

UE 11 : Principes de biologie Moléculaire et applications en génétique médicale – Fiche 2

RAPPEL PLAN

Cours 1 :

I- Analyse du matériel génétique

1- Extraction de l'ADN

2- Extraction de l'ARN

3- PCR

4- Digestion enzymatique

II- Biologie moléculaire et génétique médicale

L'achondroplasie

Cours 2 :

III. Séquençage : méthode Sanger, et séquenceurs automatiques

A. Principe

B. Méthode Sanger

C. Méthode Automatique

D. Exemple et lecture de résultat

IV. Applications en génétique médicale : le syndrome de Wolfram

A. 1ère étape : Exploration de WFS1

B. Séquençage de l'ARNm

C. Clonage Moléculaire

D. Transformation : introduction du vecteur dans une bactérie

E. Sélection, Isolement, amplification, extraction

F. Carte de Restriction

RAPPEL : primers = amorces = oligonucléotides

III. Séquençage : méthode Sanger, et séquenceurs automatiques

On ne pose pas de diagnostic avec UNE seule technique de biologie moléculaire (+)

Le séquençage détermine la succession de nucléotides d'une séquence d'intérêt.

Il nous permet de lire la séquence d'ADN et de connaître la position de chaque nucléotide.

On peut **identifier** une mutation et **diagnostiquer** une éventuelle pathologie en séquençant directement la séquence du gène concerné.

Rappel : les deux brins d'ADN sont constitués de dNTPs reliés à leur nucléotide complémentaire sur l'autre brin par des liaisons hydrogène, et relié entre eux sur un même brin par des liaisons de type phosphodiester.

A. Principe

Similarité à la PCR :

On veut partir d'ADN double brin (nos produits PCR) et obtenir un ADN simple brin, hybrider des nucléotides et faire de la polymérisation en jouant avec différentes températures. Les **étapes** du séquençage sont les **mêmes que celles de la PCR**.

On retrouve une succession de 30 à 35 cycles successifs de 3 étapes chacun :

1. Dénaturation (95°C) : l'ADN double brin est déstabilisé et rendu **simple brin** par **rupture des liaisons hydrogènes** sous l'effet de la chaleur.
2. Hybridation (50°C) : le **primer** s'hybride par complémentarité au brin. On pose **une seule amorce** (15-20nt) (car on veut synthétiser un seul brin) complémentaire d'une région connue de la séquence qui nous intéresse
3. Elongation (60°C) : l'ADN polymérase (Taq Polymérase) synthétise le brin complémentaire dans le sens 5'→3'. Elle utilise des **dNTPs** et des **ddNTPS couplés à un traceur**.

Ainsi, à la différence de la PCR :

-Le séquençage ne nécessite qu'une seule amorce/primer à partir duquel l'ADN polymérase synthétise un brin complémentaire fidèle à la séquence d'ADN à étudier, dans le sens 5'-3' !

-Utilisation de dNTPS et de ddNTPs

L'ajout d'un **ddNTP stop la synthèse** du brin fils, pourquoi ??

Ribose : O sur C2 et C3

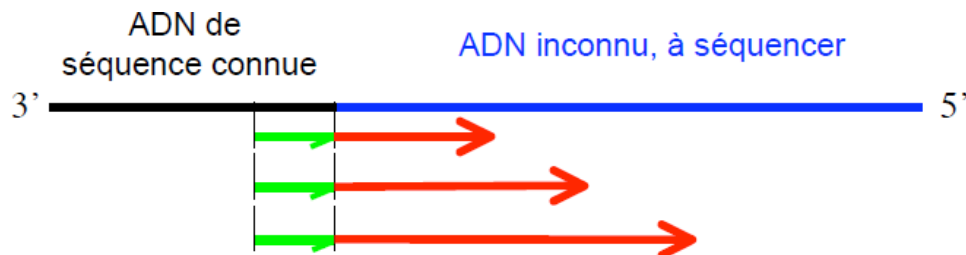
2-désoxyribose : O sur C3

Didésoxyribose : pas de O ni sur C2 ni sur C3

->Il n'y a **plus d'extrémité 3'OH disponible** la réaction s'arrête

De manière aléatoire la Taq va incorporer un dNTP ou un ddNTP pour synthétiser le brin complémentaire. Quand elle incorpore un dNTP elle pourra continuer à former des liaisons phosphodiester/covalentes par la suite si c'est un ddNTP la synthèse s'arrête.

Au bout de n cycle on a plusieurs fragments de différentes tailles qui correspondent à toutes les fois où la Taq s'est arrêtée.



On aura ainsi des brins ayant incorporé un ddNTP en tout premier et qui ne seront donc composés que d'un seul nucléotide, d'autre qui seront composés de deux

nucléotides, d'autres de trois, d'autres encore de quatre... Et ainsi de suite jusqu'à ce que l'on obtienne des fragments ayant été synthétisés entièrement et n'ayant incorporé un ddNTP qu'à la toute fin de la synthèse, pour le tout dernier nucléotide de la séquence !

B. Méthode Sanger

Nous allons voir le principe de fonctionnement de **LA méthode de référence** : la méthode de Sanger ou méthode des didésoxyribonucléotides.

Historiquement, on pouvait réaliser manuellement ce type d'expérience, qui était relativement long et peu précis par rapport au NGS actuel cependant le principe reste le même.

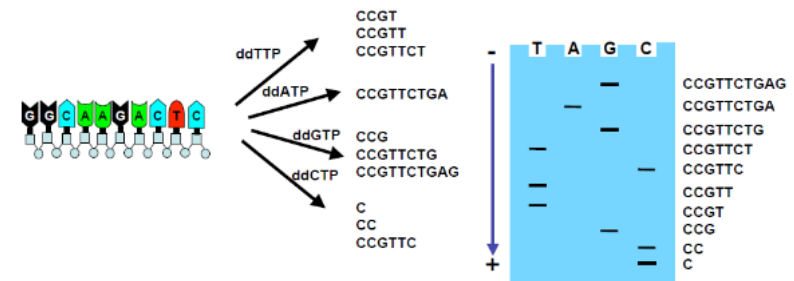
a) Principe

On faisait à l'origine quatre réactions **indépendantes** dans quatre **tubes différents** contenant chacun les 4 types de dNTPs et **un seul ddNTP radiomarqué** (A, T, C ou G). On mettait donc dans chaque tube :

- le fragment d'ADN du patient à séquencer,
- une seule amorce,
- la Taq polymérase
- les 4 types de dNTPs
- et un seul type de ddNTP radiomarqué.

Remarque : il fallait faire attention à ne pas confondre les tubes !

Dans chacun des quatre tubes, on répétait plusieurs fois les 3 étapes et on obtenait à la fin un mélange de produits générés de différentes tailles.



Ex : Si dans le 1er tube on met un ddNTP correspondant à une Thymine, tous les fragments de ce tube finiront par une thymine radiomarquée, puisque la seule façon de bloquer la synthèse est d'incorporer un ddTTP, les fragments se termineront donc obligatoirement par une thymine !

Chaque tube contient un seul type de ddNTP (A, T, C OU G) ET les quatre types de dNTP (A, T, C ET G)

b) Migration électrophorétique

On mettait ensuite ces quatre tubes à **migrer sur un gel d'acrylamide**, pour **séparer** les produits synthétisés en fonction de leur **taille** par **migration électrophorétique**. Grâce au champ électrique, l'ADN étant chargé $-$, les fragments **migrent du $-$ vers le $+$** . Les fragments les plus **gros** et donc les plus **lourds** restent **en haut** du gel et **migrent donc le moins loin**.

Pour **déterminer la séquence** du fragment, il faut lire le gel **de bas en haut**, c'est-à-dire du plus **petit** au plus **grand** des fragments !

- **Du plus petit** : les fragments ont incorporé un ddNTP rapidement, ils ont donc stoppé leur synthèse plus tôt, sont plus légers et migrent loin.
- **Au plus grand** : les fragments ont incorporé un ddNTP plus tardivement, ils ont donc stoppé leur synthèse plus tard, sont plus lourds et migrent moins loin.

On part donc du bas, puis pour **connaître l'identité du nucléotide** il suffit de **regarder** dans quelle **piste** est notre **fragment**. Sur le schéma ci-dessus on a un C puis un G etc. Sachant que la Taq synthétise de 5' en 3' les premiers nucléotides ajoutés (en bas du gel) se trouvent donc en 5'. On lit notre **gel de bas en haut et de 5' en 3'**.

