**CHAPITRE IIV : LA TRANSCRIPTION**

1. **Définition**

La transcription est un processus biologique ubiquitaire qui consiste, au niveau de la cellule, en la copie des régions dites codantes de l'ADN en molécules d'ARN. En effet, si la molécule d'ADN est le support universel de l'information génétique, ce sont les molécules d'ARN qui sont reconnues par la machinerie de traduction en séquences protéiques.

Les mécanismes généraux qui assurent la transcription des ARN (messager, ribosomique, de transfert, etc.) sont globalement identiques. Par contre, l'organisation des gènes codant ces différents ARN et le contrôle de leur expression sont très différents d'un type de gêne à l'autre.



[**Schéma global de la fabrication d'une protéine**](http://www.ebiologie.fr/upload/s/174/schema-global-de-la-fabrication-d-une-proteine)

1. **Mécanismes-généraux

1)Différentes-étapes**

Une unité de transcription s'étend toujours d'un site d'initiation de la transcription jusqu'à une région de terminaison de la transcription. Le mécanisme général de la transcription pourra donc être divisé en 3 étapes distinctes
* **l'initiation de la transcription** : reconnaissance du début de l'unité de transcription
* **l'élongation de la chaine ribonucléotidique** : polymérisation de la chaine d'ARN
* **la terminaison de la transcription** : nécessite la reconnaissance de la région de terminaison.

L'ensemble de ces 3 étapes constitue la transcription, elle est réalisée par une même enzyme : **l'ARN polymérase** qui est une enzyme complexe et très différente entre procaryotes et eucaryotes.
Toutes les ARN polymérases possèdent la même activité, fonctionnent globalement de la même façon. Ce sont des enzymes qui catalysent la synthèse d'une même molécule d'ARN à partir d'une matrice d'ADN et elles réalisent toutes sans exception l'élongation de la chaine nucléotidique dans le sens 5'3' (attention : un seul des deux brins d'ADN est transcrit).
 **2) Notion de brin matrice et brin codant**

Dans une unité de transcription, un des deux brins va servir de matrice à l'ARN polymérase, on l'appelle brin matrice. C'est sur ce brin que l'ARN polymérase va se déplacer et synthétiser la copie complémentaire de ce brin matrice qui sera donc la copie conforme du brin opposé : le brin codant (à l'exception de la Thymine qui sera de l'Uracile).



**brin codant et brin matrice**

**Attention** : sur la même molécule d'ADN on pourra trouver une unité de transcription dans le sens inverse, le brin matrice devient donc codant et vice-versa



1. **Les ARN polymérases : structure et fonctionnement

1) Propriétés générales**

Ces ARN polymérases sont très complexes dont la structure précise est très variable selon les espèces. On peut distinguer 4 grands types structuraux d'ARN polymérases :
- un type "procaryote"
- 3 types "eucaryotes"

Ces différentes ARN polymérases étudiées in vitro travaillent globalement de la même façon. Elles ont besoin pour fonctionner d'une matrice d'ADN (double brin préférentiellement), des 4 précurseurs ribonucléotidiques (ATP, UTP, CTP et GTP) et enfin d'un cofacteur apporté sous forme d'ions Mg2+. A elles seules, les enzymes sont capables de faire les 3 étapes de la transcription. Elles doivent donc avoir plusieurs propriétés.

**Initiation de la transcription :**

- reconnaissance du site d'initiation
- liaison à l'ADN
- ouverture locale de la double hélice

**l'ARN polymérase**

**Elongation** :
- activité ARN Pol



**Transcription de l'ADN**

**Terminaison** :
- reconnaissance du site de terminaison



L'ARN polymérase est une enzyme à structure quaternaire formée par un assemblage de plusieurs sous unités polypeptidiques identiques ou différentes.

**2) Cas des procaryotes**

Ici, l'ensemble des gènes (codant ARNm, ARNt, ARNr) est transcrit par la même ARN polymérase. Cette enzyme est variable en fonction des espèces de bactéries considérées, mais globalement son organisation et son fonctionnement reste à peu près comparable dans l'ensemble des procaryotes.

