

Expression du patrimoine génétique

COURS • S04 • PREMIÈRE

Spécialité de première, programme officiel. Cours original rédigé avec une IA. Le cours de ton enseignant fait foi.

Sommaire du cours

Ce que tu dois savoir faire	1
Cours	2
Exemples résolus	8
Méthodes et pièges	10
Vers le PASS	10
Exercices	10
Corrigés	12

Ce que tu dois savoir faire

- Montrer, à partir de résultats expérimentaux, que l'ARN messager est l'intermédiaire entre l'ADN du noyau et la synthèse des protéines dans le cytoplasme.
- Comparer la structure de l'ADN et celle de l'ARN.
- Décrire la transcription : rôle de l'ARN polymérase, brin transcrit, complémentarité des bases, sens de synthèse.
- Expliquer la maturation de l'ARN pré-messager chez les eucaryotes et le principe de l'épissage alternatif.
- Utiliser le tableau du code génétique et en énoncer les propriétés (universalité, redondance, codons d'initiation et d'arrêt).
- Décrire la traduction : rôle du ribosome, de l'ARN de transfert, de l'initiation, de l'élongation et de la terminaison.
- Passer d'une séquence d'ADN à une séquence d'acides aminés, et inversement discuter les séquences possibles.
- Relier la séquence d'une protéine à sa structure et à sa fonction.
- Expliquer la relation entre génotype et phénotype aux échelles moléculaire, cellulaire et de l'organisme, en intégrant l'influence de l'environnement.

Cours

1. Du gène à la protéine : la nécessité d'un intermédiaire

Un **gène** est une séquence d'ADN qui contient l'information nécessaire à la synthèse d'une molécule, le plus souvent une **protéine**. Les protéines sont des chaînes d'**acides aminés** (il en existe 20 sortes dans les protéines du vivant) dont l'ordre, la **séquence**, est déterminé par la séquence des nucléotides du gène. Le génome humain compte environ 20 000 gènes codant des protéines.

Chez les eucaryotes, l'ADN reste dans le noyau, alors que la synthèse des protéines a lieu dans le cytoplasme, au niveau des **ribosomes**. Plusieurs observations l'établissent : des cellules privées de noyau (comme des fragments d'algue ou des hématies de mammifères) peuvent encore fabriquer des protéines pendant un temps, puis cessent ; des acides aminés radioactifs fournis à une cellule se retrouvent d'abord dans le cytoplasme, associés aux ribosomes. Il existe donc un **intermédiaire** qui transporte l'information du noyau au cytoplasme : l'**ARN messenger** (ARNm).

Les expériences de marquage par l'uracile radioactif, une base présente dans l'ARN mais pas dans l'ADN, montrent qu'un ARN est d'abord synthétisé dans le noyau, puis migre vers le cytoplasme (exemple résolu 2). Sa durée de vie est limitée : il est dégradé après avoir servi, ce qui permet à la cellule d'ajuster en permanence ses productions.

L'expression d'un gène chez un eucaryote

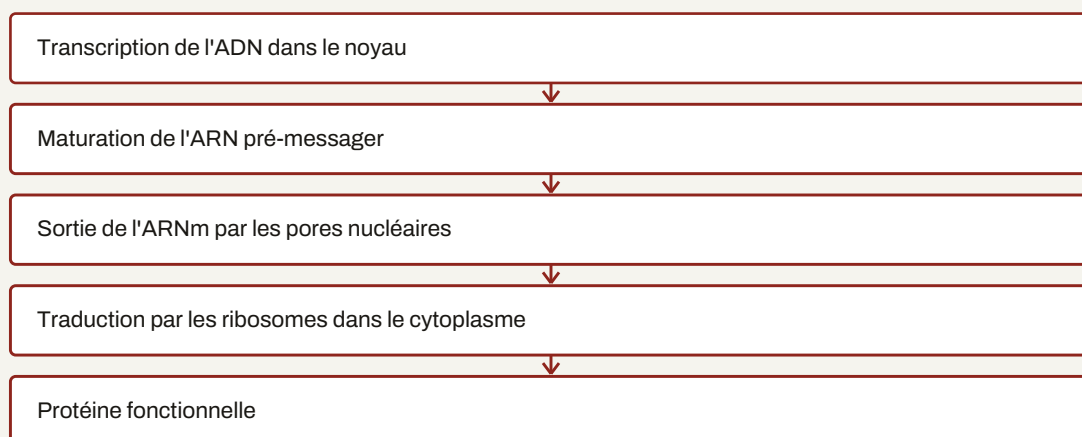


Schéma de principe original • lire les conditions et limites dans le texte

2. L'ARN, une molécule proche de l'ADN

L'**ARN** (acide ribonucléique) est, comme l'ADN, un polymère de nucléotides orienté de 5' vers 3'. Il en diffère sur plusieurs points :

Caractère	ADN	ARN
Sucre	désoxyribose	ribose
Bases	A, T, G, C	A, U, G, C (uracile à la place de la thymine)

Caractère	ADN	ARN
Nombre de brins	deux, en double hélice	un seul
Longueur	très longue (millions de nucléotides)	courte (centaines à milliers de nucléotides)
Localisation chez les eucaryotes	noyau (et mitochondries, chloroplastes)	noyau et cytoplasme
Durée de vie	stable, conservé toute la vie de la cellule	souvent brève

L'uracile s'apparie à l'adénine, comme la thymine. Il existe plusieurs catégories d'ARN : l'**ARN messager** (ARNm), copie d'un gène, qui porte l'information jusqu'aux ribosomes ; l'**ARN de transfert** (ARNt), qui apporte les acides aminés au ribosome ; l'**ARN ribosomique** (ARNr), constituant des ribosomes avec des protéines.

