

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de RELIZANE
Faculté des Sciences et de la Technologie
Département des Sciences Biologiques



POLYCOPIÉ DE COURS

Destiné aux étudiants de la 3ème année Licence Biochimie

Intitulé

Biologie moléculaire

Elaboré par :

Dr. BELGUENDOZ Amina

Année universitaire : 2022/2023

Figure 1.1 Diagramme schématique de l'expérience de Griffith qui démontre la transformation bactérienne.....	2
Figure 1.2 Diagramme schématique de l'expérience d'Avery, MacLeod et McCarty qui démontre que l'ADN est le principe de transformation.....	3
Figure 1.3 Diagramme schématique de l'expérience Hershey-Chase qui démontre que l'ADN dirige la reproduction du phage T2 dans les cellules <i>E. coli</i> infectées.	5
Figure 1.4 Diagramme schématique d'un processus typique de l'application du clonage de gènes.....	6
Figure 1.5 (A) Structure chimique des bases azotées pyrimidines et purines dans l'ADN et l'ARN. (B) Structure chimique du ribose et du 2-désoxyribose que l'on trouve respectivement dans l'ARN et l'ADN.....	8
Figure 1.6 Formation d'un nucléotide par addition de groupe(s) phosphate(s) à un nucléoside.....	9
Figure 1.7 Formation d'ADP et d'ATP par addition successive de groupements phosphate via une liaison anhydride phosphorique.	10
Figure 1.8 Exemple de trois nucléotides liés entre eux par des liaisons phosphodiester entre les groupes 5'- et 3'-hydroxyle des sucres.	11
Figure 1.9 Deux modèles de structure de l'ADN.....	12
Figure 1.10 Paires de bases de l'ADN.	13
Figure 1.11 Exemple de nature antiparallèle de la double hélice d'ADN.....	14
Figure 1.12 Présentation schématique des paires de bases empilées X/Y et X'/Y' et de l'angle de torsion.	15
Figure 1.13 Comparaison des formes A, B et Z de la double hélice d'ADN.....	16
Figure 1.14 Hélices droite et gauche sont liées l'une à l'autre de la même manière que les mains droite et gauche.	16
Figure 2.1 Circulation de l'information génétique.....	20
Figure 2.2 Hypothèses pour la réplication de l'ADN.....	21
Figure 2.3 Trois hypothèses de la réplication de l'ADN.....	22
Figure 2.4 Réplication bidirectionnelle de l'ADN.....	23
Figure 2.5 Processus de réplication de l'ADN.....	24
Figure 2.6 Synthèse d'ADN se produit toujours de 5' à 3'.....	27
Figure 2.7 Mécanisme de la réplication semi-discontinue de l'ADN.	28
Figure 2.8 Caractéristiques générales du fork de réplication.	29
Figure 2.9 Superenroulement se forme en amont de la fourche de réplication.	29
Figure 2.10 Amorces d'ARN sont formées avant la réplication de l'ADN.	31
Figure 2.11 Amorces d'ARN sont retirées avant que les entailles d'ADN ne soient fixées.....	32
Figure 2.12 Brin principal et le brin retardé sont couplés ensemble au cours du processus de réplication chez <i>E. coli</i>	33
Figure 2.13 Processus de réplication de l'ADN bactérien.	35
Figure 2.14 Diagramme schématique de la réplication en cercle roulant pour l'ADN simple brin.	36
Figure 2.15 Diagramme schématique de la réplication en cercle roulant pour l'ADN double brin.	36
Figure 2.16 Processus de réplication circulaire de l'ADN aboutit à deux molécules filles liées ou caténées.....	38
Figure 2.17 Réunion des fourches de réplication crée de l'ADN interconnecté sur des chromosomes linéaires.....	38
Figure 2.18 Télomérase se lie à 3'CG de queue riche.	40
Figure 2.19 Mécanisme de navette de l'édition dans les ADN polymérases.	43
Figure 2.20 Différents types de mutations ponctuelles dans le cadre de lecture ouvert.	44
Figure 2.21 Exemple de formation de mutation.	45
Figure 2.22 Diagramme schématique de la réparation des mésappariements dans <i>E. coli</i>	47
Figure 3.1 Diagramme schématique de la liaison de l'ARN polymérase à l'ADN et de sa transcription pour produire un nouvel ARN simple brin.	48
Figure 3.2 Exemple typique de bulle de transcription qui se compose de 14 bases non appariées, les 9 premières ont été utilisées pour transcrire un nouvel ARN simple brin.	49
Figure 3.3 Étapes d'initiation et d'élongation de la transcription.....	50
Figure 3.4 Deux modèles d'élongation de la transcription.	52
Figure 3.5 Caractéristiques intrinsèques de la séquence de terminaison - répétitions inversées et une série d'uraciles.....	52
Figure 3.6 Terminaison transcriptionnelle dépendante de Rho.	53
Figure 3.7 Exemple de contrôle temporel de la transcription chez <i>B. subtilis</i> infecté par le phage SPO1.....	55

Figure 3.8 Contrôle temporel de la transcription chez le phage T7 infecté E. coli.....	56
Figure 3.9 Unités de transcription dans les gènes N et cro du phage lambda sont des gènes précoces immédiats et sont séparées des gènes précoces retardés par des terminateurs.....	57
Figure 3.10 Aperçu de la transcription d'un opéron.	58
Figure 3.11 Régulation de base de l'expression des gènes.	59
Figure 3.12 Contrôle négatif de l'opéron lac.	60
Figure 3.13 Complexe médiateur et protéines activatrices.....	61
Figure 3.14 Répression catabolique.	62
Figure 3.15 Aperçu de la régulation de l'opéron trp.	63
Figure 3.16 Quatre éléments des promoteurs de l'ARN polymérase II.	66
Figure 3.17 Promoteurs d'ARN pol I contiennent une région centrale ainsi qu'un élément de contrôle en amont (UCE), qui interagissent avec des facteurs de transcription comme UBF et SL1 chez l'homme.	66
Figure 3.18 Exemples de promoteurs de l'ARN polymérase III.....	66
Figure 4.1 Organigramme démontrant le dogme central de la biologie dans lequel l'ADN est transcrit en ARNm, qui est ensuite traduit en une séquence d'acides aminés d'une protéine.....	68
Figure 4.2 Composition de ribosomes procaryotes par rapport à la composition de ribosomes eucaryotes.	70
Figure 4.3 Séquences de bases complémentaires liées à l'hydrogène de l'ARN ribosomal 16S maintenant la structure secondaire.....	72
Figure 4.4 Modèle de trèfle d'ARNt avec des liaisons hydrogène représentées par des lignes pointillées.....	73
Figure 4.5 Exemple de liaisons hydrogène entre les bases de l'ARNt, qui contribue à sa structure tridimensionnelle.....	74
Figure 4.6 Étapes impliquées dans la charge de l'ARNt avec un acide aminé.	75
Figure 4.7 Réaction globale de l'aminocyl-ARNt synthétase.	76
Figure 4.8 Structure tertiaire de l'ARNt représentée par un diagramme en ruban.....	78
Figure 4.9 Formation du complexe d'initiation de la traduction avec l'ARNm de liaison de la sous-unité ribosomique 30S et un ARNt initial en présence de facteurs d'initiation, IF-1, IF-2 et IF-3. La sous-unité ribosomale 50S est ensuite ajoutée pour former le complexe 70S.	79
Figure 4.10 Étapes de l'élongation translationnel.....	80
Figure 4.11 Terminaison de la chaîne peptidique dans laquelle il n'y a pas de molécules d'ARNt capables de lire les codons d'arrêt, qui sont ensuite lus par le facteur de libération (RF) au lieu de l'ARNt.	82
Figure 4.12 Polysomes procaryotes permettent la synthèse simultanée de protéines le long d'une seule molécule d'ARNm.	83
Figure 4.13 Étapes de l'initiation eucaryote de la traduction.....	84

Tableau 2.1 Propriétés de trois ADN polymérase différentes d' <i>E. coli</i>	32
Tableau 4.1 Code génétique dégénératif	68
Tableau 4.2 Appariement de bases de l'hypothèse de Wobble au troisième codon.....	68
Tableau 4.3 Classes d'aminocyl-ARNt synthétase	76

Introduction i

1	Acides nucléiques : ADN et ARN	1
1.1	Le matériel génétique.....	1
1.1.1	Transformation bactérienne	1
1.1.2	L'ADN des bactéries	2
1.1.3	L'ADN du bactériophage	3
1.1.4	L'ARN des virus	7
1.2	Les composants des acides nucleiques	7
1.2.1	Base azotée	7
1.2.2	Sucre pentose	8
1.2.3	Groupe phosphate	9
1.2.4	La formation d'un nucléotide	9
1.3	La structure des molécules d'ADN	11
1.3.1	Règles de Chargaff.....	11
1.3.2	L'arrangement de chaque sous-unité dans la structure de l'ADN.....	13
1.3.3	Les formes alternatives d'ADN.....	15
1.3.4	Les diverses formes d'ARN	17
2	Activités cellulaires à base d'acide nucléique.....	19
2.1	Réplication de l'ADN.....	19
2.1.1	L'hypothèse de la réplication de l'ADN	20
2.1.2	La réplication de l'ADN suit le mécanisme de réplication semi-conservateur	21
2.2	Le processus de réplication de l'ADN.....	23
2.2.1	Présentation des trois phases différentes de la réplication	23
2.2.2	Mécanisme de réplication semi-discontinue de l'ADN.....	25
2.2.3	Réplication de l'ADN eucaryote	25
2.3	Protéines nécessaires à la réplication de l'ADN chez <i>E. COLI</i>	26
2.3.1	Hélicase et protéines de liaison à l'ADN simple brin	26
2.3.2	ADN gyrase	28
2.3.3	Primase	30
2.3.4	ADN polymérase	31
2.4	Coupler la réplication de l'ADN	33
2.4.1	Le Réplisome	33
2.4.2	L'Assemblée du Réplisome.....	34
2.5	Modèles de réplication.....	34
2.5.1	Réplication bidirectionnelle	34

2.5.2	Réplication en cercle roulant	35
2.5.3	Arrêt de la réplication	37
2.5.4	Réplication au niveau des télomères.....	39
2.5.5	Réplication mitochondriale.....	41
2.6	Fidélité de la réplication de l'ADN	41
2.6.1	Mécanisme de relecture	41
2.6.2	Activité polymérase exonucléase.....	42
2.7	Mutation de l'ADN	43
2.7.1	Mutation ponctuelle	43
2.7.2	Insertion et suppression	44
2.7.3	Décalages tautomères	44
2.7.4	Réparation des incompatibilités.....	46
3	Expression génique : Transcription du code génétique	48
3.1	Le processus de transcription.....	48
3.1.1	Initiation.....	48
3.1.2	Élongation.....	51
3.1.3	Terminaison	52
3.2	La régulation de la transcription procaryotique	53
3.2.1	Facteurs σ alternatifs.....	54
3.2.2	Commutation de l'ARN polymérase	56
3.2.3	Anti-terminaison	56
3.2.4	Opérons.....	57
3.2.5	Améliorants.....	60
3.2.6	Atténuation de la transcription.....	62
3.2.7	Riboswitches.....	63
3.3	Transcription chez les eucaryotes	64
3.3.1	ARN polymérase eucaryote	64
3.3.2	Structure du promoteur	64
3.3.3	Transcription dirigée par l'ARN polymérase I et III.....	65
4	Expression génique : Traduction du code génétique.....	67
4.1	Le code génétique	67
4.1.1	Les bases de ribonucléotides utilisées comme lettres dans le code génétique.....	67
4.1.2	Règles d'oscillation	67
4.2	Machine translationnelle : Ribosome.....	69
4.2.1	Les composants des ribosomes	69
4.2.2	Structure ribosomique.....	70

4.3	Machine translationnelle : ARNt	72
4.3.1	La structure de l'ARNt	72
4.3.2	La structure secondaire de l'ARNt	72
4.3.3	La forme active de l'ARNt	73
4.4	Préparation à la traduction	74
4.4.1	Activation de l'ARNt	74
4.4.2	Dissociation des ribosomes.....	75
4.5	Aminoacyl-ARNt synthétases.....	75
4.5.1	Classification de l' aminoacyl-ARNt synthétase.....	75
4.5.2	Deuxième code génétique	77
4.5.3	Activité de relécture des synthétases Aminoacyl-ARNt.....	78
4.6	Traduction procaryotique.....	78
4.6.1	Initiation à la traduction.....	78
4.6.2	Elongation de traduction.....	80
4.6.3	Terminaison de la traduction	81
4.6.4	Traduction couplée	82
4.7	Traduction eucaryotique	83
4.7.1	Initiation.....	83
4.7.2	Élongation.....	84
4.7.3	Terminaison	84

Séries de TD ii

Références bibliographiques

Introduction

Depuis sa naissance dans les années 1950, la biologie moléculaire est devenue une branche dominante et exceptionnellement productive des sciences de la vie. Le terme « biologie moléculaire » a été inventé pour la première fois par le mathématicien Warren Weaver (1894-1978). Il a reconnu très tôt que les outils et les techniques de la chimie et de la physique pouvaient être adoptés pour étudier la vie au niveau moléculaire. Notamment, la proposition de Weaver d'employer les techniques modernes (implicitement de la physique et de la chimie) pour étudier la biologie à des détails moléculaires ne précisait pas les domaines de cette nouvelle branche de la science. Lorsque la biologie moléculaire a commencé à prendre racine dans les années 1950, l'un de ses premiers partisans, le physicien britannique William Astbury (1898-1961) a réitéré en 1961 que l'objectif de la biologie moléculaire était l'étude de la structure, des origines et des fonctions des biomolécules

La biologie moléculaire est l'intégration des connaissances et de la technologie en biologie moléculaire avec les techniques de laboratoire. C'est une réalité profondément enracinée dans la compréhension fondamentale de la structure du génome et des activités cellulaires au niveau moléculaire. Au cours des dernières années, le domaine du diagnostic moléculaire s'est développé rapidement en raison des développements et des découvertes révolutionnaires dans les domaines de la génétique moléculaire et de la biotechnologie, et de la nécessité de fournir une analyse au niveau moléculaire pour toutes sortes de maladies. Pour cela, les techniques d'étude et de modification des gènes et de leur expression font partie intégrante de la biologie moléculaire. La biologie moléculaire est une science qui décrit comment l'information génétique est préservée, transmise et exprimée.

Ce polycopié est destiné aux étudiants de Licence Biochimie, qui ont l'intention d'entrer dans les professions de laboratoire clinique avec des connaissances en biologie générale, en génétique et en microbiologie. Afin de fournir la compréhension la plus actuelle des techniques permettant de modifier et de transférer de façon contrôlée du matériel génétique, ce document est organisé en quatre chapitres sur les bases de la biologie moléculaire. Chapitre 1 sert d'introduction à la structure et à la fonction des acides nucléiques. Chapitres 2 à 4 contiennent une couverture de base des activités cellulaires basées sur les acides nucléiques, y compris la réplication, la réparation et l'expression des gènes.

1 Acides nucléiques : ADN et ARN

1.1 Le matériel génétique

1.1.1 Transformation bactérienne

En 1928, dans une tentative de développement d'un vaccin contre la pneumonie, Frederick Griffith est devenu le premier à identifier la transformation bactérienne, dans laquelle la forme et la fonction d'une bactérie changent. Les *Streptococcus pneumoniae* virulents et avirulents faisaient l'objet de son étude. La virulence de la bactérie est déterminée par son polysaccharide capsulaire. Les souches virulentes ont une capsule qui est enfermée dans un polysaccharide capsulaire, alors que les souches avirulentes n'en ont pas. Les bactéries non encapsulées sont facilement englouties et détruites par les cellules phagocytaires du système circulatoire de l'animal hôte. Cependant, en raison de leur capsule polysaccharidique externe protectrice, les souches virulentes ne sont pas facilement englouties par le système immunitaire de l'hôte, elles peuvent donc se multiplier et provoquer une pneumonie. La présence ou l'absence de la capsule provoque également une différence visible entre les colonies de souches virulentes et avirulentes (Downie, 1972).

Les bactéries encapsulées forment des colonies lisses et brillantes (S) lorsqu'elles sont cultivées sur une plaque de culture en gélose. En revanche, les souches non encapsulées produisent des colonies rugueuses (R). En tant que tel, il est facile d'identifier la différence entre ces deux souches grâce à une technique de culture microbiologique standard.

Dans l'expérience de Griffith, le *S. pneumoniae* virulent qui a une capsule lisse (S) dans son apparence était capable de provoquer des infections mortelles lors de l'injection chez la souris (Fig. 1.1). Du fait de leur absence de manteau protecteur, les bactéries de type R sont détruites par l'animal après l'injection, comme décrit précédemment. Ainsi, les souris sont toujours en vie après l'injection de bactéries de type R. Lorsque les bactéries de type S ont été tuées par la chaleur, elles n'étaient plus capables de provoquer une infection mortelle lors de l'injection à des souris seules. Cependant, lorsque les bactéries de type S tuées par la chaleur et les bactéries de type R vivantes ont été injectées ensemble, aucune d'entre elles ne provoquant à elle seule une infection mortelle, les souris sont mortes des suites d'une pneumonie. Il a été constaté que le trait virulent responsable de la production de la capsule de polysaccharide était passé des cellules de type S tuées par la chaleur aux cellules de type R vivantes, convertissant ainsi les bactéries de type R en bactéries de type S, lui permettant

devenir virulent et mortel en évitant la réponse immunitaire de l'hôte. Griffith a conclu que les bactéries tuées par la chaleur convertissaient d'une manière ou d'une autre les cellules avirulentes vivantes en cellules virulentes, et il a appelé le composant des bactéries mortes de type S le "principe de transformation" (Downie, 1972 ; Brown, 2002).

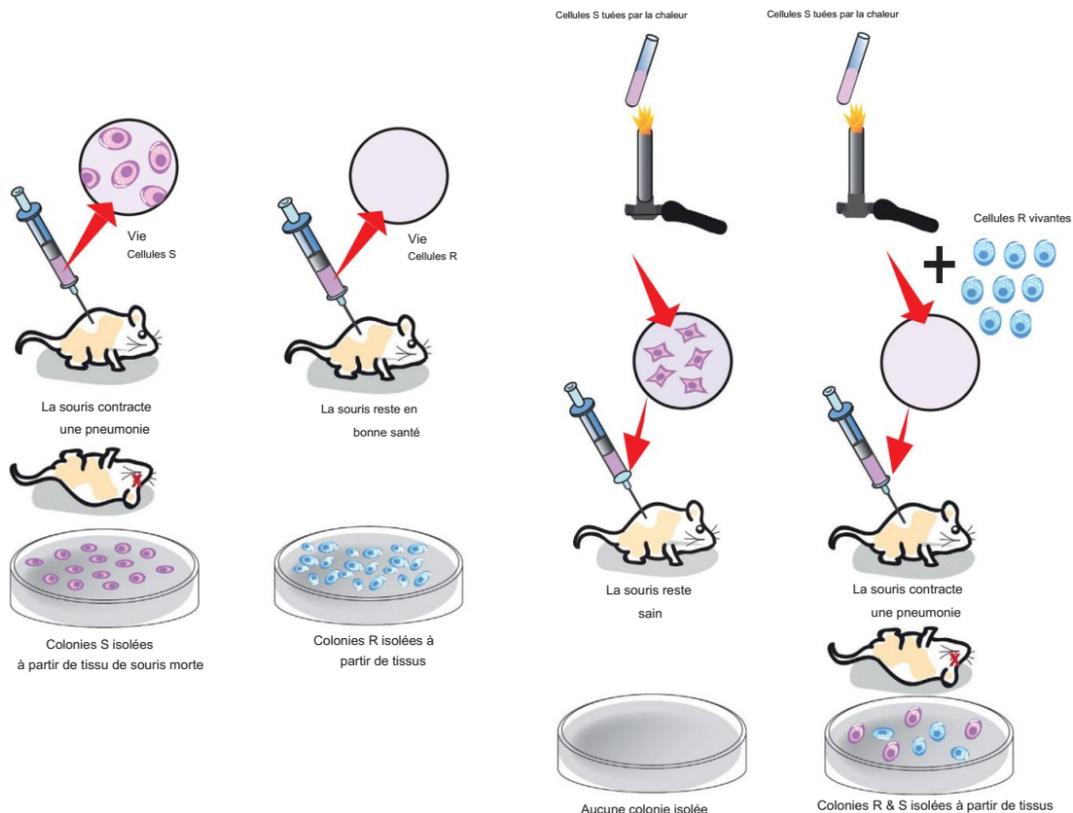


Figure 1.1 Diagramme schématisé de l'expérience de Griffith qui démontre la transformation bactérienne (Brown, 2002 ; Buckingham, 2011).

1.1.2 L'ADN des bactéries

Les travaux de Griffith ont conduit à de nouvelles recherches sur le phénomène de transformation. En 1944, Oswald Avery, Colin MacLeod et Maclyn McCarty ont publié ce qui est maintenant considéré comme un article classique dans le domaine de la génétique moléculaire. Dans ce travail, ils ont démontré que l'ADN est le principe de transformation. Le schéma de principe de leur expérience est présenté à la figure. 1.2. Dans leurs expériences, ils ont retiré la protéine de l'extrait de transformation par extraction au solvant organique. Après ce traitement, les protéines étaient absentes de l'extrait transformant. Ils ont découvert que le principe de transformation était toujours actif, ce qui signifiait que les bactéries tuées par la chaleur étaient toujours capables de convertir des cellules avirulentes vivantes en cellules virulentes. Ils ont également effectué des analyses chimiques, enzymatiques et sérologiques, ainsi que les résultats de l'électrophorèse, de

l'ultracentrifugation et de la spectroscopie ultraviolette. Ces traitements peuvent éliminer les glucides, les lipides, les protéines ou l'ARN de l'extrait. Ils ont découvert que les glucides, les lipides, les protéines et l'ARN n'étaient pas non plus la substance transformante. Les tests chimiques du produit final ont donné une forte réaction positive pour l'ADN. La confirmation finale est venue d'expériences utilisant des échantillons bruts de l'enzyme désoxyribonucléase (DNase) digérant l'ADN, qui peut dégrader l'ADN, en particulier. Ils ont démontré que le principe transformant peut être détruit par cette enzyme. Il n'y a pas eu de perte d'activité de transformation après que la chaleur a inactivé cette enzyme. Ainsi, leurs observations ont confirmé que l'ADN est la substance transformante (Brown, 2002 ; Buckingham, 2011).

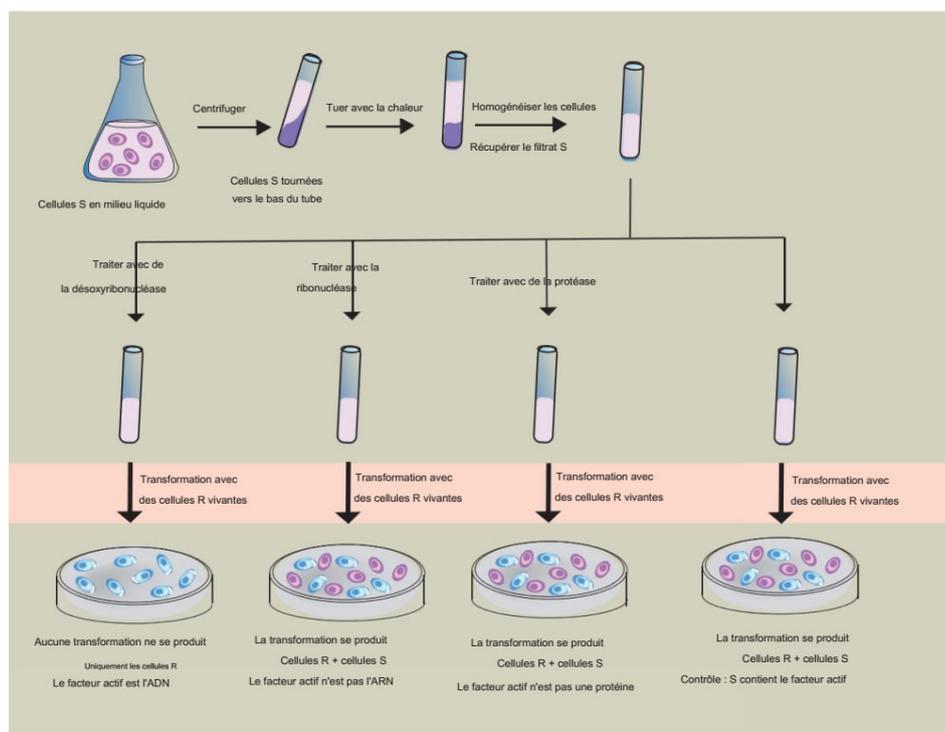


Figure 1.2 Diagramme schématisé de l'expérience d'Avery, MacLeod et McCarty qui démontre que l'ADN est le principe de transformation (Brown, 2002).

1.1.3 L'ADN du bactériophage

Le deuxième élément de preuve majeur à l'appui de l'ADN en tant que matériel génétique provenait d'expériences menées par AD Hershey et Martha Chase en 1952. Hershey et Chase ont utilisé le bactériophage T2 dans leur expérience pour déterminer si l'ADN ou la protéine est le matériel génétique. Le bactériophage peut infecter *E. coli* et utiliser l'hôte pour synthétiser de nouvelles particules de phage. Le phage est constitué d'une enveloppe protéique entourant un noyau d'ADN. Le phage se fixe à la cellule bactérienne et le composant génétique du phage pénètre dans les cellules bactériennes. Suite à l'infection,

la composante génétique virale domine la machinerie cellulaire des cellules hôtes et conduit à la reproduction virale. Par la suite, de nombreux nouveaux phages sont construits et la cellule bactérienne est lysée, libérant les virus descendants. Ce processus est normalement appelé cycle lytique.

Pour définir la fonction de l'enveloppe protéique et de l'acide nucléique dans le processus de reproduction virale, Hershey et Chase ont radiomarqué activement l'ADN de phage avec du phosphore-32 (^{32}P) et une protéine de phage marquée avec du soufre-35 (^{35}S). En effet, l'ADN contient du phosphore mais pas de soufre, alors que les protéines contiennent du soufre, mais pas de phosphore. Hershey et Chase laissent les bactériophages T2 marqués infecter les bactéries non marquées et injectent leur matériel génétique dans les cellules (Fig. 1.3). Après l'attachement et l'entrée du matériel génétique, les enveloppes de phages vides ont été éliminées par une force de cisaillement élevée dans un mélangeur. La force a enlevé les phages attachés afin que le phage et les bactéries puissent être analysés séparément. La centrifugation a séparé les particules de phage plus légères des cellules bactériennes plus lourdes. Suite à cette séparation, les cellules bactériennes, qui contenaient maintenant de l'ADN viral marqué, ont finalement été lysées au fur et à mesure que les nouveaux phages étaient produits. Ces phages descendants contenaient du ^{32}P mais pas du ^{35}S . Ces résultats suggèrent que la protéine de l'enveloppe du phage reste à l'extérieur des cellules hôtes et n'est pas impliquée dans la direction de la production de nouveaux phages. D'autre part, l'ADN du phage pénètre dans les cellules hôtes et est directement impliqué dans la reproduction du phage. Étant donné que le matériel génétique doit d'abord pénétrer dans les cellules infectées, ils ont conclu que l'ADN est le matériel génétique et qu'il contient des gènes transmis de génération en génération. Pris ensemble avec le travail qui avait été fait auparavant, le travail de Hershey et Chase a fourni des preuves définitives et solides pour prouver que l'ADN est le matériel génétique.

Bien que ces expériences aient démontré que l'ADN est le matériel génétique des bactéries et des virus, il était généralement admis que l'ADN est une substance universelle en tant que matériel génétique des eucaryotes. En effet, certaines preuves indirectes ont indiqué que l'ADN est le matériel génétique des eucaryotes. Par exemple, le matériel génétique doit résider sur le chromosome et se trouver dans le noyau. L'ADN et les protéines répondent à ces critères, mais seul l'ADN est enrichi à l'intérieur du noyau, tandis que la protéine est enrichie dans le cytoplasme. De plus, l'ADN se trouve également dans les chloroplastes et les mitochondries, qui sont également connus pour remplir des fonctions

génétiques. En tant que tel, l'ADN ne se trouve que là où les fonctions génétiques primaires se produisent. D'autre part, les protéines se trouvent partout dans la cellule.

La preuve directe que l'ADN est le matériel génétique des eucaryotes provient de la technologie de l'ADN recombinant. Par exemple, un segment d'un fragment d'ADN correspondant à un gène spécifique est isolé et ligaturé à l'ADN bactérien qui peut s'autorépliquer à l'intérieur de la cellule bactérienne. Le complexe résultant est envoyé dans une cellule bactérienne et son expression génétique est examinée.

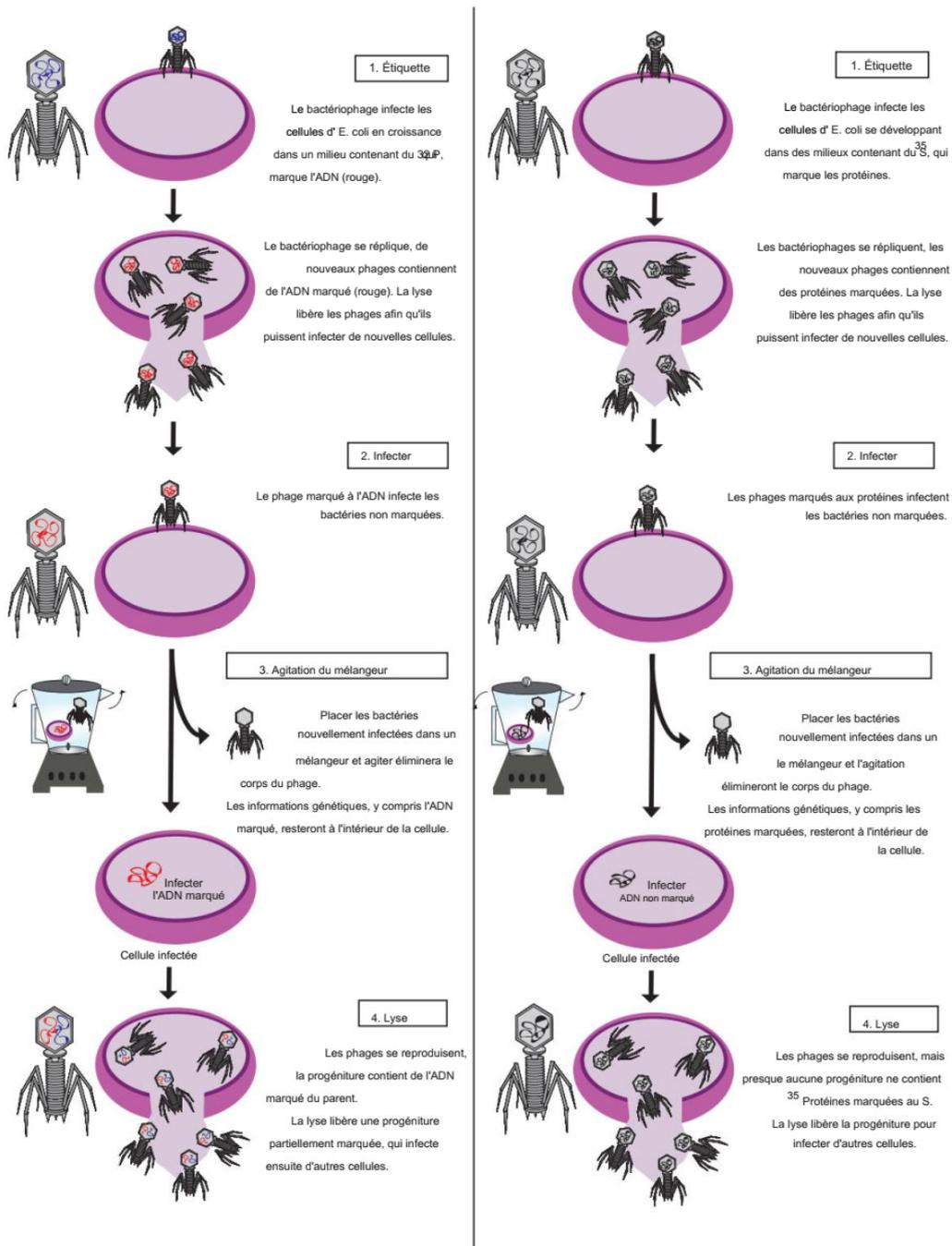


Figure 1.3 Diagramme schématisé de l'expérience Hershey-Chase qui démontre que l'ADN dirige la reproduction du phage T2 dans les cellules *E. coli* infectées (Buckingham, 2011).

La production ultérieure de la protéine eucaryote dérivée de ce segment d'ADN spécifique dans la cellule bactérienne démontre que l'ADN est le matériel génétique dans les cellules eucaryotes. Cette technique dite de clonage de gènes est aujourd'hui largement utilisée dans la recherche biomédicale et la production pharmaceutique actuelles (Fig. 1.4).

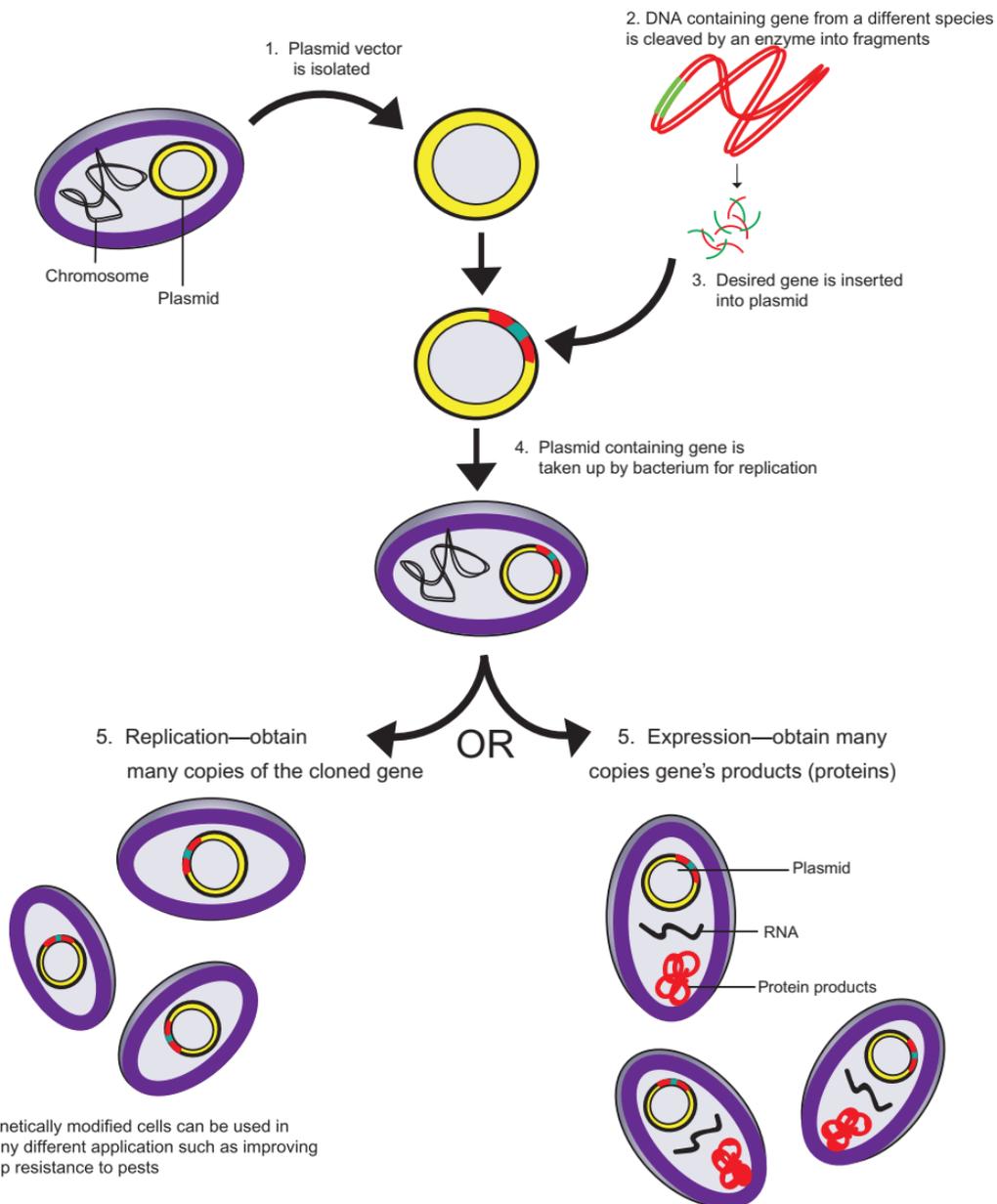


Figure 1.4 Diagramme schématisé d'un processus typique de l'application du clonage de gènes. La production de la protéine eucaryote spécifique dérivée de ce segment d'ADN eucaryote introduit prouve que l'ADN est le matériel génétique des cellules eucaryotes (Brown, 2016).

1.1.4 L'ARN des virus

Bien que l'ADN soit le matériel génétique de la plupart des organismes, il a été démontré que l'autre type d'acide nucléique, l'ARN, peut également être du matériel génétique. Il a été démontré pour la première fois que lorsque l'ARN purifié du virus de la mosaïque du tabac était répandu sur des feuilles de tabac, les feuilles présentaient des lésions d'infection virale. Ainsi, il a été conclu que l'ARN peut être utilisé comme matériel génétique dans les virus. Certains groupes de virus sont connus pour utiliser l'ADN comme matériel héréditaire, comme le bactériophage T2 dans l'expérience de Hershey et Chase. Ceux qui utilisent l'ARN comme matériel génétique sont appelés rétrovirus. Les rétrovirus utilisent une stratégie unique, la transcription inverse, pour répliquer leur matériel génétique en utilisant l'ARN comme matrice pour synthétiser l'ADN complémentaire. Cet intermédiaire d'ADN peut être incorporé dans le génome de la cellule hôte, et lorsque l'ADN hôte est transcrit, des copies de l'ARN rétroviral d'origine sont produites. Ce type de virus à ARN comprend le virus de l'immunodéficience humaine (VIH), qui cause le SIDA.

