

Université de M'sila
Faculté des Sciences
Département de Biochimie et Microbiologie
3ieme année BIOCHIMIE
module de Biologie moléculaire
Responsable du module
Boubekeur. H

L'expression de l'information génétique

La transcription et la maturation de l'ARN.

Introduction

La transcription est le processus de copie du matériel génétique (ADN ou ARN) en ARN. Chez les procaryotes une seule ARN polymérase-ADN dépendante effectue la transcription pour tous les types d'ARN, tandis que chez les eucaryotes, trois ARN polymérases (ARNpol) interviennent : l'ARNpol I ou A pour les ARN ribosomiques transcrits dans le nucléole (28S, 18S et 5,8S), l'ARNpol II ou B pour les ARNm, et l'ARNpol III ou C pour les petits ARN (ARNt, ARNr 5S, ARNsn). Pour certains virus à ARN, l'ARN est transcrit par une ARNpol-ARN dépendante appelée aussi réplicase.

Expression d'un gène = processus entier qui décode l'information portée par un gène donné et la traduit en protéine.

I - Les ARN

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

Les ARN majoritaires de la cellule sont :

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

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

I-1 - Les ARNt

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

Ils sont composés d'une seule chaîne polynucléotidique. On peut retrouver des bases inhabituelles dites « mineurs ». A son extrémité OH se trouve toujours le triplet CCA. Il se crée une liaison ester entre le COOH de l'AA et l'OH en 3' de l'ARNt. Il se forme un **amino-acyl-ARNt** se fixant au ribosome. La structure secondaire des ARNt montre une tige et trois boucles avec des bras complémentaires. Les régions fonctionnellement importantes sont les régions fixant l'AA et l'anticodon de la boucle 2 qui est caractéristique d'un ARNt donné. Il est reconnu par appariement un triplet complémentaire de l'ARNm. A un AA donné correspond un anticodon.

I-2 - Les ARNr

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

	PROCARYOTES	EUCARYOTES
Constante de sédimentation	70s	80s
Masse particulaire	$2,8 \cdot 10^6$	$4,5 \cdot 10^6$
Dissociation (quand Mg^{2+} en dessous de 1mMol)	50 + 30	60 (ARN 28s, 5s et 5'8s + 45 prot) + 40 (ARN 18s + 33 prot)
%ARN	63	50
%Protéines	37	50

I-3 - Les ARNm

Ils sont minoritaires dans la cellule. Il s'agit de la « copie » de l'information génétique contenu dans l'ADN. Ils sont synthétisés rapidement et leur demi-vie est courte (quelques minutes chez les procaryotes et quelques heures chez les eucaryotes) ; et leurs tailles sont très hétérogènes. Les ARNm eucaryotes sont synthétisés sous forme d'un précurseur, ils subissent une maturation pour donner un ARNm actif. Cette maturation se fait en partie au niveau du noyau.

II - 1- Transcription chez les procaryotes.

Les promoteurs font environ 40pb (région couverte par l'enzyme).

2 séquences conservées

- Une séquence consensus d'une dizaine de nucléotides, placée en -35 du +1 de transcription = **boite GC**.

- Une séquence en -10 du +1 de transcription = **boite TATA (Pribnow)**.

Cette dernière facilite la dissociation des deux brins d'ADN, car riche en A et T

Ces deux séquences correspondent au Promoteur minimum

ARN polymérase ADN dépendante de *E.coli* comporte 4 sous-unités de 3 sous-types différents
 β , β' , α 2 = core enzyme.

Une 5^{ème} sous-unité ou facteur **σ** est nécessaire à l'initiation de la transcription, lorsqu'elle est associée avec le core enzyme on obtient l'**Holoenzyme**.

La transcription se déroule en 5 étapes :

1 – Interaction de l'holoenzyme avec l'ADN au niveau du promoteur. Formation d'une bulle de transcription et déroulement de l'ADN sur environ 17pb.