3. La transcription

La **transcription** est la synthèse d'une molécule d'ARN à partir d'une portion d'ADN, selon la règle de complémentarité. Chez les eucaryotes, elle a lieu dans le noyau.

Elle est réalisée par une enzyme, l'**ARN polymérase**, qui se fixe sur une séquence particulière située en amont du gène, le **promoteur**. Elle ouvre localement la double hélice sur une courte longueur, puis se déplace le long du gène. Un seul des deux brins, le **brin transcrit** (ou brin matrice), sert de modèle : l'ARN polymérase lui associe des nucléotides libres à ribose selon la complémentarité (A de l'ADN en face de U de l'ARN, T en face de A, G en face de C, C en face de G). L'ARN est synthétisé dans le sens 5' vers 3' ; il est donc antiparallèle au brin transcrit. L'autre brin, le **brin non transcrit** (ou brin codant), a la même séquence que l'ARN, à ceci près qu'il porte des T à la place des U. En aval du gène, une séquence de terminaison provoque le détachement de l'ARN polymérase et de l'ARN. La double hélice se referme derrière l'enzyme.

Pour trouver l'ARNm à partir de l'ADN : on prend le brin **non transcrit** et l'on remplace T par U ; ou bien on écrit le **complémentaire** du brin transcrit avec U à la place de T.

Un même gène peut être transcrit simultanément par de nombreuses ARN polymérases qui se suivent. Au microscope électronique, on observe alors des figures « en arbre de Noël » : le long de l'ADN (le tronc), les ARN en cours de synthèse (les branches) sont de plus en plus longs vers la fin du gène. Cela permet de produire rapidement de nombreuses copies d'ARN à partir d'un seul gène.

Les deux brins de l'ADN peuvent servir de matrice, mais pour des gènes différents : pour un gène donné, c'est toujours le même brin qui est transcrit.

4. La maturation de l'ARN et l'épissage alternatif

Chez les eucaryotes, la transcription produit un **ARN pré-messager**, qui subit une **maturation** dans le noyau avant de devenir un ARNm fonctionnel.

Les gènes eucaryotes sont **morcelés** : ils alternent des séquences appelées **exons**, qui se retrouvent dans l'ARNm mature, et des séquences appelées **introns**, qui sont transcrites mais éliminées.

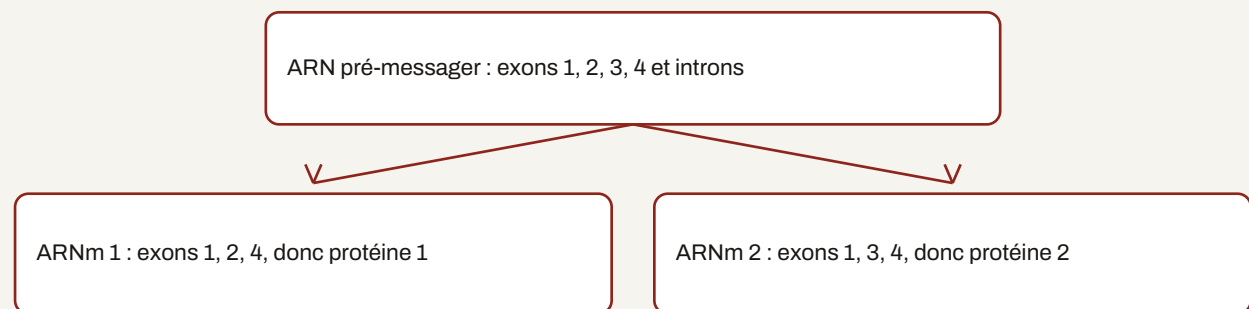
L'**épissage** est l'opération qui excise les introns de l'ARN pré-messager et raboute les exons entre eux. Il est réalisé par un complexe de protéines et d'ARN. L'ARNm mature est donc plus court que le gène

et que l'ARN pré-messager. La maturation comprend aussi l'ajout d'une « coiffe » à l'extrémité 5' et d'une longue queue de nucléotides à A à l'extrémité 3', qui protègent l'ARNm et facilitent sa traduction.

On met en évidence les introns en hybridant un ARNm mature avec le brin transcrit de son gène : les régions complémentaires s'apparient, et les introns, sans partenaire, forment des boucles d'ADN visibles au microscope électronique.

L'**épissage alternatif** désigne la possibilité, pour un même ARN pré-messager, d'être épissé de différentes manières selon le type cellulaire ou le moment : certains exons sont conservés dans un ARNm et éliminés dans un autre. **Un seul gène peut ainsi donner plusieurs ARNm, donc plusieurs protéines différentes.** Par exemple, le même gène produit la calcitonine (une hormone) dans les cellules de la thyroïde et une autre molécule, appelée CGRP, dans certains neurones, selon les exons retenus. L'épissage alternatif explique en partie qu'environ 20 000 gènes humains suffisent à produire un nombre bien plus élevé de protéines différentes.

Épissage alternatif d'un ARN pré-messager



Deux branches distinctes • ce schéma ne relie pas les deux issues entre elles

Chez les **procaryotes**, il n'y a pas de noyau et les gènes n'ont en général pas d'introns : l'ARNm est directement utilisable, et sa traduction commence avant même la fin de la transcription.

5. Le code génétique

La correspondance entre la séquence de l'ARNm et celle de la protéine est le **code génétique**. Avec 4 bases, des « mots » d'une base ne donneraient que 4 combinaisons, de deux bases $4^2 = 16$, insuffisant pour 20 acides aminés ; des mots de trois bases donnent $4^3 = 64$ combinaisons. Un **codon** est un triplet de nucléotides de l'ARNm.

