Régulation tissulaire de l'épissage alternatif : Caractérisation fonctionnelle d'une séquence activatrice de la maturation d'un exon 3' terminal

Vincent Anquetil

▶ **To cite this version:**

Vincent Anquetil. Régulation tissulaire de l'épissage alternatif : Caractérisation fonctionnelle d'une séquence activatrice de la maturation d'un exon 3' terminal. Biochimie [q-bio.BM]. Université Rennes 1, 2009. Français. NNT : . tel-00462347

HAL Id: tel-00462347
https://theses.hal.science/tel-00462347

Submitted on 9 Mar 2010

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.

N° d'ordre : 3960 **ANNÉE 2009**

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : BIOLOGIE

Ecole doctorale Vie-Agronomie-Santé

présentée par

Vincent Anquetil

préparée à l'unité de recherche UMR 6061 CNRS, IGDR
Equipe EGD de l'Institut de Génétique et Développement de Rennes
U.F.R. Sciences de la Vie et de l'Environnement

Régulation tissulaire de l'épissage alternatif : Caractérisation fonctionnelle d'une séquence activatrice de la maturation d'un exon 3' terminal

Thèse soutenue à Rennes
le 02 octobre 2009

devant le jury composé de :

Dr Laurent CORCOS
Directeur de Recherche INSERM
INSERM U613, Brest */rapporteur*

Dr Annie MOUGIN
Chargée de Recherche CNRS
UMR 5099 CNRS, Toulouse /rapporteur

Dr Cyril BOURGEOIS
Chargé de Recherche INSERM
IBPM, Rome Italie */examinateur*

Pr Denis MICHEL
Professeur à l'Université de Rennes 1
UMR 6026 CNRS, Rennes */examinateur*

Pr Serge HARDY
Professeur à l'Université de Rennes 1
UMR 6061 CNRS, Rennes */directeur de thèse*

Dr Agnès MEREAU
Chargée de Recherche CNRS
UMR 6061 CNRS, Rennes */co-directeur de thèse*

Résumé

La maturation des ARN pré-messagers est le fruit d'un ensemble de processus nucléaires interconnectés qui sont soumis à de nombreuses régulations. L'épissage alternatif des exons 3' terminaux joue un rôle majeur dans l'expression des gènes car il permet de réguler qualitativement et quantitativement leur expression.

Nous étudions les déterminants de la régulation tissulaire de l'épissage et de la polyadénylation en utilisant comme modèle le gène de la tropomyosine α de xénope. Ce gène contient, dans sa région 3' terminale, un exon composite interne/3' terminal nommé 9A9'qui est utilisé comme exon 3' terminal dans les somites et est sauté dans les tissus non musculaires dans l'embryon de xénope. L'utilisation de minigènes contenant une portion de la région 3' terminale du gène de la tropomyosine α placée sous le contrôle de promoteurs tissu-spécifiques a permis d'identifier deux éléments introniques régulant l'utilisation de l'exon 9A9'. L'un nommé 150PY est répresseur, l'autre appelé UTE est activateur. L'élément 150PY réprime l'utilisation de l'exon 9A9' dans les cellules non musculaires. Des approches biochimiques et *in vivo* ont montré que la protéine PTB se fixe sur cet élément et inhibe les réactions d'épissage et de clivage/polyadénylation de l'exon 9A9'.

Afin de caractériser la fonction de l'élément activateur UTE, nous avons bloqué son accessibilité dans les embryons à l'aide d'oligonucléotides morpholinos antisens. Nos résultats montrent que l'UTE active l'utilisation de l'exon 9A9' en tant qu'exon 3' terminal en favorisant la reconnaissance du site 3' d'épissage, du signal de polyadénylation et du point de branchement. Ces résultats suggèrent que l'UTE favorise la fixation de la snRNPU2 sur le point de branchement qui à son tour stabilise la liaison du complexe de clivage/polyadénylation sur le signal de polyadénylation permettant ainsi la définition de l'exon 9A9' en tant qu'exon terminal. La PTB prévient l'utilisation de l'UTE dans les cellules non musculaires à l'inverse de certaines protéines SR qui favorisent la sélection de l'exon 9A9' de manière dépendante de cette séquence. Pour caractériser les mécanismes moléculaires impliqués dans la fonction de l'UTE nous avons recherché les facteurs recrutés sur cette séquence. Ces expériences montrent que la protéine ESRP2 est spécifiquement recrutée sur un ARN contenant l'UTE en présence de PTB. ESRP2 est exprimée uniquement dans les cellules épithéliales et pourrait participer avec la PTB à la spécificité tissulaire de la répression de l'exon 9A9'.

L'ensemble de ces résultats suggèrent que la régulation tissulaire de l'exon 9A9' est basée sur une compétition de fixation entre des facteurs activateurs et inhibiteurs sur la séquence régulatrice UTE.

MERCISSS

Je tiens avant tout à remercier les Docteurs Annie Mougin, Laurent Corcos, et Cyril Bourgeois, ainsi que le Professeur Denis Michel, pour avoir accepté d'évaluer ce travail de thèse.

Après quatre années passées dans un laboratoire de recherche il y a tant de choses à dire, et tant de personnes à remercier ! Pour n'oublier personne, je remercie tous ceux qui ont participé, de prêt ou de loin, à la réalisation de ce travail.

Essayons pour une fois de faire les choses dans l'ordre. Serge et Agnès m'ont accueillis, m'ont transmis une partie de leur savoir, m'ont conseillé et m'ont accompagné pendant tout ce temps, pour tout ça, un grand merci. Un merci spécial à Agnès qui m'a permis d'appréhender la science dans et hors du laboratoire, qui a été présente quand il le fallait, et qui, jusqu'à la fin, aura été disponible.

La thèse est une expérience unique, certes scientifique, mais aussi humaine. Les membres de l'équipe EGD (oui cette équipe n'a plus qu'un seul nom depuis quelques mois maintenant) sont la preuve que le travail dans la joie et la bonne humeur (tout en restant sérieux) facilite grandement le quotidien de chacun. Merci donc à Bev, Luc, Carole, Yann, Vincent, Hallen, Hubert, Bernhard, Cécile, Sylvain, Serge, Agnès et à mes deux compères de thèse Olivier et Marie, ainsi qu'au troisième mousquetaire : Tanguy. Certains sont partis mais méritent un grand merci, je pense à Michelle, Dominique, Caroline, Carine, Antoine et Bertrand. Parmi toutes les personnes travaillant dans l'unité certaines sont devenues et resterons beaucoup plus que des collègues de travail, comment imaginer les pauses sans Tanguy, Marie, Olivier, Lucie, Audrey ou Stéphane. Les amis, merci pour votre soutien, vos encouragements et tous les bons moments, parfois scientifiques, que nous avons passé ensemble ! D'autres personnes rencontrées à la fac, ou un peu plus tard au cours de mes premiers stages ont participé de manière très active aux bons moments de ma vie de tous les jours, merci donc à Emilie, Yann, Damien, Xavier, Dorothée, Maël, Maud, Matthieu et Yoann sans qui rien n'aurait été pareil.

J'ai la chance d'avoir été très bien accueilli par la phénoménale Géraldine, dont l'efficacité n'a d'égale que la qualité de ses conseils et son attention. Je voudrais également remercier les habitués du week-end, qui font que le labo est moins vide et les cafés plus nombreux, merci Ariane, Franck, Najate et surtout et encore Stéphane.

La vie d'un doctorant n'est pas faite que d'expérimentations et de travail. J'ai la chance d'avoir une super bande de potes sans qui rien n'aurait été possible. Il est important de

noter que des personnes, initialement mes collègues, sont également mentionnées ici. Je pense à Lu et Tinou, à Marie, à Tanguy, à Olivier (dont le mariage a été excellent, ça n'a rien à faire ici, mais ça me fait plaisir de le dire), à Audrey, Ariane, Najate, mais aussi Maud, Matthieu, Yoann, Stéphane. Il y a les vieux amis, sur lesquels on peut toujours compter : Jeff, Emilie, Xavier, Dorothée, Mael, Morgane, Damien, Yann, Evon, Aurélie, Arnaud, Aline, Morgane, Mikael, etc… Nous nous sommes bien trouvé, nous ne nous quittons plus, et peu importe le temps passé sans se voir, les choses sont toujours aussi simples et agréables. Certains font comme ils peuvent de loin, merci Yannou pour tes coups de fils rassurants. D'autres font les choses presque au quotidien, merci Jeff et Morgane, je ne m'étendrai pas sur tout ce que vous avez fait pour moi, cette thèse est déjà suffisamment longue. Les derniers font également une thèse et nous nous posons parfois les mêmes questions, merci à Xavier (Xavsès, little daddy ou autres personnages) et à Dorothée d'avoir toujours été là quand il fallait comme il fallait !

Merci à tous ceux qui ont participé à la fin de cette thèse : Jeff (pour tout, un très gros tout), Morgane (pour tout aussi), Maël (pour plein de choses), Antoine et Lucie (pour le soutien, le manger, les figures et le reste), Matthieu (pour le soutien, les corrections, l'aide à la mise en page, le golf, etc.), Maud (pour les coups de pied aux fesses, et bien d'autres choses encore), Tanguy (pour les abréviations et tout le reste), Olivier (pour les figures, etc.), Marie (pour le soutien, etc ; aussi), Stéphane (pour beaucoup de choses également, n'oublie pas notre réunion du mois de novembre), Hubert (pour les corrections et conseils), Agnès (comme Hubert mais encore plus souvent), Serge (comme pour les deux précédents). A tous merci.

J'ai gardé le meilleur pour la fin. Rien de tout ça n'aurait été possible sans ma mère et ma sœur. Toutes deux nécessaires à tant de choses dans ma vie. Merci merci merci. Je n'en dirai pas plus car les mots ne sont pas assez forts pour décrire tout ce qu'il y a à dire à propos de vous deux !!!

MERCI à tous…

ABREVIATIONS

A :	Adénosine
aa :	acides aminés
ADN :	Acide DésoxyriboNucléique
ADNc :	Acide DésoxyriboNucléique Complémentaire
ARN :	Acide RiboNucléique
ARNm :	Acide RiboNucléique messager
ARNpol II :	ARN polymérase II
ARNpré-m :	Acide RiboNucléique pré-messager
ATP :	Adénosine TriPhosphate
C :	Cytidine
CBC :	*Cap Binding Complex*
CFIm :	*Cleavage Factor Im*
CFIIm :	*Cleavage Factor IIm*
ChIP :	*Chromatin ImmunoPrecipitation*
CPC :	*Cleavage and Polyadenylation Complex*
CPSF :	*Cleavage Polyadenylation Specificity Factor*
CstF :	*Cleavage stimulation Factor*
CT/CGRP :	*Calcitonin/Calcitonin-Gene-Related Peptide*
CTD :	*Carboxy-Terminal Domain of RNA polymerase II*
CUG-BP :	*CUG-Binding Protein*
DSE :	*Downstream Sequence Element*
dsRBP :	*double strand RNA binding protein*
ECF :	*Enhanced ChemiFluorescence*
ECL :	*Enhanced Chemiluminescence*
EJC :	*Exon Junction Complex*
EN :	Extrait Nucléaire
ESE :	*Exonic Splicing Enhancer*
ESS :	*Exonic Splicing Silencer*
EST :	*Expressed Sequence Tag*
FGFR2 :	*Fibroblast Growth Factor Receptor-2*
G :	Guanosine
hnRNP :	*heterogeneous nuclear RiboNucleoProtein*
HIV :	*Human Immunodeficiency Virus*
HSP :	*HeatShock Protein*
IgM :	Immunoglobuline M
ISE :	*Intronic Splicing Enhancer*
ISS :	*Intronic Splicing Silencer*
kb :	kilo paires de base
kDa :	kiloDalton
KH :	*hnRNP K Homology domain*
KO :	*Knock Out*

miARN :	microARN
min :	minute
NGD :	No Go Decay
ng :	nanogramme
NLS :	*Nuclear Localization Signal*
NMD :	*Nonsense-Mediated Decay*
nPTB :	*neuronal Polypyrimidine Tract Binding protein*
NSD :	Non Stop Decay
nt :	nucléotides
OH :	hydroxyl
PABP :	*Poly(A) Binding Protein*
PAP :	*Poly(A) Polymerase*
PASP :	*Pyrosequencing Analysis of Splicing Patterns*
pb :	paire de bases
PCR :	*Polymerase Chain Reaction*
poly(A) :	séquence de poly-adénosines
PPI	Pyrophosphate
PTB :	*Polypyrimidine Tract Binding protein*
RBD :	*RNA Binding Domain*
RBM :	*RNA Binding Motif*
RNA-BP :	*RiboNucleic Acid-Binding Protein*
RNase :	RiboNucléase
RNP :	RiboNucléoProtéique
RPA :	*Rnase Protection Assay*
RRM :	*RNA Recognition Motif*
RS :	arginine-sérine
RT-PCR :	*Reverse-Transcription-Polymerase Chain Reaction*
SELEX :	*Systematic Evolution of Ligand by Exponential Enrichment*
SNP	*Single Nucleotide Polymorphisms*
snRNA :	*small nucleolar RNA*
snRNP :	*small nuclear RiboNucleoProtein particle*
SR :	protéine contenant un domaine riche en dipeptides sérine/arginine
SRPK :	*SR Protein Kinase*
ss :	site d'épissage
SV40 :	*Simian Virus 40*
T :	Thymidine
U :	Uracile
UHM :	*U2AF Homology Motif*
USE :	*Upstream Sequence Element*
UTE :	*Upstream Terminal exon Enhancer*
UTR :	*UnTranslated Region*
UV :	Ultra Violet
Vg :	Vésicule germinative

