

# régulation de l'expression génétique

## I. Régulation transcriptionnelle

### Objectifs :

- Comprendre que la transcription peut être régulée à des niveaux différents
- Connaitre ces niveaux de régulation
- Connaitre les mécanismes précis de ces régulations

### Introduction

Les cellules d'un même organisme multicellulaire sont dotées d'un patrimoine génétique identique, cependant elles diffèrent les unes des autres par leur transcriptome, leur protéome et leur métabolome qui dépendent de l'expression génique. Ainsi, le niveau d'information diffère en fonction du type cellulaire, du niveau de différenciation, du stade de développement et des différents stimuli perçus par la cellule. Ces variations de l'expression génique permettent à la cellule d'assurer des fonctions très diverses, notamment en ce qui concerne les métabolismes secondaires. Cette expression est soumise à une régulation temporelle et spatiale fine et complexe. L'acide désoxyribonucléique (ADN) est le support de l'information génétique, il contient des séquences codantes qui sont transcrites en acides ribonucléiques messagers (ARNm) par des ARN polymérases ADN dépendantes, avant d'être traduits en protéines. Ces étapes intermédiaires entre le gène et la protéine constituent une opportunité de choix entre les différents gènes à transcrire et constituent donc des étapes clés dans la régulation génique. La modification transcriptionnelle des ARN est une étape importante du contrôle de l'expression des gènes, on considère généralement que les différentes modifications sont les éléments de régulation transcriptionnelle de l'expression des gènes. Ces différentes modifications peuvent influencer sur différentes caractéristiques de l'ARN, telle que sa stabilité, sa capacité à être traduit ou bien même modifier la séquence à traduire.

### 1. Définition des modifications transcriptionnelles :

Représentent l'ensemble des modifications qu'un ARN subit après avoir été transcrit. On parle également de **maturation de l'ARN**. Les modifications les plus connues sont l'ajout d'une coiffe, la polyadénylation ou encore l'épissage, mais il en existe bien d'autres. La maturation des transcrits primaires a lieu dans le noyau de la cellule.

## 2. Modifications de l'ARN après transcription chez les procaryotes :

Elles diffèrent selon la nature de l'ARN :

- L'ARNm ne subit pas de modification
  - Les ARN<sub>t</sub> et ARN<sub>r</sub> subissent une maturation qui correspond à une coupure à partir de précurseurs :
1. Pour les ARN<sub>t</sub>, le précurseur contient plusieurs séquences d'ARN<sub>t</sub> et, dans certains cas, des séquences d'ARN<sub>r</sub>.

2. Pour les ARN<sub>r</sub>, la maturation implique aussi le clivage et l'élagage d'un précurseur contenant tous les ARN<sub>r</sub> et des ARN<sub>t</sub>.

Enfin, les bases de l'ARN peuvent subir de nombreuses modifications :

- méthylation de l'uracile la transformant en thymine, formation de 5-méthylcytosine, de méthyladénine, diméthyladénine.
- Réduction de l'uracile en dihydro-uracile
- transformation de l'uridine en pseudo-uridine.

Ces modifications touchent essentiellement les ARN<sub>t</sub> et ARN<sub>r</sub> et jouent un rôle très important pour la fonction de ces molécules.

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 transcriptionnelles, catalysées chacune par des enzymes de nature protéique ou ribonucléique.

## 3. régulation transcriptionnelle chez les eucaryotes :

### 3.1. L'addition d'une coiffe en 5' (ou capping) :

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** (provenant d'un GTP hydrolysé en GMP), 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.

Cette coiffe a plusieurs rôles :

- Cet ensemble sert de coiffe protectrice à l'extrémité 5' de l'ARNm. 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.

- Elle est également nécessaire à l'exportation de l'ARNm vers le cytoplasme.

-Rendre l'ARN accessible au processus d'excision – épissage.

