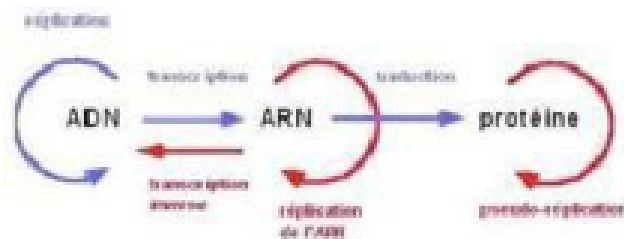


APPLICATIONS – MÉTHODES D'ÉTUDE GÉNÉTIQUE 2



On a vu que l'ADN pouvait se répliquer, être transcrit en ARN, puis l'ARN traduit en protéine. Il existe des exceptions à ce schéma :

- l'ARN peut se répliquer (ex : chez les virus à ARN)
- il existe une pseudo-réplication des protéines (connue sous le terme de prion). C'est un changement de conformation d'une protéine à partir d'une autre protéine (pas de modification de la structure primaire).
- on trouve aussi la transcription inverse (rétro-transcription), de l'ARN vers l'ADN (ex: chez les rétrovirus)
L'ADN transcrit peut alors donner plusieurs ARN qui permettront la synthèse de plusieurs protéines.
- notion d'épigénétique: C'est la modification de l'expression des gènes sans que la séquence primaire de l'ADN soit modifiée.

I. Propriétés de l'ADN

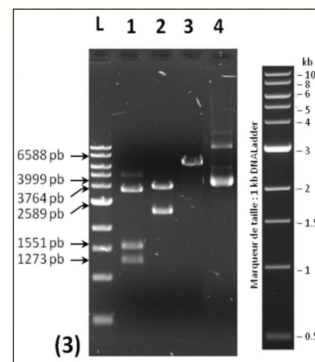
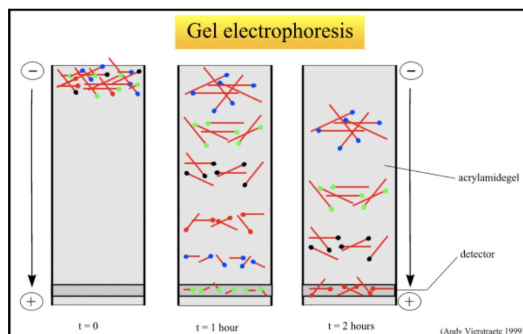
1.Charge électrique

L'ADN et l'ARN sont chargés négativement. Les bases sont orientées vers l'intérieur de la double hélice. Les phosphates (acides forts) sont à l'extérieur, ce qui confère une charge globale négative à l'ADN ou l'ARN. On utilise cette propriété pour faire migrer les molécules sur un champ

électrique (vers la cathode ++). Pour la migration, on utilise des gels d'agarose ou d'acrylamide qui sont des gels de porosité variable (maillage plus ou moins important), pour retenir les molécules en fonction de leur taille. Les plus petites avancent le plus rapidement, donc elles migrent plus loin. Cette technique s'appelle l'électrophorèse. C'est un système horizontal, on peut aussi en avoir des verticaux.

Pour pouvoir observer les molécules sur le gel, il faut un **marquage** ; on utilise ensuite un intercalant (ex : Bromure d'éthidium, qu'on remplace par d'autres molécules maintenant car il s'avère cancérigène). L'intercalant est des molécules qui émettent une fluorescence augmentée, donc plus importante lorsqu'elles se positionnent entre les bases. Il y a ensuite une détection de ces molécules fluorescentes et donc d'ADN lorsque l'on place le gel sous UV (UV excite la molécule de bromure d'éthidium). Ainsi, plus il y a d'ADN, plus c'est fluorescent.

On peut ainsi créer une échelle (comme ici sur la bande L) à partir de molécules d'ADN de taille connue pour mesurer la taille d'une molécule inconnue, ce que l'on appelle un marqueur de taille.



2. Absorption

Table 10.4 Absorption properties of nucleotides*

	λ_{\max} (nm)	$10^{-3} \epsilon_{\max}$ ($M^{-1} \text{cm}^{-1}$)
<i>Ribonucleotides</i>		
Adenosine-5'-phosphate	259	15.4
Cytidine-5'-phosphate	271	9.1
Guanosine-5'-phosphate	252	13.7
Uridine-5'-phosphate	262	10.0

Les propriétés d'absorption sont liées à la présence des doubles liaisons au niveau des hétérocycles dans les bases. Ces bases n'ont pas toutes les mêmes propriétés, comme le montre ce tableau.

Lorsqu'on mélange toutes les bases, (car chaque base a une absorbance spécifique) on a un seul pic à 260nm (pic qui correspond à l'absorption maximale de l'ADN).

L'absorbance est égale à $\epsilon \cdot L \cdot [C]$ (loi de Berr- Lambert, à connaître).

ϵ = coefficient d'extinction molaire ou d'absorption molaire, on estime que c'est une valeur constante, qui caractérise la capacité d'absorption de la lumière pour chaque molécule.

L = longueur du trajet optique

[C] = la concentration en ADN. L'absorbance augmente donc avec la concentration de l'ADN.

En pratique, on estime qu'une solution d'ADN pure (double brin) à 50µg/mL donne une absorbance =1.

L'ADN a un maximum d'absorbance à 260 nm alors que les protéines ont un maximum à 280 nm. En pratique, on mesure toujours l'absorbance à 260 nm et à 280 nm pour vérifier que la solution n'est pas contaminée par des protéines. En réalité, on mesure le rapport 260/280. A partir du moment où il est supérieur ou égal à 1,8, on considère que l'ADN est pur. Pour l'ARN on considère qu'il doit être supérieur ou égal à 2. Si le rapport est inférieur à 1,8, on considère qu'on a des protéines dans la solution.

