

# La transcription

## I - Principe généraux de la transcription

La transcription est la 1<sup>ère</sup> étape de l'expression génique. Toutes les cellules d'un organisme ont le même génome. Les transcrits produiront les protéines dont la cellule a besoin pour assurer sa fonction.

L'information génétique est stockée dans l'ADN. Cela doit être transcrit, lecture de l'info dans le double brin. Les gènes sont des éléments d'ADN et chacun code pour une protéine particulière. Le génome est un catalogue de gène avec des **espaces inter-géniques**. On a 25 000 gènes dans l'ADN humain. On a des **exons** qui portent le code de l'info (1,5% du génome humain).

La transcription se fait **avant l'interphase**, on ne peut pas transcrire en même temps que la réplication de l'ADN.

### 1) Structure de l'ARN

Les ARN sont des **polymères d'acides nucléiques**. Les nucléotides sont composés d'un **phosphate, d'un sucre et d'une base** (A, U, G et C). L'uracile remplace la thymine de l'ADN. Le sucre est un **ribose** à la différence du désoxyribose de l'ADN. Il est orienté de **5'** (phosphate) vers **3'** (groupement hydroxyle).

L'ARN est **simple brin** et forme des **liaisons hydrogènes** afin de se mettre en conformation 3D. Ils ont une capacité catalytique.

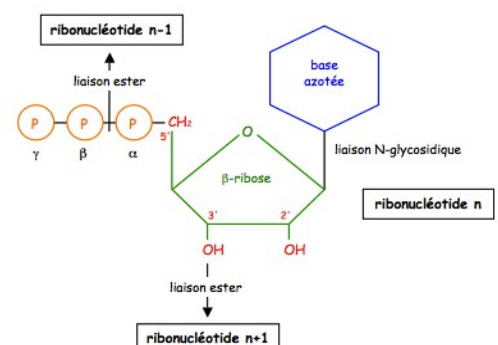
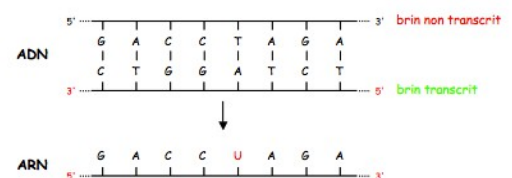
On a 3 types d'ARN : **ARNm, ARNt et ARNr**. L'ARN de transfert participe à la transcription et l'ARNm est celui qui porte l'info.

### 2) Les étapes de la transcription :

L'ARN polymérase vient reconnaître le **promoteur** du brin matrice de l'ADN et s'y fixe ce qui ouvre la double hélice. On a une **polymérisation de 3' vers 5'**. Quand l'ADN est dépareillé on parle de **bulle de transcription**. Cette étape se fait avec le facteur sigma.

L'ARN polymérase doit se désengager du promoteur. Une fois cela fait, elle peut transcrire, c'est l'**élongation** qui se fait de **5' vers 3'** par formation de liaison phosphodiester. La double hélice est suivie par le réappariement des 2 brins.

L'arrêt de la transcription chez les procaryotes fait intervenir des **séquences terminateur** et des **facteurs de terminaison** (facteur rho).



### 3) Quelques différences entre l'ARN procaryotes et eucaryotes

L'ARN poly procaryote transcrit tous les types d'ARN. Chez les eucaryotes il existe plusieurs ARN polymérases spécialisées, l'ARNm doit être exporté du noyau vers l'extérieur après maturation pour être pris en charge par les ribosomes, ce qui n'est pas vrai pour les procaryotes. Les mécanismes de recrutement de l'ARN polymérase sur les promoteurs sont plus complexes chez les eucaryotes.

Chez les eucaryotes :

- **ARN pol I** = transcrit les ARN ribosomique
- **ARN pol II** = transcrit les gènes et synthétise les ARNm
- **ARN pol III** = transcrit les ARNt

Pour les procaryotes on s'intéresse seulement à l'ARN pol II.