Chez E.Coli son poids moléculaire est de 390 000d, et est composée de 5 types de sous unités différentes.

|  |  |  |
| --- | --- | --- |
| **Sous-unité** | **PM** | **Fonction** |
| **Alpha (α)** | 2 36 500 | Liaison à l'ADN |
| **Beta (β)** | 1 150 000 | Polymérisation ARN |
| **Beta prime (β')** | 1 155 000 | Liaison à l'ADN |
| **Sigma (σ) (plusieurs types)** | 1 70 à 80 000 | Initiation |
| **Omega (ω)** | 1 11 000 | Inconnue |

La sous unité sigma est à fixation réversible, l'enzyme peut donc exister sous deux formes :
- Core enzyme = α2 β β' (ω)
- Holoenzyme = α2 β β' (ω) σ



**Core enzyme et Holo enzyme**

Il existe plusieurs types de facteurs σ qui permettent l'initiation de la transcription de différents types de gènes.

**3) Cas des eucaryotes**

Trois ARN polymérases distinctes spécialisées chacunes dans la transcription d'un type de gène :
- **ARN Pol I** : transcription des gènes de classe I (ARNr précurseurs de 5,8S ; 18S et 28S)
- **ARN Pol II** : transcription des gènes de classe II (ARNm et snRNA = Small Nuclear RNA)
- **ARN Pol III** : transcription des gènes de classe III (ARNr 5S ; ARNt et scRNA = Small Cytoplasmic RNA)

**-RNA-polymérase IV** : spécialise dans la transcription de l’ADN mitochondrial et la synthèse de l’hétérochromatine chez les plantes.

**Structure Quaternaire très complexe et beaucoup moins connue que chez les procaryotes** : la structure des ARN Pol est excessivement variable d'une espèce à l'autre chez les eucaryotes. Elles sont formées de 10 à 15 sous unités réparties en 2 grosses sous unités et 13 petites sous unités.

Ces enzymes sont en elles-mêmes peu efficaces pour la transcription. Pour fonctionner, elles nécessitent l'aide et la coopération de facteurs protéiques supplémentaires : les facteurs de transcription. Ils permettent la reconnaissance du site d'initiation mais aussi jouent sur l'élongation.

**4) Initiation de la transcription** : les régions promoteur, notion de séquences consensus



**Promoteur** = séquence d'ADN reconnue par une ARN Pol et indiquant le début d'une unité de transcription.
Il existe une très grande variabilité dans la partie "amont" des gènes. Il a fallu attendre que plusieurs centaines de promoteurs soient séquencées pour qu'on comprenne comment ils sont organisés.

**Organisation générale d'un promoteur :**
La séquence d'un promoteur est constituée de régions très variables et de régions beaucoup plus conservées, qualifiées de séquences consensus

**a) Promoteurs procaryotes**

Les promoteurs bactériens sont caractérisés par la présence de deux séquences consensus : la région -35 et la boite de Pribnow.



Ce sont des séquences reconstituées statistiquement quand on compare plusieurs promoteurs :



**Séquence consensus :**

Elles présentent des sites de reconnaissance à l'ARN Pol. C'est le facteur σ qui est responsable de la reconnaissance spécifique des séquences consensus : le core enzyme (sans s) a une certaine affinité pour l'ADN, donc il est continuellement fixé sur l'ADN où il se déplace continuellement. En présence du facteur σ, l'holoenzyme va reconnaitre spécifiquement les régions promoteurs et va se lier à ces régions avec une forte affinité.
Cette reconnaissance spécifique va être suivie d'une dénaturation locale (ouverture de la double hélice en position -10) et s'accompagne du positionnement du premier nucléotide au niveau du site d'initiation (position +1).
Le facteur σ va être relargué après positionnement d'une dizaine de nucléotides : la polymérisation s'effectue uniquement par le core enzyme. Le facteur σ et les séquences consensus qu'il reconnait permet un contrôle fin et précis de la transcription des gènes. Chez les bactéries (E.Coli) il existe plusieurs dizaines de facteurs s différents, reconnaissant des promoteurs légèrement différents (par leurs séquences consensus).
En conditions normales, chez E.Coli, on trouve un facteur σ majoritaire appelé le facteur σ 70 : c'est celui qui reconnait avec la plus forte affinité les séquences -10 et -35 qui se rapprochent le plus de ces séquences consensus (TTGACA et TATAAT) donc plus efficace.
Donc, dans les conditions normales de développement, les gènes possédant des séquences -10 et -35 se rapprochant le plus des séquences TTGACA et TATAAT seront transcrit préférentiellement.
Inversement, les gènes possédant des promoteurs dont les séquences -10 et -35 divergent par le facteur σ et donc ne seront pas transcrites.
Lorsque les conditions du milieu changent, d'autres types de facteurs s vont être synthétisés et vont permettre la reconnaissance d'autres promoteurs. Cette nouvelle reconnaissance va permettre d'activer des gènes qui étaient normalement éteints, permettant à la bactérie de réagir rapidement (soit en fabriquant des protéines de réponse au stress, soit en activant la sporulation).

