



VOS OUTILS INTERACTIFS



Consultez votre
MANUEL NUMÉRIQUE,
qui vous donne accès
aux animations,
aux exercices et à la
plateforme d'anatomie interactive.

▲ **Figure 17.1** Comment un simple gène défectueux peut-il causer l'aspect étonnant de ces ânes albinos ?

CONCEPTS CLÉS

- 17.1** Les gènes codent pour les protéines par l'intermédiaire de la transcription et de la traduction
- 17.2** La transcription est la synthèse de l'ARN à partir de l'ADN : *une étude détaillée*
- 17.3** Dans les cellules eucaryotes, l'ARN est modifié après avoir été transcrit
- 17.4** La traduction est la synthèse d'un polypeptide à partir de l'ARN messenger : *une étude détaillée*
- 17.5** Les mutations d'un ou de quelques nucléotides peuvent modifier la structure et la fonction des protéines



▲ Un raton laveur albinos.

La transmission de l'information génétique

Asinara est une île italienne au large de la Sardaigne. Tirant probablement son origine du mot latin *sinuaria* (« en forme de sinus »), son nom signifie également « habitée par des ânes ». Cette définition lui convient parfaitement puisqu'elle abrite une population sauvage d'ânes albinos (**figure 17.1**). Quels sont les facteurs responsables du phénotype de l'albinisme ?

Les caractères transmis sont déterminés par les gènes et le caractère de l'albinisme est produit par un allèle récessif du gène de la pigmentation (voir le concept 14.4). L'information contenue dans les gènes se présente sous la forme de séquences nucléotidiques précises, alignées sur les brins d'ADN, c'est-à-dire le matériel génétique. L'âne albinos possède une version défectueuse d'une protéine, une enzyme requise pour la synthèse des pigments, et cette protéine est défectueuse parce que le gène codant pour celle-ci contient des informations erronées.

Cet exemple illustre le thème principal de ce chapitre : c'est en dictant la synthèse de certaines protéines que l'ADN d'un organisme produit des caractères spécifiques. Autrement dit, les protéines représentent le lien entre le génotype et le phénotype. **L'expression génétique** est le processus par lequel l'ADN régit la synthèse des protéines (ou, dans certains cas, seulement des ARN). L'expression des gènes qui codent pour les protéines comporte deux étapes principales : la transcription et la traduction. Le présent chapitre décrit en détail la transmission de l'information des gènes aux protéines et explique comment les mutations génétiques influent sur les organismes en modifiant leurs protéines. À la fin de ce chapitre, quand vous aurez compris le processus des mutations génétiques, nous étudierons en détail le concept de gène.

Les gènes codent pour les protéines par l'intermédiaire de la transcription et de la traduction

Avant d'étudier en détail la façon dont les gènes dirigent la synthèse des protéines, prenons le temps d'examiner comment la relation fondamentale entre gènes et protéines a été découverte.

Une preuve à partir de l'étude de maladies métaboliques

En 1902, le médecin britannique Archibald Garrod a émis l'hypothèse selon laquelle les gènes déterminent les phénotypes par l'intermédiaire d'enzymes, c'est-à-dire des protéines qui catalysent certaines réactions chimiques précises dans la cellule. Il a posé comme postulat que les symptômes des maladies héréditaires reflètent une incapacité à produire une enzyme particulière. Il a qualifié celles-ci d'« erreurs innées du métabolisme ». Par exemple, les personnes atteintes d'alcaptonurie produisent une urine qui paraît noire parce qu'elle contient de l'homogentisate (un sel autrefois appelé alcaptone), qui devient foncé au contact de l'air. Garrod a supposé que les individus normaux produisent une enzyme qui dégrade l'homogentisate, tandis que les personnes alcaptonuriques ont hérité d'une incapacité à fabriquer cette enzyme, de sorte que l'alcaptone est excrétée dans l'urine.

Des recherches effectuées plusieurs décennies plus tard ont permis de confirmer l'hypothèse *un gène, une enzyme* de Garrod selon laquelle un gène dicte la production d'une enzyme spécifique. Les biochimistes ont appris que les cellules synthétisent et dégradent la plupart des molécules organiques : elles empruntent des voies métaboliques dans lesquelles chacune des réactions chimiques d'une séquence particulière est catalysée par une enzyme spécifique (voir le concept 8.1). Ce sont ces voies métaboliques qui mènent, par exemple, à la synthèse des pigments qui confèrent une couleur donnée au pelage des ânes bruns de la figure 17.1 ou aux yeux des drosophiles (voir la figure 15.3). Dans les années 1930, George Beadle, un biochimiste et généticien américain, et Boris Ephrussi, son collègue français, ont avancé que chacune des mutations affectant la couleur de yeux des drosophiles bloque la synthèse d'un pigment particulier. Ce blocage survient à un stade spécifique et empêche la production de l'enzyme catalysant l'étape correspondante. Mais on ignorait alors tout des réactions chimiques en question et des enzymes qui les catalysent.

Quelques années plus tard, à la Stanford University, George Beadle et Edward Tatum ont fait une découverte décisive. Ils ont effectué leurs recherches sur la moisissure rouge du pain, *Neurospora crassa*, une espèce haploïde. Pour qu'un changement se produise dans le phénotype d'un mutant, Beadle et Tatum devaient désactiver un seul allèle (plutôt que deux, comme pour les espèces diploïdes) d'un gène codant pour une protéine nécessaire à une activité métabolique spécifique. Ils ont donc exposé *Neurospora* aux rayons X, reconnus pour causer des modifications génétiques, puis ils ont cherché, parmi les organismes survivants, des mutants qui n'avaient pas les mêmes besoins nutritionnels que les individus du type sauvage.

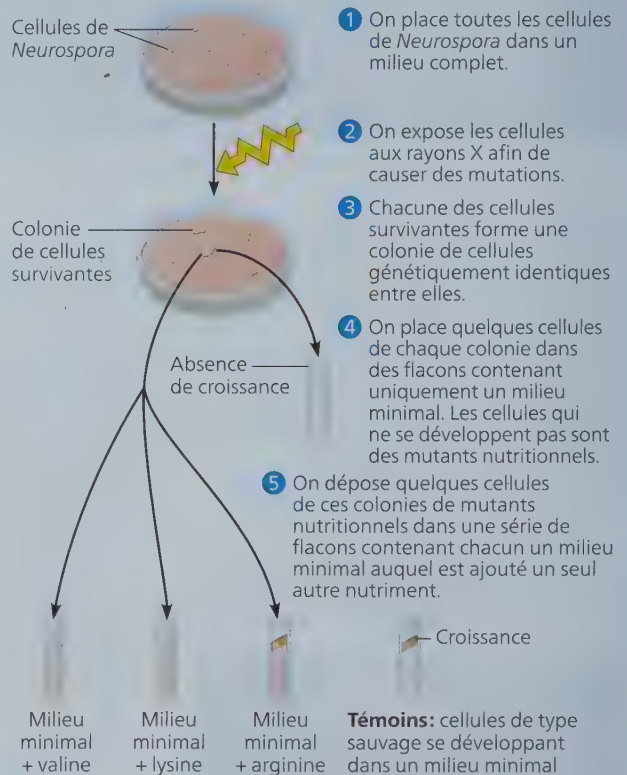
Les mutants auxotrophes de *Neurospora*

Un auxotrophe est un mutant incapable de synthétiser par lui-même un nutriment indispensable à sa croissance ; il faut donc le lui fournir.

Les besoins en nutriments de *Neurospora* de type sauvage sont très simples : en laboratoire, sa croissance ne requiert que très peu de nutriments : il lui suffit d'un mélange de sels inorganiques, de glucose et de biotine, une vitamine, incorporés à un milieu de culture solidifié par de l'agar. À partir de ce *milieu minimal*, les cellules de la moisissure de type sauvage mettent à profit leurs activités métaboliques pour produire toutes les molécules nécessaires à leur croissance, y compris les acides aminés. Ainsi, elles se divisent à plusieurs reprises et forment des colonies visibles de cellules génétiquement identiques. Comme le montre la **figure 17.2**, Beadle et Tatum ont obtenu différents « mutants nutritionnels » de *Neurospora*. Chacun de ces mutants était incapable de synthétiser un nutriment essentiel particulier. Si les cellules ne pouvaient croître dans le milieu minimal, elles pouvaient se développer dans le milieu complet contenant tous les nutriments nécessaires à leur croissance. Dans le cas de *Neurospora*, un milieu complet était un milieu minimal dans

▼ Figure 17.2 L'approche expérimentale de Beadle et Tatum.

Pour obtenir des mutants nutritionnels, Beadle et Tatum ont exposé la moisissure *Neurospora* à des rayons X causant des mutations. Ils ont ensuite analysé les mutants présentant de nouveaux besoins nutritionnels, dont celui en arginine illustré dans cette figure.



6 On examine les flacons afin de vérifier la croissance des cellules. Dans cet exemple, les cellules se sont développées uniquement dans le milieu minimal auquel on avait ajouté de l'arginine, ce qui démontre l'absence de l'enzyme assurant la synthèse de cet acide aminé chez ce mutant.

lequel on avait incorporé les 20 acides aminés et quelques autres nutriments. Beadle et Tatum ont émis l'hypothèse que chez tous les mutants nutritionnels, le gène codant pour une enzyme synthétisant un nutriment donné était désactivé.

Grâce à cette approche, Beadle et Tatum ont créé une vaste collection de souches mutantes de *Neurospora*, qu'ils ont ensuite

classées selon le type de dysfonctionnement affectant une voie métabolique donnée. Deux collègues de Beadle et Tatum, Adrian Srb et Norman Horowitz, ont utilisé une collection de mutants qui avaient besoin d'arginine pour étudier la voie biochimique de la synthèse de cet acide aminé chez *Neurospora* (figure 17.3). Srb et Horowitz ont entrepris de caractériser la

DÉMARCHE SCIENTIFIQUE

INVESTIGATION

Des gènes individuels codent-ils pour la production d'enzymes qui participent à une voie biochimique ?

■ **HYPOTHÈSE** ■ S'il est juste de penser que chacune des enzymes d'une voie métabolique est encodée dans un gène spécifique, alors des mutants pour différents gènes devraient présenter des exigences métaboliques spécifiques, selon l'enzyme affectée et le rôle qu'elle exerce dans la voie métabolique.

■ **EXPÉRIENCE** ■ Au cours de leurs travaux sur la moisissure rouge du pain, *Neurospora crassa*, Adrian Srb et Norman Horowitz, alors à la Stanford University, ont utilisé l'approche expérimentale (voir la figure 17.2) de Beadle et Tatum pour isoler des mutants nécessitant de l'arginine dans leur milieu de culture. Les chercheurs ont établi que leurs mutants se regroupaient en trois catégories, chacune de celles-ci portant une mutation sur un gène différent. En se basant sur des études menées par d'autres chercheurs et portant sur des cellules hépatiques de mammifères, ils ont supposé, d'une part, que la voie métabolique de la biosynthèse de l'arginine comportait un nutriment précurseur et, d'autre part, que l'ornithine et la citrulline constituaient des molécules intermédiaires de cette voie, comme l'illustre le schéma de droite.

Dans leur expérience la plus célèbre, que nous illustrons ci-dessous, ils ont mis à l'épreuve leur hypothèse *un gène, une enzyme* et leur postulat sur la voie de synthèse de l'arginine. Dans cette expérience, ils ont placé leurs trois classes de mutants et la souche de type sauvage dans quatre milieux de croissance différents, illustrés dans le tableau des résultats. Ils ont inclus le milieu minimal (MM) comme témoin parce qu'ils savaient que les cellules de type sauvage pouvaient croître sur le MM, mais que les cellules des mutants ne le pouvaient pas. (Voir les éprouvettes ci-dessous.)

Croissance : les cellules de type sauvage peuvent croître et se diviser.

Aucune croissance : les cellules des mutants ne peuvent pas croître et se diviser.

Témoins : milieu minimal (MM)

■ **RÉSULTATS** ■ Comme le montre le tableau ci-dessous, la souche de type sauvage peut croître dans n'importe quelles conditions expérimentales et n'exige qu'un milieu de croissance minimal. Les trois types de mutants avaient chacun leur propre ensemble d'exigences de croissance. Par exemple, les mutants de la catégorie II étaient incapables de croître en présence d'ornithine seulement, mais ils arrivaient à se développer quand on ajoutait de la citrulline ou de l'arginine.

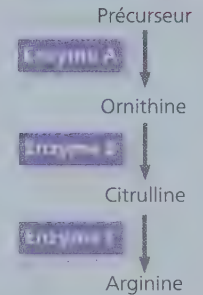
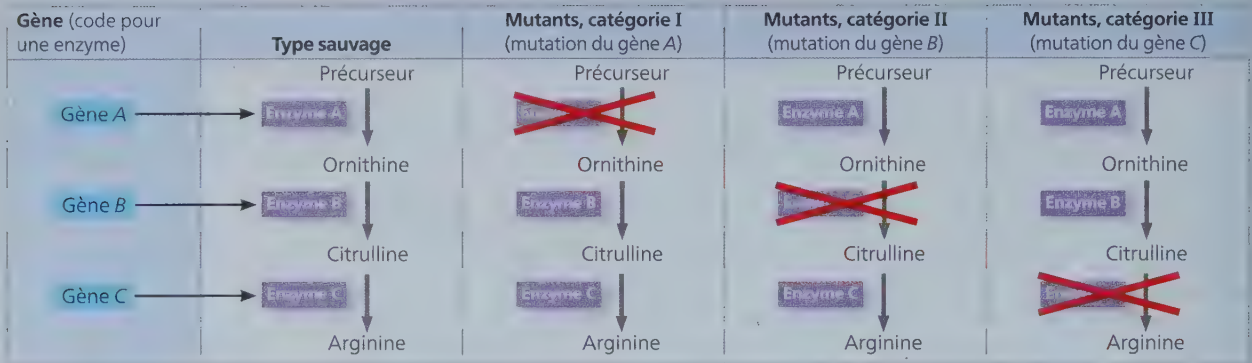


Tableau des résultats		Types de <i>Neurospora crassa</i>			
		Type sauvage	Mutants, catégorie I	Mutants, catégorie II	Mutants, catégorie III
Condition	Milieu minimal (MM) (témoin)				
	MM + ornithine				
	MM + citrulline				
	MM + arginine (témoin)				
	Résumé des résultats	Croissance avec ou sans suppléments	Croissance en présence d'ornithine, de citrulline ou d'arginine	Croissance seulement en présence de citrulline ou d'arginine	Croissance impossible sans arginine

Des gènes individuels codent-ils pour la production d'enzymes qui participent à une voie biochimique? (suite)



CONCLUSION À partir des résultats de l'expérience de croissance, Srb et Horowitz ont conclu que chaque catégorie de mutants était incapable d'accomplir l'une des étapes de la voie de synthèse de l'arginine, probablement parce que les mutants ne produisaient pas l'enzyme correspondante, comme le montre le tableau ci-dessus. Étant donné que chacun des mutants portait une mutation sur un seul gène, ils en ont conclu que chaque gène mutant devait normalement coder pour la production d'une enzyme. Leurs résultats constituaient un argument en faveur de l'hypothèse *un gène, une enzyme* proposée par Beadle et Tatum et confirmaient également que la voie métabolique de l'arginine dans le foie des mammifères fonctionne également chez *Neurospora*. (Notez dans le tableau des résultats qu'un mutant se développe seulement si on lui fournit un composé synthétisé après l'étape défectueuse, ce qui permet de contourner l'anomalie.)

Source des données: A. M. Srb et N. H. Horowitz, The ornithine cycle in *Neurospora* and its genetic control, *Journal of Biological Chemistry* 154:129-139 (1944).

ET SI ? Supposons que l'expérience démontre que les mutants de la catégorie I ne peuvent croître que dans un MM auquel on a ajouté de l'ornithine ou de l'arginine et que ceux de la catégorie II peuvent croître dans un MM enrichi de citrulline, d'ornithine ou d'arginine. Quelles conclusions les chercheurs auraient-ils dû tirer de ces résultats concernant la voie biochimique et l'anomalie chez les mutants de la catégorie I et de la catégorie II ?

déficience de chacun des mutants avec plus de précision, en effectuant d'autres tests afin d'identifier les trois catégories de mutants pour l'arginine. Les mutants de chaque catégorie avaient besoin d'un ensemble différent de composés dans la voie de synthèse de l'arginine, qui comporte trois étapes. Ces résultats, ainsi que ceux de nombreuses expériences semblables effectuées par Beadle et Tatum, semblaient indiquer que ces trois catégories de mutants présentaient un blocage à différentes étapes de la voie de synthèse de l'arginine et qu'il leur manquait l'enzyme catalysant l'étape correspondante.

Comme les conditions expérimentales déterminées par Beadle et Tatum faisaient en sorte que la déficience ne provenait que d'un seul gène dans chaque cas, les résultats obtenus constituaient dans leur ensemble un argument de taille en faveur de l'hypothèse de travail qu'ils avaient formulée. Selon cette hypothèse, appelée *un gène, une enzyme*, chaque gène a pour fonction de diriger la production d'une enzyme particulière. Des expériences biochimiques ultérieures ont permis d'identifier les enzymes dont les mutants étaient dépourvus, ce qui a permis de confirmer encore un peu plus cette hypothèse. Beadle et Tatum ont reçu conjointement un prix Nobel en 1958 pour « leur

découverte que les gènes agissent en régulant des événements chimiques définis » (selon la formulation du comité Nobel).

Aujourd'hui, de nombreux exemples nous démontrent que la mutation d'un gène peut conduire à la production d'enzymes défectueuses, et que celles-ci sont la cause de maladies particulières. Par exemple, l'albinisme de la figure 17.1 est dû à l'absence de tyrosinase, une enzyme importante qui intervient dans la voie métabolique produisant la mélanine (pigment foncé). L'absence de mélanine est à l'origine du pelage blanc de l'animal, mais aussi de la coloration rose du nez, des oreilles, des sabots et des yeux de l'âne, puisque ce pigment n'est pas présent pour masquer la coloration normalement rougeâtre des vaisseaux sanguins.

Les produits de l'expression génétique: une histoire à suivre

Au fur et à mesure que des connaissances plus précises sur les protéines se sont accumulées, il a fallu modifier l'hypothèse *un gène, une enzyme*. Tout d'abord, toutes les protéines ne sont pas des enzymes. Ainsi, la kératine, qui est la protéine structurale du poil des mammifères, et l'insuline, une hormone, sont deux protéines non enzymatiques. Comme certains gènes commandent

La synthèse de protéines qui ne sont pas des enzymes, les biologistes moléculaires ont supposé qu'un gène correspondait à une protéine. Cependant, de nombreuses protéines sont construites à partir de deux ou de plusieurs chaînes polypeptidiques différentes, chacune sous l'action de son propre gène. Par exemple, l'hémoglobine – une protéine des érythrocytes des vertébrés qui a pour fonction de transporter les molécules d'oxygène (O_2) – contient deux types de polypeptides et est produite à partir de deux gènes, un pour chaque type de polypeptide (voir la figure 5.18). Il a donc fallu reformuler l'hypothèse de Beadle et Tatum sous la forme *un gène, un polypeptide*. Toutefois, même sous cette forme, cette description n'est pas tout à fait exacte. Premièrement, dans de nombreux cas, plusieurs gènes eucaryotes codent chacun pour un ensemble de polypeptides étroitement apparentés par l'intermédiaire d'un processus appelé épissage alternatif, que vous étudierez plus loin dans le présent chapitre. Deuxièmement, un bon nombre de gènes codent pour des molécules d'ARN qui exercent des fonctions importantes dans les cellules, même s'ils ne sont jamais traduits en protéines. Pour l'instant, nous nous limiterons aux gènes qui codent pour des polypeptides. Notons ici que l'on utilise souvent le terme protéine en général, plutôt que polypeptide (plus précis), pour désigner le produit des gènes.

Les principes généraux de la transcription et de la traduction

Les gènes contiennent les instructions qui permettent de fabriquer des protéines spécifiques, mais ils ne les construisent pas directement. C'est l'acide ribonucléique, ou ARN, qui établit le lien entre l'ADN et la synthèse des protéines. L'ARN est semblable chimiquement à l'ADN. Cependant, dans l'ARN, un ribose (un glucide) remplace le désoxyribose, et l'uracile (une base azotée) remplace la thymine (voir la figure 5.23). Autrement dit, le long d'un brin d'ADN, chaque nucléotide est composé d'une base azotée qui peut être A, G, C ou T. Par contre, le long d'un brin d'ARN, chaque nucléotide est constitué d'une base azotée qui peut être A, G, C ou U. Par ailleurs, la molécule d'ARN est généralement formée d'un seul brin, beaucoup plus court que les longues molécules d'ADN.

La description du passage de l'information du gène à la protéine se rapporte souvent à la linguistique : tout comme certaines séquences précises de lettres permettent de transmettre une information dans une langue donnée, les acides nucléiques et les protéines contiennent des séquences spécifiques de monomères qui véhiculent une information. Dans l'ADN et l'ARN, les quatre types de nucléotides constituent les monomères en question : ils diffèrent par leur base azotée. Les gènes se composent généralement de centaines ou de milliers de nucléotides, et chaque gène comporte une séquence de nucléotides qui lui est spécifique. Dans les protéines aussi, chaque polypeptide présente des monomères alignés dans un ordre précis (voir la figure 5.18 pour la structure primaire des protéines) ; chacun de ceux-ci est un acide aminé. Les acides nucléiques et les protéines contiennent donc une information écrite dans deux langages chimiques différents. Le passage de l'un à l'autre se fait en deux étapes principales, appelées transcription et traduction.

La **transcription** est la synthèse d'ARN à partir de l'information contenue dans l'ADN. Les deux acides nucléiques présentent des formes différentes du même langage, et l'information est simplement transcrite, ou transposée, de l'ADN à l'ARN.

La séquence de nucléotides d'ADN constitue une matrice servant à l'assemblage d'une séquence de nucléotides d'ARN, de la même façon qu'elle constitue une matrice pour la synthèse d'un brin complémentaire pendant la réplication de l'ADN (voir le concept 16.2). Pour une protéine codée par un gène, la molécule d'ARN qui en résulte est donc une transcription fidèle des instructions fournies par un gène en vue de la construction d'une protéine. Ce type de molécule d'ARN est appelé **ARN messager (ARNm)**, parce qu'il joue le rôle de messenger génétique entre l'ADN et le dispositif de synthèse protéique de la cellule. D'une manière générale, on nomme transcription la synthèse de *tout type* d'ARN à partir d'une matrice d'ADN. Plus loin dans ce chapitre, vous verrez que l'ARNm n'est pas le seul type d'ARN produit par transcription.