La coiffe est ajoutée grâce à un complexe protéique appelé « **Cap-Binding-Complex** » qui possédant une activité triphosphatase une activité guanylyl-transférase et une activité méthyl-transférase.

La coiffe se caractérise par l'ajout d'un nucléotide à guanine à l'extrémité 5' du brin d'ARN. Ce nucléotide à guanine est relié à la chaîne nucléotidique par une liaison inhabituelle : au lieu d'être reliés par une liaison ester-phosphorique entre le groupement OH porté par le carbone 3' du ribose du nucléotide à guanine et l'acide phosphorique alpha du premier nucléotide de l'ARN natif, les deux nucléotides sont reliés par une liaison anhydride d'acide entre les acides phosphoriques des deux nucléotides.

Par la suite, le cycle imidazole de la guanine terminale est méthylé sur son azote 7.

Par ailleurs, le ribose du premier nucléotide de l'ARN natif est méthylé sur l'oxygène porté par le carbone 2'. Ce peut également être le cas du nucléotide suivant.

### 3.2. L'addition d'une queue poly A en 3'

La poly-adénylation correspond à l'ajout de jusqu'à 200 adénines à l'extrémité 3' du transcrit primaire et ceci sans matrice par la **poly-A-polymérase (PAP)** qui utilise des ATP comme donneurs d'adénine.

Le mécanisme implique la prise en charge de deux séquences de l'ARN dont l'une, **AAUAAA** (séquence consensus retrouvée à la fin de tous les gènes), appelée signal de polyadénylation, est située approximativement à **15 nucléotides** en amont d'un site de clivage endonucléasique, qui servira ensuite de site d'addition de poly-A alors que l'autre, une séquence riche en **U** ou en **G/U** est située en aval du site de coupure. Ces deux séquences sont reconnues chez les mammifères par **deux complexes protéiques** interagissant l'un avec l'autre : **CPSF** (Cleavage and Poly-adenylation Specify Factor), reconnaissant AAUAAA et **CstF** (Clivage Stimulation Factor) reconnaissant la séquence riche en G/U. À cet ensemble vient s'ajouter la poly Apolymérase et deux autres protéines probablement responsables du clivage endonucléasique CFI et CFII.

**Remarque :** Certains ARNm, en particulier ceux codant pour les séquences des histones, ne possèdent pas de queue poly A.

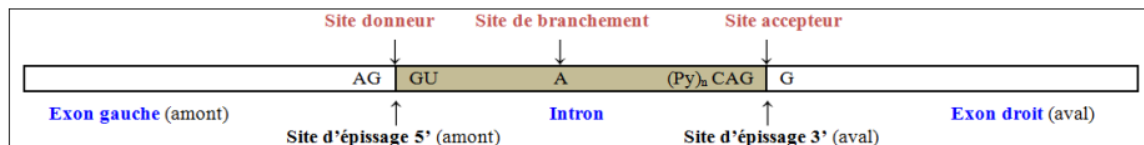
La queue poly A aurait plusieurs rôles :

- Protection de l'ARNm lors de la traduction
- Aide au passage de l'ARNm du noyau vers le cytoplasme

### 3.3. L'excision-épissage

Chez les eucaryotes, 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'ARNprém, 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.

Il faut en effet que cet épissage soit parfait, car une erreur d'un seul nucléotide aboutirait à un changement du cadre de lecture, et donc à une protéine totalement différente.



Les jonctions d'épissage sont reconnues par les **snRNP** (ou **snurps** pour *Small-NuclearRibonucleo-protein-Particules*). Les snRNP correspondent à l'association de snRNA (snRNA

**U1, U2, U4, U5, U6**) et de protéines et l'ensemble des snRNPs s'appelle le **spliceosome ou complexe d'épissage**. La taille de ces snRNA varie de 100 à 300 nucléotides chez la plupart des eucaryotes.