En résumé:

ADN "pur" -> Rapport 260/280 = 1,8

ARN "pur" -> Rapport 260/280 = 2,0

Si < à 1,8: présence de protéines

Maintenant on a des appareils qui peuvent mesurer des micros volumes (nanodrop, on y dépose juste une goutte). Ils donnent la courbe et les valeurs de concentration directement calculés.

Il existe un autre rapport. En effet, quand on veut de l'ADN pur, on fait une extraction d'ADN. On a alors recours à des solvants, qui à l'issue de la purification, peuvent rester et poser problème pour de futures études.

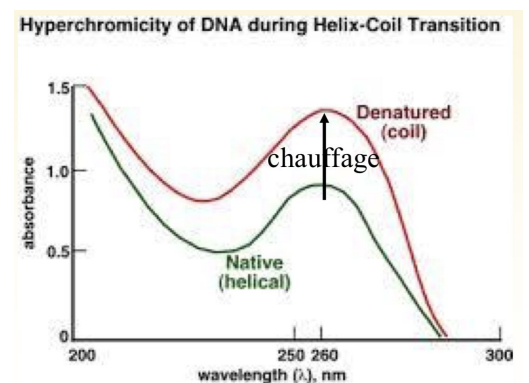
On utilise alors la mesure de la densité optique à 230 (rapport 260/230). Elle nous indique si l'échantillon est plus ou moins contaminé. Ce rapport doit normalement être aux alentours de 2.

3. Dénaturation – renaturation

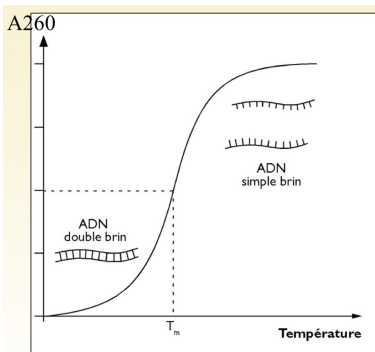
Quand on chauffe de l'ADN, l'absorbance à 260 nm augmente: c'est l'effet **hyperchromique** (ou hyperchromicité de l'ADN). Cet effet est lié aux propriétés de l'ADN (la forme simple brin absorbe plus que la forme double brin). Suite à l'échauffement l'ADN double brin devient simple brin (les liaisons hydrogènes ont été rompues), ce qui expose les bases (elles ne sont plus à l'intérieur), augmentant ainsi l'absorption.

Sous forme double brin, les bases sont masquées au centre de la double hélice et l'absorbance est plus faible. En dénaturant, les liaisons hydrogènes vont être rompues, l'ADN devenant alors simple brin.

De ce fait les bases se retrouvent exposées et absorbent plus les UV utilisés pour la mesure de l'absorbance.



La **température de fusion** (ou T_m) (m=melting) est la température à laquelle on a 50% d'effet hyperchromique. Cette T_m dépend :



- Du pourcentage en bases GC (3 liaisons hydrogène, plus difficiles à rompre)
- De la présence de mésappariement (liaisons plus faciles à rompre)
- De la force ionique du milieu
- De la taille de l'ADN (les plus gros ont plus de mal à se dissocier)

Pour avoir une idée globale : $T_m = 2 X$ (quantité de A-T) + $4 X$ (quantité de C-G). {si taille <20 nucléotides}

La dénaturation de l'ADN (par chauffage) est un **processus coopératif** car le fait de dissocier certaines régions du double brin induit la dénaturation d'autres régions.

Quand la température diminue, on a une renaturation, reformation spontanée du double brin. Ce processus est relativement lent.

Il existe d'autres formules plus précises pour calculer la T_m : $T_m (^{\circ}\text{C}) = 81.5 + 0.41(\%CG) - (675/N)$ avec N nombre de nucléotides.

4. Dégradation

Clivage chimique: hydrolyse acide qui va détruire la liaison N-osidique entre la base et le sucre. On peut même arriver à séparer les 3 composants avec des acides forts.

Pas réellement utile d'un point de vue de la recherche.

Clivage enzymatique : coupure de la liaison phosphodiester entre le phosphate et le sucre. Ce clivage est possible grâce aux nucléases.

Il existe des nucléases non spécifiques :

- **Les endonucléases** qui coupent au milieu de l'ADN et génèrent des fragments. Elles attaquent la liaison phosphodiester au niveau de l'hydroxyle et du phosphate. Elles coupent à l'intérieur du double brin de façon aléatoire.

Elles peuvent être spécifiques (enzyme de restriction) ou non (ex : DNase 1)

- **Des exonucléases** qui coupent aux extrémités de l'ADN. Elles attaquent la liaison phosphodiester et génère des fragments 5'P et 3'OH. On en trouve également des spécifiques et des non spécifiques. Ex : Lambda (qui coupe de 5' en 3') et ExoIII (qui coupe de 3' vers 5').

Le clivage enzymatique est plus contrôlé et mieux maîtrisé que le clivage chimique.

- **Des RNAses** : (ce sont des nucléases, qui s'attaquent aussi à la liaison phosphodiester)

o La RNase H coupe des molécules **hybrides** constituées d'ARN et d'ADN. Elle ne coupe pas de simple brin ou de double brin d'ADN. Elle peut être impliquée dans la destruction des amorces d'ARN présentes durant la réplication ou après la réaction de transcription inverse. Mais dans tous les cas elle ne coupe que l'ARN dans cet hybride.

o La RNase A coupe de l'ARN simple brin (monocaténaire) après les résidus pyrimidique. Petite spécificité, elle génère des brins 5'OH – 3'P.

Ces deux dernières sont utiles en laboratoire pour **purifier** des solutions d'ADN, par élimination de l'ARN résiduel éventuellement présent dans la préparation.

