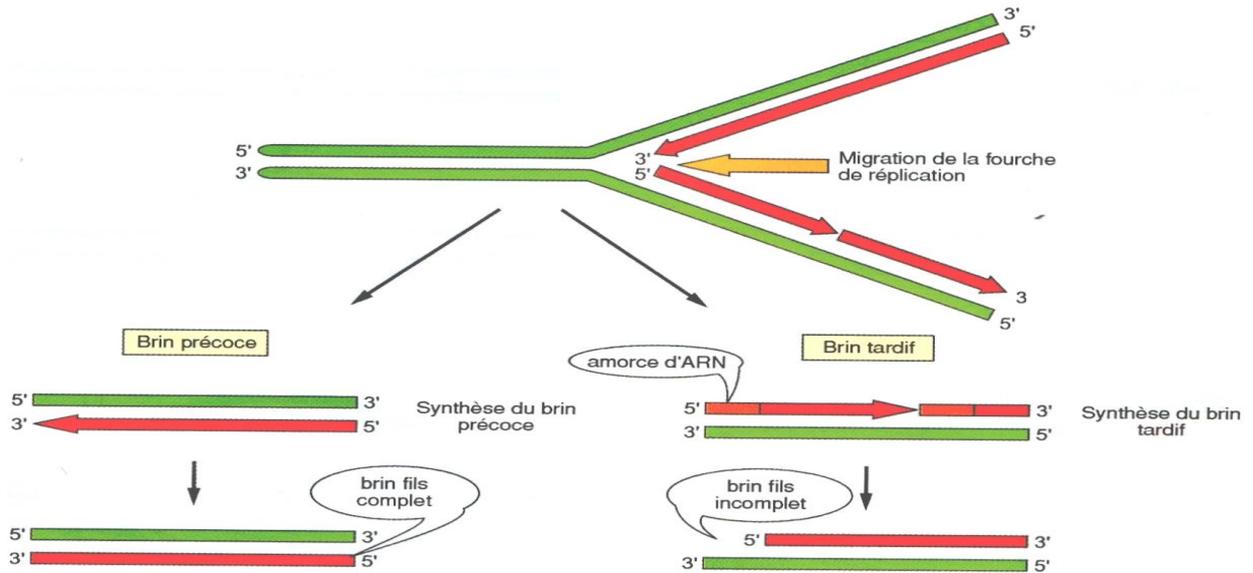
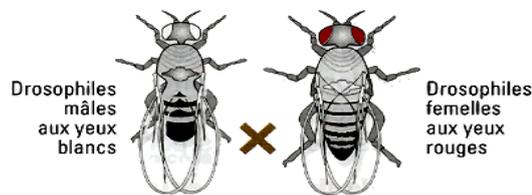


Département : Biologie
Filière : Biosciences
Semestre 4



PARENTS

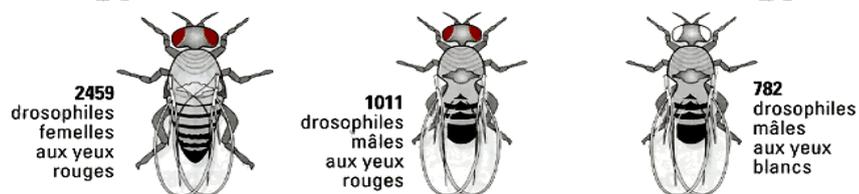
(souches pures)



F1



F2 = F1 x F1



Cours du module de BIOLOGIE MOLECULAIRE et GENETIQUE (M26)

Responsable : Pr. Khalil HAMMANI

Année Universitaire 2024-2025

Chapitre I. INTRODUCTION AU DOGME CENTRAL, STRUCTURE ET PROPRIETES DES ACIDES NUCLEIQUES

I.1. Introduction

Le phénomène de la vie sur Terre peut être envisagé, de façon simplificatrice, comme la transmission d'une génération à l'autre d'un ensemble d'informations. Le système de stockage et de transmission d'informations doit être assez robuste pour que la plupart des descendants d'un être donné soient viables, mais assez sujet à des variations pour générer avec le temps des formes vivantes de plus en plus diversifiées.

La biologie moléculaire est une discipline scientifique au croisement de la génétique, de la biochimie et de la physique, dont l'objet est la compréhension des mécanismes de fonctionnement de la cellule au niveau moléculaire. Le terme 'biologie moléculaire', utilisé la première fois en 1938 par **Warren Weaver**, désigne également l'ensemble des techniques de manipulation d'acides nucléiques (ADN, ARN), appelées aussi techniques de génie génétique.

La biologie moléculaire est apparue au XX^{ème} siècle, à la suite de l'élaboration des lois de la génétique, la découverte des chromosomes et l'identification de l'ADN comme support chimique de l'information génétique. Après la découverte de la structure en double hélice de l'ADN en 1953 par James Watson, Francis Crick, Maurice Wilkins et Rosalind Franklin la biologie moléculaire a connu d'importants développements pour devenir un outil incontournable de la biologie moderne à partir des années 1970.

Les chercheurs en biologie moléculaire utilisent des techniques spécifiques pour la biologie moléculaire (voir chapitre sur outils et techniques de la biologie moléculaire), mais les combinent de plus en plus avec les techniques et les idées provenant de la génétique et de la biochimie. Il n'y a pas de frontière bien définie entre ces disciplines, bien qu'il y en ait eu à une certaine époque.

I.2. Le Dogme Central de la biologie moléculaire

Le 'dogme central', introduit par Francis Crick (le co-découvreur de la structure de l'ADN) à la fin des années 50, veut que dans tous les êtres vivants l'information ne soit transmise que dans un sens (Figure 1): de l'ADN, où repose l'information, à l'ARN, une structure transitoire permettant sa transmission à une machine de traduction, aux protéines, les constituants de base qui font fonctionner la cellule et l'organisme entier.

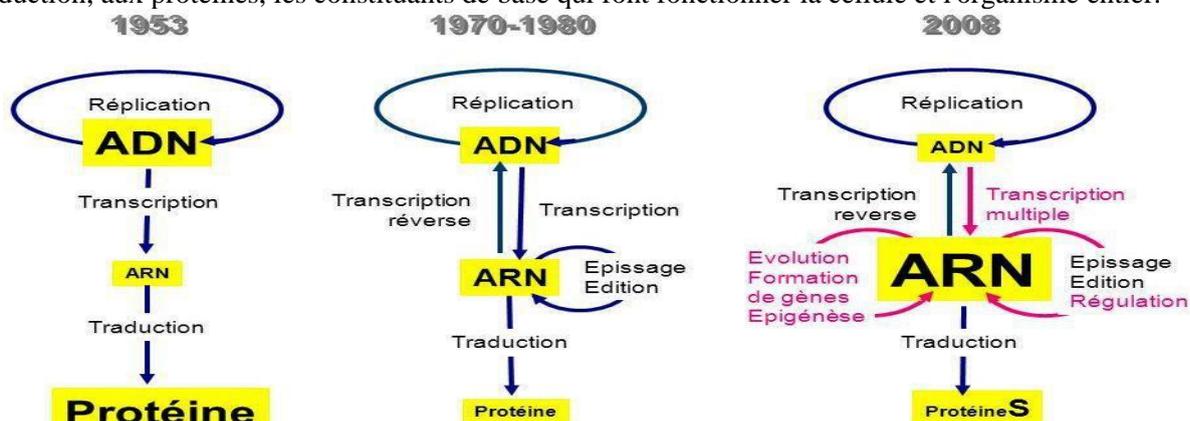


Figure 1. Schéma montrant le principe du 'Dogme Central de la biologie moléculaire'

Les flèches dans l'image ci-dessus représentent les processus de réplication de l'ADN, de transcription de l'ADN en ARN messager, et de traduction de l'ARN messager en protéine. Ces processus, qui forment l'essentiel de l'expression matérielle du dogme central, seront expliqués dans les autres chapitres de ce cours.

La biologie moléculaire est l'étude des processus de réplication, de transcription et de traduction du matériel génétique. Le dogme central de la biologie moléculaire où le matériel génétique est transcrit en ARN, puis traduit en protéines, bien qu'il soit une image très simpliste et sans fondement de la biologie moléculaire, fournit encore un bon point de départ pour comprendre ce domaine. Cette image, cependant, doit être révisée à la lumière des nouveaux rôles qu'on découvre à l'ARN (Exemple: le ribozyme). L'essentiel du travail en biologie moléculaire est **quantitatif**, et récemment beaucoup de travaux ont été faits à l'intersection de la biologie moléculaire et de l'informatique, dans la bio- informatique et dans la biologie calculatoire. Depuis les

années 2000, l'étude de la structure et de la fonction des gènes, la génétique moléculaire, fait partie des sous-domaines les plus saillants de la biologie moléculaire.

Il faut réaliser d'emblée que la quantité d'information génétique nécessaire pour construire et faire fonctionner un organisme complexe est énorme, et que chaque cellule du corps humain en contient deux copies, dérivées de nos deux parents. Les chiffres suivants sont éloquentes :

- **Virus:** 3000 nt, une page de 3000 caractères
- **Bactérie:** 3×10^6 nt, un livre de 100 pages
- **Homme:** 3×10^9 nt, 1000 livres.

I.3. De l'information génétique au programme génétique

I.3.1. Nature du matériel génétique

L'analyse génétique, qui sera traitée dans les chapitres suivants, a permis d'aller très loin dans l'étude du mode de transmission des caractères, dans certains cas d'interactions entre les gènes, leur modes d'expression peuvent parfois être compris grâce au raisonnement mendélien. Malgré cela, l'analyse, pourtant relativement simple, du dihybridisme (avec dominance) qui se traduit, en F2, par quatre phénotypes dans des proportions 9:1:3:3 alors qu'il existe plusieurs génotypes possibles, laisse prévoir que le petit pois et la drosophile sont des organismes (diploïdes) bien trop compliqués pour saisir les mécanismes fondamentaux de la recombinaison et de l'expression des gènes. C'est grâce aux génomes haploïdes tels que ceux des bactéries et des virus vont révolutionner la notion de gène.

Des séries de travaux s'étalant de 1928 à 1952 vont permettre d'associer définitivement l'ADN à la notion d'information génétique. Ces travaux sont exposés ci-dessous car ils présentent, comme ceux de Mendel, beaucoup plus qu'un intérêt historique : ils sont des modèles d'analyse faisant appel, pour la première fois, à des méthodes de la biologie moléculaire.

I.3.2. Travaux de Griffith (1928)

Griffith a travaillé avec deux types de souches de *Streptococcus pneumoniae*:

- des souches virulentes létales pour les souris qu'il appelle "S" car elles ont un aspect lisse lorsqu'elles forment des colonies sur boîte de Pétri (s = smooth).
- des souches non virulentes "R" donnant des colonies rugueuses sur boîte de Pétri (r = rough).

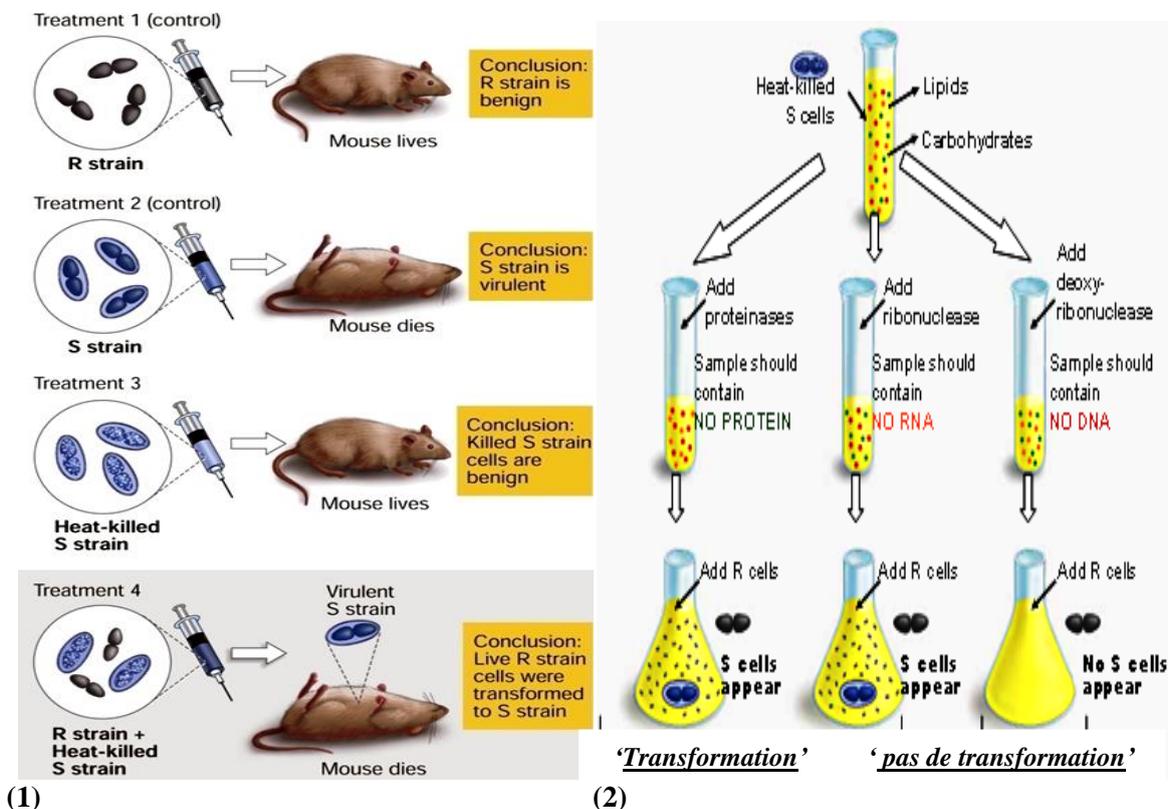


Figure 2. Travaux de Griffith (1) et Avery, McLeod et McCarty (2).

La différence de morphologie est due à la présence de polysaccharides protecteurs autour de la cellule chez les souches S. Ces polysaccharides ont différentes compositions qui spécifient le type des pneumocoques (I, II, III etc.). Ils sont facilement visualisables par l'utilisation d'anticorps spécifiques.

Lorsque Griffith chauffe les cellules S de type III et qu'il les injecte à des souris, celles-ci ne meurent pas, le chauffage ayant tué les pneumocoques (on peut le vérifier après étalement sur milieu gélosé). De même, l'injection de cellules R de type II vivantes ne provoque pas la mort des souris. Par contre s'il injecte des cellules S III chauffées avec des cellules R II, il constate la mort rapide des souris. Il extrait les bactéries des cadavres et les étale sur boîte de Pétri: les bactéries qui poussent ont un aspect lisse, elles sont S III : Quelque chose qui porte l'information "S III" dans les cellules S tuées se transmet aux cellules R vivantes et les transforme en S. Il a donc découvert le principe de la transformation. Celle-ci consiste en l'introduction à l'intérieur de la cellule d'une information exogène provenant d'une autre cellule. On exclut évidemment dans cette définition, l'introduction d'information virale (pour laquelle on utilise plutôt le terme de transfection). Quelques années plus tard un système de transformation in vitro et bien contrôlé a été mis au point, ce qui a grandement simplifié les procédures de transformation. A l'époque, Griffith ne peut que conclure à l'existence d'un facteur transformant or il s'agit bien d'une transformation génétique au sens actuel du terme.

I.3.3. Travaux d'Avery, McLeod et McCarty (1943)

C'est Oswald Avery (1877-1955) et ses collègues Colin McLeod et McCarty, qui réussissent à purifier le facteur transformant du pneumocoque après d'innombrables tentatives infructueuses de nombreux chercheurs pendant 10 ans. Ils réalisent un extrait chimique assez bien purifié de SIII (voir Figure 2 et expérience de Griffith) contenant essentiellement des acides nucléiques (ARN et ADN) et un peu de protéines, après avoir éliminé les lipides et glucides. Ils ajoutent une petite quantité de RII vivantes à une grande quantité d'extrait de SIII. Ils obtiennent des SIII vivantes. S'ils ajoutent à l'extrait de la trypsine ou de la chymotrypsine (enzymes coupant de nombreuses protéines) le résultat est le même. S'ils ajoutent une ribonucléase à l'extrait le résultat est identique. Par contre la transformation n'a pas lieu s'ils ajoutent de la désoxyribonucléase. C'est donc bien l'ADN qui est le principe transformant.

En analysant ces deux expériences dans une perspective moderne, on a donc la preuve que l'ADN est une **molécule transmissible et susceptible d'être utilisée par une cellule procaryote** pour modifier une de ses caractéristiques structurales : ici la synthèse d'une capsule.

I.3.4. Travaux de Hershey et Chase (1952)

En 1952, Hershey et Chase étudient la reproduction du bactériophage T2 dans la bactérie *Escherichia coli* (qui deviendra le monstre sacré de la génétique moderne) et emploient une technique qui va se développer rapidement : l'utilisation d'isotopes radioactifs comme traceurs, comme marqueurs permettant de suivre la destinée de macromolécules (Figure 3).

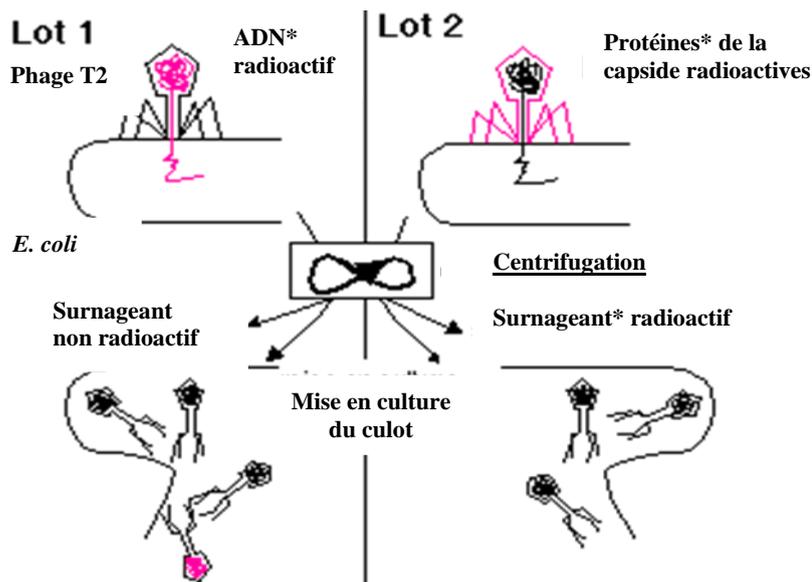


Figure 3. Expérience de Harshey et Chase.

Dans un premier temps, ils cultivent les bactéries (*E. coli*) sur un milieu contenant du phosphore 32 (³²P) et du soufre 35 (³⁵S), après un certain temps de culture, les éléments constitutifs des bactéries contiennent ces marqueurs. On infecte alors la culture avec une suspension de phages T2, ceux-ci vont réaliser un cycle lytique en utilisant les molécules radioactives de leurs cellules hôtes. La descendance phagique est recueillie et sert à infecter des bactéries normales. On sait que, dans la première étape de l'infection, les phages s'adsorbent sur la bactérie et injectent à l'intérieur une molécule informative. Après cette étape d'adsorption, les auteurs agitent violemment la suspension pour décrocher ce qui reste à l'extérieur des bactéries, après centrifugation, on obtient un culot bactérien contenant l'information phagique et un surnageant contenant la capsid (Figure 3). Or, le culot contient le ³²P et le surnageant le ³⁵S, c'est la démonstration éclatante que l'information génétique du bactériophage, qui pénètre à l'intérieur de la bactérie est de l'ADN et que la capsid protéique ne sert que d'emballage.

Cette expérience semble avoir eu un retentissement très grand notamment du fait qu'elle fut publiée à peu près en même temps que le modèle de la structure de la molécule d'ADN en 1953 par James Watson et Francis Crick.

I.4. Structure et propriétés des acides nucléiques

1.4.1. Travaux de Watson et Crick

Watson et Crick ont proposé en 1953 un modèle moléculaire en **double hélice** maintenue par des liaisons hydrogène entre des bases précises. Ils ont progressivement élaboré leur modèle moléculaire à partir d'images, souvent difficiles à interpréter, de diffraction des rayons X par la molécule d'ADN (obtenues par Maurice Wilkins et de Rosalind Franklin) laissant supposer une certaine régularité et une certaine répétition dans cette molécule très longue. Ils ont également tenu compte des observations d'Erwin Chargaff portant sur la composition en bases d'ADN provenant de différentes sources. A l'époque, il n'était pas question d'obtenir la séquence de ce polymère, mais il était possible, après hydrolyse complète (c'est à dire rupture de toutes les liaisons covalentes unissant les monomères entre eux), de séparer ceux-ci par chromatographie sur papier. On sépare ainsi quatre constituants : "A", "T", "G" et "C", que l'on peut doser afin d'évaluer leurs proportions respectives. Le tableau 1 ci-dessous, présente des caractéristiques que Chargaff a su interpréter : les purines (A et G) et les pyrimidines (C et T) sont également représentées (50% de chaque), quelque soit la source de l'ADN, la proportion de A est la même que celle de T et la même chose pour G et C, par contre, le rapport (A + T) / (G + C) semble caractéristique de la source d'ADN (Tableau 1).

Origine de l'ADN	Bases (%)				(A + T)/(G + C)	(A + G)/(T + C)
	A	G	C	T		
Bactériophage T7	26,0	23,8	23,6	26,6	1,11	0,99
<i>Escherichia coli</i> B	23,8	26,8	26,6	23,1	0,88	1,01
Neurospora	23,0	27,1	26,6	23,3	0,86	1,00
Drosophile	30,7	19,6	20,2	29,5	1,51	1,01
Saumon	28,0	22,0	20,0	27,8	1,33	1,05
Poule	28,0	22,0	21,6	28,4	1,29	1,00
Rat	28,6	21,4	21,6	28,4	1,33	1,00
Vache	27,3	22,5	22,5	27,7	1,26	0,99
Homme	29,3	20,7	20,0	30,0	1,46	1,00

Tableau 1. Composition en bases de l'ADN de différentes origines

1.4.2. Structure des acides nucléiques

Les acides nucléiques sont des polymères de nucléotides de très grande taille (macromolécules). L'enchaînement des nucléotides conduit à des molécules de très grandes tailles. Celles d'ADN se mesurent en centimètres pour la longueur avec un diamètre pour la double hélice de 20 Angström. En pratique et étant donné les moyens mécaniques mis en jeu dans les différentes étapes de purification de l'ADN, il est impossible d'obtenir des molécules intactes, on ne travaille qu'avec des morceaux. De plus, les cassures étant aléatoires, on travaille avec un mélange hétérogène de morceaux, c'est ce qui, pendant longtemps, a bloqué la progression de l'analyse fine du gène.

a. Les monomères et leurs enchainements

Le nucléotide est lui même composé de trois molécules (Figure 4):

- un **pentose** sous forme cyclique (furane): le ribose en ce qui concerne l'acide ribonucléique (ARN) et le 2'-désoxyribose pour l'acide désoxyribonucléique (ADN);
- une **base organique** est reliée au sucre en 1' : soit une base purique adénine (A) ou guanine (G) soit une base pyrimidique : cytosine (C), thymine (T) ou uracile (U), l'ensemble constitue un nucléoside;
- un **acide phosphorique** vient estérifier une fonction alcool du sucre en 3' pour constituer un nucléotide.

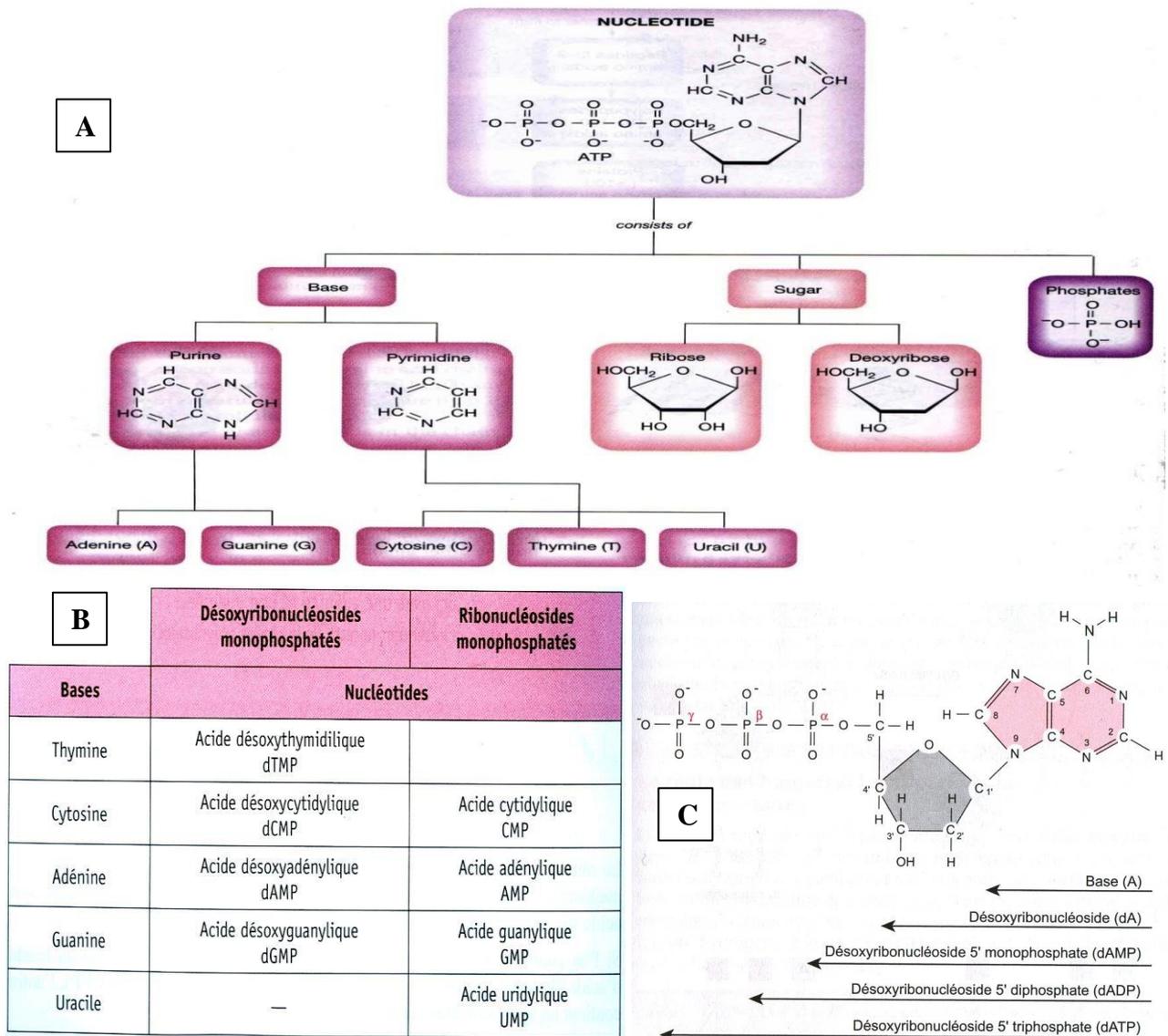
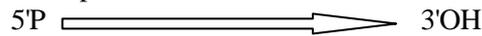


Figure 4. A) Structure des nucléotides; B) nomenclature des dérivés monophosphatés des acides nucléiques; C) structure détaillée d'un désoxyribonucléotide triphosphate

Les nucléotides précurseurs de la synthèse d'acides nucléiques sont sous forme triphosphate, deux phosphoryles (β et γ) seront éliminés au cours de la polymérisation (Figure 4). La polymérisation s'effectue par l'estérification d'un alcool situé en 3' d'un nucléotide par le phosphate d'un autre nucléotide (la liaison covalente ainsi établie est dite liaison 3'-5' **phosphodiester**). Pour un acide nucléique donné, seules les bases distinguent les différents nucléotides. Les deux acides nucléiques sont des polymères à quatre monomères possibles symbolisés par A, T, G, et C pour l'ADN et A, U, G, et C pour l'ARN.

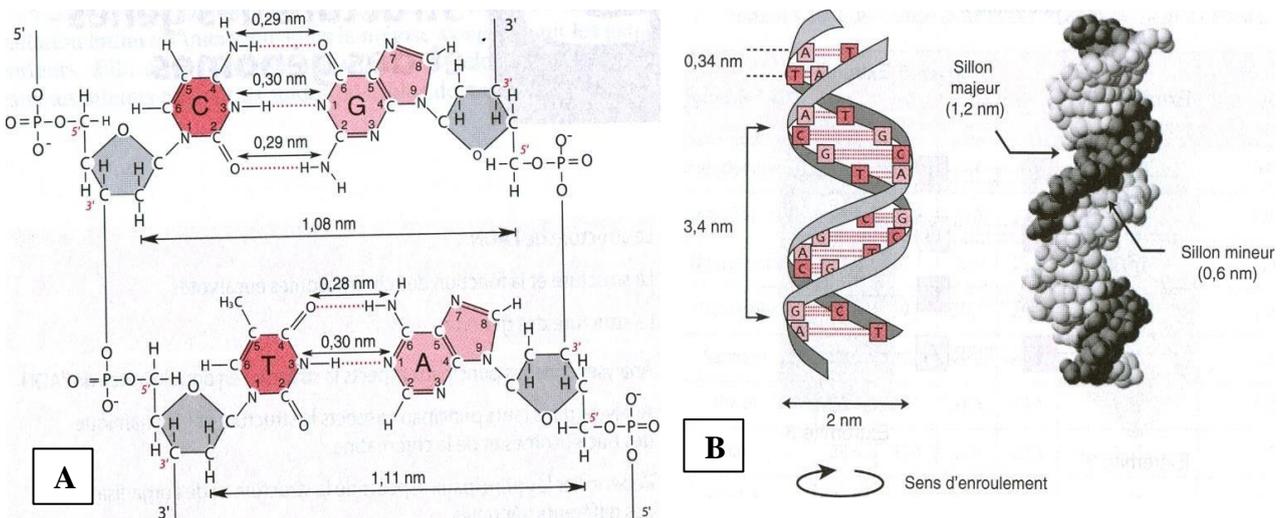
Cette succession de nucléotides nous semble aléatoire mais la séquence des monomères de tout acide nucléique est capitale pour la transmission et l'expression du matériel génétique. La séquence des nucléotides

dans le polymère impose ce que l'on appelle la structure primaire des polynucléotides. Le plus souvent, les molécules sont linéaires et possèdent donc une extrémité 5'phosphorylée et une extrémité 3'hydroxylée et qui sont orientées, d'où la représentation simplifiée suivante :



b. Structures secondaires

La structure secondaire des acides nucléiques est imposée par l'appariement des bases c'est à dire la formation de liaisons hydrogène entre deux bases organiques (Figure 5). L'ADN est constitué de deux chaînes polynucléotidiques ainsi appariées. Etant donné que, pour des raisons d'encombrement stérique, deux paires de bases seulement sont stables A..T et G..C, la séquence d'un brin de cette molécule **bicaténaire** implique celle de l'autre. Cette notion de complémentarité est fondamentale pour tout ce qui concerne le métabolisme de l'ADN in vivo et in vitro. De plus, les deux bases d'une paire ne sont suffisamment proches l'une de l'autre et ne sont capables d'établir des ponts hydrogène que si les deux chaînes nucléotidiques sont antiparallèles (Figure 5). Il en résulte, dans l'espace, une double hélice dont le pas de 3,4nm correspondant à l'empilement de 10 paires de bases. Les paires de bases sont dans un plan presque exactement perpendiculaire à l'axe de l'hélice. Les brins antiparallèles délimitent deux sillons de taille inégale, appelés sillons majeur et mineur. Il existe 3 liaisons hydrogène entre G et C et deux seulement entre A et T, ceci implique qu'un ADN riche en G-C sera plus stable qu'un ADN comportant plus de couples A-T.



I.4.3. Formes de la double hélice d'ADN

L'acide désoxyribonucléique (ADN ou DNA) est constitué de deux chaînes de nucléosides unies entre elles par des liaisons hydrogènes pour former un hybride en forme de double hélice (Figure 2, modèle de Watson et Crick) dont les deux brins sont :

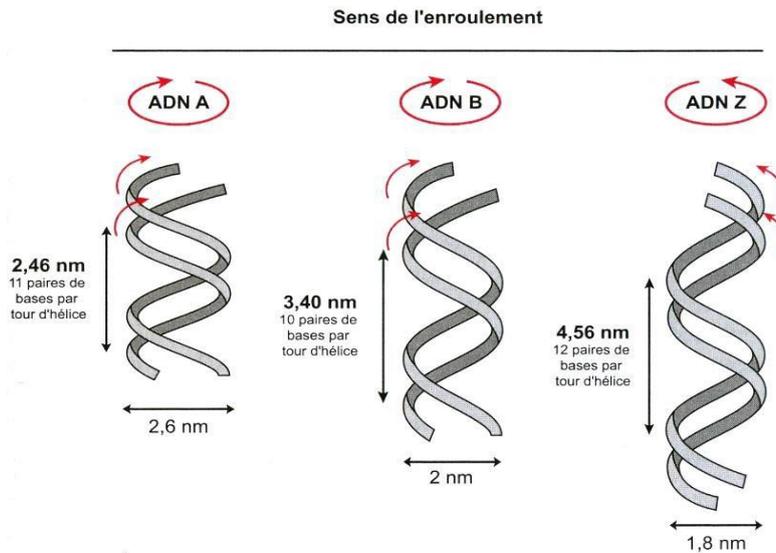
- * **antiparallèles** : l'un est constitué d'un enchaînement commençant à gauche et se poursuivant vers la droite, l'autre commençant à droite et se poursuivant vers la gauche ;

- * **complémentaires** : si l'on connaît la séquence d'un brin, on peut déduire celle du brin complémentaire.

- * **hélicoïdales** : L'ADN s'enroule suivant une double hélice à pas droit (forme **dextre**) sous les formes A et B, et sous la forme d'une double hélice à pas gauche (forme **senestre**) pour l'ADN-Z. (Figure 6). En effet, les premières formes ADN-A et ADN-B, diffèrent, entre autres, par la taille de leurs hélices. L'ADN-A a un diamètre plus large (2,6 nm) que l'ADN-B (2 nm). Il est plus compact avec une distance de 0,23 nm entre les deux plateaux de bases successifs (11 plateaux par tour). La forme A est observée notamment dans les hélices hybrides ADN-ARN. En solution et dans la cellule, c'est la forme B de l'ADN qui est prédominante. Une forme senestre de l'ADN tirée en zigzag (ADN-Z), avec un diamètre de

1,8 nm, est parfois observée mais dont la signification physiologique reste incertaine.

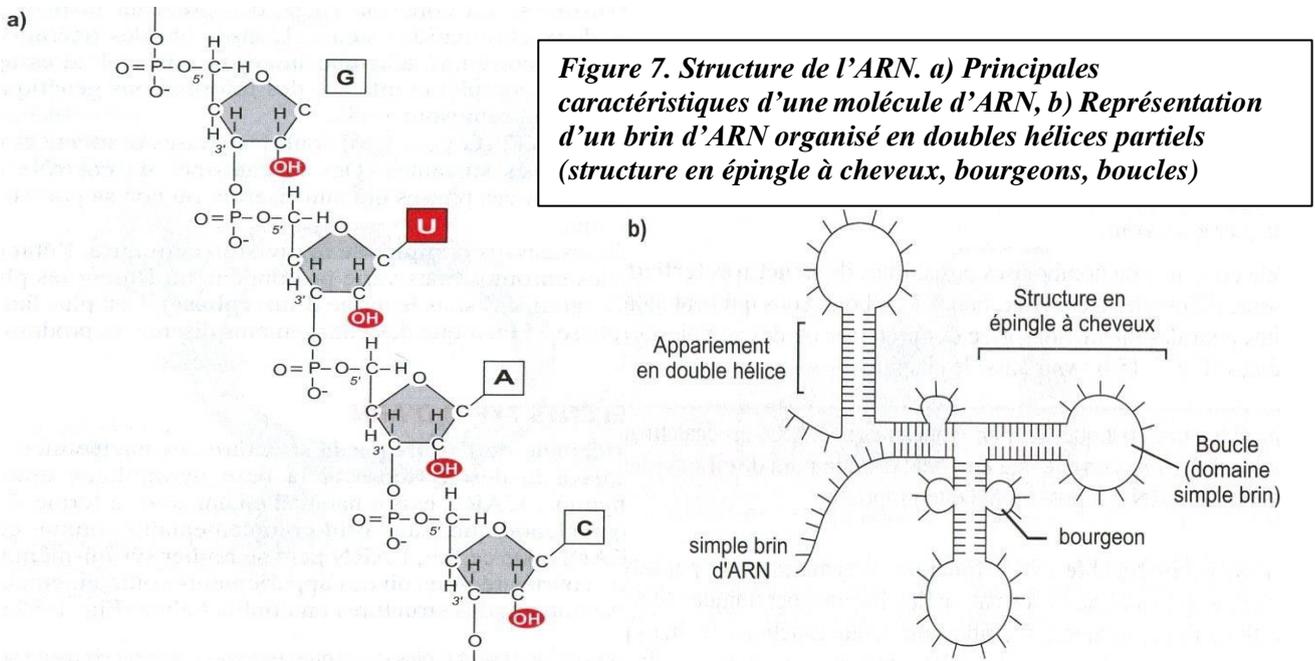
Figure 6. Représentation simplifiée des formes A, B et Z de l'ADN



Enfin, parmi les structures particulières, il faut signaler celle des molécules circulaires : il n'y a pas d'extrémité 5' ni 3' (une liaison phosphodiester les relie). De telles structures (double ou simple brin), sont fréquentes chez les virus et des éléments génétiques particuliers tels que les plasmides, le "chromosome" bactérien est également une molécule circulaire.

I.4.4. Formes et types d'ARN

L'ARN quant à lui reste souvent sous une forme simple brin et adopte souvent, localement, des structures secondaires, selon le même principe d'appariement des bases que l'ADN (Figure 7). Un brin d'ARN peut s'organiser en doubles hélices partiels (structure en épingle à cheveux, bourgeons, boucles).



L'une des caractéristiques les plus étonnantes de l'ARN est sa capacité à agir comme enzyme, d'où le nom de **ribozymes** donnés aux ARN qui possèdent cette propriété. Le tableau suivant (Tableau 2) représente les différentes classes d'ARN dans la cellule.

Classe	Les différentes classes d'ARN		
	Nom	Abrév.	Fonctions
ARN codants	ARN messagers	ARNm	dirigent la synthèse des protéines.
ARN non codants	ARN ribosomiques	ARNr	participent à la structure du ribosome et à la synthèse des protéines. Action de ribozyme (formation de la liaison peptidique).
	ARN de transfert	ARNt	adaptation des amino-acides dans la synthèse des protéines.
	Petits ARN nucléaires	ARN sn	rôle dans l'épissage des pré-ARNm.
	Petits ARN nucléolaires	ARN sno	modification des ARN ribosomiques.
	Petits ARN interférents	ARN si	actifs dans l'ARN interférence.
	Micro-ARN	ARN mi	régulent la stabilité des transcrits ARN et leur traductibilité.

Tableau 2. Les différentes classes d'ARN dans la cellule

1.5. Propriétés topologiques fondamentales de l'ADN (Figure 8)

1.5.1. Dénaturation de l'ADN et température de fusion (T_m)

Malgré sa stabilité grâce aux liaisons hydrogènes relativement fragiles, la structure de la double hélice peut être détruite, en particulier à haute température ($> 80^\circ\text{C}$) ou en milieu très alcalin ($\text{pH} > 11$), sans pour autant casser les brins. Pour l'ADN, le résultat est la séparation complète des deux brins qui le composent : il y a **dénaturation** de la molécule. En raison de la stricte complémentarité des bases, la dénaturation est réversible, les deux brins peuvent, dans des conditions appropriées de température et de force ionique, rétablir des liaisons hydrogènes entre leurs bases et reprendre la configuration en double hélice d'origine. Ce phénomène ne dépend que de la complémentarité de deux séquences nucléotidiques et non de l'origine de chaque brin. Dans le cas de renaturation, on a formation de molécules "**homoduplex**" c'est à dire de structures doubles, chaque partenaire ayant une même origine. On peut très bien mélanger des ADN dénaturés d'origines différentes, s'il y a des complémentarités de séquences, des "**hétéroduplex**" c'est à dire des molécules dont les deux brins sont d'origine différente pourront se former. Enfin, toujours par complémentarité de séquences, on peut appairer un ARN avec un brin d'ADN (il existe une possibilité de liaison A..U) on obtient alors une molécule "hybride" (ADN-ARN).

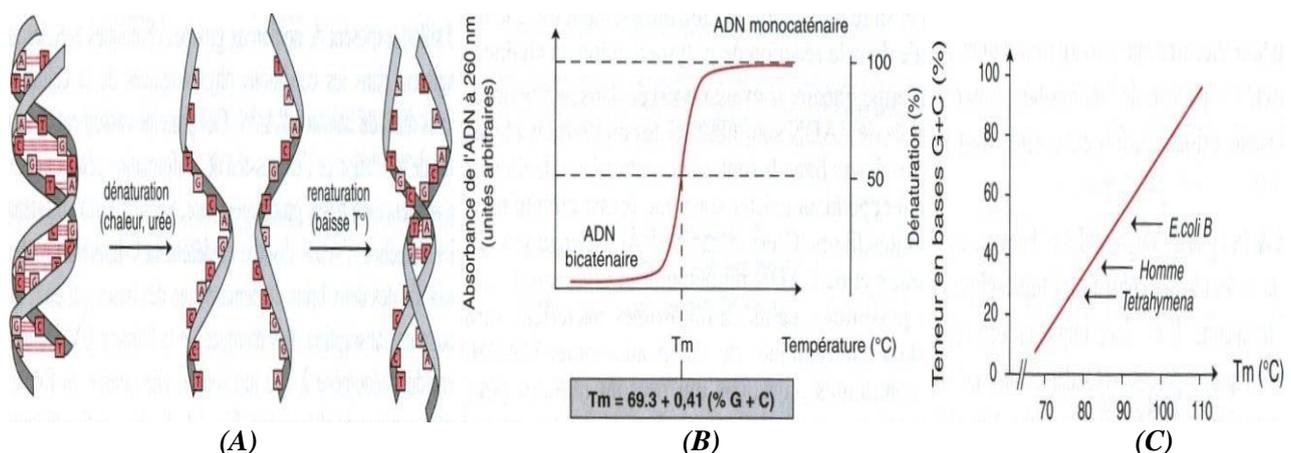


Figure 8. Propriétés topologiques de l'ADN. La dénaturation de l'ADN (A) peut être suivie par l'étude de l'absorption des U.V. (Densité Optique, DO): l'ADN absorbe spécifiquement à une longueur d'onde de 260 nm. En (B), mesure de la D.O. à 260 nm d'un échantillon d'ADN en fonction de la température donne une courbe sigmoïdale dont le point d'inflexion est T_m (température de fusion = melting). En (C), la T_m dépend de la teneur en G/C et varie d'un organisme à l'autre.

Cette propriété (appariement de séquences complémentaires) représente un outil puissant dans la reconnaissance de séquences précises à l'aide de petits fragments polynucléotidiques appelés "**sondes**" car, dans un ensemble complexe de séquences différentes, elles sont capables de trouver celle qui leur correspond,

de s'y hybrider. La détection de ces hybrides sera facilitée par l'utilisation de sondes "marquées" (Voir chapitre II).

La température de fusion est celle pour laquelle la moitié de l'ADN est dénaturé. Il est possible de mesurer directement la température de fusion d'un ADN double brin en mesurant l'augmentation de l'absorbance de la solution à 260 nm en fonction de la température. Il est possible d'évaluer la valeur de T_m grâce à l'application d'une formule qui dépend de la taille du fragment d'ADN étudié :

- Si celui-ci a une longueur égale ou inférieure à 20 nucléotides (nt) on compte 2°C par couple A/T et 4°C par couple G/C : $T_m = (nA+nT) \times 2 + (nG+nC) \times 4$

- A partir de N supérieur à 20, on corrige d'un multiplicateur proportionnel à la longueur au delà de ce chiffre : $T_m = ((nA+nT) \times 2 + (nG+nC) \times 4) \times (1 + [(N-20)/20])$
où N est le nombre total des nucléotides.

Pour être plus précis il faut tenir compte aussi de la concentration en sodium du tampon d'hybridation. Lorsque cette concentration n'excède pas 1M, on utilise la formule :

$$T_m = 81,5 + 16,6(\log_{10}[\text{Na}^+]) + 0,41[(G+C)/N] - (600/N)$$

Lorsqu'on travaille en présence de la formamide, il faut encore diminuer la T_m en fonction de la concentration de l'amide :

$$T'_m = T_m - 0,6(\% \text{ formamide})$$

I.5.2. Dosage et étude de l'ADN

La plupart des molécules biologiques contenant des noyaux aromatiques absorbent les rayons ultraviolets à différentes longueurs d'onde. Dans l'ultraviolet de très courte longueur d'onde presque toutes les solutions de molécules biologiques sont opaques. Dans la région d'émission des lampes à vapeur de mercure (254 nm) les bases puriques et pyrimidiques ont des pics d'absorption spécifiques qu'on utilise pour comparer avec les spectres d'absorption des acides aminés aromatiques dans la même région.

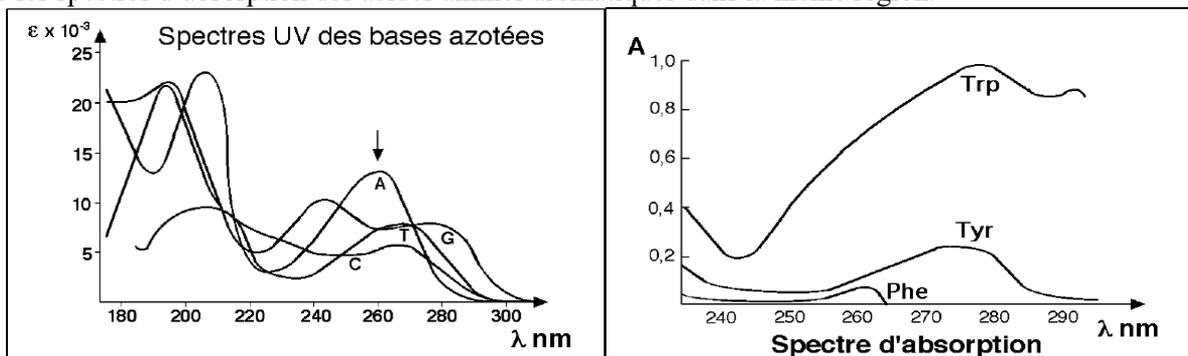


Figure 9. Spectres UV des noyaux aromatiques des bases nucléiques et des acides aminés

Les zones d'absorption des quatre bases s'étalent de 240 à 280 nm de sorte que les acides nucléiques formés de ces quatre types de nucléotides ont un maximum d'absorption à 260 nm. Les radicaux aromatiques de certains acides aminés (Phe, Tyr, et surtout Trp) ont la propriété d'absorber la lumière ultraviolette (Figure 9). L'absorption à 280 nm est due principalement aux noyaux phénols des tyrosines, parce que cet acide aminé est plus fréquent que le tryptophane, qui est pourtant beaucoup plus opaque à cette longueur d'onde. L'absorption de la lumière UV à 280 nm est caractéristique des protéines et sert à doser ces protéines lorsqu'elles sont en solution dans l'eau et qu'il n'existe pas d'autres molécules absorbant la lumière UV à

1 mole de nucléotides = 309 mg/mL	
Standard : 1 cm, 1 M, θ ambiante, $\text{Na}^+ = 0,3 \text{ M}$, $A_{260}/A_{280} \geq 1,80$	
dsDNA	$\epsilon = 6200 A_{260} \cdot \text{M}^{-1} \cdot \text{cm}^{-1}$ $1 A_{260} = 309/6200 = 50 \mu\text{g} \cdot \text{mL}^{-1}$
ssDNA (hyperchromicité)	$\epsilon = 8350 A_{260} \cdot \text{M}^{-1} \cdot \text{cm}^{-1}$ $1 A_{260} = 309/8350 = 37 \mu\text{g} \cdot \text{mL}^{-1}$
RNA	$\epsilon = 7700 A_{260} \cdot \text{M}^{-1} \cdot \text{cm}^{-1}$ $1 A_{260} = 309/7700 = 40 \mu\text{g} \cdot \text{mL}^{-1}$

cette longueur d'onde (acides nucléiques par exemple). Pour effectuer par spectrophotométrie directe le **dosage** d'une solution pure d'acides nucléiques, on utilise l'absorbance des rayons ultraviolets à 260 nm comme suit :

I.6. ADN et chromosomes (Figure 10)

La quantité importante d'ADN stockée dans la cellule (1.4 m dans un noyau de quelques microns de diamètre) nécessite des techniques d'empaquetage sophistiquées, capables de le stocker de façon compacte tout en lui permettant de se répliquer et d'être partagé entre cellules filles lors de la division cellulaire. La structure de stockage, trouvée chez tous les eucaryotes est le **chromosome**. Les chromosomes sont des structures relativement lâches pendant la plus grande partie du cycle cellulaire, permettant à l'ADN d'être accessible pour la transcription et la réplication. Pendant la division cellulaire, ils se condensent pour permettre la distribution correcte de l'ADN dans les cellules filles, et deviennent donc bien visibles.

L'ADN est d'abord empaqueté dans des **nucléosomes** composés de protéines abondantes, les **histones**, qui contiennent environ 100 paires de nucléotides d'ADN chacun dans une petite particule plus ou moins sphérique. Les nucléosomes sont ensuite arrangés de façon hélicoïdale dans une fibre de 30 nm, qui elle-même s'organise en boucles ancrées sur un échafaudage protéique. Il semble que ces boucles définissent des régions dont l'expression est contrôlée de façon coordonnée. On peut visualiser assez facilement en microscopie optique ou électronique des chromosomes en métaphase (lors de la division cellulaire), mais ils sont impossibles à voir clairement dans la plupart des cellules en interphase (sauf quelques cas exceptionnels). Néanmoins, la coloration spécifique de régions précises du chromosome permet de constater qu'une organisation spatiale persiste dans toutes les phases du cycle cellulaire.

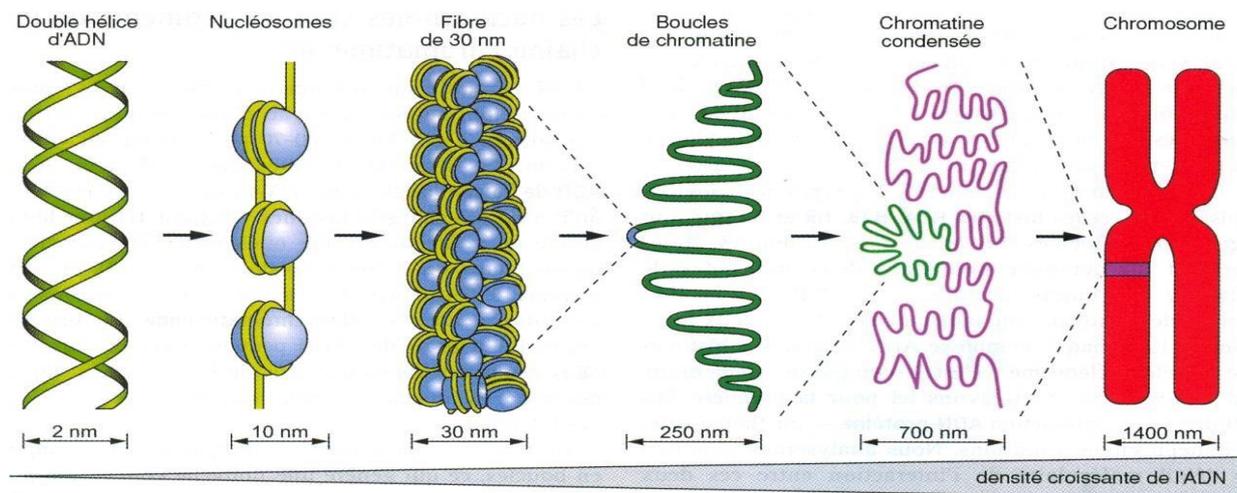


Figure 10. Niveaux d'organisation de l'ADN et de ses protéines associées pour former un chromosome.

Chapitre II. OUTILS ET TECHNIQUES DE LA BIOLOGIE MOLECULAIRE

Il existe toute une série d'outils indispensables aux biologistes moléculaires pour analyser l'ADN, le cloner, l'amplifier, pour le transporter grâce à des vecteurs de clonage, l'intégrer dans des cellules hôtes... Ces outils ont permis de mettre en place des techniques qui sont à l'origine du développement de la biologie moléculaire.

II.1. La création de molécules d'ADN recombinant

II.1.1. Clivage de l'ADN par les enzymes de restriction

Les enzymes de restriction sont des **endonucléases** et sont produits par des bactéries et participent à un mécanisme de défense des bactéries vis-à-vis des virus : système de **restriction- méthylation**. En effet les bactéries sont lysées sous l'effet de virus bactériophages dont l'ADN est répliqué par la bactérie elle-même avant sa destruction. Pour détruire l'ADN du parasite, la bactérie exprime des gènes de restriction et de méthylation. Les gènes de restriction permettent la synthèse d'endonucléases coupant l'ADN en des sites très spécifiques. Afin de protéger l'ADN bactérien de l'hydrolyse par l'enzyme, une **méthylase**, codée par le gène de méthylation, va modifier les nucléotides de l'ADN bactérien en les méthylant pour qu'ils ne soient plus reconnus par l'enzyme de restriction. L'ensemble du gène de restriction et du gène de méthylation constitue un système de défense de la Bactérie vis-à-vis des phages.

Les enzymes de restriction catalysent la coupure de l'ADN non méthylé en des endroits caractérisés par une séquence spécifique de nucléotides (site de restriction = séquence **palindromique** de 4, 6 ou 8 bases dans une orientation anti-parallèle, Tableau 1). Les produits de cette digestion sont les **fragments de restriction**, dont la longueur, toujours la même pour un ADN donné, ne dépend que de la séquence primaire de cet ADN. Les extrémités des fragments de restriction peuvent être formées de deux brins d'égale longueur (**bouts francs**) ou bien présenter un brin plus long que l'autre de quelques nucléotides (**bouts collants**).

Enz.	Organisme d'origine	Site de restriction	Nature des extrémités	Taille du site (pb)
<i>Hae</i> III	<i>Haemophilus aegyptius</i>	5'-GG↓CC-3' 3'-CC↑GG-5'	Bouts francs	4
<i>Taq</i> I	<i>Thermus aquaticus</i>	5'-T↓CGA-3' 3'-AGC↑T-5'	Bouts cohésifs 5' monocaténaire	4
<i>Hha</i> I	<i>Haemophilus haemolyticus</i>	5'-GCG↓C-3' 3'-C↑GCG-5'	Bouts cohésifs 3' monocaténaire	4
<i>Eco</i> RV	<i>Escherichia coli</i>	5'-GAT↓ATC-3' 3'-CTA↑TAG-5'	Bouts francs	6
<i>Bam</i> HI	<i>Bacillus amyloliquefaciens</i>	5'-G↓GATCC-3' 3'-CCTAG↑G-5'	Bouts cohésifs 5' monocaténaire	6
<i>Pst</i> I	<i>Providencia stuartii</i>	5'-CTGCA↓G-3' 3'-G↑ACGTC-5'	Bouts cohésifs 3' monocaténaire	6
<i>Eco</i> RII	<i>Escherichia coli</i>	5'-G↓CCTGGC-3' 3'-CGGACC↑G-5'	Bouts cohésifs 5' monocaténaire	7
<i>Not</i> I	<i>Nocardia otiditis cavarium</i>	5'-GC↓GGCCGC-3' 3'-CGCCGG↑CG-5'	Bouts cohésifs 5' monocaténaire	8

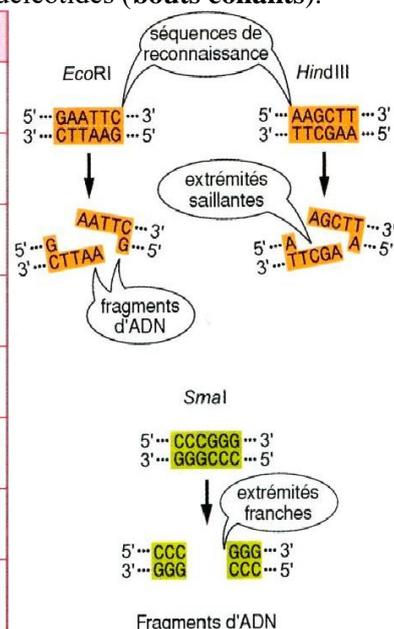


Tableau 1. Sites de reconnaissance et de clivage de quelques enzymes de restriction

Les enzymes de restriction sont des hydrolases (classe 3 de la E.C.) agissant sur des liaisons esters (sous-classe 3.1) et coupent les liaisons phosphodiester 3'-5'. Parmi les estérases on distingue celles qui hydrolysent un acide nucléique en fragments polynucléotidiques (endonucléases) et en particulier celles dont les produits gardent leur phosphate 5' initial (sous-sous-classe 3.1.21). Les enzymes de restriction nécessitent la présence de l'ion Mg^{++} dans le milieu. Le nom de chaque enzyme est dérivé du nom d'espèce ou de variété de la bactérie qui la produit. On écrit l'initiale du nom du genre, les deux initiales du nom de l'espèce, une lettre ou un nombre pour désigner la variété (ou souche) et après un espace un chiffre romain pour désigner successivement les différentes enzymes de restriction obtenues à partir de la même souche (Figure 1).