Les snRNP jouent trois rôles dans l'épissage :

- Reconnaittent le site d'épissage 5', le site de branchement et le site d'épissage en 3'.
- Rapprochent ces sites de manière adéquate.
- Catalysent des réactions de clivage et de ligation (ou participent à la catalyse) par des interactions ARN-ARN, ARN-protéine et protéine-protéine

Les séquences de début de l'intron (**GTAAGT** dans l'ADN, **GUAAGU** dans l'ARN) et de fin de l'intron (**AG**) sont invariantes. De plus, il existe toujours une **adénine** située près de l'extrémité 3' de l'intron située à environ -20 à -30 nucléotides. Ces séquences sont reconnues

dans l'ARN prémessager par différentes RNPs au sein du complexe d'épissage. En fonction de leur rôle dans le processus d'épissage, les séquences de début d'intron, de fin d'intron et celles où se trouve l'adénine sont respectivement appelées : site donneur, site accepteur et point de branchement. La fixation des RNPs s'effectue grâce à des appariements entre bases de l'ARN messager et de l'ARN de la ribonucléoprotéine.

Le snRNP U1 reconnaît le site donneur et le snRNP U2 reconnaît le site de branchement et le site accepteur.

L'excision des introns et l'épissage des exons se fait en plusieurs étapes :

Le **snRNP U1** permet la reconnaissance du site donneur d'épissage et entraîne la rupture de la liaison phosphodiester entre le premier exon et l'intron.

□ Cette rupture de la liaison **phosphodiester** entraîne la formation d'un **lasso**, qui n'est autre que l'extrémité 5' de l'intron. Ce lasso forme une liaison avec le site de branchement, lui-même situé sur le même intron qui se replie ainsi sur lui-même.

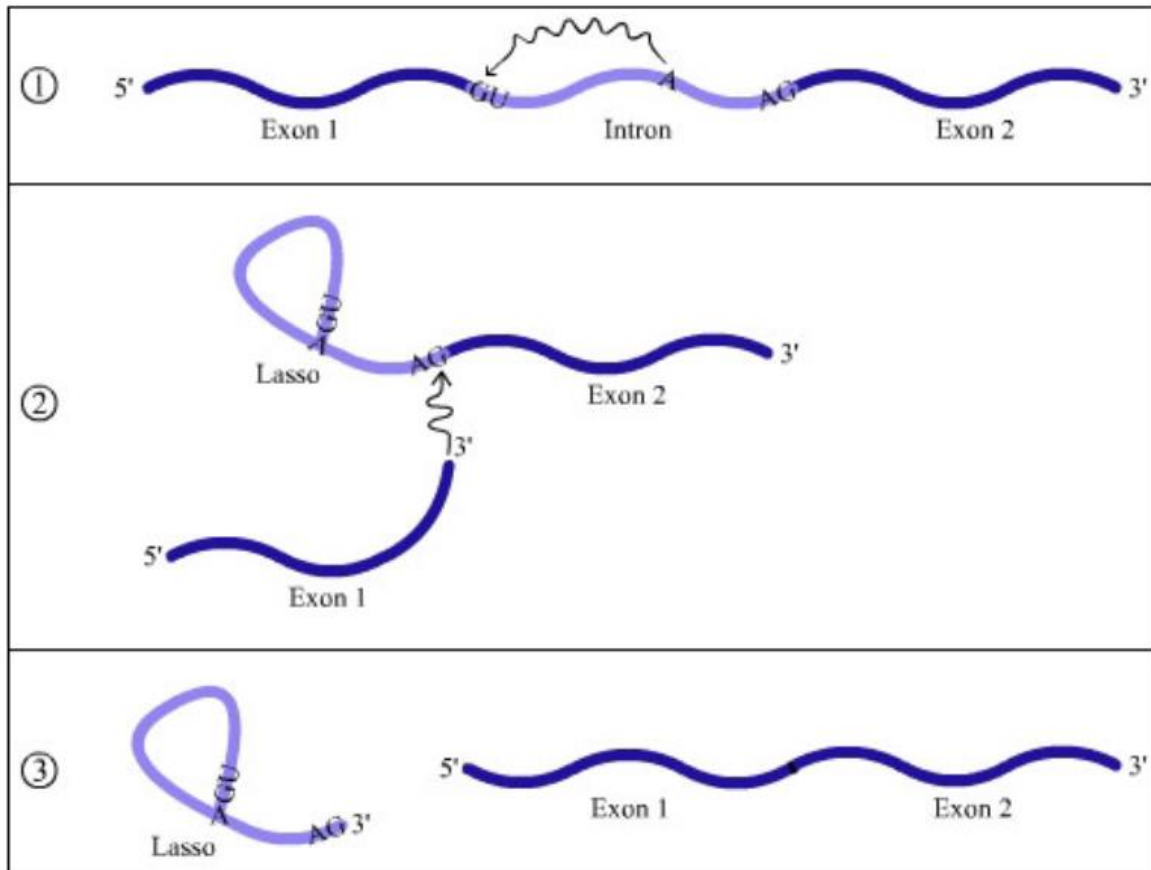
□ U2 se lie au point de branchement, puis les protéines U4 et U6 s'associent à U2 ; le **site de branchement** est reconnu par le **snRNP U2** et permet la liaison par l'intermédiaire d'une adénosine.