La **traduction** est la synthèse d'un polypeptide à partir des informations contenues dans l'ARNm. À cette étape, il y a passage d'une « langue » à une autre : la cellule doit traduire la séquence de nucléotides d'une molécule d'ARNm en une séquence d'acides aminés appartenant à un polypeptide. La traduction se déroule dans les **ribosomes**, des complexes moléculaires qui participent à la formation de chaînes polypeptidiques en permettant l'assemblage des acides aminés dans l'ordre dicté par l'ARNm.

La transcription et la traduction se déroulent dans tous les organismes. Étant donné que la plupart des études ont été effectuées sur des bactéries et des cellules eucaryotes, nous nous attarderons à ces deux catégories d'organismes dans le présent chapitre. Notre compréhension de la transcription et de la traduction chez les archées est moins avancée, mais nous savons qu'au niveau de l'expression génétique, certaines archées ont des caractéristiques communes avec les bactéries, tandis que d'autres présentent des points communs avec les cellules eucaryotes.

Le schéma général de la transcription et de la traduction est semblable chez les bactéries et les organismes eucaryotes, mais il existe une différence importante sur le plan de la transmission de l'information au sein des cellules. Les bactéries n'ont pas de noyau. Par conséquent, l'ADN et l'ARNm bactériens ne sont pas séparés des ribosomes et des autres outils essentiels à la synthèse protéique par une membrane nucléaire (**figure 17.4a**). Comme vous le verrez plus loin, cette absence de cloisonnement permet à la traduction d'un ARNm de commencer pendant que sa transcription est en cours. Par contre, les cellules eucaryotes ont un noyau. La présence de l'enveloppe nucléaire empêche la transcription et la traduction de se dérouler au même endroit et au même moment (**figure 17.4b**). La transcription a lieu dans le noyau, mais l'ARNm doit être transporté dans le cytoplasme pour que la traduction ait lieu. Cependant, avant de pouvoir quitter le noyau, les transcrits d'ARN eucaryotes produits par les gènes codant pour une protéine subissent diverses transformations. Ce n'est qu'après ces modifications qu'ils deviennent des ARNm définitifs et fonctionnels. La transcription du gène eucaryote codant pour une protéine produit de l'*ARN pré-messager* ; la maturation de l'ARN génère ensuite la version définitive de l'ARNm. De façon plus générale, on appelle **transcrit primaire** la première version d'ARN qui résulte de la transcription d'un gène, y compris ceux qui codent pour un ARN qui n'est pas traduit en protéine.

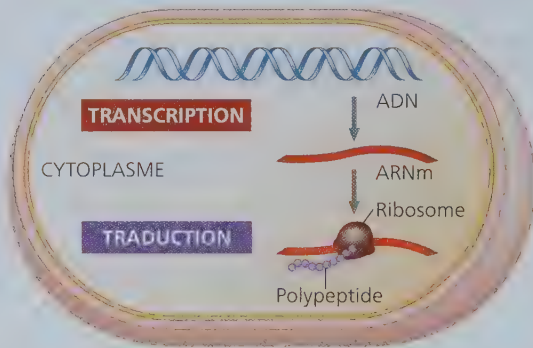
Résumons donc : les gènes programment la synthèse des protéines par l'intermédiaire de messages génétiques qui se présentent sous la forme d'ARN messagers. Autrement dit, les

cellules sont régies par une chaîne de commandement de nature moléculaire avec un flux d'information génétique directionnel, illustré ici par des flèches :

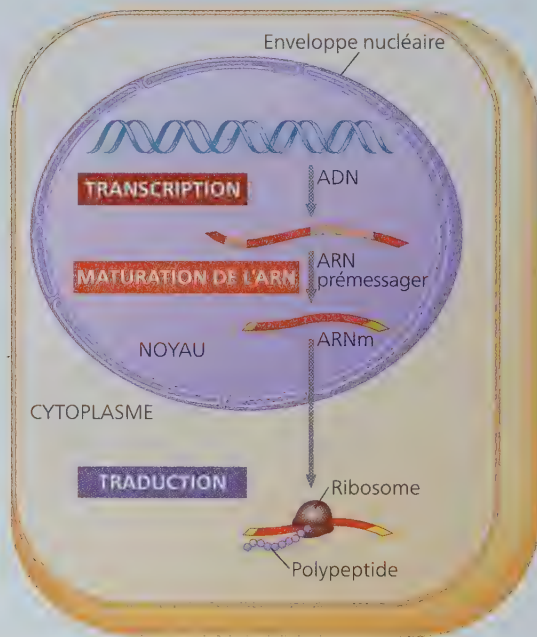


▼ **Figure 17.4** Vue d'ensemble: le rôle de la transcription et de la traduction dans la transmission de l'information génétique.

Dans la cellule, l'information génétique passe de l'ADN à l'ARN, puis de l'ARN à la protéine. Les deux étapes principales de la transmission de l'information sont la transcription et la traduction. Plusieurs figures apparaissant plus loin dans ce chapitre sont accompagnées d'une image réduite de la partie (a) ou de la partie (b); vous pourrez ainsi situer ces figures dans le processus global de l'expression génétique.



(a) **Cellule bactérienne.** Dans une cellule bactérienne, qui n'a pas de noyau, l'ARNm produit par la transcription est immédiatement traduit, sans aucune maturation.



(b) **Cellule eucaryote.** Le noyau constitue un compartiment distinct dans lequel se déroule la transcription. Le premier transcrite d'ARN, appelé ARN pré-messager, subit une maturation en plusieurs étapes, puis il quitte le noyau sous forme d'ARNm.

Cette chaîne de commandement a été baptisée *dogme central* par Francis Crick en 1956. Toutefois, dans les années 1970, les scientifiques ont découvert avec surprise que certaines enzymes utilisaient les molécules d'ARN comme matrices pour la synthèse de l'ADN (un processus que nous décrirons au concept 19.2). Cependant, ces exceptions n'invalident nullement l'idée selon laquelle, en règle générale, l'information génétique passe de l'ADN à l'ARN, puis de l'ARN à la protéine. Dans la prochaine section, nous verrons comment les acides nucléiques encodent l'ordre d'assemblage des acides aminés.

Le code génétique

Lorsque les biologistes ont commencé à se douter que l'ADN contenait les instructions pour la synthèse des protéines, ils se sont posé la question suivante : comment quatre nucléotides seulement peuvent-ils détenir le message génétique correspondant à 20 acides aminés différents ? Le code génétique ne peut constituer un langage analogue au chinois, langue dans laquelle chaque symbole d'écriture représente un mot unique. Combien de nucléotides peuvent alors correspondre à un acide aminé ?

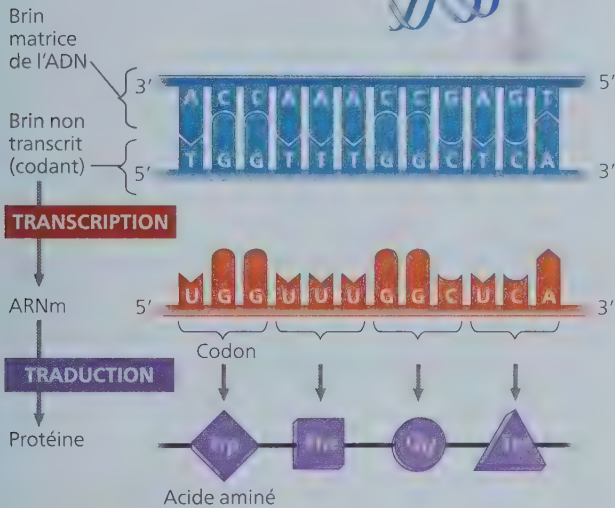
Les codons : des triplets de nucléotides

Si chaque base nucléotidique était traduite en un acide aminé, il ne pourrait y avoir que quatre acides aminés codés, un pour chacune des bases. Alors, un langage avec des mots de deux lettres suffirait-il ? Par exemple, la séquence de deux nucléotides AG désignerait un acide aminé, et GT, un autre acide aminé, etc. Était donné qu'il y a quatre bases nucléotidiques possibles dans chaque position, cela donnerait 16 (4×4 , ou 4^2) combinaisons possibles. Or, ce nombre ne suffit toujours pas à détenir le message génétique correspondant aux 20 acides aminés existants.

Les plus courtes séquences de longueur égale permettant de coder pour tous les acides aminés comprennent en fait trois bases azotées. En effet, si chaque combinaison de trois bases nucléotidiques consécutives représente un acide aminé, il est possible d'écrire 64 mots de code (4^3) ; c'est plus qu'il n'en faut pour représenter les 20 acides aminés. Des expériences ont permis de confirmer que le flux d'information allant du gène à la protéine repose sur un **code à triplets**. Les instructions génétiques pour la synthèse d'une chaîne polypeptidique se présentent sous la forme d'une série de mots composés chacun de trois nucléotides d'ADN qui ne se chevauchent pas. La série de mots dans un gène est transcrite en une série complémentaire de mots composés chacun de trois nucléotides dans l'ARNm, et celle-ci est ensuite traduite en une chaîne d'acides aminés (**figure 17.5**).

Au cours de la transcription, le gène détermine la séquence des bases nucléotidiques de la molécule d'ARN qui est synthétisée. Un seul des deux brins d'ADN de chaque gène est transcrit. Nous l'appellerons **brin matrice** (le brin non transcrit, ou brin codant, est le brin complémentaire). Ce brin matrice sert de modèle pour l'agencement des séquences de nucléotides du transcrite d'ARN. Cependant, plus loin le long de la même molécule d'ADN chromosomique, le brin complémentaire peut servir de matrice pour un autre gène. Le brin utilisé comme matrice est déterminé par l'orientation de l'enzyme assurant la transcription des gènes, laquelle dépend de la séquence d'ADN particulière associée à ce gène.

▼ **Figure 17.5 Le code à triplets.** Pour chaque gène, un seul brin d'ADN sert de matrice pour la transcription des ARN, tels que l'ARNm. Les règles de l'appariement des bases qui régissent la synthèse de l'ADN s'appliquent également à la transcription, mis à part le fait que l'ARN est constitué d'uracile (U) plutôt que de thymine (T). Pendant la traduction, l'ARNm est lu comme une séquence de triplets de nucléotides nommés codons. Chaque codon représente un acide aminé qui doit être ajouté au bout de la chaîne polypeptidique en cours de synthèse. L'ARNm est lu dans le sens 5' → 3'.



HABILETÉS VISUELLES ► Par convention, on utilise le brin d'ADN non transcrit, aussi appelé brin codant, pour représenter une séquence d'ADN. Écrivez la séquence du brin d'ARNm et du brin d'ADN non transcrit – dans les deux cas, dans la direction 5' → 3' – et comparez-les. Selon vous, pourquoi a-t-on adopté cette convention? (*Indice*: Pourquoi le brin d'ADN non transcrit est-il appelé brin codant?)

La molécule d'ARNm et sa matrice d'ADN ne sont pas identiques, mais complémentaires: les nucléotides de l'ARN s'assemblent sur la matrice suivant les règles de l'appariement des bases (voir la figure 17.5). Les paires de bases sont identiques à celles qui se forment pendant la réplication de l'ADN, à deux différences près: dans l'ARN, c'est U et non T qui s'apparie avec A, et les nucléotides contiennent le ribose au lieu du désoxyribose. Tout comme un nouveau brin d'ADN, la molécule d'ARN est synthétisée dans le sens antiparallèle du brin matrice de l'ADN. (Pour revoir ce que signifie « antiparallèle » et ce que sont les extrémités 5' et 3' d'une chaîne d'acides nucléiques, consultez la figure 16.7.) Dans l'exemple de la figure 17.5, le triplet ACC du brin matrice d'ADN (écrit sous la forme 3'-ACC-5') constitue la matrice pour 5'-UGG-3' dans la molécule d'ARNm. Les triplets de l'ARNm sont appelés **codons**, et sont habituellement écrits dans le sens 5' → 3'. Dans notre exemple, UGG est le codon de l'acide aminé appelé tryptophane (dont l'abréviation est Trp, ou W). Notons ici que le terme *codon* désigne parfois aussi

les triplets de l'ADN sur le brin *non transcrit*. Ces codons sont complémentaires au brin matrice et par conséquent de séquence identique à l'ARNm, à la différence près qu'ils comportent des T et non des U. Par conséquent, on désigne souvent sous le nom de **brin codant** le brin d'ADN non transcrit de l'ADN; par convention, on utilise la séquence du brin codant pour présenter la séquence d'un gène.

Au cours de la traduction, la séquence de codons alignés sur la molécule d'ARNm est décodée, ou traduite, en une séquence d'acides aminés constituant une chaîne polypeptidique. Le dispositif de traduction lit les codons dans le sens 5' → 3' le long de l'ARNm. Chaque codon présent sur la molécule d'ARNm détermine lequel des 20 acides aminés sera inséré à la position correspondante dans le polypeptide. Comme les codons sont des triplets de bases azotées, le nombre de nucléotides constituant le message génétique doit être trois fois plus élevé que le nombre d'acides aminés dans la protéine finale. Par exemple, il faut une séquence codante de 300 nucléotides sur un brin d'ARNm pour coder les acides aminés dans un polypeptide long de 100 acides aminés.

Le décodage du code

Les biologistes moléculaires ont décrypté le code de la vie au début des années 1960. À cette époque, une série d'expériences remarquables ont permis de connaître la traduction de chaque codon d'ARNm en un acide aminé. Marshall Nirenberg, des National Institutes of Health, aux États-Unis, a déchiffré le premier codon en 1961 avec ses collègues. Nirenberg a synthétisé un ARNm artificiel en reliant plusieurs nucléotides d'ARN identiques, dont la base était toujours l'uracile. Ainsi, peu importait où le message génétique commençait ou finissait: il ne contenait qu'un seul codon, UUU, répété plusieurs fois. Dans une éprouvette, Nirenberg a ajouté le polynucléotide « poly-U » à un mélange contenant des acides aminés, des ribosomes et les autres molécules nécessaires à la synthèse des protéines chez *Escherichia coli*. Son système artificiel a traduit l'ARNm poly-U en un polypeptide formé d'une longue chaîne de phénylalanine (Phe, ou F); Nirenberg a ainsi déterminé que le codon d'ARNm UUU représentait donc le code de la phénylalanine. Peu de temps après, on a trouvé les acides aminés correspondant aux codons AAA, GGG et CCC.

Il a fallu employer des techniques plus élaborées pour décoder des triplets mixtes, tels que AUA et CGA. Toutefois, vers le milieu des années 1960, les 64 codons étaient déchiffrés. Comme vous pouvez le voir à la **figure 17.6**, 61 des 64 triplets codent pour les acides aminés. Les trois codons qui ne désignent pas des acides aminés (UAA, UAG et UGA) codent pour des signaux d'« arrêt » marquant la fin de la traduction. Remarquez que le triplet AUG a une double fonction: il détient le message génétique correspondant à un acide aminé, la méthionine (Met, ou M), et il sert aussi de signal de « départ ». Les messages génétiques commencent généralement par le codon AUG; celui-ci indique où le dispositif de synthèse protéique doit entreprendre la traduction de l'ARNm. Étant donné que AUG code également pour la méthionine, toutes les chaînes peptidiques nouvellement synthétisées débutent par cet acide aminé (plus tard, une enzyme peut détacher la méthionine du polypeptide).

En consultant la figure 17.6, vous pouvez remarquer que le code génétique est redondant; il n'est cependant jamais ambigu.

Le tryptophane et la méthionine sont les deux seuls acides aminés désignés par un unique codon. Tous les autres acides aminés sont désignés par au moins deux codons et certains par six (redondance). Par exemple, GAA et GAG donnent tous deux l'acide glutamique. Toutefois, aucun codon ne code pour plus d'un acide aminé (il n'y a pas d'ambiguïté). La redondance du code n'est pas seulement l'effet du hasard. Dans de nombreux cas, les codons «synonymes» ne diffèrent que par la troisième base nucléotidique du triplet. Plus loin dans ce chapitre, nous verrons un des avantages de cette redondance.

Un message écrit ne peut être compris que si les symboles sont lus dans le bon ordre et selon les bons groupements; c'est ce que l'on appelle le **cadre de lecture**. Prenons, par exemple, la phrase suivante: «Ils ont élu roi mon ami qui fut ému.» Si l'on forme des groupements erronés en commençant au mauvais endroit, le message devient incompréhensible: «Iso nté lur oim ona miq uif uté mu.» Le cadre de lecture revêt une importance cruciale dans le langage moléculaire de la cellule. La synthèse du court segment polypeptidique représenté à la figure 17.5 ne se déroule correctement que si la lecture des nucléotides d'ARNm

▼ **Figure 17.6** Le tableau des codons pour l'ARNm. Dans ce tableau, on désigne les trois bases d'un codon d'ARNm par *première*, *deuxième* et *troisième* base. Elles sont lues dans le sens 5' → 3' de l'ARNm. Le codon AUG code pour l'acide aminé méthionine (Met, ou M); il constitue aussi un signal de «départ» montrant l'endroit où les ribosomes doivent commencer à traduire l'ARNm. Trois des 64 codons sont des signaux d'«arrêt» indiquant où les ribosomes terminent la traduction. Les codes à une lettre et à trois lettres sont présentés pour les acides aminés. (Pour une liste des noms complets, voir la figure 5.14.)

		Deuxième base de l'ARNm							
		U		C		A		G	
U	UUU	Phe (F)	UCU	Ser (S)	UAU	Tyr (Y)	UGU	Cys (C)	
	UUC		UCC		UAC		UGC		
	UUA	Leu (L)	UCA	UAA	Arrêt	UGA	Arrêt		
	UUG		UCG	UAG	Arrêt	UGG	Trp (W)		
C	CUU	Leu (L)	CCU	Pro (P)	CAU	His (H)	CGU	Arg (R)	
	CUC		CCC		CAC		CGC		
	CUA	Leu (L)	CCA	CAA	Gln (Q)	CGA	Arg (R)		
	CUG		CCG	CAG	CGG				
A	AUU	Ile (I)	ACU	Thr (T)	AAU	Asn (N)	AGU	Ser (S)	
	AUC		ACC		AAC		AGC		
	AUA	Met (M) ou départ	ACA	AAA	Lys (K)	AGA	Arg (R)		
	AUG		ACG	AAG	AGG				
G	GUU	Val (V)	GCU	Ala (A)	GAU	Asp (D)	GGU	Gly (G)	
	GUC		GCC		GAC		GGC		
	GUA	Val (V)	GCA	GAA	Glu (E)	GGA	Gly (G)		
	GUG		GCG	GAG	GGG				

HABILETÉS VISUELLES ► Un segment situé au milieu d'une molécule d'ARNm contient la séquence 5'-AGAGAACCGCGA-3'. À l'aide du tableau ci-dessus, traduisez cette séquence en supposant que les trois premiers nucléotides sont un codon.

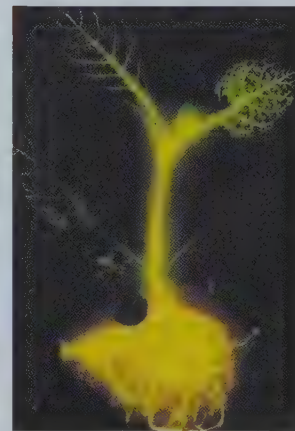
se fait de gauche à droite (5' → 3') et selon les groupements suivants: UGG UUU GGC UCA. Bien qu'aucun espace ne sépare les différents codons dans le message génétique, celui-ci est lu comme une série de mots de trois lettres par les enzymes de la synthèse protéique de la cellule. En d'autres termes, les codons ne se chevauchent pas, ils sont lus séquentiellement.

L'évolution du code génétique

ÉVOLUTION Le code génétique est presque universel; il est le même chez des organismes aussi différents que la plus simple des bactéries jusqu'aux animaux et aux végétaux les plus complexes. Par exemple, la traduction du codon CCG de l'ARNm donne l'acide aminé proline chez tous les organismes dont on a examiné le code génétique. Au cours d'expériences de laboratoire, on a réussi à traduire des ARNm et à transcrire et traduire des gènes d'une espèce après les avoir transplantés chez une autre, parfois avec des résultats très étonnants, comme le montre la **figure 17.7**. Par exemple, il est possible de programmer des bactéries en y insérant un gène humain pour leur faire produire certaines protéines utiles sur le plan médical, comme l'insuline humaine. Dans le domaine de la biotechnologie, des applications de cette nature ont abouti à de nombreux développements très intéressants, dont nous parlerons au concept 20.4.

Malgré un petit nombre d'exceptions, la signification évolutive de la quasi-universalité du code génétique est claire. Ce langage a dû apparaître assez tôt dans l'histoire de la vie pour se retrouver chez les ancêtres communs à tous les organismes actuels. L'existence d'un vocabulaire génétique commun nous rappelle les liens de parenté qui unissent toutes les formes de vie.

▼ **Figure 17.7** Un témoignage de l'évolution: l'expression des gènes de différentes espèces. Comme les diverses formes de vie possèdent un code génétique commun en raison d'une ascendance partagée, il est possible de programmer une espèce afin qu'elle produise des protéines propres à une autre espèce; pour ce faire, on y introduit de l'ADN provenant de cette dernière.



(a) **Plant de tabac exprimant un gène de luciole.** La luminescence jaune est produite au cours d'une réaction chimique catalysée par la protéine synthétisée à partir d'un gène de luciole.



(b) **Porc exprimant un gène de méduse.** Le gène d'une protéine fluorescente a été injecté dans des ovules de porc après leur fécondation. Un des embryons s'est développé pour donner ce porc fluorescent.

- FAITES DES LIENS** ► Dans un rapport de recherche portant sur l'alcaptonurie publié en 1902, Garrod affirmait que les humains recevaient deux « caractères » (allèles) pour une enzyme particulière et que les deux parents devaient contribuer à une version erronée pour que la maladie apparaisse chez le descendant. De nos jours, ce désordre serait-il appelé dominant ou récessif? Voir le concept 14.4.
- Combien d'acides aminés le polypeptide produit par un ARNm poly-G long de 30 nucléotides contiendra-t-il, et de quelle nature seront ces acides aminés?
- FAITES UN DESSIN** ► Soit un brin matrice d'un gène contenant la séquence 3'-TTCAGTCGT-5'. Imaginez que le brin codant (habituellement non transcrit) a été transcrit à la place de la séquence de la matrice. Dessinez la séquence de l'ARNm et traduisez-la à l'aide de la figure 17.6. (Faites attention aux extrémités 5' et 3'.) Dites si la protéine synthétisée à partir du brin non transcrit sera fonctionnelle, pour autant qu'elle soit synthétisée.