SOMMAIRE

ABREVIATIONS.....- 1 -

SOMMAIRE.....- 3 -

INTRODUCTION.....- 6 -

1. LA MATURATION DES ARN PRE-MESSAGERS .....- 8 -

1.1. L'EPISSAGE.....- 9 -
1.1.1. Les sites d'épissage ..... *- 11 -*
1.1.2. La réaction d'épissage ..... *- 11 -*
1.1.3. Le spliceosome ..... *- 12 -*
1.1.3.1. Les particules ribonucléoprotéiques dites snRNP ..... - 12 -
1.1.3.2. Les protéines U2AF (U2snRNP Auxiliary Factor) ..... - 13 -
1.1.3.3. Le facteur SF1 ..... - 13 -
1.1.3.4. Les protéines SR..... - 14 -
1.1.3.5. Les protéines hnRNP..... - 15 -
1.1.3.6. Les autres protéines ..... - 15 -
1.1.4. L'assemblage du spliceosome ..... *- 15 -*
1.1.5. La définition intronique versus la définition exonique..... *- 20 -*
1.2. L'ETAPE DE CLIVAGE ET DE POLYADENYLATION .....- 24 -
1.1.1. Les sites de clivage et de polyadénylation ..... *- 24 -*
1.1.2. La machinerie de clivage et de polyadénylation..... *- 25 -*
1.1.3. La réaction de clivage et de polyadénylation..... *- 26 -*

2. LA MATURATION ALTERNATIVE DES ARN PRE-MESSAGERS .....- 28 -

2.1. LES DIFFERENTS TYPES D'EPISSAGE ALTERNATIF .....- 31 -
2.2. ANALYSE GLOBALE DES EVENEMENTS D'EPISSAGE ALTERNATIF .....- 33 -
2.3. LES REGULATIONS DE L'EPISSAGE ALTERNATIF .....- 36 -
2.3.1. Les éléments régulateurs agissant en cis ..... *- 36 -*
2.3.2. Les éléments régulateurs agissant en trans..... *- 39 -*
2.3.2.1. Les protéines SR..... - 39 -
2.3.2.2. Les protéines hnRNP..... - 45 -
2.3.2.4. Les autres protéines régulatrices ..... - 54 -
2.3.3. Les différents modèles de régulation de l'épissage alternatif..... *- 58 -*
2.3.3.1. Régulation de l'épissage alternatif par des protéines spécifiques d'un environnement cellulaire : voie du déterminisme sexuel chez la drosophile..... - 60 -
2.3.3.2. Régulation de l'épissage alternatif par des complexes de protéines : l'exon N1 de l'ARNpré-m de c-src ..... - 62 -
2.3.3.3. Régulation de l'épissage alternatif par des voies de signalisation cellulaires....- 64 -
2.3.3.4. Régulation non physiologique de l'épissage alternatif ..... - 65 -

3. LA MATURATION DES EXONS 3' TERMINAUX.....- 66 -

3.1. DEFINITION DES EXONS 3' TERMINAUX.....- 67 -
3.2. REGULATION DES EXONS 3' TERMINAUX ALTERNATIFS .....- 70 -
3.2.1. Les régulations de la polyadénylation ..... *- 71 -*

3.2.1.1. La protéine UIA ..... - 72 -
3.2.1.2. La protéine UI-70K ..... - 73 -
3.2.1.3. La protéine PTB ..... - 73 -
3.2.1.4. La protéine CstF64 ..... - 74 -
3.2.2. Modèles de la régulations des exons 3' terminaux alternatifs ..... - 76 -
3.2.2.1. Le modèle du gène calcitonin/calcitonin gene-related peptide ..... - 76 -
3.2.2.2. Le modèle des chaînes lourdes des immunoglobulines ..... - 78 -

4. COUPLAGE EPISSAGE CLIVAGE/POLYADENYLATION ET TRANSCRIPTION ..... - 81 -

4.1. COUPLAGE SPATIAL ..... - 81 -
4.2. COUPLAGE TEMPOREL ..... - 83 -
4.3. ROLE DE LA TRANSCRIPTION DANS LA REGULATION DE L'EPISSAGE ALTERNATIF ..... - 84 -

OBLECTIFS DE L'ETUDE ..... - 88 -

1. L'EPISSAGE D'UN EXON 3' TERMINAL *IN VIVO* ..... - 88 -

2. LE MODELE D'ETUDE ..... - 91 -

2.1. L'ARNPRE-M DE L'A-TROPOMYOSINE ..... - 91 -
2.2. LE XENOPE ..... - 93 -
2.2.1. Utilisation de minigènes pour étudier les régulations de l'épissage ..... - 93 -
2.2.2. Utilisation des morpholinos oligonucléotides antisens ..... - 95 -

3. RESULTATS PRECEDEMMENT OBTENUS AU LABORATOIRE ..... - 97 -

4. PROJET DE RECHERCHE ..... - 98 -

RESULTATS ..... - 99 -

CHAPITRE 1. CARACTERISATION FONCTIONNELLE D'UN ELEMENT INTRONIQUE NECESSAIRE A LA RECONNAISSANCE DE L'EXON 9A9' DE L'ARNPRE-M DE L'A-TROPOMYOSINE DE XENOPE EN TANT QU'EXON 3' TERMINAL ..... - 99 -

1.1. ARTICLE 1- LA PTB EMPECHE L'ACTIVATION D'UN ELEMENT INTRONIQUE REGULATEUR QUI FAVORISE L'UTILISATION D'UN EXON 3' TERMINAL COMPOSITE ..... - 99 -
1.2. RELATION ENTRE L'ELEMENT UTE ET LES SIGNAUX QUI DEFINISSENT L'EXON 9A9' : POINT DE BRANCHEMENT, SITE 3' D'EPISSAGE ET SIGNAL DE POLYADENYLATION AE ..... - 126 -
1.2.1. Les morpholinos dirigés contre le point de branchement et le site 3' d'épissage bloquent l'épissage de l'exon 9A9' ..... - 126 -
1.2.2. Le morpholino dirigé contre le signal de polyadénylation αe bloque le clivage de l'ARNpré-m ..... - 129 -
1.2.3. Conséquences du blocage du signal de polyadénylation et du site 3' d'épissage de l'exon 9A9' sur le profil d'épissage de l'ARNpré-m α-TM ..... - 129 -
1.3. DISCUSSION ET PERSPECTIVES ..... - 132 -

CHAPITRE 2. IDENTIFICATION DE FACTEURS IMPLIQUES DANS LA REGULATION DE L'EPISSAGE DE L'EXON 9A9' ..... - 139 -

2.1. INTRODUCTION ..... - 139 -
2.2. STRATEGIES DE PURIFICATIONS ET DE CARACTERISATION DES COMPLEXES RNP ..... - 139 -
2.2.1. Utilisation de l'ARN comme « appât » ..... - 140 -
2.2.1.1. Une modification naturelle dans l'ARN ..... - 142 -
2.2.1.2. Une modification artificielle dans l'ARN ..... - 142 -

2.2.1.3. Un aptamère ou étiquette dans l'ARN ..... - 144 -
2.2.1.4. Fixation de l'ARN via un oligonucléotide biotinylé ..... - 146 -
2.2.2. Choix de la stratégie utilisée ..... *- 147 -*
2.3. MISE EN ŒUVRE DE LA PURIFICATION ..... - 149 -
2.3.1. Choix de l'ARN utilisé pour la purification des complexes ..... *- 149 -*
2.3.2. Choix des extraits nucléaires ..... *- 155 -*
2.3.2.1. Les extraits nucléaires d'ovocytes ..... - 155 -
2.3.2.2. Les extraits nucléaires de cellules HeLa ..... - 157 -
2.3.3. Obtention d'extraits nucléaires de cellules HeLa dépourvus de PTB ..... *- 160 -*
2.3.3.1. Déplétion de la PTB par chromatographie sur ARN immobilisés ..... - 162 -
2.3.3.2. Imunodéplétion en PTB ..... - 164 -
2.3.4. Purification des complexes RNP assemblés sur la séquence UTE in vitro à l'aide d'ARN modifiés avec le métapériodate de sodium ..... *- 164 -*
2.3.4.1. Formation des complexes RNP in vitro ..... - 164 -
2.3.4.2. Purification des complexes RNP par chromatographie d'affinité à l'ARN - 168 -
2.4. ANALYSE PAR SPECTROMETRIE DE MASSE DE LA FRACTION 68 A 80 KDA DES PROTEINES RECRUTEES SUR LES ARN UTEL ..... - 174 -
2.5 CONCLUSIONS ..... - 191 -

CONCLUSION GENERALE ..... - 193 -

ANNEXE 1 - MATERIELS ET METHODES ..... - 200 -

1. MATERIELS ..... - 200 -

1.1. ANTICORPS ..... - 200 -
1.2. CONSTRUCTIONS DISPONIBLES ..... - 200 -
1.3. EXTRAIT NUCLEAIRE ET TAMPON ..... - 200 -

2. METHODES ..... - 201 -

2.1. MATRICES DE TRANSCRIPTION IN VITRO ..... - 201 -
2.2. TRANSCRIPTION IN VITRO ..... - 201 -
2.3. DEPLETION EN PTB DES EXTRAITS NUCLEAIRES DE CELLULES HELA ..... - 202 -
2.3.1. Par chromatographie sur ARN immobilisés ..... *- 202 -*
2.3.2. Immunodéplétion ..... *- 202 -*
2.4. EPISSAGE IN VITRO ..... - 203 -
2.5. RETARD DE MIGRATION SUR GEL ..... - 203 -
2.6. TRANSFERT DE MARQUAGE ..... - 203 -
2.7. PURIFICATION DE COMPLEXES A L'AIDE D'ARN OXYDES PAR LE METAPERIODATE DE SODIUM SUR LES BILLES ADIPIC ACID DIHYDRAZIDE AGAROSE ..... - 204 -
2.8. COLORATION AU NITRATE D'ARGENT ..... - 205 -

ANNEXE 2 - DESCRIPTION D'UNE NOUVELLE METHODE DE DETECTION DE TRANSCRITS ALTERNATIFS PAR PYROSEQUENÇAGE : PASP (PYROSEQUENCING ANALYSIS OF SPLICING PATTERN) ..... - 206 -

BIBLIOGRAPHIE ..... - 208 -

INTRODUCTION

Chez les eucaryotes supérieurs, les gènes sont composés de parties codantes, appelées exons, intercalées de séquences non codantes, nommées introns. L'épissage permet l'excision des introns et la jonction des exons. Ce mécanisme permet d'expliquer comment un gène conduit à la production d'une protéine. Cependant, le séquençage du génome humain a mis en avant un nombre de gènes codant pour des protéines relativement faible par rapport à celui qui était attendu. Chez l'homme, 20000 à 25000 gènes codent pour plus de 100000 protéines (Human Genome Project ; (Lander *et al.*, 2001)). Cette observation a conduit à s'interroger sur les mécanismes permettant de générer une telle diversité protéique.