En lisant le gel jusqu'en haut on se retrouve avec un brin complet ici on a :

5'- CCGTTCTGAG-3'

Or depuis le début on veut connaître l'enchaînement de nucléotides (=séquencer)

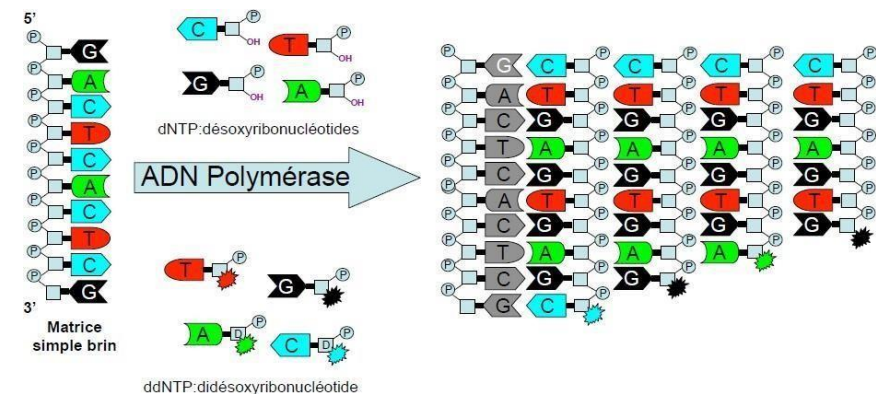
d'un brin que l'on ne connaissait pas. Avec Sanger on a synthétisé le brin complémentaire du brin qui nous intéresse.

Pour trouver la séquence qui nous intéresse il suffit de **faire le complémentaire** du brin lu sur le gel, ici :

5'- CCGTTCTGAG-3' -> 3'-GGCAAGACTC-5' (Attention les brins sont antiparallèles on n'oublie pas que les sens 5' et 3' sont inversés)

C. Méthode automatique

Avec les **séquenceurs automatiques**, on ne fait plus qu'**une seule réaction** dans 1 même tube au lieu d'avoir quatre tubes différents. Les **quatre ddNTPs** vont pouvoir être **mélangés** dans le même tube réactionnel car chacun est couplé à un **fluorochrome de couleur différente** pour les identifier. D'après la nomenclature, la **Thymine** est en **rouge**, la **Guanine** en **noir**, la **Cytosine** en **bleu** et l'**Adénine** en **vert** ! (ne pas retenir les couleurs)



Les fragments d'ADN de tailles différentes générés sont tous marqués d'un fluorochrome différent selon le ddNTP incorporé par la polymérase.

Ensuite le principe est le même : ces produits sont séparés par migration électrophorétique mais dans un automate et non dans un gel.

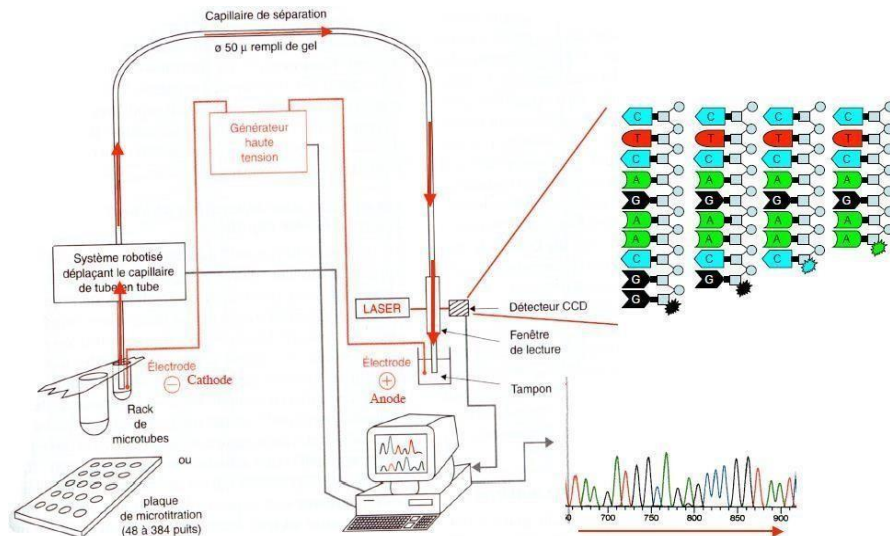
Les différences :

- Les puits deviennent des capillaires (=petits tubes). Un séquenceur à 25 capillaires peut faire migrer 25 échantillons différents (vitesse ++)
- Un polymère remplace l'agarose ou l'acrylamide

Les fragments séparés par leur taille passent devant une **caméra laser** qui va **détecter** et **lire** la couleur de chaque fragment, permettant ainsi d'**identifier le nucléotide** par lequel se termine la séquence.

Les données sont transférées de l'automate vers l'**ordinateur** qui produira un **électrophorégramme**, reconstituant le brin complémentaire séquencé selon l'ordre de détection par laser (donc selon la taille des fragments).

- ➔ On a donc une **séparation** en fonction de la **couleur** qui nous donne l'**identité du nucléotide**, et une séparation en fonction de la **taille** ce qui nous donne l'**enchaînement** des nucléotides. (+++)



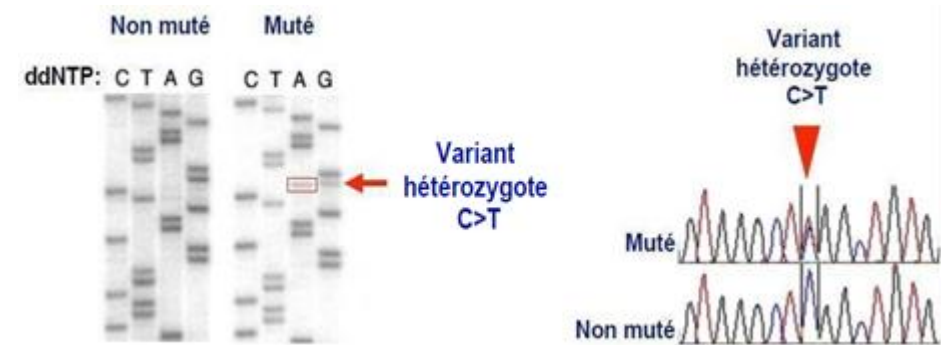
(N'apprenez pas le schéma c'est juste pour donner une idée du fonctionnement)

A savoir que la Méthode Sanger qui utilisait la radioactivité (*danger pour*

l'opérateur), longue et peu efficace (lecture de 100 à 200 pb pour 2 /3 jours de travail) était tout de même très performante jusque dans les années 90'. Néanmoins les meilleurs automates peuvent en lire plus de 300 000 en une réaction.

Les séquenceurs automatiques sont aujourd'hui LA technique de référence dans les labos en termes de séquençage même s'il y a de nouvelles techniques que nous allons voir comme la NGS qui ne sont pas la référence (Attention piège ++ !!)

D. Exemple et lecture de Résultats



En séquençage Sanger (à gauche) :

On lit le gel de bas en haut : 5'-ACTTCTT...-3' pour l'électrophorèse de gauche (non muté). On remarque ici qu'au niveau de la flèche, sur le gel du **patient muté**, il y a **deux produits de migrations dans deux puits différents mais de même taille** (au même niveau sur le gel) : il y a un fragment où un « A » a été incorporé, et un autre où c'est un « G ».

Rappel : Il ne faut pas oublier que la séquence du brin lu sur le gel est complémentaire du brin séquencé du patient.

Si on lit un « A » sur le **gel**, on a un « T » dans la **séquence d'origine**. De la même façon, un « G » lu sur le **gel** correspond à un « C » dans le **brin séquencé** du patient.

Si on reprend notre exemple, la présence de deux fragments différents de même taille, l'un se terminant par un « A » et l'autre par un « G » témoigne d'un **variant hétérozygote C>T** ! (car C s'apparie avec G et T s'apparie avec A). Le patient a un allèle sain (avec le nucléotide C) et un allèle muté (substitution de C en T).

En séquençage automatique (à droite) :

On peut également voir ce variant hétérozygote : au niveau du triangle on remarque une **superposition** de **deux pics de couleur différente** au niveau de la **même position** nucléotidique : on met en évidence le même **mélange de « C » et de « T »**.

IV. Applications en génétique médicale : le syndrome de Wolfram

Rappel :

- Dans le noyau, les gènes sont transcrits en ARN (exons + introns)
- Les préARNm sont ensuite maturés en ARNm après épissage (exons)
- Les ARNm sont ensuite exportés dans le cytoplasme
- Dans le cytoplasme, les ARNm sont traduits en protéines

Il y'a des parties non codantes en amont (5'UTR), en aval (3'UTR) et les introns : on ne connaît pas leur fonction mais leur délétion entraîne des dégradations des ARNm (queue poly-A en 3'UTR qui stabilise les ARNm).

