

1 - Notions de base de génétique

C. LEROUX¹, G. TOSSER-KLOPP²

¹INRA, Laboratoire de Génétique Biochimique et de Cytogénétique
78352 Jouy-en-Josas cedex

²INRA, Laboratoire de Génétique Cellulaire
BP 27, 31326 Castanet-Tolosan cedex

e-mail : cleroux@biotec.jouy.inra.fr

La fonction du gène : les grandes étapes de l'utilisation de l'information génétique

Résumé. L'utilisation de l'information portée par l'ADN nécessite une étape de transcription des gènes en ARN et, dans certains cas, une phase de traduction en protéine. La transcription est un phénomène complexe, régulé (un gène eucaryote contient des séquences régulatrices permettant l'initiation et la régulation de la transcription) qui produit trois types d'ARN : les ARN messagers (ARNm), les ARN de transfert (ARNt) et les ARN ribosomiques (ARNr). Les ARNt et les ARNr sont des constituants de la machinerie de traduction des ARNm. En effet, après la transcription, l'ARNm va être mûr (épissage, polyadénylation ...) puis va sortir du noyau pour aller dans le cytoplasme où sa séquence nucléotidique sera traduite en séquence d'acides aminés. Les protéines acquièrent ensuite leur structure définitive après des modifications post-traductionnelles diverses. La régulation de l'expression des gènes peut intervenir à chacune des étapes (transcription, maturation et/ou traduction).

L'activité cellulaire est coordonnée et programmée à partir d'une information stockée sous forme d'ADN, que la cellule peut recopier (réplication) et transmettre à ses cellules filles. L'utilisation de cette masse d'information, dont les principales étapes sont décrites dans ce texte, requiert un système séquentiel de copie en ARN (la transcription) et, dans certains cas, de traduction en protéines. L'article n'aborde le fonctionnement des gènes que chez les eucaryotes ; il s'inspire des ouvrages de synthèse de Berg et Singer (1993), Lewin (1994), Etienne (1998) et Maftah et Julien (1999). Après quelques généralités relatives à la première étape de l'expression d'un gène, le deuxième chapitre est consacré à la description des mécanismes concourant à la production d'une protéine et le troisième chapitre présente quelques associations des différentes étapes de l'expression génétique avec des mécanismes de régulation.

1 / La première étape de l'expression des gènes est commune à tous

1.1 / Qu'est que la transcription ?

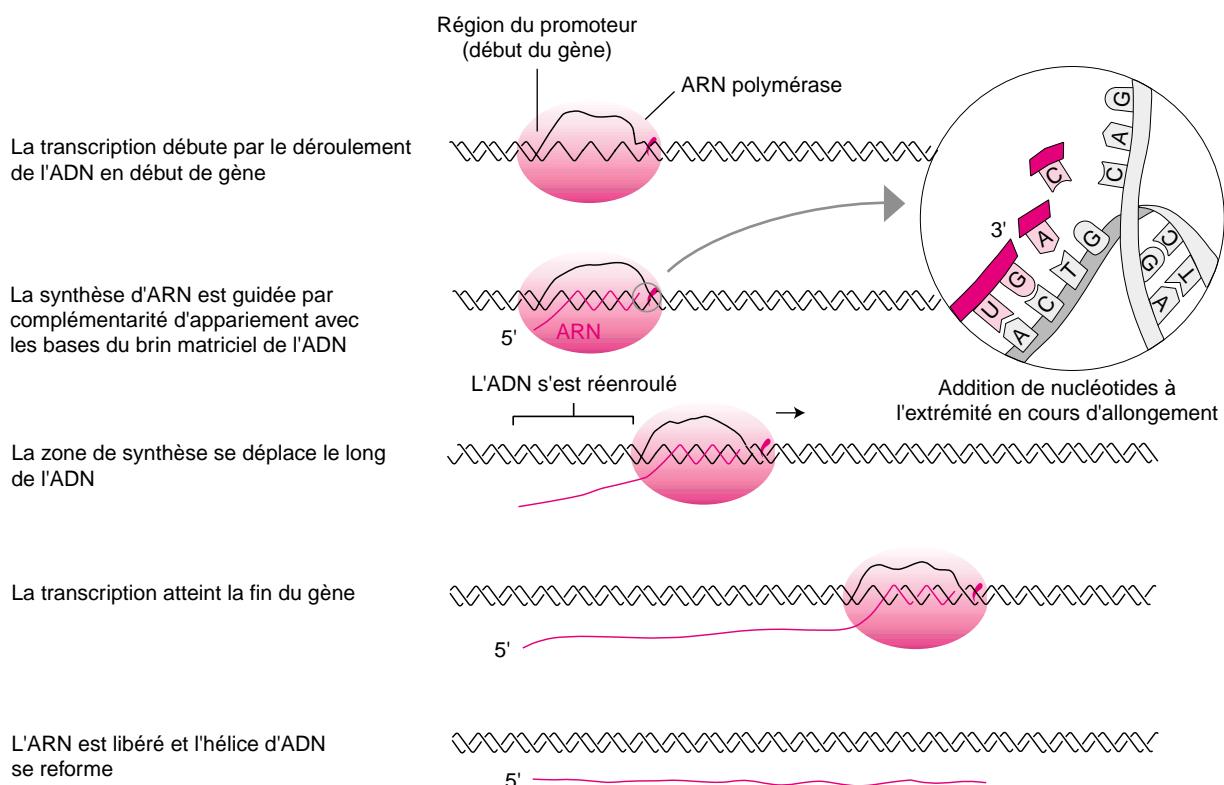
L'expression de tous les gènes cellulaires débute avec la transcription, le processus par lequel la séquence de nucléotides d'un gène est copiée sous forme d'un brin d'ARN. Chez les eucaryotes, cette étape est effectuée dans le noyau. La transcription (figure 1) est accomplie par les ARN polymérases. Pour amorcer la transcription, la molécule d'ARN polymérase se lie à des séquences d'ADN bien particulières, les promoteurs, qui se trouvent au début de