□ U5 se fixe près du site accepteur.

□ Le **snRNP U2** permet également la reconnaissance du **site accepteur** d'épissage.

Suite à cette reconnaissance il y a rupture de la liaison phosphodiester au niveau de l'extrémité 3' de l'intron.

□ Le groupement 3'OH du premier exon peut ainsi réagir avec l'extrémité 5'phosphate du deuxième exon pour former une liaison phosphodiester et permettre la libération de l'intron qui sera dégradé par des **ribonucléases**



**Figure :** Excision de l'intron par formation d'un lasso

**1<sup>re</sup> étape :** clivage du côté 5' de l'intron par l'attaque de l'hydroxyle OH situé en 2' de L'adénine du branchement sur le groupe phosphodiester (cette attaque est rendue possible par des phénomènes de repliement dans la chaîne). IL se forme alors une Liaison 2'-5' entre L'adénine et la guanine à l'extrémité de l'intron qui crée un Lasso.

**2<sup>e</sup> étape :** le 3'OH de La guanine en fin de l'exon 1 attaque le groupe phosphodiester qui relie la guanine de L'exon 2 à l'intron du côté 3', ce qui provoque le clivage de cette Liaison.

**3<sup>e</sup> étape :** il se produit alors une soudure des deux exons, et libération du Lasso qui sera dégradé par des nucléases cellulaires.

