

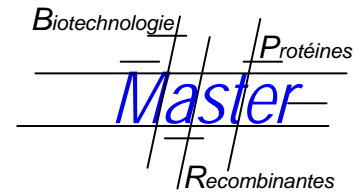
UNIVERSITE
PAUL
SABATIER



TOULOUSE III

Master Professionnel

Expression Génique et
Protéines Recombinantes



e-mail : dess.biotech@ipbs.fr / Tel : +33 (0) 5 61 17 58 59
http://www.adbt.com/DESS_EGPR/index.htm

Techniques et Stratégies en Biologie Moléculaire Première partie

Février 2007

Vincent Ecochard
Didier Fournier
Laurence Nieto
Laurent Paquereau

LES OUTILS EN BIOLOGIE MOLECULAIRE.....	3
DETECTION ET COLORATION DES ACIDES NUCLEIQUES	3
1) <i>Absorption en UV</i>	3
2) <i>Coloration</i>	3
3) <i>méthode enzymatique</i>	3
LES ENZYMES « OUTILS » DE LA BIOLOGIE MOLECULAIRE	3
1) <i>les nucléases (DNases et RNases)</i>	3
2) <i>Les polymérases</i>	3
3) <i>Les topoisomérases</i>	3
4) <i>Les ligases</i>	3
5) <i>Les phosphatases et kinases</i>	3
6) <i>Les méthylases</i>	3
7) <i>Les glycosidases</i>	3
LES VECTEURS	3
1) <i>Les plasmides</i>	3
2) <i>Les phages</i>	3
3) <i>Chromosomes</i>	3
4) <i>Systèmes hybrides</i>	3
LES MARQUEURS DE SELECTION	3
1) <i>Antibiotiques Procaryotes</i>	3
2) <i>Antibiotiques Eucaryotes</i>	3
<i>Les peptides et protéines toxiques</i>	3
<i>La complémentation</i>	3
LES SOUCHES	3
<i>Les mutations d' E. coli utilisées en biotechnologie</i>	3
<i>Les souches d' E. coli les plus utilisées</i>	3
<i>Les souches de cellules eucaryotes</i>	3
LES BANQUES.....	3
a) <i>Banque d'ADNc</i>	3
b) <i>Banque génomique</i>	3
PURIFICATION DES ACIDES NUCLEIQUES.....	3
1) PURIFICATION	3
2) TRANSFORMATION-TRANSFECTION	3
COMMENT FAIRE RENTRER UN ACIDE NUCLEIQUE DANS UNE CELLULE ?	3
COMMENT FAIRE RENTRER UN ACIDE NUCLEIQUE DANS LE NOYAU ?	3
COMMENT FAIRE RENTRER UNE PROTEINE DANS UNE CELLULE ?	3
LE MARQUAGE DES ACIDES NUCLEIQUES	3
I) LES DIFFERENTS TYPES DE SONDE ET LEUR REVELATION.....	3
<i>Sondes chaudes</i>	3
<i>Sondes froides</i>	3
II) LES DIFFERENTES TECHNIQUES DE MARQUAGE DES ACIDES NUCLEIQUES	3
<i>Marquage interne</i>	3
<i>Marquage aux extrémités</i>	3
HYBRIDATION.....	3
I) DISSOCIATION OU DENATURATION	3
<i>Calcul du Tm</i>	3
<i>Mise en évidence des états natifs et dénaturés</i>	3
II) HYBRIDATION	3
1) <i>Hybridation en phase liquide</i>	3

2) Hybridation sur support solide	3
---	---

LE SEQUENÇAGE DES ACIDES NUCLEIQUES.....3

I) LES DIFFERENTES METHODES DE SEQUENÇAGE	3
1) Méthode chimique ou méthode de Maxam et Gilbert	3
2) Méthode enzymatique ou méthode de Sanger	3
3) « Pyrosequencing »	3
4) Séquençage par hybridation	3
II) LES TECHNIQUES DE MARCHE POUR SEQUENCER DE GRANDS FRAGMENTS	3
Utilisation de nouveaux primers (primer walking)	3
Coupure au hasard	3
Transposition	3
III) LA PREPARATION DE LA MATRICE	3

AMPLIFICATION D'ACIDES NUCLEIQUES.....3

AMPLIFICATION D'ADN : PCR	3
Les composants du milieu réactionnel	3
Comment amplifier de grands fragments ?	3
Comment diminuer les erreurs lors de la polymérisation ?	3
Comment éviter les amorçages au hasard ?	3
Comment obtenir un seul des deux brins ?	3
Comment éviter les pollutions par les produits d'une amplification précédente ?	3
Comment détecter les produits d'amplification ?	3
Comment amplifier un fragment d'ADN rare ?	3
AMPLIFICATION D'ARN	3
LA PCR DE FUSION	3

LA MUTAGENESE.....3

LA MUTAGENESE DIRIGEE	3
a) A partir d'un ADN simple brin	3
b) A partir d'un ADN double brin	3
LA MUTAGENESE ALEATOIRE	3
LA MUTAGENESE SEMI-ALEATOIRE	3
LA MUTAGENESE PAR RECOMBINAISON	3
LE CRIBLAGE	3

CLONAGE DE GENES EUCARYOTES.....3

1) ON A UN ANTICORPS DIRIGE CONTRE LA PROTEINE D'INTERET	3
Fabrication d'une banque de d'ADNc	3
Criblage immunologique d'une banque d'ADNc	3
2) ON CONNAIT LA SEQUENCE DE LA PROTEINE	3
3) ON A LA SEQUENCE DE GENES PROCHES	3
Criblage d'une banque	3
PCR	3
4) ON A UN CLONE ET ON VEUT LES PARTIES FLANQUANTES	3
5) ON A UNE MUTATION, SA POSITION OU SON ORIGINE	3
7) ON VEUT CLONER UN GENE AYANT UN TAUX D'EXPRESSION VARIABLE	3
Protéomique :	3
Criblage différentiel	3
Hybridation soustractive	3
Differential display	3
Clonage des gènes les plus exprimés dans un type cellulaire, EST	3
Hybridation sur puce à ADN	3
8) GENES SYNTHETIQUES	3

CONSTRUCTIONS.....3

CONSTRUCTIONS.....	3
CONSTRUCTION SANS SITE DE RESTRICTION.....	3
<i>Par PCR :</i>	3
<i>Par recombinaison :</i>	3
INSERTION DANS UN PLASMIDE D'UN FRAGMENT D'ADN OBTENU PAR PCR A L'AIDE DE LA TAQ DNA POLYMERASE.....	3
LA RECIRCULARISATION DU VECTEUR	3
TRI DES RECOMBINANTS	3
SOUS CLONAGE PAR DES TECHNIQUES GENETIQUES	3
ANALYSE DE LA TRANSCRIPTION.....	3
1 - DETERMINATION DU SITE D'INITIATION DE LA TRANSCRIPTION.....	3
<i>Cartographie à la nucléase S1</i>	3
<i>Extension d'amorce</i>	3
<i>Transcription in vitro (type Run off)</i>	3
2 - QUANTIFICATION D'UN TRANSCRIT.....	3
<i>Northern</i>	3
<i>Protection à la RNase</i>	3
<i>RT-PCR</i>	3
3 - COMPARAISON DU TAUX DE TRANSCRIPTION DANS DEUX CELLULES DIFFERENTES.....	3
<i>Expérience de Run on (Elongation in vitro de transcrits initiés in vivo)</i>	3
4 - DETERMINATION DES REGIONS PROMOTRICES RECONNUES PAR DES FACTEURS TRANSCRIPTIONNELS.....	3
<i>Gel retard</i>	3
<i>Détermination des séquences reconnues</i>	3
<i>Détermination des bases qui interagissent avec les facteurs transcriptionnels</i>	3
5 - ETUDE FONCTIONNELLE DE LA TRANSCRIPTION EN SYSTEME CELLULAIRE.....	3
<i>Choix de la cellule</i>	3
<i>Constructions chimériques</i>	3
6 - APPROCHES IN VIVO.....	3
<i>Sites sensibles à la DNaseI</i>	3
<i>Sites hypersensibles à la DnaseI</i>	3
<i>Footprint génomique</i>	3
<i>Transfection stable</i>	3
7 - ANALYSE DE L'ABONDANCE DE LA TRANSCRIPTION (TRANSCRIPTOME).....	3
<i>EST</i>	3
<i>Array hybridization</i>	3
<i>SAGE</i>	3
8 - ANALYSE DE LA METHYLATION DES PROMOTEURS	3
ANALYSE DE LA TRADUCTION	3
ANALYSE DE L'EXPRESSION PAR ANTISENS	3
ETUDE DE LA POPULATION D'ARN EN COURS DE TRADUCTION DANS UNE CELLULE	3

Les outils en biologie moléculaire

Détection et coloration des acides nucléiques

1) Absorption en UV

On peut doser les acides nucléiques en utilisant les propriétés d'absorption des bases azotées (A, C, G, T et U) dans l'UV aux alentours de 260 nm. Les maxima et les coefficients d'absorption moléculaire (ϵ) de chaque nucléotide sont les suivants :

$$\text{AMP: } \epsilon_{259} = 15.400 \text{ M}^{-1}\text{cm}^{-1}$$

$$\text{GMP: } \epsilon_{253} = 13.700 \text{ M}^{-1}\text{cm}^{-1}$$

$$\text{CMP: } \epsilon_{271} = 9.000 \text{ M}^{-1}\text{cm}^{-1}$$

$$\text{TMP: } \epsilon_{267} = 9.600 \text{ M}^{-1}\text{cm}^{-1}$$

$$\text{UMP: } \epsilon_{262} = 10.000 \text{ M}^{-1}\text{cm}^{-1}$$

On peut donc estimer que le coefficient moyen d'absorption à 260 nm est égal, pour un nucléotide, à $10.000 \text{ M}^{-1} \text{ cm}^{-1}$.

On peut estimer la concentration en base d'une solution d'acide nucléique en appliquant la formule :

$$DO = \epsilon \times C \times l = 10.000 \times C \times l$$

Comme on travaille généralement avec des cuves de 1 cm on peut simplifier et écrire

$$DO = 10.000 \times C.$$

Les nucléotides ne sont pas libres mais associés en polymères, la concentration en mole litre⁻¹ n'a donc pas beaucoup de sens et on préfère l'exprimer en gramme par litre. Sachant que le poids moléculaire moyen d'un nucléotide est égal à 300.

$$DO = 10.000 \times C / 300 \text{ (avec } C \text{ en g/l ou mg/ml)}$$

On travaille avec des solutions moins concentrées, qui sont plutôt de l'ordre du $\mu\text{g/ml}$. Dans ce cas

$$DO = 10 \times C / 300 = C/30 \text{ (avec } C \text{ en } \mu\text{g/ml)}$$

Ainsi une concentration de 30 $\mu\text{g/ml}$ donne 1 unité DO pour des acides nucléiques sans structure secondaire, par exemple pour un oligonucléotide.

Mais ce n'est pas vrai pour l'ARN ni pour l'ADN. Expérimentalement on observe qu'une solution de 50 μg d'ADN/ml donne une absorbance de 1 unité DO. Ce qui est normal si on se souvient qu'il y a des interactions entre les bases dans les structures secondaires de l'ARN ou dans la double hélice d'ADN.

Attention, à 260nm on ne dose pas seulement l'ADN ou l'ARN mais les bases constituant ces molécules. On dose donc également les nucléotides libres. Si un échantillon absorbe à 260 nm, on dose l'ADN, l'ARN et les nucléotides libres de la solution.

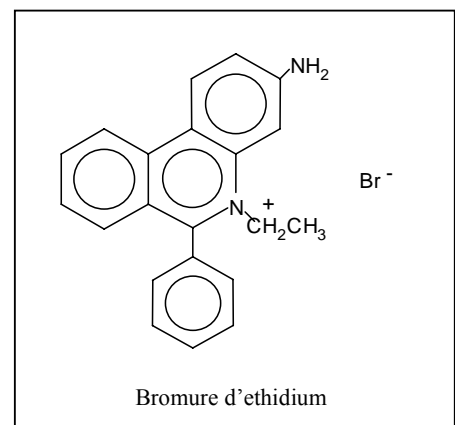
Les protéines absorbent à 280 nm ($\epsilon_{\text{trp}} = 5600 \text{ M}^{-1}\text{cm}^{-1}$ et $\epsilon_{\text{tyr}} = 1200 \text{ M}^{-1}\text{cm}^{-1}$), si bien qu'on considère que si le rapport $A_{280}/A_{260} = 0.5$, la solution est dépourvue de protéines.

2) Coloration

Le bromure d'éthidium (BET)

Le bromure d'éthidium est un intercalant qui interagit avec les plateaux des bases. Il est fluorescent lorsqu'on l'excite dans l'UV ($\lambda_{\text{exc.}}$: 302nm ou 546nm; λ_{em} : 590nm) (Le Pecq et Paoletti, 1966). Il a plus d'affinité pour les molécules double brin (ADN) que pour les molécules simple brin (ARN, oligonucléotides) mais il peut être utilisé dans les deux cas.

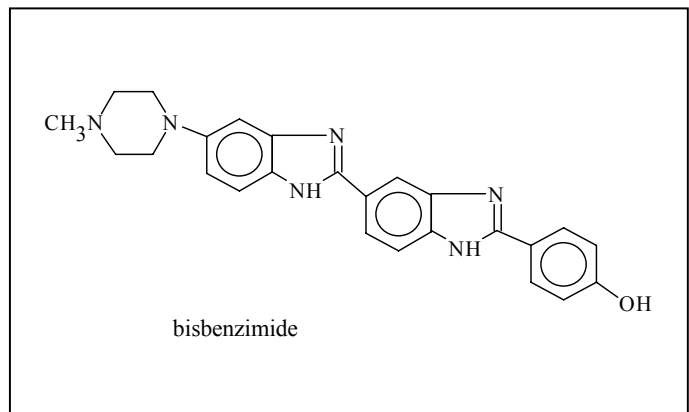
La quantité de fluorescence intercalée dans les plasmides surenroulés est plus faible que pour ce même plasmide sous forme relâchée car moins de molécules du colorant peuvent s'intercaler entre les bases à cause des contraintes conformationnelles liées au surenroulement. L'interaction du BET entre les bases les



éloigne les unes des autres et augmente le pas de l'hélice. Il s'ensuit une diminution de la densité de l'ADN d'autant plus importante que beaucoup de BET s'est fixé.

La bisbenzimidazole ou Hoechst 33258

C'est un composé fluorescent (λ exc.: 365nm, λ em: 450nm) qui ne s'intercale pas entre les bases, il se fixe sur le sillon mineur de la double hélice. Cette coloration est plus sensible que celle obtenue avec le bromure d'éthidium et elle est spécifique de l'ADN.



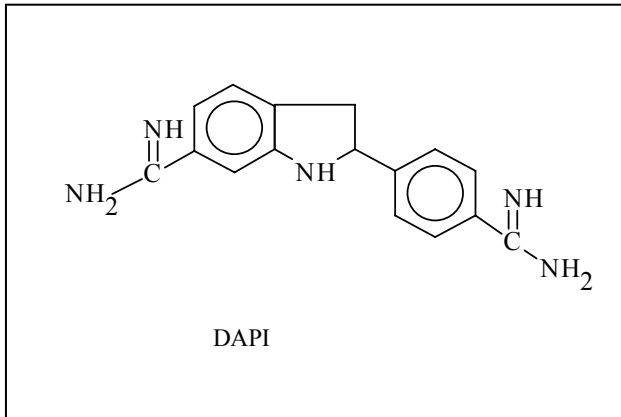
	absorbance (260 nm)	Fluorescence	
		Bisbenzimidazole	Bromure d'éthidium
Sensibilité ($\mu\text{g/ml}$)			
DNA	1-50	0.01-15	0.1-10
RNA	1-40	--	0.2-10

Toutefois la sensibilité de détection dépend de la composition en base (elle est moins sensible lorsque l'ADN comporte un fort pourcentage en GC). La sensibilité dépend également du pH qui doit être maintenu à 7,4. A des pH <6 ou >8, il y a perte du rendement de

fluorescence et augmentation du bruit de fond. La coloration n'est pas affectée par le surenroulement.

Le DAPI, 4',6-diamino-2-phenylindole

Composé fluorescent (λ exc. : 344nm, λ em: 466nm). Il permet de détecter des quantités d'ADN de l'ordre du ng.



Ce colorant est très spécifique de l'ADN, il ne réagit pas avec l'ARN et on n'observe pas de « quenching » avec les autres composés de la cellule (Brunk et al., 1979).

Comme pour la bisbenzimidine, ce colorant est sensible à la composition en base et la sensibilité de la coloration diminue avec l'augmentation du pourcentage en G-C. Cette méthode de coloration

n'est pas affectée par le surenroulement.

Le picogreen

Ce colorant est breveté et commercialisé par Molecular Probes. Il est environ 10.000 fois plus sensible que la détection par absorbance à 260nm, 400 fois plus sensible que la bisbenzimidine. Il permet de détecter jusqu'à 25 pg/ml en solution. Bien que ce colorant ne soit pas spécifique de l'ADN double brin, la fluorescence est plus importante après liaison à l'ADN double brin qu'après liaison à l'ARN ou à l'ADN simple brin. On peut donc l'utiliser pour quantifier l'ADN double brin lorsque les acides nucléiques simple brin ne sont pas trop importants.

Le picogreen n'est pas sensible à la composition en base de l'ADN.

L'oligogreen

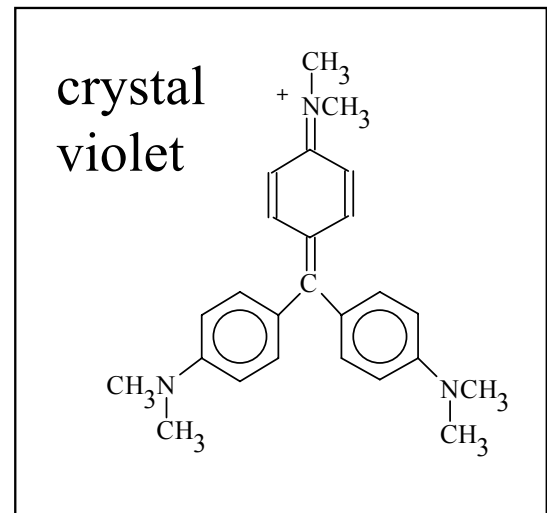
Ce colorant est breveté et commercialisé par Molecular Probes. Ce colorant est spécifique de l'ADN simple brin. Il permet de quantifier des solutions d'ADN jusqu'à 100 pg/ml, ce qui est 30.000 fois plus sensible que la quantification par absorption en UV (0,1 DO correspond à 3 µg/ml). Ce colorant permet de quantifier les oligonucléotides d'une longueur supérieure à 10 bases.

Le ribogreen

Développé et commercialisé par Molecular probes, ce colorant est utilisé pour quantifier l'ARN avec une limite de l'ordre de 1 ng/ml. Ce colorant n'est pas spécifique de l'ARN aussi si on veut s'affranchir de l'ADN le mieux est de digérer l'ADN par une DNase, avant la coloration en effet le colorant est insensible aux nucléotides.

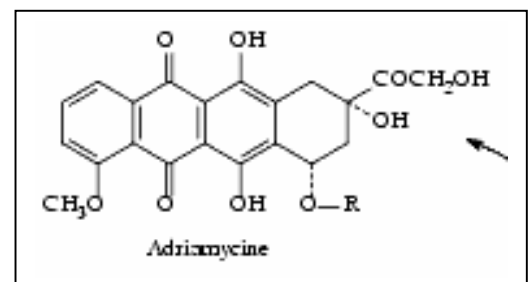
Le cristal violet

L'avantage de ce colorant est qu'il n'est pas toxique et n'abîme pas l'ADN comme le bromure d'éthidium. Toutefois cette méthode de détection n'est pas très sensible, il faut au moins 200 ng d'ADN pour révéler sa présence sur un gel alors que 10 ng suffisent pour la détection au bromure d'éthidium. Il est principalement utilisé pour colorer des gels préparatifs d'agarose ou de polyacrylamide pour faire un clonage.



L'adriamycine

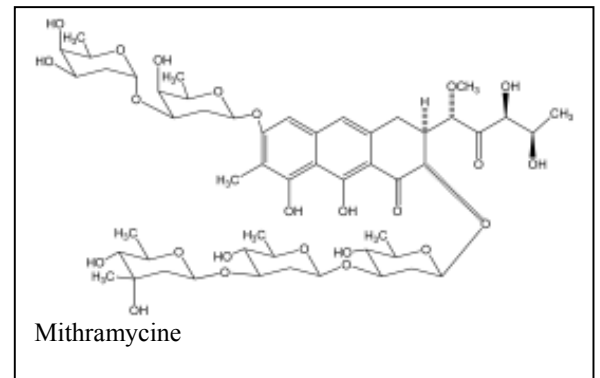
L'adriamycine est un composé antitumoral qui forme des complexes avec l'ADN. Son absorbance à 475 nm est réduite proportionnellement à la quantité d'ADN ajoutée au milieu. Cette coloration est spécifique de l'ADN et ne réagit pas avec l'ARN (Hill, 1976). Le dosage peut aussi se faire par fluorescence (λ exc.: 490nm, λ em: 555nm) avec



une sensibilité de l'ordre de 5 μ g. Dans ce dernier cas le dosage reste spécifique de l'ADN et n'est pas perturbé par les autres composants de la cellule (Kennedy et al., 1987).

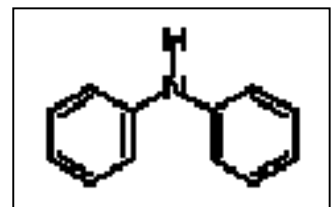
La mithramycine, la chromomycine A3 et l'olivomycine

Pour ces trois composés similaires, la fluorescence est directement proportionnelle à la concentration de l'ADN avec des spectres légèrement différents (mithramycine : λ exc: 440nm, λ em: 540nm ; chromomycine A3 λ exc: 405nm, λ em: 575nm ; olivomycine : λ exc: 440nm, λ em: 532nm). La coloration est environ 50 fois moins sensible que le bromure d'éthidium et dépend de la composition en base, elle est directement proportionnelle au pourcentage de G-C (Daxelet et al., 1989). La coloration n'est pas affectée par le surenroulement. Au moins pour la mithramycine, la coloration est spécifique de l'ADN et ne réagit pas avec l'ARN (Hill, 1976).



Le diphenylamine

L'utilisation de ce colorant est historique (Burton, 1956). L'essai est basé sur la réaction entre la diphenylamine et le désoxyribose. Mais cette réaction est sujette à des interférences avec les sucres et les protéines de la cellule.



Utilisation des colorants :

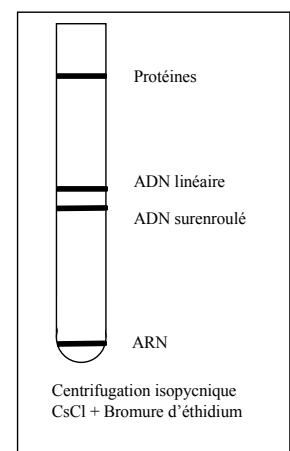
- Détermination de la quantité d'ADN dans une solution d'ADN purifié : absorption UV. Si l'ADN n'a pas été purifié, on dose aussi les ARN et les protéines absorbant à 280 nm perturbent le dosage.

- Détermination de la quantité d'ADN dans un extrait brut. On utilise le DAPI ou la bisbenzimidine. Ces composés sont spécifiques de l'ADN et le dosage n'est pas perturbé par les autres composants de la cellule. L'adriamycine peut aussi être utilisée, mais elle est moins sensible. Le picogreen peut être utilisé si les ARN ne sont pas trop important.

- Purification de l'ADN mitochondrial sur gradient de CsCl: en présence de bisbenzimidine l'ADN mitochondrial est moins dense que l'ADN nucléaire. Comme il est plus riche en AT, il y a plus de colorant fixé sur l'ADN.

- Purification de bandes sur gel avant un clonage : cristal violet.

- Purification des plasmides surenroulés. La densité de l'ADN diminuant en présence de bromure d'éthidium, l'incorporation étant diminuée lorsque l'ADN est surenroulé, on peut séparer les plasmides surenroulés de l'ADN chromosomique linéaire par centrifugation isopycniq. L'ADN du plasmide est plus dense que l'ADN linéaire.

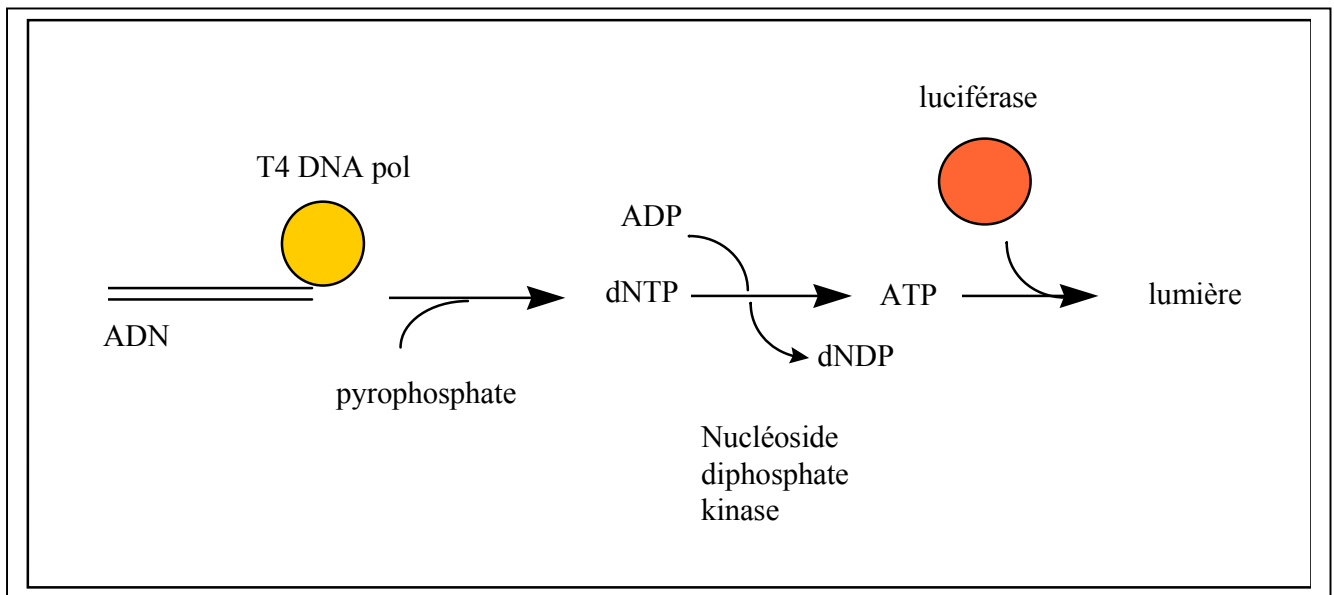


- Détermination de la composition en base : comparaison de l'estimation de la quantité d'ADN avec une méthode non affectée par la composition en bases (absorption UV), une méthode qui est plus sensible avec l'augmentation du pourcentage de A-T (bisbenzimidine) et une méthode qui est plus sensible à l'augmentation du pourcentage de G-C (mithramycine).

- Coloration des ARNs sur un filtre, dans ce cas on utilise le bleu de méthylène.

3) méthode enzymatique

On peut convertir les bases de l'ADN double brin en ATP qui est ensuite mesuré à l'aide d'une réaction enzymatique (Kinney *et al.*, 1999).



On utilise comme première réaction une pyrophosphorylation, cette réaction est l'inverse de la polymérisation et s'effectue en présence de T4 DNA polymérase et de pyrophosphate. Il y a alors production de dNTP.

Il y a ensuite conversion des dNTP en ATP par la NDPK (nucléoside diphosphate kinase). Ici le phosphate terminal des dNTP est transféré sur l'ADP pour former de l'ATP.

L'ATP produit est ensuite dosé par la luciférase en présence de son substrat, la luciférine.

Cette méthode est très sensible, elle permet de quantifier des picogrammes d'ADN double brin linéaire. L'ADN simple brin ou l'ARN n'interfèrent pas, par contre, la méthode dose aussi les nucléotides triphosphates.

Les enzymes « outils » de la biologie moléculaire

Ils permettent de travailler sur les acides nucléiques, de les modifier, de les couper, de les associer etc...

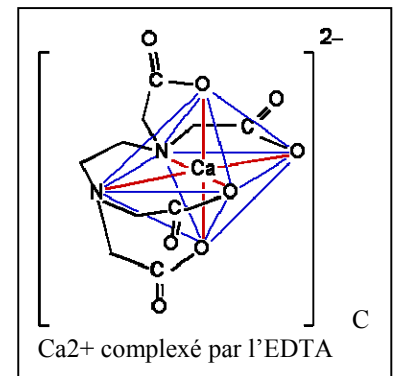
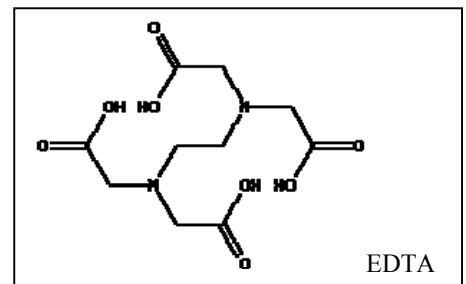
1) les nucléases (DNases et RNases)

a) Les DNases

Leur fonction consiste à couper l'ADN au niveau d'un phosphate, libérant deux fragments un se terminant par un 3'OH et l'autre par un 5' phosphate.

Toutes les DNases ont besoin d'ions divalent. Pour les inhiber, on peut donc utiliser un chélateur d'ion divalent tel que l'EDTA. Une solution d'ADN comportant de l'EDTA sera donc stable.

On distingue les exonucléases qui digèrent l'ADN en retirant les nucléotides à partir d'une extrémité, des endonucléases qui coupent un des deux brins au milieu du polymère.



α) Les exonucléases

exonucléase III est une 3'5' exonucléase, elle retire les nucléotides d'une manière séquentielle à partir de l'extrémité 3' (à la condition qu'elle ne soit pas sortante).

Exonucléase I est une 3'5' exonucléase qui clive uniquement le simple brin (à condition que l'extrémité 3' soit une extrémité OH).

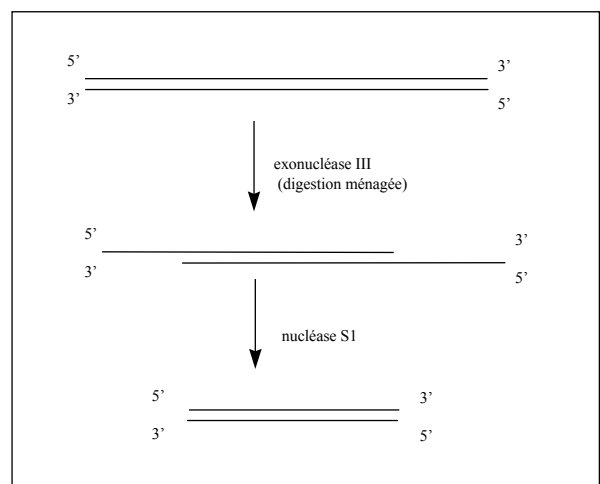
Bal 31 est aussi une exonucléase dont l'activité principale est une 3' exonucléase mais qui a aussi une activité 5'3' exonucléase. Elle a de plus une activité secondaire d'endonucléase sur le simple brin généré par l'exonucléase

La *T7Gene exonucléase* est une 5'3' exonucléase qui a comme substrat de l'ADN double brin avec un 5' phosphate ou un 5'OH.

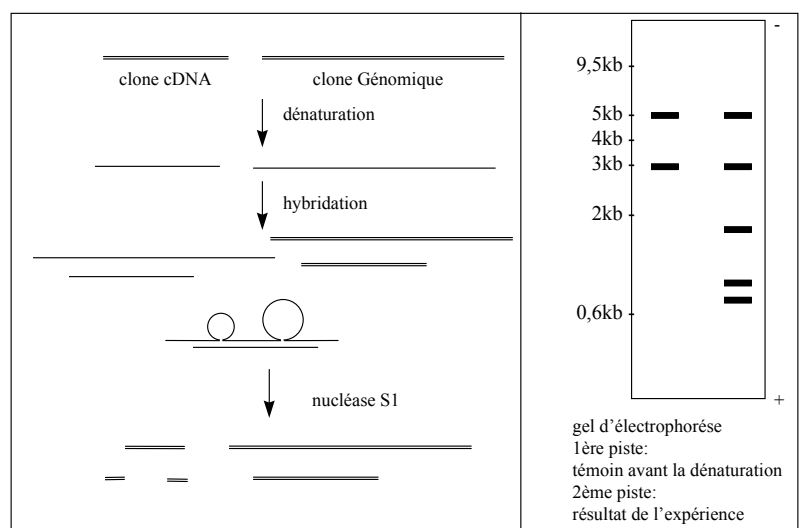
β) Les endonucléases

Les endonucléases ont diverses spécificités de substrat, certaines coupent l'ADN double brin d'autre l'ADN simple brin, enfin certaines reconnaissent et coupent au niveau de séquences caractéristiques.

DNase I : endonucléase qui coupe l'ADN double ou simple brin. On l'utilise par exemple lorsque l'on veut éliminer l'ADN d'une solution ou pour faire des « nicks » sur l'ADN comme dans la technique de marquage par déplacement de coupure (nick translation). Dans ce dernier cas on effectue une digestion ménagée. En présence de Mg^{2+} , la Dnase I coupe chaque brin de manière indépendante. En présence d'ions Mn^{2+} (10 mM), elle produit préférentiellement des coupures double brin. On l'utilise alors pour générer des fragments au hasard.



Nucléase S1 : dégrade principalement l'ADN simple brin et l'ARN. On l'utilise par exemple lorsque l'on veut réduire la taille d'un fragment d'ADN en l'attaquant par ses extrémités. Dans ce cas on fait agir l'exonucléase III qui grignote l'extrémité 3', on obtient une extrémité 5' simple brin sortante puis on fait agir la nucléase S1 pour digérer cette extrémité. Un deuxième exemple d'application est la cartographie de la structure d'un gène, de sa composition en intron et exon (voir figure ci-dessous). Pour cela, on utilise deux clones, un



provenant du génome (clone génomique) et un provenant de l'ARN (clone cDNA) que l'on place dans un tube, après dénaturation et hybridation on obtient trois sortes de molécules double brins, deux homoduplex correspondant aux réassociations des clones génomiques et cDNA et un hétéroduplex correspondant à l'hybride. Dans l'hétéroduplex, les introns présents dans le clone génomique se retrouvent simple brin et sont donc sensibles à l'attaque par la nucléase S1. Une analyse sur gel d'électrophorèse permettra de connaître le nombre d'introns, le nombre et la taille des exons.

La **mung bean nucléase** a la même action que la nucléase S1: elle digère le simple brin. Toutefois s'il y a « nick » sur un fragment double brin, elle aura du mal à digérer la liaison phosphodiester se trouvant en face du « nick » sur l'autre brin. On va l'employer par exemple pour rendre franches les extrémités de fragments d'ADN obtenus par sonication. En effet, dans ce cas on veut rendre franc le fragment sans recouper les fragments au niveau des « nicks ».

Les **DNase de restriction** (ou enzyme de restriction, restrictase) : endonucléases qui coupent au niveau d'une séquence particulière ou à proximité.

Il existe trois grandes classes (type I, II et III) seules les enzymes du type II sont utilisées en biologie moléculaire. Elles reconnaissent une séquence et coupent au niveau d'un site fixe appartenant à cette séquence ou à sa proximité. (les types I et III ne sont pas utilisés en biologie moléculaire, en effet ils portent aussi une activité de méthylation et l'activité de coupure est ATP dépendante. De plus les enzymes de type I ne coupent pas au niveau du site de reconnaissance, elles digèrent l'ADN au hasard).

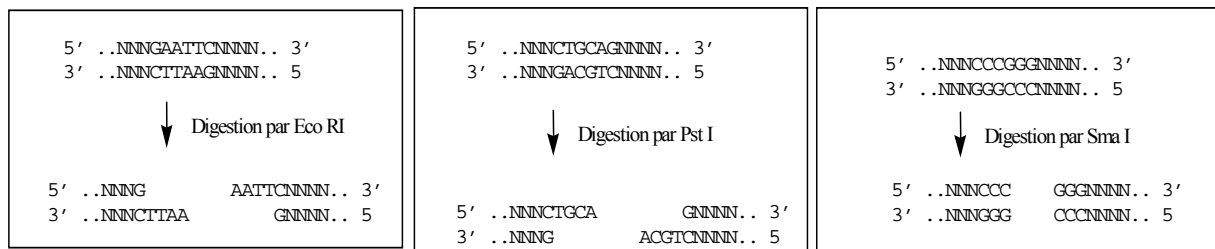
Les enzymes de restriction de type II :

Les enzymes de restriction reconnaissent des séquences le plus souvent palindromiques et libèrent des extrémités 5' phosphate et 3' OH et donnent soit des extrémités cohésives soit des extrémités franches.

Par exemple :

- Eco RI reconnaît et coupe la séquence G/AATTC (5' sortante)
- Pst I reconnaît et coupe la séquence CTGCA/G (3' sortante)

- Sma I reconnaît et coupe la séquence CCC/GGG ce qui donne des extrémités franches



La reconnaissance n'est pas toujours strictement définie [exemple : Nci I CC/ (G ou C) GG] et les palindromes peuvent être interrompus [exemple : XmnI reconnaît la séquence suivante : GAANN/NNTTC].

Fragment de restriction : fragment d'ADN généré après coupure d'une molécule d'ADN par des endonucléases de restriction.

Le nom des enzymes de restriction provient de la bactérie dont elles ont été isolées. Par exemple Eco RI provient d'*Escherichia coli* souche R, le I signifie que c'est la première enzyme qui a été découverte. Pour la bactérie, cette enzyme sert à digérer l'ADN des phages qui pourraient l'infecter.

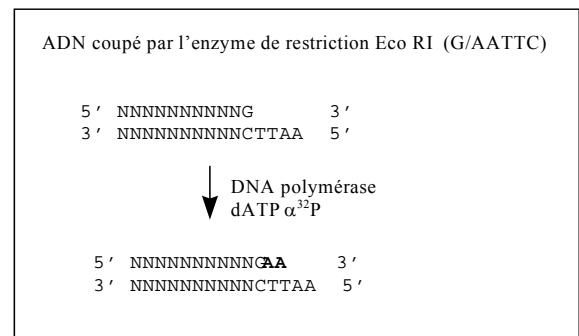
Les enzymes de restriction reconnaissent 4, 6 ou 8 bases. La fréquence de reconnaissance sera d'autant plus faible que le nombre de bases reconnues est élevé. Une enzyme reconnaissant 4 bases coupera en moyenne tous les 4⁴ bases soit tous les 256 nucléotides. De même, une enzyme reconnaissant 6 bases coupera tous les 4096 nucléotides et une enzyme reconnaissant 8 bases tous les 65536 nucléotides. Il s'agit d'une moyenne. A la suite de la coupure d'un ADN génomique, on obtient une population de molécules dont la taille moyenne est de 256 pb pour une reconnaissance de 4 bases. Par contre sur un ADN de taille moins importante tel qu'un plasmide on aura des bandes bien individualisées après séparation sur un gel d'électrophorèse.

On appelle isoschizomères des enzymes qui reconnaissent et coupent la même séquence mais qui proviennent de deux bactéries différentes

Comment transformer une extrémité cohésive 5' sortante ou 3' sortante en extrémité franche?

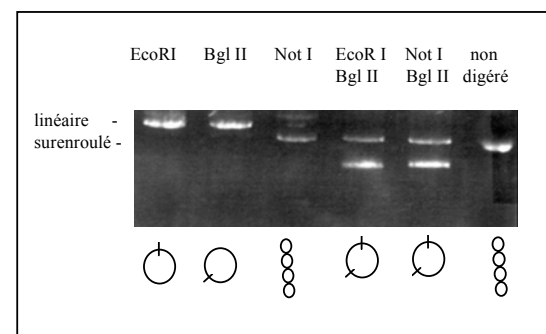
Avec une DNA polymérase en présence de deoxynucléotides tri-phosphates. Dans le cas d'une extrémité 5' sortante on aura une polymérisation et dans le cas d'une extrémité 3' sortante on aura une digestion par l'activité 3'5' exonucléase de la polymérase utilisée. Si on utilise des nucléotides marqués, dans les deux cas on aura un marquage des fragments d'ADN. Cette dernière technique permet de marquer un fragment de restriction.

Si la coupure donne une extrémité 5' sortante : en présence d'un nucléotide marqué, on aura incorporation des nucléotides marqués en faisant toutefois attention qu'ils correspondent à la séquence de reconnaissance par la DNase de restriction.



Si la coupure est 3' sortante on aura avec la même enzyme, digestion de l'extrémité 3' simple brin et remplacement du dernier nucléotide hybridé en 3'.

Les enzymes de restriction reconnaissent souvent une structure de l'ADN et non uniquement la séquence. De ce fait, certains enzymes comme Not I ne reconnaissent pas l'ADN surenroulé. Si un plasmide surenroulé est incubé en présence de Not I, on n'obtient pas de coupure, par contre si le plasmide est linéarisé on observe la coupure.



Utilisation des enzymes de restriction.

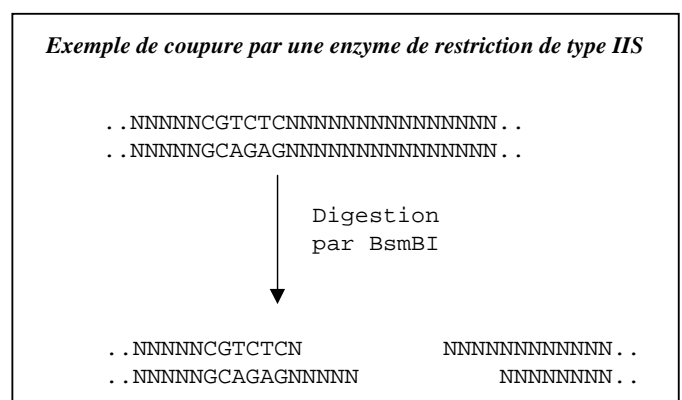
Il y a deux grands domaines d'utilisation des enzymes de restriction. Le premier est la cartographie d'un fragment d'ADN. On peut en effet identifier un fragment d'ADN grâce à sa carte de restriction, c'est à dire grâce à la position relative des sites de restriction. Le deuxième domaine d'utilisation est la construction de molécule d'ADN recombinante. Un fragment de restriction peut être recollé (ligué) à un autre fragment d'ADN si les deux extrémités sont compatibles, c'est à dire présentent les mêmes extrémités cohésives ou des extrémités franches.

Vitesse de coupure variable et digestion partielle

Souvent les enzymes de restriction ne coupent pas deux sites sur le même fragment d'ADN à la même vitesse. Certains enzymes tels que Nae I, Nar I, Sac II ou Hpa II ont une efficacité très variable en fonction des sites. Deux facteurs sont responsables de cette variabilité : les séquences flanquantes et la présence d'autres sites. L'influence du site de coupure vient du fait que certains enzymes ont deux sites de reconnaissance de l'ADN et la coupure requière une interaction simultanée aux deux sites. Dans ce dernier cas, l'ajout d'un oligonucléotide comportant le site augmente l'efficacité de la coupure.

Digestion partielle à l'aide d'une enzyme de restriction. Il arrive que l'on veuille couper un site particulier sur un fragment d'ADN alors que ce site est présent plusieurs fois sur notre molécule. La première méthode consiste à effectuer une digestion partielle (ou ménagée) et à récupérer le fragment d'intérêt sur un gel. Cependant, les coupures ne sont pas complètement aléatoires, et dépendent des séquences voisines. Aussi, certains sites sont coupés avant les autres, si ces sites sont ceux qu'on désire ne pas couper, la digestion partielle devient difficile. On peut utiliser une deuxième méthode consistant à amplifier par PCR le fragment en ajoutant dans le tampon un des dNTP méthylés. Le rôle de cette méthylation partielle est d'inhiber la coupure. Suivant la proportion de dNTP méthylés, on inhibera plus ou moins la coupure.

Les enzymes de restriction de type IIS : Ces enzymes reconnaissent une séquence et coupent à l'extérieur de cette séquence. Cette caractéristique leur donne deux propriétés intéressantes : d'une part, lorsque le site est mis à l'extrémité d'une amorce utilisée en PCR, à la suite de la coupure le site est retiré d'autre part l'extrémité générée par la digestion est unique.



Inhibitions de coupure par des méthylases: les enzymes de restriction sont sensibles à leur propre méthylase (voir plus loin) mais elles peuvent aussi être sensibles si le site de coupure recouvre un site Dam, Dcm, CpG, EcoK et EcoB. Par exemple l'enzyme Cla I reconnaît le site ATCGA/T mais elle est bloquée par la méthylase CpG qui est présente chez certains eucaryotes. Un autre exemple, la séquence de l'ADN à couper est GATCGA/T, il y a recouvrement avec la méthylation par Dam qui reconnaît le site GATC. Ainsi la coupure par Cla I n'aura lieu que si l'ADN n'est pas méthylé, s'il provient d'une réaction de PCR ou d'un plasmide originaire d'une souche Dam⁻.

b) Les RNases

Ces nucléases coupent l'ARN. Quelques RNases sont utilisées en biologie moléculaire :

- RNase A : coupe après (en 3') les résidus pyrimidiques (C, U) en donnant un 3' phosphate sur l'ARN simple brin.
- RNase T1 : coupe en 3' de la guanosine donnant une guanosine 3' phosphate.

Ces deux enzymes sont utilisées pour dégrader l'ARN, pour séquencer directement l'ARN ou pour couper les fragments d'ARN non hybridés à de l'ADN. Elles peuvent également être utilisées comme sondes de structure (ne digèrent le nucléotide ou la région concernée que s'il est sous la forme simple brin).

- RNase H : digère l'ARN dans un complexe ARN-ADN. On s'en sert pour éliminer l'ARN après avoir fabriqué un premier brin de cDNA à l'aide de la reverse transcriptase.

Les inhibiteurs de RNase : Dans les expériences nécessitant l'utilisation d'ARN, il est difficile de purifier l'ARN sans trace de RNase. Ces RNase sont des enzymes extrêmement stables, le plus souvent elles se renaturent après chauffage à 100°C. Il n'est donc pas possible de les inactiver par autoclavage. Une possibilité est d'utiliser un inhibiteur de RNase tel que la RNasin®. C'est une protéine de 50 kDa qui inhibe les RNase en interagissant avec elles selon une stœchiométrie 1:1 et une constante de dissociation de 10^{-16} M.

2) Les polymérase

Toutes les polymérase synthétisent les acides nucléiques de 5' vers 3' en utilisant des nucléotides triphosphates. L'énergie est fournie par les nucléotides triphosphates entrants. Pourquoi la polymérisation ne se fait-elle pas dans l'autre sens ? Dans le sens 5'3', en cas d'erreur il y a excision du dernier nucléotide incorporé libérant un 3' OH. La synthèse peut continuer avec l'ajout d'un nouveau nucléotide. Si la synthèse se faisait dans l'autre sens de 3' vers 5', l'énergie de la réaction serait fournie par l'extrémité 5' triphosphate de la chaîne en cours de polymérisation. En cas d'erreur, il y aurait excision du dernier nucléotide et on obtiendrait une nouvelle extrémité 5' monophosphate. On aurait donc besoin d'énergie pour repartir dans la synthèse.

a) Les ADN polymérase

On distingue les ADN polymérase ADN dépendante (ou ADN polymérase), des ADN polymérase ARN dépendante (ou reverse transcriptase) des ADN polymérase n'ayant pas besoin de matrice (les terminal transférase)

α) ADN polymérase ADN dépendante ou ADN polymérase

Elles fabriquent de l'ADN en prenant comme matrice de l'ADN. Les ADN polymérase ont besoin d'une base déjà hybridée ou amorce. Le plus souvent l'amorce est un oligonucléotide, simple brin qu'on est capable de synthétiser chimiquement. (Homonymes : primer ou, par extension, oligonucléotide).

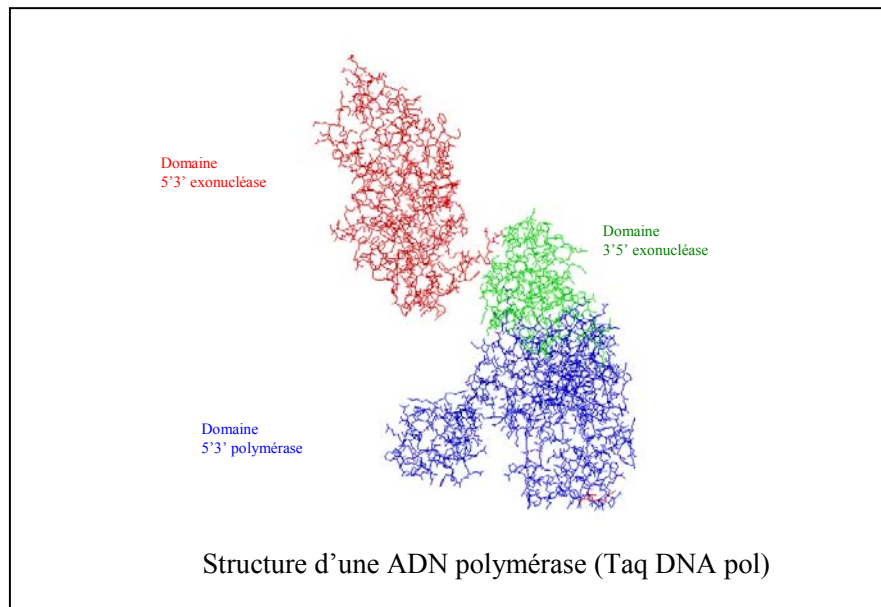
Toutes les ADN polymérase ont :

- Une activité de polymérisation de 5' vers 3', Les DNA polymérase utilisées en biologie moléculaire sont des enzymes utilisées dans la cellule pour les réparations, elles ont donc une processivité faible : elles synthétisent une vingtaine de nucléotides à la suite avant de se décrocher de l'ADN, une autre molécule reconnaît alors le substrat et continue la synthèse pour une vingtaine de nucléotides (Bambara et al., 1978).

- Une activité terminal-transférase sur double brin avec une préférence pour l'incorporation d'une adénine. Si on incube un fragment d'ADN double brin, présentant une extrémité franche avec une ADN polymérase on obtient une extrémité avec un A sortant en 3'.

Certaines DNA polymérase ont :

- Une activité 3'5' exonucléase, cette activité est souvent appelée activité de correction. En effet si la polymérase fait une erreur, le dernier nucléotide n'est plus hybridé, la polymérisation est bloquée. L'activité 3'5' exonucléase retire le nucléotide non hybridé, la polymérisation peut continuer. Cette activité 3'5' exonucléasique retire aussi le A ajouté en 3' par l'activité terminal-transférase
- une activité 5'3' exonucléase



La réaction de polymérisation peut fonctionner à l'envers. Si on ajoute une ADN polymérase, un fragment d'ADN et du pyrophosphate, on obtiendra la réaction inverse : digestion de 3' vers 5' du fragment d'ADN et formation de dNTP.

Exemple d'ADN polymérases utilisées en biologie moléculaire :

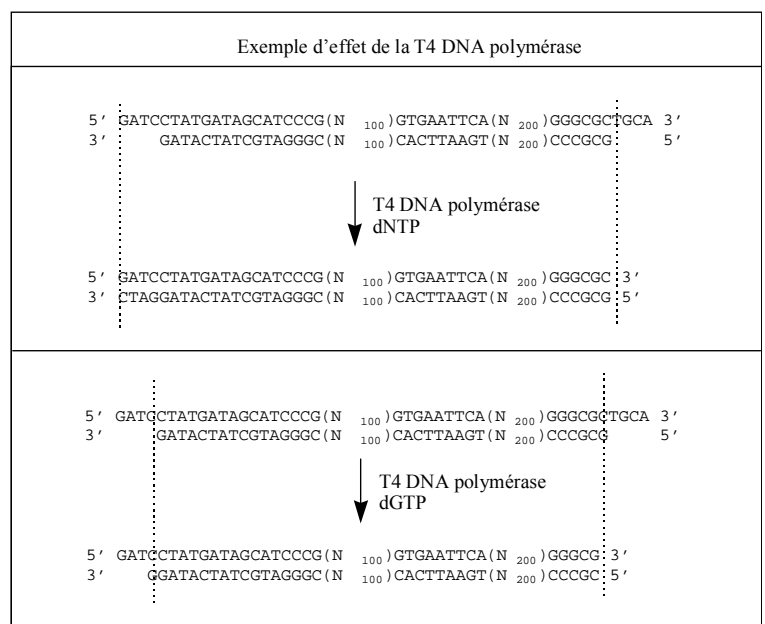
ADN polymérase I de *E. coli* : C'est une enzyme de 100 kDa qui a une activité 5'-3' polymérase, une activité 3'-5' exonucléase (activité de correction) et une activité 5'-3' exonucléase. Cette activité 5'-3' exonucléase de l'ADN polymérase I est souvent gênante. En 1970, Klenow et Henningsen ont enlevé cette activité par protéolyse (subtilisine). Le grand fragment protéique obtenu (76 kDa) est dépourvu d'activité exonucléase 5'-3' mais garde l'activité polymérase et l'activité 3'-5' exonucléase. Depuis, ce grand fragment est obtenu par expression d'un gène tronqué. C'est le fragment « Klenow » de l'ADN polymérase I.

T4 et T7 ADN polymérases :

Elles ont les mêmes activités que la Klenow (5'-3' polymérase et 3'-5' exonucléase). L'activité exonucléase de la T4 ADN polymérase est plus importante que l'activité exonucléase de la Klenow, elle sera donc préférée à la Klenow lorsque cette activité est requise.

On l'utilise par exemple pour rendre franches des extrémités sortantes, qu'elles soient 5' sortantes en utilisant l'activité de polymérisation ou 3' sortantes en utilisant l'activité exonucléase.

L'activité 3'5' exonucléase peut être retirée chimiquement pour améliorer leurs performances lorsque la correction n'a pas beaucoup d'importance, en particulier lorsqu'on ne clone pas ensuite l'ADN.



De plus une mutagenèse dirigée remplaçant une phénylalanine par une tyrosine, permet à la T7 DNA polymérase de mieux reconnaître les didésoxynucléotides. Ces polymérases modifiées commercialisées sous le nom de sequenase® sont utilisées dans les réactions de séquence.

Les ADN polymérases thermostables

La Taq ADN polymérase de *Thermus aquaticus*

Elle a les activités 5'-3' polymérase et 5'-3' exonucléase. Son avantage est d'être thermostable. Comme elle est dépourvue d'activité 3'-5' exonucléase, le taux d'erreurs est d'environ 10^{-5} par base dupliquée et l'activité terminal transférase est efficace.

L'activité 5'-3' exonucléase a été enlevée en déléant l'extrémité N-terminal de la Taq ADN polymérase d'une manière analogue à ce qui a été fait sur l'ADN polymérase I de *E. coli* pour obtenir la Klenow. Cette polymérase a été appelée KlenTaq ou Titanium®Taq (Clontech). Cette polymérase est plus efficace que la Taq ADN polymérase et les produits d'amplification sont plus faciles à insérer dans un vecteur de clonage.

La Pfu et la Pwo DNA polymérase :

Elles proviennent de *Pyrococcus furiosus*, bactérie découverte dans des sources géothermiques en Italie et de *Pyrococcus woesei*. Elles ont les même séquences et ont donc des activités identiques. Activité 5'-3' polymérase et 3'-5' exonucléase mais pas d'activité 5'-3' exonucléase. L'activité 3'-5' exonucléase dite correctrice permet de diminuer le taux d'erreurs à 10^{-6} par base dupliquée.

La Vent ADN polymérase

Elle provient de *Thermococcus littoralis* et a les activités 5'-3' polymérase et 3'-5' exonucléase. Cette dernière activité a été retirée pour permettre une meilleure amplification. Elle est commercialisée par Biolabs.

La KOD1 DNA polymérase :

Elle a été isolée de l'archaebactérie *Thermococcus kodakarensis*. Elle a les activités 5'-3' polymérase et 3'-5' exonucléase.

Comparaison des différentes polymérases thermostables

DNA polymérase	Taq	Pfu, pwo	KOD1
fidélité	$1,3 \cdot 10^{-2}$	$3,9 \cdot 10^{-3}$	$3,5 \cdot 10^{-3}$
vitesse d'élongation (base/s)	60	25	100-150
processivité	20	20	> 300

La fidélité est mesurée comme le rapport du nombre de mutants nuls de la β -galactosidase obtenus après PCR sur un plasmide contenant LacZ.

La processivité est le nombre de bases polymérisées lorsqu'une polymérase est liée à l'ADN.

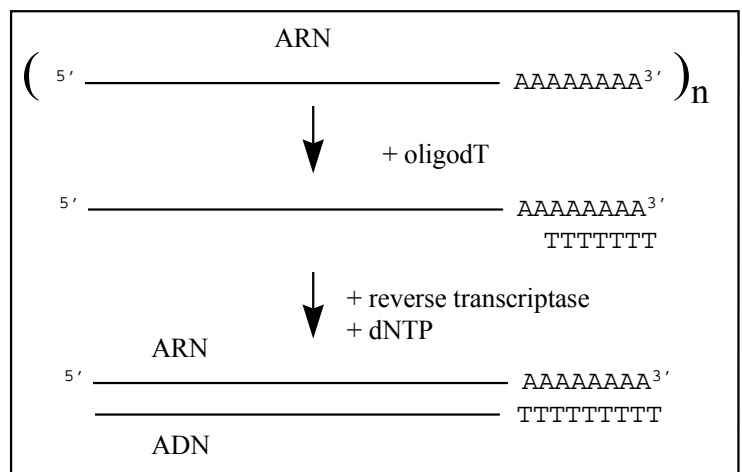
β) ADN polymérase ARN dépendante : reverse transcriptase

Cette enzyme a la même activité que l'ADN polymérase ADN dépendante, mais elle fabrique un ADN à partir d'un ARN. Elle a donc une activité 5'3' ADN polymérase sur un substrat composé d'une amorce hybridée sur un ARN.

L'utilisation principale de cette enzyme est la synthèse d'ADN complémentaire (ADNc) à partir des ARN messagers de cellules eucaryotes.

On peut utiliser trois sortes d'amorce :

- soit un oligo dT dans ce cas on obtient une population d'ADNc.
- soit une amorce au hasard, on obtient alors une population d'ADNc dont l'extrémité 5' est variable
- soit une amorce spécifique à une séquence présente sur un ARN, on obtient alors des ADNc de séquence unique correspondant à un seul gène à partir d'ARN de cellules eucaryotes.



Ces séquences pourront être amplifiées par PCR, l'association successive des deux techniques a pris le nom de RT-PCR.

La Reverse transcriptase manque d'activité 3'5' exonucléase si bien qu'il n'y pas de correction d'erreur. On peut augmenter ces erreurs en présence de Mn^{+} et en forte concentration de dNTP, dans ce cas, une base sur 500 est mal incorporée.

Les reverse transcriptases sont des enzymes fragiles. Pour améliorer la synthèse et obtenir de longs cDNA on peut ajouter des stabilisateurs de protéines tels que le tréhalose (170 mM) et le sorbitol (750 mM) (Carninci et al., 2002).

Les différentes reverse transcriptases :

- MMLV et AMV sont des reverse transcriptases virales.
- Superscript (Invitrogen) : les reverse transcriptases comme la MMLV et l'AMV ont une activité RNase H associée. Cette activité correspond à la coupure de l'ARN hybridé à de l'ADN. Cette activité est donc préjudiciable à synthèse de grand cDNA. Invitrogen a inactivé l'activité RNase H d'une reverse transcriptase et l'a appelé « superscript ».
- Fluorascript (Invitrogen) : cette reverse transcriptase dérivée de la "superscript" incorpore plus facilement les nucléotides fluorescents. Cette activité est intéressante pour fabriquer des sondes à hybrider sur les puces à ADN et ainsi déterminer des patrons d'expression.

