

Cours 04 : synthèse et séquençage de l'ADN

Techniques de clonage, de séquençage

I. Techniques de clonage

Introduction :

Le terme clone est utilisé pour la première fois en 1903 par le botaniste H.J. Webber en désignant des plantes reproduites par multiplication asexuée. Ce mot sera ensuite réutilisé par J.B.S. Haldane.

Le clonage désigne principalement le processus de multiplication d'un organisme, d'une cellule souche ou d'un gène, en grand nombre d'exemplaires identiques. Depuis plus de vingt ans, le clonage, ou la reproduction exacte de gènes particuliers et de types individuels de cellules, est une technique employée en biotechnologie afin de produire des médicaments et des vaccins pour traiter plusieurs maladies. On nomme ces deux types de clonage :

- **Le clonage cellulaire** : On fait plusieurs exemplaires de la même cellule pour pouvoir faire des recherches médicales.
- **Le clonage moléculaire** le quel on développera particulièrement dans cette partie.

1. Clonage moléculaire :

Le clonage de l'ADN est une technique puissante qui permet de séparer des séquences spécifiques d'ADN au sein d'autres séquences puis de les copier de sorte qu'on obtient des quantités importantes permettant une analyse détaillée ou des manipulations. Une importante utilisation du clonage de l'ADN est d'isoler de nouveaux gènes en leur permettant d'être explorés et caractérisés.

Toutes les expériences de clonage de l'ADN reposent sur la construction de molécules d'ADN recombinant. Ceci suppose de joindre ensemble différentes molécules d'ADN. La molécule d'ADN à cloner est insérée dans une autre molécule d'ADN généralement circulaire, appelée un vecteur. Le vecteur recombinant est introduit dans une cellule hôte, généralement la bactérie E.coli, où il produit de multiples copies de lui-même. Lorsque la cellule hôte se divise, des copies du vecteur sont ainsi produites et elles peuvent être purifiées à partir des cultures de cellules hôtes et utilisées pour l'analyse de l'insert d'ADN étranger.

2. Outils de clonage :

2.1. Les enzymes de restriction :

Joindre les molécules d'ADN ensemble pour le clonage dépend de l'utilisation d'enzymes de bactéries appelées endonucléases de restriction. Ces enzymes coupent les molécules d'ADN au niveau de séquences spécifiques, généralement 4 à 8 bases. Les séquences reconnues sont des palindromes. Ceci signifie que la séquence est la même lorsqu'on la lit dans le sens 5' vers 3' sur les deux brins. Chaque enzyme a une séquence cible spécifique. Par exemple, l'enzyme EcoRI, qui est issue de la bactérie E.coli coupera n'importe quelle molécule d'ADN pourtant la séquence GAATTC. Les autres endonucléases de restriction reconnaissent des séquences différentes. La coupure réalisée par l'enzyme de restriction n'est généralement pas franche, c'est-à-dire que les deux brins de la double hélice sont coupés à quelques bases de distance.

Ceci crée des portions simples brins qui dépassent, et qu'on appelle extrémités cohésives ou bouts collants. Certaines enzymes de restriction coupent l'ADN en laissant dépasser une extrémité 5' et d'autres laissant dépasser une extrémité 3'.

Il existe des enzymes qui coupent de façon franche les deux brins au même endroit. Noter que même lorsque les molécules sont jointes par des extrémités cohésives, elles ne sont pas liées de façon covalente. Une enzyme appelée ADN ligase, est utilisée pour catalyser la formation d'une liaison phosphodiester entre deux molécules d'ADN, les recombinant ainsi de façon permanente. Il existe plusieurs systèmes pour cloner des molécules d'ADN, en fonction du type de vecteur utilisé. Chacun présente ses propres caractéristiques et utilisations.