1.2 Les composants des acides nucléiques

L'ADN et l'ARN sont composés de diverses combinaisons d'acides nucléiques. Les acides nucléiques sont des macromolécules qui existent sous forme de polymères appelés polynucléotides. Un polynucléotide est constitué de nombreux monomères appelés nucléotides et est considéré comme la pierre angulaire de toutes les molécules d'acide nucléique. Ces unités structurales d'acides nucléiques sont constituées de trois composants essentiels : une base azotée, un groupe phosphate et un sucre pentose (un sucre à 5 carbones).

1.2.1 Base azotée

Le cycle de sucre à cinq carbones et le contenu de la base azotée entre l'ADN et l'ARN sont légèrement différents l'un de l'autre. Quatre types différents de bases azotées se trouvent dans l'ADN : l'adénine (A), la thymine (T), la cytosine (C) et la guanine (G).

Dans l'ARN, la thymine est remplacée par l'uracile (U). Les structures chimiques de A, G, C, T et U sont illustrées à la (Fig. 1.5A). En raison de leur similitude structurale, nous appelons généralement les doubles anneaux à neuf membres adénine et guanine des purines, et la thymine, l'uracile et la cytosine à un seul anneau à six membres sont des pyrimidines.

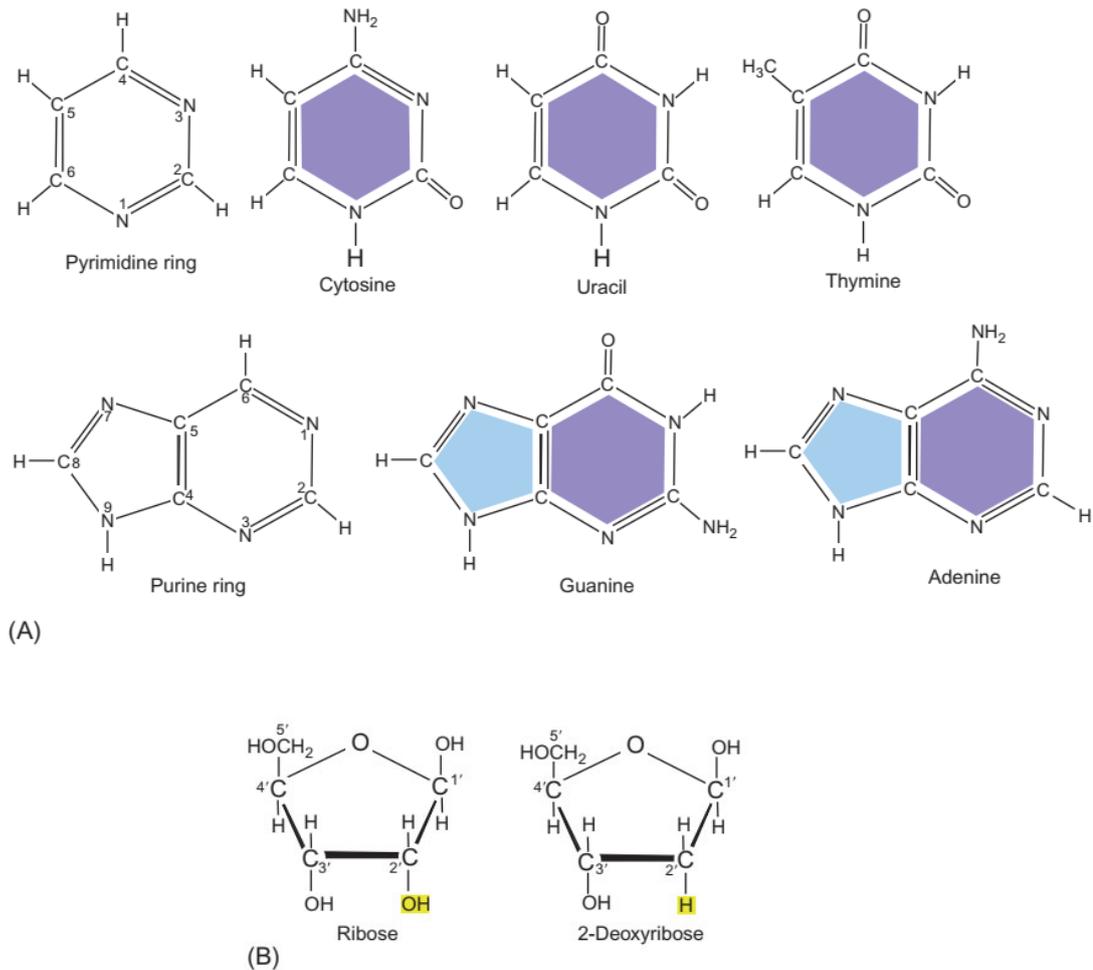


Figure 1.5 (A) Structure chimique des bases azotées pyrimidines et purines dans l'ADN et l'ARN. (B) Structure chimique du ribose et du 2-désoxyribose que l'on trouve respectivement dans l'ARN et l'ADN (Buckingham, 2011).

1.2.2 Sucre pentose

Pour l'anneau de sucre pentose, la figure 1.5B décrit la structure du sucre trouvé dans les acides nucléiques. La différence entre l'ADN et l'ARN réside dans la position C-2' du cycle de sucre ribose. Dans l'ARN, le carbone en position C-2 est attaché à un groupe hydroxyle (OH). Dans l'ADN, le carbone en C-2 ne contient pas ce groupe hydroxyle ; il est plutôt remplacé par un atome d'hydrogène (H).

Par conséquent, le cycle pentose de l'ADN est considéré comme un désoxyribose (il s'agit d'un cycle de sucre désoxygéné à cinq carbones). En l'absence du groupe hydroxyle C-2' de l'ADN, le sucre est plus spécifiquement nommé 2-désoxyribose.

1.2.3 Groupe phosphate

Si une molécule est composée d'une base purique ou pyrimidique et d'un sucre ribose ou désoxyribose, cette unité chimique est appelée nucléoside (Fig. 1.6). La base azotée et le sucre pentose sont liés par une liaison glycosidique entre la position C-1' du sucre et la base azotée. Si la base est une purine, l'atome N-9 est lié par covalence au sucre. Si la base est une pyrimidine, l'atome N-1 se lie au sucre. Lorsqu'un groupe phosphate se fixe à un nucléoside par une liaison phosphoester, l'ensemble du complexe devient un nucléotide. Cette liaison phosphoester est liée entre le groupe 5' - hydroxyle du sucre et un groupe phosphate. Parce qu'il implique un groupe phosphate et un seul sucre, il s'agit d'une liaison phosphomonoester (Buckingham, 2011).

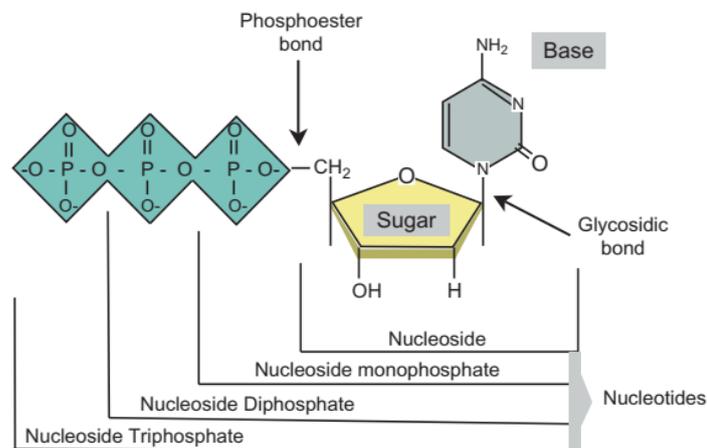


Figure 1.6 Formation d'un nucléotide par addition de groupe(s) phosphate(s) à un nucléoside

1.2.4 La formation d'un nucléotide

Les unités uniques dans les nucléotides sont également appelées monophosphates de nucléosides. L'addition d'un ou deux groupes phosphate donne respectivement des diphosphates et des triphosphates de nucléoside. La forme triphosphate est importante car elle sert de molécule précurseur lors de la synthèse d'acide nucléique dans la cellule. La figure 1.7 décrit les structures de l'adénosine-5' -triphosphate (ATP), de l'adénosine-5' -diphosphate (ADP) et de l'adénosine-5' - monophosphate (AMP). La formation d'ADP à partir d'AMP nécessite l'ajout d'une molécule de phosphate inorganique et s'accompagne de la libération d'une molécule d'eau. De même, la formation d'ATP à partir d'ADP nécessite l'ajout d'une molécule de phosphate inorganique et est accompagnée de la libération d'une molécule d'eau. D'autre part, l'hydrolyse de l'ATP en ADP, libérant une molécule de phosphate inorganique (Pi), s'accompagne de la libération d'une grande quantité d'énergie dans la cellule. Lorsque ces conversions chimiques sont couplées à d'autres réactions,

l'énergie produite est utilisée pour conduire les réactions et maintenir la vie. Au cours de la synthèse de l'ADN, les deux groupes phosphate, qui sont les phosphates β et γ éliminés du dATP, du dGTP, du dCTP et du dTTP. Ainsi, les quatre nucléotides ne contiennent que des monophosphates dans une chaîne polynucléotidique.

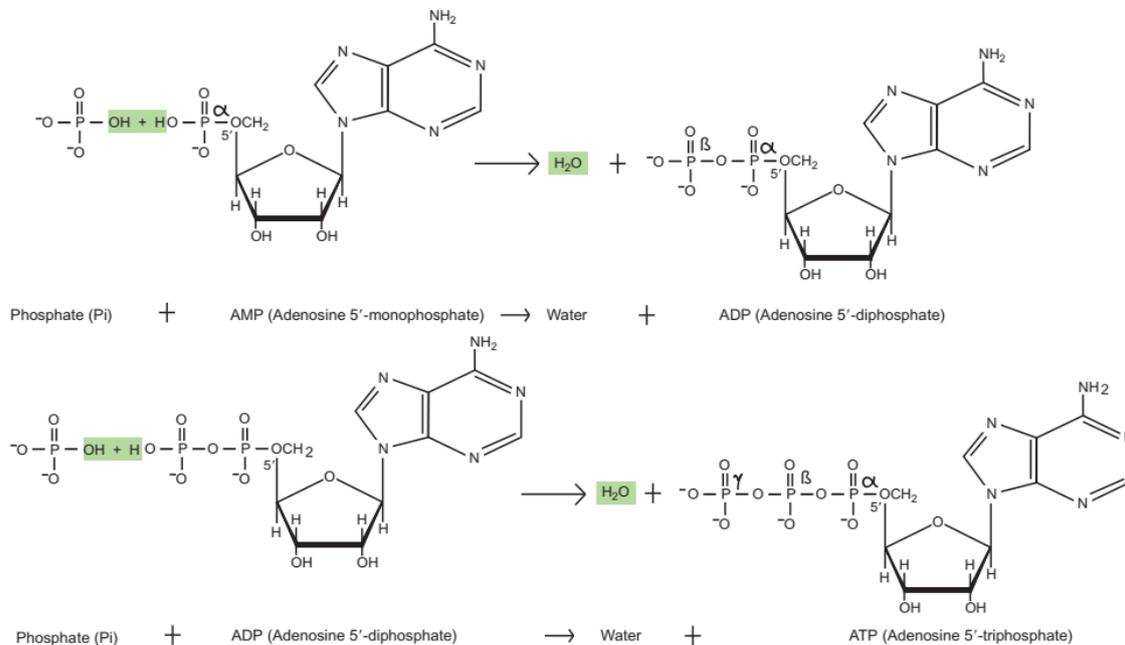


Figure 1.7 Formation d'ADP et d'ATP par addition successive de groupements phosphate via une liaison anhydride phosphorique.

Dans une chaîne polynucléotidique, les nucléotides sont reliés entre eux par une liaison phosphodiester pour former une longue chaîne de nucléotides. La formation d'une liaison phosphodiester implique une réaction de déshydratation (élimination d'une molécule d'eau) par la liaison entre un acide phosphorique et deux sucres, qui se situe entre 5' carbones d'un sucre et 3' carbones d'un autre sucre (Fig. 8). Une telle liaison phosphodiester entraîne un motif répétitif des unités sucre-phosphate appelé squelette sucre-phosphate, ce qui donne à la chaîne polynucléotidique une direction de liaison phosphodiester de 3' à 5'.

Les liaisons phosphodiester peuvent être trouvées dans l'ADN et l'ARN. Une extrémité du polynucléotide a un groupe 5' -phosphate libre, on l'appelle donc l'extrémité 5'. L'autre extrémité du polynucléotide a un groupe 3' - hydroxyle libre, et elle est appelée extrémité 3'. Il est classique d'écrire des séquences d'acides nucléiques dans la direction 5' à 3', qui va de l'extrémité 5' à gauche à l'extrémité 3' à droite. Par exemple, le polynucléotide de la Fig. 1.8 se lirait « TCA ».

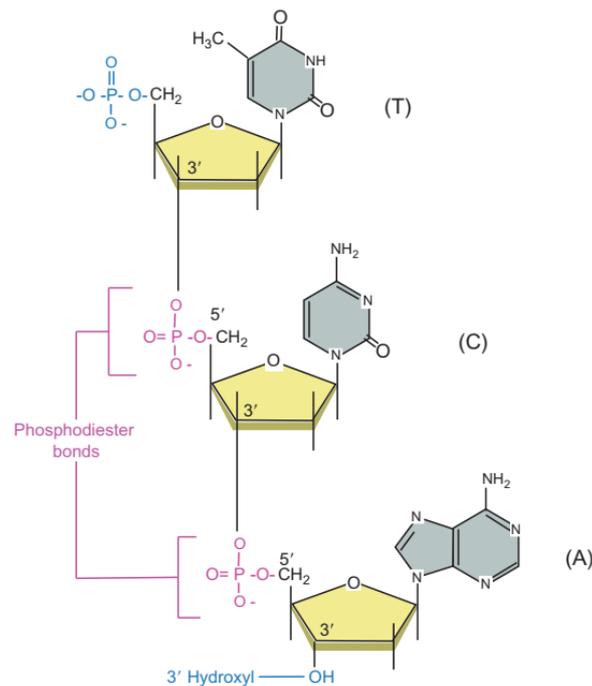


Figure 1.8 Exemple de trois nucléotides liés entre eux par des liaisons phosphodiester entre les groupes 5'- et 3'-hydroxyle des sucres.

1.3 La structure des molécules d'ADN

1.3.1 Règles de Chargaff

En 1953, James Watson et Francis Crick ont proposé la théorie selon laquelle les molécules d'ADN existent sous la forme d'une double hélice, qui a depuis été étayée par diverses études (Fig. 1.9). Le modèle de la double hélice est principalement basé sur les données de diffraction des rayons X recueillies par Rosalind Franklin et Maurice Wilkins, et sur les études de composition de l'ADN observées par Erwin Chargaff. Les données de diffraction des rayons X ont montré que l'ADN est une hélice régulière, et que la distance de répétition dans l'hélice est de 34 angströms (Å) avec un diamètre de $\sim 20\text{Å}$.

La distance entre les nucléotides adjacents est de $3,4\text{Å}$. La découverte du modèle à double hélice de l'ADN s'est appuyée sur les données critiques des découvertes de Chargaff, que l'on appelle aussi les règles de Chargaff :

1. Deux longues chaînes polynucléotidiques sont enroulées autour d'un axe central, formant une double hélice droite.
2. Les deux brins d'ADN sont antiparallèles, c'est-à-dire que leur orientation $5' \rightarrow 3'$ va dans des directions opposées.
3. La base des deux chaînes est perpendiculaire à l'axe et elles sont empilées l'une sur l'autre.

4. Les bases azotées des chaînes opposées sont appariées à la suite de la formation d'une liaison hydrogène dans l'ADN.
5. Chaque tour complet de l'hélice a une longueur de 34Å .
6. La double hélice a un diamètre de 20Å .
7. La quantité de résidus d'adénine (A) est proportionnelle à la quantité de résidus de thymine (T) dans l'ADN. Aussi, le montant de résidus de guanine (G) est proportionnel à la quantité de cytosine (C).
8. La somme des purines égale à la somme de la pyrimidine.
9. Le pourcentage de (G+C) n'est pas nécessairement égal au pourcentage de (A+T).

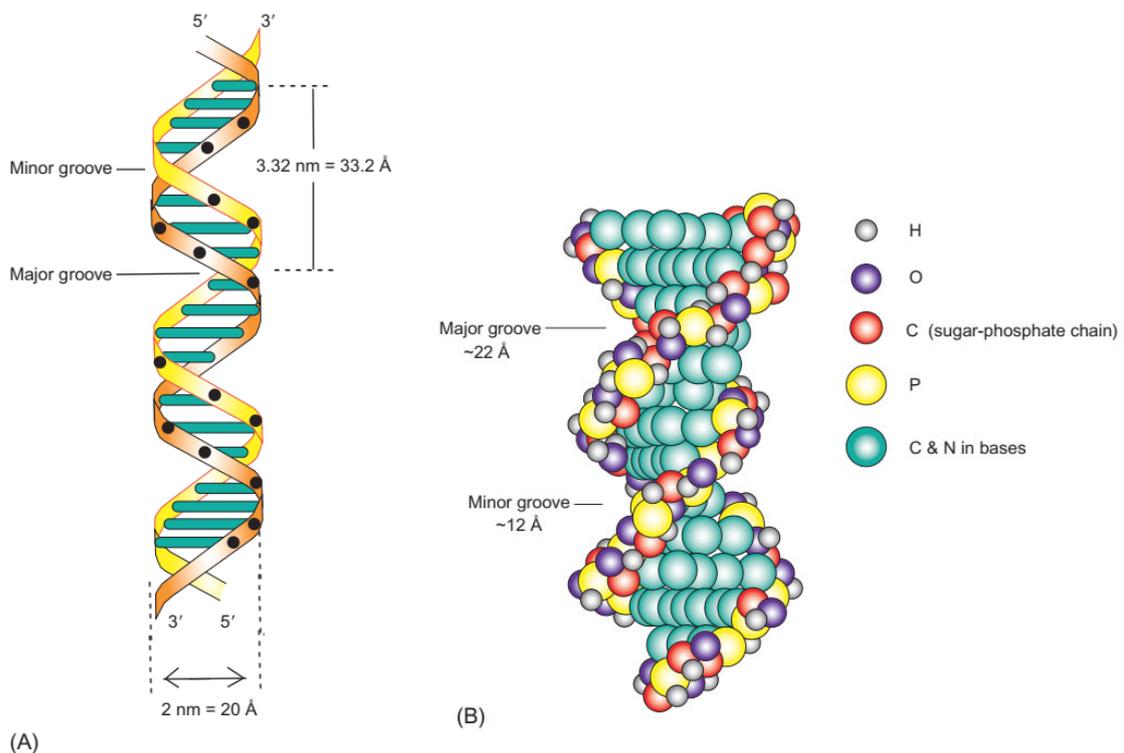


Figure 1.9 Deux modèles de structure de l'ADN.

(A) La double hélice d'ADN est présentée comme une échelle torsadée. Les deux longues barres latérales représentent les squelettes de phosphate de sucre des deux brins et les échelons sont des paires de bases. Les flèches courbes indiquent l'orientation 5' à 3' de chaque brin. (B) Un modèle de remplissage de l'espace de la structure de l'ADN (Buckingham, 2011).

Les règles de Chargaff indiquent un modèle défini de composition de bases dans les molécules d'ADN. En combinant les règles de Chargaff avec le fait que la structure de l'ADN a une périodicité de $3,4\text{Å}$, Watson et Crick ont proposé que l'ADN doit être une double hélice avec son squelette sucre-phosphate à l'extérieur et ses bases azotées à l'intérieur. Les deux chaînes polynucléotidiques de la double hélice sont maintenues par une liaison hydrogène entre les bases azotées, connue sous le nom d'appariement de bases. En général, pour l'ADN, G ne peut se lier spécifiquement qu'avec C, alors que A ne peut se lier

spécifiquement qu'avec T (Fig. 1.10). Les bases appariées sont dites complémentaires. L'appariement de bases complémentaires se produit en raison de l'emplacement géométrique et des interactions entre les groupes fonctionnels dans les bases azotées, de sorte qu'une liaison hydrogène peut se former.

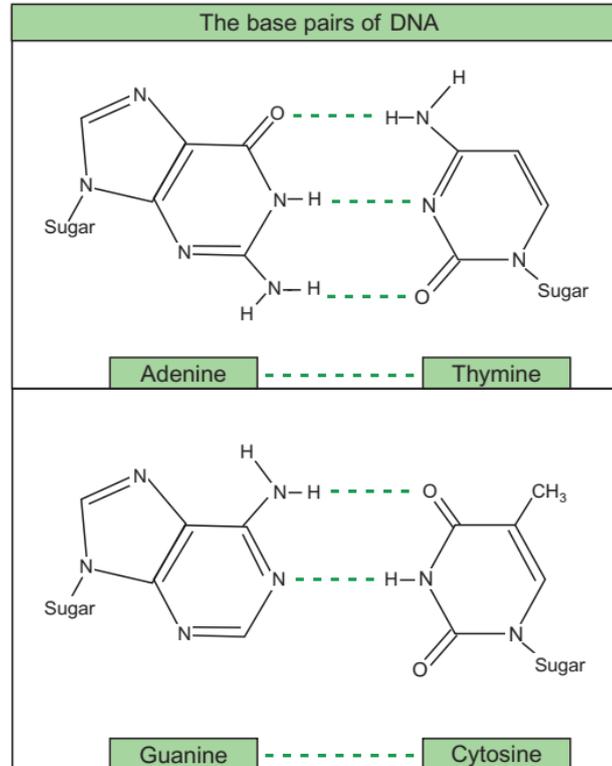


Figure 1.10 Paires de bases de l'ADN.

Une paire guanine-cytosine est maintenue par trois liaisons hydrogène (lignes pointillées) et une paire adénine thymine est maintenue par deux liaisons hydrogène).

1.3.2 L'arrangement de chaque sous-unité dans la structure de l'ADN

Le modèle Watson-Crick place les squelettes sucre-phosphate à l'extérieur de la double hélice et porte les charges négatives sur le groupe phosphate. En raison de ses charges négatives, les protéines chromosomiques chargées positivement ou les protéines régulatrices peuvent facilement fournir une force neutralisante soit pour déterminer la structure de l'ADN d'ordre supérieur, soit pour réguler l'expression des gènes. Les deux chaînes polynucléotidiques fonctionnent dans des directions opposées appelées antiparallèles. Un brin va dans la direction 5' à 3', tandis que son complément va de 3' à 5' (Fig. 1.11).

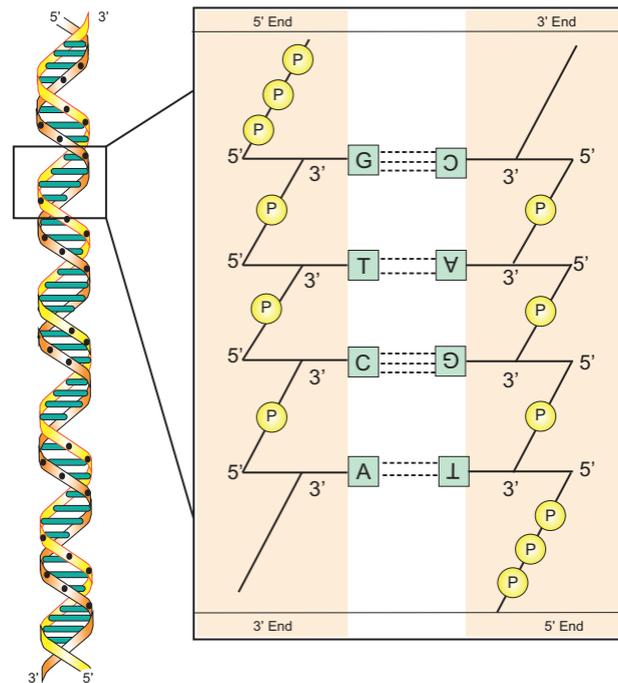


Figure 1.11 Exemple de nature antiparallèle de la double hélice d'ADN (Buckingham, 2011).

Les bases azotées sont à l'intérieur de la double hélice. Ils sont plats et perpendiculaires à l'axe de l'hélice. Les bases sont empilées les unes sur les autres autour de l'axe comme un escalier en colimaçon (Fig. 1.12). Les côtés incurvés de l'escalier en colimaçon représentent les squelettes sucre-phosphate des deux brins d'ADN; les marches sont les bases azotées. Comme mentionné précédemment, Les données de diffraction des rayons X ont montré que chaque double hélice répétée est de 34 \AA et que la distance entre les nucléotides adjacents est de $3,4 \text{ \AA}$. Celles-ci suggèrent que chaque tour d'hélice a ~ 10 paires de bases et que chaque paire de bases tourne de 36 degrés autour de l'axe de l'hélice par rapport à la paire de bases suivante, de sorte que ~ 10 paires de bases effectuent un tour complet de 360 degrés. La torsion des deux brins l'un autour de l'autre forme une double hélice avec un petit sillon et un grand sillon. Dans un petit sillon, la distance entre les deux brins d'ADN est de $\sim 12 \text{ \AA}$, tandis que la distance devient $\sim 22 \text{ \AA}$ dans un sillon majeur. La double hélice de l'ADN n'est ni généralement à droite, ce qui signifie que les tours tournent dans le sens des aiguilles d'une montre, vus le long de l'axe hélicoïdal. Il est important de savoir que ~ 10 paires de bases par tour est une structure moyenne. S'il a plus de paires de bases par tour, on dit qu'il est surbobiné. D'autre part, s'il a moins de paires de bases par tour, alors il est sous-enroulé. Le degré d'enroulement local peut être affecté par la conformation globale de la double hélice d'ADN ou par la liaison de protéines à des sites spécifiques sur l'ADN (Buckingham, 2011).

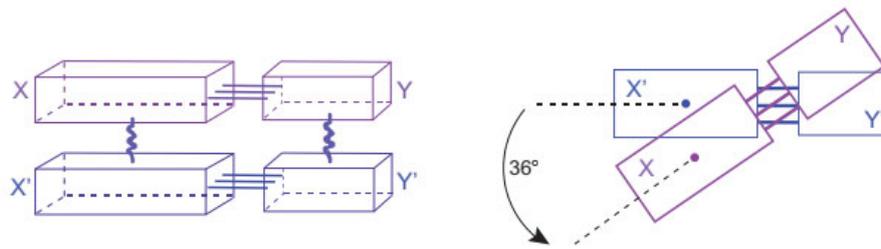


Figure 1.12 Présentation schématique des paires de bases empilées X/Y et X'/Y' et de l'angle de torsion (Buckingham, 2011).

1.3.3 Les formes alternatives d'ADN

Dans différentes conditions d'isolement, différentes conformations d'ADN sont observées. La molécule d'ADN Watson-Crick représente la molécule d'ADN en solution, qui est la molécule d'ADN qui existe dans un environnement à humidité relative très élevée (92 %). C'est ce qu'on appelle "l'ADN-B" (Fig. 1.13). Dans l'ADN-B, la double hélice est dite droite, car les tours tournent dans le sens des aiguilles d'une montre vus le long de l'axe hélicoïdal, ce qui signifie que l'hélice s'enroule vers le haut dans la direction dans laquelle les doigts de la main droite se courbent lorsque le pouce est pointant vers le haut (Fig. 1.14). Le diamètre intérieur du squelette sucre-phosphate de la double hélice est d'environ 11 Å (1,1 nm). La distance entre les points d'attache des bases aux deux brins du squelette sucre-phosphate est la même pour les deux paires de bases (AT et GC), environ 11 Å (1,1 nm), ce qui permet un empilement symétrique et une double hélice avec une colonne vertébrale lisse et sans renflements apparents. Des paires de bases autres que AT et GC sont possibles, mais elles n'ont pas un schéma de liaison hydrogène correct (paires AC ou GT) ou les bonnes dimensions (paires purine-purine ou pyrimidine-pyrimidine) pour permettre une double hélice lisse. Le diamètre extérieur de l'hélice 20 Å (2 nm). La longueur d'un tour complet de l'hélice le long de son axe est de 34 Å (3,4 nm) et contient 10 paires de bases. Les atomes qui composent les deux chaînes polynucléotidiques de la double hélice ne remplissent pas complètement un cylindre imaginaire autour de la double hélice ; ils laissent des espaces vides appelés rainures. Il y a un grand sillon majeur et un plus petit sillon mineur dans la double hélice ; les deux peuvent être des sites auxquels des médicaments ou des polypeptides se lient à l'ADN. A pH physiologique neutre, chaque groupement phosphate du squelette porte une charge négative. Les ions chargés positivement, tels que Na^+ ou Mg^{2+} , et les polypeptides avec des chaînes latérales chargées positivement doivent être associés à l'ADN afin de neutraliser les charges négatives. L'ADN eucaryote, par exemple, est complexé avec des histones, qui sont des protéines chargées positivement, dans le noyau cellulaire. L'ADN-B est la forme la plus courante *in vivo* et en solution *in vitro*.

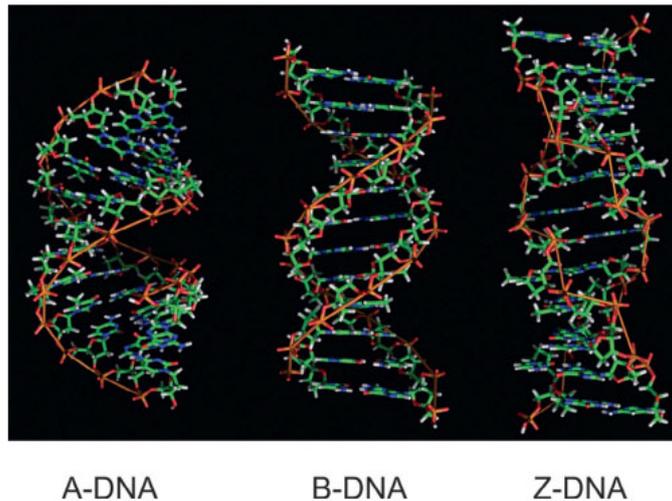


Figure 1.13 Comparaison des formes A, B et Z de la double hélice d'ADN.

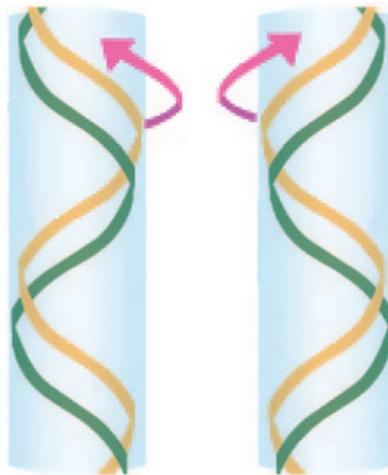


Figure 1.14 Hélices droite et gauche sont liées l'une à l'autre de la même manière que les mains droite et gauche (Brown, 2002).

Une autre forme d'ADN est l'ADN-A, qui est observé lorsque l'ADN est déshydraté ou dans des conditions de sel élevées. Une caractéristique commune importante de l'ADN-A et de l'ADN-B est que les deux sont des hélices droites. L'ADN-A est à la fois plus court et plus épais que l'ADN-B. Chaque double hélice répétée dans l'ADN-A fait $24,6\text{\AA}$, et chaque tour compte environ 11 paires de bases (pb). Le groupe majeur de l'ADN est profond et étroit, tandis que le groupe mineur est peu profond et large. Bien que la grande majorité des molécules d'ADN dans la cellule soient l'ADN-B, il existe quelques variations locales de la structure de l'ADN. Par exemple, l'ADN-A peut être trouvé dans les hybrides ADN-ARN et ARN-ARN. L'ADN-A a été trouvé à l'origine dans des échantillons d'ADN déshydraté, et de nombreux chercheurs pensaient que la forme A était un artefact de la préparation de l'ADN. Les hybrides ADN:ARN peuvent adopter une formation A parce que le 2'-hydroxyle sur le ribose (qui, comme discuté précédemment, distingue l'ADN de l'ARN)

empêche une hélice d'ARN d'adopter la forme B. Des hybrides ARN:ARN peuvent également être trouvés sous la forme A. Une troisième structure d'ADN est l'ADN-Z, qui est plus long et plus étroit que l'ADN-B et est une hélice gauche. Une hélice à gauche tourne dans le sens antihoraire loin du spectateur lorsqu'elle est vue sur son axe. Parce que le squelette formait une structure en zigzag, il est nommé Z-DNA. Dans l'ADN-Z, l'hélice répétée est de $45,6 \text{ \AA}$ et chaque tour hélicoïdal a 12 pb. Le sillon mineur est très profond et étroit. En revanche, le grand sillon est peu profond au point d'être quasiment inexistant. L'ADN-Z est formé dans des conditions de haute teneur en sel ou en présence d'alcool. L'ADN-Z est également connu pour se produire dans la nature lorsqu'il existe une séquence de purine pyrimidine alternée. Étant donné que des séquences avec la cytosine méthylée en position numéro 5 du cycle pyrimidine peuvent également être trouvées sous la forme Z, elles peuvent jouer un rôle dans la régulation de l'activation des gènes.

Comme nous l'avons vu précédemment, l'ARN diffère chimiquement de l'ADN à deux égards. Premièrement, l'ARN contient du ribose C-2' au lieu du désoxyribose C-2' dans l'ADN. Deuxièmement, l'ARN contient la base uracile au lieu de la thymine. Ces deux petites différences font que l'ARN forme une structure unique dans la cellule. Par exemple, l'ARN ne peut pas former d'ADN B double brin car le groupe 2'-hydroxyle sur le sucre ribose empêche la formation de l'ARN de forme B. Cependant, l'ARN adopte une hélice de forme A lorsqu'il forme des régions double brin (Radvanszky et al, 2015).

1.3.4 Les diverses formes d'ARN

Contrairement à l'ADN, les molécules d'ARN cellulaires sont presque toujours monocaténares. Cependant, tous contiennent généralement des régions double brin formées, lorsqu'elles sont étirées, de nucléotides, avec des séquences de bases complémentaires alignées de manière antiparallèle dont nous parlerons plus tard. Plusieurs types d'ARN jouent un rôle important dans les activités cellulaires. Trois grandes classes de molécules d'ARN cellulaires fonctionnent lors de l'expression de l'information génétique : l'ARN ribosomique (ARNr), l'ARN de transfert (ARNt) et l'ARN messager (ARNm). Ces molécules proviennent toutes de copies complémentaires de l'un des deux brins de segments d'ADN au cours du processus de transcription. L'ARN ribosomal constitue habituellement environ 80 % de tout l'ARN dans les cellules d'*E. coli*. Ce sont des composants structurels importants du ribosome, qui fonctionne comme des sites non spécifiques de synthèse des protéines pendant la traduction. Les molécules d'ARN messager transportent des informations génétiques à partir de l'ADN du gène. Les ARNm varient en taille, reflétant la gamme des tailles des protéines

codées par l'ARNm. L'ARN de transfert représente jusqu'à 15 % de l'ARN dans une cellule typique. Il transporte les acides aminés vers le ribosome pendant la traduction, aidant à la traduction de l'ADN en ARNm en protéine. Bien que les molécules d'ARN ribosomal et d'ARN de transfert soient également synthétisées par transcription de séquences d'ADN, contrairement aux molécules d'ARNm, ces ARN ne sont pas ensuite traduits pour former des protéines et restent sous forme d'ARN (Li et al, 2017).

Un seul ARNm bactérien peut contenir l'information pour la synthèse de plusieurs chaînes polypeptidiques au sein de sa séquence nucléotidique. En revanche, les ARNm eucaryotes ne codent qu'un seul polypeptide, mais sont plus complexes en ce sens qu'ils sont synthétisés dans le noyau sous la forme de molécules précurseurs beaucoup plus grosses appelées ARN nucléaires hétérogènes (nhARN). Les molécules de nhARN contiennent des tronçons de séquence de nucléotides qui n'ont aucune capacité de codage de protéine et, par conséquent, ne seront pas traduits en protéine par un ribosome. Ces régions non codantes sont appelées séquences intermédiaires ou introns car elles interviennent entre les régions codantes, appelées exons. L'intron interrompt la continuité de l'information spécifiant la séquence d'acides aminés d'une protéine et doit être épissé avant que le message puisse être traduit.

L'autre ARN important comprend le petit ARN nucléaire (snARN), le petit ARN nucléolaire (snoARN), le micro-ARN (miARN) et le petit ARN interférent (siARN). Le snRNA participe à la transformation des hnARN en ARNm et le snoRNA fonctionne principalement comme chaperons d'ARN dans le traitement de l'ARN ribosomique (ARNr). Les deux ne sont ni des ARNt ni de petites molécules d'ARNr, bien qu'ils soient de taille similaire (100 à 200 nucléotides) à ces espèces. Le snRNA se trouve toujours dans un complexe stable avec des protéines spécifiques formant de petites particules de ribonucléoprotéines nucléaires (snRNP). Les miARN et les siARN sont impliqués dans la régulation des gènes. Les siARN perturbent l'expression des gènes en bloquant la production de protéines spécifiques d'où leur nom « interférant », même si l'ARNm codant pour la protéine a été synthétisé. Les miARN contrôlent la synchronisation du développement en appariant des bases avec et en empêchant la traduction de certains ARNm, bloquant ainsi la synthèse de protéines spécifiques. Cependant, contrairement aux siARN, les miARN (longs de 22 nucléotides) ne provoquent pas de dégradation de l'ARNm (Mehta et al, 2016.).

