



HAL
open science

Interaction du snARN U1 de l'épissage avec l'ARN polymérase II

Béatrice Spiluttini

► **To cite this version:**

Béatrice Spiluttini. Interaction du snARN U1 de l'épissage avec l'ARN polymérase II. Biologie cellulaire. Université Pierre et Marie Curie - Paris VI, 2009. Français. NNT: 2009PA066109. tel-00814598

HAL Id: tel-00814598

<https://theses.hal.science/tel-00814598>

Submitted on 17 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité

Biologie cellulaire et Moléculaire
(Complexité du Vivant)

Présentée par

Mlle. Béatrice Spiluttini

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

Interaction du snARN U1 de l'épissage avec l'ARN polymérase II.

soutenue le 24 mars 2009

devant le jury composé de :

M. Olivier Bensaude, Directeur de thèse

MM. Didier Auboeuf et Edouard Bertrand, Rapporteurs

M. Jean-Michel Rossignol, Examineur

M. Pierre Netter, Président du jury

Table des matières

I	Introduction	1
1	Transcription et épissage	3
1.1	Les étapes de la transcription	3
1.1.1	La région génomique d'assemblage du complexe de transcription	6
1.1.2	Le complexe de pré-initiation et les facteurs généraux de la transcription	10
1.1.3	La transcription par la RNAPII	13
1.2	Epissage	21
1.2.1	Le mécanisme de l'épissage	24
1.2.2	Vue d'ensemble	26
1.2.3	Les complexes du splicéosome	26
1.2.4	Les UsnRNP	30
1.2.5	Définition de l'exon, éléments régulateurs de l'épissage et épissage alternatif	31
2	ARN polymérase II	37
2.1	Le domaine carboxy-terminal	37
2.1.1	Structure et modifications	37
2.1.2	Le CTD, une plate-forme de recrutement	43
2.1.3	Rôle du CTD dans des mécanismes liés à la maturation du transcrit	46
2.2	L'épissage cotranscriptionnel	48
2.2.1	Introduction	48
2.2.2	Influence de la transcription sur l'épissage	49

3 Les ARNs non codants	57
3.1 Les ARNs impliqués dans des processus généraux	57
3.2 Les ARNs non codants impliqués dans la régulation de l'expression génique	60
3.2.1 L'ARN non-codant SRA	60
3.2.2 Les ARN silenciers	60
3.3 Les ncARNs qui régulent l'activité de l'ARN polymérase II	66
3.3.1 L'ARN 6S	66
3.3.2 L'ARN B2	69
3.3.3 L'ARN Alu	72
4 Objectifs	73
II Matériels et méthodes	75
4.1 Analyse radiale	77
4.1.1 Focuser	77
4.1.2 Automatic density analysis	82
4.1.3 Présentation des résultats	84
4.2 Microscopie photonique	88
4.3 Immunoprécipitations natives	88
4.4 Immunoprécipitations en crosslink	89
4.5 Marquage des ARNs au pCp	91
4.6 Gels	92
4.7 Cellules et Anticorps	92
4.8 Anticorps	93
4.9 FISH (Fluorescence In Situ Hybridization) adapted for mammalian cells	93
4.10 Immunofluorescence	96
III Résultats	99
5 Association RNAPII-U1 indépendamment de la transcription	101

5.1	U1 immunoprécipité avec la RNAPII	102
5.2	Autres anticorps anti-RNAPII	109
5.3	U1 n'est pas tout seul	113
6	U1 recruté cotranscriptionnellement, indépendamment de l'épissage	117
6.1	La technique utilisée	117
6.1.1	Le système d'imagerie	117
6.1.2	L'analyse des images	122
6.1.3	Transcription	123
6.2	Recrutement du splicéosome	124
6.2.1	Les snARNS U4, U5, U6	124
6.2.2	Le snARN U2	128
6.2.3	Le snARN U1	128
6.2.4	La snRNP U1	129
6.3	Protéines annexes	131
6.4	Analyse radiale de l'ARN U1 et de l'ARN B2 sur les sites actine chez la souris	135
6.5	Rôle du CTD de la RNAPII dans l'enrichissement du snARN U1 au site de transcription	138
IV	Discussion et perspectives	141
6.6	Association RNAPII – snRNP U1	143
6.7	Pré-requis de l'interaction RNAPII – snRNP U1	143
6.8	La snRNP U1 présente au site de transcription indépendamment de l'épissage.	146
6.9	Rôle(s) de l'association RNAPII – snRNP U1	148
6.10	Un ARN sans intron est exprimé mais en moins grande quantité.	152
V	Annexes	155
A	Tests statistiques sur les données obtenues par analyse radiale	157
B	Représentation des données	163

Bibliographie

167

Table des figures

1.1	Processus et régulation de la transcription	5
1.2	Les classes de sites d'initiation chez les promoteurs de mammifères	8
1.3	Les motifs des cœur-promoteurs	10
1.4	Structure de l'ARN polymérase à travers les trois domaines du vivant	14
1.5	Les différents états de la RNAPII	17
1.6	Structure cristallographique de la RNAPII	19
1.7	Mécanismes élémentaires de l'épissage alternatif	23
1.8	Le cycle du splicéosome et le complexe E.	25
1.9	Transesterification	29
1.10	Définition par l'exon ou par l'intron et les séquences régulatrices de l'épissage	33
1.11	Sites et réactions catalytiques de l'épissage	35
2.1	Représentation de la RNAPII et de son CTD à l'échelle	38
2.2	Le code du CTD	40
2.3	Patron de phosphorylation du CTD de la RNAPII chez la levure	42
2.4	Les modifications dynamiques de la phosphorylation du CTD coordonnent le cycle de la RNAPII avec la maturation des pré-ARNm et la modification des histones	45
2.5	Influence de la RNAPII sur l'épissage	52
3.1	Biogenèse des petites RNPs	59
3.2	Les ARNs silencer	62
3.3	Le modèle Ping-Pong des piARNs	65

3.4	L'ARN 6S mime le promoteur ADN	68
3.5	Modèle de l'inhibition de la formation du complexe promoteur ouvert par l'ARN FC*	70
3.6	Structure de l'ARN B2	71
4.1	Plugin focuser	77
4.2	Analyse radiale	86
4.3	Analyse par bootstrap	87
5.1	Solubilité de la RNAPII dans des cellules non synchronisées et mitotiques	102
5.2	ARNs immunoprécipités avec la RNAPII	103
5.3	L'ARN U1 est co-immunoprécipité avec l'ARN polymérase II	105
5.4	L'ARN U1 est associé <i>in vivo</i> à la RNAPII.	107
5.5	Immunoprécipitation de la RNAPII par différents anticorps	112
5.6	Localisation des protéines associées à l'ARN U1 sur gradient de glycérol	114
6.1	Insertion multicopies dans le génome des cellules U2OS	120
6.2	Constructions épissées et non épissées à partir du gène de la β -globine	121
6.3	Lignées cellulaires stables WT et MUT	122
6.4	Analyse radiale	124
6.5	Différentes densités de RNAPII aux sites pour une même densité de transcrits	125
6.6	L'ARN U1 est le seul ARN de l'épissage à être enrichi au site de transcription	127
6.7	La snRNP U1 est recrutée au site de transcription indépendamment de l'épissage	130
6.8	Recrutement spécifique de certaines protéines SR au site de transcription non épissé	132
6.9	Analyse radiale de TAT-SF1-GFP	135
6.10	B2 sur des cellules NIH3T3 au site actine	137
6.11	Expression d'une RNAPII-CTD-GFP ou RNAPII- Δ 5-GFP dans les lignées MUT et WT. Analyse sur l'ARN U1	140
A.1	Bootstrap ARN	159
A.2	Bootstrap ARN	160
A.3	Bootstrap protéines	161

A.4 Bootstrap protéines	162
B.1 Choix de la représentation	164
B.2 Boxplot	166

Liste des tableaux

1.1	Facteurs généraux de la transcription.	12
1.2	Composition des 3 RNAP eucaryotes	15
1.3	Les protéines spécifiques des UsnRNPs chez l'homme et la levure	31
5.1	Les protéines de la snRNPU1	113

Première partie

Introduction

Chapitre 1

Transcription et épissage

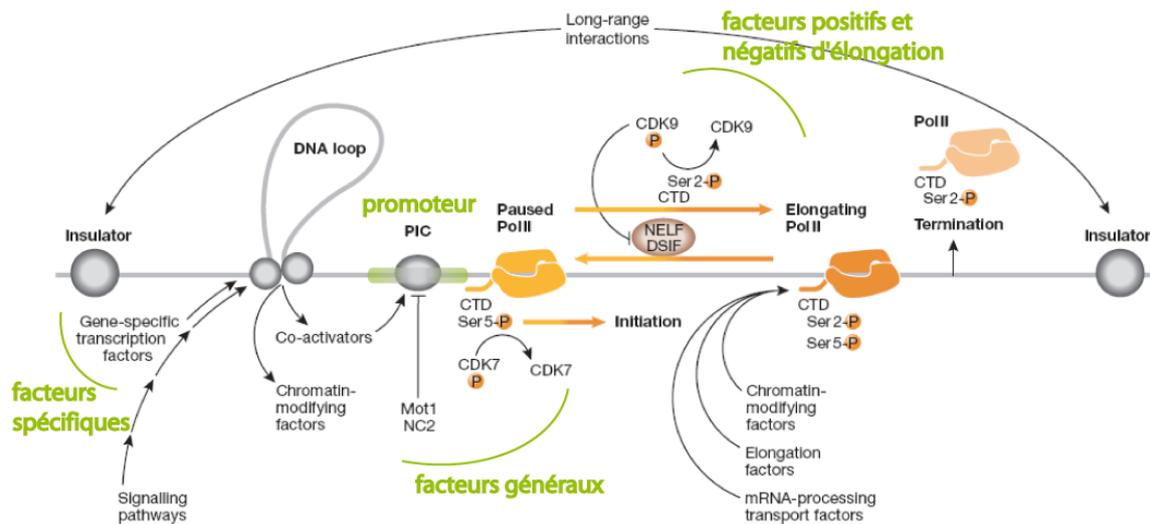
Les résultats obtenus au cours de ma thèse font intervenir deux acteurs majeurs de l'expression des gènes : l'ARN polymérase II dans la transcription et le petit ARN nucléaire (small nuclear RNA ou sn) U1 dans l'épissage. Afin de comprendre quelles sont les répercussions de leur interaction qui sera démontrée dans mes résultats, voici une présentation de ces deux mécanismes-clefs de l'expression des gènes.

1.1 Les étapes de la transcription

La transcription est le mécanisme qui permet la synthèse d'ARN à partir de l'ADN conservé dans le noyau. L'ARNm (messenger) pourra, après plusieurs modifications, être exporté vers le cytoplasme où il sera traduit en protéine qui pourra à son tour être modifiée et amenée là où elle doit exercer sa fonction. Ceci est vrai pour la transcription par l'ARN polymérase II (RNAPII). Les ARN polymérase I et III produisent des ARNs non codants qui vont subir d'autres types de modifications afin d'exercer leur fonction (comme les ARNr (ribosomiques) par exemple). Je ne développerai ici que la transcription par la RNAPII. Pour être menée à bien, la transcription requiert l'intervention constante de nombreux facteurs. En effet, certains gènes sont transcrits en permanence, ce sont les gènes dits « de ménage » qui maintiennent les fonctions vitales de la cellule. D'autres gènes, dits « spécifiques », ne sont transcrits que sous l'impulsion d'un signal (phase du cycle cellulaire, stress, différenciation, etc.). Ces deux classes de gènes font appel à une machinerie transcrip-

nelle de base identique. Cette machinerie de base est composée de l'ARN polymérase II (RNAPII) et de facteurs généraux de la transcription. Les deux classes de gènes divergent sur leur type de promoteur (ensemble de séquences permettant l'arrimage de la machinerie transcriptionnelle) et sur les facteurs périphériques de transcription (facteurs de transcription spécifiques d'un gène ou d'une classe de gènes). Ici vont être présentées les différentes étapes permettant la transcription d'un gène avec la présentation des éléments-clefs, acteurs de ce mécanisme.

Comme le décrit la figure 1.1, pour permettre la transcription d'un gène, il faut un promoteur sur lequel vont s'arrimer les facteurs généraux de la transcription. Ceux-ci vont se recruter les uns les autres dans un ordre défini par des expériences *in vitro*. L'assemblage de ces facteurs prépare l'association de la RNAPII. Intégrée à ce complexe, la RNAPII commence la transcription. A ce stade, des facteurs positifs et négatifs d'élongation vont réguler la mise en route de la transcription, permettant à la RNAPII de commencer à transcrire ou bien de faire une pause. Comme on le voit sur le schéma, les signaux qui vont permettre aux facteurs généraux de venir se fixer au promoteur sont générés par la fixation de facteurs spécifiques de la transcription d'un gène ou d'une classe de gènes, à des dizaines ou des centaines de kilobases en amont du promoteur. Ces signaux sont transmis par des coactivateurs tels que le médiateur chez la levure. Je ne détaillerai pas cette partie dans mon manuscrit bien que le médiateur sera présenté dans un tableau plus loin, car il est aussi considéré comme un facteur général de la transcription. Au cours de la transcription, la RNAPII va recruter différents facteurs pour lui permettre de progresser dans la transcription, de modifier la chromatine mais aussi de rendre mature l'ARN naissant et de l'exporter. Ces derniers événements seront traités dans les parties Epissage et Epissage co-transcriptionnel.



d'après Hahn 2008

FIGURE 1.1 – **Processus et régulation de la transcription.** Des voies de signalisation activent des facteurs de transcription spécifique de gènes qui se lient à des séquences régulatrices du gène. Des facteurs éloignés peuvent interagir avec des facteurs proches du promoteur par des repliements de l'ADN. Ces facteurs spécifiques peuvent recruter des facteurs modifiant la chromatine ou bien des complexes coactivateurs de transcription, à proximité des sites d'initiation de la transcription. Les coactivateurs recrutent et probablement stimulent l'activité du complexe de pré-initiation (PIC). L'assemblage du PIC peut être modulé par des répresseurs que sont Mot1 et NC2 chez la levure et la Drosophile (empêche la transcription dépendante de la boîte TATA et favorise la transcription dépendante de la séquence DPE [136]). A l'initiation de la transcription, le CTD de la RNAPII est phosphorylé en Ser₅ par la kinase cycline-dépendante cdk7. L'élongation de la transcription chez les eucaryotes supérieurs est souvent bloquée par l'action des facteurs négatifs d'élongation NELF et DSIF qui provoquent la pause de la RNAPII. La RNAPII sort de la pause grâce à l'action de cdk9 qui phosphoryle le CTD de la RNAPII sur les Ser2 ainsi que NELF et DSIF. La RNAPII en élongation phosphorylée en Ser 5 et Ser2 sur son CTD recrute des facteurs impliqués dans la modification de la chromatine, l'élongation de la transcription, la maturation des ARNm, le transport des ARNm et la terminaison. Les facteurs modifiant la chromatine associés au complexe d'élongation agissent de façon à générer un état chromatinien permissif en aval de la RNAPII et plus répressif après son passage pour empêcher des initiations cryptiques. Les interactions à grande échelle (long-range), comme celles entre les éléments insulateurs, isolent les domaines chromatinien pour empêcher la propagation des signaux régulateurs.

1.1.1 La région génomique d'assemblage du complexe de transcription

Le promoteur est la région génomique comprenant l'ensemble des séquences d'ADN nécessaires à la transcription d'un gène. Il est composé du cœur-promoteur et des séquences activatrices ou inhibitrices pouvant se situer à des kilobases du site d'initiation de la transcription [199]. Le cœur-promoteur est défini comme la séquence d'ADN minimale suffisante pour permettre au complexe d'initiation (PIC) de se former et permettre une transcription basale [288]. Il s'étend de part et d'autre du site d'initiation de la transcription (TSS). Il n'existe pas de longueur standard du cœur-promoteur, néanmoins on peut considérer que la moyenne s'étend jusqu'à environ 100 paires de bases (pb) de chaque côté du TSS. Les éléments *cis* sont pour la plupart conservés entre les gènes orthologues mais il existe une telle diversité des cœur-promoteurs chez les mammifères qu'il est vain de chercher des cœur-promoteurs par bioinformatique [217]. C'est donc expérimentalement que l'on définit les promoteurs, par des méthodes basées sur le séquençage d'extrémités 5' des ADNc [269]. Ces travaux ont permis de déterminer quelles étaient les séquences initiatrices de la transcription. Il en ressort que la transcription commence préférentiellement sur un dinucléotide formé d'une pyrimidine en position -1 et d'une purine en position +1. Les dinucléotides associés aux gènes les plus actifs sont CG, TG ou CA [40]. Lors de ces expériences, on trouve une majorité d'ADNc commençant au même nucléotide et quelques-uns commençant quelques nucléotides plus loin. Ce sont les gènes possédant un promoteur « à pic unique » pour le site du début de la transcription (TSS, Transcription Start Site). La figure 1.2 A montre dans la partie du haut, le schéma des ADNc obtenus et juste en-dessous leur distribution. Ici tous commencent au même nucléotide. Pour d'autres gènes, il peut y avoir plusieurs ADNc pour un même gène commençant à des nucléotides différents dans une région qui peut faire une centaine de paires de bases. C'est ce que montre la figure 1.2 B avec différents ADNc pour un même gène et la distribution observée pour chacun de ces ADNc. Ce sont les gènes avec un promoteur « à pics dispersés » : ils n'ont pas un nucléotide préférentiel pour commencer la transcription mais plusieurs nucléotides avec une fréquence d'utilisation similaire entre eux [269]. En-dessous de ces deux schémas, des cas réels sont présentés. Deux promoteurs appartenant à la première ou à la seconde catégorie sont analysés en parallèle chez la souris et chez l'homme. Sont représentés la distribution des ADNc obtenus en fonction de leur TSS. On constate d'abord que les promoteurs à pic unique ont quelques autres

TSS, mais ceux-ci sont très minoritaires comparés au pic majeur. Alors que dans la figure 1.2 B, il y a tout de même un pic majeur mais les autres TSS représentent le tiers sinon la moitié du pic majeur. Ils ne sont donc pas minoritaires. Un autre élément important est la conservation de ces TSS entre les espèces. En effet, les patrons obtenus chez l'homme et chez la souris sont très similaires. Le choix du TSS n'est certainement pas fait au hasard, bien plus il a une telle importance qu'il se transmet au cours de l'évolution. Bien-sûr, on trouve aussi des promoteurs à structure intermédiaire entre ces deux catégories. Les promoteurs « à pics dispersés » se retrouvent souvent chez les gènes de ménage alors que les gènes tissus-spécifiques semblent être plus souvent dirigés par les promoteurs « à pic unique ».

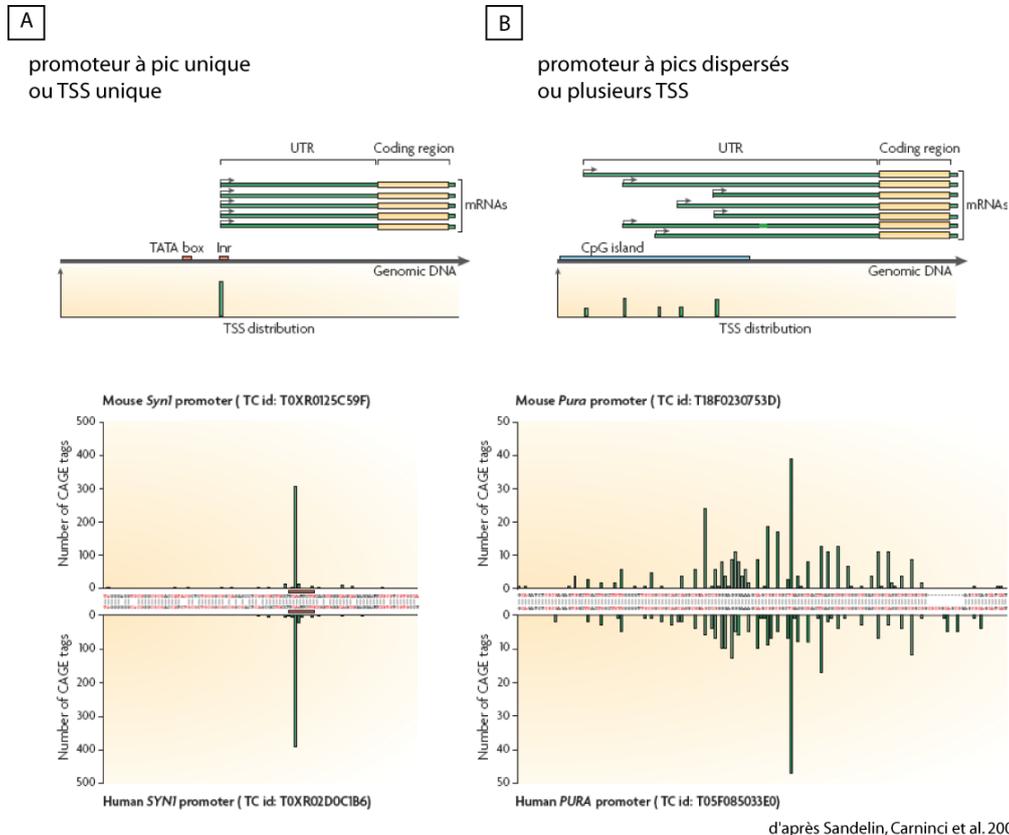
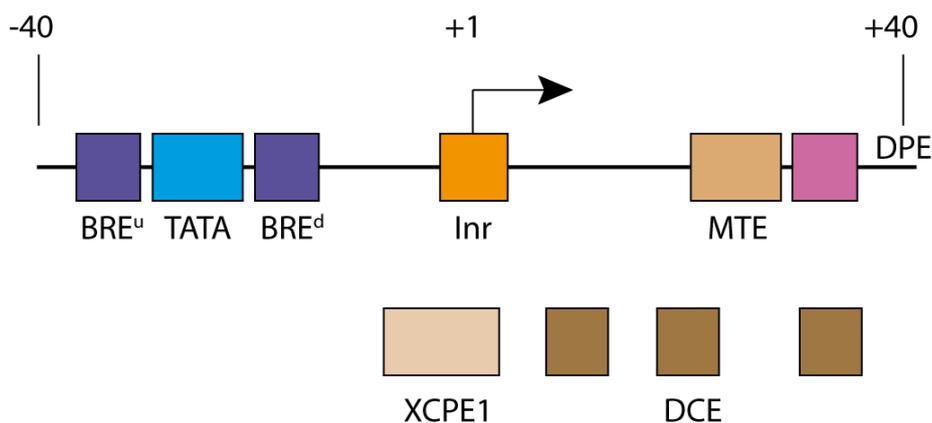


FIGURE 1.2 – Les classes de sites d’initiation chez les promoteurs de mammifères. Les promoteurs peuvent être classifiés en fonction de la distribution des sites d’initiation (TSS) qu’ils utilisent. Pour chaque classe de promoteurs, est d’abord représenté un schéma général de la classe (partie du haut) puis un exemple détaillé (partie du bas). Dans l’exemple détaillé, l’axe des ordonnées représente la distribution des TSS sur les ADNc obtenus, en haut pour la souris, en bas pour l’homme. L’axe des abscisses représente l’alignement des séquences promotrices de la souris et de l’homme. Le dinucléotide de l’initiation (Pyrimidine-Purine) est en rouge. On remarque qu’il correspond à chaque fois au pic majeur des TSS. A. Les promoteurs à pic unique utilisent un seul ou quelques nucléotides consécutifs comme TSS, ce qui revient à un seul pic majoritaire. Ces promoteurs ont souvent une boîte TATA et un Initiateur (Inr). L’exemple détaillé est celui du gène *Syn1* dans lequel presque tous les TSS sont concentrés sur quelques positions consécutives. Ce promoteur particulier a un Initiateur distinct mais pas de boîte TATA. B. Les promoteurs à pics dispersés peuvent initier la transcription sur une région d’environ 100 pb, produisant une population d’ARNm de différentes longueurs mais avec généralement la même séquence codante. Ces promoteurs sont souvent privés de boîte TATA et enrichis en îlots CpG. Le promoteur du gène *Pura* comprend plusieurs TSS qui sont répartis sur un vaste espace génomique. La plupart des TSS importants dans ce promoteur ont un dinucléotide Pyrimidine-Purine, mais pas d’Initiateur bien défini. On peut noter la correspondance entre l’homme et la souris dans l’utilisation des TSS.

En plus des divergences sur le nombre de TSS utilisés, les cœur-promoteurs se différencient par les combinaisons de séquences qui les composent (Juven-Gershon, Hsu et al. 2008). Les plus connues de ces séquences sont représentées dans la figure 1.3. Boîte TATA, BRE, Inr, DPE et DCE sont des séquences souvent associées au promoteur à « à pic unique » (figure 1.3). La boîte TATA est une courte séquence (TATAWAAR) qui recrute la TBP (TATA Binding Protein) du complexe TFIID (facteur général de la transcription) et ce faisant, l'ensemble des facteurs généraux [36, 117]. Cette boîte, que l'on croyait partie intégrante de tout promoteur, est en fait présente dans moins de 10% des cœur-promoteurs humains [84, 13]. BRE (TFIIB recognition element) est la séquence de fixation du facteur général de la transcription TFIIB. La séquence du BRE^u (upstream) est différente de la séquence du BRE^d (downstream). En fonction du contexte du promoteur, le BRE peut jouer un rôle positif ou négatif. L'initiateur (Inr) recouvre le TSS. C'est la séquence la plus communément retrouvée dans les cœur-promoteurs. Son consensus varie entre les espèces. Chez l'humain, c'est YYANWYY où l'adénine est souvent le +1. DPE (Downstream core Promoter Element) et MTE (Motif Ten Element) coopèrent avec l'Inr. La distance qui les sépare de l'Inr est critique pour une transcription optimale. Le DCE (Downstream Core Element) est composé de trois éléments (figure 1.3) et est souvent associé à la boîte TATA. Le XCPE1 (X Core Promoter Element 1) agirait en synergie avec les séquences liant des activateurs dans les îlots CpG. Les îlots CpG sont des régions génomiques dans lesquelles les dinucléotides CG sont surreprésentés. Une étude récente montre que plus des 3/4 des promoteurs des gènes de ménage contiennent des îlots CpG et ne possèdent pas de boîte TATA. Bien que traditionnellement associés à une boîte TATA, les promoteurs de gènes tissus-spécifiques sont principalement sans îlots CpG et sans boîte TATA (60%). Les autres sont à part égale CpG+/TATA- and CpG-/TATA+. Une minorité de gènes comportent des îlots CpG et une boîte TATA (4%) [356]. Chez les métazoaires, lorsqu'il y a une boîte TATA, la transcription commence 25-30 pb en aval. Sur des promoteurs sans boîte TATA, la transcription commence à une distance fixe d'autres éléments tels que le DPE [178]. Chez la levure, il a été démontré que la RNA-Pol II scanne l'ADN depuis le promoteur où elle s'est assemblée pour trouver le site +1 [162].



d'après Juven-Gershon, Hsu et al. 2008

FIGURE 1.3 – Les motifs des cœur-promoteurs. Ce schéma montre quelques-uns des éléments du cœur-promoteur pour la transcription par la RNAPII. Il n'y a pas d'éléments universels du cœur-promoteur. Chacun de ces éléments est retrouvé dans seulement une fraction (généralement estimée entre 1 et 30% en fonction du motif) de l'ensemble des cœur-promoteurs. L'Initiateur (Inr) est probablement le motif le plus récurrent. La boîte TATA, l'Inr, le MTE, le DPE et le DCE sont des sites de reconnaissance pour la liaison du facteur général de transcription TFIID. On peut noter qu'il existe de multiples formes de TFIID et de complexes de protéines reliées à TFIID qui peuvent potentiellement interagir avec le cœur-promoteur. BRE^u et BRE^d interagissent avec TFIIB. XCPE1 est retrouvé dans 1% des cœur-promoteurs, généralement sans boîte TATA. Il pourrait lier la TBP seule ou l'ensemble du complexe TFIID.

1.1.2 Le complexe de pré-initiation et les facteurs généraux de la transcription

Les séquences du cœur-promoteur sont reconnues par les facteurs généraux de la transcription. Ceux-ci s'assemblent sur le cœur-promoteur pour former le complexe de pré-initiation (PIC). Six complexes sont appelés « facteurs généraux de la transcription » chez les eucaryotes : TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIH. Ils remplacent, chez la bactérie, la sous-unité σ , partie intégrante de la RNAP bactérienne. Un complexe de pré-initiation peut être assemblé étape par étape sur des promoteurs *in vitro* [1]. L'ordre déterminé par ces expériences est le suivant : TFIID ou la TBP se lient aux éléments du promoteur, suivis de TFIIA, TFIIB, un sous-complexe de la RNAPII et TFIIIF, TFIIE en enfin TFIIH. *In vivo*, les événements ne semblent pas suivre cet ordre à la lettre. Le tableau 1.1 présente chacun de ces complexes dans leur ordre d'apparition avec leur composition, leurs fonctions et leurs liaisons [177, 117]. On retiendra que beaucoup d'entre eux se lient à l'ADN via des séquences spécifiques, qu'ils sont responsables de l'enroulement de l'ADN pour permettre

sa transcription (TFIID, TFIIA), qu'ils dirigent le promoteur dans le site actif de la RNAPII (TFIIE, TFIIH), ouvrent la double hélice (TFIIH), aident la RNAPII à trouver le site +1 (TFIIB) et qu'ils modifient la RNAPII pour l'activer (TFIIE, TFIIH).

composants	sous-unités	sous-unités essentielles	masse (kDa)	fonction	sous-unités satellites ou annexes	liaison particulière	liaison à l'ADN	remarque
TFIID (TBP et TAFs) [303]	14	TBP TAF4 et TAF6 (TAF5, TAF9, TAF12) [333]	27	enrouler la boîte TATA autour de TFIIB et de la RNAPII (devenir) [225, 164]	TLF (TBP-like transcription factors) remplacent TBP chez les métazoaires [60]	RNAPII (CTD non phosphorylé) [312]	TATA	Les TLF jouent un rôle dans la régulation de la transcription au cours du développement embryonnaire [246, 17]
		TAF4 et TAF6 (TAF5, TAF9, TAF12) [333]	749	stabilisation du complexe	Spt-Ada-Gm5-acétyltransférase (SAGA) chez la levure et le TBP-free TAF-containing complexe (TFIC) chez l'homme [110, 233, 330]		DPE [36, 279]	Transcription sans boîte TATA. Grande diversité dans les combinaisons de facteurs à relier avec la grande diversité des ceure-promoteurs [217].
TFIIA	3		69	stabilisation de l'interaction TBP-ADN [123]				Il n'est pas requis pour tous les gènes [177].
TFIIB	1	polypeptide de 316 acides aminés [256]	38	Détermination du site +1.		TBP, la RNAPII, les sous-unités de TFIIF (RAP30 and RAP74) et aussi TAF9 [103, 115, 82]	BRE [171]	Il se fixe de façon asymétrique, orientant la transcription [307].
TFIIF	4	deux grosses (TFIIF α /RAP74) et deux petites (TFIIF β /RAP30) [3]	156	recrutement de TFIIE et TFIIF dans le PIC empêche liaisons non spécifiques à l'ADN stabilise le PIC [177]		RNAPII via la sous-unité Rpb2 [50]	de part et d'autre de la boîte TATA	important pour l'enroulement de l'ADN et rôle dans la fuite du promoteur [263, 56]
TFIIE	4	deux grosses (TFIIE α) et deux petites (TFIIE β) [234]	92	recrutement de TFIIH en complant les interactions RNAPII-Promoteur. Rôle dans la phosphorylation de la RNAPII [236].		le domaine pince de la sous-unité Rpb1 de la RNAPII p62 de TFIIH [50]	dans la bulle de transcription et en dehors [148]	TFIIF et TFIIE sont diamétralement opposés autour de la crevasse centrale de la RNAPII. TFIIE et TFIIH sont indispensables à l'ouverture ATP-dépendante du double-brin d'ADN, pour former le complexe «promoteur ouvert».
TFIIH	10	coeur (XPB, p34, p44, p52, p62 et p8/TTDA) et partie kinase cycline-dépendante (CDK7, cyclin H et MAT1) reliés entre eux par XPD [275, 98]	sous-unités entre 10 et 90kDa	Ouverture du promoteur (activité hélicase), phosphorylation de la RNAPII [191]			promoteur	réparation de l'ADN (NER) phosphorylation de TFIID (TBP), TFIIE α and TFIIF α (RAP74), contrôle du cycle cellulaire [72, 83]
Médiateur	20		>1000	transduction de signaux régulateurs (coactivateurs ou co-répresseurs) [297] rôle dans la stabilisation du PIC (aide à recruter TFIIH) [132, 12], dans la stimulation de la phosphorylation de la RNAPII [78], dans la fuite du promoteur		RNAPII [202] Rad3 (TFIIH) Facteurs spécifiques de transcription (cycle cellulaire, développement, stress, etc.)	non	complexes homologues chez l'homme : ARC/DRIP, TRAP/SMCC, NAT, CRSP, and PC2 [43, 42]

TABLE 1.1 – Présentation des facteurs généraux de la transcription participant au complexe de pré-initiation et à l'activation de la RNAPII.

1.1.3 La transcription par la RNAPII

Une fois recrutée, l'ARN polymérase II doit commencer la transcription. Avant de rentrer dans le détail du mécanisme, voici une description de cette enzyme.

L'ARN polymérase II

Chez les eucaryotes, l'ARN polymérase II est un complexe de 10 à 12 sous-unités appelées Rpb1, Rpb2... , Rpb12. L'ARN polymérase II est l'homologue de l'ARN polymérase chez les procaryotes [61]. Beaucoup de régions sont conservées entre les RNAP des archéobactéries, des bactéries et des eucaryotes comme le montre la figure 1.4, notamment les sous-unités Rpb1 et Rpb2 qui forment la « pince de crabe » au centre de laquelle se trouve le site actif (figure 1.4 A). En superposant les structures cristallographiques, on remarque que la zone du centre actif est la plus conservée entre les espèces (en vert sur la figure 1.4 B). Parmi les divergences au fil de l'évolution, la plus importante est sans doute l'absence de domaine carboxy-terminal chez les procaryotes [101]. Il existe aussi une conservation entre les trois RNAP eucaryotes (tableau 1.2). Chacune transcrit une classe d'ARN : l'ARN polymérase I synthétise les ARNs de classe I, à savoir de longs ARN ribosomiques (ARN 18S ; 28S ; 5,8S) ; l'ARN polymérase II transcrit les ARNs de classe II : les ARNs messagers et les ARNs non-codant de l'épissage ; l'ARN polymérase III : les ARN de transfert et certains ARN ribosomiques (ARN 5S) ainsi que d'autres ARNs non codants et le snARN U6. Comme le montre le tableau 1.2, cinq sous-unités sont communes entre ces trois RNAP : les sous-unités Rpb5, Rpb6, Rpb8, Rpb10 et Rpb12 [57].

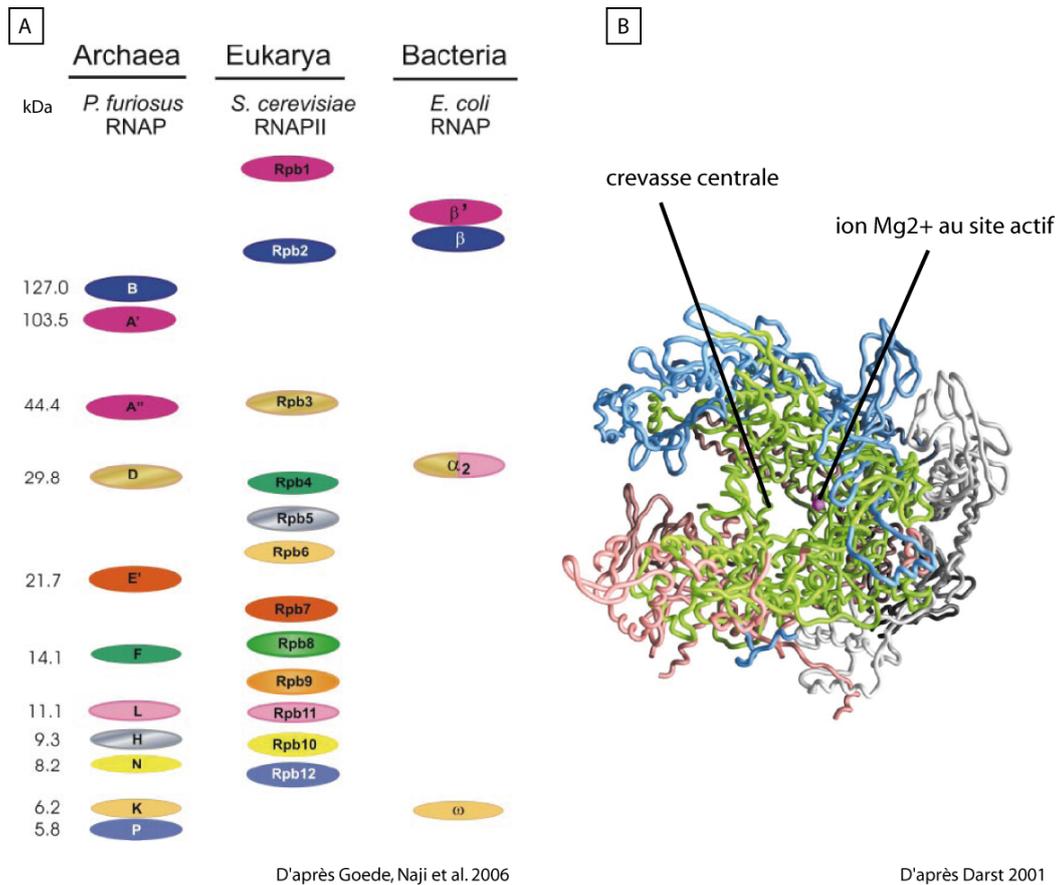


FIGURE 1.4 – **Structure de l'ARN polymérase à travers les trois domaines du vivant.** A. Sous-unités homologues de la RNAP entre les bactéries, les archéobactéries et les eucaryotes. Les couleurs signifient l'homologie. B. Structure de la RNAP bactérienne. En rose, la sous-unité β' , en bleu la sous-unité β . En vert, les zones d'homologie sur ces deux sous-unités avec la RNAPII eucaryote. On note que l'homologie de structure concerne essentiellement le site catalytique.

Parmi les 12 sous-unités de la RNAPII, l'hétérodimère Rpb4/Rpb7 est nécessaire à l'initiation mais associé au cœur de façon moins stable que les autres sous-unités. Les sous-unités principales Rpb1 et Rpb2 portent l'activité catalytique et ont des poids moléculaires de 220 et 140kDa respectivement. La sous-unité Rpb1 possède un domaine carboxy-terminal, appelé CTD, essentiel à sa fonction de transcription et jouant le rôle de plate-forme pour d'autres facteurs de transcription et de maturation des ARNm [58, 249]. Le CTD de la RNAPII est composé de motifs répétés ($Y_1S_2P_3T_4S_5P_6S_7$) 52 fois chez l'homme. Au cours de la transcription, ce motif est phosphorylé, principalement en Ser₅ et Ser₂. Ces phosphorylations vont permettre la transcription par la RNAPII. Bien que non systématiques, les phosphorylations vont avoir lieu sur tous les motifs répétés du

RNA polymerase	Pol I	Pol II	Pol III
Ten-subunit core	A190	Rpb1	C160
	A135	Rpb2	C128
	AC40	Rpb3	AC40
	AC19	Rpb11	AC19
	A12.2	Rpb9	C11
	Rpb5 (ABC27)	Rpb5	Rpb5
	Rpb6 (ABC23)	Rpb6	Rpb6
	Rpb8 (ABC14.5)	Rpb8	Rpb8
	Rpb10 (ABC10 α)	Rpb10	Rpb10
	Rpb12 (ABC10 β)	Rpb12	Rpb12
Rpb4/7 subcomplex	A14	Rpb4	C17
	A43	Rpb7	C25
TFIIF-like subcomplex ^a	A49	(Tfg1/Rap74)	C37
	A34.5	(Tfg2/Rap30)	C53
Pol III-specific subcomplex	–	–	C82
	–	–	C34
	–	–	C31
Number of subunits	14	12	17

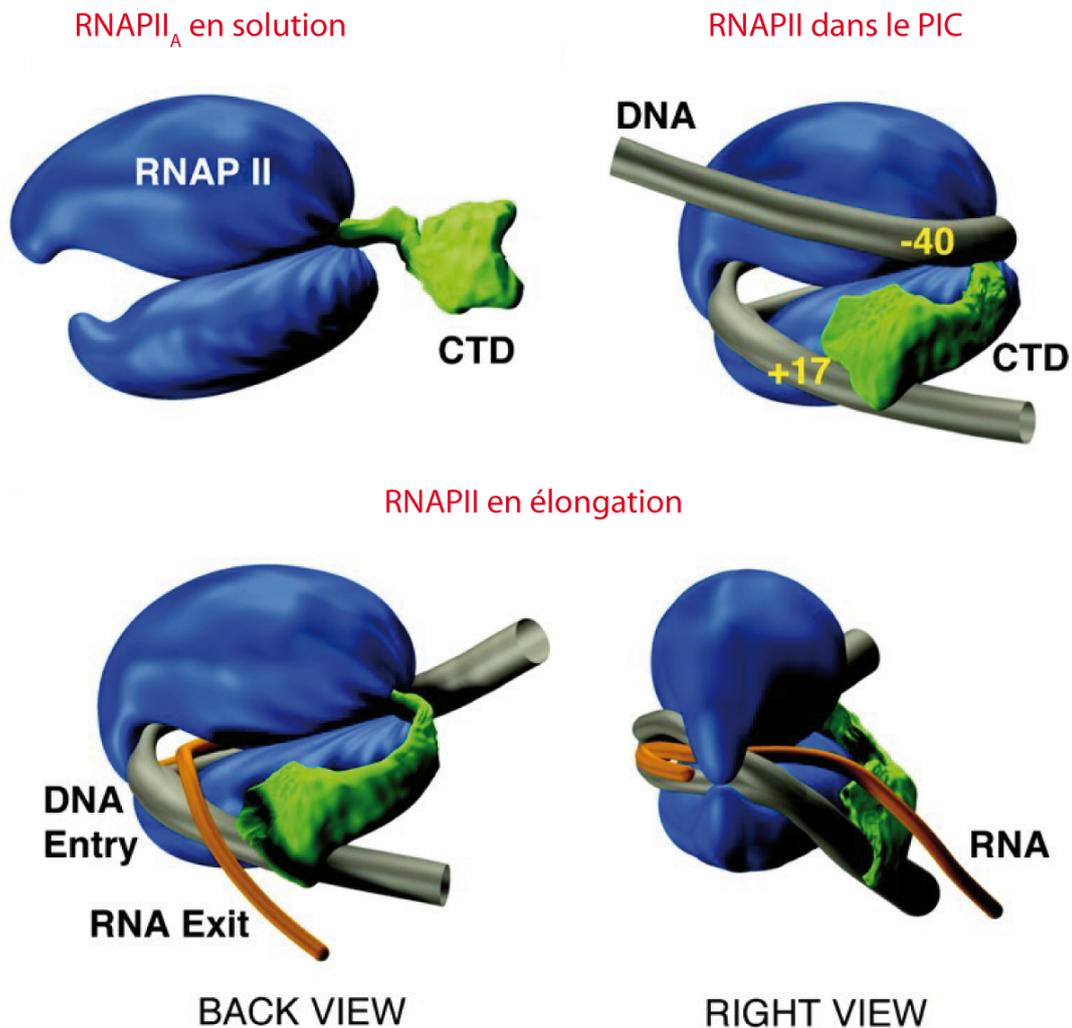
TABLE 1.2 – **Composition des 3 RNAP eucaryotes.** Les sous-unités homologues sont mises les unes en face des autres. ^a Les deux sous-unités dans la Pol I et la Pol III sont prédites pour former des hétérodimères qui ressemblent en partie au facteur d'initiation et d'élongation TFIIF, qui est composé des sous-unités Tfg1, Tfg2 et Tfg3 chez la levure et des sous-unités Rap74 et Rap30 chez l'homme.

CTD, conduisant ainsi à une modification du poids moléculaire de la RNAPII et donc à une migration différente sur gel. On appelle la forme hyperphosphorylée ou RNAPII_O, la RNAPII phosphorylée sur toutes ou presque toutes ces répétitions. La forme dite hypophosphorylée ou RNAPII_A, correspond à la RNAPII peu ou pas phosphorylée sur son CTD. Le poids de Rpb1 varie avec les phosphorylations entre 220 et 240 kDa. Entre ces deux formes, il existe bien-sûr tout un éventail de combinaisons de phosphorylation et donc de poids moléculaires. J'en parlerai plus en détail dans la partie portant sur l'épissage cotranscriptionnel.

Structure moléculaire de la transcription

La transcription est une succession de réarrangements pour que la RNAPII puisse se déplacer le long de l'ADN et transcrire l'ARN. La figure 1.5 montre l'état de la RNAPII_A libre. On voit qu'une fois assemblé dans le PIC, l'ADN est enroulé autour d'elle grâce à l'action des facteurs généraux de

la transcription présents dans le PIC. Il faut ensuite un changement conformationnel pour l'activer et initier la transcription. Ce changement consiste en l'ouverture du promoteur. Pour comprendre comment cela se passe, il faut décrire la structure des éléments en place. La structure cristallographique de la RNAPII est représentée dans la figure 1.6. Les deux plus grosses sous-unités Rpb1 et Rpb2 forment une pince de crabe à laquelle les autres sous-unités participent (mâchoire supérieure (essentiellement Rpb2) et mâchoire inférieure (essentiellement Rpb1)). Le site actif de l'enzyme est formé par Rpb1 et comporte deux ions Mg^{2+} visibles par deux billes bleues sur le schéma A [58].



d'après Douziech, Forget et al. 1999

FIGURE 1.5 – **Les différents états de la RNAPII.** Représentation en 3 dimensions de la RNAPII libre (RNAPII_A), assemblée dans le PIC et en élongation. En bleu la RNAPII, en vert son CTD, en gris l'ADN, en orange l'ARN naissant. Les facteurs généraux d'initiation de la transcription ne sont pas représentés pour simplifier le modèle. Sur la RNAPII en élongation, on remarque la proximité du CTD avec le canal de sortie de l'ARN.

Entre 11 et 15 pb autour du site +1, l'ADN se fond dans la RNAPII en empruntant le sillon formé par Rpb1 et le brin matrice rentre dans la crevasse qui forme le site actif (schéma B; site actif en orange sur le schéma A). L'ADN est ouvert par l'activité hélicase de TFIIH et forme une bulle de transcription de 11 nucléotides. En prenant le site d'initiation comme +1, la bulle se forme de -9 à +2 [174]. *In vitro*, en utilisant des radicaux hydroxides sur un promoteur en présence d'un PIC de levure, il a été démontré que l'ADN suit une ligne droite dans la crevasse centrale de la RNAPII [207]. Il interagit avec la sous-unité rad25 de TFIIH immédiatement après la bulle de transcription et s'étend vers la sous-unité Rpb3 de la RNAPII en amont de la boîte TATA. La sous-unité XPB (Rad25 chez la levure) de TFIIH possède une activité hélicase. Elle induit une torsion ATP-dépendante du site de liaison pour générer un ADN simple brin qui peut descendre dans la crevasse de la RNAPII [148]. XPB/Rad25 pourrait continuer à enfiler l'ADN simple brin dans la crevasse, permettant ainsi à la RNAPII de scanner la matrice ADN pour trouver des sites +1 acceptables. Etant donné que les TSS sont parfois à plus de 100 pb du lieu de fixation du PIC, ce modèle prédit la formation de boucles d'ADN simple brin émergeant de la crevasse centrale alors que la RNAPII reste ancrée sur le promoteur [118]. L'ouverture du double-brin d'ADN est ATP-dépendante et requiert TFIIE et TFIIH [148].

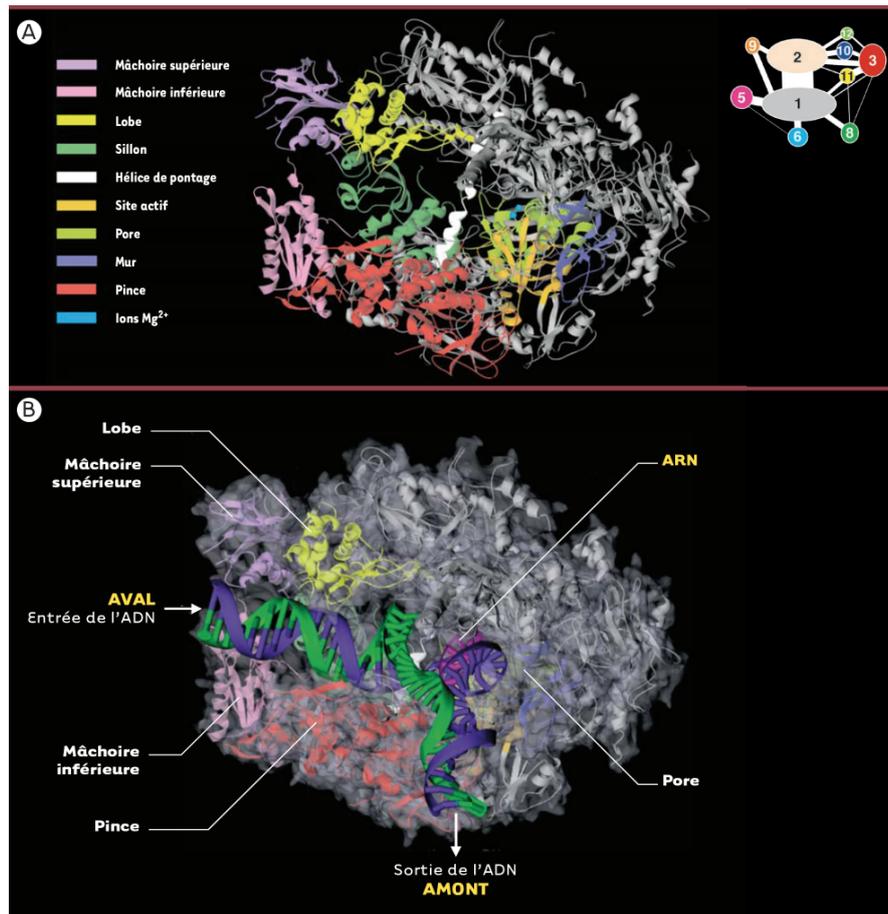


FIGURE 1.6 – **Structure cristallographique de la RNAPII.** **A.** Structure de la RNAPII avec en couleur les différentes parties du complexe. En haut, à droite, une représentation schématique des sous-unités dans cette structure. **B.** Ouverture de l'ADN dans le site actif de la RNAPII et synthèse du brin d'ARN qui fait contact avec le mur formé par Rpb2.

L'initiation commence avec la synthèse de la première liaison phospho-diester entre les deux premiers nucléotides [117, 57]. Au fur et à mesure que la transcription avance, le +1 est considéré comme le dernier résidu ajouté au pré-ARNm et la bulle se déplace le long du gène. Un élément de structure appartenant à Rpb1 et appelé hélice de pontage jouerait un rôle important dans la translocation de l'enzyme (schéma A). Après l'ajout du quatrième rNTP, il se produit un changement de conformation qui stabilise le complexe en conformation ouverte ou pinces « ouvertes ». Dès lors, TFIIF n'est plus nécessaire au maintien de la bulle de transcription [144]. A l'intérieur de la bulle, l'appariement des rNTPs produit un hybride ARN-ADN de 8 à 9 nucléotides calé entre l'hélice de pontage (Rpb1) et le mur (Rpb2) (schéma B, mur en bleu foncé sur le schéma A), in-

teragissant avec une quinzaine de régions protéiques appartenant à Rpb1 et Rpb2 [100]. Ces interactions créent une zone de haute complémentarité qui ne doit cependant pas être un frein au déplacement de la RNAPII. Entre 3 et 10 nucléotides polymérisés, la transcription est dite abortive car la RNAPII laisse échapper le pré-ARNm et recommence la transcription au site d'initiation. La RNAPII n'avancerait pas sur l'ADN mais le ferait se compacter à l'intérieur de la crevasse. Cette structure n'étant évidemment pas stable, la transcription ne peut se faire que sur quelques nucléotides. Pour changer d'état la RNAPII a deux choix possibles : libérer l'ARN pour libérer l'ADN de la crevasse ou maintenir l'ARN mais changer de conformation en cassant ces liaisons avec le promoteur. La dernière solution mène à ce qu'on appelle la fuite du promoteur et à l'élongation de la transcription [138, 258]. On ne sait pas bien quels sont les causes de ce phénomène et comment la RNAPII « choisit » entre ces deux solutions. Néanmoins, TFIIB y joue certainement un rôle car il possède un doigt B inséré dans la crevasse par le canal de sortie de l'ARN. Il doit être délogé pour permettre à l'ARN de sortir [29].

Après la synthèse d'une trentaine de bases, la RNAPII est sensée relâcher ses contacts avec le cœur-promoteur et le reste de la machinerie transcriptionnelle c'est ce qu'on appelle la « fuite du promoteur ». Pour permettre cette fuite, la RNAPII est phosphorylée sur les Ser₅ de son CTD par l'activité kinase de la sous-unité cdk7 de TFIIH [191, 320].

Facteurs négatifs et positifs de transcription, fin de la transcription

La RNAPII a donc initié la transcription, elle a quitté le promoteur. Mais d'autres facteurs interviennent pour stopper la transcription. Ce sont des facteurs négatifs de transcription qui la mettent en pause transcriptionnelle. La RNAPII est, à ce stade phosphorylée en Ser₅. Elle a initié la transcription et est stoppée à 20–40 nt en aval du TSS. Cette pause est en partie due aux effets négatifs du facteur négatif d'élongation NELF et du complexe sensible au DRB (5,6-dichloro-1- β -D-ribofuranosylbenzimidazole ; inhibiteur de la transcription) DSIF/Spt4,5 [54]. NELF comprend 4 sous-unités NELF-A, B, C/D et E. Il est conservé chez les mammifères mais est absent chez *Caenorhabditis elegans*, *Saccharomyces cerevisiae* et *Arabidopsis thaliana* [220]. NELF se lie à DSIF. DSIF est composé des facteurs d'élongation Spt4 et Spt5, conservés de la levure à l'homme. DSIF s'associe à la RNAPII. *In vivo*, DSIF et NELF sont présents sur les gènes de choc thermique non induits de

la drosophile et ces facteurs maintiendraient la pause transcriptionnelle dans la région proche du promoteur du gène Hsp70 puisqu'une immunodéplétion de NELF abolit la pause [6, 334]. Ils pourraient interagir avec l'ARN car la sous-unité NELF-E se lie à l'ARN [92]. Cette pause est importante car elle permet la pose de la coiffe sur l'ARN naissant. Pour sortir de la pause et continuer l'élongation de la transcription, le P-TEFb (Positive Transcription Elongation Factor b) phosphoryle le CTD de la RNAPII sur les Ser_2 et la sous-unité SPT5 de DSIF et NELF-E. Le P-TEFb est composé de 2 sous-unités : cycline T et sa kinase associée cdk9 [247]. DSIF phosphorylé, NELF perd son interaction et se décroche du complexe. DSIF devient un facteur positif d'élongation. Le PTEF-b est recruté sur les gènes transcrits par la RNAPII par divers mécanismes pour la transcription d'une majorité de gènes, incluant les gènes de choc thermique et le gène c-Myc, et il est aussi requis pour la transcription du VIH en étant recruté par la protéine Tat [252, 354]. Même si le P-TFb reste associé avec le complexe d'élongation, son activité kinase n'est plus requise une fois que le complexe est sorti de la pause [224].

La transcription se termine par la libération de l'ARN et de l'ADN [253]. La libération de l'ARN a lieu avec la coupure après le signal polyA et l'ARNm est polyadénylé. La RNAPII se détache de l'ADN grâce à la présence de séquences terminatrices. Une réinitiation de la transcription plus rapide est possible par le fait que certains facteurs restent associés au promoteur [344].

1.2 Epissage

L'ARNm transcrit n'est pas encore mature. Il subit plusieurs modifications (coiffage, épissage, coupure et polyadénylation) qui lui permettront d'être ensuite véhiculé et traduit. Parmi ces modifications, l'épissage est un événement indispensable de la maturation du pré-ARNm multi-exonique. Celui-ci ne pourra être ni exporté ni traduit s'il n'est pas épissé.

L'ARN pré-messager est la copie parfaite du gène et contient de façon alternée exons (parties codantes) et introns (parties non codantes). L'épissage consiste en l'excision des introns et en la jonction des exons les uns à la suite des autres. L'épissage possède une autre propriété : celle de créer une grande quantité, parfois phénoménale, de transcrits à partir d'un seul et même gène. Par exemple, le gène Dscam (Down syndrome cell-adhesion molecule) pourrait générer 38 016

variants d'épissage [273] et tout autant de protéines différentes. Cela est possible car il existe diverses manières de faire de l'épissage alternatif. Les événements de base sont représentés dans la figure 1.7. Et pour un même gène multi-exonique, ces différents modes peuvent se combiner (figure 1.7 H). L'épissage alternatif est loin d'être un cas exceptionnel mais plutôt une règle. En effet, 92% des gènes multi-exons peuvent donner au minimum deux transcrits distincts par l'épissage alternatif. En ne comptant que les gènes dont chaque isoforme représente au moins 15% dans un tissu, il reste encore 86% de gènes subissant l'épissage alternatif [321]. La plupart des épissages alternatifs affectent la séquence codante : la moitié altère la phase ouverte de lecture et le tiers conduit à la dégradation de l'ARNm par le système NMD (nonsense mediated decay) [200]. Le système NMD est un système de surveillance des ARNm qui détecte les ARNm possédant un codon-stop prématuré. Lorsqu'il est dérégulé, l'épissage alternatif peut être la source de maladies, notamment de cancers [322]. La dérégulation peut provenir de mutations en *cis* (perturbation du code) ou en *trans* (perturbation de la machinerie d'épissage). Dans une étude récente, il a été proposé que 60% des mutations géniques causaient des maladies par un dérèglement de l'épissage (mutations en *cis*) [190]. Les mutations en *trans* provoquent des surexpressions ou des sous-expressions de facteurs d'épissage qui ont pour effet de modifier les quantités relatives d'isoformes d'un ARNm. Ainsi, le facteur d'épissage SF2/ASF a été trouvé surexprimé dans plusieurs tumeurs humaines, affectant l'épissage alternatif de plusieurs gènes dont le suppresseur de tumeur BIN1, ce qui le conduit à la perte son activité suppresseur de tumeur. Les kinases MNK2 et S6K1 sont aussi affectées et l'isoforme résultante de MNK2 stimule la phosphorylation de eIF4E (protéine associée à la coiffe qui permet l'initiation de la traduction) indépendamment de la voie MAPK. L'isoforme 2 de S6K1 est surexprimée se met à arborer des propriétés oncogéniques alors que cette protéine est impliqué dans la croissance cellulaire et dans l'apoptose [140]. A l'inverse, le facteur d'épissage U2AF35 est sous-exprimé dans d'autres types de cancer [112].

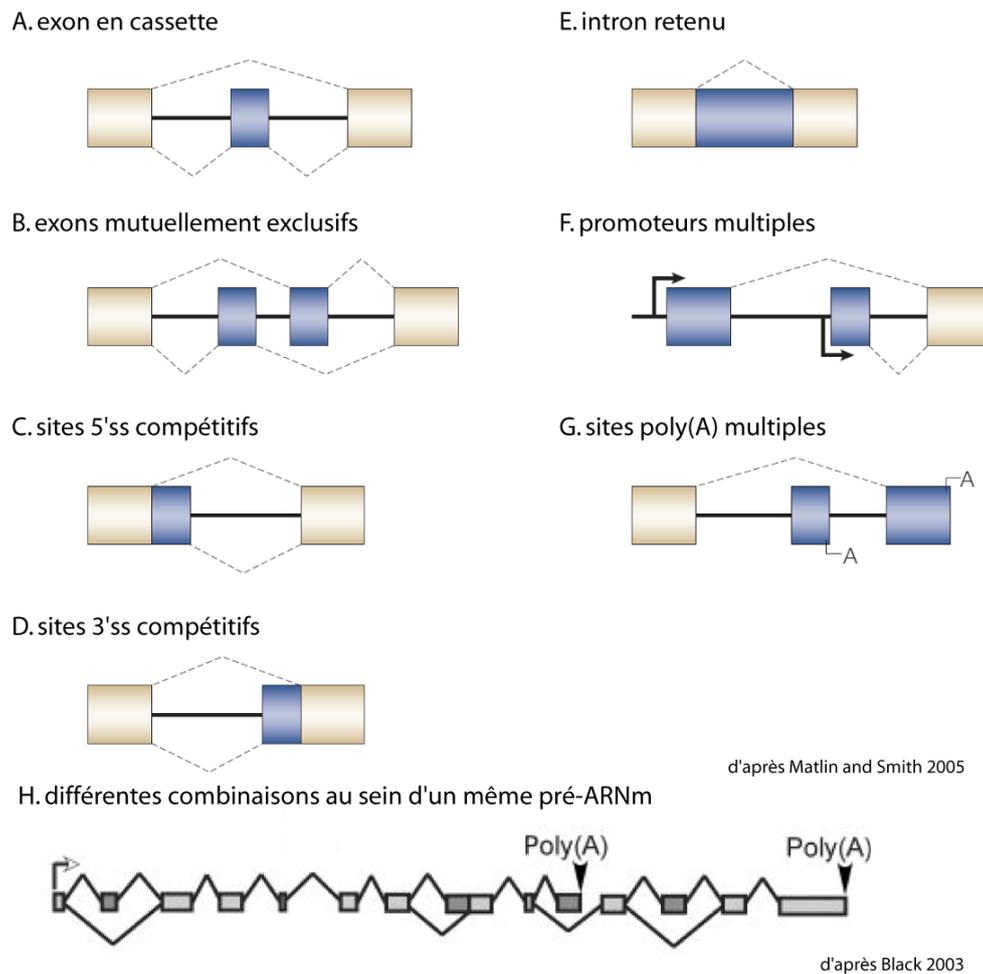


FIGURE 1.7 – **Mécanismes élémentaires de l'épissage alternatif.** **A.** Les exons en cassette sont des exons discrets qui peuvent être indépendamment inclus ou exclus de l'ARNm. Ils peuvent être ensuite sous-divisés dans les catégories "exon sauté" ou "exon cryptique" selon le type majeur de variants observés (variant excluant ou incluant l'exon). **B.** Les exons mutuellement exclusifs impliquent la sélection d'un seul exon parmi plusieurs exons variants. **C** et **D.** Les sites 5' et 3' ss compétitifs représentent les modifications d'exons. **E.** Un intron retenu peut être excisé du pré-ARNm ou bien retenu dans l'ARNm mature. **F** et **G.** Les promoteurs et les sites poly(A) alternatifs changent le 5' ou le 3'ss des exons dans le transcrit. **H.** Un seul pré-ARNm peut avoir de multiples sites d'épissage alternatif en utilisant différents patrons d'inclusion. Ceux-ci sont souvent utilisés en combinaison pour produire plusieurs ARNm finaux différents.

1.2.1 Le mécanisme de l'épissage

L'épissage est réalisé par le spliceosome. C'est un complexe gigantesque estimé à 1 mégaDalton, composé de plus de 300 protéines dont 100 protéines formant le cœur [355, 134, 52]. Il est essentiellement composé de 5 complexes ribonucléoprotéiques, les UsnRNPs (Uridine-rich small nuclear ribonucleoprotein particles) [325]. Ces complexes comprennent chacun un petit ARN nucléaire (small nuclear RNAs ou snARNs) : U1, U2, U4, U5, U6. L'épissage repose sur 2 transestérifications consécutives impliquant 3 sites dans l'intron : un site donneur en 5' de l'intron (ou 5' splicing site, 5'ss), un site accepteur en 3' de l'intron (ou 3' splicing site, 3'ss) et un point de branchement (BP) au milieu de l'intron, tous trois représentés dans la figure 1.11 avec les séquences consensus (figure 1.11 A). La première réaction consiste en l'attaque nucléophile du BP sur le 5'ss, ce qui conduit l'intron à former un lasso et au clivage de l'exon en 5' (figure 1.11 B). Dans la seconde réaction, l'exon en 5' attaque le 3'ss, produisant ainsi l'ARNm épissé d'une part et le lasso intronique d'autre part. Pour permettre ces réactions les 5 UsnRNPs vont se réarranger en plusieurs complexes schématisés sur la figure 1.8 A qui vont être détaillés dans la suite de ce manuscrit.

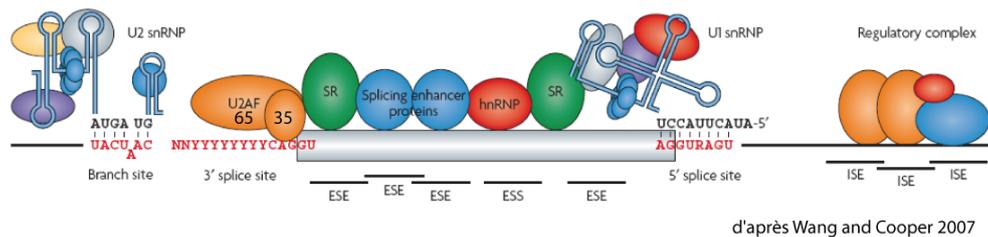
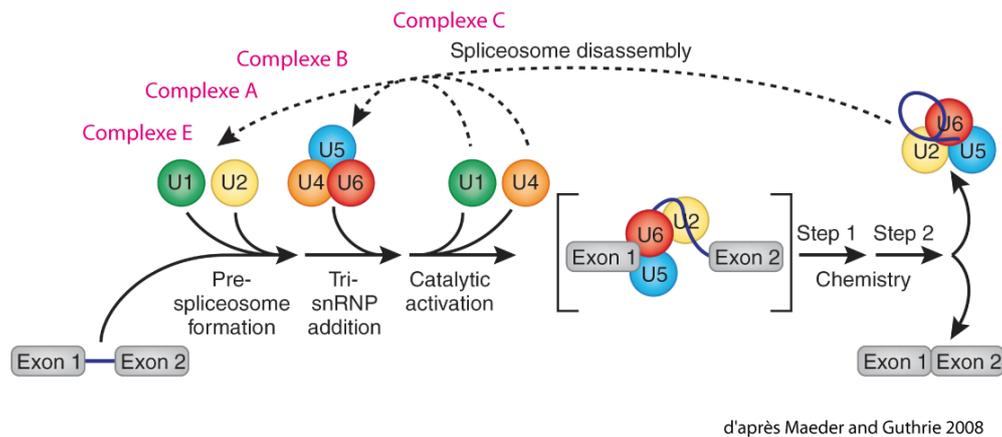


FIGURE 1.8 – **Le cycle du splicéosome et le complexe E.** **A.** Le modèle le plus connu de l'épissage est celui d'un assemblage en plusieurs étapes avec une succession de complexes : complexes E (ATP-indépendant), A, B et C (tous trois ATP-dépendants). Le complexe C aboutit à la conformation qui va permettre les réactions catalytiques, moyennant des réarrangements entre les deux réactions. A la fin du cycle, le splicéosome se dissocie pour être recyclé et recommencer un nouvel épissage. **B.** Schéma du pré-ARNm dans le complexe E avec le point de branchement, les sites 3'ss et 5'ss. ESE : Exonic Splicing Enhancer, ESS : Exonic Splicing Silencer, ISE : Intronic Splicing Enhancer, ISS : Intronic Splicing Silencer. La U1 snRNP se fixe sur le site 5'ss, la U2snRNP reconnaît le point de branchement. U2AF65 reconnaît la séquence polypyrimidine et U2AF35 le dinucléotide AG du 3'ss. Les séquences régulatrices de l'épissage sont liées par des protéines hnRNP (ESS, ISS) ou SR (ESE, ISE). On note que ce schéma correspond à la représentation de la définition de l'exon.

1.2.2 Vue d'ensemble

Le schéma B de la figure 1.8 représente les premières étapes de l'épissage. La première étape est réalisée par la snRNP U1 qui reconnaît le 5'ss (site donneur) [293]. Dans le même temps, la protéine SF1/BBP (Branchpoint Binding Protein) interagit avec le point de branchement (BP) (non représenté sur le schéma) et le complexe U2AF (U2snRNP auxiliary factor) établit des contacts avec la séquence polypyrimidine (polypyrimidine tract dans le schéma) (par sa sous-unité U2AF65) et le dinucléotide AG du 3'ss (par sa sous-unité U2AF35) [346, 335]. La snRNP U2 va s'immiscer dans ces complexes en faisant d'abord une interaction avec U2AF65. Cet ensemble forme le complexe E [63]. La snRNP U2 va déplacer SF1 et permettre l'hybridation du snARN U2 avec le BP. Cette étape nécessite l'hydrolyse d'ATP. On obtient ainsi le complexe A [27, 18]. L'association ultérieure du U4/U6•U5 trisnRNP va conduire au complexe B [70]. Le complexe catalytique C est obtenu par des réarrangements, notamment des modifications des interactions ARN/ARN [293, 135]. Beaucoup de protéines sont impliquées dans le splicéosome. Je n'en présente ici que quelques unes, les plus importantes et étudiées pour le moment.

1.2.3 Les complexes du splicéosome

Le complexe E

Les UsnRNP sont des complexes ribonucléoprotéiques composés d'un UsnARN (uridine-rich small nuclear ARN), d'un cœur de 7 protéines Sm et de protéines spécifiques de chaque UsnRNP. La snRNP U1 se lie au 5'ss. Cette interaction est le fruit de la liaison de la protéine U1C de la U1snRNP avec le 5'ss, qui laisse ensuite sa place au snARN U1 (figure 1.8 B) [357, 277, 266]. In vitro, ce recrutement semble stabilisé par la protéine TIA-1 (T-cell intracellular antigen-1) dont l'homologue Nam8p chez la levure fait partie intégrante de la U1snRNP [90, 134]. Au 5'ss, TIA-1 interagit avec U1C. La snRNP U1 est aussi plus facilement recrutée sur des sites d'épissage « faibles » grâce à U2AF65 [89]. Enfin, l'affinité de la snRNP U1 est stimulée ou au contraire inhibée par des facteurs intervenant dans l'épissage alternatif comme ASF/SF2 (stimulation) ou hnRNPA1 (répression) [77]. Parallèlement à l'interaction de la snRNP U1 au 5'ss, la protéine SF1 (BBP ou Msl5 chez la levure) se lie au point de branchement [160, 244]. Cette liaison se fait avec U2AF65 (Mud2p chez la le-

vure) qui interagit directement avec SF1 et la séquence polypyrimidine [169, 286]. La liaison SF1-U2AF65 se fait par l'intermédiaire du domaine RRM3 de U2AF65 [276]. Il permet une meilleure reconnaissance du BP car la liaison de SF1 sur le BP est faible alors que celle de U2AF65 sur la séquence polypyrimidine est plus forte [21]. Cette liaison serait favorisée par des phosphorylations de proline sur SF1 [196]. Chez la levure, on trouve des interactions entre la snRNPU1 et le complexe BBP-Mud2, ce qui pourrait permettre un meilleur recrutement de la snRNP U1 [324]. Des travaux montrent néanmoins que SF1 ne serait pas indispensable pour l'épissage chez l'homme mais serait plutôt l'apanage de certains gènes [298]. Dans ce cas, la snRNP U1 reconnaît directement le point de branchement. Il est en tout cas indispensable à la viabilité des cellules. Lorsque la U2snRNP arrive, elle fait une interaction avec le domaine RS (Arginine-Serine) de U2AF65 par l'intermédiaire de sa sous-unité SAP155 (SF3b155) [346, 108]. Cette interaction perturbe la liaison U2AF65-SF1 et SF1 quitte l'ARN. Comme le domaine RS de U2AF65 interagit avec la séquence du BP lorsqu'il est lié à la séquence polypyrimidine, il permet au snARN U2 de venir s'hybrider à la séquence BP [316]. D'autres interactions permettent certainement l'arrimage de U2snRNP sur le BP, étant donné que des travaux montrent un rôle important pour les sous-unités SF3b de U2snRNP au BP et une interaction entre U2AF65 et l'ATPase UAP56 [85, 323].

Le complexe A

La snRNP U1 interagit avec U2AF65 qui est lui-même en interaction avec la snNP U2. Pour que la snRNP U2 s'associe de façon stable à l'ARNm par l'hybridation du snARN U2, deux ATPase sont requises : UAP56 et Prp5. UAP56 est une ATPase à boîte DEXD/H (ARN hélicase ATP-dépendante) qui, lorsqu'elle est absente, empêche la formation du complexe A [85]. On ne sait pas à l'heure actuelle comment agit UAP56. UAP56 pourrait jouer son rôle d'hélicase et réarranger les structures secondaires du snARN U2 pour lui permettre de se lier au BP. Cependant l'activité hélicase de UAP56 est faible, non spécifique et ne se fait que sur de très courts ARNs double brin. De même son activité ATPase est très faible et pourrait nécessiter l'intervention d'autres facteurs [282]. Une autre possibilité serait la stimulation du départ de U2AF65 du 3'ss par UAP56, ce qui faciliterait la liaison de la U2 snRNP car lorsque Mud2 (SF1) est absent, Sub2 (UAP56 de levure) n'est pas nécessaire [152]. On note qu'outre ses fonctions dans le complexe A, UAP56 reste dans le splicéosome

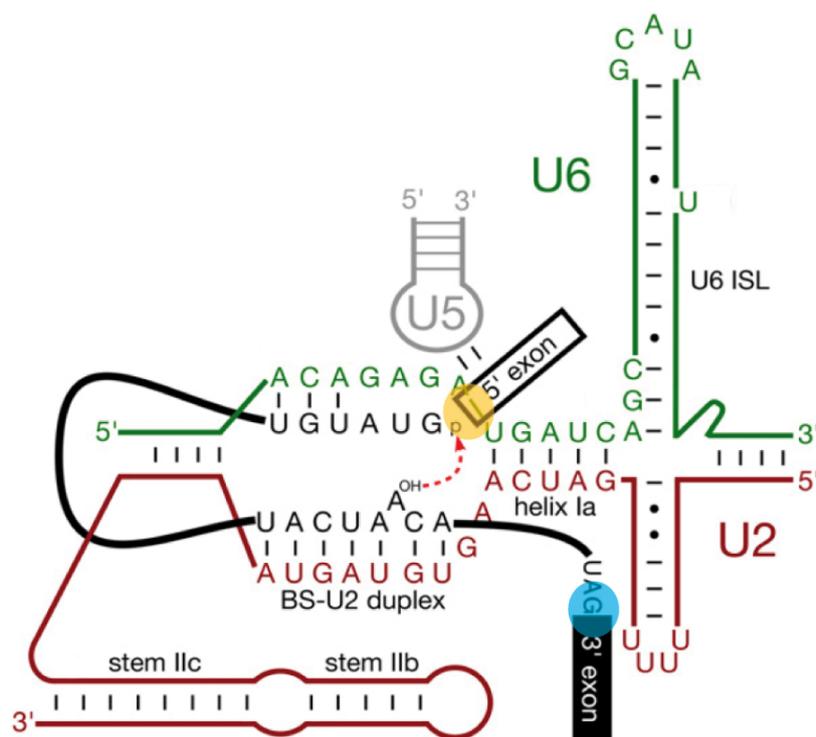
jusqu'à l'export où il recrute le facteur d'export Aly/REF de façon ATP-dépendante, y compris pour des ARNs sans introns [158, 300]. L'autre ATPase impliquée dans la formation du complexe A est Prp5. Au contraire de UAP56, Prp5 montre une forte spécificité pour le snARN U2 et est capable de changer sa conformation [230, 2]. Prp5 est physiquement lié à la di-snRNP U1-U2, sans besoin d'hydrolyse d'ATP [338]. L'ATP permet les changements de conformation de l'ARN U2. Mais ceux-ci, s'ils sont trop efficaces, empêchent l'épissage. Prp5 pourrait donc jouer un rôle régulateur sur l'épissage via son activité ATPase [339].

Le complexe B

L'arrivée de la tri-snRNP U4/U6•U5 conduit à l'obtention du complexe B [70]. Dans la tri-snRNP, les snARNs U4 et U6 sont hybridés entre eux par deux tiges [293]. Pour former un complexe catalytique, des réarrangements vont être nécessaires pour rompre l'hybridation U4-U6, rompre également l'hybridation de U2 sur lui-même et former une hybridation U6-U2, et enfin échanger l'ARN U1 pour l'ARN U6 au site 5'ss. Ce dernier évènement est coordonné par les NTPases (GPase et ATPase) de la snRNP U5 [309]. Prp28p (famille DEXD/H) joue un rôle dans la déstabilisation de l'hybride U1-5'ss, probablement en perturbant l'interaction avec la protéine U1C [51]. La cassure de l'hybride U4-U6 serait aussi le fait de la snRNP U5 via Brr2p (famille DEXD/H), Snu114p (GTPase) et Prp8p mais nécessiterait aussi l'intervention de UAP56 [189, 289, 281]. Ces réarrangements se font en plusieurs étapes et font intervenir d'autres sous-unités de la tri-snRNP et d'autres facteurs d'épissage comme Tat-SF1 (Cus2p chez la levure) [245, 116]. A ce stade, la première réaction catalytique va pouvoir avoir lieu car le cœur catalytique est prêt. Il est formé de l'hybride U2-U6 [260]. Ce cœur fait intervenir des ions Mg²⁺ [343]. La figure transesterification représente les hybridations nécessaires à cette première réaction. Pour la réaliser, l'ARN U6 s'hybride au site 5'ss [157].