Le code a été déchiffré à partir de 1961 : Marshall Nirenberg et Heinrich Matthaei fournissent à un extrait cellulaire capable de traduire un ARN artificiel formé uniquement d'uracile (poly-U) et obtiennent une chaîne formée uniquement de phénylalanine : UUU code la phénylalanine. Toutes les correspondances ont été établies en quelques années.

1 ^{re} base ↓ 2 ^e base →	U	C	A	G
U	UUU, UUC : Phe ; UUA, UUG : Leu	UCU, UCC, UCA, UCG : Ser	UAU, UAC : Tyr ; UAA, UAG : stop	UGU, UGC : Cys ; UGA : stop ; UGG : Trp

1 ^{re} base ↓ 2 ^e base →	U	C	A	G
C	CUU, CUC, CUA, CUG : Leu	CCU, CCC, CCA, CCG : Pro	CAU, CAC : His ; CAA, CAG : Gln	CGU, CGC, CGA, CGG : Arg
A	AUU, AUC, AUA : Ile ; AUG : Met	ACU, ACC, ACA, ACG : Thr	AAU, AAC : Asn ; AAA, AAG : Lys	AGU, AGC : Ser ; AGA, AGG : Arg
G	GUU, GUC, GUA, GUG : Val	GCU, GCC, GCA, GCG : Ala	GAU, GAC : Asp ; GAA, GAG : Glu	GGU, GGC, GGA, GGG : Gly

Abréviations : Phe phénylalanine, Leu leucine, Ser sérine, Tyr tyrosine, Cys cystéine, Trp tryptophane, Pro proline, His histidine, Gln glutamine, Arg arginine, Ile isoleucine, Met méthionine, Thr thréonine, Asn asparagine, Lys lysine, Val valine, Ala alanine, Asp aspartate, Glu glutamate, Gly glycine.

Propriétés du code génétique :

- sur les 64 codons, **61 codent un acide aminé** et **3 sont des codons stop** (UAA, UAG, UGA), qui ne codent aucun acide aminé et signalent la fin de la traduction ;
- le codon **AUG** code la méthionine et sert de **codon d'initiation** : toutes les protéines commencent par une méthionine, souvent retirée ensuite ;
- le code est **redondant** (ou dégénéré) : la plupart des acides aminés sont codés par plusieurs codons (de 1 pour Met et Trp à 6 pour Leu, Ser et Arg) ;
- il est **non ambigu** : un codon donné ne code jamais qu'un seul acide aminé ;
- il est lu sans chevauchement, codon après codon, à partir du codon initiateur, qui fixe le **cadre de lecture** ;
- il est **universel** : à quelques rares exceptions près, il est le même chez tous les êtres vivants, ce qui témoigne de leur origine commune et permet de faire exprimer un gène humain par une bactérie (production d'insuline humaine par génie génétique, par exemple).

6. La traduction

La **traduction** est la synthèse d'une protéine à partir de l'information portée par l'ARNm. Elle a lieu dans le cytoplasme, sur les **ribosomes**.

Les acteurs.

- Le **ribosome** est un complexe d'ARN ribosomiques et de protéines formé de deux sous-unités. Il se déplace le long de l'ARNm, codon par codon, et catalyse la formation des liaisons entre acides aminés.
- L'**ARN de transfert** (ARNt) est une petite molécule d'ARN repliée. Elle porte à une extrémité un acide aminé précis et, sur une boucle, un triplet de nucléotides, l'**anticodon**, complémentaire d'un codon de l'ARNm. C'est l'ARNt qui fait la correspondance entre codon et acide aminé : c'est le « traducteur » du code. Des enzymes spécifiques chargent chaque ARNt avec l'acide aminé correct.

Les étapes.

- **Initiation** : la petite sous-unité du ribosome se fixe sur l'ARNm près de son extrémité 5' et se déplace jusqu'au codon d'initiation **AUG**, où se place l'ARNt portant la méthionine ; la grande

sous-unité s'associe alors.

- **Élongation** : un ARNt dont l'anticodon est complémentaire du codon suivant apporte son acide aminé ; le ribosome établit une **liaison peptidique** entre la chaîne en cours de croissance et ce nouvel acide aminé. Le ribosome avance d'un codon dans le sens 5' vers 3' de l'ARNm, l'ARNt libéré repart se charger, et le cycle recommence. La chaîne s'allonge ainsi d'un acide aminé par codon.
- **Terminaison** : lorsque le ribosome atteint un **codon stop**, aucun ARNt ne lui correspond ; la chaîne d'acides aminés est libérée et les sous-unités du ribosome se séparent.

Les étapes de la traduction

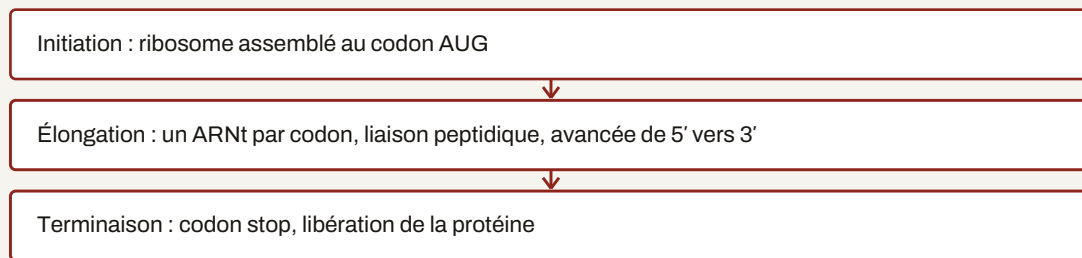


Schéma de principe original • lire les conditions et limites dans le texte

Un même ARNm est lu simultanément par plusieurs ribosomes qui se suivent : l'ensemble forme un **polysome**. Une seule molécule d'ARNm, elle-même produite en nombreux exemplaires par la transcription, permet ainsi de synthétiser rapidement de nombreuses protéines identiques : l'expression d'un gène est **amplifiée** à chaque étape.

