

## Matière Physiologie et Génétique Bactérienne (PGB)

### Chapitre IV : La génomique microbienne

#### 1. La génomique microbienne

La génomique est l'étude de l'organisation moléculaire des génomes, de l'information qu'ils contiennent et des produits codés par leurs gènes. C'est une vaste discipline que l'on peut diviser en trois domaines généraux :

- ❖ **La génomique structurale** : est l'étude de la nature physique des génomes. Son but premier est de déterminer et d'analyser la séquence de l'ADN du génome.
- ❖ **La génomique fonctionnelle** : s'attache à la façon dont le génome fonctionne. C'est-à-dire qu'elle examine les transcrits produits par le génome et l'ensemble des protéines codées par ces transcrits.
- ❖ **La génomique comparative** : qui consiste à comparer les génomes de différents organismes pour y relever les différences et les similarités significatives.

#### 2. La détermination des séquences d'ADN :

Le séquençage de l'ADN est inventé dans la deuxième moitié des années 1970. Deux méthodes sont développées indépendamment, l'une par l'équipe de Walter Gilbert, aux États-Unis, et l'autre par celle de Frederick Sanger (en 1977), au Royaume-Uni. Ces deux méthodes sont fondées sur des principes diamétralement opposés : l'approche de Sanger est une méthode par synthèse enzymatique sélective, tandis que celle de Maxam et Gilbert est une méthode par dégradation chimique sélective. Pour cette découverte, Gilbert et Sanger sont récompensés par le prix Nobel de chimie en 1980.

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

La séquence d'ADN contient l'information nécessaire aux êtres vivants pour survivre et se reproduire. Déterminer cette séquence est donc utile aussi bien pour les recherches visant à savoir comment vivent les organismes que pour des sujets appliqués. En médecine, elle peut être utilisée pour identifier, diagnostiquer et potentiellement trouver des traitements à des maladies génétiques et à la virologie. En biologie, l'étude des séquences d'ADN est devenue un outil important pour la classification des espèces.

**Le séquençage** de l'ADN constitue une méthode dont le but est de déterminer la succession linéaire des bases A, C, G et T prenant part à la structure de l'ADN. La lecture de cette séquence permet d'étudier l'information biologique contenue par celle-ci.

## **2.1. Acteurs du séquençage :**

### **2.1.1. Acide désoxyribonucléique:**

L'acide désoxyribonucléique (ADN) est une molécule présente dans toutes les cellules vivantes, elle renferme l'ensemble des informations nécessaires au développement et au fonctionnement d'un organisme. C'est aussi le support de l'hérédité car il est transmis lors de la reproduction, il porte l'information génétique et constitue le génome des êtres vivants.

#### **2.1.1. Structure de l'ADN :**

La molécule d'ADN est constituée de quatre bases : La cytosine (C), la thymine (T), dérivent de la pyrimidine tandis que l'adénine (A) et la guanine (G) dérivent de la purine. Cette molécule est composée d'un double brin orientés dans le sens 5' vers 3' et reliés entre eux par des liaisons hydrogènes entre les bases purines d'un brin et les bases pyrimidines de l'autre brin. Les bases sont complémentaires deux à deux, deux liaisons hydrogènes relient l'adénine et la thymine tandis que trois liaisons hydrogènes relient la cytosine et la guanine, les deux brins sont complémentaires et antiparallèles.

