

MODULE : GENETIQUE GENERALE ET MOLECULAIRE / PARTIE II

CHAPITRE I : L'ADN : MATERIEL GENETIQUE DU VIVANT

L'ADN est une macromolécule qui est le constituant essentiel des chromosomes. Il représente l'information génétique qui englobe les gènes. Un gène est un segment d'ADN qui contient l'information nécessaire pour la synthèse d'un produit biologique fonctionnel, une protéine ou un ARN. Donc, la fonction unique de l'ADN est le stockage et la transmission de l'information génétique.

Découverte de l'ADN.

Une expérience réalisée par Avery McLeod et McCarty en 1944 a apporté la preuve directe que l'ADN est le porteur de l'information génétique de la cellule. Cette expérience, détaillée dans le chapitre précédent, est partagée en trois étapes :

- a). la première étape consiste à injecter une souris avec une souche virulente S d'une bactérie *Streptococcus pneumoniae* ce qui a causé sa mort.
- b). la deuxième étape consiste à injecter une autre souris avec une souche non virulente R de cette même bactérie. Cette souris a survécu.
- c). la troisième étape est de mélanger la souche non-virulente vivante avec une souche virulente tuée par la chaleur et d'injecter ce mélange à une souris. Le résultat est la mort de cette souris. La conclusion de cette expérience est qu'il y a un agent transformant qui a provoqué la conversion de la souche non-virulente en une souche virulente. Les chercheurs ont extrait l'ADN de la souche virulente tuée par la chaleur, ont éliminé le plus de protéines possibles et l'ont ajouté aux bactéries non-virulentes. Ces dernières ont été transformées de façon permanente en des bactéries virulentes. L'agent transformant est donc l'ADN. Aucune des autres fractions cellulaires, protéines, lipides ou glucides, n'a pu aboutir à cette transformation.

Une deuxième expérience importante a donné la preuve définitive que l'ADN porte l'information génétique. En 1952, Hershey et Chase ont utilisé du phosphore radioactif P^{32} et du soufre radioactif S^{35} pour démontrer que lorsque le bactériophage T_2 infecte son hôte *E. coli*, c'est l'ADN contenant le phosphore et non les protéines de la paroi cellulaire de T_2 contenant le soufre qui pénètre dans la cellule hôte et fournit l'information génétique pour la réplication du virus.

La composition chimique de l'ADN

L'ADN est formé de plusieurs unités qui sont les nucléotides. Chaque nucléotide est formé de 3 composés caractéristiques a) une base azotée, b) un pentose, le désoxyribose et c) un groupe phosphate. La molécule sans le phosphate est appelée un nucléoside. Les bases sont des dérivés de composés appelés des purines et des pyrimidines. Les purines sont l'adénine (A) et la guanine (G) et les pyrimidines sont la cytosine (C) et la thymine (T) (figure 1).

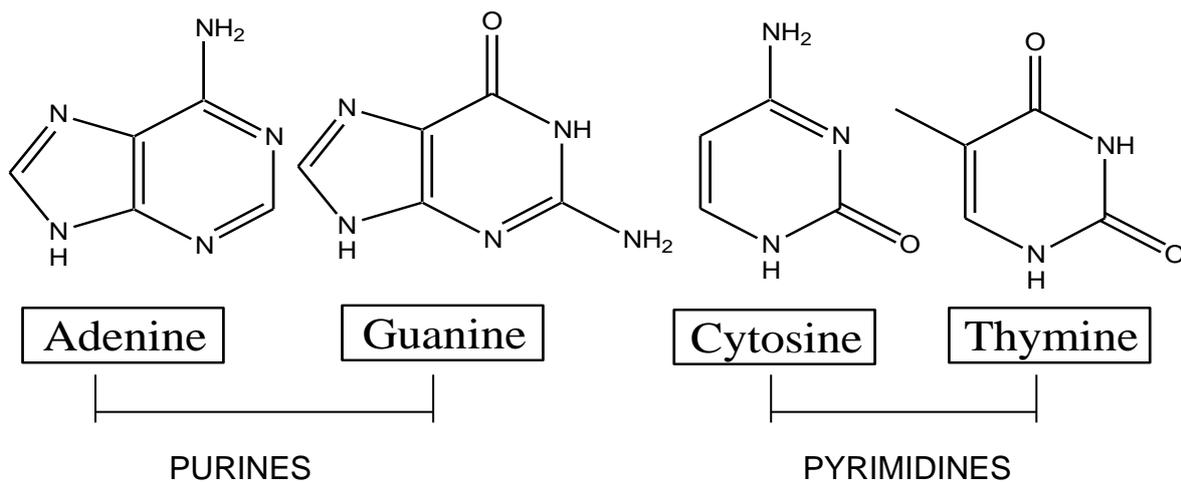


Figure 1 : les bases de l'ADN

La liaison entre une base et le sucre est une liaison covalente entre le N-1 de la pyrimidine ou le N-9 de la purine et le C1' du pentose. Ce type de liaison est appelée N-glycosidique. Le phosphate est estérifié au carbone 5' par une liaison phosphodiester (figure 2).

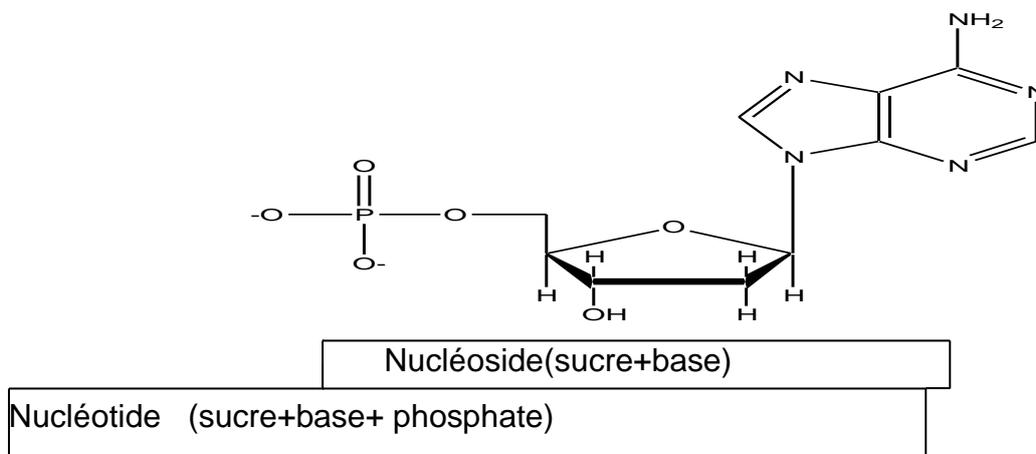


Figure 2 : l'unité de l'ADN

Ainsi, chaque chaîne d'ADN a une extrémité 5' et une autre extrémité 3' et donc une orientation de 5' vers 3' (figure 3). L'axe principal de la chaîne appelé aussi «le squelette carboné» est formé de résidus de phosphates et de pentoses alternés. Ils sont hydrophiles car les groupes hydrogènes des pentoses forment des liaisons hydrogènes avec l'eau et les phosphates sont négativement chargés à pH 7. Ces charges sont neutralisées par les interactions avec des charges positives portées par les protéines, les métaux et les polyamines. Les groupes attachés à cet axe principal à des intervalles réguliers sont les bases azotées. Un acide nucléique court (contenant 50 nucléotides ou moins) est appelé un oligonucléotide. Celui qui est plus long est appelé un polynucléotide.

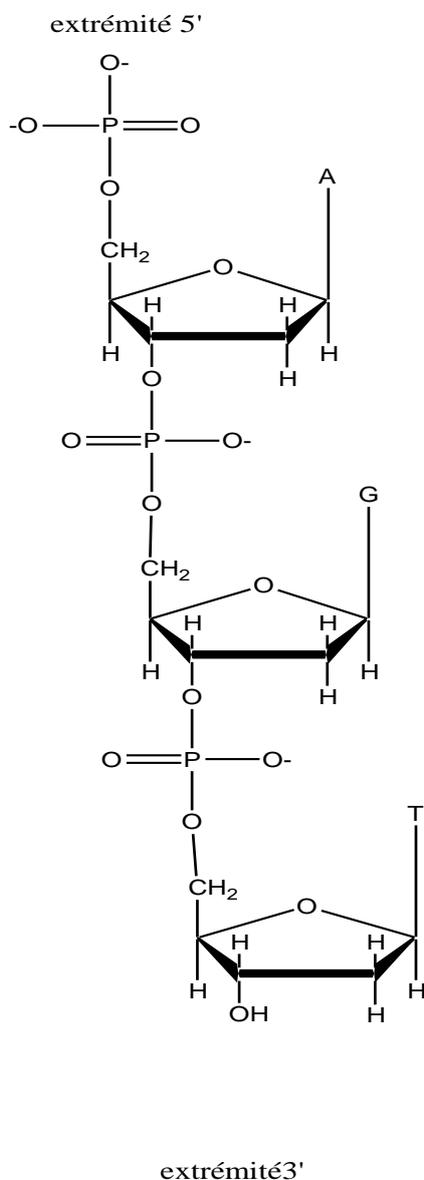


Figure 3 : Chaîne d'ADN

Propriétés des bases

Les purines et les pyrimidines sont des bases faibles. Leurs propriétés chimiques peuvent affecter la structure et éventuellement la fonction des acides nucléiques. Les purines et les pyrimidines de l'ADN sont des molécules hautement conjuguées, une propriété très importante pour la structure, la distribution électronique, et l'absorption de lumière des acides nucléiques. La résonance des atomes dans la structure cyclique des bases donne à la plupart des liaisons un caractère de double-liaison partielle. Ces dernières peuvent changer de place ce qui fait que les bases peuvent avoir des formes tautomériques. Certaines formes tautomériques prédominent à un pH de 7.0. Le résultat de la résonance est que toutes les bases absorbent les rayons ultra-violet (UV) à des longueurs d'onde de 260nm.

Les purines et les pyrimidines sont hydrophobes et relativement insolubles dans l'eau au pH neutre de la cellule. A des pH acides et basiques, les bases deviennent chargées et leur solubilité dans l'eau augmente. L'entassement des bases dans un acide nucléique implique la combinaison des forces de Van der Waals qui sont très importantes dans la stabilité de la structure tri-dimensionnelle des acides nucléiques.

Propriétés physiques de l'ADN

L'ADN peut subir une fusion réversible si on le chauffe à 98 °C ; la chaleur rompt les liaisons hydrogènes entre les bases, aboutissant ainsi à la séparation des deux brins. La fusion est subite à une certaine température appelée température de fusion (T_m) définie par la température à laquelle la moitié de la structure hélicale est perdue. Le fait que la fusion soit subite montre que la structure de l'ADN est hautement coopérative.

La fusion de l'ADN est aussi accompagnée par un hyperchromisme qui est une augmentation de l'absorbance en allant d'une structure double à une structure simple. Si l'ADN est riche en GC, T_m augmente. Si par contre il est riche en AT T_m diminue.

Les lois de Chargaff:

En 1950, Erwin Chargaff a étudié la quantité de bases dans une large sélection d'ADN provenant d'organismes différents. Ces études ont mené aux déductions suivantes appelées les lois de Chargaff :

- 1). La quantité totale des pyrimidines (T+C) est toujours égale à la quantité totale des purines (A+G).

2). La quantité de T est toujours égale à la quantité de A, et la quantité de C est toujours égale à la quantité de G. Cependant, la quantité de A+T n'est pas toujours égale à G+C. Ce rapport est différent selon les organismes.

La structure des acides nucléiques

La structure de l'ADN a été mise en évidence en 1953 par James Watson et Francis Crick. Cette découverte a été faite après une série de recherches sur l'ADN, exploitant ainsi les résultats obtenus auparavant concernant la composition des bases de Chargaff et les résultats de la diffraction aux rayons X de Wilkins. A partir de ces données et utilisant leur propres constatations, Watson et Crick élaborèrent le concept de la double hélice d'ADN :

- 1). Deux chaînes polynucléotidiques de directions opposées et qui tournent autour d'un axe commun pour former une double hélice droitère.
- 2). Les purines et les pyrimidines sont à l'intérieur de l'hélice, alors que les unités de phosphate et de sucre sont à l'extérieur, formant ainsi le squelette carboné de la molécule d'ADN.
- 3). Le diamètre de l'ADN est de 20Å . Les bases adjacentes sont séparées par 3.4Å au long de l'axe de l'hélice et reliées par une rotation de 36° . La structure hélicoïdale est répétée tous les 10 résidus ou à des intervalles de 34Å , représentant le pas de l'hélice.
- 4). L'adénine est appariée à la thymine par deux liaisons hydrogènes alors que la guanine est appariée à la cytosine par trois liaisons hydrogènes. La double hélice est aussi stabilisée par des interactions entre les bases entassées sur la même chaîne. Cet appariement de bases est fortement soutenu par les études préliminaires de Chargaff.
- 5). La séquence de bases d'une chaîne polynucléotidique n'est pas restreinte. Elle porte l'information génétique.

Les différentes formes tri-dimensionnelles de l'ADN.

Le modèle proposé par Watson et Crick est celui de l'ADN B. Il a été déduit sur la base de la diffraction aux rayons X de fibres d'ADN. D'autres informations structurales peuvent être obtenues d'analyses de rayons X sur des cristaux d'ADN. Ces dernières ont révélé que l'ADN peut avoir beaucoup plus de diversité et de variabilité qu'envisagé avant.

L'analyse par Dickerson d'un dodécamère d'ADN montre que celui-ci ressemble beaucoup à l'ADN B de Watson et Crick à l'exception du fait qu'il présente des variations

locales. Celles-ci sont dues à des angles de rotation entre les bases qui varient de 28 à 42 degré au lieu d'un angle de rotation constant de 36° dans le modèle de Watson et Crick. L'autre variation est due au phénomène de «propeller twist» qui est une rotation opposée des deux bases sur le long de leur axe, ce qui augmente l'entassement des bases de chaque chaîne. Une autre source de variabilité est le «tilting» qui représente l'enroulement des paires de base relativement à leurs voisines. Ces variations locales de la double hélice dépendent de la séquence de bases. Une autre caractéristique de l'ADN B est qu'il peut être plié lentement en un arc ou superenroulé sans changer la structure locale de l'ADN. Cette déformation facile est très importante biologiquement parce qu'elle permet la formation de l'ADN circulaire et permet à l'ADN de s'enrouler autour des protéines. Cela permet aussi la condensation de l'ADN pour qu'il rentre dans un volume restreint telle une cellule. L'ADN peut être aussi plié à certains sites tels une séquence de quatre adénines ou l'attachement d'une protéine.

Une autre caractéristique est la présence de sillons majeur et mineur. Ils sont formés car les liaisons glycosidiques d'une paire de bases ne sont pas diamétralement opposées. La largeur du sillon majeur fait qu'il est plus accessible aux interactions avec les protéines qui reconnaissent des séquences spécifiques d'ADN.

La forme de l'ADN-A (fig 4 b).

L'ADN-A est formé quand l'humidité relative tombe jusqu'à 75%. Il est plus large et plus court que l'ADN-B et ses paires de bases sont courbées par rapport à l'axe de l'hélice. Le sillon mineur presque disparaît. Les groupes phosphates dans l'ADN-A s'attachent à moins de molécules d'eau que les phosphates dans l'ADN-B.

La forme de l'ADN-Z (fig 4c).

L'ADN-Z a été découvert quand des cristaux d'une chaîne courte d'ADN (un hexanucléotide) avec des cytosines et des guanines alternées (CGCGCG...) ont été analysés par diffraction aux rayons X par Rich et ses collaborateurs. Ils ont trouvé que cette double hélice est gauchère contrairement aux doubles hélices d'ADN A et B qui sont droitières. Une autre différence est que les phosphates sont en forme de zigzag du fait que l'unité répétée est un dinucléotide au lieu d'un mononucléotide, d'où l'appellation ADN-Z. La troisième différence réside dans le fait que l'ADN-Z contient un sillon profond. Il a été constaté que la structure Z peut exister à l'intérieur de l'ADN-B. La conversion de la forme B en Z peut se faire en tournant les paires de base de 180° et en tournant les

sucres des résidus de purine. Cependant, on notera que la formation de l'ADN Z est thermo dynamiquement défavorable à cause de la répulsion entre les groupes phosphates qui sont plus rapprochés dans la forme Z que dans la forme B. La transition vers la forme Z est cependant favorisée par la méthylation de la cytosine en C-5, une modification très connue chez les eucaryotes.

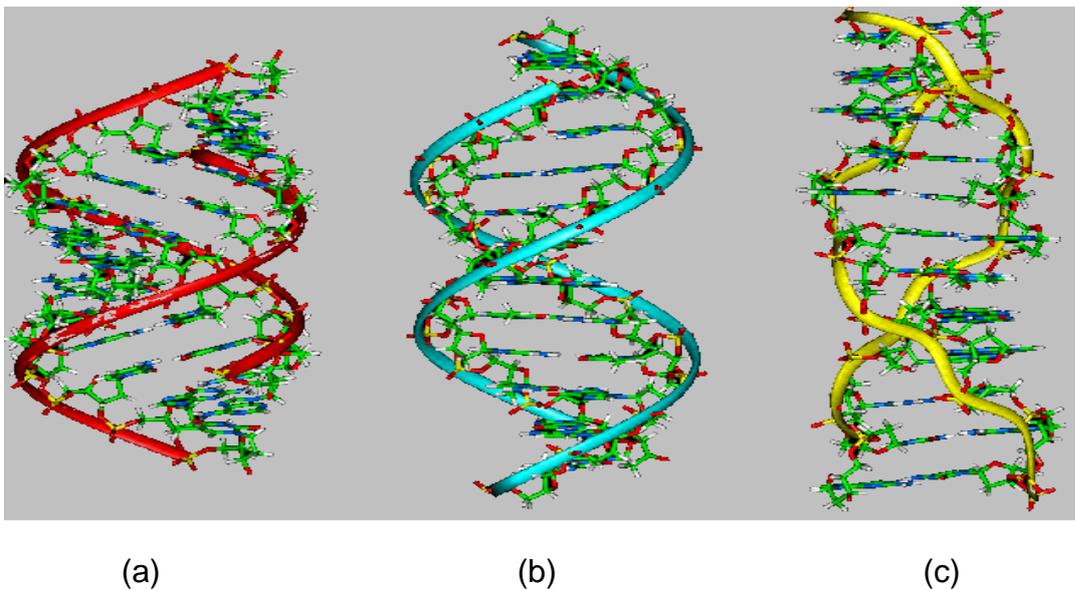


Fig 4 : les trois formes de l'ADN : la forme A (a), la forme B (b), et la forme Z (c)

La chromatine: organisation fonctionnelle du génome

I- Introduction

Dans les cellules eucaryotes, le matériel génétique est organisé en une structure complexe constituée d'ADN et de protéines et il est localisé dans un compartiment spécialisé, le noyau. Cette structure a été baptisée **chromatine**. Environ deux mètres d'ADN dans chaque cellule doivent être contenus dans un noyau de quelques μm de diamètre. En plus de cet énorme degré de compaction, l'ADN doit être rapidement accessible afin de permettre son interaction avec les machineries protéiques régulant les fonctions de la chromatine: la réplication, la réparation et la recombinaison. Ainsi l'organisation dynamique de la structure chromatinienne influence, potentiellement, toutes les fonctions du génome. L'unité fondamentale de la chromatine est appelée le **nucléosome** qui est composé d'ADN et d'histones. Il constitue le premier niveau de compaction de l'ADN dans le noyau. Cette structure est ensuite régulièrement répétée pour former le nucléofilament qui peut, lui-même, adopter des niveaux d'organisation plus compacts (Figs 1 et 3), le niveau de

condensation le plus élevé étant atteint au sein du chromosome métaphasique. Au sein du noyau interphasique, la chromatine est organisée en territoires fonctionnels.

La chromatine a été divisée en: **euchromatine** et **hétérochromatine**.

L'hétérochromatine a été définie comme une structure qui ne change pas d'état de condensation au cours du cycle cellulaire tandis que l'euchromatine apparaît décondensée pendant l'interphase. L'euchromatine se colorie de façon légère alors que l'hétérochromatine se colore de façon plus dense parce qu'elle est plus condensée que l'euchromatine. La plupart du génome est formé d'euchromatine dans laquelle la plupart des gènes sont actifs. L'hétérochromatine est localisée principalement en périphérie du noyau et du nucléole tandis que l'euchromatine est répartie à l'intérieur du nucléoplasme.

On distingue:

L'hétérochromatine constitutive qui contient peu de gènes, formée principalement de séquences répétées et dont les plus grandes régions sont situées à proximité des centromères et des télomères, de

L'hétérochromatine facultative qui contient des régions codantes pouvant adopter les caractéristiques structurale et fonctionnelle de l'hétérochromatine, comme le chromosome X inactif chez la femelle des mammifères.

Dans cette revue, nous rappèlerons les composants de la chromatine et nous présenterons les différents niveaux d'organisation de la chromatine, en partant du niveau du nucléosome jusqu'à celui de l'organisation en domaines dans le noyau. Nous discuterons ensuite comment les variations dans la composition fondamentale de la chromatine peuvent influencer son activité et comment les facteurs capables de faciliter la formation de ces structures jouent un rôle critique pour transmettre des caractères de diversité dans la structure chromatinienne dynamique. Finalement, nous exposerons comment la chromatine influence l'organisation du génome à l'échelle du noyau.

II- Le nucléosome

La digestion partielle de l'ADN organisé en chromatine génère des fragments de 180 à 200 paires de bases. La régularité de cette structure a été ensuite confirmée par analyse en microscopie électronique révélant une chromatine constituée de particules régulièrement espacées, dont l'aspect rappelle celui d'un "collier de perles". La stochiométrie ADN-histones est de 1/1 en masse.

Le nucléosome est l'unité fondamentale de la chromatine. Il est composé d'une particule cœur et d'une région de liaison (ou région internucléosomale) qui relie les particules cœurs adjacentes (Fig 1).

