

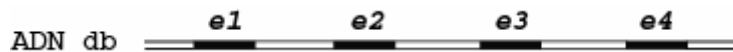
ANNEXE XV SYSTÈMES DE TRANSCRIPTION ET DE MATURATION DIFFÉRENTIELLES DE L'ARN

Les eucaryotes utilisent différents mécanismes pour enrichir l'information génétique contenue dans leurs chromosomes. À partir d'un même gène, ils synthétisent plusieurs protéines. En général, ces protéines ne diffèrent que partiellement. Par exemple, les trois produits du gène *shc1*** de la souris (P46, P52 et P66) ont simplement des longueurs différentes.

Mais il y a des exceptions. C'est le cas pour les deux produits du gène *ink4a***, dont la séquence n'a rien de commun. Le système de traitement de l'information génétique nous intéresse pour une autre raison : il fonctionne quelquefois de manière incorrecte à la suite de mutations, et peut causer des maladies à évolution catastrophique, comme celle de Hutchinson-Gilford (*annexe XVIII*).

XV.A. STRUCTURE DES GÈNES

Plus de 95 % des gènes humains spécifiant des protéines sont morcelés. L'ADN qui les forme juxtapose deux types de séquences, appelées exons (*e*) et introns, représentés ci-dessous par des espaces pleins et vides :

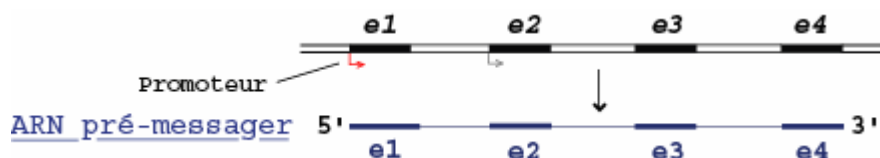


Les exons sont inclus dans le produit final du gène (l'ARN messager), tandis que les introns en sont exclus.

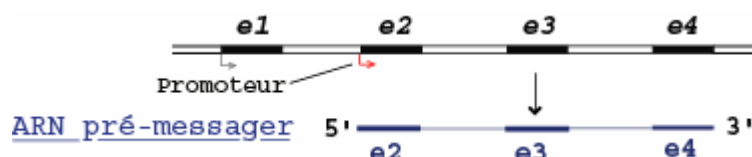
XV.B. TRANSCRIPTION DIFFÉRENTIELLE

En général, la transcription démarre au niveau du promoteur situé à proximité du premier exon et se poursuit jusqu'au dernier, créant ainsi un produit primaire, appelé ARN pré-messager, qui réunit l'ensemble des exons et des introns.

Mais certains gènes possèdent plusieurs promoteurs optionnels, Si la transcription commence au niveau du premier exon, elle crée un ARN pré-messager long :



Si elle commence au niveau d'un promoteur interne, elle crée un ARN messager amputé d'un ou de plusieurs exons :

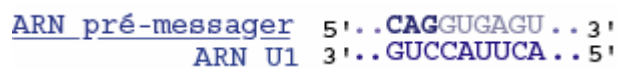


XV.C. ÉPISSAGE DIFFÉRENTIEL

À l'intérieur même du noyau, l'ARN pré-messager subit une série de remaniements qui concernent les deux extrémités ainsi que l'intérieur de la molécule. Les remaniements internes consistent à éliminer les introns et à raccorder les exons en une chaîne continue, mais plus courte. Cette opération est dénommée épissage. Un ARN pré-messager comportant n introns doit subir $2n$ coupures, suivies de n épissures, ou ligations. Les coupures/ligations ont lieu de manière coordonnée, grâce à un complexe associant cinq ARN guides* (U1, U2, U4, U5 et U6) et environ 300 protéines. Elles se font en des endroits précis de l'ARN pré-messager, appelés sites d'épissage, formés par de courtes séquences, dites consensus, qui chevauchent les jonctions exons/introns. Chez les animaux, ces séquences comportent neuf unités du côté 5' des introns et quatre du côté 3', chaque position étant presque toujours occupée par une même base ou par deux bases optionnelles (A ou C ; A ou G ; U ou C) :