Voir les réponses proposées à l'appendice A.

La transcription est la synthèse de l'ARN à partir de l'ADN: une étude détaillée

Après ces considérations sur l'aspect linguistique et la signification du code génétique du point de vue de l'évolution, abordons plus en détail la transcription, la première étape de l'expression génétique.

Les composants moléculaires de la transcription

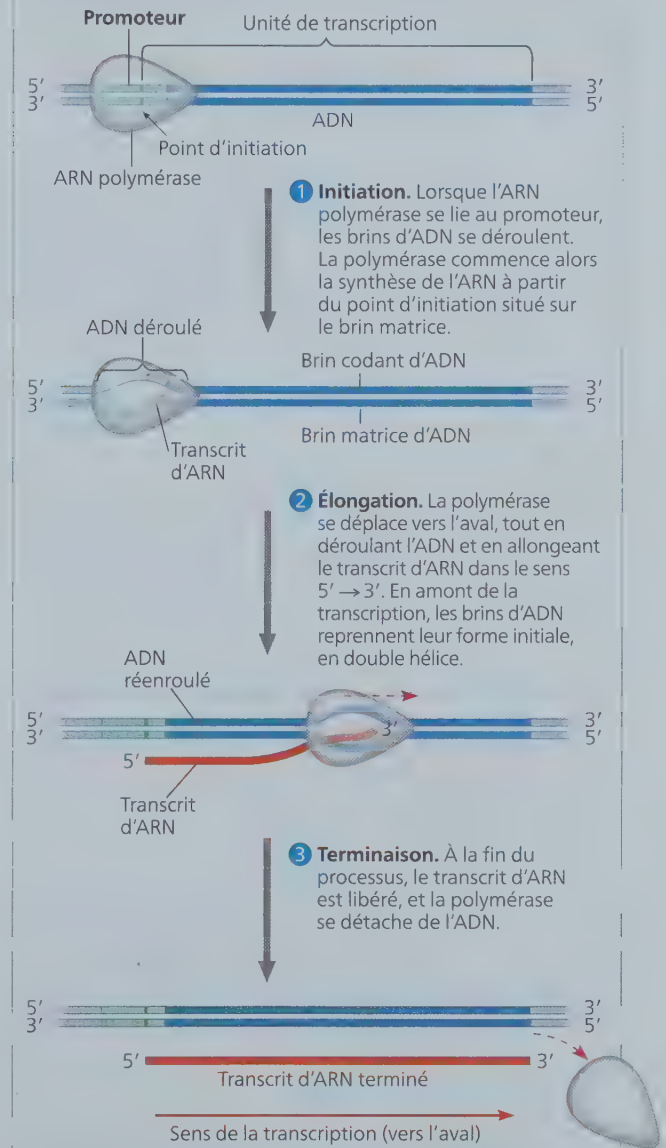
L'ARN messager est transcrit à partir du brin matrice d'un gène. Il transmet l'information de l'ADN aux structures cellulaires assurant la synthèse des protéines. Une enzyme appelée **ARN polymérase** écarte les deux brins d'ADN et assemble les nucléotides complémentaires de l'ARN au brin matrice de l'ADN (**figure 17.8**). À l'instar des ADN polymérases, qui assurent la réplication de l'ADN, les ARN polymérases ne peuvent assembler un polynucléotide que dans le sens 5' → 3', en ajoutant des nucléotides à l'extrémité 3'. Par contre, contrairement aux ADN polymérases, les ARN polymérases peuvent commencer la synthèse d'une chaîne directement sur la matrice; elles n'ont pas à ajouter le premier nucléotide à une amorce préexistante.

Des séquences particulières de nucléotides dans l'ADN marquent le début et la fin de la transcription du gène. La séquence d'ADN à laquelle l'ARN polymérase se lie pour commencer la transcription est appelée **promoteur**. Celui-ci agit comme un guide qui oriente le travail de l'ARN polymérase. Chez les bactéries, la séquence qui marque la fin de la transcription est appelée **terminateur**. (Le mécanisme de la terminaison chez les eucaryotes est différent; nous le décrirons plus loin.) En biologie moléculaire, on comme « aval » le sens dans lequel s'effectue la transcription et « amont » le sens opposé. Ces termes désignent également les positions relatives des séquences de

nucléotides de l'ADN et de l'ARN. Par conséquent, la séquence du promoteur dans l'ADN se situe en amont du terminateur. L'**unité de transcription** est le segment d'ADN transcrit en molécule d'ARN.

Chez les bactéries, il n'existe qu'un seul type d'ARN polymérase, et cette enzyme synthétise l'ARNm et d'autres sortes d'ARN intervenant dans la synthèse des protéines, comme l'ARN

▼ **Figure 17.8** Les étapes de la transcription: initiation, élongation et terminaison. Cette description générale de la transcription s'applique aussi bien aux bactéries qu'aux organismes eucaryotes; cependant, les détails de la terminaison diffèrent, comme nous le décrivons dans le texte. De plus, chez une bactérie, l'ARN ainsi transcrit est aussitôt utilisé comme ARNm; chez un eucaryote, par contre, il doit d'abord passer par l'étape de la maturation.



FAITES DES LIENS ► Comparez l'utilisation d'un brin matrice pendant la transcription et lors de la réplication. Voir la figure 16.17.

ribosomal. Par contre, les noyaux des eucaryotes renferment au moins trois types d'ARN polymérase; c'est l'ARN polymérase II qui effectue la synthèse de l'ARNm. Les autres ARN polymérases transcrivent les molécules d'ARN qui ne sont pas traduites en protéines. Dans cette section, consacrée à la transcription, nous commencerons par les aspects de la synthèse de l'ARNm qui sont communs aux bactéries et aux organismes eucaryotes. Ensuite, nous verrons quelques-unes des différences principales entre ces deux groupes.

La synthèse d'un transcrit d'ARN

Les trois étapes de la transcription illustrées à la figure 17.8 et dont il est question plus loin sont l'initiation, l'élongation et la terminaison de la chaîne d'ARN. À l'aide de la figure 17.8, apprenez à les reconnaître et familiarisez-vous avec les termes qui s'y rapportent.

La liaison de l'ARN polymérase et l'initiation de la transcription

Le promoteur d'un gène inclut le **point de départ** de la transcription, ou point d'initiation (le nucléotide à partir duquel l'ARN polymérase commence la synthèse de l'ARNm). Il couvre habituellement plusieurs douzaines de paires de nucléotides, voire plus, en amont de ce point (**figure 17.9**). Selon les interactions avec les protéines, que nous aborderons sous peu, l'ARN polymérase se lie au promoteur à un endroit et dans un sens précis. Cette liaison marque le début de la transcription et détermine lequel des deux brins de l'hélice d'ADN sera transcrit.

Certaines parties du promoteur jouent un rôle particulièrement important dans la liaison de l'ARN polymérase puisqu'elles veillent à ce que la transcription commence au bon endroit. Chez les bactéries, c'est une partie de l'ARN polymérase elle-même qui reconnaît le promoteur et qui s'y lie. Chez les eucaryotes, un ensemble de protéines appelées **facteurs de transcription** servent d'intermédiaires: ce sont elles qui permettent la liaison de l'ARN polymérase et le début de la transcription. Ce n'est que lorsque les facteurs de transcription ont été fixés au promoteur que l'ARN polymérase II se lie à celui-ci. L'ensemble constitué par l'ARN polymérase II et les facteurs de transcription liés au promoteur est nommé **complexe d'initiation de la transcription**. La figure 17.9 montre la fonction des facteurs de transcription et d'une séquence essentielle de l'ADN du promoteur, appelée **boîte TATA** – parce qu'elle présente une forte concentration de thymine (T) et d'adénine (A) –, lors de la formation du complexe d'initiation sur un promoteur chez les eucaryotes.

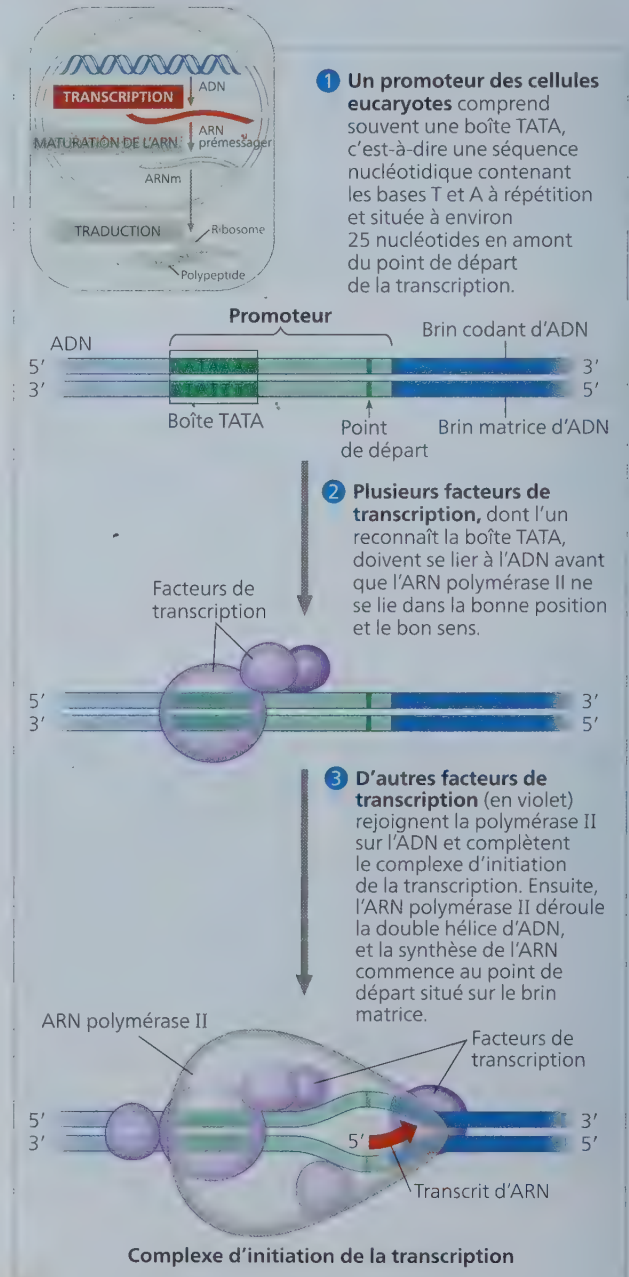
L'interaction entre l'ARN polymérase II d'un eucaryote et les facteurs de transcription illustre bien l'importance particulière que revêtent les interactions protéine-protéine dans le contrôle de la transcription chez les eucaryotes. Une fois que les facteurs de transcription appropriés sont fermement liés au promoteur et que la polymérase est bien orientée sur l'ADN, l'enzyme déroule les deux brins de la molécule et commence à transcrire le brin matrice au point de départ.

L'élongation du brin d'ARN

Pendant qu'elle se déplace le long de l'ADN, l'ARN polymérase déroule la double hélice; elle expose de 10 à 20 nucléotides environ à la fois. Elle permet ainsi leur appariement avec les nouveaux nucléotides d'ARN présents dans le milieu sous la forme

de ribonucléosides triphosphates (**figure 17.10**). Précisons qu'elle ajoute ceux-ci à l'extrémité 3' de la molécule d'ARN en cours de synthèse, tout en avançant le long de la double hélice. Notons aussi que l'ARN polymérase n'effectue pas de « correction d'épreuves » comme le fait l'ADN polymérase si des erreurs

▼ **Figure 17.9** L'initiation de la transcription sur un promoteur d'eucaryote. Dans les cellules eucaryotes, des protéines appelées facteurs de transcription jouent un rôle d'intermédiaire lors de l'initiation de la transcription par l'ARN polymérase II.



❓ Expliquez en quoi l'interaction de l'ARN polymérase avec le promoteur serait différente si l'initiation de la transcription chez les bactéries était représentée dans la figure.

surviennent au cours de l'appariement des paires de bases. Dans le prolongement de la synthèse du polypeptide qui progresse, la nouvelle molécule d'ARN se détache progressivement du brin matrice d'ADN et la double hélice d'ADN se reconstitue. Chez les eucaryotes, la vitesse de progression de la transcription est d'environ 40 nucléotides par seconde. Chez les procaryotes, l'ARN polymérase est légèrement plus rapide, progressant au rythme de 50 à 100 nucléotides par seconde.

Un même gène peut être transcrit simultanément par plusieurs molécules d'ARN polymérase, qui se suivent tel un convoi de camions. De chaque molécule émerge un nouveau filament d'ARN en formation : sa longueur reflète la distance parcourue par l'enzyme sur le brin matrice depuis le point de départ (voir les molécules d'ARNm à la figure 17.23). La transcription simultanée d'un même gène par un grand nombre de molécules de polymérase accroît le nombre d'ARNm fabriqués ; cela permet à la cellule de produire la protéine correspondante en grande quantité.

La terminaison de la transcription

Il existe certaines différences dans le mécanisme de la terminaison de la transcription entre les bactéries et les organismes eucaryotes. Chez les bactéries, la transcription se poursuit jusqu'au terminateur dans l'ADN. Le terminateur transcrit (une séquence d'ARN) joue le rôle de signal de terminaison : lorsqu'elle atteint cet endroit, l'ARN polymérase se détache de l'ADN et libère le transcrit, qui ne requiert généralement pas d'autres modifications avant la traduction. Par contre, dans la cellule eucaryote, l'ARN polymérase II transcrit une séquence sur l'ADN, appelée séquence de polyadénylation, qui code pour un signal de polyadénylation (AAUAAA) dans l'ARN prémessager. On dit qu'il s'agit d'un signal, car, dès sa formation, cette séquence

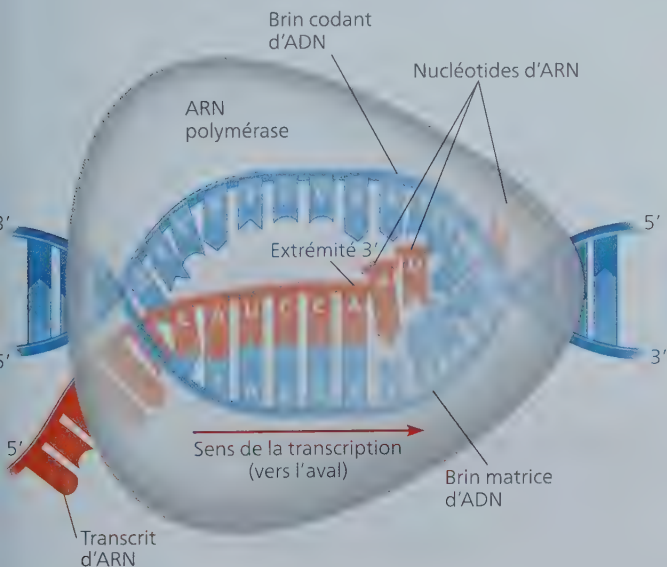
de six nucléotides d'ARN se lie immédiatement à certaines protéines situées dans le noyau. Ensuite, en un point situé de 10 à 35 nucléotides en aval de cette séquence, les protéines séparent le transcrit d'ARN en croissance de la polymérase, libérant l'ARN prémessager. L'ARN prémessager subit alors la maturation, un sujet que nous aborderons dans la prochaine section. Même si cette séparation marque la fin de l'ARN prémessager, la transcription de l'ARN polymérase II se poursuit sur l'ADN matrice. Les enzymes commencent à dégrader l'ARN de la polymérase à partir de l'extrémité 5' nouvellement exposée. L'ARN polymérase continue son processus de transcription, tout en étant poursuivie par les enzymes de dégradation. Quand ces enzymes rattrapent l'ARN polymérase, cette dernière est forcée de se dissocier de l'ADN matrice.

RETOUR SUR LE CONCEPT 17.2

1. Qu'est-ce qu'un promoteur ? Où est-il situé : en amont ou en aval de l'unité de transcription ?
2. Qu'est-ce qui permet à l'ARN polymérase de commencer à transcrire un gène au bon site sur l'ADN, dans une cellule bactérienne ? Dans une cellule eucaryote ?
3. **ET SI ?** ► Supposons que des rayons X modifient la séquence dans la boîte TATA du promoteur d'un gène particulier. Comment cela influera-t-il sur la transcription du gène ? (Voir la figure 17.9.)

Voir les réponses proposées à l'appendice A.

▼ **Figure 17.10** L'élongation de la transcription. L'ARN polymérase se déplace le long du brin matrice de l'ADN, liant les nucléotides complémentaires de l'ARN à l'extrémité 3' du transcrit d'ARN en formation. Dans le prolongement de la polymérase, la nouvelle molécule d'ARN se détache progressivement du brin matrice, qui reconstitue la double hélice avec le brin codant.



CONCEPT 17.3

Dans les cellules eucaryotes, l'ARN est modifié après avoir été transcrit

Dans le noyau de la cellule eucaryote, des enzymes apportent des modifications spécifiques à l'ARN prémessager avant que l'information génétique ne soit envoyée vers le cytoplasme. Pendant cette **maturation de l'ARN**, les deux extrémités du transcrit primaire subissent des modifications. Ensuite, dans la plupart des cas, certaines parties de l'intérieur de la molécule d'ARN sont excisées, et les parties restantes sont réunies. Ces transformations produisent une molécule d'ARNm prête pour la traduction.

La modification des extrémités de l'ARN prémessager

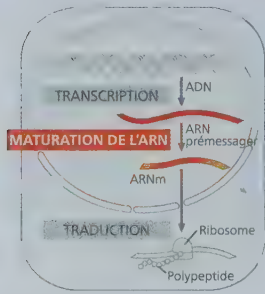
Chaque extrémité de la molécule d'ARN prémessager subit une transformation (**figure 17.11**). L'extrémité 5', qui est synthétisée en premier, reçoit une **coiffe 5'**, une forme modifiée d'un nucléotide de guanine (G) ajoutée à l'extrémité 5' après la transcription des 20 à 40 premiers nucléotides. Quant à l'extrémité 3' de la molécule d'ARN prémessager, elle est elle aussi modifiée avant que l'ARNm quitte le noyau. Rappelez-vous que l'ARN prémessager est excisé et libéré peu après la transcription du signal de polyadénylation, AAUAAA. À l'extrémité 3', une enzyme, la

▼ **Figure 17.11** La maturation de l'ARN: l'ajout de la coiffe 5' et de la queue poly-A.

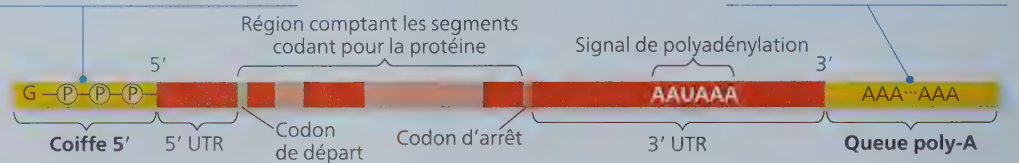
Dans la cellule eucaryote, les enzymes modifient les deux extrémités de la molécule d'ARN pré-messager. Les extrémités ainsi modifiées interviennent dans la sortie de l'ARNm du noyau

et contribuent à protéger l'ARN de la dégradation. Lorsque l'ARNm parvient dans le cytoplasme, les extrémités modifiées, conjointement avec certaines protéines cytosoliques, facilitent la liaison aux ribosomes. La coiffe 5' et la queue poly-A ne sont pas traduites en protéines,

ni les séquences appelées séquence 5' non traduite (5' UTR) et séquence 3' non traduite (3' UTR). Les segments roses sont les introns, que nous décrirons sous peu (voir la figure 17.12).



Ajout d'un nucléotide de guanine à l'extrémité 5'



poly-A polymérase, produit alors une **queue poly-A** formée de 50 à 250 nucléotides d'adénine (A). La coiffe 5' et la queue poly-A ont en commun plusieurs fonctions importantes: premièrement, elles semblent faciliter le transport de l'ARNm mature vers l'extérieur du noyau; deuxièmement, elles contribuent à protéger l'ARNm de la dégradation par les enzymes hydrolytiques; et troisièmement, elles facilitent la fixation des ribosomes à l'extrémité 5' de l'ARNm lorsque celui-ci parvient dans le cytoplasme. La figure 17.11 illustre également les séquences non traduites (UTR: *UnTranslated Region*, en anglais) aux extrémités 5' et 3' de l'ARNm (on les désigne par 5' UTR et 3' UTR). Ces séquences sont des parties de l'ARNm qui ne seront pas traduites en protéines; par contre, elles remplissent d'autres fonctions, comme la liaison aux ribosomes.

Les gènes discontinus et l'épissage de l'ARN

Dans le noyau de la cellule eucaryote, une étape étonnante de la maturation de l'ARN est l'**épissage de l'ARN** (figure 17.12). Durant ce processus, qui s'apparente au montage d'une bande vidéo, une grande partie des molécules d'ARN sont éliminées, alors que les parties restantes sont reliées. La longueur moyenne

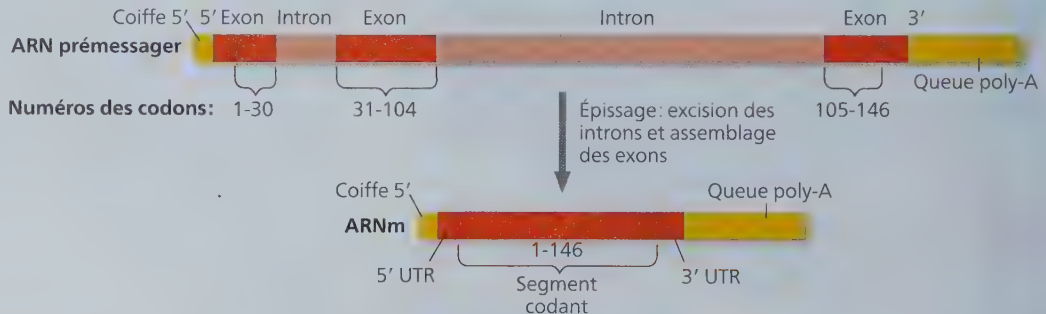
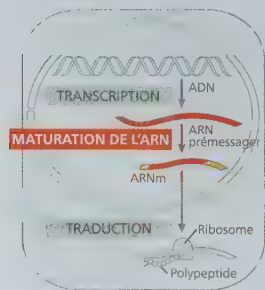
d'une unité de transcription d'une molécule d'ADN d'humain est d'environ 27 000 nucléotides, comme celle du transcrit primaire d'ARN. Cependant, environ 1 200 nucléotides dans l'ARN suffisent à coder pour une protéine de taille moyenne de 400 acides aminés (souvenez-vous que chaque acide aminé est codé par un *triplet* de nucléotides). La plupart des gènes d'eucaryotes et leurs transcrits d'ARN ont donc de longues séquences nucléotidiques non codantes qui échappent à la traduction. Ce qui est encore plus surprenant, c'est que la plupart de ces séquences sont dispersées entre les segments codants d'un gène, donc entre les segments codants de l'ARN pré-messager. Autrement dit, chez les eucaryotes, la séquence de nucléotides d'ADN qui détient le message génétique correspondant à un polypeptide n'est généralement pas continue; elle est séparée en segments. Les segments d'acide nucléique qui interrompent la séquence codante sont appelés **introns**. Chaque intron peut contenir plusieurs centaines, voire quelques milliers de nucléotides, et il y aurait une dizaine d'introns par gène. Quant aux régions codantes, elles sont appelées **exons**, parce qu'elles sont destinées à être exprimées: habituellement, elles sont traduites en des séquences d'acides aminés. (Les UTR des exons, situées aux extrémités de l'ARN, font exception: elles font partie de

▼ **Figure 17.12** La maturation de l'ARN: l'épissage.

