

Biologie Moléculaire



La synthèse des protéines.

I. Généralités

Le matériel génétique ou génome contient les gènes

Un **gène** contient une **information** sous la forme d'une **suite de nucléotides** et **s'exprime** lorsque cette information est **utilisée**.

GENES CODANTS	GENES NON CODANTS
Leur information sert à la synthèse des protéines	Leur information ne sert qu'à la synthèse des autres ARNs (<i>ARNr, ARNt, petits ARN nucléaires ou nucléolaires</i>)
Transcrits en pré-ARNm puis modifiés en ARNm mature (maturation)	Uniquement transcrit dans le noyau (non traduit)
L'ARNm rejoint le cytosol et sa séquence est traduite en acides aminés	Certains de ces ARNs restent dans le noyau , d'autres rejoignent le cytosol Tous participent à l'expression des gènes codants !

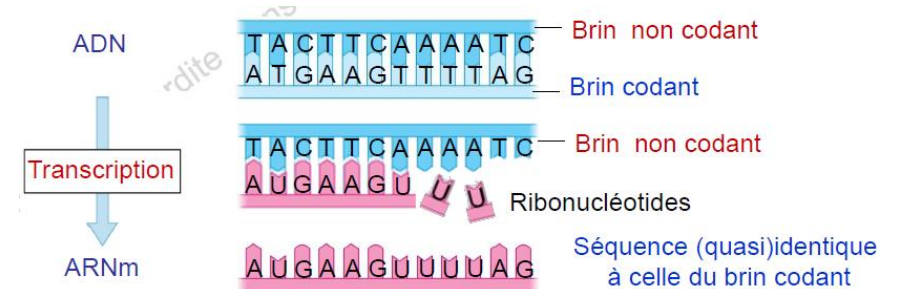
L'expression d'un gène codant **début**e par sa **transcription en ARNm** dans le **noyau** et **s'achève** par la **traduction** de l'ARNm en **protéine**.

Un gène est une séquence d'ADN **double brin**. L'un des brins contient l'information du gène à retranscrire dans l'ARNm (**brin codant**), l'autre brin ne contenant pas d'information (**brin non codant**).

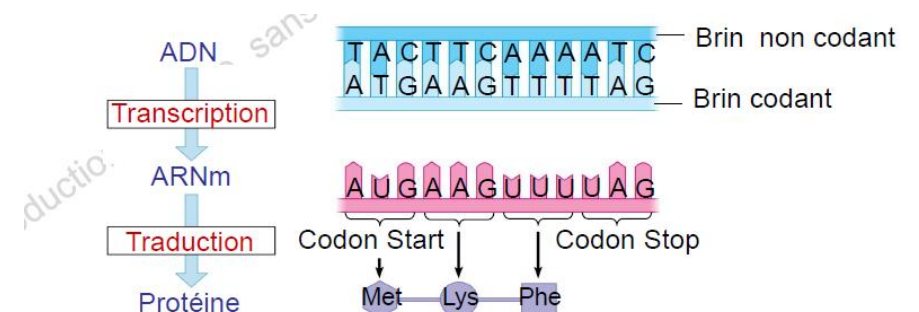
Comme la transcription repose sur le **principe de complémentarité**, le brin **non codant** sert de **matrice** pour former l'ARNm à partir de rNTPs.

L'information du **brin codant** est celle qui figure dans le **brin d'ARNm**.

C'est donc le brin complémentaire **non codant** qui sert de **matrice** pour la transcription selon le principe de complémentarité.



L'ARNm rejoint le **cytosol** où la séquence de **ribonucléotides** est traduite en une séquence d'**acides aminés** pour former la **protéine**. Cette étape de traduction repose sur un code appelé **code génétique** qui indique à quel **acide aminé** correspond chaque **triplet de nucléotides (codons)** de l'ARNm.



ATTENTION : Le brin **codant** contient l'**information** mais le brin **non codant** sert de **matrice** lors de la **transcription**.

II. Structure d'un gène codant eucaryote

Un gène codant eucaryote comprend deux régions :

A. Une région destinée à être transcrite (unité de transcription)

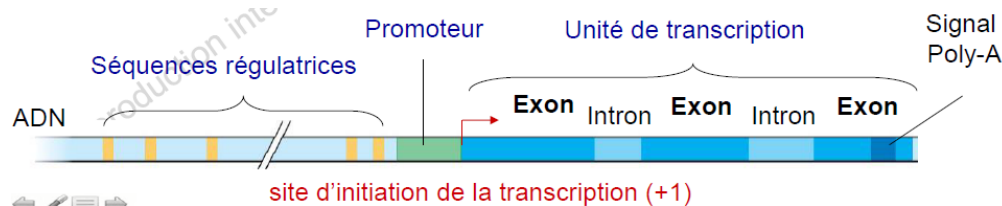
C'est un ensemble de séquences **transcrites** mais **pas forcément traduites** !

Cette succession de séquences **codantes (EXONS)** et **non codantes (INTRONS)** est transcrite du signal d'initiation (nucléotide +1) au signal de terminaison (signal Poly-A).

ATTENTION : Transcription et traduction commencent et ne finissent pas au même endroit !

B. Des régions situées en amont et non transcrites

1. **Le promoteur** : situé **près du site d'initiation** de la transcription et constitué de la séquence TATAA (TATA box), il fixe le complexe assurant la transcription
2. **Des séquences distales et proximales** : plus éloignées, elles assurent la **régulation de la transcription**.



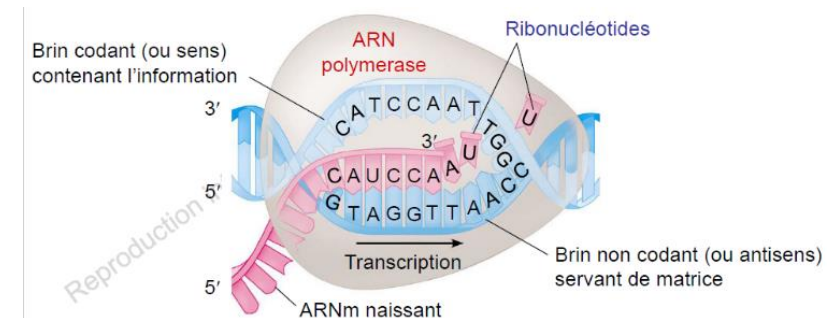
C. L'ARN polymérase II

Chez les eucaryotes, c'est l'**ARN polymérase II** qui transcrit les gènes codants.