2 – Initiation de la polymérisation de la chaîne d'ARN.

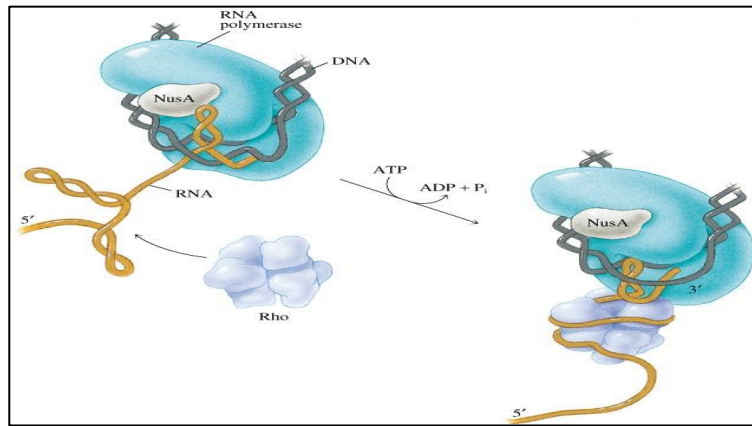
3 – Libération du facteur σ .

4 – Elongation de la chaîne ARN.

5 – Terminaison de la transcription.

5 -1- la terminaison Rho-dépendante

Une courte séquence d'ADN, riches en paires G-C (3 liaisons hydrogènes donc plus fortement liées) suivie de plusieurs U, est transcrite en "épingles à cheveux" ce qui stoppe la progression de l'ARN polymérase. Une protéine de terminaison appelée Rho (une hélicase ARN-ADN / ATP-dépendante sous forme d'homohexamère se fixe sur l'ARN et le complexe d'élongation est dissocié. Le brin d'ARN néo-synthétisé est libéré du brin d'ADN matrice. Le site d'enroulement de l'ARN autour de Rho est une région d'environ 70 nucléotides



la terminaison Rho-dépendante

5 -2-La terminaison Rho-indépendante (ou terminaison intrinsèque)

Elle a lieu au niveau d'une séquence répétée inversée (également riche en paires G-C) suivie de 6 A, située après la séquence codante. La transcription de cette séquence inversée aboutit à la formation d'une structure en "épingle à cheveux" dans l'ARN en cours de biosynthèse qui bloque le complexe de transcription. Les liaisons H entre la séquence poly A et le brin complémentaire poly-U néo-synthétisé sont rompues et le brin d'ARN est dissocié du brin d'ADN matrice.

Chez les procaryotes les transcrits d'ARNm servent de matrice à la traduction bien avant que la transcription soit terminée, certainement car la transcription n'est pas compartimentée dans le noyau. Par conséquent les ARN procaryotes subissent peu de modifications post-transcriptionnelles.

Alors que les ARNr et ARNt primaires sont beaucoup plus long que la molécule finale. Ils nécessitent donc une transformation avant d'être fonctionnelle.

II – 2- Transcription chez les eucaryotes.

Intéressons-nous plus particulièrement à la transcription des gènes codant pour les chaînes polypeptidiques chez les eucaryotes. Leur transcription se déroule en deux étapes : formation d'un ARN prémessager puis maturation de cet ARNprémessager pour former un ou plusieurs ARN messagers.

2.1. L'initiation de la transcription

Tout l'ADN n'est pas transcrit, seules les régions correspondant à des gènes le sont. Et encore, cette expression peut être régulée selon le stade de développement, le type cellulaire,

l'environnement, etc... Dès lors, un acteur doit intervenir pour déterminer à quel endroit une région d'ADN doit commencer à être transcrite : c'est le rôle du promoteur.

