser C.A, Krieger M., Scott M.P, Zipursky L. et Darnell J. (2003). *Molecular Cell Biology. 6th Edition Molecular Cell Biology*. 937 p.
- Loenen W.A.M. (2019). *Restriction Enzymes : A history. Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, New York. pp. 67.
- Lucotte G. (1983). *ABC de génie génétique - Initiation aux techniques de clonage, de séquençage et d'expression des gènes eucaryotes. InterEditions*, Paris. 190 p.
- Luger K., Mäder A.W., Richmond R.K., Sargent D.F. et Richmond T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389, 251-260.
- Maxam A.M. et Gilbert W. (1977). A new method for sequencing DNA. *Proc Natl Acad Sci USA*. 74 : 560–564.
- McClintock B. (1948). Mutable Loci in Maize. *Carnegie Institution of Washington Year Book*, 47: 155-169.
- Michel B. et Baldacci G. (1998). Réplication. *médecine/sciences*, 14 : 1422-7.
- Miller S.A, Dykes D.D et Polesky H.F. (1988). A simple salting-out for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16(3) : 1215.
- Milman G. et Herzberg M. (1981). Efficient DNA transfection and rapid assay for thymidine kinase activity and viral antigenic determinants. *Somat. Cell Genet*, 7 :161-70
- Mullis K., Faloona F., Scharf S., Saiki R. K., Horn G. et Erlich H. (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposium on Quantitative Biology*. 51: 263–273.
- Orratia E. et Chang P.L. (1990). Intracellular distribution of DNA internalized through calcium phosphate precipitation. *Exp. Cell Res*. 190 : 170-4.
- Pallier C. (2001). Les régions non traduites des ARN messagers et leur rôle dans la synthèse protéique. *médecine/sciences*, 17 : 23-f32.
- Passarge E. (2007). *Color Atlas of Genetics. 3rd Ed. Flexibook. Thieme Stuttgart*, NY. USA.
- Pierce B.A. (2012). *L'essentiel de la génétique. 1^{ère} Ed. De boeck*, Bruxelles. 458 p.
- Pray L.A. (2008). Transposons: The Jumping Genes. *Nature Education*, 1(1) : 204.
- Primrose S., Twyman R. et Old R. (2004). *Génie génétique. 1^{ère} Edition De Boeck*, Bruxelles, 400 p.
- Primrose S.B et Twyman R.M. (2006). *Principles of Gene Manipulation and Genomics. Seventh edition Blackwell Publishing*. 672 p.

- Riquet J. et Pitel F. (2000). 1- Notions de base de génétique : Les techniques de base de la génétique moléculaire. *INRA Prod. Anim., numéro hors série "Génétique moléculaire : principes et application aux populations animales"*, 29-35.
- Rogel-Gaillard C. (2000). Les banques de grands fragments d'ADN. *INRA Prod. Anim., numéro hors-série "Génétique moléculaire : principes et application aux populations animales"*, 79-85.
- Rossignol J.L., Berger R., Deutsch J., Fellous M., Lamour-Isnard C., Ozier-Kalogeropoulos O., Picard M. et De Vienne D. (2000). Génétique : Gènes et génomes - Cours et questions de révision. *Edition Dunod*, Paris. pp. 57.
- Sambrook J., Fritsch E.F. et Maniatis T. (1989). *Molecular cloning: A Laboratory Manual, 2nd éd. Cold Spring Harbor, Laboratory press, N. J., USA.*
- Sanger F., Nicklen S. et Coulson A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 74 (12) : 5463-7.
- Schindelhauer D. (1999). Construction of mammalian artificial chromosomes: prospects for defining an optimal centromere. *Bioessays*, 21(1) : 76-83.
- Shizuya H., Birren B., Kim U.J., Mancino V., Slepak T., Tachiri Y. et Simon M. (1992). Cloning and stable maintenance of 300-kilobasepair. fragments of human DNA in *Escherichia coli* using F-factorbased vector. *Proceedings of the National Academy of Science USA*, 89 : 8794-8797.
- Smith H.O. et Nathans D. (1973). Letter: a suggested nomenclature for bacterial host modification and restriction systems and their enzymes. *J Mol Biol.*, 81: 419-423.
- Smithies O., Gregg R.G., Boggs S.S., Koralewski M.A. et Kucherlapati R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature*, 317 : 230-4.
- Solignac M., Periquet G., Anxolabéhère D. et Petit C. (1995). Génétique et évolution, L'espèce, l'évolution moléculaire. Tome II. *Ed. Hermann*, Paris. 307 p.
- Southern E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis Southern. *Journal of molecular biology*, 98 : 503-508.
- Sternberg N. (1990). Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc. Nati. Acad. Sci. USA*, 87 : 103-107.
- Strachan T et Read A.P. (1998). Génétique Moléculaire Humaine. *Edition Flammarion-Médecine-Sciences*, Paris. 596 p.

- Strachan T. et Read A. (2012). Génétique Moléculaire humaine. 4^{ème} édition *Médecine Sciences Publications Lavoisier*, Paris. pp. 198-199.
- Tagu D. et Moussard C. (2003). Principes des techniques de biologie moléculaire. 2^{ème} édition *INRA*, Paris. 176 p.
- Tagu D., Jaubert-Possamai S. et Méreau A. (2018). Principes des techniques de biologie moléculaire et génomique. 3^{ème} édition *Quae*, Versailles, France. 328 p.
- Thiry M., Bourmeyster N., Dommès J., Lebrun M. et Rigo P. (2016). Biologie moléculaire- Exercices et méthodes. *Edition. Dunod*, Malkoff, Paris. 352 p.
- Trotureau A., Kern J., Viardot A., Thiriet A. et Schouler C. (2018). Explorer le potentiel thérapeutique des coliphages : approche curative des infections causées par *Escherichia coli* chez les oiseaux. *Innovations Agronomiques* 66 : 9-17.
- Van Der Meer M., Baumans V. et Van Zuphten L.F.M. (1996). Use and welfare aspects of transgenic animals, In : Proceedings of the Sixth Symposium of the Federation of European Laboratory Animal Science Associations, Harmonization of Laboratory Animal Husbandry, Basel, Switzerland, London : *The Royal Society of Medicine Press Limited*, 49-51.
- Varani G. (1997). A cap for all occasions. *Structure*, 5 : 855-858.
- Wahle E. et Kuhn U. (1997). The mechanism of 3' cleavage and polyadenylation of eukaryotic pre-mRNA. *Prog Nucleic Acids Res Mol Biol*, 57 : 41-71.
- Warburg O. et Christian W. (1942). Isolation and crystallization of enolase. *Biochem. Z.* 310 : 384-421.
- Watson J.D. et Crick F.H.C. (1953). Molecular Structure of Nucleic Acids. *Nature* 171, 737-738.
- Watson J.D., Baker T.A., Bell S.P., Gann A., Levine M. et Losick R. (2012). Biologiémoléculaire du gene. 6^{ème} édition *Pearson Education*, France. 588 p.
- Wilmut I., Schnieke P., McWhir J., Kind A.J et Campbell K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* 385 : 810–813.
- Winter P.C., Hickey G.I. et Fletcher H.L. (1999). L'essentiel en génétique. *Edition Berti*, Paris. 401 p.
- Winter S., Simboeck E., Fischle W., Zupkovitz G., Dohnal I., Mechtler K., Ammerer G. et Seiser C. (2008). 14-3-3 proteins recognize a histone code at histone H3 and are required for transcriptional activation. *Embo Journal*, 27(1): 88-99.