Même si on connaît la séquence entière du génome humain on ne connaît pas tous les gènes ni leurs fonctions.

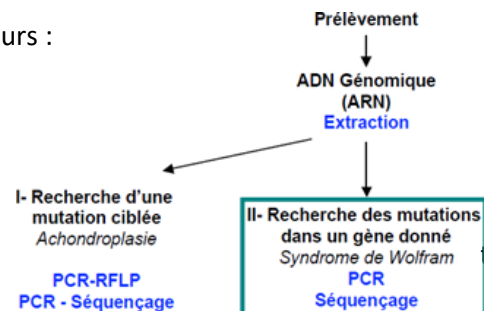
Savoir traduire un ARNm en protéine n'implique pas connaître la fonction de celle-ci.

➔ Difficile de savoir si un variant identifié est bien responsable de la pathologie (certains variant ne sont pas pathogènes)

Vous savez désormais comment détecter une mutation ciblée et connue : PCR-RFLP, puis confirmation avec un séquençage (ex : Achondroplasie)

Que faire dans la situation, plus compliquée, où l'on doit rechercher plusieurs mutations dans un même gène ? Ou plusieurs gènes comme les maladies multigéniques ? Ou encore quand le séquençage est ininterprétable ou ne donne pas assez d'informations, comme dans le cas du Syndrome de Wolfram ?

C'est l'objet de la suite du cours :



Pour poser un diagnostic, on ne séquence que les séquences codantes (=exome)

On ne peut pas interpréter les mutations des **régions non codantes** car, ne donnant normalement pas d'acide aminé, on ne pourra pas faire le lien entre modification de l'ADN et altération des protéines. Or, même si on ne les cherche pas, des **mutations** dans ces régions peuvent avoir des **effets délétères** par ex en modifiant les séquences consensus d'épissage, en créant de **nouveau site accepteur d'épissage (=site cryptique d'épissage AG)** ... Si le préARNm est épissé différemment, l'ARNm est différent et la protéine aussi : elle peut être défectueuse !

Maintenant, mettez-vous dans la peau d'un médecin généticien. Un patient enfant (qu'on appellera Jimmy) arrive dans votre cabinet, il vous a été adressé par son médecin généraliste qui suspecte, après examen clinique, un Syndrome de Wolfram. On ne peut pas se contenter de signes cliniques il va falloir faire des examens complémentaires, ici en génétique. On va se préoccuper en particulier de WFS1 (gène codant, ARNm, protéine finale) car on sait que ce Syndrome y est associé.

Syndrome de Wolfram :

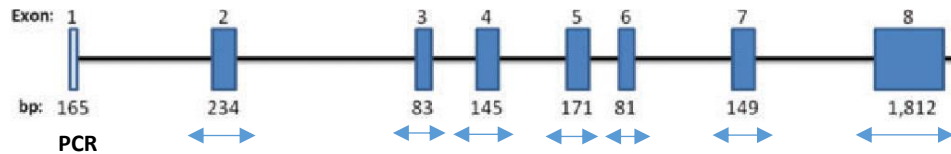
- Diabète, atrophie optique, surdité, troubles neurologiques** variables
- Pathologie **autosomique RECESSIVE ++**
- Gène responsable : **WFS1**. Petit gène de 8 exons, dont le 1^{er} est non codant (ATG se situe sur le 2^{ème})
- Code pour la **wolframine** protéine dont la fonction est inconnue. Elle aurait un rôle dans le flux calcique permettant la communication entre le Réticulum endoplasmique et la mitochondrie.

A. 1^{ère} étape : Exploration de WFS1

On réalise un séquençage sanguin en s'intéressant aux exons codants.

- 1) On va faire 7 amplifications par PCR des 7 exons codant (pas le 1 du coup) et de leurs jonctions avec les introns. On place nos primers de part et

d'autre des exons, on séquence l'exon et les 5 premiers/derniers nucléotides des introns : ce sont des sites d'épissage.

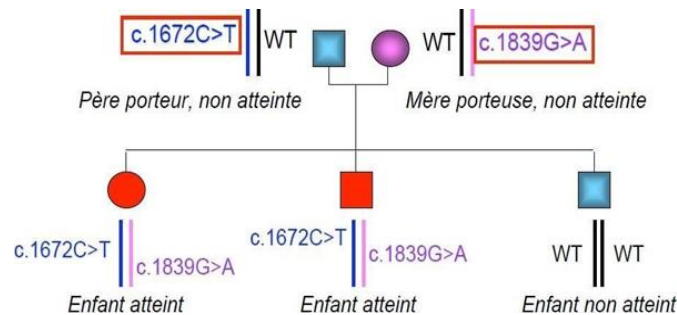


2) On vérifie ces 7 PCR sur gel et on contrôle les négatifs.

3) On les séquence avec la technique des didésoxyribonucléotides

a) Première situation : mutation dans l'ADN codant

On trouve à Jimmy un variant hétérozygote au niveau du 1672^{ème} nucléotide situé sur une séquence codante (mutation : substitution C->T notée c.1672 C>T) et sur l'autre allèle un variant c.1832 G>A



Explication : Le père et la mère sont **hétérozygotes muté donc porteurs sains** (rappel : maladie récessive ++) mais n'ont pas le même variant (la même mutation). Ici, l'enfant a hérité du variant 1672 du père, 1832 de la mère. Il n'est pas homozygote pour la même mutation mais malade quand même parce que ses 2 allèles codent pour une wolframine défectueuse. C'est un cas **d'hétérozygotie composite**.

- ✓ Il a tous les signes cliniques
- ✓ La littérature (=articles, études, docs scientifiques sur le sujet) possède les preuves de la pathogénicité de ces variant : ces mutations sur ces

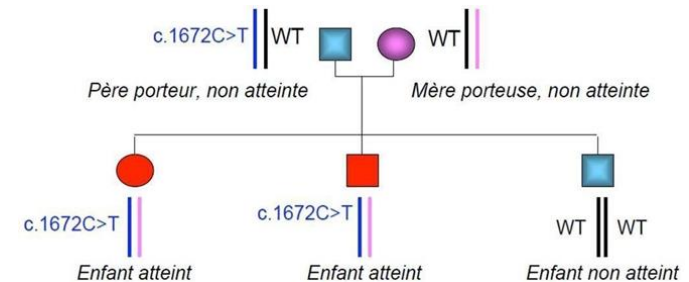
nucléotides sont déjà connus dans d'autre cas de patients. Et on avait déjà fait des tests montrant l'impact de cette mutation sur la protéine finale (cf. vecteur d'expression plus loin dans le cours)

- ✓ On a pu justifier la transmission de la mutation (on a trouvé l'allèle paternelle défectueux en séquençant le génome paternelle, pareil pour l'allèle maternelle)

➔ On peut poser le diagnostic

b) Deuxième situation : mutation dans l'ADN non codant

C'est moins « facile », si l'enfant a tous les **signes cliniques** mais après séquençage des 7 exons on a identifié le **variant du père**, on a un allèle muté c.1672C>T mais rien sur l'**allèle maternel**, l'allèle **semble être Wild Type** (=non muté=sauvage). Hors si elle c'était le cas l'enfant devrait être porteur sain (car c'est autosomique récessif la prof insiste) mais ce n'est pas le cas.

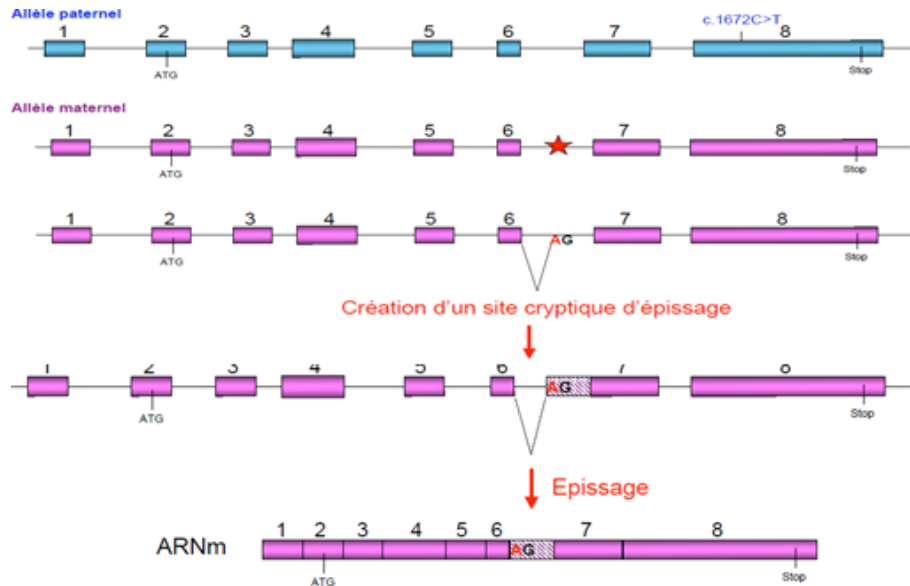


Hypothèses d'explication possibles :

Si nous n'avons pas trouvé de mutation dans les séquences codantes il y'en a peut-être dans les séquences non codantes, les introns, que nous n'avons pas séquencés.