2.1.1. Le promoteur

Le promoteur correspond à une région non transcrite de l'ADN, généralement juste en amont du début de la région transcrite, dont la séquence permet le recrutement de l'ARNpol II. Certaines séquences du promoteur (surnommées "boîtes") ont une importance particulière dans ce processus, essentiellement parce que ces séquences sont reconnues spécifiquement par différentes protéines appartenant au complexe d'initiation (voir paragraphe suivant) :

- la "boîte TATA" riche en thymine et adénine, la plus importante, est située vers -25 à -30 nt du site de démarrage de la transcription (noté +1);
- des éléments proximaux :
 - la "boîte CAAT" (facultative), contenant de la cytosine, est située vers -120 à -80 nt du site de démarrage de la transcription.
 - la "boîte GC" (facultative également), riche en guanine et cytosine, peut être présente entre la boîte CAAT et la boîte TATA..

Signalons que si ces séquences sont souvent bien conservées, une variabilité non négligeable peut également être observée. Ainsi, il existe des promoteurs sans "boîte TATA"

2.1.2. Le complexe d'initiation

Contrairement à ce qui se passe chez les procaryotes, l'ARNpol II des eucaryotes ne reconnaît pas seul le promoteur proximal. 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. Ces facteurs sont notés TFIIA, TFIIB, etc... pour Transcription Factor for RNA polymerase II. Ils correspondent aux facteurs généraux de la transcription car ils s'assemblent sur tous les promoteurs utilisés par l'ARNpol II. La séquence d'assemblage du complexe d'initiation est décrite sur la fig. 1

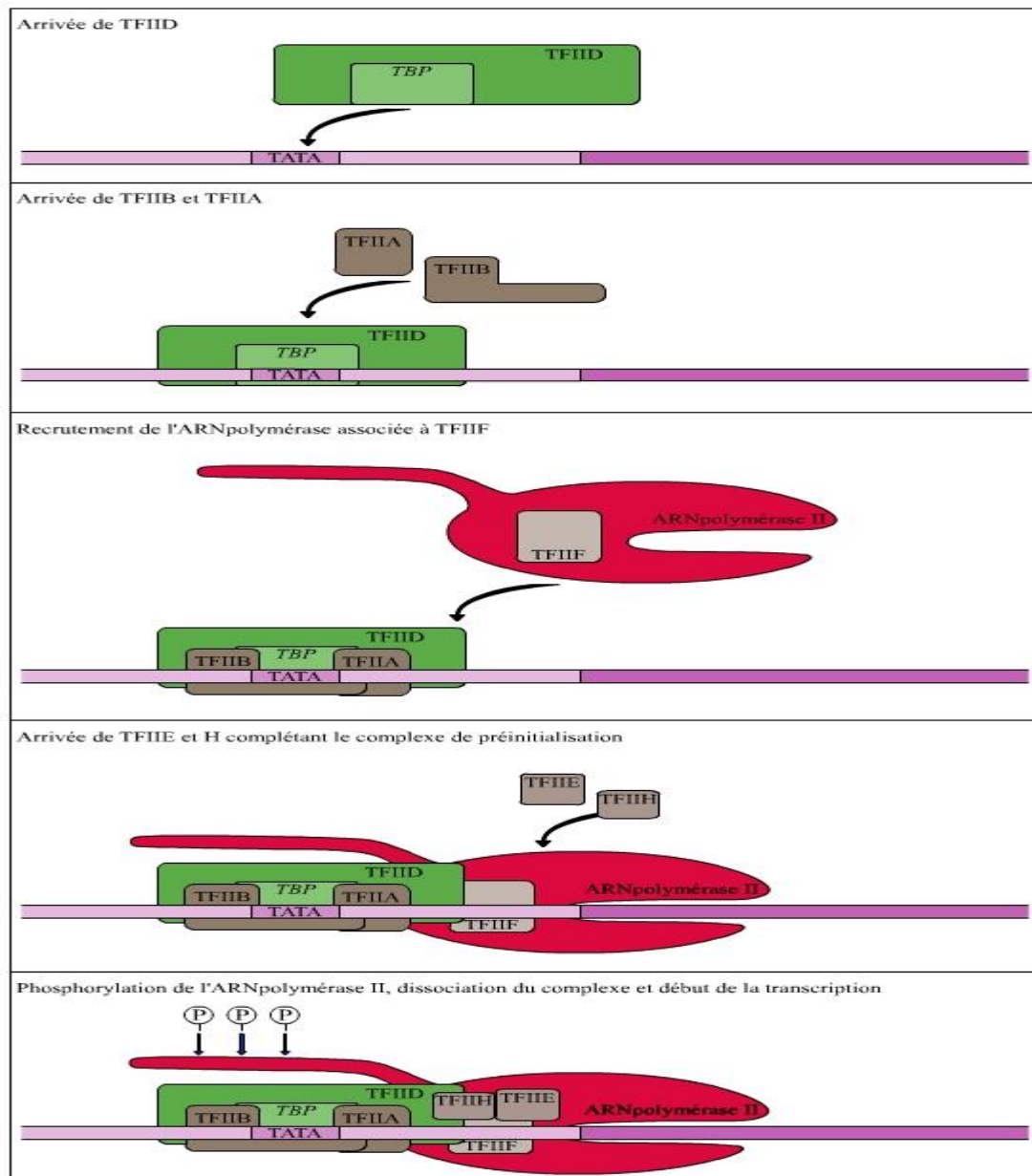
La TBP est la première protéine qui reconnaît une séquence spécifique de l'ADN initiateur de la transcription (la boîte TATA). Le facteur TFIIB semble impliqué dans la sélection précise du site d'initiation (nucléotide à partir duquel se déroule la transcription).

