

IV.1. Banques d'ADN

Lors du clonage moléculaire, la multiplication de la cellule hôte contenant le gène d'intérêt va produire plusieurs copies qui lui sont tout à fait identiques, dont chaque copie représente un **clone recombinant**. Toutes les cellules filles constituant le clone, contiennent une copie du gène ayant été introduit dans la cellule mère qui a donné naissance à ce clone. Ce clonage a donc permis d'obtenir un bon nombre de copies identiques à la séquence initialement insérée dans la cellule receveuse, donc il s'agit d'une **amplification *in vivo***. L'ensemble de ces clones est appelé : **banque d'ADN** (ou librairie : *library*). Une banque d'ADN est caractérisée par : le type de vecteur, le pourcentage de vecteurs possédant un insert, la taille moyenne des inserts, la taille de la banque (la quantité de clones) et la nature des inserts. Selon ce dernier point, il existe deux types de banques d'ADN: banque d'ADN génomique (ADNg) et banque d'ADN complémentaire (ADNc).

IV.1.1. Banque d'ADN génomique (ADNg) : Elle se définit par l'ensemble de vecteurs recombinants contenant chacun un morceau différent du génome de l'individu étudié. L'obtention d'une banque génomique est nécessaire pour la caractérisation d'une séquence génomique particulière, établir la carte physique d'un génome ou séquencer à grande échelle des portions ou la totalité d'un génome. Deux points essentiels sont pris en considération lors de la construction d'une banque d'ADN, dont le principe général de sa construction a été étudié dans le chapitre du clonage moléculaire :

- Le choix de l'enzyme de restriction : Le génome doit être digéré par l'enzyme de restriction qui permet de libérer des fragments de taille et d'extrémités compatibles avec le vecteur choisi. Egalement, il est préférable d'obtenir des fragments chevauchants (Fig.15) que des fragments contigus (non chevauchants), afin d'éviter que la séquence recherchée ne soit malencontreusement coupée en plusieurs morceaux saturant la banque (éviter de saturer la banque par de clones de petite taille). L'une des façons de répondre à ces exigences est de réaliser une **digestion partielle**.
- Le choix du vecteur de clonage : Le choix du vecteur présente une étape cruciale au moment de l'établissement de la banque, car on se trouve généralement confronté au problème de la taille des gènes (principalement les gènes des eucaryotes) qui risquent de ne pas être acceptés par le vecteur. Le tableau 03 présente la taille de l'insert accepté par chaque vecteur.

La formule statistique de Clark et Carbon donne le nombre de clones nécessaires pour construire une banque d'ADNg :

p : probabilité de présence d'une séquence ($p = 0.99$).
n: rapport entre la longueur du génome et la longueur moyenne des fragments insérés

$$N = \frac{\log(1-p)}{\log(1-\frac{1}{n})}$$

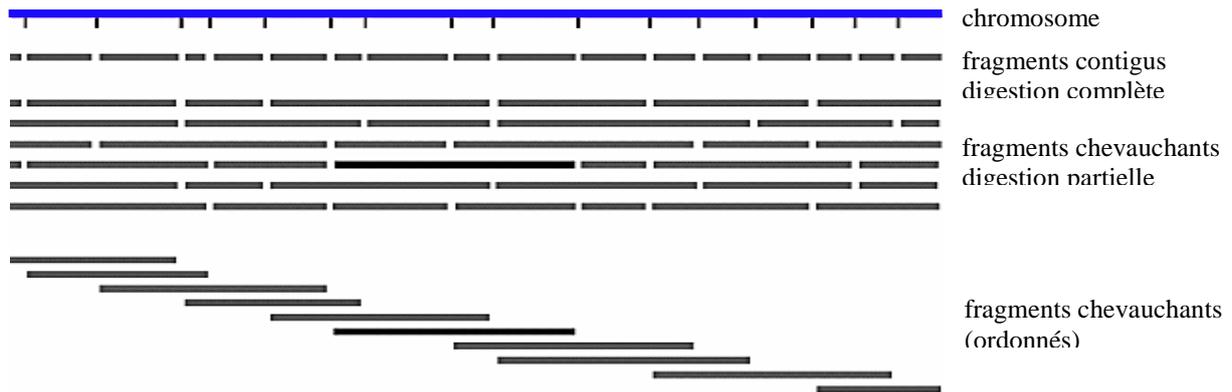


Fig.15 : Fragments contigus et fragments chevauchants

Tab.03 : Différents vecteurs et taille de l'insert accepté

Vecteur de clonage	Taille de l'insert
Plasmide bactérien	<20 kb
Bactériophage	9-25 kb
Cosmides	30-44 kb
BAC (chromosome bactérien artificiel)	Jusqu'à 300 kb
YAC (chromosome artificiel de levure)	0.2-2.0 Mb

IV.1.2. Banque d'ADN complémentaire (ADNc) : Comme la manipulation des ARNm est délicate (faible quantité, sensibilité aux nucléases), il est nécessaire de les transformer en ADN bicaténares aisément manipulables et adaptés à des clonages dans des vecteurs bactériens. Une banque d'ADNc est donc représentative des populations d'ARNm présents dans un tissu. Contrairement à une banque d'ADNg, une banque d'ADNc sera spécifique d'un tissu. Elle peut être considérée comme une photographie instantanée des populations d'ARNm représentées dans un organe, la construction d'une banque d'ADNc nécessite de nombreuses étapes, décrites ci-dessous et résumées dans la Figure 16.