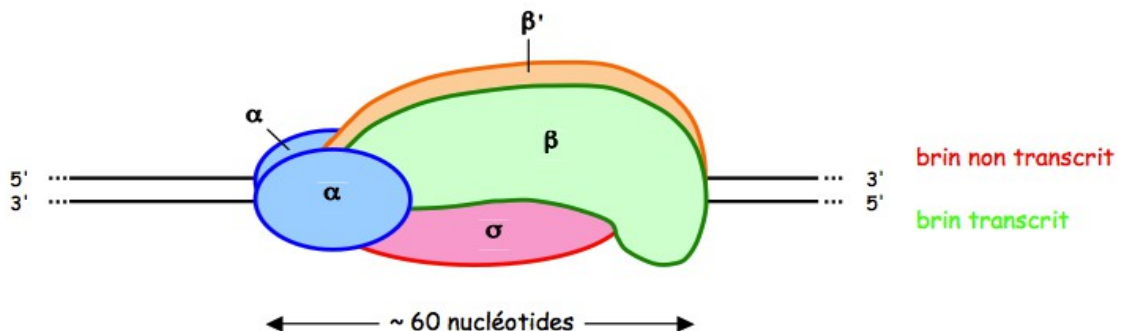
## II - L'ARN polymérase

Cette enzyme a pour rôle de créer la **liaison phosphodiester** entre 2 nouveaux ribonucléotides. Elle est **ARN polymérase-ADN dépendante**. De part sa forte affinité pour l'ADN, elle s'y fixe n'importe où, elle n'a pas de spécificité de séquence.

Elle est composée de plusieurs sous-unités ; chez les procaryotes les sous-unités sont :

- **$\alpha$ 1 et  $\alpha$ 2** ont 2 domaines importants, le domaine N-ter est important pour l'assemblage de l'enzyme et le domaine C-ter est important pour la liaison avec les promoteurs.
- **Oméga** facilite l'assemblage de l'enzyme et stabilise le complexe.
- **$\beta'$ ,  $\beta$**  forment une pince qui guide la molécule.

**Sigma** est un **facteur d'initiation** de la transcription, il augmente l'affinité de l'ARN Polymérase pour le promoteur, permettant à la transcription de démarrer à une position correcte. Elle s'associe avec le facteur sigma et forme une **holoenzyme** d'environ 400 kDa.



## III - Les différentes étapes de la transcription

### 1) Chez les procaryotes

#### a) L'initiation

L'ARN polymérase se combine avec le **facteur sigma** et forment l'**holoenzyme**. Celui-ci se fixe sur l'ADN grâce aux régions  $\beta$  et  $\beta'$  de l'ADN.

Le facteur sigma reconnaît une séquence spécifique de l'ADN appelée promoteur qui se situe au début du gène à transcrire. Chez *E.coli*, les promoteurs ont 2 régions importantes : l'une à -10 nucléotides qui a la **séquence TATATT** (boîte TATA), l'autre à -35 nucléotides ayant la **séquence TTGACA**. La transcription commence à un **nucléotide +1** (site TSS), puis la séquence codante arrive.

Le **facteur sigma** contient **4 domaines d'activités enzymatiques**. Les domaines 2 reconnaît l'ARN poly et s'y fixe et le 4 est la liaison à l'ADN. On parle du **facteur  $\sigma$  70** car il reconnaît le promoteur de tous les gènes quand il est fixé sur l'ARN polymérase (protéines orthologue).

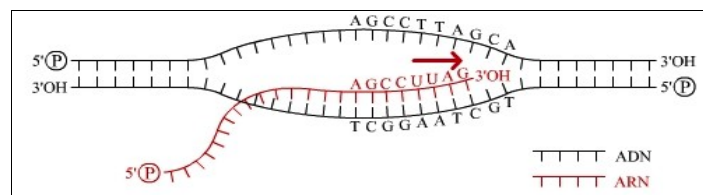
Une fois fixé, et le promoteur reconnu, l'holoenzyme sépare les 2 brins d'ADN (dénaturation) cela entraîne une ouverture dans la région -10 nucléotides. Cette ouverture permet à l'holoenzyme de présenter le 1<sup>er</sup> nucléotide complémentaire du brin codant.

### L'échappement du promoteur

Lorsque l'élongation atteint une dizaine de nucléotides, ce complexe est stabilisé. La phase d'initiation prend fin avec la **libération du facteur sigma**. On a une **dissociation stochastique** dû à une plus faible affinité de sigma pour l'ARNP. Cela permet à l'ARN polymérase d'engager l'élongation.

### b) L'élongation

Plusieurs ARN polymérases transcrivent simultanément le même gène. La polymérisation a une vitesse moyenne de 50-60 nucléotides par seconde. Dans la bulle de transcription, l'ARN est hybridé à l'ADN. On a une fidélité de la transcription et l'ARNp a aussi une activité de correction.