Différents tampons sont utilisés pour les enzymes de restriction permettant l'optimisation des réactions de multiples enzymes avec le même tampon. Toutes les enzymes de restriction nécessitent la présence de l'ion Mg^{++} . Le pH (7,0-7,9), le potentiel d'oxydo-réduction (dithiothréitol 1 mM), la force ionique (NaCl ou CH_3COOK de 0 à 100 mM) varient d'un tampon à l'autre (Figure 1). De rares enzymes nécessitent des conditions spéciales précisées par leur mode d'emploi.

<p>3 = Hydrolase 3.1 = Esterase 3.1.21 = Endonucléase produisant un 5'-phosphate 3.1.21.4 = Enzyme avec Mg (seul cofacteur)</p> <p style="text-align: center;">ECOR I</p> <p>Genre : <i>Escherichia</i> Espèce : <i>coli</i> Variété : <i>R</i></p> <p style="text-align: right;">Numéro d'ordre (s'il y a plusieurs enzymes)</p>	<p>10 mM Tris Propane-HCl (pH 7,0 à 25°C) 10 mM $MgCl_2$ (1 mM DTT)</p>	<p>10 mM Tris-HCl (pH 7,9 à 25°C) 10 mM $MgCl_2$ 50 mM NaCl (1 mM DTT)</p>
	<p>50 mM Tris-HCl (pH 7,9 à 25°C) 10 mM $MgCl_2$ 100 mM NaCl (1 mM DTT)</p>	<p>20 mM Tris-CH_2COOH (pH 7,9 à 25°C) 10 mM $Mg(CH_2COO)_2$ 50 mM CH_2COOK (1 mM DTT)</p>

Figure 1. Nomenclature des enzymes de restriction et composition des tampons de digestion

L'analyse de la longueur des fragments de restriction (**carte de restriction**), à la recherche de variations individuelles (polymorphismes de longueur des fragments de restriction = **RFLP**) est une des techniques d'analyse de la séquence primaire de l'ADN à la recherche de substitutions, d'insertions ou de délétions qui modifient le nombre de sites de restriction et donc la longueur des fragments de restriction.

II.1.2. Ligature de l'ADN

Les extrémités d'ADN générées par un clivage enzymatique peuvent spontanément se réassocier. Ceci est plus facile avec les extrémités qui sont débordantes car elles offrent des séquences monocaténaires complémentaires pouvant s'hybrider par des liaisons hydrogène (Figure 2).

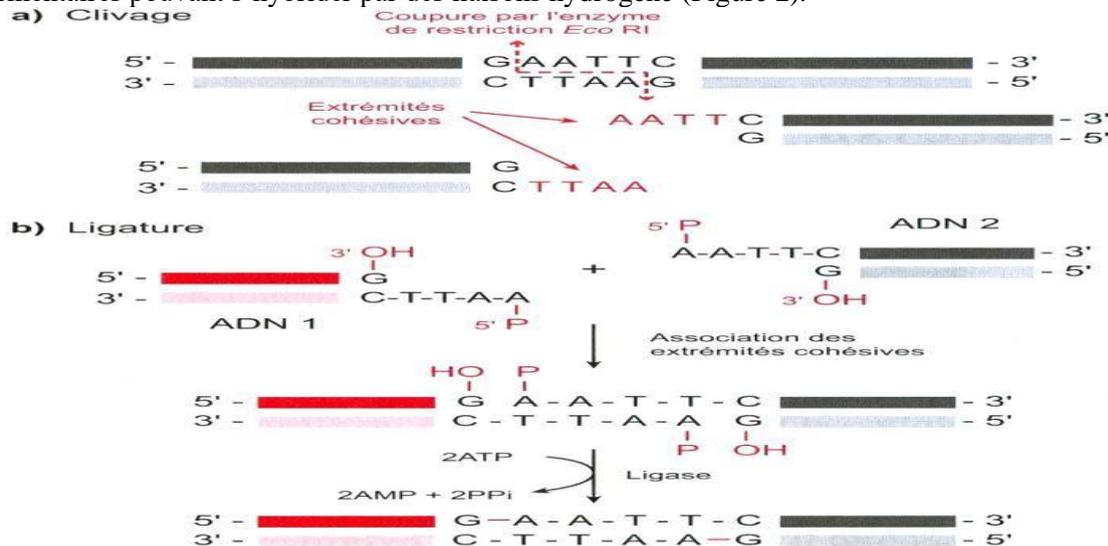


Figure 2. Les extrémités cohésives produites par clivage enzymatique (a) peuvent s'associer puis l'ADN ligase établit des liaisons covalentes entre les brins adjacents (b)

Les clivages par des enzymes de restriction différentes peuvent parfois fournir les mêmes extrémités débordantes identiques qui peuvent s'associer par hybridation entre les bases complémentaires (Figure 3). Cette association instable peut être rendue permanente par une réaction de ligature grâce à l'ADN ligase, qui lie de manière covalente les fragments d'ADN en créant des liaisons phosphodiester 3'-5' au niveau des jonctions cohésives ou francs.

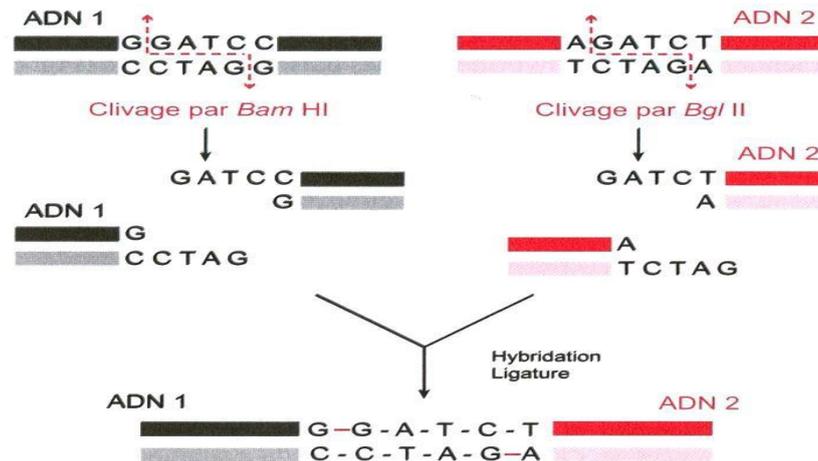


Figure 3. Ligation de deux fragments issus de clivage par deux enzymes différentes (Bam HI et Bgl II). Malgré que les séquences palindromiques de 6 nucléotides diffèrent, elles possèdent la même centrale d'où la complémentarité entre les extrémités 3' et 5' générées

Dans certains cas, de courtes séquences, appelés **adaptateurs**, possédant un site de restriction sont ajoutées à chaque extrémité franche du fragment d'ADN (Figure 4). Cela apporte cependant des séquences supplémentaires qui peuvent affecter la fonction de la molécule recombinante lorsqu'elle est introduite dans une cellule. Pour cette raison, la réalisation de molécules d'ADN recombinant s'effectue préférentiellement à l'aide d'extrémités franches ou cohésives générées par les enzymes de restriction.

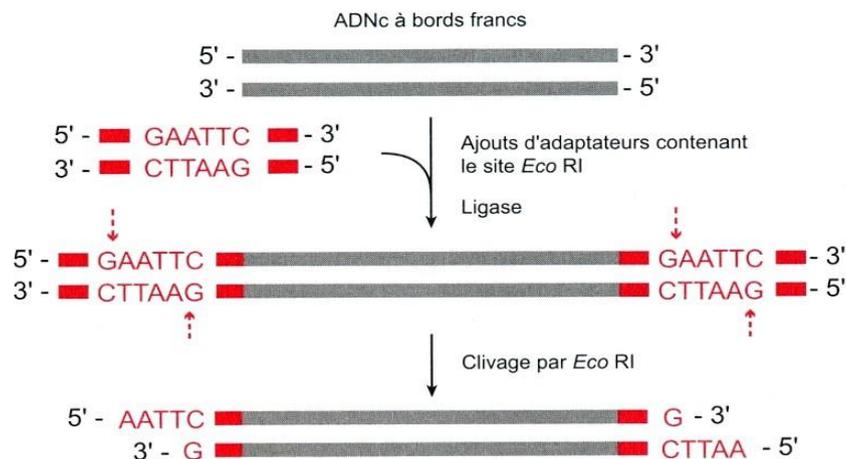


Figure 4. Addition d'adaptateurs à des fragments d'ADN à bords francs et production d'un ADN à bords cohésifs

II.2. Les vecteurs de clonages

Un vecteur est un transporteur capable de conduire une molécule à pénétrer dans une cellule alors qu'elle n'est pas captée spontanément par cette cellule. Les vecteurs sont principalement étudiés pour transporter des médicaments dans un type déterminé de cellules de l'organisme d'un malade, tout en mettant ce médicament à l'abri des enzymes de détoxification du sujet. En biologie moléculaire, on obtient l'**ADN recombinant** en insérant un fragment d'ADN étranger dans un **vecteur de clonage**. Ces vecteurs sont étudiés pour faire entrer un acide nucléique dans une cellule et permettre la réplication ou l'expression de cet acide nucléique dans la cellule qui le reçoit. Les plasmides bactériens, les bactériophages et de nombreux virus sont des vecteurs naturels permettant la **transfection** de différents types de cellules-hôtes. D'autres vecteurs encore plus performants ont été obtenus artificiellement à partir de ceux-là grâce à des manipulations génétiques comme les **cosmides**.

Les cellules qui sont habituellement transfectées par des vecteurs contenant des fragments d'ADN, sont destinées à permettre d'**amplifier** ce vecteur en même temps que leur croissance. Des bactéries comme *Escherichia coli* sont utilisées pour recevoir des plasmides. Certaines souches de cette espèce sont plus particulièrement développées pour permettre la transfection de certains vecteurs : souche DH5α pour les plasmides pUC18 ou pUC19, HB 101 pour le plasmide pBR322, souche JM 109 dépourvue d'opéron lactose

pour les vecteurs à β -galactosidase. Des levures comme *Saccharomyces cerevisiae* sont utilisées pour des vecteurs eucaryotiques. Les cellules animales en culture du hamster doré : CHO (*Chinese hamster ovaries*) ou BHK (*baby hamster kidneys*) ou de la souris : lignée 3T3 peuvent recevoir des vecteurs spécifiques.

En règle générale, les vecteurs sont de petite taille, et se répliquent de façon efficace en produisant de nombreuses copies du fragment d'ADN inséré. Il existe de nombreux vecteurs de clonage, et le choix s'effectue selon la taille du fragment à cloner et la cellule dans laquelle il sera transféré (Tableau 2).

Tableau 2. Vecteurs de clonage chez *Escherichia coli*; kb : kilobases

Type de vecteur	Taille maximum du fragment d'ADN cloné
Plasmide	10 kb
Phage lambda (λ)	15 kb
Cosmide	45 kb
BAC (« Bacterial Artificial Chromosome »)	300 kb

II.2.1. Les plasmides

Les plasmides sont des molécules d'ADN circulaire, extrachromosomiques, présents chez la plupart des procaryotes et chez quelques eucaryotes. Ils se multiplient de façon indépendante au sein des cellules-hôtes et sont transmis de manière régulière au fur et à mesure des divisions. Chaque plasmide contrôle de façon autonome son propre nombre de copies (variable entre 1 à 60 copies) c'est-à-dire le nombre de répliques dans la cellule-hôte. La taille des plasmides varie entre 5 et 400 kilo paires de bases (kpb).

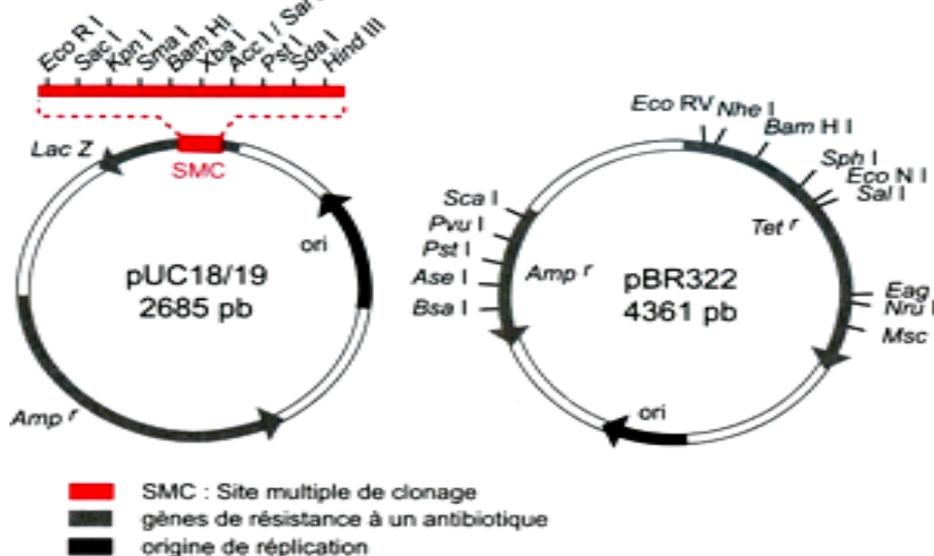


Figure 5. Principales caractéristiques des plasmides bactériens. Le gène de résistance à l'ampicilline (Amp^r) permet de sélectionner les bactéries transformées par pUC18/19. Idem pour pBR322, sauf si un site de restriction présent dans la séquence des gènes Amp^r et Tet^r (résistance à la tétracycline) est utilisé pour le clonage.

Les plasmides constituent le véhicule idéal pour introduire dans les bactéries des gènes non bactériens grâce aux techniques de clonage moléculaire. Les plasmides utilisés en génie génétique dérivent de plasmides naturels, préalablement modifiés pour faciliter leur utilisation comme outils de clonage. Un vecteur plasmidique typique comporte (Figure 5) :

- Une **origine de réplication** pour assurer la multiplication autonome du vecteur ;
- Une petite région contenant de nombreux sites uniques de restriction, appelée région à site **multiple de clonage** (SMC) ou **polylinker** qui offre diverses possibilités de clonage d'un ADN étranger sans affecter l'autoréplication du plasmide ;
- Un ou deux **marqueurs de sélection** : ce sont des gènes de résistance à des antibiotiques (Ampicilline, Streptomycine, Tétracycline,...) ;

- Dans certains vecteurs (pBR322) les gènes de résistance aux antibiotiques possèdent un répertoire important de sites de clonage. L'interruption de la séquence entraîne la perte de l'activité de la protéine conférant la résistance à l'antibiotique. La sélection des clones recombinants exploite les changements de sensibilité-résistance aux antibiotiques qui résultent du clonage.

II.2.2. Les vecteurs viraux

Les bactériophages permettent le clonage de grands fragments d'ADN car environ 1/3 de l'ADN du bactériophage peut être remplacé par un ADN étranger. De plus, comme les virus infectent les cellules avec une très grande efficacité, le gène cloné sera introduit avec une fréquence plus élevée que celle qui est observée par simple transformation de bactéries avec les plasmides. Les vecteurs phagiques utilisés pour le clonage sont dérivés du phage lambda (λ) (Figure 6). Seuls les gènes de structure du phage, regroupés aux extrémités de son génome, sont conservés. Les séquences médianes, non indispensables pour la réplication du phage, sont éliminées et remplacées par des sites de restriction pour faciliter le clonage. Aux extrémités du phage, existent des séquences appelées **sites cos**, nécessaires pour l'intégration de l'ADN du phage dans le génome de son hôte durant la phase de lysogénie. Des ADN étrangers peuvent être introduits dans les bactériophages après élimination de la partie centrale à l'aide d'une enzyme de restriction. L'ADN génomique que l'on désire ainsi cloner est également coupé par une enzyme créant des extrémités cohésives compatibles. Des fragments d'ADN jusqu'à 15 kpb seront les plus facilement empaquetés. Après une encapsidation in vitro, et infection des bactéries, des plages de lyse apparaissent sur le tapis bactérien et correspondent à des **phages recombinants** possédant un fragment d'ADN étranger.

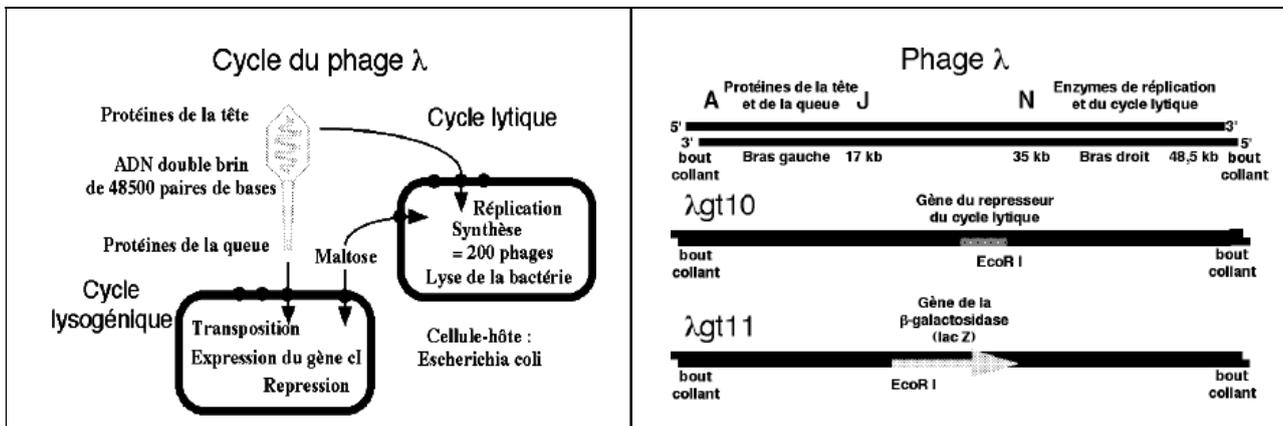


Figure 6. Cycle du bactériophage λ et structure de son génome

Le génome du bactériophage λ dans les particules virales est représenté par un ADN double brin de 48,5 kb, dont les extrémités sont cohésives entre elles (bouts collants). Le virus entre dans les cellules bactériennes par un récepteur (lamB) qui sert normalement au transport du maltose. Dès qu'il pénètre dans la bactérie, l'ADN du phage se circularise par hybridation de ses extrémités et l'ADN ligase de la bactérie achève de fermer les brèches. L'évolution se poursuit dans deux directions différentes en fonction des promoteurs activés de l'ADN du phage. La plupart du temps, l'expression des premiers gènes déclenche la réplication intense de l'ADN du phage et l'expression de protéines permettant la synthèse de particules virales. La bactérie meurt, éclate et libère environ

200 nouvelles particules virales (cycle lytique). Plus rarement, le promoteur activé au départ transcrit le gène cI , qui réprime le cycle lytique et commande la transposition de l'ADN du phage dans le chromosome de la bactérie où il reste peu exprimé (prophage), mais répliqué à chaque division de la cellule-hôte (cycle lysogénique). Une élévation de la température déclenchera à nouveau l'excision du prophage et sa réplication comme dans le cycle lytique.

Les gènes transcrits à partir de cet ADN sont désignés par des lettres de l'alphabet : de A à J du côté 5' terminal et de N à R dans la partie 3' terminale. Le génome du phage λ (λ gt10) peut être modifié en supprimant le gène du répresseur du cycle lytique dans la partie centrale, ce qui favorise la réplication du phage, et en ajoutant le gène de la β -galactosidase après le gène J (λ gt11).

II.2.3. Les cosmides

Les cosmides sont des molécules d'ADN circulaires hybrides construites à partir des phages λ et des

plasmides. Ils possèdent comme les plasmides, une origine de réplication plasmidique, un ou plusieurs marqueurs de sélection, une région à site multiple de clonage, et un site *cos* (d'où leur nom cosmides). Ils permettent le clonage de fragments d'ADN de grande taille, trois fois plus grande que la taille de ceux qui sont intégrés par le phage λ (Tableau 4). Le cosmide est pour cela ouvert par l'action d'une enzyme de restriction. L'ADN à cloner est lui aussi fragmenté par une enzyme qui fournit des extrémités compatibles. On obtient alors par ligatures successives des répétitions en tandem (concatémères) du vecteur lié à des fragments d'ADN étranger. Le système d'encapsidation *in vitro* permet d'introduire dans les têtes des phages des fragments d'ADN à cloner compris entre deux sites *cos* distants de 35 à 45 kb. Ces fragments d'ADN sont introduits dans la bactérie par infection, ils s'y circularisent et se répliquent comme les plasmides.

II.2.4. Les chromosomes artificiels

Les chromosomes artificiels bactériens également appelés 'BAC' sont construits à partir du plasmide F de *E. coli* qui est capable d'intégrer de très grands fragments d'ADN. Le vecteur possède les gènes du facteur F (*oriS*, *repE*, *parA* et *parB*) impliqués dans la réplication et la régulation du nombre de copies du plasmide (Figure 7). Utilisés comme vecteurs, ils peuvent porter des inserts d'ADN étranger d'une taille proche de 300 kpb. Les BAC sont généralement introduits dans les bactéries par électroporation.

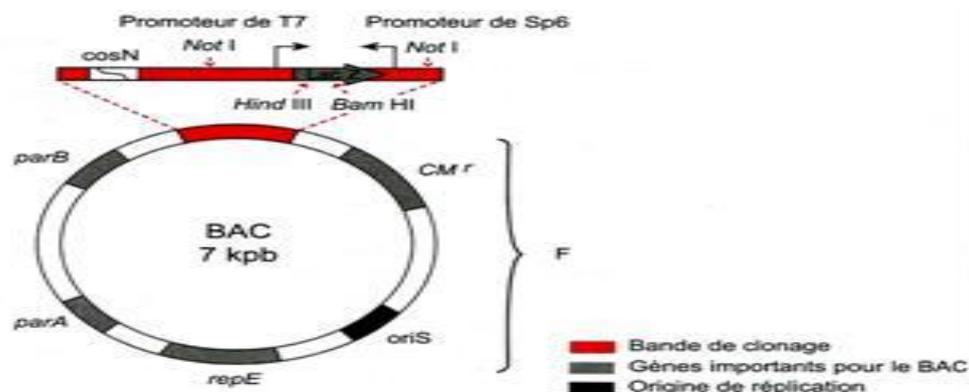


Figure 7. Principales caractéristiques d'un chromosome artificiel bactérien (BAC)

II.3. Les banques d'ADN

II.3.1. Les banques d'ADN génomiques

Une banque d'ADN génomique est constituée d'un ensemble de fragments d'ADN représentatifs d'un génome entier ou d'une fraction de génome. Les fragments sont clonés dans un très grand nombre de copies d'un même vecteur. Pour constituer la banque, l'ADN génomique est coupé soit par des enzymes de restriction à sites peu fréquents, soit à l'aide d'un dispositif provoquant un cisaillement physique libérant des fragments de taille variable. Dans les deux cas, les différents fragments ainsi obtenus sont insérés de façon aléatoire dans les vecteurs de clonage plasmidiques ou autre avant de transformer des bactéries (Figure 8). Chaque bactérie transformée par un plasmide qui contient fragment d'ADN est sélectionnée par culture sur un milieu contenant un antibiotique, elle est à l'origine d'un clone bactérien. L'ensemble de l'information génétique se trouve alors dispersé dans les différentes copies du vecteur.

Les banques génomiques les plus intéressantes sont celles qui sont constituées de fragments de taille suffisante pour contenir des gènes entiers. Aussi, il existe différents types de banques, classifiées par le vecteur utilisé et l'origine de l'ADN. Pour les petits génomes on utilisera des vecteurs plasmidiques ou phagiques; les BACs demeurent les plus employés en raison de la grande taille des fragments qu'ils peuvent intégrer (Tableau 4). Après transformation, chaque colonie obtenue correspond à des cellules contenant un des vecteurs recombinants constitutifs de la banque. On utilise la méthode de criblage pour rechercher le (ou les) clone(s) contenant la partie du génome qui nous intéresse.

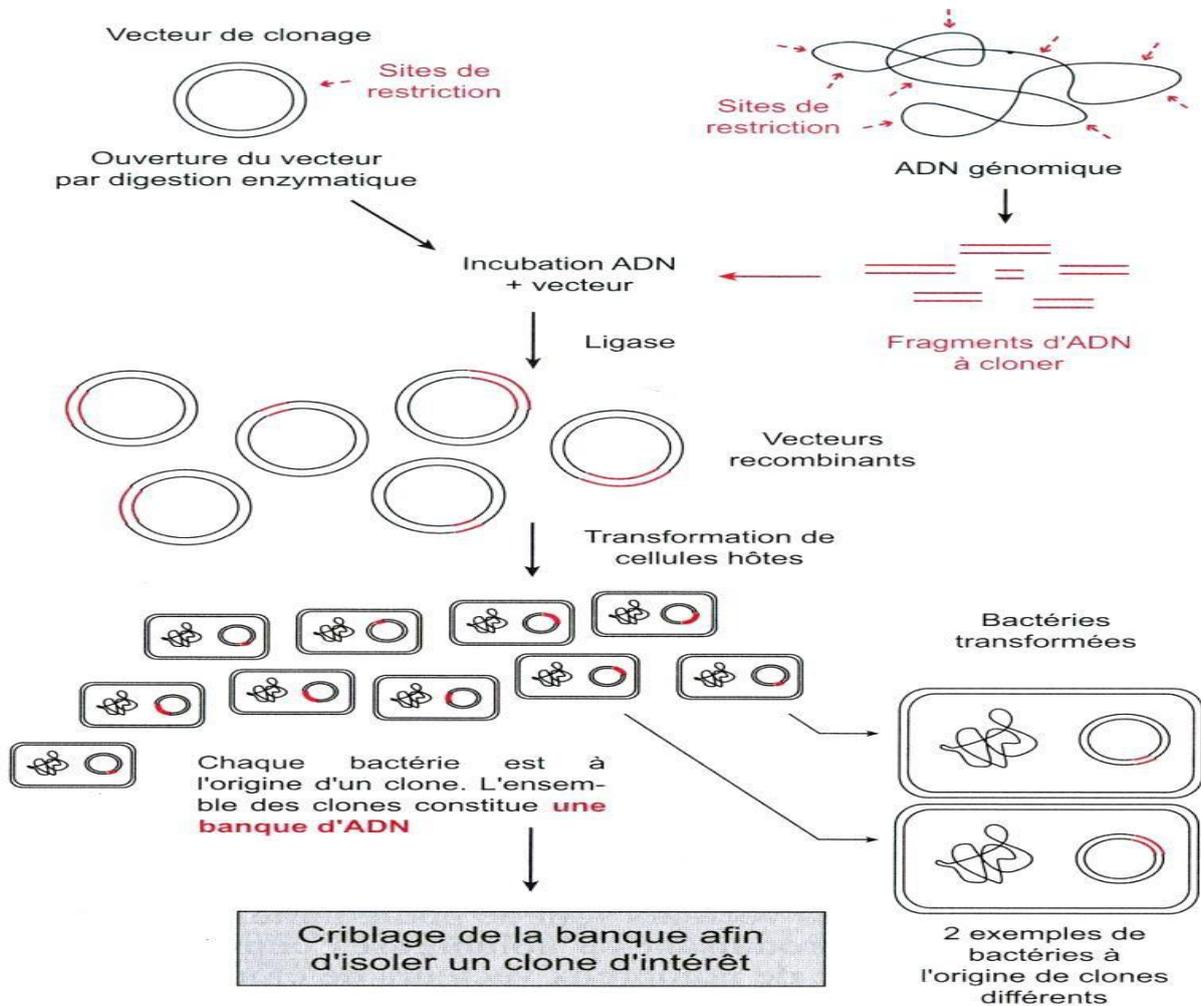


Figure 8. Les principales étapes de clonage pour construire une banque d'ADN génomique.

II.3.2. Les banques d'ADN complémentaires

La banque construite à partir d'ADN complémentaire (ADNc) constitue une banque d'ADNc. Dans ce cas, les ARNm sont le point de départ de la construction de la banque. Quand l'ARNm a une origine eucaryote, son extrémité 3'OH est généralement **polyadénylée**. Elle est alors utilisée pour séparer l'ARNm, par un piégeage à l'aide d'un support polydT, des autres ARN cellulaires (principalement ARNr et ARNt). Pour synthétiser l'ADNc à partir de l'ARNm, on utilise la **transcriptase inverse** de rétrovirus qui est capable de fabriquer un brin d'ADNc à partir de l'ARNm. L'enzyme pourra utiliser comme amorce un oligonucléotide poly dT complémentaire de la queue poly A. On obtient ainsi un hétéroduplex ARN/ADN qu'il faut transformer en un double brin d'ADN (Figure 9).

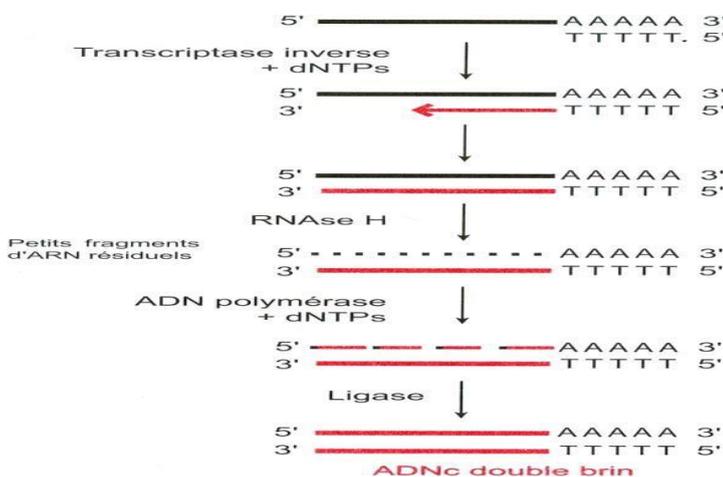


Figure 9. Synthèse d'un ADN complémentaire à partir d'un ARNm polyadénylé. Noter le rôle de la RNase H qui dégrade l'ARN de l'hétéroduplex ADN/ARN. La synthèse du 2e brin d'ADN peut s'effectuer à partir des petits fragments d'ARN résiduels ou d'amorces de séquences aléatoires (hexamères).

Pour cloner correctement le fragment d'ADNc il est souhaitable que ses extrémités soient cohésives. Cela peut être obtenu soit par une **transférase terminale** (catalyse l'addition de désoxynucléotides sur une fonction alcool 3'OH terminale de l'ADN et préfère les molécules d'ADN dont l'extrémité 3'OH est sortante, mais il existe des conditions qui favorisent la catalyse sur l'ADN à bouts francs, voire sur les extrémités 3'OH rentrantes) ou par l'addition d'adaptateurs aux deux extrémités. Ces adaptateurs contiennent des sites de clivage pour les enzymes de restriction qui seront utilisés pour le clonage dans des plasmides adéquats afin de créer la banque d'ADNc.

II.4. Les techniques d'analyse d'ADN

II.4.1. Séparation des fragments d'ADN

L'ADN étant une molécule chargée négativement (au niveau des groupements phosphates), il est possible de la faire migrer sous l'effet d'un champ électrique vers le pôle positif. La migration se fait de plus en plus dans un gel d'agarose ou de polyacrylamide, les grosses molécules seront plus freinées que les petites par le milieu d'électrophorèse. On observera donc une séparation des fragments d'ADN en fonction de leur masse : les petites molécules migreront le plus loin du dépôt, les plus grosses le plus près (Figure 10) :

- Le gel de polyacrylamide permet de séparer de petits fragments d'ADN inférieurs à quelques centaines de paires de bases (pb), en position verticale. Ce support permet l'obtention de meilleurs résultats (séquençage d'ADN).
- Le gel d'agarose permet de séparer les fragments de plusieurs milliers de pb, en position horizontale.

On fait migrer en même temps que les fragments d'ADN des marqueurs de poids moléculaire de taille connue. Les bandes d'ADN obtenues par électrophorèse sont invisibles dans un premier temps ; on les rend visibles en utilisant le **bromure d'éthidium**, qui s'intercale entre les paires de bases de l'ADN et fluoresce sous un rayonnement UV.

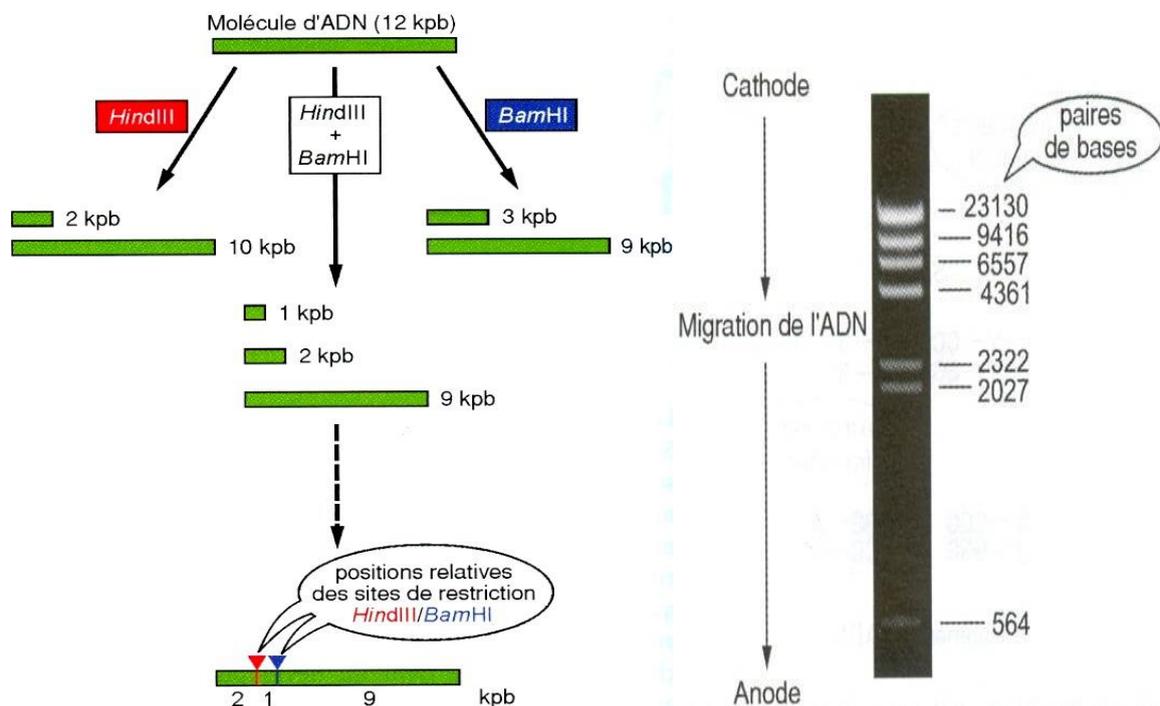


Figure 10. A gauche, Carte de restriction d'une molécule d'ADN en combinant deux enzymes de restriction. Les fragments d'ADN de restriction sont séparés par électrophorèse pour déterminer la position des sites des endonucléases. A droite, électrophorèse sur gel d'agarose (1%) de l'ADN du phage lambda digéré par *HindIII*, qui est utilisé comme marqueur de poids moléculaire

II.4.2. Marquage des ADN- Réactions de Phosphorylation-déphosphorylation (Figure 11)

a. Polynucléotide kinase T4

La polynucléotide kinase catalyse le transfert du phosphate γ de l'ATP sur une fonction alcool du carbone 5' d'un ADN ou d'un ARN. Le substrat requiert une déphosphorylation préalable si la fonction est

estérifiée. L'enzyme catalyse aussi l'échange du phosphate 5' terminal d'un ADN ou d'un ARN avec le phosphate γ de l'ATP, en présence d'ADP comme accepteur de phosphate. La polynucléotide kinase du commerce est extraite d'*E. coli* infectée par le bactériophage T4. Cette enzyme est utilisée au laboratoire pour incorporer du phosphate radioactif (^{32}P) sur l'extrémité 5' d'un acide nucléique, soit par transfert, soit par échange.

b. Phosphatase alcaline

Les phosphatases sont des enzymes à très large spécificité, hydrolysant le radical phosphoryle porté par une liaison ester ou anhydride. On les divise en plusieurs classes en fonction de leur pH optimum d'action : phosphatases acides, phosphatases alcalines. Certaines phosphatases ont une spécificité de substrat plus étroite : 5' nucléotidase. La réaction catalysée par les phosphatases est irréversible, mais certaines sont capables d'échanger le radical phosphate à partir de molécules plus riches en énergie comme le pyrophosphate. La phosphatase alcaline du commerce est celle extraite de l'intestin de veau.

La phosphatase alcaline est utilisée au laboratoire pour déphosphoryler les extrémités 5' des acides nucléiques, soit pour préparer un marquage en 5' par une polynucléotide kinase, soit pour empêcher la réannélation d'un ADN circulaire linéarisé (vecteurs de clonage).

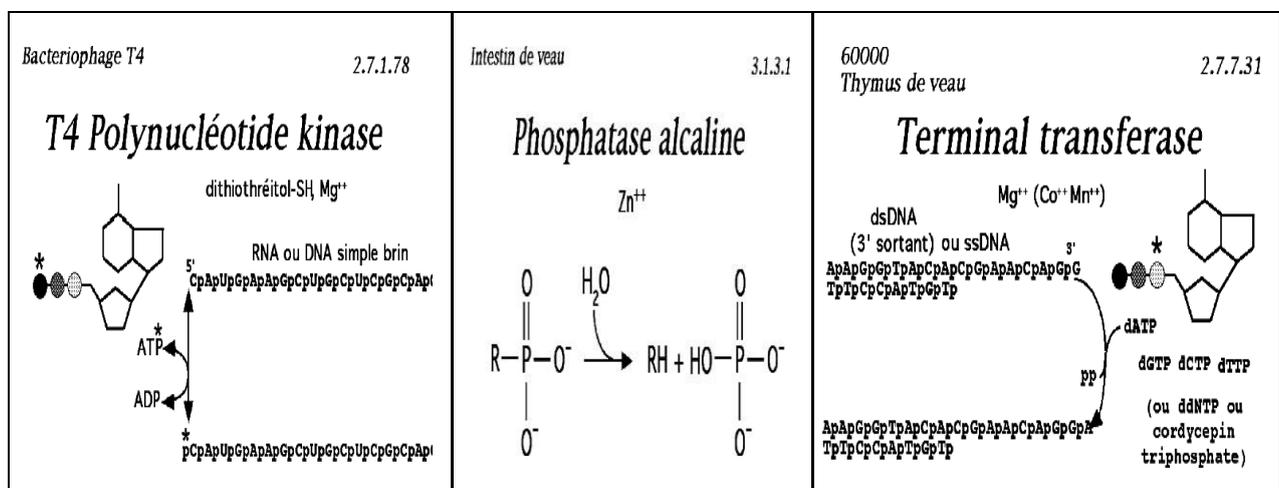


Figure 11. Enzymes utiles dans les réactions de Phosphorylation-déphosphorylation de l'ADN

c. Terminal transférase

La transférase terminale catalyse l'addition de désoxynucléotides sur une fonction alcool 3'OH terminale de l'ADN. L'enzyme préfère les molécules d'ADN dont l'extrémité 3'OH est sortante, mais il existe des conditions qui favorisent la catalyse sur l'ADN à bouts francs, voire sur les extrémités 3'OH rentrantes. Elle incorpore plus spécifiquement les purines ou les pyrimidines en fonction du cation divalent qu'on lui donne comme cofacteur : Mg^{++} pour les purines, Co^{++} pour les pyrimidines ou encore Mn^{++} . La transférase terminale du commerce est extraite du thymus de veau et elle est utilisée au laboratoire pour :

- * construire des séquences polymérisées de nucléotides (queues) à l'extrémité 3'OH de l'ADN (en vue du clonage des fragments d'ADN);
- * marquer les extrémités 3' de l'ADN avec un nucléotide marqué sur le phosphate α ;
- * ajouter un nucléotide à la fin d'une séquence pour induire une mutation dirigée.

II.4.3. Le séquençage de l'ADN

Pour connaître la séquence des ADN, on fait synthétiser un brin d'ADN par une enzyme spécifique. La méthode de séquençage la plus utilisée est celle qui a été développée par **Fred Sanger**, appelé aussi méthode des **didésoxyribonucléotides**. Elle met en œuvre, dans sa version initiale, une succession d'étapes comportant la séparation des deux brins d'ADN (dénaturation), l'hybridation d'une **amorce marquée** par un élément radioactif ou un dérivé fluorescent et une polymérisation par une ADN polymérase en présence de didésoxyribonucléosides triphosphate (ddNTP). Ces nucléotides particuliers (Figure 12a-b) sont normalement incorporés dans un brin d'ADN en cours de synthèse grâce à leur extrémité 5' triphosphate, mais ils ne peuvent pas former de liaison phosphodiester en 3' avec le nucléoside phosphate suivant en raison de l'absence du groupe OH à cette position. L'introduction d'un ddNTP provoque donc l'arrêt de la synthèse.

Dans chacun des quatre tubes contenant la matrice d'ADN à séquencer, une amorce marquée et les quatre dNTP, on introduit une petite quantité d'un des quatre ddNTP suivants: ddATP, ddCTP, ddGTP ou ddTTP. À la fin de la réaction de polymérisation, chaque tube contient des fragments d'ADN de taille variable, correspondant à l'arrêt de la synthèse pour chacun des sites d'incorporation du ddNTP spécifique du tube. En séparant sur gel de polyacrylamide les fragments d'ADN de chaque réaction de polymérisation, il est possible de déduire la séquence du fragment d'ADN initial (Figure 12c).

Aujourd'hui, de puissants séquenceurs automatiques sont capables d'effectuer ce travail en routine. Ils déterminent la séquence nucléotidique d'un ADN inconnu par la séparation automatisée des divers fragments au sein d'un capillaire rempli de gel et la lecture directe, à l'aide d'un faisceau laser, des signaux de fluorescence émis par les différents ddNTP eux-mêmes.

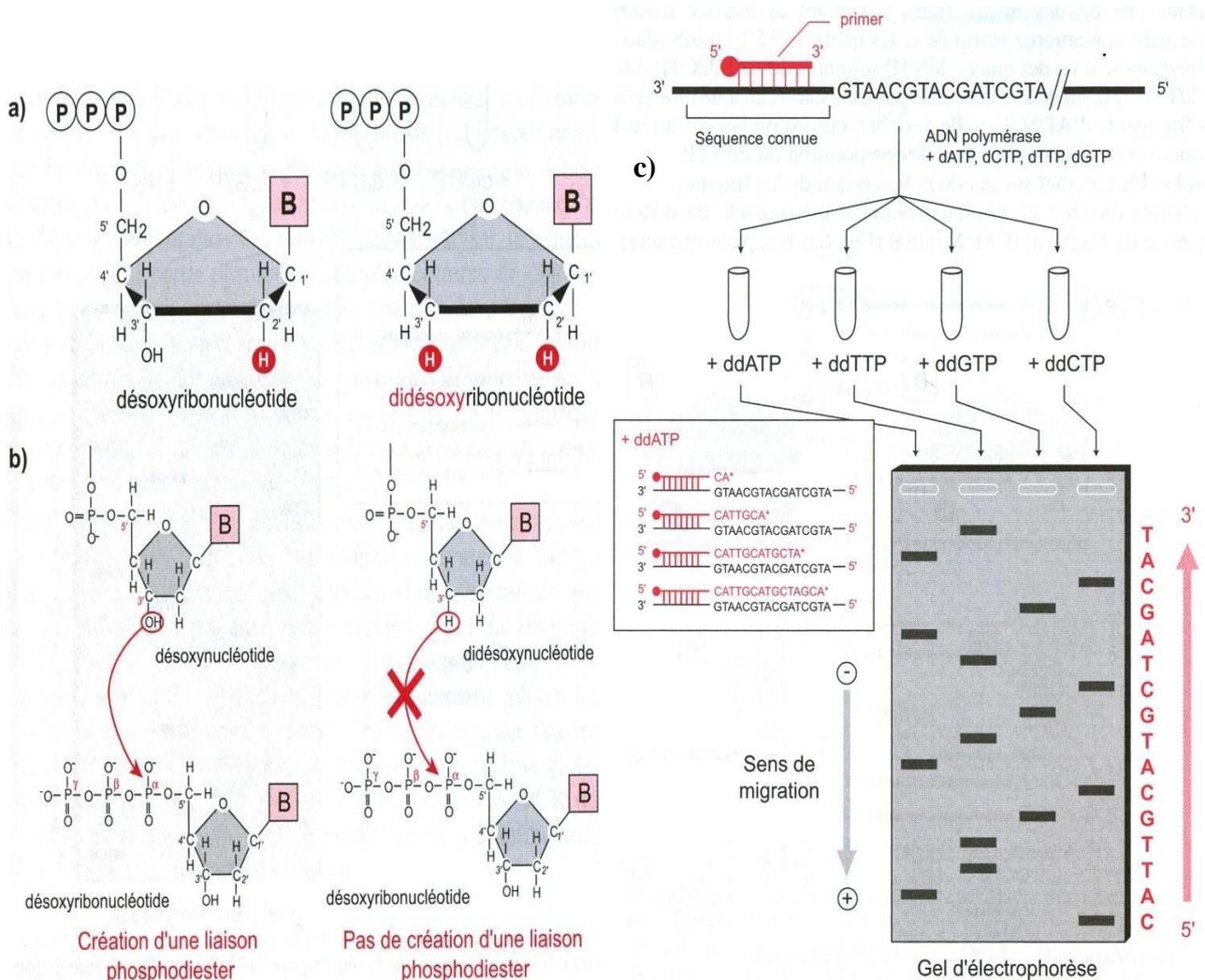


Figure 12. a) structure d'un didésoxyribonucléoside triphosphate (ddNTP) ; b) les ddNTP ne peuvent pas créer de liaison phosphodiester 3'-5'. c) principe du séquençage de l'ADN selon la méthode de Sanger en présence des 4 ddNTP. Noter que la séquence établie expérimentalement est la séquence complémentaire de la séquence initiale à déterminer

De nombreuses informations de séquence sont ainsi accumulées et cataloguées dans des banques de données informatiques, librement consultées par la communauté scientifique. Des comparaisons de séquences nucléotidiques permettent de déterminer si une séquence est déjà connue ou si, encore inconnue, elle appartient à une famille répertoriée. L'application du code génétique (Voir Chapitre V) dans le respect de l'usage des codons propre à chaque grand groupe d'êtres vivants, conduit ainsi à l'identification de séquences protéiques déduites des séquences nucléiques et à la possibilité d'une analyse comparative approfondie.

II.4.4. La réaction de Polymérisation en Chaîne (PCR)

Mise au point dans les années 1980, la technique de réaction de polymérisation en chaîne ou PCR (**Polymerisation Chain Reaction**) a ouvert de nouvelles voies dans l'étude et l'analyse des acides nucléiques et des gènes. Cette technique consiste à réaliser un nombre très élevé de copies d'une séquence d'ADN choisie par l'expérimentateur ce qui facilite grandement son étude et son exploitation. Une ADN polymérase, la **polymérase Taq**, extraite de la souche bactérienne *thermophile Thermus aquaticus*, est utilisée pour catalyser la réaction de polymérisation à partir d'amorces complémentaires de chaque brin d'ADN à amplifier. La Taq polymérase allonge les amorces dans le sens 5'→3'.

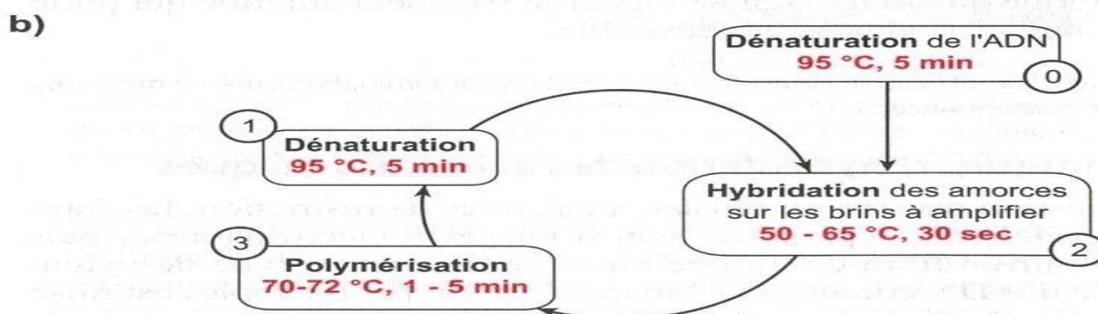
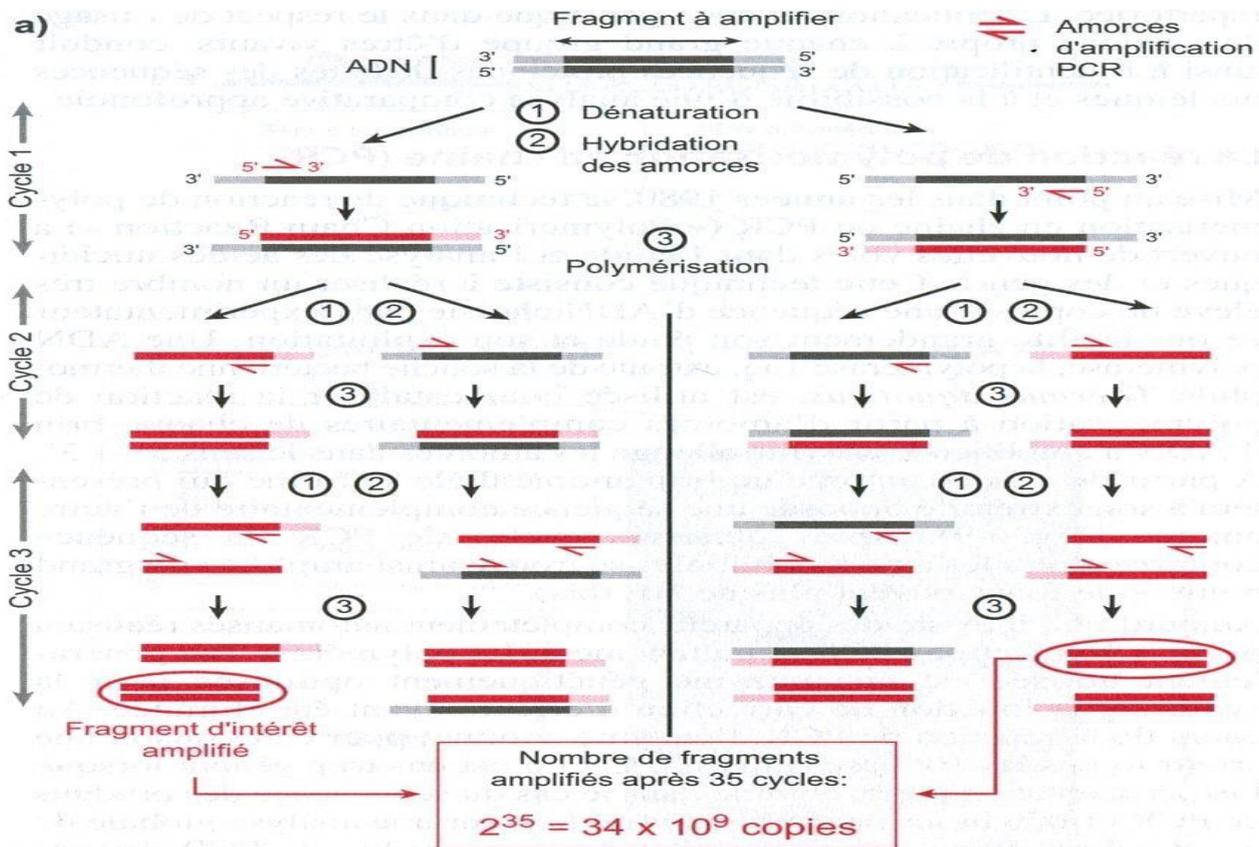


Figure 13. Principe de l'amplification d'un fragment d'ADN par PCR

La PCR permet d'amplifier spécifiquement une région de l'ADN double brin de quelques centaines de paires de bases. Cet ADN doit d'abord être séparé en simples brins (dénaturation à 95°C). On ajoute à l'ADN de départ une large quantité d'amorces (oligonucléotides synthétiques complémentaires des deux extrémités de la région à amplifier) qui vont s'hybrider à 50-60°C environ avec la séquence complémentaire sur chacun des brins de l'ADN, et les quatre dNTP qui serviront de substrats. On soumet le tout à l'activité d'une ADN polymérase (Taq polymérase) qui synthétise à 72°C un brin complémentaire antiparallèle à partir du 3'OH de l'amorce hybridée. On obtient quatre brins d'ADN. On recommence à dénaturer ces 4 brins, puis on les laisse s'hybrider avec les amorces (toujours en excès) et la polymérase entre en action pour aboutir à 8 brins d'ADN. On recommence à dénaturer ces 8 brins, puis on les laisse s'hybrider avec les amorces (toujours

en excès) et la polymérase entre en action pour aboutir à 16 brins d'ADN. Après plusieurs cycles de PCR, la séquence comprise entre les amorces initiales se trouve ainsi amplifiée un grand nombre de fois (souvent plus de 10^6 fois, Figure 13).

Aujourd'hui, il existe des appareils complètement automatisés réalisant jusqu'à 96 réactions PCR simultanément. La polymérase Taq généralement utilisée est une enzyme génétiquement optimisée pour la synthèse, la fonction de correction d'erreurs ayant été éliminée. Au cours de la

réaction de PCR, l'enzyme commet pour cette raison une erreur toutes les 10^4 bases environ. Ceci n'est pas trop gênant lorsque l'expérimentation porte, comme dans le cas du séquençage des produits de PCR (1000 bases par produit amplifié), sur une analyse globale de ces derniers. Par contre une erreur dans un produit de PCR devient problématique quand il s'agit de cloner le fragment amplifié qui porte alors une mutation aléatoire indésirable.

II.4.5. Les techniques d'hybridation des acides nucléiques

L'une des techniques majeures dans l'analyse d'un fragment d'ADN est la technique dite de transfert selon **Southern**, du nom de son inventeur, E.M. Southern. Après digestion par une ou plusieurs enzymes de restriction, les fragments d'ADN sont séparés au sein d'un gel d'électrophorèse, puis transférés sur un filtre de nitrocellulose ou une membrane de nylon. Une sonde d'ADN connu (ADN marqué au ^{32}P par exemple) est mise au contact de la réplique et hybridée par complémentarité aux fragments d'ADN étudiés (Figure 14).