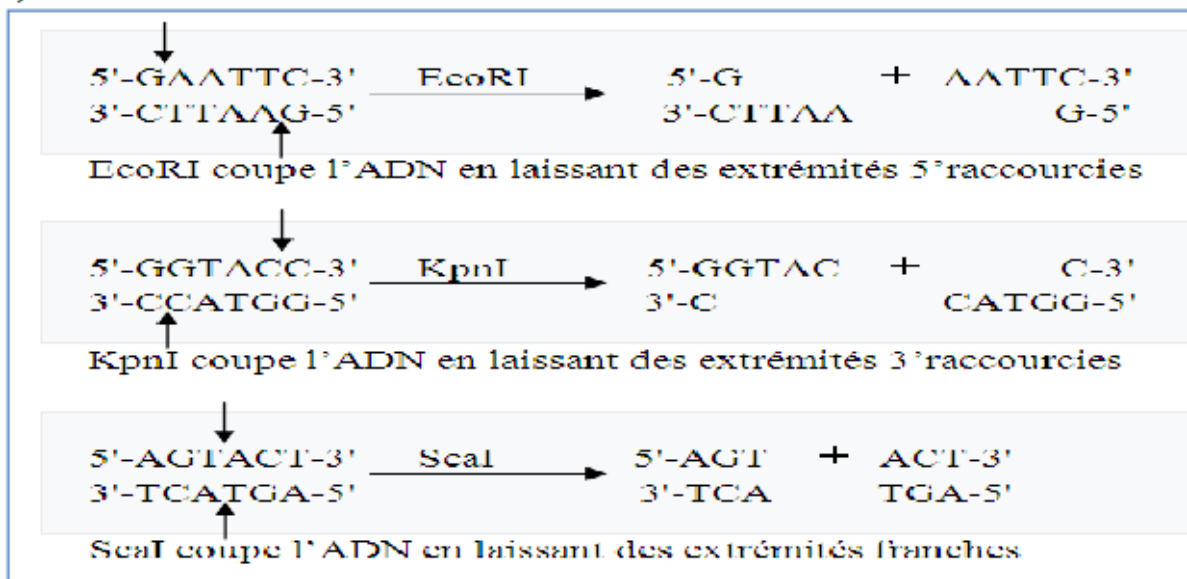


Figure : Action d'enzymes de restriction sur l'ADN.

La nomenclature des enzymes de restriction est précise. Le nom comporte 3 ou 4 lettres correspondant entre autres à la bactérie à partir de laquelle on a extrait cette enzyme :

- La première lettre est en majuscule cela correspond à la première lettre du genre de la Bactérie.
- La seconde et troisième lettre sont en minuscule et correspondent aux deux premières lettres de l'espèce bactérienne.
- La quatrième lettre correspond à la souche bactérienne, elle est écrite en majuscule mais n'est pas présente dans toutes les enzymes de restriction.
- Enfin un chiffre romain donne le numéro d'ordre de caractérisation de l'enzyme.

EcoRI : Enzyme de restriction d'*Escherichia* (genre) *coli* (espèce) souche RY317 la première à être caractérisée (I).

2.2 Procédure de clonage selon les vecteurs utilisés :

- **Plasmide :**

Les plasmides forment le type le plus commun de vecteur de clonage. La procédure de clonage peut être divisée en plusieurs étapes comme suit (Figure) :

- Le plasmide est digéré par une enzyme de restriction qui le coupe en un site unique et transforme la molécule circulaire qu'il était en une molécule linéaire avec des extrémités cohésives.
- L'ADN étranger à cloner est lui aussi digéré par la même enzyme de restriction pour produire les mêmes bouts collants,
- Le plasmide et l'ADN étranger sont mixés, des molécules plasmides sont jointes à des molécules d'ADN étranger via leurs extrémités cohésives communes et on obtient des plasmides recombinants circulaires,
- L'ADN ligase est utilisé pour joindre de façon covalente les deux,
- Le plasmide recombinant est introduit dans la bactérie hôte (généralement E. coli) par un processus appelé transformation,
- Les bactéries transformées sont ensuite étalées sur des plaques d'agar et les colonies des bactéries composées de cellules ayant correctement intégré le plasmide recombinant croissent,
- Des colonies individuelles sont séparées et cultivées en milieu liquide. De grandes quantités de plasmides peuvent alors être purifiées à partir des cultures et on peut récupérer l'ADN cloné pour l'analyse.