- A.** Extraction d'ARN de l'organe.
- B.** Purification des ARNm polyadénylés, s'il est d'origine eucaryote (Exp. : par chromatographie d'affinité sur une colonne polyT).
- C.** Copie de ces ARNm en ADN complémentaire par l'action de l'enzyme **transcriptase inverse (RT)**. Cette enzyme fait spontanément une boucle lorsqu'elle atteint la fin de sa matrice.
- D.** Elimination spécifique des ARNm par l'**ARNase H** ou la soude.
- E.** Synthèse du second brin d'ADN par une **ADN polymérase**.
- F.** Ligature d'oligonucléotides (adaptateurs=linkers) par une **ligase** pour créer des sites de restriction. Les étapes de synthèse du second brin d'ADN et de ligature d'adaptateurs sont de plus en plus souvent réalisées à l'aide de la PCR.

G. Insertion dans un vecteur de clonage (généralement un phage λ).

H. Intégration des vecteurs recombinant dans une bactérie.

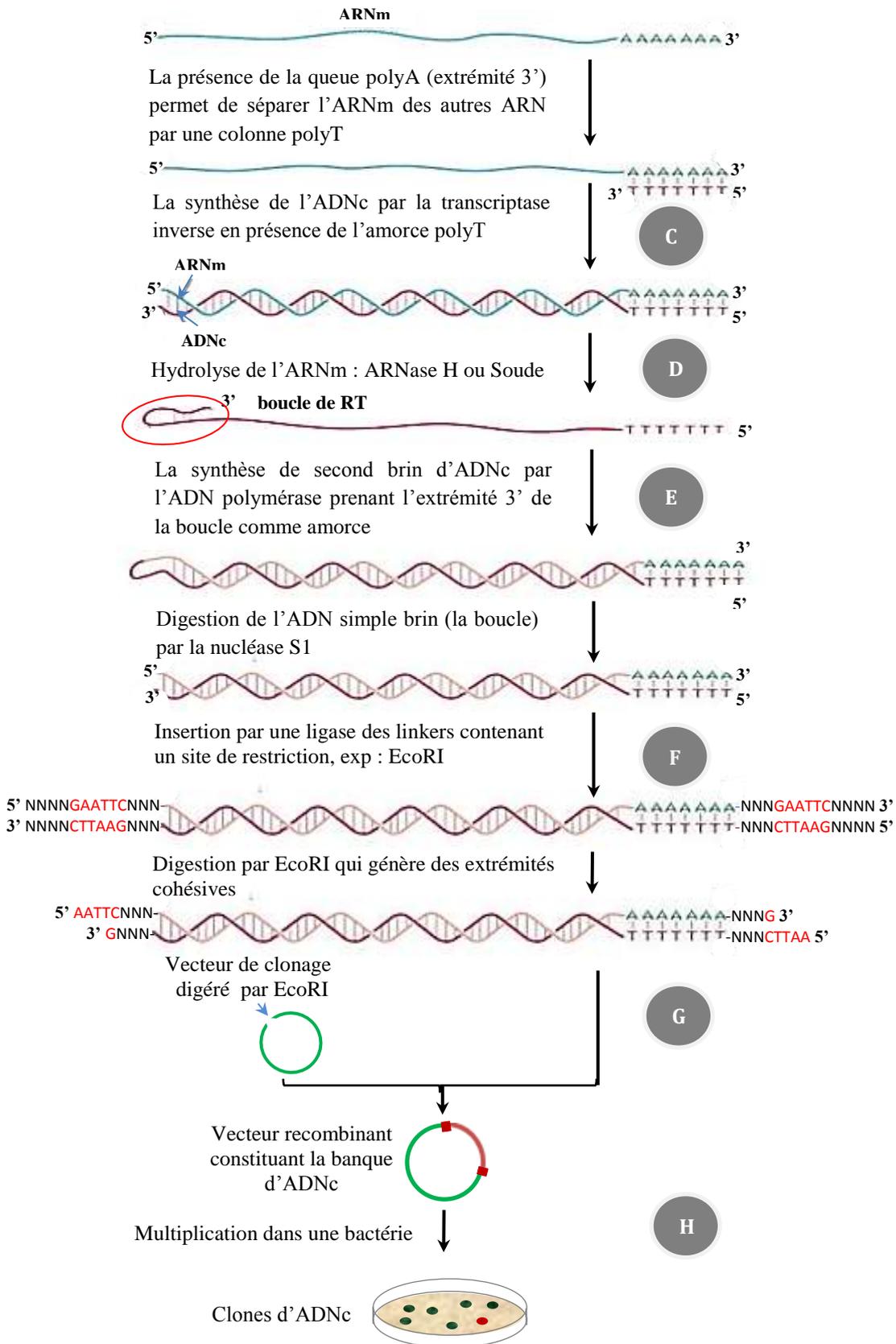


Fig.16 : Construction d'une banque d'ADNc

IV.1.1.3. Analyse de la banque (criblage)

Après multiplication des cellules receveuses, on peut isoler un clone donné grâce à une propriété donnée : c'est le criblage. Le criblage d'une banque génomique ou d'ADNc consiste à sélectionner spécifiquement des clones contenant la séquence du gène recherché : clones recombinés, sans éliminer les clones non intéressants. Nombreuses méthodes de criblage existent : le criblage par clonage par marche sur chromosome, moléculaire (hybridation), immunologique (western blot) et le criblage biologique qui se base sur la présence d'un gène intéressant (marqueur génétique) qui confère un phénotype particulier à l'hôte tel qu'une résistance à un antibiotique particulier auxquels les cellules sauvages sont sensibles ou la production d'une enzyme. En effet, le criblage par hybridation de l'ADN des clones transformés avec une sonde spécifique marquée constitue une méthode de choix (Fig.17).