Le complexe C

La première réaction catalytique aboutit au complexe C qui va permettre la deuxième réaction. Des réarrangements sont encore nécessaires pour permettre celle-ci [156, 157].



Smith, Query et al. 2008

FIGURE 1.9 – **Transesterification.** Schéma des interactions ARN-ARN contribuant à la première réaction catalytique de l'épissage. Le pré-ARNm est en noir, le snARN U2 est en rouge, le snARN U5 en gris et le snARN U6 en vert. Représentation de l'attaque nucléophile par le point de branchement sur le site 5'ss par laquelle va être générée la première liaison phospho di-ester. En jaune, le groupe hydroxyle de l'exon en 5' généré après la première réaction.

Après la première transestérification, le groupe hydroxyle de l'exon en 5' (en orange sur le schéma) se trouve physiquement à côté de la liaison phospho-diester 2'-5'. Pour pouvoir effectuer la deuxième transestérification, ce groupe hydroxyle doit être positionné à proximité du site accepteur (en bleu sur le schéma) afin de pouvoir l'attaquer. Un ensemble de travaux mettent en évidence que le lasso généré après la première réaction est libéré du site catalytique afin que le 3'ss puisse y rentrer [290]. Au cours de ces réactions de nombreux facteurs d'épissage interviennent, ce qui explique pourquoi on trouve autant de composants lors des purifications de complexes [311, 355, 52]. Parmi ces protéines, une semble jouer un rôle à part entière : Prp8. D'une part Prp8 régulerait l'activité de Brr2p et de Prp28p [163], d'autre part on le trouve accroché aux snARNs U5 et U6 dans le cœur catalytique ainsi qu'aux trois sites essentiels de l'ARN : le site donneur, le site accepteur et le point de branchement [285, 35, 109, 262].

Fin de l'épissage, recyclage des composants

La libération de l'ARNm mature, le désassemblage du splicéosome et son recyclage vont impliqués de nouveaux réarrangements ARN/ARN et ARN/protéine. Des protéines ont été identifiées pour jouer un rôle dans ce recyclage. Prp24 dont l'orthologue humain est p110, permet la ré-hybridation des ARNs U4 et U6 [254, 19]. Prp43 catalyse la dissociation de l'intron en lasso et le désassemblage simultané du splicéosome [7]. Il est pour cela aidé de deux autres protéines Ntr1/Ntr2 [299, 308]. Des protéines ayant déjà participé aux réactions pourraient aussi jouer un rôle dans le recyclage comme le complexe Prp19 [49]. Les protéines impliquées dans les réarrangements, les réactions catalytiques et le recyclage sont principalement des ATPases à boîte DExD/H [239]. En attendant d'être réutilisés, les facteurs d'épissage vont dans les speckles. Les speckles sont des structures nucléaires visualisés par microscopie photonique en utilisant des anticorps dirigés contre des facteurs d'épissage, notamment SC35. Il y a environ de 25 à 50 speckles observés par noyau en interphase [172]. Les speckles changent lorsque l'on bloque l'épissage [235, 209]. Ces structures comprennent à la fois des granules de stockage des facteurs d'épissage (Interchromatin Granule Cluster ou IGC) et des zones d'activité d'épissage autour ds ARNs naissants (périfibrilles) . Néanmoins, des simulations montrent que les facteurs d'épissage n'y résident pas longtemps, ils y rentrent et en sortent selon un mode Brownien [261].

1.2.4 Les UsnRNP

Suprasplicéosome

Le détail de leur biogenèse est décrit dans la troisième partie. On peut retenir que ce sont des ARNs transcrits par la RNAPII, sauf U6 qui est transcrit par la RNAPIII. Après leur synthèse, ils sont exportés dans le cytoplasme pour y subir des modifications et une fois revenus dans le noyau, ils vont dans les Cajal bodies pour y être modifiés par de petits ARNs (scaARNs, small Cajal bodies RNAs). Les Cajal Bodies peuvent être aussi le lieu de super-assemblages de UsnRNP. En effet, des arguments montrent que le processus d'assemblage du splicéosome étape par étape cité ci-dessus n'est pas la réalité mais plutôt un artefact *in vitro*. Des penta-snRNP ou supraspliceosomes ont ainsi été mises en évidence [226, 294, 195, 11]. Un supra-splicéosome est composé de quatre spli-

céosomes sur un ARNm. Cette structure a été isolée *in vivo* et la contradiction entre les deux modèles pourrait s'expliquer par le fait qu'*in vitro*, l'assemblage se fait sur un ARNm déjà transcrit et non en cours de transcription comme dans la cellule. Le modèle controversé du suprasplicéosome permet aussi de donner une réponse à l'énigme des ARNm multiexoniques. Quatre nucléosomes feraient l'épissage en même temps. De plus, les réarrangements nécessaires seraient moins nombreux et moins importants [292].

Composition des UsnRNP

UsnRNP	U1 snRNP	U2 snRNP	U4 snRNP	U5 snRNP	U6 snRNP
snARN	U1	U2	U4	U5	U6
transcription	RNAPII	RNAPII	RNAPII	RNAPII	RNAPIII
longueur	164	186	142	116	106
cœur protéique	B/B'	B/B'	B/B'	B/B'	Lsm2
	D1	D1	D1	D1	Lsm3
	D2	D2	D2	D2	Lsm4
	D3	D3	D3	D3	Lsm5
	E	E	E	E	Lsm6
	F	F	F	F	Lsm7
	G	G	G	G	Lsm8
protéines spécifiques	U1A/Mud1	U2-A'/Lea1	15,5kD/Snu13	Prp8	cf. U4
	U1C/Yhc1	U2-B'/Msl1	HPRP3/Prp3	U5-200kD/Brr2	
	U1 70k/Snp1	SF3a60/Prp9	HPRP4/Prp4	U5-116kD/Snu114	
		SF3a66/Prp11	RY-1	U5-102kD/Prp6	
		SF3a120/Prp21	USA-Cyp/Cpr1	U5-100kD/Prp28	
		SF3b49/Hsh49		U5-40kD	
		SF3b145/Cus1		U5-15kD/Dib1	
		SF3b130/Rse1			
		SF3b155/Hsh155			
		p14/Snu17			

TABLE 1.3 – **Les protéines spécifiques des UsnRNPs chez l'homme et la levure.** Sont décrites les 5 UsnRNP des eucaryotes. Les noms correspondent aux noms chez l'homme suivis du nom chez la levure. Les snRNP U4 et U6 formant un complexe, les protéines spécifiques sont communes [34, 134, 242].

1.2.5 Définition de l'exon, éléments régulateurs de l'épissage et épissage alternatif

Les exons chez les mammifères sont plutôt courts (50-250bp, la moyenne étant autour de 150pb) alors que les introns sont très longs (la moyenne est de 3365 pb, certains font plus de 10kb) [173] et représentent plus de 90% des transcrits [325]. Les premières expériences ont démontré que la mutation d'un site d'épissage (figure 1.11 A) entraînait l'exclusion de l'exon. Il a alors été

supposé que l'épissage reposait sur une « définition par l'exon » : il existe une interaction initiale et une stabilisation à travers l'exon des facteurs liés au 3'ss et au 5'ss avant la formation d'interactions à travers l'intron [200]. Une représentation schématique de la définition de l'exon est présentée dans la figure 1.10 A. Ce système est retrouvé chez les mammifères. Il s'oppose à la « définition par l'intron » plus commune chez les invertébrés, les plantes et les champignons [337], également schématisé dans la figure 1.10 A. L'une ou l'autre de ces définitions est le résultat du croisement d'informations portées par des séquences régulatrices de l'épissage (SRE). Ces séquences se distinguent en 2 groupes en fonction de leur localisation (figure 1.10 B) : ESE et ESS (exonic splicing enhancer/silencer) séquences contenues dans l'exon et stimulant ou inhibant l'inclusion de l'exon ; ISE et ISS (intronic splicing enhancer/silencer) séquences contenues dans l'intron et stimulant ou inhibant l'usage des sites d'épissage adjacents [325].

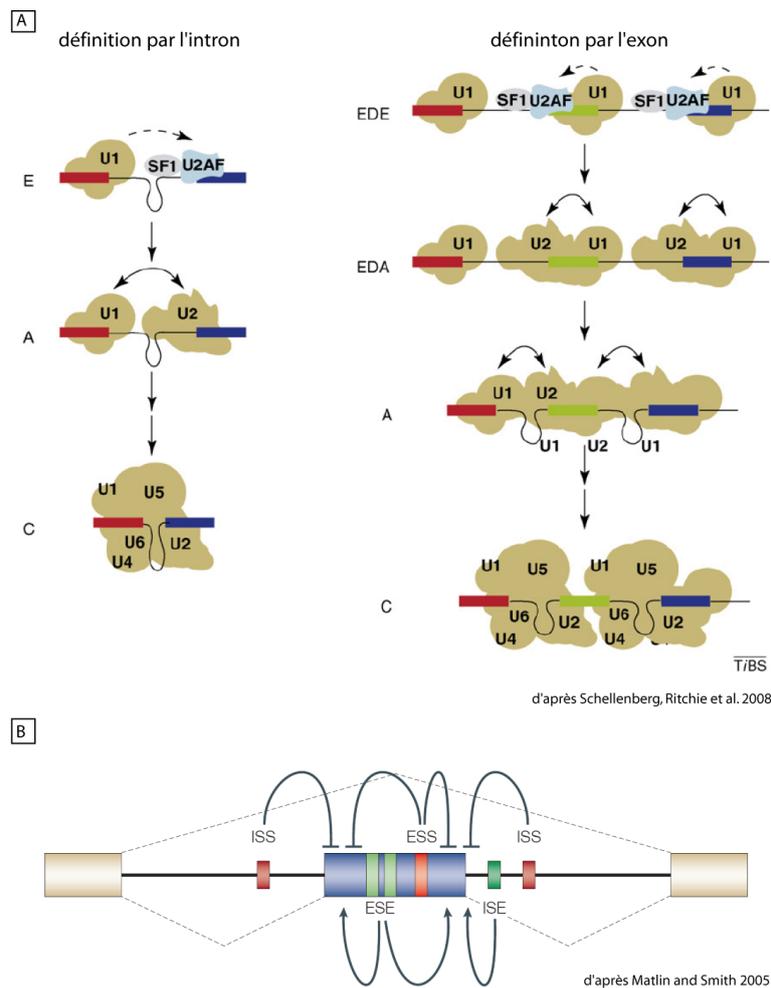


FIGURE 1.10 – Définition par l'exon ou par l'intron et les séquences régulatrices de l'épissage. A. Il existe deux modes pour le splicéosome de reconnaître les sites actifs pour l'épissage : par la définition de l'intron (à gauche) ou par la définition de l'exon (à droite). Dans la définition de l'intron, ce sont les sites 5'ss (ou donneur) et 3'ss (ou accepteur) qui encadrent l'intron qui permettent l'assemblage. Ce mode est utilisé chez les invertébrés, les plantes et les champignons. La définition de l'exon consiste à reconnaître les sites 5'ss et 3'ss encadrant l'intron. Ce mode est privilégié chez les mammifères. Dans les deux cas, le splicéosome passe par les mêmes complexes, à la différence que dans la définition par l'exon, le complexe E et A sont légèrement différents de leurs homologues dans la définition par l'intron. Dans le complexe EDE (Exon-defined E), c'est la snRNP U1 en aval qui interagit avec U2AF. Et dans le complexe EDA (Exon-defined A), l'interaction se fait entre la snRNP U2 en amont et la snRNP U1 en aval de l'exon. Avec la formation du lasso, les interactions changent et se font dans le même ordre 5'-3' que dans la définition par l'intron. **B.** Les séquences régulatrices de l'épissage se trouvent dans les exons et dans les introns. Elles peuvent être inhibitrices de l'épissage : ESS (Exonic Splicing Silencer) et ISS (Intronic Splicing Silencer) et favoriser l'exclusion de l'exon ou l'inclusion de l'intron respectivement. Elles peuvent aussi être activatrices : ESE (Exonic splicing Enhancer) et ISE (Intronic Splicing Enhancer) et entraîner l'inclusion de l'exon ou l'exclusion de l'intron respectivement.

Il n'existe pas de séquences consensus à proprement parler. Pour autant, il existe des familles de protéines associées à ces séquences. Les ESE sont très divers par leurs séquences et presque tous les exons en contiennent [81]. La plupart recrutent des protéines SR (Ser-Arginine) par leurs domaines RRM qui établissent des interactions protéiques et ARN par leurs domaines RS. Les protéines les plus connues sont SF2/ASF et SC35 [187, 41]. Les ESS au contraire, sont plutôt associés aux protéines de la famille des hnRNPs (heterogeneous nuclear RNP). Les modes d'action de ces dernières sont variés. Par exemple, le PTB (Polypyrimidine Tract Binding protein ou hnRNP I) se lie à la séquence polypyrimidine et empêche les snRNP U1 et U2 de se fixer en interférant avec U2AF65 [271]. hnRNPA1, lui, peut se lier de chaque côté d'un exon et l'exclure de l'ARNm en cours d'épissage [221]. Les ISE et ISS sont moins étudiés, donc moins connus. Néanmoins, les données actuelles montrent que les ISS recrutent les mêmes protéines que les ESS [272].

Pour comprendre l'épissage alternatif, il faut ajouter à ces interactions, leur variabilité dans l'espace et dans le génome. En effet, ces séquences peuvent n'avoir aucun rôle ou bien un rôle opposé en fonction de la localisation : si la séquence est dans l'intron ou dans l'exon, proche d'un 3' ou d'un 5', son effet pourra être activateur ou inhibiteur. Ou bien en fonction du gène : la séquence n'a pas d'effet dans un autre gène. Un exemple du premier état est le cas du triplet GGG qui est lié par des protéines hnRNPH. Cette séquence, située dans un intron, stimule l'épissage des exons qui l'encadrent alors que, située dans un exon, cette séquence inhibe son épissage [38]. On comprend bien que ces séquences jouent un rôle capital dans l'épissage alternatif. Elles entraînent plusieurs possibilités d'épissage alternatif (figure 1.7). Mais elles ont aussi un rôle essentiel dans la fiabilité et l'exactitude de l'épissage qui repose sur des séquences dégénérées (5'ss, 3'ss et BP) (figure 1.11) [128].

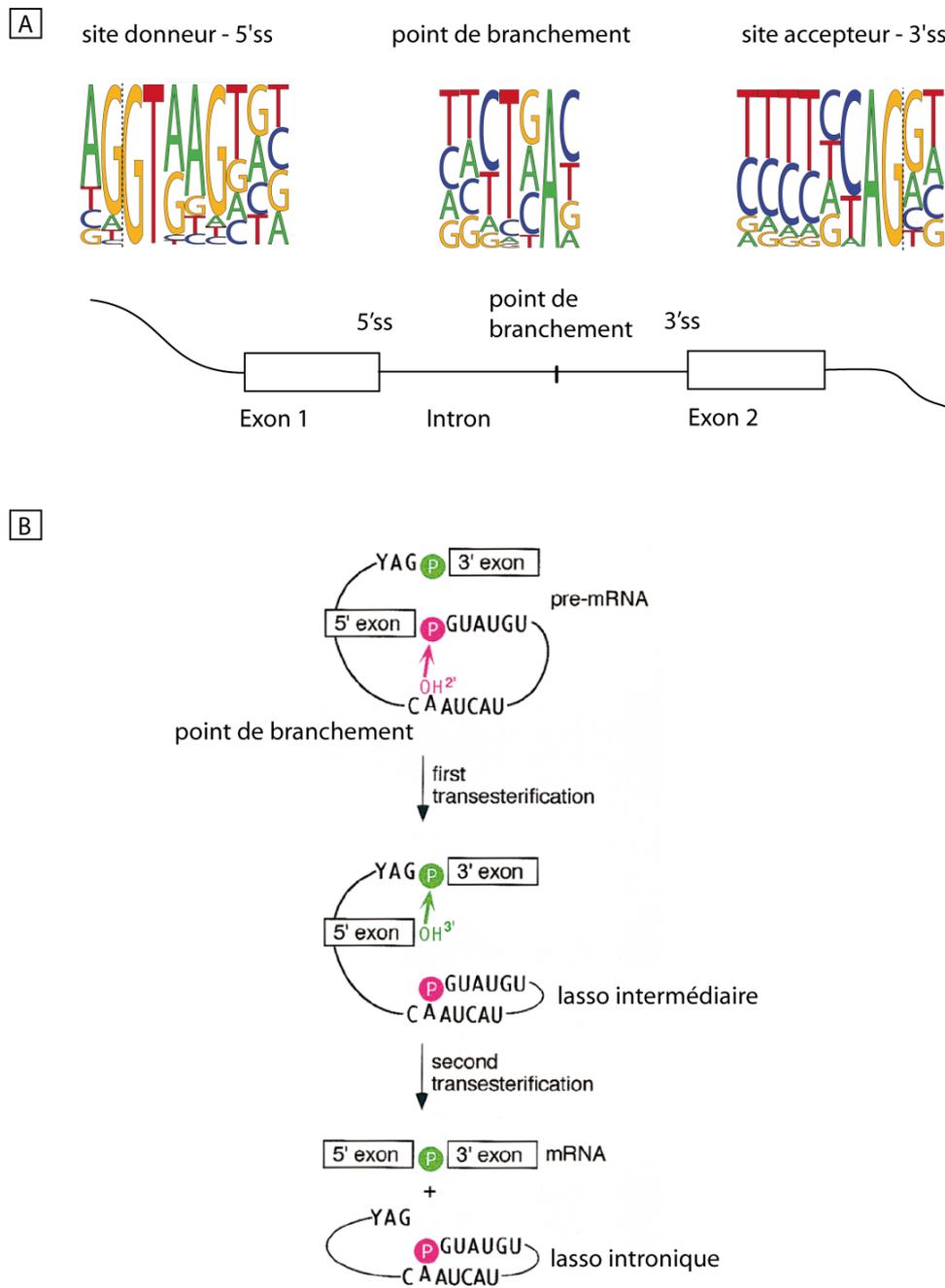


FIGURE 1.11 – **Sites et réactions catalytiques de l'épissage.** **A.** L'épissage fait intervenir 3 sites : le site 5'ss ou site donneur, le point de branchement et le site 3'ss ou site accepteur. Chaque site est déterminé par une séquence consensus qui repose essentiellement sur un ou deux nucléotides très conservés. Les autres sont relativement dégénérés. **B.** Attaque nucléophile du point de branchement sur le site 5'ss. L'intron forme un lasso et l'exon en 5' est coupé. L'exon en 5' attaque le site 3'ss. Formation de l'ARNm épissé et du lasso intronique.

La transcription et l'épissage sont deux étapes indispensables dans la production de protéines (au moins chez les métazoaires) et sont très proches spatialement et dans le temps. L'existence d'interactions entre ces deux mécanismes est une hypothèse facile à admettre et dont des preuves de plus en plus nombreuses sont apportées par la littérature. Nous allons y venir en nous attardant sur la protéine majeure de la transcription : l'ARN polymérase II.

Chapitre 2

ARN polymérase II

2.1 Le domaine carboxy-terminal

2.1.1 Structure et modifications

Structure

Le domaine carboxy-terminal de la sous-unité Rpb1 se compose de 52 répétitions du motif Tyr₁-Ser₂-Pro₃-Thr₄-Ser₅-Pro₆-Ser₇ chez les mammifères à 26 chez la levure (32 chez les nématodes, 45 chez la drosophile) [76]. Sa longueur est directement liée à la viabilité des cellules car en-dessous de 10 répétitions, les levures ne survivent pas [228]. Chez les mammifères, les 3 premières répétitions et la dernière sont indispensables à la vie [87, 48, 44]. Pourtant le CTD ne participe pas au processus de la transcription. Son caractère indispensable lui est conféré par ses multiples interactions avec les facteurs de transcription et de maturation de l'ARNm. La figure 2.1.1 représente à l'échelle la RNAPII et son CTD. Le CTD est lié à Rpb1 par le linker (chaîne de 80 résidus). Le début du linker est associé à Rpb7 qui forme avec Rpb4 une sous-unité qui sort de l'enzyme [8]. Le CTD ne possède pas de structure définie. Dépliés, le CTD mesurerait 650 Å et le linker 250 Å (l'enzyme a un diamètre de 150 Å) [58]. Linker et CTD semblent cependant être compactés aux environs de la RNAPII, au moins dans la forme déphosphorylée : 100 Å. Phosphorylé, il est plus étendu, plus désordonné et plus sensible aux protéases, probablement à cause des charges négatives qui se repoussent [349]. Les études structurales penchent plutôt vers une conformation en spirale qui peut

être plus ou moins compactée comme représenté sur la figure 2.1.1. Cette structure désordonnée lui permet d'interagir avec de nombreuses protéines [206].

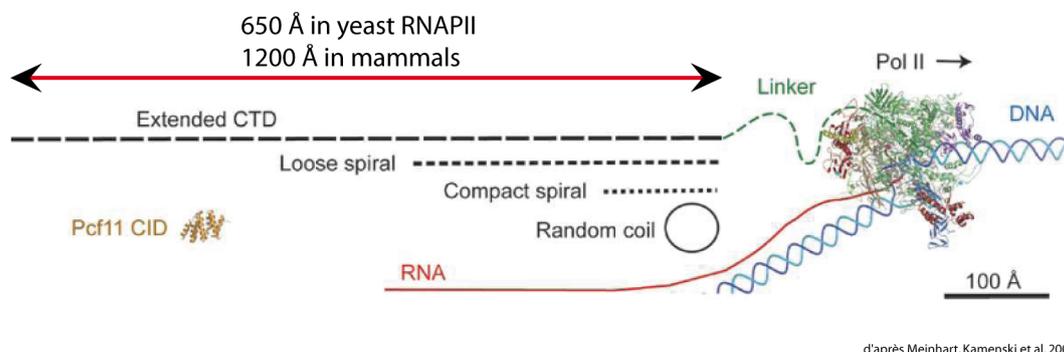


FIGURE 2.1 – **Représentation de la RNAPII et de son CTD à l'échelle.**

Le CTD est très peu structuré. Les études cristallographiques le représentent dans une conformation en spirale qui, selon sa compaction, peut lui permettre de tenir dans un espace de 100 Å ou de s'étendre jusqu'à 1200 Å chez les mammifères. La figure représente la RNAPII de levure, le CTD déplié représente 650 Å. Le domaine CID de Pcf11 interagissant avec le CTD est mis à titre indicatif pour représenter l'échelle.

De nombreuses protéines participant à différents mécanismes cellulaires interagissent avec le CTD : maturation de l'ARN, transcription, modification de la structure de la chromatine, lésion / réparation de l'ADN, dégradation de protéines, modification des snARNs, biogenèse des snoRNP [253, 250, 249, 126, 66]. La liste n'est d'ailleurs peut-être pas finie. Comme le CTD est une suite de répétitions d'un même motif, l'existence d'un code permettant l'association sélective de protéines est une théorie qui fait son chemin, d'autant que la conservation est importante au sein des eucaryotes. On peut d'ailleurs diviser le CTD en trois parties : le linker et ses séquences de type heptapeptide, la région des répétitions heptapeptidiques et les séquences C-terminales divergentes. La taille et la présence de ces trois régions varient en fonction des organismes [47]. Si l'on prend les mammifères, le CTD y est à la fois le plus long et le moins conservé. Dans la deuxième région, les 26 premières répétitions sont complètement conservées alors que les 26 dernières ne comptent que 3 séquences consensus [75]. En revanche, chez *Drosophila melanogaster*, très peu de séquences sont totalement identiques entre elles. Parmi les résidus les plus conservés, les deux prolines et la Tyr₁ sont les plus représentées sur les 52 répétitions [47]. A l'inverse, le résidu le plus variant est la Ser₇. Cette divergence en partie terminale peut avoir un rôle crucial pour la

régulation de l'expression génétique. Le code du CTD est porté par cette séquence Tyr₁-Ser₂-Pro₃-Thr₄-Ser₅-Pro₆-Ser₇. Néanmoins, il existe des travaux qui montrent l'implication de deux heptapeptides à la fois [295, 182]. Ainsi le motif essentiel du CTD serait le motif Ser₂-Ser₅-Ser₂ [188]. En réponse à ces travaux, Chapman pense que le vrai motif serait en fait Ser₅-Pro₆-Ser₇-Tyr₁-Ser₂-Pro₃-Thr₄ [47, 206].

Modifications

Plusieurs modifications entrent en jeu dans le code du CTD. Celles qui sont connues à ce jour sont des phosphorylations, des glycosylations et des peptidyl-prolyl isomérisations. Celles-ci sont représentées sur la figure 2.2. Il a aussi été découvert une méthylation de la Lys₇ exclusivement sur la *RNAPII_A*.

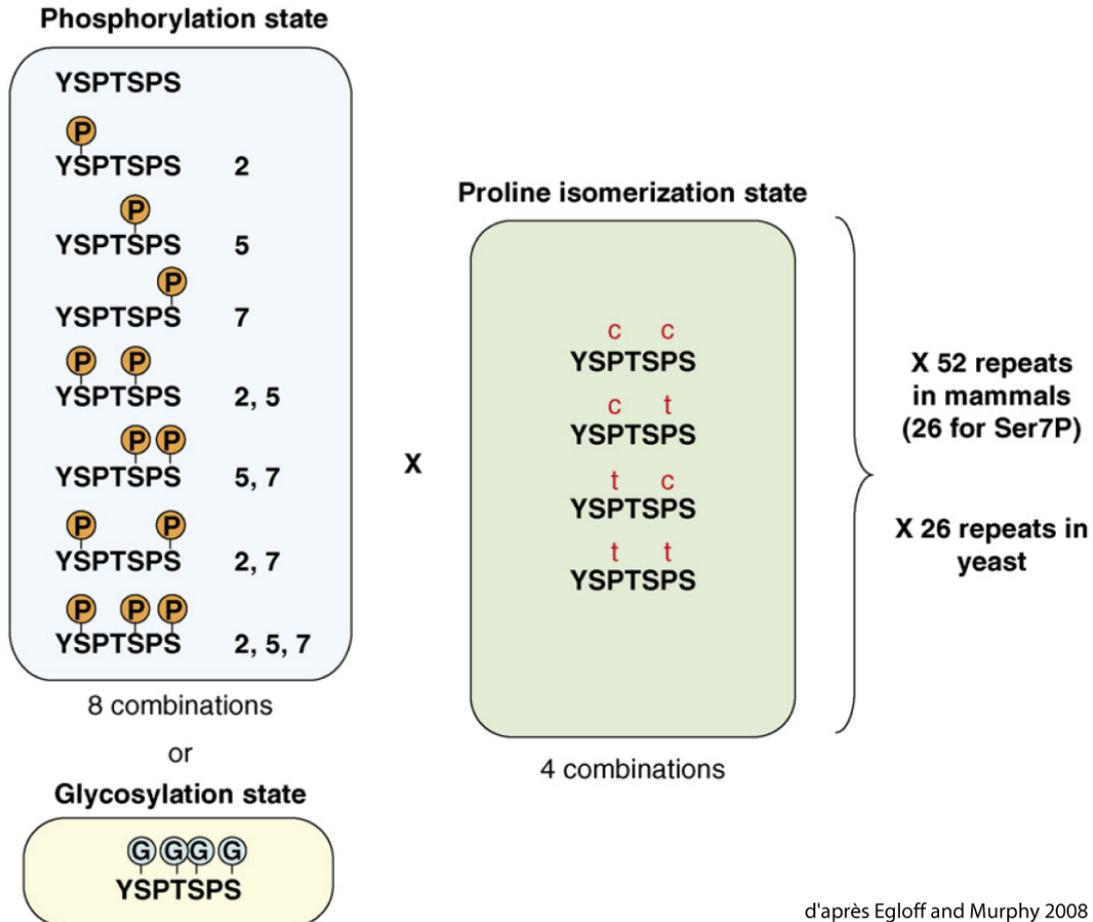
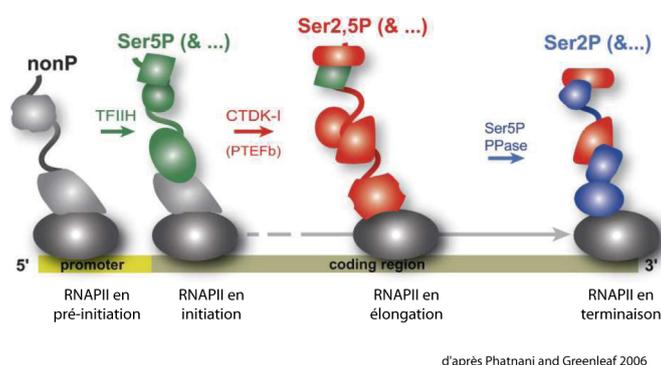


FIGURE 2.2 – **Le code du CTD.** Représentation de toutes les phosphorylations de sérine et les isomérisations de proline. Il n'a pas été déterminé si les phosphorylations des Tyr₁ et Thr₄ coexistaient avec les phosphorylations de sérine. L'insertion d'un acide aminé supplémentaire entre deux paires d'heptapeptides est toléré alors qu'entre deux heptapeptides pas du tout. Ceci complique un peu plus le code car cela voudrait dire que l'unité de reconnaissance serait une paire d'heptapeptide, ou bien un autre motif (voir texte). Même si le code est très complexe, avec 52 répétitions ou 26 paires, chacune ayant son patron de modifications, en réalité le nombre de combinaisons est restreint par le recrutement de différents ensembles de d'enzymes de modification aux endroits appropriés. De plus, certaines modifications stimuleront ou en orienteront d'autres.

Les phosphorylations concernent principalement les Ser₂ et Ser₅. Récemment a aussi été mise en évidence la phosphorylation de la Ser₇, mais également la Tyr₁ et la Thr₄. La Ser₅ phosphorylée est plus largement représentée sur la RNAPII au début du gène, dans les environs du promoteur. Elle est phosphorylée par l'activité kinase du complexe TFIIH portée par Cdk7 (kin28 chez la levure). Cette phosphorylation entraîne le recrutement des facteurs négatifs d'élongation (DSIF et NELF) qui provoquent une pause transcriptionnelle. Pendant cette pause, les enzymes de la coiffe sont recrutées sur les Ser₅P [155]. Au contraire, on trouve la Ser₂ phosphorylée aux extrémités 3' des gènes [215]. Celle-ci est phosphorylée par la kinase Cdk9 (Cdk-1 chez la levure) du complexe P-TEFb [252, 133]. Cette phosphorylation jouerait un rôle dans le passage à l'élongation [247]. Ces deux phosphorylations ont pendant longtemps focalisé l'attention et apporté la notion de «code». Aujourd'hui un consensus est accepté, selon lequel la RNAPII est recrutée au promoteur dans un état non phosphorylé. Puis elle subit une première phosphorylation par TFIIH sur la Ser₅ ce qui lui permet d'initier la transcription. Et l'étiquette Ser₂P (Ser₂ phosphorylée) permet l'élongation de la transcription. Il existe donc deux gradients opposés le long du gène décrits sur la figure 2.3 : un gradient décroissant de la Ser₅P de 5' vers 3' et inversement pour la Ser₂P [249]. Cependant si cette observation est vraie chez la levure, elle ne semble pas être universelle car on trouve de la Ser₅P tout le long des gènes chez les mammifères [215].



d'après Phatnani and Greenleaf 2006

FIGURE 2.3 – **Patron de phosphorylation du CTD de la RNAPII chez la levure.** Le long d'un gène en cours de transcription, la RNAPII modifie son patron de phosphorylation sur son CTD. Dans le complexe de pré-initiation, la RNAPII est hypophosphorylée. A l'initiation, le CTD est phosphorylé sur les Ser₅ par le facteur TFIIF (Ser₅P). Pour passer en mode d'élongation, la RNAPII va ensuite être phosphorylée sur les Ser₂ du CTD par le P-TEFb (Ser₂P). Au fur et à mesure que la RNAPII se déplace sur le gène, son patron de phosphorylation va se modifier. Il va ainsi se créer un double gradient aux sens contraires : une plus forte présence de Ser₅P au début du gène et de Ser₂P à la fin du gène.

Les modifications permanentes du patron de phosphorylation sont le fruit d'une balance entre kinases et phosphatases [231, 237]. Outre Cdk7 et Cdk9, les Ser₂ et Ser₅ peuvent aussi être phosphorylées par Cdk8 (Srb10 chez levure) et cdc2 (Cdk1 chez levure). La Ser₅ peut aussi être phosphorylée par la kinase ERK-1/2 [30]. Les déphosphorylations sont en grande partie le fruit de l'action de la phosphatase FCP1, surtout sur les Ser₂, et SCP1 (Ssu72 chez levure) sur Ser₅ [121, 120]. Elles sont indispensables à la transcription et au recyclage de la RNAPII [259]. L'absence de FCP1 provoque d'ailleurs de graves problèmes chez l'homme [318]. D'autres phosphorylations, moins fréquentes, ciblent les autres acides aminés de l'heptapeptide. Les Ser₇, Tyr₁ et Thr₄, sont elles aussi phosphorylées par respectivement les kinases DNA-PK et Abl1/2 pour les deux premières. La Ser₇ s'enrichit vers le milieu et l'extrémité 3' du gène, suggérant un rôle dans la terminaison et/ou la maturation des ARNm [46, 76]. C'est pourtant le résidu le moins conservé de l'heptapeptide. La Thr₄ est indispensable chez les mammifères mais on ne sait pas pour quelle raison puisque sa mutation n'affecte ni la transcription, ni l'épissage. La Tyr₁ n'est phosphorylée que dans le contexte *Lys7-Tyr1*, pour autant aucune fonction clairement définie n'a encore été avancée. Les peptidyl-prolyl isomérisations ne modifient pas des résidus mais changent l'orientation des prolines. Il y a deux liaisons peptidyl-prolyl dans chaque heptapeptide qui peuvent être soit en conformation *trans*, soit en *cis*.

Ces réactions seraient effectuées par une peptidyl-prolyl *cis/trans* isomérase (PPIase) *in vivo*, Pin1 chez les mammifères (ESS1 chez la levure). Pin1 est une PPIase qui se lie à des résidus phosphorylés, elle est donc le candidat idéal pour effectuer ces modifications sur le CTD. Pin1 influencerait l'état de phosphorylation du CTD en inhibant la phosphatase FCP1 [340]. Il bloquerait la RNAPII sous forme hyperphosphorylée, inhibant ainsi l'initiation de la transcription [341]. D'autre part, Pin1 joue un rôle essentiel dans le contrôle du cycle cellulaire, plus précisément dans la mitose. Pin1 stimule l'hyperphosphorylation des phosphoprotéines mitotiques, permettant l'assemblage des chromosomes [341]. La phosphorylation de la RNAPII et sa régulation via Pin1 jouerait donc un rôle dans le contrôle du cycle cellulaire [341]. Le CTD peut aussi subir des glycosylations. Une équipe a effectivement montré que les résidus sérine et thréonine peuvent recevoir sur leurs groupes hydroxyles des monosaccharide-N-acétylglucosamine (O-GlcNAc) [142]. Glycosylations et phosphorylations semblent mutuellement exclusives, ce qui suggère que la RNAPII est recrutée au promoteur sous une forme glycosylée et qu'il lui faut l'activité d'une N-acétylglucosaminase pour enlever les groupes O-GlcNAc et permettre la phosphorylation [75].

L'ensemble de ces modifications d'acides aminés sur l'heptapeptide laisse entrevoir de nombreuses combinaisons responsables du recrutement de telle ou telle protéine comme le suggère la figure 2.2. A titre d'exemple, on sait que les enzymes de la coiffe sont sensibles à la Ser₅P, Pin1 à la combinaison Ser₂P-Ser₅P, Spt6 (chaperone des histones H3) à la Ser₂P et Pcf11 (enzyme de maturation du 3') à une Ser₂P accompagnée des deux prolines en trans [75]. La figure 2.4 montre dans quelle mesure les modifications du CTD permettent la progression de la RNAPII dans la transcription.

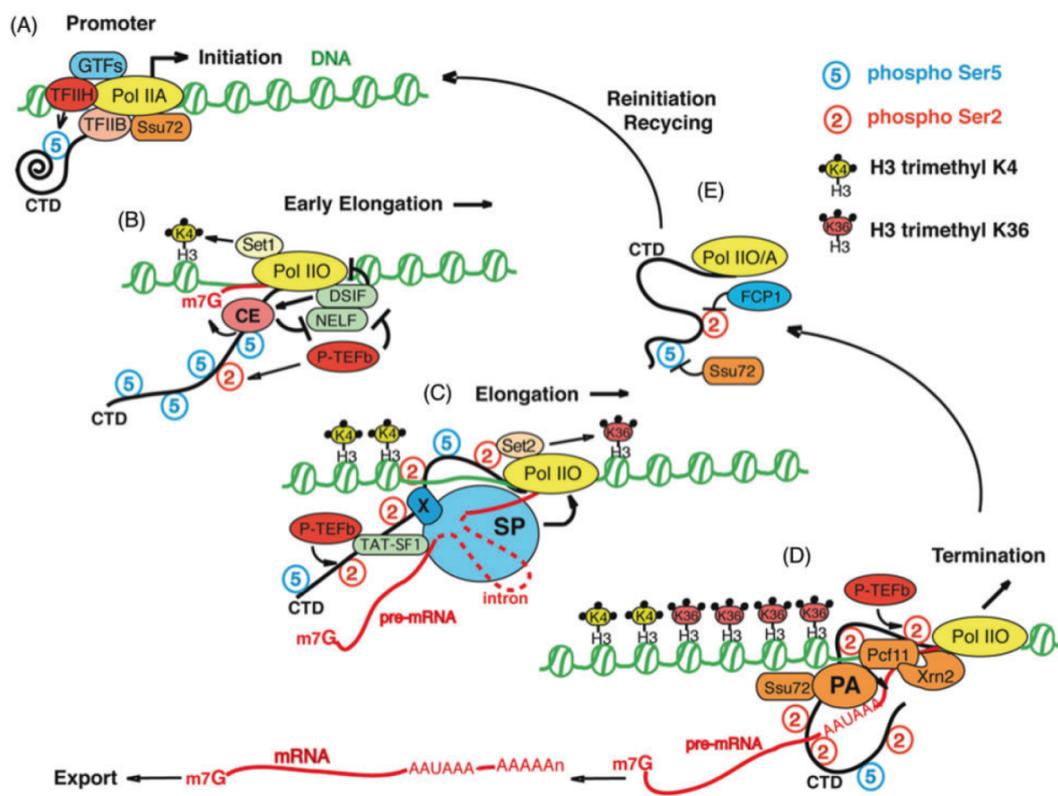
2.1.2 Le CTD, une plate-forme de recrutement

Du fait de sa taille, de sa structure désordonnée et des nombreuses combinaisons qu'il offre, le CTD est une plate-forme formidable pour le recrutement de facteurs divers et variés, jouant un rôle dans la transcription ou dans l'éventail des processus de maturation des ARNm.

Rôle du CTD dans la transcription

Modification de la chromatine

Le lien de la RNAPII avec des enzymes de modification de la chromatine est évident puisque l'ADN doit être accessible pour faire l'élongation de la transcription. Plusieurs histones acétylases sont d'ailleurs recrutées à l'initiation par les facteurs d'initiation [55]. Mais plusieurs autres enzymes de modification de la chromatine sont associées au CTD phosphorylé [250]. Set2 est une histone lysine méthyltransferase [296]. Elle méthyle les histones H3 sur la Lys₃₆. C'est donc un répresseur transcriptionnel. On la trouve associée à un complexe d'histones déacétylases et impliquée dans la régulation négative de la transcription [143]. La figure 2.4 représente un gène en cours de transcription. On voit dans C et D que Set2 est recrutée pendant l'élongation pour méthyliser les histones H3 afin d'empêcher une réinitiation de la transcription. Elle joue aussi un rôle dans l'élongation de la transcription car mutée ou absente, l'élongation est perturbée, les cellules meurent [161]. On la retrouve associée à la RNAPII phosphorylée en Ser₂ par son domaine SRI (Set2 Rpb1 interacting), un peu partout le long des gènes [181]. Lorsque Set2 est mutée sur son domaine d'interaction avec la RNAPII, on observe une perturbation de la localisation de la RNAPII le long des gènes [153]. Cela suggère que Set2 a un rôle dans le nombre de RNAPII recrutées sur un gène. Comment expliquer cette ambivalence ? Il est possible que Set2 puisse jouer les deux rôles en fonction de l'étape de la transcription à laquelle elle se trouve. En effet, on trouve plutôt le rôle répresseur de Set2 quand celle-ci se trouve à l'initiation [25]. En revanche, Set2 associée à la RNAPII en élongation a un rôle important pour l'activation de la transcription car elle recrute des histones acétylases. D'ailleurs, les gènes hautement transcrits sont associés à une méthylation de H3 sur K36 [180].



d'après Hirose and Ohkuma 2007

FIGURE 2.4 – Les modifications dynamiques de la phosphorylation du CTD coordonnent le cycle de la RNAPII avec la maturation des pré-ARNm et la modification des histones. A. Les facteurs de transcription généraux forment un complexe avec la RNAPII_A au promoteur. La transcription commence avec la phosphorylation des Ser₅ du CTD par TFIIF. B. Peu après l'initiation, les enzymes de la coiffe sont recrutées par la RNAPII_O phosphorylée en Ser₅. Le complexe contenant l'histone méthyltransférase Set1 est aussi recruté et ajoute trois méthyles sur la lysine 4 des histones H3 (H3K4). La pause transcriptionnelle induite par DSIF/NELF est abolie par la phosphorylation du CTD réalisée par le P-TEFb. C. La RNAPII_O en élongation est de plus en plus phosphorylée par le P-TEFb sur les Ser₂ et s'associe à l'histone méthyltransférase Set2 qui triméthyle la lysine 36 des histones H3 (H3K36). La RNAPII_O permet aussi le recrutement de la machinerie d'épissage (SP) via un facteur inconnu (X) qui facilite l'excision des introns (ARN en rouge). D. Près de l'extrémité 3' du gène, les facteurs de maturation du 3' (PA) sont abondamment recrutés sur la RNAPII_O par une interaction directe entre Pcf11 et les Ser₂ phosphorylées. Après avoir transcrit le signal poly(A) AAUAAA, les facteurs de maturation du 3' se transfèrent sur l'ARN pour catalyser la coupure endonucléolytique (flèche noire) et induisent la terminaison de la transcription que l'on pense être aidée par les exonucléases 5'-3' Xrn2 et Pcf11. E. Après dissociation de l'ADN, la RNAPII_O est probablement déphosphorylée par l'action de phosphatases du CTD, FCP1 et Ssu72, avant d'être recyclée ou de réinitier une transcription.

Facteurs d'initiation et d'élongation

Parmi les protéines associées au CTD phosphorylé (phosphoCTD associated proteins ou PCAPs), on trouve plusieurs facteurs de transcription.

Le complexe Médiateur s'associe à la RNAPII pour stimuler l'initiation de la transcription par le CTD non phosphorylé. Lorsque TFIIF phosphoryle le CTD en Ser₅, il dissocie le médiateur [219, 202]. Un homologue humain du Médiateur de la levure, le CSRP interagit lui aussi avec le CTD et change de conformation à son contact. Un excès de peptides CTD inhibe le rôle activateur de CSRP [218].

2.1.3 Rôle du CTD dans des mécanismes liés à la maturation du transcrit

Les premières études qui ont porté sur le CTD en pratiquant des délétions d'heptapeptides ont montré qu'on obtenait plus de problèmes de maturation d'ARN que de transcription. Effectivement, la plupart des protéines recrutées par le CTD jouent un rôle dans la maturation de l'ARNm.

Le coiffage

L'ajout de la coiffe sur le 5' des ARNm est la première étape dans la maturation et consiste en une série d'activités enzymatiques : RNA-5' triphosphatase, guanylyltransferase et RNA (guanine-N7) méthyltransferase. Chez les métazoaires, l'enzyme de la coiffe est bifonctionnelle avec à la fois les activités RNA-5' triphosphatase et guanylyltransferase. Alors que chez la levure, l'enzyme de la coiffe est un hétérodimère d'une RNA-5' triphosphatase (Cet1) et d'une guanylyltransferase (Ceg1) [126]. La coiffe est ajoutée cotranscriptionnellement très peu de temps après l'initiation de la transcription quand l'ARN naissant mesure entre 20 et 30 nucléotides [125]. Les enzymes de la coiffe sont recrutées dès la transition initiation-élongation par le CTD phosphorylé [203, 155, 274]. Cette association se fait par la partie N-terminale du CTD [86] et requiert une phosphorylation de type S₂-S₅-S₂ [188]. Mce1, la guanylyltransférase chez les mammifères, est capable de se lier au CTD et d'être activée par le même motif. Elle peut se lier aux Ser₂P comme aux Ser₅P mais seule la Ser₅P l'active [127].

La maturation du 3' : clivage et polyadénylation

Chez les eucaryotes, l'extrémité 3' du pré-ARNm est formée en deux étapes : une coupure endonucléolytique 10 à 30 nucléotides en aval d'une séquence signal poly(A) très conservée, suivie de l'ajout d'un poly(A) à l'extrémité 3' de ce produit. Cette maturation du 3' requiert plusieurs facteurs protéiques comprenant le facteur de coupure/polyadénylation (CPSF), le facteur de stimulation de coupure (CstF), les facteurs de coupure CFIm et CFIIIm et la polymérase poly(A) (PAP) chez les mammifères, et le facteur de coupure-polyadénylation CPF, les facteurs de coupure CFIA et CF1B et la PAPA chez la levure [253]. Le phosphoCTD recrute beaucoup de facteurs de clivage et polyadénylation comme CstF, CFIA, Rna14, etc. [14, 126]. Chez la levure Pcf11, s'associe de façon spécifique aux Ser₂P par son domaine CID (140 acides aminés) [345, 184]. Son recrutement au CTD dissocie le complexe d'élongation [350]. S'il est muté, la RNAPII ne fait pas de terminaison et continue sur le gène [268].

Ces trois événements, coiffage, épissage (décrit dans la partie suivante) et coupure / polyadénylation, sont intimement liés à la transcription mais également entre eux. Par exemple, les protéines liant la coiffe en 5' du pré-ARNm interagissent aussi avec les facteurs d'épissage pour leur assemblage et favorisent la reconnaissance du premier site d'épissage en 5' [53, 179, 104]. Le cap-binding proteins complex (CBC) recrute la U1snRNP au site 5'ss. L'interaction de l'ARNm avec ces protéines de la coiffe est aussi importante pour son export [227]. Les facteurs d'épissage qui s'associent avec le dernier intron en 3', interagissent aussi avec les facteurs de polyadénylation et promeuvent la coupure en 3' et la polyadénylation [205]. Par exemple, la PAP stimule l'épissage par son interaction avec le facteur U2AF65 [315]. Inversement, la protéine U1A de l'épissage inhibe la fixation de la PAP [114].

Ces interactions jouent un rôle important en garantissant la reconnaissance des sites corrects d'épissage et leur engagement dans la réaction d'épissage, ainsi que le timing, la précision et l'efficacité du coiffage et de la polyadénylation [197]. Le couplage entre transcription et maturation peut influencer les réactions de maturation selon trois modes [20]. D'abord, comme on vient de le voir, en fonction de la localisation car le couplage permet d'augmenter localement, là où on en a besoin, les facteurs de maturation. Deuxièmement, la cinétique : la vitesse de transcription peut avoir des conséquences sur le repliement de l'ARNm naissant et son assemblage avec les facteurs

de maturation (la vitesse d'élongation influe sur le choix des sites d'épissage [67]). Plus la RNA-Pol II est lente, plus les sites faibles d'épissage peuvent être sélectionnés. Enfin, l'allostérie car les contacts entre le complexe d'élongation et les facteurs de maturation peuvent activer ou inhiber ces derniers de façon allostérique (les guanylyltransférases du CBC sont de plus en plus stimulées avec le nombre croissant de Ser₅P sur le CTD [127]).

2.2 L'épissage cotranscriptionnel

2.2.1 Introduction

L'épissage, une idée intuitive

L'épissage cotranscriptionnel est une idée qui a surgi peu de temps après la découverte de l'épissage car chez l'homme les gènes sont souvent très longs (46 380 pb en moyenne par transcrit, le plus long mesure $2 \cdot 10^6$ pb, et la médiane se situe à 12690 pb) avec en moyenne 8,8 exons [173] et pour certains, un cycle cellulaire ne suffit pas à leur transcription (l'ARNm de la dystrophine mesure environ 8 kb et sa séquence codante environ 6Mb [273]). Difficile alors de concevoir que l'épissage attende la fin de la transcription de ces immenses gènes. D'une façon générale, toute la maturation des ARNm naissants peut se faire de façon cotranscriptionnelle, en partie grâce au CTD, en partie par leur synergie mutuelle.

Deux événements : même lieu, même moment

Beaucoup de systèmes ont été mis à contribution pour comprendre le mécanisme de l'épissage. Des embryons de Drosophile aux noyaux de cellules de mammifères, l'existence d'un mécanisme d'excision des introns est apparue comme un système universel chez les eucaryotes. Dès la fin des années 80, alors que l'épissage est établi depuis peu, des équipes obtiennent des résultats *in vivo* démontrant l'existence d'un épissage cotranscriptionnel [23, 336, 348]. Dans ces expériences, on met en évidence la simultanéité de l'épissage et de la polyadénylation avec la transcription par des méthodes de colocalisation et copurification. Quelques années plus tard, le CTD de la RNA polymérase II est directement impliqué dans l'interdépendance entre épissage et polyadénylation avec la transcription [216, 204]. On a déjà vu que le coiffage et les étapes de maturation du 3' de

l'ARNm étaient co-transcriptionnels, au moins pour le recrutement de ces facteurs. L'épissage est aussi cotranscriptionnel et plusieurs travaux démontrent qu'il s'agit plutôt d'une règle que de cas ponctuels [4, 240]. Néanmoins l'épissage post-transcriptionnel existe aussi et le choix de l'un ou de l'autre est probablement dépendant de différents facteurs (du gène, de sa localisation génomique, nucléaire, etc.).

2.2.2 Influence de la transcription sur l'épissage

Des facteurs d'épissage recrutés par la machinerie transcriptionnelle et rôle du CTD

Tout comme les enzymes de coiffage, les facteurs d'épissage sont recrutés sur un gène comportant des introns et transcriptionnellement actif [210, 243]. Leur présence est directement lié à l'épissage car un gène sans introns est moins enrichi en facteurs d'épissage [213, 243]. Par exemple, U2AF65, U170k et U5-116k ne s'accumulent pas sur un gène sans intron et ne sont plus enrichis sur un gène privé d'introns [159, 186]. Nombre d'associations ont été mises en évidence entre le complexe transcriptionnel et des facteurs d'épissage. Dans sa revue, Eric Allemand en présente une liste non exhaustive [4]. Parmi ces facteurs, on retiendra U2AF65 (Mud2), SF1 (sub2), des sous-unités SF3 de la snRNP U2 ou encore SF2/ASF ou Srp20. On voit qu'il s'agit aussi bien de facteurs directement impliqués dans le splicéosome comme des facteurs un peu plus éloignés. Comment ces facteurs sont-ils recrutés au site de transcription? Là encore, la RNAPII et son CTD sont impliqués. Dans un système *in vitro*, la *RNAPII_O* immunoprécipite le splicéosome (U1 et U2snRNP, protéines associées Prp5, Prp28, TAT-SF1) et des facteurs de maturation du 3' sur un ARN naissant synthétique court comprenant uniquement le 5'ss [137]. Inversement, des immunoprécipitations de protéines SR co-immunoprécipitent de la *RNAPII_O* [216]. D'autres complexes constitués de composants du splicéosome et de facteurs d'épissage avec la RNAPII tels que la snRNP U1 (ARN U1, protéines U1A et U170k) ou des protéines SR telles que SRp20 sont également mis en évidence *in vivo* [270, 65]. Peu ou prou, il semble que l'on peut trouver toutes les protéines de l'épissage associées au CTD : SC35, SF2/ASF, SRp40, Prp40 (FBP11 chez les mammifères [95]), U2AF65 [146, 210, 214, 310].

Effets de la transcription sur l'épissage

Quel est l'impact de cet épissage cotranscriptionnel ? Il y a plusieurs réponses à cela. Tout d'abord, cette synergie jouerait un rôle sur l'efficacité de la réaction d'épissage. Des expériences *in vitro* montrent en effet l'importance de la RNAPII dans l'assemblage du splicéosome et dans l'efficacité de l'épissage. Pour une même quantité d'ARN transcrits par la RNAP T7 (ne possède pas de CTD) ou par la RNAPII, le rapport ARNm épissés / pré-ARNm sera plus grand avec la RNAPII [64]. Le rôle du CTD semble important dans ces mécanismes puisqu'un CTD de 52 répétitions ajouté à un milieu réactionnel augmente le nombre de transcrits épissés. Et lorsqu'on enlève la RNAPII d'un milieu réactionnel, on constate une baisse de la quantité de transcrits épissés. Quantité qui ré-augmente lorsqu'on rajoute du CTD [347]. *In vivo*, on constate des comportements similaires : un rapport ARNm épissés/pré-ARNm plus important lorsqu'il y a une augmentation de la quantité de RNAPII sur le gène [26]. Pour expliquer cet effet, les avis divergent. Certains avancent que l'épissage se fait plus rapidement lorsqu'il est couplé à la transcription [64], d'autres mettent en évidence que la quantité de pré-ARNms associés aux splicéosomes est plus importante et donc l'efficacité est augmentée [122]. Quoi qu'il en soit, le CTD de la RNAPII est un activateur de cette efficacité *in vivo* car une RNAPII avec un CTD de 5 répétitions seulement continue de transcrire (beaucoup moins que son homologue sauvage) mais les ARNm ne sont plus épissés [96, 204]. De façon intéressante, l'épissage post-transcriptionnel, découplé de la transcription n'est pas affecté. Et de la même façon, si on perd la RNAPII ou si la transcription est inhibée, l'épissage découplé n'est pas affecté alors que l'épissage cotranscriptionnel est perdu [24]. De plus, les facteurs d'épissage recrutés sur un site transcrit par une RNAPII sauvage (U2snRNP, SC35, SF2/ASF, SRp40), sont absents sur un site transcrit par une RNAPII Δ CTD [210]. Dans chacun de ces travaux, c'est le CTD phosphorylé qui est responsable du couplage de l'épissage avec la transcription et de la liaison des protéines de l'épissage par leur domaine SR [208].

Un deuxième corollaire à la synergie transcription-épissage est la régulation de l'épissage alternatif. Cela se traduit sous différentes manières, à commencer par la définition de l'exon. Plusieurs travaux montrent un épissage cotranscriptionnel dans lequel les exons sont reconnus et restent associés au complexe transcriptionnel afin de permettre l'épissage et cela sans que l'intron ait besoin d'être continu [73]. Ainsi, une coupure peut survenir dans l'intron sans que l'épissage soit affecté.

Cela a un effet direct sur l'épissage alternatif car si une coupure survient entre la séquence régulatrice d'épissage (SRE) et l'exon qu'il régule, l'exon n'est pas inclus dans l'ARNm [111]. Le CTD joue là aussi un rôle important car *in vitro*, ajouté à un milieu réactionnel, il n'augmente l'efficacité de l'épissage que si les sites d'épissage encadrant les exons sont corrects (3'ss et 5'ss), et n'a aucun effet lorsqu'un des sites d'épissage encadrant un intron est absent (5'ss et 3'ss) [347].

Une autre façon de moduler l'épissage alternatif par la transcription est d'influer sur la sélection des sites d'épissage. Trois exemples sont représentés dans le schéma B de la figure 2.5. Il existe en effet des sites donneurs et accepteurs « forts » et des sites « faibles ». Comme la RNAPII et ses facteurs de transcriptions s'associent à des facteurs d'épissage capables de se lier à des SRE, en fonction du facteur recruté, le patron d'épissage ne va pas être le même (a). Par exemple, le CTD recrute directement ou indirectement la protéine SRp20, protéine SR favorisant l'exclusion de l'intron [194, 68]. Cette sélection de sites dépend aussi de leur localisation par rapport au promoteur (b). En effet, si le choix s'offre entre deux sites 3'ss pour faire la paire avec le premier 5'ss, c'est celui qui est le plus proche du promoteur qui a la préférence [122]. De cette façon, plus l'intron est long et plus l'épissage alternatif est favorisé [264]. Enfin, la vitesse d'élongation va aussi jouer sur cette sélection de sites d'épissage. Plus la RNAPII est lente et plus les exons sont inclus dans l'ARNm (c) [67]. On pourrait alors supposer que les facteurs d'épissage pourraient accélérer ou ralentir la RNAPII pour faire tel ou tel patron d'épissage. Une autre façon de moduler la vitesse de la RNAPII est de lui faire faire des pauses. Le gène *FGFR2* subit un épissage alternatif qui conduit à deux formes mutuellement exclusives et typiques d'un type cellulaire [265]. Lorsque l'élément *MAZ4*, capable de faire pauser la RNAPII, est inséré entre l'exon alternativement épissé et les ISS qui le répriment, l'inclusion de cet exon augmente. De façon surprenante, la pause de la RNAPII permet l'inclusion de l'exon alors qu'augmenter la distance entre l'exon et ses ISS ne modifie pas le patron d'épissage.

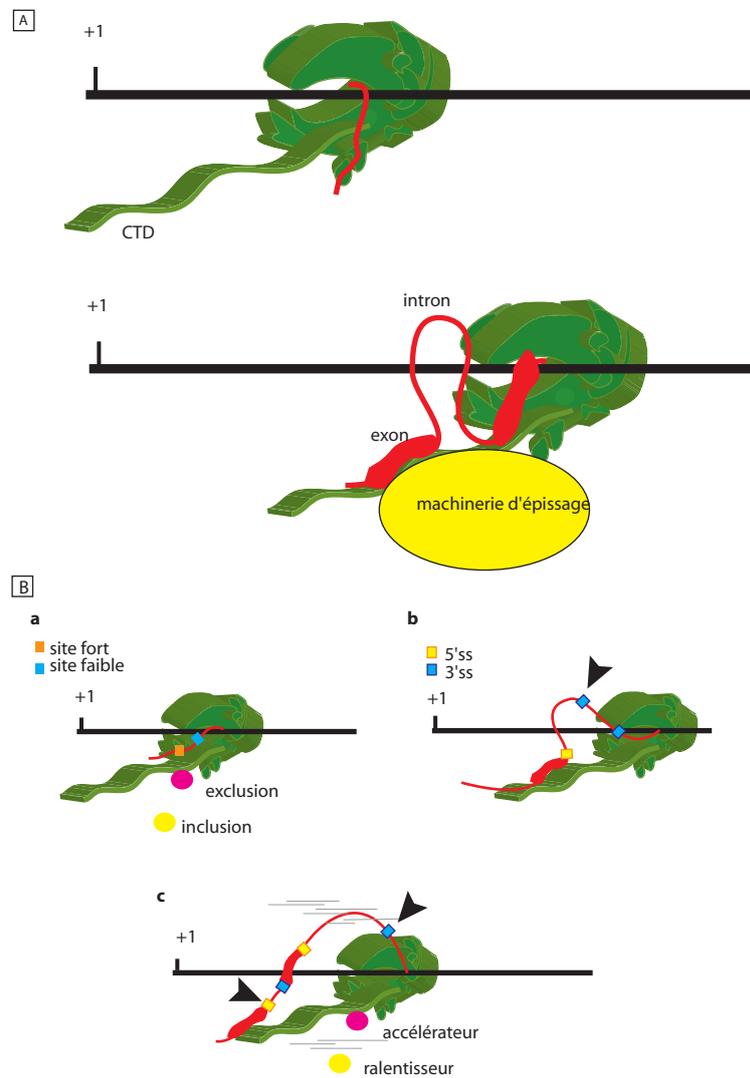


FIGURE 2.5 – **Influence de la RNAPII sur l'épissage.** **A.** Dans la définition par l'exon, les exons sont associés au CTD alors que les introns forment des boucles. Ainsi, les sites d'épissage 5'ss et 3'ss sont les uns à côté des autres pour permettre une meilleure efficacité. **B.** L'épissage alternatif est dépendant de plusieurs paramètres. **Ba.** Les protéines des SRE. Si, pour un même site donneur ou accepteur, il y a le choix entre un site fort et un site faible, la sélection se fera en fonction du type de protéines qui se liera aux SRE. Parmi toutes les protéines susceptibles de se lier à ces séquences, si l'un d'elle est déjà liée au CTD, elle sera la première à se fixer à l'ARN et donc à influencer ce choix. **Bb.** La distance vis-à-vis du promoteur. S'il y a compétition entre deux sites 3'ss pour un site 5'ss, le site 3'ss le plus proche du promoteur sera le plus souvent choisi. **Bc.** La vitesse d'élongation. Des facteurs d'épissage peuvent accélérer ou ralentir la RNAPII (voire provoquer une pause) pour exclure ou inclure des exons en sélectionnant des sites forts ou faibles.

Influence de l'épissage sur la transcription

Si la transcription exerce une influence sur l'épissage, on peut s'attendre au corollaire. Il existe une règle totalement empirique selon laquelle un gène auquel on enlève des introns est moins transcrit. Cette diminution du nombre de transcrits n'est pas liée à la stabilité des ARNm. Des études ont été menées *in vivo* chez un animal transgénique et plusieurs tissus ont été étudiés avec les mêmes conclusions. Étonnamment, lorsque l'expérience est effectuée sur des lignées cellulaires stables, l'effet disparaît [33]. Réciproquement, si on inclut des introns hétérologues provenant d'autres gènes, on augmente la transcription du gène. L'effet est d'autant plus important que l'intron est placé en 5' du gène. Plus on s'en éloigne, moins l'effet est visible. Cet effet semble dépendant du type de promoteur. Un promoteur CMV paraît moins sensible à la présence d'un intron [238]. Ces constatations sont de prime importance dans la thérapie génique et une équipe a testé la transcription du gène de la GFP et du gène Cre, inséré stablement dans des lignées cellulaires et chez l'animal, en ajoutant ou pas un court intron en 5'. La transcription est augmentée trente fois dans les lignées possédant un intron comparées aux lignées avec la construction sauvage. Un résultat similaire est obtenu chez l'animal [170]. Aujourd'hui, seules des hypothèses permettent d'expliquer ces résultats. L'intron pourrait jouer un rôle dans le réarrangement de la chromatine ou dans la stabilisation du pré-ARNm au cours de sa transcription. On sait, d'autre part, qu'il existe un mécanisme de répression de l'expression des gènes de la cellule vis-à-vis d'ADN étranger. Chez les plantes, c'est un phénomène courant [329] mais on le trouve aussi ailleurs. Méthylations *de novo* ou ARN interférence ont été observés sur des gènes rétroviraux ou des transgènes chez la souris [22]. Une structure de forme exon-intron-exon pourrait paraître plus « familière » pour la cellule et donc être acceptée plus facilement. Cette « reconnaissance » pourrait se faire via la machinerie d'épissage et diminuerait la répression de l'expression des gènes. D'autre part, la présence d'un intron recrute des facteurs d'épissage qui sont importants ensuite dans l'export des ARNm [317]. Cela permet un export plus rapide. UAP56 en est l'exemple le plus flagrant puisqu'il recrute le facteur d'export Aly/REF de façon ATP-dépendante [176, 158]. Sa présence est d'ailleurs requise sur des gènes sans introns pour faciliter leur export [300]. Les gènes possédant un intron recrutent plus facilement ces facteurs et pourraient donc être plus facilement exportés et traduits. En retour, les facteurs d'export pourraient empêcher une dégradation prématurée de

ces ARNs.

Le premier intron peut augmenter la quantité de transcrits mais si celui-ci est muté au site d'épissage 5'ss ou sur le 3'ss ou encore sur le polypyrimidine tract, l'épissage est aboli mais on diminue aussi le nombre de pré-ARNms transcrits [88]. L'intégrité des sites d'épissage affecte donc aussi la transcription.

Il existe aussi des protéines de l'épissage qui ont un rôle dans la transcription. Par exemple, Cus2 chez la levure, (Tat-SF1 chez l'homme), joue un rôle très important dans les réarrangements de l'ARN U2 entre sa fixation au point de branchement et la première réaction catalytique. Il intervient avec l'ATPase Prp5 dans le changement conformationnel de la structure tige II qui doit passer de la structure IIa à la structure IIc [342, 245]. Cus2 et Prp5 sont tous deux impliqués dans la fidélité de la première réaction catalytique. L'orthologue humain de Cus2 est la protéine Tat-SF1. Tat-SF1 possède la même propriété que Cus2 à lier l'ARN U2. Cependant sa découverte n'a rien à voir avec l'épissage puisqu'il a été mis en évidence pour son rôle dans la transcription dépendante du complexe Tat-TAR du VIH [353]. En plus, d'activer l'élongation du VIH, Tat-SF1 a un rôle plus général puisqu'il s'associe avec le facteur d'élongation cycline T1 du complexe P-TEFb [88]. Si aucun lien direct n'a été fait avec son rôle dans l'épissage, on remarque que, traité à la nucléase micrococcalle, Tat-SF1 n'a plus d'effet sur la transcription. Les auteurs proposent que Tat-SF1 soit amené par le P-TEFb sur le gène et recrute ensuite les facteurs d'épissage.

L'épissage peut aussi inhiber la transcription. Par la méthode SELEX, un ARN court a été découvert, comportant une séquence très proche de la séquence 5'ss reconnue par le snARN U1 [302]. Cet ARN est capable d'inhiber l'activité de l'ARN polymérase II. Cette inhibition n'est possible qu'avant la formation du complexe de pré-initiation. L'auteur postule que la snRNP U1 est associée à la RNAPII. Si trop d'ARNs naissants contenant un 5'ss sont transcrits et associés au complexe RNAPII-U1snRNP, ceci est perçu comme le signal qu'il y a trop d'ARNs et l'initiation de la transcription est abolie. Un tel complexe RNAPII-U1snRNP pourrait faire passer la machinerie transcriptionnelle pour un pseudo complexe d'élongation (puisque'il est lié à de l'ARN et à des facteurs d'épissage) même s'il n'est pas engagé sur l'ADN. Et de ce fait, cela pourrait empêcher l'activation de la RNAPII pour l'initiation de la transcription. D'autres auteurs ont remarqué un effet positif du snARN U1 sur la transcription et notamment sur la réinitiation de la transcription. En effet, lorsque l'ARN

U1 est présent, on constate *in vitro* une réinitiation plus efficace. Les auteurs proposent que cet effet de U1 se fasse par l'intermédiaire de TFIIH, par l'association de U1 avec la cycline H [166]. Cette association aurait pour conséquence l'augmentation de l'activité kinase de cdk7 [232].

Positives ou négatives, les interactions entre transcription et épissage sont nombreuses et la liste présentée ici n'est certainement pas exhaustive mais elle prétend refléter l'éventail des possibilités connues à l'heure actuelle.

Chapitre 3

Les ARNs non codants

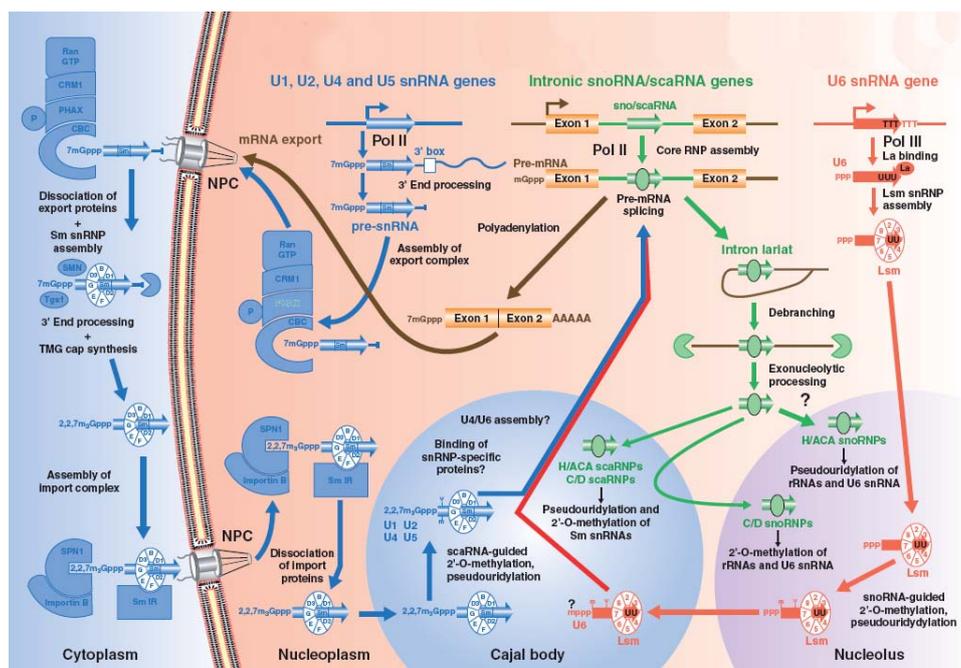
Nous avons vu dans les deux précédentes parties comment fonctionnaient la transcription et l'épissage et comment était régulé l'épissage cotranscriptionnel. On a pu constater que ces mécanismes font intervenir de nombreuses protéines aux activités variées. Outre les protéines, d'autres molécules jouent des rôles importants dans ces événements : les ARNs non codants (ncARNs). Ces ARNs se caractérisent par l'absence de phase codante. Ils peuvent être uniques et produits à partir d'un gène spécifique ou bien provenir de séquences intergéniques transcrites ou bien encore de l'ARNm découpé avec précision d'un gène codant pour une protéine. Ils ont des rôles variés : de reconnaissance, structuraux, activateurs d'enzymes. Mon sujet de thèse avait comme point de départ des travaux concernant des ARNs non codants. Par la suite, j'ai découvert un rôle potentiel de l'ARN U1 de l'épissage avec la RNAPII. Voici donc une introduction sur cette classe particulière d'ARNs que sont les ARNs non codants.

3.1 Les ARNs impliqués dans des processus généraux

Historiquement, les premiers rôles attribués aux ARNs non codants sont des rôles à la fois structuraux et de reconnaissance. C'est le cas de l'épissage, comme nous avons pu le voir dans la première partie. C'est également le cas de la traduction avec les ARNs ribosomiques (ARNr) et les ARNs de transfert (ARNt). Pour leur permettre de s'assembler et d'exercer leur fonction, ces ARNs non codants doivent subir des modifications préalables. Certaines de ces modifications sont réa-

lisées par d'autres ARNs non codants tels que les snoARNs (small nucleolar ARNs) ou les scaARNs (small Cajal bodies ARNs). Ces ARNs font partie de ribonucléoprotéines dans lesquels ils jouent les rôles de guide pour la réaction de pseudouridylation ou la 2'-O-méthylation en fonction de leur séquence. Les ARNs responsables de la pseudouridylation font partie des H/ACA snoRNPs, ceux de la 2'-O-méthylation, des C/D snoRNPs [150]. Les cibles directes de ces ARNs sont les ARNr, les ARNt et les UsnARNs. Des snoARNs dont la séquence n'a pas été retrouvée dans les ARNr, t ou UsnARNs ont été mis en évidence et certains sont impliqués dans la régulation de l'épissage alternatif de certains gènes. Leur perte entraînerait le syndrome de Prader-Willi. Un exemple décrit ci-dessous la biogenèse des UsnARNs ayant à subir des modifications par les snoARNs et les scaARNs. Les Uridine-rich small nuclear ribonucleoproteins sont au nombre de 5, chacun composé d'un snARN, de 7 protéines Sm et de protéines spécifiques à l'UsnRNP. Tous les UsnARNs sont transcrits par la RNAPII, sauf U6 qui est synthétisé par la RNAPIII. Il existe également d'autres UsnRNAs connus pour de l'épissage secondaire (U11, U12...) de certains introns mais nous ne les développerons pas dans ce mémoire. Les 5 snARNs majeurs ont une coiffe monométhylée, m7GpppG (m7G) [151]. Après leur synthèse par la RNAPII, U1, U2, U4 et U5 sont exportés dans le cytoplasme où commence l'assemblage des UsnRNP (figure 3.1). Cette coiffe m7G est indispensable à leur export. En effet, le complexe d'export est composé du m7G cap binding complexe (CBC), du récepteur d'export (CRM1/Xpo1), de RanGTP et de PHAX (phosphorylated adaptor for RNA export) [332]. Une fois dans le cytoplasme, l'UsnARN se dissocie de ces facteurs d'export, et s'associe avec 7 protéines Sm : B/B' (B et B' sont deux protéines pouvant se remplacer), D3, D2, D1, E, F, G qui s'assemblent en anneau autour du domaine A de l'UsnARN composé d'un site Sm encadré par deux tiges-boucles. Les protéines Sm et l'UsnARN forment ainsi le cœur du complexe snRNP [242]. Toujours dans le cytoplasme, la m7G est convertie en m3G (2,2,7-tri-méthylée) et les UsnRNAs subissent une maturation du 3' (coupure pour leur donner leur taille standard). La m3G permet la reconnaissance du UsnARN par la snuportin-1 et son import dans le noyau. Chez les vertébrés, SMN (Survival of Motor Neurons, le gène de la maladie Spinal Muscular Atrophy) facilite l'assemblage des UsnRNPs. SMN participe aux trois événements de maturation dans le cytoplasme : l'assemblage des protéines Sm, l'hyperméthylation de la coiffe et la maturation de l'extrémité 3' des UsnARNs. De retour dans le noyau, les UsnRNAs sont encore modifiés, sur

plusieurs positions (pseudouridylation et 2'-O-méthylation) dans des structures nommées Cajal bodies par les scaARNs pour les UsnARNs U1, U2, U4 et U5 [151]. Le snARN U6 suit un chemin un peu différent. Après sa transcription, il est aussitôt associé à la protéine La et dirigé vers les nucléoles où il est associé à un anneau de protéines Lsm par son extrémité 3' (Like Sm). Il subit les mêmes modifications que les autres UsnARNs par les snoARNs [242]. Dans les étapes ultérieures avant l'assemblage du spliceosome, la snRNP U6 interagit avec la snRNP U4, via une hybridation entre les snARNs U4 et U6. Les protéines spécifiques des UsnRNPs sont assemblées dans le noyau, possiblement dans les Cajal Bodies.



d'après Kiss 2004

FIGURE 3.1 – Biogenèse des petites RNPs. Présentation de la biogenèse des snARNs U1, U2, U4, U5 (transcription RNAPII), du snARN U6 (transcription RNAPIII) et des sno/scaARNs (transcription RNAPII). Les snARNs sont impliqués dans l'épissage et font partie du spliceosome. Les snoARNs se trouvent dans les nucléoles et les scaARNs dans les Cajal bodies. Les snoARNs sont responsables des modifications de l'ARN U6 et les scaARNs des 4 autres snARNs de l'épissage.

3.2 Les ARNs non codants impliqués dans la régulation de l'expression génique

D'autres ARNs non codants sont directement impliqués dans la régulation de l'expression des gènes soit en activant ou réprimant leur transcription, soit en dégradant les ARNm. Dans la première catégorie, on trouve des ARNs uniques aux fonctions bien particulières comme l'ARN 7SK qui est inhibiteur général du P-TEFb et des ARNs plus spécifiques comme l'ARN Xist ou SRA. La deuxième catégorie regroupe les ARNs silenciers que l'on peut diviser en trois classes : small interfering RNA (siARN), microRNA (miARN), piwiRNA (piRNA). Les paragraphes suivants prennent un exemple de chaque catégorie.

