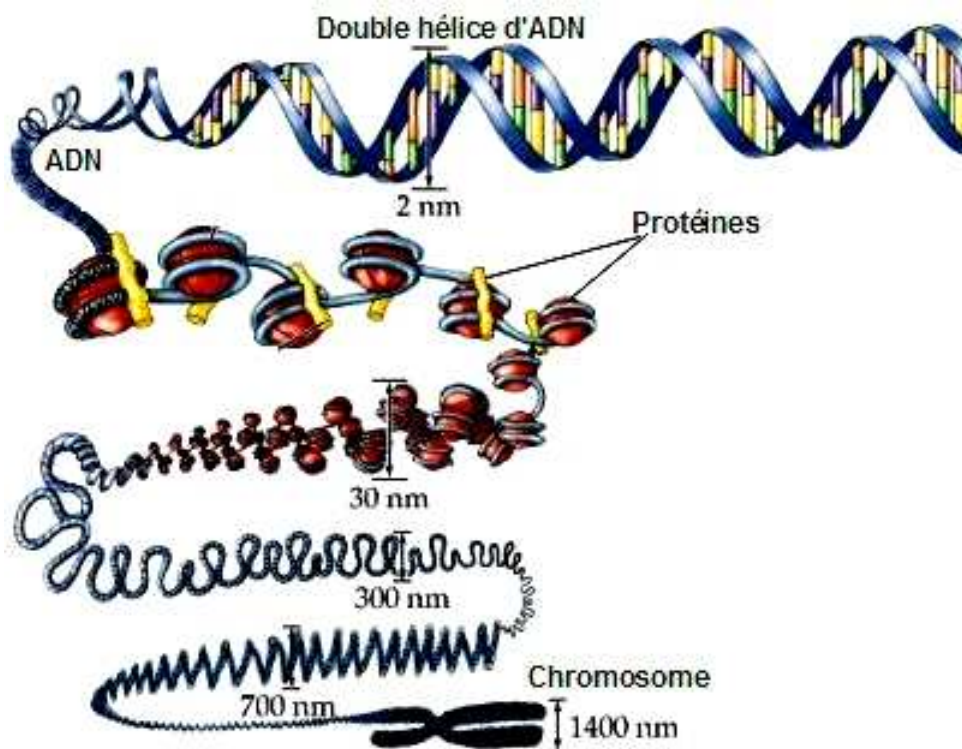


FORMATION BIOLOGIE MOLECULAIRE

DAKAR – Septembre 2006



SOMMAIRE

La naissance de la biologie moléculaire

I- L'émergence de la génétique formelle	p4
II- Le chromosome, support de l'hérédité	p4
III- La convergence de la biochimie et de la génétique	p4
IV- L'ADN, support de l'information génétique	p4
V- La structure de l'ADN	p5

La structure des acides nucléiques

I- Les nucléotides	p6
1- Les différents constituants des nucléotides	p6
2- Nomenclature	p6
II- Caractéristiques générales des acides nucléiques	p7
1- Les liaisons reliant les nucléotides	p7
2- Sens de lecture d'un acide nucléique	p7
III- L'ADN	p7
1- Caractéristique de l'ADN	p7
2- Propriétés physico-chimiques de l'ADN	p8
3- Hybridation des acides nucléiques et des sondes nucléiques	p9
IV- Les ARN	p9
1- Caractéristiques générales des ARN	p9
2- Les différents types d'ARN	p9
a- Les ARN ribosomiques (ou ARNr)	p9
b- Les ARN de transfert (ou ARNt)	p9
c- Les ARN messager (ou ARNm)	p10
d- Les petits ARN nucléaires (ou snRNA pour small nuclear RNA)	p10

La réplication de l'ADN

I- Caractéristiques générales de la réplication	p12
--	------------

II- Éléments nécessaires à la réplication de l'ADN	p12
III- Les différentes étapes de la réplication de l'ADN	p12
IV- Discontinuité de la réplication entre les deux brins d'ADN	p13
V- Réparation des erreurs	p14

La transcription de l'ADN en ARN messager
--

I- Caractéristiques générales de la transcription	p15
II- Les éléments nécessaires à la transcription	p15
III- Les différentes étapes de la transcription	p15
1- L'initiation de la transcription	p15
2- L'élongation de la chaîne poly-nucléotidique au cours de la transcription	p16
3- La terminaison ou fin de la transcription	p16
IV- La maturation des produits de transcription	p17
1- L'addition de la coiffe à l'extrémité 5'	p17
2- L'addition de poly(A) à l'extrémité 3'	p17
3- L'élimination des introns	p17

La traduction

I- Caractéristiques générales de la traduction	p18
II- Les éléments nécessaires à la traduction	p18
III- Mécanisme de la traduction	p18
1- Initiation de la chaîne peptidique	p19
2- Elongation de la chaîne peptidique	p19
3- Terminaison de la chaîne peptidique	p20
IV- Le code génétique	p20
1- Le code génétique est défini par trois lettres	p20
2- Le code génétique est universel	p21
3- Le code génétique est dégénéré	p21
4- Code non chevauchant	p21
5- Notion de cadre de lecture	p21

Le polymorphisme de l'ADN

I- Types de polymorphisme	p22
1- Macrolésions de l'ADN : les remaniements chromosomiques	p22
2- Microlésions de l'ADN : les mutations ponctuelles	p22
3- Séquences répétées en tandem	p23
II- Mécanismes mis en œuvre dans la génération du polymorphisme	p23
1- Erreurs de réparations	p23
2- Recombinaisons inégales	p23
3- Conversion génique	p24
4- Insertions de séquences mobiles (transposons) ou virales	p24
5- Glissement intra-chromatidien	p24

Les outils et techniques de la biologie moléculaire

I- Les outils de la biologie moléculaire	p25
1- Les enzymes	p25
2- Les amorces et sondes	p26
3- Les vecteurs et cellules hôtes	p26
II- Les techniques de la biologie moléculaire	p26
1- Extraction –Purification de l'ADN	p27
2- Migration électrophorétique et visualisation des fragments d'ADN	p28
3- PCR	p29
4- PCR-RFLP	p31
5- Purification de produits PCR	p31
6- Séquençage	p31
7- Microsatellites	p32
8- AFLP	p33
9- SSCP	p34
10- Clonage	p35

Annexes

Annexe I : Les différentes phases de la mitose	p38
Annexe II : Glossaire	p38

La naissance de la biologie moléculaire

Le XXe siècle coïncide avec la naissance génétique : débutant avec la redécouverte des travaux de Mendel précisément en 1900, se poursuivant par l'élaboration de la théorie chromosomique de l'hérédité au début du siècle, la découverte de l'ADN comme support biochimique de l'information génétique, l'élucidation de sa structure, et l'explosion de la biologie moléculaire à partir des années 70.

I- L'émergence de la génétique formelle

La génétique classique débute au milieu du XIXe siècle avec les travaux de **Gregor Mendel** (1822-1884), qui établit dès 1866 les premières **lois de l'hérédité** en étudiant la transmission au cours des générations d'un certain nombre de caractères simples à observer. Ce sont des caractères à versions alternatives tranchées : fruits lisses ou rugueux, verts ou jaunes, graines rondes ou irrégulières, jaunes ou vertes, tige haute ou petite. Ces travaux passent alors inaperçus et seront redécouverts en 1900.

II- Le chromosome, support de l'hérédité

En 1910, les travaux de **Thomas Morgan** (1866-1945) sur la transmission à la descendance de caractères mutés chez la mouche du vinaigre (*Drosophila melanogaster*), conduisent au développement de la **théorie chromosomique de l'hérédité**. Les gènes sont alors localisés sur les chromosomes, et avec **Alfred Sturtevant** (1891-1970), ils pourront même y être ordonnés, constituant les premières cartes génétiques en 1913.

III- La convergence de la biochimie et de la génétique

Si la présence des gènes sur les chromosomes est alors établie, rien n'est connu de la nature biochimique des gènes ou de leur mode d'action. La première relation entre un gène et un enzyme est établie en 1902 par **Archibald Garrod** (1857-1936), à partir d'une observation portant sur une maladie génétique humaine (déficience enzymatique transmissible à la descendance). **George Wells Beadle** (1903-1989) approfondit cette relation sur un système accessible à l'expérimentation, le champignon *Neurospora crassa* sur lequel il réalise des mutations par irradiation. L'ensemble de ces travaux aboutissent finalement à la conclusion que les gènes contrôlent la synthèse des enzymes, et que **chaque protéine est codée par un gène différent**.

IV- L'ADN, support de l'information génétique

Le premier phénomène qui allait permettre de progresser dans l'identification du support de l'hérédité est celui de la transformation bactérienne, rapporté en 1928 par l'anglais **Fred Griffith** (1877-1941). Ce phénomène représente alors un test d'activité biologique (injection de bactéries virulentes ou non à des souris), grâce auquel il est possible de déterminer la nature du matériel génétique. Ce test ne sera pas mis à profit par Griffith lui-même, mais par **Oswald Avery** (1877-1955) qui l'utilise pour élucider **la nature biochimique du matériel génétique : l'ADN**. Cette découverte est toutefois accueillie avec beaucoup de scepticisme. Il faudra de nombreux autres travaux pour que cette réalité soit acceptée : en particulier ceux

d'**Erwin Chargaff** (1905-1992) ou de **Alfred Hershey** (1908-1997). L'acceptation définitive ne viendra qu'avec l'élucidation de la structure de l'ADN par **Watson et Crick**.

V- La structure de l'ADN

Finale­ment, c'est avec l'élucidation de la **structure en double hélice de l'ADN** par **James Dewey Watson** (1928-) et **Francis Harry Compton Crick** (1916-2004) en 1953 que la biologie moléculaire connaît son apothéose. Ils établissent en effet un modèle moléculaire à l'aide de l'ensemble des données acquises sur l'ADN depuis le début du siècle. Cette structure est aujourd'hui connue de tous, elle est devenue l'emblème de la biologie moléculaire (voire la couverture de ce document!

En outre, Crick a apporté d'importantes contributions à plusieurs domaines de la biologie moléculaire : c'est à lui que l'on doit notamment **le dogme central de la biologie**, qui stipule que le flux d'information depuis les acides nucléiques vers les protéines est à sens unique.

La structure des acides nucléiques

Les acides nucléiques ont été isolés initialement des noyaux des cellules. On peut en distinguer deux grands types : les acides désoxyribonucléiques (**ADN**) et les acides ribonucléiques (**ARN**). Les premiers sont essentiellement localisés dans le noyau des cellules et les seconds dans le cytoplasme cellulaire. Les acides nucléiques (ADN et ARN) sont des macromolécules et constitués de sous-unités appelées **nucléotides**.

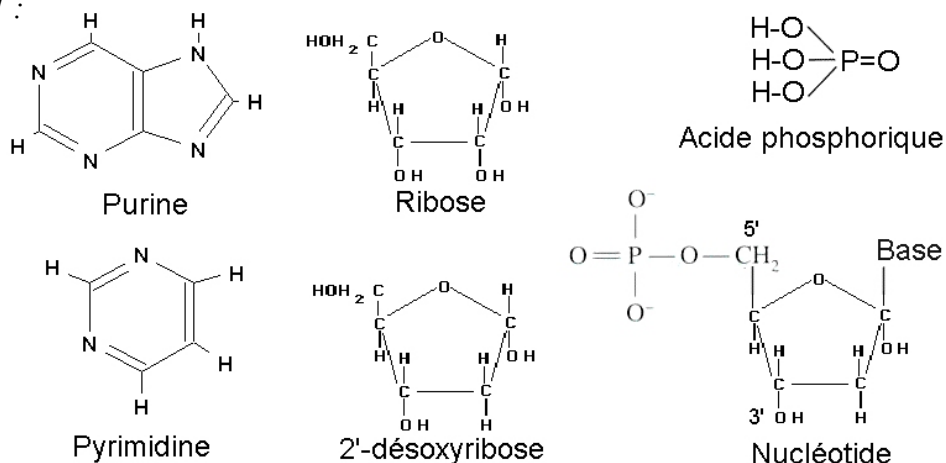
I- Les nucléotides

1- Les différents constituants des nucléotides

Un nucléotide comporte trois composants (figure 1) :

- **une base** : les bases sont classées en pyrimidines et en bases purines. Les principales bases pyrimidiques sont: l'uracile, la cytosine et la thymine. Les principales bases puriques sont l'adénine et la guanine.
- **un ose** (sucre) : il en existe deux types, le ribose et le 2'-désoxyribose. Ces deux sucres sont des oses à cinq atomes de carbone ou pentoses. On les numérote avec des chiffres accompagnés de l'indication prime pour éviter des confusions avec les numérotations des bases. Le 2'-désoxyribose est un ribose dans lequel il manque un OH en 2' (remplacé par un H).
- **un acide phosphorique** (H_3PO_4) : il possède trois fonctions acide. Deux de ces fonctions sont estérifiées dans les ADN et les ARN. La troisième fonction acide est libre.

Figure 1 :



2- Nomenclature (tableau 1)

L'association ose-base est appelée **nucléoside**.

L'association acide phosphorique-ose-base est appelée **nucléotide**.

Les nucléosides et nucléotides à base purique (A et G) ont respectivement une terminaison "ine" et "ylique".

Les nucléosides et nucléotides à base pyrimidique (C, T et U) ont respectivement une terminaison "idine" et "idylique".

Tableau 1 :

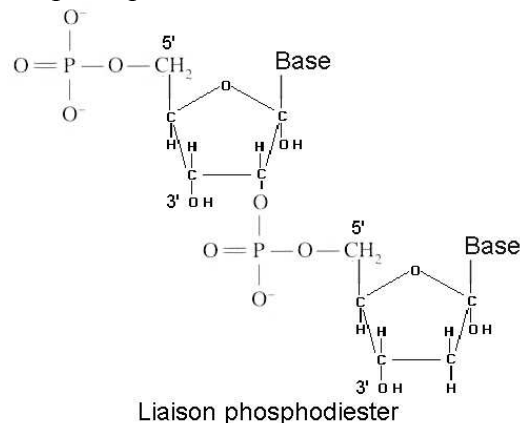
Base	Nucléoside (ose+base)	Nucléotide (ac P+ose+base)	Code
adénine	adénosine	acide adénylique	A
guanine	guanosine	acide guanylique	G
cytosine	cytidine	acide cytidylique	C
thymine	thymidine	acide thymidylique	T
uracile	uridine	acide uridylique	U

II- Caractéristiques générales des acides nucléiques

1- Les liaisons reliant les nucléotides

Dans un acide nucléique (ou polymère de nucléotides), les nucléotides sont assemblés entre eux par des **liaisons phosphodiester** (figure 2). Une molécule d'eau est donc éliminée entre un OH de l'acide phosphorique et l'H de la fonction alcool située en 3' de l'ose. Quand l'acide phosphorique présente ses deux fonctions acides bloquées dans la formation d'ester, on parle de liaison phosphodiester. On peut indiquer les chiffres a et b pour préciser la liaison phosphodiester concernée. La liaison phosphodiester a correspond à la liaison ester entre l'acide phosphorique et l'OH en 3' de l'ose. La liaison phosphodiester b correspond à la liaison ester entre l'acide phosphorique et l'OH en 5' de l'ose.

Figure 2 :



2- Sens de lecture d'un acide nucléique

Par convention, on lit toujours un acide nucléique dans le **sens de l'extrémité 5'** (comportant un groupement phosphate) **vers l'extrémité 3'** (comportant un groupement OH libre). La séquence des bases d'un ADN par convention sera écrite dans le sens horizontal en précisant les extrémités 5' et 3' et on indique seulement les bases correspondantes (A, T, G ou C).