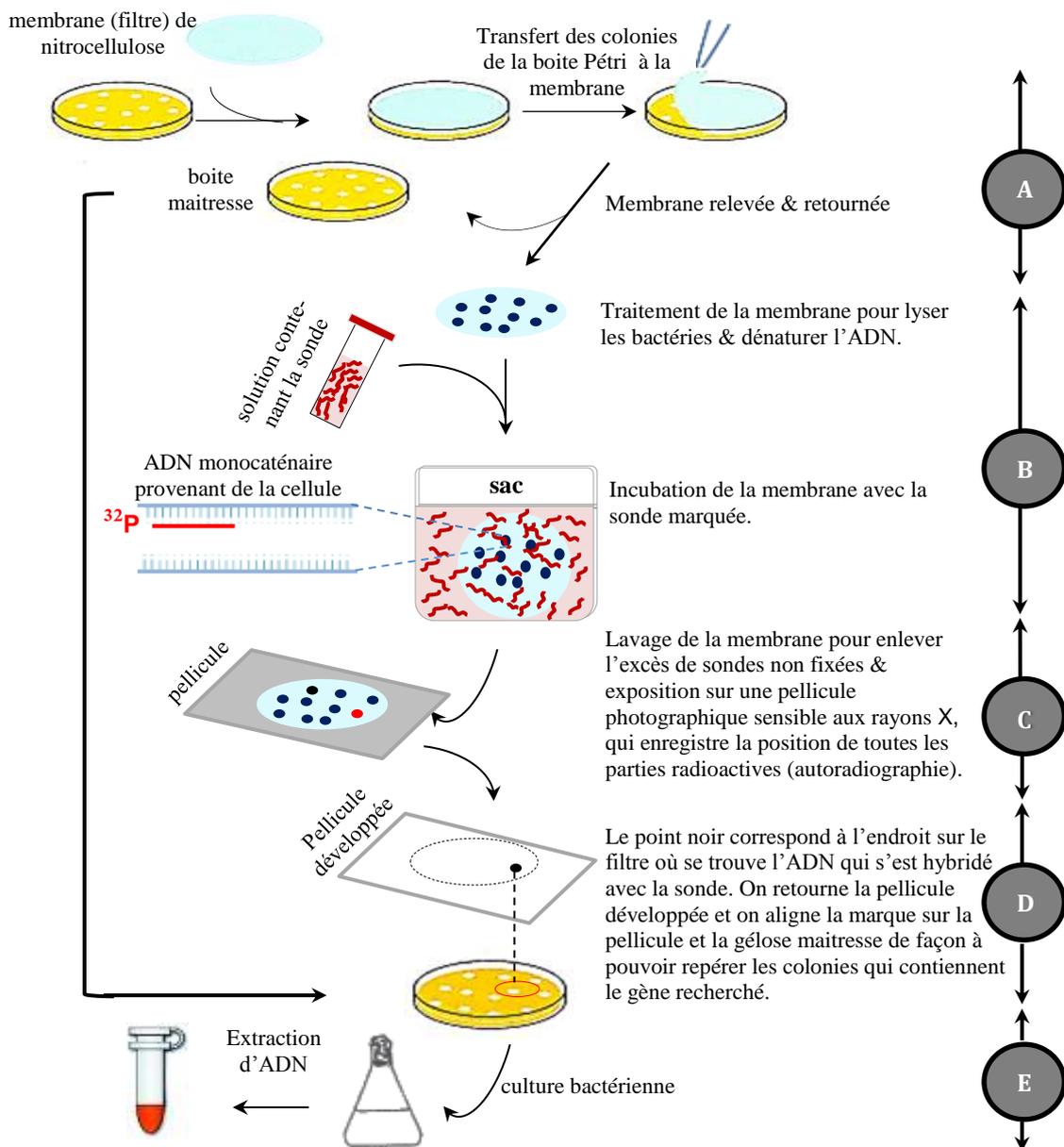


Fig.17 : Criblage de cellules recombinées à l'aide d'une sonde nucléique. **A** : Répliques, **B** : Hybridation moléculaire, **C** : Révélation, **D** : Identification des clones, **E** : Collecte d'ADN d'intérêt.

IV.2. Amplification génique *in vitro*

IV.2.1. PCR classique

La PCR (*Polymerisation Chaine Reaction*) mise au point dans les années 80, est une technique permettant d'amplifier (obtenir un nombre très élevés de copies) *in vitro* une séquence particulière d'ADN (ou parfois d'ARN) ce qui facilite grandement son **étude** et son **exploitation**. La découverte d'une bactérie thermophile : *Thermus aquaticus* vivant dans les sources chaudes (70 à 75°C), et l'utilisation par la suite de son ADN polymérase (**Taq polymérase**), **stable** jusqu'à des températures de **100°C**, est à l'origine du développement de cette technique. La technique de PCR a littéralement révolutionné les recherches en biologie moléculaire et trouve de nombreuses applications aussi bien dans le **clonage** et l'étude de l'**expression des gènes** que dans la recherche d'un **polymorphisme génétique**.

