

# Chapitre 5 : Séquençage des acides nucléiques

## Introduction

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

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

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

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

### 1- Préparation de la séquence cible

Les méthodes de séquençage utilisent des bases communes :

- La molécule d'ADN à séquencer est extraite, amplifiée, puis dénaturée (on sépare le double brin d'ADN en deux), et on utilise la polymérase, une enzyme dont le rôle est de dupliquer un élément.
- L'amplification, c'est à dire la réplication de la séquence, est réalisée à partir d'une technique appelée PCR ou "Polymerase chain reaction" (réaction de polymérisation en chaîne). Cela permet de dupliquer à plusieurs millions d'exemplaires un fragment d'ADN grâce à l'ADN polymérase.
- La dénaturation (ou déshybridation), c'est à dire la séparation des deux brins composant la molécule d'ADN, est obtenue par élévation de température ce qui a pour effet de fragiliser les liaisons hydrogènes reliant les bases azotées

### 2- Méthodes de séquençage

#### 2-1 Méthode de Maxam et Gilbert

La technique de Maxam et Gilbert repose sur un procédé chimique qui coupe une molécule d'ADN marquée radioactivement à son extrémité 5' ou 3' au niveau d'une base ou d'une famille de base spécifique. Les conditions utilisées sont adaptées pour conduire à une

coupure partielle. Par conséquent la longueur des fragments marqués identifie la position de la base. Les réactions chimiques effectuées clivent préférentiellement l'ADN aux guanines, aux adénines, aux cytosines et thymines et aux cytosines. Les produits des quatre réactions sont résolus par électrophorèse. La séquence d'ADN peut être lue directement à partir du profil des bandes radioactives. Cette technique s'est peu développée car elle nécessite des réactifs chimiques toxiques, n'est pas facile à automatiser et reste limitée quant à la taille des fragments d'ADN qu'elle permet d'analyser ( $\leq 250$  nucléotides)

## 2-2 Méthode de Sanger

La technique de Sanger utilise le principe de la synthèse enzymatique de l'ADN à séquencer en présence d'inhibiteurs d'élongation de l'ADN polymérase les 2',3' di-déoxynucléotides (ddNTP). La réaction de séquençage s'effectue grâce à quatre réactions enzymatiques menées parallèlement. Dans quatre tubes à essais distincts, on introduit en parallèle :

- des brins d'ADN cibles dénaturé en grand nombre,
- des amorces (sur laquelle se fixe la polymérase),
- la polymérase elle-même
- des nucléotides normaux (dATP, dTTP, dCTP, dGTP).
- On ajoute ensuite un nucléotide modifié ou didéshydro-nucléotide (ddATP, ddTTP, ddCTP, ddGTP) de type différent dans chaque tube. Ceux-ci sont identiques aux bases classiques à la différence qu'il leur manque un groupement-HO nécessaire à l'assimilation du nucléotide suivant par la polymérase (liaison phospho-diester).