III- L'ADN

1- Caractéristique de l'ADN

Le ADN présente plusieurs caractéristiques propres et qui l'opposent aux ARN :

- **l'ose : le 2'-désoxyribose** (remplacé par le ribose dans les ARN);
- **les bases : A, C, G et T**. Dans les ARN, T est remplacé par U;

- la molécule d'ADN est constituée de deux chaînes (ou brins) de nucléotides. Les molécules d'ARN sont le plus souvent sous forme simple brin.

D'autre part, la double chaîne d'ADN possède les caractéristiques suivantes :

- elle est **antiparallèle** : les deux brins de la molécule d'ADN sont disposés dans des directions opposées. Un brin est orienté dans une direction 5'→3', et le second brin sera parallèle au premier et dans la direction inverse 3'→5';
- elle est **complémentaire** : l'appariement des bases des deux brins d'une molécule d'ADN se fait suivant la règle de complémentarité: A apparié avec T, C apparié avec G. Cette complémentarité repose sur des raisons stériques (encombrement dans l'espace) et sur la formation de liaisons hydrogène. Les liaisons hydrogène sont formées par l'interaction entre un atome d'hydrogène et deux autres atomes dits électronégatifs (oxygène et azote). Les liaisons hydrogène sont au nombre de deux entre A et T et de trois entre C et G (figure 3);
- elle est **hélicoïdale** : dans l'espace les deux chaînes présentent une configuration hélicoïdale. Elles s'enroulent autour d'un axe imaginaire pour constituer une double hélice à rotation droite.

Figure 3 : Thymine

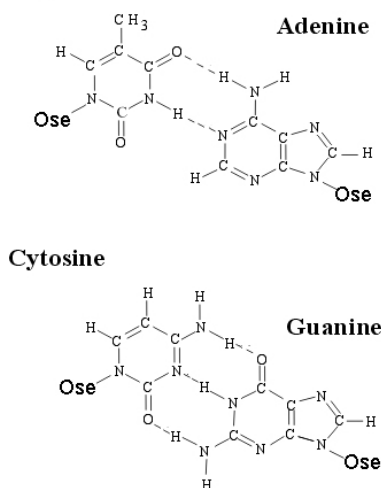
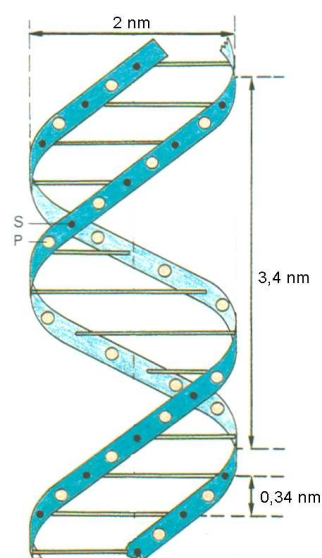


Figure 4 :



Les caractéristiques de l'hélice (figure 4) :

- environ 10 paires de bases par tour de spire;
- l'espacement entre deux bases consécutives est de 0,34 nm;
- le pas de l'hélice est de 3,4 nm;
- le diamètre de l'hélice est de 2 nm;
- les bases puriques et pyrimidiques sont à l'intérieur de l'hélice, les groupements phosphates et les désoxyriboses sont à l'extérieur.

2- Propriétés physico-chimiques de l'ADN

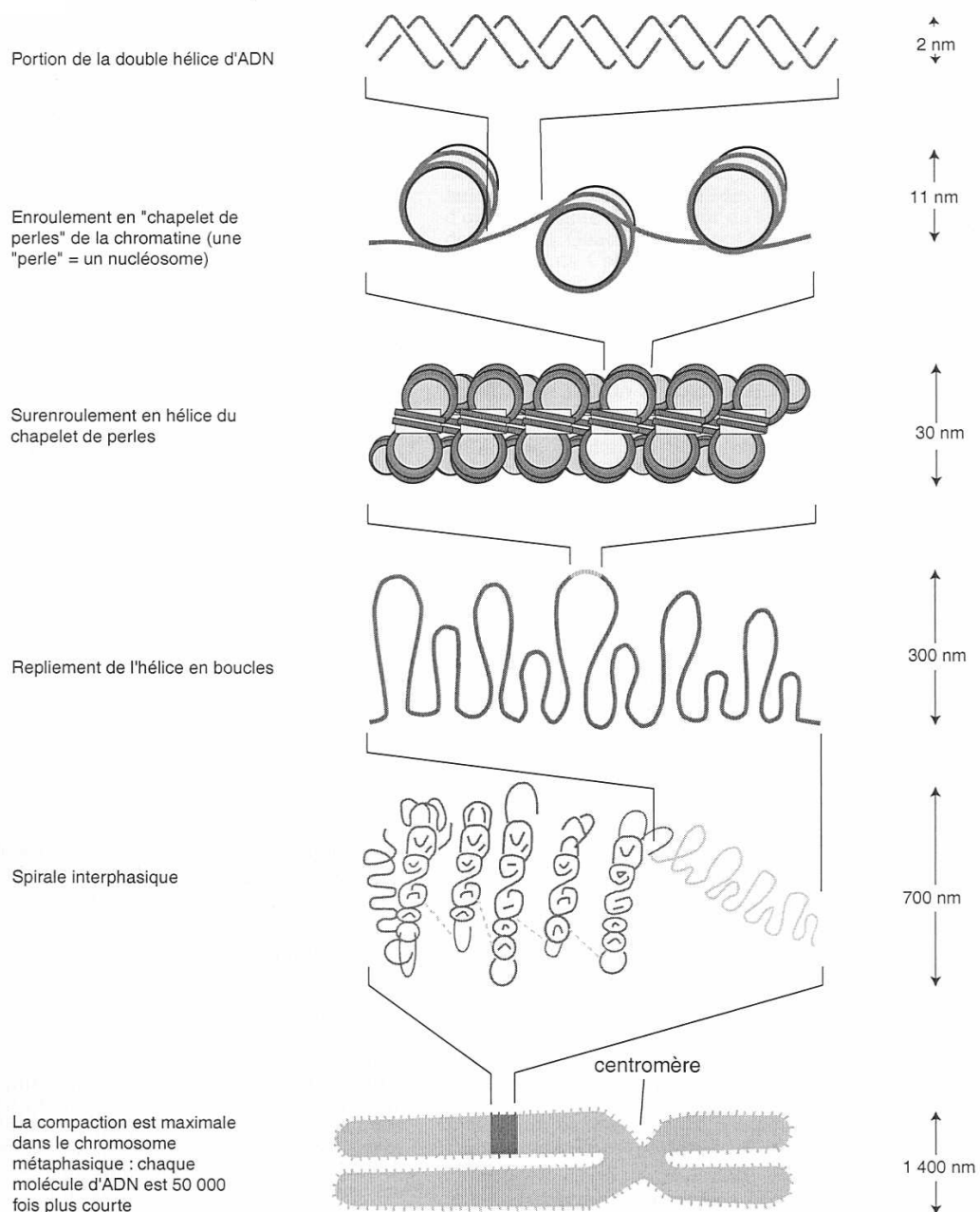
Si on chauffe une solution d'ADN, à une certaine température, les liaisons hydrogène qui assurent la cohésion des deux brins appariés se rompent, les deux brins se séparent, on parle de fusion de l'ADN, caractérisée par la température de fusion (ou T_m). L'ADN est dénaturé. Cette **dénaturation** est cependant réversible dans certaines conditions, les deux brins peuvent se réassocier suivant les règles de complémentarité. La température de fusion varie selon l'ADN étudié. Elle augmente lorsque le pourcentage de bases (G+C) augmente. Ceci est lié au nombre de liaisons hydrogène possibles formées par les bases G et C (3 liaisons hydrogène entre C≡G au lieu de 2 entre A=T).

La présence des bases puriques et pyrimidiques permet aux acides nucléiques (ADN et ARN) d'absorber dans l'ultra-violet (UV) à 260 nm. Les protéines absorbent un peu à 260 nm, mais

surtout à 280 nm. Cette **absorption dans l'UV permet de doser les acides nucléiques** et d'estimer la contamination par les protéines lors de la purification des acides nucléiques.

Dans les cellules, la molécule d'ADN nucléaire présente différents niveaux d'**enroulement et de compaction** (figure 5). Elle est fortement associée à des protéines pour constituer la chromatine. L'image la plus classique est celle du collier de perles. La molécule d'ADN relie les « perles » qui sont des complexes protéines-ADN appelés **nucléosomes**. Le nucléosome contient environ 200 paires de bases d'ADN associées à des protéines appelées **histones**. Les histones sont des protéines de petit poids moléculaire (11-14 kDa), riches en acides aminés basiques. Dans un nucléosome, elles sont au nombre de 8 avec deux exemplaires de chacune des histones: H2A H2B, H3 et H4. Au niveau d'un nucléosome, l'ADN (200 pb) est donc associé à un octamère d'histone. L'histone H1 n'appartient pas au nucléosome, mais interviendrait dans le contact entre deux nucléosomes. Ce collier de perles peut subir des enroulements successifs avec formation de structures de type solénoïde.

Figure 5 : Enroulement et compaction de l'ADN dans le chromosome.



3- Hybridation de l'ADN avec des sondes nucléiques

L'hybridation est une propriété fondamentale des acides nucléiques qui repose sur les règles de complémentarité. Il est possible d'apparier des brins d'ADN (ou d'ARN) avec des oligonucléotides ou polynucléotides qui reconnaissent spécifiquement des séquences sur les brins d'ADN de manière anti-parallèle et complémentaire. Ces oligo- ou polynucléotides sont appelés sondes nucléiques ou amorces selon leur utilisation.

IV- Les ARN

1- Caractéristiques générales des ARN

Les ARN présentent plusieurs caractéristiques propres et qui les opposent à l'ADN :

- **l'ose : le ribose** (à la place du 2'-désoxyribose présent dans les ADN);
- **les bases pyrimidiques et puriques qui sont A, C, G et U.** U remplace donc le T présent dans l'ADN;
- **une seule chaîne nucléotidique** au lieu de deux dans les ADN. On dit que les ARN sont monocaténaire (ou simple brin). Cette unique chaîne est plus courte que les chaînes d'ADN. Cependant, dans une même chaîne d'ARN des portions peuvent être sous forme bicaténaire (ou double brin) avec un appariement suivant la règle : A apparié avec U (deux liaisons hydrogène) et C apparié avec G (trois liaisons hydrogène). Dès lors, au niveau des appariements, on aura la constitution de sortes de tiges alors dans les régions non appariées, des boucles apparaîtront (voir schéma des ARNt).

2- Les différents types d'ARN

Les cellules contiennent essentiellement quatre types d'ARN.

a- Les ARN ribosomiques (ou ARNr)

Les ribosomes sont des organites intra-cellulaires situés dans le cytoplasme et qui sont l'usine de fabrication des protéines de la cellule (voir la partie du cours sur la traduction). Les ARNr jouent un rôle essentiel dans la structure et le maintien de l'intégrité des ribosomes en association avec les protéines ribosomales.

b- Les ARN de transfert (ou ARNt)

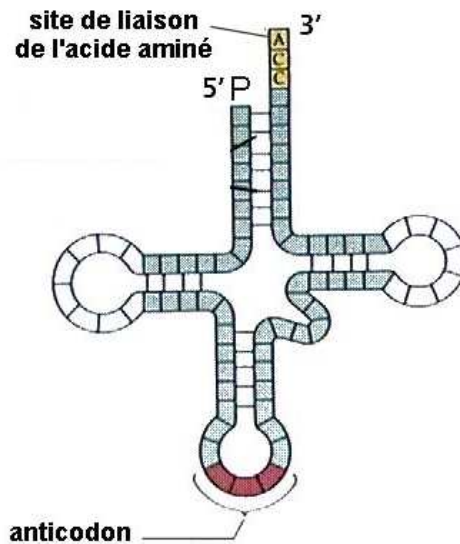
Les ARNt présentent bien entendu la structure générale des ARN, mais ils possèdent en plus quelques particularités propres : certaines portions de la séquence nucléotidique s'apparient entre elles pour former un ARN double brin. En structure spatiale, on présente souvent les ARNt sous forme de trèfle.

Les sites importants dans les tARN (figure 6) :

- l'extrémité 3'-OH est caractérisée par trois nucléotides CCA-3'-OH. C'est par cette extrémité que sera fixé l'acide aminé qui sera véhiculé par le ARNt.
- l'anticodon qui correspond à un groupe de trois nucléotides (ou triplet) situé sur une boucle du ARNt. Ce triplet présente un rôle très important puisqu'il doit s'apparier avec le codon correspondant présent sur l'ARN messager. Cet appariement entre l'anti-codon et le codon se fait par des liaisons hydrogène et suivant les règles de complémentarité. Le codon et l'anti-codon sont disposés de manière antiparallèle.

Les ARNt jouent un rôle capital dans la biosynthèse protéique. La fixation de l'acide aminé à transporter sur le ARNt spécifique sera décrite en détail dans la partie du cours concernant la traduction.

Figure 6 :



a- Les ARN messager (ou **ARNm**)

Les ARN messagers constituent le support essentiel de l'information génétique entre l'ADN et le ribosome où s'effectuera la synthèse protéique. Les ARNm sont très rapidement synthétisés et dégradés, leur durée de vie est très courte.

Ils sont formés d'une seule chaîne de nucléotides avec les bases communes aux ARN : A, U, C et G. Cette chaîne comporte une succession de triplets nucléotidiques. Chaque triplet nucléotidique constitue un codon spécifique d'un acide aminé donné (voir la partie du cours sur la traduction).

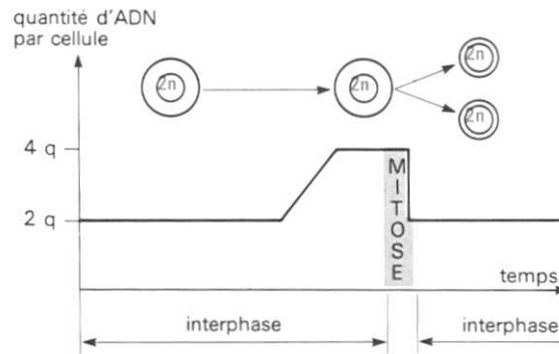
b- Les petits ARN nucléaires (ou **snRNA** pour small nuclear RNA)

Ils sont présents dans le noyau des cellules et sont impliqués dans certaines étapes de la transcription (copie de l'ADN en ARN messager).

La réplication de l'ADN

Lors de la **mitose** (ou division cellulaire, voire Annexe I), une cellule-mère donne deux cellules-filles, et il est essentiel que l'ADN présent dans les cellules-filles soit la copie identique de l'ADN présent dans la cellule-mère. Cette copie de l'ADN est indispensable à réaliser avant la mitose : c'est la **réplication** de l'ADN. Préalablement à toute division cellulaire, la quantité d'ADN est multipliée par deux (figure 7). Les mécanismes de réplication conditionnent donc le déroulement de la division cellulaire.

Figure 7 :



I- Caractéristiques générales de la réplication

La réplication est :

- **semi-conservatrice**. Ceci signifie que sur les deux brins d'ADN, on a toujours un brin d'ADN qui provient d'un des deux brins de l'ADN parental et un brin nouvellement formé. A chaque réplication, les deux brins d'ADN parental se séparent, chacun de ces deux brins sert de matrice pour la synthèse d'un brin complémentaire;
- **bidirectionnelle**, elle va progresser dans les deux sens à partir du point d'ouverture de la double hélice;
- **complémentaire** et progresse dans le sens 5' → 3';
- **discontinue** entre les deux brins d'ADN.