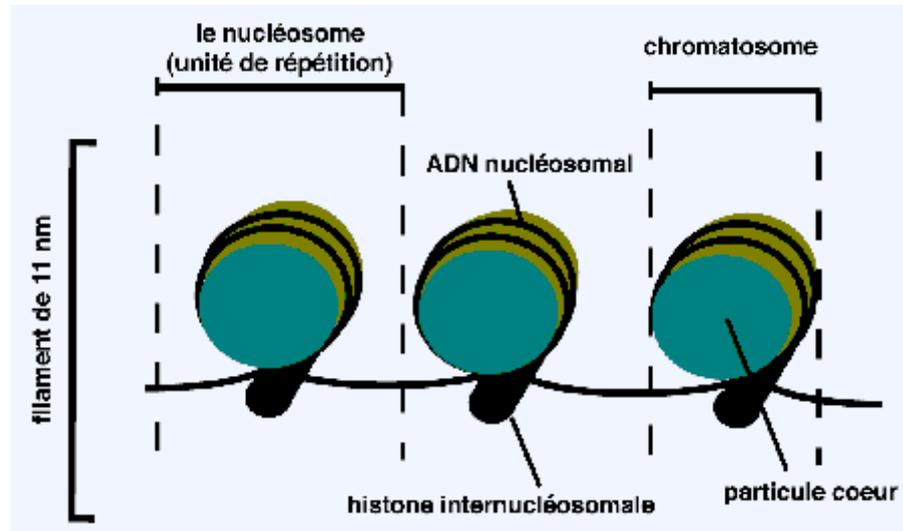


Figure 1. Eléments de définition du nucléosome et du chromatosome.

La particule cœur, dont la structure est très conservée parmi les espèces, est composée de 146 paires de bases d'ADN enroulées selon environ 1,7 tour autour d'un octamère protéique comprenant deux exemplaires de chacune des histones H3, H4, H2A et H2B. La longueur de la région d'ADN internucléosomale varie selon l'espèce et le type cellulaire. C'est au niveau de cette région que les histones internucléosomales également variables, sont incorporées. Ainsi, la longueur d'ADN caractéristique d'un nucléosome peut varier selon les espèces entre 160 et 241 paires de bases. Des analyses ont révélé d'une part, l'enroulement de l'ADN autour de l'octamère d'histones et d'autre part, les interactions entre histones/ADN et histones/histones par leur "motif histone fold" selon une configuration en poignée de mains.

III. Les histones

III.1. Les histones de la particule cœur

Les histones de la particule cœur, H3, H4, H2A et H2B sont de petites protéines basiques très conservées au cours de l'évolution (Fig 2). La région la plus conservée de ces histones est leur domaine central structuré composé du "motif histone fold" qui comprend trois hélices séparées par deux boucles. En revanche, les extrémités N-terminales de ces histones sont plus variables et sont dépourvues de structure secondaire. Ces extrémités

sont particulièrement riches en résidus lysine et arginine et donc très basiques. Elles sont la cible de nombreuses modifications post-traductionnelles pouvant affecter leurs charges mais aussi l'accessibilité à l'ADN et les interactions protéines/protéines avec le nucléosome. Il est important de noter que d'autres protéines impliquées dans les interactions avec l'ADN présentent ce type de "motif histone fold".

III. 2. Les histones internucléosomales

Les histones internucléosomales sont formées de H1, les protéines qui s'associent à la région d'ADN de liaison entre deux nucléosomes. Contrairement aux histones de la particule cœur, elles sont peu conservées parmi les espèces. Chez les eucaryotes supérieurs, elles sont composées de trois domaines: un domaine globulaire central non polaire très conservé pendant l'évolution, essentiel pour les interactions avec l'ADN, et deux extrémités N- et C- terminales non structurées, hautement basiques, et soumises à des modifications post-traductionnelles. Dans certains tissus, comme les érythrocytes aviaires, H1 est remplacée par H5 une autre protéine histonique. Les histones internucléosomales joueraient un rôle dans l'espacement des unités nucléosomales et dans la compaction de l'ADN au sein du nucléosome en créant une région d'interaction entre les nucléosomes adjacents.

IV. Les différentes étapes dans l'assemblage de la chromatine

L'assemblage de l'ADN en chromatine comprend plusieurs étapes qui commencent par la formation de son unité fondamentale, le nucléosome, et finissent par des niveaux d'organisation supérieurs en domaines spécifiques dans le noyau. Les différentes étapes de cet assemblage sont décrites schématiquement dans la Fig 3.

- La première étape comprend la mise en place, sur l'ADN, d'un tétramère d'histones (H3-H4)₂ nouvellement synthétisées formant la particule sub-nucléosomale, à laquelle viennent s'adjoindre deux dimères H2A-H2B. L'ensemble constitue la particule nucléosomale cœur composée de 146 paires de base d'ADN enroulées autour de l'octamère d'histones. Cette particule cœur et l'ADN de liaison forment le nucléosome. Les histones nouvellement synthétisées sont spécifiquement modifiées (ex: acétylation de l'histone H4).
- L'étape suivante est une étape de maturation nécessitant la présence d'ATP, au cours de laquelle les nucléosomes sont régulièrement espacés et forment le nucléofilament. Pendant cette étape, les histones nouvellement incorporées sont désacétylées.

- Ensuite l'incorporation des histones internucléosomales est accompagnée par le repliement du nucléofilament en fibre de 30 nm dont la structure n'est pas élucidée à ce jour. Deux modèles principaux existent : le modèle de type solénoïde et le modèle de type zig zag.

- Finalement, un grand nombre de repliements successifs conduisent à des niveaux d'organisation supérieurs en domaines spécifiques dans le noyau.

A chacune des étapes décrites ci-dessus, des variations dans la composition et l'activité de la chromatine peuvent être obtenues en modifiant soit ses composants élémentaires, soit l'activité de facteurs impliqués dans les processus d'assemblage et de désassemblage.

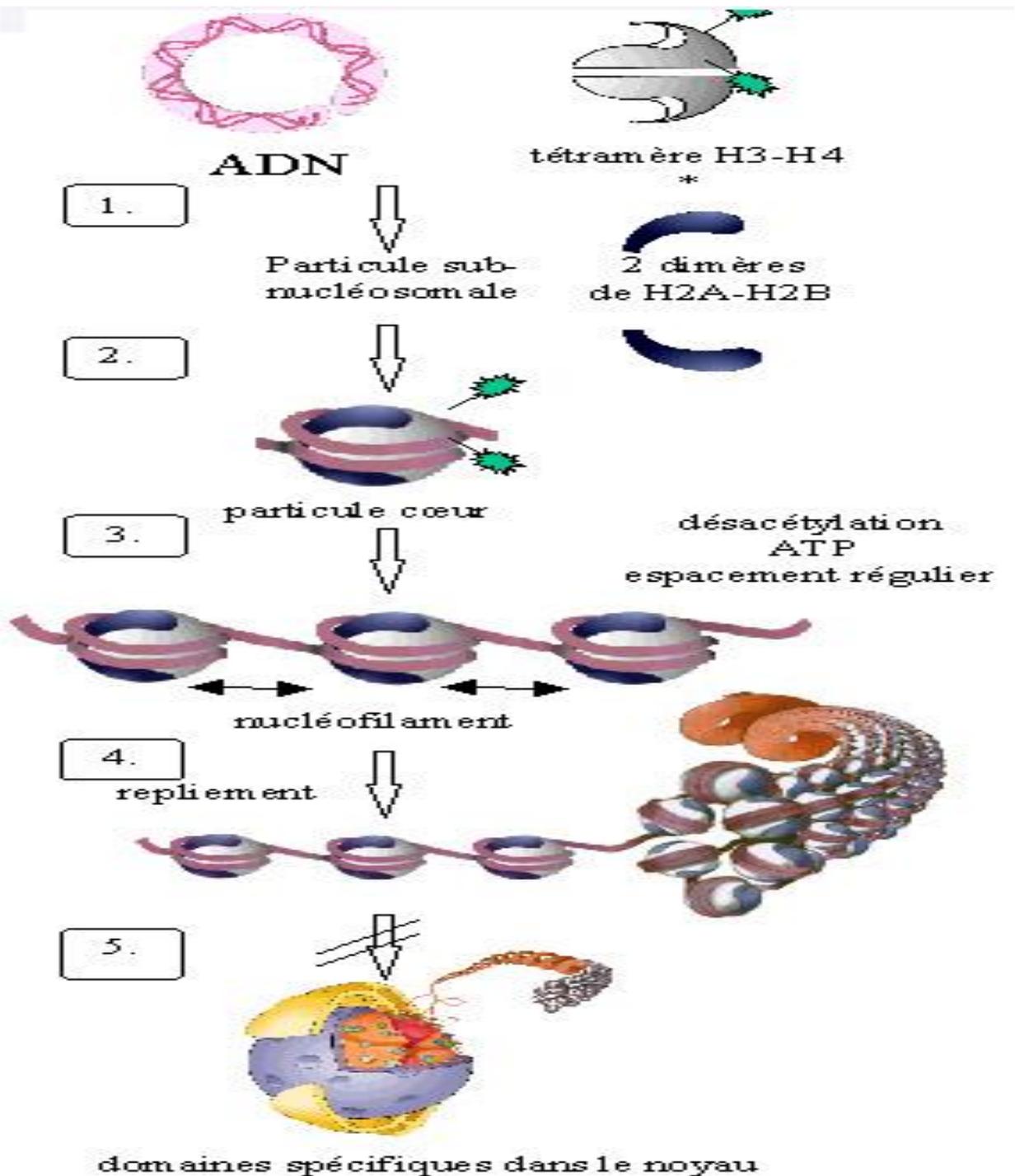


Figure 3. Les principales étapes de l'assemblage de la chromatine.

L'assemblage commence avec la mise en place d'un tétramère (H3-H4)₂ d'histones nouvellement synthétisées (1), à laquelle viennent s'adjoindre deux dimères H2A-H2B (2) pour former la particule nucléosomale cœur. Les histones nouvellement synthétisées sont spécifiquement modifiées; la modification la plus conservée est l'acétylation de l'histone H4 sur les lysines 5 et 12 (H3-H4*). L'étape de maturation nécessite la présence d'ATP afin

d'établir un espacement régulier des nucléosomes et les histones nouvellement incorporées sont désacétylées (3). L'incorporation des histones internucléosomales est accompagnée par le repliement du nucléofilament. Ici est présenté le modèle de type solénoïde dans lequel il y a 6 nucléosomes par tour (4). Finalement, plusieurs repliements successifs conduisent à des niveaux d'organisation supérieurs en domaines spécifiques dans le noyau (5).

VII. Organisation du génome dans le noyau

Au-delà des nucléosomes, la chromatine présente des niveaux d'organisation supérieurs encore peu caractérisés. Le nucléofilament se compacte pour former la fibre de 30 nm qui s'organise en boucles de 150 à 200 kpb (250 nm pendant l'interphase) pour atteindre un niveau de compaction maximum dans le chromosome métaphasique (850 nm).

En interphase, l'organisation du génome est reliée à la structure des chromosomes dont les différentes régions organisées en bandes peuvent être visualisées par l'utilisation de colorant comme le Giemsa.

Les principales bandes sont: les **bandes G** et **C** qui sont répliquées tardivement en phase S correspondent à l'hétérochromatine, les bandes G sont riches en adénine et thymine qui sont stables et les **bandes R** qui sont répliquées précocement en phase S correspondent à l'euchromatine. Les bandes R sont enrichies en histones acétylées qui sont conservées entre l'interphase et la mitose, suggérant un rôle de marqueur fonctionnel de cette modification pouvant servir de mémoire de l'organisation de ces domaines au cours du cycle cellulaire.

Une autre technique de coloration est la Q qui utilise le quinacrine qui s'attache préférentiellement aux régions riches en A et T ; les bandes Q sont visualisés avec le microscope à fluorescence. Les bandes Q et G sont équivalentes sur le chromosome.

La technique la plus moderne est celle de FISH. Qui utilise une sonde d'un gène fluorescente qui s'hybride à un gène pour le localiser sur le génome.

Un karyotype est le groupe de tous les chromosomes métaphasiques. Le karyotype est spécifique à l'espèce. La paire de chromosomes la plus grande est la première. Seul le chromosome 21 est plus petit que le 22 ; le karyotype peut détecter certaines mutations.

La localisation des chromosomes dans le noyau interphasique révèle que chaque chromosome occupe un espace bien défini. Chez les mammifères, l'organisation des chromosomes dans le noyau varie en fonction du type cellulaire. Pendant l'interphase, les bandes des chromosomes métaphasiques sont localisées dans le noyau en fonction de leur

réplication: les bandes G répliquées tardivement occupent la périphérie du noyau tandis que les bandes R contenant la majorité des gènes résident dans la partie interne du nucléoplasme.

Ainsi, bien que chaque chromosome occupe un territoire différent, des parties appartenant à des chromosomes distincts peuvent se réunir pour former des domaines fonctionnels. La localisation de régions codantes ou non codantes suggère que les gènes tendent à être localisés à la surface des territoires chromatinien.

D'après un modèle reposant sur la localisation de quelques gènes, les transcrits seraient alors libérés dans des canaux interchromosomiques, transférés au niveau des sites d'épissage, puis exportés dans le cytoplasme après maturation.

Toutes ces études ont permis de proposer que le noyau est organisé en domaines. La localisation de l'ADN dans ces domaines peut être, en partie, une conséquence des activités de la chromatine. Le ciblage de protéines pourrait aider à apporter des protéines spécifiques dans des domaines particuliers du noyau. Dans un modèle hypothétique, les protéines associées à l'hétérochromatine (par exemple HP1, Polycomb, Sir3/pSir4p, ATRX), les facteurs de transcription (comme Ikaros) et les facteurs d'assemblage de la chromatine (comme CAF-1) sont des candidats pour l'établissement et le maintien des domaines nucléaires.

La valeur C est le nombre haploïde de chromosomes dans le génome. Il est caractéristique de chaque espèce. La valeur C d'un noyau eucaryotique est 1000 fois plus élevée que celui des bactéries et 100000 fois plus élevée que celui d'un virus.

Les classes cinétiques de l'ADN eucaryotique

Les séquences à copies uniques de gènes

C'est le cas de la plupart des gènes. Quand cet ADN est fragmenté, dissocié et incubé dans des conditions de réassociation la cinétique de réassociation est très lente. Chez la plupart des eucaryotes l'ADN à copie unique représente à peu près 70% de l'ADN chromosomal.

Les séquences répétées modérément Le génome humain contient 25% à 30% de ces séquences. Il est formé de familles de gènes répétées de façon dispersée dans le génome. On prend comme exemple les histones ou transposons qui peuvent être répétées un million de fois. L'organisation de ces séquences est faite de la manière suivante : une séquence de 300 p.b. répétée de façon modérée suivie par une séquence unique de 1200p.b. Certaines de ces séquences sont des séquences régulatrices. Certaines de 100 à 500 p.b. sont appelées SINES, d'autres de 5000p.b. sont appelées LINES. La famille Alu est la plus

étudiée parmi les SINES. Cette dernière de 300p.b. est répétée $9 \cdot 10^5$. Pour cela ces séquences sont considérées comme étant une classe hétérogène d'ADN eucaryotique.

Séquences hautement répétées.

Ils constituent 10% du génome Ils sont répétés en tandem comme les gènes de l'ARNr et ARNt . Ils constituent aussi les centromères et les télomères. Ils forment une bande distincte après centrifugation dans un gradient de CsCl ; cette bande est appelée satellite Ils peuvent ne pas être centromériques suivant l'espèce.

CHAPITRE 2 : La réplication de l'ADN

Watson et Crick ont proposé un modèle de réplication de l'ADN un mois après avoir proposé le modèle de la structure de l'ADN. Ce modèle est celui de la réplication semi-conservative. Une des chaînes filles de l'ADN est nouvellement synthétisée à partir d'une des chaînes parentales qui s'est séparée de l'autre chaîne. Néanmoins, deux autres modèles ont été proposés: le modèle conservateur et le modèle dispersif. L'expérience réalisée par Messelson et Stahl (1957) a permis de fournir la preuve en faveur du modèle semi-conservatif.

Ces chercheurs ont utilisé *E. coli* car sa culture dans un milieu minimal est facile et rapide et son temps de génération est court, ce qui permet de suivre l'évolution de l'ADN à travers les générations successives. *E. coli* est cultivé pendant plusieurs générations dans un milieu contenant comme seule source d'azote le NH_4Cl contenant du N^{15} . Le résultat est que tous les composés cellulaires de la bactérie contenant de l'azote entre autres les purines et les pyrimidines de l'ADN seront marquées par le N^{15} qui fait que son ADN est plus dense que l'ADN normal contenant le N^{14} . La bactérie est ensuite transférée rapidement dans un milieu contenant le N^{14} .

La distribution de N^{14} et N^{15} est révélée par la technique de centrifugation sur gradient de densité. Elle consiste à dissoudre une petite quantité d'ADN dans une solution concentrée de Chlorure de Césium qui a une densité proche de celle de l'ADN ($1,7\text{g}/\text{cm}^3$). Cette solution est centrifugée jusqu'à atteindre l'équilibre. Un gradient stable de densité est établi qui varie de $1,66$ à $1,76\text{g}/\text{cm}^3$. Les molécules d'ADN formeront une bande à l'endroit du gradient qui correspond à leur densité. Un mélange d'ADN N^{14} et d'ADN N^{15} a donné des bandes séparées parce que leur densité est différente de 1%. Si on extrait l'ADN une génération après l'avoir transférée dans un milieu contenant le $\text{N}^{14}\text{H}_4\text{Cl}$, une seule bande apparaît qui est intermédiaire entre la bande contenant seulement le N^{15} et celle contenant seulement le N^{14} . Ceci indique que cette bande est formée d'un hybride de N^{15} et de N^{14} , ce qui veut dire que la moitié de son ADN est parental et l'autre moitié est nouvellement synthétisée. Si on extrait l'ADN deux générations après le transfert des bactéries dans un milieu contenant le N^{14} , on trouve deux bandes : l'une est la bande intermédiaire qu'on a trouvée après la première génération et l'autre correspond à la bande qui contient entièrement du N^{14} (voir figure 5).

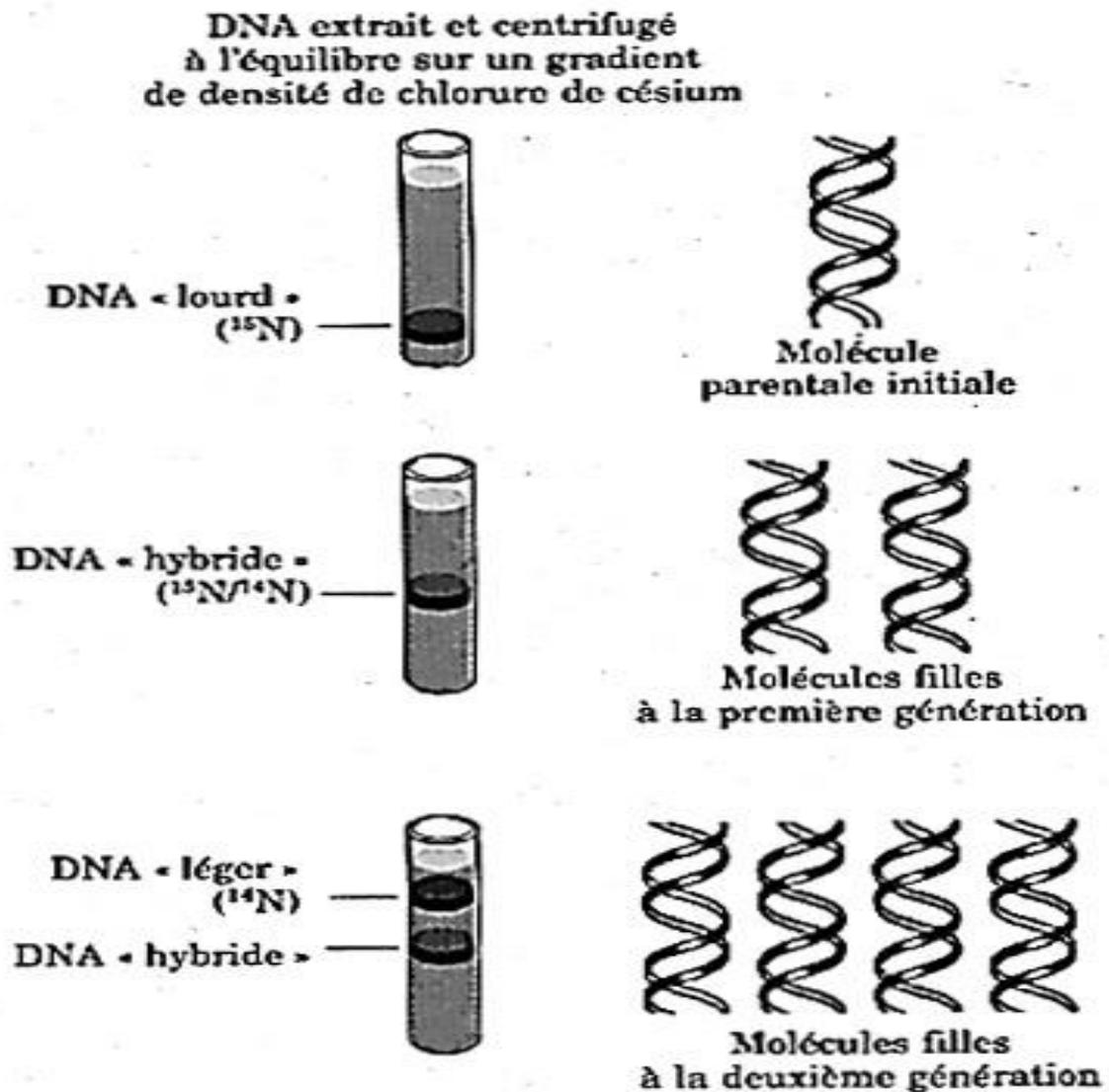
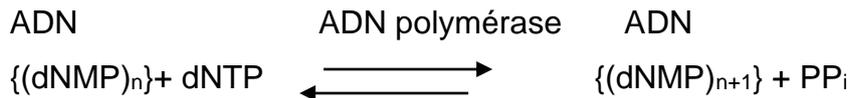


Fig 5. Expérience de Messelson et Stahl

Les enzymes de la réplication de l'ADN.