γ) Terminal transférase: Cette polymérase n'a pas besoin de matrice comme les autres polymérases, elle ajoute des nucléotides en 3' à l'extrémité du brin d'ADN. C'est une enzyme rare (on n'en parle pas dans les mécanismes généraux de la cellule), elle ajoute des nucléotides en 3' en présence de dNTP. Si on veut en ajouter qu'un seul, on ajoute un ddNTP.

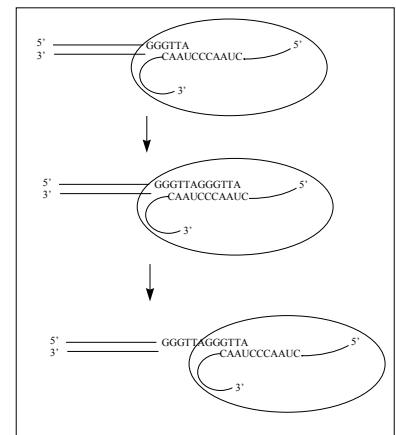
utilisation :

Fabrication d'une queue homopolymérique

Fabrication d'une extrémité cohésive en 3'

δ) Les télomérases

Ce sont des ribonucléoprotéines qui catalysent l'addition de motifs répétés à l'extrémité des chromosomes de séquence TTAGGG. L'ARN présent dans la télomérase sert de matrice pour la synthèse de l'ADN. La télomérase a donc une activité reverse transcriptase. La plupart des cellules somatiques n'ont pas d'activité télomérase et le chromosome se raccourci peu à peu alors que les cellules embryonnaires, les cellules souches ou les cellules cancéreuses ont cette activité.



2) Les ARN polymérase ADN dépendantes

L'ARN polymérase reconnaît un promoteur et synthétise un ARN complémentaire au brin en aval de ce promoteur. La synthèse s'effectue dans le sens 5'-3' en présence de ribonucléotides triphosphates.

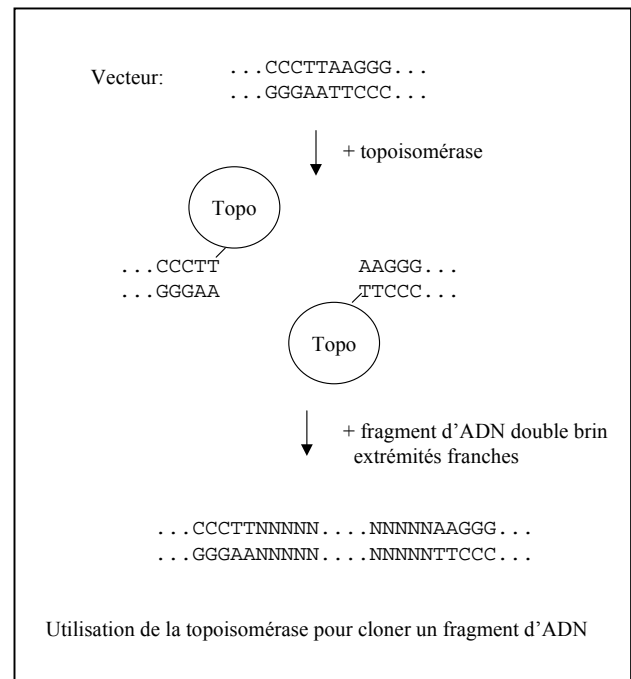
Le gène qui nous intéresse doit donc être cloné derrière un promoteur spécifique pour l'ARN polymérase que l'on veut utiliser. On utilise trois principales ARN polymérases, la T7 ARN polymérase, la T3 ARN polymérase et la SP6 ARN polymérase. Ces trois polymérases sont issues des phages T7, T3 ou SP6 et reconnaissent chacune un promoteur spécifique, ainsi la T7 ARN polymérase reconnaît le promoteur T7 mais pas les promoteur T3 ou SP6.

3) Les topoisomérases

La topoisomérase I du virus de la vaccine coupe en 3' de la séquence (C ou T) CCTT en formant un 3' phosphate-enzyme et un 5' OH. Elle reforme ensuite la liaison phosphodiester.

Si on ajoute deux séquences (C ou T)CCTT sur les deux brins et à proximité, on a coupure des deux brins, la topoisomérase reste attachée par une liaison covalente au brin d'ADN.

En présence d'un fragment d'ADN, elle peut reformer la liaison à condition que la séquence du fragment d'ADN soit 5'OH.



4) Les ligases

T4 ADN ligase

Elle catalyse la formation d'un pont phosphodiester entre un 3' OH et un 5' phosphate, elle a besoin d'ATP et d'ions divalents. Elle ligue de l'ADN double brin.

Utilisation : ligation d'extrémités cohésives ou d'extrémités franches de fragments de restriction. Dans le cas d'extrémités franches, la ligation est souvent effectuée à 15-20°C pendant 4 à 16 heures pour favoriser l'interaction entre les extrémités.

Si les deux extrémités sont déphosphorylées, la ligation ne peut avoir lieu, par contre si une seule est déphosphorylée, la ligation a lieu sur un des deux brins, l'autre reste avec un nick.

Généralement les extrémités cohésives sont générées en coupant par des enzymes de restriction. Le plus souvent, les extrémités à liquer seront générées par la même enzyme mais ce n'est pas obligatoire. Par exemple les enzymes Bam HI (G/GATCC), Bgl II (A/GATCT) et Mbo I (N/GATCN) donnent des extrémités cohésives qui peuvent être ligaturées car elles sont compatibles.

Toutefois après certaines ligations comme celles d'une extrémité Bam HI et d'une extrémité Bgl II on ne pourra pas redigérer le produit par l'une ou l'autre des deux enzymes.

La Taq ADN ligase a la même activité que la T4 ADN ligase, mais elle nécessite du NAD à la place de l'ATP. Elle est plus stable et fonctionne à 45°C. Elle s'utilise dans des méthodes de détection de mutation, lorsqu'on veut faire plusieurs ligations successives entrecoupées d'une étape de dénaturation

T4 ARN ligase

Catalyse la jonction entre un 5' phosphate d'un ARN ou d'un ADN simple brin avec un 3' OH (ions divalents et ATP)

ADN ligase d'E. coli

Ligue des extrémités cohésives double brin d'ADN. L'ADN ligase d'E. coli est inefficace pour liguer des extrémités franches ou pour liguer de l'ARN. Elle utilise du NAD comme cofacteur.

5) Les phosphatases et kinases

Phosphatases

La plus utilisée est la phosphatase alcaline bovine (CIP). Elle catalyse le retrait du phosphate en 5'. Elle a besoin de zinc et d'un pH entre 9 et 10 pour fonctionner et est stimulée par le magnésium. Elle est thermolabile, elle peut être inactivée par incubation à 65°C pendant une heure.

On peut aussi utiliser une phosphatase alcaline de crevette (SAP), cette phosphatase n'a pas besoin de zinc pour être active et peut donc être utilisée dans la plupart des tampons compatibles avec les enzymes de restriction. Elle est encore moins stable que la CIP et est inactivée par une incubation de 15 mn à 65°C.

Les phosphatases alcalines sont utilisées par exemple pour déphosphoryler les vecteurs avant de liguer un insert pour éviter sa recircularisation.

T4 polynucléotide kinase

Transfert le phosphate en γ de l'ATP sur le 5'OH de l'ADN (double ou simple brin) ou sur l'ARN.

Utilisation : marquage de l'extrémité 5' de l'ADN, marquage des oligonucléotides

6) Les méthylases

Les enzymes de restriction sont produites par des bactéries pour se protéger des phages. Pour que l'ADN bactérien ne soit pas lui-même digéré, la bactérie produit aussi des méthylases spécifiques qui inhibent la digestion. Ainsi Eco RI ne coupe pas $GA^m ATTC$. On pourra utiliser ces méthylases pour inhiber la digestion d'un fragment.

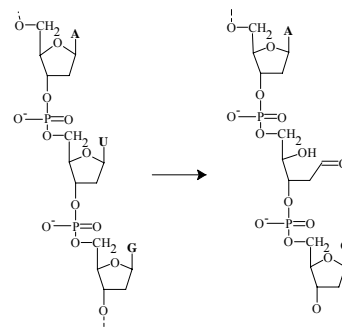
Les enzymes de restriction reconnaissant la même séquence ne sont pas sensibles aux mêmes méthylations. On pourra donc détecter des zones sur l'ADN qui sont méthylés en digérant avec deux enzymes, une sensible et l'autre non. Les produits seront déposés sur un gel d'électrophorèse et la digestion sera analysée après un Southern. Si la bande est détectée uniquement avec l'enzyme insensible à la méthylation, on peut en déduire que la zone sondée est méthylée.

7) Les glycosidases

L'Uracile DNA-glycosidase enlève les uraciles présents dans l'ADN.

Cette enzyme est principalement utilisée dans les typages moléculaires à l'aide de PCR. Dans ces expériences, une des principales limitations est la contamination par les produits d'amplification des PCR précédentes. Si les amplifications sont faites à partir d'uracile à la place de thymine, l'ADN

néosynthétisé contient de l'uracile alors que l'ADN à analyser ne contient pas d'uracile. Comme



l'ADN des amplifications précédentes peut polluer les nouveaux échantillons à analyser, un traitement à l'uracile glycosidase permet de l'éliminer.

Les vecteurs

Rôle d'un vecteur : capable de transporter l'ADN d'intérêt (d'où le nom de vecteur), c'est un morceau d'ADN capable d'auto-réplication. Il doit être petit de façon à être manipulé *in vitro*

Pour satisfaire à ces deux exigences, on utilise soit des plasmides soit des virus, soit des petits chromosomes.

1) Les plasmides

Les plasmides sont de petits morceaux d'ADN circulaires, double brin que l'on trouve dans les bactéries en dehors du chromosome (épisomes). Ils se répliquent grâce aux enzymes présents dans la bactérie.

On est capable de les introduire dans une bactérie par des méthodes chimiques (CaCl_2) ou des méthodes physiques (électroporation). C'est la transformation bactérienne (notion différente de la transformation des cellules eucaryotes).

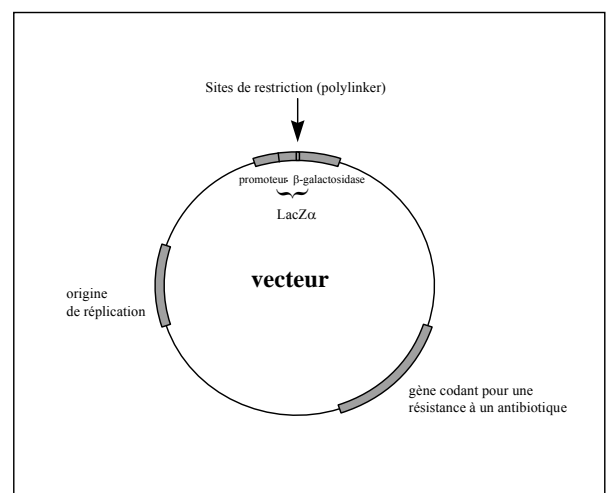
Les parties essentielles :

Une origine de réplication, importante pour l'initiation de la réplication.

Un site de restriction pour pouvoir insérer un fragment d'ADN (ce fragment d'ADN est souvent appelé insert).

Un gène de résistance aux antibiotiques : permet de

sélectionner les bactéries ayant incorporés le plasmide de celles qui ne l'ont pas incorporé. La transformation est un événement rare, il n'y a qu'un faible nombre de bactéries qui sont transformées,



il n'y a donc qu'un plasmide par bactérie. Si on isole cette bactérie et qu'on la fait pousser, on obtient un clone c'est à dire un grand nombre de bactéries identiques portant le même plasmide. Par extension, le terme de clone est souvent donné au plasmide lui-même.

Les parties accessoires :

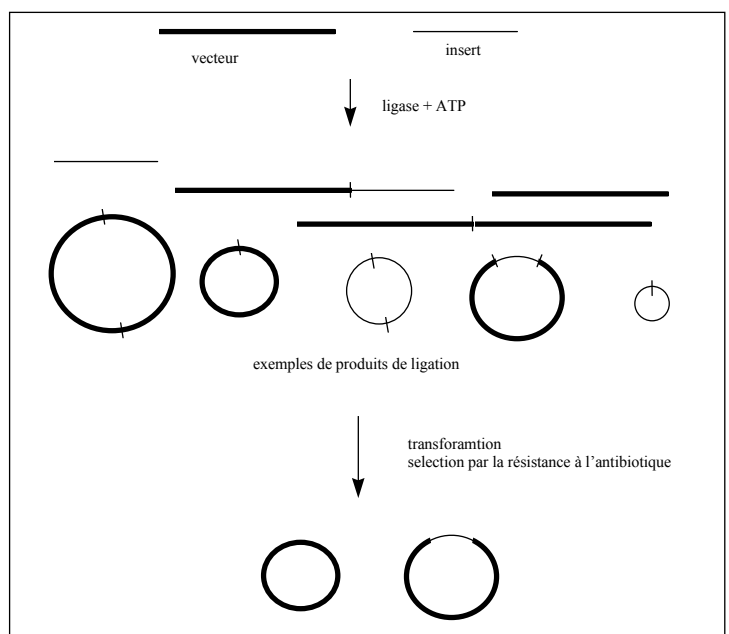
-> Un polylinker ou site de clonage multiple.

L'insertion d'un insert dans un plasmide lorsqu'il n'y a qu'un seul site de restriction n'est pas facile. En effet, l'insert n'est pas obligatoirement borné par le site de restriction présent sur le plasmide. De plus, le vecteur peut se recirculariser (à moins qu'on le déphosphoryle) et l'insert n'est pas orienté. Pour faciliter les clonages, on a inséré dans le vecteur une séquence présentant une série de sites de restriction uniques dans le vecteur, les uns à la suite des autres. A l'origine, pour fabriquer ces polylinker, on ajoutait par ligation une série de petits oligonucléotides double brin contenant un ou plusieurs sites de restriction. Ces oligonucléotides ont le nom de linker parce ce qu'on peut les utiliser pour ajouter des sites de restriction aux extrémités d'un fragment d'ADN, les sites de clonage multiples ont donc pris le nom de polylinker.

-> Un gène codant pour le peptide α de la β -galactosidase.

Lorsqu'on veut insérer un fragment d'ADN dans un plasmide, on digère le vecteur avec un ou deux enzymes de restriction donnant des extrémités compatibles avec les extrémités de l'insert. Mais cette digestion n'est jamais totale. On a donc dans la solution des molécules de vecteur non digérées.

De plus, lors de la ligation des deux ADN, vecteur et insert, une partie du vecteur se recircularise si les deux extrémités sont compatibles et si la déphosphorylation n'a pas été totale. Après la transformation, on a



donc des bactéries qui ont un vecteur seul et des bactéries qui ont le plasmide recombinant. Pour trier ces bactéries, on a inséré le site de clonage à l'intérieur d'un gène codant pour le peptide α de la β -galactosidase. Le clonage de l'insert crée une mutation par insertion, inactivant le peptide. Si le plasmide est introduit dans une souche déficiente pour le peptide α , le vecteur seul permet la synthèse de galactosidase par contre le vecteur plus l'insert ne le permet pas. Les deux colonies peuvent être facilement triées en utilisant des substrats colorés de la β -galactosidase (X-gal par ex).

De plus, si on clone un fragment d'ADN codant pour une protéine, on a une chance sur trois que la phase de lecture de l'insert soit en concordance avec celle du peptide α . Une fois sur six, si le clonage est non orienté, on aura production d'une protéine de fusion comprenant en N-terminal le peptide α suivi du peptide codé par l'insert. Ce peptide pourra alors être reconnu par un anticorps.

-> Promoteurs d'ARN polymérases.

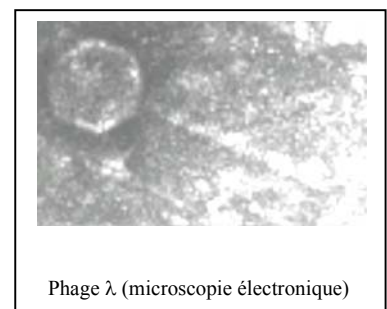
De chaque côté du site de clonage, on a souvent rajouté les promoteurs de la T3 ou T7 ARN polymérase. On peut donc fabriquer des ARN sur des plasmides ouverts en aval de l'insert en incubant l'ADN avec l'ARN polymérase en présence de ribonucléotides (transcription *in vitro* de type « run off »). Ces promoteurs servent donc entre autre à faire des sondes ARN ou pour faire des ARNc qui pourront être traduits *in vitro*. On peut également faire ainsi des ARN anti-sens.

-> Plasmide d'expression (voir plus loin).

2) Les phages

Le phage λ

Ce phage est un des modèles de laboratoire au même titre que la souris, la drosophile ou le nématode *Caenorhabditis elegans*. Les connaissances accumulées sur ce phage ont permis de l'utiliser comme vecteur. C'est un virus à ADN double brin, linéaire de 50 kb avec à ses deux extrémités 12 nucléotides simple brin complémentaires (extrémité cohésives ou cos). Il se fixe sur le



Phage λ (microscopie électronique)

récepteur lamB qui permet l'absorption de maltose par la bactérie. Seul l'ADN rentre dans la bactérie. Une fois dans la bactérie, les extrémités cohésives s'hybrident et l'ADN est ligué par une ligase d'*E. coli*. Là, il y a deux possibilités, soit un cycle lysogénique avec intégration de l'ADN du phage dans le génome soit un cycle lytique. C'est ce cycle lytique qui est utilisé dans les vecteurs. On incube donc une solution de phage avec une solution de bactérie. Après une quinzaine de minutes, les phages sont rentrés dans les bactéries et on étale, sur un milieu de culture solide dans une boîte de pétri, la solution comportant ces bactéries dont certaines ont incorporé un phage. Le phage lyse la bactérie, puis attaque les bactéries les plus proches par diffusion. Lorsque la culture bactérienne arrive à confluence (tapis bactérien sur la boîte), les bactéries n'expriment plus certaines protéines de surface comme les porines ou le récepteur maltose. En l'absence de récepteur, les phages ne peuvent plus attaquer les bactéries. On obtient un tapis bactérien avec une plage de lyse correspondant à la descendance d'un seul phage, c'est à dire un clone.

Le phage λ en tant que vecteur :

Dans le phage λ , la longueur de 50 kb de son ADN est importante pour l'empaquetage. Les 20 kb au milieu du génome ne sont pas importants pour le cycle lytique mais nécessaires pour le cycle lysogénique. Pour l'utiliser en tant que vecteur on retire ces 20 kb et on insère à la place le fragment d'ADN d'intérêt. On a alors un bras gauche 20 kb et un bras droit 10 kb. La purification des bras s'effectue par centrifugation zonale, les brins d'ADN ont en effet la même densité mais des masses différentes.

La sélection des recombinants peut être effectuée par plusieurs méthodes:

Physique : si on purifie les bras droit et gauche, ils sont trop petits pour être insérés dans la capsid qui n'admet que des faibles variations de taille (40-55kb).

Génétique: on peut toujours avoir les 20 kb d'origine si la purification des bras n'a pas été bien faite. Plusieurs gènes ont été utilisés pour éviter cela, visant à trier les non recombinants dans un cycle lysogénique et les recombinants dans un cycle lytique. Par exemple lors d'un clonage avec un phage contenant dans sa partie centrale le gène cI, les non recombinants sont lysogènes (dans une souche hfl, high frequency of lysogenization) et les recombinants rentrent dans un cycle lytique.

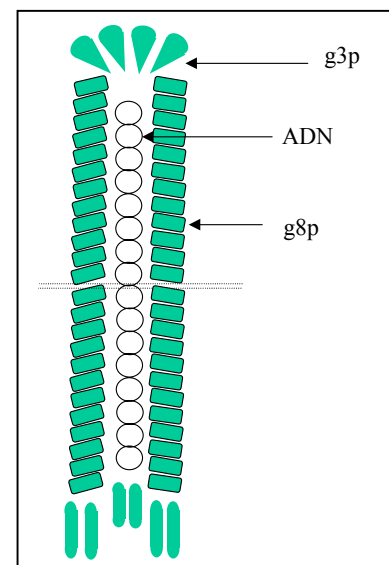
Intérêt du phage λ :

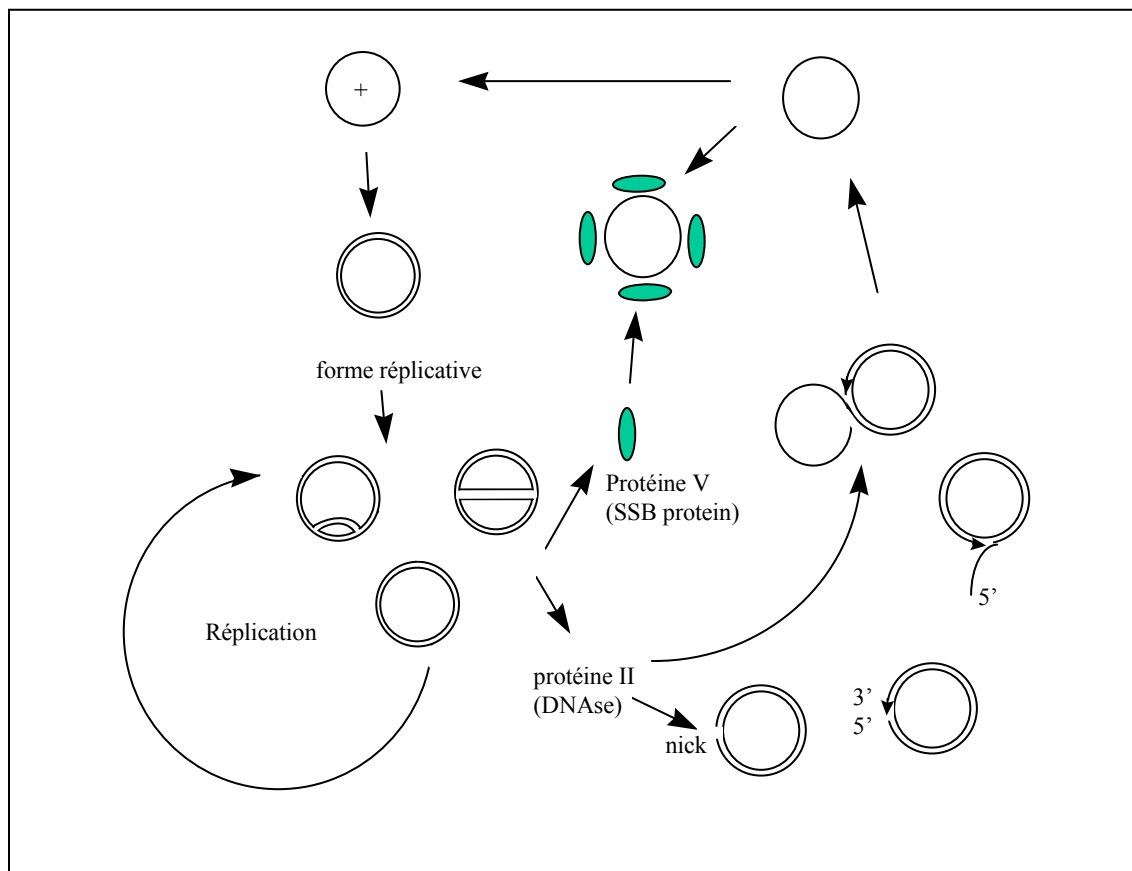
- On dispose d'un moyen très efficace pour faire rentrer l'ADN dans la bactérie. En effet, un virus infecte une cellule bien plus efficacement qu'une méthode de transformation physico-chimique. Il faut toutefois faire au préalable un empaquetage *in vitro* qui consiste à rassembler dans un tube l'ADN, la tête et la queue du phage.
- Le criblage des banques avec un anticorps ou avec une sonde nucléotidique est plus aisé que lorsqu'on utilise un plasmide. Les plages de lyse sont plates et il est plus facile de faire une réplique sur ces boîtes que sur une population de colonies bactériennes.

Utilisation du phage λ : principalement pour faire des banques, on profite ainsi des plages de lyse qu'il est plus facile à cribler que des colonies. On l'utilise pour des banques génomiques dans ce cas les inserts font de 15 à 20 kb, ou pour des banques d'ADNc. Dans ce dernier cas, les inserts sont plus petits (<10 kb) et n'influencent pas l'empaquetage *in vitro*.

Le phage M13

C'est un bactériophage de 6.4 kb, ADN simple brin, qui contient une dizaine de gènes. Il infecte seulement les bactéries qui expriment le pilus sexuel codé par le facteur F. Il pénètre par le pilus, la protéine de structure g8p est retirée et l'ADN est transféré dans le corps principal de la bactérie. Ce brin d'ADN (appelé brin +) est alors converti en double brin circulaire appelé ADN RF (forme répliquative). Cette conversion est due à des enzymes bactériens, une ARN polymérase initie la réplication qui est effectuée par une ADN polymérase. La transcription des gènes viraux peut alors avoir lieu. La protéine produit du gène II introduit un nick à un site spécifique du brin +, plus il y a de formes répliquatives, plus il y a de protéines produit du gène II, si bien qu'à partir d'une certaine quantité de plasmide, ceux-ci sont tous ouverts. Une ADN polymérase ajoute des nucléotides en 3' en déplaçant le brin + original (Rolling circle). Une fois qu'un tour a été fait, le produit du gène II coupe au même endroit, libérant un brin linéaire qui est alors recircularisé.





Au début de l'infection ce brin est de nouveau transformé en forme répliquative mais lorsqu'il y a beaucoup de forme répliquative, le produit du gène V (SSB single strand binding protein) s'accumule. Il y a production presque uniquement de forme simple brin. La balance entre les deux formes, double brin et simple brin dépend donc de la concentration en ADN double brin du phage, qui après transcription et traduction, produit la protéine II responsable du nick à l'origine de la réplication en rolling circle et produit les SSB protéines qui stabilisent le simple brin empêchant la synthèse du brin complémentaire.

L'ADN n'est pas encapsidé dans une structure préformée comme les autres bactériophages. Ils sont simplement couverts des protéines de la capsid lorsqu'ils sortent de la bactérie. Ceci implique qu'il n'y a pas de limitation dans la taille de l'ADN simple brin. Il n'y a pas de lyse de la bactérie si bien qu'elle continue à pousser (plus doucement) en produisant des phages qui atteignent le nombre de 10^{12} par ml de culture.

Tous les gènes sont essentiels mais on peut insérer un fragment d'ADN dans la forme répliquative. Un polylinker a été introduit avec la séquence codante pour le peptide α de la β galactosidase comme pour les plasmides. La construction s'effectue dans la forme répliquative avec les méthodes utilisées pour les plasmides, en utilisant les DNases de restriction et la ligase.

Utilisation : lorsqu'on a besoin d'ADN simple brin, séquence (méthode enzymatique), mutagenèse dirigée, obtention de sondes sens spécifique, phage display...

3) Chromosomes

Plusieurs chromosomes peuvent être utilisés comme vecteur

- les YAC (yeast artificial chromosome) sont dérivés des chromosomes de *Saccharomyces cerevisiae*. Ils contiennent une origine de répliquaison, un centromère et deux télomères. Ces vecteurs ont été principalement utilisés pour réaliser des clonages positionnels mais cette application est aujourd'hui abandonnée car elle est difficile à réaliser et qu'il y avait trop de recombinaisons.
- Les BAC (Bacterial artificial chromosome) sont des dérivés du facteur F. Leur taille est d'environ 100 kb, et ils se maintiennent chez *E. coli* comme simple copie.
- Les PAC (P1 artificial chromosome) sont des dérivés du phage P1.

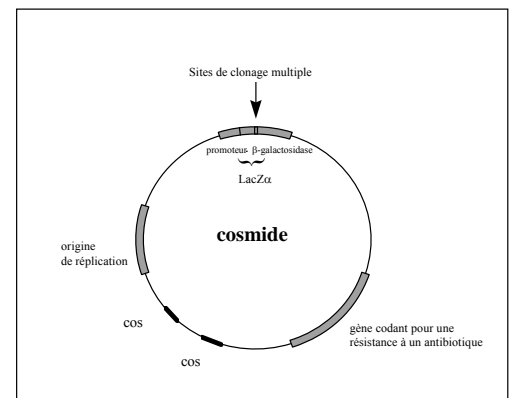
4) Systèmes hybrides

Les trois vecteurs vus jusqu'à présent ont chacun leur avantage, les plasmides sont faciles à manipuler, le phage M13 permet d'obtenir de l'ADN simple brin et le phage λ permet une transformation efficace des bactéries.

Plasmide + M13 = phagemide : On peut introduire une origine de répliquaison simple brin dans un plasmide, on obtient un phagemide (comme la plupart des plasmides ont maintenant une origine de M13, cette dernière appellation tombe en désuétude)

Le phagemide est un plasmide qui contient l'origine de réplique de M13. Cette région de M13 contient le site de coupure de la protéine du gène II pour l'initiation de la réplique en rolling circle, une séquence suffisante pour le packaging de l'ADN simple brin dans les particules du phage, une séquence utile pour la synthèse du brin -, et une séquence de terminaison, site de coupure de l'ADN simple brin et site de recircularisation par la protéine du gène II. On les cultive comme des plasmides mais si on veut du simple brin on infecte la bactérie avec un autre M13. Le M13 produit toutes les protéines pour la réplique en simple brin et l'encapsidation si bien que les deux ADN, du M13 et du plasmide sont encapsidés. En purifiant le virus on obtient les deux ADN. Pour obtenir une plus grande quantité de simple brin du plasmide par rapport à celui du phage on utilise un phage muté dans la zone de coupure par la protéine du gène II comme le M13K07. On appelle ces phages « phages Helper ». Ils ne peuvent pas eux-mêmes être encapsidés.

Plasmide + phage λ = cosmide : On ajoute deux extrémités cos à un plasmide. Si on clone dans ce plasmide un grand fragment d'ADN (d'environ 45 kb) de telle façon que plasmide et insert font 50 kb, cet ADN pourra être encapsidé *in vitro* dans le phage λ . Il pourra donc être introduit avec une grande efficacité dans la bactérie. Par contre une fois à l'intérieur, il se répliquera comme un plasmide. Les extrémités cos permettent donc de sélectionner les recombinant avec de grands inserts (45 kb) et à introduire le plasmide dans la bactérie.



Plasmide + phage λ + phage M13 = λ Zap : (Short et coll., 1988) On a ici un hybride entre trois vecteurs.

Un plasmide est inséré dans le phage, il est inséré au niveau de l'origine de réplique de M13 entre les domaines d'initiation et de terminaison de la réplique simple brin. En présence de M13 ou de fl, la protéine du gène II initie la réplique en coupant le brin + (nick) et la polymérisation a lieu jusqu'au terminateur. A ce site l'ADN est coupé et circularisé par la protéine du gène II. L'ADN simple brin est converti en double brin comme M13. Pour récupérer le plasmide soit on effectue une préparation d'ADN plasmidique que l'on transforme ensuite soit on récupère les virus M13, on effectue une infection d'E. coli à une température non permissive pour le phage helper. Dans ce

dernier cas on utilise un phage helper thermosensible comme R408 (Russel et coll., 1986). Le virus injecte son ADN dans la bactérie, son ADN est transformé en double brin (forme répliquative) mais en absence de phage helper, n'est plus converti en simple brin.

Les marqueurs de sélection

On utilise souvent des antibiotiques pour sélectionner les bactéries transformées. Selon les molécules, elles seront généralement efficaces pour sélectionner les procaryotes ou les eucaryotes, mais certaines d'entre elles seront efficaces pour les deux types cellulaires et pourront donc être utilisées dans les vecteurs utilisés successivement en procaryote et en cellule eucaryote (on parle de vecteur navette).

On peut utiliser des antibiotiques toxiques pour les bactéries et un gène de résistance porté par le plasmide pour sélectionner les bactéries qui ont été transformées. On peut aussi utiliser des protéines toxiques.

1) Antibiotiques Procaryotes

En dehors des micro-organismes du genre *Penicillium*, les bactéries du genre *Streptomyces* produisent environ deux tiers de tous les antibiotiques connus. Presque toutes les classes d'antibiotiques sont synthétisées : les macrolides, les cyclines, les aminoglycosides, les β -lactamines, etc... Depuis les années 60, de nombreux antibiotiques sont obtenus par synthèse ou semi-synthèse.

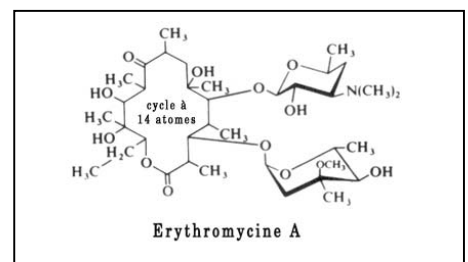
a) Antibiotiques dirigés contre la sous-unité 50S

α) Les Macrolides

Les macrolides sont constitués par un macrocycle porteur d'une fonction lactone, sur laquelle viennent se greffer deux ou plusieurs sucres dont l'un est aminé. En raison de la présence d'une et parfois de deux amines, les macrolides sont des molécules basiques.

Il se fixe au niveau du site P. Ils empêchent ainsi le transfert du complexe peptidyl-ARNt depuis le site P vers le site A, ce qui entraîne une inhibition de l'élongation de la chaîne peptidique.

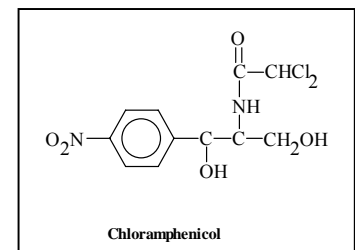
L'érythromycine est un composé de type macrolide, isolé aux Philippines de *Streptomyces erythreus*. D'autres composés sont semi-synthétiques : azithromycine, clarithromycine, roxithromycine...



β) Chloramphénicol

Il se lie aux ribosomes et inhibe la peptidyl transférase. Il ne reconnaît pas les ribosomes eucaryotes sauf ceux des cellules érythroïétiques des mammifères mais il reconnaît les ribosomes des mitochondries.

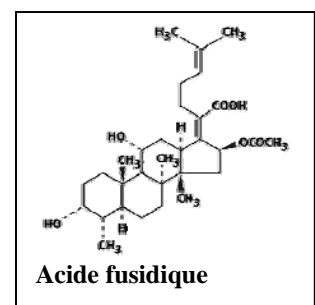
Le mécanisme de résistance est métabolique, la chloramphénicol-acétyltransférase (CAT) inactive le chloramphénicol en utilisant l'acétyl-CoA comme cofacteur.



γ) L'acide fusidique

Initialement isolé à partir du champignon *Fusidium coccineum* (Deuteromycète) d'où son nom, on le trouve aussi chez *Mucor ramannianus* (Zygomycète) et *Isaria kogana* (Deuteromycète).

Il se fixe au facteur d'élongation EF-G et bloque ainsi la fixation des amino-acyl-tRNA.

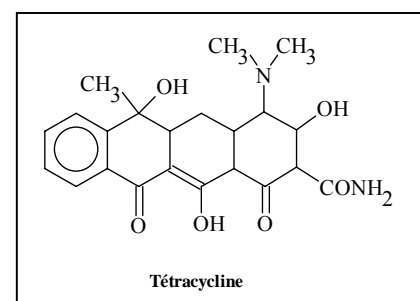


b) Antibiotiques dirigés contre la sous-unité 30S

α) Les Tétracyclines

(tétracycline, doxycycline, minocycline, oxtétracycline, ...)

Elles bloquent la traduction en empêchant la liaison de l' aminoacyl-ARNt à la sous unité 30S. La spécificité d'action provient de la présence de transporteurs actifs présents uniquement sur la membrane cytoplasmique des procaryotes.

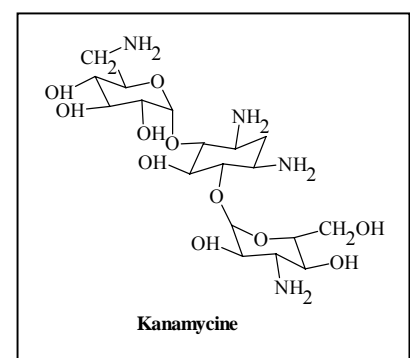


β) Les Aminoglycosides

(Streptomycine, kanamycine, néomycine, tobramycine, amikacine, gentamicine...)

Les premiers aminosides découverts sont des molécules naturelles, produites par des souches de *Streptomyces* (streptomycine, néomycine, kanamycine, tobramycine- avec un "y") ou d'*Actinomyces* (gentamicine, sisomicine – avec un "i").

Mode d'action : ils se lient au ribosome procaryote au site A et inhibent la phase d'élongation. La liaison entraîne des erreurs dans



la traduction.

La kanamycine.

Le gène de résistance code pour une kanamycine nucléotidyltransférase qui transfère un nucléotide sur le résidu amino-glucose et ainsi inactive l'antibiotique.

γ) Viomycine

Cet antibiotique se lie comme les aminoglycosides au site A du ribosome procaryote. C'est un peptide cyclique composé d'acides aminés non codés et d'arginine, sérine et lysine, des acides aminés souvent impliqués dans les interactions protéines-ARN.

c) Antibiotiques dirigés contre la paroi bactérienne

α) β -lactamines

(Penicillines, cephalosporines...)

Mode d'action : inhibiteur de la synthèse des parois bactérienne.

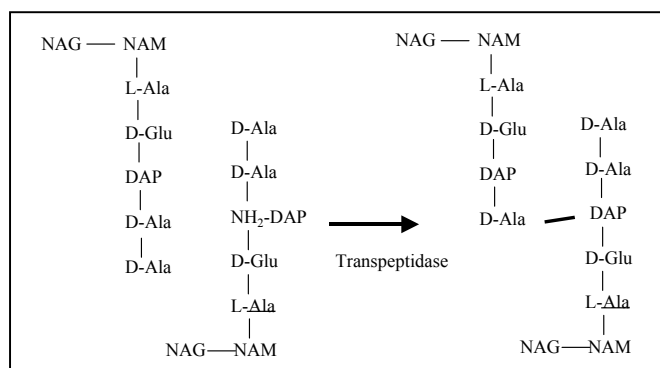
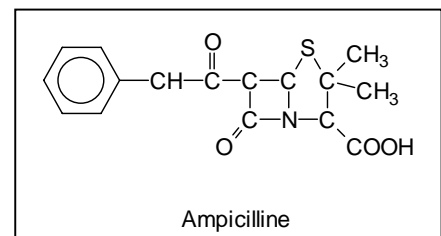
Le noyau β -lactame mime la structure 3D du motif D alanyl-alanine de l'extrémité de la chaîne pentapeptidique du

peptidoglycane de la paroi. L'ampicilline se fixe sur la transpeptidase et empêche la polymérisation de la paroi.

Comme la paroi n'est synthétisée qu'en phase exponentielle, cet antibiotique n'est actif que pendant la croissance bactérienne.

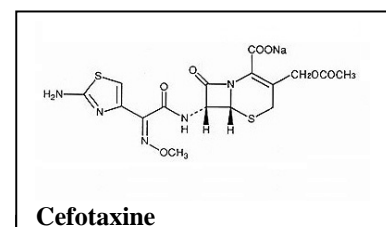
Le gène de résistance (gène *bla*) code pour une

β -lactamase qui coupe la liaison amide du noyau β -lactame ce qui l'inactive.



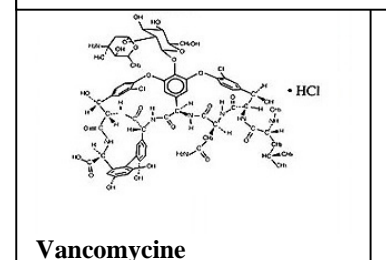
β) Cefotaxime

Mode d'action : inhibe la synthèse des mucopeptides de la paroi bactérienne. Sensible aux β -lactamases.



γ) Vancomycine

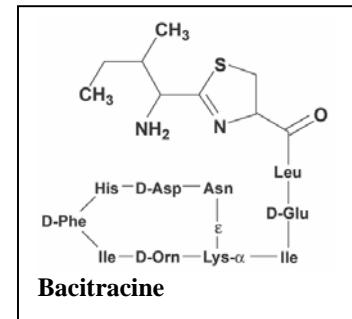
Glycopeptide qui inhibe la formation des polymères de peptidoglycane de la paroi bactérienne.



Il est utilisé principalement pour les agrobactériums qui ont une forte expression de β -lactamase.

δ) *Bacitracine*

La bacitracine inhibe la synthèse de la paroi bactérienne chez de nombreuses bactéries Gram+.

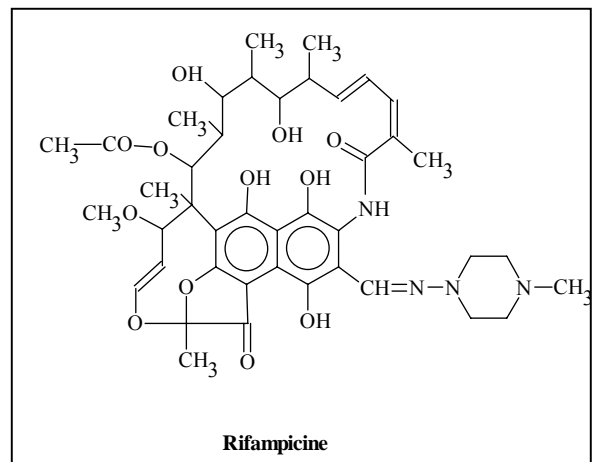


d) Antibiotiques dirigés contre la RNA polymérase

Rifampicine

La rifampicine B est produite par *Streptomyces mediterranei*. Elle se fixe sur la sous unité β de l'ARN polymérase ADN dépendante et bloque l'initiation de la transcription. La résistance est due à une activité ADP-ribosyl transférase.

On l'utilise principalement lorsqu'on fait une production de protéine dans *E. coli* en utilisant l'ARN polymérase T7. En effet l'ARN polymérase T7 est insensible à la rifampicine. Dans ce système,

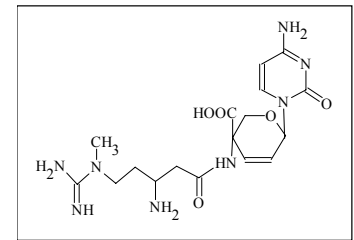


le gène d'intérêt est sous le contrôle d'un promoteur de l'ARN polymérase T7 et cette polymérase est elle-même sous le contrôle d'un promoteur inductible. Lorsque l'on fait une production, on produit tout d'abord l'ADN polymérase T7 puis on inhibe les ARN polymérases bactériennes à l'aide de la rifampicine. Les bactéries ne transcrivent plus que le gène d'intérêt et ne traduisent donc que la protéine recherchée.

2) Antibiotiques Eucaryotes

Aminopteridine et acide mycophénolique

Gène de résistance : xanthine-guanine phosphoryl transférase
l'aminopteridine et l'acide mycophénolique bloquent la synthèse de GTP.



acide mycophénolique

Blasticidine

Utilisé pour sélectionner les cellules d'insectes

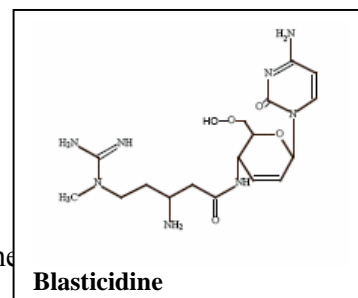
Bromodesoxyuridine, glancicovir

Gène de résistance : thymidine kinase. La thymidine kinase catalyse la synthèse de la thymidine monophosphate à partir de la thymidine et donc de la thymidine triphosphate, nucléotide indispensable pour la synthèse de l'ADN et donc pour la réplication et la viabilité de la cellule.

Elle phosphoryle la drogue qui s'incorpore dans l'ADN.

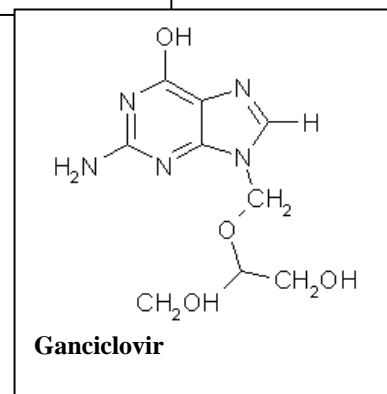
Si on utilise des cellules TK⁻ (mutées sur leur gène cellulaire), les recombinants seront TK⁻ et les parents seront TK⁺ (due à l'expression de la thymidine kinase virale). Or on sait trier les cellules TK⁻ des cellules TK⁺, les cellules TK⁻ sont résistantes à la 5-bromodésoxyuridine. En effet pour être active, cette drogue doit être phosphorylée par la thymidine kinase. Le problème est que la mutation de TK⁺ vers TK⁻ a une fréquence de 10⁻⁴, aussi on associe souvent une autre méthode.

Le ganciclovir est un substrat pauvre pour la thymidine kinase de mammifère.



Blasticidine

une des voies de

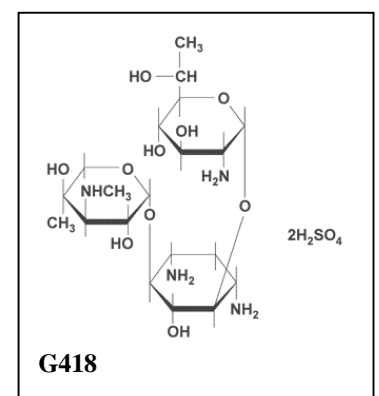


Ganciclovir

G418

Mode d'action : le G418 interfère avec la sous unité 80S du ribosome et bloque la synthèse protéique.

Gène de résistance : Le gène de résistance qui est utilisé est celui de la néomycine qui est proche du G418. Il code pour une aminoglycosyl



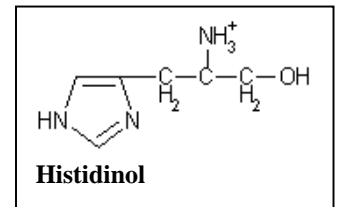
G418

phosphotransférase qui dégrade le G418. Cette enzyme dégrade aussi la kanamycine permettant d'utiliser le même gène pour la sélection dans les cellules procaryotes et eucaryote.

Histidinol

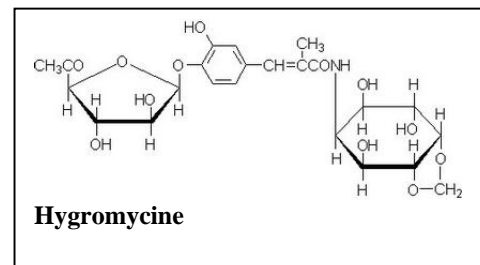
Gène de résistance : histidinol deshydrogénase (*hisD*)

l'histidinol est cytotoxique, l'histidinol déshydrogénase oxyde l'histidinol en histidine.



Hygromycine B

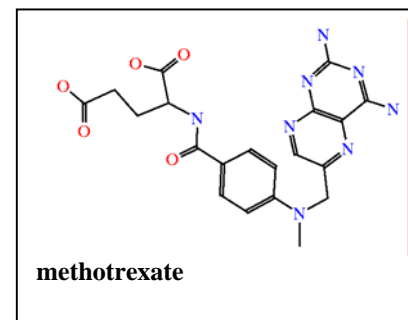
L'hygromycine B bloque la synthèse des protéines en bloquant la translocation. Elle fonctionne sur de nombreuses souches bactériennes, des champignons filamenteux et des cellules eucaryotes. Le gène de résistance est le gène *hph* (hygromycine phosphotransférase).



Methotrexate

Gène de résistance : dihydrofolate réductase (*DHFR*)

- résistante au méthotrexate qui peut alors être ajouté au milieu de culture (la DHFR résistante a été trouvée dans les cellules cancéreuses résistantes au méthotrexate).



Mitomycine C

C'est un antibiotique produit par des bactéries du sol du genre des Streptomyces qui exerce sa toxicité en méthylant et en effectuant des liaisons covalentes sur l'ADN. Le gène de résistance *mcrA* code pour une hydroquinone oxydase qui oxyde la forme inactive réduite de la mytomicine.

Oubaine

gène de résistance : Na⁺, K⁺-ATPase α subunit)

L'oubaine bloque le transport des ions Na⁻ et K⁺ par la membrane cellulaire

PALA (N-(phosphonoacetyl)-L-Aspartate)

Gène de résistance : cytosine désaminase

Le PALA bloque la synthèse *de novo* des pyrimidines, la cytosine désaminase converti la cytosine en uracile.

Phléomycine

Gène de résistance *Sh ble* : se fixe sur l'antibiotique et bloque son action.

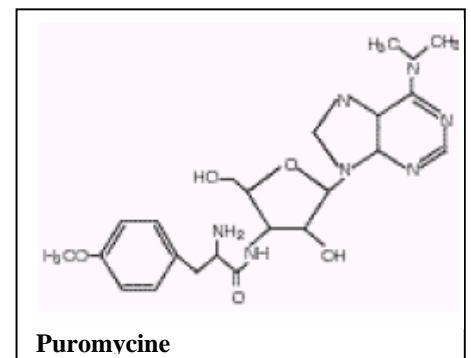
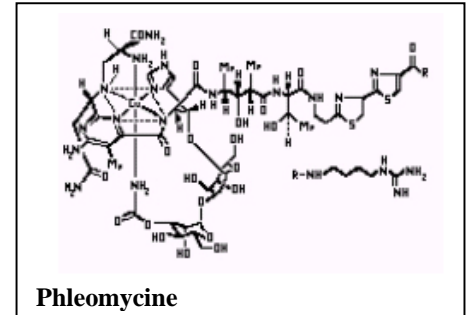
La zéocyne a les mêmes propriétés que la phléomycine.

Puromycine

Gène de résistance : puromycine N-acétyl transférase (*pac*) (*puro*)

La puromycine bloque la synthèse des protéines. Certaines cellules sont intrinsèquement résistantes à la puromycine *via* la surexpression de gènes codant pour sa MDR (multi drug résistance).

Dans ce cas, on peut ajouter de la cyclosporine qui inhibe les MDR et ainsi restore la sensibilité de la cellule à la puromycine (Zhen et Lyons, 2001).



Xyl-A, b-D-Xylofuranosyl adénine

gène de résistance adénosine désaminase (*ada*)

Xyl A est converti en Xyl-ATP qui endommage les acides nucléiques

3) Protéines toxiques

Certains gènes codent pour des protéines toxiques pour *E. coli* et sont utilisés comme moyen de sélection positive. Si la protéine est produite la bactérie ne peut pas se développer par contre si un fragment d'ADN est inséré dans le gène, la protéine est inactivée et la bactérie peut se développer.

- *ccdA* et *ccdB* sont des gènes qui contribuent à la stabilité du plasmide F chez *E. coli* en tuant les bactéries filles qui n'ont pas hérité du facteur F. *ccdB* code pour un peptide de 101 acides aminés qui agit en augmentant le clivage de l'ADN par la gyrase.
- La Barnase est une RNase.

Pratiquement la préparation d'un vecteur comportant un gène de sélection positive nécessite une souche résistante et son utilisation nécessite une souche sensible.

4) La complémentation

Pour trier les transformant, on peut utiliser une complémentation d'auxotrophie. On utilise une souche incapable de synthétiser un acide aminé ou un nucléotide. Sur milieu minimum, la souche ne pourra pas pousser. Si on la transforme avec un vecteur comportant le gène déficient, la souche pourra pousser.

Cette méthode est plus lourde, car il faut utiliser des milieux de culture définis, on obtient des faux positifs dus à des révertants, des mutants qui ont corrigé la mutation de départ. Toutefois elle présente l'avantage de ne pas utiliser de molécules toxiques, ce qui est intéressant lorsqu'on désire produire une protéine à but thérapeutique.

Les souches

La plupart des souches *d'Escherichia coli* utilisées en biologie moléculaire proviennent de la souche K-12 isolée en 1922 d'un patient atteint de diphtérie à Pablo Alto. Cette souche a tout d'abord servi à des études génétiques parce qu'elle était facile à cultiver, qu'elle avait un temps de génération court. Son maintien en laboratoire pendant une cinquantaine d'année s'est accompagné d'une dérive génétique si bien que les antigènes K et O caractéristiques de *E. coli* étaient perdus en 1950. En 1975, il a été montré que la souche était incapable de coloniser le tractus digestif humain (Smith, 1975). Cette souche a été étudiée extensivement en génétique si bien que de nombreuses mutations ont été isolées.

Les mutations d' *E. coli* utilisées en biotechnologie

dam⁻ : méthylation sur les séquences GATC abolies. Les souches *dam*⁻ sont difficiles à transformer avec des plasmides méthylés provenant de souche *dam*⁺.

(DE3) : Insertion du gène de la l'ARN polymérase T7 sous le contrôle du promoteur LacZ (dans le site d'attachement de lambda).

dcm⁻ : méthylation à la séquence CCWGG abolie.

dut⁻ : les souches porteuses de cette mutation sont déficientes en UTPase et contiennent beaucoup d'UTP qui entre en compétition avec le TTP. Cette mutation est utilisée en association avec la mutation ung lorsque l'on désire avoir un ADN comportant des uraciles à la place des thymines, par exemple dans certaines méthodes de mutagenèse.

$\Delta(lac-proAB)$: Délétion chromosomique de l'opéron lac et des gènes impliqués dans la synthèse de la proline. Les souches portant cette mutation ne peuvent utiliser le lactose comme source de carbone et ont besoin de proline pour leur croissance.

$\Delta(proAB)$: Déletion des gènes intervenant dans la biosynthèse de la proline. ProAB est souvent complémenté par un gène sur F', ce qui permet de sélectionner les souches ayant gardé le facteur F' en les étalant sur milieu minimum.

$EndA^-$: le gène EndA code pour une endonucléase qui coupe l'ADN de manière non spécifique. La quantité de plasmide produite par une souche $endA^-$ est donc augmentée.

gor 522 : Tn10 (Tc^R) : inactivation de la glutathion reductase par Tn10. Le glutathion n'est plus réduit par la réductase, ce qui abaisse le pouvoir réducteur du milieu intracellulaire. Cette mutation favorise l'établissement de ponts disulfures des protéines recombinantes produites dans E. coli (en association avec *trxB*).

hsdR : Une souche qui porte la mutation *hsdR* code pour la méthylase de EcoK mais pas pour la DNase. On utilise cette mutation pour cloner des fragments d'ADN non méthylés tels que les fragments de PCR.

lon : protéase responsable de la dégradation des protéines aberrantes abolie (*E. coli B* n'a pas Lon)

lacI^q : Le gène *lacI* code pour le répresseur lac. La mutation *lacI^q* augmente la concentration en répresseur dans la bactérie, en l'absence de lactose l'opéron lactose est mieux régulé.

lacZ Δ M15 : *lacZ* est le gene d'*E.coli* qui code pour la β -galactosidase. *LacZ Δ M15* code pour une protéine inactive dans laquelle il manque les acides aminés 11-45. Le peptide manquant peut être produit par un vecteur, et s'associer à la protéine produite par *LacZ Δ M15* pour former une protéine active (α -complémentation).

mcrA : mutation utilisée pour cloner des fragments d'ADN génomique méthylés provenant par exemple de génome.

mutS, *mutD5*, *mutT* : mutation dans des gènes impliqués dans la réparation de l'ADN. Ces mutations sont utilisées pour faire de la mutagenèse au hasard.

recA⁻ : *recA* est le gène principal impliqué dans la recombinaison en aidant le transfert des brins entre les molécules d'ADN.

rne : gène codant pour la RNase E qui est impliquée dans la dégradation des ARNm. Un mutant de la RNase E peut permettre dans certains cas une meilleure expression (si la dégradation de l'ARN est le facteur limitant).

rpoH : aussi connue sous le nom de *htpR*, gène codant pour facteur de transcription qui induit la production de protéases. La mutation *ropHam* augmente la stabilité de certaines protéines.

supE : gène codant pour un ARNt suppresseur de la mutation ambre (TAG). *supE* insère une glutamine à la place du stop.

supF : gène codant pour un ARNt suppresseur de la mutation ambre (TAG). *supF* insère une tyrosine à la place du stop.

tonA : cette mutation procure la résistance au phage T1. Elle est utilisée pour réduire les risques de contamination.

TraD36 : supprime le transfert du facteur F par conjugaison.

TrxB : Délétion de la thioredoxine réductase. Cette mutation favorise la formation des ponts disulfures des protéines recombinantes produites dans *E. coli*.

ung⁻ : les souches *ung*⁻ manquent d'uracyl N-glycosylase, enzyme qui normalement enlève l'uracile incorporé dans l'ADN. Cette mutation est utilisée avec *dut*⁻ pour fabriquer de l'ADN ayant incorporé de l'uracile. Cet ADN est utilisé dans une méthode de mutagenèse dirigée ; le brin sauvage

comporte de l'uracile et est ensuite réparé lors de son incorporation dans une souche dut^+ , ung^+ . Cette réparation utilise le brin muté comme matrice, la mutation se retrouve donc sur les deux brins.

Les souches d'*E. coli* les plus utilisées

AD494 (DE3) :

Génotype : *trxB*

Utilité : Cette souche contient une copie du gène codant pour la T7 RNA polymérase en aval du promoteur *lacUV5*. Elle permet donc l'expression de gène en aval d'un promoteur T7. La délétion du gène codant pour la thioredoxine reductase (*trxB*) facilite la formation de ponts disulfures dans le cytoplasme.

AG1 :

Génotype : *recA1 endA1 gyrA96 thi-1 hsdR17 (rK- mK+) supE44 relA1*

BL21 :

Génotype : *E. coli B F⁻ dcm ompT hsdS_B (r_B⁻ m_B⁻) gal*

Utilité : expression de gènes

BL21 star :

Commercialisation : Invitrogen.

Cette souche dérivée de BL21 a en plus une mutation de *rne* qui abaisse la dégradation des ARNm.

BL21 (DE3) :

Cette souche dérivée de BL21 a intégré le phage λ DE3. Ce phage contient le gène codant pour l'ARN polymérase T7 en aval du promoteur *lacUV5*.

Utilité : production de protéines recombinantes dont le gène a été cloné en aval d'un promoteur reconnu par l'ARN polymérase T7. La production est induite par ajout d'IPTG.

BL21 (DE3) pLysS :

Cette souche dérivée de BL21 (DE3) contient en plus un plasmide codant pour le lysozyme.

Utilité : production de protéines recombinantes toxiques pour *E. coli*. La production est inhibée en l'absence d'induction.

C600 :

Génotype : *F' thr-1 leuB6 thi-1 glnV44 rfbD1 fhuA21*

Utilité : clonage dans le phage λ

DH5 α

Génotype : *supE44 DlacU169 f80 lacZDM15 hsdR17 recA1 'endA1 gyrA96 thi-1 relA1*

DH5 α F' :

Génotype : *F' / endA1 hsdR17(r_K⁻m_K⁺) glnV44 thi-1 recA1 gyrA (Nal^r) relA1 Δ (lacIZYA-argF) U169 deoR (ϕ 80dlac Δ (lacZ)M15)*

DH5 α T1^r :

Commercialisation : Invitrogen

DH5 α avec le génotype *tonA* qui confère la résistance au phage T1.b

GM2163

Commercialisation : NEB

Utilisée pour préparer des plasmides lorsque les méthylation Dam ou Dcm posent des problèmes de coupure par des enzymes de restriction.