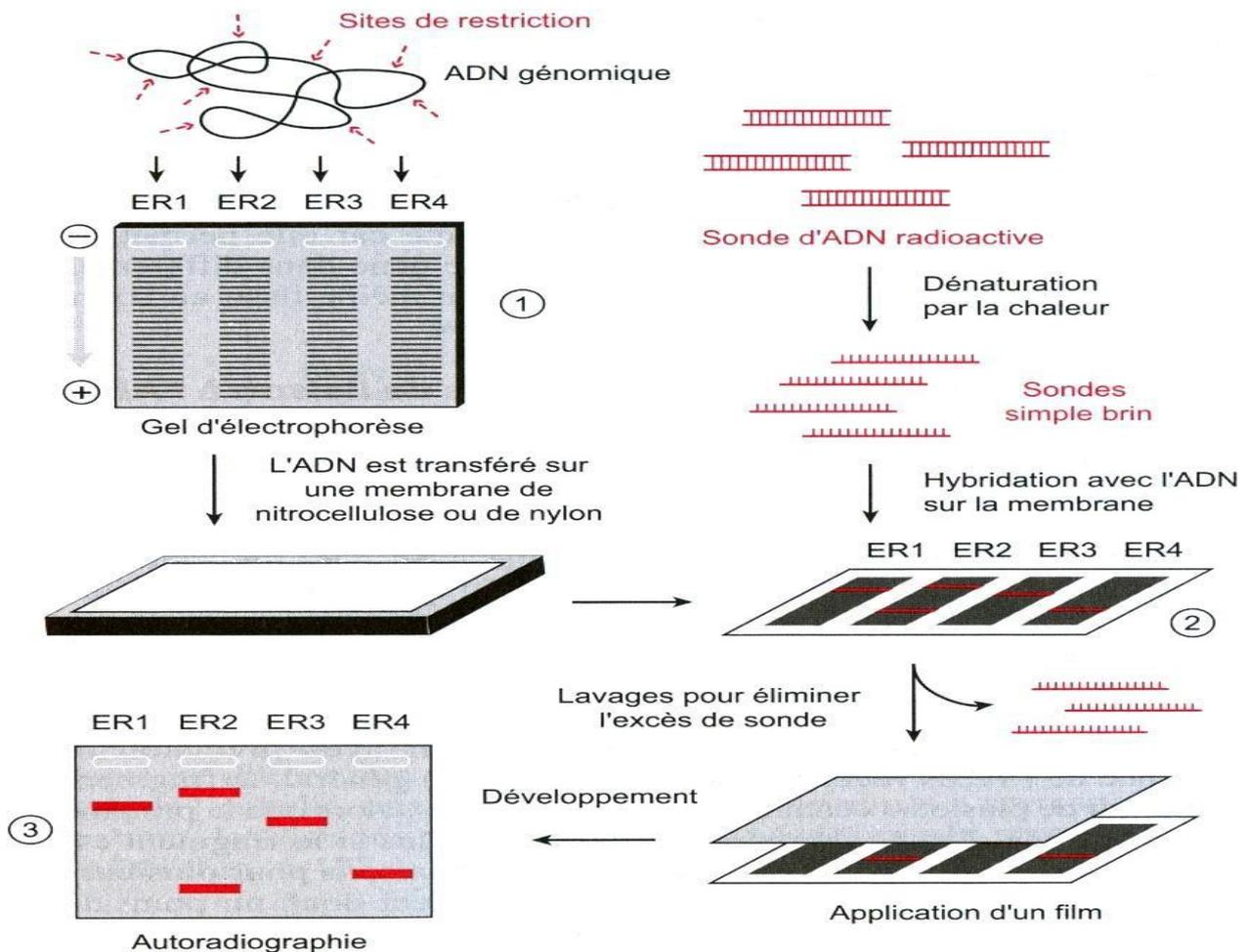


Figure 14. Principe de la technique de transfert-hybridation de Southern. Les très nombreux fragments d'ADN sont préalablement séparés par électrophorèse et repérés avant d'être transférés sur la membrane où ils sont invisibles, sauf en lumière UV où ils apparaissent sous la forme d'une traînée continue (1). Seuls les fragments reconnus après hybridation par la sonde (2) seront révélés par autoradiographie (3).
ER1 à ER4 : différents enzymes de restriction.

L'autoradiographie de la membrane révèle des bandes correspondant aux fragments d'ADN qui se sont hybridés avec la sonde. Par cette technique de transfert et d'hybridation, il est possible de réaliser une carte de restriction d'un fragment d'ADN c'est-à-dire de déterminer la position des sites de restriction au sein de cet

ADN. Cette technique est intéressante pour mettre en évidence les remaniements génétiques (délétions par exemple), l'organisation des gènes les uns par rapport aux autres sur un chromosome et donc la disposition des groupes de gènes fonctionnellement apparentés.

La technique de transfert-hybridation de Southern est également intéressante pour la recherche de séquences apparentées.

La même technique de transfert peut s'appliquer aux ARN. On parle dans ce cas de transfert de type **Northern** (par référence à Southern). Elle s'avère très intéressante car elle permet de comparer les niveaux d'expression d'un même gène dans différentes situations (différents tissus d'un organisme, un même tissu au cours du développement, etc.).

II.4.6. Les techniques de localisation des sites de liaison à l'ADN

De nombreuses protéines comme les facteurs de transcription se lient à de très courtes séquences d'ADN. Plusieurs méthodes ont été développées pour localiser des sites de liaison de protéines sur des fragments d'ADN. Ces techniques s'avèrent très efficaces, même si la protéine est inconnue. L'une d'entre elles utilise les propriétés de migration sur gel de fragments d'ADN lié ou non à une ou plusieurs protéines. En effet, si à un fragment d'ADN s'associent des protéines, la migration électrophorétique du complexe ADN-protéine sera retardée comparativement à la migration de l'ADN nu. Cette technique est appelée pour cela **retard sur gel** (Figure 15).

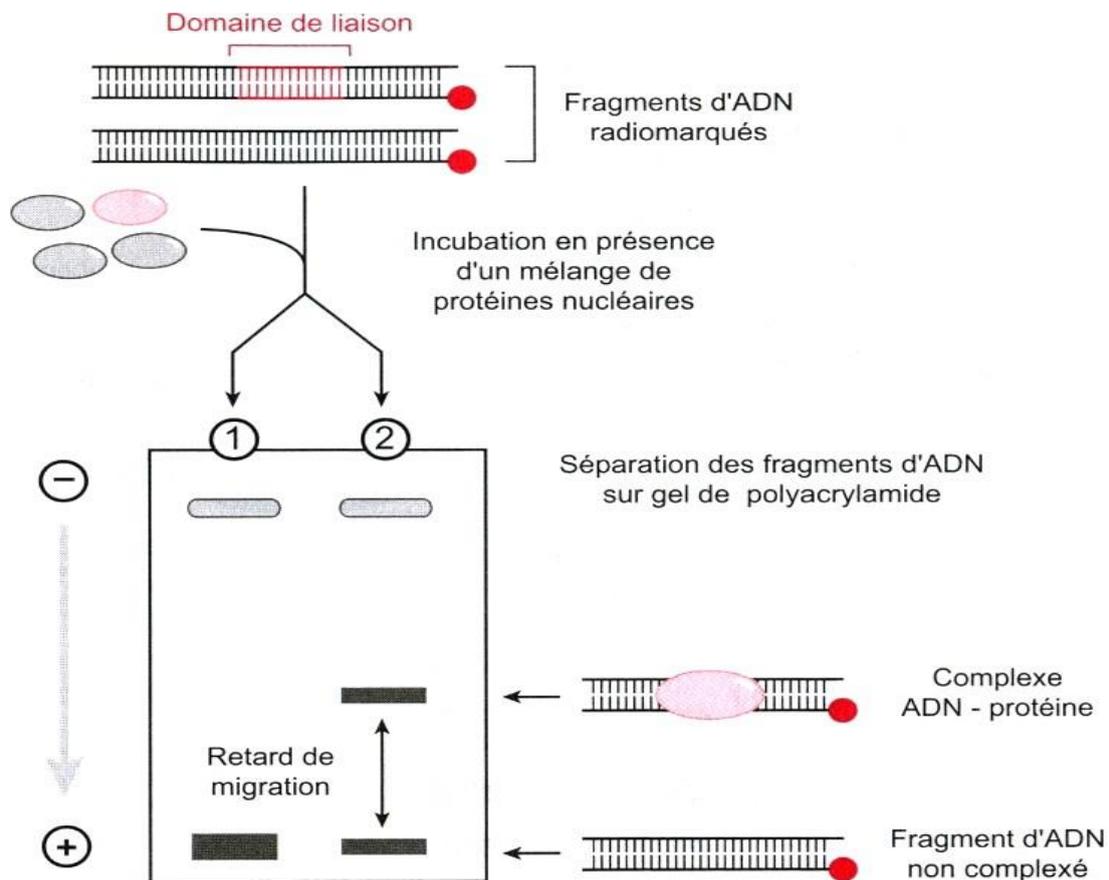


Figure 15. Analyse par retard sur gel. Un extrait de protéines nucléaires a été mélangé à des fragments d'ADN radiomarqués dont un est une cible potentielle d'une des protéines nucléaires. Le complexe formé de masse moléculaire supérieure à celle de l'ADN «nu» migre plus lentement (piste 2). Le fragment correspondant est donc retardé. Dans une expérience similaire non représentée ici, la protéine est connue et repérée à l'aide d'un anticorps. C'est le fragment d'ADN portant le site de liaison qui est recherché.

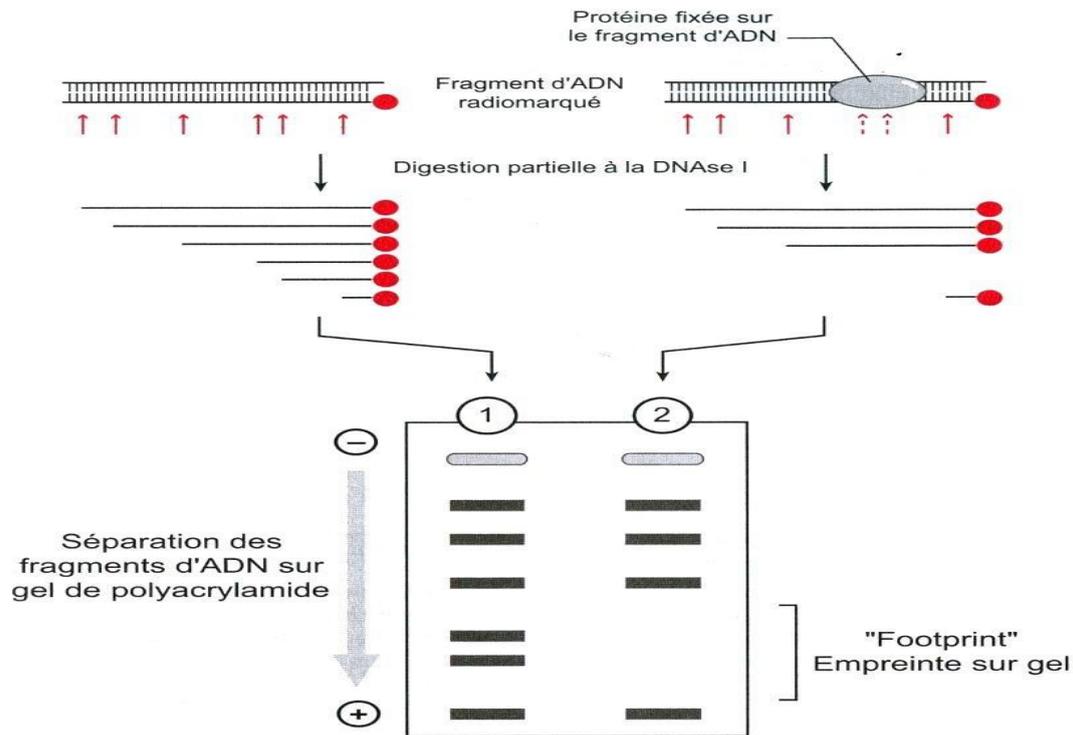


Figure 16. Technique d'empreinte à la DNase I «footprinting». Les fragments utilisés sont marqués à leurs extrémités et le traitement à la DNase I s'effectue en présence de sels de manganèse. Ainsi, elle clive au hasard (flèches verticales) les deux brins de molécule cible. La zone sans bande, «footprint» correspond à la position des sites protégés par la protéine de liaison à l'ADN.

Le retard sur gel ne permet pas de définir avec précision la séquence de l'ADN reconnue par la protéine. En général, le fragment retardé est de plusieurs centaines de paires de bases (pb), alors que la protéine se fixe sur une dizaine de pb. De plus si le fragment est long, il est susceptible de contenir des sites de liaisons pour plusieurs protéines. Les études de retard sur gel constituent donc un point de départ et sont complétées par d'autres techniques (par exemple celle dite de la protection à la DNase I ou du «**footprinting**»), qui apportent des informations plus précises (Figure 16).

Cette technique de protection est basée sur le principe suivant: une protéine liée à une molécule d'ADN protège une partie des nucléotides contre l'action de la DNase I, une nucléase à large spectre qui clive aléatoirement les liaisons phosphodiester, et libère de petits fragments d'ADN, variables en taille. Après élimination de la protéine, les fragments sont séparés sur gel. On obtient une échelle dont la taille des fragments ne diffère que par une ou un petit nombre de bases. À l'intérieur de l'échelle, une zone blanche ou '**empreinte**' correspond à la position des liaisons phosphodiester protégées par la protéine liée à l'ADN.

II.4.7. Criblage à l'aide d'une sonde d'acide nucléique

Un type de criblage consiste à chercher dans une banque les cellules ayant intégré la séquence d'ADN d'intérêt de celles ayant intégré le plasmide non recombinant. Une des techniques communément employées est l'hybridation de l'ADN des cellules avec des sondes nucléotidiques marquées.

Pour cela, on réalise une réplique en transférant par contact quelques cellules de chaque colonie sur un filtre de nitrocellulose ou une membrane de nylon. Les filtres sont traités pour éliminer les protéines, fixer l'ADN et le rendre accessible à l'hybridation par la sonde radioactive. Celle-ci est placée au contact du filtre où elle s'hybride avec la séquence complémentaire de l'ADN recherché. Les colonies positives sont identifiées par autoradiographie (Figure 17).

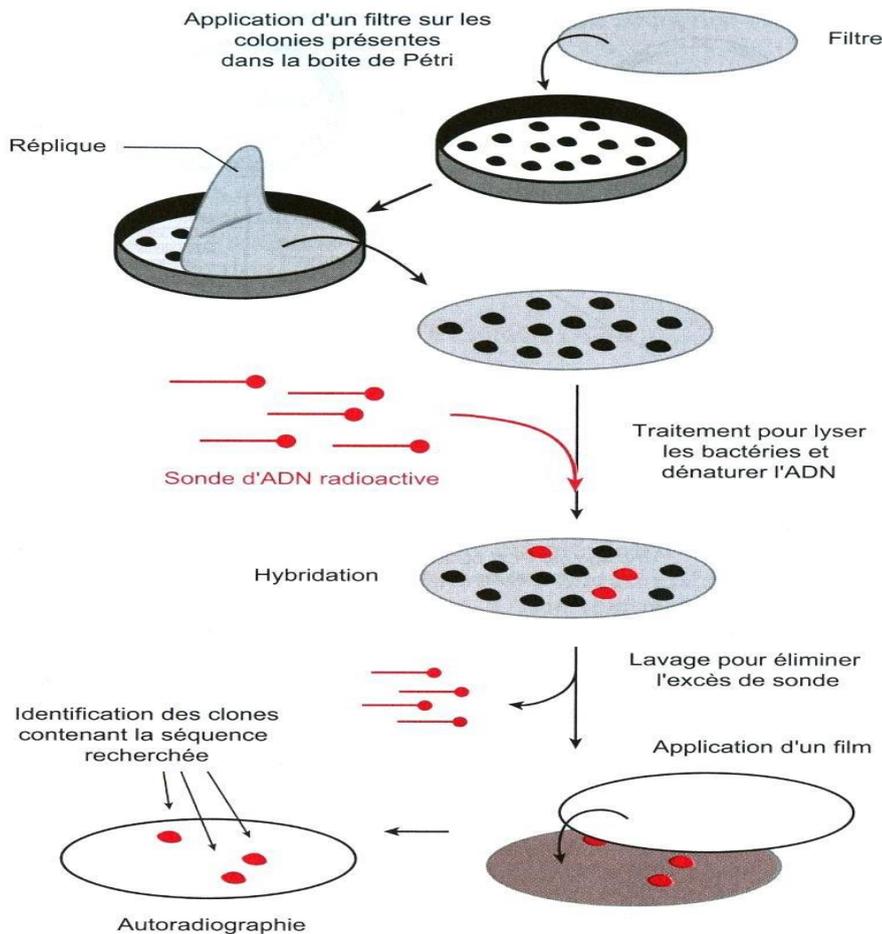


Figure 17. Criblage de cellules recombinées à l'aide d'une sonde nucléotidique. Après transformation, il est possible de rechercher les colonies ayant intégré un plasmide recombinant par hybridation avec une sonde homologue du gène d'intérêt. Une réplique des colonies sur boîte est effectuée sur un filtre. Des repères sont portés sur le filtre et sur la boîte afin d'assurer une correspondance exacte entre les cellules de la boîte et celles qui ont été transférées sur le filtre.

II.4.8. Puces à ADN et réseaux génétiques

Une **puce à ADN** est un ensemble de molécules d'ADN fixées en rangées ordonnées sur une petite surface qui peut être du verre, du silicium ou du plastique. Les puces à ADN permettent de tester simultanément la transcription d'une grande quantité de gènes d'un organisme. Cette biotechnologie récente permet d'analyser le niveau d'expression des gènes (transcrits) dans une cellule, un tissu, un organe, un organisme ou encore un mélange complexe, à un moment donné et dans un état donné par rapport à un échantillon de référence. Les puces à ADN sont aussi appelées **puces** à gènes, **biopuces** ou par les termes anglais « DNA chip, **DNA-microarray**, **biochip** ». Les termes français microréseau d'ADN et micromatrice d'ADN sont aussi des termes utilisés chez les chercheurs francophones.

Le principe de la puce à ADN repose sur la propriété que possède l'ADN dénaturé (simple brin) de reformer spontanément sa double hélice lorsqu'il est en présence d'un brin complémentaire (réaction d'hybridation). Techniquement parlant, après avoir déposé sur une plaque de verre ou de nylon des ADN spécifiques d'un gène particulier, on vient hybrider sur cette plaque de l'ARN fluorescent préalablement exprimé par une cellule dans une condition particulière. Une plaque peut contenir plusieurs milliers d'échantillons (Figure 18), ce qui donne accès à de nombreux gènes et permet de mettre en évidence des corrélations et donc les réseaux génétiques qui sous-tendent le fonctionnement de la cellule. Les puces à ADN peuvent être utilisées pour mesurer/détecter les ARN qui seront ou pas traduits en protéines. Les scientifiques parlent d'**analyse d'expression** ou de **profil d'expression**. Puisqu'il est possible de fixer jusqu'à un million de sondes sur une **biopuce**, les puces à ADN constituent ainsi une approche massive et ont contribué à la révolution de la **génomique** (discipline de la biologie moderne qui étudie le fonctionnement d'un organisme, d'un organe, d'un cancer, etc. à l'échelle du génome, au lieu de se limiter à l'échelle d'un seul gène. La génomique se divise en deux branches : la génomique structurale, qui se charge du séquençage du génome entier et la génomique fonctionnelle, qui vise à déterminer la fonction et l'expression des gènes séquencés en caractérisant le **transcriptome** et le **protéome**), puisqu'elles permettent en une seule expérience d'avoir une estimation sur l'expression de plusieurs dizaines de milliers de gènes. Mais il existe

également un grand nombre d'applications différentes qui font intervenir la technologie des puces à ADN (criblage de mutations, reséquençage, interactions ADN/protéine, écologie microbienne).

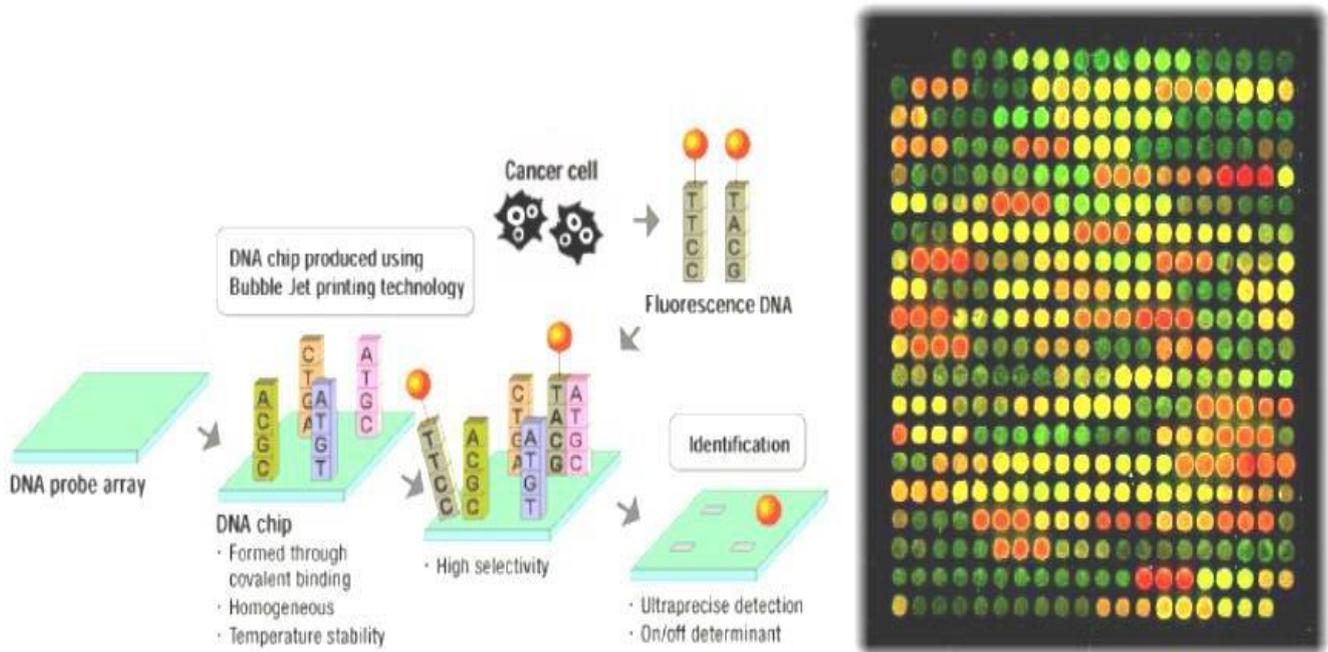


Figure 18. Principe et exemple de puce à ADN

II.4.9. Les applications de la technologie de l'ADN recombinant

a. La mutagenèse

La mutagenèse est l'introduction d'un changement dans la séquence nucléotidique d'un gène, autrement dit d'une mutation. Cette technique très puissante offre la possibilité d'agir sur la relation entre la séquence d'un gène et la fonction de la protéine codée. La mutagenèse *in vitro* est utilisée aujourd'hui de manière courante. Elle est aléatoire ou dirigée. La mutagenèse aléatoire engendre des mutations en des emplacements non prédéterminés d'un gène. L'une des méthodes consiste à produire une petite délétion en éliminant le fragment libéré par une coupure de l'ADN au niveau de deux sites de restriction. De même, au niveau d'un site de restriction unique se situant dans la séquence du gène à muter (Figure 19), une délétion de quelques nucléotides s'avère possible à l'aide d'une nucléase qui élimine les extrémités monocaténares (cas valable quand l'enzyme de restriction génère des bords cohésifs).

Une mutation par addition peut être également réalisée en provoquant la synthèse des séquences complémentaires des extrémités monobrin. La mutagenèse **aléatoire** produit des mutants qui comptent plusieurs modifications comparativement à la séquence de départ. Il devient alors difficile d'associer le phénotype muté à telle ou telle mutation. Cette technique reste tout de même très utile pour réaliser une première approche sur la relation entre la séquence d'une protéine, codée par le gène, et sa fonction biologique.

La mutagenèse **dirigée** porte sur un emplacement très précis de l'ADN (un nucléotide par exemple). Elle permet de modifier d'une manière prédéterminée la séquence d'un gène. Quand le site à muter est encadré par des sites de restriction uniques, il est possible d'enlever le fragment d'ADN du gène sauvage et de le remplacer par un autre fragment constitué d'oligonucléotides mutés. Ce processus est appelé mutagenèse par 'cassette'.

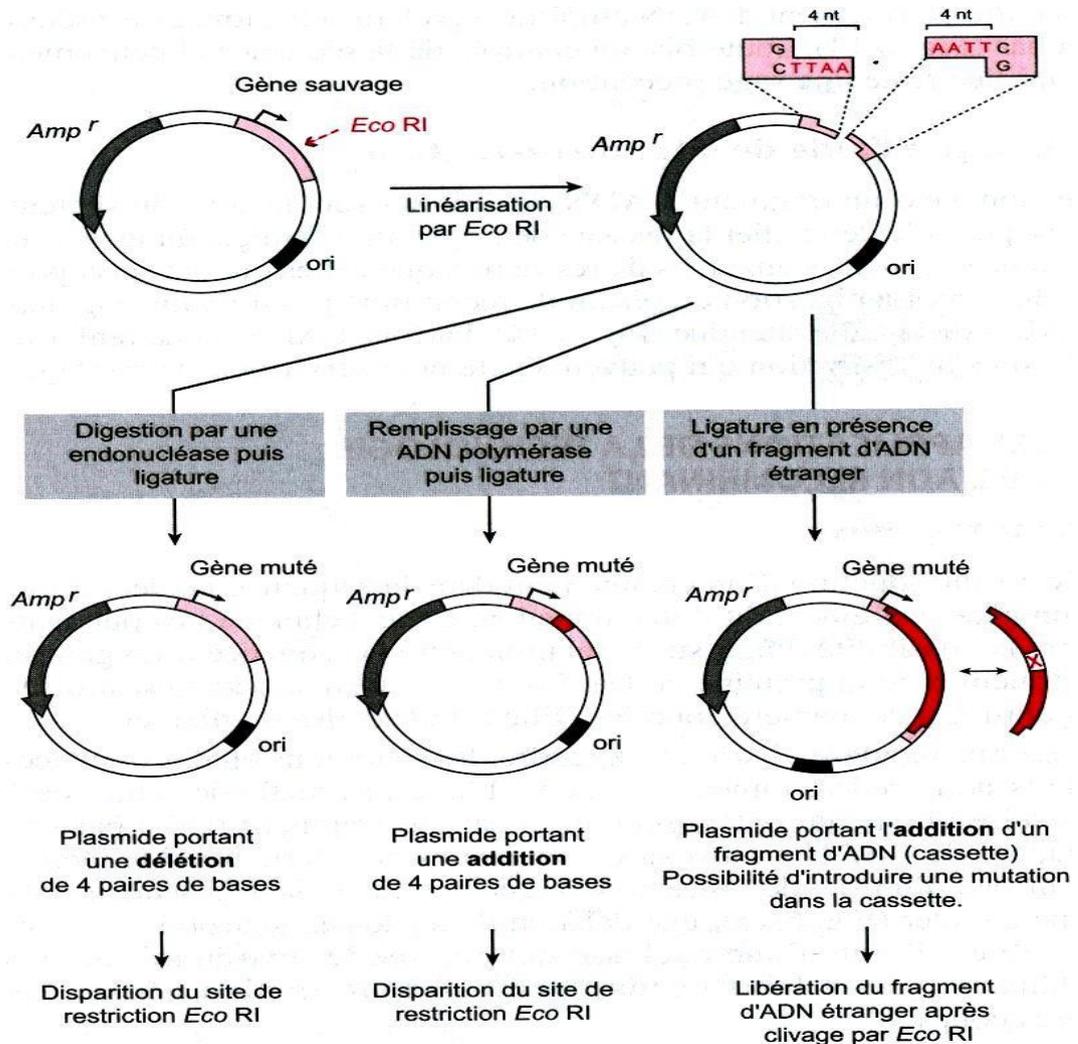


Figure 19. Mutagenèse par délétion, addition, ou par changement de cassette. L'enzyme de restriction à site unique libère des extrémités cohésives qui peuvent être éliminées par une nucléase entraînant après ligation une délétion de 4 nucléotides. Le gène peut également être muté par création de bouts francs après remplissage puis ligation (addition de 4 nucléotides) ou par remplacement d'une partie de sa séquence par un fragment d'ADN synthétique ou cassette. Cette dernière qui a des extrémités compatibles avec le site *Eco RI* peut également être mutée. La cassette devient un moyen de mutagenèse dirigée.

Une autre méthode de mutagenèse dirigée peut être employée même quand le gène à muter ne porte pas de sites de restriction. Elle consiste à insérer le gène d'intérêt dans un vecteur phagique simple brin, telle phage M13 (Figure 20). Un oligonucléotide synthétique, portant une mutation sur une seule base (substitution, insertion, délétion), s'hybride à la région homologue complémentaire et sert d'amorce pour la production complète du brin complémentaire. Après ligation de l'hétéroduplex (un brin muté, un brin sauvage), une souche bactérienne est transformée, et les clones portant la mutation sont sélectionnés. La cellule ne contient généralement qu'un seul type de plasmide. Lors de la réplication, l'un ou l'autre des deux plasmides (plasmides sauvage et muté) sera conservé par la cellule, l'autre sera détruit.

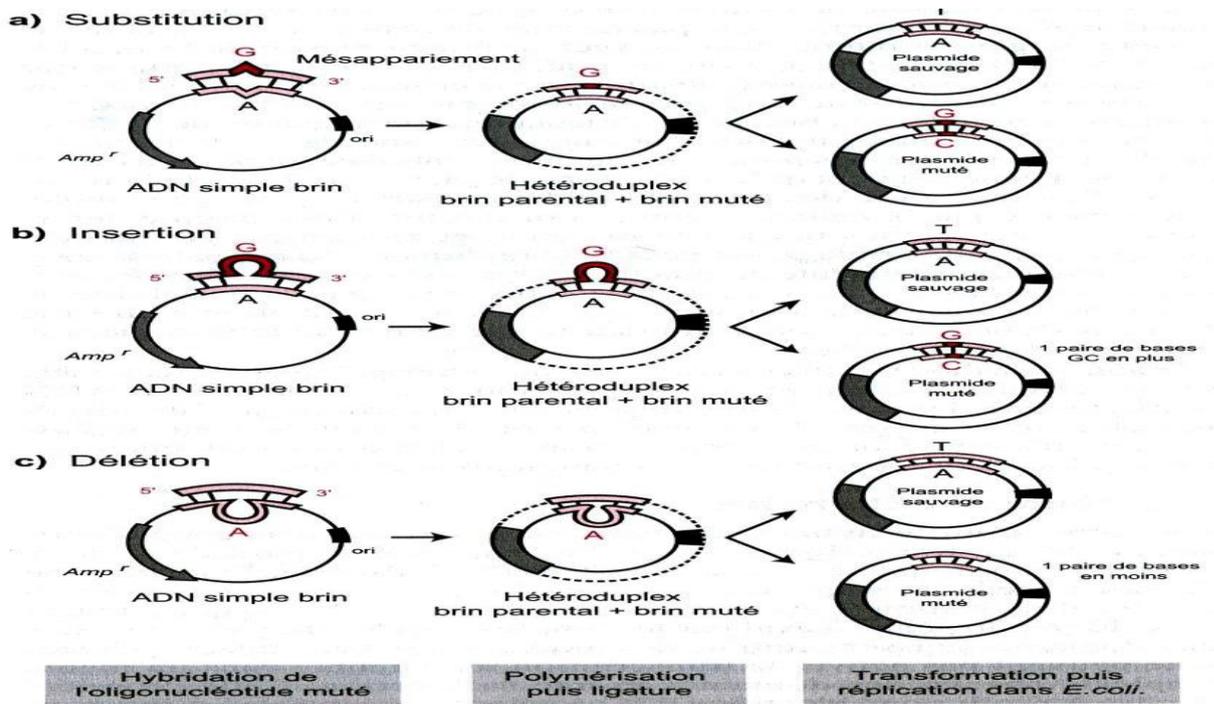


Figure 20. Principe de la mutagenèse dirigée utilisant un oligonucléotide muté et un vecteur phagique simple brin

b. Transfert et expression de gènes eucaryotes chez les procaryotes

Tout organisme ayant intégré dans son génome un fragment d'ADN étranger est appelé organisme **transgénique**. Le gène étranger est un transgène. Les bactéries dans lesquelles on a introduit des gènes eucaryotes sont donc des bactéries transgéniques. De toutes les cellules hôtes, *E. coli* et la levure sont les deux systèmes d'expression les plus utilisés en expérimentation de laboratoire et dans les applications biotechnologie en particulier pour la production à l'échelle industrielle de protéines recombinantes.

L'expression d'un gène cloné est possible grâce à des vecteurs d'expression possédant les signaux de transcription et de traduction appropriés. Ils possèdent tous une origine de réplication et un marqueur de sélection, souvent un gène de résistance à un antibiotique. Ils portent aussi un site multiple de clonage et une séquence codant un site de fixation au ribosome reconnu par les enzymes de la cellule qui héberge le vecteur. Les promoteurs forts des vecteurs d'expression sont, dans la plupart des cas, issus de promoteurs naturels (*trpE* par exemple) ou phagiques (T7 par exemple). Ces derniers sont particulièrement performants puisqu'ils détournent naturellement le métabolisme bactérien au bénéfice de la prolifération virale. Le promoteur souvent inducible permet de contrôler l'expression du transgène. Il sera induit par un choc thermique ou l'ajout dans le milieu de culture d'un inducteur chimique. Un exemple de plasmide d'expression est fourni par les vecteurs pET (Figure 21). Dans pET, la séquence nucléotidique d'intérêt est insérée dans la région des sites multiples de clonage. Elle se trouve ainsi en aval d'un promoteur T7 contrôlé par un opérateur lac et en amont d'une séquence codant une polyhistidine (6 résidus histidine) utile pour la purification du produit final. Ce plasmide favorise une expression fortement régulée grâce à un mécanisme en forme de cascade. En présence de galactosides (IPTG), une première induction, au niveau du génome de la bactérie, permet la synthèse de l'ARN polymérase de T7 à partir d'un promoteur lac modifié et rendu très efficace. Cette dernière grâce à la levée par l'IPTG (analogue chimique du lactose) de la répression du gène d'intérêt porté par le plasmide pET (deuxième induction) peut démarrer la transcription à partir du promoteur T7.

Pour permettre l'expression, le vecteur recombinant est introduit dans la bactérie hôte par le processus de la transformation. Les systèmes bactériens de transcription et de traduction produisent alors la protéine hétérologue codée par le gène étranger, comme ils le feraient pour une protéine de la cellule hôte. La protéine étrangère peut s'accumuler à des niveaux très élevés (jusqu'à 20 % des protéines totales de la bactérie). Elle est souvent produite sous forme d'une protéine de fusion en phase avec une protéine bactérienne.

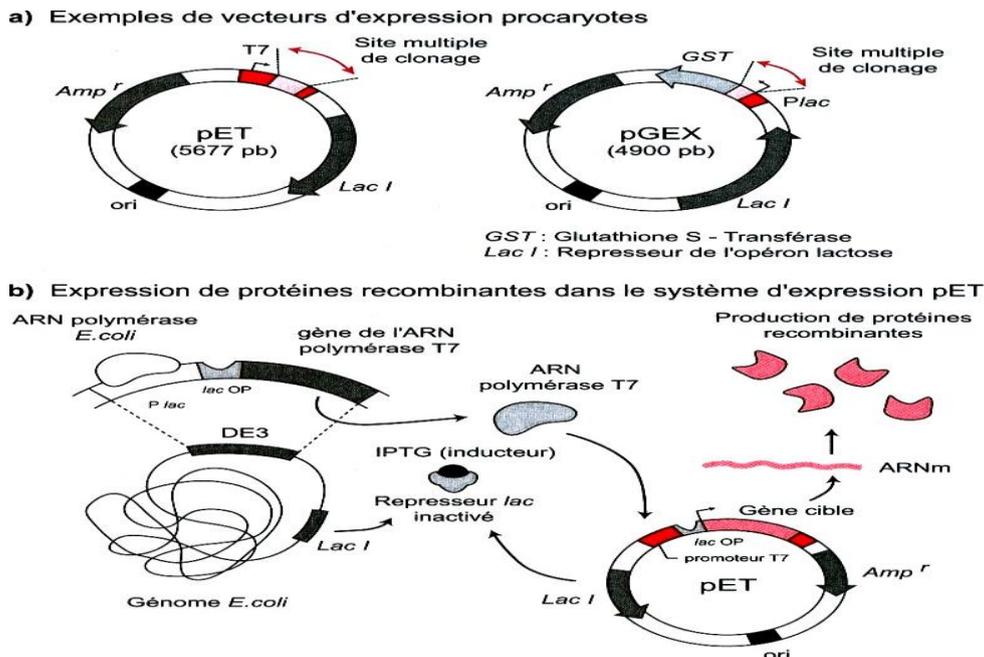


Figure 21. Production d'une protéine recombinante dans *E. coli* transformée par un vecteur d'expression. a) Schémas de deux vecteurs d'expression. *pGEX* porte la séquence codante de la glutathion S-transférase ; b) surexpression d'une protéine recombinante par une cascade inductive associant un vecteur *pET* et une souche bactérienne adaptée.

II.4.10. Le déchiffrement du génome humain

En 2001 la séquence nucléotidique du génome humain a été publiée suite à la réalisation du plus ambitieux projet de l'histoire de l'humanité par le consortium international appelé HUGO (Human Genome Organisation). La première version du génome humain comprend au total 3,2 milliards de paires de bases sur 22 autosomes et deux hétérosomes (X, Y). Une des surprises de cette découverte est le nombre relativement réduit de gènes qui ne dépasse probablement pas 30.000 gènes. Parmi eux, environ 75% spécifient des protéines (22.500) alors que les 25% restants vont à des gènes ne codant pas de protéines (pas de phase ouverte de lecture, 7.500 gènes). La figure 22 montre les 12 catégories de classes fonctionnelles de protéines.

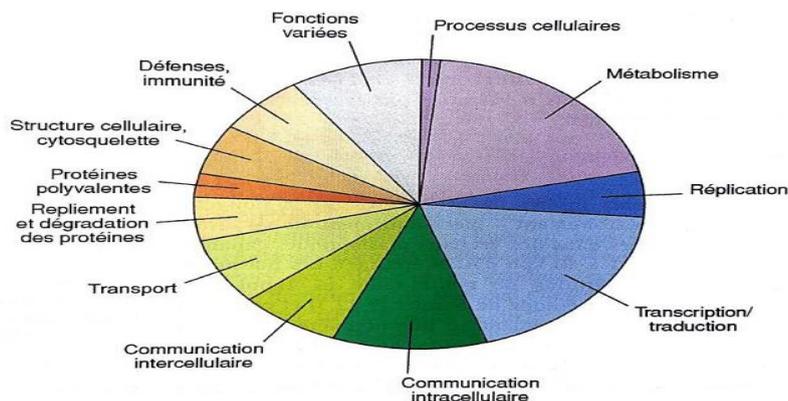


Figure 22. Fonctions biologiques des protéines du génome humain, regroupés en 12 catégories

Le séquençage des énormes bibliothèques d'ADN de l'homme a favorisé le développement de nouvelles stratégies de séquençage comme le **séquençage aléatoire** (Shotgun sequencing). Dans cette technique on utilise souvent de l'ADN extrait des lymphocytes qui est ensuite fragmenté (mécaniquement ou par des enzymes) et les fragments obtenus sont utilisés pour la construction de banques d'ADN génomiques dans des vecteurs BAC (bacterial artificial chromosomes) capables de contenir jusqu'à 150 kpb. Par la suite, l'ordre des fragments insérés dans les BAC est déterminé par des techniques comme la marche sur le chromosome et les séquences de ces fragments sont assemblées en **contigs** (Figure 23).

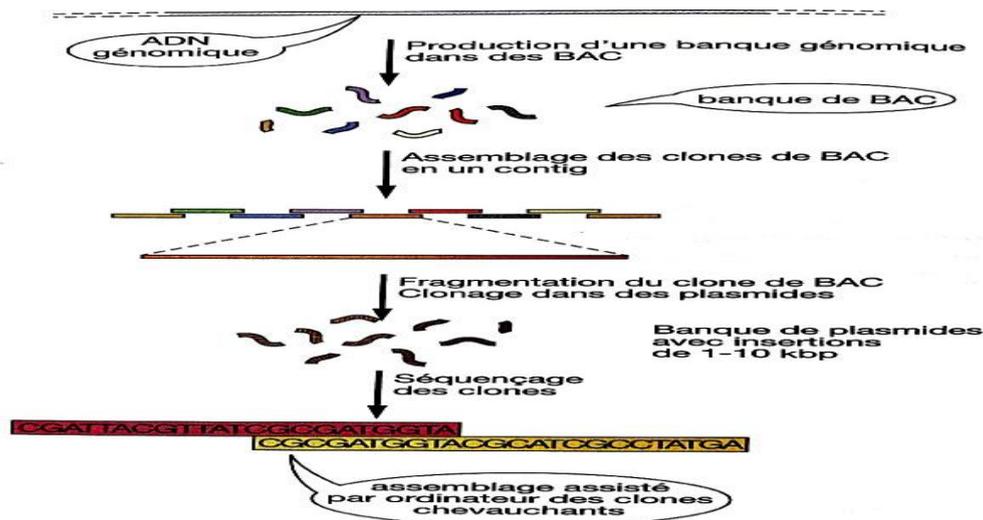


Figure 23. Séquençage aléatoire (Shotgun).

Il ressort de ce travail que seuls 25% environ du génome humain sont occupés par des gènes et les 75% restants sont des séquences intergéniques qui contiennent des régions régulatrices (Figure 24).

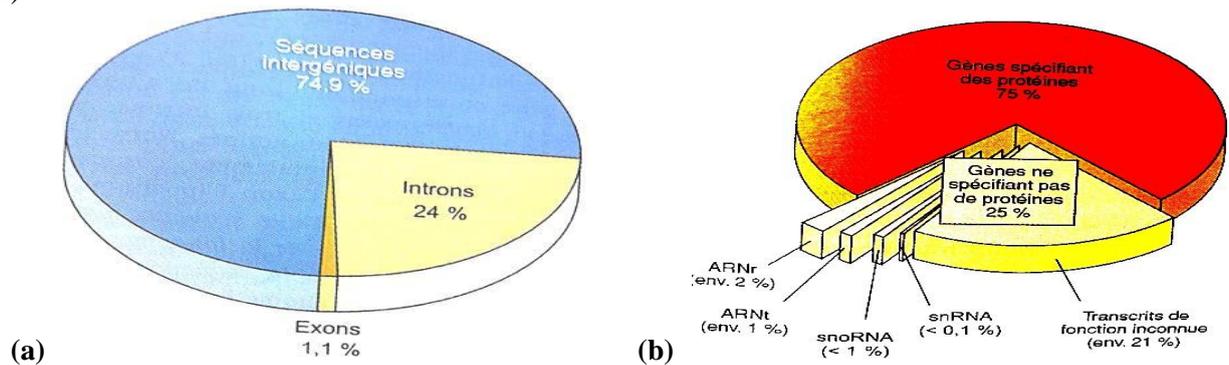


Figure 24. Organisation du génome humain. En (a), proportion relative de séquences codantes avec environ 226.000 exons identifiés. En (b), répartition entre gènes spécifiant ou non des protéines.

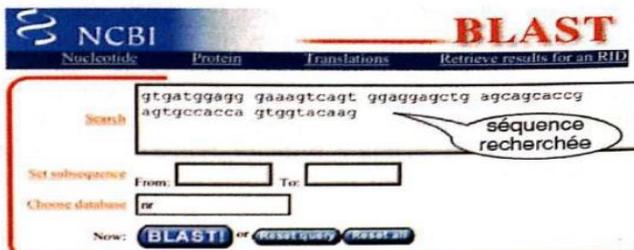
II.4.11. L'analyse des gènes *in silico*

La recherche biomédicale moderne s'appuie d'une part sur des procédés expérimentaux, et d'autres part sur l'interrogation à l'aide de l'outil informatique (ordinateurs et internet), ainsi que d'immenses banques de données qui conservent des données structurales et fonctionnelles concernant les gènes, les protéines et d'autres biomolécules, c'est l'analyse *in silico*. Le séquençage et l'identification de séquences de gènes (séquences génomiques ou d'ADN complémentaire (ADNc) sont enregistrées et rassemblées dans des banques de données comme celle du National Center for Biotechnology Information (NCBI), pour lesquelles on leur donne un numéro d'accès (accession number). Supposons qu'on a isolé une séquence (partielle et complète), on peut utiliser des programmes de recherche (**BLAST** pour **B**asic **L**ocal **A**lignment **S**earch **T**ool) pour rechercher, par alignement de séquences, des coïncidences avec la séquence isolée et la base de données NCBI (Figure 25). L'interrogation de ces banques de données permettent de donner beaucoup d'informations comme : séquences de gènes codantes ou non, localisation chromosomique, présences de polymorphisme de nucléotide simple ou **SNP** (prononcé 'snip', Single Nucleotide Polymorphism), ...

Si l'on suppose que nous recherchons des nouveaux partenaires d'interaction de la guanylate cyclase (GC) sensible au monoxyde d'azote (NO) chez l'homme, et que nous ayons identifié plusieurs protéines candidates. A partir des données des ADN complémentaires, on synthétise des primers à partir desquels on effectue des empreintes Northern Blot. Celles-ci révèlent un candidat potentiel intéressant dans un tissu donné où GC est bien abondant, comme montré par l'ARNm. Alors on se sert de la séquence de l'ADNc de ce candidat pour interroger la banque de données du NCBI, qui contient les gènes, ADNc et marqueurs génétiques

enregistrés par leur numéro d'accès. Comme montré dans la figure 25 (à gauche), et par alignement de séquences, le moteur de recherche BLAST trouve une coïncidence parfaite entre la séquence partielle de notre candidat et une région de la séquence L36861, qui spécifie une protéine humaine de fonction inconnue. Il s'agit d'un gène de longueur 4.054 nucléotides, avec quatre exons et trois introns, pour donner une protéine de 201 acides aminés. La séquence partielle trouvée expérimentalement correspond à une partie de la séquence de l'exon 1 (postions 241-480 nucléotides) du gène L36861. En conclusion, à partir d'une séquence partielle qui ne couvre qu'environ 1,5% de la longueur totale du gène, on peut avoir *in silico* la séquence entière du gène.

Par ailleurs, on peut localiser dans le génome humain le gène identifié et ce, en interrogeant la **banque de données de séquences génomiques humaines**, en partant de la séquence L36861 (Figure 25 à droite). Les résultats de la recherche montrent que cette séquence est semblable à 99,1% avec une région séquencée du chromosome 6, ce qui veut dire pas d'alignement parfait, du fait que notre séquence de départ contient encore des régions incomplètement séquencées dans les introns du gène L36861 (séquences marquées par 'n'). De plus deux autres gènes potentiels qui présentent plus de 94% d'identité avec L36861 et qui peuvent s'agir de gènes apparentés qui spécifieraient par exemple d'autres protéines associées à GC encore inconnues.



Résultats après comparaison de séquences

ACCESSION L36861

1 cccgttggtga agaaaggat taqaagattg aaaaatgggt taaaaaatcc tctcaatcc 60
 61 agctcaaggc tgaataccta agatctctct ctctctcaag cctctgctca tttcccaacc 120
 121 tttctccag ggtacagctt cccctggggg tgaaggattt tctctctcag cctcaaacac 180
 181 agtctccagc cactctgag ctgagctgga ggaactcact tctctctcag cctcaaacac 240
 241 tctccactt tcccgttacc atctatcaca tcaaggctcc ctctctctag tcttgaaggt 300
 301 acccaagcct gtgagagagg agggcccctt tctgggcccc gacagctttt tggcaggagg 360
 361 acccaaaagg tccctccact ctcaaaatc agcccctcga agggctccag aagggccctt 420
 421 gttatggagg gaaagtcaag tgggtggggc agaaacaaag aggtctccca agggctccag 480
 481 aggtctccga ctgagggccc ctctctctct atgactccct atgactccct ccaagctctc 540
 541 ggcctcaaga aactgagccc gtccggcagg cagtctctgt aaccagatgt ttgagatctt 600
 601 gactctcaaca agttgagag gggcccagtg gcagcagg agagtgctct agggaccctt 660
 661 ccccaagctt accagctctg tggctggagg gacgaacaa aggtctccca agggctccag 720
 721 ccccaagctt accagctctg tggctggagg gacgaacaa aggtctccca agggctccag 780
 781 agctcaaggc tgaataccta agatctctct ctctctcaag cctctgctca tttcccaacc 840
 841 agtctccagc cactctgag ctgagctgga ggaactcact tctctctcag cctcaaacac 900
 901 tctccactt tcccgttacc atctatcaca tcaaggctcc ctctctctag tcttgaaggt 960
 961 acccaagcct gtgagagagg agggcccctt tctgggcccc gacagctttt tggcaggagg 1020
 1021 acccaaaagg tccctccact ctcaaaatc agcccctcga agggctccag aagggccctt 1080
 1081 gttatggagg gaaagtcaag tgggtggggc agaaacaaag aggtctccca agggctccag 1140
 1141 aggtctccga ctgagggccc ctctctctct atgactccct atgactccct ccaagctctc 1200
 1201 ggcctcaaga aactgagccc gtccggcagg cagtctctgt aaccagatgt ttgagatctt 1260
 1261 gactctcaaca agttgagag gggcccagtg gcagcagg agagtgctct agggaccctt 1320
 1321 ccccaagctt accagctctg tggctggagg gacgaacaa aggtctccca agggctccag 1380
 1381 agctcaaggc tgaataccta agatctctct ctctctcaag cctctgctca tttcccaacc 1440
 1441 agtctccagc cactctgag ctgagctgga ggaactcact tctctctcag cctcaaacac 1500
 1501 tctccactt tcccgttacc atctatcaca tcaaggctcc ctctctctag tcttgaaggt 1560
 1561 acccaagcct gtgagagagg agggcccctt tctgggcccc gacagctttt tggcaggagg 1620
 1621 acccaaaagg tccctccact ctcaaaatc agcccctcga agggctccag aagggccctt 1680
 1681 gttatggagg gaaagtcaag tgggtggggc agaaacaaag aggtctccca agggctccag 1740
 1741 aggtctccga ctgagggccc ctctctctct atgactccct atgactccct ccaagctctc 1800
 1801 ggcctcaaga aactgagccc gtccggcagg cagtctctgt aaccagatgt ttgagatctt 1860
 1861 gactctcaaca agttgagag gggcccagtg gcagcagg agagtgctct agggaccctt 1920
 1921 ccccaagctt accagctctg tggctggagg gacgaacaa aggtctccca agggctccag 1980
 1981 agctcaaggc tgaataccta agatctctct ctctctcaag cctctgctca tttcccaacc 2040
 2041 agtctccagc cactctgag ctgagctgga ggaactcact tctctctcag cctcaaacac 2100
 2101 tctccactt tcccgttacc atctatcaca tcaaggctcc ctctctctag tcttgaaggt 2160
 2161 acccaagcct gtgagagagg agggcccctt tctgggcccc gacagctttt tggcaggagg 2220
 2221 acccaaaagg tccctccact ctcaaaatc agcccctcga agggctccag aagggccctt 2280
 2281 gttatggagg gaaagtcaag tgggtggggc agaaacaaag aggtctccca agggctccag 2340
 2341 aggtctccga ctgagggccc ctctctctct atgactccct atgactccct ccaagctctc 2400
 2401 ggcctcaaga aactgagccc gtccggcagg cagtctctgt aaccagatgt ttgagatctt 2460
 2461 gactctcaaca agttgagag gggcccagtg gcagcagg agagtgctct agggaccctt 2520
 2521 ccccaagctt accagctctg tggctggagg gacgaacaa aggtctccca agggctccag 2580
 2581 agctcaaggc tgaataccta agatctctct ctctctcaag cctctgctca tttcccaacc 2640
 2641 agtctccagc cactctgag ctgagctgga ggaactcact tctctctcag cctcaaacac 2700
 2701 tctccactt tcccgttacc atctatcaca tcaaggctcc ctctctctag tcttgaaggt 2760
 2761 acccaagcct gtgagagagg agggcccctt tctgggcccc gacagctttt tggcaggagg 2820
 2821 acccaaaagg tccctccact ctcaaaatc agcccctcga agggctccag aagggccctt 2880
 2881 gttatggagg gaaagtcaag tgggtggggc agaaacaaag aggtctccca agggctccag 2940
 2941 aggtctccga ctgagggccc ctctctctct atgactccct atgactccct ccaagctctc 3000
 3001 ggcctcaaga aactgagccc gtccggcagg cagtctctgt aaccagatgt ttgagatctt 3060
 3061 gactctcaaca agttgagag gggcccagtg gcagcagg agagtgctct agggaccctt 3120
 3121 ccccaagctt accagctctg tggctggagg gacgaacaa aggtctccca agggctccag 3180
 3181 agctcaaggc tgaataccta agatctctct ctctctcaag cctctgctca tttcccaacc 3240
 3241 agtctccagc cactctgag ctgagctgga ggaactcact tctctctcag cctcaaacac 3300
 3301 tctccactt tcccgttacc atctatcaca tcaaggctcc ctctctctag tcttgaaggt 3360
 3361 acccaagcct gtgagagagg agggcccctt tctgggcccc gacagctttt tggcaggagg 3420
 3421 acccaaaagg tccctccact ctcaaaatc agcccctcga agggctccag aagggccctt 3480
 3481 gttatggagg gaaagtcaag tgggtggggc agaaacaaag aggtctccca agggctccag 3540
 3541 aggtctccga ctgagggccc ctctctctct atgactccct atgactccct ccaagctctc 3600
 3601 ggcctcaaga aactgagccc gtccggcagg cagtctctgt aaccagatgt ttgagatctt 3660
 3661 gactctcaaca agttgagag gggcccagtg gcagcagg agagtgctct agggaccctt 3720
 3721 ccccaagctt accagctctg tggctggagg gacgaacaa aggtctccca agggctccag 3780
 3781 agctcaaggc tgaataccta agatctctct ctctctcaag cctctgctca tttcccaacc 3840
 3841 agtctccagc cactctgag ctgagctgga ggaactcact tctctctcag cctcaaacac 3900
 3901 tctccactt tcccgttacc atctatcaca tcaaggctcc ctctctctag tcttgaaggt 3960
 3961 acccaagcct gtgagagagg agggcccctt tctgggcccc gacagctttt tggcaggagg 4020
 4021 acccaaaagg tccctccact ctcaaaatc agcccctcga agggctccag aagggccctt

exon 1

exon 2

exon 3

exon 4

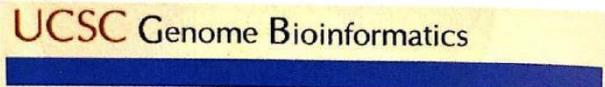
séquence recherchée

séquence du gène

région codante

région non traduite

séquence en acides aminés du produit du gène



Human BLAT Search

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

Please paste in a query sequence to see where it is located in the the genome.
 Multiple sequences can be searched at once if separated by a line starting with > and

```

3541 eatggtggca gcagagagge aagaatggga gtcaagacca ggaacagtga
3591 aggatgggtc ctggcccttc tgagtgcacag ctggtggcag cactccttgc
3641 tggggggcac tgttcaacat tctctctcag tctggtgacc ctacgctctc
3691 tgactccttc cagcttttcc cagctttccc cactgagctt ctccagctat
3741 gctctctcga cgtgactctc tgagcagaac tgaactttcc agccctctat
3791 ggaatcctgc agatcragtg gctgcagctt caatccagc gctgcaatca
3841 caactccat ctgctgggg accctggagc ctacttctgc gctttgact
3891 tcattgattg acgctcctt tcaacaagna ttactgagg cgtctacta
3941 gtactaatgc tagatgtag atgcaaacg aagcagcttt tcatctcta
3991 ggaactcata ggtcaatggt gacacacaa gacaacatc attataataa
4041 eaatggttaa gag
  
```

séquence : GCAP n° d'accès 236861

Comparaison avec la séquence du génome Recherche BLAT

Résultat

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDEN.	CHRO	ST.	START	END
bcr_details	YourSeq	3844	1	4053	4053	99.1%	6	+	48561395	48568245
bcr_details	YourSeq	104	1262	1406	4053	94.2%	5	-	5984116	5984341
bcr_details	YourSeq	101	1261	1383	4053	94.9%	6	-	63616159	63616289

Figure 25. A gauche, identification des gènes *in silico*. Grâce au programme BLAST, il est possible de comparer les séquences d'ADN obtenues expérimentalement et non connues avec les gènes humains (les nucléotides des introns désignés par 'n' indiquent non séquencés). A droite, recherche d'une région génomique dans la séquence du génome humain avec la séquence complète L36861 longue de 4053 nucléotides.