2 Activités cellulaires à base d'acide nucléique

2.1 Réplication de l'ADN

Le flux d'informations génétiques dans une cellule peut être démontré par l'exemple d'une bactérie dans laquelle un seul chromosome circulaire est présent, et l'information génétique peut être utilisée à l'intérieur et également transmise entre les cellules (Fig. 2.1). Lors de la transmission de l'information génétique, toutes ces activités cellulaires, y compris la réplication de l'ADN, la recombinaison et l'expression des gènes, sont des activités basées sur les acides nucléiques. Par exemple, chaque cellule a besoin d'un ensemble complet et fidèle de matériel génétique. En conséquence, une cellule parentale doit dupliquer tout son génome avant de se diviser. Par la suite, chacune des deux nouvelles cellules filles créées reçoit ensuite une copie du génome via un processus appelé division cellulaire. Ainsi, la réplication de l'ADN est un processus permettant de s'assurer que l'ADN est dupliqué avant la division cellulaire afin que chaque cellule de progéniture reçoive un ou plusieurs chromosomes identiques à ceux du parent. Cependant, dans les organismes multicellulaires, la division cellulaire n'entraîne pas automatiquement la création du nouvel organisme. Au lieu de cela, ils augmentent la taille et la complexité de l'organisme d'origine. Dans chaque cycle cellulaire, les cellules eucaryotes doivent répliquer de grandes quantités d'ADN génomique réparties sur plusieurs chromosomes. Pour accomplir cet exploit dans un laps de temps raisonnable, la réplication s'initie tout au long de la phase S (une étape du cycle cellulaire) à de multiples origines le long de chaque chromosome. L'initiation à partir de ces origines doit être coordonnée afin qu'aucune région du génome ne soit laissée non répliquée et qu'aucune région ne soit répliquée plus d'une fois. La réplication de l'ADN doit également être coordonnée avec la ségrégation des chromosomes pour s'assurer que chaque cellule fille reçoit un complément complet et inaltéré d'informations génétiques. Des erreurs dans la réplication de l'ADN ou la ségrégation des chromosomes peuvent entraîner la perte ou la duplication de ces informations génétiques, événements qui peuvent jouer un rôle important dans la genèse des cellules cancéreuses ou des maladies (Nigro et al, 2015).

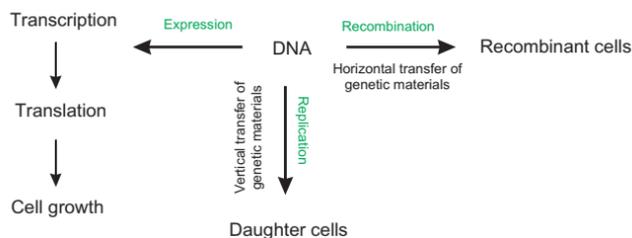


Figure 2.1 Circulation de l'information génétique.

L'ADN peut être obtenu soit à partir d'une cellule mère pour la division, soit à partir d'une autre cellule de la même génération. Il est alors exprimé au sein de cette cellule, ou transféré dans une autre.

2.1.1 L'hypothèse de la réplication de l'ADN

À l'origine, trois modèles de réplication de l'ADN ont été proposés : le modèle semi-conservateur, le modèle conservateur et le modèle dispersif. Dans le modèle semi-conservatif, les deux brins parentaux se séparent, permettant à chaque brin séparé de servir de matrice pour la synthèse d'un brin complémentaire (Fig. 2.2A). Dans ce mécanisme de réplication, chaque molécule d'ADN fille double brin aura un brin d'ADN conservé dérivé de l'ADN parental et un brin nouvellement synthétisé. Dans le modèle conservateur, les deux brins de la double hélice se déroulent au niveau du site de réplication dans la mesure nécessaire pour synthétiser une nouvelle séquence de bases. Les deux brins d'origine restent entrelacés après la réplication de sorte que l'une des deux molécules d'ADN présentes après la réplication contient les deux brins d'origine et l'autre molécule d'ADN est constituée des deux nouveaux brins (Fig. 2.2B). Le modèle dispersif partage certaines des caractéristiques du modèle conservateur, mais prédit que chaque brin des molécules d'ADN filles et des molécules d'ADN parentales comporte des sections intercalées d'ADN ancien et nouveau (Fig. 2.2C). Bien que ces trois modèles prédisent des résultats différents en termes de composition des molécules d'ADN filles, ils suivent tous la règle d'appariement de bases complémentaires pour la réplication.

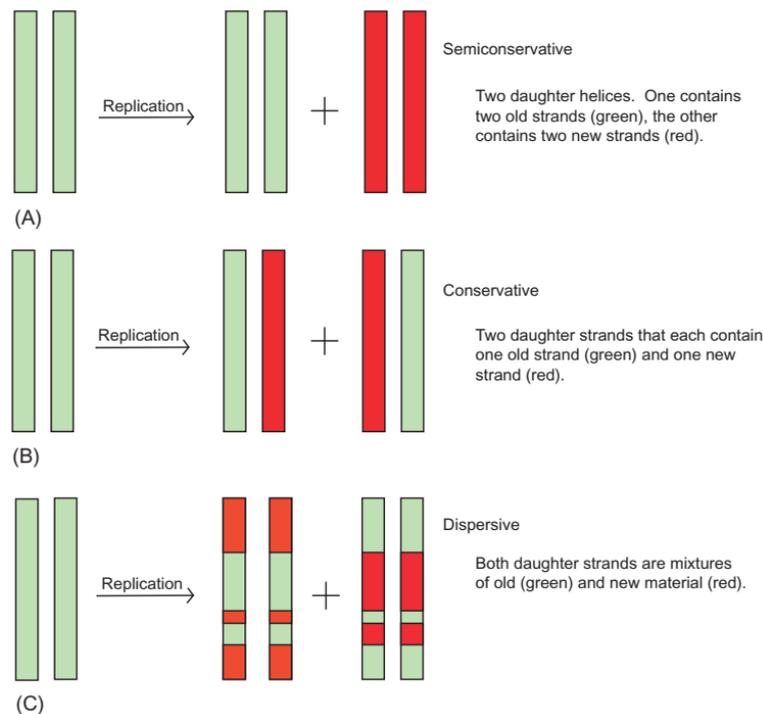


Figure 2.2 Hypothèses pour la réplication de l'ADN.

La réplication peut être semi-conservatrice, conservatrice ou dispersive (Buckingham, 2011).

2.1.2 La réplication de l'ADN suit le mécanisme de réplication semi-conservateur

En 1958, Matthew Meselson et Franklin Stahl ont utilisé des expériences de centrifugation à gradient de densité d'équilibre pour démontrer que la réplication de l'ADN est semi-conservatrice. Dans leurs expériences, ils ont cultivé des cellules d'*E. coli* pendant plusieurs générations dans un milieu de croissance contenant du $^{15}\text{NH}_4\text{Cl}$ comme seule source d'azote (Fig. 2.3). De cette manière, toutes les bases azotées de l'ADN ont été marquées avec l'isotope lourd ^{15}N . Par la suite, ils ont transféré les bactéries dans un nouveau milieu de croissance contenant du $^{14}\text{NH}_4\text{Cl}$ comme seule source d'azote.

Les cellules bactériennes ont été retirées à divers moments pour l'extraction de l'ADN, suivie d'une centrifugation pour le gradient de densité d'équilibre de chlorure de césium. Si le mécanisme de réplication est conservateur, il devrait y avoir deux bandes dans le gradient après une génération, une bande représentant le duplex parental [$^{15}\text{N}/^{15}\text{N}$] et l'autre bande représentant l'ADN nouvellement synthétisé [$^{14}\text{N}/^{14}\text{N}$] (Fig. 2.3). Si la réplication suit un mécanisme semi-conservateur, les deux brins parentaux [$^{15}\text{N}/^{15}\text{N}$] se sépareront pour servir de matrice pour la réplication, et chacun recevra un nouveau brin [^{14}N]. Une bande hybride [$^{15}\text{N}/^{14}\text{N}$] apparaîtrait dans le gradient.

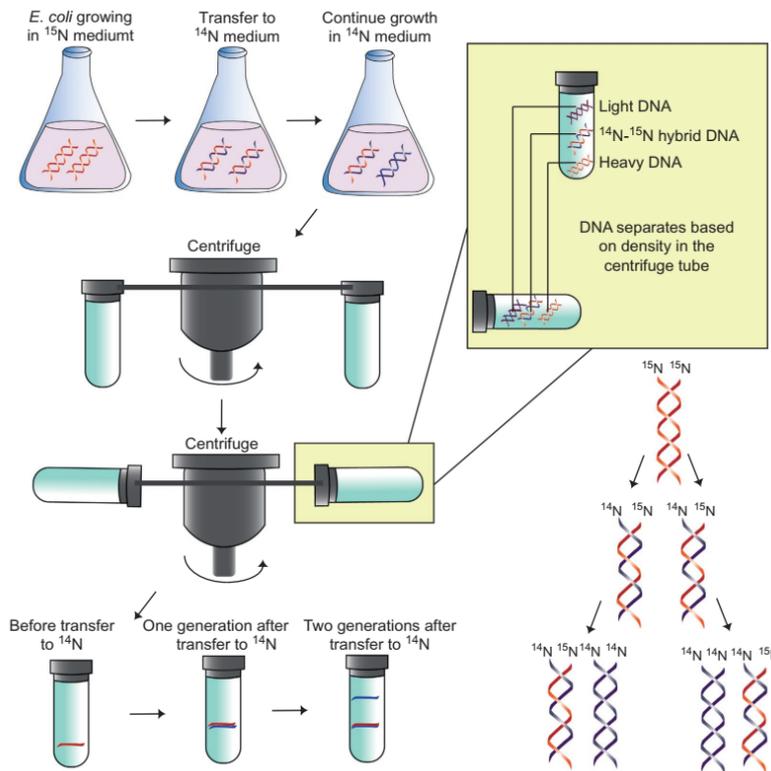


Figure 2.3 Trois hypothèses de la réplication de l'ADN (Brown, 2016).

La critique pour le succès de cette expérience, l'ADN contenant ¹⁵N peut être distingué de l'ADN contenant ¹⁴N. La procédure expérimentale implique l'utilisation d'une technique appelée centrifugation à l'équilibre de sédimentation ou centrifugation à gradient de densité flottant. Les échantillons sont forcés par centrifugation à travers un gradient de densité de chlorure de césium. Les molécules d'ADN atteindront l'équilibre lorsque leur densité sera égale à la densité du milieu de gradient. En tant que tel, l'ADN-¹⁵N atteindra le point à une position plus proche du fond du tube que l'ADN-¹⁴N.

Leurs résultats ont montré qu'il existe une bande d'ADN représentant [¹⁵N/¹⁴N] dans le gradient pour les cellules incubées dans le milieu contenant du ¹⁵NH₄Cl après une génération d'incubation dans le milieu contenant du ¹⁴NH₄Cl. En tant que tel, ce résultat exclut l'hypothèse conservatrice qui devrait avoir deux bandes.

En deux cycles de réplication, le mécanisme de dispersion donnerait un produit hybride avec un quart [¹⁵N] et trois quarts [¹⁴N]. D'autre part, l'hypothèse semi-conservatrice prédit que [¹⁵N/¹⁴N] se séparerait et servirait de matrice pour la réplication et produirait des ADN [¹⁵N/¹⁴N] et [¹⁴N/¹⁴N]. Cela devrait donner un rapport de 1:1 pour ces deux molécules d'ADN. Après deux générations d'incubation dans le milieu contenant ¹⁴NH₄Cl, Meselson et Stahl ont observé qu'il existe deux bandes dans le gradient dont l'une représente [¹⁵N/¹⁴N] et

l'autre représente [$^{14}\text{N}/^{14}\text{N}$]. Par conséquent, cette expérience prouve que la réplication de l'ADN suit le mécanisme semi-conservateur.

2.2 Le processus de réplication de l'ADN

2.2.1 Présentation des trois phases différentes de la réplication

Le processus de réplication de l'ADN peut être divisé en trois phases : initiation, élongation et terminaison. La réplication de l'ADN commence à des sites spécialisés appelés origines de réplication et s'éloigne d'une origine dans les deux sens, créant une structure connue sous le nom de bulle de réplication (Fig. 2.4). La double hélice d'ADN est ouverte aux origines de la réplication et déroulée des deux côtés de l'origine pour former deux structures appelées fourches de réplication qui déroulent la double hélice dans des directions opposées. Les fourches de réplication sont les sites sur lesquels l'ADN simple brin est exposé et sur lesquels se produit la synthèse de l'ADN.

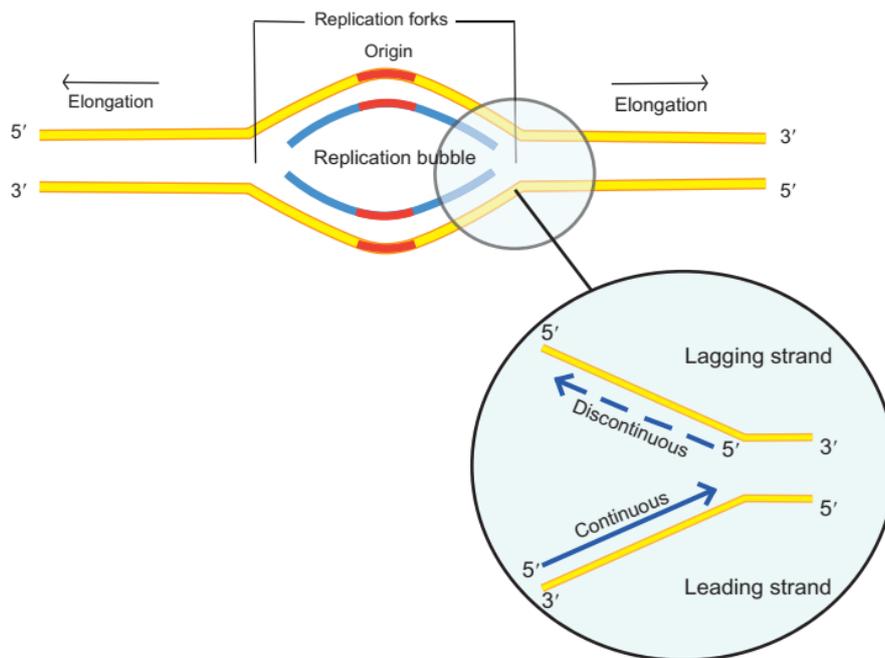


Figure 2.4 Réplication bidirectionnelle de l'ADN.

L'origine (rouge) indique le site d'initiation de la réplication de l'ADN. Les deux brins sont copiés de l'origine vers l'extérieur, créant la bulle de réplication. Des fourches de réplication se forment à chaque extrémité de la bulle et déroulent en continu la double hélice. Le brin principal est toujours le brin synthétisé dans le sens 5' vers 3'. L'autre est considéré comme le brin en retard et est répliqué de manière discontinue sous forme de fragments qui sont ensuite joints pour former un brin continu (voir la ligne pointillée) (Buckingham, 2011).

Une fois la double hélice d'ADN ouverte, les enzymes de réplication et les protéines sont chargées sur le simple brin, et celles-ci formeront les matrices des brins filles à synthétiser (Fig. 2.5). Une fois que la machinerie de réplication est en place et que l'ADN a

été ouvert, la réplication entre dans la phase d'élongation. Pendant la phase d'élongation, la machinerie de réplication se déplace le long des brins d'ADN parents et forme les brins filles au fur et à mesure. La résiliation de la réplication d'ADN se produit lorsque les deux fourches de réplication se déplaçant dans des directions opposées se rencontrent et que les complexes de réplication sont désassemblés. Chez les bactéries, cela se produit sur un site spécifique appelé *ter*, alors que chez les eucaryotes, aucune séquence de terminaison spécifique n'est requise.

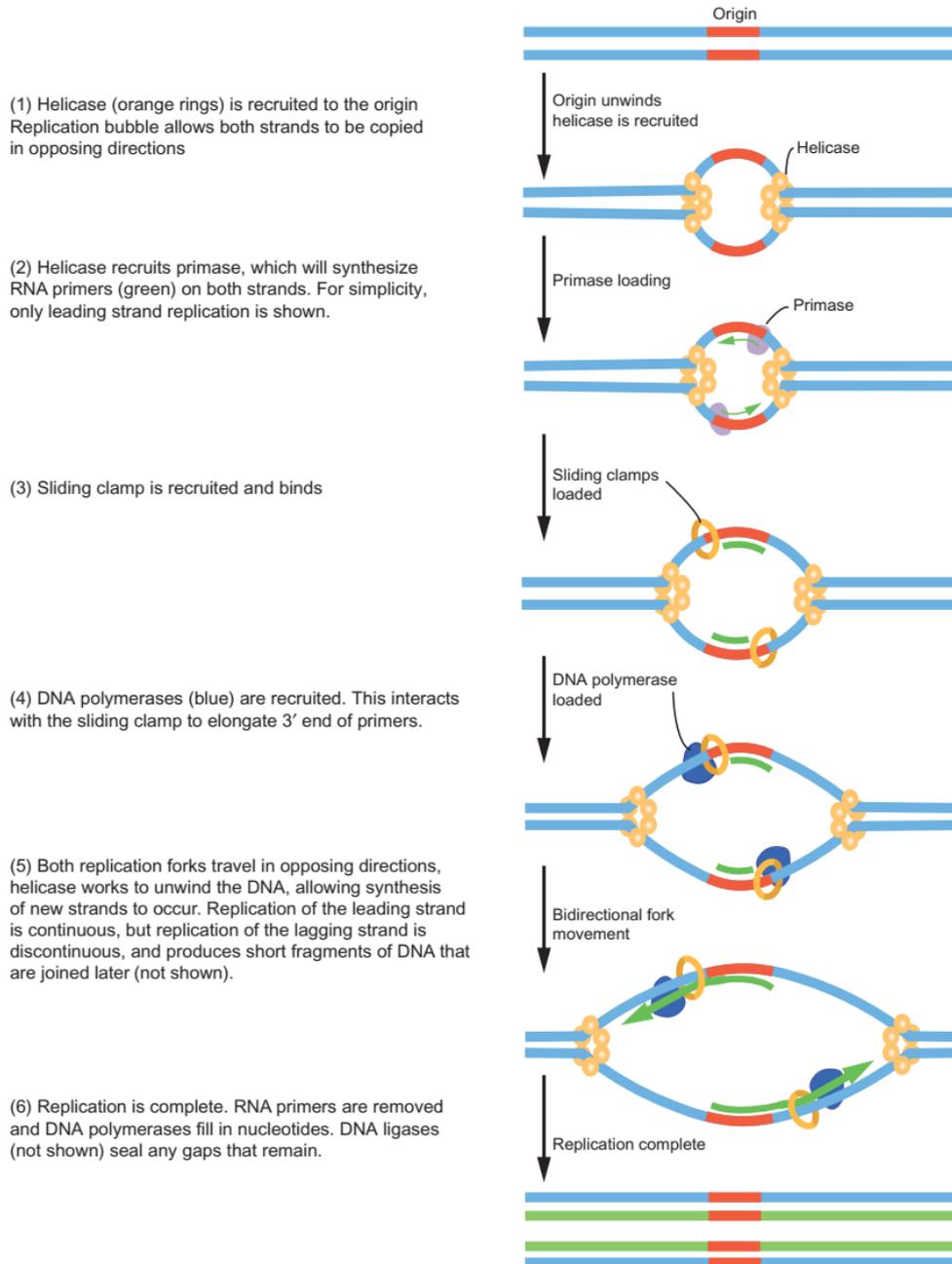


Figure 2.5 Processus de réplication de l'ADN (Weaver, 2011).

2.2.2 Mécanisme de réplication semi-discontinue de l'ADN

En raison de la nature de la réaction de synthèse de l'ADN, toute la synthèse des chaînes de nucléotides se produit dans la direction $5' \rightarrow 3'$. Dans une chaîne en croissance, le nucléotide entrant à un 5'-triphosphate sur son sucre. Le dernier nucléotide ajouté à une chaîne en croissance possède un groupe 3'-hydroxyle sur le sucre. Parce que ce groupe 3-hydroxyle est un nucléophile, il attaque le groupe phosphate α , entraînant l'élimination du pyrophosphate et la formation d'une nouvelle liaison phosphodiester, qui relie deux nucléotides ensemble et permet la synthèse de brins d'ADN (Fig. 2.6). Cette réaction est catalysée par l'ADN polymérase. Toutes les ADN polymérases connues jusqu'à présent ne peuvent ajouter des nucléotides que dans une direction $5' \rightarrow 3'$. En tant que telle, la réplication de l'ADN ne peut se produire que dans les années $5' \rightarrow 3'$ directions.

Pendant la réplication, l'ADN à double hélice se déroulerait pour créer une fourche. En raison de la nature antiparallèle des deux brins d'ADN, si l'on en court $5' \rightarrow 3'$ direction de gauche à droite, l'autre doit courir $3' \rightarrow 5'$ direction de gauche à droite. Par conséquent, la synthèse des brins se produit dans des directions opposées pour les deux matrices en double hélice. Suite au fait que la synthèse d'ADN ne peut faire qu'un sens $5' \rightarrow 3'$, Okazaki et ses collègues ont proposé un mécanisme de réplication semi-discontinue. Dans ce modèle, les deux brins ne pouvaient pas se répliquer en continu. L'ADN polymérase pourrait fabriquer un brin qui est le brin principal en continu dans les $5' \rightarrow 3'$ direction à la fourche de réplication sur le $3' \rightarrow 5'$ brins de gabarit (Fig. 2.7). Sa direction de synthèse est la même que la direction dans laquelle se déplace la fourche de réplication.

L'autre brin, qui est le brin en retard, devrait être fabriqué de manière discontinue en petits fragments - fragments d'Okazaki. Ces fragments sont généralement longs de 1 000 à 2 000 nucléotides, comme le montre la figure 2.7. La discontinuité de synthèse du brin en retard est due au fait que sa direction de synthèse est opposée à la direction de déplacement de la fourche de réplication. Les extrémités $5'$ de chacun de ces fragments d'Okazaki est plus proche de la fourche de réplication que l'extrémité $3'$. Les petits fragments d'Okazaki du brin retardé sont ensuite liés entre eux par une enzyme appelée ADN ligase, et ensemble, ils constitueront le deuxième brin fille.

2.2.3 Réplication de l'ADN eucaryote

La réplication de l'ADN eucaryote est plus complexe que l'ADN procaryote, en raison de plusieurs caractéristiques de l'ADN eucaryote. Par exemple, les cellules eucaryotes

contiennent beaucoup plus d'ADN et leur ADN est complexé avec des nucléosomes. De plus, les chromosomes eucaryotes sont linéaires plutôt que circulaires. Pour surmonter ces problèmes, la réplication à partir d'une seule origine sur un chromosome eucaryote typique prendrait des jours. Cependant, la réplication de génomes eucaryotes entiers est généralement accomplie en quelques minutes à quelques heures. Pour faciliter la synthèse rapide de grandes quantités d'ADN, les chromosomes eucaryotes contiennent de multiples origines de réplication. Les origines de réplication eucaryotes agissent non seulement comme des sites d'initiation de la réplication, mais contrôlent également le moment de la réplication de l'ADN (Klug et al, 2011).

2.3 Protéines nécessaires à la réplication de l'ADN chez *E. COLI*

La réplication de l'ADN est réalisée dans tous les organismes par un complexe multiprotéique appelé réplisome. La réplication de l'ADN commence au réplicon, où les deux brins d'ADN se déroulent à la fourche de réplication. Par conséquent, nous partons de la séparation de deux brins d'ADN pour discuter de la participation de chaque enzyme et protéine à chaque étape du processus.

2.3.1 Hélicase et protéines de liaison à l'ADN simple brin

Au cours de la phase d'initiation, l'ADN parental en double hélice est ouvert afin de donner aux enzymes de réplication et à d'autres protéines l'accès aux simples brins qui formeront la matrice des brins filles à synthétiser. En fait, une protéine initiatrice spécifique peut reconnaître l'origine de la réplication et recruter l'hélicase à l'origine de la réplication. L'hélicase est l'enzyme qui exploite l'énergie chimique de l'hydrolyse de l'ATP pour séparer les deux brins d'ADN à la fourche de réplication (Fig. 2.8). Lorsque les deux brins sont séparés, les protéines de liaison à l'ADN simple brin (SSB) se lient sélectivement à l'ADN simple brin dès sa formation. La liaison des SSB peut stabiliser l'ADN simple brin afin qu'il ne recuit pour reformer la double hélice et protéger l'ADN simple brin de l'hydrolyse par les nucléases. Cela permet également aux enzymes de se fixer au brin unique nouvellement ouvert et d'initier l'allongement. Les SSB sont éliminés au fur et à mesure de la synthèse de l'ADN.

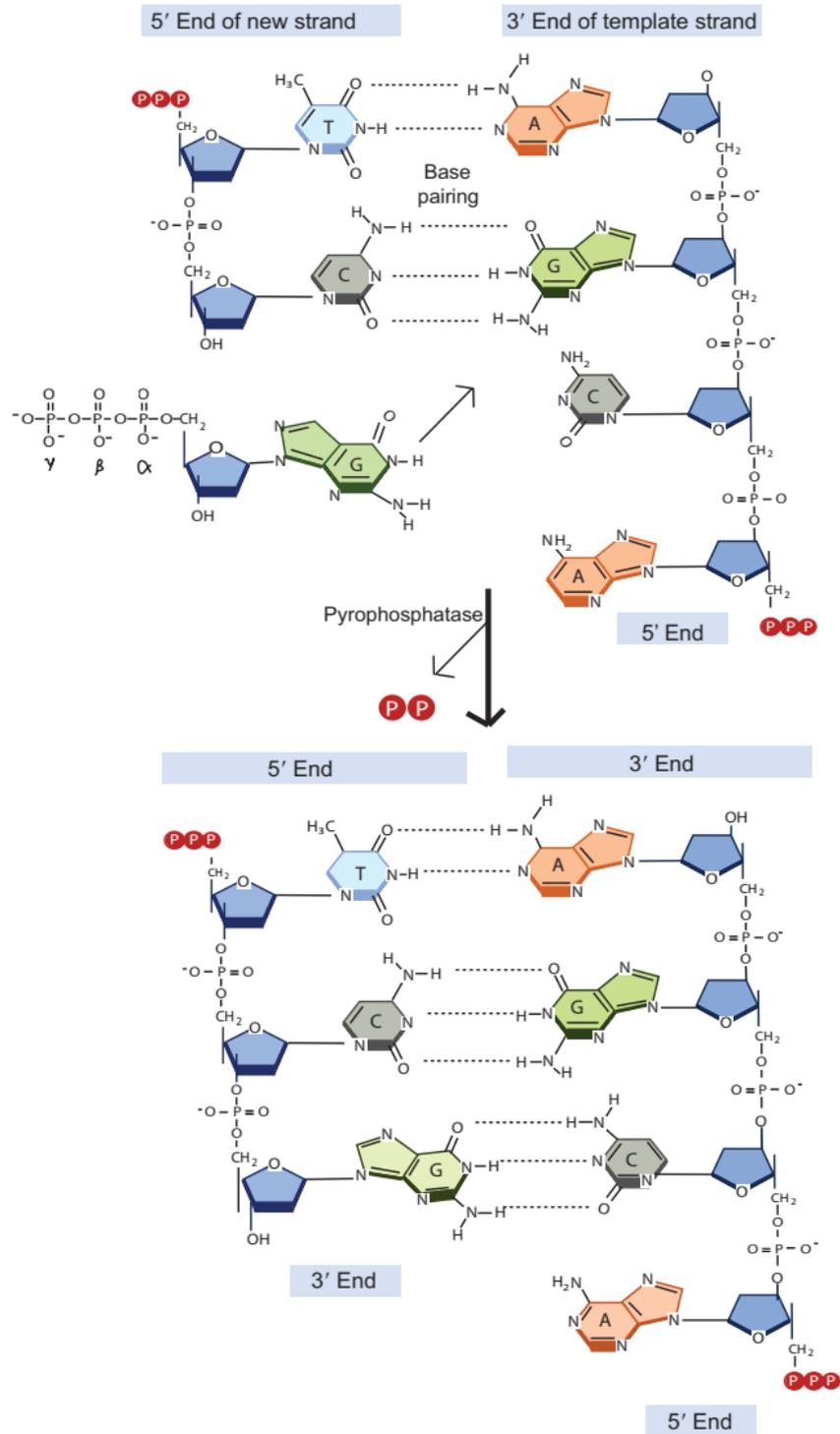


Figure 2.6 Synthèse d'ADN se produit toujours de 5' à 3'.
 Le brin matrice dirige l'ajout du dNTP approprié au brin en croissance. La synthèse de l'ADN nouvellement répliqué commence une fois que l' α -phosphate du dNTP entrant, qui allongera l'extrémité 3' et libérera un pyrophosphate, qui décompose le pyrophosphate en deux molécules de phosphate (Weaver, 2011).

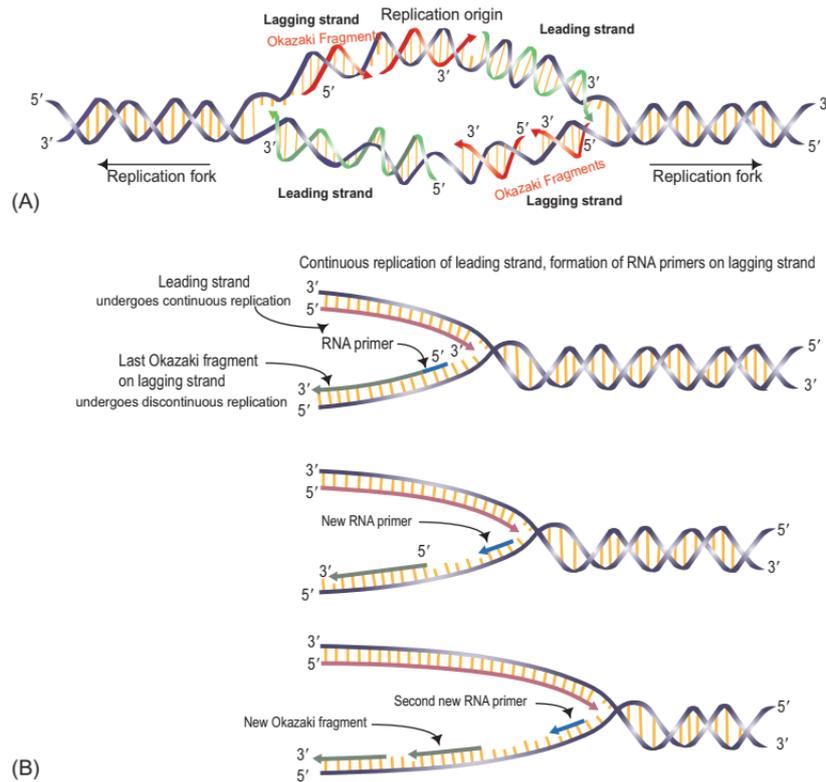


Figure 2.7 Mécanisme de la réplication semi-discontinue de l'ADN.

La réplication de l'ADN se produit de manière bidirectionnelle. (A) Deux fourches de réplication se forment à l'origine de la réplication et se déplacent dans des directions opposées. (B) Le schéma ne montre qu'une seule direction de réplication. La réplication du brin principal est continue, mais la réplication du brin retardé est discontinue en ce sens qu'il forme de courts fragments d'ADN qui sont ensuite scellés ensemble. Dans les deux brins, la réplication se produit toujours 5' à 3' (Brown, 2016).

2.3.2 ADN gyrase

La gyrase est une enzyme de la famille des ADN topoisomérases de classe II, qui est une protéine spécifique des bactéries, qui est essentielle à la réplication de leur chromosome circulaire. Chez les procaryotes, l'ADN existe sous une forme de cercle fermé négativement surenroulé. Lorsque deux brins d'ADN sont séparés lors de la réplication, des supercoils positifs sont introduits avant la fourche de réplication. En effet, les brins d'ADN duplex se déroulent pour permettre à chaque brin d'ADN de servir de matrice pour la synthèse d'un brin complémentaire ; un réplisome qui est empêché de tourner accumulera des superbobines positives devant la fourche de réplication (Fig. 2.9), tordant davantage l'ADN sur lui-même. Si la fourche de réplication continuait à se déplacer, la contrainte de torsion des superbobines positives finirait par rendre toute réplication impossible, car l'ADN serait trop compacté. À ce moment, l'ADN gyrase (topoisomérase de type II) combat ces superbobines positives en plaçant les superbobines négatives devant la fourche de réplication. Ainsi, la contrainte de torsion est libérée et la réplication peut continuer.

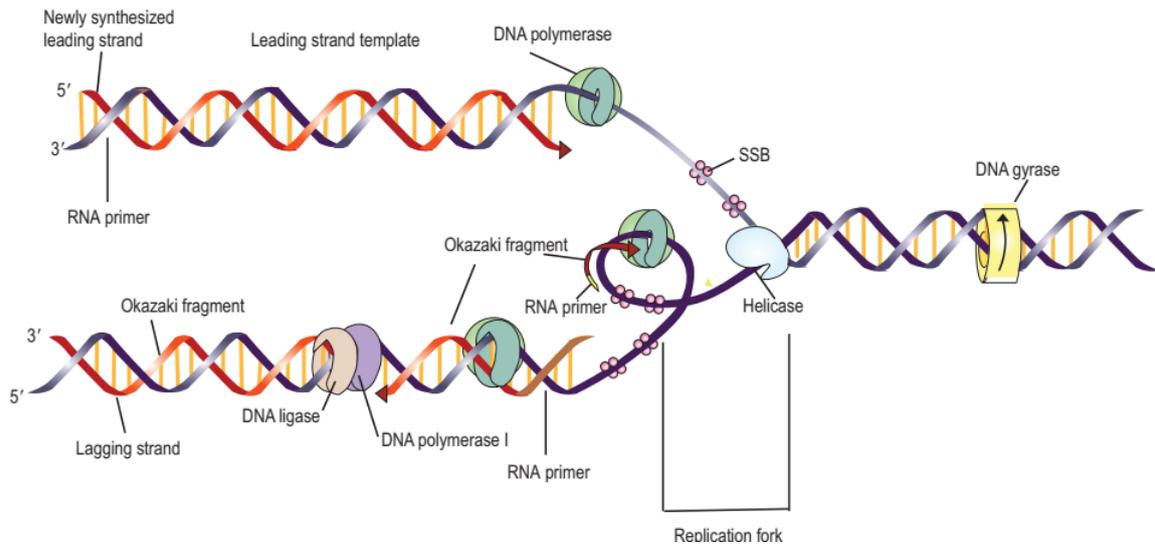


Figure 2.8 Caractéristiques générales du fork de réplication (Brown, 2016).

L'ADN gyrase et l'hélicase fonctionnent pour dérouler l'ADN afin qu'il puisse être répliqué. Les brins simples nouvellement déroulés sont recouverts d'une protéine de liaison à l'ADNss, SSB en abrégé. L'ADN polymérase répliquative est liée au brin matrice (avant ou arrière) via la sous-unité β . En aval, l'ADN polymérase II et l'ADN ligase éliminent les amorces d'ARN, les remplacent par les fragments d'ADN corrects et ligaturent également (scellent ensemble) les fragments d'Okazaki sur le brin en retard.

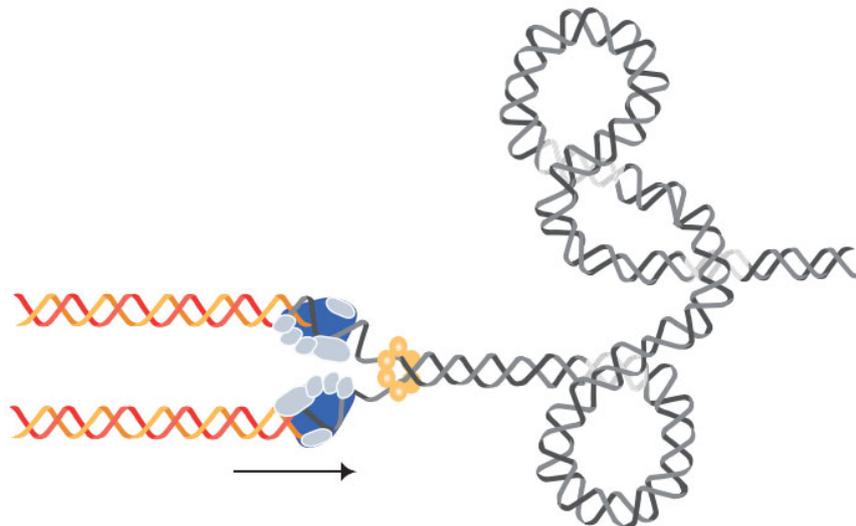


Figure 2.9 Superenroulement se forme en amont de la fourche de réplication.

Au fur et à mesure que l'hélicase déroule l'ADN, la torsion devant la fourche augmente en tension. Cela revient à tordre un élastique, puis à utiliser deux doigts pour séparer les brins. Le supertour est résolu par la topoisomérase I ou II, permettant le mouvement continu de l'hélicase (Weaver, 2011).

2.3.3 Primase

Une caractéristique inhabituelle de l'ADN polymérase est qu'elle ne peut pas synthétiser un nouveau brin d'ADN dès le début du brin parent. Il ne peut ajouter des nucléotides qu'à l'extrémité 3' d'un fragment de nucléotide qui existe déjà. Il est nécessaire d'avoir un court brin d'ARN appelé amorce pour le début de la réplication de l'ADN. L'amorce doit avoir un 3'-hydroxyle libre auquel la chaîne en croissance peut se fixer. Primase est responsable de la copie d'un court tronçon du brin de matrice d'ADN pour produire la séquence d'amorce d'ARN (Fig. 2.10). L'utilisation d'une amorce d'ARN pour la réplication de l'ADN est commune à tous les procaryotes. Le brin principal nécessite une amorce. D'autre part, chaque fragment d'Okazaki dans le brin retardé aurait besoin d'une amorce. L'amorce est liée à l'hydrogène à la matrice afin qu'elle puisse fournir un cadre stable dans lequel la chaîne naissante commence à se développer. Chez les bactéries, une courte amorce d'ARN d'environ 10 à 30 bases est synthétisée par une primase, suivie de la réplication dirigée par l'ADN polymérase III.

Le complexe primase eucaryote contient quatre sous-unités. Deux d'entre eux fonctionnent comme une primase, puis il y a une sous-unité catalytique α et une sous-unité accessoire. Les primases et la sous-unité catalytique α se lient dans un complexe avec l'ADN polymérase. Normalement, le complexe sous-unité polymérase- α -primase synthétise une étendue de 10 à 30 nucléotides d'ARN. Par la suite, la sous-unité catalytique α continue de synthétiser une courte portion d'ADN avant que l'ADN polymérase ne prenne en charge le processus de réplication. Ce phénomène est appelé commutation de polymérase.

