

## L' EXPRESSION DES GENES ET SON CONTROLE

### Introduction

Expression du matériel génétique : de l'ADN aux protéines. La synthèse protéique sera vue plus tard. On étudiera ici la synthèse des ARN, qui se pose en des termes différents chez les Eucaryotes (sites distants de transcription et traduction) et les Procaryotes (sites identiques car pas de compartimentation).

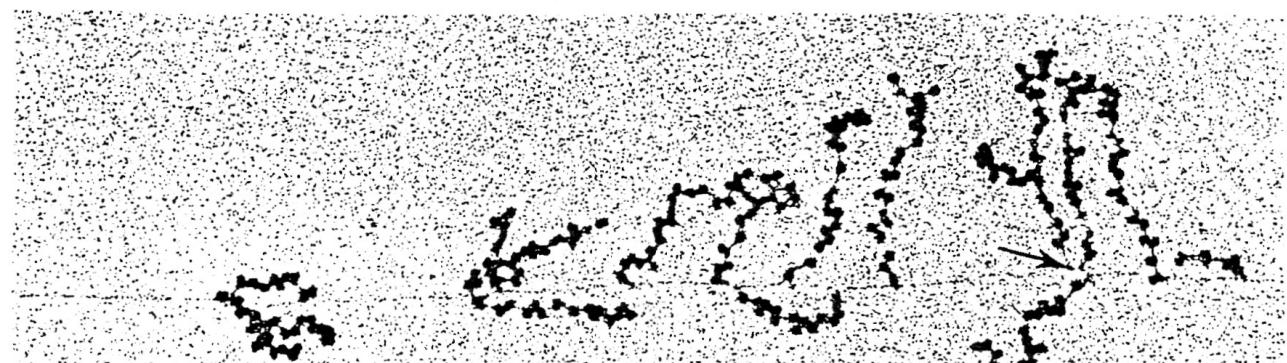
De même, le contrôle de l'expression des gènes ne se pose pas de la même manière chez les Procaryotes (adaptation rapide aux conditions du milieu) que chez les Eucaryotes (spécialisation cellulaire d'une part, et développement de l'œuf à l'organisme fonctionnel d'autre part)

### I. Observations microscopiques et mise en évidence

Le concept d'ARN messager est pressenti par JACOB et MONOD en 1967 à partir de considérations structurales simples et des résultats de Palade. De plus, CASPERSON et BRACHET avaient montré en 1940 que la quantité de protéines synthétisées dans le cytosol était proportionnelle à la teneur cytoplasmique en ARN.

Une preuve définitive du rôle d'intermédiaire de l'ARNm est l'utilisation d'uridine tritiée, montrant la synthèse nucléaire et la migration vers le cytosol.

× 70 000



Observation de nucléoïde de E. Coli

Noter la coexistence de la transcription et de la traduction (ribosomes bien visibles)

× 300 000



Observation de chromatine décondensée dans le noyau d'une cellule eucaryote

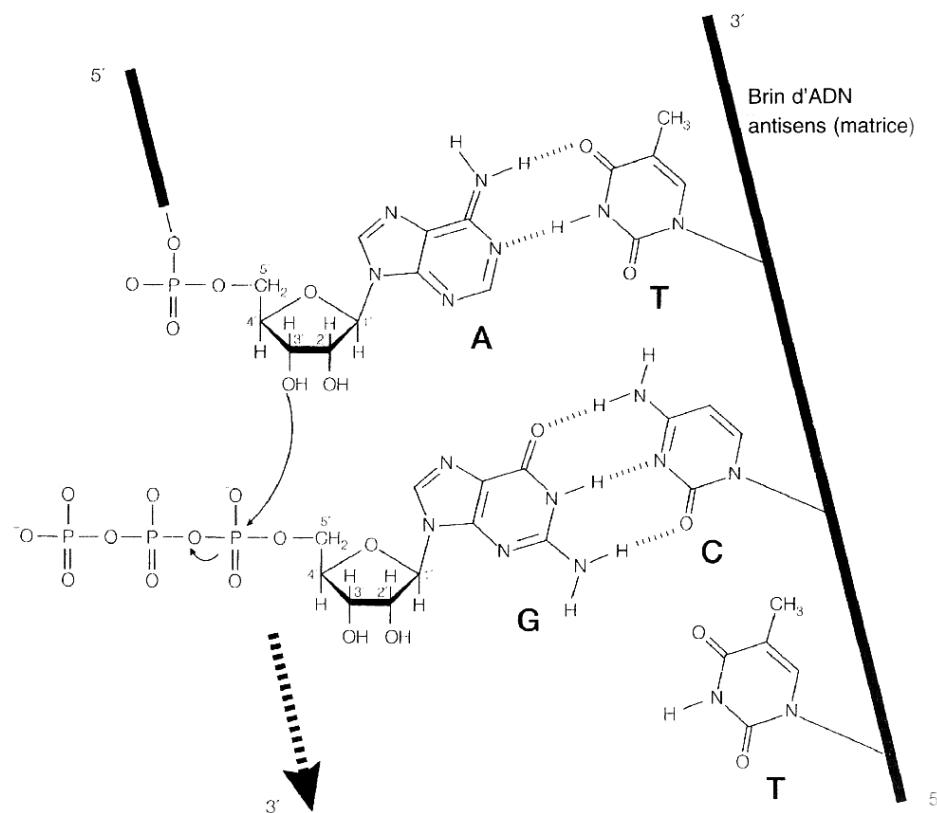
Forte densité d'ARNm en cours de synthèse dans un noyau eucaryote. Les tailles des ARNm donnent le sens de synthèse.

## II. Déroulement de la transcription chez les Procaryotes

### II.A. L'ARN Polymérase de E. Coli

Plus de 450 kb. 4 sous-unités et une sous-unité  $\sigma$  nécessaire pour le début du processus. Couvre environ 60 paires de bases. Nécessite un déroulement en amont et un réenroulement en aval.

Cette enzyme complexe catalyse l'addition de nucléotides triphosphorylés sur un ARN à partir d'un brin matrice d'ADN dans le sens 5' → 3'.



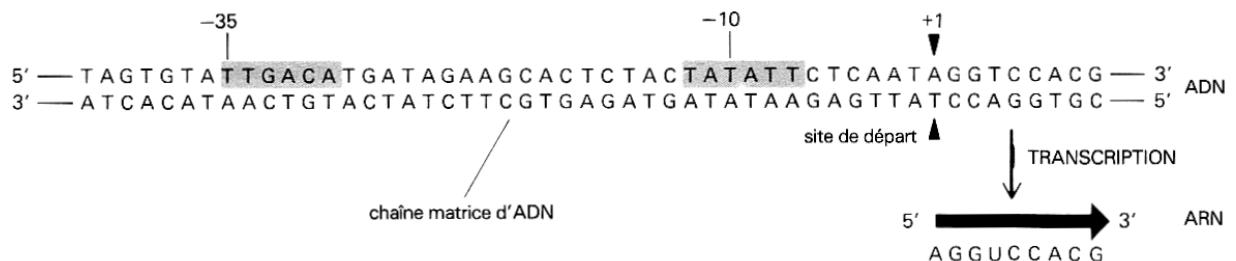
Formation de la liaison phosphodiester au cours de l'élargissement

## II.B. Le promoteur $\sigma$

La chaîne  $\sigma$  de l'ARN polymérase a une MM de 70 kDa en moyenne. Elle est nécessaire à l'initiation du processus de transcription mais pas à l'élargissement. Cette sous-unité est capable de reconnaître une séquence de nucléotides de l'ADN appelée promoteur. De nombreux facteurs  $\sigma$  existent chez les bactéries, de même que de nombreux promoteurs, mais le plus fréquent reste le promoteur  $\sigma^{70}$ .