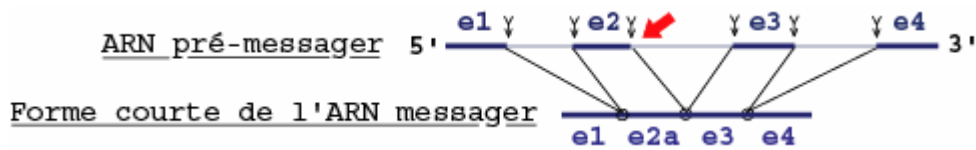
La séquence consensus de la jonction exon/intron 5' est reconnue par une séquence interne de l'ARN U1, qui sert de guide pour introduire une coupure entre les séquences **CAG** et GUGAGU :



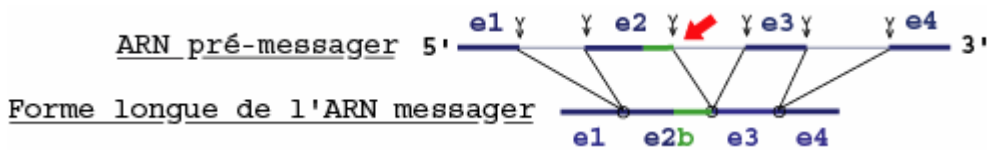
Le système d'épissage commet parfois des erreurs. Chez l'homme, on estime qu'environ 10 % des mutations provoquent un épissage incorrect de l'ARN pré-messager, ce qui entraîne la synthèse de protéines malformées.

Le système d'épissage le plus simple produit à partir d'un même ARN pré-messager un seul messager, que l'appareil de traduction utilise comme matrice pour agencer dans un ordre précis les acides aminés qui forment une protéine. Mais il arrive que le système crée, suivant les cellules ou les circonstances, des ARN différents. Chez les mammifères, la grande majorité des gènes spécifiant des protéines peuvent engendrer plusieurs produits.

Plusieurs mécanismes d'épissage différentiel ont été découverts, dont un seul nous concerne ici. Un exon peut posséder deux sites d'épissage, situés à l'intérieur sur les bords. Si c'est le premier site qui est utilisé, on obtient un ARN messager court :



Si c'est l'autre site qui est utilisé, on obtient un ARN messager plus long, contenant une séquence additionnelle, qui fait défaut dans l'ARN messager court :



En général, le système d'épissage raboute les exons en respectant la phase de lecture, c'est-à-dire la succession de groupes de trois bases (codons) commandant l'insertion d'un acide aminé dans une protéine. De la sorte, les protéines spécifiées par deux ARN messagers produits par un même gène diffèrent par une séquence interne, figurées ci-dessous par des lettres représentant des suites d'acides aminés :

<u>Forme courte de la protéine</u>	ABCDEFGHIJKLMN e1 e2b e3 e4
<u>Forme longue de la protéine</u>	ABCDEFGHIJKLMN e1 e2a e3 e4

Mais cette règle n'est pas absolue. Par exemple, le gène *ink4a*** peut engendrer à partir de deux promoteurs différents un ARN pré-messenger court et un ARN pré-messenger long, que le système d'épissage raccorde suivant deux phases de lecture décalées, si bien que les deux produits de ce gène, (les protéines INK4a et Arf**) n'ont aucune séquence en commun (*fig. XV*).