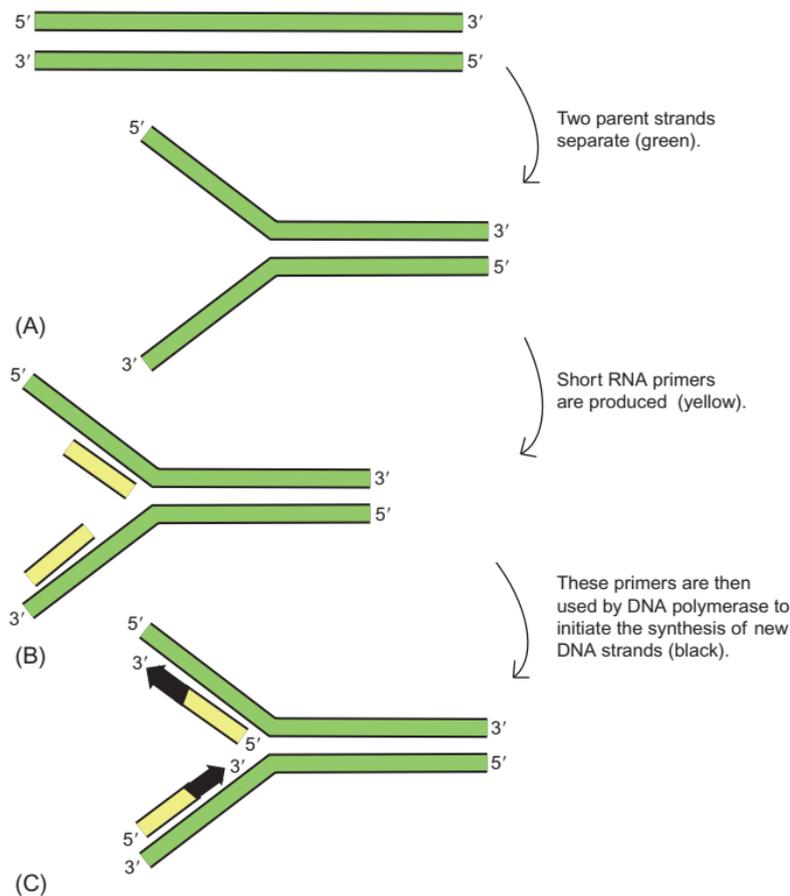


Figure 2.10 Amorces d'ARN sont formées avant la réplication de l'ADN (Weaver, 2011).

2.3.4 ADN polymérase

Une fois la réplication initiée, la synthèse de nouveau brin d'ADN est dirigée par l'une des polymérase -ADN polymérase III. Le premier nucléotide est ajouté au 3'-hydroxyle de l'amorce d'ARN, et la synthèse se déroule de l'extrémité 5' à l'extrémité 3' sur le brin avant et le brin en retard. Au fur et à mesure que la fourche de réplication se déplace, l'amorce d'ARN est retirée et remplacée par des désoxynucléotides nouvellement formés. Ces réactions sont réalisées par l'activité exonucléase et l'activité de polymérisation de l'ADN polymérase I. C'est ce qu'on appelle le processus de maturation d'Okazaki. Juste après l'élimination de l'ARN, l'ADN polymérase I utilise son activité polymérase pour combler le vide laissé par l'ARN avec un nouvel ADN. Lorsqu'elle atteint le début du segment d'ADN suivant, l'ADN polymérase ne peut pas sceller l'entaille finale dans le squelette d'ADN qui reste. Enfin, ADN ligase est l'enzyme responsable du scellement de la coupure entre les nouveaux brins synthétisés par la polymérase III et la polymérase I.

Trois ADN polymérase chez *E. coli* ont été étudiées de manière intensive (Tableau 2.1). Les trois ADN polymérase ont le 5'→3' activité de polymérisation et 3'→5' activité

exonucléase. Le 5'→3' L'activité de polymérisation peut ajouter des nucléotides à une chaîne en croissance lors de la synthèse de l'ADN. En revanche, le 3'→5' L'activité de l'exonucléase fait partie de la fonction de relecture qui peut supprimer les nucléotides incorrects un par un au cours du processus de réplication et les remplacer par le bon. Seule l'ADN polymérase I a l'activité exonucléase 5'→3' qui peut éliminer de courtes portions de nucléotides pendant la réparation. Une fois que la polymérase III a produit la nouvelle chaîne polynucléotidique, l'ADN polymérase I élimine l'amorce d'ARN jusqu'à 5'→3' activité exonucléase. Il vient alors se substituer derrière lui avec son activité polymérase (Fig. 2.11). Il s'agit d'un processus de coupure et de correction également appelé traduction de pseudo. Polymerase I utilise également la traduction de pseudo dans le processus de réparation.

Tableau 2.1 Propriétés de trois ADN polymérases différentes d'*E. coli*

Propriété	Pol I	Pol II	Pol III
Masse (kDa)	103	90	830
Gène structurel	PolA	polB	PolC
Polymérisation 5' → 3'	Oui	Oui	Oui
Exonucléase 5' → 3'	Oui	Non	Non
Exonucléase 3' → 5'	Oui	Oui	Oui
Nombre de sous-unités	1	Plus de 4	Plus que 10

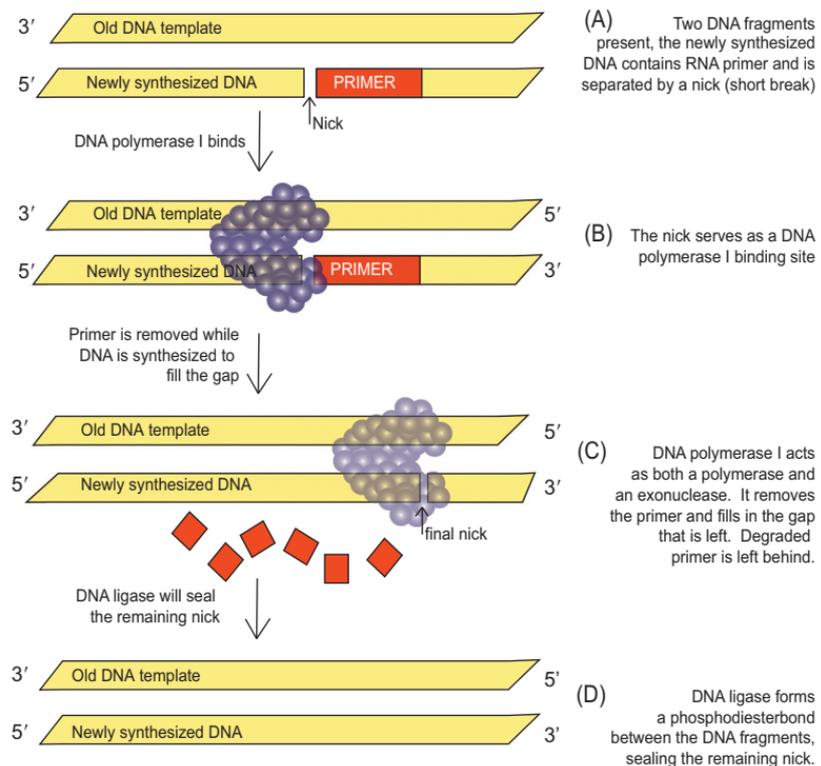


Figure 2.11 Amorces d'ARN sont retirées avant que les entailles d'ADN ne soient fixées (Weaver, 2011).

2.4 Coupler la réplication de l'ADN

Jusqu'à présent, nous avons discuté du fait que la synthèse du brin principal se déroule de manière continue et que la synthèse du brin retardé est discontinue en raison de la formation de fragments d'Okazaki. Le mouvement des polymérasés est coordonné pour permettre aux polymérasés des deux brins de se déplacer ensemble dans le sens du mouvement de la fourche. Par conséquent, si la réplication du brin principal s'arrête, la réplication du brin en retard sera également interrompue.

2.4.1 Le Réplisome

Le couplage de la réplication de l'ADN sur les brins avant et arrière des bactéries est réalisé en associant physiquement les protéines répliquant chaque brin en une grande protéine appelée complexe de réplication ou réplisome. Le réplisome s'associe à l'ADN, et ce complexe se déplace ensemble lorsque la fourche de réplication se déplace. Les protéines du complexe peuvent s'engager et se désengager de l'ADN lui-même, mais restent associées à la fourche. Chez *E. coli*, l'assemblage multi-sous-unités de l'ADN polymérase, de la pince coulissante et du chargeur de pince est appelé l'holoenzyme de l'ADN polymérase III ou le réplisome (Fig. 2.12). Il contient deux copies de l'ADN polymérase III multi-sous-unités, qui contient la sous-unité α (l'activité de l'ADN polymérase), la sous-unité ϵ (l'exonucléase de relecture 30-50) et la sous-unité θ (qui stimule l'exonucléase) avec des pinces coulissantes et un chargeur de pinces qui recharge en continu des pinces coulissantes sur le brin calorifuge.

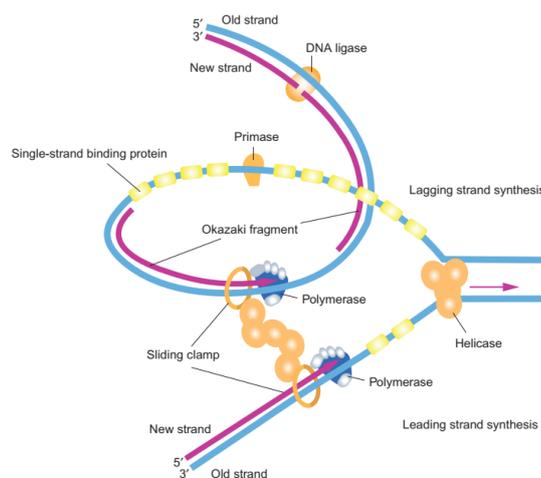


Figure 2.12 Brin principal et le brin retardé sont couplés ensemble au cours du processus de réplication chez *E. coli* (Klug et al, 2011).

La fourche de réplication se déplace vers la droite. Le brin avant est synthétisé en continu et le brin en retard est synthétisé en discontinu. Deux molécules d'ADN polymérase III (bleues) sont liées aux pinces coulissantes (anneaux orange) (Weaver, 2011).

2.4.2 L'Assemblée du Réplisome

Le chargeur de pince et la pince coulissante jouent un rôle important dans le recrutement de l'ADN polymérase à l'emplacement approprié sur la matrice d'ADN. Ils se localisent spécifiquement dans la région où la synthèse de l'ADN doit commencer (Fig. 2.5, étape 3). La pince coulissante est responsable du maintien des noyaux catalytiques sur leurs brins de gabarit. Le chargeur de pince place la pince sur l'ADN.

Une polymérase réplique le brin avant et une réplique le brin retard lorsque la fourche se déplace le long de la double hélice. Les polymérases sont liées entre elles par une protéine appelée tau (τ). La protéine Tau est associée au clamp loader et relie ce complexe polymérases-clamp loader à l'hélicase. Par conséquent, le complexe de réplication maintient la polymérase du brin retardé associée à la fourche même lorsqu'elle est libérée de l'ADN à la fin de chaque fragment d'Okazaki. De plus, grâce à cette association de polymérase, de chargeur de pince et de pince coulissante, le chargement d'une nouvelle pince et d'une polymérase sur l'ADN au début de chaque fragment d'Okazaki peut être efficacement coordonné avec la progression de la polymérase sur le brin principal.

2.5 Modèles de réplication

2.5.1 Réplication bidirectionnelle

La plupart des ADN eucaryotes et bactériens se répliquent de manière bidirectionnelle (Fig. 2.4). L'ADN sous le contrôle de l'origine de réplication est appelé un réplicon. Le chromosome circulaire d'*E. coli* a un réplicon unique car il se réplique à partir d'un point de départ unique, l'origine de la réplication. La réplication de l'ADN commence par la création d'une bulle, qui est une petite région où les brins parentaux se sont séparés et les brins descendants ont été synthétisés. Deux fourches naissent de cette bulle et se déplacent dans des directions opposées. Au fur et à mesure que la bulle se dilate, l'ADN qui se réplique commence à prendre la forme thêta (θ) jusqu'à ce que les deux fourches de réplication se rencontrent de l'autre côté du cercle (Fig. 2.13). Les chromosomes eucaryotes ont de nombreux réplicons et la réplication dans ces réplicons commence simultanément. Encore une fois, les fourches de réplication se déplacent dans les deux sens et continuent de se déplacer jusqu'à ce qu'elles atteignent l'extrémité du chromosome, appelée télomère. La réplication au niveau du télomère adopte un mécanisme spécial dont nous parlerons plus tard.

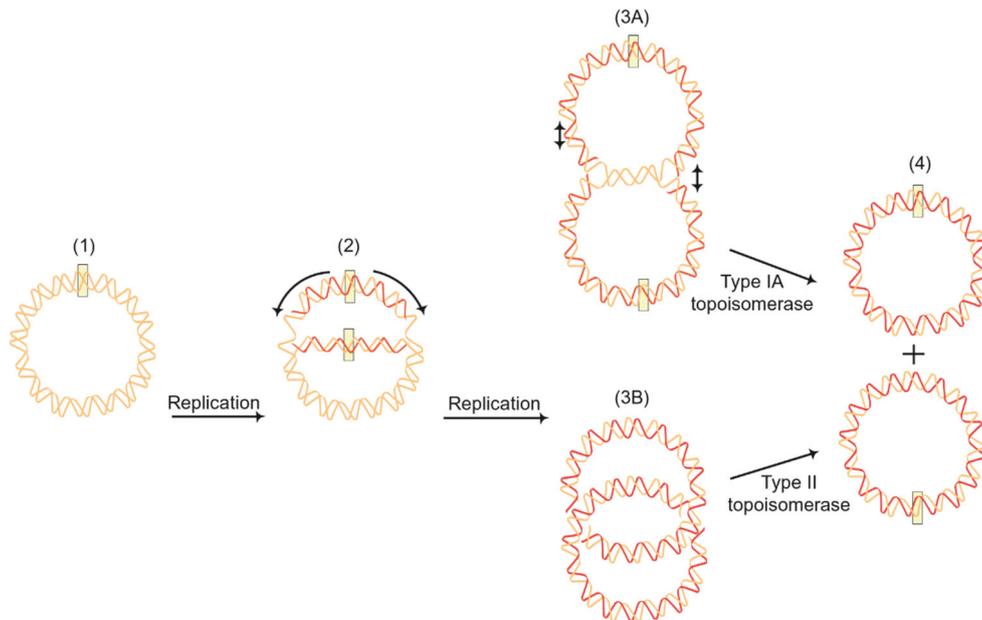


Figure 2.13 Processus de réplication de l'ADN bactérien. Deux molécules filles bicaténares interconnectées sont produites par réplication d'ADN circulaire (Weaver, 2011).

2.5.2 Réplication en cercle roulant

Certains ADN circulaires sont répliqués par un mécanisme de cercle roulant au lieu du mode θ (Fig. 2.14). La réplication par cercles roulants est courante chez les bactériophages. Le phage bactérien nommé ϕ X174 a un génome d'ADN circulaire simple brin et adopte une forme simple de réplication en cercle roulant. Fondamentalement, la réplication d'un seul brin est utilisée pour générer des copies de certaines molécules circulaires. Par exemple, une forme répliquative double brin donne lieu à de nombreuses copies d'un ADN de descendance simple brin. Un brin d'un ADN double brin est coupé et l'extrémité 3'-OH libre générée par la coupure est prolongée par l'ADN polymérase. La chaîne est allongée autour du gabarit de brin circulaire jusqu'à ce qu'elle atteigne le point de départ. Par la suite, le brin nouvellement synthétisé déplace le brin parental d'origine. Pendant ce temps, le brin déplacé est libre comme un cercle. Cela déplace la fin 5'. Pendant le processus de réplication, les intermédiaires qui ont la partie double brin de l'ADN répliquant peuvent être considérés comme roulant dans le sens antihoraire et traînant l'ADN monocaténaire de la descendance. Pendant la phase de réplication de l'infection virale, il peut être utilisé comme matrice pour synthétiser le brin complémentaire. Le cercle duplex peut alors être utilisé comme cercle roulant pour générer plus de descendance. Au cours de la morphogénèse du phage, le brin déplacé est encapsidé dans le virion du phage.

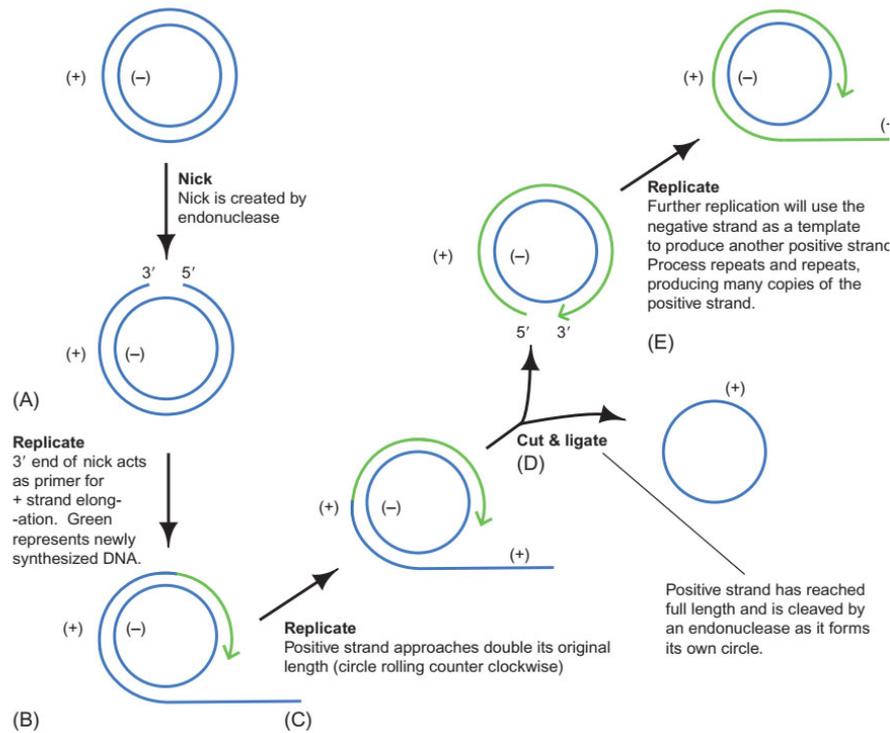


Figure 2.14 Diagramme schématique de la réplication en cercle roulant pour l'ADN simple brin (Weaver, 2011).

Le mécanisme du cercle roulant peut être utilisé pour répliquer l'ADN double brin. Par exemple, le phage λ adopte le mode de réplication θ pour produire plusieurs copies d'ADN circulaire (Fig. 2.15). Par la suite, ces ADN circulaires servent de matrice pour la synthèse en cercle roulant de molécules d'ADN λ linéaires. Le brin d'ADN intact sert de matrice pour la synthèse du brin principal, tandis que le brin déplacé sert de matrice pour la synthèse du brin retardé.

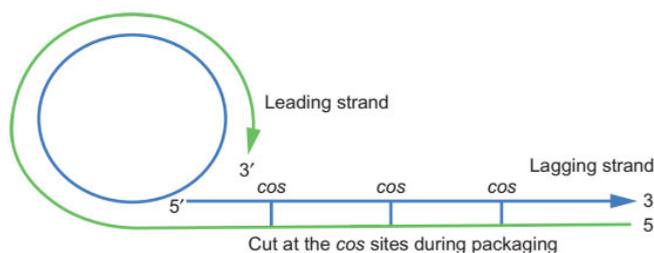


Figure 2.15 Diagramme schématique de la réplication en cercle roulant pour l'ADN double brin. Le brin avant s'allonge en continu, et le brin en retard s'allonge de manière discontinue, et utilise le brin avant déroulé comme gabarit. Les amorces d'ARN permettent à l'ADN polymérase de créer des fragments d'Okazaki. L'ADN db de la descendance contiendra de nombreuses longueurs du génome et sera coupé et emballé dans une tête de phage (Weaver, 2011).

2.5.3 Arrêt de la réplication

L'arrêt de la réplication de l'ADN se produit lorsque les deux fourches de réplication s'éloignant de l'origine unique de réplication se rencontrent du côté opposé du cercle dans le chromosome circulaire des bactéries (Fig. 2.16). Lorsque les fourches se rapprochent d'autre part, le complexe de réplication se désassemble et deux brins en croissance se rejoignent, ce qui entraîne l'interconnexion des deux nouveaux ADN filles. Les deux molécules d'ADN circulaires sont séparées par des topoisomérases. Les séquences impliquées dans la terminaison sont appelées sites ter. Un site ter contient une courte séquence de 23 pb. Les séquences de terminaison sont unidirectionnelles et peuvent être reconnues par une contrehélicase unidirectionnelle. Cette hélicase connue sous le nom de Tus chez *E. coli* peut reconnaître les séquences consensus et empêcher la fourche de réplication de se poursuivre. Chez les eucaryotes, la terminaison est terminée en plusieurs points. En effet, les chromosomes linéaires ont de multiples origines de réplication, dont chacune initie une bulle de réplication. Lors de l'élongation de réplication, les deux fourches de réplication forment des bulles différentes qui vont se rencontrer de la même manière que les deux fourches bactériennes. Aux fins de terminaison, les chromosomes interconnectés résultants seront dissociés par la topoisomérase (Fig. 2.17).

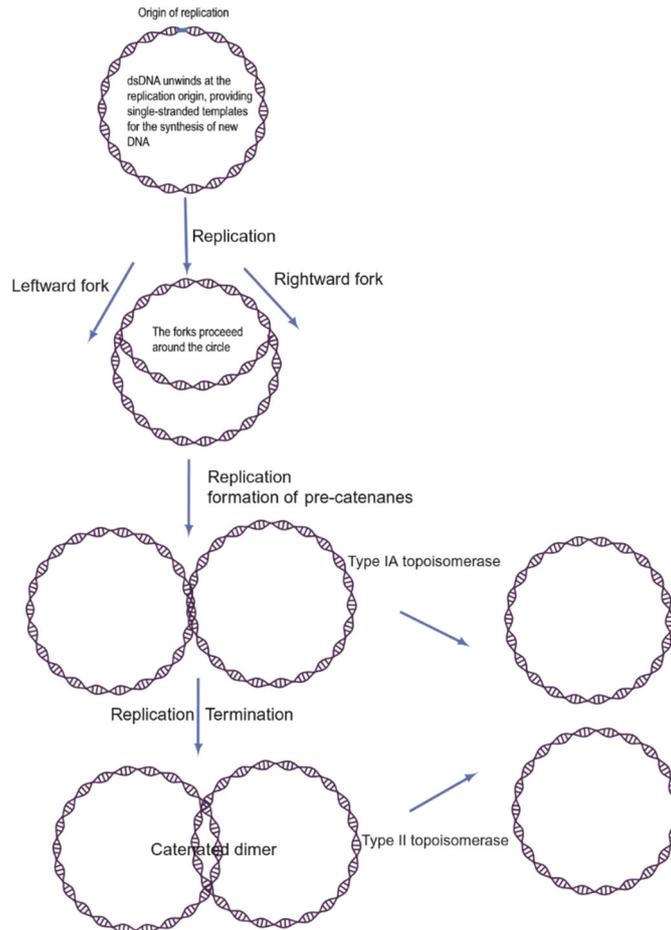


Figure 2.16 Processus de réplication circulaire de l'ADN aboutit à deux molécules filles liées ou caténées.

Le chromosome circulaire peut être non lié ou décaté par les topoisomérases.

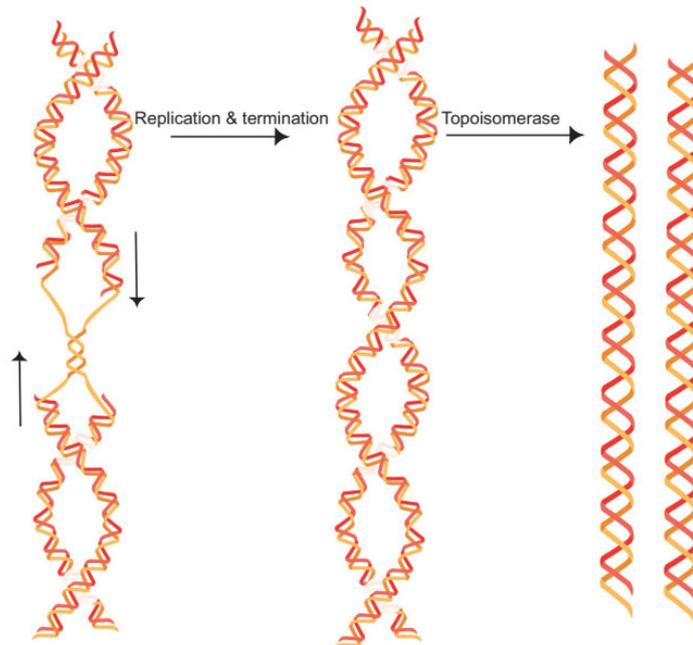


Figure 2.17 Réunion des fourches de réplication crée de l'ADN interconnecté sur des chromosomes linéaires (Weaver, 2011).

Lorsque deux fourches de réplication convergent, l'ADN entre elles est répliqué et les deux nouveaux brins filles s'entrelacent normalement. Celui-ci est fixé par la topoisomérase avant leur séparation.

2.5.4 Réplication au niveau des télomères

Les chromosomes eucaryotes se terminent par des séquences distinctes appelées télomères qui aident à préserver l'intégrité et la stabilité du chromosome. Les télomères sont constitués de séquences répétées simples. La longueur de l'ensemble des répétitions des télomères varie d'environ 100 bp à 20 000 bp, selon les espèces. Habituellement, l'ADN aux extrémités chromosomiques du protozoaire consiste en la courte séquence répétée en tandem TTGGGG, connue sous le nom de brin riche en G. D'autre part, son brin complémentaire affiche la séquence répétée AACCCC, appelée brin riche en C. De manière similaire, tous les vertébrés contiennent la séquence TTAGGG à l'extrémité du brin riche en G. Le brin riche en G est toujours à l'extrémité 3', ce qui a une signification particulière lors de la réplication des télomères.

Au cours de la réplication, la synthèse 5' à 3' sur la matrice du brin principal peut se poursuivre jusqu'à la fin. Cependant, il y a un problème sur le brin retardé lorsque l'amorce d'ARN est retirée. Parce que l'espace ne peut pas être comblé en l'absence du groupe 3'-OH, il y aura un espace à la fin de la réplication, qui laissera donc une région simple brin d'ADN. Avec l'aide de la télomérase, une ribonucléoprotéine, l'extrémité 3' du brin riche en G peut être allongée en utilisant l'ARN comme matrice pour ajouter des séquences télomériques à l'extrémité 3' de l'ADN télomérique (Fig. 2.18). Cela se traduit par un allongement net des télomères et permet de maintenir les télomères autour d'une longueur d'équilibre.

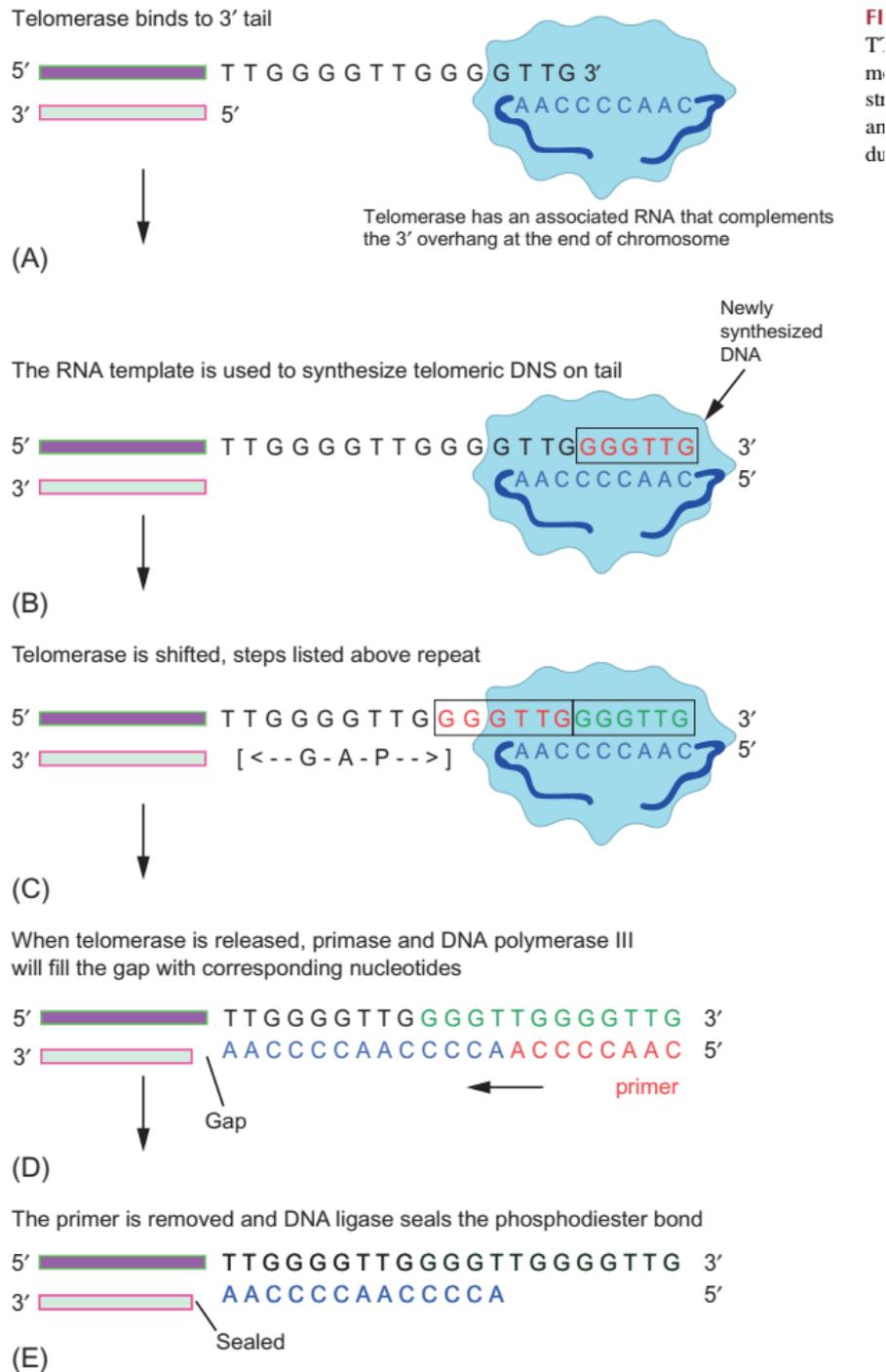


Figure 2.18 Télomérase se lie à 3'CG de queue riche (Brown, 2016).

Les séquences répétées de TTGGGG sont synthétisées par l'enzyme télomérase, facilitant la synthèse de l'ADN sur le brin opposé. Cela allonge également les extrémités des chromosomes et empêche la dégradation de matériel génétique important pendant la réplication.

La télomérase est une ADN polymérase inhabituelle car elle contient à la fois un composant protéique et un composant ARN essentiel. Dans le composant ARN, il possède une courte région matrice qui spécifie la séquence des répétitions de télomères qui sont ajoutées aux extrémités des chromosomes. D'autre part, le composant protéique synthétise l'ADN

télomérique à travers la matrice d'ARN. Dans les cellules souches humaines, la télomérase est active et les télomères sont maintenus. Dans de nombreux tissus humains, en revanche, la télomérase n'est pas aussi active que dans les cellules souches, ce qui entraîne un raccourcissement progressif des télomères.

2.5.5 Réplication mitochondriale

Les mitochondries doivent être dupliquées au cours du cycle cellulaire et séparées des cellules filles. Le processus à chaque étape de la duplication des mitochondries (y compris la réplication de l'ADN, la ségrégation de l'ADN dans les mitochondries dupliquées et la ségrégation des organites dans les cellules filles) est stochastique, ce qui signifie qu'il n'y a aucun contrôle sur les copies particulières qui sont répliquées. Elle est régie par une distribution aléatoire de chaque exemplaire. Les mitochondries se divisent en développant un anneau autour de l'organite qui se resserre pour le pincer en deux moitiés. Ce mécanisme est similaire aux principes de la division bactérienne. La combinaison des mécanismes de réplication et de ségrégation peut entraîner une attribution stochastique d'ADN à chacune des copies. Ainsi, la distribution des génomes mitochondriaux aux mitochondries filles ne dépend pas de leurs origines parentales.

2.6 Fidélité de la réplication de l'ADN

2.6.1 Mécanisme de relecture

La réplication de l'ADN a lieu une fois par génération dans chaque cellule. En tant que tel, il est essentiel que la fidélité du processus de réplication soit aussi élevée que possible pour éviter les mutations, qui sont des erreurs de réplication et peuvent entraîner des maladies. Pour la réplication du bactériophage et d'*Escherichia coli* en l'absence de réparation des mésappariements d'ADN et de stress environnemental externe, le taux d'erreur de substitution de bases de la machinerie de réplication in vivo est de l'ordre d'une fois toutes les 10^7 à 10^8 paires de bases. La réplication de l'ADN eucaryote est susceptible d'être au moins aussi précise. En règle générale, les erreurs de liaison hydrogène conduisent à l'incorporation d'un nucléotide incorrect dans une chaîne d'ADN en croissance une fois toutes les 10^5 paires de bases. À ce rythme, la réplication introduirait des erreurs dans un pourcentage important de gènes à chaque génération chez *E. coli*. Heureusement, l'activité exonucléase $3' \rightarrow 5'$ peut éliminer le nucléotide incorrect et la réplication reprend lorsque le nucléotide correct est ajouté. Les erreurs de réplication se produisent environ une fois toutes

les 10^9 à 10^{10} paires de bases après les résultats combinés de l'activité polymérase et de l'activité de relecture. En tant que telle, l'activité de relecture peut améliorer considérablement la fidélité de la réplication (Zaher et Green, 2009).

2.6.2 Activité polymérase exonucléase

L'ADN existant peut également être réparé par l'activité exonucléase polymérase I $5' \rightarrow 3'$. Par exemple, si une ou plusieurs bases ont été endommagées par un agent externe, ou si une incompatibilité a été manquée par l'activité de relecture, la polymérase I peut utiliser son processus de traduction de pseudo pour réparer une telle erreur d'ADN lorsqu'elle se déplace le long de l'ADN. En tant que telles, les polymérases peuvent réaliser une copie fidèle du modèle. À partir d'études structurales et biochimiques de la polymérase I, il est démontré que la fidélité découle à la fois des contraintes imposées à l'appariement des bases au site actif de la polymérase ainsi que de l'édition des paires de bases mésappariées au site actif de la $3'$ -exonucléase. Le site actif de la polymérase et le site actif de l'exonucléase sont généralement spatialement séparés sur la polymérase. L'avantage de cette disposition est que lorsqu'une base correcte est ajoutée, la polymérase passe simplement à la position de matrice suivante et procède à la polymérisation. En revanche, lorsqu'un nucléotide mésappariement est incorporé, la vitesse d'incorporation du nucléotide suivant par la polymérase est ralentie. A ce moment, l'exonucléase peut corriger l'erreur avant que la polymérase ne reprenne sa vitesse de polymérisation habituelle.

Le mécanisme par lequel le domaine exonucléase exerce sa fonction d'édition est basé sur le mécanisme de navette (Fig. 2.19). Le domaine exonucléase implique une compétition avec les sites actifs de la polymérase pour l'extrémité $3'$ du brin matrice de l'amorce et une navette rapide de l'extrémité de l'amorce entre eux. Le site actif de la $3'$ -exonucléase se lie à l'ADN simple brin, tandis que le site actif de la polymérase se lie à l'ADN duplex avec un rapport d'environ 1 à 10 pour une connexion correcte d'ADN duplex à paires de bases Watson-Crick. Les paires de bases mésappariées déstabilisent l'ADN duplex et améliorent ainsi la liaison des $3'$ ADN simple brin au site actif de l'exonucléase. De plus, la polymérisation est bloquée après l'incorporation de paires de bases mésappariées, probablement en raison d'une mauvaise orientation du groupe $3'$ -hydroxyle de l'extrémité de l'amorce sur laquelle le nucléotide suivant doit être ajouté. Encore une fois, ce blocage de la réaction de polymérisation sert à augmenter la probabilité d'excision par l'activité exonucléase (Yan et al, 2017).

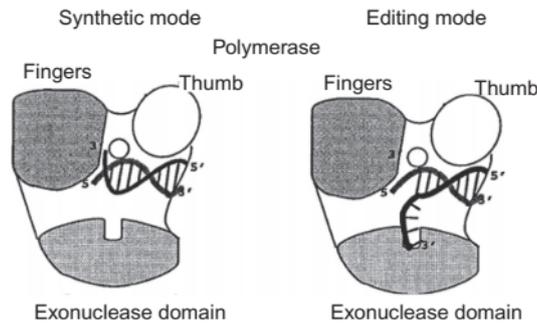


Figure 2.19 Mécanisme de navette de l'édition dans les ADN polymérases.

L'équilibre entre l'extrémité 3' du brin d'amorce étant lié en tant que brin unique dans le site actif de l'exonucléase (à droite) et lié en tant que duplex au site actif de la polymérase (à gauche) est déplacé vers le mode d'édition par des paires de bases non appariées, qui déstabilisent le duplex ADN et retarder l'addition du nucléotide suivant. La navette de l'extrémité 3' entre les deux sites actifs est rapide par rapport à la vitesse d'ajout de nucléotide suivant (Yan et al, 2017).

2.7 Mutation de l'ADN

2.7.1 Mutation ponctuelle

Des changements peuvent se produire dans la séquence nucléotidique d'une molécule d'ADN si les changements échappent à la relecture et à la réparation. Un tel changement génétique s'appelle une mutation. Les mutations sont classées selon le type de changement dans la molécule d'ADN. Des mutations ponctuelles surviennent lorsqu'une base s'apparie avec un partenaire inapproprié lors de la réplication de l'ADN. Les transitions, dans lesquelles une purine (ou pyrimidine) est remplacée par une autre, et la transversion, dans laquelle une purine est substituée à une pyrimidine ou vice versa, sont les deux types possibles de mutations ponctuelles. Un changement d'un nucléotide d'un triplet dans une région codant pour une protéine d'un gène peut entraîner la création d'un nouveau triplet qui code pour un acide aminé différent dans le produit protéique. Si cela se produit, la mutation est connue sous le nom de mutation faux-sens (Fig. 2.20). Un deuxième résultat possible est que le triplet est transformé en codon d'arrêt, ce qui entraîne l'arrêt précoce de la synthèse protéique. C'est ce qu'on appelle la mutation non-sens. La troisième possibilité est que la mutation ponctuelle modifie le nucléotide, mais que le triplet code toujours pour le même acide aminé ; c'est ce qu'on appelle la mutation silencieuse, car ici, il n'y a aucun effet sur le produit protéique final (Agrawal et Biswas, 2015).