II. Méthodes d'étude de l'ADN

1. Amplification par PCR (réaction en chaîne de la polymérase)

C'est une technique qui comprend 3 étapes :

- **dénaturation** (par dissociation des 2 brins par chauffage),
- **hybridation** (appariement des amorces *{primers}* spécifiques de la région à amplifier avec le brin matrice d'ADN, afin de borner la région) (on descend la température pour arriver à une température optimale d'hybridation)
- et **élongation** (=polymérisation) (synthèse d'ADN) → on utilise une polymérase qui synthétise de 5' en 3'.

Elle permet une amplification **exponentielle** de la quantité d'ADN car chaque nouvelle copie est réutilisée pour un autre cycle.

Cette méthode a connu un essor particulier lorsqu'on a découvert des polymérases résistantes à des hautes températures (Taq polymérase), pouvant ainsi être utilisées dans des cycles successifs de PCR.

Tout d'abord, on effectue une dénaturation par la chaleur (~95°C) qui dissocie les doubles brins. On refroidit ensuite (50°C) pour permettre l'hybridation d'amorces sur l'ADN simple brin. Ces amorces sont spécifiques et bornent la région qu'on souhaite amplifier. La connaissance de la T_m est importante pour le choix des températures d'hybridation des amorces. On met des amorces en excès pour favoriser leur hybridation (sinon il y a un risque que les brins d'ADN s'hybrident entre eux). La position des primers définit la taille de l'amplicon.

On passe ensuite à l'élongation, faisant intervenir la DNA polymérase à *environ 70°C (dépend de la polymérase)*.

La quantité de produit amplifié par PCR, en théorie, est : $X_n = X_0 \cdot 2^n$

Il y a différents types de polymérase, dont la Taq polymérase, issue d'une bactérie thermophile vivant dans des conditions extrêmes. C'est une enzyme qui résiste à la chaleur. Ce qui permet de la conserver intacte tout au long des cycles.

Toutes les polymérase synthétisent en 5'→3'.

Certaines sont capables d'avoir une **capacité d'autocorrection** (activité exonucléatique 3'→5')

Caractéristiques des polymérase:

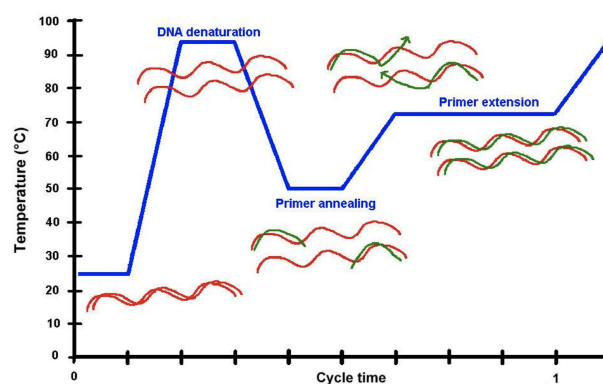
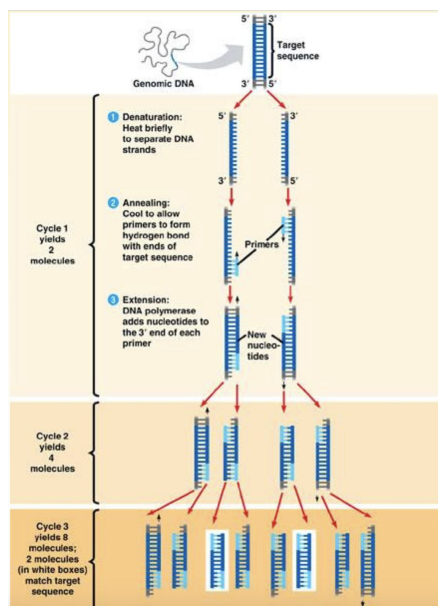
- Notion de processivité: les polymérase sont plus ou moins rapides en fonction de leur nature
- Fidélité
- Température d'élongation qui peut varier selon leur nature

Certaines polymérase ont, par ailleurs, une activité exonucléatique 5'→3'.

A l'issue de la PCR, on fera une électrophorèse pour vérifier que l'amplification s'est bien faite.

Cette méthode est utile en recherche mais aussi en clinique pour identifier des bactéries en milieu infectieux, en recherchant l'ADN d'une bactérie.

On aura une amplification bordée par les amorces. Plus l'enzyme est capable de rester accrochée à l'ADN, plus elle va être rapide.

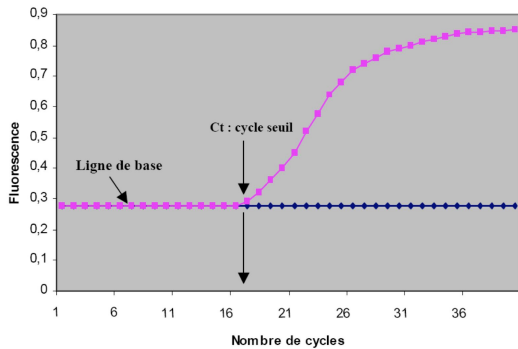


PCR quantitative (PCR en temps réelle) :

Elle permet de polymériser et détecter simultanément ce qui est produit au cours de la réaction (d'où "en temps réel").

La PCR quantitative est une méthode qui utilise les caractéristiques de la PCR classique mais, dans ce cas on va la suivre en temps réel grâce à l'utilisation d'intercalants spécifiques qui vont émettre de la

fluorescence beaucoup plus importante quand ils sont liés à l'ADN double brin (mélange avec un agent intercalant associé à un fluorochrome). On utilise toujours un thermocycleur mais cette fois-ci associé à des sondes, permettant un système de détection de fluorescence (qui se fait simultanément à la PCR). La **Sonde Sybr Green** est un exemple de sonde fluorescente qui se fixe sur n'importe quel ADN double brin. On peut ainsi suivre l'évolution de la quantité d'ADN synthétisé en temps réel. Ces fluorophores sont non spécifiques. La fluorescence augmente avec l'amplification (car augmentation de l'adhésion du fluorochrome sur l'ADN).