Elle se fixe au promoteur du gène et recopie l'unité de transcription

Elle relie entre eux les **rNTPs** complémentaires du brin non codant dans le sens **5' → 3'**, du site d'initiation de la transcription au signal Poly-A.

Sa liaison au **promoteur** et son **activation** requièrent d'autres **protéines**.

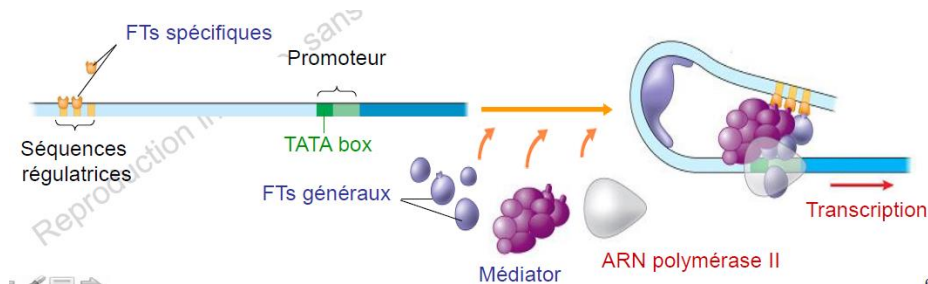


D. La séquence TATA box

La TATA box **recrute la machinerie basale** de transcription qui comprend :

- **L'ARN polymérase II**, dont l'extrémité **C-term** peut être phosphorylée
- Les facteurs généraux de transcription (**TFII A, B, D, E, F, H**) qui permettent à l'ARN polymérase II de se fixer au promoteur et l'activent

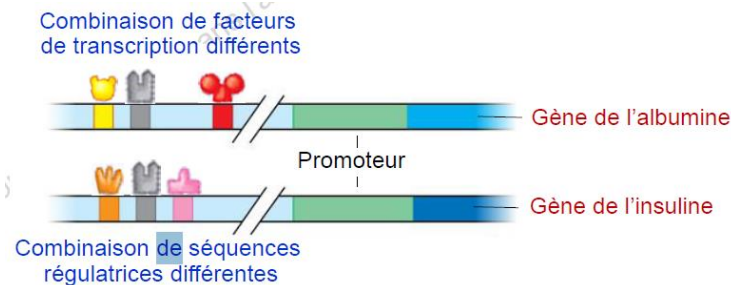
E. Les séquences régulatrices



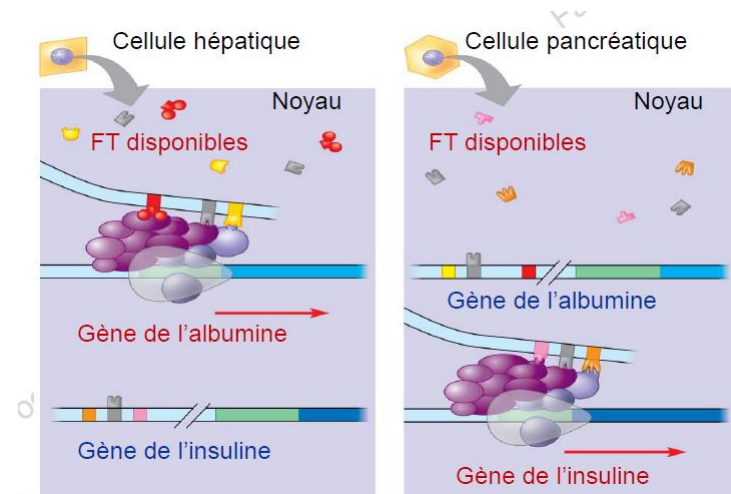
Les **séquences régulatrices** d'amont fixent d'autres protéines. Ces protéines **activent** (*enhancer*) ou **répriment** (*repressor*) la transcription.

Les séquences régulatrices des gènes codants varient

Chaque gène possède sa **propre combinaison** de séquences régulatrices. Une séquence **donnée** ne peut fixer qu'un facteur de transcription **donné**. L'ensemble des séquences régulatrices d'un gène permet la fixation d'une **combinaison particulière** de **facteurs de transcription**. Ainsi, chaque gène est régulé par une **combinaison** de facteurs de transcription.



**Le gène ne s'exprimera qu'en leur présence,
Variable selon le type cellulaire**



Chaque gène recrute donc une **combinaison variable** de **FT spécifiques** qui permettent l'assemblage de la machinerie basale de transcription.

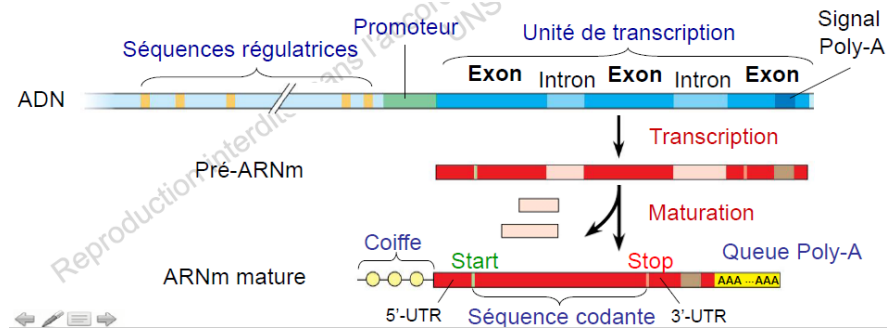
F. L'ARN pré-messager

**Un gène codant eucaryote est transcrit
en ARN pré-messager**

Ce transcrit primaire subit une **maturation** pour pouvoir être traduit.