L'étude des modifications post-transcriptionnelles, mécanismes nucléaires qui permettent la maturation des ARN pré-messagers (ARNpré-m) en ARN messagers (ARNm), a permis de montrer à la fin des années 70 l'existence de l'épissage alternatif (Gilbert, 1978). L'épissage est dit alternatif quand différentes combinaisons d'exons sont sélectionnées dans un ARNpré-m. La sélection des exons diffère selon le stade de développement, le type cellulaire ou les stimuli reçus par la cellule. L'épissage alternatif permet ainsi de produire à partir d'un seul ARNpré-m, plusieurs ARNm pouvant conduire à la synthèse de protéines structurellement et/ou fonctionnellement différentes. Celui-ci a tout d'abord été considéré comme un mécanisme mineur impliquant environ 5% des gènes (Sharp, 1994). Aujourd'hui, grâce au développement de techniques d'étude à « grande échelle », il est considéré comme un mécanisme majeur dans la maturation des ARNpré-m. En effet, l'analyse combinée des EST (*Expressed Sequence Tag*) et de banques d'ADNc indique que 40 à 60% des gènes humains sont épissés de manière alternative (Modrek et Lee, 2002). Ce chiffre monte à 75% suite aux études par puces à ADN dédiées à l'épissage et couvrant les jonctions exons-exons (Johnson *et al.*, 2003; Clark *et al.*, 2007). A ce jour, l'épissage alternatif est donc considéré comme le mécanisme contribuant le plus à l'augmentation de la diversité protéique.

La sélection différentielle des exons induite par l'épissage alternatif implique que la machinerie d'épissage différencie les exons des introns. Celle-ci est capable de reconnaître les jonctions exon-intron des exons « centraux », et utilise d'autres particularités pour reconnaître les exons terminaux. Ainsi, la machinerie d'épissage reconnaît le complexe mis en place sur la coiffe et la jonction exon-intron des exons 5' terminaux, et la jonction intron-exon et le complexe mis en place au niveau du site de clivage/polyadénylation des exons 3' terminaux.

Cette observation met en avant le fait que les exons terminaux peuvent également être soumis à une maturation alternative. Ainsi la sélection différentielle d'un exon 5' terminal pourra conduire à la production d'ARNm possédant des régions 5' terminales distinctes, et la sélection de différents sites de clivage/polyadénylation produira des ARNm aux extrémités 3' terminales distinctes.

La polyadénylation est une réaction composée de deux étapes. La première correspond à un clivage endonucléolytique de l'ARNpré-m en un site spécifique déterminé par la fixation de facteurs de polyadénylation. La seconde consiste en l'ajout de résidus adénosine à l'extrémité 3' terminale produite après clivage. La polyadénylation alternative concerne 29 à 54% des gènes humains (Beaudoing *et al.*, 2000; Tian *et al.*, 2005; Zhang *et al.*, 2005). Les protéines résultant des ARNpré-m différentiellement maturés au niveau du site de clivage/polyadénylation peuvent posséder des régions carboxy terminales différentes.

Les mécanismes d'épissage et de clivage/polyadénylation sont fortement interconnectés et couplés avec la transcription (Moore et Proudfoot, 2009). Ces connexions sont gouvernées par des interactions ARN-protéines. La formation de ces complexes ribonucléoprotéiques (RNP), leur composition et leur remaniement au cours de la synthèse et de la maturation des ARNpré-m, puis de l'export et de l'initiation de la traduction des ARNm va orienter leur devenir (Wang et Burge, 2008; Wahl *et al.*, 2009). Les RNP sont donc des éléments clés de la régulation de l'expression des gènes. L'épissage alternatif peut moduler l'expression d'un gène en faisant apparaître des codons stop précoces dans sa région traduite. Cet ARNm va alors être entraîné vers des mécanismes de dégradation comme le *nonsense-mediated decay* (NMD). En modifiant les régions 5' et 3' non traduites (UTR), l'épissage alternatif peut également modifier la stabilité des ARNm (Stamm *et al.*, 2005). Enfin, la composition des complexes RNP peut moduler l'export et la traductibilité des ARNm (Sawicka *et al.*, 2008; Long et Caceres, 2009).

Vue l'importance de ces mécanismes dans l'expression génétique chez l'homme, il n'est pas étonnant qu'un nombre croissant de maladies aient été montrées comme étant associées à des perturbations de leurs régulations (Srebrow et Kornblihtt, 2006; Lukong *et al.*, 2008). Plus de la moitié des mutations induisant des maladies touchent l'épissage (Wang et Cooper, 2007). Ces anomalies peuvent avoir pour origine des mutations des sites d'épissage ou des séquences régulatrices, qui conduisent à la production de protéines aberrantes ou non

fonctionnelles. De plus, les facteurs *trans* régulateurs sont généralement surexprimés dans les tissus tumoraux (Kirschbaum-Slager *et al.*, 2004). Les différents éléments permettant la régulation de la maturation des ARNpré-m sont donc des cibles de choix pour la mise en place de thérapies visant à corriger leurs défauts impliqués dans de nombreux cancers et autres maladies génétiques (Corcos et Solier, 2005; Tazi *et al.*, 2005). Par exemple, l'utilisation d'oligonucléotides antisens permet de reprogrammer les décisions d'épissage en ciblant les sites d'épissage ou les séquences sites régulatrices. Cette stratégie, très spécifique d'un ARNpré-m donné, a été utilisée *in vitro* pour corriger les défauts d'épissage d'un ARNpré-m défectueux (Cartegni et Krainer, 2003). Des drogues ciblant spécifiquement les protéines SR, régulatrices de l'épissage, ont également été développées, comme par exemple des dérivés indole qui bloquent l'assemblage précoce du spliceosome au niveau de séquences exoniques activatrices (Soret *et al.*, 2005). Une autre stratégie est en cours de développement. Celle-ci utilise des vecteurs dérivés de virus afin d'exprimer des séquences antisens permettant de restaurer le cadre de lecture des ARNm modifiés chez les patients atteints de dystrophie musculaire de Duchenne (Benchaouir *et al.*, 2007).

L'épissage alternatif est donc un acteur majeur de la régulation au niveau quantitatif et qualitatif de l'expression des gènes chez l'homme. Il semble donc important de comprendre les mécanismes impliqués dans la régulation des processus d'épissage et de clivage/polyadénylation. Cette introduction débutera par une présentation générale de ces deux processus. Les mécanismes moléculaires impliqués dans leur régulation seront ensuite présentés. Enfin, les liens entre les différents mécanismes de maturation des ARNpré-m seront abordés.

1. La maturation des ARN pré-messagers

Chez les eucaryotes la maturation des ARNpré-m a lieu de manière co-transcriptionnelle. Avant qu'un ARNpré-m nouvellement transcrit puisse être transporté hors du noyau où il sera traduit, celui-ci doit subir trois processus majeurs permettant la production d'un ARNm mature. Si chacune de ces réactions est biochimiquement distincte, elles sont interconnectées et couplées à la transcription (Proudfoot *et al.*, 2002).

Dès le début de la transcription par l'ARN polymérase II (ARNpol II), une coiffe composée d'un résidu 7 méthyl-guanosine est ajoutée à l'extrémité 5' phosphate de l'ARN en cours de synthèse. Une fois cette étape effectuée, l'ARNpré-m va être épissé permettant ainsi l'excision des introns et la jonction des exons. De manière concomitante, un complexe de jonction exon-exon (*Exon Junction Complex* ou EJC) est ajouté dans une région de 24 nucléotides en amont de la jonction. Ce complexe est impliqué dans différents processus cellulaires comme l'export nucléocytoplasmique, la localisation subcellulaire, ou encore le contrôle de la qualité des ARNm (Le Hir et Seraphin, 2008). La dernière étape de maturation des ARNpré-m est constituée d'un clivage endonucléolytique de leur extrémité 3' et de l'ajout d'une queue de polyadénosines (poly(A)) d'environ 200 résidus. Le clivage permet le décrochage de l'ARNpol II de l'ADN et le relargage du complexe de transcription (West *et al.*, 2008).

Au cours de ces processus de maturation, les ARNpré-m ne sont pas « nus » dans le noyau, mais recouverts d'un certain nombre de protéines. Cet assemblage est appelé complexe ribonucléoprotéique (RNP). La composition des RNP, qui évolue au cours de la maturation des ARN, va orienter le devenir de l'ARNpré-m puis de l'ARNm.

1.1. L'épissage

Les introns sont présents dans les trois règnes du monde du vivant : les eucaryotes, les procaryotes et les « archaes » (Belfort *et al.*, 1995). Chez les eucaryotes, les introns sont présents dans le génome nucléaire, mais également dans le génome des mitochondries et des chloroplastes. Chez les métazoaires, la majorité des gènes possèdent des introns. Au contraire, chez les eucaryotes unicellulaires comme *S. cerevisiae* moins de 5% des gènes en contiennent (Lopez et Seraphin, 1999), et il existe rarement plus d'un intron par gène (Bon *et al.*, 2003).

L'épissage consiste en l'élimination des introns et à la jonction des exons. Chez l'homme, environ 90% de l'ARNpré-m, correspondant aux introns, est éliminé, seuls les 10% restants, correspondant aux exons, sont présents dans l'ARNm (Stamm *et al.*, 2005). Les introns sont excisés suite à deux réactions de transestérification qui ont lieu au sein d'un macrocomplexe ribonucléoprotéique, le spliceosome. Chez les eucaryotes les introns sont majoritairement délimités par les signaux d'épissage GT/AG localisés respectivement aux

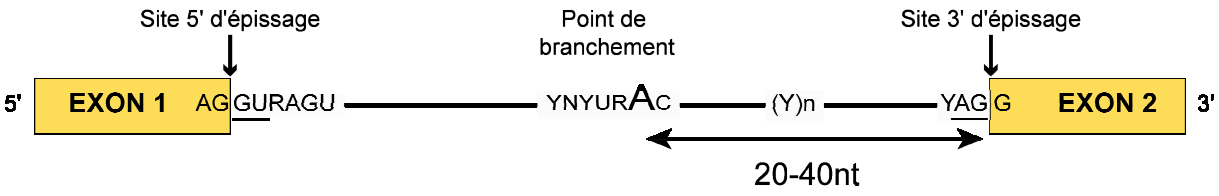


Figure 1- Séquences d'épissage exoniques et introniques consensuelles présentes au niveau d'un ARNpré-m chez les vertébrés. Les dinucléotides introniques invariants GU et AG sont indiqués. Le A correspond au point de branchement. La distance optimale entre celui-ci et le site 3' d'épissage est indiqué. Y : pyrimidines C, U. N : A, C, G, U. R : purine A, G.

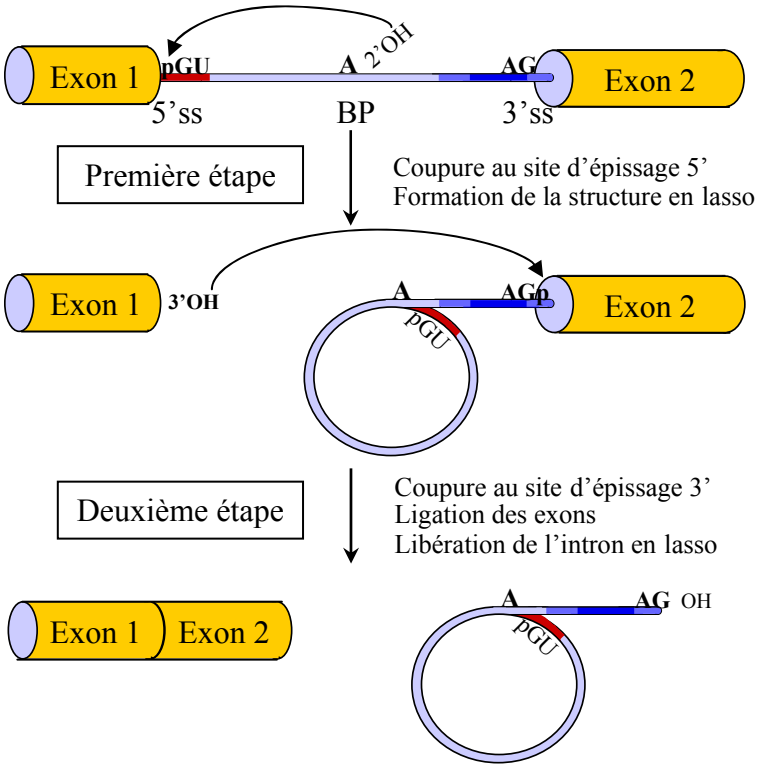


Figure 2- Les étapes catalytiques de l'épissage des ARNpré-m. Les exons sont représentés par des rectangles et les introns par des traits. Les groupements conservés dans les ARNpré-m de mammifères, les groupements phosphate (p) et hydroxyle (OH) sont représentés. 5' ss et 3' ss correspondent respectivement aux site 5' et 3' d'épissage, et BP au point de branchement (A). Les flèches représentent les attaques nucléophiles des phosphates des sites d'épissage par les groupements hydroxyles.

extrémités 5’ et 3’ de l’intron et excisés par le spliceosome dépendant des snRNP U1/U2 (Kreivi et Lamond, 1996). Une très faible proportion d’introns, 1 sur 300 chez l’homme, est éliminée de l’ARNpré-m par un autre complexe, appelé spliceosome « mineur » qui est dépendant des snRNP U11/U12 et qui reconnaît principalement les dinucléotides introniques AT/AC (Steitz *et al.*, 2008).
Dans cette introduction, je m’attacherai à décrire les introns GT/AC, reconnus par le spliceosome « majeur » ou « U2 dépendant ».