Les protéines destinées à être sécrétées ou insérées dans la membrane sont traduites par des ribosomes associés au réticulum endoplasmique rugueux, puis modifiées dans l'appareil de Golgi (ajout de glucides, coupures), étape de **maturation des protéines**.

7. De la séquence à la fonction de la protéine

Une protéine est un polymère d'acides aminés reliés par des liaisons peptidiques. Elle possède une extrémité portant un groupement amine libre (extrémité N-terminale, synthétisée en premier) et une extrémité portant un groupement carboxyle libre (extrémité C-terminale). Les 20 acides aminés diffèrent par leur **chaîne latérale** (ou radical), qui peut être hydrophile, hydrophobe, chargée positivement ou négativement.

La **séquence** des acides aminés détermine la façon dont la chaîne se **replie** dans l'espace : les interactions entre chaînes latérales (liaisons hydrogène, attractions entre charges opposées, regroupement des radicaux hydrophobes à l'abri de l'eau, ponts entre deux cystéines) stabilisent une **structure tridimensionnelle** précise. Certaines protéines associent plusieurs chaînes (l'hémoglobine en compte quatre). C'est cette structure qui permet à la protéine de remplir sa **fonction** : site actif d'une enzyme qui reconnaît son substrat, site de fixation de l'oxygène de l'hémoglobine, forme d'un canal membranaire, d'une hormone ou d'un anticorps.

Les fonctions des protéines sont très variées : enzymes (catalyse), protéines de structure (collagène, kératine), transporteurs (hémoglobine, canaux membranaires), hormones (insuline), récepteurs, anticorps, protéines motrices (myosine).

Une modification de la séquence, même d'un seul acide aminé, peut modifier le repliement, donc la fonction : c'est le lien entre mutation et phénotype.

8. Génotype et phénotype à différentes échelles

Le **génotype** est l'ensemble des allèles d'un individu (pour un ou plusieurs gènes). Le **phénotype** est l'ensemble de ses caractères observables. On le décrit à trois échelles emboîtées :

- **phénotype moléculaire** : les protéines produites et leur activité ;
- **phénotype cellulaire** : la forme, le contenu ou le fonctionnement des cellules ;
- **phénotype macroscopique** (ou de l'organisme) : les caractères visibles ou mesurables de l'individu.

Exemple de l'albinisme. La mélanine, pigment de la peau, des poils et des yeux, est fabriquée dans des cellules appelées mélanocytes, par une chaîne de réactions dont la première est catalysée par l'enzyme **tyrosinase**. Certains allèles mutés du gène de la tyrosinase produisent une enzyme non fonctionnelle. Une personne qui porte deux allèles non fonctionnels (homozygote) présente :

- à l'échelle moléculaire, une tyrosinase inactive ;
- à l'échelle cellulaire, des mélanocytes qui ne produisent pas de mélanine ;
- à l'échelle de l'organisme, une peau et des cheveux très clairs, des yeux peu pigmentés et une sensibilité accrue aux UV.

Un hétérozygote, avec un seul allèle fonctionnel, produit assez d'enzyme pour un phénotype pigmenté : l'allèle fonctionnel est **dominant**, l'allèle muté **récessif**.

L'influence de l'environnement. Le phénotype dépend du génotype et de l'environnement. Chez le chat siamois, la tyrosinase est codée par un allèle qui produit une enzyme active seulement à basse température : le pelage n'est foncé qu'aux extrémités du corps, plus froides (oreilles, pattes, queue, museau). Chez les végétaux, l'exposition à la lumière déclenche l'expression de certains gènes. Chez l'être humain, l'exposition au soleil augmente la production de mélanine.

Une même information, des expressions différentes. Toutes les cellules d'un organisme possèdent le même génome, mais chaque type cellulaire n'exprime qu'une partie de ses gènes : les cellules du pancréas expriment le gène de l'insuline, les hématies en formation celui de la globine. Cette **expression différentielle** des gènes, contrôlée par des signaux internes et externes, explique la diversité des cellules d'un même individu.