### c) La terminaison

L'ARN pol reconnaît des séquences spécifiques afin d'arrêter la transcription.

#### ➤ La terminaison intrinsèque :

A la fin du gène il existe 1 **séquence d'ADN** particulière **palindromique**. On a un signal de terminaison G-C, en forme de **boucle** dans la bulle de transcription, qui est suivit de plusieurs A donnant une série de U, ce qui baissera l'énergie de part ces 2 liaisons hydrogènes. L'hybride ARN-ADN ne sera plus stable. La transcription s'arrête.

#### ➤ La terminaison rho dépendant :

Certains gènes ont des sites de terminaison différents. La molécule d'ARN est prise en charge par les ribosomes des procaryotes. Ces gènes sont reconnus par une protéine rho qui libère l'ARN de l'ADN en utilisant l'ATP. Elle se rembobine jusqu'à l'ARN polymérase. La molécule d'ARN s'en va et la transcription s'arrête.

## 2) Chez les eucaryotes

### a) Différents complexes d'enzymes

Les eucaryotes ont différents complexes d'enzymes, la reconnaissance des promoteurs est différente et l'étape d'élongation se fait **en couplage** avec la maturation de l'ARNm.

L'ARN polymérase II possède 12 sous-unités (Rpb). Rpb1 et Rpb2 sont homologues et forment le site catalytique. On a 4 sous-ensembles : une contenant Rpb1, Rpb2, Rpb3 et la tige.

L'ARN pol II ne reconnaît pas le promoteur toute seule, elle effectue ce travail en compagnie de nombreux co-facteurs protéiques qui se recrutent les uns les autres et qui forment avec elle un complexe d'initiation. Ce sont les protéines enzymatiques GTF, appelée (TFII). Une fois assemblé ils recrutent l'ARN polymérase.

### b) Complexe d'initiation (PIC)

**TFIID** reconnaît le promoteur et sa sous-unité **TBP** reconnaît la boîte TATA ce qui pliera l'ADN afin de recruter d'autres facteurs. (*TBP en association avec les facteurs ancillaires TAF forment le complexe TFIID*).

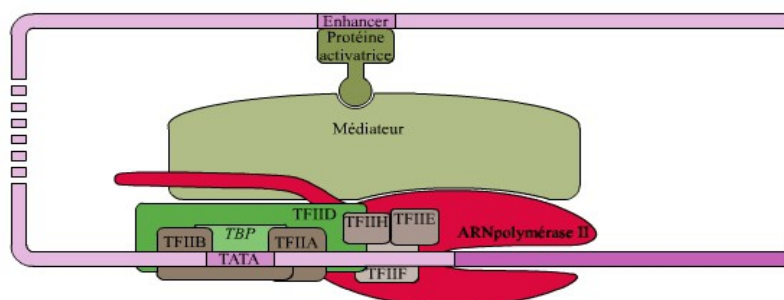
**TFIIA** (stabilise le complexe), **TFIIB**, **TFIIE** (ouverture de la bulle), **TFIIF** (même rôle que facteur sigma), et **TFIIH** se fixent ensuite au TFIID et à la région du promoteur. (*TFIIF est une protéine qui guide l'ARN polymérase II vers le début du gène à transcrire. TFIIH sera importante pour libérer l'ARN du promoteur, elle a une activité hélicase*). Ceci forme le **complexe de pré-initiation** de la transcription.

Ce complexe a besoin d'un co-activateur : **médiateur**. Il permet l'assemblage de l'appareil transcriptionnel. Les **séquences enhancer**, situé en amont du promoteur, augmente le taux de transcription et active la protéine. Il existe d'autres boîtes : TFIIB reconnaît la boîte **BRE** via ses liaisons G-C avant le boîte TATA. Puis on a la boîte **DPE et INR** reconnut par le TFIID.

Il y a des gènes qui n'ont pas de boîte TATA mais ont à la place une boîte GC avec une protéine Sp1 qui se fixe à la boîte et recrute la TBP qui recrutera le reste du complexe de pré-initiation.

### Initiation

Elle commence, après assemblage de tous les facteurs, par l'ouverture de l'ADN au niveau du site d'initiation (bulle), puis allongement de celle-ci et elle finit par le relargage des facteurs d'initiation. La **queue CTD** est phosphorylée au début de l'initiation.