JM101 :

Génotype : *supE thi-1 D(lac-proAB) [F' traD36 proAB lacIqZDM15]*

JM110 :

Génotype : *rpsL (Strr) thr leu thi-1 lacY galK galT ara tonA tsx dam dcm supE44 D(lac-proAB) [F' traD36 proAB lacIqZDM15]*

Utilité : comme cette souche est déficiente pour les méthylases Dam et Dcm, l'ADN fabriqué par cette souche peut être digéré par des enzymes de restriction sensibles à la méthylation

MC1061

Génotype : $\Delta(\text{araA-leu})7697 \text{ araD139 } \Delta(\text{codB-lac})3=\Delta\text{lac74 galK16 galE15 mcrA0 relA1 rpsL150 spoT1 mcrB9999 hsdR2 } \lambda^- \text{ F}^-$

Orgami B DE3 :

Commercialisation : Novagen

Génotype : $F^- \text{ ompT hsdS}_B \text{ (rB}^-\text{mb}^-) \text{ gal dcm lacY1 gor 522 :Tn10 (TcR) trxB: :kan (DE3)}$

Utilité : expression de protéine recombinante ayant des ponts disulfures.

SCS110 :

Génotype : $rpsL \text{ (Strr)} \text{ thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44D (lac-proAB) [F}^- \text{ traD36, proAB lacIqZDM15]}$

Utilité : comme cette souche est déficiente pour les méthylases Dam et Dcm, l'ADN fabriqué par cette souche peut être digéré par des enzymes de restriction sensibles à la méthylation

TG1 :

Génotype : $supE \text{ thi-1 } \Delta(\text{lac-proAB}) \Delta(\text{mcrB-hsdSM})5 \text{ (rK}^- \text{ mK}^-) \text{ [F}^- \text{ traD36 proAB lacIqZ}\Delta\text{M15]}$

Vendeur : Stratagene

XL1-Blue :

Commercialisation : Stratagène

Génotype : $recA1 \text{ endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lacI [F}^- \text{ proAB lacIqZD M15 Tn10 (Tetr)]}$

Utilité : préparation de plasmides en grande quantité due à la présence de *endA1*.

XL1-Red strain:

Commercialisation : Stratagène

Génotype : $endA1 \text{ gyrA96 thi-1 hsdR17 supE44 relA1 lac mutD5 mutS mutT Tn10 (Tetr)}$

Utilité : mutagenèse aléatoire

XL12-Blue :

Commercialisation : Stratagène

Génotype : *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lacI* [*F'* *proAB lacIqZD M15 Tn10* (Tet^r) Amy Cam^r]

Utilité : réalisation de transformation avec des efficacités supérieures à 5.10^9 colonies par μg d'ADN.

Les souches de cellules eucaryotesEtablissement d'une lignée cellulaire

Lorsqu'on explante un tissu d'un mammifère et qu'on le cultive en laboratoire, il pousse rarement de lui-même. Les tissus embryonnaires poussent pendant quelques générations pour finalement s'arrêter. Par contre des cellules cancéreuses peuvent se maintenir indéfiniment en culture. Les premières cellules issues de tissus, sont trop spécialisées. Les cellules embryonnaires sont moins spécialisées mais manquent de certains composants comme par exemple de télomères (ou télomère terminal transférase) qui ajoutent des séquences répétées à l'extrémité du chromosome. A chaque cycle cellulaire, il y a perte d'une partie de l'extrémité du chromosome et mort de la cellule lorsque cette perte atteint un gène essentiel.

Une expérience menée en 1911 par Peyton ROUS a montré qu'il était possible d'induire des tumeurs en injectant un extrait acellulaire de tumeurs à des poulets, on a réalisé plus tard qu'il avait découvert un virus tumorigène. Pour établir des lignées immortelles, soit on prélève des tumeurs soit on traite des cellules avec des agents carcinogènes tels que les produits chimiques, des radiations ou des virus tumorigènes. Ces cellules sont appelées transformées (sens différent du terme employé pour les bactéries). Ce travail a été fait par de nombreux laboratoires et on dispose à l'heure actuelle d'une panoplie importante de lignées cellulaires utilisables pour produire des protéines.

Les banques

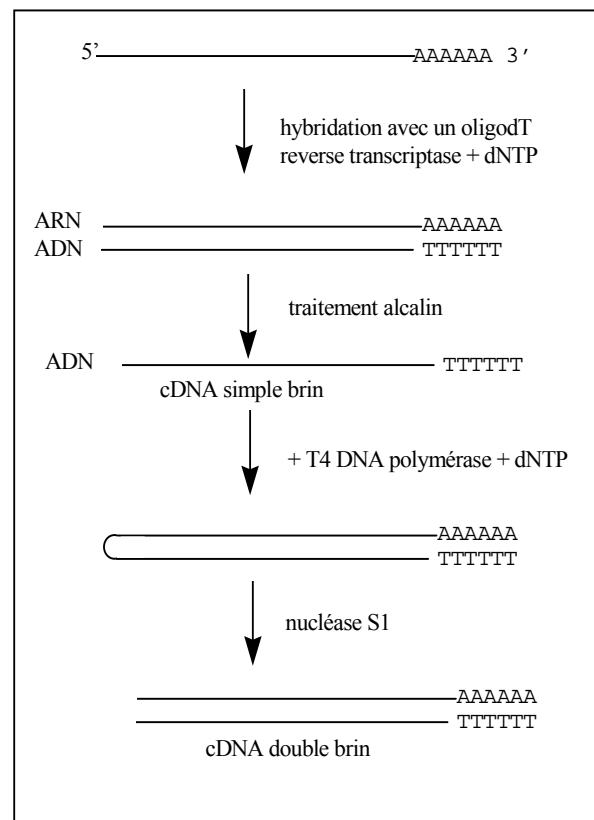
On distingue des banques d'ADNc et des banques d'ADN génomique. La banque d'ADNc représente la population des ARNs d'une cellule, d'un tissu ou d'un organisme à un moment donné. La seconde, la banque génomique, doit représenter l'ensemble de l'ADN génomique d'un individu. Une banque est donc la conversion d'une population d'ARN ou d'un génome en clones.

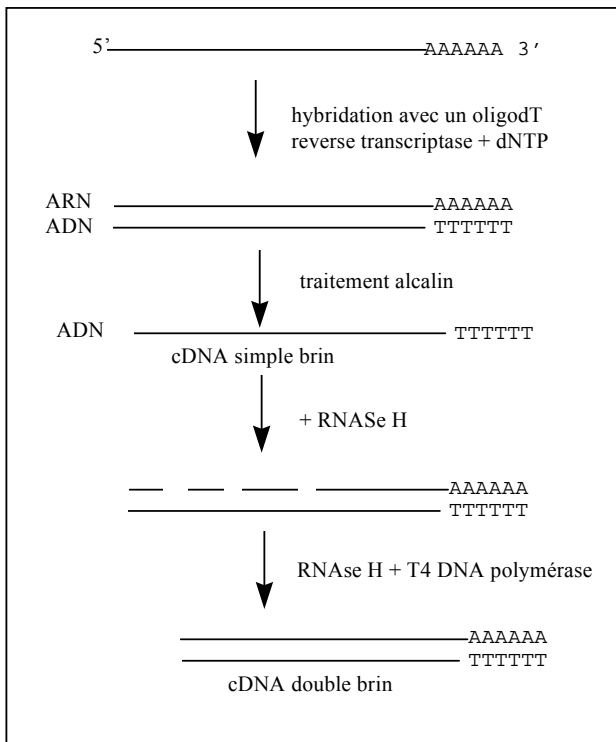
a) Banque d'ADNc

Il faut tout d'abord transformer les ARNs en ADNc qui sont alors appelés ADN complémentaires ou ADNc. On distingue deux grandes techniques pour fabriquer un ADNc.

Technique à la nucléase S1

Pour la synthèse du premier brin, on utilise une reverse transcriptase (ADN polymérase ARN dépendante) et, comme pour toutes les ADN polymérases, une amorce, qui est ici un oligonucléotide (oligodT). On détruit ensuite l'ARN par un traitement alcalin. Lorsque l'extrémité 5' de l'ADN se retourne et s'hybride sur lui-même en faisant une boucle en épingle à cheveux, on a une extrémité double brin qui peut servir d'amorce à une DNA polymérase ADN dépendante. On fait donc agir cette enzyme. On obtient donc un ADNc double brin avec une extrémité en épingle à cheveux. Une telle structure ne peut pas être intégrée dans un vecteur. Pour avoir une extrémité franche on fait agir une DNase dégradant le simple brin (nucléase S1).





Technique de Gubler et Hoffman

Ici on synthétise un premier brin d'ADNc comme précédemment avec une reverse transcriptase après avoir hybridé une amorce, mais l'ARN est éliminé par digestion ménagée à la RNase H. Les fragments restants servent alors d'amorce pour une DNA polymérase qui comble les trous au fur et à mesure que la RNase H continue son action. Les fragments d'ADN sont alors reliés grâce à l'action d'une ligase. Noter que les bouts sont ici francs, grâce à l'action 3'5' exonucléase de la polymérase.

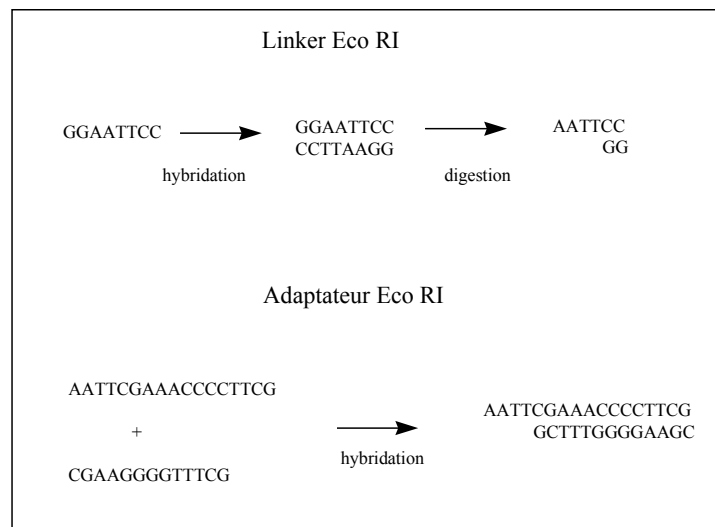
Avec les deux techniques, on n'a généralement pas de copie de l'extrémité 5' de l'ARN. D'autres techniques permettront de l'obtenir. Ce n'est souvent pas grave si on s'intéresse à la partie codante, la partie 5' situé en amont de l'ATG est dans ce cas moins importante puisque non codante.

Pour les deux techniques, on a comme alternative l'utilisation d'amorces au hasard à la place de l'oligo dT. Cette méthode se rapproche de la méthode de marquage par "random priming". Les oligonucléotides se fixent au hasard sur l'ARN et l'ADNc résultant correspond à uniquement une partie du gène, on perd donc l'extrémité 3'. Cette méthode est principalement utilisée lorsqu'on attend un cDNA long, en effet la reverse transcriptase n'est souvent pas assez active pour synthétiser plusieurs kilobases. Les ADNc amorcés par l'oligo dT, n'étant pas assez longs, la partie 5' du gène manque. Ce manque peut être comblé si l'amorce est moins éloignée de l'extrémité 5'.

Il faut maintenant intégrer l'ADNc double brin à un vecteur.

On pourrait directement liguer la population d'ADNc à un vecteur coupé par une extrémité franche et déphosphorylée mais cette ligation est relativement inefficace et on préfère faire une ligation avec une extrémité cohésive. Il faudra donc ajouter des extrémités cohésives aux deux bouts du cDNA.

Une première technique consiste à inactiver par méthylation un site de restriction (Eco RI, par exemple) en incubant la population d'ADNc avec la méthylase de Eco RI. On ligue ensuite des



linkers Eco RI. Ce sont des petits oligonucléotides palindromiques qui deviendront double brin après hybridation. Ces linkers comportent un site EcoRI si bien qu'une digestion par cette enzyme permet de générer des extrémités cohésives sans couper les sites internes qui seront protégés par la méthylation. La méthylation est souvent partielle, ce qui nous amène à cloner des fragments d'ADNc lorsque la protection n'a pas été totale. Pour éviter cette étape, on peut utiliser des adaptateurs. Il s'agit d'un couple oligonucléotides partiellement complémentaires générant une extrémité franche qui sera compatible avec l'extrémité de l'ADNc et une extrémité cohésive qui sera compatible avec les extrémités du vecteur.

Les ADNc seront ensuite ligués avec un vecteur coupé par EcoRI et déphosphorylé pour éviter sa recircularisation. On sera à cette étape en présence d'un ensemble de plasmides ou phages recombinants, chacun portant un insert différent, représentant l'ensemble de la population d'ARN de départ.

La banque d'ADNc obtenue représentera l'ensemble des ARN utilisés au départ. La banque sera donc représentative uniquement du matériel de départ. Comme la population d'ARN dépend du patron d'expression des gènes de la cellule de départ, chaque banque sera représentative du type de cellule utilisé.

b) Banque génomique.

Une banque génomique représente l'ensemble du génome. On prépare donc de l'ADNg et on le coupe par une enzyme de restriction de façon à obtenir des fragments de 20 kb se recouvrant. Pour ce faire, on utilise une enzyme qui coupe fréquemment dans le génome telle que Sau3A qui reconnaît la séquence /GATC. Cette enzyme reconnaît 4 bases et coupe donc en moyenne toutes les 4⁴ bases soit environ toutes les 200 bases. Si la digestion est complète, les fragments sont trop petits pour être utilisables dans une banque, on préfère avoir des fragments plus grands et surtout se recouvrant. On effectue donc une digestion partielle en utilisant peu d'enzyme de façon à obtenir des fragments de 20 kb. Même si la digestion est bien calibrée, une digestion partielle ne donne pas que des fragments de 20 kb mais aussi des fragments plus petits et plus grands. Le produit de la digestion est alors soumis à une centrifugation zonale, dans une solution de saccharose ou de NaCl permettant de séparer les molécules en fonction de leurs tailles. Les molécules les plus lourdes, les plus grandes, sédimentent plus vite que les molécules les plus petites. Si on récolte le gradient on peut déterminer par électrophorèse la fraction comportant de l'ADN de 20 kb. Cet ADN est borné par des extrémités 5' sortantes GATC générées par la digestion par Sau3A. Cette population de fragments d'ADN sera ensuite liguée aux bras droit et gauche d'un lambda obtenu par digestion par BamH1 (G/GATCC) générant des extrémités GATC, compatibles avec Sau3A. Après transformation par empaquetage *in vitro*, on obtient une population de phages dont les inserts représentent l'ensemble de l'ADN génomique.

Dans certains cas la digestion de l'ADN est difficile du fait de la méthylation de l'ADN ou de la présence d'inhibiteur de la digestion qu'il n'a pas été possible d'éliminer lors de la purification de l'ADN. Dans ce cas on coupe l'ADN au hasard par sonication. Les extrémités seront rendues franches par action de la mung bean nucléase. Cette enzyme est préférée aux polymérases ou à la nucléase S1 car elle ne reconnaît pas les nicks qui dans ce cas sont nombreux. Les fragments d'ADN sont alors ligués dans un vecteur soit directement soit après avoir ajouté des extrémités cohésives.

Selon la longueur des inserts que l'on veut obtenir, on utilise différents vecteurs

< 5kb : plasmide, 16-22 kb : phage λ , 42-50 kb : cosmide, 100-300 kb : BAC, PAC, > 300 kb : YAC

Purification des acides nucléiques

1) Purification

Dégradation des protéines

Dans le cas de la purification de l'ADN, la protéinase K est le plus souvent utilisée, et ce, en présence d'un détergent dénaturant. La digestion est effectuée en présence d'EDTA pour éviter l'action des DNases.

Pour l'ARN, des agents fortement dénaturants tels que les sels de guanidium sont utilisés. Ils ont comme rôle de dénaturer les protéines et en particulier les RNases.

Précipitation et élimination des autres composants de la cellule

Le phénol précipite les protéines

Le CTAB (cetyltriméthylammonium bromide) est un détergent cationique qui précipite les protéines et les polysaccharides. Ces polysaccharides sont présents en quantité plus ou moins grande chez les plantes.

Chromatographies

L'ADN peut s'accrocher sur des résines échangeuses d'ions, sur de la silice ou de la magnétite (Davies et al., 1998)

- sur colonne d'échange d'ions, l'ADN s'accroche à faible force ionique et est élué à forte force ionique.
- sur silice (SiO_2) comme sur magnétite (Fe_3O_4), l'ADN est fixé à forte force ionique (2M NaI), la colonne est lavée par de l'alcool 50% et l'ADN peut être élué par de l'eau. L'ADN est donc directement utilisable pour les réactions suivantes.

L'ARN messager peut s'accrocher par hybridation sur colonne d'affinité sur laquelle on a fixé un oligodT comme ligand.

Concentration des acides nucléiques

La méthode la plus utilisée est la précipitation en présence d'éthanol. Le précipité se fait en présence d'ions monovalents en rajoutant 2,5 volumes d'éthanol à 95%. On utilise le NaCl (0.1M), le LiCl (0.4M), l'acétate d'ammonium (0.25M, pH5.2) ou l'acétate de sodium (0.25M). On utilise généralement indifféremment l'un ou l'autre de ces sels.

Toutefois, dans certains cas, on doit effectuer un choix lorsqu'on veut faire des précipitations sélectives. Le NaCl est utilisé lorsqu'il y a du SDS dans le tampon, le SDS restant alors dans la solution. Le LiCl est utilisé lorsqu'on ne veut précipiter que les grandes molécules d'ARN sans précipiter les petites. L'acétate d'ammonium 2,5M est utilisé si on ne veut pas précipiter les nucléotides.

Les sels d'acide nucléiques formés sont insolubles dans l'alcool.

Certains sels seront évités s'ils sont inhibiteurs des réactions suivantes. Les ions Cl⁻ sont inhibiteurs de la reverse transcriptase et de l'initiation de la traduction, le LiCl et le NaCl seront donc évités si après la précipitation, on veut faire un ADNc ou faire une traduction *in vitro*. Les ions ammonium sont des inhibiteurs de la T4PNkinase.

A la suite de la précipitation, le culot d'ADN devra être le plus souvent lavé avec de l'éthanol 70%. Ce lavage permet de réhydrater en partie les acides nucléiques et facilite leur reprise en solution aqueuse.

Récupération d'un fragment d'ADN d'un gel d'électrophorèse

« Freeze-squeeze » méthode: la bande est découpée, refroidie à -80°C et centrifugée. On récupère environ 50 % de l'ADN pour des fragments d'ADN d'environ 500 pb.

Papier DEAE : une bande de papier DEAE est placée en aval de la bande à récupérer sur le gel. La migration continue mais lorsque l'ADN arrive sur le papier, il reste bloqué sur le papier. Le papier est

alors récupéré, lavé avec un tampon à faible force ionique puis élué avec un tampon contenant 1M NaCl.

Electroélution : dans les protocoles les plus simples, la bande est découpée puis placée en présence de tampon dans un boudin de dialyse. L'ensemble est placé dans un appareil d'électrophorèse, la bande migre hors du morceau de gel mais reste dans le tampon du boudin de dialyse. Des appareils ont été fabriqués pour récupérer les fragments d'ADN par électroélution. La bande est mise dans le réservoir de la cathode, l'ADN migre, sort de la bande, est pigée dans un réservoir intermédiaire contenant un tampon à haute force ionique.

2) Transformation-transfection

Comment faire rentrer un acide nucléique dans une cellule ?

Il existe différentes techniques qui permettent de faire pénétrer de l'ADN dans une cellule eucaryote. Dans tous les cas, le but est de traverser la membrane cytoplasmique. Le choix de l'une ou l'autre de ces techniques est essentiellement fonction de la sensibilité des cellules, de l'efficacité de transfection recherchée et de l'objectif exact de l'expérience. Une cellule dans laquelle on a fait entrer un ADN est une cellule transfectée.

- La microinjection. Des microaiguilles sont fabriquées par étirement d'un tube de verre à la chaleur. Cette technique est principalement utilisée pour injecter de l'ADN dans des ovocytes ou des œufs.
- Le dextran/DEAE. On a couplé un groupement chimique chargé positivement (diéthylaminoéthyl) au dextran qui est un hydrate de carbone polymérique. Le mécanisme par lequel le dextran/DEAE permet à l'ADN d'atteindre les noyaux des cellules est mal connu. Conventionnellement, on admet que les complexes d'ADN-dextran/DEAE collent à la surface des cellules et sont internalisés par endocytose.
- Le phosphate de calcium. La plus ancienne des techniques de transfection. Cette méthode donne de meilleurs rendements que le dextran/DEAE. Elle consiste à faire rentrer l'ADN via une co-précipitation PO_4Ca /ADN qui adhère à la surface des cellules. Cette technique est basée sur le fait que les cations divalents tels que le calcium ou le magnésium favorisent la pénétration de l'ADN dans les cellules. Elle nécessite d'autre part d'utiliser des cellules adhérentes qui présentent une forte activité d'endocytose. Les principaux problèmes liés à cette technique sont de trois ordres : i) le calcium est toxique pour certaines cellules ii) beaucoup de copies sont transfectées par cellule – plusieurs dizaines iii) l'expression du gène rapporteur est retardée d'environ 72h .
- L'électroporation (ou électroperméabilisation) : Cette technique consiste à placer les cellules dans un champ électrique qui perturbe la membrane des cellules, on pense que l'ADN rentre par des pores par diffusion et par électrophorèse. En effet, l'ADN étant chargé négativement, il migre vers l'anode dans le champ électrique. Cette méthode s'applique aux

cellules adhérentes ou en suspension ; il n'y a pas de retard d'expression du gène rapporteur ; peu de copies sont transfectées.

- Les liposomes : on fait des liposomes en utilisant des lipides chargés positivement. Ces liposomes forment des agrégats avec l'ADN (chargé négativement) qui sont souvent appelés « lipoplexes ». Les liposomes ayant une charge globale positive sont attirés par la membrane cellulaire de charge globalement négative. On a tout d'abord pensé que le liposome rentre dans la cellule par fusion des membranes du liposome et de la cellule, mais il a été montré que le liposome rentre dans la cellule par endocytose. Après destruction des endosomes, l'ADN se retrouve dans le cytoplasme et lors d'une division cellulaire peut se retrouver dans le noyau (Zabner et al., 1995). Le premier liposomes qui ont été mis au point étaient constitués d'un mélange de DOTMA (N-[1-(2, 3-dioleoyloxy)propyl]-N,N,N-triméthylammonium chloride) et de DOPE (1,2-dioleoyl-3-phosphatidylethanolamine). Il y a actuellement de nombreux lipides capables de former des liposomes cationiques sur le marché et des polymères peuvent être substitués aux lipides.

- Les peptides : Un peptide amphiphile dérivé de la troisième hélice de la protéine antennapedia de la drosophile est capable d'être internalisé dans les cellules sans passer par la voie de l'endocytose. Si on synthétise ce peptide avec une queue polylysine, le peptide accroche l'ADN par liaison électrostatique et rentre dans la cellule avec l'ADN (Citti et al., 2002).

Comment faire rentrer un acide nucléique dans le noyau ?

Dans une cellule en division, la membrane nucléaire disparaît à chaque mitose et est reconstituée dans chaque cellule fille. Aussi, une fois dans le cytoplasme de la cellule, l'ADN se retrouvera dans le noyau lors de la reconstitution de la membrane nucléaire. Si l'entrée dans le noyau est un problème comme par exemple d'entrée de SiRNA dans des cellules qui ne se divisent pas, on peut utiliser comme vecteur des protéines qui condensent l'ADN et qui sont pourvues d'une séquence d'adressage au noyau. L'histone H2 a ainsi été utilisée comme vecteur de SiRNA.

Comment faire rentrer une protéine dans une cellule ?

- Microinjection : on peut injecter des protéines dans les cellules en utilisant des aiguilles effilées, cependant cette technique demande de l'habileté et on doit injecter les cellules une à une.
- Electroporation : l'électroporation utilise un fort voltage qui produit des pores transitoires. Cette méthode peut être utilisée sur des populations cellulaires mais n'est pas spécifique, les autres composants du milieu de culture rentrent dans la cellule. Un autre inconvénient de l'électroporation est son efficacité très variable en fonction du type cellulaire.
- Liposomes : Les liposomes sont très efficaces pour transporter l'ADN, mais pour la plupart, restent peu efficaces pour les protéines. Cependant Gene Therapy System Inc a développé une formulation de lipides qui interagit avec les protéines et permet leur internalisation. De tels lipides sont commercialisés par Perbio.
- Les peptides : Il a été observé que quelques protéines peuvent entrer dans les cellules. L'analyse du mécanisme a mis en évidence que l'internalisation était due à des petits peptides qui peuvent traverser la membrane plasmique et entraîner le reste de la protéine. Les trois peptides les plus utilisés sont dérivés du facteur de transcription Antennapedia de la drosophile, de la protéine VP22 du virus de l'Herpes et de l'activateur transcriptionnel Tat du virus HIV. Ce sont de petits peptides de 10-16 résidus qui comportent de nombreux acides aminés basiques. Par exemple le peptide de Tat présente la séquence suivante : RKKRRQRRR. A partir de ces séquences (Ho *et al.*, 2001) de nombreux peptides synthétiques ont été construits. L'entrée des protéines est indépendante de l'endocytose comme de la taille de la protéine à la condition que la protéine soit liée de façon covalente au peptide ; ces peptides sont donc utilisés en protéines de fusion.
- Les protéines transportrices : Pep-1 est une protéine de 21 résidus qui est composé de 2 domaines, un domaine hydrophobe, riche en tryptophane il favorise l'interaction avec la protéine à transporter et un domaine riche en lysine dérivé du virus SV40 qui permet l'entrée dans la cellule et permet la solubilité du peptide.

Le marquage des acides nucléiques

Certains types d'expériences (Southern, northern, hybridation in situ, ...) nécessitent de visualiser une hybridation. Pour cela, il est nécessaire de marquer un acide nucléique (ADN ou ARN simple brin) qui prendra le nom de sonde.

1) Les différents types de sonde et leur révélation

Sondes chaudes

Comment marquer une sonde ? On dispose de plusieurs isotopes radioactifs qu'on pourra incorporer dans la sonde. On peut utiliser :

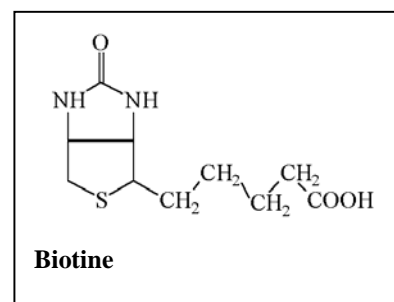
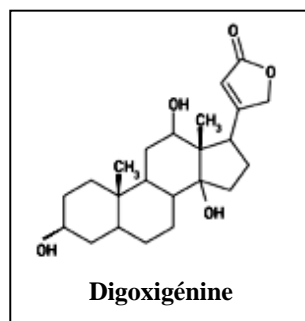
Isotope	demi-vie	énergie
^{32}P β -	14,3 jours	1.7 MeV
^{35}S , β -	87.2 jours	0.17 MeV
^{14}C , β -	5700 ans	0.16 MeV
^3H , β -	12.3 ans	0.02 MeV

Le soufre n'est normalement pas présent dans les acides nucléique mais comme il est proche de l'oxygène on peut faire des analogues soufrés des nucléotides dans lesquels un oxygène est remplacé par un soufre. Le choix de l'isotope dépendra de l'utilisation de la sonde. On emploiera un rayonnement fort pour une détection sur un filtre, et un rayonnement moins énergétique pour une détection en microscopie.

La détection des radioisotopes peut se faire de plusieurs manières : i) en utilisant un compteur à scintillation, ii) par radioautographie (film sensible ou émulsion) iii) par un scanner spécial, le phosphoimager.

Sondes froides

On peut utiliser des marquages non radioactifs. Dans ce dernier cas, on parle de sonde froide. Les deux types de marquage froid les plus souvent utilisés sont les marquages à la biotine et à la digoxigénine. Ces deux composés sont fixés sur un nucléotide, lui-même incorporé à la sonde.

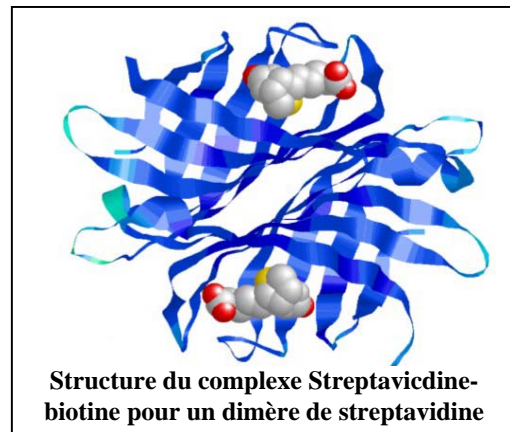


La biotine est la vitamine H. C'est un cofacteur de réactions enzymatiques impliquées dans l'incorporation ou le transfert de CO_2 (si on mange trop de blanc d'œuf, on inhibe l'incorporation de CO_2 dans le foie).

La détection des sondes froides

La biotine est détectée avec de la streptavidine. L'avidine est une protéine du blanc d'œuf, elle comporte 4 sites de liaison pour la biotine. Elle est généralement produite par génie génétique dans *Streptomyces avidinii*, sous une forme non glycosylée et prend alors le nom de streptavidine.

La digoxigénine se révèle à l'aide d'un anticorps.



La streptavidine ou les anticorps n'ont pas d'activité enzymatique propre, on utilise des protéines de fusion pour les détecter. L'anticorps ou la streptavidine sont couplés avec une protéine ayant une activité enzymatique facilement détectable comme la phosphatase alcaline ou la peroxydase.

La peroxydase du raifort (radis noir) (HRP, horseradish peroxidase) est une protéine de 40 kDa. Elle catalyse la réaction : $H_2O_2 + \text{donneur d'électron} \rightarrow \text{produit coloré} + H_2O$
on utilise un chromogène comme donneur d'électrons

- soit du tétrachlorure de diamino-3-3'-benzidine (DAB) qui donne un précipité brun
- soit de l' amino-3-ethyl-9-carbazole (AEC) qui donne un précipité rouge.

La phosphatase alcaline

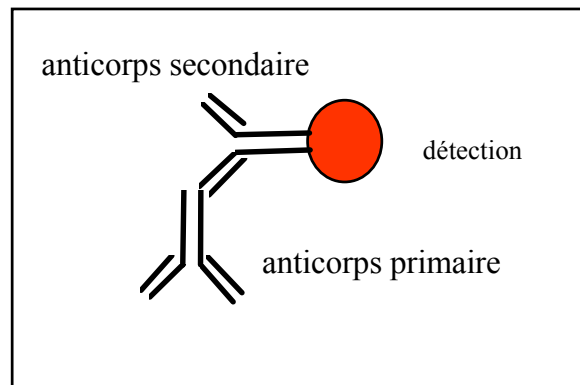
C'est une protéine de 40 kDa isolé d'intestin de veau.

Plusieurs chromogènes peuvent être utilisés: new fushine, fast red, BCIP/NBT...

L'amplification du signal : pour augmenter la sensibilité des sondes froides on cherche à amplifier le signal détecté. Il existe de nombreuses variantes pour atteindre ce but.

- Méthode en deux couches

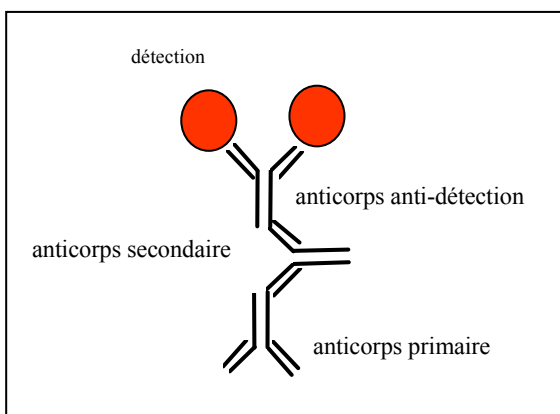
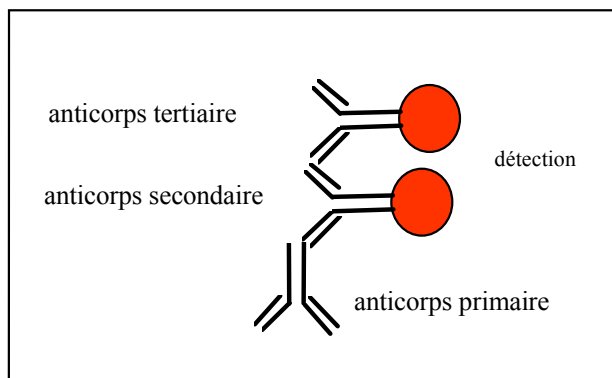
Elle permet une simple détection de l'anticorps primaire sans amplification du signal., sauf si l'anticorps secondaire est polyclonal et reconnaît donc plusieurs épitopes. L'anticorps secondaire est conjugué avec l'enzyme ou le marqueur de fluorescence.



L'anticorps secondaire peut être remplacé par de la protéine A.

- Méthodes en trois couches

On utilise un anticorps primaire qui reconnaît l'antigène, un anticorps secondaire qui reconnaît l'anticorps primaire et un anticorps tertiaire qui reconnaît l'anticorps secondaire. Les anticorps secondaires et tertiaires sont conjugués au même système de détection.



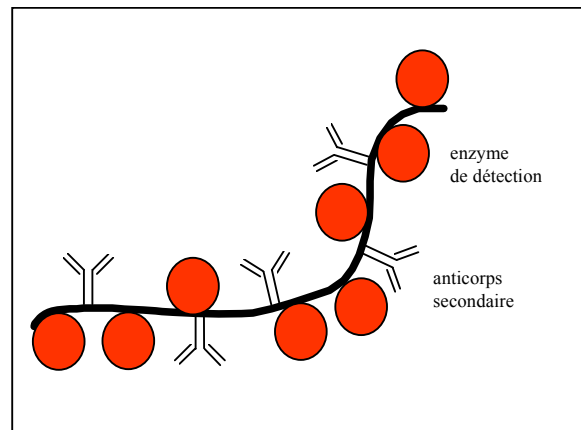
- Méthode utilisant un complexe entre l'enzyme de détection et un anticorps.

Cette méthode est parfois appelée méthode des anticorps non marqués, ou suivant l'enzyme de détection PAP (peroxydase antiperoxydase) ou APAAP (alkaline phosphatase anti alkaline phosphatase)

L'anticorps utilisé contre l'enzyme de détection doit être de même nature (obtenu dans la même espèce animale) que l'anticorps primaire.

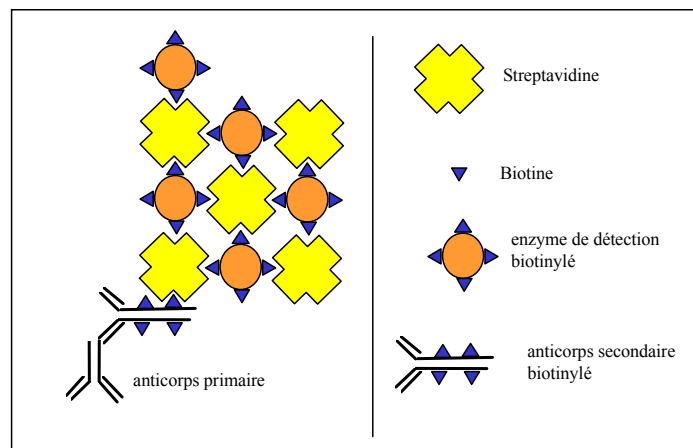
- Méthode EPOS (Enhanced Polymer One Step Staining) et Envision (Polymer two step staining) (DAKO)

Les anticorps primaires (EPOS) ou secondaires (Envision) sont fixés sur un polymère de dextran avec la protéine servant de révélation (peroxydase, phosphatase alcaline..)



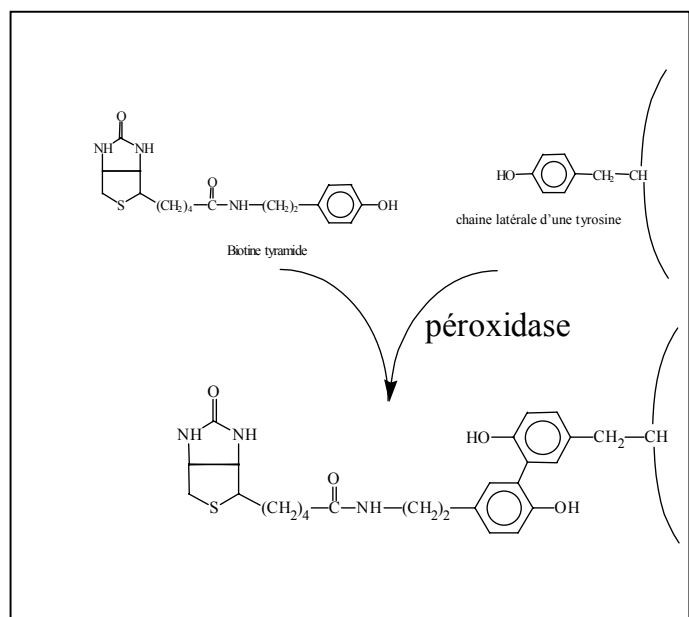
- Méthodes utilisant les complexes avidine-biotine. (AB Complex)

L'anticorps secondaire est biotinylé comme la protéine de révélation. En présence d'avidine qui a quatre sites de liaisons à la biotine, on obtient un empilement de l'enzyme de détection.



- CSA (Catalysed Signal Amplification)

Utilisation de la biotinylyl-tyramide. Les tyramides sont convertis par la peroxydase en un intermédiaire qui est réactif avec les résidus aromatiques riches en électrons tel que la tyrosine, présente dans les protéines. La streptavidine-péroxydase reconnaît la biotine sur la sonde. La peroxydase active la biotine-tyramide qui se fixe sur la protéine (ou sur une protéine proche) et, ainsi, on a une nouvelle biotine fixée à proximité de la



sonde. Cette biotine peut alors être reconnue par la streptavidine-peroxidase.

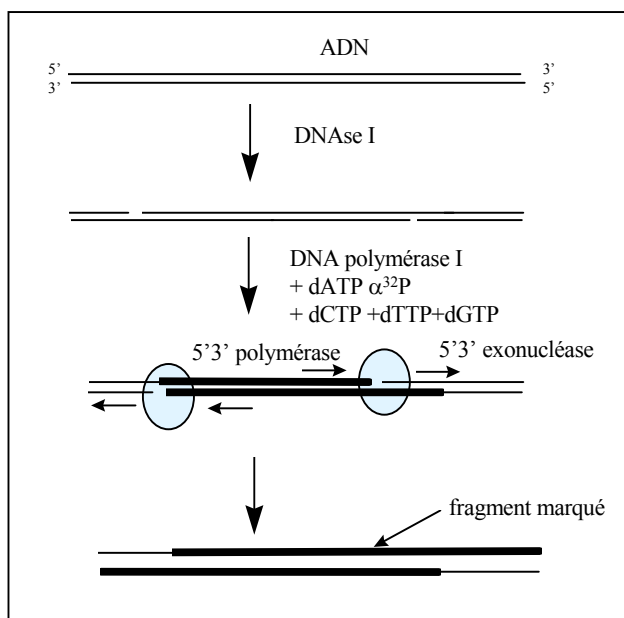
Cette technique peut être utilisée pour les sondes biotinylées mais aussi pour toutes les sondes qui sont reconnues par un anticorps. Dans ce dernier cas on utilisera un anticorps secondaire marqué à la biotine.

II) Les différentes techniques de marquage des acides nucléiques

Les méthodes enzymatiques sont les plus employées, on incorpore le nucléotide modifié à l'aide d'une enzyme.

Marquage interne

Marquage de l'ADN. Deux méthodes peuvent être utilisées, la « nick translation » et le « random priming ».



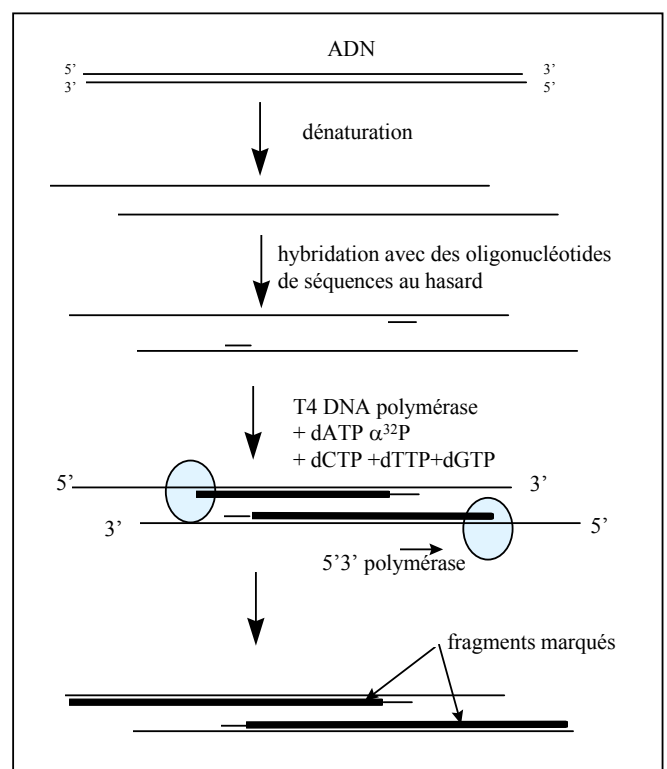
La nick translation ou déplacement de coupure vise à marquer un ADN double brin. On digère le fragment d'ADN par la DNase I de *E. coli* qui coupe après les pyrimidines, mais d'une façon ménagée de façon à ne faire que quelques coupures au hasard sur le fragment. Ces coupures sur un seul des brins sont appelées des nicks, elles libèrent des extrémités 5' OH à l'intérieur du fragment. On utilise ensuite l'ADN polymérase I de *E. coli* qui présente plusieurs activités dont une activité 5'-3' exonucléase et une activité 5'-3' polymérase. (Il y a en plus une activité 3'-5' exonucléase moins

rapide que l'activité polymérase en présence de nucléotides). Les deux brins sont ainsi marqués sur

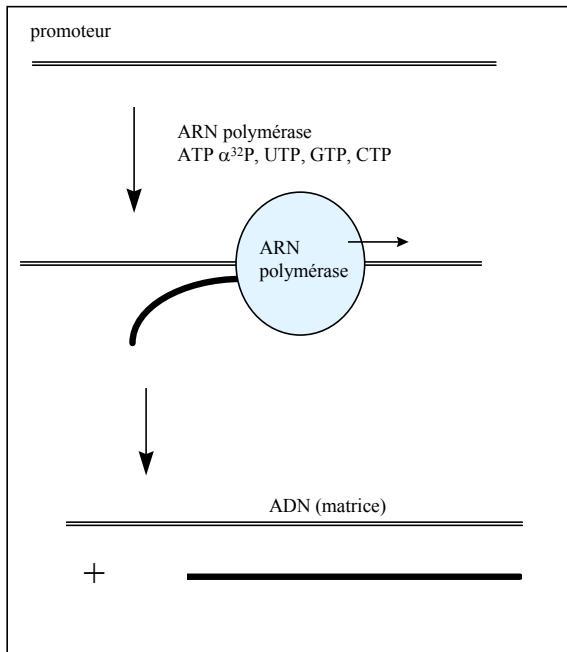
toute leur longueur à l'exception de leur extrémité 5'. La difficulté de cette méthode est de réaliser la digestion ménagée pour initier la polymérisation. Si la digestion est trop forte, il y a trop de nick, certains se trouvent à proximité l'un de l'autre sur les deux brins et on obtient uniquement des petits morceaux d'ADN, si elle est trop faible, un grand nombre de molécules ne sont pas marquées. De plus l'activité 5'-3' exonucléase de l'ADN polymérase I attaque le fragment d'ADN par l'extrémité et on obtient un raccourcissement du fragment. Cette méthode « historique » n'est plus utilisée et a été remplacée par le « Random priming »

Random priming ou amorçage au hasard.

On commence par dénaturer le fragment d'ADN. Puis on hybride des oligonucléotides au hasard. Pour ce-faire on utilise des petits oligonucléotides (6 à 8 mer) qui sont obtenus par coupure d'un ADN au hasard suivi d'une purification des fragments de 6 à 8 paires de base. Une fois ces oligonucléotides hybridés, on va synthétiser la séquence en aval, en 3', à l'aide d'une polymérase n'ayant pas d'activité 5'3' exonucléase.



Dans les deux cas, les deux brins sont marqués et la sonde est double brin. Il faudra donc la dénaturer avant de s'en servir dans une hybridation.



Pour obtenir une sonde spécifique d'un seul brin on utilisera plutôt une sonde ARN synthétisée *in vitro*. Pour faire une transcription *in vitro*, on aura besoin i) d'un ADN linéaire double brin servant de matrice positionné en aval d'un promoteur ii) d'une ARN polymérase reconnaissant ce promoteur iii) de nucléotides dont un est marqué en α .

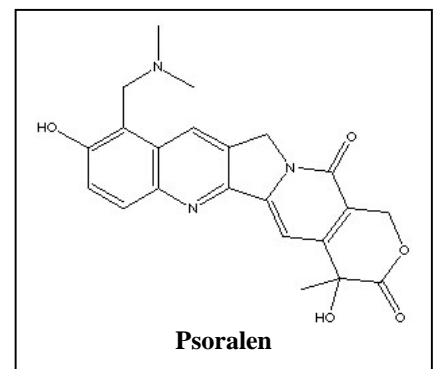
Le promoteur peut être présent dans un vecteur ou placé en amont du fragment à marquer par PCR.

Le nucléotide marqué est généralement en quantité limitante, à une concentration de l'ordre de 3 à 5 μM ce qui a pour effet de ralentir la polymérisation. Ce ralentissement n'est pas important sauf pour les 8-12 premières bases. Pour obtenir un bon rendement on placera en début de transcription une séquence dépourvue de A si on utilise de l'UTP marqué.

Cette méthode de marquage donne de meilleures sondes que la méthode d'amorçage au hasard. En effet on n'a pas de brin non marqué. En effet l'ADN matrice peut être éliminé à l'aide d'une DNase à la fin de l'expérience. Il n'y a donc qu'un seul brin si bien que la sonde n'a pas tendance à s'hybrider sur elle-même. L'inconvénient est qu'il s'agit d'ARN et l'ARN est plus difficile à manipuler que l'ADN (il est facile d'éliminer les DNases par un traitement thermique, par contre la plupart des RNases sont thermostables et il est difficile de les éliminer).

On peut utiliser une méthode non enzymatique pour marquer une sonde ARN. On utilise du psoralen qui est une molécule plane. Le psoralen s'intercale entre les bases et une irradiation UV à 356 nm le lie de façon covalente à l'acide nucléique

L'avantage des sondes froides sur les sondes chaudes est



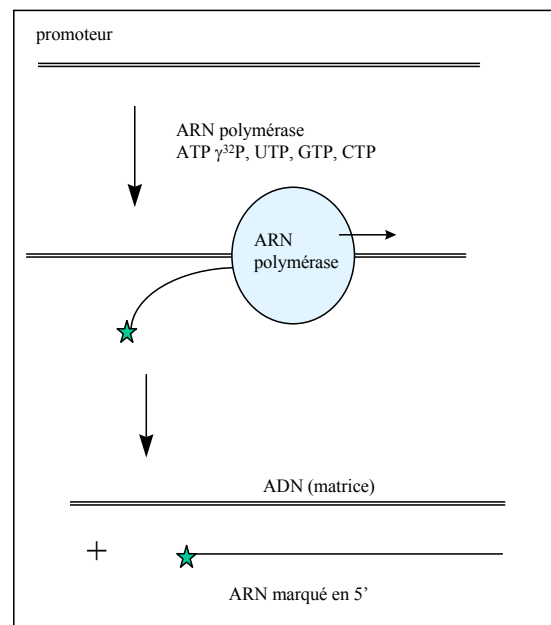
qu'elles sont stables et donc réutilisables. Toutefois elles sont un peu moins puissantes que les sondes chaudes et si on veut détecter par exemple un transcrit rare en northern, la sensibilité de révélation n'est pas suffisante.

Marquage aux extrémités

Dans certains cas on a besoin que la sonde soit marquée uniquement à une extrémité. On peut marquer soit l'extrémité 5', soit l'extrémité 3'. C'est pour ainsi dire la seule méthode de marquage utilisable pour les oligonucléotides, ils sont en effet trop petit pour envisager un marquage interne.

Marquage en 5' des ARN par polymérisation

La première base des ARN est généralement une adénine. Si on effectue une transcription *in vitro* en présence de γ ATP ^{32}P (ou β ATP ^{32}P), le seul A marqué sera celui qui est en 5'. Les autres adénines incorporées dans l'ARN néosynthétisé ne seront pas marquées, en effet seul le phosphate en α restera dans le polymère.



Marquage en 5' des acides nucléiques (ADN ou ARN)

A partir d'une extrémité 5'OH on peut ajouter un phosphate marqué à l'aide de la T4 Polynucléotide kinase et de γ ^{32}P ATP. Il faudra au préalable obtenir l'acide nucléique avec un 5'OH.

Pour les oligonucléotides synthétisés chimiquement, la synthèse s'effectue de 3' vers 5' et l'extrémité 5' est OH. On n'a donc aucun traitement préliminaire à faire.

Pour les ADN obtenus par digestion par une endonucléase de restriction, on a un 5' phosphate. Ce phosphate peut être éliminé à l'aide d'une phosphatase. On obtient une extrémité 5'OH,

on détruit ensuite la phosphatase (chauffage, extraction phénolique par exemple), avant de faire agir la T4 PNkinase sur le produit de la réaction.

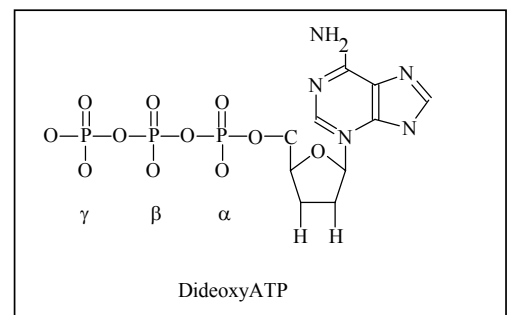
Pour les ARN messagers eucaryotes qui ont un cap en 5', il faut détruire ce cap, par action d'une pyrophosphatase puis laisser agir la phosphatase sur le phosphate restant (voir plus loin pour la structure du cap).

Cette méthode est généralement employée pour les oligonucléotides pour lesquels les marquages internes ne sont pas possibles du fait de leur petite taille.

Marquage en 3':

Dans ce cas il va falloir ajouter un ou plusieurs nucléotides à l'extrémité 3' de la sonde.

- Pour les ARN on utilisera une ARN ligase en présence d'un nucléotide monophosphate et d'ATP pour fournir de l'énergie. Le nucléotide sera ajouté en 3' OH. Pour qu'il n'y ait pas qu'un seul nucléotide ajouté, on utilisera un nucléotide ayant son extrémité 3' bloquée par exemple par un phosphate.
- Pour les ADN, on peut utiliser une deoxynucléotidyl terminal transférase avec un α ^{32}P nucléotides. Il y a alors ajout de plusieurs nucléotides en 3'. Si on veut n'en rajouter qu'un, il faut utiliser un didéoxynucléotide marqué à la place d'un déoxynucléotide, la polymérisation sera ainsi bloquée après l'ajout du premier nucléotide.
- Pour les ADN, on peut aussi utiliser les DNA polymérasés. Si l'ADN présente une extrémité 5' sortante, il y aura polymérisation de la séquence complémentaire de l'extrémité sortante. Si l'extrémité est franche il faudra distinguer deux cas : soit la DNA polymérase possède une activité 3'5' exonucléase soit elle en est dépourvue. Dans le premier cas, l'activité 3'5' exonucléase transforme l'extrémité franche en une extrémité 5' sortante qui est ensuite remplie par l'activité de polymérisation. Il y a alors marquage comme précédemment. Dans le cas où la DNA polymérase est



dépourvue d'activité 3'5' exonucléase, le marquage se fait grâce à l'activité terminale transférase. La plupart des polymérases présente en effet une activité terminale transférase lorsqu'elles ont comme matrice un ADN double brin. Comme l'activité est spécifique du double brin, un seul nucléotide est rajouté en 3' même si on utilise un deoxynucléotide (en effet ensuite la matrice devient simple brin). C'est une adénine qui est préférentiellement ajoutée, on fera donc le marquage avec du dATP $\alpha^{32}\text{P}$. Si l'ADN présente une activité 3'5' sortante, il faut utiliser une polymérase ayant une activité 3'5' exonucléase. Cette activité retire l'extrémité sortante, génère une extrémité 5'3' sortante qui est remplie par l'activité polymérase.

Hybridation

1) Dissociation ou dénaturation

Lorsque l'ADN double brin est chauffé à une température dite de fusion, ou T_m (pour « melting temperature »), les deux brins se séparent suite à la rupture des liaisons hydrogènes qui les maintiennent appariés. La double hélice se défait, il y a perte de la structure secondaire, on dit que l'ADN est dénaturé. Ce terme est employé pour toutes les macromolécules. Par exemple, lorsqu'on détruit les liaisons faibles d'une protéine, il y a perte de la structure tridimensionnelle, la protéine est dénaturée.

Le T_m dépend de deux facteurs principaux : du nombre de liaisons hydrogènes et de la composition du milieu.

Le nombre de liaisons hydrogène lui-même dépend :

- de la longueur du fragment : le T_m augmente avec la longueur. Toutefois, le nombre de liaisons hydrogènes est important en dessous de 150-200 liaisons. Au-delà, la dénaturation devient principalement un phénomène coopératif, et le nombre de bases n'est plus important. On tiendra donc compte de la longueur des fragments principalement pour l'hybridation des oligonucléotides.
- de la composition en bases : l'augmentation de la proportion en GC augmente le T_m . Il y a en effet 3 liaisons hydrogènes entre les bases G et C et 2 entre les bases A et T. Donc plus la proportion en GC est importante, plus la température de fusion sera élevée.
- de la présence de mésappariements : Les mésappariements abaissent le T_m , puisque au niveau du mésappariement, il n'y a pas de liaison hydrogène.

Plusieurs éléments du milieu contribuent à l'établissement des liaisons hydrogènes et seront utilisés pour faire varier le T_m .

- la force ionique. L'augmentation de la concentration en cations monovalents tel que le NaCl joue sur le T_m . Le NaCl masque les charges négatives des phosphates et ainsi diminue les forces de répulsion électrostatique entre les deux brins. Ainsi moins il y a de sel, plus les forces de répulsion sont importantes, plus le T_m est bas.

- Certains composés tels que la formamide ou l'urée abaissent le T_m . Ces composés forment des liaisons hydrogènes avec les bases et donc entrent en compétition avec les liaisons interbases.
- le pH est aussi important. Aux pH extrêmes, l'ADN est dénaturé. A température ambiante, on utilise souvent le NaOH pour dénaturer l'ADN.

Calcul du T_m

- Pour les fragments de plus d'un kilobase (kb) on utilise souvent l'équation suivante pour estimer le T_m :

$$T_m = 81,5 + 16,6 \log M + 41 (G+C)$$

Où M est la concentration en cation monovalent et (G+C) représente la proportion de bases G et C et (Schildkraut et Lifton , 1965).

- Pour des fragments plus petits, pour les oligonucléotides, on utilise l'expression :

$$T_m = 81,5 + 16,6 \log M + 41 (G+C) - 500/L$$

L représente la longueur de l'oligonucléotide

- Si on utilise un agent dénaturant, abaissant le T_m tel que la formamide. On utilise l'équation suivante

$$T_m = 81,5 + 16,6 \log M + 41 (G+C) - 500/L - 0,62 F$$

F représente la concentration molaire en formamide.

- Ces méthodes ne donnent que des approximations du T_m et la plupart des logiciels et serveurs utilisent le model du « plus proche voisin » et des données thermodynamiques pour estimer le T_m . L'équation la plus utilisées est (SantaLucia et coll., 1996, Panjkovich et Melo, 2004)

$$T_m = \frac{\sum (\Delta H_d) + \Delta H_i}{\sum (\Delta S_d) + \Delta S_i + \Delta self + R \ln \frac{C_T}{b}} + 16.6 \log [Na^+]$$

La somme des enthalpies (ΔH_d) et des entropies (ΔH_i) sont calculés pour tout les doublets. $\Delta self$ est le désavantage entropique pour les séquences autoccomplémentaires, ΔH_i et ΔS_i sont les sommes des enthalpies et entropies d'initiation, $\Delta self$ est utilisé pour les séquences

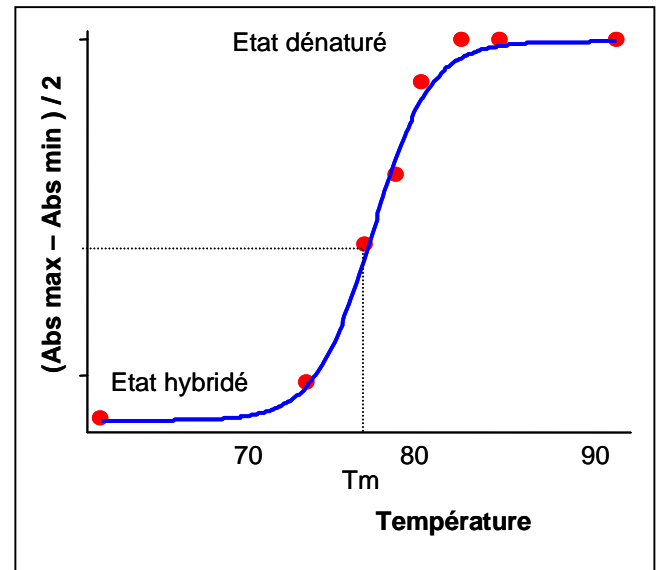
autocomplémentaires, R est la constante des gaz parfaits (1.987 cal/K.mol), C_T est la concentration en mole.Litre⁻¹, T_m est la température de fusion en Kelvin, la constante b prends la valeur 1 lorsque l'oligonucléotide est en excess par rapport à sa cible.

Mise en évidence des états natifs et dénaturés

La dénaturation de l'ADN s'accompagne de modifications importantes des propriétés physico-chimiques:

- la viscosité augmente
- l'absorption à 260 nm augmente car il n'y a plus d'interactions entre les bases. Cette méthode permet d'estimer le T_m .
- la densité augmente.

On peut séparer l'ADN simple brin de l'ADN double brin par plusieurs méthodes :



- par chromatographie sur colonne d'hydroxyapatite: ce sont des cristaux de phosphate de calcium. En forte concentration saline, seuls les acides nucléiques double brin se fixent, les acides nucléiques simple brin ne sont pas retenus. Les ADN double brin peuvent être ensuite élués par du phosphate.
- par filtration sur nitrocellulose, en effet seul l'ADN simple brin se fixe, l'ADN double brin passe à travers.
- en utilisant la différence de densité, en gradient isopycnique.

II) Hybridation

L'hybridation correspond à l'association de deux brins d'acide nucléiques complémentaires. Cette hybridation peut se faire entre deux brins d'ADN, deux brins d'ARN ou un brin d'ARN et un brin d'ADN.

La dénaturation est réversible. Quand la température est abaissée progressivement jusqu'au point de fusion (T_m), les molécules peuvent s'hybrider selon la règle de complémentarité des bases. La réassociation, comme la dénaturation, dépend du T_m : pour qu'il y ait hybridation, il faut que la

température soit inférieure au T_m , mais l'hybridation dépend également de la concentration en ADN et du temps. Ces deux derniers paramètres sont importants pour la rencontre entre les deux brins.

Ainsi, si on veut garder l'ADN sous forme dénaturée, on laisse l'ADN soit à une température élevée, soit on le refroidit brusquement, on évite ainsi les mouvements moléculaires et donc la probabilité de rencontre entre deux brins complémentaires.