En 1955, Arthur Kornberg et ses collègues avaient entrepris des recherches afin de purifier les enzymes de la réplication pour pouvoir comprendre les réactions impliquées dans ce processus. En mettant un mélange de réaction contenant des fragments d'ADN, tous les 4 précurseurs nucléotidiques (dNTP) marqués radioactivement, et un lysat préparé de cellules de *E. coli*, ils ont accompli avec succès pour la première fois une synthèse d'ADN. En réalisant que le lysat de *E. coli* contient le composé crucial de la synthèse qui est l'enzyme, il l'ont analysé et ont pu isoler cette enzyme nécessaire à la synthèse. Cette enzyme a été originalement appelée l'enzyme de Kornberg mais elle est maintenant appelée l'ADN polymérase I parce qu'elle a été isolée en premier. Les chercheurs ont conclu que toutes les ADN polymérases ont besoin de 4 ingrédients :

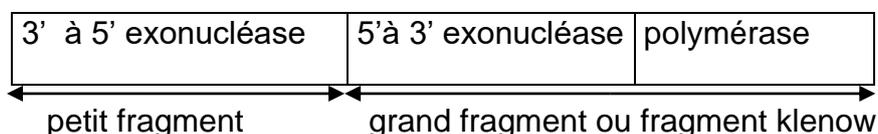
- 1). Les 4 dNTPs qui sont les précurseurs activés des nucléotides.
- 2). Les ions magnésium (Mg^{2+}).
- 3). Un fragment d'ADN ou d'ARN qui représente une amorce ayant une extrémité 3'OH libre.
- 4). Une matrice. Cette matrice peut être de l'ADN simple brin ou double brin. Ce dernier est une matrice efficace seulement quand l'une de ses chaînes est cassée à un site ou plus. La réaction de polymérisation est la suivante:



Cette réaction représente la catalyse par l'ADN polymérase d'une réaction d'addition d'un nucléotide à une chaîne d'ADN existante avec la libération d'un pyrophosphate inorganique (PPi). C'est une attaque nucléophyllique de l'extrémité 3'OH de l'amorce sur l'atome de phosphore le plus interne du désoxyribonucléoside triphosphate. La liaison formée est une liaison phosphodiester. L'hydrolyse subséquente du pyrophosphate va donner l'énergie suffisante pour la polymérisation. L'élongation de l'ADN se fait toujours dans la direction 5' → 3' puisque les ADN polymérases connues jusqu'à présent ne peuvent allonger la chaîne d'ADN que dans cette direction.

L'ADN polymérase I

L'ADN polymérase I est codée par le gène *polA* et consiste en un seul polypeptide. Une mutation de ce gène a été découverte en 1969 quand De Lucia et Cairns ont isolé le mutant *polA1*. Ce mutant qui est déficient pour la polymérase I avait la capacité de se reproduire et de croître normalement ; cependant, quand il est exposé aux rayons UV ou à des mutagènes chimiques, son taux de mutation est très élevé. Ceci montre que la fonction primaire de l'ADN polymérase I n'est pas la synthèse de l'ADN mais une réparation spécifique de l'ADN appelée «nick translation». En fait, cette fonction est assurée par l'activité exonucléasique 5' → 3' absente chez la plupart des autres polymérases. Cette activité est localisée dans un domaine structural (appelé le fragment Klenow) qui peut être séparé du reste de l'enzyme par un traitement protéolytique léger (fig).



L'activité 5' → 3' est aussi responsable de l'hydrolyse de l'amorce d'ARN pendant la réplication. L'ADN polymérase I a une faible vélocité qui est estimée à 16-20 nucléotides par seconde.

Des études menées sur plusieurs mutants ont pu démontrer l'existence d'autres enzymes polymérisantes dans la cellule. Il s'agit notamment des ADN polymérases II et III. Des recherches encore plus récentes ont pu mettre en évidence l'existence d'autres polymérases: les ADN polymérases III et IV.

L'ADN polymérase II.

Elle a été découverte en 1970. Elle est codée par le gène *polB* et consiste en quatre sous-unités.

L'ADN polymérase III.

Elle a été découverte en 1971 et s'est avérée être l'enzyme principale de la réplication de l'ADN dans la cellule et donc la plus étudiée. Elle est formée de 10 sous-unités codées par plusieurs gènes. Trois de ces sous-unités constituent le centre catalytique de l'enzyme: 1) la sous-unité alpha (α) codée par le gène *dnaE* ou *polC* responsable de l'activité polymérasique 2) la sous-unité epsilon (ϵ) codée par le gène *dnaQ* possède une activité exonucléasique 3' → 5' qui a le rôle de lecture d'épreuves pour assurer la fidélité de la réplication et 3) la sous-unité theta (θ) codée par le gène *holE* qui s'associe aux deux autres sous-unités citées ci-dessus pour former ce qu'on appelle une polymérase «core» qui peut synthétiser l'ADN mais avec une activité catalytique limitée. Deux polymérases «core» peuvent être liées en un complexe par un dimère de sous-unité tau (τ). L'activité catalytique de l'enzyme est augmentée par l'association de 4 sous-unités beta (β). Elles permettent l'accrochage du centre catalytique à la matrice en s'associant en paires et forment une structure de «donut» qui entoure l'ADN et agit comme une pince. Ceci permet une activité catalytique de 500000 et une vitesse de réplication de 250-1000 nucléotides par seconde. Chaque cellule de *E. coli* contient 200 molécules de cette enzyme alors que le nombre de molécules de la polymérase I est de 10-20 molécules seulement. La fonction de l'ADN polymérase II et le nombre de molécules de cette enzyme dans la cellule ne sont pas encore connus. On croit qu'elle est concernée par la réparation de l'ADN.

Toutes les trois enzymes contiennent une activité exonucléasique 3' → 5' responsable du mécanisme de «lecture d'épreuves» ou correction des erreurs. Si un polymère contient un mésappariement de C au lieu d'un T devant un A, l'addition d'une polymérase I et de dTTP excise le C misapparié avant la continuation de la polymérisation. Ceci montre que

les ADN polymérasés assurent une fidélité de réplication très élevée car elles examinent le résultat de chaque polymérisation avant de procéder à la suivante, ce qui permet la stabilité du matériel génétique.

Deux autres polymérasés, les polymérasés IV et V, ont été découvertes en 1999. Il semble qu'elles sont concernées par une réparation inhabituelle de l'ADN.

La réplication de l'ADN

-chez les procaryotes.

La réplication de l'ADN requiert la présence de plus de 20 protéines, ayant chacune une fonction spécifique. Tout le système est appelé l'ADN réplicase ou le replisome. La réplication de l'ADN est divisée en trois phases : l'initiation, l'élongation et la terminaison.

L'initiation

L'autoradiographie et la microscopie électronique ont révélé la présence de structures en forme de theta (θ) chez *E. coli*, suggérant ainsi que la molécule d'ADN maintient sa forme circulaire pendant la réplication. Il est clair aussi qu'il y a une absence de segments longs d'ADN simple brin. Donc, les brins parentaux ne se séparent pas complètement pour servir comme matrice. En fait, la synthèse d'ADN nouveau est couplée au déroulement de l'ADN parental. Le site de déroulement et de synthèse spontanés est appelé une fourche de réplication.

La réplication étant un processus rigoureusement contrôlé, les chercheurs ont supposé qu'il doit y avoir un site d'initiation spécifique. Ceci est en fait démontré par des études sur les quantités relatives de certains gènes dans des cultures de *E. coli* où la réplication est rapide. Les fréquences relatives de ces gènes ont été déterminées par l'hybridation avec des sondes complémentaires de ces gènes (figure).

Au moment où les chercheurs tentaient d'expliquer le déroulement de la réplication au niveau moléculaire, un certain nombre de points sont restés ambigus : il s'agit, entre autres, de :

- a). Origine de réplication : est-elle fixe ou mobile ?
- b). Direction de la réplication : est-elle uni- ou bidirectionnelle ?

Deux expériences menées séparément ont permis de lever tous les doutes concernant l'origine et la direction de la réplication.

Origine de réplication

Sur le chromosome circulaire de *E.coli*, on considère deux gènes a et b tels que a est proche de l'origine de réplication alors que b est loin de cette origine.

Premier cas : Dans le cas où l'origine de réplication est fixe, le gène a se répliquera avant le gène b. Dans une culture bactérienne en croissance rapide, il y aura deux fois plus de gène a que de gène b.

Deuxième cas : Par contre, dans le cas où l'origine de réplication est mobile, c'est-à-dire aléatoire, le nombre de gènes a sera égal au nombre de gènes b puisque la réplication démarrera tantôt à côté de a tantôt à côté de b.

Le résultat expérimental est en faveur d'une origine fixe.

Direction de la réplication

Afin de mettre en évidence la direction de la réplication, les chercheurs ont encore une fois fait appel à *E.coli*. Des bactéries sont cultivées dans un milieu radioactif contenant de la thymine (.) à faible densité. Dès qu'elles ont entamées la réplication, elles ont été transférées dans un autre milieu radioactif mais contenant de la thymine à haute densité (.)

a). Si la réplication se faisait d'une manière unidirectionnelle (qu'elle se dirige à droite ou à gauche), on constatera après autoradiographie la présence de radioactivité à faible densité suivie de la radioactivité à haute densité.

ori.....**ou**.....**ori**

b). Si la réplication se faisait d'une manière bidirectionnelle, on verra la radioactivité à faible densité au milieu flanquée de la radioactivité à haute densité de part et d'autre.

.....**ori**.....

Les preuves expérimentales ont été en faveur de la réplication bidirectionnelle.

Ces études ont montré que les fréquences relatives de ces gènes dépendent de leur position dans le génome et que :

1). La réplication est initiée à un site unique identifié comme oriC localisé à 83mn sur la carte génétique standard de *E. coli* qui est de 100mn.

2). La réplication procède simultanément dans les deux directions opposées, à la même vitesse. En d'autres termes, on trouve deux fourches de réplication, une se déplace dans

la direction des aiguilles d'une montre, l'autre se déplace dans la direction opposée des aiguilles d'une montre.

3). Les deux fourches de réplication se rencontrent à «ter» (31' sur la carte) ou la région de terminaison, qui est opposée à la région de l'origine de réplication.

L'origine de réplication oriC est constituée de 245pb contenant des séquences hautement conservées parmi les origines de réplication bactériennes. Les séquences clés sont formées de deux séries de répétitions courtes : 3 répétitions de 13pb et 4 répétitions de 9pb. Au moins 9 enzymes ou protéines participent dans l'initiation de la réplication. Ils forment un complexe et séparent les deux brins de l'hélice à l'origine. Le composé clé dans le processus d'initiation est la protéine DnaA. Un complexe unique de 20 DnaA s'associe aux 4 répétitions de 9pb, puis reconnaissent et dénaturent avec succès l'ADN dans la région des trois répétitions de 13 pb, qui sont riches en AT (fig).

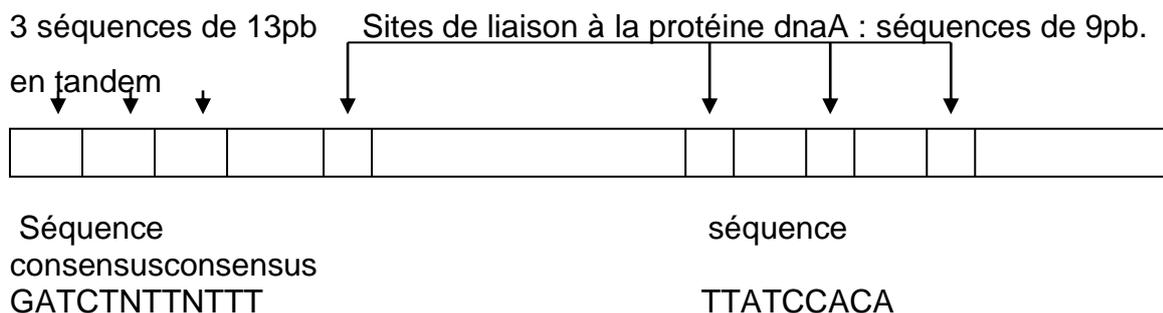


Figure . L'arrangement des séquences de l'origine de réplication oriC de *E. coli*.

Ce processus requiert la présence de l'ATP et HU qui ressemble à une histone. La protéine DnaB s'attache ensuite à la région de séparation par une réaction impliquant la protéine DnaC. Deux hexamères de DnaB s'attachant chacun à un brin de l'ADN agissent comme des hélicases, déroulant l'ADN bidirectionnellement et créant deux fourches de réplication potentielles (fig).

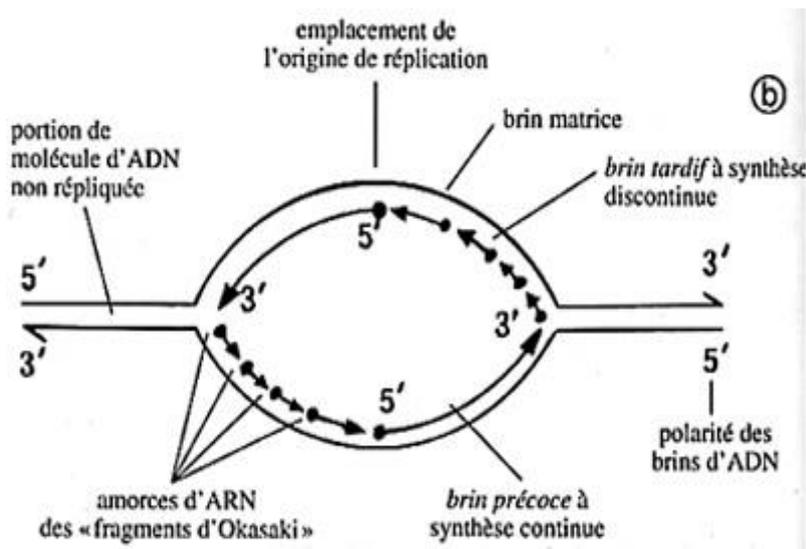
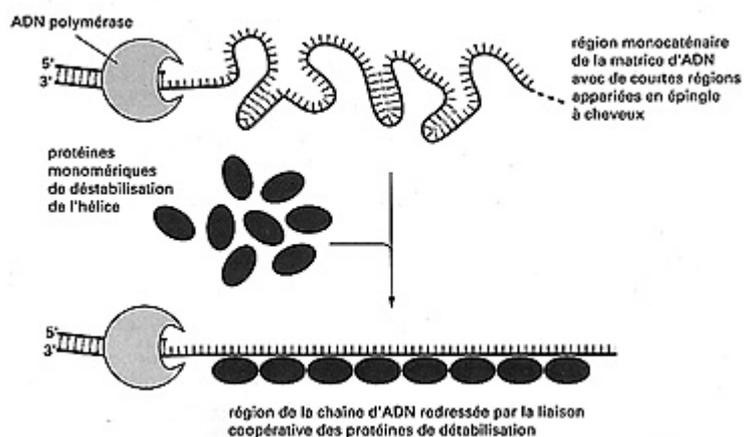


Fig 6 : Modèle d'initiation de la réplication à l'origine oriC de E.coli.

La protéine SSB (single-strand binding protein) ou protéine de liaison à l'ADN monobrin, s'attache ensuite coopérativement à chacun des deux monobrins et les empêche de se réannuler, alors que la gyrase (ADN topoisomérase II) libère le stress topologique créé par la DnaB hélicase. La gyrase agit en convertissant l'énergie de l'ATP en un superenroulement. Elle a une forme de losange sur laquelle 200pb d'ADN peuvent s'enrouler. Le déroulement de l'ADN engendré par la DnaB est couplé à la réplication quand d'autres protéines de réplication sont ajoutées.



L'initiation est la seule phase de la réplication de l'ADN qui est régulée. Cette régulation est telle que la réplication se fait une seule fois dans le cycle cellulaire. Le mécanisme de cette régulation n'est pas encore totalement connu.

Elongation

Elle inclut deux opérations distinctes : la synthèse du brin continu et du brin discontinu.

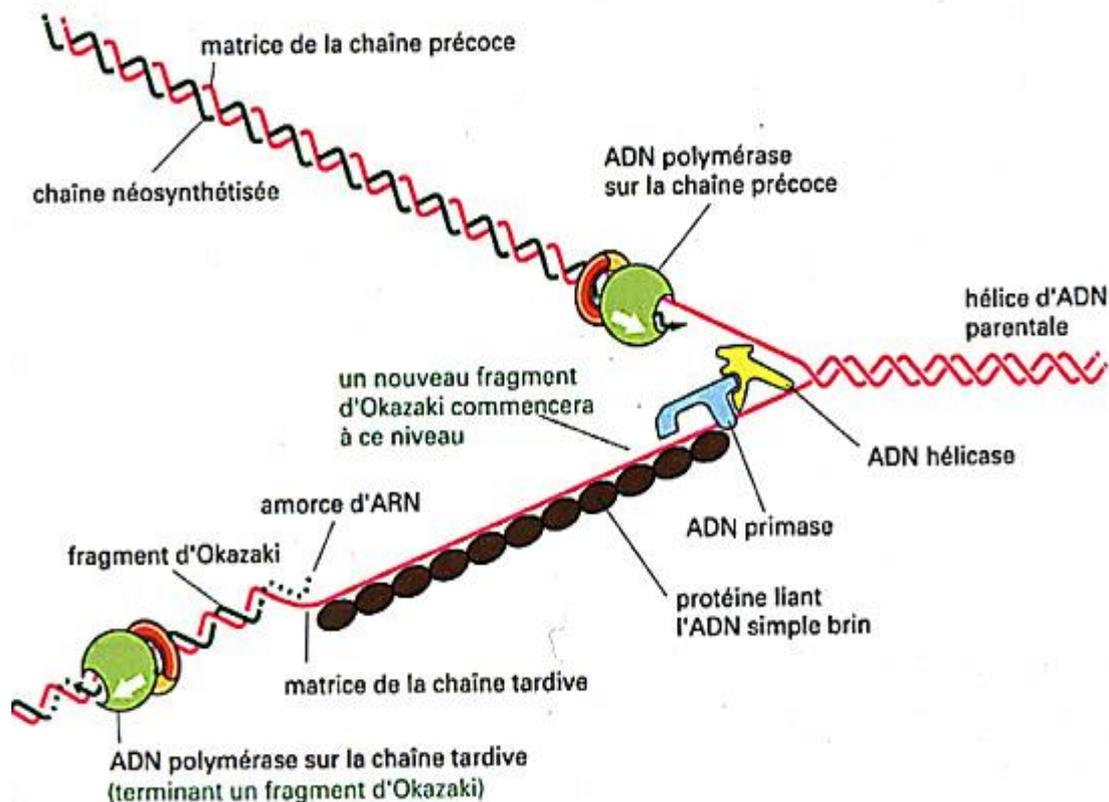
La première est plus directe. En raison de l'incapacité des ADN polymérases à assurer une synthèse *de novo* d'ADN, la réplication doit commencer par la synthèse d'une courte amorce d'ARN dotée d'une extrémité 3'OH (10 à 60 nucléotides) à l'origine de la réplication par la primase (DnaG). La synthèse de l'ADN par la polymérase III se fait par l'addition par cette enzyme de désoxyribonucléotides à l'extrémité 3'OH de l'amorce de façon continue.

La synthèse de l'autre chaîne a posé un problème aux chercheurs puisque la polymérase III ne peut pas synthétiser l'ADN dans la direction 3' → 5'. Ce problème a été résolu par Okazaki (1979) qui a découvert que la synthèse de cette chaîne se fait de façon discontinue. En effet, quand Okazaki a marqué le milieu de culture par la thymine tritiée (H^3) pendant des périodes de temps très courtes (5 secondes), il a trouvé de la radioactivité dans des petits fragments de 1000 à 2000 nucléotides. Cependant, quand il a marqué le milieu durant des périodes de temps plus longues, il a trouvé de la radioactivité dans des fragments plus longs. Ces recherches ont prouvé que l'ADN est synthétisé en petits fragments au début de la réplication qui sont ensuite liés pour former des fragments plus longs. En fait, des petits fragments d'ARN sont synthétisés par la primase. Cette enzyme est aidée par un complexe de six protéines qui forment le primosome et qui se déplace tout le long de la chaîne parentale qui a une polarité de 5' 3'. Les fragments d'ARN ainsi synthétisés sont utilisés ensuite comme des amorces pour la synthèse de petits fragments d'ADN par la polymérase III. A ce niveau, la synthèse des deux brins est coordonnée. Elle est réalisée par un dimère de polymérase III unique de façon simultanée. Cela est accompli par la formation d'une boucle par le brin parental servant de matrice au brin discontinu (fig 7). Les amorces d'ARN sont ensuite hydrolysées par l'activité exonucléasique 5' à 3' de l'ADN polymérase I et le vide est comblé par la même enzyme. Les extrémités de ces fragments d'ADN sont ensuite liées par une enzyme découverte en 1967: l'ADN ligase. La réaction catalysée par cette enzyme est la suivante :

Brin d'ADN -3'OH+ -O-P-O-5'-Brin d'ADN Brin d'ADN -3'-O-P-O-5'-Brin d'ADN

L'ADN ligase catalyse la formation d'une liaison phosphodiester entre le groupement 3'OH à l'extrémité d'une chaîne d'ADN et le groupement 5'-phosphate de l'extrémité de l'autre chaîne en utilisant l'énergie. Chez *E. coli* et d'autres bactéries, le NAD^+ fournit

cette énergie alors que chez les cellules animales, l'ATP est la source d'énergie. Cependant, Il faut noter que l'ADN ligase ne peut pas lier deux molécules d'ADN simple brin. Seuls les duplexes d'ADN sont liés par cette enzyme.

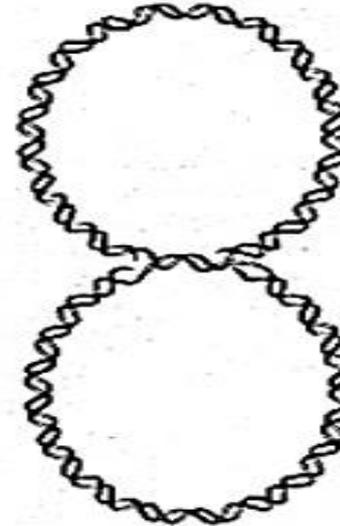
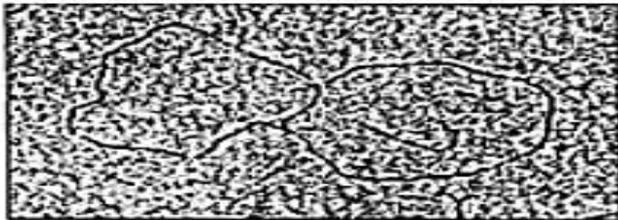
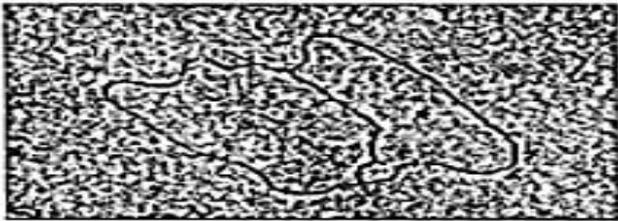
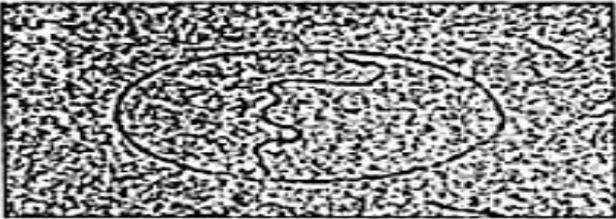
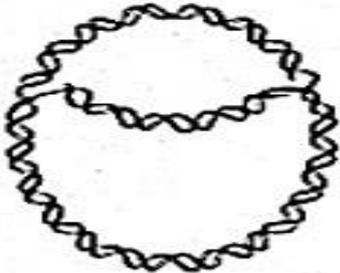
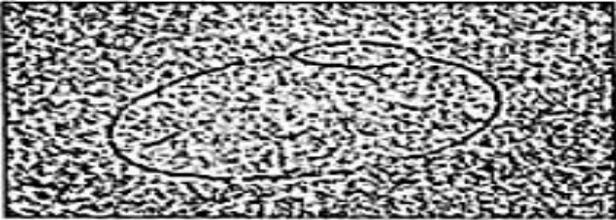


Terminaison de la réplication

Les deux fourches de réplication du chromosome circulaire de *E. coli* se rencontrent dans une région terminale contenant des copies multiples d'une séquence de 20 paires de bases appelée «Ter». Les séquences «Ter» sont arrangées sur le chromosome pour créer un piège qu'une fourche de réplication ne peut pas dépasser. Les séquences «Ter» sont des sites d'attachement d'une protéine appelée «Tus» (terminus utilization substance). Un seul complexe «Ter-Tus» fonctionne par cycle de réplication, celui qui est rencontré en premier par l'une des deux fourches de réplication.