- ❖ Il y'a eu une **modification au niveau des sites d'épissage** entrainant la transcription d'un ARNm défailant donc une protéine défailante. On retombe sur une sorte d'hétérozygotie composite (2 mutations différentes mais pathologique au final).

Exemple : il suffit qu'il y ait un AG qui se crée par mutation, substitution, en plein milieu de l'intron ça va générer un **site cryptique d'épissage**. L'épissage ne se fera plus de la fin de l'exon 6 au début du 7 mais entre la fin du 6 et cet AG en plein milieu de l'intron : on gardera le bout entre le AG et l'exon 7, bout d'intron normalement non codant. Ce **bout d'intron** serait **traduit** et engendrerai une **protéine anormale**.



- ❖ Une **mutation** peut aussi faire « **sauter** » un **exon**. Par exemple on peut avoir un épissage de la fin de l'exon 6 au début de l'exon 8 (l'exon 7 est éjecté) ou même parfois du 4 au 7.

Ces zones codantes sont très peu prévisibles, on interprète mal leurs variations. De plus on a une base de données de cas mais les cas possibles sont trop nombreux.

Comment faire dans ce cas-là ?

On ne peut pas donner de diagnostic sans avoir clairement identifier les mutations chez le père ET chez la mère !

Important : Ici on a commencé par séquencer les exons : on fait à chaque fois cette étape car c'est :

- plus simple/rapide
- les exons sont plus courts que les introns

On va chercher plus loin. Nos hypothèses impliquent une modification génétique qui n'est pas analysable facilement au niveau de la séquence du gène. En effet, on pourrait séquencer tout le génome et non l'exome uniquement mais se serait **trop long** et même si on trouve un variant **on ne saurait pas ce qu'il implique sur la future protéine**, si on est « sur » qu'il soit pathogène. Par contre on peut observer une possible modification au niveau de l'ARNm reflet ++ de la protéine et donc meilleur reflet de l'activité cellulaire (la prof insiste++). On va donc séquencer l'ARNm.

B. Séquençage de l'ARNm

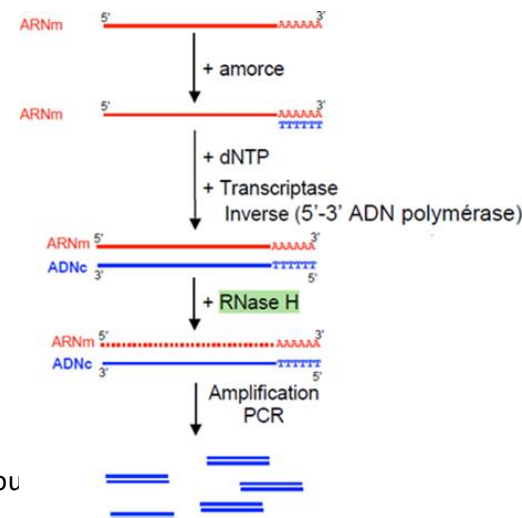
1) Obtention de l'ARNm (cf.cours 1)

Même principe que l'ADN : on lyse, on récupère les ARNm du cytoplasme dans du **phénol à pH acide** (attention ! différence avec l'ADN). On isole les ARNm en utilisant la complémentarité de leur queue poly-A, caractéristique du type messager, avec un support solide d'oligoDT.

2) Transcription inverse

La Taq-polymérase est une ADN polymérase, elle a besoin d'une matrice d'ADN pour fonctionner. On va donc vouloir transformer notre brin d'ARNm en brin d'ADN. Comment ?

On ajoute à notre ARNm, une **amorce** d'ADN (une queue oligoDT), des **dNTP** et la **TRANSCRIPTASE INVERSE** (=reverse transcriptase #coucouLesTéломèresde laBiomol 😊)



Transcriptase inverse :

5'-3' ADN polymérase **d'origine virale** qui synthétise un ADNc (ADN complémentaire) à partir d'une amorce d'ADN hybridée sur un ARNm.

NTB : Enzyme de restriction + Taq polymérase = origine bactérienne vs Transcriptase inverse = origine virale

On obtient un **hybride (=hétéroduplex) ARN/ADNc**. Enfin on ajoute des **RNase H** (enzyme dégradant l'ARN uniquement s'il est hybridé à un ADN).

3) Amplification par PCR

Maintenant qu'on a un brin d'ADN simple on amplifie directement par PCR (pas besoin d'étape de dénaturation on a déjà du simple brin). Ici on suspectait qu'un fragment de l'intron 6 (entre exon 6 et 7) avait été conservé. On peut mettre des primers partout et recommencer nos 7 PCR, mais quand on « sait » comme ici (peut-être par la littérature ?) on va mettre 2 primers : **un au milieu de l'exon 6 et un au milieu de l'exon 7**. La PCR amplifiera les exons **et ce qu'il y a entre les deux**.

NTB : On va rajouter après notre amorce une certaine séquence, reconnaissable par une enzyme, pour qu'elle soit répliquée et ajouté à nos amplicons (elle nous servira plus tard).

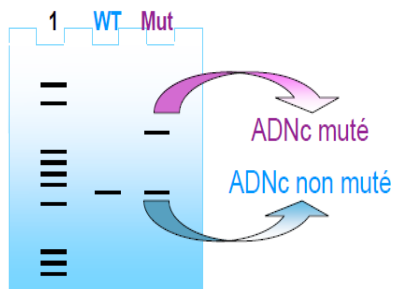
4) Vérification sur gel analytique

On compare la masse des deux **amplicons** (les 2 allèles de l'enfant : paternel et maternel) par électrophorèse. S'ils sont de **même taille** (même **nombre** de nucléotides et pas **type** de nucléotides : une substitution ne change pas significativement le poids du brin) ils migrent à la **même hauteur**. Ici, selon notre hypothèse, le variant d'épissage rajoutant une séquence codante, le produit PCR d'**ADNc maternel** devrait être **plus lourd** et donc **migrer moins loin** que l'ADNc du père.

NTB : on observe ça si le variant se trouve bien entre exon 6 et 7 si on n'observe rien on réessaie avec d'autres exons.

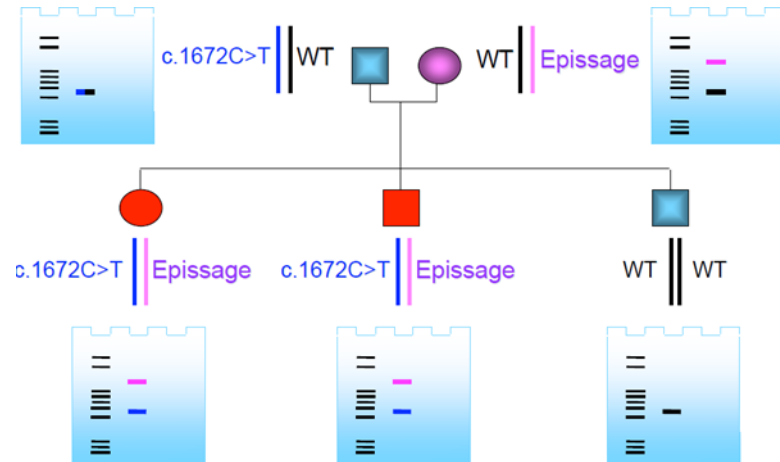
Résultat :

Analyse des produits PCR
après migration électrophorétique



Piste WT (Wild Type) : correspondrait à la situation du **père**. La **mutation ponctuelle** 1672C>T au niveau de son exon 8 ne **modifie pas la taille** du fragment PCR car il s'agit juste d'une **substitution** de base.

Allèle maternel (Mut) : correspondrait à la situation de la **mère** et des deux **enfants atteints** (le brin muté est plus lourd et migre moins loin).