## Elongation et terminaison

On a des facteurs protéiques d'élongation qui facilite la progression de l'ARN pol II au travers d'une chromatine. Un **ARN pré-messenger complémentaire du brin matrice** de l'ADN (antisens), identique au brin codant de l'ADN (sens) commence à être synthétisé selon la direction **5'-3'**.

Au cours de l'élongation, l'ARN polymérase II rencontre des facteurs protéiques assurant la **maturation** des ARN avant qu'ils quittent le noyau pour aller vers le cytoplasme.

La terminaison est différente des procaryotes car il n'y a **pas de terminateur**, l'ARN polymérase libère l'ARNpm qu'elle vient d'assembler.

### c) Maturation de l'ARN messenger

La molécule d'ARNm subit plusieurs modifications avant de quitter le noyau. La maturation a lieu pendant l'élongation.

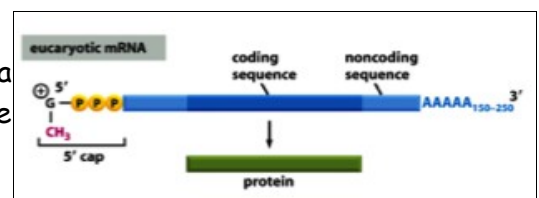
#### ➤ Ajout d'une coiffe en 5'

La coiffe se caractérise par l'**ajout d'une guanine méthylée** en C7 à l'**extrémité 5'** du brin d'ARN. Celui-ci est relié par une **liaison anhydride d'acide** (au lieu d'une liaison phosphodiester). Cela permet d'avoir un GMP à l'extrémité au lieu de phosphates.

#### ➤ Ajout d'une queue poly AA en 3'

Cela permet de couper la chaîne d'ARN au niveau d'une séquence particulière de nucléotides et de rendre l'**ARNm mature**. L'ARN transcrit **riche en GU** enlève la dernière région en 3' (qui sera dégradée) puis ajoute une queue poly AAA.

Ces 2 modifications augmentent la stabilité de la molécule d'ARNm eucaryote. Mais la plupart subissent une modification supplémentaire pour être fonctionnel.



#### ➤ L'épissage

C'est l'étape de maturation des ARN au cours de laquelle une séquence non codante interne de l'ARN est enlevée. Le gène est composé d'**introns** qu'il faut enlever et d'**exons**, qui sont des séquences codantes, sont reliés entre eux.

Après l'addition de la coiffe, l'épissage commence. Au sein d'un ARNm les deux premières bases faisant suite à un exon sont toujours GU et les deux dernières bases de l'intron sont AG.

L'ARNm subit une **excision des introns** grâce à leur site donneurs en 5' et accepteurs en 3' suivit d'un **épissage**, c'est-à-dire la réunion des exons restants qui constituent l'ARNm. La conséquence de l'excision des introns est la formation d'un **lasso** et l'obtention d'une extrémité 3'OH du premier exon. *L'exon 1 réagit avec l'extrémité 5' de l'exon 2 permettant l'épissage des 2 exons et la libération du lasso qui sera dégradé par les ribonucléases.*

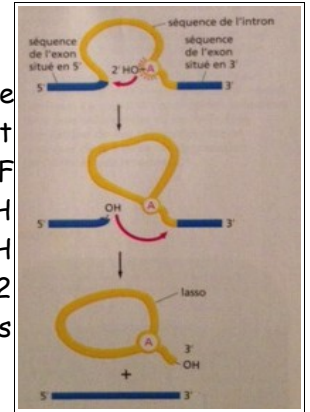


L'intron est libéré sous forme d'un petit ARN circulaire qui vont ensuite être dégradés par l'activité exonucléase des ARN polymérases.

L'épissage n'est pas toujours constitutif, il y a des **épissages alternatifs** dans lequel l'élimination des introns peut faire se lier entre eux des exons différents. Un tel épissage peut produire de nombreuses protéines différentes à partir d'un même gène.

L'épissage de l'ARN est effectué par des molécules d'ARN appelées **petits ARN nucléaire** (ARNsn) et sont regroupés pour former de petites particules nucléaires ribonucléoprotéique, les **RNPsn**. Ces RNPsn forment le **spliceosome**.