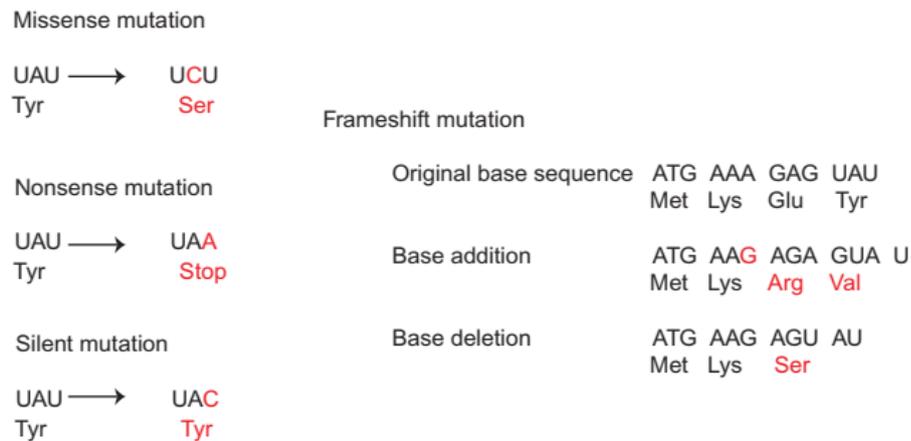


Figure 2.20 Différents types de mutations ponctuelles dans le cadre de lecture ouvert.

2.7.2 Insertion et suppression

En plus de la mutation ponctuelle, la réplication de l'ADN peut conduire à l'introduction de petites insertions ou délétions. L'ajout ou la suppression d'une ou plusieurs paires de bases conduit respectivement à des mutations d'insertion ou de délétion. Comme illustré dans la figure 2.20, la perte ou l'ajout d'un seul nucléotide entraîne la modification de tous les codons à trois lettres suivants. Celles-ci sont appelées mutations de décalage de cadre car le cadre de la lecture du triplet pendant la traduction est altéré. Une mutation de décalage de cadre se produira lorsqu'un nombre quelconque de bases sont ajoutées ou supprimées, à l'exception des multiples de trois, qui rétabliront le cadre de lecture initial et ajouteront simplement un autre cadre de lecture dans la séquence, sans modifier la composition de ces cadres de lecture par la suite. Il est possible, cependant, que le décalage de cadre provoque une terminaison précoce de la traduction, par exemple, si un codon d'arrêt de séquence de 3 nucléotides est introduit. Par conséquent, les résultats de la mutation par décalage de cadre peuvent être très graves s'ils se produisent au début de la séquence codante (Rajan-Babu et Chong, 2016).

2.7.3 Décalages tautomères

Le fait que les bases puissent prendre plusieurs formes, appelées tautomères, augmente le risque de mésappariement lors de la réplication de l'ADN. Les tautomères ne diffèrent que par un seul déplacement de proton dans la molécule. Les décalages modifient la structure de liaison de la molécule et permettent la liaison hydrogène avec des bases non complémentaires. Par conséquent, le changement tautomérique peut entraîner des modifications et des mutations permanentes des paires de bases (Fig. 2.21).

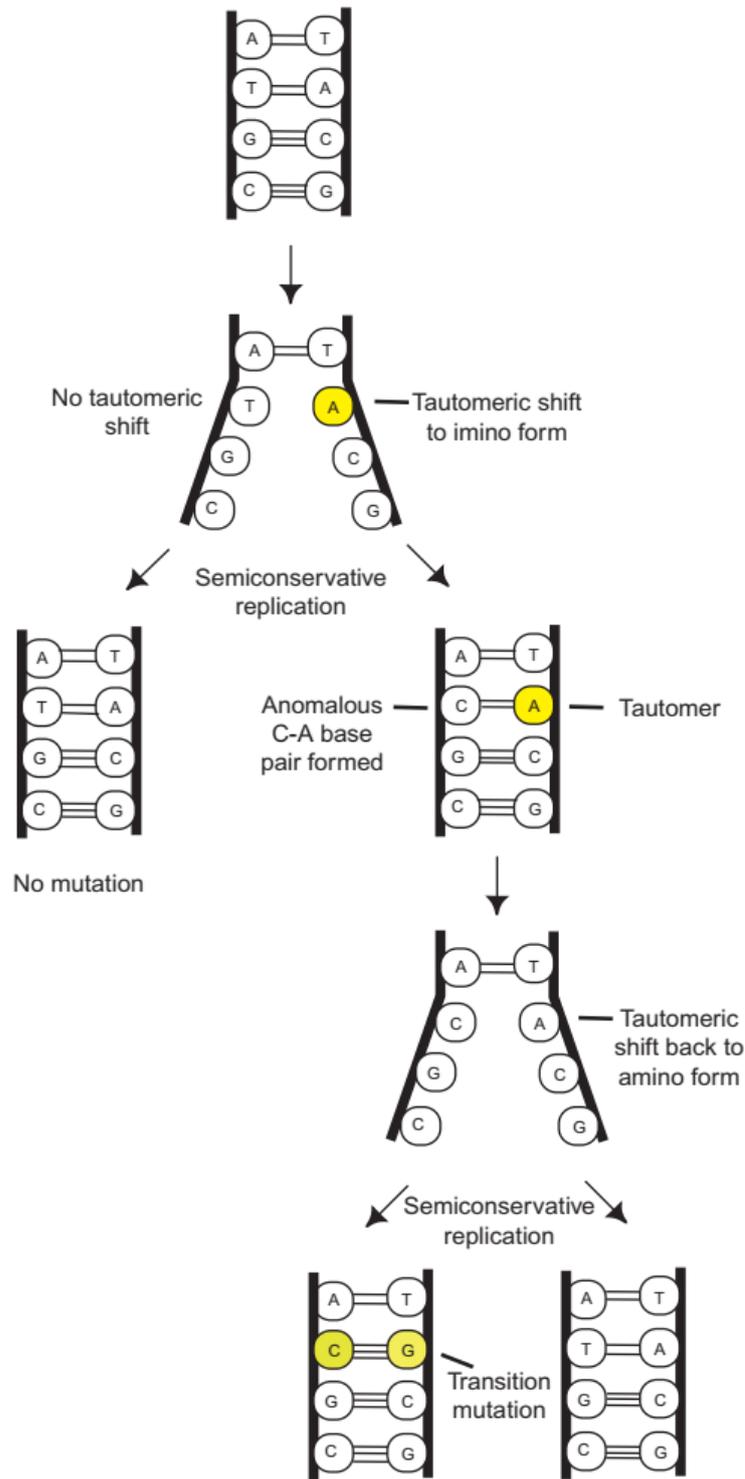


Figure 2.21 Exemple de formation de mutation.

Une mutation se produit pendant la réplication de l'ADN lorsqu'un tautomère formé de manière transitoire dans le brin matrice s'apparie avec une base non complémentaire. Lors du cycle de réplication suivant, les membres « mésappariés » de la paire de bases sont séparés, et chacun devient la matrice de sa base complémentaire normale et le résultat final est une mutation ponctuelle (Klug et al, 2011).

2.7.4 Réparation des incompatibilités

Lorsque le mésappariement a réussi à échapper aux activités normales d'exonucléase de l'ADN polymérase I et III, *E. coli* dispose d'un système de réparation des mésappariements pour corriger l'erreur. La clé du système de réparation des mésappariements est d'identifier le brin qui contient de l'adénine méthylée dans la séquence GATC en tant que brin parental, de sorte qu'il laisse ce brin seul et corrige le mésappariement à proximité dans le brin non méthylé. La voie de réparation des mésappariements dirigée par le méthyle d'*E. coli* repose sur le schéma de méthylation dans l'ADN pour déterminer quel brin est le brin nouvellement synthétisé et lequel était le brin parental (modèle). La méthylation de l'ADN, qui consiste à ajouter des groupes méthyle à certains nucléotides le long du nouveau brin d'ADN, se produit souvent juste après la réplication de l'ADN. Cependant, une fenêtre d'opportunité existe entre la fin de la réplication et le début de la méthylation où seuls les brins parentaux sont méthylés. Le système suppose que le brin portant la base méthylée est le brin parental et sa correcte séquence, et le brin sans méthylation est le nouveau brin. Par conséquent, cette fenêtre offre la possibilité au système de réparation des mésappariements d'examiner les brins nouvellement synthétisés pour toute erreur de réplication.

Dans ce processus, les enzymes MutH, MutS et MutL forment une boucle entre l'erreur et un site méthylé (Fig. 2.22). L'hélicase H de l'ADN aide à dérouler l'ADN. L'exonucléase I supprime la section d'ADN contenant l'erreur. Les protéines de liaison simple brin protègent le brin matrice de la dégradation. Enfin, la polymérase III peut combler la pièce manquante (Andersson et al, 2015).

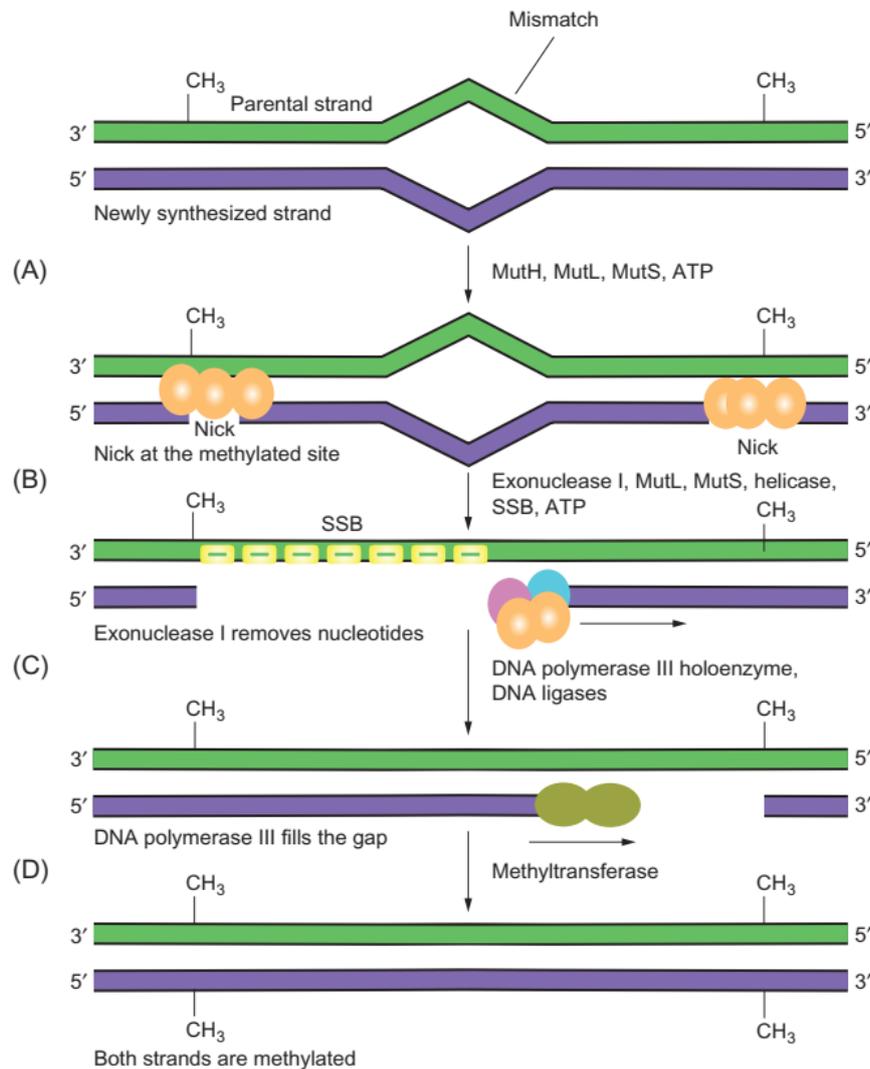


Figure 2.22 Diagramme schématisé de la réparation des mésappariements dans *E. coli*. (A) Les produits des gènes représentés combinés à l'ATP reconnaîtront les nucléotides incompatibles (bande surélevée). Ils utilisent l'absence de groupes méthyle pour identifier le brin nouvellement synthétisé (violet) et introduiront une entaille sur ce brin en face d'un groupe méthyle sur le brin d'origine (vert). (B) MutL, MutS, exonucléase I, hélicase, ATP supprimeront la région d'ADN sur ce brin (violet) qui abrite le nucléotide de la paire de mésappariement. (C) L'holoenzyme ADN polymérase III comblera le vide (avec l'aide de SSB). Ensuite, l'ADN ligase est utilisée pour combler l'écart restant. (D) La méthyltransférase méthyle les séquences GATC dans le brin fille (violet). Après cette action, les brins d'origine et fils ne pourront plus être distingués et, par conséquent, la réparation des décalages ne fonctionnera plus (Buckingham, 2011).

3 Expression génique : Transcription du code génétique

3.1 Le processus de transcription

3.1.1 Initiation

Le processus de transcription se décompose généralement en trois phases : initiation, élongation et terminaison. Il y a généralement quatre étapes dans l'initiation de la transcription :

- (1) formation d'un complexe promoteur fermé,
- (2) conversion du complexe promoteur fermé en un complexe promoteur ouvert,
- (3) polymérisation des premiers nucléotides (jusqu'à 10) tandis que la polymérase reste au niveau du promoteur, et
- (4) la clairance du promoteur, dans laquelle le transcrit devient suffisamment long pour former un hybride stable avec le brin matrice (Fig. 3.1).

Cela aide à stabiliser le complexe de transcription, et la polymérase s'éloigne du promoteur et la transcription commence l'étape d'élongation. Vous trouverez ci-dessous le détail de ces quatre étapes de la phase d'initiation.

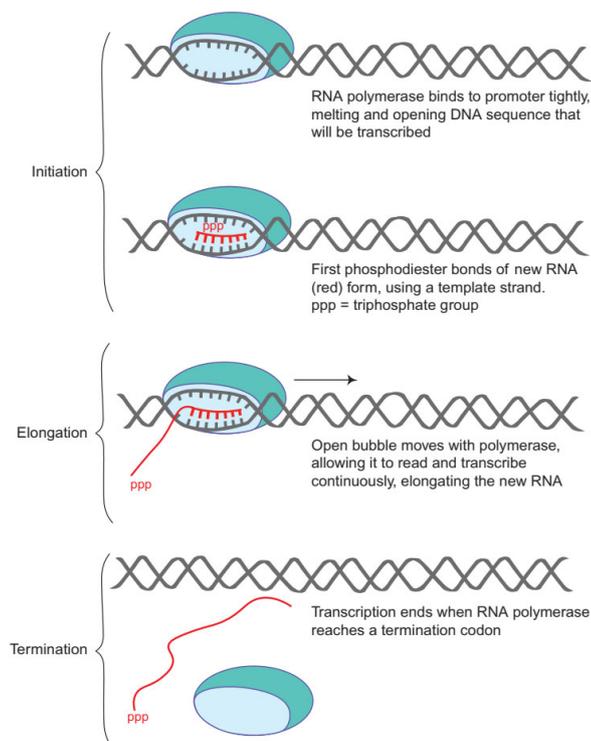


Figure 3.1 Diagramme schématisé de la liaison de l'ARN polymérase à l'ADN et de sa transcription pour produire un nouvel ARN simple brin (Brown, 2016).

L'initiation commence lorsque l'ARN polymérase se lie au promoteur et forme le complexe fermé. La sous-unité σ dirige la polymérase vers le promoteur en se liant à des séquences spécifiques en amont du site d'initiation de la transcription. Par la suite, les sous-unités β' et σ initient la séparation des deux brins d'ADN en aval, fondant environ 10 à 17 pb entourant le TSS, pour former la bulle de transcription complexe ouverte également connue sous le nom de bulle de transcription (Fig. 3.2).

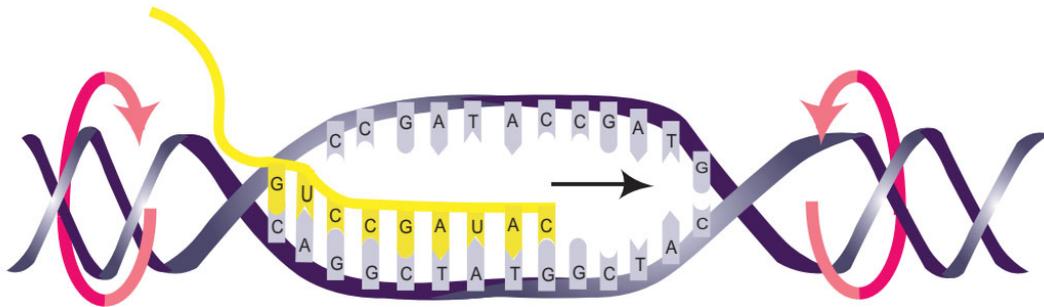


Figure 3.2 Exemple typique de bulle de transcription qui se compose de 14 bases non appariées, les 9 premières ont été utilisées pour transcrire un nouvel ARN simple brin (Buckingham, 2011).

Ensuite, la polymérase commence à construire la chaîne d'ARN. La première base de l'ARN est un ribonucléoside triphosphate de purine et A a tendance à apparaître plus souvent que G (Fig. 3.3). Une fois le premier nucléotide en place, la polymérase joint un deuxième nucléotide au premier, formant la liaison phosphodiester initiale dans la chaîne d'ARN. Les sous-unités β et β' sont impliquées dans la formation de liaisons phosphodiester ainsi que dans la liaison à l'ADN. Plusieurs nucléotides peuvent être joints avant que la polymérase ne quitte le promoteur et que l'élongation ne commence. Dans la bulle de transcription se trouvent les neuf ribonucléotides les plus récemment ajoutés du transcrit d'ARN, qui restent appariés en bases au brin matrice. L'association de la polymérase avec l'hybride ARN-ADN au sein de la bulle de transcription contribue à la stabilité du complexe d'élongation.

Toutes les cellules ont un facteur sigma primaire qui dirige la transcription à partir des promoteurs du gène codant pour les protéines essentielles nécessaires à la croissance des cellules. Ces gènes sont généralement appelés gènes de ménage. D'autre part, les bactéries ont également une variété de facteurs sigma alternatifs. Ces niveaux ou activités de facteurs sigma alternatifs sont régulés par la réponse à des signaux spécifiques ou à des conditions de stress. Par conséquent, ces facteurs sigma alternatifs dirigent la transcription des gènes qui ne sont nécessaires que dans certaines conditions. Nous discuterons plus tard du facteur sigma alternatif.

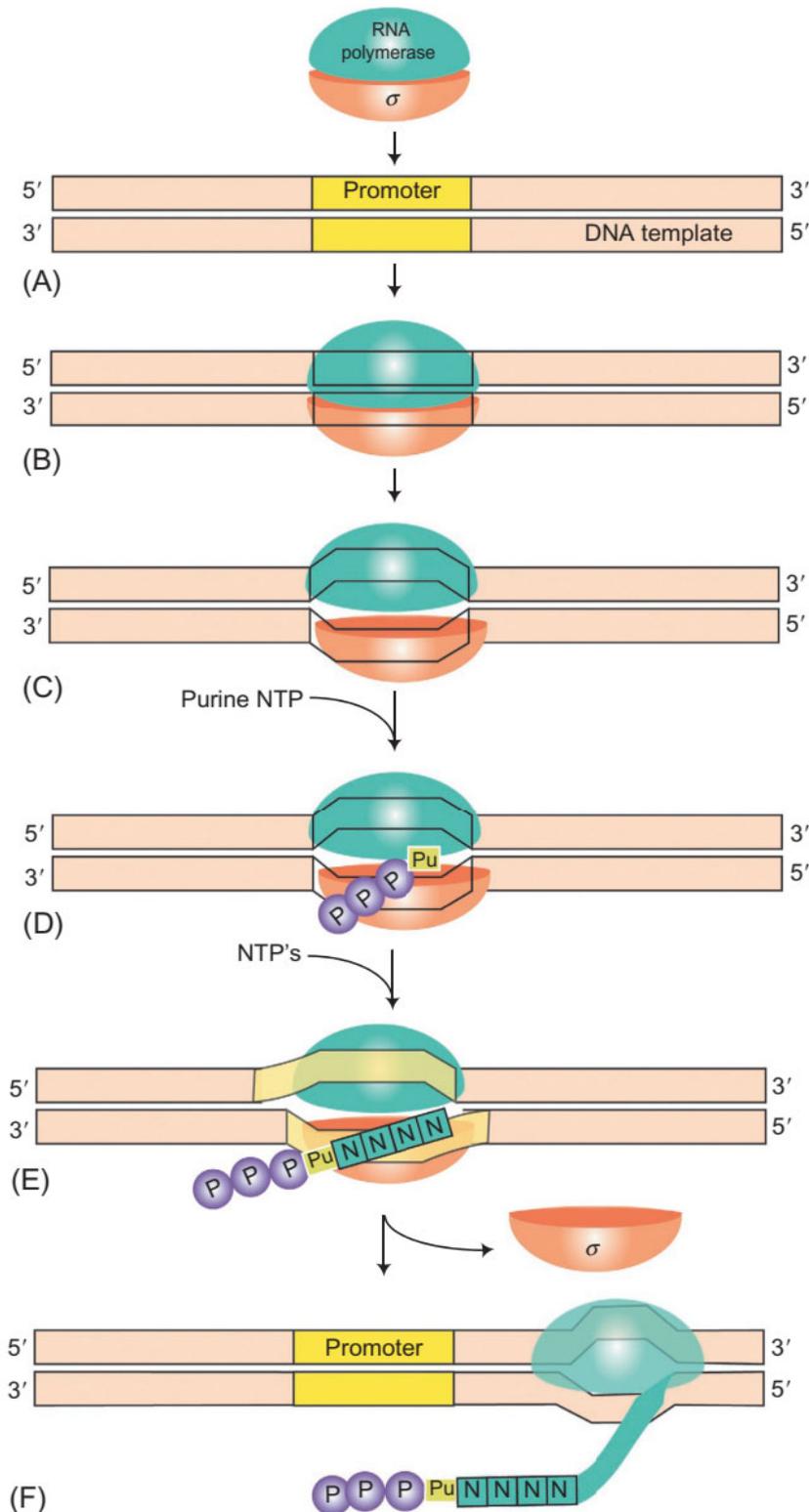


Figure 3.3 Étapes d'initiation et d'élongation de la transcription.

(A) La sous-unité Sigma de l'ARN polymérase reconnaît le promoteur et permet à l'enzyme de se lier à l'ADN. (B) Formation d'un complexe promoteur fermé ARN polymérase. (C) Au niveau du promoteur, l'ADN se déroule et forme un complexe promoteur ouvert. (D) L'ARN polymérase synthétise l'ARNm, commence presque toujours par une purine. (E) L'holoenzyme ARN polymérase catalyse l'allongement de l'ARNm d'environ 4 nucléotides. (F) La sous-unité sigma est libérée au fur et à mesure que le noyau procède à l'allongement du transcrit d'ARN.

3.1.2 Élongation

Pendant la phase d'élongation, l'ARN polymérase dirige la liaison séquentielle des ribonucléotides à la chaîne d'ARN en croissance dans la direction $5' \rightarrow 3'$, tandis que l'ARN polymérase et la bulle de transcription se déplacent le long de l'ADN matrice dans la direction $3' \rightarrow 5'$. Lorsque l'ARN polymérase se déplace le long de l'ADN matrice, la bulle de transcription se déplace également avec elle. Cette région fondue expose les bases de l'ADN matrice une par une afin qu'elles puissent s'apparier avec les bases des ribonucléotides entrants.

Lorsqu'environ 9 à 10 nucléotides ont été incorporés, la sous-unité σ se dissocie de l'holoenzyme et est ensuite recyclée pour se lier à une autre enzyme centrale pour un autre processus d'initiation. L'enzyme centrale continue d'allonger l'ARN, ajoutant un nucléotide après l'autre à la chaîne d'ARN en croissance. La sous-unité centrale β se trouve près du site actif de l'ARN polymérase où les liaisons phosphodiester sont formées et implique la formation de liaisons. Au fur et à mesure que l'ARN polymérase se déplace le long de l'ADN matrice, la polymérase maintient une courte région fondue d'ADN matrice. Cela nécessite que l'ADN se déroule devant la polymérase en progression et se referme derrière elle. Au cours de ce processus, un surenroulement positif est produit avant la bulle de transcription et un surenroulement négatif est produit derrière la bulle de transcription (Fig. 3.4). En tant que telles, les topoisomérases interviennent pour détendre les supercoils devant et derrière la bulle de transcription en progression. Dès que la machinerie de transcription passe, les deux brins d'ADN s'enroulent à nouveau, reformant la double hélice.

Le processus d'élongation est loin d'être uniforme et régulier. Au lieu de cela, l'ARN polymérase s'arrête fréquemment, voire revient en arrière, tout en allongeant une chaîne d'ARN. La pause est physiologiquement importante pour deux raisons : premièrement, elle permet à la traduction de suivre le rythme de l'ARN polymérase. Ceci est important pour l'atténuation et l'abandon de la transcription en cas d'échec de la traduction. L'atténuation est un mécanisme permettant de réguler l'expression en provoquant l'arrêt prématuré de la transcription de l'opéron lorsque les produits de l'opéron sont abondants. Le deuxième aspect important de la pause est qu'il s'agit de la première étape de la terminaison de la transcription. Le retour en arrière de l'ARN polymérase facilite la relecture en extrudant l'extrémité $3'$ de l'ARN hors de la polymérase, où les nucléotides mal incorporés peuvent être éliminés par une activité nucléase inhérente de la polymérase, stimulée par des facteurs auxiliaires.

3.1.3 Terminaison

La terminaison de la transcription de l'ARN implique des séquences spécifiques en aval du gène réel pour que l'ARN soit transcrit. Il existe deux types de mécanismes de terminaison - terminaison intrinsèque et rho (ρ) - terminaison dépendante. Pour la terminaison intrinsèque, les régions à l'extrémité des gènes appelées sites de terminaison ou de terminaison peuvent signaler la terminaison de la transcription. Les sites de terminaison sont caractérisés par deux répétitions inversées espacées de quelques autres bases (Fig. 3.5). L'ADN code alors une série d'uraciles. Lorsque l'ARN est créé, les répétitions inversées forment une boucle en épingle à cheveux. Cela tend à bloquer l'avancement de l'ARN polymérase. Pendant ce temps, la présence des uraciles provoque une série de paires de bases AU entre le brin matrice et l'ARN. Parce que les paires de bases AU entre le brin matrice et l'ARN sont faiblement liées à l'hydrogène par rapport aux paires GC, l'ARN se dissocie de la bulle de transcription et met fin à la transcription.

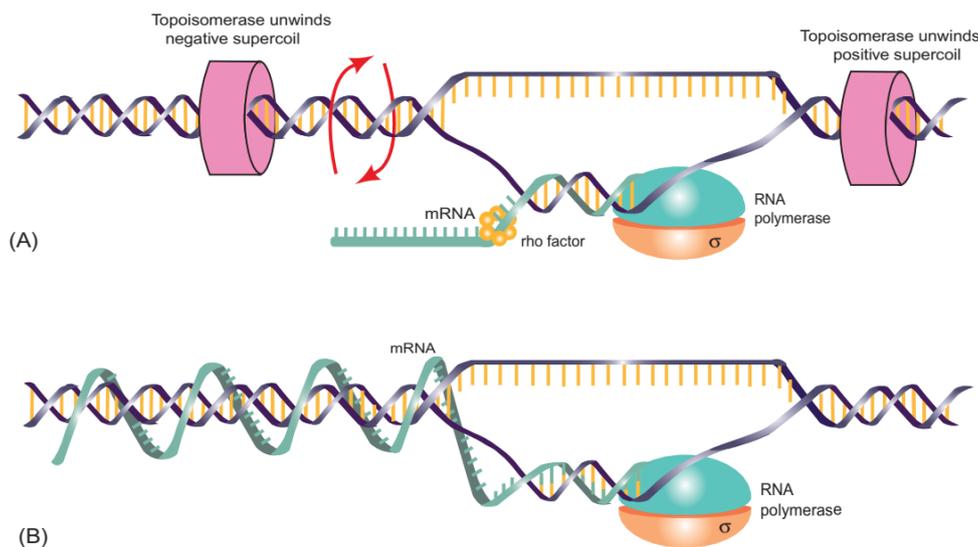


Figure 3.4 Deux modèles d'élongation de la transcription.

(A) La topoisomérase détend à la fois le superenroulement positif (avant la bulle de réplication) et négatif (après) de l'ADN. (B) L'ARN polymérase suit le brin matrice autour du duplex, de sorte qu'il peut éviter tout surenroulement de l'ADN (Buckingham, 2011).

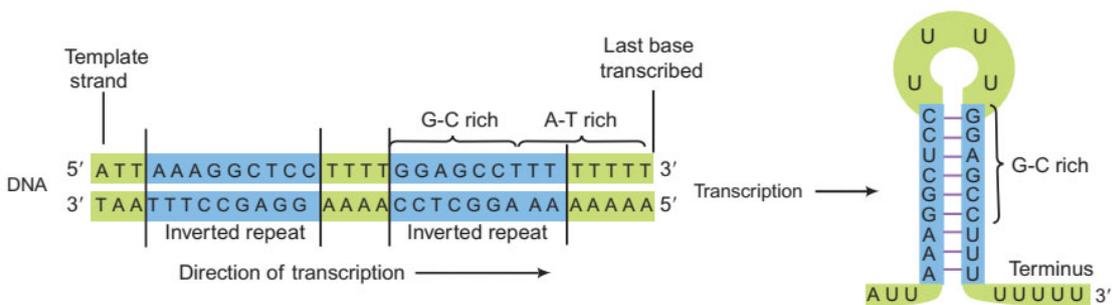


Figure 3.5 Caractéristiques intrinsèques de la séquence de terminaison - répétitions inversées et une série d'uraciles (Buckingham, 2011).

La terminaison ρ -dépendante implique une répétition inversée, elle provoque donc également la formation d'une boucle en épingle à cheveux mais pas de chaîne de U. Dans ce cas, la protéine ρ se lie à l'ARN et chasse la polymérase (Fig. 3.6). Lorsque la polymérase transcrit l'ARN qui forme une boucle en épingle à cheveux, elle se bloque, donnant à la protéine σ l'occasion de rattraper son retard. Lorsque la protéine σ atteint le site de terminaison, elle facilite la dissociation de la machinerie de transcription. Le résultat de la dissociation termine la transcription.

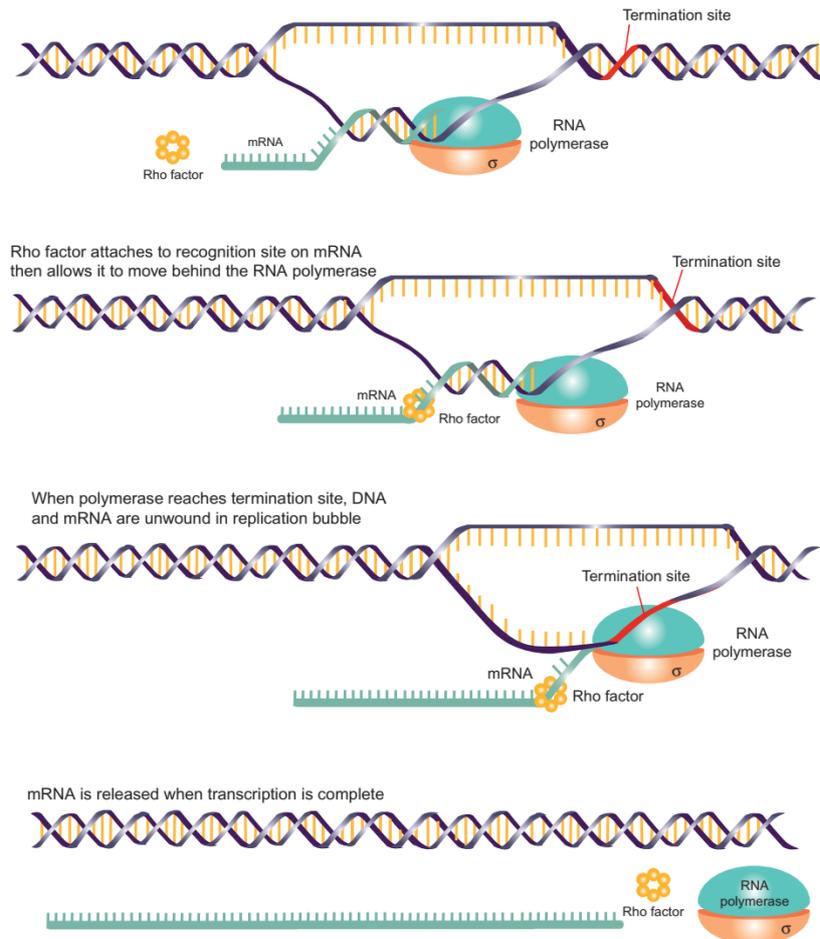


Figure 3.6 Terminaison transcriptionnelle dépendante de Rho (Buckingham, 2011).

3.2 La régulation de la transcription procaryotique

Le génome d'*E. coli* compte plus de 3000 gènes. Certains d'entre eux sont actifs tout le temps parce que leurs produits sont en demande constante. Mais certains d'entre eux ne sont pas transcrits la plupart du temps car leurs produits sont rarement nécessaires. La raison pour laquelle la cellule ne laisse pas tous ses gènes allumés en permanence est que l'expression des gènes est un processus coûteux. Il faut beaucoup d'énergie pour produire de l'ARN et des protéines. Ainsi, le contrôle de l'expression des gènes est essentiel à la vie.

Chez les procaryotes, la transcription est contrôlée par sept stratégies majeures : les facteurs σ alternatifs, la commutation de l'ARN polymérase, l'anti-terminaison, les activateurs, l'opéron, l'atténuation de la transcription et le riboswitch. Les facteurs σ alternatifs, la commutation de l'ARN polymérase et l'anti-terminaison impliquent le changement de la machinerie de transcription pendant l'opéron, l'atténuation de la transcription et le riboswitch impliquent la régulation de groupes de gènes fonctionnellement apparentés. Les activateurs sont les éléments d'ADN qui lient les facteurs protéiques.

3.2.1 Facteurs σ alternatifs

Lorsqu'un phage infecte une bactérie, il subvertit généralement la machinerie de transcription de l'hôte à son propre usage. Dans le processus, il établit un programme de transcription dépendant du temps ou temporel en produisant différentes sous-unités σ qui dirigent l'ARN polymérase vers différents gènes. SPO1 est un virus qui infecte la bactérie *Bacillus subtilis*. Le virus possède un ensemble de gènes appelés gènes précoces, qui sont transcrits par l'ARN polymérase de l'hôte, en utilisant sa sous-unité σ régulière (Fig. 3.7). L'un des gènes viraux précoces code pour une protéine gp28, qui est une sous-unité σ . Cette sous-unité σ dirige l'ARN polymérase pour transcrire davantage de gènes viraux dans la phase intermédiaire. Les produits de deux gènes intermédiaires sont gp33 et gp34, qui forment ensemble une autre sous-unité σ qui dirige la transcription des gènes tardifs. Au final, les sous-unités σ virales contrôlent la machinerie de transcription du virus au lieu de la bactérie.

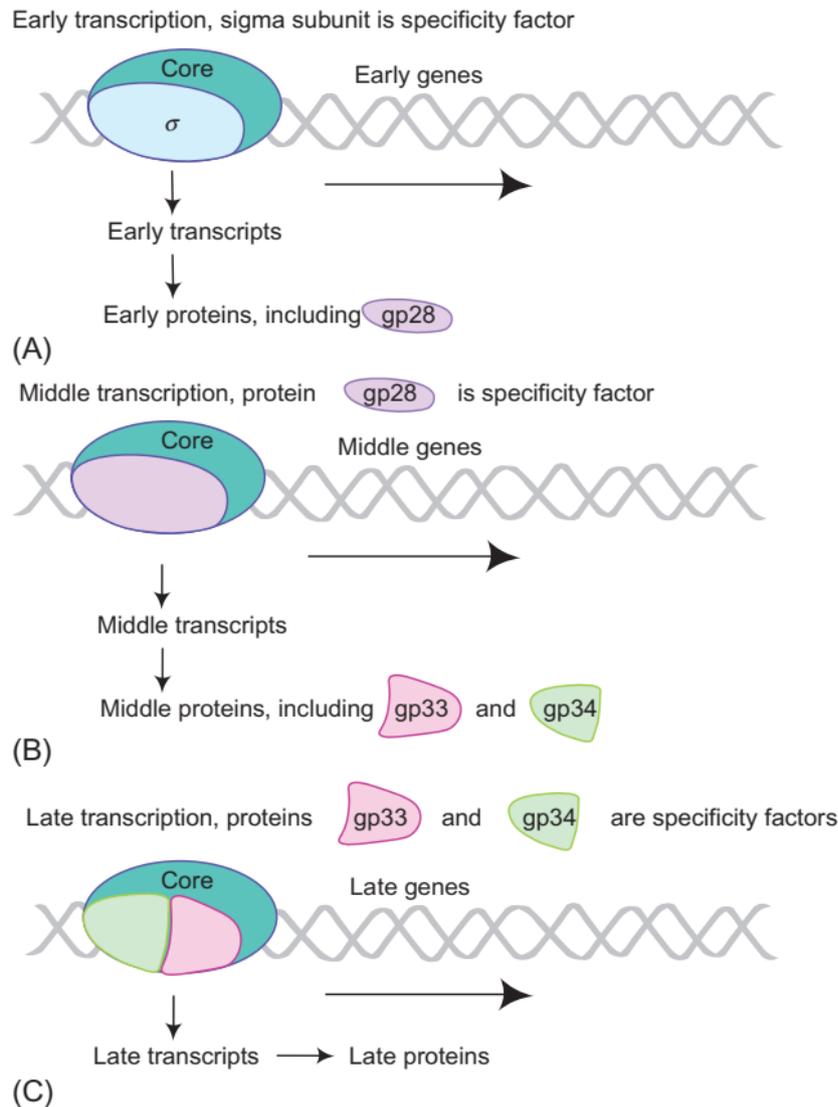


Figure 3.7 Exemple de contrôle temporel de la transcription chez *B. subtilis* infecté par le phage SPO1.