IV.2.1.1. Principe : Le principe de réplication d'ADN utilisé par les cellules (*in vivo*) est exploité pour l'amplification des fragments d'acides nucléiques par PCR. Ce principe repose sur la répétition de trois phases constituant un cycle de PCR (**Fig.18**):

- **Dénaturation à 95°C):** La dénaturation des deux brins d'ADN à température très élevée afin d'obtenir des molécules d'ADN monocaténares.
- **Hybridation à 40-65°C) :** L'hybridation des amorces sur les brins matrices monocaténares est effectuée à une température bien précise appelée la **température d'hybridation** (Th) ou **Ta** en anglais « *annealing temperature* ». Elle dépend des tailles des amorces et de leurs compositions. Pour calculer la Th d'une amorce on utilise la formule suivante : **$Th (^{\circ}C) = 4(C+G) + 2(A+T) - 5$**
- **Elongation à 72°C) :** L'élongation par la Taq polymérase à partir des amorces, réalisée à la température optimale de cette enzyme.

IV.2.1.2. Réaction de PCR : cinq ingrédients sont nécessaires pour une réaction de PCR, ils sont mélangés dans un microtube réactionnel (tube PCR):

- L'extrait d'ADN possédant la séquence cible à amplifier (en faible quantité : moins de 1µg).
- Les deux amorces (*primers*): sens et anti-sens (*forward & reverse primers*) qui encadrent la séquence cible à amplifier. Sachant que l'amorce sens s'hybride avec le brin anti-sens et l'amorce anti-sens s'hybride avec le brin sens.
- Les désoxynucléotides triphosphates (dATP, dCTP, dGTP, dTTP).
- La Taq polymérase thermostable.
- Le tampon (TrisHCl à pH basique 8,5 à 9) qui sert à maintenir stable le pH du milieu réactionnel au niveau optimal pour la Taq polymérase, ce tampon est additionné du magnésium « Mg^{++} », cofacteur indispensable au fonctionnement de de cette polymérase.

Les microtubes contenant le mélange réactionnel sont placés dans la machine de PCR, appelée également : thermocycleur. Cette machine permet d'exposer les tubes à des températures et durées choisies et déterminées par le manipulateur. Les températures de dénaturation et de d'élongation (polymérisation) sont fixes, seule la température d'hybridation devra être calculée pour chaque nouvelle PCR. Un thermocycleur permet d'effectuer les transitions de températures **automatiquement**, dont le produit d'un cycle sert de matrice pour le cycle suivant, où le nombre de copies du fragment d'ADN croît selon une progression géométrique de raison 2, par la suite 2^n molécules sont ainsi obtenues après n cycles.