Chapitre III. LA REPLICATION

III.1. La réplication de l'ADN

La réplication d'ADN n'a lieu que pendant une période limitée du cycle cellulaire = phase S (pour synthèse) de l'interphase. Le premier postulat du dogme central, amplement vérifié, est que l'ADN est capable d'auto-réplication. Ceci avait été immédiatement déduit par Watson et Crick du fait que chaque brin d'ADN spécifie univoquement son complément, mais il a fallu du temps pour que les détails du mécanisme de la réplication soient élucidés.

La réplication aboutit à la formation de 2 lots d'ADN rigoureusement identiques tant sur le plan **quantitatif** que **qualitatif**. Les 2 branches de la molécule d'ADN mère se séparent par rupture de liaisons hydrogènes. Chaque chaîne sert de **matrice** à la synthèse d'une chaîne complémentaire par association entre bases complémentaires (A=T et C=G). Les 2 molécules obtenues sont strictement identiques à la molécule mère et chacune en a conservé une chaîne avec:

$$(n1 A + n2 G / n1 T + n2 C) = 1$$

III.1.1. Réplication sémi-conservative

La première question fut celle de la conservation du mécanisme de réplication. Les expériences maintenant classiques de **Meselson et Stahl** prouvèrent que la réplication de l'ADN est **semi-conservative** (Figure 1), c'est à dire chacune des molécules filles hérite d'un brin de l'ADN parental. Il aurait été parfaitement possible d'imaginer un mécanisme conservatif, où les deux brins de la molécule fille auraient été synthétisés *de novo*, en utilisant les brins parentaux comme matrice.

Dans ce modèle de réplication semi-conservative, chaque brin sert de matrice pour la synthèse d'une nouvelle chaîne complémentaire par les *ADN-polymérases* et se fait toujours dans le sens 5'p → 3'OH et nécessite une extrémité 3'OH libre en plus d'une amorce d'ARN. Il y a des avantages à une réplication semi-conservative, puisqu'elle permet de réparer *post facto* des erreurs commises pendant la réplication, dans la mesure où le brin parental peut être identifié.

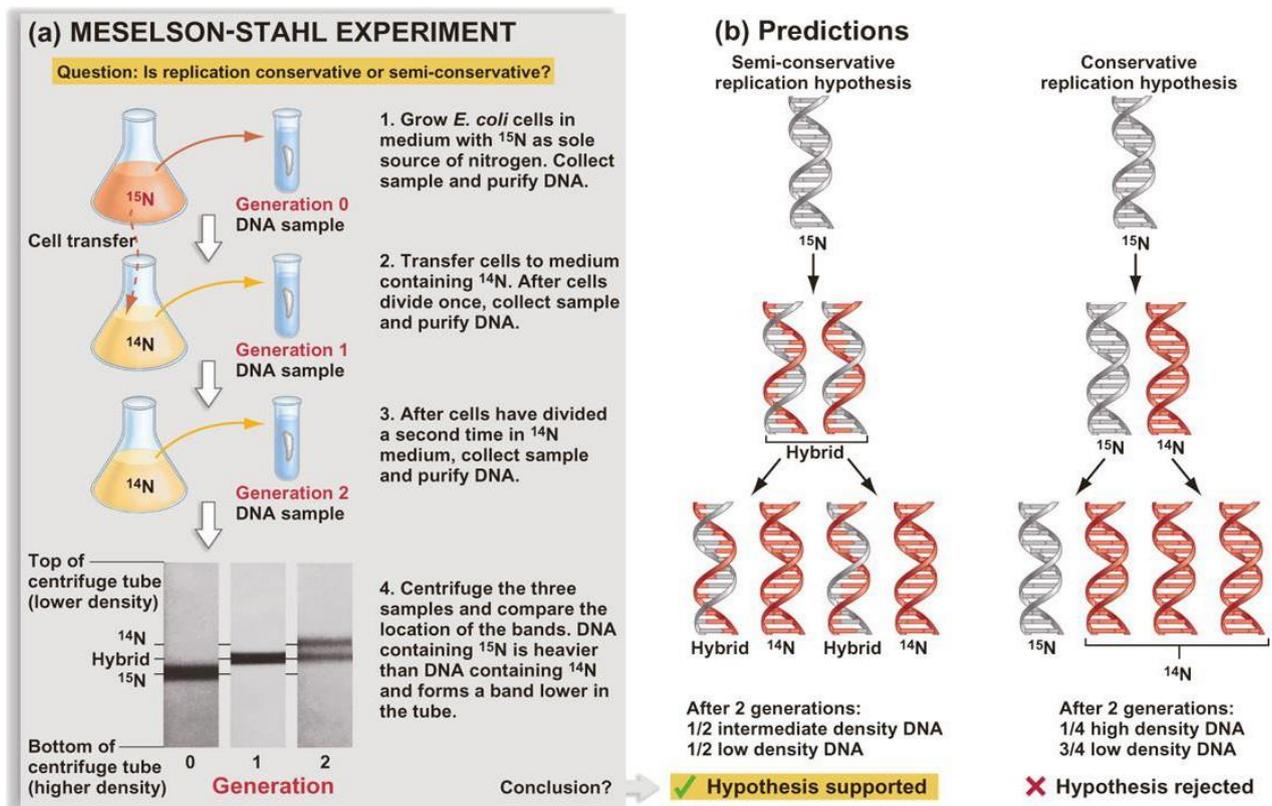


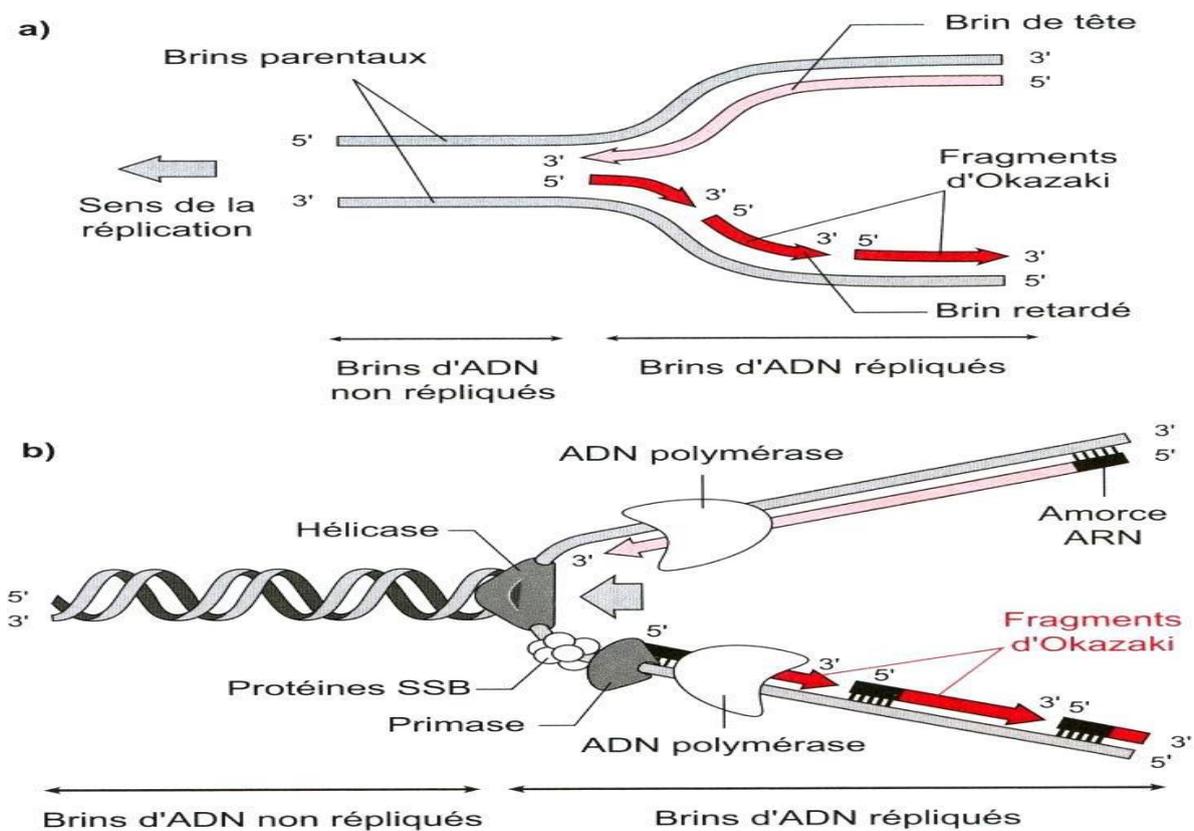
Figure 1. Expérience de Meselson et Stahl montrant que la réplication est sémi-conservative

III.1.2. Mécanisme de la réplication

Il faut ensuite identifier et purifier des enzymes capables de polymériser des nucléotides sur une matrice d'ADN. Un grand nombre d'**ADN polymérases** sont maintenant connues, qui toutes ont besoin d'une **amorce** à laquelle ajouter des nucléotides, et toujours à son extrémité 3'. La nécessité d'une amorce et la polarité 5' → 3' obligatoire posent un sérieux problème: comment peut-on synthétiser de façon coordonnée deux brins à la polarité antiparallèle? La solution a été d'utiliser des mécanismes **différents** pour les deux brins: alors que sur le brin tête ou conducteur la polymérisation se fait en continu, sur le brin retardé ou suiveur elle est discontinue. De courtes amorces sont d'abord préparées par une **ARN polymérase** (qui peut démarrer sans amorce), puis allongées par une ADN polymérase, puis enlevées, les trous restants sont réparés, et les brins sont recollés par une "**ADN ligase**".

La plupart des molécules d'ADN sont soit très longues, soit circulaires. Il faut donc prévoir des points de démarrage ou **origines de réplication**. Un processus complexe sépare les deux brins d'ADN à ces origines, met en place les premières amorces, et démarre des **fourches** de réplication dans les deux sens. On obtient donc des **bulles de réplication** qui sont caractéristiques de l'ADN en phase réplivative.

Figure 2. Mécanisme de la réplication et formation de la fourche de réplication



La formation de la fourche de réplication se fait par déroulement de la molécule d'ADN par l'ADN-hélicase et son maintien sous forme de simple brin par des protéines stabilisatrices (DBP = **DNA Binding protéines**). Puisque les 2 brins sont antiparallèles, la croissance des brins néosynthétisés n'est pas la même sur les 2 branches (Figure 2):

* *Synthèse continue* sur le brin de la fourche ayant l'extrémité 3'OH libre = brin conducteur ou maître brin.

* *Synthèse discontinue* sur le brin ayant l'extrémité 5'p libre = brin suiveur ou brin retardé: sens opposé à la progression de la fourche par des fragments successives = fragments d'OKASAKI (1000 à 2000 nucléotides).

Les données accumulées montrent que le phénomène de la réplication est réalisé d'une manière similaire chez les eucaryotes et les procaryotes et que celui des procaryotes est le bien élucidé. Les différences entre ces 2 systèmes portent principalement sur les ADN-polymérases, formées souvent de plusieurs sous unités (Tableau 1):

* ADN-polymérase Eucaryotes: α , β , γ , δ et ϵ .

* ADN-polymérase Procaryotes: I, II et III.

Chez les Procaryotes c'est l'ADN-polymérase III qui fait le gros du travail sur les 2 brins et nécessite chaque fois pour démarrer un *facteur d'initiation* = séquence d'ARN amorce (10 nucléotides) synthétisée par l'ARN primase. Chez les Eucaryotes, l'ADN-polymérase α associée à une ADN primase travaille sur le brin retardataire alors que sur le brin conducteur, une autre ADN-polymérase δ spécifique. L'ADN polymérase γ , monomérique, assure la réplication de l'ADN mitochondriale. Chez les Procaryotes, il y a hydrolyse des amorces d'ARN entre les fragments d'OKAZAKI par une *ribonucléase H* et l'ADN-polymérase I va combler ce vide. L'ADN ligase va ensuite lier ces différents fragments.

ADN polymérase	structure quaternaire	activité enzymatique	fonction physiologique
<i>procaryotes</i>			
ADN polymérase I	1 polypeptide (103 kDa)	polymérase exonucléase 5'-3' exonucléase 3'-5'	réplication de l'ADN : élimination des amorces, relecture réparation de l'ADN
ADN polymérase II	1 polypeptide (88 kDa)	polymérase exonucléase 3'-5'	réparation de l'ADN
ADN polymérase III	10 sous-unités $\alpha, \beta, \gamma, \tau, \delta, \epsilon, \theta, \psi, \chi$ dimérique	polymérase exonucléase 3'-5'	réplication de l'ADN relecture
<i>eucaryotes</i>			
ADN polymérase α	5 sous-unités (chez l'Homme)	polymérase	synthèse de courts fragments d'ADN
ADN polymérase β	1 polypeptide	polymérase	réparation de l'ADN
ADN polymérase γ	1 polypeptide	polymérase exonucléase 3'-5'	réplication mitochondriale
ADN polymérase δ	2 sous-unités	polymérase exonucléase 3'-5'	réplication de l'ADN nucléaire réparation de l'ADN
ADN polymérase ϵ	5 sous-unités	polymérase exonucléase 3'-5'	réplication de l'ADN nucléaire réparation de l'ADN

Tableau 1. ADN polymérase procaryotes et eucaryotes

III.1.3. Enzymes et protéines de la réplication

a. Réactions de l'ADN polymérase

Les ADN-polymérase sont des enzymes du noyau cellulaire qui agissent en phase S du cycle pour doubler systématiquement l'ensemble du génome diploïde. D'autres ADN polymérase interviennent dans la synthèse des amorces ARN/ADN, dans la réplication du génome mitochondrial ou dans la réparation de l'ADN (Figure 3). Elles ne peuvent démarrer la condensation des nucléotides que sur la fonction alcool en 3'-OH du ribose du dernier nucléotide d'une amorce, nucléotide qui doit être hybridé avec le nucléotide. Elles utilisent comme substrats des désoxyribonucléotides triphosphates (dATP, dCTP, dGTP et dTTP) et des amorces d'ARN et de l'ADN synthétisées par une primase (ARN polymérase capable de synthétiser un brin d'ARN complémentaire d'un brin d'ADN sur 10 nucléotides) suivie d'une ADN polymérase α (qui prolonge l'amorce d'ARN de 20 désoxyribonucléotides). Elles synthétisent l'ADN nouveau par fragments qui, après excision des amorces, sont liés entre eux par une ADN-ligase. Elles ont aussi une activité exonucléasique, qui leur permet en particulier d'hydrolyser le dernier nucléotide en 3' du brin synthétisé si celui-ci ne s'apparie pas correctement avec le nucléotide complémentaire du brin modèle.

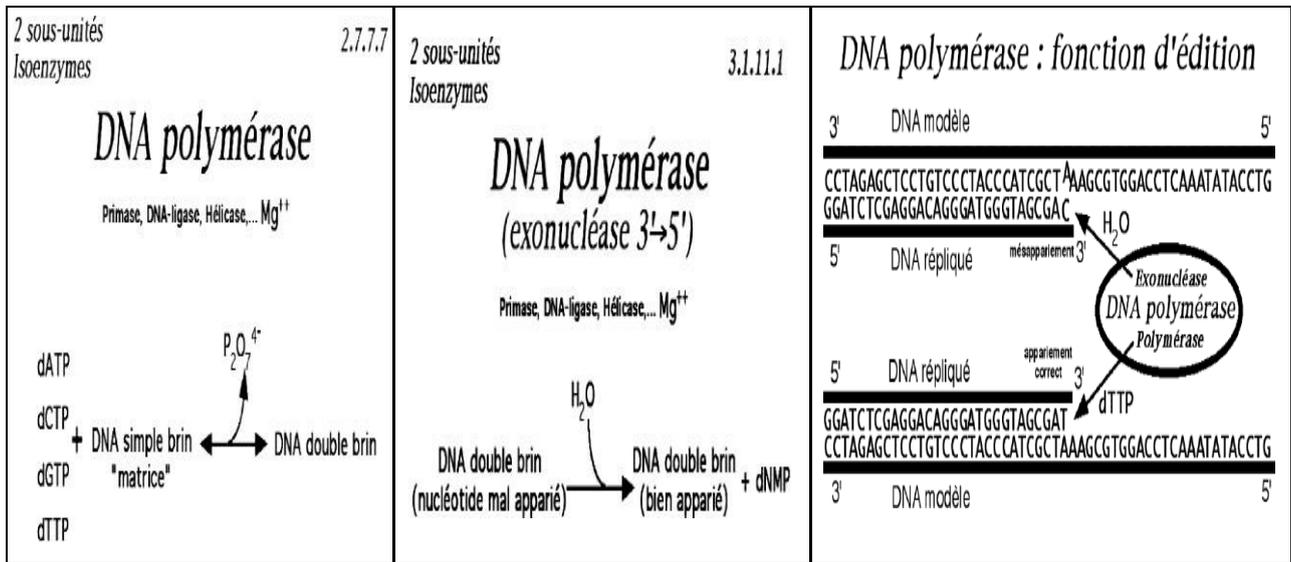


Figure 3. ADN polymérase : réactions de polymérisation, exonucléasique et d'édition

b. Fonction d'édition de l'ADN polymérase

Les ADN polymérases ont à la fois une activité de polymérase pour ajouter des nucléotides au nouveau brin de l'ADN, et une activité de 3' → 5' exonucléase pour hydrolyser le dernier nucléotide incorporé (Figure 3). Si le nucléotide incorporé est complémentaire du nucléotide du brin modèle et donc que l'hybridation des bases se fait normalement, l'activité de polymérase est plus rapide que celle d'exonucléase et le nucléotide suivant va être incorporé. Si le nucléotide incorporé n'est pas complémentaire du nucléotide du brin modèle et donc qu'il y a **mésappariement**, l'activité d'exonucléase est plus rapide que celle de polymérase et ce nucléotide sera hydrolysé.

c. Types d'ADN polymérases

Les ADN polymérases sont des enzymes qui catalysent la synthèse de l'ADN à partir des désoxyribonucléosides triphosphates en recopiant la séquence d'une matrice d'ADN. Certaines d'entre elles ont aussi des activités d'exonucléase qui peuvent s'exercer de 5' vers 3' ou dans l'autre sens, afin de corriger les erreurs d'incorporation (fonction d'édition, Figure 4).

DNA polymérases	MM	Autres activités		pH opt
		5'→3'	3'→5'	
<i>E. coli DNA pol I</i>	109000	+	+	7,4
<i>E.coli Klenow fragment</i>	76000	-	+	8,4 (Tris)
<i>T4 DNA pol</i>	114000	-	+	8,5
<i>T7 DNA pol (Sequenase)</i>	92000	-	-	7,7
<i>Th. aquat. DNA pol (Taq)</i>	94000	-	-	8,3 (Tris)
<i>Reverse transcriptase</i>	84000	-	-	7,6

Figure 4. Les Types d'ADN polymérases.

Les ADN polymérases sont spécifiques de l'ADN double brin et incorporent les nucléotides pour faire la synthèse d'un deuxième brin en utilisant le premier comme matrice. Les ADN polymérases ont besoin d'une amorce (extrémité 3'OH « rentrante » d'un deuxième brin) pour initier la réaction. Tous les êtres vivants ont besoin d'une ADN polymérase pour la répllication de leur ADN. Ce sont en général des protéines d'environ 100.000 daltons. Celles des êtres vivants les plus simples (virus, archéobactéries) sont souvent dépourvues de fonctions d'édition. Toutes les ADN polymérases exigent la présence de l'ion magnésium. Beaucoup d'entre elles fonctionnent en milieu alcalin. Leur température optimale d'action se situe habituellement dans une fourchette de 20 à 40 °C.

d. La reverse transcriptase

Les transcriptases reverses sont des ADN polymérases qui peuvent synthétiser un brin d'ADN complémentaire (ADNc ou cDNA) en prenant un ARN comme matrice, pour former un hybride ADN:ARN (Figure 5). Elles catalysent donc la réaction inverse de la transcription, d'où le nom de transcriptases reverses. Les transcriptases reverses sont produites par des cellules infectées par des rétrovirus, virus à ARN qui font synthétiser un ADNc par la cellule-hôte afin de permettre leur réplication. La transcriptase reverse est utilisée pour l'étude des ARN. Après la synthèse de l'ADN complémentaire, on détruit l'ARN matrice, puis on soumet l'ADNc à l'amplification par la PCR, qui fournit une grande quantité d'ADNc hybridé avec une copie ADN de la séquence de l'ARN de départ.

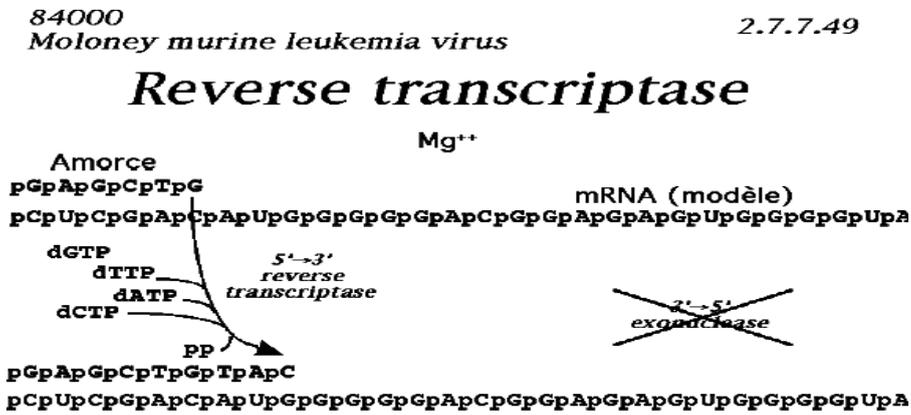
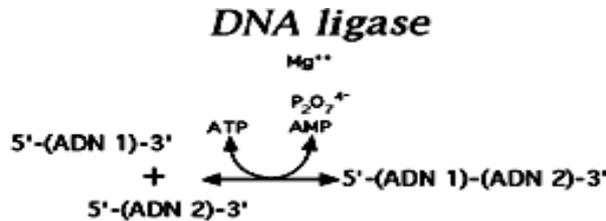


Figure 5. La reverse transcriptase

e. Les ADN ligases

Les ADN ligases sont des enzymes qui sont capables de reconstituer la liaison phosphoester entre le carbone 3'-OH et le phosphate-5' de deux nucléotides voisins sur un brin d'ADN selon la réaction :



f. Les Topoisomérases

La réplication ou la transcription de l'ADN nécessite une fusion partielle de la double hélice et modifie l'enroulement des deux brins. Afin de permettre ces réactions, la topoisomérase est capable de modifier l'enroulement en hydrolysant un brin de l'ADN et en le reconstituant après avoir fait le tour de l'autre brin. Après cette opération, la torsion de la double hélice tend à se rapprocher de la valeur normale : pas de l'hélice = 10 paires de nucléotides par tour. Au cours de la réplication et de la traduction le pas de l'hélice diminue en avant des polymérases et l'hélicase (une des topoisomérases) travaille alors pour augmenter le pas. Au contraire en arrière des polymérases les topoisomérases ajoutent des tours pour reconstituer la double hélice (Figure 6).

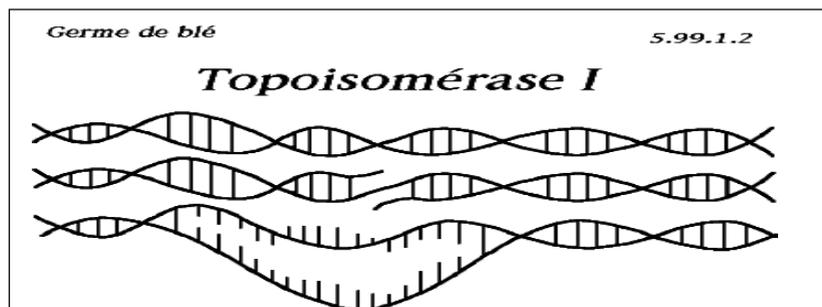


Figure 6. Action des topoisomérases

III.2. Erreurs de réplication de l'ADN et leur réparation

III.2.1. Origines des mutations

Pour survivre et se reproduire, les organismes vivants doivent répliquer fidèlement leur ADN et le protéger des altérations physico-chimiques. Chez l'homme par exemple, des mutations peuvent dérégler le contrôle strict de la division cellulaire, entraîner la prolifération continue de certaines cellules et former une tumeur cancéreuse. Si les mutations touchent l'ADN du spermatozoïde ou de l'ovule des parents (lignée germinale), des maladies génétiques peuvent en résulter chez l'enfant. Des mécanismes moléculaires de correction et de réparation des défauts de l'ADN minimisent ces risques et agissent pour essayer d'éliminer les dommages causés à l'ADN par les erreurs de réplication ou les agents mutagènes apportés par les nuisances environnementales. Il faut admettre cependant que sans ces erreurs de fidélité au cours de la réplication de l'ADN ou de mutations, l'évolution des espèces vivantes à l'origine de ce que l'on nomme la biodiversité ne serait pas cette réalité admise par tous le monde aujourd'hui. En conclusion, les détériorations de séquence de l'ADN d'un individu à un temps donné sont le résultat de la balance qui s'établit entre les événements de mutations et réarrangements et l'efficacité des mécanismes de réparation de ces altérations.

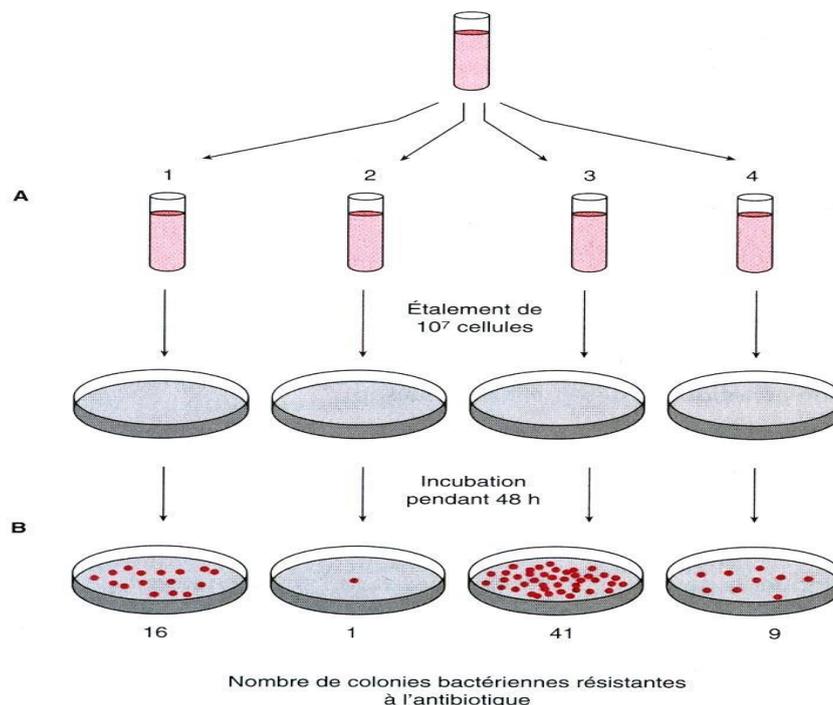


Figure 7. Mise en évidence des mutations spontanées

Les taux de mutations varient selon les espèces. Un taux élevé des mutations peut permettre à certains organismes d'évoluer et donc de s'adapter plus rapidement à leur environnement. La figure 7 montre un exemple avec des mutations bactériennes entraînant l'apparition de résistance aux antibiotiques.

III.2.2. Les mutations liées à des erreurs de duplication de l'ADN

Les mutations les plus simples portent sur la modification d'une base seulement ou mutations **ponctuelles**. Ce sont les **transitions** ou **transversions** (Figure 8). Elles sont l'origine de la notion de polymorphisme d'un seul nucléotide (**SNP** pour *Single Nucleotide Polymorphism*). A ce type de mutations se rattachent aussi les **insertions** ou les **délétions** d'un seul nucléotide.

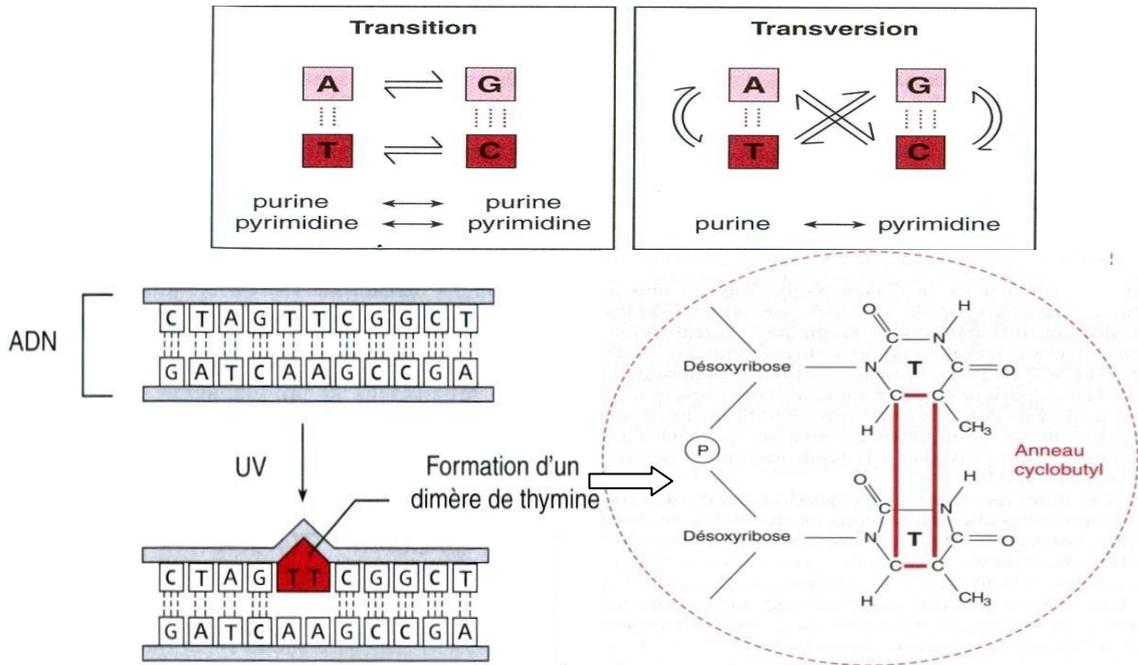


Figure 8. Changements de base pyrimidique ou purique par transition ou transversion et formation de dimères de thymine

Parfois le changement de position d'un atome d'hydrogène dans une base (**tautomérisation**) peut conduire aussi à de faux appariements et à des changements de base au deuxième cycle de réplication (Figure 9). D'autres modifications touchant des fragments d'ADN de taille plus importante, par des événements de **recombinaison** ou de **transposition**, entrent dans la catégorie des mutations chromosomiques.

La fréquence de mutations quel que soit leur type varie de 10^{-6} à 10^{-11} par cycle de réplication de l'ADN avec des régions chromosomiques particulièrement instables où la fréquence est très élevée (= **points chauds** de mutations). Parmi ces régions, les séquences répétées de deux à quatre nucléotides, appelées **microsatellites**. Ce sont des répétitions de bases cytosine et adénine sur un brin d'ADN qui s'écrivent $(CA)_n$ ou en sens inverse sur l'autre brin $(TG)_n$. Lors de la réplication de ces microsatellites, des dérapages de la machinerie de duplication se produisent fréquemment, entraînant des variations dans le nombre 'n' de répétitions effectivement copiées. Pour un microsatellite donné, il est fréquent d'observer parmi les individus d'une même espèce, un fort polymorphisme du nombre de répétitions (ou taille de microsatellites). Il existe plusieurs centaines de milliers de microsatellites dans le génome des eucaryotes. Chez l'homme, le choix approprié de certains d'entre elles et la combinatoire qui en découle dans leur polymorphisme de longueur, constitue un moyen technique efficace (**empreintes génétiques**) pour identifier les caractéristiques génétiques de tout individu de l'espèce concernée.

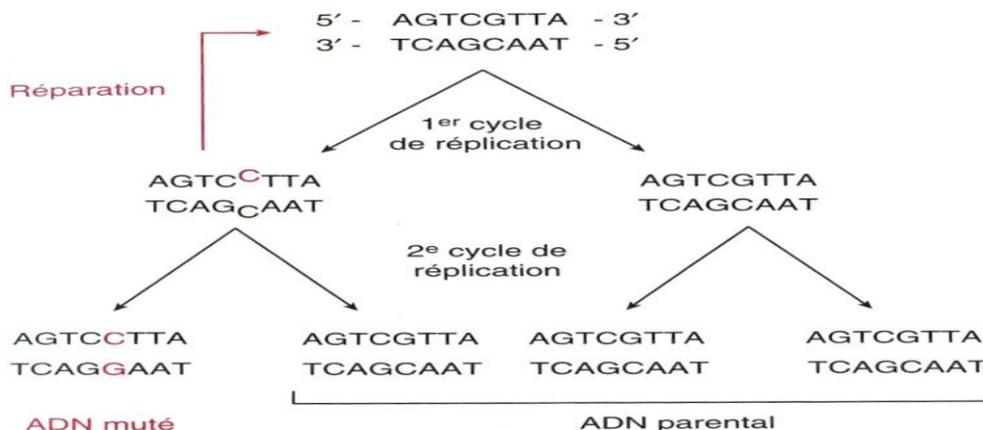


Figure 9. Deux cycles sont nécessaires pour qu'une mutation soit définitivement introduite dans l'ADN

III.2.3. Les mutations provoquées par des agents mutagènes

En plus des erreurs spontanées de réplication, des **agents mutagènes** divers provoquent des détériorations de l'ADN comme les radiations ionisantes et ultraviolettes, certaines substances chimiques (agents alkylants, méthylants, intercalants).

L'hydrolyse spontanée des groupes NH_2 de la cytosine ou de la 5-méthyl-cytosine donne naissance respectivement à l'uracile et à la thymine. Elles s'apparieront alors après un cycle de réplication avec l'adénine et non plus avec la guanine créant ainsi une mutation ponctuelle. L'hydrolyse spontanée de la liaison N-glycosidique unissant les bases puriques au désoxyribose crée des **pseudo-nucléotides** sans guanine ou adénine, dits **apuriniques** qui bloquent la réplication. L'alkylation des bases puriques ou pyrimidiques crée souvent des bases éthylées ou méthylées dont l'appariement naturel est modifié (Ex. la O^6 -méthylguanine qui peut s'apparier avec la thymine provoquant à la réplication un changement de paires G-C en paires A-T).

Les radicaux libres (O^\cdot , OH^\cdot) engendrés à partir de l'oxygène par les radiations ionisantes ou certains agents chimiques attaquent l'ADN et provoquent des **adducts** transformant les bases ciblées en sites hautement mutagènes (Ex. la 7,8-dihydro-8-oxoguanine (oxoG) qui peut s'apparier aussi bien avec la cytosine que l'adénine, provoquant une des mutations par transversion les plus fréquemment rencontrées à l'origine de certains cancers humains).

La lumière ultraviolette (proche UV à 260 nm) est fortement absorbée par les purines et pyrimidines, en raison de leur système de doubles liaisons conjuguées. Cela peut donner un dimère de thymine (Figure 8) qui ne peut pas efficacement s'apparier avec les adénines complémentaires bloquant ainsi l'ADN polymérase lors de la réplication. Les radiations ionisantes (rayons γ et rayons X) provoquent des cassures aléatoires de la double hélice d'ADN, ce qui conduit souvent à la mort cellulaire. Bien maîtrisé, leur emploi thérapeutique sert à limiter la prolifération des cellules tumorales dans le traitement de certains cancers.

Les agents intercalants (polycycliques : bromure d'éthidium, acridine orange) ont des formes aplaties **mimant** les bases azotées appariées. Ces agents hydrophobes s'intercalent facilement entre les plateaux formés par les bases au sein de la double hélice, ce qui provoque lors de la réplication, des insertions ou des délétions de nucléotides. Ces détériorations interfèrent avec la réplication ou la transcription et mettent en cause la survie cellulaire et des systèmes d'identification et de réparation de ces dommages aient été identifiés. Parmi eux des mécanismes d'excision de la base ou du nucléotide défectueux, de réparation par recombinaison quand les deux brins de l'ADN sont en cause. La réparation par excision d'une base incorrecte utilise une **ADN glycosylase** qui reconnaît et élimine la base indésirable en hydrolysant la liaison glycosidique unissant le pentose à la base.

Lorsque la détérioration concerne les deux brins lors d'une cassure complète de l'ADN, la réparation ne peut pas s'effectuer en utilisant l'un des brins comme matrice modèle. Dans ce cas, un ADN homologue est utilisé comme une matrice modèle de secours, à la condition qu'il soit effectivement présent dans la cellule. Si l'ADN polymérase est bloquée au cours de la réplication par une détérioration non réparée, un mécanisme de forçage de la lésion peut alors se déclencher. Il est réalisé par une ADN polymérase spéciale (famille Y), qui synthétise l'ADN directement en ignorant la lésion, sans le recours à l'appariement conforme à une matrice pour incorporer les nucléotides. Cela assure la poursuite de la réplication, essentielle pour la survie de la cellule, même avec un taux élevé de mutations.

Chapitre V. EXPRESSION DE L'INFORMATION GENETIQUE : TRANSCRIPTION ET TRADUCTION

IV.1. Introduction

On désigne par le terme 'expression génétique' le processus moléculaire et cellulaire qui consiste à convertir la séquence ADN des gènes en protéines fonctionnelles. L'expression d'un gène est une suite de synthèses chimiques et de réactions aboutissant à la production d'un ARN ou d'une protéine (Figure 1). Dans un premier temps a lieu la synthèse d'un ARN, dont la séquence est complémentaire d'un des deux brins du gène donc identique à celle de l'autre brin : c'est la **transcription**. La séquence de l'ARN est identique à celle du brin sens et orientée dans le même sens. Elle se construit de 5' vers 3' en complémentarité de celle qui est 'lue' sur le brin antisens. Après des réactions de maturation, le transcrit primaire (ARN) devient ARN messager (ARNm) et sort du noyau vers le cytoplasme. Dans le second temps, l'ARN messager est 'lu' par groupe de trois nucléotides (lettres) grâce à un ARN complémentaire qui porte l'acide aminé correspondant. La 'lecture' de l'ARN se fait de 5' vers 3' et la synthèse de la protéine se fait en même temps de l'extrémité NH₂ terminale vers l'extrémité COOH terminale. Après des réactions de maturation, la protéine 'mature' est prête à remplir sa fonction dans la cellule.

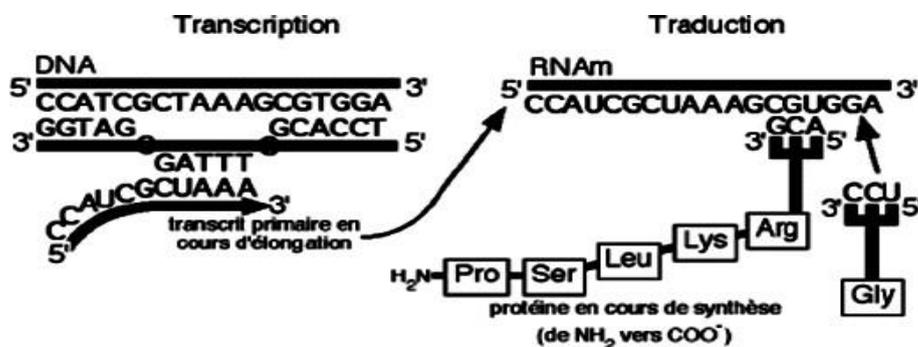


Figure 1. Expression d'un gène

IV.2. Transcription

La transcription est la lecture d'un gène par une ARN-polymérase qui synthétise un acide ribonucléique ou ARN dont la structure primaire reproduit celle du brin 'sens' de ce gène. La transcription est conduite par plusieurs enzymes :

- * ARN-polymérase I qui synthétise les ARN cytoplasmiques : ARN ribosomiques (18 S, 5,8 S et 28 S);
- * ARN-polymérase II qui synthétise les ARN messagers qui contiennent l'information destinée à la traduction et certains des snARN;
- * RNA-polymérase III qui synthétise les petits RNA (ARNt, ARNr 5S, snARN et 7SLRNA).

IV.2.1. Promoteur et Brins sens ou antisens

Les ARN polymérases se fixent à des régions de l'ADN désignées sous le terme de promoteurs, souvent avec l'aide de protéines additionnelles. La dernière ligne de la séquence nucléotidique de la figure 40 est la séquence du premier exon du gène de l'apolipoprotéine A-II. Les 900 nucléotides qui précèdent contiennent de nombreuses séquences reconnues par des protéines nucléaires qui permettent le début de la transcription ou la régulation de cette étape. On distingue en particulier :

- * vers 20 à 30 nucléotides en amont de l'exon 1 en direction de l'extrémité 5' du brin sens une séquence TATATA appelée '**boîte TATA**', qui est spécifiquement reconnue par la protéine TFIID, cofacteur de la RNA-polymérase II;
- * vers -80 une séquence GGCCATCCAT à la fois '**boîte CAT** ou **CAAT**' et '**boîte GC**', sur laquelle viennent se fixer d'autres protéines de la transcription (CTF et Sp-1);
- * de nombreuses autres séquences où se fixent des protéines régulatrices de la transcription. Par exemple, une séquence TRE (TGACTCA) au début de la troisième ligne de cette image, sur laquelle vont se fixer des protéines de la famille AP-1 pour activer la transcription.

TATTTAACTGATTTACCCCAAATGCTTTGAACCTGGGAATGTACCTCTC
 CCCCTCCCCACCCCAACAGGAGTGAGACAAAGGGCCAGGGCTATTGCC
 CTGCTGACTCAATATTGGGCTAATCACTGCCTAGAAGCTGATAAGGTGATC
 AAATGACCAGGTGCCTTCAACCTTTACCTGGTAGAAGCCTCTTATTCA
 CCTCTTTTCTGCCAGAGCCCTCCATTGGGAGGGGACGGGCGGAAGCTG
 TTTTCTGAATTTGTTTTACTGGGGGTAGGGTATGTTTCAGTGATCAGCAT
 CCAGGTCATTCTGGGCTCTCCTGTTTTCTCCCCGCTCTCATTACACATTA
 ACTCAAAACGGACAAGATCATTACACTTGCCTCTTACCCGACCCCTC
 ATTCCCCTAACCCCATAGCCCTCAACCTGTCCCTGATTTCAATTCCCT
 TTCCTCTTCTCTGCTCCCAATATCTCTCGCCAAGTTGCAGTAAAG
 TGGGATAAGGTTGAGAGATGAGATCTACCCATAATGGAATAAAGACACC
 ATGAGCTTTCCATGGTATGATGGGTTGATGGTATTCATGGGTTGATAT
 GTCAGAGCTTTCCAGAGAAATAACTTGAATCCTGCTTCCCTGTTGCATT
 CAAGTCCAAGGACCTCAGATCTAAAAGAAATGAACCTCAAATATACCTGA
 AGTGATCCCTTAGCTCCACTAAGAGCTGTACCCCTGCCTCTCACC
 CCATCACCATGAGTCTTCCATGTCTTGTCTCTCCTCCCCATTCTC
 CAACTTGTATTATCCTCACATAATCCTGCCCACTGGCCCATCCATAGT
 CCCTGTACCTGACAGGGGTGGGTAACAGACAGGTATATAGCCCTTC
 CTCTCCAGCCAGGGCAGGCACAGACACCAAGGACAGAGACGCTGGCTAG

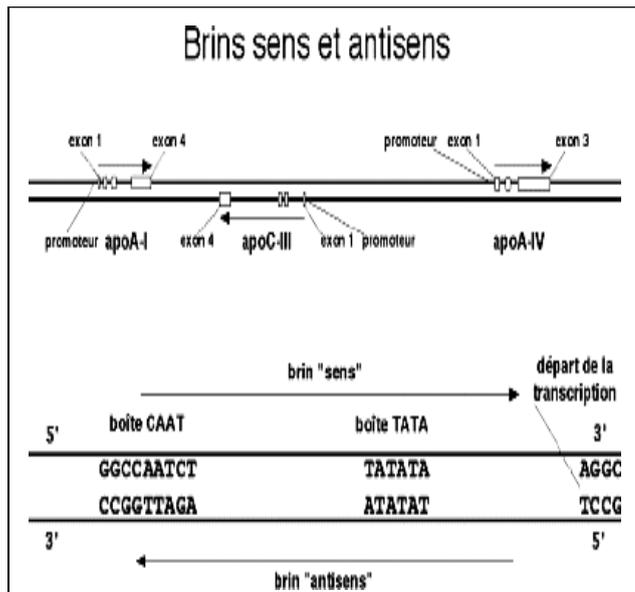


Figure 2. Séquences régulatrices du promoteur et brins sens et antisens

Les gènes sont situés sur les deux brins du DNA, de gauche à droite ou de droite à gauche, indifféremment. Ainsi les gènes des apolipoprotéines A-I et C-III (Figure 2) qui sont voisins sont orientés en sens opposés. Le début du message (exon 1) est à gauche pour l'apoA-I et à droite pour l'apoC-III. Le message codé doit être lu en commençant par le début donc dans un sens différent pour ces deux gènes. Pour servir de modèle la transcription doit faire la copie de la séquence d'un brin de l'ADN, qualifié de brin 'sens'; l'autre brin en est le complémentaire on le dit 'antisens'. La transcription se fait en synthétisant un RNA complémentaire de la séquence du brin antisens, donc identique à celle du brin sens. Lors de la formation du complexe d'initiation de la transcription la fixation de TFIID sur la boîte TATA se fait en premier et sur les deux brins de façon symétrique. Lors de la fixation des autres protéines du complexe (NF-I ou NF-Y sur la boîte CAAT) il y a un brin sur lequel la boîte TATA est en aval (côté 3') de la boîte CAAT, c'est le brin 'sens' qui contient le message et un brin sur lequel la boîte CAAT est en aval (côté 3') de la boîte TATA, c'est le brin 'antisens' qui doit servir de modèle pour la transcription.

IV.2.2. Régulation de l'expression des gènes chez les procaryotes

La cellule bactérienne est capable d'adapter très rapidement son métabolisme aux modifications de l'environnement. Prenons une suspension de cellules d'*Escherichia coli*. Une partie de la suspension est cultivée dans un milieu minimal contenant du lactose (disaccharide composé de deux molécules de glucose) comme seule source de carbone (milieu A). Une autre partie est cultivée dans le même milieu minimal mais contenant du glucose comme source de carbone et auquel on a ajouté les 20 acides aminés (milieu B). On constate que les cellules du milieu A croissent (prolifèrent), mais moins vite que celles du milieu B. Elles doivent en effet assurer le clivage du lactose en deux monosaccharides (glucose et galactose) assimilables et toute la biosynthèse des acides aminés. L'étude biochimique et génétique de l'utilisation du lactose par *E. coli* va nous montrer comment l'économie réalisée par les cellules cultivées dans le milieu B (de génotype identique à celles cultivées sur le milieu A) se fait, par adaptation du programme d'expression des gènes, dès le niveau transcriptionnel. Ces résultats sont le fruit des travaux de Jacob, Monod et leurs collaborateurs dans les années 60.

a. Opéron 'inductible'

Ainsi, la présence de lactose semble déclencher la synthèse d'enzymes spécifiques : la β -galactosidase, codée par le gène *LacZ*, mais également une **perméase** (codée par le gène *LacY*), capable d'accélérer la pénétration du lactose dans la cellule et une **transacétylase** codée par le gène *LacA*, qui purge la cellule des **thiogalactosides** toxiques introduits par la perméase (Figure 3). La première fonction de la β -galactosidase est de cliver le lactose en glucose et galactose, sa seconde fonction est de convertir la liaison 1-4 entre glucose et galactose (au sein du lactose) en une liaison 1-5 dans l'**allolactose**, qui joue un rôle dans la régulation de l'opéron Lac. Ces trois activités enzymatiques sont **co-régulées** (elles apparaissent en réponse à une même induction). Des méthodes très fines de cartographie montrent que ces gènes sont adjacents dans une région 'Lac' du chromosome bactérien. Plusieurs types de mutation peuvent interférer avec l'utilisation du lactose.

* Des mutations dans le gène Z empêchent le catabolisme du lactose par absence de β -galactosidase, des mutations dans le gène Y empêchent la pénétration active du lactose à l'intérieur de la cellule. Dans ces deux cas le phénotype, symbolisé par *lac⁻*, c'est à dire impossibilité d'utiliser le lactose, résulte d'un défaut enzymatique.

* D'autres mutations perturbent d'une façon tout à fait différente l'utilisation du lactose : elles affectent la production de trois enzymes en même temps (β -galactosidase, perméase et transacétylase). Plusieurs de ces mutations entraînent un phénotype constitutif c'est-à-dire production d'enzymes en absence d'inducteur.

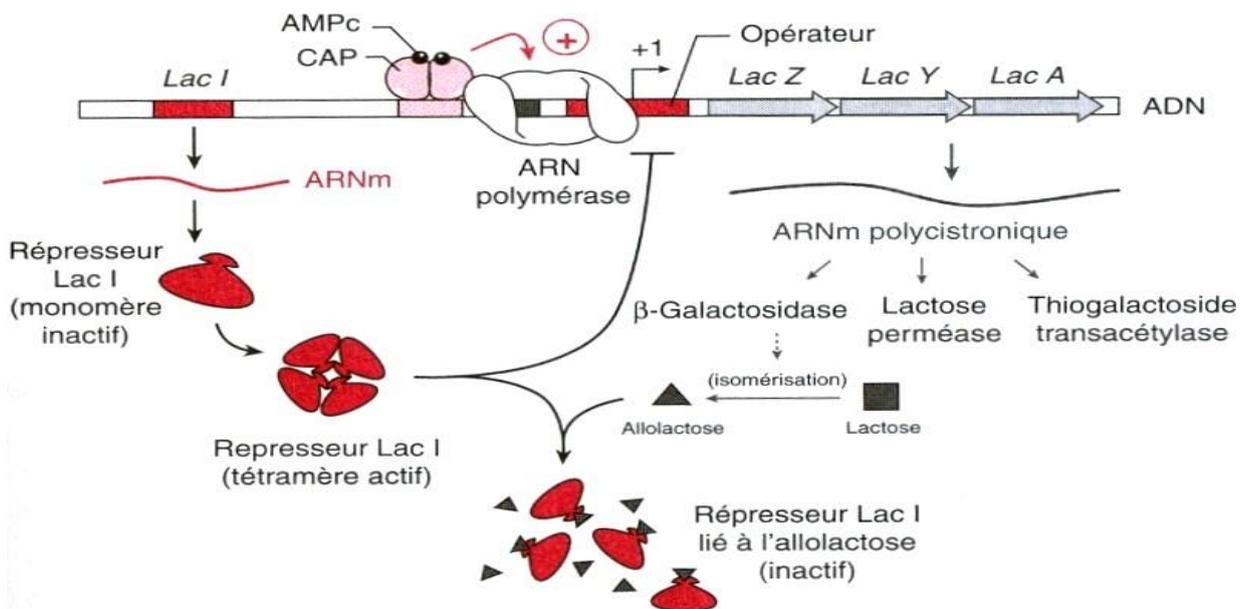


Figure 3. Positionnement des divers partenaires fonctionnels de l'opéron lac. La transcription des 3 gènes de l'opéron n'est effective que lorsque le lactose est présent, quand le répresseur est inactif et l'opérateur inoccupé. La protéine CAP complexée à l'AMPc stimule le recrutement de la polymérase uniquement en l'absence du glucose.

Etant donné que les trois gènes cartographient côte à côte sur le chromosome, on peut supposer que leur expression est régulée par un même système et que les mutations constitutives affectent non pas une production d'enzyme mais un élément de contrôle. Cet ensemble forme ce que l'on appelle un **opéron bactérien**. Les premières mutations constitutives étudiées ont été appelées **I⁻** et sont localisées près du gène Z, par la suite on a caractérisé des mutants constitutifs **O^c**, ces mutations sont situées encore plus près de Z. L'analyse de ces mutants par des expériences très élégantes faisant appel à des diploïdes partiels a permis au groupe de **Monod et Jacob** d'élaborer le célèbre modèle de régulation de l'opéron lactose par répression de la transcription. Le tableau 1 nous rappelle les caractéristiques phénotypiques des mutants utilisés.

Génotype	Phénotype
$I^+ O^+ Z^+ Y^+$	sauvage (inductible)
$I^+ O^+ Z^- Y^+$ $I^+ O^+ Z^+ Y^-$	Lac^- (non inductible)
$I^- O^+ Z^+ Y^+$ $I^+ O^c Z^+ Y^+$	constitutif

Tableau 1. Mutants phénotypiques de l'opéron lac utilisés par Monod et Jacob

Une confirmation d'un mode de régulation par répression a été apportée par des mutations I^s ('s' pour 'super-réprimé'). Ces mutants sont incapables d'utiliser le lactose car l'opéron est réprimé en permanence, le lactose ne peut induire la transcription. Les mutations affectent bien le gène I, mais dans une région importante pour la formation du complexe répresseur-inducteur. La région essentielle pour la liaison du répresseur à la séquence opératrice étant intacte, l'état réprimé est stable. Ces observations permettent d'aborder un aspect plus général des protéines de régulation, celui de domaines fonctionnels spécialisés. L'analyse détaillée de la protéine après clonage du gène **I** dans un vecteur d'expression et surproduction par des clones bactériens transformés confirmera ce concept. Les mutations I^- , I^s et d'autres ne sont pas disposées de façon aléatoire dans le gène mais leur cartographie reflète le fait qu'une partie de la protéine codée est essentielle dans la reconnaissance du site opérateur, une autre région est nécessaire pour la fixation de l'inducteur et la transition allostérique, une autre permet la formation d'un tétramère. Cette notion de protéines de régulation modulaires, séparables en domaines fonctionnels, se retrouvera chez les Eucaryotes. L'opéron lactose ne fonctionne que s'il est induit par une molécule effectrice (le lactose) capable de lever l'effet du répresseur, il est dit **inductible**.

L'analyse des allèles du promoteur ont également permis de mieux comprendre le fonctionnement de l'opéron Lactose (P^s : augmente l'affinité de reconnaissance pour ARN Polymérase, augmente le niveau de transcription de l'opéron lac; P^l ^{cr}: affecte le site de fixation CAP-AMPc et réduit le niveau d'expression des gènes de l'opéron lac d'environ 10% par rapport et type sauvage (P^+); i cr insensible à la répression du catabolisme).

b. Contrôle négatif et contrôle positif : cas des opérons lactose ou tryptophane

Les gènes soumis à un contrôle négatif ne sont pas transcrits si un répresseur est lié à l'opérateur. Les gènes soumis à un contrôle positif ne sont transcrits efficacement que si une protéine régulatrice favorise l'initiation. L'opéron lactose est également soumis à un contrôle positif. Il a été précisé au début de ce chapitre que la transcription est induite si l'on remplace le glucose par du lactose comme seule source de carbone dans le milieu minimal. En effet, si l'on ajoute du lactose, l'opéron n'est pas transcrit tant que le glucose n'est pas épuisé. Ce phénomène, qui concerne de nombreux opérons du catabolisme est appelé '**effet glucose**' ou encore 'répression catabolique'. Il repose, comme les autres modes de régulation sur une protéine à régulation allostérique et une molécule effectrice. Deux types de mutations abolissent l'effet glucose: le premier concerne les gènes impliqués dans la formation de l'adénosine monophosphate cyclique (AMPc) notamment celui de l'adénylate cyclase, le second dans un gène codant pour une protéine appelée '**CAP**' comme protéine activatrice du catabolisme (Figure 4). On a pu montrer que le glucose freine la production d'AMPc à partir de l'ATP et maintient un très faible niveau d'AMPc. Lorsque le glucose diminue, la concentration en AMPc augmente, or cette molécule peut former un complexe spécifique avec la protéine CAP. Il en résulte une modification de la structure tridimensionnelle et le complexe est capable de se fixer sur l'ADN, au niveau d'une séquence particulière appelée site CAP, située un peu en amont du promoteur. La liaison entraîne une contrainte topologique de la double hélice d'ADN qui favorise l'initiation de la transcription.

