

Structure et fonction des Gènes

Le Gène: Un gène est une partie d'un chromosome formant une unité d'information génétique. Un gène détermine la mise en place et la transmission d'un caractère observable. Unité fonctionnelle et physique élémentaire de l'hérédité qui transmet l'information d'une génération à la suivante.

Un fragment d'ADN, constitué d'une région transcrite et de séquences régulatrices.

Le génome : est l'ensemble des gènes portés par les chromosomes d'une espèce.

- Le génome (séquence complète d'ADN) humain est composé de 3,3 milliards de paires de bases, le nombre des gènes est estimé à environ 60.000.
- La taille des gènes est très variable, le plus petit gène comprenant quelques centaines de Paires de bases, alors que le plus grand peut avoir plus d'un million de paires de bases de longueur; un kilobase (1kb) équivaut à 1000 paires de bases (pb).
- Le génome contenant le plus petit nombre de gène est celui de *Mycosplasma genitalium*, une bactérie parasite intracellulaire appartenant au groupe des Mollicutes. Il contient environ 520 gènes dont 480 codent pour des protéines/
- La plante herbacée *Paris japonica* a le plus grand génome connu : il comporte près de 150 milliards de paires de bases, soit près de 50 fois la taille du génome humain.

Ressemblance de séquences

- –Familles de gènes si > 50% de similitude entre séquences

Gènes identiques

- –Cluster ARN ribosomal type 5S (2000 gènes sur chromosome 1)

Gènes fonctionnellement identiques

- –Histones

Gènes fonctionnellement similaires

- –Globines

Gènes fonctionnellement apparentés

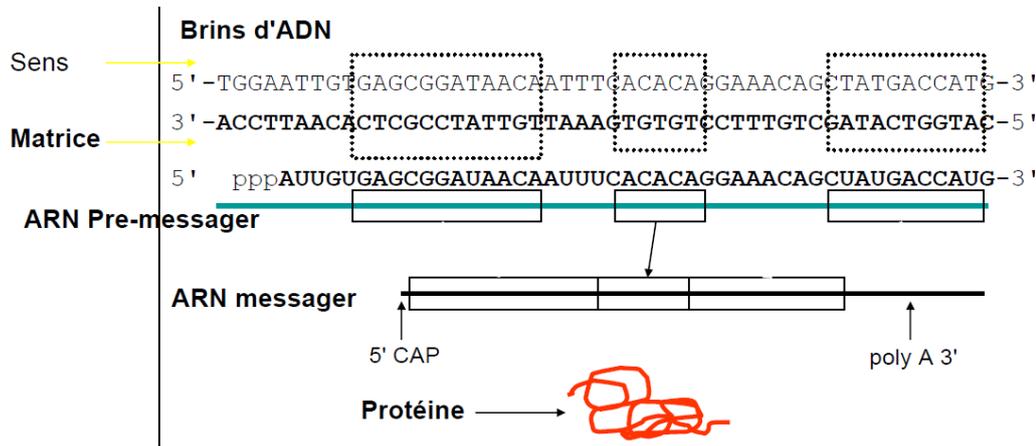
- –Récepteurs nucléaires des hormones stéroïdes

Organisation des familles de gènes :

- Clusters
- Dispersés
- Et toute la gamme de variation entre clusters et dispersés.

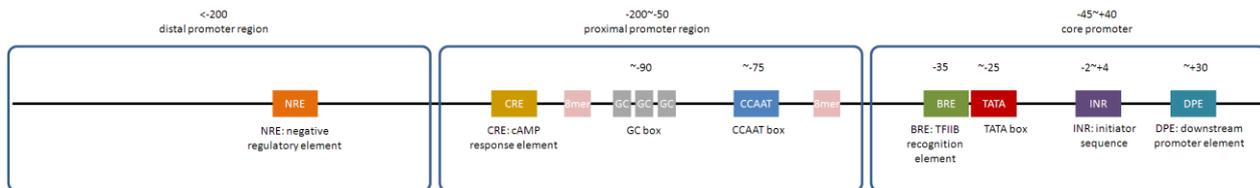
Structure des Gènes : un gène est défini comme un enchaînement de désoxyribonucléotides (dit aussi séquence), c'est-à-dire comme une portion d'acide désoxyribonucléique (séquence d'ADN), qui peut être transcrit en acide ribonucléique (ARN). S'il est ensuite traduit en protéine, la séquence est dite « codante ». La plupart du temps, un gène est composé d'un promoteur, des exons et des introns.