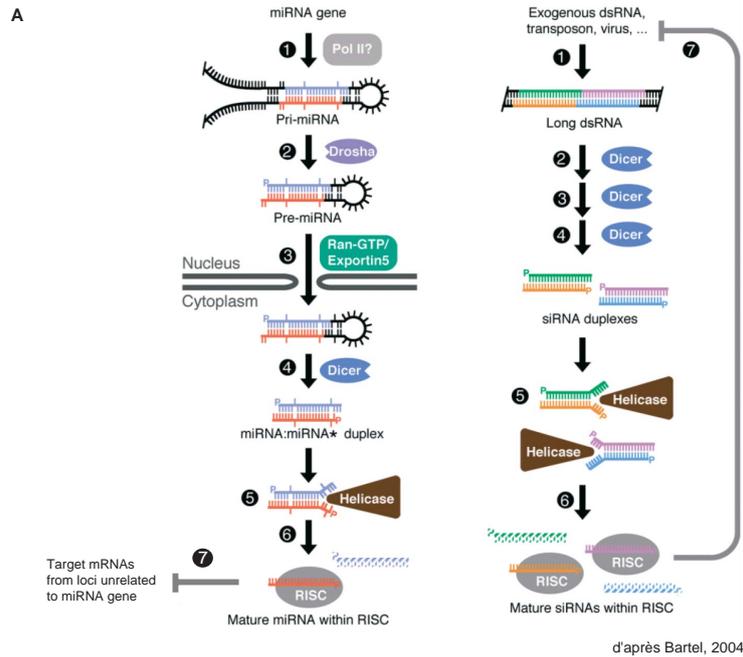
3.2.1 L'ARN non-codant SRA

Cet ARN non-codant de 883 nucléotides est un co-activateur des récepteurs stéroïdes [175]. Sa surexpression accroît la transactivation de la transcription par les récepteurs stéroïdes. Inversement lorsque le SRA est absent, il y a suppression de l'activité des récepteurs androgènes dans les cellules du cancer de la prostate [165]. L'ARN précurseur du SRA soumis à l'épissage alternatif peut donner la protéine SRAP (SRA protein) [130]. Contrairement à SRA, la protéine SRAP réprime l'activité des récepteurs œstrogènes lorsqu'elle est surexprimée. Les proportions de SRA et de SRAP sont variables dans les différentes lignées de cancer du sein et pourraient avoir un lien avec leurs phénotypes. SRA interagit directement ou indirectement avec plusieurs protéines sans que l'on sache comment il agit sur la co-activation. Ces protéines sont impliquées de près ou de loin dans la régulation des récepteurs stéroïdes (SRC-1, p68/72 RNA hélicase, SHARP, SLIRP, MAPK) [283, 328, 69, 119, 39]. Son activité est également dépendante de son état de pseudouridylation (réalisée par les pseudouridines synthases Pus3p et Pus1p). Une mutation ponctuelle sur le résidu U206 entraîne le renversement du rôle de SRA en co-répresseur dominant négatif [351, 352].

3.2.2 Les ARN silenciers

La répression de l'expression génique a deux fonctions majeures : la défense contre les virus [94], très étudiée chez la Drosophile et la répression de l'expression génique. La défense contre

les virus est assurée par les siRNAs (small interfering RNAs) qui sont des ARNs double-brins de ≈ 21 nucléotides de long, modifiés en 3', provenant de longs transcrits exogènes double-brin (dsRNAs). Les micro-ARNs répriment l'expression de transcrits RNAPII; $\approx 30\%$ de gènes humains seraient régulés par des miRNAs [251]. Ils sont double-brin également, longs de 21 nucléotides et proviennent de la coupure de transcrits structurés en épingle à cheveux (figure 3.2 A). Ces deux catégories d'ARNs proviennent de transcrits plus long coupés par des ribonucléase III Droscha et Dicer. L'homme, la souris et *Caenorhabditis elegans* possèdent un seul gène Dicer. Chez la drosophile, il existe deux gènes : Dicer-2 pour les siRNAs et Dicer-1 pour les microRNAs. Une fois coupés et déshybridés, ces petits ARNs s'associent à des protéines de la famille Argonaute pour former les complexes RISC et MiRNP respectivement de répression de l'expression génique [141]. La protéine Argonaute possède deux domaines principaux : un domaine de liaison à l'ARN PAZ en N-terminal et un domaine Piwi RNase-like en C-terminal. Comme pour Dicer, chez les mammifères, les membres de la famille AGO contribuent aux deux voies miARN et siARN alors que chez la Drosophile, Ago1 s'occupe des miARNs et Ago2 des siARNs. Dans ces complexes ribonucléoprotéiques, ces ARNs vont reconnaître les transcrits cibles et permettre leur dégradation [16].



d'après Bartel, 2004

B

	<i>Arabidopsis thaliana</i>	<i>Caenorhabditis elegans</i>	<i>Drosophila melanogaster</i>	<i>Mus musculus</i>	<i>Homo sapiens</i>
Drosha		drsh-1	Drosha	Rnase	Rnase
Dicer	DCL1 [miRNA] DCL2 [antiviral siRNA] DCL3 [endogenous siRNA, 24nt] DCL4 [siRNA, 21nt]	dcr-1 [miRNA, siRNA]	Dcr-1 [miRNA] Dcr-2 [siRNA]	Dicer [miRNA, siRNA]	Dicer [miRNA, siRNA]
Argonaute					
Ago subfamily	AGO1 [miRNA, endogenous siRNA] AGO4 [endogenous siRNA, DNA methylation] AGO2,3,5-10	alg-1 [miRNA] alg-2 [miRNA] T23D8.7 ZK757.3 T22B3.2	Ago1 [miRNA] Ago2 [siRNA]	Ago1-4 Ago5 [siRNA, miRNA]	Ago1,2,3 Ago4 [siRNA, miRNA]
Piwi subfamily		ergo-1 [exogenous, siRNA] D2030.6 C01G5.2	Piwi Aub Ago3	Miwi Mili Miw2	Hiwi Hili Piwil3 Hiwi2

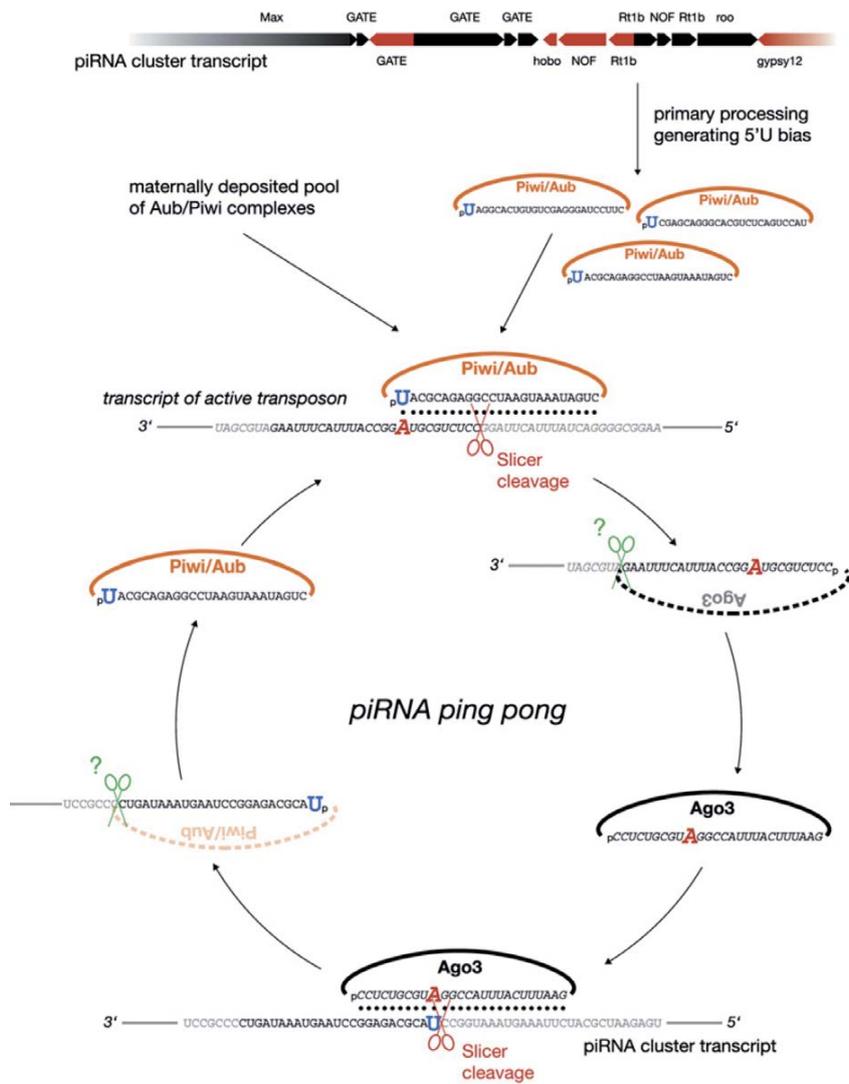
Code couleur
 Related to miRNA
 Related to siRNA
 Related to both, or unknown

d'après Kawaji and Hayashizaki, 2008

FIGURE 3.2 – Les ARNs silencier. A. Biogenèse des miARN chez les métazoaires. Le précurseur est synthétisé par la RNAPII le plus souvent et s'hybride sur lui-même, c'est le pri-miARN. Les extrémités sont ensuite coupées par l'enzyme Droscha et aboutissent au pré-miARN. Exporté dans le cytoplasme, la boucle du pré-miARN est coupée par l'enzyme Dicer. Une hélicase vient ensuite séparer les deux brins et le miARN est incorporé au complexe RISC. Les gènes ciblés ne sont pas les mêmes que ceux qui ont produit les miARNs. Le brin opposé est beaucoup moins représenté dans ces complexes. **Biogenèse des siARN.** Le précurseur est un ARN double brin, provenant d'un ARN étranger. Il est coupé par Dicer en petits morceaux de 21 nucléotides et les duplex sont séparés par une hélicase. Chacun des deux brins peut être ensuite incorporé dans un complexe RISC. **B. Protéines associées aux ARN silenciers.** Les protéines associées aux ARNs silenciers peuvent être rangées en familles et sous-familles. Les voies sont indiquées entre parenthèses.

Une nouvelle classe de petits ARNs a été mise en évidence : les repeat-associated small interfering RNAs ou rasi-RNAs. Ils forment des complexes avec les protéines Piwi et Aubergine de la famille Argonaute plutôt que Ago-1 ou Ago-2 comme les microARNs et siARNs (figure 3.2 B). Les protéines Piwi sont impliquées dans l'intégrité des lignées germinales. On les trouve dans la lignée germinale de la Drosophile, de *Caenorhabditis elegans* ainsi que chez *Arabidopsis thaliana* [185]. Ils se différencient des autres siARNs par le choix d'un brin préférentiel, seul un brin sert à faire les rasiARN alors que pour les siARNs, les deux brins sont utilisés de façon aléatoire. Ils répriment l'expression d'éléments répétés (SINes, LINes, transposons, etc.) et contrôlent leur mobilisation dans la lignée germinale de Drosophile et de *Caenorhabditis elegans* [287, 313]. Ils pourraient également réguler l'expression en modifiant la chromatine, notamment chez les plantes [141]. En plus des sous-familles Piwi et Aubergine, il existe la famille Argonaute-3. Ces protéines ont un profil d'expression similaire aux protéines Piwi et Aubergine et s'associent à des ARNs de 23 à 26 nt. Ces trois sous-familles d'Argonaute sont souvent associées à des rasiARNs. Ils mesurent entre 24 et 29 nucléotides et ne requièrent ni Dicer-1 ni Dicer-2 pour leur coupure. Parmi les ARNs associés à ces protéines Piwi/Aubergine et Ago3. Une observation a été faite sur le biais du brin associé. Les piARNs (associés à Piwi, comprenant majoritairement des rasiARNs) sont des brins antisens alors que les ARNs associés à Ago3 sont sens. Ces deux complexes fonctionnent, en fait, en synergie pour permettre la répression des transposons. Les protéines Piwi reconnaissent les piARNs (antisens) possédant un biais uridine en 5' (par la protéine Slicer) et s'hybrident avec des transposons inactifs. Ces complexes vont ensuite s'hybrider avec les ARNs de transposons et les couper en 5' pour former de nouveaux complexes Ago3 avec un brin sens. Ce complexe va alors chercher une nouvelle cible, probablement un pré-ARNm d'un locus contrôle qui contient des séquences antisens de transposon et les couper en 3'. Cela va permettre de créer de nouveaux piARNs qui pourront être chargés sur des protéines Piwi (figure 3.3). Le système ne cesse ainsi de s'amplifier au cours des cycles, ce qui explique la persistance du système dans les lignées germinales de Drosophile et de mammifères. Il reste encore des points à éclaircir, notamment la provenance originels des piARNs (ARNs maternels ?) ainsi que l'origine de la coupure en 3' [32, 113]. Outre les siARNs exogènes, des endo-siARNs ont été découverts chez les animaux. Déjà connus chez les plantes, des siARNs endogènes ont été mis en évidence dans des cellules en culture de Drosophile ainsi que

dans les têtes de *Drosophile*. De même taille que les siARNs exogènes et modifiés en 3', ces ARNs proviennent à 80% et 33% de transposons respectivement dans les cellules somatiques en culture et dans les têtes de *Drosophile*. Les 67% restants dans les têtes de *Drosophile* proviennent à part égale de régions codantes et de régions intergéniques. Pour le moment, on ne connaît pas leur biogenèse (ils ne sont pas complètement sensibles à une délétion de Dicer-2), ni leur fonction [97] mais les auteurs proposent un rôle similaire aux piARNs des cellules germinales.



d'après Brennecke, Aravin et al. 2007

FIGURE 3.3 – **Le modèle Ping-Pong des piARNs.** Ce schéma illustre la boucle d'amplification basée sur les complexes Piwi/Aub, et Ago3, le groupe de piARNs et les transcrits des transposons actifs. Les coupures de nucléotides sont représentées avec des ciseaux. Les sources potentielles de piARNs primaires sont les transcrits piARNs et les complexes piARNs hérités de la mère.

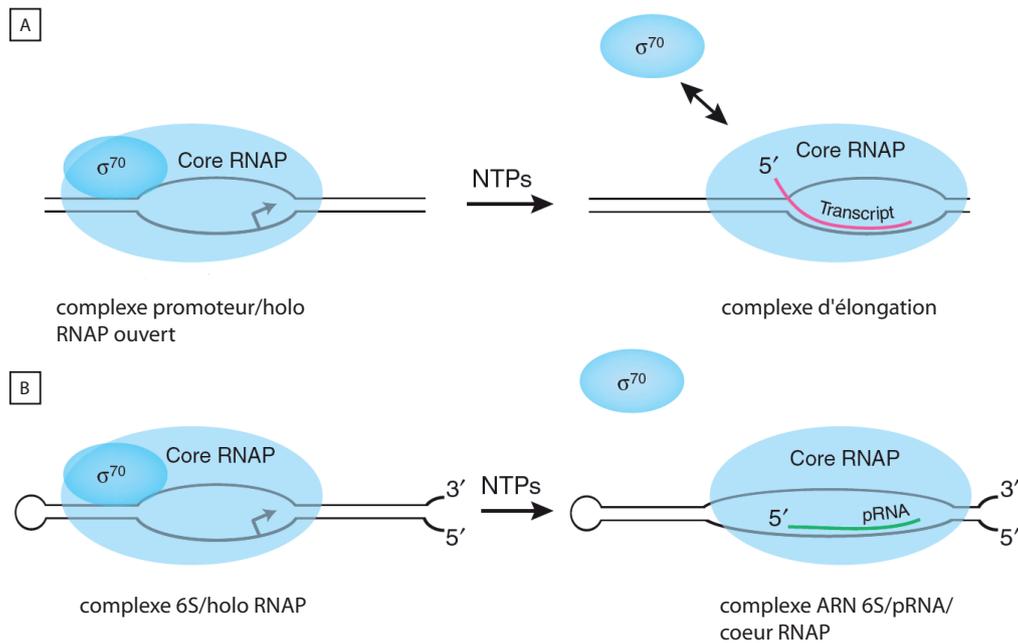
3.3 Les ncARNs qui régulent l'activité de l'ARN polymérase II

Parmi les ARNs non-codants, il en existe dont la cible directe est l'ARN polymérase II. Deux cas ont été particulièrement mis en évidence : l'ARN 6S chez la bactérie et l'ARN B2 chez la souris.

3.3.1 L'ARN 6S

Ce fut l'un des premiers ARNs non-codants découverts [124]. Chez *E. coli*, cet ARN a la particularité de voir son nombre de copies multiplié par 10 lors de la phase stationnaire. Il est codé par le gène *ssrS*. Celui-ci forme un opéron avec le gène *ygfA* [129, 147]. L'ARN 6S s'associe à l'ARN polymérase formée du cœur protéique des 3 sous-unités $\alpha/\beta/\beta'$ et de la sous-unité variable σ^{70} . La RNAP bactérienne possède en plus de ses sous-unités formant le cœur protéique, une sous-unité σ qui sert à l'initiation de la transcription. Il en existe plusieurs, chacune ayant des promoteurs de prédilection [37, 331]. Cette sous-unité a été remplacée chez les eucaryotes par un large panel de polypeptides pour chacune des trois RNAP (I, II et III) [74]. L'association de l'ARN 6S n'a lieu qu'avec l'holoenzyme E^{70} et entraîne l'inhibition de son activité transcriptionnelle. Sans ARN 6S, la transcription par la RNAP commence avec l'initiation de l'holoenzyme E^{70} et la formation du complexe promoteur ouvert (figure 3.4 A). Puis la sous-unité σ^{70} est libérée pour permettre l'élongation de la transcription. Lorsque l'ARN 6S est présent, il se lie aux sous-unités β/β' en présence de la sous-unité σ^{70} (figure 3.4 B) [326, 99]. L'ARN 6S ne s'associe pas uniquement avec le cœur $\alpha/\beta/\beta'$ ni avec σ^{70} ni avec l'holoenzyme E^S contenant σ^S . Cette association ARN 6S/ E^{70} inhibe l'activité transcriptionnelle de plusieurs promoteurs dépendants de σ^{70} [99] et active certains promoteurs dépendants de σ^S [304]. L'action de l'ARN 6S semble directement liée à sa structure en forme de bulle qui ressemble à la double hélice de l'ADN lorsqu'elle est ouverte dans l'ARN polymérase [319, 15, 305]. La liaison de l'ARN 6S semble être en compétition avec la liaison du promoteur au site actif de la RNAPII, induisant ainsi l'inhibition de la transcription. Cette association, d'importance mineure en phase exponentielle, devient omniprésente en phase stationnaire avec l'augmentation du nombre de copies de l'ARN 6S [326]. Les ARNs 6S d'autres espèces bactériennes (*Pseudomonas aeruginosa*, *Bacillus subtilis*, *Geobacillus stearothermophilus*, *Bordetella pertussis*, *Bordetella parapertussis*, *Bordetella bronchiseptica*) partagent des homologies de séquence et de

structure et sont également co-immunoprécipités avec l'ARN polymérase [15, 305]. La présence de l'ARN 6S joue un rôle sur la survie des bactéries *E. coli*. Lorsque celles-ci sont privées de nutriments sur de longues périodes, les bactéries ayant perdu le gène *ssrS* sont moins résistantes et meurent [304]. Un résultat inverse est obtenu en augmentant le pH [306]. Dans ces conditions de pH élevé, l'ARN 6S est très abondant et le gène *pspF* (phage-stock protein F) réprimé. En effet, la protéine *pspF* catalyse la formation du complexe « promoteur ouvert » et facilite l'initiation de la transcription par l'ARN polymérase E54 [255]. Cette ARN polymérase est requise pour la transcription d'une sous-classe de gènes qui empêcheraient la déplétion de ressources d'énergie dans certains environnements ou neutraliseraient partiellement des conditions défavorables [257].



d'après Kugel and Goodrich 2007

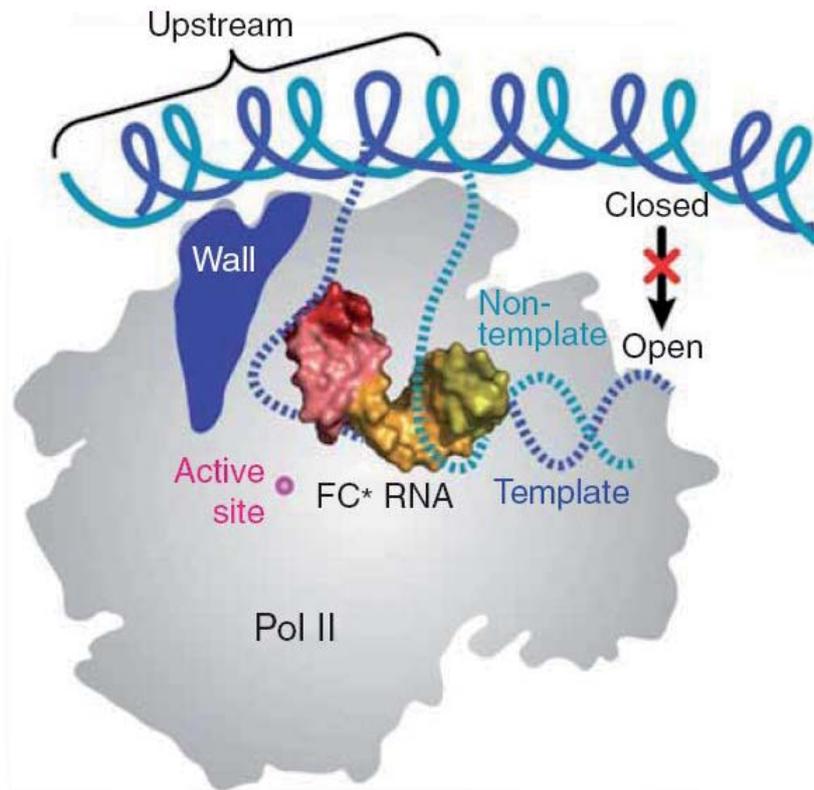
FIGURE 3.4 – **L'ARN 6S mime le promoteur ADN.** (a) Transcription par la RNAP de *E. coli*. Les complexes ouverts contiennent l'holoRNAP et le promoteur ADN. La transcription commence avec l'addition des NTPs, σ^{70} peut se dissocier et le complexe d'élongation se forme. Pendant ce temps, la région en amont de la bulle se réhybride et l'extrémité 5' du transcrit ARN se détache du promoteur ADN. (b) Régulation transcriptionnelle par l'ARN 6S. Pendant la phase stationnaire, l'ARN 6S se lie à l'holoenzyme E^{70} de la même façon que le promoteur ADN dans le complexe ouvert. Ce faisant, il bloque la transcription. L'inhibition est relâchée quand les cellules sortent de la phase stationnaire car la concentration en NTPs augmente et permet à la RNAPII de synthétiser des pARNs en utilisant l'ARN 6S comme matrice. Cela aboutit à la libération de σ^{70} . Les complexes 6S-pARN-RNAP qui restent sont instables. Cette instabilité serait provoquée par le fait que la bulle formée par l'ARN 6S ne peut pas se réhybrider derrière la RNAPII, comme le fait le promoteur.

L'ARN 6S possède de surcroît une qualité étonnante, celle de diriger sa propre transcription ARN dépendante. L'ARN 6S est sa propre matrice, produisant des ARNs de 14 à 20 nucléotides de long (pARNs). Cette transcription relâche l'ARN 6S et la sous-unité σ^{70} de l'ARN polymérase par la déstabilisation de l'hybride ARN 6S-pARN (figure 3.4 B). Comme l'ARN 6S est simple brin et hybridé sur lui-même avec une boucle, il ne peut pas y avoir de réhybridation permettant à la matrice d'avancer et ceci conduit à la libération du pARN. Cette transcription a lieu lorsque la concentration en rNTPs est élevée, condition réalisée lorsque la transcription est réduite. In vivo, cette transcription a lieu en sortant de la phase stationnaire mais ni pendant ni en phase exponentielle.

Un modèle est proposé dans lequel l'ARN 6S est produit en grande quantité, inhibe l'ARN polymérase et où les rNTPs s'accumulent, faute de transcription. Lorsque les conditions redeviennent bonnes, la transcription peut reprendre, produisant ainsi l'ARN 6S, piégé dans l'ARN polymérase, et permettant sa dissociation [327].

3.3.2 L'ARN B2

B2 est un ARN non codant de 178 nucléotides, transcrit par la RNAPIII et provenant des SINES (short interspersed elements). Il est abondamment exprimé dans les cellules murines ayant subi un choc thermique. Dans ces conditions, cet ARN s'associe à la RNAPII [5]. In vitro, l'ARN B2 est capable d'inhiber l'activité transcriptionnelle de la RNAPII. Un autre ARN inhibiteur de la RNAPII, l'ARN FC*, entre lui aussi en compétition avec B2. Or, le complexe ARN FC*-RNAPII avec ses 12 sous-unités a été cocrystallisé à une résolution de 3,8 Å et il en ressort que l'ARN FC* se place dans le sillon de l'ADN (figure 3.5). De cette façon, l'ARN FC* empêche l'ADN d'entrer et donc inhibe la transcription [145]. L'ARN B2 est constitué de quatre tiges-boucles et d'une partie non structurée entre la deuxième et la troisième tiges-boucles et d'une queue riche en adénosines (figure 3.6 structure). L'ARN B2 privé de sa queue riche en adénosines s'associe toujours au PIC et inhibe la transcription. La partie 3-149 comprenant les quatre tiges-boucles est suffisante à elle-seule pour se fixer à la RNAPII et inhiber la transcription. Cette inhibition n'est pas directement liée à sa fixation sur la RNAPII puisque la région 99-131 (partie non structurée et troisième tige-boucle) s'associe fortement à la RNAPII sans l'inhiber, alors que le segment minimal permettant l'inhibition va du nucléotide 81 au nucléotide 131 (deuxième et troisième tige-boucle avec la partie non structurée). Il faut d'ailleurs remarquer que l'ARN B2 81-131 est capable d'inhiber la formation des complexes d'élongation alors que l'ARN B2 99-131 n'a aucun effet [80].



d'après Kettenberger, Eisenfuhr et al. 2006

FIGURE 3.5 – **Modèle de l'inhibition de la formation du complexe promoteur ouvert par l'ARN FC***. La région en amont du promoteur ADN a été placée sur la RNAPII sur la base de sa localisation dans le complexe RNAP-promoteur chez la bactérie. La région en aval du promoteur ADN et le brin matrice dans la bulle du complexe ouvert sont modélisés selon le complexe d'élongation de la RNAPII. L'ARN FC* est représenté comme une surface moléculaire. Il s'hybride au promoteur, empêchant toute transcription. On comprend sur ce schéma que l'ARN FC* ajouté après le début de la transcription, n'aura pas d'effet inhibiteur.

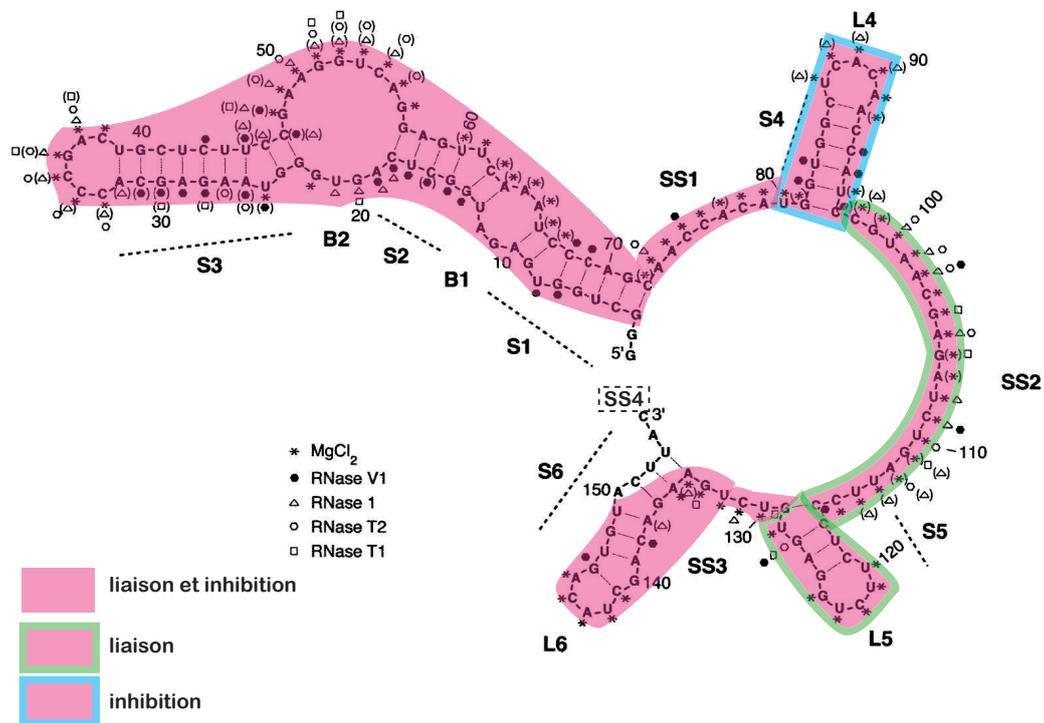


FIGURE 3.6 – **Structure de l'ARN B2.** Les différentes parties de l'ARN B2 ont été testées pour l'association à la RNAPII et pour l'inhibition de son activité transcriptionnelle. En rose, les parties nécessaires pour la liaison et l'inhibition, en rose cerclé de vert, la partie indispensable à la liaison, en rose cerclé de bleu, la partie responsable de l'inhibition de la RNAPII.

3.3.3 L'ARN Alu

L'ARN Alu serait l'homologue humain de l'ARN B2. Les gènes Alu sont les SINEs les plus représentés. On en compte plus d'un million de copies, ce qui représente plus de 10% du génome. Ils sont plutôt présents dans les gènes codant pour des protéines, le plus souvent dans les régions introniques. L'ARN Alu est composé de 2 bras similaires mais distincts, le bras gauche et le bras droit, réunis par une chaîne riche en adénine et se termine par une queue poly(A). Les deux bras se différencient par deux séquences : une séquence de 31 pb présente sur le bras droit et un promoteur interne pour la RNAPIII sur le bras gauche. Les gènes Alus ne sont exprimés qu'en cas de stress. Lors d'un choc thermique les ARNs Alu sont donc abondamment exprimés. L'équipe de Kugel et Goodrich a mis en évidence la restauration de la transcription de gènes éteints par le choc thermique grâce à l'inhibition de cet ARN par antisens (ADNc clone TS 103, connu pour être exprimé sous la forme d'un transcrit de 281 nucléotides dans les cellules humaines [278]). In vitro, cet ARN s'associe à la RNAPII et un ARN Alu pourrait se fixer à deux molécules de RNAPII. Les deux bras de cet ARN sont deux longues tiges-boucles séparées par une région non structurée. Cette dernière région et une région située sur la deuxième boucle semblent suffisantes pour l'inhibition de la transcription indépendamment l'une de l'autre. Comme l'ARN B2, l'ARN Alu s'associe à la RNAPII pendant la formation du PIC. Cette RNAPII en complexe avec B2 ou avec Alu est bloquée sur les promoteurs des gènes dont l'expression est réprimée [198].

Chapitre 4

Objectifs

Le laboratoire d'Olivier Bensaude travaille sur les ARNs non codants. Mon sujet de master a pris comme point de départ les travaux sur l'ARN 6S chez la bactérie et sur l'ARN B2 chez la souris. Ces ARNs non codants n'ont rien en commun d'un point de vue de la séquence mais possèdent plusieurs caractéristiques en commun. Ils sont tous deux exprimés lors d'un stress. L'ARN 6S est transcrit lorsque les bactéries arrivent en phase stationnaire et l'ARN B2 est produit par les cellules murines à la suite d'un choc thermique. Tous deux sont abondamment exprimés dans ces situations et sont responsables d'une inhibition de la transcription dans ces cellules. Cette inhibition est la conséquence de leur liaison à l'ARN polymérase. L'ARN 6S se fixe dans la crevasse de la RNAP, bloquant l'accès au promoteur. L'ARN B2 inhibe la RNAPII avant la formation du PIC, potentiellement par un mécanisme similaire. Ces deux exemples d'ARN inhibiteurs de la RNAPII nous ont poussés à chercher des ARNs non codants, associés à la RNAPII humaine en-dehors de tout contexte de stress.

L'objectif étant de répondre à une question la plus ouverte qui soit, il fallait avoir accès à toutes les formes de RNAPII, à savoir phosphorylée ou non. Pour réaliser cela, je suis partie de cellules HeLa, particulièrement résistantes au traitement nocodazole (85 % d'entre elles sont viables après 16-18h de traitement). En effet, en mitose, la RNAPII est détachée de l'ADN et donc facilement extractible sans avoir besoin d'utiliser des conditions trop dénaturantes qui casseraient les liaisons potentielles avec des partenaires ARNs. Les ARNs obtenus de cette façon ont été purifiés et marqués radioactivement. Parmi ces ARNs, le snARN U1 a été identifié.

La co-immunoprécipitation de l'ARN U1 avec une RNAPII non transcriptionnellement active a orienté la suite de mon projet sur la recherche du rôle de cette association et par conséquent à d'abord la caractériser. Nous avons cherché à déterminer si cette interaction était spécifique de l'ARN U1, des ARNs de l'épissage ou bien d'un ARN abondant en cherchant la présence de l'ARN U2 et de l'ARN 7SK. En cherchant biochimiquement quels étaient les partenaires associés, nous nous sommes rendus compte que la snRNP U1 était entièrement et spécifiquement associée à la RNAPII non transcriptionnellement active. Afin de savoir si cette liaison avait un lien avec l'épissage, j'ai créé deux lignées cellulaires comportant le même gène en multiples copies mais distinct entre les deux lignées par sa propriété à être épissé ou non. Ces copies sont insérées à un endroit unique du génome. Par microscopie à fluorescence et analyse d'images, j'ai ensuite observé l'enrichissement des snARNs de l'épissage sur ce macro-site de transcription. Alors que les ARNs U2, U4, U5 et U6 ne sont plus enrichis sur un site dont l'ARNm n'est pas épissé, le snARN U1 est recruté d'une façon similaire dans les deux cas. Les protéines de la snRNP U1 sont également présentes sur le macro-site de l'ARNm non épissé. Cet enrichissement de la snRNP U1 indépendamment de l'épissage a peut-être pour vocation de préparer l'épissage. Dans cette hypothèse, plusieurs protéines ont été testées pour voir si elles étaient toujours présentes avec la snRNP U1 : des protéines SR directement impliquées dans l'épissage ou associées à des granules de stockage des facteurs d'épissage (speckles). Les résultats montrent que la snRNP U1 est recrutée sans autres protéines impliquées dans les premières étapes de l'épissage et que cet enrichissement est indépendant des speckles. Le profil de protéines co-enrichies avec la snRNP U1 sur un gène dont l'ARNm n'est pas épissé semble donc précis et reflète probablement le rôle de ce recrutement.

La suite de ce manuscrit va présenter les résultats énoncés ci-dessus.

Deuxième partie

Matériels et méthodes

4.1 Analyse radiale

Plugins écrits sous ImageJ, Béatrice Spiluttini. Les images sont prises en x, y et z. Le pas du z est de 200nm.

4.1.1 Focuser

L'utilisateur indique les couleurs de ses images et précise la couleur du site de transcription (figure 4.1). Les cellules sont sélectionnées par leur site de transcription. L'image en MS2 est sommée sur tous les plans et le bruit de fond est soustrait.

Couleur du site de transcription

Couleurs supplémentaires sur lesquelles on veut faire l'analyse.

Pour chaque couleur cochée, une image sommée sur trois plans sera réalisée

Combien de plans doivent être sommés ?

Quelle est l'image sur laquelle on va sélectionner le site contrôle ?

Si on ne veut pas être trop influencé on peut décider de sommer deux images pour choisir le site contrôle.

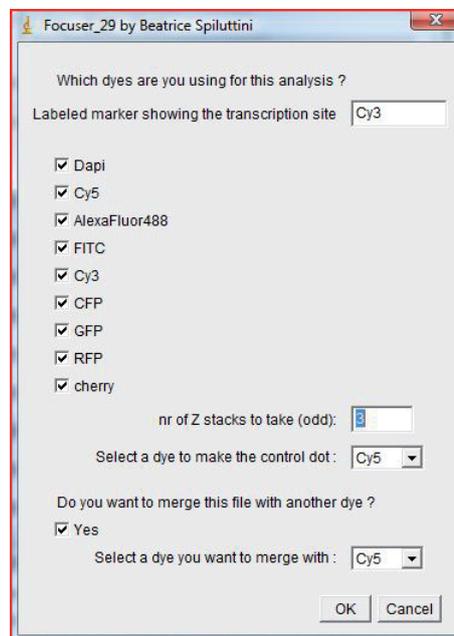


FIGURE 4.1 – Plugin focuser.

```
run("Z Project...", "start=1 stop=21 projection=[Average Intensity]");
getStatistics(Z_avint_area, Z_avint_mean, Z_avint_min, Z_avint_max, Z_avint_
    std);
close();
setBatchMode(false);
p_x_TS = new_xrect;
p_y_TS = new_yrect;
p_width = new_widthrect;
p_height = new_heightrect;
background_subtract = Z_avint_mean + 0.5 * Z_avint_std;
```

```
run("Subtract...", "stack value="+background_subtract);
```

L'utilisateur entoure d'un ovale les sites de transcriptions sur cette image. Le programme analyse les pixels contenus dans l'ovale sur l'ensemble des images en z. Il détermine le pixel aux coordonnées x, y, z ayant la plus forte intensité.

```
// find the optimal Z slice

w = getWidth();
h = getHeight();
the_max_pixel = -10;
for (i=p_x_TS; i<(p_x_TS+p_width); i++) {
    for(ii=p_y_TS;ii<(p_height+p_y_TS);ii++){
        curr_intensity = getPixel(i,ii);
        if(curr_intensity>=the_max_pixel){
            the_max_pixel = curr_intensity;
            pixel_maxX = i;
            pixel_maxY = ii;
        }
    }
}
close();
// on cherche le pixel de maximum intensite sur tous les stacks et on
determine le stack ou il se trouve = maxZ donc on ouvre le stack Cy3
file_Nr=substring(file, lengthOf(file)-8, lengthOf(file)-4);
open(MS2_STK_file_path);
ts_size=5;
intensity=0;
for (z=1; z<=nSlices; z++) {
    setSlice(z);
    pix_intensity = getPixel(pixel_maxX,pixel_maxY);
    if (pix_intensity>intensity) {
        intensity=pix_intensity;
        optimal_Z_slice=z;
    }
}
print("Optimal Z slice =", optimal_Z_slice);
```

Le stack est ouvert sur le plan en z contenant ce pixel d'intensité maximale entouré d'un cercle de 20 pixels de rayon. L'utilisateur peut changer ce plan s'il juge que celui-ci n'est pas en focus sur le site de transcription. Une fois le plan focus choisi, le plan focus est sommé avec le plan

immédiatement supérieur et le plan immédiatement inférieur.

```
run("Z Project...", "start=" + (optimal_Z_slice - ((ZP-1)/2)) + " stop=" + (
    optimal_Z_slice + ((ZP-1)/2)) + " projection=[Sum Slices]");
```

A partir de l'ovale dessiné par l'utilisateur, le programme calcule le pixel barycentre (fonction de l'intensité et de la position) sur cette nouvelle image. Formule du barycentre : $\frac{\sum X_i \cdot I_i + \sum Y_i \cdot I_i}{\sum I_i}$

```
barycentre_x_numerator = 0;
barycentre_x_denominator = 0;
barycentre_y_numerator = 0;
barycentre_y_denominator = 0;
// on calcule de barycentre et en meme temps on trouve les pixels du contour.
nb_pixels_contour = 0;
for( x = p_x_TS ; x < (p_x_TS + p_width) ; x++){
    for( y = p_y_TS ; y < (p_y_TS + p_height) ; y++){
        current_intensity = getPixel(x,y);
        if(current_intensity == 0){
            s = getPixel(x-1,y-1) + getPixel(x-1,y) + getPixel(x-1,y
                +1) + getPixel(x,y-1) + getPixel(x,y+1) + getPixel(x+1,
                y-1) + getPixel(x+1,y+1);
            if(s > 0){
                nb_pixels_contour += 1;
            }
        }
        barycentre_x_numerator += x * current_intensity;
        barycentre_x_denominator += current_intensity;
        barycentre_y_numerator += y * current_intensity;
        barycentre_y_denominator += current_intensity;
    }
}
print("NB pixels contour:", nb_pixels_contour);
pixels_contour_x = newArray(nb_pixels_contour) ;
pixels_contour_y = newArray(nb_pixels_contour);
tmp = 0;
for( x = p_x_TS ; x < (p_x_TS + p_width) ; x++){
    for( y = p_y_TS ; y < (p_y_TS + p_height) ; y++){
        current_intensity = getPixel(x,y);
        if(current_intensity == 0){
            s = getPixel(x-1,y-1) + getPixel(x-1,y) + getPixel(x-1,y
                +1) + getPixel(x,y-1) + getPixel(x,y+1) + getPixel(x+1,
                y-1) + getPixel(x+1,y+1);
```

```

        if(s > 0){
            pixels_contour_x[tmp] = x;
            pixels_contour_y[tmp] = y;
            tmp += 1;
        }
    }
}
}
barycentre_x = round(barycentre_x_numerator / barycentre_x_denominator);
barycentre_y = round(barycentre_y_numerator / barycentre_y_denominator);

```

Ensuite l'image est recentrée sur le site de transcription (=autant d'images que de sites de transcription) en coupant 170 pixels de part et d'autres du pixel barycentre. Les stacks correspondants aux autres couleurs sélectionnées (une même cellule est prise en plusieurs couleurs) sont sommés de la même façon au plan focus et les coordonnées du recadrage sont reportées sur chacune de ces nouvelles images. Une de ces images dans la couleur choisie (celle-ci pouvant être également une somme de deux images, avec un DAPI le plus souvent) est ouverte. Le pixel barycentre y est indiqué par une cible et l'utilisateur doit choisir un site contrôle en dessinant un carré. Le programme choisit ensuite aléatoirement un pixel à l'intérieur de ce carré. Celui-ci devient le pixel contrôle.

```

// dans le repertoire sum of 3 planes, pour faire le merge de deux images de
// couleurs differentes
file_target_fluo = replace(MS2_Zsum_file,TS_id+".tif",target_fluo+".tif");
sum_file_path_target_fluo= sum_path+ "/" +file_target_fluo;
if(do_merge){
    file_control_target = replace(MS2_Zsum_file,TS_id+".tif",control_target+".
    tif");
    sum_file_path_control_target= sum_path+ "/" +file_control_target;
    open(sum_file_path_target_fluo);
    ch_name=getTitle();
    open(sum_file_path_control_target);
    chcontrol_name=getTitle();
    mix_name = replace(chcontrol_name, control_target+".tif", "mix.tif");
    imageCalculator("Add create", chcontrol_name,ch_name);
    save(sum_path+"/"+mix_name);
}else{
    open(sum_file_path_target_fluo);
};

```

```
// fait une cible autour du site de transcription
run("RGB Color");
run("Colors...", "foreground=red background=white selection=yellow");
lenght=8;
label_diameter=11;
setFont("SansSerif", 20, "bold");
xt=5; yt=26;
setColor(200, 100, 200);
drawString("X="+x_TS+" :Y="+y_TS, xt, yt);
setTool(4);
// trait droit
x1=newx_TS+label_diameter;
y1=newy_TS;
x2=x1+lenght;
y2=newy_TS;
makeLine(x1, y1, x2,y2);
run("Draw");
// trait gauche
x1=newx_TS-label_diameter;
y1=newy_TS;
x2=x1-lenght;
y1=newy_TS;
makeLine(x1, y1, x2,y2);
run("Draw");
// trait bas
x1=newx_TS;
y1=newy_TS+label_diameter;
x2=newx_TS;
y2=y1+lenght;
makeLine(x1, y1, x2,y2);
run("Draw");
// trait haut
x1=newx_TS;
y1=newy_TS-label_diameter;
x2=newx_TS;
y2=y1-lenght;
makeLine(x1, y1, x2,y2);
run("Draw");
//faire le carre controle sur l'image mix
setTool(0);
```

```

showMessage("you will have 6 seconds to select an ROI for the control outside
a nucleolus");
wait(4000);
getSelectionBounds(x_cont, y_cont, width_cont, height_cont);
rnd=random;
x_cont_position=round(x_cont+(width_cont*rnd));
rnd=random;
y_cont_position=round(y_cont+(height_cont*rnd));
close();
close();

```

Ces images sont enregistrées en format tif. Un fichier log est créé pour chaque site de transcription, comportant entre autres : le nom du stack ayant permis la sélection du site de transcription, le nom des différentes images créées, les coordonnées du pixel barycentre dans le stack et dans la nouvelle image, le z sélectionné, les coordonnées du pixel contrôle dans la nouvelle image.

4.1.2 Automatic density analysis

L'utilisateur indique le grossissement de l'objectif, le rayon sur lequel s'étendra l'analyse (par défaut 15 pixels) et la couleur de l'image à analyser. Les fichiers logs sont ensuite lus, et les images correspondantes (à savoir image du site de transcription et image de la couleur choisie) sont ouvertes. L'analyse se fait sur le site de transcription pour l'image du site de transcription et sur le site de transcription et sur le site contrôle pour la couleur choisie. Cette analyse procède comme suit : le programme dessine des ovales de rayons croissants (incrémentés de 1 pixel à chaque fois) et mesure l'intensité moyenne sur l'aire ainsi délimitée.

```

//on fait les cercles autour du site de transcription en mesurant l'intensite
du signal MS2
prev_oval_intensity=0;
prev_area=0;
intensities_TS = newArray(radius);
density_TS = newArray(radius);
pixels= newArray(radius);
line1= newArray(radius);
for (i=1; i<(radius+1); i++) {
w=1+2*(i-1);
x=xTS_value-i+1;
y=yTS_value-i+1;

```

```

makeOval(x, y, w, w);
getStatistics(area, mean);
if(i==radius_average) { final_mean_TS=mean;}
oval_intensity=(area*mean);
density_TS_value=((oval_intensity-prev_oval_intensity)/(area-prev_area));
prev_area=area;
prev_oval_intensity=oval_intensity;
        intensities_TS[i-1]=density_TS_value;
    }

```

Ensuite, chaque intensité moyenne est retranchée à l'intensité moyenne de l'aire suivante. Ces intensités moyennes par couronne sont ensuite divisées par l'aire correspondante, donnant une densité de signal.

```

max_density_TS_norm=0;
min_density_TS_norm=1;
max_density_tg_norm=0;
min_density_tg_norm=1;
max_density_ctl_norm=0;
min_density_ctl_norm=1;
print("Nr; radius in nm; raw intensity TS; intensity TS; raw intensity target;
        intensity target; raw intensity control; intensity TS control");
wait(100);
for (i=1; i<(radius+1); i++) {
    density_TS_value=intensities_TS[i-1];
    density_tg_value=intensities_tg[i-1];
    density_ctl_value=intensities_ctl[i-1];
    pixels[i-1]=((i-1)*pix_size);
    line1[i-1]=1;
    density_TS_norm = (density_TS_value/final_mean_TS);
    density_tg_norm = (density_tg_value/final_mean_tg);
    density_ctl_norm = (density_ctl_value/final_mean_ctl);
    density_TS[i-1] = density_TS_norm;
    density_tg[i-1]=density_tg_norm;
    density_ctl[i-1]=density_ctl_norm;
    max_density_TS_norm=maxOf(max_density_TS_norm, density_TS_norm);
    min_density_TS_norm=minOf(min_density_TS_norm, density_TS_norm);
    max_density_tg_norm=maxOf(max_density_tg_norm, density_tg_norm);
    min_density_tg_norm=minOf(min_density_tg_norm, density_tg_norm);
    max_density_ctl_norm=maxOf(max_density_ctl_norm, density_ctl_norm);
    min_density_ctl_norm=minOf(min_density_ctl_norm, density_ctl_norm);
}

```

```
radiuss=((i-1)*pix_size);
print(i+"; "+radiuss+"; "+density_TS_value+"; "+density_TS_norm+"; "+
      density_tg_value+"; "+density_tg_norm+"; "+density_ctl_value+"; "+
      density_ctl_norm);
}
miny_ts=min_density_TS_norm-0.1;
maxy_ts=max_density_TS_norm+0.1;
miny=min_density_tg_norm-0.1;
maxy=max_density_tg_norm+0.1;
miny_ct=min_density_ctl_norm-0.1;
maxy_ct=max_density_ctl_norm+0.1;
maxx=radius*pix_size;
```

Ces données sont enregistrées dans un fichier csv et un graphique est produit pour chaque site de transcription avec la densité de signal rapportée au rayon pour l'image du site de transcription au TS, pour l'image dans la couleur choisie au TS et au RS.

4.1.3 Présentation des résultats

A chaque courbe correspond un site de transcription. Chaque valeur de la courbe est divisée par la valeur la plus périphérique car on admet que la dernière couronne est totalement en-dehors du site de transcription (figure 4.2 A). Cette étape est indispensable pour faire ressortir le signal par rapport au bruit de fond comme on peut s'en rendre compte dans la figure 4.2 B. La division par la couronne la plus excentrée donne des courbes qui se rejoignent toutes à 1 pour le dernier rayon. Toutes les courbes d'un site de transcription d'une part et du site contrôle d'autre part, sont ensuite moyennées pour une même condition (ex : le signal MS2) et une même lignée et les moyennes du site de transcription et du site contrôle sont reportées sur un graphique (figure 4.2 C). On peut ensuite reporter sur un même graphique les moyennes obtenues au site de transcription et au site contrôle pour différentes lignées ou conditions (figure 4.2 D). Les barres d'erreurs dessinées sur les graphiques représentent l'erreur standard (écart-type/racine(taille échantillon)). Pour déterminer si un résultat est significatif, j'ai à chaque fois réalisé un bootstrap à partir des données (figure 4.3). Les valeurs de chaque site contrôle et de chaque site de transcription sont mélangées et deux lots de 50 cellules sont tirés au sort, le lot A et le lot B. La courbe moyenne est calculée pour ces deux lots et on calcule la somme des moindres carrés à partir de chacun des points de la courbe. Ce

processus est réalisé plusieurs fois (1000 ou 10 000 fois). Les valeurs obtenues pour la somme des moindres carrées se disposent selon une loi normale. On compare ensuite la somme de moindres carrés des courbes moyennes des sites de transcription et des sites contrôles avec cette distribution normale. La p-value obtenue et indiquée en rouge sur le graphe, donne l'aire correspondant au cas où sites de transcription et sites contrôles sont identiques. Plus cette valeur est faible, plus le résultat est significativement différent entre les sites contrôles et les sites de transcription. On fixe le pourcentage d'erreur à 5% (p-value = 0,05). Si on obtient un résultat de 5%, on considère qu'il n'y a pas de différences. En général, les p-value sont inférieures à 10^{-2} , 10^{-3} .

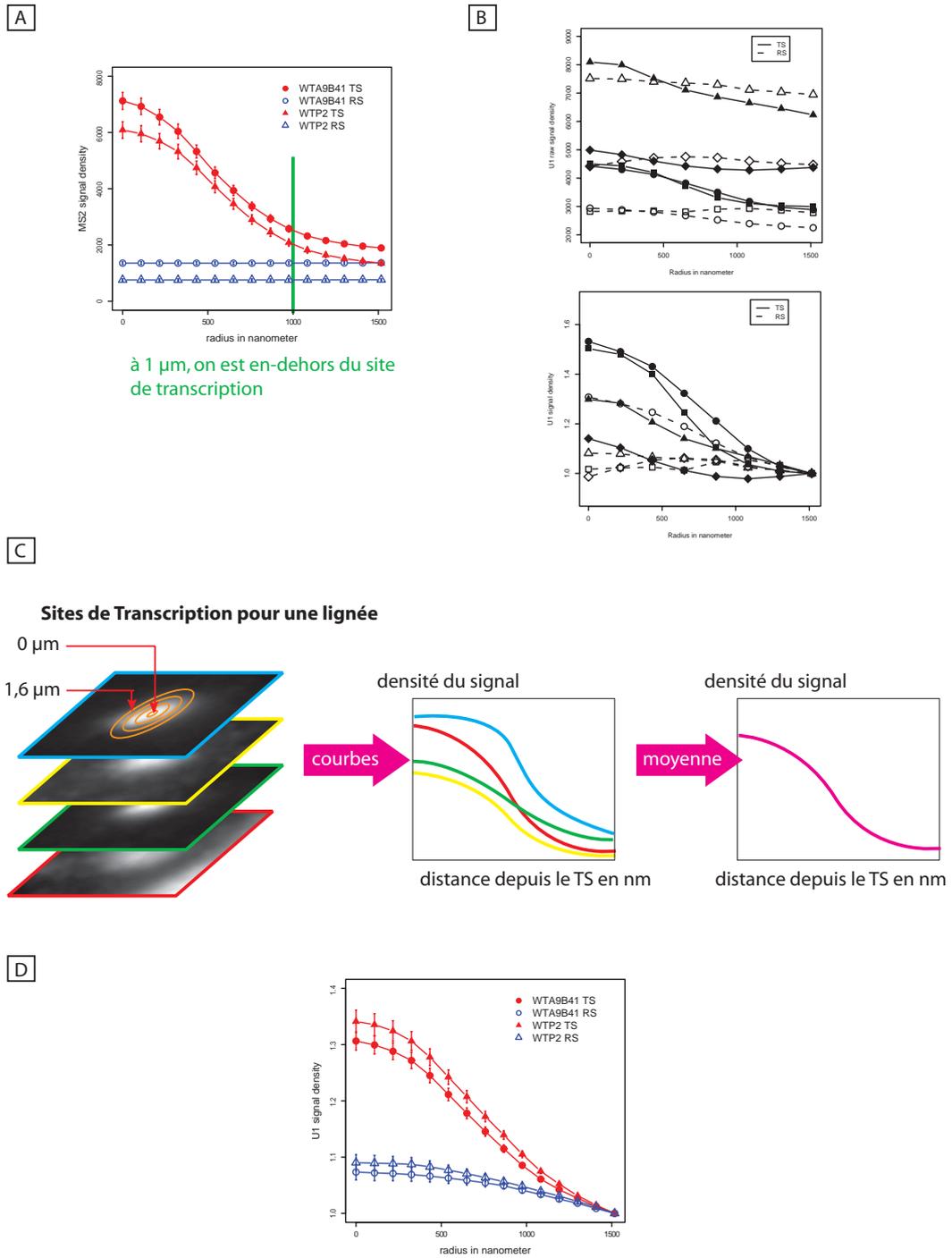


FIGURE 4.2 – Analyse radiale.

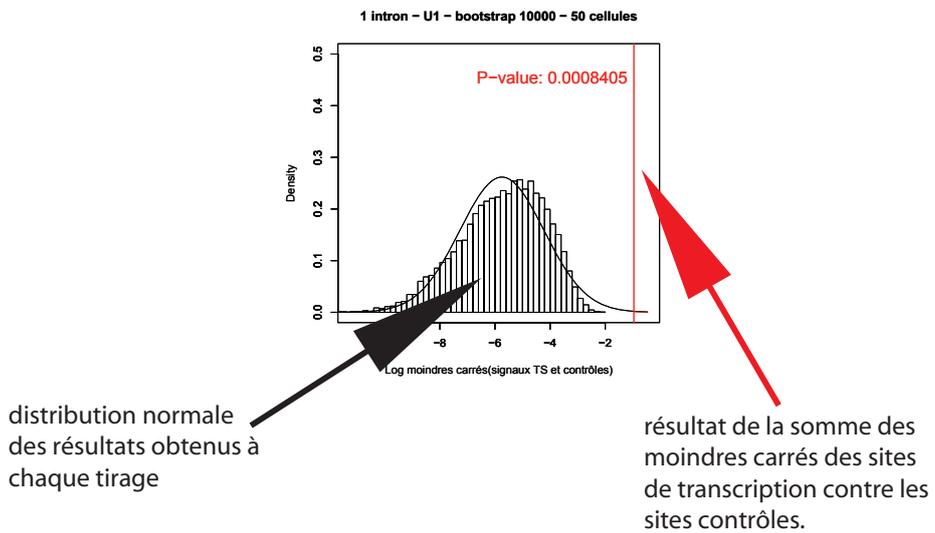
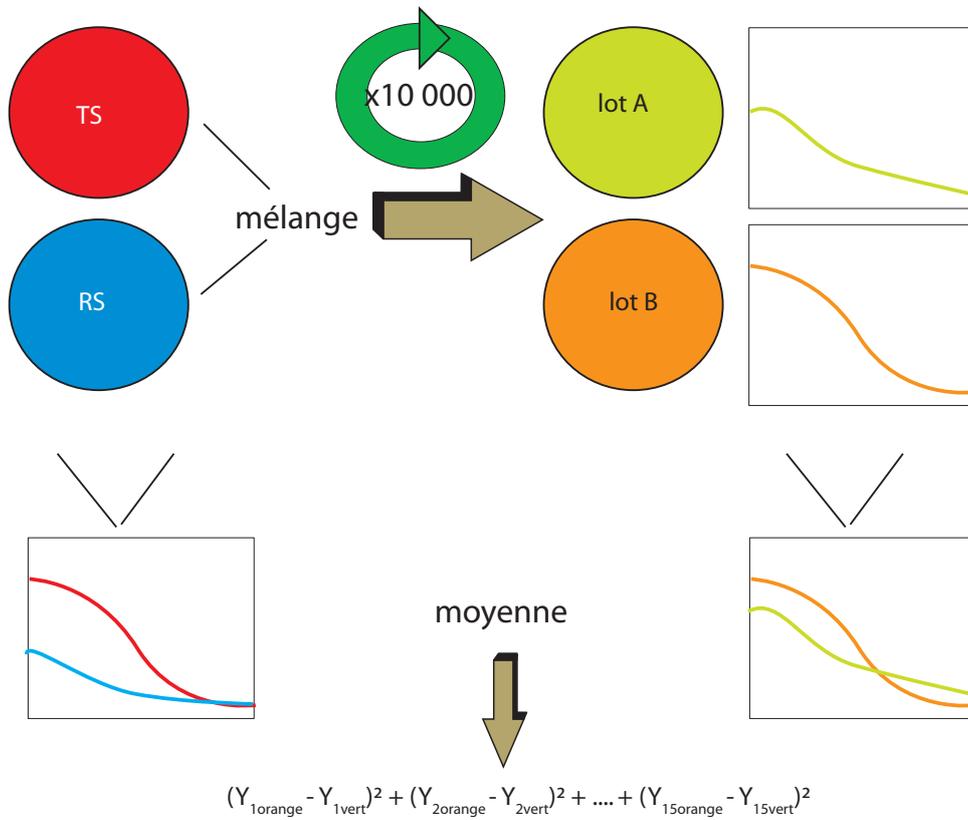


FIGURE 4.3 – Analyse par bootstrap.

4.2 Microscopie photonique

Le microscope utilisé est un Olympus IX 80. Les filtres utilisés sont : Cyan GFP, quad Semrock BrightLine DA/FI/TR/Cy5-4X4M-B. Les images sont prises avec l'objectif 60x. Le logiciel d'acquisition d'image est Metamorph version 6.3r7. La prise d'image est manuelle ou automatisée avec l'enregistrement de positions en x, y et z. Les temps d'exposition sont déterminés avant chaque acquisition afin d'avoir la meilleur dynamique de signal possible sans saturation de l'image. Les images obtenues sont des piles d'images par position (x,y) espacées de 200nm en z. Elles sont en format 16 bits.

4.3 Immunoprécipitations natives

Les cellules lavées au PBS sont râclées (pour les cellules non synchronisées) ou vortexées (pour les cellules mitotiques) dans le volume utile de tampon de lyse, préparé à partir d'un tampon 5X. Pendant 20 minutes, elles sont maintenues dans la glace en vortexant toutes les 2 minutes.

Tampon 5X

produits	concentration finale dans le 5X	quantité pour 10mL (mL)	quantité pour 50mL (mL)
Hepes 0,5M, pH 7,8	50mM	1	5
MgCl ₂ 1M	7,5M	0,075	0,375
KCl 1 M	50mM	0,5	2,5
EDTA 0,1M	1mM	0,1	0,5
NaCl 4M	250mM	0,625	3,125

Tampon de lyse

qsp (mL)	1	5	10
TP 5X	200	1000	2000
NP40 10%	50	250	500
eau	750	3750	7500
DTT	1	5	10
RNAse Inhibitor	1	5	10
Protease Inhibitor SIGMA	3,3	16,5	33
PMSF 200mM	5	25	50

4.4 Immunoprécipitations en crosslink

4 ou 2 boîtes diamètre 15cm

- 2 non traitées (1 cross-linkée ; 1 non cross-linkée)
- + 2 traitées DRB 100 μ M 1h à 37°C (1 cross-linkée ; 1 non cross-linkée)

Avant tout : précoater les billes (voir IP)

1. Conditions de cross-link

- Ajouter dans le milieu 1/10e Volume de tampon FA (formaldéhyde 11%)
- Agiter à température ambiante pendant 10 min
- Ajouter 1/20e de glycine 2,5M pH7 final dans le milieu
- Agiter à température ambiante pendant 5 min

2. Lyse

- Faire deux lavages sur boîtes dans du PBS froid.
- Gratter les cellules dans du PBS et les collecter dans tubes 50mL
- Centrif 5 min, 700g, 4°C
- Remettre en suspension les culots avec 4mL de RIPA /culot et aliquoter 5.10⁷-1.10⁸ cells par tube. AJOUTER protease inhibitor!!!!

3. Sonication

- Travailler dans des tubes de 15mL en Polypropylène

- Soniquer 30 sec puis pause 90 sec pendant 30 min dans la glace en rajoutant de la glace toutes les 5 min
- Ajuster à 25mM MgCl₂ (0,1mL de 1M) et 5 mM CaCl₂ (70μL de 2M au 1/7e). Ajouter 60 unités DNase I et 120 unités de RNase Inhibitor. Laisser à 37°C pour 15 min. Optionnel : Ajouter 1/10e Volume Triton X-100
- Transférer en eppendorfs et Eliminer les débris en centrifugeant 10 min à 20 000g à 4°C
Optionnel : les lysats peuvent être congelés à -80°C et utilisés le jour suivant

4. Préclearing

- Le lysat est préincubé avec 50μL de billes pA-sépharose avec un tARN (100μg/mL) pendant 1h à 4°C.
- Conserver 100μL pour western (+100μL de laemmli 2X) pour l'analyse prot
- Prélever 50 →200 μL à congeler à -20°C. Cet échantillon servira pour les ARNs input

5. Immunoprécipitation

- Préparer les billes protéineA-sépharose. Incuber les anticorps dans 1mL de tp RIPA avec les billes. Agiter pendant plusieurs heures à 4°C. OU BIEN faire le complexe immun avant de mettre les billes.
- Ajouter 800μL de lysat pour 20μL de billes. Agiter 1h →overnight à 4°C.
- Récupérer les surnageants : 100μL pour western (+100μL de laemmli 2X) et 700μL pour extraction ARN
- Laver 4 fois avec 800μL de tampon RIPA haut-sel
- Prélever 100μL du dernier lavage+billes (1/8 du total de billes) pour analyse en western + 100μL de laemmli 2X
- Remettre en suspension dans 300μL de tampon d'élution

6. Réversion du cross-link

- Incuber 2h →overnight à 65°C

7. Extraction et purification des ARNs

- Prendre 200μL pour les inputs
- Ajouter 800μL de solution toute prête de Qiazol (guanidine+phénol) (Invitrogen) à chaque tube

- Vortexer 15 sec
- Laisser 5 min sur la paillasse
- Ajouter 200 μ L chloroforme/mL de Qiazol (soit 160 μ L pour 800). Secouer vigoureusement
- Laisser 2 à 3 min à température ambiante
- Centrif 15 min à 12 000g à 4°C
- Récupérer la phase aqueuse (60% Volume Qiazol pipetté) et peser
- Ajouter 500 μ L de chloroforme
- Centrif 5 min à 12 000g à 4°C
- Récupérer la phase aqueuse, peser et recommencer un lavage chloroforme
- Ajouter 0,5mL d'isopropanol (ou d'EtOH 100% froid) pour 1mL de Qiazol +10 μ g glycogène
+ 1/10e Volume total de sodium acetate
- Placer overnight à -20°C OU 1h à -80°C
- Centrifuger à 20 000g pendant 20 mn.
- Enlever le surnageant et laisser sécher (attention à ce que le culot ne sèche pas trop, impossible à récupérer ensuite).
- Elution dans 20 μ L pour les inputs et pour les billes et nanodrop. Réajustement des volumes en fonction de la concentration et des pertes durant les lavages chloroforme.

4.5 Marquage des ARNs au pCp

Préparer le pCMP marqué au γ 32P ATP

(CMP = Cytidine 3'-MonoPhosphate, produit Sigma Réf C-1133)

1 μ l	CMP
1 μ l	T4 polynucleotide kinase buffer
1 μ l	DTT 50mM
1 μ l	eau
5 μ l	γ 32P ATP
1 μ l	T4 polynucleotide kinase (Fermentas, 10 U/ml)

Incubation 37°C 1h, inactiver à 95°C 5mn. Congeler à -20°C.

Ligation CMP marqué aux ARN

1 μ l	ARN
0,5 μ l	10x RNA ligase buffer
0,5 μ l	ATP 0,25 mM
0,5 μ l	DTT 33 mM
0,5 μ l	DMSO
0,5 μ l	CMP marqué
0,5 μ l	T4 RNA ligase 15U/ μ l

Mélanger puis mettre une nuit à la chambre froide.

Le lendemain ajouter 2 μ l de RNA loading buffer

900 μ l	formamide déionisé
50 μ l	Tris 1M pH 7,5
2 μ l	EDTA 0,5 M pH 8,0
25 μ l	bleu BMP 10%
25 μ l	xylène cyanol 10%

Chauffer à 60°C pendant 2mn puis mettre sur glace avant de déposer.

4.6 Gels

Les gels de Western blot et de Northern blot sont des gels en acrylamide et Tris. Pour la RNAPII, le pourcentage d'acrylamide utilisé est de 8% et le transfert est liquide (Biorad). Pour les Northern blots, les gels sont à 12% acrylamide et le transfert est semi-sec.

4.7 Cellules et Anticorps

Cellules : En biochimie, les cellules sont des HeLa données par M. Rousselle. Elles sont cultivées en présence de Dubelco's modified Eagles medium (DMEM) Glutamax (4,5g/L de glucose) avec du sérum de veau foetal à 10%. Pour la microscopie, les cellules sont des U2OS transfectées avec un plasmide tet-on et l'une ou l'autre des deux constructions (voir résultats). Elles sont cultivées en présence de Dubelco's modified Eagles medium (DMEM) Glutamax (1g/L de glucose) avec du sérum de veau foetal à 10%. La sélection des clones stables a été faite en puromycine.

4.8 Anticorps

Anti-RNAPII : N20 (Santa Cruz), 8WG16 (ascites), 8WG16 (Eurogentec), H224 (Santa Cruz), A10 (Santa Cruz), CTD4H8 (Millipore), pol3/3 (ascites), H5 et H14 (Covance). snRNP U1 : Anti-U170K CB7 (don de D.L. Black, University of California à Los Angeles), anti-U1A (Abcam). Anti-TFIIH : p62, cycline H, cdk7 (don de J.M. Egly au C.U. de Strasbourg). Epissage : anti-SC35 (Santa Cruz), anti-U2AF65 (Abcam), anti-SF2/ASF clone 103 (ZYMED Laboratories), anti-FLAG (SIGMA).

4.9 FISH (Fluorescence In Situ Hybridization) adapted for mammalian cells

Inspired from : Sensitive and high-resolution detection of RNA in situ (2000) Pascal Chartrand, Edouard Bertrand, Robert H. Singer and Roy M. Long Methods in Enzymology Volume 318, Pages 493-506H.

Preparation of FISH probes

1. 20 ng of 50-mer oligodeoxynucleotides including with 5 (amino-allyl modified thymidines) (EUROGENTEC) is used. Probe design : 50% GC is optimal. Between 48 and 52% is tolerable. The amino-allyl thymidines are scattered through the sequence, a minimum of 8 base spacing is required. The remaining thymidines are unmodified. If required, a non-hybridizing amino-allyl T can be added in 5' or an amino-allyl T-A in 3'. Avoid 3' or 5' hairpins. Check specificity by Blast : less than unspecific 18 nucleotide hybridization acceptable to avoid unspecific binding (10 base in ribosomal genes).
2. The DNA was dried in a Speed-Vac.
3. The DNA was resuspended in 20 μ L NaCO₃ (0.1M pH 9.3).
4. The solution was transferred to a vial opened at the last minute of CyTM5 (LQ15108/AC) or CyTM3 (LQ13108/AB) mono-Reactive Dyes from GE Healthcare and incubated overnight in the DARK!

5. The probes were fractionated by exclusion chromatography (Sephadex G-50) or (FPLC – Superdex 75) and equilibrated with a 10mM Triethylamine Buffer (TEAB) saturated overnight with CO₂ (the first peak coming out is the colored probe and the next one is uncoupled dye).
6. The probes are speed-dried in a Speed-VAC overnight.
7. The pellets are resuspended in 40 μ L (less if doubts on efficiency of the reaction) deionized distilled water and the concentrations is measured with a Nanodrop.
8. The concentrations are adjusted to 40 ng/ μ L per each probe with TE buffer (Tris-EDTA).
9. Probes are kept at -20° C for storage.

Coverslip preparation

1. Soak 18 mm diameter round coverslips in a beaker containing 0.11 M HCl
2. Boil for 10 minutes
3. Dry
4. Autoclave

Cells are plated on cover-slips in a 12-well tissue-culture plate at day 0.

Day 1, fixing cells

1. Cells are quickly washed twice in 1X PBS
2. The cells are fixed in 4% Paraformaldehyde (PFA) in PBS for 15 min at room temperature.
3. Quick wash in 1X PBS
4. Place in 70% ethanol
5. Store in 70% ethanol at 4° C at least 6 hours before proceeding.

Day 2, washing

1. The coverslips are washed twice for 10 min in 1X PBS.

2. Then washed once for 10 min in 0.5% Triton in PBS (This time can be reduced or eliminated if cells are transfected with GFP proteins as there is a possibility that the fluorescent protein will be lost during permeabilization).
3. Following, the coverslips are washed for 10 min in 1X PBS.
4. In the end the coverslips are washed twice at room temperature for 5 min in FAWASH buffer (60 ml 4X SSC + 40 ml 100% formamide).

During the washing solution One and Two are made :

Solution one (vol. for 10-12 c/s)

Use 2 μ l (80 ng) fluorescent probe per coverslip (probes stocks are usually 40 ng/ μ L) ; 3.6 μ L of 20X SSC ; 2 μ L of 5 mg/mL ssDNA/tRNA (100X excess over probe) (Mix equal volume of 10 mg/ml ssDNA (Sigma D-7656) and 10 mg/ml tRNA (Roche 109541) 160 μ L 100% formamide ; Adjust to 200 μ L with deionized water. Just before hybridization the solution is boiled in an Eppendorf tube for 5 min. and cooled on ice for 5 min.

Solution two

198 μ L deionized water 2 μ L BSA 50 μ L 20X SSC Add 200 μ L Solution Two to 200 μ L boiled Solution One and keep on ice.

Incubation chambers

A 10 cm \varnothing Petri dish is labeled on the bottom with the name of cell and probe. Place inside a reservoir (use the cap of a 50 ml Falcon tube) containing 2ml of 40% formamide to keep a humid atmosphere. Make drops of 25-40 μ L of probe solution in the dish (one drop per coverslip). When the washes are finished, the coverslips are gently dried and put face down (cell side) on one drop. The dish is sealed with parafilm and placed at 37°C in an incubator for 3 hours (or overnight)

Washes

1. Preheat FAWASH at 37°C
2. The c/s were replaced, face up, back into a 12 well dish and rinsed twice 15 min in warm FAWASH in an incubator.

3. The c/s were rinsed twice for 1 hour in 1X PBS at room temperature.
4. After that washes the cells on the coverslips are stained for 5 min. with DAPI (0.1 ug/ml stock solution) diluted 20X in 1X PBS.
5. Then they are washed for 5 min. in 1X PBS and the coverslips are mounted on slides for microscopy in mounting solution.

4.10 Immunofluorescence

1. Seed cells on glass slides
2. Prepare paraformaldehyde stock solution stored at -20°C .

Heat 4gr/40 ml distilled water, heat at 60°C , stir magnetically. Add NaOH - 1M drop by drop (2 to 3) until the solution is clear. Bring to 50 ml with distilled water. To 25 ml of this solution add 25 ml of PBSx2.

Fixation

3. Wash slides twice in PBS rapidly
4. Fix in paraformaldehyde 4% in PBS - 15 min.
5. Wash 3 times 5 min. in PBS.
6. Permeabilize in 0.2 % Triton X100 in PBS 10 min.

For GFP fluorescence, you may go directly to DNA staining step 14.

7. Wash in 100 mM glycine in PBS, 10 min.
8. Wash in 3% BSA in PBS 30 min.
9. Incubate first antibody in 0.1 % Tween20, 0.3 % BSA in PBS, one hour. (One drop on parafilm paper) in humidified atmosphere.
10. Wash in 0.1% Tween 20 in PBS.
11. Wash in PBS 2 times 10 min.

12. Incubate with 2nd antibody in 0.1 % Tween20, 0.3 % BSA in PBS, one hour. (One drop on parafilm paper) in humidified atmosphere. If 2nd antibody is fluorescent labelled (DANGER PHOTOBLEACHING = keep protected from light).
13. Repeat wash steps 10 and 11.

DNA STAINING

14. Stain during 5min with DAPI in PBS (DAPI 20X stock solution (0.1 ug/ml) stored light-protected at 4°C).

Alternatively, stain DNA with Hoechst 0,1% or propidium iodide (10 mg/ml) in PBS for 15 minutes. Propidium iodide stains both RNA and DNA - an RNase step may be performed to strip off RNA.

15. Wash in PBS 2 times 10 min.
16. Mount on slides with Mounting media (stored at -20°C).
17. Eventually seal with nail-varnish/nail-polish for long-term storage.

Troisième partie

Résultats

Chapitre 5

Association RNAPII-U1 indépendamment de la transcription

Les ARNs 6S et B2 sont tous deux des ARNs non codants, capables de se lier à la RNAP en conditions de stress et d'inhiber son activité. Au début de ma thèse, nous voulions savoir s'il existait des ARNs non codants associés à la RNAPII humaine, en-dehors de tout contexte de stress. Ces expériences m'ont d'abord conduite à observer une interaction entre le snARN U1 de l'épissage et la RNAPII dans un contexte où la RNAPII ne transcrit pas. Cette association semble être le reflet de l'interaction de la snRNP U1 au complet avec la RNAPII. Par imagerie sur des super sites de transcription, j'ai cherché à déterminer le lien entre cette association et le rôle du snARN U1 dans l'épissage. Deux lignées ont donc été réalisées par mutations ponctuelles, l'une avec un intron, l'autre sans intron. Alors que les UsnARNs U2, U4, U5 et U6 de l'épissage sont absents du site de transcription d'un gène sans intron, le snARN U1 est le seul à toujours être présent, dans des proportions identiques à ceux d'un gène avec intron. Ces résultats suggèrent que la snRNP U1 joue un rôle avant l'épissage, peut-être pour le préparer, et que cette fonction requiert son association avec la RNAPII.

5.1 U1 immunoprécipité avec la RNAPII

Nous avons d'abord cherché des ARNs non codants associés à la RNAPII humaine en-dehors de conditions de stress. Nous souhaitons pour cela extraire la RNAPII tout en conservant ses éventuelles associations avec des ARNs non codants. Or, dans une cellule non stressée, non synchronisée, la RNAPII est accrochée à l'ADN en cours de transcription.

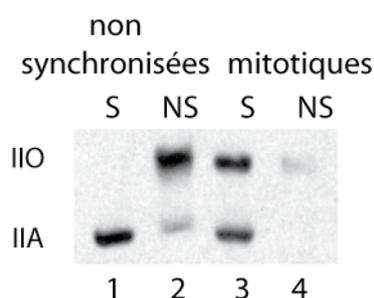


FIGURE 5.1 – **Solubilité de la RNAPII dans des cellules non synchronisées et mitotiques.** Extrait cellulaire après lyse 150mM NaCl, 0,5% NP40 et centrifugation 10000rpm. NS : non soluble ; S : soluble.

Pour l'extraire, il faut utiliser des conditions très sévères (haut sel et fort détergent) [229, 241]. Ce faisant, on perd une quantité importante d'associations potentielles avec d'autres molécules. Pour contourner ce problème, j'ai utilisé des cellules bloquées en mitose. Durant la mitose, la transcription est réduite [149, 107], notamment la transcription par la RNAPII [241, 291]. Les complexes d'élongation sont pour la plupart relâchés de l'ADN et la RNAPII est ainsi soluble. De plus, l'absence d'enveloppe nucléaire est un obstacle de moins à son extraction. On peut ainsi extraire la RNAPII de cellules mitotiques avec une faible concentration en sel et des détergents doux et ainsi préserver un maximum d'associations. Les cellules HeLa (R) ont donc été traitées environ 16h au nocodazole et les cellules mitotiques ont été collectées en les décrochant mécaniquement. Après une lyse douce avec 150mM NaCl et 0,5 % NP40, on obtient très peu de RNAPII soluble dans le lysat clarifié (à 10 000 rpm) de cellules non synchronisées (figure 5.1, pistes 1 et 2). De plus, la RNAPII extraite n'est pas représentative de la RNAPII cellulaire puisqu'on ne solubilise que la forme hypophosphorylée RNAPII_A. Au contraire, l'extrait mitotique contient la presque totalité de la RNAPII, sous ses deux formes : la forme hyperphosphorylée RNAPII_O et la forme RNAPII_A (pistes 3 et 4).