Quand l'une des deux fourches rencontre le complexe «Ter-Tus», elle s'arrête. L'autre fourche s'arrête quand elle rencontre la première fourche déjà arrêtée. Les dernières centaines de paires de bases d'ADN entre ces deux complexes sont répliquées par un mécanisme inconnu complétant deux chromosomes circulaires liés topologiquement et

appelés des caténanes. Les deux cercles sont séparés par la topoisomérase IV, une topoisomérase de type II.



La réplication chez les eucaryotes

L'ADN des cellules eucaryotiques est beaucoup plus complexe que celui des procaryotes. Il est organisé en structures nucléoprotéiques complexes, ce qui rend sa réplication plus difficile et compliquée. Cependant, les caractéristiques essentielles de la réplication sont les mêmes chez les procaryotes et les eucaryotes.

Les origines de la réplication appelées réplicateurs ou ARS (séquences répliquatives autonomes) ont été identifiées et le mieux étudiées chez la levure. Ces réplicateurs peuvent s'étendre sur 150pb et contiennent plusieurs séquences conservées. Il y aurait 400 réplicateurs distribués parmi les 17 chromosomes du génome haploïde de levure. L'initiation de la réplication chez tous les eucaryotes requiert une protéine formée de plusieurs sous-unités appelée l'ORC (le complexe de reconnaissance de l'origine) qui s'associe à plusieurs séquences du réplicateur. ORC interagit et est régulé par d'autres protéines impliquées dans le contrôle du cycle cellulaire. Le taux de mouvement de la fourche de réplication chez les eucaryotes (50/s) est de 1/20 celui de *E. coli*. Les chromosomes eucaryotes sont beaucoup plus grands que les chromosomes procaryotes, ce qui requiert la présence d'origines multiples de réplication. En effet, les expériences d'autoradiographie ont montré que la réplication est initiée à plusieurs sites simultanément de façon bidirectionnelle. Chaque unité de réplication est appelée un réplicon et il existerait 30000 réplicons dans le génome haploïde humain.

Les chromosomes harlequins

Dans cette procédure, les chromosomes sont mis dans un milieu contenant de la Bromodéoxyuridine (BUdR). Les chromosomes sont ensuite colorés avec de la fluorescence et du Giemsa ; ceci distingue les chromatides hybrides qui contiennent un brin contenant du BUdR et un brin qui ne le contient pas et qui est coloré en noir qui donne une apparence de chromosome harlequin et démontre le mode de réplication semi-conservatif chez les eucaryotes.

A l'instar des bactéries, les organismes eucaryotes possèdent plusieurs types d'ADN polymérases; quelques-unes sont impliquées dans la réplication de l'ADN mitochondrial. La réplication du chromosome nucléaire requiert l'ADN polymérase alpha (α) associée à l'ADN polymérase delta (δ). L'ADN polymérase alpha est formée de sous-unités multiples avec des structures et des caractéristiques similaires dans toutes les cellules eucaryotes.

Une des sous-unités possède l'activité de la primase et la plus grande sous-unité contient l'activité de la polymérase. Néanmoins, cette polymérase ne possède pas l'activité de lecture d'épreuve de l'exonucléase 3' 5', ce qui veut dire qu'elle n'est pas l'enzyme principale de la réplication. On croit que sa fonction essentielle est de synthétiser des amorces pour les fragments d'Okazaki dans le brin discontinu. D'autre part, l'ADN polymérase delta possède l'activité exonucléasique 3' 5' et synthétise ainsi l'ADN des deux brins continu et discontinu en formant un complexe analogue à la polymérase III dimérique des bactéries. Cette enzyme est aussi associée et stimulée par une protéine PCNA (antigène nucléaire de prolifération cellulaire) qu'on trouve à des taux élevés dans les noyaux des cellules en prolifération. Cette protéine a une structure et une fonction similaires à la sous unité bêta (β) de la polymérase III de *E. coli*, formant une pince circulaire autour de l'ADN et augmentant ainsi son activité catalytique. Une autre polymérase, la polymérase epsilon (ϵ) remplace l'ADN polymérase delta dans quelques situations, comme lors de la réparation de l'ADN. Cette polymérase peut aussi fonctionner au niveau de la fourche de réplication, jouant ainsi un rôle analogue à celui de la polymérase I en assurant l'hydrolyse des fragments d'Okazaki sur le brin discontinu.

La polymérase bêta (β) participe dans la réparation de l'ADN. Elle est constituée d'une chaîne polypeptidique unique alors que la polymérase gamma (γ) est responsable de la réplication de l'ADN mitochondrial. Ces deux dernières enzymes comme la polymérase alpha n'ont pas d'activité nucléasique. Seuls alpha delta et gamma ont des activités nucléasiques de lecteurs d'épreuves.

Deux autres complexes de protéines fonctionnent aussi dans la réplication des ADN des cellules eucaryotes. RPA est une protéine qui s'attache à l'ADN simple-brin pendant la réplication et qui a donc une fonction équivalente à celle de la protéine SSB de *E. coli*. RFC fonctionne comme une pince qui accroche la PCNA et facilite l'assemblage de complexes de réplication actives.

La linéarité des chromosomes eucaryotes pose un problème de terminaison. En effet, la terminaison de la réplication sur ces chromosomes implique la synthèse de structures appelées télomères aux extrémités des chromosomes.

Chapitre 3 : Mutations de l'ADN

L'ADN peut être endommagé par une variété d'agents physiques et chimiques. Les bases peuvent être altérées ou perdues, et la chaîne d'ADN peut être cassée. Ces mutations peuvent donner naissance à des phénotypes pathologiques.

Les mutations de l'ADN peuvent être de différents types :

Les mutations ponctuelles

a). substitutions ou misappariements: c'est la mutation la plus commune. Elle se produit quand une base est remplacée par une autre. Elle est appelée une transition quand une purine est remplacée par une autre purine ou une pyrimidine est remplacée par une pyrimidine. Elle est appelée une transversion quand une purine est remplacée par une pyrimidine ou l'inverse, c'est-à-dire une pyrimidine est remplacée par une purine.

b). mutations silencieuses: dans ce cas, la mutation n'induit pas de changement d'acide aminé à cause de la dégénérescence du code génétique.

c). mutation non-sens: la mutation aboutit à un codon de terminaison (Stop) qui provoque la terminaison prématurée de la traduction d'une protéine.

d). addition ou délétion d'une base unique: c'est quand une seule base est ajoutée ou délétée, ce qui provoque un décalage du cadre de lecture, aboutissant à la synthèse d'une protéine complètement différente de la protéine du type sauvage.

Il existe d'autres mutations qui touchent des régions beaucoup plus étendues du chromosome (mutations dans la structure) et d'autres qui touchent le nombre de chromosomes ou même le niveau de ploïdie. Celles-ci sont appelées des anomalies ou aberrations chromosomiques.

Ces mutations peuvent être spontanées ou induites. Les mutations spontanées peuvent être la conséquence de plusieurs types d'évènements :

Elles peuvent être provoquées par des erreurs des enzymes pendant la réplication. La fréquence de ces erreurs est très basse vu la grande fidélité des enzymes de réplication. Elle est de l'ordre de 10^{-9} - 10^{-11} . Ces mutations peuvent être ponctuelles ou peuvent être des additions et des délétions.

-Les mutations ponctuelles sont dues à la tautomérisation des bases qui fait que ces bases forment des liaisons hydrogènes différentes, ce qui donne naissance à des misappariements (fig). Par exemple, pour les pyrimidines, la cytosine dans son état tautomérique rare s'apparie avec l'adénine au lieu de la guanine et la thymine

tautomérisées s'apparie avec la guanine au lieu de l'adénine. Pour les purines, la forme tautomérisée de la guanine s'apparie avec la thymine et la forme tautomérisée de l'adénine s'apparie avec la cytosine.

-Les additions et délétions sont provoquées spontanément pendant la réplication quand une boucle est formée dans un brin parental ou dans le brin nouvellement synthétisé, généralement dans des régions où une des bases est répétée. Si la boucle est formée dans le brin matrice, une délétion en résulte. Si, par contre, la boucle est formée dans le brin nouvellement synthétisé, une addition en résulte (fig)(page 602 Russel).

Elles peuvent être aussi le résultat de changements chimiques spontanés : la dépurination et la désamination de bases particulières. Dans la dépurination, une purine, l'adénine ou la guanine, est enlevée du brin d'ADN quand la liaison glycosidique entre le sucre et la base est rompue (fig

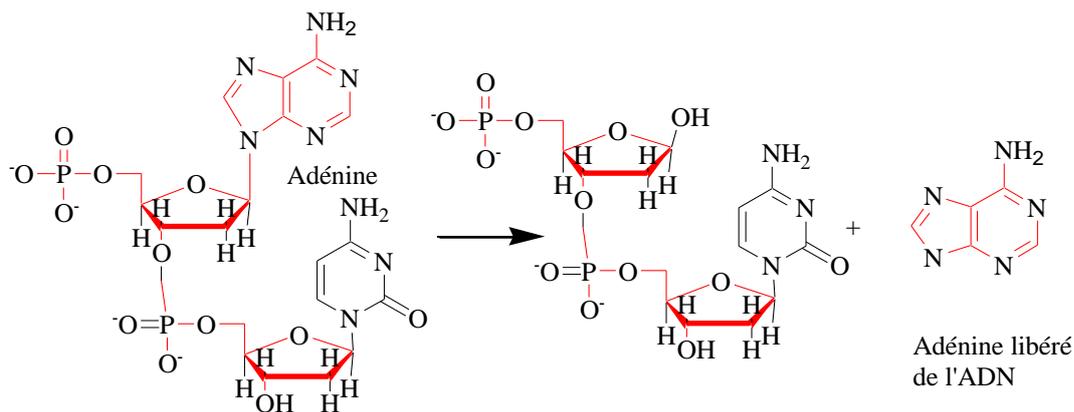


Fig : Dépurination d'un simple-brin d'ADN.

Des milliers de purines sont perdues en une génération dans une cellule de mammifère en culture tissulaire. La désamination est l'enlèvement d'un groupe amine d'une base. Comme exemple, la désamination de la cytosine donne l'uracile (fig). L'ADN des procaryotes et des eucaryotes contient une quantité relativement petite de la base modifiée 5-méthylcytosine (5^mC) à la place de la cytosine. La désamination de la 5^mC donne la thymine résultant en des transitions de CG à TA (fig). Comme la thymine est une base normale de l'ADN, il n'y a aucun mécanisme de réparation qui peut détecter et réparer ces mutations. La localisation de 5^mC dans le génome est appelée ainsi des «hot spots» ou des régions chaudes car la fréquence des mutations dans ces régions est plus élevée que la moyenne.

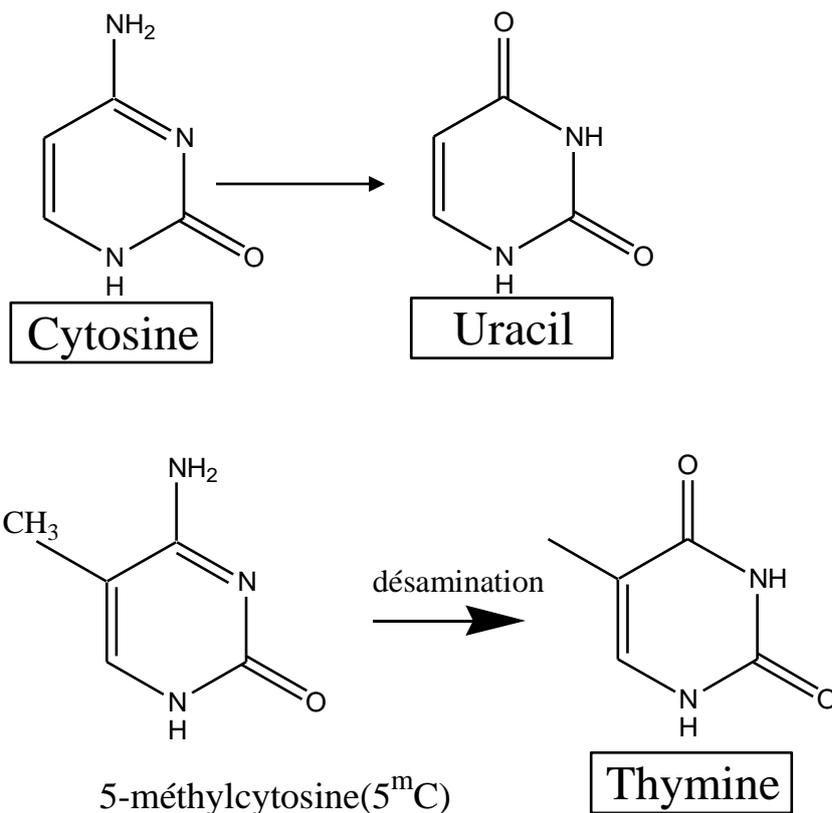


Fig : la désamination de la 5^mC en thymine.

Les mutations peuvent aussi être induites par des agents physiques ou chimiques. Ces agents capables d'induire une mutation sont appelés des mutagènes et le processus est appelé la mutagenèse.

Les mutagènes :

Les agents physiques.

Les mutagènes physiques sont représentés par les rayons UV et les rayons X. Les UV causent la formation des dimères de pyrimidines c'est à dire la formation de deux liaisons covalentes entre deux pyrimidines adjacentes surtout entre deux thymines. Cette dimérisation provoque la distorsion du brin touché à cause de l'impossibilité de la formation de liaisons hydrogènes entre les bases complémentaires (fig).

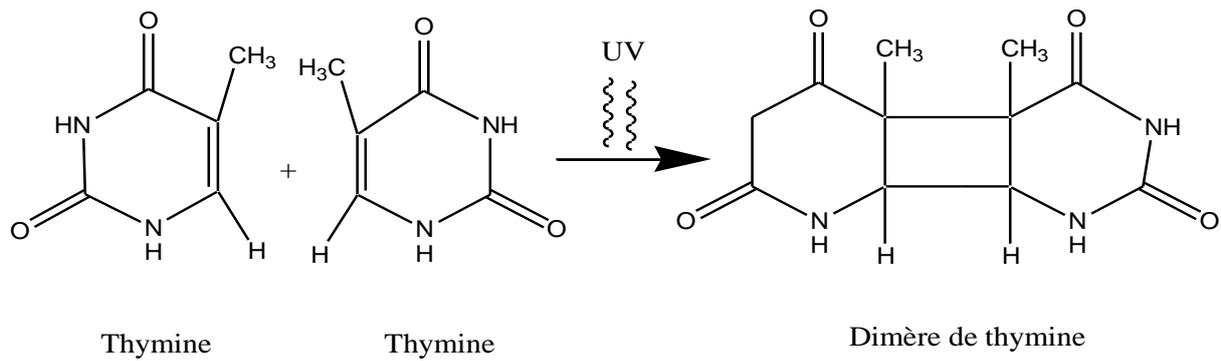


Fig : Production de dimère de thymine par UV. Les deux composés du dimère sont liés de façon covalente.

Les rayons X sont des radiations ioniques qui peuvent provoquer des cassures et des réarrangements chromosomiques importants comme les inversions et les translocations et d'autres dégâts comme des mutations ponctuelles.

Les mutagènes chimiques :

Les analogues de base :

a) Le 5-bromouracil (5-BU) : c'est un analogue de la thymine donc il peut le remplacer pendant la réplication. Son tautomère s'apparie avec la Guanine au lieu de l'Adénine, ce qui cause une transition AT \longrightarrow GC (fig).

b) Le 2-aminopurine (2-AP) : c'est un analogue de l'adénine mais dans sa forme tautomérique, il ressemble à la guanine et s'apparie ainsi avec la cytosine (fig). Il provoque ainsi des transitions AT \longrightarrow GC ou GC \longrightarrow AT.

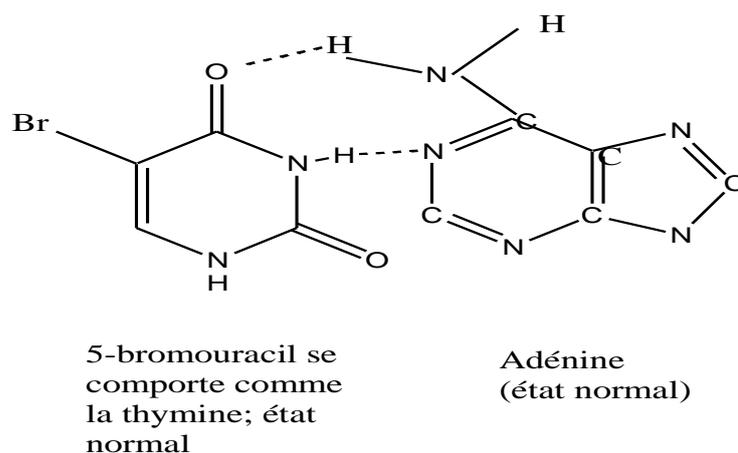


Fig : l'appariement de base de 5-bromouracil dans son état normal

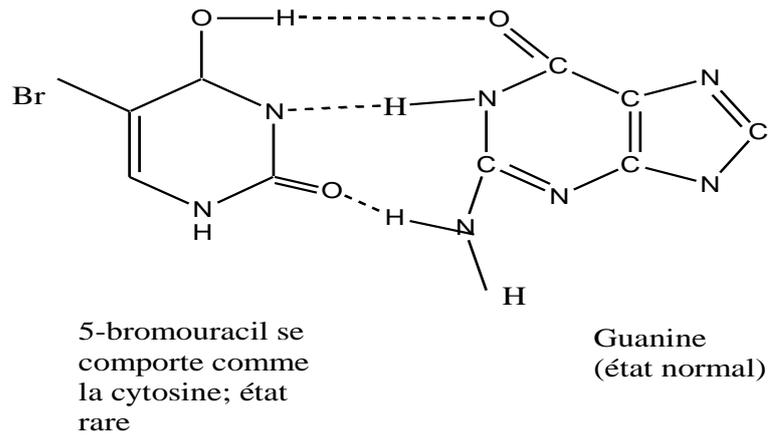
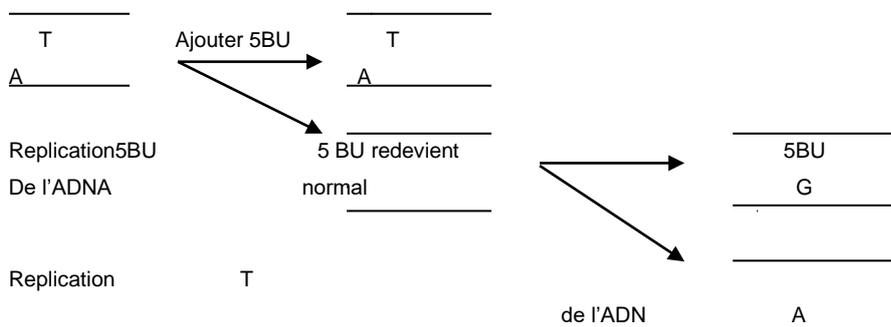


Fig : l'appariement de base de la 5-bromouracil dans son état rare



Les modificateurs de base. Ces agents agissent directement en modifiant directement la structure chimique ainsi que les propriétés des bases.