1) Hybridation en phase liquide

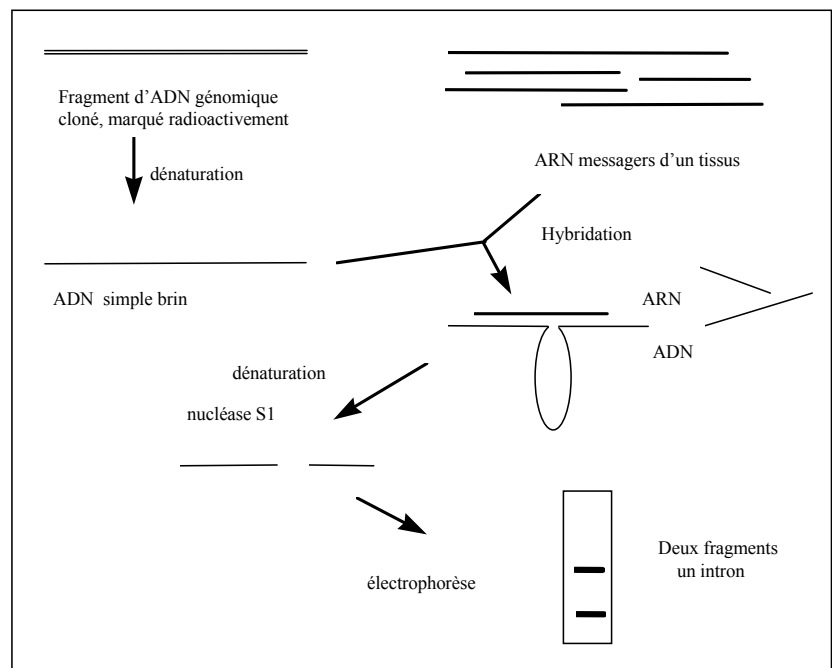
Dans ce cas les deux brins sont en solution. On utilise l'hybridation en phase liquide dans plusieurs cas, par exemple :

a) Pour mesurer le T_m , en utilisant le fait que les acides nucléiques simples brins absorbent plus que les doubles brins.

b) Pour délimiter les introns et les exons d'un gène.

On utilise ici un fragment d'ADN génomique qui est cloné (inséré dans un plasmide). On peut donc en avoir une grande quantité.

- L'ADN est marqué de manière spécifique en général à une extrémité. Il est ensuite dénaturé puis hybridé avec une population d'ARN messager provenant d'une population de cellules (tissus, individu ou culture cellulaire)
- Seules les parties complémentaires s'hybrident, les



autres restent simple brin.

- L'ADN simple brin est digéré par la nucléase S1 qui ne digère que le simple brin (ARN ou ADN). Les fragments d'ADN sont visualisés sur gel après autoradiographie.

c) Comparaison des tailles de génome sans séquence répétitive. On met dans un tube une préparation d'ADN d'un organisme. Du fait des manipulations effectuées lors de la purification, le génome est représenté par des fragments d'environ 100 kilobases (kb). On chauffe l'échantillon au-dessus du T_m puis on le refroidit à une température en dessous du T_m . Plus le génome est grand, ou plus exactement plus il est complexe, moins deux brins complémentaires auront de chance de se rencontrer pour une concentration d'ADN donnée. A l'inverse, moins le génome est complexe, plus la concentration des brins complémentaires sera élevée, plus l'hybridation sera rapide. On a donc ici un moyen d'estimer la complexité relative d'un ADN.

d) Hybridation d'une amorce (les DNA polymérases ont besoin d'une amorce. *In vivo*, cette amorce est fournie par la primase qui est une RNA polymérase. *In vitro*, l'amorce est le plus souvent un oligonucléotide qui est synthétisé chimiquement).

- Hybridation d'extrémités cohésives.

Les DNases de restriction génèrent souvent des fragments complémentaires. Ces fragments peuvent s'hybrider ce qui va faciliter une ligation.

Le phage λ présente à ses extrémités des bouts complémentaires (extrémités cohésives ou cos qui peuvent s'hybrider sur 12 bases)

2) Hybridation sur support solide

L'immobilisation de l'une des séquences complémentaires facilite certaines manipulations quoiqu'elle soit souvent moins sensible du fait que le support masque souvent une partie des bases.

a) Fixation sur gel de chromatographie :

On utilise par exemple la chromatographie d'affinité sur oligodT pour purifier les ARN polyA⁺. On utilise un support de chromatographie sur lequel on a fixé des oligonucléotides ne comportant que des T (oligodT).

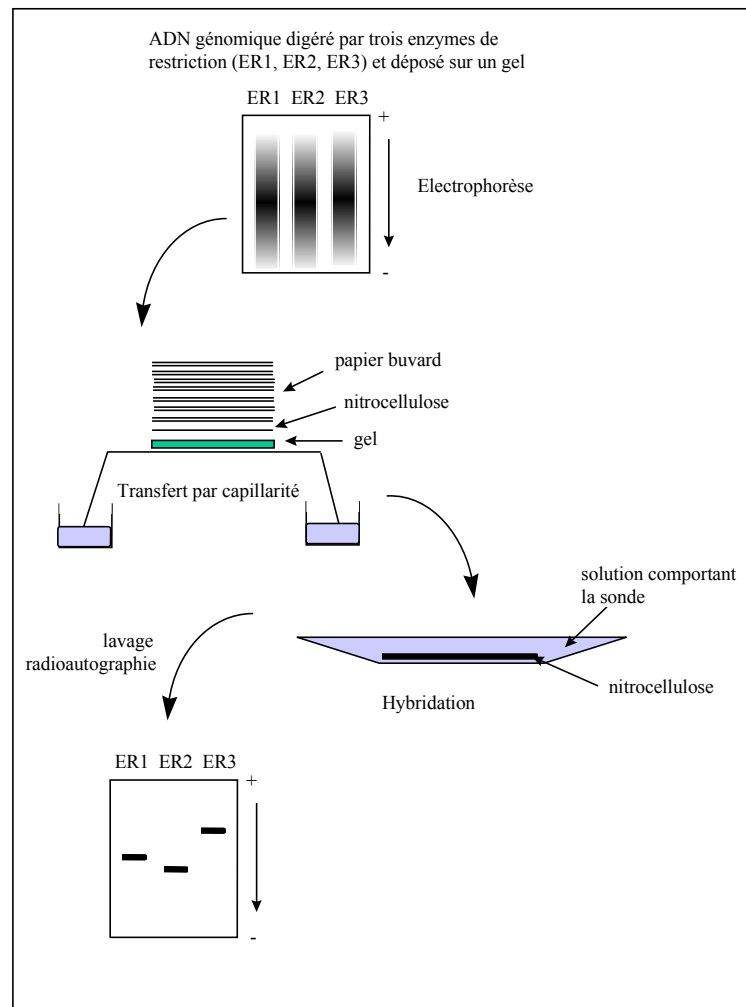
On passe sur la colonne une solution d'ARN totaux eucaryotes, dans des conditions inférieures au T_m. Les ARN messagers polyadénylés en 3' s'hybrident aux oligodT de la colonne. Par contre, les ARN ribosomiques et ARN de transfert passent à travers sans être retenus. Pour décrocher les ARN messager hybridés dans la colonne, il suffit d'augmenter le T_m par exemple en diminuant la concentration de sel de la phase mobile et en chauffant. A partir d'une solution d'ARN totaux, on obtient une solution fortement enrichie en ARN polyadénylés c'est à dire en ARN messager. La fraction retenue correspond aux ARN polyA⁺, la fraction non retenue aux polyA⁻.

b) Fixation sur membrane:

Il y a plusieurs sortes de membranes : la nitrocellulose, le Nylon ou la lamelle de verre. La fixation de l'ADN sur la nitrocellulose s'effectue par cuisson à 80°C, et l'irradiation aux UV (254 nm) permet de réaliser des liaisons covalentes entre les acides nucléiques et le Nylon. La fixation sur lamelle de verre s'effectue grâce à des linkers, des molécules fonctionnalisées qui permettent de lier des fragments d'ADN par leur extrémité à l'aide de liaisons chimiques covalentes.

Le Southern, le northern et la puce à ADN sont quelques exemples d'utilisation de ce type de support.

* Southern



L'ADN est coupé par des enzymes de restriction qui coupent à des séquences précises (la digestion doit être totale)

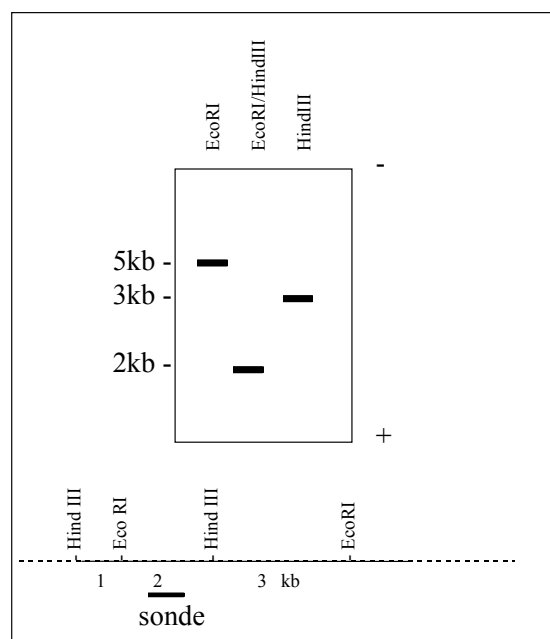
- Le milieu de digestion est déposé sur un gel électrophorèse. Les fragments d'ADN se séparent, ils migrent plus ou moins vite en fonction de leur taille, les plus petits migrent le plus vite.
- Les fragments séparés sur le gel sont ensuite transférés sur une membrane de nitrocellulose ou de Nylon puis dénaturés (alternativement la dénaturation peut se faire avant le transfert).
- Les différents fragments d'ADN sont ensuite fixés sur la membrane et la membrane est saturée avec de l'ADN simple brin.
- La membrane est ensuite plongée dans une solution contenant un fragment d'ADN marqué radioactivement simple brin (sonde). La membrane ayant été saturé à l'étape précédente, la sonde ne s'accrochera pas à la membrane.

- Le bain est maintenu à une température inférieure au T_m . L'hybridation a lieu lorsqu'il y a une séquence complémentaire à la sonde sur la membrane.
- La membrane est lavée à une température proche du T_m de façon à retirer les hybridations non spécifiques.
- On effectue enfin une autoradiographie pour voir où la sonde s'est accrochée

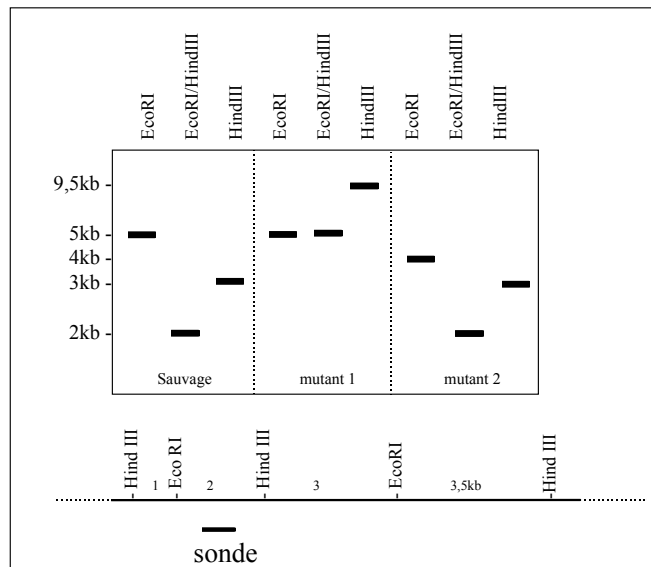
Informations obtenues par le Southern :

- Nombre de gènes, dans ce cas il faut que la sonde provienne d'un ADN génomique.
- Cartographie
- Détection de mutations

Exemple de cartographie obtenue avec un Southern.



L'ADN génomique a été digéré par les enzymes de restriction Eco RI et/ou HindIII. Les produits de la digestion ont été déposés sur un gel d'électrophorèse puis transférés sur nitrocellulose. A la suite de l'hybridation avec la sonde, on observe une seule bande. Il n'y a donc qu'une seule région dans le génome reconnue par la sonde. La taille des fragments hybridés permet de positionner les sites de restriction dans cette région correspondants aux deux enzymes. On peut donc établir une carte de la région.



Exemple d'utilisation de la technique de Southern pour détecter des mutations.

Dans le mutant 1, le site Hind III situé dans le fragment Eco RI de 5 kb est muté et n'est plus coupé par l'enzyme. Dans le mutant 2, une délétion de 1 kb peut être détectée dans le fragment Hind III-EcoRI de 3 kb.

Il y a des informations qui ne sont pas obtenues par le Southern comme par exemple la taille d'un gène. Il n'y a en effet aucune relation entre l'emplacement d'un site de restriction et la position d'un promoteur ou d'une séquence transcrite.

Les températures d'hybridation et de lavage de la membrane sont importantes pour l'hybridation. Si on travaille à une température légèrement inférieure au T_m , on n'obtient que des hybridations homologues, spécifiques. Si on travaille à une température plus élevée que le T_m , on n'a aucune hybridation. Si on travaille à une température plus basse que le T_m on va pouvoir hybrider des séquences présentant des mésappariements avec la sonde. On va donc hybrider des gènes qui ont des séquences proches et qui peuvent par exemple être des gènes codant pour des protéines de la même famille ou des promoteurs ayant les mêmes séquences de régulation. Généralement on travaille

en dessous du T_m pour l'hybridation, on la favorise puis on lave dans des conditions plus proches du T_m de façon à ne garder que les hybridations homologues.

* Northern

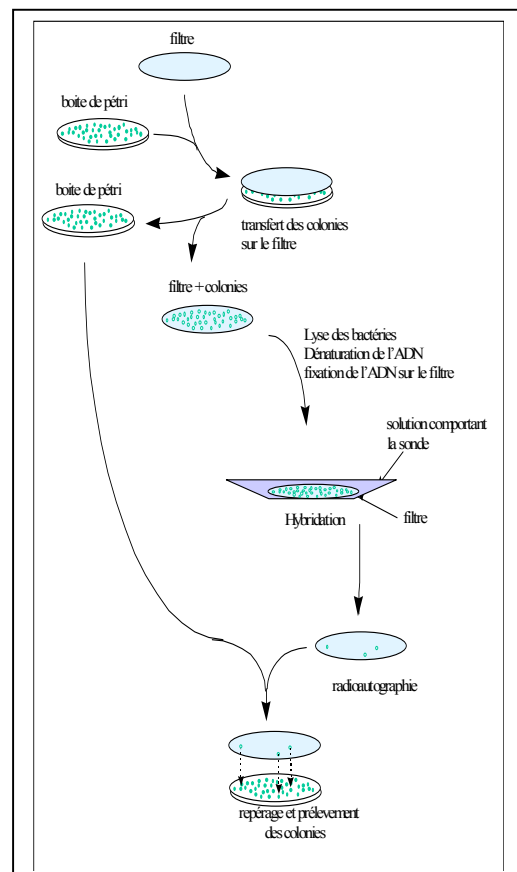
C'est exactement le même principe que le Southern, mais avec de l'ARN comme cible. Les ARN messagers ont des tailles variables, on peut donc les séparer sur gel.

Information obtenue par le northern :

- nombre de transcrits
- taille d'un ARN

* Hybridation sur colonie

Des colonies sont étalées sur une boîte de pétri. Une fois qu'elles ont poussé, on fait une réplique sur une membrane de nitrocellulose. Les colonies sur la membrane de nitrocellulose sont lysées par la soude, l'ADN est ainsi libéré et dénaturé. Il se fixe sur la nitrocellulose à l'emplacement de la colonie. Après avoir saturé la membrane avec de l'ADN neutre, qui est différent de l'ADN présent dans les colonies, on hybride la réplique avec une sonde marquée. L'hybridation aura lieu seulement sur les séquences complémentaires. L'emplacement de l'hybridation permettra de connaître la présence d'une séquence complémentaire de la sonde dans une des colonies étalées sur la boîte. Cette méthode permet de cribler des banques de plasmide.



* Hybridation sur puce à ADN.

Une puce à ADN est une lamelle de verre sur laquelle on a fixé des brins d'ADN qui sont souvent des oligonucléotides. Les brins sont arrangés de telle manière à connaître la séquence correspondant à chaque spot. Le nombre de spots peut être très élevé, jusqu'à 100 000 par lamelle de la taille d'une lamelle utilisée conventionnellement en microscopie. La sonde est marquée à l'aide d'un composé fluorescent et la détection s'effectue à l'aide d'un laser. La présence d'une hybridation indique qu'il existe une séquence complémentaire dans la sonde.

Par exemple si on spotte toutes les séquences des gènes connus chez l'homme et qu'on hybride avec les ARN d'un type cellulaire, les gènes qui hybrident correspondent à ceux qui sont exprimés dans les cellules ayant servi à préparer la sonde.

On peut aussi spotter les gènes correspondant à une bactérie qui contamine les aliments. On prépare une sonde à l'aide d'ADN extrait de l'aliment, lorsqu'il y a hybridation on en déduit que la bactérie est présente dans l'aliment.

* Hybridation in situ

On peut considérer deux sortes d'hybridation in situ, l'hybridation sur chromosome et l'hybridation des ARN. Dans le premier cas on localise le locus où s'hybride la sonde, dans le deuxième cas on localise le patron d'expression d'un ARN dans un tissu, un organe ou même un organisme ...

Hybridation sur chromosome

L'hybridation est effectuée sur une préparation de chromosomes. Plusieurs applications peuvent être envisagées.

→ La localisation d'un clone. Lorsqu'on a cloné un gène et que l'on ne connaît pas sa localisation chromosomique, cette localisation peut être effectuée par hybridation *in situ*.

→ Une application récente de ce type d'hybridation est le coloriage des chromosomes (FISH). On fabrique des sondes correspondant à chacun des chromosomes mais avec des fluorophores différents.

Après hybridation on peut distinguer chaque chromosome les uns des autres. Cette méthode permet d'une part de distinguer des chromosomes de tailles proches et d'autre part, elle permet de caractériser des remaniements chromosomiques. Le nombre de fluorophores doit être a priori aussi grand que le nombre de chromosome à identifier, mais on peut aussi combiner plusieurs chromophores sur la même sonde ce qui diminue fortement le nombre de chromophores à utiliser de n à $2^n - 1$. Autrefois, cette hybridation se faisait sur des chromosomes en métaphase (ou, dans certains cas, sur des chromosomes polythènes). Maintenant, on réalise ces hybridations sur des chromosomes étirés (peignés).

Le séquençage des acides nucléiques

1) Les différentes méthodes de séquençage

1) Méthode chimique ou méthode de Maxam et Gilbert

1) Marquage de l'ADN en 5'. Il faut ici que l'ADN ne soit marqué qu'à une seule extrémité.

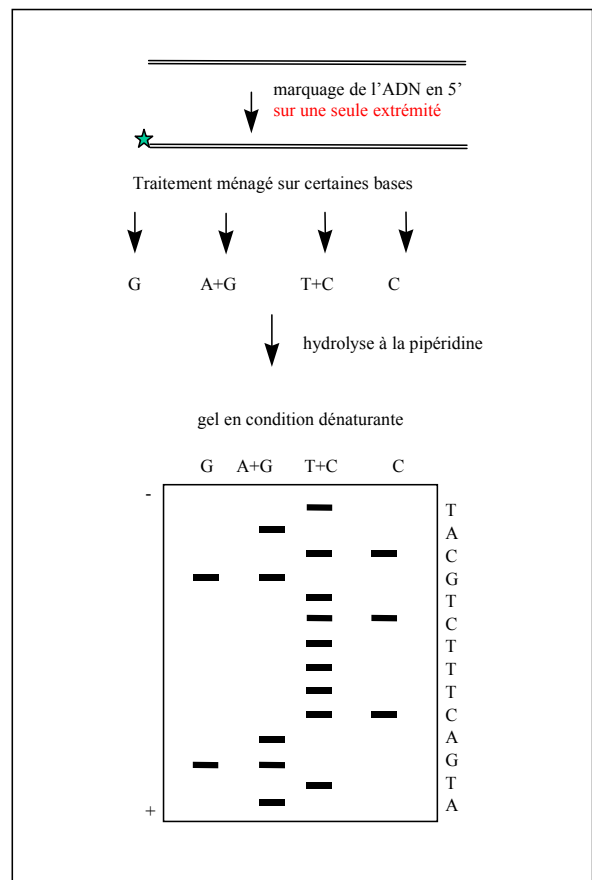
2) coupure de l'ADN par des réactions spécifiques à certaines bases, par exemple après toutes les guanosines. La coupure est partielle si bien qu'on obtient toute une série de molécules dont l'extrémité 3' est un G, parmi celles-ci certaines d'entre elles sont marquées en 5' par du ^{32}P .

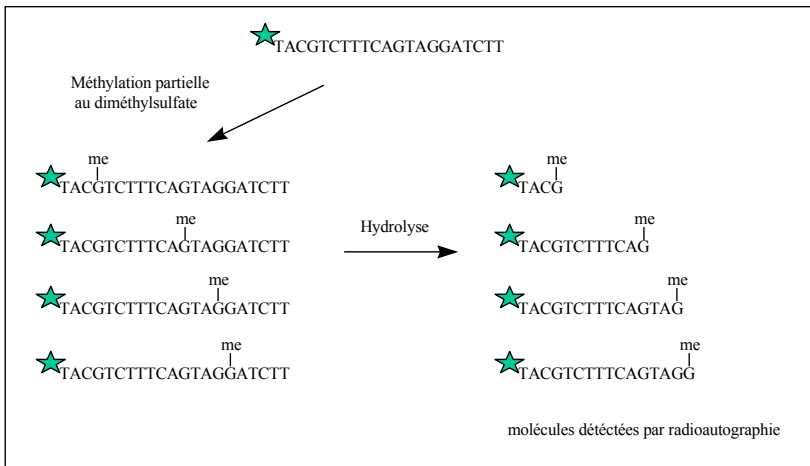
3) dénaturation de l'ADN par la chaleur

4) dépôt sur un gel électrophorèse en condition dénaturante (température haute et présence d'urée dans le gel). Les fragments d'ADN migrent en fonction de leurs tailles et sous forme simple brin.

5) Radioautographie, seuls les fragments marqués, c'est à dire seul ceux qui portent l'extrémité 5'- et donc commencent au même endroit-, sont visualisables. Certaines molécules s'arrêtent au premier G, d'autre au second etc...

Si au départ on a fait des coupures séparément et différentes avec plusieurs agents, on obtient pour chaque tube, des molécules de taille correspondant à la séquence. On ne dispose pas de schéma réactionnel permettant de couper après chacune des bases. Mais on dispose de moyens pour couper après G, après A+G, après C+T et après A. Si on charge ces quatre séquences sur le gel, on peut lire la séquence.



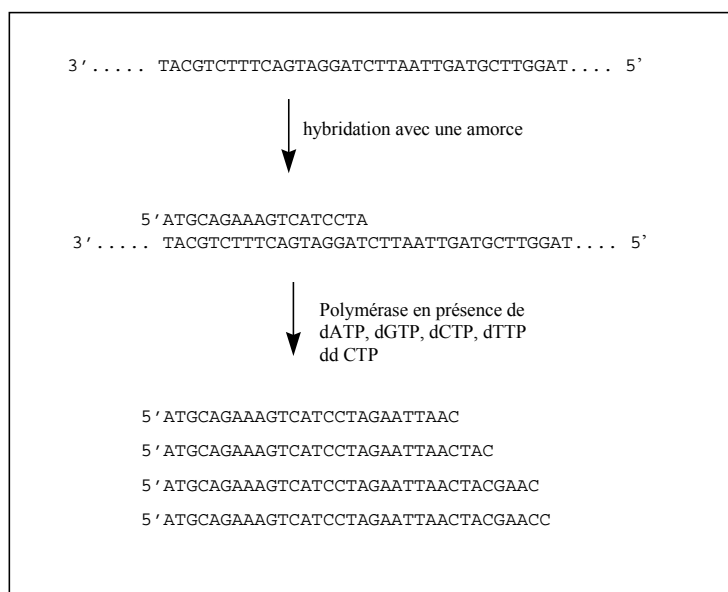


Cette méthode n'est plus guère utilisée sauf dans certains cas particuliers, par exemple lorsqu'on veut détecter la liaison d'une protéine sur l'ADN. En effet dans ce cas la liaison de la protéine empêche la coupure chimique, certains fragments sont absents et on observe des blancs sur le gel. Cette technique a été appelée "foot printing".

2) Méthode enzymatique ou méthode de Sanger

La première étape est l'obtention d'un ADN simple brin. On peut dénaturer l'ADN double brin par la chaleur ou par un traitement à la soude. On peut aussi utiliser une matrice simple brin (voir plus loin). La deuxième étape est l'hybridation d'une amorce sur l'ADN simple brin. La troisième étape est la polymérisation du brin complémentaire dont l'extrémité 5' correspondra à l'extrémité 5' de l'amorce.

Le principe consiste à ajouter des didéoxynucléotides (ddNTP) aux désoxynucléotides lors de la polymérisation. Ainsi, si on ajoute du ddATP, aux dNTP, la polymérisation s'arrêtera lorsque la polymérase ajoutera un ddATP, mais continuera lorsqu'elle ajoutera un dATP. On aura donc une série de molécules s'arrêtant toutes à un A, ces molécules seront plus ou moins longues.



Parallèlement, dans trois autres tubes on ajoutera aux dNTP du ddCTP, du ddTTP ou du ddGTP. On obtiendra ainsi des molécules s'arrêtant soit aux C soit aux T soit aux G en fonction du ddNTP utilisé.

Dans les quatre tubes on a donc des molécules dont la taille dépend de la séquence. Pour les séparer on peut faire un gel d'électrophorèse en condition dénaturante de façon à ne voir que les molécules néosynthétisées.

La quantité d'ADN est faible et ne peut donc pas être directement visible sur le gel. Aussi pour le détecter, on marque le brin synthétisé en ajoutant un nucléotide radioactif lors de la polymérisation.

Quelle ADN polymérase utiliser ?

La plupart des ADN polymérases catalysent l'addition des ddNTP à environ 0.02 -0.1 % de la vitesse d'incorporation des dNTP. La T7 DNA polymérase est une exception. Elle incorpore les ddNTP plus efficacement, avec un taux de 20 % comparé aux dNTP. En comparant les séquences des ADN polymérases, on s'est aperçu qu'une phénylalanine conservée était mutée en tyrosine chez l'ADN polymérase T7. La différence d'activité provient du groupement hydroxyle de la tyrosine. En effet le remplacement de la tyrosine 526 de l'ADN polymérase T7 en phénylalanine diminue l'efficacité de l'utilisation des ddNTP de plus de 1000 fois et à l'inverse la mutation de la phénylalanine 667 de l'ADN polymérase Taq en tyrosine augmente son efficacité d'incorporation des ddNTP de 8000 fois (Tabor et Richardson, 1995). La présence de ce groupement OH compense l'absence du même groupement sur le didésosyribose.

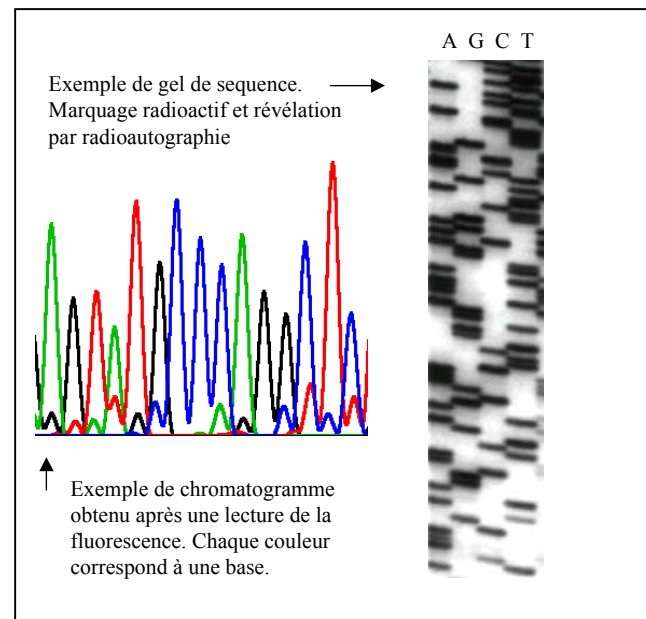
Les ADN polymérases de la famille de l'ADN polymérase I de E. coli ne sont que peu processives et n'ajoutent à chaque événement qu'une vingtaine de nucléotides. L'ADN polymérase T7 est, elle aussi peu processive, mais cette processivité est augmentée en ajoutant de la thioredoxine qui se complexe à cette polymérase. Ce complexe ADN polymérase T7 -thioredoxine a été nommé Sequenase.

A partir de cette technique, plusieurs variantes ont été développées :

a) Le marquage, à l'origine radioactif, peut aussi être fluorescent.

On peut utiliser une amorce fluorescente (méthode du « dye primer »).

La réaction est toujours faite dans quatre tubes différents. Pour chaque ddNTP on utilise une amorce marquée avec un fluorophore différent. Les réactions sont chargées sur un gel et les bandes sont détectées par un laser. On peut utiliser des ddNTP fluorescents (méthode du «dye terminator»). Dans ce cas les réactions se font toutes dans le même tube. L'avantage de cette méthode par rapport au marquage de l'amorce est qu'on peut utiliser n'importe quelle amorce pour la réaction de séquence sans avoir à la modifier au préalable.



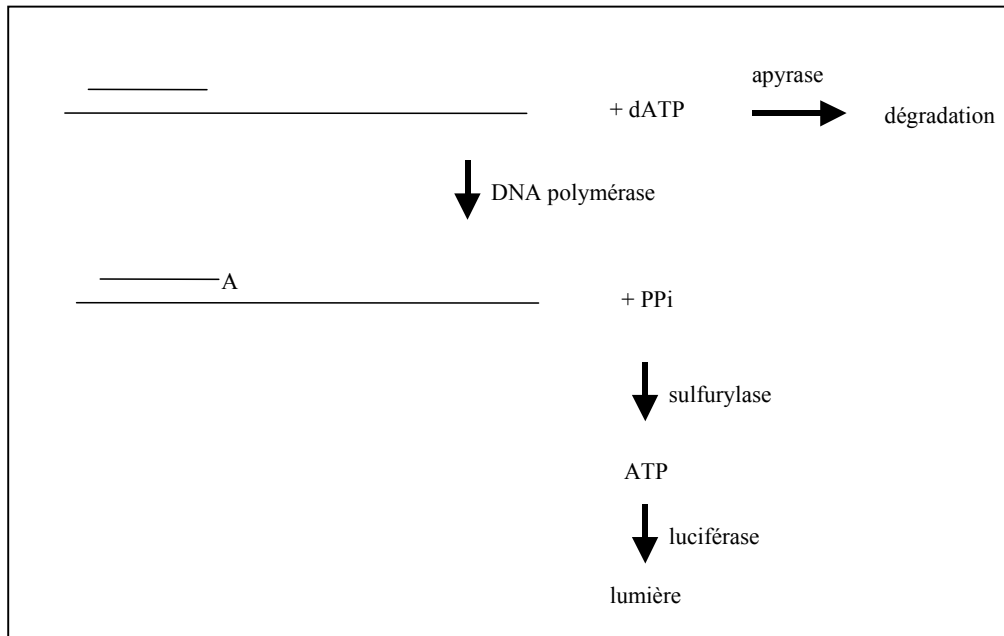
b) Plusieurs techniques peuvent être utilisées pour déterminer la longueur des brins. A l'origine la détection se faisait après migration sur gel d'électrophorèse, on peut utiliser d'autres techniques comme l'électrophorèse capillaire ou la spectrométrie de masse.

c) Dans certains cas on observe des "compressions de bandes" qui sont dues à des structures secondaires intra-chaîne et qui ne sont pas complètement dénaturées lors de l'électrophorèse. La présence de ces compressions peut être diminuée en utilisant des analogues de nucléotides durant la polymérisation. Le 7-deaza dGTP (C^7 dGTP) ou le dITP sont par exemple utilisés à la place du dGTP (Barr *et al.*, 1986).

3) « Pyrosequencing »

Cette méthode est basée sur la détection du pyrophosphate lors de l'incorporation d'un nucléotide par une polymérase (Ronaghi *et al.*, 1998). Le fragment d'ADN d'intérêt (amorce hybridée à une matrice simple brin) est incubé avec une DNA polymérase, une ATP sulfurylase, une luciférase et une enzyme de dégradation des nucléotides telle que l'apyrase. Des nucléotides sont ensuite ajoutés par cycle, tout d'abord du dATP, puis du dTTP, du dGTP et du dCTP. Lorsque le nucléotide complémentaire est ajouté, il y a libération du pyrophosphate (PPi). Celui-ci est dosé à

l'aide de la luciférase qui produit de la lumière en présence d'ATP. Les nucléotides non incorporés sont dégradés entre chaque cycle par une apyrase, mais le temps de dégradation est plus long que celui de l'incorporation de façon à ne pas interférer avec la réaction de polymérisation.



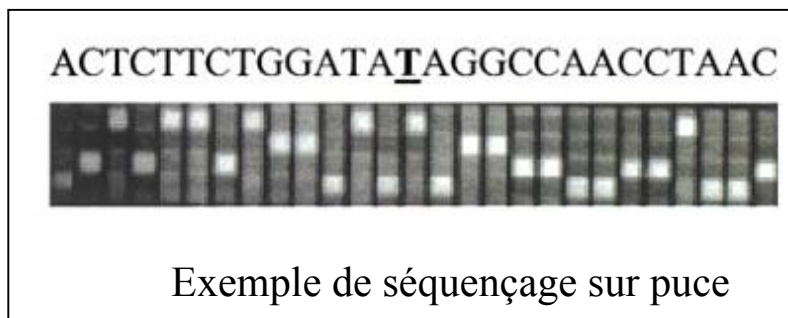
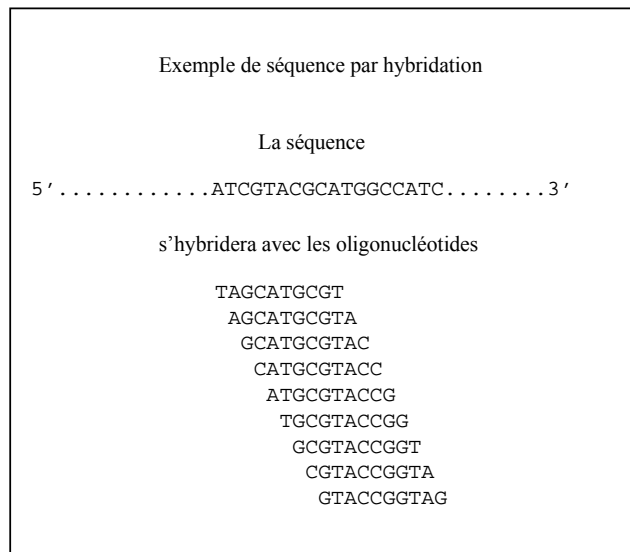
Une des difficultés est la lecture de domaines homopolymériques (par exemple si plusieurs T se suivent). En effet dans ce cas le signal est plus intense mais n'est pas directement proportionnel au nombre de bases.

Cette méthode permet de séquencer uniquement des petits fragments, mais sa simplicité permet d'en séquencer un grand nombre en parallèle.

4) Séquençage par hybridation

Cette méthode utilise les puces à ADN (DNA chips). Une puce à ADN est une petite surface sur laquelle on dépose une très grande quantité (plusieurs milliers) de molécules d'ADN. Dans le cas présent on dépose des oligonucléotides qui sont directement synthétisés sur la surface de silicium ou de polymère d'acrylamide. Le fragment d'ADN à séquencer est marqué à l'aide d'un fluorophore et on effectue une hybridation. La lecture de la

puce est réalisée en liaison avec des lecteurs optiques qui balayent la surface des puces grâce à des techniques de micro-optique.



II) Les techniques de marche pour séquencer de grands fragments

La lecture d'un gel séquence est limitée par la taille et la résolution du gel et, suivant les techniques, on peut lire jusqu'à 300 ou 700 pb. Pour lire au-delà, on utilise i) soit une deuxième amorce ii) soit on fabrique des clones plus courts à l'aide d'une exonucléase iii) soit on insère une séquence connue dans la séquence à séquencer.

Utilisation de nouveaux primers (primer walking).

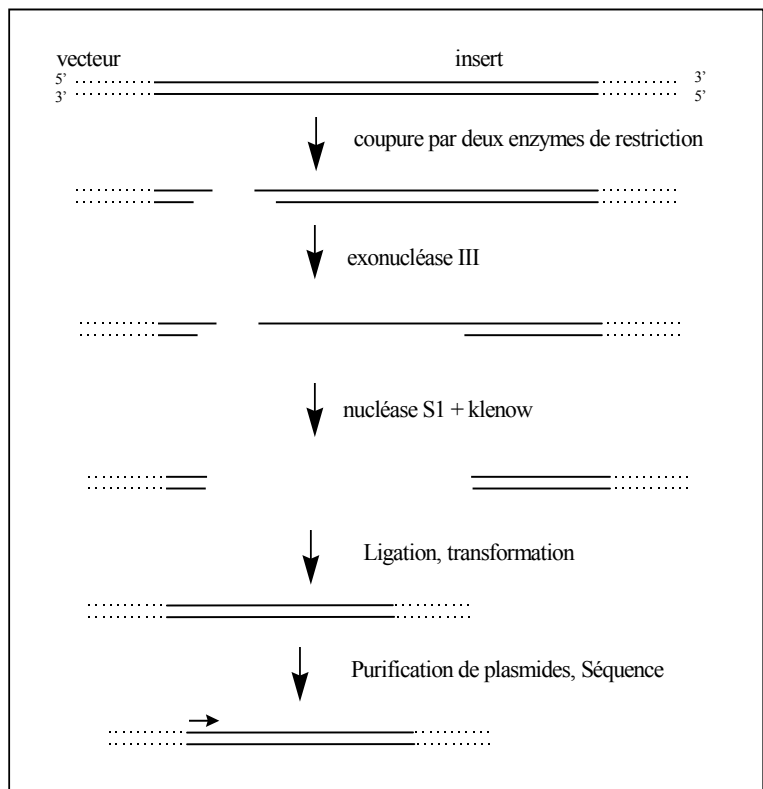
Lorsqu'une première séquence a été effectuée, on peut se servir de la séquence lue à l'extrémité comme point d'amorçage de la séquence suivante en synthétisant une nouvelle amorce. C'est la méthode la plus pratique et de plus en plus utilisée avec la baisse du coût de synthèse des amorces et l'augmentation de taille des séquences à réaliser. Toutefois cette méthode nécessite d'effectuer les séquences séquentiellement ce qui allonge d'autant la durée de l'expérience lorsque les séquences à découvrir sont longues.

Coupure au hasard

L'insert est ici purifié puis coupé au hasard par les ultrasons. Les extrémités sont rendues franches par un traitement à la Klenow puis l'insert est sous cloné dans un vecteur. Comme alternative on peut couper l'insert par des enzymes de restriction. Toutefois, il faudra faire attention que les fragments ne soit pas chevauchants et, de plus, il peut toujours y avoir deux sites de restriction très proches l'un de l'autre.

Délétion ménagée.

Le plasmide est coupé dans le polylinker par deux enzymes de restrictions en amont de l'insert, une donnant une extrémité 3' sortante et l'autre donnant une extrémité franche ou une extrémité 5' sortante. Le produit de la digestion est alors incubé avec l'exonucléase III qui ne digère que l'insert du côté où il y a une extrémité 5' sortante. Le vecteur est lui protégé par son extrémité 3' sortante. La réaction est stoppée puis les extrémités sont rendues franches par la nucléase S1 suivie par un traitement par une polymérase (la nucléase S1 est peu efficace lorsqu'il ne reste qu'une base sortante).



Les plasmides sont ensuite recircularisés et une amorce hybridant sur le vecteur peut être utilisée pour continuer la séquence.

La réaction de l'exonucléase III doit être ménagée. En fait, toutes les 200-400 bases, on prélève un aliquote et on arrête la réaction. On dispose ainsi d'un ensemble de clone plus ou moins délétés qui peuvent tous être séquencés avec la même amorce.

Il y a deux difficultés dans cette technique.

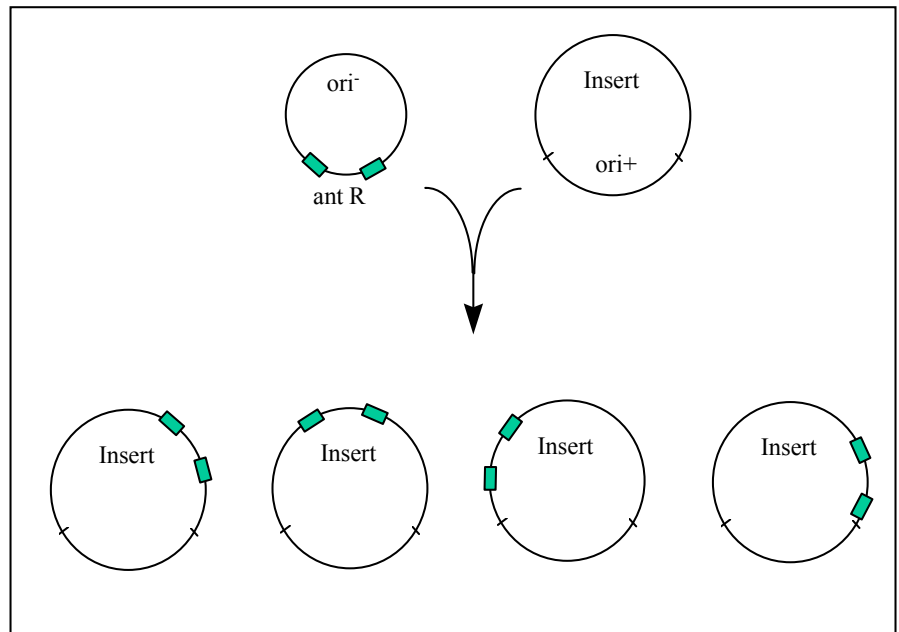
1) La digestion par les enzymes de restriction doit être totale. En effet autrement on se retrouve avec un grand nombre de clone non délétés.

2) La digestion par l'exonucléase III doit être bien calibrée pour enlever le nombre de base désiré. Avant la ligation, on peut charger un aliquote de chaque tube sur un gel pour vérifier l'étendue de la digestion.

Transposition

On transforme le clone bactérien avec un second plasmide qui induit la transposition d'une séquence (un transposon) dans la séquence désirée. La séquence transposée étant connue, on peut s'en servir pour l'amorçage.

Le transposon doit s'insérer au hasard, il ne doit pas avoir de spécificité d'insertion. On utilise la transposase TnsABC de Tn7. Le plasmide donneur ne doit pas pouvoir se répliquer. Le transposon est donc sur un



plasmide qui ne peut pas se répliquer dans la souche d'*E. coli* utilisée.

Le transposon contient un gène de résistance à un antibiotique ce qui permet de sélectionner les clones pour lesquels la transposition a eu lieu.

Pour chaque clone on peut donc faire deux séquences, une de chaque côté de l'insertion.

III) La préparation de la matrice

Le séquençage par la méthode enzymatique requiert l'hybridation d'une amorce sur un ADN simple brin. Il y a deux méthodes pour obtenir un simple brin, soit utiliser un phage simple brin soit dénaturer l'ADN double brin par la chaleur.

Les premières techniques de séquençage se faisaient en sous clonant l'insert dans un phage M13, lorsqu'il est sous sa forme répliquative. Les phages étaient purifiés et on extrayait leur ADN simple brin puisqu'un seul des deux brins (+) se retrouve dans le phage infectieux. Cette méthode

n'est plus employée, mais les séquences en amont des inserts ont été recopiées sur la plupart des plasmides et les amorces ont pris les noms d'universal primer ou de reverse primer.

Dans un deuxième temps l'origine de réplication de M13 (ou d'un autre phage proche, le phage f1) ont été ajoutées dans les plasmides permettant de fabriquer un ADN simple brin à l'aide d'une infection par un phage se répliquant peu (phage helper).

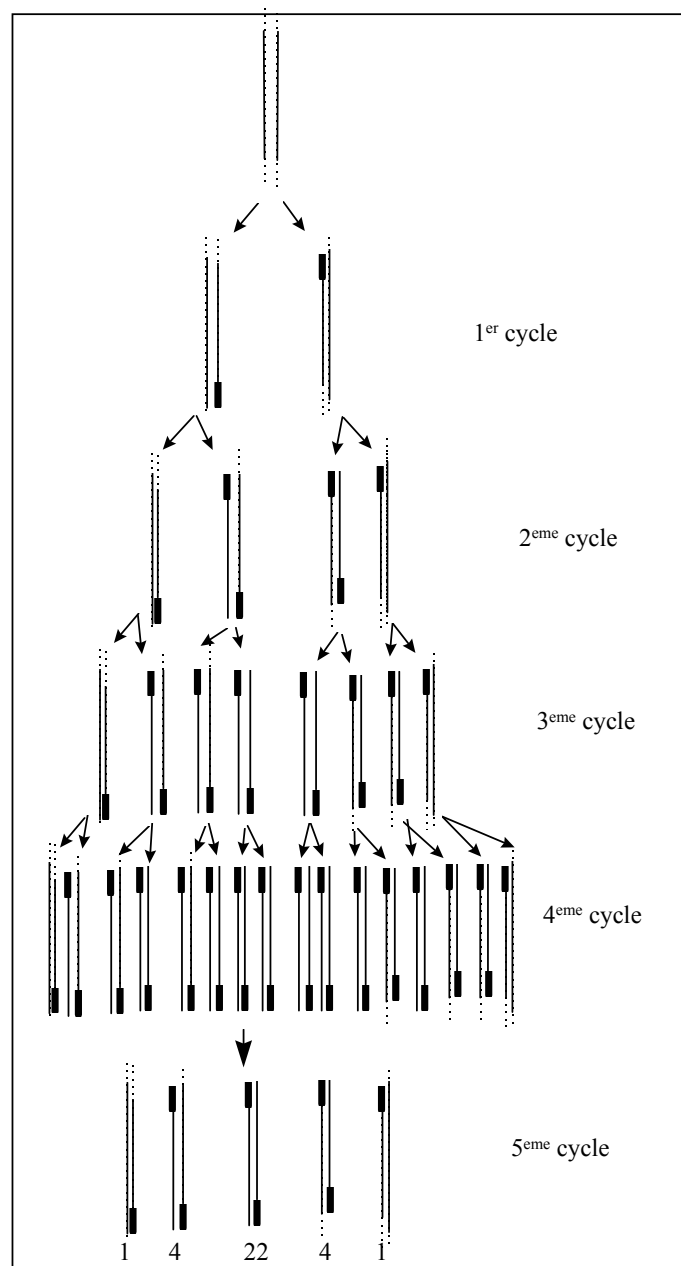
Ces méthodes ont été abandonnées avec l'arrivée de techniques de séquençage double brin (on travaille avec un plasmide double brin que l'on dénature à la chaleur) ainsi que grâce à des souches bactériennes plus adaptées.

Amplification d'acides nucléiques

Amplification d'ADN : PCR

Réaction de base

On ajoute dans un tube de l'ADN qui sert de matrice, deux oligonucléotides hybridant sur chacun des deux brins de l'ADN, une polymérase thermostable et les 4 desoxynucléotides triphosphates. On chauffe le tube à 95°C, les deux brins d'ADN se séparent. On abaisse la température en dessous du T_m des deux oligonucléotides, ils s'hybrident à l'ADN. L'hybridation est concentration dépendante, comme la concentration des deux oligonucléotides est beaucoup plus élevée que la concentration de l'ADN, l'hybridation des oligonucléotides aura lieu alors que l'hybridation de l'ADN n'aura pas le temps de se faire. On élève ensuite la température à 72°C, température de fonctionnement de l'ADN polymérase. L'ADN néoformé peut servir de matrice pour un nouveau cycle, si on répète cette alternance de température une quarantaine de fois, la région située entre les deux oligonucléotides présente en une seule copie au début de l'expérience se retrouvera en 2^{40} soit 10^{12} copies à la fin de l'expérience. Comme la



polymérisation a toujours lieu de 5' vers 3', les brins matrices issus de la polymérisation sont bornés en 5' et en 3' par les oligonucléotides (en 3') ou leurs séquences complémentaires (en 5'). On obtiendra donc un fragment d'ADN de longueur égale à la distance les séparant et à laquelle il faut ajouter à taille des deux oligonucléotides qui font partie du produit amplifié. Les fragments non bornés, provenant de la polymérisation sur l'ADN de départ n'ont pas une croissance exponentielle mais linéaire, (2 x 40) et leur quantité sera rapidement négligeable au cours des cycles successifs.

Les composantes du milieu réactionnel

Mg^{2+} : Les cations et plus particulièrement le magnésium influencent l'hybridation des amorces. Les ions magnésium se lient aux groupements chargés négativement (groupements phosphate du squelette) et ainsi diminuent les forces de répulsion entre l'amorce et la matrice. Si la concentration en magnésium est trop faible, il n'y a pas d'hybridation, si la concentration est trop forte, il y a diminution de la spécificité. D'une manière générale le magnésium est utilisé entre 1 et 5 mM. Le Mg^{2+} est indispensable à la reconnaissance des dNTP et de la matrice par l'enzyme.

Amorces : la concentration des amorces influence la spécificité et l'efficacité de l'amplification. Une concentration trop élevée diminue la spécificité et une concentration trop faible diminue l'efficacité de l'amplification. Les amorces sont souvent utilisées aux alentours de 1 μ M.

Solvant organiques : L'addition de solvants organiques miscibles tels que le DMSO (0,8 M), la bétaine (0,3 M) , le polyéthylène glycol, le glycérol ou la formamide est souvent utilisé pour augmenter la spécificité ou le rendement de la réaction principalement pour les régions riches en GC (Bachman et al., 1990). Le plus efficace serait le tetramethylene sulfoxide (0,5 M) (Chakrabarti et Schutt, 2002). Ces solvants organiques agiraient en se liant aux sillons mineur et majeur de l'ADN ce qui déstabiliserait la double hélice (Cheng et al., 1994).

Comment amplifier de grands fragments ?

La Taq DNA polymérase est rapide mais fait des erreurs. Lorsqu'il y a une erreur de synthèse, elle reste bloquée et l'amplification ne peut pas continuer. Le risque d'erreur étant proportionnel à la

longueur du fragment, les grands fragments (> 2 kb) sont difficiles à obtenir. Les ADN polymérases ayant une activité de correction (3'5' exonucléase) sont peu rapides et l'amplification des grands fragments se fait mal. Le problème a été résolu en mélangeant les deux types de polymérases et ainsi des fragments de 10-20 kb peuvent être obtenus (Barnes, 1994).

Les ADN polymérases ayant une activité correctrice sont inhibées par le dUTP qui est produit lors de l'amplification par désamination oxydative du dCTP. Une solution consiste à ajouter une dUTPase au mélange réactionnel. Cette solution est commercialisée sous le nom de Pfu-turbo® par Stratagene.

Comment diminuer les erreurs lors de la polymérisation ?

La technique est d'utiliser une ADN polymérase pourvue d'une activité 3'5' exonucléasique. Toutefois la plupart de ces polymérases sont moins performantes que la Taq DNA polymérase. On effectuera donc un mélange de Taq et d'une polymérase pourvue d'activité correctrice, suivant la concentration respective des deux enzymes, on privilégiera soit l'amplification, soit la fidélité.

Comment éviter les amorçages au hasard ?

Pour que les amplifications soient spécifiques, il faut que l'hybridation des amorces ne se fasse qu'au niveau de la séquence désirée. Pour ce faire, on choisit des amorces assez longues, d'au moins 18 nucléotides et on fait l'hybridation à une température proche du T_m (T_m-5°C). Toutefois même dans ces conditions, on peut observer des amplifications qui ne correspondent pas à la séquence désirée. Elles peuvent provenir d'une hybridation à température ambiante suivie par une polymérisation lors de la préparation du mélange réactionnel.

Pour éviter ces amplifications non spécifiques, il faut donc éviter la polymérisation lors de la préparation de l'échantillon. Plusieurs techniques ont été développées et ont pris le nom de "hot start".

- le plus simple est de faire toutes les manipulations à 4°C (dans la glace) de façon à bloquer la polymérase puis à ne mettre les échantillons dans le thermocycleur que lorsqu'il a atteint la température de dénaturation (95°C).

- On peut aussi ajouter tous les composants de la PCR dans le tube à l'exception de la polymérase, effectuer une première étape de dénaturation, redescendre la température au T_m et ajouter la polymérase.
- Une deuxième technique consiste à incorporer dans le mélange réactionnel un inhibiteur de la polymérase thermolabile. Avant la dénaturation, la polymérase sera inactive, par contre après la première dénaturation, l'inhibiteur sera dénaturé et donc inefficace. Cet inhibiteur est un anticorps, qui bloque l'activité et qui est dénaturé par la chaleur.

Des hybridations non spécifiques peuvent aussi se produire pendant la réaction de PCR. Dans ce cas on augmente la température d'hybridation pour chercher la température qui permet une hybridation efficace des amorces sans hybridation non spécifique. On peut aussi utiliser une protéine commercialisée par Invitrogen qui diminue les hybridations non spécifiques.

Comment obtenir un seul des deux brins ?

Une première méthode consiste à faire une PCR asymétrique, en utilisant une des deux amorces à faible concentration et l'autre à forte concentration. Au début de l'amplification la PCR se fera normalement, les deux brins seront amplifiés. Puis avec l'épuisement d'une des deux amorces, un seul des deux brins sera synthétisé.

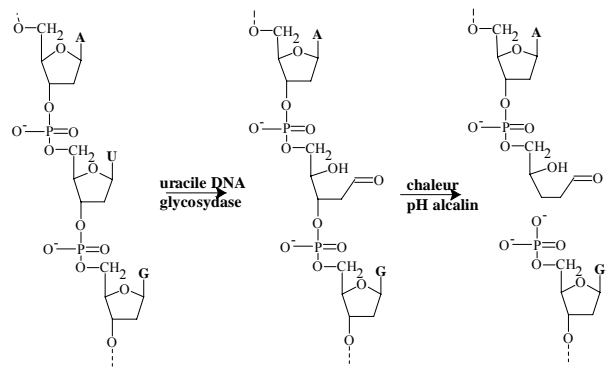
Une deuxième méthode consiste à utiliser une amorce modifiée, comportant un phosphorothioate. Cette modification rend l'ADN résistant à la digestion par une exonucléase. Après l'amplification, le produit est digéré par une exonucléase, le brin portant l'amorce modifiée est protégée.

Comment éviter les pollutions par les produits d'une amplification précédente ?

La réaction de PCR peut amplifier une seule molécule plus d'un milliard de fois, ainsi les contaminants sont facilement amplifiables. Ces contaminants sont souvent des résidus des amplifications précédentes qui sont présents par exemple dans le corps des pipettes

- On peut utiliser des pipettes à déplacement positif ou des pointes à filtre ou pour éviter le passage des molécules du corps de la pipette au mélange réactionnel.
- On peut utiliser des jeux de pipette différents, un jeu pour la préparation des réactions et un jeu pour l'analyse des produits de la réaction.

Une méthode est de substituer le dUTP au dTTP à l'amplification, ce qui produit un ADN qui contient de l'uracile. Les réactions suivantes seront préparées en présence d'uracile-DNA- Glycosydase qui retirera l'uracile des contaminants lors de la préparation des échantillons. L'uracile DNA-glycosydase sera dénaturée lors de la première dénaturation, et les sites apurique-apyrimidique



seront hydrolysés à 95°C en condition alcaline lors cette même première dénaturation.

Utilisation : Cette stratégie peut s'appliquer lorsque la PCR est utilisée pour le diagnostic ou lorsque le produit de la PCR sera séquencé. Si le produit de la PCR doit être cloné, il faudra alors utiliser une souche ung-. Par contre on ne pourra pas l'utiliser si on veut faire des cartes de restriction (certains enzymes sont sensibles à l'uracile) ou si on étudie des interactions ADN/protéines.

Comment détecter les produits d'amplification ?

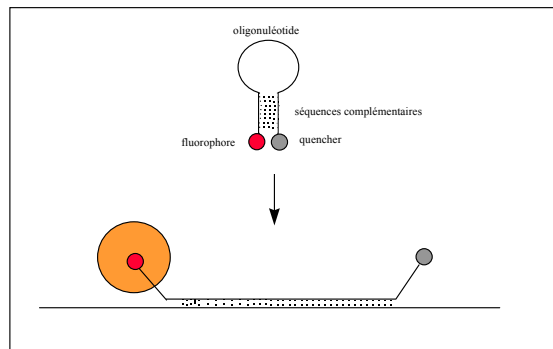
- Electrophorèse

La technique la plus simple est de faire un gel d'électrophorèse révélé au bromure d'éthidium. Cette méthode permet de contrôler la spécificité de l'amplification, de vérifier que le produit de l'amplification correspond bien à la taille attendue.

- Molecular beacons

On peut utiliser des oligonucléotides ayant à leurs extrémités deux séquences complémentaires.

Cet oligonucléotide adopte donc une structure en tige-boucle. A l'extrémité de la tige, il y a un fluorophore et un quencher qui se retrouvent à proximité si bien que les photons émis par le fluorophore sont absorbés par le quencher. En présence d'une séquence cible, la sonde se dénature, et le fluorophore se retrouve éloigné du quencher et la sonde émet de la fluorescence (Vet et al. 1999).



Cette méthode permet de quantifier le produit de la réaction : un fluorimètre est incorporé à l'appareil PCR, il permet de suivre l'amplification au cours de la réaction.

Comment amplifier un fragment d'ADN rare ?

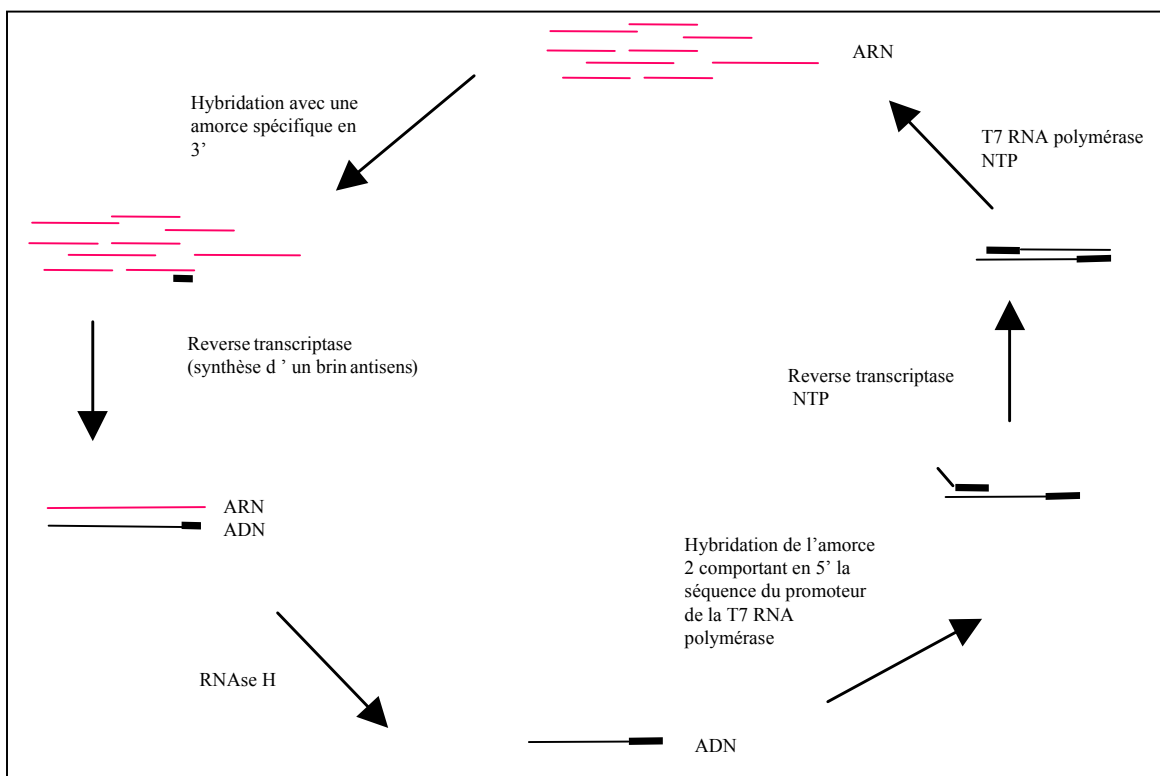
Lorsqu'une séquence est très faiblement représentée dans la réaction de PCR, on n'observe généralement pas de produit d'amplification après dépôt sur un gel d'électrophorèse. On peut alors faire une deuxième PCR avec le produit de la première, mais pour éviter d'amplifier les produits non spécifiques, on utilise des amorces internes, on dit alors qu'on fait des PCR emboîtées ou « nested PCR »

Amplification d'ARN

Le principe de la méthode consiste à synthétiser à partir d'un ARN particulier un fragment d'ADN contenant un promoteur du phage T7 en 5'. Une fois cet ADN obtenu, l'ARN polymérase T7 synthétisera des ARN qui à leur tour serviront de matrice à la synthèse d'ADNc comportant le promoteur. On a donc une amplification en chaîne, les produits de la réaction servant de substrat pour les amplifications suivantes. Cette méthode a pris le nom de NASBA (Nucleic Acid Sequence-Based Amplification) (Compton, 1991).