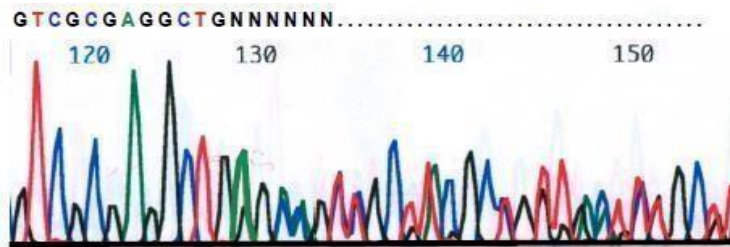


La migration électrophorétique nous permet de vérifier la présence d'un variant d'épissage qui rend l'ADNc plus lourd. Notre **hypothèse** semble donc être la bonne : il y'a bien un variant d'épissage entre les exons 6 et 7 de l'ADNc chez la mère. On va pouvoir faire une réaction de **séquence** des produits PCR obtenus précédemment pour **identifier ce variant d'épissage** !

Tout le raisonnement précédent était **théorique**, dans notre tête, on a **imaginé** que la mutation était **entre les exons 6 et 7** à titre d'**exemple** pour mieux visualiser les mécanismes impliqués, mais en réalité elle peut se trouver dans **n'importe quel intron** d'un patient à l'autre, d'où la nécessité de faire un séquençage ! On est **obligés** de trouver précisément **les deux mutations** (via le séquençage) pour avoir le droit de poser un diagnostic car le syndrome est de **transmission récessive** et implique que **deux allèles soient mutés** !

5) Séquençage et identification de la mutation

On n'a pas d'autre choix que de séquencer nos **deux allèles simultanément** car les produits PCR sont mélangés ! En effet chez les parents comme les enfants quand on extrait l'ARNm on obtient les deux allèles car ceux-ci sont ensemble en suspension dans le cytoplasme. Ici on aura l'allèle paternel et maternel (plus long) dans le même tube. On obtient cela si on séquence :



A partir de la jonction exon6 / 7 (~ position 130) la séquence devient illisible, on a une superposition de deux signaux.

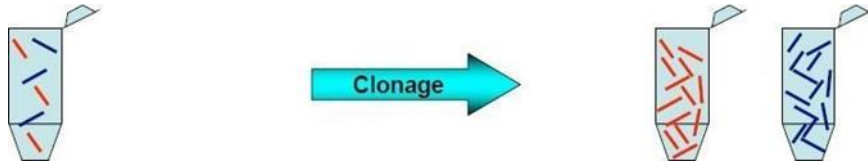
Ceci est dû au fait qu'on a un mélange de nos deux produits PCR de taille différente (à cause du variant d'épissage qui est plus long et qui décale la suite de la séquence). Ainsi, on a une superposition de la séquence de l'allèle paternel et de l'allèle maternel.

➔ Il faut **séparer les deux PCR** (1^{er} étape) puis les séquencer individuellement (2^{ème} étape). Pour cela on fait un **clonage moléculaire**.

C. Clonage Moléculaire

a) Principe

But = obtenir un **grand nombre de copies identiques** et **pures** d'une **séquence donnée d'ADN**



4étapes :

1. On crée un **ADN recombinant (=vecteur +insert)**.
2. On introduit ces vecteurs dans la bactérie = **transformation bactérienne**
3. On **cultive** les bactéries sur une boîte de Pétri (**amplification**) + **antibio (sélection)**
4. On « **repique** » : on prélève des bactéries de chaque colonie pure et on les met chacune dans des **tubes différents** (on a **isolé** les 2 populations) on les fait proliférer de nouveau

➔ On récupère des fragments purs en grande quantité

Pour ce faire, on va utiliser des **vecteurs** (ADN circulaire double brin) dans lesquels on va insérer à chaque fois **un seul fragment d'ADN** appelé **insert** (notre produit PCR d'origine paternelle **ou** maternelle)

b) Les vecteurs

VECTEUR

Il doit être :

- De **l'ADN circulaire double brin** / plasmidique
- De taille réduite pour avoir la place d'y insérer un fragment d'ADN étranger
- Avoir une **réplication épisomiale** (autonome indépendante du cycle du génome nucléaire de la cellule hôte)
- Posséder des **gènes de sélection** (généralement gène de résistance à l'Ampicilline = antibio), afin de trier les bactéries qui auront vraiment intégré le vecteur

2 types de vecteurs :

Vecteurs de clonage : pour isoler et amplifier un fragment d'ADN (ce qu'on utilise ici)

Vecteurs d'expression : pour transférer un gène dans une cellule eucaryote afin que la cellule exprime ces protéines, on pourra observer si la mutation qui influe la position d'une protéine dans la cellule, etc...

Le clonage moléculaire ne se fait qu'avec des cellules procaryotes (bactéries) !

Il y a **différents vecteurs de clonage selon la taille du fragment à cloner**, au-delà d'une certaine taille on n'utilise plus des bactéries mais des levures (YAC). Dans ce cours on s'intéressera uniquement au **PLASMIDE**.

Remarque : Les bactéries contiennent un unique chromosome libre dans le cytoplasme ET des petites molécules d'ADN circulaires appelées plasmides.

La séquence du plasmide est connue, ils sont commercialisés, on a :

- ✓ Un **polylinker (site multiple de clonage)** : courte **séquence connue** où il y a de nombreux **sites pour des enzymes de restriction** qui seront digérés afin d'introduire l'insert.
- ✓ Une **origine de réplication** : permet une réplication indépendante du chromosome de la cellule hôte
- ✓ Un **gène de sélection** : Il en existe plusieurs, mais le plus courant apporte une résistance à l'Ampicilline (antibio), donc si on étale les bactéries sur une boîte de Pétri avec de l'Ampicilline, seules celles avec le gène et donc le plasmide (donc intéressantes) vont se multiplier : on donne un avantage sélectif aux bactéries avec plasmide.

Plasmide = polylinker + origine de Réplication + gène de sélection

c) Création de l'ADN Recombinant

1. Linéarisation du plasmide, digestion du plasmide et de l'insert

On va essayer de digérer l'insert et le plasmide avec la même Enzyme de restriction. On la choisit en fonction de la manière dont on veut couper nos séquences il existe 3 **stratégies de clonage** : extrémités **franches**, **cohésives** ou **combinés**.

Digestion du plasmide :

- Création **d'extrémités cohésives** uniquement : la meilleure solution (ex : enzyme EcoRI) car les bases complémentaires entre l'insert et le plasmide pourront s'hybrider spontanément par liaisons hydrogènes

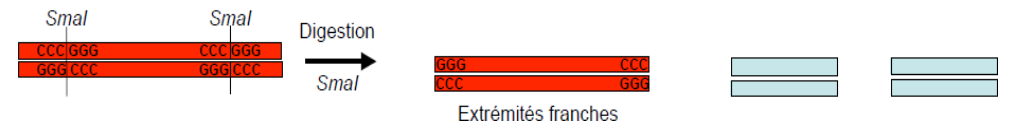


On a bien notre séquence palindromique, l'enzyme à pour travail de **reconnaître** son site de digestion et de **couper** entre le G et la séquence AATTC au niveau des **2 brins** (++) diffère avec une endonucléase.

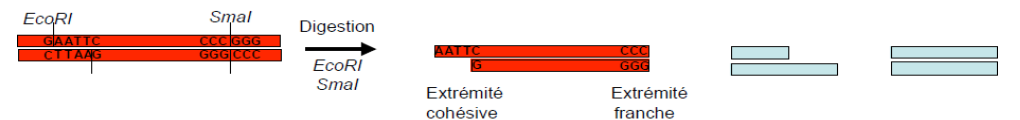


→ Ici on choisit cette solution

- **Extrémités franches** uniquement (ex : SmaI) -> pas top, car le vecteur à tendance à se refermer sur lui-même



- Les deux à la fois : **on peut les combiner**



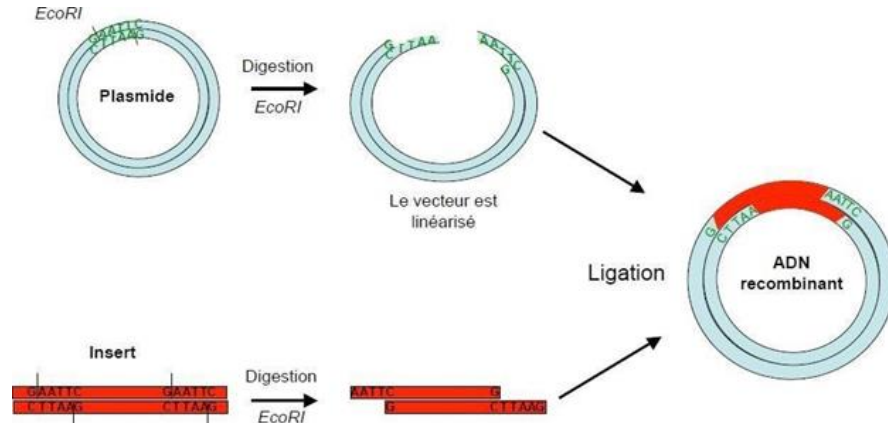
Digestion du vecteur :

On a coupé le plasmide il est prêt à accueillir l'insert.