II- Eléments nécessaires à la réplication de l'ADN

- une **matrice d'ADN** constituée par un brin parental;
- des **amorces ARN**;
- des **nucléotides** propres à l'ADN, c'est-à-dire contenant du 2'-désoxyribose, des bases A, T, G et C et sous forme de nucléotides triphosphates (dATP, dTTP, dCTP et dGTP);
- la présence de **nombreux enzymes, dont l'ADN polymérase**;
- la présence de **certaines ions** (notamment le Mg^{2+} , cation indispensable à l'activité de l'ADN polymérase).

III- Les différentes étapes de la réplication de l'ADN (figure 8)

La réplication de l'ADN commence simultanément en de nombreux points précis appelés **origines de réplication**.

En ces points, une enzyme, l'ADN hélicase va catalyser le **déroulement de la double hélice** d'ADN, ce qui définira la future **fourche de réplication**.

Les **brins séparés de l'ADN sont stabilisés sous forme simple brin** grâce à la fixation de protéines appelées SSB (pour « single strand binding »). Ces protéines SSB empêchent les deux brins d'ADN de se réapparier.

Une enzyme appelée primase **synthétise une petite amorce d'ARN** (environ 10 nucléotides). Cette amorce va servir de point de départ à l'ADN polymérase pour la **synthèse du brin complémentaire d'ADN dans le sens 5'→3'**. En effet, l'ADN polymérase est incapable de commencer la synthèse d'une chaîne d'ADN sans amorce.

Les **amorces d'ARN seront ensuite détruites** et hydrolysées par l'intervention d'une enzyme, la RNase H. Elles seront remplacées par des fragments d'ADN. **Les lacunes engendrées par l'enzyme RNase H sont comblées** par une ADN polymérase. Finalement, les **fragments d'ADN deviennent contigus et sont soudés** les uns aux autres par l'intervention d'une enzyme qui est une DNA-ligase.

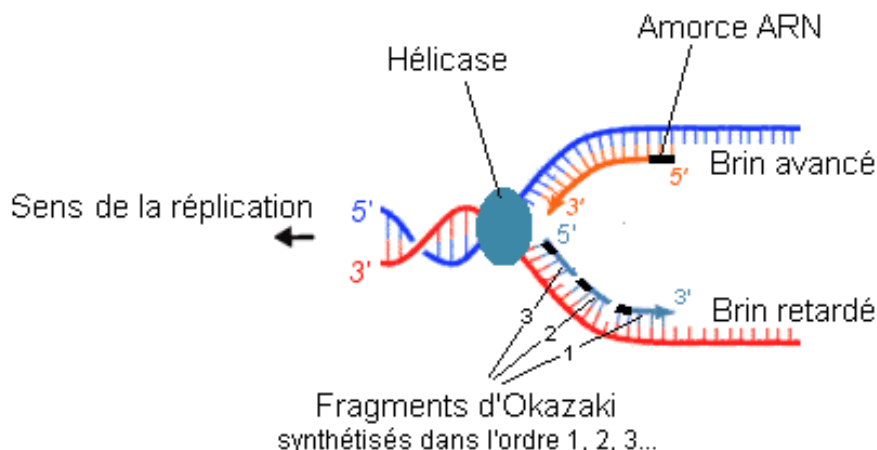
IV- Discontinuité de la réplication entre les deux brins d'ADN (figure 8)

La réplication n'est pas identique entre les deux brins d'ADN. En effet, sur l'un des brins la réplication s'effectue de manière **continue**, on parle de **brin avancé**, alors que sur l'autre brin elle s'effectue de manière **discontinue**, on parle de **brin retardé**.

Sur le brin avancé, la progression de la réplication se fait dans le sens 5'→3' en utilisant comme modèle le brin d'ADN orienté dans le sens 5'→3'. Sur le brin retardé, des petits fragments d'ADN sont synthétisés, encore appelés **fragments d'Okazaki**. Chaque fragment est synthétisé dans le sens 5'→3', le brin modèle correspondant de l'ADN est orienté de manière anti-parallèle 3'→5'. On voit donc que l'allongement discontinu de ce brin retardé se fait dans le sens de la propagation de la réplication.

Il est important de comprendre que le brin qui sert de matrice de lecture pour la constitution du brin retardé doit former une boucle autour d'un des deux sites actifs de l'ADN polymérase. Dans ces conditions, l'intervention de la primase et de l'ADN polymérase permet la synthèse d'un brin nouveau d'ADN dans le sens 5'→3' par copie au niveau de cette boucle. Après la synthèse d'ADN sur le brin retardé, la boucle est défaits et une nouvelle est reformée au niveau de l'ouverture de la fourche de réplication. Finalement, des fragments d'ADN sont ainsi synthétisés sur le brin retardé; de manière discontinue.

Figure 8 :



V- Réparation des erreurs

Des altérations de l'ADN peuvent provenir d'**erreurs commises par la polymérase durant la réplication**. L'ADN polymérase présente, outre l'activité de synthèse $5' \rightarrow 3'$, une activité exonucléasique $3' \rightarrow 5'$ qui lui permet de vérifier si la dernière base introduite correspond bien aux conditions classiques d'appariement. Cette activité d'édition diminue considérablement les erreurs possibles. Au cours de la réplication, il est estimé que l'ADN polymérase incorpore une base incorrecte (erreur d'appariement) tous les 10^4 à 10^5 nucléotides, mais avec les fonctions d'édition, finalement le taux d'erreur passe à 1 pour 10^9 nucléotides incorporés.

La transcription de l'ADN en ARN messenger

La transcription constitue l'ensemble des mécanismes par lequel l'ARN messenger (ARNm) est synthétisé. **L'ARNm est une copie d'une portion de l'ADN.** Seules certaines portions de l'ADN sont transcrites, ces séquences d'ADN sont appelées **gènes**. Enfin, seul l'un des deux brins d'ADN est copié en ARNm. La transcription constitue l'étape préliminaire essentielle pour la biosynthèse protéique (ou traduction).

I- Caractéristiques générales de la transcription

La synthèse d'un ARNm à partir d'ADN s'effectue toujours :

- dans le **sens 5'→3'**;
- de manière **anti-parallèle** par rapport à la portion d'ADN copiée;
- de façon **complémentaire** (appariements G/C et A/U).

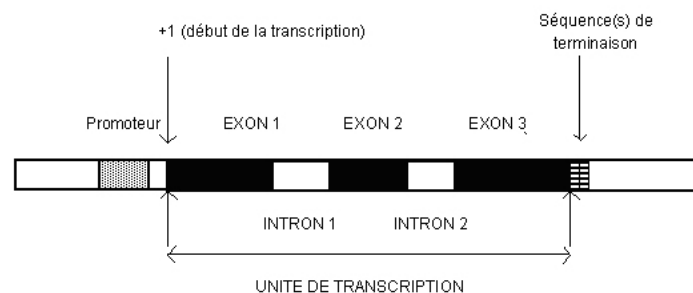
II- Les éléments nécessaires à la transcription

- une **matrice d'ADN** servant de modèle. Ce modèle est indispensable à l'enzyme, on parle souvent d'ARN polymérase-ADN dépendante;
- des nucléotides propres au ARNm, les **ribonucléotides**, contenant du ribose, des bases A, U, G et C sous forme de ribonucléotides triphosphates (ATP, UTP, CTP et GTP);
- une enzyme : l'**ARN polymérase**;
- des **ions Mg²⁺** indispensables au fonctionnement de l'ARN polymérase.

III- Les différentes étapes de la transcription (figure 10)

La transcription ne concerne qu'une portion de l'ADN, les gènes. La séquence d'ADN qui va être transcrite (ou **unité de transcription**) comporte une **origine** (départ) et une **séquence de terminaison** (fin) de la transcription. L'ARN synthétisé qui correspond à cette unité de transcription s'appelle le **transcrit primaire** ou **ARN pré-messager** (pré-ARNm) (figure 9).

Figure 9 :



La synthèse du pré-ARNm comprend trois phases successives : l'**initiation** (reconnaissance de la séquence du promoteur et du point de départ de la transcription), l'**élongation** (synthèse de l'ARNm), et la **terminaison** (reconnaissance de la séquence de fin de transcription).

1- L'initiation de la transcription

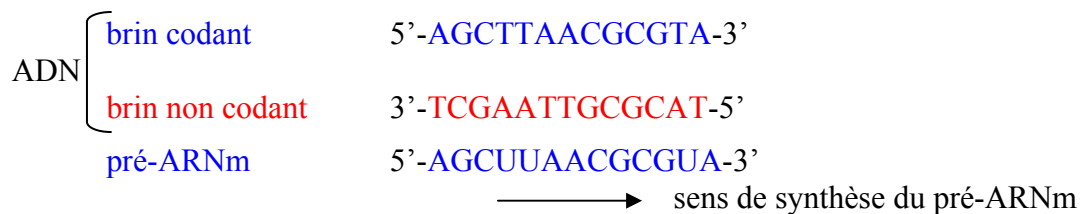
Cette étape implique la reconnaissance par l'ARN polymérase d'une région de l'ADN, le **promoteur**, qui se trouve juste avant le début de la région d'ADN à transcrire. Par

convention, on donne le chiffre +1 au premier nucléotide à partir duquel la transcription commence. L'ARN polymérase se fixe ensuite sur l'ADN et sépare localement les deux brins, créant ainsi une courte région simple brin qui servira de matrice pour l'appariement des premiers ribonucléotides.

2- L'élongation de la chaîne poly-nucléotidique au cours de la transcription

Une fois les premiers ribonucléotides incorporés, l'ARN polymérase va se déplacer le long de l'ADN en séparant les deux brins. L'incorporation des ribonucléotides continue dans le sens 5'→3', et le pré-ARNm se détache de la matrice ADN au fur et à mesure de sa synthèse. Les liaisons hydrogène se reforment après passage de l'ARN polymérase et les deux brins d'ADN reprennent leur forme hélicoïdale.

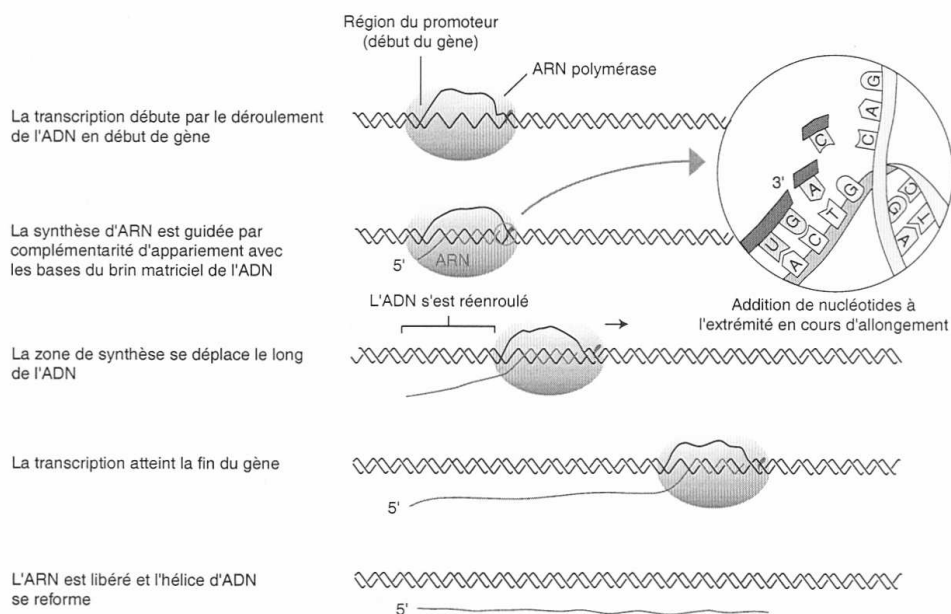
Remarque : l'un des brins d'ADN est appelé **brin codant** (ou brin +, ou brin sens, ou brin non transcrit), l'autre est appelé **brin non codant** (ou brin -, ou brin anti-sens, ou brin transcrit ou brin matrice). C'est le brin non codant qui va être parcouru par l'ARN polymérase et qui va servir de matrice pour la transcription, de telle sorte que, par le jeu de la complémentarité, le pré-ARNm synthétisé sera une copie du brin d'ADN codant :



3- La terminaison ou fin de la transcription

Un signal de poly-adénylation (séquence 5'-AATAAA-3' sur le brin sens) annonce la terminaison de la transcription d'un gène. L'ARN polymérase reconnaît ce signal sur l'ADN mais poursuit la transcription. Une enzyme (endonucléase) reconnaît la séquence correspondante sur le pré-ARNm (5'-AAUAAA-3') et coupe la séquence environ 20 bases en aval. Le pré-ARNm sera ensuite remodelé (modifications post-transcriptionnelles).

Figure 10 : La transcription de l'ADN en ARN (d'après Berg et Singer 1993).



IV- La maturation des produits de transcription

La plupart des gènes codant pour des protéines présentent une structure discontinue, comportant des **exons** et des **introns**. Les exons sont les séquences d'ADN qui seront traduites en protéines (on dit aussi qu'elles seront exprimées). Les introns sont des séquences non codantes d'ADN intercalées entre les exons.

La transcription correspond à l'étape de copie d'un gène et à la formation d'un transcrit primaire (ou pré-ARNm). Ce transcrit primaire correspond à une copie intégrale des exons et des introns d'un gène. Pour que ce pré-ARNm devienne un ARNm et puisse servir de matrice pour la synthèse de protéines, il doit subir plusieurs phases de **maturation** : addition d'une coiffe, addition d'une queue polyA et élimination des introns.

1- L'addition de la coiffe à l'extrémité 5'

Cette phase survient dès la formation des premières liaisons nucléotidique et consiste en un **blocage de l'extrémité 5' du pré-ARNm** en voie de synthèse par fixation d'un ribonucléotide GTP en position inversée (liaison 5'-5' au lieu de 3'-5'). Ce ribonucléotide particulier est appelé coiffe. L'extrémité 5' du pré-ARNm coiffé est ensuite l'objet de **plusieurs méthylations** qui protègent l'extrémité 5' du pré-ARNm de l'attaque par des enzymes de dégradation (exonucléases).

2- L'addition de poly(A) à l'extrémité 3'

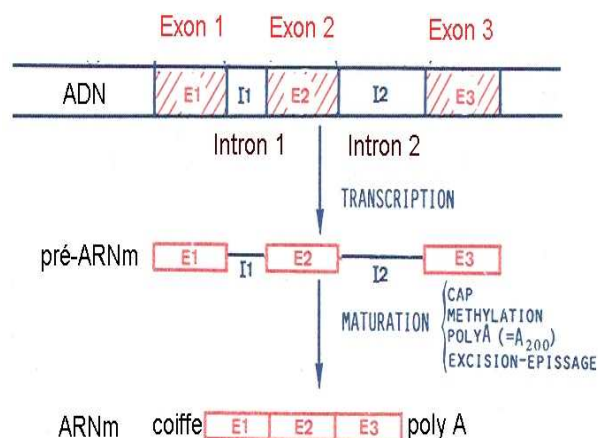
La deuxième phase intervient après la terminaison de la transcription du pré-ARNm et correspond à une **coupeure de l'extrémité 3'**. Cette coupeure est immédiatement suivie par le **transfert de 30 à 300 ribonucléotides ATP à l'extrémité 3'**. On parle de polyadénylation de l'extrémité 3'. La présence de cette queue polyA aurait une fonction de protection de l'extrémité 3' du pré-ARNm contre l'attaque des exonucléases.

3- L'élimination des introns (figure 11)

La troisième phase, appelée **excision-épissage**, permet la maturation du transcrit primaire en ARNm par élimination des introns :

Figure 11 :

1. coupeure de l'intron à son extrémité 5';
2. formation d'une boucle ou lasso par formation d'une liaison entre l'extrémité 5' libérée et un résidu adénylique interne de l'intron (site de branchement);
3. jonction des deux exons et élimination de l'intron.