chaque gène. A chaque promoteur, l'enzyme déroule les deux brins de l'ADN et sélectionne celui qui servira de matrice pour la copie. Le choix des nucléotides qui seront ajoutés au brin d'ARN est déterminé par leur complémentarité d'appariement avec les nucléotides successifs du brin d'ADN (A:T et C:G). Le brin d'ARN (monocaténaire) ainsi synthétisé possède la même séquence de nucléotides et la même polarité que l'un des deux brins d'ADN, excepté que U y remplace T. Des séquences de nucléotides particulières placées en fin de gène signalent l'arrêt de la transcription et provoquent le détachement du brin d'ARN. On dit que l'ARN est synthétisé dans le sens 5' vers 3', 5' au début et 3' à la fin.

1.2 / Les intervenants

a / Les différents types d'ARN

La transcription engendre différents types fonctionnels d'ARN. Les gènes qui spécifient les séquences d'acides aminés des protéines sont transcrits en ARN messagers (ARNm). Certains gènes produisent d'autres sortes d'ARN, les plus abondants étant les ARN ribosomiques (ARNr) et les ARN de transfert (ARNt). Les ARNr qui entrent dans la constitution des ribosomes (voir le paragraphe consacré à la traduction) et les ARNt qui assurent le transport des acides aminés, ne codent pas pour des protéines mais composent plutôt la machinerie qui traduit les ARNm en protéines. Les ARNr et les ARNt représentent, respectivement, 90 % et un peu moins de 10 % des ARN totaux d'une cellule. Les ARNm qui servent, eux, d'intermédiaire entre les gènes et les protéines, représentent moins de 1 % des ARN totaux.

Figure 1. La transcription de l'ADN en ARN (d'après Berg et Singer 1993).

b / Les ARN polymérases des eucaryotes

La polymérase I sert à la transcription des ARNr dont les gènes existent en de nombreuses copies. Ces gènes sont transcrits en un précurseur 45S qui va se scinder pour donner les ARN 18, 28 et 5,8S. L'ARNr 18S entouré de protéines sort du noyau et forme la petite sous-unité des ribosomes. Les ARNr 28 et 5,8S associés à un autre petit ARN (5S) et à différentes protéines vont constituer la grande sous-unité des ribosomes.

La polymérase II assure, quant à elle, la transcription des pré-messagers (pré-ARNm), ainsi que d'autres petits ARN (snRNA) entrant dans la constitution des complexes d'épissage (voir plus loin). Les ARN messagers seront lus ou traduits pour aboutir à la synthèse de protéines.

Enfin, la polymérase III transcrit les ARNr 5S, les ARNt ainsi que certains petits ARN (snRNA) impliqués dans la formation du complexe d'épissage. Les gènes des ARNt sont, également, très répétés.

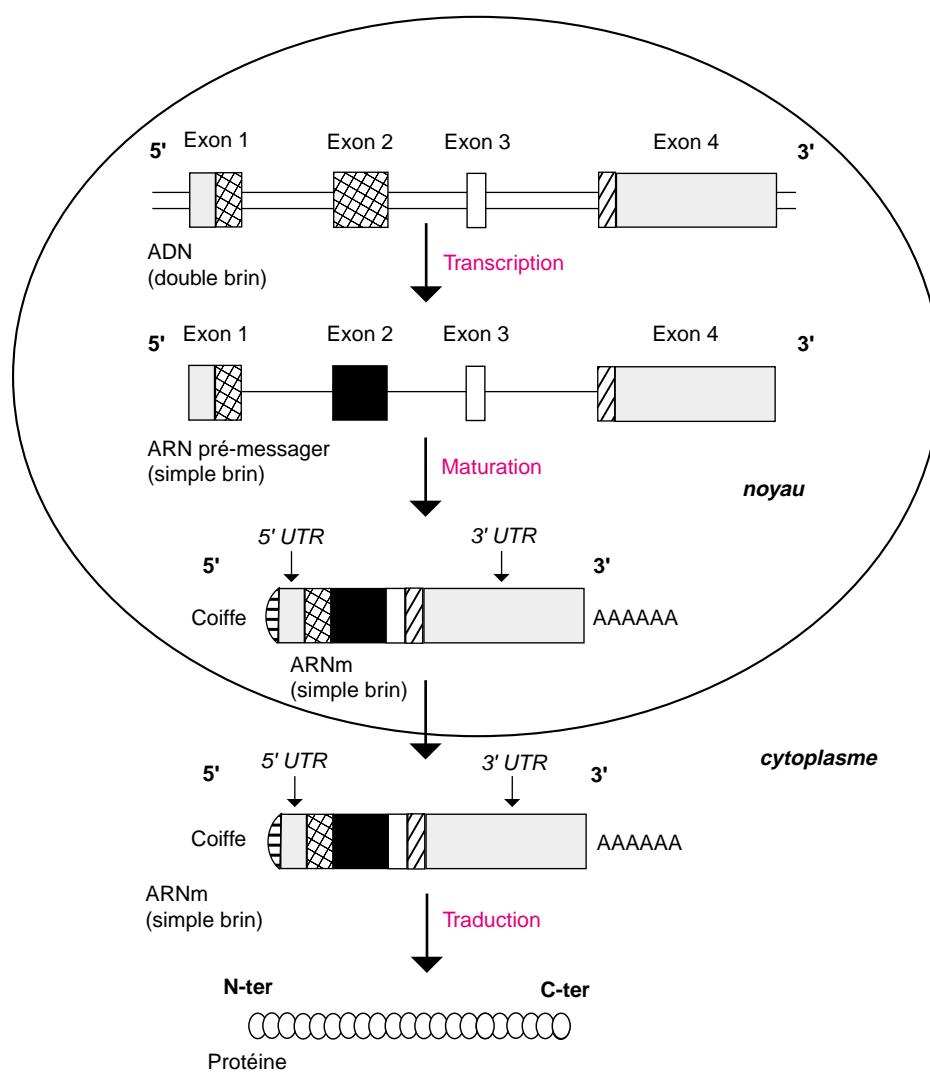
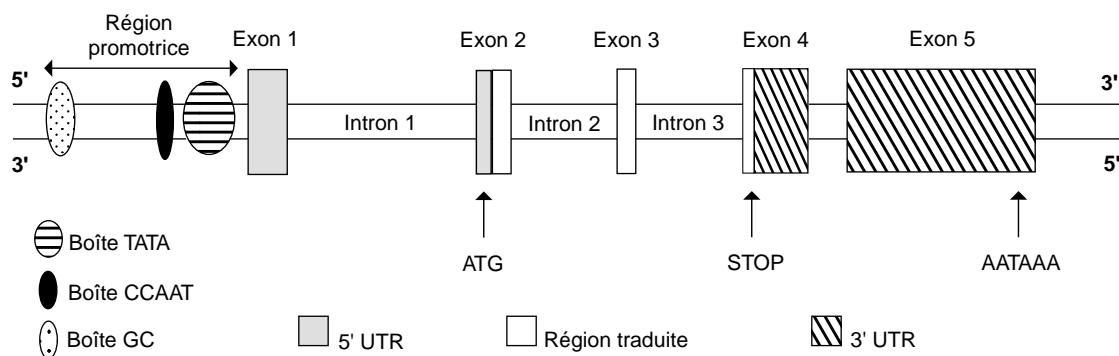
2 / Du gène à la protéine