La molécule d'ARN illustrée ici code pour le β -globine, un des polypeptides de l'hémoglobine. Les nombres qui figurent sous l'ARN correspondent à des codons. La β -globine est longue de 146 acides aminés. Le gène et son

transcrit d'ARN pré-messager possèdent trois exons, correspondant à des séquences qui vont quitter le noyau sous forme d'ARNm. (La 5' UTR et la 3' UTR font partie des exons parce qu'elles sont incluses dans l'ARNm. Cependant, elles ne codent pas pour des protéines.) Pendant

la maturation de l'ARN, les introns sont excisés et les exons sont réunis par épissage. Dans de nombreux gènes, les introns sont beaucoup plus volumineux que les exons.



FAITES UN DESSIN ► Sur l'ARNm, indiquez où se trouvent les codons de départ et d'arrêt.

l'ARNm, mais ne sont pas traduites en protéines. Par conséquent, il est utile de se souvenir que les exons sont les séquences de l'ARN qui parviennent à l'extérieur du noyau.) Les termes *intron* et *exon* s'appliquent à la fois aux séquences de l'ARN et à celles de l'ADN qui leur correspondent.

Lors de la synthèse du transcrite primaire à partir d'un gène, l'ARN polymérase II transcrit les introns et les exons de l'ADN. Toutefois, la molécule d'ARNm ne parvient dans le cytoplasme qu'après avoir été tronquée. Les introns sont retirés de la molécule, et les exons sont réunis par épissage, de sorte que la molécule d'ARNm ne comporte plus qu'une seule séquence codante continue. C'est ce processus d'excision et de réarrangement de l'ARN prémessager que l'on nomme **épissage**.

Comment l'épissage de l'ARN prémessager se déroule-t-il ? L'excision des introns est assurée par le **complexe d'épissage** (ou spliceosome), un important complexe constitué de protéines et de petites molécules d'ARN. Ce complexe se lie à plusieurs courtes séquences de nucléotides le long d'un intron, y compris les séquences clés situées à chaque extrémité (**figure 17.13**). L'intron est ensuite libéré (et rapidement dégradé), et le complexe d'épissage réunit les deux exons qui le bordaient. Il s'avère qu'en plus d'intervenir dans l'assemblage du complexe d'épissage et dans la reconnaissance des sites d'épissage, les petites molécules d'ARN de ce complexe catalysent également le processus d'épissage lui-même.

Les ribozymes

L'ARN peut jouer le rôle d'une enzyme. C'est par exemple le cas de l'ARN ribosomal, qui peut catalyser l'excision de ses propres introns ; ces ARN catalytiques sont nommés ribozymes.

POUR APPROFONDIR ■ L'idée selon laquelle le petit ARN nucléaire a une fonction catalytique dans le complexe d'épissage est née de la découverte des **ribozymes**, des molécules d'ARN agissant comme des enzymes. Chez certains organismes, il arrive

que l'épissage de l'ARN prémessager se déroule en l'absence de toute protéine et même de toute autre molécule d'ARN, parce que l'ARN de l'intron joue le rôle de ribozyme et catalyse lui-même sa propre excision ! Par exemple, chez le protiste cilié *Tetrahymena*, il y a autoépissage lors de la production d'ARN ribosomique (ARNr), un composant des ribosomes. En fait, l'ARN préribosomique libère ses propres introns. La découverte des ribozymes a donc rendu caduc le principe voulant que tous les catalyseurs biologiques soient des protéines.

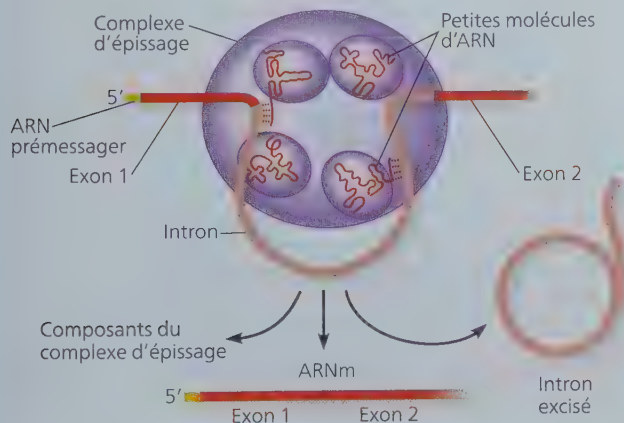
Grâce à trois propriétés de l'ARN, certaines molécules d'ARN ont la capacité de jouer le rôle d'enzymes. Premièrement, la structure monocaténaire (à simple brin) de l'ARN permet aux bases d'une région de la molécule d'ARN de s'apparier, selon un arrangement antiparallèle, avec celles d'une région complémentaire située ailleurs dans la même molécule ; cela confère une structure tridimensionnelle particulière à la molécule d'ARN. Cette structure tridimensionnelle est essentielle au rôle de catalyseur des ribozymes, tout comme pour les protéines enzymatiques. Deuxièmement, à l'instar de certains acides aminés dans une protéine enzymatique, certaines bases dans l'ARN comportent des groupements fonctionnels qui peuvent participer à la catalyse. Troisièmement, la capacité de l'ARN à former des liaisons hydrogène avec d'autres molécules d'acides nucléiques (soit l'ARN, soit l'ADN) ajoute de la spécificité à son activité catalytique. Par exemple, l'association par appariement de bases complémentaires entre l'ARN du complexe d'épissage et l'ARN d'un transcrite d'ARN primaire permet une localisation précise de la région où le ribozyme catalyse l'épissage. Plus loin dans le présent chapitre, vous apprendrez comment ces propriétés particulières confèrent également à l'ARN la capacité de jouer des rôles non catalytiques importants dans la cellule, comme la reconnaissance des codons à trois nucléotides sur l'ARNm. ■

L'importance des introns du point de vue de la fonction et de l'évolution

ÉVOLUTION L'épissage de l'ARN et la présence des introns constituent-ils ou non des avantages de sélection au cours de l'histoire évolutive ? Cette question est toujours matière à débat. Quoi qu'il en soit, il est instructif de considérer leurs possibles avantages adaptatifs. La plupart des introns ne semblent pas exercer de fonctions spécifiques, mais au moins certains d'entre eux contiennent des séquences qui assurent la régulation de l'expression génétique, et la plupart influent sur les produits des gènes.

La présence des introns dans les gènes a une conséquence importante : ils permettent à un même gène de coder pour plusieurs polypeptides différents. On sait que de nombreux gènes, que l'on nomme parfois *gènes en mosaïque*, mènent à la synthèse de deux chaînes polypeptidiques différentes ou plus, selon les segments traités comme des exons pendant la maturation de l'ARN. Ce phénomène est appelé **épissage différentiel** de l'ARN, ou encore épissage alternatif (voir la figure 18.13). Grâce à ce processus d'épissage différentiel, des cellules peuvent produire des protéines qui varient légèrement selon le type de tissu où le gène est présent, ou selon le stade de développement de l'organisme. Les résultats du Projet génome humain (dont il est question au concept 21.1) permettent de penser que l'épissage différentiel de l'ARN est l'une des raisons qui font que les humains possèdent à peu près le même nombre de gènes qu'un nématode (ver rond).

▼ **Figure 17.13** L'épissage de l'ARN prémessager par un complexe d'épissage. Ce schéma montre une partie du transcrite d'ARN prémessager ainsi qu'un intron (pâle) bordé de deux exons (foncés). De petites molécules d'ARN situées dans le complexe d'épissage s'apparient à des nucléotides positionnés à des endroits précis le long de l'intron. Ensuite, les petites molécules d'ARN de ce complexe catalysent la coupure de l'ARN prémessager ainsi que l'assemblage des exons. Cet épissage libère l'intron, qui est alors rapidement dégradé.

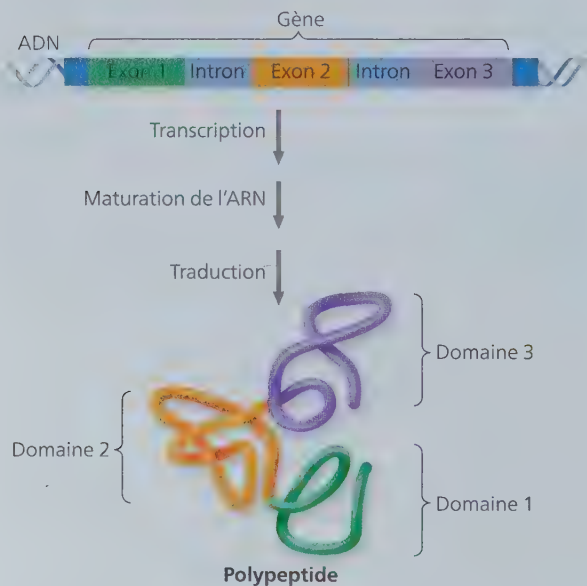


L'épissage différentiel de l'ARN explique également pourquoi un organisme produit un nombre de protéines différentes beaucoup plus élevé que le nombre de ses gènes.

Les protéines ont souvent une architecture modulaire comportant des régions structurales et fonctionnelles discontinues, appelées **domaines**. Par exemple, un des domaines d'une protéine enzymatique peut comprendre le site actif de celle-ci, alors qu'un autre permet à l'enzyme de se fixer à une membrane cellulaire. Dans bon nombre de cas, différents exons codent pour chacun des domaines d'une protéine donnée (**figure 17.14**).

Par ailleurs, il se pourrait que la présence des introns dans un gène favorise l'apparition de nouvelles protéines utiles grâce à un processus appelé *échange ou brassage d'exons* (voir la figure 21.16). Les introns rendraient plus probables les enjambements entre les exons des allèles d'un gène (par la simple création de nouveaux sites possibles d'enjambement sans interruption des séquences codantes). Il en résulterait de nouvelles combinaisons d'exons et de protéines présentant des modifications de leur structure et de leur fonction. On peut également imaginer que des exons puissent se mélanger sur un même gène ou s'intégrer à des gènes complètement différents. Ces deux sortes d'échanges peuvent mener à l'apparition de protéines ayant des fonctions nouvelles. Alors que la plupart des échanges ou brassages n'aboutissent pas à des changements positifs, il arrive qu'une variante génétique bénéfique apparaisse.

▼ **Figure 17.14** La correspondance entre les exons et les domaines des protéines.



RETOUR SUR LE CONCEPT **17.3**

1. On compte environ 20 000 gènes humains codant pour des protéines. Comment les cellules humaines peuvent-elles alors produire de 75 000 à 100 000 protéines différentes ?

2. En quoi l'épissage de l'ARN ressemble-t-il au visionnement d'une émission de télévision préalablement enregistrée ? Dans cette analogie, à quoi pourraient correspondre les introns ?
3. **ET SI ?** ► Si on traitait des cellules avec un agent qui enlève la coiffe des ARNm, quelles seraient les conséquences ?

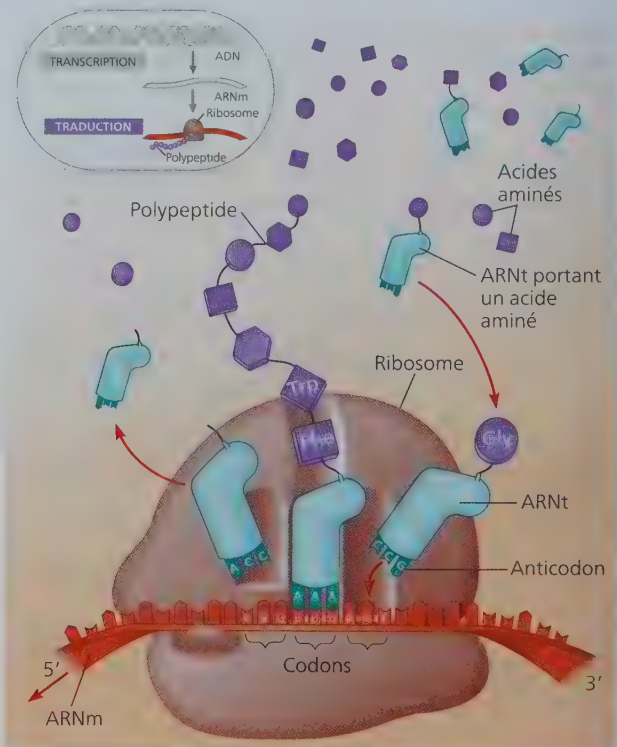
Voir les réponses proposées à l'appendice A.

CONCEPT **17.4**

La traduction est la synthèse d'un polypeptide à partir de l'ARN messenger: une étude détaillée

Nous allons étudier maintenant la traduction, c'est-à-dire le mode de transmission de l'information génétique de l'ARNm à la protéine (**figure 17.5**). Nous nous intéresserons avant tout aux principales étapes de la traduction chez les bactéries et les organismes eucaryotes, tout en soulignant les différences essentielles qui existent entre ces deux groupes.

▼ **Figure 17.15** Le concept de base de la traduction. Les codons sont traduits en acides aminés un par un, à mesure que la molécule d'ARNm traverse le ribosome. Ce sont des molécules d'ARNt qui les traduisent ou les interprètent. Chaque type d'ARNt porte un anticodon donné à une de ses extrémités et un acide aminé correspondant à l'autre extrémité. Lorsque son anticodon se lie à un codon complémentaire situé sur l'ARNm, l'ARNt ajoute son acide aminé à l'extrémité de la chaîne polypeptidique en cours de synthèse. Il y a une relation de spécificité entre l'anticodon et l'acide aminé que l'ARNt transporte.



Les composants moléculaires de la traduction

Au cours de la traduction, la cellule construit une protéine à partir des instructions qu'elle « lit » dans le message génétique. Ce message consiste en une série de codons alignés sur une molécule d'ARNm, et le traducteur est une molécule d'ARN d'un autre type, qui porte le nom d'**ARN de transfert (ARNt)**. Cet ARN a pour fonction d'acheminer des molécules d'acides aminés présentes dans le cytosol vers un polypeptide en cours de synthèse dans un ribosome. Il faut savoir que la cellule garde en réserve une quantité des 20 acides aminés dans son cytosol, soit en les synthétisant à partir d'autres composés, soit en les prélevant dans la solution entourant la cellule. Le ribosome, une structure constituée de protéines et d'ARN, ajoute les acides aminés que chacun des ARNt lui apporte à l'extrémité de la chaîne polypeptidique en cours de synthèse (figure 17.15).

Dans son principe, la traduction est simple; en réalité, elle repose sur des phénomènes biochimiques et des mécanismes complexes, notamment chez les cellules eucaryotes. Nous allons l'analyser plus en détail en nous intéressant avant tout à ce qui se produit chez les bactéries, où le mécanisme est un peu plus simple. Nous verrons d'abord quels sont les principaux acteurs dans ce processus cellulaire, et ensuite comment ils agissent conjointement pour fabriquer un polypeptide.

La structure et la fonction de l'ARN de transfert

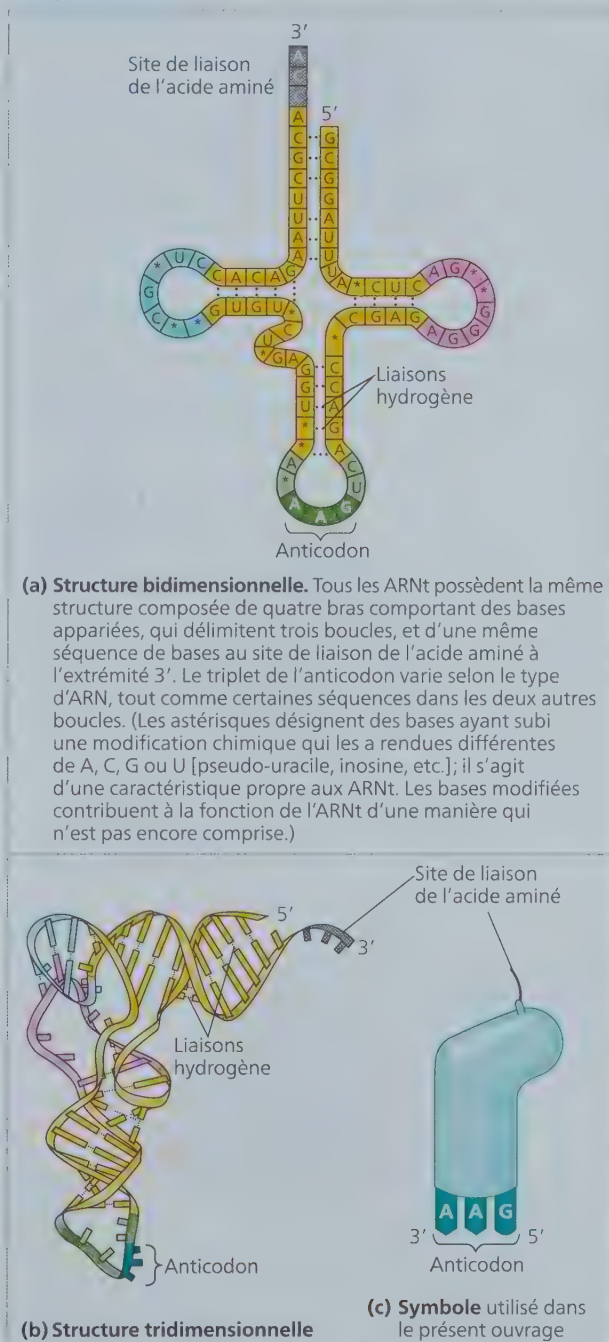
Le fait que les molécules d'ARNt ne sont pas toutes identiques constitue la clé de la traduction du message génétique en une séquence d'acides aminés. Chacun de ces ARNt permet la traduction d'un codon spécifique de l'ARNm en un acide aminé particulier. Tout cela est possible parce que chaque ARNt porte un acide aminé précis à une extrémité de sa structure tridimensionnelle, alors que son autre extrémité porte un triplet de nucléotides qui s'apparie au codon complémentaire sur l'ARNm.

La molécule d'ARNt est formée d'un seul brin d'ARN dont la longueur ne dépasse pas 80 nucléotides environ. (Par comparaison, la plupart des molécules d'ARNm contiennent des centaines de nucléotides.) À cause de la présence de séquences de bases nucléotidiques complémentaires qui peuvent s'associer les unes aux autres par des liaisons hydrogène, ce brin d'ARN se replie sur lui-même et forme une molécule dont la structure est tridimensionnelle. Si l'on représente la molécule d'ARNt à plat pour montrer l'emplacement des zones d'appariement des bases, on voit qu'elle prend la forme d'une feuille de trèfle (**figure 17.16a**). Cependant, en réalité, l'ARNt se tord et se replie de façon à former une structure tridimensionnelle compacte, qui ressemble vaguement à un L inversé (**figure 17.16b**), les extrémités 5' et 3' de l'ARNt linéaire étant toutes deux réunies près d'une extrémité de la structure (correspondant au bras court du L). L'extrémité 3' proéminente agit en tant que site de fixation d'un acide aminé. La boucle qui dépasse à l'extrémité de la partie longue du L inversé porte l'**anticodon** (soit le triplet de bases spécialisé qui se lie à un codon d'ARNm spécifique). La structure de l'ARNt reflète donc sa fonction.

Par convention, on écrit les anticodons dans le sens 3' → 5' pour pouvoir les aligner avec les codons, qui sont écrits dans le sens 5' → 3' (voir la figure 17.15). Pour que les bases azotées puissent s'apparier, il faut que les brins d'ARN soient antiparallèles, comme dans le cas de l'ADN. Prenons, par exemple, le codon d'ARNm 5'-GGC-3', qui commande l'acide aminé glycine. L'ARNt qui se lie à ce codon par des liaisons hydrogène

porte l'anticodon 3'-CCG-5' à une de ses extrémités et la glycine à l'autre (voir l'ARNt qui s'approche du ribosome dans la figure 17.15). Lorsque la molécule d'ARNm avance à travers le ribosome, une glycine est ajoutée à l'extrémité de la chaîne polypeptidique chaque fois que le codon 5'-GGC-3' se présente au

▼ **Figure 17.16** La structure de l'ARN de transfert (ARNt).



HABILETÉS VISUELLES ► Examinez l'ARNt présenté dans cette figure. Indiquez à quel codon l'anticodon se lierait et quel acide aminé il transporterait.

site de traduction. Notons que les ARNt traduisent le message génétique un codon à la fois : ils déposent les acides aminés dans l'ordre requis, et le ribosome les ajoute à la chaîne. La molécule d'ARNt est un traducteur au sens où, grâce au contexte du ribosome, elle lit un mot (le codon) de l'ARNm et l'interprète en termes de protéine (en acides aminés).