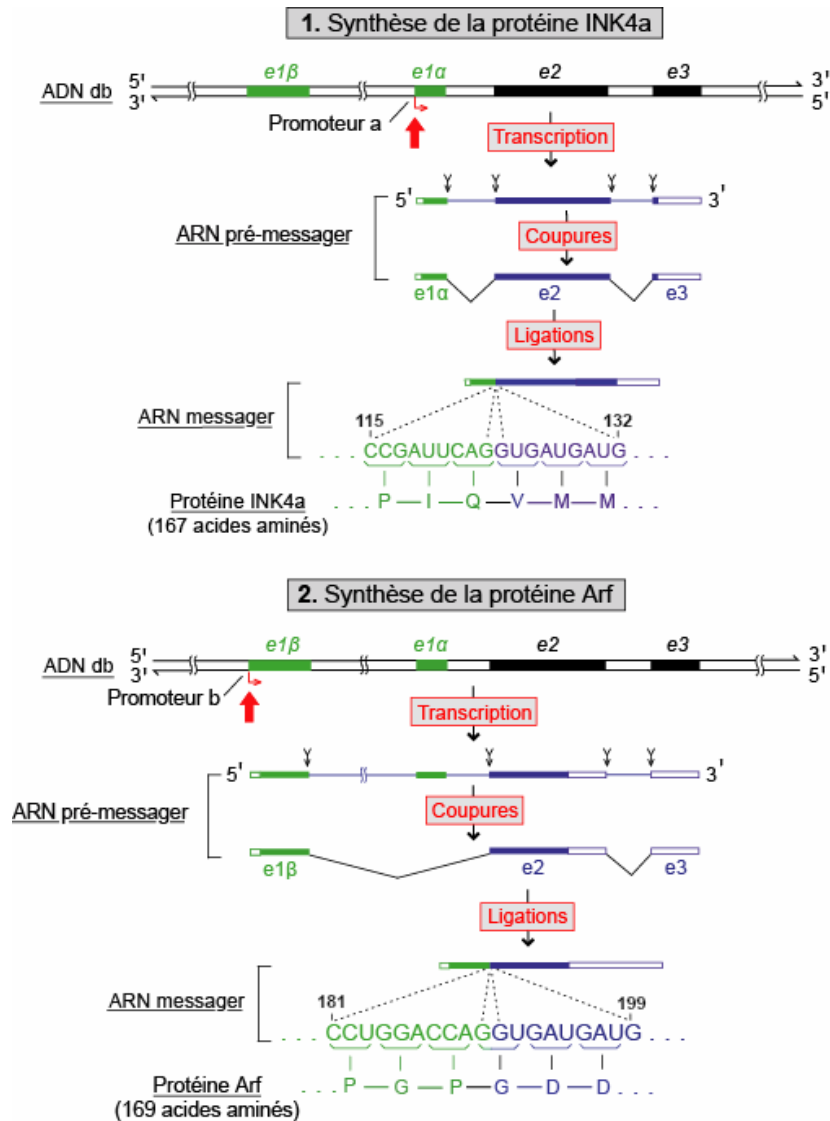


Fig. XV. Transcription et épissage différentiels de l'ARN *ink4a* de souris.

Le gène *ink4a* comporte deux promoteurs (a et b), trois introns et quatre exons, dont les deux premiers (*e1α* et *e1β*) sont optionnels.

1. Pour synthétiser la protéine INK4a, les cellules utilisent le promoteur interne (a). Seuls les trois derniers exons (*e1α*, *e2* et *e3*) sont transcrits. Le système d'épissage raccorde l'exon 1α à l'exon 2 de l'ARN pré-messenger suivant la même phase de lecture (représentée par des rectangles pleins). L'exon 1α spécifie les acides aminés n° 1 à 41 de la protéine, tandis que les exons 2 et une partie de l'exon 3 spécifient les acides aminés n° 42 à 167. Le codon STOP se trouve au début de l'exon 3.

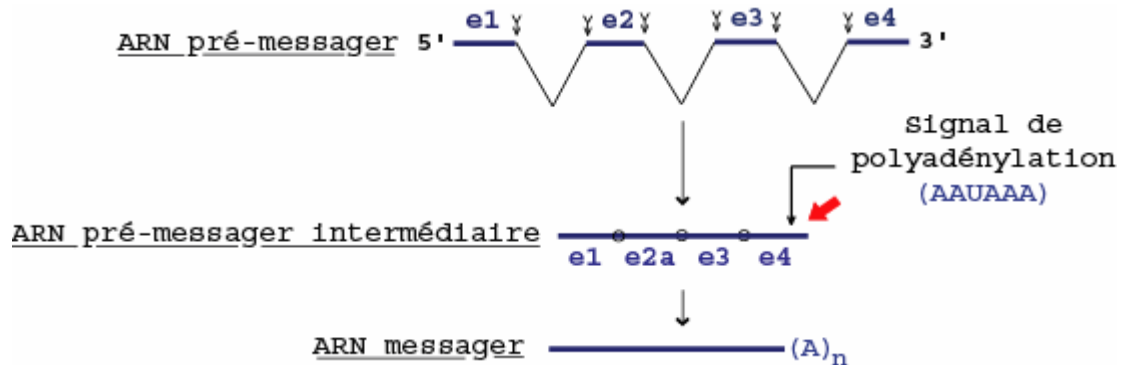
2. Pour synthétiser la protéine Arf, les cellules utilisent le promoteur externe (b). Tous les exons sont inclus dans l'ARN pré-messenger, mais le système d'épissage élimine l'exon 1α en même temps que les introns. Le rabotage entre les exons 1β et 2 se fait de telle manière qu'ils sont lus suivant des phases différentes. Cette particularité vient du fait que l'exon 1β contient 190 bases, ce qui n'est pas un multiple de trois. Les 189 premières bases spécifient les acides aminés n° 1 à 63 de la protéine. La dernière (G) est incluse dans un codon GGU, qui ouvre une phase de lecture décalée d'une unité, si bien que la séquence d'acides aminés spécifiée par l'exon 2 est complètement changée. Suivant cette phase de lecture, le codon STOP se trouve à l'intérieur de l'exon 2. Il interrompt la synthèse de la protéine au niveau de l'acide aminé n°169.