On ajoute dans un tube

- L'ARN,
- les 4 désoxyribonucléotides
- les 4 ribonucléotides
- deux oligonucléotides flanquant la séquence à amplifier
- Une reverse transcriptase
- Une RNase H
- Une ARN polymérase T7



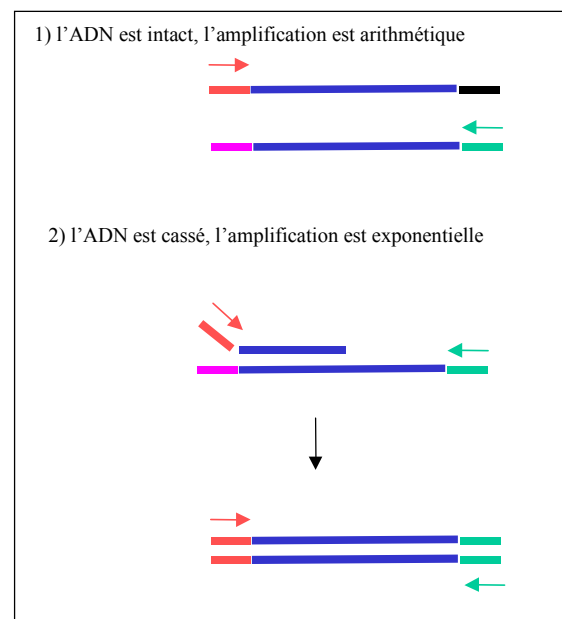
L'oligonucléotide sens comportant en 5' la séquence du promoteur de l'ARN polymérase T7 s'hybride à l'ARN cible, et un premier brin de ADNc est synthétisé par la reverse transcriptase. On obtient donc un hétéroduplex, ADN-ARN substrat de la RNase H qui dégrade l'ARN hybridé. A partir de l'ADNc simple brin sur lequel l'amorce sens peut s'hybrider, la reverse transcriptase synthétise le second brin d'ADN, bien que la matrice ne soit pas de l'ARN. On obtient un ADNc double brin avec un promoteur de la T7 RNA polymérase à une extrémité. L'ARN polymérase T7 synthétise donc un ARN à partir de cette nouvelle matrice.

Avantage de la méthode:

- L'amplification est très rapide puisque la T7 RNA polymérase fabrique de 10 à 100 copies par cycle
- L'ADN n'est pas amplifié puisqu'il n'y a pas d'étape de dénaturation.

La PCR de fusion

La PCR de fusion (Bridging PCR, Jumping PCR) a été découverte lors de l'analyse d'ADN anciens ou abîmés (Paabo et al., 1990). On obtient seulement une amplification arithmétique lorsqu'il y a deux amorces qui hybrident sur deux matrices différentes. Dans le cas où les deux matrices contiennent une région centrale commune (en bleu) bordées par des régions de séquence différentes (en rouge, noir, rose et vert) on a amplification seulement si une des deux matrices est interrompue. Le brin coupé sert d'amorce au premier cycle de PCR permettant la fusion.



Application :

- construction de nouveaux allèles par recombinaison

- détection de cassures de l'ADN ou d'autres facteurs qui empêchent la polymérisation (dans ce cas on utilisera une polymérase sans activité 5'3' exonucléase pour éviter de fabriquer des molécules tronquées)

La mutagenèse

Il y a deux grands types de mutagenèse, la mutagenèse dirigée et la mutagenèse aléatoire.

La mutagenèse dirigée

La mutagenèse dirigée a pour objet de muter une séquence d'ADN à un endroit précis. Un exemple de mutagenèse dirigée simple consiste à éliminer un site de restriction donnant des extrémités cohésives dans un plasmide. On ouvre le plasmide au site de coupure et on ajoute une polymérase (dépourvue d'activité 5'3' exonucléase) et des dNTP. On effectue ensuite une ligation des extrémités devenues franches. Dans cette expérience, on ajoute ou on retire généralement soit 2 soit 4 nucléotides. Le cadre de lecture est donc modifié et doit être pris en compte si l'expérimentation a pour but de produire une protéine.

Lorsqu'on veut changer une ou plusieurs bases dans un plasmide, on peut partir soit d'un ADN simple brin soit d'un ADN double brin.

a) A partir d'un ADN simple brin

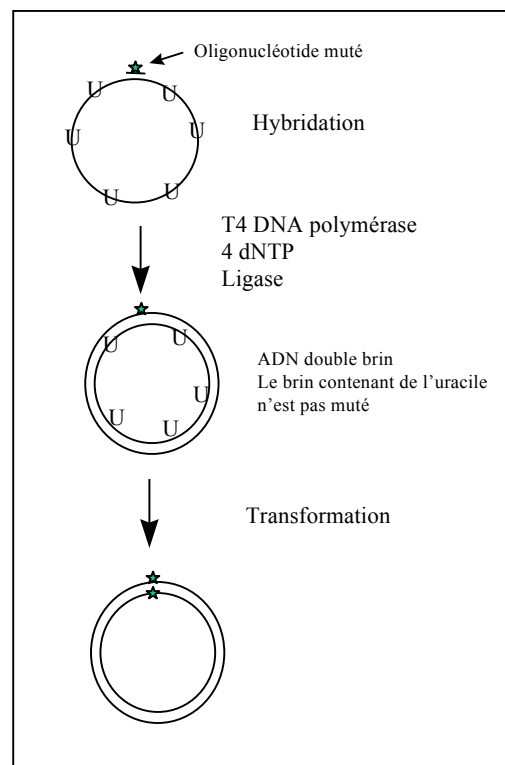
Pour obtenir un ADN simple brin, on peut cloner le gène à muter dans un phage simple brin comme M13. En effet, la phase répliquative étant double brin, on peut la manipuler comme un plasmide et faire agir les enzymes de restriction. Une méthode plus simple consiste à introduire dans le plasmide utilisé, une origine de répllication simple brin. On utilise les origines de répllication des phages M13 ou f1. Pour obtenir un ADN simple brin, on infecte une bactérie comportant ce plasmide avec des phages qui produisent les enzymes nécessaires au développement du phage (répllication, protéines de l'enveloppe...). L'ADN du plasmide se retrouve dans les enveloppes du phage et il peut être purifié comme un virus. De plus, si on utilise comme phage "helper" un mutant déficient pour sa répllication, l'ADN du plasmide sera prédominant dans la préparation du virus.

Technique de base : on hybride un oligonucléotide phosphorylé en 5', comportant la mutation désirée en son centre. La polymérisation du brin manquant amorcée par l'oligonucléotide est effectuée à l'aide d'une T4 DNA polymérase et une ligation entre l'extrémité 5' de l'oligonucléotide et le brin néosynthétisé est éventuellement effectuée avec la T4 DNA ligase. On est donc à ce stade en présence d'un hybride ADN sauvage, ADN muté.

Plusieurs techniques ont été mis au point pour éliminer le brin sauvage :

- Une première technique utilise lors de la polymérisation un analogue soufré d'un nucléotide (Nakamaye et Eckstein, 1986). On effectue ensuite une coupure par Nsi I qui ne coupe pas le brin muté, néosynthétisé et comportant donc l'analogue soufré. Par contre, Nsi I coupe le brin matrice, sauvage qui ne comporte pas d'analogue soufré. Nsi I reconnaissant uniquement 4 bases, le brin sauvage est toujours coupé en plusieurs endroits. L'ADN est alors digéré par une exonucléase pour retirer tout le brin parental et resynthétisé par la T4 DNA polymérase.

- Une deuxième technique consiste à préparer de l'ADN simple brin dans une souche bactérienne *dut*⁻ *ung*⁻ (Kunkel, 1985). Ces souches *dut*⁻ sont déficientes en UTPase et contiennent beaucoup d'UTP qui rentre en compétition avec le TTP. Les souches *ung*⁻ manquent d'uracyl N-glycosylase, enzyme qui normalement enlève l'uracile incorporé dans l'ADN. Ainsi, les souches *dut*⁻, *ung*⁻ incorporent plusieurs uraciles dans l'ADN et donc dans le simple brin. Lors de la synthèse *in vitro*, l'UMP ne gêne pas, il n'est ni mutagène ni inhibiteur. Après hybridation de l'oligonucléotide et polymérisation, on obtient donc un ADN hybride, un brin muté et un brin sauvage contenant de l'uracile. Si cet hybride est injecté dans une souche *ung*⁺, les uracyl N-glycosylases retirent les uraciles, produisant des sites apuriniques. Ces sites bloquent la synthèse d'ADN et seul le brin muté se



réplique. De plus, les sites apuriques sont coupés et la DNA polymérase I répare. Comme cette polymérase a une activité 5'3' exonucléase, on obtient un déplacement de coupure (nick-translation) au cours duquel la mutation est introduite dans le brin matrice.

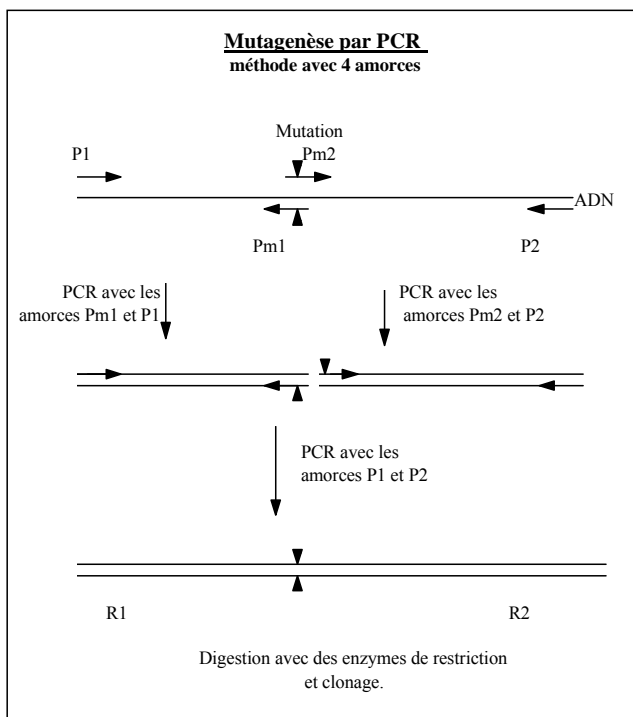
- Une autre technique utilise deux oligonucléotides, un pour la mutation et un pour modifier le gène de résistance à l'antibiotique lui donnant ainsi une nouvelle spécificité, une résistance à un nouvel antibiotique. Après hybridation avec les deux oligonucléotides et remplissage, la molécule hybride est directement utilisée pour une transformation. Après la première répllication, on a deux catégories de plasmides, sauvage et muté. Si on prépare les plasmides et qu'on effectue une nouvelle transformation seuls les plasmides mutés pourront se développer en présence de l'antibiotique.

Pour l'ensemble de ces techniques on utilisera *in vitro* une ADN polymérase n'ayant pas d'activité 5'3' exonucléase. En effet cette activité occasionne un "remplacement de brin" par digestion du brin néosynthétisé en commençant par l'oligonucléotide. Un tel remplacement résulte en la perte de la mutation introduite au départ. Toutefois, même si on utilise une polymérase dépourvue d'activité 5'3' exonucléase, on observe le remplacement de brin. On le détecte en utilisant des oligonucléotides porteurs de deux mutations et après la transformation on observe des colonies qui ont le plasmide muté avec les deux mutations, des colonies qui ont un plasmide sauvage, et des colonies qui ont un plasmide porteur d'une seule mutation. Dans ce dernier cas, seule la mutation en 5' est absente ou autrement dit, on ne trouve pas de plasmide avec uniquement la mutation en 5'. Il y a donc eut digestion de l'amorce de 5' vers 3' et non l'inverse. Ce phénomène peut paraître paradoxal si on a utilisé une polymérase dépourvue d'activité 5'3' exonucléase. Toutefois, les problèmes rencontrés dans ce type d'expérience ne s'arrêtent pas là, un nombre élevé de plasmide peut ne pas se liguer soit parce que la ligase n'a pas été efficace soit parce que la phosphorylation de l'oligonucléotide ne l'a pas été. La réparation du nick se fait *in vivo*, dans la bactérie par la ADN polymérase I, qui a une activité 5'3' exonucléase et qui effectue une nick translation.

b) A partir d'un ADN double brin

Utilisation des techniques de PCR en présence d'un oligonucléotide muté.

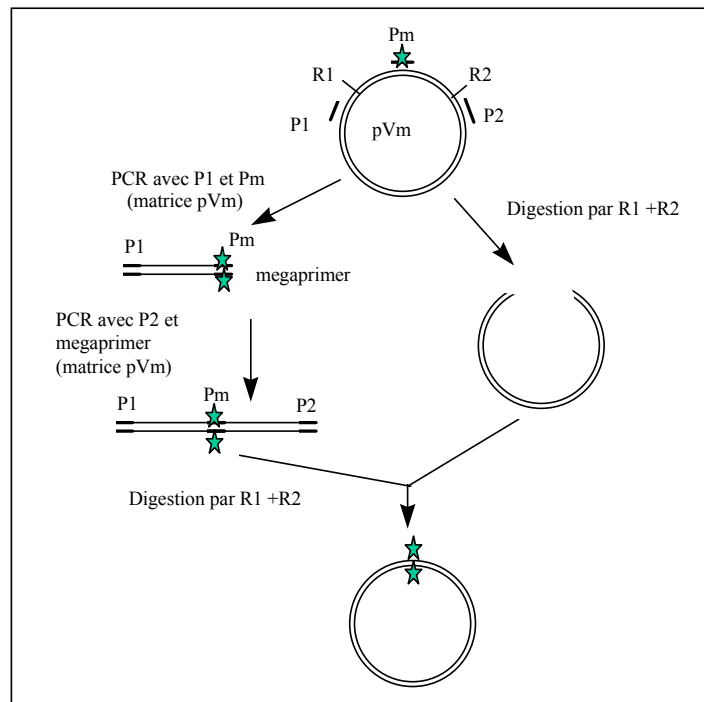
Il existe de nombreuses méthodes qui ont toutes le même principe, la mutation désirée est introduite dans une des amorces.

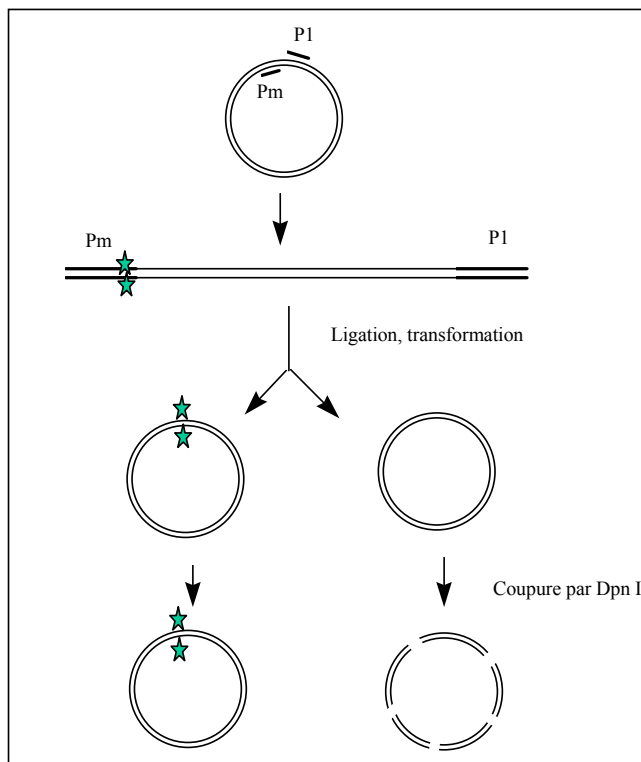


Une méthode utilise quatre oligonucléotides. On fait deux amplifications séparées chacune avec un oligonucléotide externe et un oligonucléotide portant la mutation. Les deux fragments obtenus séparément sont ensuite amplifiés ensemble avec les deux amorces les plus externes. On digère ensuite les deux extrémités par des enzymes de restriction et on ligue dans le plasmide original.

La méthode du "megaprimer". Cette méthode utilise le résultat de la première amplification comme amorce pour la deuxième amplification (Sarkar et Sommer, 1990; Chen et Przybyla, 1994). Une première amplification est faite avec deux amorces (P1 + Pm), dont une comporte la mutation désirée (Pm). Le produit de PCR est purifié et utilisé directement comme amorce appelée "megaprimer" pour une deuxième réaction de PCR utilisant une troisième amorce (P2). Comme

précédemment, le produit de la deuxième amplification est digéré par une enzyme de restriction et introduit dans un plasmide.





Dans les méthodes décrites jusqu'à présent, une construction doit être faite à la fin pour introduire le fragment muté dans le plasmide ce qui implique une digestion par des enzymes de restriction et une ligation. Une alternative consiste à utiliser deux oligonucléotides adjacents de sens opposés et phosphorylés (Fisher et Guo, 1997). Un des deux oligonucléotides (Pm) porte la mutation. L'ensemble du plasmide sert ici de matrice. Le produit est circularisé par ligation et transformé. La matrice sauvage doit être éliminée avant la transformation, en effet cet ADN circulaire est en grande quantité et donnera un grand nombre de transformants, comparé à un produit de

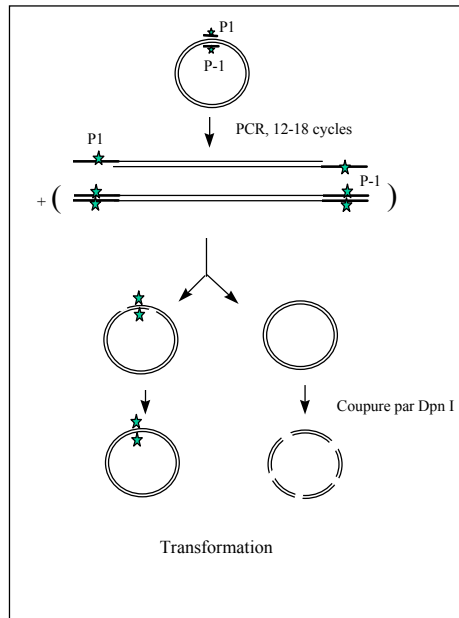
ligation.

Une technique utilisable pour éliminer le brin matrice consiste à utiliser une enzyme sensible à la méthylation de l'ADN comme Dpn I, cet enzyme coupe G^m6ATC et ne coupe pas la séquence GATC non méthylée. Si l'ADN a été préparé dans une bactérie qui effectue cette méthylation (Dam^+), une digestion par Dpn I coupera l'ADN sauvage, mais non l'ADN néosynthétisé. Lors de la transformation, les coupures seront réparées comme précédemment et le déplacement de brin introduira la mutation dans le brin parental.

Une autre possibilité est de méthyler le plasmide parental avant la polymérisation puis de transformer le produit de la mutagenèse dans une souche $mcrBC^+$. Ces souches ont l'endonucléase *mrcB* qui enlève le brin méthylé (kit commercialisé par Invitrogen sous le nom de GeneTailor[®]).

Un des inconvénients de la technique de PCR en chaîne est d'accumuler les mutations au cours des cycles successifs. Une solution pour remédier à ce problème est d'utiliser deux oligonucléotides complémentaires pour la réaction de PCR. Dans ce cas seul le brin original sert de matrice (et il ne

s'agit plus de PCR à proprement parler). Ici les brins néoformés ne peuvent pas servir de matrice à de nouvelles amplifications.



Dans cette technique on utilise deux oligonucléotides complémentaires. On obtient des molécules ayant des extrémités cohésives en utilisant uniquement le vecteur d'origine comme matrice. Ces molécules ne peuvent pas servir de matrice aux amplifications suivantes, il ne s'agit donc pas à proprement parler d'une PCR. Comme les molécules ont des extrémités cohésives, on peut avoir circularisation.

L'élimination de l'ADN parent peut se faire par deux techniques. On peut digérer le produit de la polymérisation par l'enzyme de restriction DpnI, seul l'ADN parental méthylé s'il a été purifié à partir d'une souche Dam^+ sera coupé et les nicks seront réparés *in vivo* par la DNA polymérase I après

transformation. Son activité 5'3' exonucléase permettra d'introduire la mutation dans le brin parental suivant le principe de la nick translation.

Après quelques cycles, la population de molécules synthétisées augmente et on peut avoir remplissage des extrémités cohésives par la polymérase. Mais comme il n'y a pas d'étape de ligation, ces molécules ne donnent pas de plasmide.

A l'origine de cette méthode on utilisait deux oligonucléotides complémentaires mais on s'est aperçu que l'utilisation d'une seule amorce était suffisante. On se retrouve alors avec une mutagenèse qui ressemble à la mutagenèse simple brin (Makarova *et al.*, 2000).

Dans ces méthodes, il est préférable d'utiliser une ADN polymérase thermostable qui n'a pas d'activité terminale transférase du fait de la présence d'une activité 3'5' exonucléase comme par exemple la *pfu* DNA polymérase. Il est aussi préférable d'utiliser une polymérase qui n'a pas d'activité 5'3' exonucléase pour éviter les déplacements d'amorces.

Cette méthode est actuellement la plus utilisée car elle est la plus simple à mettre en œuvre.

Le transfert d'un fragment muté d'un plasmide vers un autre plasmide

Lorsqu'on veut transférer plusieurs mutations d'un plasmide vers un autre on peut utiliser les sites de restriction, coupure et ligation. Mais les sites ne sont pas toujours disponibles. On peut faire les mutagenèses dirigées une à une, mais s'il y a plusieurs mutations, l'expérience sera longue. Une méthode dérivée de la mutagenèse utilisant deux amorces complémentaires a été développée (West et Wilson, 2002). Un petit fragment d'ADN de moins de 300 pb contenant la partie à transférer est généré par PCR. Ce fragment est utilisé comme amorce pour la mutagenèse dirigée sur le plasmide receveur.

Que modifie t-on par mutagenèse dirigée ?

L'oligonucléotide peut présenter la substitution d'une base, mais il peut aussi présenter une insertion ou une délétion. La longueur de l'insertion (ou de la délétion) n'a pas d'importance, il faudra simplement utiliser un oligonucléotide assez grand de telle sorte que les deux extrémités puissent s'hybrider sur la matrice. Etant donné qu'on peut synthétiser des oligonucléotides jusqu'à 120 pb, et qu'il faut entre 15 et 20 bases pour obtenir une hybridation spécifique ($n^4 >$ grandeur du génome) une insertion sera actuellement limitée à environ 80 bases. Par contre une délétion ne sera pas limitée. Seules les techniques utilisant l'élimination des simples brins par filtration sur nitrocellulose seront évitées, l'hybride étant en effet partiellement simple brin au niveau de l'hybridation.

Comment détecter les colonies portant les mutations des colonies portant un insert non muté ?

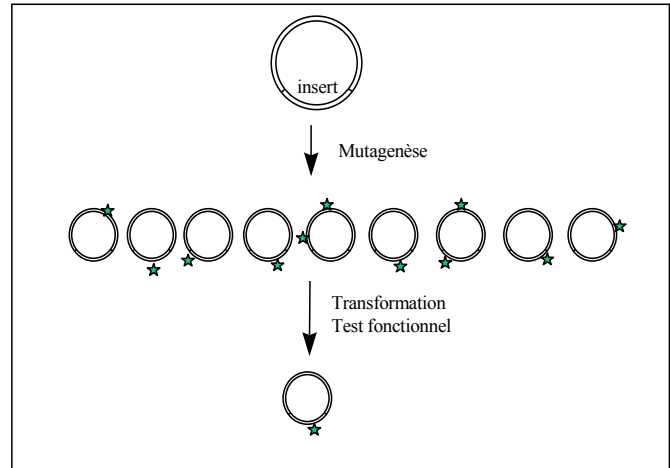
Quelle que soit la méthode, après la transformation on a deux sortes de colonies, celles qui portent la mutation et celles qui correspondent à la matrice. On peut trier les colonies en préparant l'ADN des plasmides et en séquençant la région mutée. Toutefois afin de faire un minimum de séquence on préfère souvent faire un premier tri en utilisant les méthodes de détection des mutations (voir chapitre détection des mutations).

La mutagenèse aléatoire

La mutagenèse aléatoire génère des mutations n'importe où dans l'ADN. Les mutants sont ensuite triés.

On peut générer des mutants *in vivo* ou *in vitro*. Dans le premier cas, les bactéries sont soumises directement au traitement. Dans le deuxième cas, le plasmide portant le gène est purifié, modifié puis réintroduit dans une nouvelle bactérie.

On peut utiliser des produits qui modifient les bases appelées des agents mutagènes ou utiliser des enzymes.

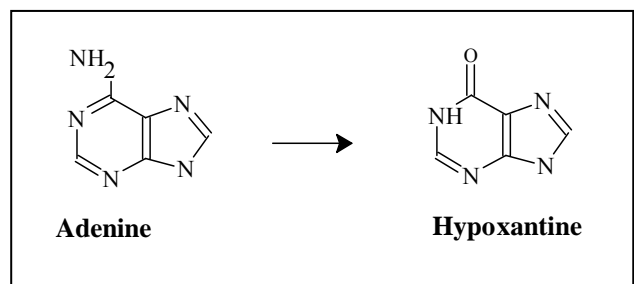


a) Utilisation d'agents mutagènes

De nombreuses molécules abîment l'ADN et peuvent être utilisées pour effectuer de la mutagenèse.

- Désamination des bases

La désamination est une des modifications les plus courantes de l'ADN. Les groupements amines de l'adénine, de la cytosine et de la guanine peuvent être retirés soit spontanément soit à l'aide de nombreux produits chimiques.



Lorsque l'adénine est désaminée, on obtient l'hypoxanthine, lorsque la guanine est désaminée on obtient la xantine, et lorsque la cytosine est désaminée on obtient l'uracile.

La désamination des bases entraîne des mutations car elle entraîne simultanément des mésappariements. Par exemple, l'hypoxanthine, provenant de l'adénosine s'apparie avec la cytosine au

lieu de la thymine. De même, l'uracile provenant de la désamination de la cytosine s'apparie avec l'adénine au lieu de la guanine.

Si la base reste modifiée jusqu'à la réplication, il y a mutation. Par exemple, si une adénine reste désaminée durant la réplication, il y aura un C sur le brin néosynthétisé à la place du T. Dans les réplications suivantes, on aura un G à la place du A et en fait, on aura une transition AT vers GC.

Plusieurs molécules peuvent être utilisées pour désaminer les bases de l'ADN :

- l'hydroxylamine désamine la cytosine et donc est responsable de transition GC vers AT. Comme l'hydroxylamine ne rentre pas dans les cellules, cette molécule ne peut être utilisée qu'*in vitro*.
- le bisulfite de sodium désamine les cytosines mais l'ADN doit être simple brin.
- l'acide nitreux désamine la cytosine, l'adénine et la guanine. Il peut causer des transitions AT vers GC comme des transitions GC vers AT. De plus il est peu spécifique et peut aussi causer des délétions.

Ces mutations ne s'effectuent pas complètement au hasard et on peut localiser des points chauds. La plupart des mutations sont réparées : les bases modifiées sont repérées par les enzymes de réparation de l'ADN, tout d'abord une glycosilase coupe la liaison sucre-base puis une AP-endonucléase coupe les liaisons phosphodiester, une ADN polymérase comble les trous. Cette réparation est due au fait que les bases modifiées n'existant pas dans l'ADN, leur présence entraîne un mésappariement qui est reconnu par les enzymes de réparation.

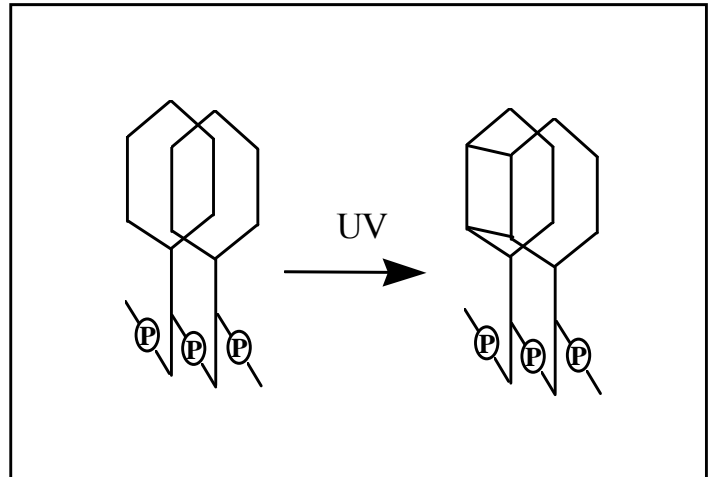
Il existe toutefois des exceptions, dans l'ADN certaines bases sont méthylées pour ne pas être coupées par les enzymes de restriction ou pour réguler l'expression génique. Parmi les méthylations, on peut trouver la 5 méthyl-cytosine ou la désamination des 5 méthyl cytosines donne une thymine qui n'est pas considérée comme une base modifiée et donc n'est pas réparée.

- Alkylation

Les agents alkylants ajoutent des groupes alkyls (CH_3 , $\text{CH}_3\text{CH}_2\dots$) aux bases. On peut citer l'éthyle méthane sulfonate (EMS) ou la nitrosoguanine. Les atomes les plus réactifs sont le N^7 de la guanine, le N^3 de l'adénine ou le O^6 de la guanine. Cette alkylation n'est souvent pas reconnue par les systèmes de réparation et entraîne une différence d'appariement.

- Irradiation UV

L'ADN absorbe à 260 nm du fait des doubles liaisons conjuguées présentes sur les bases. Les photons absorbés augmentent l'énergie des bases et les doubles liaisons peuvent réagir avec d'autres atomes présents à proximité. L'altération la plus souvent rencontrée est la formation de dimère de pyrimidine entre deux bases consécutives.



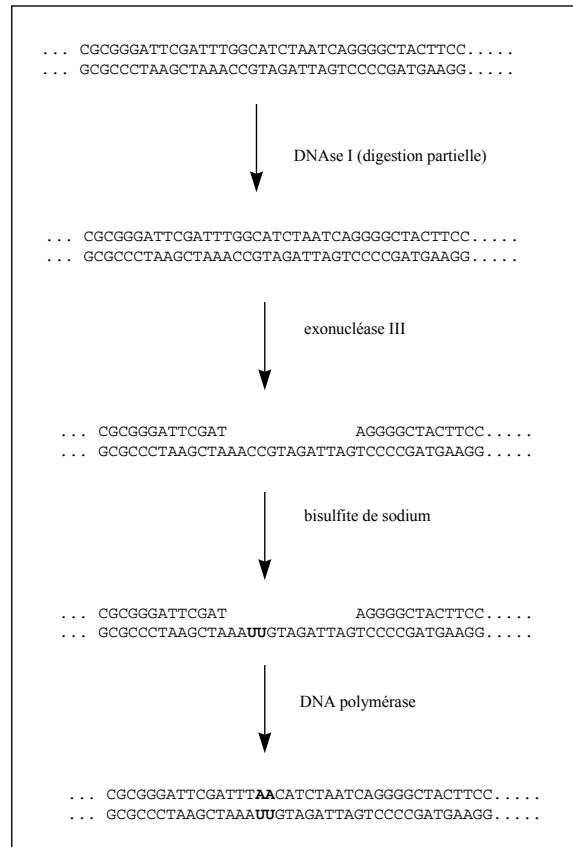
Les modifications des bases sont réparées et il n'y a généralement pas de mutation. Cependant, si la réplication a lieu avant la réparation, la base modifiée peut servir de matrice à une base différente et ainsi occasionner une mutation.

b) Utilisation d'enzymes

Les méthodes historiques :

- On a tout d'abord utilisé une polymérase en présence d'analogues de nucléotides. On fait des nicks sur l'ADN par exemple avec la DNase I de *E. coli*. On effectue ensuite une "nick translation" avec la DNA polymérase I de *E. coli*, les quatre dNTP et du N⁴-hydroxy dCTP. L'analogue se fixe à la place du TMP et on obtient donc des mutations de T en C (Müller *et al.*, 1978). On obtient ici un brin muté et un brin sauvage. Ces deux brins peuvent être triés facilement par transformation, purification des plasmides suivi d'une deuxième transformation.

Une deuxième méthode peut être utilisée, le nick est agrandi à l'aide d'une exonucléase ce qui libère des simples brins. Il y a ensuite traitement par le bisulfite de sodium qui converti les cytosines en uracile puis synthèse du brin manquant à l'aide d'une polymérase (Shortle et Nathans, 1978).

*Les méthodes actuellement utilisées :*« error prone PCR » (Vartamian et al. 1996)

On peut utiliser les erreurs faites par une polymérase dépourvue d'activité 3'5' exonucléase encore appelée activité de correction. L'ADN polymérase Taq est une de ces enzymes. On clone ensuite le fragment de PCR dans un plasmide. Deux méthodes peuvent être utilisées pour augmenter le nombre de mutations :

- 1. utilisation d'un milieu réactionnel favorable aux erreurs tels que la forte concentration en ions manganèse (500 μ M)
- 2. Utilisation d'un biais dans la concentration en dNTP. On utilise dans ce dernier cas la relativement bonne stabilité du mésappariement G:T, et on met dans la solution une forte

concentration en dTTP (1 mM) et une faible concentration en dCTP (1 μ M). L'inconvénient de cette deuxième méthode est que l'efficacité de la réaction de PCR est diminuée ce qui rend difficile le clonage des produits de la mutagenèse.

Ces mutagenèses ne peuvent se faire que si on dispose d'un crible facile à mettre en œuvre pour cribler la banque et trouver les mutations intéressantes. Par exemple, si on cherche une meilleure résistance à un antibiotique ou si on regarde une meilleure métabolisation d'un substrat dont le produit est chromogène.

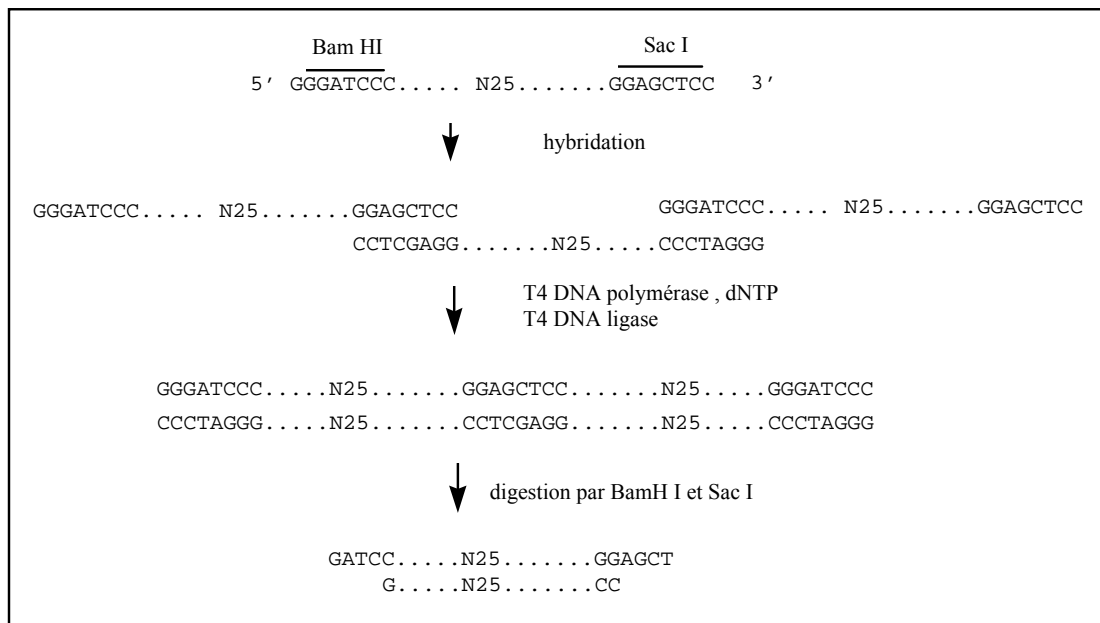
Utilisation d'analogues de base

Une autre solution est d'effectuer une amplification par PCR à l'aide d'analogues de nucléotides tel que le 5'-triphosphate de 6-(2-déoxy- β -D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one qui est efficacement incorporé lors de la polymérisation soit à la place du TTP soit à la place du dCTP. Cette méthode permet de contrôler le nombre de mutations en contrôlant la proportion d'analogues (Zaccolo et al., 1996).

La mutagenèse semi-aléatoire

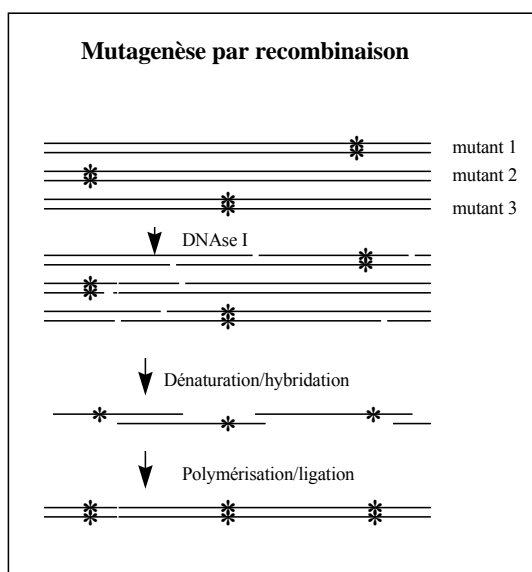
Dans cette mutagenèse, on remplace une partie de la séquence par une séquence au hasard. On obtient une banque de clones que l'on trie pour leurs fonctions.

Une première méthode consiste à insérer un fragment entre deux sites de restriction. On synthétise un oligonucléotide avec des sites de restriction à chaque extrémité. Ces oligonucléotides simple brin sont convertis en doubles brins à l'aide d'une polymérase après hybridation des extrémités (Oliphant et Struhl, 1988; 1989)



Une deuxième méthode consiste à effectuer une mutagenèse dirigée mais au lieu d'utiliser un seul oligonucléotide, on utilise un mélange d'oligonucléotides (encore appelé oligonucléotide dégénéré). Dans ce cas, il faut utiliser une méthode de mutagenèse dirigée mais comme il faut que les deux séquences soient identiques on ne peut pas utiliser les méthodes de mutagenèse sur double brin. Par contre, les méthodes de mutagenèse sur simple brin sont efficaces.

La mutagenèse par recombinaison



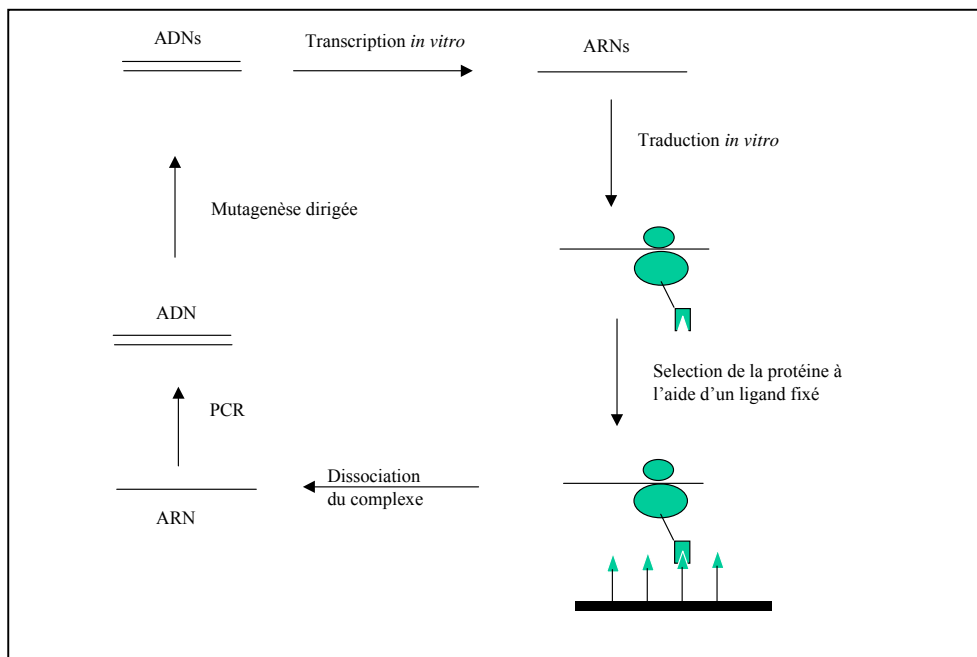
Lorsqu'on dispose de plusieurs mutants (ou d'une librairie de mutants), il peut être intéressant de recombiner ces mutations dans un même gène ou de rechercher une combinaison de mutations. Pour ce faire, on isole les fragments d'ADN, on les coupe par la DNase 1, puis on les réunit au hasard par une série de dénaturation/hybridation et polymérisation. On fait ici une PCR sans amorce. Les gènes complets sont ensuite récupérés en faisant une PCR en utilisant deux amorces hybridant à chaque extrémité.

Le criblage

Le criblage d'une mutagenèse aléatoire se fait généralement après transformation dans *E. coli*, expression des protéines et criblage des clones.

Une méthode permet de s'affranchir de la transformation et du criblage des clones, le ribosome display (Hanes et Plückthun, 1997).

La banque d'ADN provenant de la mutagenèse aléatoire est amplifiée par PCR avec un oligonucléotide portant le promoteur de la T7 RNA polymérase. La banque est ensuite transcrite *in vitro*. Après purification, les ARN sont traduits *in vitro* à l'aide du système S30 de *E. coli* en présence de facteurs augmentant la stabilité du complexe ribosome-ARN-protéine. La traduction est arrêtée en refroidissant la réaction et en augmentant la concentration en ion magnésium. Le complexe désiré est sélectionné par affinité sur le ligand immobilisé de la protéine recherchée. Le complexe lié est dissocié en ajoutant de l'EDTA, l'ARN est rétrotranscrit en ADNc et amplifié par PCR pour pouvoir subir une nouvelle mutagenèse dirigée.



Clonage de gènes eucaryotes

Comment peut-on cloner un gène ? Comment isoler la séquence correspondant à un ARN sur les 10 000 ARNs transcrits par une cellule ? Comment isoler un fragment d'ADN de quelques kilobases sur un génome qui en contient plus d'un million ? Comment disposer du fragment isolé en grande quantité pour pouvoir l'étudier ?

Les stratégies pouvant être mises en œuvre dépendent principalement des informations de départ. Comme ces informations sont très variables en fonction des problématiques abordées, il existe un grand nombre de stratégies différentes.

1) On a un anticorps dirigé contre la protéine d'intérêt

Si on a la protéine, on peut avoir un anticorps dirigé contre cette protéine, simplement en l'injectant à un lapin et en récupérant son sérum un à trois mois après. On obtient ainsi un anticorps polyclonal, c'est à dire un mélange d'anticorps dirigés contre un ensemble d'épitopes de la protéine. Dans le sérum, il y a aussi d'autres anticorps fabriqués par le lapin avant l'injection, ces anticorps peuvent aussi réagir contre d'autres protéines, aussi, souvent on récupère un sérum avant d'injecter la protéine au lapin. Ce sérum pré-immun est alors testé sur un extrait protéique pour vérifier l'absence de réactions non désirées.

Pour vérifier la spécificité de l'anticorps on peut faire un «western blot». On fait migrer un extrait protéique dans un gel dénaturant, on transfère les protéines sur une feuille de nitrocellulose ou de nylon à l'aide d'un courant électrique, on sature la feuille par des protéines puis on l'incube avec l'anticorps. Après lavage de la feuille pour éliminer les anticorps non fixés, on révèle la présence de l'anticorps à l'aide d'un anticorps secondaire pourvu d'un marqueur. Si on ne voit qu'une seule bande, l'anticorps est spécifique par contre si on en voit une multitude, il ne l'est pas. Dans ce dernier cas, il est toujours possible de purifier les anticorps spécifiques de la protéine. Pour ce faire, on fixe la protéine sur un substrat solide qui peut être de la nitrocellulose ou des billes d'agarose. On fait passer le sérum sur la protéine, seul les anticorps spécifiques se lient. Après lavage, on récupère les anticorps en dissociant la liaison protéine-anticorps le plus souvent en abaissant le pH.

Il n'est pas possible d'isoler un ARN particulier et de lui faire exprimer une protéine pour voir si l'anticorps réagit. Il va donc falloir exprimer l'ensemble des protéines correspondant à une population d'ARN par des bactéries de telle façon que chaque colonie bactérienne ne fabrique qu'un seul type de protéine recombinante. On va donc fabriquer une banque d'expression à partir d'ADNc.

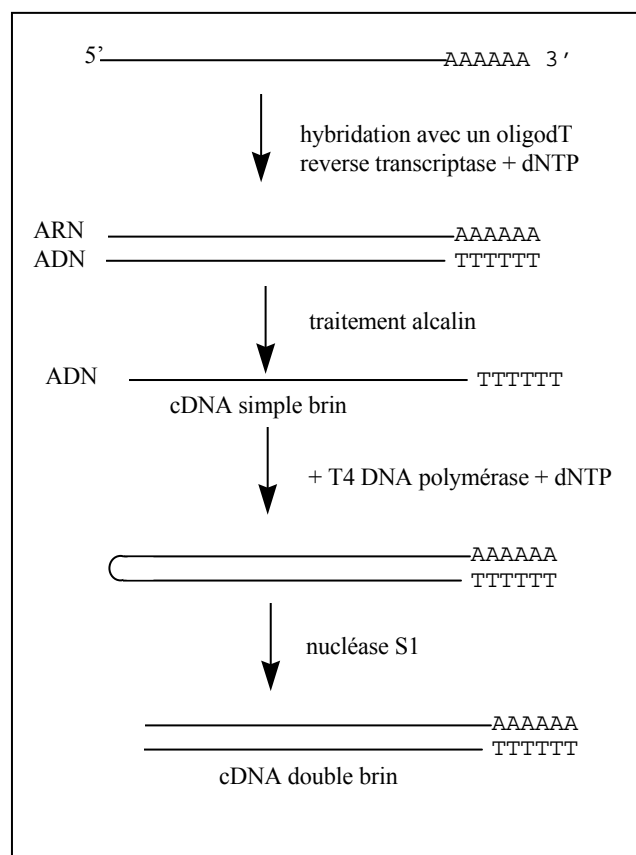
Fabrication d'une banque de d'ADNc

On a à notre disposition une préparation d'ARN messagers d'une cellule qu'il va falloir transformer en ADN. Ils seront alors appelés ADN complémentaires ou ADNc. Comme on a une population d'ARN, on obtiendra une population d'ADNc.

On distingue deux grandes techniques pour fabriquer un ADNc

Technique à la nucléase S1

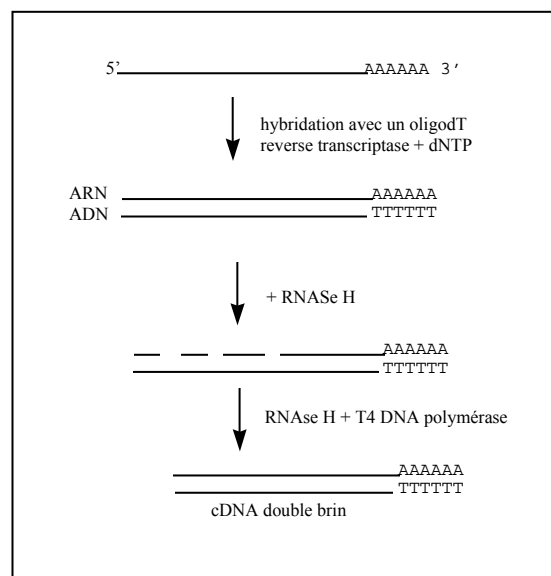
Pour la synthèse du premier brin, on utilise une reverse transcriptase (ADN polymérase ARN dépendante). Comme pour toutes les ADN polymérase, on a besoin d'une amorce pour initier la polymérisation. On utilise un oligonucléotide qui s'hybridera sur tous les ARN messagers, un oligodT qui s'hybridera sur toutes les queues polyA. On détruit l'ARN par un traitement alcalin. Lorsque l'extrémité 5' de l'ADN se retourne et s'hybride sur lui-même en faisant une boucle en épingle à cheveux, on a une extrémité double brin qui peut servir d'amorce à une ADN polymérase (ADN dépendante). On fait donc agir



cette enzyme. On obtient alors un ADNc double brin avec une extrémité en épingle à cheveux. Une telle structure ne peut pas être intégrée dans un vecteur. Pour avoir une extrémité franche on fait agir une nucléase dégradant le simple brin, la nucléase S1 (l'ADN polymérase n'est pas indispensable, en effet la reverse transcriptase peut synthétiser de l'ADN à partir d'une amorce ADN, toutefois cette activité n'est souvent pas suffisante).

Technique de Gubler et Hoffman

Ici on synthétise un premier brin d'ADNc comme précédemment avec une reverse transcriptase après avoir hybridé une amorce. Mais l'ARN est éliminé par digestion ménagée à la RNase H, les fragments restants servent alors d'amorces pour une ADN polymérase qui comble les trous au fur et à mesure que la RNase H continue son action. Les fragments d'ADN sont alors reliés grâce à l'action d'une ADN ligase. Noter que les bouts sont ici francs, grâce à l'action 3'5' exonucléase de la polymérase.



Avec les deux techniques, on n'a pas de copies des extrémités 5' des ARNs. D'autres techniques permettront de les obtenir. Ce n'est souvent pas grave au départ si on s'intéresse uniquement à la partie codante, la partie 5' située en amont de l'ATG est dans ce cas moins importante.

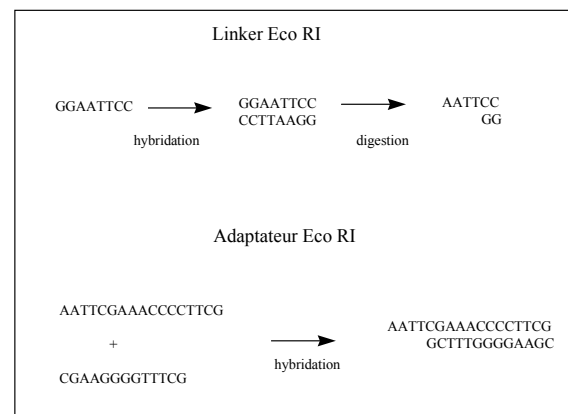
Pour les deux techniques, on a comme alternative d'utiliser des amorces au hasard à la place de l'oligo(dT). Cette méthode se rapproche de la méthode de marquage par hybridation aléatoire ("random priming"), les oligonucléotides se fixent au hasard sur l'ARN et l'ADNc résultant correspond à uniquement une partie du gène, on perd donc également l'extrémité 3'. Cette méthode est principalement utilisée lorsqu'on attend un ADNc long, en effet la reverse transcriptase n'est souvent pas assez active pour synthétiser plusieurs kilobases. Les ADNc amorcés par l'oligo dT, n'étant pas

assez long, la partie 5' du gène manque. Ce manque peut être comblé si l'amorce est moins éloignée de l'extrémité 5'.

Il faut maintenant intégrer l'ADNc double brin dans un vecteur de façon à pouvoir l'amplifier, et à obtenir des clones.

On peut liguer directement la population de ADNc à un vecteur coupé par une enzyme de restriction générant des extrémités franches. Le vecteur est alors déphosphorylé en utilisant une phosphatase alcaline pour éviter sa recircularisation et donc pour éviter la présence d'un grand nombre de clones non recombinants. Toutefois, cette ligation est relativement inefficace et on préfère faire une ligation avec une extrémité cohésive. Il faudra donc ajouter des extrémités cohésives aux deux bouts.

Une première technique pour ajouter des extrémités cohésives consiste à liguer des linkers aux extrémités des ADNc. Les linkers sont des oligonucléotides palindromiques, qui s'hybrident en solution. Ils comportent une séquence correspondant à un site de restriction. Ils sont phosphorylés à l'aide de la kinase du phage T4 puis ligués aux extrémités des ADNc. Une digestion par l'enzyme de restriction reconnaissant la séquence présente sur le linker permet alors de générer des extrémités cohésives. Le site de restriction est présent sur certains ADNc, ils seront alors coupés par l'enzyme. Pour



éviter cette coupure interne, on méthyle les sites de restriction internes avant l'étape de ligation. Pour chaque enzyme de restriction, on dispose d'une méthylase qui reconnaît le même site et qui ainsi protège le site. On a donc les étapes suivantes : méthylation des ADNc, ligation des linkers phosphorylés, digestion par une enzyme de restriction et ligation des ADNc avec un vecteur linéarisé et déphosphorylé pour éviter la recircularisation du vecteur.

La méthylation est souvent partielle, ce qui amène à cloner des fragments d'ADNc. Pour éviter cette étape, on peut utiliser des adaptateurs. Il s'agit d'un couple d'oligonucléotides partiellement complémentaires et générant une extrémité franche qui sera compatible avec l'extrémité de l'ADNc et une extrémité cohésive qui sera compatible avec les extrémités du vecteur. Un seul des deux oligonucléotides sera phosphorylé, si bien que lors de la ligation des adaptateurs avec les ADNc, un

seul adaptateur se liguera à chaque extrémité. Après élimination des adaptateurs surnuméraires par filtration sur gel, les ADNc seront phosphorylés.

On utilise généralement comme vecteur un phage λ pour la facilité de criblage, pour la possibilité de faire un empaquetage *in vitro* souvent plus efficace que la transformation de plasmide mais, comme l'électroporation donne des rendements proches de l'empaquetage *in vitro*, certaines banques sont également construites avec des plasmides.

A cette étape, on est en présence d'une banque d'ADNc, c'est à dire d'un ensemble de bactéries ou de phages recombinants, chacun portant un insert différent, représentant l'ensemble de la population d'ARN de départ. La banque sera donc représentative uniquement du matériel de départ. Comme la population d'ARN dépend de la transcription, chaque banque sera représentative du type de cellule utilisé.

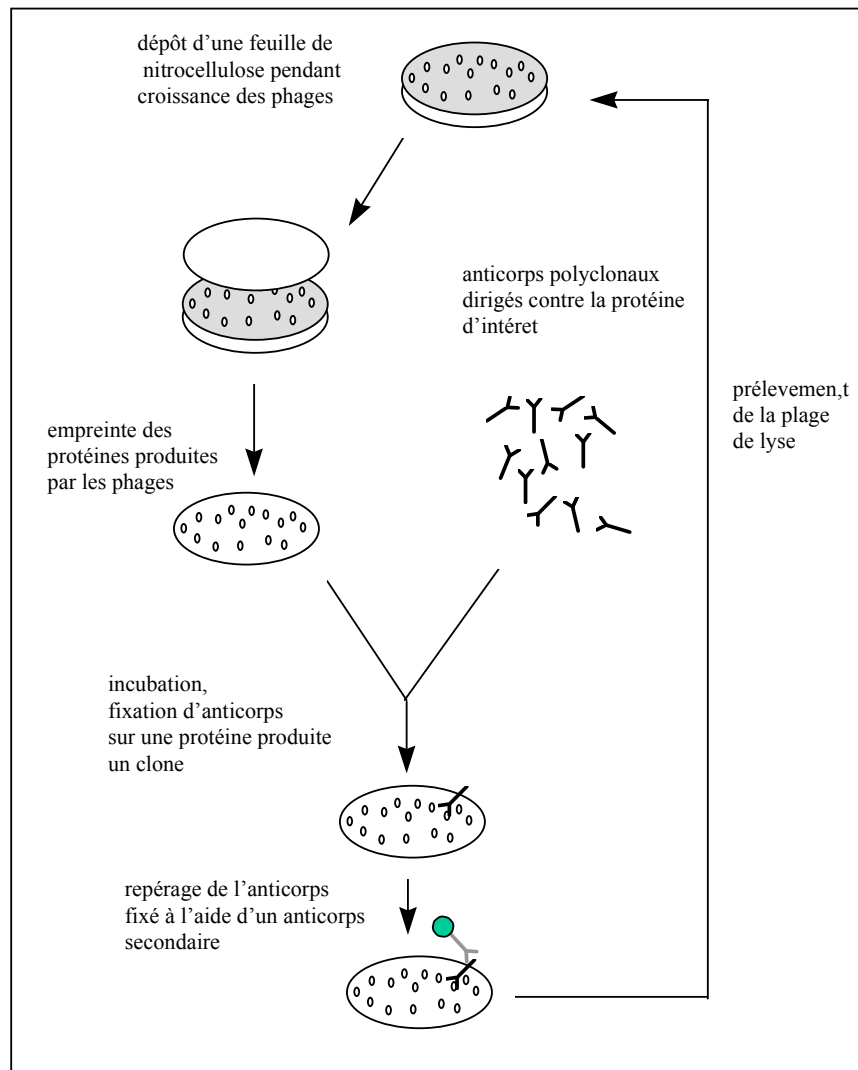
Les ARN les plus fréquents seront les plus représentés et à l'inverse les ARN les plus rares seront les moins représentés dans la banque. Dans certains cas on désire diminuer cette différence de représentativité, on réalise une banque dite normalisée. On hybride les ARN messagers à de l'ADN génomique fixé sur une membrane. Pour chaque gène, un seul ARN peut s'hybrider, les ARN non hybridés sont éliminés par lavage de la membrane. Après dénaturation, les ARN messagers sont récupérés et une banque d'ADNc est construite.

Le criblage est effectué à l'aide d'un anticorps réagissant avec la protéine provenant de la transcription de l'insert. Il faut donc qu'il y ait expression des protéines dans *E. coli*, la banque d'ADNc doit être faite dans un vecteur permettant la fusion avec une protéine bactérienne. En effet, pour que la protéine soit produite, qu'il y ait un ATG précédé d'une séquence de liaison au ribosome (RBS) or, du fait des différences de traduction entre les procaryotes et les eucaryotes, cette séquence est absente dans les gènes eucaryotes. La fusion la plus utilisée comprend le peptide α de la β -galactosidase. On obtiendra donc une protéine de fusion formée en N terminal du peptide α suivie des protéines provenant des inserts.