(A) L'holoenzyme de l'ARN polymérase (vert), y compris la sous-unité sigma (bleu clair), dirige la transcription précoce. Le nouveau produit protéique de phage gp28 (violet) est montré, qui agit comme le nouveau facteur sigma. (B) Le nouveau produit protéique gp28 dirige la transcription médiane. Deux protéines de phage médian gp33 et gp34 (rose et verte) sont produites. (C) La transcription tardive se produit en fonction de la polymérase centrale de l'hôte et de gp33 et gp34 (Brown, 2016).

Lorsque les cellules bactériennes subissent une augmentation de la température, ou une variété d'autres insultes environnementales, elles montent une réponse de défense appelée réponse au choc thermique pour minimiser les dommages. Un ensemble de gènes de choc thermique sont exprimés pour protéger les cellules. Dans cette situation, la réponse au choc thermique chez *E. coli* est régie par une sous-unité σ alternative, σ^{32} (σ^H) qui déplace σ^{70} (σ^A) et dirige l'ARN polymérase vers les promoteurs du gène du choc thermique. En plus du mécanisme alternatif de la sous-unité σ , *E. coli* a développé des moyens de contrôler la transcription en utilisant l'anti-sous-unité σ . Lorsque les cellules sont stressées

par des insultes telles que la perte de nutriments, une osmolarité élevée ou une température élevée, elles arrêtent de croître et entrent dans la phase stationnaire. À ce stade, une protéine Rsd est produite. Il peut se lier à la sous-unité σ^{70} et bloquer sa liaison à la polymérase centrale. Ainsi, cette sous-unité anti- σ peut compléter le mécanisme de remplacement de σ en inhibant l'activité de la sous-unité σ^{70} .

3.2.2 Commutation de l'ARN polymérase

Le phage T7 est un simple phage *E. coli*, et il a trois phases de transcription : une phase précoce appelée Classe I et deux phases tardives appelées Classes II et III. Au lieu de créer une nouvelle sous-unité σ , un gène spécifique de classe I code pour une nouvelle ARN polymérase avec une spécificité absolue pour les gènes de classe II et III du phage T7 (Fig. 3.8). Ainsi, le mécanisme de commutation dans ce phage est assez simple.

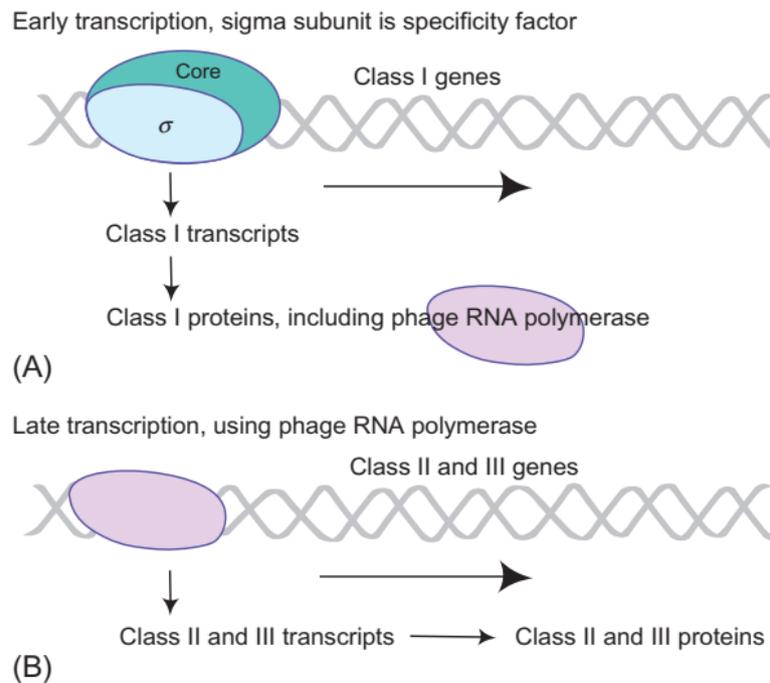


Figure 3.8 Contrôle temporel de la transcription chez le phage T7 infecté *E. coli*.

(A) *E. coli* ARN polymérase holoenzyme et sous-unité sigma contrôlent la transcription précoce (classe I) (B) T7 ARN polymérase contrôle la transcription tardive (classes II et III) (Brown, 2016).

3.2.3 Anti-terminaison

Le phage λ peut se répliquer dans un cycle lytique ou lysogénique. En mode lytique, presque tous les gènes du phage sont transcrits et traduits, et l'ADN du phage est répliqué, conduisant à une nouvelle descendance. En mode lysogénique, l'ADN λ est incorporé dans le génome de l'hôte. Seul *cI*, qui code pour la protéine répresseur λ , est exprimé et empêche alors la transcription de tout le reste des gènes du phage. La commutation transcriptionnelle

précoce immédiate, précoce retardée ou tardive dans le cycle lytique du phage λ est contrôlée par l'anti-termination. L'un des deux gènes précoces immédiats est *cro*, qui code pour un répresseur du gène *cI* qui permet la poursuite du cycle lytique (Fig. 3.9). L'autre, *N*, code pour un anti-terminateur, *N*, qui remplace les terminateurs après les gènes *N* et *cro*. La transcription se poursuit ensuite dans les gènes précoces retardés. L'un des gènes retardés précoces code pour un autre anti-terminateur qui permet la transcription des gènes tardifs. Ces anti-terminateurs agissent pour modifier la polymérase afin qu'elle lise à travers les terminateurs pour continuer la transcription des gènes dans la phase suivante.

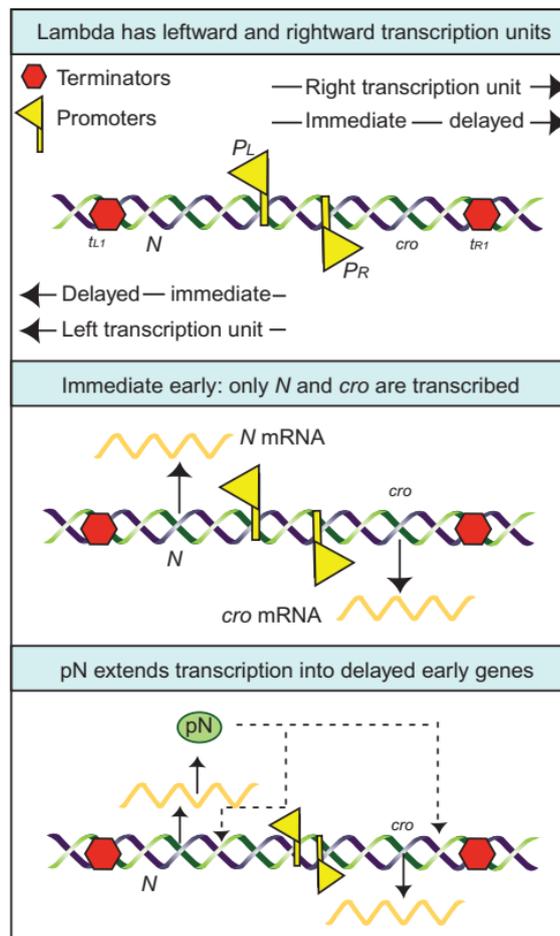


Figure 3.9 Unités de transcription dans les gènes *N* et *cro* du phage lambda sont des gènes précoces immédiats et sont séparées des gènes précoces retardés par des terminateurs. La protéine *N* (pN) permet à l'ARN polymérase de faire passer les terminateurs t_{L1} vers la gauche et t_{R1} vers le serré (Brown, 2016).

3.2.4 Opérons

Chez les procaryotes, les gènes qui codent pour les enzymes de certaines voies métaboliques sont souvent contrôlés en tant que groupe, les gènes codant pour les protéines de la voie étant proches les uns des autres et sous le contrôle d'un promoteur commun. Un tel groupe de gènes contigus et contrôlés de manière coordonnée est appelé un opéron (Fig.

3.10). Habituellement, ces gènes ne sont pas transcrits tout le temps. L'expression de ces gènes peut être déclenchée par la présence d'une substance appropriée appelée inducteur. Ce phénomène est appelé induction. Parce que tous les gènes d'un opéron sont transcrits ensemble pour produire un message, il est également appelé message polycistronique, qui est simplement un message contenant des informations provenant de plus d'un gène. Chaque cistron de l'ARNm a son propre site de liaison au ribosome, de sorte que chaque cistron peut être traduit par des ribosomes séparés qui se lient indépendamment les uns des autres.

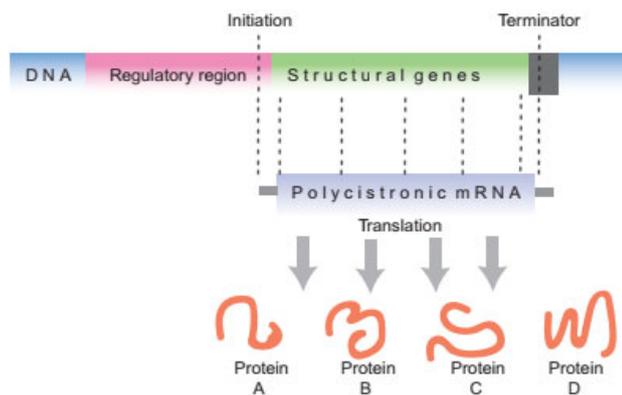


Figure 3.10 Aperçu de la transcription d'un opéron.

Dans un opéron, les gènes structuraux sont situés les uns à côté des autres et sont transcrits en un seul ARNm polycistronique. Au cours de la traduction, plusieurs protéines qui ont généralement une fonction similaire sont produites à partir de cet ARNm unique (Brown, 2016).

Les opérons peuvent être contrôlés par un mécanisme de régulation positif ou négatif. Selon la façon dont ils répondent aux molécules qui contrôlent leur répression, ils peuvent être inductibles, répressibles ou les deux. Il existe quatre possibilités générales (Fig. 3.11). Dans un système inductible contrôlé négativement, une protéine répresseur arrête la transcription lorsqu'elle se lie au promoteur. Il est inductible lorsqu'il libère la répression en présence d'inducteur ou de co-inducteur. Si le répresseur est muté d'une manière qui arrête sa fonction, l'opéron est toujours exprimé. Les gènes qui sont toujours exprimés sont dits constitués. Pour un système inductible contrôlé positivement, la protéine de contrôle est un inducteur qui se lie au promoteur, stimulant la transcription, mais il ne fonctionnera que lorsqu'il est lié à son co-inducteur. Si le gène de l'inducteur est muté, il devient inductible. Dans un système répressible contrôlé négativement, un répresseur arrête la transcription, mais ce répresseur ne fonctionne qu'en présence d'un co-répresseur, qui peut être une protéine ou une petite molécule. Dans un système répressible contrôlé positivement, une protéine inductrice se lie au promoteur, stimulant la transcription. Cependant, l'inducteur est inactivé en présence du co-répresseur.

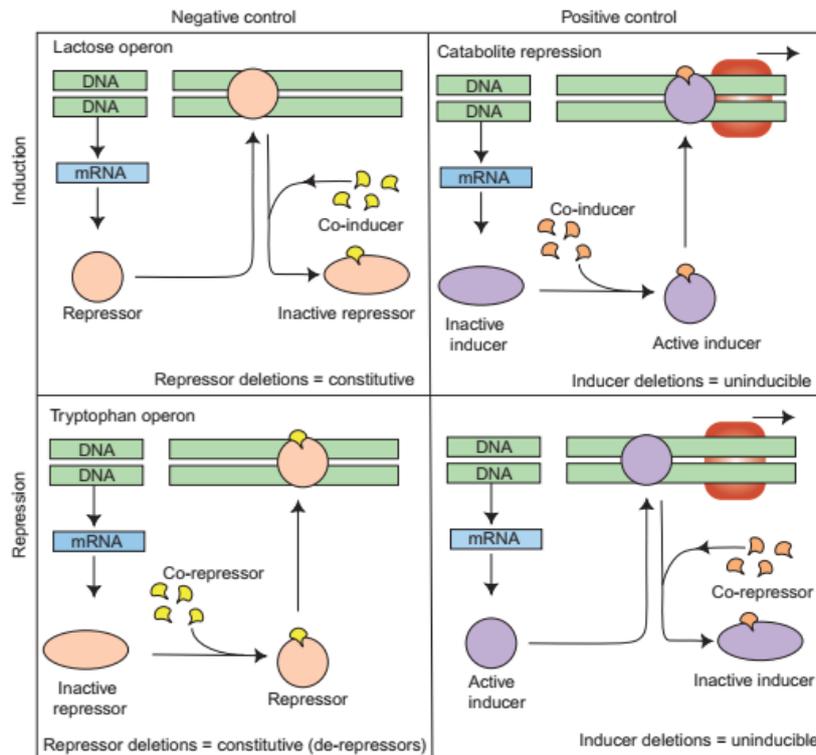


Figure 3.11 Régulation de base de l'expression des gènes.

Les mécanismes de contrôle peuvent être inductibles ou répressibles, contrôlés négativement ou positivement. Les co-répresseurs peuvent être des inducteurs inactifs ou des répresseurs actifs, tandis que les co-inducteurs peuvent inhiber les répresseurs ou stimuler les inducteurs (Brown, 2016). Le métabolisme du lactose dans *E. coli* est assuré par trois enzymes structurales. Les gènes structuraux codent pour les produits géniques impliqués dans la voie biochimique de l'opéron. Les trois gènes sont *lacZ*, qui code pour l'enzyme β -galactosidase, *lacY*, qui code pour l'enzyme galactoside perméase, et *lacA*, qui code pour l'enzyme galactoside transacétylase. L'opéron *lac* est un système inductible contrôlé négativement. L'expression de ces gènes de structure est sous le contrôle d'un gène régulateur (*lac I*) qui code pour le répresseur. Le répresseur peut se lier à l'opérateur (O). En l'absence de lactose, l'opérateur est lié par le répresseur et l'ARN polymérase ne peut pas se lier à la région promotrice adjacente, ce qui facilite l'expression des gènes de structure. En tant que tel, l'opéron *lac* est réprimé tant qu'il n'y a pas de lactose, ce qui est une régulation négative (Fig. 3.12).

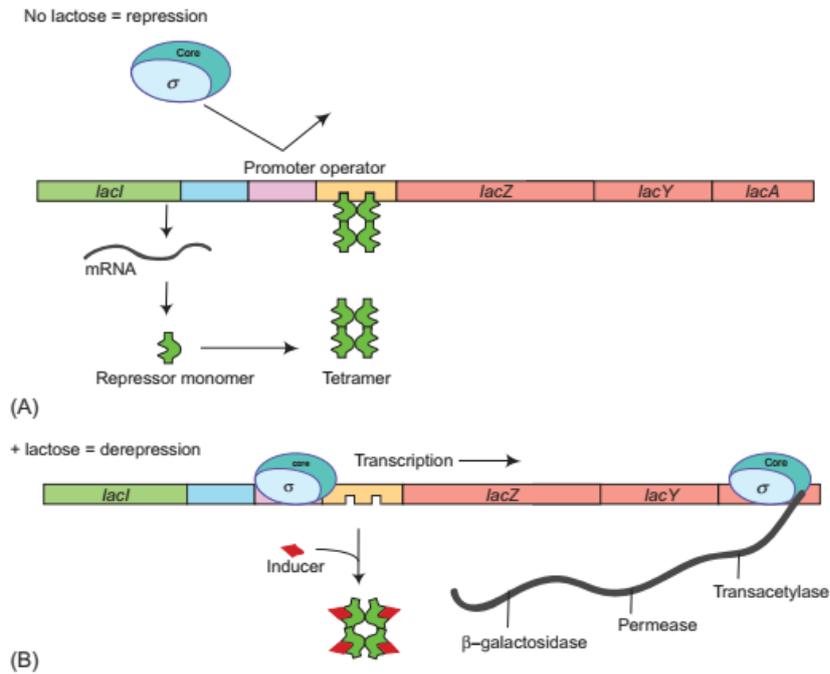


Figure 3.12 Contrôle négatif de l'opéron lac.

(A) En l'absence de lactose, le gène *lacI* produit un répresseur qui empêche l'ARN polymérase de transcrire les gènes. (B) En présence de lactose, l'inducteur se lie au répresseur et change de forme pour ne plus pouvoir se lier à l'opérateur. Cela permet à l'ARN polymérase de transcrire les gènes *lacZ*, *lacY* et *lacA* (Brown, 2016).

L'opéron lac devient inductible lorsque le lactose est la seule source de carbone. En présence de lactose, l'inducteur de l'allolactose déclenche l'expression de *lacZ*. En effet, l'inducteur se lie au répresseur, produisant un répresseur inactif qui ne peut pas se lier à l'opérateur. L'ARN polymérase peut maintenant se lier au promoteur, et la transcription et la traduction des gènes de structure peuvent avoir lieu. Le produit β -galactosidase peut décomposer le lactose en galactose et glucose. D'autre part, la galactoside perméase peut transporter le lactose dans la cellule. La galactoside transacétylase peut être responsable de l'inactivation de certains antibiotiques qui peuvent pénétrer dans la cellule par la lactose perméase. L'opéron lac est pleinement actif lorsque le lactose est présent et que le glucose est absent. Comme nous le verrons plus tard, l'opéron lac est particulièrement faible, et la liaison à l'ARN polymérase est minimale en présence de glucose et en présence de lactose. La forte liaison de l'ARN polymérase au promoteur nécessite une autre protéine régulatrice dont nous parlerons également plus tard.

3.2.5 Améliorants

Les activateurs sont des éléments qui stimulent la transcription et sont des éléments d'ADN agissant en cis qui ne font pas strictement partie du promoteur. Ils peuvent se trouver

jusqu'à plusieurs kilobases de distance et soit en amont soit en aval des promoteurs qu'ils contrôlent. La régulation se fait en bouclant l'ADN afin que les protéines activatrices liées à l'amplificateur puissent entrer en contact avec la machinerie de transcription via une protéine connue sous le nom de complexe médiateur (Fig. 3.13).

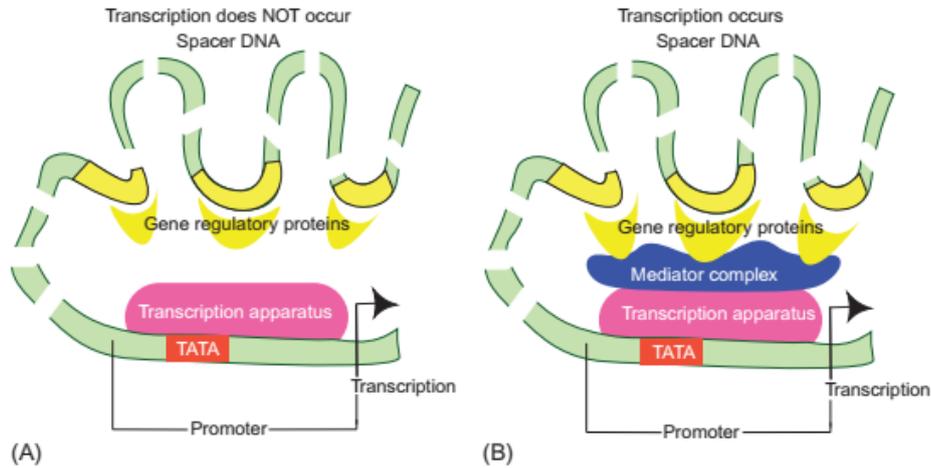


Figure 3.13 Complexe médiateur et protéines activatrices (Klug et, 2011).

Parce que l'ADN est replié, les protéines activatrices peuvent atteindre l'appareil de transcription. Le complexe médiateur permet aux activateurs et/ou aux protéines répressives de se lier à l'ARN polymérase.

Pour l'exemple de l'opéron lac, lorsque le glucose et le lactose sont présents, la cellule ne fabrique pas les protéines lac. La répression de la synthèse des protéines lac par le glucose est appelée répression catabolique, qui implique les deux régions dans le promoteur. L'un est le site de liaison de l'ARN polymérase et l'autre est le site de liaison d'une autre protéine régulatrice, la protéine activatrice des catabolites (CAP). Lorsque le glucose n'est pas présent, CAP forme un complexe avec 3', 5'-AMP cyclique (cAMP) et se lie au site CAP sur le promoteur (Fig. 3.14). Cette liaison peut aider l'ARN polymérase à se lier au promoteur et l'opéron est pleinement actif. En présence de glucose, le taux d'AMPc est faible, seuls très peu de complexes CAP-AMPc peuvent se former. Il ne suffit pas d'activer l'opéron lac. La présence de lactose est nécessaire à la transcription de l'opéron, mais pas suffisante. L'opéron lac n'est actif que lorsque le glucose est absent et que le lactose est présent. En tant que tel, le site CAP est un exemple d'élément activateur et le complexe CAP-AMPc est un facteur de transcription. La modulation de la transcription par CAP est un type de régulation positive (Buckingham, 2011).

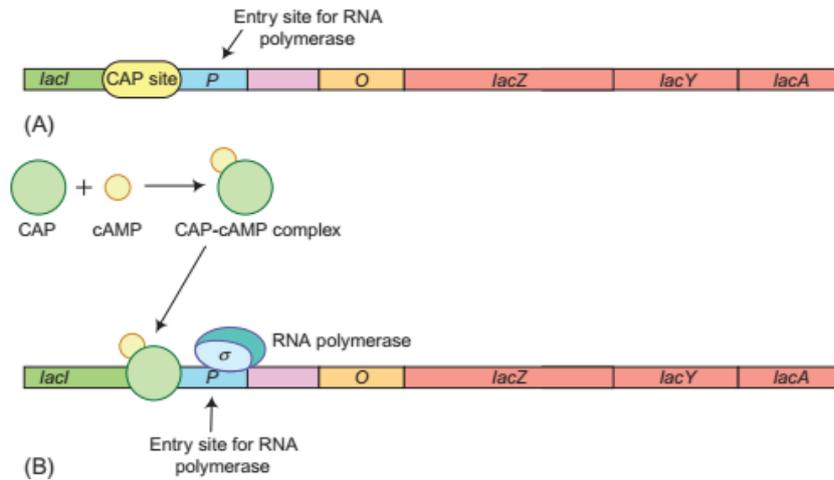


Figure 3.14 Répression catabolique (Klug et al, 2011).

(A) Le complexe CAP-AMPc doit se lier au site CAP du promoteur avant que l'ARN polymérase puisse se lier. (B) En l'absence de glucose, l'AMPc forme un complexe avec CAP, qui se lie au site CAP, permettant à l'ARN polymérase de se lier au promoteur et de transcrire les gènes.

3.2.6 Atténuation de la transcription

L'opéron *trp* d'*E. coli* code pour les enzymes dont la bactérie a besoin pour fabriquer l'acide aminé tryptophane. Comme l'opéron *lac*, l'opéron *trp* est un mécanisme de contrôle négatif. L'opéron *lac* répond à un inducteur qui provoque la dissociation du répresseur de l'opérateur, dérèprimant l'opéron. L'opéron *trp* répond à une protéine répresseur qui se lie à deux molécules de tryptophane. Lorsque le tryptophane est abondant, ce complexe répresseur-tryptophane se lie à l'opérateur *trp*. Cette liaison empêche la liaison de l'ARN polymérase, de sorte que l'opéron n'est pas transcrit (Fig. 3.15). D'autre part, lorsque les niveaux de tryptophane sont réduits, le répresseur ne se lie pas à l'opérateur, de sorte que l'opéron est transcrit. C'est un exemple de système répressible et sous régulation négative. En plus de la régulation négative standard, l'opéron *trp* est régulé par un autre mécanisme de contrôle appelé atténuation de la transcription. Ce mécanisme opère en provoquant l'arrêt prématuré de la transcription de l'opéron lorsque le tryptophane est abondant. Comme représenté sur la figure 3.15, il y a deux loci, le leader *trp* et l'atténuateur *trp*, entre l'opérateur et le gène *trpE*. Des structures secondaires formées dans l'ARNm de la séquence leader sont responsables de cette terminaison prématurée. La formation de ces structures secondaires provient des signaux d'arrêt de la transcription -- une répétition inversée et une chaîne de 8 paires AT dans l'atténuateur. Lorsque le tryptophane est rare, l'opéron est traduit normalement. Lorsqu'elle est abondante, la transcription est interrompue prématurément après la transcription des séquences de tête. Ainsi, l'atténuation impose un niveau supplémentaire de contrôle sur un opéron, au-delà du système répresseur-opérateur.

3.2.7 Riboswitches

Un riboswitch est une région, généralement dans la région 5' non traduite (5'-UTR) d'un ARNm, qui contient deux modules : un aptamère et une plate-forme d'expression. Un aptamère est une région qui se lie à un ligand. D'autre part, la plate-forme d'expression peut être un terminateur, un site de liaison au ribosome ou un autre élément d'ARN qui affecte l'expression du gène lors du changement de conformation. Par exemple, l'opéron *ribD* chez *B. subtilis* contrôle la synthèse et le transport de la vitamine riboflavine et de l'un de ses produits, la flavine mononucléotide (FMN). L'opéron *rib* contient un élément conservé dans leurs 5'-UTR connu sous le nom d'élément RFN, qui est un aptamère. Il existe deux conformations alternatives de l'élément RFN, la conformation d'anti-terminaison et de terminaison. FMN se lie directement à l'élément RFN et l'amène à former une conformation de terminaison, bloquant l'expression. En absence de FMN, l'élément forme une conformation antiterminaison qui peut compléter la transcription. Dans un autre exemple, l'ARNm *glmS* de *B. subtilis* contient un riboswitch qui est une ribozyme-RNase. Il peut répondre au produit de l'enzyme codé par l'ARNm. Lorsque ce produit s'accumule, il se lie au riboswitch, modifiant la conformation de l'ARN pour devenir une RNase efficace. La RNase détruit l'ARNm, donc moins d'enzyme est produite. En tant que tels, ces exemples de riboswitches démontrent que le riboswitch peut fonctionner en diminuant l'expression des gènes. Lorsque ce produit s'accumule, il se lie au riboswitch, modifiant la conformation de l'ARN pour devenir une RNase efficace. La RNase détruit l'ARNm, donc moins d'enzyme est produite. En tant que tels, ces exemples de riboswitches démontrent que le riboswitch peut fonctionner en diminuant l'expression des gènes. Lorsque ce produit s'accumule, il se lie au riboswitch, modifiant la conformation de l'ARN pour devenir une RNase efficace. La RNase détruit l'ARNm, donc moins d'enzyme est produite. En tant que tels, ces exemples de riboswitches démontrent que le riboswitch peut fonctionner en diminuant l'expression des gènes (Rallapalli et al, 2014).

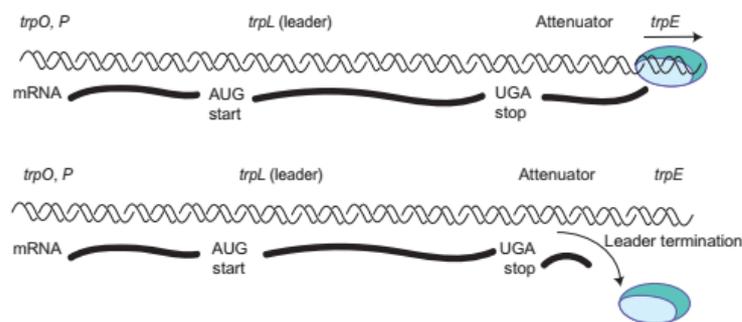


Figure 3.15 Aperçu de la régulation de l'opéron *trp*.

Lorsque le tryptophane est faible, l'ARN polymérase (bleu) lit à travers l'atténuateur et les gènes sont transcrits. Atténuation de l'opéron *trp* d'*E. coli*. Lorsque le tryptophane est élevé, l'atténuateur provoque l'arrêt prématuré de la transcription, de sorte que les gènes qui produisent plus de tryptophane ne sont pas transcrits (Klug et al, 2011).

3.3 Transcription chez les eucaryotes

3.3.1 ARN polymérase eucaryote

Le processus de transcription est plus compliqué chez les eucaryotes. Trois ARN polymérases avec des activités différentes sont connues, et chacune transcrit un ensemble différent de gènes et reconnaît un ensemble différent de promoteurs. L'ARN polymérase synthétise les précurseurs de la plupart des ARN ribosomiques. L'ARN polymérase II synthétise des précurseurs d'ARNm. L'ARN polymérase III se trouve dans le nucléoplasme et synthétise l'ARNt, les précurseurs de l'ARN ribosomal 5S et certaines petites molécules d'ARN impliquées dans le traitement de l'ARNm et le transport des protéines.

Les trois ARN polymérases eucaryotes sont de grands complexes et contiennent 10 sous-unités ou plus. Ils sont tous liés à la polymérase procaryote et les uns aux autres. Les sous-unités qui sont uniques à une polymérase eucaryote particulière sont vraisemblablement responsables de fonctions spécifiques à chaque type de polymérase, y compris la régulation de la transcription. Parmi les trois ARN polymérases, l'ARN polymérase II est la plus étudiée. Dans la levure *Saccharomyces cerevisiae*, l'ARN polymérase II comporte 12 sous-unités appelées RPB1 à RPB12. Les sous-unités centrales sont RPB1, RPB2 et RPB3, qui sont nécessaires à l'activité enzymatique. Ils sont homologues aux sous-unités β' , β et α , respectivement, de l'ARN polymérase d'*E. coli*.

L'ARN polymérase II peut coupler la transcription au traitement du transcrit d'ARN pour produire l'ARNm mature qui est une fonction unique qui ne se trouve pas dans d'autres polymérases. Cette activité de couplage dépend du domaine fonctionnel - domaine C-terminal (CTD). CTD comprend l'extrémité C-terminale du RPB1 et se compose de répétitions de sept acides aminés. Le CTD a le rôle central dans la transition de l'initiation à l'élongation et dans le traitement du transcrit d'ARN.

3.3.2 Structure du promoteur

Trois ARN polymérases eucaryotes ont des structures différentes et transcrivent différentes classes de gènes. Il est donc raisonnable de prédire que ces polymérases reconnaîtraient des promoteurs différents. Les promoteurs reconnus par l'ARN polymérase

II (promoteurs de classe II) sont les plus complexes et les mieux étudiés. Les promoteurs de classe II ont généralement deux parties : le promoteur central et le promoteur proximal (Fig. 3.16). Le promoteur central peut attirer des facteurs de transcription généraux et l'ARN polymérase II, déterminant ainsi le site de départ et la direction de la transcription. Il est modulaire et contient à la fois la boîte initiateur et la boîte TATA situées à environ 37 pb du site de début de transcription, de chaque côté. La boîte initiateur est une séquence trouvée au site où la transcription commence. Environ 25 paires de bases en amont de celle-ci se trouve la boîte TATA, une séquence riche en AT, qui est reconnue par les protéines de liaison TATA (TBP). Les éléments promoteurs proximaux se trouvent généralement en amont des promoteurs centraux. Ces éléments en amont sont les sites de reconnaissance de facteurs de transcription spécifiques. Par exemple, la boîte GC qui a la séquence consensus de GGGCGG, se lie au facteur de transcription Sp1. En dehors des éléments mentionnés ci-dessus,

3.3.3 Transcription dirigée par l'ARN polymérase I et III

Les ARN polymérases eucaryotes I et III reposent sur un ensemble distinct de protéines pour initier la transcription. Bien que les ARN polymérases I et III partagent plusieurs sous-unités enzymatiques centrales identiques avec l'ARN polymérase II, elles reconnaissent des séquences promotrices très différentes et possèdent des facteurs de transcription généraux uniques. Les promoteurs reconnus par l'ARN polymérase I ne sont pas bien conservés en séquence d'une espèce à l'autre. Cependant, ils ont tous une architecture générale similaire du promoteur car il se compose d'un élément central entourant le site de démarrage de la transcription et d'un élément promoteur en amont, qui se trouve à environ 100 pb plus en amont. L'ARN polymérase I, qui transcrit les gènes de l'ARNr, se lie au promoteur contenant un élément promoteur central et un élément de contrôle en amont (UCE). Le TBP, qui fait partie d'un plus grand complexe appelé SL1, aide l'ARN polymérase I à reconnaître le promoteur principal (Fig. 3.17). Les promoteurs classiques de l'ARN polymérase III sont de type I et de type II qui ont des promoteurs qui se trouvent entièrement dans les gènes. Ces gènes de type I et de type II comprennent une variété de gènes d'ARN tels que l'ARNt, la sous-unité d'ARN 5S du ribosome et les gènes d'ARN VA de l'adénovirus (Fig. 3.18). Les promoteurs d'ARN de type III de la polymérase III sont non classiques et ressemblent à ceux des gènes de classe II, notamment le gène snRNA U6, le gène ARN 7SL et le gène ARN 7SK.



Figure 3.16 Quatre éléments des promoteurs de l'ARN polymérase II.

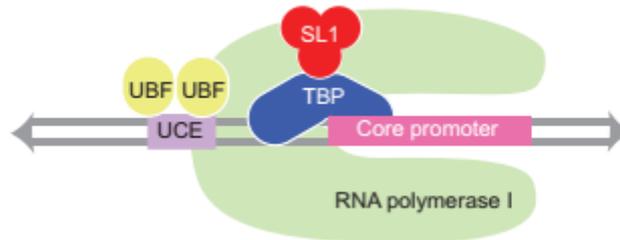


Figure 3.17 Promoteurs d'ARN pol I contiennent une région centrale ainsi qu'un élément de contrôle en amont (UCE), qui interagissent avec des facteurs de transcription comme UBF et SL1 chez l'homme (Shapiro et Hofreiter, 2011).

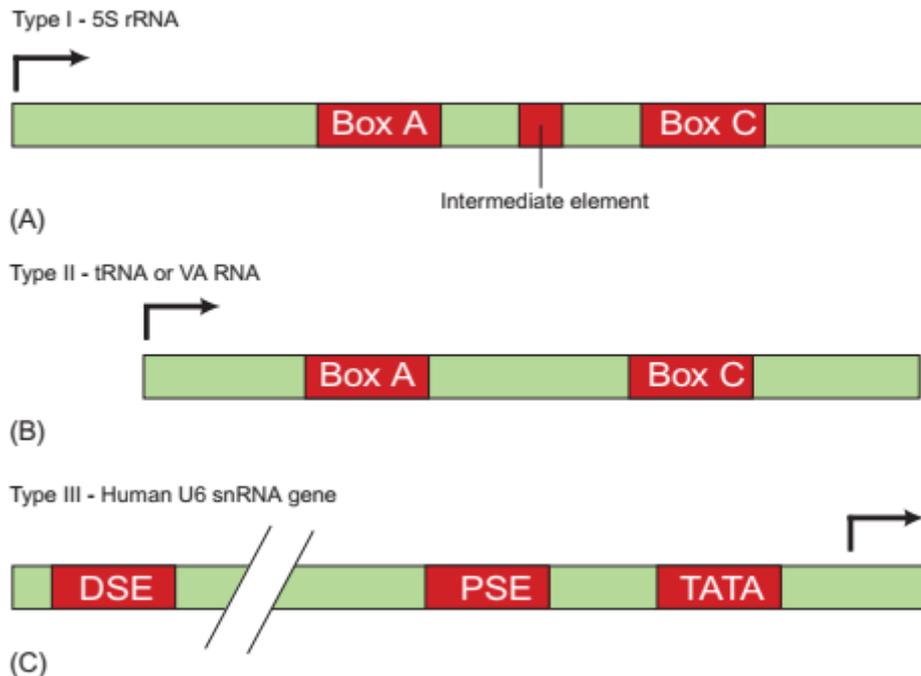


Figure 3.18 Exemples de promoteurs de l'ARN polymérase III.

Les boîtes rouges signifient les promoteurs des gènes 5S rRNA, tRNA et U6 snRNA. DSE représente un élément de séquence distal, PSE représente un élément de séquence proximal. Les promoteurs de la polymérase III sont non classiques et ressemblent à ceux des gènes de classe II, notamment le gène snRNA U6, le gène ARN 7SL et le gène ARN 7SK (Steitz, 1999).

4 Expression génique : Traduction du code génétique

4.1 Le code génétique

4.1.1 Les bases de ribonucléotides utilisées comme lettres dans le code génétique

Le code génétique est écrit sous forme linéaire, en utilisant les ribonucléotides qui composent les molécules d'ARNm sous forme de lettres. La séquence ribonucléotidique est dérivée des bases nucléotidiques complémentaires dans le brin matrice d'ADN. Par conséquent, la séquence nucléotidique est exactement la même que le brin codant de l'ADN. Chaque code génétique se compose de trois lettres ribonucléotidiques, ainsi appelées code triplet. En tant que tel, un code génétique est un code triplet dans lequel une séquence de trois bases est nécessaire pour spécifier un acide aminé. Le code génétique traduit les séquences d'ARN dans la séquence d'acides aminés (Fig. 4.1). Chaque groupe de trois ribonucléotides, appelé codon, spécifie un acide aminé. Ces codes sont sans ambiguïté, car chaque triplet ne spécifie qu'un seul acide aminé. Ainsi, on pourrait imaginer qu'un codon aurait au moins trois bases de long. Avec trois bases, il y a $4^3=64$ codons, ce qui est plus que suffisant pour coder les 20 acides aminés. Par conséquent, le code génétique est dégénéré, ce qui signifie que plusieurs triplets peuvent coder le même acide aminé. Chaque acide aminé peut avoir plus d'un codon, mais aucun codon ne peut coder plus d'un acide aminé. De plus, le code génétique est universel, car le code peut être utilisé par tous les virus, procaryotes, archées et eucaryotes.