Des modifications **co-transcriptionnelles** assurent sa maturation en ARNm :

- L'ajout d'une « **coiffe** » à l'extrémité 5' et d'une « **queue** » **Poly-A** en 3'
- L'**excision** (élimination) des **introns** et l'**épissage** des **exons** (ligation) de telle sorte que sa séquence codante entre les signaux **Start/Stop** soit **ininterrompue**.



ATTENTION : Gènes codants et non codants sont transcrits par la même ARN polymérase !

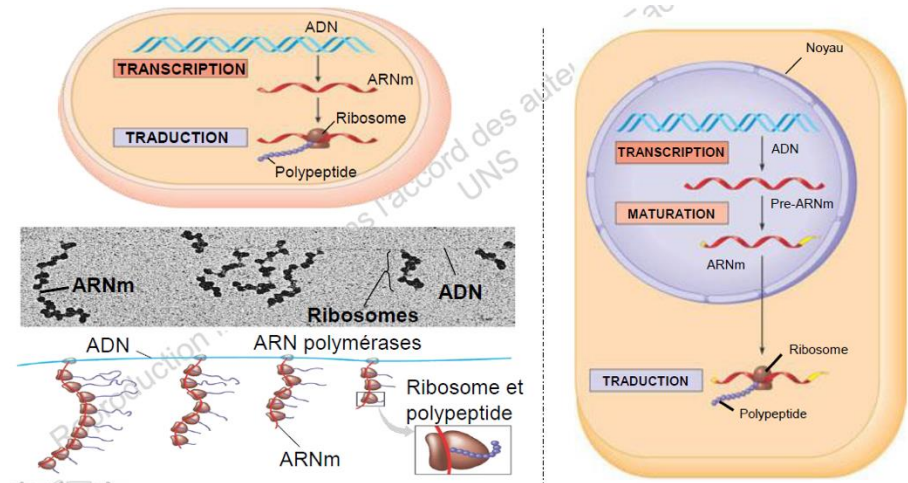
Elle est assistée du facteur σ (sigma) chargé de reconnaître le promoteur.

Il n'existe pas de facteurs généraux de transcription.

III. L'expression des gènes chez les procaryotes

L'expression des gènes est différente chez les procaryotes

PROCARYOTES	EUCARYOTES
Les gènes sont compacts et regroupés (absence d'introns) et régulés par les mêmes séquences régulatrices	Les gènes sont morcelés (grâce aux introns) et régulés individuellement
ADN non associé à des protéines histones La transcription débute sans décompaction des nucléosomes	ADN associé à des protéines histones
Une séquence régulatrice unique contrôle un ensemble de gènes (= opéron)	Plusieurs séquences régulatrices proximales et distales
Opéron transcrit en un long ARNm ne nécessitant pas de maturation	Gène transcrit en un pré-ARNm nécessitant des maturations
Transcription et traduction simultanées Pas de membrane séparant le nucléoïde du cytosol	La transcription précède la traduction Membrane séparant le noyau du cytosol



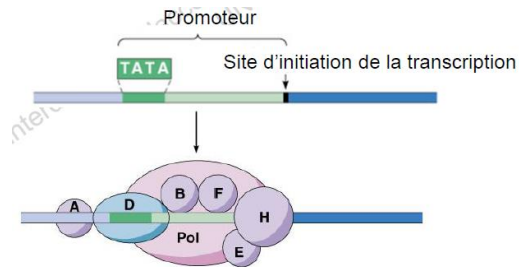
IV. Les étapes de la transcription

A. Initiation de la transcription

L'initiation de la transcription nécessite plusieurs étapes

- Elle débute par la fixation du complexe TFIID sur le promoteur. Il se lie à la TATA Box grâce à l'une de ses sous-unités appelée TBP !
- Les autres complexes et l'ARN polymérase II sont ensuite recrutés.

L'ensemble forme la machinerie basale de transcription encore **INACTIVE** tant que l'**extrémité C-terminale** de l'ARN polymérase n'est pas **phosphorylée** !



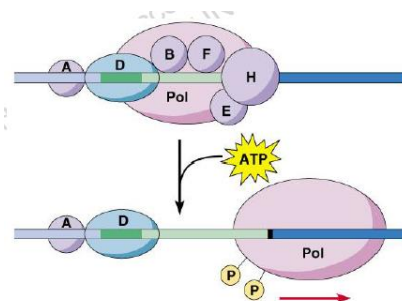
L'ARN polymérase II ne peut pas se fixer seule au promoteur !

-Le complexe **TFIIH** **phosphoryle** l'extrémité C-terminale de l'ARN Pol II.

Elle est **activée** et la transcription de l'ARNm débute (phase d'**élongation**).

-La maturation de l'ARN pré-messager débute **simultanément** !
D'autres **phosphorylations** de l'extrémité C-terminale recrutent les **enzymes de maturation** du pré-ARNm (couplage élongation-maturation).

La transcription débute avec la phosphorylation de la polymérase



B. Élongation de la transcription

La maturation de l'ARNm débute **SIMULTANEMENT** à la transcription (couplage élongation/maturation) : d'autres **phosphorylations** de l'extrémité C-term de l'ARN Pol II recrutent **successivement** les enzymes de **maturation du pré-ARNm**.

L'ARN polymérase utilise le **brin non codant (=antisens)** comme **matrice**, afin de **transcrire** le message qui est celui du **brin sens** par complémentarité des bases puisque les deux brins sont ... **complémentaires** ! L'ARN Pol II relie donc entre eux les rNTPs complémentaires du brin non codant.

La synthèse se fait dans le sens 5' - 3' et s'arrête au signal Poly-A

C. Terminaison de la transcription

Les gènes eucaryotes ne possèdent pas de séquence signalant à la machinerie basale de transcription où s'arrêter.

La transcription s'arrête vers la **séquence de polyadénylation AAUAAA**.

D. Modifications co-transcriptionnelles

Un gène codant eucaryote est transcrit dans un premier temps en **ARN pré-messager** (transcrit primaire).