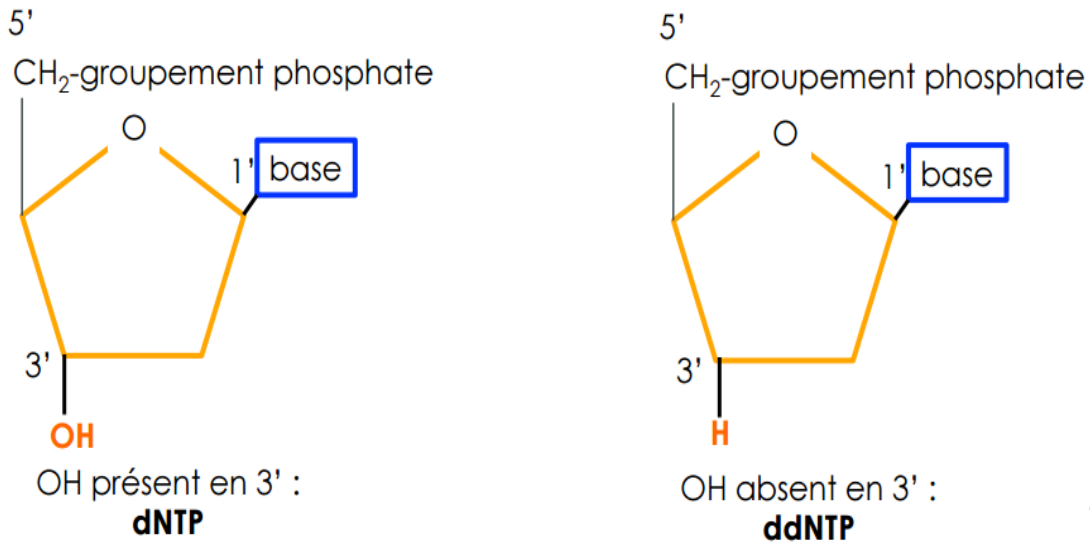
L'ADN cible est extrait de l'échantillon qu'on veut analyser (cultures cellulaires ou d'organismes, prélèvements sanguins, squames, biopsies, salive, cheveux ...) puis, il sera purifié en ADN que l'on souhaite amplifier et qui peut être utilisé dans le séquençage.

#### **2.1.2. Désoxyribonucléotides dNTPs :**

Les désoxyribonucléotides sont les briques de l'ADN (A, C, G ou T), ils sont attachés les uns aux autres grâce à des liaisons chimiques nécessitant la présence d'un groupement OH particulier sur le carbone 3' du ribose. Les nucléotides sont composés d'un acide phosphorique, d'un aldopentose et d'une base hétérocyclique azotée (purine ou pyrimidine). Ils sont utilisés par la Taq polymérase pour la synthèse du nouveau brin d'ADN complémentaire. Ces nucléotides doivent être en concentration très faible; car plus leur concentration est élevée plus on observe d'amplifications parasites et surtout plus la polymérase commet des erreurs de réplication.

### 2.1.3. Didésoxyribonucléotides ddNTPs :

Sont des nucléotides privés du groupement OH en position 3', leur incorporation dans une chaîne d'ADN interrompt définitivement sa synthèse qui continue normalement sur le 3-OH.



**Figure 1** : Structure de dNTP et de ddNTP.

### 2.1.4. Amorces :

Ce sont des fragments courts d'ADN capables de s'hybrider de façon spécifique grâce à la complémentarité des bases sur l'un des deux brins d'ADN. La taille de ces amorces est généralement d'une vingtaine de désoxyribonucléotides.

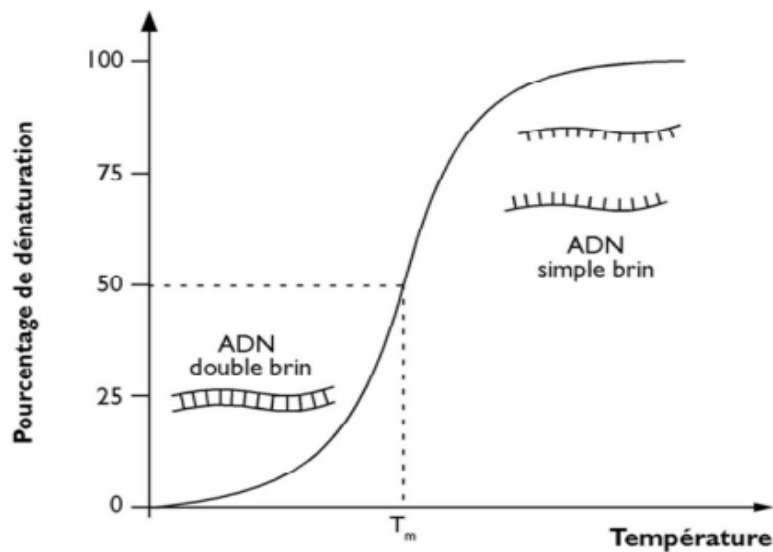
### 2.1.5. ADN polymérase :