4.1.2 Règles d'oscillation

Les 64 codons ont tous reçu une signification, 61 d'entre eux codant pour les acides aminés et les 3 autres servant de signaux de terminaison, également appelés codons non-sens (Tableau 4.1). Plusieurs codons pour un seul acide aminé ne sont pas distribués au hasard mais ont une ou deux bases en commun. Les bases communes à plusieurs codons sont généralement les premières et la deuxième base, avec plus de place pour la variation dans la troisième base, qui est appelée la base oscillante. Les règles d'oscillation indiquent qu'un anticodon de première base U pourrait reconnaître un A ou un G dans la position de troisième base du codon, et qu'un anticodon de première base G pourrait reconnaître U ou C dans la position de troisième base du codon (Tableau 4.2). Parce que les codons dégénérés pour un

acide aminé donné différent dans la troisième base, un ARNt donné peut s'apparier avec plusieurs codons. Ainsi, moins d'ARNt différents sont nécessaires (Klug et al, 2011).

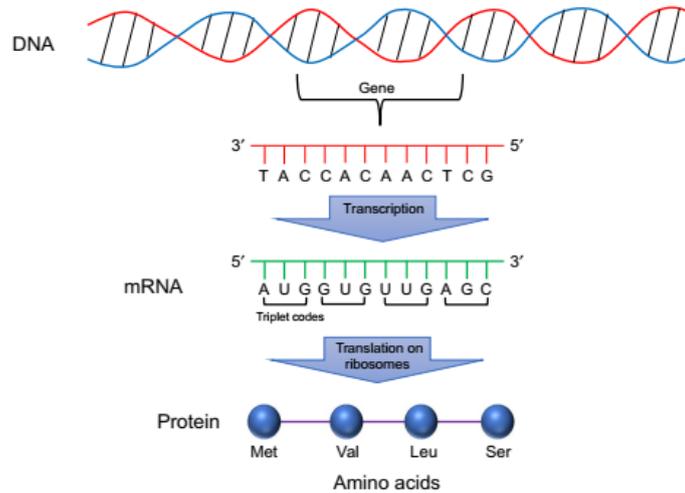


Figure 4.1 Organigramme démontrant le dogme central de la biologie dans lequel l'ADN est transcrit en ARNm, qui est ensuite traduit en une séquence d'acides aminés d'une protéine.

Tableau 4.1 Code génétique dégénératif

		Second letter							
		U	C	A	G				
First letter	U	UUU } Phe	UCU } Ser UCC } UCA } UCG }	UAU } Tyr	UGU } Cys	U C A G	Third letter		
		UUA } Leu		UAA Stop	UGA Stop				
		C		CUU } Leu	CCU } Pro			CAU } His	CGU } Arg
				CUC } CUA } CUG }	CCC } CCA } CCG }			CAC } CAG }	CGC } CGA } CGG }
	A	AAU } Ile	ACU } Thr ACC } ACA } ACG }	AAU } Asn	AGU } Ser	U C A G			
		AUA }		AAA } Lys	AGA } Arg				
		AUG Met		AAG }	AGG }				
	G	GUU } Val	GCU } Ala GCC } GCA } GCG }	GAU } Asp	GGU } Gly	U C A G			
		GUC }		GAA } Glu	GCC }				
		GUA }		GAC }	GGA }				
		GUG }		GAG }	GGG }				

Tableau 4.2 Appariement de bases de l'hypothèse de Wobble au troisième codon

Base on the Anticodon	Base Recognized on the Codon
U	A,G
C	G
A	U
G	U,C
I (inosine; a modified base in tRNA)	U,C,A

4.2 Machine translationnelle : Ribosome

La traduction de l'ARNm est la polymérisation biologique des acides aminés en chaînes polypeptidiques. Ce processus nécessite de l'ARNm, de l'ARNt et un ribosome, composé d'ARNr. En association avec un ribosome, l'ARNm présente un codon triplet qui appelle un acide aminé spécifique. Une molécule d'ARNt spécifique contient dans sa séquence nucléotidique trois ribonucléotides consécutifs complémentaires du codon, appelés anticodon, qui peuvent s'apparier avec le codon. Une autre région de cet ARNt est liée de manière covalente à l'acide aminé correspondant du codon. La liaison hydrogène de l'ARNt à l'ARNm maintient les acides aminés à proximité les uns des autres afin qu'une liaison peptidique puisse être formée. Ce processus se produit à plusieurs reprises lorsque les ARNm traversent le ribosome et que les acides aminés sont polymérisés en un polypeptide.

4.2.1 Les composants des ribosomes

Un ribosome bactérien est une machine moléculaire complexe et se compose de deux sous-unités, une grande et une petite. Les deux sous-unités sont constituées d'une ou plusieurs molécules d'ARNr et de protéines ribosomiques. Ces deux sous-unités distinctes effectuent les tâches disparates de lecture d'ARNm, de reconnaissance d'ARNt et d'allongement de la chaîne polypeptidique. La petite sous-unité (30S chez les procaryotes et 40S chez les eucaryotes) contient le site de décodage où l'interaction codon-anticodon est déchiffrée, tandis que la grande sous-unité (50S chez les procaryotes et 60S chez les eucaryotes) contient le site actif où se produisent les réactions de transfert de peptidyle et d'hydrolyse. Malgré leur fonction moléculaire conservée, les sous-unités ribosomiques eucaryotes et procaryotes diffèrent considérablement en taille et en complexité. Ces différences peuvent refléter une régulation supplémentaire de la traduction et sont également à la base de plusieurs antibiotiques qui bloquent spécifiquement la fonction des sous-unités procaryotes. En général, les deux tiers de la masse du ribosome sont constitués d'ARNr et l'autre tiers est composé de protéines ribosomiques. Les composants de la sous-unité et de l'ARNr sont plus facilement isolés et caractérisés sur la base de leur comportement de sédimentation dans un gradient de saccharose. La vitesse de sédimentation a été normalisée en unités appelées Svedbergs (S). Parce que la masse et la forme d'une molécule déterminent sa migration, plus la masse est grande, plus la vitesse de sédimentation est grande. La grande sous-unité des procaryotes consiste en une molécule d'ARNr 23S, une molécule d'ARNr 5S et 34 protéines ribosomiques ; la petite sous-unité est constituée d'un composant

ARNr 16S et de 21 protéines ribosomiques (Fig. 4.2). D'autre part, la grande sous-unité chez les eucaryotes est constituée d'un ARNr 28S, d'un ARNr 5,8S, d'un ARNr 5S et de 46 protéines ribosomiques, et la petite sous-unité est constituée d'ARNr 18S et de 33 protéines ribosomiques. La petite sous-unité assure la médiation des interactions entre l'ARNm et l'ARNt. La sous-unité large catalyse la formation de liaisons peptidiques. Ces événements distincts sont intégrés à l'interface entre les deux sous-unités, où le mouvement brut de l'interface peut déplacer la position du complexe ARNt-ARNm lorsque des acides aminés sont ajoutés séquentiellement à la chaîne polypeptidique en croissance. Ces mouvements sont coordonnés par des éléments de pont spécifiques entre les deux sous-unités, qui sont composées d'ARN et de protéines.

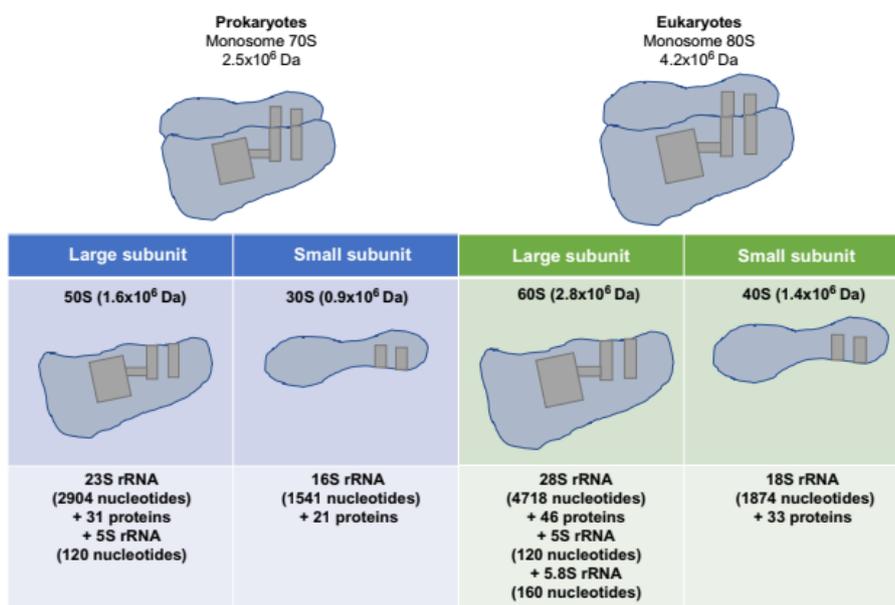


Figure 4.2 Composition de ribosomes procaryotes par rapport à la composition de ribosomes eucaryotes (Jeon, 2012).

4.2.2 Structure ribosomique

Chacun des principaux ARNr a une structure secondaire avec plusieurs domaines discrets. Quatre domaines généraux sont formés par l'ARNr 16S, dans lequel un peu moins de la moitié de la séquence est appariée. Six domaines généraux sont trouvés par l'ARNr 23S. La structure tridimensionnelle des sous-unités ribosomiques 30S et 50S montre que les formes générales des sous-unités ribosomiques sont déterminées par la conformation des molécules d'ARNr en leur sein. Les protéines ribosomiques jouent un rôle largement structurel dans les ribosomes, et leur fonction principale est de renforcer et de stabiliser les conformations d'ARNr au sein des sous-unités ribosomiques.

L'organisation des domaines d'ARNr dans les deux sous-unités ribosomiques est étonnamment différente. Dans la petite sous-unité, chaque domaine d'ARNr est largement limité à une zone de la sous-unité, tandis que dans la grande sous-unité, les éléments d'ARNr sont intimement entrelacés. Cette organisation reflète les différences fonctionnelles des sous-unités. Dans la petite sous-unité, la flexibilité interdomaine est adaptée aux mouvements à grande échelle. En revanche, la stabilité apportée par les domaines d'ARN entrelacés est adaptée à la protection du site actif de synthèse protéique par la grande sous-unité. Par conséquent, le domaine central du ribosome lie l'ARNm et l'extrémité tige-boucle de l'anticodon des aminoacyl-ARNt, fournissant le cadre pour décoder l'information génétique dans l'ARNm par reconnaissance codon-anticodon.

Le domaine central de la sous-unité 30S qui sert de centre de décodage est composé uniquement d'ARNr 16S. Les molécules d'ARNr 16S se replient en structures secondaires caractéristiques à la suite d'interactions intramoléculaires d'appariement de bases (Fig. 4.3). L'extrémité 30 de l'ARNr 16S interagit directement avec l'ARNm à l'initiation et les régions spécifiques de l'ARNr 16S interagissent directement avec les régions anticodon de l'ARNt dans le site A et le site P. D'autre part, la sous-unité 50S lie les extrémités accepteurs d'aminoacyl des ARNt et est responsable de la catalyse de la formation de la liaison peptidique entre les acides aminés successifs de la chaîne polypeptidique. Ce centre catalytique est appelé peptidyl transférase et est situé au fond d'une fente profonde. L'ARNr 23S interagit avec l'extrémité CCA du peptidyl-ARNt dans le site P et le site A. Par conséquent, ces deux sous-unités distinctes effectuent les tâches disparates de la reconnaissance de l'ARNt et d'élongation de la chaîne polypeptidique. Lors de l'initiation de la traduction, les deux sous-unités se rejoignent pour former le ribosome 70S (80S chez les eucaryotes) et lancer le cycle d'élongation.

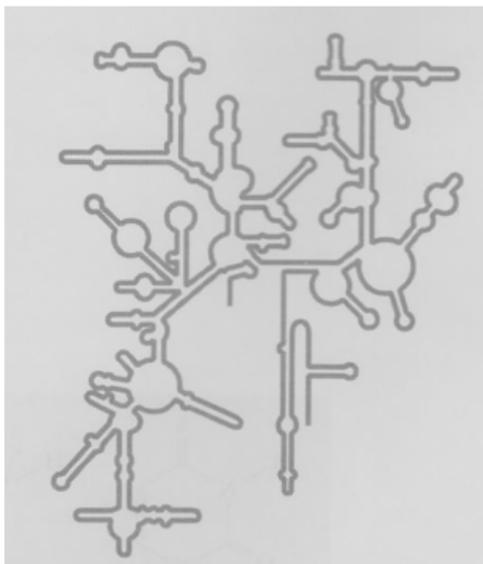


Figure 4.3 Séquences de bases complémentaires liées à l'hydrogène de l'ARN ribosomal 16S maintenant la structure secondaire (Buckingham, 2011).
Les régions alignées représentent des séquences de bases complémentaires à liaison hydrogène.

4.3 Machine translationnelle : ARNt

4.3.1 La structure de l'ARNt

Pour comprendre comment un ARNt remplit ses fonctions, nous devons connaître la structure de la molécule. L'ARNt est composé de seulement 75 à 90 nucléotides, affichant une structure presque identique chez les procaryotes et les eucaryotes. Tout comme une protéine a une structure primaire, secondaire et tertiaire, il en va de même pour un ARNt. L'ARNt est transcrit à partir de l'ADN sous forme de grands précurseurs. La structure primaire est la séquence linéaire de bases dans l'ARN ; la structure secondaire est la structure tige-boucle issue de l'appariement de bases entre différentes régions de la séquence ; et la structure tertiaire est la forme tridimensionnelle globale de la molécule (Klug, 2011).

4.3.2 La structure secondaire de l'ARNt

En général, la structure secondaire de l'ARNt adopte une forme de feuille de trèfle. Le trèfle à quatre tiges appariées qui définissent les quatre régions principales de la molécule (Fig. 4.4). Le premier, vu en haut du diagramme, est la tige acceptrice. Le bras accepteur a les extrémités 5' et 3' de l'ARNt, et ces deux extrémités sont appariées l'une à l'autre. L'extrémité 3', portant la séquence invariante CCA, dépasse de l'extrémité 5'. Sur la gauche se trouve la boucle dihydrouracile (boucle D). Ceci est nommé pour les bases d'uracile modifiées dans cette région. En bas se trouve la boucle anticodon. Cette région s'apparie

avec un codon d'ARNm et permet donc le décodage de l'ARNm. À droite se trouve la boucle T en raison de sa séquence invariante de TΨC. Le Ψ est une uridine modifiée. La région entre la boucle anticodon et la boucle T est appelée boucle variable en raison de sa longueur variable. La structure tertiaire de l'ARNt provient des interactions d'appariement de bases entre les bases de la boucle D avec les bases des boucles variable et TΨC (Fig. 4.5). Ces interactions d'appariement de bases impliquent les nucléotides invariants des ARNt. Ces interactions plient les bras D et TΨC ensemble et plient la feuille de trèfle dans la forme tertiaire en forme de L stable. Beaucoup de ces interactions d'appariement de bases impliquent des paires de bases qui ne sont pas des appariements canoniques A:T ou G:C.

4.3.3 La forme active de l'ARNt

Pour qu'une traduction précise se produise, la reconnaissance des codons doit être obtenue par l'aminoacyl-ARNt, qui est une forme chargée et activée d'ARNt. Le processus d'activation de l'ARNt, catalysé par les aminoacyl-ARNt synthétases, est discuté dans la prochaine section. L'aminoacyl-ARNt approprié lit le codon par appariement de bases via sa boucle anticodon. Par la suite, l'acide aminé spécifique est ajouté passivement à la chaîne peptidique en croissance. Par conséquent, il est important que les synthases soient capables de faire la distinction entre les différents ARNt et d'aminoacyler leurs ARNt apparentés.

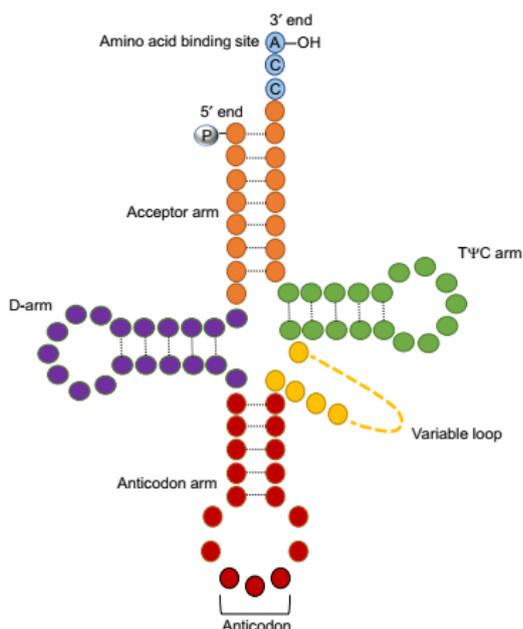


Figure 4.4 Modèle de trèfle d'ARNt avec des liaisons hydrogène représentées par des lignes pointillées (Buckingham, 2011).

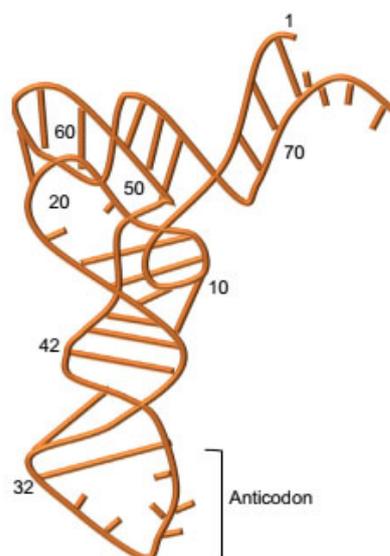


Figure 4.5 Exemple de liaisons hydrogène entre les bases de l'ARNt, qui contribue à sa structure tridimensionnelle (Buckingham, 2011).

4.4 Préparation à la traduction

Comme nous l'avons vu précédemment, l'ARNt lui-même ne porte pas d'ainoacyle et un acide aminé doit être incorporé dans l'ARNt pour qu'il devienne fonctionnel. Ainsi, deux événements importants doivent se produire avant que l'initiation de la traduction puisse avoir lieu. L'une consiste à générer un apport d'ainoacyl-ARNt, qui est l'activation de l'ARNt, et l'autre est la dissociation des ribosomes en leurs deux sous-unités.

4.4.1 Activation de l'ARNt

Avant qu'un acide aminé puisse être incorporé dans une chaîne protéique en croissance, il doit d'abord être activé. Ce processus est également appelé charge d'ARNt, qui implique à la fois des synthétases d'ARNt et d'ainoacyl-ARNt. Tous les ARNt ont les mêmes trois bases (CCA) à leurs extrémités 3', et l'adénosine terminale est la cible de la charge. Dans la première étape, l'acide aminé forme une liaison covalente avec un nucléotide adénine, fournissant un ainoacyl-AMP (Fig.4.6). L'énergie libre d'hydrolyse de l'ATP fournit de l'énergie pour la formation de liaisons, qui est piégée dans l'ainoacyl-AMP. C'est pourquoi nous appelons cela un acide aminé activé. Dans la deuxième partie de la réaction, une liaison ester est formée entre l'acide aminé et le 3'-hydroxyle ou le 2'-hydroxyle du ribose à l'extrémité 3' de l'ARNt, formant l'ainoacyl-ARNt. L'ainoacyl-ARNt est maintenant prêt à rejoindre le ribosome pour la formation de peptides.

Chez *Escherichia coli*, les ribosomes 70S sont constitués d'une sous-unité 30S et d'une sous-unité 50S. La sous-unité 30S lie l'ARNm et les extrémités anticodon des ARNt. C'est l'agent

de décodage du ribosome qui lit le code génétique dans l'ARNm et permet la liaison avec les aminoacyl-ARNt appropriés. La sous-unité 50S lie les extrémités des ARNt chargées d'acides aminés et possède l'activité peptidyl transférase qui relie les acides aminés entre eux par des liaisons peptidiques. Ainsi, les ribosomes d'*E. coli* doivent se dissocier en sous-unités à la fin de chaque cycle de traduction pour qu'un nouveau complexe d'initiation se forme. Les bactéries ont un facteur de libération de ribosomes qui agit conjointement avec un facteur d'allongement pour séparer les sous-unités. De plus, un facteur d'initiation, IF3, se lie à la petite sous-unité et l'empêche de se réassocier à la grande sous-unité avant la formation du complexe d'initiation (Watson et al, 1998).

4.4.2 Dissociation des ribosomes

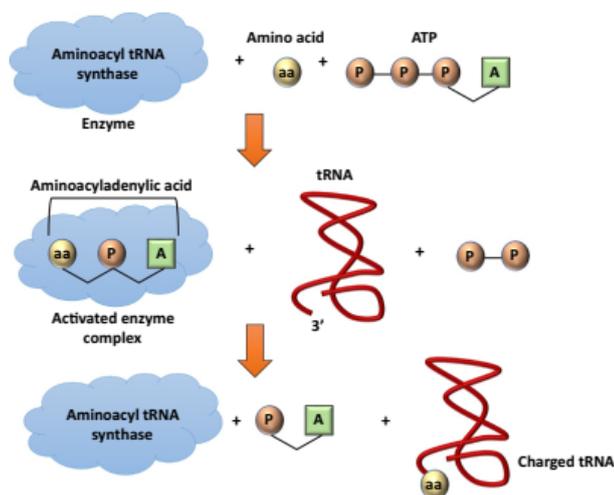


Figure 4.6 Étapes impliquées dans la charge de l'ARNt avec un acide aminé (Klug et al, 2011).

4.5 Aminoacyl-ARNt synthétases

4.5.1 Classification de l' aminoacyl-ARNt synthétase

Bien que toutes les aminoacyl-ARNt synthétases catalysent des fonctions enzymatiques communes, elles constituent un groupe diversifié de protéines en termes de taille et de séquence d'acides aminés. Ils peuvent être divisés en familles de classe I et de classe II basées sur des ensembles mutuellement exclusifs de motifs de séquence, de domaines structuraux et de fonction d'acylation (Tableau 4.3). Les synthétases de classe I et de classe II sont fonctionnellement différenciées de plusieurs façons. Les synthétases d' aminoacyl-ARNt de classe I ajoutent d'abord l'acide aminé au 2'-OH du résidu adénylate terminal de l'ARNt avant de le déplacer vers le 3'-OH. D'autre part, les synthétases d' aminoacyl-ARNt de classe II ajoutent directement l'acide aminé au 3'-OH (Fig. 4.7). Ces

deux classes d'enzymes semblent sans rapport et indiquent une évolution convergente (Wang et al, 2012).

Tableau 4.3 Classes d'aminocyl-ARNt synthétase

Class I	Class II
Arg	Ala
Cys	Asn
Gln	Asp
Glu	Gly
Ile	His
Leu	Lys
Met	Phe
Trp	Pro
Tyr	Ser
Val	Thr

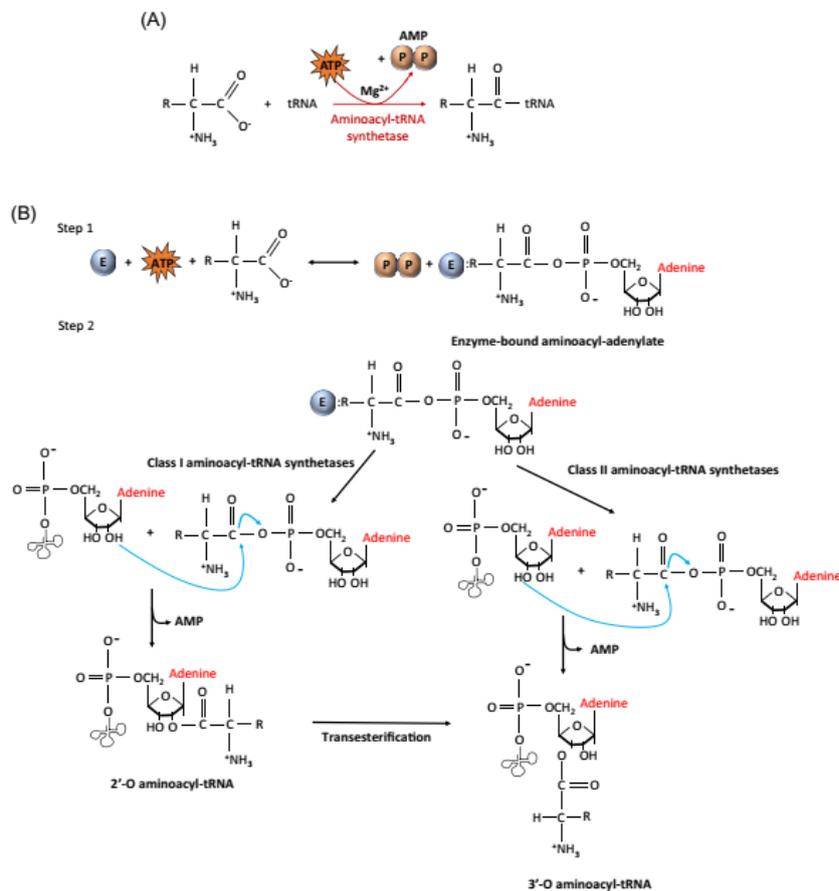


Figure 4.7 Réaction globale de l'aminocyl-ARNt synthétase (Buckingham, 2011). (A) la réaction globale. (B) La formation d'aminocyl-ARNt se déroule en deux étapes : (étape 1) formation d'un aminoacyl-adenylate et (étape 2) transfert de la fraction d'acide aminé activé de l'anhydride mixte vers les 2'-OH (aminocyltARN synthétases de classe I) ou les 3'-OH (aminocyl-ARNt synthétases de classe II) du ribose sur l'acide adénylique terminal à l'extrémité 3'-CCA commune à tous les ARNt. Ces aminoacyl-ARNt formés au niveau des esters 2'-aminoacyle subissent une transestérification qui déplace la fonction aminoacyle vers le 3'-O des ARNt. Seuls les 3'-esters sont des substrats pour la synthèse des protéines .

4.5.2 Deuxième code génétique

Plusieurs ARNt peuvent exister pour chaque acide aminé, mais un ARNt donné ne se lie pas à plus d'un acide aminé. L'enzyme aminoacyl-tRNA synthétase nécessite Mg^{2+} et est hautement spécifique à la fois pour l'acide aminé et pour l'ARNt. Chaque synthétase existe pour chaque acide aminé, et cette synthétase fonctionne pour toutes les différentes molécules d'ARNt pour cet acide aminé. Les cellules eucaryotes et certaines bactéries possèdent 20 synthétases d' aminoacyl-ARNt différentes, une pour chaque acide aminé. Chacune de ces enzymes catalyse la fixation dépendante de l'ATP de son acide aminé spécifique à l'extrémité 3' de ses molécules d'ARNt apparentées. La synthétase garantit que le bon acide aminé s'apparie avec le bon ARNt. Une fois que l' aminoacyl-ARNt a été synthétisé, la partie acide aminé ne contribue pas à une traduction précise de l'ARNm. Ceci est réalisé par la reconnaissance des codons par appariement de bases via l'anticodon de l' aminoacyl-ARNt. Par conséquent, il est important que le bon acide aminé soit associé à son ARNt approprié. Un deuxième code génétique est le code par lequel chaque aminoacyl-ARNt synthétase associe son acide aminé à des ARNt qui peuvent interagir avec des codons spécifiant son acide aminé. La présence d'un deuxième code génétique peut aider une aminoacyl-ARNt synthétase à faire la distinction entre les 20 acides aminés et les nombreux ARNt et à choisir de manière unique son substrat approprié. Les ARNt appropriés sont ceux ayant des anticodons qui peuvent s'apparier avec les codons spécifiant l'acide aminé particulier. Le deuxième code génétique est crucial pour la fidélité du transfert d'acides aminés. Par conséquent, la spécificité de l'enzyme contribue à la précision du processus de traduction. Pour la plupart des ARNt, un ensemble d'éléments de séquence est reconnu par son aminoacyl-ARNt synthase spécifique, plutôt que par un seul nucléotide distinctif ou une paire de bases. Ces éléments comprennent un ou plusieurs des éléments suivants :

- (1) au moins une base dans l'anticodon ;
 - (2) une ou plusieurs des trois paires de bases dans la tige acceptrice; et
 - (3) la base non appariée précédant l'extrémité CCA (position canonique 73), appelée base discriminatrice car cette base est invariante dans les ARNt pour un acide aminé particulier.
- L'emplacement des nucléotides qui contribuent à la reconnaissance spécifique par les aminoacyl-ARNt synthases respectives pour chacun des 20 acides aminés est indiqué dans la figure 4.8. Ces caractéristiques d'ARNt sont des exemples du deuxième code génétique qui peut aider les aminoacyl-ARNt synthétases à faire la distinction entre les divers ARNt.

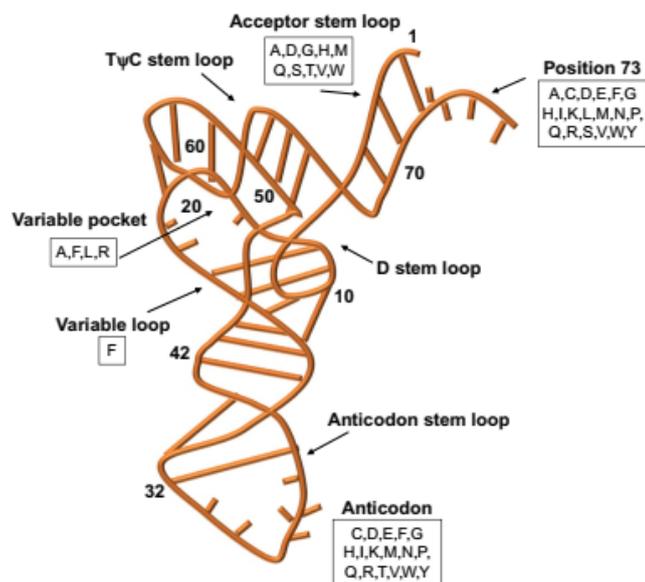


Figure 4.8 Structure tertiaire de l'ARNt représentée par un diagramme en ruban. Les nombres représentent les nucléotides reconnus par les aminoacyl-ARNt synthétases, tandis que les lettres représentent les acides aminés de la synthétase interagissant à cet endroit (Buckingham, 2011).

4.5.3 Activité de relecture des synthétases Aminoacyl-ARNt

Les synthétases d' aminoacyl-ARNt affichent un taux d'erreur global d'environ un sur 10 000. Cette très faible fréquence d'erreurs est obtenue par deux mécanismes. Premièrement, les aminoacyl-ARNt synthétases établissent une série complexe de contacts avec à la fois leur acide aminé et leur ARNt dans le site actif de l'enzyme, ce qui garantit en grande partie que seuls les substrats corrects sont sélectionnés parmi le grand pool cellulaire de candidats similaires. Deuxièmement, les enzymes possèdent une variété d'activités de relecture (édition) qui servent à hydrolyser l'acide aminé mésapparié avant ou après le transfert à l'ARNt. Le mécanisme de relecture implique une hydrolyse médiée par l'eau de l'ARNt mal chargé. Le produit correct n'est pas hydrolysé en raison de l'exclusion stérique. Ainsi, le taux d'erreur est très faible (Malentacchi et al, 2016).

4.6 Traduction procaryotique

4.6.1 Initiation à la traduction

La synthèse des protéines nécessite la liaison spécifique de l'ARNm et des aminoacyl-ARNt au ribosome. Les étapes de la traduction peuvent être divisées en initiation de chaîne, allongement de chaîne et terminaison de chaîne. Pour que l'ARNm soit traduit correctement, le ribosome doit être placé au bon emplacement de départ. L'appariement de bases entre une séquence riche en pyrimidine à l'extrémité 3' de l'ARNr 16S et des voies complémentaires

riches en purine à l'extrémité 5' des ARNm positionne la sous-unité ribosomale 30S dans un alignement approprié avec un codon d'initiation sur l'ARNm. La séquence d'ARNm riche en purine est appelée la séquence Shine-Dalgarno (5'-GGAGGU-3') et est le site de liaison du ribosome (Fig. 4.9). Le début de la traduction nécessite la formation d'un complexe d'initiation, comprenant l'ARNm, la sous-unité ribosomique 30S, le fmet-ARN^tfmet, le GTP (guanosine-5'-triphosphate) et trois facteurs d'initiation protéique, nommés IF-1, IF-2 et IF-3. Chez les procaryotes, l'acide aminé N-terminal initial de toutes les protéines est la N-formylméthionine (fmet), qui est généralement éliminée après la synthèse. Il existe deux ARN^t différents pour la méthionine dans *E. coli*, un pour la méthionine non modifiée (ARN^tmet) et un pour la N-formylméthionine (ARN^tfmet). Le ARN^tfmet reconnaît le signal de départ (AUG) lorsqu'il se produit au début de la séquence d'ARNm qui dirige la synthèse du polypeptide. Le tRNA^{met} reconnaît l'AUG lorsqu'il se trouve dans une position interne dans la séquence d'ARNm. Les aminoacyl-ARN^t qu'ils forment avec la méthionine sont appelés respectivement met-ARN^tfmet et met-ARN^tmet. Le met-ARN^tfmet est en outre formylé en fmet-ARN^tfmet.

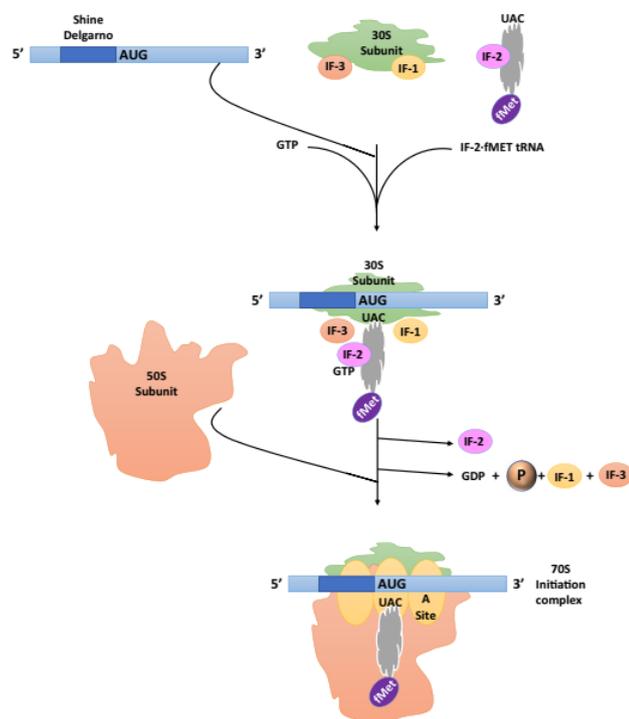


Figure 4.9 Formation du complexe d'initiation de la traduction avec l'ARNm de liaison de la sous-unité ribosomique 30S et un ARN^t initial en présence de facteurs d'initiation, IF-1, IF-2 et IF-3. La sous-unité ribosomale 50S est ensuite ajoutée pour former le complexe 70S (Chanderbali et al, 2017).

La protéine IF3 facilite la liaison de l'ARNm à la sous-unité ribosomale 30S (Fig. 4.9). Il empêche également la liaison prématurée de la sous-unité 50S. IF2 se lie au GTP et aide à la sélection de l'ARN^t initiateur (fmet-ARN^tfmet). Bien que la fonction de IF-1 soit moins

claire, il semble se lier à IF-3 et IF-2 et catalyser la séparation des sous-unités ribosomiques 30S et 50S recyclées pour un autre cycle de traduction. La combinaison de l'ARNm, de la sous-unité ribosomale 30S et du fmet-ARN^tfmet est le complexe d'initiation 30S. Une sous-unité ribosomale 50S se lie au complexe d'initiation 30S pour produire le complexe d'initiation 70S.

4.6.2 Elongation de traduction

L'allongement de la chaîne commence par l'ajout du deuxième acide aminé spécifié par l'ARNm au complexe d'initiation 70S (Fig. 4.10). La phase d'élongation utilise le fait que trois sites de liaison pour l'ARN^t sont présents sur la sous-unité 50S du ribosome 70S. Les trois sites de liaison à l'ARN^t sont appelés site P (peptidyle), site A (aminoacyle) et site E (sortie) (Coligan et al, 1995).

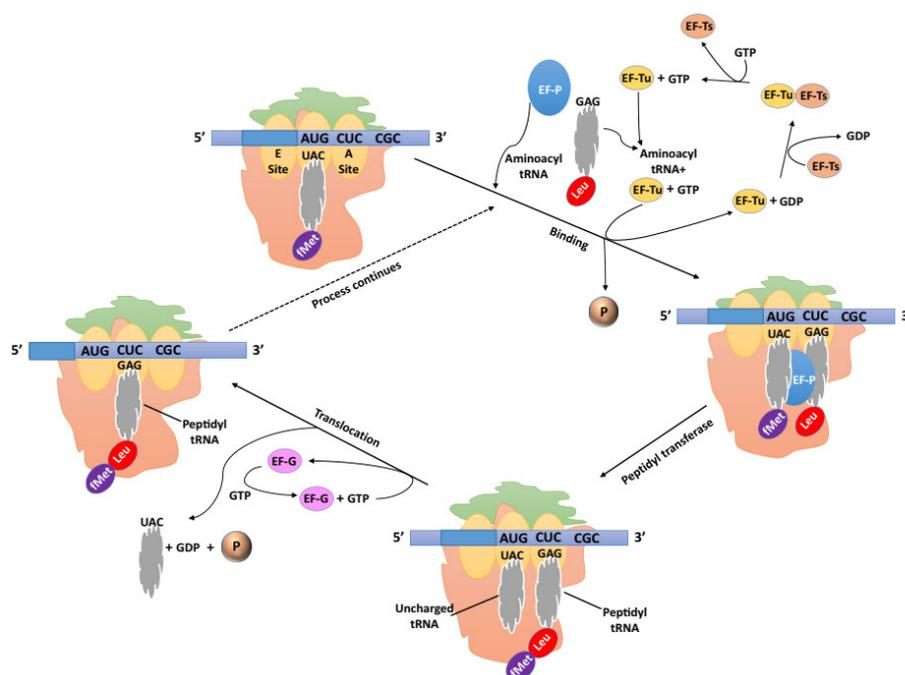


Figure 4.10 Etapes de l'élongation translationnel.