La TBP est la première protéine qui reconnaît une séquence spécifique de l'ADN initiateur de la transcription (la boîte TATA). Le facteur TFIIB semble impliqué dans la sélection précise du site d'initiation (nucléotide à partir duquel se déroule la transcription).

Le facteur TFIIF comporte plusieurs activités enzymatiques dont une activité hélicase permettant l'ouverture de la double hélice d'ADN au niveau du promoteur, et une activité kinase responsable de la phosphorylation de la queue C-terminale de l'ARN polymérase II. Cette phosphorylation entraîne une modification de la structure tridimensionnelle de l'ARN polymérase entraînant la dissociation du complexe d'initiation et le début de la transcription.

TFII = Transcription Factor for RNA polymérase II (Facteurs généraux de la transcription pour l'ARN polyméraseII).

TBP = TATA box-Binding Protein (Protéine de liaison à la boîte TATA). La liaison du complexe de transcription au promoteur proximal provoque l'ouverture et le déroulement des deux brins de son ADN, tout en indiquant le brin qui va être transcrit



. **Figure 1 : Mise en place du complexe d'initiation**

2.1.3. Intervention de facteurs spécifiques de la transcription

Le complexe d'initiation composé de l'ARNpol II et des différents TFII est suffisant pour obtenir une activité transcriptionnelle *in vitro* mais à très faible taux. L'augmentation de cette activité basale (ou sa répression) est sous la dépendance de facteurs spécifiques qui vont interagir avec le complexe d'initiation. Ces protéines activatrices ou inhibitrices se lient à des promoteurs distaux, séquences spécifiques de l'ADN, appelées enhanceurs lorsqu'il recrute des cofacteurs activateurs, ou silencer lorsqu'ils recrutent des cofacteurs inhibiteurs. Ces

promoteurs distaux peuvent être situés à des milliers de nucléotides du promoteur proximal et agissent sur le promoteur proximal par le jeu de courbures de l'ADN (voir fig. 2).