Relation génotype et phénotype

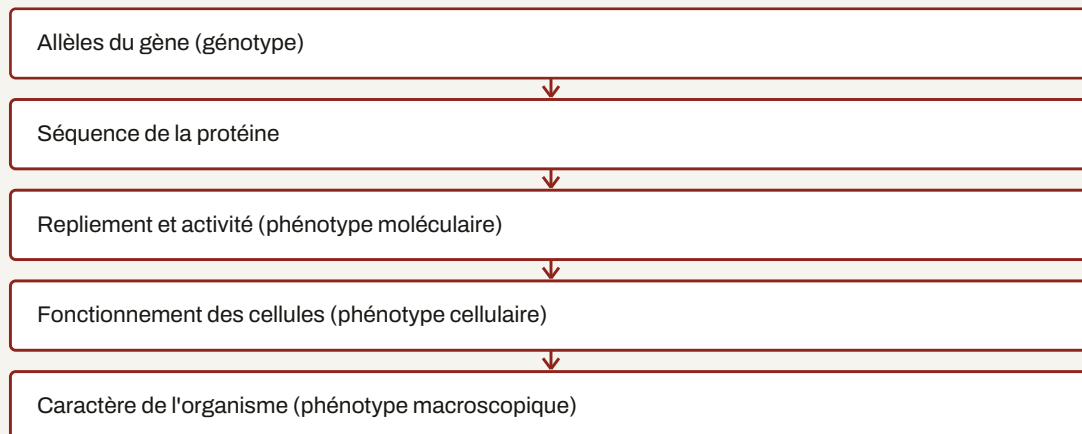


Schéma de principe original • lire les conditions et limites dans le texte

Exemples résolus

Exemple 1 : de l'ADN à la protéine

Énoncé. On donne la séquence du brin transcrit d'un fragment de gène, sans intron : 3' TAC AAA CGT GGT TTA ACT 5' a) Écris la séquence du brin non transcrit et celle de l'ARNm, en précisant leurs extrémités. b) Traduis l'ARNm à l'aide du tableau du code génétique. c) Indique ce que deviendrait la protéine si le 9^e nucléotide du brin transcrit, un T, était remplacé par un C.

Résolution. a) Le brin non transcrit est complémentaire et antiparallèle au brin transcrit : 5' ATG TTT GCA CCA AAT TGA 3'. L'ARNm est complémentaire du brin transcrit, avec U à la place de T ; il a donc la même séquence que le brin non transcrit, U remplaçant T : 5' AUG UUU GCA CCA AAU UGA 3'.

b) On lit les codons de 5' vers 3' à partir de AUG : AUG Met, UUU Phe, GCA Ala, CCA Pro, AAU Asn, UGA stop. Protéine : **Met Phe Ala Pro Asn**, soit 5 acides aminés (le codon stop ne code aucun acide aminé).

c) Le 9^e nucléotide du brin transcrit est le T du triplet CGT (3^e triplet). CGT devient CGC ; le codon de l'ARNm, qui était GCA, devient GCG. GCG code aussi l'alanine : la protéine est **inchangée**, la mutation est **silencieuse**, grâce à la redondance du code génétique.

Exemple 2 : l'ARN messager, intermédiaire entre le noyau et le cytoplasme

Énoncé. On fournit à des cellules en culture de l'uracile radioactif pendant 15 minutes, puis on les transfère dans un milieu contenant de l'uracile non radioactif. On mesure, au cours du temps, la radioactivité du noyau et du cytoplasme (en pourcentage de la radioactivité totale des cellules).

Temps après le début du marquage	15 min	30 min	60 min	90 min
Noyau	95 %	70 %	30 %	15 %

Temps après le début du marquage	15 min	30 min	60 min	90 min
Cytoplasme	5 %	30 %	70 %	85 %

a) Explique pourquoi on utilise de l'uracile radioactif. b) Exploite les résultats. c) Relie ces résultats à la localisation de l'ADN et à celle des ribosomes.

Résolution. a) L'uracile est une base présente dans l'ARN et absente de l'ADN. L'uracile radioactif est incorporé dans les ARN synthétisés pendant le marquage, et seulement dans eux : il permet de **suivre la localisation des ARN** nouvellement formés.

b) *Je constate* qu'à 15 min, 95 % de la radioactivité est dans le noyau : les ARN sont **synthétisés dans le noyau**. *Je constate* qu'ensuite la radioactivité du noyau diminue (de 95 % à 15 %) tandis que celle du cytoplasme augmente d'autant (de 5 % à 85 %). Après le transfert dans le milieu non radioactif, aucun nouvel ARN marqué n'est formé : ce sont les mêmes ARN qui changent de place. *J'en déduis* que les ARN formés dans le noyau **migrent vers le cytoplasme**.

c) L'ADN, qui porte l'information génétique, est dans le noyau, alors que la synthèse des protéines a lieu sur les ribosomes du cytoplasme. Un ARN synthétisé dans le noyau, par transcription de l'ADN, puis exporté dans le cytoplasme, peut **transporter l'information** du gène jusqu'aux ribosomes : c'est l'ARN **messenger**.

Exemple 3 : un gène, deux protéines

Énoncé. Le gène de la calcitonine comporte 6 exons. On compare l'ARN pré-messager et les ARNm obtenus dans deux types cellulaires.

Molécule	Exons présents	Longueur (nucléotides, valeurs arrondies)
ARN pré-messager (toutes cellules)	1, 2, 3, 4, 5, 6 et les 5 introns	5 700
ARNm des cellules de la thyroïde	1, 2, 3, 4	1 000
ARNm de certains neurones	1, 2, 3, 5, 6	1 200

a) Nomme le mécanisme responsable de la différence entre l'ARN pré-messager et les ARNm. b) Explique la différence entre les deux ARNm. c) Les deux cellules produisent-elles la même protéine ? Conclue sur la relation entre nombre de gènes et nombre de protéines.

Résolution. a) Les ARNm sont beaucoup plus courts que l'ARN pré-messager (1 000 et 1 200 nucléotides contre 5 700) : les **introns** ont été éliminés et les exons raboutés. C'est l'**épissage**, étape de la maturation de l'ARN pré-messager dans le noyau.

b) Dans la thyroïde, l'ARNm conserve les exons 1 à 4 et a perdu les exons 5 et 6 ; dans les neurones, il conserve les exons 1, 2, 3, 5 et 6 et a perdu l'exon 4. Le même ARN pré-messager a donc été épissé **de deux façons différentes** selon le type cellulaire : c'est un **épissage alternatif**.

c) Les deux ARNm ont des séquences différentes au-delà de l'exon 3 : leur traduction donne **deux protéines différentes** (la calcitonine dans la thyroïde, le CGRP dans les neurones), qui partagent une partie commune codée par les exons 1 à 3. Un seul gène peut donc coder **plusieurs protéines** : le

nombre de protéines différentes d'un organisme est **supérieur** à son nombre de gènes.

