

FLUX D'INFORMATION GÉNÉTIQUE

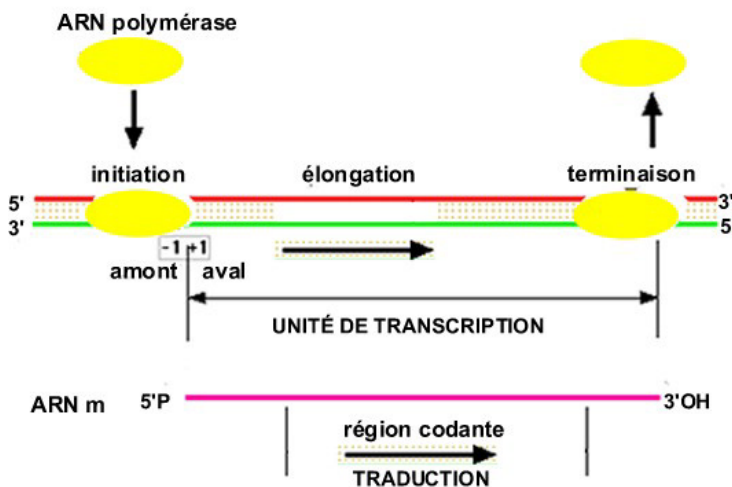
2 MÉCANISME GÉNÉRAL DE LA TRANSCRIPTION

La transcription est une biosynthèse d'ARN qui repose, comme celle de l'ADN, sur la **complémentarité des bases**.

Ce processus présente des analogies avec celui de la **réplication** mais également des différences fondamentales:

- Contrairement à la réplication qui intéresse la totalité du génome à chaque cycle, le programme de transcription n'est pas fixe : seules, de petites portions du génome sont transcrites à une époque donnée de la vie de la cellule et ces portions varient en fonction du développement, de l'environnement etc...

La transcription commence en un point précis de l'ADN pour se terminer en un point également précis, l'espace entre les deux constitue une **unité de transcription**, notion proche du cistron mais pas tout à fait identique.



- **Un seul brin** de l'ADN est transcrit, c'est à dire sert de modèle à la polymérisation des ribonucléotides. En effet, en un endroit donné, un seul brin de l'ADN a un sens en termes de protéine, c'est pourquoi l'on écrit généralement une séquence d'ADN sous forme d'une succession de bases de 5' à 3'. Dans cette convention on représente effectivement le brin qui possède le code (brin codant) mais, c'est l'autre brin (3' - 5' dans cette convention) qui est transcrit. Le résultat est une molécule d'ARN dont l'orientation 5' 3' correspond à l'orientation NH₂ - COOH de la protéine. La lecture du code (la traduction) se fait dans le même sens que la transcription.

* *Remarque* : Tout le produit de transcription ne correspond pas à la séquence qui code de la protéine, côté 5' (amont) une séquence guide ou «leader» permet la fixation de l'ARN messager au ribosome.

- La transcription est assurée par une **ARN polymérase** qui utilise l'ADN simple brin (comme pour la réplication, une dénaturation locale de la molécule d'ADN est nécessaire) mais elle polymérise des ribonucléotides en regard des désoxyribonucléotides.

* *Remarque* : Contrairement à l'ADN polymérase qui ne fait qu'allonger des chaînes préexistantes, l'ARN polymérase peut réaliser un dinucléotide au point d'initiation de la transcription .

De nombreux facteurs protéiques interviennent également pour assurer l'initiation, l'élongation et la terminaison mais ce ne sont pas les mêmes que ceux intervenant dans la réplication.

Il existe des différences sensibles entre la transcription chez les procaryotes et celle des eucaryotes et des différences encore plus importantes dans son contrôle, c'est pourquoi leur étude sera séparée.