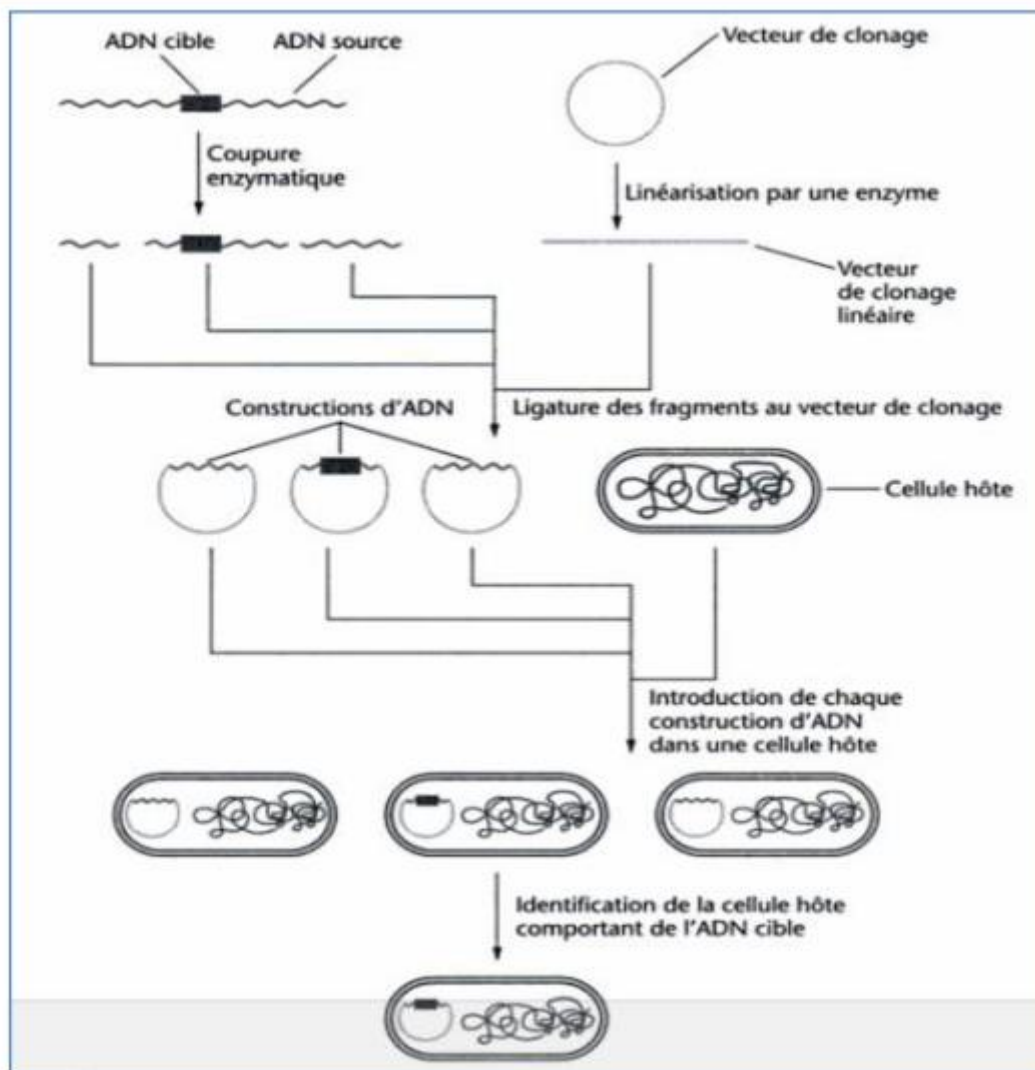


Figure : Représentation schématique du clonage d'ADN via des plasmides

- **Le phage lambda (λ) :**

Le phage λ a été adapté à une utilisation comme vecteur de clonage. La portion centrale de l'ADN de λ , qui n'est pas essentielle à l'infection, est détruite, laissant des fragments 5' et 3' appelés bras (Figure 19). La région ayant subi la délétion peut être remplacée par l'ADN étranger pour produire un phage à ADN recombinant (Figure). Celui-ci est inséré dans les capsides phagiques in vitro par un processus appelé empaquetage qui suppose de mélanger l'ADN du phage recombinant avec un extrait d'empaquetage contenant les protéines de la capsid du phage et des enzymes nécessaires à la fabrication. Des particules de phages recombinants sont produites et elles infectent E.coli avec une grande efficacité. Les cellules infectées sont étalées sur une plaque d'agar et produisent un feuillet continu de bactéries appelé une couche qui contient de petites zones de la taille d'une épingle à cheveux. Ces zones correspondent aux plages de lyse produites par l'infection phagique. Des plages individuelles peuvent être isolées et utilisées pour générer des quantités importantes d'ADN cloné par infection de cultures fraîches d'E.coli.

Le principal avantage de λ en tant que vecteur de clonage est que la taille des fragments qui peuvent y être insérés est bien supérieure à celle des inserts plasmidiques. Le phage λ peut en effet héberger des fragments long jusqu'à 25Kpb, à comparer aux 10Kpb maximales des inserts dans les vecteurs plasmidiques. Cette possibilité de cloner de plus larges fragments a conduit au développement de la principale utilisation de λ dans la construction de banques d'ADN.