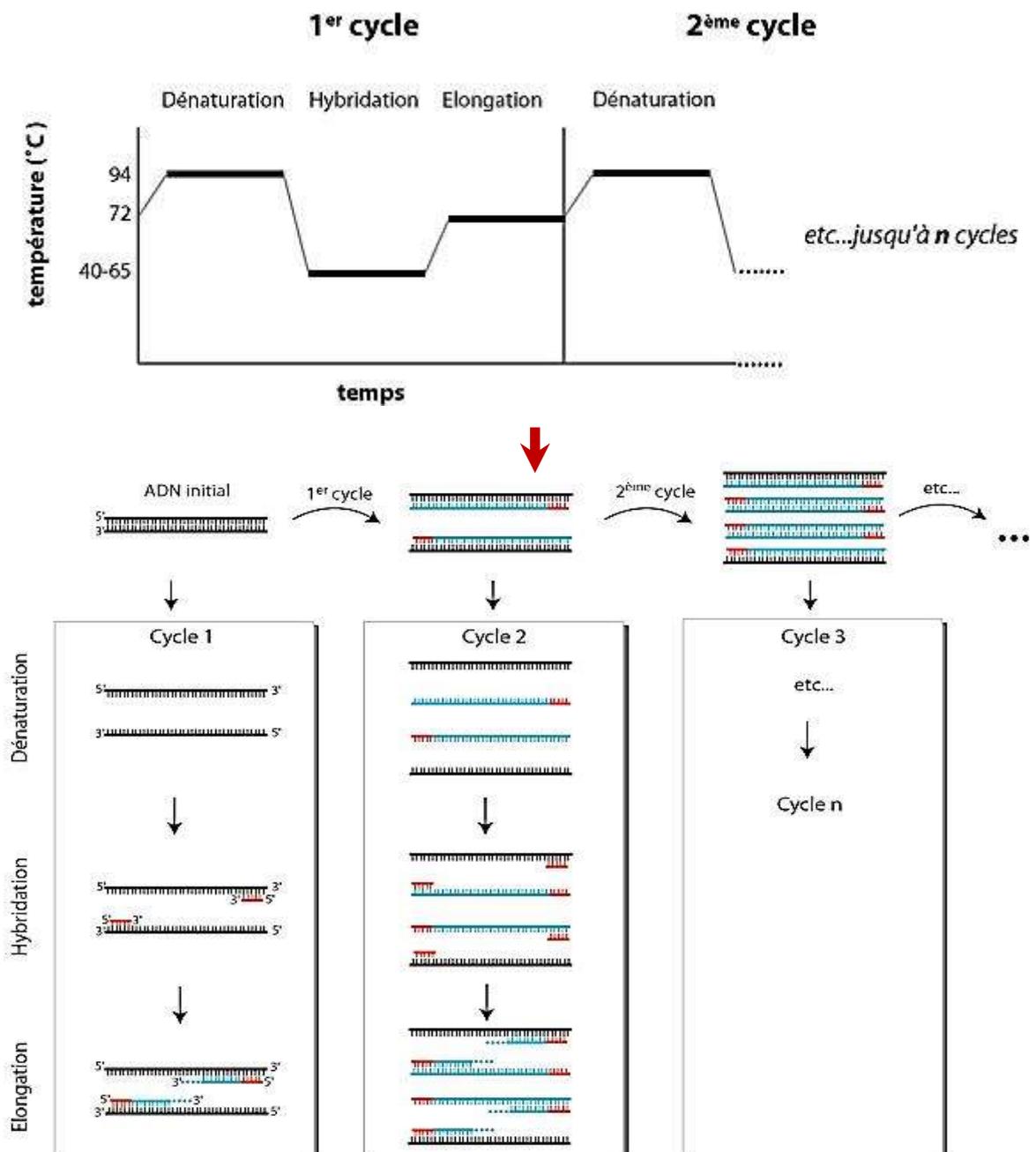


Fig.18 : Schéma illustratif de l'amplification par PCR avec les trois phases constituant un cycle.

IV.2.1.3. Facteurs clés pour la réussite d'une PCR : La spécificité de PCR dépend de façon critique des amorces. Les facteurs énumérés ci-dessous sont déterminants dans le choix d'amorces performantes.

- La longueur des amorces doit être comprise entre 17 et 30 nucléotides.
- Un contenu en GC de 50% est idéal. Si les amorces correspondent à des séquences pauvres en GC, il convient de les allonger pour éviter des températures de fusion trop basses.
- Des séquences contenant de longues répétitions d'un même nucléotide (c'est-à-dire plus de trois ou quatre d'affiliée) doivent être évitées.
- Des amorces susceptibles de former des structures secondaires sont inutilisables.
- Les deux amorces ne doivent pas présenter de complémentarité de séquence.

La grande majorité des amorces conformes à ces règles donnent de bons résultats, bien que leur efficacité puisse varier d'une paire à l'autre, même dans des conditions de réactions optimisées.

IV.2.2. PCR quantitative (q-PCR) ou PCR en temps réel

Pour de nombreuses applications de la PCR, il est important de pouvoir estimer la quantité du matériel de départ. Théoriquement, il y a une relation directe entre la quantité du matériel de départ (la séquence cible) et la quantité de produit formé après un nombre donné de cycles de réaction. En pratique, on constate que les cycles successifs ont des rendements variables, ce qui rend l'estimation théorique peu fiable. Une PCR n'est plus quantitative si on analyse le résultat de l'amplification à la fin du processus d'amplification, c'est-à-dire lorsque les échantillons se trouvent dans la phase stationnaire (Fig.19). Pour analyser le déroulement de l'amplification d'ADN dans la phase exponentielle, il est nécessaire de déterminer la quantité d'ADN produite par la réaction de PCR à la fin de chaque cycle. Higuchi et *al.* (1992, 1993) furent les premiers à mesurer la quantité des produits de PCR au fur et à mesure de leur accumulation en utilisant le bromure d'éthidium. Le principal inconvénient de l'utilisation de bromure d'éthidium est que les produits non spécifiques génèrent aussi un signal. C'est pour quoi une méthode basée sur l'utilisation de sonde a été développée pour mesurer l'accumulation du produit (Livak et *al.*, 1995). A la différence d'une PCR classique, la PCR en temps réel permet, parallèlement à l'amplification (les mêmes étapes qu'une PCR classique), la détection et la quantification de l'ADN. Elle repose sur l'utilisation d'un marqueur fluorescent (par exemple SYBR Green ou sonde Taqman®) dont le signal augmente proportionnellement à la quantité d'ADN synthétisée au cours de la réaction. Ce système de détection - système fermé - c'est à dire sans ouverture des tubes réactionnels, rend cette PCR plus rapide et plus pratique qu'une PCR de base, en limitant considérablement les risques de contamination à partir des produits d'amplification générés au cours de la réaction enzymatique. Ce système présente également l'avantage d'éviter l'étape de la préparation du gel pour l'analyse et le contrôle des résultats (nécessaire dans une PCR classique).