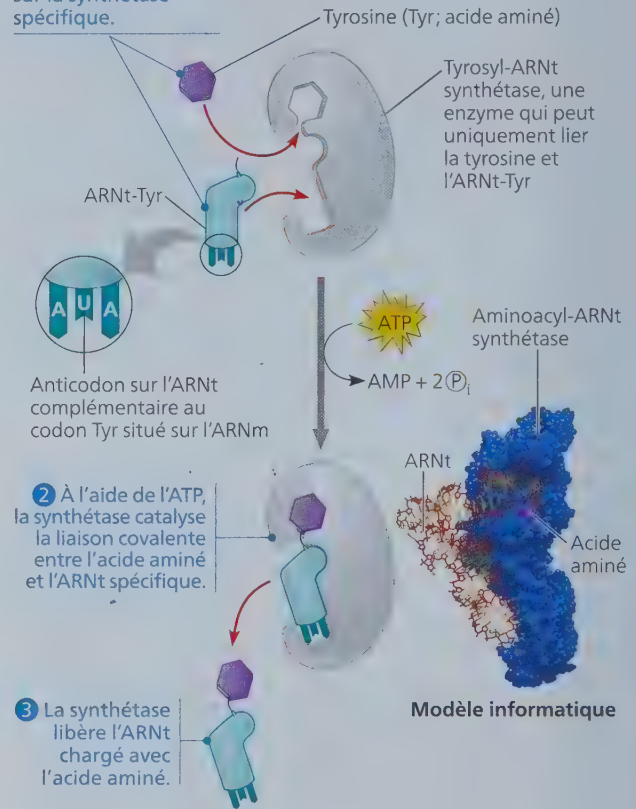
À l'instar de l'ARNm et des autres types d'ARN de la cellule, les molécules d'ARNt sont transcrites à partir de matrices d'ADN. Dans la cellule eucaryote, l'ARNt, tout comme l'ARNm, est produit dans le noyau et doit passer de celui-ci au cytoplasme, où il participera au processus de traduction. Que ce soit dans les cellules bactériennes ou eucaryotes, chaque molécule d'ARNt sert de nombreuses fois : elle se lie d'abord à l'acide aminé qui lui correspond dans le cytosol ; elle le cède ensuite à une chaîne polypeptidique au ribosome, puis elle quitte ce dernier pour aller chercher un nouvel acide aminé identique dans le cytosol.

La traduction exacte du message génétique nécessite une double reconnaissance moléculaire. Premièrement, un ARNt qui s'associe à un codon d'ARNm commandant un acide aminé précis ne doit pouvoir apporter que cet acide aminé au ribosome, et aucun autre. L'appariement correct de l'ARNt et de l'acide aminé est effectué par une famille d'enzymes apparentées nommées **aminoacyl-ARNt synthétases** (figure 17.17). Le site actif de chaque type d'enzyme ne peut former qu'une seule combinaison d'acide aminé et d'ARNt. Il existe 20 types d'aminoacyl-ARNt synthétase dans la cellule, soit une par sorte d'acide aminé. Une synthétase lie un acide aminé donné à un ARNt approprié ; une seule synthétase peut lier chacun des différents ARNt correspondant à un acide aminé spécifique. L'aminoacyl-ARNt synthétase catalyse la liaison covalente d'un acide aminé et de son ARNt suivant un processus alimenté par l'hydrolyse d'ATP. Le complexe acide aminé-ARNt ainsi formé (aussi appelé ARNt chargé) se détache ensuite de l'enzyme et est alors prêt à ajouter son acide aminé au bout d'une chaîne polypeptidique en cours de formation sur un ribosome.

Deuxièmement, l'anticodon d'un ARNt doit s'apparier avec le codon approprié d'un ARNm. Des expériences ont montré que même si on modifie l'acide aminé porté par un ARNt, cela ne change pas l'appariement de l'ARNt et de l'ARNm : l'ARNt se placera toujours sur l'ARNm à l'endroit où devrait se lier l'acide aminé original, car c'est l'anticodon que porte l'ARNt et non son acide aminé qui dicte l'appariement. Si un ARNt différent correspondait à chaque codon d'ARNm commandant un acide aminé, il y aurait 61 ARNt (voir la figure 17.6). Chez les bactéries, on n'en compte toutefois soit qu'environ 45, ce qui signifie que certains ARNt pourraient se lier à plus d'un codon. Une telle souplesse est rendue possible parce que les règles qui dictent l'appariement de la troisième base nucléotidique d'un codon et de la base correspondante de l'anticodon d'ARNt sont plus souples que celles qui régissent d'autres positions de codons. Par exemple, la base nucléotidique U à l'extrémité 5' d'un anticodon d'ARNt peut s'associer soit à la base A, soit à la base G en troisième position (à l'extrémité 3') d'un codon d'ARNm. Cet appariement flexible des bases à cette position des codons est appelé **oscillation**. Le phénomène d'oscillation permet d'expliquer pourquoi les codons synonymes, codant pour un même acide aminé, ne diffèrent souvent que par leur troisième base nucléotidique. Par exemple, un ARNt avec l'anticodon 3'-UCU-5' peut s'apparier aussi bien avec le codon 5'-AGA-3' que 5'-AGG-3' de l'ARNm, les deux codant pour l'arginine (voir la figure 17.6).

▼ Figure 17.17 L'appariement spécifique des acides aminés et de leur ARNt par l'aminoacyl-ARNt synthétase. La formation d'une liaison entre l'ARNt et son acide aminé est un processus endergonique. Au cours de la réaction, l'ATP perd deux groupements phosphate et se transforme en AMP (adénosine monophosphate).

1 L'acide aminé et l'ARNt approprié pénètrent le site actif sur la synthétase spécifique.



La structure des ribosomes et leur fonction

Les ribosomes permettent l'appariement des anticodons d'ARNt avec les codons d'ARNm au cours de la synthèse des protéines. Un ribosome est formé d'une grande sous-unité et d'une petite sous-unité. Chacune d'elles est constituée de protéines et d'un ou de plusieurs **ARN ribosomiques (ARNr)**. Chez les eucaryotes, les sous-unités sont produites dans le nucléole, une région du noyau un peu plus dense (ce qui la rend visible au microscope photonique dans certaines conditions). Les gènes de l'ARN ribosomique sont transcrits, puis subissent une maturation et les ARNr sont assemblés avec des protéines importées du cytosol. Les sous-unités ribosomiques ainsi fabriquées sont ensuite exportées dans le cytosol par les pores nucléaires. Chez tous les organismes, la petite et la grande sous-unité ne s'unissent pour former un ribosome fonctionnel qu'au moment où elles se fixent à une molécule d'ARNm. Près du tiers de la masse d'un ribosome est constitué de protéines ; le reste se compose de trois molécules d'ARNr chez les bactéries et de quatre chez les eucaryotes. Comme la plupart des cellules contiennent des milliers de ribosomes, l'ARNr est le type d'ARN cellulaire le plus abondant.

Si les ribosomes des cellules bactériennes et eucaryotes ont une structure et des fonctions très semblables, ceux des eucaryotes sont un peu plus gros, et leur composition moléculaire est quelque peu différente. Cette différence est importante du point de vue médical pour lutter contre les infections causées par les bactéries. En effet, il est possible d'utiliser des antibiotiques, dont la tétracycline et la streptomycine, qui paralysent les ribosomes des cellules bactériennes sans nuire aux ribosomes eucaryotes.

La structure d'un ribosome reflète sa fonction, qui est de rapprocher un ARNm des ARNt porteurs d'acides aminés. Chaque ribosome comprend, outre un site de liaison à l'ARNm, trois sites de liaison à l'ARNt (**figure 17.18**). Le **site P** (site de liaison du peptidyl-ARNt) retient l'ARNt qui porte la chaîne polypeptidique en cours de synthèse. Le **site A** (site de liaison de l'aminocyl-ARNt) retient l'ARNt portant le prochain acide aminé qui viendra se joindre à la chaîne. C'est à partir du **site E** (site de sortie, *exit*) que l'ARNt quitte le ribosome. L'ensemble des sites du ribosome permet de rapprocher les ARNt et l'ARNm, de placer le nouvel acide aminé de façon à l'ajouter à l'extrémité carboxyle du polypeptide en cours de synthèse et de catalyser ensuite la formation de la liaison peptidique. Le polypeptide, qui continue de s'allonger, émerge alors par un *tunnel de sortie* dans la grande sous-unité ribosomique. Une fois sa synthèse terminée, c'est par cet orifice que le polypeptide rejoint le cytosol.

D'après le modèle généralement accepté, la structure et la fonction des ribosomes relèvent principalement de l'ARNr plutôt que des protéines ribosomales. Les protéines, principalement localisées en périphérie du ribosome, assistent les molécules d'ARNr dans leur modification de structure lorsqu'elles effectuent la catalyse au cours de la traduction. L'ARN ribosomique est le principal composant des sites A et P ainsi que de l'interface entre les deux sous-unités. C'est également lui qui catalyse la formation de la liaison peptidique. On peut donc considérer le ribosome comme un énorme ribozyme !

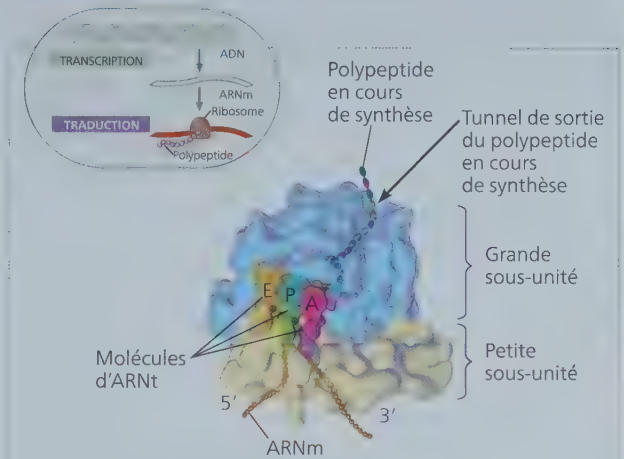
La synthèse d'un polypeptide

La traduction, ou synthèse d'un polypeptide, comprend trois étapes principales : il s'agit de l'initiation, de l'élongation et de la terminaison. Celles-ci ne peuvent se dérouler qu'en présence de « facteurs » protéiques qui les assistent pendant la traduction. Certaines étapes de l'initiation et de l'élongation de la chaîne nécessitent également un apport énergétique fourni par l'hydrolyse de la guanosine triphosphate (GTP).

La liaison au ribosome et l'initiation de la traduction

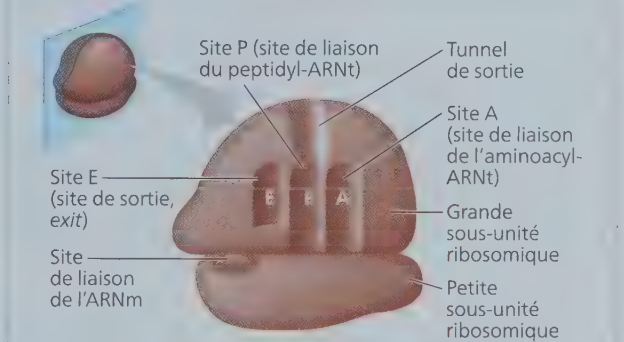
Chez les bactéries comme chez les eucaryotes, le codon de départ (AUG) précise le point de départ de la traduction ; cette étape est déterminante, car elle sert d'assise à la phase de lecture de l'ARNm. Pendant la première étape de la traduction, la petite sous-unité ribosomique s'attache à la fois à un ARNm et à un ARNt spécifique d'initiation qui porte un acide aminé méthionine. Chez les bactéries, la petite sous-unité peut se lier à ces deux acides ribonucléiques dans n'importe quel ordre ; elle se fixe à l'ARNm sur une séquence spécifique, juste en amont du codon AUG de départ. Dans l'exercice de la rubrique **Habilités scientifiques**, vous aurez à travailler avec les séquences d'ADN codant les sites de liaison des ribosomes situés sur l'ARNm d'un groupe de gènes d'*E. coli*. Chez les eucaryotes, la petite sous-unité déjà associée à l'ARNt d'initiation se fixe à la coiffe 5' de l'ARNm et se déplace

▼ Figure 17.18 L'anatomie d'un ribosome fonctionnel.

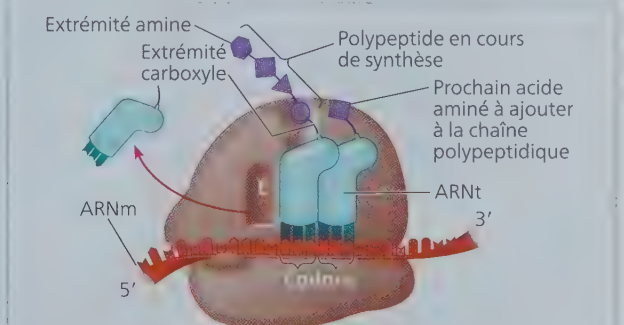


(a) Modèle informatique d'un ribosome fonctionnel.

Ce modèle montre la forme générale d'un ribosome bactérien. Les ribosomes des eucaryotes sont à peu près semblables. Chaque sous-unité est un complexe de molécules d'ARN ribosomique et de protéines.



(b) Schéma montrant les sites de liaison. Un ribosome comprend un site de liaison de l'ARNm ainsi que trois sites de liaison de l'ARNt, appelés A, P et E. Nous reverrons ce schéma dans d'autres illustrations.



(c) Schéma montrant l'ARNm et l'ARNt en interaction.

Un ARNt s'unit à un site de liaison lorsque les bases de son anticodon s'apparient avec celles d'un codon d'ARNm. Le site P retient l'ARNt attaché au polypeptide en cours de synthèse. Le site A retient l'ARNt portant le prochain acide aminé qui s'ajoutera à la chaîne polypeptidique. L'ARNt libéré se détache du ribosome par le site E. C'est par l'extrémité carboxyle (extrémité C-terminale) que s'allonge le polypeptide.



Interpréter un logo de séquence

■ COMMENT PEUT-ON UTILISER UN LOGO DE SÉQUENCE D'ADN POUR CIBLER LES SITES DE LIAISON DES RIBOSOMES SUR LES ARNm D'UNE BACTÉRIE ? ■

Lorsqu'ils amorcent la traduction, les ribosomes se fixent au site de liaison des ribosomes d'une molécule d'ARNm, en amont du codon de départ AUG. Comme les ARNm des différents gènes se lient tous à un ribosome, il est probable que les gènes qui encodent ces ARNm présentent une séquence de bases comparable au site de liaison des ribosomes. Il est donc possible d'identifier les sites de liaison potentiels des ribosomes sur l'ARNm en comparant les séquences d'ADN (et, par conséquent, les séquences d'ARNm) des différents gènes d'une espèce et en examinant la région en amont du codon de départ afin de déceler les séquences de bases communes (conservées). Dans cet exercice, vous analyserez les séquences d'ADN de différents gènes représentées sous forme de logo de séquence, un outil de représentation graphique.

■ MÉTHODE ■ Les séquences d'ADN de 149 gènes du génome d'*E. coli* ont été alignées à l'aide d'un logiciel. L'exercice avait pour but d'identifier des séquences de bases similaires – à l'endroit approprié dans chaque gène – et constituant des sites de liaison potentiels des ribosomes. Plutôt que de présenter les données sous forme de séries de 149 séquences alignées dans une colonne (alignement de séquences), les chercheurs ont créé un logo de séquence.