Notion de cycle seuil :

La fluorescence augmente avec le nombre de cycle. A partir d'un certain nombre de cycles, la fluorescence dépasse le « bruit de fond » et commence à croître réellement : c'est ce qu'on appelle cycle seuil, qu'on va utiliser pour mesurer la quantité d'ADN présente dans le milieu. Plus il y a d'ADN matrice à amplifier dans l'échantillon, plus le cycle seuil est faible (car s'il y a moins d'ADN il faudra alors un plus grand nombre de cycle pour l'amplifier).

Le thermocycleur peut changer la température très rapidement et voir la fluorescence.

Le problème est que l'intercalant n'est **pas spécifique**, on peut mesurer la fluorescence pour n'importe quel ADN présent. Il y a des problèmes s'il y a eu contamination ou si amorces peu spécifiques. On fait donc une étape de courbe de fusion, on refait le cycle et on va mesurer la fluorescence tous les 1°C dans la dernière étape pour vérifier si l'amplification est bien spécifique (dépend de la séquence d'ADN).

Cette courbe de fusion donne le T_m , on aura pour chaque produit d'amplification une courbe de fusion, on doit donc en avoir une. Si on a d'autres pics, on a eu une contamination.

On va monter à 95°C pour dissocier brins puis on adapte progressivement la température pour avoir T_m .

On peut aussi faire la dérivée de cette courbe où c'est plus facile de voir s'il y a une contamination ou non (cela permet d'avoir l'assurance d'avoir amplifié un seul produit).

Il faut faire beaucoup de contrôles pour être sûr du résultat.

Pour quantifier la réaction:

- Dans le cas où on a une gamme étalon (une solution de référence dont on connaît les valeurs de concentration), on peut faire une courbe d'étalonnage en fonction de la quantité d'ADN.
- Souvent on n'a pas de solution standard, on fait donc une quantification relative (permet de comparer l'échantillon à un échantillon témoin). Cela nous donne juste un rapport. (Se fait souvent lors de l'étude de l'expression des gènes, permettre de connaître le niveau d'expression d'un gène par rapport à un autre). On va utiliser la méthode des Delta Ct.

Méthode des Delta Ct :

1- Normalisation par « un gène de ménage » : pour être sûr d'avoir la même quantité de matériel.

$$\Delta Ct = Ct_{\text{gène d'intérêt}} - Ct_{\text{gène de ménage}}$$

2- Normalisation par un échantillon témoin

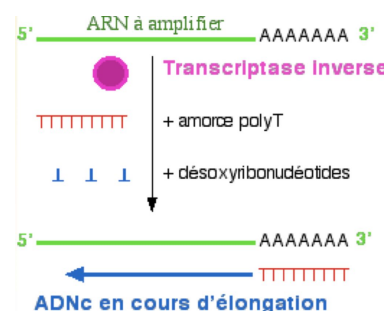
$$\Delta \Delta Ct = \Delta Ct_{(\text{échantillon à tester})} - \Delta Ct_{(\text{échantillon témoin})}$$

3 - Calcul du ratio d'expression : $2^{-\Delta \Delta Ct}$

On peut aussi utiliser des sondes TaqMan, qui sont des sondes d'hydrolyse (complémentaires de la région que l'on veut amplifier). C'est un système comportant un reporter fluorescent (un fluorochrome quoi) et un Quencher qui empêche la fluorescence du reporter lorsqu'elles sont relativement proches. Lorsque la polymérase atteint cette sonde, elle la détruit et sépare donc le reporter du Quencher, ce qui permet l'émission de fluorescence par le reporter. La polymérase possède une action exonucléasique 5'-3', spécifique.

RT-PCR : synthèse d'ADNc+PCR.

Quand on a de l'ARN et non de l'ADN : on utilise la transcriptase inverse (ou reverse transcriptase) sur un fragment d'ARN. Elle a besoin d'une amorce 3'OH libre. C'est une ADN polymérase ARN dépendante issue de rétrovirus (synthétise de l'ADN à partir de l'ARN). Pour l'utiliser il faut une amorce, possédant une extrémité 3'OH-libre, qui sera de type poly-T pour s'hybrider à la partie poly-A chez les eucaryotes (parfois on utilisera des amorces aléatoires pour les procaryotes). On peut ensuite utiliser la méthode de la PCR sur le fragment d'ADN complémentaire obtenu et travailler ainsi sur l'expression des gènes. On peut faire aussi de la PCR quantitative avec ça.



2. Digestion enzymatique de l'ADN

Cette étape est réalisée grâce à des endonucléases spécifiques de type enzyme de restriction. Ces enzymes de restriction sont issues des bactéries et archéobactéries. Elles ont été découvertes en utilisant des bactériophages sur différentes souches de bactéries (E.Coli) dont certaines étaient capable de couper l'ADN des bactériophages (= virus des bactéries). Le bactériophage envahit la bactérie et s'y multiplie. C'est ce qu'on appelle le système de restriction.

L'ADN du bactériophage B ne se reproduit pas dans la souche K (autre souche d'E.Coli). Après multiples passages, le bactériophage réussit à se reproduire. Ce système fonctionne grâce à une méthylation, la bactérie méthyle son ADN qui n'est pas coupé par le système de restriction, il est protégé. Le système de restriction empêche la bactérie d'être infectée par des virus (ou tout ADN étranger). Le bactériophage utilise le même système pour s'adapter et infecter la souche K.