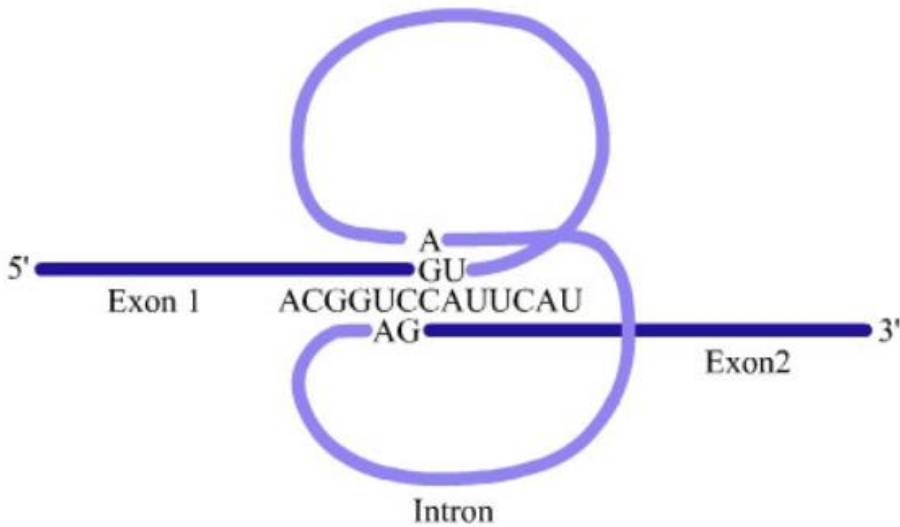
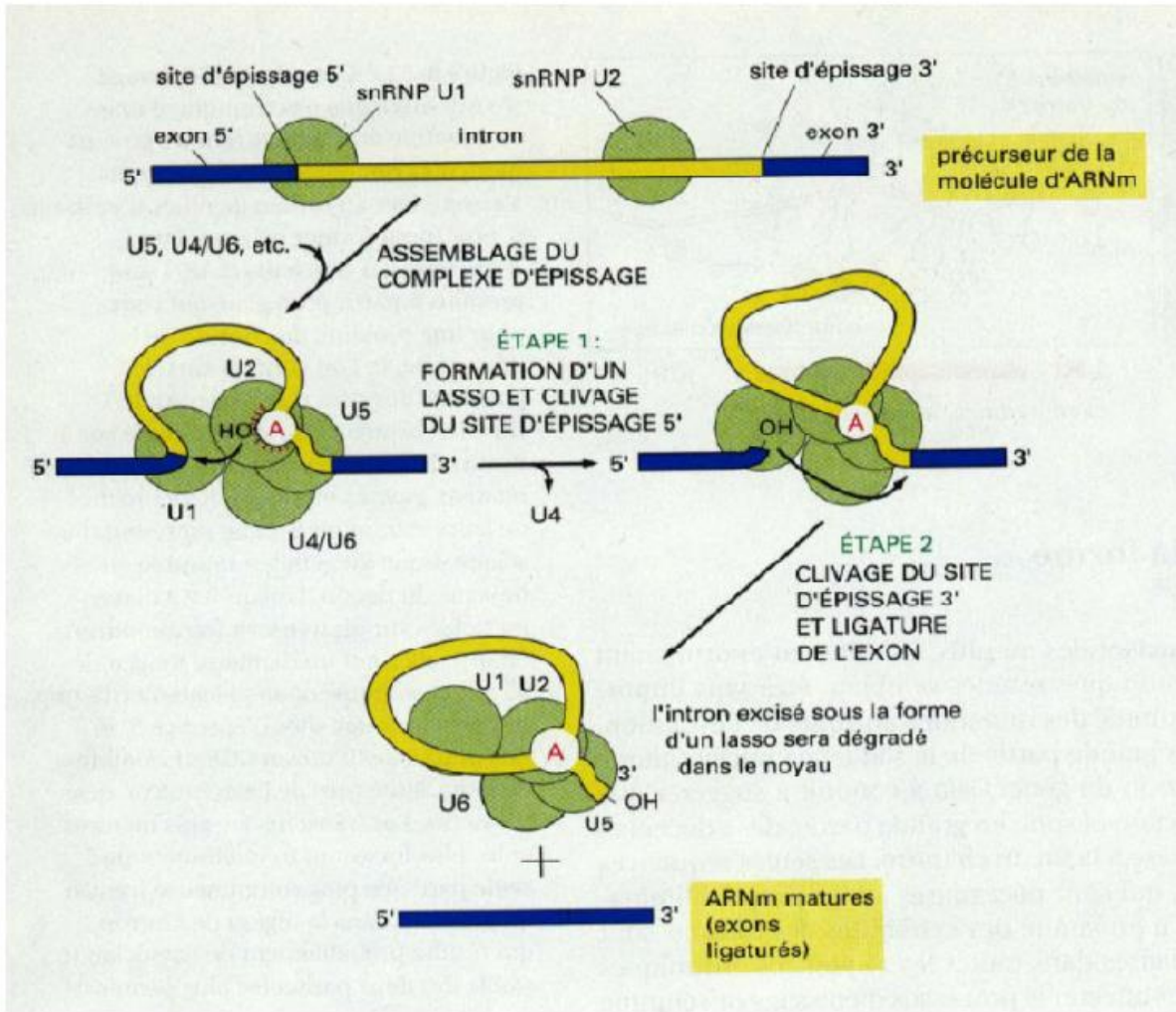