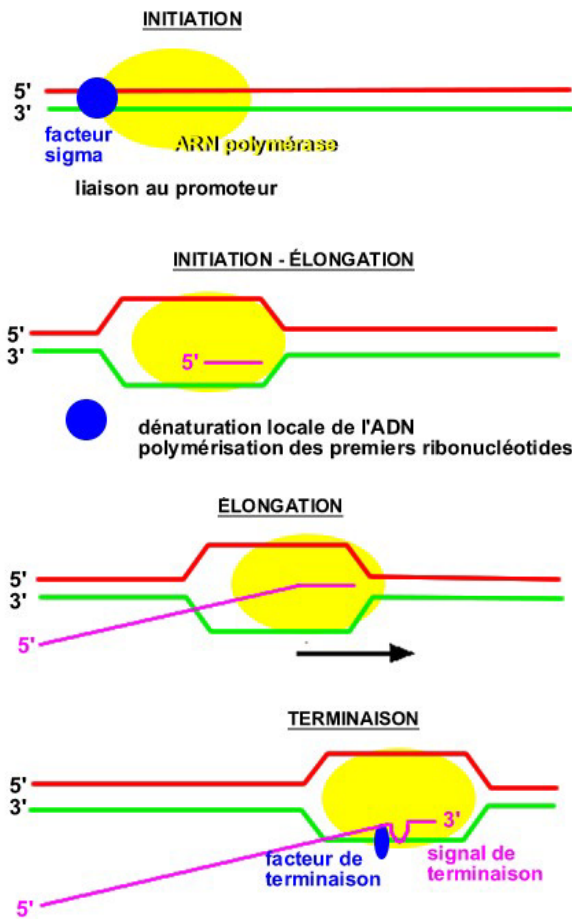
2.1 TRANSCRIPTION CHEZ LES PROCARYOTES

L'ARN polymérase d'*E. coli* est une enzyme assez complexe composée de quatre sous unités : deux α , une β , une β' et d'autres éléments protéiques qui établissent des liaisons temporaires avec ce complexe. Cette enzyme réalise la polymérisation de ribonucléotides face à un brin d'ADN mais, *in vitro*, commence n'importe où et s'arrête n'importe où.

Une protéine particulière, le **facteur sigma**, est capable, *in vivo* d'assurer la reconnaissance de séquences spécifiques situées légèrement en amont du point d'initiation de la transcription qui constituent le **promoteur**. L'installation de l'ARN polymérase au niveau précis de ces promoteurs (grâce au facteur σ) va permettre la transcription du bon brin en commençant par le bon nucléotide. C'est le point crucial de l'**initiation**.

Au démarrage de l'**élongation**, le facteur σ quitte le complexe.

Cette étape marque le passage à l'allongement de la chaîne d'ARN de 5' vers 3' jusqu'au site de **terminaison**.



Les schémas ci-dessus impliquent que la transcription correcte, dans l'espace, d'une molécule d'ARN, représentative d'une **unité d'information**, repose sur le point de départ et sur le point de terminaison.

Comment ces nucléotides précis sont-ils déterminés?

Ce sont les **séquences** promotrices et les sites de terminaison qui contiennent les informations nécessaires.

Plus d'une centaine de promoteurs d'*E.coli* ont été séquencés et l'alignement des séquences par rapport au point de départ de la transcription fait ressortir ce que l'on appelle des **séquences consensus** c'est à dire qui se retrouvent très fréquemment, au même endroit dans les différents **promoteurs**.

A C A T G C A G T A A A T A C A A A T C G A T
 T G G C G G T G A T A A T G G T T G C A T C
 C G G G T T G T A T G T T G T G T G T A A
 T C T T T G T T A T G A T A T G G T T A T
 A A C T C T C T A C T G T T T C T C T A T
 G C G T T T T T G T C A T G G C T T T G G
 G A A C T A G T T A A A T A G T A C G C A
 T T C G G T G T A G A C T T G T A A A C C
 A A A G A T T T A G G A T T A C A A G T C

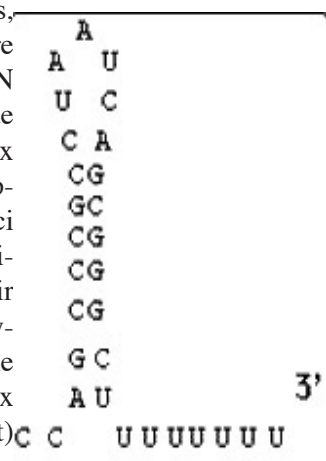
consensus : T A T A A T

Une première, en position -10 est appelée « cassette » -10 ou cassette **TATA** (TATA box en anglais) en raison de sa composition ; une autre est une **cas-**

sette -35 dont le consensus est TTGACA. Toute mutation dans ces régions affecte l'initiation de la transcription, soit dans le choix du brin à transcrire, soit dans la position du point de départ. Des délétions qui modifient les distances de ces séquences, entre elles ou par rapport au point de départ affectent également la transcription.