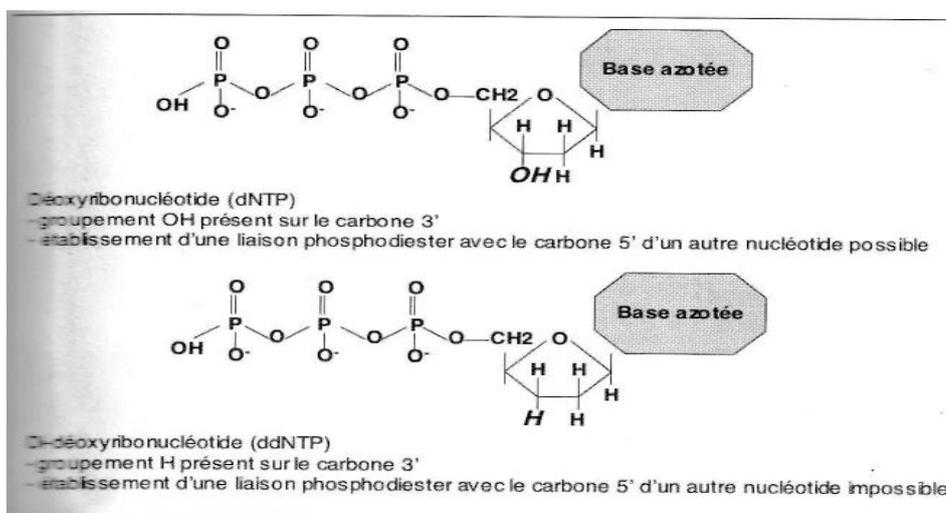


Figure 1 : Les déoxynucléotides et les di-déoxynucléotides

La quantité de nucléotides modifiés incorporés dans chaque tube est bien inférieure au nombre de nucléotides classiques afin de ne pas stopper l'assimilation trop tôt. Il faut qu'il y ait moins de ddNTP que de dNTP pour que dans le tube, on ait toutes les molécules possibles qui soient synthétisées. Ainsi, lors de l'assimilation d'un de ces nucléotides par la polymérase, la réplication du brin est stoppée. Une fois les réactions terminées, on obtient dans chacun des quatre tubes des doubles brins d'ADN de tailles variables, en fonction de leur arrêt par les nucléotides modifiés. Les "réactions de séquence" (avec ddNTP) sont rapides (moins de 15 minutes).

On place alors le contenu des tubes dans quatre puits distincts (correspondant aux quatre nucléotides possibles) de gel d'électrophorèse. L'ADN étant chargée négativement, plus la taille de la séquence est importante et plus la molécule migre vers le pôle positif du gel d'électrophorèse. Après leurs migrations, dans ce gel on peut aisément retrouver l'ordre des nucléotides de la séquence concernée. Une simple lecture horizontale de ce gel (une fois révélé) permet de connaître l'ordre des bases de la séquence.

Afin de voir les fragments d'ADN sur le gel d'électrophorèse, on utilise la plupart du temps des marqueurs fluorescents. On peut aussi marquer radioactivement les nucléotides puis les exposer à un film photographique: des bandes sombres apparaissent là où se trouvait de l'ADN sur le chromatogramme ; c'est l'autoradiographie.

### **3- Automatisation du séquençage**

Aujourd'hui, la plupart des séquençages sont réalisés par des séquenceurs industriels entièrement automatisés. Ceux-ci utilisent la technique de Sanger mais avec des méthodes de révélation différentes.

Les fragments d'ADN sont marqués par des marqueurs fluorescents; leur taille est ensuite déterminée par chromatographie ou électrophorèse assistée par ordinateur.

Avec ces techniques, on peut séquencer jusqu'à 1000 bases avec les meilleurs séquenceurs contre 200 à 300 via une méthode manuelle comme celles exposées ci-dessus. En effet, lors de l'électrophorèse manuelle, le nombre de bases est limité afin de ne pas rendre le chromatogramme illisible par la surcharge des bandes et ainsi ne plus permettre une lecture horizontale.

### **4- Les nouvelles techniques de séquençage**

Depuis 2004, de nouvelles techniques de séquençage sont disponibles sur le marché. Par contraste avec les techniques traditionnelles, elles ont été développées par des industriels qui commercialisent les plateformes automatisées permettant d'utiliser ces techniques. Un autre

point commun très important à toutes ces nouvelles technologies est que l'amplification des banques d'ADN matrice ne passe pas par multiplication clonale, mais par réactions de PCR. Les deux types de réaction PCR utilisées sont :

- **La PCR en émulsion (emulsion PCR or EmPCR)** ; les fragments d'ADN sont liés à des billes d'agarose
- **La PCR par pontage (bridge amplification)** ; les fragments d'ADN sont fixés sur une lame de verre appelée flow-cell.

Ces amplifications par PCR évitent tout biais de représentation des fragments. En effet, lors d'un clonage bactérien, certains fragments d'ADN peuvent présenter une toxicité pour les cellules bactériennes. Ces nouvelles techniques reposent sur :

- La synthèse d'ADN (pyroséquençage et Solexa)
- L'hybridation sur des puces à ADN (SOLiD pour Sequencing by Oligo Ligation and detection développé par Applied Biosystems)
- La détection en temps réel de molécules

### **3- Pyroséquençage luminométrique en temps réel**

L'ADN à séquencer est amplifié par PCR en émulsion. Chaque bille d'agarose portant un exemplaire d'ADN de la banque est déposée dans le puits d'une lame à alvéole analysable par fibre optique. Dans chaque alvéole se déroule la réaction de séquence.