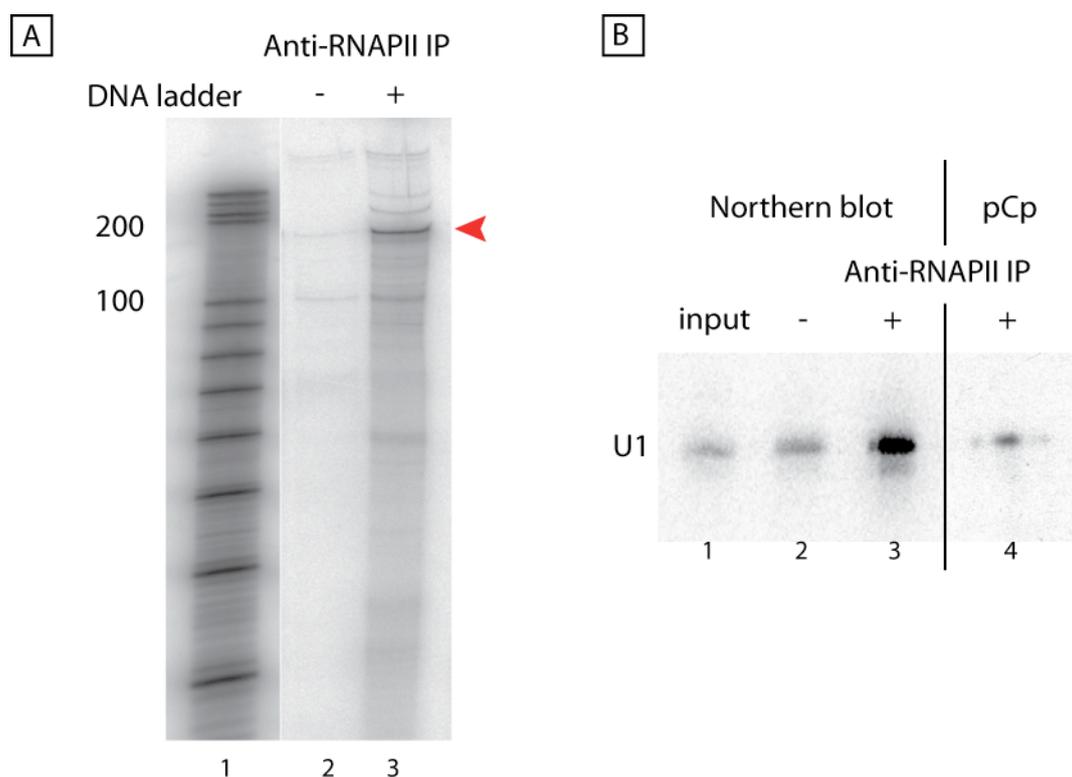


FIGURE 5.2 – **ARNs immunoprécipités avec la RNAPII.** **A.** Marquage au pCp des ARNs co-immunoprécipités avec la RNAPII. 1 : marqueur ADN de taille ; 2 : IP anticorps contrôle ; 3 : IP avec anticorps N20. **B.** Co-migration de la bande RNAPII majeure co-immunoprécipitée avec U1. 1-3 : Northern blot avec la sonde U1 ; 4 : pCp labelling ; 1 : input ; 2 : IP avec anticorps contrôle ; 3-4 : IP avec anticorps N20.

La RNAPII est un complexe de plusieurs sous-unités, capables d'interagir avec un large panel de protéines. Afin de se débarrasser de ces protéines satellites et de pré-purifier la RNAPII, les lysats d'extraits cellulaires mitotiques ont été séparés sur un gradient de glycérol (figure 5.6). Les fractions positives à la RNAPII ont été sélectionnées et une immunoprécipitation anti-RNAPII (IP RNAPII) a été réalisée. L'anticorps choisi pour faire cette immunoprécipitation est un anticorps anti-Rpb1 qui reconnaît le domaine N-terminal de cette sous-unité. Ainsi, on conserve dans l'immunoprécipitation les deux formes de la RNAPII, II_O et II_A. Les ARNs de l'immunoprécipitat ont ensuite été extraits et purifiés. Afin de les visualiser, ces ARNs ont été marqués radioactivement au pCp. Cette technique présente l'avantage d'ajouter une molécule unique de cytidine monophosphate radioactive par ARN. De cette façon, le signal radioactif obtenu est directement corrélé à la

quantité de molécules d'ARN. Les ARNs ainsi marqués ont été déposés sur gel (figure 5.2 A). Une de nos craintes était d'obtenir une traînée continue d'ARNs qui s'accrocheraient non spécifiquement à la RNAPII. Les résultats montrent au contraire plusieurs bandes distinctes dans la piste de l'IP RNAPII (piste 3). Certaines de ces bandes se retrouvent dans le contrôle (piste 2). Une bande cependant se dénote particulièrement des autres par un fort signal dans la piste de l'IP RNAPII. Cet ARN est compris entre 150 et 200 nucléotides. L'expérience répétée sur un autre échantillon provenant d'un autre gradient et d'un autre extrait cellulaire mitotique nous a, à nouveau, donné cette bande très enrichie dans l'immunoprécipitation RNAPII. Parmi les ARNs non codants connus, les snARNs du splicéosome sont compris entre 100 et 200 nucléotides, dont les deux plus grands, le snARN U1 (164 nucléotides) et le snARN U2 (186 nucléotides) [242]. Pour vérifier s'il pouvait s'agir de l'un ou de l'autre de ces snARNs, j'ai déposé sur le même gel l'échantillon marqué au pCp à côté de ce même échantillon non marqué. La partie non marquée du gel a été révélée par Northern blot avec une sonde U2 (donnée non présentée) puis U1 (figure 5.2 B). Cela nous a permis de constater que cette bande majoritaire co-migrerait avec l'ARN U1 (164 nt). Le Northern blot a de plus révélé que l'ARN U1 est également co-immunoprécipité avec la RNAPII. On voit en effet un très net enrichissement de l'ARN U1 dans la piste de l'IP RNAPII (piste 3) par rapport à la piste de l'anticorps contrôle (piste 2). Pour vérifier si cette bande enrichie dans l'immunoprécipitation RNAPII est bel et bien l'ARN U1, j'ai procédé à des tests RNase H sur des échantillons d'extraits cellulaires mitotiques. J'ai cependant été confrontée à des problèmes de dégradation d'ARN et de contamination qui ne m'ont pas permis de retrouver ces résultats. Il serait néanmoins intéressant de refaire l'expérience en refaisant un gradient et en modifiant les conditions de purification des ARNs pour confirmer l'ARN U1 comme ARN majeur associé à la RNAPII ou découvrir un autre ARN abondamment enrichi sur la RNAPII.

La découverte de la co-immunoprécipitation de l'ARN U1 avec la RNAPII dans des cellules mitotiques était un fait majeur. J'ai donc réalisé d'autres immunoprécipitations en partant de lysats cellulaires mitotiques totaux, pré-purifiés ou non sur un gradient de glycérol. Par Northern blot, ceux-ci ont montré à chaque fois une co-immunoprécipitation de l'ARN U1 avec la RNAPII (figure 5.3).

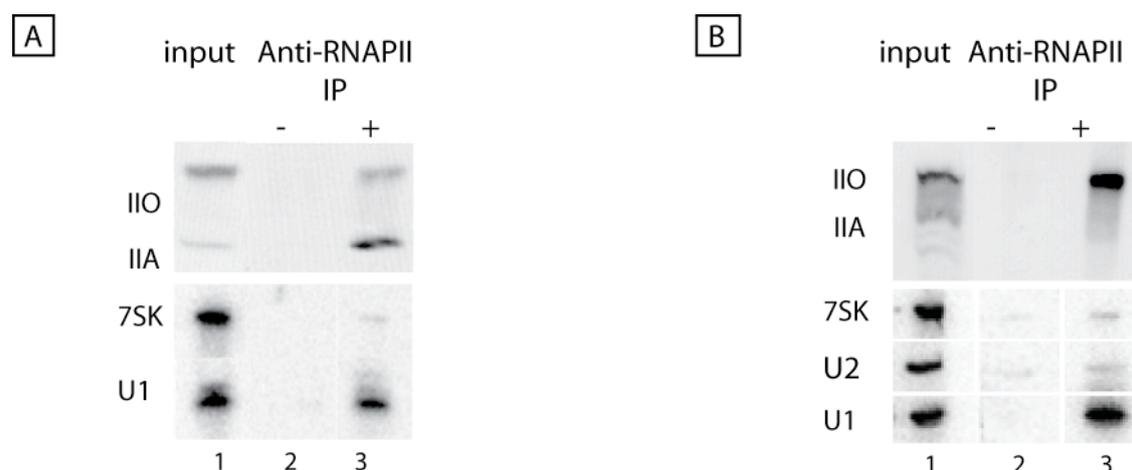


FIGURE 5.3 – **L'ARN U1 est co-immunoprécipité avec l'ARN polymérase II.** **A.** Immunoprécipitation sur des extraits cellulaires. L'input est dilué par 2. 1 : input ; 2 : anticorps contrôle ; 3 : anticorps 8WG16. **B.** Immunoprécipitation sur des extraits cellulaires. L'input est dilué par 2,4. 1 : input ; 2 : anticorps contrôle ; 3 : anticorps N20.

Deux anticorps différents ont été utilisés. Tous deux reconnaissent la sous-unité Rpb1 de la RNAPII. L'anticorps N20 en reconnaît la partie N-terminale, l'anticorps 8WG16, la partie C-terminale sur les répétitions non phosphorylées du CTD. Chacun de ces anticorps permet d'observer à chaque fois un très net enrichissement de l'ARN U1 avec la RNAPII, comparé à une immunoprécipitation contrôle (comparer les pistes 2 et 3). L'ARN U1 est donc associé, directement ou indirectement, à la RNAPII dans un contexte où la transcription est réduite et la RNAPII immunoprécipitée non transcriptionnellement active. Cette association est spécifique de l'ARN U1 puisque d'autres ARNs nucléaires non codants et abondants ne sont pas immunoprécipités avec la RNAPII. En effet l'ARN 7SK est un ARN très abondant dans le noyau (2.10^5 copies par noyau) tout comme le snARN U2 (5.10^5 copies par noyau) [358]. Pourtant, l'ARN 7SK n'est pas enrichi dans les immunoprécipitations N20 et 8WG16 (figure 5.3 A et B). De même, le snARN U2 n'est pas co-immunoprécipité avec la RNAPII (figure 5.3 B). Il est pourtant possible de trouver parfois de l'ARN 7SK sur ces immunoprécipitations (données non présentées). Cette observation n'étant pas constante, il est difficile de savoir si elle a une signification biologique. L'ARN 7SK inhibe l'activité du P-TEFb, facteur d'élongation de la transcription [223]. Sa présence sur la RNAPII pourrait donc ne pas être anodine. Il serait intéressant de comprendre quelles sont les conditions qui permettent de co-immunoprécipiter

l'ARN 7SK avec la RNAPII. Il pourrait être associé via d'autres facteurs et rester accroché si les lavages n'ont pas été assez forts. Ou bien il pourrait s'accrocher non spécifiquement à la RNAPII ou aux billes d'immunoprécipitation en fonction de faibles variations dans les conditions d'extraction. J'ai en effet parfois observé la présence mineure d'ARN 7SK sur les billes contrôles (billes pré-couplées avec un sérum pré-immun ou extrait cellulaire incubé avec un sérum pré-immun). Si l'association de l'ARN U1 sur la RNAPII n'est pas due à son abondance, elle n'est pas non plus le reflet d'une association générale des snARNs du splicéosome. L'ARN U2 est, en effet, partie intégrante du splicéosome et intervient juste après la reconnaissance du site donneur par la snRNP U1. L'abondance de l'ARN U1 et son association au splicéosome ne semblent donc pas intervenir dans son interaction avec la RNAPII. L'ARN U1 est donc spécifiquement associé à la RNAPII dans un contexte non transcriptionnel et indépendamment du splicéosome.

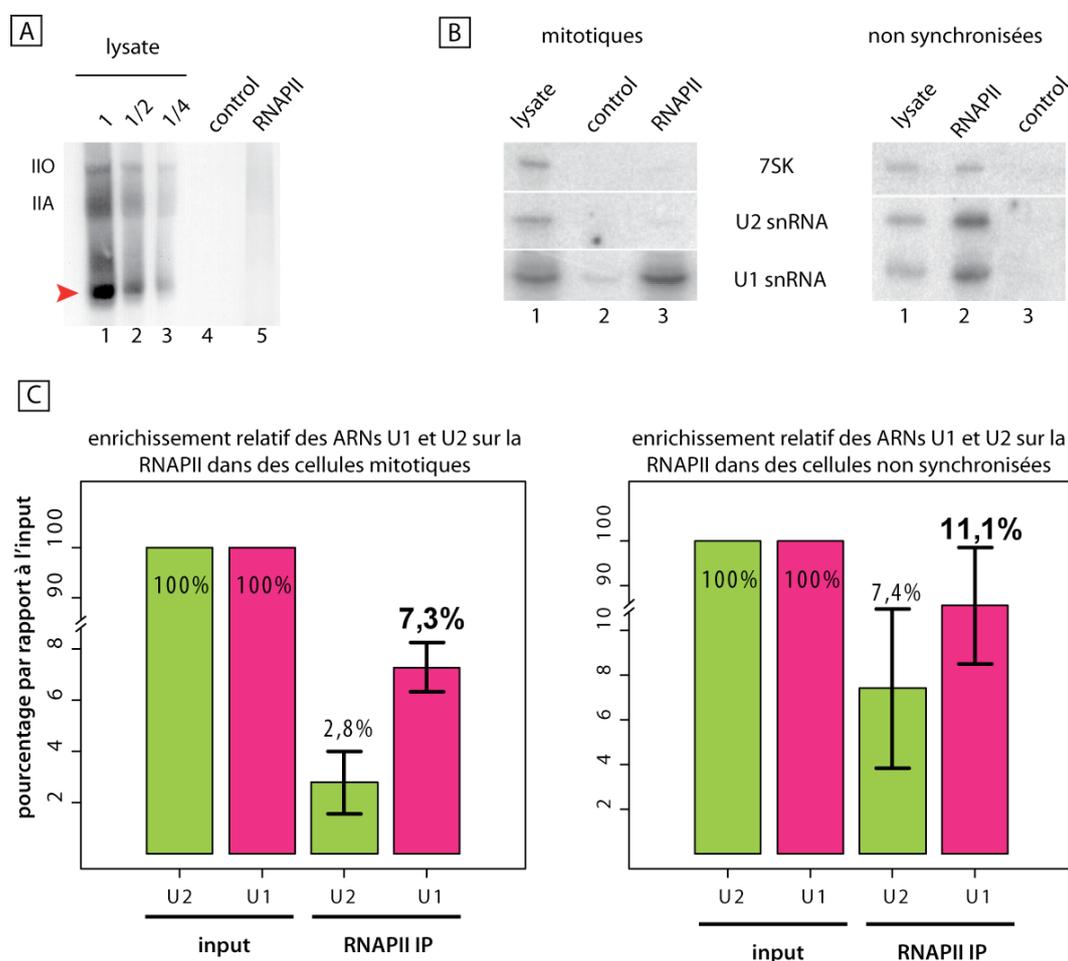


FIGURE 5.4 – **L'ARN U1 est associé *in vivo* à la RNAPII.** **A.** WB anti-RNAPII sur cellules mitotiques traitées au formaldéhyde, l'input a été déposé en 3 dilutions : 1 ; 1/2 et 1/4 (piste 1, 2 et 3) à côté de l'IP contrôle (piste 4) et de l'IP RNAPII (piste 5). **B.** Panneau de gauche : NB U1, U2, 7SK sur cellules mitotiques traitées au formaldéhyde. Panneau de droite : NB U1, U2, 7SK sur cellules non synchronisées traitées au formaldéhyde. **C.** Quantification à partir de NB sur 3 immunoprécipitations RNAPII et contrôle indépendantes. Les pourcentages indiqués représentent les rapports IP/input des signaux obtenus.

Afin de vérifier que cette association n'était pas le fruit d'une reconstitution *in vitro* lors de la lyse cellulaire, j'ai réalisé des couplages covalents au formaldéhyde sur ces cellules mitotiques. Après sonication, les lysats sont clarifiés, et on procède à l'immunoprécipitation. Les ARNs purifiés sont détectés par Northern blot. Les rendements de l'immunoprécipitation sont très diminués avec le couplage covalent, on immunoprécipite en effet qu'environ un quart de la RNAPII totale (figure 5.4 A). Néanmoins, on observe toujours sur ces extraits l'enrichissement de l'ARN U1 avec la RNAPII (figure 5.4 B). En revanche, dans les cellules mitotiques, l'ARN 7SK n'est pas particulièrement enrichi avec la RNAPII. De même, l'ARN U2 n'est pas largement co-immunoprécipité avec la RNAPII. On en trouve tout de même un peu. Sur trois expériences indépendantes en cellules mitotiques, on co-immunoprécipite un peu moins de 3% de l'ARN U2 sur environ un quart de RNAPII immunoprécipitée (figure 5.4 C). La proportion d'ARN U1 co-immunoprécipité s'élève à 7%. Si on ramène ces valeurs à la RNAPII totale, on peut dire qu'il y a environ 25% ($4 \times 7\% = 28\%$) d'ARN U1 associé à la RNAPII dans la cellule. Ce rapport est 2 à 3 fois inférieur à ce qu'on obtient sans formaldéhyde. Cela peut s'expliquer de deux façons : soit l'ARN U1 est associé à la RNAPII par une protéine intermédiaire et cette association indirecte est partiellement perdue au cours des sonications, soit l'ARN U1 interagit avec la RNAPII par son CTD. Or, celui-ci a tendance à être perdu avec les sonications (figure 5.4 A, indiqué par la flèche). En ce qui concerne la présence d'ARN U2 sur la RNAPII mitotique totale, on remarque qu'elle est inférieure à celle de l'ARN U1 (de l'ordre de 10% puisqu'il y a 3% co-immunoprécipité avec un quart de la RNAPII). Il semble plutôt s'agir d'une association mineure, ce qui expliquerait pourquoi on ne la voit pas en l'absence de couplage covalent. Cette association est d'autant plus mineure que sur 4 expériences, deux d'entre elles ne présentent presque pas d'ARN U2 co-immunoprécipité. Cette variabilité entre les expériences pourrait provenir en partie de l'anticorps utilisé. En effet, les plus importantes co-immunoprécipitation d'ARN U2 ont été observées avec des ascites de 8WG16, alors que l'anticorps commercial donne un enrichissement moindre en U2 et plus net en U1. Cette fraction d'ARN U2 associé à la RNAPII pourrait néanmoins avoir une signification biologique. La RNAPII est effectivement présente dans les granules d'interchromatine mitotique (MIG). Les MIG sont les speckles durant la phase mitotique. Ils concentrent la RNAPII sous sa forme hyperphosphorylée et des facteurs d'épissage dont les snARNs. En utilisant des conditions de couplage covalent, on peut imaginer que les molécules

présentes dans ces MIG soient couplées les unes aux autres du fait de leur proximité et qu'ainsi on puisse co-immunoprécipiter l'ARN U2 avec la RNAPII.

Dans les cellules non synchronisées, la RNAPII transcrit et la majeure partie des transcrits subissent un épissage cotranscriptionnel. Dans ces conditions, on s'attend à immunoprécipiter les snARNs de l'épissage se trouvant dans le voisinage immédiat de la RNAPII en cours de transcription. Sur des immunoprécipitations RNAPII, on obtient en moyenne 11% d'ARN U1 et 7% d'ARN U2 associés à moins d'un quart de RNAPII immunoprécipitée (figure 5.4 B et C), soit environ 40% d'ARN U1 et 25% d'ARN U2 totaux associés à la RNAPII cellulaire. Les différences entre les quantités immunoprécipitées sur une RNAPII transcriptionnellement active et inactive sont plutôt faibles (de l'ordre de 2) et peu significatives. Si on prend cependant uniquement les expériences faites avec l'anticorps 8WG16 commercial, ces écarts sont plus prononcés et globalement on peut dire qu'il existe une interaction plus fréquente entre les ARNs de l'épissage et la RNAPII transcriptionnellement active.

Les Northern blots nous ont donné des quantifications approximatives de l'ARN U1 associé à la RNAPII. Nous avons cherché dans un premier temps à être plus précis en purifiant les ARNs immunoprécipités et en amplifiant les snARNs par RT-PCR et PCR quantitative. Les nombreuses tentatives sont restées infructueuses du fait de la non-reproductibilité des résultats.

Les résultats ci-dessus démontrent l'interaction spécifique de l'ARN U1 avec la RNAPII dans des cellules mitotiques. Deux anticorps ont été utilisés dans ces expériences, l'un anti-NTerminal, l'autre anti-CTerminal de la sous-unité Rpb1. Pour essayer de déterminer si U1 s'associait préférentiellement à la forme hyperphosphorylée II_O ou la forme hypophosphorylée II_A , plusieurs anticorps ont été utilisés.

5.2 Autres anticorps anti-RNAPII

L'immunoprécipitation de la RNAPII sur cellules mitotiques a été réalisée avec plusieurs anticorps différents tous dirigés contre la sous-unité Rpb1. Les premières immunoprécipitations ont été réalisées avec l'anticorps polyclonal N20 reconnaissant le domaine N-terminal de Rpb1 et ont permis de trouver U1 associé à la RNAPII. Par la suite, les autres lots de cet anticorps commercial

n'ont plus sensiblement enrichi l'ARN U1 par rapport à une immunoprécipitation contrôle (donnée non présentée). L'autre anticorps anti-RNAPII, enrichissant très nettement l'ARN U1 par rapport au contrôle, est l'anticorps 8WG16. Celui-ci reconnaît le peptide YSPTSPS non phosphorylé du CTD. Or, la RNAPII n'est jamais complètement phosphorylée ou déphosphorylée. Il permet donc d'observer et d'immunoprécipiter les deux formes II_O et II_A de la RNAPII (figure 5.5, piste 4). Pour autant, il immunoprécipite plutôt la forme II_A que la forme II_O . Les autres anticorps anti-RNAPII utilisés immunoprécipitant correctement la RNAPII enrichissent faiblement l'ARN U1 (entre 5 et 10%) (figure 5.5). On notera cependant que les anticorps H14 et H5, bien qu'immunoprécipitant moins de RNAPII, donnent un signal U1 comparable aux autres anticorps (pistes 5 et 6). Cette tendance a été observée dans 50% des expériences. Les anticorps H5 et H14 reconnaissent les heptapeptides phosphorylés en Ser_2 et Ser_5 respectivement. L'efficacité d'immunoprécipitation de ces anticorps n'est pas très bonne, de ce fait les résultats en Northern blot sont variables. Un autre anti-CTD, l'anticorps 4H8 immunoprécipite très bien la RNAPII (piste 7). Bien qu'étant un anti-CTD, cet anticorps reconnaît les deux formes de la RNAPII. Cependant, il n'enrichit pas l'ARN U1. De même, d'autres anticorps anti-NTerminaux (pistes 1, 2 et 3) ne co-immunoprécipitent pas largement l'ARN U1 avec la RNAPII. D'un côté, nous avons donc l'ARN U1 co-immunoprécipité avec des anticorps anti-CTD (8WG16, et dans une certaine mesure par H5 et H14) aux propriétés contraires. Ces résultats sont en accord avec les précédents travaux de Kameoka. Des immunoprécipitations sont réalisés avec les anticorps N20, 8WG16, H5 et H14 sur des extraits nucléaires [137]. Toutes les quatre co-immunoprécipitent U1 mais 8WG16 est celui qui est le plus spécifique de cet ARN : il l'enrichit plus et les autres anticorps immunoprécipitent aussi les autres snARNs. De l'autre côté, ce même ARN U1 est également co-immunoprécipité par un anticorps anti-NTerminal (N20). Pourquoi un seul lot de N20 a-t-il permis d'enrichir U1 ? Probablement à cause de sa polyclonalité. La distribution des anticorps au sein du lot est différente à chaque nouveau lot et malheureusement pour nous, la distribution « gagnante » ne s'est réalisée qu'une seule fois. A ces anticorps anti-NTerminaux, il faut rajouter le cas à part du 4H8 qui est un anti-CTD mais qui reconnaît les deux formes. Au vu des résultats 8WG16, H5, H14 d'une part et 4H8 d'autre part, cette co-immunoprécipitation ne semble pas être en lien directe avec la phosphorylation du CTD de la RNAPII. Pourtant, il serait précipité de laisser cette hypothèse car ces résultats ne reposent que sur

un type d'expérience qu'est l'immunoprécipitation et une type de visualisation de la RNAPII qui est le Western blot. Outre les efficacités des anticorps, les interprétations sont restreintes par les limites techniques du Western blot. Le poids moléculaire de la sous-unité Rpb1 variant de 220 à 240 kDa en fonction de sa phosphorylation, son transfert total sur membrane est délicat et l'efficacité difficile à estimer. D'autre part, l'article de Kameoka examine des conditions particulières dans lesquelles elle trouve un complexe formé, entre autres, de RNAPII et de la snRNP U1. Une séquence 5'ss photoactivable est utilisée pour assembler le splicéosome *in vitro* avec des extraits nucléaires. Les complexes formés autour de cette séquence contiennent de nombreux facteurs du splicéosome (dont la snRNP U1) ainsi que de la transcription (dont la RNAPII_o). Les anticorps N20, H14 et H5 immunoprécipitent fortement ces complexes alors que l'anticorps 8WG16 très peu. Pourtant, dans un extrait nucléaire, 8WG16 co-immunoprécipite beaucoup d'ARN U1 et de façon spécifique. Il semble donc exister différentes associations entre l'ARN U1 et la RNAPII. Celles-ci pourraient avoir des rôles différents en fonction de l'état de phosphorylation du CTD.

Sur ces immunoprécipitations, on peut voir également la présence de l'ARN U2 et de l'ARN U5 (données non présentées). Par rapport à d'autres données obtenues auparavant, on peut dire que la présence de l'ARN U2 est de l'ordre du bruit de fond. La présence de l'ARN U5 sur des immunoprécipitations RNAPII (et pas sur les contrôles) a déjà été observée dans des expériences précédentes (données non présentées) sans que l'on sache si elle a une réelle signification biologique. L'expérience doit être menée à nouveau avec un anticorps contrôle.

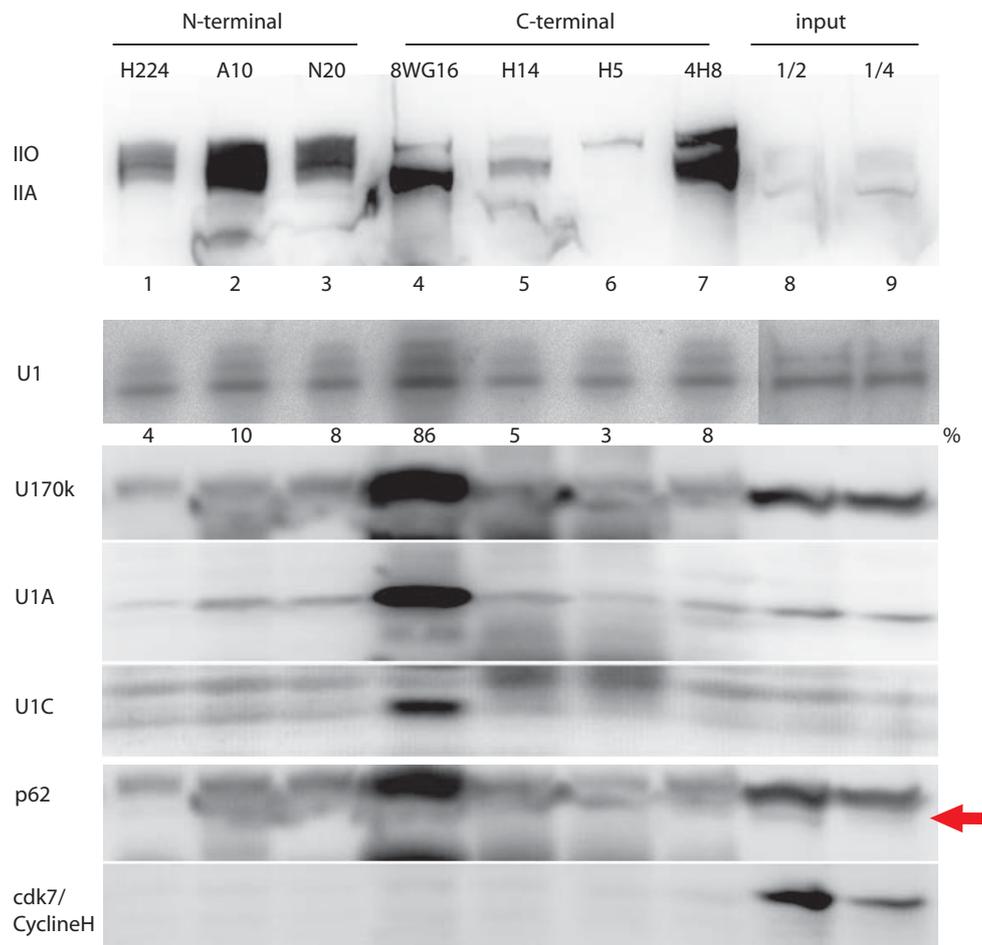


FIGURE 5.5 – **Immunoprécipitation de la RNAPII par différents anticorps.** Les immunoprécipitations ont été faites sur des lysats cellulaires mitotiques en conditions non dénaturantes. WB pol 3/3. Les chiffres correspondent aux rapports IP/input. Remarque : les protéines U170k et p62 migrent l'une en-dessous de l'autre. La flèche indique la taille attendue de p62, juste en-dessous du signal U170k visible sur le même WB. Les protéines cdk7 et cycline H ont des tailles très proches (37 et 39kDa) il est difficile de les différencier sur ce gel.

Sur ces immunoprécipitations, on peut voir également la présence de l'ARN U2 et de l'ARN U5. Il manque à cette figure une immunoprécipitation avec un anticorps contrôle. Par rapport à d'autres données obtenues auparavant, on peut dire que la présence de U2 est de l'ordre du bruit de fond. La présence de l'ARN U5 sur des immunoprécipitations RNAPII (et pas sur les contrôles) a déjà été observée dans des expériences précédentes (données non présentées) sans que l'on sache si elle a une réelle signification biologique. L'expérience doit être menée à nouveau avec un anticorps contrôle. L'ARN U1 est en complexe avec d'autres protéines au sein de la snRNP U1. Il était donc important de savoir si ces protéines étaient elles aussi associées à la RNAPII.

5.3 U1 n'est pas tout seul

Sur un gradient de glycérol, on observe une colocalisation de l'ARN U1 et de la RNAPII dans environ 3 fractions chevauchantes (figure 5.6 A et B). L'ARN U1 fait partie de la ribonucléoprotéine snRNP U1. Il y est associé avec les protéines Sm et des protéines spécifiques (table 5.1) (Hochleitner, Kastner et al. 2005). L'une d'elles, la U170k, a été analysée sur gradient par Western blot (figure 5.6 A). On observe ainsi une colocalisation de l'ARN U1 et U170k dans le gradient, signifiant que la snRNP U1 est conservée dans l'expérience. Pour vérifier si l'ARN U1 interagit avec la RNAPII accompagné par la snRNP dans son ensemble, des Western blots anti-U170k, anti-U1A et anti-U1C ont été réalisés sur les immunoprécipitations de RNAPII (figure 5.5). On observe la présence de ces trois protéines spécifiques sur chaque immunoprécipitation anti-RNAPII. Ces protéines sont éga-

humain	levure	Dans la snRNPU1 chez l'homme
U170kD	Snp1	oui
U1A	Mud1	oui
U1C	Yhc1	oui
FBP11/Fnbp3	Prp40	non
	Prp39	
	Snu56	
TIA-1	Nam8	non
	Snu71	
	Snu65	

TABLE 5.1 – Les protéines de la snRNPU1. Adapté de [134].

lement associées *in vivo* avec la RNAPII puisqu'on les observe sur des immunoprécipitations en couplage covalent (données non présentées). On peut toutefois noter que, de même que l'anticorps 8WG16 est l'anticorps qui co-immunoprécipite le plus d'ARN U1, c'est aussi celui qui co-immunoprécipite le plus les protéines de la snRNP U1 (piste 4). On a donc une relation entre la quantité d'ARN U1 immunoprécipitée et la quantité de snRNP U1 présente. Parallèlement à ce travail, une autre équipe a montré la co-immunoprécipitation des protéines U1A et U170k ainsi que de l'ARN U1 avec la RNAPII en présence d'ATP (Das, Yu et al. 2007). L'association de l'ARN U1 avec la RNAPII semble donc être le reflet de l'association de la snRNP U1 en entier avec la RNAPII.

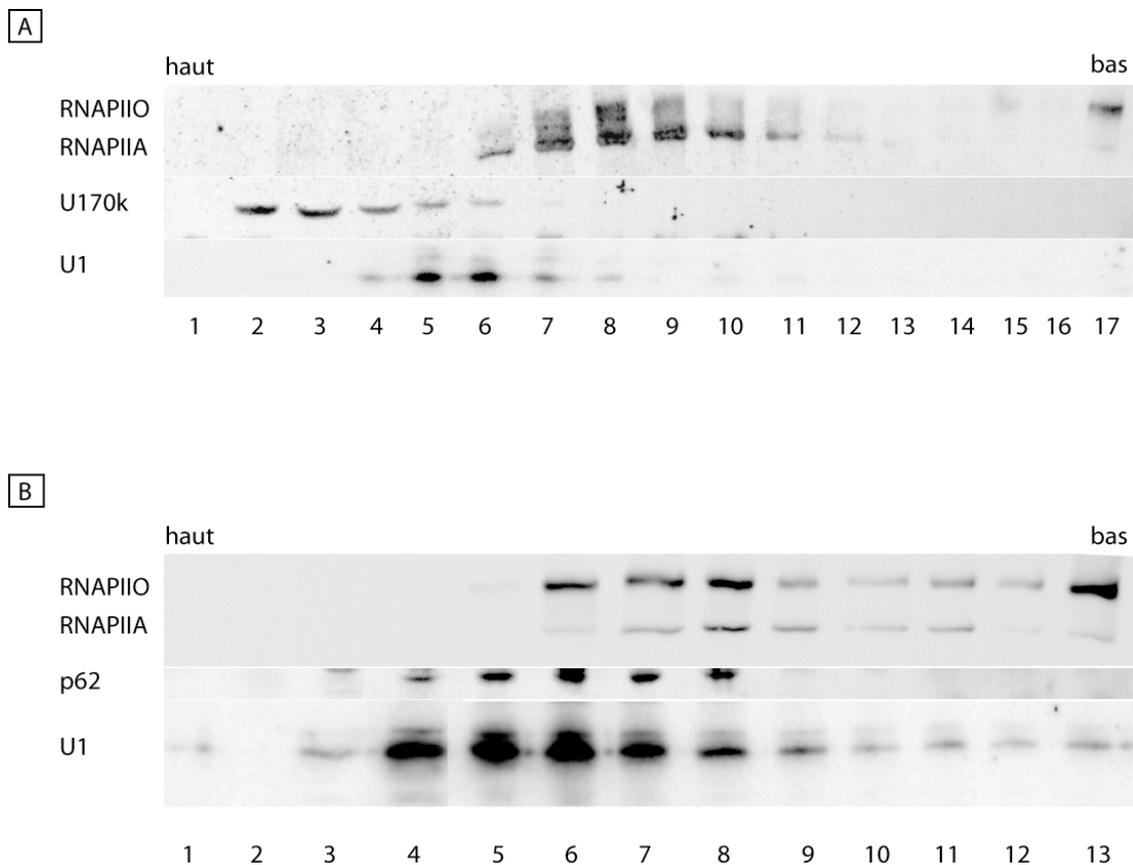


FIGURE 5.6 – Localisation des protéines associées à l'ARN U1 sur gradient de glycérol. Extrait cellulaire mitotique en NaCl 150mM, 0,5% NP40, déposé sur gradient de glycérol (10-40%) 150mM NaCl. WB anti-RNAPII (pol3/3), anti-p62, anti-U170k (H111).

Il existe par ailleurs d'autres travaux qui montrent l'interaction de l'ARN U1 avec le facteur d'initiation TFIIH, notamment la cycline H [166, 232]. Sur un gradient, on constate que la cycline

H (complexe cycline-dépendant) et p62 (complexe cœur) colocalisent effectivement avec l'ARN U1 (p62, cycline H et cdk7) (figure 5.6 B). En revanche, je n'ai pas trouvé ces protéines sur les immunoprécipitations anti-RNAPII en couplage covalent ou non (figure 5.5 et données non présentées). Contrairement aux protéines de la snRNP U1, elles ne sont pas plus enrichies sur l'immunoprécipitation 8WG16. Ces résultats proposent une association entre l'ARN U1 et la RNAPII, probablement avec la snRNP U1 en entier. Il faut souligner que cette association se fait dans un contexte non transcriptionnel et sans épissage [284] et avec une RNAPII soluble, détachée de l'ADN. De cette observation est née la question de savoir ce qui se passait sur une RNAPII active en cours de transcription et dans un contexte où la snRNP U1 ne serait pas requise pour de l'épissage cotranscriptionnel.

Chapitre 6

U1 recruté cotranscriptionnellement, indépendamment de l'épissage

6.1 La technique utilisée

6.1.1 Le système d'imagerie

Pour appréhender l'ARN U1 sur une RNAPII active, nous ne pouvions pas utiliser la biochimie. Il fallait trouver un moyen d'accéder à la RNAPII en cours de transcription sans pour autant perturber ses associations. Pour cela, nous avons utilisé un système d'imagerie mis au point par Xavier Darzacq et Yaron Shav-Tal dans le laboratoire de Robert Singer à New-York [62, 131]. Ce système consiste en l'insertion en multiples copies d'un gène à un endroit unique du génome. Les copies insérées en tandem forment alors un gigantesque site de transcription (figure 6.1). Ce site de transcription est visible en microscopie à fluorescence par FISH (Fluorescence In Situ Hybridization) : des oligonucléotides fluorescents dirigés contre l'ARNm produit permettent de visualiser les ARNs et donc leur site de production. En utilisant des sondes contre des parties exoniques ou introniques ou encore des jonctions d'exons, on peut observer l'ARN pré-mature et l'ARN mature. Ce site est également une cible sur laquelle on peut regarder la présence ou l'absence d'autres molécules, par FISH pour les acides nucléiques, par immunofluorescence pour les protéines.

Le système comprend le gène tronqué de la β -globine, fusionné au gène de la protéine fluo-

rescente cyan (CFP) comprenant un tripeptide qui dirige les protéines aux péroxysomes. La fluorescence de la CFP dans les péroxysomes permet de vérifier la bonne traduction de la protéine. Le gène se termine par des répétitions MS2. Ces séquences MS2 proviennent du bactériophage et n'interfèrent ni n'interagissent avec les composants cellulaires et ont la capacité de se lier fortement aux protéines MS2. Lorsqu'elles sont transcrites dans l'ARNm, elles se replient en deux boucles et lient spécifiquement des dimères de protéines MS2 fluorescentes. Par exemple, s'il y a 24 répétitions MS2 insérées dans le gène (comme ici), la molécule d'ARN est théoriquement liée à 48 protéines MS2. Ce faisant, le signal est très fort et fournit un rapport élevé de signal sur bruit [280, 93]. Ce gène est sous le contrôle d'un promoteur minimal CMV fusionné à sept répétitions du promoteur inductible Tet-on. Le Tet-on est un système dérivé lui-même du système Tet-off où la protéine tTA (Activateur Transcriptionnel en réponse à la tétracycline ou tetracycline-responsive transcriptional activator) est une protéine de fusion entre la protéine répressive tet, dérivée d'*Escherichia coli*, et le domaine d'activation VP16 du virus herpes simplex. Dans le cas du Tet-off, la transcription est réprimée en présence de tétracycline [105]. Dans le système Tet-on, la protéine rtTA (Activateur transcriptionnel inverse en réponse à la tétracycline ou reverse tetracycline-responsive transcriptional activator) a été créée en modifiant quatre acides aminés dans la protéine tTA. Cette protéine rtTA se lie à l'élément de réponse au tet et induit la transcription en présence de tétracycline [106, 211]. Dans les expériences présentées, la tétracycline a été remplacée par la doxycycline.

De cette observation est née la question de savoir ce qui se passait sur une RNAPII active en cours de transcription et dans un contexte où l'épissage est très réduit. Pour savoir si la snRNP U1 était requise dans le contexte transcriptionnel avec ou sans épissage, j'ai modifié ce système de sorte à avoir deux constructions similaires dont la seule différence provient de l'absence ou présence d'épissage (figure 6.2 A). Pour parvenir à cela, l'intron 1 a été muté ponctuellement sur un nucléotide dans les deux constructions afin d'être dans la phase ouverte de lecture. La construction standard comprend donc 2 exons de la β -globine, séparés d'un intron. La construction mutante « MUT » comporte en plus trois mutations ponctuelles, deux sur le site 5'ss et une sur le site 3'ss, qui invalident l'épissage (figure 6.2 C). Les constructions ont été transfectées dans des cellules U2OS afin de créer des lignées stables. *In vivo*, l'ARNm « non épissé » ne subit pas d'épissage dû

à des sites cryptiques, comme on peut le constater par western blot et on peut vérifier la bonne traduction de la protéine par le signal CFP dans les péroxysomes (figure 6.3). Un marquage FISH des exons (exon 1 de la β -globine ou MS2) permet de visualiser les ARNs pré-matures et matures. Un marquage FISH de l'intron ne permet d'observer que les ARNs pré-matures, pas encore épissés. Dans la lignée WT. α 1.A9B41 (WT), on voit le marquage exonique très fort au niveau du site de transcription et plus dispersé dans le noyau et dans le cytoplasme (figure 6.3). Dans la lignée MUT.t2.5A (MUT), on a une image similaire. De même le marquage intronique marque le site de transcription avec un signal très fort mais les ressemblances s'arrêtent là entre les lignées. En effet, le signal intronique disparaît très vite en-dehors du site de transcription dans la lignée WT alors qu'il demeure présent partout dans le noyau et dans le cytoplasme dans la lignée MUT, comme le signal exonique. L'imagerie est en corrélation avec les données biochimiques : l'ARNm de la lignée MUT n'est pas épissé. Dans la lignée WT, puisque l'intron n'est visible qu'au site de transcription, il doit donc être rapidement dégradé. L'épissage dans la lignée WT est donc cotranscriptionnel.

Nous avons deux moyens à notre disposition pour visualiser les sites de transcription : le FISH sur l'ARNm ou la protéine MS2 fluorescente. Quand cela a été possible, j'ai toujours privilégié le FISH car il est plus facile à mettre en œuvre. Lorsque cela a été nécessaire, une transfection transitoire MS2-YFP a été réalisée sur les lignées. La protéine MS2-YFP a été utilisée à chaque fois qu'une immunofluorescence devait être menée et dont l'anticorps ne supportait pas le traitement préalable du FISH. Aucune différence de résultats n'a été observée entre les deux méthodes.

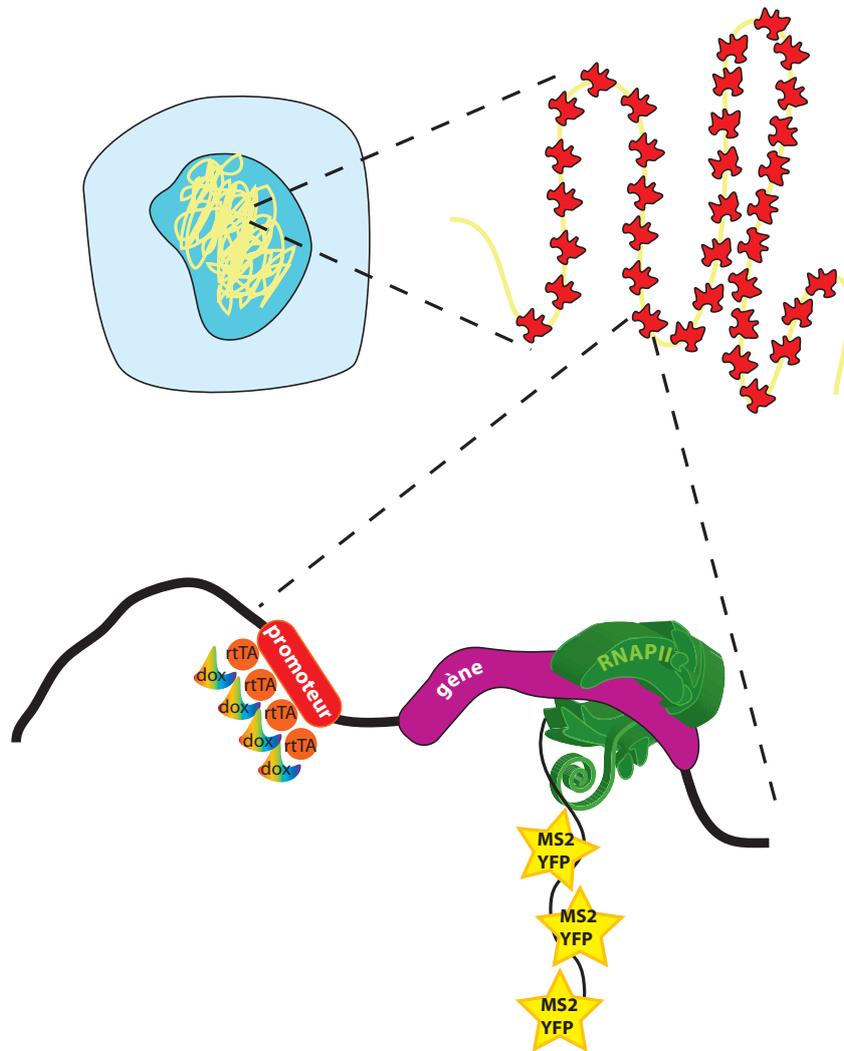


FIGURE 6.1 – **Insertion multicopies dans le génome des cellules U2OS.** Les constructions géniques sont insérées en tandem en un seul endroit dans le génome de cellule U2OS possédant le répresseur Tet-On. Les protéines MS2-YFP reconnaissent les séquences MS2 sur mes ARNs. A la place de ces protéines fluorescentes, on peut utiliser des oligonucléotides fluorescents s'hybridant avec les séquences choisies. rtTA : activateur transcriptionnel inverse en réponse à la tétracycline. dox : doxycycline.

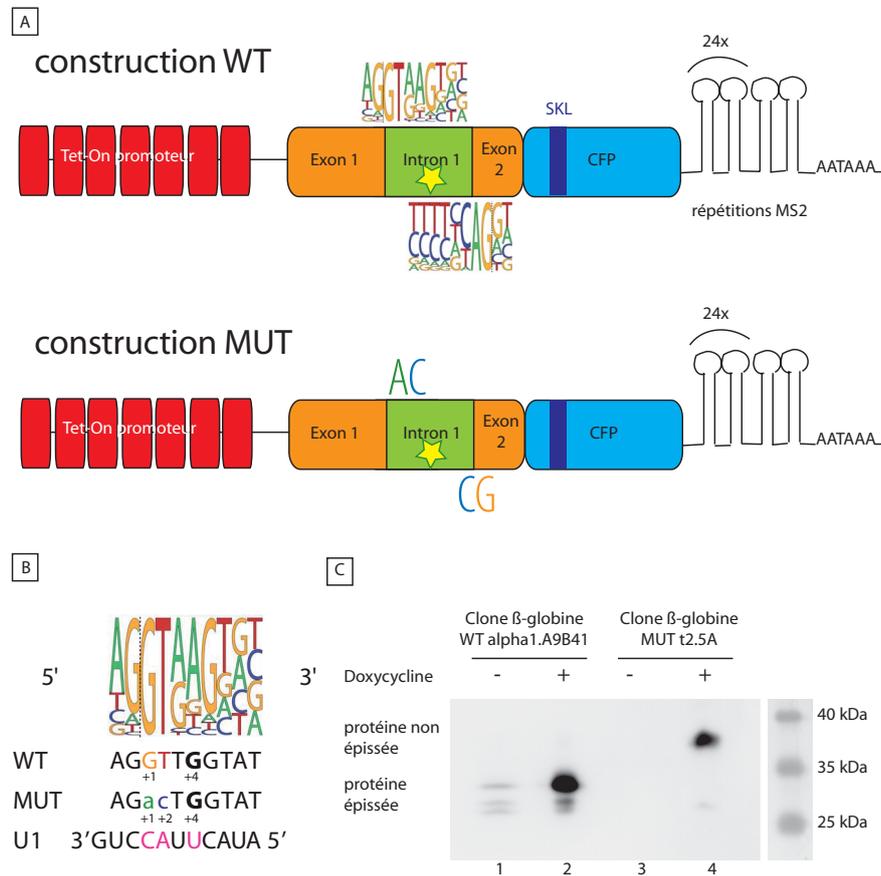


FIGURE 6.2 – **Constructions épissées et non épissées à partir du gène de la β-globine.** A. Le gène comporte 7 éléments de réponse Tet-on fusionnés à un promoteur CMV. L'intron 1 a subi une délétion d'un nucléotide pour être en phase ouverte de lecture (*). L'exon 2 a été coupé pour être fusionné à la séquence codant pour la CFP, contenant le signal SKL de localisation aux peroxy-somes. Le gène se termine par 24 répétitions de la séquence MS2. La construction WT possède des sites d'épissage intègres. La construction MUT a subi en plus deux mutations ponctuelles (substitution de 3 nucléotides) pour abolir les sites d'épissage. B. Séquence consensus du site donneur, séquence mutée et séquence de l'ARN U1 s'hybridant avec le site donneur. C. WB anti-GFP sur lysats des clones U2OS WT alpha1.A9B41 et MUT t2.5A, induits (pistes 2 et 4) ou non à la doxycycline (pistes 1 et 3).

6.1.2 L'analyse des images

Les images obtenues ont d'abord été les résultats de différents FISH dirigés contre les snARNs de l'épissage. Or ces ARNs présentent deux inconvénients : ils sont très abondants [358] et s'accumulent dans de petites structures appelées speckles qui sont des lieux de stockage de molécules de l'épissage [172]. De ce fait, une hybridation U1 donne une image où le noyau contient du signal partout avec de petits foci de signal plus fort (les speckles) (figure 6.4). Une augmentation locale du signal passe alors inaperçue. Pour déterminer s'il y a un réel enrichissement au site de transcription, j'ai mis au point une technique d'analyse d'images basée sur l'étude du site de transcription et d'un site contrôle pris au hasard dans le noyau en-dehors des nucléoles. Sur chaque cellule analysée, la densité du signal est calculée au site de transcription et au site contrôle sur des couronnes concentriques de 1 à 1,6 μm de rayon. Ces données sont ensuite reportées sur un graphique. Une moyenne de ces courbes est faite par site (transcription ou contrôle) et à son tour reportée sur un graphique qui nous permet de visualiser l'enrichissement d'une molécule donnée au site de transcription par rapport à une zone en-dehors de ce super site de transcription.

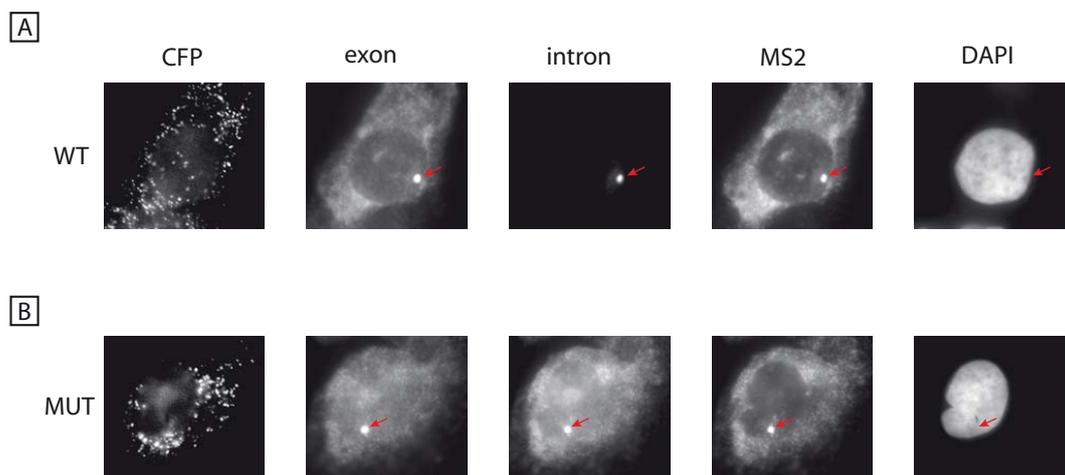


FIGURE 6.3 – **Lignées cellulaires stables WT et MUT.** La protéine exprimée va aux péroxysomes, on la localise par son signal CFP. L'exon, l'intron et les répétitions MS2 sont visualisées par FISH (Fluorescent In Situ Hybridization). La flèche indique le site de transcription.

6.1.3 Transcription

Plusieurs copies de chaque construction sont insérées au même endroit en nombre variable. Ainsi la construction « épissée » est présente en environ 40 copies dans la lignée WT α 1A9B41 et la construction « non épissée » environ 25 copies dans la lignée MUTt.2.5.A. La transcription dans ces lignées aux sites de transcription est quantifiable par le signal MS2 obtenu en FISH (figure 6.5). On voit qu'au site de transcription, le signal MS2 est très fort. Plus on s'éloigne du site de transcription, plus celui-ci diminue, revenant au taux basal dans le noyau. Malgré la différence du nombre de copies, on constate que le niveau de transcrits au site de transcription entre les deux lignées est tout à fait comparable. Si l'on observe par immunofluorescence la RNAPII sur ces sites, on se rend compte que son enrichissement semble être différent en fonction des lignées. On voit en effet moins de RNAPII au site dans la lignée MUT que dans la lignée WT. Cette différence est peut-être le reflet du nombre de copies insérées. Alors comment expliquer que la quantité de transcrits est la même dans les deux lignées ? Deux explications sont proposées :

- la RNAPII de la lignée MUT est beaucoup plus processive, à cause de l'environnement chromatinien de l'insertion, de l'absence de facteurs liés à l'épissage, de l'absence d'épissage lui-même, de la RNAPII elle-même, etc.
- les transcrits quittent moins vite le site de transcription car ils n'ont pas l'étiquette de l'Exon Junction Complex (EJC) qui les valide pour l'export.

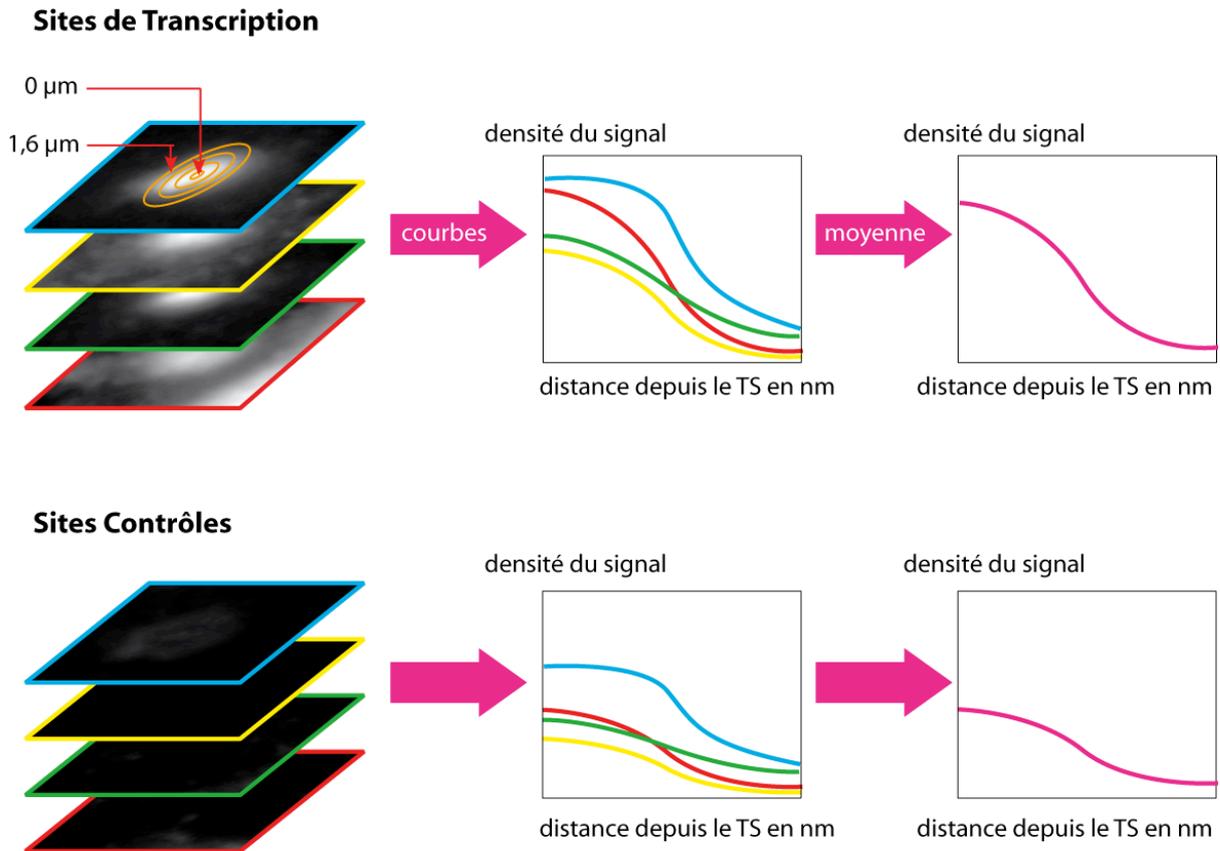


FIGURE 6.4 – **Analyse radiale.** L'intensité du signal est mesurée depuis le centre du site de transcription jusqu'à rayon de $1,6 \mu\text{m}$ en couronnes croissantes de 1 pixel. Cette intensité est ensuite rapportée à la surface pour obtenir la densité du signal. On procède de la même façon pour le site contrôle. Une courbe est obtenue par cellule et par site. La moyenne de la densité du signal est calculée sur toutes les courbes pour un site donné.

6.2 Recrutement du splicéosome

6.2.1 Les snARNs U4, U5, U6

La lignée WT semble faire un épissage co-transcriptionnel. J'ai donc regardé si les éléments du splicéosome étaient recrutés sur le site de transcription (figure 6.6). Par un FISH dirigé conjointement contre les snARNs U4, U5 et U6, on constate que ces 3 snARNs sont largement enrichis au site de transcription dans la lignée WT par rapport à un site contrôle dans la cellule. Cette observation corrobore le fait que l'intron disparaît très vite en s'éloignant du site de transcription : le

spliceosome recruté au site de transcription épisse efficacement l'intron sur place. La même hybridation dans la lignée MUT montre un enrichissement nul au site de transcription. Le spliceosome n'est donc pas présent au site de transcription du gène non épissé. Plus surprenant, ces snARNs semblent sous-enrichis par rapport à un site aléatoire dans le noyau. Le spliceosome serait moins abondant autour d'un gène non épissé que ce que l'on attend par le hasard. Il faut remettre cela dans le contexte de la cellule humaine dont plus de 90% des gènes possèdent des introns (données EnsEmbl).

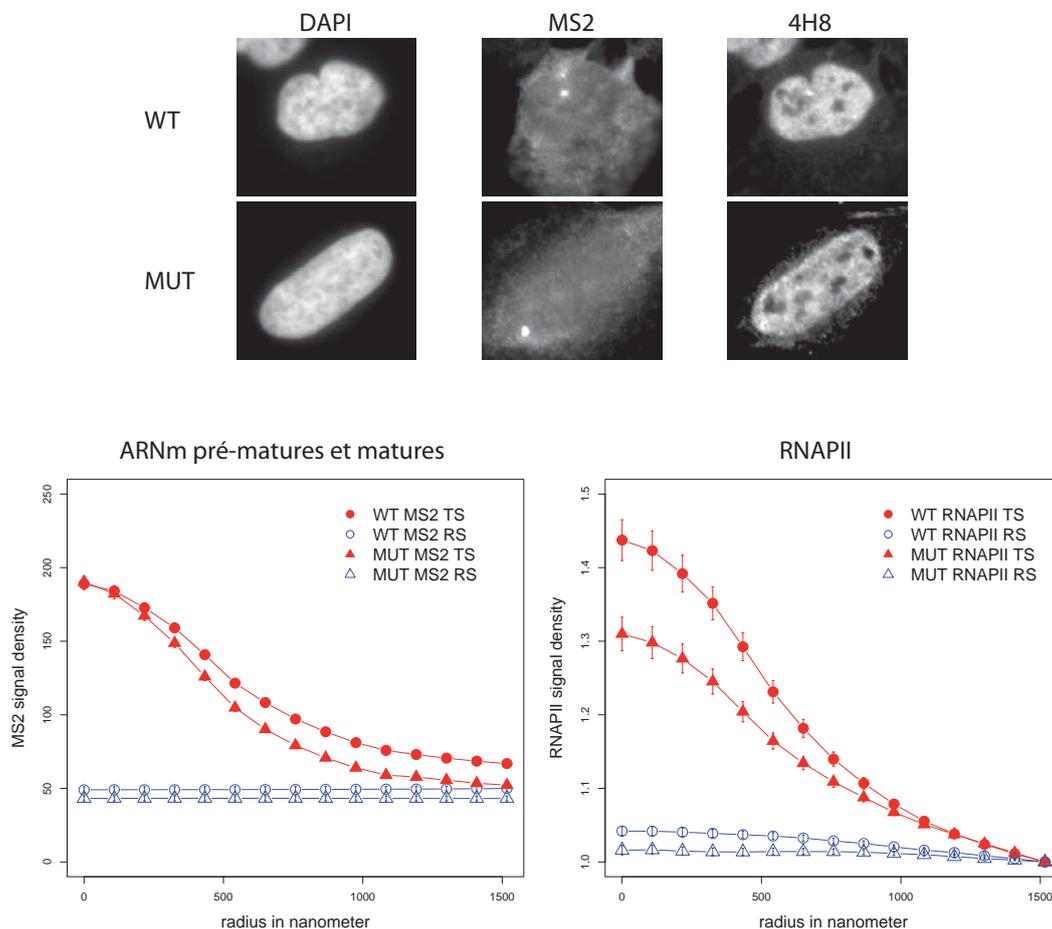


FIGURE 6.5 – Différentes densités de RNAPII aux sites pour une même densité de transcrits. 197 cellules WT et 174 cellules MUT ont été analysées par FISH pour le MS2 ; 95 cellules WT et 116 cellules MUT ont été analysées par IF (anticorps 4H8) pour la RNAPII. En haut, les images du FISH et de l'immunofluorescence anti-RNAPII. En bas, les graphiques des analyses. Les ronds correspondent à la lignée WT, les triangles à la lignée MUT. Les ronds et triangles ouverts sont les sites contrôles, les ronds et triangles pleins, les sites de transcriptions.

Par le hasard, le site contrôle tombe donc le plus souvent à proximité d'un gène transcrit et épissé, cotranscriptionnellement ou pas. Cela explique pourquoi on trouve une déplétion de ces snARNs au site de transcription non épissé par rapport au contrôle. Ces mêmes snARNs U4 et U6 regardés séparément donnent un résultat identique pour la lignée WT et légèrement différent pour la lignée MUT (figure 6.6 et données non présentées). En effet, on n'observe pas d'enrichissement au site de transcription non épissé, sans pour autant voir une déplétion. Cela peut s'expliquer par le fait qu'en imagerie, l'enrichissement en un point donné est dépendant de la qualité du signal. Un FISH dirigé contre 3 snARNs (4 sondes) donne un signal beaucoup plus fort et net qu'un FISH contre un seul ARN (une sonde). Dans le cas des snARNs U4 et U6 seuls, on est limité par la dynamique de l'image qui n'est pas très grande. De ce fait, l'enrichissement observé est moins fort (comparer U4 et trisnRNP) et de même le sous-enrichissement est sans doute sous-évalué.

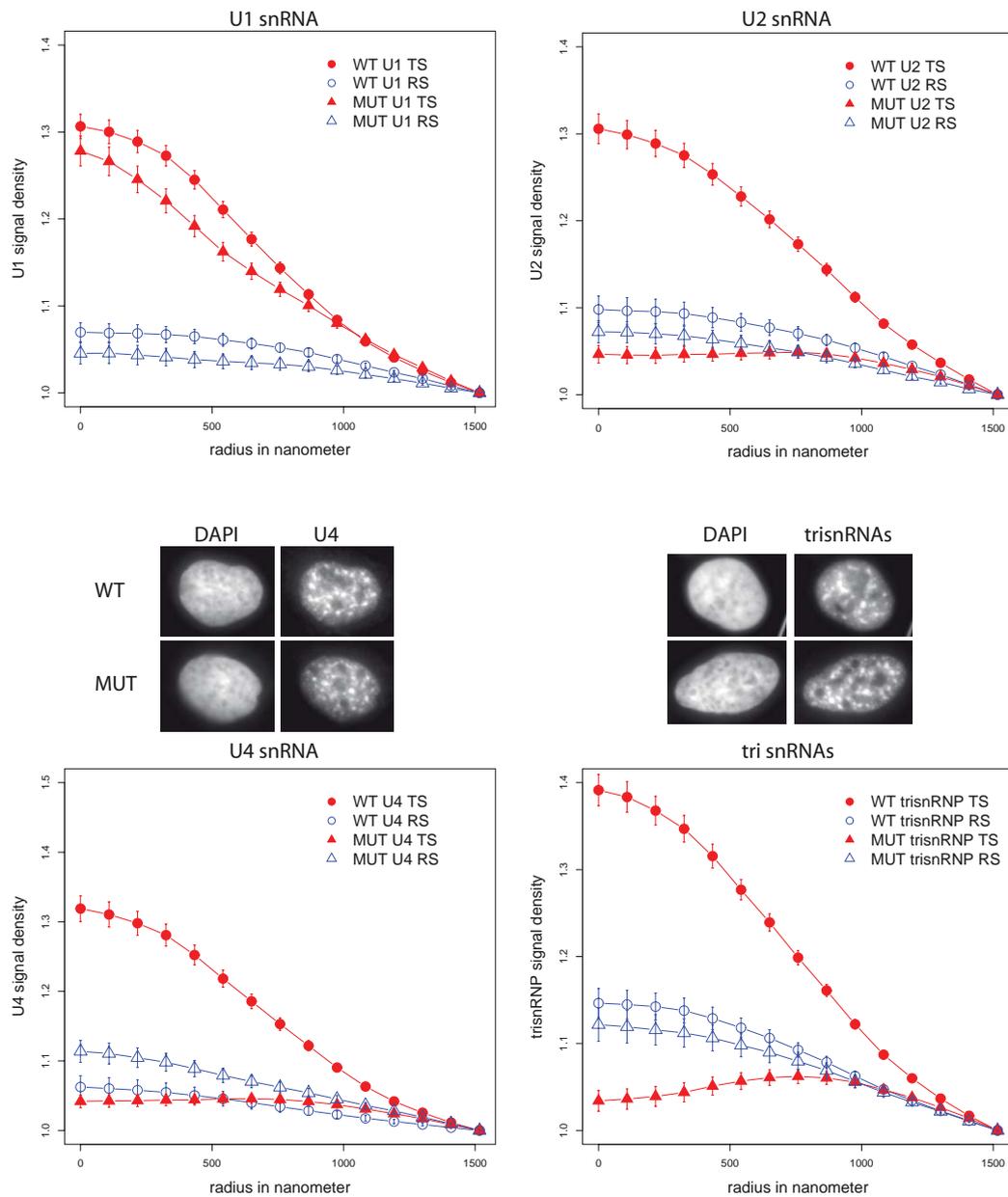


FIGURE 6.6 – L'ARN U1 est le seul ARN de l'épissage à être enrichi au site de transcription. 146 cellules WT et 118 cellules MUT ont été analysées en FISH U1. 138 cellules WT et 126 cellules MUT ont été analysées en FISH U2. 136 cellules WT et 157 cellules MUT ont été analysées en FISH U4. 167 cellules WT et 127 cellules MUT ont été analysées en FISH U4, U5 et U6.

6.2.2 Le snARN U2

Le snARN U2 a été observé également par FISH. Deuxième étape de l'épissage et première étape nécessitant de l'énergie, sa présence n'est possible que si la première étape engagée par U1 est réalisée. U2 est effectivement bien recruté sur un site de transcription épissé (figure 6.6). Son enrichissement au site est comparable à celui de U1. En revanche, le snARN U2 n'est pas plus enrichi au site de transcription que sur un site contrôle nucléaire. Le FISH U2 étant réalisé avec une seule sonde, cet écart pourrait s'avérer plus grand en réalité (cf l'explication donnée pour U4) et U2 pourrait être également sous-représenté sur un site de transcription non épissé. Cette absence d'enrichissement n'est pas due à une mutation du point de branchement, la mutation ayant permis de mettre l'intron dans la phase ouverte de lecture se trouvant bien en amont. Quoi qu'il en soit, ni U2, ni U4, ni U5, ni U6 ne sont enrichis sur un site de transcription non épissé, ils pourraient même y être sous-représentés par rapport à un autre endroit du noyau, en-dehors des nucléoles.

6.2.3 Le snARN U1

Sur un site de transcription qui épisse, l'ARN U1 est enrichi par rapport à un site contrôle dans le noyau. Etant donné que tous les autres snARNs y sont présents, ce résultat semble aller de soi. Mais contrairement à tous les autres UsnARNs de l'épissage, l'ARN U1 est toujours enrichi sur un site de transcription qui n'épisse pas. Cet enrichissement est d'autant plus surprenant qu'il est comparable à celui observé sur la lignée WT. Cette expérience a été menée plusieurs fois et l'enrichissement a toujours été observé sur les deux lignées. Parfois, on a observé un peu moins d'enrichissement dans la lignée MUT. La dynamique des images ne semble pas être le problème car elle a une grande amplitude et ce, pour deux raisons : deux sondes sont dirigées contre le snARN U1 et celui-ci est le plus abondant de tous les snARNs (1 million de copies par noyau [358]). La présence de l'ARN U1 au site de transcription rejoint l'observation donnée par les résultats biochimiques, à savoir U1 est associé à l'ARN polymérase II transcriptionnellement inactive. Cette présence au site n'est pas due à la persistance de l'hybridation de l'ARN U1 au site 5'ss. Le site 5'ss de l'intron de la construction MUT contient déjà un mésappariement par rapport à la séquence consensus en position +4 (figure 6.2 B). En rajoutant deux mutations ponctuelles sur les deux nucléotides principaux, on fragilise encore plus l'hybridation. Or, des travaux ont montré d'autres

mutations ponctuelles sur le site 5'ss qui invalident fortement la reconnaissance par U1 [169, 59]. Dans ces expériences, un mésappariement en position +3 dans la séquence du site donneur abolie complètement l'épissage par la fragilisation de l'hybridation 5'ss-U1. Les mutations des positions les plus conservées du site donneur et le mésappariement naturel de l'intron 1 de la β -globine portent au nombre de 3 les mésappariements qui déstabilisent sévèrement l'hybridation 5'ss-U1. De plus, si l'hybridation 5'ss muté-U1 persistait faiblement malgré tout, on s'attendrait à avoir un taux d'ARNm produits beaucoup plus important dans le cas de la lignée MUT pour maintenir un enrichissement de l'ARN U1 comparable à celui du site de transcription épissé. Or, la production d'ARNm au site de transcription est identique entre les deux lignées. La présence de l'ARN U1 sur le site de transcription non épissé ne peut donc pas provenir de la persistance d'une hybridation 5'ss muté-U1. Les résultats des différents FISH montrent que les snARNs de l'épissage sont tous recrutés sur un site de transcription épissé. Ces résultats confirment les premières données de microscopie montrant l'intron rapidement dégradé à la périphérie du site. L'épissage du gène de la lignée WT est bien cotranscriptionnel. D'autre part, le splicéosome n'est pas recruté sur un gène ne nécessitant pas d'épissage. Le seul snARN toujours présent est l'ARN U1.

6.2.4 La snRNP U1

Afin de vérifier si, comme pour la biochimie, l'ARN U1 est accompagné des protéines de la U1snRNP, j'ai testé par immunofluorescence la présence des trois protéines spécifiques de la U1snRNP. La figure 6.7 montre un enrichissement de la protéine U170k au site de transcription dans la lignée WT comme dans la lignée MUT. De même, la protéine U1A est enrichie au site qu'il y ait épissage ou pas. On observe donc un comportement identique des protéines de la snRNP U1 à celui de l'ARN U1. La snRNP U1 semble donc être recrutée au complet sur le site de transcription en absence d'épissage. Pourtant, on constate une différence non négligeable d'enrichissement au site entre le site de transcription épissé et non épissé. Cela est délicat à interpréter dans la mesure où l'interaction entre l'ARN U1 et la protéine U1A est très forte ($K_d \approx 10^{-9}$) [193]. On sait par ailleurs que toutes les protéines U1A ne sont pas en complexe avec la snRNP U1 mais qu'on les trouve aussi associées avec le facteur PSF (polypyrimidine-tract binding protein-associated factor), p54nrb et p68 sous la forme du complexe SF-A (snRNP-free U1A) [192, 183]. Ce complexe jouerait un rôle

dans la polyadénylation des ARNm. Si beaucoup de U1A est recruté sur le site de transcription épissé (dans certaines expériences le recrutement atteint 1,5 unités arbitraires), peut-être est-ce dû en partie à la snRNP et en partie au complexe SF-A. Il faudrait vérifier la présence de PSF et p54nrb sur le site de transcription non épissé.

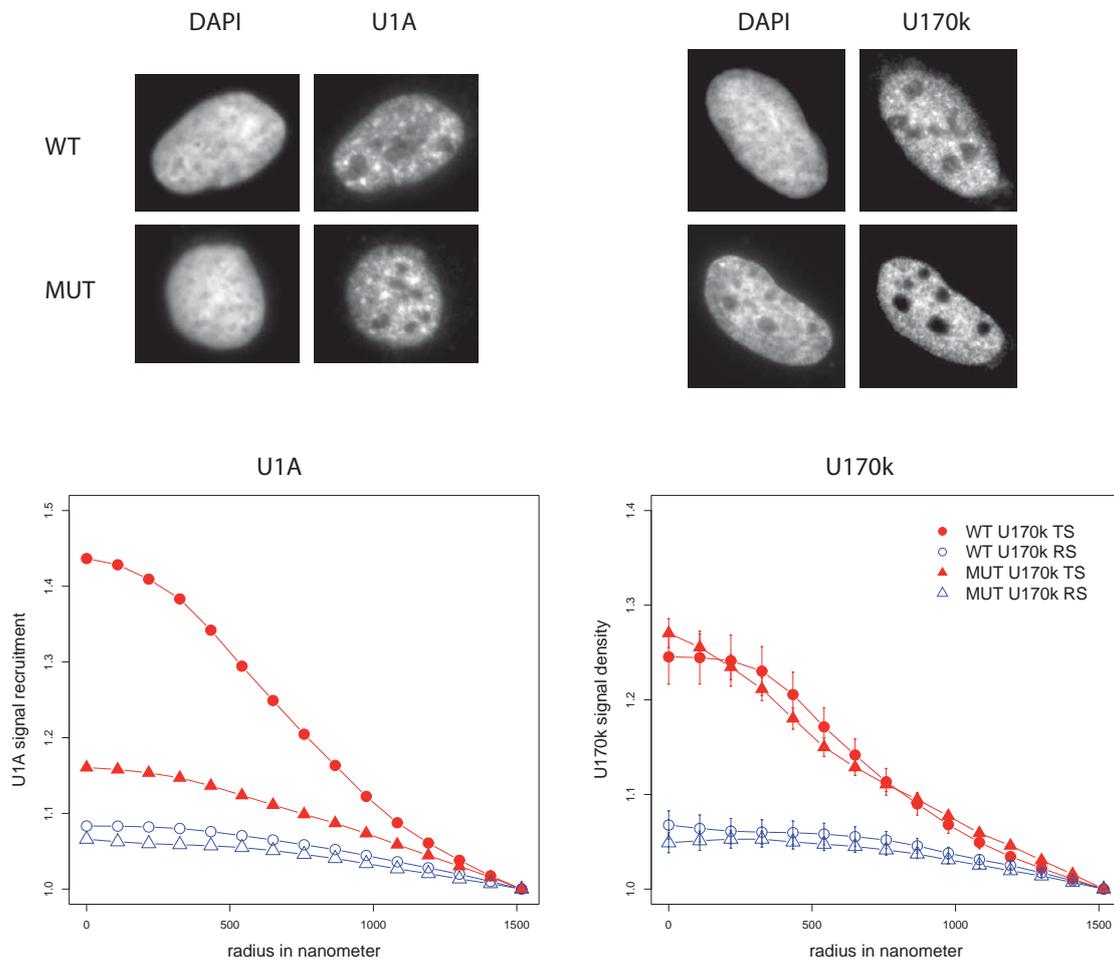


FIGURE 6.7 – La snRNP U1 est recrutée au site de transcription indépendamment de l'épissage. 373 cellules WT et 273 cellules MUT ont été analysées en immunofluorescence anti-U1A. 51 cellules WT et 120 cellules MUT ont été analysées en immunofluorescence anti-U170k. En haut, les images des immunofluorescences. En bas, les graphiques correspondants aux analyses.