Méthodes et pièges

- Pour passer de l'ADN à l'ARNm, identifie d'abord le brin transcrit ; l'ARNm a la séquence du brin non transcrit, avec U à la place de T.
- Lis toujours l'ARNm de 5' vers 3', à partir du premier codon AUG, et découpe-le en codons sans chevauchement ; arrête-toi au premier codon stop en phase.
- Ne traduis jamais un codon stop en acide aminé, et ne compte pas le stop dans la longueur de la protéine.
- Ne confonds pas codon (sur l'ARNm) et anticodon (sur l'ARNt) : ils sont complémentaires et antiparallèles.
- La redondance du code permet de déduire la protéine à partir de l'ARNm, mais pas l'ARNm à partir de la protéine : plusieurs séquences sont possibles.
- Distingue bien les lieux : transcription et maturation dans le noyau, traduction dans le cytoplasme (chez les eucaryotes).
- Un intron est transcrit mais n'est pas traduit ; il est présent dans le gène et dans l'ARN pré-messager, absent de l'ARNm mature.
- Pour relier génotype et phénotype, décris les trois échelles dans l'ordre (moléculaire, cellulaire, organisme) et n'oublie pas l'éventuel rôle de l'environnement.

Vers le PASS

En PASS, la transcription est étudiée en détail : différentes ARN polymérases chez les eucaryotes, structure des promoteurs, facteurs de transcription, mécanisme moléculaire de l'épissage et de ses sites, maturation des extrémités de l'ARNm, contrôle de sa stabilité. La régulation de l'expression des gènes (séquences régulatrices et protéines qui s'y fixent) y occupe une place centrale, avec le raisonnement expérimental associé.

La traduction est reprise à l'échelle moléculaire (sites du ribosome, facteurs d'initiation et d'élongation, adressage des protéines vers le réticulum, le noyau ou la mitochondrie), de même que l'architecture des protéines (structures primaire à quaternaire) et le lien entre repliement et fonction. Ces notions servent en pharmacologie (cibles des antibiotiques sur le ribosome bactérien) et en génétique médicale (variants qui perturbent l'épissage).

Suite en PASS : Transcription et maturation des ARN

Exercices

Niveau 1 : s'entraîner

1. Recopie et complète par ADN, ARN ou les deux : a) contient du ribose ; b) contient de la thymine ; c) est formé de nucléotides ; d) est simple brin ; e) est présent dans le cytoplasme des cellules eucaryotes.
2. Le brin non transcrit d'un fragment de gène a pour séquence 5' ATG GCT TGG AAA CAT TAG 3'.
a) Écris la séquence de l'ARNm. b) Écris celle du brin transcrit en précisant ses extrémités. c) Traduis l'ARNm.

3. Traduis l'ARNm 5' GCAUGCCAGAUUGCAAAUAGGC 3' en repérant le codon d'initiation, et indique le nombre d'acides aminés de la protéine.
4. Donne : a) le nombre de codons du code génétique ; b) le nombre de codons stop et leurs séquences ; c) les acides aminés codés par un seul codon ; d) le nombre de codons de la leucine.
5. Remets dans l'ordre les étapes suivantes et indique le lieu de chacune dans une cellule eucaryote : traduction ; épissage ; transcription ; sortie de l'ARNm par les pores nucléaires ; repliement de la protéine.

Niveau 2 : type bac

6. Une protéine commence par la séquence Met Trp Lys Phe. a) Écris toutes les séquences d'ARNm possibles pour ce fragment. b) Combien de séquences différentes peut-on écrire ? c) Explique pourquoi on ne peut pas retrouver la séquence exacte d'un gène à partir de celle de sa protéine.
7. On observe au microscope électronique la transcription d'un gène : le long d'un filament d'ADN, des filaments d'ARN de plus en plus longs sont fixés, tous d'un même côté. a) Explique cette disposition. b) Indique où se trouve le début du gène. c) Montre l'intérêt de cette organisation pour la cellule.
8. On hybride l'ARNm mature d'un gène d'ovalbumine (une protéine du blanc d'œuf) avec le brin transcrit de ce gène, puis on observe la molécule hybride au microscope électronique : on voit 7 boucles d'ADN simple brin qui dépassent d'une structure double. a) Explique la formation des boucles. b) Déduis-en le nombre d'introns et d'exons du gène, sachant que le début et la fin de l'ARNm sont appariés. c) Le gène compte environ 7 700 paires de bases et l'ARNm mature environ 1 900 nucléotides : calcule la proportion du gène occupée par les introns.
9. Chez le lapin himalayen, le pelage est blanc sur le corps et noir aux extrémités (oreilles, pattes, queue). Si l'on rase une zone du dos et qu'on y maintient une poche de glace, les poils repoussent noirs. a) Formule une hypothèse sur l'enzyme de synthèse du pigment noir. b) Montre que le phénotype dépend à la fois du génotype et de l'environnement. c) Décrit le phénotype aux échelles moléculaire, cellulaire et de l'organisme pour une zone du dos non refroidie.