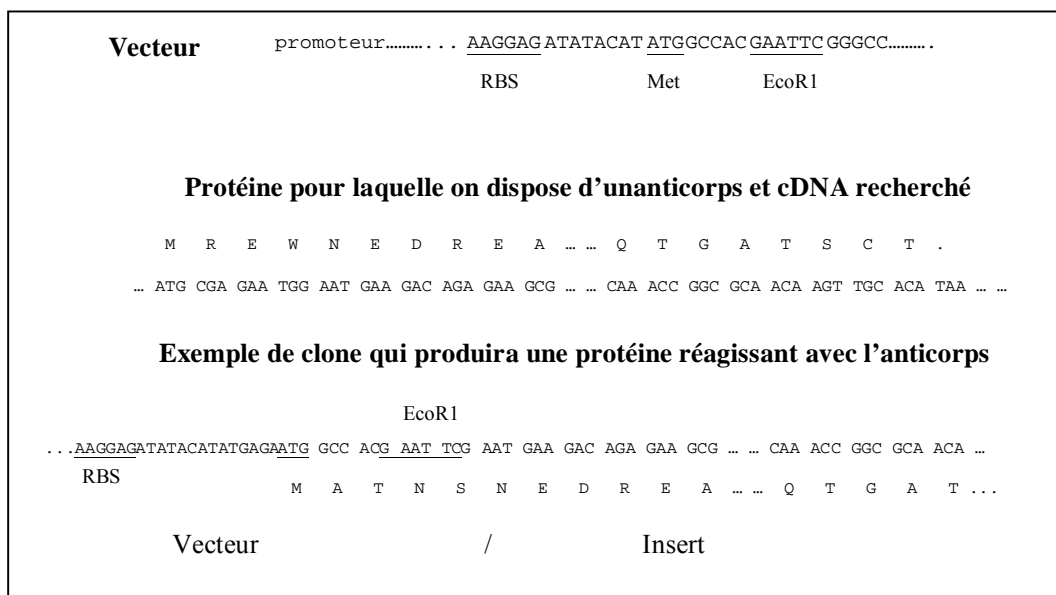
Criblage immunologique d'une banque d'ADNc

La banque est étalée sur une boîte de pétri et répliquée sur des filtres de nitrocellulose. La protéine hybride est repérée en incubant le filtre avec une solution contenant l'anticorps qui lui-même est repéré par la suite à l'aide d'un anticorps secondaire lui-même porteur d'une activité enzymatique.

Dans un premier tour, les plages de lyse seront étalées à forte densité, il sera donc impossible de bien viser la plage qui a exprimé la protéine réagissant avec l'anticorps. On carottera donc la région, et on étalera les phages prélevés de telle façon que les plages soit bien individualisées.

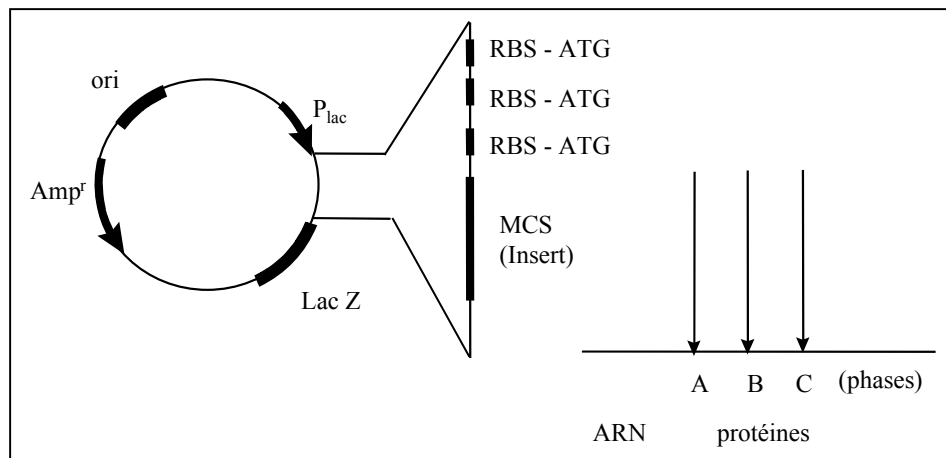


Pour effectuer un criblage immunologique, il faut que la protéine, ou plus exactement une partie de la protéine, soit exprimée par la bactérie. La traduction est initialisée grâce à des séquences présentes dans le vecteur, le site de liaison du ribosome (RBS) composé de la boîte de Shine-Dalgarno et l'AUG à proximité. Il y a donc synthèse d'une protéine de fusion, la partie N-terminale correspond à une séquence codée par le vecteur, la séquence suivante correspond à la traduction de l'insert. Comme le clonage est le plus souvent non orienté, l'insert peut se mettre dans un sens comme dans l'autre. Aussi un insert sur deux sera dans le bon sens, avec l'extrémité correspondant à l'extrémité 5' de l'ARN à proximité du promoteur et des séquences d'initialisation de la traduction. Comme il y a trois phases de lecture possible et que l'extrémité 5' de l'ADNc est variable, un clone sur 3 dans le bon sens sera en phase et codera pour un peptide de séquence identique à la protéine qui a servi à obtenir l'anticorps. Ainsi, un clone sur six sera en phase avec l'ATG présent dans le vecteur.

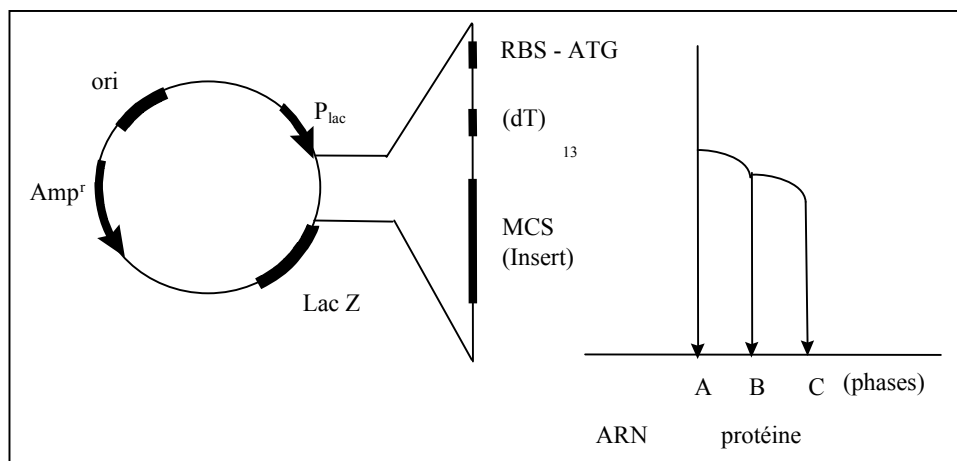


Pour augmenter les chances d'être dans la bonne phase on a trois solutions :

- Faire un clonage orienté, on peut par exemple ajouter un site de restriction en 5' de l'oligo dT servant d'amorce à la reverse transcriptase. Ce site sera différent du site ajouté avec le linker en amont du gène. On passe alors à un clone sur 3 qui sera dans la bonne phase.
- Mettre plusieurs ATG en amont du site de clonage, chacun doit être dans un environnement favorable, avec un site de liaison du ribosome procaryote en amont. Chaque ATG initiera une phase différente.

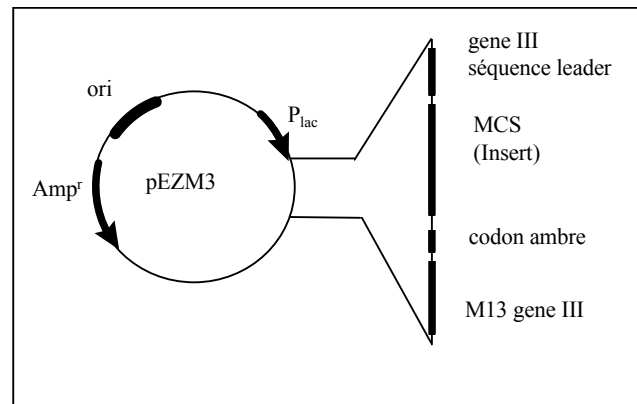


- Intercaler une série de T entre l'ATG et les sites de clonage. Cette série de T fait dériver l'ARN polymérase et on obtient des traductions dans les trois phases.



Pour exprimer les protéines, les différences d'initiation de la traduction entre les eucaryotes et les procaryotes obligent à faire une fusion. La fusion avec la β -galactosidase est la plus courante, mais on peut choisir d'autres fusions comme par exemple l'expression des protéines à la surface du virus M13 pour faciliter le criblage.

Cette méthode est appelée « phage display ». On utilise un vecteur de type pEZM3. Les ADNc sont clonés en fusion du gène III de M13 en aval de la séquence codant pour le peptide leader. Cette protéine, produit du gène III, est normalement présente à l'extrémité du virus lors de l'infection par M13. La fusion en aval du peptide leader suit le même adressage, la protéine de fusion se retrouve à l'extrémité du virus. Pour effectuer le criblage, l'anticorps est fixé au fond d'un



puits de plaque de micro-titration. On prépare une banque dans le plasmide, et on transforme les bactéries. On obtient une banque où chaque clone bactérien fabrique une protéine de fusion protéine III – protéine de la banque. On infecte cette banque par un virus M13 tel que le M13K07. Ce virus est appelé « phage helper », il va se répliquer dans la bactérie, fournir l'ensemble des protéines composant le virus mais peu d'ADN viral, son originien de répllication étant peu efficace.

Chaque bactérie fabrique du virus, composé des protéines structurales provenant de la transcription de l'ADN du phage helper, la protéine de fusion provenant de notre plasmide, et de l'ADN simple brin de notre plasmide, on a en effet ajouter dans le plasmide une origine de répllication du phage M13 (M13 ori). Les virus formés dans une bactérie qui expriment la protéine de fusion se collent à l'anticorps, les autres sont éliminés par lavage. On rajoute des bactéries dans le puits, les virus colonisent les bactéries et on retrouve les plasmides contenant un insert dont le produit a réagit avec l'anticorps.

Plusieurs tours successifs sont quelquefois nécessaires si la protéine est très peu représentée dans la banque. La stratégie habituelle consiste alors à augmenter la spécificité de l'interaction entre chaque tour.

(Il y a d'autres éléments intéressants dans le vecteur utilisé. En exemple : Un codon stop ambre, permet d'obtenir ou non la fusion avec la protéine III : si on utilise une bactérie contenant le supresseur de mutation ambre, on obtient ainsi la fusion nécessaire au criblage, si on utilise une bactérie déficiente pour le supresseur, on obtient une protéine plus proche de la protéine naturelle).

2) On connaît la séquence de la protéine

Si on a purifié la protéine, on peut aussi en séquencer des morceaux après l'avoir coupé à l'aide d'une protéase. Trois stratégies ont été successivement développées pour cloner des gènes à partir de ces séquences peptidiques. On peut cribler une banque d'ADNc à l'aide d'oligonucléotides de séquences déduites de la séquence en acide aminé ; on peut amplifier par PCR le fragment d'ADN compris entre deux oligonucléotides déduits de la séquence protéique enfin on peut cribler *in silico* des banques de données.

Criblage d'une banque avec des oligonucléotides dérivés de la séquence protéique

Le code génétique est dégénéré mais si on trouve une séquence composée uniquement de méthionines et de tryptophanes, on pourra faire un oligonucléotide spécifique de cette séquence. La phénylalanine est codée par deux codons. S'il y a une phénylalanine avec les méthionines et tryptophane, on fera un mélange des deux oligonucléotides. S'il y a une sérine codée par 6 codons, on aura un mélange d'un nombre encore plus grand d'oligonucléotides. Si à la limite il n'y a que des sérines, on fera un mélange d'un très grand nombre d'oligonucléotides. Dans ce dernier cas uniquement un seul des oligonucléotides hybridera à la bonne séquence. Mais la fréquence sera trop faible pour obtenir une hybridation visible et de plus les autres oligonucléotides n'hybridant pas à la séquence recherchée risquent d'hybrider à une séquence n'ayant rien à voir avec la séquence voulue. On utilisera donc parmi les séquences de la protéine disponibles, une région qui donnera le moins d'oligonucléotides possibles.

Acide aminé	Codons possibles					
Met	AUG					
Trp	UGC					
Asp	GAU	GAC				
Glu	GAA	GAG				
Phe	UUU	UUC				
Tyr	UAU	UAC				
His	CAU	CAC				
Lys	AAA	AAG				
Asn	AAU	AAC				
Gln	CAA	CAG				
Cys	UGU	UGC				
Ile	AUA	AUC	AUT			
Gly	GGA	GGC	GGG	GGT		
Ala	GCU	GCC	GCA	GCG		
Val	GUU	GUC	GUA	GUG		
Pro	CCA	CCG	CCC	CCT		
Thr	ACA	ACG	ACC	ACT		
Arg	CGA	CGG	CGC	CGT	AGA	AGG
Ser	UCU	UCC	UCA	UCG	AGU	AGC
Leu	UUA	UUG	CUA	CUG	CUC	CUT

Ces oligonucléotides sont marqués radioactivement avec du γ - ^{32}P ATP et une kinase. La banque est étalée sur des boîtes de pétri, répliquée sur nitrocellulose. Les bactéries (dans le cas d'une banque de plasmide) ou les phages sont répliqués sur nitrocellulose, lysés par la soude ce qui a aussi pour effet de dénaturer l'ADN qui se fixe sur la nitrocellulose. On effectue alors une hybridation que l'on révèle par radioautographie. La position du spot est repérée et permet de retrouver la position de la bactérie ou du phage.

Avant d'hybrider les oligonucléotides sur la banque, on peut les tester pour voir si l'hybridation est efficace et spécifique en les hybridant sur un Southern ou un northern. Si on observe une bande et une seule, il y a de grandes chances que le mélange d'oligonucléotide permettra de cloner le gène recherché. Cependant, il n'est pas possible de récupérer l'ARN hybridé et de fabriquer un ADNc avec cet ARN. Il n'est pas non plus possible de récupérer l'ARN sur le gel en notant la position de l'hybridation. En effet, les gels ne sont pas assez résolutifs et à une position donnée, il y a un grand nombre d'ARN différents.

Une alternative à l'emploi d'un mélange d'oligonucléotides consiste à employer un oligonucléotide plus grand que 20mer (30-40 mer). L'hybridation à plus basse température que le T_m , à faible stringence, permet alors une hybridation sur le gène recherché.

Clonage direct par PCR

Si on a purifié la protéine, on peut la couper à l'aide d'une protéase et séquencer plusieurs fragments. A partir de ces séquences on peut préparer plusieurs oligonucléotides, sens et antisens. On ajoute dans un tube une préparation d'ADNc, deux oligonucléotides hybridant sur chacun des deux brins de l'ADN, une polymérase thermostable et les 4 nucléotides triphosphates. On effectue alors une réaction de PCR. Si les deux oligonucléotides ont reconnu des séquences peu éloignées et s'ils se faisaient face, on aura amplification d'un fragment d'ADN. Pour savoir si ce fragment correspond au gène recherché, il est souvent plus sûr de le cloner dans un plasmide afin de le séquencer.

Le clonage des fragments de PCR dans un plasmide peut se faire par plusieurs méthodes :

- Clonage bord franc: le vecteur est coupé par une enzyme générant des bords francs tels que Sma I ou EcoR V et les extrémités du fragment de PCR sont rendues franches à l'aide d'une ADN polymérase ayant une activité 3'5' exonucléase. De façon à éliminer la recircularisation du vecteur, il sera déphosphorylé, par contre le fragment de PCR sera phosphorylé.
- Clonage par extrémités cohésives. Les amorces servant à l'amplification peuvent comporter des sites de restriction. Le fragment de PCR et le vecteur seront coupés par les mêmes enzymes de façon à obtenir des extrémités cohésives. Toutefois, l'introduction de ces sites diminuera la spécificité des amorces et on a toujours le risque d'avoir un site de coupure interne dans le fragment amplifié.
- Clonage T/A. Les ADN polymérases présentent toutes une activité terminale transférase sur une extrémité double brin. Elles ajoutent donc un seul nucléotide en 3'. Cette activité n'est généralement pas visible car le nucléotide est enlevé par l'activité 3'5' exonucléase de la polymérase. Toutefois si la polymérase est dépourvue de cette activité exonucléase, on obtient un nucléotide sortant en 3'. Comme c'est le dATP qui est le meilleur substrat pour l'activité terminal transférase, on a le plus souvent un A en 3'. La Taq DNA polymérase n'a pas d'activité 3'5' exonucléase, les fragments amplifiés à l'aide de cette polymérase comportent donc des extrémités avec un A sortant. On peut profiter de cette particularité pour cloner le fragment de PCR dans un vecteur ayant un T sortant. Plusieurs

méthodes peuvent être utilisées pour faire de tels vecteurs (voir le chapitre « construction »).

Dans certains cas, il sera difficile d'obtenir plusieurs séquences de la protéine et on aura qu'une seule séquence utilisable. On peut dans ce cas s'en sortir en criblant une banque en utilisant une amorce spécifique et une amorce hybridant sur le vecteur.

Lorsque les amorces sont très dégénérées ou lorsque les températures d'hybridation sont très basses, on peut s'attendre à cloner des gènes non désirés. On obtient dans ce cas plusieurs bandes en PCR ou même si la PCR est peu efficace, on n'obtient aucune bande visible sur le gel. Dans ces deux cas, on effectue une PCR emboîtée ou "nested PCR" : Un aliquote de la première amplification est utilisée comme matrice avec des amorces internes au fragment amplifié.

Criblage de banques de donnée ou criblage *in silico*. Tous les gènes clonés et séquencés sont réunis dans des banques informatiques. Certains génomes des espèces modèles tels que le génome humain, de certains insectes (drosophile), de certaines plantes (*Arabidopsis thaliana*), du nématode (*Caenorhabditis elegans*) ont été complètement séquencés. De plus, il existe des banques de séquences partielles d'ADNc (EST, expressed sequence tag), ces banques ont été obtenues par séquençage des deux extrémités de clones d'ADNc dont on ne connaît rien à priori. Lorsqu'on a une séquence de protéine et que l'on veut obtenir des gènes comportant cette séquence, la première chose à faire est de cribler ces banques informatiques.

3) On a la séquence de gènes proches.

De plus en plus de gènes ont été clonés et leurs séquences sont disponibles sur des serveurs informatiques. Toutefois, les séquences sont principalement disponibles pour des organismes modèles tels que la souris, la drosophile ou l'arabidopsis. Ces séquences peuvent être utiles si on veut cloner un gène d'un organisme tout en connaissant la séquence du même gène mais provenant d'autres organismes. De même, on peut vouloir cloner un gène appartenant à une famille multigénique si les séquences d'autres gènes de la famille sont disponibles. De même, on peut avoir un clone cDNA et

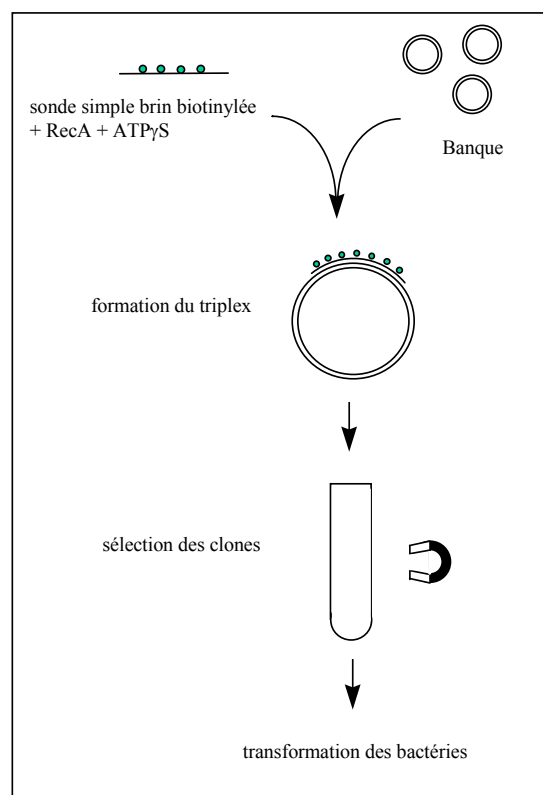
vouloir le clone génomique. Le clone génomique nous permet d'avoir les introns ou les séquences régulatrices présentes en cis (et en particulier le promoteur).

Criblage d'une banque.

Une première méthode consiste à cribler une banque avec un clone connu provenant d'une autre espèce. Comme on ne s'attend pas à une identité parfaite, l'hybridation se fera à faible stringence pour permettre une hybridation en présence de mésappariements. S'il y a trop de positifs, on en déduira que la stringence est trop faible et on l'augmentera soit en augmentant la température soit en diminuant la concentration en sel. L'inconvénient de cette méthode est qu'elle est très laborieuse. Il faut en effet étaler la banque, la dupliquer sur nitrocellulose, dénaturer l'ADN, hybrider, laver et révéler l'hybridation par radioautographie. De plus, on obtient de nombreux faux positifs et il est difficile de trier les vrais positifs (clones qui correspondent au gène recherché) des faux positifs (clones correspondants à des inserts qui se sont hybridés à la sonde mais qui ne correspondent pas au gène recherché).

Pour simplifier le criblage, on peut former un triplex entre la sonde et l'ADN cible. Dans ce cas, la sonde s'hybride avec l'ADN double brin. Le triplex est formé grâce à la protéine RecA. Cette protéine provient du système de réparation d'*E. coli*. Elle favorise la formation d'un triplex suivie d'une recombinaison en présence d'ATP. Toutefois, en présence d'ATP γ S, la catalyse n'a pas lieu et le triplex est stabilisé. Ainsi pour cribler une banque, on ajoute la sonde en présence de la protéine RecA et d'ATP γ S.

Pour récupérer les triplex on utilise une sonde biotinylée. Après hybridation, les clones positifs sont récupérés à l'aide de billes magnétiques sur lesquelles on a fixé de la streptavidine. Le



plasmide double brin peut ensuite être récupéré par un traitement alcalin. (A priori on pourrait faire

une hybridation avec de l'ADN de la banque dénaturé. Toutefois, on récupérerait à la fin de l'expérience de l'ADN simple brin qui est peu transformant comparé à l'ADN double brin utilisé dans cette technique).

PCR.

On choisira dans ce cas des régions conservées entre les différents gènes appartenant à la famille. Ces boîtes conservées sont souvent des séquences importantes pour l'activité de la protéine ou pour sa structure. Parmi les différentes séquences possibles, on choisira celles qui sont le moins dégénérées comme dans le cas d'un clonage à partir des séquences de la protéine. Des oligonucléotides seront déduits de ces séquences, synthétisés et utilisés comme amorces pour des amplifications. Les PCR seront souvent effectuées à faible stringence et les produits d'amplifications seront soumis à des PCR emboîtés ("nested PCR"). Cette méthode initiée par Gould et ses collaborateurs en 1989 est actuellement une des méthodes les plus utilisées pour cloner des gènes dans des organismes pour lesquels on n'a que peu d'information. Le clonage s'effectuera alors dans les mêmes conditions que précédemment, lorsque les oligonucléotides dérivent de la séquence protéique.

Criblage de banques de données.

Comme précédemment, la première chose à faire est de cribler des banques informatiques lorsque l'on a une séquence consensus et que l'on veut obtenir des gènes comportant cette séquence. Par exemple quelqu'un était intéressé par des gènes codant pour les galectines (lectines reconnaissant le galactose). En criblant les banques de données avec une séquence consensus des galectines, il a trouvé un gène dont le produit était mal identifié et donc mal nommé. L'expression *in vitro* du gène a permis de vérifier cette nouvelle identification.

4) On a un clone et on veut les parties flanquantes

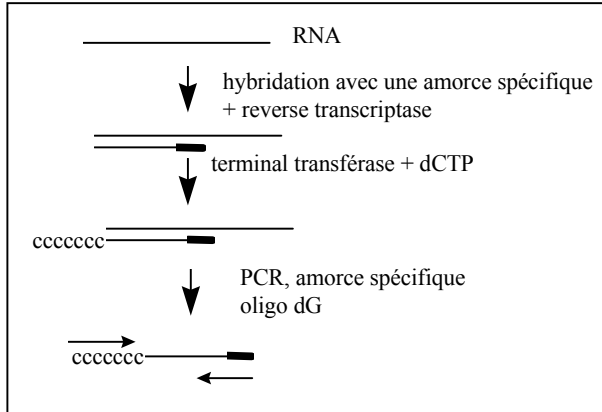
Les criblages immunologiques donnent des ADNc incomplets, pour lesquels il manque la partie 5' du fait des méthodes de synthèse du deuxième brin. Lors du clonage par PCR en prenant des amorces internes, il manquera les extrémités 5' et 3' de l'ADNc. Il nous faudra donc les obtenir, plus particulièrement si la phase de lecture n'est pas complète. Notre but sera donc de cloner ces séquences pour reconstituer un ADNc comportant au moins toute la partie codante.

Pour avoir la séquence correspondant à l'extrémité 3' de l'ARNm, c'est relativement facile. On fabrique un ADNc simple brin à l'aide d'une amorce oligodT puis on effectue une PCR avec un oligonucléotide sens spécifique et l'oligodT comme amorce antisens (pour les ARN eucaryotes).

- Mais, l'oligodT se fixe fréquemment sur des séquences internes et on peut obtenir un grand nombre de fragments provenant de deux amorces oligodT sens et antisens. Pour atténuer ce problème, lors de la fabrication de l'ADNc avec la reverse transcriptase, on amorce avec un oligodT suivi du côté 5' par une séquence quelconque. Un oligonucléotide comportant uniquement cette séquence sera utilisé dans la réaction de PCR.
- Dans certains cas, on obtient plusieurs bandes, du à la faible spécificité de l'amorce interne et à la faible représentation de l'ADNc recherché. Dans ce cas on fait des PCR emboîtées. Une première PCR est faite avec un oligonucléotide spécifique et l'oligodT, le produit de cette amplification est alors utilisé pour effectuer une deuxième PCR avec un autre oligonucléotide spécifique hybridant en 3' de l'oligonucléotide utilisé pour la première PCR et l'oligodT.

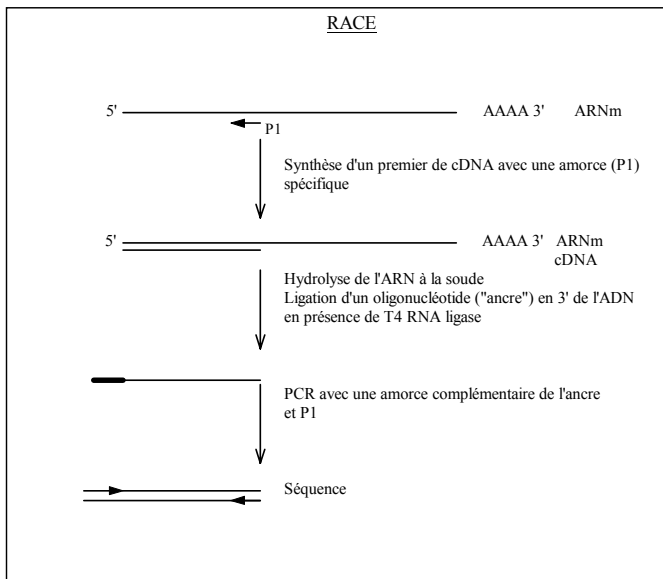
Pour obtenir l'extrémité 5', on peut encore faire une réaction de PCR. Dans certains cas particuliers, on connaît la séquence 5' de l'ARN, lorsqu'il y a trans-épissage comme chez *Trypanosoma brucei* ou chez *Caenorhabditis elegans*. l'ARN messenger mature provient de la ligation de deux ARN et l'ARN en 5' est toujours le même. Pour trouver l'extrémité 5', on effectue simplement une PCR avec une amorce spécifique de la séquence rajoutée et une amorce spécifique de la séquence du gène d'intérêt. Mais généralement on ne connaît pas la séquence en 5' de l'ARN, il va donc falloir ajouter une séquence connue en 5' du gène. Il existe plusieurs méthodes appelées RACE (Rapid Amplification of cDNA End).

- On peut ajouter une queue homopolymérique (par exemple polydC) à l'aide d'une terminal



transférase et d'un seul oligonucléotide (dCTP) à l'extrémité 3' du premier brin de cDNA. Un oligonucléotide complémentaire à cette queue (oligodG) pourra alors être utilisée avec une amorce spécifique de la séquence connue pour amplifier la partie 5' du gène. L'amorce spécifique sera un oligonucléotide hybridant en 3' du premier brin de l'ADNc. L'inconvénient majeur de cette méthode est que les ADNc non complets servent

de matrice à la réaction de PCR, on obtient donc plusieurs produits d'amplification, le plus grand étant censé représenter le clone complet.



- On peut ajouter un oligonucléotide à l'extrémité 3' du cDNA (LA-PCR: ligation-anchored PCR ou SLIC: single-strand ligation to ss DNA ends). On utilise la T4 RNA ligase qui peut liquer des acides nucléiques simple brin pour liquer un oligonucléotide à l'extrémité 3' de l'ADNc. La PCR peut ensuite se faire en utilisant une amorce spécifique et une amorce

complémentaire à l'oligonucléotide ajouté en 3' de l'ADNc.

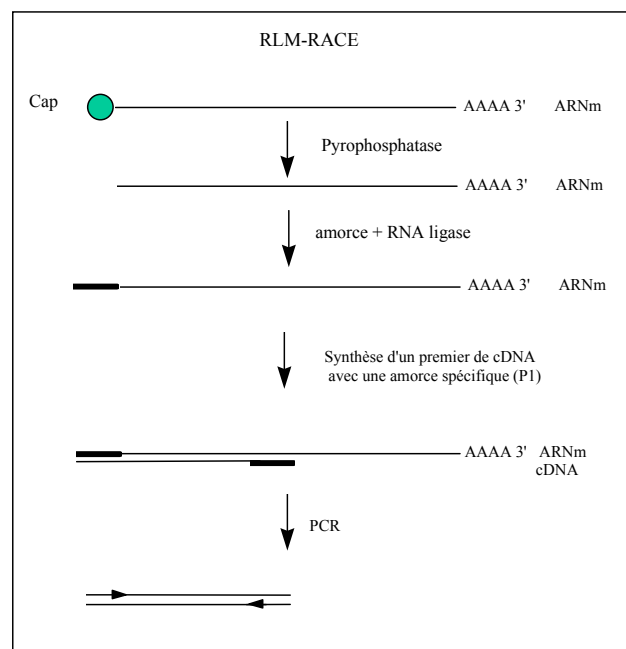
Pour avoir uniquement la ligation entre l'oligonucléotide et l'ADNc il va falloir :

- Eliminer l'ARN matrice de la reverse transcription qui serait un bon substrat pour la T4 ARN ligase. Il sera détruit par un traitement alcalin qui converti l'ARN en ribonucléotides 3' phosphate.
- Phosphoryler l'oligonucléotide en 5' pour permettre la ligation.
- Bloquer l'oligonucléotide en 3' pour empêcher la ligation de concatèmes. Le blocage en 3' peut être effectué en incorporant un analogue de nucleotides (aminonucléotide) comme première base lors de la synthèse ou en ajoutant un didéoxynucléotide en 3' de l'oligonucléotide déjà synthétisé à l'aide d'une terminal transférase.
- L'amorce qui sert à démarrer la réaction de reverse transcription doit avoir un 5' OH de façon à ne pas servir de substrat à la ligation.

L'inconvénient de cette technique est le même que pour les queues homopolymériques, les ADNc non complets peuvent servir de matrice à la ligase. De plus la réaction de ligation à l'aide de l'ARN ligase est peu efficace, en effet si l'ADN peut servir de donneur (c'est à dire amener le phosphate) à la ligation, il est un accepteur peu efficace or, ici, le donneur (l'oligonucléotide) comme l'accepteur (ADNc) sont des ADN.

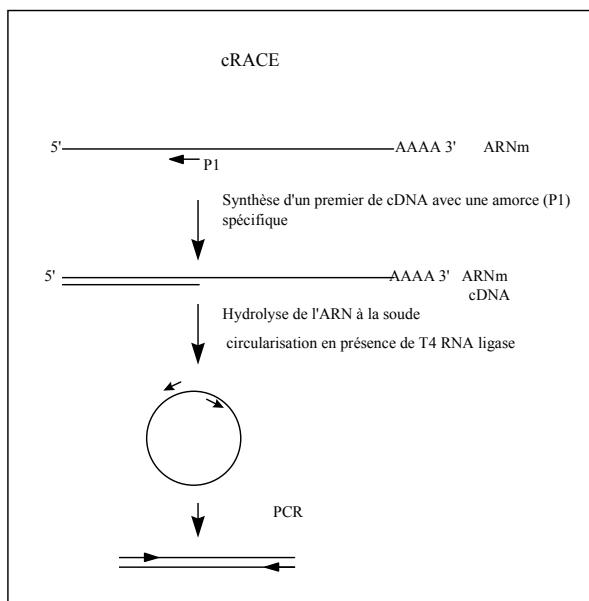
- On peut ajouter un oligonucléotide en 5' de l'ARN (RLM-RACE : RNA ligase-mediated RACE, ou RLPCR : reverse ligation-mediated PCR).

Il faut tout d'abord retirer le Cap. Pour ce faire, on utilise une pyrophosphatase (tobacco acid pyrophosphatase, TAP) qui hydrolyse les liaisons anhydres de l'acide phosphorique et laisse un 5' phosphate à l'extrémité 5' des ARNm décapés. Un oligonucléotide est alors attaché en 5' de l'ARN à l'aide de la T4 ARN ligase. L'amorce utilisée ne doit pas être phosphorylée pour éviter d'avoir des concatèmes. Elle est en excès ce qui évite les liaisons entre plusieurs ARN ou les circularisations. Elle peut être un



oligodésoynucléotide mais l'ADN étant un pauvre accepteur dans la ligation, il est préférable d'utiliser une amorce ARN. L'ARN est ensuite reverse transcrit et le fragment désiré est amplifié par PCR.

L'amorce peut à priori s'accrocher sur toutes les extrémités 5' phosphate, qu'elles proviennent d'ARN partiellement dégradés ou du produit du décapage. Pour éviter les premiers, on peut faire agir une phosphatase alcaline avant d'utiliser la pyrophosphatase. Ainsi, les ARN dépourvus de Cap avant l'action de la pyrophosphatase, ne pourront pas être des substrats de la ligation (Schaeffer, 1995).



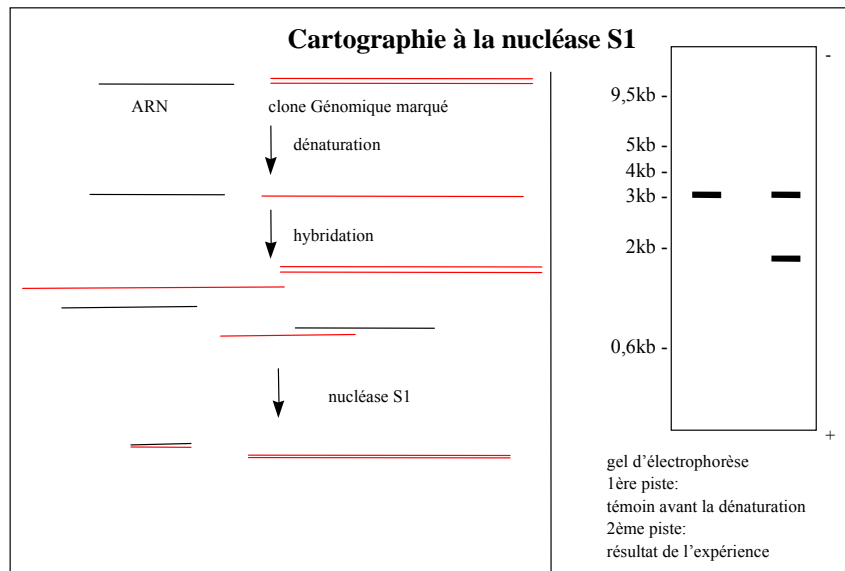
- On peut circulariser l'ADNc (circular-RACE ou cRACE). Une amorce spécifique du gène est phosphorylée puis utilisée dans une reverse transcription. L'ARN est détruit à la soude et l'ADNc est ligué avec la T4 RNA ligase. Comme la concentration de l'ADNc est faible, on favorise la circularisation. Cet ADNc circulaire est alors utilisé comme matrice en utilisant deux amorces en sens inverse (Maruyama et al., 1995).

Une autre stratégie consiste à fabriquer une banque d'ADNc complets. Dans ce cas le deuxième brin d'ADNc n'est pas initialisé par l'extrémité 3' du premier brin comme dans la technique utilisant la nucléase S1 ni par les amorces ARN provenant de la digestion partielle de l'ARN comme dans la technique utilisant la RNase H. On utilise pour faire le deuxième brin, une amorce complémentaire à une séquence qui a été rajouté en 3' du premier brin par les mêmes techniques que celles utilisées pour les RACE. Par exemple, on peut ajouter une queue homopolymérique en 3' du premier brin à l'aide d'une terminal transférase, ou ajouter une oligonucléotide à l'aide d'une ARN ligase.

Avant l'utilisation de la PCR, d'autres méthodes permettaient d'obtenir l'extrémité 5' :

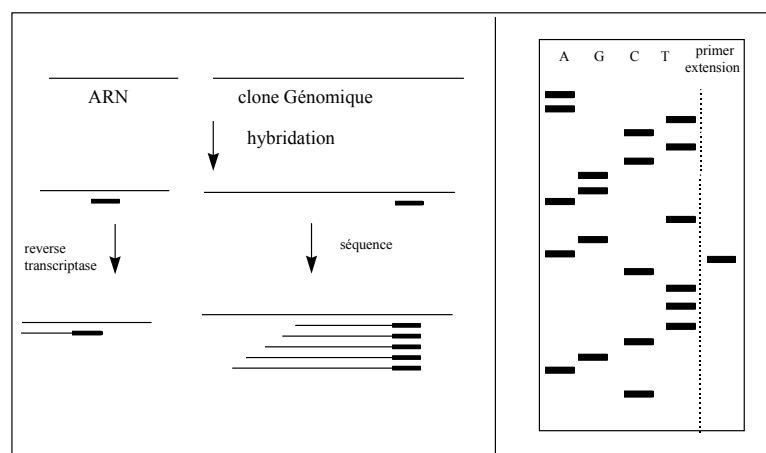
elles consistaient à trouver sur l'ADN la région correspondante au début de l'ARN, le nucléotide d'initiation de la transcription. Trois méthodes principales peuvent être utilisées: cartographie à la nucléase S1, RNase protection et primer extension.

La technique de la nucléase S1 consistait à hybrider un clone génomique radioactivement marqué à une préparation d'ARN. Le produit de la ligation est ensuite incubé avec la nucléase S1 pour digérer les fragments simples brins. La taille de l'hétéroduplex ADN-ARN peut ensuite être estimée sur un gel.



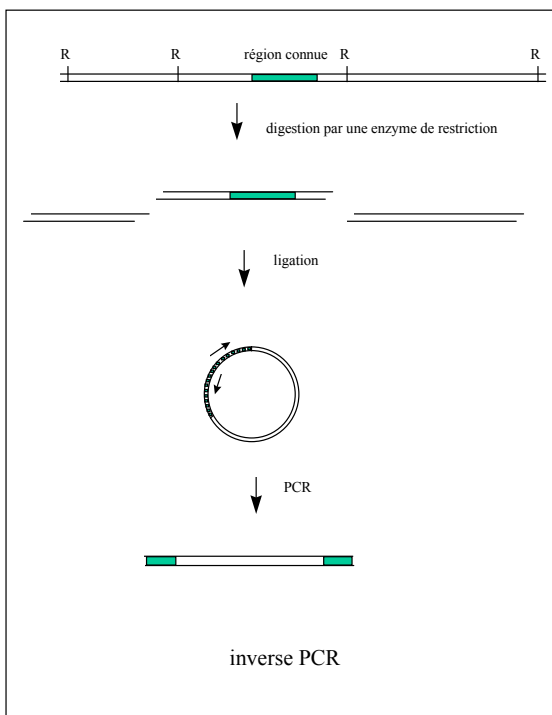
La RNase protection suit le même principe, mais, au lieu d'utiliser un ADN marqué, on utilise un ARN antisens (cRNA), transcrit *in vitro* et marqué, obtenu à partir du clone génomique. L'ARN simple brin de la sonde est alors digéré par un mélange de RNase A et T1. Cette méthode diffère de la précédente par la méthode de marquage du clone génomique et par la digestion du simple brin, mais le principe reste le même.

La technique d'extention d'amorce (primer extension) consiste à utiliser une amorce antisens hybridant dans la région 5' de l'ARN connue. Une polymérisation est effectuée en présence de reverse transcriptase et de desoxynucléotides marqués. La polymérisation s'effectue jusqu'au bout de l'ARN. Le produit de la réaction, correspondant à l'ADN complémentaire de l'extrémité 5'



de l'ARN est alors déposé sur un gel de séquence à côté d'une réaction de séquence (réalisée par la méthode enzymatique) utilisant un clone génomique et la même amorce.

On peut aussi vouloir cloner les séquences en amont ou en aval d'un clone génomique. On a par exemple un clone et on veut connaître les parties flanquantes de ce clone. Une première stratégie consiste à effectuer une marche (voir paragraphe suivant), mais dans certains cas on ne veut qu'une petite région, par exemple lorsqu'on veut connaître la séquence d'insertion d'un élément transposable. Pour cloner cette région, il existe plusieurs méthodes quelquefois dénommées RAGE (pour rapid amplification of genomic ends).

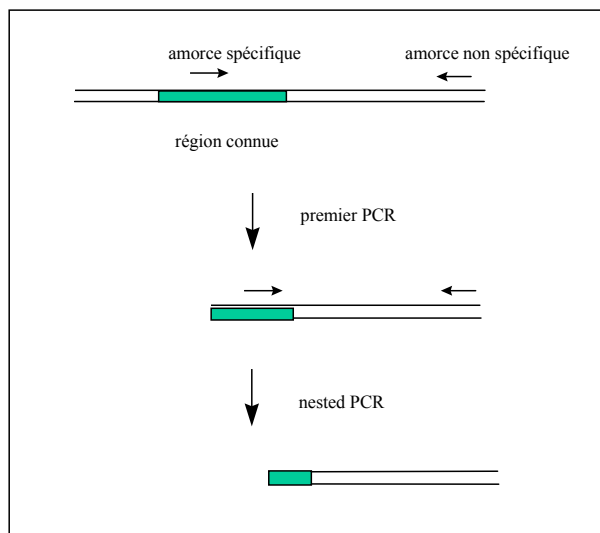
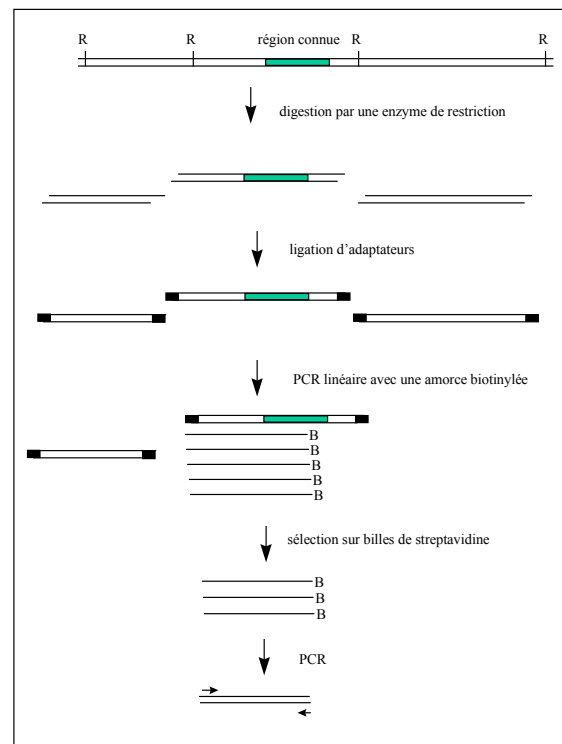


La PCR inverse : on digère l'ADN par une enzyme de restriction, on effectue une ligation sur le produit de digestion en diluant assez l'ADN pour favoriser les ligations intramoléculaires, puis on effectue une PCR en prenant deux amorces en sens opposés (Oachman et al., 1988).

On utilise généralement une enzyme de restriction qui coupe souvent pour que la région inconnue soit petite (< 2kb) et donc facilement amplifiable par la réaction de PCR.

Une deuxième technique consiste à couper l'ADN avec une enzyme de restriction puis à liguer des adaptateurs aux extrémités des fragments de restriction. On amplifie le fragment d'intérêt par PCR avec deux amorces, une spécifique du fragment qui hybride sur la région connue et une amorce complémentaire à l'adaptateur.

Si on applique ce protocole, on aura aussi des amplifications entre les amorces hybridant aux deux extrémités de tous les fragments de restriction. Pour éviter cette amplification, on peut faire une première PCR avec uniquement l'amorce qui hybride sur le fragment connu, cette PCR est linéaire. On utilise une amorce biotinylée, ce qui permet sélectionner les fragments simples brins de cette amplification. Ces fragments servent alors de matrice pour l'amplification avec les deux amorces, une spécifique et une hybridant sur l'adaptateur (Rosenthal et Jones, 1990)



- Une troisième technique consiste à amorcer au hasard sur l'ADN. Pour la première amplification on utilise une amorce spécifique et une série d'amorces non spécifiques, dégénérées. Le produit de l'amplification est alors utilisé pour faire une « nested PCR » avec une amorce spécifique interne. Cette deuxième amplification permet de sélectionner les bons produits dans la première amplification (Parker et al., 1991).

5) On a une mutation, sa position ou son origine.

On a un clone pas trop éloignée de notre gène : marche sur le chromosome.

Le clone recherché coségrège avec des marqueurs génomiques. A partir de ce marqueur il va falloir obtenir le clone recherché. On utilise dans ce cas des banques génomiques construites soit dans le phage λ , soit dans un cosmide soit dans un YAC.

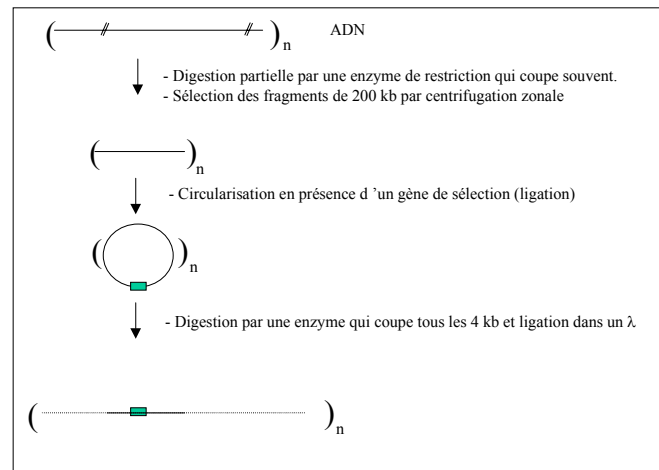
Le premier clone est cartographié par digestion à l'aide de quatre enzymes de restriction. On isole le fragment d'ADN correspondant à une extrémité de l'insert. Cet ADN est marqué, puis utilisé comme sonde pour cribler la banque génomique, les clones obtenus sont à nouveau cartographiés. Le fragment se situant le plus à l'extrémité de la région, est marqué et utilisé comme sonde pour à nouveau cribler la banque génomique. On obtient ainsi une carte de restriction de la région et une série de clones qu'il est possible d'arranger sur la carte.

Des marches ont déjà été effectuées chez plusieurs organismes et on dispose par exemple chez *Caenorhaditis elegans* de filtres avec tous les YAC arrangés dans l'ordre, pour chaque YAC on dispose de cosmides eux-mêmes arrangés.

Problème : Si la sonde utilisée dans un des tours tombe sur un ADN répété, un grand nombre de clones sont hybridés. La plupart d'entre eux hybrident dans d'autres endroits du génome et la marche ne peut être continuée. Pour résoudre ce problème on a deux solutions. La première est de travailler sur deux génomes simultanément, deux génomes provenant de deux souches différentes. Les éléments répétés n'étant pas toujours à la même place, le problème ne se répétera pas aux même endroits sur les deux génomes. La deuxième solution consiste à prendre comme sonde un fragment d'ADN un peu en amont de l'extrémité de la région déjà clonée.

Variante : si on n'a pas de clone dans la région on peut utiliser un clone d'une autre région et faire un saut en utilisant des translocations.

Banque de saut : la marche sur le chromosome est une technique relativement laborieuse, avec λ on avance de 10 à 15 kb à chaque tour et il faut environ une semaine pour faire ce tour. Pour avancer plus vite on peut faire une banque de saut. Comme pour toutes les banques génomiques, on prépare de l'ADN de haut poids moléculaire, on le coupe avec une enzyme de restriction qui coupe souvent comme par exemple Sau3A qui reconnaît 4 bases (GATC) et coupe donc en moyenne tous les 4^4 bases soit toutes les 256 bases. La digestion est ménagée pour pouvoir obtenir des grands fragments se recouvrant. On trie les fragments de 200 kb par centrifugation zonale. Ces fragments sont ligés en



présence d'un gène de sélection. On obtient des fragments circulaires comportant un seul fragment d'ADN de départ de 200 kb et au moins un gène de sélection. Ce fragment d'ADN n'est pas un plasmide puisqu'il n'y a pas d'origine de réplication. On digère le fragment circularisé par une enzyme de restriction qui reconnaît 6 bases et qui coupe donc en moyenne tous les 4100 bases. Et on insère les fragments dans un λ . Tous les fragments peuvent s'insérer mais seuls ceux qui portent le gène de sélection vont permettre à la colonie de pousser. Les inserts comportent donc uniquement les deux extrémités des fragments de 200 kb de départ. La sélection d'un clone par hybridation permettra donc d'isoler un fragment d'ADN situé à 200 kb de la sonde.

Une marche permet d'avoir une région, si on cherche un clone particulier, une mutation, on devra encore l'identifier dans l'ensemble de la région par exemple en faisant la carte de transcription. Pour ce faire on analyse la présence de transcrit par Northern ou par "exon trapping". Cette dernière technique consiste à faire transcrire des fragments d'ADN et à rechercher les ARNm les plus grands.

La mutation est due à un élément transposable,

Pour cloner certains gènes, on peut mobiliser des éléments transposables pour induire la mutation. On induit donc la transposition et on trie les mutants ayant le phénotype désiré dans la descendance. On

fait ensuite les croisements nécessaires pour éliminer tous les éléments transposables présents dans le génome sauf celui responsable de la mutation désirée. Il suffit ensuite de faire une banque génomique du mutant puis de la cribler avec une sonde correspondant à l'élément transposable mobilisé au début de l'expérience. Cette technique a été utilisée avec succès pour cloner des gènes chez la drosophile.

On peut maintenant effectuer des nano-dissections de chromosomes métaphasiques en utilisant un microscope à champ proche encore appelé microscope à sonde locale (Thalhammer et al. 1997). On utilise dans ce cas un microscope de force atomique (Atomic Force Microscopes, AFM) qui exploite les interactions entre particules virtuelles. Lorsque la pointe du microscope touche l'objet, la tige qui porte la pointe se courbe. On analyse le rayon de courbure nanométrique d'une pointe au moyen d'un tube piézo-électrique. On balaye le chromosome avec la pointe du microscope en mode contact avec une faible force, de l'ordre de quelques picoNewtons. Les variations de position de la pointe sont suivies avec un faisceau laser, ce qui permet d'obtenir une image du chromosome. On effectue alors un balayage localisé perpendiculaire à l'axe du chromosome sur uniquement une région, mais ici on effectue une pression plus importante de l'ordre de 100 picoNewtons. La pointe déchire le chromosome et une partie du chromosome se retrouve sur la pointe. Cette pointe est directement utilisée pour une réaction de PCR utilisant des amorces au hasard. Les produits PCR sont ensuite clonés dans un plasmide.

6) On a une fonction sélectionnable

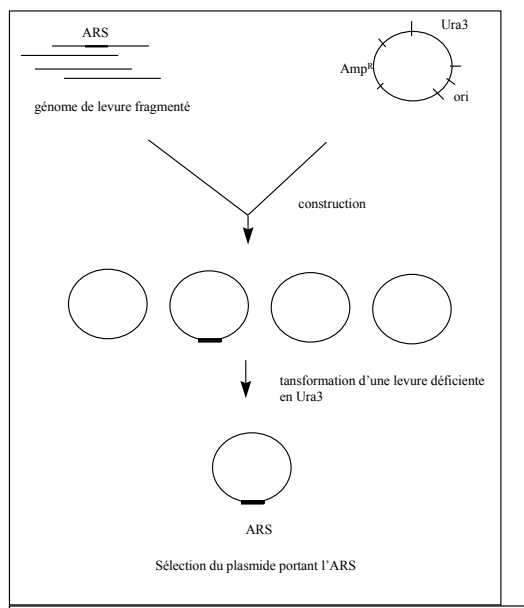
Dans ce cas le gène d'intérêt est cloné parce que son produit confère à la cellule hôte une caractéristique décelable. On va donc faire une banque d'expression et repérer le clone comportant le gène d'intérêt grâce à son activité. La banque d'expression peut être faite dans *E. coli* mais aussi dans des vecteurs eucaryotes, dans des levures ou des cellules d'eucaryotes supérieurs.

Le produit du gène a une activité enzymatique qu'on peut détecter.

Exemple du clonage de la paraoxonase de *Pseudomonas sp.* Cette enzyme coupe le parathion pour donner comme produit du p-nitrophenate qui est coloré. Pour cloner le gène, il a suffi de faire une

banque d'expression, les inserts étaient insérés en fusion avec le peptide α de la β -galactosidase en aval d'un promoteur procaryote. La banque a été étalée sur un milieu contenant du parathion. Les colonies jaunes synthétisaient l'enzyme et donc comportaient le plasmide contenant le gène.

Le produit du gène confère la viabilité de la cellule hôte.



Cette méthode permet de cloner des éléments indispensables des plasmides soit des origines de réplication soit des gènes de résistance à des drogues. Cette technique a par exemple été utilisée pour cloner des origines de réplication autonome de la levure (ARS). Chez la levure, il existe des fragments d'ADN circulaire doués d'autoréplication, on les appelle des épisomes et sont l'équivalent des plasmides chez les bactéries. Les séquences responsables de la réplication des épisomes (ARS) ont été clonées et identifiées en transfectant une banque génomique faite à partir d'un vecteur dépourvu d'origine de réplication. Ce vecteur n'avait pas d'origine de réplication mais comportait un gène de sélection.

Seules les colonies poussant en présence du facteur de sélection comportaient une origine de réplication dans le plasmide.

Cette méthode a aussi permis de cloner les gènes conférant la résistance à un composé toxique. Par exemple le gène codant pour la Dihydrofolate réductase (DHFR) responsable de la résistance à une drogue anticancéreuse, le méthothrexate, qui a été cloné de cette façon.

Le maintien du plasmide peut résulter d'une complémentation de la cellule hôte. Dans ce cas là, on utilise le plus souvent un hôte muté. Cette stratégie est très utilisée avec la levure. En effet, on dispose d'un grand nombre de mutants conditionnels qui peuvent être complémentées avec des gènes provenant d'autres espèces. Par exemple, pour cloner un gène impliqué dans le stress oxydatif chez *Arabidopsis thaliana* dans *E. coli*, on a utilisé une souche de *E. coli* qui ne peut pas pousser en

présence de plus de 100 μM H_2O_2 . Cette souche a été transformée par une banque d'expression d'*Arabidopsis* et les transformants ont été maintenus en présence d' H_2O_2 qui est toxique pour la bactérie. Cette méthode a permis de cloner une peroxydase. Cette technique est aussi très utilisée chez les procaryotes, par exemple le clonage du gène responsable de la motilité chez *Flavobacterium* a été effectué en transformant une souche mutante avec une banque d'individus sauvages.

Le produit du gène recherché réagit avec un ligand.

Dans ce cas, on peut faire une banque d'expression en cellules eucaryotes. Si la protéine est exprimée à la surface de la cellule, on pourra sélectionner les recombinants en immobilisant le ligand sur un support solide. Pour exprimer la protéine à la surface de la cellule, les gènes seront clonés en aval d'une séquence codant pour un peptide signal permettant la signalisation vers la voie de sécrétion. En aval, on ajoutera une séquence codant pour un ancrage à la surface de la membrane cellulaire. Pour ce faire on peut par exemple employer le domaine transmembranaire de la PDGFR (platelet-derived growth factor receptor). Il y a toujours un danger dans cette technique. On peut en effet cloner un gène régulateur qui active un endogène, ce dernier réagissant avec le ligand.

La protéine codée par le gène recherché interagit avec un acide nucléique ou une autre protéine

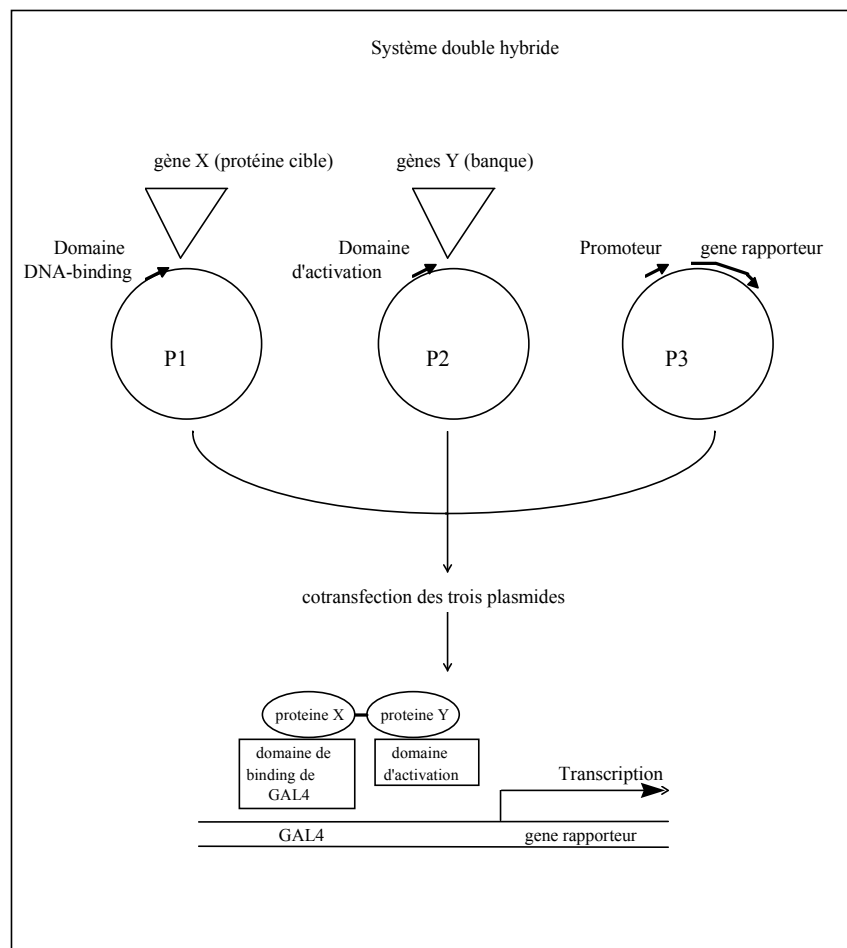
- La protéine interagit avec une séquence d'ADN

Certaines protéines, telles que les facteurs de transcription, ont une affinité pour l'ADN, et en particulier pour certaines séquences d'ADN qui ont un rôle de régulation. Pour cloner un gène codant pour une protéine ayant une affinité pour une séquence d'ADN, on peut étaler une banque d'expression et la cribler avec l'ADN marqué. Au lieu d'effectuer une hybridation ADN/ADN comme pour le clonage à partir d'une sonde, on effectue une reconnaissance protéine/ADN. Il est nécessaire dans ce cas là que la protéine exprimée par un clone de la banque soit active, c'est à dire soit bien repliée et n'ait pas besoin de modification post-traductionnelle pour fonctionner.

- La protéine interagit avec une autre protéine

Pour cloner un gène codant pour une protéine interagissant avec une autre protéine, on peut utiliser le système double hybride. Dans ce système un facteur de transcription est coupé fonctionnellement et physiquement en deux parties, un domaine de liaison à l'ADN et un domaine d'activation de la

transcription. Si ces domaines sont réunis en trans, leur activité est restaurée. La protéine cible est exprimée en fusion avec le domaine de liaison à l'ADN et la protéine recherchée est exprimée en fusion avec le domaine d'activation de la transcription.