Généralement, une cellule peut épisser le transcrit primaire de différentes façons, et donner ainsi plusieurs protéines différentes à partir d'un seul gène (ainsi une protéine A peut être constituée des exons 1, 2 et 3 et une protéine B des exons 1 et 3 uniquement). Ce processus est appelé **épissage alternatif** ou **épissage différentiel**.

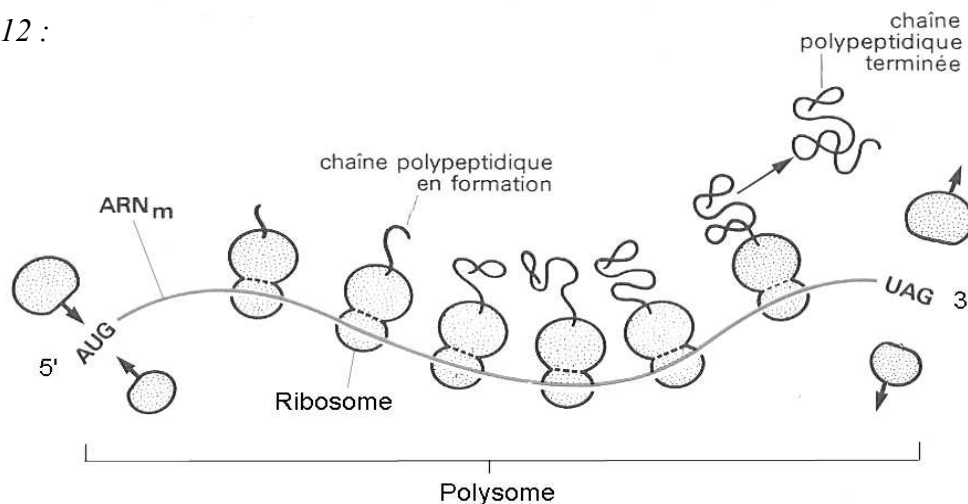
La traduction

La traduction est le **mécanisme de biosynthèse des protéines**. La traduction s'effectue dans le cytoplasme au niveau du ribosome qui traduira l'information contenue sur les ARNm et synthétisera les protéines correspondantes.

I- Caractéristiques générales de la traduction

- les ribosomes lisent l'ARNm dans le sens 5'→3';
- la synthèse protéique se déroule de l'extrémité N-terminale à l'extrémité C-terminale de la protéine;
- la traduction se réalise au niveau des **polysomes** (figure 12). Un polysome correspond à un ensemble comprenant un ARNm et plusieurs ribosomes. Chaque ribosome réalise sa propre synthèse protéique en décodant l'ARNm. Chaque ARNm peut donc être décodé par plusieurs ribosomes à la fois.

Figure 12 :



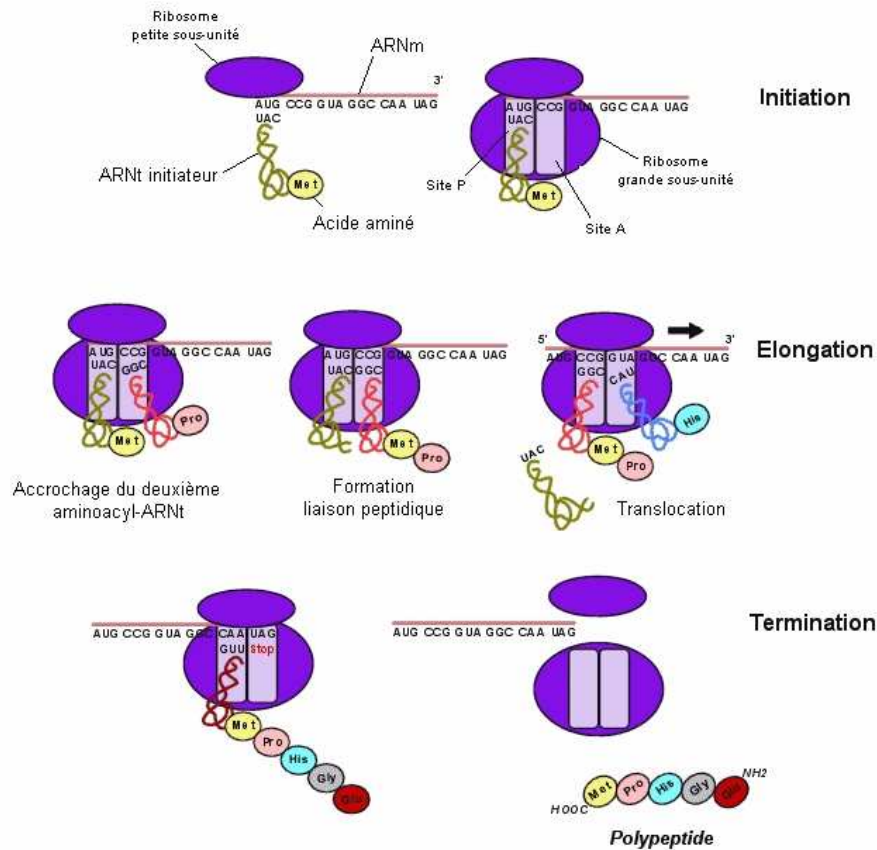
II- Les éléments nécessaires à la traduction

- l'**ARNm** qui indique, par un codage nucléotidique, l'ordre (ou séquence) des acides aminés qui seront incorporés dans la chaîne polypeptidique par le ribosome.
- les **ribosomes** qui sont des organites cytoplasmique constitués par l'assemblage de deux sous-unités protéiques (une grande et une petite) et d'ARNr;
- les **ARNt** qui servent d'adaptateurs entre l'ARNm et l'acide aminé. Ils possèdent un site d'accrochage des acides aminés à leur extrémité 3' (séquence 5'-CCA-3') et un site appelé anti-codon qui reconnaît le codon complémentaire sur l'ARNm ;
- les **acides aminés** qui vont se liés les uns aux autres (liaison peptidique) pour constituer la protéines;
- une enzyme, l'**aminoacyl-ARNt synthase**, qui lier l'aminoacyl (ou acide aminé) à l'ARNt qui lui correspond. C'est ce complexe aminoacyl-ARNt qui participera à la synthèse du peptide par le ribosome.

III- Mécanisme de la traduction

La traduction comprend trois étapes successives : l'**initiation**, l'**élongation** et la **terminaison** (figure 13).

Figure 13 :



1- Initiation de la chaîne peptidique

Il existe sur l'ARNm, près de l'extrémité 5' phosphate, un **codon initiateur AUG**. Ce codon, qui code pour la méthionine, est le premier à être reconnu par le ribosome. La méthionine est donc le premier acide aminé à être incorporé dans la chaîne polypeptidique. Il sera plus tard supprimé de la chaîne polypeptidique.

Ainsi, au début de l'initiation de la traduction, les deux sous-unités du **ribosome** sont dissociées. Puis, la petite sous-unité forme un complexe avec l'ARNm (au niveau du codon initiateur) et l'ARNt portant la méthionine initiale (**ARNt initiateur**). La grande sous-unité va alors se fixer à cet ensemble. Le ribosome est alors complet. Les ribosomes possèdent deux sites pour fixer les ARN de transfert :

- un **site A** (A pour acide aminé) sur lequel l'ARNt porteur du nouvel l'acide aminé à incorporer viendra se fixer;
- un **site P** (P pour peptide) sur lequel l'ARNt porteur de la chaîne peptidique en cours d'élongation est fixé.

Initialement, la méthionine est fixée par son ARNt dans le site P du ribosome.

2- Elongation de la chaîne peptidique

L'élongation de la chaîne peptidique se déroule en trois étapes:

- première étape : **accrochage du deuxième aminoacyl-ARNt sur le ribosome.**

Un deuxième aminoacyl-ARNt (le premier étant la méthionyl-ARNt) se fixe sur le site A de la grande sous-unité du ribosome. Le codon situé après le codon d'initiateur AUG détermine l'anticodon qui va se fixer, et donc l' aminoacyl-ARNt;

- deuxième étape : **formation d'une liaison peptidique.**

Il y a tout d'abord rupture de la liaison entre la formyl-méthionine et le premier ARNt. Ce dernier est éjecté du ribosome. Puis, il y a formation d'une liaison peptidique entre le groupement -COO^- (carboxylique) de l'acide aminé 1 (méthionine) et le groupement -NH_3^+ du deuxième de l'acide aminé 2. L'enzyme qui catalyse la formation de cette liaison peptidique

est une peptidyl-transférase. Cette activité enzymatique est portée par la grande sous-unité du ribosome. Finalement à la fin de cette phase, on a un dipeptide (peptide constitué de deux acides aminés) qui est logé dans le site A de la grande sous-unité du ribosome. Le site P du ribosome est alors libre;

- troisième étape : **la translocation**

Après la formation de la nouvelle liaison peptidique, le deuxième acide aminé (toujours lié à son ARNt) se déplace du site A vers le site P. Le ribosome se déplace donc d'un cran (d'un codon) sur l'ARNm dans la direction 5'→3'. Un nouveau codon (codon 3) se trouve alors en face du site A.

3- Terminaison de la chaîne peptidique

La fin de la traduction se produit quand le ribosome qui se déplace le long de l'ARNm rencontre **codon-stop** : UAG, UGA ou UAA qui ne codent pour aucun acide aminé (voir paragraphe III- Le code génétique). Il se produit alors une coupure entre le dernier ARNt et le dernier acide aminé de la chaîne polypeptidique. La chaîne polypeptidique est libérée et les deux sous-unités du ribosome se dissocient.

IV- Le code génétique

Le code génétique correspond à l'enchaînement ordonné de trois bases nucléotidiques (**triplet ou codon**) permettant de définir un acide aminé. L'enchaînement des triplets le long de l'ARNm définit la séquence en acides aminés de la protéine qui est synthétisée lors de la traduction.

1- Le code génétique est défini par trois lettres (tableau 2)

Le code génétique représente un alphabet de quatre lettres (ribonucléotides) permettant d'écrire des mots de trois lettres (codon ou triplet). Le nombre de codons possibles est donc de $4 \times 4 \times 4 = 64$. Sur ces **64 codons disponibles** :

- 3 codons (UAA, UAG, UGA) sont des codons qui ne peuvent pas être traduits en acides aminés. Ces codons sont appelés non-sens ou codons-stop et provoquent l'arrêt de la synthèse peptidique par les ribosomes.

- 61 codons correspondant aux 20 acides aminés. Mis à part le tryptophane et la méthionine qui disposent d'un seul codon, les 18 autres acides aminés sont définis individuellement par plusieurs codons (2 à 6).

Tableau 2 :

1ère base	2ème base				3ème base
	U	C	A	G	
U	UUU } <i>Phe</i> UUC } UUA } <i>Leu</i> UUG }	UCU } <i>Ser</i> UCC } UCA } UCG }	UAU } <i>Tyr</i> UAC } UAA } Stop UAG } Stop	UGU } <i>Cys</i> UGC } UGA } Stop UGG } <i>Trp</i>	U C A G
C	CUU } <i>Leu</i> CUC } CUA } CUG }	CCU } <i>Pro</i> CCC } CCA } CCG }	CAU } <i>His</i> CAC } CAA } <i>Gln</i> CAG }	CGU } <i>Arg</i> CGC } CGA } CGG }	U C A G
A	AUU } <i>Ile</i> AUC } AUA } <i>Met</i> AUG }	ACU } <i>Thr</i> ACC } ACA } ACG }	AAU } <i>Asn</i> AAC } AAA } <i>Lys</i> AAG }	AGU } <i>Ser</i> AGC } AGA } <i>Arg</i> AGG }	U C A G
G	GUU } <i>Val</i> GUC } GUA } GUG }	GCU } <i>Ala</i> GCC } GCA } GCG }	GAU } <i>Asp</i> GAC } GAA } <i>Glu</i> GAG }	GGU } <i>Gly</i> GGC } GGA } GGG }	U C A G

2- Le code génétique est universel

Le code génétique est dit **universel** : il le même pour tous les organismes aussi bien chez les eucaryotes que chez les procaryotes.

3- Le code génétique est dégénéré

Le code génétique est dit **dégénéré** car un même acide aminé peut être codé par plusieurs codons différents (cf tableau). Dans beaucoup de cas, les triplets nucléotidiques codant pour un même acide aminé ne diffèrent que par la troisième base. C'est là un facteur de stabilité génétique car une mutation ponctuelle sur la troisième base du codon n'entraînera le plus souvent aucun changement d'acide aminé (on parle alors de mutation silencieuse).

Les deux premières bases d'un codon de l'ARNm sont rigoureusement complémentaires de l'anticodon de l'ARNt. En revanche, pour la troisième base du codon l'appariement avec la première base de l'anticodon peut se faire de manière non classique (on parle alors de flottement). Il n'est donc pas nécessaire qu'il y ait, dans la cellule, autant d'ARNt différents qu'il y a de codons différents : un même ARNt peut reconnaître plusieurs codons correspondants à un même acide aminé.

4- Code non chevauchant

Le code génétique est dit **non chevauchant**. La partie codante d'un ADN est transcrite en ARNm puis traduite régulièrement triplet par triplet. Les triplets nucléotidiques sont donc lus d'un point de départ (initiation de la synthèse protéique) jusqu'à un point de terminaison et ceci sans chevauchement.

5- Notion de cadre de lecture

Le **cadre de lecture** représente l'enchaînement des triplets le long d'une portion d'ARNm. Pour une séquence ribonucléotidique donnée, il existe donc trois cadres de lecture différents.

Si on considère par exemple la séquence suivante d'un ARNm :

ACG ACG ACG ACG ACG ACG ACG ACG ...

Les trois cadres de lecture possibles (en bleu) sont :

ACG ACG ACG ACG ACG ACG ACG ACG ...

A CGA CGA CGA CGA CGA CGA CGA CGA ...

AC GAC GAC GAC GAC GAC GAC GAC GAC ...

Un cadre de lecture est dit **ouvert** (ou ORF "Open Reading Frame") si tous les triplets qui le constituent codent pour des acides aminés. Une séquence ribonucléotidique qui est traduite en protéine a un cadre de lecture qui commence avec un codon d'initiation (AUG codant pour la méthionine). Elle comprend ensuite une série de triplets jusqu'à l'un des trois codons de terminaison possibles (codons-stop).

Un cadre de lecture qui ne peut pas être lu en protéine parce que les codons-stop se présentent fréquemment est dit **bloqué ou fermé**. Si cette séquence est bloquée dans les différents cadres de lecture possibles, elle ne peut pas avoir la fonction de coder pour une protéine.

En conséquence, une séquence qui code pour une protéine possède obligatoirement un cadre de lecture ouvert parmi les trois possibles. Quand la séquence d'une région d'ADN de fonction inconnue est obtenue, chaque cadre de lecture doit être analysé pour déterminer s'il est ouvert ou bloqué.

Le polymorphisme de l'ADN

Le polymorphisme peut être défini comme l'existence de deux ou plusieurs formes alléliques pour un même locus entre individus d'une même espèce. Il est possible d'identifier des polymorphismes de l'ADN à différentes échelles, allant du niveau chromosomique (remaniements) au niveau nucléotidique (mutations ponctuelles).

I- Types de polymorphisme

Le polymorphisme provient de modifications de la séquence d'ADN, les mutations, qu'on peut classer selon l'étendue de la lésion de l'ADN : les macrolésions et les microlésions.

1- Macrolésions de l'ADN : les remaniements chromosomiques

Ce type de polymorphisme concerne des altérations de zones plus ou moins importantes des chromosomes. On peut distinguer :

- les **translocations**;
- les **inversions** (changement d'orientation tête-bêche d'un segment variable d'ADN);
- les **fusions** de gènes;
- les **délétions** ou **insertions** de séquences d'ADN (amputations ou additions de matériel génétique d'amplitude variable).