Au cours de la 1ère étape, reconnaissance du site 5' par la **RNPsn U1**. Une Adénine du point de branchement (rouge) de l'intron attaque l'extrémité 5' du point d'épissage et coupe le squelette sucre-P de l'ARN en ce point. **BBP et U2 AF** reconnaît le A. L'extrémité 5' coupée de l'intron se lie de façon covalente au 2'OH du ribose de l'Adénine A pour former une structure branchée. L'extrémité 3'OH libre de la séquence de l'exon réagit avec le début de l'exon suivant, reliant les 2 exons ensemble pour former une séquence codante continue et libérant l'intron sous forme d'une structure en lasso qu sera ensuite dégradée.



Il peut se produire des erreurs lors de l'épissage

- Le spliceosome ne reconnaît pas un exon. C'est pas régulé
- Mauvaise reconnaissance des sites d'épissage : élimination d'un exon par erreur

Ces erreurs ne se produisent pas souvent grâce à un mécanisme de sortage de ces erreurs. Dans le génome eucaryote, les exons ont tous à peu près la même taille, ce n'est pas le cas pour les introns. C'est donc **la taille qui permet de reconnaître un exon**. La protéine SR se fixe sur un exon. C'est la quantité de SR fixé qui définit la taille de l'exon, et U1 est recrutée. D'autres protéines se fixent sur les introns : hnRNP pour définir leur taille.

Autres type de spliceosome : Épissage en **trans** 2 gènes peuvent s'échanger des exons. Finalement certains ARNm sont capable de faire de l'auto-épissage sans avoir besoin de spliceosome.

(**CstF** : facteur qui favorise la coupure et **CPSF** : Facteur qui active la coupure + ajout de la queue)

CstF et CPSF permet le recrutement de facteurs additionnel pour couper l'ARNm. Puis recrutement de la poly-A polymérase pour ajouter la queue poly AAA. Elle n'a pas besoin de matrice. La queue poly A est une série de A en 3' qui permet le recrutement de protéine qui se fixent sur la queue. La taille de la queue est régulée, et selon sa taille, l'ARNm n'aura pas le même destin.

On a fabrication d'une molécule ARNm matrice qui sort du noyau grâce à la reconnaissance par des pores nucléaires.

## IV - Régulation de la transcription

En plus du promoteur, presque tous les gènes, qu'ils soient bactériens ou eucaryotes, contiennent des **séquences ADN régulatrices**, utilisées pour activer ou inhiber le gène. Ces séquences doivent être reconnues par des **protéines régulatrices de la transcription**, qui se lient à l'ADN.

Dans la plupart des cas, la protéine s'insère dans le grand sillon de la double hélice d'ADN et établit une série de contacts moléculaires avec les paires de bases.

## 1) Chez les procaryotes

Un **opéron** est ensemble linéaire de gènes de structure dont le fonctionnement est contrôlé par un promoteur et un opérateur. L'activité de l'opéron est déterminée par une molécule répresseur dont la synthèse dépend d'un gène régulateur séparé de l'opéron.

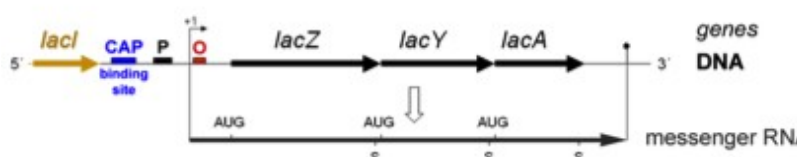
Le contrôle coordonné de l'expression de ces gènes est possible grâce à des **protéines régulatrices**, appelées **facteurs de transcription**. Elles régulent le taux de transcription en se liant à l'ADN au niveau de séquences spécifiques. Dans ce cas, l'ARNm est **polycistronique**, c'est-à-dire qu'il contient l'information nécessaire à la synthèse des différentes protéines.

Les **protéines de régulation** reconnaissent le promoteur qui contient la courte séquence d'ADN. Quand cette protéine se lie à cette séquence nucléotidique, appelée **Opérateur**, elle bloque l'accès de l'ARN polymérase au promoteur, ceci empêche donc la transcription de l'opéron et la production des enzymes responsables de la transcription.

La bactérie trouve sa source de carbone dans le catabolisme des sucres. Si le glucose est la source de carbone "préférée", le **lactose** (qui est un  $\beta$ -galactoside) peut également être consommé par la bactérie et métabolisé en galactose et glucose. Les enzymes nécessaires à l'utilisation du lactose ne seront synthétisées qu'en **présence de ce substrat**. Le lactose a besoin d'une  $\beta$ -galactosidase et d'une perméase.