Alors que, chez les procaryotes, la transcription et la traduction se déroulent dans le même compartiment cellulaire, chez les eucaryotes, la transcription est effectuée dans le noyau et la traduction dans le cytoplasme. Les différentes étapes depuis la transcription jusqu'à la synthèse protéique (figure 2) sont décrites dans les paragraphes suivants.

2.1 / De l'ADN à l'ARNm

a / La structure des gènes chez les eucaryotes

La transcription consiste donc à copier l'ADN en ARN. Cependant, tout l'ADN n'est pas transcrit, mais seulement certaines parties. Ces séquences transcris, ou gènes, contiennent les signaux nécessaires à la transcription (figure 3). Ainsi, le signal de début de transcription est contenu dans la région promotrice. La centaine de paires de bases immédiatement en amont du site d'initiation de la transcription est appelée promoteur proximal. Dans cette région, sont trouvées essentiellement la boîte TATA, présente dans presque tous les gènes, ainsi que les boîtes CCAAT et GC moins systématiquement retrouvées. Consécutivement à la région promotrice proximale, le site d'initiation de la transcription délimite la partie 5' du gène qui sera copiée en ARN. Les gènes des eucaryotes sont pour la plupart "discontinus" ou "fragmentés". En effet, ils contiennent des exons qui seront conservés dans l'ARNm mature et des introns qui, eux, seront absents de l'ARNm cytoplasmique consécutivement à leur élimination lors de l'épissage. Si la majorité des pré-ARNm subissent l'étape d'épissage, ce n'est pas une règle absolue car certains gènes sont monoexoniques et ne contiennent donc pas d'intron. Enfin, la région 3' du gène contient le signal de polyadénylation ainsi que des signaux de terminaison de transcription moins bien connus. Dans certains cas, la fin de la transcription peut avoir lieu plus de 1000 paires de bases en aval du site de polyadénylation qui délimite l'extrémité 3' des ARNm.

Figure 2. Les grandes étapes de l'expression d'un gène.**Figure 3.** Exemple de l'organisation structurale d'un gène eucaryote.**b / L'initiation de la transcription**

L'initiation de la transcription nécessite la présence, dans la région promotrice, d'un complexe multiprotéique. Ce complexe est, notamment, formé de facteurs transcriptionnels (TF). Parmi ces protéines, seule TFIID se fixe sur la boîte TATA. TFIIA stabiliserait le complexe alors que

TFIID-ADN, TFIIB et TFIIE assureraient la liaison avec l'ARN polymérase. Enfin TFIIF aurait une activité hélicase permettant de dénaturer la double hélice de l'ADN au niveau du site d'initiation de la transcription. Ce groupe de facteurs généraux ainsi que l'ARN polymérase constituent donc le complexe d'initiation. La transcription démarre quelques nucléotides en aval de la boîte TATA au

site d'initiation de la transcription (+1). La boîte TATA permet une initiation précise de la transcription. Ainsi, par exemple, l'hétérogénéité des extrémités 5' des ARNm du lysozyme de poulet serait consécutive à l'absence de boîte TATA au niveau du gène les spécifiant.