Mais d'autres participants existent encore. On trouve ainsi des promoteurs proximaux alternatifs et des promoteurs distaux, situés à des milliers de nucléotides du promoteur proximal. Malgré la distance qui sépare les promoteurs proximaux des promoteurs distaux, ces derniers agissent sur le promoteur proximal comme activateurs (enhancers) ou répresseurs (silencers) par le jeu de courbures de l'ADN, des facteurs de transcription et du médiateur qui maintient liés tous ces acteurs.

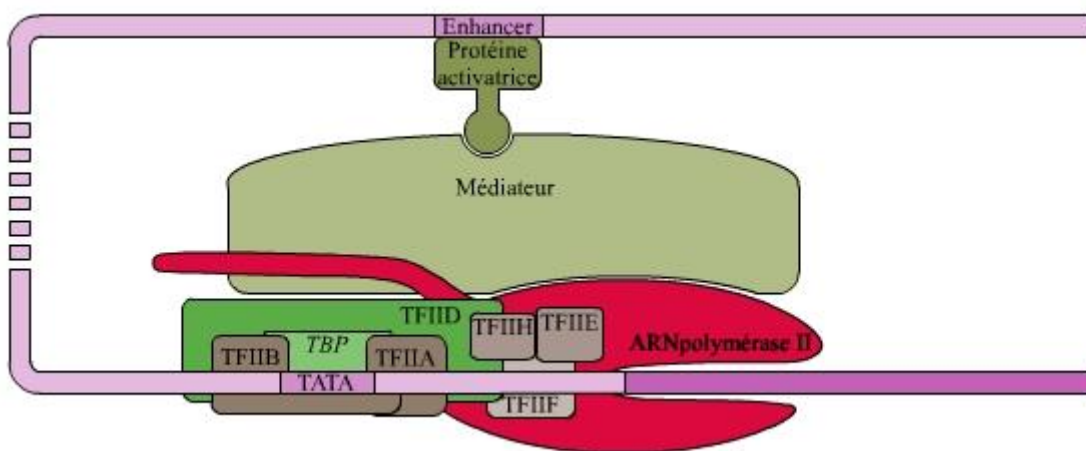


Figure 2 : La régulation spécifique de la réplication

Les enhancers (activateurs) sont des séquences situées parfois à plusieurs milliers de nucléotides d'un promoteur qui activent ce dernier par un jeu d'interactions protéiques qui stabilisent le complexe d'initiation. Ces interactions sont rendues possibles par la courbure de l'ADN qui rapproche des éléments situés à de grandes distances les uns de autres. Cette activation est spécifique du gène et utilise une multitude de facteurs de transcriptions spécifiques (les protéines activatrices) qui agissent généralement sous forme dimérique). Une répression spécifique peut agir selon le même type de modalité.

2.2. L'élongation

L'ARN pol II est équipé de facteurs protéiques d'élongation qui facilitent sa progression au travers d'une chromatine dont ils relâchent la structure. Un ARN pré-messager complémentaire du brin matrice de l'ADN (brin antisens), donc identique au brin codant de l'ADN (brin sens) aux riboses et uraciles près, commence à être synthétisé selon la direction 5'-3' (voir fig. 3).

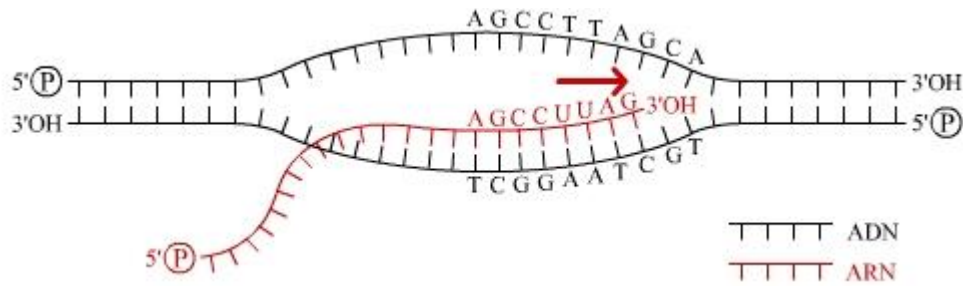


Figure 3 : La phase d'élongation de la transcription

L'ARN polymérase lit le brin patron (ou brin antisens) qui est complémentaire du brin codant (ou brin sens) depuis son extrémité 3' vers son extrémité 5'. Le brin d'ARN néosynthétisé est donc identique au brin codant d'ADN, aux uraciles et riboses près.