Ce promoteur est constitué de 40 paires de bases en moyenne. Il est constitué de séquences d'ADN qui, bien que variables, présentent des similitudes statistiques : on parle de séquences consensus : ce sont les séquences TTGACA et TATAAT situées sur le brin codant respectivement en position -35 à -30 et -12 à -7 paires de base avant la séquence transcrive.

Notons d'ores et déjà que l'efficacité des promoteurs est éminemment variable (facteur 1000) et participe bien entendu aux phénomènes de régulation des transcriptions.

Séquences consensus du promoteur  $\sigma^{70}$  de E. Coli

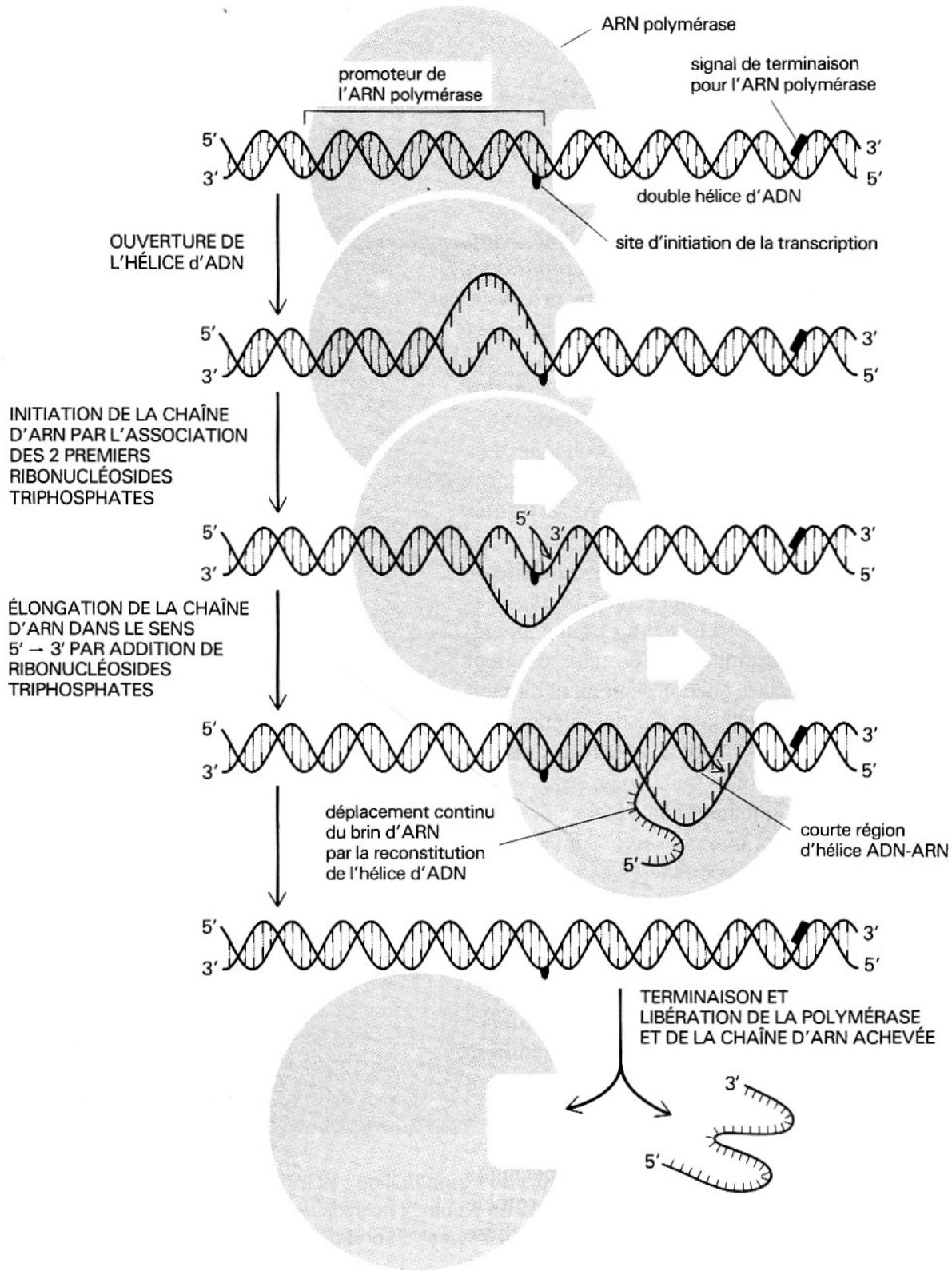
### II.C. Initiation, élongation et terminaison

Initiation : fixation du facteur  $\sigma$  favorisant la reconnaissance du promoteur. Ceci provoque l'ouverture au niveau de l'ADN d'une « bulle de transcription » de 17 paires de bases.

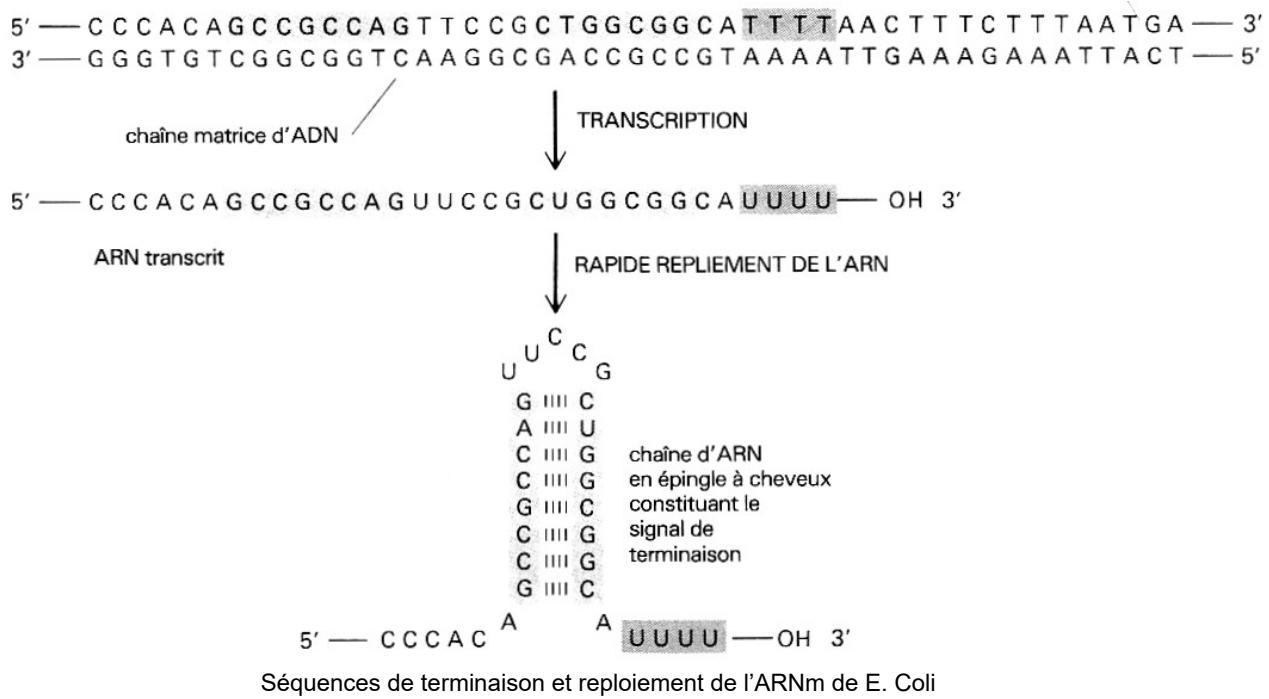
Elongation : l'ARN polymérase libère le facteur  $\sigma$  et progresse le long de l'ADN en maintenant cette bulle ouverte grâce à des hélicases (famille des topoisomérases) et catalyse l'ajout de nucléotides à la vitesse d'environ 40 par seconde dans le sens 5'  $\rightarrow$  3'. Une dizaine de nucléotides de l'ARN en cours d'élongation restent appariés à l'ADN ouvert.