Dans l'opéron lactose, on trouve les 3 gènes indispensables à la dégradation du lactose. Ils codent :

- *lac Z* code pour la  **$\beta$ -galactosidase** ==> dégradation du lactose en glucose et galactose
- *lac Y* code pour la **lactose perméase** ==> permet l'entrée du lactose dans la cellule.
- *Lac A* code pour la **transacétylase** ==> Pas nécessaire pour la  $\beta$ -galactosidase



**Mutants auxotrophes :**

**LacZ<sup>-</sup>** : la  $\beta$ -galactosidase n'est pas fonctionnelle => pas de dégradation du lactose.

**LacY<sup>-</sup>** : la  $\beta$ -galactoside perméase n'est pas fonctionnelle => pas d'absorption du lactose à l'intérieur de la cellule.

**Lac A<sup>-</sup>** : la  $\beta$ -thiogalactoside acétyltransférase n'est pas fonctionnelle => pas nécessaire pour le métabolisme du lactose.

Mutant	- inducteur	+ inducteur
WT	-	+
Lac P <sup>-</sup>	-	-
Lac O <sup>c</sup>	+	+
Lac I <sup>-</sup>	+	+
Lac I <sup>S</sup>	-	-

•Inducteur : lactose ou IPTG en absence de glucose

•WT: phénotype inductible sauvage

•+: colonie bleue : synthèse de la  $\beta$ -Galactosidase => opéron transcrit

• -: colonie blanche : pas de synthèse de la  $\beta$ -galactosidase => opéron non transcrit

•O<sup>c</sup>: mutation constitutive

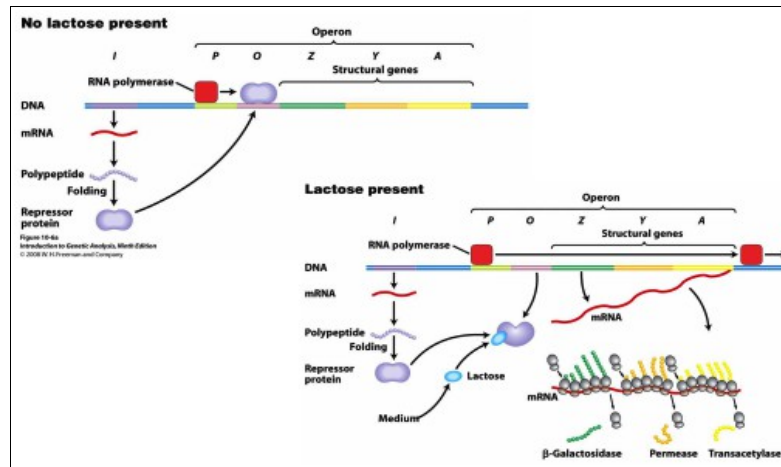
•I<sup>-</sup> et P<sup>-</sup>: mutations nulles

•I<sup>S</sup>: mutation superrepresseur gain de fonction

Ces trois gènes de structure sont précédés par une région responsable de la régulation de leur expression. Cette région régulatrice comprend le **promoteur P** et l'**opérateur O**. On trouve également en amont de l'opéron lactose, le **gène régulateur** (*lacI*) qui code un répresseur. Celle-ci inhibe l'expression des gènes de l'opéron lactose par trans-activation en se liant spécifiquement sur l'ADN au niveau de l'opérateur. L'expression de ce répresseur est constitutive, il est exprimé quelque soient les conditions de croissance de la bactérie.

En **absence de lactose**, la molécule de répresseur se fixe sur l'opérateur et empêche le fonctionnement de l'ARN-polymérase.

En **présence de lactose**, c'est l'**allolactose**, un isomère du lactose, qui va jouer le rôle d'**inducteur** celui-ci se lie au répresseur et l'inactive. Le déplacement de l'ARN-polymérase permet la synthèse d'un ARN messager polycistronique.



L'activité d'un promoteur peut souvent être contrôlée par 2 régulateurs de la transcription différents. L'opéron Lactose de E.Coli, par exemple est contrôlée par le **répresseur Lac** et par la **protéine activatrice CAP**.