Le promoteur est donc défini par des séquences d'initiation types en positions précises.

La terminaison semble pouvoir se faire selon au moins deux mécanismes selon les renseignements apportés par des expériences de transcription *in vitro*. Pour certains gènes, elle dépend de la structure secondaire prise par l'ARN lui-même. La séquence de l'extrémité 3' de nombreux messagers bactériens se rapproche de celle présentée ci contre, le signal de terminaison est ici reconnu après avoir été transcrit, la séquence polyuridilique en 3' implique une série de paires A-U (à deux liaisons hydrogène seulement)



Dans d'autres cas, un facteur protéique spécifique, le facteur rho, est nécessaire pour la terminaison et la libération de la molécule d'ARN.

**Remarque : Les trois classes d'ARN : ARN ribosomiques, ARN de transfert et ARN messagers sont synthétisés selon ce mécanisme de transcription. Il faut noter que, sur le plan qualitatif, les cistrons ribosomiques et de transfert sont peu diversifiés et occupent une petite partie du génome bien que sur le plan quantitatif, leurs produits puissent représenter plus de 90% de l'ARN cellulaire alors que l'ARN messager, peu représenté dans le cytoplasme peut l'être à partir de gènes très variés qui se répartissent sur une grande partie du génome. Ce paradoxe est accentué dans la cellule eucaryotique.*

2.2 LA TRANSCRIPTION CHEZ LES EUKARYOTES

Le principe est bien entendu le même : il s'agit de transcrire un code contenu dans la séquence de l'ADN dans une molécule d'ARN mais les modalités diffèrent.

2.2.1 LES ARN POLYMÉRASES

Alors qu'une seule enzyme synthétise toutes les catégories d'ARN des procaryotes, dans la cellule eucaryotique, trois ARN polymérases sont mises en jeu dans la transcription d'ensembles de cistrons différents.

- **L'ARN polymérase A** (ou I) transcrit les **gènes ribosomiques** (qui restent souvent appelés cistrons ribosomiques), elle assure la synthèse des ARN des ribosomes. Son lieu d'action est le **nucléole**, en effet, les très nombreux cistrons ribosomiques répétés (c'est une de leur caractéristique), sont groupés au niveau de *loci* précis : les **organismes nucléolaires**. A l'interphase, la transcription active de ces cistrons se manifeste sous forme de structures cytologiques : les nucléoles.

- **L'ARN polymérase B** (ou II) réalise la synthèse de tous les **ARN messagers** nucléaires, qui seront traduits en protéines. Ces gènes sont regroupés dans ce que l'on appelle «gènes de la classe II».

- **L'ARN polymérase C** (ou III) assure la synthèse des petits ARN nucléaires (ARN de transfert et ARN ribosomique 5S) et la transcription de gènes cytoplasmiques (contenus dans les mitochondries et les plastides des végétaux).

Toutes ces polymérases sont des édifices protéiques complexes composés de quatre sous unités principales, rappelant celles des procaryotes, auxquelles s'ajoutent jusqu'à une douzaine de protéomères. De plus, ces complexes ne fonctionnent, *in vivo* que grâce à l'interaction de nombreux autres facteurs protéiques établissant des relations temporaires avec le complexe enzymatique de base.

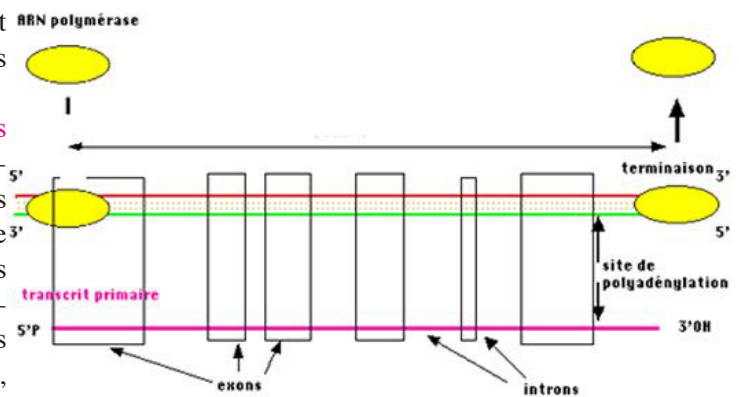
Ce que l'on connaît de la transcription des cistrons ribosomiques rappelle la transcription de l'ADN procaryotique et nous ne nous y attarderons pas.