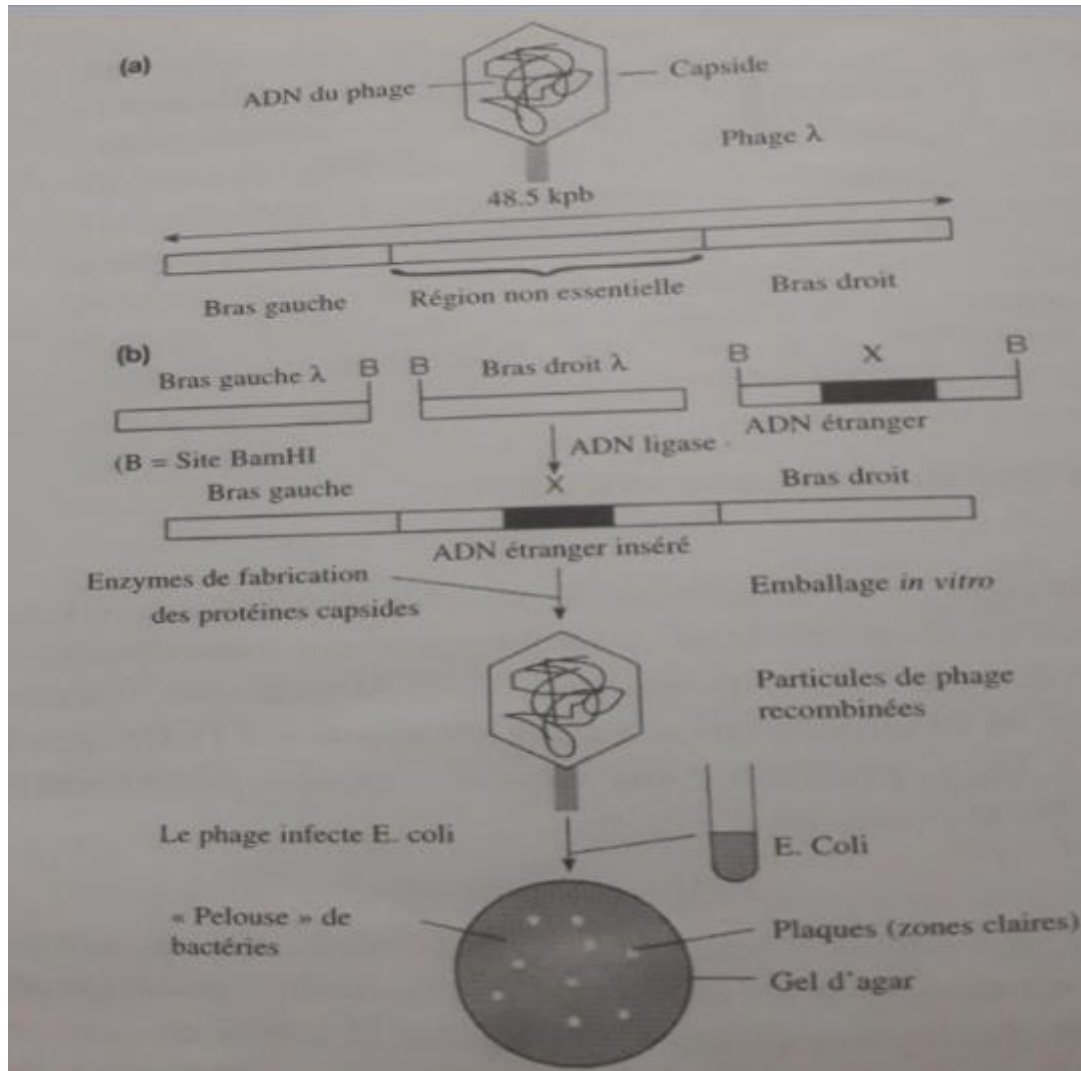


Figure : (a) Le phage λ (b) Utilisation du phage λ comme vecteur de clonage

- **Cosmides :**

Ce type de vecteur combine des caractéristiques rencontrées chez les plasmides et le phage λ .

Les cosmides ont tous les traits normaux des plasmides, y compris un MCS et des gènes conférant une résistance à des antibiotiques, mais ils incluent aussi des séquences issues de λ appelées séquences cos. Elles sont situées aux deux extrémités de la molécule d'ADN de λ et sont responsables de son insertion dans la capsid du phage. Cloner avec des cosmides combine des caractéristiques associés à l'utilisation de λ et de plasmides comme vecteur de clonage. L'ADN des cosmides est clivé par une enzyme de restriction et lié à l'ADN étranger. Le cosmide recombinant est ensuite empaqueté dans les capsides de λ et utilisé pour infecter E.coli. Les cosmides ne contiennent pas tous les gènes de λ et ne forment donc pas de plages après infection. Au lieu de cela, les cellules infectées sont cultivées sur de l'agar contenant des

antibiotiques et les colonies résistantes contenant des cosmides recombinants sont isolées et peuvent être amplifiées de la même façon que des plasmides. Les cosmides ont l'avantage d'être capable d'héberger de très gros inserts. Sachant que les cosmides sont petits, typiquement 8Kpb et que la capsid de λ peut accueillir 52Kpb, les inserts cosmidiqes peuvent mesurer jusqu'à 44Kpb.