Il y a 2 parties dans ce système: une partie qui méthyle l'ADN (l'ADN méthylé est protégé des enzymes de restriction). La bactérie va méthyler son propre ADN qui va faire que cet ADN méthylé est protégé des enzymes de restriction de la bactérie produit : des enzymes capables de dégrader l'ADN étranger mais pas méthylé. Au départ le bactériophage n'a pas le profil de méthylation de la bactérie donc il est dégradé puis petit à petit quand il commence à se reproduire dans la bactérie, progressivement son ADN va être modifié et sera donc protégé de l'enzyme de restriction et pourra ensuite se multiplier. Dans ce système il y a une enzyme de modification et l'enzyme de restriction.

On trouve 4 types connus d'enzymes de restriction. Les différences entre les types se situent au niveau du site de reconnaissance et du site de coupure et la nécessité ou non d'ATP.

- Le type 1 reconnaît un site asymétrique et l'enzyme coupe ailleurs, au moins 1000 paires de base (pb) plus loin. Elles ont un site de reconnaissance mais la coupure se fait très à distance de ce site de reconnaissance.
- Les types 2 sont les plus utilisées. ne sont pas capables de méthyler seuls, ont un site de reconnaissance spécifique assez court et coupent dans le site ou juste à côté, 2 sites sont séparés. Le site de clivage spécifique est le même que le site de reconnaissance. Ne nécessite pas d'ATP.
- Les types 3 reconnaissent une séquence asymétrique et coupent 24 à 26 pb après le site de reconnu.
- Les types 4 coupent l'ADN modifié (souvent méthylé).
- Type 5 : Il est constitué par des endonucléases à ARN guide : Cas9 du système CRISPR-Cas (permet de générer des mutations beaucoup plus facilement qu'apparemment par clonage). Elles ont besoin d'un guide ARN qui reconnaît une séquence cible pour avoir coupure, ce qui les rend particulières.

Les plus intéressantes et les plus utilisées sont les types 2 dites « classiques ». Elles présentent une séquence nucléotidique de reconnaissance souvent continue et palindromique (symétrie de lecture). Elles effectuent des coupures à bouts francs (blunt) ou cohésifs (5' ou 3').

Il existe des sous-types 2 non classiques :

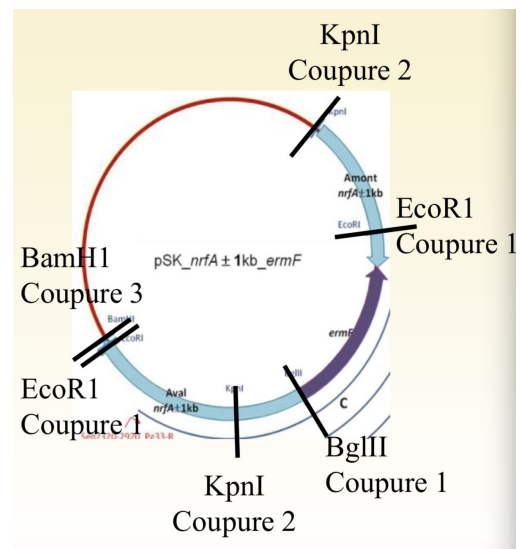
- Séquence nucléotidique de reconnaissance peut être discontinue, asymétrique.
- Coupure possible hors du site de reconnaissance. (pas plus loin de 20pb)
- Activité méthylase possible.
- Certaines ont jusqu'à 4 sites de coupure.
- enzymes qui ont besoin d'une séquence activatrice de la coupure.

Applications :

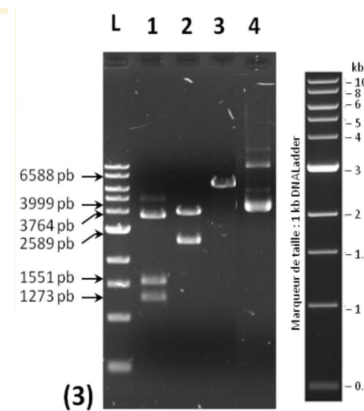
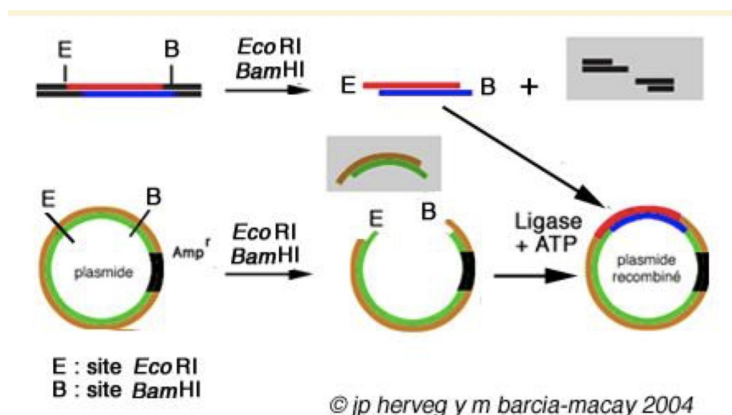
- Clonage
- Recherche de mutations
- Empreintes génétiques

Exemple d'application : le clonage.

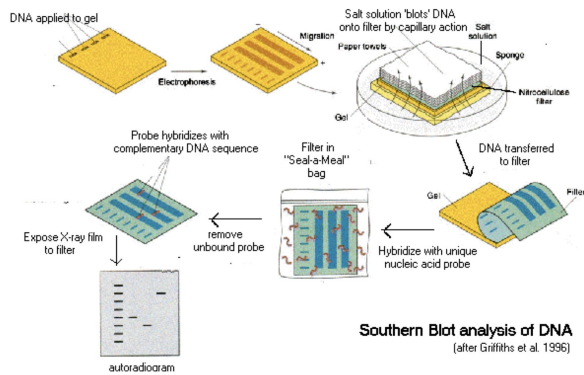
On va couper l'ADN de séquence connue à insérer dans un plasmide avec séquence voulue grâce à des enzymes de restriction (ici *EcoRI* et *BamHI*). Dans un même temps on coupe le plasmide avec ces enzymes. Ensuite on utilise une ligase qui va permettre l'incorporation du fragment à l'intérieur du plasmide. Ensuite, on va pouvoir utiliser ce plasmide recombiné pour réaliser le clonage ou pour transformer des bactéries.