On va vouloir obtenir un insert avec des extrémités dont les nucléotides soient complémentaires au vecteur afin qu'ils s'apparient par complémentarité des bases. Mais comment faire ?

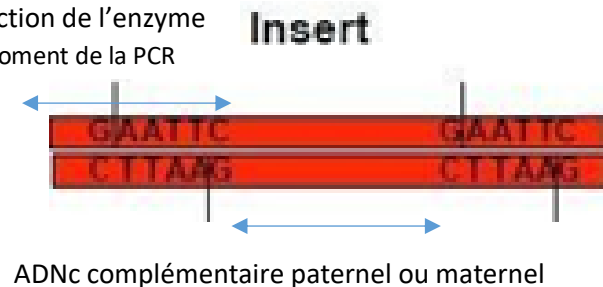
On sait que les extrémités simple brin libres de notre vecteur sont la séquence AATT sur un brin et TTAA sur le brin d'en face : obtenue après digestion d'EcoRI.

En fait on va s'arranger pour que notre **insert** possède le **site de restriction de l'enzyme choisie** (ici EcoRI) à ses extrémités, afin que l'enzyme puisse venir digérer l'insert et que les bords vecteurs/insert puisse se compléter par complémentarité (comme les pièces d'un puzzle).



Pour cela, on se rappelle qu'au moment de la **PCR** on a **ajouté une séquence à l'amorce** afin que cette séquence soit amplifiée avec l'ADNc de la Wolframine. Cette séquence était justement la suite de nucléotides GAATTC : site de restriction d'EcoRI. Attention normalement on n'a pas touché à l'information qu'on veut séquencer = ADNc de l'enfant !! On obtient après PCR :

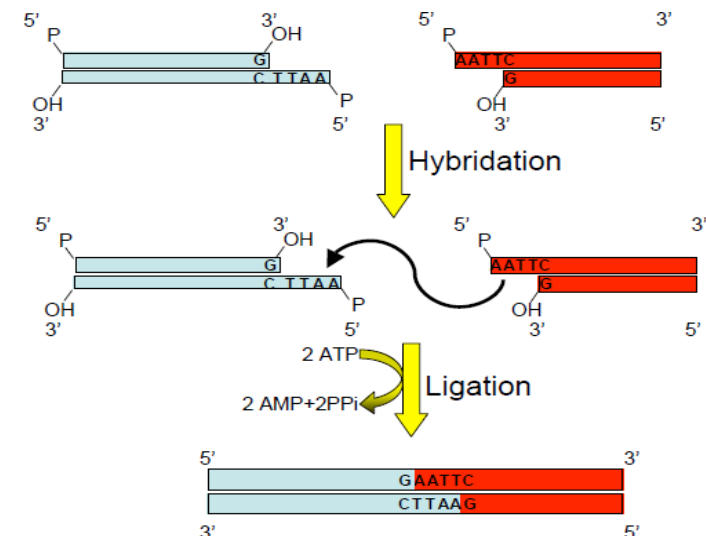
Site de restriction de l'enzyme rajoutée au moment de la PCR



2. Ligation du plasmide et de l'insert via la T4 DNA ligase

Par **complémentarité des bases** le brin 5'-3'OH de l'insert s'hybride au brin 3'OH-5'P du vecteur -> création de **liaisons hydrogènes** entre deux brins face à face

Puis afin de lier les brins 5'-3' de l'insert et 5'-3' du vecteur (liaison sur le même brin = liaison phosphodiester) on a besoin de la **T4 DNA ligase**. Enzyme qui catalyse la formation d'une liaison **phosphodiester** entre un 3'OH et un 5'Phosphate en **présence d'ATP** et d'ions divalents = liaison covalente entre deux fragments d'ADN.



NTB : les liaisons hydrogènes :

- création : naturellement par complémentarité des bases
- destruction : par la chaleur, d'où l'étape de dénaturation de la PCR à 95°

Les liaisons phosphodiesters = liaisons covalentes :

- création : ADN polymérase ou ligase (enzyme)
- destruction : la chaleur ne marche pas, il faut aussi une enzyme : nucléase (endo ou exo)

L'étape de ligation selon les 3 stratégies de clonage :

✚ Extrémités cohésives : pas de soucis

Rien de particulier à faire la complémentarité des bases attire l'insert dans le plasmide (liaisons H) puis la T4 DNA ligase se charge des liaisons phosphodiester.

✚ Extrémités franches : étape supplémentaire de déphosphorylation (facultatif)

On rajoute une étape car la ligase risque de refermer le vecteur sur lui-même avant même d'avoir mis l'insert ce qui n'est pas intéressant. On ajoute des phosphatases au vecteur uniquement. Ce sont des enzymes bactériennes déphosphorylantes, elles remplacent les phosphates en 5' par un -OH. Ainsi la ligase ne peut plus agir sans utiliser l'insert, qui sera alors intégré.

Explication : La T4 DNA ligase n'a qu'une **seule liaison phosphodiester** à former dans le cas où **l'insert n'est pas présent**, alors qu'elle doit en faire **deux** si l'insert est **intégré dans le vecteur**.

L'enzyme aura donc tendance à **privilégier** le cas le plus simple, c'est-à-dire celui où elle n'a qu'une **seule liaison à faire**, et donc où l'insert n'est **pas intégré** dans le vecteur !

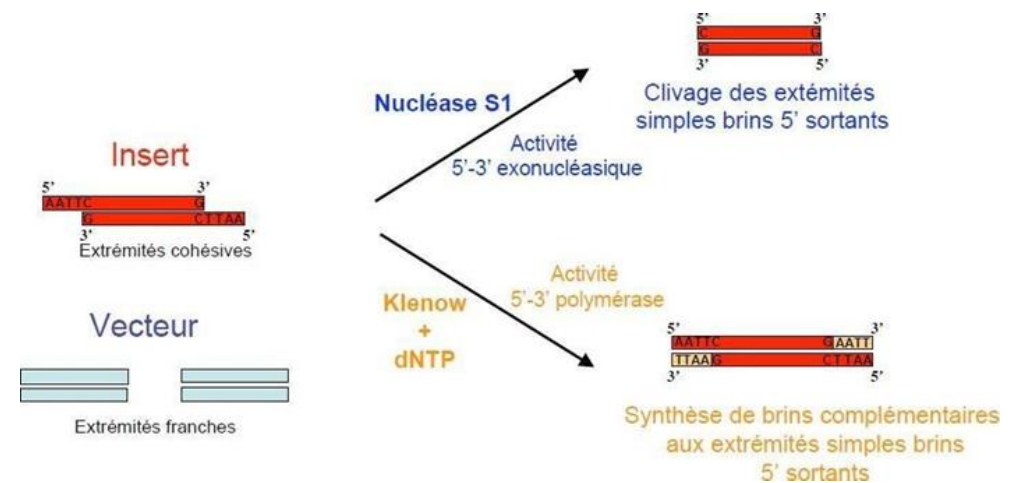
✚ Extrémité franches (vecteur) et cohésives (insert) ou l'inverse :

On utilise **de préférence le même type d'enzymes de restriction sur le vecteur et sur l'insert** pour avoir le même type d'extrémité pour les deux.

Cependant, il existe des situations où **on ne peut pas** utiliser les mêmes enzymes de restriction pour le vecteur (pour lequel on dispose d'un grand nombre d'enzymes utilisables) et pour l'insert.

On a donc recours à une **stratégie de clonage** car le vecteur et l'insert ne présentent pas le même type d'extrémités (on peut avoir un insert à bords francs et un vecteur à bords cohésifs ou l'inverse, dans les deux cas les deux ne sont pas complémentaires et ne pourront pas s'apparier/s'hybrider).

Exemple : dans cette situation on a un insert à bords cohésifs et un vecteur à bords francs, 2 solutions :



- Soit **on coupe** les extrémités 5' sortantes de l'insert :

Nucléase S1 : elle possède une **activité 5'-3' exonucléasique** lui permettant de grignoter les extrémités simple brin sortantes : l'enzyme s'arrête de grignoter les nucléotides dès qu'elle rencontre de l'ADN double brin !

- Soit **on allonge** l'extrémité 3' de l'insert pour faire du double brin partout :

Klenow + dNTP : elle possède une **activité 5'-3' polymérase** qui lui permet de combler l'ADN simple brin en synthétisant les brins complémentaires en présence de dNTP !

Remarque : Dans les 2 cas on ne touche qu'à l'insert !