Des modifications co-transcriptionnelles assurent sa maturation en ARNm :

- L'ajout d'une **coiffe** à l'extrémité 5' (coiffe) et de la **queue Poly-**

A en 3'

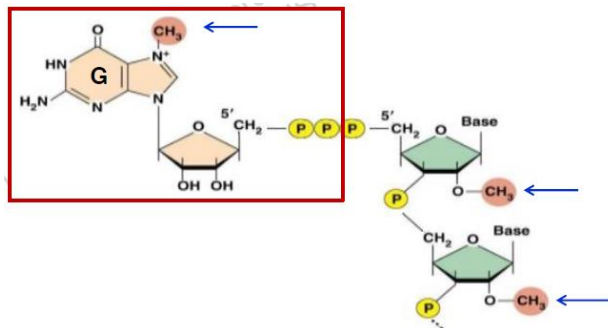
- L'**excision** (élimination) des introns et l'**épissage** des exons (ligation)

L'ARNm mature aura une séquence codante ininterrompue et encadrée par les signaux Start/Stop

1. La coiffe de l'ARNm

La coiffe comprend plusieurs modifications de l'**extrémité 5'** :

- Ajout d'un nucléotide à **guanine** à l'**extrémité 5'-P** du transcrit et **méthylation** ($-\text{CH}_3$) de la **guanine** ajoutée et du **ribose** des **deux premiers** nucléotides.
- La coiffe **protègera** le transcrit de la **dégradation**, augmentant sa durée de vie, et est **nécessaire** à sa **reconnaissance** par les **ribosomes**.

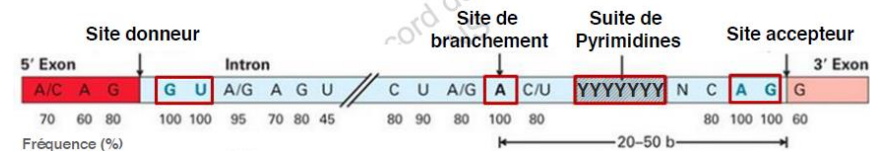


2. L'épissage des introns (régions transcrites mais non codantes)

L'épissage fait intervenir des séquences introniques appelées consensus

Elles sont **quasi invariables** et retrouvées dans **tous les gènes codants** :

- Site **donneur d'épissage (GU)** au début et **accepteur (AG)** à la fin de l'intron.
- Site de **branchement (A)** et **suite de pyrimidine (Y)** avant la fin de l'intron.



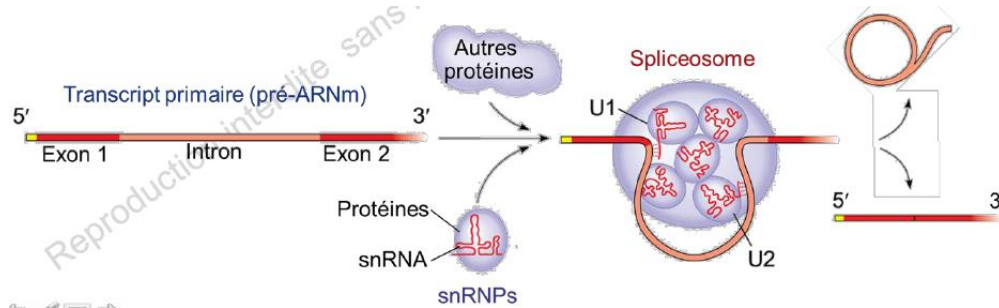
Le spliceosome est le complexe enzymatique qui assure l'épissage

Il est formé par les **ribonucléoprotéines U1, U2, U4, U5 et U6** constituées de **diverses protéines** et des **petits ARNs nucléaires (snRNAs)** correspondants.

Les ribonucléoprotéines snRNPs « repèrent » et définissent les introns.

Le site **donneur** et de **branchement** fixent respectivement le complexe **U1** et **U2** via leur **snRNA** **respectif** qui s'apparie par complémentarité avec l'ARNm.

Les **complexes U1 et U2** recrutent les **autres complexes** ce qui permet de **rapprocher les exons** pour les réactions suivantes : l'**intron est éliminé** sous la forme d'un lasso (ou lariat) et les **exons sont reliés** entre eux !

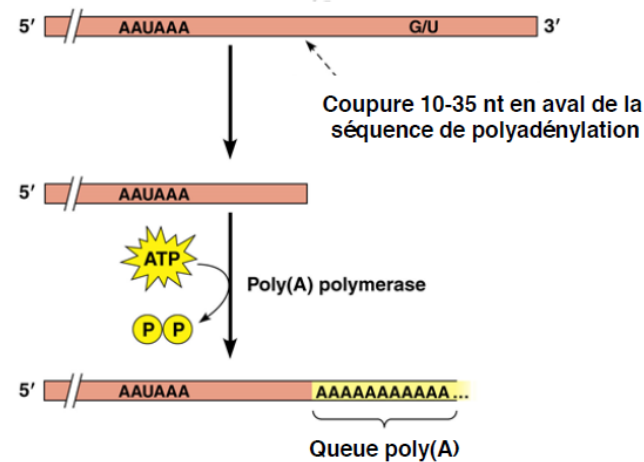


3. La polyadénylation de l'ADN

La transcription s'interrompt vers la **séquence de polyadénylation AAUAAA**.

Des complexes coupent le **pré-ARNm** quelques nucléotides après ce signal.

La **Poly (A) polymérase** vient ajouter une **queue poly A** d'environ 250 nucléotides.

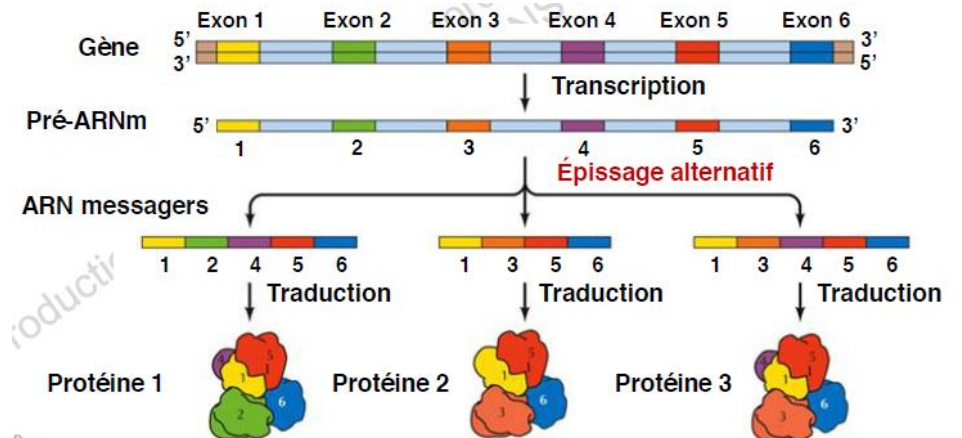


La polyadénylation **ralentit** aussi la **dégradation** du transcrit mature.