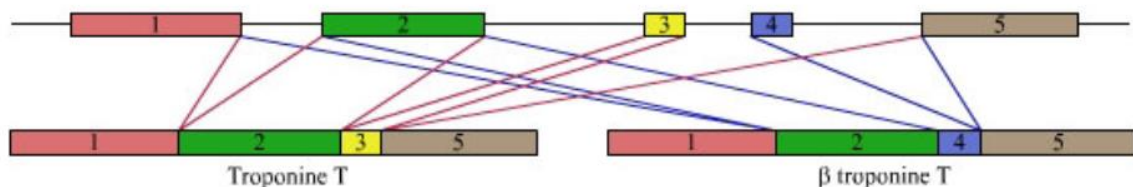
Figure : Excision de l'intron par formation d'un lasso : le splicéosome

### 3.4. L'épissage alternatif

A partir d'un transcrit primaire on peut avoir deux ou plus ARNm matures qui seront à l'origine de la formation des protéines-isoformes. Ceci est possible grâce à l'épissage alternatif qui consiste en **l'élimination de certains exons**. En effet certains exons sont constants au niveau des différents ARNm matures et d'autres sont variables et spécifiques du tissu dans lequel se trouve la protéine isoforme.

#### Exemple :

Au niveau du gène de la tropomyosine on met en évidence 12 exons au total dont 7 sont constants et 5 sont alternatifs



**Figure** : 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.

### 4. Structure des facteurs protéiques de la transcription

La régulation de la transcription des gènes est assurée en partie par l'intervention de facteurs protéiques agissant en « trans ». Certains de ces facteurs sont capables de se lier directement à l'ADN et d'autres interviennent dans des complexes protéiques sans se lier à l'ADN.

L'association de l'ensemble de ces protéines en complexes de haut poids moléculaire permet de réguler la transcription du gène en fonction des signaux cellulaires.

Ces protéines nucléaires sont capables de reconnaître et de se lier avec des régions promotrices de gènes cibles. Deux types de facteurs de transcription participent à la régulation de l'expression des gènes chez les eucaryotes : les facteurs généraux de la transcription qui interagissent avec la machinerie basale de la transcription au niveau du promoteur central et les facteurs de transcriptions spécifiques interagissant avec les régions distales des promoteurs.

Les facteurs de transcription possèdent généralement deux domaines caractéristiques : un site de liaison direct à une séquence spécifique d'ADN et un domaine d'activation ou de



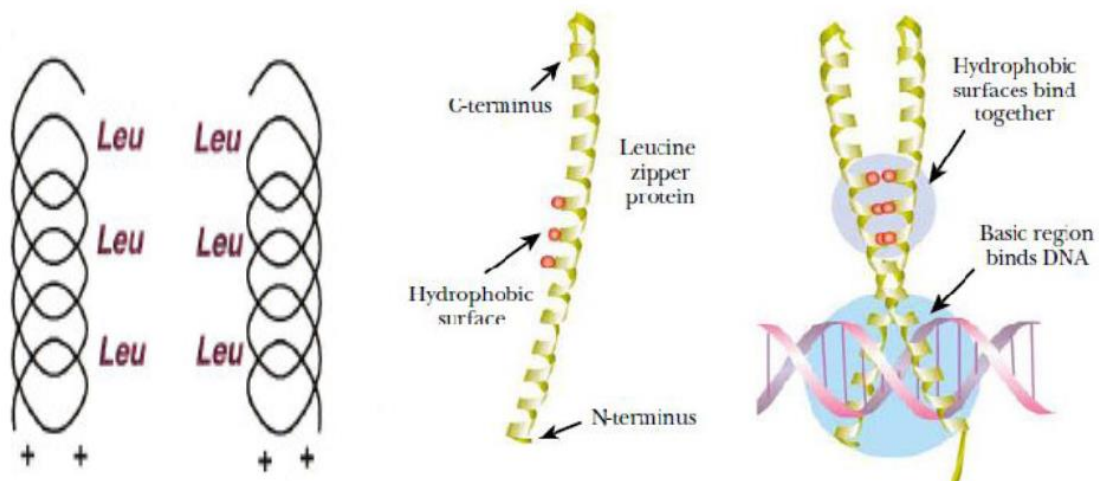
répression qui permet au facteur de transcription d'interagir avec des cofacteurs pour contrôler (activer ou réprimer) l'activation de la machinerie basale.