**Remarque : Le mécanisme de transcription «ressemble à celui des Procaryotes» ne veut pas dire que les systèmes de régulation soient identiques.*

2.2.3 TRANSCRIPTION DES GÈNES EXPRIMÉS EN PROTÉINES

La transcription des gènes de la classe II est assurée par la polymérase B mais, parmi eux, il faut distinguer plusieurs types d'organisation.

Quelques rares gènes présentent des unités de transcription semblables à celles des procaryotes, le plus souvent, les gènes sont structurés en **mosaïque** de portions codantes (les **exons**) et de portions n'ayant pas de signification protéique (les **introns**). L'organisation générale d'un messenger eucaryotique est présentée ci-dessous.



La synthèse d'ARN chez les eucaryotes donne généralement naissance à un produit de transcription «primaire» (ou **prémessenger**) qui devra subir une **maturation** pour fournir le messenger cytoplasmique fonctionnel.

Nous allons donc distinguer les deux étapes dans la réalisation d'un ARN messenger fonctionnel : transcription et maturation.

2.2.3.1 TRANSCRIPTION.

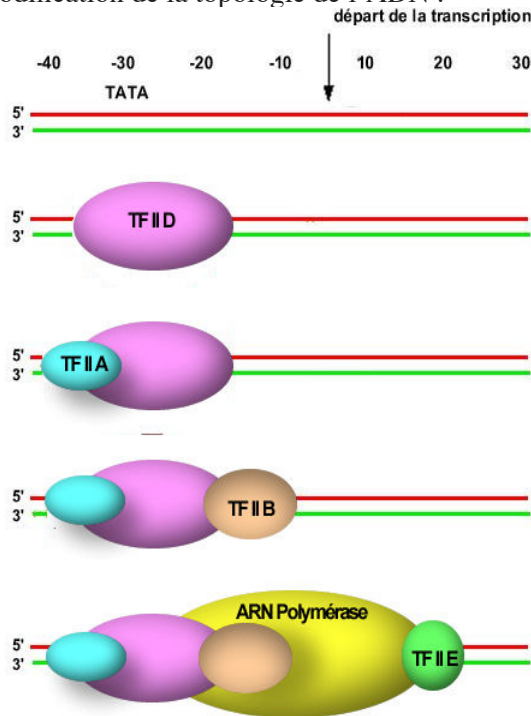
Elle est assurée par la polymérase B mais l'initiation (en particulier) fait intervenir d'autres facteurs protéiques du groupe des **TF II**.

** Remarque : Les ARN polymérases A, B et C avaient été initialement baptisées I, II et III en raison de leur ordre d'éluition de certains supports chromatographiques, TF II signifie «Transcription Factor» nécessaire au fonctionnement de la polymérase II (ou B).*

La formation du complexe d'initiation nécessaire à la fixation précise de l'ARN polymérase B est résumée dans la figure suivante.

Un premier complexe protéique, TFII-D reconnaît le promoteur et permet la fixation de TFII-A puis des interactions protéiques entre cet assemblage,

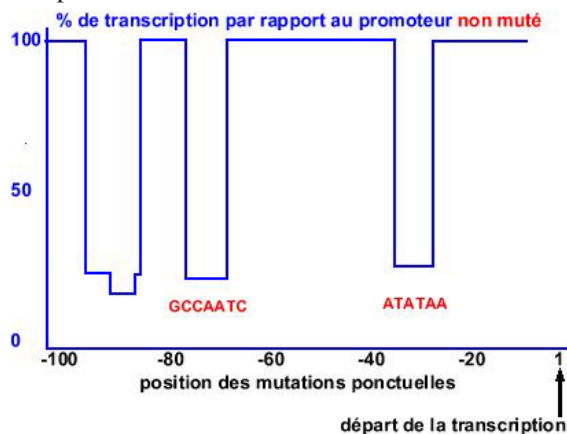
TFII-B et la polymérase B permettent la fixation de celle-ci, son maintien est assuré par TFI-EI . D'autres facteurs (TFII-H et J) participent à la modification de la topologie de l'ADN .



C'est au prix de cette complexité que la transcription est correctement initiée !

Ici encore, tout repose sur l'interaction de protéines spécifiques avec des séquences déterminées. Les promoteurs des eucaryotes, reconnus par l'ARN polymérase B, ont été étudiés à l'aide des outils les plus performants de la génétique moléculaire.