Terminaison : La structure finale est systématiquement riche en nucléotides de type G ou C et provoque la formation d'un boucle en épingle à cheveux par appariement interne à l'ARN néoformé. Il se termine enfin par quelques (4 en moyenne) nucléotides U. Il semblerait en tout cas que l'épingle à cheveux provoque des tensions au sein de la bulle de réPLICATION, suffisamment pour que l'ARN se détache.

Un autre mode de terminaison fait intervenir une ATP-ase hexamérique rompant les ponts ADN-ARN appelée protéine rho.



Déroulement schématique de la transcription chez E. Coli



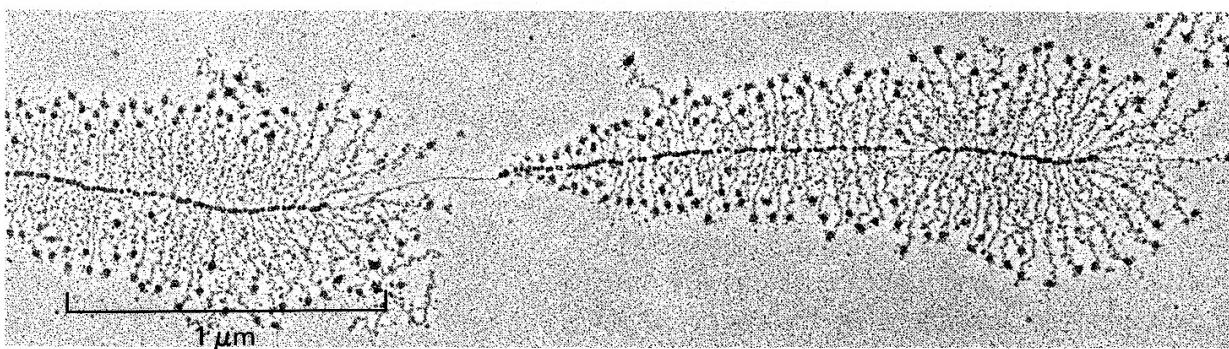
### III. Les ARN polymérases des Eucaryotes

Les cellules eucaryotes ont pour particularité de présenter une nette compartimentation. Il en résulte que les processus de transcription et de traduction sont distincts dans l'espace et dans le temps.

Mécanisme similaire à celui des Prokaryotes, mais présence de trois ARN polymérases : une pour les ARN ribosomiaux (I), une pour les ARN messagers (II) et une pour les ARN de transfert (III).

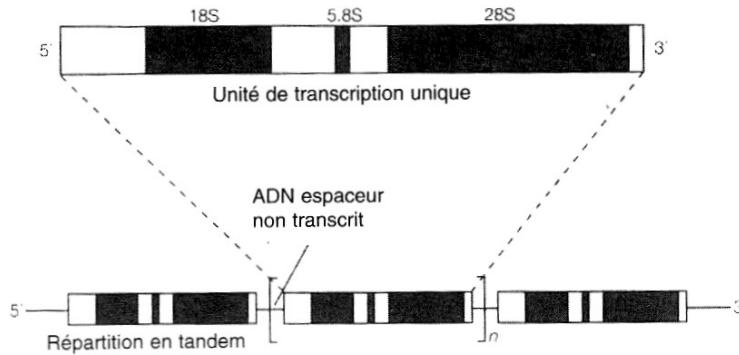
Les trois ARN polymérases eucaryotes reconnaissent différents promoteurs et nécessitent différents facteurs d'initiation appelés TF I, II et III respectivement.

#### III.A. ARN polymérase I et ribosomes



Les ARN ribosomiaux, au nombre de 3, sont distingués par leur vitesse de sédimentation : 28S, 18S et 5.8S. Ils sont codés par des séquences d'ADN répétées en tandem formant 5 groupes d'environ 40 copies disséminées tout au long du génome. Ils forment 70% de la masse totale des ribosomes.

Ces groupes de copies sont organisées en boucles tournées vers une zone centrale du noyau appelée organisateur nucléolaire.



Chaque gène est transcrit en un pré-ARN 45S. Il subira une maturation (voir TP2) en fin de transcription formant les 3 ARNr en quantités égales.

### III.B. ARN polymérase III et transcription de l'ARNt

Facteurs d'initiation nombreux et complexes, se fixant sur des promoteurs répétés le long de l'ADN, et codant pour les boucles spécifiques de l'ARNt. L'ARN polymérase III transcrit également l'ARNr 5S de la grosse sous-unité ribosomiale.

Les ARNt représentent 10 à 15% des ARN totaux dans la cellule.

### III.C. Facteurs de transcription et ARN polymérase II

Le facteur de transcription TF II D (pour facteur de transcription par l'ARN polymérase II) est appelé souvent « facteur TATA » ou TBP (TATA Binding Protein) car il est capable de se lier à une séquence consensus des promoteurs appelée également « TATA box », centrée environ à 25 nucléotides en amont du site de début de transcription.

N.B. : le promoteur des gènes eucaryotes est en général plus long (une centaine de paires de bases) et comporte plusieurs séquences consensus : TATA vers -25, CAAT vers -75 et GC vers -90.

Le TBP est systématiquement requis pour l'initiation. Cette protéine se fixe au niveau du petit sillon de l'ADN à hauteur de la boîte TATA, induisant un changement de conformation local de l'ADN.

D'autres facteurs de transcription (TF II A, B, C...etc.) jouent également un rôle majeur dans le contrôle de la transcription des ARNm. Tous ces facteurs de transcription des Eucaryotes ont donc un rôle similaire au facteur  $\sigma$  des Prokaryotes. Le dernier facteur à se fixer est le TF II H : il s'agit de l'hélicase qui ouvre la double hélice).

La transcription eucaryote se déroule ensuite selon des modalités identiques aux Prokaryotes. Toutefois, la terminaison se réalise par une séquence spécifique sur l'ADN (terminateur), qui recrute une protéine détachant l'ensemble du complexe de transcription.

Le résultat final est un ARN pré-messager, contenant des séquences UTF et une séquence ORF. Cette dernière contient des introns et des exons (voir organisation des génomes), les introns devant être éliminés : c'est la maturation.

### III.D. Maturation des ARN pré-messagers

Les ARN synthétisés par la polymérase II sont appelés ARNnh (nucléaires hétérogènes) ou pré-messagers. Leur fonction ultérieure de messagers impose des modifications covalentes, notamment aux extrémités.

### A.1. Capsulage 5'

Ajout en début de transcription (au bout une trentaine de nucléotides ajoutés) d'un 7 méthyl guanosine par une liaison 5'-5' à l'extrémité 5' donc. Il est ainsi en position inverse par rapport aux autres nucléotides. Cet ajout est nécessaire à la stabilité de l'ARN et à l'initiation de la traduction (reconnaissance par les ribosomes). N. B. : certains virus à ARN+ possèdent aussi cette capsule ; elle semble protéger des ARNs.