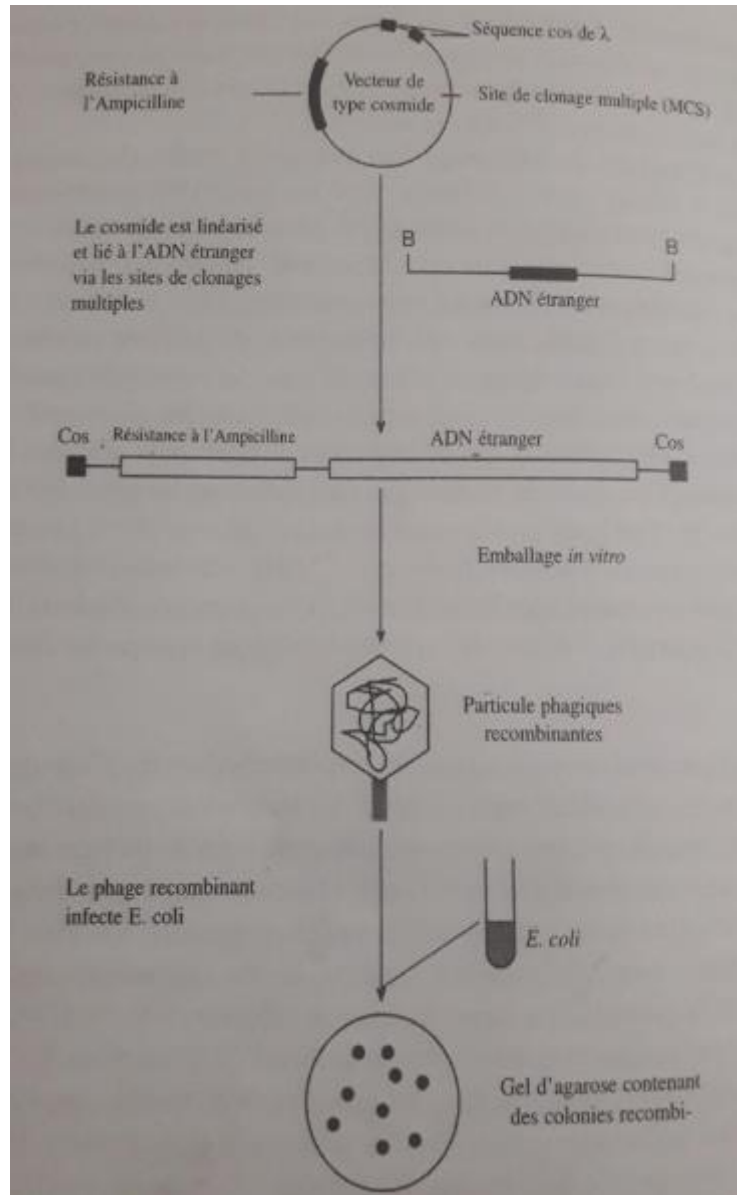


Figure : Cloner avec des vecteurs cosmidiqes

3. Application du clonage de l'ADN :

Le clonage de l'ADN est une technique puissante qui a apporté d'importantes contributions dans de nombreux domaines de la recherche biologique, principalement :

- L'identification de gènes impliqués dans des maladies ;
- La cartographie des génomes ;
- La production de protéines recombinantes ;

- Les organismes génétiquement modifiés.

II. Le séquençage de l'ADN

Introduction :

Le séquençage de l'ADN, consiste à déterminer l'ordre d'enchaînement des nucléotides d'un fragment d'ADN donné.

Le séquençage de l'ADN est une technique qui a révolutionné la biologie moléculaire seulement une vingtaine d'années après l'établissement de la structure de l'ADN par Watson et Crick. Les premières techniques de séquençage ont été développées en parallèle au milieu des années 1970. Les méthodes de Gilbert (Etats-Unis) et Sanger (Grande-Bretagne) ont toutes deux été récompensées d'un prix Nobel de chimie en 1980.

Le séquençage est devenu un outil essentiel en biologie moléculaire tant en médecine que dans de nombreuses autres disciplines des sciences de la vie. Le séquençage a été décrit il y a environ 30 ans par la méthode de Sanger et la méthode de Maxam et Gilbert, depuis cette période cette technique n'a cessé d'évoluer. La méthode de détermination des chaînes par un didésoxy c'est-à-dire la méthode de Sanger a aujourd'hui surpassé la méthode chimique car elle est à la fois plus efficace et plus facile à mettre en œuvre. C'est devenu une technique courante dans les laboratoires de biologie moléculaire. Les connaissances acquises grâce à cette méthode et la possibilité de séquencer des génomes de grande taille, tel que le génome humain, ont amené les chercheurs à développer des techniques de séquençage de plus en plus sophistiquées.

1. Séquençage 1ère génération : méthode de Sanger

Méthode mise au point par Frederick SANGER dans les années 1970, grâce à cette technique, son équipe a identifié la première séquence complète d'un virus, celui du bactériophage phiX174.

L'ADN à séquencer est appelé l'ADN matrice. Il doit être obtenu sous forme purifiée et être homogène ; c'est-à-dire qu'il ne doit contenir que des molécules d'ADN ayant la même séquence. On utilise couramment comme matrice de l'ADN cloné ou produit par PCR.

L'ADN utilisé dans cette méthode doit être simple brin afin d'être copié par l'ADN polymérase. Il est produit par dénaturation de la matrice d'ADN double brin via un traitement thermique ou alcalin. De nombreuses ADN polymérases sont utilisées pour la réaction de séquençage. Les principales sont le fragment de Klenow (une partie de l'ADN polymérase I d'E.coli), une ADN polymérase modifiée tirée du phage T appelée la Séquenase et la Taq ADN polymérase, qui est aussi utilisée en PCR.

Toutes ces ADN polymérase ont besoin d'une région d'ADN double brin pour pouvoir initier la synthèse d'ADN à partir d'une matrice simple brin. Celle-ci est fournie par addition, dans le milieu réactionnel d'amorces qui s'apparient à l'ADN matrice et déterminent ainsi le point où commencera le séquençage.



Figure : Point de départ pour le séquençage.

On prépare quatre milieux réactionnels différents, tous contiennent :

L'ADN matrice sous forme simple brin, de l'ADN polymérase, des amorces, les quatre désoxynucléotide triphosphate (dNTP) nécessaires à la construction du nouvel ADN. Chacun contient un nucléotide modifié différent, appelé un didésoxynucléotide (ddNTP). Il y a quatre ddNTP qui correspondent aux quatre bases de l'ADN. Ils diffèrent des dNTP par l'absence de groupe hydroxyle sur le carbone 3' du sucre. Ils sont incorporés à l'ADN en croissance par la polymérase, mais cette absence de l'hydroxyle en 3' empêche de former un lien phosphodiester avec le nucléotide qui devrait être ajouté ensuite. Ceci interdit de poursuivre l'élongation du polynucléotide d'ADN en cours de synthèse. Les ddNTP jouent ainsi le rôle d'inhibiteurs de la synthèse de l'ADN.