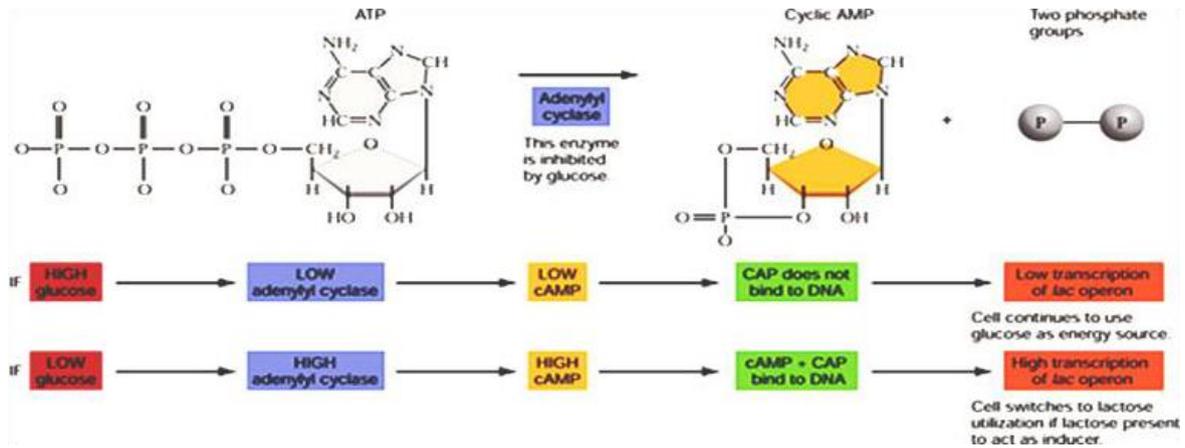


Figure 4. Régulation de l'opéron lac par le glucose

c. Opéron 'Répressible'

L'opéron lactose est **inductible** car ne fonctionne que s'il est induit par une molécule effectrice capable de lever l'effet du répresseur. Dans d'autres cas la molécule effectrice provoque la répression de la transcription, l'opéron est dit répressible. Un exemple typique nous est fourni par un opéron intervenant dans la biosynthèse du tryptophane. Il est à remarquer que l'opéron lactose intervient dans le catabolisme, l'**opéron tryptophane** intervient dans l'anabolisme. Cet opéron comporte cinq gènes codant pour des enzymes impliquées dans la synthèse du tryptophane (les cistrons trpE, D, C, B et A) groupés, sous la dépendance d'un seul système promoteur-opérateur (Figure 5). Un gène, trp R, qui ne fait pas partie de l'opéron, code pour un répresseur spécifique de l'opéron tryptophane. Cette protéine est incapable de se lier au site opérateur, et par conséquent inactive, tant qu'elle n'est pas complexée avec une molécule effectrice : le tryptophane lui-même. Il agit donc comme un co-répresseur dans ce mécanisme de régulation en retour par le produit final de la chaîne métabolique de l'opéron.



Figure 5. L'opéron tryptophane

IV.2.3. Régulation de l'expression des gènes chez les eucaryotes : Cis/Trans régulateurs et DNA binding proteins

La régulation de toute voie métabolique se fait principalement à la première réaction pour ne pas synthétiser d'intermédiaires inutiles. Pour l'expression d'un gène, la régulation a lieu à 4 niveaux (Figure 6) : lors de transcription, maturation du transcrit, traduction ou lors de l'activation de la protéine mature. Le point de **contrôle principal** est la transcription concerne l'ARN polymérase, l'enzyme-clé de l'expression d'un gène.

Les séquences de nucléotides du promoteur (facteurs **cis-régulateurs**) sont reconnues par une classe de protéines spécifiques (facteurs **trans-régulateurs**) dont la structure permet une liaison avec l'ADN (Figure 7, *DNA binding proteins*). Les facteurs trans-régulateurs ont des effets sur la vitesse de la transcription : activateurs (séquences **enhancer**) ou inhibiteurs (séquences **silencer**).

Leur liaison avec l'ADN dépend le plus souvent de circonstances physiologiques qui induisent ou répriment l'expression du gène. Il existe des éléments essentiels et présents dans la plupart des gènes : * élément principal comme la boîte TATA,

* éléments de base comme l'octamère reconnu par le facteur OCT-1 présent dans presque toutes les cellules.

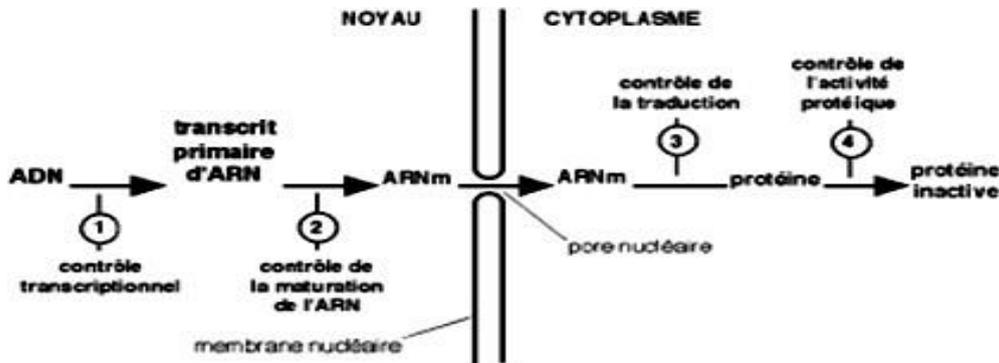


Figure 6. Niveaux de régulation de l'expression des gènes eucaryotes

Il existe des éléments qui répondent à des facteurs dont la présence dépend des signaux qui parviennent à la cellule, principalement les hormones comme l'insuline ou les seconds messagers comme l'AMP cyclique. Il y a encore des éléments spécifiques d'un type cellulaire permettant l'expression différente des gènes dans chaque tissu. Ainsi, le promoteur de la lipoprotéine lipase contient la boîte TATA, des éléments du promoteur de base, des éléments de réponse aux hormones et des éléments d'expression tissulaire spécifique. Dans le promoteur des gènes il y a souvent 20 ou 30 sites de fixation pour des facteurs trans-régulateurs dont beaucoup ont un effet inducteur ou répresseur sur l'expression du gène.

Les protéines qui reconnaissent et se lient à l'ADN (*DNA binding proteins*) sont des enzymes, des protéines de structure, des facteurs de transcription et des éléments trans-régulateurs. Dans la structure spatiale de ces protéines, on reconnaît des domaines propres à cette liaison spécifique :

* Le domaine **hélice-boucle-hélice** permet le lien spécifique des hélices α avec les nucléotides dans le grand sillon sur une longueur de 10 paires de bases environ.

* Les **doigts de Zinc** sont des domaines formés d'un ion Zn^{++} tétracoordonné avec deux histidines (H) et deux cystéines (C) de la protéine. Chaque doigt de Zinc se lie à cinq nucléotides dans le grand sillon de l'ADN. Il y a souvent plusieurs doigts de Zinc à la suite dans la même protéine.

* Certaines protéines ont un domaine riche en acides aminés basiques (Lys,Arg) qui se lie aux phosphates des nucléotides. Ces protéines se lient à des séquences répétées inverses sur l'ADN, lorsqu'elles sont associées en dimères par la liaison hydrophobe de deux domaines riches en leucine (fermeture 'Eclair' à 'leucines').

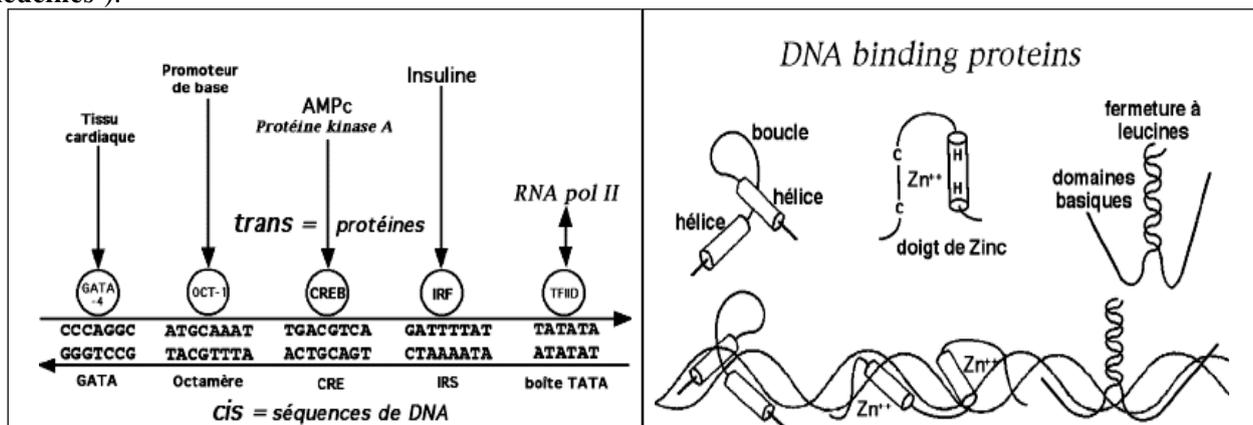


Figure 7. Séquences du promoteur et les DBP (DNA binding proteins)

IV.2.4. Mécanisme de la transcription

L'ARN polymérase II est l'enzyme de la transcription des gènes exprimés sous forme de protéines. Elle est présente dans tous les noyaux cellulaires. Sa masse moléculaire est de 500.000 daltons pour 10 sous-unités. La transcription qu'elle catalyse nécessite des ribonucléosides triphosphates comme substrats (ATP,

CTP, GTP et UTP), plusieurs cofacteurs protéiniques (*transcription factors* TFIIA à TFIIJ). L'énergie de la réaction est fournie par l'hydrolyse des liaisons riches en énergie des nucléosides triphosphates (42 kJ/mol). Le mécanisme de la transcription se fait en 3 étapes (initiation, élongation et terminaison) dont l'initiation est l'étape cruciale.

a. Initiation de la transcription

L'initiation de la transcription sur un promoteur humain implique plusieurs facteurs de transcription, connus sous les noms de TFIIA, TFIIB, TFIID, TFII E, TFII F, TFII H et TFII J en plus de l'ARN polymérase II elle-même. Au centre de ce mécanisme initial, il y a l'association de TBP, la sous-unité reconnaissant l'ADN du facteur TFIID, avec la boîte TATA du promoteur. La protéine TFIID se fixe sur l'ADN, en reconnaissant l'élément TATA et déforme la double hélice de façon importante et constitue le complexe d'initiation de la transcription. Cet ensemble va servir de point d'ancrage pour les autres facteurs d'initiation de la transcription. L'adjonction successive de TFIIB et de RAP30, petite sous-unité de TFII F, sont ensuite nécessaires pour la fixation de la polymérase et déterminent à la fois le brin transcrit et le sens de la transcription. A ce complexe ADN-TFIID-TFIIB-RAP30-polymérase II, s'ajoutent encore successivement la grande sous-unité de TFII F (RAP74), TFII E et TFII H. Alors la transcription peut commencer si les ribonucléosides substrats sont présents.

Le complexe d'initiation de la transcription recouvre une séquence d'environ 100 nucléotides du brin antisens de l'ADN. TFII H provoque l'ouverture et le déroulement partiel de la double hélice à cet endroit. La transcription commence à environ 22 nucléotides de la boîte TATA en direction opposée à celle des éléments GC-CAAT et se poursuit en remontant le brin antisens en direction de son extrémité 5'. La formation de ce complexe est activée ou inhibée par les facteurs trans-régulateurs liés aux séquences cis-régulatrices du même promoteur.

Les activateurs ou inhibiteurs de la transcription (facteurs trans-régulateurs) sont des protéines produites par d'autres gènes qui interviennent pour activer ou pour inhiber l'ARN polymérase II et par suite induire ou réprimer l'expression du gène à transcrire. Ces facteurs trans-régulateurs se lient au DNA (*DNA binding proteins*) dans la région du gène située en amont de la partie transcrite. Les sites de fixation de ces facteurs trans-régulateurs sur l'ADN sont des séquences spécifiques appelées éléments cis-régulateurs. Le couple élément cis-régulateur plus le facteur trans-régulateur lié entre en contact avec le complexe d'initiation de l'ARN polymérase par l'intermédiaire d'un ensemble de protéines appelé médiateur.

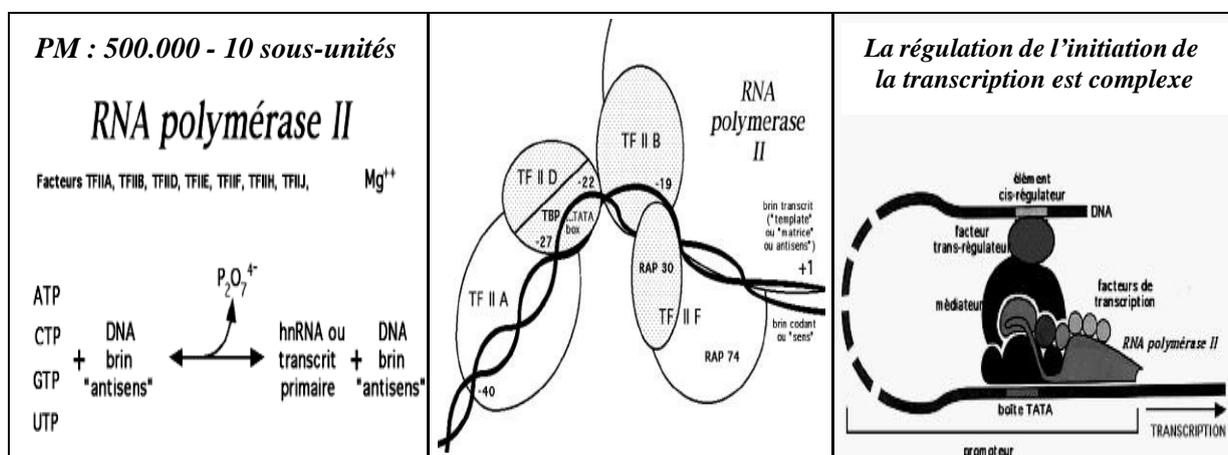


Figure 8. Régulation de l'initiation de la transcription par l'ARN polymérase

L'ARN polymérase peut donc démarrer la transcription lorsqu'elle est activée par les facteurs de transcription dont certains sont liés à ces éléments comme les boîtes TATA ou CAAT et lorsque par l'intermédiaire du médiateur elle reçoit un signal d'activation ou d'inhibition. Ce signal dépend de la présence d'un facteur trans-régulateur lié à un autre élément du promoteur. La liaison du régulateur et du médiateur nécessite un repliement de l'ADN du promoteur pour permettre la liaison de ces protéines (Figure 8).

b. Elongation de la transcription

Chaque nucléoside triphosphate est choisi spécifiquement pour être complémentaire de la base du brin antisens qui va lui faire face. La liaison riche en énergie entre le premier phosphate estérifiant le carbone 5' du ribose et les deux autres phosphates est hydrolysée, libérant un pyrophosphate qui sera hydrolysé ensuite par

une **pyrophosphatase**. Le phosphate restant est lié par une liaison ester au carbone 3' libre du dernier nucléotide du transcrit en cours de synthèse (Figure 9). L'ARN polymérase construit un RNA hybridé avec le brin antisens de l'ADN, dont la séquence primaire est la copie du brin sens mais composée de ribonucléotides au lieu des désoxyribonucléotides et d'uracile à la place des thymines. Au cours de cette élongation l'ARN polymérase II est accompagné d'une série de facteurs d'élongation qui modifient la chromatine pour permettre l'avancée de l'enzyme, empêchent la formation d'obstacles comme des épingles à cheveux ou facilitent l'avancée de la polycondensation.

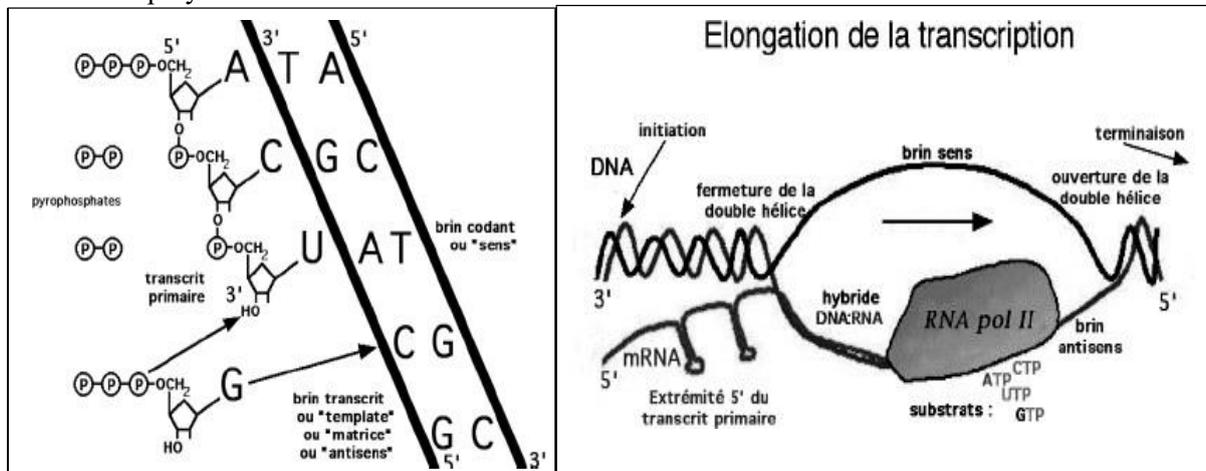


Figure 9. Elongation de la transcription

La transcription commence au point d'initiation (environ 25 nucléotides avant la boîte TATA sur le brin antisens) par l'hybridation et la condensation des premiers ribonucléotides du transcrit primaire (pré-ARNm). La double hélice est ouverte en une boucle où se place l'ARN polymérase II. Le transcrit primaire reste hybridé sur quelques nucléotides avec le brin antisens puis s'en détache pour permettre à la double hélice de se reformer après le passage de la polymérase. L'extrémité 5' du transcrit primaire libérée se condense en diverses structures secondaires (épingles à cheveux). Les substrats de l'ARN polymérase sont les quatre ribonucléosides triphosphates. La polymérase lit le brin antisens en allant dans le sens 3' → 5'. Elle poursuit la synthèse du transcrit primaire en condensant les ribonucléotides jusqu'à ce qu'elle rencontre les signaux de terminaison. A la rencontre des signaux de terminaison, la polymérase se détache de l'ADN, libère le transcrit primaire et la double hélice se referme.

c. Fin de la transcription

Vers la fin du gène à transcrire, des facteurs liés à l'ARN polymérase reconnaissent sur le brin antisens une séquence 3'-TTATTT-5' suivie dans la plupart des gènes d'un autre signal 3'- ATACAAAC-5' qui libèrent l'ARN polymérase. La transcription s'arrête en effet peu après le premier signal et la fin du transcrit s'écrit 5'-AAUAAAUGCUGAAUGAAUCC-3'OH. L'ARN polymérase ayant libéré l'ADN et le transcrit primaire qui contient la copie de l'information génétique qui va permettre l'expression du gène sous forme de protéine.

IV.2.5. Régulations post-transcriptionnelles

a. Discontinuité des gènes et notion d'exon-intron

Les molécules de l'ADN de chaque chromosome sont très longues (jusqu'à plusieurs centaines de millions de paires de nucléotides et contiennent plusieurs milliers de gènes). Chez l'homme, les gènes sont très dispersés et ne représentent que 10 % de l'ADN génomique total. Chaque gène comprend des régions qui contiennent les séquences codées qui seront traduites, les **exons**; mais aussi, des régions de régulation comme le promoteur et des régions non traduites appelées **introns** qui séparent les exons. Un gène peut ne contenir qu'un seul exon et pas d'intron, mais il y a des gènes de plus de cent exons. Après la transcription l'ARN produit de l'ARN polymérase ou transcrit primaire contient encore les introns et les exons, mais pas le promoteur. Après l'excision des introns et l'épissage des exons, le messenger ne contient plus que les exons réunis en une seule séquence codante.

Les exons sont des fragments de séquence primaire d'un gène qui seront recopiés dans la structure primaire du RNA messenger, après l'épissage. Il existe des exons de longueur très variable : courts (34 nucléotides pour l'exon 1 de l'apoAII) ou longs (7572 nucléotides pour l'exon 26 de l'apoB). Un exon code le plus souvent pour un domaine fonctionnel de la protéine ou une partie d'un tel domaine, de sorte que les

protéines des eucaryotes formées de plusieurs domaines, sont codées par des gènes possédant au moins autant d'exons. Au cours de l'évolution, le gène peut être modifié par la substitution, l'insertion ou la délétion d'un exon entier (conversion de gènes) ce qui modifie, apporte ou retire à la protéine un domaine fonctionnel en entier.

Après la fixation de l'ARN polymérase sur la boîte TATA par l'intermédiaire du facteur TFIID, la transcription va commencer 23 nucléotides au delà de cette boîte. A ce point, la séquence du brin sens est 5'AGGCACAGA...3', celle du brin antisens est donc 3'TCCGTGTCT...5' (complémentaire et antiparallèle). En choisissant comme substrats les ribonucléotides complémentaires de ce brin antisens, l'ARN polymérase va composer 5'AGGCACAGA...3' qui sera le début de la séquence de l'ARN transcrit ou exon 1. Lorsque la polymérase rencontre un A sur le brin antisens, elle incorpore un nucléotide à Uracile dans l'ARNm: 5'...GCUGGCUAG...3'.

La transcription commence au point d'initiation (environ 25 nucléotides avant la boîte TATA sur le brin antisens) par l'hybridation et la condensation des premiers ribonucléotides du transcrite primaire (pré-ARNm). La double hélice est ouverte en une boucle où se place l'ARN polymérase II. Le transcrite primaire reste hybridé sur quelques nucléotides avec le brin antisens puis s'en détache pour permettre à la double hélice de se reformer après le passage de la polymérase. L'extrémité 5' du transcrite primaire libérée se condense en diverses structures secondaires (épingles à cheveux). Les substrats de l'ARN polymérase sont les quatre ribonucléosides triphosphates. La polymérase lit le brin antisens en allant dans le sens 3' → 5'. Elle poursuit la synthèse du transcrite primaire en condensant les ribonucléotides jusqu'à ce qu'elle rencontre les signaux de terminaison. A la rencontre des signaux de terminaison, la polymérase se détache de l'ADN, libère le transcrite primaire et la double hélice se referme.

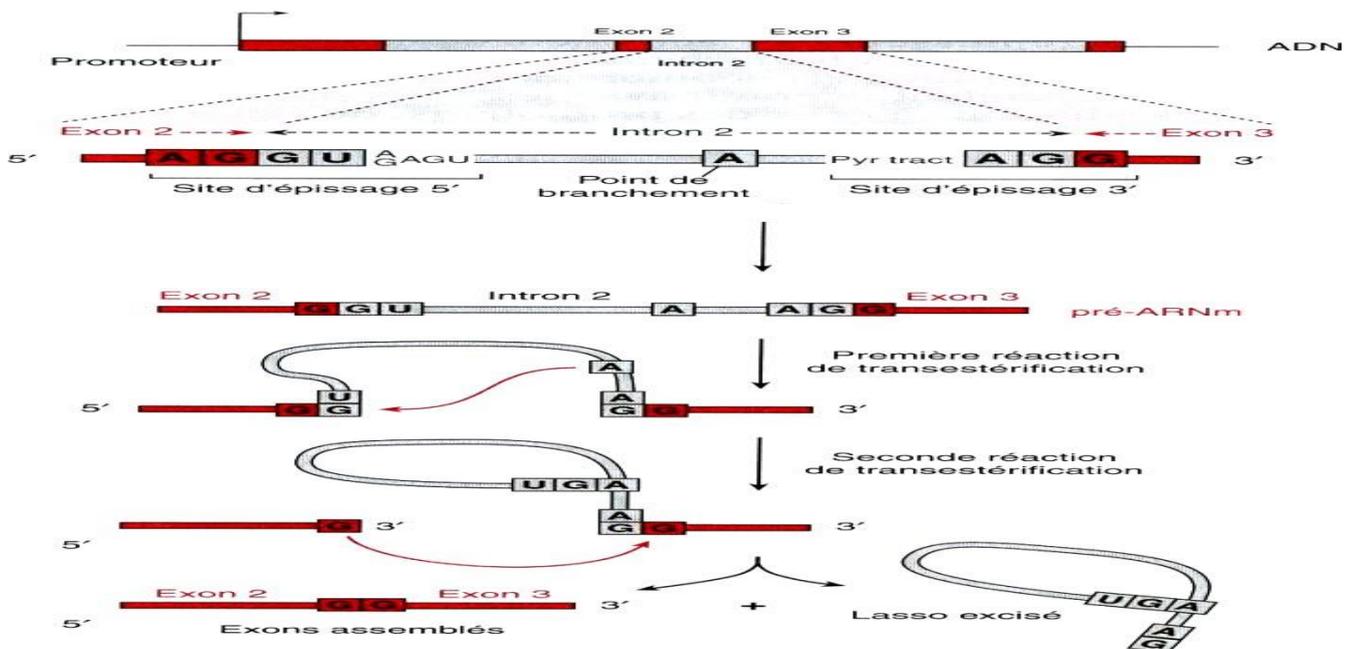
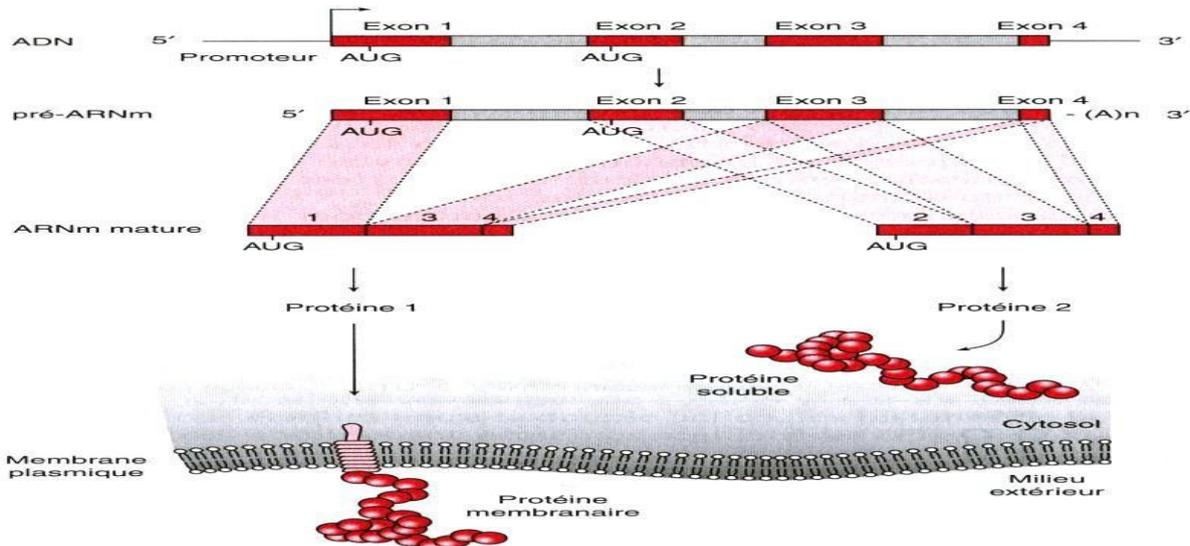


Figure 10. Gènes morcelés et excision-épissage des ARN premessagers

Après libération de l'ARN, les introns, parties non codantes des gènes des eucaryotes, sont coupés (excision) de la structure primaire des ARN au cours de la maturation des messagers. Les exons, parties codantes, sont ensuite liés entre eux bout à bout (épissage), pour établir la séquence primaire de l'ARN messager. L'excision et l'épissage représentent donc l'action d'enzymes et de ribozymes qui catalysent la coupure du RNA (endoribonucléase, figure 10) et la fermeture de la brèche (ARN ligase). L'excision des introns et l'épissage des exons est l'étape la plus importante de la maturation du transcrite dans le noyau cellulaire. Elle fait appel à des facteurs spécifiques : ribonucléoprotéines contenant des petits ARN (snRNP), ribozymes (ARN ayant des propriétés catalytiques), enzymes spécifiques (maturases).

La structure secondaire du transcrite met en contact trois séquences de l'intron : la séquence du début de l'intron (extrémité 5' ou site donneur), la séquence de la fin de l'intron (extrémité 3' ou site accepteur) et un nucléotide à adénine (A du branchement) environ 40 nucléotides avant le site receveur. Le site donneur

commence habituellement par un nucléotide à guanine dont le phosphate est détaché de l'exon précédent pour être transféré sur le carbone 2' de l'A du branchement. Le dernier nucléotide du site accepteur, aussi une guanine, est détaché de l'exon suivant, dont le premier nucléotide est lié par son phosphate 5' au carbone 3'



du dernier nucléotide de l'exon précédent. L'intron, libéré sous forme de **lasso**, est détruit par des nucléases. Le transcrit perd successivement tous ses introns et les exons épissés constituent la séquence codante du message. L'épissage peut être **alternatif** (se fait de façon différente d'une cellule à l'autre), c'est à dire peut conduire à plusieurs structures de l'ARNm: les ARNm alternatifs ont chacun en propre certains exons ainsi que des exons en commun (Figure 11).

Figure 11. Epissage alternatif d'un même ARNm donnera des protéines de fonctions différentes.

b. Modifications du transcrit

Après l'épissage, l'ARNm subit d'autres modifications dont la coiffe de cet ARN et la formation d'une queue polyAdényl ou poly(A) (Figure 12).

Les RNA messagers sont détruits par des ribonucléases dans le cytoplasme. Leur hydrolyse libère des nucléotides qui seront réemployés à la synthèse de nouveaux acides nucléiques. Pour protéger les RNA messagers de ce catabolisme, une structure particulière dissimule l'extrémité 5' terminale de ces ARN aux exonucléases spécifiques de cette partie de l'ARN. Cette structure est appelée coiffe du message (CAP). Lorsque l'élongation du transcrit primaire commence, l'extrémité COOH terminale de la polymérase est phosphorylée. Cette phosphorylation permet de libérer les protéines du complexe d'initiation et de commencer aussitôt les modifications sur la partie du RNA qui est déjà transcrite. Pour ajouter la coiffe, un complexe multienzymatique exerce trois activités:

- l'hydrolyse du phosphate γ du nucléotide 5' terminal du transcrit primaire;
- le transfert sur le phosphate β restant d'un guanylate à partir d'un GTP (GMP);
- la méthylation de ce GMP sur l'azote n°7 (Si le nucléotide initial comprend une adénine celle-ci peut être méthylée sur l'azote n°6. Les riboses des nucléotides initiaux sont parfois méthylés sur l'oxygène de la fonction alcool en 2').

Les deux premières activités sont catalysées par une protéine de 68 kDa et la méthylation par une autre sous-unité de 57 kDa.

Des signaux de fin de transcription se retrouvent à la fin de la plupart des gènes des Mammifères : TATGTTTC. A la lecture de ces signaux, l'ARN-polymérase II s'arrête de transcrire et quitte l'ADN en libérant le transcrit primaire. Aussitôt une endonucléase coupe la fin du transcrit environ une quinzaine de nucléotides après un autre signal : AAUAAA dit boîte de polyadénylation (ou boîte poly(A)). Sur l'extrémité 3'OH de cette coupure, une poly(A) polymérase va condenser un grand nombre de nucléotides, tous à adénine. Le transcrit primaire se trouve allongé d'une '**queue poly(A)**' de plus de mille nucléotides. Cette queue poly(A) est indispensable à la maturation et à l'activité de l'ARN messageur qui la porte. Elle sera lentement digérée par les exonucléases du cytoplasme lorsque le messageur sera actif. Lorsqu'elle sera réduite à quelques centaines de nucléotides le messageur vieilli sera détruit totalement pour être remplacé par un messageur neuf.

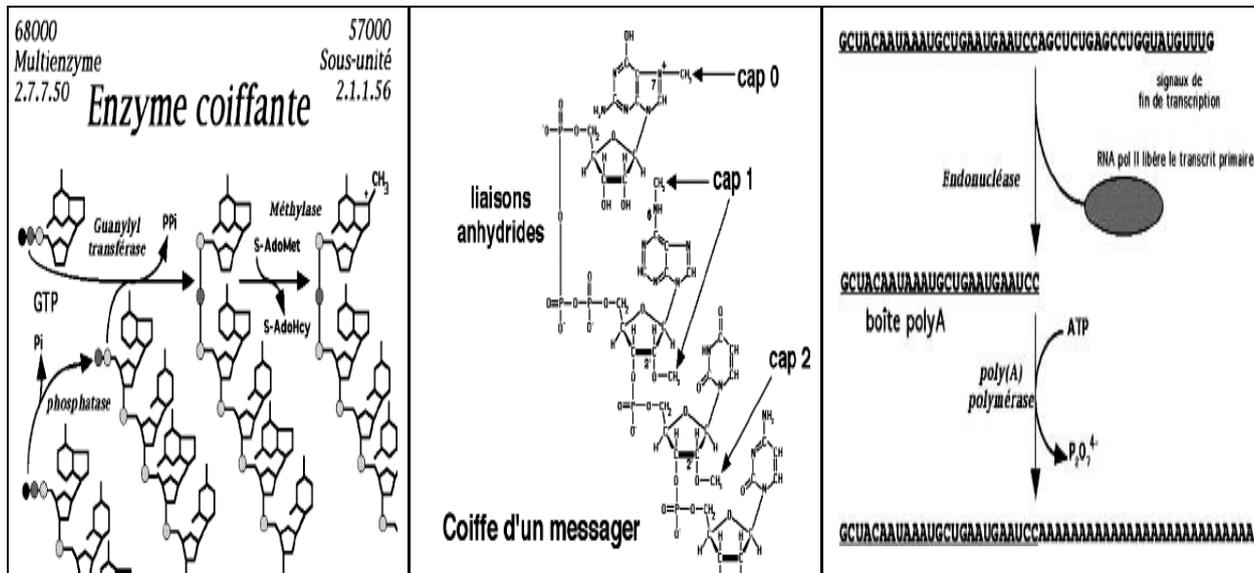


Figure 12. Modifications de l'ARNm: coiffe ou cap et queue poly(A)

c. Stabilité du messager

L'ARN messager mature est une copie de travail du gène destiné à diriger la traduction catalysée par les ribosomes. Son extrémité 5'-phosphate est protégée par la coiffe sans laquelle l'ARN est dégradé dès la transcription par des 5' exonucléases. Son extrémité 3'-OH est constituée d'une longue queue poly(A) qui est lentement digérée par des 3' exonucléases. Lorsque cette hydrolyse est avancée, le messager est alors inapte à la traduction et sera détruit par les endonucléases (Figure 13). Les messagers qui portent peu de ribosomes (parce que l'initiation de la traduction est ralentie ou inhibée) sont détruits directement par les endonucléases.

Certains messagers ont une durée de vie courte : beaucoup de ceux qui interviennent dans les mécanismes post-prandiaux (après les repas) ont une durée de vie de l'ordre de 2 heures et doivent être entièrement resynthétisés à chaque repas. D'autres messagers, dans le système nerveux central par exemple, sont complètement stables durant des années.

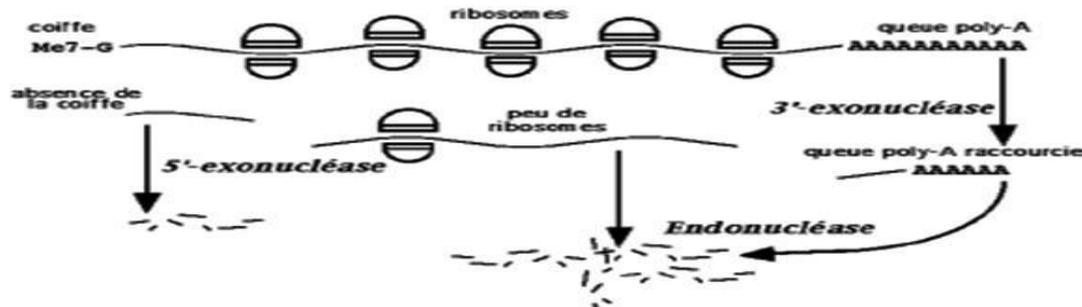


Figure 13. Stabilité du messager

d. Maturation des RNA ribosomiques

Pour transcrire les ARN des ribosomes, la RNA polymérase I synthétise un transcrite primaire de 13000 nucléotides (nt), et l'ARN polymérase III synthétise le petit ARN ribosomal (ARNr) 5 S de 120 nt. La maturation du transcrite primaire précurseur des ARNr se fait par excision de trois fragments : l'ARNr 28 S de 4518 nt, le 18 S de 1874 nt et le 5,8 S de 160 nt (Figure 14). Les ARNr 28 S, 5,8 S et 5 S vont s'associer avec 49 protéines pour former la grande sous particule. L'ARNr 18 S formera la petite sous-particule avec 33 protéines.

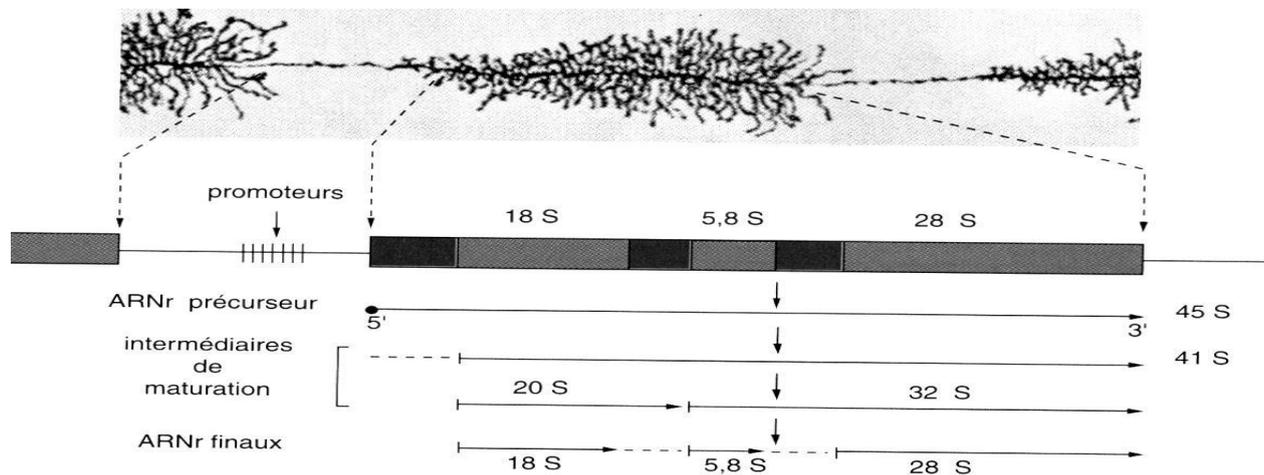


Figure 14. Maturation des ARN ribosomiques

IV.3. La traduction

La traduction est la lecture d'un ARNm par des ribosomes qui synthétisent des protéines dont la structure primaire est déterminée par celle de cet ARN messager. L'ARN messager mature comprend plusieurs éléments de structure influençant la traduction (Figure 15):

- * Les structures situées aux deux extrémités de l'ARNm : la structure en 5', constituée par la 7-méthyl-guanosine triphosphate (ou **CAP** ou chapeau) et la structure en 3', constituée d'une séquence additionnelle polyadénylée (queue poly(A)). La communication entre ces deux structures est essentielle pour l'efficacité du démarrage de la traduction.

- * Les séquences internes du messager assurant pour le démarrage de la traduction une entrée des ribosomes (ou IRES) indépendante de l'extrémité 5' (CAP) du messager.

- * Les petits cadres de lecture (uORFs), situés en amont du cadre principal (ou **ORF**, 'open reading frame') représenté par la séquence codante du gène, qui amoindrissent la traduction par avortement répété en raison de leur taille.

- * Les structures secondaires et tertiaires de l'ARNm, en forme d'épingles à cheveux ou de nœuds, qui bloquent normalement la traduction, mais peuvent faire partie des IRES et ainsi stimuler une traduction indépendante de l'extrémité 5'.

- * Les sites de liaisons spécifiques de complexes protéiques qui sont des déterminants cruciaux de la traduction (Séquence de Kozack de démarrage avec codon AUG d'initiation et le codon de terminaison).

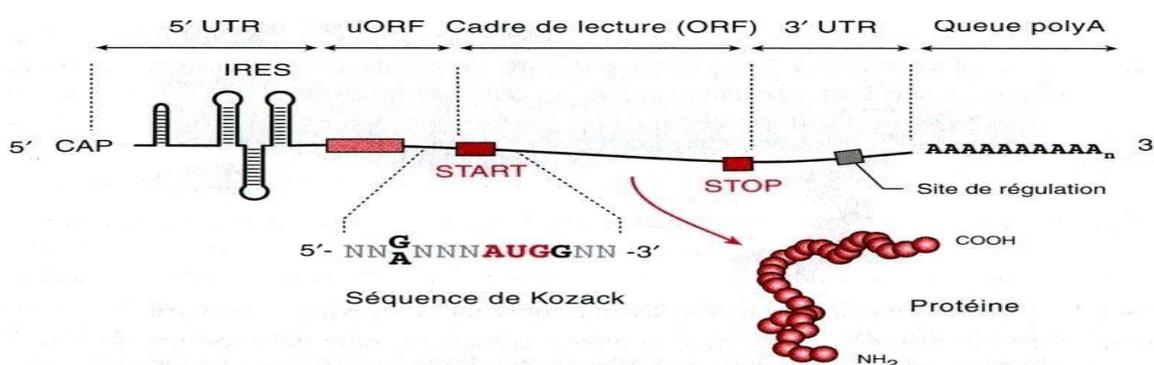


Figure 15. Structure des ARNm influençant la traduction. **START**: région contenant le codon de démarrage et **UTR**: région non traduite

IV.3.1. Langage et outils de la traduction

a. Codon et code génétique

Le code génétique est donc fondé sur des mots de trois lettres, les **codons**. Le codon est un ensemble de trois nucléotides de la séquence d'un acide nucléique portant l'information génétique permettant l'incorporation d'un acide aminé dans la séquence primaire d'une protéine. Il y a plus de codons dans le code

génétique que d'acides aminés et de ponctuations dans les protéines, il y aura donc plusieurs codons traduits par le même acide aminé (homonymes). Ces homonymies représentent une perte d'information entre le langage nucléique (64 signifiants pour les mots de 3 lettres nucléiques) et le langage protéique (au moins 20 signifiants pour les 20 acides aminés plus les ponctuations de début et fin du message) : le **code génétique est dégénéré** (Tableau 2).

Code dégénéré	Ile = AUA, AUC, AUU (AUH) STOP = UAA, UAG, UGA (UGH)	1er	2ème nucléotide				3ème
		nucléotide	U	C	A	G	nucléotide
Arg = CGA, CGC, CGG, CGU (CGN) AGA, AGG (AGR)	Asn = AAC, AAU (AAY) Asp = GAC, GAU (GAY) Cys = UGC, UGU (UGY)	U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
			Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
			Leu (L)	Ser (S)	Stop	Stop	A
			Leu (L)	Ser (S)	Stop	Trp (W)	G
Leu = CUA, CUC, CUG, CUU, (CUN) UUA, UUG (UUR)	His = CAC, CAU (CAY) Phe = UUC, UUU (UUY) Tyr = UAC, UAU (UAY)	C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
			Leu (L)	Pro (P)	His (H)	Arg (R)	C
			Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
			Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
Ser = UCA, UCC, UCG, UCU (UCN) AGC, AGU (AGY)	Gln = CAA, CAG (CAR) Glu = GAA, GAG (GAR) Lys = AAA, AAG (AAR)	A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
			Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
			Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
			Met (M)	Thr (T)	Lys (K)	Arg (R)	G
Ala = GCA, GCC, GCG, GCU (GCN) Gly = GGA, GGC, GGG, GGU (GGN) Pro = CCA, CCC, CCG, CCU (CCN) Thr = ACA, ACC, ACG, ACU (ACN) Val = GUA, GUC, GUG, GUU (GUN)	Met = AUG Trp = UGG	G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
			Val (V)	Ala (A)	Asp (D)	Gly (G)	C
			Val (V)	Ala (A)	Glu (E)	Gly (G)	A
			Val (V)	Ala (A)	Glu (E)	Gly (G)	G

Tableau 2. Le code génétique et sa dégénérescence

La séquence codante est une suite de codons dont chacun permet d'incorporer spécifiquement un acide aminé dans la synthèse d'une protéine. Le code génétique est le même pour tous les êtres vivants de la biosphère (universel). Il existe quelques variations (codons propres à la biosynthèse des protéines dans les mitochondries). Le code génétique comporte 61 codons pour signifier les 20 acides aminés qui participent à la synthèse des protéines : chaque acide aminé peut être codé par plusieurs codons (de un à six) qui diffèrent en général par leur troisième nucléotide. On dit que le code est dégénéré.

Le code génétique est le fruit d'une longue évolution et sa disposition n'est pas due au hasard. Les correspondances entre les codons et les acides aminés ont été sélectionnées pour que les changements de bases aient le moins d'effet possible sur la protéine exprimée. Ainsi, tous les codons dont la deuxième lettre est un U correspondent à des acides aminés hydrophobes, donc ayant des propriétés physiques proches. De même, les acides aminés acides correspondent aux codons commençant par GA, de telle sorte que le changement de la troisième base ne fera pas disparaître la charge anionique du radical. Le codon le plus proche des codons Stop est celui du tryptophane. Un changement de chacun ou des deux G engendre un codon Stop et donc un arrêt de la traduction. Mais ce codon correspond à l'acide aminé le plus rare des protéines habituellement traduites.

En écrivant le code génétique dans l'autre sens, de l'acide aminé vers les codons, on montre que beaucoup d'acides aminés ont des codons homonymes. Certains ont six codons différents, d'autres quatre, trois ou deux. Ces codons homonymes ne sont pas employés au hasard parce que les ARNt correspondants n'existent pas dans toutes les cellules aux mêmes concentrations. De sorte que certains codons auront moins de chances de s'exprimer dans les tissus où l'ARNt correspondant est rare. Il existe de nombreux ARNt dont l'anticodon ne se lie pas de façon spécifique avec la troisième base du codon. Cette base est donc reconnue moins spécifiquement ou pas du tout, ce qui explique la dégénérescence habituelle de cette troisième lettre.

b. Ribosome eucaryote et polyribosome

Les ribosomes du cytoplasme des cellules eucaryotes sont des complexes multienzymatiques qui associent 82 chaînes d'acides aminés et 4 acides ribonucléiques. Ces molécules sont associées entre elles pour former deux particules distinctes : la sous-unité 60S (*Large* = 2800000 daltons) et la sous-unité 40S (*Small* = 1400000 daltons) qui peuvent se dissocier facilement. Les acides ribonucléiques ribosomiaux et les protéines ont des sites de fixation pour la séquence du message, pour les RNA de transfert qui portent l'acide aminé à incorporer (site A) et le peptide en cours de synthèse (site P), un site catalytique pour former les liaisons

peptidiques, des sites de fixation pour les cofacteurs protéiques de l'initiation (eIF2, eIF3), de l'élongation et de la terminaison et des sites de régulation (protéine S6).

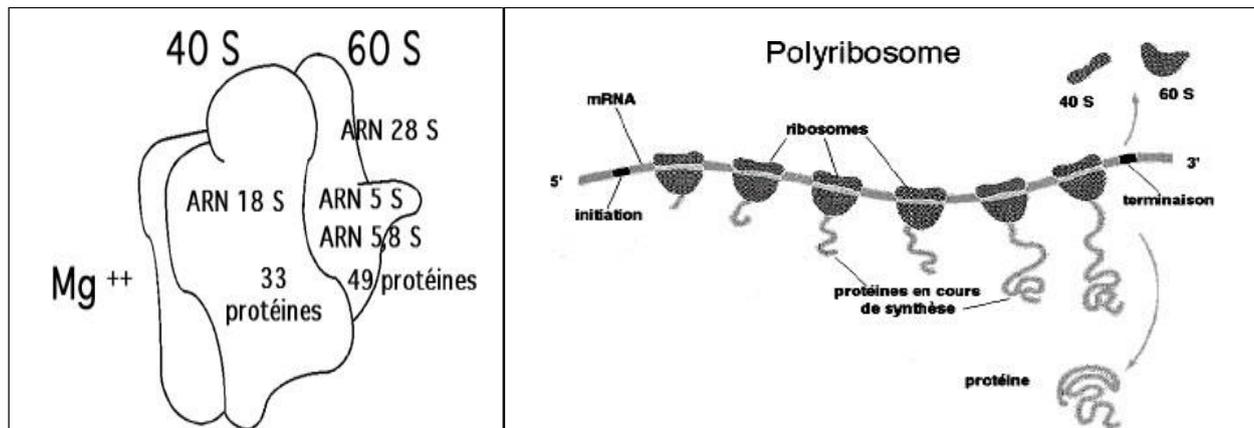


Figure 16. Structure du ribosome eucaryote et le polyribosome

Sur le même messenger plusieurs ribosomes effectuent la traduction de la même protéine les uns après les autres : le tout constitue un polyribosome (Figure 16). L'initiation est un phénomène permanent à l'extrémité 5' d'un RNA messenger et les ribosomes se succèdent sur le messenger à raison d'un tous les 100 nucléotides environ. Le premier acide aminé incorporé constitue l'extrémité NH₂ terminale de la protéine. A l'extrémité 3', en arrivant au codon de terminaison, les sous-particules du ribosome se séparent et libèrent la protéine synthétisée. Le dernier acide aminé incorporé constitue l'extrémité COOH terminale de la protéine. Les polyribosomes présentent un aspect différent selon la régulation des différentes étapes de la traduction. Plus l'initiation est active plus les ribosomes sont nombreux sur le messenger. Si l'élongation est lente, les ribosomes vont mettre plus de temps à lire la séquence codante. Une activation brutale de la terminaison dissocie tous les ribosomes du messenger. De nombreux antibiotiques sont capables d'interférer avec chacune des étapes de la synthèse des protéines (streptomycine, cycloheximide, puromycine).

c. L'ARN de transfert

La liaison de l'acide ribonucléique de transfert (ARNt, Figure 17) avec l'ARNm porteur de l'information, se fait par complémentarité entre 3 nucléotides de chacun de ces deux ARN. Les 3 nucléotides de l'ARNm constituent un codon et les 3 nucléotides de l'ARNt un anticodon. Au cours de la traduction l'anticodon et le codon se lient de manière antiparallèle, et l'acide aminé porté par l'ARNt est incorporé à la protéine en cours de synthèse. La séquence primaire de l'ARNm est donc traduite par groupes de 3 nucléotides (codons). Un nucléotide de l'ARNm au cours de la traduction peut se trouver en position 1, 2 ou 3 dans un codon si le nombre de nucléotides qui le séparent du codon d'initiation AUG est ou n'est pas un multiple de 3. Cette position porte le nom de **cadre de lecture**.

Les ARN de transfert constituent le lien chimique nécessaire entre la structure du codon, reconnu par l'anticodon, et l'acide aminé spécifique porté par l'ARNt. C'est en quelque sorte le **dictionnaire** de la traduction. L'anticodon est une séquence de trois nucléotides situés à l'extrémité de la boucle inférieure de l'ARNt, complémentaire et antiparallèle de la séquence du codon de l'acide aminé correspondant. Chaque acide aminé est lié spécifiquement (code génétique) par une **amino-acyl-tRNA-synthétase** à l'extrémité 3' de l'ARNt dont l'anticodon lui correspond (ARNt chargé). Les ribosomes lient les ARNt chargés sur le site A de l'élongation si leur anticodon s'apparie avec le codon du messenger à cet endroit. L'élongation transfère alors le peptide sur l'acide aminé nouveau, lui-même porté par l'ARN de transfert.

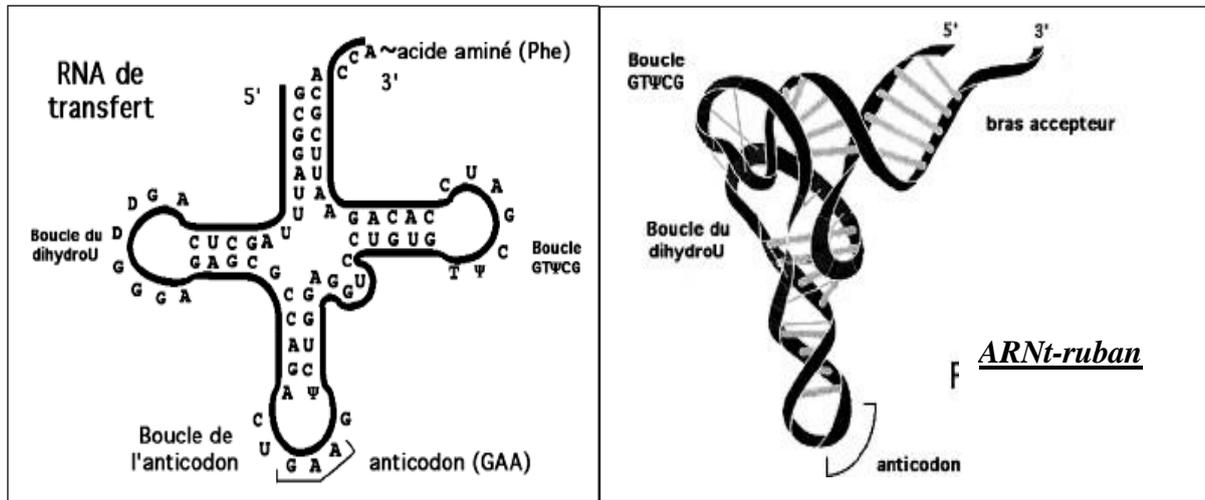


Figure 17. Structures de l'ARN de transfert: secondaire (trèfle) et tertiaire (ruban)

L'analyse cristallographique des ARNt de transfert a montré que les extrémités des deux bras, boucle des dihydrouraciles et boucle GTΨCG, se referment l'une sur l'autre en échangeant des liaisons hydrogène. En sorte que dans ce modèle l'ARNt affecte une forme en L où seuls le bras de l'anticodon et le bras accepteur de l'acide aminé sont aux extrémités de la molécule.

IV.3.2. Mécanisme de la traduction

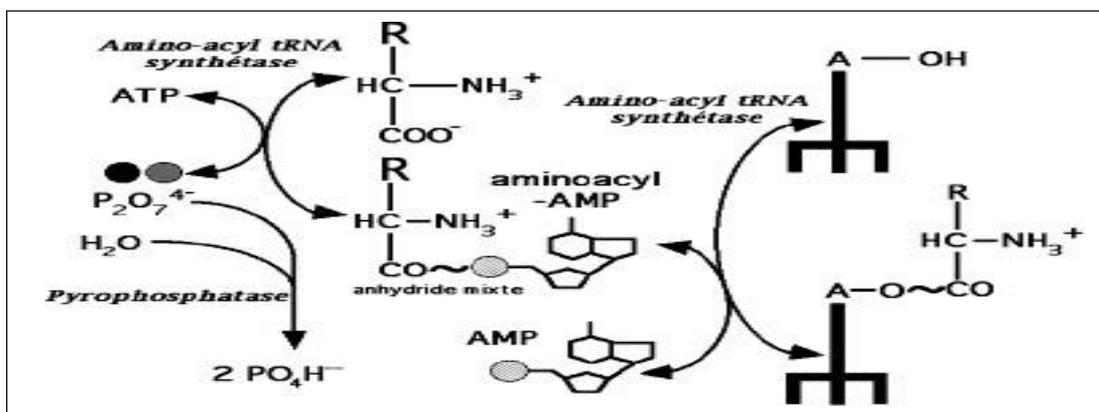
Le processus d'assemblage de la chaîne polypeptidique est une réaction ordonnée consistant en un transfert étape par étape d'acides aminés en suivant la séquence d'ARNm. Ce processus peut être divisé en trois étapes (initiation, élongation et terminaison) encadrées par deux autres étapes qui sont l'activation des acides aminés et les modifications post-traductionnelles.

a. Activation d'un acide aminé

Les acides aminés libres du cytoplasme sont les substrats de la synthèse des protéines. Pour y participer ils doivent être activés. L'activation des acides aminés est catalysée par des enzymes spécifiques : les aminoacyl-tRNA synthétases. Il en existe au moins une pour chacun des 20 acides aminés. Ces enzymes ont une double spécificité : elles reconnaissent spécifiquement un acide aminé et elles reconnaissent spécifiquement l'ARNt non chargé correspondant.