**b) Promoteurs eucaryotes**

la structure des promoteurs eucaryotes est beaucoup plus complexe que chez les procaryotes, et bien moins connue. Les 3 ARN Pol vont reconnaître 3 grandes classes de promoteurs réparties en plusieurs milliers de types différents. Dans la plupart des cas, notamment les gènes de classe 2 (codant pour les ARNm) le promoteur est situé en amont de l'unité de transcription (comme chez lez procaryotes). Dans d'autres cas, comme dans les gènes de classe 3 (ARNt...) le promoteur peut se trouver dans la séquence codante (unité de transcription) donc difficile de dresser un schéma général de la structure des promoteurs eucaryotes.
D'une façon générale : 3 séquences consensus différents



**" TATA box" : séquence consensus = TATAAAA
" GC box" : séquence consensus = GGGCGG
" CAT box" : séquence consensus = GCCAAT**

Ces promoteurs sont mal reconnus par l'ARN Pol. Leur reconnaissance spécifique nécessite très souvent des séquences supplémentaires permettant d'activer ces promoteurs : on parle de séquence enhancer, ou les inhiber : les séquences silencer (elles font intervenir des facteurs de transcription).

**5) La phase d'élongation**

La transcription se fait dans le sens 5'3' ce qui signifie qu'elle s'agrandit en 3'.
Une boucle de transcription comporte environ 17 bases ouvertes. La vitesse de polymérisation est d'environ 30 à 80 nucléotides par seconde.

**6) La terminaison de la transcription**
La transcription d'un gène s'achève quand l'ARN Pol arrive à la fin de l'unité de transcription. Le complexe de transcription se désassemble, l'ARN polymérase se décroche de la matrice et la boucle de transcription se referme. Cette terminaison de transcription se produit en des régions appelées "sites terminateurs".
- d'une façon générale, les terminateurs eucaryotes sont caractérisés par la présence d'une suite de "T" sur le brin sens, on les appelle "T - rum"
- chez les procaryotes, la structure des terminateurs est bien connue, les mécanismes qui assurent la terminaison sont bien compris. Chez E.Coli : 2 types de terminateurs (ρ dépendant et ρ indépendant)
ρ dépendants nécessitent pour être reconnus en tant que terminateurs la présence d'une protéine particulière : la protéine ρ
ρ indépendants ne nécessitent pas de facteurs supplémentaires pour être reconnus

Dans les deux cas, la terminaison de la transcription résulte d'une fragilisation, déstabilisation et dissociation du complexe formé entre l'ADN matrice et l'ARN en cours de transcription.

organisation d'un terminateur sigma indépendant chez escherichia coli

**7) Modifications post transcriptionnelles des ARN**

Pour aboutir à la formation d'un ARN fonctionnel. On appelle aussi ces réactions de maturation de l'ARN. Ces phénomènes concernent les 3 classes d'ARN. On les observe à la fois chez le procaryotes et les eucaryotes.

**a) Maturation des ARNr**

Le principe général de maturation des ARNr est comparable chez les eucaryotes et procaryotes : on trouve dans le génome un gène qui va coder un précurseur préribosomique qui va être clivé au cours de ce phénomène de maturation pour libérer 2 ou 3 ARNr.

**b) Maturation des ARNt**

* clivage de quelques bases aux extrémités 5' et 3'
* modifications de certaines bases par méthylation, désamination, réduction...
* parfois, excision d'introns.

**c) Maturation des ARNm**

Chez les procaryotes, du fait de l'absence de compartimentation interne, la traduction d'un ARNm en protéine débute avant même la fin de la transcription. Il n'y a donc pas le temps de modifier les ARN avant que la transcription commence, donc peu de maturation des ARNm.
Chez les eucaryotes, les ARNm vont subir de profonds remaniements qui vont conduire à la formation d'ARNm matures. 3 types de grandes réactions :
- Excision des introns et réunion des exons (= EPISSAGE)
- CAPPING de l'extrémité 5'
- POLYADENYLATION de l'extrémité 3'

L'ensemble de ces modifications dans le noyau (nucléole plus exactement) dure environ une heure.