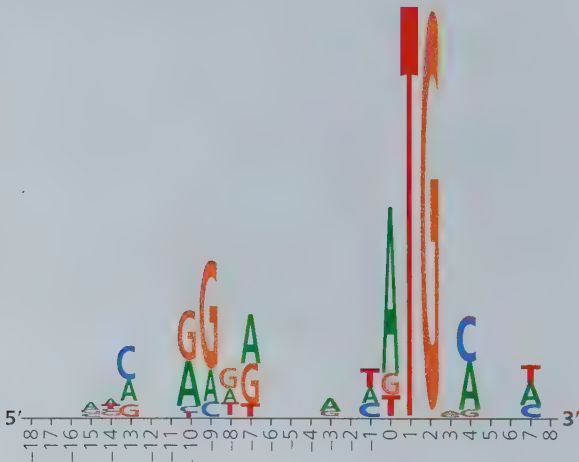
■ RÉSULTATS ■ Afin d'illustrer comment on crée un logo de séquence, nous avons présenté ci-dessous les sites de liaison potentiels des ribosomes de 10 gènes d'*E. coli* dont les séquences ont été alignées, puis dans un logo de séquence réalisé à partir de ces alignements. Il est à noter que la partie de l'ADN qui est présentée ici est le brin d'ADN codant (non transcrit) puisque c'est ainsi que sont généralement présentées les séquences d'ADN.

```

thrA  G G T A A C G A G G T A A C A A C C A T G C G A G T G
lacA  C A T A A C G G A G T G A T C G C A T T G A A C A T G
lacY  C G C G T A A G G A A A T C C A T T A T G T A C T A T
lacZ  T T C A C A C A G G A A A C A G C T A T G A C C A T G
lacI  C A A T T C A G G G T G G T G A A T G T G A A A C C A
recA  G G C A T G A C A G G A G T A A A A A T G G C T A T C
galR  A C C C A C T A A G G T A T T T T C A T G G C G A C C
metJ  A A G A G G A T T A A G T A T C T C A T G G C T G A A
lexA  A T A C A C C C A G G G G G C G G A A T G A A A G C G
trpR  T A A C A A T G G C G A C A T A T T A T G G C C C A A
5'-----18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -2 3 4 5 6 7 8 3'
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

▲ Alignement des séquences.



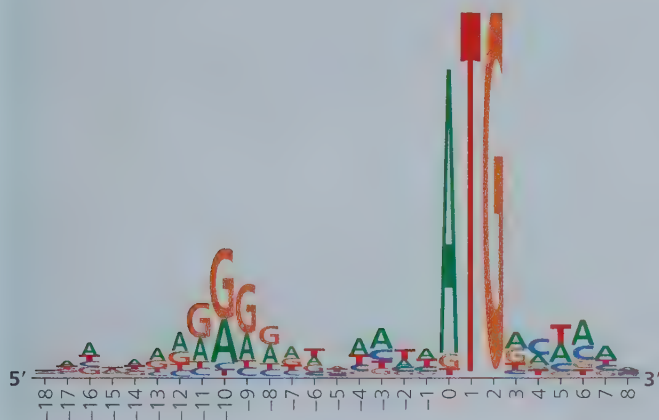
▲ Logo de séquence.

INTERPRÉTEZ LES DONNÉES ▼

1. Dans le logo de séquence (en bas, à gauche), l'axe horizontal montre la séquence principale de l'ADN par position nucléotidique. Les lettres correspondant à chacune des bases sont empilées les unes par-dessus les autres selon leur fréquence relative à cette position, parmi les séquences alignées. La hauteur d'une lettre représente la fréquence relative d'une base à cette position. La plus grosse lettre, placée au sommet de chaque pile, indique donc la base la plus fréquente. (a) Dans l'alignement de séquences, comptez le nombre de bases de chaque type à la position -9, puis classez ces bases en ordre décroissant, soit de la plus fréquente à la moins fréquente. Comparez vos résultats à la hauteur et à l'emplacement des bases de chaque type à la position -9 dans le logo. (b) Répétez l'exercice pour les positions 0 et 1.
2. Dans un logo, la hauteur d'une pile de lettres indique la puissance prédictive de cette pile (établie au moyen d'une méthode statistique). Plus une pile est haute, plus il est possible de prévoir avec certitude quelle base occupera cette position dans l'éventualité où une nouvelle séquence serait ajoutée au logo. Par exemple, à la position 2 dans l'alignement de séquences, la lettre G est présente dans les 10 séquences; les chances que la lettre G se trouve dans une nouvelle séquence sont donc très élevées, comme le montre la hauteur de la pile dans le logo de séquence. Lorsqu'une pile est peu élevée, les bases présentent toutes environ la même fréquence. Par conséquent, il est difficile de prévoir quelle base occupera cette position. (a) Examinez le logo de séquence et indiquez quelles sont les deux positions pour lesquelles les bases sont les plus prévisibles. Selon vous, quelles bases occuperont ces positions dans un gène nouvellement séquencé ? (b) Quelles sont les 12 positions pour lesquelles les bases sont le moins prévisibles ? Justifiez votre réponse. Comment cela reflète-t-il les fréquences relatives des bases occupant ces positions dans l'alignement des séquences ? Dans votre réponse, utilisez les deux positions les plus à gauche, parmi les 12 positions, à titre d'exemple.

Interpréter un logo de séquence (suite)

3. Dans la véritable expérience, les chercheurs ont utilisé 149 séquences pour créer le logo de séquence présenté ci-dessous. Comme le nombre de données est plus élevé dans un logo de séquence, on y trouve une pile par position, même si cette pile peut être moins élevée. (a) Dans ce logo de séquence, quelles sont les trois positions pour lesquelles les bases sont le plus prévisibles ? Nommez les bases les plus fréquentes à chaque position. (b) Quelles sont les quatre positions pour lesquelles les bases sont le moins prévisibles ? Justifiez votre réponse.



4. Une séquence consensuelle représente la base la plus fréquente à chaque position pour l'ensemble des séquences. (a) Quelle est la séquence consensuelle de ce brin codant d'ADN (non transcrit) ? Placez un tiret vis-à-vis des positions pour lesquelles il est impossible de prévoir la base. (b) Entre la séquence consensuelle et le logo de séquence, quelle méthode procure le plus de renseignements ? Quels sont les renseignements en moins avec la méthode la moins détaillée ?

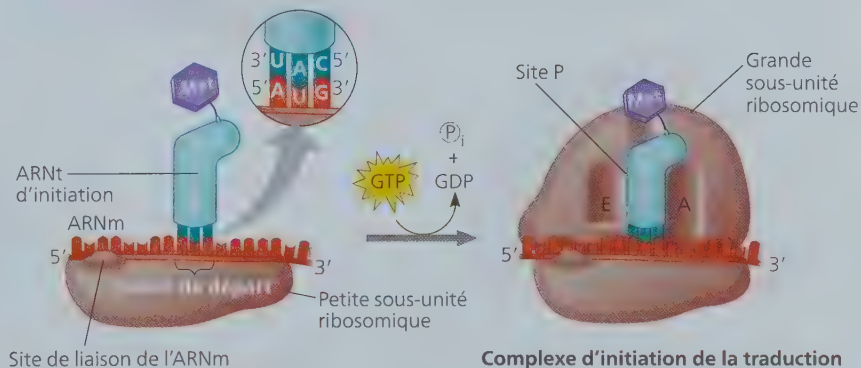
5. (a) D'après le logo, quelles sont les cinq positions de bases adjacentes dans la région 5' UTR qui sont le plus susceptibles d'intervenir dans la fixation ribosomique ? Expliquez votre réponse. (b) Que représentent les bases aux positions 0 à 2 ?

Pour en savoir plus : T. D. Schneider et R. M. Stephens, Sequence logos: A new way to display consensus sequences, *Nucleic Acides Research* 18: 6097-6100 (1990).

ensuite sur l'ARNm vers l'aval – elle effectue un balayage –, jusqu'à ce qu'elle atteigne le codon d'initiation ; l'ARNt d'initiation établit alors des liaisons hydrogène avec le codon AUG de départ.

Par conséquent, les premiers composants à s'associer les uns aux autres pendant la phase d'initiation de la traduction sont l'ARNm, un ARNt portant le premier acide aminé du polypeptide et la petite sous-unité ribosomique (figure 17.19). L'union de l'ARNm, de l'ARNt d'initiation et de la petite sous-unité ribosomique est suivie de l'arrivée de la grande sous-unité ribosomique, ce qui achève la constitution du complexe d'initiation de la traduction. L'assemblage de tous ces composants ne peut se faire qu'en présence de protéines appelées *facteurs d'initiation*. Pour former un complexe d'initiation, la cellule obtient de l'énergie grâce à l'hydrolyse de la molécule de GTP. À la fin du processus d'initiation, l'ARNt d'initiation se retrouve sur le site P du ribosome, et le site A, vacant, est prêt à recevoir le prochain aminoacyl-ARNt.

▼ Figure 17.19 L'initiation de la traduction.



- 1 Une petite sous-unité ribosomique se lie à une molécule d'ARNm. Dans la cellule bactérienne, le site de cette sous-unité auquel se fixe l'ARNm reconnaît une séquence nucléotidique spécifique située sur l'ARNm, à peu de distance en amont du codon de départ. Les bases de l'ARNt d'initiation, qui portent l'anticodon UAC, s'apparient avec le codon de départ AUG. Cet ARNt d'initiation porte l'acide aminé méthionine (Met).
- 2 L'organisation du complexe d'initiation est complétée par l'arrivée de la grande sous-unité ribosomique. Des protéines appelées *facteurs d'initiation* (non représentées ici) permettent de regrouper tous ces éléments en vue de la traduction. L'hydrolyse de la GTP fournit l'énergie nécessaire à l'assemblage. L'ARNt d'initiation se trouve au site P, et le site A est prêt à recevoir l'ARNt portant le prochain acide aminé.

Notez qu'un polypeptide est toujours synthétisé dans le même sens, à partir de la méthionine à l'extrémité amine, ou extrémité N-terminale, vers l'acide aminé final à l'extrémité carboxyle, ou extrémité C-terminale (voir la figure 5.15).

L'élongation de la chaîne polypeptidique

L'élongation est l'étape de la traduction au cours de laquelle les acides aminés sont ajoutés un par un à l'extrémité carboxyle de la chaîne en cours de synthèse. Chaque ajout suppose la participation de plusieurs protéines appelées *facteurs d'élongation* et se déroule selon un cycle comptant trois phases, décrites à la **figure 17.20**. La dépense énergétique s'effectue à la première et à la troisième étape. La reconnaissance du codon nécessite l'hydrolyse d'une molécule de GTP, ce qui renforce l'exactitude et l'efficacité de cette étape. Une autre molécule de GTP est hydrolysée à l'étape de la translocation.

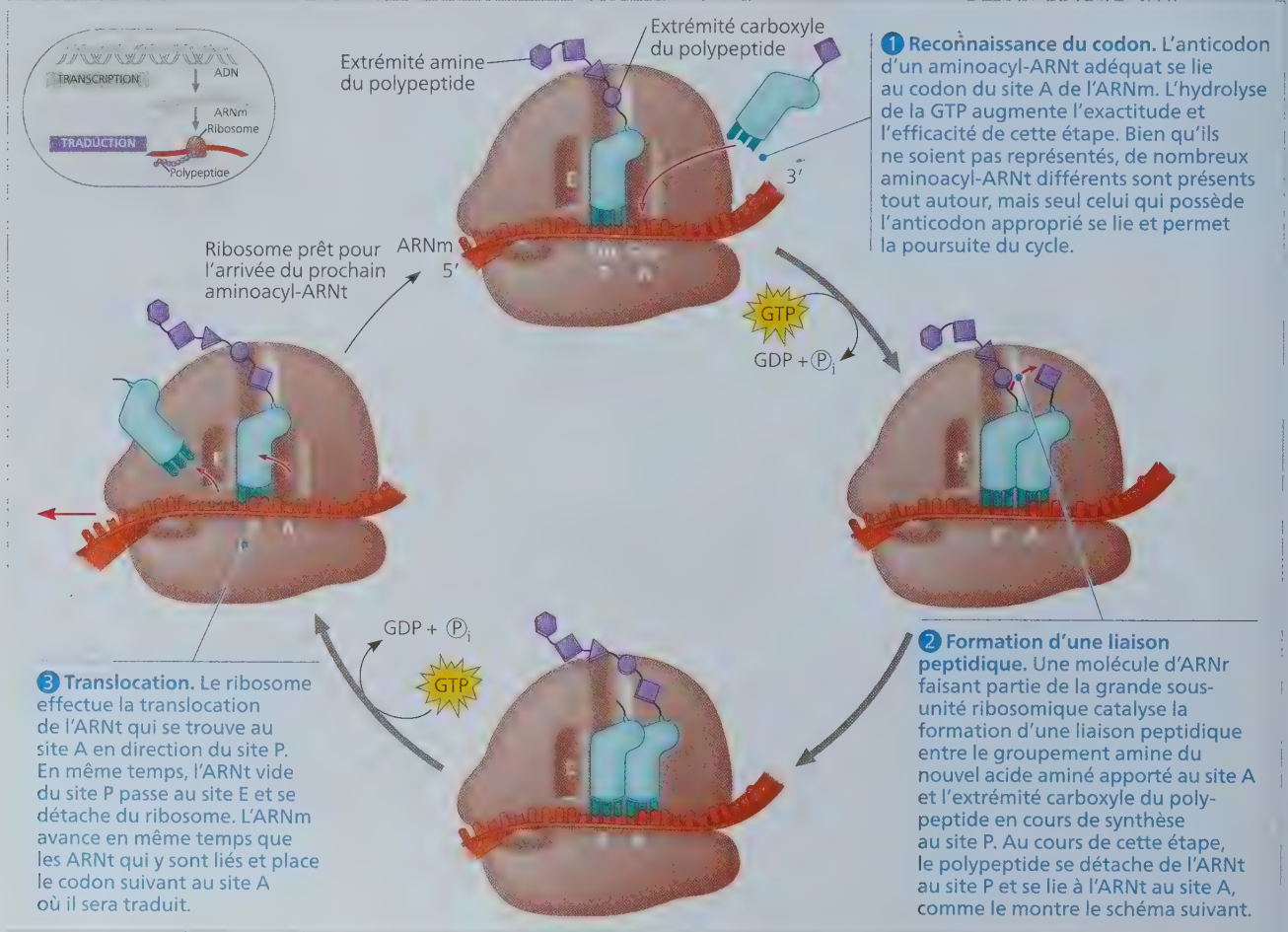
L'ARNm traverse toujours le ribosome dans la même direction, c'est-à-dire en commençant par l'extrémité 5'. Cela revient à dire que le ribosome se déplace dans le sens 5' → 3' sur l'ARNm. Il suffit de se souvenir que le ribosome et l'ARNm

bougent l'un par rapport à l'autre, en sens inverse et codon par codon. Le cycle d'élongation dure moins d'un dixième de seconde chez les bactéries et se répète chaque fois qu'un acide aminé est ajouté à la chaîne polypeptidique, et ce, jusqu'à ce que celle-ci soit complète. Les ARNt vides qui sont libérés du site E retournent dans le cytoplasme où ils seront à nouveau chargés de l'acide aminé approprié par l'aminoacyl-ARNt synthétase (voir la figure 17.17).

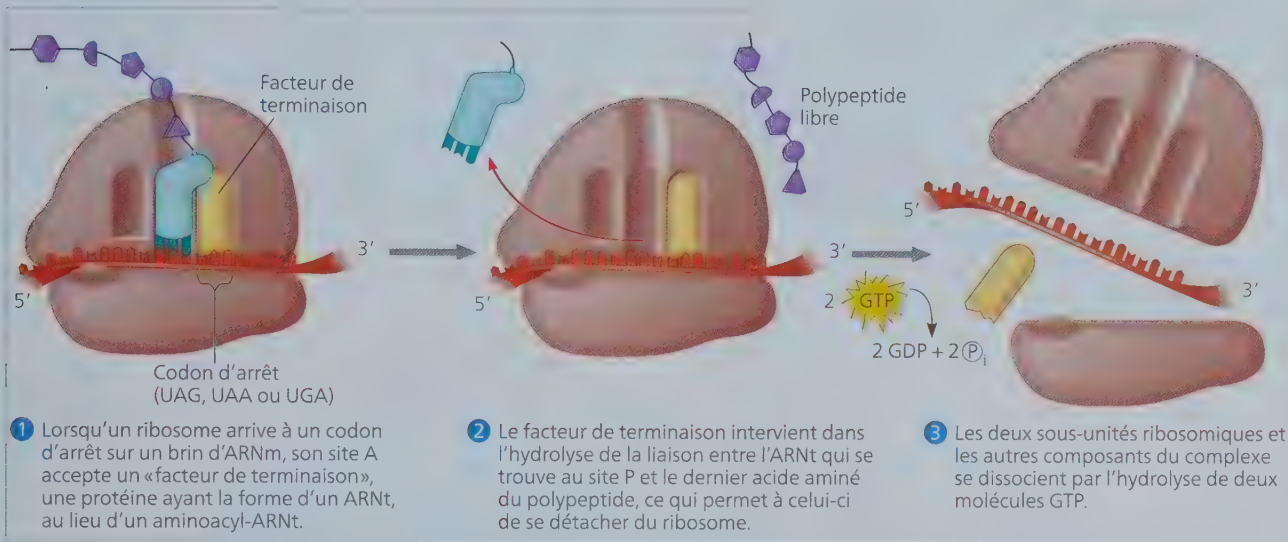
La terminaison de la traduction

La dernière étape de la traduction est la terminaison (**figure 17.21**). L'élongation se poursuit jusqu'à ce que l'un des codons d'arrêt de l'ARNm atteigne le site A du ribosome. Les triplets de bases azotées (5' → 3') UAG, UAA et UGA de l'ARNm ne codent pas pour des acides aminés (ces codons n'ont pas d'anticodons complémentaires), mais servent de signal de fin de la traduction. Un *facteur de terminaison*, une protéine ayant la forme d'un aminoacyl-ARNt, se lie alors directement au codon de terminaison du site A et ajoute une molécule d'eau au lieu d'un acide aminé à la chaîne polypeptidique enfin terminée.

▼ **Figure 17.20** Le cycle d'élongation de la traduction. L'hydrolyse de la GTP joue un rôle important dans le processus d'élongation. Ces schémas ne montrent pas les facteurs d'élongation.



▼ **Figure 17.21** La terminaison de la traduction.



(Les molécules d'eau sont nombreuses dans le cytosol.) La liaison entre la chaîne polypeptidique et l'ARNt qui se trouve au site P est ainsi rompue (hydrolysée), et le polypeptide se détache de la grande sous-unité ribosomique en passant par le tunnel de sortie. Le reste du complexe de traduction se dissocie alors dans un processus en plusieurs étapes, assisté par d'autres facteurs protéiques. La dissociation du complexe de traduction nécessite l'hydrolyse de deux autres molécules de GTP.

L'achèvement et l'acheminement de la protéine fonctionnelle

Souvent, le processus de traduction ne peut synthétiser une protéine fonctionnelle sans aide. Dans la présente section, vous étudierez les modifications que les chaînes de polypeptides subissent après le processus de traduction ainsi que quelques mécanismes utilisés pour acheminer la protéine achevée vers des sites spécifiques dans la cellule.

Le repliement des protéines et les modifications posttraductionnelles

Pendant la synthèse, la chaîne polypeptidique s'enroule et se replie spontanément à cause de sa séquence d'acides aminés (structure primaire), en formant une protéine dotée d'une structure spécifique. En d'autres termes, elle devient une molécule tridimensionnelle possédant une structure secondaire et tertiaire (voir la figure 5.18). Par conséquent, un gène détermine la structure primaire, et cette dernière détermine à son tour les niveaux de structure supérieurs de la molécule.

Avant de pouvoir remplir sa fonction dans la cellule, la protéine doit parfois subir des *modifications posttraductionnelles*, c'est-à-dire des modifications qui s'effectuent après la traduction proprement dite. Certains acides aminés sont modifiés chimiquement par l'ajout de glucides, de lipides, de groupements phosphate ou d'autres substances. Des enzymes peuvent détacher un ou plusieurs acides aminés de l'extrémité amine de la

chaîne polypeptidique. Dans certains cas, une chaîne polypeptidique est découpée en plusieurs fragments par voie enzymatique. Dans d'autres cas, plusieurs polypeptides synthétisés séparément peuvent s'unir si la protéine est pourvue d'une structure quaternaire. L'hémoglobine en est un exemple (voir la figure 5.18).

L'acheminement des polypeptides vers des cibles spécifiques

L'observation au microscope électronique de cellules eucaryotes synthétisant des protéines permet de mettre en évidence deux populations de ribosomes : des ribosomes libres et d'autres qui sont liés (voir la figure 6.10). Les premiers sont en suspension dans le cytosol et synthétisent surtout des protéines qui restent dans le cytosol, où elles remplissent leurs fonctions. Les seconds sont fixés à la face cytoplasmique du réticulum endoplasmique (RE) rugueux ou à celle de l'enveloppe nucléaire; ils synthétisent les protéines du réseau de membranes intracellulaires (l'enveloppe nucléaire, le RE, le complexe golgien, les lysosomes, les vacuoles et la membrane plasmique; voir la figure 6.15) ainsi que celles qui doivent être sécrétées à l'extérieur de la cellule (comme l'insuline). Il est important de noter que tous les ribosomes sont identiques et que, d'une utilisation à l'autre, ils peuvent passer de l'état libre à l'état lié.

Quel est le facteur qui détermine si un ribosome se trouve libre dans le cytosol ou lié au réticulum endoplasmique? Il faut savoir que la synthèse de tout polypeptide commence dans le cytosol lorsqu'un ribosome libre entame la traduction d'une molécule d'ARNm. Le processus se poursuit entièrement dans ce milieu par défaut, *sauf* si le polypeptide en cours de synthèse incite, par une séquence particulière d'acides aminés, le ribosome à se fixer au RE. Par exemple, les protéines destinées au réseau de membranes intracellulaires ou devant être sécrétées portent la **séquence signal** (ou peptide signal) qui les oriente vers le RE (**figure 17.22**). Cette séquence compte environ 20 acides aminés et se situe à l'extrémité amine du polypeptide ou près de

▼ **Figure 17.22** Le mécanisme de signalisation pour l'acheminement des protéines au RE.

1 La synthèse du polypeptide commence sur un ribosome libre dans le cytosol.

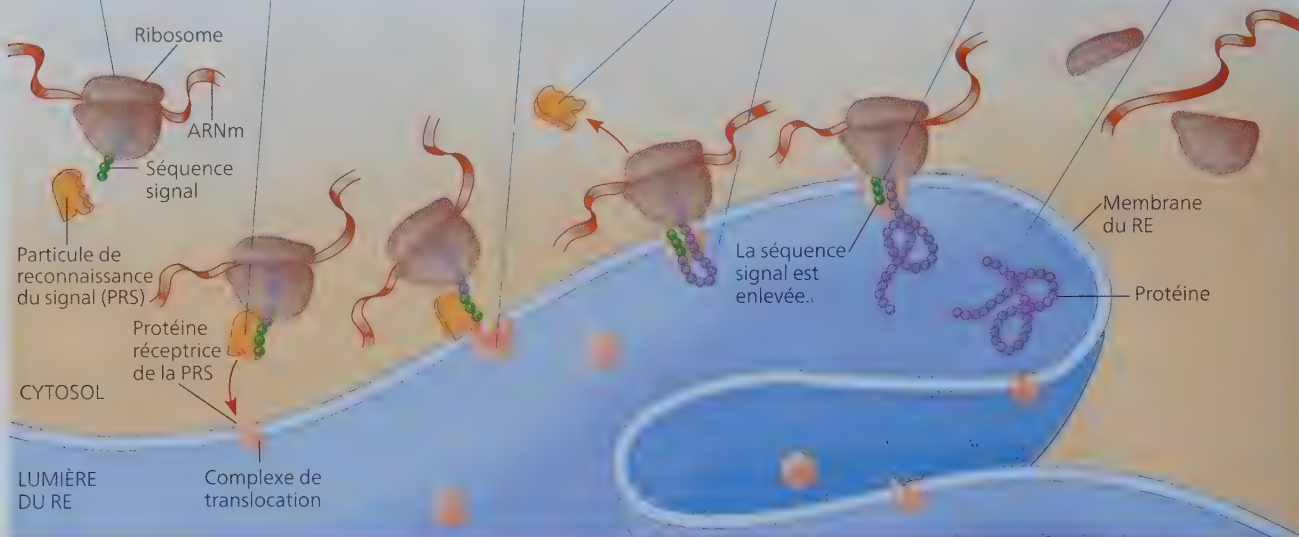
2 Une particule de reconnaissance du signal (PRS) se lie à la séquence signal et interromp momentanément la synthèse polypeptidique.

3 La PRS se lie à une protéine réceptrice de la membrane du RE, une partie d'un complexe protéique (complexe de translocation) qui forme un pore.

4 La PRS se détache, et le polypeptide traverse la membrane pendant que sa synthèse se poursuit.

5 L'enzyme de clivage située dans le complexe protéique coupe la séquence signal.

6 Le polypeptide enfin terminé se détache du ribosome et se plie de façon à prendre sa conformation définitive.



FAITES DES LIENS ► Si la protéine doit être sécrétée, qu'advient-il d'elle une fois sa synthèse terminée? Voir la figure 7.9.

celle-ci. Au moment où la séquence signal émerge du ribosome, elle est reconnue par un complexe appelé **particule de reconnaissance du signal (PRS)**. Celle-ci est constituée de six protéines et d'un petit ARN. Elle agit alternativement comme un interrupteur de la synthèse protéique et elle escorte le ribosome jusqu'à une protéine réceptrice dans la membrane du RE. Cette protéine réceptrice fait partie d'un complexe de translocation multiprotéique. La synthèse du polypeptide se poursuit à cet endroit, et la molécule en voie de formation commence à passer dans la lumière du RE en se faufilant à travers un canal protéique de la membrane. S'il est destiné à devenir une protéine de sécrétion, le polypeptide enfin complété est libéré dans la solution qui remplit la lumière du RE. S'il doit devenir une protéine membranaire, il reste partiellement enchâssé dans la membrane du RE. Dans l'un ou l'autre des cas, une vésicule de sécrétion l'achemine vers sa destination (voir, par exemple, la figure 7.9).

D'autres types de séquences signal dirigent les polypeptides vers les mitochondries, les chloroplastes, l'intérieur du noyau et d'autres organites qui ne font pas partie du réseau intracellulaire de membranes. Dans ces cas, la principale particularité est que la traduction prend fin dans le cytosol, avant que le polypeptide soit exporté vers l'organite auquel il est destiné. Les mécanismes de translocation sont également variables; mais, dans tous les cas qui ont été étudiés jusqu'à aujourd'hui, on a mis en évidence divers types de séquence signal faisant office de « codes postaux » qui sont chargés d'acheminer les protéines vers un

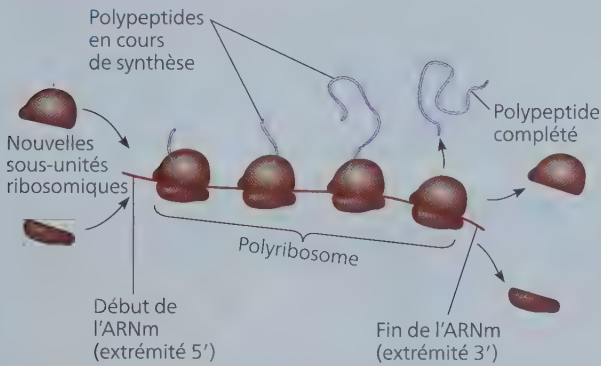
endroit particulier de la cellule ou qui les destinent à devenir des protéines de sécrétion. Les bactéries emploient également des séquences signal pour acheminer les protéines destinées à devenir des protéines de sécrétion vers la membrane plasmique.

La production de multiples copies d'un polypeptide chez les bactéries et les eucaryotes

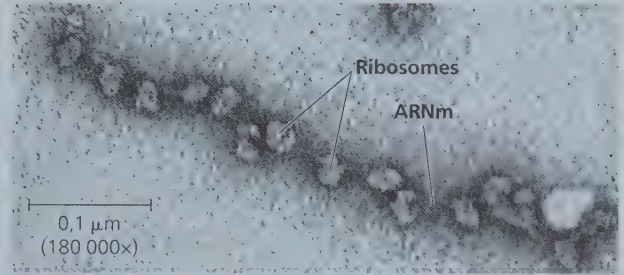
Dans les sections précédentes, vous avez appris comment un polypeptide est synthétisé à l'aide de l'information encodée dans une molécule d'ARNm. Toutefois, lorsqu'une cellule a besoin d'un polypeptide, une seule copie ne suffit pas: il lui en faut beaucoup.