### A.2. Clivage 3' et polyadénylation

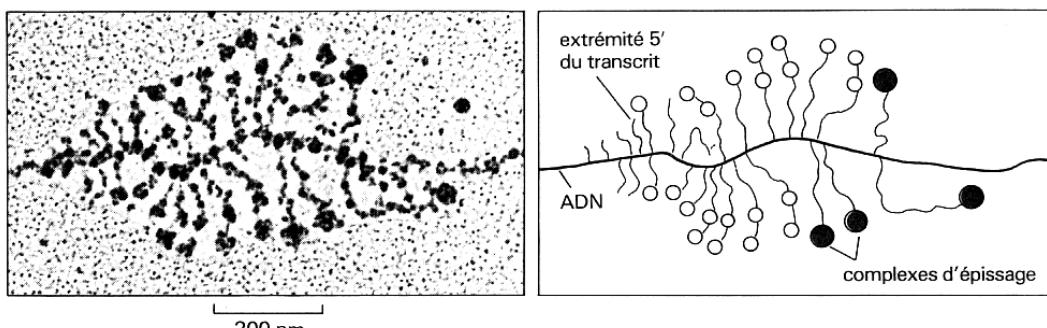
Ajout d'une chaîne polyA de 100 à 250 nucléotides immédiatement après le clivage final (terminaison) au niveau d'un site spécifique. Cette séquence est donc ajoutée après clivage par une polymérase différente.

### A.3. Excision et épissage

La comparaison des séquences du brin d'ADN matrice et d'un ARN ou d'un ADNc synthétisé à partir de la protéine après traduction montre que de nombreuses séquences de l'ADN sont transcrites mais non codantes. En effet, seuls environ 5% de la masse totale de l'ARN synthétisé atteint le cytoplasme, le reste étant dégradé rapidement dans le noyau.

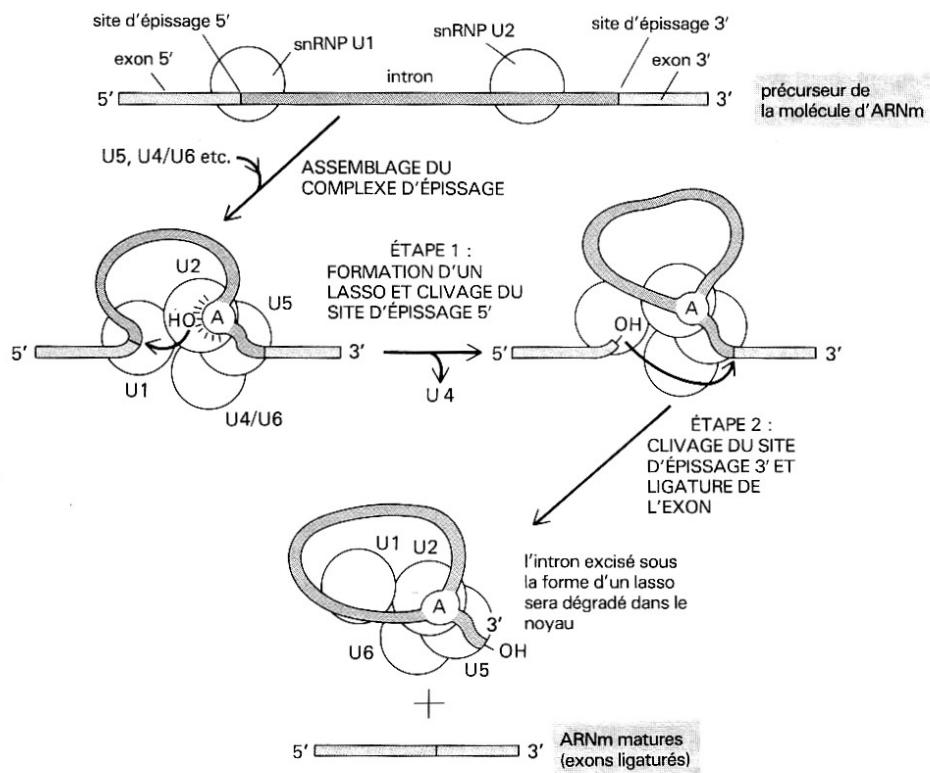
Ces séquences, situées au milieu de la molécule d'ARN pré-messager et éliminées, ont été baptisées introns, alors que les séquences subsistant sont des exons.

L'ARN néosynthétisé (ARNnh) est immédiatement empaqueté dans le noyau, formant des particules denses de protéines et d'ARN appelées particules RNPnh (RiboNucleoParticles Nucléaires Hétérogènes ou Spliceosomes) ou snRNP. Ces complexes sont presque aussi nombreux que les histones et possèdent une structure voisine des nucléosomes. Leur taille est de 25nm environ (moins qu'un ribosome). Ils contiennent des ARN à activité catalytique.



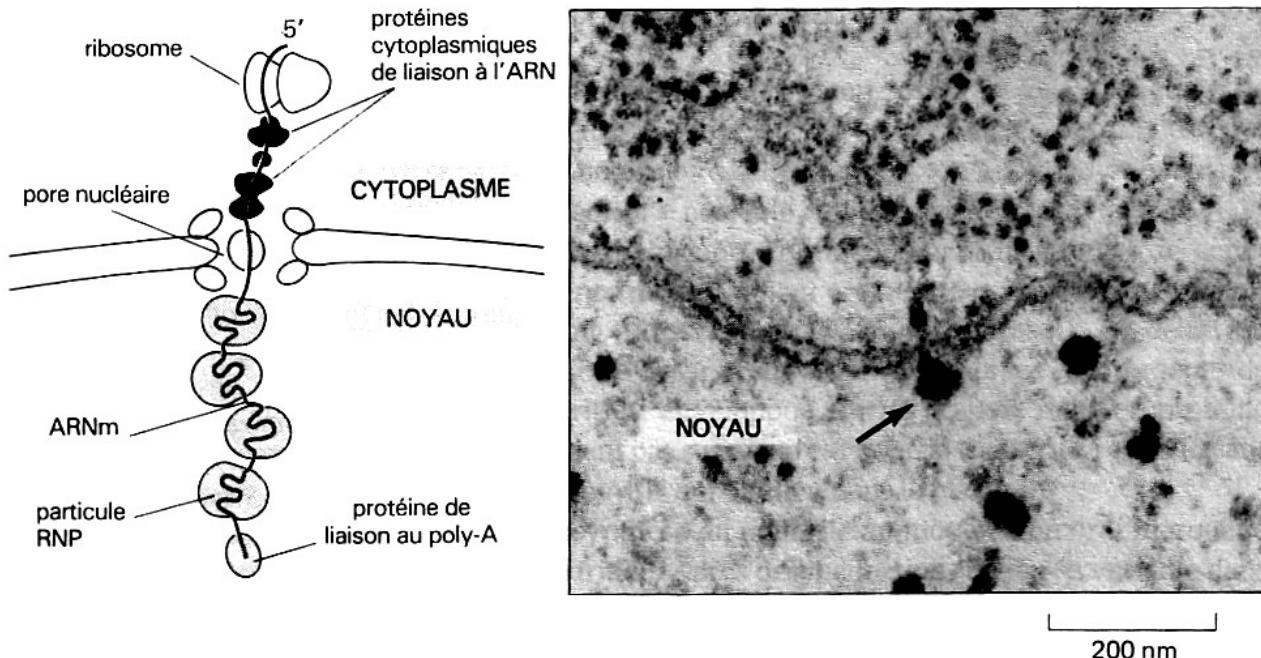
Observation de complexes d'épissages (RNP) au niveau de transcrits

Les introns sont excisés en 3' et 5' au niveau de séquences consensus reconnues par des protéines des RNPnh sous forme de brins d'ARN en lasso et les brins d'ARN restants (les exons) sont épissés. Attention toutefois, tous les gènes ne sont pas morcelés (ex. : histones).