On note l'effet de mutations induites à presque chaque position sur les 100 paires de bases en amont du point de début de transcription sur l'efficacité de l'initiation *in vitro*. La figure suivante montre que la plupart des modifications introduites artificiellement n'affectent pas la reconnaissance du promoteur, seuls, dans l'exemple choisi (promoteur du gène de la globine), trois modules très restreints aux environs de -30, -75 et -90 sont indispensables.



Des expériences du même genre mais en introduisant des délétions entre ces modules ont montré que, pour un gène donné, leur position joue également un rôle.

Le terme de modules (ou cassette) prend toute sa signification quant on sait que les différents promoteurs en contiennent des assortiments variés (certains en plusieurs exemplaires et certains en orientation inverse).

- La cassette TATA vers -30 est probablement la seule séquence consensus située à une position fixe présente dans la quasi totalité des gènes de la classe II, elle semble jouer un rôle dans la précision du choix du premier nucléotide transcrit.

- La cassette CAAT est située à -80 dans l'exemple donné du gène de la globine mais sa position peut varier selon les gènes.

- La cassette GC (en raison de sa composition en bases) est souvent en plusieurs copies et dans les deux orientations.

Ces modules de base représentent donc un signal pour le complexe d'initiation de la polymérase B mais leur assemblage en différentes combinaisons permet une diversification qui interviendra dans la régulation de la transcription.

Dans un chapitre suivant, on verra que d'autres modules, parfois très éloignés du promoteur «de base», jouent également un rôle dans la régulation.

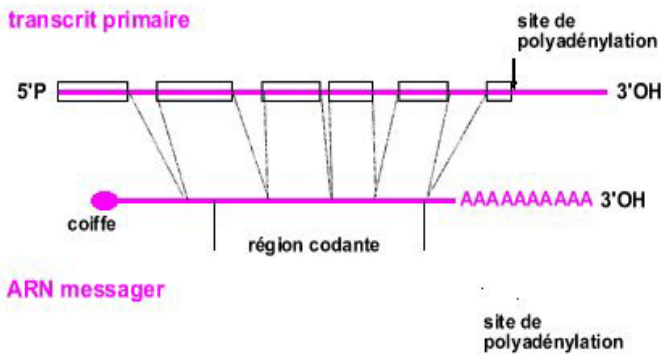
Bien entendu, l'étape d'initiation de la transcription est cruciale, mais il existe également des éléments nécessaires à l'élongation et à la terminaison qui peuvent jouer un rôle dans le contrôle de la synthèse d'ARN messagers.

**Remarque : Les signaux de terminaison sont moins bien connus que chez les procaryotes, par contre on connaît des signaux de polyadénylation (l'ajout d'une séquence polyadénylée en 3' OH des messagers eucaryotiques fait partie de la régulation).*

2.2.3.2 MATURATION

La maturation de la plupart des transcrits primaires d'ARN porte sur 3 points :

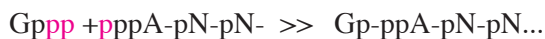
- la formation d'une structure particulière (coiffe) en 5'.
- l'adjonction d'une séquence polyadénylée en 3'
- l'épissage : excision des introns et jonction des exons.



Le **coiffe** est formée par addition d'une guanosine triphosphate. Le premier nucléotide du messenger est généralement une purine, A ou G et représente théoriquement l'extrémité 5' triphosphorylée de la molécule:



en fait, une guanine est ajoutée par une liaison inhabituelle 5'-5':



Diverses méthylations peuvent se produire ensuite notamment une en position 7 de la guanine et sur le ribose pour compléter cette structure que l'on retrouve dans tous les messagers eucaryotiques.

La **polyadénylation** est un ajout **post-transcriptionnel** de nucléotides adényliques au niveau d'un site de polyadénylation du transcrit primaire. Le site est reconnu par un complexe protéique comportant une poly-A polymérase. Cette structure va former l'extrémité 3' du messenger, elle peut aller jusqu'à 200 nucléotides .



L'épissage :

Les **gènes mosaïques** sont transcrits depuis le point d'initiation jusqu'au signal de terminaison, les introns de ce pré-messenger (ou transcrit primaire) doivent donc être éliminés et la jonction des exons se faire avec précision. L'ensemble se fait simultanément par un mécanisme d'**épissage**.