- les exons sont les fragments d'un ARN primaire qui se retrouvent dans l'ARN cytoplasmique après épissage (fragments d'ARN primaire éliminés au cours de l'épissage)(Les **exons** sont les régions du gène qui contiennent l'information codante..
- Un intron est une portion de gène, le plus souvent non codante, qui ne se retrouve pas dans l'ARN cytoplasmique après épissage (Les **introns** sont des segments d'ADN localisés au sein du gène, transcrit en ARN, puis épissés avant que l'ARN messager ne soit traduit en protéine.) Il s'oppose à l'exon.
- La taille moyenne des gènes transcrits est de 27.900 pb. La fraction codante couvre généralement 1.340 pb, répartis en 8-9 exons, et la taille moyenne des exons est de 145 pb.
- La taille moyenne des introns est d'environ 3.300 pb. Plus de 35% des gènes présentent un épissage alternatif.
- Le gène de plus grande taille est celui codant la dystrophine, qui s'étend sur 2,4 Mb.
- Le gène de la titine présente d'autres records: il code le plus grand messenger humain connu (80.780 pb), le plus important nombre d'exons (178), parmi lesquels figure le plus grand exon connu (17.106 pb).
- La région 5' et la région 3', situées aux deux extrémités du gènes ont des régions non codantes (non traduites).
- Les gènes sont parfois rassemblés en groupes ,les gènes qui partagent une même fonction étant proche, les uns des autres.
- Les **pseudogènes** sont des séquences d'ADN qui ont certaines des caractéristiques de structure des gènes exprimés ont probablement été fonctionnels, mais ils ont acquis au cours de l'évolution, un nombre de mutations inactivatrices, les empêchant de produire une protéine.
 - Les promoteurs initient et contrôlent la transcription.
 - Les promoteurs sont des séquences d'ADN au niveau desquelles s'assemblent le complexe d'initiation.
 - La transcription de chaque gène démarre à son promoteur particulier.
 - L'efficacité de l'initiation varie selon que la séquence du promoteur est plus ou moins proche (grande efficacité) ou éloignée (faible efficacité) de la séquence consensus.



Le Promoteur :

- Les promoteurs initient et contrôlent la transcription.
- Les promoteurs sont des séquences d'ADN au niveau desquelles s'assemblent le complexe d'initiation.
- La transcription de chaque gène démarre à son promoteur particulier.
- L'efficacité de l'initiation varie selon que la séquence du promoteur est plus ou moins proche (grande efficacité) ou éloignée (faible efficacité) de la séquence consensus.



Régions non traduites

Les **régions non traduites**, ou **régions UTR** (*untranslated regions*), sont les parties de l'ARNm issues de la transcription de l'ADN qui ne sont pas traduites en protéines.

On distingue : les régions 5'UTR et 3'UTR.

Les régions UTR — que l'on retrouve tant chez les eucaryotes que chez les procaryotes, virus inclus — ont une grande importance dans la régulation de l'expression d'un gène :

- Le transport/localisation de l'ARNm ;
- La régulation de sa stabilité ;
- La régulation de sa traduction.

Séquences répétées

. Séquences localisées répétées en tandem :

- Séquences microsatellites : des séquences de 1 à 5 pb sont répétées un grand nombre de fois, répétitions longues de plusieurs Kb, il y a plus de 10 000 zones de répétition dans le génome
- Séquences minisatellites : des séquences de 15 à 25 pb sont répétées un grand nombre de fois, répétitions inférieures à 150 pb

eucaryotes) et consiste à copier des régions dites codantes de l'ADN et de les transcrire en molécules d'ARN.