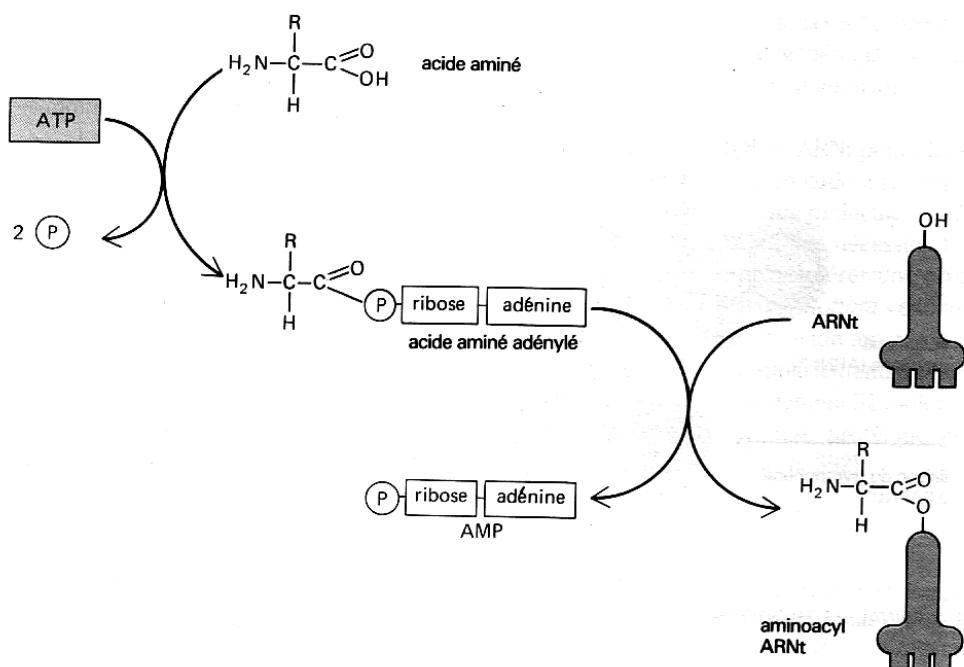
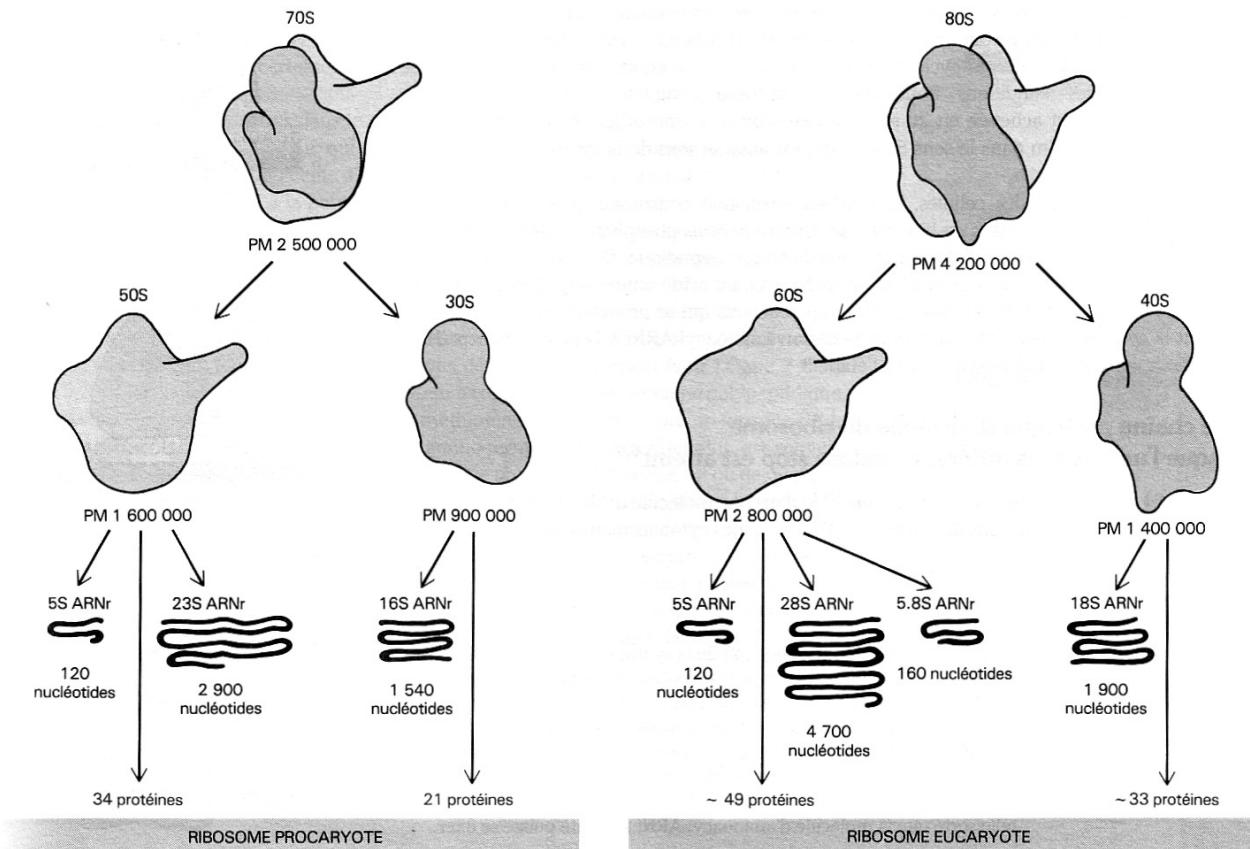
## Mécanisme d'action probable des RNPh au niveau de l'excision des introns et de l'épissage des exons

Le transport des ARN messagers matures jusqu'au cytosol est en général retardé jusqu'à ce que les complexes d'épissage aient terminé la maturation. En effet, **la présence de RNP liées aux ARN pré-messagers empêche toute translocation de ceux-ci à travers les pores nucléaires.**

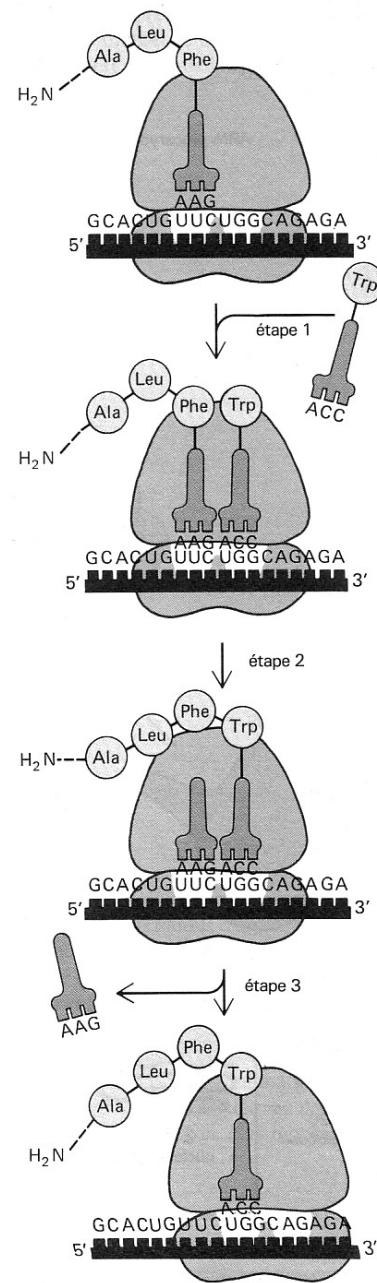
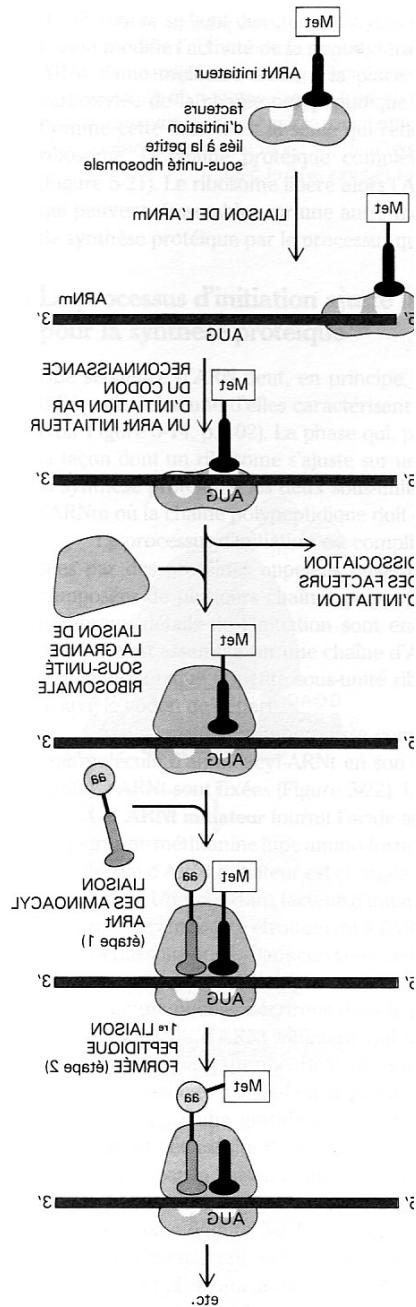