E. Un seul gène pour plusieurs ARNm

Plusieurs ARNm différents sont issus d'un seul gène

- Le transcrit **primaire** (pré-ARNm) lui-même peut être **variable** : utilisation de **sites alternatifs** d'initiation/terminaison de la transcription.
- Le transcrit **mature** (ARNm) peut varier selon les exons qu'il contient : le phénomène d'épissage alternatif aboutit à différents ARN messagers !
- Ces différents ARNm sont traduits en protéines différentes



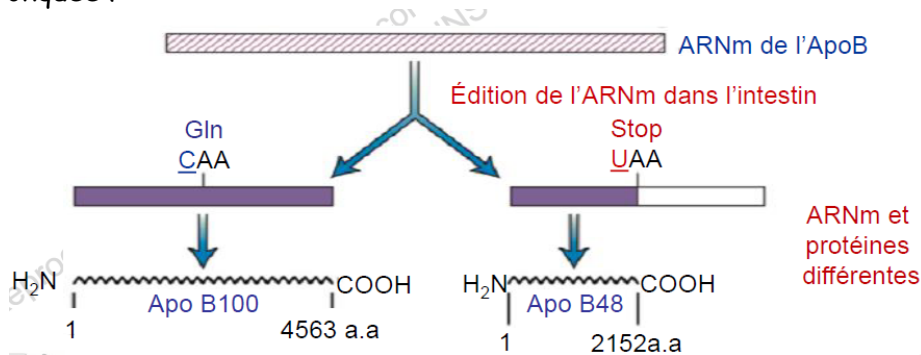
F. Modifications post-transcriptionnelles

La séquence d'un ARNm mature peut encore être changée (édition)

Exemple de l'ARNm de l'apolipoprotéine B (ApoB) :

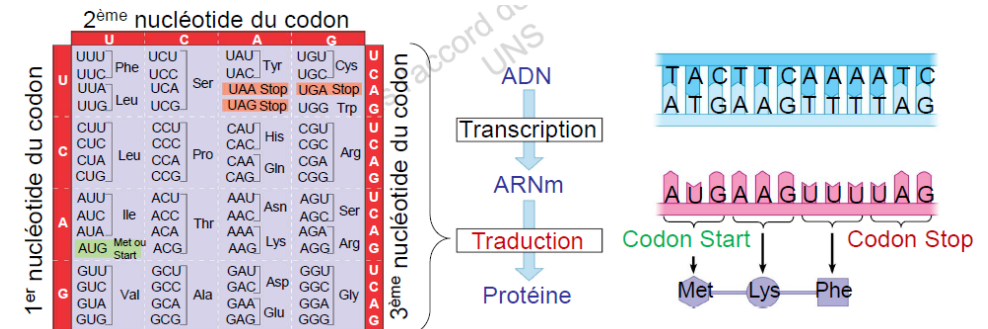
- Dans le **foie**, cet ARNm n'est **pas modifié** et est traduit en **ApoB100**.
- Dans l'**intestin**, une **cytosine** de l'ARNm est **désaminée** en **uracile**.

Il y **arrêt de la traduction** (codon **Stop**) et production d'**ApoB48**, tronquée !



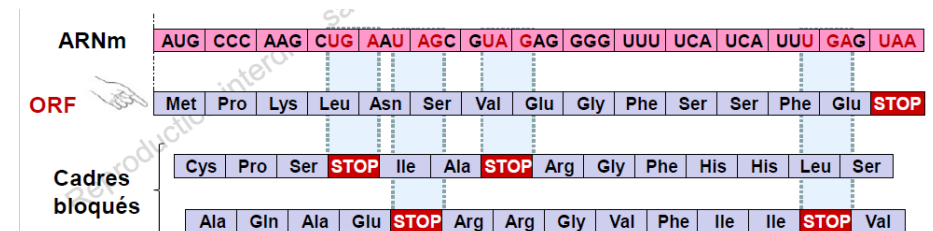
- **Non-ambigu** : un codon donné correspond toujours au même AA.
- **Dégénéré** : plusieurs codons spécifient le même AA sauf pour la méthionine et le tryptophane.

ATTENTION : il y a **61 combinaisons** pour **20 AA** !



Il existe 3 cadres de lecture de l'ARNm en théorie

Un **seul** aboutit à la synthèse de la protéine attendue : le **cadre ORF**, débutant au codon **AUG** repéré grâce à la **séquence Kozak**. Les **2 autres** sont **décalés** par rapport au cadre ORF, les protéines formées sont **différentes** et souvent stoppées par un **codon STOP prématuré**.



V. Le code génétique et ses caractéristiques

Le code génétique assure la **correspondance codon/ acide aminé**. Il existe **4 puissance 3 combinaisons** de **3 nucléotides** pour former un **codon**.

On retrouve 1 codon **Start (AUG)** qui **initie** la traduction (code pour une **méthionine**), et **3 codons Stop (UAA / UAG / UGA)** qui **terminent** la traduction.

Ce code est :

- **Quasi-universel** : toutes les espèces utilisent la même correspondance codon/AA. Il y a quelques rares exceptions comme les **mitochondries**.
- **Non-chevauchant** : chaque nucléotide de l'ARNm appartient à un seul codon.

VI. Mutations du code génétique

Certaines mutations de l'ADN modifient le code génétique.