L'inactivation du répresseur pour des systèmes comme le lac se nomme induction. Les dérivés du lactose qui inactivent le répresseur et provoquent l'expression des gènes lac sont des **inducteurs**. Le système lac représente un exemple du contrôle négatif puisqu'il repose sur l'inactivation du répresseur.

Il existe un contrôle positif : si le lactose et le glucose sont présents dans le milieu, il n'y a pas de dégradation du lactose tant que tout le glucose n'a pas été utilisé. Si la concentration du glucose décroît, la concentration de l'AMP cyclique augmente. Une forte concentration est nécessaire à l'activation de l'opéron lactose. Cet AMPc forme un complexe avec la protéine CAP (ou CRP), (CAP ne se fixe pas s'il y a trop de glucose). Ce **complexe CAP-AMPc** se lie à l'ADN en amont du site de fixation de l'ARN polymérase. Ce complexe agit comme un **inducteur**. CAP augmente l'affinité de l'ARN poly pour le promoteur ce qui augmente l'expression.

Certaines mutations peuvent exercer leur action sur d'autres gènes adjacents dans l'opéron. Un effet connu : la **dominance en cis**. Elle implique l'interaction physique d'un élément, avec des gènes directement en contact avec lui. Par opposition la **dominance en trans** implique l'action d'un produit diffusible.

La mutation  $I^s$  élimine la réponse à l'inducteur en modifiant le site de fixation de l'inducteur.

## Principe de régulation génique

L'ADN poly se fixe au promoteur avec l'aide du **facteur sigma70**. Une protéine se fixe sur l'ADN en présence d'un ligand, ici CAP, se fixe quand il y a de l'AMPc.

### 1) Recrutement régulé

La protéine activatrice et l'ARN poly doivent se toucher. Si x reconnaît y, cela suffit-il à induire la transcription ? Oui c'est le cas pour la protéine CAP et l'ARN poly.



## 2) Activation de la polymérase

L'holoenzyme ARN polymérase +  $\sigma 70$  se fixe de manière stable (complexe fermé) sur des gènes spécifiques du métabolisme de l'azote mais sans initier de transcription. **L'activation de l'ARN polymérase nécessite un changement de conformation** de l'holoenzyme, induit par un activateur.

L'holoenzyme reconnaît son promoteur et  $\sigma 54$  est fixé sur l'ADN mais ne déclenche pas la transcription. Pour qu'il y est transcription, il faut que **l'holoenzyme ait un changement de conformation**. C'est un facteur qui induit ce changement de conformation.

Une **protéine activatrice NtrC** se fixe sur sa zone régulatrice mais rien ne se passe. L'azote phosphoryle la protéine NtrC ce qui va induire l'interaction entre la poly et l'activateur NtrC.

Le contact entre NtrC et l'ARN poly induit un changement de conformation de l'ARN poly ce qui induit la transcription. La fixation de NtrC sur l'ADN n'est pas obligatoire mais il faut alors une très forte concentration d'azote pour induire la transcription.

## 3) Activation du promoteurs

A certains promoteur (gène de résistance au mercure) l'ARN polymérase +  $\sigma 70$  se fixe également de manière stable mais inactive. **L'activation de la polymérase requiert que le promoteur subisse un changement de conformation**. Ce changement est induit par un activateur.

## 2) Chez les eucaryotes

Tout interagit (élément régulateur) grâce a un complexe médiateur (plusieurs protéines). La levure est capable d'utiliser le galactose comme source de C. Elle le transforme en glucose. Elle possède donc les gène gal.

Si elle a du glucose, elle n'a pas besoin de dégrader le galactose mais s'il y a que du galactose, elle l'utilise. Il existe une protéine activatrice **gal4** du gène gal qui se fixe sur la séquence **UAS** (séquence d'activation en amont). (2 domaines dans la protéine gal4 un pour se fixer a l'ADN et un domaine régulateur.) Il faut que l'activateur soit pas loin du promoteur. La communication se fait entre l'activateur et les facteurs généraux de transcription.

Chez les eucaryotes, il va y avoir pour un gène donné plusieurs séquence régulatrices auxquelles se fixent plusieurs protéines régulatrice. La fixation des protéines régulatrices sur l'ADN induit des co-activateurs par dessus et recrute la fixation de l'ARN poly.

L'ADN va devoir se replier pour que tous ces facteurs interagissent correctement. Principe de combinatoire de facteur de transcription. La régulation transcriptionnelle peut être étudié lors du développement.