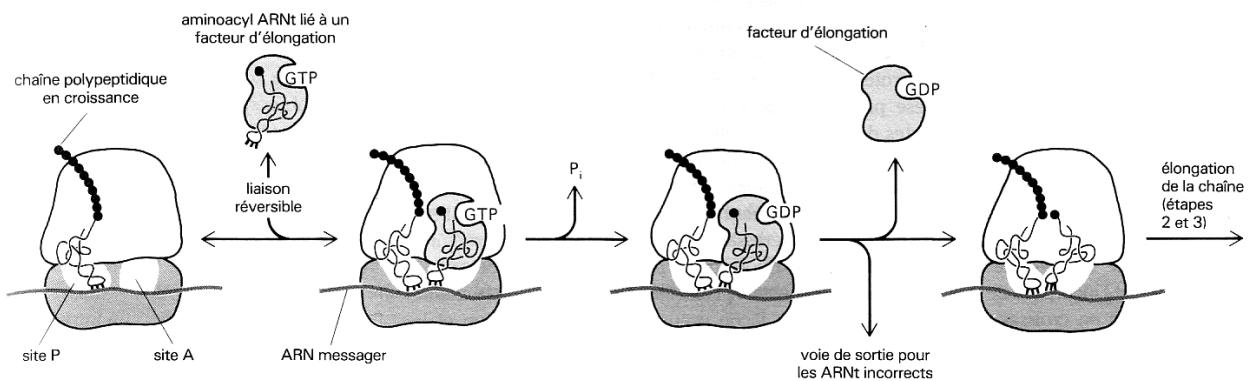
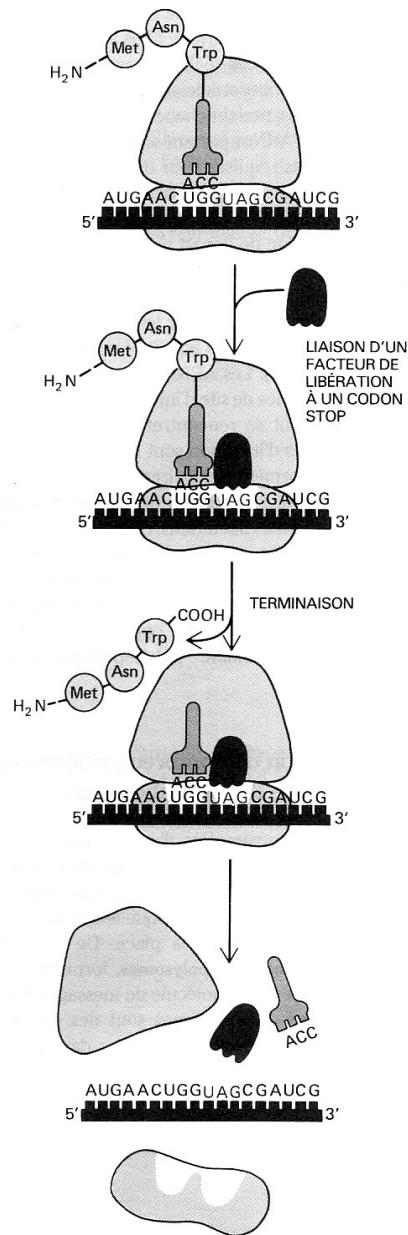
## IV Des ARN au protéines : la traduction et la maturation

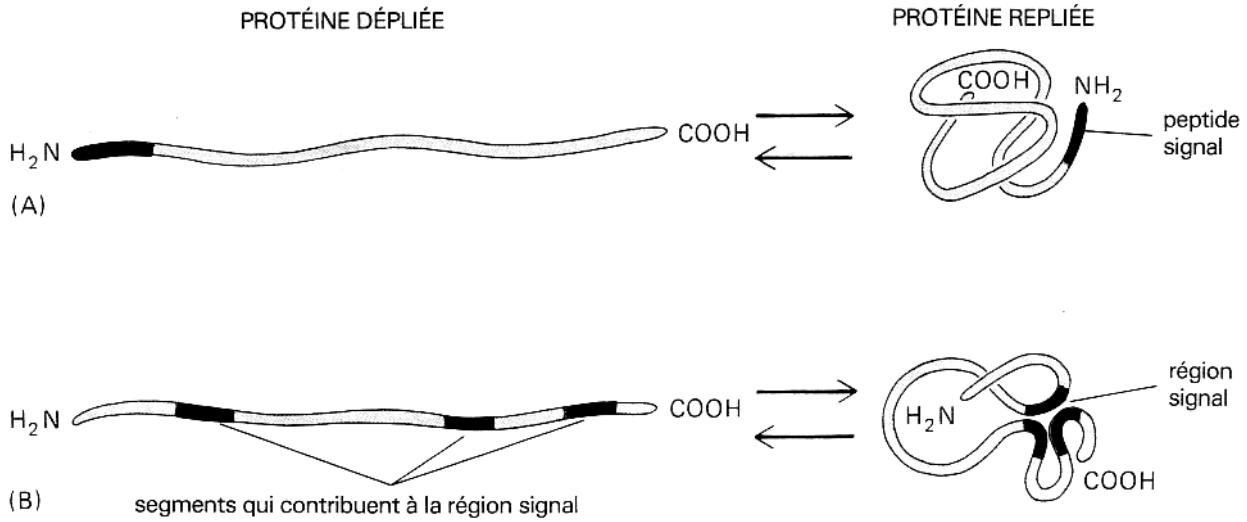
- III.E. A partir d'un exemple : ultrastructure de la cellule acineuse de pancréas
- III.F. Approche expérimentale : l'expérience de Palade
- III.G. Mécanisme de la synthèse et de l'exportation des protéines