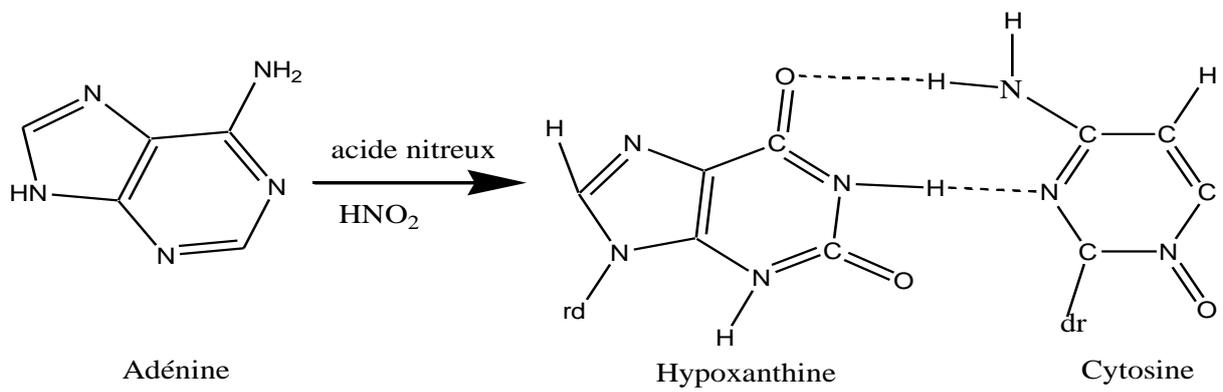
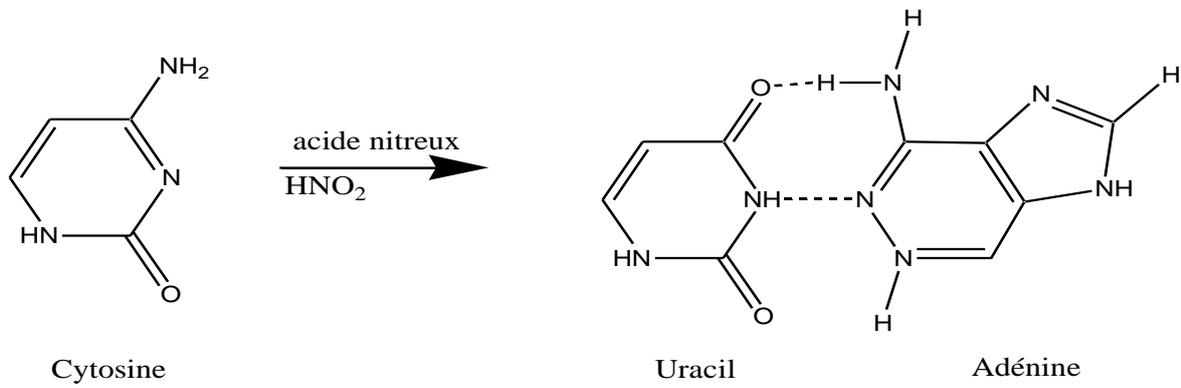
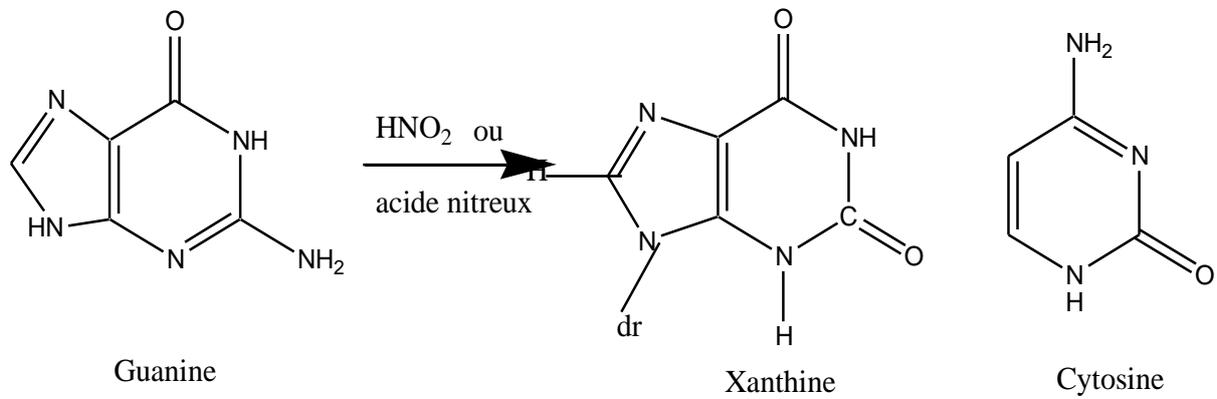
a) L'acide nitreux (HNO_2): il agit en désaminant la guanine, la cytosine et l'adénine. La désamination de la guanine donne la xanthine qui s'apparie avec la cytosine, ce qui ne donne aucune mutation. La désamination de la cytosine donne l'uracile qui s'apparie avec l'adénine et cause une transition $\text{CG} \rightarrow \text{TA}$. La désamination de l'adénine donne l'hypoxanthine qui s'apparie avec la cytosine et donne des transitions $\text{AT} \rightarrow \text{GC}$.

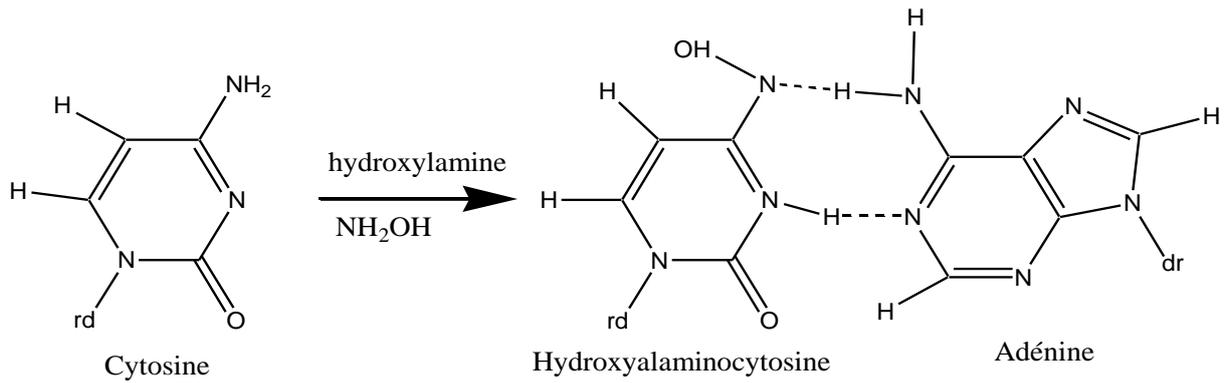
b) L'hydroxylamine (NH_2OH) : Il réagit avec la cytosine en ajoutant un groupe hydroxyl (OH) de façon à ce qu'il s'apparie avec l'adénine au lieu de la guanine induisant ainsi une mutation $\text{CG} \rightarrow \text{TA}$.

c) les agents alkylants représentent le groupe le plus large de mutagènes. On en cite le gaz moutarde, l'époxide, le diméthylsulfonate et le méthylméthanesulfonate (MMS). Ce dernier introduit des groupes alkyls ($-\text{CH}_3$, $-\text{CH}_2\text{CH}_3$) à certaines bases comme les guanines et les thymines qui deviennent méthylées. Du coup, la guanine méthylée s'apparie avec la thymine au lieu de la cytosine, donnant des transitions de $\text{GC} \rightarrow \text{AT}$



et une thymine méthylée s'apparie avec la guanine au lieu de l'adénine donnant des transitions de TA CG.





Les agents intercalaires : On en cite la proflavine (fig), l'acridine orange, le bromure d'éthidium et l'ICR-170. Ils agissent en s'intercalant entre deux bases adjacentes dans une ou les deux chaînes de la double hélice d'ADN. Le résultat est une addition ou une délétion suivant à ce que l'agent intercalaire s'insère dans la chaîne qui joue le rôle de matrice pendant la réplication ou dans la chaîne nouvellement synthétisée.

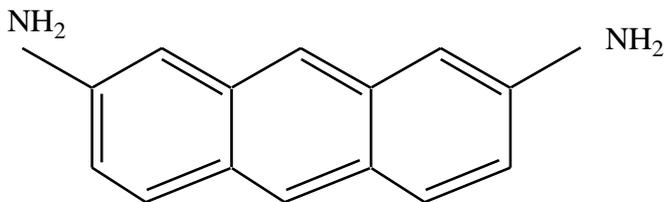


Fig Proflavine

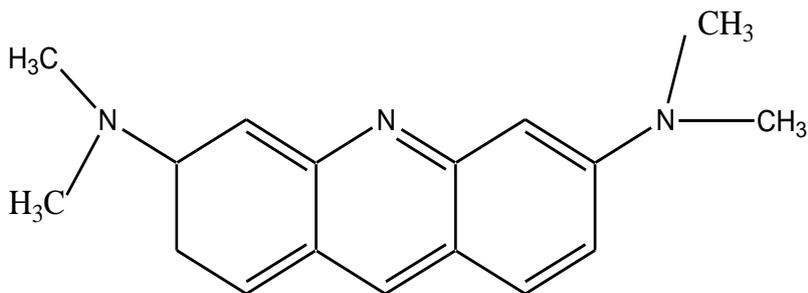


Fig Acridine orange

Chapitre 3 : Mécanismes de réparation des lésions.

L'excision des dimères de pyrimidine formés par l'exposition de l'ADN aux rayons UV peut être réalisée par plusieurs systèmes de réparation.

Systèmes exempts d'erreur :

1) La photoréactivation: C'est un système qui utilise la lumière visible d'une longueur d'onde qui varie entre 320 et 370nm pour stimuler une enzyme spécifique appelée photolyase ou enzyme photoréactivante (PRE) qui s'associe au dimère et provoque sa dissociation. La photolyase est une enzyme ubiquitaire et apparemment très efficace puisque très peu de dimères de thymine sont laissés non corrigés après son passage.

2) Réparation des lésions dues à des agents alkylants: Dans ce cas, des systèmes de réparation spécifiques interviennent. On prend comme exemple, une enzyme codée par le gène «ada» appelée 6-méthyl-guanine méthyl transférase, qui reconnaît le 6-méthyl-guanine dans l'ADN et enlève le groupe méthyl, le changeant à sa forme originale.

3) Réparation par excision: C'est le mécanisme le mieux connu parmi les mécanismes de réparation. Il a été découvert en 1964 par Setlow et Carrier. Les dimères de pyrimidine bloquent la réplication et l'expression des gènes. Un complexe enzymatique codé par *uvrABC* détecte la distorsion créée par le dimère de thymine. Ce complexe enzymatique, appelé excinuclease est une endonucléase qui coupe des deux côtés de la lésion de 8 nucléotides du côté de 5' et de 4 nucléotides du côté de 3'. Le résultat est l'excision d'un fragment d'ADN de 12 paires de bases et le vide est rempli par la polymérase I et liés par la ligase. Ce système opère seulement dans l'obscurité et est appelé système «darkrepair». Il répare non seulement les dimères de thymine mais toute lésion qui crée une distorsion dans l'ADN comme celles qui sont créées par l'aflatoxine B (AFB) et les époxydes de benzopyrène. Ce système existe dans la plupart des organismes étudiés jusqu'à maintenant. Chez la levure et les mammifères 12 gènes sont connus qui codent des protéines impliquées dans le système de réparation par excision. Des mutations dans ces gènes qui donnent des enzymes de réparation défectueuses peuvent être à l'origine de maladies graves, comme la Xérodémie pigmentée. C'est une maladie à transmission autosomique récessive causée par une déficience dans l'un des gènes codant pour les protéines du système d'excision. Les symptômes cliniques de cette

maladie montrent l'importance des systèmes de réparation de l'ADN. La peau est très sensible à la lumière du soleil ou UV. Des ulcérations sur la peau et autour de la cornée de l'œil commencent à apparaître dès l'enfance. Ces ulcérations peuvent donner le cancer de la peau et les malades meurent en général avant l'âge de trente ans à cause des métastases.

- Réparation par les ADN glycosylases

Chaque cellule possède une classe d'enzymes appelées «ADN glycosylases» qui reconnaissent des lésions d'ADN communes comme les produits de la désamination de l'adénine et la cytosine. Ces enzymes clivent la liaison N-glycosidique entre le sucre et la base libérant ainsi la base altérée et créant un site apurinique ou apyrimidique AP. Celui-ci est ensuite réparé par une AP endonucléase. On trouve plusieurs sortes de glycosylases : uracile glycosylase, par exemple, qu'on trouve dans la plupart des cellules, enlève de l'ADN l'uracile altéré. Cette glycosylase n'enlève pas des uraciles altérés de l'ARN ou des thymine de l'ADN. Son action est très spécifique. D'autres ADN glycosylases reconnaissent et enlèvent l'hyoxanthine et les bases alkylées comme la 3-méthyladénine et la 7-méthylguanine. Des glycosylases qui reconnaissent des dimères de pyrimidines ont été aussi identifiées. Il faut noter que des sites apuriniques sont aussi générées par une mutation spontanée, une hydrolyse lente des liaisons N-glycosyl.

3) Systèmes post-réplicatifs ou prônes à l'erreur.

Les systèmes de réparation par excision fonctionnent quand un seul des deux brins d'ADN est endommagé. L'autre brin peut alors agir comme matrice pour réparer le brin endommagé. Cependant, quand les deux brins sont endommagés créant des cassures de l'ADN double brin, ou quand l'ADN est simple brin, la cellule doit recourir à d'autres mécanismes de réparation.

Deux systèmes peuvent entrer en jeu.

a). réparation par recombinaison génétique homologue: dans ce système la région où se trouve la lésion est appariée au chromosome homologue avec lequel un échange de fragments s'effectue appelé recombinaison génétique homologue d'où un échange d'informations pour réparer le vide. Ce modèle de recombinaison a été proposé en 1964 par Holliday. Dans ce modèle, deux duplexes homologues sont alignés. Un des brins d'un duplexe et le brin correspondant de l'autre duplexe sont coupés par une endonucléase. L'extrémité de chaque brin coupé quitte son propre duplexe et se déplace vers l'autre

duplexe. Les brins des deux différents duplexes sont ensuite joints l'un à l'autre pour former un intermédiaire de recombinaison qui est dynamique. L'échange de brins peut continuer, permettant le déplacement du crossing-over, un processus appelé migration de branche ou «branch migration». Cet intermédiaire peut être ensuite clivé et rejoint de deux façons différentes pour donner deux produits de recombinaison différents (voir figure).

Le processus de recombinaison homologue chez *E. coli* dépend des gènes rec. La protéine RecA, produit du gène recA, catalyse la réaction d'assimilation de brins avec l'hydrolyse d'un ATP. Le produit de la réaction est un simple brin déplacé et un double-brin, ayant la forme de la lettre D, d'où l'appellation de structure de D-boucle. Le fonctionnement de cette protéine est le suivant: la protéine recA s'attache à une région d'ADN simple brin. Un filament de molécules recA s'enroule autour de l'ADN simple brin avec un monomère de recA par 4 nucléotides. Ce filament recA-simple brin d'ADN se lie ensuite à l'ADN double brin. La protéine recA se déplace le long de l'ADN double brin pour trouver une séquence complémentaire à celle de l'ADN double brin. L'ADN double brin est en partie déroulé pour faciliter la lecture de sa séquence de base. Quand une séquence complémentaire est trouvée, le duplexe est déroulé encore plus pour permettre un changement de pair de base. Quand l'ADN simple brin est apparié avec le brin complémentaire du duplexe ciblé. Une fois initiée de cette façon, l'échange de brins continue, un processus appelé «branch migration» Les actions de recombinaison de la recA sont aidées par l'hydrolyse de l'ATP. Le filament recA a une polarité de 5' à 3'. Cette unidirectionnalité va permettre à l'ADN simple brin de dépasser les régions endommagées du duplexe qui ne lui sont pas complémentaires. L'hydrolyse de l'ATP aide le «branch migration» en enroulant l'ADN duplexe autour du filament. La recombinaison requiert que l'ADN invasif soit simple brin. Les études ont montré que les gènes recB, recC et recD sont importants pour la recombinaison. Ils codent des protéines qui forment un complexe recBCD de 328Kd. Le recB catalyse le déroulement de l'ADN doublet-brin avec l'hydrolyse de l'ATP la recD est une nucléase. Le complexe s'attache à une extrémité de l'ADN double brin linéaire et le déroule à un taux de 300 nucléotides par seconde.

Le clivage se fait à une séquence appelée chi. 5'GCTGGTGG3'

Ex de recombinaison

Les gènes des immunoglobulines sont assemblés par recombinaison. Quelques réarrangements d'ADN sont programmés pendant le développement chez les eucaryotes. Un mammifère est capable de produire des millions d'immunoglobulines différents avec des spécificités différentes même si le génome humain contient seulement 100000 gènes. Les études ont démontré un mécanisme de recombinaison très proche de la transposition et suggèrent que ce système a évolué probablement d'une ancienne invasion de transposons. Les gènes humains codant les immunoglobulines G illustre un exemple de recombinaison. Les immunoglobulines sont générées par la combinaison de différentes régions variables des chaînes lourdes avec les chaînes kappa ou lambda de la chaîne légère. Dans les cellules non-différenciées les gènes sont divisés en 3 segments qui codent pour la chaîne polypeptidique légère kappa. Le segment V code les premiers 95 résidus d'acides aminés de la région variable, le segment J les 12 acides aminés restants de la région variable et le segment C qui code pour la région constante. Le génome contient à peu près 300 V segments différents, 4 segments J différents et un segment C.

Quand une cellule de moelle osseuse non-différenciée se différencie pour former une cellule osseuse mature, un segment V et un segment J sont combinés par un système de recombinaison spécialisé. Pendant cette délétion d'ADN programmé, l'ADN délété est dégradé. Il y a $300 \times 4 = 1200$ combinaisons V-J possibles. Ce processus n'est pas aussi précis que la recombinaison dirigée. Une variabilité additionnelle est générée aux jonctions V-J de l'ordre de 2,5 ce qui fait que le nombre total de combinaisons V-J est de $2,5 \times 1200 = 3000$.

La combinaison finale de la V-J avec la région C est accomplie par un épissage. Les gènes des chaînes légères et lourdes de lambda se fait par le même processus. Pour les chaînes lourdes, il y a plus de segments de gènes et plus de 5000 combinaisons. Une autre variabilité est créée parce que les séquences V subissent beaucoup de mutations. Chacune des lymphocytes B produit un seul type d'anticorps mais le nombre d'anticorps produits par toutes les cellules est énorme.

a). La réparation par le système SOS: celui-ci est stimulé quand l'exposition aux rayons UV est très forte et que les systèmes d'excision sont saturés et n'arrivent pas à réparer tous les dimères de pyrimidine. Ce système dépend d'une protéine RecA qui, activée par la présence de dimères, stimule l'autoclivage d'un répresseur de 17 gènes: le LexA. Le

clivage de ce dernier l'inactive et stimule l'expression des 17 gènes qu'il réprime. Les protéines codées par ces gènes sont donc exprimées à un niveau élevé. Parmi eux, les protéines UvrA, UvrB impliquées dans les systèmes de réparation par excision. Une autre protéine, UmuD est clivée en une forme plus courte, UmuD', qui forme un complexe avec UmuC pour créer la polymérase V. Cette dernière enzyme est capable de répliquer sans enlever les dimères de pyrimidine, mais avec un niveau élevé d'erreurs. Une autre enzyme, dont le gène *dinB* est stimulé par ce système, est produite. C'est la polymérase IV qui réplique avec une fréquence élevée d'erreurs. Ce système est donc hautement mutagène, ce qui tue les cellules parfois. Seulement, il est nécessaire à cause du fait que c'est un système d'urgences où la cellule est en condition de désespoir.

b) Système de réparation mismatch. Ce système opère après la réplication chez *E. coli* pour réparer les erreurs dues à la réplication. Il augmente ainsi la fidélité de réplication par un facteur de 10^2 à 10^3 . Ce système doit distinguer entre la matrice et le brin d'ADN nouvellement synthétisé. Cela est accompli grâce à une protéine, la Dam méthylase, codée par le gène *dam*, et qui méthyle toutes les adénines d'une séquence 5'GATC3' chez *E. coli* dans la chaîne matrice. Ceci est accompli pendant une période de temps très courte après la réplication. La réparation des erreurs mismatch du brin nouvellement synthétisé se fait toujours dans le voisinage de la séquence GATC jusqu'à une distance maximale de 1000 paires de base de cette séquence. Le fonctionnement de ce système chez *E. coli* est sous l'influence de trois gènes: *mutS*, *mutL* et *mutH* comme suit:

-la protéine MutL forme un complexe avec MutS qui s'associe aux paires de bases misappariées (sauf C-C).

On remarque que la structure de l'ARN est très similaire à celle de l'ADN sauf que le sucre est le ribose au lieu du désoxyribose et la base uracile remplace la thymine. L'ARN est aussi simple brin en général avec des régions à structure double brin : les épingles. Ces régions double brin peuvent former jusqu'à 50% de la molécule d'ARN.

Les cellules contiennent différents types d'ARN. L'ARN messager est la matrice pour la synthèse de protéines. Une molécule d'ARN est produite pour chaque gène ou groupe de gènes exprimés. Donc, l'ARNm est une classe très hétérogène de molécules. Chez *E. coli*, la longueur moyenne d'un ARNm est de 1,2Kb. L'ARN de transfert (ARNt) porte des acides aminés en une forme activée au ribosome pour la formation d'une liaison peptidique. Il existe au moins un ARNt pour chacun des 20 acides aminés. L'ARNt consiste de à peu près 75 nucléotides qui le rend le plus petit des molécules d'ARN. L'ARNr (ribosomal) est le composé majeur des ribosomes, mais son rôle précis dans la synthèse des protéines n'est pas encore connu. Chez *E. coli*, il existe trois types d'ARNr, appelés 23S, 16S et 5S ARN à cause de leur comportement pendant la sédimentation. Chaque ribosome contient une molécule de ces trois types d'ARNr. L'ARNr est le plus abondant. L'ARNt vient après et ensuite l'ARNm, qui constitue seulement 5% de l'ARN total. Les cellules eucaryotes contiennent des petites molécules d'ARN additionnelles, comme les molécules d'ARNsn ou petites nucléaires, qui participent dans l'épissage des exons. Une petite molécule d'ARN dans le cytosol joue un rôle dans le ciblage des protéines nouvellement synthétisées.

L'ARN messager

Le concept d'ARN messager a été formulé par François Jacob et Francis Monod en 1961. Ils ont préconisé la présence d'un intermédiaire qui fait passer le message du noyau dans lequel on trouve les gènes au cytoplasme, le site de synthèse des protéines. Les points les plus importants sont venus de leur expérience sur l'induction d'un système enzymatique par la présence du lactose. Ils ont proposé le concept du messager avec les propriétés suivantes:

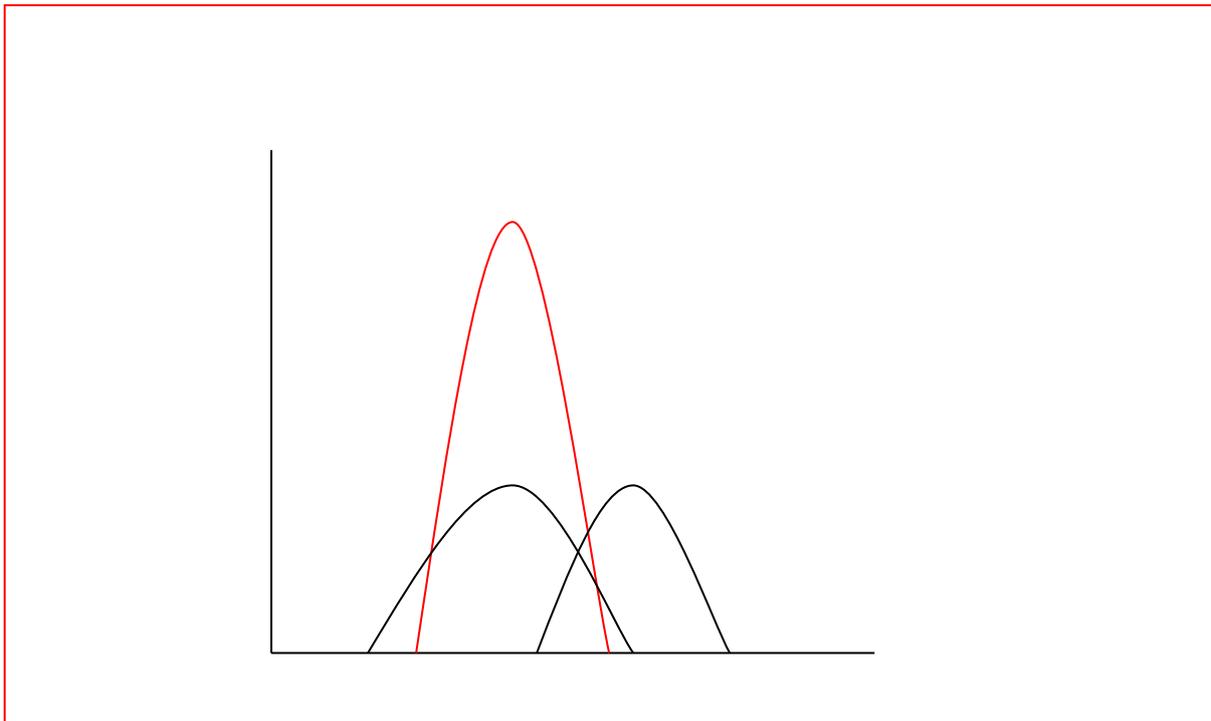
- 1). Le messager doit être un polynucléotide.
- 2). La composition de base du messager doit être similaire à celle de l'ADN qui le spécifie.
- 3). Les messagers doivent être de tailles différentes car les gènes ou les groupes de gènes sont de tailles différentes.