Parmi les substitutions (c'est-à-dire celles qui changent un nucléotide) on trouve :

- ✓ Les mutations silencieuses (elles sont **neutres**) : ne changent pas l'acide aminé codé.
- ✓ Les mutations faux sens : remplacent un AA par un autre.
- ✓ Les mutations non-sens : introduisent un **codon STOP prématuré**.

Parmi les insertions/délétions (celles qui modifient le nombre de nucléotides) on trouve :

- ✓ Les multiples de 3 : on peut avoir ajout d'un AA ou d'un STOP mais le cadre de lecture est respecté.
- ✓ Les non multiples de 3 : le cadre de lecture en aval peut être décalé et il peut y avoir la présence de mutations **faux sens multiples** ainsi que des **Stop prématurés**. Elles peuvent également former une mutation non -sens et dans ce cas il y a une **absence totale** de la synthèse de la protéine.

Le code génétique est organisé en **16 boîtes de 4 codons**, où seul le **3^{ème} nucléotide change**. Dans la majorité des cas, l'AA codé par la boîte est le même. Si ce n'est pas le cas, il est de **même polarité**.

Diagramme du code génétique (tableau des codons) :

		2 ^{ème} nucléotide du codon				3 ^{ème} nucléotide du codon
		U	C	A	G	
1 ^{er} nucléotide du codon	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr Stop Stop	Cys Cys Stop Trp	U C A G
	C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
	A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
	G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

Boîte Phénylalanine-Leucine (codons UUU, UUC, UUA, UUG)

Boîte Glycine (codons GGU, GGC, GGA, GGG)

L'organisation du code permet de **minimiser les effets des mutations** : l'importance du nucléotide varie selon sa position :

- Une mutation du **3^{ème} nucléotide** est souvent **neutre** et sans conséquence sur la protéine ;
- Une mutation du **2^{ème} nucléotide** induit souvent une **mutation faux sens non conservative** (les conséquences sont donc plus sévères) ;
- Une mutation du **1^{er} nucléotide** induit souvent une **mutation faux-sens conservative**.

VII. La synthèse des protéines

La synthèse des protéines nécessitent **différents acteurs** qui peuvent avoir un rôle :

❖ **Structural :**

- **L'ARN messager (ARNm) :** sa structure primaire contient les instructions pour la synthèse des protéines.
- **L'ARN de transfert (ARNt) :** se fixe sur l'ARNm par complémentarité de la séquence codon/anticodon et apporte les AA grâce à sa tige acceptrice.
- **L'ARN ribosomal (ARNr) :** s'associent avec des protéines pour former des ribosomes.

❖ **Fonctionnel :** (ce sont des enzymes)

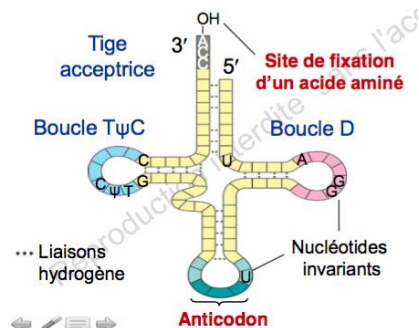
- Les **AminoAcyls ARNt synthétases** : fixent l'ARNt sur les protéines. Elles sont **spécifiques** d'un AA classique, il n'y a donc pas d'aaRs pour la sélénocystéine !
Elles peuvent fixent l'AA sur des ARNt isoaccepteurs grâce à l'ATP, et possèdent également une activité de **proofreading**.
- Un **ARN ribosomal** : permet de relier les AA par des **liaisons peptidiques** pour former la protéine.

Les ARN de transfert : ils sont formés d'une tige acceptrice et de 3 boucles. L'AA est fixé à l'extrémité 3'OH de la tige acceptrice.

Attention, chaque ARNt est spécifique d'un AA.

Ils sont d'abord transcrit sous la forme d'un pré-ARNt qui va subir 10 à 25% de modification de base. L'ARNt contient donc des bases qu'on appelle mineures, telles que :

- **L'Inosine** : formée par la désamination fréquente de l'adénine en hypoxanthine ;
- La **Pseudo-uridine**, la **Dihydrouridine**, la **Méthyluridine** : ce sont des dérivés de l'uridine ;
- La **thymine** (celle de l'ADN).

**A. Spécificités du code génétique**

La fiabilité de la traduction est assurée par 2 mécanismes :

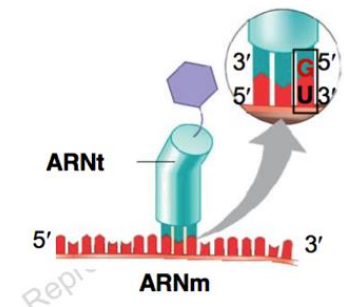
- La spécificité de l'appariement entre l'anticodon de l'ARNt et le codon de l'ARNm : c'est l'association inhabituelle de la 1ère base de l'anticodon et de la 3ème base du codon

Il devrait exister **61 ARNt** mais un **appariement flexible en 5' (wobble)** réduit ce nombre à **48** et permet de **minimiser** l'effet des mutations.

Ce wobble **ne respecte donc pas le principe de complémentarité**.

De nouvelles paires de bases peuvent se former entre :

Anticodon (1 ^{ère} base)	Codon (3 ^{ème} base)
A	U
C	G
U	A et G
G	C et U
I	A, U et C



Mais attention, la **règle purine-pyrimidine** est le plus souvent **respectée** !

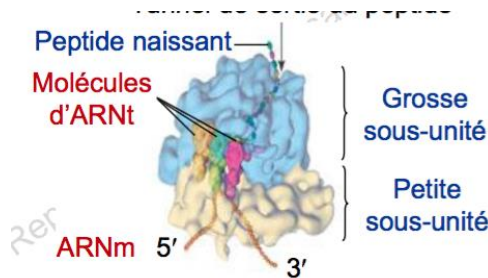
- La spécificité de l'appariement entre la protéine et son ARNt : elle est assurée par des enzymes (20) : les **aminoacyls ARNt synthétase (aaRs)**.