La Taq Polymérase est une enzyme purifiée à partir d'une bactérie thermophile, *Thermus aquaticus* vivant à proximité des sources d'eau chaude à 80°C. Cette enzyme possède une activité exonucléase 5'→3' mais sans activité 3'→5' ce qui rend impossible la correction d'erreur lors de la copie. L'ADN polymérase a pour rôle de copier l'ADN en synthétisant un brin complémentaire au brin matrice, elle ne fonctionne qu'en ajoutant des nucléotides à une amorce déjà présente (un A est toujours positionné en face d'un T et un C toujours en face d'un G).

## 2.2. Notion de température de fusion ou $T_m$ :

Quand un ADN double brin est chauffés au-delà d'une certaine température, les deux brins se séparent par rupture des liaisons hydrogènes et deviennent simple brin, la valeur de la température correspondant à ce phénomène s'appelle la température de fusion ou  $T_m$  ("melting temperature").

En pratique, la  $T_m$  correspond à la température où 50 % de l'ADN est sous forme simple brin et 50 % sous forme double brin.



**Figure 2 :** Température de fusion moléculaire d'un fragment d'ADN.

Cette  $T_m$  dépend de plusieurs facteurs:

- 1) La composition en bases : Les appariements AT (deux liaisons hydrogènes) sont moins stables que les appariements GC (trois liaisons hydrogène) en milieu salin (NaCl).
- 2) La composition en sels du milieu: Une diminution de la concentration en sels du milieu diminue la force ionique et par conséquent diminue le  $T_m$  et inversement.
- 3) La quantité de misappariements (ou mismatches) : On considère qu'il y a une diminution du  $T_m$  de 1°C pour 1 % de mismatches dans l'ADN.
- 4) La longueur du fragment : Cet effet est minime quand la taille des fragments hybridés est supérieure à 500 Pb.

- 5) La présence d'agents "déstabilisants" (par exemple la formamide et l'urée). La  $T_m$  est calculé par des logiciels informatiques (bioinformatiques) ou approché par une formule plus pratique :  $2^\circ\text{C}$  par base nucléotidique A ou T +  $4^\circ\text{C}$  par base G ou C.

### 2.3. Etapes du séquençage :

- le séquençage se passe dans un tube à essai en présence des acteurs de la synthèse d'ADN: -ADN à séquencer,
- L'ADN polymérase utilise aléatoirement les nucléotides présents dans le milieu pour copier le brin matrice en synthétisant un ADN de séquence complémentaire.
- Lorsque l'ADN polymérase choisit par hasard un didésoxynucléotide (ce qui est rare puisqu'il y en a moins que des nucléotides) et qu'elle l'incorpore dans la chaîne en synthèse, celle-ci s'interrompt prématurément. Chaque didésoxynucléotide est marqué par un fluorochrome différent (A vert, T rouge, G violet et C bleu), une chaîne qui se termine par exemple par un A sera verte.
- Puisqu'un grand nombre de réactions de synthèse ont lieu dans le tube, il existe statistiquement des chaînes de toutes les tailles (correspondant à un arrêt de la synthèse à chaque nucléotide) et beaucoup de fragments d'une même taille. Ces chaînes commencent toutes au même endroit sur l'ADN matrice, toutes celles qui possèdent la même longueur se terminent par le même didésoxynucléotide marqué.
- Il est alors possible de séparer les chaînes d'ADN obtenues en fonction de leur taille sur un gel d'acrylamide en présence d'un courant électrique. Plus les chaînes sont courtes, plus elles migrent loin et tous les fragments d'une même taille migrent à la même distance.

On obtient alors une succession de bandes colorées, chacune correspondant au dernier nucléotide incorporé. Il suffit de lire la succession des couleurs pour connaître l'ordre des nucléotides.