Toutes insertions ou délétions d'un nombre de base non multiple de trois entraîne une modification du cadre de lecture. Ce type de mutation fait généralement apparaître en aval un codon non-sens conduisant à la formation d'une protéine tronquée. Une variation d'un nombre de bases multiple de trois maintient le cadre de lecture, mais entraîne la perte ou l'addition d'acides aminés.

2- Microlésions de l'ADN : les mutations ponctuelles

Ce type de mutation, qui n'intéresse qu'une région très limitée du génome (quelques nucléotides maximum), est la source principale de polymorphisme (une étude chez l'homme donne une estimation de une base polymorphe tous les 200 à 1000 nucléotides).

Les modifications peuvent être de différentes natures :

- **transitions** : substitution d'une base purique par une autre base purique (A/G), ou d'une base pyrimidique par une autre base pyrimidique (C/T);
- **transversions** : substitution d'une base purique par une base pyrimidique, ou d'une base pyrimidique par une base purique (A,G/C,T);
- **délétions** de base;
- **insertions** de base.

Il existe 4 transitions et 8 transversions possibles. Dans les régions codantes, ces substitutions peuvent modifier la nature de l'acide aminé codé, en fonction principalement de la position de la substitution dans le codon. En effet, compte tenu de la dégénérescence du code génétique, 95% des mutations de la première base, 100% des mutations de la seconde et 28% des mutations de la troisième entraînent un changement d'acide aminé. Dans le cas contraire on parle de mutation silencieuse (pour la troisième base, 90% des transitions et 30% des transversions sont silencieuses).

Les délétions ou insertions d'une base conduisent à une modification du cadre de lecture.

3- Séquences répétées en tandem (VNTR : Variable Number of Tandem Repeats)

Les VNTR sont constituées d'un nombre variable de répétitions en tandem d'un motif de taille variable. En fonction de la taille du motif et de la répétition, on distingue :

- les **satellites** : le motif comporte 2 à plus de 100 nucléotides et le nombre de répétition peut dépasser les 1000;
- les **minisatellites** : le motif comporte de 6 à 24 nucléotides et le nombre de répétition est de 20 à 50;
- les **microsatellites** : le motif comporte 1 à 4 nucléotides et le nombre de répétition ne dépasse généralement pas 25.

La fonction des VNTR reste inconnue, sauf dans le cas des minisatellites de l'ADN télomérique où une fonction supposée est de protéger l'extrémité des chromosomes de la dégradation et d'en permettre la réplication.

II- Mécanismes mis en œuvre dans la génération du polymorphisme

1- Erreurs de réparations

Ce mécanisme explique l'apparition de mutations ponctuelles.

L'ADN subit des modifications de sa composition nucléotidiques principalement dues :

- aux erreurs commises par l'ADN polymérase lors de la réplication. En effet, l'ADN polymérase commet des erreurs d'appariement lors de la synthèse du nouveau brin d'ADN, et bien qu'un système de réparation des erreurs existe, celui-ci est imparfait et on estime le taux d'erreur (après correction) à environ 1 base tous les 10^9 nucléotides incorporés.

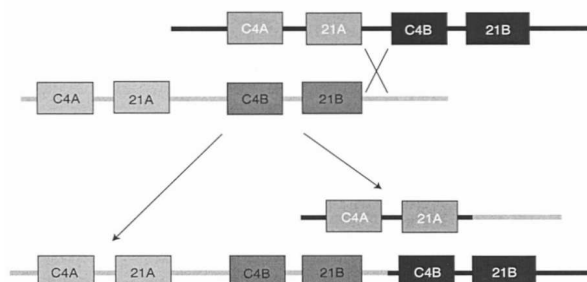
A titre d'exemple, au cours d'une vie humaine, on a environ 10^{17} divisions cellulaires soit environ 10^{17} mutations ponctuelles somatiques. En ce qui concerne la lignée germinale, les mutations ponctuelles sont estimées à environ 20 pour un ovule et à environ 250 pour un spermatozoïde (potentiellement transmises à la descendance !!!).

- à des agressions chimiques ou physiques. L'ADN génomique peut être altéré par exposition à de multiples mutagènes (UV, rayons X, agents chimiques...) qui peuvent provoquer des cassures des brins d'ADN, la formation de liaisons entre nucléotides, la suppression d'une base... La plupart de ces altérations sont généralement reconnues et corrigées par les systèmes enzymatiques de réparation de la cellule, mais ceux-ci ne sont pas infaillibles.

2- Recombinaisons inégales (figure 14)

Les délétions et les duplications peuvent résulter de recombinaisons entre séquences très semblables mais non homologues. Ces séquences favorisent l'apparition d'erreurs d'alignement, conduisant à des crossing-over inégaux. Ce mécanisme aboutit à l'élimination d'une région sur l'un des chromosomes et à la duplication sur l'autre chromosome. Ces recombinaisons sont favorisées par l'existence de séquences semblables contiguës.

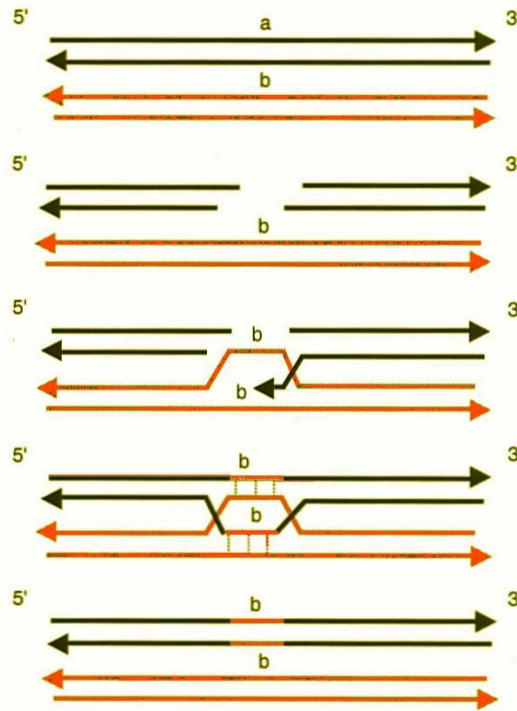
Figure 14 :



3- Conversion génique

La conversion génique est un mécanisme de recombinaison permettant un transfert d'information génétique qui conduit au remplacement d'une séquence d'ADN par une autre apparentée. Ce remplacement a lieu lors de la réparation d'une lésion de l'ADN par la polymérase (figure 15 : c'est un fragment apparenté "b" qui sert de matrice pour la réparation de la partie manquant "a")

Figure 15 :



4- Insertions de séquences mobiles (transposons) ou virales

Certaines séquences d'ADN répétitif possèdent une structure de transposon (éléments d'ADN instable) capable de migrer vers différentes régions du génome par rétro-transposition (les transcrits ARN peuvent être convertis dans la cellule en ADN complémentaire par une enzyme reverse transcriptase, et ces ADN s'insèrent au sein de l'ADN génomique). C'est le même phénomène qui se produit avec les virus à ARN.

5- Glissement intra-chromatidien

Le glissement intra-chromatidien modifie le nombre de répétition des VNTR. Il se produit par glissement de l'ADN polymérase consécutif à un mésappariement lors de la réplication : la polymérase "dérage" et est alors incapable de reproduire fidèlement le nombre de répétitions d'origine.

En conclusion, l'ADN recèle donc un important polymorphisme résultant soit d'accidents purement aléatoires, soit de mécanismes favorisés par certaines structures du génome (VNTR). Ce polymorphisme génère une importante variabilité et est le moteur de l'évolution.

Les outils et techniques de la biologie moléculaire

La biologie moléculaire appliquée à la génétique a pour but l'étude d'une région du génome et la recherche d'un gène, voire d'une mutation. Les techniques de génétique moléculaire, reposant sur un nombre limité d'outils, ont donc pour objectif de cibler le fragment intéressant au sein d'une molécule d'ADN des millions de fois plus grande. Le facteur "quantité" est le plus souvent limitant : la quantité d'ADN contenue dans une cellule.

I- Les outils de la biologie moléculaire




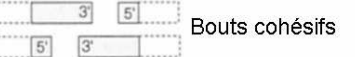
1- Les enzymes

Les principales enzymes utilisées en biologie moléculaire sont :

- **les enzymes de restriction** (tableau 3)

Les enzymes de restriction sont issues des bactérie (l'abréviation enzymatique est déterminée par la source bactérienne). Elles reconnaissent un site spécifique de l'ADN (souvent 4 ou 6 paires de bases palindromiques), et coupent la molécule d'ADN dans ce site. Elles permettent d'obtenir, à partir d'une molécule entière, des fragments d'ADN dont la longueur et le nombre vont dépendre de la fréquence et de la répartition des sites de restriction sur la molécule d'ADN d'origine. La coupure peut générer des bouts francs ou des bouts cohésifs. Elles sont utilisées dans de nombreuses techniques de biologie moléculaire, comme la PCR-RFLP, les AFLP, le clonage...

Tableau 3 :

Source bactérienne	Abréviation enzymatique	Séquence 5' -> 3' 3' -> 5'	Extrémités générées
<i>Haemophilus aegyptius</i>	<i>Hae</i> III	GG CC CC GG	 Bouts francs
<i>Staphylococcus aureus</i> 3A	<i>Sau</i> 3AI	GATC CTAG	
<i>Bacillus amyloliquefaciens</i> H	<i>Bam</i> HI	G GATC C C CTAG G	
<i>Providencia stuartii</i>	<i>Pst</i> I	C TGCA G G ACGT C	 Bouts cohésifs

- **les enzymes de digestion autres que les enzymes restriction**

Ces enzymes coupent les molécules d'acides nucléiques (ADN ou ARN), mais contrairement aux enzymes de restriction, elles ne nécessitent pas la reconnaissance de sites spécifiques. On distinguera les exonucléases qui digèrent les acides nucléiques par les extrémités, et les endonucléases qui digèrent les acides nucléiques par l'intérieur. Les endonucléases sont spécifiques d'un type d'acide nucléique : ARN simple brin, ARN hybridé à de l'ADN, ADN simple brin, ADN double brin.

- **les polymérases**

les polymérases synthétisent le brin complémentaire d'un brin d'ADN à partir de nucléotides libres présents dans le milieu réactionnel. Elles sont utilisées dans des techniques telles que la PCR, le marquage de l'ADN (séquençage, génotypage)...

- **les ligases**

Les ligases assemblent des fragments d'ADN linéaires entre eux. Elles créent une liaison entre les extrémités, de séquences compatibles, de deux fragments d'ADN. Elles sont utilisées

notamment dans les techniques de clonage et dans l'addition d'adaptateurs à de l'ADN génomique digéré (AFLP par ex.).

- **les phosphatases**

Les phosphatases éliminent le groupement phosphate à l'extrémité 5' d'un acide nucléique.

- **les kinases**

Les kinases ajoutent un groupement phosphate à l'extrémité 5' d'un acide nucléique.

2- Les amorces et sondes

Les **amorces** sont de courte séquence d'environ 20 nucléotides, complémentaires du début d'une matrice et servant de point de départ à son recopiage par l'ADN polymérase dans la technique d'amplification par PCR.

Les **sondes** sont des séquences d'au moins 15 nucléotides, homologues à une séquence d'ADN avec laquelle elle s'hybride de façon stable et spécifique par association entre bases complémentaires. Les sondes sont marquées à l'aide d'une molécule détectable (radioactivité, biotine, enzyme...) et permettent ainsi d'identifier les séquences d'ADN auxquelles elles s'hybrident.

Divers facteurs affectent l'hybridation moléculaire des sondes et des amorces :

- le temps : plus le temps de réaction est long et plus la probabilité d'appariements pour des séquences complémentaires augmente;
- la concentration en acides nucléiques (ADN, sondes, amorces) : plus elle augmente et plus la probabilité de rencontre augmente;
- autres facteurs : longueur des fragments, présence d'agents chimiques (ex. : DMSO);
- la température : basses températures favorise hybridation non spécifique;
- la complémentarité.

3- Les vecteurs et cellules hôtes

Les **vecteurs** sont des molécules d'ADN circulaire et double brin dans lesquelles un fragment d'ADN exogène (appelé **insert**) peut être inséré. Ces vecteurs, une fois introduit dans une **cellule hôte**, ont la capacité de se répliquer de façon autonome, indépendamment de la réplication de la cellule elle-même. Ils possèdent également des propriétés permettant de sélectionner les cellules qui l'ont incorporées. Ainsi, le vecteur pourra posséder un gène de résistance à un antibiotique et un gène codant pour une protéine facilement détectable. Ces deux gènes permettront de sélectionner (1) les cellules hôtes qui ont incorporées un vecteur, et (2) parmi celles-ci, celles dont le vecteur contient un insert (voir la technique de clonage). Les vecteurs les plus couramment utilisés sont les **plasmides**.

Les **cellules hôtes**, généralement des **bactéries**, subissent un traitement chimique de façon à les rendre **compétente**. La compétence des bactéries peut être définie comme leur capacité à incorporer de l'ADN exogène. La technique consistant à faire pénétrer un fragment d'ADN exogène dans une bactérie s'appelle la **transformation bactérienne**.

II- Les techniques de la biologie moléculaire

Les techniques de biologie moléculaire peuvent être classées selon leurs objectifs, à savoir :

- fragmenter l'ADN (digestion enzymatique, PCR);
- séparer les fragments (électrophorèse, clonage);
- repérer les fragments d'intérêt (coloration, marquage à l'aide de sondes);
- multiplier le fragment à étudier (PCR, clonage);
- lire le fragment (séquençage, génotypage).

1- Extraction –Purification de l'ADN

Pour pouvoir étudier le polymorphisme de l'ADN, il faut dans un premier temps accéder à la molécules d'ADN. Pour ce faire l'ADN doit impérativement être purifié à partir du matériel biologique dans des conditions optimales de **qualité et de quantité**.

La première technique à mettre en œuvre est donc l'extraction-purification d'ADN. Cette technique comporte deux étapes.

La première étape consiste à libérer l'ADN de la cellule, c'est l'étape d'**extraction**

Pour cela on va traiter l'échantillon dont on veut extraire l'ADN avec des procédés qui vont permettre de :

- dissocier les tissus et individualiser les cellules : la **digestion**. La digestion des tissus est réalisée à l'aide d'enzymes vont détruire les protéines liant les cellules les unes aux autres (collagénase, trypsine, protéinase K);
- détruire les membranes cytoplasmique et nucléaire : la **lyse** cellulaire. La lyse des cellules est réalisée à l'aide (1) de détergents anioniques (ex. : SDS) qui vont permettre de désorganiser la structure des membranes (de nature lipoprotéique) et (2) d'enzymes protéolytiques (ex. : protéinase K) qui vont détruire les protéines. On peut également faire éclater les cellules grâce à un tampon hypertonique ou par la chaleur.

L'extraction réalisée, on aura un mélange d'ADN libre et des autres constituants cellulaires (protéines, lysozymes, organites, débris membranaires etc.)