Encore une fois, on utilise un brin dénaturé, une amorce, les quatre types de nucléotide normaux. On ajoute également quatre enzymes : la polymérase, la sulfurylase, la luciférase, et l'apyrase. Les nucléotides, les enzymes et leurs substrats sont introduits dans une cartouche adaptée, et le brin d'ADN dans une micro cuvette. Le tout est mis dans un pyroséquenceur, appareil entièrement informatisé.

Le pyroséquenceur propose alors tour à tour un type de base. Si la base correspond, elle est alors incorporée par la polymérase au brin. Celui-ci rejette alors un phosphate inorganique (PPi) transformé en **ATP** par la **sulfurylase** puis en **lumière** par la **luciférase**. Cette émission lumineuse est alors captée par un détecteur photosensible qui transmet un signal à l'ordinateur. Dans le cas où deux bases identiques se suivent, celles-ci sont incorporées dans le même temps ; l'énergie lumineuse a donc une intensité double se qui se traduit graphiquement sur l'écran. Dans le cas où la base n'est pas incorporée, celle-ci est alors dégradée par l'**apyrase**. Un autre type de nucléotide est alors proposé.

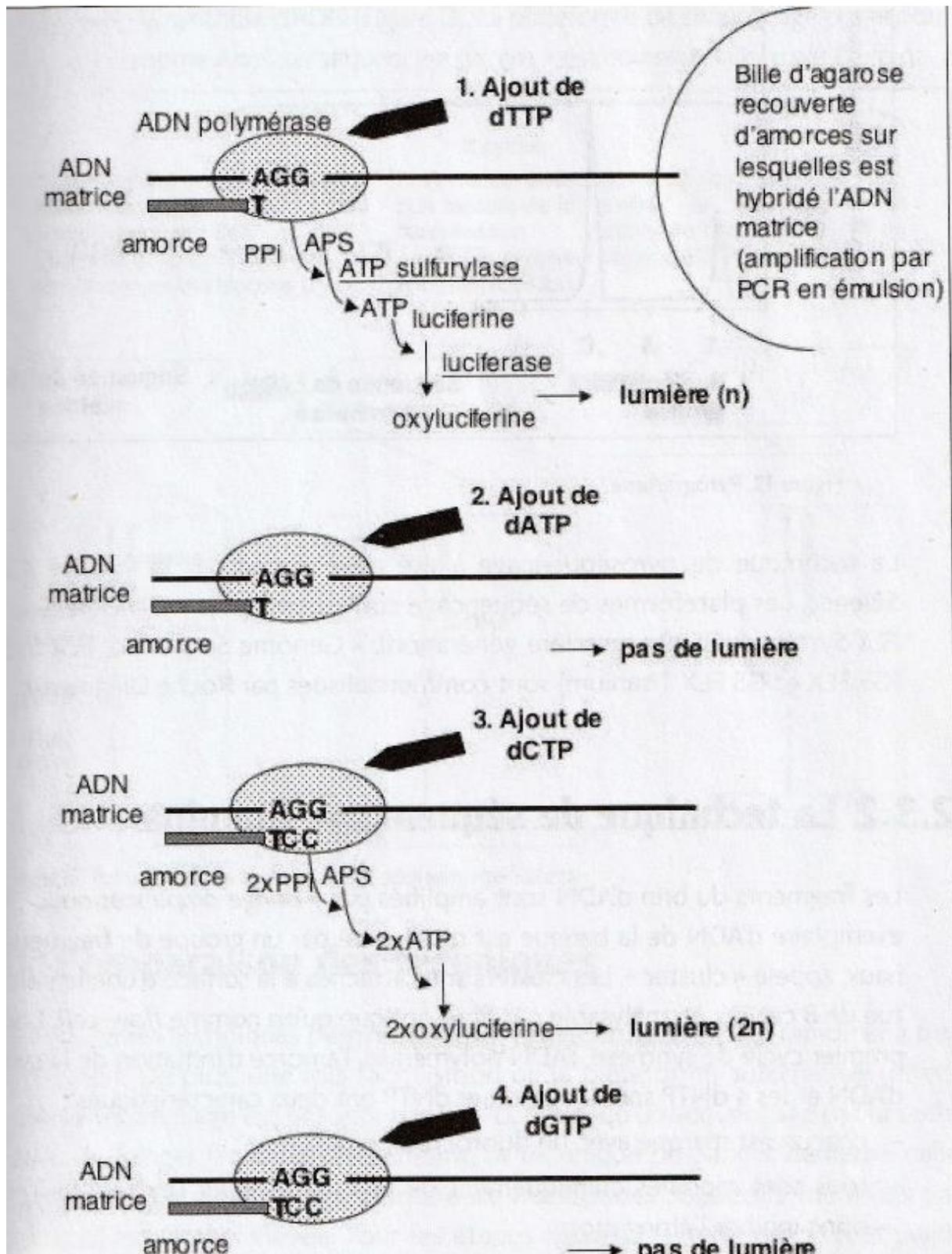


Figure 2 : Technique de pyroséquence

#### **4 - Technique Illumina**

Les fragments du brin d'ADN sont amplifiés par la PCR par pontage, chaque exemplaire d'ADN de la banque est représenté par un groupe de fragments clonaux appelé cluster.

Les clusters sont attachés à la surface d'une lame constituée de 8 canaux analysable par fibre optique qu'on nomme flow-cell. Lors d'un premier cycle de synthèse, l'ADN polymérase, l'amorce d'initiation de la synthèse d'ADN et les 4 bases sont ajoutés.

Ces bases sont marquées avec des fluorochromes différents et tous sont modifiés chimiquement de façon à bloquer l'extrémité 3' OH et donc inhibé l'élongation. La fluorescence de la dernière base incorporée est mesurée avant qu'une réaction chimique n'enlève le fluorochrome et débloque l'extrémité 3'OH à fin de permettre la poursuite de la synthèse d'ADN.