6.3 Protéines annexes

Un gène qui n'épisse pas recrute tout de même et uniquement la U1snRNP. D'autres facteurs en-dehors des UsnRNPs sont requis pour participer à l'épissage, notamment plusieurs protéines SR. SF2/ASF est une protéine que l'on retrouve dans les complexes A et B du splicéosome [70, 18]. C'est une protéine SR qui stimule la liaison de la snRNP U1 et module le choix du 5'ss. Cette activité serait portée par l'interaction entre ASF/SF2 et U170k [154]. Outre ses interactions protéine-protéine, SF2/ASF a la capacité de se lier à des SRE sur l'ARN, le plus souvent des ESE mais aussi parfois des ISS [200]. Puisque SF2/ASF interagit avec U170k, j'ai regardé son enrichissement au site de transcription. Il s'avère que cette protéine est toujours présente qu'il y ait épissage ou non (figure 6.8). On voit en effet un fort recrutement au site de transcription, bien que celui-ci soit différent en fonction de la lignée cellulaire.

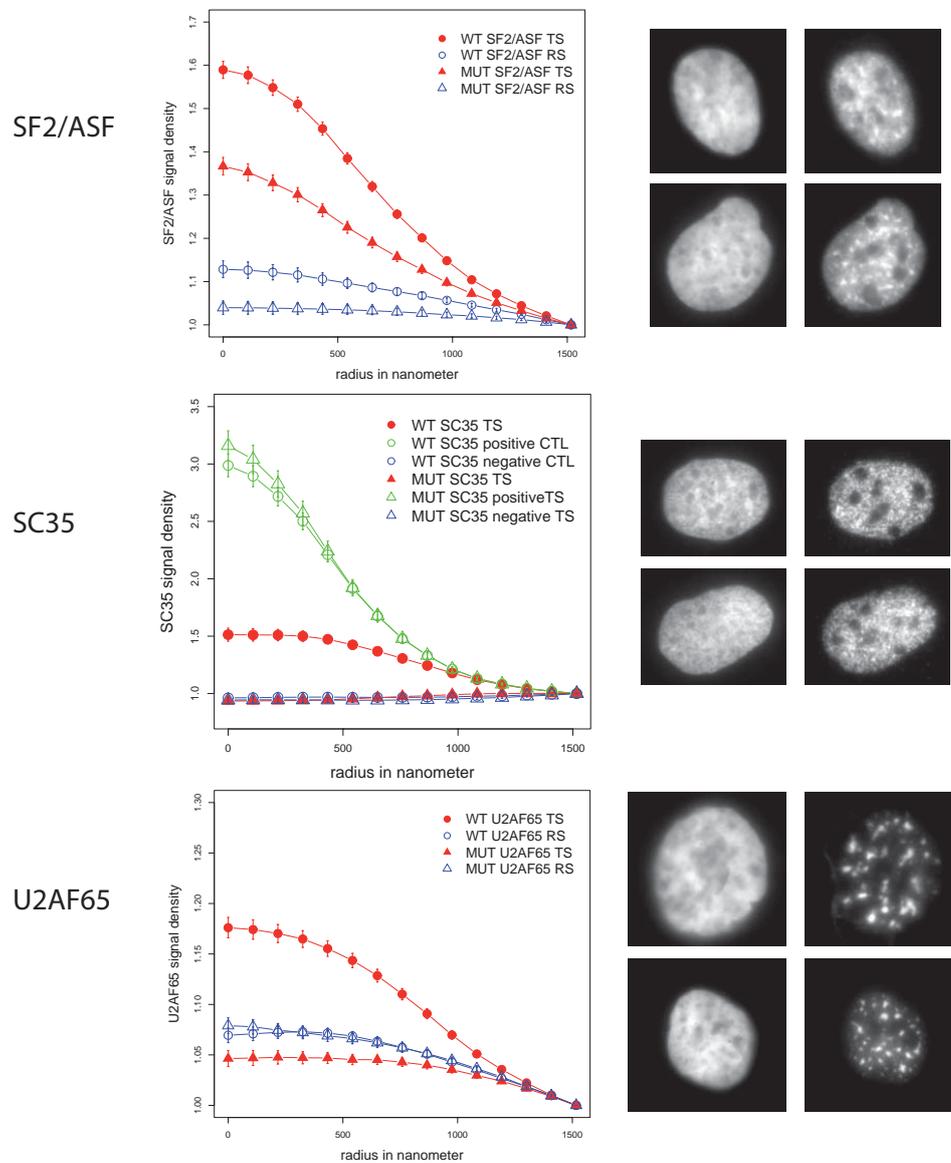


FIGURE 6.8 – **Recrutement spécifique de certaines protéines SR au site de transcription non épissé.** 111 cellules WT et 88 cellules MUT ont été analysées en immunofluorescence anti-SC35. 109 cellules WT et 110 cellules MUT ont été analysées en immunofluorescence anti-SF2/ASF. 195 cellules WT et 217 cellules MUT ont été analysées en immunofluorescence anti-U2AF65.

La protéine SC35 (Splicing Component 35) est aussi une protéine SR impliquée dans l'épissage, faisant partie des complexes A et B du splicéosome [70, 18]. Comme ASF/SF2, SC35 se lie à des ESE pour favoriser l'inclusion des exons lors de l'épissage. D'autre part, SC35 est un composant des speckles, ces structures de stockage de composants d'épissage [172]. L'analyse a été réalisée sous un mode un peu différent puisque le signal SC35 forme des foci (les speckles) et entre ces foci, des zones où il n'y a pas de signal. J'ai donc choisi deux contrôles, un absolument négatif et un autre absolument positif pour le signal SC35. On constate de cette façon que la protéine SC35 est recrutée au site de transcription épissé. Le contrôle positif représentant 100% de speckles et le négatif 0%, cet enrichissement en SC35 au site signifie que le site de transcription épissé est souvent mais pas toujours associé à un speckle. En revanche, le site de transcription non épissé n'accumule aucun signal SC35. Ce site n'est donc jamais associé à un speckle. Étonnamment, ASF/SF2 est aussi un facteur que l'on trouve associé aux speckles. Pour autant le comportement de ces deux protéines, SC35 et ASF/SF2, est différent lorsqu'il n'y a pas d'épissage. Néanmoins, si ASF/SF2 est toujours enrichi au site non épissé, il l'est dans une moindre mesure comparé à la lignée WT. Cette perte d'ASF/SF2 provient peut-être de l'absence de speckles autour du site non épissé. La protéine U2AF (U2 small nuclear ribonucleoprotein auxiliary splicing factor) n'est pas une protéine SR mais elle possède un domaine RS (Arginine-Sérine). Elle est composée de deux sous-unités de 35 et 65kDa. U2AF65 reconnaît la séquence polypyrimidine et recrute U2snRNP au point de branchement alors que U2AF35 marque le dinucléotide AG au site accepteur [316, 212]. U2AF65 interagit avec les snRNP U1 et U2 et aussi avec le point de branchement. Il a été démontré que U2AF65 était associé avec la RNAPII dans le complexe d'initiation avant le début de la transcription [310]. Il pourrait ainsi scanner les ARNs naissants à la recherche d'une séquence polypyrimidine. La présence de U2AF65 dans ce contexte aurait pour effet de diminuer le temps de pause de la RNAPII. U2AF65 permet aussi l'export des ARNs avec ou sans introns chez la *Drosophile* [28]. En regardant par immunofluorescence la protéine U2AF65, on constate qu'elle est bien recrutée sur le gène épissé (figure 6.8). En revanche, U2AF65 est absent du site de transcription non épissé. Pour autant, il n'y a pas de problème d'export des ARNs non épissés. Peut-être que U2AF65 n'est pas requis pour l'export des ARNs sans introns. D'autre part, la présence de la U1snRNP sur un gène non épissé ne suffit pas à faire venir U2AF65 et la U2snRNP. La présence de U1snRNP ne semble donc pas en lien

direct avec l'épissage.

Tat-SF1 est l'homologue humain de Cus2 (U2snRNP) chez la levure. Il fait des interactions avec la snRNP U2. Chez la levure, il est indispensable pour les réarrangements de l'ARN U2 fixé au point de branchement qui doit se déplier et se replier pour créer le site catalytique. Nous avons donc voulu regarder si Tat-SF1 était recruté au site de transcription avec ou sans intron pour voir s'il pouvait être présent avec la snRNP U1 pour faire venir la snRNP U2. Comme nous ne disposons pas d'anticorps performants en immunofluorescence, nous avons transfecté le plasmide codant pour la protéine Tat-SF1-FLAG. La figure 6.9 montre les résultats d'une immunofluorescence contre l'étiquette FLAG dans les lignées MUT ou WT. Tout d'abord l'expression est variable d'une cellule à l'autre du fait de la transfection qui n'est pas homogène. Les images ont été à chaque fois prises sur des cellules exprimant un taux de protéines transfectées au moins 4 fois supérieur au bruit de fond. Sur ces images, le recrutement est significatif quoique plutôt faible au site de transcription, quelle que soit la lignée. Il semble donc qu'il y ait du TAT-SF1 en faible quantité sur les sites de transcription, indépendamment ou non de l'épissage. Ces résultats doivent tout de même être pris avec précaution car ils reflètent uniquement le comportement de la protéine transfectée et donc surexprimée.

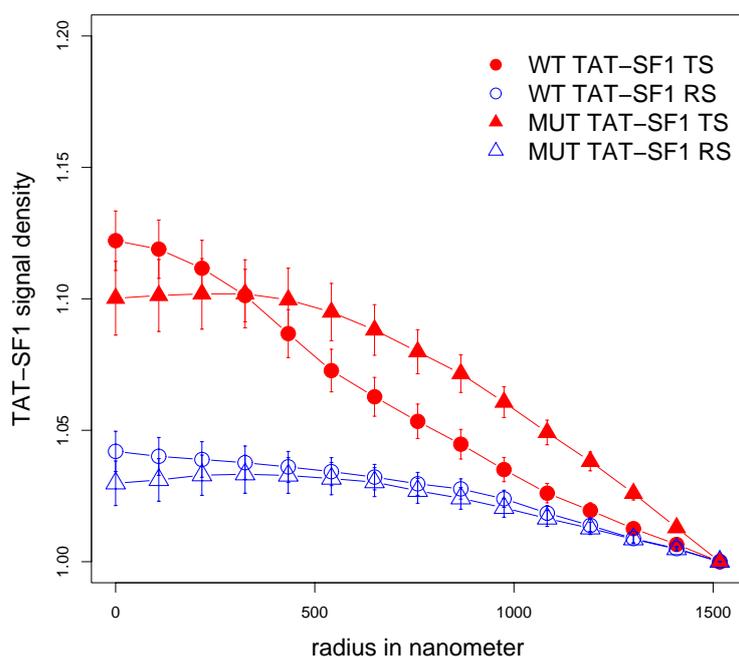


FIGURE 6.9 – **Analyse radiale de TAT-SF1-GFP.** Les cellules ont été transfectées avec le plasmide TAT-SF1-GFP le matin et induites à la doxycycline sur la nuit. Elles ont été fixées le lendemain matin et un FISH MS2 a été réalisé le soir. 60 cellules WT et 61 cellules MUT ont été analysées. En rouge, les sites de transcription, en bleu, les sites contrôles. Les ronds représentent les cellules WT et les triangles, les cellules MUT.

6.4 Analyse radiale de l'ARN U1 et de l'ARN B2 sur les sites actine chez la souris

L'ARN B2 est un ARN non codant de la souris qui est exprimé lors d'un stress comme le choc thermique [5]. Cet ARN a la propriété de lier la RNAPII et d'inhiber son activité avant la formation du complexe de pré-initiation [80, 79]. L'aptamère FC* inhibe aussi la RNAPII et entre en compétition avec l'ARN B2 pour la liaison à la RNAPII [145]. Or, cet ARN inhibe la RNAPII en se glissant dans le sillon de l'ADN, à l'intérieur de celle-ci. L'ARN U1, quant à lui, aurait plutôt un effet activateur sur la transcription [166]. Dans la perspective d'étudier le comportement de l'ARN U1 en présence de l'ARN B2 dans le contexte de stress, j'ai d'abord réalisé un FISH contre ces deux ARNs sur un gène aisément visible par microscopie, le gène de la β -actine sur des cellules NIH3T3. Pour visualiser les sites de transcription des gènes actine, il a fallu priver les cellules de sérum sur une nuit puis rajouter le lendemain matin du sérum à hauteur de 20% pendant 30 mn. On voit que

dans ces conditions, l'ARN B2 est exprimé par rapport à une cellule non activée (figure 6.10 A) indiquant que cette activation est sans doute ressentie comme un signal de stress. Mais on observe également que les quantités de B2 exprimé sont variables d'une cellule à l'autre.

Les images ont été analysées en U1 et en B2 (figure 6.10 B). On observe sur le graphique, le fort recrutement de l'ARN U1 (triangle) sur les sites de transcription actine. L'ARN U1 est particulièrement enrichi sur ces sites, ce qui paraît logique car les gènes actine contiennent 5 introns. L'ARN B2 (rond), bien qu'il ne s'agisse pas d'un choc thermique, est transcrit (figure 6.10) et il est même présent au site actine en cours de transcription. Certes, l'enrichissement n'est pas grand mais il est significatif (p-value : 0,002). Les cellules ayant été soumises à la privation en sérum puis à une double dose de sérum sont sans doute en condition de stress, ce qui peut expliquer la présence du transcrit B2 dans la cellule. Ce recrutement de B2 au site actine est plus surprenant car l'actine est le gène utilisé comme contrôle négatif dans les expériences démontrant le rôle de B2 au cours du choc thermique. Bien-sûr, l'actine ne se comporte pas de la même façon en choc thermique et dans les conditions de stress que j'ai utilisées. Mais si B2 est exprimé, il va sans doute se lier à la RNAPII, notamment sur un gène sensé être réprimé. Deux explications peuvent alors être proposées : soit la condition d'activation est effectivement un stress mais « n'active pas » le rôle inhibiteur de B2 ; soit son inhibition est dose-dépendante et dans ces conditions, on est à un niveau trop faible pour voir une inhibition de la RNAPII. Il faudrait faire un FISH B2 sur le gène de choc thermique hsp70 pour voir si celui-ci est beaucoup plus transcrit dans la cellule et s'il est beaucoup plus recruté au site de transcription Hsp70. Un FISH U1 sur les mêmes cellules permettrait également de voir quel est le comportement de l'ARN U1 sur ces gènes sans introns.

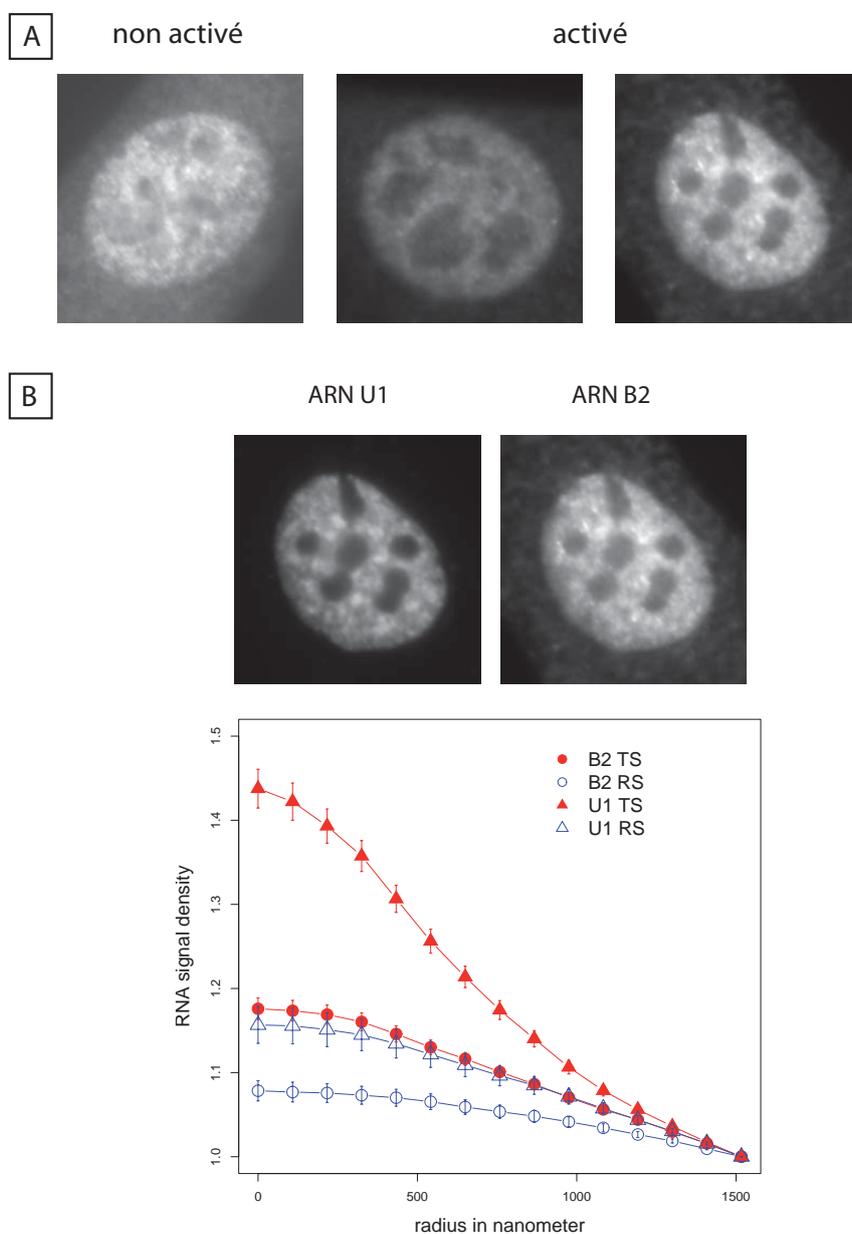


FIGURE 6.10 – **B2 sur des cellules NIH3T3 au site actine.** Les cellules NIH 3T3 ont été privées de sérum sur la nuit puis mis en présence de 20% sérum pendant 30mn. Après avoir été fixées, un FISH contre l'ARN B2, l'ARN U1 et l'ARN α -actine a été pratiqué sur les cellules NIH 3T3. A. FISH B2 sur des cellules non activées (conditions normales de sérum) et activées. B. FISH U1 et B2 sur une cellule activée. 90 cellules ont été analysées sur les sites de transcription de l'actine. En rouge, les sites de transcription, en bleu, les sites contrôles. Les ronds représentent le signal B2 et les triangles, le signal U1.

6.5 Rôle du CTD de la RNAPII dans l'enrichissement du snARN U1 au site de transcription

Le CTD de la RNAPII est impliqué dans le recrutement de nombreux facteurs de maturation de l'ARN pré-messager (cf. intro). Pour savoir si l'ARN U1 pourrait être enrichi au site de transcription via le CTD de la RNAPII, des analyses d'images ont été faites avec une RNAPII dont le CTD ne comporte que 5 répétitions (RNAPII- $\Delta 5$) [96]. Cette RNAPII possède une étiquette GFP en C-terminal de la sous-unité Rpb1 et surtout une mutation ponctuelle qui lui confère une résistance à l' α -amanitine [139]. Les cellules ont été transfectées soit avec la RNAPII-GFP sauvage, soit avec la RNAPII- $\Delta 5$. En présence d' α -amanitine, la RNAPII endogène est bloquée et la cellule ne possédant pas la RNAPII résistante meurt [222, 267, 102]. La RNAPII- $\Delta 5$ est capable de transcrire un gène inséré en multicopies dans un génome [210].

Après un traitement à l' α -amanitine de 4h à 100 μ g/mL, les cellules n'ayant pas été transfectées sont en train de mourir et ne présentent aucun site de transcription. Les cellules fixées sont analysées en FISH contre l'ARN U1. Le signal GFP donne la localisation de la sous-unité Rpb1. Celle-ci, après un tel traitement, se trouve dans le noyau et dans le cytoplasme. En effet, il y a sans doute plus de sous-unités Rpb1 produites que de complexes RNAPII possibles et la sous-unité Rpb1 ne possède pas de signal de rétention nucléaire. Ce patron est légèrement différent entre la Rpb1 sauvage et la Rpb1- $\Delta 5$. En effet, on voit un signal plus nucléaire pour la Rpb1 sauvage. Néanmoins, le signal GFP reste très faible dans les deux cas. Il faut remarquer que les cellules présentant un fort signal GFP, et donc une forte expression, sont toutes en apoptose. Une trop forte expression de Rpb1 muté ou non entraîne donc un état trop fortement perturbé pour la cellule. D'autre part, la proportion de cellules en apoptose est beaucoup plus importante dans le cas de la transfection avec la sous-unité Rpb1- $\Delta 5$. Il est évident qu'un CTD tronqué empêche la cellule de produire des ARNm matures et donc conduit la cellule à la mort [45].

Par un FISH sur l'ARN U1, on a regardé son enrichissement au site en fonction du type de RNAPII transfectée. Les premiers résultats ne sont pas très concluants comme le montre la figure 6.11. On voit un recrutement positif de l'ARN U1 avec la RNAPII sauvage résistante à l' α -amanitine sur la lignée WT (figure 6.11 Aa). Cependant, les valeurs mesurées sont très faibles. L'enrichissement

de l'ARN U1 est significatif dans la lignée WT mais peu dans la lignée MUT (voir annexe A sur les résultats significatifs). En revanche, il n'y a aucune différence entre le site de transcription et un site contrôle lorsqu'on utilise la RNAPII $\Delta 5$, quelle que soit la lignée observée (figure 6.11 Ab). Il est difficile de porter du crédit à ces premières expériences car il y a très peu de RNAPII au site de transcription comme on le constate sur la figure 6.11 B. D'une façon générale, le signal GFP est très faible dans la cellule et il y a peu de différence entre le site de transcription et un site contrôle. On peut d'ailleurs remarquer que le signal U1 obtenu dans la lignée WT avec le CTD intégral, seul résultat significatif, correspond au signal GFP le plus fort. La faible concentration de RNAPII-GFP n'empêche pas la production d'ARNm comme le présente la figure 6.11 C. Les densités de signaux ARNm sont comparables à ce qu'on trouve dans les autres expériences (la différence d'échelle provient de du type d'image analysée : de 1 à 10 000 en 16 bits et de 1 à 500 en 8 bits). Cette expérience doit être renouvelée avec comme contrôles, les autres snARNs de l'épissage pour étudier leur comportement avec la RNAPII $\Delta 5$. L'ARN U2 ou la tri-snRNP ne sont en effet pas recrutés sur le gène sans intron. Les densités obtenues nous permettront de comparer avec l'ARN U1. Si on ne peut conclure sur le rôle du CTD à travers cette expérience, on peut tout de même dire que l'ARN U1 n'est pas recruté par l'ARNm puisque dans la lignée MUT avec la RNAPII GFP sauvage, la densité de signal MS2 est très forte et pourtant le recrutement de l'ARN U1 n'est pas significatif. L'ARN U1 n'est donc définitivement pas enrichi sur un site de transcription sans épissage par une hybridation persistante avec un site donneur sur l'ARNm.

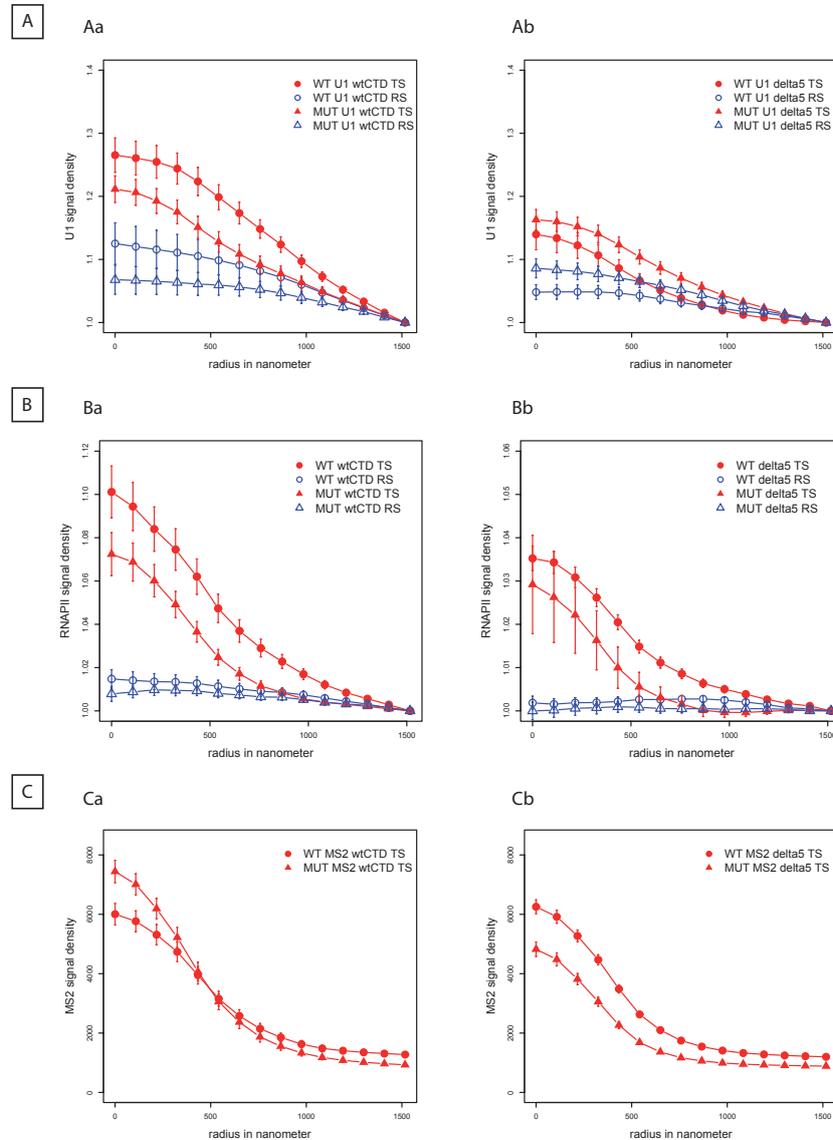


FIGURE 6.11 – **Expression d'une RNAPII-CTD-GFP ou RNAPII- Δ 5-GFP dans les lignées MUT et WT.** Analyse sur l'ARN U1. Les cellules ont été transfectées avec la construction Rpb1-wtCTD ou Rpb1- Δ 5CTD, puis traitées à l' α -amanitine et induites à la doxycycline. Après avoir été fixées, un FISH contre les répétitions MS2 et contre l'ARN U1 a été réalisé. Les images obtenues ont fait l'objet d'une analyse radiale. Les figures (a) représentent les transfections RNAPII-CTD-GFP et les figures (b), les transfections RNAPII- Δ 5-GFP. A. Densité de signal U1 au site de transcription et aux sites contrôles. B. Densité de signal GFP. C. Densité de signal MS2, représentant l'ARN pré-mature et mature.

Quatrième partie

Discussion et perspectives

6.6 Association RNAPII – snRNP U1

Les résultats présentés démontrent l'association de l'ARN U1 avec la RNAPII. Par immunoprécipitation anti-RNAPII, U1 est co-immunoprécipité. Cette association ne concerne pas l'ARN U1 seul mais l'ensemble de la snRNP U1. Cette association a également été démontrée par Das et al. dans des extraits nucléaires de cellules HeLa en présence d'ATP [65]. L'ATP est utilisé pour désengager les spliceosomes potentiellement engagés sur les pré-ARNm de l'extrait nucléaire. Leur immunoprécipitation RNAPII co-immunoprécipite beaucoup de protéines, identifiées par spectroscopie de masse et Western blot. Ils trouvent, entre autres, les protéines de la snRNP U1, une protéine de la snRNP U2 (U2-A'), des protéines SR, des protéines hnRNP, des facteurs de transcription ou reliés. D'autre part, Tian montrait déjà en 2001 que la protéine U1-70k co-immunoprécipitait la RNAPII [302]. Par deux méthodes différentes, on constate l'association de la snRNP U1 avec la RNAPII. Mes expériences de crosslink et celles de Kameoka confirment également l'existence d'un complexe RNAPII-snRNP U1 *in vivo* [137]. Dans ces expériences, Kameoka utilise un ARN comprenant un site 5'ss pour assembler le spliceosome. Cet ARN comprend un groupe benzo-phénone photoactivable qui permet de coupler de façon covalente les éléments qui s'y associent. De cette manière, il obtient deux types de complexes, chacun contenant la RNAPII et la snRNP U1 et d'autres facteurs d'épissage et de transcription sans nécessité d'ATP. De plus, mes résultats démontrent que cette association a lieu indépendamment de l'état d'activité de la RNAPII puisque la RNAPII utilisée dans mes expériences est transcriptionnellement inactive. On remarquera d'ailleurs que Das et al. n'ont pas eu de problème à extraire le complexe RNAPII-snRNP U1. Peut-être en ont-ils perdu au cours de leurs purifications, néanmoins cela signifie que la liaison qui unit la RNAPII à la snRNP U1 est suffisamment forte pour ne pas être perdue dans un protocole d'extraction de RNAPII transcriptionnellement active.

6.7 Pré-requis de l'interaction RNAPII – snRNP U1

Pour le moment, nous n'avons que quelques éléments de réponse. Premièrement, on ne sait pas si ce complexe RNAPII-snRNP U1 existe pour lui-même ou n'est que la conséquence d'une association plus grande. Cela est tout à fait concevable dans le contexte transcriptionnel. En mi-

tose, avec une RNAPII inactive, les associations sont en théorie plus restreintes. Pour autant, la RNAPII des cellules mitotiques se trouve aussi sous ses deux formes *RNAPII_O* et *RNAPII_A* et pléthore de protéines sont associées au CTD de la RNAPII dont des protéines de l'épissage telles que SF2-ASF. Or on sait que SF2-ASF se lie à la protéine U1-70k [154, 65]. Cette association permet le recrutement plus efficace de la snRNP U1 au site 5'ss [77]. Si U1-70k est présent sur les immunoprécipitations RNAPII en mitose, je n'ai pas regardé SF2/ASF. Mais on sait qu'il est présent sur les immunoprécipitations de l'article de Das et.al. Il faudrait donc déterminer si SF2/ASF joue un rôle dans l'association RNAPII – snRNP U1. Pour cela, on pourrait rajouter la protéine SF2/ASF sur des extraits nucléaires en concentrations croissantes et observer la proportion de RNAPII liée à l'ARN U1. Parmi les composants du splicéosome, je n'ai de résultats biochimiques qu'avec l'ARN U2 qui est peu enrichi. Par spectroscopie de masse, Das ne trouve que la protéine U2-A' [65]. Dans mes Northern blots et dans les résultats de spectroscopie de masse, les quantités sont faibles par rapport à la snRNP U1 et il n'est pas toujours aisé de savoir s'il s'agit d'un bruit de fond ou d'un réel signal U2. Dans les cellules synchronisées, on peut penser que l'interaction du snARN U2 avec la RNAPII existe car elle pourrait témoigner d'un voisinage avec le splicéosome. Dans les cellules mitotiques, le jugement est plus délicat. D'une part, les résultats de l'anticorps commercial sont certes meilleurs que les ascites (et enrichissent moins l'ARN U2), d'autre part, on voit quand même du U2. Cependant l'association de la snRNP U1 ne peut être la conséquence de l'association des UsnRNPs du splicéosome avec la RNAPII, d'abord parce que les extractions de splicéosomes ne contiennent pas de RNAPII [18, 52] et inversement [65, 301]. Ensuite la RNAPII est déjà un énorme complexe à elle-seule, il semble difficile qu'elle soit associée avec un complexe tout aussi gros qu'elle, sinon plus, dans la cellule. La présence du snARN U2 dans les couplages covalents dans les cellules mitotiques pourraient provenir des granules d'interchromatine mitotique (MIG). En mitose, la *RNAPII_O* peut se retrouver dans les MIG qui sont les speckles mitotiques. Elle y côtoie entre autres les snARNs de l'épissage. Les couplages covalents pourraient maintenir des interactions dans ces granules et permettre d'observer une interaction de l'ARN U2, qui est l'ARN le plus abondant après l'ARN U1, avec la RNAPII. Deuxièmement, la phosphorylation du CTD pourrait jouer un rôle dans l'interaction RNAPII – snRNP U1 bien que celui-ci reste à élucider. En effet, l'article de Das a basé sa recherche sur des immunoprécipitations avec l'anticorps 8WG16. Les

auteurs l'ont choisi car de tous les autres anticorps testés (N20, H5, H14), le 8WG16 a été le seul à permettre de voir les protéines co-immunoprécipitées sur un gel de Coomassie. De même, Kameoka montre une immunoprécipitation 8WG16 qui enrichit spécifiquement l'ARN U1, plus que les anticorps N20, H5 et H14 qui co-immunoprécipitent les autres snARNs [137]. Or, dans mes expériences, c'est ce même anticorps qui me permet d'enrichir largement l'ARN U1 et ses protéines par rapport aux autres anticorps. Cet anticorps immunoprécipite les deux formes de la RNAPII (*RNAPII_O* et *RNAPII_A*), tout comme l'anticorps N20. A la différence près que l'anticorps 8WG16 reconnaît le CTD sur les peptides non phosphorylés et le N20 sur la partie N-terminale de Rpb1. Se pourrait-il que l'anticorps 8WG16 modifie la conformation de Rpb1 et/ou de son CTD et que celui-ci devienne plus attractif pour la snRNP U1 ? Actuellement, Bo Gu au laboratoire, tente de faire exprimer une RNAPII-GFP avec un CTD pleine longueur (WT-CTD) ou un CTD réduit à 5 répétitions ($\delta 5$ -CTD). Il serait intéressant de faire une immunoprécipitation anti-GFP après formaldéhyde et regarder si U1 s'associe à la polymérase tronquée. L'implication du CTD dans l'association de la snRNP U1 à la RNAPII est difficile à résoudre par les anticorps mais ce n'est pas le seul problème rencontré. En effet, dans les expériences de couplages covalents au formaldéhyde, on fixe les cellules dans un état mais on utilise aussi beaucoup de détergents, des sonications, pour casser les cellules et réaliser les immunoprécipitations. Au laboratoire, Praveen Belagal a mis en évidence une perte du CTD de la RNAPII au fil des sonications. Pour ma part, j'ai déjà observé ce phénomène parfois même sans sonication avec des conditions de lyse douces. Dans ce dernier cas, cette perte est peut-être due à des protéases encore actives. En conditions natives, on a aussi remarqué que la quantité d'ARN U1 immunoprécipité était beaucoup plus importante que sur des cellules fixées. Les ARNs U1 ou U2 immunoprécipités dans ces extraits sont-ils le reflet de la réalité ou bien en a-t-on perdu au cours du protocole ? Si le CTD est dégradé, il n'est pas reconnu par l'anticorps 8WG16. S'il est coupé, il l'est potentiellement. L'un ou l'autre cas peuvent avoir une influence directe sur la quantité d'ARN U1 immunoprécipité. On ne sait donc pas à l'heure actuelle quelle est la réelle proportion d'ARN U1 immunoprécipité avec la RNAPII (certainement au moins 1 pour 1, voire plus) et si cette RNAPII a pu conserver l'intégrité de ses associations. Une autre tentative de savoir si le CTD joue un rôle dans cette interaction est d'utiliser l'imagerie avec le plasmide codant pour Rpb1 $\delta 5$ cité ci-dessus. Nous avons cependant été confrontés à des problèmes techniques

d'expression de ce plasmide et pour le moment, nous n'avons pas de résultats fiables.

6.8 La snRNP U1 présente au site de transcription indépendamment de l'épissage.

L'association de la snRNP U1 avec la RNAPII est très spécifique. Cette spécificité peut révéler une fonction, d'autant que mes expériences sont faites sur des cellules mitotiques. La RNAPII y est transcriptionnellement inactive et l'épissage très nettement réduit [284]. J'ai donc voulu investiguer la présence de l'ARN U1 et de ses protéines sur des sites de transcription, visibles par imagerie. Cette technique permet de regarder à un moment donné, les interactions dans la cellule. Ainsi, sur un site de transcription dont l'ARN est épissé, on voit un enrichissement local des snARNs du spliceosome. Sur un site de transcription dont l'ARNm ne possède pas d'introns, les snARNs de l'épissage ne sont pas présents alors que la snRNP U1 est présente au complet. De plus, la protéine U2AF65 qui reconnaît la séquence polypyrimidine avant la jonction ATP-dépendante de la snRNP U2 est également absente sur un site de transcription sans intron. Un résultat un peu similaire est obtenu sur un système identique avec un gène rapporteur VIH et des sites donneur et accepteur d'épissage sous le contrôle de LTR. Ce gène ne fait pas d'épissage cotranscriptionnel mais la snRNP U1 est présente au site de transcription. Elle semble y être recrutée par le site 5'ss car si celui-ci est muté, la snRNP U1 n'est plus enrichie (Bertrand E., communication personnelle). Dans mes expériences, la présence de l'ARN U1 n'est vraisemblablement pas due à une reminiscence de la reconnaissance ARN U1 – 5'ss. L'hybridation repose sur 7 nucléotides discontinus et la mutation complète du dinucléotide donneur quand une mutation sur la thymidine (uracile dans l'ARNm) du dinucléotide donneur suffit à abolir l'interaction avec l'ARN U1 [59]. La snRNP U1 est donc recrutée au site de transcription indépendamment de l'épissage. Ce résultat est en contradiction avec d'autres travaux qui montrent l'absence de la protéine U1-70k sur des gènes sans introns transcriptionnellement actifs par CHIP et l'absence de co-immunoprécipitation de RNAPII par U1-70k [186]. D'autres travaux, chez la levure, montrent également l'absence d'enrichissement de la snRNP U1 sur des gènes sans introns mais constatent son fort recrutement en 5' des gènes épissés [159]. Ces résultats sont paradoxaux et difficiles à expliquer. Chez la levure,

les auteurs concluent à un résultat négatif concernant la présence de la snRNP U1 sur les gènes sans introns. Pourtant, la snRNP U1 n'est pas complètement absente. On en voit, certes beaucoup moins que sur les gènes épissés, mais tout le long de ces gènes sans introns. Bien que faible, cet enrichissement a peut-être une signification biologique. Il se peut également que la comparaison des systèmes levure et homme ne soit pas appropriée concernant l'épissage car les ChIPs sur cellules HeLa n'enrichissent pas U2AF65 sur des gènes sans introns [186], contrairement à ce qu'on trouve chez la levure sur le gène sans intron PDR5 [104], à moins que ces différences ne soient en fait que des différences liées aux gènes. D'autre part, les méthodes utilisées n'ont rien en commun. L'imagerie me permet de voir si j'ai une colocalisation de différents éléments mais ne me permet pas de dire s'il y a interaction ou non. L'absence d'U1-70k sur un gène sans intron ne nous dit pas ce qu'il se passe sur l'ARN naissant ou aux alentours. Il est vrai que Listerman ne co-immunoprécipite pas de RNAPII avec un anticorps anti-U170k. Mais cet anticorps est différent de celui de Tian [302]. De plus, les auteurs ne recherchent qu'une RNAPII phosphorylée alors que Tian fait un Western blot avec N20 pour visualiser toutes les formes de RNAPII et dans mes résultats, on constate que tous les anticorps anti-RNAPII ne se comportent pas de la même façon. Ces différents résultats, s'ils apparaissent en contradiction au premier abord, ne le sont peut-être qu'à cause des méthodes utilisées et de nos lacunes actuelles. Car d'autres travaux montrent, comme mes résultats, une interaction de la snRNP U1 avec des ARNs naissants indépendamment de l'épissage dans des ovocytes amphibiens [243]. Les auteurs injectent des snARNs fluorescents dans le cytoplasme des ovocytes et ceux-ci viennent colocaliser sur les ARNs naissants. Lorsqu'un antisens du snARN U2 est injecté, l'épissage est bloqué. Les snARNs U1, U4 et U5 fluorescents injectés par la suite viennent tout de même se localiser au niveau des ARNs naissants alors que la protéine U2B" de la snRNP U2 se concentrent dans les nucléoles. Certes, dans cet article, il n'y a pas modifications de séquences et l'ARN U1 se dirige peut-être vers les ARNs naissants car il reconnaît le 5'ss d'un intron. D'ailleurs, si on enlève la tige-boucle I et le site de reconnaissance du 5'ss, l'ARN U1 ne va plus se localiser auprès des ARNs naissants. Néanmoins, cette association est indépendante de la snRNP U2, tout comme le recrutement de l'ARN U1 sur mes constructions.

La présence de l'ARN U1 au site de transcription avec ou sans intron pourrait être le fruit de son interaction avec la RNAPII. Mais elle pourrait être aussi la conséquence du recrutement via le

complexe de la coiffe. En effet, le complexe de la coiffe stimule le recrutement de la snRNP U1 [179]. Sur les sites de transcription de ces lignées, le complexe de la coiffe peut effectivement faciliter le recrutement de la snRNP U1 sur l'ARN quand celui-ci est fixé sur la RNAPII. S'il ne fait aucun doute que ce complexe soit présent sur les sites de transcription étudiés, je n'ai en revanche pas regardé la présence des enzymes de la coiffe sur les immunoprécipitations de la RNAPII. Or, on sait que les enzymes de la coiffe se lient au CTD phosphorylé de la RNAPII [203].

6.9 Rôle(s) de l'association RNAPII – snRNP U1

Les résultats obtenus permettent aussi une réflexion sur le contexte de la transcription de ces gènes. On constate en effet la présence de SC35 et SF2/ASF sur ces sites de transcription avec épissage. Sans épissage, SC35 disparaît et SF2/ASF diminue. On peut en déduire que, comme tous deux sont marqueurs des speckles, un site de transcription faisant de l'épissage est associé une fois sur deux avec un speckle alors qu'un site sans épissage ne l'est jamais. Les speckles regroupent plusieurs complexes nucléaires contenant la protéine SC35 mais ce terme est plus souvent associé aux groupes de granules interchromatiniens (IGC interchromatin granule cluster) qui sont considérés comme des lieux de stockage des facteurs d'épissage [172]. On y trouve aussi un peu de RNAPII [31, 216]. La RNAPII que l'on y trouve n'est pas active. Ces IGC (ou speckles) sont connectés les uns aux autres par les fibrilles périchromatiniennes (FP) qui ne sont autres que des ARNs naissants entourés de facteurs d'épissage actifs. IGC et FP sont donc voisins. Un site de transcription a donc beaucoup de chances d'être proche d'un speckle. On peut imaginer dans ce cas que la RNAPII en cours de recyclage passe par un IGC car le temps de résidence n'y est pas long [248]. Dans mes résultats, si mes sites de gènes non épissés ne sont pas associés aux speckles, alors cela suppose une organisation nucléaire particulière où les speckles « se positionnent » ou « se créent » à proximité des gènes qui ont besoin d'épissage. Ceux-ci se créeraient en quelque sorte par défaut, là où les facteurs s'accumulent car ils ne cessent d'être requis et relâchés. Ces résultats nous apprennent également que la snRNP U1, recrutée sur un site de transcription sans intron, n'est pas associée aux speckles. Il existe donc bien une forme de snRNP U1 différente de la snRNP U1 active dans le spliceosome et différente de celle associée aux speckles.

L'ensemble de ces résultats nous amènent à considérer différentes hypothèses, pas nécessairement exclusives. Le complexe formé par la RNAPII et la snRNP U1 pourrait d'abord avoir une influence sur la RNAPII. L'association est déjà présente en mitose ce qui pourrait signifier que la RNAPII en a besoin avant de commencer son activité. Cette hypothèse est en accord avec les résultats de Tian qui montrent une association de la snRNP U1 avec la RNAPII avant la formation du complexe d'initiation. En effet, un ARN non codant homologue à la séquence du 5'ss peut inhiber l'activité transcriptionnelle de la RNAPII s'il est ajouté avant la formation du complexe, mais pas s'il est ajouté après. De manière réciproque, la présence d'un site 5'ss valide permet d'augmenter la transcription de la RNAPII, indépendamment de l'épissage [59]. Ainsi, en plus de scanner l'ARN, la snRNP U1 pourrait signaler à la RNAPII qu'elle peut s'engager dans la transcription car elle n'est pas entourée d'ARNs pré-matures, en cours de transcription. Dans cette même idée de rôle sur la transcription, on a cherché à savoir si Tat-SF1 était lui aussi recruté au site. En effet, Tat-SF1 interagit avec la snRNP U2 mais il est aussi sensé faire des interactions avec DSIF (Handa, Cold Spring Harbor 2008). Or, DSIF est un facteur négatif d'élongation de la transcription. Si Tat-SF1 était recruté avec la snRNP U1, cela pourrait avoir un impact sur la liaison de DSIF et/ou de la snRNP U2. Aucun anticorps n'étant compatible avec une immunofluorescence, nous avons utilisé des protéines Tat-SF1-GFP. Les résultats n'ont cependant pas été convaincants, le recrutement de Tat-SF1 sur un gène avec introns, quoique significatif, étant déjà bien faible. Si Tat-SF1 exerce un rôle, c'est probablement à faible dose ou bien sur des gènes particuliers. La snRNP U1 pourrait aussi utiliser la RNAPII comme vecteur afin d'être amenée au site de transcription pour s'y associer à des facteurs et les activer. Le facteur de transcription TFIID serait en effet associé à l'ARN U1 via la cycline H et stimulerait l'activité kinase de cdk7 [166, 232]. La présence de U1 associé à TFIID permettrait une réinitiation plus rapide de la transcription. Même si je n'ai pas trouvé traces de cycline H ou cdk7 sur mes immunoprécipitations RNAPII, il se peut tout de même qu'une interaction ait lieu entre la snRNP U1 et TFIID. En effet, si la snRNP U1 est associée à la RNAPII, TFIID est aussi en interaction avec la RNAPII. TFIID n'est pas le seul candidat potentiel à une association avec l'ARN U1. Sur une construction du gène HIV-1 avec un 5'ss muté comparé à un 5'ss normal, TFIID, TFIIE et TFIIIB sont enrichis de la même façon partout le long du gène sur la construction au 5'ss muté [59]. En revanche, sur le 5'ss sauvage, TFIIIB n'est plus enrichi à 200 pb du 5'ss alors

que TFIID et TFIIF le sont jusqu'au 5'ss (560pb). Les auteurs en concluent que le 5'ss est capable de stimuler la transcription en recrutant au moins TFIID et TFIIF pour permettre une réinitiation plus rapide. L'ARN U1 ou la snRNP U1 pourraient être les médiateurs de cet effet du 5'ss.

Etant donné la fonction de la snRNP U1, ce complexe pourrait aussi avoir des répercussions sur l'épissage cotranscriptionnel. Amenée par la RNAPII, la snRNP U1 serait à proximité de l'ARN naissant pour scanner ce pré-ARNm et vérifier si celui-ci contient des sites d'épissage. La snRNP U1 étant le premier élément à reconnaître le site donneur, il est indispensable à l'association ultérieure du splicéosome. Sa présence en avant-garde permettrait une association plus rapide du splicéosome et donc un épissage cotranscriptionnel plus efficace. On connaît déjà des protéines de la snRNP U1 chez la levure qui se lient à la RNAPII (FBP11/Prp40) [214] ou qui stimulent l'épissage (TIA-1/Nam8) [90]. Bien que l'association de Prp40 avec la RNAPII ait été caractérisée [95], aucune fonction particulière n'a été mise en évidence. TIA-1 recrute quant à lui la protéine U1C de la snRNP U1 pour permettre une reconnaissance plus efficace du site 5'ss [90]. La snRNP U1 est donc un bon candidat pour stimuler l'épissage cotranscriptionnel. Pour vérifier l'hypothèse émise plus haut, il faudrait utiliser un système de transcription et épissage *in vitro* avec une RNAPII purifiée en ajoutant séquentiellement l'ARN U1 ou bien la snRNP U1 avant d'ajouter l'extrait nucléaire, avant ou après l'initiation sur un promoteur à cassette G-less. On mesurerait la vitesse d'épissage et le nombre de transcrits épissés.

On sait également que le complexe de la coiffe facilite le recrutement de la snRNP U1 pour la reconnaissance du site 5'ss. La RNAPII est un gros complexe et l'ARN U1 est très petit en comparaison. Si l'ARN U1 se lie à un endroit éloigné de la sortie de l'ARN, il peut avoir besoin d'aide pour rejoindre l'ARN naissant. Ainsi, l'ARN U1 pourrait être facilité dans sa fonction de reconnaissance du 5'ss, d'abord en étant amené au site de transcription par la RNAPII, ensuite en trouvant l'ARN néosynthétisé par le complexe de la coiffe.

Enfin, le complexe RNAPII – snRNP U1 pourrait également avoir un rôle dans la stimulation d'une autre étape de maturation de l'ARN. En effet, il a été récemment montré que la snRNP U2 stimule la coupure endonucléolytique du 3' des pré-ARNm des histones par la snRNP U7 par la liaison de deux de ses sous-unités (SF3b155 et hPrp43) à une séquence d'ARN particulière. Or les histones sont des gènes sans introns qui ne recrutent a priori pas le splicéosome [91]. On savait

déjà que la snRNP U2 stimulait la maturation du 3' par une interaction avec CPSF [167] mais c'est la première fois que l'on voit U2 recruté dans ce but sur un gène sans intron. La snRNP U1 est plutôt connue pour son rôle inhibiteur vis-à-vis de la maturation du 3' sur les gènes humains [114, 314] et sur le VIH [10, 9]. Elle inhibe la polyadénylation des gènes dont le 3' est proche du 5' du gène suivant. Artificiellement, on crée des sites de liaisons pour l'ARN U1 afin d'inhiber la polyadénylation. La snRNP U1 pourrait être recrutée avec la RNAPII qui se place en amont du gène sur le promoteur et donc plus près de l'extrémité 3' d'un gène contigu afin de réguler les facteurs de coupure et polyadénylation. D'autres mécanismes pourraient être régulés par la présence de la snRNP U1 : l'exportation ou le NMD sont des événements en lien avec l'épissage. Cependant aucun travail dans la littérature ne vient aujourd'hui étayer cette hypothèse.

Pour le moment, nous ne savons pas si l'association RNAPII – snRNP U1 est basée sur l'interaction avec l'ARN U1 ou avec l'une de ses protéines. Cette question est restée de côté pendant ma thèse mais elle est suffisamment importante pour mériter une réponse. Si l'ARN U1 est l'élément qui se lie à la RNAPII, il faudrait rechercher sur quel(s) aspect(s) cette liaison repose : un motif de séquence particulier ou la structure ? L'article de Tian évoque l'importance de la séquence correspondant au site de reconnaissance du 5'ss. Cependant, cette séquence n'est probablement pas en interaction physique avec la RNAPII car elle doit s'hybrider avec les ARNs naissants pour inhiber l'initiation. En lien avec la séquence et la structure, il existe des variants d'ARN U1 [168]. Ils sont beaucoup moins exprimés (entre 2 et 7% de l'ARN U1 majoritaire) et comportent des variations de séquences, notamment sur la séquence de reconnaissance du 5'ss et le dinucléotide GU. Certains de ces variants comportent des SNPs, y compris sur le dinucléotide GU. Ces variations de séquences ne les empêchent pas de former des snRNPs et leur permettent probablement de reconnaître des sites 5'ss non canoniques. On pourrait vérifier si l'ARN U1 en association avec la RNAPII est l'ARN U1 majoritaire ou bien n'importe lequel ou bien un variant spécifique. Cette dernière hypothèse donnerait alors un rôle très spécifique à cet ARN U1 variant car, moins nombreux que l'ARN majoritaire, il pourrait avoir besoin d'être concentré auprès des ARNs naissants possédant des sites 5'ss non canoniques. Il se pourrait également que l'ARN U1 immunoprécipité avec la RNAPII dans des cellules mitotiques et celui accumulé au site de transcription ne soient pas les mêmes et que leurs rôles soient différents : l'un pour inhiber une initiation (ARN U1 majoritaire),

l'autre pour scanner les ARNs naissants (un variant). Les hypothèses sont multiples, la difficulté résidant dans la capacité du protocole choisi d'arriver à voir 2% d'un variant. De rigoureux moyens devront être mis en œuvre pour faire aboutir des protocoles de RT-PCR et PCR quantitative par exemple.

Dans le système murin, l'équipe de Kugel et Goodrich a démontré d'une part l'activation de la transcription de l'ARN B2 par la RNAPII en conditions de choc thermique et d'autre part son effet inhibiteur sur les gènes RNAPII, hormis les gènes de choc thermique [5, 79]. En réalisant un FISH sur des cellules NIH3T3 contre U1 et B2 et en analysant leurs enrichissements respectifs au site actine, je constate l'expression de l'ARN B2 et un petit recrutement au site actif. L'actine étant un des gènes contrôlés étudiés dans leur article, il se peut que cet enrichissement n'en soit pas vraiment un ou bien que les conditions de choc thermique et celles que j'ai utilisées ne provoquent pas les mêmes réactions dans la cellule. Cette expérience était un test préliminaire. L'enjeu est de savoir si l'ARN B2 et l'ARN U1 sont compétiteurs pour un site de liaison sur le RNAPII. Pour le vérifier, l'étude comparative des gènes histones en conditions normales et celle des gènes de choc thermique par la même méthode pourra nous informer sur plusieurs points. Premièrement, le recrutement de B2 sur des gènes sensibles ou non au choc thermique corroborera ou non avec les résultats de l'équipe de Kugel et Goodrich. Deuxièmement, on pourra évaluer l'enrichissement de l'ARN U1 et celui de l'ARN B2, proportionnellement à l'enrichissement de la RNAPII sur ces différents sites. Cela nous permettra de comparer les conditions de stress et de non-stress d'une part, et un site endogène avec introns et sans introns en conditions de non stress pour l'ARN U1. Enfin, des signaux opposés entre l'ARN U1 et l'ARN B2 seront un argument pour la compétition de ces deux ARNs sur la RNAPII. Bien évidemment, c'est supposer que les résultats, et donc les différences, seront très nettes. Cela peut ne pas être le cas et quelle qu'en soit l'issue, des expériences d'EMSA seront indispensables pour valider l'imagerie.

6.10 Un ARN sans intron est exprimé mais en moins grande quantité.

Outre la découverte du nouveau complexe RNAPII – snRNP U1 et les indices d'un rôle de la snRNP U1 en-dehors de l'épissage, les expériences menées ont montré qu'un gène sans intron

pouvait tout à fait être transcrit, exporté et traduit. Pourtant, il faut remarquer que cette traduction est plus faible en l'absence d'intron. En effet, il y a moins de protéines produites dans la lignée MUT que dans la lignée WT (jugé par Western blot et par le signal CFP). Cela ne vient pas d'un problème du plasmide puisque les transfections transitoires donnaient des quantités de protéines équivalentes entre les plasmides WT et les plasmides MUT. Cet effet de l'intron ne semble pas venir de l'export des ARNm puisque le signal MS2 est identique entre les deux lignées. Pourtant, c'est la raison la plus communément admise : le complexe EJC (Exon Junction Complexe) déposé par la machinerie d'épissage facilite l'export des ARNm [71]. Bien que je n'ai pas cherché à comprendre les causes de cette différence de traitement de l'expression entre les deux lignées, ces résultats confirment l'idée générale selon laquelle un intron améliore l'expression d'un gène [201].

Ce projet comporte bien des pistes à creuser et c'est une joie pour moi de voir autant de possibilités à l'issue de mes recherches. Pour la suite, des reconstitutions *in vitro*, des EMSAs ou bien la technique Biacore seront à un moment donné indispensables pour caractériser l'association RNAPII-snRNP U1. En fonction des résultats, il faudra également tester l'importance de la séquence de l'ARN U1 ou bien de sa structure secondaire dans cette association. Le point le plus important est bien entendu la découverte de la raison de cette association. Des tests d'épissage et de transcription permettront d'en savoir plus car le système *in vivo* ne permet pas de tout faire, notamment du siRNA. J'ai construit au cours de ma thèse d'autres plasmides, similaires aux constructions présentées et utilisées mais cette fois-ci mutés soit en 5'ss, soit en 3'ss. Ces plasmides ont permis d'établir des lignées cellulaires qui n'ont pour l'instant pas été étudiées. Par imagerie, on pourrait voir sur celles-ci comment se comporte l'ARN U1 sur ces sites de transcription. La prédiction étant de trouver de l'ARN U1 sur les sites mutés en 3' mais aussi sur les sites mutés en 5' du fait de leur présence sur les constructions MUT (mutés en 5 et 3'). Cette fois-ci cependant, les autres facteurs et ARNs de l'épissage se comporteraient sans doute différemment et cela pourrait également influencer le ratio d'ARN U1 recruté. Cela permettrait peut-être d'évaluer la part de l'ARN U1 qui viendrait avec la RNAPII de celle qui viendrait avec la machinerie d'épissage. Sur

ces 4 constructions différentes, on pourra envisager de faire du CHIP et par Northern Blot ou PCR quantitative, d'évaluer la quantité d'ARN U1 et d'autres snARNs sur la RNAPII. Des tests sur des gènes endogènes sans introns, comme les gènes d'histones, seraient également les bienvenus par la suite.

Le sujet de ma thèse est très intéressant à l'heure actuelle. Cela n'a pas toujours été le cas, notamment au début où des problèmes de gels et de transfert m'ont beaucoup ralenti. Les tentatives de quantifications également. Aujourd'hui, le sujet doit continuer en donnant la priorité à la fonction de l'association observée car d'autres équipes compétitives ont déjà mis en évidence cette association sans y accorder pour le moment beaucoup d'importance. La publication de mon travail va certainement donner un grand coup d'accélérateur dans le domaine, d'autant que les pistes sont nombreuses et très intéressantes.

Cinquième partie

Annexes

Annexe A

Tests statistiques sur les données obtenues par analyse radiale

Le détail de ces tests est expliqué dans matériel et méthodes. Chaque analyse a été testée en bootstrap et le résultat de ces tests est présenté dans les figures suivantes. Un résultat est dit significatif en comparant la p-value à un taux que l'on fixe arbitrairement. En plaçant l'erreur à 5% (p-value = 0,05), un résultat est significatif si la p-value de celui-ci est inférieure à 0,05. Plus la p-value est en-dessous de cette valeur et plus le résultat est significatif. Un résultat proche de 5% (par exemple p-value = 0,03) est considéré comme peu significatif et ne sera pas pris en compte. Les données représentées dans les figures A.1 et A.2 montrent que le recrutement au site de transcription est très significativement différent pour l'ensemble des snARNs de l'épissage sur le gène avec intron et uniquement pour le snARN U1 sur le gène sans intron. On constate que les p-values pour l'ARN U1 sont approximativement dans le même ordre de grandeur dans les deux lignées. L'enrichissement en U1 est donc équivalent sur les sites MUT et sur les sites WT.

Les figures A.3 et A.4 montrent les bootstraps réalisés sur les données concernant les protéines analysées sur les constructions. On constate que toutes les protéines sont significativement enrichies au site de transcription par rapport au site contrôle sur la construction WT et dans des ordres de grandeur identiques entre les lignées. Au site de transcription muté pour l'épissage, U2AF65 n'est pas enrichi. Les graphiques SC35 représentent le site de transcription par rapport au site contrôle négatif (en-dehors des speckles). À côté des graphiques SC35, la p-value obtenue en bootstrap des données du contrôle positif (dans un speckle) par rapport au contrôle négatif. On constate que la p-value du site de transcription et du site contrôle positif sont dans le même ordre de grandeur dans les cellules WT ainsi que la p-value du contrôle positif dans les cellules MUT. En revanche, le signal SC35 obtenu au site de transcription des cellules MUT n'est pas significativement différent du contrôle négatif. SC35 n'est donc pas enrichi au site de transcription des cellules MUT.

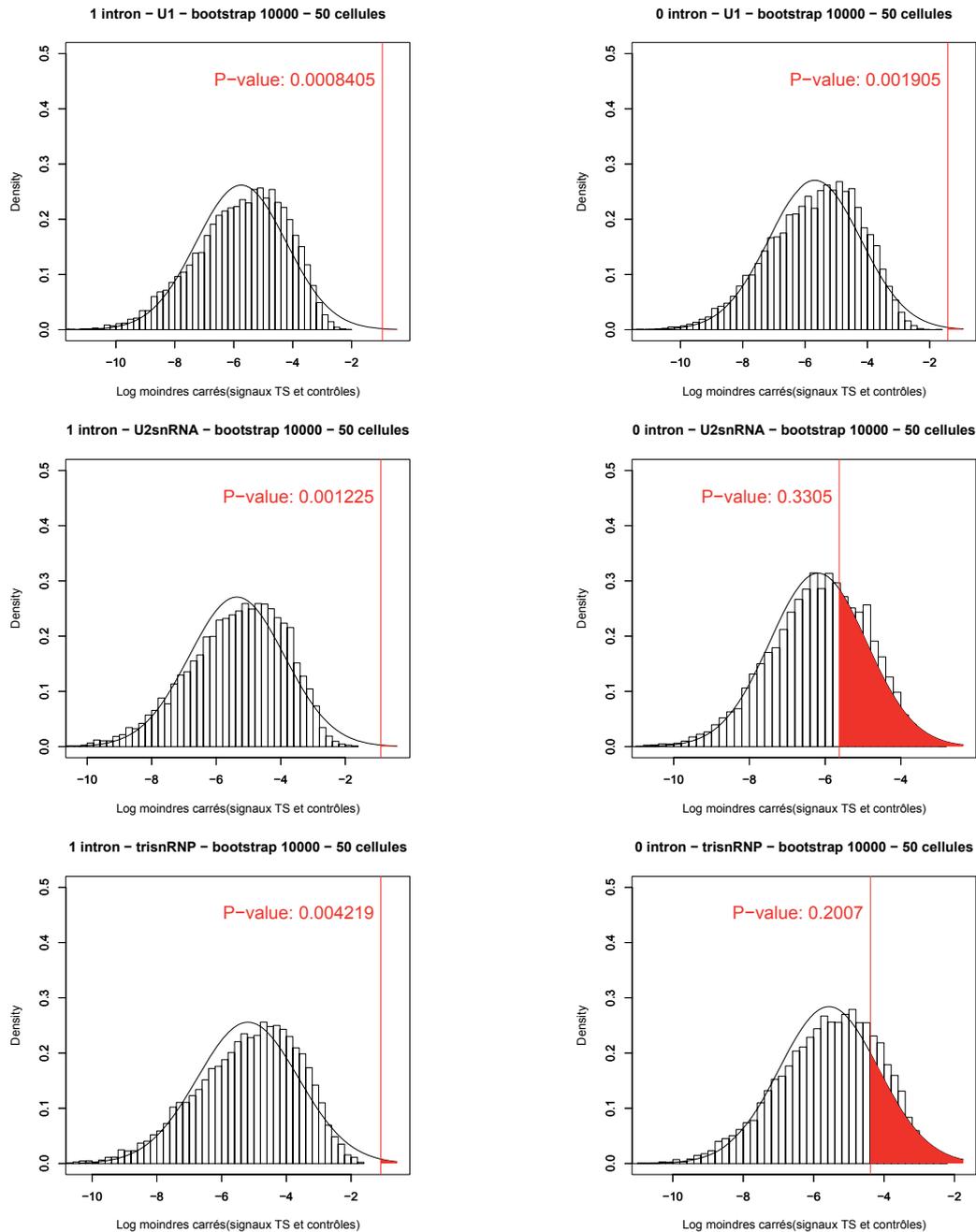


FIGURE A.1 – **Bootstrap ARN.** Représentation des bootstraps réalisés sur les données obtenues avec les snARNs au site de transcription par rapport au site contrôle sur les lignées WT et MUT.

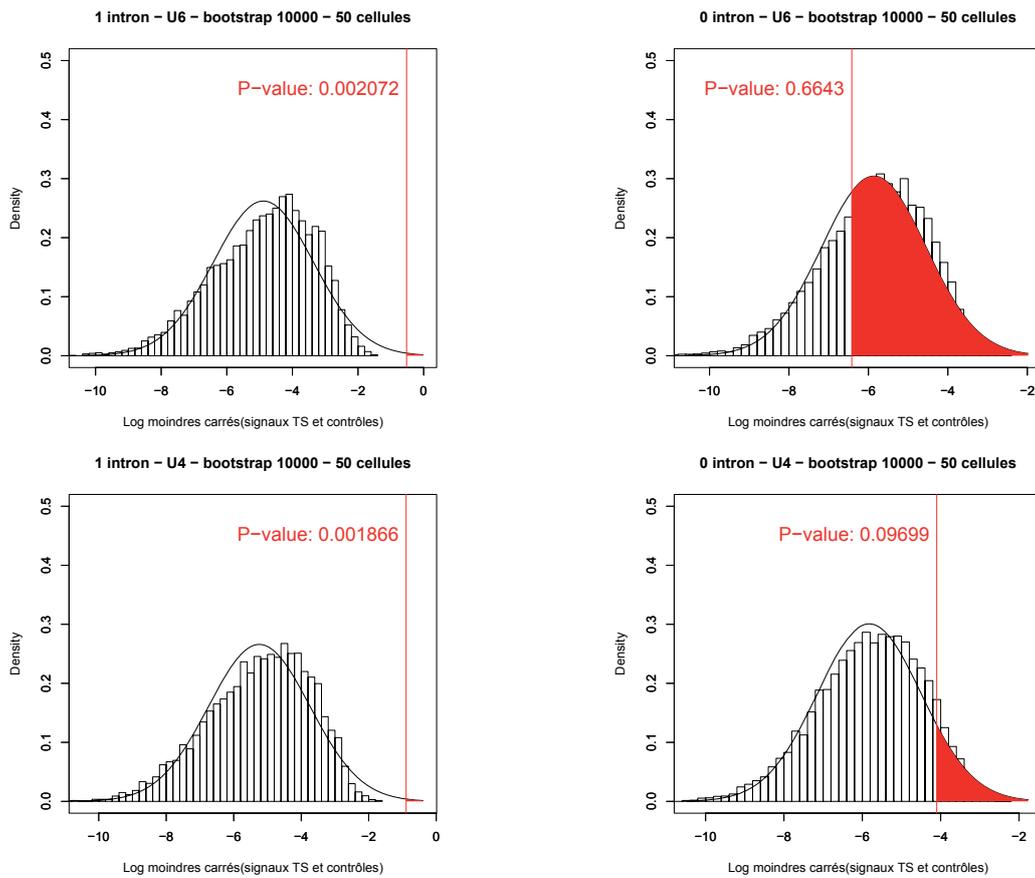


FIGURE A.2 – Bootstrap ARN. Suite de la figure

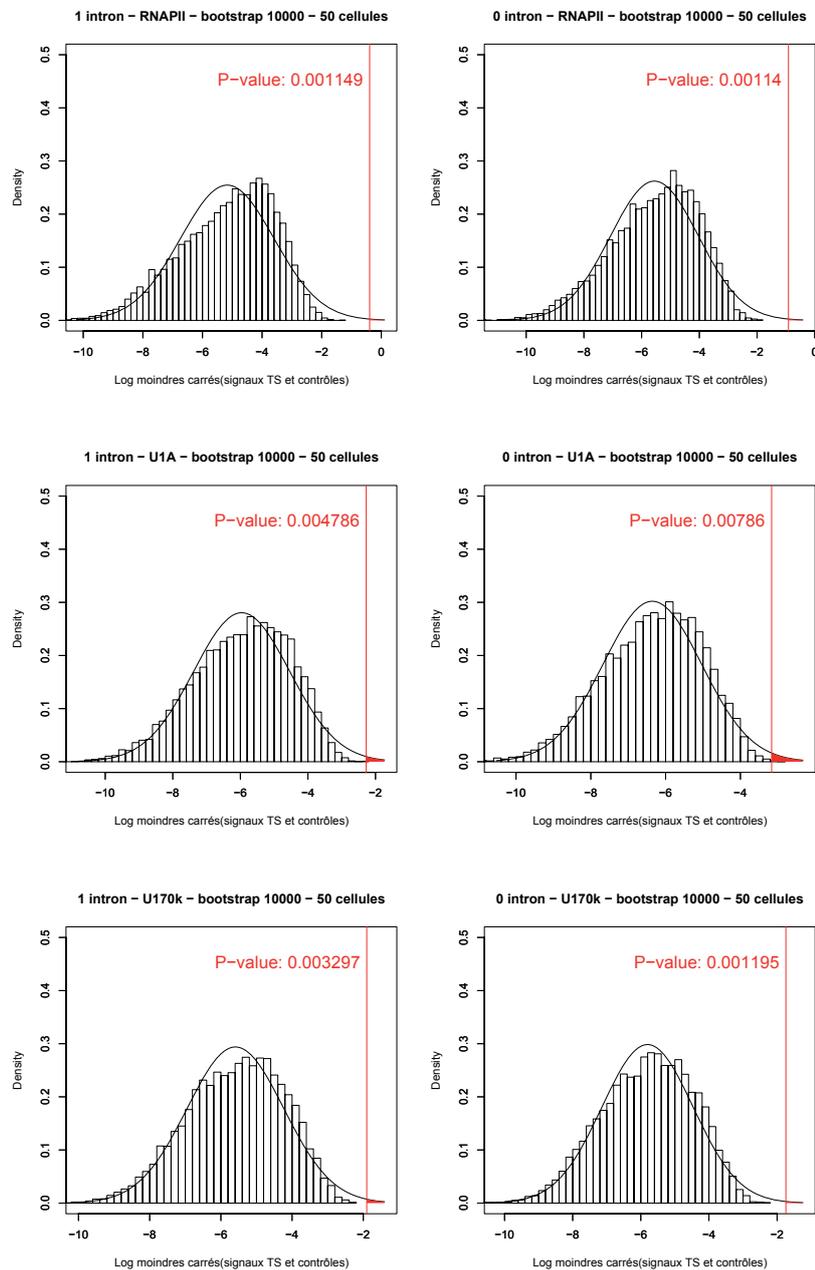
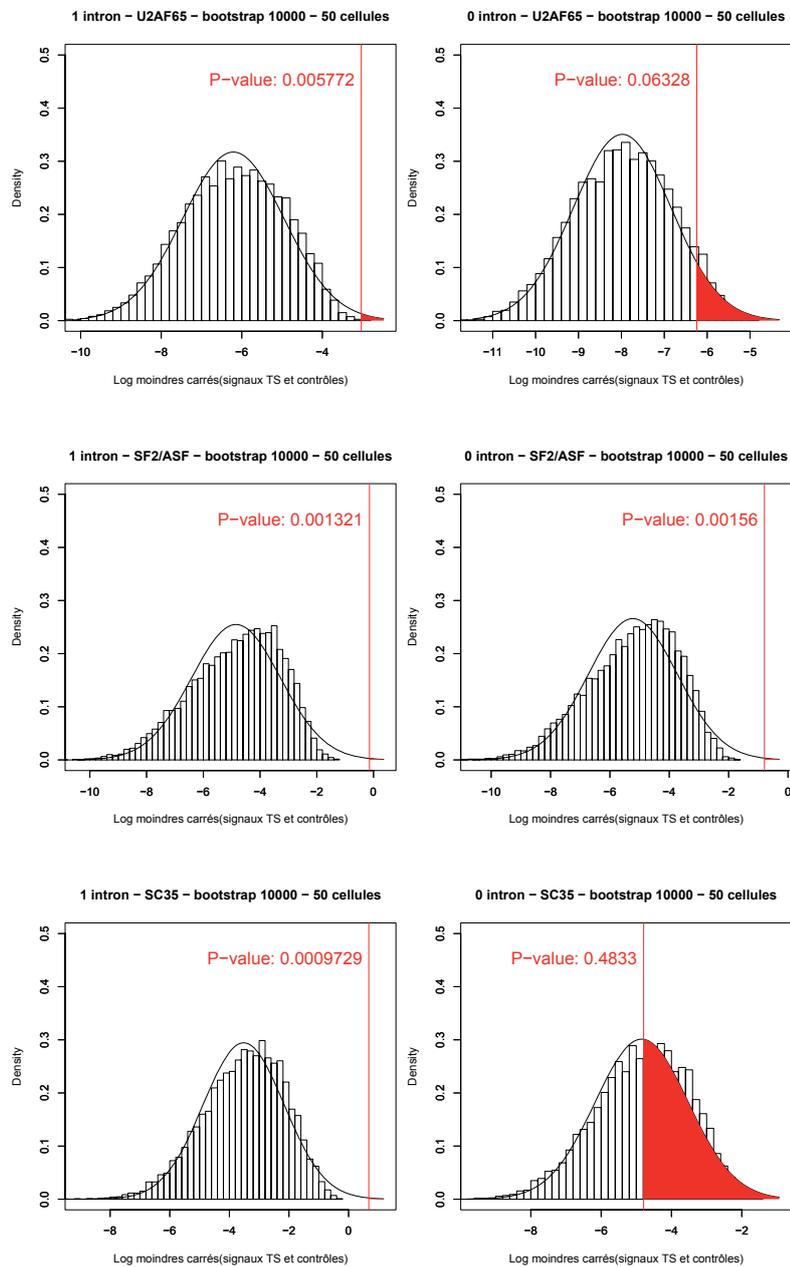


FIGURE A.3 – **Bootstrap protéines.** Représentation des bootstraps réalisés sur les données obtenues avec les protéines au site de transcription par rapport au site contrôle sur les lignées WT et MUT.



positive control P-value: 0.001029

positive control P-value: 0.0008162

FIGURE A.4 – Bootstrap protéines. Suite de la figure

Annexe B

Représentation des données

J'ai décrit dans les matériels et méthodes la représentation utilisée pour représenter les résultats dans ma thèse. Une autre méthode a été utilisée, basée sur la différence de signal entre le centre du site de transcription et la périphérie. Dans la représentation utilisée dans ce manuscrit, on évalue le rapport du signal au centre du site de transcription sur le signal à sa périphérie (figure B.1) avec la formule indiquée en A. On fait ensuite la moyenne, rayon par rayon, sur plusieurs cellules d'une même lignée. L'autre analyse consiste à mesurer la différence $TS_{1.6m} - TS_{rm}$. Si on se contente de faire la moyenne de cette différence sur plusieurs cellules, rayon par rayon, on ne peut pas comparer plusieurs lignées différentes, car les niveaux de signal ne sont pas les mêmes d'une cellule à une autre (expression de l'ARN variable, différence du temps d'exposition, intensité lumineuse de l'exposition, qualité du FISH ou de l'immunofluorescence. . .). Il faut donc rapporter ces signaux à une valeur référence propre à la cellule étudiée ou à la lignée. Dans le cas de la première représentation, chaque signal est rapporté au signal de la périphérie qui représente le niveau de base de la cellule. Celui-ci est différent pour chaque cellule. Dans la deuxième représentation, si on rapporte le signal obtenu au centre du site sur celui de la périphérie, on obtient la même opération que dans la première représentation. Chaque signal est rapporté à la valeur du signal de base dans la cellule : $\frac{\sum \frac{TS - TS_{1.6}}{TS_{1.6}}}{n} = \sum \frac{TS - TS_{1.6}}{n} - 1$. Or, la deuxième représentation rapporte cette différence, non pas à la valeur de base de la cellule, mais à la valeur de base moyenne obtenue dans les cellules de la lignée donnée (figure B). Cela suppose que les images d'une cellule à une autre constituent un groupe homogène, contrairement à la première représentation qui y est moins sensible.

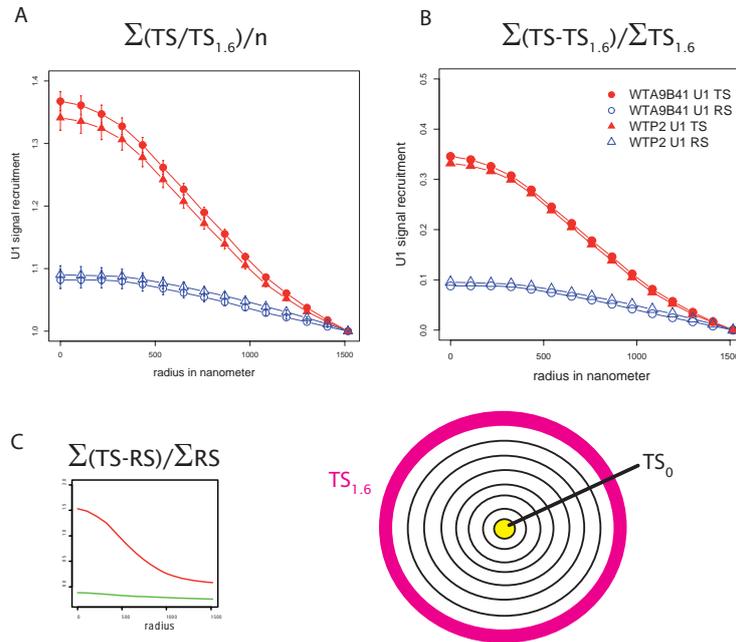


FIGURE B.1 – **Choix de la représentation.** Les deux formules choisies pour la représentation des résultats sont démontrées ici. La figure A correspond à la représentation utilisée dans le manuscrit. Pour l'explication, voir le texte. Le schéma du bas représente l'analyse radiale et les couronnes correspondantes.

Une autre méthode envisagée était de faire la différence, rayon par rayon, entre le site de transcription et le site contrôle, rapporté au RS (figure C). Ce calcul n'aboutit à rien dans la mesure où le site de transcription correspond à un locus précis et donc à un environnement nucléaire précis alors que le site contrôle est totalement aléatoire et donc son environnement également. Si le site se trouve dans un locus proche d'un gène hautement exprimé par exemple, il sera plus souvent associé à des facteurs de transcription que le site aléatoire. La différence $TS-RS/RS$ donnera alors un enrichissement surévalué, car le point RS du site contrôle représentera un panel de loci alors que le site TS est associé à un locus unique. Réciproquement, si le site TS est dans un locus peu transcrit, le site contrôle paraîtra plus riche en facteurs de transcription.