Outre le promoteur proximal évoqué dans le paragraphe précédent il existe d'autres séquences régulatrices constituant les séquences distales du promoteur. Ces éléments cis-régulateurs sont généralement constitués de quelques nucléotides et peuvent être éloignés (en amont ou même en aval) du site d'initiation de la transcription. C'est sur ces séquences que viennent se fixer les facteurs "trans-régulateurs". Parmi ces séquences cis-régulatrices, on note les "enhancers" ou "activateurs" ainsi que les "silencers" ou "séquences extinctrices". Leurs particularités sont d'être parfois localisées à de très grandes distances (1 à 30 kilobases) et d'être capables d'agir indépendamment de leur position (qu'ils soient en amont, en aval ou même dans le gène) ou de leur orientation (sur l'un ou l'autre des brins d'ADN).

De nombreux éléments cis-régulateurs sont maintenant connus. Ils sont retrouvés "dans plusieurs gènes", mais chaque gène, dans une cellule animale, possède une combinaison particulière d'éléments cis-régulateurs uniques dans leur type, leur nombre ou leur position. Citons juste quelques exemples à titre indicatif :

- la boîte GC est formée d'éléments riches en bases G et C tels que par exemple l'hexanucléotide GGGCGG, répété plusieurs fois. Cette boîte se trouve dans la zone promotrice proximale de certains gènes à un emplacement variable, environ -110 à -40 pb. Il semble que les gènes constitutifs (housekeeping

gènes) possèdent le plus souvent des boîtes GC comparativement aux gènes régulés. Cette séquence est reconnue par une protéine appelée "sp1" qui possède une structure dite en doigt de zinc (3 doigts) ;

- une boîte CRE (cAMP responsive element) dont la séquence est l'octamère palindromique TGACGT-CA, a été retrouvée dans la région promotrice de gènes inducibles par l'AMPc. Cette séquence serait reconnue par une protéine CREB (CRE-binding protein) qui présente une région dite "Leucine Zipper" responsable de sa dimérisation (par des liaisons hydrophobes évoquant un système à "fermeture éclair") nécessaire à son activation ;

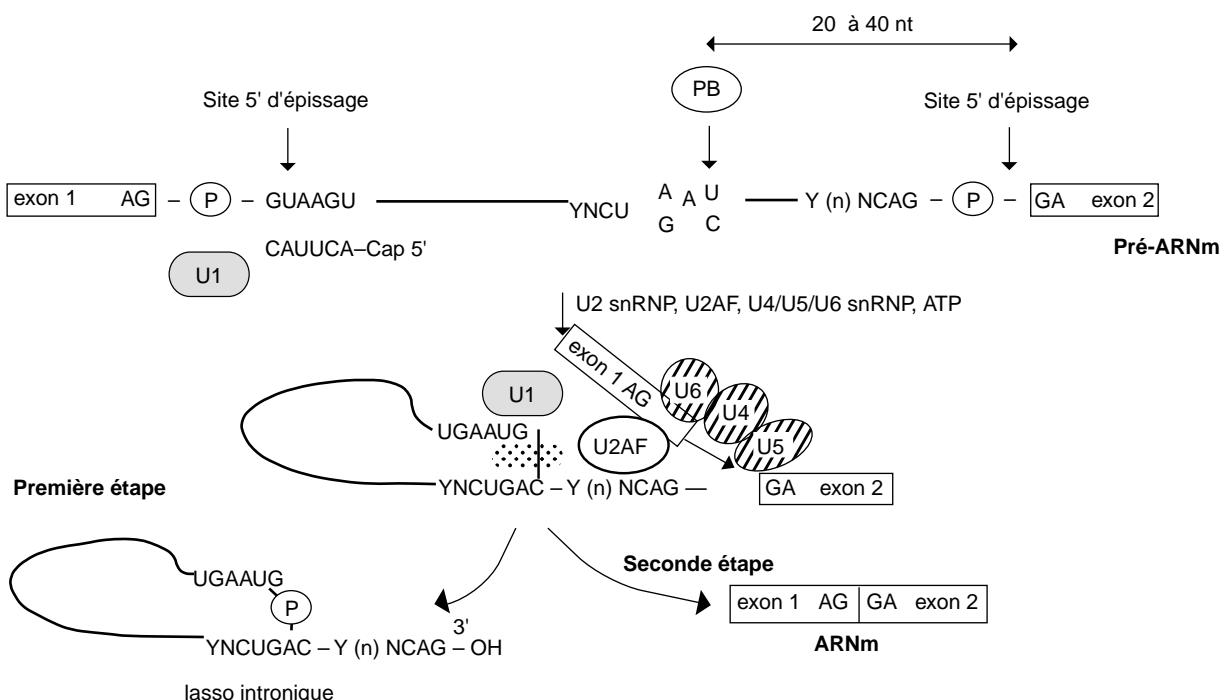
- parmi les séquences recevant les complexes "hormones+récepteurs", citons les GRE (glucocorticoid response element) dont la séquence est AGAACANNNTGTTCT. Il semblerait que le récepteur complexé se fixe sur cette séquence nucléotidique par sa partie centrale ou N-terminale faisant intervenir une structure en doigt de zinc.

c / La maturation des ARNm

L'épissage

Au cours de la transcription, l'ARN polymérase II synthétise un transcript primaire, continu, colinéaire du gène complet. Le transcript cytoplasmique, généralement plus court, ne contient que les séquences dites exoniques. Avant de sortir du noyau, le transcript doit subir une étape de maturation qui consiste à éliminer les séquences intervenantes (ou introns) et à "rabouter" les exons pour former le transcript mature. Cette opération est appelée "épissage" (figure 4).

Figure 4. Les étapes du processus d'épissage des ARNm.