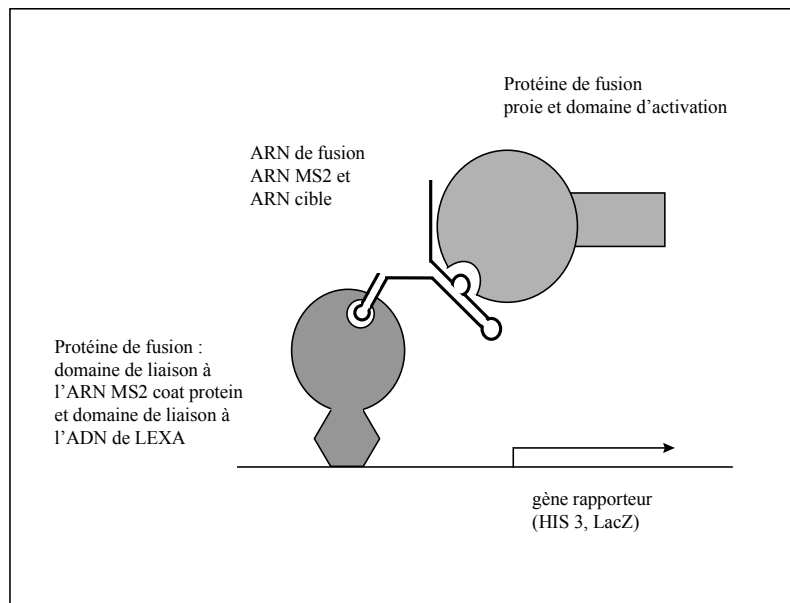
- On construit un plasmide (P1) portant un gène codant pour une fusion de la protéine cible et d'un domaine DNA-binding d'un activateur transcriptionnel.
- On construit un plasmide (P3) qui comporte un gène rapporteur en aval d'un promoteur inductible par l'activateur transcriptionnel.
- On construit la banque dans un plasmide (P2) en fusion avec le domaine d'activation du domaine d'activation de l'activateur transcriptionnel.

On construit un clone de levure comportant P1 et P2 et on transforme cette souche avec la banque faite dans P3. Lorsqu'un clone de P3 fabrique une protéine qui a une forte affinité pour la protéine cible fabriquée par P1, les deux protéines se lient, on a rapprochement des domaines DNA-binding et activation transcriptionnelle grâce à l'interaction des deux protéines de fusion. On a alors transcription et traduction du gène rapporteur.

Dans ce système, on utilise un activateur transcriptionnel qui a deux domaines fonctionnels. Le premier domaine a une fonction de DNA-binding qui permet à la protéine de se fixer au bon endroit sur l'ADN, ce domaine est codé par le plasmide P1. Le deuxième domaine a une fonction d'activateur transcriptionnel, il est codé par P2. Ces deux domaines doivent être présents pour avoir une activation de la transcription. Ils sont normalement présents sur la même protéine. Mais pour les besoins du test on les exprime séparément, et dans ce cas il n'y a activation transcriptionnelle que lorsque les deux domaines sont réunis. Ce système appelé double hybride, est utilisé chez la levure en utilisant par exemple GAL4 qui est un activateur transcriptionnel dont on peut séparer les deux domaines, DNA binding et activation de la transcription par expression de gène tronqué.

- La protéine interagit avec un ARN

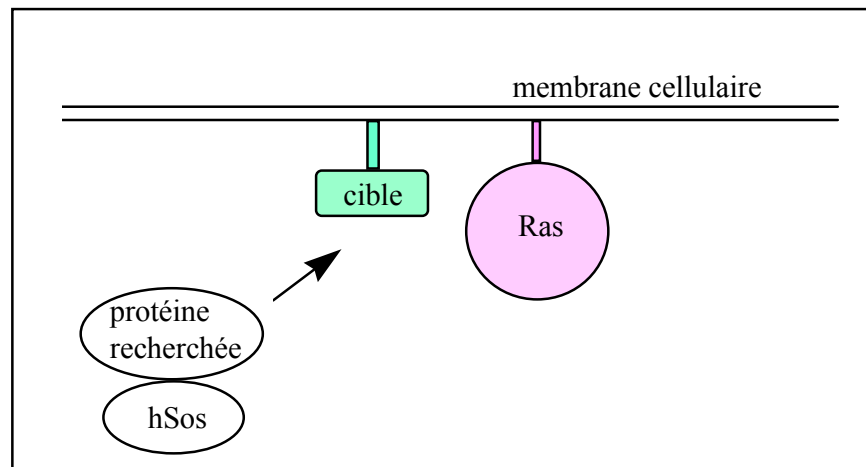
Le système est identique si on veut cloner un gène codant pour une protéine interagissant avec un ARN. Ces protéines ont un rôle majeur dans de nombreux processus biologiques tels que la maturation des ARN, la traduction ou l'infection par les virus. Ici encore, on va rapprocher un domaine d'activation de la transcription (domaine d'activation de B42) à un promoteur. Cette proximité va induire une activité transcriptionnelle d'un gène rapporteur. L'ARN cible est lié au domaine DNA binding via une fusion avec l'ARN MS2 qui lui-même est reconnu par une protéine hybride qui a deux domaines, un qui reconnaît le promoteur et un qui reconnaît l'ARN MS2. Ainsi l'ARN cible se trouve à proximité du promoteur. La protéine cible sera fusionnée avec le domaine d'activation. La liaison de la protéine reconnaissant l'ARN cible rapprochera les domaines de binding et d'activation et se traduira par une activité transcriptionnelle repérable grâce à un gène rapporteur.



La méthode de double hybride a permis de cloner un grand nombre de gènes comme par exemple des récepteurs de virus, des récepteurs reconnaissant des signaux d'adressage, des protéines interagissant au niveau des promoteurs.

Toutefois, elle présente certaines limitations, les macromolécules doivent interagir dans le noyau. De plus on risque de cloner des gènes dont le produit est lui-même un activateur transcriptionnel. Pour résoudre ces problèmes, on peut utiliser une interaction protéine-protéine dans le cytoplasme qui active la voie de signalisation de Ras.

On utilise une levure, *Saccharomyces cerevisiae* mutante sur le gène *cdc25*, la mutation est thermosensible, à 37°C, la protéine ne se fixe pas sur Ras et la levure ne pousse pas. Le gène *cdc25* est analogue au gène *hSos* humain qui code pour un facteur d'échange de nucléotides guanylyl, on peut compléter la mutation par la protéine hSos produite du gène humain *hSos*. La complémentation ne marche que si la protéine hSos est ancrée sur la membrane. On va donc exprimer la protéine cible sur la membrane en y ajoutant un signal de myristylation. La banque sera construite dans un plasmide de levure (épisode) pour avoir une fusion des cDNA avec le gène hSos. Si une protéine produite par la banque interagit avec la protéine cible, on aura liaison de hSos à la membrane, donc interaction entre hSos et Ras et la cellule produisant cette protéine pourra pousser à 37°C.



La séquence recherchée est une séquence régulatrice.

Il s'agit ici de trouver des séquences régulatrices, dans ce cas la banque sera faite en amont du gène rapporteur.

Exemple : "enhancer trap" chez la drosophile. La technique consiste à transformer l'insecte avec un gène rapporteur tel que la β -galactosidase. L'activité de l'enzyme dépend du promoteur derrière lequel le gène sera intégré. On trouvera ainsi des mouches exprimant la β -galactosidase que dans certaines cellules, en criblant une banque génomique du transformant avec le gène codant pour la β -galactosidase, on obtiendra les séquences responsables de l'expression recherchée.

La séquence recherchée interagit avec une protéine.

Dans ce cas on a une protéine et on veut identifier son ligand : un fragment d'ADN ou d'ARN qui se lie à la protéine. Il existe plusieurs méthodes, par exemple :

- On produit la protéine *in vitro*, on la purifie et on la fixe sur un support solide. Un des moyens est de d'accrocher une protéine sur un support est de la fabriquer *in vitro* en présence de lysine biotinylée puis de l'accrocher sur de la streptavidine elle-même liée à un support solide comme des billes. Les ARN (ou l'ADN) de la cellule sont incubés avec les billes portant la protéine, les ARN non fixés sont éliminés par lavage et les ARN fixés sont récupérés à l'aide d'une solution dénaturante des protéines. Les ARN sont convertis en ADNc double brin puis coupés avec une enzyme de restriction qui coupe souvent. On ligue un adaptateur aux extrémités cohésives ce qui

permet de rajouter des séquences connues aux extrémités de l'ADN qui est alors amplifié par PCR. La séquence des fragments de PCR permet alors d'identifier l'ARN qui s'est accroché à la protéine.

- On incube la protéine avec des oligonucléotides de 70 bp comportant aux extrémités des séquences connues et au centre des séquences aléatoires, dégénérées. On incube la protéine avec les oligonucléotides, puis on immunoprécipite la protéine liée aux oligonucléotides. Les fragments d'ADN sont amplifiés par PCR à l'aide d'oligonucléotides complémentaires aux séquences connues présentes aux extrémités de l'oligonucléotide dégénéré. Le produit d'amplification est alors incubé avec la protéine, immunoprécipité et le ligand est amplifié par PCR. Cette expérience peut être effectuée plusieurs fois de suite pour sélectionner les séquences qui sont les plus affines pour la protéine. A chaque étape, l'affinité peut être estimée par un retard sur gel. Cette méthode (SELEX) permet de trouver les séquences les plus affines qui n'existent pas obligatoirement *in vivo*. Pour trouver la séquence reconnue dans l'organisme, on peut soit cribler une banque avec les oligonucléotides soit comparer la séquence avec une banque de donnée (criblage *in silico*)
- On produit une protéine *in vitro* et on crible une banque génomique. La liaison est révélée en utilisant une protéine marquée radioactivement ou un anticorps.

7) On veut cloner un gène ayant un taux d'expression variable

Les eucaryotes contiennent environ 10.000 gènes différents, mais seulement une petite partie est exprimée dans une cellule donnée à un moment donné. L'expression de certains gènes est variable, elle dépend du développement, de la différenciation, de la réponse à un événement, de l'âge etc..

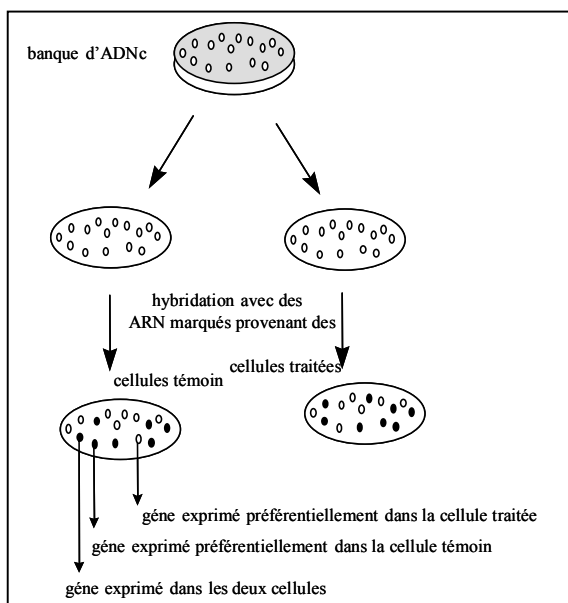
Si on veut cloner de tels gènes, spécifiquement transcrits dans une cellule, on dispose de plusieurs méthodes : protéomique, criblage différentiel, hybridation soustractive et differential display.

Protéomique :

-> Séparation des protéines par électrophorèse bi-dimensionnelle

On peut séparer les protéines de la cellule à l'aide d'une électrophorèse bidimensionnelle (électrophorèse 2D). L'ensemble des protéines d'une cellule est parfois appelé protéome et l'étude est appelée protéomique par analogie avec les études du génome qui sont appelées génomique. On fait l'électrophorèse 2D pour séparer les protéines de deux populations de cellule puis on compare les patterns obtenus. La protéine spécifique sera éluee et on remontera au gène soit à l'aide d'un anticorps soit à l'aide de la séquence. Cette méthode est en voie d'automatisation : Les protéines sont digérées *in situ*, dans le gel par une protéase spécifique – trypsine par exemple -, les peptides sont élués à l'aide d'un laser et séquencés en cours d'éluion en déterminant leurs masses.

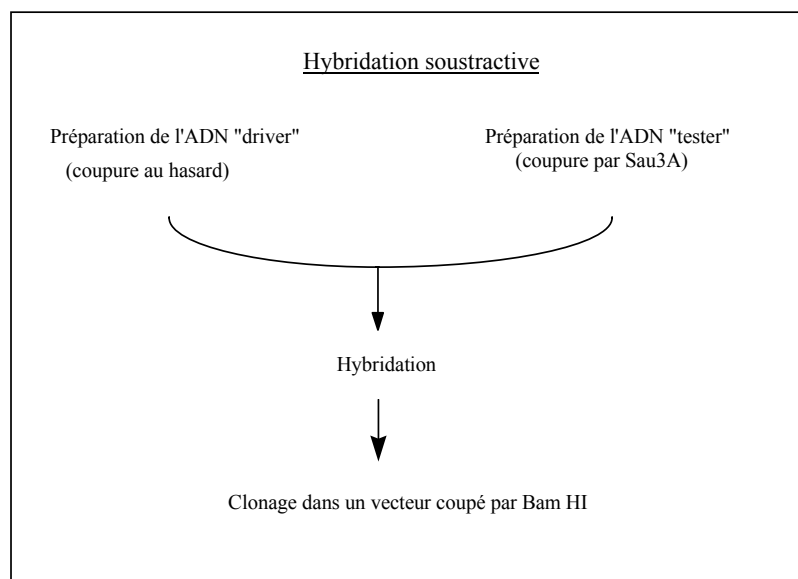
Criblage différentiel



- On peut effectuer un criblage différentiel. On fait une banque d'ADNc et on la crible avec deux sondes, une représentant les ARN de la cellule traitée et l'autre représentant les ARN de la cellule témoin. Les colonies hybridant avec uniquement avec la première sonde, comportent un gène recherché. Le problème de cette technique est qu'on ne peut pas détecter des ARN dont la représentativité est en dessous de 0.1%. La sensibilité de la sonde spécifique peut être

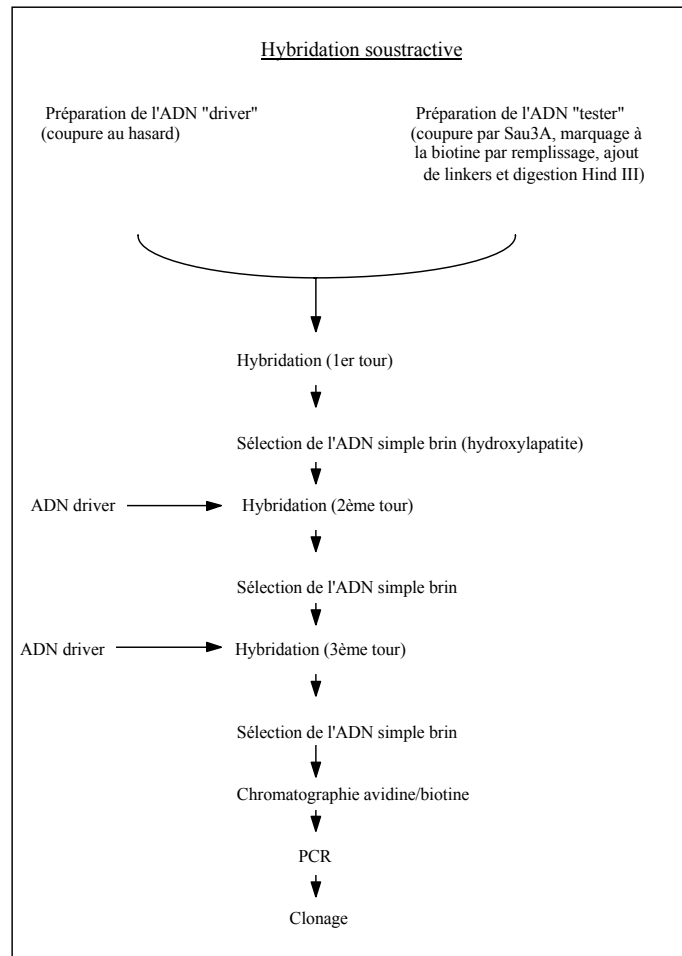
améliorée en enlevant au préalable la plupart des ARN communs aux deux cellules (Scott *et al.*, 1983). Les ADNc de la banque de la cellule témoin sont extraits et couplés sur un support solide tel que de la nitrocellulose. Des ADNc de la cellule d'intérêt, sont préparés, marqués radioactivement et hybridés plusieurs fois consécutives à l'ADN provenant de la banque. On obtient ainsi une sonde plus spécifique des séquences exprimées dans la cellule et non dans le témoin. Cette dernière technique est proche de l'hybridation soustractive.

Hybridation soustractive



On peut faire une hybridation soustractive (Wieland *et al.*, 1990). On fabrique des cDNA des deux types cellulaires et on distingue l'ADN "tester" qui contient la séquence désirée et l'ADN "driver" qui ne contient pas la séquence recherchée. L'ADN "tester" est digéré par Sau3A (/GATC, séquence compatible avec BamHI G/GATCC) et l'ADN driver est coupé au hasard (ultrasons ...). On mélange les deux ADN avec un excès important de l'ADN "driver", l'ADN driver est 200 fois plus concentré que l'ADN "tester". On dénature et on hybride à une température élevée. Les séquences présentes dans l'ADN "tester" qui sont aussi présentes dans l'ADN "driver" vont s'hybrider avec l'ADN "driver" puisque cet ADN driver est présent en excès. Par contre les séquences de l'ADN "tester" qui ne sont pas présentes dans l'ADN "driver" ne peuvent pas s'hybrider avec de l'ADN driver et ne peuvent s'hybrider qu'avec elles même. Dans ce dernier cas, l'hybridation reconstruit des sites de

clonage par Sau3A ou BamHI. Ainsi, seules les hybridations ADN "tester" sur lui-même sont insérables dans un site BamHI. Les hybridations "driver"/"driver" et "driver"/"tester" ne génèrent pas de fragments insérables. Une ligation des produits de l'hybridation avec un vecteur linéarisé par BamHI, permet de ne récupérer que les fragments de cDNA qui sont spécifique de l'ADN "tester".



Cette méthode ne donne que des rendements faibles, en effet, la concentration en ADN "tester" doit être faible pour que les séquences communes s'hybrident préférentiellement avec l'ADN "driver", or l'hybridation étant dépendante de la concentration, elle est faible.

- Pour améliorer le rendement on peut sélectionner les ADN simple brin après hybridation à l'aide d'une colonne d'hydroxylapatite, l'ADN double brin est retenu par la colonne alors que l'ADN simple brin passe à travers. Cet ADN est majoritairement de l'ADN "tester" enrichi en séquences

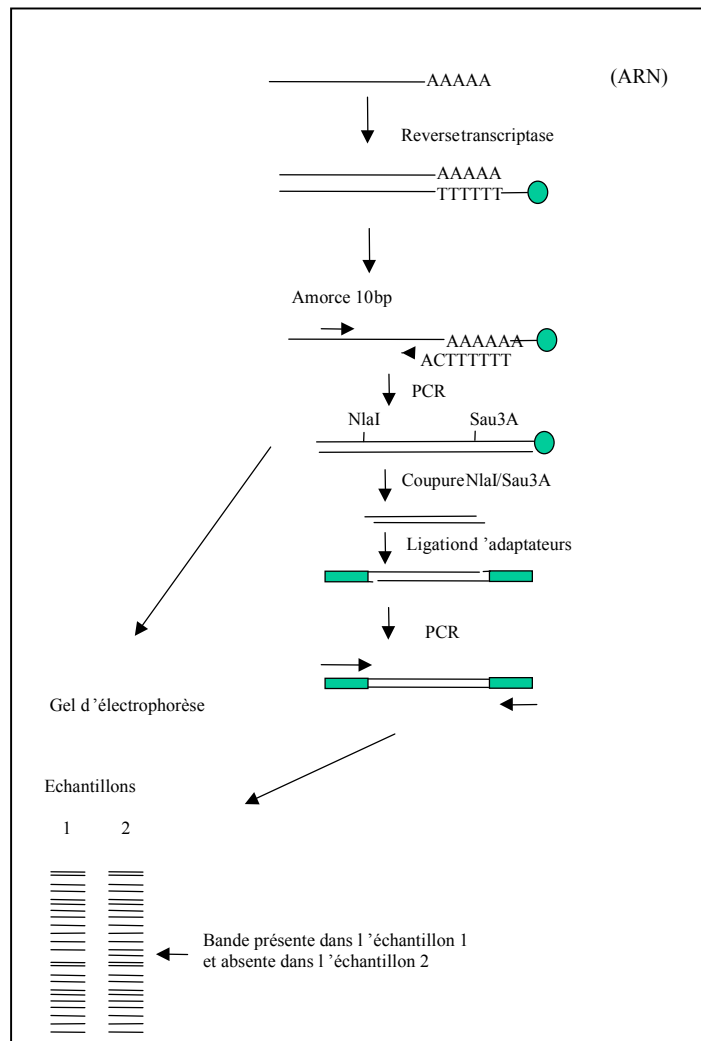
non présentes dans l'ADN "driver". On peut donc rajouter de l'ADN "driver" en excès et recommencer l'hybridation plusieurs fois pour continuer l'enrichissement en séquences "tester" non présentes dans l'ADN "driver". Ainsi si on a 90% des séquences du driver hybridées à chaque tour, après trois tours on aura un enrichissement des séquences de "tester" recherchées de 1000 fois.

- Après ces hybridations successives, on reste contaminé par de l'ADN "driver" simple brin. Pour l'éliminer, on peut au départ marquer l'ADN "tester" par la biotine. Pour ce faire, on fait un remplissage de l'extrémité Sau 3A par des nucléotides dont un est lié à la biotine. Après les hybridations, l'ADN "tester" restant sera retenu sur une colonne de chromatographie utilisant l'avidine. Pour récupérer une extrémité cohésive nécessaire au clonage à la fin de l'expérience, il suffit d'utiliser des linkers aux extrémités.
- Après les hybridations, l'ADN "tester" sélectionné est présent en très faible quantité, trop faible pour pouvoir être cloné. On peut l'amplifier en utilisant des amorces correspondant aux linkers.

Differential display

On peut faire un "differential display" (Liang et Pardee, 1992) ou DDPCR. On commence par amplifier un petit fragment de nombreux ARN des deux échantillons. Les produits d'amplification entre les deux échantillons sont ensuite comparés par migration sur un gel. Si une bande est présente dans un seul échantillon, on peut penser que l'ARN correspondant y est préférentiellement transcrit. La zone du gel est découpée, et le fragment d'ADN est amplifié avec les deux amorces. Il faut donc choisir les amorces de telle façon à ce qu'il n'y ait qu'un seul produit d'amplification par gène.

On effectue tout d'abord une synthèse d'ADNc avec un oligodT, tous les ARN polyadénylés présents dans l'échantillon sont alors convertis en ADNc. Puis on amplifie par PCR une partie des ADNc avec une amorce non



spécifique et un oligodT. De façon à ne pas amplifier tous les ARNm on peut choisir un oligodT terminé en 3' par un G un C ou un A, cet oligonucléotide s'hybridera sur un tiers des ADNc. Habituellement, on utilise un oligodT terminé en 3' par deux bases comme par exemple 5' T₁₁CA 3', cet oligonucléotide s'hybridera sur 1/12 des ARNm (3 bases x 4; un T avant le A donnerait 1/4). Comme oligonucléotide non spécifique en sens, on choisit un petit oligonucléotide de 10 bases. Cet oligonucléotide va s'hybrider sur de nombreux ADNc à une distance de l'autre amorce qui dépend de la séquence. Un desoxynucléotide marqué radioactivement au ³⁵S sera ajouté au milieu réactionnel pour suivre les produits d'amplification qui seront chargés sur un gel de séquence, en condition dénaturante. Deux chargements seront effectués en parallèle, le produit d'amplification des ADNc de

la cellule étudiée et le produit d'amplification de la cellule témoin. Par comparaison des deux lignées, on pourra détecter les amplifications spécifiques de l'échantillon. Le fragment d'intérêt sera alors élué du gel, soumis à une nouvelle amplification avec les mêmes amorces, et cloné.

Pour améliorer la technique on peut couper les produits d'amplification par des enzymes qui coupent souvent, liguer des adaptateurs et amplifier les produits de digestion en choisissant comme amorces les séquences des adaptateurs (Kondoh *et al.*, 1999).

Parmi les problèmes rencontrés, il y a l'amplification entre deux oligonucléotides non spécifiques un en sens et l'autre en antisens. Dans ce cas le differential display se résume à un RAPD qui risque de ne pas être spécifique. Il faudra donc faire l'expérience plusieurs fois pour vérifier que l'amplification résulte bien d'une différence d'expression et non d'un hasard du aux PCR.

Clonage des gènes les plus exprimés dans un type cellulaire, EST.

Chaque laboratoire fait des efforts pour cloner un ou un faible nombre de gènes d'un organisme. Ces efforts peuvent être regroupés pour identifier tous les transcrits d'un même organisme et ainsi obtenir une série de sonde. Ces sondes ont été appelées des EST (Express Sequence Tag, Adams et al., 1991). Des ADNc sont fabriqués par amorçage au hasard puis séquencés. L'amorçage au hasard sert à obtenir des séquences codantes plutôt que les extrémités 3' non codantes. Les gènes sont alors identifiés par comparaison des séquences obtenues avec les séquences de gènes homologues d'autres organismes qui sont présentes dans les banques.

Hybridation sur puce à ADN.

Si on a la séquence complète d'un génome, on peut effectuer une hybridation sur puce (Array hybridization) pour identifier tous les gènes qui sont transcrits à un moment donné, par exemple aux différentes étapes du cycle cellulaire. On fabrique une puce à ADN avec des oligonucléotides s'hybridant sur chacun des gènes. On utilise une trentaine d'oligonucléotides pour un même gène ce qui fait pour la levure *Saccharomyces cerevisiae* 260 000 oligonucléotides pour 6200 gènes. Les cellules sont synchronisées et les ARN sont extraits toutes les 10 min, convertis en ADNc, coupés en morceaux de 50 pb par une DNase puis marqués à une extrémité par la biotine. Les sondes sont hybridées à la puce à ADN et l'hybridation est révélée à l'aide de streptavidine liée à un marqueur

fluorescent. Dans cette technique, plusieurs sondes sont utilisées pour éviter les problèmes de stringence, pour éviter les problèmes de fabrication des ADNc.

Cette méthode peut aussi être utilisée pour cloner des gènes qui sont surexprimés par exemple à la suite d'un traitement par une drogue même si on n'a pas encore la séquence du génome. On hybride les ARN sur une puce composés d'oligonucléotides de séquence au hasard. Si on a une hybridation différentielle, on peut utiliser cet oligonucléotide comme sonde pour rechercher le gène par exemple par PCR.

8) Gènes synthétiques.

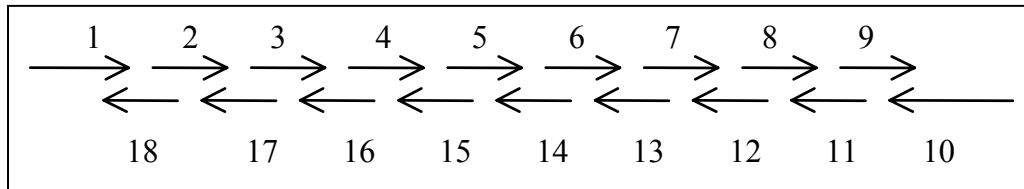
La méthode traditionnelle d'obtention d'un gène consiste à l'isoler à partir d'une source naturelle. Mais on peut aussi le synthétiser chimiquement.

La synthèse chimique présente l'avantage de pouvoir altérer la séquence du gène d'intérêt pour par exemple :

- Optimiser les signaux d'initiation et de terminaison
- Modifier des régions 5' et 3' non traduites par addition d'éléments permettant d'augmenter l'expression ou suppression d'éléments induisant une diminution d'expression.
- Introduire des sites de restriction
- Introduire des caractères spécifiques
- Optimiser l'utilisation des codons
- Supprimer/insérer des domaines protéiques
- Introduire des sites de mutagenèse

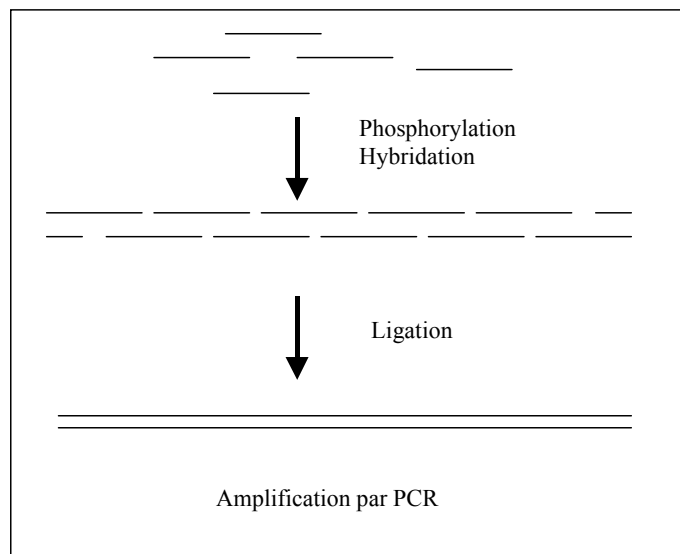
Il existe une limitation, on ne peut pas synthétiser des oligonucléotides de plus de 120 bases en routine. Il va donc falloir les assembler par PCR. Il existe deux méthodes d'assemblage :

1) On synthétise des oligonucléotides se chevauchant et on fait une seule PCR pour remplir les trous (Stemmer et al., 1995 ; Feng *et al.*, 2000)



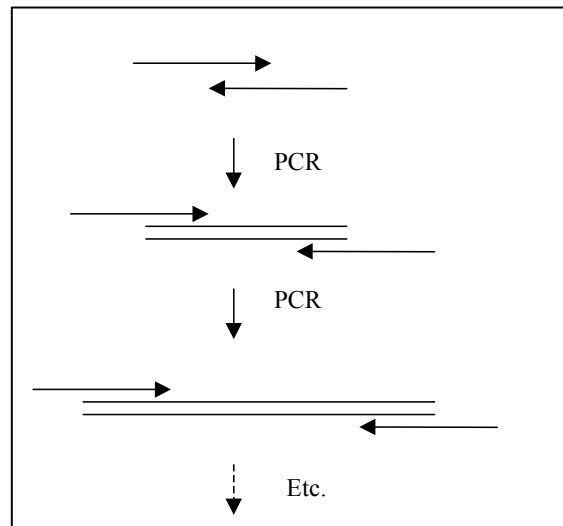
L'ensemble des oligonucléotides sont mis dans un tube mais en faible quantité à l'exception des oligonucléotides 1 et 10.

Pour améliorer le rendement de la méthode, on peut utiliser des oligonucléotides se chevauchant totalement. On les phosphoryle puis on les ligue à l'aide de la Taq DNA ligase de façon à faire plusieurs cycles d'hybridation/ligation et ainsi augmenter les chances d'obtenir le fragment complet. On effectue ensuite une réaction de PCR avec deux amorces situées aux extrémités.



2) On synthétise deux oligonucléotides au centre de la séquence, on remplit par PCR puis on allonge la séquence en ajoutant deux nucléotides aux extrémités de la séquence. La séquence complète est ainsi obtenue par allongements successifs (Majunder, 1992 ; Di Donato et al., 1993).

Si les oligonucléotides sont dégénérés (dans les parties qui ne sont pas hybridées), on obtient une banque de gène synthétique et donc une banque de mutants.

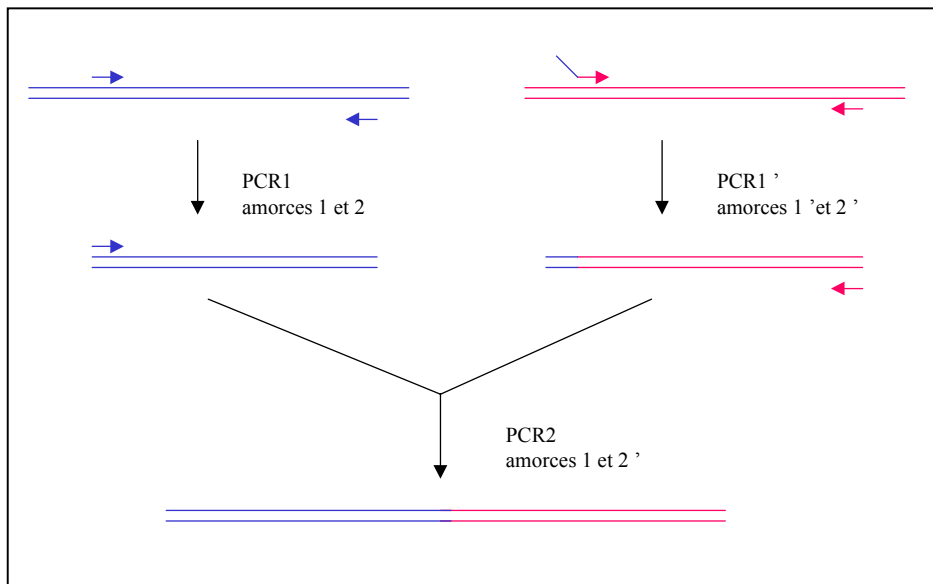


Constructions

Construction sans site de restriction

Par PCR :

La technique la plus habituelle pour relier deux fragments d'ADN est d'effectuer une ligation avec la T4 DNA ligase sur des extrémités compatibles. Cependant, il n'est pas toujours possible d'avoir un site unique entre les deux fragments à réunir. Une autre méthode consiste à faire une PCR de fusion. Les deux fragments sont amplifiés avec une partie commune.



Les deux fragments à relier sont amplifiés indépendamment, avec une partie commune générée en incorporant la séquence d'un fragment dans l'amorce servant à amplifier l'autre fragment. On effectue alors une deuxième PCR avec les deux amorces les plus extrêmes et les produits des deux amplifications précédentes.

Par recombinaison :

Une autre possibilité est d'effectuer une recombinaison *in vivo*, dans la bactérie. Une bactérie contenant le plasmide cible en une seule copie est transformée avec le plasmide donneur portant une

origine de répllication thermosensible. Lorsqu'il y a recombinaison, homologue, il y a inactivation de la β -galactosidase du plasmide cible et les colonies sont blanches en présence de X-gal.

Cette méthode est principalement utilisée pour insérer des gènes dans un vecteur de grande taille. On clone tout d'abord le gène d'intérêt dans un petit plasmide comportant de chaque côté les séquences bordantes du plasmide cible pour permettre la recombinaison homologue.

Cette technique est aussi utilisée en cellule eucaryote, par exemple pour fabriquer des baculovirus recombinants, on transfecte des cellules avec le virus et avec un plasmide comportant le gène d'intérêt flanqué de séquence du baculovirus.

Insertion dans un plasmide d'un fragment d'ADN obtenu par PCR à l'aide de la Taq DNA polymérase.

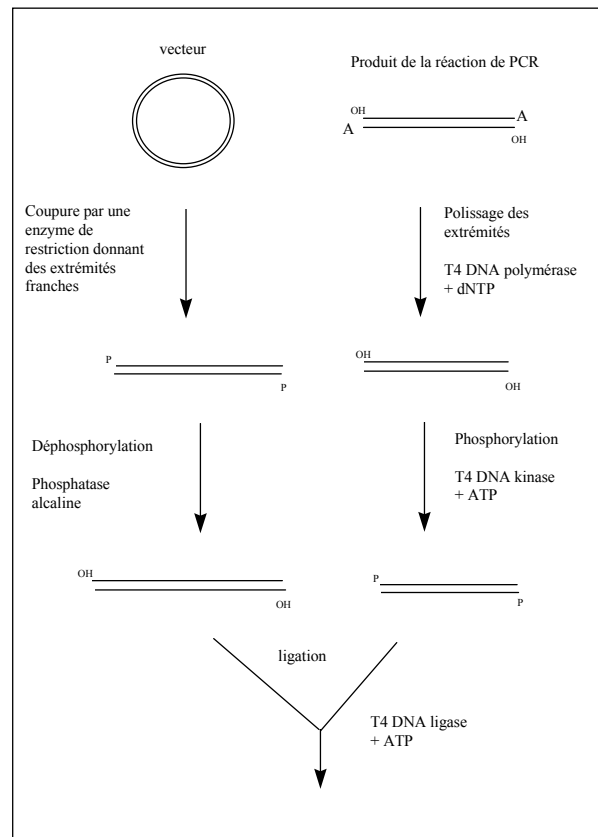
L'ADN polymérase Taq, comme toutes les ADN polymérases, a une activité terminal-transférase sur l'ADN double brin et la base qui est le plus souvent utilisée est l'adénine. Si bien que les fragments obtenus ont généralement un A en extrémité 3' sortante (Clark, 1988). Cette activité est assez générale chez les ADN polymérases mais est souvent masquée par l'activité 3'5' exonucléase qui est absente dans le cas de l'ADN polymérase Taq.

Il existe plusieurs solutions pour cloner de tels fragments.

1) Retirer le A du fragment amplifié

On peut le faire en utilisant l'activité 3' 5' exonucléase d'une DNA polymérase. Généralement on utilise la T4 DNA polymérase en présence des quatre nucléotides triphosphate.

Le fragment obtenu pourra alors être inséré dans un vecteur ayant des extrémités franches. Comme les amorces n'ont pas de groupement phosphate en 5', il faudra les ajouter si on veut utiliser un vecteur déphosphorylé.



2) Fabriquer des extrémités compatibles.

a) En ajoutant des sites de restriction sur les amorces

On ajoute une séquence en amont (en 5') des amorces, cette séquence ne s'hybride pas sur la séquence cible mais comporte un site de restriction. A la fin de la PCR, le fragment est digéré par les enzymes de restriction et inséré dans le vecteur approprié. Il faut faire attention de rajouter quelques bases en 5' en amont du site, en effet, certaines enzymes ont du mal à digérer lorsque le site se trouve trop près du bord. Par exemple Hind III ne coupe pas s'il y a une ou deux bases en amont de la séquence A/AGCTT, et coupe avec une faible efficacité (10%) s'il y a trois bases. Cette inhibition de la coupure est très variable selon les DNAses, Eco RI coupe à plus de 90% même s'il n'y a qu'une seule base en 5' du site.

b) A l'aide d'une exonucléase

On peut par exemple ajouter la séquence AATTC en amont des deux amorces. Après PCR on aura la séquence :

AATTCNNNNNNNNGAATTA
 ATTAAGNNNNNNNNCTTAA

Si on fait agir une ADN polymérase T4 en présence de GTP uniquement on aura :

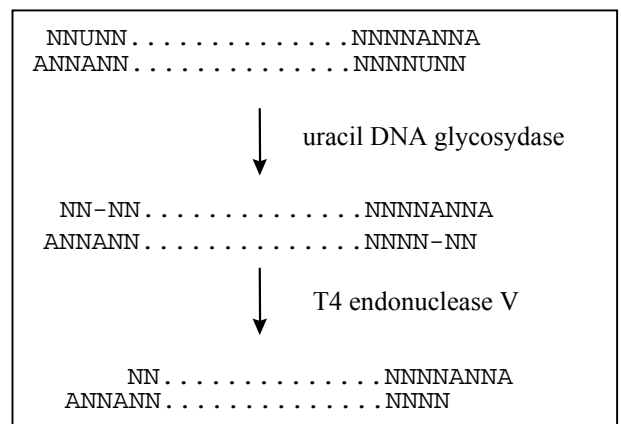
AATTCNNNNNNNNG
 GNNNNNNNNCTTAA

Ce fragment peut être ligué dans un vecteur coupé par Eco RI après une phosphorylation si le vecteur est déphosphorylé.

Si on veut faire un clonage orienté, on peut utiliser deux sites différents, par exemple Xma I (C/CCGGG) et EagI (C/GGCCG). On ajoutera donc les deux extensions suivantes sur chacune des deux amorces : CCGGT pour le site XmaI et GGCCT pour le site EagI, les deux sites seront générés par incubation avec l'ADN polymérase T4 en présence d'ATP.

c) A l'aide de l'uracil DNA glycosydase

On ajoute un uracil en 5' dans l'amorce (Rashtchian *et al.*, 1992). Après la réaction de PCR, on fait agir l'uracil DNA glycosidase. Cette enzyme clive la liaison base-sucre ce qui laisse un site apurinique/apyrimidique. On fait ensuite agir la T4 endonuclease V. Cette enzyme a deux activités, elle catalyse l'excision des dimères de pyrimidines (cette activité n'est pas utilisée ici) et elle clive le site apurinique/apyrimidique pour laisser un 5' phosphate (Watson et Bennett, 1997).



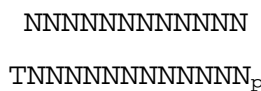
Dans ce cas on obtient des extrémités 3' sortantes qui peuvent être clonés dans un vecteur coupé par une enzyme donnant des extrémités compatibles.

4) Utiliser un vecteur linéarisé ayant un T en 3'

Dans ce cas on peut augmenter la proportion de fragment avec un A en 3' en faisant une incubation longue (15 min) à la fin de la dernière polymérisation, au dernier cycle de la réaction de PCR. De plus on peut faire une incubation du fragment à amplifier en présence d'ADN polymérase Taq et uniquement de 3mM dATP (Bielefeldt-Ohmann et Fitzpatrick, 1997)

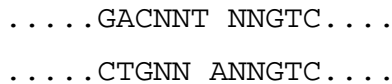
Mais il faut aussi obtenir un vecteur avec un T sortant en 3'

a) On peut liquer un adaptateur sur des extrémités franches

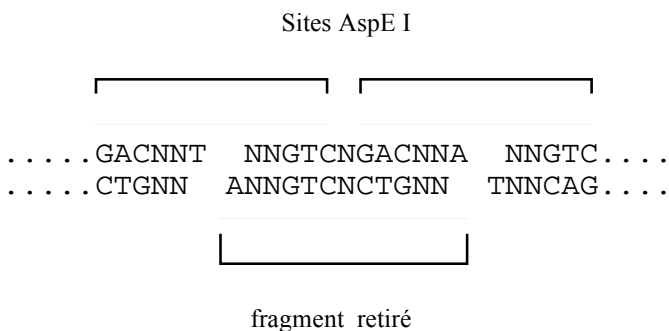


b) On peut utiliser une polymérase sur des extrémités franches. Soit, on fait agir une terminal transférase en présence de didesoxy thymidine triphosphate (ddTTP) ou de la Taq DNA polymérase en présence de dTTP uniquement.

c) On peut générer un T sortant en 3' par une coupure à l'aide d'un site de restriction (Cha *et al.*, 1993). Pour ce faire on peut utiliser plusieurs enzymes comme AspE I (GACNNN/NGGTC). Dans la séquence, le nucléotide juste avant la coupure doit être un T ce qui donne comme extrémités :



On obtient un T d'un coté mais un A de l'autre. Pour résoudre ce problème on met deux sites adjacents sur le vecteur :



On peut utiliser d'autres enzymes tels que Xcm I (CCAN₅/N₄TTG).

5) Effectuer une hybridation

On peut ajouter une queue homopolymérique en 3' avec une terminal transférase, dans ce cas on doit utiliser un vecteur dans lequel on a ajouté une queue homopolymérique avec le nucléotide complémentaire. La queue homopolymère peut être rajoutée à l'aide de la terminal-transférase.

On peut ajouter en 3' des amorces, une séquence dépourvue de A. Après action d'une ADN polymérase T4 en présence uniquement de dATP, on aura une extrémité 3' sortante. Il faut que le Tm de cette extrémité soit supérieur à 40°C pour ne pas se dehybrider lors de la transformation.

Exemple :

On coupe un vecteur par Sma I (GGG/CCC)

```
5' ..NNACTGGTTCCGGGG CCCGCCGGAGCGGAGTN...
   ..NNTGACCAAGGCCCC GGGCGGCCTCGCCTCAN... 5'
```

on incube avec de l'ADN polymérase T4 en présence de dATP uniquement

```
5' ..NNA          CCCGCCGGAGCGGAGTN...
   ..NNTGACCAAGGCCCC          AN... 5'
```

Pour avoir des extrémités longues, il faut trouver ou fabriquer un vecteur avec un A éloigné en 5' d'un site de restriction.

PCR en présence des amorces : avec la séquence complémentaire suivie d'un A et de la séquence spécifique

```
5' CTGGTTCCGGGGANNNNNN-amorce spécifique
   et amorce spécifique-NNNNAGGGCGGCCTCGCCTC 5'
```

```
5' CTGGTTCCGGGGANNNNNN gène d'intérêt NNNTCCCGCCGGAGCGGAGA
   AGACCAAGGCCCTNNNNNN..... NNNAGGGCGGCCTCGCCTC 5'
```


Incubation en présence de T4 DNA polymérase et de dTTP uniquement

5' CTGGTTCGGGGANNNNNN gène d'intérêt NNNT
 TNNNNNN..... NNNAGGGCGGCCTCGCCTC 5'

Il existe d'autres méthodes pour fabriquer des extrémités hybridantes. Par exemple en utilisant dans les amorces un pont phosphorothioate. On digère le fragment de PCR par l'exonucléase du gène 6 du phage T7 qui digère de 5' vers 3' jusqu'au pont phosphorothioate ou elle est inefficace (Zhou et Hatanet, 1995).

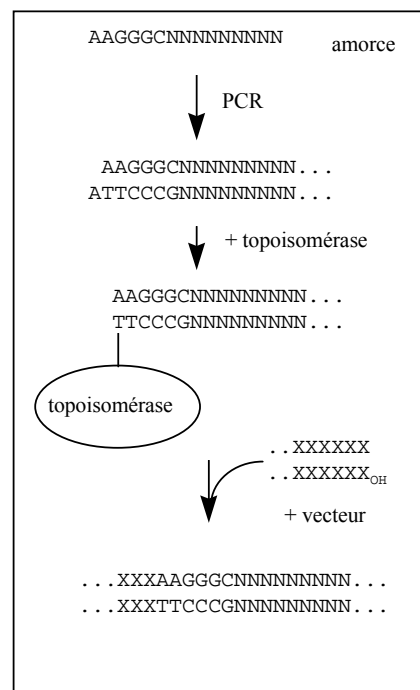
6) Utiliser la topoisomérase du virus de la vaccine

La topoisomérase du virus de la vaccine coupe en 3' de la séquence (CouT)CCTT. Elle peut activer ces séquences et les transférer sur n'importe quelle séquence à condition qu'elle soit 5'OH.

Cette propriété peut être utilisée pour tous les clonages (Shuman, 1994).

Si on ajoute en 5' des deux amorces les séquences AAGGG on obtiendra un site de coupure de la topoisomérase. La topoisomérase reste alors sur le fragment d'ADN et est capable de refermer l'ADN à condition que l'accepteur soit 5'OH. On prépare donc un vecteur à extrémités franches déphosphorylées. L'incubation du vecteur avec la topoisomérase coupe après le T et permet d'accepter le vecteur.

Pour favoriser la ligation, on peut ajouter une extrémité cohésive. Par exemple si on utilise l'amorce TTAAAAGGGNNNN. On obtiendra une extrémité compatible avec un fragment EcoRI.



La construction peut donc se faire en utilisant trois

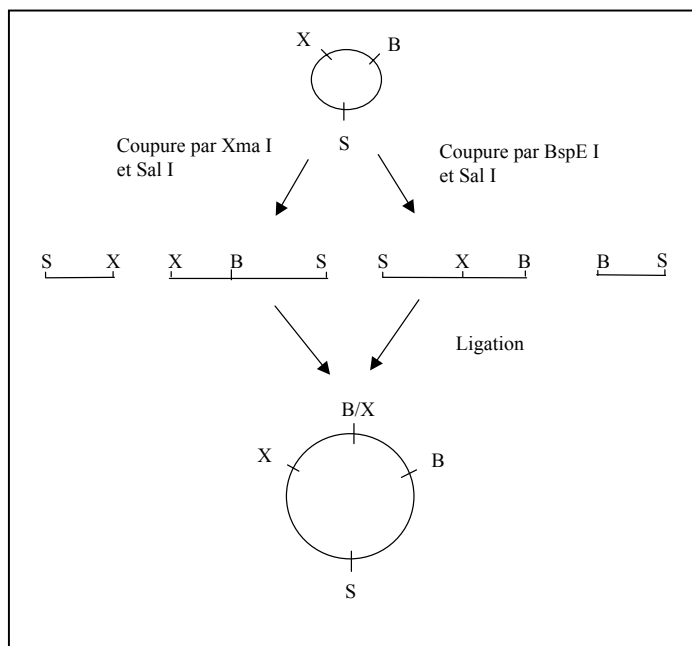
grandes méthodes.

- 1) On utilise une ligation en présence de T4 DNA ligase et de l'ATP. Les extrémités doivent être compatibles et au moins une des deux extrémités 5', sur le vecteur ou sur l'insert, doit être phosphorylé.
- 2) On utilise une hybridation. Dans ce cas, les extrémités doivent être compatibles et assez longue, au moins d'une dizaine de nucléotides pour permettre l'hybridation à 37°C, température de la transformation.
- 3) On utilise la topoisomérase du virus de la vaccine. Le vecteur ou l'insert doit comporter une séquence de reconnaissance du site de l'enzyme.

Construction d'unités répétées dans un plasmide

On a un plasmide avec un insert et on veut que cet insert soit présent en plusieurs copies en tandem. Une méthode a été développée par Lewis et al. 1996.

On place de chaque côté de l'insert deux sites de restriction compatibles dont le produit de ligation ne peut pas être coupé par les deux enzymes, on peut par exemple utiliser Xma I (C/CCGGG) et BspE I (T/CCGGA), le produit de ligation des deux sites CCCGGA ne peut pas être coupé ni par Xma I ni par BspE I. On utilise un autre site de restriction dans le plasmide par exemple Sal I.



On coupe le plasmide par Sal I et Xma I et on récupère le fragment de restriction portant l'insert à amplifier, on coupe le plasmide par Sal I BspEI et on récupère le fragment de restriction portant le fragment à amplifier.

Les deux fragments de restriction sont ligués, on obtient un plasmide avec deux répétitions.

Pour obtenir plus de répétitions on recommence la même expérience avec deux plasmides portant deux répétitions. La combinaison des différents plasmides

permet d'obtenir le nombre de combinaisons désiré.

La recircularisation du vecteur

Lorsqu'on veut cloner un fragment d'ADN, le vecteur peut se recirculariser. Pour éviter cette recircularisation, on peut :

- Déphosphoryler le vecteur coupé à l'aide d'une phosphatase
- Effectuer une double digestion avec des sites non compatibles

- Effectuer une coupure après la ligation. On ajoute lors de la ligation l'enzyme de restriction qui a servi à couper le vecteur. Il faut dans ce cas que le clonage ne restitue pas le site de restriction (Bielefeldt-Ohmann et Fitzpatrick, 1997). Cette méthode est employée par exemple pour les clonages de fragments à extrémités franches obtenus par PCR à l'aide d'une polymérase ayant une activité 3'5'exonucléase (activité de correction).

Tri des recombinants

1) Repérage des recombinants

Par sélection négative.

- Le gène est inséré dans un gène conférant une résistance à un antibiotique. Les bactéries sensibles à cet antibiotique comportent donc l'insert. L'inconvénient de cette méthode est qu'il faut répliquer toutes les colonies pour tester leur sensibilité à l'antibiotique. Cette méthode est historique et n'est plus employée.

Par une activité enzymatique

Le fragment à cloner est inséré dans un marqueur.

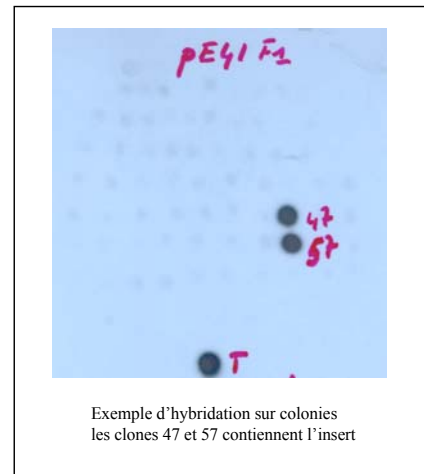
- Le plus couramment utilisé est le peptide α de la β -galactosidase en utilisant une bactérie déficiente pour la synthèse de ce peptide. Les recombinants ne fabriquent plus de β -galactosidase et peuvent être visualisés par son activité. Cette activité peut être visualisée soit en utilisant un substrat chromogène, le X-Gal soit en utilisant un milieu qui réagit au pH, le Mac-Conkey.
- Une autre possibilité est d'utiliser la GFP (Green Fluorescent Protein) qui permet de détecter les colonies par fluorescence.
- L'endoglucanase A (ou cellulase A) catalyse l'hydrolyse des liaisons glycosidiques de la cellulose. La Cellulase forme un cercle blanc autour de la colonie sur les plaques d'agar contenant le substrat chromogène, « Cellomix » (Song et al., 2002).

Par hybridation

- Le tri peut se faire par hybridation, les colonies sont transférées sur nitrocellulose et hybridées avec le fragment d'ADN à insérer.

Par PCR

- Le tri peut se faire par PCR en utilisant deux oligonucléotides. Le plus sûr est souvent d'utiliser un oligonucléotide hybridant sur le vecteur et un oligonucléotide hybridant sur l'insert. La réaction de PCR peut se faire directement sur les colonies, et dans un premier temps, elles peuvent même être groupées.



Par cartographie

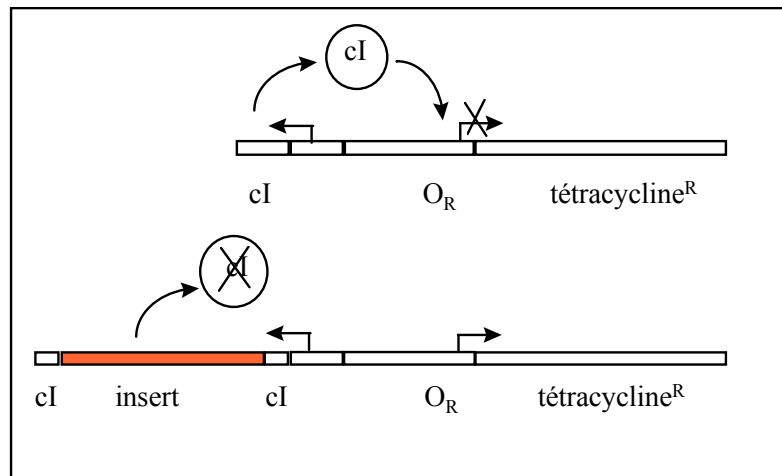
- Le tri peut se faire par cartographie. L'ADN plasmidique de chaque colonie est purifié puis digéré par des enzymes de restriction. Cette dernière méthode est généralement utilisée pour confirmer le tri effectué par une des méthodes précédentes.

2) Sélection positive

Dans ce cas les vecteurs non recombinants ne sont pas amplifiés.

On peut insérer le fragment d'ADN dans un gène codant pour une protéine toxique. Les non recombinants produisent la protéine toxique et ne poussent pas, les recombinants ne produisent pas la protéine toxique et poussent. Pour faire cette sélection positive, on peut utiliser le gène *ccd B* du plasmide F qui code pour une protéine qui inhibe la gyrase. Pour préparer le vecteur seul, on utilise une souche de bactérie Ccd B résistante qui a une gyrase mutée (Bernard, 1995).

On peut insérer le fragment d'ADN dans un éléments régulateur de la transcription d'un gène de résistance à un antibiotique. On utilise le répresseur cI du phage λ , ce répresseur se fixe sur un opérateur (O_R du gène *cro*) en amont du gène de résistance à la tétracycline. Lorsqu'un insert est présent au milieu de cI, le



répresseur n'est plus produit, le gène de résistance à la tétracycline est transcrit, la bactérie est résistante.

Sous clonage par des techniques génétiques.

Les clonages génétiques peuvent être utilisés pour faire passer un insert d'un phage λ à un plasmide.

Les banques sont souvent faites dans le phage λ pour deux raisons

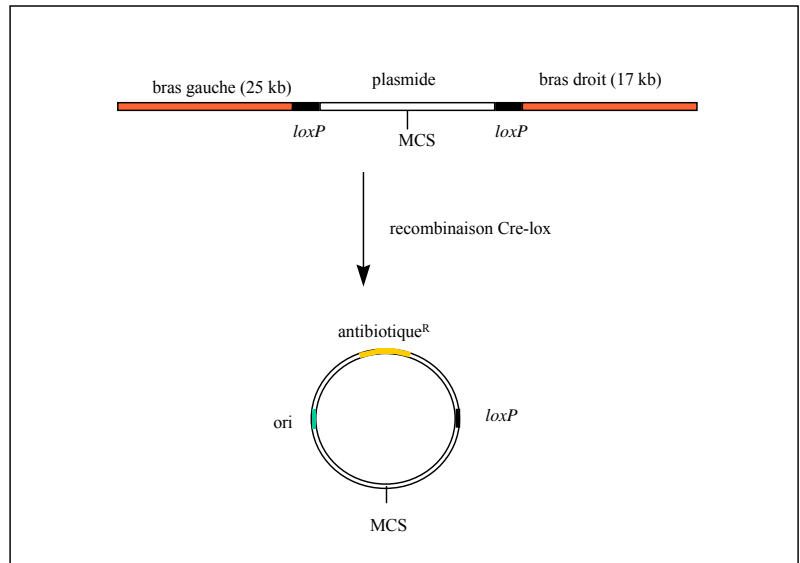
- l'empaquetage *in vitro* donne entre 10^8 et 10^9 clones par μg d'ADN ce qui permet d'obtenir beaucoup de clones avec peu de matériel de départ.
- les plages de lyse sont souvent plus faciles à étaler et à cribler que les colonies bactériennes.

Toutefois, la préparation d'ADN à partir d'un clone est plus longue avec un phage qu'avec un plasmide. Aussi, lorsqu'un clone d'ADNc est obtenu en λ , on s'empresse de récupérer l'insert et de le sous cloner dans un plasmide. Pour éviter cette étape de sous-clonage (digestion, récupération de l'insert, ligation dans un plasmide) on peut utiliser des phages qui ont déjà un plasmide cloné entre les deux bras.

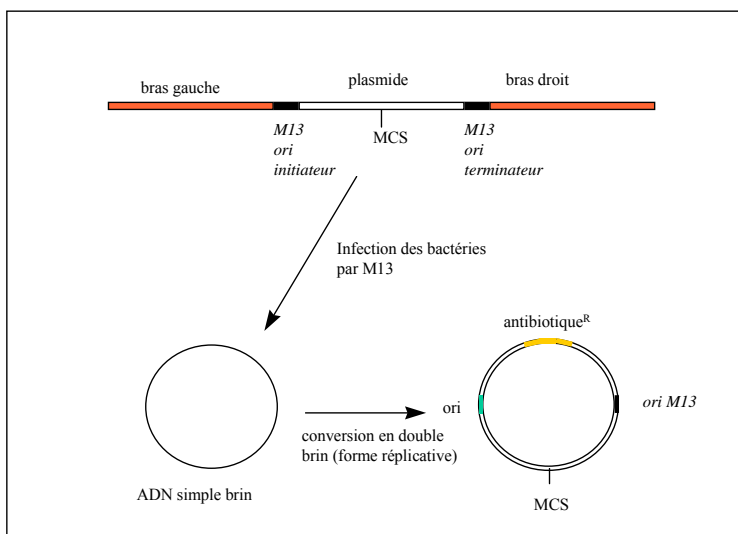
Le clonage génétique peut être aussi utilisé pour l'expression, pour faire passer un insert d'un plasmide dans un autre plasmide pour, par exemple, changer de système d'expression.

Recombinaison

La recombinaison est utilisée par exemple dans le cas de λ TripExc2. Le plasmide est borné par les séquences de *loxP*. La transduction du lysat dans la souche de *E. coli* BM25.8 induit la circularisation du plasmide via la recombinase *Cre*. Le plasmide se retrouve dans les bactéries qui sont triées grâce à la résistance aux antibiotiques.