La séquence des six acides aminés spécifiés par les bases chevauchant les jonctions *e1α* - *e2* et *e1β* - *e2* est présentée suivant le code à une lettre (D = aspartate ; G = glycine ; I = isoleucine ; M = méthionine ; P = proline ; Q = glutamine ; V = valine).

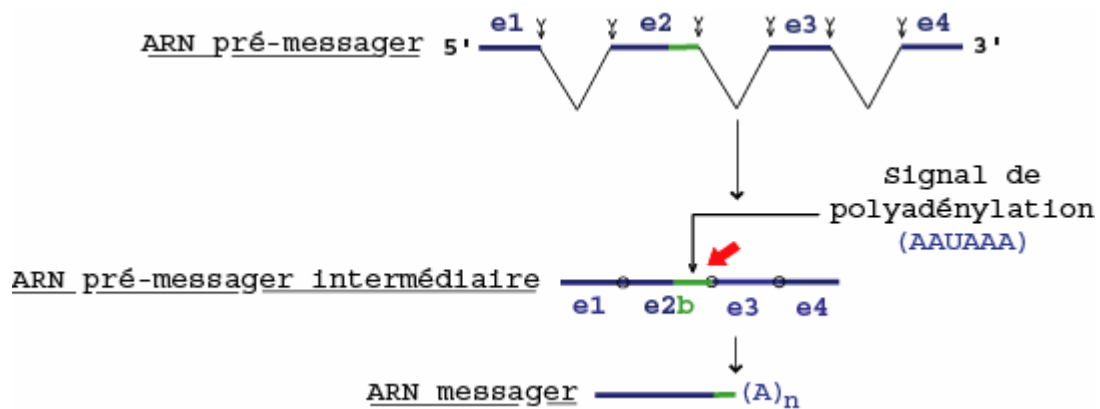
D'après Quelle DE *et al. Cell* 1995 ; **83** : 993-1000.

XV.C. POLYADÉNYLATION DIFFÉRENTIELLE

L'extrémité 3' des ARN messagers porte une queue poly(A) plus ou moins longue, qui contrôle sa stabilité ainsi que l'efficacité de sa traduction. Le processus de polyadénylation est mené à bien par un complexe enzymatique qui reconnaît un signal proche de l'extrémité 3' de l'ARN pré-messager, Le complexe coupe l'ARN à quelque 20 bases du côté 3' par rapport au signal, puis ajoute une série d'adénylates. Le signal est généralement situé dans le dernier exon, si bien que tous les exons présents dans l'ARN pré-messager se retrouvent dans l'ARN messager :



Si l'ARN pré-messager subit un épissage différentiel, il arrive qu'un signal de polyadénylation apparaisse dans un exon interne, ce qui entraîne l'ablation de tous les exons situés en aval :



De la sorte, la version courte d'un ARN pré-messager peut donner naissance à un ARN messager et à une protéine plus courts que la version longue de ce même ARN. C'est par ce mécanisme que sont produites les lamines A et C (*annexe XVIII.B*).

BIBLIOGRAPHIE

Corcus L, Solier S. Épissage alternatif, pathologie et thérapeutique moléculaire. *Med Sci* 2005; **21**: 253-60.

Quelle DE, Ashmun RA, Hannon GJ, Rehberger PA *et al.* Cloning and characterization of murine p16^{INK4a} and p15^{INK4b} genes. *Oncogene* 1995; **11**: 635-45.