L'élimination des introns est complexe et requiert de nombreux facteurs. La première étape correspond à un clivage de l'extrémité 5' de l'intron (site donneur) qui va se lier au résidu «A» du point de branchement situé entre 20 à 40 nucléotides en amont de l'extrémité 3' de l'intron (site accepteur). Cette réaction génère une structure dite en "lasso". Suite au clivage de l'extrémité 3' de l'intron et au raboutage des exons, l'intron ainsi libéré va rapidement être dégradé. Cette opération s'effectue au sein d'une structure nucléaire complexe appelée "spliceosome" (Maniatis et Reed 1987). Ce dernier est composé de protéines ainsi que de particules ribonucléoprotéiques appelées U snRNPs (pour Uridine-rich small nuclear RiboNucleoprotein Particles) ou "snurps" qui sont composées de protéines et de petits ARN nucléaires (snRNA). Différents facteurs du «spliceosome» reconnaissent les séquences consensus d'épissage que sont le site donneur, le point de branchement et le site accepteur. Ainsi, par exemple, la reconnaissance du site donneur est assurée par complémentarité de base avec l'extrémité 5' de l'ARN constituant de la particule U1 (U1 snRNP). Les séquences autres que les séquences consensus peuvent jouer un rôle déterminant dans le mécanisme d'épissage. Ainsi, l'influence des séquences exoniques sur la sélection des sites d'épissage a notamment été démontrée ; de même, des structures secondaires peuvent jouer un rôle déterminant dans l'accessibilité de certains sites.

L'épissage doit être précis car une erreur d'une base lors de la reconnaissance de la séquence exonique produirait une insertion ou une délétion au niveau de l'ARNm qui, en décalant le cadre de lecture, aurait pour conséquence d'entraîner une modification des résidus d'acides aminés suivants et/ou l'apparition prématûrée d'un codon stop. Mais précision n'est pas synonyme de rigidité. En effet, le processus d'épissage est suffisamment souple pour moduler l'utilisation de certains exons : on parle alors d'épissage différentiel ou d'épissage alternatif. La machinerie d'épissage exerce parfois un choix et certains exons ou parties d'exons (utilisation de site cryptique d'épissage) sont alors utilisés de façon différentielle, selon le contexte dans lequel est le gène. L'épissage alternatif d'un exon peut être "contrôlé" (ex : exons mutuellement exclusifs), constitutif (ex : présence de transcrits multiples) ou consécutif à une mutation allélique. Ce phénomène «alternatif» peut également intervenir dans la région promotrice (utilisation alternative de promoteurs) ou dans la région 3' non traduite (utilisation alternative de signaux de polyadénylation). Ces phénomènes, encore peu connus il y a une dizaine d'années, sont maintenant décrits pour de nombreux gènes.

Le "raboutage" des exons, après excision des introns, concerne généralement des exons appartenant au même transcript primaire ("cis-splicing"). Il a été toutefois décrit, chez quelques organismes (par exemple un trypanosome et un nématode), un épissage d'exons provenant de transcrits primaires différents ("trans-splicing") et donnant donc un ARNm mosaïque.

Les modifications des extrémités

L'une des premières modifications du pré-ARNm est l'ajout d'une "coiffe" ou "chapeau" à son extrémité 5'. En effet, cette coiffe (7-méthylguanosine) se met en place dès le début de la transcription, avant que les 30 premiers nucléotides ne soient assemblés. L'ARNm n'aura donc plus une extrémité 5' phosphate libre. Le rôle de la coiffe est

double. D'une part, elle protège cette région contre les nucléases et phosphatasées et, d'autre part, elle favorise la fixation du complexe d'initiation de la traduction, jouant de ce fait un rôle important dans la lecture du message.

L'extrémité 3' des ARNm va également subir des modifications. Cette extrémité n'est pas le résultat direct de la terminaison de la transcription, elle est le résultat d'une coupure qui se produit dans le pré-messager environ une vingtaine de nucléotides au-delà d'une séquence spécifique (AAUAAA) appelée signal de polyadénylation. Après cette coupure, l'extrémité de l'ARN est modifiée par l'addition de 50 à 300 adénines qui forment une queue poly(A). Cette étape requiert une endonucléase pour couper l'ARN et une poly(A) polymérase pour synthétiser la queue poly(A) présente chez la plupart des ARNm des eucaryotes, mais qui n'existe pas dans le gène. Cependant, certains messagers, tel que les ARNm des histones, ne subissent pas cette modification. Dans leur cas, la formation de leur extrémité 3' dépend d'une structure secondaire (structure en épingle à cheveux) formée dans cette région.

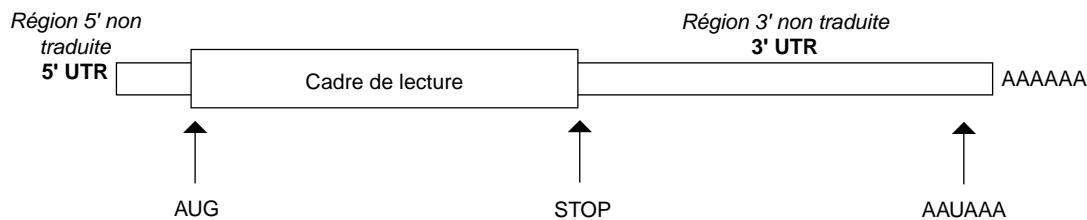
Les modifications post-transcriptionnelles (« Editing »)

Outre les opérations de maturation, quasi systématiques, qui viennent d'être évoquées, d'autres modifications post-transcriptionnelles des ARNm ont également été décrites dans la littérature. Ainsi, «l'editing», ou correction de l'ARN, concerne tout traitement d'un transcrit (addition, suppression, substitution d'un ou plusieurs nucléotides) autre que l'épissage, aboutissant à un ARNm dont la séquence diffère de celle du brin d'ADN sens. Ces modifications peuvent avoir une incidence considérable sur l'expression des gènes. Par exemple, chez l'Homme, le remplacement de C par U modifie la séquence de certains ARNm de l'apolipoprotéine B et provoque l'apparition d'un codon stop prématûré. De ce fait, la protéine synthétisée à partir des ARNm "modifiés" sera environ deux fois plus courte. Un même gène peut donc donner naissance à deux ARNm différents, traduits en apoB100 dans le foie ou en apoB48 dans l'intestin (Greeve *et al* 1993).