Ici on peut voir un résultat de clonage. On a inséré toute la partie bleue et violette dans le plasmide. Les enzymes de restriction ont permis de vérifier la bonne taille et le fait que les différentes parties soient bien insérées au bon endroit, si on a bien la taille attendue ... Le résultat est présenté sur le gel juste à côté, avec toujours un marqueur de taille pour avoir une référence de taille et, différentes pistes montrant les différents fragments du plasmide non digéré : on voit 2 formes, surenroulé et long. Sur la piste 3, on réalise une simple coupure avec *BamHI*. Quand on fait une simple coupure on linéarise le plasmide, on a donc une seule bande sur le gel qui va permettre de déterminer la taille précise du plasmide : on va générer 2 bandes. Sur la piste 1, le plasmide a été digéré par les enzymes *EcoRI* et *BglIII* : il y a 3 sites de coupure et donc 3 fragments. Sur Piste 4, on a plusieurs bandes du plasmide. Plus il est enroulé, plus il va migrer vite.



3. Hybridation de l'ADN



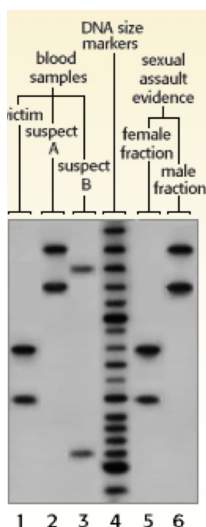
Techniques qui consistent à réaliser des hybridations par complémentarité des bases. On a donc besoin d'ADN monocaténaire.

Southern Blot (méthode d'hybridation couplée à une digestion de l'ADN) :

qui permet de détecter des fragments particuliers de l'ADN.

On dépose l'ADN sur un gel d'électrophorèse et on le fait migrer. Ce gel va ensuite être utilisé pour faire un transfert sur une membrane de cellulose (blot) : on a alors un transfert de l'ADN du gel à la membrane. On hybride ensuite cette membrane avec des sondes particulières d'acides nucléiques dont on connaît la séquence qui sont marquées soit de manière fluorescente, soit de manière radioactive. On fait un lavage pour enlever tout ceux non-utilisés. On peut alors faire une radiographie pour voir la position des morceaux hybridés spécifiquement avec la sonde. Il ne faut pas oublier l'étape de dénaturation après l'électrophorèse par la chaleur ou par des agents dénaturants.

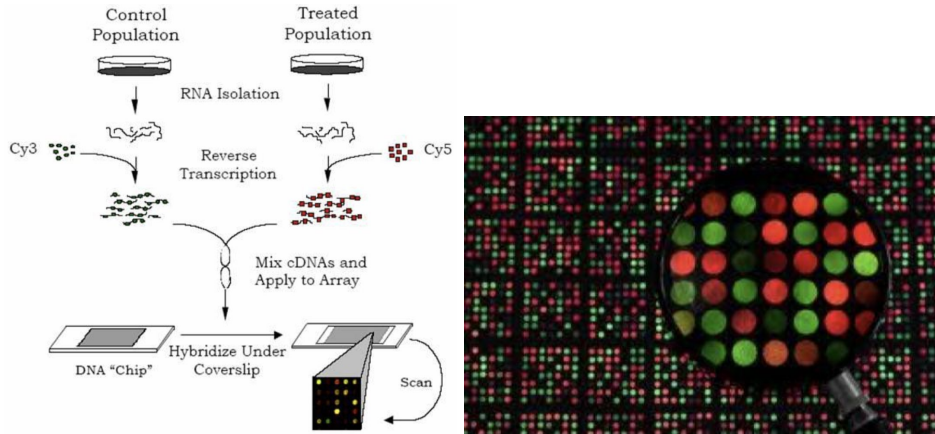
Cette méthode peut être couplée avec une digestion enzymatique par des enzymes de restriction. Cela permet de connaître le profil d'ADN. Avec la RFLP (Restriction Fragment Length Polymorphisme, ou Polymorphisme de Longueur des Fragments de Restriction pour le groupe X). Chaque individu possède un profil de restriction qui lui est propre. En réalisant des coupures sur certains fragments d'ADN on va pouvoir réaliser un profil type de l'individu.



Exemple : on peut identifier un suspect grâce aux résultats ci-contre. A gauche, ce sont les échantillons sanguins de la victime et de deux suspects. On a coupé l'ADN avec la même enzyme de restriction, et on s'aperçoit bien que les deux profils sont différents. On réalise avec une sonde, comme dans Southern Blot, une hybridation pour obtenir la révélation suivante. Donc en faisant des prélèvements sur la scène du crime ou sur la victime, on va voir que la coupure par des enzymes de restriction va définir un profil de restriction qui est celui du suspect A.

On peut aussi utiliser le couplage du southern blot et de la digestion des enzymes de restriction pour la recherche de paternité.

Une autre méthode d'hybridation est la miniaturisation sur des **puces à ADN** (micro-array). On utilise également les propriétés d'hybridation de l'ADN mais cette fois-ci on est sur des lames. On fixe les sondes sur des lames : on va avoir énormément d'échantillons analysables sur une seule lame. L'intérêt est qu'on peut cibler un nombre très important de gènes, sur une puce miniaturiste en une seule hybridation.