4). Le messenger doit être associé de façon transitoire avec les ribosomes, les sites de synthèse de protéines.

5). Le messenger doit être synthétisé et dégradé rapidement.

Après avoir énoncé le concept d'ARN messenger, une expérience réalisée par Brenner, Jacob et Messelson a prouvé son existence. Leur étude faite sur des cellules de *E. coli* infectées par le bactériophage T₂ a montré qu'une fraction mineure d'un ARN avec une demi-vie de courte durée apparaît juste après l'infection. Les bactéries ont été cultivées dans un milieu contenant des isotopes lourds (N^{15} et C^{13}), ensuite elles ont été infectées par les phages et transférées dans un milieu contenant des isotopes légers (N^{14} et C^{12}).

Les ribosomes synthétisés avant et après l'infection sont séparés par la technique de centrifugation sur gradient de densité à cause de la différence dans leur densité (lourds et légers). En plus, l'ARN nouveau a été marqué par le P^{32} ou C^{14} -uracile et les protéines nouvellement synthétisées par S^{35} . Les résultats montrés par la courbe ci-dessous montrent que:



1). Les ribosomes ne sont pas synthétisés après l'infection, ce qui est montré par l'absence de ribosomes «légers».

2). L'ARN est synthétisé après l'infection. La plupart de l'ARN marqué se trouve dans le pic des ribosomes «lourds». Donc, il est associé avec les ribosomes préexistants et non pas nouvellement synthétisés.

3). L'isotope S^{35} est apparu dans le pic des ribosomes «lourds», ce qui a montré que les protéines nouvellement synthétisées existent le sont dans des ribosomes préexistants.

Ces expériences ont montré que les ribosomes sont des structures non spécialisées, qui synthétisent, à un certain temps, des protéines dictées par le message qu'ils contiennent.

En 1961, Sol Spiegelman a développé une technique appelée l'hybridation en menant l'expérience suivante :

- 1). Un échantillon d'ARNm de T_2 a été marqué par le P^{32} . L'ADN de T_2 marqué avec H^3 est préparé séparément.
- 2). Un mélange d'ARNm et d'ADN de T_2 est chauffé jusqu'à $100^{\circ}C$, ce qui aboutit à la dénaturation de la double hélice d'ADN en simple brin. Cette solution d'ADN et d'ARN simple brin a été refroidi lentement à la température ambiante.
- 3). Le mélange refroidi est ensuite analysé par centrifugation sur gradient de densité.

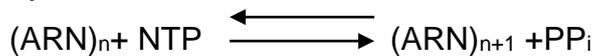
Le résultat de l'expérience a révélé trois bandes. La plus dense est constituée de l'ARN simple brin. La deuxième contient de l'ADN double brin. La troisième contient des molécules hybrides d'ADN et d'ARN à côté de la bande d'ADN double brin. Ainsi, l'ARNm de T_2 a formé un hybride avec l'ADN de T_2 . D'autres expériences ont montré que l'ARNm ne s'hybride pas avec l'ADN dérivé d'une variété de bactéries et d'autres organismes. Ces expériences ont révélé que la séquence de bases de l'ARNm est complémentaire à celle de l'ADN matrice.

La technique d'hybridation a permis de démontrer que les l'ARNr et l'ARNt sont aussi synthétisés en utilisant une matrice d'ADN. Les hybrides ADN-ARN ou ADN-ADN sont retenus par un filtre de nitrocellulose alors que les ADN ou ARN simple-brin ne le sont pas. Plusieurs types d'ARN marqués par le P^{32} ont été ajouté à de l'ADN non marqué de la même espèce. Ces mélanges sont chauffés, ensuite refroidis lentement, puis filtrés à travers le filtre de nitrocellulose. La radioactivité retenue par le filtre a été mesurée. Les résultats ont montré que des hybrides sont formés avec les trois types d'ARNr, et avec l'ARNt, montrant que ces molécules contiennent des séquences complémentaires à ces molécules d'ADN qui sont présentes dans le génome de *E. coli*.

L'ARN polymérase de *E. coli*.

L'ARN polymérase, l'enzyme qui synthétise l'ARN à partir d'une matrice d'ADN a été découverte indépendamment en 1960 par deux chercheurs: J. Hurwitz et S. Weiss. Cette enzyme requiert la présence des composés suivants :

- 1). Une matrice. La matrice préférée est l'ADN double-brin mais l'ADN simple-brin peut agir comme matrice. L'ARN, ou l'hybride ARN-ADN ne sont pas des matrices efficaces.
- 2). Des précurseurs activés. Tous les quatre nucléosides triphosphates: ATP, GTP, UTP, et CTP doivent être présents.
- 3). Les ions divalents Mg^{++} et Mn^{++} sont efficaces, mais les ions Mg^{++} sont requis in vivo.
- 4). L'ARN polymérase catalyse l'initiation et l'élongation de chaînes d'ARN. La réaction de catalyse est la suivante :



La synthèse d'ARN ressemble beaucoup à celle d'ADN.

1° La direction de synthèse est de 5' à 3'. 2°. Le mécanisme d'élongation est le même. Le groupe 3'OH produit une attaque nucléophile sur le phosphate interne d'un nucléoside triphosphate. 3° La synthèse est commandée par l'hydrolyse du pyrophosphate. D'autre part, l'ARN polymérase ne requiert pas une amorce. Une autre différence est que l'ARN polymérase ne possède pas l'activité nucléasique de lecture d'épreuves que possède l'ADN polymérase.

L'ARN polymérase de *E. coli* est une holoenzyme de 450kd qui est formée de 6 sous unités: $\alpha_2\beta\beta'\omega\sigma$. La sous unité σ s'attache au site promoteur où la transcription débutera. Une fois la synthèse de l'ARN est initiée, cette sous-unité se dissocie du reste de l'enzyme. L'ARN polymérase démunie de cette sous-unité est appelée enzyme «core» qui catalyse la réaction de synthèse de l'ARN. La sous-unité β' s'attache à la matrice, et la sous-unité β aux précurseurs de ribonucléosides triphosphates. L'ARN polymérase a plusieurs fonctions:

- 1). Elle cherche les sites d'initiation. Il existe 2000 sites promoteur dans le génome de *E. coli* de 4×10^6 paires de base.
- 2). Elle déroule un petit segment d'ADN double-brin qui servira de matrice pour la transcription.
- 3). Elle sélectionne le ribonucléoside triphosphate correcte et catalyse la formation d'une liaison phosphodiester. Cette enzyme a une potence catalytique très élevée, ce qui lui permet de synthétiser un transcrit du début à la fin.

- 4). Elle détecte des signaux de terminaison qui spécifient la fin de la transcription.
- 5). Elle interagit avec les répresseurs et les activateurs qui modulent le taux de transcription. La régulation des gènes se fait surtout au niveau de la transcription.

La transcription se fait en trois étapes: l'initiation, l'élongation et la terminaison.

L'initiation

L'ADN matrice contient des régions appelées sites promoteurs qui se lient spécifiquement à l'ARN polymérase et déterminent le début de la transcription. Ces sites ont été identifiés de la façon suivante:

On ajoute l'ARN polymérase, ensuite l'ADNase à l'ADN matrice. En l'absence de ribonucléosides triphosphates, l'ARN polymérase reste attachée aux promoteurs et les protège de la digestion par l'ADNase. Des fragments de 60 paires de base sont produits après une digestion extensive. La position et la séquence de ces fragments sont ensuite étudiées. Chez les bactéries, trois promoteurs sont identifiés: la boîte Pribnow, qui possède la séquence consensus 5'TATAAT3' et centrée à -10. La deuxième séquence, appelée séquence -35, a la séquence consensus 5'TTGACA3'. Une troisième séquence riche en AT, appelée UP (upstreampromoter), ou promoteur en amont, se trouve entre les positions -40 et -60 dans les promoteurs de certains gènes hautement exprimés. L'élément UP se lie à la sous-unité α de l'ARN polymérase. Le premier nucléotide est en général une purine.

Chez les eucaryotes, un site promoteur similaire à la boîte Pribnow existe à -30. Cette séquence, appelée boîte TATA ou boîte Hogness, a pour consensus TATAAA. Les promoteurs eucaryotes exhibent aussi une boîte CAAT centrée à -75 mais qui peut exister dans d'autres régions chez certains gènes. Elle possède une séquence consensus 5'-GGCCAATCT-3'. La boîte GC a la séquence consensus 5' GGGCGG3' qui peut exister en plusieurs copies dans un promoteur et peut fonctionner dans les deux orientations. Cette boîte existe surtout dans les gènes constitutifs et peut aider l'attachement de l'ARN polymérase à côté du point d'initiation de la transcription. Des séquences «enhancers» existent aussi qui peuvent opérer à distance en amont ou en aval du site d'initiation et qui stimulent la transcription.

L'efficacité avec laquelle l'ARN polymérase s'attache à un promoteur pour initier la transcription dépend beaucoup sur les séquences consensus, la distance entre eux et la distance qui les sépare du site d'initiation de la transcription. Les promoteurs forts

causent des initiations fréquentes chaque deux secondes chez *E. coli* et possèdent des régions -10 et -35 qui correspondent très fortement à la séquence consensus. Les promoteurs faibles ont des substitutions multiples dans leur séquence consensus. La fréquence de la transcription qui dépend de ces promoteurs faibles peut aller jusqu'à une fois chaque 10 minutes. La distance entre ces séquences conservées est importante, avec une distance optimale de 17 nucléotides. Aussi, les protéines régulatrices peuvent influencer fortement la fréquence de la transcription.

L'attachement de l'ARN polymérase au promoteur se fait en deux étapes: la première étape consiste à un faible attachement à la région -35 du promoteur pendant que l'ADN est toujours double-brin. La deuxième étape implique un changement d'un attachement faible à un attachement plus fort entre l'ARN polymérase et l'ADN. Ce changement est accompagné par un déroulement local d'à peu près 17 paires de base autour de la région -10, ce qui est équivalent à deux tours d'hélice de l'ADN. Une fois que l'ARN polymérase s'attache à cette région, elle est correctement orientée pour initier la transcription (fig).

Le facteur sigma (σ) est essentiel pour la reconnaissance de la séquence promotrice sans lequel aucune initiation efficace de la transcription n'est possible. En fait, il existe plusieurs facteurs sigma différents chez *E. coli* qui jouent un rôle important dans la régulation de l'expression des gènes. Chaque facteur sigma s'associe à l'enzyme «core» et lui permet de reconnaître un promoteur spécifique. La plupart des promoteurs avec les séquences consensus discutées ci-dessus sont reconnus par un facteur sigma de poids moléculaire 70000 da, appelé σ^{70} . Sous des conditions extrêmes, comme la chaleur, la quantité d'azote limitée et l'infection avec des virus, des facteurs sigma différents s'attachent à l'ARN polymérase «core» qui s'associent à des promoteurs différents pour stimuler la transcription de gènes spécifiques en réponse du stress.

La plupart des chaînes d'ARN nouvellement synthétisées commencent par pppG ou pppA. Ceci a été découvert en marquant le phosphore en position gamma (γ) de l'ATP avec le P^{32} dans un mélange d'incubation synthétisant l'ARN. Cet ATP a été détecté dans la chaîne d'ARN, signifiant ainsi que le nucléotide terminal est un pppA.

Elongation

La dénaturation de l'ADN dans la région de synthèse de l'ARN forme une bulle de transcription qui contient toujours 17 paires de base déroulées (fig). Après la transcription de 8 ou 9 nucléotides d'ARN, le facteur sigma quitte la polymérase et peut être utilisé dans d'autres réactions d'initiation. Dans la région de dénaturation de l'ADN, quelques

bases d'ARN sont liées à l'ADN et forment un hybride ADN-ARN d'une longueur de 12pb alors que le reste de l'ARN est déplacé loin de l'ADN. La vitesse de transcription moyenne est de 30-50 nucléotides /seconde. Comme pour la réplication de l'ADN, la tension créée par le déroulement de l'ADN est libérée par la topoisomérase II. La transcription se fait toujours dans la direction 5' → 3', comme dans la synthèse de l'ADN. Ce fait est montré par l'expérience suivante: quand on utilise le GTP ou l'ATP marqué avec le γP^{32} comme substrat dans un milieu d'incubation qui synthétise l'ARN, on trouve que le rapport de l'incorporation de P^{32} sur le total des nucléotides incorporés est le plus élevé juste après le mélange des composés puis diminue progressivement.

Il est à noter que puisque l'ARN polymérase ne possède pas d'activité nucléasique, la fidélité de transcription est moindre que celle de l'ADN polymérase. Le taux d'erreur de la synthèse d'ARN est de $1/10^4$ ou de $1/10^5$. Ce taux d'erreur plus élevé que celui de l'ADN polymérase est toléré parce qu'il n'est pas hérité par la progéniture. En plus, il y a plusieurs transcrits d'un même gène qui ne contiennent pas tous des erreurs.

Terminaison de la transcription.

La phase de terminaison est aussi précisément contrôlée que l'initiation. L'arrêt de la transcription dépend de la présence de signaux de terminaison. On distingue au moins deux classes de signaux chez *E. coli*. La première classe ne dépend pas d'une protéine appelée ρ et la deuxième classe dépend de cette protéine.

Terminaison-indépendante de la protéine ρ .

La première classe est la plus simple. C'est une région palindromique riche en GC suivie par une région riche en AT. Le transcrit d'ARN contenant une région auto-complémentaire va former une structure en épingle stable à cause de sa richesse en GC. Cette structure est suivie par quatre ou plus de U. Elle va contribuer à la terminaison de la transcription de la façon suivante :

- 1). Il est probable que l'ARN polymérase ralentit quand elle rencontre la structure en épingle.
- 2). L'hybride ADN-ARN qui se trouve après la structure en épingle est instable à cause de son contenu en A-U pairs de base.
- 3). Donc, il y aura dissociation de l'ARN de sa matrice d'ADN et ensuite de l'enzyme.

L'ADN double brin se reformera dans la région de la bulle de transcription après la libération de l'ARN nouvellement synthétisée. L'enzyme «core» a beaucoup moins d'affinité pour l'ADN double-brin que pour l'ADN simple brin, d'où son détachement.

Terminaison- dépendante de la protéine Rho.

Pour certains gènes, la terminaison de la transcription dépend de la présence d'un facteur additionnel: la protéine Rho. Ceci a été démontré par le fait que certaines molécules d'ARN synthétisées *in vitro* par l'ARN polymérase seule sont plus longues que celles synthétisées *in vivo* du même gène. Par exemple, l'ARN synthétisé de l'ADN du phage filamenteux M13 en la présence de « ρ » a un coefficient de sédimentation de 10S alors que celui qui est synthétisé en son absence a un coefficient de sédimentation de 23S. Une information additionnelle sur l'action de « ρ » a été obtenue en ajoutant ce facteur de terminaison à un mélange d'incubation à des temps variables après l'initiation de la transcription. On a obtenu de l'ARN de 13S, 17S, et 23S après l'addition de ρ quelques secondes après le début de la transcription, deux minutes et 10 minutes respectivement (fig). Ce résultat indique que l'ADN contient trois sites de terminaison qui répondent à ρ et un site de terminaison qui ne répond pas à ρ .

Le mécanisme par lequel ρ provoque la terminaison de la synthèse est indiqué par le fait que cette protéine hydrolyse l'ATP en présence d'ARN simple brin mais pas en la présence de l'ADN ou l'ARN double-brin. Rho est un hexamère autour duquel un fragment d'ARN simple brin de 72 nucléotides s'enroule. Elle possède une activité ATPasique qui lui permet de se déplacer de façon unidirectionnelle au long de l'ARN jusqu'à arriver à la bulle de transcription. Elle dissocie ensuite l'hybride ADN-ARN en déplaçant l'ARN loin de l'ADN. D'autres protéines chez *E. coli* ont été découvertes et qui causent la terminaison de la transcription. Par exemple, la protéine nusA permet à l'ARN polymérase de *E. coli* de reconnaître une classe spécifique de sites de terminaison.

Modifications post-transcriptionnelles:

Chez les procaryotes, les molécules d'ARNm subissent très peu ou aucune modification post-transcriptionnelle. En fait, la plupart sont traduites pendant la transcription. Cependant, l'ARNr et l'ARNt sont produits par clivage et modification de l'ARN natif. Par exemple, chez *E. coli*, trois types d'ARNr et un type d'ARNt sont produits par le clivage d'ARN primaire qui contient aussi des espaces «spacers». D'autres transcrits primaires

contiennent plusieurs types d'ARNt ou plusieurs copies d'un même ARNt (fig). Les nucléases qui clivent ces molécules primaires d'ARN sont hautement précises.

Un deuxième type de modification est l'addition de nucléotides aux extrémités de certaines chaînes d'ARN. Par exemple, un CCA est ajouté à l'extrémité 3' des molécules d'ARNt qui ne possèdent pas encore cette extrémité.

Un troisième type de modification est celui de bases et du ribose des molécules d'ARNr. Chez les procaryotes, certaines bases sont méthylées, alors que chez les eucaryotes le groupe 2'-OH de 1/100 riboses est méthylé. Des bases étranges sont présentes dans toutes les ARNt, comme le pseudouridylate et ribothymidylate. Ces modifications génèrent de la diversité.

Transcription chez les eucaryotes

Chez les eucaryotes, on trouve trois types d'ARN polymérases qui diffèrent dans la spécificité de la matrice, la localisation et la susceptibilité aux inhibiteurs. L'ARN polymérase I est localisée dans le nucléole, et elle transcrit une série de gènes pour l'ARNr 18S, 5,8S et 28S. L'ARNr 5S et tous les ARNt sont synthétisés par l'ARN polymérase III, qui est localisée dans le nucléoplasme. Les précurseurs de l'ARNm sont synthétisés par la polymérase II, localisée aussi dans le nucléoplasme. Plusieurs d'autres petites molécules comme le U1sn ARN de l'épissage sont aussi synthétisées par l'ARN polymérase II.

Les facteurs de transcription de base

L'ARN polymérase II a été étudiée de façon extensive. Elle est beaucoup plus complexe que celle des procaryotes. Elle est constituée de 12 sous-unités et ne peut pas transcrire *in vitro* sans l'intervention de certaines protéines appelées les facteurs de transcription de base (TFII) avec lesquelles elle forme un complexe de transcription. Ces facteurs de transcription sont hautement conservés parmi les eucaryotes. La transcription est divisée en quatre phases : l'assemblage, l'initiation, l'élongation et la terminaison.

L'assemblage :

La formation d'un complexe fermé commence quand la protéine TBP (TATA binding protein) s'associe à la boîte TATA; le TBP est ensuite lié par le facteur de transcription TFIIB qui se lie aussi à l'ADN des deux côtés de TBP. L'association de TFIIA qui n'est pas essentielle, peut stabiliser le complexe TFIIB-TBP surtout sur des promoteurs non-consensus où l'association de TBP avec l'ADN est faible. Un autre complexe, formé de

TFIIF et l'ARN polymérase II s'associe avec celui-là. En interagissant avec le TFIIB, le TFIIF aide l'ARN polymérase II à trouver ses promoteurs. Finalement, le TFIIH et le TFIIIE s'associent pour former un complexe fermé (fig). Le TFIIH agit comme une hélicase qui va dérouler l'ADN pour initier la transcription créant ainsi un complexe ouvert.

L'initiation

TFIIH possède une fonction supplémentaire pendant l'initiation. Elle possède une activité de kinase qui phosphoryle le terminal carboxyle de l'ARN polymérase II qui causera un changement de conformation du complexe et stimulera ainsi l'initiation de la transcription. Après 60-70 nucléotides d'ARN, le TFIIIE et le TFIIH quittent le complexe et marquent ainsi le début de la phase d'élongation.

Elongation, terminaison et libération. TFIIF reste associé à l'ARN polymérase II pendant toute la phase de l'élongation. L'activité de l'ARN polymérase II est renforcée par des facteurs d'élongation. Une fois le transcrit est complété, la terminaison de la transcription se fait par des mécanismes qui sont encore mal compris. L'ARN polymérase II est déphosphorylée et recyclée.

Modifications post-transcriptionnelles chez les eucaryotes:

L'ARNm est modifié en général aux deux extrémités 5' et 3' par des enzymes spécifiques en plus de l'épissage qui coupe les introns des transcrits primaires.

Capping de 5'. Elle implique l'addition de m⁷G (7-méthyl guanosine) au nucléotide terminal de 5' par une liaison 5'-5' et l'addition de deux groupes méthyls aux deux premiers nucléotides de la chaîne d'ARN(fig). Ce capping est fait quand le transcrit est de 20-30 nucléotides de long. Le cap ou coiffe est essentiel pour l'association du ribosome à l'extrémité 5' de l'ARN pendant la phase initiale de la traduction. Il est aussi essentiel pour les réactions d'épissage suivantes. Il contribue aussi à la stabilité de l'ARNm en protégeant leur extrémité 5' des phosphatases et des nucléases.

L'addition d'une queue 3' polyA.