2.3. La terminaison

L'ARN pol II est également équipée de facteurs protéiques de terminaison, et reconnaît ainsi un ou plusieurs signaux de terminaison portés par le brin progressivement parcouru et qui annoncent la fin de la transcription sur le brin d'ADN matrice (TTATTT par exemple, parfois aussi plus en aval ATACAAC...). Elle arrête bientôt son travail de transcription et libère l'ARNpm qu'elle vient d'assembler.

3. La formation d'un ou plusieurs ARNmessenger(s)

Le transcrit primaire n'est pas utilisé tel quel pour la synthèse protéique (la traduction). Il doit subir des modifications qui répondent à plusieurs impératifs (augmentation de la demi-vie, modification de la séquence). Toutes ces modifications sont réalisées au fur et à mesure de la progression de la synthèse du préARNm dans le nucléoplasme. Il existe trois grands types de modifications, catalysées chacune par des enzymes de nature protéique ou ribonucléique.

3.1. L'addition d'une coiffe en 5'

Elle a lieu dès le début de la transcription avant que la chaîne ne compte plus de 30 nucléotides. Elle consiste en l'ajout d'un nucléotide à guanine sur l'extrémité 5' de l'ARN suivi de sa méthylation sur l'azote 7 de la base, ainsi que de la méthylation en 2' du ribose des un ou deux premiers nucléotides du transcrit primaire. La particularité de cet ajout consiste dans le type de liaison mis en jeu (une liaison anhydride d'acide entre les acides phosphoriques des deux nucléotides). Il en résulte que l'extrémité 5' de l'ARNm n'est pas

porteuse des trois acides phosphoriques libres habituels, mais d'un GMP, ce qui limite la réactivité de cette extrémité et sa reconnaissance par les exonucléase (protection contre la dégradation). Cet ensemble sert de coiffe protectrice à l'extrémité 5' de l'ARNm. Elle est également nécessaire à l'exportation de l'ARNm vers le cytoplasme et à la liaison de ce dernier avec la petite sous-unité du ribosome lors de l'étape d'initiation de la traduction.

3.2. L'excision-épissage

Chez les eucaryotes, les archéobactéries et les cyanobactéries, les gènes sont morcelés : constitués d'une alternance d'exons (parties codantes du gène) et d'introns (parties non codantes, bornées par des séquences de bases spécifiques : 5'GU et 3'AG), ils sont d'abord intégralement recopiés dans l'ARNpm, puis subissent une opération d'excision des introns (ainsi que celle parfois de petits morceaux d'exons) suivi d'un épissage (splicing), c'est à dire la réunion bout à bout des exons restants qui constituent l'ARNm. Ce remaniement se déroule au fur et à mesure de la progression de la transcription.

L'excision des introns s'opère par l'entremise d'une formation dite "en lasso" (voir fig. 5 et 6).