#### **Séquençage de génome complet**

Deux stratégies de séquençage sont utilisées pour le séquençage des génomes complets. La stratégie aléatoire ou shot-gun utilisée pour les génomes relativement simple et de petite taille, et la stratégie ordonnée ou clone par clone utilisée pour les génomes complexes et de grande taille.

##### **a- La construction des banques**

Plusieurs banques de fragments d'ADN génomique sont constituées : le plus souvent, une banque de larges fragments (50-100Kb) dans des bactériales artificiel chromosomes (BAC) et une banque de fragments de taille plus petite (1-10Kb) dans des plasmides.

##### **1-Stratégie aléatoire ou Shot-gun**

Après fragmentation de l'ADN bactérien et la construction de banques de fragments dans des vecteurs de clonage (plasmides pour les fragments de petites tailles et des BAC pour de larges inserts) une amplification des banques par clonage bactérien est effectuée.

La méthode de Sanger est appliquée pour séquencer les inserts de petite taille et les extrémités des larges inserts.

Les extrémités des séquences de larges inserts et les séquences des petites fragments sont ensuite assemblées : par chevauchement des séquences contiguës.

Cette opération d'assemblage est effectuée par des programmes informatiques.

##### **2-Stratégie ordonnée ou clone par clone**

Pour cela on construit des banques de fragments d'en moyenne 100 à 200 Kb dans des BAC. Ces clones sont ensuite ordonnés, c'est-à-dire positionnés les uns par rapport aux autres.

Ceci est effectué par analyse de restriction, hybridation des clones entre eux ou des séquences uniques préalablement identifiées. Les groupes de grands clones chevauchants ainsi constitués sont en suite positionnés le long des chromosomes grâce aux cartes de liaison qui disposent des marqueurs ordonnés le long des chromosomes par la mesure de leur liaison deux à deux.

Par la suite un ensemble de grands clones chevauchants est sélectionné en vue de les séquences.

Cette sélection permet de minimiser les régions de chevauchement entre les grands clones ; puis pour chaque grand clone, une banque de fragments de quelques Kb est produites dans les plasmides. La méthodologie est alors la même que pour la stratégie aléatoire précédemment décrite.