L'étape suivante consistera donc à séparer l'ADN de ces autres constituants cellulaires, c'est l'étape de **purification**. Pour la réaliser, différentes techniques sont disponibles :

- purification **phénol/chloroforme**. Cette technique est basé sur le principe de la solubilité différentielle des molécules entre deux phases non miscibles : l'ADN est soluble dans la phase aqueuse et les contaminants (protéines, lipides) dans la phase organique. Le phénol est un déprotéinisant puissant qui va dénaturer et solubiliser les protéines, et dans lequel l'ADN n'est pas soluble. Le chloroforme va quant à lui permettre d'éliminer les traces de phénol qui auraient ou être emportées avec la phase aqueuse. L'ADN est ensuite récupérer sous forme solide à partir de la phase aqueuse par précipitation à l'alcool (éthanol ou isopropanol). Cette opération permet d'éliminer les molécules qui n'ont pas précipité en même temps que l'ADN (notamment les sels). Le culot ainsi obtenu par précipitation est resuspendu à la concentration voulue dans un tampon Tris/EDTA ou de l'eau.
- purification par utilisation de **kits commerciaux**. Les kits présentés ici sont le kit Puregene de la société Gentra et DNA easy Tissus kit de la société Qiagen. Le principe du kit Puregene est basé sur la précipitation des protéines (qui permet leur élimination), puis la précipitation de l'ADN (élimination des sels et des sucres) et sa resuspension dans un tampon Tris/EDTA ou de l'eau. Le kit Qiagen quant à lui est basé sur le principe des interactions ioniques : l'ADN est d'adsorbé sur une membrane de silice en présence de sels chaotropiques (qui déshydrate les molécules d'acides nucléiques). Les protéines, les lipides et les polysaccharides ne sont pas retenus par la membrane. Après lavage de la membrane, les acides nucléiques sont ensuite élués avec de l'eau ou avec le tampon d'éluion fourni dans le kit (solution aqueuse contenant très peu de sel). On obtient un rendement d'éluion 15 à 20% supérieur en chauffant le tampon d'éluion à 70°C).

Tableau 4 : comparatif des 3 méthodes (chaque critère est noté sur une échelle de 1 à 3, 1 étant le moins bon et 3 le meilleur) :

	Phénol chloroforme	Kit Puregene	Kit Qiagen
Quantité/rendement	2	3	1
Pureté de l'extrait	3	1	2
Rapidité	1	2	3
Coût	3	2	1
Sécurité du manipulateur	1	3	3

Remarque : la technique phénol/chloroforme en elle-même a un faible coût, mais celui augmente considérablement si on intègre les coûts d'achat et de maintenance des équipements de sécurité et la gestion des déchets (le phénol et chloroforme sont très toxiques pour le manipulateur et l'environnement).

Le **dosage** de l'ADN extrait peut être :

- effectué par une mesure au **photomètre**, l'ADN absorbant fortement dans l'ultra-violet à 260 nm. Une unité de densité optique à 260 nm (DO_{260}) correspond à $50\mu\text{g}/\mu\text{l}$ pour de l'ADN double brin. Une éventuelle contamination protéique peut être mesurée à 280 nm (DO_{280}), et le rapport DO_{260}/DO_{280} est alors calculé. Pour une solution d'ADN pure, ce rapport doit être compris entre 1,8 et 2,1. Une contamination par le phénol peut également être recherchée par mesure de l'absorbance à 270 nm (le phénol inhibe la PCR);
- estimé par **migration électrophorétique sur gel d'agarose** et comparaison avec un ADN de concentration connue.

2- Migration électrophorétique et visualisation des fragments d'ADN

L'électrophorèse consiste à **séparer des fragments d'ADN**, qui migre dans un gel soumis à un champ électrique, **en fonction de leur taille** (plus la taille est élevée et moins le fragment migrera loin dans le gel et inversement, les plus petits fragments auront la distance de migration la plus importante). Ceci est possible car la molécule d'ADN est chargée négativement, elle se déplacera donc vers le pôle positif (anode) de la cuve de migration. La détection de l'ADN est ensuite réalisée par **coloration** au bromure d'éthidium (BET, agent chimique qui s'intercale entre les brins de l'ADN), puis par exposition aux UV. Les UV excitent le colorant qui émet alors une fluorescence rose/orange.

La migration et la séparation des fragments d'ADN vont être réalisés dans un gel d'**agarose** ou d'**acrylamide** selon la résolution souhaitée (tableau 5) :

- gel d'agarose : en fonction du pourcentage d'agarose dans le gel, des fragments d'ADN de quelques dizaines à quelques milliers de paires de bases peuvent être discriminés. Il est utilisé pour la visualisation de produits PCR;
- gel d'acrylamide : il est plus résolutifs que le gel d'agarose et permet de séparer, en fonction du pourcentage d'acrylamide dans le gel, des fragments des quelques dizaines à quelques centaines de paires de bases, à la base près. Il est utilisé pour la séparation fine de fragments d'ADN de taille voisines : séquençage, analyses de polymorphismes de longueur (microsatellites, AFLP).

Tableau 5 : résolution des gels d'agarose et d'acrylamide en fonction du % respectivement d'agarose et d'acrylamide :

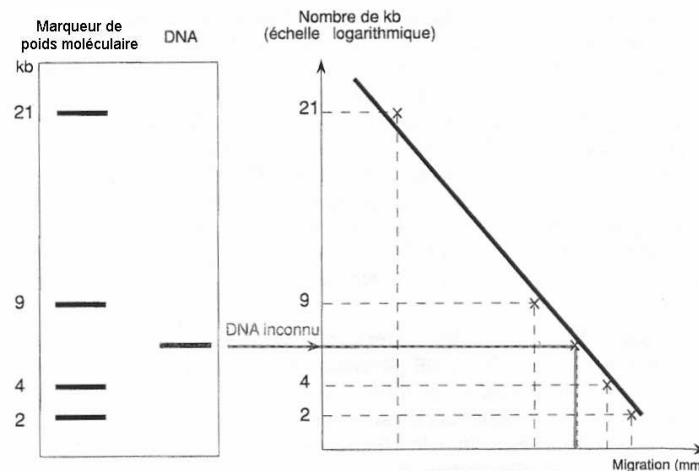
Agarose (%)	Size of fragments separated (kb)*	Acrylamide* (%)	Size of fragments separated (bp)
0.5	1-30	3.5	100-1000
0.7	0.8-12	5.0	80-500
1.0	0.5-10	8.0	60-400
1.2	0.4-7	12.0	40-200
1.5	0.2-3	15.0	25-150
2.0	0.05-2	20.0	6-100

Après coloration, cette technique permet de **vérifier quantité, qualité et la taille de l'ADN**. C'est une technique de contrôle qui peut être mise en œuvre à diverses étapes des différents protocoles de biologie moléculaire. Ainsi, elle va permettre de visualiser l'ADN :

- après extraction-purification : évaluation de la quantité et de la qualité;
- après amplification par PCR : évaluation de la taille, de la quantité et de la spécificité de l'amplification;
- obtenus après digestion par des enzymes de restriction : nombre de fragments, évaluation de leur taille.

La **détermination de la taille** se fait à l'aide d'un marqueur de poids moléculaire qui est composé de plusieurs fragments d'ADN de taille connue. Il est déposé dans le gel et migre en même temps que les échantillons (figure 16)

Figure 16 :



L'évaluation de la **quantité** d'un ADN est réalisée par comparaison entre l'intensité du fragment à doser et les intensités d'une gamme de concentration connues d'un fragment témoin.

L'évaluation de la **qualité** d'un d'ADN est réalisée en observant le profil de migration. Un extrait d'ADN génomique non dégradé (coupé) doit présenter, sur gel, une seule bande de haut poids moléculaire. Si l'ADN est dégradé, on observera des bandes supplémentaires de plus petit poids moléculaire.

3- PCR (Polymerase Chain Reaction ou réaction de polymérisation en chaîne)

La PCR est l'**amplification in vitro** d'une séquence d'ADN double-brin (**ADN matrice**) par extension de deux **amorces**, situées de part et d'autre de la région considérée, grâce à une

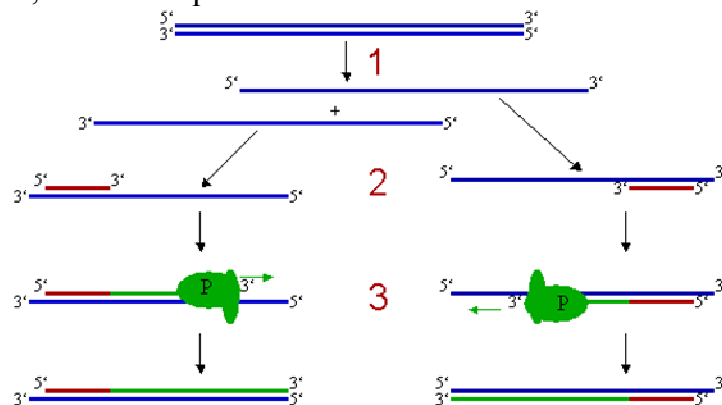
ADN polymérase, en présence de **nucléotides** et d'ions Mg^{2+} . L'amplification est effectuée par la **répétition de cycles** qui assure une duplication exponentielle de chaque brin.

Cette technique permet d'augmenter considérablement la quantité d'ADN ciblé. Elle nécessite de connaître les séquences en nucléotides qui délimitent la région d'ADN à amplifier. Ces séquences serviront à définir des amorces complémentaires de celles-ci, et qui seront les points de départ de l'amplification par la polymérase

La PCR comporte des cycles successifs. **Chaque cycle comprend une succession 3 phases** (figure 17) :

- (1) une phase de **dénaturation** par la chaleur qui sépare les deux brins d'ADN;
- (2) une phase d'**hybridation des amorces**;
- (3) une phase d'**élongation** au cours de laquelle la polymérase synthétise, à partir d'une amorce, le brin complémentaire.

Figure 17 :



Cette technique a pris un essor considérable avec la découverte d'une polymérase thermorésistante issue d'une bactérie thermophile *Thermus aquaticus*, d'où son nom de *Taq* polymérase. Elle a permis une automatisation des différents cycles (thermocycleur). Cette polymérase n'a pas d'activité exonucléase 3'-5', elle n'est donc pas capable de corriger les erreurs d'incorporation de nucléotides (un A au lieu d'un G en face d'un C par exemple). On estime le taux d'erreur à 1 mauvaise incorporation toutes les 10 000 ou 100 000 bases.

De **nombreux paramètres peuvent influencer le résultat de l'amplification** :

- la température d'hybridation et la spécificité des amorces;
- la concentration en ions Mg^{2+} ;
- le nombre de cycle;
- la présence d'inhibiteur;
- l'ajout d'adjuvants.

Pour plus d'information, se reporter au protocole " CBGP-BM0009_Polymerase Chain Reaction".

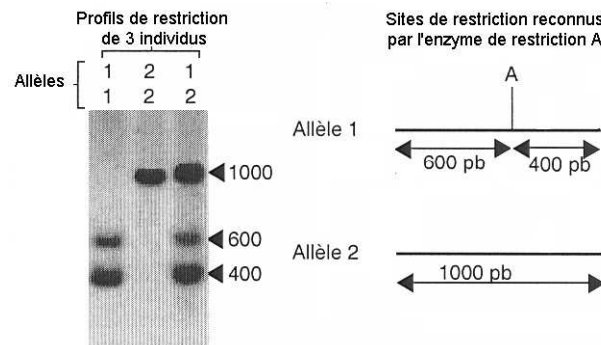
Les utilisations des produits PCR sont très variées. En voici quelques unes qui vont être détaillées par la suite :

- analyse de restriction : PCR-RFLP
- introduction dans un vecteur pour isoler les différentes formes alléliques d'un gène : clonage de produits PCR
- séquençage
- génotypage microsatellite
- génotypage AFLP
- génotypage SSCP

4- PCR-RFLP (Polymerase Chain Reaction-Restricted Fragment Length Polymorphism ou polymorphisme de longueur des fragments de restriction de produit PCR)

Le produit PCR est soumis à une digestion enzymatique par une enzyme de restriction. Si une mutation ponctuelle modifie le site de restriction initialement présent, la taille des fragments d'ADN obtenus après digestion sera modifiée et observable après migration des fragments d'ADN sur gel (figure 18).

Figure 18 :



5- Purification de produits PCR

La purification des produits PCR consiste à **éliminer les réactifs** ayant servi à réaliser la PCR (amorces, nucléotides, sels, polymérase) et qui risquent d'interférer lors de l'utilisation de ces produits PCR dans d'autres techniques (ex. : séquençage, clonage).

Cette purification peut être réalisée :

- sur amplification PCR mono-bande par colonne (kit PCR Purification Qiagen) ou réaction enzymatique (ExoSap-IT USB Corporation);
- sur amplification multi-bande par colonne (kit PCR Gel Extraction kit de chez Qiagen ou de chez Millipore) après migration et découpe sur gel des fragments d'intérêt dans le cas d'une amplification PCR multi-bande.

Le principe des kits colonne (Qiagen et Millipore) est le même que pour l'extraction-purification d'ADN : une membrane de silice retient spécifiquement l'ADN mais pas les nucléotides, la polymérase, les sels et les amorces (qui sont trop petites pour être adsorbées). Une récupération des fragments d'intérêt sur gel peut être nécessaires si l'amplification par PCR a générée de nombreuses bandes (problème de spécificité, duplication de gène, contamination).

La purification enzymatique est réalisée à l'aide de deux enzymes : l'exonucléase I et une phosphatase alcaline. La première dégrade l'ADN simple brin et la seconde hydrolyse les dNTP. Seul l'ADN double brin reste intact.

6- Séquençage

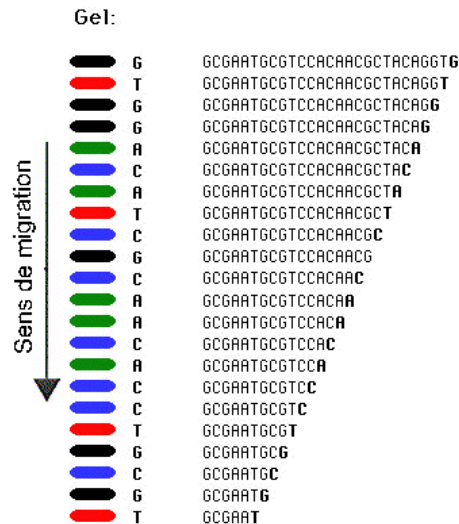
Cette technique enzymatique est basée sur la copie d'un fragment d'ADN simple-brin, que l'on désire séquencer, par une ADN polymérase. On utilise des didésoxynucléotides (ddNTP) couplés à des fluorochrome (4 couleurs différentes, une par ddNTP). Ces dérivés de désoxynucléotides ne possèdent pas de OH à l'extrémité 3' du désoxyribose. L'incorporation de ce ddNTP par la polymérase bloque l'allongement de la molécule d'ADN en cours de copie. En effet, l'allongement d'une chaîne polynucléotidique par une polymérase nécessite qu'un groupement OH soit disponible en 3' (voir les liaisons phosphodiester entre les nucléotides).

Lors de la réaction de séquence, les fragments d'extension sont marqués en 3' par l'incorporation des didésoxynucléotides (ddNTPs) fluorescents (4 couleurs de marquages

différents pour les 4 ddNTPs). Ces ddNTPs bloquent également la réaction d'élongation. Les fragments de différentes tailles ainsi synthétisés sont ensuite séparés par électrophorèse. La fluorescence de chaque fragment est lue à l'aide d'un laser, la succession des couleurs permet alors de déterminer la séquence nucléotidique.

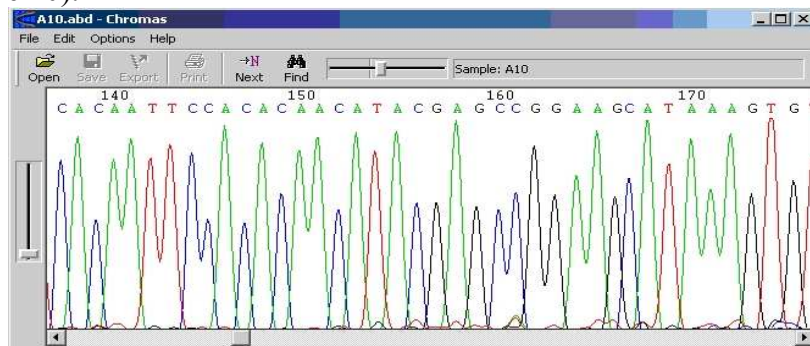
Les fragments de différentes tailles ainsi synthétisés vont ensuite migrer dans un gel d'acrylamide en fonction de leur taille (figure 19).