Par exemple on va comparer l'expression d'un gène entre une population traitée avec une molécule et une population contrôle. On extrait l'ARN des deux populations, on va faire une transcription reverse pour générer l'ADN complémentaire. Couplé à cette étape on ajoute des marqueurs fluorescents : vert pour les témoins, rouge pour les traités. On va mélanger les ADN complémentaires issus de la transcription reverse, puis on les met sur la puce et on réalise une hybridation. Tous les ADN avec la sonde hybridée sont visibles grâce à la fluorescence. En fonction de la couleur, on va savoir si c'est plutôt issu de la population témoin, ou issu de la population traitée. Si les deux populations expriment le gène considéré, cela va donner une couleur plutôt jaune. Ensuite on va pouvoir quantifier cette valeur et déterminer des ratios sur les populations.

On peut observer l'expression différentielle des gènes qui sont exprimés dans l'une et l'autre population. On obtient donc un grand nombre d'informations.

Par rapport à une analyse par PCR en temps réel, on amplifie toute une panoplie de gènes choisis alors qu'en PCR on amplifie un gène particulier. Au final on obtient le même type d'informations.

4.Séquençage de l'ADN

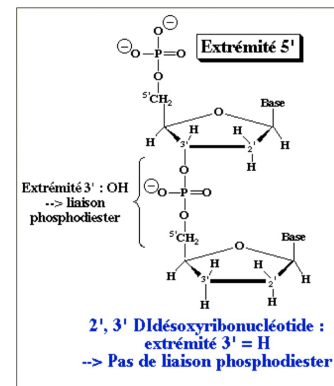
Le séquençage représente la majorité des techniques utilisées en recherche et au laboratoire.

Le principe est basé sur les propriétés de polymérisation. On travaille sur un seul brin, une seule amorce. La différence avec la poly classique

On utilise des désoxyribonucléotides modifiés. Ce sont des didésoxyribonucléotides, qui ne possèdent pas de fonction OH en 2' ni en 3' et bloquent car pas de possibilité de création de liaison phosphodiester alors la **réaction de polymérisation**. La principale différence avec une polymérisation classique est qu'on utilise des didésoxyribonucléotides.

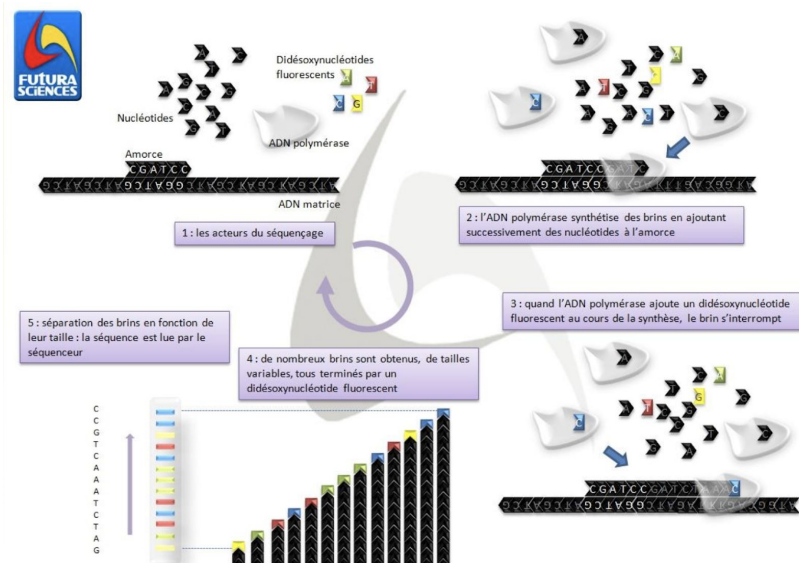
→ Technique de Sanger, prix nobel

On utilise à la fois des désoxyribonucléotides et des didésoxyribonucléotides. On a alors un arrêt de la polymérisation aléatoire (car on ne peut plus avoir d'accrochage de bases sur la liaison phosphodiester) et des fragments d'ADN de différentes tailles. On fait ensuite migrer ces fragments sur un gel par électrophorèse et on peut lire la séquence grâce aux différentes bases modifiées qui auront été incorporées. Contrairement à la PCR ici, on a une polymérisation que sur un brin.



La méthode a été améliorée et aujourd'hui, on marque chacun des 4 nucléotides pouvant être incorporés avec un fluorophore différent. On peut ainsi identifier la dernière base grâce à la couleur du fragment. La technique est automatisée en utilisant Dye terminator sequencing.

Cette méthode permet de mettre tous les didésoxyribonucléotides en même temps dans le même tube au lieu de faire un tube par type de bases comme dans la technique d'origine. On a incorporation aléatoire d'un desoxy ou didésoxy, jusqu'à l'obtention d'un dye terminator.



didésoxy: ça s'arrête

désoxy: ça continue

On fait alors une électrophorèse capillaire couplée à une détection des fluorochromes, ainsi on obtient des signaux attribués à chaque base.

On obtient un **chromatogramme** sous forme de pics de couleurs qui permet de vérifier la qualité du séquençage effectué par le séquenceur. On fait sa sur capillaire notamment pour miniaturiser l'expérience. Attribution en fonction de la couleur de la base associée.

Dans ce type de séquençage, on peut séquencer jusqu'à 1000 nucléotides par réaction.

Depuis quelques années, on peut faire du **séquençage haut débit** ou NGS (next generation sequencing). Il permet de séquencer des quantités d'ADN immenses en un temps record, qui seront traitées par des méthodes bio-informatiques.