1.1.1. Les sites d’épissage

Quatre séquences sont importantes pour la reconnaissance des introns par le spliceosome : le site 5’ d’épissage, la région du point de branchement une région riche en pyrimidines et le site 3’ d’épissage. Chez la levure *S. cerevisiae* les séquences d’épissage introniques sont fortement conservées, tandis que chez les eucaryotes supérieurs, ces motifs particuliers sont plus dégénérés (Burge *et al.*, 1999). Cependant les sites 5’ et 3’ d’épissage présentent respectivement de manière très fréquente les dinucléotides GT et AG qui définissent les extrémités de l’intron. Une autre région importante pour l’épissage est la région du point branchement. Cette séquence est localisée généralement entre 20 et 40 nucléotides (nt) en amont du site 3’ d’épissage et contient une adénosine qui correspond au point de branchement. Chez les eucaryotes supérieurs, une région riche en pyrimidines, de taille variable, est présente entre le point de branchement et le site 3’ d’épissage.
Les motifs consensus de ses différentes régions présents chez les eucaryotes supérieurs sont représentés dans figure 1.

1.1.2. La réaction d’épissage

Au cours de l’épissage, l’excision des introns de l’ARNpré-m par le spliceosome se fait en deux étapes correspondant à deux réactions de transestérification (figure 2) (Moore et Sharp, 1993; Staley et Guthrie, 1998). La première consiste en une attaque nucléophile du groupement phosphate de la guanosine du site 5’ d’épissage par le groupement 2’ hydroxyle libre de l’adénosine du point de branchement. Cette première réaction aboutit à la libération de l’exon 5’ linéaire et à la production d’un intermédiaire en « lasso » (ou *lariat*) composé de l’exon 3’ lié à l’intron formant une boucle au niveau du point de branchement. La seconde attaque nucléophile a lieu entre le groupement 3’ hydroxyle libre de l’exon 5’ et le

groupement phosphate de l'exon 3'. Cette réaction libère l'intron en lasso qui va alors être dégradé, et va permettre la formation d'une liaison phosphodiester entre les exons 5' et 3'. Ces deux réactions sont réalisées au sein du complexe catalytique mature du spliceosome appelé complexe C.

1.1.3. Le spliceosome

Le spliceosome est un macrocomplexe ribonucléoprotéique qui est le siège de l'épissage. Celui-ci est composé de cinq petits ARN nucléaires (*Small Nuclear RNAs* ou snRNA) associés à des protéines pour former des particules snRNP (*Small Nuclear RibonucleoProtein*), et de plus de 150 protéines distinctes de celles contenues dans les snRNP (Zhou *et al.*, 2002; Jurica et Moore, 2003). Une étude récente réalisée sur des cellules en culture a permis la purification de plus de 300 protéines associées au spliceosome, ce qui en fait la machinerie cellulaire la plus complexe jamais caractérisée (Chen *et al.*, 2007). Son assemblage sur l'ARNpré-m a lieu au cours de la transcription et est hautement dynamique. Au cours de cet assemblage dit séquentiel, apparaissent différents complexes transitoires au sein desquels de nombreux réarrangements ATP-dépendants ont lieu, aboutissant ainsi à la formation du spliceosome actif.

1.1.3.1. Les particules ribonucléoprotéiques dites snRNP

Les cinq snRNP, U1, U2, U4, U5 et U6, composants majeurs du spliceosome (Will et Luhrmann, 2001; Patel et Bellini, 2008), sont composés de U snRNA (*Uridine rich Small Nuclear RNA*) et d'un cortège de protéines associées, dont les protéines Sm qui participent à la formation du corps de la snRNP. Les U snRNA sont des ARN de petite taille (entre 100 et 200 nt chez l'homme) fortement modifiés post-transcriptionnellement. Les snRNA U1, U2, U4, U5 sont transcrits par l'ARNpol II. Ils possèdent une coiffe de type 2,2,7 triméthyl guanosine ainsi qu'un motif Sm, riche en uridine encadré de deux structures en tige boucle. Le snRNA U6 est transcrit par l'ARNpol III, possède une coiffe monométhylée et est complexé à des protéines LSm (pour *Sm like*).

Le domaine central des snRNP U1, U2, U4 et U5 est constitué de huit protéines communes appelées protéines Sm. Le domaine périphérique est composé de protéines spécifiques comme les protéines U1-A, U1-C et U1-70K pour la snRNP U1, ou les protéines

15,5K et 61K pour la snRNP U4 (Patel et Bellini, 2008). La snRNP U6 est composée de sept protéines LSm (numérotées de 2 à 8) et de trois protéines communes à la di-snRNP U4/U6 (Patel et Bellini, 2008). Si les snRNP U1, U2 et U5 existent seules, U4 et U6 existent sous forme de di-snRNP qui peut former une tri-snRNP avec U5.

1.1.3.2. Les protéines U2AF (U2snRNP Auxiliary Factor)

Parmi les protéines non snRNP requises pour l'épissage, le facteur U2AF va se lier à l'ARN pour favoriser la reconnaissance des séquences d'épissage. Ce facteur est composé de deux polypeptides nommés selon leur masse moléculaire U2AF65 et U2AF35 (Zamore et Green, 1989). Ces deux protéines sont dites « *SR like* » car elles présentent en plus de leur domaine de liaison à l'ARN de type RRM (*RNA Recognition Motif*), un domaine RS riche en répétitions d'un dipeptide arginine-sérine qui est caractéristique des protéines SR. U2AF65 se fixe sur la région riche en pyrimidines et U2AF35 sur le dinucléotide AG du site 3' d'épissage. L'interaction entre ces deux facteurs a lieu entre le domaine amino terminal de U2AF65 et le domaine central UHM (*U2AF Homology Motif*) de U2AF35 (Kielkopf *et al.*, 2004). Les sous-unités U2AF interagissent avec la snRNP U2 par l'intermédiaire de protéines spécifiques de cette snRNP (Spadaccini *et al.*, 2006; Thickman *et al.*, 2006). Ces interactions sont nécessaires à la reconnaissance du point de branchement par la snRNP U2, permettant ainsi la première étape d'épissage (Zamore *et al.*, 1992; Wu *et al.*, 1999).

1.1.3.3. Le facteur SF1

La protéine SF1 est une protéine majeure du spliceosome qui reconnaît le site du point de branchement. Ce facteur de 75 kDa qui appartient à la famille STAR (*Signal Transduction and Activation of RNA*), est constitué de cinq domaines. La région amino terminale interagit avec le facteur U2AF65. L'interaction de ces deux facteurs, est dépendante de la phosphorylation de SF1 au niveau de la sérine 20 et favoriserait le recrutement de la snRNP U2 au point de branchement (Kramer et Utans, 1991; Wang *et al.*, 1999). Deux domaines sont impliqués dans la fixation de SF1 au niveau du point de branchement : un domaine KH (*hnRNP K Homology domain*) et un domaine de type « doigt de zinc » (*zinc knuckle*) (Berglund *et al.*, 1998; Liu *et al.*, 2001). Le domaine KH est nécessaire et suffisant pour la fixation de SF1 sur le point de branchement, alors que le domaine « zinc knuckle » stabilise cette interaction en se liant de manière non spécifique à l'ARN. Une région riche en proline

est responsable d'interactions protéines-protéines. Enfin aucune fonction n'a encore été caractérisée pour la région carboxy terminale. De nombreuses isoformes de SF1 existent selon les tissus, celles-ci présentent des régions riches en prolines et une région carboxy terminale de taille variable.

Bien que SF1 semble nécessaire à l'épissage des ARNpré-m *in vitro* ainsi qu'à la viabilité cellulaire, des expériences réalisées *in vitro* (Guth et Valcarcel, 2000) et *in vivo* (Tanackovic et Kramer, 2005) montrent que SF1 ne serait pas indispensable à l'épissage de tous les introns de mammifères. Ainsi SF1 serait spécifiquement impliqué dans l'épissage de certains exons alternatifs et non dans l'épissage constitutif (Tanackovic et Kramer, 2005).

1.1.3.4. Les protéines SR

Les membres de la famille des protéines riches en sérine et en arginine (SR) ont de multiples fonctions dans l'épissage des ARNpré-m. Celles-ci sont présentes chez tous les métazoaires, pour lesquels il existe une dizaine de membres, mais pas chez tous les eucaryotes. Par exemple, deux protéines SR ont été caractérisées chez *S. pombe*, mais aucune chez *S. cerevisiae*. La classification d'une protéine dans la famille des SR repose sur trois critères principaux (Graveley, 2000; Bourgeois *et al.*, 2004). Tout d'abord elles doivent avoir une structure modulaire présentant au moins un domaine RRM dans la partie amino terminale, et un domaine RS dans la partie carboxy terminale. Ensuite, leur fonction étant essentielle dans l'épissage, l'ajout d'un des membres de la famille suffit à complémenter un extrait nucléaire de cellules HeLa S100 incompétent pour l'épissage *in vitro*. Enfin, elles contiennent au moins un épitope phosphorylé quand elles sont actives. Cette classification a quelque peu évolué puisque certaines protéines présentant un domaine RRM et un domaine RS, sont dites SR alors qu'elles ne sont pas capables de complémenter un extrait cytoplasmique de cellules HeLa S100. C'est le cas par exemple de SRp38 (Feng *et al.*, 2008). Ces protéines sont appelées selon les articles « *SR like* » ou SR.

Si ces protéines sont indispensables à l'épissage constitutif, elles sont également impliquées dans la régulation de l'épissage alternatif du fait de leur interaction avec des séquences régulatrices présentes sur l'ARNpré-m. Les fonctions régulatrices des protéines SR seront développées dans le chapitre consacré aux régulations de l'épissage alternatif.

1.1.3.5. Les protéines hnRNP

Les protéines hnRNP (*Heterogeneous Nuclear RiboNucleoProtein*) ont surtout été décrites pour leur rôle de régulateurs spécifiques dans l'épissage alternatif, développé dans le chapitre consacré aux régulations de l'épissage alternatif. Cette famille d'au moins 20 protéines a la capacité de se fixer aux ARNpré-m (via un ou plusieurs domaine de fixation à l'ARN) et de former des interactions protéine-protéine (via des domaines auxiliaires). Certaines hnRNP se fixent de manière non spécifique sur les transcrits issus de l'ARNpol II (Matunis *et al.*, 1993) et participent à la mise en place du complexe H, qui précède la formation du premier complexe d'épissage, protégeant ainsi les ARN de la dégradation par des nucléases. Toutefois, le rôle des hnRNP dans l'épissage constitutif reste flou (Dreyfuss *et al.*, 1993; Chabot *et al.*, 2003).