**1. l'épissage.**

On distingue différents types d'introns en fonction des mécanismes qui assurent leur excision. Dans la plupart des cas, l'épissage nécessite l'intervention d'un complexe catalytique particulier faisant intervenir de petits ARN nucléaires (snRNA) qui s'associent à des protéines. Ces complexes snRNA-protéines forment des snRNP (smallnuclearribonucleoparticle) ou "snurps". Ils s'associent à l'ARN à modifier pour former un complexe de maturation qu'on appelle "spliceosome".

On connait un autre groupe d'introns "autocatalytiques" pour lesquels l'excision ne nécessite pas la présence de facteurs protéiques mais peut être réalisée (in vitro) en la seule présence d'un nucléotide guanylique (GTP, GMP ou GDP) donc certains ARN ont une activité enzymatique, on les appelle "ribozymes".

**Mécanisme d’excision -épissage**

C’est un mécanisme de précision, une erreur d’un seul nucléotide au moment d’excision change le cadre de lecture et aboutira à la formation d’une fausse protéine. Deux sites sont importants : la jonction exon-intron et le site de branchement.

**a-La jonction exon-intron :** tous les introns d’un transcrit primaire commencent par GU situé à l’extrémité 5’ de l’intron et appelé site donneur d’épissage, et ils se terminent par AG (site 3’) appelé site accepteur d’épissage. Le site de branchement (A) n’est pas loin de l’extrémité 3’ d’intron environ -30 nucléotides se trouve une adénine appelée adénine de branchement. Des ARN nucléaires spéciaux, les snRNA, reconnaissent les extrémités 5’ et 3’ des introns et participent activement à une réaction de transestérification qui échange les liaisons phosphate de façon à laisser l’intron sous forme de lasso alors que les exons sont reliés entre eux.

**b-Schéma du mécanisme excision-épissage** : On aura deux clivages :

1. Clivage à l’extrémité 5’ d’intron entre la fin de l’exon situé en amont et début de l’intron, l’attaque de cette liaison ester est due à l’OH en 2’ de l’adénine de branchement. L’extrémité 5’ de l’intron GMP vient se soudée par une liaison covalente avec l’A de branchement formé ainsi un lasso.
2. Formation de deuxième clivage en 3’ d’intron entre la fin d’intron et début d’exon situé en aval, il se produit simultanément une soudure des deux exons (entre 3’OH en de l’exon amont et 5’p de l’exon aval et libération de lasso qui sera dégradé par des nucléases cellulaires.

**Rôle des snRNA :** chez les eucaryotes il existe des snRNA qui jouent un rôle important dans l’excision-épissage, ce sont des petites molécules d’environ 100 à 300 nucléotides situées dans le noyau, riche en uracile et agissent sous forme de complexe avec des protéines d’où le nom de snRNP, elles contiennent un snRNA et plusieurs protéines. Il existe plusieurs types différents entre eux par leurs structure on distingue : U1, U2, U4, U5, U6.

**Le spliceosome :** c’est un assemblage moléculaire complexe qui se place au niveau de l’intron à éliminer, il contient plusieurs types de snRNA. Certain snRNA sont impliqués dans la reconnaissance des 3 principaux sites (par complémentarité) sur un petit nombre de nucléotides, ce sont les U1 qui reconnait la jonction 5’ à cliver (GU), U2 reconnait le site de branchement (A), U5 reconnait la jonction 3’ (AG). D’autre snRNA sont impliqués dans les réactions enzymatiques d’excision-épissage comme le U4 et le U6.

2**. le capping.**

C'est la première modification qui affecte les ARNm juste à la fin de la transcription. Cette réaction consiste en la perte de 2 groupements phosphates à l'extrémité 5' triphosphate de la

molécule, suivie de la fixation d'un GDP par une liaison 5'-5'.



**3. la polyadénylation**

Elle consiste à rajouter une suite d'adénine au niveau de l'extrémité 3' des ARNm. Cette suite constituera la "queue poly A" caractéristique de tous les ARNm eucaryotes..

Cette queue est reconnue par des transporteurs. Un des points de contrôle de durée des ARNm réside dans la longueur de cette queue