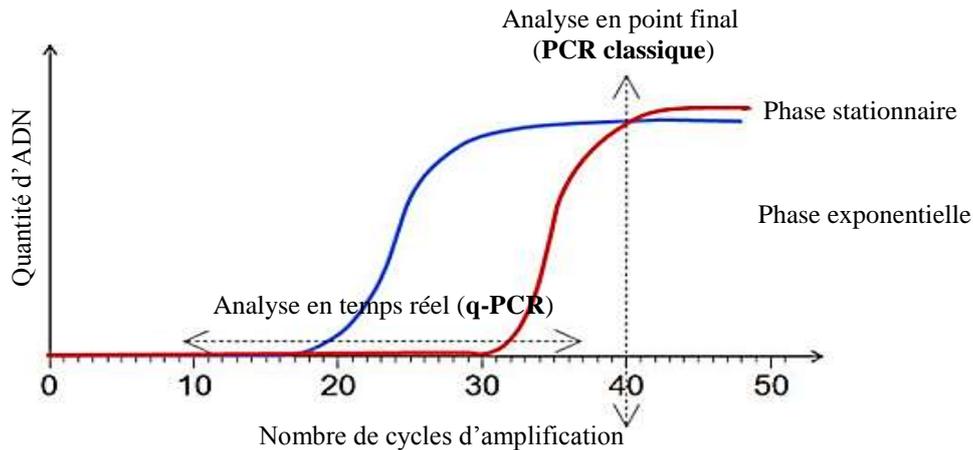


Fig.19 : Les deux modes d'analyse du résultat de l'amplification par PCR

Dans une PCR en temps réel avec une sonde, et parallèlement à la paire d'amorces, une troisième sonde spécifique (dans l'exemple présenté dans la **Fig.20** elle est de type Taqman®) se fixe au centre de la région à amplifier. Cette sonde possède un fluorophore rapporteur « *reporter* » et un désactivateur « *suppresseur : quencher* » qui inhibe l'émission de fluorescence par le rapporteur lorsqu'il est situé à proximité. Au cours de l'élongation, l'activité exonucléase (5' vers 3') de l'ADN polymérase dégrade la sonde hybridée au brin matrice, permettant la libération du rapporteur et l'expression de sa fluorescence. Le rapporteur ainsi séparé du suppresseur émet un signal proportionnel au nombre de sondes hydrolysées, mesurable au moment de l'élongation. Par conséquent, la cinétique d'émission de la fluorescence est proportionnelle à la quantité d'amplicons produits au cours de la réaction.

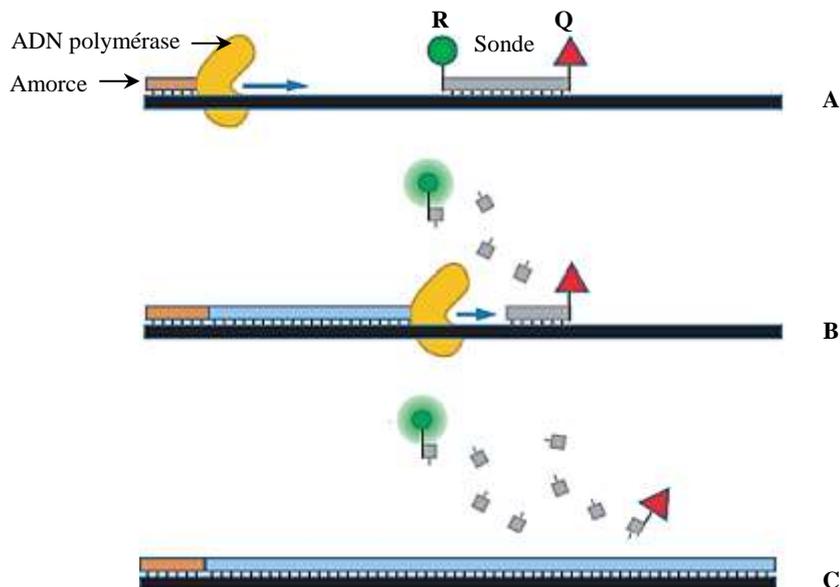


Fig.20 : PCR en temps réel (PCR quantitative : q-PCR) avec sonde spécifique. R : *reporter*. Q : *quencher*. **A :** La polymérisation, dont le rapporteur « R » est situé à proximité du suppresseur « Q ». **B :** Dégradation de la sonde, **C :** Libération du rapporteur et émission de la fluorescence.

IV.3. Détermination des séquences des acides nucléiques (séquençage)

Le séquençage d'un acide nucléique se définit par la détermination du nombre, de la nature et de l'ordre (succession) des nucléotides qui le composent. Déterminer la succession des nucléotides composant une séquence d'ADN est un élément indispensable des manipulations génétiques modernes. En médecine, elle peut être utilisée pour identifier, diagnostiquer et potentiellement trouver des traitements à des maladies génétiques et virales. En biologie, l'étude des séquences d'ADN est devenue un outil important pour la classification des différentes espèces. Les études ont d'abord concernées des fragments d'ADN puis des gènes entiers et enfin l'intégralité d'un chromosome. Deux méthodes sont développées indépendamment, l'une par Maxam et Gilbert, et l'autre par Sanger. Ces deux méthodes inventées dans les années soixante-dix sont fondées sur des principes diamétralement opposés, dont l'approche de Maxam et Gilbert repose sur la dégradation chimique sélective, tandis que celle de Sanger repose sur la synthèse enzymatique sélective. L'inconvénient majeur de la première méthode était l'utilisation de composés chimiques hautement cancérigènes. Cette méthode n'est plus utilisée depuis longtemps au profit de la méthode de Sanger qui est à la base du fonctionnement des séquenceurs automatiques actuels. C'est pour quoi seule cette méthode sera étudiée.

IV.3.1. Principe

Le principe de cette méthode repose sur l'utilisation des nucléotides légèrement modifiés : les **didésoxyribonucléotides triphosphate (ddNTP)** qui sont des analogues structuraux aux nucléotides habituels (dNTP), mais ils sont dépourvus du groupement OH en position 3', d'où le nom de « didésoxy » (**Fig.21**). L'absence du groupement 3'OH n'autorise pas une nouvelle liaison phosphodiester, autrement lorsqu'une ADN polymérase incorpore un ddNTP, le nucléotide suivant (dNTP) n'est plus capable de s'incorporer, ce qui conduit à l'arrêt de la synthèse du brin d'ADN (blocage de la synthèse).