Pour savoir si une population d'images est homogène, un test simple consiste à faire un boxplot (figure B.2). Le boxplot représente la population en la divisant en 2 parties par la médiane, valeur partageant les données à 50% de part et d'autre (figure A). Entre le 1er quartile et le 3e quartile, on trouve 50% des valeurs. Une population homogène est représentée dans le schéma A. La figure B montre 6 expériences indépendantes de FISH U1. Les données correspondent aux signaux trouvés

au rayon 0 nm au site de transcription de la lignée WT. Les valeurs en ordonnée représentent le niveau de gris en 32 bits. La saturation des images est à 12285. On voit pour l'expérience 1 que les valeurs sont comprises entre environ 2200 et 8200, ce qui représente une bonne dynamique du signal. En revanche, l'expérience n° 4 est très ramassée, les valeurs sont entre 200 et 4500. De plus, la médiane est vraiment décalée par rapport aux données (les quartiles ne sont pas équivalents). Cette expérience n'est donc pas bonne dans le sens où la dynamique d'images n'est pas bonne et la population des images peu homogène.

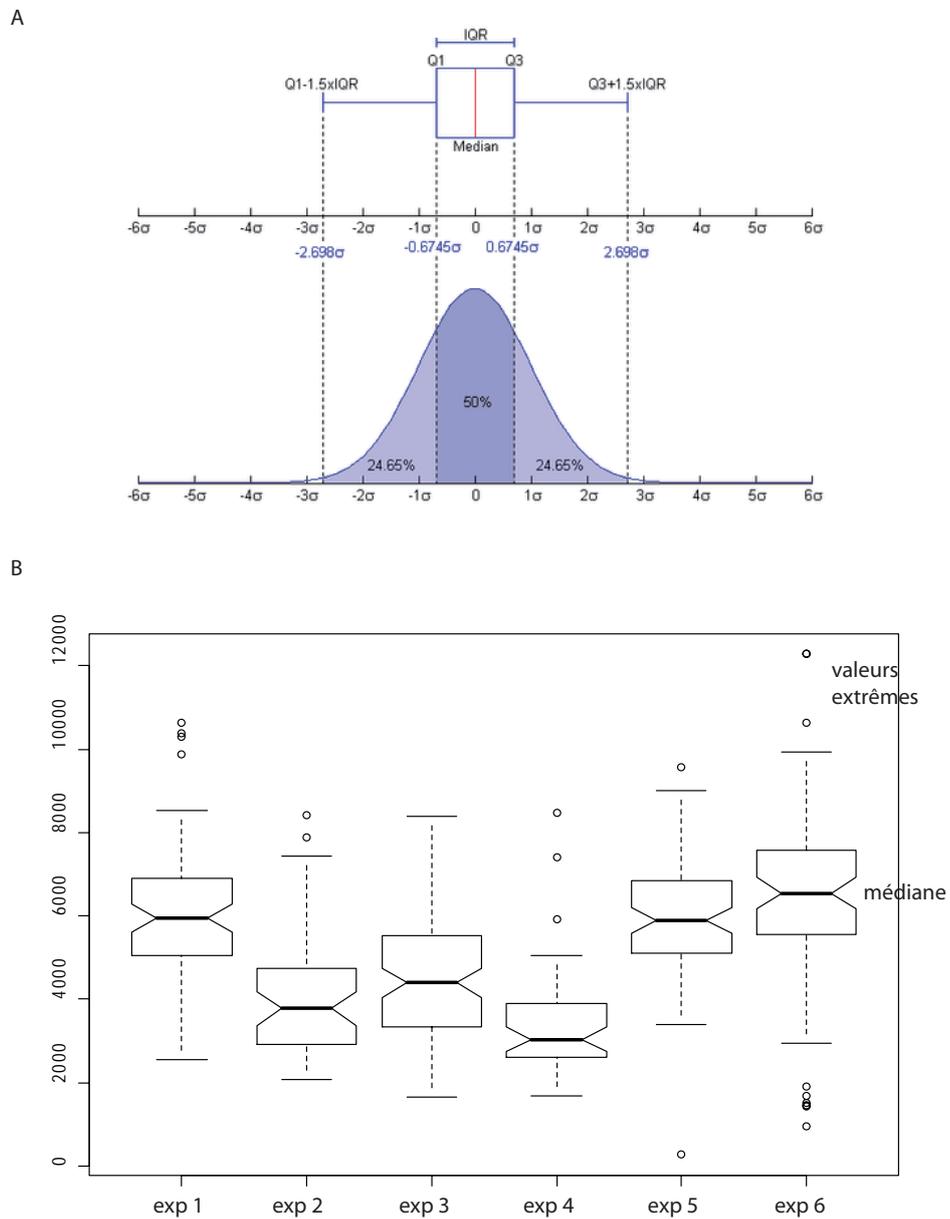


FIGURE B.2 – **Boxplot.** **A.** Représentation d'une population homogène standard en box-plot. **B.** Box-plot de 6 expériences indépendantes d'images de FISH U1 au rayon 0 nm du site de transcription.

Bibliographie

- [1] Abraham Abraham and Martin Pelchat. Formation of an rna polymerase ii preinitiation complex on an rna promoter derived from the hepatitis delta virus rna genome. *Nucleic Acids Res*, 36(16) :5201–5211, 2008.
- [2] Barham K Abu Abu Dayyeh, Tiffani K Quan, Marygrace Castro, and Stephanie W Ruby. Probing interactions between the u2 small nuclear ribonucleoprotein and the dead-box protein, prp5. *J Biol Chem*, 277(23) :20221–20233, 2002.
- [3] Satoko Akashi, Shinjiro Nagakura, Seiji Yamamoto, Masahiko Okuda, Yoshiaki Ohkuma, and Yoshifumi Nishimura. Structural characterization of human general transcription factor ttf1 in solution. *Protein Sci*, 17(3) :389–400, Mar 2008.
- [4] Eric Allemand, Eric Batsche, and Christian Muchardt. Splicing, transcription, and chromatin : a manage a trois. *Curr Opin Genet Dev*, 18(2) :145–151, 2008.
- [5] Tiffany A Allen, Sandra Von Kaenel, James A Goodrich, and Jennifer F Kugel. The sine-encoded mouse b2 rna represses mrna transcription in response to heat shock. *Nat Struct Mol Biol*, 11(9) :816–821, 2004.
- [6] E D Andrusis, E Guzman, P Doring, J Werner, and J T Lis. High-resolution localization of drosophila spt5 and spt6 at heat shock genes in vivo : roles in promoter proximal pausing and transcription elongation. *Genes Dev*, 14(20) :2635–2649, 2000.
- [7] J E Arenas and J N Abelson. Prp43 : An rna helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci U S A*, 94(22) :11798–11802, 1997.
- [8] Karim-Jean Armache, Simone Mitterweger, Anton Meinhart, and Patrick Cramer. Structures of complete rna polymerase ii and its subcomplex, rpb4/7. *J Biol Chem*, 280(8) :7131–7134, 2005.
- [9] M P Ashe, A Furger, and N J Proudfoot. Stem-loop 1 of the u1 snrnp plays a critical role in the suppression of hiv-1 polyadenylation. *RNA*, 6(2) :170–177, Feb 2000.
- [10] M P Ashe, L H Pearson, and N J Proudfoot. The hiv-1 5' ltr poly(a) site is inactivated by u1 snrnp interaction with the downstream major splice donor site. *EMBO J*, 16(18) :5752–5763, Sep 1997.
- [11] Maia Azubel, Naomi Habib, Ruth Sperling, and Joseph Sperling. Native spliceosomes assemble with pre-mrna to form supraspliceosomes. *J Mol Biol*, 356(4) :955–966, 2006.
- [12] Hwa Jin Baek, Yun Kyoung Kang, and Robert G Roeder. Human mediator enhances basal transcription by facilitating recruitment of transcription factor iib during preinitiation complex assembly. *J Biol Chem*, 281(22) :15172–15181, Jun 2006.
- [13] Vladimir B Bajic, Sin Lam Tan, Alan Christoffels, Christian Schonbach, Leonard Lipovich, Liang Yang, Oliver Hofmann, Adele Kruger, Winston Hide, Chikatoshi Kai, Jun Kawai, David A Hume, Piero Carninci, and Yoshihide Hayashizaki. Mice and men : their promoter properties. *PLoS Genet*, 2(4) :e54, 2006.
- [14] D Barilla, B A Lee, and N J Proudfoot. Cleavage/polyadenylation factor ia associates with the carboxyl-terminal domain of rna polymerase ii in saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, 98(2) :445–450, 2001.

- [15] Jeffrey E Barrick, Narasimhan Sudarsan, Zasha Weinberg, Walter L Ruzzo, and Ronald R Breaker. 6s rna is a widespread regulator of eubacterial rna polymerase that resembles an open promoter. *RNA*, 11(5) :774–784, 2005.
- [16] David P Bartel. Micrnas : genomics, biogenesis, mechanism, and function. *Cell*, 116(2) :281–297, Jan 2004.
- [17] Richard Bartfai, Carolin Balduf, Traci Hilton, Yvonne Rathmann, Yavor Hadzhiev, Laszlo Tora, Laszlo Orban, and Ferenc Muller. Tbp2, a vertebrate-specific member of the tbp family, is required in embryonic development of zebrafish. *Curr Biol*, 14(7) :593–598, Apr 2004.
- [18] Nastaran Behzadnia, Monika M Golas, Klaus Hartmuth, Bjoern Sander, Berthold Kastner, Jochen Deckert, Prakash Dube, Cindy L Will, Henning Urlaub, Holger Stark, and Reinhard Luhrmann. Composition and three-dimensional em structure of double affinity-purified, human prespliceosomal a complexes. *EMBO J*, 26(6) :1737–1748, 2007.
- [19] Mathias Bell, Silke Schreiner, Andrey Damianov, Ram Reddy, and Albrecht Bindereif. p110, a novel human u6 snrnp protein and u4/u6 snrnp recycling factor. *EMBO J*, 21(11) :2724–2735, 2002.
- [20] David L Bentley. Rules of engagement : co-transcriptional recruitment of pre-mrna processing factors. *Curr Opin Cell Biol*, 17(3) :251–256, 2005.
- [21] J A Berglund, N Abovich, and M Rosbash. A cooperative interaction between u2af65 and mbbp/sf1 facilitates branchpoint region recognition. *Genes Dev*, 12(6) :858–867, 1998.
- [22] T H Bestor. Gene silencing as a threat to the success of gene therapy. *J Clin Invest*, 105(4) :409–411, 2000.
- [23] A L Beyer and Y N Osheim. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev*, 2(6) :754–765, 1988.
- [24] Gregory Bird, Diego A R Zorio, and David L Bentley. Rna polymerase ii carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mrna splicing and 3'-end formation. *Mol Cell Biol*, 24(20) :8963–8969, 2004.
- [25] Debabrata Biswas, Rinku Dutta-Biswas, Doyel Mitra, Yoichiro Shibata, Brian D Strahl, Tim Formosa, and David J Stillman. Opposing roles for set2 and yfact in regulating tbp binding at promoters. *EMBO J*, 25(19) :4479–4489, 2006.
- [26] Danielle Bittencourt, Martin Dutertre, Gabriel Sanchez, Jerome Barbier, Lise Gratadou, and Didier Auboeuf. Cotranscriptional splicing potentiates the mrna production from a subset of estradiol-stimulated genes. *Mol Cell Biol*, 28(18) :5811–5824, 2008.
- [27] Douglas L Black. Mechanisms of alternative pre-messenger rna splicing. *Annu Rev Biochem*, 72 :291–336, 2003.
- [28] Marco Blanchette, Emmanuel Labourier, Richard E Green, Steven E Brenner, and Donald C Rio. Genome-wide analysis reveals an unexpected function for the drosophila splicing factor u2af50 in the nuclear export of intronless mrnas. *Mol Cell*, 14(6) :775–786, 2004.
- [29] Hinrich Boeger, David A Bushnell, Ralph Davis, Joachim Griesenbeck, Yahli Lorch, J Seth Strattan, Kenneth D Westover, and Roger D Kornberg. Structural basis of eukaryotic gene transcription. *FEBS Lett*, 579(4) :899–903, 2005.
- [30] F Bonnet, M Vigneron, O Bensaude, and M F Dubois. Transcription-independent phosphorylation of the rna polymerase ii c-terminal domain (ctd) involves erk kinases (mek1/2). *Nucleic Acids Res*, 27(22) :4399–4404, 1999.
- [31] D B Bregman, L Du, S van der Zee, and S L Warren. Transcription-dependent redistribution of the large subunit of rna polymerase ii to discrete nuclear domains. *J Cell Biol*, 129(2) :287–298, 1995.
- [32] Julius Brennecke, Alexei A Aravin, Alexander Stark, Monica Dus, Manolis Kellis, Ravi Sachidanandam, and Gregory J Hannon. Discrete small rna-generating loci as master regulators of transposon activity in drosophila. *Cell*, 128(6) :1089–1103, Mar 2007.

- [33] R L Brinster, J M Allen, R R Behringer, R E Gelinas, and R D Palmiter. Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci U S A*, 85(3) :836–840, 1988.
- [34] David A Brow. Allosteric cascade of spliceosome activation. *Annu Rev Genet*, 36 :333–360, 2002.
- [35] S E Butcher and D A Brow. Towards understanding the catalytic core structure of the spliceosome. *Biochem Soc Trans*, 33(Pt 3) :447–449, 2005.
- [36] Jennifer E F Butler and James T Kadonaga. The rna polymerase ii core promoter : a key component in the regulation of gene expression. *Genes Dev*, 16(20) :2583–2592, 2002.
- [37] Elizabeth A Campbell, Lars F Westblade, and Seth A Darst. Regulation of bacterial rna polymerase sigma factor activity : a structural perspective. *Curr Opin Microbiol*, 11(2) :121–127, 2008.
- [38] M Caputi and A M Zahler. Determination of the rna binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) h/h'/f/2h9 family. *J Biol Chem*, 276(47) :43850–43859, 2001.
- [39] Giuseppina Caretti, Elissa P Lei, and Vittorio Sartorelli. The dead-box p68/p72 proteins and the non-coding rna steroid receptor activator sra : eclectic regulators of disparate biological functions. *Cell Cycle*, 6(10) :1172–1176, 2007.
- [40] Piero Carninci, Albin Sandelin, Boris Lenhard, Shintaro Katayama, Kazuro Shimokawa, Jasmina Ponjavic, Colin A M Semple, Martin S Taylor, Par G Engstrom, Martin C Frith, Alistair R R Forrest, Wynand B Alkema, Sin Lam Tan, Charles Plessy, Rimantas Kodzius, Timothy Ravasi, Takeya Kasukawa, Shiro Fukuda, Mutsumi Kanamori-Katayama, Yayoi Kitazume, Hideya Kawaji, Chikatoshi Kai, Mari Nakamura, Hideaki Konno, Kenji Nakano, Salim Mottagui-Tabar, Peter Arner, Alessandra Chesi, Stefano Gustincich, Francesca Persichetti, Harukazu Suzuki, Sean M Grimmond, Christine A Wells, Valerio Orlando, Claes Wahlestedt, Edison T Liu, Matthias Harbers, Jun Kawai, Vladimir B Bajic, David A Hume, and Yoshihide Hayashizaki. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, 38(6) :626–635, 2006.
- [41] Luca Cartegni and Adrian R Krainer. Disruption of an sf2/asf-dependent exonic splicing enhancer in smn2 causes spinal muscular atrophy in the absence of smn1. *Nat Genet*, 30(4) :377–384, 2002.
- [42] Amelia Casamassimi and Claudio Napoli. Mediator complexes and eukaryotic transcription regulation : an overview. *Biochimie*, 89(12) :1439–1446, Dec 2007.
- [43] James Z Chadick and Francisco J Asturias. Structure of eukaryotic mediator complexes. *Trends Biochem Sci*, 30(5) :264–271, May 2005.
- [44] Rob D Chapman, Marcus Conrad, and Dirk Eick. Role of the mammalian rna polymerase ii c-terminal domain (ctd) nonconsensus repeats in ctd stability and cell proliferation. *Mol Cell Biol*, 25(17) :7665–7674, 2005.
- [45] Rob D Chapman, Marcus Conrad, and Dirk Eick. Role of the mammalian rna polymerase ii c-terminal domain (ctd) nonconsensus repeats in ctd stability and cell proliferation. *Mol Cell Biol*, 25(17) :7665–7674, 2005.
- [46] Rob D Chapman, Martin Heidemann, Thomas K Albert, Reinhard Mailhammer, Andrew Flatley, Michael Meisterernst, Elisabeth Kremmer, and Dirk Eick. Transcribing rna polymerase ii is phosphorylated at ctd residue serine-7. *Science*, 318(5857) :1780–1782, 2007.
- [47] Rob D Chapman, Martin Heidemann, Corinna Hintermair, and Dirk Eick. Molecular evolution of the rna polymerase ii ctd. *Trends Genet*, 24(6) :289–296, 2008.
- [48] Rob D Chapman, Benoit Palancade, Andreas Lang, Olivier Bensaude, and Dirk Eick. The last ctd repeat of the mammalian rna polymerase ii large subunit is important for its stability. *Nucleic Acids Res*, 32(1) :35–44, 2004.
- [49] Chun-Hong Chen, Der-I Kao, Shih-Peng Chan, Tsu-Chung Kao, Jui-Yen Lin, and Soo-Chen Cheng. Functional links between the prp19-associated complex, u4/u6 biogenesis, and spliceosome recycling. *RNA*, 12(5) :765–774, 2006.
- [50] Hung-Ta Chen, Linda Warfield, and Steven Hahn. The positions of tffi and tffie in the rna polymerase ii transcription preinitiation complex. *Nat Struct Mol Biol*, 14(8) :696–703, Aug 2007.

- [51] J Y Chen, L Stands, J P Staley, R R Jr Jackups, L J Latus, and T H Chang. Specific alterations of u1-c protein or u1 small nuclear rna can eliminate the requirement of prp28p, an essential dead box splicing factor. *Mol Cell*, 7(1) :227–232, 2001.
- [52] Yen-I G Chen, Roger E Moore, Helen Y Ge, Mary K Young, Terry D Lee, and Scott W Stevens. Proteomic analysis of in vivo-assembled pre-mrna splicing complexes expands the catalog of participating factors. *Nucleic Acids Res*, 35(12) :3928–3944, 2007.
- [53] H V Colot, F Stutz, and M Rosbash. The yeast splicing factor mud13p is a commitment complex component and corresponds to cbp20, the small subunit of the nuclear cap-binding complex. *Genes Dev*, 10(13) :1699–1708, 1996.
- [54] Leighton J Core and John T Lis. Transcription regulation through promoter-proximal pausing of rna polymerase ii. *Science*, 319(5871) :1791–1792, 2008.
- [55] Maria Pia Cosma. Ordered recruitment : gene-specific mechanism of transcription activation. *Mol Cell*, 10(2) :227–236, 2002.
- [56] B Coulombe and Z F Burton. Dna bending and wrapping around rna polymerase : a "revolutionary" model describing transcriptional mechanisms. *Microbiol Mol Biol Rev*, 63(2) :457–478, Jun 1999.
- [57] P Cramer, K-J Armache, S Baumli, S Benkert, F Brueckner, C Buchen, G E Damsma, S Dengl, S R Geiger, A J Jasiak, A Jawhari, S Jennebach, T Kamenski, H Kettenberger, C-D Kuhn, E Lehmann, K Leike, J F Sydow, and A Vannini. Structure of eukaryotic rna polymerases. *Annu Rev Biophys*, 37 :337–352, 2008.
- [58] P Cramer, D A Bushnell, and R D Kornberg. Structural basis of transcription : Rna polymerase ii at 2.8 angstrom resolution. *Science*, 292(5523) :1863–1876, 2001.
- [59] Christian Kroun Damgaard, Soren Kahns, Soren Lykke-Andersen, Anders Lade Nielsen, Torben Heick Jensen, and Jorgen Kjems. A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell*, 29(2) :271–278, 2008.
- [60] J C Dantonel, J M Wurtz, O Poch, D Moras, and L Tora. The tbp-like factor : an alternative transcription factor in metazoa? *Trends Biochem Sci*, 24(9) :335–339, Sep 1999.
- [61] S A Darst. Bacterial rna polymerase. *Curr Opin Struct Biol*, 11(2) :155–162, 2001.
- [62] Xavier Darzacq, Robert H Singer, and Yaron Shav-Tal. Dynamics of transcription and mrna export. *Curr Opin Cell Biol*, 17(3) :332–339, 2005.
- [63] R Das, Z Zhou, and R Reed. Functional association of u2 snrnp with the atp-independent spliceosomal complex e. *Mol Cell*, 5(5) :779–787, 2000.
- [64] Rita Das, Kobina Dufu, Ben Romney, Megan Feldt, Mark Elenko, and Robin Reed. Functional coupling of rnap ii transcription to spliceosome assembly. *Genes Dev*, 20(9) :1100–1109, 2006.
- [65] Rita Das, Jiong Yu, Zuo Zhang, Melanie P Gygi, Adrian R Krainer, Steven P Gygi, and Robin Reed. Sr proteins function in coupling rnap ii transcription to pre-mrna splicing. *Mol Cell*, 26(6) :867–881, 2007.
- [66] Sergio F de Almeida and Maria Carmo-Fonseca. The ctd role in cotranscriptional rna processing and surveillance. *FEBS Lett*, 582(14) :1971–1976, 2008.
- [67] Manuel de la Mata, Claudio R Alonso, Sebastian Kadener, Juan P Fededa, Matias Blaustein, Federico Pelisch, Paula Cramer, David Bentley, and Alberto R Kornblihtt. A slow rna polymerase ii affects alternative splicing in vivo. *Mol Cell*, 12(2) :525–532, 2003.
- [68] Manuel de la Mata and Alberto R Kornblihtt. Rna polymerase ii c-terminal domain mediates regulation of alternative splicing by srp20. *Nat Struct Mol Biol*, 13(11) :973–980, 2006.
- [69] Genevieve Deblois and Vincent Giguere. Ligand-independent coactivation of eralpha af-1 by steroid receptor rna activator (sra) via mapk activation. *J Steroid Biochem Mol Biol*, 85(2-5) :123–131, 2003.
- [70] Jochen Deckert, Klaus Hartmuth, Daniel Boehringer, Nastaran Behzadnia, Cindy L Will, Berthold Kastner, Holger Stark, Henning Urlaub, and Reinhard Luhrmann. Protein composition and electron microscopy structure of affinity-purified human spliceosomal b complexes isolated under physiological conditions. *Mol Cell Biol*, 26(14) :5528–5543, 2006.

- [71] Michael D Diem, Chia C Chan, Ihab Younis, and Gideon Dreyfuss. Pym binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat Struct Mol Biol*, 14(12) :1173–1179, 2007.
- [72] R Drapkin, J T Reardon, A Ansari, J C Huang, L Zawel, K Ahn, A Sancar, and D Reinberg. Dual role of TFIID in DNA excision repair and in transcription by RNA polymerase II. *Nature*, 368(6473) :769–772, Apr 1994.
- [73] Michael J Dye, Natalia Gromak, and Nick J Proudfoot. Exon tethering in transcription by RNA polymerase II. *Mol Cell*, 21(6) :849–859, 2006.
- [74] R H Ebricht. RNA polymerase : structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol*, 304(5) :687–698, 2000.
- [75] Sylvain Egloff and Shona Murphy. Cracking the RNA polymerase II CTD code. *Trends Genet*, 24(6) :280–288, 2008.
- [76] Sylvain Egloff, Dawn O'Reilly, Rob D Chapman, Alice Taylor, Katrin Tanzhaus, Laura Pitts, Dirk Eick, and Shona Murphy. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, 318(5857) :1777–1779, 2007.
- [77] I C Eperon, O V Makarova, A Mayeda, S H Munroe, J F Caceres, D G Hayward, and A R Krainer. Selection of alternative 5' splice sites : role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol*, 20(22) :8303–8318, 2000.
- [78] Cyril Esnault, Yad Ghavi-Helm, Sylvain Brun, Julie Soutourina, Nynke Van Berkum, Claire Boschiero, Frank Holstege, and Michel Werner. Mediator-dependent recruitment of TFIID modules in preinitiation complex. *Mol Cell*, 31(3) :337–346, Aug 2008.
- [79] Celso A Espinoza, Tiffany A Allen, Aaron R Hieb, Jennifer F Kugel, and James A Goodrich. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol*, 11(9) :822–829, 2004.
- [80] Celso A Espinoza, James A Goodrich, and Jennifer F Kugel. Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA*, 13(4) :583–596, 2007.
- [81] William G Fairbrother, Ru-Fang Yeh, Phillip A Sharp, and Christopher B Burge. Predictive identification of exonic splicing enhancers in human genes. *Science*, 297(5583) :1007–1013, 2002.
- [82] S M Fang and Z F Burton. RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. *J Biol Chem*, 271(20) :11703–11709, May 1996.
- [83] Robert P Fisher. Secrets of a double agent : Cdk7 in cell-cycle control and transcription. *J Cell Sci*, 118(Pt 22) :5171–5180, Nov 2005.
- [84] Peter C FitzGerald, Andrey Shlyakhtenko, Alain A Mir, and Charles Vinson. Clustering of DNA sequences in human promoters. *Genome Res*, 14(8) :1562–1574, 2004.
- [85] J Fleckner, M Zhang, J Valcarcel, and M R Green. U2AF65 recruits a novel human dead box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev*, 11(14) :1864–1872, 1997.
- [86] N Fong and D L Bentley. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II : different functions for different segments of the CTD. *Genes Dev*, 15(14) :1783–1795, 2001.
- [87] Nova Fong, Gregory Bird, Marc Vigneron, and David L Bentley. A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3' end processing. *EMBO J*, 22(16) :4274–4282, 2003.
- [88] Y W Fong and Q Zhou. Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, 414(6866) :929–933, 2001.
- [89] Patrik Forch, Livia Merendino, Concepcion Martinez, and Juan Valcarcel. U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor of 65 kDa, U2AF65, can promote U1 snRNP recruitment to 5' splice sites. *Biochem J*, 372(Pt 1) :235–240, 2003.

- [90] Patrik Forch, Oscar Puig, Concepcion Martinez, Bertrand Seraphin, and Juan Valcarcel. The splicing regulator tia-1 interacts with u1-c to promote u1 snrnp recruitment to 5' splice sites. *EMBO J*, 21(24) :6882–6892, 2002.
- [91] Kyle Friend, Alexander F Lovejoy, and Joan A Steitz. U2 snrnp binds intronless histone pre-mrnas to facilitate u7-snrnp-dependent 3' end formation. *Mol Cell*, 28(2) :240–252, 2007.
- [92] Koh Fujinaga, Dan Irwin, Yehong Huang, Ran Taube, Takeshi Kurosu, and B Matija Peterlin. Dynamics of human immunodeficiency virus transcription : P-tefb phosphorylates rd and dissociates negative effectors from the transactivation response element. *Mol Cell Biol*, 24(2) :787–795, 2004.
- [93] Dahlene Fusco, Nathalie Accornero, Brigitte Lavoie, Shailesh M Shenoy, Jean Marie Blanchard, Robert H Singer, and Edouard Bertrand. Single mrna molecules demonstrate probabilistic movement in living mammalian cells. *Curr Biol*, 13(2) :161–167, Jan 2003.
- [94] Delphine Galiana-Arnoux, Catherine Dostert, Anette Schneemann, Jules A Hoffmann, and Jean-Luc Imler. Essential function in vivo for dicer-2 in host defense against rna viruses in drosophila. *Nat Immunol*, 7(6) :590–597, 2006.
- [95] Alexander Gasch, Silke Wiesner, Pau Martin-Malpartida, Ximena Ramirez-Espain, Lidia Ruiz, and Maria J Macias. The structure of prp40 ff1 domain and its interaction with the crn-tpr1 motif of clfl gives a new insight into the binding mode of ff domains. *J Biol Chem*, 281(1) :356–364, 2006.
- [96] H P Gerber, M Hagmann, K Seipel, O Georgiev, M A West, Y Litingtung, W Schaffner, and J L Corden. Rna polymerase ii c-terminal domain required for enhancer-driven transcription. *Nature*, 374(6523) :660–662, 1995.
- [97] Megha Ghildiyal, Herve Seitz, Michael D Horwich, Chengjian Li, Tingting Du, Soohyun Lee, Jia Xu, Ellen L W Kittler, Maria L Zapp, Zhiping Weng, and Phillip D Zamore. Endogenous sirnas derived from transposons and mrnas in drosophila somatic cells. *Science*, 320(5879) :1077–1081, 2008.
- [98] Giuseppina Giglia-Mari, Frederic Coin, Jeffrey A Ranish, Deborah Hoogstraten, Arjan Theil, Nils Wijgers, Nicolaas G J Jaspers, Anja Raams, Manuela Argentini, P J van der Spek, Elena Botta, Miria Stefanini, Jean-Marc Egly, Ruedi Aebersold, Jan H J Hoeijmakers, and Wim Vermeulen. A new, tenth subunit of tfiih is responsible for the dna repair syndrome trichothiodystrophy group a. *Nat Genet*, 36(7) :714–719, Jul 2004.
- [99] Nina Gildehaus, Thomas Neusser, Reinhild Wurm, and Rolf Wagner. Studies on the function of the riboregulator 6s rna from e. coli : Rna polymerase binding, inhibition of in vitro transcription and synthesis of rna-directed de novo transcripts. *Nucleic Acids Res*, 35(6) :1885–1896, 2007.
- [100] A L Gnatt, P Cramer, J Fu, D A Bushnell, and R D Kornberg. Structural basis of transcription : an rna polymerase ii elongation complex at 3.3 a resolution. *Science*, 292(5523) :1876–1882, 2001.
- [101] Bernd Goede, Souad Naji, Oliver von Kampen, Karin Ilg, and Michael Thomm. Protein-protein interactions in the archaeal transcriptional machinery : binding studies of isolated rna polymerase subunits and transcription factors. *J Biol Chem*, 281(41) :30581–30592, 2006.
- [102] Xue Q Gong, Yuri A Nedialkov, and Zachary F Burton. Alpha-amanitin blocks translocation by human rna polymerase ii. *J Biol Chem*, 279(26) :27422–27427, 2004.
- [103] J A Goodrich, T Hoey, C J Thut, A Admon, and R Tjian. Drosophila tafii40 interacts with both a vp16 activation domain and the basal transcription factor tfiib. *Cell*, 75(3) :519–530, Nov 1993.
- [104] Janina Gornemann, Kimberly M Kotovic, Katja Hujer, and Karla M Neugebauer. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell*, 19(1) :53–63, 2005.
- [105] M Gossen and H Bujard. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A*, 89(12) :5547–5551, 1992.
- [106] M Gossen, S Freundlieb, G Bender, G Muller, W Hillen, and H Bujard. Transcriptional activation by tetracyclines in mammalian cells. *Science*, 268(5218) :1766–1769, 1995.
- [107] J M Gottesfeld and D J Forbes. Mitotic repression of the transcriptional machinery. *Trends Biochem Sci*, 22(6) :197–202, 1997.

- [108] O Gozani, J Potashkin, and R Reed. A potential role for u2af-sap 155 interactions in recruiting u2 snrnp to the branch site. *Mol Cell Biol*, 18(8) :4752–4760, 1998.
- [109] Richard J Grainger and Jean D Beggs. Prp8 protein : at the heart of the spliceosome. *RNA*, 11(5) :533–557, 2005.
- [110] P A Grant, D Schieltz, M G Pray-Grant, J R 3rd Yates, and J L Workman. The atm-related cofactor tral is a component of the purified saga complex. *Mol Cell*, 2(6) :863–867, Dec 1998.
- [111] Natalia Gromak, Gabriele Talotti, Nicholas J Proudfoot, and Franco Pagani. Modulating alternative splicing by cotranscriptional cleavage of nascent intronic rna. *RNA*, 14(2) :359–366, 2008.
- [112] Ana Rita Grosso, Sandra Martins, and Maria Carmo-Fonseca. The emerging role of splicing factors in cancer. *EMBO Rep*, 9(11) :1087–1093, 2008.
- [113] Lalith S Gunawardane, Kuniaki Saito, Kazumichi M Nishida, Keita Miyoshi, Yoshinori Kawamura, Tomoko Nagami, Haruhiko Siomi, and Mikiko C Siomi. A slicer-mediated mechanism for repeat-associated sirna 5' end formation in drosophila. *Science*, 315(5818) :1587–1590, 2007.
- [114] S I Gunderson, S Vagner, M Polycarpou-Schwarz, and I W Mattaj. Involvement of the carboxyl terminus of vertebrate poly(a) polymerase in u1a autoregulation and in the coupling of splicing and polyadenylation. *Genes Dev*, 11(6) :761–773, 1997.
- [115] I Ha, S Roberts, E Maldonado, X Sun, L U Kim, M Green, and D Reinberg. Multiple functional domains of human transcription factor iib : distinct interactions with two general transcription factors and rna polymerase ii. *Genes Dev*, 7(6) :1021–1032, Jun 1993.
- [116] Irina Hacker, Bjoern Sander, Monika M Golas, Elmar Wolf, Elif Karagoz, Berthold Kastner, Holger Stark, Patrizia Fabrizio, and Reinhard Luhrmann. Localization of prp8, brr2, snu14 and u4/u6 proteins in the yeast tri-snrnp by electron microscopy. *Nat Struct Mol Biol*, 15(11) :1206–1212, 2008.
- [117] Steven Hahn. Structure and mechanism of the rna polymerase ii transcription machinery. *Nat Struct Mol Biol*, 11(5) :394–403, 2004.
- [118] Michael Hampsey. The pol ii initiation complex : finding a place to start. *Nat Struct Mol Biol*, 13(7) :564–566, 2006.
- [119] Esme C Hatchell, Shane M Colley, Dianne J Beveridge, Michael R Epis, Lisa M Stuart, Keith M Giles, Andrew D Redfern, Lauren E C Miles, Andrew Barker, Louisa M MacDonald, Peter G Arthur, James C K Lui, Jemma L Golding, Ross K McCulloch, Cecily B Metcalf, Jackie A Wilce, Matthew C J Wilce, Rainer B Lanz, Bert W O'Malley, and Peter J Leedman. Slirp, a small sra binding protein, is a nuclear receptor corepressor. *Mol Cell*, 22(5) :657–668, 2006.
- [120] Stephane Hausmann, Hisashi Koiwa, Shankarling Krishnamurthy, Michael Hampsey, and Stewart Shuman. Different strategies for carboxyl-terminal domain (ctd) recognition by serine 5-specific ctd phosphatases. *J Biol Chem*, 280(45) :37681–37688, 2005.
- [121] Stephane Hausmann and Stewart Shuman. Characterization of the ctd phosphatase fcp1 from fission yeast. preferential dephosphorylation of serine 2 versus serine 5. *J Biol Chem*, 277(24) :21213–21220, 2002.
- [122] Martin J Hicks, Chin-Rang Yang, Matthew V Kotlajich, and Klemens J Hertel. Linking splicing to pol ii transcription stabilizes pre-mrnas and influences splicing patterns. *PLoS Biol*, 4(6) :e147, 2006.
- [123] Aaron R Hieb, Wayne A Halsey, Meredith D Betterton, Thomas T Perkins, Jennifer F Kugel, and James A Goodrich. Tfia changes the conformation of the dna in tbp/tata complexes and increases their kinetic stability. *J Mol Biol*, 372(3) :619–632, Sep 2007.
- [124] J Hindley. Fractionation of 32p-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *J Mol Biol*, 30(1) :125–136, 1967.
- [125] Y Hirose and J L Manley. Rna polymerase ii and the integration of nuclear events. *Genes Dev*, 14(12) :1415–1429, 2000.
- [126] Yutaka Hirose and Yoshiaki Ohkuma. Phosphorylation of the c-terminal domain of rna polymerase ii plays central roles in the integrated events of eucaryotic gene expression. *J Biochem*, 141(5) :601–608, 2007.

- [127] C K Ho and S Shuman. Distinct roles for ctd ser-2 and ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mrna capping enzyme. *Mol Cell*, 3(3) :405–411, 1999.
- [128] Amy E House and Kristen W Lynch. Regulation of alternative splicing : more than just the abcs. *J Biol Chem*, 283(3) :1217–1221, 2008.
- [129] L M Hsu, J Zagorski, Z Wang, and M J Fournier. Escherichia coli 6s rna gene is part of a dual-function transcription unit. *J Bacteriol*, 161(3) :1162–1170, 1985.
- [130] Florent Hube, Jimin Guo, Shilpa Chooniedass-Kothari, Charlton Cooper, Mohammad K Hamedani, Alexander A Dibrov, Anne A A Blanchard, Xuemei Wang, George Deng, Yvonne Myal, and Etienne Leygue. Alternative splicing of the first intron of the steroid receptor rna activator (sra) participates in the generation of coding and noncoding rna isoforms in breast cancer cell lines. *DNA Cell Biol*, 25(7) :418–428, 2006.
- [131] Susan M Janicki, Toshiro Tsukamoto, Simone E Salghetti, William P Tansey, Ravi Sachidanandam, Kannanganattu V Prasanth, Thomas Ried, Yaron Shav-Tal, Edouard Bertrand, Robert H Singer, and David L Spector. From silencing to gene expression : real-time analysis in single cells. *Cell*, 116(5) :683–698, Mar 2004.
- [132] Kristina M Johnson, Jin Wang, Andrea Smallwood, Charina Arayata, and Michael Carey. Tfid and human mediator coactivator complexes assemble cooperatively on promoter dna. *Genes Dev*, 16(14) :1852–1863, Jul 2002.
- [133] Janice C Jones, Hemali P Phatnani, Timothy A Haystead, Justin A MacDonald, S Munir Alam, and Arno L Greenleaf. C-terminal repeat domain kinase i phosphorylates ser2 and ser5 of rna polymerase ii c-terminal domain repeats. *J Biol Chem*, 279(24) :24957–24964, 2004.
- [134] Melissa S Jurica and Melissa J Moore. Pre-mrna splicing : awash in a sea of proteins. *Mol Cell*, 12(1) :5–14, 2003.
- [135] Melissa S Jurica, Duncan Sousa, Melissa J Moore, and Nikolaus Grigorieff. Three-dimensional structure of c complex spliceosomes by electron microscopy. *Nat Struct Mol Biol*, 11(3) :265–269, 2004.
- [136] Tamar Juven-Gershon, Jer-Yuan Hsu, Joshua Wm Theisen, and James T Kadonaga. The rna polymerase ii core promoter - the gateway to transcription. *Curr Opin Cell Biol*, 20(3) :253–259, 2008.
- [137] Sei Kameoka, Paula Duque, and Maria M Konarska. p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *EMBO J*, 23(8) :1782–1791, 2004.
- [138] Achillefs N Kapanidis, Emmanuel Margeat, Sam On Ho, Ekaterine Kortkhonjia, Shimon Weiss, and Richard H Ebricht. Initial transcription by rna polymerase proceeds through a dna-scrunching mechanism. *Science*, 314(5802) :1144–1147, 2006.
- [139] Craig D Kaplan, Karl-Magnus Larsson, and Roger D Kornberg. The rna polymerase ii trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell*, 30(5) :547–556, 2008.
- [140] Rotem Karni, Elisa de Stanchina, Scott W Lowe, Rahul Sinha, David Mu, and Adrian R Krainer. The gene encoding the splicing factor sf2/asf is a proto-oncogene. *Nat Struct Mol Biol*, 14(3) :185–193, 2007.
- [141] Hideya Kawaji and Yoshihide Hayashizaki. Exploration of small rnas. *PLoS Genet*, 4(1) :e22, Jan 2008.
- [142] W G Kelly, M E Dahmus, and G W Hart. Rna polymerase ii is a glycoprotein. modification of the cooh-terminal domain by o-glcnaac. *J Biol Chem*, 268(14) :10416–10424, 1993.
- [143] Michael-Christopher Keogh, Siavash K Kurdistani, Stephanie A Morris, Seong Hoon Ahn, Vladimir Podolny, Sean R Collins, Maya Schuldiner, Kayu Chin, Thanuja Punna, Natalie J Thompson, Charles Boone, Andrew Emili, Jonathan S Weissman, Timothy R Hughes, Brian D Strahl, Michael Grunstein, Jack F Greenblatt, Stephen Buratowski, and Nevan J Krogan. Cotranscriptional set2 methylation of histone h3 lysine 36 recruits a repressive rpd3 complex. *Cell*, 123(4) :593–605, 2005.
- [144] Hubert Kettenberger, Karim-Jean Armache, and Patrick Cramer. Complete rna polymerase ii elongation complex structure and its interactions with ntp and tfiis. *Mol Cell*, 16(6) :955–965, 2004.

- [145] Hubert Kettenberger, Alexander Eisenfuhr, Florian Brueckner, Mirko Theis, Michael Famulok, and Patrick Cramer. Structure of an rna polymerase ii-rna inhibitor complex elucidates transcription regulation by noncoding rnas. *Nat Struct Mol Biol*, 13(1) :44–48, 2006.
- [146] E Kim, L Du, D B Bregman, and S L Warren. Splicing factors associate with hyperphosphorylated rna polymerase ii in the absence of pre-mrna. *J Cell Biol*, 136(1) :19–28, 1997.
- [147] Kwang-sun Kim and Younghoon Lee. Regulation of 6s rna biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res*, 32(20) :6057–6068, 2004.
- [148] T K Kim, R H Ebright, and D Reinberg. Mechanism of atp-dependent promoter melting by transcription factor iih. *Science*, 288(5470) :1418–1422, 2000.
- [149] D W King and M L Barnhisel. Synthesis of rna in mammalian cells during mitosis and interphase. *J Cell Biol*, 33(2) :265–272, 1967.
- [150] T Kiss. Small nucleolar rna-guided post-transcriptional modification of cellular rnas. *EMBO J*, 20(14) :3617–3622, Jul 2001.
- [151] Tamas Kiss. Biogenesis of small nuclear rnps. *J Cell Sci*, 117(Pt 25) :5949–5951, 2004.
- [152] A L Kistler and C Guthrie. Deletion of mud2, the yeast homolog of u2af65, can bypass the requirement for sub2, an essential spliceosomal atpase. *Genes Dev*, 15(1) :42–49, 2001.
- [153] Kelby O Kizer, Hemali P Phatnani, Yoichiro Shibata, Hana Hall, Arno L Greenleaf, and Brian D Strahl. A novel domain in set2 mediates rna polymerase ii interaction and couples histone h3 k36 methylation with transcript elongation. *Mol Cell Biol*, 25(8) :3305–3316, 2005.
- [154] J D Kohtz, S F Jamison, C L Will, P Zuo, R Luhmann, M A Garcia-Blanco, and J L Manley. Protein-protein interactions and 5'-splice-site recognition in mammalian mrna precursors. *Nature*, 368(6467) :119–124, 1994.
- [155] P Komarnitsky, E J Cho, and S Buratowski. Different phosphorylated forms of rna polymerase ii and associated mrna processing factors during transcription. *Genes Dev*, 14(19) :2452–2460, 2000.
- [156] Maria M Konarska and Charles C Query. Insights into the mechanisms of splicing : more lessons from the ribosome. *Genes Dev*, 19(19) :2255–2260, 2005.
- [157] Maria M Konarska, Josep Vilardell, and Charles C Query. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol Cell*, 21(4) :543–553, 2006.
- [158] Krishna P Kota, Stefan R Wagner, Elvira Huerta, Jean M Underwood, and Jeffrey A Nickerson. Binding of atp to uap56 is necessary for mrna export. *J Cell Sci*, 121(Pt 9) :1526–1537, 2008.
- [159] Kimberly M Kotovic, Daniel Lockshon, Lamia Boric, and Karla M Neugebauer. Cotranscriptional recruitment of the u1 snrnp to intron-containing genes in yeast. *Mol Cell Biol*, 23(16) :5768–5779, 2003.
- [160] A Kramer and U Utans. Three protein factors (sf1, sf3 and u2af) function in pre-splicing complex formation in addition to snrnps. *EMBO J*, 10(6) :1503–1509, 1991.
- [161] Nevan J Krogan, Minkyu Kim, Amy Tong, Ashkan Golshani, Gerard Cagney, Veronica Canadien, Dawn P Richards, Bryan K Beattie, Andrew Emili, Charles Boone, Ali Shilatifard, Stephen Buratowski, and Jack Greenblatt. Methylation of histone h3 by set2 in *saccharomyces cerevisiae* is linked to transcriptional elongation by rna polymerase ii. *Mol Cell Biol*, 23(12) :4207–4218, 2003.
- [162] Jason N Kuehner and David A Brow. Quantitative analysis of in vivo initiator selection by yeast rna polymerase ii supports a scanning model. *J Biol Chem*, 281(20) :14119–14128, 2006.
- [163] Andreas N Kuhn, Elizabeth M Reichl, and David A Brow. Distinct domains of splicing factor prp8 mediate different aspects of spliceosome activation. *Proc Natl Acad Sci U S A*, 99(14) :9145–9149, 2002.
- [164] L Kuras and K Struhl. Binding of tbp to promoters in vivo is stimulated by activators and requires pol ii holoenzyme. *Nature*, 399(6736) :609–613, Jun 1999.
- [165] T Kurisu, T Tanaka, J Ishii, K Matsumura, K Sugimura, T Nakatani, and H Kawashima. Expression and function of human steroid receptor rna activator in prostate cancer cells : role of endogenous hsra protein in androgen receptor-mediated transcription. *Prostate Cancer Prostatic Dis*, 9(2) :173–178, 2006.

- [166] Kon Yew Kwek, Shona Murphy, Andre Furger, Benjamin Thomas, William O’Gorman, Hiroshi Kimura, Nick J Proudfoot, and Alexandre Akoulitchev. U1 snrna associates with tfiih and regulates transcriptional initiation. *Nat Struct Biol*, 9(11) :800–805, 2002.
- [167] Andrea Kyburz, Arno Friedlein, Hanno Langen, and Walter Keller. Direct interactions between subunits of cpsf and the u2 snrnp contribute to the coupling of pre-mrna 3’ end processing and splicing. *Mol Cell*, 23(2) :195–205, 2006.
- [168] Christina Kyriakopoulou, Pontus Larsson, Lei Liu, Jens Schuster, Fredrik Soderbom, Leif A Kirsebom, and Anders Virtanen. U1-like snrnas lacking complementarity to canonical 5’ splice sites. *RNA*, 12(9) :1603–1611, Sep 2006.
- [169] Scott A Lacadie and Michael Rosbash. Cotranscriptional spliceosome assembly dynamics and the role of u1 snrna :5’ss base pairing in yeast. *Mol Cell*, 19(1) :65–75, 2005.
- [170] A Lacy-Hulbert, R Thomas, X P Li, C E Lilley, R S Coffin, and J Roes. Interruption of coding sequences by heterologous introns can enhance the functional expression of recombinant genes. *Gene Ther*, 8(8) :649–653, 2001.
- [171] T Lagrange, A N Kapanidis, H Tang, D Reinberg, and R H Ebright. New core promoter element in rna polymerase ii-dependent transcription : sequence-specific dna binding by transcription factor iib. *Genes Dev*, 12(1) :34–44, Jan 1998.
- [172] Angus I Lamond and David L Spector. Nuclear speckles : a model for nuclear organelles. *Nat Rev Mol Cell Biol*, 4(8) :605–612, 2003.
- [173] E S Lander, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, K Dewar, M Doyle, W FitzHugh, R Funke, D Gage, K Harris, A Heaford, J Howland, L Kann, J Lehoczky, R LeVine, P McEwan, K McKernan, J Meldrim, J P Mesirov, C Miranda, W Morris, J Naylor, C Raymond, M Rosetti, R Santos, A Sheridan, C Sougnez, N Stange-Thomann, N Stojanovic, A Subramanian, D Wyman, J Rogers, J Sulston, R Ainscough, S Beck, D Bentley, J Burton, C Clee, N Carter, A Coulson, R Deadman, P Deloukas, A Dunham, I Dunham, R Durbin, L French, D Grafham, S Gregory, T Hubbard, S Humphray, A Hunt, M Jones, C Lloyd, A McMurray, L Matthews, S Mercer, S Milne, J C Mullikin, A Mungall, R Plumb, M Ross, R Shownkeen, S Sims, R H Waterston, R K Wilson, L W Hillier, J D McPherson, M A Marra, E R Mardis, L A Fulton, A T Chinwalla, K H Pepin, W R Gish, S L Chissoe, M C Wendl, K D Delehaunty, T L Miner, A Delehaunty, J B Kramer, L L Cook, R S Fulton, D L Johnson, P J Minx, S W Clifton, T Hawkins, E Branscomb, P Predki, P Richardson, S Wenning, T Slezak, N Doggett, J F Cheng, A Olsen, S Lucas, C Elkin, E Uberbacher, M Frazier, R A Gibbs, D M Muzny, S E Scherer, J B Bouck, E J Sodergren, K C Worley, C M Rives, J H Gorrell, M L Metzker, S L Naylor, R S Kucherlapati, D L Nelson, G M Weinstock, Y Sakaki, A Fujiyama, M Hattori, T Yada, A Toyoda, T Itoh, C Kawagoe, H Watanabe, Y Totoki, T Taylor, J Weissenbach, R Heilig, W Saurin, F Artiguenave, P Brottier, T Bruls, E Pelletier, C Robert, P Wincker, D R Smith, L Doucette-Stamm, M Rubenfield, K Weinstock, H M Lee, J Dubois, A Rosenthal, M Platzer, G Nyakatura, S Taudien, A Rump, H Yang, J Yu, J Wang, G Huang, J Gu, L Hood, L Rowen, A Madan, S Qin, R W Davis, N A Federspiel, A P Abola, M J Proctor, R M Myers, J Schmutz, M Dickson, J Grimwood, D R Cox, M V Olson, R Kaul, C Raymond, N Shimizu, K Kawasaki, S Minoshima, G A Evans, M Athanasiou, R Schultz, B A Roe, F Chen, H Pan, J Ramser, H Lehrach, R Reinhardt, W R McCombie, M de la Bastide, N Dedhia, H Blocker, K Hornischer, G Nordsieck, R Agarwala, L Aravind, J A Bailey, A Bateman, S Batzoglou, E Birney, P Bork, D G Brown, C B Burge, L Cerutti, H C Chen, D Church, M Clamp, R R Copley, T Doerks, S R Eddy, E E Eichler, T S Furey, J Galagan, J G Gilbert, C Harmon, Y Hayashizaki, D Haussler, H Hermjakob, K Hokamp, W Jang, L S Johnson, T A Jones, S Kasif, A Kasprzyk, S Kennedy, W J Kent, P Kitts, E V Koonin, I Korf, D Kulp, D Lancet, T M Lowe, A McLysaght, T Mikkelsen, J V Moran, N Mulder, V J Pollara, C P Ponting, G Schuler, J Schultz, G Slater, A F Smit, E Stupka, J Szustakowski, D Thierry-Mieg, J Thierry-Mieg, L Wagner, J Wallis, R Wheeler, A Williams, Y I Wolf, K H Wolfe, S P Yang, R F Yeh, F Collins, M S Guyer, J Peterson, A Felsenfeld, K A Wetterstrand, A Patrinos, M J Morgan, P de Jong, J J Catanese, K Osoegawa, H Shizuya, S Choi, and Y J Chen. Initial sequencing and analysis of the human genome. *Nature*, 409(6822) :860–921, 2001.
- [174] MF ;Trinh V. Langelier. Gros plan sur l’arn polymérase ii. *Médecine/Sciences*, 18 :210–216, 2002.

- [175] R B Lanz, N J McKenna, S A Onate, U Albrecht, J Wong, S Y Tsai, M J Tsai, and B W O'Malley. A steroid receptor coactivator, sra, functions as an rna and is present in an src-1 complex. *Cell*, 97(1) :17–27, 1999.
- [176] H Le Hir, E Izaurralde, L E Maquat, and M J Moore. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mrna exon-exon junctions. *EMBO J*, 19(24) :6860–6869, 2000.
- [177] T I Lee and R A Young. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet*, 34 :77–137, 2000.
- [178] Brian A Lewis, Robert J 3rd Sims, William S Lane, and Danny Reinberg. Functional characterization of core promoter elements : Dpe-specific transcription requires the protein kinase ck2 and the pc4 coactivator. *Mol Cell*, 18(4) :471–481, 2005.
- [179] J D Lewis, E Izaurralde, A Jarmolowski, C McGuigan, and I W Mattaj. A nuclear cap-binding complex facilitates association of u1 snrnp with the cap-proximal 5' splice site. *Genes Dev*, 10(13) :1683–1698, 1996.
- [180] Bing Li, Madelaine Gogol, Mike Carey, Samantha G Pattenden, Chris Seidel, and Jerry L Workman. Infrequently transcribed long genes depend on the set2/rpd3s pathway for accurate transcription. *Genes Dev*, 21(11) :1422–1430, 2007.
- [181] Bing Li, LeAnn Howe, Scott Anderson, John R 3rd Yates, and Jerry L Workman. The set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of rna polymerase ii. *J Biol Chem*, 278(11) :8897–8903, 2003.
- [182] Ming Li, Hemali P Phatnani, Ziqiang Guan, Harvey Sage, Arno L Greenleaf, and Pei Zhou. Solution structure of the set2-rpb1 interacting domain of human set2 and its interaction with the hyperphosphorylated c-terminal domain of rpb1. *Proc Natl Acad Sci U S A*, 102(49) :17636–17641, 2005.
- [183] Songchun Liang and Carol S Lutz. p54nrb is a component of the snrnp-free u1a (sf-a) complex that promotes pre-mrna cleavage during polyadenylation. *RNA*, 12(1) :111–121, 2006.
- [184] Donny D Licatalosi, Gabrielle Geiger, Michelle Minet, Stephanie Schroeder, Kate Cilli, J Bryan McNeil, and David L Bentley. Functional interaction of yeast pre-mrna 3' end processing factors with rna polymerase ii. *Mol Cell*, 9(5) :1101–1111, 2002.
- [185] Zachary Lippman, Anne-Valerie Gendrel, Michael Black, Matthew W Vaughn, Neilay Dedhia, W Richard McCombie, Kimberly Lavine, Vivek Mittal, Bruce May, Kristin D Kasschau, James C Carrington, Rebecca W Doerge, Vincent Colot, and Rob Martienssen. Role of transposable elements in heterochromatin and epigenetic control. *Nature*, 430(6998) :471–476, Jul 2004.
- [186] Imke Listerman, Aparna K Sapra, and Karla M Neugebauer. Cotranscriptional coupling of splicing factor recruitment and precursor messenger rna splicing in mammalian cells. *Nat Struct Mol Biol*, 13(9) :815–822, 2006.
- [187] H X Liu, S L Chew, L Cartegni, M Q Zhang, and A R Krainer. Exonic splicing enhancer motif recognized by human sc35 under splicing conditions. *Mol Cell Biol*, 20(3) :1063–1071, 2000.
- [188] Pengda Liu, Arno L Greenleaf, and John W Stiller. The essential sequence elements required for rna polymerase ii carboxyl-terminal domain function in yeast and their evolutionary conservation. *Mol Biol Evol*, 25(4) :719–727, 2008.
- [189] Sunbin Liu, Reinhard Rauhut, Hans-Peter Vornlocher, and Reinhard Luhrmann. The network of protein-protein interactions within the human u4/u6.u5 tri-snrnp. *RNA*, 12(7) :1418–1430, 2006.
- [190] Nuria Lopez-Bigas, Benjamin Audit, Christos Ouzounis, Genis Parra, and Roderic Guigo. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett*, 579(9) :1900–1903, 2005.
- [191] H Lu, L Zawel, L Fisher, J M Egly, and D Reinberg. Human general transcription factor iih phosphorylates the c-terminal domain of rna polymerase ii. *Nature*, 358(6388) :641–645, 1992.
- [192] C S Lutz, C Cooke, J P O'Connor, R Kobayashi, and J C Alwine. The snrnp-free u1a (sf-a) complex(es) : identification of the largest subunit as psf, the polypyrimidine-tract binding protein-associated splicing factor. *RNA*, 4(12) :1493–1499, 1998.

- [193] C Lutz-Freyermuth, C C Query, and J D Keene. Quantitative determination that one of two potential rna-binding domains of the a protein component of the u1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop ii of u1 rna. *Proc Natl Acad Sci U S A*, 87(16) :6393–6397, 1990.
- [194] Kristen W Lynch. Cotranscriptional splicing regulation : it's not just about speed. *Nat Struct Mol Biol*, 13(11) :952–953, 2006.
- [195] Hadar Malca, Noam Shomron, and Gil Ast. The u1 snrnp base pairs with the 5' splice site within a penta-snrnp complex. *Mol Cell Biol*, 23(10) :3442–3455, 2003.
- [196] Valerie Manceau, Matthew Swenson, Jean-Pierre Le Caer, Andre Sobel, Clara L Kielkopf, and Alexandre Maucuer. Major phosphorylation of sf1 on adjacent ser-pro motifs enhances interaction with u2af65. *FEBS J*, 273(3) :577–587, 2006.
- [197] Tom Maniatis and Robin Reed. An extensive network of coupling among gene expression machines. *Nature*, 416(6880) :499–506, 2002.
- [198] Peter D Mariner, Ryan D Walters, Celso A Espinoza, Linda F Drullinger, Stacey D Wagner, Jennifer F Kugel, and James A Goodrich. Human alu rna is a modular transacting repressor of mrna transcription during heat shock. *Mol Cell*, 29(4) :499–509, Feb 2008.
- [199] Glenn A Maston, Sara K Evans, and Michael R Green. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*, 7 :29–59, 2006.
- [200] Arianne J Matlin, Francis Clark, and Christopher W J Smith. Understanding alternative splicing : towards a cellular code. *Nat Rev Mol Cell Biol*, 6(5) :386–398, 2005.
- [201] K Matsumoto, K M Wassarman, and A P Wolffe. Nuclear history of a pre-mrna determines the translational activity of cytoplasmic mrna. *EMBO J*, 17(7) :2107–2121, 1998.
- [202] T Max, M Sogaard, and Jesper Q Svejstrup. Hyperphosphorylation of the c-terminal repeat domain of rna polymerase ii facilitates dissociation of its complex with mediator. *J Biol Chem*, 282(19) :14113–14120, 2007.
- [203] S McCracken, N Fong, E Rosonina, K Yankulov, G Brothers, D Siderovski, A Hessel, S Foster, S Shuman, and D L Bentley. 5'-capping enzymes are targeted to pre-mrna by binding to the phosphorylated carboxy-terminal domain of rna polymerase ii. *Genes Dev*, 11(24) :3306–3318, 1997.
- [204] S McCracken, N Fong, K Yankulov, S Ballantyne, G Pan, J Greenblatt, S D Patterson, M Wickens, and D L Bentley. The c-terminal domain of rna polymerase ii couples mrna processing to transcription. *Nature*, 385(6614) :357–361, 1997.
- [205] Susan McCracken, Mark Lambermon, and Benjamin J Blencowe. Srm160 splicing coactivator promotes transcript 3'-end cleavage. *Mol Cell Biol*, 22(1) :148–160, 2002.
- [206] Anton Meinhart, Tomislav Kamenski, Sabine Hoepfner, Sonja Baumli, and Patrick Cramer. A structural perspective of ctd function. *Genes Dev*, 19(12) :1401–1415, 2005.
- [207] Gail Miller and Steven Hahn. A dna-tethered cleavage probe reveals the path for promoter dna in the yeast preinitiation complex. *Nat Struct Mol Biol*, 13(7) :603–610, 2006.
- [208] Scott Millhouse and James L Manley. The c-terminal domain of rna polymerase ii functions as a phosphorylation-dependent splicing activator in a heterologous protein. *Mol Cell Biol*, 25(2) :533–544, 2005.
- [209] T Misteli, J F Caceres, and D L Spector. The dynamics of a pre-mrna splicing factor in living cells. *Nature*, 387(6632) :523–527, 1997.
- [210] T Misteli and D L Spector. Rna polymerase ii targets pre-mrna splicing factors to transcription sites in vivo. *Mol Cell*, 3(6) :697–705, 1999.
- [211] H Mizuguchi and T Hayakawa. Characteristics of adenovirus-mediated tetracycline-controllable expression system. *Biochim Biophys Acta*, 1568(1) :21–29, 2001.
- [212] M J Moore. Intron recognition comes of age. *Nat Struct Biol*, 7(1) :14–16, 2000.

- [213] Michael J Moore, Elissa M Schwartzfarb, Pamela A Silver, and Michael C Yu. Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mrna splicing. *Mol Cell*, 24(6) :903–915, 2006.
- [214] D P Morris and A L Greenleaf. The splicing factor, prp40, binds the phosphorylated carboxyl-terminal domain of rna polymerase ii. *J Biol Chem*, 275(51) :39935–39943, 2000.
- [215] Daniel P Morris, Gregory A Michelotti, and Debra A Schwinn. Evidence that phosphorylation of the rna polymerase ii carboxyl-terminal repeats is similar in yeast and humans. *J Biol Chem*, 280(36) :31368–31377, 2005.
- [216] M J Mortillaro, B J Blencowe, X Wei, H Nakayasu, L Du, S L Warren, P A Sharp, and R Berezney. A hyperphosphorylated form of the large subunit of rna polymerase ii is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci U S A*, 93(16) :8253–8257, 1996.
- [217] Ferenc Muller, Mate A Demeny, and Laszlo Tora. New problems in rna polymerase ii transcription initiation : matching the diversity of core promoters with a variety of promoter recognition factors. *J Biol Chem*, 282(20) :14685–14689, 2007.
- [218] Anders M Naar, Dylan J Taatjes, Weiguo Zhai, Eva Nogales, and Robert Tjian. Human crsp interacts with rna polymerase ii ctd and adopts a specific ctd-bound conformation. *Genes Dev*, 16(11) :1339–1344, 2002.
- [219] Dhanalakshmi Nair, Yeejin Kim, and Lawrence C Myers. Mediator and tfiih govern carboxyl-terminal domain-dependent transcription in yeast extracts. *J Biol Chem*, 280(40) :33739–33748, 2005.
- [220] Takashi Narita, Yuki Yamaguchi, Keiichi Yano, Seiji Sugimoto, Sittinan Chanarat, Tadashi Wada, Dongki Kim, Jun Hasegawa, Masashi Omori, Naoto Inukai, Masaki Endoh, Tomoko Yamada, and Hiroshi Handa. Human transcription elongation factor nelf : identification of novel subunits and reconstitution of the functionally active complex. *Mol Cell Biol*, 23(6) :1863–1873, 2003.
- [221] Faiz-Ul Hassan Nasim, Stephen Hutchison, Melanie Cordeau, and Benoit Chabot. High-affinity hnrnp a1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA*, 8(8) :1078–1089, 2002.
- [222] V T Nguyen, F Giannoni, M F Dubois, S J Seo, M Vigneron, C Kedinger, and O Bensaude. In vivo degradation of rna polymerase ii largest subunit triggered by alpha-amanitin. *Nucleic Acids Res*, 24(15) :2924–2929, 1996.
- [223] V T Nguyen, T Kiss, A A Michels, and O Bensaude. 7sk small nuclear rna binds to and inhibits the activity of cdk9/cyclin t complexes. *Nature*, 414(6861) :322–325, 2001.
- [224] Zhuoyu Ni, Abbie Saunders, Nicholas J Fuda, Jie Yao, Jose-Ramon Suarez, Watt W Webb, and John T Lis. P-tefb is critical for the maturation of rna polymerase ii into productive elongation in vivo. *Mol Cell Biol*, 28(3) :1161–1170, 2008.
- [225] D B Nikolov, H Chen, E D Halay, A Hoffman, R G Roeder, and S K Burley. Crystal structure of a human tata box-binding protein/tata element complex. *Proc Natl Acad Sci U S A*, 93(10) :4862–4867, May 1996.
- [226] Timothy W Nilsen. The spliceosome : no assembly required? *Mol Cell*, 9(1) :8–9, 2002.
- [227] Takayuki Nojima, Tetsuro Hirose, Hiroshi Kimura, and Masatoshi Hagiwara. The interaction between cap-binding complex and rna export factor is required for intronless mrna export. *J Biol Chem*, 282(21) :15645–15651, 2007.
- [228] M Nonet, D Sweetser, and R A Young. Functional redundancy and structural polymorphism in the large subunit of rna polymerase ii. *Cell*, 50(6) :909–915, 1987.
- [229] T O'Brien and J T Lis. Rna polymerase ii pauses at the 5' end of the transcriptionally induced drosophila hsp70 gene. *Mol Cell Biol*, 11(10) :5285–5290, 1991.
- [230] C L O'Day, G Dalbadie-McFarland, and J Abelson. The *saccharomyces cerevisiae* prp5 protein has rna-dependent atpase activity with specificity for u2 small nuclear rna. *J Biol Chem*, 271(52) :33261–33267, 1996.

- [231] Thomas Oelgeschlager. Regulation of rna polymerase ii activity by ctd phosphorylation and cell cycle control. *J Cell Physiol*, 190(2) :160–169, 2002.
- [232] William O’Gorman, Benjamin Thomas, Kon Yew Kwek, Andre Furger, and Alexandre Akoulitchev. Analysis of u1 small nuclear rna interaction with cyclin h. *J Biol Chem*, 280(44) :36920–36925, 2005.
- [233] V V Ogryzko, T Kotani, X Zhang, R L Schiltz, T Howard, X J Yang, B H Howard, J Qin, and Y Nakatani. Histone-like tails within the pcaf histone acetylase complex. *Cell*, 94(1) :35–44, Jul 1998.
- [234] Y Ohkuma, H Sumimoto, M Horikoshi, and R G Roeder. Factors involved in specific transcription by mammalian rna polymerase ii : purification and characterization of general transcription factor tflie. *Proc Natl Acad Sci U S A*, 87(23) :9163–9167, Dec 1990.
- [235] R T O’Keefe, A Mayeda, C L Sadowski, A R Krainer, and D L Spector. Disruption of pre-mrna splicing in vivo results in reorganization of splicing factors. *J Cell Biol*, 124(3) :249–260, 1994.
- [236] Masahiko Okuda, Aki Tanaka, Manami Satoh, Shoko Mizuta, Manabu Takazawa, Yoshiaki Ohkuma, and Yoshifumi Nishimura. Structural insight into the tflie-tfiih interaction : Tflie and p53 share the binding region on tfiih. *EMBO J*, 27(7) :1161–1171, Apr 2008.
- [237] Benoit Palancade and Olivier Bensaude. Investigating rna polymerase ii carboxyl-terminal domain (ctd) phosphorylation. *Eur J Biochem*, 270(19) :3859–3870, 2003.
- [238] R D Palmiter, E P Sandgren, M R Avarbock, D D Allen, and R L Brinster. Heterologous introns can enhance expression of transgenes in mice. *Proc Natl Acad Sci U S A*, 88(2) :478–482, 1991.
- [239] Shatakshi Pandit, Bert Lynn, and Brian C Rymond. Inhibition of a spliceosome turnover pathway suppresses splicing defects. *Proc Natl Acad Sci U S A*, 103(37) :13700–13705, 2006.
- [240] Shatakshi Pandit, Dong Wang, and Xiang-Dong Fu. Functional integration of transcriptional and rna processing machineries. *Curr Opin Cell Biol*, 20(3) :260–265, Jun 2008.
- [241] G G Parsons and C A Spencer. Mitotic repression of rna polymerase ii transcription is accompanied by release of transcription elongation complexes. *Mol Cell Biol*, 17(10) :5791–5802, 1997.
- [242] Snehal Bhikhu Patel and Michel Bellini. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res*, 36(20) :6482–6493, 2008.
- [243] Snehal Bhikhu Patel, Natalya Novikova, and Michel Bellini. Splicing-independent recruitment of spliceosomal small nuclear rnp to nascent rna polymerase ii transcripts. *J Cell Biol*, 178(6) :937–949, 2007.
- [244] H Peled-Zehavi, J A Berglund, M Rosbash, and A D Frankel. Recognition of rna branch point sequences by the kh domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. *Mol Cell Biol*, 21(15) :5232–5241, 2001.
- [245] Rhonda J Perriman and Manuel Jr Ares. Rearrangement of competing u2 rna helices within the spliceosome promotes multiple steps in splicing. *Genes Dev*, 21(7) :811–820, 2007.
- [246] Stephan P Persengiev, Xiaochun Zhu, Bharat L Dixit, Glenn A Maston, Ellen L W Kittler, and Michael R Green. Trf3, a tata-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proc Natl Acad Sci U S A*, 100(25) :14887–14891, Dec 2003.
- [247] B Matija Peterlin and David H Price. Controlling the elongation phase of transcription with p-tefb. *Mol Cell*, 23(3) :297–305, 2006.
- [248] R D Phair and T Misteli. High mobility of proteins in the mammalian cell nucleus. *Nature*, 404(6778) :604–609, 2000.
- [249] Hemali P Phatnani and Arno L Greenleaf. Phosphorylation and functions of the rna polymerase ii ctd. *Genes Dev*, 20(21) :2922–2936, 2006.
- [250] Hemali P Phatnani, Janice C Jones, and Arno L Greenleaf. Expanding the functional repertoire of ctd kinase i and rna polymerase ii : novel phospho-ctd-associating proteins in the yeast proteome. *Biochemistry*, 43(50) :15702–15719, 2004.
- [251] Ramesh S Pillai, Suvendra N Bhattacharyya, and Witold Filipowicz. Repression of protein synthesis by mirnas : how many mechanisms? *Trends Cell Biol*, 17(3) :118–126, 2007.

- [252] D H Price. P-tefb, a cyclin-dependent kinase controlling elongation by rna polymerase ii. *Mol Cell Biol*, 20(8) :2629–2634, 2000.
- [253] Nick J Proudfoot, Andre Furger, and Michael J Dye. Integrating mrna processing with transcription. *Cell*, 108(4) :501–512, 2002.
- [254] P L Raghunathan and C Guthrie. A spliceosomal recycling factor that reanneals u4 and u6 small nuclear ribonucleoprotein particles. *Science*, 279(5352) :857–860, 1998.
- [255] Mathieu Rappas, Daniel Bose, and Xiaodong Zhang. Bacterial enhancer-binding proteins : unlocking sigma54-dependent gene transcription. *Curr Opin Struct Biol*, 17(1) :110–116, 2007.
- [256] D Reinberg and R G Roeder. Factors involved in specific transcription by mammalian rna polymerase ii. purification and functional analysis of initiation factors iib and iie. *J Biol Chem*, 262(7) :3310–3321, Mar 1987.
- [257] L Reitzer and B L Schneider. Metabolic context and possible physiological themes of sigma(54)-dependent genes in escherichia coli. *Microbiol Mol Biol Rev*, 65(3) :422–444, 2001.
- [258] Andrey Revyakin, Chenyu Liu, Richard H Ebright, and Terence R Strick. Abortive initiation and productive initiation by rna polymerase involve dna scrunching. *Science*, 314(5802) :1139–1143, 2006.
- [259] Mariela Reyes-Reyes and Michael Hampsey. Role for the ssu72 c-terminal domain phosphatase in rna polymerase ii transcription elongation. *Mol Cell Biol*, 27(3) :926–936, 2007.
- [260] Britta M Rhode, Klaus Hartmuth, Eric Westhof, and Reinhard Luhrmann. Proximity of conserved u6 and u2 snrna elements to the 5' splice site region in activated spliceosomes. *EMBO J*, 25(11) :2475–2486, 2006.
- [261] Jose Rino, Teresa Carvalho, Jose Braga, Joana M P Desterro, Reinhard Luhrmann, and Maria Carmo-Fonseca. A stochastic view of spliceosome assembly and recycling in the nucleus. *PLoS Comput Biol*, 3(10) :2019–2031, 2007.
- [262] Dustin B Ritchie, Matthew J Schellenberg, Emily M Gesner, Sheetal A Raithatha, David T Stuart, and Andrew M Macmillan. Structural elucidation of a prp8 core domain from the heart of the spliceosome. *Nat Struct Mol Biol*, 15(11) :1199–1205, 2008.
- [263] F Robert, M Douziech, D Forget, J M Egly, J Greenblatt, Z F Burton, and B Coulombe. Wrapping of promoter dna around the rna polymerase ii initiation complex induced by tfiif. *Mol Cell*, 2(3) :341–351, Sep 1998.
- [264] G C Roberts, C Gooding, H Y Mak, N J Proudfoot, and C W Smith. Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res*, 26(24) :5568–5572, 1998.
- [265] Nicole D Robson-Dixon and Mariano A Garcia-Blanco. Maz elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon iiib. *J Biol Chem*, 279(28) :29075–29084, 2004.
- [266] F Rossi, T Forne, E Antoine, J Tazi, C Brunel, and G Cathala. Involvement of u1 small nuclear ribonucleoproteins (snrnp) in 5' splice site-u1 snrnp interaction. *J Biol Chem*, 271(39) :23985–23991, 1996.
- [267] M D Rudd and D S Luse. Amanitin greatly reduces the rate of transcription by rna polymerase ii ternary complexes but fails to inhibit some transcript cleavage modes. *J Biol Chem*, 271(35) :21549–21558, 1996.
- [268] Martin Sadowski, Bernhard Dichtl, Wolfgang Hubner, and Walter Keller. Independent functions of yeast pcf1p in pre-mrna 3' end processing and in transcription termination. *EMBO J*, 22(9) :2167–2177, 2003.
- [269] Albin Sandelin, Piero Carninci, Boris Lenhard, Jasmina Ponjavic, Yoshihide Hayashizaki, and David A Hume. Mammalian rna polymerase ii core promoters : insights from genome-wide studies. *Nat Rev Genet*, 8(6) :424–436, 2007.
- [270] Shigeo Sato, Chieri Tomomori-Sato, Tari J Parmely, Laurence Florens, Boris Zybaylov, Selene K Swanson, Charles A S Banks, Jingji Jin, Yong Cai, Michael P Washburn, Joan Weliky Conaway, and Ronald C Conaway. A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol Cell*, 14(5) :685–691, 2004.

- [271] Jerome Sauliere, Alain Sureau, Alain Expert-Bezancon, and Joelle Marie. The polypyrimidine tract binding protein (ptb) represses splicing of exon 6b from the beta-tropomyosin pre-mrna by directly interfering with the binding of the u2af65 subunit. *Mol Cell Biol*, 26(23) :8755–8769, 2006.
- [272] Matthew J Schellenberg, Dustin B Ritchie, and Andrew M MacMillan. Pre-mrna splicing : a complex picture in higher definition. *Trends Biochem Sci*, 33(6) :243–246, 2008.
- [273] D Schmucker, J C Clemens, H Shu, C A Worby, J Xiao, M Muda, J E Dixon, and S L Zipursky. Drosophila dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, 101(6) :671–684, 2000.
- [274] S C Schroeder, B Schwer, S Shuman, and D Bentley. Dynamic association of capping enzymes with transcribing rna polymerase ii. *Genes Dev*, 14(19) :2435–2440, 2000.
- [275] P Schultz, S Fribourg, A Poterszman, V Mallouh, D Moras, and J M Egly. Molecular structure of human tfiif. *Cell*, 102(5) :599–607, Sep 2000.
- [276] Philipp Selenko, Goran Gregorovic, Remco Sprangers, Gunter Stier, Zakaria Rhani, Angela Kramer, and Michael Sattler. Structural basis for the molecular recognition between human splicing factors u2af65 and sf1/mbbp. *Mol Cell*, 11(4) :965–976, 2003.
- [277] B Seraphin and M Rosbash. Identification of functional u1 snrna-pre-mrna complexes committed to spliceosome assembly and splicing. *Cell*, 59(2) :349–358, 1989.
- [278] T H Shaikh, A M Roy, J Kim, M A Batzer, and P L Deininger. cdnas derived from primary and small cytoplasmic alu (scalu) transcripts. *J Mol Biol*, 271(2) :222–234, 1997.
- [279] Hanshuang Shao, Merav Revach, Sandra Moshonov, Yael Tzuman, Kfir Gazit, Shira Albeck, Tamar Unger, and Rivka Dikstein. Core promoter binding by histone-like taf complexes. *Mol Cell Biol*, 25(1) :206–219, Jan 2005.
- [280] Yaron Shav-Tal, Xavier Darzacq, and Robert H Singer. Gene expression within a dynamic nuclear landscape. *EMBO J*, 25(15) :3469–3479, 2006.
- [281] Haihong Shen, Xuexiu Zheng, Jingping Shen, Lingdi Zhang, Rui Zhao, and Michael R Green. Distinct activities of the dexd/h-box splicing factor huap56 facilitate stepwise assembly of the spliceosome. *Genes Dev*, 22(13) :1796–1803, 2008.
- [282] Jingping Shen, Lingdi Zhang, and Rui Zhao. Biochemical characterization of the atpase and helicase activity of uap56, an essential pre-mrna splicing and mrna export factor. *J Biol Chem*, 282(31) :22544–22550, 2007.
- [283] Y Shi, M Downes, W Xie, H Y Kao, P Ordentlich, C C Tsai, M Hon, and R M Evans. Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev*, 15(9) :1140–1151, 2001.
- [284] Chanseok Shin and James L Manley. The sr protein srp38 represses splicing in m phase cells. *Cell*, 111(3) :407–417, 2002.
- [285] M Siatecka, J L Reyes, and M M Konarska. Functional interactions of prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev*, 13(15) :1983–1993, 1999.
- [286] E Allen Sickmier, Katherine E Frato, Haihong Shen, Shanthi R Paranawithana, Michael R Green, and Clara L Kielkopf. Structural basis for polypyrimidine tract recognition by the essential pre-mrna splicing factor u2af65. *Mol Cell*, 23(1) :49–59, 2006.
- [287] Titia Sijen and Ronald H A Plasterk. Transposon silencing in the caenorhabditis elegans germ line by natural rna. *Nature*, 426(6964) :310–314, Nov 2003.
- [288] Stephen T Smale and James T Kadonaga. The rna polymerase ii core promoter. *Annu Rev Biochem*, 72 :449–479, 2003.
- [289] Eliza C Small, Stephanie R Leggett, Adrienne A Winans, and Jonathan P Staley. The ef-g-like gtpase snu114p regulates spliceosome dynamics mediated by brr2p, a dexd/h box atpase. *Mol Cell*, 23(3) :389–399, 2006.
- [290] Duncan J Smith, Charles C Query, and Maria M Konarska. "nought may endure but mutability" : spliceosome dynamics and the regulation of splicing. *Mol Cell*, 30(6) :657–666, 2008.