Point Vocab :

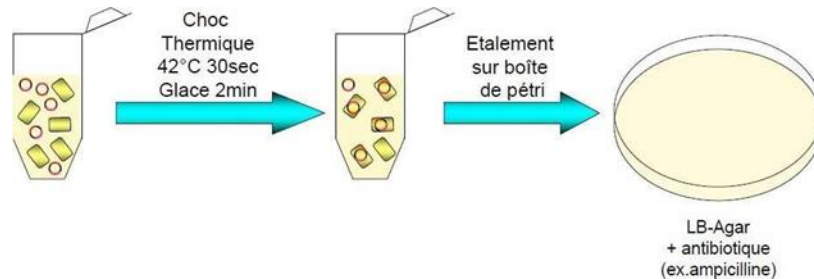
- Enzyme de restriction = ENDOnucléase (à l'intérieur) différence avec EXONucléase (aux extrémités), attention aux pièges elle a insisté dessus
- Copier/synthétiser = polymérase
- Reformuler les liaisons = ligases.

Conclusion du clonage moléculaire :

On se retrouve avec **un seul produit PCR dans un vecteur** : 2 molécules ne pouvant pas s'insérer dans le vecteur. On a ainsi créé un ADN recombinant afin **d'isoler UN produit PCR**, (que ce soit le maternel ou paternel on ne sait pas encore mais c'était pas le but, on verra plus tard en séquençant).

D. Transformation : introduction du vecteur dans une bactérie

On se sert de souche issue d'E.Coli. On place les ADN recombinants et les bactéries dans un tube. Pour que l'ADN recombinant puisse rentrer dans la cellule hôte on doit rendre les **bactéries compétentes** : leur **membrane** a été **perméabilisée** avec un **choc thermique** (42° 30sec puis dans la glace pendant 2 min) ou **électrique**.



On aura alors :

- Des bactéries ayant intégrées un ADN recombinant contenant l'**ADNc paternel**
- Des bactéries ayant intégrées un ADN recombinant contenant l'**ADNc maternel**
- Des bactéries ayant intégré le **vecteur sans insert** (vecteur qui s'est refermé tout seul)
- Des bactéries **sans ADN recombinant** = pas de gène de résistance situé sur le vecteur -> sensible aux antibiotiques

Introduction d'ADN dans une cellule **bactérienne**
= **TRANSFORMATION**

Introduction d'ADN dans une cellule **eucaryote**
= **TRANSFECTION**

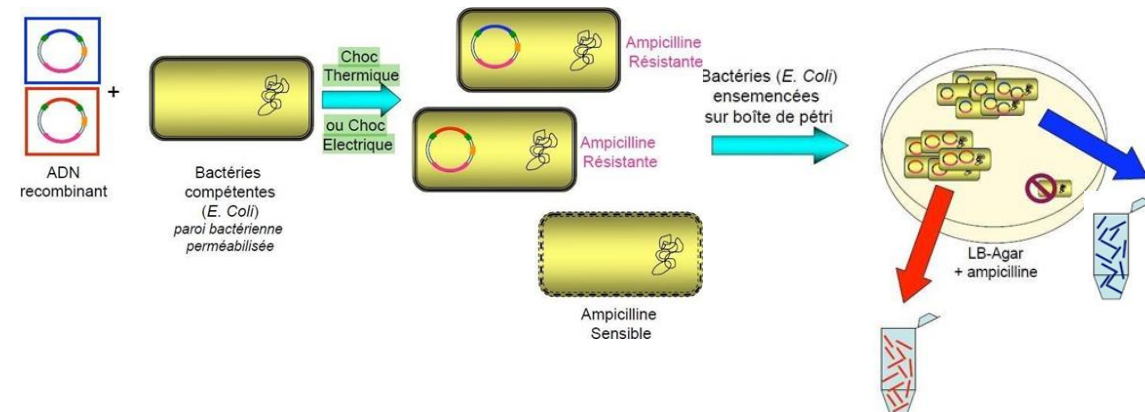
Qu'est ce qu'il nous reste à faire pour Jimmy ? Il va falloir maintenant sélectionner les bactéries d'intérêts (on élimine celles sans insert ou sans ADN recombinant), les isoler, les amplifier et extraire l'ADN Recombinant puis l'ADNc afin de le séquencer.

E. Sélection, isolement, amplification, extraction

a) Sélection

Contiennent-ils un **ADN recombinant** ? -> sélection à l'**Ampicilline** (antibio)

On va alors étaler les bactéries sur une boîte de Pétri dans du LB-Agar : c'est là que le gène de résistance à l'Ampicilline entre en jeu, il permet **d'éliminer les bactéries n'ayant pas intégré de vecteur** (elle insiste beaucoup sur ce point depuis le début du cours). Les boîtes sont stockées à 37°C, au bout de 16h on a des colonies. Il faut faire attention à pouvoir les séparer par la suite pour le repiquage, donc ne pas en mettre trop dans la même boîte.



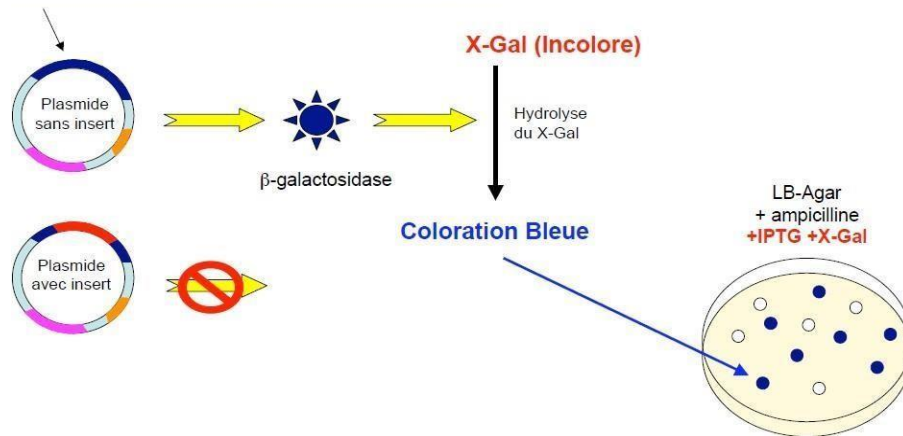
Contiennent-ils un **insert** ? -> **sélection blanc/bleu**

Parfois malgré certaines précautions (comme la déphosphorylation pour les bords francs) on a des **vecteurs qui se referment sur eux-mêmes sans insert**.

Pour savoir s'il y'a un insert on va utiliser notre connaissance de la position des gènes du plasmide. Le **poly-linker** est au sein du **gène** codant pour la **Bêta-galactosidase**, l'insert étant intégré à ce niveau il va éteindre le gène ... s'il est bien présent !

Donc, pour voir si la B-Galactosidase est exprimée on ajoute dans la boîte de Pétri : de l'**IPTG** (qui induit l'expression de la protéine) et **X-Gal** (substrat incolore qui devient bleu si la B-Galactosidase l'hydrolyse : on ne repiquera que les colonies **BLANCHES** !

IPTG : Induit l'expression de la β -galactosidase

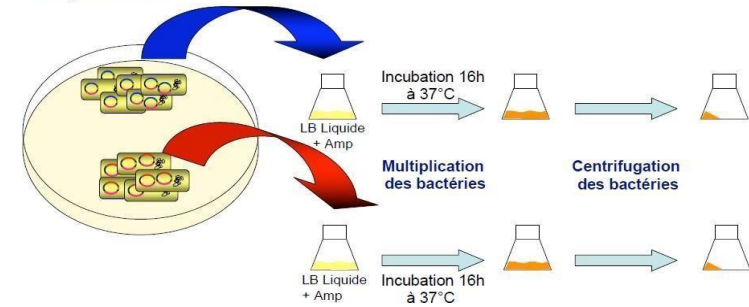


AVEC INSERT	SANS INSERT
Gène bêta-galactosidase inactivé	Gène bêta galactosidase activé
X-Gal incolore non hydrolysé	X-Gal hydrolysé colorant le milieu en bleu
COLONIES BLANCHES	COLONIES BLEUES

b) Isolement et amplification

On récupère donc uniquement les colonies blanches pour les repiquer en milieu liquide à 37° pendant une nuit avec de l'ampicilline pour pas que d'autres bactéries se développent, la multiplication de notre ADN d'intérêt va continuer.

Une colonie est mise en culture dans un milieu liquide de culture adapté à la croissance bactérienne



Remarque : on veut séparer les deux populations d'ADN (paternel et maternel) en isolant les bactéries sur une boîte de pétri et en les étalant de façon à obtenir plein de petites colonies. Bien comprendre qu'une bactérie = un ADN Recombinant = Un insert = ADNc paternel **OU** maternel



UNE SEULE bactérie est à l'origine de toute une colonie

Un seul « point » correspond à une colonie et non pas à une bactérie ! Il s'agit en réalité d'une bactérie qui s'est multipliée pour donner une colonie : c'est la **même population**. Par contre d'une colonie à l'autre les bactéries ne contiendront pas forcément le même ADN recombinant (paternel ou maternel).