Synthèse des aminoacyl-ARNt







Fonction du peptide signal	Exemples de peptide signal
Importation dans le RE	$\text{H}_3\text{N}^+$ -Met-Met-Ser-Phe-Val-Ser-[Leu-Leu-Leu-Val- $\text{Gly}^{\ominus}$ -Ile-Leu-Phe-Trp-Ala]-Thr-Glu-Ala-Glu- $\text{Gln}^{\oplus}$ -Leu-Thr-Lys-Cys-Glu-Val-Phe-Gln- $\text{Lys}^{\ominus}$ -Asp-Glu-Leu-COO $^-$
Maintien dans la lumière du RE	-Lys-Asp-Glu-Leu-COO $^-$
Importation dans les mitochondries	$\text{H}_3\text{N}^+$ -Met-Leu-Ser-Leu-Arg-Gln-Ser-Ile-Arg-Phe- $\text{Phe}^{\oplus}$ -Lys-Pro-Ala-Thr-Arg-Thr-Leu-Cys-Ser- $\text{Ser}^{\oplus}$ -Arg-Tyr-Leu-Leu- $\text{Pro}^{\oplus}$ -Pro-Lys-Lys-Lys-Arg-Lys-Val- $\text{H}_3\text{N}^+$ -Gly-Ser-Ser-Lys-Ser-Lys-Pro-Lys-
Importation dans le noyau	
Attachement aux membranes via la liaison covalente d'un acide myristique à l'extrémité aminée	

## V Le Contrôle de l'expression des gènes

Rappel des principales étapes de l'expression des gènes eucaryotes, de la chromatine à la protéine mature, et présentation des cibles potentielles de contrôle (elles sont très nombreuses)

N. B. : le contrôle de l'expression des opérons procaryotes n'est plus au programme mais mérite d'être cité,

### III.H. L'ADN des Procaryotes est organisé en opérons

#### H.1. Opérons et chaînes métaboliques

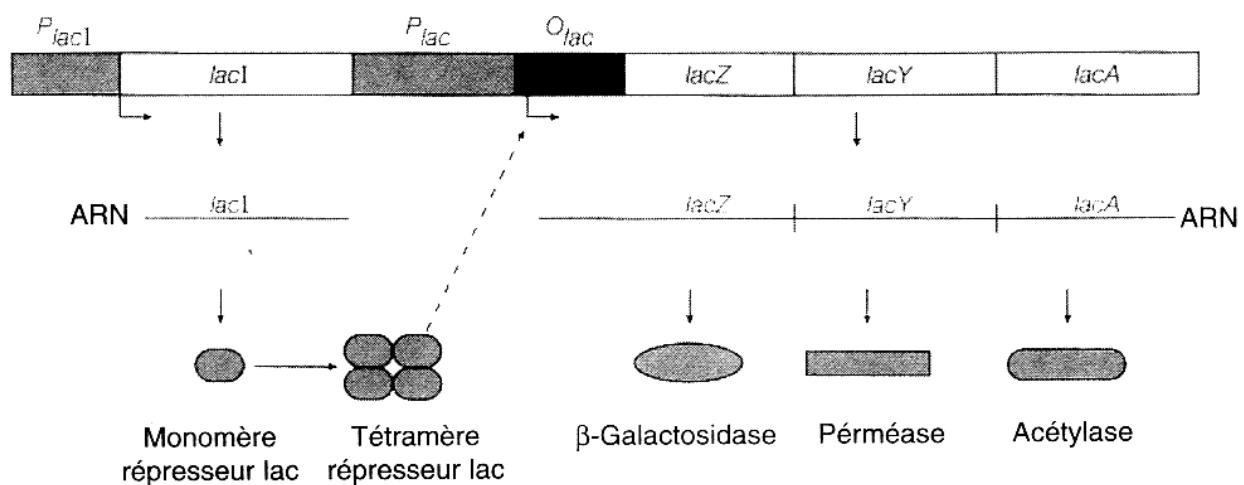
Concept proposé en 1961 par Jacob et Monod. Il consiste en une unité d'expression génétique codant le plus souvent pour plusieurs protéines impliquées dans une voie métabolique et partageant des processus de contrôle communs. Par exemple, l'opéron Lac (pour lactose) code pour plusieurs enzymes Z, Y, A, étant respectivement la  $\beta$ -galactosidase, la galactoside perméase et la thiogalactoside transacétylase. Ces trois enzymes sont nécessaires à l'entrée dans la cellule bactérienne du lactose extérieur et à son utilisation métabolique.

#### H.2. L'opéron Lac est un opéron induit

Entre le promoteur et l'opéron lactose, il existe une courte séquence d'ADN capable de fixer un répresseur. Cette séquence est appelée Opérateur et possède donc des facultés de régulation de l'expression de l'opéron lac.

Le gène *Plac1*, situé en amont de l'opéron, code pour une protéine qui, formant des holotétramères, se fixe sur l'opérateur et empêche l'initiation de la transcription par l'ARN polymérase sans pour autant empêcher sa fixation.

Ce n'est qu'en présence de lactose dans le milieu (donc a fortiori dans la cellule bien qu'en faibles quantités) que le tétramère répresseur fixe le lactose et change de conformation : il se détache alors de l'opérateur et la transcription de l'opéron peut débuter.



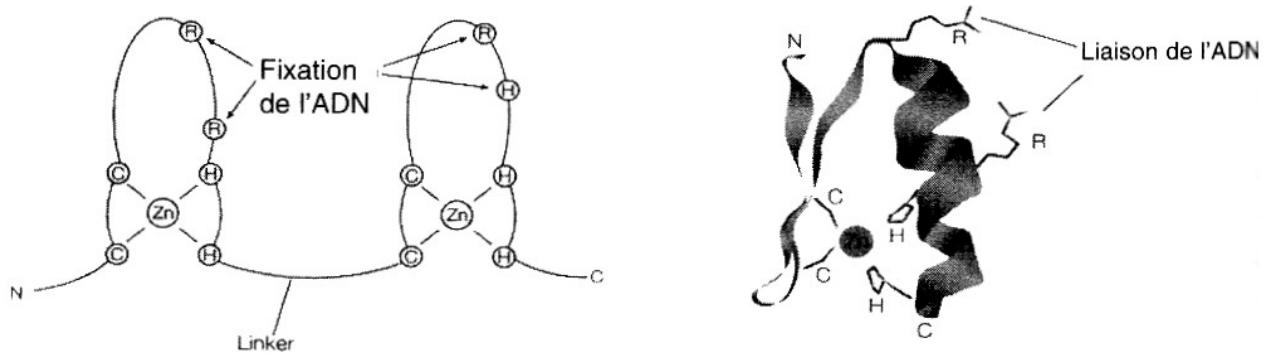
N. B. : l'opéron lactose est un opéron catabolique et son promoteur n'est pas très efficace. L'efficacité de ce dernier peut être renforcée par la présence d'AMPc, dont la concentration est faible en présence d'ATP et de glucose. Ainsi, ma voie catabolique du lactose est réprimée tant que d'autres sources de carbone sont présentes. L'effet de l'AMPc passe par un couple AMPc-protéine CAP qui se fixe sur le promoteur et active les transcriptions, le tout en l'absence de glucose.

Au final, en présence de glucose et absence de lactose, la transcription des 3 gènes est fortement réprimée, elle est modérément activée en présence de glucose et de lactose (50

$\beta$ -galactosidases par cellule) et elle est fortement activée en présence de lactose seul (3000 unités de  $\beta$ -galactosidase par cellule).

### H.3. Motifs tridimensionnels des protéines régulatrices de transcription

Motifs revenant souvent : hélice/tour/hélice (exemple : protéine CAP), hélice/boucle/hélice, doigts de zinc (exemple : récepteur aux hormones thyroïdiennes), fermeture à leucine. Fixation au niveau des grands sillons de l'ADN.



### III.I. Aperçu de la régulation des transcriptions chez les Eucaryotes

La meilleure preuve d'une régulation de l'expression des gènes est bien entendu **l'existence de cellules différenciées dans un même organisme, cellules possédant pourtant toutes le même patrimoine génétique**. En fait, l'immense majorité des protéines présentes dans une cellule (environ 2000 à plus de 50 000 copies chacune, 10 000 à 20 000 en tout) sont présentes dans toutes les cellules d'un organisme. La spécialisation d'une cellule est finalement le résultat de la synthèse importante d'une infime fraction des protéines (exemple du muscle strié).