L' aminoacyl-tRNA synthétase hydrolyse un ATP en AMP (liaison riche en énergie) puis active l'acide aminé en liant sa fonction acide avec la fonction acide du phosphate α de l'AMP (liaison anhydride mixte riche en énergie). Le pyrophosphate est aussitôt détruit par une pyrophosphatase. L'acide aminé ainsi activé est transféré ensuite avec sa liaison riche en énergie sur une des fonctions alcool secondaires du ribose de l'AMP 3'-terminal de l'ARNt (Figure 18). L'ARNt chargé se lie ensuite au ribosome pour la synthèse de la protéine.

Figure 18. Activation des acides aminés



b. Initiation de la traduction

L'initiation de la traduction est l'étape limitante de la traduction. Les sous-unités des ribosomes sont dissociées dans le cytoplasme. Une cascade d'événements va former un complexe d'initiation (Figure 19). Au repos, le facteur eIF2 (*eucaryotic initiation factor 2*) est porteur d'un GDP, coenzyme qu'il a hydrolysé au cours du cycle d'une initiation précédente. En présence du facteur eIF2B, un nouveau GTP est substitué à ce GDP. Le facteur eIF2 ainsi activé, peut alors lier l'ARNt chargé d'une méthionine dont l'anticodon est complémentaire du codon d'initiation (AUG) du messenger. En présence du cofacteur eIF4C, la petite sous-unité va fixer le facteur eIF3 et le facteur eIF2 activé qui porte l'ARNt chargé de la méthionine initiale. L'énergie de la formation de ce complexe a été fournie par l'hydrolyse de la liaison riche en énergie du GTP porté par le facteur eIF2. La séquence 5' non traduite du RNA messenger est reconnue par les cofacteurs eIF4A, eIF4B et eIF4F qui s'y fixent. Grâce à l'hydrolyse d'un ATP pour fournir l'énergie, le messenger est alors transféré sur la petite sous-unité, en regard du site P, de façon à hybrider les nucléotides du codon d'initiation avec ceux de l'anticodon de l'ARNt de la méthionine initiale.

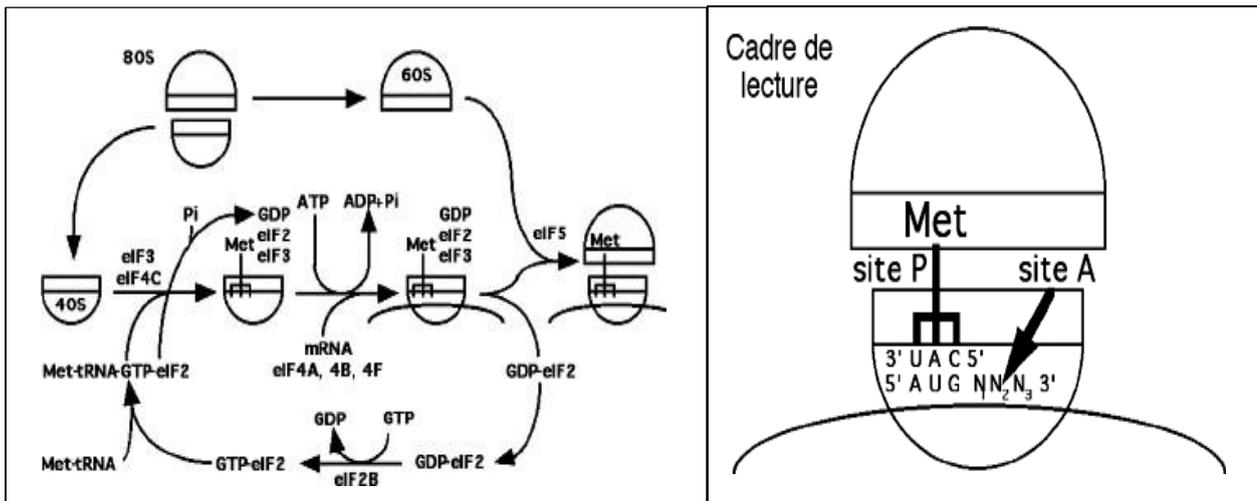


Figure 19. Initiation de la traduction et cadre de lecture

En présence du dernier cofacteur eIF5, le complexe va se lier à une grande sous-unité pour constituer un ribosome fonctionnel. Les cofacteurs d'initiation sont libérés et la traduction commence. Le cofacteur eIF2 toujours porteur de son GDP, se libère pour recommencer un nouveau cycle d'initiation.

Le positionnement du messenger par rapport à l'ARNt de la Méthionine initiale détermine le cadre de lecture de la traduction du messenger. L'ARNt de la méthionine occupe un des deux sites de fixation des ARNt sur le ribosome : site P (ou **peptidyl**, car il contiendra le peptide en cours de synthèse). Le codon AUG du messenger, en s'hybridant dans le site P avec l'anticodon CAU de l'ARN de transfert de la méthionine, place le messenger de telle sorte que le codon suivant ($N_1N_2N_3$) apparaisse dans l'autre site de fixation : site A (ou **aminoacyl**, car il servira à la fixation des nouveaux acides aminés incorporés). Le nucléotide N_1 sera toujours le premier de chaque codon.

c. Elongation de la traduction

Les ribosomes initiés ont leur site A vacant. Le facteur d'élongation eEF1B catalyse l'échange du GDP par un GTP sur le facteur eEF1A. Celui-ci, activé, va recevoir un ARNt chargé qu'il viendra fixer sur ce site A, en hydrolysant le GTP en GDP.

Dès que le codon du messenger au fond du site A a pu se lier complémentaiement avec l'anticodon de l'ARNt apporté, le facteur eEF1A est libéré avec son GDP. Le ribosome catalyse alors le transfert du peptide situé sur l'ARNt du site P sur la fonction amine de l'acide aminé de l'ARNt du site A. Il utilise pour cela, l'énergie de l'hydrolyse de la liaison ester riche en énergie entre le peptide et l'ARNt du site P. Enfin, grâce au facteur eEF2 et à l'hydrolyse d'un autre GTP, l'ARNt du site P est libéré, le messenger, l'ARNt restant et le peptide en cours de synthèse sont alors déplacés (**translocation**) du site A vers le site P, sans qu'il y ait de séparation entre le codon et l'anticodon (Figure 20). Le site A est à nouveau libre pour recevoir l'ARNt de l'acide aminé suivant.

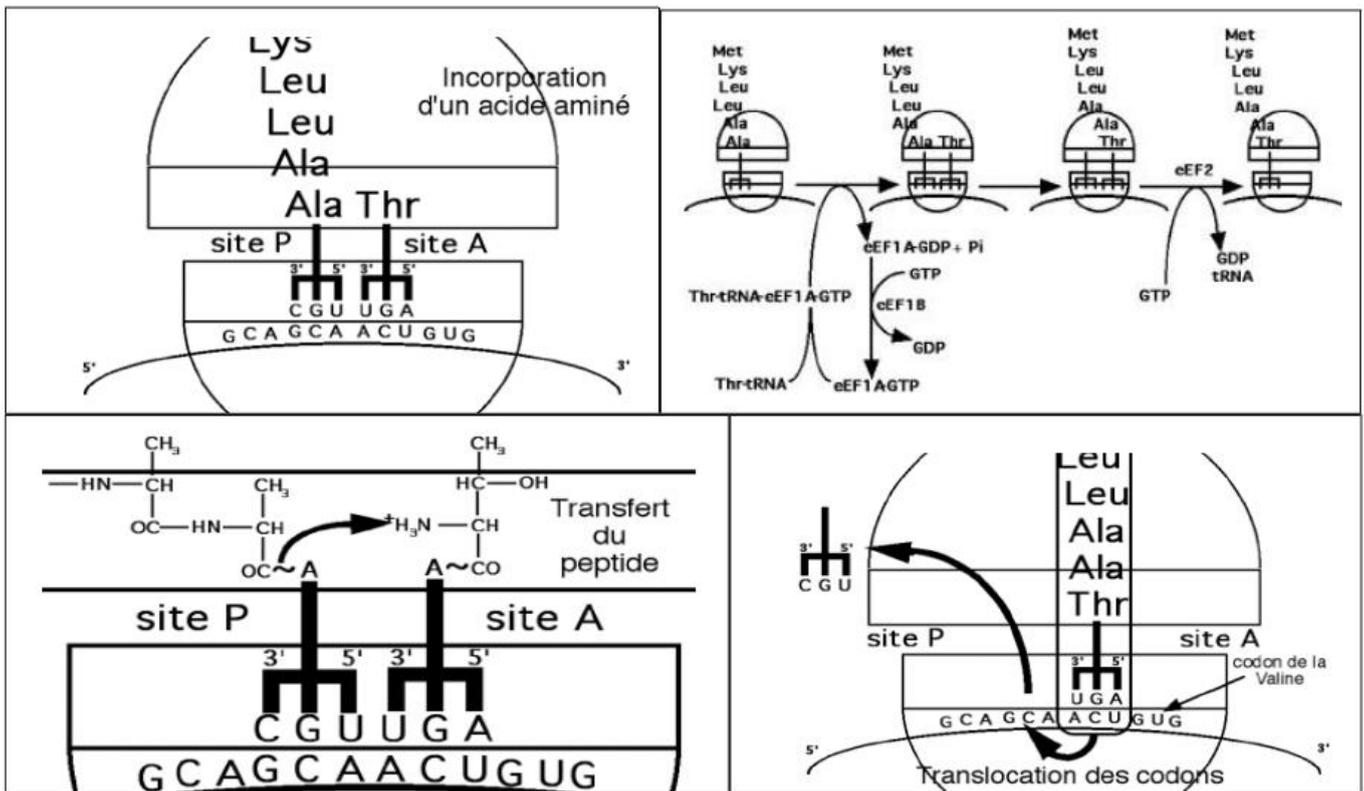


Figure 20. Elongation de la traduction: incorporation d'un acide aminé, formation de la liaison peptidique et translocation des codons

Le bilan énergétique de la traduction dépend du nombre d'acides aminés de la protéine synthétisée. Pour chaque acide aminé, on utilise d'abord un $\text{ATP} \rightarrow \text{AMP}$ pour la synthèse de l'ARNt chargé (aminoacyl-tRNA synthétase). L'AMP produit est réactivé en ADP par un autre ATP (nucléoside-P2 kinase). L'ensemble équivaut à la consommation de deux liaisons riches en énergie $\text{ATP} \rightarrow \text{ADP}$. Pour l'incorporation de cet ARNt chargé dans le site acide aminé du ribosome, le facteur eEF1B utilise un $\text{GTP} \rightarrow \text{GDP}$. Pour la translocation du peptidyl-tRNA du site acide aminé au site peptidique, le facteur eEF2 utilise encore un $\text{GTP} \rightarrow \text{GDP}$. En tout chaque acide aminé incorporé dans la protéine coûte à la cellule quatre liaisons riches en énergie.

*** Signal-peptide et polyribosomes liés**

Lorsqu'une protéine est synthétisée, sa structure secondaire ou tertiaire se construit au fur et à mesure de la traduction et elle est libérée dans le cytoplasme. Lorsque cette protéine est destinée à être incorporée dans les membranes ou les organites de la cellule, la partie codante de l'ARN messager commence par une séquence de quelques acides aminés (15 à 20 résidus) qui sert d'adresse pour l'incorporation de cette protéine dans la membrane. Ces peptides orientent la destinée de la protéine : incorporation dans les membranes (réticulum endoplasmique, appareil de Golgi, membrane plasmique, lysosomes...), entrée dans les mitochondries, excréation hors de la cellule via l'appareil de Golgi, etc... Le signal-peptide est un peptide d'adressage situé à l'extrémité NH₂-terminale des protéines à excréter. Parce que sa fonction est de pénétrer dans la membrane du réticulum endoplasmique, sa structure est riche en acides aminés hydrophobes (Phe, Leu, Ile, Met, Val).

Les polyribosomes qui se forment sur un messager comportant un signal peptide ont une évolution légèrement différente (Figure 21). La traduction s'arrête peu après la synthèse du signal-peptide, dès que celui-ci apparaît hors de la structure du ribosome et se lie avec la SRP (SRP = *Signal Recognition Particle*), qui inhibe l'élongation.

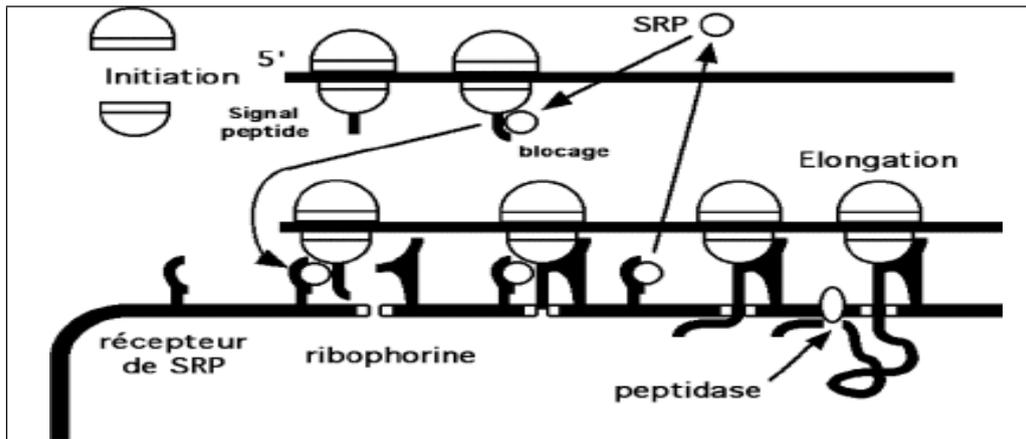


Figure 21. Formation du signal peptide avec les polysomes liés

Pour que l'élongation puisse se poursuivre, il faut que la **SRP** soit reconnue spécifiquement par un récepteur de la membrane du réticulum endoplasmique. Dès que cette liaison est établie, le ribosome est lié par la **ribophorine**, toujours dans la membrane du réticulum, près d'une protéine qui ouvre un pore à travers cette membrane. Le SRP écarté, l'élongation va reprendre. Le peptide en cours de synthèse est alors dirigé à travers cette membrane pour se développer dans la lumière du réticulum endoplasmique. A la face interne de la membrane une endopeptidase spécifique va couper le signal peptide et la synthèse se poursuivra jusqu'à la terminaison. La protéine sera enfin incorporée dans une membrane ou exportée, à travers l'appareil de Golgi, vers l'extérieur de la cellule. L'adressage des protéines vers les mitochondries fait appel à un autre type de peptide d'adressage qui conduit les peptides à traverser la membrane mitochondriale.

* Carboxylation de l'acide glutamique

L'acide carboxyglutamique est un acide aminé propre aux protéines fixant le calcium: protéines des os, protéines de la coagulation du sang. Cet acide aminé est créé par une modification post-traductionnelle de ces protéines (Figure 22). L'enzyme qui effectue la carboxylation a pour coenzyme la vitamine K (ou naphthoquinone), aliment indispensable qu'on trouve dans les feuilles vertes. Cette vitamine K est un transporteur d'hydrogène qui en s'oxydant, va retirer un hydrogène au carbone n°4 d'un acide glutamique, et transférer à cet endroit une molécule de gaz carbonique issue d'un ion bicarbonate. La vitamine K oxydée doit être réduite par une **diaphorase** à NADH pour retrouver sa forme initiale avant de participer à nouveau à cette carboxylation. Les inhibiteurs de la réduction de la vitamine K (antivitamines K) vont empêcher la réaction de carboxylation, faute de vitamine K réduite, et par conséquent empêcher la coagulation du sang.

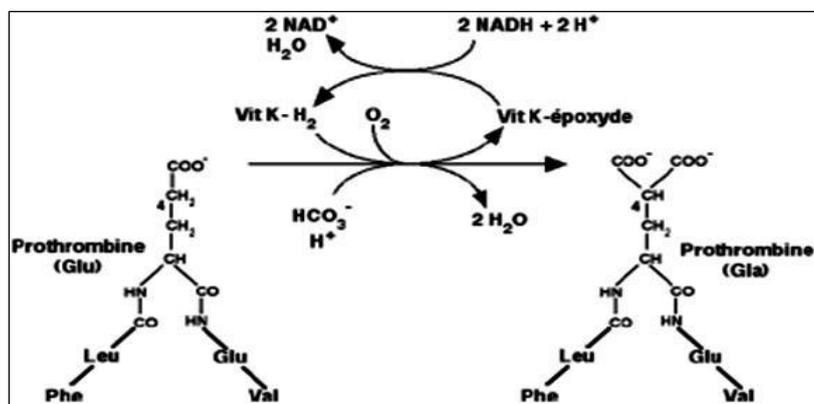


Figure 22. Carboxylation du glutamate

d. Terminaison de la traduction

Lorsque le site A se trouve en regard d'un codon non-sens annonçant la fin de la traduction, le complexe va se dissocier du messager en présence d'un dernier cofacteur eRF (*Releasing factor*). Les deux sous-unités du ribosome se dissocient, la protéine synthétisée est libérée, ainsi que le dernier ARNt (Figure 23).

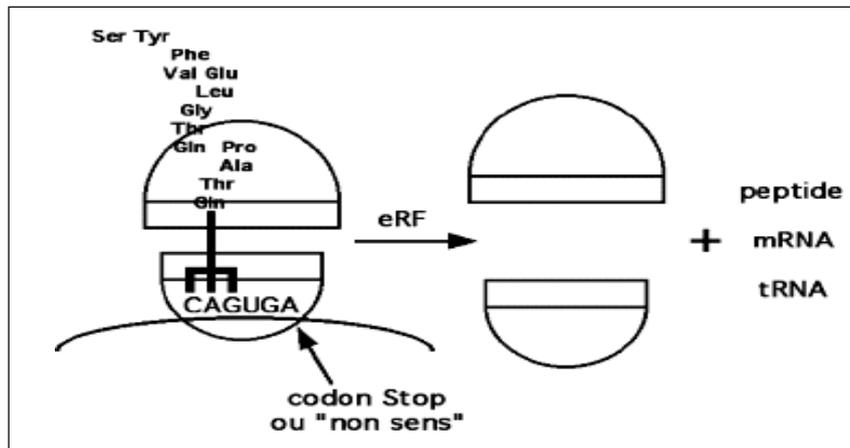


Figure 23. *Terminaison de la traduction*

e. Modifications post-traductionnelles

De nombreuses modifications chimiques se produisent après l'incorporation des acides aminés dans la structure primaire de la protéine (traduction) ; on les appelle modifications posttraductionnelles. On distingue des modifications co-traductionnelles qui se produisent alors que la traduction se poursuit encore et que le peptide naissant est encore attaché au ribosome qui l'a construit, des modifications post-traductionnelles proprement dites qui ont lieu dans la cellule, dans les organites ou hors de la cellule. On appelle protéine **mature** la forme chimique définitive que la protéine montrera au moment où elle remplira sa fonction dans l'organisme. Parmi ces modifications post-traductionnelles:

- Protéolyse : signal-peptide ;
- Glycosylation : Ser-O, Asn-N ;
- Acylation : farnésyl, géranyl, myristyl, palmityl ;
- Hydroxylation (OH-Pro, OH-Lys, OH-Glu), méthylation (CH₃-His), carboxylation (CO₂-Glu);
- Phosphorylation : P-Ser, P-Thr, P-Tyr, P-His ;
- Liaison d'un cofacteur : métal, flavine, hème ;
- Blocage des extrémités (PyroGlu, carboxyamide) ;

Chapitre V. DIVISION CELLULAIRE ET BRASSAGES GENETIQUES AU COURS DE LA REPRODUCTION SEXUEE

V.1. Les divisions cellulaires

La mitose et la méiose constituent les deux modes de divisions chez les eucaryotes, il s'agit des divisions indirectes par différences à la division directe chez les procaryotes (division par bipartition chez les bactéries).

V.1.1. Mitose

La mitose concerne à la fois les cellules haploïdes et diploïdes et aboutit à deux cellules filles génétiquement identiques à la cellule mère. Toute cellule provient de la division d'une autre cellule et tout organisme vivant provient de la division (ou des divisions) d'une cellule. Lorsqu'une cellule se divise, son noyau se divise lui-même en deux selon un processus très précis appelé: Caryocinèse ou mitose.

La mitose doit son nom (du grec: "mitos" = filament c-à-d chromosome) à l'apparition des chromosomes. Les chromosomes de la cellule mère se séparent en deux lots égaux quantitativement et qualitativement (Figures 1 et 2). Les cellules filles reçoivent chacune un lot de chromosomes absolument identique aux chromosomes de la cellule mère. La duplication préalable de l'ADN assure ce processus. Le phénomène est continu mais les cytologistes, pour en faciliter la description, distinguent conventionnellement 4 phases appelées successivement: prophase, métaphase, anaphase et télophase.

Dans les cellules animales et certaines cellules végétales (algues et champignons), la mitose se fait avec l'intervention d'un centrosome (Hyaloplasme + centrioles), ce qui n'est pas le cas chez les plantes supérieures. Ces différentes phases se caractérisent comme suit:

* Prophase:

- Début: disparition des nucléoles et spiralisation des chromatides en filaments courts, épais et colorables.
- Fin: disparition de la membrane nucléaire et les deux chromatides sœurs sont accolés sur toute leur longueur, particulièrement au niveau du centromère (= région d'hypercondensation de la chromatine) pour former les chromosomes scindés en deux (phénomènes invisibles au microscope photoniques). Ensuite il y a formation du fuseau achromatique (complexes protéiques ou kinétochores) entre les deux calottes polaires (zones claires aux extrémités de la cellule).

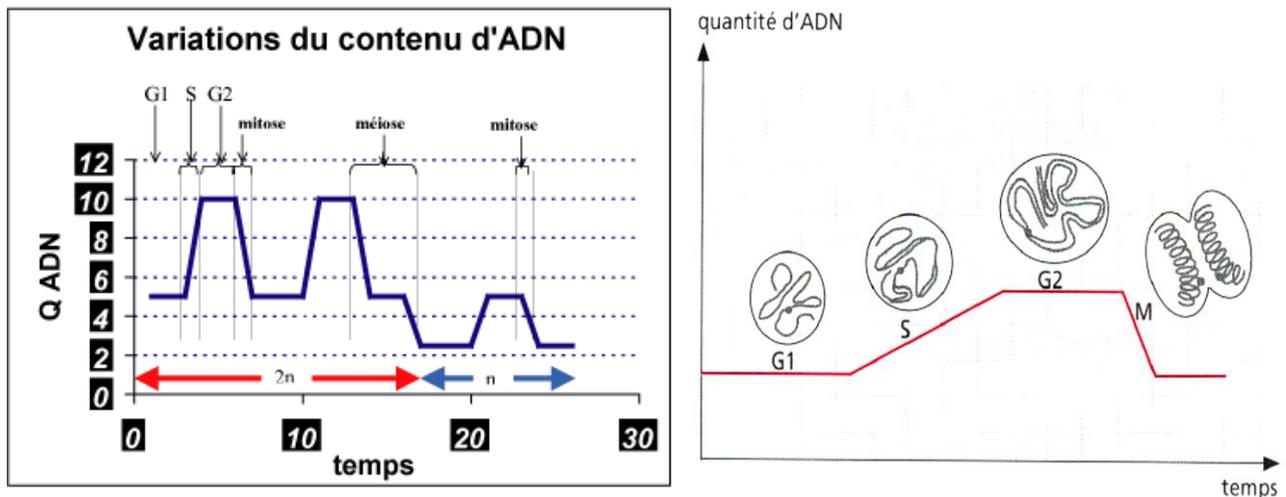


Figure 1. A gauche, quantité d'ADN et ploïdie d'une cellule lors des divisions cellulaires ; à droite, structure des chromosomes au cours du cycle cellulaire.

* Métaphase:

- Individualisation complète des chromosomes en désordre.
- Les chromosomes forment la plaque équatoriale.
- Les chromosomes en forme de V et longitudinalement fissurés s'attachent sur les fibres au niveau du centromère (zone incolore).

* **Anaphase:**

- Contraction des fibres fusoriales.
- Clivage des centromères.
- Séparation et glissement symétrique des 2 lots de chromosomes frères identiques vers les pôles du fuseau.
- Les branches en V restent tournées vers l'équateur.

* **Télophase:** caractérisée par des phénomènes inverses de ceux de la prophase.

- Despiralisation des chromosomes réunis aux pôles.
- Enchevêtrement des chromosomes les uns dans les autres en masses granuleuses peu colorables.
- Reformation de la membrane nucléaire.
- Réapparition des nucléoles et disparition du fuseau achromatique.
- Apparition, au niveau de l'ancienne plaque équatoriale, de la future membrane qui sépare les cellules filles (phragmoplasme).
- Les noyaux retrouvent le même aspect qu'à la prophase mais les cellules filles sont moitié moins grandes que la cellule mère.

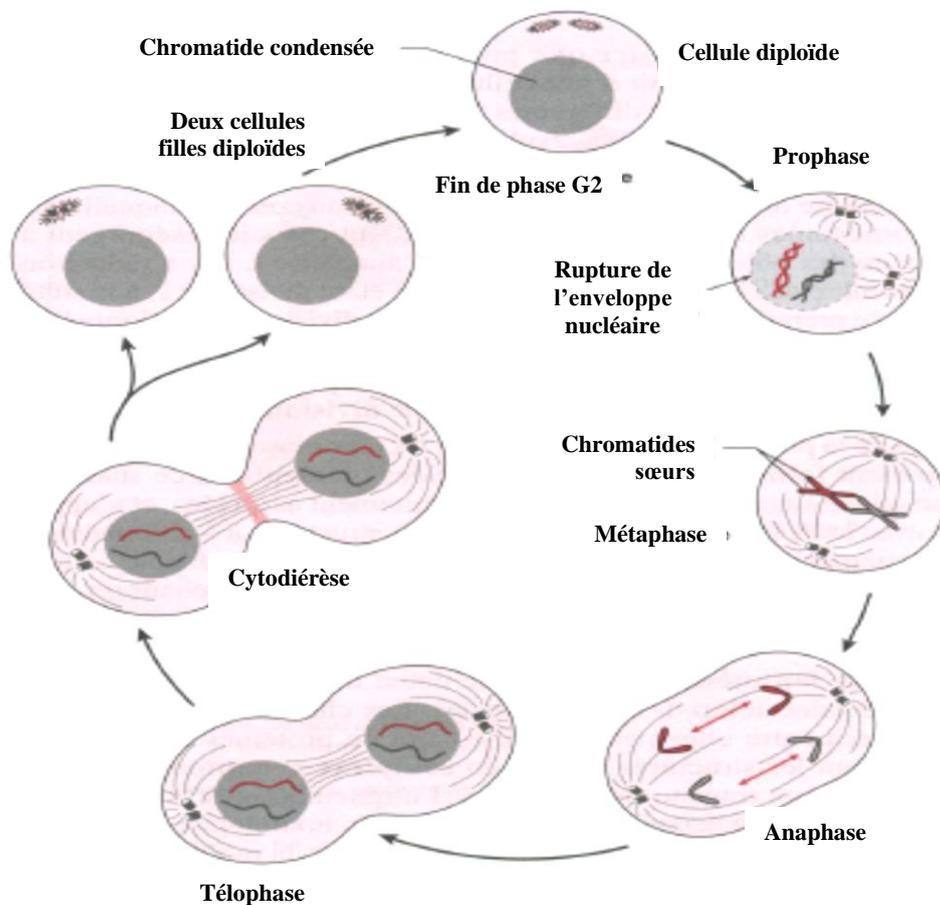


Figure 2. Représentation schématique des différentes phases de la mitose.

		Méiose	Mitose 2 cas possibles	
ADN, ploïdie	au début	1 cellule 2n	1 cellule 2n	1 cellule n
	à la fin	4 cellule n	2 cellule 2n	2 cellule n
	Q d'ADN après la synthèse	4x	4x	2x
	Qd'ADN à la fin	4 fois x	2 fois 2x	2 fois x
Conséquence		2 divisions successives sans synthèse d'ADN entre les deux	une seule division	
Métaphase	Chromosomes homologues	Métaphase I appariement	Métaphase pas d'appariement	
	centromère	pas de clivage : chaque chromosome (avec ses deux chromatides) va à un pôle	clivage de chaque centromère : les 2 chromatides d'un chromosome donnent les deux chromosomes fils.	

Tableau 1. Différences entre méiose et mitose.

V.1.2. Méiose

La méiose est un mode de division qui fait passer une cellule de l'état diploïde obligatoirement (deux exemplaires d'information génétique, 2n) à 4 cellules haploïdes (un seul stock, n) en général appelées les gamètes mais aussi les spores chez les champignons, les algues,...

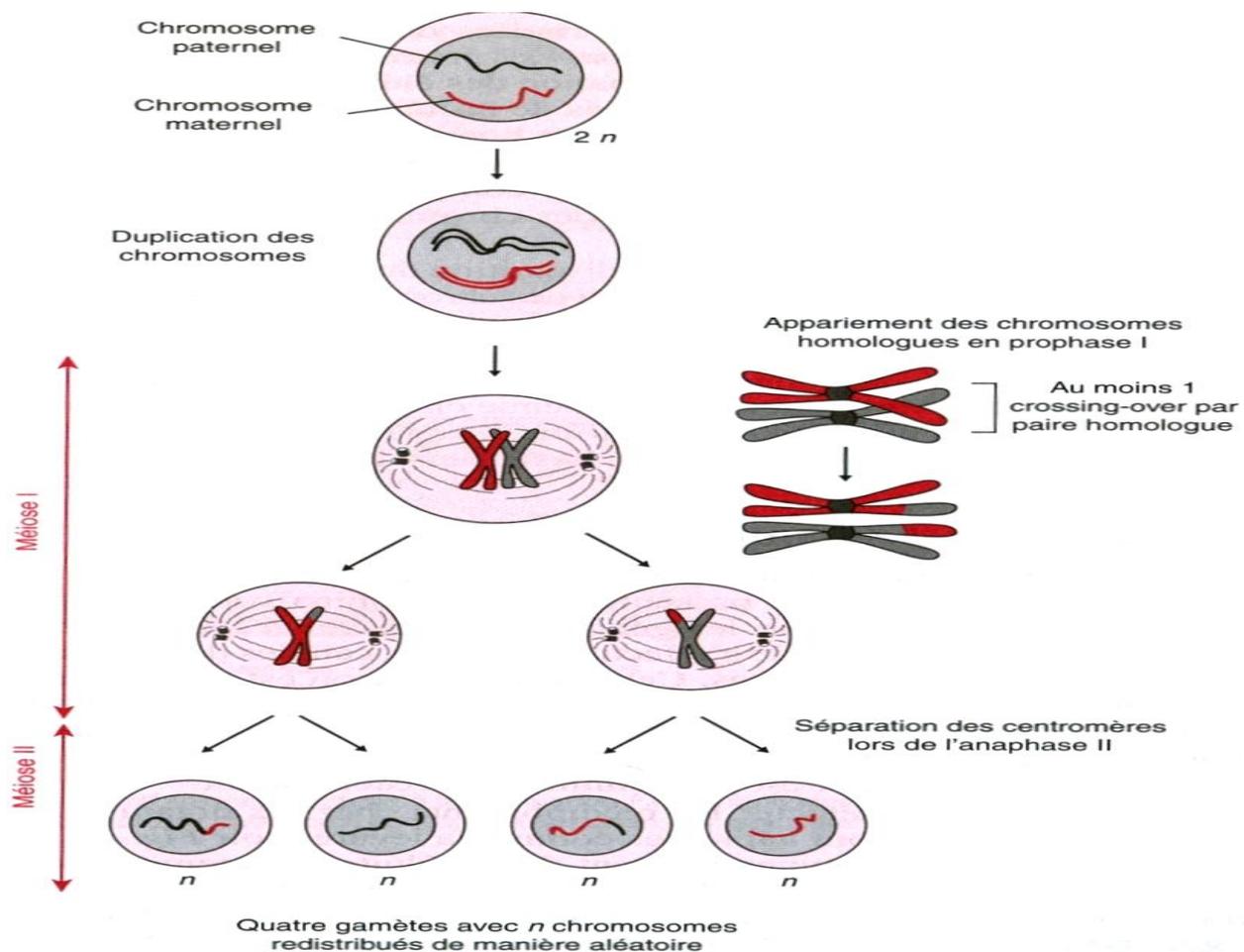


Figure 3. Représentation schématique des différentes étapes de la méiose.

Cette division comporte deux étapes (Figure 3). Dans la première division méiotique ou **méiose I**, les éléments constitutifs de chacun des stocks chromosomiques sont recombinaison d'une façon aléatoire sans clivage de centromère et les chromatides restent associés ségrégent à chacun des pôles de la cellule dans l'anaphase I. Il y a réduction du

nombre de chromosomes par 2 à la suite de cette division ou **division réductionnelle**. Dans la seconde division ou **méiose II**, les chromatides sœurs se séparent et migrent aux pôles de la cellule selon un mode égal à celui de la mitose ou **division équationnelle**. Alors que le nombre des chromosomes est caractéristique de la ploïdie d'un organisme, la quantité d'ADN que contient son noyau ne l'est pas car elle varie suivant le moment où la cellule est observée (Figure 4).

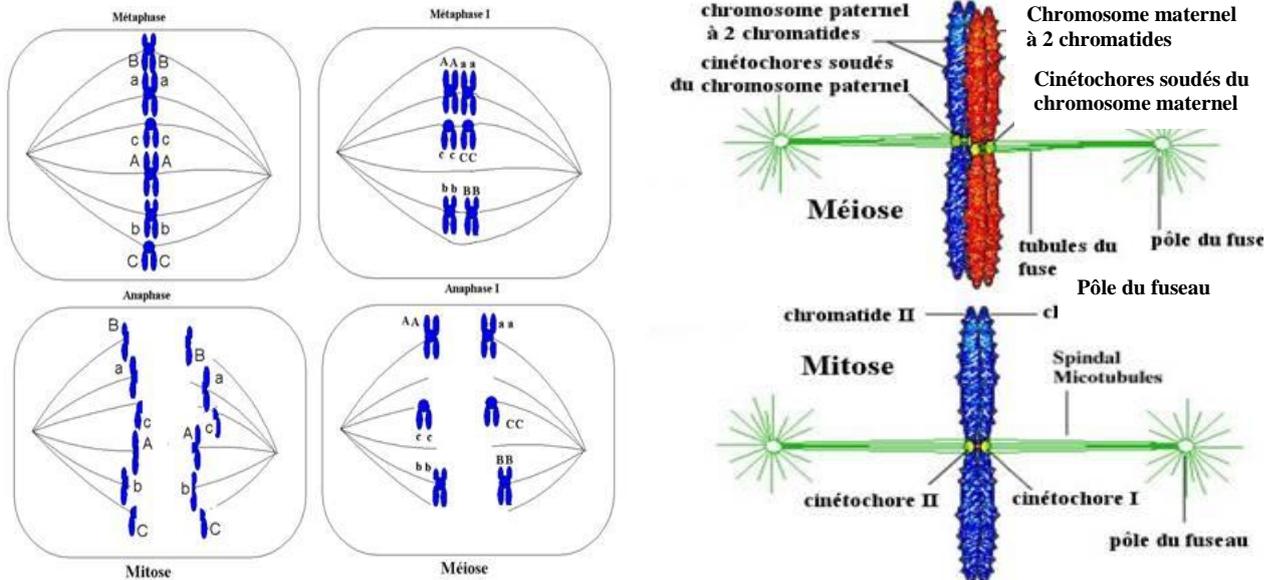


Figure 4. Ce qui diffère à la métaphase et position différente des cinétochores à la méiose et à la mitose.

V.2. Brassages génétiques au cours de la reproduction sexuée

Les brassages génétiques ou recombinaison génétique désignent tout processus permettant d'obtenir un assemblage nouveau d'informations génétiques à partir d'ensembles différents. La recombinaison génétique est responsable de la mobilité et la plasticité de la séquence d'ADN, cause de la biodiversité des êtres vivants. On distingue deux types de recombinaison méiotique (recombinaison produite lors de la méiose).

- recombinaison **intrachromosomique** : recombinaison issue de crossing-over, chiasma ou **recombinaison homologue**.
- recombinaison **interchromosomique** ou ségrégation: répartition aléatoire des chromosomes homologues dans les cellules filles lors de la première division méiotique.

En plus de la recombinaison méiotique, il peut y avoir de la recombinaison in vitro par ligation (voir cours de Biologie moléculaire), union de molécules d'ADN de deux origines différentes, (intégration d'un prophage dans une bactérie) ou in vivo entre le génome d'une cellule et un fragment d'ADN reçu par transformation. Les mécanismes de recombinaison in vivo mettent en jeu des enzymes qui exigent l'appariement des molécules d'ADN qui vont recombiner. Elles doivent donc présenter de l'**homologie**. Un autre type de recombinaison est également illustré par les réarrangements et les déplacements de séquences connues sous le nom de **transposition**, qui se réalise par une recombinaison non homologue (Figure 5).

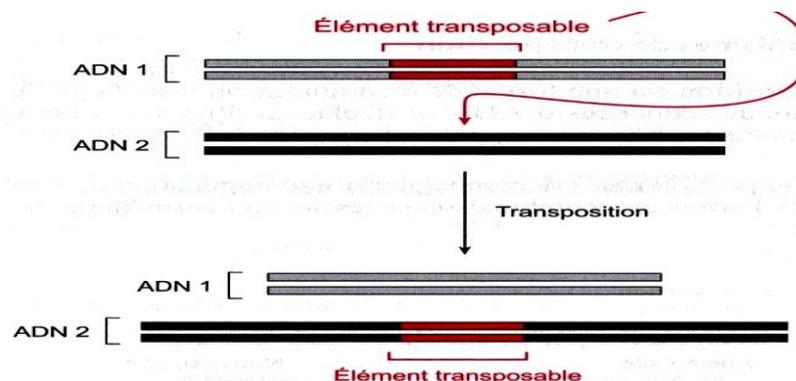


Figure 5. Transposition entre deux sites d'ADN présents ou non sur le même chromosome.

V.2.1. Recombinaison homologue

La recombinaison homologue existe chez tous les organismes et consiste en la rupture suivie de la réunion de deux molécules d'ADN homologues, c'est-à-dire identiques sur des régions d'au moins une centaine de bases, ce qui n'exclut pas la possibilité de différences comme il en existe entre deux allèles d'un même gène. Les deux molécules recombinantes sont porteuses de cassures symétriques sur l'un ou les deux brins et un appariement croisé s'établit entre les brins des deux molécules qui se connectent ainsi en une structure appelée **jonction de Holliday** (Figure 6).

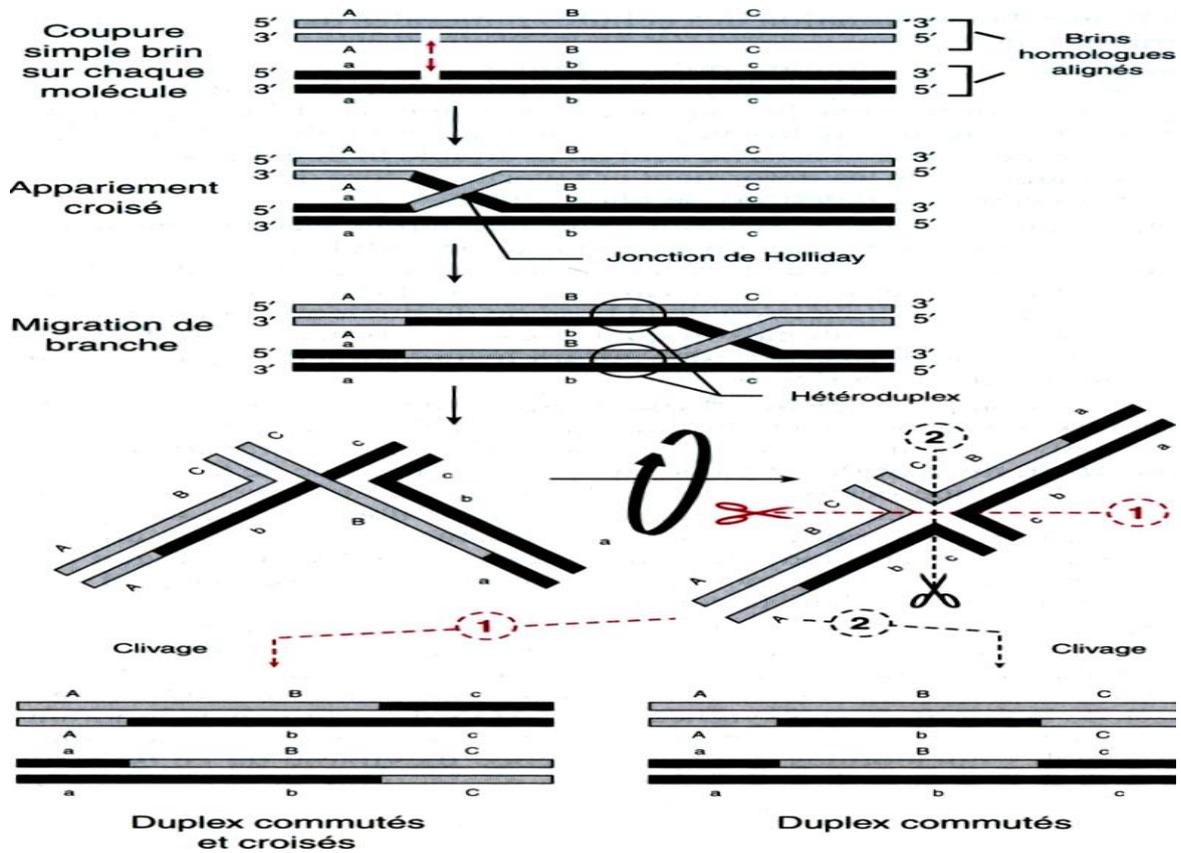


Figure 6. Mécanisme de la recombinaison homologue selon le modèle de Holliday. Les lettres ABC et abc désignent des régions homologues des ADN où les séquences sont cependant légèrement différentes. Les duplex hybrides résultant de la résolution de la jonction diffèrent selon le type de clivage.

Cette structure de jonction migre le long des deux molécules d'ADN par séparation progressive des simples brins ou **migration de branche**. Il assure une hybridation croisée des brins des deux origines. Finalement, des clivages interviennent. Ils libèrent deux duplex d'ADN ayant échangé des régions de taille variable selon les circonstances. Selon le type de clivage, le réassortiment des régions entre les deux ADN concernés par la recombinaison fournit soit des duplex **commutés et croisés**, soit des duplex **commutés** seulement.

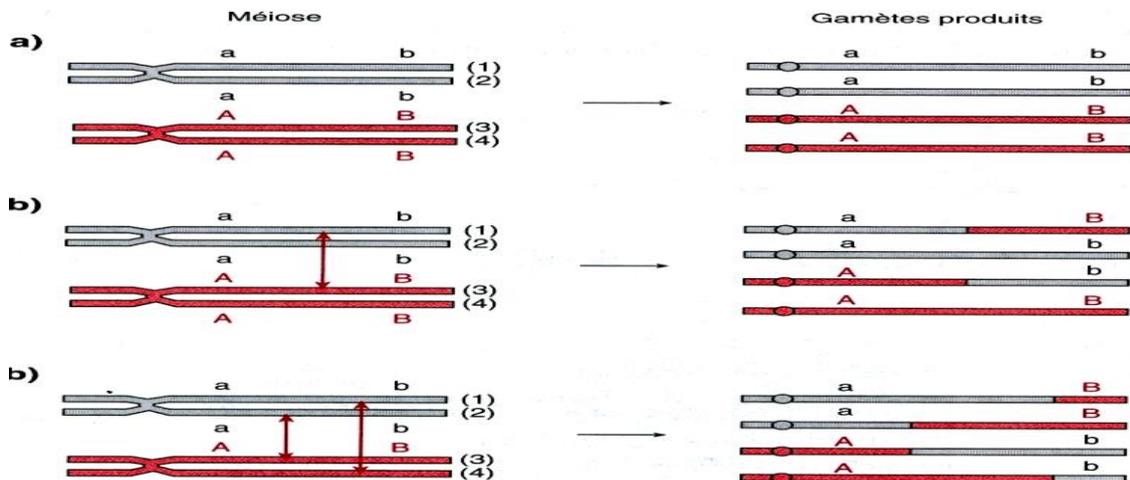


Figure 7. Exemples de possibilités de recombinaison homologue lors de la production des gamètes.

Comme chaque molécule d'ADN composant un chromosome, contient des séquences informatives ou gènes disposés linéairement (Figure 7), la fréquence de recombinaison intrachromosomique permettrait de nous informer sur la liaison génétique entre les caractères étudiés (gènes et allèles) et de savoir la carte génétique d'un génome.

V.2.2. Recombinaison interchromosomique ou ségrégation

Au moment de la première division de méiose, chaque chromosome homologue migrera vers un pôle du fuseau, vers le nord ou le sud de façon aléatoire. Donc pour une cellule d'une espèce où $n=3$, il y a deux façons de se disposer pour chaque chromosome homologue soit en tout 2^3 résultats possibles (ou gamètes) en dehors de tout événement de recombinaison (Figure 8).

Dans la figure 8 seuls 4 cas ont été représentés, les haploïdes de la première ligne peuvent à chaque fois se retrouver au pôle sud ou au pôle nord. Les 8 types de ségrégations se réduisent à 4 si l'on ne tient pas compte de l'orientation nord sud.

V.2.3. Les mutations chromosomiques

Les mutations chromosomiques se présentent sous la forme de duplications (ou amplifications) et de délétions de régions entières qui conduisent souvent à la fusion de gènes d'origine séparés. Il peut s'agir de translocations ou d'inversions de fragments entre deux chromosomes ce qui provoque des fusions géniques dont les produits sont des protéines de **fusion**, parfois responsables de processus de **tumourisation** et de **cancers** (Figure 9). C'est le cas par exemple de la délétion chromosomique conduisant à la fusion des gènes FIG et ROS et à la production d'une protéine de fusion possédant une activité permanente de tyrosine kinase, ce qui a pour effet la transformation tumorale des cellules gliales et de provoquer un cancer du cerveau.

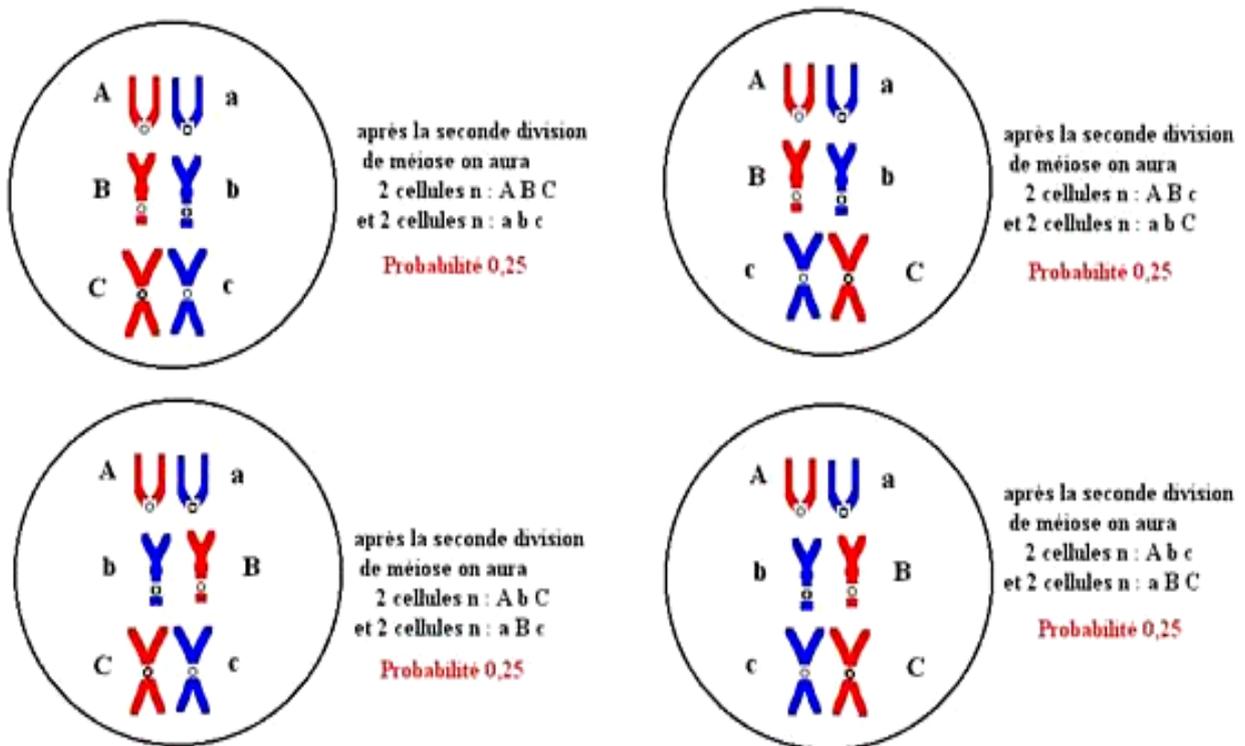


Figure 8. Ségrégation indépendante de 3 chromosomes homologues.

Dans un organisme qui présente deux allèles différents pour un même gène (état hétérozygote : porteurs de deux allèles différents, un sur chaque chromosome homologue), la délétion ou un événement de recombinaison spécifique peuvent conduire à la perte de l'un des deux allèles et à la perte de l'hétérozygotie.

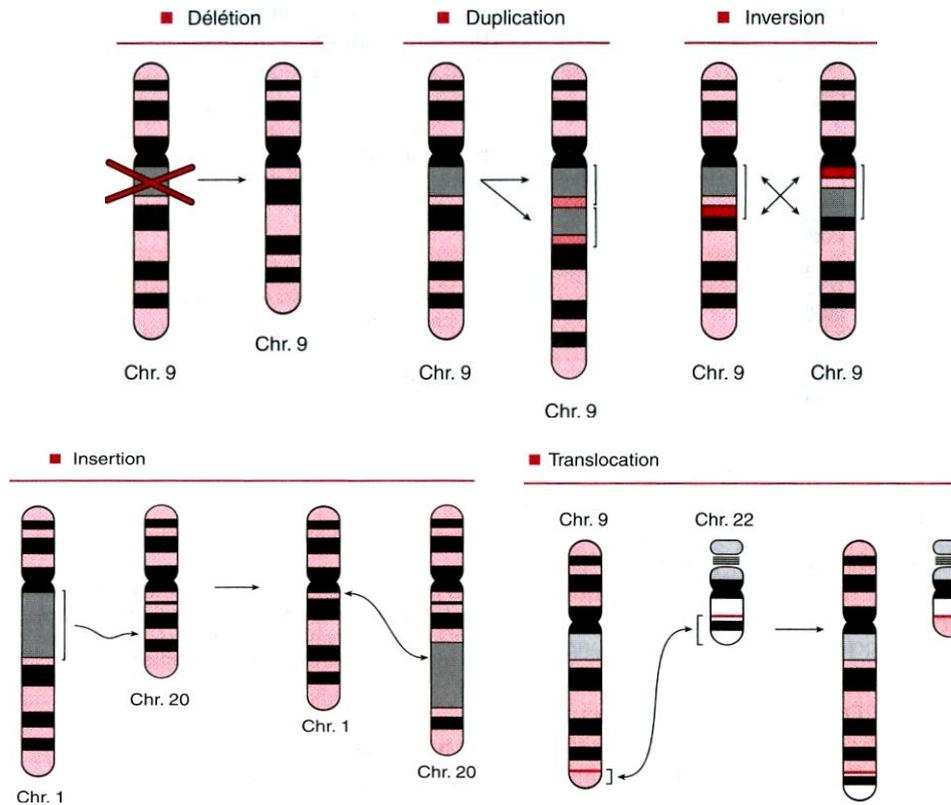


Figure 9. Les cinq principaux types de mutations chromosomiques.

V.3. Ultrastructure du chromosome

Au cours de la division cellulaire la chromatine se transforme en bâtonnet qui ont une grande affinité pour les colorants basiques, ce sont les chromosomes. Au microscope photonique, le chromosome en métaphase apparaît sous forme de 2 bâtonnets de même longueur (= chromatides) réunis en un point: le **centromère** (constriction primaire) qui peut être en position centrale ou excentrée. Le chromosome est l'instrument de transmission des caractères parentaux, de l'information génétique. Il provient de la condensation de l'ADN (ADN + histones + protéines non histoniques) par spiralisation au début de la division cellulaire. Les chromosomes ont une forme définie en fonction de l'apparence et de la position du centromère (Figure 10):

- * En bâtonnet, en V et punctiforme.
- * Métacentrique, submétacentrique ou télacentrique.

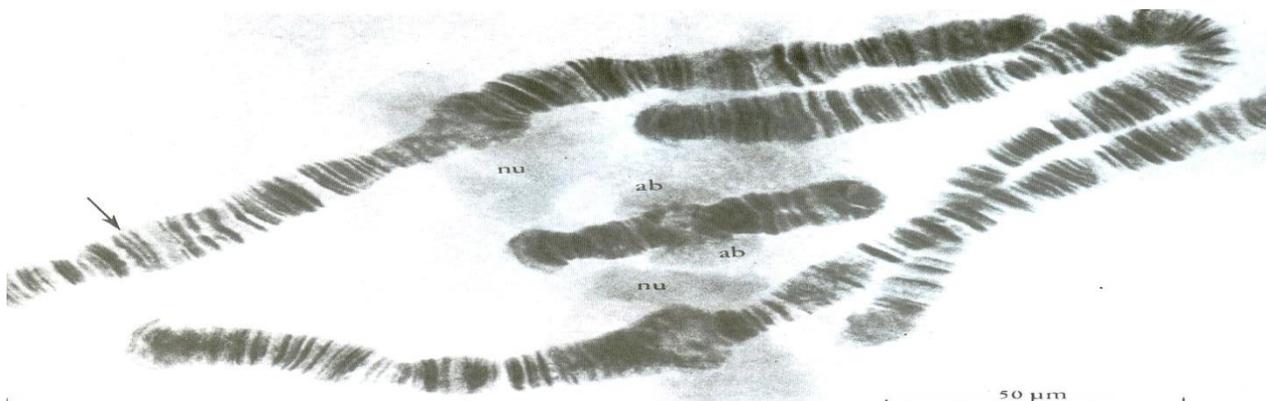


Figure 10. Chromosomes polyténiques.

Les chromosomes sont formés en générale de 2 chromatides. En cas particulier, ils sont dits:

- * En **écouvillons** ou chromosomes plumeux quand les chromatides sœurs sont juxtaposées (cellules germinales, ovocytes).

* **Polyténiques** quand deux chromosomes homologues accolés ayant plusieurs centaines de nucléofilaments (larve de mouche). Ces chromosomes sont visibles en interphase (Figure 10).

Chez les cellules diploïdes tous les chromosomes sont représentés en doubles exemplaires (sauf pour les chromosomes sexuels) et leur nombre varie en fonction des espèces (Voir tableau .2).

Organisme	Position systématique	Nombre de chromosomes par jeu haploïde	Taille du génome (en Mb)
<i>E. coli</i>	Bactérie	Circulaire unique	4,7
<i>S. cerevisiae</i>	Ascomycète	16	14
<i>A. thaliana</i>	Crucifère	5	100
<i>C. elegans</i>	Nématode	6	100
<i>D. melanogaster</i>	Insecte	4	165
<i>M. musculus</i>	Mammifère	20	3 000
<i>H. sapiens</i>	Mammifère	23	3 000

Tableau 2. Taille du génome et nombre de chromosomes chez certains organismes vivants.

a. Le caryotype

Le caryotype est une caractéristique de l'espèce. C'est l'ensemble ordonné de chromosomes d'une cellule bloquée à la métaphase de la mitose. Il est défini par le nombre de chromosomes ($2n = 46$ chez l'homme) et la forme (Méta-, Subméta- et acrocentrique) de ces chromosomes.