1.1.3.6. Les autres protéines

Comme exposé au début de cette introduction, le spliceosome semble être la machinerie cellulaire la plus complexe avec quelques 300 facteurs protéiques impliqués dans sa mise en place et sa composition (Chen *et al.*, 2007). Ainsi sont associées au spliceosome des protéines de structure des ARN (ARN *Helicase-like*), des protéines d'export, et des protéines participant aux mécanismes de surveillance des ARN (comme des protéines du NMD). Des protéines impliquées dans la transcription et dans le processus de clivage/polyadénylation ont également été identifiées lors de la purification de spliceosome (Chen *et al.*, 2007). Du fait de leur caractérisation récente, de leur nombre, du dynamisme de la réaction d'épissage et de certaines interactions, l'implication de tous ces facteurs dans l'épissage n'a pas encore été résolue.

1.1.4. L'assemblage du spliceosome

La conformation catalytique du spliceosome est l'aboutissement de nombreux réarrangements ATP dépendants de complexes intermédiaires. L'association des différents composants repose sur des interactions ARN-ARN, ARN-protéines et protéines-protéines (House et Lynch, 2008).

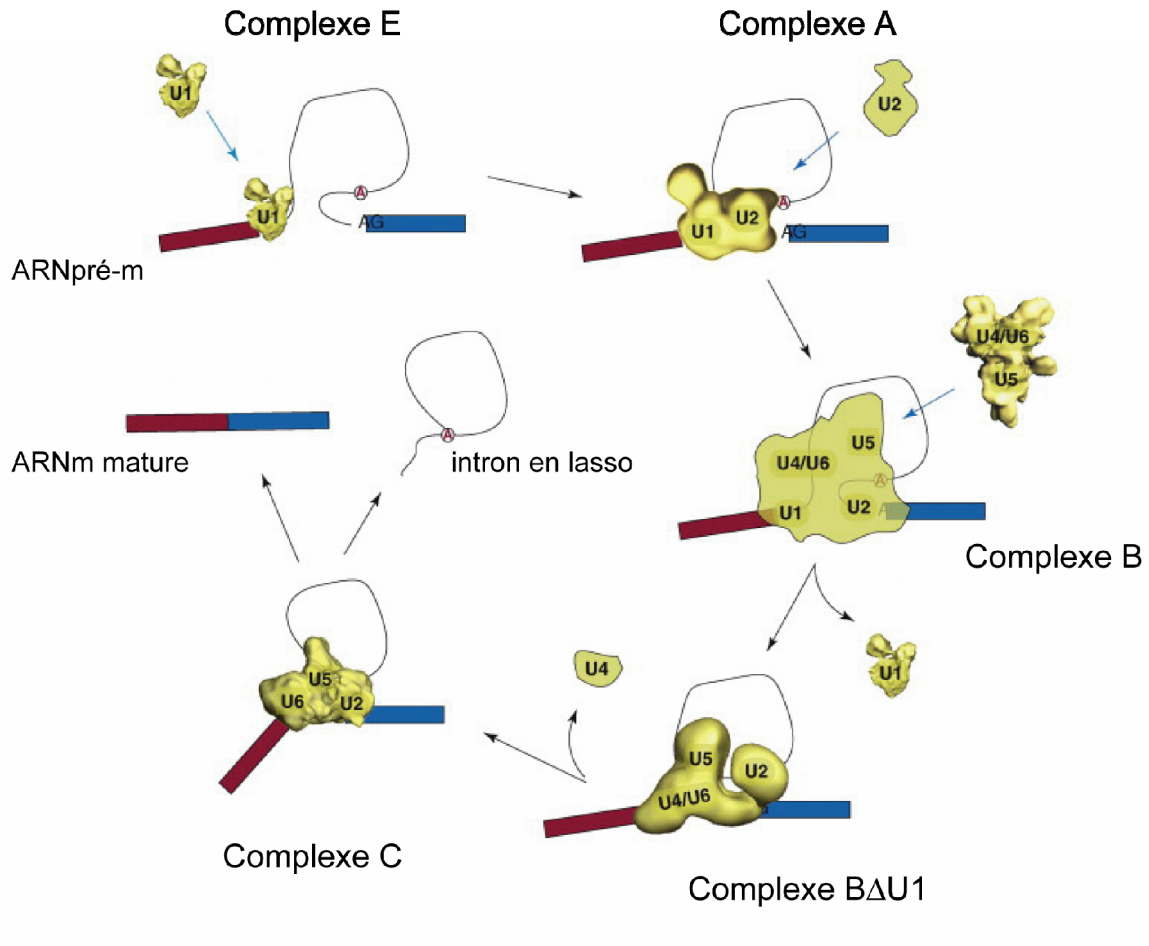


Figure 3- Modèle d'assemblage séquentiel du spliceosome. D'après {Rino, 2009 #1879}. La maturation d'un ARNpré-m contenant deux exons et un intron est représentée. L'épissage conduit à la jonction des exons et à la libération de l'intron structuré en « lasso ». Les complexes macromoléculaires correspondant aux intermédiaires d'épissage sont indiqués : complexes E, A, B, BΔU1, et C.

Il existe deux modèles expliquant l'assemblage du spliceosome sur l'ARNpré-m. Le premier, qui a longtemps été admis, montre que le spliceosome s'assemble de manière séquentielle sur l'intron à épisser. L'épissage s'accompagnerait alors de la libération des différents composants afin de renouveler un cycle d'assemblage sur un nouvel intron. Le second modèle repose sur l'existence d'un spliceosome déjà pré-assemblé dans le noyau.

Le premier modèle, dit de l'assemblage séquentiel du spliceosome (schématisé dans la figure 3) est basé essentiellement sur des études réalisées chez la levure et *in vitro* à l'aide d'extraits nucléaires de cellules de mammifères (Seraphin et Rosbash, 1989; Fabrizio *et al.*, 1994; Chiara *et al.*, 1996; Das et Reed, 1999). Un premier complexe est formé sur l'ARNpré-m de manière co-transcriptionnelle. Celui-ci, appelé complexe H, est composé notamment de protéines hnRNP, et se forme indépendamment de la présence de signaux d'épissage fonctionnels. La composition en hnRNP du complexe H est variable d'un ARNpré-m à l'autre. La fixation de ces protéines sur l'ARNpré-m est importante pour les régulations précoces de l'épissage alternatif (Bennett *et al.*, 1992). La mise en place du spliceosome proprement dit débute par la formation du complexe d'engagement, appelé complexe E. Le site 5' d'épissage est reconnu par la snRNP U1, le point de branchement par la protéine SF1, et la région 3' d'épissage par le complexe U2AF (Reed, 2000). Depuis la publication des travaux de (Guth et Valcarcel, 2000; Tanackovic et Kramer, 2005), la nécessité de la fixation de SF1 au point de branchement pour avancer dans la formation des complexes spliceosomaux est discutée. SF1 pourrait jouer un rôle dans la cinétique de l'assemblage du spliceosome en favorisant le recrutement de la snRNP U2. Il a également été proposé que SF1 favoriserait le rapprochement entre la snRNP U1 et le site 3' d'épissage au sein d'un complexe appelé E', précédant le complexe E, et ne contenant pas l'hétérodimère U2AF (Kent *et al.*, 2005). Le recrutement de la snRNP U2 est favorisé par l'hétérodimère U2AF et la snRNP U1. A ce stade, l'association de la snRNP U2 est faible (Das *et al.*, 2000).

En présence d'ATP, l'interaction entre la snRNP U2 et le point de branchement est stabilisée et la protéine SF1 est exclue. Le complexe A est ainsi formé. L'entrée dans le complexe A de la tri-snRNP U4/U6.U5 conduit à la formation du complexe B. De nombreux réarrangements ATP dépendants conduisent au relargage des snRNP U1 et U4 du spliceosome. Le complexe C est formé suite à l'interaction de la snRNP U6 avec le site 5' d'épissage, remplaçant ainsi la snRNP U1, et avec la snRNP U2 (Villa *et al.*, 2002). Ce

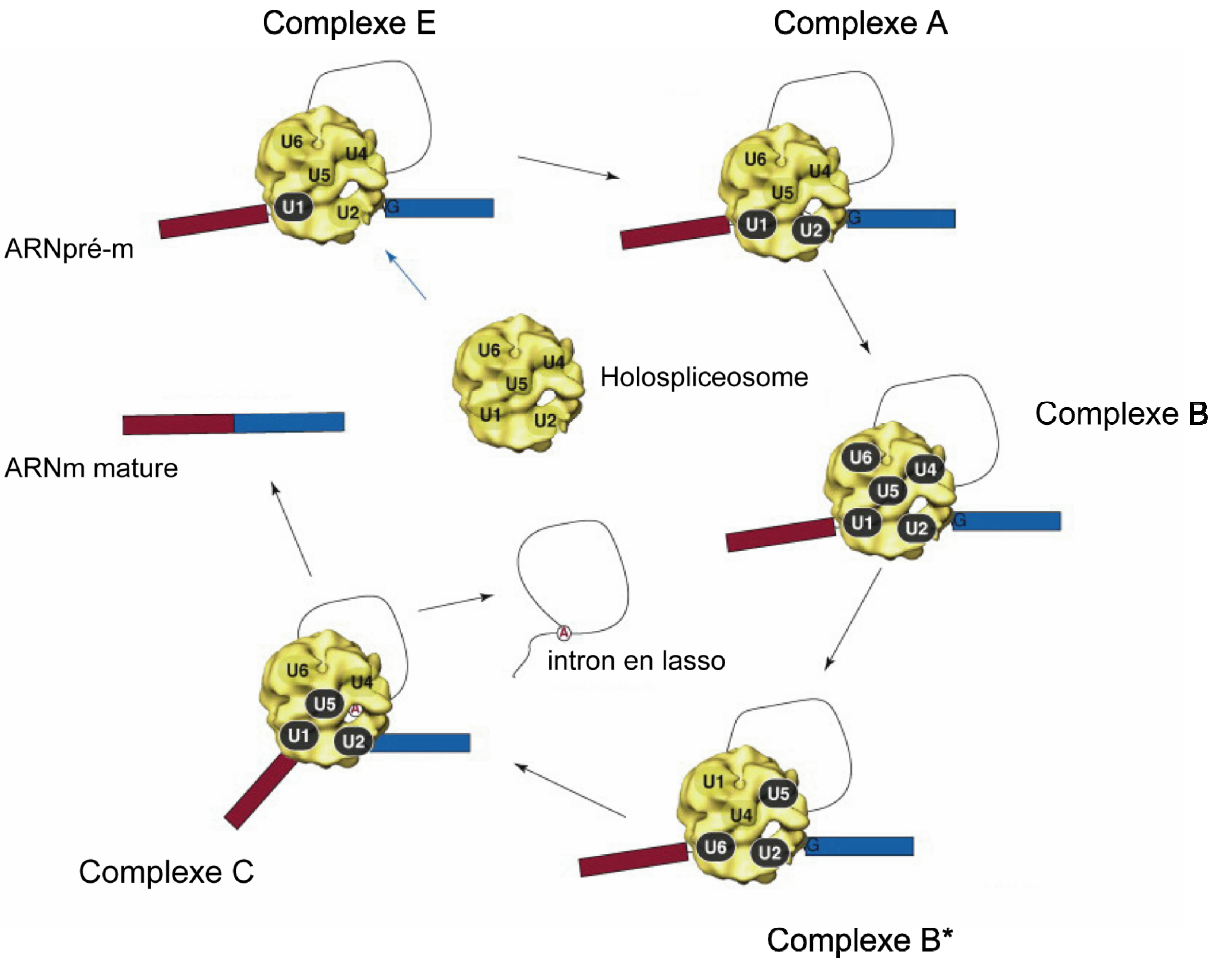


Figure 4- Modèle de la penta-snRNP. D'après {Rino, 2009 #1879}. Il existerait un spliceosome pré-assemblé ou holospliceosome. Dans ce cas, les cinq snRNP présentes au sein d'un même complexe sont recrutées sur l'ARNpré-m. L'épissage est alors le fait de réarrangements conformationnels qui permettent de moduler les interactions entre l'ARN et les composants du complexe.

complexe qui permet les deux réactions de transestérification contient donc les snRNP U2, U5 et U6. Les snRNP U2 et U6 forment le complexe catalytique du spliceosome en permettant les deux étapes de clivage de l'ARNpré-m (Butcher et Brow, 2005; Mefford et Staley, 2009). Il a longtemps été admis que le spliceosome permettait à lui seul le clivage endonucléolytique de l'ARNpré-m, mais des travaux récents montrent que la protéine Prp8 est impliquée de manière très active dans cette catalyse, notamment par l'intermédiare de son domaine RNase H (Turner *et al.*, 2006; Valadkhan *et al.*, 2007; Abelson, 2008; Ritchie *et al.*, 2008). La protéine Prp8 est également associée à la snRNP U5 qui permet de rapprocher les exons 5' et 3' (Dix *et al.*, 1998; Staley et Guthrie, 1998). Le relargage de l'intron en lasso et le désassemblage du spliceosome est favorisé par des ARN hélicases (Wen *et al.*, 2008).