Chacun est **spécifique d'un AA** mais peut le fixer sur un ou plusieurs ARNt, ce sont les **ARNt isoaccepteurs**.

B. La traduction

Elle est assurée par le **ribosome**. Il est composé de **2 sous-unités** ayant des fonctions spécifiques :

- La petite sous-unité : se lie à l'ARNm et **décode** l'information en assurant la **correspondance codon/anticodon**. Elle est appelée **30s** chez les **procaryotes** et **40s** chez les **eucaryotes**.
- La grosse sous-unité : se lie à la petite s.u et **fabrique** la protéine. Elle a un rôle **structural** et **fonctionnel**. Elle est appelée **50s** chez les **procaryotes** et **60s** chez les **eucaryotes**.



Elle se divise en **3 sites** :

- le **A** qui accueille l'ARNt avec l'AA
- le **P** qui forme le peptide
- le **E** qui éjecte l'ARNt.

C'est l'**ARN 28s** (enzyme) qui forme la liaison peptidique.

La petite et la grosse s.u **diffèrent** par leurs contenus en ARNr et en protéines.

C. Reprogrammation :

Le code génétique peut être **reprogrammé**. On note que les **sélénoprotéines** sont des protéines contenant la **sélénocystéine**. C'est un **AA rare** pour lequel il n'existe **pas de codon**, mais qui peut être incorporé durant la traduction.

C'est la reprogrammation du **codon Stop UGA** qui est à son origine. La sélénocystéine possède son propre ARNt : **ARNt^{(ser)sec}** mais il n'existe pas d'enzyme pouvant le fixer sur son ARNt ...

Il est d'abord chargé avec la **sérine** puis après réaction chimique, la **sélénocystéine** est chargée sur son ARNt.

Cette reprogrammation est liée à l'ARNm des sélénoprotéines.

Leurs **extrémités 3'UTR** contiennent une séquence particulière, appelée **SECIS**.

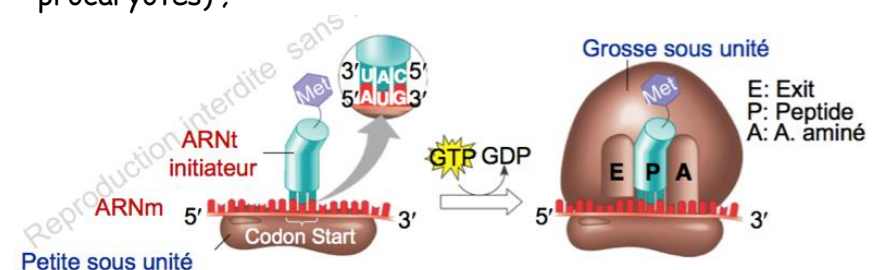
Cette extrémité se replie en **épingle à cheveu**, et elle est alors reconnue par 2 protéines qui apportent l'ARNt^{(ser)sec}.

La sélénocystéine est alors **incorporée au codon UGA**.

La traduction se fait en 3 étapes successives :

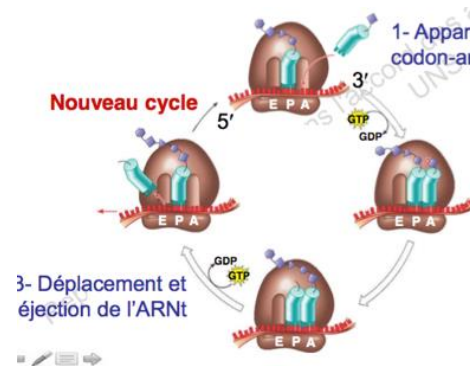
1. L'initiation en 2 étapes :

- 1) Fixation du **complexe de pré-initiation** composé de la **petite sous-unité** et d'un **ARNt initiateur** portant la **méthionine** (se fixe sur la coiffe pour les eucaryotes, sur le codon AUG pour les procaryotes) :



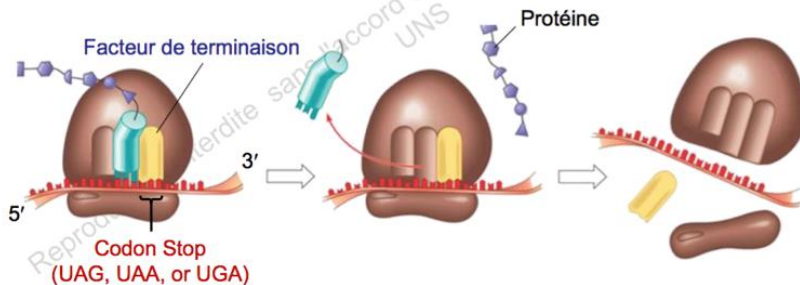
- 2) Fixation de la **grosse s.u** sur la petite s.u après reconnaissance du codon initiateur par l'ARN, formation du **ribosome**.

L'**élongation** (succession de cycles), le ribosome se déplace de codon en codon si l'appariement codon/ anticodon est correct.



La **terminaison** s'effectue lorsque le ribosome rencontre un codon STOP.

Il n'y a **pas d'ARNt** correspondant au **codon STOP**, c'est un facteur de terminaison qui se fixe à la place. La protéine est libérée et le ribosome **se dissocie**.



De **nombreux ribosomes** se fixent sur un ARNm, sa traduction est donc assurée **simultanément** à différents endroits. C'est ce qu'on appelle un **polyribosome** ou **polysome**.