Figure 19 :



La lecture du gel est réalisée par le balayage automatique d'un laser qui permet de discriminer les 4 bases qui sont couplées à 4 fluorochromes différents (A-Hex vert; T-Rox rouge; C-Fam bleu; G-Ned jaune). Le résultat de ce balayage se présente sous la forme d'un électrophorégramme représentant la succession des bases composant le fragment d'ADN séquencé (figure 20).

Figure 20 :



7- Microsatellites

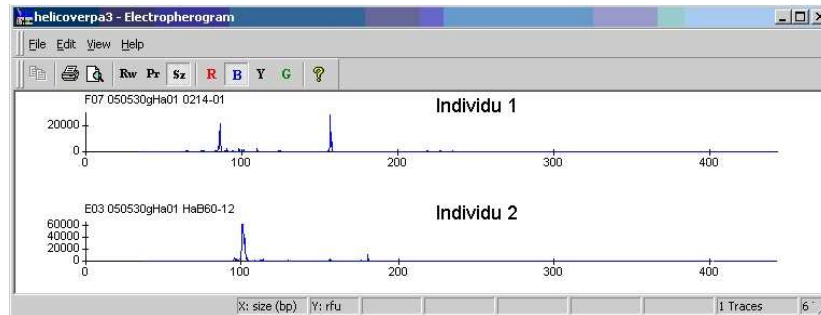
Les microsatellites sont des segments d'ADN contenant des répétitions en tandem de courts motifs di, tri ou tétra-nucléotidiques (le plus typique étant CA/GT). Le polymorphisme des microsatellites réside dans le nombre de répétitions du motif.

La technique consiste à amplifier par PCR une région d'ADN contenant un motif microsatellite. L'une des deux amorces utilisées pour réaliser la PCR est marquée avec un fluorochrome. Le fragment fluorescent obtenu migre ensuite dans un gel d'acrylamide et, comme pour la séquence, la fluorescence est détectée à l'aide d'un laser. On pourra alors observer des différences de taille d'allèle (en fonction du nombre de répétitions du microsatellite) d'un individu à l'autre et déterminer l'état homo ou hétérozygote de ces individus. Les microsatellites sont donc des **marqueurs co-dominants**. Un marqueur co-dominant est un marqueur qui permet de détecter simultanément les différentes versions alléliques d'un gène chez un individu hétérozygote (ex. : PCR-RFLP, microsatellites). Un marqueur dominant ne permet pas de distinguer un individu hétérozygote d'un individu

homozygote car les différentes versions alléliques ne sont pas simultanément détectables (ex. : AFLP).

Dans l'exemple (figure 21) suivant, pour un même locus, deux individus présentent des profils microsatellites différents : l'individu 1 est hétérozygote (un allèle 90 et allèle 150), l'individu 2 est homozygote (2 allèles 100) :

Figure 21 :



8- AFLP (Amplified Fragment Length Polymorphism ou polymorphisme de longueur de fragments amplifiés) (figure 22)

La technique AFLP permet de générer des marqueurs génétiques sans développement préalable (pas besoin de savoir sur quelles portions d'ADN on travaille, donc pas d'obligation de définir des amorces spécifiques pour l'amplification). Son principe est basé sur la détection de bandes polymorphes entre individus dans un **profil multi-bandes** obtenu par digestion de restriction et puis amplification PCR.

Cette technique consiste donc à digérer l'ADN génomique à l'aide de deux enzymes de restriction reconnaissant des sites de restriction différents. Les fragments obtenus, de longueurs différentes, sont ensuite liés à chacune de leurs extrémités à des adaptateurs de séquences nucléotidiques connues. Ces adaptateurs sont spécifiques : ils ne sont compatibles qu'avec les extrémités générées par la digestion d'une enzyme de restriction donnée (par exemple, l'adaptateur EcoRI ne peut se fixer qu'à l'extrémité d'un fragment d'ADN qui a été coupé par l'enzyme de restriction EcoRI). Ces adaptateurs permettent ensuite l'hybridation d'amorces spécifiques (la séquence de l'amorce est complémentaire à celle de l'adaptateur) et l'amplification par une série de PCR. La première **PCR est dite pré-sélective** : on ajoute 1 base d'ancrage aux amorces en plus de la séquence de l'adaptateur, ce qui va permettre de n'amplifier qu'un sous-ensemble des fragments d'ADN présents dans le mélange (la première base du fragment d'ADN, juste après l'adaptateur, est soit un A, un T, un C ou un G, et si on ajoute un C à l'amorce, seuls les fragments commençant par G pourront être amplifiés). La deuxième **PCR est dite sélective** : on ajoute 3 bases d'ancrage aux amorces (dont la première est la même que celle ajoutée pour la PCR sélective) et l'une des amorces est marquée à l'aide d'un fluorochrome. On diminue ainsi la complexité du mélange de fragments, de manière à obtenir un profil constitué de seulement quelques dizaines de bandes distinctes. Les fragments fluorescents obtenus migrent ensuite dans un gel d'acrylamide et leur fluorescence est détectée à l'aide d'un laser. On pourra alors observer des profils multi-bandes caractéristiques des individus (ces profils sont fonction de la présence ou de l'absence des sites de restriction) (figure 23). La technique AFLP génère des **marqueurs dominants**.

Figure 22 :

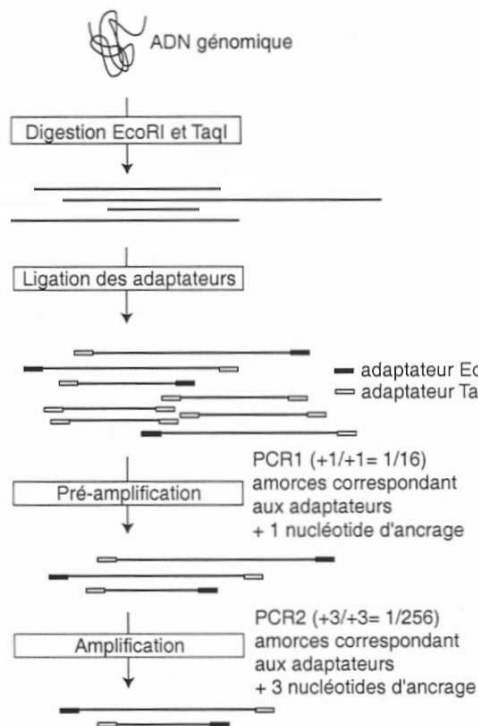
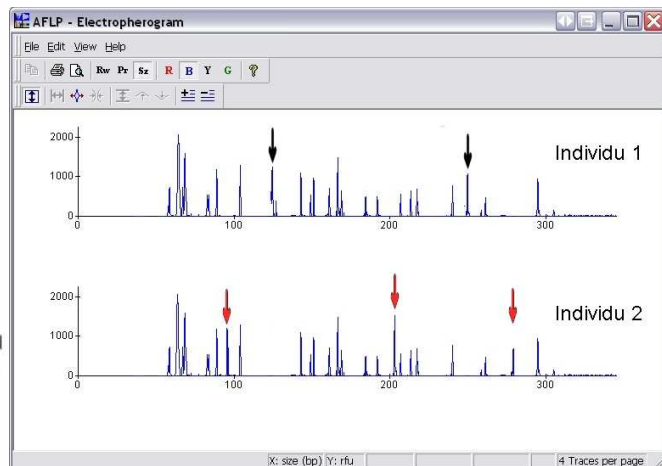


Figure 23 :



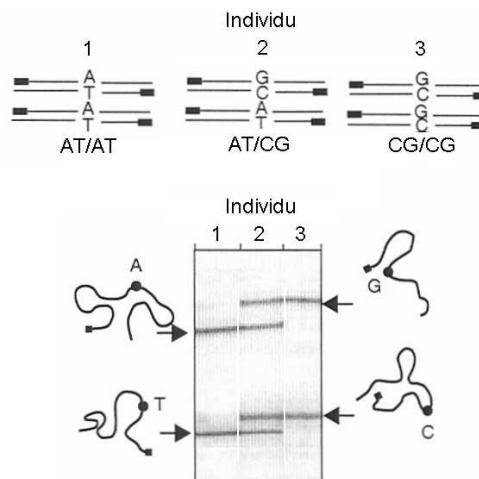
Dans l'électrophorogramme ci-dessus, 2 bandes sont présentes chez l'individu 1 et pas chez l'individu 2 (flèches noires) et inversement 3 bandes sont présentes chez l'individu 2 et pas chez l'individu 1 (flèches rouges).

9- SSCP (Single Strand Conformation Polymorphism ou polymorphisme de conformation simple brin)

La technique SSCP repose sur le fait que la **conformation tridimensionnelle** d'une molécule d'ADN simple brin est fonction de sa séquence nucléotidique : deux allèles d'un même marqueur ne se comporteront pas de la même manière dans l'espace, à l'état simple brin. Ils auront donc des vitesses de migration différentes dans un gel d'électrophorèse non dénaturant (respectant donc la conformation des molécules). Cette propriété est mise à profit pour détecter les **mutations ponctuelles**.

Cette technique consiste à amplifier par PCR une région d'ADN contenant une ou plusieurs mutations ponctuelles. Les deux amorces utilisées pour réaliser la PCR sont marquées avec deux fluorochromes différents. Les fragments fluorescents obtenus migrent ensuite dans un gel d'acrylamide et leur fluorescence est détectée à l'aide d'un laser. On pourra alors observer des différences de vitesse de migration (figure 24). Comme les microsatellites, les SSCP sont des **marqueurs co-dominants**.

Figure 24 :



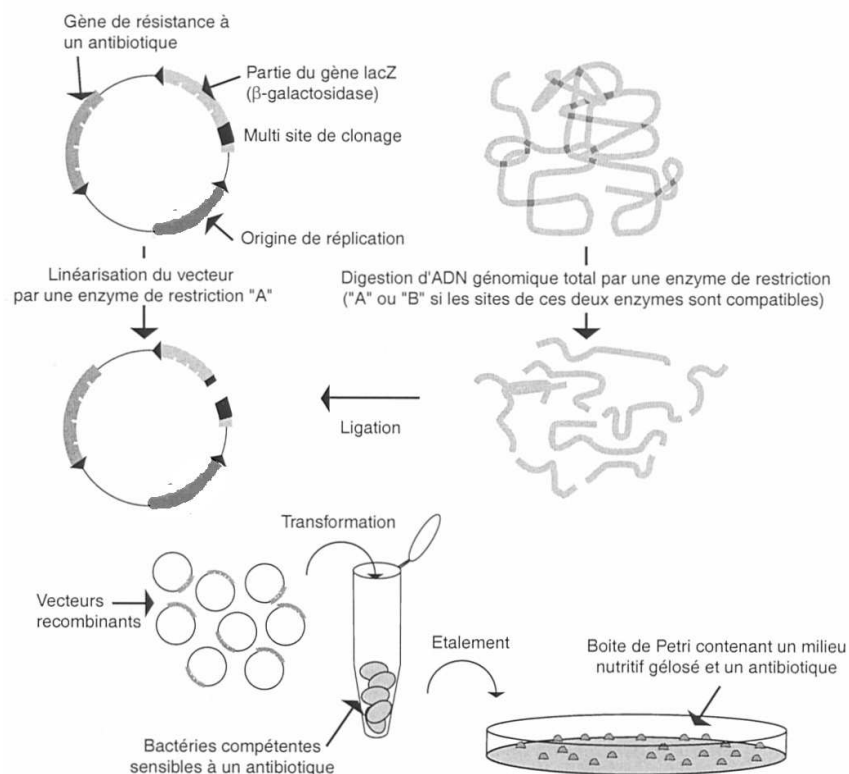
10- Clonage

La technique de clonage est basée sur la **fragmentation de l'ADN génomique** par des enzymes de restriction, puis l'insertion de ces fragments (**inserts**) dans un **vecteur**. Ce vecteur permet alors l'introduction et la multiplication de la séquence dans un **microorganisme hôte** (généralement des bactéries). Chaque cellule hôte intégrera un fragment d'ADN différent et l'ensemble de ces clones constituera une **banque d'ADN génomique**. Le clonage permet donc d'**isoler des fragments d'ADN** et de les **amplifier** en quantités illimitées.

Les différentes étapes du clonage sont :

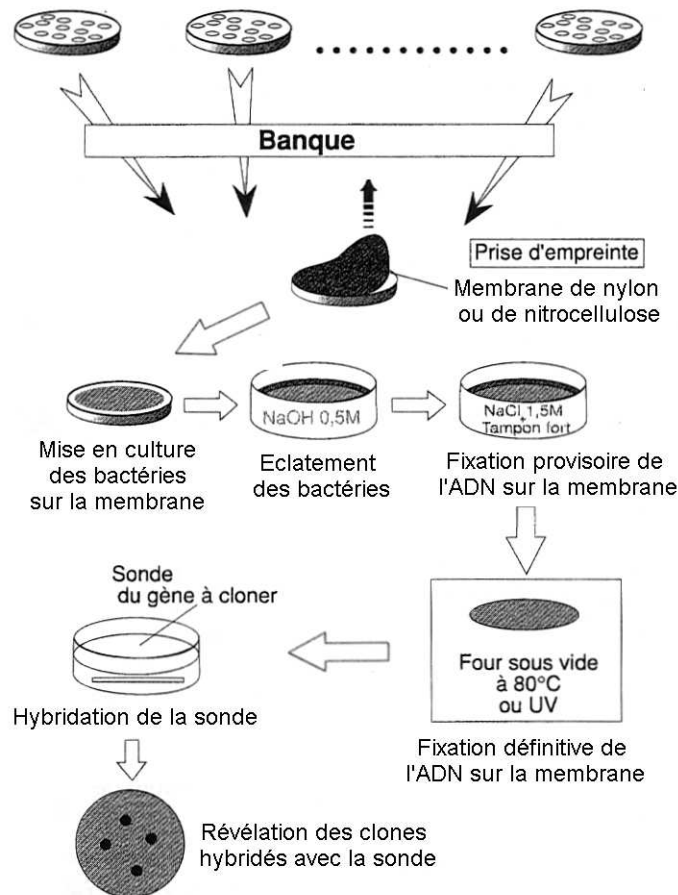
- **la préparation du vecteur (digestion)** (figure 25). Le vecteur doit être linéarisé par coupure à l'aide d'une enzyme de restriction à l'endroit où l'on désire insérer le fragment d'ADN. Les extrémités ainsi libérées doivent être traitées par la phosphatase alcaline de telle sorte que celui-ci ne puisse pas se refermer sur lui-même, seul l'ADN à insérer permettant le raboutage;
- **la préparation de l'ADN à insérer (digestion)** (figure 25). L'ADN est découpé en fragments par une enzyme de restriction créant des extrémités compatibles avec celles du vecteur;
- **la réalisation du recombinant (ligation)** (figure 25). Le vecteur et l'ADN à insérer sont mélangés en présence d'une ligase qui permet la ligation entre les extrémités du vecteur et celles de l'ADN à insérer;
- **l'incorporation à l'hôte (transformation)** (figure 25). Pour se répliquer, le vecteur doit être incorporé dans une cellule hôte;
- **l'isolement des différent clones cellulaires** (figure 25), contenant des inserts différents, par étalement sur un milieu nutritif dans des boîtes de Pétri;
- **la sélection des clones ayant incorporés un fragment d'ADN génomique** (figure 25). Cette sélection est rendue possible par la présence de deux **gènes de sélection** dans le vecteur : (1) un gène qui confère à la cellule hôte une résistance à un antibiotique : les cellules n'ayant pas incorporé un vecteur meurent en présence de l'antibiotique, (2) un gène codant pour une protéine facilement détectable (gène lacZ en générale) : si un insert est présent dans le vecteur, il se trouve au milieu du gène de sélection et la protéine n'est pas synthétisée.