Niveau 3 : vers le PASS

10. Une protéine humaine est formée de 146 acides aminés après retrait de la méthionine initiale. a) Calcule le nombre minimal de nucléotides de son ARNm codant (de AUG au codon stop inclus). b) Son gène comporte 3 exons et 2 introns de 130 et 850 paires de bases. On néglige les régions non traduites. Calcule la longueur de la portion transcrite du gène. c) Explique pourquoi la masse d'une protéine ne permet pas de prédire la longueur de son gène chez un eucaryote.
11. Dans un extrait cellulaire capable de traduire, on introduit des ARN artificiels. a) Avec un ARN poly-U (UUUU...), on obtient de la polyphénylalanine. Avec un ARN poly-A, on obtient de la polylysine. Vérifie ces résultats à l'aide du tableau. b) Avec un ARN formé de la répétition de UC (UCUCUC...), on obtient une chaîne qui alterne deux acides aminés. Lesquels ? Justifie. c) Avec un ARN formé de la répétition de UUC (UUCUUCUUC...), on obtient trois types de chaînes homogènes. Lesquelles, et pourquoi ? d) Montre que ces résultats confirment que le code est lu par triplets sans chevauchement.
12. QCM : indique pour chaque proposition si elle est vraie ou fausse et justifie.

A. L'ARN polymérase lit le brin non transcrit. B. L'ARNm a la même séquence que le brin transcrit, avec U à la place de T. C. L'anticodon de l'ARNt qui reconnaît le codon 5' AUG 3' est 3' UAC 5'. D.

L'épissage alternatif permet à une cellule de produire des protéines différentes à partir d'un seul gène. E. Deux cellules d'un même individu qui ne produisent pas les mêmes protéines ont des génomes différents.

13. La phénylcétonurie est due à des allèles non fonctionnels du gène de la phénylalanine hydroxylase (PAH), enzyme du foie qui transforme la phénylalanine en tyrosine. Chez une personne homozygote pour un allèle muté, la phénylalanine s'accumule dans le sang et, sans prise en charge, perturbe le développement du cerveau. Un régime pauvre en phénylalanine, débuté dès la naissance, empêche ces troubles. a) Décris le phénotype de cette personne aux trois échelles. b) Un des allèles mutés porte une substitution qui remplace le codon CGG (arginine) par UGG au codon 408. Nomme le type de mutation et prévois son effet. c) Explique pourquoi le régime modifie le phénotype sans modifier le génotype. d) Explique pourquoi les parents, hétérozygotes, n'ont pas de phénylcétonurie.

Corrigés

1. a) Ribose : ARN. b) Thymine : ADN. c) Formé de nucléotides : **les deux**. d) Simple brin : ARN. e) Présent dans le cytoplasme : ARN (ARNm, ARNt, ARNr ; l'ADN cytoplasmique des mitochondries et chloroplastes est enfermé dans ces organites). On attend la réponse ARN.
2. a) L'ARNm a la séquence du brin non transcrit avec U à la place de T : 5' AUG GCU UGG AAA CAU UAG 3'.
- b) Brin transcrit, complémentaire et antiparallèle : 3' TAC CGA ACC TTT GTA ATC 5'. c) AUG Met, GCU Ala, UGG Trp, AAA Lys, CAU His, UAG stop : **Met Ala Trp Lys His**.
3. On cherche le premier AUG : GC **AUG** CCA GAU UGC AAA UAG GC. Codons à partir de AUG : AUG Met, CCA Pro, GAU Asp, UGC Cys, AAA Lys, UAG stop. Protéine : **Met Pro Asp Cys Lys**, soit **5 acides aminés**. Les nucléotides GC du début et GC de la fin ne sont pas traduits.
4. a) **64 codons** (4^3). b) **3 codons stop** : UAA, UAG, UGA. c) **La méthionine (AUG) et le tryptophane (UGG)**. d) **6 codons** pour la leucine : UUA, UUG, CUU, CUC, CUA, CUG.
5. **Transcription (noyau), épissage (noyau), sortie de l'ARNm par les pores nucléaires (enveloppe nucléaire), traduction (cytoplasme, ribosomes), repliement de la protéine (cytoplasme ou réticulum endoplasmique)**.
6. a) Met : AUG ; Trp : UGG ; Lys : AAA ou AAG ; Phe : UUU ou UUC. Les séquences possibles sont : **AUG UGG AAA UUU ; AUG UGG AAA UUC ; AUG UGG AAG UUU ; AUG UGG AAG UUC**.
- b) $1 \times 1 \times 2 \times 2 = 4$ **séquences**. c) Le code est **redondant** : la plupart des acides aminés sont codés par plusieurs codons. Pour une protéine de plusieurs centaines d'acides aminés, le nombre de séquences possibles est immense. De plus, chez les eucaryotes, la protéine ne renseigne pas sur les **introns**, absents de l'ARNm. On ne peut donc pas retrouver la séquence exacte du gène.
7. a) Le gène est transcrit simultanément par **plusieurs ARN polymérases** qui se suivent. Chacune porte l'ARN qu'elle est en train de synthétiser : plus une polymérase est avancée sur le gène, plus son ARN est long. Les ARN sont d'un même côté car c'est toujours le **même brin** qui est transcrit.
- b) Le début du gène (promoteur) est du côté des **ARN les plus courts**, puisque les polymérases y ont transcrit le moins de nucléotides. c) De nombreux ARN sont produits en même temps à partir d'**un seul gène** : la cellule obtient rapidement beaucoup de copies d'ARNm, donc beaucoup de protéines (amplification de l'expression).