Un ribosome peut synthétiser à lui seul un polypeptide de taille moyenne en moins d'une minute. Toutefois, pour produire rapidement un polypeptide donné en plus grande quantité, on observe chez les bactéries et chez les eucaryotes que plusieurs ribosomes s'associent et traduisent simultanément le même ARNm (**figure 17.23**); autrement dit, une même molécule d'ARNm sert en général à synthétiser simultanément un grand nombre de copies de ce polypeptide. En effet, dès qu'un ribosome s'éloigne suffisamment du codon de départ, un autre peut se lier à l'ARNm à son tour, et ainsi de suite. Le résultat est que de nombreux ribosomes forment une file le long d'une même molécule d'ARNm. On peut observer au microscope électronique des chapelets de ribosomes, appelés **polyribosomes** (ou parfois

▼ **Figure 17.23** Les polyribosomes.



(a) Une molécule d'ARNm est généralement traduite simultanément par plusieurs ribosomes. L'ensemble de ces ribosomes forme un polyribosome.



(b) Cette micrographie montre un polyribosome dans une cellule bactérienne. Les polypeptides en cours de synthèse ne sont pas visibles ici (MET).

polysomes), que l'on trouve à l'état libre ou à l'état lié. Les cellules bactériennes et eucaryotes contiennent des polyribosomes, mais ceux-ci sont en général plus longs chez les premières. Grâce à eux, une cellule a la capacité de synthétiser de nombreuses copies d'un polypeptide très rapidement.

La transcription de plusieurs ARNm d'un même gène est un autre moyen qui permet aux bactéries et aux eucaryotes d'augmenter le nombre de copies d'un polypeptide. Cependant, ces deux groupes se distinguent par la façon dont ils coordonnent les processus de transcription et de traduction. Les différences principales entre les bactéries et les eucaryotes concernant l'expression génétique sont dues à l'absence de compartiments dans les cellules bactériennes. Ainsi, dans la cellule bactérienne, l'activité se déroule de façon ininterrompue par le couplage de la transcription et de la traduction. Comme il n'y a pas de noyau, un même gène peut être le siège simultané de ces deux processus (**figure 17.24**), et la protéine qui vient d'être synthétisée peut atteindre rapidement son site fonctionnel par diffusion.

Par contre, dans la cellule eucaryote, l'enveloppe du noyau délimite deux compartiments, à l'intérieur desquels se déroulent respectivement la transcription et la traduction. L'un de ces compartiments est également le siège d'une maturation très élaborée de l'ARN. Le processus de maturation inclut des étapes

supplémentaires, dont la régulation permet de coordonner les activités complexes de la cellule eucaryote. La **figure 17.25** résume l'ensemble de ce processus dans une cellule eucaryote, du gène au polypeptide.

RETOUR SUR LE CONCEPT

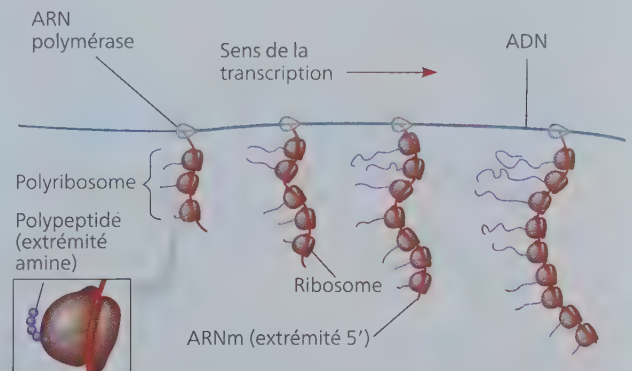
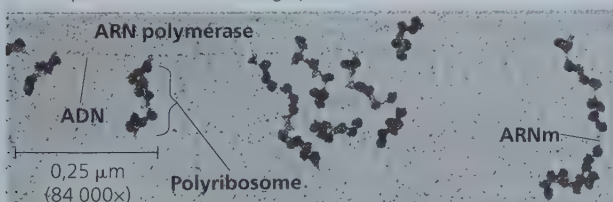
17.4

1. Quels sont les deux processus qui garantissent que l'acide aminé approprié est ajouté à une chaîne polypeptidique en cours de synthèse ?
2. Expliquez comment un polypeptide destiné à être sécrété à l'extérieur de la cellule atteint le réseau intracellulaire de membranes.
3. **FAITES UN DESSIN** ► Dessinez un ARNt avec l'anticodon 3'-CGU-5'. Quels sont les deux codons différents auxquels il peut se lier ? Dessinez chacun d'eux sur un ARNm, en identifiant toutes les extrémités 5' et 3'. Ajoutez l'acide aminé transporté par l'ARNt.
4. **ET SI ?** ► Dans les cellules eucaryotes, on a observé que les ARNm ont un arrangement circulaire dans lequel des protéines retiennent une queue poly-A près d'une coiffe 5'. Comment cet arrangement peut-il accroître l'efficacité de la traduction ?

Voir les réponses proposées à l'appendice A.

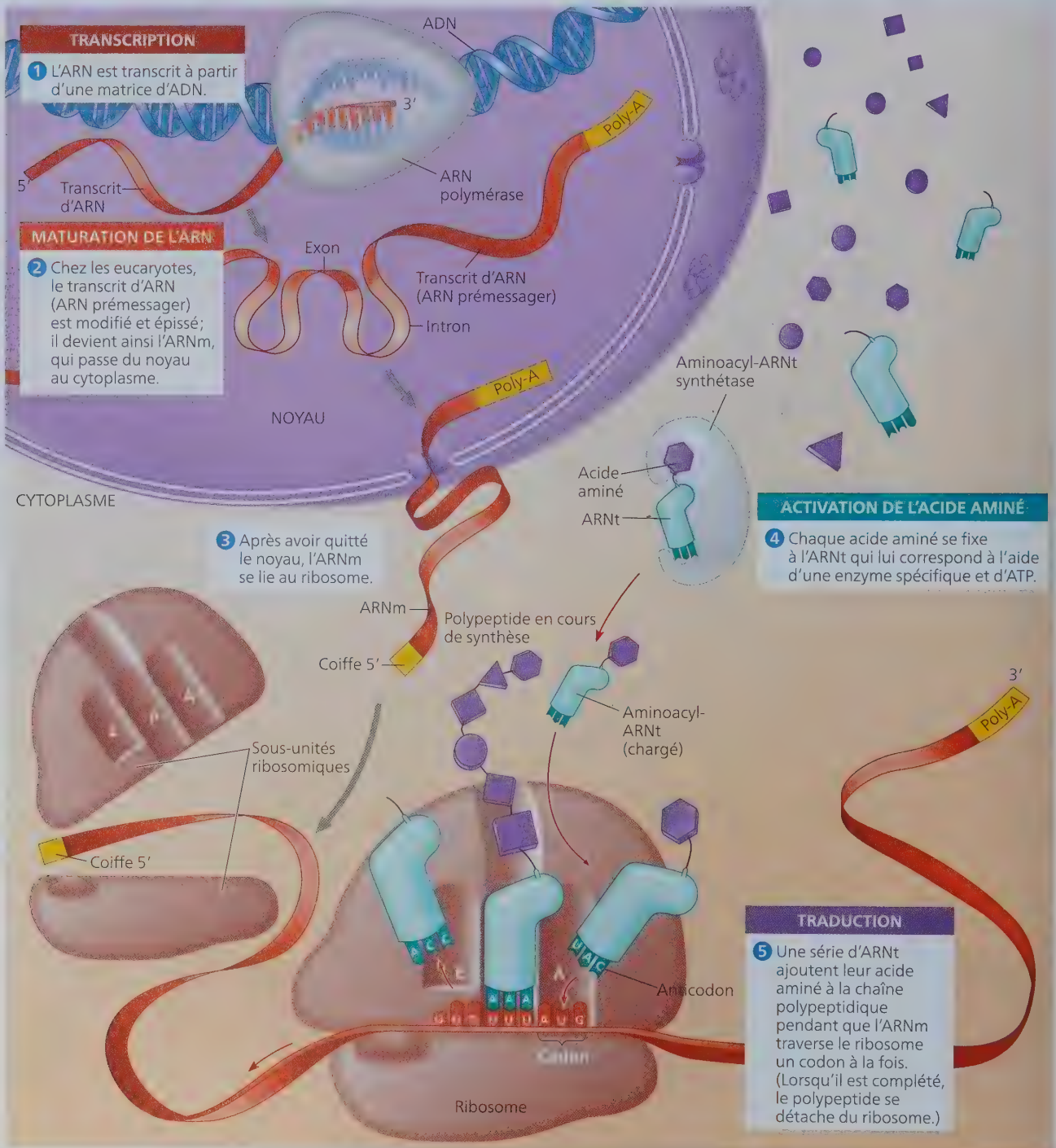
▼ **Figure 17.24** Le couplage de la transcription et de la traduction chez les bactéries.

Dans les cellules bactériennes, la traduction de l'ARNm peut commencer dès que la première extrémité (5') de la molécule d'ARNm se détache de la matrice d'ADN. La micrographie (MET) montre la transcription d'un brin d'ADN d'*E. coli* par des molécules d'ARN polymérase. Chacune d'elles engendre un brin d'ARNm déjà en cours de traduction par les ribosomes. Les polypeptides nouvellement synthétisés ne sont pas visibles dans la micrographie, mais sont illustrés dans le schéma.



HABILITÉS VISUELLES ► Laquelle des molécules d'ARNm a commencé la transcription en premier (à gauche ou à droite) ? Sur cet ARNm, quel ribosome a commencé la traduction en premier (en haut ou en bas) ?

▼ **Figure 17.25 Résumé de la transcription et de la traduction dans une cellule eucaryote.** Ce schéma illustre le processus de synthèse d'un polypeptide à partir du gène qui détient le message génétique correspondant. Chaque gène peut être transcrit à maintes reprises en molécules d'ARNm identiques et chaque ARNm peut être traduit en polypeptides identiques de nombreuses fois. (Souvenez-vous également que le produit final de certains gènes n'est pas un polypeptide, mais une molécule d'ARN qui n'est pas traduite et qui peut être un ARNt ou un ARNr.) De façon générale, les étapes de la transcription et de la traduction sont semblables dans les cellules bactériennes, archéennes et eucaryotes. La différence principale est l'étape de la maturation de l'ARNm qui se déroule dans le noyau de la cellule eucaryote. Les autres différences importantes concernent les étapes de l'initiation de la transcription et de la traduction ainsi que la terminaison de la transcription. Pour remettre ces processus dans leur contexte cellulaire, voir la figure 6.32.



Les mutations d'un ou de quelques nucléotides peuvent modifier la structure et la fonction des protéines

Après avoir étudié le processus de l'expression génétique, vous êtes maintenant prêt à comprendre les effets des modifications sur l'information génétique d'une cellule (ou d'un virus). Ces modifications, appelées **mutations**, sont à l'origine de l'immense diversité des gènes présents parmi les organismes parce que les mutations sont la source première de nouveaux gènes. Précédemment, nous avons parlé des remaniements chromosomiques touchant de longs segments d'ADN (voir la figure 15.14) qu'on peut considérer comme des mutations à grande échelle. Maintenant, nous abordons les mutations à petite échelle d'une ou de quelques paires de nucléotides, notamment les **mutations ponctuelles**. Il s'agit de modifications chimiques touchant une seule paire de bases nucléotidiques d'un gène.

Si elle apparaît dans un gamète ou dans une cellule productrice de gamètes, une mutation ponctuelle peut être transmise à la descendance immédiate et aux générations suivantes. Quand elle a des effets nocifs sur le phénotype d'un organisme, on parle d'anomalie génétique ou de maladie héréditaire. Par exemple, on a trouvé la cause génétique de la drépanocytose (aussi nommée anémie à hématies falciformes, ou plus simplement anémie falciforme) : cette maladie est due à la mutation touchant une seule paire de nucléotides du gène qui code pour la β -globine, l'un des polypeptides de l'hémoglobine. Cette modification d'un seul nucléotide de la matrice d'ADN entraîne l'altération d'un codon de l'ARNm et la production d'une protéine anormale (figure 17.26 ; voir aussi la figure 5.19). L'hémoglobine ainsi transformée des personnes homozygotes pour le gène mutant donne aux érythrocytes une forme de faucille qui réduit substantiellement la fixation de l' O_2 . Cela fait apparaître les nombreux symptômes de l'anémie falciforme (voir les figures 14.17 et 23.18). La myocardiopathie familiale, un trouble cardiaque, est une autre anomalie causée par une mutation ponctuelle ; cette maladie est responsable de certains cas tragiques de mort subite chez de jeunes athlètes. On a découvert

plusieurs mutations ponctuelles de gènes encodant des protéines musculaires et susceptibles de provoquer cette affection.

Les catégories de mutations à petite échelle

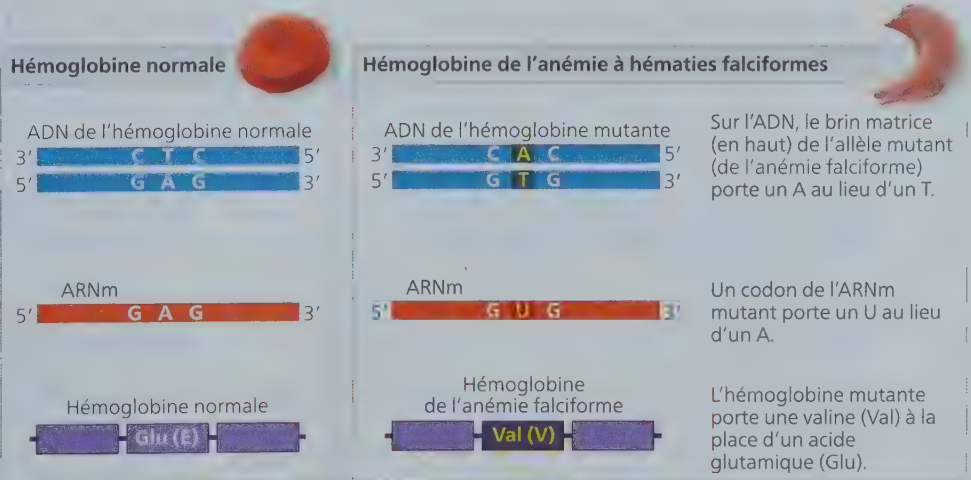
Voyons maintenant de quelle façon les mutations à petite échelle modifient les protéines. Notons d'abord que plusieurs mutations surviennent dans des régions extérieures aux gènes codant pour des protéines et que les effets potentiels de ces mutations sur les phénotypes peuvent être subtils et difficiles à déceler. C'est pourquoi nous mettrons ici l'accent sur les mutations dans les gènes codant pour une protéine. On peut classer les mutations ponctuelles survenant à l'intérieur d'un gène en deux grandes catégories : (1) les substitutions d'une seule paire de nucléotides ; (2) les insertions ou les délétions de paires de nucléotides. Les insertions et les délétions peuvent faire intervenir une ou plusieurs paires de nucléotides.

Les substitutions

La **substitution d'une paire de bases** (ou de nucléotides) est le remplacement de la base d'un nucléotide et de son vis-à-vis par une paire de bases différentes (figure 17.27a). Certaines substitutions n'ont aucun effet sur la protéine synthétisée à cause de la redondance du code génétique. Par exemple, si 3'-CCG-5' sur le brin matrice devient 3'-CCA-5' à la suite d'une mutation, le codon d'ARNm GGC devient GGU ; or, ces deux derniers commandent l'ajout d'une glycine à l'endroit correspondant de la protéine (voir la figure 17.6). Autrement dit, la modification d'une paire de nucléotides peut résulter en un codon dont la traduction donne le même acide aminé que celui pour lequel le codon initial aurait codé. Une telle modification est un exemple de **mutation silencieuse**, qui n'a aucun effet observable sur le phénotype. (Les mutations silencieuses peuvent se produire également à l'extérieur du gène.) Fait à noter, on a démontré que certaines mutations silencieuses peuvent avoir une incidence indirecte sur le lieu ou le degré d'expression d'un gène, même si la protéine demeure la même.

Les substitutions qui aboutissent au remplacement d'un acide aminé par un autre sont appelées **mutations faux-sens**. Il arrive qu'elles n'aient pas d'effets notables sur la protéine synthétisée, parce que le nouvel acide aminé a des propriétés

► **Figure 17.26** L'origine moléculaire de l'anémie falciforme, une mutation ponctuelle. La différence entre l'allèle qui entraîne l'anémie falciforme et l'allèle normal consiste en une modification d'une seule paire de nucléotides de l'ADN. Les deux micrographies représentent des érythrocytes observés en microscopie électronique. Celle de gauche montre un érythrocyte normal, et celle de droite, un érythrocyte falciforme ; chacun d'eux provient d'un homozygote pour les allèles de type sauvage et mutants, respectivement.



semblables à celles de l'ancien ou encore parce que la séquence exacte des acides aminés d'une certaine région de la protéine n'est pas essentielle à sa fonction.

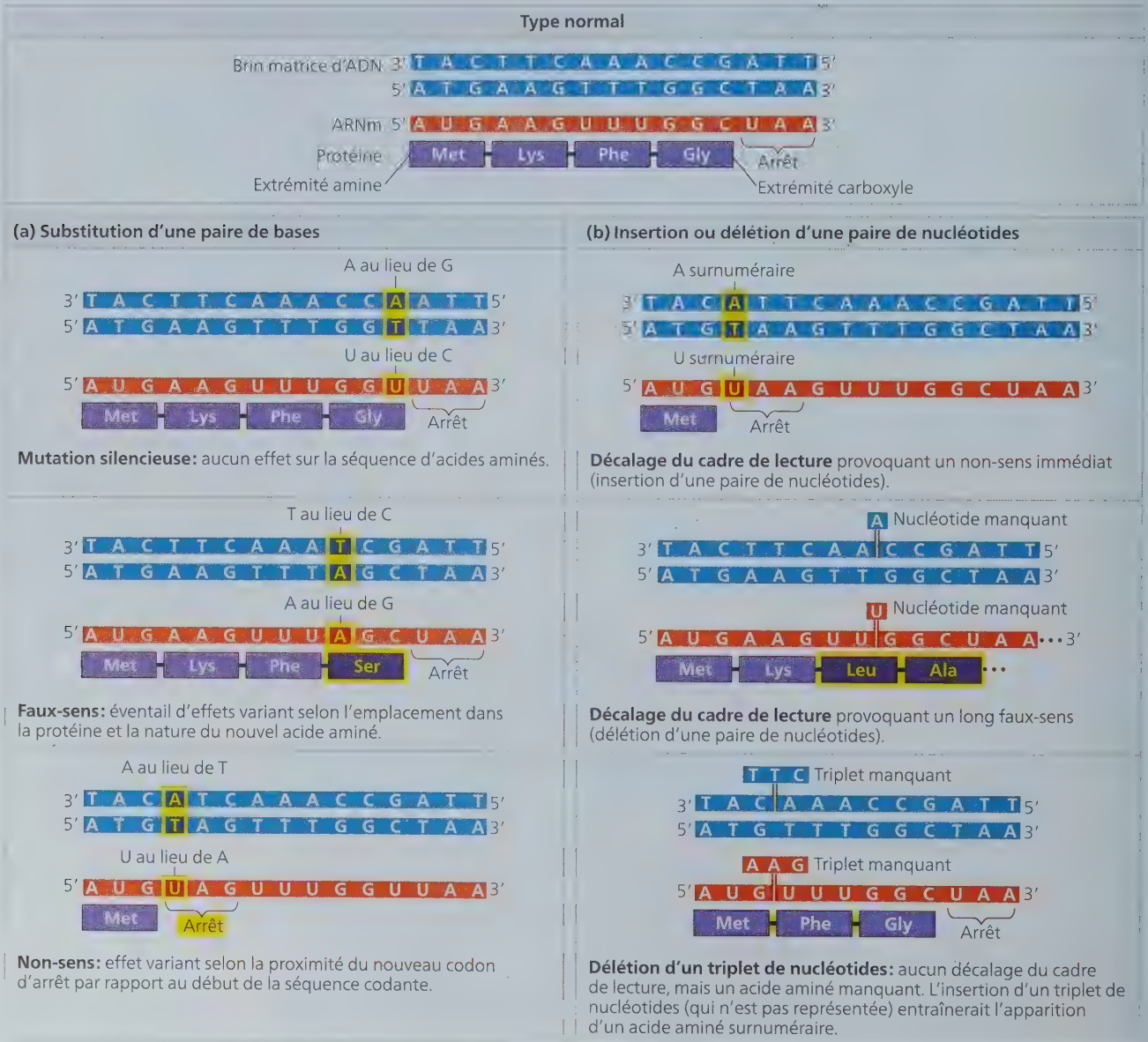
Bien plus intéressantes pour les généticiens sont les substitutions de paires de bases qui occasionnent un changement important dans la composition de la protéine. L'altération d'un seul acide aminé dans une région essentielle de la protéine peut avoir des répercussions importantes sur son activité (voir par exemple la partie de la sous-unité β -globine de l'hémoglobine illustrée à la figure 17.26 ou l'effet sur le site actif d'une enzyme tel qu'il est présenté à la figure 8.18). De temps à autre, une telle mutation crée une protéine améliorée ou ayant de nouvelles propriétés.

Cependant, la plupart du temps, les mutations exercent un effet neutre ou néfaste parce qu'elles engendrent une protéine inutile ou moins active qui entrave le fonctionnement de la cellule.

Les substitutions provoquent le plus souvent des mutations faux-sens; les codons touchés codent encore pour des acides aminés et ont donc un sens, mais celui-ci est *erroné*. Toutefois, une mutation ponctuelle peut transformer un codon correspondant à un acide aminé en un codon d'arrêt; ces altérations sont appelées **mutations non-sens**, et elles conduisent à la fin prématurée de la traduction; le polypeptide synthétisé est plus court que celui qui est encodé par le gène normal. La plupart des mutations non-sens conduisent à la synthèse de protéines non fonctionnelles.

▼ **Figure 17.27** Les catégories de mutations à petite échelle qui modifient la séquence de l'ARNm.

Toutes les catégories illustrées ici, à l'exception d'une seule, modifient la séquence des acides aminés du polypeptide encodé.



Dans l'exercice de la rubrique **Résolution de problème**, vous travaillerez avec quelques substitutions courantes d'une seule paire de bases dans le gène encodant l'insuline. Quelques-unes de ces substitutions, sinon la totalité, peuvent causer le diabète. Vous devrez déterminer à quelle catégorie de mutations appartiennent ces substitutions et décrire le changement dans la séquence des acides aminés.

Les insertions et les délétions

Les **insertions** et les **délétions** correspondent à l'ajout ou à la perte d'une ou de plusieurs paires de nucléotides dans un gène (**figure 17.27b**). Elles ont généralement des conséquences plus désastreuses que les substitutions. En effet, l'insertion ou la délétion de nucléotides peut modifier le cadre de lecture du message

RÉSOLUTION DE PROBLÈME

Les mutations du gène codant pour l'insuline sont-elles responsables du diabète néonatal de trois nourrissons ?

L'insuline est une hormone qui agit comme régulateur principal de la glycémie. Dans certains cas de diabète néonatal, il s'est produit une substitution d'une paire de bases dans le gène codant pour l'insuline, ce qui entraîne une modification suffisamment importante de la structure de la protéine pour la rendre non fonctionnelle. Comment pouvez-vous déterminer quelle paire de bases a été touchée par la substitution et quels en sont les effets sur la séquence d'acides aminés ?

Maintenant qu'il est possible de séquencer le génome entier d'une personne, les médecins peuvent utiliser l'information fournie par des séquences d'ADN pour diagnostiquer certaines maladies et pour mettre au point de nouveaux traitements. Par exemple, chez un enfant souffrant de diabète néonatal, il est possible d'analyser la séquence du gène codant pour l'insuline afin de déterminer si ce gène présente une mutation et, le cas échéant, établir quel est l'effet de cette mutation.



Dans cet exercice, vous devez déterminer quels sont les effets des mutations présentes dans une partie des séquences du gène codant pour l'insuline chez les patients diabétiques.

Votre méthode

Supposons que vous êtes un généticien médical et que vous avez affaire à trois nourrissons qui présentent une substitution d'une paire de bases dans le gène codant pour l'insuline. Vous devez analyser la mutation qui affecte chacun des trois bébés afin de déterminer quel effet elle produit sur la séquence d'acides aminés de l'insuline. Pour identifier la mutation dans chaque patient, vous devrez comparer sa séquence d'ADN complémentaire (ADNc) de l'insuline à celle de l'ADNc de type sauvage. L'ADNc est une molécule d'ADN à double brin correspondant à une séquence d'ARNm qui, par conséquent, ne contient que la partie traduite d'un gène – les introns ne sont pas inclus. (On utilise souvent les séquences d'ADNc pour comparer les régions codantes des gènes.) En repérant les codons modifiés, vous pourrez déterminer quels acides aminés sont altérés dans la protéine d'insuline.

Vos données

Comme vous analyserez les codons d'ADNc pour les acides aminés 35 à 54 (parmi les 110 acides aminés) de chaque protéine d'insuline, le codon de départ (AUG) est absent. Les séquences d'ADNc de type sauvage et d'ADNc des patients sont présentées ci-dessous sous forme de codons.

ADNc de type sauvage 5'-CTG GTG GAA GCT CTC TAC CTA GTG TGC GGG GAA CGA GGC TTC TTC TAC ACA CCC AAG ACC-3'

ADNc du patient 1 5'-CTG GTG GAA GCT CTC TAC CTA GTG TGC GGG GAA CGA GGC TGC TTC TAC ACA CCC AAG ACC-3'

ADNc du patient 2 5'-CTG GTG GAA GCT CTC TAC CTA GTG TGC GGG GAA CGA GGC TCC TTC TAC ACA CCC AAG ACC-3'

ADNc du patient 3 5'-CTG GTG GAA GCT CTC TAC CTA GTG TGC GGG GAA CGA GGC TTC TCG TAC ACA CCC AAG ACC-3'

Source des données : N. Nishi et K. Nanjo, Insulin gene mutations and diabetes, *Journal of Diabetes Investigation* 2: 92-100 (2011).

Votre analyse

1. Comparez la séquence d'ADNc de chaque patient à la séquence d'ADN de type sauvage, puis encercler les codons dans lesquels une paire de nucléotides a été substituée.
2. Utilisez un tableau des codons (voir la figure 17.6) pour identifier l'acide aminé auquel correspond le codon présentant la mutation dans chaque séquence d'ADN de l'insuline, et comparez-le à l'acide aminé commandé par le codon de la séquence de type sauvage correspondante. Conformément à la pratique standard de séquençage de l'ADN, le brin d'ADNc *codant* (d'ADN non transcrit) est présenté afin que vous puissiez le convertir en ARNm et l'utiliser avec le tableau des codons. Vous n'avez qu'à remplacer le T par un U. Pour chaque patient, classez la substitution : s'agit-il d'une mutation silencieuse, d'un faux-sens ou d'un non-sens ? Expliquez chaque réponse.
3. Comparez la structure de l'acide aminé identifié dans la séquence d'ADN de l'insuline de chaque patient à celle de l'acide aminé de la séquence d'ADN de type sauvage correspondante (voir la figure 5.14). Étant donné que les trois patients souffrent de diabète néonatal, expliquez comment la modification d'un acide aminé peut altérer la protéine d'insuline et, par conséquent, causer la maladie.