L'ARN^t initiateur, fMet, est lié au site P ribosomique, tandis qu'un second ARN^t portant l'acide aminé suivant arrive au site A. La peptidyl transférase catalyse ensuite l'élimination de l'acide aminé initiateur vers le deuxième acide aminé, la chaîne en croissance étant maintenant située dans le site A. Dans l'étape de translocation, le ribosome subit maintenant un déplacement de sorte que l'ARN^t non chargé en sortie, tandis que la chaîne peptidyle en croissance se situe au site P. Le site A est maintenant prêt à recevoir le prochain ARN^t chargé (Chanderbali et al, 2017).

Le site P lie un ARN^t qui porte une chaîne peptidique. Initialement, le site P est occupé par le fmet-ARN^tfmet dans le complexe d'initiation 70S. Parce que le site A se lie à un aminoacyl-ARN^t entrant, le second aminoacyl-ARN^t se lie au site A. Un triplet de bases d'ARN^t forme des liaisons hydrogène avec un triplet de bases d'ARNm. De plus, GTP et trois facteurs d'allongement, EF-P, EF-Tu et EF-Ts, sont requis. EF-Tu guide l'aminoacyl-

ARNt dans une partie du site A et aligne l'anticodon avec le codon d'ARNm. Lorsque la correspondance est correcte, l'aminoacyl-ARNt s'insère complètement dans le site A. Le GTP est hydrolysé et EF-Tu se dissocie. EF-P est lié adjacent au site P et au site E et on pense qu'il aide à la formation de la première liaison peptidique. EF-Ts aide à la régénération de EF-Tu-GTP. Une liaison peptidique est formée par une réaction catalysée par la peptidyl transférase. Par la suite, l'ARNt non chargé se déplace du site P vers le site E, et le peptidyl-ARNt se déplace du site A vers le site P libéré. Le site E porte donc un ARNt non chargé qui est sur le point d'être libéré du ribosome. De plus, l'ARNm se déplace par rapport au ribosome. Un autre facteur d'allongement, EF-G, est nécessaire à ce stade, et une fois de plus le GTP est hydrolysé en GDP et Pi. C'est l'étape de translocation. Par conséquent, l'allongement de la chaîne comporte trois étapes : la liaison de l'aminoacyl-ARNt, la formation de liaisons peptidiques et la translocation (Buckingham, 2011).

Lors de l'élongation en translation, le centre de décodage est la sous-unité 30S. Ce centre de décodage est l'endroit où les boucles d'anticodon des ARNt des sites A et P et les codons de l'ARNm sont appariés. Il s'agit principalement d'une propriété de l'ARNr 16S. Les changements conformationnels de l'ARNr 16S déclenchés par l'interaction entre l'ARNr 16S et l'ARNm sont les clés de la reconnaissance des codon-anticodons. D'autre part, le transfert de peptidyle est la réaction centrale de la synthèse des protéines, la véritable étape de formation de la liaison peptidique. Aucun apport d'énergie n'est nécessaire et cette réaction est une propriété de l'ARNr 23S dans la sous-unité 50S.

4.6.3 Terminaison de la traduction

La fin de la chaîne nécessite un signal d'arrêt. Les codons UAA, UAG et UGA sont les signaux d'arrêt. Ces codons sont reconnus par les facteurs libérés RF-1 ou RF-2, qui se lient respectivement à UAA et UAG ou à UAA et UGA (Fig. 4.11). D'autre part, RF-3 facilite l'activité de RF-1 et RF-2. RF-1 ou RF-2 est lié près du site A du ribosome lorsqu'un des codons de terminaison est atteint. Le facteur libéré bloque non seulement la liaison d'un nouvel aminoacyl-ARNt mais affecte également l'activité de la peptidyl transférase de sorte que la liaison entre l'extrémité carboxyle du peptide et l'ARNt est hydrolysée. Par la suite, l'ensemble du complexe se dissocie, mettant libérer les facteurs de libération, l'ARNt, l'ARNm et les sous-unités ribosomiques 30S et 50S. Tous ces composants peuvent être réutilisés dans d'autres synthèses de protéines.

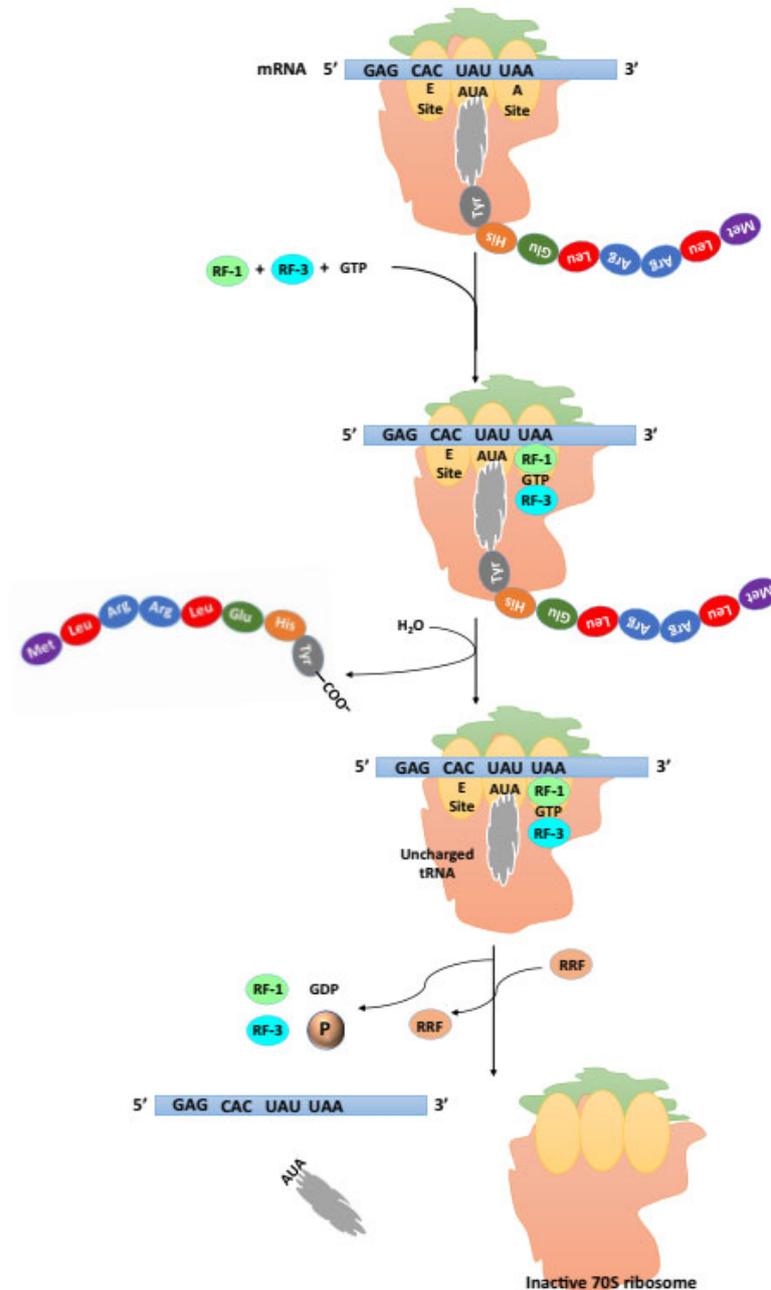


Figure 4.11 Terminaison de la chaîne peptidique dans laquelle il n'y a pas de molécules d'ARNt capables de lire les codons d'arrêt, qui sont ensuite lus par le facteur de libération (RF) au lieu de l'ARNt (Chanderbali et al, 2017).

4.6.4 Traduction couplée

Chez les procaryotes, la traduction commence très peu de temps après la transcription de l'ARNm. Le gène est simultanément transcrit et traduit, un processus appelé traduction couplée. Dans cette situation, plusieurs molécules d'ARN polymérase sont attachées à un même gène, donnant lieu à plusieurs molécules d'ARNm qui sont transcrites. Alors que chaque molécule d'ARNm est toujours en cours de transcription, un certain nombre de

ribosomes s'y attachent qui sont à divers stades de traduction de cet ARNm. Ce complexe d'ARNm à plusieurs ribosomes est appelé polysome ou polyribosome (Fig. 4.12).

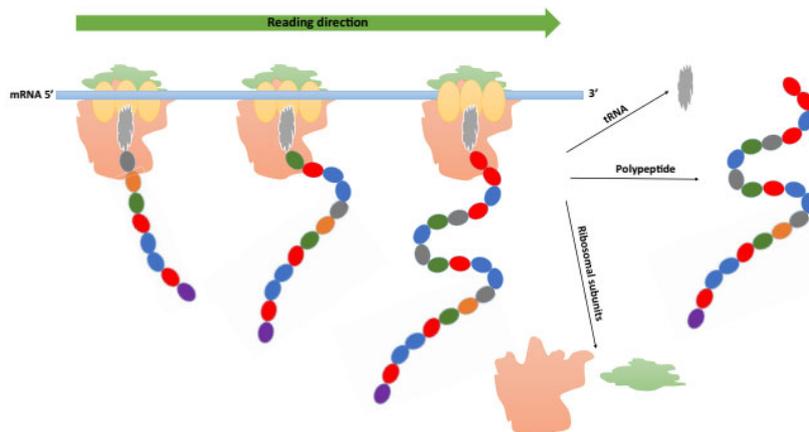


Figure 4.12 Polysomes procaryotes permettent la synthèse simultanée de protéines le long d'une seule molécule d'ARNm (Chanderbali et al, 2017).

4.7 Traduction eucaryotique

4.7.1 Initiation

Les principales caractéristiques de la traduction sont les mêmes chez les procaryotes et les eucaryotes, à l'exception des deux principales modifications post-transcriptionnelles : 5' cap et 3' poly A tail. La chaîne d'initiation de la traduction eucaryote est la partie la plus différente des procaryotes. L'étape 1 de l'initiation de la chaîne implique l'assemblage d'un complexe de préinitiation 43S composé de met-ARNt, de sous-unité 40S, de GTP et de facteurs d'initiation eucaryotes, notamment eIF1A, eIF2 et eIF3 (Fig. 4.13). A l'étape 2, l'ARNm est recruté. Le capuchon 5' oriente le ribosome vers le bon AUG via un mécanisme de balayage. Les eIF4A, eIF4B, eIF4E et eIF4G aident le complexe de préinitiation 43S à se lier à l'ARNm. La sous-unité 40S, l'ARNm et les sept eIF constituent le complexe de préinitiation 48S. Il se déplace vers l'aval jusqu'à ce qu'il rencontre le premier AUG dans le bon contexte. À l'étape 3, le ribosome 60S est recruté, formant le complexe d'initiation 80S. Le GTP est hydrolysé et les facteurs d'initiation sont libérés (Klug et al, 2011).

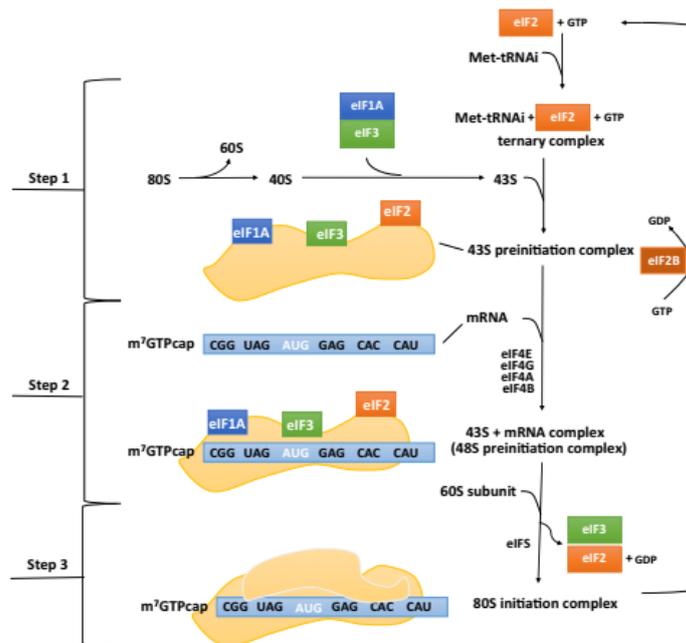


Figure 4.13 Etapes de l'initiation eucaryote de la traduction (Chanderbali et al, 2017).

4.7.2 Élongation

L'élongation de la chaîne chez les eucaryotes est très similaire à celui des procaryotes. Le même mécanisme de peptidyl transférase et de translocation des ribosomes est observé. Cependant, la structure du ribosome eucaryote est différente de celle du ribosome procaryote en ce qu'il n'y a pas de site E chez les eucaryotes, seulement les sites A et P. Il existe deux facteurs d'élongation eucaryotes, eEF1 et eEF2. L'eEF1 se compose de deux sous-unités, eEF1A et eEF1B. La sous-unité 1A est la contrepartie de EF-Tu chez les procaryotes, et la sous-unité 1B est l'équivalent des EF-T chez les procaryotes. La protéine eEF2 est la contrepartie de l'EF-G chez les procaryotes qui provoque la translocation.

4.7.3 Terminaison

Comme dans la terminaison procaryote, le ribosome rencontre un codon stop : UAA, UAG ou UGA. Contrairement aux procaryotes, où trois facteurs de libération sont impliqués, chez les eucaryotes, un seul facteur de libération (RF) se lie aux trois codons stop et catalyse l'hydrolyse de la liaison entre l'acide aminé C-terminal et l'ARNt. La liaison du RF eucaryote au site ribosomique A dépend du GTP et le RF:GTP se lie à ce site lorsqu'il est occupé par un codon de terminaison. Ensuite, l'hydrolyse de la liaison ester peptidyl-ARNt, l'hydrolyse du GTP, la libération du polypeptide naissant et de l'ARNt désacylé et la dissociation du ribosome de l'ARNm se produisent en séquence (Klug et al, 2011).

Exercice 01 :

Hershey et Chase ont démontré que l'ADN était porteur de l'information génétique. Leurs expériences portaient sur le bactériophage T2, composé d'un noyau d'ADN entouré d'une capsidie protéique. Lors de l'infection, le virus se fixe à la cellule, injecte son matériel génétique mais ne pénètre pas. Afin d'identifier la nature du matériel génétique injecté, les chercheurs ont radiomarqué de manière spécifique les acides nucléiques ou les protéines du virus.

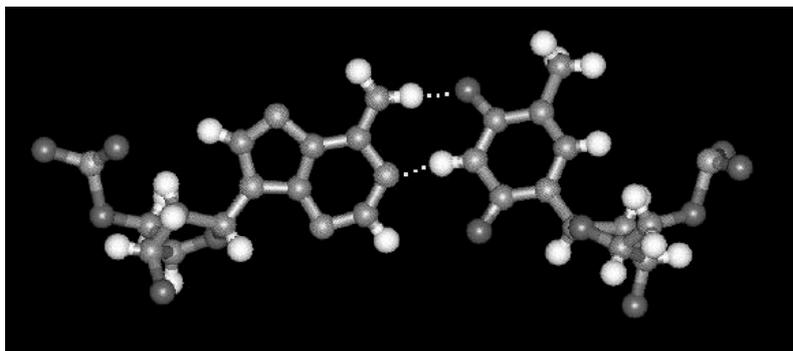
Quels radioisotopes et quelles molécules précurseurs radioactives vont-ils utiliser pour marquer l'ADN ou les protéines ?

Des bactéries sont infectées durant un temps court avec des phages T2 marqués au niveau des protéines ou de l'ADN puis la suspension est fortement agitée afin de détacher le virus de la cellule. Les bactéries sont ensuite séparées des virus.

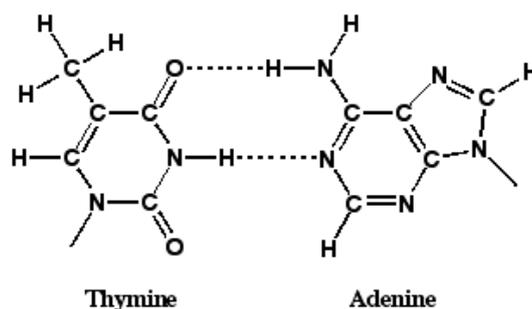
Quelle technique est utilisée pour cette séparation ? Le contenu en radioactivité des bactéries et des phages est ensuite mesuré. Donnez le résultat de cette expérience.

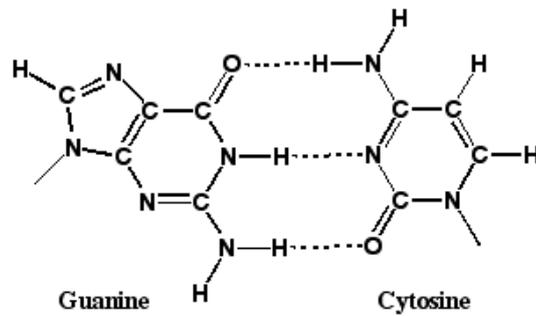
Exercice 02 :

1. Analysez le modèle de paires de bases de la figure. Identifiez la position des purines et des pyrimidines, du sucre et des phosphates. De quelle paire de nucléotides s'agit-il ?



2. Observez la structure des paires de bases AT et GC.





- Indiquez la position de la liaison N- glycosidique
- Indiquez avec un trait la localisation, du grand et du petit sillon, sur la structure.
- Quels sont les atomes présents dans les sillons et susceptibles d'établir des liaisons hydrogènes avec des protéines ?

Exercice 03 :

L'analyse de la composition de l'ADN isolé à partir du bactériophage M13 donne comme résultat : 25%A ; 33%T ; 22%C et 20%G.

Ce résultat vous semble-t-il cohérent ?

Quelles informations vous suggèrent ces résultats sur la structure de cet ADN ?

Comment cet ADN peut-il se répliquer ?

Exercice 04 :

Après une hydrolyse complète d'une solution d'ARNt, on a pu identifier dans l'hydrolysat : le nucléoside A, le nucléoside diphosphate pGp et des nucléotides Ap, Gp, Cp, Up. Quelles sont les extrémités 5' et 3' de cet ARNt ?

Exercice 05 :

Une fibre d'ADN est transférée d'un environnement à forte humidité dans un environnement à faible humidité et passe donc de la forme B à la forme A. Quel est son pourcentage de raccourcissement sachant que dans la forme A le pas de l'hélice n'est plus que de 2,8 nm et couvre une longueur de 11 nucléotides ?

Exercice 01 :

Représentez à l'aide d'un schéma les caractéristiques principales de la réplication chez *E.coli* en indiquant :

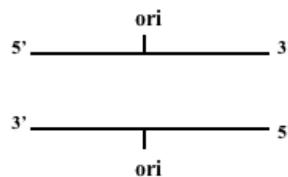
- La direction de la synthèse de l'ADN par rapport à l'orientation du brin d'ADN matrice.
- Les propriétés des brins avancés, retardés et des fragments d'Okazaki.
- Les enzymes impliquées dans chaque étape.

Exercice 02 :

L'ADN de *E. coli* contient 4,5.10⁶ paires de bases. Si l'ADN polymérase progresse à la vitesse de 1000 nucléotides par seconde, en combien de temps le chromosome circulaire entier de *E. coli* est-il répliqué ? Dans certaines conditions optimales de croissance, le temps de génération peut être de 20 minutes seulement. Etant donné que la vitesse de polymérisation reste la même et qu'il n'y a qu'un seul site d'origine de réplication sur l'ADN circulaire parental, comment peut-on envisager une telle rapidité ? Dans ces conditions de croissance rapide, certaines régions du chromosome bactérien sont présentes en un plus grand nombre de copies que d'autres. Lesquelles ?

Exercice 03 :

Une origine de réplication est représentée ci-dessous sur l'ADN double brin. Indiquez la position de chaque nouveau brin en cours de synthèse ainsi que la direction de la synthèse répllicative.



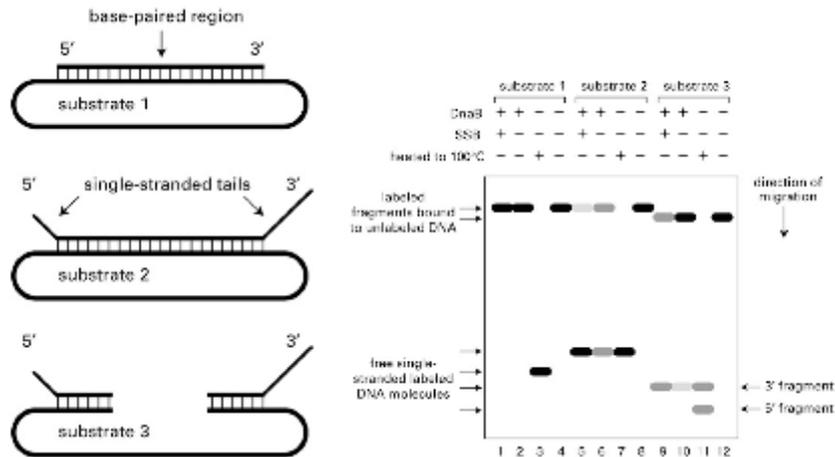
Exercice 04 :

Si la fourche de réplication progresse à une vitesse de 500 nucléotides par seconde, combien de rotation (tour d'hélice) par minute doit effectuer la double hélice d'ADN, qui précède le site de la réplication, pour se dérouler ?

Exercice 05 :

DnaB code pour une hélicase qui déroule l'ADN à la fourche de la réplication. Ses propriétés ont été analysées *in vitro* en incubant pendant le même temps la protéine purifiée,

de l'ATP et les substrats comme ceux qui sont montrés dans la figure de gauche. Pour identifier chaque substrat, les fragments 5' et 3' d'ADN sont marqués au ³²P et incubés avec l'hélicase en présence ou en absence de la protéine SSB. Les produits d'ADN obtenus après la réaction sont par la suite séparés d'après leurs tailles et formes sur un gel d'agarose. Le résultat de l'autoradiographie est montré dans la figure de droite :



A partir de vos connaissances, déterminer : -

- Quels substrats sont utilisés efficacement par l'hélicase ?
- Dans quelle direction l'hélicase se déplace sur l'ADN ?
- Pourquoi l'hydrolyse de l'ATP est-elle nécessaire pour la réaction ? –
- Si on rajoute à la réaction de la protéine SSB avant de rajouter DnaB, on inhibe la réaction. Si, au contraire, on rajoute SSB après DnaB, la réaction de déroulement par l'hélicase est stimulée. Expliquer ces résultats.

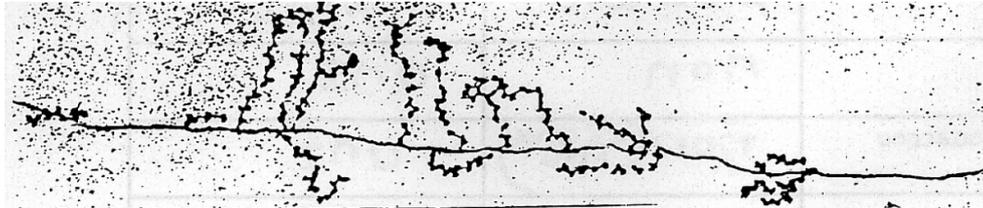
Exercice 06 :

Les mutations conditionnelles sont très utiles pour caractériser un processus biologique. Dans le cas de la réplication, un grand nombre de mutants thermosensibles (*ts*) ont été isolés et caractérisés. Ces mutants conduisent à deux types de phénotypes d'arrêt de la division cellulaire : l'arrêt rapide de la synthèse de l'ADN ou un arrêt lent.

- Donnez votre prédiction sur le type de phénotype d'arrêt (rapide ou lent) pour les mutants suivants : DNA topoisomérase, DnaA, SSB, DnaB, DnaG, DNA ligase en justifiant votre réponse.
- Si vous préparez un extrait cellulaire à partir d'un mutant *ts* qui arrête la synthèse rapidement et d'un mutant qui arrête la synthèse lentement et si vous les mélangez dans un système de réplication *in vitro* quel effet allez-vous obtenir sur la réplication de l'ADN ?

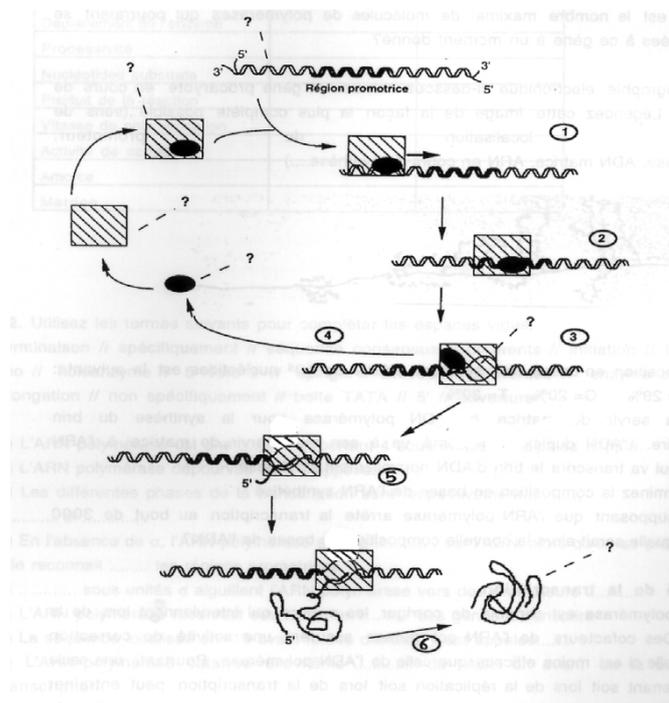
Exercice 01 :

La photographie prise au microscope électronique montre un gène procaryote en cours de transcription. Légendez cette image de la façon la plus complète possible (sens de transcription, localisation du promoteur, ARN polymérase, ADN matrice, ARN en cours de synthèse)



Exercice 02 :

Complétez ce schéma représentant les différentes étapes de la transcription. Légendez chacune des étapes numérotées. Nommez les éléments marqués par un “ ? ”.



Exercice 03 :

Prédisez l'effet des mutations suivantes :

- 1) une mutation du facteur sigma empêchant sa dissociation de l'enzyme core.
- 2) une mutation de sigma permettant à ce facteur de se lier aux séquences promoteur en l'absence des autres sous unités de l'ARN polymérase.

Exercice 01 :

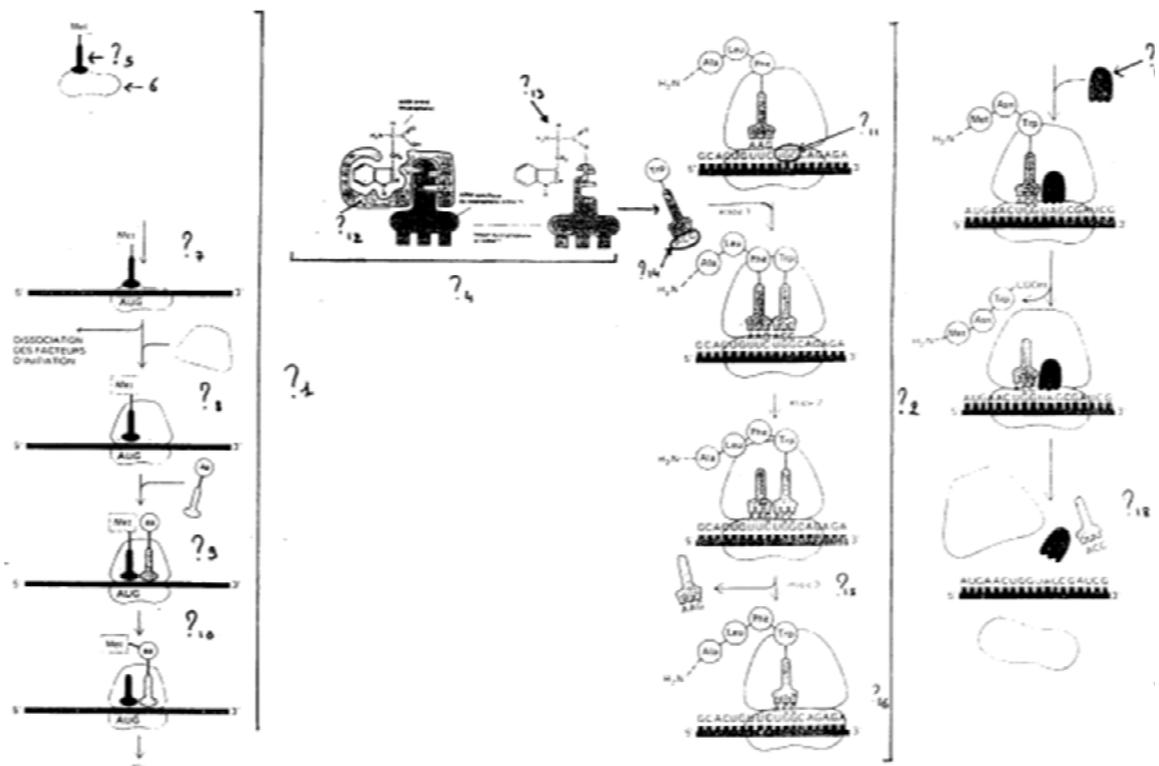
La séquence en bases d'un ARNm peut-elle être prédite grâce à la séquence en aminoacides du polypeptide unique lui correspondant ? Par exemple, combien d'ARNm peuvent coder pour la séquence tripeptidique LeuMet-Tyr ? Un seul ARNt peut-il reconnaître tous les codons de la leucine ?

Exercice 02 :

Une mutation au sein d'un gène bactérien conduit à l'apparition d'un codon stop UGA au milieu de la séquence codant pour la protéine. Une mutation suppresseur dans un tRNA permet cependant de produire une protéine de taille correcte. Ce tRNA muté introduit un acide aminé tryptophane en regard du codon UGA. Quelle est la nature de la mutation sur le tRNA ? Quelle sera la conséquence de cette mutation sur la traduction des autres gènes bactériens ?

Exercice 03 :

Schéma récapitulatif à légender. Donnez le nom des éléments et des étapes réactionnelles marquées d'un " ? "



1. Agrawal, M., Biswas, A., 2015. Molecular diagnostics of neurodegenerative disorders. *Front. Mol. Biosci.* 2, 54. DOI: 10.3389/fmolb.2015.00054
2. Andersson, D.I., Jernstrom-Hultqvist, J., Nasvall, J., 2015. Evolution of new functions de novo and from preexisting genes. *Cold Spring Harb. Perspect. Biol.* 7, a017996. DOI: 10.1101/cshperspect.a017996
3. Brown, T.A., 2002. *Genomes*, second ed. Wiley-Liss, Oxford. 600 p.
4. Brown, T.A., 2016. *Gene Cloning and DNA Analysis: An Introduction*, seventh ed. Wiley-Blackwell. 376 p.
5. Buckingham, L., 2011. *Molecular Diagnostics: Fundamentals, Methods and Clinical Applications*, second ed. F.A. Davis Company. 558 p.
6. Catalina, P., Cobo, F., Cortés, J.L., Nieto, A.I., Cabrera, C., Montes, R., Concha, A., Menendez, P., 2007. Conventional and molecular cytogenetic diagnostic methods in stem cell research: a concise review. *Cell Biol. Int.* 31, 861–869. <https://doi.org/10.1016/j.cellbi.2007.03.012>
7. Chandrabali, A.S., Berger, B.A., Howarth, D.G., Soltis, D.E., Soltis, P.S., 2017. Evolution of floral diversity: genomics, genes and gamma. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 372(1713). DOI: 10.1098/rstb.2015.0509
8. Cheema, M.S., Ausio, J., 2015. The structural determinants behind the epigenetic role of histone variants. *Genes* 6, 685–713. DOI: 10.3390/genes6030685
9. Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W., Wingfield, P.T., 1995. *Current Protocols in Protein Science*. John Wiley & Sons, Inc. 864 p.
10. Daban, J.R., 2011. Electron microscopy and atomic force microscopy studies of chromatin and metaphase chromosome structure. *Micron* 42, 733–750. DOI: 10.1016/j.micron.2011.05.002
11. Dogini, D.B., Pascoal, V.D., Avansini, S.H., Vieira, A.S., Pereira, T.C., Lopes-Cendes, I., 2014. The new world of RNAs. *Genet. Mol. Biol.* 37(1 Suppl.), 285–293. DOI: 10.1590/s1415-47572014000200014
12. Downie, A.W., 1972. Pneumococcal transformation—a backward view. Fourth Griffith Memorial Lecture. *J. Gen. Microbiol.* 73, 1–11. DOI: 10.1099/00221287-73-1-1
13. Hayes, P.C., Roland Wolf, C., Hayes, J.D., 1989. Blotting techniques for the study of DNA, RNA, and proteins. *BMJ* 299, 965–968. DOI: 10.1136/bmj.299.6705.965
14. Jamali, S., Eskandari, N., Aryani, O., Salehpour, S., Zaman, T., Kamalidehghan, B., Houshmand, M., 2014. Three novel mutations in Iranian patients with Tay-Sachs disease. *Iran. Biomed. J.* 18 (2), 114–119. Doi: 10.6091/ibj.1137.2013
15. Jeon, K. (Ed.), 2012. *International Review of Cell and Molecular Biology*, first ed., vol. 293. Academic Press, 328 p.
16. Kleckner, N., Fisher, J.K., Stouf, M., White, M.A., Bates, D., Witz, G., 2014. The bacterial nucleoid: nature, dynamics and sister segregation. *Curr. Opin. Microbiol.* 22, 127–137.
17. Klug, W.S., Cummings, M.R., Spencer, C.A., Palladino, M.A., 2011. *Concepts of Genetics*, tenth ed. Benjamin Cummings. 895 p.
18. Li, X., Zhao, J., Yuan, Q., Xia, N., 2017. Detection of HBV covalently closed circular DNA. *Viruses* 9 (6), 139. DOI: 10.3390/v9060139
19. Long, M., VanKuren, N.W., Chen, S., Vibranovski, M.D., 2013. New gene evolution: little did we know. *Annu. Rev. Genet.* 47, 307–333. DOI: 10.1146/annurev-genet-111212-133301
20. Malentacchi, F., Pizzamiglio, S., Wyrich, R., Verderio, P., Ciniselli, C., Pazzagli, M., Gelmini, S., 2016. Effects of transport and storage conditions on gene expression in blood samples. *Biopreserv. Biobank.* 14 (2), 122–128.
21. Maze, I., Noh, K.M., Soshnev, A.A., Allis, C.D., 2014. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nat. Rev. Genet.* 15 (4), 259–271. DOI: 10.1038/nrg3673
22. Mehta, N., Lazarin, G.A., Spiegel, E., Berentsen, K., Brennan, K., Giordano, J., Haque, I.S., Wapner, R., 2016. Tay-Sachs carrier screening by enzyme and molecular analyses in the New York City Minority population. *Genet. Test. Mol. Biomarkers* 20 (9), 504–509.
23. Meisner, L.F., Johnson, J.A., 2008. Protocols for cytogenetic studies of human embryonic stem cells. *Methods* 45, 133–141. DOI: 10.1016/j.ymeth.2008.03.005
24. Mutz, K.-O., Heilkenbrinker, A., Léonne, M., Walter, J.-G., Stahl, F., 2013. Transcriptome analysis using next-generation sequencing. *Curr. Opin. Bio-technol.* 24, 22–30.

25. Nigro, E., Imperlini, E., Scudiero, O., Monaco, M.L., Polito, R., Mazzarella, G., Orru', S., Bianco, A., Daniele, A., 2015. Differentially expressed and activated proteins associated with non small cell lung cancer tissues. *Respir. Res.* 16, 74. <https://doi.org/10.1186/s12931-015-0234-2>
26. Persing, D.H., Tenover, F.C., Tang, Y.-W., Nolte, F.S., Hayden, R.T., Van Belkum, A. (Eds.), 2011. *Molecular Microbiology: Diagnostic Principles and Practice*, second ed. ASM Press. 936p.
27. Radvanszky, J., Surovy, M., Nagyova, E., Minarik, G., Kadasi, L., 2015. Comparison of different DNA binding fluorescent dyes for applications of high- resolution melting analysis. *Clin. Biochem.* 48, 609–616. DOI: 10.1016/j.clinbiochem.2015.01.010
28. Rainen, L., Oelmueller, U., Jurgensen, S., Wyrich, R., Ballas, C., Schram, J., Herdman, C., Bankaitis-Davis, D., Nicholls, N., Trollinger, D., Tryon, V., 2002. Stabilization of mRNA expression in whole blood samples. *Clin. Chem.* 48 (11), 1883–1890.
29. Rajan-Babu, I.-S., Chong, S.S., 2016. Molecular correlates and recent advancements in the diagnosis and screening of FMR1-related disorders. *Genes (Basel)* 7 (10), 87. DOI: 10.3390/genes7100087
30. Rallapalli, G., Kemen, E.M., Robert-Seilaniantz, A., Segonzac, C., Etherington, G.J., Sohn, K.H., MacLean, D., Jones, J.D.G., 2014. EXPRSS: an Illumina based high-throughput expression-profiling method to reveal transcriptional dynamics. *BMC Genomics* 15, 341.
31. Shapiro, B., Hofreiter, M. (Eds.), 2011. *Ancient DNA: Methods and Protocols*, *Methods in Molecular Biology*. vol. 840. https://doi.org/10.1007/978-1-61779-516-9_11.
32. Steitz, T.A., 1999. DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* 274, 17395–17398. DOI: 10.1074/jbc.274.25.17395
33. Wang, Y., Hayatsu, M., Fujii, T., 2012. Extraction of bacterial RNA from soil: challenges and solutions. *Microbes Environ.* 27 (2), 111–121. DOI: 10.1264/jsme2.ME11304
34. Watson, A., Mazumder, A., Stewart, M., Balasubramanian, S., 1998. Technology for microarray analysis of gene expression. *Curr. Opin. Biotechnol.* 9, 609–614. DOI: 10.1016/s0958-1669(98)80138-9
35. Weaver, R.F., 2011. *Molecular Biology*, fifth ed. McGraw-Hill Education. 912 pp.
36. Yan, M., Bai, W., Zhu, C., Huang, Y., Yan, J., Chen, A., 2016. Design of nuclease-based target recycling signal amplification in aptasensors. *Biosens. Bioelectron.* 77, 613–623. DOI: 10.1016/j.bios.2015.10.015
37. Zaher, H.S., Green, R., 2009. Fidelity at the molecular level: lessons from protein synthesis. *Cell* 136 (4), 746–762. DOI: 10.1016/j.cell.2009.01.036