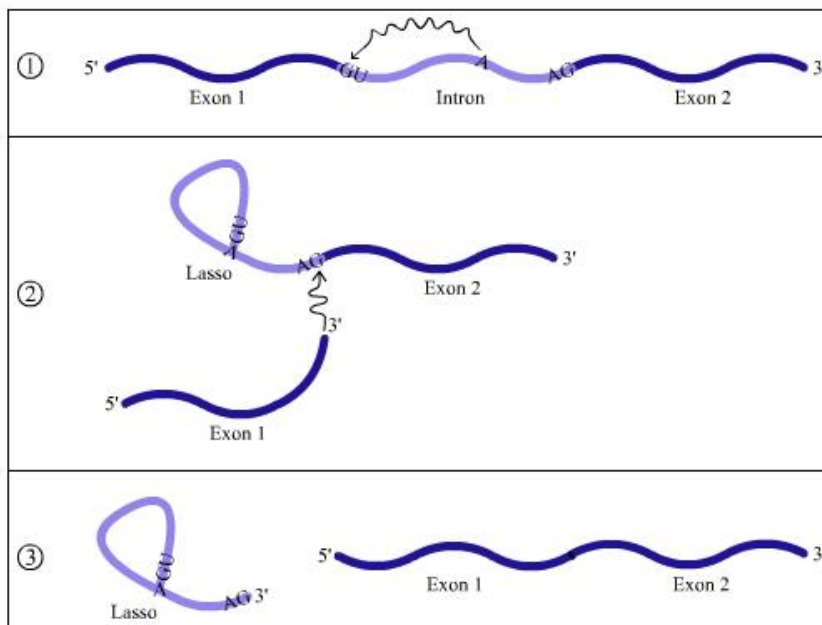


Figure 5 : Excision de l'intron par formation d'un lasso : vue d'ensemble du mécanisme

L'excision-épissage est réalisée par réaction d'un nucléotide à adénine (A) situé dans l'intron avec un nucléotide à guanine situé en 5' de l'intron. Cela entraîne la séparation de l'intron d'avec l'exon 1 (situé en amont) et la formation d'une structure en lasso interne à l'intron. Ensuite, l'extrémité 3' de l'exon 1 réagit avec l'extrémité 5' de l'exon 2 permettant l'épissage des deux exons et la libération du lasso qui sera dégradé par des ribonucléases

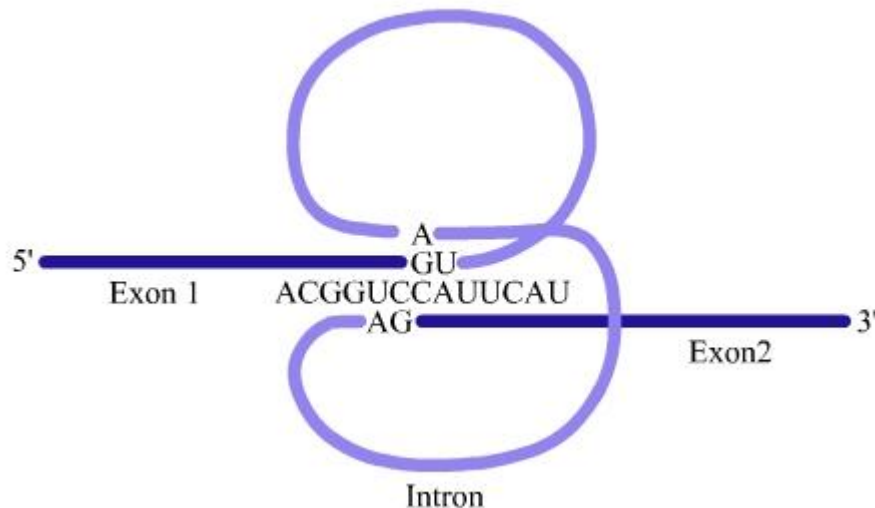


Figure 6 : Excision de l'intron par formation d'un lasso : le splicéosome

Le splicéosome fait appel à des facteurs spécifiques : ribonucléoprotéines contenant des petits ARN (snRNP), ribozymes (ARN ayant des propriétés catalytiques), des enzymes spécifiques (maturases)... La structure secondaire du transcrit I met en contact trois séquences de l'intron : la séquence du début de l'intron (extrémité 5'- GU ou site donneur), la séquence de la fin de l'intron (extrémité 3'-AG ou site accepteur) et un nucléotide à adénine (A du branchement) 40 nucléotides avant le site receveur. Mais l'épissage n'est pas seulement constitutif: il existe également des épissages alternatifs, dans lesquels l'élimination des introns (ou des portions d'exons) peut faire se lier entre eux des exons différents. C'est ce mécanisme qui démultiplie les capacités codantes d'un gène (voir un exemple d'épissage alternatif dans la fig. 7)