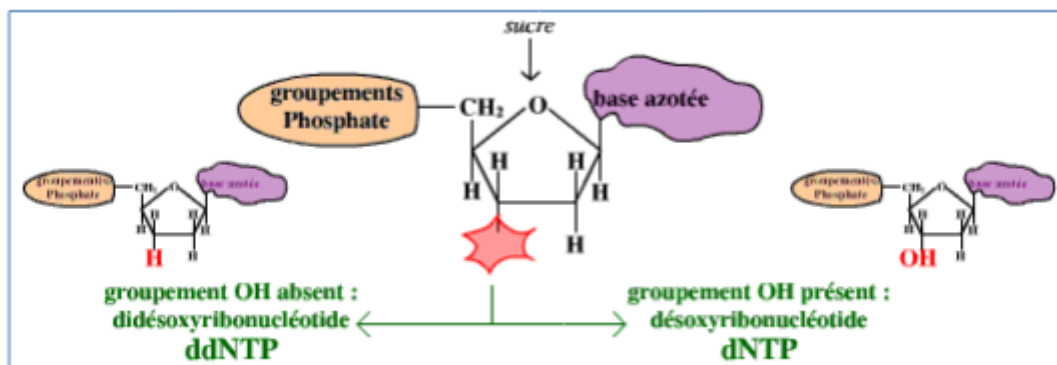


Figure : Deux types de nucléotides triphosphates.

Chacun des quatre milieux réactionnels préparés pour le séquençage contient un ddNTP différent. Suivant le ddNTP qui est présent, la synthèse se termine à l'endroit où ce nucléotide modifié est incorporé. A tout moment, durant la synthèse d'ADN, la polymérase peut incorporer à la chaîne en élongation soit un dNTP soit un ddNTP. Dans le premier cas, la synthèse continue tandis que dans le second, la synthèse s'achève en cette position. L'effet global est que dans chacune des quatre préparations, est produite une série de molécules d'ADN de différentes longueurs qui se terminent en une position où dans la matrice est présente la base complémentaire de celle dont la forme de didésoxynucléotide est présente dans le milieu réactionnel considéré.

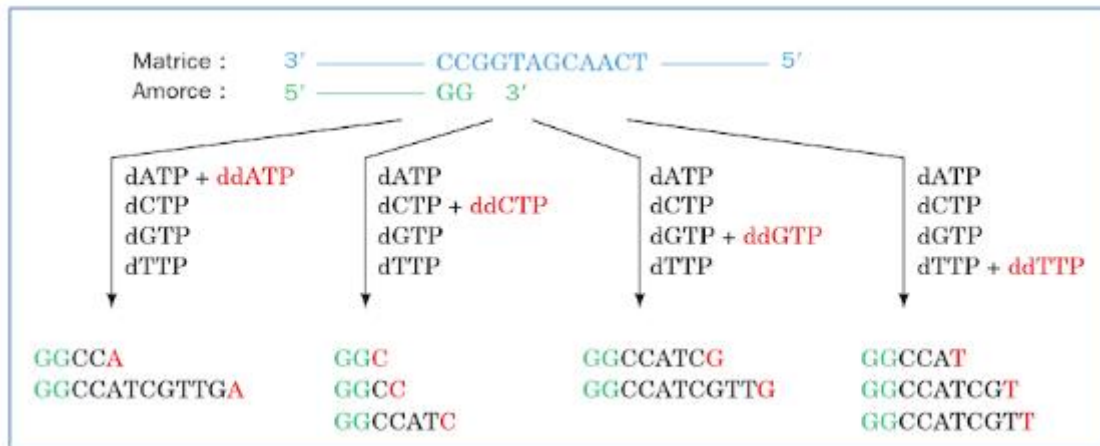


Figure : Séquençage par la méthode enzymatique de terminaison de la chaîne par un Didésoxy

A l'issue de la réaction de séquençage, les molécules d'ADN synthétisées sont séparées suivant leur taille par électrophorèse sur gel de polyacrylamide avec les quatre préparations versées sur quatre lignes parallèles du gel. Si on ajoute des dNTP contenant du phosphore ou du soufre radioactif dans les préparations de séquençage, les ADN synthétisés deviennent radioactifs ce qui permet de les détecter par autoradiographie. Lorsque le film est développé, une série de bande est observée dans chacune des lignes du gel. La séquence de l'ADN matrice peut être déterminée en identifiant les plus petites bandes et les bandes de plus en plus grosses et associant la ligne dans le gel à la base terminale.