La synthèse de l'ARN se fait par polymérisation de ribonucéotides triphosphate dans le sens 5' > 3', sens opposé à celui du déroulement du brin matrice (3' > 5').

- La chromatine doit au préalable avoir été décompactée (euchromatine) pour permettre à la machinerie protéique d'accéder à l'ADN
- L'enzyme responsable de la transcription est l'ARN polymérase.
- Il existe 3 types d'ARN polymérase chez les eucaryotes: I, II, III
- L'ARN polymérase est accompagnée par des facteurs de transcription généraux (protéines) : TF_I, TF_{II}, TF_{III}... ce qui constitue un complexe protéique constitué de 8 à 14 sous unités.

I - Les ARN

Les ARN sont des polyribonucléotides. Par rapport à l'ADN le T est remplacé par le U (même si dans certain ARNt on peut trouver du T). On les trouve dans le noyau et le cytoplasme. Un ARN est monocaténaire, mais sa chaîne peut se replier, ie former une structure secondaire stable (épingle à cheveu) en formant des liaisons hydrogène entre les bases.

Les ARN majoritaires de la cellule sont :

- ARNr 83%
- ARNt + petits ARN 15%
- ARNm 2%.

L'ARN est un produit de la transcription de l'ADN par l'**ARN polymérase ADN dépendante**.

1 - Les ARNt

Ils sont présents dans le cytoplasme. Ils transfèrent les acides aminés (aa) sur les chaînes protéiques en constitution au niveau du ribosome. Ils ont un rôle **adaptateurs** en amenant les aa à la bonne place de la séquence polypeptidique. Il existe **70 ARNt différents**, mais il n'y a que 20aa et 64 codons, il y a donc plus d'ARNt que de aa ou de codons. Ils sont composés d'une seule chaîne polynucléotidique.

2 - Les ARNr

Ils forment quasiment de suite après leur synthèse des ribosomes en s'associant à des protéines ribosomales (ribonucléoprotéines).

	PROCARYOTES	EUCARYOTES
Constante de sédimentation	70s	80s
Masse particulaire	2,8.10 ⁶	4,5.10 ⁶
Dissociation (quand Mg ²⁺ en dessous de 1mMol)	50 + 30	60 (ARN 28s, 5s et 5'8s + 45 prot) + 40 (ARN 18s + 33 prot)
%ARN	63	50
%Protéines	37	50

Les ARNm

Ils sont minoritaires dans la cellule. Il s'agit de la « copie » de l'information génétique contenu dans l'ADN. Ils sont synthétisés rapidement et leur demi-vie est courte (quelques minutes chez les procaryotes et quelques heures chez les eucaryotes) ; et leurs tailles sont très hétérogènes.

La transcription chez les eucaryotes.

Il existe plusieurs ARN polymérase.

Classe de l'enzyme	Type d'ARN transcrit
I (A)	ARNr
II (B)	ARNm (précurseur)
III (C)	ARNt

Chez les eucaryotes les ARN polymérase sont incapables à elles seules, d'initier la transcription. Il faut pour cela tout un assortiment de protéines appelées facteurs de transcription généraux qui vont se lier au promoteur. Cet assemblage fournit différentes possibilités de régulation de l'initiation de la transcription. La présence de ces facteurs de transcription généraux est nécessaire au fonctionnement de l'ARN polymérase.

Dans la plupart des cas, plusieurs facteurs de transcription s'assemblent sur l'ADN au niveau du promoteur et recrutent l'ARN polymérase. Il existe un grand nombre de facteurs de transcription.

Le promoteur des eucaryotes est composé de plusieurs séquences permettant de définir entre autre : le début de transcription, la fréquence, le type cellulaire ou le type d'environnement nécessitant cette transcription... Il existe une boîte TATA ressemble beaucoup à celle des procaryotes, fonctionnellement apparenté (ie qui définit où débute la transcription) et situé en -32 du +1 de transcription. D'autres séquences, tel que les boîtes GC et CAAT, +/- en amont définissent la fréquence de transcription. Toute variation de séquence ou de positionnement de ces éléments entraîne de forte variation de fréquence de transcription. Ces éléments agissent en *cis* car ils sont sur la même molécule d'ADN que le gène contrôlé, alors que les facteurs protéiques agissent en *trans*.