On imagine aisément qu'il existe divers niveaux possibles de régulation des traductions :

- Au niveau des gènes transcrits, tant pour le moment de la transcription que pour sa fréquence.
- Au niveau de la maturation des ARN, donc des introns excisés.
- Par sélection des ARNm matures qui gagneront le cytoplasme.
- Par sélection dans le cytoplasme des ARNm qui seront traduits ou détruits.
- Par contrôle de l'activité des protéines après synthèse.

Dans le cadre de ce paragraphe, nous nous intéresserons qu'aux trois premiers points.

#### I.1. Protéines régulatrices et activité de transcription

Chaque gène possède en amont, ainsi que nous l'avons vu précédemment, une séquence consensus appelée promoteur qui contrôle le taux et la vitesse à laquelle le gène est transcrit.

Cette activité de transcription peut par ailleurs être modulée par la présence de séquences situées loin en amont (100 à 300 nucléotides) appelées amplificateurs (*enhancers*). De telles séquences ont été découvertes en 1981, à partir du génome d'un virus, le SV 40, dont certains amplificateurs étaient capables de multiplier par 100 l'activité transcriptionnelle d'un gène même s'il était situé 300 paires de bases en amont.

De très nombreuses protéines, appelées protéines régulatrices, sont capables de se fixer au niveau de ces séquences promotrices ou amplificatrices. Chacune de ces protéines régulatrices est présente en quantités infimes dans les cellules, et certaines sont ubiquistes (l'amplificateur SV40) alors que d'autres sont beaucoup plus spécifiques de types cellulaires donnés.

Les très nombreuses études récentes sur ces protéines régulatrices ont livré les conclusions générales suivantes :

- **Les séquences régulatrices sur l'ADN sont spécifiques et courtes : 8 à 15 nucléotides.** Elles sont capables de se lier à une série de protéines régulatrices.
- Certaines protéines régulatrices activent la transcription alors que d'autres l'inhibent. L'effet global d'un élément régulateur dépend de la **combinaison des protéines** qui y sont fixées.
- Les amplificateurs et promoteurs lient le plus souvent les mêmes protéines : ils doivent agir par des mécanismes similaires.
- Un nouvel élément régulateur étudié s'avère le plus souvent influencé par la liaison à des protéines régulatrices déjà connues : ceci suggère que le contrôle de la transcription de gènes eucaryotes est le fait d'un petit nombre de protéines régulatrices.

**Notons enfin que l'activité de ces protéines régulatrices peut être elle-même modifiée, notamment par des facteurs externes comme par exemple les hormones stéroïdes, ou par phosphorylation** (par exemple sous l'influence d'hormones peptidiques agissant par la voie de l'adényl-cyclase).

#### I.2. Contrôle de la maturation des ARNm : l'excision des introns

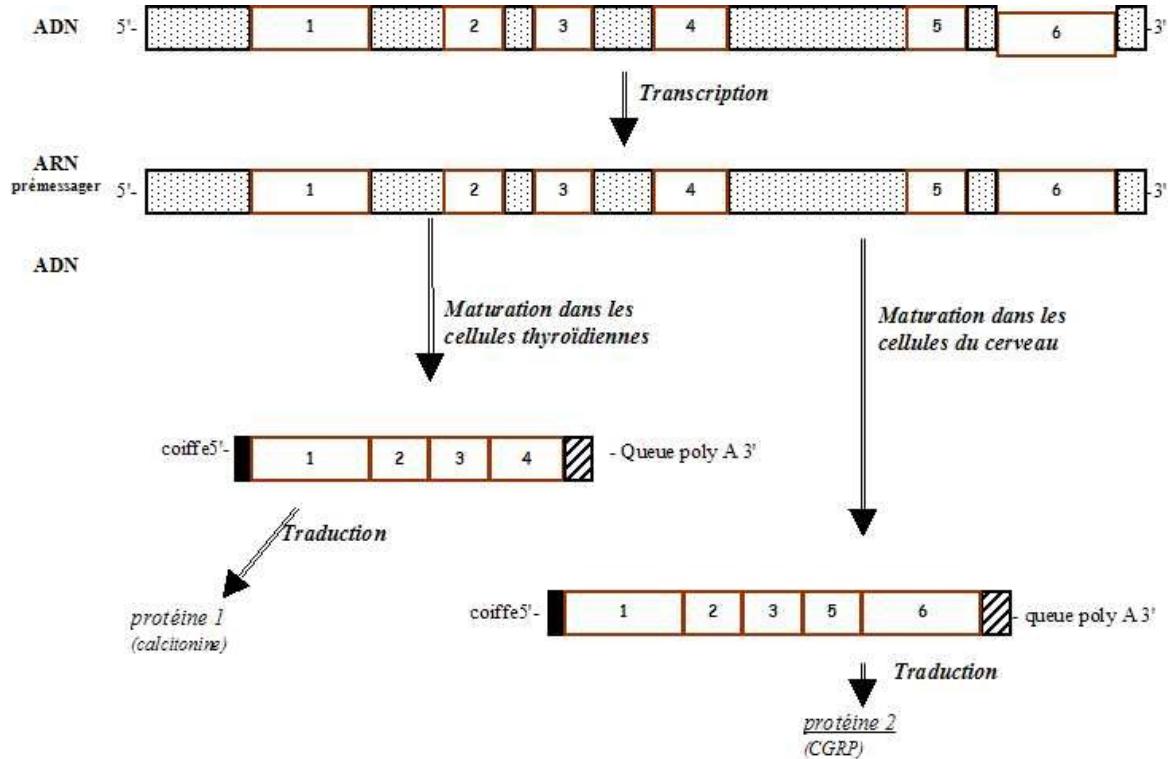
L'épissage fut d'abord découvert chez un virus qui, à partir d'un transcrit primaire, était capable de produire diverses protéines. On parle d'épissage alternatif.

**Dans certains cas, l'épissage alternatif est le fait d'une ambiguïté** dans les appariements possibles des sites d'épissage 5' et 3', de sorte que les choix dans l'épissage sont le fruit du hasard. Cette forme d'épissage alternatif est la cause de formation d'ARN anormaux chez des mutants pour le gène de la  $\beta$ -globine atteints de  $\beta$ -thalassémie. Pour d'autres gènes, une telle ambiguïté est normale et produit plusieurs versions différentes d'une protéine (immunoglobulines par exemple).

**Dans de nombreux cas, l'épissage alternatif est contrôlé**, produisant des protéines différentes dans des cellules différentes à partir d'un même transcrit selon les besoins de l'organisme. De nombreuses protéines sont fabriquées selon ce mode, notamment les myosines.

En général, les changements d'exons ne produisent pas de protéines radicalement différentes, mais des protéines analogues spécifiques de tissus particuliers.

Notons que l'épissage alternatif peut être un mode d'activation ou d'inhibition de l'expression de certains gènes, la protéine fabriquée étant du coup fonctionnelle ou non.



### I.3. Le transport de l'ARN hors du noyau peut être réglé

Un transcript primaire est en moyenne 10 fois plus long qu'un ARNm mature. Or, on peut démontrer que seulement 1/20 des transcrits primaires deviennent matures et gagnent le noyau. Une fraction est donc éliminée dans le noyau après maturation.

On pense actuellement que ce phénomène résulte d'interactions complexes entre l'ARNm et le complexe du pore nucléaire, à travers lequel le passage est un phénomène actif.

### I.4. La traduction peut être contrôlée par de petits ARN : mécanisme de l'interférence.

cf exercice sur les gènes de la floraison (ENS 2007) et l'ARN MIR

## Conclusion