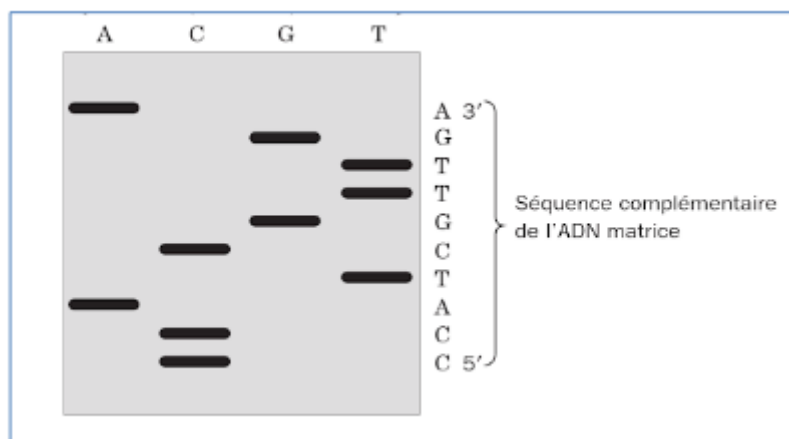
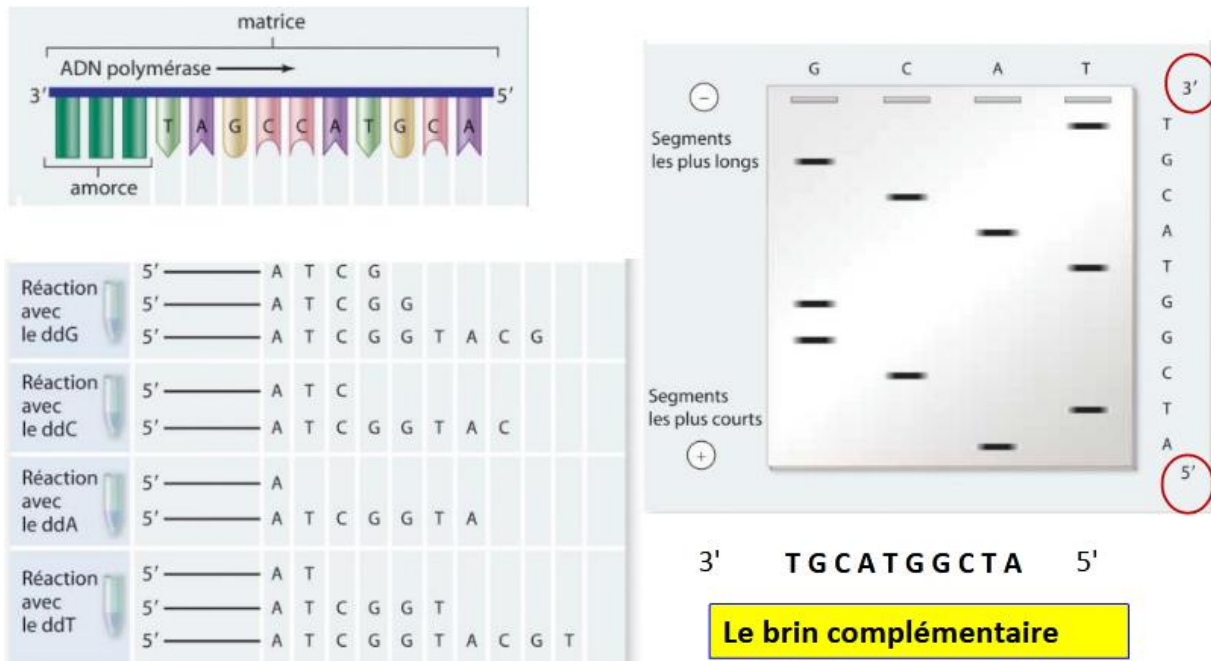


Figure : Motif en échelle obtenu après séquençage d'ADN et électrophorèse

Les didésoxynucléotides peuvent être marqués plutôt que les amorces, ce qui permet d'utiliser un seul mélange au lieu de quatre. Chaque didésoxyribonucléotide est marqué par un fluorophore spécifique. Les fragments d'ADN synthétisés portent ce fluorophore terminal. Ils sont nommés des terminateurs d'élongation ou Big Dye terminators ou Dye-labeled terminator. Cette adaptation a été proposée par Smith et al. nommer séquençage par arrêt de synthèse à l'aide d'un terminateur marqué (dye terminator sequencing).

C'est la technique actuellement utilisée dans certains séquenceurs automatisés.



2. Séquençage automatisé de l'ADN :

La méthode de Sanger a été développée durant ces 25 dernières années, et cela revient à plusieurs avancées technologiques importantes :

- La mise au point de vecteurs de séquençage, comme le phage M13 développé par Joachim Messing au début des années 1980.
- Le développement de la sythèse chimique automatisée des oligonucléotides qui sont utilisés comme amorces dans la sythèse.
- L'introduction de traceurs fluorescents à la place des marqueurs radioactifs utilisés initialement.
- L'adaptation de la technique PCR pour le séquençage.
- L'utilisation de séquenceurs automatiques de gènes
- L'utilisation de l'électrophorèse capillaire pour séparation et l'analyse

Les séquenceurs en gel plat et les séquenceurs capillaires ont permet une automatisation de la techniques dans les laboratoires utilisant couramment le séquençage. La technique de Sanger est celle qui est mise en œuvre dans les premiers séquenceurs automatiques (Le premier séquenceur automatisé a été mis au point en 1987, par la compagnie applied Biosystem ABI).