Il existe une troisième classe de séquences régulant le taux de transcription : élément **stimulateur** ou **extincteur** = **enhancer** ; ils peuvent être situés en amont ou en aval du gène, même très loin. Les éléments régulateurs hormonaux, certains métaux, des chocs thermiques ou encore certaines toxines agissent comme tel.

Mécanismes généraux de la Transcription par l'ARN polymérase II

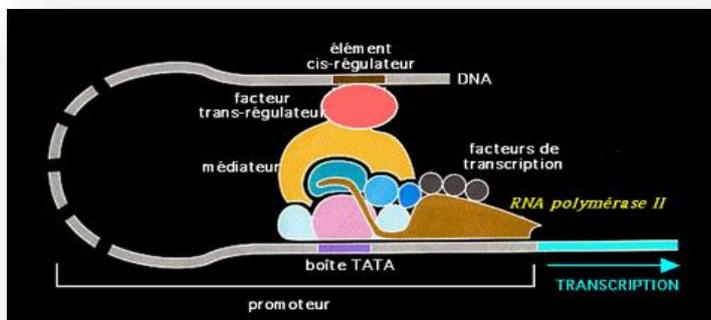
La transcription des gènes se déroule selon trois phases principales : phases d'initiation, d'élongation et de terminaison.

Phase d'initiation

L'initiation de la transcription par l'ARN polymérase II est assurée par des facteurs transcriptionnels généraux dénommés **TFII A-B-C-D-E-F- et H**. Ces facteurs s'assemblent sur une région située en amont de

l'ADN à transcrire qui porte le nom de promoteur. La plupart des promoteurs de l'ARN polymérase II contiennent une région riche en nucléotides A et T appelée **TATA box** (boîte TATA). Elle contient la séquence consensus TATAT/AAT/AA. Seul le facteur **TFIID** (qui est en fait un complexe multiprotéique dont l'élément de liaison à la TATA box porte le nom de **TBP** pour "TATA Box Binding Protein") reconnaît spécifiquement la boîte TATA et s'y lie. La liaison de TBP à l'ADN est peu usuelle car elle se lie au niveau du petit sillon de l'ADN, alors que toutes les autres protéines se liant à l'ADN, se lient au niveau du grand sillon. De plus son angle de liaison à l'ADN (80°) provoque une **distorsion** de celui-ci. Cette structure, en changeant l'organisation spatiale autour de la boîte TATA va permettre aux facteurs transcriptionnels et à l'ARN polymérase d'être plus étroitement associés qu'ils ne l'auraient été sur de l'ADN linéaire. En fait la liaison à la boîte TATA déroule l'hélice d'environ 1/3 de tour. Cet assemblage entraîne la formation d'un complexe d'activation qui permet la liaison de l'ARN polymérase II. L'ensemble forme le complexe d'initiation de la transcription.

La transcription débute à 20-30 paires de base (pb) en aval de la boîte TATA, au site d'initiation de la transcription (ATG), qui par convention est appelé +1. Le complexe d'initiation de la transcription va catalyser la formation de la première liaison phosphodiester entre les deux premiers nucléotides de l'ARN messenger (ARNm). L'ARN polymérase II se déplace ensuite le long de l'ADN en ouvrant une partie de la molécule d'ADN par déroulement de la double hélice sur une courte distance en formant une boucle de transcription.



Phase d'élongation

La boucle de transcription se déplace dans le sens 3'-5' du brin matriciel et la chaîne d'ARNm s'allonge dans le sens 5'-3'.

L'élongation de la molécule d'ARNm se fait par l'appariement des bases complémentaires et par l'addition successive de nucléotides 5' triphosphates. Cette élongation nécessite des facteurs supplémentaires appelés facteurs d'élongation nécessaires au déplacement de l'ARN polymérase II. L'ADN lu se rebobine immédiatement après la lecture.