Il existe plusieurs méthodes mais les étapes communes (principalement de préparation) aux différentes technologies sont :

- Fragmentation enzymatique de l'ADN, on le coupe aléatoirement pour avoir des tailles exploitables.
- Préparation d'une banque d'ADN par ligation d'adaptateurs (aux fragments) qui sert de base pour l'amplification ensuite.
- Amplification clonale : étape de polymérisation.
- Séquençage générant des signaux
- Détection des signaux émis et conversion en séquence

Parmi ces méthodes il y a :

- Séquençage haut débit NGS terminateurs réversibles : ici une fois que le fluorochrome du dernier nucléotide ajouté est détecté, on lui enlève la fonction terminateur (et son fluorochrome) et il peut ensuite être réutilisé pour continuer la polymérisation.

Ce qu'on supprime par rapport à l'autre méthode c'est la terminaison finale, ce qui permet de gagner beaucoup de temps. La synthèse et la détection sont simultanées.

- La méthode Illumina, basée sur le principe de Sanger, fait appel à des terminateurs réversibles. C'est-à-dire que par une réaction dans l'appareil, on peut détacher le terminateur et ainsi la réaction va continuer. On a incorporation des nucléotides fluorescents, détection du signal en fonction ensuite clivage de la partie terminator puis lavage et la réaction peut continuer, la taille de l'amplification augmente au fur et à mesure. Permet d'avoir de plus en plus de rides (courtes séquences par cellules de lectures) avec des tailles de lecture qui peuvent varier. On arrive à poursuivre la polymérisation plutôt que s'arrêter au terminator.

L'étape commune est la préparation de l'échantillon puis ensuite on ajoute des adaptateurs qui permettent l'attachement de tous les fragments, ensuite polymérisation puis PCR. Le fragment va s'attacher de l'autre côté de la puce pour réaliser un pont, puis dissociation. Ce qui génère des clones ou l'obtention des mêmes fragments avec un nombre élevé de copies.

Grâce à ce séquençage, on a remarqué qu'il y avait beaucoup de séquences de régulations. Sur les prochaines méthodes on étudie sur des micro-billes et il n'y a pas de fluorescence, on se base sur la détection du phosphate inorganique et de l'ion hydrogène.

a. Pyroséquençage

Il va y avoir émission d'un phosphate inorganique lorsqu'un nucléotide va être incorporé. Le phosphate inorganique sert de co facteur). Ce phosphate inorganique, on va le convertir en ATP (par l'intermédiaire d'une enzyme) qui va servir dans une réaction générant de la **luminescence** qui est la transformation de la luciferin par la luciférase (En fait dès qu'on génère un ATP, on peut transformer la luciferin en oxyluciferin et on va avoir une génération de lumière). Il y a autant de lumière que de phosphates inorganiques émis.

Quand il y a incorporation séquentiellement d'un nucléotide, il y a de la lumière (cela se traduit par l'apparition d'un pic d'intensité de lumière) on a donc émission d'un Pi, qui va générer un ATP ce qui génère une augmentation de l'intensité de lumière par la réaction de luminescence. Il faut alors laver à chaque fois qu'un nucléotide est incorporé (réaction séquentielle, on donne dans l'appareil des nucléotides séquentiellement).

b. Détecter le proton H⁺ (qui est libéré lors de la réaction) : ion torrent ou séquençage par semi-conducteur

Le H⁺ va servir à détecter l'incorporation de nucléotides par une mesure du **changement du pH** local. De la même manière que la technique d'avant, on va introduire **séquentiellement** les nucléotides et quand un nucléotide est incorporé il y a un signal qui est détecté parce que le H⁺ est détecté localement (le H⁺ a été émis) par une variation de pH qui est transformé en courant. A noter que s'il y a deux fois le même nucléotide, le signal sera deux fois plus important.

c. Une 3ème génération qui arrive

Ce sont des techniques qui sont en train d'arriver et qui ne sont, pour l'instant, pas aussi fidèles que celles décrites précédemment, elles font encore un certain nombre d'erreurs mais elles ont l'avantage, notamment pour la technique « nanopore », d'être miniaturisées et transportables étant donné que l'outil est petit. Souvent, ce sont des techniques qui ne font pas appel à la polymérisation. Par exemple, pour la technique « nanopore », la chaîne d'ADN passe à travers des nanopores, la chaîne qui est passée bloque le pore ce qui génère une diminution de courant. Cet abaissement de courant est caractéristique d'un nucléotide donné (selon la force de courant qu'il reste, on sait quel est le nucléotide bloquant le nanopore).

Le coût pour le séquençage d'un génome a nettement diminué (notamment depuis les années 2005-2006). Quelques données sur le génome humain : la séquence d'ADN codant est très réduite en comparaison à celle de l'ADN non codant.

PacBio: on excite avec une source de lumière pour détecter une fluorescence qui correspond à un nucléotide incorporé

Oxford Nanopore: pas d'étape de polymérisation. L'ADN traverse un nanopore. Et en fonction du nucléotide qui a traversé le nanopore, on a une signature électrique. Pas d'étape de synthèse.

Quelques données sur le génome humain

Régions transcrites en ARN : 0,05%	<ul style="list-style-type: none"> •Tailles moyennes : Gènes : 45 kb •Séquence codante : 1500 nucléotides <ul style="list-style-type: none"> •Exon : 145 nucléotides •Intron : 5200 nucléotides •5' UTR : 210 nucléotides •3' UTR : 740 nucléotides •Introns - exons • Nombre d'introns : 6 ± 3 introns / 1000 pb séquence codante • Introns / [introns + séquence codante] : 92% • Epissage alternatif pour plus de 90% des gènes
Régions codant pour des protéines : 1,2%	
Introns : 31%	
ADN intergénique : 61%	
ADN satellite : 6 - 7%	
Pseudogènes : 1 - 1,2%	
Eléments transposables : 42 - 46 %	
Source : Duret, L. (2011) "Bioinformatique: Annotation des génomes (eucaryotes)" Université Lyon 1	