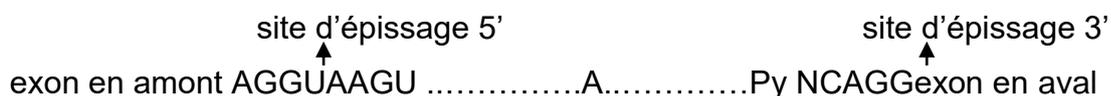
Chez la plupart des gènes eucaryotiques, on a une addition d'une queue polyA de 50 à 250 nucléotides à la pré-ARNm. On ne trouve pas cette queue sur toutes les ARNm eucaryotes. Par exemple, les ARNm des histones dans les cellules de mammifères ne possèdent pas la queue poly A. Elle n'est pas codée par l'ADN. En fait, il existe un site d'une séquence AAUAAA reconnu par une endonucléase qui clive le transcrit primaire pour signaler l'extrémité 3' de l'ARNm.

Cependant, cette séquence ne signale pas la fin de la transcription qui peut continuer pour des centaines de nucléotides après. Le clivage n'aura pas lieu si cette séquence ou un segment de 20 nucléotides du côté de 3' est délété. Quelques ARNm matures contiennent la séquence AAUAAA. Donc, cette séquence fait seulement partie d'un signal de clivage. Après le clivage par une endonucléase, une polymérase poly A va ajouter la queue polyA. L'ATP est le donneur de l'Adénine dans cette réaction.

Le rôle de cette queue poly A est une énigme. Le blocage de cette synthèse n'affecte pas le transport de l'ARNm du noyau vers le cytoplasme. On croit que sa présence est importante pour allonger la demi-vie de la molécule d'ARNm.

L'épissage de L'ARNm.

Les introns sont épissés de façon précise chez les eucaryotes. Les jonctions entre les introns et les exons ont été révélatrices. Ces séquences qui varient en complexité des levures aux mammifères ont un motif en commun. Ils commencent toujours par GU et finissent par AG. La séquence consensus chez les vertébrés est la suivante: 5'AGGUAAGU3' et à l'extrémité 3' la séquence consensus est de 10 pyrimidines (U ou C) suivi par n'importe quelle base et se terminant par AG. Les introns contiennent aussi un site de branchement de 20-30 nucléotides en amont du site d'épissage 3'. Chez la levure, le site de branchement est presque toujours UACUAAC, alors que chez les mammifères il est variable. La figure ci-dessous résume :



La jonction intron-exon 5' impliquent au moins 7 nucléotides La séquence consensus chez les vertébrés est la suivante : GUAAGUDU. Celle de l'extrémité 3' implique 10 nucléotides et un site de branchement de 20-30 nucléotides. Chez les mammifères, la séquence consensus du site de branchement est YNCURAY (Y représente une pyrimidine et R représente une purine et N une base quelconque.) Les introns ont une longueur de 50 –10000 nucléotides.

Les évènements qui amènent à l'épissage sont les suivants:

- a). clivage à la jonction 5' qui mène à la séparation de l'exon I du reste de l'ARNm qui contient l'intron I et l'exon 2.
- b). L'extrémité libre 5' de l'intron forme une boucle qui va ensuite joindre un nucléotide A qui fait partie du site de branchement. La structure qui apparaît est appelée le lariat.
- c). Clivage à la jonction 3' et ligation des deux exons.

d). Le lariat est ensuite rendu linéaire et dégradé par des enzymes spécifiques.

L'épissage se fait par des complexes appelés spliceosomes qui sont formés par le pré-ARNm associé à des particules petites ribonucléaires (snRNP) ou snurps. Ces ARN nucléaires sont associés à des protéines. Il existe 6 snRNA : U1-U6, chacun d'eux est associé à 6-10 protéines pour former les snRNP.

On trouve U4 et U6 dans le même snRNP alors que chacun des autres est associé à un seul snRNP. Un modèle pour l'assemblage d'un spliceosome et son rôle dans l'épissage est le suivant :

- a). U1 snRNP s'attache au site d'épissage 5' par un appariement.
- b). U2 snRNP s'attache au site d'embranchement
- c). Une particule préassemblée U4/U6/U5 joint le complexe par l'association de la particule avec les U1 et U2 snRNPs.
- d). U4 snRNP se dissocie du complexe

Il n'est pas clair encore comment le spliceosome enlève l'intron.

Modifications de l'ARNt eucaryotique

Chez les eucaryotes, les modifications de l'ARNt en ARNt mature impliquent : le clivage de la séquence leader, l'épissage pour enlever les introns, le remplacement du terminal 3'UU par CCA et la modification de plusieurs bases.

Au moins quelques gènes d'ARNt des eucaryotes contiennent des introns qui sont présents la plupart du temps dans la région de l'anticodon. L'épissage de cet intron se fait par une endonucléase spécifique et la ligation des fragments d'ARN qui restent se fait par une ARN ligase.

Les autres modifications de l'ARNt eucaryote impliquent :

- a). l'addition d'un 5' CCA3' à l'extrémité 3'.
- b). La modification chimique extensive de certains nucléotides à plusieurs endroits de la chaîne.

Le type et l'étendue de ces modifications varient d'un ARNt à un autre, mais elle implique typiquement la méthylation de bases spécifiques, la réduction de certains uridines pour donner par exemple la dihydrouridine ou le réarrangement de certains uridines pour produire des pseudouridines (Ψ) (fig). Ces modifications spécifiques, créées par des enzymes spécifiques donnent à chaque ARNt une configuration bi ou tri-dimensionnelle spécifique qui lui permet de s'attacher à un acide aminé spécifique et au ribosome.

Modifications post-transcriptionnelles de l'ARNr.

La plupart des eucaryotes qui ont été examinés ont un large nombre de copies des gènes pour chacun des espèces d'ARNr : 18S, 5.8S, 28S et 5S. On les trouve adjacents l'un à l'autre dans l'ordre 18S, 5.8S, et 28S, avec chaque groupe répété plusieurs fois pour former des unités répétées en tandem et appelées unités répétées rADN. Ils représentent les gènes modérément répétés. Chez la plupart des organismes, les gènes de 5S ARNr sont localisés à un site ou à des sites différents de ceux des autres gènes d'ARNr. Chez les humains, ils sont groupés en une seule location. Très peu d'organismes ont des introns dans les gènes d'ARNr (*Drosophila*, *Tetrahymena*). L'épissage de ces introns est très différent de celui qu'on trouve dans l'épissage des introns des ARNm et des ARNt.

Les inhibiteurs antibiotiques

L'actinomycine et la rifamycine sont deux antibiotiques qui inhibent la transcription de deux façons différentes. La rifamycine, un dérivé d'une souche de *Streptomyces*, et la rifampicine, un dérivé semi-synthétique, inhibent spécifiquement l'initiation de la transcription. La rifampicine bloque la formation de la première liaison phosphodiester dans la chaîne d'ARN. L'élongation de la chaîne n'est pas affectée. Le site d'action de la rifampicine est la sous-unité β de l'ARN polymérase. Quelques mutants (appelés rif-r) ont une chaîne β altérée et sont résistants à la rifampicine.

L'actinomycine D, un antibiotique contenant un polypeptide, qu'on obtient d'une souche différente de *Streptomyces*, inhibe la transcription en se liant fortement et spécifiquement à l'ADN double-brin et l'empêchant ainsi de devenir une matrice efficace pour la synthèse d'ARN. Cet antibiotique ne se lie jamais à l'ADN ou à l'ARN simple brin, à l'ARN double-brin ou aux hybrides d'ADN-ARN. Cette liaison est renforcée par la présence de résidus de guanine. L'actinomycine agit comme un agent intercalaire qui se glisse entre deux bases adjacentes dans l'ADN à cause de sa structure cyclique (anneau de phénoxazone).

Chapitre 6 : La synthèse des protéines.

Le code génétique

L'évidence que le code génétique est formé de 3 nucléotides est venu d'expériences faites par Francis Crick, Leslie Barnett, Sidney Brenner et R. Watts-Tobin au début des années 60. Les expériences ont utilisé le bactériophage T4 dont les mutants RII produisent des plaques claires sur *E. coli* B, alors que les souches sauvages produisent des plaques turbides. En plus, les mutants RII sont incapables de se reproduire sur *E. coli* K12.

Crick et ses collègues ont pris une souche r+ qui a été traité avec un mutant la proflavine qui cause une addition ou une délétion d'une paire de base dans l'ADN. Ils ont raisonné que si les mutants RII sont le résultat d'une addition ou une délétion, leur traitement avec de la proflavine peut causer une inversion de ces mutants en r+. Ce qui veut dire que si la mutation originale était une addition, elle peut être corrigée par une délétion et l'inverse est vrai. Ainsi, ils ont pu isoler un nombre de révertants r+ en cultivant une population de rII traités avec la proflavine sur une boîte de *E. coli* K12 (λ).

1). Un code formé s'une seule base peut spécifier seulement 4 types d'acides aminés parce qu'il y a 4 types de bases dans l'ADN. Si le code est formé de deux bases, il y aurait 16 codons (4^2), alors que si le code est formé de 3 bases, on aurait 64 codons (4^3). Comme on trouve 20 acides aminés dans la nature, il est évident que le code doit être formé de trois bases ou plus. Les expériences de génétique ont montré ensuite que le code est en fait formé de 3 bases.

2). Est-ce que le code est chevauchant ou non ? Dans un code non-chevauchant, chaque groupe de 3 bases dans une séquence ABCDEF spécifie seulement un acide aminé. ABC spécifie un acide aminé et DEF spécifie un autre, alors que dans le code chevauchant, ABC spécifie le premier acide aminé BCD spécifie un autre et CDE spécifie un autre.

Ces alternatives ont été étudiées par des mutations. Supposons que la base C est mutée en C'. Les études ont montré qu'un seul acide aminé est altéré ce qui veut dire que le code est non-chevauchant. Si le code était chevauchant, les acides aminés 1, 2 et 3 seraient altérés.

3). Le code doit être dégénéré car quelques acides aminés sont codés par plus d'un triplet.

La dégénérescence du code est résumée par l'hypothèse de wobble:

- 1). Les deux premières bases du codon s'apparient de façon précise. Si deux codons diffèrent dans les deux premières bases, ils sont reconnus par des ARNt différents. Par exemple, UUA et CUA codent pour la leucine mais sont lus par des ARNt différents.
- 2). La première base de l'anticodon détermine si une molécule d'ARNt particulière va lire un, deux ou trois types de codons. Le tableau ci-dessous résume l'appariement de chaque base. Donc, une partie de la dégénérescence du code vient de l'imprécision ou wobble dans l'appariement de la troisième base du codon.

Le déchiffrement du code génétique.

Cela a été l'œuvre de Nirenberg qui a découvert que l'addition de poly-U à un système acellulaire a mené à la synthèse de la polyphénylalanine. Poly-U a probablement servi comme ARNm. Le système acellulaire a été obtenu de la façon suivante : les cellules bactériennes ont été broyées délicatement. Les débris cellulaires (les membranes cellulaires et les parois) sont ensuite enlevés par centrifugation. L'extrait acellulaire résultant contient de l'ADN, de l'ARNm, de l'ARNt, des ribosomes, des enzymes et des constituants cellulaires. Les protéines ont été synthétisées par cet extrait acellulaire quand on ajoute de l'ATP, du GTP et des acides aminés. Au moins l'un des acides aminés est marqué radioactivement de façon à le détecter lors de son incorporation dans une protéine. Ce mélange a été incubé à 37° pour une heure. On ajoute ensuite de la TCA pour arrêter les réactions et précipiter les protéines, laissant les acides aminés libres dans le surnageant. Le précipité a été lavé et sa radioactivité mesurée. Une caractéristique importante de ce système est que la synthèse des protéines peut être arrêtée par l'addition d'une désoxyribonucléase, qui détruit la matrice pour la synthèse d'ARNm. Cette synthèse recommencera par l'addition d'une fraction crue d'ARNm. Un autre composé critique de cette expérience est la présence d'un polyribonucléotide synthétique, qui a été synthétisé par la poly-nucléotide phosphorylase, une enzyme découverte en 1955 par Manago et Ochoa. La réaction de synthèse est la suivante :



Une différence critique entre cette enzyme et l'ARN polymérase est qu'elle n'a pas besoin d'une matrice et qu'elle utilise des ribonucléosides diphosphates au lieu de triphosphates. L'ARN synthétisée par cette enzyme aura une composition dictée seulement par les rapports des ribonucléotides dans le milieu d'incubation. Par exemple, des copolymères

de deux ribonucléotides, U et A auraient des séquences arbitraires en incubant l'enzyme avec ADP et UDP. Les séquences peuvent être les suivantes : AAA, AAU, AUA, UAA, AUU, UAU, UUA, UUU.

Les expériences de Khorana ont complété celles de Nirenberg pour déchiffrer le code génétique. Il a réussi à synthétiser par une combinaison de techniques organo-chimiques et enzymatiques une variété de copolymères de deux, trois et quatre types de bases. Considérons par exemple la stratégie de synthèse de poly(GUA). Ce copolymère a la séquence suivante :

GUAGUAGUAGUAGUA....

En premier, Khorana a synthétisé par des méthodes organo-chimiques deux désoxyribonucléotides complémentaires, chacun formé de neuf résidus : d(TAC)₃ et d(GTA)₃. Des duplexes partiellement doubles sont ensuite formés en mélangeant ces deux oligonucléotides et ensuite utilisés comme des matrices pour la synthèse de longues chaînes d'ADN par des ADN polymérases. Une des chaînes est ensuite sélectionnée pour la synthèse de polyribonucléotides par l'ARN polymérase. La sélection se fait en ajoutant seulement les ribonucléotides nécessaires pour la transcription d'une des deux chaînes. Par exemple, quand GTP, UTP et ATP sont ajoutés au mélange d'incubation, le polyribonucléotide poly(GUA) est synthétisé de la matrice poly d(TAC). L'autre brin n'est pas synthétisé car un des substrats nécessaires, CTP, était absent.

La molécule d'ARNt

La première séquence de bases d'un ARNt, celle de l'ARNt de l'alanine de levure a été déterminée par Holley en 1965. Plus tard, les séquences d'une centaine d'autres molécules d'ARNt ont été trouvées. La caractéristique commune des ARNt est qu'ils peuvent tous avoir une structure en trèfle dans laquelle la moitié des résidus sont appariés. En fait, tous les ARNt connus ont plusieurs caractéristiques communes du fait qu'ils doivent tous interagir avec le ribosome et l'ARNm:

- 1). Ce sont des chaînes simple brin contenant 73-93 ribonucléotides.
- 2). Ils contiennent plusieurs bases étranges, typiquement entre 7-15 par molécule. Plusieurs sont méthylées, ce qui les rend hydrophobes, une caractéristique très importante pour leur interaction avec les protéines ribosomales et les synthétases. La méthylation empêche aussi leur appariement à d'autres bases pour qu'ils restent disponibles à d'autres interactions.
- 3). L'extrémité 5' de l'ARNt est phosphorylée, en général un pG.

4. L'extrémité 3' est CCA auquel l'acide aminé activé est attaché.

5. Presque 50% des nucléotides de l'ARNt sont en forme double-brin. 5 groupes de bases ne sont pas appariés : la région terminale 3'CCA, la boucle TΨC qui a acquis son nom de la séquence ribothymine-pseudouracil-cytosine ; le bras extra, qui contient un nombre variable de résidus ; la boucle DHU, qui contient plusieurs résidus de dihydrouracil ; et la boucle anticodon.

6. La boucle anticodon contient 7 bases avec la séquence suivante: 5' Pyrimidine-Pyrimidine-X-Y-Z- purine modifiée-base variable-3'.

Les ribosomes :

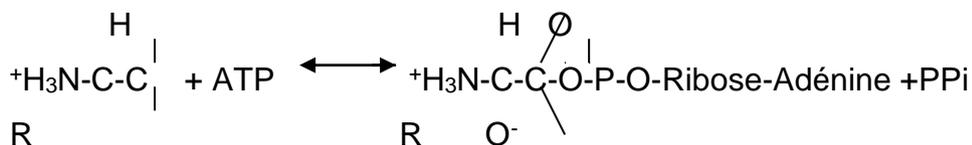
Un ribosome de E. coli est une particule de ribonucléoprotéine qui a un coefficient de sédimentation de 70S. Il peut être dissocié en deux sous-unités : une large de 50S et une petite de 30S. Ces sous-unités sont formées de protéines et d'ARN. La sous-unité 30S contient 21 protéines (appelées S1 à S21) et une molécule d'ARN de 16S.

La sous-unité 50S contient 34 protéines (L1-L34) et deux molécules d'ARN, une de 23S et une de 5S (fig).

Schéma

Une protéine est synthétisée dans la direction amine à carboxyle par l'addition séquentielle des acides aminés au terminal carboxyle.

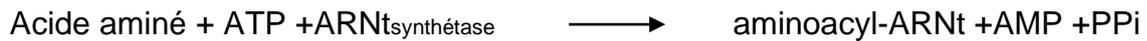
La synthèse des protéines se fait en trois étapes: l'initiation, l'élongation et la terminaison. L'initiation : implique l'association de l'ARNt au signal d'initiation de l'ARNm et au ribosome. Les acides aminés sont activés par des synthétases spécifiques, appelées aminoacyl-ARNt synthétases par les deux chercheurs qui les ont découvertes en 1957, Zamecnik et Hoagland. La première étape de cette activation est la formation d'un aminoacyl-adénylate (ou aminoacyl-AMP) d'un acide aminé et d'un ATP.



La deuxième étape est le transfert du groupe aminoacyl à l'ARNt pour former l' aminoacyl-ARNt. Pour quelques ARNt, ce groupe est transféré au 2'-OH du ribose et pour d'autres au 3'-OH du ribose. La réaction est la suivante :



La réaction totale de l'activation et de transfert se résume à :



Les deux réactions sont catalysées par la même synthétase. Au moins une synthétase existe pour chaque acide aminé. Ces enzymes diffèrent l'une de l'autre par leur taille et leur structure.

L'initiation.

La traduction ne commence pas immédiatement à l'extrémité 5' de l'ARNm. En fait, le premier codon traduit est d'à peu près plus de 25 nucléotides de l'extrémité 5' de l'ARNm. En plus, plusieurs ARNm sont polycistroniques. Chacune des protéines codées par l'ARNm polycistronique a un signal d'initiation et un signal de terminaison sur l'ARNm. La première indication du fait que l'ARNm contient des signaux d'initiation est le fait que la moitié des protéines chez *E. coli* ont un terminal amine qui commence par la méthionine, qui n'est pas commun dans le reste de la protéine. En plus, le terminal amine des protéines de *E. coli* est normalement modifié, ce qui suggère qu'un dérivé de la méthionine participe dans l'initiation. Il a été constaté que la synthèse des protéines chez les bactéries commence par la formylméthionine (fMet). Un ARNt spécifique ramène la formylméthionine au ribosome pour initier la synthèse des protéines. Cet ARNt est différent de celui qui place la méthionine interne. Une enzyme spécifique met un groupe formyl au groupe amine de la méthionine attaché à l'ARNt (fig).

Le codon d'initiation de l'ARNm est AUG ou GUG. Comment cette région est-elle distinguée d'une région intérieure qui contient un AUG? L'isolement des séquences d'initiation d'un certain nombre d'ARNm a pu répondre à cette question. Ceci a été fait par la digestion de complexes ribosome-ARNm par une ribonucléase pancréatique. Dans chaque cas, une séquence de 30 nucléotides a été protégée de la digestion. Cette séquence contient en plus du codon AUG ou GUG, une séquence appelée Shine-Dalgarno, riche en purine et centrée sur 10 nucléotides à l'extrémité 5' du codon d'initiation. Elle est complémentaire à l'extrémité 3' de l'ARNr 16S du ribosome. Ceci a été démontré par l'isolement d'un complexe de l'extrémité 3' de 16S et du ribosome après la digestion avec une ribonucléase. L'appariement entre l'ARNm et 16S ARNr implique 3 à 9 nucléotides suivant la protéine.

L'initiation se fait en trois étapes: 1) la sous-unité 30S s'associe à deux facteurs d'initiation IF1 et IF3. Ce dernier a pour rôle d'empêcher l'association entre 30S et 50S avant l'association de l'ARNm à la sous-unité 30S. Les ribosomes bactériens contiennent trois sites, le site A ou aminoacyl, le site P ou peptidyl, et le site E ou exit. L'ARNt-fmet est le seul amino-acyl ARNt qui s'attache au site P. Tous les autres ARNt se lient au site A. Le site E est celui sur lequel s'attachent des ARNt non chargés. Le facteur d'initiation IF-1 s'attache au site A et empêche l'attachement d'un ARNt à ce site pendant la phase d'initiation. 2) le complexe qui consiste en la sous-unité 30S, IF-3 et l'ARNm est joint par IF-2 GTP et fmet-ARNt^{fmet}. 3). Ce large complexe se combine avec la sous-unité 50S et simultanément, le GTP attaché à l'IF-2 est hydrolysé en GDP et Pi, qui sont libérés du complexe. Tous les facteurs d'initiation sont ensuite libérés du ribosome. L'achèvement de ces étapes nous donne un complexe appelé le complexe d'initiation.

Initiation chez les cellules eucaryotes : La traduction chez les eucaryotes est en général similaire à celle des procaryotes mais avec la plupart des différences significatives dans l'initiation. L'ARNm eucaryotique forme un complexe avec le ribosome et un certain nombre de protéines spécifiques. Plusieurs de ces protéines sont supposées lier les deux extrémités du message. A l'extrémité 3', l'ARNm est lié par une protéine appelée PAB ou poly A binding protein. Les cellules eucaryotes contiennent au moins neuf facteurs d'initiation.

L'élongation.

Dans le premier stade de l'élongation, l'aminoacyl-ARNt approprié est lié à un complexe de EF-Tu et GTP. Le complexe formé se lie au site A du complexe d'initiation 70S. Le GTP est hydrolysé et un complexe EF-Tu-GDP est libéré du ribosome 70S.

Dans le deuxième stade de l'élongation, une liaison peptidique, catalysée par la peptidyl-transférase se forme entre les deux acides aminés liés par leur ARNt au site A et au site P du ribosome. Cette dernière activité enzymatique peut faire partie de l'ARNr 23S du ribosome. Cette réaction produit un dipeptidyl-ARNt dans le site A, laissant le site P avec un ARNt^{met} non chargé.

Dans le troisième stade de l'élongation, le ribosome se déplace d'un codon vers l'extrémité 3' de l'ARNm. L'ARNt chargé par le dipeptidyl se retrouve donc dans le site P. L'ARNt non chargé se déplace vers le site E et ensuite dans le cytosol. Le mouvement du ribosome est appelé une translocation et requiert une translocase (EF-G) qui est aussi une GTPase. Le mouvement est rendu possible par un changement de la conformation

tri-dimensionnelle du ribosome entier. La structure de EF-G est similaire à celle du complexe EF-Tu/ARNt, suggérant que EF-G peut s'associer au site A et déplacer l'ARNt-peptidyl.