La figure 11.a représente un classement des chromosomes humains d'après leur taille et la position du centromère; 7 groupes sont ainsi définies: A: 1-3, B: 4-5, C: 6-12, D: 13-15, E: 16-18, F: 19-20 et G: 21-22. Le protocole de coloration utilise (bandes R) fait apparaître un profil caractéristique pour chaque chromosome; les télomères sont en particulier bien mis en évidence avec cette technique. Le point marque la place du centromère. Le document de la figure 11.b représente le caryotype humain avec l'étalement des chromosomes d'une cellule humaine en **métaphase**.

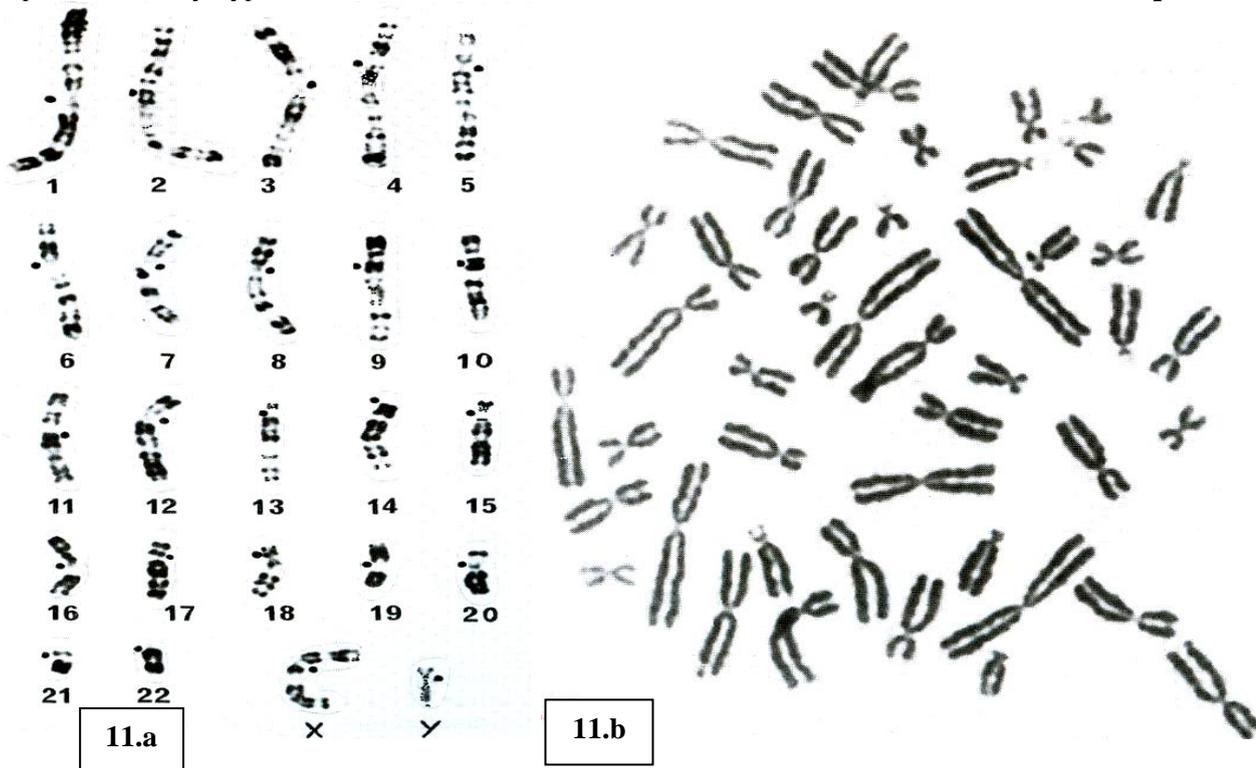


Figure 11. Chromosomes humains et caryotype.

b. Régions et bandes chromosomiques

Chaque chromosome contient deux bras (court 'p' et long 'q'), et plusieurs techniques de colorations (à la Giemsa, moutarde de quinacrine,...) permettent de distinguer des zones caractéristiques sombres et clairs sur les chromosomes.

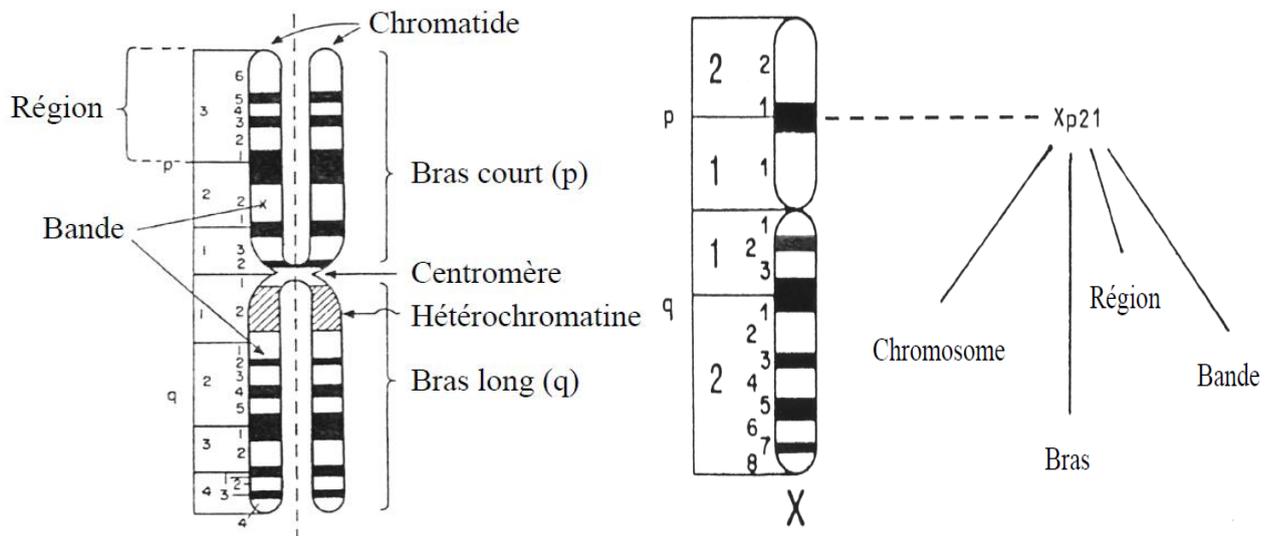


Figure 12. Régions et bandes chromosomiques.

Chapitre VI. ANALYSE GENETIQUE CHEZ LES ORGANISMES DIPLOÏDES

VI.1. Mendel et la notion de gènes et transmission de l'information génétique

VI.1.1. Mendel et la génétique

Le premier chercheur ayant utilisé de véritables techniques génétiques fut **Gregor Mendel** vers 1856 malgré qu'à son époque on ne connaît ni gènes ni chromosomes. Ses résultats sont remarquables puisqu'il proposa une "théorie particulière de l'hérédité" pleinement confirmée par la démonstration récente selon laquelle l'information génétique est codée d'une façon discontinue, en gènes, dans la molécule d'ADN.

Il n'est pas question ici de retracer toute l'œuvre de Mendel ni de consacrer une part importante de ce cours à la génétique **formelle** (utilise certains termes aigus comme «caractère» morphologique) mais la méthode d'analyse mendélienne est d'autant plus d'actualité que l'étude du polymorphisme de l'ADN telle qu'on la pratique maintenant fait appel aux mêmes principes. De très nombreux biologistes avant Mendel avaient observé le résultat de croisements de plantes ou d'animaux sans en tirer le moindre enseignement. Pour comprendre l'essentiel du mécanisme de l'hérédité, Mendel a réalisé simplement des croisements de pois, *Pisum sativum*, et n'a observé que quelques caractères phénotypiques (Figure 1). Sa démarche d'analyse est simple et sert de modèle jusqu'à nos jours :

- * il a choisi, comme matériel d'étude, une espèce à cycles relativement courts, autogame (dont les fleurs s'autofécondent) mais malgré tout à fleur accessible et pollinisable manuellement.
- * il existait de nombreuses variétés de pois aisément discernables les unes des autres.
- * il s'est intéressé à des caractères qualitatifs bien tranchés : aspect **lisse** ou **ridé** du grain, couleur de la fleur, des cotylédons... dont l'expression n'est pas soumise à l'environnement. Les caractères quantitatifs, qui peuvent varier d'une façon continue (taille de la plante, poids des grains ...) sont beaucoup plus difficiles à appréhender bien que se transmettent selon les mêmes règles que les autres.
- * il n'a étudié que peu de caractères à la fois (un seul au début).
- * il a été le premier à utiliser l'outil mathématique en biologie.
- * il avait un grand sens de l'analyse et de la synthèse et ne s'enlisait pas dans les détails

Graine		Fleur	Cosse		Tige	
Forme	Cotylédons	Couleur	Forme	Couleur	Emplacement	Taille
						
Gris & lisse	Jaune	Blanc	Plein	Jaune	Cosse axiale Fleur tout du long	Long (~3m)
						
Blanc & Ridé	Vert	Violet	Étroit	Vert	Cosse terminales Fleurs en haut	Court (~30 cm)

Figure

1 : Les phénotypes des petits pois comestibles (*Pisum sativum* L.).

Les travaux de Mendel lui ont permis de proposer trois principes ou lois de Mendel:

- Loi de dominance
- Loi de disjonction des caractères
- Loi de ségrégation indépendante des caractères

VI.1.2. Analyse mendélienne de la transmission d'un caractère (monohybridisme, modèle 3:1)

Cette analyse va se faire par croisement de variétés ne différant que par un caractère (d'où le terme de monohybridisme). L'exemple le plus célèbre est celui du caractère déformé de la graine de pois, **lisse** (=ronde) chez la plupart des pois mais **ridé** pour plusieurs variétés établies à l'époque de Mendel. Une **variété établie** veut dire ici que les graines produites par des plantes issues de graines rondes, par autofécondation, présentent la même caractéristique de production de graines rondes. Il en est exactement de même pour des "lignées" à grains ridés. C'est cette notion de continuité génétique qui permet l'obtention de ce que nous allons appeler des **lignées pures**.

Ayant observé et sélectionné ce caractère nettement discriminant, Mendel va croiser des plantes de ces deux lignées, plantes définies comme parents (**P**). Le résultat est l'obtention de plantes (**F1** ou filiation de 1^{ère} génération)

qui toutes produisent des grains lisses. L'homogénéité des produits de première génération est remarquable et va constituer un élément important de l'analyse. Une génération **F2** obtenue par autofécondation d'individus F1 présente à nouveau les deux phénotypes : grains lisses et ridés, le dénombrement fait ressortir dans cette population de seconde génération trois quarts de grains lisses pour un quart de grains ridés : les deux caractères parentaux se répartissent (ségrégent) dans des proportions $\frac{3}{4}$ et $\frac{1}{4}$ (Figure 2). Mendel a pu généraliser cette observation en croisant d'autres lignées ne différant que par un caractère, Plusieurs "traits" qualitatifs ont été suivis avec les résultats dans le Tableau 1.

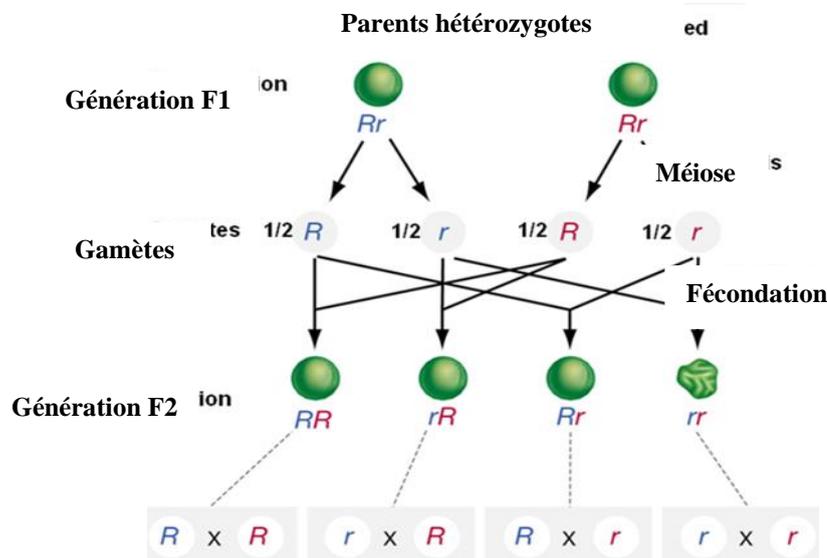


Figure 2. Croisements de pois avec le caractère "aspect du grain": 'R' pour allèle lisse dominant et 'r' pour allèle ridé récessif.

Phénotypes parentaux en croisement	Phénotype de F1	Phénotype de F2	Rapport des 2 phénotypes de F2
Graines lisses × ridées	100 % lisses	5 474 lisses 1 850 ridées	2,96:1
Graines jaunes × vertes	100 % jaunes	6 022 jaunes 2 001 vertes	3,01:1
Pétales violets × blancs	100 % violets	705 violets 224 blancs	3,15:1
Gousses jaunes × vertes	100 % vertes	428 vertes 152 jaunes	2,82:1
Fleurs axiales × terminales	100 % axiales	651 axiales 207 terminales	3,14:1

Tableau 1. Résultats de Mendel dans les croisements de lignées pures de pois différant par un caractère.

L'observation des produits de l'autofécondation de la génération F2 était la suite logique du travail. Le résultat pour le caractère "aspect du grain", a été le suivant : les plantes issues des pois à phénotype grain ridé reproduisent invariablement le même phénotype. Il en va autrement des pois de phénotype grain lisse, bien que certains reproduisent invariablement le caractère grain lisse, d'autres produisent un mélange des deux phénotypes dans des proportions de $\frac{3}{4}$ de ronds pour $\frac{1}{4}$ de ridés. Il fallait par la suite procéder méthodiquement en tirant parti de l'aspect **qualitatif** puis **quantitatif** des résultats. Sur le plan qualitatif, il est clair qu'une information est transmise et non un caractère: aucune plante ne présente le caractère ridé en F1 cependant il réapparaît en F2. Il faut bien admettre que ces informations sont conservées intactes (pures) à travers les générations. Toute plante possède en fait deux informations pour un même caractère (deux formes alléliques), l'une d'entre elles (ridé pour la forme du grain ou blanc pour la couleur de la fleur...) n'étant pas exprimée en F1, on dira que l'allèle (lisse ou violet) est **dominant** sur l'autre. L'allèle masqué en F1 est appelé **récessif**. Mendel a perçu la **diploïdie** de la cellule somatique et l'a symbolisé par des lettres: R (majuscule) va représenter l'allèle lisse dominant, r (minuscule), l'allèle ridé, récessif. L'information (double) pour une plante de lignée pure de phénotype grain lisse sera écrite R/R c'est ce qui représente le **génotype** de la plante pour le caractère

forme du grain. De la même façon, le génotype de la lignée récessive pour ce même caractère sera écrit r/r. Ces génotypes sont dits **homozygotes**.

Mendel a également compris qu'au cours de la formation des gamètes, s'il y a disjonction de ces deux informations, ce que l'on connaît actuellement de **la méiose** et de la séparation des chromosomes homologues (porteurs d'éléments d'information homologues), apporte un support à l'hypothèse de Mendel. Il est clair qu'une recombinaison va avoir lieu au moment de la fécondation, les combinaisons parentales d'information sont : R avec R pour l'un avec production de gamètes R et r avec r pour l'autre et production de gamètes r, l'information d'un individu issu de ce croisement regroupera R et r et le génotype s'écrira R/r (**Hétérozygote**). Ceci représente une combinaison nouvelle par rapport à R/R ou r/r et constitue une première prise de contact avec la notion fondamentale en génétique de recombinaison ("Re- combinaison").

VI.1.3. Analyse mendélienne de la transmission de deux caractères autosomiques et indépendants (dihybridisme, modèle 9:3:1:1)

Comme on l'a vu, Mendel a montré expérimentalement que les mêmes règles régissent la transmission de plusieurs caractères phénotypiques chez le pois. Or, s'il envisage des variétés différant par deux de ces caractères, par exemple une à grains lisses de couleur jaune et l'autre avec des grains ridés et verts, le croisement de ces lignées parentales donne une F1 homogène avec des grains lisses et jaunes (ceci confirme la dominance de ces caractères), et une F2 dont les phénotypes sont les suivants : lisse-jaune : 9/16^{ème} de la population ; ridé-vert ; 1/16^{ème} lisse-vert : 3/16^{ème} ; ridé-jaune : 3/16^{ème}. (Modèle **9:3:3:1**, Tableau 2).

Phénotypes parentaux	Phénotype de F1	Phénotype de F2	Rapport des phénotypes de F2	Rapport des phénotypes de F2 pour le caractère lisse/ridé	Rapport des phénotypes de F2 pour le caractère jaune/vert
Graines Lisses et jaunes (RR JJ) X Graines ridées et vertes (rr jj)	100 % de Graines Lisses et jaunes (Rr Jj)	315 lisses et jaunes	9	423 / 133 (3,18:1)	416 / 140 (2,97:1)
		108 lisses et vertes	3		
		101 ridées et jaunes	3		
		32 ridées et vertes	1		

Tableau 2. Résultats obtenus par Mendel dans les croisements de deux lignées pures différant par deux caractères (en F2, on obtient des types parentaux R,J et r,j et recombinés R,j et r,J).

Pour les deux caractères pris séparément, les proportions de phénotypes sont 12/16^{ème} et 4/16^{ème} soit le modèle **3:1**, les gènes se comportent donc d'une façon indépendante, les unités d'information pour différents caractères sont séparées les unes des autres : on dira qu'il y a **ségrégation indépendante** de ces unités à la méiose et recombinaison à la fécondation comme le fait apparaître l'analyse des génotypes. La méthode de calcul des rapports génétiques dans les hybrides est basée sur le produit des rapports et s'applique aux phénotypes comme aux génotypes (Figure 3). Lorsque qu'on écrit (R-) ou (J-) cela signifie que l'allèle dominant (R ou J) qui impose le phénotype peut être indifféremment apparié avec un autre allèle dominant (RR ou JJ) ou avec un allèle récessif (Rr ou Jj) sans modification du phénotype.

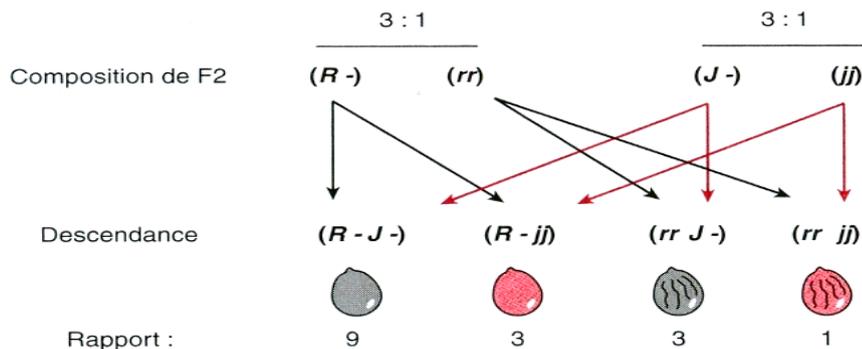


Figure 3. Produits des rapports génétiques pour un double hybride chez le pois.

Un système pour résoudre les croisements dihybrides est donné par l'échiquier de Punnett. Cette méthode utilise un tableau à double entrée pour montrer les génotypes de tous les gamètes possibles de chaque parent, l'un sur la rangée supérieure, et l'autre sur la colonne de gauche. Les combinaisons génotypiques possibles de chaque combinaison de gamète de la descendance sont montrées au niveau des carrés centraux. De la même manière on peut présenter un échiquier phénotypique. Si l'on prend l'exemple du croisement chez les pois (Tableau 2) :

	RRJJ (graines lisses et jaunes) x rrjj (graines ridées et vertes)
F1 :	RrJj x RrJj (tous 100% lisses et jaunes)

En considérant seulement le locus R, **Rr x Rr** produit 1/4 Rr, 1/2 Rr et 1/4 rr. Alors que pour le locus J, **Jj x Jj** produit 1/4 JJ, 1/2 Jj et 1/4 jj. Ces probabilités génotypiques sont placées dans le tableau de l'échiquier génotypique et combinées par multiplication :

F2	1/4 JJ	1/2 Jj	1/4 jj
1/4 RR	1/16 RRJJ	1/8 RRJj	1/16 RRjj
1/2 Rr	1/8 RrJJ	1/4 RrJj	1/8 Rrjj
1/4 rr	1/16 rrJJ	1/8 rrJj	1/16 rrjj

Pour l'échiquier phénotypique, si on considère le locus R, **Rr x Rr** produit 3/4 de graines lisses et 1/4 de graines ridées. Alors que le locus J, **Jj x Jj** produit 3/4 de graines jaunes et 1/4 de graines vertes. Ces probabilités phénotypiques sont placées dans un tableau et combinées par multiplication :

F2	3/4 de graines jaunes	1/4 de graines vertes
3/4 de graines lisses	9/16 graines lisses et jaunes	3/16 graines lisses et vertes
1/4 de graines ridées	3/16 graines ridées et jaunes	1/16 graines ridées et vertes

Pour vérifier expérimentalement les modèles 3:1 et 9:3:3:1, on réalise un **croisement test** ou **test cross** qui s'opère entre l'hétérozygote de la génération F1 avec un homozygote récessif (Par différence, le **rétrocroisement** ou **back cross** est le croisement entre les individus de la descendance F1 avec l'un de leurs parents). Ceci permet d'analyser le génotype sous-jacent à un phénotype dominant puisque le parent homozygote récessif n'apporte que les allèles récessifs à sa descendance. Le croisement test prédit donc un rapport de phénotypes égal à 1:1 (Figure 4).

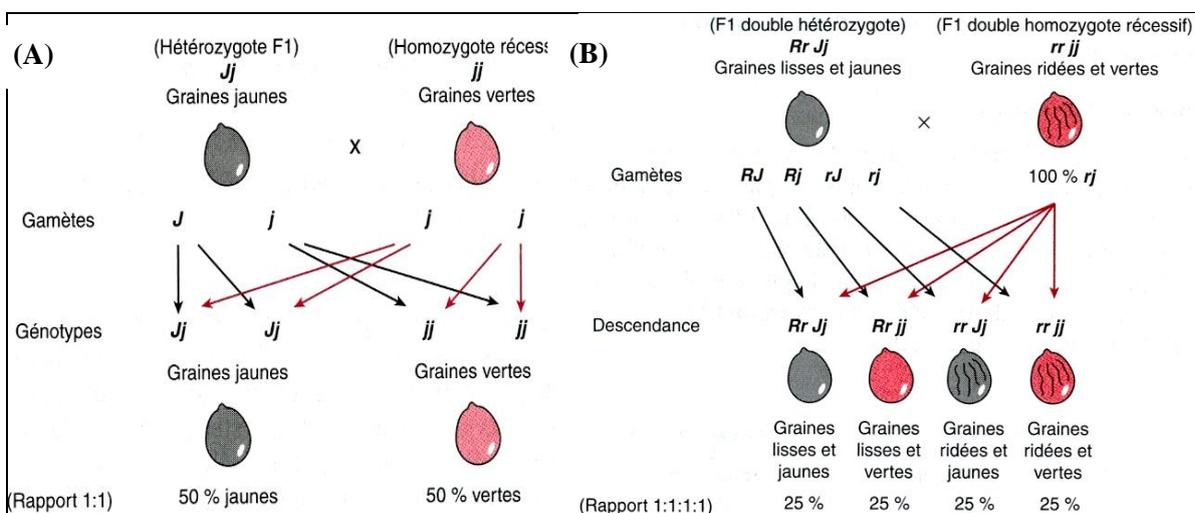


Figure 4. Croisement test d'un pois hétérozygote (hybride F1) avec un pois homozygote récessif pour un caractère (A) ou deux caractères (B).

VI.1.4. Généralisation pour un diagramme d'un triple hybride ou plus

Si l'on admet, comme dans les analyses précédentes, qu'il y a ségrégation indépendante de trois gènes correspondants aux trois couples de caractères avec dominance de l'un des deux allèles on obtiendra à la génération F2 une descendance correspondante au croisement ($AaBbCc \times AaBbCc$) comme représenté dans le diagramme de la figure 5.

Si l'on suit les flèches du haut vers le bas et on effectue les produits successifs, on aura le modèle **27:9:9:3:3:3:1**, qui est une variante combinée du modèle de départ de Mendel 3:1, avec 8 phénotypes et 27 génotypes (Figure 4). En augmentant le nombre de gènes appariés participant aux croisements hybrides, avec toujours le principe de ségrégation indépendante, le nombre de phénotypes et de génotypes deviendra rapidement très élevé (Tableau 3).

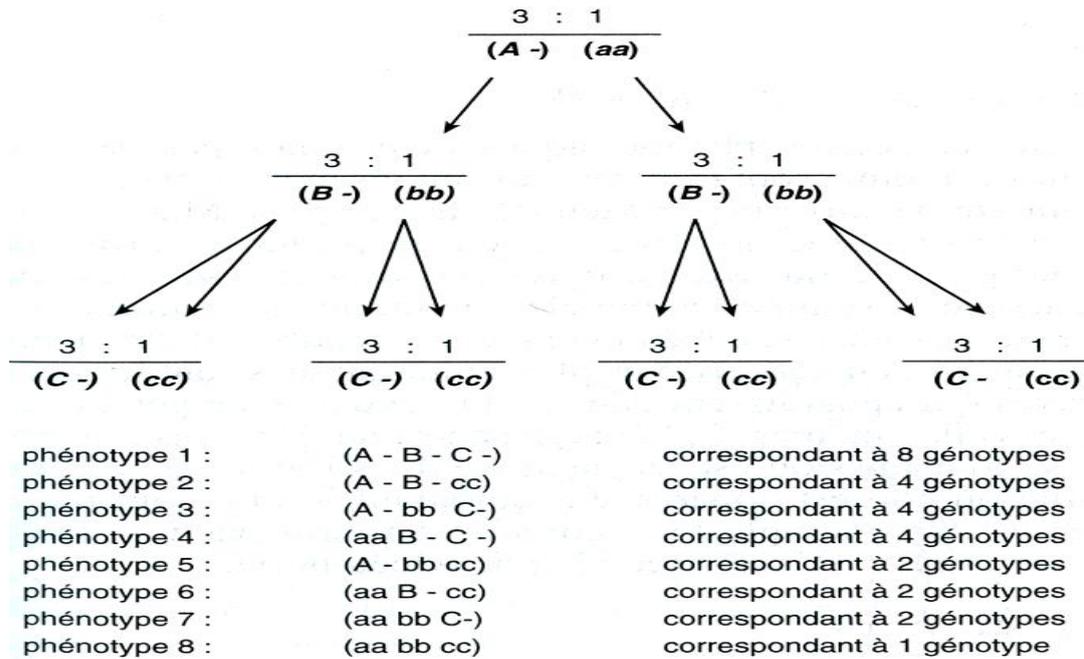


Figure 5. Généralisation du calcul des rapports génétique dans les hybrides: Trihybridisme.

Nombre de gènes appariés	Nombre de phénotypes	Nombre de génotypes
1	2	3
2	4	9
3	8	27
4	16	81
-	-	-
n	2^n	3^n

Tableau 3. Nombre de phénotypes et génotypes en fonction du nombre de gènes apparentés dans les croisements hybrides.

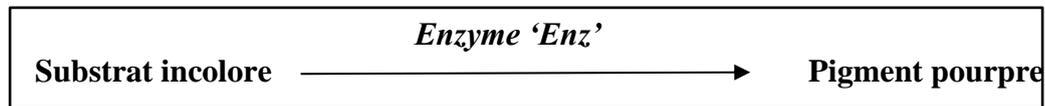
VI.2. Extension des lois de Mendel

VI.2.1. De l'analyse mendélienne aux gènes

Dans la mémoire de Mendel en 1866 on ne trouve nulle part référence aux gènes ou à la génétique, deux mots qui n'apparaissent qu'en 1905. La même chose pour les chromosomes ou l'hérédité. Un pas considérable sera franchi lorsque l'on découvrira que derrière ces majuscules et minuscules (R et r par exemple), qui symbolisent des formes alléliques d'un même gène, se cachent des séquences d'ADN, presque semblables, mais pas tout à fait, situées au même endroit (au même **locus**), dans deux chromosomes

homologues. De la même façon, les réponses à la question de l'**expression** du matériel génétique vont éclairer des notions restées longtemps abstraites telle que la dominance par exemple.

Les travaux postérieurs ont montré que l'expression d'un gène passe par la réalisation d'une protéine spécifique, le gène, en tant qu'unité d'information "code" une protéine. On peut en rester là : un caractère phénotypique peut très bien être une protéine de structure. Il peut également s'agir d'une protéine enzymatique et un exemple très simple de la réalisation du caractère phénotypique "**fleur pourpre**" étudié par Mendel, pourrait être le suivant :



L'enzyme 'Enz' représente le produit direct d'un gène. En reprenant la symbolique déjà utilisée, si E représente l'information pour 'Enz' : **E/E** représente un génotype homozygote avec, sur chaque chromosome homologue une information permettant la synthèse de l'enzyme fonctionnelle et donc la possibilité de transformer le substrat incolore en pigment. **e/e** représente un génotype, homozygote également mais, pour chaque chromosome homologue, le locus est occupé par une information erronée ne permettant pas la synthèse d'une enzyme fonctionnelle, en conséquence, le substrat ne sera pas métabolisé et la fleur apparaîtra blanche (absence de pigment). L'hétérozygote **E/e** possède à la fois une information normale et une information erronée, l'information correcte permet la réalisation d'une enzyme fonctionnelle et la synthèse du pigment car elle "**complémente**" l'information incorrecte ce qui se traduit par: l'allèle 'E' qui est dominant sur l'allèle 'e' (récessif) car chez un hétérozygote E/e c'est le phénotype pigmenté qui se manifeste.

Cette incursion dans le domaine de l'expression de l'information génétique va nous permettre d'aborder, à l'aide d'un seul exemple, deux notions importantes (par leurs conséquences), la **codominance** et le **pluriallélisme**.

VI.2.2. Codominance et pluriallélisme

Prenons un exemple en génétique humaine. Les groupes sanguins sont définis par la présence d'antigènes spécifiques à la surface des hématies. Dans le système ABO (Tableau 4), nous allons nous intéresser aux antigènes 'A' et 'B'. La production de ces molécules dépend d'informations génétiques situées à un locus précis d'un chromosome (I). Ce locus peut porter le code de l'enzyme responsable de la synthèse de l'antigène A et sera symbolisé par I_A , l'individu de génotype I_A/I_A (homozygote) va donc synthétiser des antigènes de type A, l'individu I_B/I_B porte l'information pour des antigènes de type B et va les produire. Un hétérozygote I_A/I_B possède deux informations différentes correspondant à deux allèles différents pour un même locus, le phénotype de cet individu est AB, les deux informations sont exprimées (les deux antigènes A et B sont présents), on dit qu'il y a **codominance**.

Dans une cellule diploïde, il est évident que deux allèles seulement peuvent coexister car il n'existe que deux chromosomes homologues, ceci ne veut pas dire qu'il n'existe que deux allèles possibles capables d'occuper ce locus, on verra que théoriquement il peut en apparaître une infinité. Dans le cas des groupes sanguins, il existe un allèle qui ne spécifie aucune enzyme capable de produire un antigène (la fonction est perdue, **allèle nul**), on symbolise cet allèle par 'i'. En conséquence, un individu I_A/i portera une information "antigène A" sur un chromosome et une information "aucun antigène" sur le chromosome homologue. Ceci nous permet de comprendre que l'individu sera du groupe A (on dira que I_A est dominant) de la même façon que l'individu de génotype I_B/i sera de groupe B, seul un homozygote récessif i/i sera de groupe zéro ou 0 (avant l'informatique on ne distinguait pas la lettre O du zéro).

La suite du travail de Mendel, l'analyse d'autres variants, a permis d'aboutir à une notion de gènes en tant qu'unités d'informations **discontinues** et **indépendantes**. Il s'agit d'étudier des variants se distinguant par 2, 3 ou plusieurs caractères (cas de di, tri, et polyhybridisme).

Dans certains cas, la codominance peut conduire à un nouveau phénotype qui a des traits intermédiaires entre les phénotypes des parents (par exemple le **caractère couleur**).

Groupe (phénotype)	Génotype
H (O, ni A, ni B)	ii
A	I ^A I ^A ou I ^A i
B	I ^B I ^B ou I ^B i
AB	I ^A I ^B

Tableau 4. Groupes sanguins ABO humains avec quatre phénotypes et trois allèles principaux.

VI.2.3. L'hérédité liée au sexe

C'est en désirant valider ces théories sur les mutations que **Thomas Morgan** (1866-1945) travaille sur la mouche du vinaigre, *Drosophila melanogaster*.

La Drosophile, *Drosophila melanogaster*, est l'un des organismes diploïdes les mieux connus des généticiens. C'est en étudiant la transmission des caractères héréditaires dans cette espèce que Morgan a pu établir dans les années 1900 la **'théorie chromosomique de l'hérédité'** qui peut se résumer de la façon suivante : les gènes sont des éléments matériels alignés bout à bout le long des chromosomes et dont la transmission reflète fidèlement le comportement des chromosomes lors de la méiose et de la fécondation.

Plusieurs centaines de mutations sont connues chez la Drosophile et une carte détaillée de son génome a été établie et corrélée précisément à l'organisation physique des chromosomes. Les lois statistiques de la transmission des caractères héréditaires ont pu ainsi être confirmées et étendues. Aujourd'hui, le séquençage de son génome est fini et la drosophile constitue un des principaux organismes modèles en biologie du développement.

a. Reconnaissance des sexes et des différentes mutations étudiées

La figure 6 présente les principales différences phénotypiques entre le mâle et la femelle, observées attentivement à la loupe binoculaire chez quelques individus :

- a- Les mâles sont un peu plus petits que les femelles.
- b- L'extrémité des abdomens des mâles en vue dorsale est arrondie et presque noire alors que celle des femelles est pointue et plus claire (2).
- c- Les mâles possèdent un "peigne sexuel" (3) situé sur les pattes antérieures.

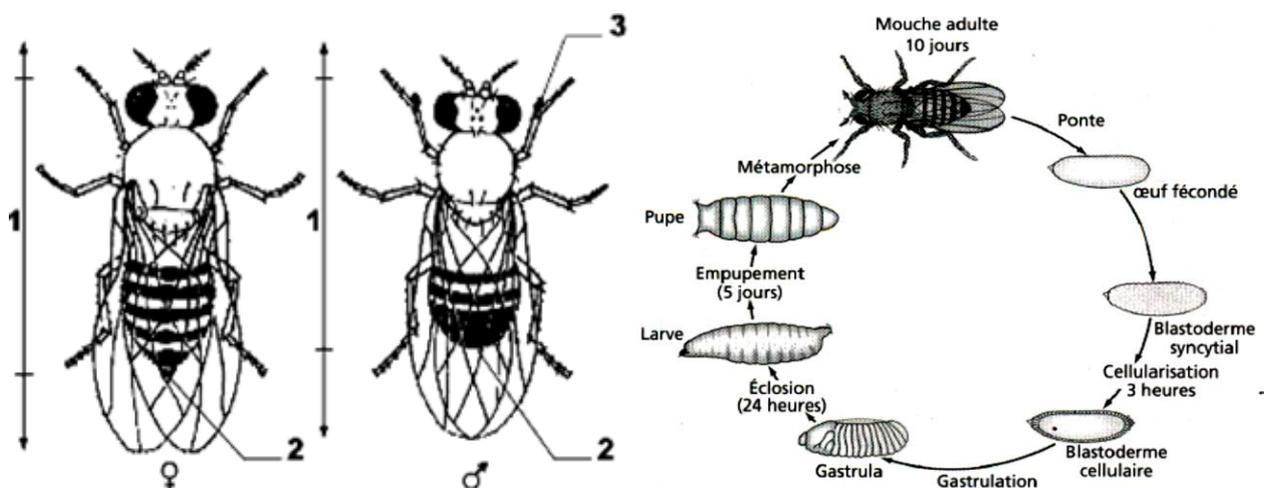


Figure 5. A gauche, différences phénotypiques entre mâles et femelles de drosophile d'une souche sauvage ; à droite, cycle de la vie de la drosophile.

Les mutations étudiées porteront sur les caractères suivants :

- * **Couleur des yeux** : w = œil blanc (mutant) ; w+ = œil rougeâtre (sauvage).
- * **Forme des ailes** : vg = aile atrophiée (mutant) ; vg+ = aile sauvage.
- * **Couleur du corps** : e = couleur ébène (mutant) ; e+ = couleur sauvage.

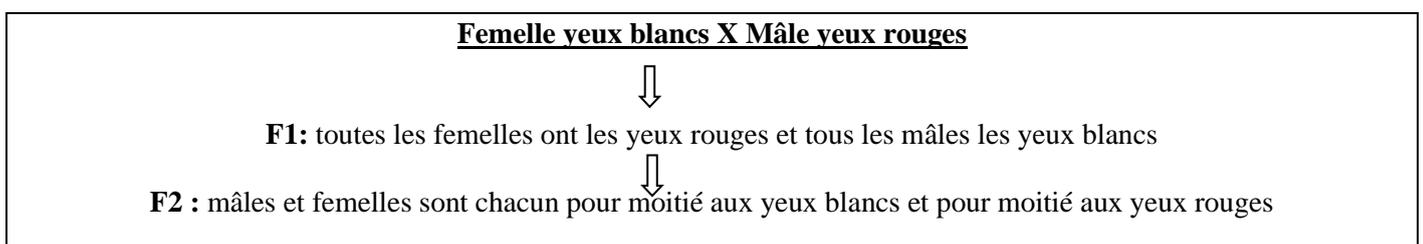
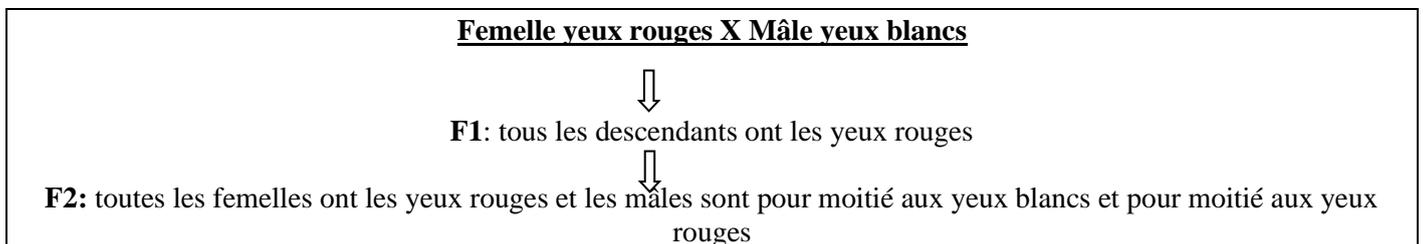
La théorie chromosomique de l'hérédité a permis d'ouvrir la boîte noire de la génétique (entre 1920 et 1922) en montrant que les gènes postulés par la génétique formelle (de Mendel) sont portés par les chromosomes, les généticiens trouvent dans la ségrégation des chromosomes à la méiose et la formation de gamètes haploïdes, suivi de l'union aléatoire des gamètes à la fécondation, les mécanismes concrets, objectifs, cytologiques qui permettent d'expliquer la loi de pureté des gamètes (un chromosome donc un seul allèle de chaque gène par gamète) et les fréquences des divers types de descendants d'un croisement (en phénotype et en génotype) résultant de l'union aléatoire des gamètes à la fécondation. En plus, cette théorie permet d'expliquer la plupart des **exceptions** au mendélisme observées depuis le début du 20^{ème} siècle, par l'introduction du concept de liaison génétique et de recombinaison génétique par **crossing-over** qui vient compléter le concept de recombinaison génétique par réassortiment aléatoire des chromosomes (Figure 7). Il fallait attendre le milieu du 20^{ème} siècle, entre 1942 et 1950, pour que la boîte noire de la génétique s'ouvre enfin à la connaissance du mode d'action des gènes, quand la collaboration entre généticiens et biochimistes aboutit au dogme «un gène-une chaîne peptidique».

La première mutation obtenue par Morgan est un mâle aux yeux blancs dans une population aux yeux rouges (caractère sauvage). Comme il n'apparaît que chez le mâle, l'idée lui vient de l'associer à un chromosome sexuel: c'est le premier modèle d'hérédité liée au sexe. En effet, chez la drosophile, comme chez de nombreux organismes, il existe une différence chromosomique entre les deux sexes: la différence la plus commune étant le remplacement de l'un des chromosomes d'une paire d'homologue par un chromosome différent dans l'un ou l'autre des sexes (comme pour l'homme qui est XY et la femme XX, XX est dit **homogamétique** et XY **hétérogamétique**).

Ce déterminisme chromosomique du sexe n'est pas le seul type de déterminisme mais son étude est fort intéressante. La drosophile a 8 chromosomes dont 6 autosomes et 2 hétérosomes (Chromosomes sexuels). Le déterminisme du sexe est différent chez les espèces:

- * Y détermine le sexe mâle chez l'Homme (X0 est féminin, monosomie X ou syndrome de Turner).
- * Chez la drosophile c'est pas clair car X0 sont des males.
- * Chez les oiseaux, le sexe mâle est homogamétique (Z/Z), celui femelle est hétérogamétique (Z/W).

b. Les expériences de Morgan



c. Analyse formelle

- On utilisera la notation classique (+) désignant l'allèle sauvage conduisant au phénotype yeux rouges.
- (w) désignant l'allèle muté conduisant au phénotype yeux blancs (white).
- L'absence d'allèle porté par le chromosome homologue chez le mâle sera désigné par la lettre y.
- On dit que le gène porté par la partie propre de l'X est hémizygoté chez le mâle et que le caractère yeux blancs (en fait l'allèle) est lié au sexe (liaison au sexe).

On peut représenter par un tableau à double rentrée (**Echiquier de Punnett**), les génotypes des gamètes possibles des parents et les combinaisons génotypiques dans la descendance.



Femelle (+//+) x (w//y) mâle

	w	y
+	+//w	+//y

F1: toutes les femelles ont les yeux rouges (+//w) ainsi que tous les mâles (+//y)

	+	y
+	+//+	+//y
w	w//+	w//y

**F2: femelles à yeux rouges (pour moitié +//+ et pour moitié w//+)
mâles pour moitié à yeux rouges (+//y) et pour moitié à yeux blancs (w//y)**



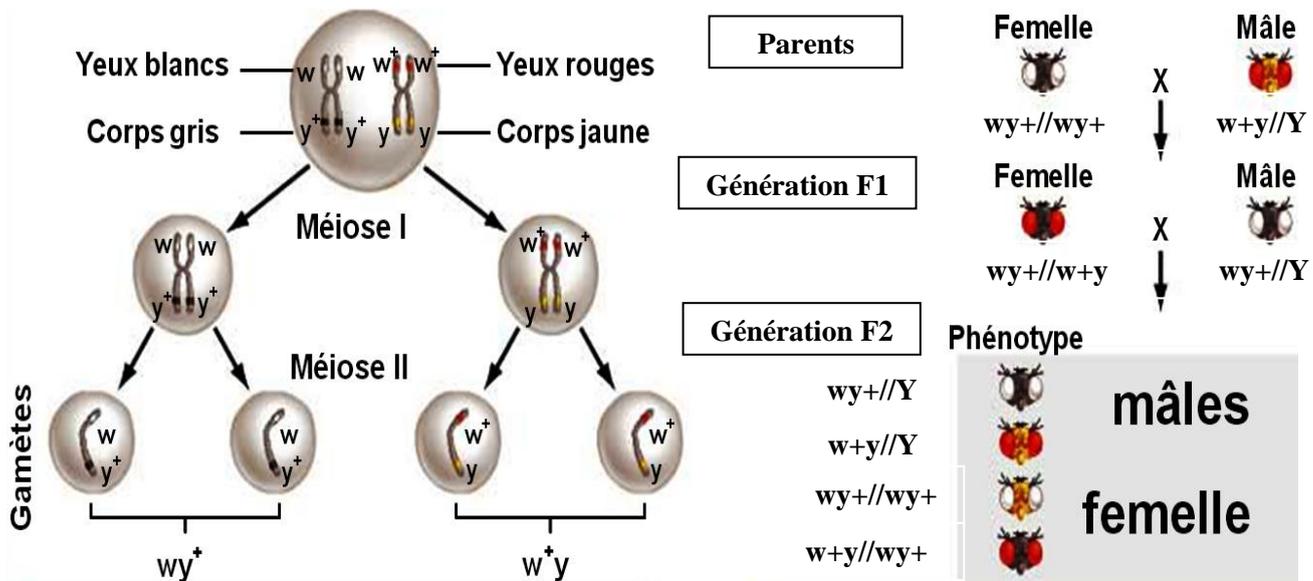
Femelle (w//w) x (+//y) mâle

	+	y
w	w//+	w//y

F1: toutes les femelles ont les yeux rouges (w//+) alors que tous les mâles ont les yeux blancs (w//y)

	w	y
w	w//w	w//y
+	+//w	+//y

**F2: femelles pour moitié à yeux rouges (+//w) et pour moitié à yeux blancs (w//w)
mâles pour moitié à yeux rouges (+//y) et pour moitié à yeux blancs (w//y)**



Quand deux loci se trouvent sur un même chromosome, La méiose conduit à deux gamètes différents seulement car pas de ségrégation indépendante

Résultats s'il n'y a pas de recombinaison homologue

Figure 7. Un exemple de vérification de l'hypothèse par Morgan.

VI.2.4. Distinction entre phénotype et génotype

Le phénotype d'un être vivant correspond à son apparence physique d'ensemble ou bien à un aspect particulier (on dit aussi une caractéristique, un trait) mesurable de cette apparence générale, comme sa taille, son poids, sa couleur, une forme, une propriété, une aptitude ou un comportement qui varient suivant les individus examinés. Le phénotype dépend principalement du génotype, c'est-à-dire des gènes et surtout de leurs allèles présents dans les chromosomes de l'individu. La plupart des phénotypes sont déterminés par de nombreux gènes et sont influencés par les facteurs physico-chimiques et biologiques de l'environnement dans lequel vit l'individu ou la population à laquelle il appartient.

La description du phénotype d'un individu s'avère en général plus facile à réaliser - par de simples mesures structurales, biochimiques, physiologiques ou comportementales - que celle de son génotype qui nécessite d'accéder à la séquence de l'ADN. La détermination de cette séquence n'a été possible que depuis quelques dizaines d'années. Auparavant les généticiens déduisaient la fonction des gènes à partir du seul examen des phénotypes. La génétique classique ou génétique directe part donc des phénotypes pour aller vers les génotypes, à l'aide souvent de mesures et d'analyses quantitatives portant sur les phénotypes (**génétique quantitative**). Avec les progrès de la biologie moléculaire, une génétique inverse se développe qui part des génotypes et parfois des génomes entiers, pour connaître les phénotypes. Adeptes de la première génétique, cultivateurs et éleveurs, intègrent aujourd'hui les démarches de la génétique moléculaire dont la puissance prédictive des phénotypes s'accroît avec les progrès réalisés dans la connaissance approfondie des génomes entiers (séquençage et génotypage).

Les différences constatées entre les phénotypes individuels sont révélatrices de la variabilité génétique existant entre les individus. La sélection naturelle (Voir Chapitre IV) agit sur les phénotypes en favorisant ou non les individus dont les phénotypes assurent la plus ou moins bonne adaptabilité à l'environnement. Sans différences phénotypiques, il ne pourrait y avoir de différences dans l'adaptabilité et par voie de conséquence pas de sélection naturelle et pas d'évolution.

Le jeu complexe qui s'établit entre phénotype, génotype et milieu peut se représenter:

$$\text{Génotype} + \text{Environnement} = \text{Phénotype}$$

Ou, mieux encore afin de tenir compte des variations aléatoires, par la relation:

$$\text{Génotype} + \text{Environnement} + \text{Variation aléatoire} = \text{Phénotype}$$

Ces relations expriment d'une manière simpliste les interactions qui s'établissent entre les gènes et l'environnement. Elles expriment également les termes du débat classique et toujours d'actualité entre ce qui appartient à l'inné (héritable par les gènes) et ce qui appartient à l'acquis (apporté ou non par l'environnement).

VI.3. Liaison génétique et cartographie

VI.3.1. Distance génétique et carte génétique

Les gènes sont disposés sur le chromosome suivant un ordre linéaire. La cartographie génétique a deux caractéristiques principales:

- a. détermination de l'ordre linéaire selon lequel les gènes se succèdent le long du chromosome (ordre des gènes),
- b. détermination des distances relatives entre chaque locus (distances entre gènes).

L'unité de distance est l'expression de la **probabilité** avec laquelle un crossing-over apparaît entre deux gènes considérés (fréquence de recombinaison). Par conséquent, un **centimorgan (cM = unité de distance génétique)** correspond à la distance nécessaire pour, qu'entre deux gènes, la probabilité d'avoir un crossing-over soit de 1 %. Chez l'homme, on considère qu'une distance d'un cM équivaut en moyenne à une séquence d'ADN de 10^6 pb.

Chaque chiasma donne 50 % de recombinants. 50 % de crossing-over équivaut à 50 unités de distance. Si, pour une paire de chromosomes, le nombre de chiasmas est connu, la longueur totale de la carte, pour cette liaison, peut être estimée par :

Longueur totale = nombre estimé de chiasmas x 50

VI.3.2. Le test deux points

Pour détecter les gamètes recombinants, chez le **dihybride**, on utilise la technique du **croisement test**. Supposons qu'un croisement test soit effectué sur un double hétérozygote en phase couplée (AC/ac), et que l'on obtienne dans les phénotypes de la descendance: 37 % de dominants aux deux loci, 37 % de récessifs aux deux loci, 13 % de dominants pour le 1^{er} locus et récessifs pour le second, 13 % de dominants pour le second locus et récessifs pour le 1^{er}. Les deux derniers groupes (de génotype **Ac/ac** et **aC/ac**) sont produits par des gamètes recombinants de parents doubles hétérozygotes.

En résumé, 26 % (13 + 13) de tous les gamètes sont de type recombinant, donc la distance entre les loci A et C est estimée à 26 unités.

VI.3.3. Le test trois points

Habituellement, il n'y a pas de double crossing-over entre des gènes distants de moins de cinq unités. Pour des gènes plus éloignés, l'utilisation d'un **troisième marqueur**, entre les deux premiers, permet de mettre en évidence les doubles crossing-over. Supposons que l'on effectue un croisement test sur un individu triple hétérozygote de génotype **ABC/abc** et que la descendance soit la suivante:

36% ABC/abc	9% Abc/abc	4% ABc/abc	1% AbC/abc
36% abc/abc	9% aBC/abc	4% abC/abc	1% aBc/abc
<u>72% type parental</u>	<u>18% simple crossing-over entre A et B (région I)</u>	<u>8% simple crossing-over entre B et C (région II)</u>	<u>2% double crossing-over</u>

Pour trouver la distance A-B, on doit compter tous les crossing-over (simples et doubles) qui se produisent dans la région I: $18\% + 2\% = 20\%$, donc 20 unités séparent les loci A et B. De la même façon, pour trouver la distance B-C, il faut compter tous les crossing-over qui se produisent dans la région II: $8\% + 2\% = 10\%$, donc 10 unités séparent les loci B et C. La distance A-C est donc de 30 unités quand les doubles crossing-over sont comptabilisés.

Sans le marqueur central B, les doubles crossing-over pourraient être assimilés à des types parentaux, et la distance intergénique A-C serait donc sous-estimée. Dans ce cas, les 2 % de doubles crossing-over s'ajouteraient aux 72 % de types parentaux, faisant un total de 74 % de types parentaux et 26 % de types recombinants. Donc, pour trois

gènes liés, dont les distances sont connues, le nombre de recombinants détectables entre les deux marqueurs externes A et C, quand le marqueur du centre, B, est absent, est:

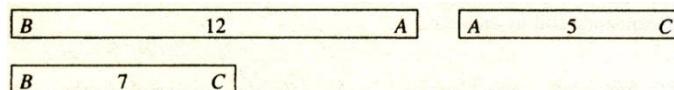
$$(\% \text{ des crossing-over entre A et B}) + (\% \text{ des crossing-over entre B et C}) - (2 \times \% \text{ de double crossing-over})$$

Cette méthode n'est valable que si un crossing-over dans la région A-B a lieu indépendamment d'un crossing-over dans la région B-C.

VI.3.4. Ordre des gènes

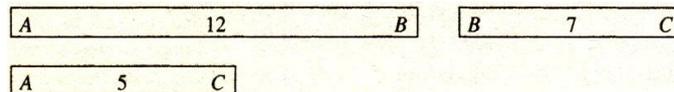
L'additivité des distances génétiques permet de placer les gènes selon un **ordre linéaire**. Trois gènes liés, peuvent être disposés de trois façons différentes en fonction du gène central. Dans un premier temps, on ne tiendra pas compte de l'alternative gauche-droite. Si aucun double crossing-over ne se produit, les distances sont strictement additives. Si A-B = 12 cM, B-C = 7 cM et A-C = 5 cM, il est possible de déterminer l'ordre correct.

• **Cas 1** : Supposons que A soit au centre.



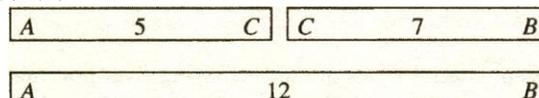
La distance B-C n'est pas correcte. Donc A ne peut pas être au centre.

• **Cas 2** : Supposons que B soit au centre.



La distance A-C n'est pas correcte. Donc B ne peut pas être au centre.

• **Cas 3** : Supposons que C soit au centre



La distance A-B est correcte. Donc C doit être au centre.

• **Relations de liaison à partir d'un test 2 points** : Les combinaisons parentales ont tendance à être majoritaires dans la descendance et les recombinants restent la classe la moins fréquente. À partir de ces informations il est possible de déterminer le mode de liaison (phase couplée ou phase de répulsion) chez les parents dihybrides.

• **Relations de liaison à partir d'un test 3 points** : Dans un test-cross faisant intervenir trois gènes liés, les types parentaux sont les plus fréquents et les doubles crossing-over sont les moins représentés. L'ordre des gènes est déterminé en ajustant les combinaisons parentales, dans un ordre adéquat permettant l'obtention de doubles crossing-over.

VI.3.5. Liaison génétique et sélection

Contrairement aux observations de G. Mendel, certains gènes ne ségrègent pas à la méiose de façon indépendante. Quand les locus de deux gènes sont liés ou très proches l'un de l'autre sur le même chromosome, on dit qu'il existe une liaison génétique. Au cours de la formation des gamètes, un brassage du matériel génétique s'opère grâce à des croisements par recombinaison (**crossing over**). Plus les allèles (gènes) sont proches les uns des autres sur un chromosome, plus faible est la probabilité qu'un événement de recombinaison entre chromosomes vienne à les séparer. Il en résulte que si l'un des gènes est l'objet d'une sélection (naturelle ou artificielle), les autres gènes qui lui sont proches ont toutes les chances d'être eux aussi sélectionnés. Au final, la sélection agit sur le potentiel de variation du génome et suivant les circonstances et l'échelle de temps considérée, elle peut ou non appauvrir la variabilité intrinsèque du génome, objet de la sélection.

Par une sélection systématiquement dirigée, il peut arriver qu'un gène (un allèle de ce gène), responsable d'un

phénotype recherché par l'éleveur ou le cultivateur, voit augmenter sa fréquence au sein de la population. Les allèles qui lui sont liés au locus chromosomique qu'il occupe verront du même coup également leur fréquence augmenter, quelle que soit leur action effective (neutre, positive ou négative). Poussée à l'extrême, une telle sélection conduira à fixer dans la population et pour la région génomique considérée, un seul **haplotype**, c'est-à-dire une seule combinaison allélique (l'allèle du gène cible de la sélection et les allèles des gènes liés voisins).