Les protéines SR interviennent au cours de la mise en place du spliceosome (Graveley, 2000; Bourgeois *et al.*, 2004). Précocement, elles facilitent la mise en place du complexe E en favorisant le recrutement et la stabilisation de la snRNP U1 (via la U1-70K) et de l'hétérodimère U2AF respectivement aux sites 3' et 5' d'épissage (Kohtz *et al.*, 1994; Staknis et Reed, 1994). Elles sont également impliquées dans la transition du complexe A vers le complexe B en facilitant le recrutement de la tri-snRNP U4/U6.U5 sur l'ARNpré-m (Roscigno et Garcia-Blanco, 1995).

Le second modèle est basé sur l'existence d'un spliceosome pré-assemblé (figure 4). Un complexe nommé « pseudospliceosome » composé des snRNP U2, U4, U5 et U6 assemblé sans ARNpré-m a été identifié dans des extraits nucléaires de cellules HeLa (Konarska et Sharp, 1988). D'autres expériences réalisées chez la levure *S. cerevisiae* ont permis de mettre en évidence l'existence d'un complexe contenant les cinq snRNP ainsi qu'une soixantaine de protéines associées (Stevens *et al.*, 2002). Ainsi, les différents complexes spliceosomaux intermédiaires E, A, B et C correspondraient à des étapes de stabilisation d'interactions plutôt qu'au recrutement séquentiel des facteurs d'épissage (Nilsen, 2002; Rino et Carmo-Fonseca, 2009). Des données récentes montrent que ce modèle peut être reconstitué *in vitro*, conduisant ainsi à la formation de quatre spliceosomes actifs autour d'un ARNpré-m et visible par microscopie électronique (Azubel *et al.*, 2006; Sperling *et al.*, 2008). L'existence d'un spliceosome pré-assemblé permettrait de scanner l'ARNpré-m en cours de synthèse facilitant ainsi l'excision des introns (Azubel *et al.*, 2006).

Ces deux modèles permettent d'expliquer la mise en place du spliceosome sur les ARNpré-m et soulèvent la question de l'utilisation des modèles mis en place *in vitro* qui

permettent d'étudier de manière dissociée la biochimie des réactions de maturation des ARNpré-m sans tenir compte de l'organisation nucléaire qui semble jouer un rôle important dans la régulation de ces évènements de maturation (Misteli, 2001; Misteli, 2005; Rino *et al.*, 2007).

1.1.5. La définition intronique versus la définition exonique

La reconnaissance des sites d'épissage « authentiques » est une étape critique de la réaction d'épissage, indispensable à la conservation du message transmis entre les gènes et les protéines *via* les molécules d'ARNm. Chez la levure *S. cerevisiae* la taille limitée des introns (inférieure à 550 nt) ainsi que les séquences introniques sont nécessaires à l'épissage (Langford et Gallwitz, 1983; Klinz et Gallwitz, 1985). Au contraire, la longueur des exons ne semble pas être une contrainte (Zhang et Marr, 1994). Ces observations ont conduit à proposer un mécanisme de reconnaissance des sites d'épissage au travers de l'intron (figure 5) (Talerico et Berget, 1994). L'interaction des facteurs d'épissage de part et d'autre de l'intron permet ensuite l'assemblage du spliceosome. Le modèle de la définition intronique s'appliquerait aux ARNpré-m possédant des introns de petite taille, comme c'est le cas pour 50 % des ARNpré-m de la drosophile (Guo *et al.*, 1993), ou encore chez les eucaryotes unicellulaires (Romfo *et al.*, 2000), et potentiellement chez l'homme comme par exemple dans l'intron 2 de l'α-globine (McCullough et Berget, 1997).

Le modèle de la définition de l'intron ne permet pas d'expliquer la majeure partie de l'épissage chez les eucaryotes supérieurs dont les ARNpré-m présentent des sites d'épissage faiblement conservés et une forte disparité de taille entre les exons et les introns. En effet, chez l'homme la taille moyenne des exons est de 145 nt contre 3365 nt pour les introns (Stamm *et al.*, 2005). Si l'on regarde l'ensemble du génome humain, les séquences exoniques représentent moins de 10 % de l'ARNpré-m. Si l'on considère en plus la distribution de taille des introns qui varie de moins de 100 nt à plus de 100000 nt, il devient difficile d'imaginer comment le spliceosome peut reconnaître avec précision les jonctions exons/introns. De ce fait, un second modèle a été proposé suite à des expériences d'épissage effectuées *in vitro*, qui ont permis de montrer que la présence d'un site 5' d'épissage favorise l'épissage de l'exon localisé en amont (Robberson *et al.*, 1990). Ce modèle dit de la définition de l'exon propose

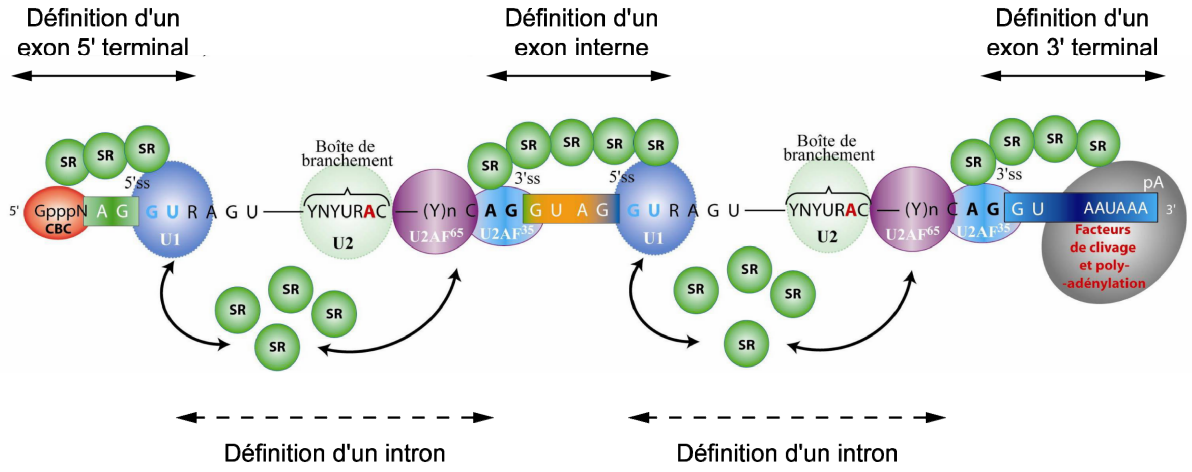


Figure 5- Exon définition ou intron définition ? L'intron définition est le modèle permettant la mise en place du spliceosome suite à la reconnaissance des sites d'épissage au travers des introns. Ce modèle est valable pour les ARNpré-m possédant des introns de petite taille. A l'inverse, le modèle de la définition de l'exon permet d'expliquer la reconnaissance des sites d'épissage au travers de l'exon des ARNpré-m humains qui possèdent de grands introns. Les sites 5' et 3' d'épissage sont représentés par 5'ss et 3'ss, le point de branchement présent dans la boite de branchement est représenté par un A en rouge, et la région riche en pyrimidines par (Y)n. La définition de l'exon interne débute par la fixation du facteur U2AF au niveau du site 3' d'épissage et par la reconnaissance de site 5' d'épissage par la snRNP U1. Les protéines SR SC35 et ASF/SF2 peuvent faciliter la communication au travers de l'exon en faisant le lien entre U2AF et la snRNP1. La fixation de SF1 au niveau du point de branchement facilite ensuite le recrutement de la snRNP U2 qui interagit également avec U2AF65. Les protéines SR permettraient également la communication entre les complexes définissant les exons 3' terminaux.

que l'interaction entre les facteurs reconnaissant les sites 3' et 5' d'épissage localisés de part et d'autre d'un exon favorise la mise en place du spliceosome (figure 5). La définition de l'exon implique que les exons aient une taille comprise entre 50 et 500 nt, limites au delà desquelles des contraintes stériques ne permettraient plus les interactions à travers ceux-ci (Berget, 1995; Sterner *et al.*, 1996). Selon ce modèle, la reconnaissance des sites d'épissage par le spliceosome est favorisée par la présence de séquences exoniques appelées ESE (*Exonic Splicing Enhancers*). La fixation de protéines SR sur ces ESE permet d'activer le recrutement des complexes d'épissage sur les sites 3' et 5' d'épissage (Graveley, 2000; Black, 2003).

Le modèle de la définition exonique s'applique également aux exons 5' et 3' terminaux bien que ceux-ci ne présentent pas de site 3' d'épissage et de site 5' d'épissage respectivement. La définition de l'exon 5' terminal repose sur une interaction entre le complexe associé à la coiffe de l'ARNpré-m (nommé CBC pour *Cap Binding Complex*) et les facteurs d'épissage reconnaissant le site 5' d'épissage (Lewis et Izaurralde, 1997). De la même manière, l'exon 3' terminal est défini par une interaction entre le complexe de clivage/polyadénylation (nommé CPC pour *Cleavage and Polyadenylation Complex*) mis en place sur l'exon 3' terminal et les facteurs d'épissage reconnaissant le site 3' d'épissage (Zhao *et al.*, 1999).

Une fois l'exon défini, une réorganisation des complexes est nécessaire pour que les facteurs reconnaissant deux exons distincts puissent interagir au travers de l'intron et permettre l'épissage. Le passage de la communication au travers de l'exon à une interaction au travers de l'intron reste largement méconnu. Les protéines SR pourraient participer au recrutement du complexe SRm160/300 qui serait impliqué dans la communication entre les complexes d'épissage à travers l'intron (Eldridge *et al.*, 1999; Blencowe *et al.*, 2000). A l'inverse nous verrons plus tard que la protéine PTB (*Polypyrimidine Tract Binding protein*) peut empêcher cette communication (Sharma *et al.*, 2008).

Nous pouvons noter que l'épissage *in vitro*, bien qu'utilisant des ARNpré-m et des extraits nucléaires de cellules eucaryotes, semble répondre au mécanisme de la définition de l'intron. En effet, la taille des introns doit être limitée à moins de 600 nt pour observer un épissage efficace *in vitro* à l'aide d'extraits nucléaires (Hernandez et Keller, 1983; Klinz et Gallwitz, 1985). De plus, il n'est pas nécessaire que le site 5' d'épissage de l'exon en 3' de l'ARNpré-m soit présent pour observer un épissage.

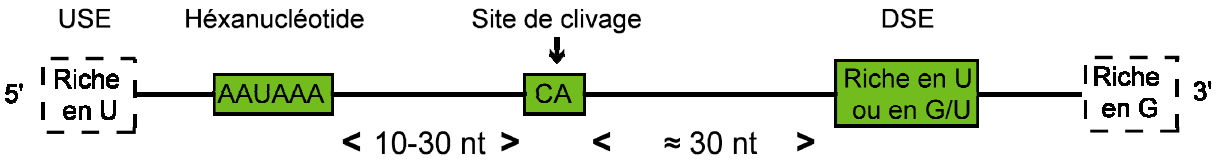


Figure 6- Signaux de clivage/polyadénylation chez les eucaryotes supérieurs. Les motifs conservés sont présentés en vert (DSE *Downstream Sequence Element*). Les séquences accessoires sont représentées en pointillés (*USE Upstream Sequence Element*). La distance optimale entre les différents signaux est également indiquée.