L'élongation chez les eucaryotes est très similaire à celle des prokaryotes. Cependant, les ribosomes eucaryotes ne possèdent pas de site E. Les ARNt non chargés quittent directement le site P.

Terminaison.

C'est le dernier stade de la synthèse des protéines. Le signal pour la terminaison est l'un des codons de terminaison dans l'ARNm : UAA, UAG et UGA. Chez les bactéries, trois facteurs de libération, RF1, RF2 et RF3 contribuent à 1) l'hydrolyse de la liaison terminale entre le peptidyl et l'ARNt, 2) la libération du polypeptide libre et le dernier ARNt non chargé du site P et 3) la dissociation du ribosome 70S en 50S et 30S. RF1 reconnaît les codons de terminaison UAG et UAA, et le RF2 reconnaît les codons UGA et UAA. Ces deux facteurs de terminaison se lient au codon de terminaison et induit la peptidyl transférase à transférer le polypeptide à une molécule d'eau au lieu d'un autre acide aminé. Ces facteurs de terminaison ont une structure qui est semblable à celle de l'ARNt. La fonction spécifique du facteur de terminaison RF-3 n'est pas encore bien établie, mais on croit qu'il libère la sous-unité ribosomale. Chez les eukaryotes, un seul facteur de terminaison appelé eRF reconnaît tous les codons de terminaison.

Quelques différences et similarités entre la synthèse des protéines chez les prokaryotes et chez les eukaryotes :

1. Les ribosomes : Ils sont plus larges. Ils consistent d'une large sous-unité 60S et une petite sous-unité 40S, pour former une particule de 80S. La sous-unité 40S contient une ARN 18S homologue à l'ARNr procaryotique 16S. La sous-unité 60S contient 3 ARNs : les ARNs 5S et 28S sont homologues à ceux des prokaryotes mais l'ARN 5,8S est propre aux eukaryotes.
2. L'ARNt initiateur. Chez les eukaryotes, l'acide aminé initiateur est la méthionine au lieu de la N-formyl méthionine. Cependant un ARNt spécifique participe dans l'initiation. Celui-ci est appelé Met-ARNtf ou Met-ARNti (formylé in vitro ou initiation).
3. Le signal d'initiation. Le codon d'initiation chez les eukaryotes est toujours AUG. Cependant, les eukaryotes ne possèdent pas une séquence riche en purine à l'extrémité 5' comme chez les prokaryotes. Au lieu de cela, les ribosomes 40S s'attachent au cap 5' de l'ARNm et cherche le AUG initiateur en se déplaçant vers

l'extrémité 3'. Ceci est accompagné par l'hydrolyse de l'ATP. L'ARNm eukaryotique a un seul signal d'initiation (monocistronique) alors que l'ARNm prokaryotique peut en avoir plusieurs.

4. Les complexes d'initiation. Les eukaryotes contiennent plus de facteurs d'initiation que les prokaryotes (au nombre de neuf).
5. Les facteurs d'élongation et de terminaison : les facteurs d'élongation eucaryotiques EF1 α , et EF1 $\beta\gamma$ remplacent les facteurs procaryotiques EF-Tu et EF-Ts. EF2 est l'analogue eucaryotique de EF-G qui catalyse la translocation. La terminaison chez les eucaryotes est faite par un seul facteur de libération, eRF, une GTP-ase.

Les inhibiteurs antibiotiques de la synthèse des protéines.

Les antibiotiques sont des inhibiteurs hautement spécifiques de la synthèse des protéines. La streptomycine, la tetracycline, le chloramphenicol et l'érythromycine sont des agents antibactériens qui agissent en bloquant des phases différentes de la synthèse des protéines. La streptomycine, un trisaccharide hautement basique interfère avec l'association de l'ARNt-formylmethionine aux ribosomes et empêche l'initiation de la synthèse des protéines. La streptomycine provoque la mauvaise lecture de l'ARNm. Un polyU est lu par exemple comme l'isoleucine en plus de la phenylalanine.

La puromycine est l'analogue de la portion terminale aminoacyl-adénosine de l' aminoacyl-ARNt. Il s'associe au site A du ribosome et inhibe l'entrée de l' aminoacyl-ARNt. En plus, la puromycine contient un α amino groupe, qui, comme celui qui se trouve sur l' aminoacyl-ARNt forme une liaison peptide avec le groupe carboxyl de la chaîne peptidique croissante dans une réaction catalysée par la peptidyl transférase. Ce peptidyl puromycine se détache ensuite du ribosome.

La compartimentalisation des protéines

Cette compartimentalisation des protéines eucaryotiques est sous contrôle génétique. Les Protéines à sécréter sont empaquetés dans des vésicules sécrétoires qui sont formés dans l'appareil de Golgi. L'hypothèse du signal, énoncée par Bobbel et co en 1975, qui énonce qu'une protéine qui doit être sécrétée, doit contenir une séquence signal qui est un terminal amine hydrophobe qui va ancrer cette protéine dans la membrane.

Chapitre 7 : Régulation de l'expression des gènes chez les procaryotes

Le contrôle de l'expression des gènes est crucial pour tous les organismes. Chez les procaryotes, la régulation de l'expression des gènes permet aux cellules bactériennes de répondre rapidement à ses besoins nutritifs. Chez les cellules supérieures, il y a une différenciation de cellules telle que seulement certains gènes sont exprimés et d'autres sont réprimés de façon permanente.

L'opéron lactose chez *E. coli* :

E. coli peut utiliser le lactose comme seule source de carbone. Une enzyme essentielle pour le métabolisme du lactose est la β galactosidase, qui hydrolyse le lactose en glucose et galactose. Une cellule de *E. coli* qui croît sur un milieu qui contient le lactose contient plusieurs milliers de molécules de β galactosidase. D'autre part, si les cellules de *E. coli* sont cultivées sur un milieu qui contient une autre source de carbone que le lactose, comme le glucose ou le glycérol, le nombre de molécules de β galactosidase par cellule est moins que 10. Cela veut dire que la présence du lactose dans le milieu induit l'expression du gène de la β galactosidase pour produire un grand nombre de molécules de cette enzyme. Deux autres enzymes sont synthétisées en même temps et dans les mêmes proportions que la β galactosidase : la galactoside perméase et la thiogalactoside transacétylase. La perméase a pour rôle le transport du lactose à travers la membrane bactérienne. La transacétylase n'est pas essentielle pour le métabolisme du lactose ; son rôle physiologique n'est pas certain. *In vitro*, elle catalyse le transfert d'un groupe acétyl d'un acétyl-CoA à un groupe hydroxyl C-6 d'un thiogalactoside. Chez *E. coli*, l'inducteur physiologique est l'allolactose.

Les études de mutants ont montré que la β galactosidase, la perméase et la transacétylase sont codés par trois gènes contigus, *z*, *y* et *a* respectivement (fig). Les mutants défectueux dans l'une de ces protéines ont été isolés. Le plus intéressant est un mutant où les trois protéines sont affectées. Ce mutant constitutif synthétise de larges quantités de β -galactosidase, de perméase et de transacétylase en la présence ou en l'absence de l'inducteur. François Jacob et Jacques Monod ont déduit que le taux de synthèse de ces trois protéines est gouverné par un élément commun. Le gène pour cet élément régulateur commun est appelé *i*. La bactérie sauvage inductible a le génotype $i^+z^+y^+a^+$, alors que les mutants constitutifs ont le génotype $i^-z^+y^+a^+$. Pour expliquer cela, les deux chercheurs ont énoncé l'hypothèse que le gène i^+ détermine la synthèse d'une substance cytoplasmique appelé un répresseur, qui n'existe pas chez les mutants i^- .

Cette idée a été testée sur des mérozygotes ou des diploïdes partiels. Un groupe de mérozygotes était $i^+z^-/F_i^+z^+$ où F_i^+ est facteur sexuel introduit par conjugaison qui contient les deux gènes i^+ et z^+ . Un autre groupe était $i^-z^+/F_i^+z^-$. Ces mérozygotes sont-ils inductibles ou constitutifs ?

Les résultats expérimentaux étaient clairs. Ils sont inductibles, signifiant que le répresseur est diffusible et spécifié par le gène i . Ce répresseur diffusible est un exemple d'un facteur trans, codé par un locus sur l'ADN différent de celui qui contient la cible.

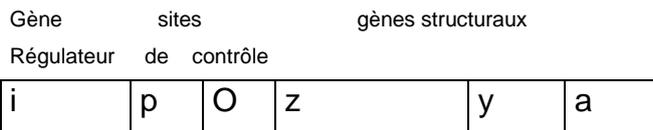


Fig : carte de l'opéron lactose et son gène régulateur

Le répresseur de l'opéron lac a été purifié. C'est un tétramère de sous-unités identiques de 47 Kd, chacune contenant un site d'attachement à l'inducteur. Il s'associe très rapidement et fortement à l'opérateur. L'opérateur a été isolé par Gilbert et ses associés qui ont fait subir à l'ADN d'un phage contenant l'opéron lac une sonication qui a coupé l'ADN en des fragments de 1000 paires de bases de longueur. Le répresseur a été ensuite ajouté au mélange et filtré sur des membranes de nitrocellulose. Les complexes ADN-répresseur se sont attachés au filtre alors que le reste a passé à travers. L'ADN a été ensuite libéré en ajoutant IPTG. Ces fragments libérés ont été digérés avec l'ADNase pancréatique en présence du répresseur, qui a protégé l'opérateur de la digestion.

La séquence de base de l'opérateur a été analysée ; elle est constituée de 28 paires de bases avec une symétrie rotationnelle (fig). Donc, la symétrie du répresseur correspond à celle de l'opérateur.



fig : séquence de l'opéron lac

E. coli qui est cultivée dans un milieu ne contenant que le glucose comme source d'énergie, a des niveaux très bas d'enzymes cataboliques comme la β galactosidase, la galactokinase, l'arabinose isomérase, et la tryptophanase. La base moléculaire de l'inhibition de ces enzymes par le glucose a été élucidée. L'observation que le glucose baisse la concentration de l'AMP cyclique chez *E. coli* a été révélatrice. On a trouvé alors que l'AMP cyclique exogène peut lever la répression exercée par le glucose. Les études

biochimiques et génétiques qui ont suivi ont montré que l'AMP cyclique stimule l'initiation de la transcription de plusieurs opérons inductibles.

Dans les cellules bactériennes, l'AMPc s'associe à une protéine appelée CAP (catabolite geneactivatingprotein). Cette protéine est un dimère de deux sous-unités identiques, chacune d'elles contient deux domaines, un domaine d'attachement à l'ADN et un domaine d'attachement à l'AMPc. Le complexe CAP-AMPc stimule la transcription en s'attachant à certains sites promoteurs. Dans l'opéron lac, le CAP s'attache à côté du site d'attachement de l'ARN polymérase. Plus spécifiquement, le CAP protège les nucléotides de -87 à -49 de la digestion, alors que l'ARN polymérase protège -48 à +5, deux sites non-chevauchants. Le CAP exhibe une symétrie rotationnelle qui correspond à celle de son site d'attachement à l'ADN.

Le CAP stimule l'initiation de la synthèse du lac ARNm par un facteur de 50. L'arrangement contigu et non-chevauchant des sites d'attachement de l'ARN polymérase et de celui de CAP a suggéré que l'attachement de CAP crée un site d'attachement additionnel de l'ARN polymérase. En fait, l'attachement de l'ARN polymérase au promoteur est favorisé par son contact énergétique avec le CAP. En contraste, le répresseur du lac s'attache aux nucléotides -3 à +21, qui chevauche significativement au site de l'ARN polymérase. Donc, le répresseur inhibe de façon stérique l'attachement de l'ARN polymérase. Tous les opérons inductibles par le complexe CAP-AMPc contiennent la séquence TGTGA en amont du site promoteur. Une autre caractéristique commune est que les séquences -10 et -35 sont très différentes des séquences consensus des promoteurs forts. L'évolution a probablement affaibli ces promoteurs pour rendre leur opéron dépendant sur une protéine qui facilite l'initiation efficace de la transcription. Donc, les opérons cataboliques inductibles sont sous un contrôle double. Une expression élevée de la transcription requiert la présence simultanée de l'AMPc et l'inducteur.

L'opéron tryptophane :

L'opéron tryptophane (trp) est constitué de cinq gènes qui codent pour les enzymes qui convertissent le chorismate en tryptophane (fig). Un ARNm polycistronique est produit en quatre minutes et ensuite dégradé rapidement. Cette durée de vie courte permet à la bactérie de répondre rapidement à ses besoins en tryptophane. La traduction de cet ARNm qui suit la transcription va produire de façon séquentielle et en quantités équimolaires les cinq enzymes. En fait, quand la quantité de tryptophane diminue dans la cellule, la synthèse de ces cinq enzymes peut augmenter de 700 fois.

a

a



Fig: diagramme de l'opéron trp montrant les sites de contrôle du promoteur (p), l'opérateur(o), et l'atténuateur(a), les gènes pour la séquence tête (L), et les enzymes de la synthèse du tryptophane (E, D, C, B, A).

D'un autre côté, quand le niveau de tryptophane est élevé, il s'attache à un répresseur causant un changement de conformation de ce dernier, qui va lui permettre de s'associer à l'opérateur et inhiber ainsi l'expression de cet opéron. Le répresseur est un homodimère codé par le gène trpR , qui est loin de l'opéron trp. Le répresseur seul ne peut pas s'attacher à l'opérateur. Seul le complexe répresseur-trp peut s'attacher à ce site, montrant ainsi que le trp est un corépresseur.

La cible du complexe est un ADN possédant une symétrie rotationnelle (fig). Ce site est chevauchant avec celui du promoteur, pour empêcher l'association de l'ARN polymérase à ce dernier.

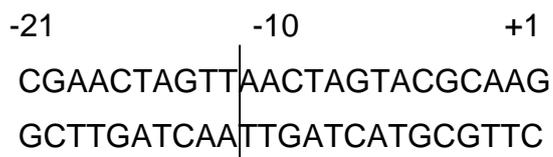


Fig : séquence de bases de l'opérateur trp. La symétrie rotationnelle est montrée par la ligne verticale. La paire de base marquée +1 est le début de la transcription de l'opéron.

Un autre processus, appelé atténuation, contrôle l'expression de l'opéron trp. En effet, quand le trp est abondant dans le milieu, les transcrits d'ARN sont courts de 140p.b. seulement qui se sont terminés au site d'atténuation a dans la région trpL ; cette dernière contient une séquence de terminaison qui peut former une épingle de cheveux. Elle contient aussi deux codons adjacents de trp qui jouent un rôle important dans l'atténuation.

Quatre régions dans l'ARNm peuvent former des structures secondaires avec des appariements complémentaires L'appariement des régions 1 et 2 peut donner un signal de pause et l'appariement de 3 et 4 est un signal de terminaison de la transcription.

L'opéron arabinose

C'est un mécanisme de régulation chez *E. coli* plus complexe que les autres.

La protéine Ara C exerce un contrôle à la fois positif et négatif. Elle possède deux conformations, l'une répressive qui se transforme en la forme active après la liaison avec une molécule signal, Chacune de ces deux formes a un site d'attachement sur l'ADN.

Cette protéine réprime aussi la transcription de son gène, un mécanisme d'auto-régulation.

Cet opéron est aussi sous le contrôle de séquences régulatrices distantes de son promoteur. Ce contrôle se fait par le pliement de l'ADN, sous l'effet d'interactions protéines –protéines et protéines –ADN, semblables à certains mécanismes régulateurs chez les eucaryotes.

E. coli peut utiliser l'arabinose comme source de carbone en le convertissant d'abord en xylulose 5-phosphate, un intermédiaire dans le cycle du pentose. Cette conversion requiert la présence de trois enzymes, codés par trois gènes : *araB*, *araA* et *araD*. Ces trois gènes se trouvent dans l'opéron arabinose, en plus d'un site régulateur avec deux opérateurs, *AraO1* et *AraO2*, un autre site de liaison de la protéine AraC appelé *araI* (I pour inducteur) et un promoteur adjacent à *araI* appelé *PBAD*. L'opérateur *AraO2* a un seul site de liaison à la protéine AraC. L'opérateur *araO1* et *araI* ont chacun deux sites de liaison à cette protéine dans la même orientation. Le gène *araC* est transcrit sous le contrôle de son propre promoteur mais dans l'orientation opposée que les gènes *araB*, *A* et *D*. Un site de liaison de CRP est adjacent au promoteur *PBAD*, et a un mécanisme d'action similaire au CRP de l'opéron *lac*.

Le rôle de la protéine AraC dans le contrôle de l'opéron arabinose est complexe.

Premièrement, quand sa concentration excède 40 copies par cellule, il réprime sa propre synthèse en se liant à l'opérateur *araO1*.

Deuxièmement, elle exerce un contrôle positif et négatif sur les gènes *araBAD*, en se liant à *araO2* et *araI*.

Le mécanisme de régulation est résumé de la façon suivante :

1. Glucose est abondant et arabinose est absent. Sous ces conditions, la protéine dimérique s'attache à *araO1* ainsi qu'à la moitié de *araI*, formant une boucle de 210 p.b., et réprimant la transcription des gènes *araBAD*.
2. Le glucose est présent seulement à des niveaux très bas, et l'arabinose est disponible. Sous ces conditions, CRP-cAMP devient abondant, et occupe le site CRP adjacent à *araI*. Arabinose se lie aussi à AraC et change sa conformation. La boucle

d'ADN formée par l'homodimère araC s'ouvre. AraC s'associe aux deux demi-sites de aral et devient un activateur, agissant en même temps que CRP-AMPc pour activer les gènes araBAD.

Quand l'arabinose et le glucose sont tous les deux abondants, ou quand ils sont tous les deux absents, l'opéron arabinose reste réprimée. Le mécanisme détaillé de cette répression n'est pas encore bien connu. La réponse de cet opéron est très rapide.

Le bactériophage lambda :

Le premier système de recombinaison dirigée étudié in vitro est celui du bactériophage lambda. Quand l'ADN du phage lambda entre une cellule de E. coli, une série d'évènements régulateurs dirigent la cellule vers un de deux destinées : ou bien l'ADN du phage lambda se réplique pour produire plus de bactériophages (phase lytique), ou bien il s'intègre dans le chromosome de l'hôte, se répliquant en même temps que le chromosome de l'hôte pour plusieurs générations. L'intégration est accomplie par une recombinase (intégrase) codée par le génome de lambda, qui agit à des sites de recombinaison appelés attP et attB , respectivement.

La décision se fait par un « switch génétique » qui implique une compétition entre les produits du gène Ci (répresseur) et le gène cro (protéine Cro). Si le répresseur domine, la phase lysogène suivra. Si la protéine Cro domine, c'est la phase lytique suivra.

Différents gènes sont exprimés suivant à ce que la phase lysogène ou la phase lytique est suivie. Premièrement, le chromosome linéaire devient circulaire avec l'appariement de sites cos qui représentent des extrémités cohésives de 12 p.b. et l'action de la ligase qui les lient de façon covalente ; puisqu'aucun répresseur n'a été formé pour réprimer les gènes du cycle lytic, la transcription commence aux promoteurs PL et PR PL étant le promoteur pour la transcription à gauche de l'opéron early, et PR est le promoteur pour la transcription à droite de l'opéron early. Le promoteur PR est sur un autre brin que PL, ainsi les deux promoteurs ont des orientations opposées,

De PR, le premier gène à être transcrit est le cro, dont le produit est la protéine Cro. Cette protéine joue un rôle dans le « switch génétique » vers la phase lytique. De PL, le premier gène à être transcrit est le gène N. La protéine N est un antiterminateur qui permet à la synthèse de l'ARN à passer par certains sites de terminaisons. Ce phénomène est appelé antitermination ou transcription à travers. Nprotéine va permettre à l'ARN polymérase de passer à gauche du gène N et à droite du gène cro, incluant tous les gènes précoces. Un des gènes transcrits sous l'influence de la protéine N est le gène cII . La protéine cII active les gènes suivants : cl code le répresseur de lambda, O et P codent

deux protéines de réplication de l'ADN, et Q (qui code une protéine qui active les gènes tardifs pour la lyse et les protéines du phage lambda). La protéine Q fonctionne comme une protéine d'anti-terminaison, permettant aux gènes tardifs de se transcrire. Cependant, seulement quand le switch est établi vers la phase lytique et que la transcription continue de PR pour un temps suffisant que la protéine Q s'accumule pour fonctionner efficacement.

La phase lysogène.

Après la transcription early, ou la phase lytique ou la phase lysogène est suivie. Le switch pour la phase lysogène se fait de la façon suivante :

La protéine cII stabilisée par la protéine cIII active la transcription de du gène cI localisé entre le promoteur PL et PR d'un promoteur appelé PRE (promoter for repressor establishment), localisé à la droite de cI-cro région. Le produit du gène cI, le répresseur, bloque l'expression des gènes nécessaires pour la réplication du chromosome, l'assemblage de la progéniture du phage, et la lyse de la cellule, maintenant la phase lysogène. Le répresseur de lambda consiste de deux domaines, le domaine amine et le domaine carboxyl. Le répresseur est un dimère formé par le contact entre les deux domaines carboxyls, quand il y a assez de répresseur dans la cellule, il s'associe à deux régions régionales opératrices, OL et OR, dont les séquences chevauchent avec celles des promoteurs, ce qui empêche la transcription précoce contrôlée par PL et PR. La transcription des gènes N et cro est bloquée et puisque les deux protéines codées par ces deux gènes sont instables, les niveaux de ces deux protéines dans la cellule diminuent dramatiquement. L'association du répresseur bloque aussi la transcription des gènes O, P et Q contrôlés par la protéine N. En plus, un répresseur attaché à O stimule la synthèse de plus de répresseur d'un promoteur différent, le PRM, pour maintenir la concentration du répresseur dans la cellule.

Les opérateurs OL et OR chacun a trois sites d'association au répresseur, désignés par OL1, OL2 et OL3 et OR1, OR2 et OR3 respectivement. Le répresseur ne s'attache pas de la même force à ces sites. La force d'association du répresseur à ces sites est dans l'ordre suivant : OL1>OL2>OL3. Pour le site d'opérateur OR, la force relative d'association est dans l'ordre suivant : OR1>OR2>OR3.

La régulation de l'expression des gènes chez les eucaryotes :

Il existe plusieurs niveaux de contrôle chez les eucaryotes.

1. Le niveau des chromosomes
2. Le niveau transcriptionnel
3. Le niveau post-transcriptionnel
4. Le niveau traductionnel
5. le niveau post-traductionnel