#### 4.1. Structure des activateurs

Il faut distinguer dans ces facteurs le domaine de liaison à l'ADN (noté DBD pour DNA Binding Domain), et un ou plusieurs domaines d'activation.

##### 4.1.1 Protéines à Leu Zipper

Il s'agit de deux hélices  $\alpha$  formant ainsi un dimère, reliées entre elles par des liaisons impliquant des AA hydrophobes tels que la leucine, ce qui crée des liaisons faibles susceptibles d'être assez facilement rompues puis recréées, ce qui évoque le système d'une fermeture éclair (Figure 5). En amont de cette région « Leu Zipper », on trouve une zone riche en AA basiques (chargés positivement) qui vont interagir avec les charges négatives de l'ADN, là encore par des liaisons faibles. C'est principalement par cette région que la liaison à l'ADN se fera.



**Figure:** structure des Leu-Zipper (Aouf, 2016)

##### 4.1.2 Protéines à doigts de zinc

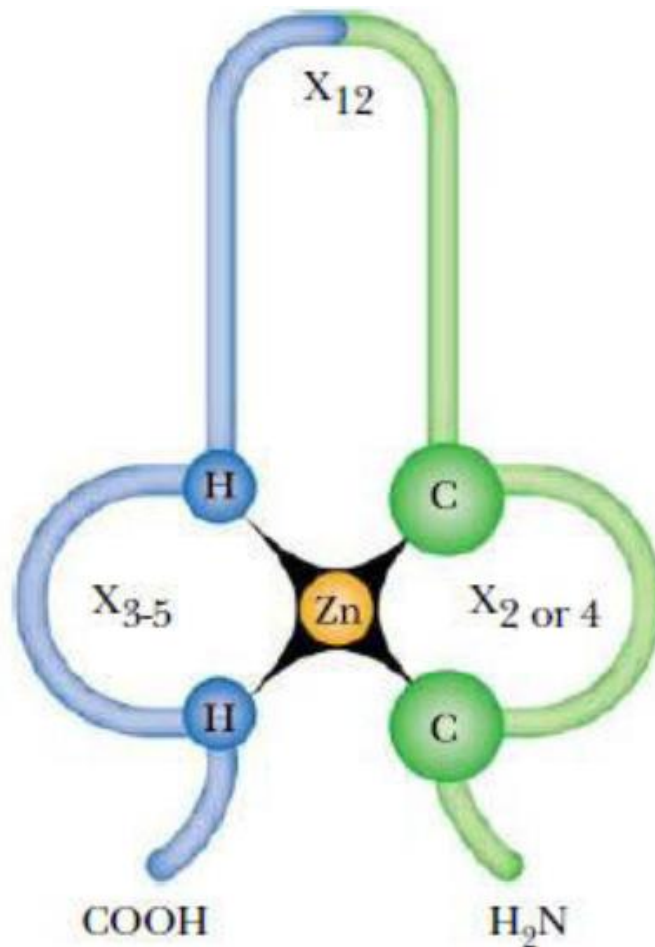
Il s'agit d'une chaîne peptidique repliée sur elle-même en boucle sur une trentaine d'AA, stabilisé par un ion  $Zn^{+2}$  situé à la base du doigt (figure 6). Cet ion échange des liaisons de coordination avec 4 AA eux aussi situés à la base du doigt (cystéines ou histidine). Ceci stabilise le reste du doigt qui entre en interaction avec l'ADN par l'intermédiaire d'AA hydrophobes chargés positivement.