Figure 25 :



- **la recherche du fragment d'intérêt (hybridation)** (figure 26). Parmi tous les clones cellulaires ayant reçu un vecteur contenant un insert, il faut trouver le clone qui contient le fragment d'ADN que l'on veut étudier. Pour ce faire, les clones cellulaires sont transférés sur une membrane et lysés de façon à libérer l'ADN. Un traitement fixe l'ADN vecteur sur la membrane, puis on hybride une sonde marquée (radioactivité, biotine, enzyme...) et complémentaire du fragment recherché. Seuls les clones qui ont incorporé le fragment recherché, et sur lequel la sonde s'est fixée, seront visibles après révélation (clones positifs).

Figure 26 :



L'étude du polymorphisme de l'ADN repose sur une série d'outils en nombre limité, mais qui combinés entre eux permettent d'imaginer et de créer un nombre sans cesse croissant de techniques de biologie moléculaire. Nous n'avons vu ici qu'un certain nombre d'entre elles : cette discipline est en perpétuelle évolution et un grand nombre de nouvelles techniques ont fait leur apparition ces dernières années.

Annexes

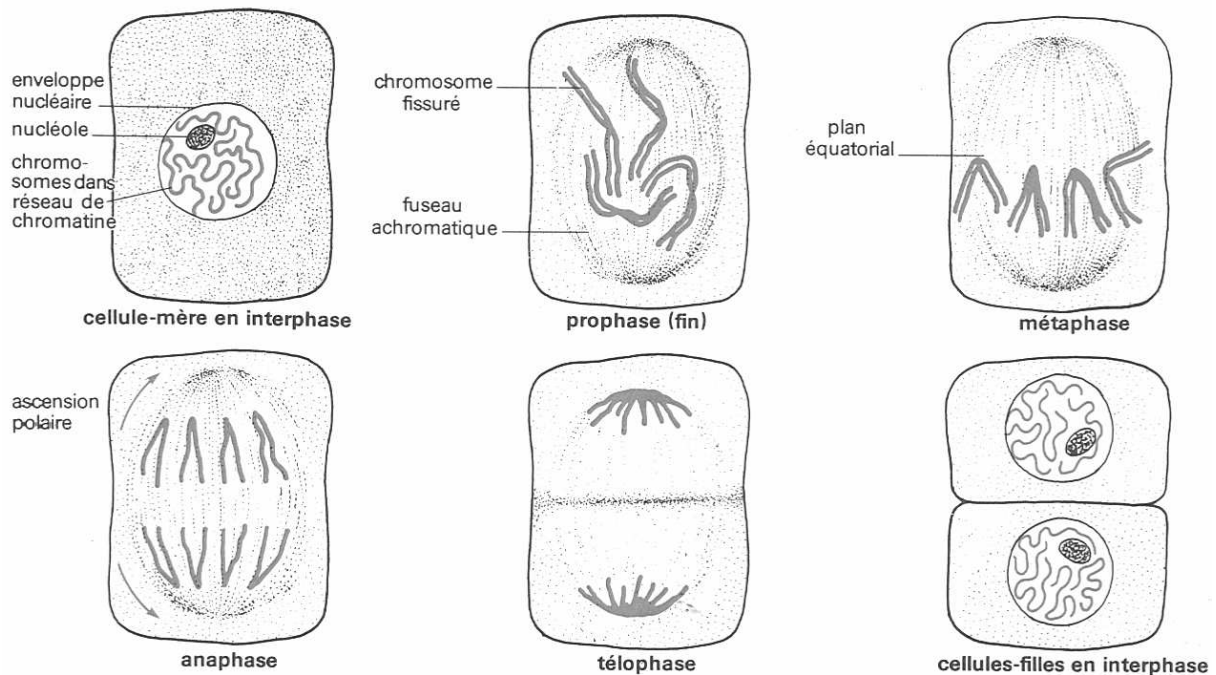
Annexe I : Les différentes phases de la mitose

La mitose, ou division cellulaire, est le mécanisme par lequel se transmet l'information génétique de la cellule mère aux cellules filles. Elle comprend quatre phases :

- Prophase;
- Métaphase;
- Anaphase;
- Télaphase.

Durant l'interphase qui précède la mitose, chaque molécule d'ADN est répliquée mais les deux copies ne se séparent pas totalement; elles restent unies au niveau du centromère.

Au début de la mitose, en **prophase**, on assiste à une extraordinaire condensation des molécules d'ADN. Chaque molécule d'ADN forme alors un bâtonnet très colorable appelé chromatide. Les deux chromatides correspondant aux deux copies d'une même molécule d'ADN demeurent unies au niveau du centromère et forment le chromosome mitotique (ou chromosome fissuré). En **métaphase**, tous les chromosomes se rangent au milieu du noyau puis en **anaphase**, les deux chromatides de chaque chromosome se séparent et s'éloignent l'une de l'autre pour se retrouver aux deux pôles opposés de la cellule. Durant la **télaphase**, le cytoplasme se divise entre les deux lots de chromatides ce qui aboutit à la formation de deux cellules filles ayant reçu chacune une chromatide de chaque chromosome, c'est-à-dire une copie de chacune des molécules d'ADN de la cellule mère, c'est-à-dire encore la totalité de l'information génétique de la cellule mère.



Annexe II : Glossaire

AFLP (Amplified Fragment Length Polymorphism) : polymorphisme de la longueur des fragments d'ADN obtenus après digestion par des enzymes de restriction et amplification par PCR.

Allèles : versions alternatives d'un même gène qui diffèrent par leur séquence nucléotidique.

Amorce : courte séquence de nucléotides complémentaire du début d'une matrice et servant de point de départ à son recopiage par une polymérase.

Anti-parallèle : se dit de l'orientation inverse des brins complémentaires de l'ADN. Il en résulte que l'extrémité 5' d'un brin est en regard de l'extrémité 3' de l'autre brin.

Autosome : chromosome non sexuel.

Bactérie compétente : se dit d'une bactérie ayant subi un traitement lui permettant d'incorporer de l'ADN exogène (transformation).

Banque : collection de clones cellulaires (bactérie par exemple) où chaque cellule renferme un exemplaire différent d'ADN recombiné à un vecteur.

Carte de restriction : ordre de la disposition des sites de restriction présents sur un segment déterminé d'ADN.

Chromatide : copie conforme de chaque chromosome se matérialisant au cours de la métaphase et se séparant de sa copie parentale lors de l'anaphase.

Chromatine : structure associant l'ADN à des protéines (histone) dans le noyau pendant l'interphase.

Clonage moléculaire : recombinaison in vitro d'un fragment d'ADN avec un vecteur se répliquant de manière autonome à l'intérieur d'une cellule hôte (bactérie par ex). La culture de cette cellule contenant le vecteur recombiné permet d'isolement de l'ADN inséré à l'état pur et en quantités illimitées.

Code génétique ou codon : triplet de nucléotides codant pour un acide aminé donné.

Codominance : désigne les caractères héréditaires dont les différentes versions sont simultanément détectables chez un hétérozygote.

Complémentarité : règle universelle d'appariement des bases des acides nucléiques, selon laquelle A s'associe avec T et G avec C.

Crossing-over : échange réciproque de matériel génétique entre chromosomes homologues, survenant en général au moment de la méiose.

Délétion : perte d'une ou plusieurs paires de bases consécutives par rupture de continuité de la molécule d'ADN.

Dénaturation de l'ADN : passage de la forme double-brin à la forme simple brin. Elle est obtenue le plus souvent par la chaleur. Se dit également pour l'abolition de la structure second-tertiaire de l'ADN simple brin.

Déséquilibre de liaison : situation dans laquelle deux allèles correspondant à deux loci distincts d'un même chromosome sont plus fréquemment associés dans une population que ne le voudrait le hasard (voir Equilibre de liaison). Une telle association allélique préférentielle est favorisée par : (1) la proximité physique des loci sur le chromosome; (2) le caractère récent de la mutation ayant produit l'un des allèles; (3) l'existence d'un avantage sélectif.

Diploïde : jeu de chromosome où chaque autosome est en double exemplaire et comportant deux chromosomes sexuels.

Distance génétique : (1) sur une carte génétique désigne un intervalle entre deux loci, calculé d'après la fréquence des recombinaisons observées; (2) sur un arbre phylogénétique désigne l'intervalle de temps écoulé pour permettre d'accumuler les différences observées entre deux séquences homologues dans deux espèces différentes.

Divergence : pourcentage de différences de séquence nucléotidique entre deux ADN apparentés.

Dominant : se dit d'un allèle ou d'une mutation qui, à l'état hétérozygote, conditionne le phénotype.

Empreinte génétique (DNA finger printing) : assortiment d'allèles mis en évidence dans l'ADN génomique grâce à des sondes explorant simultanément plusieurs loci hautement polymorphes (ex. : microsattellites). Permet de caractériser le génotype de chaque individu à des fins d'identification ou de filiation. Encore appelée carte d'identité génétique.

Endonucléases : enzymes coupant la liaison entre deux nucléotides (liaison phosphodiester) à l'intérieur d'un acide nucléique. Ces enzymes sont spécifiques d'un type d'acide nucléique : ARN simple brin, ARN hybridé à de l'ADN, ADN simple brin, ADN double brin.

Enzyme de restriction : endonucléase bactérienne coupant spécifiquement les deux brins d'ADN au niveau d'une séquence, en générale palindromique, parfaitement définie (de 4 à 8 nucléotides).

Elongation d'amorce : extension dans le sens 5' vers 3' par une polymérase d'une amorce d'ADN permettant la copie d'un brin matrice.

Epissage : se dit du mécanisme d'excision des introns et de raboutage des exons lors de la maturation des transcrit.

Equilibre de liaison : se dit de l'absence d'association allélique préférentielle pour deux loci d'un même chromosome, attestée par une distribution conforme aux lois statistiques.

Exon : séquence de gène dont le transcrit persiste dans l'ARN messenger après épissage. Chaque exon représente une séquence codante.

Exonucléases : enzymes détruisant les acides nucléiques de proche en proche à partir d'une ou des deux extrémités.

Gène : séquence de nucléotides contenant l'information pour la production d'une protéine particulière.

Génotype : constitution génétique d'un individu.

Haploïde : jeu de chromosome où il n'existe qu'un seul exemplaire de chaque autosome et un seul chromosome sexuel.

Hétérozygotie : situation génotypique où deux loci homologues d'une même paire de chromosome portent chacun un allèle différent.

Homozygotie : présence du même allèle sur les deux chromosomes d'une même paire.

Hybridation : appariement par complémentarité des bases (A-T et G-C) de deux séquences nucléotidiques complémentaires.

Insert : fragment d'ADN exogène inséré dans un vecteur.

Intron : séquence d'ADN transcrite puis éliminée par épissage lors de la maturation d'un ARN.

Ligation : union par une liaison 3'OH-5'phosphate de deux fragment d'ADN au niveau de deux nucléotides adjacents.

Locus (pluriel : loci) : emplacement d'un segment d'ADN sur un chromosome.

Marqueur génétique : trait génotypique ou phénotypique permettant de distinguer des individus différents.

Méiose : désigne les deux divisions cellulaires particulières (division réductionnelle et division équationnelle) qui constituent le stade ultime de la gamétogenèse.

Microsatellites : segments d'ADN contenant des répétitions en tandem de courts motifs di, tri ou tétra-nucléotidiques (le plus typique étant CA/GT). Le polymorphisme des microsatellites réside dans le nombre de répétitions du motif.

Mutation : désigne n'importe quel changement intervenu dans la séquence d'ADN. S'il ne concerne qu'une seule base, on parle de mutation ponctuelle.

Palindrome : brins d'ADN complémentaire dont la séquence est identique quand elle est lue de gauche à droite sur un brin et de droite à gauche sur l'autre (donc dans les deux cas de 5' vers 3'), ex. :
5' CCATGG 3'
3' GGTACC 5'

PCR (Polymerase Chain Reaction) : amplification in vitro d'une séquence d'ADN double-brin par extension de deux amorces, situées de part et d'autre de la région considérée, grâce à une ADN polymérase. L'amplification est effectuée par la répétition de cycles de dénaturation/hybridation/élongation qui assure une duplication exponentielle de chaque brin.

Phénotype : manifestation apparente de la constitution du génome sous la forme d'un trait morphologique.

Plasmide : ADN circulaire présent dans les bactéries et qui se réplique de façon autonome. Il peut porter des gènes de résistance aux antibiotiques, et est utilisé comme vecteur dans la technique de transformation bactérienne.

Polymorphisme génétique : présence dans une population d'au moins deux variantes alléliques d'un même gène.

Polymorphisme de restriction : variation individuelle de la séquence en bases du génome modifiant un ou plusieurs sites de restriction. Elle donne lieu à des versions alternatives de la taille des fragments d'ADN obtenus avec une enzyme de restriction donnée (techniques PCR-RFLP, AFLP).

Polymorphisme de séquence : toute variation individuelle de la séquence nucléotidique en un site donné du génome.

Recombinaison génétique in vitro : assemblage expérimental de séquence d'ADN (ex. : insertion d'un ADN exogène dans un vecteur).

Renaturation (ou réassociation) de l'ADN : appariement des brins complémentaires d'un ADN préalablement dénaturé.

Récessif : se dit d'un allèle ou d'une mutation n'influençant pas le phénotype à l'état hétérozygote.

Réplication : processus de duplication à l'identique d'une molécule d'ADN en deux molécules filles.

Semi-conservative : désigne le fait qu'au cours de la réplication, chaque brin d'un ADN double-brin sert de matrice pour former une copie complémentaire.

Site de restriction : séquence d'ADN double-brin reconnue et coupée par une enzyme de restriction donnée.

Sonde : séquence d'acide nucléique, d'au moins 15 nucléotides, homologue à une séquence d'ADN avec laquelle elle s'hybride de façon stable et spécifique par association entre bases complémentaires.

SSCP (Single Strand Conformation Polymorphism) : variation de conformation tridimensionnelle d'un brin d'ADN induite par une variation de séquence nucléotidique. Cette propriété est mise à profit pour détecter les mutations ponctuelles.

SSR (Simple Sequence Repeat) : voir Microsatellite.

STR (Short Tandem Repeat) : voir Microsatellites.

Taq polymérase : ADN polymérase thermorésistante extraite de la bactérie *Thermus aquaticus* et utilisée pour l'amplification in vitro de l'ADN (PCR) à température élevée (environ 70°C).

Température de fusion (T_m) : point d'inflexion de la courbe de fusion d'un segment d'ADN, correspondant virtuellement à une dénaturation de la moitié de la séquence.

Température d'hybridation (T_a) : point d'inflexion de la courbe d'hybridation d'une sonde ou d'une amorce sur un segment d'ADN, correspondant virtuellement à une hybridation de la moitié des sondes/amorces.

Traduction : synthèse d'une protéine par les ribosomes à partir d'un ARN messager mûré.

Transcription : synthèse d'ARN par une ARN polymérase à partir d'une matrice d'ADN.

Transcrit : ARN produit par la transcription d'un gène.

Transformation bactérienne : technique expérimentale consistant à faire pénétrer un fragment d'ADN exogène dans une bactérie.

Transversion : mutation ponctuelle entraînant la substitution d'une base purique par une base pyrimidique ou vice versa.

Vecteur : séquence nucléotidique capable de s'autorépliquer, utilisée pour la recombinaison in vitro de l'ADN et son amplification extra-chromosomique (clonage) (ex. : plasmide, bactériophage).

Zygote : cellule diploïde résultant de la fusion de deux gamètes parentaux (cellules haploïdes).