La structure des transcrits matures et leur sortie du noyau

Conséutivement à ces différentes étapes, l'ARNm mature d'un eucaryote est composé de séquences exoniques et contient trois parties (figure 5) : une région 5' non traduite portant une coiffe, une région traduite débutant par un codon d'initiation de la traduction AUG et se terminant par un codon stop, et une région 3' non traduite se terminant par une queue poly(A).

Chez les eucaryotes, la transcription et la traduction se déroulent dans deux compartiments différents puisque la première a lieu dans le noyau et la seconde dans le cytoplasme. La continuité entre le noyau et le cytoplasme est assurée par les pores nucléaires. Ces structures multiprotéiques, ancrées dans la membrane, se comportent comme des sites de transport actif permettant les échanges macromoléculaires entre le noyau et le cytoplasme. Il semble de plus en plus évident que cette étape de transport pourrait être limitante pour l'expression génétique, au moins pour certains gènes. De manière inattendue, de nombreuses études effectuées dans différents modèles expérimentaux indiquent

Figure 5. Structure d'un ARNm mature.

que les événements nucléaires (transcription, épissage) ne peuvent être dissociés du devenir cytoplasmique des ARNm (stabilité, traduction). Ces observations ont conduit Urlaub *et al* (1989) à proposer un modèle dans lequel le transport depuis le noyau vers le cytoplasme et l'épissage des ARNm sont couplés à leur traduction.

2.2 / La biosynthèse d'une protéine

La biosynthèse d'une protéine comporte deux étapes principales : la première correspond à la synthèse de la chaîne polypeptidique par décodage de l'ARNm - c'est l'étape de traduction ; la seconde consiste à maturer la chaîne polypeptidique néosynthétisée qui peut subir diverses modifications avant de devenir une protéine fonctionnelle.

a / La traduction

Les caractéristiques fondamentales de la traduction sont les suivantes : la traduction procède de 5' en 3' sur l'ARNm, et la protéine croît de l'extrémité aminée vers l'extrémité carboxylée. Il est utile de se rappeler que, dans la lecture d'une phrase, le sens importe également. Le début de la "phrase" - l'extrémité 5' de l'ARNm et l'extrémité aminée du polypeptide - est représenté, par convention, à gauche. La fin - extrémité 3' de l'ARNm et l'extrémité carboxylée du polypeptide - l'est à droite. Le message est donc lu de gauche à droite.

Le mécanisme de la traduction d'une séquence de codons en une chaîne polypeptidique est complexe

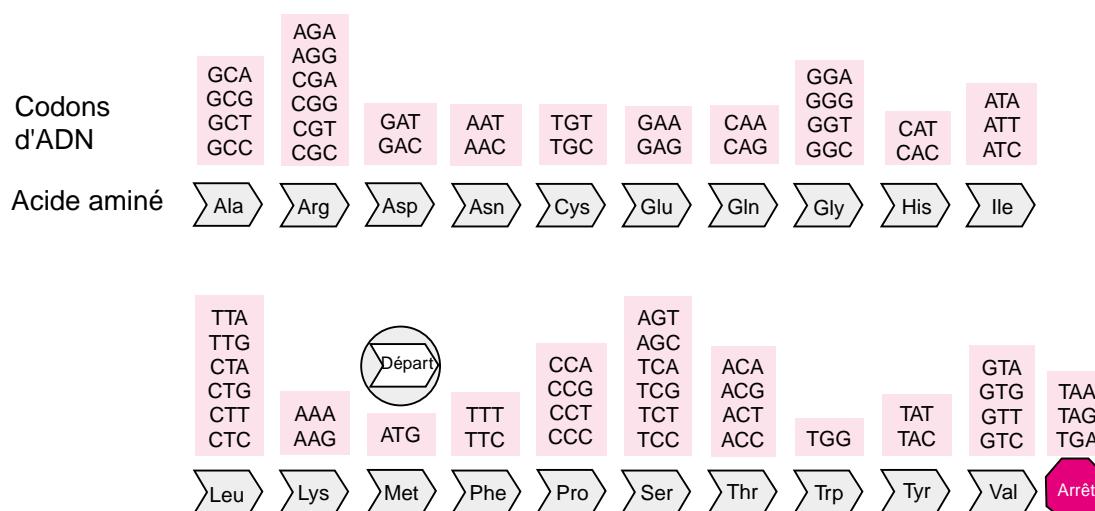
et implique un grand nombre d'étapes répétitives. La traduction s'accomplit dans le cytoplasme sur des petites particules appelées ribosomes. De concert avec les ARNt, les ribosomes constituent la machinerie qui convertit les séquences de nucléotides des ARNm en séquences d'acides aminés dans les protéines. Cependant, ce processus met également en jeu une variété d'intervenants.