- [291] C A Spencer, M J Kruhlak, H L Jenkins, X Sun, and D P Bazett-Jones. Mitotic transcription repression in vivo in the absence of nucleosomal chromatin condensation. *J Cell Biol*, 150(1) :13–26, 2000.
- [292] Joseph Sperling, Maia Azubel, and Ruth Sperling. Structure and function of the pre-mrna splicing machine. *Structure*, 16(11) :1605–1615, 2008.
- [293] J P Staley and C Guthrie. Mechanical devices of the spliceosome : motors, clocks, springs, and things. *Cell*, 92(3) :315–326, 1998.
- [294] Scott W Stevens, Daniel E Ryan, Helen Y Ge, Roger E Moore, Mary K Young, Terry D Lee, and John Abelson. Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell*, 9(1) :31–44, 2002.
- [295] John W Stiller and Matthew S Cook. Functional unit of the rna polymerase ii c-terminal domain lies within heptapeptide pairs. *Eukaryot Cell*, 3(3) :735–740, 2004.
- [296] Brian D Strahl, Patrick A Grant, Scott D Briggs, Zu-Wen Sun, James R Bone, Jennifer A Caldwell, Sahana Mollah, Richard G Cook, Jeffrey Shabanowitz, Donald F Hunt, and C David Allis. Set2 is a nucleosomal histone h3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol*, 22(5) :1298–1306, 2002.
- [297] Yuichiro Takagi and Roger D Kornberg. Mediator as a general transcription factor. *J Biol Chem*, 281(1) :80–89, Jan 2006.
- [298] Goranka Tanackovic and Angela Kramer. Human splicing factor sf3a, but not sf1, is essential for pre-mrna splicing in vivo. *Mol Biol Cell*, 16(3) :1366–1377, 2005.
- [299] Naoko Tanaka, Anna Aronova, and Beate Schwer. Ntr1 activates the prp43 helicase to trigger release of lariat-intron from the spliceosome. *Genes Dev*, 21(18) :2312–2325, 2007.
- [300] Ichiro Taniguchi and Mutsuhito Ohno. Atp-dependent recruitment of export factor Aly/REF onto intronless mRNAs by rna helicase UAP56. *Mol Cell Biol*, 28(2) :601–608, 2008.
- [301] Daniel F Tardiff, Katharine C Abruzzi, and Michael Rosbash. Protein characterization of *Saccharomyces cerevisiae* rna polymerase ii after in vivo cross-linking. *Proc Natl Acad Sci U S A*, 104(50) :19948–19953, 2007.
- [302] H Tian. Rna ligands generated against complex nuclear targets indicate a role for U1 snRNP in coordinating transcription and rna splicing. *FEBS Lett*, 509(2) :282–286, 2001.
- [303] Laszlo Tora. A unified nomenclature for tata box binding protein (TBP)-associated factors (TAFs) involved in rna polymerase ii transcription. *Genes Dev*, 16(6) :673–675, Mar 2002.
- [304] Amy E Trotochaud and Karen M Wassarman. 6S rna function enhances long-term cell survival. *J Bacteriol*, 186(15) :4978–4985, 2004.
- [305] Amy E Trotochaud and Karen M Wassarman. A highly conserved 6S rna structure is required for regulation of transcription. *Nat Struct Mol Biol*, 12(4) :313–319, 2005.
- [306] Amy E Trotochaud and Karen M Wassarman. 6S rna regulation of PspF transcription leads to altered cell survival at high pH. *J Bacteriol*, 188(11) :3936–3943, 2006.
- [307] F T Tsai and P B Sigler. Structural basis of preinitiation complex assembly on human pol ii promoters. *EMBO J*, 19(1) :25–36, Jan 2000.
- [308] Rong-Tzong Tsai, Chi-Kang Tseng, Pei-Jung Lee, Hsin-Chou Chen, Ru-Huei Fu, Kae-Jiun Chang, Fu-Lung Yeh, and Soo-Chen Cheng. Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. *Mol Cell Biol*, 27(23) :8027–8037, 2007.
- [309] I A Turner, C M Norman, M J Churcher, and A J Newman. Roles of the U5 snRNP in spliceosome dynamics and catalysis. *Biochem Soc Trans*, 32(Pt 6) :928–931, 2004.
- [310] Andrea Ujvari and Donal S Luse. Newly initiated rna encounters a factor involved in splicing immediately upon emerging from within rna polymerase ii. *J Biol Chem*, 279(48) :49773–49779, 2004.
- [311] J G Umen and C Guthrie. Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mrna splicing. *RNA*, 1(6) :584–597, 1995.

- [312] A Usheva, E Maldonado, A Goldring, H Lu, C Houbavi, D Reinberg, and Y Aloni. Specific interaction between the nonphosphorylated form of rna polymerase ii and the tata-binding protein. *Cell*, 69(5) :871–881, May 1992.
- [313] Vasily V Vagin, Alla Sigova, Chengjian Li, Herve Seitz, Vladimir Gvozdev, and Phillip D Zamore. A distinct small rna pathway silences selfish genetic elements in the germline. *Science*, 313(5785) :320–324, 2006.
- [314] S Vagner, U Ruegsegger, S I Gunderson, W Keller, and I W Mattaj. Position-dependent inhibition of the cleavage step of pre-mrna 3'-end processing by u1 snrnp. *RNA*, 6(2) :178–188, 2000.
- [315] S Vagner, C Vagner, and I W Mattaj. The carboxyl terminus of vertebrate poly(a) polymerase interacts with u2af 65 to couple 3'-end processing and splicing. *Genes Dev*, 14(4) :403–413, 2000.
- [316] J Valcarcel, R K Gaur, R Singh, and M R Green. Interaction of u2af65 rs region with pre-mrna branch point and promotion of base pairing with u2 snrna [corrected]. *Science*, 273(5282) :1706–1709, 1996.
- [317] Patricia Valencia, Anusha P Dias, and Robin Reed. Splicing promotes rapid and efficient mrna export in mammalian cells. *Proc Natl Acad Sci U S A*, 105(9) :3386–3391, 2008.
- [318] Raymonda Varon, Rebecca Gooding, Christina Steglich, Lorna Marns, Hua Tang, Dora Angelicheva, Kiau Kiun Yong, Petra Ambrugger, Anke Reinhold, Bharti Morar, Frank Baas, Marcel Kwa, Ivailo Tournev, Velina Guerguelcheva, Ivo Kremensky, Hanns Lochmuller, Andrea Mullner-Eidenbock, Luciano Merlini, Luitgard Neumann, Joachim Burger, Maggie Walter, Kathryn Swoboda, P K Thomas, Arpad von Moers, Neil Risch, and Luba Kalaydjieva. Partial deficiency of the c-terminal-domain phosphatase of rna polymerase ii is associated with congenital cataracts facial dysmorphism neuropathy syndrome. *Nat Genet*, 35(2) :185–189, 2003.
- [319] D W Vogel, R K Hartmann, J C Struck, N Ulbrich, and V A Erdmann. The sequence of the 6s rna gene of pseudomonas aeruginosa. *Nucleic Acids Res*, 15(11) :4583–4591, 1987.
- [320] Joseph T Wade and Kevin Struhl. The transition from transcriptional initiation to elongation. *Curr Opin Genet Dev*, 18(2) :130–136, 2008.
- [321] Eric T Wang, Rickard Sandberg, Shujun Luo, Irina Khrebtkova, Lu Zhang, Christine Mayr, Stephen F Kingsmore, Gary P Schroth, and Christopher B Burge. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221) :470–476, 2008.
- [322] Guey-Shin Wang and Thomas A Cooper. Splicing in disease : disruption of the splicing code and the decoding machinery. *Nat Rev Genet*, 8(10) :749–761, 2007.
- [323] Qiang Wang, Jin He, Bert Lynn, and Brian C Rymond. Interactions of the yeast sf3b splicing factor. *Mol Cell Biol*, 25(24) :10745–10754, 2005.
- [324] Qiang Wang, Li Zhang, Bert Lynn, and Brian C Rymond. A bbp-mud2p heterodimer mediates branch-point recognition and influences splicing substrate abundance in budding yeast. *Nucleic Acids Res*, 36(8) :2787–2798, 2008.
- [325] Zefeng Wang and Christopher B Burge. Splicing regulation : from a parts list of regulatory elements to an integrated splicing code. *RNA*, 14(5) :802–813, 2008.
- [326] K M Wassarman and G Storz. 6s rna regulates e. coli rna polymerase activity. *Cell*, 101(6) :613–623, 2000.
- [327] Karen M Wassarman and Ruth M Saecker. Synthesis-mediated release of a small rna inhibitor of rna polymerase. *Science*, 314(5805) :1601–1603, 2006.
- [328] M Watanabe, J Yanagisawa, H Kitagawa, K Takeyama, S Ogawa, Y Arao, M Suzawa, Y Kobayashi, T Yano, H Yoshikawa, Y Masuhiro, and S Kato. A subfamily of rna-binding dead-box proteins acts as an estrogen receptor alpha coactivator through the n-terminal activation domain (af-1) with an rna coactivator, sra. *EMBO J*, 20(6) :1341–1352, 2001.
- [329] P M Waterhouse, M B Wang, and T Lough. Gene silencing as an adaptive defence against viruses. *Nature*, 411(6839) :834–842, 2001.

- [330] E Wieczorek, M Brand, X Jacq, and L Tora. Function of taf(ii)-containing complex without tbp in transcription by rna polymerase ii. *Nature*, 393(6681) :187–191, May 1998.
- [331] Sivaramesh Wigneshweraraj, Daniel Bose, Patricia C Burrows, Nicolas Joly, Jorg Schumacher, Mathieu Rappas, Tillmann Pape, Xiaodong Zhang, Peter Stockley, Konstantin Severinov, and Martin Buck. Modus operandi of the bacterial rna polymerase containing the sigma54 promoter-specificity factor. *Mol Microbiol*, 68(3) :538–546, 2008.
- [332] C L Will and R Luhrmann. Spliceosomal usnrnp biogenesis, structure and function. *Curr Opin Cell Biol*, 13(3) :290–301, 2001.
- [333] Kevin J Wright, Michael T 2nd Marr, and Robert Tjian. Taf4 nucleates a core subcomplex of tfiid and mediates activated transcription from a tata-less promoter. *Proc Natl Acad Sci U S A*, 103(33) :12347–12352, Aug 2006.
- [334] Chwen-Huey Wu, Chanhyo Lee, Ruopeng Fan, Meghan J Smith, Yuki Yamaguchi, Hiroshi Handa, and David S Gilmour. Molecular characterization of drosophila nelf. *Nucleic Acids Res*, 33(4) :1269–1279, 2005.
- [335] S Wu, C M Romfo, T W Nilsen, and M R Green. Functional recognition of the 3' splice site ag by the splicing factor u2af35. *Nature*, 402(6763) :832–835, 1999.
- [336] J Wuarin and U Schibler. Physical isolation of nascent rna chains transcribed by rna polymerase ii : evidence for cotranscriptional splicing. *Mol Cell Biol*, 14(11) :7219–7225, 1994.
- [337] Xinshu Xiao, Zefeng Wang, Minyoung Jang, and Christopher B Burge. Coevolutionary networks of splicing cis-regulatory elements. *Proc Natl Acad Sci U S A*, 104(47) :18583–18588, 2007.
- [338] Yong-Zhen Xu, Catherine M Newnham, Sei Kameoka, Tao Huang, Maria M Konarska, and Charles C Query. Prp5 bridges u1 and u2 snrnps and enables stable u2 snrnp association with intron rna. *EMBO J*, 23(2) :376–385, 2004.
- [339] Yong-Zhen Xu and Charles C Query. Competition between the atpase prp5 and branch region-u2 snrna pairing modulates the fidelity of spliceosome assembly. *Mol Cell*, 28(5) :838–849, 2007.
- [340] Yu-Xin Xu, Yutaka Hirose, Xiao Zhen Zhou, Kun Ping Lu, and James L Manley. Pin1 modulates the structure and function of human rna polymerase ii. *Genes Dev*, 17(22) :2765–2776, 2003.
- [341] Yu-Xin Xu and James L Manley. Pin1 modulates rna polymerase ii activity during the transcription cycle. *Genes Dev*, 21(22) :2950–2962, 2007.
- [342] D Yan, R Perriman, H Igel, K J Howe, M Neville, and M Jr Ares. Cus2, a yeast homolog of human tat-sf1, rescues function of misfolded u2 through an unusual rna recognition motif. *Mol Cell Biol*, 18(9) :5000–5009, 1998.
- [343] Faqing Yuan, Laura Griffin, LauraJane Phelps, Volker Buschmann, Kenneth Weston, and Nancy L Greenbaum. Use of a novel forster resonance energy transfer method to identify locations of site-bound metal ions in the u2-u6 snrna complex. *Nucleic Acids Res*, 35(9) :2833–2845, 2007.
- [344] N Yudkovsky, J A Ranish, and S Hahn. A transcription reinitiation intermediate that is stabilized by activator. *Nature*, 408(6809) :225–229, 2000.
- [345] A Yuryev, M Patturajan, Y Litingtung, R V Joshi, C Gentile, M Gebara, and J L Corden. The c-terminal domain of the largest subunit of rna polymerase ii interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A*, 93(14) :6975–6980, 1996.
- [346] P D Zamore, J G Patton, and M R Green. Cloning and domain structure of the mammalian splicing factor u2af. *Nature*, 355(6361) :609–614, 1992.
- [347] C Zeng and S M Berget. Participation of the c-terminal domain of rna polymerase ii in exon definition during pre-mrna splicing. *Mol Cell Biol*, 20(21) :8290–8301, 2000.
- [348] G Zhang, K L Taneja, R H Singer, and M R Green. Localization of pre-mrna splicing in mammalian nuclei. *Nature*, 372(6508) :809–812, Dec 1994.
- [349] J Zhang and J L Corden. Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse rna polymerase ii largest subunit. *J Biol Chem*, 266(4) :2297–2302, 1991.

- [350] Zhiqiang Zhang, Jianhua Fu, and David S Gilmour. Ctd-dependent dismantling of the rna polymerase ii elongation complex by the pre-mrna 3'-end processing factor, pcf11. *Genes Dev*, 19(13) :1572–1580, 2005.
- [351] Xiansi Zhao, Jeffrey R Patton, Shannon L Davis, Brian Florence, Sarah J Ames, and Remco A Spanjaard. Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor rna activator. *Mol Cell*, 15(4) :549–558, 2004.
- [352] Xiansi Zhao, Jeffrey R Patton, Sajal K Ghosh, Nathan Fischel-Ghodsian, Ling Shen, and Remco A Spanjaard. Pus3p- and pus1p-dependent pseudouridylation of steroid receptor rna activator controls a functional switch that regulates nuclear receptor signaling. *Mol Endocrinol*, 21(3) :686–699, 2007.
- [353] Q Zhou and P A Sharp. Tat-sf1 : cofactor for stimulation of transcriptional elongation by hiv-1 tat. *Science*, 274(5287) :605–610, 1996.
- [354] Qiang Zhou and Jasper H N Yik. The yin and yang of p-tefb regulation : implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. *Microbiol Mol Biol Rev*, 70(3) :646–659, 2006.
- [355] Zhaolan Zhou, Lawrence J Licklider, Steven P Gygi, and Robin Reed. Comprehensive proteomic analysis of the human spliceosome. *Nature*, 419(6903) :182–185, 2002.
- [356] Jiang Zhu, Fuhong He, Songnian Hu, and Jun Yu. On the nature of human housekeeping genes. *Trends Genet*, 24(10) :481–484, 2008.
- [357] Y Zhuang and A M Weiner. A compensatory base change in u1 snrna suppresses a 5' splice site mutation. *Cell*, 46(6) :827–835, 1986.
- [358] G W Zieve. Two groups of small stable rnas. *Cell*, 25(2) :296–297, 1981.

SPLICING-INDEPENDENT RECRUITMENT of U1 snRNP on a TRANSCRIPTION UNIT

Béatrice Spiluttini,

Bo Gu, Praveen Belagal, Van Trung Nguyen,

Xavier Darzacq & Olivier Bensaude

² Ecole Normale Supérieure, Paris, France

³ CNRS UMR 8541 Régulation de l'Expression Génétique, Paris, France

ABSTRACT

Non-coding RNAs (ncRNAs) associate and regulate several components of the transcriptional machinery. We now report that U1 snRNA is co-immunoprecipitated with RNA polymerase II. To examine whether these ncRNA are recruited on genes along with the transcription machinery, we have established cell lines with transcription units consisting of multiple copies of a tetracycline-inducible reporter gene integrated at a single locus. The recruitment of ncRNA was examined by RNA-FISH. The major splicing snRNAs (U1, U2, U4, U5 and U6) were found to be recruited on a “wild-type” gene array containing a single intron in agreement with the idea of co-transcriptional splicing. We next generated cell lines with a reporter gene array that differed from the previous one by three point mutations that abolished splicing of the pre-mRNA. U2, U4, U5 and U6 were not recruited on these intron-less gene arrays as expected. In contrast, U1 snRNA recruitment was not affected by the lack of pre-mRNA splicing. Our finding is consistent with a constitutive association of U1 snRNA with RNA polymerase and suggests that U1 bound to the transcription machinery might scan the nascent transcript.

INTRODUCTION

During the past decade, numerous non-coding RNAs (ncRNAs) have been identified as regulators of transcription {Prasanth, 2007 #2153}. A few examples of ncRNAs have been shown to directly interact with and regulate components of the transcription machinery {Barrandon, 2008 #2235}. In bacteria, 6S RNA, one of the first ncRNA to be described, directly targets bacterial RNA polymerase (RNAP) {Wassarman, 2007 #2269; Wassarman, 2007 #2270}. The 6S RNA copurifies with sigma70-RNAP holoenzyme extracted from stationary phase cells. Association of 6S RNA inhibits transcription from several but not all sigma70-dependent promoters. It has been suggested that its interaction with RNAP contributes to alter the gene expression profile when bacteria enter the stationary phase. The

6S RNA binds into the active site of RNAP and inhibits transcription by blocking its binding to promoter DNA {Wassarman, 2006 #2165}. A similar situation might occur in stressed eukaryotic cells, when the mouse B2 {Allen, 2004 #2026; Espinoza, 2004 #2027} or the human Alu RNAs {Mariner, 2008 #2271} bind to and inhibit RNA polymerase II (RNAPII) initiation. B2 and Alu RNAs might contribute to the transcriptional shut-off occurring during stress. These RNAs might bind within the polymerase cleft surrounding the active site like the FC aptamer. This aptamer had been isolated in a pioneering study using yeast RNAPII as target molecule {Thomas, 1997 #2156}. Indeed, a central fragment of FC RNA, the FC* RNA, forms a double stem-loop inserted within the polymerase cleft that accommodates the DNA template {Kettenberger, 2006 #2158}. Thus, FC* RNA prevents the DNA template to enter the cleft and inhibits initiation of transcription. As described for B2 and Alu RNAs, elongation is not inhibited by FC* RNA as it cannot displace out the DNA template from the cleft.

A number of unrelated studies, within the line of splicing coupled to transcription, have suggested that RNAPII associates U snRNAs {Kim, 1997 #1264; Tian, 2001 #2297 ; Robert, 2002 #2276; Das, 2007 #2191}. Given the diversity of the published RNA associated to RNAPII, we decided to reinvestigate this question without prejudice. We first looked for RNAs that would co-purify with RNAPII from a mitotic cell lysate. Following these lines, U1 snRNA was identified as a major RNA co-purifying with RNAPII. An assay based on integrated gene arrays was set up next and indicated that U1 snRNP was efficiently recruited on an active transcription unit whether or not it had an intron. A constitutive U1 snRNP/RNAPII association might constitute a preloaded platform to initiate cotranscriptional splicing.

RESULTS

U1 snRNA is a major RNA associated with RNAPII

To look for RNAs that might associate and regulate RNAPII, we investigated RNAs that co-immunoprecipitate with RNAPII from a mitotic cell extract. The choice of a mitotic cell lysate was based on two statements : a) there is virtually no transcription during mitosis {Gottesfeld, 1997 #2136; Jiang, 2004 #2225}, hence nascent transcripts should not interfere; b) as the nuclear membrane breaks down during mitosis most of RNAPII was extracted in low salt non-denaturing buffers from mitotic cells (Figure 1A, lane 1) whereas at least half of

RNAPII remained in the nuclear pellet from an asynchronous cell extract (compare lanes 3 and 4). Knowing that RNA protein interactions are salt sensitive, it was not desirable to increase the ionic strength of the lysis buffer to improve RNAPII extractability. Thus, mitotic cell extracts were fractionated on a glycerol gradient and immunoprecipitation was carried out in fractions containing RNAPII. A major radioactive band migrating like a 160 +/- 10 base RNA was observed when RNAs retained by anti-Rpb1 were isolated and 3'-labeled with pCp and T4 RNA ligase (Figure 1B – lane 3). The size of this RNA prompted us to test for the presence of U1 snRNA (164 bases) by Northern blot. Indeed, when Rpb1, the largest RNAPII subunit, was immunoprecipitated from a mitotic cell extract using two different antibodies, a strong hybridization signal corresponding to U1 snRNA was detected (Figure 1C, lanes 3 and 6). Other RNAs present in the extract, such as 7SK snRNA or U2 snRNA remained within background levels in both control and anti-Rpb1 immunoprecipitates.

Reassortment between RNA and RNA-binding proteins has been shown to occur in cell extracts after cell disruption {Mili, 2004 #2145}. Therefore, Northern blot analysis was repeated on RNAs co-immunoprecipitated with Rpb1 from extracts of cells that had been cross-linked with formaldehyde prior to lysis in a dissociating buffer (Figure 1D). Again, U1 snRNA was markedly co-immunoprecipitated from mitotic cells (lane 3). U2 or 7SK snRNA were found in higher amounts than in mock antibody immunoprecipitates. ~~The extract contained 25 fold more U1 and around 3 fold more RNAPII molecules than in immunoprecipitates. HeLa cells are thought to contain about 10^6 U1 {Zieve, 1981 #2286} and $1.8 \cdot 10^5$ RNAPII molecules on average {Kimura, 1999 #2287}. Hence cells should contain about 6 fold more U1 than RNAPII molecules. Such proportions are compatible with a 1/1 U1/RNAPII stoichiometry.~~ Hence association of U1 snRNA with RNAP II molecules not engaged in transcription likely reflects a real in vivo interaction.

A model gene array to study the influence of splicing on a transcription unit.

The constitutive association of U1 snRNA with RNAPII suggested it might be recruited on a gene independently of splicing. To test this possibility, we turned to gene arrays using microscopy techniques {Misteli, 1999 #2288}. Gene arrays were made using either plasmid pTet-on-CFP-MS2 WT that harbored a tetracycline-inducible promoter driving the synthesis of a transcript with a single intron placed between a translation start codon and the CFP cDNA or plasmid pTet-on-CFP-MS2 MUT that differed from pTet-on-CFP-MS2 WT by mutagenesis of three nucleotides in the splicing consensus sequences (Figure 2A). WTA9B41

cells having integrated multiple copies of pTet-on-CFP-MS2 WT at a single locus (gene array) were isolated and cloned. Following exposure to doxycyclin, these cells expressed CFP proteins localized in peroxisomes (Figure 2B). The doxycyclin-induced RNAs were analyzed by FISH using probes labeled with different stains. Exon 1 (E1) and MS2 probes stained more the cytoplasm than the nuclei of WTA9B41 cells. Yet, a bright spot (arrow) was visible in the nucleoplasm of most cells. This bright spot was also stained with the intron probe (I1). However, the intron probe neither stained the rest of the nucleoplasm, nor the cytoplasm. This observation was an indication of co-transcriptional splicing and established that the nuclear bright spots corresponded to transcribed gene arrays. MUT2.5A cells having integrated multiple copies of the pTet-on-CFP-MS2 MUT at a single locus were next isolated and cloned. Again, exon 1 (E1) and MS2 probes stained more the cytoplasm than the nuclei and a bright spot was visible in the nucleoplasm of most cells (Figure 2B). This bright spot was also stained by the intron probe (I1). But in contrast to the WTA9B41 cells, the intron probe stained the cytoplasm of the MUT2.5A cells. Such cytoplasmic staining was indicative of a splicing deficiency likely due to the splicing impairment. In both pTet-on-CFP-MS2 plasmids, one nucleotide had been deleted in the β -globin first intron sequence to suppress intronic stop codons and keep the translation frame. Thus, MUT2.5A cells also expressed CFP proteins localized in peroxisomes (Figure 2B). To evaluate splicing efficiency, the CFP proteins were characterized by Western blot. Indeed, the protein made by the T2.5A cell line was larger than the one made by WTA9B41 cells as expected from intron inclusion (Figure 2C). Importantly, the Western blot suggested a very sharp all or none splicing phenotype that should facilitate interpretation of the data.

Radial analysis of image intensities to investigate recruitment of snRNAs on gene arrays

To monitor factor recruitment on gene arrays several oligonucleotide and antibody probes labeled with distinct colors were hybridized to doxycyclin-induced WTA9B41 cells. MS2 probes were generally used to localize the transcription sites. Colocalization of U1 or U2 snRNA with the transcription sites was hardly convincing by eye (Figure 3). Data averaging was expected to provide more significant observations. To this aim, U1 FISH images were obtained for a collection of cells and centered on the transcription site (Transcription Site – TS) determined using the MS2 hybridization image and summed up. The averaged U1 or U2 transcription site images also appeared as a bright spots but broader than the MS2 ones (Figure 3 – bottom line). In contrast, averaging images taken at random within the nuclei of

the same cells (Random Site – RS) provided a rather homogeneous image with a smooth decrease in light intensity with increasing distance from the image center. This decrease was most likely due to the contribution of nucleoli and cytoplasm that are not stained and have an increased probability to be present and contribute with increasing distance from the center of the random site (RS). A quantitative procedure was developed next (Figure 4A). The light densities provided by the U1 FISH were determined as a function of distance from the transcription site center : rings were centered on the transcription site and the light density at a given distance was defined as the ratio of the sum of pixel intensities in an annulus between two consecutive rings to the area of the same annulus. To define a background reference, the same analysis was repeated on a site taken at random in the nucleoplasm (RS). There was a trend to observe a higher increase in U1 density in the center of the transcription sites (Figure 4B – closed symbols) than in the center of the random sites (open symbols). As the difference between transcription and random sites was hardly convincing, an averaging procedure was developed. A normalization was performed to give the same statistical weight to each cell: light densities at a distance of $1.6 \mu\text{m}$ from TS or RS were set to 1 for all cells (Figure 4C). The last step consisted in summation and averaging the normalized data obtained for a collection of cells. Such radial analysis was applied first to WTA9B41 cells (Figure 4D - circles). The averaged signal density curve at the random sites (RS) adopted a dome shape with a monotonic decrease of densities with increasing distance from the center. Densities at the center were $5 \pm 3 \%$ higher than densities at $1.6 \mu\text{M}$ from the center likely because partial overlap with unstained nucleoli or cytoplasm. The averaged U1 signal densities at transcription sites (TS) were clearly above background (RS) of the same cell set. Densities at the center were $25 \pm 5 \%$ higher than those of the random site average thereby indicating an enrichment of U1 snRNA on transcription sites. The shape of the curve was very similar to the shape of the averaged MS2 signal density (Figure 4E). To establish that U1 snRNA enrichment was not due to some particularity of plasmid integration locus in the genome, the same analysis was repeated with WTP2 cells that had integrated the same plasmid and had been selected and subcloned following an independent transfection. Profiles obtained using images taken with WTP2 cells were very similar to those obtained with WTA9B41 cells (Figure 4D - triangles). This finding indicated that U1 snRNA enrichment was not influenced by the plasmid insertion locus and illustrated the robustness of the method.

Recruitment of U1 snRNP on gene arrays is splicing-independent

The gene arrays in WTA9B41 cells used to establish the radial analysis procedure had an intron that was spliced cotranscriptionally. Hence, enrichment of U1 snRNA at the transcription site might be attributed to cotranscriptional recruitment of the spliceosome. In contrast, although the transcription site signals (TS) obtained with U2, U4 and U6 snRNA probes were clearly above background (in the 25% range) in WTA9B41 cells having a functional intron

The radial analysis was therefore repeated with MUT2.5A cells that contained gene arrays with disrupted splice sites. Once again, U1 enrichment at transcription site (20 ± 5 %) was clearly above background (Figure 5A). In contrast, although the transcription site signals (TS) obtained with U2, U4 and U6 snRNA probes were clearly above background (in the 25% range) in WTA9B41 cells having a functional intron, it was within background range (RS) in splicing deficient MUT2.5A cells hybridized with the same probes (Figure 5B, C and D). However, the density curves observed using these probes on transcription sites adopt a S shape remarkably different from the monotonic dome shape of random site shapes. U4, U5, U6 snRNAs are loaded on introns as a trimer, the tri snRNP {Will, 2001 #2290}. Therefore, hybridization was repeated using simultaneously the three probes altogether labelled with the same colour (Figure 4E). The tri snRNP was recruited on WTA9B41 cells but not on MUT2.5A cells transcription sites. Using the three probes improved the signal to noise ratio and indicated that U4, U5, U6 density was on average 10 ± 4 % lower at transcription sites than at random sites. When data obtained independently from U4 and U6 images were averaged altogether, it became obvious that U snRNA density was also lower at the transcription site than at the random site (Supplemental Figure X). The S shape of the density curve observed with splicing deficient (MUT2.5A cells) transcription sites was remarkable and reproducibly observed when averaging distinct set of cell images. On a methodological point of view, this observation demonstrates the reliability of the radial analysis as U2, U4, U5, U6 probes provided a result opposite to the U1 probes although MUT2.5A cells provided very similar FISH images. Moreover, the S shape of the density curves below the random site densities indicates a lower concentration of these RNA's at the transcription site. It is concluded that recruitment on splicing-deficient gene arrays is a characteristic of U1 snRNA.

Recruitment of U1 snRNP on transcription sites.

As constitutive RNAPII/U1 snRNP association was the working hypothesis, immunofluorescence of RNAPII was questioned on the same gene arrays. The radial analysis

method could be applied indistinctly to FISH or immunofluorescence images. Some antibodies were incompatible with FISH procedures. In this case, the transcription sites were determined in cells transfected with MS2-GFP protein (Supplemental figure XX). As a control U1 FISH images were analyzed following such procedure and provided strictly identical results (Supplemental figure XX). Recruitment of RNAPII was 35% above background in gene arrays of intron-containing WTA9B41 cells and weaker in those of intron-deficient MUT2.5A ones (Figure 6A). U1 snRNA is associated with several companion proteins such as U1-A and U1-70K that form the U1 snRNP {Will, 2001 #2290 ; Hochleitner, 2005 #2289}. U1-70K recruitment on intron-deficient arrays was almost as efficient as on intron-containing arrays (Figure 6B). In contrast, although U1A was still recruited on intron-deficient arrays, its recruitment was largely less efficient than on intron-containing arrays (Figure 6C). Collectively, the data were consistent with RNAPII/U1 snRNP co-recruitment on transcription sites.

Role of the CTD in U1 snRNP recruitment?

The carboxyl-terminal domain (CTD) of RNAPII largest subunit plays a critical role in cotranscriptional RNA processing and surveillance {de Almeida, 2008 #2275}. RNAP II or recombinant CTD associate pre-mRNA and stimulate splicing in vitro and in vivo {Hirose, 1999 #1613; Zeng, 2000 #2272}.

PRESENT DATA UNCONCLUSIVE

Splicing-dependent recruitment of SR proteins.

SR proteins are a highly conserved family of essential splicing factors that share a bipartite structure composed of an N-terminal RNA binding domain and a C-terminal arginine-serine-rich domain {Sanford, 2005 #2309}. Recruitment of the U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor of 65 kDa, U2AF65 was investigated because it binds the intronic polypyrimidine tract sequence and can promote U1 snRNP recruitment to 5'-splice sites {Hoffman, 1992 #2295; Förch, 2003 #2274}. Although the polypyrimidine tract sequence had not been modified in our intron-lacking construct, the recruitment of U2AF65 was found to be intron-dependent (Figure 7A). SR proteins such as the alternative splicing factor/splicing factor 2 (ASF/SF2) and SC35 interact with the 70k protein of the U1 small nuclear ribonucleoprotein {Cao, 1998 #2292; Golovkin, 1999 #2291; Ellis, 2008 #2273}. ASF/SF2 was found to be recruited on both intron-containing and intron-lacking gene arrays

as well (Figure 7B). In contrast, but in agreement with a previous report {Huang, 1996 #2279}, SC35 was only recruited on intron-containing gene arrays, the SC35 density was lower than average at intron-lacking transcription sites (Figure 7C). To summarize, recruitment of some but not all SR proteins, was intron-independent.

DISCUSSION

Using mitotic cell extracts, the present study shows a highly significant U1 snRNA association with RNAPII not engaged in transcription. In contrast, association of the other splicing U snRNAs was not significantly detected. In agreement with our findings, a recent proteomic study reported U1 as the sole U snRNA to be co-immunoprecipitated with RNAPII from a nuclear extract of interphasic cells {Das, 2007 #2191}. Using an *in vivo* RNA/protein cross-linking strategy, we now observe a specific and highly significant U1 snRNA/RNAPII association. This finding eliminates a possible reassortment between RNA and RNA-binding proteins as previously reported to occur in cell extracts {Mili, 2004 #2145}. It establishes that U1 snRNA and non-transcribing RNAPII are constitutively associated within living cells. The U1/RNAPII stoichiometry was estimated to be in the 1/1 range thereby strongly supporting the *in vivo* significance of the U1 snRNA/RNAPII interaction.

Radial analysis of microscopic images, a sensitive assay to study colocalizations

Assuming that U1 snRNA/RNAPII association was constitutive implied that U1 snRNA was expected to be recruited on genes even when lacking introns. To verify this prediction, an assay based on FISH using gene arrays was set up. Just looking at colocalization of U snRNAs with transcription sites was not fully convincing in part because of the punctuated distribution of U snRNAs. Therefore, a quantitative analysis of light intensities was developed. Fluorescence microscopic images obtained with a number of cells and centered on the transcription site were averaged. The evolution of averaged light densities as a function of distance to the transcription center was compared to the evolution of light densities in averaged images chosen at random in the nucleoplasm of the same cells. This radial analysis method proved to be robust as different clonal cell lines provided identical results. It was highly reliable and sensitive as distinct probes provided distinct results although images were indistinguishable. U1 snRNA was recruited on both intron-containing and intron-lacking gene arrays whereas U2, U4, U5 and U6 snRNA are only recruited on intron-containing gene arrays. More, in the vicinity of intron-lacking transcription sites, the density of these RNAs is

lower than in the average nucleoplasm. The radial analysis was as efficient to process data from immunofluorescence. Some proteins (Rpb1, U1-70K, ASF/SF2) were found to be recruited on both intron-containing and intron-lacking gene arrays and others were only recruited on intron-containing gene arrays (SC35, U2AF65). The fact that different probes give different results strongly supports the reliability of the radial analysis method. The radial analysis reveals the inhomogeneity of the nucleoplasm, some factors are recruited while others can be excluded from the vicinity of gene arrays.

U1 snRNP recruitment, the first step in cotranscriptional splicing?

U1 snRNA and at least two U1-associated proteins were recruited on an intron-deficient gene array as efficiently as on its intron-containing counterpart. Such findings are consistent with previous reports that U1 snRNP is recruited on intron-less *Drosophila* heat-shock puffs {Martin, 1987 #2277 ; Matunis, 1993 #2278}. Taking advantage that in vitro synthesized snRNA injected into amphibian oocytes assemble into functional snRNPs {Hamm, 1987 #2308}, an assay based on injection of fluorescently labeled snRNA into amphibian oocyte showed that U1 snRNAs target nascent transcripts independently of splicing {Patel, 2007 #2307}. A central argument in that study was that a U1 snRNA lacking the 5' splice site recognition sequence still targeted nascent transcripts. In apparent discrepancy with our findings, U2 snRNA lacking the branch point recognition sequence was also targeted to nascent transcripts {Patel, 2007 #2307}. However, disrupting intron recognition elements in U2 snRNA may not prevent U2 snRNP association with U1 snRNP and other factors involved in U2 snRNP recruitment. Splicing-independent recruitment of U1 snRNP on transcription units might be understood in the broader context of coupling between transcription and post-transcriptional maturation of transcripts {Hirose, 2000 #1731; Proudfoot, 2002 #1867; Bentley, 2005 #2042 ; de Almeida, 2008 #2275}. FISH using an intron probe only stains transcription sites with the wild-type version of our gene arrays. This finding demonstrates this reporter gene undergoes co-transcriptional splicing. One might expect that U1 snRNP riding RNAPII improves the efficiency of 5' intron recognition through nascent transcript scanning.

U1 snRNP a structural component of the transcription/post-transcription factories

Association between U1 snRNA and the cyclin subunit of the TFIIF general transcription factor has been reported {Kwek, 2002 #1884; O'Gorman, 2005 #2030}. This association

might contribute in splicing-dependent stimulation of transcription. It is remarkable that sequences corresponding to the U1 5' splice recognition sequence were found in four out of eight RNA aptamers selected for their capacity to inhibit transcription using a HeLa nuclear extract {Tian, 2001 #2297}. Addition of synthetic U1 snRNA stimulated the formation of the first phosphodiester bond in an assay using only purified factors {Kwek, 2002 #1884}. Furthermore, U1 snRNA stimulated reinitiation of transcription in this assay when a 5' splice site was present close the initiation site. As purified RNAPII lacks U1 snRNP because of the harsh salt conditions used throughout its purification, one might conceive that an interaction of U1 snRNA with RNAPII and TFIIF among multiple other components favors the assembly of an active transcription initiation complex in vitro. Indeed, the nuclear cap-binding complex has been reported to facilitate association of U1 snRNP with the cap-proximal 5' splice site {Lewis, 1996 #2304}. U1-70K binds the 25-kD cleavage factor I subunit, CF-I m, binds to U1-70K {Awasthi, 2003 #2186}. The U1-A protein interacts with the 160-kD subunit of the cleavage-polyadenylation specificity factor and increases polyadenylation efficiency in vitro {Lutz, 1996 #2305}. Gene arrays are good candidates to promote the formation of such factories. U1 snRNP might play a structural role in assembling the transcription/post-transcription machinery in vitro. This role might be generalized in vivo to the formation of transcription factories {Carter, 2008 #2310}.

METHODS

Plasmids. Plasmid pTet-on CFP-MS2 wt was generated by sequentially inserting PCR fragments into pTet-globin-CFP-18MS2-2 {Darzacq, 2006 #2242}. A single nucleotide was deleted in the β -globin intron 1 sequence to maintain the translational frame in the absence of splicing and the 18 MS2 repeats were brought up to 24. pTet-on CFP-MS2 mut was derived from pTet-on CFP-MS2 wt by targeted point mutagenesis. Both plasmids comprised a puromycin resistance gene. Plasmids driving the expression RNAPII largest subunit Rpb1wt-EGFP or Rpb1 Δ 5-EGFP were generously provided by Dirk Eick {Chapman, 2004 #2300}. pMS2-GFP {Darzacq, 2006 #2242} ???, *pSC35-GFP?*,

Cells. U2OS Tet-on cells (CLONTECH) and HeLa cells were propagated respectively in low and high glucose DMEM (INVITROGEN) supplemented with 10 % foetal calf serum. HeLa cells arrested in mitosis after exposure to nocodazole (50 ng ml⁻¹) for 18 h were detached by shaking and pelleted by centrifugation at 200 g. To obtain gene arrays, U2OS cells were co-transfected with pBABE-puro (Addgene) and pTet-on CFP-MS2 wt or pTet-on CFP-MS2

mut by standard calcium phosphate procedure. Individual colonies were selected in culture medium containing puromycin (10 µg/ml), screened for CFP expression after doxycycline induction and subcloned. FuGene6 (Roche) was used for transient transfection assays. CFP transcription was induced using 5 µg/ml of doxycycline overnight.

Northern blot assays and protein-RNA cross-linking. RNAs were electrophoresed in polyacrylamide gels, transferred onto Hybond N (Amersham), and hybridized to nick end-translated ³²P-labeled DNA probes. The membranes were washed in 0.1X SSC buffer (0.015 M NaCl, 0.015 M sodium citrate) at 65°C. Radioactivity was quantified with a Fuji FLA 3000 phosphorimager using the Image Gauge software. For protein-RNA cross-linking, formaldehyde (1% final concentration) was added to the culture medium for 10 min and then neutralized with addition of glycine (125 mM, pH 7) for 5 min. Cells were washed in PBS, lysed in RIPA buffer (50 mM Tris [pH 7.9], 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl) and sonicated. The lysate was clarified by centrifugation at 10,000g. Protein-RNA complexes were immunoprecipitated from supernatants and the cross-link was reversed by incubation for 2 h at 65°C in a mixture of 50 mM Tris [pH 7.9], 5 mM EDTA, 10 mM dithiothreitol, and 1% SDS. RNAs were isolated with the RNeasy micro kit (QIAGEN).

Antibodies. 8WG16 was provided by Nancy E. Thompson {Thompson, 1989 #658}, Pol3/3 provided by Ekkehard K. Bautz {Krämer, 1980 #522}, N20 (sc-899, Santa Cruz Biotechnology), CTD4H8 (Upstate 05-623), U1-70K provided by Douglas Black (Irvine, CA), U1-A (AbCAM ab55751), U2AF65 (XXXX), ASF/SF2 (XXXX), SC35 (XXXX), prp40 (sc-68081, Santa Cruz Biotechnology), PSF/SFPQ (ab 11825, AbCAM), SF1/BBP (ab 58077, Abcam), p54nrb (ab 45359, AbCAM)

Primary antibodies were detected with horseradish peroxidase-labeled (Promega), Cy3- or Cy5-conjugated (GE Healthcare) secondary antibodies.

Fluorescence In Situ Hybridization (FISH) and immunofluorescence.

Immunofluorescence and FISH was performed as described previously on paraformaldehyde-fixed cells {Darzacq, 2006 #2242}. Oligodeoxynucleotides (supplementary Table 1) containing 5-amino-allyl thymidines (Eurogentec) were labeled with Cy3 or Cy5 mono-reactive dyes (GE Healthcare). In some cases, an immunostaining was performed after FISH using the standard protocol.

Supplementary Table 1 – Sequences of oligodeoxynucleotides used for FISH. 5-amino-allyl thymidines are noted “t”s in the sequences.

```

U1-1      CtG GGA AAA CCA Cct TCG TGA TCA tGG TAT CTC CCC tGC CAG GTA AGt AT
U1-2      CGA ACG CAG tCC CCC Act ACC ACA AAT TAt GCA GTC GAG tTT CCC ACA tT
U2        AGt GGA CGG AGC AAG CtC CTA TTC Cat CTC CCT Gct CCA AAA ATC CAT tT
U4-1      tAG CAA TAA tCG CGC CTC GGA tAA ACC TCA TtG GCT ACG ATA CtG CCA CT
U4-2      AGA CtG TCA AAA ATt GCC AAT GCC GAC tAT ATT GCA AGt CGT CAC GGC Ggt A
U5-1      GGC AAG Gct CAA AAA ATt GGG TTA AGA CtC AGA GTT GTt CCT CTC CAC Ggt A
U6        CGt GTC ATC CtT GCG CAG GGG CCA tGC TAA TCT Tct CTG TAT CGT tCC AA
beta-glob-E1  AGG AGt CAG GTG CAC CAt GGT GTC TGT TtG AGG TTG CTA GtG AAC ACA GtA
beta-glob-II  tCA GTG CCT AtC AGA AAC CCA AGA GtC TTC TCT GTC tCC ACA TGC CCA GtA
MS2 NBX   CtA GGC AAT tAG GTA CCT tAG GAT CTA AtG AAC CCG GGA ATA CtG CAG AC

```

Radial analysis of site images. Imaging was performed at room temperature using a 60X/1.4 NA planapo ??? objective (???) on an inverted microscope (Olympus 1X81) equipped with a z-piezo (???) and automatic excitation/emission filter wheels connected to an air-cooled charge-coupled ??? digital camera (???) run by Metamorph (???) software. Images were taken sequentially in planes separated by 200 nm and at different wavelengths corresponding to specific probes. Image processing used the Image-J software. The MS2 FISH or MS2-GFP fluorescence images were used to accurately define the transcription site. Briefly, all MS2 planes were stacked. An ellipse fitting the transcription site was drawn manually on this stack and defined a column of planes. The brightest pixel in this column was determined and used to assist in defining manually the transcription site focal plane. The image in this plane was stacked with the images of the planes immediately above and below. The barycenter of the MS2 FISH or MS2-GFP intensities defined in this stack of three planes was taken as the transcription site center. The signal density obtained with the various probes was defined as the sum of pixel intensities in an annulus between rings separated by 1 pixel (108,3 nm) divided by the area of the same annulus and determined as a function of distance from transcription center. The random site was selected automatically at random within a circle of 15 pixels positioned manually in the three-plane stack avoiding as much as possible nucleoli and cytoplasm.

ACKNOWLEDGEMENTS

Work in the authors' laboratory has been supported by grants from the Association pour la Recherche sur le Cancer (ARC 3692), Agence Nationale pour la Recherche (06-BLAN-0072 and JC-XXXXX).

REFERENCES

FIGURE LEGENDS

Figure 1 : **U1 snRNA is selectively associated to RNAPII in mitotic cells.** (A) HeLa cells lysed in a low salt buffer (0.15 M NaCl) were fractionated by 10,000 g centrifugation providing supernatants (S) (lanes 1 and 3) and nuclear pellets (P) (lanes 2 and 4). Lanes 1 and

2 correspond to asynchronous cells whereas lanes 3 and 4 correspond to mitotic cells. (B) RNAs were immunoprecipitated from a low salt supernatant with mock or N20 anti-RNAPII antibody and 3'-end labeled with radioactive pCp and T4 RNA ligase. (C) RNAs were immunoprecipitated from a low salt supernatant of mitotic cells (In) with mock (lanes 2, 5), N20 (lane 3) or 8WG16 (lane 6) anti-Rpb1 antibody and probed for U1, U2 or 7SK snRNA by Northern blot. Immunoprecipitates were probed for Rpb1 by Western blot with Pol3/3 antibody. (D) RNAs immunoprecipitated from a cross-linked mitotic cell extract (lanes 1-3) with mock or 8WG16 anti-RNAPII antibodies were probed by Northern blot.

Figure 2 : Splicing-dependent gene arrays. (A) pTet-on CFP-MS2 plasmid schemes. For pTet-on CFP-MS2 wt, a tetracyclin-inducible promoter (Tet-on PR) drives the synthesis of a mRNA carrying an AUG initiation codon, the β -globin exon 1, intron 1 and the beginning of exon 2 fused to the CFP protein coding sequence terminated with the peroxisome targeting peptide Ser-Lys-Leu (SKL) at its COOH terminus. The translation stop codon (*) is followed by 24 repeats of the MS2 RNA hairpin and a polyadenylation signal (pA). A single nucleotide was deleted (Δ) in the β -globin intron 1 sequence to maintain the CFP translation frame in case of splicing deficiency. pTet-on CFP-MS2 mut plasmid was made splicing-deficient by GT>AC replacement of the 5' splice site and G>C replacement of the 3' splice site. Positions of labeled oligonucleotide probes used for FISH in the β -globin first exon (E1) and first intron (I1) and for MS2 sequences are indicated. (B) Characterization of splicing-dependent gene arrays. WTA9B41 cells have integrated pTet-on CFP-MS2(wt) plasmid. MUT2.5A cells have integrated the intron-mutated pTet-on CFP-MS2(mut) plasmid. Both cells expressed CFP localized in peroxisomes, stained with DAPI, were hybridized with fluorescent DNA probes of different colors for exon 1 (E1), intron 1 (I1) and MS2 sequences. Transcription sites are indicated by arrows. (C) Western blot analysis of CFP proteins made by WTA9B41 and MUT2.5A cells induced with doxycycline.

Figure 3. snRNA distribution around a transcription site. (A) A WTA9B41 cell was hybridized with fluorescent DNA probes of different colors for U1, U2 snRNA or MS2 sequences. The transcription site (TS) and a site selected at random (RS) in the nuclei were framed by squares. (B) Enlarged pictures (3.25 μ m x 3.25 μ m squares) of the transcription (TS) and random (RS) sites in the nuclei of 5 different cells. The bottom line shows images resulting from the summation (Σ) of images taken in **50** different cells.

Figure 4. Radial analysis of snRNA distribution around a transcription site. (A) Cells were hybridized with probes labeled with different colors. Rings separated by 1 pixel (108,3 nm) were drawn, centered on the transcription sites (TS) determined with a MS2 probe image or centered on random sites (RS); (B) Averaged intensities in annuli between rings were determined for individual cells and plotted as a function of distance to the center for 4 different cells (triangles, diamonds, circles and squares); (C) Densities were determined setting a density of 1 at 1.6 μm ; (D) Densities were averaged for 92 WTA9B41 (circles) or 95 WTP2 (triangles) cells that both contain a gene array of the same pTet-on CFP-MS2(wt) plasmid but have originated from distinct transfections and cloning; (E) Averaged MS2 intensities for the same cell sets. Closed symbols correspond to densities around transcription sites (TS), whereas open symbols correspond to densities around random sites (RS). Error bars represent standard errors.

Figure 5. Splicing-independent recruitment of U1 snRNA on transcription sites. Cells were hybridized with MS2 probes to determine transcription sites (TS). FISH images were processed by radial analysis of light density using WTA9B41 cells (circles) and MUT2.5A cells (triangles) hybridized with probes for (A) U1 (146 and 118 cells), (B) U2 (138 and 126 cells), (C) U4 (136 and 157 cells), (D) U6, and (E) for altogether U4, U5 and U6 snRNAs (tri-snRNP) (167 and 127 cells). Closed symbols correspond to densities around transcription sites (TS), whereas open symbols correspond to densities around random sites (RS).

Figure 6. Splicing-independent recruitment of U1 snRNP on transcription sites. Cells were hybridized with MS2 probes to determine transcription sites. Light densities of immunofluorescence images obtained with antibodies against RNAPII largest subunit (Rpb1) or U1 snRNPs were processed by radial analysis. (A) U1-70K using 51 WTA9B41 cells and 121 MUT2.5A cells; (B) U1A using 374 WTA9B41 cells and 274 MUT2.5A cells ; (C) Rpb1 (CTD4H8 antibody) using 95 WTA9B41 cells (circles) and 116 MUT2.5A cells (triangles). Closed symbols correspond to densities around transcription sites (TS), whereas open symbols correspond to densities around random sites (RS).

Figure 7. Splicing-dependent recruitment of SR proteins on transcription sites. Cells were hybridized with MS2 probes to determine transcription sites. Light densities of immunofluorescence images obtained with antibodies against SR proteins were processed by

radial analysis. (A) SC35 using 111 WTA9B41 cells (circles) and 88 MUT2.5A cells (triangles); (B) U2AF65 using 195 WTA9B41 cells (circles) and 217 MUT2.5A cells (triangles); (C) ASF/SF2 using 109 WTA9B41 cells (circles) and 110 MUT2.5A cells (triangles). Closed symbols correspond to densities around transcription sites (TS), whereas open symbols correspond to densities around random sites (RS).

*human ASF/SF2 siRNA : ACGAUUGCCGCAUCUACGU : Goina E, Skoko N, Pagani F. Mol Cell Biol. 2008 Jun;28(11):3850-60.
for SF2/ASF knockdown, siSF2/ASF, 5'-CCAAGGACAUUGAGGACGUdT-3'
de La Mata and Kornblihtt NSMB (2006)*

These findings contradict previous reports that major U snRNAs (U1, U2, U4, U5 and U6) are co-immunoprecipitated with RNAPII from both mitotic and interphasic cell extracts {Kim, 1997 #1264} and that a RNAPII complex associated with several U snRNAs including U1, U2 and U4 has been affinity-purified from a whole cell extract with the binding of RNAPII on a transcription elongation factor SII (TFIIS) column {Robert, 2002 #2276}. The use of different buffers and procedures might explain discrepancies with other studies.

In contrast, chromatin immunoprecipitation did not detect U1-70K on the intron-less human heat-shock genes activated with sodium arsenite {Listerman, 2006 #2284}. Satellite III transcripts also immunoprecipitate together with small nuclear ribonucleoproteins (snRNPs) in vivo whereas the intronless hsp70 transcripts do not {Metz, 2004 #2285}. The discrepancy between studies might be due to the different methodologies. Splicing is impaired in heat-shocked cells {Yost, 1986 #285}. SRp38 is dephosphorylated in response to heat shock in mammalian cells and interaction of dephosphorylated SRp38 with U1 snRNP is associated with repression of the splicing machinery {Shin, 2004 #2303}. Furthermore, heat-shock leads to dramatic alterations of snRNPs in human cells {Bond, 1988 #2302}.

Recruitment of SR-proteins

Consistent with this result, several previous studies had reported that a subset of the SR proteins show preferential binding to the CTD [38–41] {Mortillaro, 1996 #1219 ; Kim, 1997 #1264}. It is therefore conceivable that the U1 snRNP and SR proteins pre-loaded onto the CTD {Das, 2007 #2191} bind to the nascent transcript as the 5'ss and ESE sequences emerge from the polymerase (Fig. 4). Assuming that SR proteins bridge their exonic binding sites to the CTD, SR proteins may provide the molecular ties that tether adjoining exons to RNA pol

II as recently reported by the Proudfoot laboratory [34]. Consistent with this view, binding of SR proteins to ESEs was previously shown to play a critical role in ensuring the correct 5' to 3' linear order of exon joining during splicing [43], and the CTD was found required to mediate the effect of an SR protein on alternative exon inclusion [44]. {de Almeida, 2008 #2275}

purified U1 snRNP and ASF/SF2 form a ternary complex with pre-mRNA, which is dependent on a functional 5' splice site {Jamison, 1995 #2293}.

U1 snRNP interacts specifically with both the 5' and 3' splice sites in the presence of ATP and that SC35 is required for these ATP-dependent interactions. Significantly, the SC35-dependent interaction between U1 snRNP and the 3' splice site requires U2 snRNP but not the 5' splice site {Fu, 1992 #2294}.

Prp40 associates with the U1 snRNA {Kao, 1995 #2282; Neubauer, 1997 #2283}. Prp40 binds to in phospho-CTD in vitro {Morris, 2000 #2281}.

U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. Mol Cell. 1998 Jan;1(2):255-64.

A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo {Damgaard, 2008 #2280}.

SRPK1 is predominantly associated with U1 small nuclear ribonucleoprotein (snRNP) Mathew et al. 2008

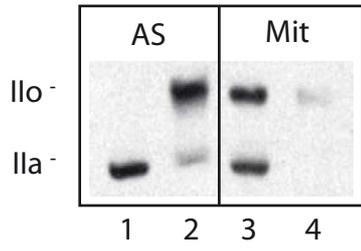
PSF and p54nrb/NonO copurified with a holoenzyme form of pol II containing hypophosphorylated CTD and could be coimmunoprecipitated with antibodies specific for this and the hyperphosphorylated form of pol II {Emili, 2002 #2299}.

p54nrb, a transcription/splicing factor, associates in vitro with the 5' splice site (SS) within large complexes present in HeLa cell nuclear extracts, in which the hyperphosphorylated form of RNA polymerase II (RNAPII) is associated with U1 or U1 and U2 snRNPs. {Kameoka,

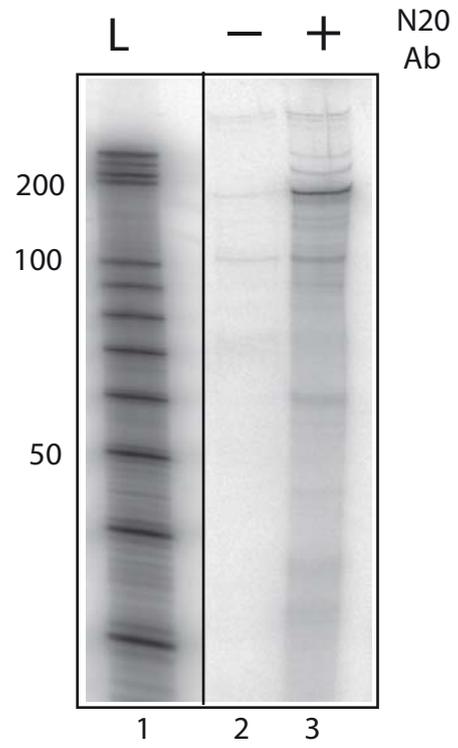
2004 #1972}

The SMN complex is coimmunoprecipitated with RNAPII and GST-CTD interacts with in vitro translated SMN {Pellizzoni, 2001 #2076}.

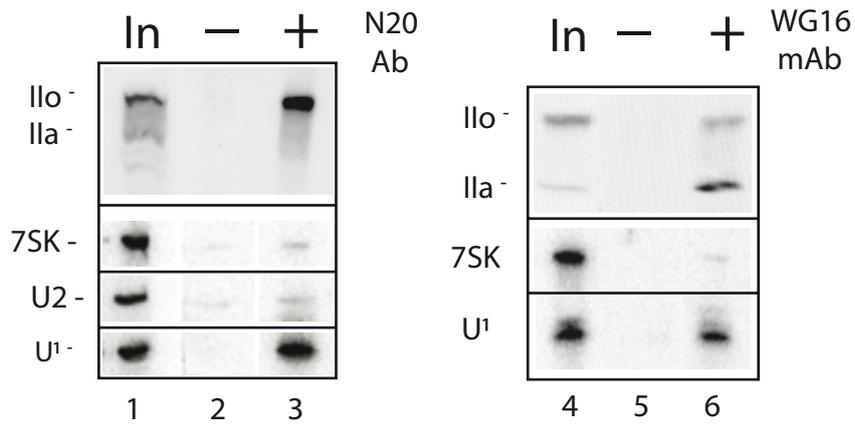
A



B



C



D

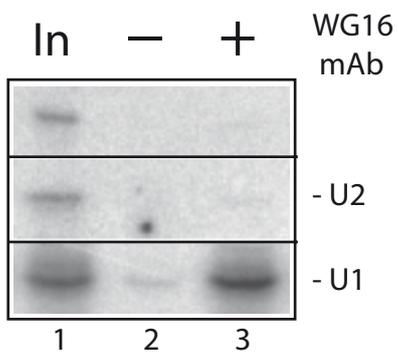
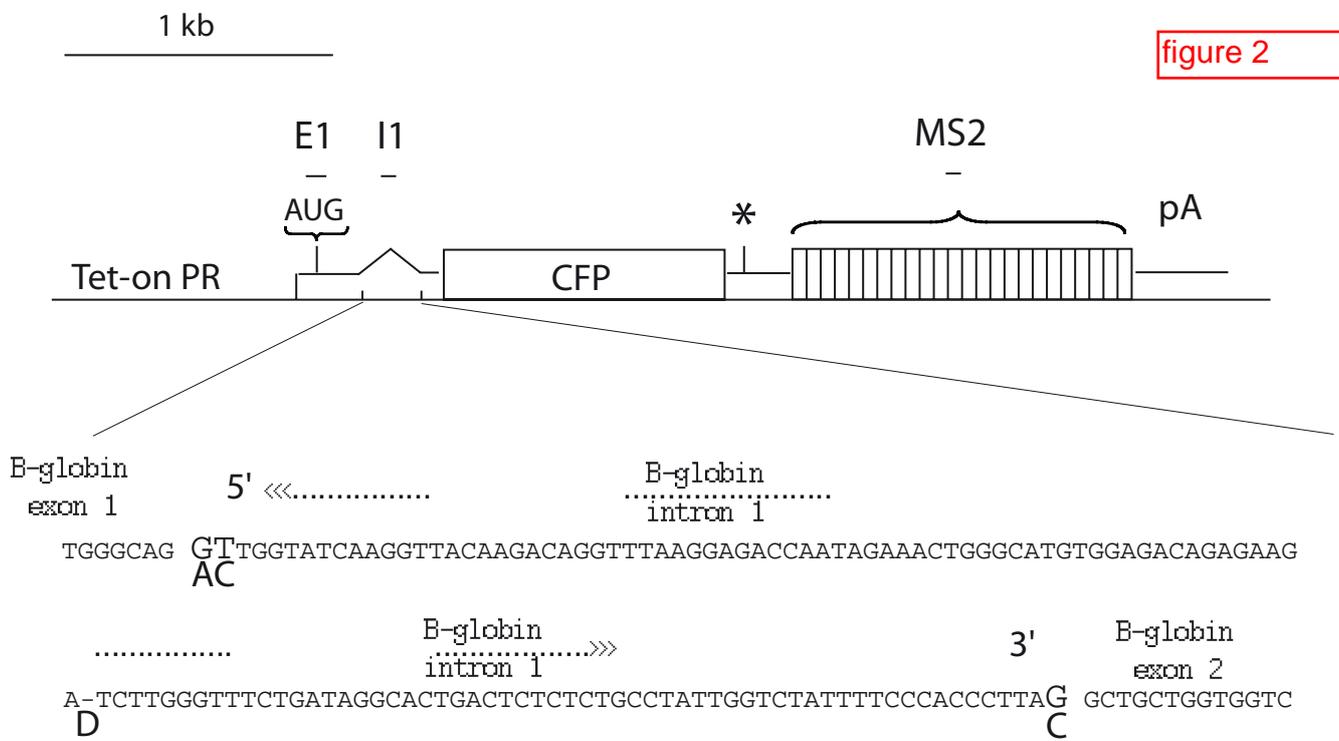
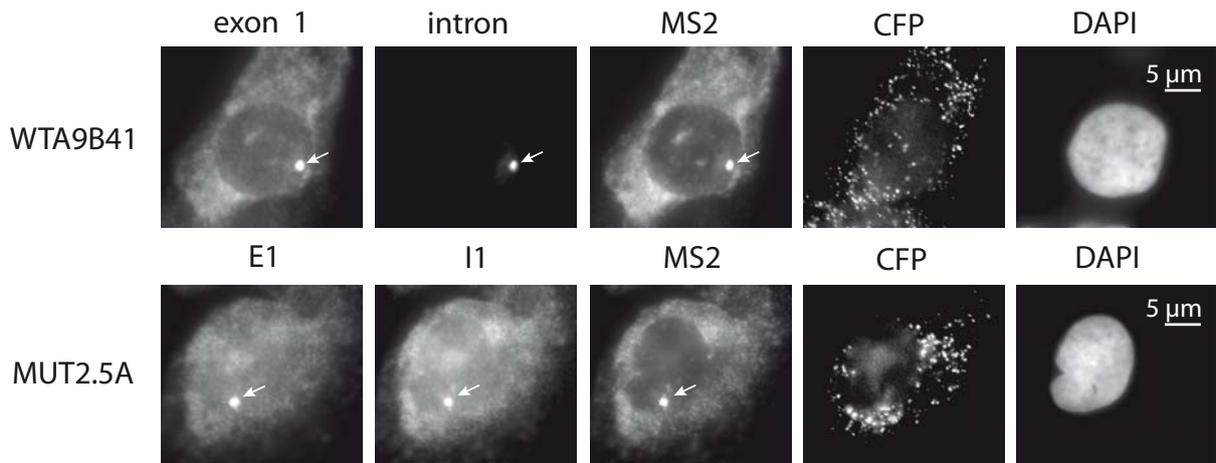
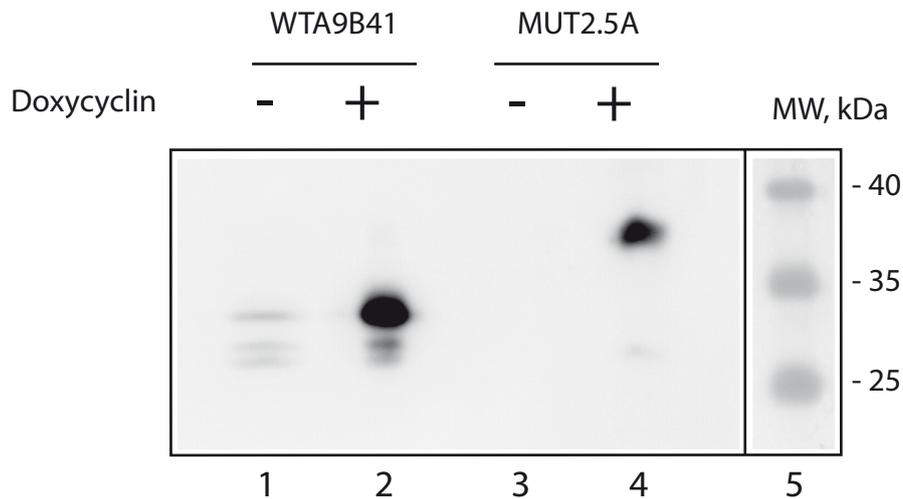
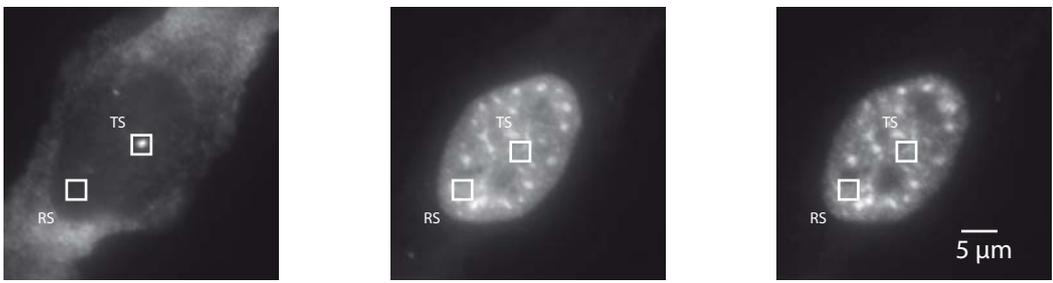


Figure 1

A**figure 2****B****C**



MS2

U1

U2

TS

RS

TS

RS

TS

RS

#1



#2



#3



#4



#5



...

...

...

...

...

...

...