L'utilisation du zinc dans les structures biologiques est assez rare, et on pense

d'ailleurs que c'est pour cela qu'il est utilisé ici ; en effet, ne faisant pas partie de chaînes d'oxydoréduction « classiques », cet ion zinc ne risque pas d'interagir avec l'ADN et d'y provoquer éventuellement des mutations.

On peut noter que certaines protéines possèdent plusieurs de ces structures en doigt de zinc, comme les hormones stéroïdiennes qui en comptent deux.

Certains doigts de zinc sont formés avec quatre cystéines, d'autres avec un mélange de cystéine et d'histidine (2 à 2).



**Figure :** structure des doigts de zinc (Aouf, 2016)

Les protéines présentant des motifs en doigts de zinc permettent d'activer la transcription en se liant à des régions riches en GC.

#### 4.1.3 Protéines à homeodomains

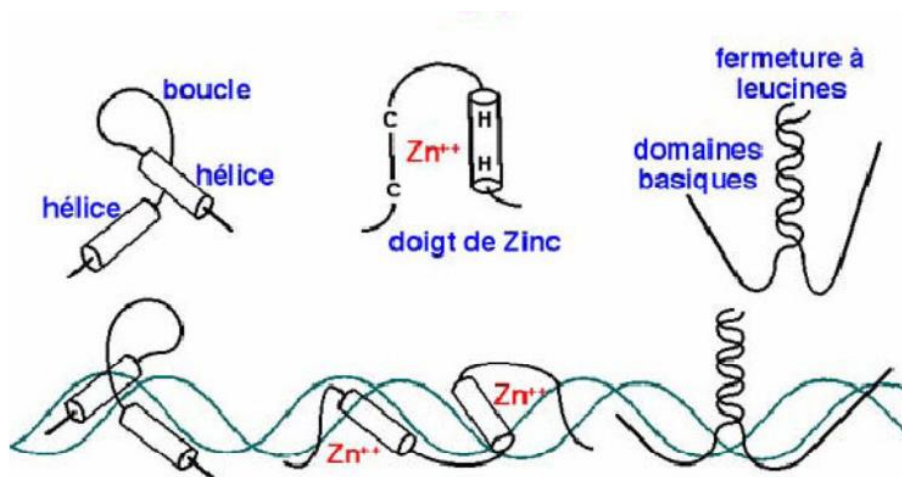
Ces protéines ont été tout d'abord rencontrées comme facteurs de régulation de gènes impliqués dans le développement embryonnaire, puis dans d'autres facteurs.

Leur structure est du type hélice tour — hélice (hélice  $\alpha$  — tour  $\beta$  — hélice  $\alpha$ ), dont une des

chaînes interagissent grâce à ses charges positives avec les charges négatives du grand sillon de l'ADN. Il semble que la structure contienne au total trois hélices  $\alpha$  et deux coudes  $\beta$  et que ce soit l'hélice n° 3 qui interagisse principalement avec le grand sillon de l'ADN par des charges positives.

Les protéines contenant des motifs HD reconnaissent généralement des structures d'ADN riches en bases AT.

Ces facteurs protéiques sont codés par des gènes dits homéobox, et sont donc impliqués dans le développement, mais également dans la forme du corps chez les vertébrés.



**Figure : Représentation schématique de domaines de liaison à l'ADN couramment rencontrés.**