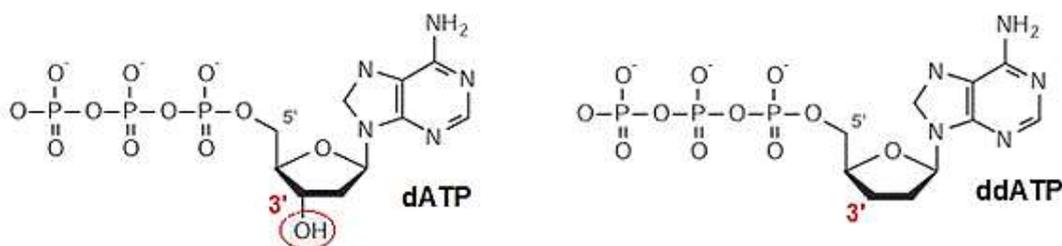


Fig.21 : Structure chimique d'un nucléotide (dATP) et son analogue structural (ddATP).

IV.3.2. Acteurs de séquençage : Pour faire le séquençage selon la méthode de Sanger appelée encore la méthode par terminaison de chaîne, on doit disposer des éléments suivants :

- L'**ADN** à séquencer qui doit être préalablement dénaturé pour servir de matrice.
- L'**amorce** d'ADN c'est un court fragment d'ADN simple brin (une vingtaine de nucléotides), qui doit être complémentaire à l'ADN à séquencer. Elle permettra de fournir une extrémité 3'OH libre à l'ADN polymérase.
- L'**ADN polymérase** est une enzyme qui permet de faire la synthèse du brin complémentaire au brin à séquencer. L'ADN polymérase ne fonctionne qu'en ajoutant des nucléotides à une **amorce** déjà présente (elle ne sait qu'allonger une chaîne préexistante), en fonction de la succession de nucléotides du brin **matrice**.
- **Les didésoxynucléotides :** ddATP, ddTTP, ddCTP, ddGTP, dont leur incorporation dans une chaîne d'ADN interrompt définitivement la synthèse de l'ADN.
- **Les désoxynucléotides :** les quatre types de nucléotides (dATP, dTTP, dCTP, dGTP) qui sont les briques de l'ADN.

IV.3.3. Réaction de séquençage et révélation des fragments

L'amorce nucléotidique est hybridée au fragment d'ADN monocaténaire (dénaturé) dont on veut déterminer la séquence. A partir de l'extrémité 3'OH de l'amorce appariée, une ADN polymérase synthétise le brin complémentaire de l'ADN matrice en présence de désoxyribonucléotides. L'un des désoxyribonucléotides doit être marqué (radioactif ou fluorescent). La polymérisation se réalise dans 4 tubes, chacun de ces tubes contient, en plus des dNTP, le ddNTP correspondant (Fig.22). Chaque fois qu'un ddNTP est incorporé à une position, l'élongation de la chaîne est stoppée. En effet, puisqu'un grand nombre de réactions de synthèse ont lieu dans chaque tube, il existe statistiquement des fragments de différentes tailles. Ces chaînes commencent toutes au même endroit sur l'ADN matrice (déterminé par l'amorce utilisée) et se terminent toutes par le même ddNTP (dans le même tube).

Le contenu des 4 tubes est ensuite analysé par électrophorèse en gel de polyacrylamide-urée en condition dénaturante. Les différents fragments marqués néosynthétisés migrent en fonction de leur taille et sont séparés à la **base près**. Une autoradiographie (cas des nucléotides radioactifs) ou une lecture en fluorescence (cas des nucléotides fluorescents) est réalisée après migration des molécules dans le gel. La séquence 5'P → 3'OH du brin néosynthétisé peut être lue du bas vers le haut, en comparant les positions relatives des bandes dans les 4 pistes.

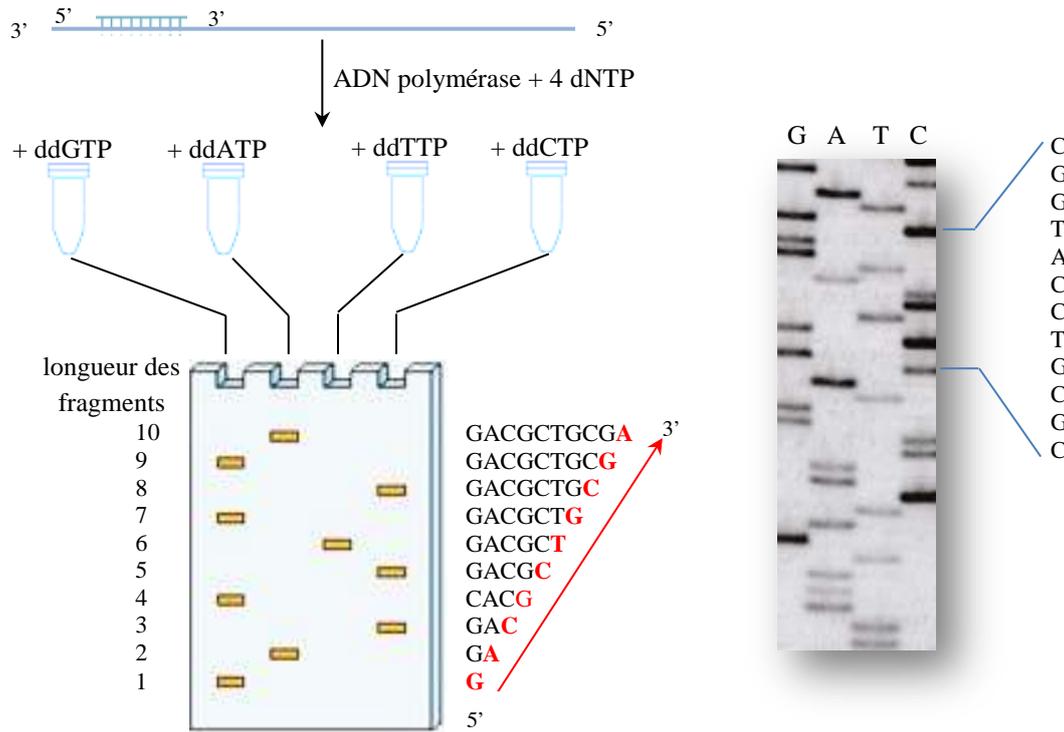


Fig.22. Séquençage selon la méthode de terminaison de chaîne (cas de révélation par autoradiographie).