Le code génétique

Dès 1964, le code génétique était totalement connu (figure 6). Chaque codon est constitué de trois nucléotides adjacents dans un brin d'ARNm. Soixante et un codons spécifient les 20 acides aminés et chacun d'entre eux n'en spécifie qu'un seul. Le triplet ATG (ou son équivalent AUG dans l'ARN) exerce deux fonctions. D'une part, il code pour l'acide aminé méthionine, d'autre part, il indique le début d'une séquence codante pour une protéine - le codon «départ». Trois triplets - TAG (UAG), TAA (UAA) et TGA (UGA) - ne spécifient aucun acide aminé mais signalent plutôt la fin d'une séquence codante (codon «arrêt» ou codon «stop»). On dit du code génétique qu'il est dégénéré parce que le même acide aminé peut être spécifié par plus d'un codon mais, point très important, il n'est nullement ambigu puisqu'un codon donné ne spécifie qu'un seul acide aminé.

Le mécanisme de traduction

La traduction de l'ARNm fait intervenir un grand nombre de macromolécules dont des ARN de transfert (ARNt), des enzymes et des ribosomes. Ces der-

Figure 6. Le code génétique.

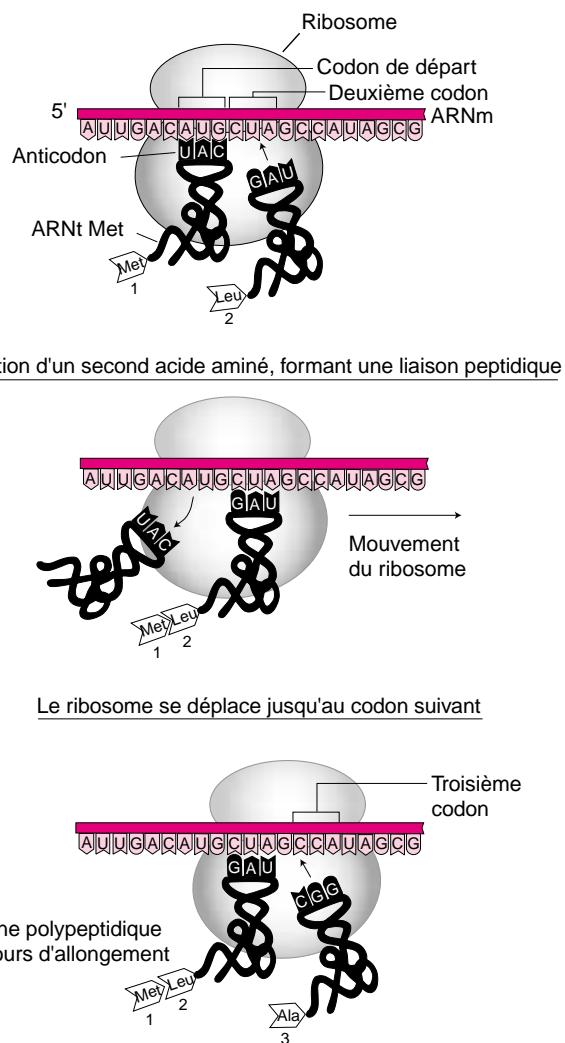
niers, chez les eucaryotes, comportent 2 sous-unités : 60S et 40S. La sous-unité 60S est constituée de 3 ARNr (5, 5,8 et 28S) et de 49 protéines. La sous-unité 40S est formée de l'assemblage d'un seul ARNr (18S) et de 33 protéines. Les ribosomes ont deux sites pour lier les ARNr : le site A (site acide aminé) où viendra l'ARNt porteur de l'acide aminé et le site P (site peptidique) pour l'ARNt porteur de la chaîne peptidique en cours d'elongation.

La traduction se déroule en trois étapes : l'initiation, l'elongation et la terminaison. L'initiation débute par la reconnaissance et la fixation de l'extrémité 5' (chapeau ou coiffe) de l'ARNm et un complexe "ARNt.Met initiateur + facteurs d'initiation eIF + sous-unité ribosomique 40S". L'ensemble va ensuite migrer dans le sens 5' vers 3' jusqu'au codon d'initiation AUG. Les nucléotides entourant le triplet initiateur sont très comparables d'un ARNm à l'autre. Kozak (1987) a défini, à partir de 699 ARNm de vertébrés, la séquence consensus suivante : GCCGCC(A/G)CCAUGG. Lorsque, dans la partie 5' de l'ARNm, deux triplets AUG sont présents, la comparaison des nucléotides entourant ces codons peut aider à prévoir lequel aura le plus de chance d'être initiateur. Une fois le site d'initiation atteint, les facteurs d'initiation sont libérés et la sous-unité ribosomique 60S s'associe à la 40S pour donner un ribosome fonctionnel. La phase d'elongation commence alors (figure 7). L'acide aminé suivant dans la chaîne polypeptidique est amené au site ribosomique de traduction par un ARNr dont l'anticodon reconnaît par appariement de bases le deuxième codon de l'ARNm. La première liaison peptidique est alors formée entre la méthionine et le deuxième acide aminé : la chaîne peptidique est commencée. Au fur et à mesure que les codons sont traduits, les acides aminés sont ajoutés à la chaîne naissante. Ce processus se répète jusqu'à ce que tous les codons de la séquence codante aient été traduits. Cette étape d'elongation fait intervenir des facteurs d'elongation dont eEF-1 et eEF-2 respectivement nécessaires à la fixation de l'amino acylARNt sur le site A et à la translocation du peptidyl-ARNt du site A vers le site P. Cette translocation se traduit par un glissement par triplet du ribosome sur l'ARNm. Enfin, lorsque la machinerie de traduction atteint le signal de fin de traduction (l'un des trois codons stop UAA, UAG ou UGA), la chaîne polypeptidique complète est libérée après l'hydrolyse, par une peptidyl-transférase, de la liaison ester unissant le dernier ARNr au dernier acide aminé de la chaîne. A cette dernière étape de la traduction, la fixation du facteur protéique de terminaison sur le codon stop provoque la libération du dernier ARNr et de la chaîne peptidique. Le ribosome se dissocie en deux sous-unités qui pourront recommencer une nouvelle lecture d'ARNm.