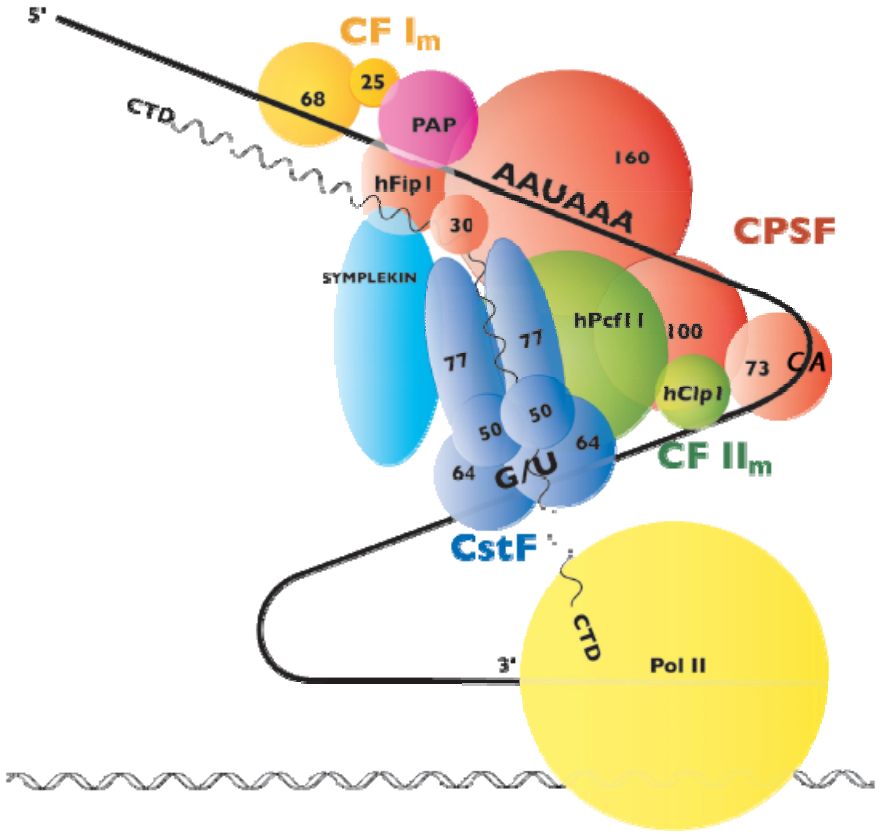


Figure 7- Complexe de clivage/polyadénylation chez les mammifères. D'après (Mandel *et al.*, 2008). Les différents facteurs du complexe de clivage/polyadénylation sont représentés. La séquence riche en GU (G/U), le site de clivage (CA) et l'hexanucléotide (AAUAAA) sont indiqués. Cette schématisation tient compte des interactions ARN/protéines et protéines/protéines actuellement décrites.

1.2. L'étape de clivage et de polyadenylation

La formation de l'extrémité 3' des ARNm est réalisée par deux réactions étroitement couplées *in vivo*. La première réaction est un clivage endonucléolytique de l'ARNpré-m. L'extrémité 3' ainsi générée sert de point d'ancrage à la seconde réaction : la polyadénylation nucléaire. Le processus de clivage/polyadénylation est également intimement lié à l'épissage et à la transcription. Ce processus peut affecter la terminaison de la transcription, le transport, et la stabilité des ARNm (Zhao *et al.*, 1999; Tian *et al.*, 2007; Mandel *et al.*, 2008; Moore et Proudfoot, 2009). Comme pour l'épissage, cette étape de maturation des ARNpré-m est guidée par un certain nombre d'éléments permettant l'assemblage de la machinerie de clivage et de polyadénylation.

1.1.1. Les sites de clivage et de polyadénylation

Le signal de clivage et de polyadénylation des eucaryotes supérieurs comporte trois motifs conservés (figure 6) :

- Un hexanucléotide est présent 10 à 30 nt en amont du site de clivage. Le motif le plus fréquemment rencontré est AAUAAA. Depuis longtemps identifié (Proudfoot et Brownlee, 1976), celui-ci est présent dans 53 % des ARNpré-m (Tian *et al.*, 2005).

- Un site de clivage, qui n'est pas strictement conservé. Chez les vertébrés, 70 % des sites de clivage sont localisés en 3' d'un résidu adénosine, avec une préférence nucléotidique A>U>C>>G (Sheets *et al.*, 1990). Le nucléotide précédant majoritairement le A est un résidu cytosine (Sheets *et al.*, 1990). Le site de clivage est donc le plus souvent constitué du dinucléotide CA (Zhao *et al.*, 1999), et le clivage a généralement lieu après le résidu adénosine (Sheets *et al.*, 1990).

- Une séquence DSE (*Downstream Sequence Element*) est située une trentaine de nucléotides en aval du site de clivage. Cet élément présente une faible conservation de séquence et est caractérisé par une richesse en U et/ou en GU.

Deux séquences auxiliaires permettant de stimuler la réaction de clivage/polyadénylation ont également été décrites. La première nommée USE (*Upstream Sequence Element*) est située en amont de l'hexanucléotide AAUAAA, et riche en U (Zhao *et al.*, 1999). La seconde est riche en G et est localisée en aval du DSE (Arhin *et al.*, 2002).

1.1.2. La machinerie de clivage et de polyadénylation

La machinerie de clivage et de polyadénylation est constituée uniquement de facteurs protéiques. Chez les mammifères, celle-ci s'organise autour de six éléments protéiques qui reconnaissent les séquences présentées précédemment et d'une protéine faisant la liaison entre les deux complexes principaux (figure 7).

- **CPSF** (*Cleavage and Polyadenylation Specificity Factor*) est composé de cinq sous-unités nommées CPSF30, CPSF73, CPSF100, CPSF160 et hFip1 (Kaufmann *et al.*, 2004). CPSF160 reconnaît l'hexanucléotide AAUAAA, interagit avec CstF et avec la poly(A) polymérase (PAP) (Murthy et Manley, 1995). CPSF73 est responsable de l'activité de clivage au niveau du dinucléotide CA de manière dépendante d'ions Zn^{2+} (Ryan *et al.*, 2004). La sous-unité hFip1 contacte CPSF160, PAP et l'ARNpré-m au niveau de l'USE, facilitant ainsi le recrutement de la PAP sur l'ARN. CPSF30 interagit avec PABPN1 (*Poly(A) Binding Protein Nuclear 1*) qui favorise la fixation de la PAP sur l'ARN. CPSF participe donc à l'étape de clivage ainsi qu'à celle de polyadénylation.

- **CstF** (*Cleavage stimulation Factor*) est composé de trois sous-unités de 50, 64 et 77 kDa. CstF64 interagit avec la séquence DSE (MacDonald *et al.*, 1994; Perez Canadillas et Varani, 2003) et Cstf 77 interagit avec la grosse sous-unité de CPSF (Murthy et Manley, 1995). CstF 50 interagit avec le domaine carboxy terminal (CTD) de l'ARN polymérase II (ARNpol II) permettant un lien entre la transcription et la maturation des ARNpré-m (McCracken *et al.*, 1997). CstF est impliqué dans le clivage et pourrait également participer à la polyadénylation quand un élément de type USE est présent dans l'ARNpré-m (Moreira *et al.*, 1998).

- **CF Im** (*Cleavage Factor Im*) est composé de trois polypeptides de 25, 59 et 68 kDa et d'un quatrième moins abondant de 72 kDa. CF Im est fonctionnel sous la forme d'un hétérodimère composé du polypeptide de 25 kDa et d'un des trois autres polypeptides. La fixation de CF Im sur l'ARN (Brown et Gilmartin, 2003) stabilise le complexe CPSF-CstF-PAP sur l'ARNpré-m (Ruegsegger *et al.*, 1996). De plus, le polypeptide de 25 kDa permettrait une interaction avec la PAP et avec PABPN1 (Dettwiler *et al.*, 2004). CF Im, est nécessaire pour le clivage, favorise la polyadénylation en stimulant la reconnaissance de l'hexanucléotide que sa séquence soit consensuelle ou non (Venkataraman *et al.*, 2005).

- **CF IIm** (*Cleavage Factor IIm*) est composé de deux sous-facteurs CF IIAm et CF IIBm. CF IIAm contient deux sous-unités : hPcf11 interagit avec le CTD de l'ARNpol II (Barilla *et al.*, 2001), et hClp1 qui, tout comme hPcf11, interagit avec CF Im et CPSF (de

Vries *et al.*, 2000; Gross et Moore, 2001). Le facteur CF IIAm est essentiel à la réaction de clivage, par contre, le rôle du facteur CF IIBm n'est pas encore déterminé.

- La protéine **symplekin** interagit à la fois avec CPSF et CstF permettant ainsi de favoriser l'assemblage de la machinerie de clivage/polyadénylation (Takagaki et Manley, 2000).

- **PAP** est l'enzyme catalysant la formation de la queue poly(A) à l'extrémité 3' de l'ARNpré-m. Son interaction avec CPSF160 et hFip1 permet sa fixation sur l'ARN (Kaufmann *et al.*, 2004).- **PABPN1** est également appelée PABP2. Cette protéine n'est pas indispensable à la polyadénylation mais stimule cette réaction (Kerwitz *et al.*, 2003) et régule la longueur de la queue poly(A) ajoutée à l'ARNm (Bienroth *et al.*, 1993; Minvielle-Sebastia *et al.*, 1997). Au cours de l'élongation de la queue poly(A), PABPN1 stabiliserait l'interaction entre la PAP et CPSF, tout en maintenant CPSF au niveau de l'hexanucléotide AAUAAA. Le maintient de ce complexe tripartite est nécessaire à l'ajout des résidus adénosines sur l'ARNm (Kuehn *et al.*, 2009).

1.1.3. La réaction de clivage et de polyadénylation

Il existe une coopérativité de fixation des différents facteurs de la machinerie de clivage/polyadénylation qui permet une stabilisation globale du complexe sur l'ARNpré-m. Il n'est donc pas aisé d'appréhender l'ordre réel de l'arrivée de chaque facteur sur l'ARNpré-m. Les facteurs CF Im et CPSF ont tous deux été identifiés à partir de la purification de complexes d'épissage (Rappsilber *et al.*, 2002), ces résultats suggèrent qu'ils sont tous deux associés précocement à l'ARNpré-m. Toutefois, la haute affinité de CF Im pour l'ARN et sa capacité à accélérer l'assemblage de la machinerie sur l'ARNpré-m indique qu'il serait le premier à se lier (Ruegsegger *et al.*, 1998). L'assemblage serait ensuite rapidement complété par la fixation de CPSF au niveau de l'hexanucléotide AAUAAA. Ensuite CstF s'associe au niveau de la séquence DSE (Ruegsegger *et al.*, 1998), et dans le même temps CF IIm rejoint le complexe (Nag *et al.*, 2007). Les interactions de ces facteurs entre eux et de la simplekin avec CPSF et CstF permettent de stabiliser ce complexe. Enfin CPSF favoriserait le recrutement de la PAP (Kaufmann *et al.*, 2004).