**Remarque : ce que nous savons du code génétique et du cadre de lecture permet de comprendre à quel point l'épissage doit être précis, la jonction exon - intron ne coïncidant d'ailleurs pas forcément à une limite entre deux triplets.*

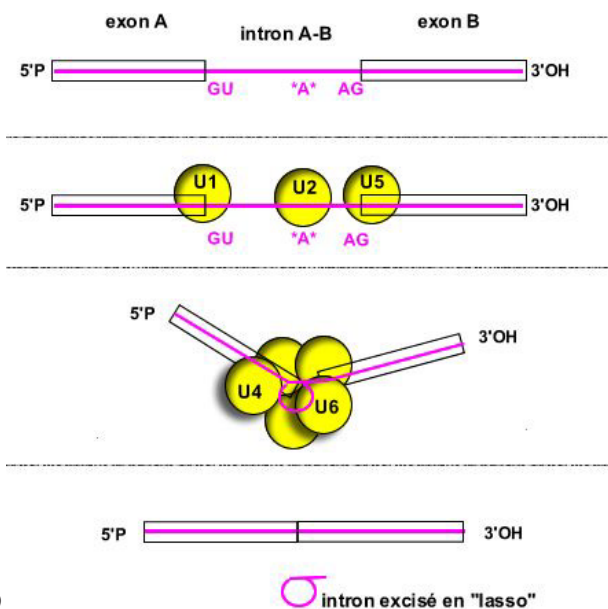
Selon les unités de transcription envisagées, il existe plusieurs mécanismes d'épissage, tous font intervenir d'autres molécules d'ARN que le transcrit primaire, dans certains cas, l'ARN seul (non associé à des protéines) suffit à catalyser la réaction.

*. * Remarque : La découverte d'une activité catalytique liée à un acide nucléique (les ribozymes) a un impact considérable en génétique évolutive.*

Le mécanisme décrit ici est celui utilisé pour la plupart des messagers nucléaires, il fait appel à des **ribonucléoprotéines** dont les deux composantes ont certainement leur importance, mais c'est la molécule d'ARN qui joue probablement un rôle prépondérant.

L'analyse de nombreux introns de levure a fait apparaître trois séquences consensus que l'on a pu généraliser : la première, GU est appelée consensus gauche car elle représente l'extrémité 5' de l'intron, de même que la droite (AG) représente la jonction 3' intron - 5' exon suivant. Une autre, UACUAAC chez les messagers de levure (plus généralement Py N Py Py Pu A Py chez les eucaryotes supérieurs (Py représente une pyrimidine, N un nucléotide quelconque, Pu une purine), l'adénine étant remarquablement constante) est située peu avant l'extrémité 3' de l'intron et appelée séquence de branchement pour les raisons évoquées ci-dessous.

Des mutations dans les nucléotides codant pour ces trois séquences d'un intron empêchent l'épissage correct des exons qui l'encadrent. Ces séquences sont reconnues par des ribonucléoprotéines (en jaune sur le schéma ci-dessous) qui vont former un complexe nécessaire à l'épissage, appelé **spliceosome** (de splicing = épissage), visible en microscopie électronique.



Les molécules d'ARN de ces complexes, connues depuis longtemps sans qu'on ait pu, au départ, leur attribuer une fonction, appartiennent au groupe des «petits ARN nucléaires» : classes discrètes de 100 à 1000 nucléotides, qui ne sortent pas du noyau, ont un turn-over très lent, sont riches en uridines, et dont chaque molécule est associée à une dizaine de protéines pour former un complexe parfois appelé un Snurp (Small Nuclear Uridine-rich RibonucleoProtein).

L'édifice représenté sur la figure, imposé par la liaison des différents Snurps, va replier l'intron et permettre une liaison curieuse de l'extrémité 5' de l'intron à l'hydroxyle 2' de l'adénine de la séquence consensus dite «de branchement». L'extrémité 3' va être détachée de la 5' de l'exon suivant et une liaison phosphodiester peut se créer entre les exons.

**Remarque : Le branchement en 2' de l'adénosine fait prendre une structure dite en «lasso» à l'intron. In vivo, ce lasso est immédiatement dégradé mais dans des expériences in vitro on peut l'observer en microscopie électronique.*

Ce mécanisme complexe est génétiquement contrôlé et joue un rôle important dans la régulation de l'expression des gènes chez les eucaryotes.