L'automatisation du séquençage consiste l'emploi de :

- Système d'électrophorèse piloté par ordinateur,
- Marqueurs fluorescents de différentes couleurs qui sont révélés après excitation par laser à l'aide d'une caméra CCD

- Des logiciels permettant l'analyse des signaux sortant de l'appareil et leur mise en forme sous forme de résultats (électrophorégramme et séquence).
- Un robot passeur d'échantillon permettant d'enchaîner les échantillons les uns à la suite des autres (passage de plaques de réaction à 96 puits).

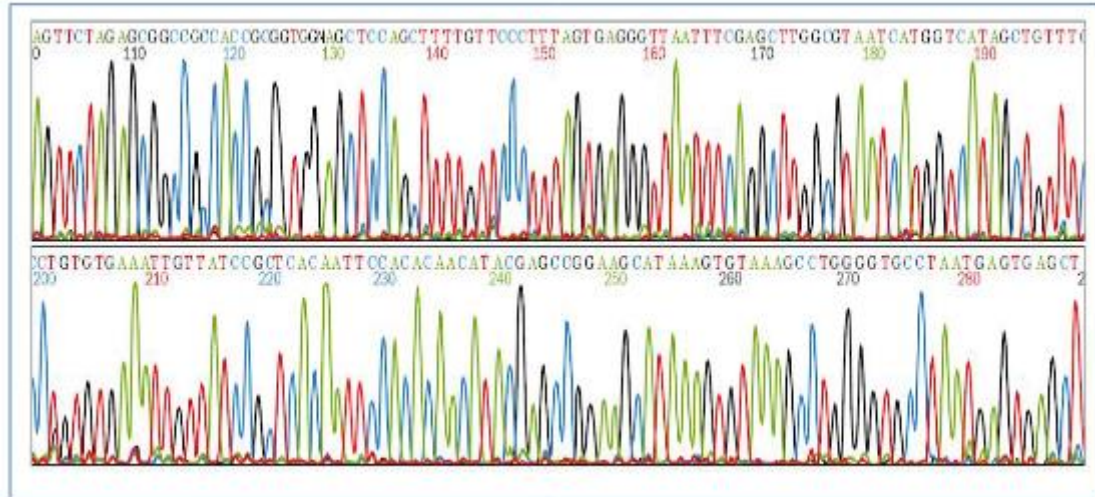


Figure : Exemple d'enregistrement obtenu à partir d'un séquenceur automatique

3.Méthode de Maxam et Gilbert

Cette technique est pratiquement abandonnée de nos jours. Nous la décrivons brièvement pour des raisons historiques. Cette méthode, publiée parallèlement à celle de Sanger en 1977, par son caractère révolutionnaire a grandement contribué à l'histoire de la biologie moléculaire.

Il s'agit, d'une méthode chimique de séquençage. Il est basée sur une dégradation chimique de l'ADN et utilise les réactivités différentes des quatre bases G, C, [A +G] et [C+T] pour réaliser des coupures sélectives. Cette technique permettait d'analyser des fragments allant jusqu'à 500 pb. En reconstituant l'ordre des coupures, on peut remonter à la séquence des nucléotides de l'ADN correspondant. On peut décomposer ce séquençage chimique en six étapes successives :

1. Dénaturer un ADN double brin à un seul brin en augmentant la température.
2. Étiqueter radioactivement une extrémité 5' du fragment d'ADN à séquencer par une réaction de kinase en utilisant gamma- ^{32}P .
3. Cliver le brin d'ADN à des positions spécifiques en utilisant des réactions chimiques. Par exemple, nous pouvons utiliser l'un des deux produits chimiques suivis par pipéridine. Le sulfate de diméthyle attaque sélectivement purine (A et G), tandis que l'hydrazine attaque sélectivement pyrimidines (C et T). Les traitements chimiques décrits dans le document de Maxam-Gilbert clivé à G, A + G, C et C + T. A + G signifie qu'il clive à A, mais de temps en temps à G aussi.
4. Maintenant, dans quatre tubes de réaction, nous aurons plusieurs brins d'ADN de différentes tailles.
5. Les fragments sont soumis à une électrophorèse dans des gels d'acrylamide à haute résolution pour la séparation de taille.

6. Ces gels sont placés sous un film à rayons X, ce qui donne alors une série de bandes sombres qui montrent l'emplacement des molécules d'ADN radiomarqués. Les fragments sont triés par taille et donc on peut déduire la séquence de la molécule d'ADN. Seuls les fragments marqués, c'est à dire seul ceux qui portent l'extrémité 5'- et donc commencent au même endroit-, sont visualisables. Certaines molécules s'arrêtent au premier G, d'autre au second etc..

Si au départ on a fait des coupures séparément et différentes avec plusieurs agents, on obtient pour chaque tube, des molécules de taille correspondant à la séquence. On ne dispose pas de schéma réactionnel permettant de couper après chacune des bases. Mais on dispose de moyens pour couper après G, après A+G, après C+T et après A. Si on charge ces quatre séquences sur le gel, on peut lire la séquence.

Cette méthode n'est plus guère utilisée sauf dans certains cas particuliers, par exemple lorsqu'on veut détecter la liaison d'une protéine sur l'ADN. En effet dans ce cas la liaison de la protéine empêche la coupure chimique, certains fragments sont absents et on observe des blancs sur le gel. Cette technique a été appelée "foot printing".