8. a) L'ARNm s'apparie par complémentarité aux régions du brin transcrit qui correspondent aux **exons**. Les **introns**, éliminés lors de l'épissage, n'ont pas de séquence complémentaire dans l'ARNm : ils restent simples brins et forment des **boucles**.
- b) Chaque boucle correspond à un intron : **7 introns**. Le début et la fin étant appariés, il y a un exon de part et d'autre de chaque intron : **8 exons**. c) Longueur des introns $\approx 7\,700 - 1\,900 = 5\,800$ paires de bases. Proportion $= 5\,800 / 7\,700 \approx 0,75$: les introns occupent **environ 75 % du gène**.
9. a) Hypothèse : l'enzyme de synthèse du pigment noir (une forme de tyrosinase) est **thermosensible** : elle n'est active qu'à **basse température**. Aux extrémités, plus froides, elle fonctionne et les poils sont noirs ; sur le corps, plus chaud, elle est inactive et les poils sont blancs.
- b) Toutes les cellules du lapin ont le **même génotype**, donc le même allèle de l'enzyme ; pourtant la couleur diffère selon la zone. Le refroidissement artificiel d'une zone du dos y fait apparaître des poils noirs : un facteur de l'**environnement**, la température, modifie le phénotype. Le phénotype résulte donc de l'interaction entre le **génotype** (allèle codant une enzyme thermosensible) et l'**environnement** (température locale). c) Zone du dos non refroidie : **moléculaire**, enzyme présente mais inactive à la température du corps ; **cellulaire**, cellules productrices de pigment des follicules pileux sans pigment noir ; **organisme**, pelage blanc sur le dos.
10. a) 146 acides aminés + la méthionine initiale = 147 codons traduits, plus 1 codon stop = 148 codons. Nombre de nucléotides $= 148 \times 3 = 444$ **nucléotides**.
- b) Portion transcrite = exons (444 nucléotides, régions non traduites négligées) + introns (130 + 850 = 980) = **1 424 paires de bases**. c) Chez un eucaryote, le gène contient des **introns** dont la longueur, très variable, n'a aucun lien avec la taille de la protéine (ici les introns représentent $980/1\,424 \approx 69\%$ de la portion transcrite). De plus, la protéine peut être modifiée après traduction (méthionine retirée, coupures, ajout de glucides). La masse de la protéine ne permet donc de prédire que la **longueur de la partie codante**, pas celle du gène.
11. a) Lu par triplets, poly-U ne contient que des codons UUU, qui codent **Phe** ; poly-A ne contient que des codons AAA, qui codent **Lys**. Les résultats sont conformes au tableau.
- b) UCUCUC... lu par triplets donne la suite UCU CUC UCU CUC... : UCU code la **sérine** et CUC la **leucine**. On obtient une chaîne qui alterne **Ser et Leu**, quel que soit le point de départ de la lecture (s'il commence au C, on lit CUC UCU..., soit Leu Ser...). c) UUCUUCUUC... peut être lu selon trois cadres de lecture selon le nucléotide de départ : UUC UUC... (**Phe**), UCU UCU... (**Ser**) ou CUU CUU... (**Leu**). Chaque cadre donne une chaîne homogène : **polyphénylalanine, polysérine, polyleucine**. d) Un code par doublets donnerait, pour UCUCUC..., la suite UC UC UC..., donc une chaîne homogène, ce qui n'est pas observé. Un code chevauchant lirait UCU, CUC, UCU... en se décalant d'un seul nucléotide, et pour UUC répété on obtiendrait dans une même chaîne une alternance de plusieurs acides aminés au lieu de trois chaînes homogènes. Les résultats correspondent uniquement à une lecture **par triplets, sans chevauchement**, à partir d'un point de départ qui fixe le cadre de lecture.

12. A. **Faux.** L'ARN polymérase utilise le brin transcrit comme matrice ; l'ARN a la séquence du brin non transcrit.

B. **Faux.** L'ARNm est complémentaire du brin transcrit ; c'est au brin non transcrit qu'il est identique, U remplaçant T. C. **Vrai.** L'anticodon est complémentaire et antiparallèle au codon : en face de 5' A U G 3', on a 3' U A C 5'. D. **Vrai.** En conservant des exons différents, l'épissage alternatif produit plusieurs ARNm, donc plusieurs protéines, à partir d'un même gène. E. **Faux.** Toutes les cellules d'un individu (issues de la cellule-œuf par mitoses) ont le même génome ; elles n'expriment pas les mêmes gènes (expression différentielle).

13. a) **Moléculaire** : enzyme PAH non fonctionnelle, la phénylalanine n'est pas transformée en tyrosine. **Cellulaire** : les cellules du foie ne dégradent pas la phénylalanine, qui s'accumule dans le sang et, en excès, perturbe le fonctionnement des cellules nerveuses en développement. **Organisme** : sans prise en charge, troubles du développement du cerveau.

b) CGG (arginine) devient UGG (tryptophane) : c'est une **substitution faux-sens**. L'arginine, chargée positivement, est remplacée par le tryptophane, gros et hydrophobe : le repliement de l'enzyme est modifié, ce qui peut rendre l'enzyme **non fonctionnelle**, comme l'indique son association avec la maladie. c) Le régime apporte peu de phénylalanine : même sans enzyme fonctionnelle, celle-ci ne s'accumule pas et ne perturbe pas le développement du cerveau. Le **génotype** est inchangé (deux allèles mutés), mais l'**environnement** (l'alimentation) modifie le **phénotype** aux échelles cellulaire et de l'organisme. d) Un hétérozygote possède un allèle fonctionnel, qui produit assez d'enzyme pour transformer la phénylalanine : il n'y a pas d'accumulation. L'allèle fonctionnel est **dominant**, l'allèle muté **récessif** : seuls les **homozygotes** pour un allèle muté sont atteints.