Utilisation de M13



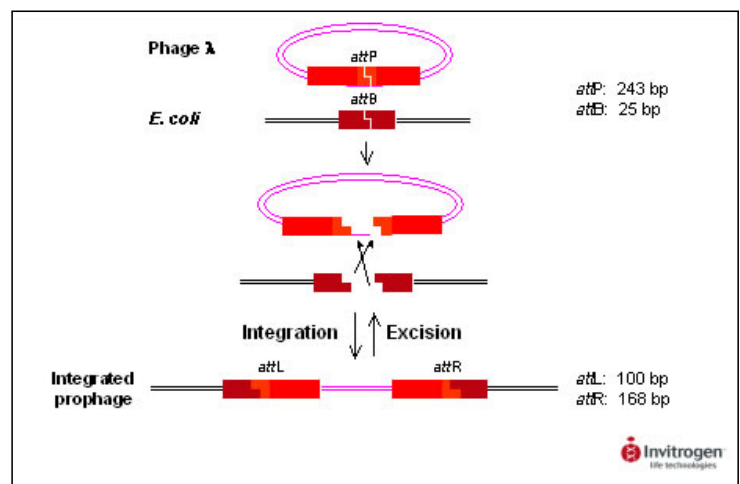
Cette technique a été développée par Short *et al.* (1988) qui ont construit le vecteur λ ZAP. Le plasmide à exciser est cloné entre l'initiation et la terminaison de l'origine de réplication du phage M13. Les bactéries sont infectées par un phage M13 "helper". Il y a formation d'un ADN simple brin circulaire qui est converti en ADN double brin dans la bactérie.

Le phage helper est ensuite éliminé par

traitement thermique si on utilise un phage helper possédant une mutation conditionnelle, ou en purifiant l'ADN double brin et en transformant des bactéries.

Utilisation du site d'attachement de lambda.

Chez Lambda, le site d'intégration est appelé *attP*. Chez *E. coli* le site d'intégration de lambda est appelé *attB*. C'est une séquence courte de 25 bp. Ces sites att contiennent les sites de liaison pour les protéines responsables de la recombinaison, une intégrase et un facteur d'intégration. Lorsque l'intégration a eut lieu, deux nouveaux sites sont créés *attL* et *attR*.

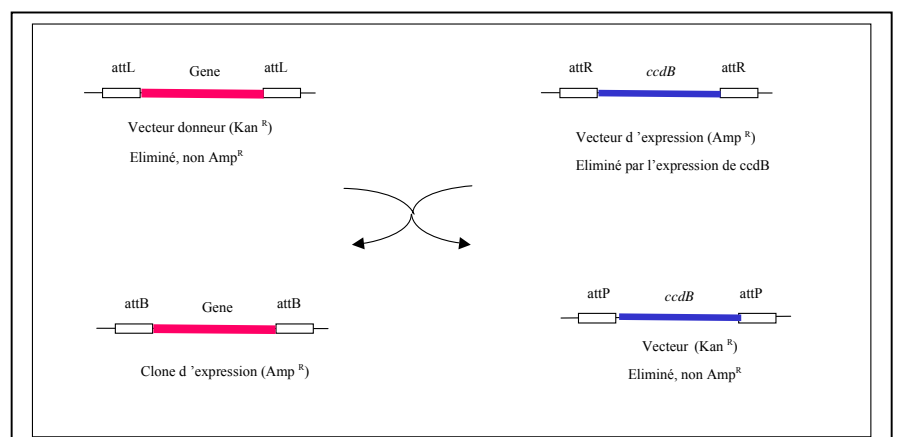


La réaction est réversible, elle implique les mêmes protéines: intégrase (Int), un facteur d'intégration (IHF) et une excisionase (Xis).

L'insert est cloné entre les deux sites *attL* (*attL1* et *attL2*) et on veut le transférer dans un autre vecteur comme un vecteur d'expression. Dans le vecteur d'expression, on a deux sites *attR1* et *attR2*.

On incube les deux plasmides avec les enzymes responsables

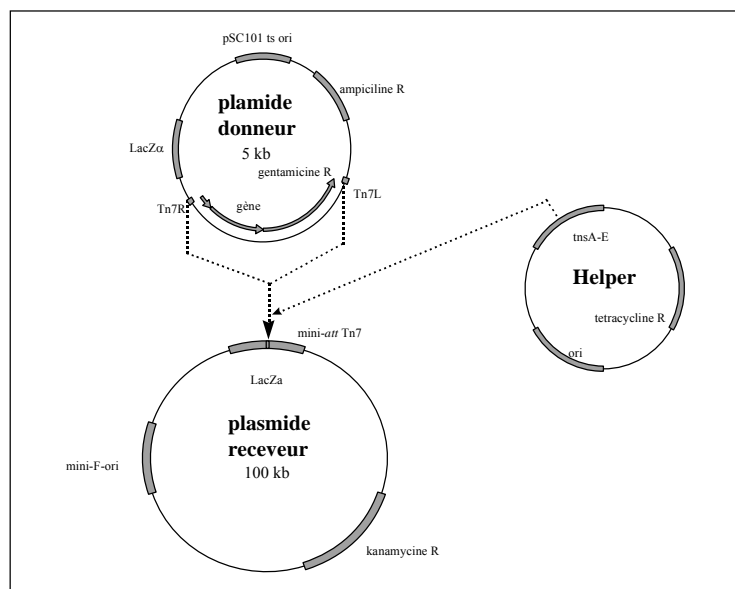
de la recombinaison (Int, IHF et Xis) il y a recombinaison entre les site *attL* et *attR*.



Pour sélectionner les événements de recombinaison, on transforme des bactéries et on étale sur un milieu sélectionnant le plasmide receveur à l'aide d'un antibiotique. Le plasmide donneur est éliminé. Pour éliminer le plasmide receveur n'ayant pas reçu le gène d'intérêt, on met un gène de sélection positive (ccdB) entre les deux sites du vecteur accepteur. Seul le plasmide d'expression recombinant. Ce système est commercialisé par Invitrogen sous le nom de Gateway[®].

Transposition

Cette méthode est par exemple utilisée pour insérer des inserts dans de très grands plasmides. En effet la construction est difficile à effectuer *in vitro*, les sites de restriction uniques sur 100 kb sont rares et la manipulation de grands fragments d'ADN sans les casser est difficile.



On utilise une bactérie dans laquelle il y a un plasmide codant pour une transposase. On transforme cette bactérie par le plasmide receveur. Ce plasmide contient les séquences de réception de l'insert. Ce plasmide contient une origine de réplication et un gène de sélection différent de ceux présents dans le plasmide codant pour la transposase. Dans les bactéries contenant les deux plasmides, on transforme le plasmide donneur.

La transposition est repérée en cotransposant un gène de résistance et en inactivant un gène LacZ. Le plasmide donneur est éliminé en utilisant une origine de réplication thermosensible.

Analyse de la transcription

1 - Détermination du site d'initiation de la transcription

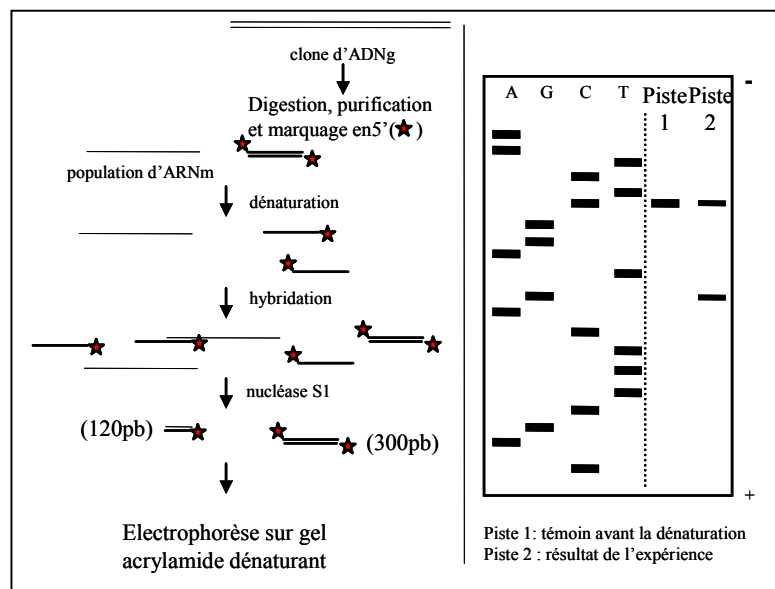
Le site d'initiation de la transcription (*tsp-transcription start point* noté +1) correspond au premier nucléotide situé à l'extrémité 5' de l'ARN et ne peut être déterminé qu'expérimentalement. La molécule permettant de déterminer ce *tsp* est donc l'ARN et non le gène lui-même. Dans le cas des ARNm il s'agit bien sûr de l'ARNm non coiffé. Ce nucléotide est une purine et, dans 80% des cas, il s'agit d'une adénine.

Il est nécessaire de disposer d'un gène cloné ou d'un clone contenant au moins la jonction entre le promoteur et le 1^{er} exon, pour déterminer le *tsp*.

Cartographie à la nucléase S1.

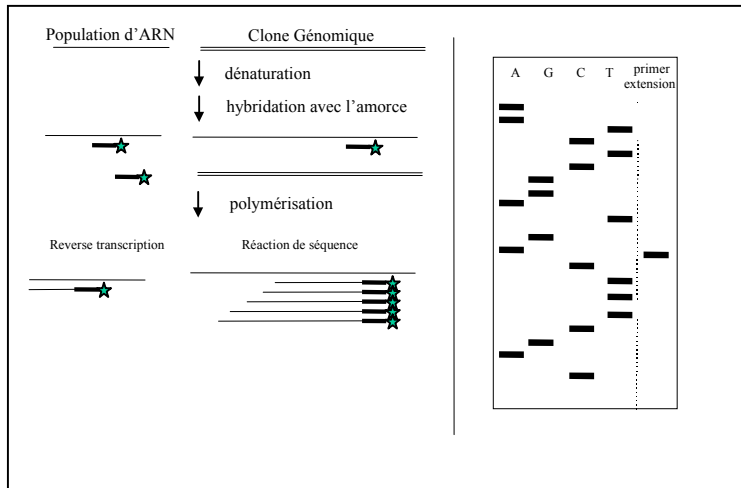
(Cette méthode permet également de déterminer la structure générale d'un gène : voir DEUG).

Le principe de cette technique consiste à hybrider un fragment d'ADNg marqué, que l'on suppose recouvrir le point +1, avec une population de messagers. L'ARNm correspondant au gène dont est issue la sonde, va protéger une partie de cette sonde contre la digestion à la nucléase S1. Les résultats sont ensuite analysés sur gel et on mesure la diminution de taille de la sonde au nucléotide près. On fait donc migrer



en parallèle une réaction de séquence selon Sanger qui sert de marqueur de taille.

Extension d'amorce.

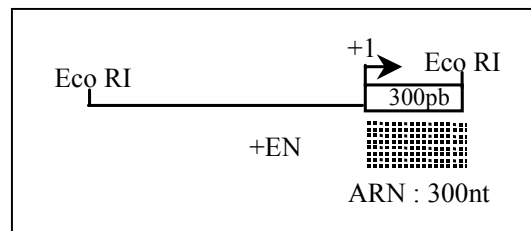


Dans cette expérience on choisit d'allonger la sonde plutôt que de la dégrader comme dans l'expérience précédente. On réalise en fait une reverse-transcription avec une amorce spécifique issue de l'ADNg. Cette amorce correspond à une région située en aval du point +1. On mesure ensuite sur gel polyacrylamide dénaturant l'augmentation de la taille de l'amorce

par comparaison avec une réaction de séquençage selon Sanger réalisée avec la même amorce sur l'ADNg.

Transcription *in vitro* (type Run off)

La transcription *in vitro* permet en premier lieu de tester des extraits nucléaires (EN) pour vérifier leur activité fonctionnelle au niveau de l'initiation de la transcription. Elle permet également de déterminer le site d'initiation de la transcription. On utilise un clone linéarisé qui est transcrit en présence d'EN et de $\text{CTP}\alpha^{32}\text{P}$. Les résultats sont analysés sur gel dénaturant sur lequel on détermine la taille de l'ARN obtenu et ainsi la position du tsp.



Cette expérience permet en outre de savoir si la région promotrice est complète. Cependant il est difficile de mettre en évidence une régulation par cette méthode. Cette absence de régulation est souvent liée à la perte ou la modification de facteurs transcriptionnels lors de la préparation des EN ainsi qu'à la conformation de l'ADN *in vitro*.

Cette méthode n'est donc pas adaptée, sauf dans de rares cas, à l'étude fonctionnelle de la régulation de la transcription. Si on veut déterminer fonctionnellement le rôle d'un facteur régulateur ou d'une région promotrice particulière on utilise un système de cellules en culture.

2 - Quantification d'un transcrit

L'abondance d'un transcrit dépend de la transcription et de sa stabilité, c'est à dire de l'équilibre entre l'étape de synthèse et l'étape de dégradation. Plusieurs méthodes ont été développées pour estimer l'abondance d'un transcrit.

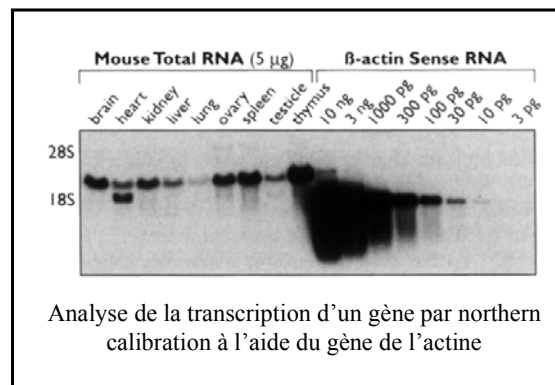
Northern

La première méthode est le Northern. Cette méthode présente plusieurs avantages : i) elle permet une comparaison directe de l'abondance d'un transcrit dans plusieurs échantillons sur un seul filtre, ii) elle permet de déterminer simultanément la taille d'un transcrit et la présence d'épissages alternatifs.

Pour estimer la quantité d'un transcrit, on doit tout d'abord normaliser la quantité d'ARNm présente dans les différentes pistes. On hybride le filtre avec deux sondes, la sonde du gène d'intérêt et une sonde qui hybride avec un ARN dont la quantité est invariante d'une cellule à

une autre. Cette sonde invariante est par exemple la β -actine qui est transcrite d'une manière homogène dans toutes les cellules. Cependant les sondes invariantes sont loin d'être parfaites. L'ARN ribosomique peut être utilisé lorsque aucune autre sonde n'est disponible.

Inconvénients de la méthode : i) elle est peu sensible et est donc difficile à utiliser pour les transcrits rares, ii) de plus elle est très sensible à la dégradation de l'ARN. Chaque coupure dans l'ARN entraîne une perte d'information puisque la molécule ne migre plus à la bonne place.



Protection à la RNase

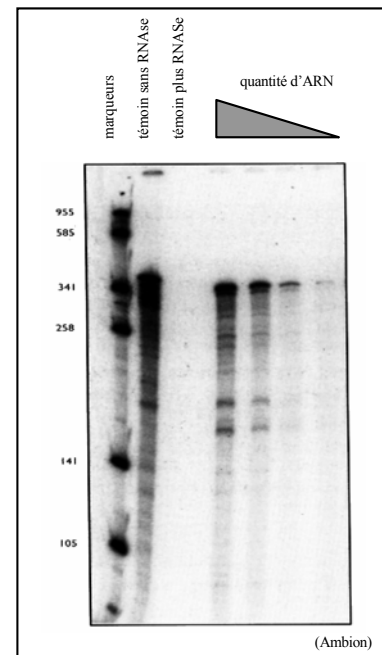
Une deuxième méthode est la protection à la RNase (ribonuclease protection assay, RPA). On synthétise *in vitro* un ARN antisens marqué. Cet ARN est hybridé avec la solution d'ARNm et l'ensemble est digéré par la RNase A. Le produit est ensuite chargé sur un gel et on quantifie l'ARN antisens protégé.

Avantages: Cette méthode est 10 à 100 fois plus sensible que le Northern. En effet on peu utiliser une quantité importante au départ (par exemple 100 µg alors que pour le Northern on est limité par la quantité d'ARN qui peut être déposé sur un gel - 20 à 30µg)

La méthode est plus tolérante à la dégradation de l'ARN puisqu'on n'en analyse qu'une petite partie.

On peut quantifier plusieurs ARN simultanément.

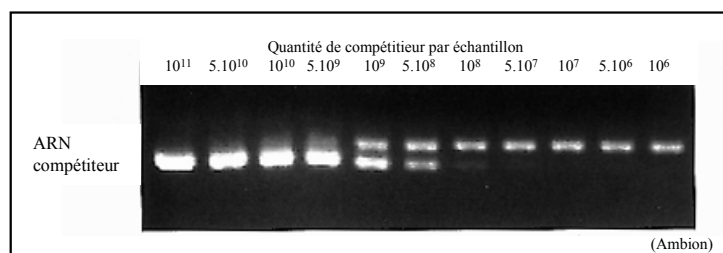
Inconvénients : il ne faut pas qu'il y ait un polymorphisme sur l'ARN, il faut que la sonde corresponde exactement à l'ARN, en effet un mésappariement entraînerait une coupure de la sonde.



RT-PCR

Une troisième méthode est la RT-PCR. On peut distinguer la RT-PCR compétitive de la RT-PCR relative.

La RT-PCR compétitive utilise une référence interne pour normaliser la quantité d'ADNc utilisée dans la réaction. La normalisation est encore plus difficile à faire qu'en Northern. En effet les conditions d'amplification peuvent varier d'une manière indépendante entre l'amplification de la référence et l'amplification du gène d'intérêt.



La technique de normalisation a consisté à utiliser comme référence un ARNc (ARN synthétique, fabriqué *in vitro*) mais de taille légèrement différente de la taille de l'ARN " naturel ". Cette méthode a été appelée quantitative compétitive RT-PCR (Wang et al., 1989).

Avantage, cette méthode est très sensible et est principalement employée pour les transcrits rares.

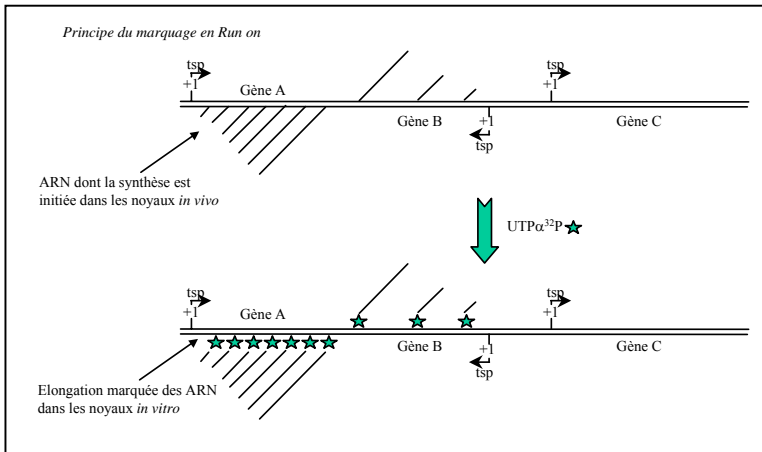
Inconvénient : cette méthode est lourde si on veut comparer un grand nombre d'échantillons, en effet il faut un gel (ou plusieurs pistes) par échantillon. Comme pour toutes les méthodes utilisant la PCR, la sensibilité de la méthode la rend aussi très sensible aux pollutions de l'échantillon par des ADN (ces ADN polluant les réactions proviennent d'amplification faites avant ou sur des clones présents dans le laboratoire, ils sont présents en très faible quantité dans les pipettes ou dans les solutions utilisées). Un autre inconvénient est la formation d'hétéroduplex, produits de l'hybridation entre la séquence dosée et la séquence du compétiteur. Cet hétéroduplex forme une troisième bande qu'il faut aussi quantifier (Hoff *et al.*, 1999).

La RT-PCR relative n'utilise pas de compétiteurs mais la réaction est seulement comparée entre les différents échantillons. L'information est donc comme pour le Northern uniquement relative. Comme pour le Northern, les quantités d'ARN utilisées sont vérifiées en co-amplifiant un gène transcrit d'une manière identique dans tous les tissus. On utilise par exemple le gène codant pour l'actine ou celui codant pour la GAPDH.

3 - Comparaison du taux de transcription dans deux cellules différentes

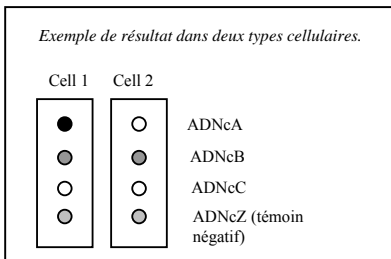
Expérience de Run on (Elongation in vitro de transcrits initiés in vivo)

On purifie des noyaux sur coussin de sucrose à 70%. Cette forte molarité en sucre empêche les petits facteurs de sortir des noyaux et sépare ces derniers du reste du lysat cellulaire. Le culot de noyaux est récupéré et incubé en présence d'UTP α ³²P pendant 30min.



Ce précurseur radioactif va permettre le marquage des ARNm en cours de transcription et ces derniers vont subir une élongation d'environ 200 à 500nt. Dans cette élongation réalisée *in vitro*, la vitesse de l'ARN Polymérase II ne dépasse pas à 10 ou 20 nt/min. Les rares initiations qui auront lieu au cours de

cette expérience donneront des ARN beaucoup trop courts et trop peu nombreux pour être détectés par la suite. Cette technique permet donc d'avoir un instantané de l'état transcriptionnel d'un type cellulaire au moment de la préparation des noyaux. Une fois l'élongation terminée, les noyaux sont lysés et les ARNm marqués sont extraits. La ou les sondes utilisées sont en général des ADNc correspondant aux gènes dont on désire analyser la transcription. Ces ADNc sont déposés sur une membrane en quantité identique (technique du dot blot : « taches »).

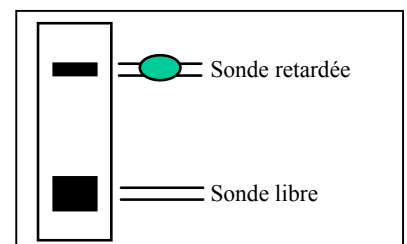


La membrane est ensuite incubée avec les tous ARNm radioactifs, lavée et autoradiographiée. Par cette expérience, on ne détermine pas une valeur absolue de la transcription d'un gène mais une valeur relative entre l'expression de différents gènes.

4 - Détermination des régions promotrices reconnues par des facteurs transcriptionnels

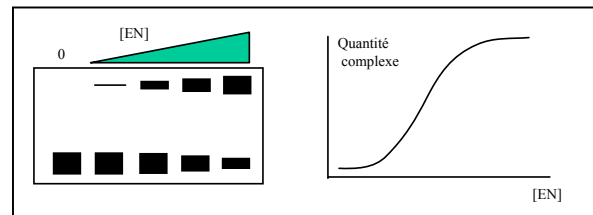
Gel retard

Cette expérience permet de mettre en évidence l'interaction d'un ou de plusieurs facteurs transcriptionnels avec une séquence d'ADN. On utilise des fragments d'ADN courts ou des oligonucléotides double brin (20pb à 40pb). Ces fragments sont marqués en 5' au ³²P. Ils sont ensuite incubés avec un extrait nucléaire (EN). Cet extrait



protéique nucléaire est préparé en lysant des noyaux. Il est ensuite dialysé, précipité au sulfate d'ammonium et resuspendu dans un volume équivalent au volume initial des noyaux. Après l'incubation, le mélange ADN/E.N. est déposé sur un gel d'acrylamide en condition non dénaturante. Le gel est séché et mis en exposition. On compare dans cette expérience le retard de la sonde du à l'interaction avec un facteur particulier présent dans l'extrait nucléaire.

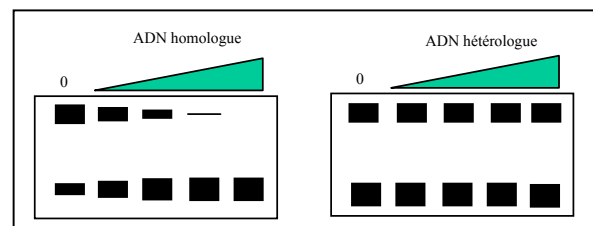
La spécificité de cette interaction peut être démontrée par la saturation du phénomène. L'expérience est réalisée avec une quantité constante d'ADN et une quantité croissante d'EN. Si l'interaction est spécifique on doit observer une courbe de saturation.



On peut également vérifier la spécificité par des expériences de compétition.

1- le compétiteur peut être homologue, c'est à dire que l'on va rajouter en plus de l'ADN marqué et de l'EN (qui sont en quantités constantes) une quantité croissante du même ADN non marqué.

2- le compétiteur peut être hétérologue : ADN marqué constant + EN constant + quantité croissante de poly-dIdC qui ne doit pas avoir d'effet.



Détermination des séquences reconnues

Footprint *in vitro* (empreinte à la DNaseI)

On cherche ici à identifier les séquences avec lesquelles vont interagir des facteurs transcriptionnels. On utilise des fragments de restriction (100 à 300pb) plus grands que dans les expériences de gel retard. La molécule que l'on va suivre est le fragment d'ADN marqué à une seule extrémité.

Comment faire ce marquage unilatéral ? Si le fragment d'ADN est purifié, il est nécessaire qu'il contienne un site de coupure asymétrique. Le fragment est marqué en 5', coupé puis purifié sur gel. Si ce fragment contient deux extrémités différentes on peut par exemple en marquer une par remplissage sans toucher à l'autre. Si le fragment est contenu dans un vecteur, ce dernier est linéarisé,

marqué en 5' puis l'insert est libéré par une seconde coupure enzymatique et purifié sur gel. Enfin on peut marquer par PCR en utilisant un oligonucléotide marqué en 5'.

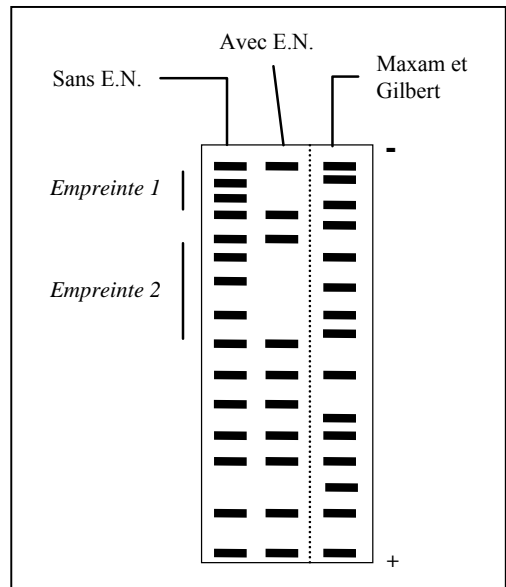
Le fragment est incubé en présence d'extrait nucléaire puis soumis à une digestion ménagée par la DNaseI. Un témoin est réalisé en incubant le fragment sans E.N. avec la DNaseI. L'analyse est effectuée sur gel d'acrylamide en conditions dénaturantes (gel de séquence) et le profil obtenu permet de déterminer les régions qui ont été protégées de l'action de la DNaseI par l'extrait nucléaire.

Si les fragments utilisés sont reconnus par plusieurs facteurs on pourra mettre en évidence des Footprint multiples.

Enfin, la protection mise en évidence sur un brin n'est pas nécessairement la même sur l'autre brin. On réalise donc en général le Footprint sur les deux brins, simplement en

changeant le brin marqué, pour mettre en évidence l'asymétrie de fixation sur les deux brins.

Pour déterminer les positions exactes des séquences protégées, une réaction de séquençage chimique (séquençage selon Maxam et Gilbert) est réalisée et déposée en parallèle sur le gel. Cette technique permet de séquencer le même fragment marqué que celui utilisé pour faire le Footprint. On réalise par exemple un séquençage G+A. Le fragment est méthylé chimiquement par du diméthylsulfate (méthyle les purines : G sur azote 7 et A sur azote 3). On fait ensuite agir la pipéridine qui va enlever la base méthylée (liaison glycosidique très instable à cause de la méthylation). D'autre part, la pipéridine attaque le désoxyribose ayant perdu sa base et sépare les deux nucléotides adjacents, mais cette réaction a un rendement très faible (2%) et par conséquent chaque copie d'ADN ne sera statistiquement coupée qu'une seule fois. Compte tenu du nombre de copies utilisées on va obtenir tous les fragments possibles.



Détermination des bases qui interagissent avec les facteurs transcriptionnels

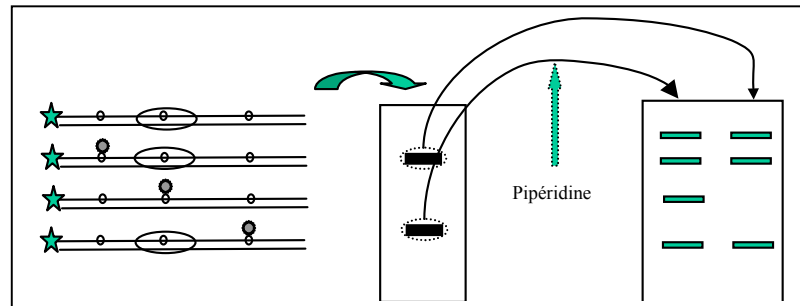
Interférence de méthylation

La méthylation est un des systèmes de régulation de la transcription des gènes chez les Eucaryotes. Chez l'espèce humaine, par exemple, on trouve 60 à 90% de 5-Méthyl cytosine dans les doublets CG.

Il n'y a pas de fixation des facteurs transcriptionnels sur les gènes méthylés et donc pas d'expression. (maintient du même état de différenciation d'une cellule fille par rapport à la cellule mère)

On peut donc rechercher les bases en interaction avec un facteur transcriptionnel particulier en utilisant des séquences cibles partiellement méthylées.

On marque un oligonucléotide en 5' au ^{32}P puis on l'hybride avec un second oligonucléotide non marqué (on peut bien sûr inverser



le marquage des deux oligonucléotides). L'ADN double brin est incubé en présence de faible quantité de DMS (diméthylsulfate) pour ne méthyler que quelques purines. Le fragment est ensuite incubé avec l'EN et analysé par gel retard. Les sondes libres et les sondes retardées sont récupérées et incubées avec de la pipéridine. Les produits de digestion sont ensuite analysés sur gel d'acrylamide en conditions dénaturantes. On détermine ainsi les bases impliquées dans cette interaction.

5 - Etude fonctionnelle de la transcription en système cellulaire

On se place ici dans un système cellulaire qui permet de faire les expériences dans un système proche d'un contexte *in vivo*. Les cellules étant en culture et donc hors de l'organisme dont elles sont issues, on parle de système *ex vivo*.

Le principe de ces expériences est de faire pénétrer dans le noyau de cellules en culture une construction chimérique qui permet de tester le rôle, dans la transcription et sa régulation, des différentes séquences et facteurs mis en évidence précédemment.

Choix de la cellule

On utilise soit des lignées cellulaires (cellules immortalisées) soit des cultures primaires. Ces dernières conservent des caractéristiques de différenciation parfois indispensables, permettant par exemple de répondre à différents effecteurs hormonaux. Certaines d'entre elles ne se divisent pas et présentent l'inconvénient de devoir être préparées extemporanément. A l'inverse, les lignées cellulaires, qui peuvent être maintenues en culture pendant beaucoup plus longtemps, ont

l'inconvénient d'avoir perdu des caractéristiques de différenciation que présentaient les cellules dont elles sont dérivées.

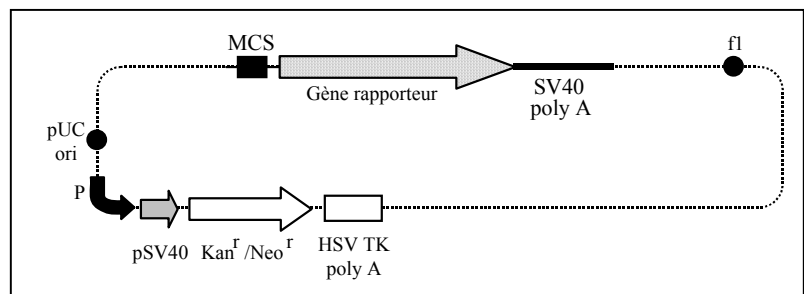
Constructions chimériques

Constructions homologues

Dans une construction homologue la totalité de la partie promotrice provient d'un même et unique gène. Ce promoteur ou une partie de ce dernier est sous cloné dans un plasmide en amont d'un gène rapporteur.

Le plasmide utilisé doit permettre de réaliser le clonage en système bactérien puis de pouvoir dans certains cas être sélectionné en système eucaryote.

Dans ce système la présence de la séquence de polyadénylation de SV40 (séquence normalement issue d'une modification post-transcriptionnelle) permet d'avoir un messager dont la stabilité et la traductibilité est invariable. Il est en effet indispensable qu'il n'y ait aucune régulation post transcriptionnelle ni traductionnelle si l'on veut avoir un reflet direct de l'activité transcriptionnelle du promoteur testé.



Gènes rapporteurs

Les gènes rapporteurs sont des ADNc qui codent pour des protéines qui doivent remplir deux conditions :

- être facilement détectables
- ne pas être confondues avec une protéine endogène

On utilise en général une enzyme dont on détecte l'activité après avoir lysé les cellules. Cette enzyme peut être d'origine Procaryote (CAT : *Chloramphenicol Acetyl Transferase*, β Galactosidase) ou Eucaryote (Luciférase, GFP : *Green Fluorescent Protein*, SAP : *Secreted human Alkaline Phosphatase*). L'activité est détectée soit par chimiluminescence, fluorescence, radioactivité ou colorimétrie.

Transfection

La transfection est en générale effectuée de manière transitoire, c'est à dire que le plasmide reste sous forme d'épisome dans le noyau. On peut utiliser un vecteur avec une origine de répllication eucaryote qui permet de maintenir le plasmide dans les cellules, si ces dernières se divisent bien sûr.

Constructions hétérologues

On réalise également des constructions hétérologues. C'est à dire que le promoteur de ces constructions est constitué de deux parties. Un promoteur minimal comme celui de la thymidine kinase (tk) par exemple. Ce promoteur présente des sites de fixation NF1, Sp1 et la TATA box et ne subit aucune régulation transcriptionnelle dans une cellule eucaryote. Il présente donc une activité transcriptionnelle de base que l'on va pouvoir augmenter ou diminuer en lui greffant en 5' des séquences régulatrices. Ces constructions hétérologues permettent donc de tester des séquences régulatrices dans un contexte qui ne subit normalement aucune régulation. C'est par ce type d'approche que l'on peut mettre en évidence des Enhancer ou des Silencer, séquences régulatrices indépendantes de leur distance et de leur orientation par rapport au promoteur minimal.

6 - Approches *in vivo*

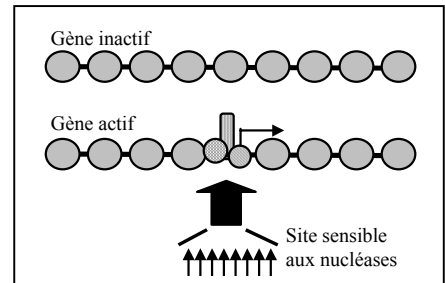
Dans toutes les approches précédentes, l'ADN utilisé est sous forme d'ADN double brin adoptant une simple structure d'hélice B. Dans une structure chromatinienne, l'ADN prend une toute autre conformation qui modifie son accessibilité vis à vis de la machinerie transcriptionnelle. Sommairement, la chromatine se présente sous deux formes : l'euchromatine ou chromatine claire qui correspond à des régions transcrites et l'hétérochromatine ou bandes sombre qui correspond à des régions non transcrites. Ces deux structures reflètent des régions de condensation différentes de l'ADN (variation de 1 à 1000).

Il existe d'autres paramètres qui modulent l'expression génique *in vivo* : l'acétylation des histones et la méthylation des gènes par exemple.

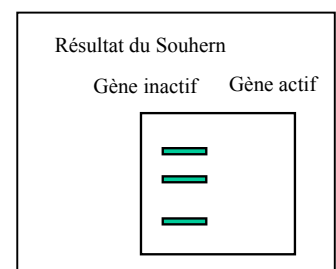
Il est donc important d'avoir accès à des expériences réalisées *in vivo* pour valider les interprétations obtenues *in vitro*.

Sites sensibles à la DNaseI

Cette expérience permet de mettre en évidence la structure particulière de la chromatine en corrélation avec l'état transcriptionnel de gènes particuliers. On purifie des noyaux (ou on utilise des cellules traitées au NP40-détergent non ionique) que l'on traite à la DNaseI **faiblement concentrée**. L'ADNg est extrait et analysé par Southern blot avec un ADNc correspondant



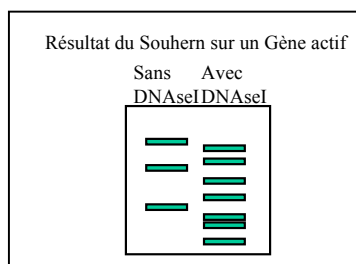
à un gène exprimé par la cellule ou qui l'a été récemment. Dans ce cas on n'obtient pas de signal. Si l'hybridation est réalisée avec un ADNc non exprimé dans ce type cellulaire on obtient un signal. Une absence de signal signifie que l'ADNg a été digéré par la DNaseI. Si l'accessibilité à la DNase est permise dans une région particulière elle l'est aussi pour les facteurs transcriptionnels et le gène est donc probablement transcrit ou susceptible de l'être.



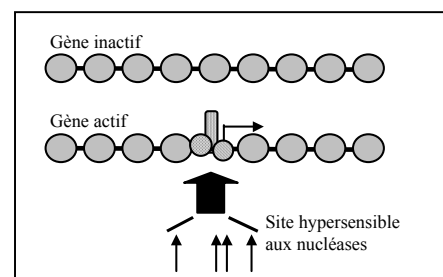
Biologiquement, la différenciation cellulaire « installe » les gènes exprimés ou non dans une structure chromatinienne particulière.

Sites hypersensibles à la DnaseI

Si on utilise la DNaseI de manière **excessivement diluée**, cette dernière se comporte comme une enzyme de restriction. On met ici en évidence des sites hypersensibles qui produisent des fragments de tailles définies et reproductibles. Les coupures sont analysées par Southern avec une



sonde assez grande pour recouvrir le promoteur. Une expérience témoin est réalisée sans DNaseI. Les sites hypersensibles caractérisent



les bases particulièrement exposées au niveau d'un promoteur et qui sont impliquées dans les mécanismes de régulations transcriptionnelles.

Footprint génomique

Dans ce Footprint les noyaux (ou cellules traitées NP40) sont incubés avec de la DNaseI **très diluée** de manière à couper dans des conditions de digestion ménagée.

La détection des fragments est réalisée par une PCR particulière : Ligation Mediated PCR ce qui permet d'augmenter la sensibilité de la détection des fragments obtenus.

Ligation Mediated PCR

Après la lyse des noyaux, l'ADNg est extrait, dénaturé et hybridé avec un oligonucléotide (Pg1) spécifique du gène d'intérêt. Pg1 sert d'amorce dans une réaction d'extension d'amorce dont l'extrémité 3' est délimitée par les coupures engendrées par la DNaseI. Les produits d'ADN db servent de cibles pour liguer un linker. Une PCR est réalisée avec l'oligonucléotide P1 (hybridant sur le linker) et un oligonucléotide emboîté Pg2, c'est à dire un oligo spécifique du gène étudié et hybridant en amont de Pg1 (PCR « semi nested » i.e. semi emboîtée). Cette PCR permet d'amplifier spécifiquement la région promotrice du gène quelle que soit la position de coupure par la DNaseI et sur une très faible quantité de matériel. Le fait d'utiliser un oligo emboîté augmente considérablement la spécificité de la réaction et permet en limitant la taille du produit amplifié d'éviter l'obtention d'une PCR entre les deux polylinker. Enfin, on réalise une extension d'amorce avec un oligonucléotide Pg3 marqué, emboîté par rapport à Pg2. L'analyse des résultats est effectuée sur gel d'acrylamide et se présente comme un Footprint *in vitro*, la fixation de facteurs transcriptionnels protégeant l'ADN de la digestion par la DnaseI.

Transfection stable

Pour placer une construction dans un contexte chromatinien on peut chercher à intégrer la construction chimérique (homologue ou hétérologue) dans le génome cellulaire. Trois conditions doivent être remplies pour favoriser cette expérience :

1-utiliser des lignées cellulaires ou des cultures primaires de cellules ayant conservées une activité mitotique.

2- linéariser le plasmide pour favoriser son intégration

3-ajouter un marqueur de sélection eucaryote (G418/néomycine)

Les cellules sont sélectionnées pendant 3 semaines en présence de l'agent sélectif puis, selon l'information que l'on cherche, soit :

- on mesure l'activité sur l'ensemble des clones sélectionnés-->activité moyenne de la population cellulaire

- on isole par dilution limite différents clones cellulaires et on compare l'activité obtenue pour chaque clone --> dépend du site d'insertion au sein du génome.

Outre le fait que l'ADN exogène est dans une conformation chromatinienne, l'intérêt de cette technique est de pouvoir conserver le clone obtenu indéfiniment et l'expérience devient entièrement indépendante de la technique de transfection utilisée ainsi que de son efficacité.

7 - Analyse de l'abondance de la transcription (transcriptome)

Le but est ici de connaître quels gènes sont transcrits de façon importante dans une population de cellule. Il faut au préalable connaître tous les gènes (ou la plupart des gènes) de l'espèce.

EST

(Express Sequence Tag, Adams *et al.*, 1991). Des ADNc sont fabriqués par amorçage au hasard puis séquencés. Les gènes sont alors identifiés par comparaison des séquences obtenues avec les séquences présentes dans les banques. Cette méthode est lourde puisqu'il faut faire une séquence par clone, elle présente surtout un intérêt pour cloner de nouveaux gènes.

Array hybridization

La principe technologique des puces à ADN ou bio-puces est particulièrement simple et utilise les propriétés d'hybridation des acides nucléiques. Par contre la mise en œuvre de ces puces est lourde et complexe.

L'utilité de cet outil est scientifiquement très large. La connaissance du taux d'expression d'un gène dans des situations différentes constitue une source d'information importante quant à la

connaissance de sa fonction. Lorsqu'on analyse l'expression simultanée d'un ensemble de gènes (« transcriptome » : ensemble des ARN transcrits), plusieurs milliers voire un génome entier, dans des dizaines de conditions physiopathologiques différentes, on accède à une information globale et intégrée du profil d'expression d'un type cellulaire dans des conditions variées.

Cette approche peut également servir à cribler de nouvelles molécules à visée thérapeutique en analysant leur impact sur le profil d'expression.

D'autres applications des Bio-puces concernent la recherche de SNP (*Single Nucleotide Polymorphism*), le séquençage par hybridation, le diagnostic.

Quelle que soit la technologie utilisée pour fabriquer les puces, le principe général est le même. De nombreuses sondes (oligonucléotides, produit PCR ou ADNc) sont immobilisées sur un minuscule support : la puce (sa taille est actuellement de l'ordre du centimètre carré). Les acides nucléiques à étudier sont amplifiés et marqués au cours de cette amplification en général avec un fluorochrome. Le produit d'amplification est ensuite hybridé avec la puce. Après lavage, les régions où l'hybridation s'est produite sont repérées et quantifiées au moyen d'un système d'acquisition d'image ce qui permet de déterminer avec quelle(s) sonde(s) le produit d'amplification s'est hybridé. Les résultats sont ensuite analysés et validés grâce à des logiciels informatiques dédiés à cette analyse et interprétés d'un point de vue biologique.

Cette étape est certainement la plus délicate. La quantité d'information engendrée par cette technologie est énorme et il n'est pas encore facile d'exploiter ces informations sur le plan biologique.

Fabrication de la puce:

On peut distinguer trois types de Bio-puces : macroarray (ou filtre à haute densité), microarray et puce à oligonucléotides.

Le support est l'élément principal de la puce. Sur ce support vont être répartis régulièrement et en grand nombre les unités d'hybridation (« plot »).

On utilise des supports de nature différente :

Le verre (peu cher, inerte et transparent) est recouvert de polymère (L-polylysine, silane) et les sondes sont fixées par liaisons électrostatiques. On utilise aussi du silicium, des surfaces Or/Platine, des fibres optiques, du gel d'acrylamide, du nylon.

Chaque plot correspond à une séquence d'ADN (oligonucléotides, cDNA, produit PCR) en plusieurs milliers d'exemplaires. Ces séquences sont fixées soit par dépôt soit par synthèse *in situ*.

En fixation par dépôt, une micropointe robotisée permet de déposer des plots avec une précision de 10µm soit par simple contact avec la surface, soit par jet d'encre contrôlé par un système piezo-électrique, par électrochimie ou de manière mécanique dans certains cas.

Lorsque la synthèse est réalisée *in situ*, elle se fait soit par photolithographie ou par électrochimie.

	Macroarray	Microarray	Puce à oligonucléotides
Support	nylon	Verre, silicium, Or/platine	Quartz couvert de silane
Taille de la puce	12cm x 8cm	5.4cm x 0.9cm	1.28cm x 1.28cm
Marquage des cibles	Radioactif	Fluorescent	Fluorescent
Nature des sondes	CDNA ou PCR	cDNA ou PCR	Oligonucléotides
Fixation des sondes	Dépôt direct	Dépôt direct	Photolithographie
Nombre de spot (sondes différentes)	2400	10 000 à 12 000	300 000 à 400 000
Nombre de sondes par spot	Variable	variable	10 ⁷
Nombre de conditions expérimentales pouvant être testées simultanément	1	2	1
Analyse	Autoradiographie	Microscopie confocale	Microscopie confocale
Réutilisation	Oui	Non	Non

Les cibles

Les cibles sont marquées de différentes manières : fluorochrome, radioactivité.

Ces cibles sont des cDNA reverse transcrit à partir d'une population de mRNA. Au cours de cette reverse transcription on marque les cDNA soit avec un précurseur ³²P ou ³³P, soit avec un fluorochrome. Dans le cas des microarrays sur lame de verre, on peut directement comparer deux conditions expérimentales en marquant les cDNA avec deux fluorochromes différents.

Le marquage radioactif n'est utilisé que si le support est autofluorescent (nylon).

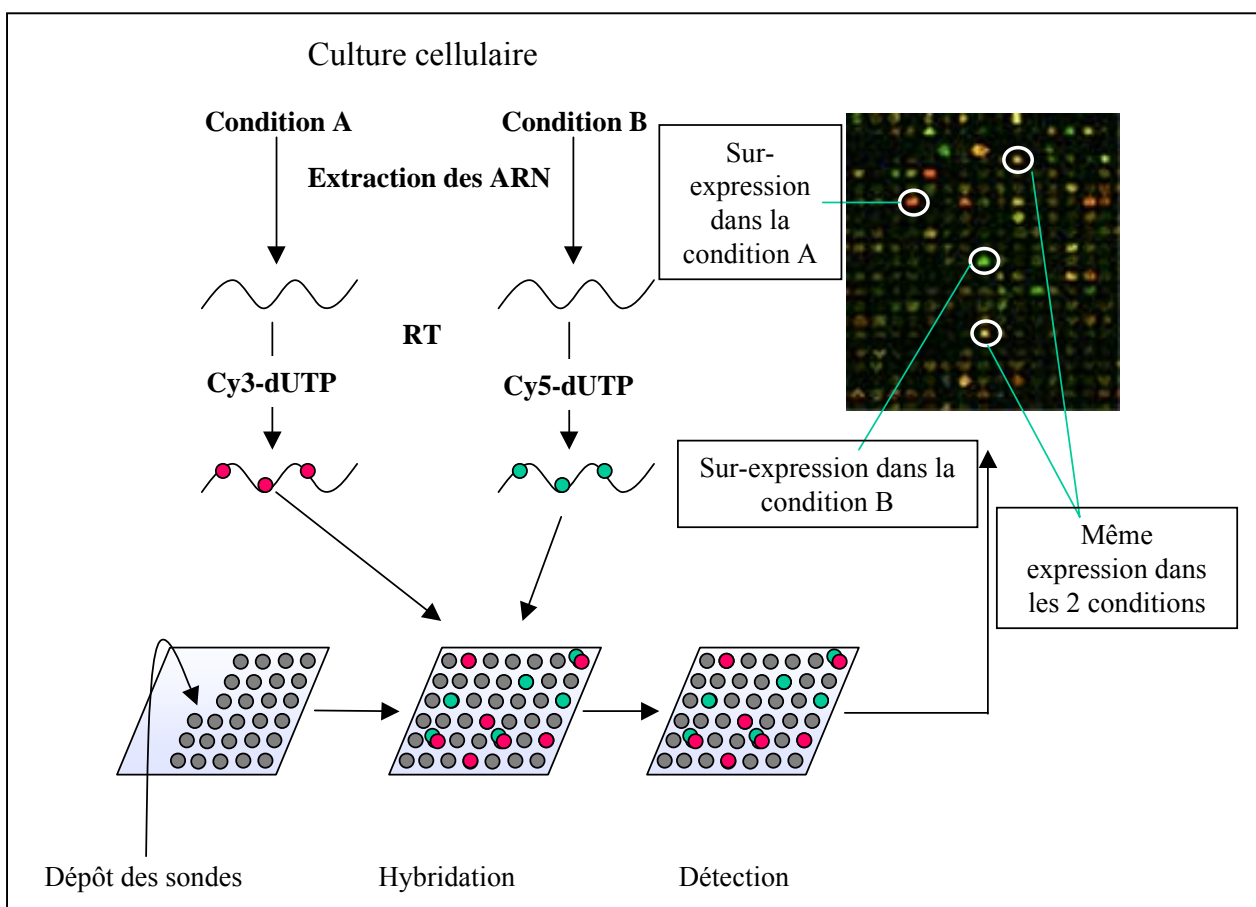
La majorité des puces sont analysées par fluorescence soit en marquant directement la cible lors de sa synthèse par un fluorochrome [Cy5-dUTP (Indodicarbocyanine, vert λabs 550nm/λem 570nm), Cy3-UTP (Indodicarbocyanine, rouge λabs 650nm/λem 670nm)] ou indirectement par de la biotine détectée par la suite grâce à de la streptavidine couplée à un fluorochrome.

L'hybridation:

On essaie en général d'avoir sur la puce des oligonucléotides ayant un T_m proche. Si ce n'est pas le cas on peut contrôler l'hybridation en appliquant une différence de potentiel variable sur les spots de la puce.

Détection :

Chaque spot est excité par un laser et on récupère la fluorescence émise via un photo-multiplicateur couplé à un microscope confocal. On obtient en superposant les images en fausses couleurs des spots allant du vert au rouge en passant par le jaune (de l'ADN des deux conditions fixé en quantité égal).



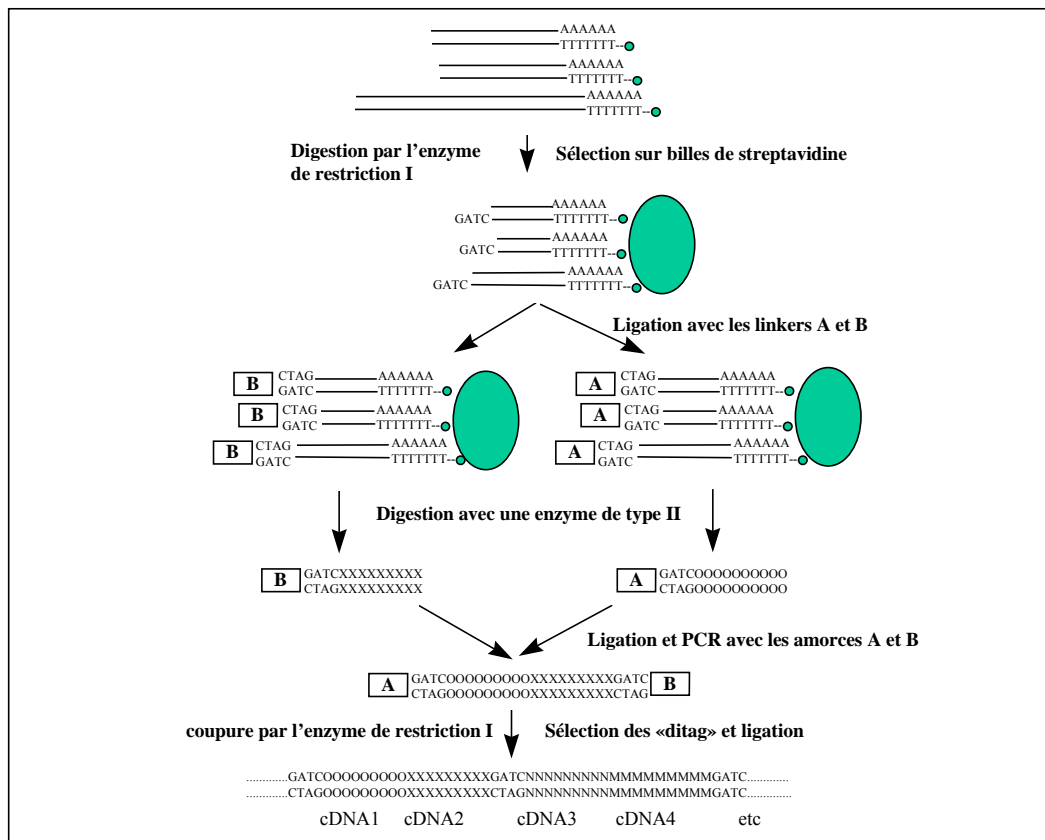
SAGE

Comme ce n'est pas la peine de séquencer un grand nombre de base, une autre technique peut être employée, la SAGE (Serial Analysis of Gene Expression, Velculescu *et al.*, 1995). En effet le génome humain code pour environ 80 000 transcrits, une séquence de 9 bases est donc suffisante

puisqu'elle peut distinguer 4^9 soit 262 144 transcrits. On veut donc avoir pour chaque clone une seule séquence de 9 bases. On va les liquer à la suite les uns des autres et les séquencer simultanément.

Des ADNc sont fabriqués à l'aide d'amorces oligodT biotinylées puis coupés à l'aide d'une enzyme de restriction qui coupe souvent (qui reconnaît 4 bases et donc qui coupe tous les 4^4 bases en moyenne). Les produits de digestions sont purifiés grâce à des billes de streptavidine. La solution est alors divisée en deux et chaque aliquote est ligué à un adaptateur (A et B) ayant une extrémité complémentaire au fragment de restriction obtenu sur les ADNc. Ces adaptateurs ont un site de restriction d'une enzyme de type IIS. Ces enzymes reconnaissent une séquence mais coupent en aval, le site sur l'adaptateur est positionné de telle sorte que la coupure laisse une dizaine de bases de l'ADNc. Par exemple, BsmF coupe 14 pb en aval de son site de reconnaissance. Après digestion par l'enzyme de type IIS, on obtient une série de fragments d'ADNc de petite taille. Comme on n'en a pas assez, on rend les extrémités franches à l'aide de la klenow, on effectue une ligation suivie d'une réaction de PCR utilisant les deux amorces A et B.

Il va ensuite falloir associer ces différentes séquences pour pouvoir les séquencer. Les produits d'amplification peuvent être directement ligués les uns aux autres mais il est préférable de les digérer par le premier enzyme de restriction, puis les produits de la digestion sont purifiés et ligués pour former des concatémères. On obtient généralement 30-50 ditags par concatémères, l'ensemble sera inséré dans un vecteur puis séquencé.



Lors de la formation des concatémères on peut avoir des contaminations par les linkers. Pour éviter cette contamination, les amorces utilisées dans la réaction de PCR sont biotinylées et séparées des ditags à l'aide de streptavidine liée à des billes magnétiques (Powell, 1998)

8 - Analyse de la méthylation des promoteurs

mise en évidence :

digestion de l'ADN par des isoschizomères sensibles et non sensibles à la méthylation. Les produits de digestion sont analysés en Southern pour voir si le site était coupé ou non.

Etude de la fonction de la méthylation

Traitement des cellules par la 5 azacytosine. Au cours des réplifications, la 5 azacytosine est incorporée dans l'ADN qui ne peut plus être méthylé.

Analyse de la traduction

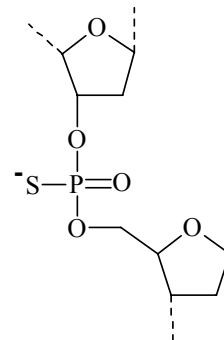
Analyse de l'expression par antisens

Cette méthode a fait son apparition il y a une vingtaine d'années (Zamecnik and Stephenson, 1978) et s'est développé principalement en modifiant l'oligonucléotide. On fabrique un oligonucléotide qui s'hybridera à l'ARN messenger ce qui aura deux effets: l'hybridation inhibera la traduction et induira une coupure de l'ARN par la RNase H. L'oligonucléotide doit donc être résistant aux nucléases, il doit traverser la membrane cytoplasmique, il doit avoir une bonne affinité pour sa cible et doit être un bon substrat pour la RNase H.

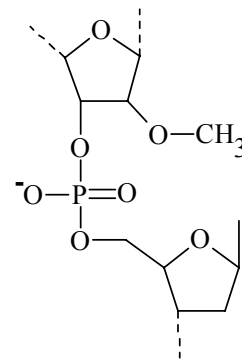
a) La résistance aux nucléases

La résistance aux nucléases peut être obtenue en modifiant soit le squelette phosphodiester soit en rajoutant un cap en 3'

On peut modifier le squelette en incorporant des phosphorothioates



On peut incorporer un 2'-O methyl nucléotide ((2'-OMethyl RNA).
 Cette modification protège des RNase comme des DNase sans altérer l'hybridation.

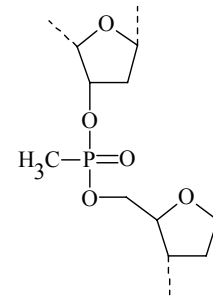


On peut ajouter un cap en 3' comme de la biotine ou un amino propyl

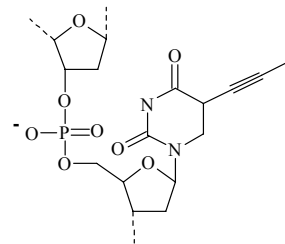
b) Le passage à travers la membrane

Il peut être obtenu :

- En utilisant un transporteur comme le cholestérol (Letsinger, 1989)
- Par modification du squelette en le rendant plus lipophile par exemple en utilisant un methylphosphonate.



- Par modification des bases, en utilisant un 5-(1-propynyl) desoxy pyrimidine



c) L'augmentation de l'hybridation

L'augmentation de l'hybridation est difficile à réaliser car les liaisons hydrogènes sont déjà optimisées.

toutefois on a remarqué que

- Les oligonucléotides contenant des 2'-Ome nucléotides forment des hybrides avec l'ARN plus stables que les oligonucléotides ADN ou ARN (Monia, 1993). De plus la formation de triple hélices sont fortement stabilisées par les 2'-OMe-RNA.
- Les oligonucléotides contenant des phosphorothioates forment des hybridations plus stables.
- Les pyrimidines modifiées à la position 5 (avec un groupe methyl ou propynyl) forment des hybridations plus stables.

d) L'augmentation de l'activité RNase H

Les oligonucléotides composés d'oligodéoxynucléotides phosphate ou thiophosphates sont de bons substrats pour la RNase H. Mais les 2' O-Me oligonucléotides ou les phosphonates ne sont pas des substrats. Si on veut utiliser ces modifications, on fabrique des oligonucléotides chimériques composés d'au moins cinq nucléotides substrat se suivant dans la séquence pour induire l'activité RNase H.

Etude de la population d'ARN en cours de traduction dans une cellule

Analyse des polysomes

On purifie les polysomes d'une cellule par centrifugation zonale en gradient de saccharose, les polysomes, qui sont constitués d'ARN liés à plusieurs ribosomes, se retrouvent dans les fractions contenant les complexes de haute masse moléculaire. Parallèlement on peut purifier les ARN qui ne sont pas impliqués dans une traduction, qui ne sont pas liés aux ribosomes. Les deux lots d'ARN (liés et libres) sont alors utilisés dans une hybridation sur une puce et la différence d'hybridation permet de connaître la traductibilité de l'ensemble des gènes d'une cellule dans une condition particulière.