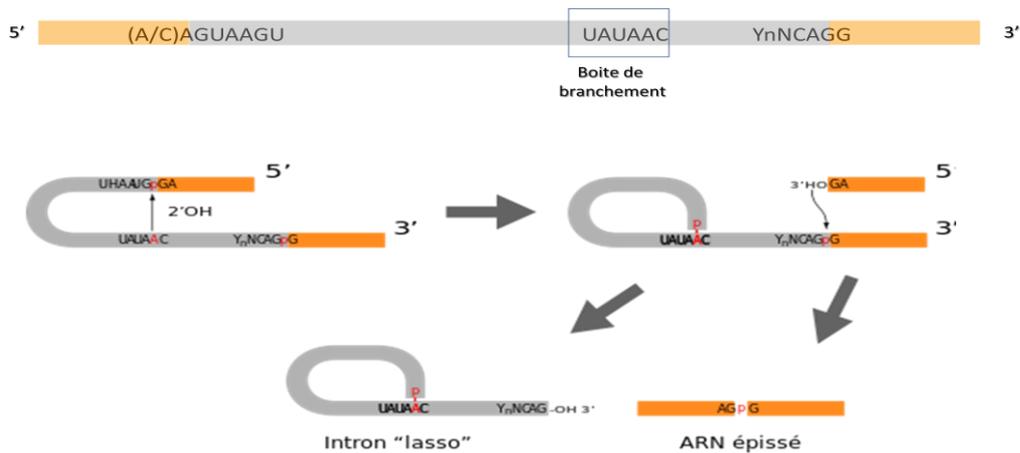
Phase de terminaison

- La terminaison est assurée par des signaux spécifiques dont le signal de polyadénylation 3'-TTATTT-5' suivie dans la plupart des gènes d'un autre signal 3'-ATACAAAC-5'.

- La polymérase continue sa transcription un peu après ce motif puis est libérée sous l'action de divers facteurs.

Maturation de l'ARNm

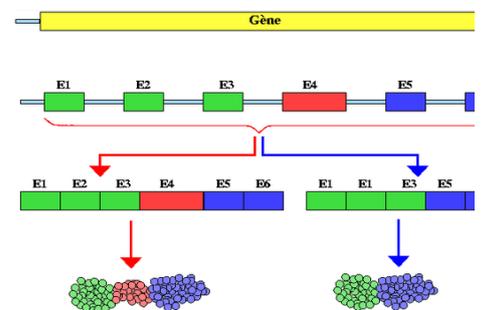
- L'épissage dépend de la présence de séquences signal dans le pré-ARNm :
 - Dans la quasi-totalité des gènes, les deux premiers nucléotides de l'extrémité 5' des introns sont GU et les deux derniers à l'extrémité 3' sont AG.
 - On trouve une autre séquence signal dans les introns appelée la séquence de branchement ~40 nucléotide en amont de l'extrémité 3'.
- L'épissage se déroule dans le noyau et se résume en 2 étapes :
 - clivage de l'intron à son extrémité 5' et son attachement à la séquence de branchement.
 - les exons sont mis bout à bout au même moment du clivage de l'extrémité 3' de l'intron et sa libération sous forme de « lasso ».
- La catalyse de l'épissage est assurée par un complexe enzymatique appelé Splicéosome, il est formé par de petites ribonucléoprotéines (snRNP).



Épissage alternatif :

- Uniquement chez les eucaryotes.
- Les signaux d'épissage :
 - Fort (constitutifs)
 - Faibles (épissage alternatif)
- Estimations :
 - 70% des gènes humains subissent un épissage alternatif
 - en moyenne, un gène donne naissance à 4 variantes issus

d'un tel épissage



- Mise en place de la coiffe (capping) :

Le coiffage de l'ARN implique une modification de l'extrémité 5' du transcrit primaire. L'extrémité 5' est coiffée par addition d'un nucléotide atypique. Dès la transcription il y a ajout d'une **méthyl 7 guanosine** par une transférase (la guanyltransférase et la méthyltransférase). Elle est rajouté du côté 5' comportant un triphosphate. **5' me7GpppApXpX3'**. Cette coiffe participe à **l'initiation de la traduction** et **empêche la dégradation** de l'ARN par une exonucléase 5'3' (ce qui augmente sa demi-vie). C'est aussi un signal autorisant le ribosome à reconnaître le début de la molécule d'ARNm.