génétique formé par le regroupement des triplets de nucléotides sur l'ARNm qui est lu pendant la traduction. Ce type de mutation, appelé **décalage du cadre de lecture**, apparaît chaque fois que le nombre de nucléotides insérés ou enlevés n'est pas un multiple de trois. Tous les nucléotides situés en aval de la modification sont alors regroupés en des codons erronés. Il en résulte un long faux-sens qui aboutit tôt ou tard à un non-sens et à une terminaison prématurée ou retardée. À moins que le décalage du cadre de lecture survienne très près de la fin du gène, il est presque certain que la protéine ne sera pas fonctionnelle. Les insertions et les délétions surviennent aussi à l'extérieur des régions codantes; le cas échéant, il ne peut y avoir décalage du cadre de lecture. Toutefois, ce type de modification peut agir sur le phénotype, par exemple en modifiant la façon dont un gène est exprimé.

Les nouvelles mutations et les mutagènes

Les mutations peuvent avoir des causes diverses. Les erreurs survenues lors de la réplication ou de la recombinaison de l'ADN peuvent engendrer des substitutions de paires de bases, des insertions, des délétions ou des mutations touchant des parties plus longues de l'ADN. Si un nucléotide incorrect est ajouté à la chaîne en cours de formation pendant la réplication, par exemple, la base sur ce nucléotide sera désappariée avec la base du nucléotide de l'autre brin. Dans de nombreux cas, l'erreur sera corrigée par la correction d'épreuves de l'ADN et par les systèmes de réparation (voir le concept 16.2). Sinon, la base incorrecte sera utilisée comme matrice dans la réplication suivante, causant une mutation. Les mutations résultant de ce genre d'erreurs sont appelées *mutations spontanées*, et il est difficile d'en calculer le taux d'occurrence. Des estimations approximatives effectuées sur le taux de mutation au cours de la réplication de l'ADN chez *E. coli* et chez les eucaryotes ont donné des valeurs similaires: environ 1 nucléotide sur 10^{10} est modifié, et le changement est transmis à la génération suivante de cellules.

Certains agents physiques ou chimiques appelés **mutagènes** interagissent avec l'ADN et provoquent des changements. Dans les années 1920, Hermann Muller, un étudiant de T. H. Morgan et Prix Nobel de médecine en 1946, a découvert que les rayons X causaient des modifications génétiques chez les drosophiles. En exposant des mouches à ces rayonnements, il a obtenu des drosophiles mutantes, qu'il a ensuite utilisées au cours de ses recherches. Mais il s'est rendu compte des conséquences inquiétantes de sa découverte: les rayons X et les autres formes de radiations à haute énergie sont dangereux, tant pour le génome humain que pour celui des organismes de laboratoire. Le rayonnement ultraviolet fait partie des mutagènes physiques; il contribue à la formation de dimères de thymine dans l'un ou l'autre des brins d'ADN (voir la figure 16.19).

Il existe plusieurs catégories de mutagènes chimiques. Les analogues des nucléotides (comme la 5-bromo-uracile, un analogue de la thymine) sont des substances qui ressemblent aux nucléotides normaux de l'ADN et qui s'insèrent dans la molécule pendant sa réplication. Ils modifient ponctuellement l'information génétique, car ils sont plus susceptibles d'entraîner de mauvais appariements que les nucléotides normaux. D'autres mutagènes chimiques entravent la réplication en s'insérant dans l'ADN et en déformant la double hélice. Enfin, certains mutagènes modifient chimiquement les bases en altérant aussi leur capacité d'appariement.

Des chercheurs ont mis au point plusieurs méthodes pour tester *in vitro* l'activité mutagène de diverses substances chimiques. Le principal domaine d'application de ces tests est le dépistage préliminaire des substances chimiques susceptibles de causer le cancer. Cette approche est valable parce que la plupart des agents cancérigènes (qui provoquent le cancer) sont des mutagènes, et, inversement, la plupart des mutagènes sont cancérigènes.

Qu'est-ce qu'un gène? Reconsidérons la question

Au cours des derniers chapitres, notre définition du gène a progressé. Nous avons commencé par le concept mendélien, selon lequel le gène est une unité héréditaire discontinue définissant un caractère phénotypique (chapitre 14). Puis, nous avons vu que Morgan et ses collaborateurs ont associé les gènes à des locus spécifiques situés sur les chromosomes (chapitre 15). Ensuite, nous avons montré qu'un gène est une région d'une molécule d'ADN d'un chromosome portant une séquence nucléotidique précise (chapitre 16). Enfin, dans le présent chapitre, nous avons examiné une définition fonctionnelle du gène: il s'agit d'une séquence d'ADN qui code pour une chaîne polypeptidique spécifique ou une molécule d'ARN fonctionnelle (ARNt ou ARNr). Toutes ces définitions peuvent être utiles selon le contexte dans lequel on étudie les gènes.

Comme nous l'avons mentionné, il est réducteur de dire qu'un gène code pour un polypeptide. Ainsi, chez les eucaryotes, la plupart des gènes comportent des segments non codants (comme les introns), de sorte qu'une grande partie de la chaîne d'ADN ne correspond à aucun segment au niveau des polypeptides. Les spécialistes de la biologie moléculaire considèrent souvent que les promoteurs et certaines régions régulatrices de l'ADN font partie du gène même s'ils ne sont pas transcrits. Ces séquences d'ADN ne sont pas transcrites, mais il est possible de considérer qu'elles font partie du gène fonctionnel puisqu'elles sont nécessaires à la transcription. Notre définition du gène doit aussi être assez large pour englober les segments d'ADN qui sont transcrits en ARNr, en ARNt et en d'autres types d'ARN qui ne sont pas traduits. Ces gènes ne produisent aucun polypeptide, mais jouent des rôles essentiels dans la cellule. On en arrive à la définition suivante: *un gène est une région de l'ADN qui peut être exprimée pour produire un produit final fonctionnel, soit un polypeptide, soit une molécule d'ARN.*

Cependant, quand on considère les phénotypes, il est souvent utile de commencer par s'intéresser aux gènes qui codent pour des polypeptides. Dans le présent chapitre, nous avons appris comment un gène ordinaire est exprimé au niveau moléculaire, à savoir par transcription en ARNm, puis par traduction en un polypeptide qui forme une protéine dotée d'une structure et d'une fonction spécifiques. Les protéines, pour leur part, expriment le phénotype observable de l'organisme.

En tout temps, un type cellulaire donné n'exprime qu'un petit sous-ensemble de ses gènes. C'est là une caractéristique essentielle chez les organismes multicellulaires: vous auriez des ennuis si les cellules du cristallin de vos yeux se mettaient à exprimer les gènes des protéines des cheveux, qui sont normalement exprimés seulement dans les cellules des follicules pileux! L'expression génétique est donc soumise à une régulation précise, que nous aborderons dans le prochain chapitre, en commençant par celle des bactéries, qui est relativement simple, pour continuer ensuite avec les eucaryotes.

1. Que se passe-t-il lorsqu'une paire de nucléotides est enlevée au milieu de la séquence codante d'un gène ?
2. **FAITES DES LIENS** ▶ Les individus hétérozygotes pour l'allèle de l'anémie falciforme sont généralement en bonne santé, mais ils présentent des effets phénotypiques de l'allèle dans certaines circonstances (voir la figure 14.17). Expliquez cette observation en termes d'expression génétique.

3. **FAITES UN DESSIN** ▶ Le brin matrice d'un gène contient la séquence nucléotidique suivante : 3'-TACTTGTCCGATATC-5'. À la suite d'une mutation, il est modifié en 3'-TACTTGTCCAATATC-5'. Dessinez le double brin de l'ADN pour les deux séquences (type sauvage et mutante) ainsi que la séquence d'acides aminés encodée par chacune. Quel est l'effet de cette mutation sur la séquence des acides aminés ?

Voir les réponses proposées à l'appendice A.

RÉVISION DU CHAPITRE 17



Consultez votre MANUEL NUMÉRIQUE, qui vous donne accès aux **animations**, aux **exercices** et à la plateforme d'**anatomie interactive**.

Résumé des concepts clés

CONCEPT 17.1

Les gènes codent pour les protéines par l'intermédiaire de la transcription et de la traduction (p. 370 à 377)

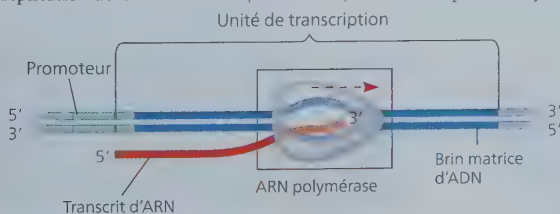
- Les études de Beadle et Tatum de souches mutantes de *Neurospora* ont abouti à l'hypothèse dite « un gène, un polypeptide ». Au cours de l'**expression génétique**, l'information encodée dans les gènes est utilisée pour produire des chaînes polypeptidiques spécifiques (des enzymes et d'autres protéines) ou des molécules d'ARN.
- La **transcription** est la synthèse d'ARN complémentaire à un **brin matrice** de l'ADN permettant le passage de l'information de désoxyribonucléotides à des ribonucléotides. La **traduction** est la synthèse d'un polypeptide dont la séquence d'acides aminés est codée par la séquence de nucléotides dans l'**ARN messager (ARNm)**.
- L'information génétique dans l'ADN est encodée sous forme de séquence de triplets de nucléotides qui ne se chevauchent pas, les **codons** (ou **génons**). Un codon est un triplet de nucléotides qui, dans l'ARNm, peut coder pour un acide aminé (61 des 64 codons codent pour les acides aminés) ou servir de signal d'arrêt (3 codons). Les codons doivent être lus dans le bon **cadre de lecture** (dans le bon sens).

? Décrivez le processus de l'expression génétique par lequel un gène peut déterminer le phénotype d'un organisme.

CONCEPT 17.2

La transcription est la synthèse de l'ARN à partir de l'ADN : une étude détaillée (p. 377 à 379)

- La synthèse de l'ARN est catalysée par l'**ARN polymérase**, qui assemble les nucléotides complémentaires de l'ARN au brin matrice de l'ADN. Ce processus obéit aux mêmes règles d'appariement des bases que la réplication de l'ADN. Toutefois, dans l'ARN, l'uracile remplace la thymine.



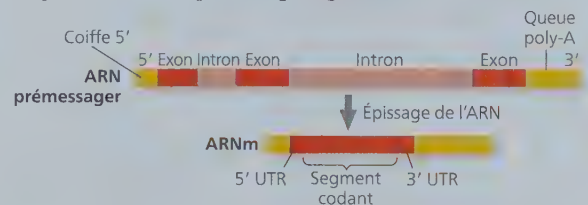
- Les trois étapes de la transcription sont l'initiation, l'élongation et la terminaison. Des **promoteurs**, comportant souvent une **boîte TATA** chez les eucaryotes, déterminent l'endroit de l'initiation de la synthèse de l'ARN. Chez les eucaryotes, les **facteurs de transcription** aident l'ARN polymérase à reconnaître les séquences du promoteur, en formant un **complexe d'initiation de la transcription**. La terminaison se déroule différemment chez les bactéries et les eucaryotes.

? Quelles sont les similitudes et les différences dans l'initiation de la transcription des gènes chez les bactéries et les eucaryotes ?

CONCEPT 17.3

Dans les cellules eucaryotes, l'ARN est modifié après avoir été transcrit (p. 379 à 382)

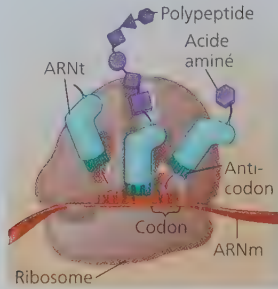
- Chez les eucaryotes, les molécules d'ARN pré-messager subissent une **maturation de l'ARN**. Cette maturation comprend l'épissage de l'ARN, l'ajout à l'extrémité 5' d'une **coiffe 5'** et l'ajout à l'extrémité 3' d'une **queue poly-A**. L'ARNm mature compte une région non traduite (5' UTR ou 3' UTR) à chaque extrémité du segment codant.
- La plupart des gènes d'eucaryotes sont séparés en segments : ils contiennent des **introns** intercalés entre les **exons** (régions contenues dans l'ARNm). Pendant l'**épissage de l'ARN**, les introns sont enlevés et les exons sont réunis. L'épissage de l'ARN est catalysé par le **complexe d'épissage**. Dans certains cas, l'ARN catalyse seul son propre épissage. La capacité catalytique de certaines molécules d'ARN, appelées **ribozymes**, provient de propriétés inhérentes à l'ARN. La présence d'introns permet l'**épissage différentiel de l'ARN**.



? Quelle est la fonction de la coiffe 5' et de la queue poly-A dans l'ARNm des eucaryotes ?

La traduction est la synthèse d'un polypeptide à partir de l'ARN messenger: une étude détaillée (p. 382 à 392)

- Une cellule traduit le message de l'ARNm en polypeptides avec l'aide de l'**ARN de transfert (ARNt)**. Après avoir été fixée à l'acide aminé qui lui correspond par une **aminoacyl-ARNt synthétase**, chaque molécule d'ARN de transfert s'aligne sur le codon complémentaire de l'ARNm par l'intermédiaire de son **anticodon**. Un **ribosome**, constitué d'**ARN ribosomique (ARNr)** et de protéines, facilite cet appariement grâce à son site de liaison pour l'ARNm et à son site de liaison pour l'ARNt.
- Les ribosomes coordonnent les trois étapes de la traduction, qui sont l'initiation, l'élongation et la terminaison. L'ARNr catalyse la formation des liaisons polypeptidiques entre les acides aminés pendant que les ARNt se déplacent à travers les **sites A et P** et sortent au **site E**.



- Après la traduction, les protéines peuvent être modifiées au cours de leur maturation soit par épissage, soit par la fixation de glucides, de lipides, de groupements phosphate ou d'autres groupements chimiques.
- Les ribosomes libres dans le cytosol amorcent la synthèse de toutes les protéines; cependant, celles qui sont dotées d'une **séquence signal** sont synthétisées sur le RE.
- Un gène peut être transcrit par plusieurs ARN polymérases en même temps. De plus, plusieurs ribosomes peuvent traduire une même molécule d'ARNm simultanément et ainsi former un **polyribosome**. Chez les bactéries, ces processus sont couplés. Par contre, chez les eucaryotes, la membrane nucléaire sépare ces processus dans l'espace et le temps.

❓ Décrivez l'action de l'ARNt pendant la synthèse d'un polypeptide du point de vue du ribosome.

CONCEPT 17.5

Les mutations d'un ou de quelques nucléotides peuvent modifier la structure et la fonction des protéines (p. 393 à 397)

- Les **mutations** à petite échelle comprennent les **mutations ponctuelles** consistant en des modifications d'une paire de nucléotides de l'ADN, ce qui peut entraîner la production d'une protéine non fonctionnelle. Les **substitutions de paires de bases** peuvent provoquer une **mutation faux-sens** ou une **mutation non-sens**. L'**insertion** et le **délétion** de paires de nucléotides peuvent provoquer le **décalage du cadre de lecture**.
- Des mutations spontanées peuvent apparaître pendant la réplication, la recombinaison ou la réparation de l'ADN. Des **mutagènes** chimiques ou physiques causent des dommages à l'ADN qui peuvent modifier les gènes.

❓ Quels seraient les résultats de la modification chimique d'une base nucléotidique d'un gène? Quel rôle jouent les systèmes de réparation de l'ADN dans la cellule?

Évaluation

NIVEAU 1: CONNAISSANCES ET COMPRÉHENSION

1. Dans les cellules eucaryotes, la transcription ne peut commencer tant que:
 - a) les deux brins d'ADN ne se sont pas complètement séparés pour exposer le promoteur.
 - b) plusieurs facteurs de transcription ne sont pas liés au promoteur.
 - c) la coiffe 5' n'a pas été enlevée de l'ARNm.
 - d) les introns d'ADN n'ont pas été enlevés de la matrice.
2. Parmi les affirmations suivantes concernant le codon, laquelle est *fausse*?
 - a) Il peut coder pour le même acide aminé qu'un autre codon.
 - b) Il ne code jamais pour plus d'un acide aminé.
 - c) Il s'allonge à partir de l'une des extrémités de la molécule d'ARNt.
 - d) C'est l'unité fondamentale du code génétique.
3. L'anticodon d'une molécule d'ARNt:
 - a) et le codon correspondant sur l'ARNm sont complémentaires.
 - b) et le triplet correspondant sur l'ARNr sont complémentaires.
 - c) est la partie de l'ARNt qui se lie à un acide aminé spécifique.
 - d) est un catalyseur, ce qui fait de l'ARNt un ribozyme.
4. Parmi les affirmations suivantes concernant la maturation de l'ARN, laquelle est *fausse*?
 - a) Les exons sont coupés et dégradés avant que l'ARNm ne quitte le noyau.
 - b) Des nucléotides peuvent être ajoutés aux deux extrémités de l'ARN.
 - c) Les ribozymes peuvent jouer un rôle dans l'épissage de l'ARN.
 - d) L'épissage de l'ARN peut être catalysé par les complexes d'épissage.
5. Quel est le composant qui *n'intervient pas directement* dans le mécanisme appelé traduction?
 - a) La GTP.
 - b) L'ADN.
 - c) L'ARNt.
 - d) Les ribosomes.

NIVEAU 2: APPLICATION ET ANALYSE

6. À l'aide de la figure 17.6, désignez une séquence possible de nucléotides (que vous lirez dans le sens 5' → 3') de la matrice d'ADN qui produit un ARNm codant pour la séquence de polypeptides Phe-Pro-Lys.
 - a) 5'-UUUCCCAAA-3'
 - b) 5'-GAACCCCTT-3'
 - c) 5'-CTTCGGGAA-3'
 - d) 5'-AAACCCUUU-3'
7. Parmi les mutations suivantes, laquelle risque *le plus* d'avoir un effet nocif sur l'organisme touché?
 - a) La délétion de trois nucléotides près du milieu d'un gène.
 - b) La délétion d'un seul nucléotide au milieu d'un intron.
 - c) La délétion d'un seul nucléotide près de la fin de la séquence codante.
 - d) L'insertion d'un seul nucléotide en aval et près du début d'une séquence codante.
8. Dans une cellule eucaryote, pourrait-on observer le couplage des processus, comme ils sont illustrés à la figure 17.24? Expliquez votre réponse.
9. **FAITES UN DESSIN** ▶ Remplissez le tableau suivant:

Type d'ARN	Fonctions
ARN messenger (ARNm)	
ARN de transfert (ARNt)	
Transcrit primaire (ARN pré-messager)	Dans un ribosome, joue un rôle structural et surtout, en tant que ribozyme, joue un rôle catalytique.
Petit ARN dans le complexe d'épissage	

Voir les réponses proposées à l'appendice A.