Plusieurs ribosomes peuvent être engagés simultanément dans la traduction du même ARNm, d'où le nom de polysome donné au complexe ARNm-ribosome. En effet, dès qu'un ribosome a franchi la première étape dans la traduction de l'ARNm, il s'éloigne du codon de départ AUG et un deuxième ribosome peut recommencer l'opération à ce même codon. Et ainsi de suite, de sorte qu'un troisième ribosome, un quatrième, voire d'autres encore, s'engagent l'un après l'autre dans la traduction du messager au même site de départ, amorçant chacun l'assemblage d'une chaîne polypeptidique nouvelle. De cette façon, la même molécule d'ARNm peut donner

naissance à plusieurs polypeptides identiques en très peu de temps. Les polypeptides néosynthétisés peuvent commencer à se replier en structures actives avant même que leur synthèse ait atteint son terme.

Figure 7. Le démarrage de la traduction d'un ARNm (d'après Berg et Singer 1993).



b / Les modifications post-traductionnelles

La structure primaire de la chaîne polypeptidique peut subir diverses modifications pendant et/ou après sa synthèse. Dans 95 % des cas, la protéine finale a une structure plus ou moins modifiée par rapport à celle de la chaîne néosynthétisée. Ces modifications covalentes ont un rôle biologique souvent crucial car elles peuvent affecter les propriétés intrinsèques du polypeptide sur le plan de la conformation, stabilité, assemblage, activité, régulation, reconnaissance, transport, localisation, exportation, etc. Les types de modifications sont nombreux, mais nous n'en présenterons qu'un seul. La caséine kappa est synthétisée dans le réticulum endoplasmique des cellules épithéliales mammaires des ruminants en lactation (Mepham *et al* 1993). Au cours de la phase de traduction, son peptide signal ou "code d'adressage" (correspondant aux 21 premiers acides aminés N-terminaux de la préprotéine) est excisé par une peptidase signal située du côté interne de la membrane du réticulum endoplasmique. Ensuite, cette caséine poursuit sa maturation au cours de

son transport vers les vésicules de sécrétion via l'appareil de Golgi. Durant ce transit cellulaire, elle subit les étapes de *O*-phosphorylation et de *O*-glycosylation. *In fine*, cette caséine est trouvée dans le lait, associée avec les trois autres caséines (alpha-S1, alpha-S2 et beta) sous forme de particules sphériques appelées micelles.

3 / La régulation de l'expression des gènes

Les quantités de protéines synthétisées par une cellule ne sont pas nécessairement fixes, mais varient en fonction des besoins de la cellule. Cette variation du niveau d'expression d'un gène peut résulter d'une régulation de chacune des étapes qui viennent d'être décrites (cf figure 2). Par exemple, un système de régulation peut être associé à la transcription (activation d'un gène par la transduction d'un signal hormonal), à l'épissage (par exemple, régulation du déterminisme du sexe chez la drosophile), à la stabilité des ARNm (séquences stabilisatrices dans la région 3' UTR et/ou longueur

de la queue poly(A) des ARNm).. De ce fait, il est concevable qu'une variabilité génétique des séquences jouant un rôle important dans ces régulations puisse altérer l'expression du gène associé. Les articles de la deuxième partie de cet ouvrage présentent des exemples de polymorphismes associés à des variations de l'expression des gènes ayant une incidence sur un caractère d'intérêt. De même, le polymorphisme génétique peut intervenir sur les étapes se déroulant dans le cytoplasme. Ainsi, le site de glycosylation de la mucine épithéliale MUC1 est répété un nombre variable de fois (VNTR = Variable Number Tandem Repeat) et cette différence de degré de glycosylation semble avoir une incidence sur la protection des épithéliums sécrétant cette mucine.

En conclusion, après avoir brièvement présenté les grandes étapes de l'expression des gènes, nous venons d'évoquer la complexité de leur régulation. De fait, on conçoit aisément l'importance de l'étude de la variabilité de l'expression des gènes d'intérêt, tant au niveau des ARNm qu'au niveau des protéines.

Références

- Berg P., Singer M., 1993. Comprendre et maîtriser les gènes. Le langage de l'hérédité. Editions Vigot, Paris.
- Etienne J., 1998. Biochimie génétique - Biologie moléculaire. Editions Masson, Paris.
- Greeve J., Altkemper I., Dieterich J.H., Greten H., Windler E., 1993. Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins. Journal of Lipid Research, 34, 1367-1383.
- Kozak M., 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Research, 15, 8125-8148.
- Lewin B., 1994. Genes V. Oxford University Press, UK.
- Maftah A., Julien R., 1999. Biologie moléculaire. Editions Dunod, Paris.
- Maniatis T., Reed R., 1987. The role of small nuclear ribonucleoprotein particules in pre-mRNA splicing. Nature, 325, 673-678.
- Mepham T.B., Gaye P., Martin P., Grosclaude F., 1993. Biosynthesis of milk protein. In : Fox (eds) Advanced dairy chemistry, Vol. 1. Elsevier Science Publishers, Amsterdam.
- Urlaub G., Mitchell P.J., Ciudad C.J., Chasin L.A., 1989. Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. Molecular Cellular Biology, 9, 2868-2880.