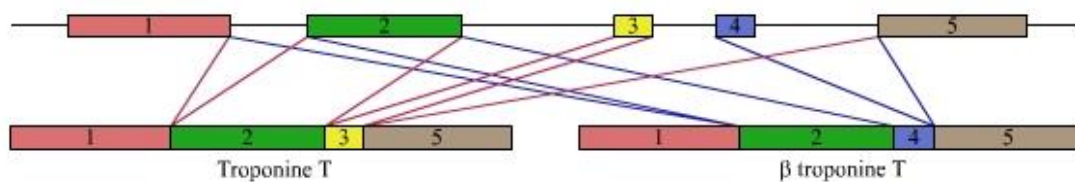


Figure 7 : Exemple d'épissage alternatif

A partir du même préARNm peuvent être obtenus deux ARNm matures différents codant pour deux isoformes de la troponine, une molécule impliquée dans la structure des myofibrilles.

3.3. L'addition d'une queue polyA en 3'

Le site de polyadénylation est codé au niveau du gène. Le site de clivage déterminé par le dinucléotide CA est entouré par une séquence AAUAAA très conservée située 10 à 30 nucléotides en amont du site de clivage, et par une séquence DSE (Down Stream Element) riche en U ou en GU situé une trentaine de nucléotides en aval du site de clivage. La séquence AAUAAA est reconnue spécifiquement par un complexe protéique appelé CPSF (Cleavage and Polyadenylation Specific Factor), et la séquence DSE par un complexe CstF (Cleavage Stimulation Factor). Ces deux complexes et d'autres composants, comme l'ARN polymérase II et la poly(A) polymérase (PAP), vont interagir en formant le complexe de clivage qui va cliver la molécule de préARN au niveau du site de clivage (voir fig. 8).

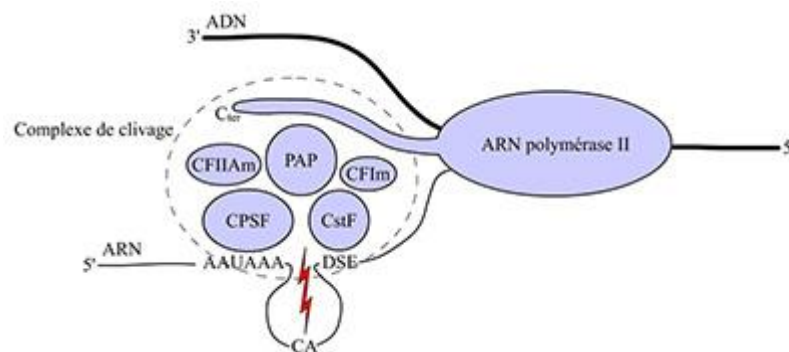


Figure 8 : Le complexe de clivage impliqué dans la maturation en 3' du préARNm

CPSF = Cleavage and Polyadenylation Specific Factor.

CstF = Cleavage Stimulation Factor.

PAP = Poly(A) Polymerase.

CF = Cleavage Factor

DSE = Down stream Element.

La PAP (Poly(A) Polymerase) va alors ajouter environ 200 nucléotides, une poly(A) binding protein II (PABP2) qui interagit avec la PAP se chargeant d'activer l'enzyme et de contrôler la longueur de cette queue poly(A). Il est intéressant de noter que cette polymérase n'utilise pas

de matrice ADN pour créer cette séquence poly(A). La queue Poly(A) n'est donc pas codée par le génome. Cette queue poly(A) confère de la stabilité au futur ARNm et se perd au fur et à mesure qu'il est traduit.