On s'arrange aussi pour que les bactéries soient suffisamment éloignées les unes des autres pour que les clones soient purs et qu'on en ait un seul type par colonie !

c) Extraction

On récupère ensuite les bactéries par centrifugation. De ce culot bactérien on va extraire l'ADN, avec le même principe que pour l'extraire d'une cellule :

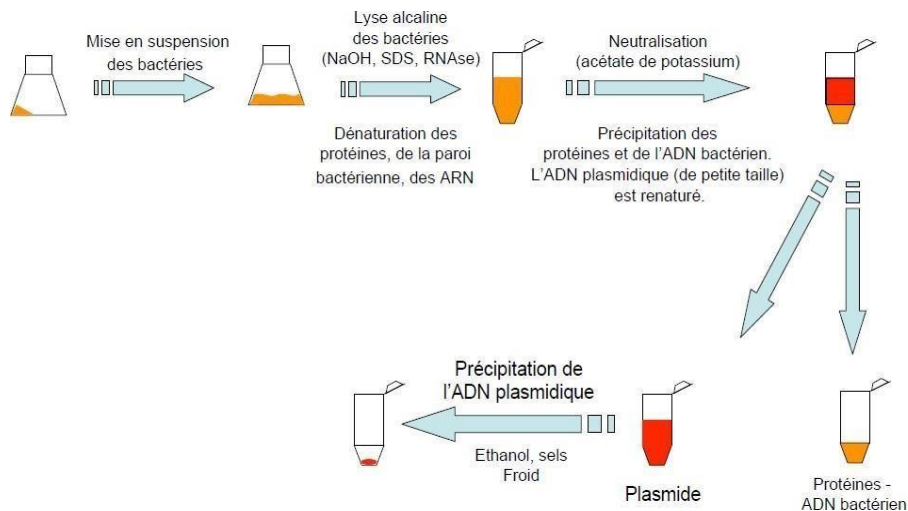
1) **Mise en suspension** des bactéries dans un tampon particulier : **NaOH** (lyse alcaline des protéines), détergents (**SDS**) pour la paroi bactérienne et **RNases** pour éliminer l'ARN bactérien.

2) **Neutralisation** de la lyse (acétate de potassium)

3) **Précipitation** des protéines et de l'ADN bactérien. L'ADN plasmidique (de petite taille) est renaturé.

4) On **centrifuge** et on sépare protéines/ADN bactérien des plasmides (dans le surnageant).

5) **Précipitation** de l'ADN plasmidique : on récupère le surnageant à part et on rajoute éthanol et NaCl et on congèle à -20°C . Plus qu'à récupérer le culot d'ADN.

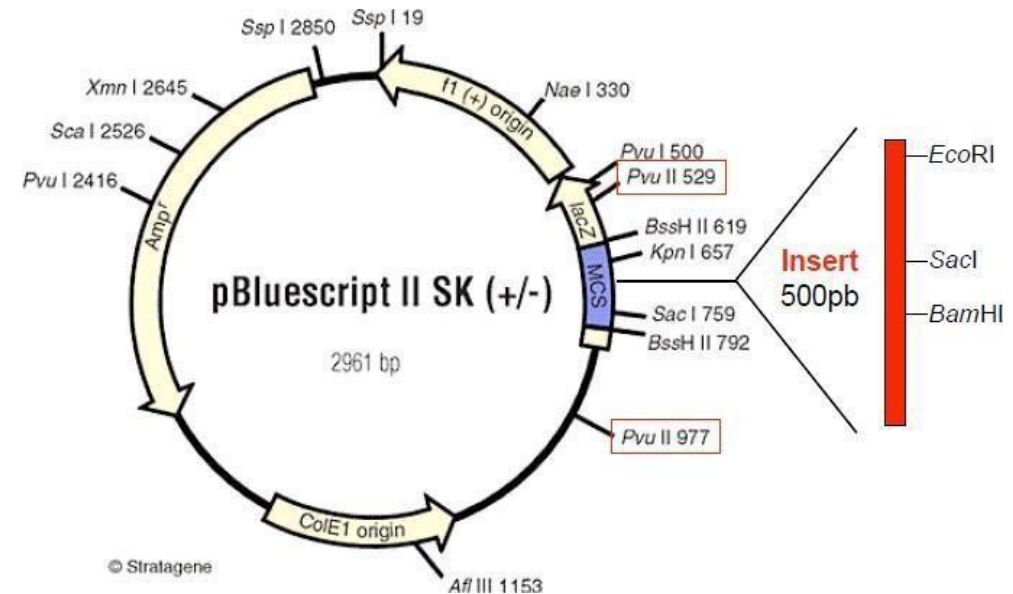


Maintenant on a notre ADN Recombinant, il ne nous reste plus qu'à isoler l'insert et le séquencer ! Mais avant, comme pour notre PCR, **on se doit de vérifier que la séquence que l'on veut séquencer est intacte** (imaginez si on séquence un insert abîmé et qu'on fait à cause de ça un faux diagnostic !). Pour la PCR on faisait une électrophorèse pour vérifier que notre fragment faisait bien le bon nombre de paires de base. Ici on utilise les **cartes de restrictions**.

F. Carte de Restriction

Ces cartes sont en fait des **schémas** sur lesquels on retrouve la **séquence complète du plasmide** et où sont tous les **sites de restriction**. Ils ne sont pas tous figurés mais généralement il y a ceux qui servent (peu fréquents). On peut retrouver les autres dans la documentation du plasmide si jamais. On va alors **digérer le plasmide** avec différentes enzymes de façon à **vérifier qu'il y a bien l'insert** et qu'il est bien de la **bonne taille**, car même avec la sélection par B-Gal on n'est pas complètement surs.

Exemple du cours :

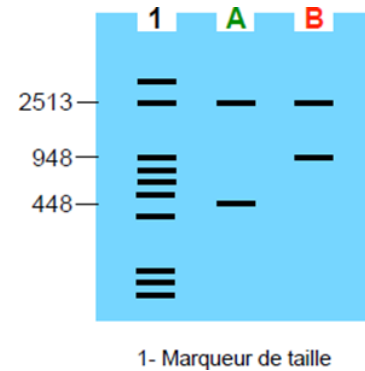


❖ Ici on a un vecteur qui fait 2961 pb

Si on coupe le vecteur avec **Pvu II**, l'enzyme agit en 529 et en 977. On se retrouve alors avec un fragment de 2513 et un de 448 pb s'il n'y a pas d'insert (piste A),

contre un fragment à 2513 et un à $448+500 = 948$ pb s'il y a l'insert (qui fait 500pb) (piste B).

On peut donc savoir s'il y a un **insert ou non** en faisant simplement **migrer sur gel** d'agarose : selon la **taille** observée on sait déterminer la présence ou non de l'insert.



❖ **Attention, vous pouvez avoir des sites de restriction dans l'insert.**

En concours, on vous donne un cas de figure en vous précisant la taille de l'insert et en vous disant si telle ou telle enzyme est présente ou non dans l'insert. On vous donne également un résultat de gel d'agarose et c'est à vous d'interpréter. Il faut donc toujours bien calculer les fragments qui sont coupés par telle ou telle enzyme.

Est-ce qu'on tient compte aux sites de restrictions dans l'insert ? Ce sera précisé.
Par exemple ici « l'insert fait 500 pb et il contient un site de restriction à l'enzyme *SacI* qui nous intéresse, cette enzyme coupe l'insert en 2 fragments de 200 et 300 pb ».

Voilà c'est finiii ! Le reste sortira en fiche de cours (ça représente environ 12 pages de ronéo donc vous avez fait une grosse partie de l'UE 11)

Bossez à fond les cartes de restrictions les petits potes ça tombe 2 QCMs mini par an, sur 9 ça fait beaucoup. Une séance méthodo est prévue sur ça, on commencera dès la rentrée à vous entraîner ++ pour que vous soyez bouillant sur ça. Des bisous courage pour ce S2 qui commence, vous verrez c'est un régal <3 PS : on sait que reprendre sans avoir les résultats c'est chaud mais faut vraiment s'y mettre que vous soyez deg ou non du cc du S1

Bisous Aubrée, Natacha, Linda, Andréa Labrunie et Campoverde et mes fillots et fillotes <3

Ps : Natacha on est un peu écolo donc pas la place pour des photos mais dans la fiche de cours complète tu seras pas déçue