TS

RS

TS

RS

TS

RS

S



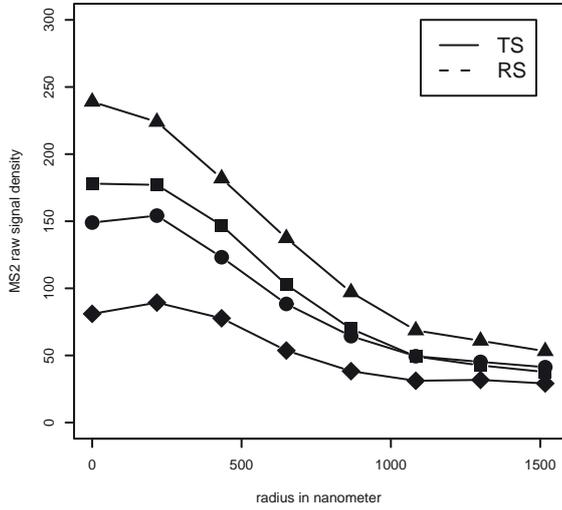
figure 3

figure 4

MS2

TS

RS



U1

TS

RS

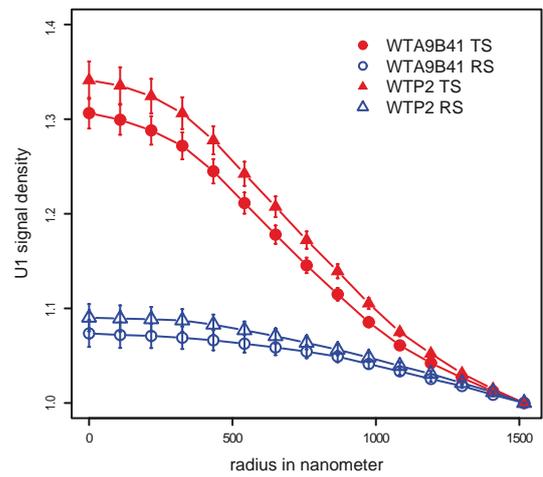
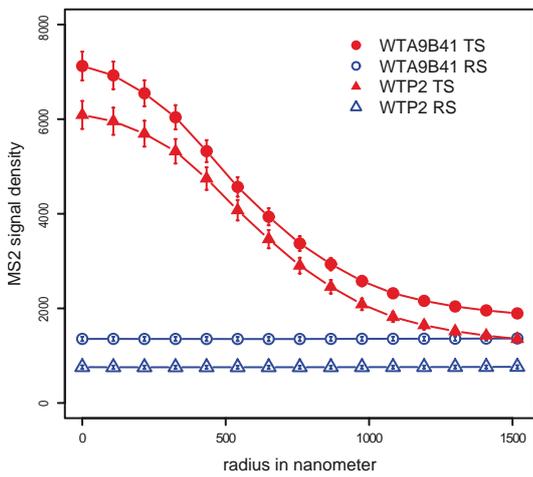
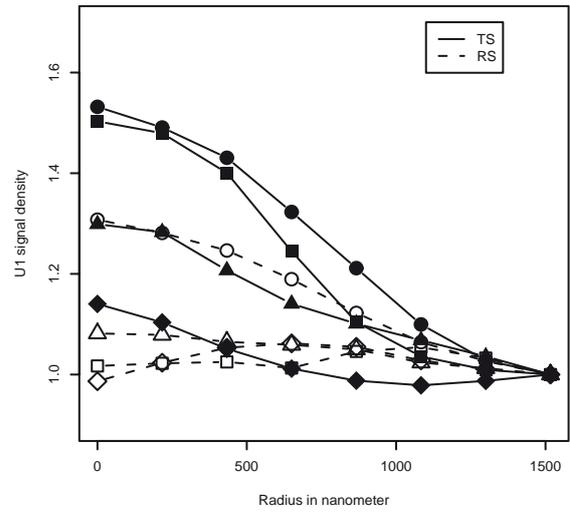
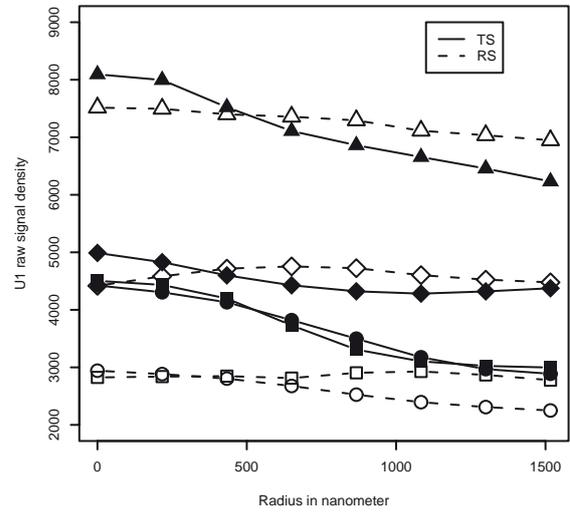
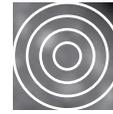
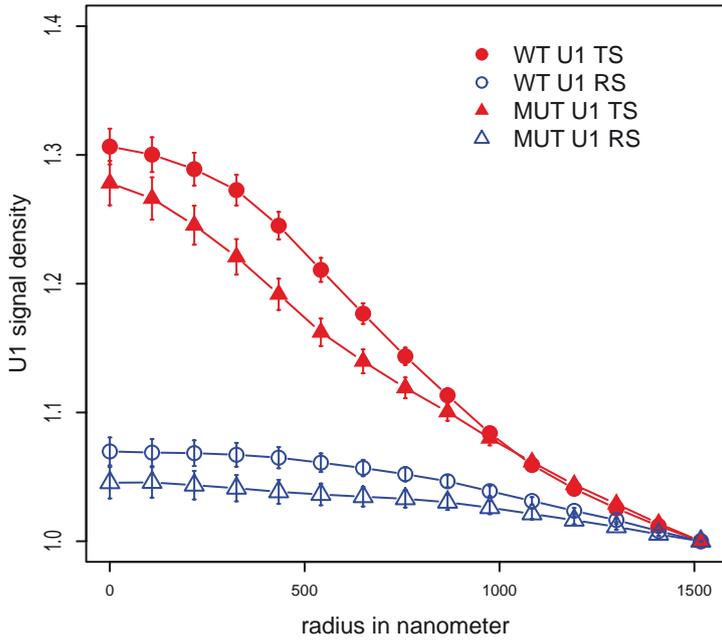
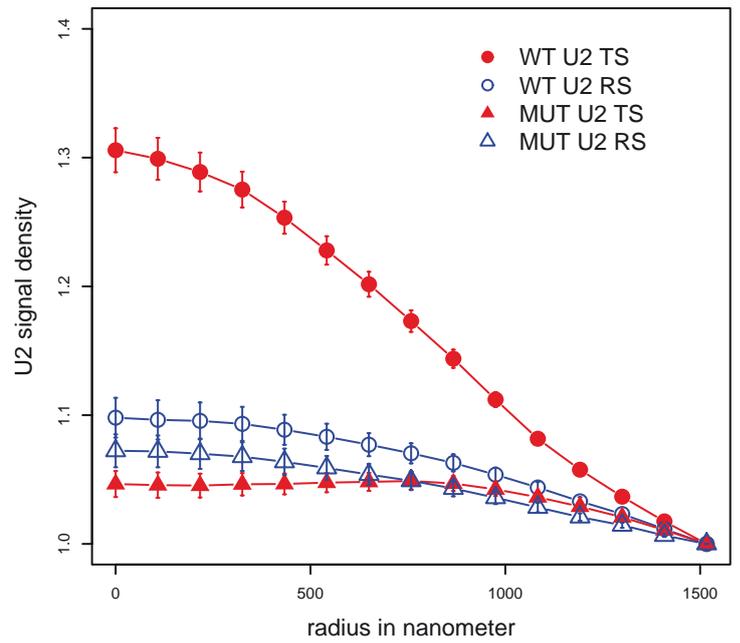


figure 5

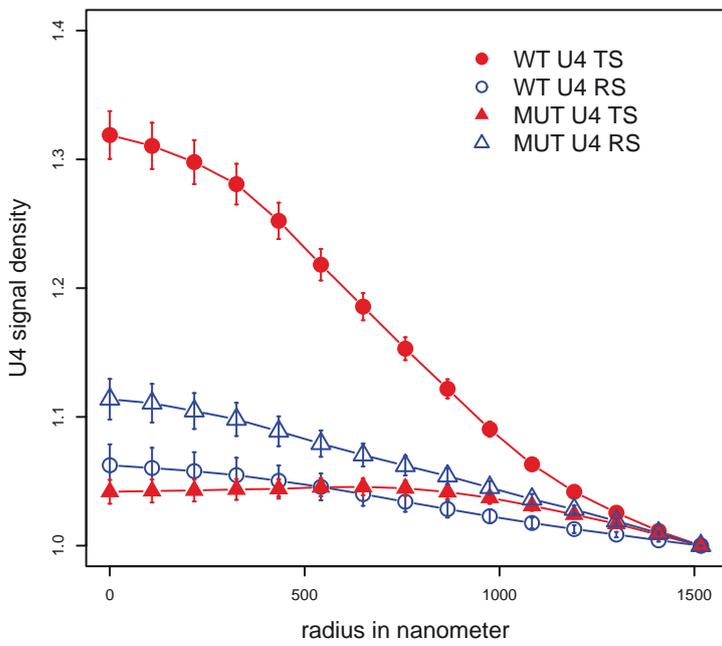
U1 snRNA



U2 snRNA



U4 snRNA



tri snRNAs

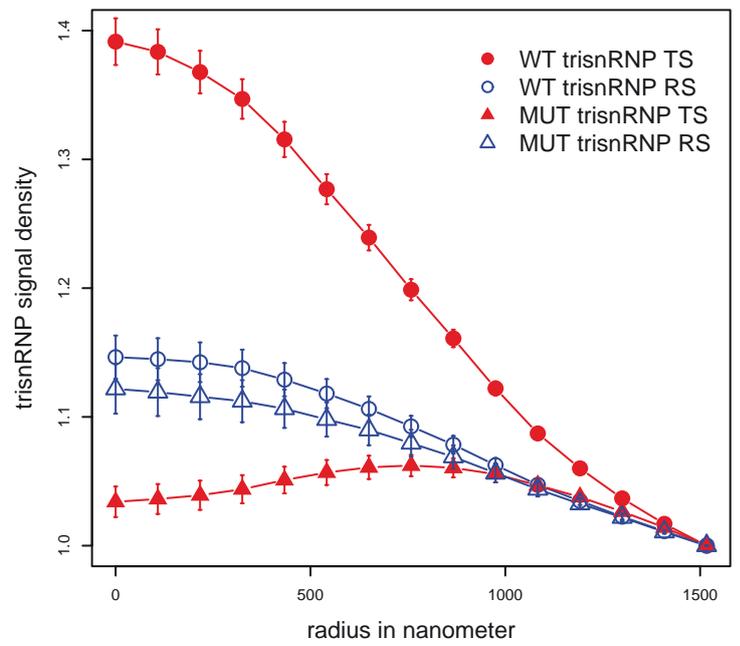


figure 6

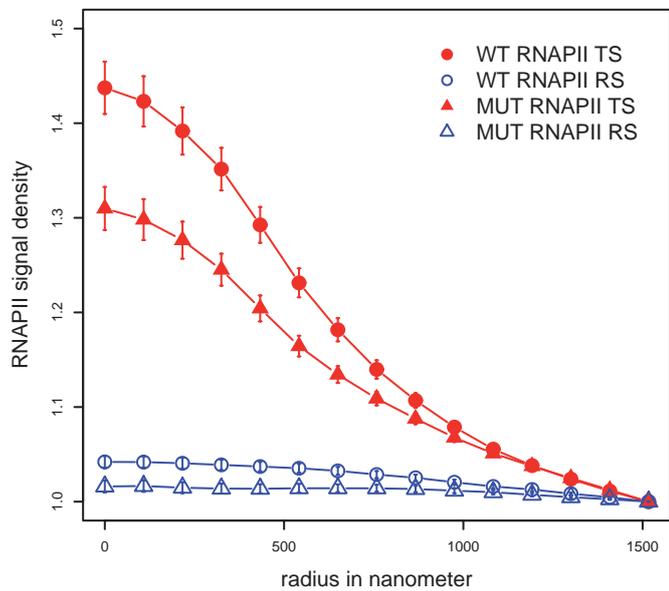
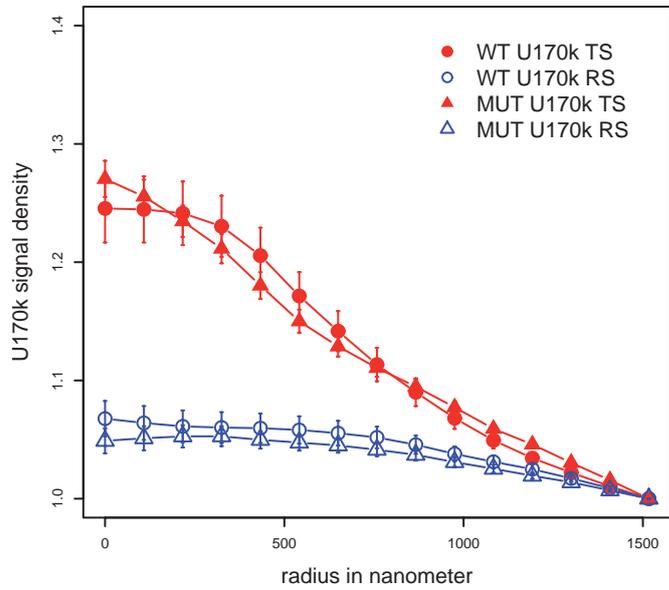
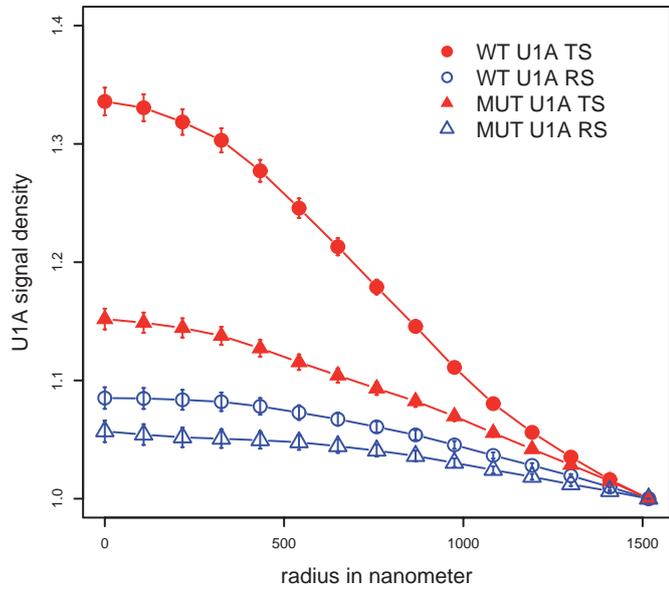
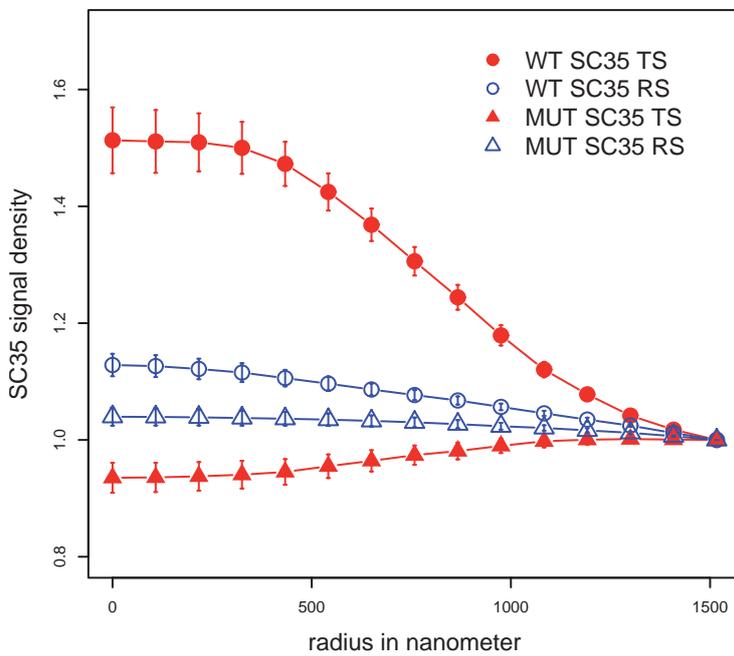
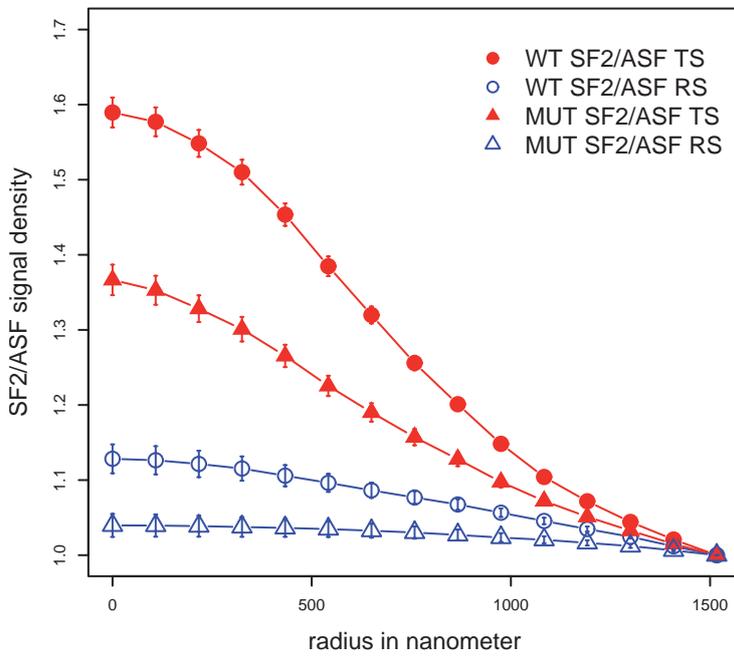
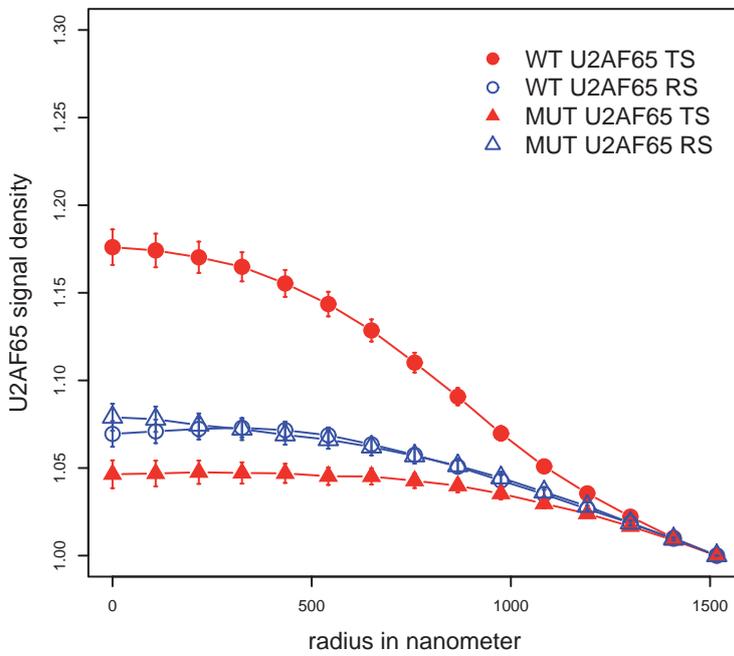


figure 7



Non-coding RNAs regulating the transcriptional machinery

Charlotte Barrandon, Béatrice Spiluttini and Olivier Bensaude¹

CNRS UMR 8541, Ecole Normale Supérieure, 46 rue d'Ulm, Paris, Cedex 05, France

During the past decade, numerous ncRNAs (non-coding RNAs) have been identified as regulators of transcription. This review focuses on a few examples of ncRNAs that directly interact with and regulate components of the transcription machinery. Artificial RNA aptamers have been selected against components of the transcriptional machinery. The bacterial 6S RNA and the eukaryotic B2 RNA directly target RNA polymerases. The 7SK RNA, U1 snRNA (small nuclear RNA) and SRA (steroid receptor RNA activator) RNA bind to and regulate the activity of transcription factors. Xist (X-inactive-specific transcript) and roX (RNA on the X) RNAs are involved in epigenetic regulation of transcription through the recruitment of histone-modifying enzymes.

Introduction

During the past decade, numerous ncRNAs (non-coding RNAs) have been identified as regulators of transcription (Prasanth and Spector, 2007). Short (21–23 nt) double-stranded RNA molecules, such as siRNAs (silencing RNAs) and shRNAs (small hairpin RNAs), are today among the most well-characterized ncRNAs (Rana, 2007). siRNAs and shRNAs have become very useful tools for studying gene inactivation, and have revolutionized genetics. Endogenous short (19–25 nt), natural ncRNAs, which are known as miRNAs (microRNAs), appear to play an important role in gene regulation during development (Bushati and Cohen, 2007; Zhao and Srivastava, 2007). Short, as well as long, non-coding transcripts are involved in epigenetic regu-

lation (Yang and Kuroda, 2007; Zaratiegui et al., 2007). In this review, we will focus on a few examples of ncRNAs that directly interact with and regulate components of the transcription machinery. Artificial RNA aptamers that have been selected against components of the transcriptional machinery serve as useful models. The bacterial 6S RNA is one of the first ncRNAs to be described. Similar to the eukaryotic B2 RNA, it directly targets RNA polymerase. The 7SK RNA, U1 snRNA (small nuclear RNA) or SRA (steroid receptor RNA activator) RNA bind to and regulate the activity of transcription factors. IGS (intergenic spacer), Xist (X-inactive-specific transcript) and roX (RNA on the X) RNAs are involved in epigenetic regulation of transcription (Figure 1).

RNAs that bind and inhibit RNA polymerases

In a pioneering study, RNA aptamers were isolated using yeast RNA polymerase II as target molecule (Thomas et al., 1997). It was expected that aptamers directed against protein surfaces involved in transcription complex formation would become useful tools in functional studies. One of these aptamers, FC RNA, strongly binds to and inhibited the yeast *Saccharomyces cerevisiae* RNA polymerase II. This effect is highly specific, as neither *S. cerevisiae* RNA polymerase I and III nor *Schizosaccharomyces pombe* and wheat-germ RNA polymerase II are sensitive. FC RNA inhibits assembly of RNA polymerase into a PIC (pre-initiation complex) for transcription on promoter DNA. However, it does not bind RNA

¹To whom correspondence should be addressed (email bensaude@biologie.ens.fr).

Key words: non-coding RNA, 7SK RNA, 6S RNA, steroid receptor RNA activator (SRA), transcription.

Abbreviations used: AR, acidic region; ARM, arginine-rich motif; BR, basic region; CDK, cyclin-dependent kinase; CTD, C-terminal domain; DHFR, dihydrofolate reductase; FC* RNA, central fragment of FC RNA; HEXIM, hexamethylene bis-acetamide-inducible mRNA; hnRNP, heterogeneous nuclear ribonucleoprotein; IGS, intergenic spacer; LARP, La-related protein; ncRNA, non-coding RNA; NoRC, nucleolar chromatin-remodelling complex; PcG, Polycomb group; PIC, pre-initiation complex; PIP7S, protein interacting with P-TEFb for 7SK stability; PRC, Polycomb repressive complex; pRNA, RNA product; pspF, phage-shock protein F; P-TEFb, positive transcription elongation factor b; RNAi, RNA interference; roX, RNA on the X; SHARP, SMRT1/HDAC1-associated repressor protein; shRNA, small hairpin RNA; siRNA, silencing RNA; SLIRP, SRA stem-loop interacting protein; snRNA, small nuclear RNA; SRA, steroid receptor RNA activator; SRAP, SRA protein; SRC, steroid receptor co-activator complex protein; TAR, *trans*-activating region; TBP, TATA-box-binding protein; TF, transcription factor; TRM, Tat/TAR recognition motif; Xist, X-inactive-specific transcript; Tsix, Xist antisense.

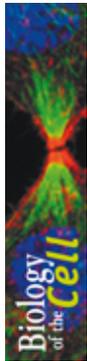
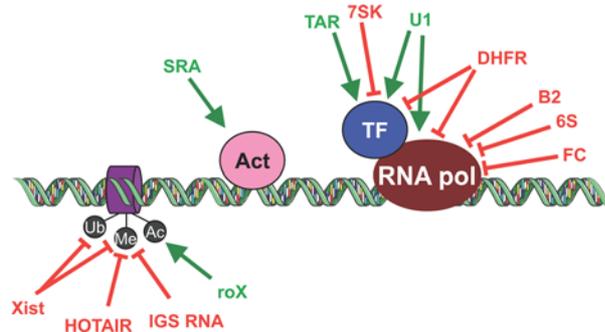


Figure 1 | ncRNAs regulate the transcription machinery at various steps

ncRNAs regulate transcription positively (green arrows) or negatively (red lines). ncRNAs regulate activity of histone-modification enzymes. Ac, acetylation; Act, specific transcription activators; Me, methylation; RNA pol, RNA polymerase; Ub, ubiquitination.



polymerase molecules engaged in transcription. This characteristic was understood recently when the structure of a complex formed by yeast RNA polymerase II and FC* RNA (a central fragment of FC RNA) was determined at 3.8 Å resolution (Kettenberger et al., 2006). The FC* RNA forms a double stem-loop, which is inserted within the polymerase cleft that accommodates the DNA template. Its positioning overlaps the positioning of the DNA template and the nascent transcript in an elongation complex for transcription. Thus FC* RNA prevents the DNA template from entering the cleft and inhibits the initiation of transcription. In contrast, elongation is not inhibited by FC* RNA, as it cannot displace the DNA template from the cleft. These structural data might be useful in the understanding of the mechanism of RNA polymerase inhibition by natural ncRNAs, as discussed below for the bacterial 6S RNA and the murine B2 RNA.

6S RNA: a template for bacterial DNA-dependent RNA polymerase

The *Escherichia coli* 6S RNA was among the first ncRNAs to be described (Hindley, 1967). It accumulates, upon entry into stationary phase, from 1000 to 10 000 copies per cell. The *ssrS* (6S) and *ygfA* genes form an operon (Hsu et al., 1985; Kim and Lee, 2004). Two distinct transcripts, initiated from two distinct promoters, generate a unique 6S RNA after

processing by endonucleases RNase E and RNase G (Kim and Lee, 2004). Transcription from the proximal promoter is σ^{70} -dependent, whereas the more distal promoter is both σ^{70} - and σ^S -dependent. The *E. coli* 6S RNA associates and represses bacterial RNA polymerase activity (Wassarman and Storz, 2000). It co-purifies with the major RNA polymerase holoenzyme that comprises the σ^{70} subunit. It is not associated with holoenzymes comprising another σ subunit. 6S RNA accumulates upon entry into stationary phase and is associated with most RNA polymerase- σ^{70} holoenzymes purified from stationary-phase cells. *In vitro*, 6S RNA association with RNA polymerase is recapitulated with a holoenzyme comprising σ^{70} . 6S-RNA-binding stabilizes the RNA polymerase- σ^{70} holoenzyme, as less σ^{70} is co-immunoprecipitated with RNA polymerase from extracts of bacterial cells lacking the *ssrS* gene. 6S RNA does not associate with the core polymerase, isolated σ^{70} subunit or the stationary-phase holoenzyme that contains the σ^S subunit (Wassarman and Storz, 2000; Gildehaus et al., 2007). Association of 6S RNA inhibits transcription from several, but not all, σ^{70} -dependent promoters. It has been suggested that its interaction with RNA polymerase contributes to alter the gene-expression profile when bacteria enter the stationary phase.

High constitutive 6S RNA overexpression causes reduced cell viability (Wassarman and Storz, 2000). Disruption of the *ssrS* gene in *E. coli* genome results in a decreased ability to persist during long-term nutrient deprivation in a time frame of weeks (Trotochaud and Wassarman, 2004), as well as a increased survival at a high pH value (pH 9.3) (Trotochaud and Wassarman, 2006). The *pspF* (phage-shock protein F; an AAA+ transcriptional activator) gene is one of the few indisputably repressed *in vivo* by high levels of 6S RNA. The increased resistance at high pH of bacterial cells lacking the *ssrS* gene is linked to the restored expression of the *pspF* gene. The *pspF* protein catalyses open promoter complex formation and facilitates initiation of transcription by the RNA polymerase- σ^{54} holoenzyme (Rappas et al., 2007). RNA polymerase- σ^{54} is required to transcribe a subclass of genes that may prevent depletion of energy resources in certain environments or partially neutralize adverse conditions (Reitzer and Schneider, 2001).

The 6S RNA is widely spread in eubacteria (Barrick et al., 2005; Trotochaud and Wassarman,

2005; Willkomm et al., 2005). The linkage between *ssrS* and *ygfA* genes has been conserved in evolution. Some species, such as *Bacillus subtilis*, contain two distinct 6S RNA genes (initially described as *bsrB* and *bsrA*). According to a recently proposed nomenclature, these genes would now be designated as *ssrSA* and *ssrSB*, which code for 6Sa and 6Sb RNA respectively. The *B. subtilis* 6S RNAs show distinct regulations. The 6Sa RNA peaks in mid-log phase. In contrast, the 6Sb RNA behaves like the genuine *E. coli* 6S RNA, with its concentration increasing markedly (~18-fold) between early log phase and stationary phase.

6S RNA: an RNA template mimicking an open promoter structure

The 6S RNAs probably fold into closely related structures that have been conserved across various bacterial species (Barrick et al., 2005; Trotochaud and Wassarman, 2005). Two hairpin/stalk structures are separated by a large bulge or 'bubble' of two single strands, which extends over 10–15 nt. Chemical cleavage analysis support this folding hypothesis. The 6S RNA sequence surrounding the bubble makes contacts directly with both the σ^{70} and β/β' polymerase subunits within the holoenzyme (Gildehaus et al., 2007). The separated single-stranded bubble between the stalk and the hairpin evokes the melted DNA bubble that forms in transcribing polymerases (Barrick et al., 2005; Trotochaud and Wassarman, 2005). The 6S RNAs may bind within the RNA polymerase cleft that contains the active site, as reported above for the FC* RNA aptamer. Binding of the 6S RNA probably competes for promoter DNA binding. In a striking support for this analogy, the bacterial RNA polymerase uses the 6S RNA as a template to synthesize short (14- to 20-nt) pRNAs (RNA products) that initiate within the bubble (Gildehaus et al., 2007; Wassarman and Saecker, 2006). The pRNAs can form a triple-helix hybrid with the 6S RNA hairpin. This hybrid might destabilize the RNA polymerase–6S RNA complex, thereby allowing restoration of polymerase activity. *In vitro* pRNA synthesis from 6S RNA requires higher concentrations of NTPs than that for transcription from DNA promoters. Hence, it is suggested that low NTP concentrations at stationary phase stabilize the inactive RNA polymerase–6S RNA complex. Although cells lacking the 6S RNA display relatively

subtle phenotypes, the 6S RNA has been conserved in many bacterial species, and 6S-dependent transcription inhibition has been demonstrated *in vitro*. Inhibition of bacterial RNA polymerase by 6S RNA therefore constitutes a useful model system.

Murine B2 RNA: regulating RNA polymerase II during stress

The murine B2 RNA is likely to be a eukaryotic counterpart of the bacterial 6S RNA. B2 RNA is an RNA polymerase III transcript of 178 nt that accumulates in rodent cells following various stress, including heat shock (Fornace et al., 1989). It has been reported to inhibit transcription *in vitro* and to be co-immunoprecipitated with RNA polymerase II in nuclear extracts prepared from heat-shocked cells (Allen et al., 2004; Espinoza et al., 2004).

Similar to the FC RNA, B2 RNA associates directly with core RNA polymerase II. As the binding of B2 is competitive with that of the FC* RNA aptamer, it has been proposed that B2 binds within the polymerase cleft surrounding the active site, similar to the FC RNA (Kettenberger et al., 2006). Like 6S RNA, B2 RNA does not bind RNA polymerase once the latter has assembled into a PIC on a DNA template (Espinoza et al., 2004). Unlike the FC aptamer and 6S RNA, the region of the B2 RNA that is important for transcriptional repression is single-stranded and not structured (Espinoza et al., 2007). Deletion of a 33-nt region did not alter B2 RNA binding to RNA polymerase, but nullified the transcription inhibition effect. Hence, B2 RNA binding can be distinguished from transcription inhibition.

B2 RNA has been suggested to contribute to the repression of the transcription of house-keeping genes, such as actin and hexokinase II, following a heat-shock stress. In contrast, heat-shock genes might escape this inhibition. This is reminiscent of the bacterial 6S RNA that accumulates during stress and represses a subclass of genes.

ncRNAs targeting transcription initiation factors

In a recent study, aptamers were generated against a general transcription factor, the TATA-box-binding protein (TBP) (Shi et al., 2007). Two classes of aptamers were isolated. Both classes inhibit transcription initiation *in vitro* through distinct mechanisms. A first round of selection generated

exclusively aptamers that recognize the DNA-binding site and hence interfere with formation of TBP–DNA complexes. However, the DNA-binding domain is positively charged, and this feature introduces a bias in the aptamer selection, as RNAs are anions and tend to bind positively charged domains. It is notable that the aptamers mentioned previously that are directed against RNA polymerase were also directed against a DNA-binding domain. To bypass this bias, aptamers selected through the first round of selection were used to ‘mask’ the DNA-binding site, and a second round of selection using TBP–aptamer complexes then provided a new class of aptamers that were directed against other parts of the protein. This second class of aptamers competes with TBP binding to the general transcription factor TFIIA (transcription factor IIA) that is required to assemble PICs on promoters.

A natural non-coding transcript that prevents the assembly of the PIC of transcription has been described recently in human cells. Repression of the *DHFR* (dihydrofolate reductase) gene in quiescent cells relies upon a short transcript that accumulates upon serum starvation (Martianov et al., 2007). This short transcript initiates from a minor promoter upstream of the major promoter that directs the *DHFR* pre-mRNA. It terminates within the second intron of the *DHFR* pre-mRNA. It forms a triple-helix complex with its double-stranded DNA template sequence. Furthermore, this short transcript binds the general transcription factor TFIIB. As a result, the PIC that forms on the *DHFR* promoter can become dissociated and its transcription would be inhibited. Such mechanism of transcription inhibition might be very general, as 5′ and 3′ ends of almost half of human protein coding genes generate short non-coding transcripts (Kapranov et al., 2007).

TAR (*trans*-activating region) RNA recruits P-TEFb (positive transcription elongation factor b) on the HIV LTR (long terminal repeat) promoter

P-TEFb can be considered as a paradigm for transcription factors regulated by ncRNAs. It comprises a kinase subunit, CDK9 (cyclin-dependent kinase 9), and its associated cyclin T1 or T2 (Peterlin and Price, 2006). Studies of HIV genome transcription provided the first clues that an RNA regulated the function of this general transcription factor. Activ-

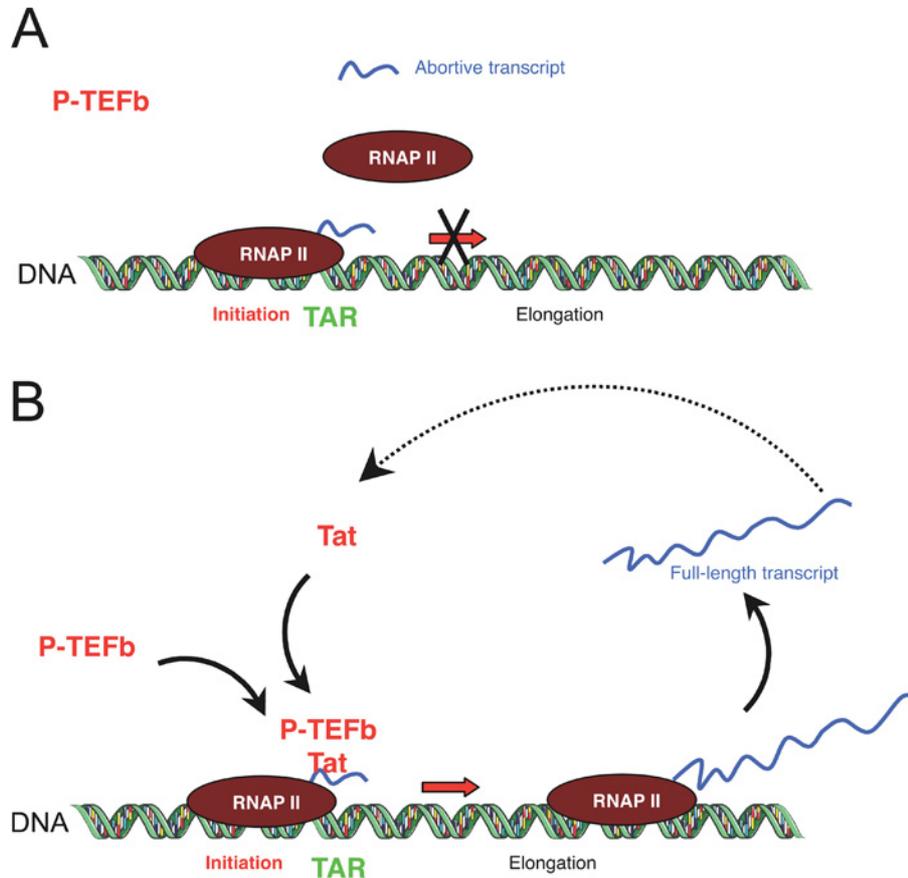
ation of the HIV 5′ LTR (long terminal repeat) promoter transcription relies upon the recruitment of P-TEFb to the transcription complex by Tat, a small protein coded by the viral genome and the TAR RNA, an RNA loop that folds within the first 50 nt of the viral transcript (Bannwarth and Gatignol, 2005) (Figure 2). When one of these elements is deficient or absent, transcription initiates efficiently, but aborts after 60–80 nt and RNA polymerase II is dissociated from the complex. Phosphorylation of the RNA polymerase II CTD (C-terminal domain) of negative elongation factors Spt5 (suppressor of Ty 5) and NELF (negative elongation factor) by the Tat/TAR-recruited P-TEFb is required for a productive elongation of transcription (Mancebo et al., 1997; Zhu et al., 1997; Ping and Rana, 2001; Fujinaga et al., 2004; Phatnani and Greenleaf, 2006). Furthermore, Tat/TAR-enhanced phosphorylation of the CTD stimulates co-transcriptional capping of HIV pre-mRNA (Zhou et al., 2003). As Tat is coded by the viral genome, the recruitment of Tat/TAR constitutes the first step in a positive feed-back loop controlling HIV genome transcription (Weinberger and Shenk, 2007).

Recruitment of P-TEFb on the HIV promoter relies on the binding of Tat to both TAR RNA and human cyclin T1. A short sequence (amino acids 251–265) placed after the cyclin box in cyclin T1, the TRM (Tat/TAR recognition motif), is critical for transactivation (Garber et al., 1998). A single amino-acid substitution in the TRM, C261Y in rodent cyclin T1, suppresses the affinity of P-TEFb for Tat. This cysteine residue interacts with Cys²² in the Tat sequence. Other residues in the TRM are involved in TAR binding.

The TAR RNA interacts with a stretch of basic amino acids, the ARM (arginine-rich motif) in the Tat sequence. Acetylation of Lys⁵⁰ in the ARM, possibly by p300 histone acetyl transferase, adds an additional level of regulation, as it interferes with TAR binding to Tat (Bres et al., 2002; Kaehlcke et al., 2003). It is suggested that this acetylation contributes to the release of the viral transcript from P-TEFb once elongation proceeds. A characteristic loop structure in HIV-1 TAR RNA is required for the formation of the cyclin T1–Tat–TAR ternary complex. Cyclin T1 remodels the structure of Tat to enhance its affinity for TAR RNA. Conversely, TAR RNA further enhances the interaction between Tat and CycT1

Figure 2 | HIV LTR positive feedback loop

Transcription from the HIV promoter is abortive when the viral protein Tat is missing (A). Tat binds the TAR RNA structure that folds within the abortive transcript and binds the cyclin T1 subunit of P-TEFb. As a result, P-TEFb is recruited on the transcription machinery, phosphorylates RNA polymerase and negative elongation factors, thereby allowing full-length HIV transcripts to be made (B). Translation of the viral transcripts provides the viral proteins including Tat.



(Zhang et al., 2000). It has been proposed that the TAR RNA nucleates the formation of the Tat–P-TEFb complex through an induced fit mechanism. CDK9 usually phosphorylates Ser² in the repeats of the RNA polymerase II CTD. Remarkably, in the presence of TAR and Tat, the substrate specificity of CDK9 is altered, so that the kinase now phosphorylates both Ser² and Ser⁵ in the repeat regions (Zhou et al., 2000). Although the TRM is relatively far apart from the putative CDK9-binding interface on cyclin T1 (Anand et al., 2007), the conformational changes promoted by Tat/TAR RNA binding influences CDK9 kinase activity. Nevertheless, targeting an essential transcription factor on RNA polymerase is the major function of the TAR RNA.

7SK RNA: a negative cellular regulator of P-TEFb

Two major complexes of P-TEFb are found in mammalian cells and have been separated by ultracentrifugation on glycerol gradients (Nguyen et al., 2001). The smaller complex corresponds to the active form of P-TEFb. It comprises a kinase CDK9 and cyclin T1 or T2 subunits. The larger complex is inactive and comprises a small nuclear RNA, the 7SK RNA (Nguyen et al., 2001; Yang et al., 2001), as well as HEXIM1 (hexamethylene bis-acetamide-inducible mRNA1) or HEXIM2 proteins in addition to CDK9 and cyclin T1 or T2 (Michels et al., 2003; Yik et al., 2003; Byers et al., 2005; Dulac et al., 2005; Yik et al., 2005).

The 7SK RNA is an abundant non-coding 331 nt RNA polymerase III transcript, which is conserved in vertebrate genomes (Wassarman and Steitz, 1991; Gürsoy et al., 2000). Similar to U6 snRNA, but unlike RNA polymerase transcripts, 7SK RNA is mono-methylated on the γ phosphate of its 5' end guanosine (Gupta et al., 1990). The methylase, BCDIN3 (bicoid-interacting 3), has been identified recently in a proteomic screen and shown to remain stably bound to the methylated 7SK RNA associated with P-TEFb (Jeronimo et al., 2007). The 7SK RNA is predicted to fold into a long stem-loop at the 5' end and a smaller stem-loop at the 3' end (Wassarman and Steitz, 1991). Numerous snRNAs possess a 3' stem-loop and terminate with a stretch of uridines that associate the La protein. The La protein provides protection against degradation and displays RNA chaperone-like activity (Maraia and Bayfield, 2006). The 3' end of 7SK RNA shows weak binding for the La protein and strong binding towards LARP7/PIP7S (La-related protein 7/protein interacting with P-TEFb for 7SK stability) (Q. Zhou, personal communication). Indeed, knock-down of LARP7/PIP7S results in degradation of 7SK RNA.

The long 5' stem-loop that extends from nt 1 to 108 in 7SK RNA, more specifically between nt 24 and 87, binds to the HEXIM1 protein (Michels et al., 2004; Egloff et al., 2006). Alternatively, data *in vitro* suggested that nt 10 to 48 in 7SK RNA might fold into a distinct hairpin that also binds to HEXIM1 (Li et al., 2007). The 3' stem-loop from nt 296 to 331 binds to HEXIM1 *in vitro* (Sedore et al., 2007). It evokes the TAR RNA stem-loop and might directly bind *in vivo* P-TEFb without HEXIM1 bridging (Egloff et al., 2006). On the other hand, TAR RNA may also bind HEXIM1 (Sedore et al., 2007). Indeed, human HEXIM1 has been reported to be a promiscuous double-stranded RNA-binding protein and to co-immunoprecipitate with miRNAs, such as miR-16-1 (Li et al., 2007). However, 7SK RNA is probably the major HEXIM1 ligand.

Binding of the 7SK snRNA changes the HEXIM1 protein into a P-TEFb inhibitor

Addition of synthetic 7SK RNA and purified recombinant HEXIM1 to purified recombinant P-TEFb recapitulated its inhibition *in vitro* (Michels et al., 2004). 7SK RNA binds HEXIM1 and its cognate

HEXIM2 protein through the BR (basic region), a stretch of basic amino acids, mostly arginine residues, which evokes the ARM motif in Tat (Michels et al., 2004; Yik et al., 2004). The C-terminal domain of HEXIM proteins contacts the cyclin box in the N-terminal domain of cyclin T1 or T2 (Michels et al., 2003; Schulte et al., 2005). Binding of 7SK RNA to HEXIM proteins unveils its cyclin T-binding capacity. It has been proposed that the BR interacts with the AR (acidic region) (Barboric et al., 2005). Binding of 7SK RNA would neutralize the positive charges of the BR, the resulting disruption of the AR-BR interaction would unmask the cyclin T-binding domain in HEXIM. Purified HEXIM proteins form highly anisotropic oligomers (probably dimers) (Dulac et al., 2005; Li et al., 2005; Schulte et al., 2005; Yik et al., 2005). The structure of the HEXIM1 dimerization domain has been solved by NMR (Dames et al., 2007). It forms an extended coiled-coil structure that may account for the anisotropy of the HEXIM protein. HEXIM oligomerization is required for P-TEFb binding, but is not affected by it, although it involves a region in its C-terminal domain that overlaps the cyclin T-binding domain. The 'large' P-TEFb complex comprises several (probably two) CDK9/cyclin T modules, in addition to a single 7SK RNA molecule.

HEXIM1 overexpression inhibited the basal level of HIV LTR expression (Yik et al., 2003; Barboric et al., 2007). As HEXIM1 also binds TAR RNA, it might be recruited to the 5' end of the viral transcript and maintain P-TEFb inhibited in the absence of Tat (Sedore et al., 2007). Both HIV Tat and HEXIM proteins interact with the N-terminal domain of cyclin T1 (Michels et al., 2003; Schulte et al., 2005). This association is competitive. *In vitro*, Tat caused the release of P-TEFb from a preformed P-TEFb-HEXIM1-7SK RNA complex (Barboric et al., 2007; Sedore et al., 2007). Conversely, HEXIM1 overexpression inhibited Tat-induced HIV LTR transcription (Fraldi et al., 2005; Shimizu et al., 2007).

Dynamic exchange of 7SK RNA between P-TEFb and hnRNPs (heterogeneous nuclear ribonucleoproteins)

The small and large P-TEFb complexes are in dynamic exchange in exponentially growing cells (Nguyen et al., 2001). Within minutes following

inhibition of transcription with drugs or UV irradiation, the P-TEFb–HEXIM–7SK complexes are disrupted and core P-TEFb accumulates (Nguyen et al., 2001; Yang et al., 2001; Michels et al., 2003; He et al., 2006; Biglione et al., 2007). However, core P-TEFb does not accumulate in cells arrested during mitosis, although transcription is arrested (Barrandon et al., 2007). As a result of P-TEFb–HEXIM–7SK complex disruption, an increase in CDK9 activity is detected (Nguyen et al., 2001). This increase might contribute to hyperphosphorylating the CTD of RNA polymerase II and activating the transcription of HIV genome in response to stress (in the absence of Tat) (Valerie et al., 1988; Dubois et al., 1994; Cassé et al., 1999; Imamichi et al., 2005)

7SK RNA released from P-TEFb–HEXIM is not degraded, it is captured by a subclass of hnRNPs (hnRNPs A1, A2, K and Q/R), as well as the RHA (RNA helicase A) RNA helicase (Barrandon et al., 2007; Hogg and Collins, 2007; Van Herreweghe et al., 2007). However, inhibition of transcription leads to an increase in the amount of 7SK RNA bound to these hnRNPs, suggesting that 7SK RNA shuttles from P-TEFb–HEXIM to the hnRNP complexes. hnRNPs associate with nascent transcripts (Dreyfuss et al., 2002; Bomsztyk et al., 2004). In response to an arrest in transcription, nascent transcripts are suppressed, and hnRNPs are released from pre-mRNA. Two experiments suggest that binding of the released hnRNP to 7SK RNA might contribute to P-TEFb–HEXIM–7SK complex dissociation. First, RNAi (RNA interference) knockdown of hnRNP K increases the amount of P-TEFb–HEXIM–7SK complexes (Hogg and Collins, 2007). RNAi knockdown of hnRNP A1 and A2 attenuates the dissociation of P-TEFb–HEXIM–7SK complexes (Barrandon et al., 2007). Furthermore, hnRNPs bind 7SK RNA through a hairpin located in the middle of the sequence (Van Herreweghe et al., 2007). When this sequence is deleted, binding of the resulting ‘truncated 7SK’ to P-TEFb is insensitive to transcriptional arrest.

Our current knowledge is summarized with a simple model. Shortly after initiation, when P-TEFb is loaded on to the transcription machinery, 7SK dissociates from P-TEFb and is trapped on hnRNPs which have just been released from a mature transcript (Figure 3A). Conversely, P-TEFb–HEXIM–7SK would reassemble during the elonga-

tion/termination step, as 7SK RNA is competed out by nascent transcripts (Figure 3B). Thus in exponentially growing cells, 7SK RNA trapping by hnRNPs might act as a sensor for transcription, thereby controlling P-TEFb activity and fine-tuning the overall class II gene transcription efficiency.

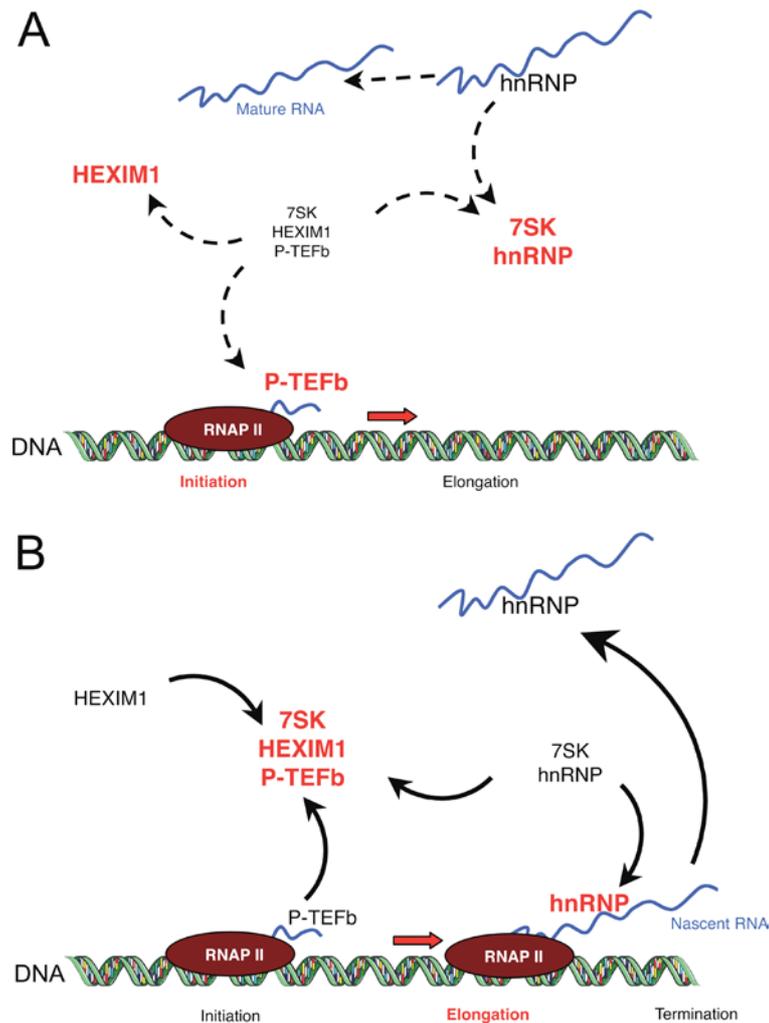
U1 snRNA: linking transcription and post-transcriptional RNA processing

snRNAs form the functional core of the spliceosome. Splicing is a co-transcriptional process, and spliceosomes associate with RNA polymerase II (Bentley, 2005; Listerman et al., 2006). Studies suggest a direct and specific involvement of the U1 snRNA in transcription regulation, although it is primarily known as a major spliceosome component. U1-dependent recognition of the 5' splice site in pre-mRNA constitutes the first step in assembly of the spliceosome (Lacadie and Rosbash, 2005; Nilsen, 2005). U1 snRNP also forms a complex with CPSF (cleavage and polyadenylation specific factor) (Awasthi and Alwine, 2003). U1 snRNA co-purified with a fraction of TFIIF and was cross-linked to its cyclin H subunit (Kwek et al., 2002). TFIIF is a multisubunit factor that plays a critical role in transcription initiation of both class I and class II genes (Zurita and Merino, 2003). An MS analysis of proteolytic fragments generated after cross-linking indicated that two stem-loops of synthetic U1 RNA made contact with peptides both in the N- and C-termini of purified recombinant cyclin H (O'Gorman et al., 2005). These peptides appear in close vicinity in the three-dimensional structure of cyclin H (Andersen et al., 1997). U1 snRNA stimulated the rate of formation of the first phosphodiester bond by RNA polymerase II (Kwek et al., 2002). Furthermore, addition of synthetic U1 RNA stimulated the CDK7 kinase activity *in vitro*. However, although CTD phosphorylation by the CDK7 subunit of TFIIF is an early step (Phatnani and Greenleaf, 2006), its function in transcription remains unclear (Kanin et al., 2007). In an *in vitro* reconstitution assay using purified factors, addition of synthetic U1 RNA stimulated abortive initiation, as well as re-initiation of transcription from a promoter preceding a 5' splice site recognized by U1 snRNA (Kwek et al., 2002).

In contrast with the bacterial 6S and the murine B2 RNAs, U1 snRNA stimulates transcription. U1

Figure 3 | Coupled hnRNP-7SK and P-TEFb-HEXIM1-7SK RNA complex formation: a model

(A) As P-TEFb is recruited on to the transcription initiation complex, 7SK may be released from P-TEFb-HEXIM1-7SK RNA complex and transferred on to hnRNP. (B) hnRNP bind nascent RNA extruded from elongating RNA polymerases and 7SK RNA is displaced. 7SK RNA then binds to HEXIM proteins and associates with P-TEFb released from elongation complex. Reproduced with permission from Barrandon, C., Bonnet, F., Nguyen, V.T., Labas, V. and Bensaude, O. The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. *Mol. Cell. Biol.* **27**, 6996-7006 © American Society for Microbiology.



snRNP might be specifically recruited by RNA polymerase on to the transcribing unit as the whole U1 snRNP (comprising U1 snRNA, U1-70K, U1A and U1C proteins) (Das et al., 2007). It is tempting to speculate that U1 snRNA brought about by RNA polymerase facilitates the initiation of transcription assisted by TFIIF. U1 snRNA 'piggy-backed' on RNA polymerase might scan the nascent transcript and account for the enhanced transcription direc-

ted by promoter proximal splice sites (Furger et al., 2002).

SRA: a transcriptional co-regulator ncRNA

The transcriptional activity of the steroid receptor is modulated by co-activators and co-repressors (McKenna and O'Malley, 2002). One of the

identified co-activators, the SRA is a ncRNA of 883 nt (Lanz et al., 1999). The SRA was unexpectedly discovered in a yeast two-hybrid screen for proteins that interact with the N-terminal domain of the human progesterone receptor. It was discovered that the short open reading frames in the SRA sequence are not required for its activity. The SRA RNA folds under multiple secondary structure motifs, which all contribute to its activity (Lanz et al., 2002). Overexpression of SRA enhances transcription transactivation by steroid receptors. Conversely, knockdown of SRA by siRNA suppressed androgen receptor activity in prostate cancer cells (Kurisu et al., 2006). The SRA precursor is subjected to an alternative splicing which generates an mRNA that codes for SRAP (SRA protein) (Hube et al., 2006). In contrast with SRA RNA, SRAP overexpression represses oestrogen receptor activity. The proportions of SRA and SRAP RNAs vary within breast cancer cell lines and might contribute to their phenotypes.

Little is known about the mechanism of co-activation by SRA. MAPK (mitogen-activated protein kinase) activation and phosphorylation of a specific serine in the oestrogen receptor ($ER\alpha$) is required for SRA co-activation (Deblois and Giguère, 2003). SRA RNA co-purifies with SRC-1 (steroid receptor co-activator complex protein 1) (Lanz et al., 1999). SRC-1 is not an RNA-binding protein, but SRA RNA associates the DEAD box p68/p72 RNA helicase that binds SRC-1 (Watanabe et al., 2001; Carretti et al., 2007). The p68/p72 RNA helicase acts as an oestrogen receptor co-activator.

Several SRA-binding proteins have been identified and are involved in steroid receptor co-regulation. Two steroid receptor co-repressors have also been shown to associate with SRA, SHARP (SMRT1/HDAC1-associated repressor protein) (Shi et al., 2001) and SLIRP (SRA stem-loop interacting protein) (Hatchell et al., 2006). SHARP and SLIRP appear to compete with the co-activator complex for SRA binding. This competition might contribute to the co-repressing effect of SHARP and SLIRP. As SHARP expression is hormone inducible, it appears to be part of a negative feedback loop to attenuate the hormone response.

Pseudouridylation (isomerization of specific uridine residues) of SRA adds an additional level of complexity in the regulation of steroid receptor activity. The Pus1p and Pus3p pseudo-uridine synthases bind

nuclear receptors and act as co-activators (Zhao et al., 2004, 2007). Pus1p and Pus3p pseudo-uridylate both common and specific nucleotide residues on SRA. Although SRA tolerates large deletions, replacement of U206 (a common Pus1p and Pus3p target) by adenosine switches SRA from a co-activator to a dominant-negative co-repressor, thus suggesting an important function for pseudo-uridylation.

Several other less-characterized examples of ncRNAs regulating specific transcription factors have been reported (Kuwabara et al., 2004; Willingham et al., 2005; Feng et al., 2006; Shamovsky et al., 2006). It is likely that a large number of ncRNAs behave as transcription co-regulators and contribute to the fine tuning of gene expression.

Recruiting or interfering with recruitment of chromatin-modifying machinery

X chromosome inactivation is a fascinating paradigm for long non-coding transcripts involved in epigenetic regulation. In mammalian female cells, one X chromosome is inactivated to keep a single copy active, as found in male cells. The inactivated chromosome is coated with the Xist ncRNA, deposited in the vicinity of its transcription site. The Xist RNA signals epigenetic histone modifications associated with gene silencing (Heard and Distèche, 2006). Tsix (Xist antisense), a non-coding antisense transcript of Xist, and Xist expression are mutually exclusive on a given chromosome. Tsix expression protects the active X chromosome from inactivation. As an initial step in X chromosome inactivation, the transient recruitment of the PcG (Polycomb group) complexes PRC2 (Polycomb repressive complex 2) and PRC1 on the inactivated chromosome follows and requires the appearance of the Xist RNA (Heard and Distèche, 2006). The PcG proteins assemble into three different classes of complexes involved in epigenetic gene silencing. PRC methylates histone H3 on Lys²⁷. This marks recognition for PRC1, which is presumed to mono-ubiquitinate histone H2A on Lys¹¹⁹. A third complex, PhoRC, binds to mono- and di-methylated Lys⁹ and Lys²⁰ on histones H3 and H4 respectively.

Silencing of homeotic genes in the human *HOX* loci also involves the PcG complex, as well as ncRNAs (Rinn et al., 2007). Indeed HOTAIR (Hox transcript antisense intergenic RNA) is a 2.2 kb non-coding transcript that interacts with PRC2 and is required for PRC2 recruitment to the *HOXD* locus

and subsequent trimethylation of histone H3 on Lys²⁷. The chromatin reorganization generated by the PcG proteins might also involve the RNAi machinery (Schuettengruber et al., 2007). In both cases mentioned above, ncRNAs might serve as recruiting agents for the enzymatic machinery installing chromatin repression.

However, epigenetic marking by ncRNAs does not necessarily lead to gene silencing. For example, *Drosophila* X chromosome dosage compensation involves stimulation of gene activity in male cells. This process is mediated by the roX ncRNAs, the NuRF (nucleosome remodelling factor) chromatin-remodelling complex and the MSL (male-specific lethal) complex (Deng and Meller, 2006; Bai et al., 2007). The roXs target the MSL complex specifically to genes along the male X chromosome, resulting in acetylation of histone H4 on Lys¹⁶, leading to an approx. 2-fold increase in the transcription of X-linked genes in males (X) relative to females (XX), thereby compensating for the difference in X chromosome number between the sexes.

RNA polymerase I is also subjected to ncRNA regulation (Mayer et al., 2006). Eukaryotic cells contain multiple tandem copies of rRNA genes, a significant fraction of them is repressed in heterochromatin. The key factor is a non-coding transcript originating from the IGS that separates rRNA genes and contains the promoters. This transcript (150–300 nt) is complementary to the DNA promoter sequence and binds TIP5 (transcription termination factor-I-interacting peptide 5), the largest subunit of NoRC (nucleolar chromatin-remodelling complex). RNA binding stabilizes NoRC association with rDNA promoters, and serves as a platform to recruit repressive histone modification and DNA methylation machineries.

In these three examples, the ncRNA marks the target DNA sequence and recruits protein complexes involved in silencing. Several other ncRNAs associated with dosage compensation have been reported and involve similar mechanisms (Yang and Kuroda, 2007; Zaratiegui et al., 2007).

Concluding remarks

A growing number of nuclear ncRNAs are associated with human disease (Cao et al., 2006; Mehler and Mattick, 2006; Mehler and Mattick, 2007). The increasing variety of ncRNAs being identified suggests a strong connection between complexity of organ-

isms, biogenesis, dynamics of action and the combinatorial regulatory potential of ncRNAs.

Acknowledgements

We are indebted to Xavier Darzacq for fruitful discussions. Work in the authors' laboratory is supported by grants from the Association pour la Recherche sur le Cancer, Agence Nationale de Recherche sur le SIDA, Agence Nationale pour la Recherche and Fondation pour la Recherche Médicale.

References

- *Articles of special interest
- Allen, T.A., Von Kaenel, S., Goodrich, J.A. and Kugel, J.F. (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol. Biol.* **11**, 816–821
- Anand, K., Schulte, A., Fujinaga, K., Scheffzek, K. and Geyer, M. (2007) Cyclin box structure of the P-TEFb subunit cyclin t1 derived from a fusion complex with ElAV Tat. *J. Mol. Biol.* **370**, 826–836
- Andersen, G., Busso, D., Poterszman, A., Hwang, J.R., Wurtz, J.M., Ripp, R., Thierry, J.C., Egly, J.M. and Moras, D. (1997) The structure of cyclin H: common mode of kinase activation and specific features. *EMBO J.* **16**, 958–967
- Awasthi, S. and Alwine, J.C. (2003) Association of polyadenylation cleavage factor I with U1 snRNP. *RNA* **9**, 1400–1409
- Bai, X., Larschan, E.N., Kwon, S.Y., Badenhorst, P. and Kuroda, M.I. (2007) Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in *Drosophila melanogaster*. *Genetics* **176**, 1491–1499
- *Bannwarth, S. and Gagnon, A. (2005) HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. *Curr. HIV Res.* **3**, 61–71
- *Barboric, M., Kohoutek, J., Price, J.P., Blazek, D., Price, D.H. and Peterlin, B.M. (2005) Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb. *EMBO J.* **24**, 4291–4303
- Barboric, M., Yik, J.H., Czudnochowski, N., Yang, Z., Chen, R., Contreras, X., Geyer, M., Peterlin, M.B. and Zhou, Q. (2007) Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. *Nucleic Acids Res.* **35**, 2003–2012
- *Barrandon, C., Bonnet, F., Nguyen, V.T., Labas, V. and Bensaude, O. (2007) The transcription-dependent dissociation of P-TEFb–HEXIM1–7SK RNA relies upon formation of hnRNP–7SK RNA complexes. *Mol. Cell. Biol.* **27**, 6996–7006
- Barrick, J.E., Sudarsan, N., Weinberg, Z., Ruzzo, W.L. and Breaker, R.R. (2005) 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* **11**, 774–784
- *Bentley, D.L. (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* **17**, 251–256
- Biglione, S., Byers, S.A., Price, J.P., Nguyen, V.T., Bensaude, O., Price, D.H. and Maury, W. (2007) Inhibition of HIV-1 replication by P-TEFb inhibitors DRB, seliciclib and flavopiridol correlates with release of free P-TEFb from the large, inactive form of the complex. *Retrovirology* **4**, 47
- Bomsztyk, K., Denisenko, O. and Ostrowski, J. (2004) hnRNP K: one protein multiple processes. *Bioessays* **26**, 629–638
- Bres, V., Tagami, H., Peloponese, J.M., Loret, E., Jeang, K.T., Nakatani, Y., Emiliani, S., Benkirane, M. and Kiernan, R.E. (2002) Differential acetylation of Tat coordinates its interaction with the co-activators cyclin T1 and PCAF. *EMBO J.* **21**, 6811–6819
- Bushati, N. and Cohen, S.M. (2007) microRNA Functions. *Annu. Rev. Cell Dev. Biol.* **23**, 175–205

- Byers, S.A., Price, J.P., Cooper, J.J., Li, Q. and Price, D.H. (2005) HEXIM2, A HEXIM1 related protein, regulates P-TEFb through association with 7SK. *J. Biol. Chem.* **280**, 16360–16367
- Cao, X., Yeo, G., Muotri, A.R., Kuwabara, T. and Gage, F.H. (2006) Noncoding RNAs in the mammalian central nervous system. *Annu. Rev. Neurosci.* **29**, 77–103
- Caretti, G., Lei, E.P. and Sartorelli, V. (2007) The DEAD-box p68/p72 proteins and the noncoding RNA steroid receptor activator SRA: eclectic regulators of disparate biological functions. *Cell Cycle* **6**, 1172–1176
- Cassé, C., Giannoni, F., Nguyen, V.T., Dubois, M.-F. and Bensaude, O. (1999) The transcriptional inhibitors, actinomycin D and α -amanitin, activate the HIV-1 promoter and favor phosphorylation of the RNA polymerase II C-terminal domain. *J. Biol. Chem.*, **274**, 16097–16106
- Dames, S.A., Schonichen, A., Schulte, A., Barboric, M., Peterlin, B.M., Grzesiek, S. and Geyer, M. (2007) Structure of the Cyclin T binding domain of Hexim1 and molecular basis for its recognition of P-TEFb. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14312–14317
- Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P. and Reed, R. (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* **26**, 867–881
- Deblois, G. and Giguère, V. (2003) Ligand-independent coactivation of ER α AF-1 by steroid receptor RNA activator (SRA) via MAPK activation. *J. Steroid Biochem. Mol. Biol.* **85**, 123–131
- Deng, X. and Meller, V.H. (2006) roX RNAs are required for increased expression of X-linked genes in *Drosophila melanogaster* males. *Genetics* **174**, 1859–1866
- Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* **3**, 195–205
- Dubois, M.-F., Bellier, S., Seo, S.-J. and Bensaude, O. (1994) Phosphorylation of the RNA polymerase II largest subunit during heat-shock and inhibition of transcription in HeLa cells. *J. Cell Physiol.* **158**, 417–426
- Dulac, C., A.A., M., Fraldi, A., Bonnet, F., Nguyen, V.T., Napolitano, G., Lania, L. and Bensaude, O. (2005) Transcription-dependent association of multiple P-TEFb units to a HEXIM multimer. *J. Biol. Chem.* **280**, 30619–30629
- *Egloff, S., Van Herreweghe, E. and Kiss, T.** (2006) Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. *Mol. Cell Biol.* **26**, 630–642
- *Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F. and Goodrich, J.A.** (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat. Struct. Mol. Biol.* **11**, 822–829
- Espinoza, C.A., Goodrich, J.A. and Kugel, J.F. (2007) Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA* **13**, 583–596
- Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P. and Kohtz, J.D. (2006) The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470–1484
- Fornace, A.J., Alamo, I., Hollander, M.C. and Lamoreaux, E. (1989) Induction of heat shock protein transcripts and B2 transcripts by various stresses in Chinese hamster cells. *Exp. Cell Res.* **182**, 61–74
- Fraldi, A., Varrone, F., Napolitano, G., Michels, A.A., Majello, B., Bensaude, O. and Lania, L. (2005) Inhibition of Tat activity by the HEXIM1 protein. *Retrovirology* **2**, 42
- Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T. and Peterlin, B.M. (2004) Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell Biol.* **24**, 787–795
- Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A. and Proudfoot, N.J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev.* **16**, 2792–2799
- *Garber, M.E., Wei, P., KewalRamani, V.N., Mayall, T.P., Herrmann, C.H., Rice, A.P., Littman, D.R. and Jones, K.A.** (1998) The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev.* **12**, 3512–3527
- Gildehaus, N., Neusser, T., Wurm, R. and Wagner, R. (2007) Studies on the function of the riboregulator 6S RNA from *E. coli*: RNA polymerase binding, inhibition of *in vitro* transcription and synthesis of RNA-directed *de novo* transcripts. *Nucleic Acids Res.* **35**, 1885–1896
- Gupta, S., Busch, R.K., Singh, R. and Reddy, R. (1990) Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of γ -monomethyl-GTP cap structure in 7SK and several other human small RNAs. *J. Biol. Chem.* **265**, 19137–19142
- Gürsoy, H.C., Koper, D. and Benecke, B.-J. (2000) The vertebrate 7S K RNA separates hagfish (*Myxine glutinosa*) and lamprey (*Lampetra fluviatilis*). *J. Mol. Evol.* **50**, 456–464
- Hatchell, E.C., Colley, S.M., Beveridge, D.J., Epis, M.R., Stuart, L.M., Giles, K.M., Redfern, A.D., Miles, L.E., Barker, A., MacDonald, L.M., Arthur, P.G., Lui, J.C., Golding, J., McCulloch, R.K., Metcalf, C.B., Wilce, J.A., Wilce, M.C., Lanz, R.B., O'Malley, B.W. and Leedman, P.J. (2006) SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. *Mol. Cell* **22**, 657–668
- He, N., Pezda, A.C. and Zhou, Q. (2006) Modulation of a P-TEFb functional equilibrium for the global control of cell growth and differentiation. *Mol. Cell Biol.* **26**, 7068–7076
- Heard, E. and Disteché, C.M. (2006) Dosage compensation in mammals: fine tuning the expression of the X chromosome. *Genes Dev.* **20**, 1848–1867
- Hindley, J. (1967) Fractionation of ³²P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *J. Mol. Biol.* **30**, 125–136
- Hogg, J.R. and Collins, K. (2007) RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA* **13**, 868–880
- Hsu, L.M., Zagorski, J., Wang, Z. and Fournier, M.J. (1985) *Escherichia coli* 6S RNA gene is part of a dual-function transcription unit. *J. Bacteriol.* **161**, 1162–1170
- Hube, F., Guo, J., Chooniedass-Kothari, S., Cooper, C., Hamedani, M.K., Dibrov, A.A., Blanchard, A.A., Wang, X., Deng, G., Myal, Y. and Leygue, E. (2006) Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines. *DNA Cell Biol.* **25**, 418–428
- Imamichi, T., Conrads, T.P., Zhou, M., Liu, X., Adelsberger, J.W., Veenstra, T.D. and Lane, H.C. (2005) A transcription inhibitor, actinomycin D, enhances HIV-1 replication through an interleukin-6-dependent pathway. *J. Acquir. Immune Defic. Syndr.* **40**, 388–397
- *Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J., Chabot, B., Poirier, G.G., Hughes, T.R., Blanchette, M., Price, D.H. and Coulombe, B.** (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell* **27**, 262–274
- Kaehlecke, K., Dorr, A., Hetzer-Egger, C., Kiermer, V., Henklein, P., Schnoelzer, M., Loret, E., Cole, P.A., Verdin, E. and Ott, M. (2003) Acetylation of Tat defines a cyclin T1-independent step in HIV transactivation. *Mol. Cell* **12**, 167–176
- Kanin, E.I., Kipp, R.T., Kung, C., Slattery, M., Viale, A., Hahn, S., Shokat, K.M. and Ansari, A.Z. (2007) Chemical inhibition of the TFIIF-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5812–5817

- *Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H. and Gingeras, T.R. (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488
- *Kettenberger, H., Eisenfuhr, A., Brueckner, F., Theis, M., Famulok, M. and Cramer, P. (2006) Structure of an RNA polymerase II-RNA inhibitor complex elucidates transcription regulation by noncoding RNAs. *Nat. Struct. Mol. Biol.* **13**, 44–48
- Kim, K.S. and Lee, Y. (2004) Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res.* **32**, 6057–6068
- Kurusu, T., Tanaka, T., Ishii, J., Matsumura, K., Sugimura, K., Nakatani, T. and Kawashima, H. (2006) Expression and function of human steroid receptor RNA activator in prostate cancer cells: role of endogenous hSRA protein in androgen receptor-mediated transcription. *Prostate Cancer Prostatic Dis.* **9**, 173–178
- Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K. and Gage, F.H. (2004) A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* **116**, 779–793
- *Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O’Gorman, W., Kimura, H., Proudfoot, N.J. and Akoulitchev, A. (2002) U1 snRNA associates with TFIH and regulates transcriptional initiation. *Nat. Struct. Biol.* **9**, 800–805
- Lacadie, S.A. and Rosbash, M. (2005) Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5’s base pairing in yeast. *Mol. Cell* **19**, 65–75
- *Lanz, R.B., McKenna, N.J., Onate, S.A., Albrecht, U., Wong, J., Tsai, S.Y., Tsai, M.-J. and O’Malley, B.W. (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17–27
- Lanz, R.B., Razani, B., Goldberg, A.D. and O’Malley, B.W. (2002) Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16081–16086
- Li, Q., Price, J.P., Byers, S.A., Cheng, D., Peng, J. and Price, D.H. (2005) Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *J. Biol. Chem.* **280**, 28819–28826
- Li, Q., Cooper, J.J., Altwerger, G.H., Feldkamp, M.D., Shea, M.A. and Price, D.H. (2007) HEXIM1 is a promiscuous double-stranded RNA-binding protein and interacts with RNAs in addition to 7SK in cultured cells. *Nucleic Acids Res.* **35**, 2503–2512
- Listerman, I., Sapra, A.K. and Neugebauer, K.M. (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* **13**, 815–822
- Mancebo, H.S., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D. and Flores, O. (1997) P-TEFb kinase is required for HIV Tat transcriptional activation *in vivo* and *in vitro*. *Genes Dev.* **11**, 2633–2644
- Maraia, R.J. and Bayfield, M.A. (2006) The La protein–RNA complex. *Mol. Cell* **21**, 149–152
- *Martianov, I., Ramadass, A., SerraBarros, A., Chow, N. and Akoulitchev, A. (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, **445**, 666–670
- *Mayer, C., Schmitz, K.M., Li, J., Grummt, I. and Santoro, R. (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol. Cell* **22**, 351–361
- McKenna, N.J. and O’Malley, B.W. (2002) Combinational control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465–474
- Mehler, M.F. and Mattick, J.S. (2006) Non-coding RNAs in the nervous system. *J. Physiol.* **575**, 333–341
- Mehler, M.F. and Mattick, J.S. (2007) Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol. Rev.* **87**, 799–823
- Michels, A.A., Nguyen, V.T., Fraldi, A., Labas, V., Edwards, M., Bonnet, F., Lania, L. and Bensaude, O. (2003) MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol. Cell. Biol.* **23**, 4859–4869
- *Michels, A.A., Fraldi, A., Li, Q., Adamson, T.E., Bonnet, F., Nguyen, V.T., Sedore, S.C., Price, J.P., Price, D.H., Lania, L. and Bensaude, O. (2004) Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J.* **23**, 2608–2619
- *Nguyen, V.T., Kiss, T., Michels, A.A. and Bensaude, O. (2001) 7SK snRNA binds to and inhibits the activity of Cdk9/cyclin T complexes. *Nature* **414**, 322–325
- Nilsen, T.W. (2005) Spliceosome assembly in yeast: one ChIP at a time? *Nat. Struct. Mol. Biol.* **12**, 571–573
- O’Gorman, W., Thomas, B., Kwek, K.Y., Furger, A. and Akoulitchev, A. (2005) Analysis of U1 small nuclear RNA interaction with cyclin H. *J. Biol. Chem.* **280**, 36920–36925
- *Peterlin, B.M. and Price, D.H. (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305
- Phatnani, H.P. and Greenleaf, A.L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936
- Ping, Y.H. and Rana, T.M. (2001) DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation. *J. Biol. Chem.* **276**, 12951–12958
- *Prasanth, K.V. and Spector, D.L. (2007) Eukaryotic regulatory RNAs: an answer to the ‘genome complexity’ conundrum. *Genes Dev.* **21**, 11–42
- *Rana, T.M. (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell. Biol.* **8**, 23–36
- Rappas, M., Bose, D. and Zhang, X. (2007) Bacterial enhancer-binding proteins: unlocking sigma54-dependent gene transcription. *Curr. Opin. Struct. Biol.* **17**, 110–116
- Reitzer, L. and Schneider, B.L. (2001) Metabolic context and possible physiological themes of sigma(54)-dependent genes in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **65**, 22–44
- *Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugnann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. and Chang, H.Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323
- *Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. and Cavalli, G. (2007) Genome regulation by Polycomb and trithorax proteins. *Cell* **128**, 735–745
- Schulte, A., Czudnochowski, N., Barboric, M., Schonichen, A., Blazek, D., Peterlin, B.M. and Geyer, M. (2005) Identification of a cyclin T-binding domain in Hexim1 and biochemical analysis of its binding competition with HIV-1 Tat. *J. Biol. Chem.* **280**, 24968–24977
- Sedore, S.C., Byers, S.A., Biglione, S., Price, J.P., Maury, W.J. and Price, D.H. (2007) Manipulation of P-TEFb control machinery by HIV: recruitment of P-TEFb from the large form by Tat and binding of HEXIM1 to TAR. *Nucleic Acids Res.* **35**, 4347–4358
- Shamovsky, I., Ivannikov, M., Kandel, E.S., Gershon, D. and Nudler, E. (2006) RNA-mediated response to heat shock in mammalian cells. *Nature* **440**, 556–560
- Shi, Y., Downes, M., Xie, W., Kao, H.Y., Ordentlich, P., Tsai, C.-C., Hon, M. and Evans, R.M. (2001) Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev.* **15**, 1140–1151

- *Shi, H., Fan, X., Sevilimedu, A. and Lis, J.T. (2007) RNA aptamers directed to discrete functional sites on a single protein structural domain. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3742–3746
- Shimizu, S., Urano, E., Futahashi, Y., Miyauchi, K., Isogai, M., Matsuda, Z., Nohtomi, K., Onogi, T., Takebe, Y., Yamamoto, N. and Komano, J. (2007) Inhibiting lentiviral replication by HEXIM1, a cellular negative regulator of the CDK9/cyclin T complex. *AIDS*, **21**, 575–582
- *Thomas, M., Chédin, S., Carles, C., Riva, M., Famulok, M. and Sentenac, A. (1997) Selective targeting and inhibition of yeast RNA polymerase II by RNA aptamers. *J. Biol. Chem.* **272**, 27980–27986
- Trotochaud, A.E. and Wassarman, K.M. (2004) 6S RNA function enhances long-term cell survival. *J. Bacteriol.* **186**, 4978–4985
- Trotochaud, A.E. and Wassarman, K.M. (2005) A highly conserved 6S RNA structure is required for regulation of transcription. *Nat. Struct. Mol. Biol.* **12**, 313–319
- Trotochaud, A.E. and Wassarman, K.M. (2006) 6S RNA regulation of *pspF* transcription leads to altered cell survival at high pH. *J. Bacteriol.* **188**, 3936–3943
- Valerie, K.A., Delers, Bruck, C., Thiriart, H., Rosenberg, C., Debouck, C. and Rosenberg, M. (1988) Activation of human immunodeficiency virus type 1 by DNA damage in human cells. *Nature* **333**, 78–81
- *Van Herreweghe, E., Egloff, S., Goiffon, I., Jady, B.E., Froment, C., Monsarrat, B. and Kiss, T. (2007) Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *EMBO J.* **26**, 3570–3580
- Wassarman, D.A. and Steitz, J.A. (1991) Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol. Cell. Biol.* **11**, 3432–3445
- *Wassarman, K.M. and Saecker, R.M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* **314**, 1601–1603
- *Wassarman, K.M. and Storz, G. (2000) 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* **101**, 613–623
- Watanabe, M., Yanagisawa, J., Kitagawa, H., Takeyama, K., Ogawa, S., Arai, Y., Suzawa, M., Kobayashi, Y., Yano, T., Yoshikawa, H., Masuhiro, Y. and Kato, S. (2001) A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor α coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO J.* **15**, 1341–1352
- Weinberger, L.S. and Shenk, T. (2007) An HIV feedback resistor: auto-regulatory circuit deactivator and noise buffer. *PLoS Biol.* **5**, 67–81
- Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B. and Schultz, P.G. (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309**, 1570–1573
- Willkomm, D.K., Minnerup, J., Huttenhofer, A. and Hartmann, R.K. (2005) Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog. *Nucleic Acids Res.* **33**, 1949–1960
- Yang, P.K. and Kuroda, M.I. (2007) Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell* **128**, 777–786
- *Yang, Z., Zhu, Q., Luo, K. and Zhou, Q. (2001) The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**, 317–322
- Yik, J.H., Chen, R., Nishimura, R., Jennings, J.L., Link, A.J. and Zhou, Q. (2003) Inhibition of P-TEFb (CDK9/cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol. Cell* **12**, 971–982
- Yik, J.H., Chen, R., Pezda, A.C., Samford, C.S. and Zhou, Q. (2004) A human immunodeficiency virus type 1 Tat-like arginine-rich RNA-binding domain is essential for HEXIM1 to inhibit RNA polymerase II transcription through 7SK snRNA-mediated inactivation of P-TEFb. *Mol. Cell. Biol.* **24**, 5094–5105
- Yik, J.H., Chen, R., Pezda, A.C. and Zhou, Q. (2005) Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive P-TEFb complexes for control of transcription. *J. Biol. Chem.* **280**, 16368–16376
- *Zaratiegui, M., Irvine, D.V. and Martienssen, R.A. (2007) Noncoding RNAs and gene silencing. *Cell* **128**, 763–776
- Zhang, J., Tamilarasu, N., Hwang, S., Garber, M.E., Huq, I., Jones, K.A. and Rana, T.M. (2000) HIV-1 TAR RNA enhances the interaction between Tat and cyclin T1. *J. Biol. Chem.* **275**, 34314–34319
- *Zhao, Y. and Srivastava, D. (2007) A developmental view of microRNA function. *Trends Biochem. Sci.* **32**, 189–197
- Zhao, X., Patton, J.R., Davis, S.L., Florence, B., Ames, S.J. and Spanjaard, R.A. (2004) Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol. Cell* **27**, 549–558
- Zhao, X., Patton, J.R., Ghosh, S.K., Fischel-Ghodsian, N., Shen, L. and Spanjaard, R.A. (2007) Pus3p- and Pus1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling. *Mol. Endocrinol.* **21**, 686–699
- Zhou, M., Halanski, M.A., Radonovich, M.F., Kashanchi, F., Peng, J., Price, D.H. and Brady, J.N. (2000) Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. *Mol. Cell. Biol.* **20**, 5077–5086
- Zhou, M., Deng, L., Kashanchi, F., Brady, J.N., Shatkin, A.J. and Kumar, A. (2003) The Tat/TAR-dependent phosphorylation of RNA polymerase II C-terminal domain stimulates cotranscriptional capping of HIV-1 mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12666–12671
- Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M.B. and Price, D.H. (1997) Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation *in vitro*. *Genes Dev.* **11**, 2622–2632
- Zurita, M. and Merino, C. (2003) The transcriptional complexity of the TFIID complex. *Trends Genet.* **19**, 578–584

Received 25 July 2007/1 October 2007; accepted 4 October 2007

Published on the Internet 21 January 2008, doi:10.1042/BC20070090

Résumé

Interaction du snARN U1 de l'épissage avec l'ARN polymérase II.

Les ARNs non codants sont des régulateurs de l'expression génétique à plusieurs niveaux. Chez la bactérie et chez la souris, des ARNs non codants (6S et B2) ont la propriété de se lier à l'ARN polymérase et d'inhiber son activité. Afin de déterminer si l'ARN polymérase II (RNAPII) humaine était associée à des ARNs non codants, une immunoprécipitation anti-RNAPII a été réalisée sur des cellules HeLa mitotiques. Les ARNs co-immunoprécipités ont été purifiés et marqués et l'ARN U1 s'est trouvé particulièrement enrichi par rapport au contrôle. Cette co-immunoprécipitation reflète l'association de la snRNP U1 avec la RNAPII. Pour vérifier cette association sur un site de transcription actif, des lignées ont été établies avec l'insertion en multiples copies d'un gène à un site unique, créant ainsi un unique super site de transcription visualisable par FISH (Fluorescence In Situ Hybridization). Deux lignées distinctes ont été créées, l'une avec un gène comportant un intron, l'autre avec le même gène où l'intron comporte trois mutations ponctuelles abolissant l'épissage. Alors que les snARNs U2, U4, U5 et U6 sont absents du site non épissé, l'ARN U1 est enrichi de la même façon indépendamment de l'épissage. La présence des protéines spécifiques de la snRNP U1 indique que la snRNP U1 est recrutée au complet au site de transcription. Ces résultats laissent supposer un rôle pour l'association RNAPII – U1snRNP dans l'épissage cotranscriptionnel.

Mot-clefs : ARN polymérase II, snRNP U1, épissage, FISH, mitose.

Summary

U1 snRNA spliceosomal component interaction with RNA polymerase II.

Non-coding RNAs are multi-step genetic expression regulators. Bacterial 6S RNA and murine B2 RNA are non-coding RNAs able to bind and inhibit RNA polymerase II (RNAPII). Seeking for non-coding RNAs bound to human RNAPII, immunoprecipitations against RNAPII were carried on mitotic HeLa cells. Co-immunoprecipitated RNAs were purified and labelled and U1 RNA was found to be peculiarly enriched versus the control IP. This co-immunoprecipitation reflects the association of the U1 snRNP with RNAPII. In order to verify this association on an active transcription site, stable cell lines were established with multiple copies insertion of a gene at one single site, allowing its visualisation by FISH (Fluorescent In Situ Hybridization). Two distinct cell lines were made with two similar constructs only differing by three point mutations abolishing the splicing in one construct. Whereas 4 snRNAs were absent on a non splicing transcription site, U1 snRNA is recruited both on spliced and non-spliced transcription sites. The whole U1snRNP seems to be recruited independently of splicing because of the presence of U1 specific proteins. These results may suggest a role for the RNAPII – U1 snRNP interaction in cotranscriptional splicing.

Key words: RNA polymerase II, U1 snRNP, splicing, FISH, mitosis.