- L'extrémité 3' polyadénylation

Une queue polyA est ajoutée grâce à la polyA polymérase. Les seuls ARN échappant à cette maturation sont ceux codant pour les histones. La polyadénylation **protège l'extrémité 3'** de l'ARN des exonucléases. Cette queue polyA est une aubaine pour les biochimistes car elle est très utile pour séparer des ARNm d'autres ARN grâce à des colonnes d'oligo-dT (ou U) Sépharose. Elle permet la synthèse d'ADNc en appariant un fragment dT qui sert d'amorce à la reverse transcriptase.

Traduction :

C'est le mécanisme par lequel le flux d'information va passer de la forme acide nucléique (alphabet à 4 lettres) à la forme protéine (alphabet à 20 lettres) selon un code universel ou presque.

LA TRADUCTION CHEZ LES EUCARYOTES

Comme pour la réplication, la transcription, la transcription inverse et la réplication de l'ARN, la lecture de la matrice (ici l'ARNm) se fait de 3'→5' et l'ensemble se passe en 3 étapes : initiation, élongation, terminaison. Ces 3 étapes sont catalysées par des facteurs.

Les sous-unités du ribosome (40S et 60S) s'assemblent au niveau du codon initiateur (5'AUG3'), c'est l'initiation. Le tRNA^{met} se fixe au niveau du site P (Peptidyl) du ribosome (80S).

Ensuite vient l'élongation, le ribosome se déplace vers l'extrémité 3' de l'ARNm, les aatRNA se logent dans le site A (Aminoacyl) du ribosome en respectant évidemment la séquence de l'ARNm (code génétique). Le peptide fixé sur le tRNA (site P) se détache et se fixe sur l'acide aminé qui est lié au tRNA (site A), c'est l'élongation. Le ribosome se déplace vers le 3' de l'ARNm. Le tRNA du site P est libéré, celui qui était dans le site A vient dans le site P. Le site A est alors libre et peut accueillir le prochain aatRNA.

Arrivé au codon terminateur (soit UAA, UAG ou UGA) comme il n'existe pas de tRNA correspondant, le ribosome se disloque, c'est la terminaison. Mais elle se passe grâce à des facteurs de terminaison RF (Release Factor) qui imitent la structure 3D d'un tRNA (on parle de mimétisme structural) permettant ainsi le décrochage du peptide lié au tRNA (site P) donnant ainsi la protéine.

INITIATION.

Le methionine-tRNAⁱ ne reconnaît que AUG initiateur. En fait le ribosome ne se fixe pas directement sur AUG mais avant et il " scanne " l'ARNm. Dans certains cas, il ne s'arrête pas au premier AUG mais au deuxième ou encore après.

Nécessiter de facteurs d'initiation eIF (2 à 6), de GTP et d'ATP.

1 – Interaction du ARN^{ti}-met avec la petite sous unité du ribosome et de eIF2, 3 et 4 au niveau de la coiffe de l'ARN^m.

2 – Le complexe ARN^t-ribosome- eIF2, 3 et 4 se place au niveau du codon initiateur l'AUG.

3 – Recrutement de la grande sous unité du ribosome, libération des trois eIF. Formation du complexe d'initiation : ARN^{ti}-met, ribosome 80s et ARN^m.

ELONGATION

1 - "Activation "de l'aa tRNA^{aa} avec du GTP et 2 facteurs d'élongation eEF1 (eucaryote Elongation Factor).

2 - Liaison de l' aa tRNA^{aa} dans le site A

3 - Transfert du peptide : spontané

4 - Translocation grâce à un facteur eEF2, consommation d'un GTP par translocation d'un codon

5 - Libération du tRNA (site P).

TERMINAISON

La terminaison se fait grâce à un facteur RF qui a une structure tridimensionnelle très proche de celle d'un tRNA. Le RF ne se fixe que lorsque le codon est un codon stop (UAA, UAG ou UGA).