VI.3.6. L'interférence

Dans la plupart des organismes supérieurs, la formation d'un crossing-over réduit la probabilité qu'un autre crossing-over se produise dans la région immédiatement adjacente du chromosome, c'est ce qu'on appelle le phénomène d'**interférence**. Cette réduction de formation du chiasma peut être due à l'inaptitude physique des chromatides de se lier l'une à l'autre à l'intérieur d'une distance minimale. Le résultat final de cette interférence est que l'on observe moins de doubles crossing-over par rapport à ce qui pourrait être attendus compte tenu de la distance cartographique liant les gènes. L'importance de l'interférence varie en fonction des différents segments du chromosome et est exprimée de manière usuelle par un **coefficient de coïncidence**. Ce coefficient correspond au rapport entre les doubles chiasmats observés et les doubles chiasmats attendus.

$\text{Coefficient de coïncidence} = \frac{\% \text{ doubles crossing-over observés}}{\% \text{ doubles crossing-over attendus}}$

La coïncidence et le complément de l'interférence avec : **Coïncidence + Interférence = 1**

Lorsque l'interférence est totale (1,0) aucun double crossing-over se seront observés et la coïncidence est égale à 0. Inversement si l'on observe tous les chiasmats attendus la coïncidence est de 1 et l'interférence sera nulle. Lorsque l'interférence est de 40% cela implique que la coïncidence devient 60%.

VI.4. Les interactions génétiques

Le phénotype résulte de nombreux produits de gènes exprimés dans un environnement donné. L'environnement inclut en plus des facteurs externes (température, quantité et qualité de la lumière, ..) des facteurs internes comme les enzymes et les hormones. Les enzymes, qui sont des protéines, sont douées d'activités catalytiques responsables de la cassure ou de l'union de molécules variées. Ce sont des réactions métaboliques (anabolique ou catabolique) permettant le maintien de l'intégrité de la cellule vivante, souvent organisées dans une voie métabolique où les réactions biochimiques sont liées.

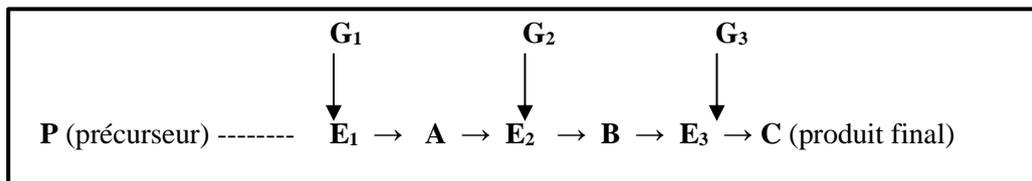


Figure 8. Interactions génétiques dans une voie de biosynthèse métabolique .G, gène ; E, enzyme ; A et B, métabolites intermédiaires.

Plusieurs gènes sont nécessaires pour coder les enzymes impliquées dans une voie métabolique (Figure 8). Chaque métabolite intermédiaire (A, B) est produit par l'action catalytique de divers enzymes (E_n) codées par différents gènes sauvages (G_n). On parle d'interaction génétique lorsque deux ou plusieurs gènes codent des enzymes catalysant les étapes d'une même voie. Si la substance finale C est essentielle à l'obtention du phénotype normale et si les mutants récessifs g_1 , g_2 et g_3 produisent des enzymes déficientes (ou non fonctionnelles), alors un phénotype mutant résulterait d'un génotype récessif homozygote à l'un ou l'autre des trois loci. On distingue plusieurs cas d'interactions génétiques.

VI.4.1. Pléiotropie

Le pléiotropisme concerne toutes les formes d'expression phénotypique d'un même gène c'est-à-dire que l'expression phénotypique d'un gène concerne plus d'un caractère. Parfois un caractère semblera évident (effet majeur) et d'autres, apparemment sans relation (effets secondaires) le seront moins. Dans d'autres cas, l'ensemble des modifications peut être considéré comme un **syndrome** (Ensemble de signes, de symptômes, de troubles dont les causes sont inconnues ou multiples, par opposition à maladie).

Un exemple de maladie génétique chez l'homme de type pléiotropie, le syndrome appelé 'anémie falciforme' (Figure

9) est dû à une hémoglobine anormale. Il se caractérise, chez les porteurs homozygotes, par un retard mental et par diverses déformations incluant des anomalies osseuses de la cornée. Une analyse enzymatique indique la déficience simultanée de plusieurs enzymes lysosomiales différentes. C'est l'effet primaire de la mutation d'un gène codant la N-acétylglucosamine-1-phosphotransférase, enzyme qui intervient dans la phosphorylation des enzymes lysosomiales. Cette modification post-traductionnelle de ces enzymes lysosomiales est nécessaire à leur activité.

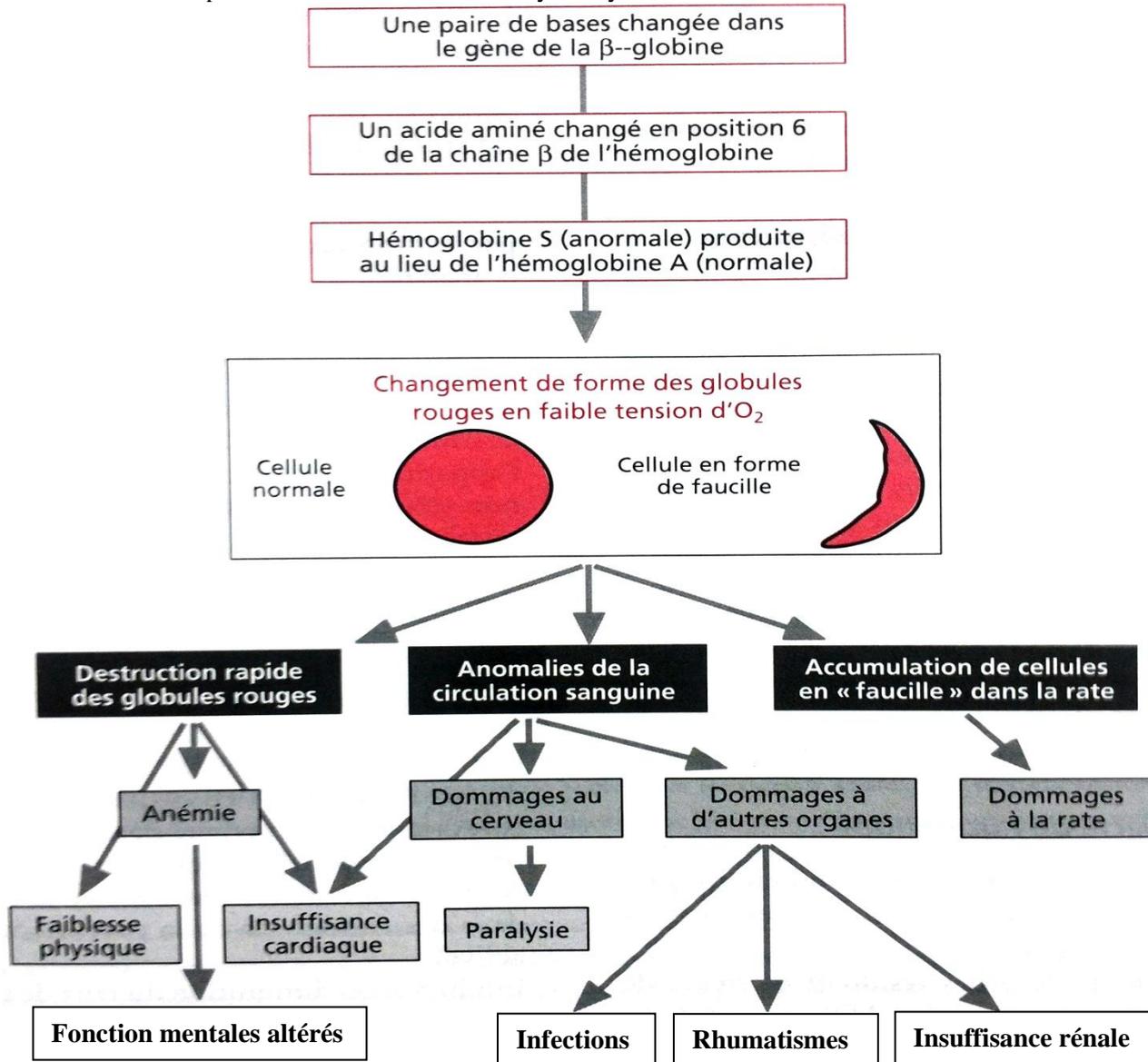


Figure 9. Différents niveaux de changements phénotypiques consécutifs à une mutation dans l'anémie falciforme.

VI.4.2. Epistasie

C'est un cas d'interaction entre gènes lorsque le nombre de phénotypes, apparaissant dans la descendance de parents hybrides, est inférieur à quatre. Il y a plusieurs cas d'épistasie (dominante, récessive, ..). Un exemple d'épistasie est observé lorsque les effets d'un gène ne peuvent être observés en présence d'un allèle particulier d'un autre gène (on parle de masquage comme déduit par Bateson, 1908).

Un exemple d'épistasie récessive est celui observé entre deux gènes impliqués dans la couleur du pelage chez la souris (Figure 10). Lorsque l'on croise des souris au pelage noir et des souris au pelage blanc, on obtient en F1 des souris au pelage noir. La F2 (issue des croisements F1x F1) est composée de 9/16 de souris noires, 3/16 de souris brunes et 4/16 de souris blanches (rapport 9:3:4). Deux gènes indépendants génétiquement, B et C, interviennent dans ce croisement. Le gène B contrôle la couleur proprement dite, noire si l'allèle dominant B est présent, brune chez les homozygotes récessifs b/b. Le gène C, sous sa forme dominante, permet l'expression de la couleur, qu'il interdit l'allèle récessif c. Trois classes phénotypiques sont donc obtenues, dont les proportions suivantes :

- a- couleur noire (9/16) : B/- C/- = B/B C/C ; B/B C/c ; B/b C/C ; B/c C/c
 b- couleur brune (3/16) : b/b C/- = b/b C/C ; b/b C/c
 c- couleur blanche (4/16) : B/- c/c = B/B c/c ; B/b c/c
 d-

En conclusion, l'allèle récessif c est épistatique sur l'expression du gène B, puisqu'il impose un phénotype blanc, quel que soit l'allèle du gène B présent.

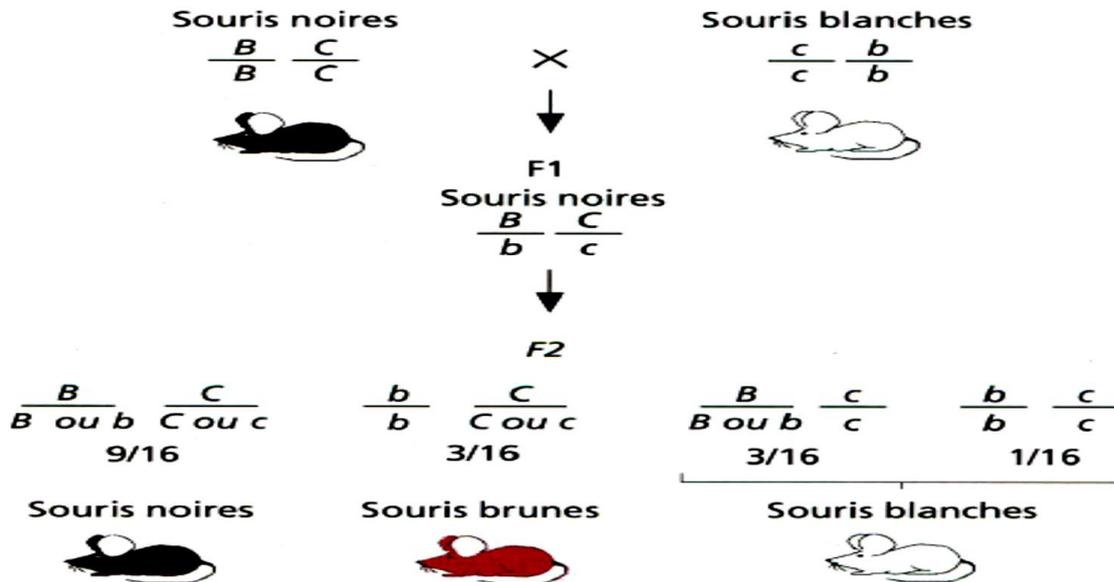


Figure 10. Epistasie entre deux gènes impliqués dans la couleur du pelage chez la souris.

VI.4.3. Létalité

La létalité est la manifestation phénotypique de certains gènes par la mort de l'organisme avant même sa maturité sexuelle. Ceci est dû à certains allèles létaux. L'apparition d'un véritable allèle dominant (c'est-à-dire qu'il tue aussi bien dans des conditions homozygotes qu'hétérozygotes) n'arrive qu'occasionnellement par mutation d'un allèle sauvage. Les individus possédant un allèle dominant léthal meurent avant d'avoir pu se reproduire. Ainsi l'allèle mutant, dominant léthal, est éliminé de la population dans la même génération que son apparition. Les allèles létaux qui tuent uniquement à l'état homozygote peuvent être de deux types :

- a- soit ils n'ont pas d'effet phénotypique évident chez les hétérozygotes.
 b- soit ils induisent un phénotype distinct chez les hétérozygotes.

Chapitre VII. ANALYSE GENETIQUE CHEZ LES ORGANISMES HAPLOÏDES

VII.1. Cycle de croissance chez les Champignons – Analyse des tétrades

L'un des fondements de l'analyse génétique repose sur l'analyse des produits de la méiose. La ségrégation des gènes permet entre autres de les dénombrer, de les cartographier, parfois même de préciser leur type d'interaction. La ségrégation allélique à la méiose donne des gamètes dont le contenu génétique n'est pas directement déductible, sauf si on étudie des organismes qui ont une phase haploïde qui ne concerne pas que les gamètes, c'est le cas de champignons ou les fougères.

Le mécanisme de la méiose et ses conséquences génétiques, dont la recombinaison, ont pu être précisés par l'étude génétique de champignons particuliers appelés **ascmycètes**. Chez les ascmycètes, les spores haploïdes résultant de la méiose (méiospores ou ascospores), restent regroupées et enfermées dans un sac appelé **asque**. L'isolement des quatre spores d'un asque permet d'entreprendre l'analyse isolée des quatre produits d'une même méiose, la **tétrade**.

VII.1.1. Analyse des tétrades de la levure de boulangerie

La levure de boulangerie ou *Saccharomyces cerevisiae* est l'ascmycète le plus simple, unicellulaire et présente un cycle de vie **haplobiontique**. Sa reproduction asexuée est réalisée par bourgeonnement, une division mitotique caractérisée généralement avec une cytokinèse inégale. Le cycle sexuel implique l'union de cellules entières de types sexuels opposés (MAT α et MATa) pour former un zygote diploïde (Figure 1). La cellule diploïde donne soit une descendance diploïde par bourgeonnement (reproduction asexuée, végétative) ou une descendance haploïde par méiose (reproduction sexuée). Ces quatre noyaux haploïdes sont des ascospores qui restent enfermées dans l'asque, dont la rupture libérera les spores haploïdes qui germent en cellules nouvelles de levure (2 colonies de signe α et 2 colonies de signe a).

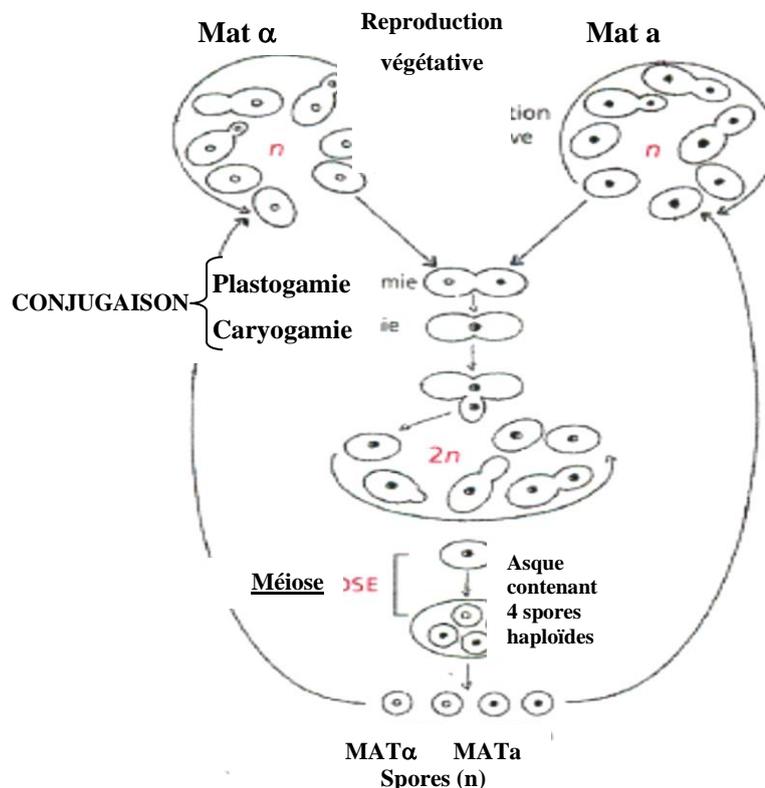


Figure 1. Cycle de vie de la levure de boulangerie, *Saccharomyces cerevisiae*.

Les généticiens ont pu développer dans les laboratoires de nombreux mutants de la levure dont les étapes du cycle de reproduction sont bien contrôlées. Ces chercheurs ont pu caractériser des mutants **conditionnels** (phénotype sauvage viable dans les conditions **permissives** et un phénotype muté létal dans les conditions **non permissives**). Les généticiens ont recherché ainsi des mutants **thermosensibles** (*ts*) : phénotype sauvage à une température permissive

(20°C) où la levure se divise normalement et phénotype muté à une température non permissive (37°C) où la levure arrête toute division.

VII.1.2. Analyse des tétrades de la moisissure du pain

Un autre Ascomycète intéressant pour les généticiens est la moisissure du pain *Neurospora crassa* (Figure 2). Les filaments fongiques (**mycélium**) sont composés de filaments entrelacés appelés hyphes. L'extrémité des **hyphes** peu se couvrir de spores asexués (**conidies**) qui donnent de nouveaux hyphes après germination. Les hyphes végétatifs sont segmentés, et chaque segment contient plusieurs noyaux haploïdes. Les hyphes forment un mycélium qui peut s'anastomoser (union de deux hyphes) avec des hyphes d'un autre mycélium pour former un mélange de noyaux et de cytoplasme appelé **hétérocaryon**. Un couple d'allèles A et a déterminé les deux types reproductifs et la reproduction sexuée ne se réalise qu'entre des cellules de types opposés. Des régions spécialisées du mycélium produisent une fructification immature femelle (**protopérithèque**) à partir de laquelle sont libérés des filaments réceptifs appelés **trichogynes**. Une conidie ou hyphe provenant d'un type sexuel opposé fusionne avec le trichogyne, transfère plusieurs noyaux et conduit après des divisions successifs à plusieurs caryogamies.

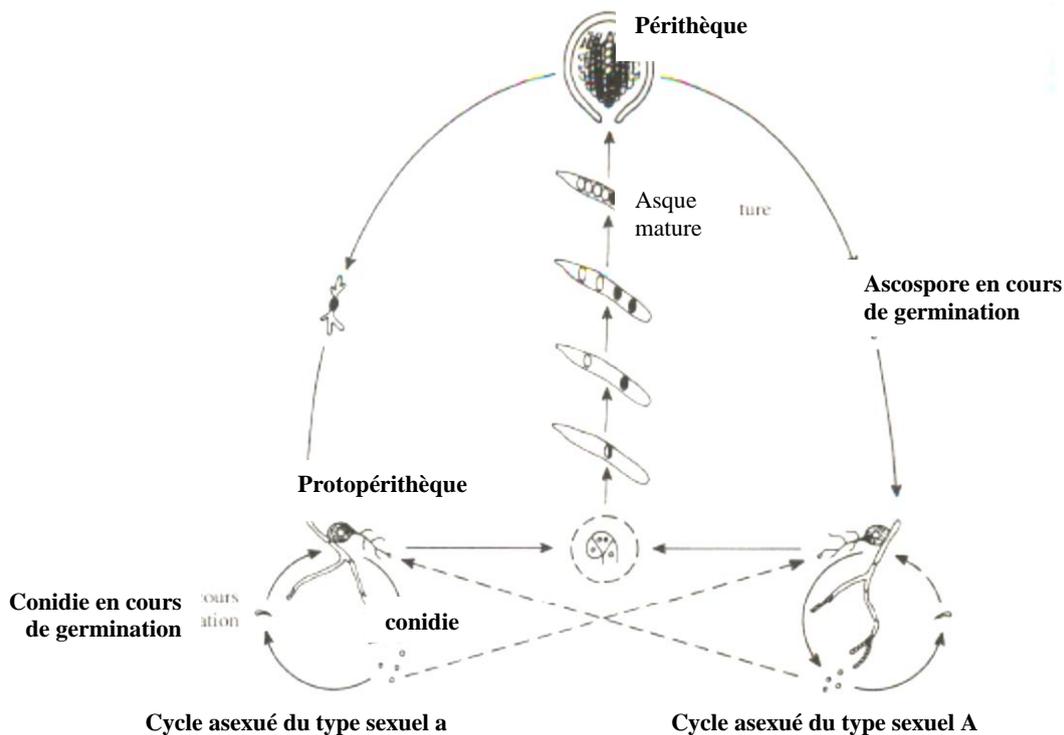


Figure 2. Cycle de vie de la moisissure du pain, *Neurospora crassa*.

Chacun des zygotes diploïdes résultant de cette fusion repose dans un sac (**asque**), qui se divise par une méiose (pour donner 4 noyaux) suivi par une mitose qui met en place quatre paires de noyaux qui mûrissent en 8 ascospores (**octade**). Une fructification mature ou périthèque peut contenir plus de cent asques, contenant chacun 8 ascospores orientées en longueur dans l'asque et préviennent les recouvrements des produits de méiose et de la mitose (chacune des 4 chromatides de la prophase I de méiose sont représentées par paire d'ascospore en tandem ordonnée dans l'asque).

Si les 4 ascospores de la levure ne sont pas ordonnées, les 8 ascospores de *Neurospora crassa* sont disposées linéairement dans l'asque avec la même séquence de celle des chromatides dans la plage métaphasique de la méiose I. Le comptage et l'investigation de tous les produits de méiose est appelé **analyse des tétrades**. En effet avec *Neurospora crassa* les centromères peuvent être cartographiés car les 8 ascospores sont réparties dans l'asque en fonction du profil de ségrégation résultant des 2 divisions méiotiques.

VII.2. Hérité monogénique et distance locus – centromère

Les centromères (régions des chromosomes qui ne présentent pas d'hétérozygotie) sont difficiles à cartographier. Cependant chez *Neurospora crassa*, chez qui les produits de la méiose sont ordonnés dans l'asque, les centromères peuvent être cartographiés. Ainsi on peut utiliser la cartographie du centromère pour estimer la distance qui sépare un locus du centromère d'un chromosome. Cette technique est basée sur le fait qu'à la suite d'un chiasma lors de la méiose l'ordonnancement imposé aux ascospores sera modifié. Supposons qu'on a un croisement entre deux

individus ayant chacun un allèle différent (A ou a) à un locus. Chaque asque de *Neurospora crassa* quand on analyse la ségrégation de ce couple d'allèles appartient à l'une des deux catégories d'asques :

a- les asques pré-disjoints de type **4 : 4** (octade comportant deux blocs adjacents de 4 ascospores, l'un de génotype A et l'autre de génotype a), résultant d'une ségrégation des allèles en **première division** de méiose (Figure 3).

b- les asques post-disjoints de type **2 : 2 : 2 : 2** ou **2 : 4 : 2**, résultant d'une ségrégation des allèles en **deuxième division** de méiose (Figure 4).

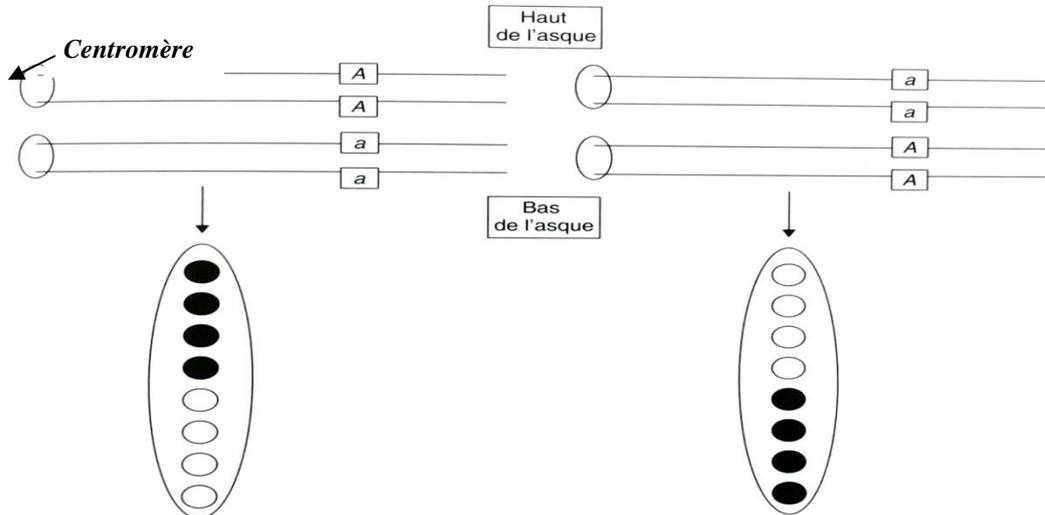


Figure 3. Les deux types d'asques issus d'une méiose sans crossing-over (ségrégation en première division) chez *Neurospora Crassa*.

Concernant les profils de ségrégation de seconde division obtenus lorsqu'il y a un crossing-over, ou aura quatre différents types d'asques (Figure 4). La fréquence des octades ayant ces profils doit être proportionnelle à la distance qui sépare le locus du centromère et elle peut servir pour estimer la distance génétique qui sépare le centromère du locus étudié. Ainsi, après un croisement entre individus A et a, on dénombre les octades de profil de ségrégation des allèles en **première division** de méiose (M_1) et celles de profil de ségrégation des allèles en **seconde division** de méiose (M_2). Ceci nous permettra de savoir le pourcentage de recombinants comme suit :

$$\frac{\text{Nombre de } M_2}{\text{Nombre total d'ascospores } (M_1 + M_2)}$$

En réalité, cette valeur trouvée représente un nombre de méiose et non le nombre de chromatides recombinantes. Comme dans les méioses, seulement 50% des chromatides sont recombinantes, on doit diviser la fréquence des M_2 par 2 pour la convertir en unités génétiques (centimorgan).

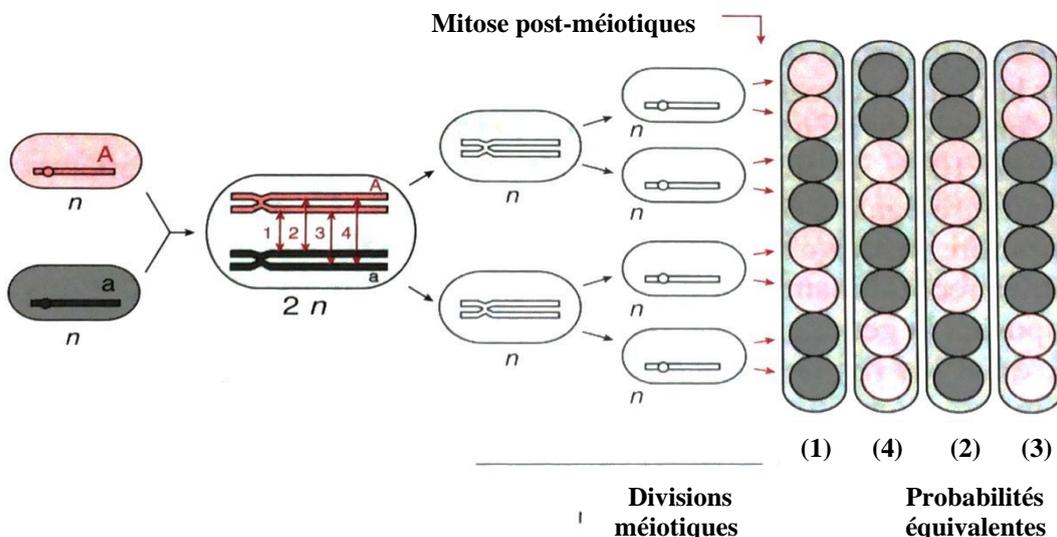


Figure 4. Les quatre différents profils de ségrégation de seconde division chez *Neurospora Crassa*.

VII.3. Transmission de deux gènes indépendants

L'analyse des tétrades, pour une méiose affectant 2 gènes représentés par 2 couples d'allèles indépendants, permet de définir toutes les configurations possibles de méioses inaccessibles dans l'étude de gamètes en désordre et de préciser certains aspects de la méiose souvent mal connus ou trop simplifiés.

Si l'on croise une souche haploïde de *N. Crassa* [A ; B] avec une autre souche [a ; b] pour donner une souche diploïde [A/a ; B/b], pour chacun des deux gènes, on va obtenir 6 configurations possibles (voir § IV.2) : 2 pré-réduites et 4 post-réduites. Sous l'hypothèse qu'une paire de chromosomes homologues ségrège indépendamment d'une autre paire d'homologues, on peut prévoir $6 \times 6 = 36$ configurations possibles non superposables (Tableau 1). On ne va pas montrer dans ce tableau les différentes configurations obtenues après la mitose supplémentaire car il n'aura pas d'effet sur chacune des configurations. Le taux ou fréquence de post-réduction pour les allèles A/a et B/b sont respectivement représentés par p et q .

L'analyse des tétrades ordonnées (en tenant compte des positions haut et bas des asques), pour deux gènes physiquement indépendants, permet d'observer ces 36 types d'asques non superposables, avec des fréquences égales aux valeurs prévues sous l'hypothèse d'une ségrégation indépendante des paires d'homologues, soit le produit des fréquences marginales fonction de p et q .

La cohérence entre les prévisions et observations constitue bien une validation expérimentale directe du fait qu'à la méiose les paires d'homologues différents ségrégent indépendamment l'une de l'autre.

Il faut noter que les 36 types d'asques peuvent être, si l'on tient compte de l'ordre des spores dans l'asque, regroupés en 3 types différents :

- a- Les ditypes parentaux (DP) : quatre spores deux à deux identiques à chacun des deux parents, [A ; B] et [a ; b].
- b- Les ditypes recombinés (DR) : quatre spores deux à deux identiques à chacun des deux types recombinés, [A ; b] et [a ; B].
- c- Les tétratypés (T) : quatre spores toutes de génotypes différents, deux de type parental, [A ; B] et [a ; b] et deux de type recombinés possibles, [A ; b] et [a ; B].

		Préréduction (1-p)		Postréduction p			
		(1-p)/2	(1-p)/2	p/4	p/4	p/4	p/4
		A	a	A	a	a	A
		A	a	a	A	A	a
		a	A	A	a	A	a
		a	A	a	A	a	A
Préréduction (1-q)/2	B	DP	DR	T	T	T	T
	B						
	b						
	b						
Préréduction (1-q)/2	b	DR	DP	T	T	T	T
	B						
	B						
	B						
Postréduction q/4	B	T	T	DP	DR	T	T
	b						
	B						
Postréduction q/4	b	T	T	DR	DP	T	T
	B						
	b						
Postréduction q/4	b	T	T	T	T	DP	DR
	B						
Postréduction q/4	B	T	T	T	T	DR	DP
	b						
	b						
	B						

Tableau 1. Configurations possibles de la méiose suite à l'analyse des tétrades pour une transmission de deux gènes indépendants. DP, ditypes parentaux ; DR, ditypes recombinés et T, tétratypés. Si on a des asques ordonnés

sans haut ni bas, les asques seront identiques 2 à 2, sauf le dernier carré en bas à droite du tableau, ce qui réduit à 20 le nombre de configurations différentes de la méiose au lieu de 36

VII.4. Ségrégation de deux gènes physiquement liés, distance génétique et carte factorielle

Si deux gènes sont physiquement liés, les configurations possibles de la méiose dépendent d'abord de la disposition aléatoire des deux centromères de la paire d'homologues, ensuite de la survenue ou non d'un ou plusieurs crossing-over et enfin, dans le cas de plusieurs chiasmata, des chromatides impliqués.

Dans l'analyse de tétrades pour les deux gènes physiquement liés, on doit tenir compte des événements survenant entre les locus des deux gènes et on va négliger la position du centromère qui ne sera figuré en pointillés que pour rappeler qu'il unit en un site les deux chromatides sœurs (Figure 5). Dans ces conditions, si l'on croise une souche haploïde de *N. Crassa* [D ; E] avec une autre souche [d ; e], avec les gènes D et E physiquement liés, on obtiendra une cellule diploïde [DE // de] dont la méiose présente plusieurs configurations possibles en fonction de la fréquence des crossing-over (Figure 5) :

- a- Si aucun crossing-over n'a eu lieu entre les locus des deux gènes (Figure 5-a), la méiose donne exclusivement un ditype parental (DP) avec deux gamètes de type parental [D ; E] et deux gamètes de type parental [d ; e].
- b- Si un crossing-over a eu lieu entre les locus des 2 gènes (Figure 5-b), la méiose donne exclusivement un tétratype (T) avec un gamète de type parental [D ; E], un gamète de type parental [d ; e], un gamète de type recombiné [D ; e] et un gamète de type recombiné [d ; E], quelle que soit la paire de chromatide impliquée par crossing-over.
- c- S'il y a deux crossing-over entre les locus des deux gènes (Figure 5-c), si le premier chiasma survient entre les chromatides 2-3, le type d'asque résultant de la méiose dépendra des chromatides impliquées dans le second crossing-over. On obtiendra ainsi :
 - Un DP si le 2^{ème} chiasma touche les mêmes chromatides 2-3 ;
 - Un DR si le 2^{ème} chiasma touche les deux autres chromatides 1-4 ;
 - Une T si le 2^{ème} chiasma touche un même chromatide, 2 ou 3, et un 3^{ème} chromatide non sœur, autrement dit si le 2^{ème} chiasma est 2-4 ou 3-1

Ainsi dans le cas d'un double crossing-over entre deux locus, on obtient statistiquement $\frac{1}{4}$ de DP + $\frac{1}{2}$ de T et $\frac{1}{4}$ de DR.

- d- S'il y a plus de deux chiasmata entre les locus des deux gènes, on peut montrer que les DP et les DR ont également la même fréquence.

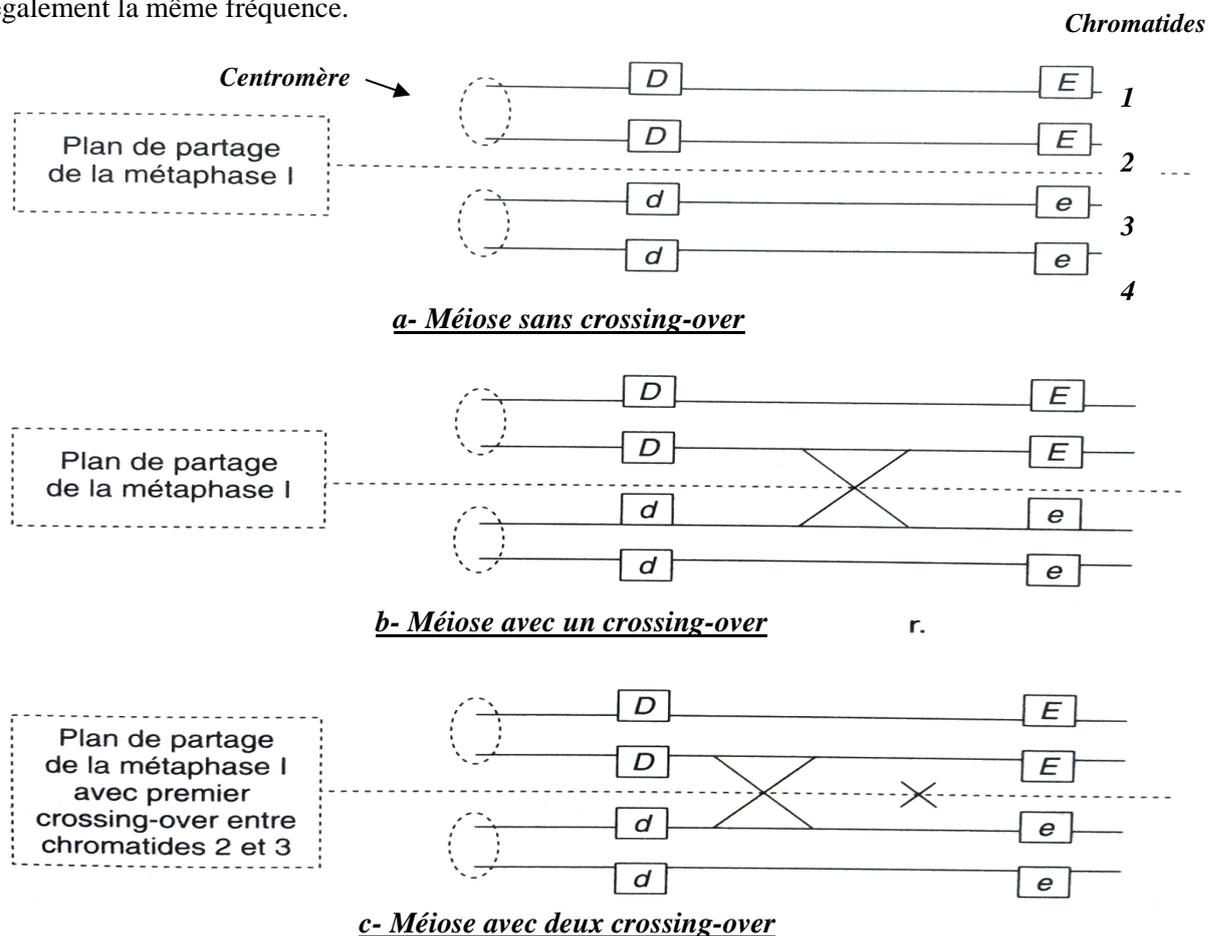


Figure 5. Les possibilités des produits de la méiose dans la distribution de deux gènes physiquement liés chez *Neurospora Crassa*.

En résumé, la fréquence des crossing-over entre deux locus étant une fonction de leur **distance génétique**, on peut alors estimer cette distance par l'estimation de la fréquence des chiasmas (Nombre de produits recombinés/Nombre total des produits, voir § III.3). L'analyse de plusieurs marqueurs dans un organisme donné permettra de dresser une carte factorielle d'un être vivant.

VII.5. Les groupes de complémentation et dénombrement des gènes

La génétique des champignons est essentiellement basée sur la possibilité de croiser deux souches haploïdes portant des mutations différentes pour former une souche diploïde. Garrod était le premier chercheur à exprimer la possible relation entre un gène et une enzyme en utilisant un **test de complémentation** (test de cis-transport) qui est un essai employé pour déterminer si deux mutations différentes se situent dans le même ou différents gènes (deux mutations dans un même gène ne peuvent pas se compléter). Ce chercheur travaille sur une anomalie métabolique humaine, l'**alcaptonurie**, qui affecte le métabolisme de la tyrosine et de la phénylalanine. Garrod propose en 1909 de justifier des déficiences enzymatiques héréditaires chez l'homme (albinisme, cystinurie, pentosurie...) par des anomalies génétiques dues à l'inactivation de gènes codant pour certaines enzymes. Un gène erroné produit une enzyme non fonctionnelle, et sans l'enzyme, le substrat de l'enzyme s'accumule et produit la maladie.

Beadle et Tatum isolent des souches mutantes de *N. crassa* qui ne se développent pas sans apport supplémentaire de tryptophane lorsque les cultive sur un milieu minimum (un acide aminé qu'elles synthétisent habituellement) obtenues par irradiation aux rayons X (mutagenèse dirigée). Plusieurs souches sont obtenues pour lesquelles l'apport de **tryptophane** peut être remplacé par celui d'**acide anthranilique** ou encore d'**indole**. Ces souches sont ensuite croisées entre elles (tests de complémentation).

- Si les mutations sont situées dans la même unité fonctionnelle (le même gène donc), la fécondation de deux souches donne des mycéliums qui possèdent la même déficience métabolique que les souches croisées.
- Si les mutations sont situés dans des gènes différents (mais qui conduisent au même phénotype), les déficiences peuvent se compenser (les étapes métaboliques se complètent ou se complémentent) et il ya obtention d'un mycélium qui ne présente plus de déficience apparente (Figure 6).

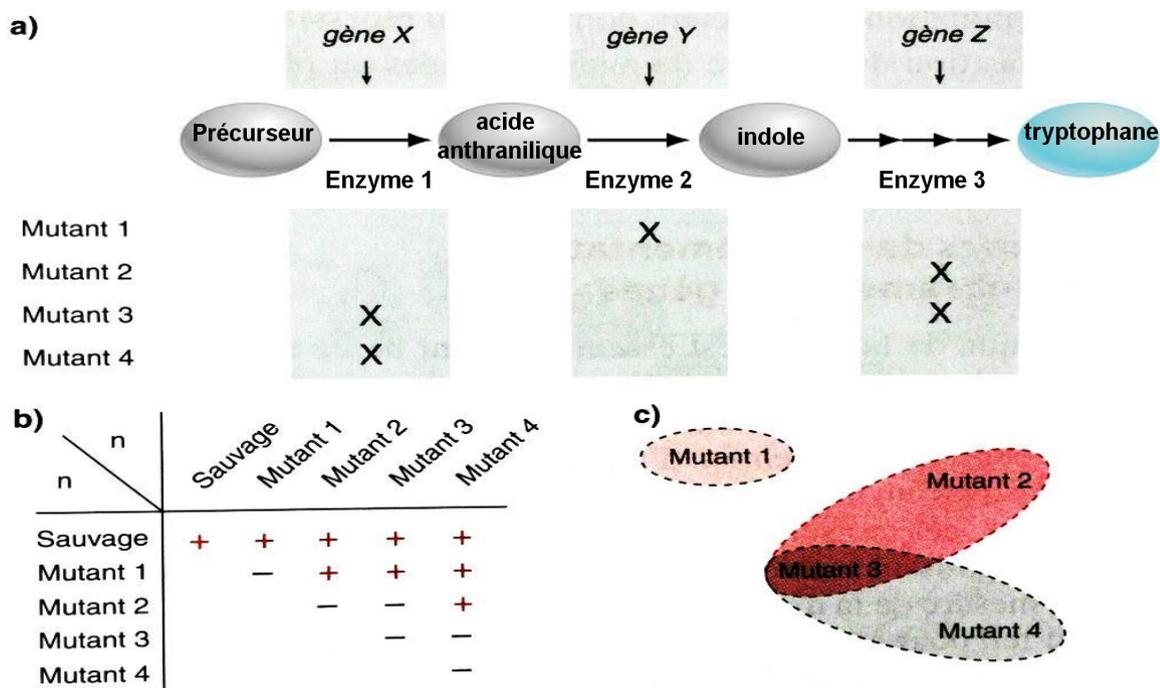


Figure 6. Détermination du nombre de groupes de complémentation et par analogie du nombre minimal de gènes impliqués dans une voie métabolique (voie de biosynthèse tryptophane, a). Les 4 mutants auxotrophes pour Trp sont croisés entre eux et avec un haploïde sauvage prototrophe pour Trp. Les diploïdes obtenus sont mis en culture et les résultats sont reportés dans le tableau (b). On observe 3 groupes de complémentations (c) et donc trois gènes différents impliqués dans cette voie métabolique. (+) et (-) complémentations et non complémentations respectivement.

Les quatre groupes de mutants (1, 2, 3 et 4) sont des souches qui sont auxotrophes au tryptophane mais chacune

présente une mutation dans un gène spécifique au moins puisque toutes complémentent. La chaîne métabolique la plus simple à proposer pour interpréter ces résultats (figure 6-a). En conclusion, chaque gène (unité mutable) étant associé à la fonction enzymatique (unité fonctionnelle) permettant le passage d'une substance à l'autre.

VII.6. Carte physique d'un génome

De la même façon que l'on réalise une carte de restriction d'un plasmide avec différentes combinaisons d'enzymes de restriction et de sondes, il est possible d'établir une **carte physique** d'un génome complexe. La distance entre deux locus peut être déterminée directement par analyse de fragments de restriction : la taille du plus petit fragment contenant les deux locus permet d'estimer la distance qui les sépare. En règle générale, les locus utilisés correspondent à des marqueurs génétiques espacés de plusieurs cM, soit plusieurs mégabases. Il est donc nécessaire de réaliser des digestions partielles et/ou d'utiliser des enzymes de restriction à sites rares (*Not I*, *Mlu I*) afin de libérer des fragments d'ADN suffisamment grands. Compte tenu de leur taille importante, les fragments générés peuvent être séparés par électrophorèse en champ pulsé. Les procédés de cartographie du génome ont beaucoup évolués avec l'apparition de la biologie moléculaire (Figure 7). Cependant les méthodes classiques d'étude de ségrégation des caractères à partir d'un sujet diploïde hétérozygote gardent toute leur actualité, leur mise en œuvre étant beaucoup plus simple que l'utilisation de la biotechnologie moléculaire.

Pour placer les différents gènes sur les chromosomes humains on a eu besoin de repères. Un bon type de repère doit être régulièrement réparti sur tout le champ à couvrir et chaque repère doit être unique pour éviter toute confusion. Deux types de repères ont présenté ces propriétés et sont utilisés pour baliser le génome humain :

- les RFLP (Restriction Fragment Length Polymorphism).
- les microsatellites.

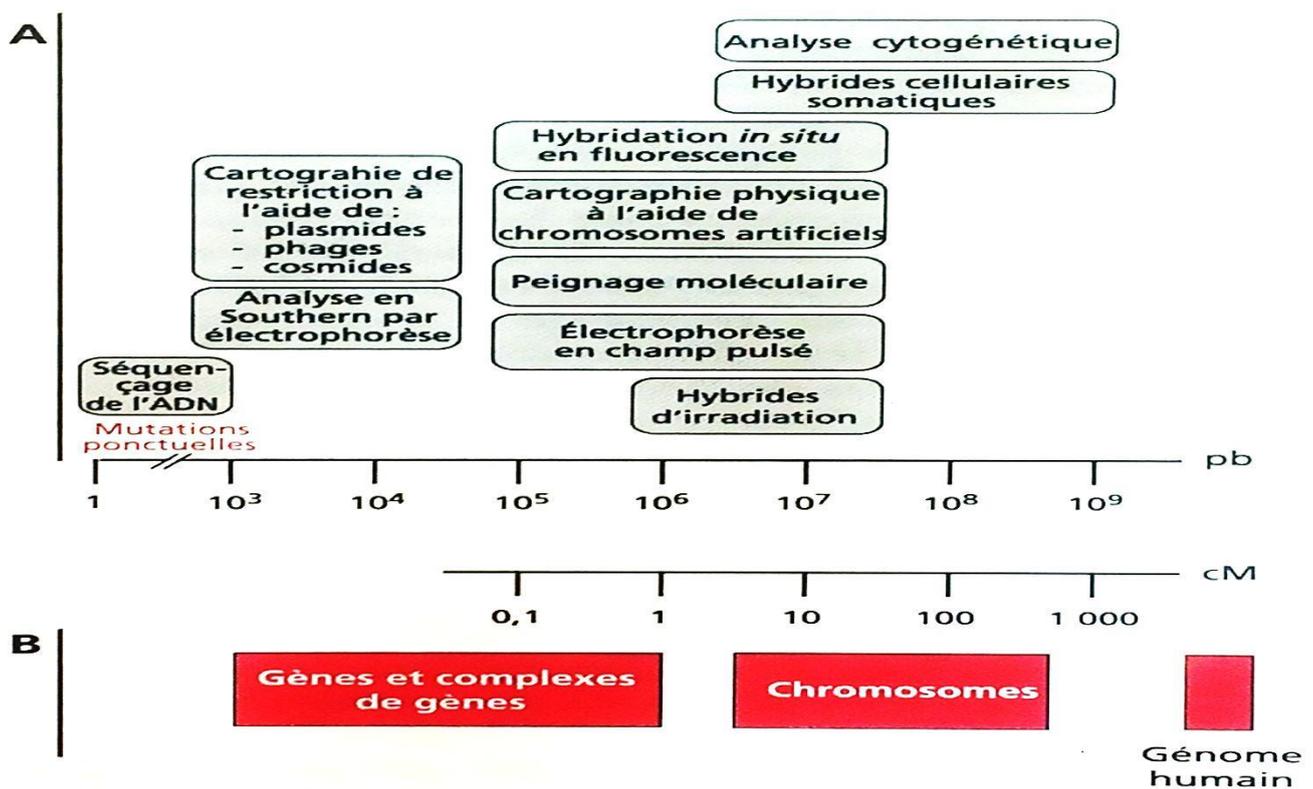


Figure 7. Ordres de précision obtenus par les différentes techniques de cartographie du génome.

Références bibliographiques

- Biochimie de Lehninger A.L , Eds. Flammarion, 2004.
- Biologie Moléculaire. Cours, exercices, annales et QCM corrigés. S. Beaumont. Eds Dunod. 2007.
- Mini manuel de Biologie Moléculaire. A. Maftah, J-M. Petit et R. Julien. Eds Dunod. 2007.

- Biochimie et Biologie Moléculaire. W. Müller-Esterl. Eds Dunod. 2007.
- Molecular Biology : Principles and Practice. Michael M. Cox, Jennifer A. Doudna & Michael O'Donnell, Eds W.H. Freeman & Company, 2015.
- Mini manuel de Génétique. J-M. Petit et R. Julien. Eds Dunod. 2007.

Pr Khalil Hammani, Février 2025

Khalil.hammani@usmba.ac.ma

Sommaire

Chapitre I. INTRODUCTION AU DOGME CENTRAL, STRUCTURE ET PROPRIETES DES ACIDES NUCLEIQUES. 2

I.1. Introduction	2
I.2. Le Dogme Central de la biologie moléculaire.	2
I.3. De l'information génétique au programme génétique.	3
I.3.1. Nature du matériel génétique.	3
I.3.2. Travaux de Griffith.	3
I.3.3. Travaux d'Avery, McLeod et McCarty (1943).	4
I.3.4. Travaux de Hershey et Chase (1952).	4
I.4. Structure et propriétés des acides nucléiques.	5
I.4.1. Travaux de Watson et Crick	5
I.4.2. Structure des acides nucléiques	5
I.4.3. Formes de la double hélice d'ADN	7
I.4.4. Formes et types d'ARN	7
I.5. Propriétés topologiques fondamentales de l'ADN	8
I.5.1. Dénaturation de l'ADN et température de fusion (T _m)	8
I.5.2. Dosage et étude de l'ADN	9
I.6. ADN et chromosomes	10

Chapitre II. OUTILS ET TECHNIQUES DE LA BIOLOGIE MOLECULAIRE 12

II.1. La création de molécules d'ADN recombinant	12
II.1.1. Clivage de l'ADN par les enzymes de restriction	12
II.1.2. Ligature de l'ADN	13
II.2. Les vecteurs de clonages	14
II.2.1. Les plasmides	15
II.2.2. Les vecteurs viraux	16
II.2.3. Les cosmides	16
II.2.4. Les chromosomes artificiels	17
II.3. Les banques d'ADN	17
II.3.1. Les banques d'ADN génomiques	17
II.3.2. Les banques d'ADN complémentaires.	18
II.4. Les techniques d'analyse d'ADN	19
II.4.1. Séparation des fragments d'ADN	19
II.4.2. Marquage des ADN-Réactions de Phosphorylation-déphosphorylation	19
II.4.3. Le séquençage de l'ADN	20
II.4.4. La réaction de Polymérisation en Chaîne (PCR)	22
II.4.5. Les techniques d'hybridation des acides nucléiques	23
II.4.6. Les techniques de localisation des sites de liaison à l'ADN	24
II.4.7. Criblage à l'aide d'une sonde d'acide nucléique	25
II.4.8. Puces à ADN et réseaux génétiques.	26

II.4.9. Les applications de la technologie de l'ADN recombinant.	27
II.4.10. Le déchiffrement du génome humain	30
II.4.11. L'analyse des gènes <i>in silico</i>	31

Chapitre III. LA REPLICATION 33

III.1. La réplication de l'ADN	33
III.1.1. Réplication sémi-conservative.	33
III.1.2. Mécanisme de la réplication	34
III.1.3. Enzymes et protéines de la réplication	35
III.2. Erreurs de réplication de l'ADN et leur réparation.	38
III.2.1. Origines des mutations.	38
III.2.2. Les mutations liées à des erreurs de duplication de l'ADN	38
III.2.3. Les mutations provoquées par des agents mutagènes.	40
IV.2.4. Les mutations provoquées par recombinaison ou transposition	41

Chapitre IV. EXPRESSION DE L'INFORMATION GENETIQUE : TRANSCRIPTION ET TRADUCTION 41

IV.1. Introduction	41
IV.2. Transcription	43
IV.2.1. Promoteur et Brins sens ou antisens	41
IV.2.2. Régulation de l'expression des gènes chez les procaryotes.	42
IV.2.3. Régulation de l'expression des gènes chez les eucaryotes: Cis/Trans régulateurs et DNA binding proteins	45
IV.2.4. Mécanisme de la transcription	49
IV.2.5. Régulations post-transcriptionnelles.	48
IV.3. La traduction	51
IV.3.1. Langage et outils de la traduction	52
IV.3.2. Mécanisme de la traduction	54

Chapitre V. DIVISION CELLULAIRE ET BRASSAGES GENETIQUES AU COURS DE LA REPRODUCTION SEXUEE 59

V.1. divisions cellulaires.	59
V.1.1. Mitose.	59
V.1.2. Méiose.	61
V.2. Brassages génétiques au cours de la reproduction sexuée.	62
V.2.1. Recombinaison homologue.	63
V.2.2. Recombinaison interchromosomique ou ségrégation.	64
V.2.3. Les mutations chromosomiques.	64

Chapitre VI. ANALYSE GENETIQUES CHEZ LES ORGANISMES DIPLOIDES. 68

IV.1. Mendel et la notion de gènes et transmission de l'information génétique.	68
IV.1.1. Mendel et la génétique.	68

VI.1.2. Analyse mendélienne de la transmission d'un caractère (monohybridisme, modèle 3:1)....	68
VI.1.3. Analyse mendélienne de la transmission de deux caractères autosomiques et indépendants (dihybridisme, modèle 9:3:1:1).....	70
VI.1.4. Généralisation pour un diagramme d'un triple hybride ou plus.....	72
VI.2. Extension des lois de Mendel.....	72
VI.2.1. De l'analyse mendélienne aux gènes.....	72
VI.2.2. Codominance et pluriallélisme.....	73
VI.2.3. L'hérédité liée au sexe.....	74
VI.2.4. Distinction entre phénotype et génotype.....	77
VI.3. Liaison génétique et cartographie.....	78
VI.3.1. Distance génétique et carte génétique.....	78
VI.3.2. Le test deux points.....	78
VI.3.3. Le test trois points.....	78
VI.3.4. Ordre des gènes.....	79
VI.3.5. Liaison génétique et sélection.....	79
VI.3.6. L'interférence.....	80

VI.4. Les interactions génétiques.....	80
VI.4.1. Pléiotropie.....	80
VI.4.2. Epistasie.....	81
VI.4.3. Létalité.....	82

Chapitre VII. ANALYSE GENETIQUES CHEZ LES ORGANISMES HAPLOIDES.....83

VII.1. Cycle de croissance chez les Champignons – Analyse des tétrades.....	83
VII.1.1. Analyse des tétrades de la levure de boulangerie.....	83
VII.1.2. Analyse des tétrades de la moisissure du pain.....	84
VII.2. Hérité monogénique et distance locus – centromère	84
VII.3. Transmission de deux gènes indépendants.....	86
VII.4. Ségrégation de deux gènes physiquement liés, distance génétique et carte Factorielle.....	87
VII.5. Les groupes de complémentation et dénombrement des gènes.....	88
VII.6. Carte physique d'un génome.....	89
Références bibliographiques.....	89