IV.3.4. Adaptation de la technique à la fluorescence (Séquençage automatisé)

Dans le séquençage manuel, les fragments d'ADN sont marqués par **radioactivité**, et séparés sur quatre pistes d'un gel de séquençage, puis détectés par **autoradiographie**. Ce protocole se prête mal à l'automatisation. Pour l'automatiser, il est préférable d'obtenir les données de séquence en temps réel, en détectant les bandes d'ADN dans le gel pendant l'électrophorèse. Cela est rendu possible grâce au marquage par fluorescence, où chaque ddNTP est marqué par un fluorochrome différent, dont le spectre d'émission est spécifique. Dans le séquençage automatisé et pendant la migration, lorsqu'une bande d'ADN passe devant le faisceau, un signal de fluorescence est émis. Celui-ci est capté par un détecteur (photodiode) situé en regard du gel. Le signal est amplifié puis transmis à l'ordinateur de contrôle et analysé par un logiciel spécialisé (Fig.23).

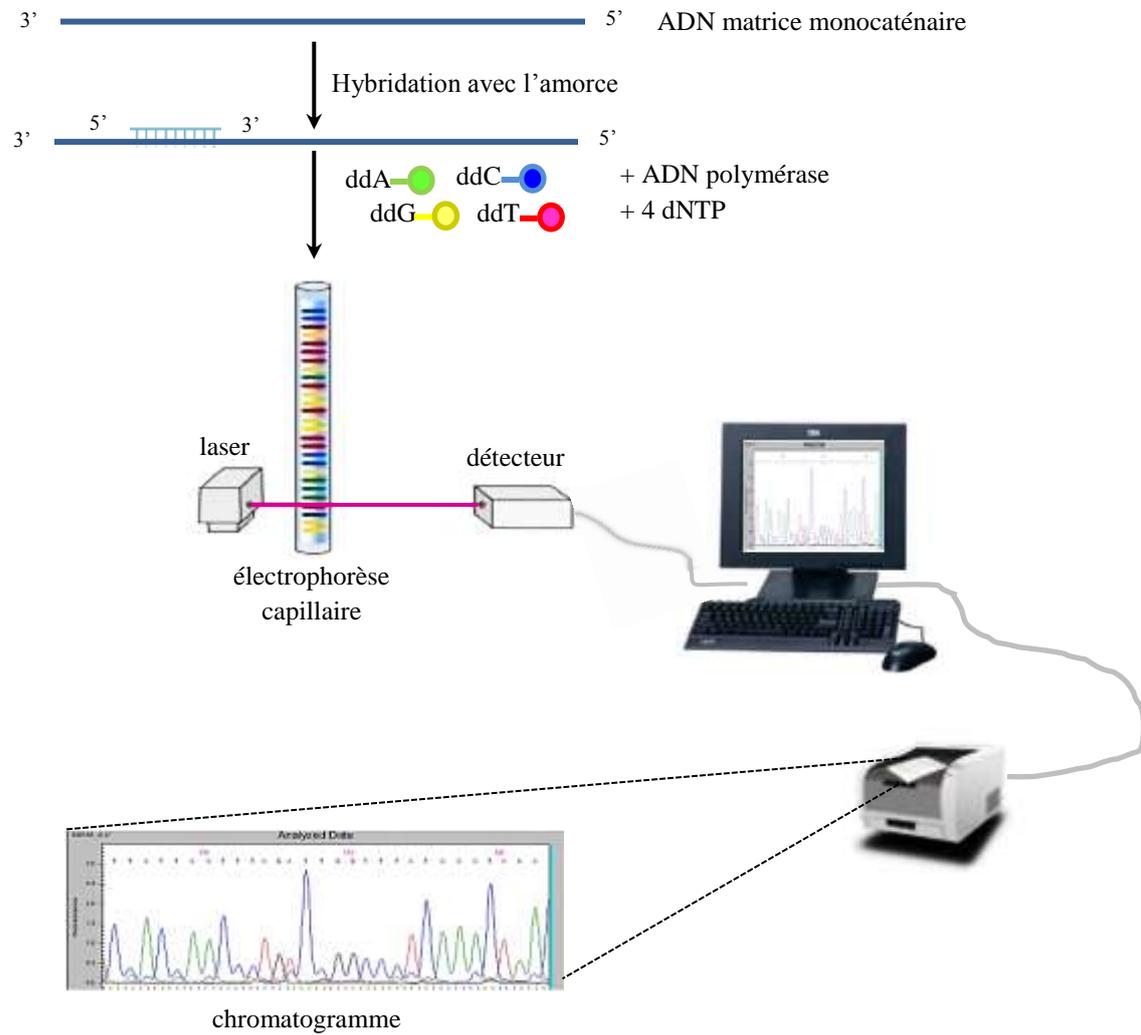


Fig.23 : Séquençage automatisé (méthode *Dye terminator*).