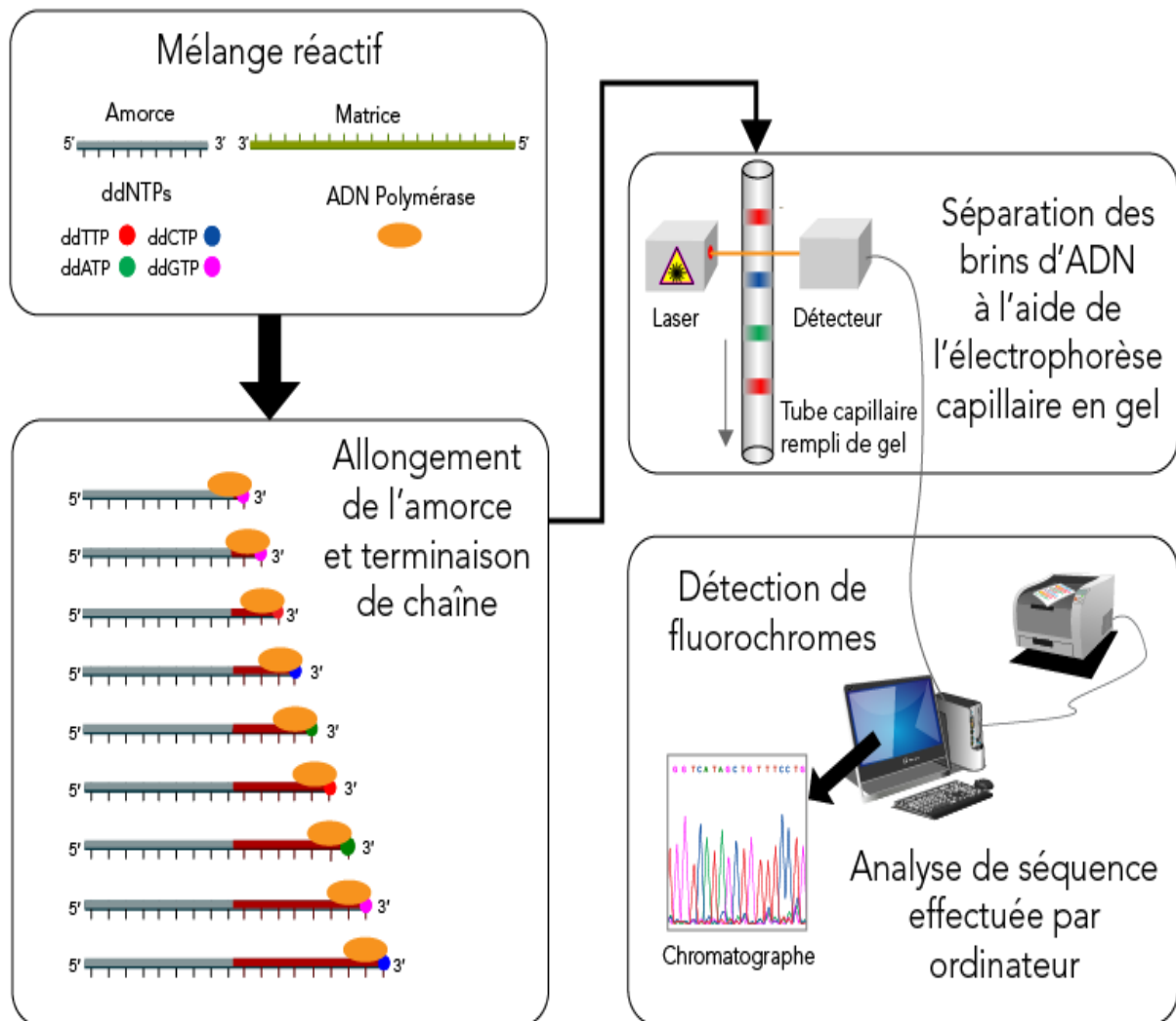


Figure 3 : Différentes étapes du séquençage

**Question :**

C'est pourquoi la technique de Sanger est fréquemment appelée méthode de séquençage de l'ADN par terminaison de chaîne.