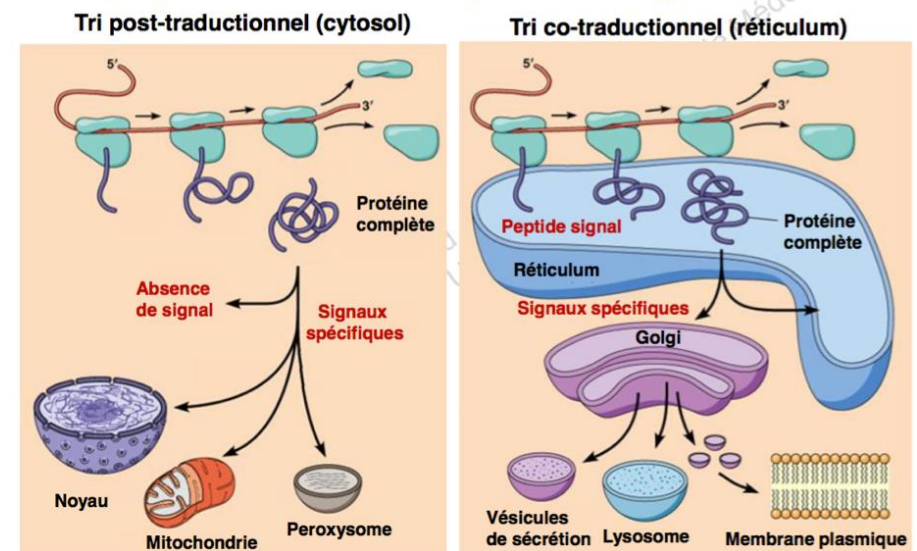
L'**efficacité** et la **rapidité** de la traduction est ainsi **augmentée**.

D. Adressage des protéines

C'est un **tri sélectif** de la protéine vers son site d'action. Il se fait grâce à un **signal protéique** ou **non** situé dans la séquence. Chaque compartiment possède un **signal spécifique**.

On a **2 mécanismes** de tri :

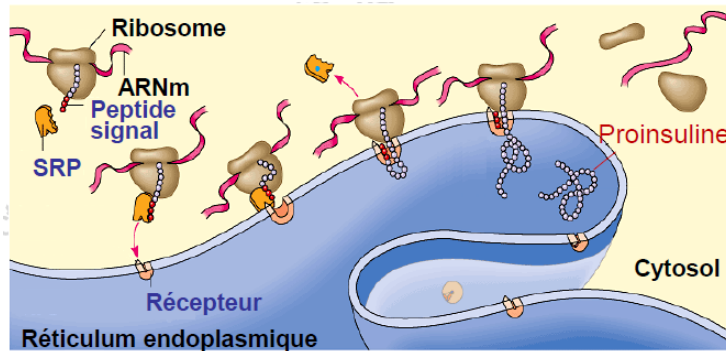
- **Post-traductionnel cytosolique** : (quand la protéine est **finie**). La protéine reste dans le cytosol ou va dans le noyau, les mitochondries ou les peroxysomes en fonction de son signal.



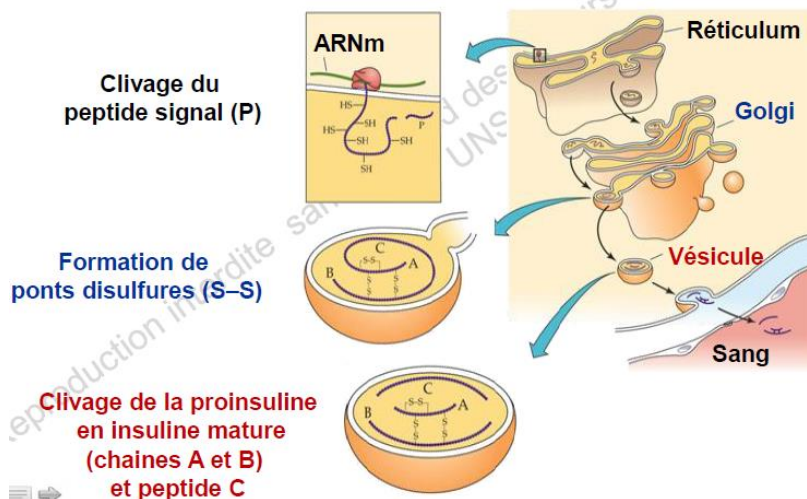
- **Co-traductionnel** : (se fait **au cours de la synthèse** de la protéine, par un ribosome libre présent grâce à un **peptide signal**). Le ribosome se fixe au REG où la synthèse va s'achever. Ensuite si **d'autres signaux** sont présents, la protéine peut aller dans le golgi, le lysosome ou la membrane.

E. Synthèse de l'insuline

L'**insuline** possède un **peptide signal** et doit être sécrétée. Sa synthèse est achevée dans le **REG**, où son ribosome et ARNm sont venus se fixer pendant l'élongation grâce à sa protéine de reconnaissance (SRP) et à son récepteur qui forme un canal membranaire.



La synthèse s'achève à ce canal jusqu'à formation de la **pro insuline**. Cette pro insuline possède un signal pour le **golgi** où elle subit une **maturation**. Puis elle est sécrétée dans la circulation par **exocytose**.



VIII. Points clés

Les gènes codants eucaryotes permettent la synthèse des protéines :

- Ils possèdent un promoteur et des séquences régulatrices **non transcrits** !
- Le **promoteur minimal** est **constant** dans la plupart des gènes, il est constitué par la **TATA box** qui fixe la machinerie basale.
- Les **séquences régulatrices** sont **variables** selon les gènes, elles fixent les **FT spécifiques** qui régulent la machinerie basale.
- Leur séquence transcrite est **morcelée** (présence d'introns).

Ils sont transcrits par l'ARN Pol II chez les eucaryotes :

- Elle utilise le principe de **complémentarité des bases**.
- Le transcrit primaire subit une **maturation** (coiffe, polyadénylation, épissage).
- Plusieurs **ARNm** et **protéines** peuvent provenir d'un **seul gène**.

La traduction d'un ARNm en protéine repose sur le code génétique :

- Il est quasi-universel, non ambigu, non chevauchant et dégénéré.
- La traduction respecte le cadre ORF par le codon AUG.
- Une mutation faux-sens, non-sens, avec décalage perturbe le message.

Elle fait intervenir les ARNt chargés et les ARN ribosomaux :

- Chaque aminoacyl-ARNt synthétase est spécifique d'un AA.
- Elle le fixe sur un ou plusieurs ARNt isoaccepteurs.
- L'appariement de l'anticodon d'un ARNt est flexible en 5'.
- Le ribosome se fixe à l'ARNm et relie entre eux les AA.

Chaque protéine subit une maturation et un tri sélectif :

- Son tri repose sur la présence ou l'absence de signal spécifique.