Le complexe ainsi formé est positionné de part et d'autre du site de clivage et permet à CPSF73 de catalyser le clivage endonucléolytique de l'ARNpré-m (Ryan *et al.*, 2004). Deux fragments d'ARN sont alors générés. Le fragment aval possédant une extrémité 5' phosphate est dégradé par l'exonucléase 5' → 3' Xrn2 (West *et al.*, 2004), et le fragment amont

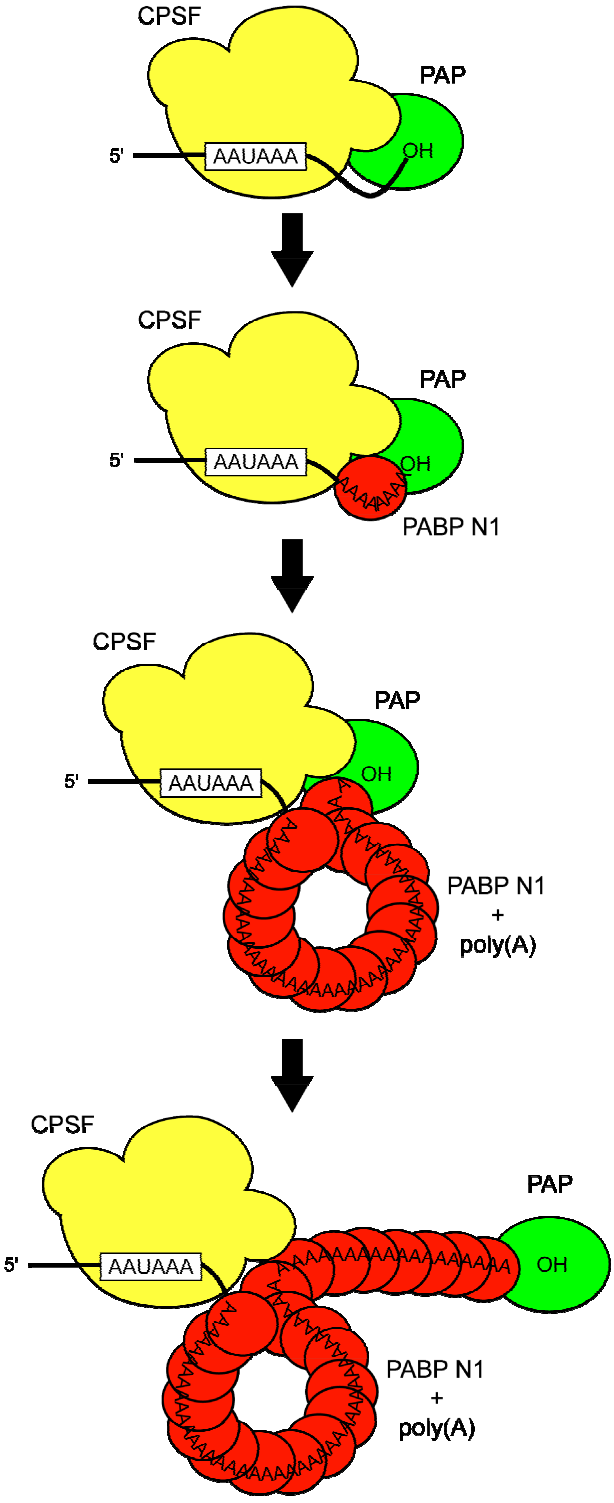


Figure 8- Modèle d'allongement de la queue poly(A) sur les ARNpré-m de mammifères. D'après (Kuehn *et al.*, 2009). L'allongement de le queue poly(A) est possible tant que le complexe CPSF-PAP-PABPN1 reste formé. L'hexanucléotide AAUAAA est représenté.

présentant une extrémité 3'-OH est polyadénylé. Seuls les facteurs CPSF, PAP et PABN1 sont nécessaires à la polyadénylation (Kaufmann *et al.*, 2004). Suite au clivage endonucléolytique, un remaniement du complexe est effectué s'accompagnant probablement du départ des facteurs CstF, CF Im et CF IIm. L'interaction simultanée de PAP avec les facteurs CPSF et PABPN1 permet la processivité de la réaction de polyadénylation (Kuehn *et al.*, 2009). Comme présenté dans la figure 8 une boucle de poly(A) recouverte de PABPN1 est alors formée. Quand la queue poly(A) atteint 250 résidus, la structure est déstabilisée et conduit à la rupture de la liaison entre CPSF et la PAP (Kuehn *et al.*, 2009). La fin de cette interaction marque l'arrêt de l'élongation.

Le processus de clivage/polyadénylation influence l'arrêt de la transcription par l'ARNpol II. Le couplage entre ces deux mécanismes sera présenté dans le dernier chapitre de cette introduction.

2. La maturation alternative des ARN pré-messagers

L'épissage alternatif est défini comme la sélection différentielle des sites d'épissage ou de polyadénylation et permet, à partir d'une séquence génomique unique, de produire plusieurs ARNm correspondant à des protéines distinctes selon le tissu ou le stade de développement. Considéré comme exceptionnel dans les années 1990 (Sharp, 1994), il est aujourd'hui admis que la maturation alternative des transcrits concerne la majorité des ARNpré-m des eucaryotes supérieurs. Chez l'homme les analyses d'EST ou par puces à ADN indiquent que 40% à 75% des gènes seraient épissés de manière alternative (Lander *et al.*, 2001; Modrek et Lee, 2002; Johnson *et al.*, 2003; Clark *et al.*, 2007), 29% à 54% seraient concernés par la polyadénylation alternative (Beaudoing *et al.*, 2000; Tian *et al.*, 2005; Zhang *et al.*, 2005), et 20% des ARNpré-m humains contiendraient des exons 3' terminaux alternatifs, qui peuvent être produits à la fois par épissage alternatif et polyadénylation alternative (Yan et Marr, 2005; Tian *et al.*, 2007).

La maturation alternative des ARNpré-m permet non seulement un contrôle qualitatif de l'expression des gènes, en modulant les propriétés d'interaction, la localisation intracellulaire, l'activité enzymatique, la stabilité d'un grand nombre de protéines (Stamm *et al.*, 2005), mais aussi un contrôle quantitatif. Le contrôle quantitatif de l'expression des gènes

par l'épissage alternatif peut être du à la modification des régions 5' et 3' non traduites des ARNm, orientant ainsi leur stabilité, leur localisation ou leur traductibilité, ou encore à l'introduction de codons stop précoces, dirigeant ainsi les ARNm vers les mécanismes de dégradation (Lejeune et Maquat, 2005; Stamm *et al.*, 2005).

Les ARNm qui possèdent des codons stop prématurés sont dirigés vers la voie de dégradation du NMD (*Nonsense-Mediated Decay*) (Chang *et al.*, 2007; Shyu *et al.*, 2008). Au cours de la maturation des ARNpré-m, la machinerie d'épissage dépose un complexe multi-protéique, nommé EJC (*Exon Junction Complex*) qui permet de « marquer » les jonctions exon-exon de l'ARNm (Le Hir *et al.*, 2000). Ce complexe accompagne l'ARNm dans le cytoplasme (Le Hir *et al.*, 2001) jusqu'à sa traduction en protéine (Le Hir et Seraphin, 2008). Le mécanisme de NMD est activé si la machinerie traductionnelle rencontre un codon stop placé à plus de 50 à 55 nt en amont d'une jonction exon-exon marquée par un EJC. Ce mécanisme de contrôle permet donc d'éviter la synthèse de protéines tronquées qui ne seraient pas fonctionnelles. Celui-ci permet également de réguler de manière intégrée l'expression de certains gènes (McGlincy et Smith, 2008). Dans ce cas, des évènements d'épissage alternatifs conservés permettent l'introduction de codons stop précoces, ce qui permettrait ainsi d'exploiter le NMD pour réguler quantitativement l'expression des gènes au niveau post-transcriptionnel. Un tel mécanisme est impliqué dans la régulation de la production de protéines régulatrices de l'épissage. Par exemple, la protéine PTB, autorégule l'épissage alternatif de son propre ARNpré-m (Wollerton *et al.*, 2004). Lorsque la quantité de PTB est importante, celle-ci se fixe sur l'ARNpré-m et favorise l'exclusion de l'exon 11, conduisant à l'apparition d'un codon stop prématuré dans l'ARNm qui est dégradé par la voie du NMD (Wollerton *et al.*, 2004). Une étude considérant les évènements d'épissage alternatif de type saut d'exon conservés entre l'homme et la souris, montre que 24% des isoformes analysées seraient dirigées vers la voie de NMD. Les séquences contrôlant ces évènements d'épissage alternatif sont conservées de la souris à l'homme, indiquant qu'il s'agit d'un mécanisme actif et non d'erreurs d'épissage (Baek et Green, 2005; McGlincy et Smith, 2008).

Dans cette partie de l'introduction les différents mécanismes permettant la régulation de l'épissage alternatif seront présentés. Les régulations conduisant à la sélection différentielle des exons 3' seront décrites dans la troisième partie de l'introduction. Enfin le couplage des mécanismes d'épissage et de clivage/polyadénylation avec la transcription sera

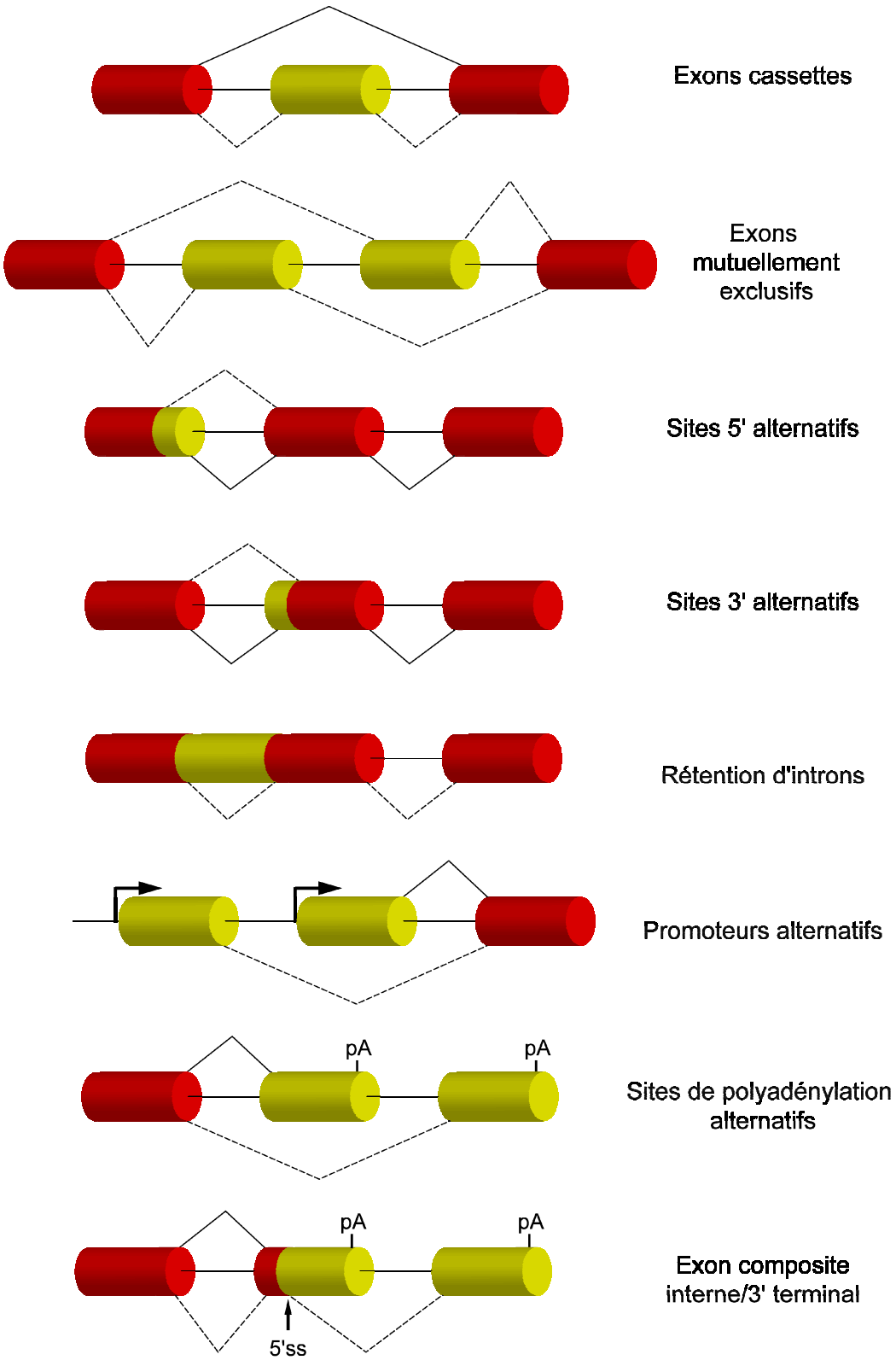


Figure 9- Les différents types d'épissage alternatif. Les exons constitutifs sont représentés en rouge et les exons alternatifs en jaune. Les traits pleins et pointillés indiquent les différents évènements d'épissage. pA : signal de polyadénylation ; 5'ss : site 5' d'épissage ; flèches : promoteurs.

abordé. Dans un premier temps les différents types de maturation alternative des ARNpré-m sont présentés.

2.1. Les différents types d'épissage alternatif

Chez les eucaryotes supérieurs plusieurs types d'exons alternatifs ont été décrits (figure 9)

- **Les exons cassettes.** Les exons cassettes sont soit inclus soit exclus de l'ARNm par la machinerie d'épissage. Ces exons sont généralement de petite taille comme l'exon N1 de l'ARNpré-m c-src d'une longueur de 18 nucléotides et qui est spécifiquement inclus dans les cellules neuronales (Black, 1992).

- **Les exons mutuellement exclusifs.** Les exons mutuellement exclusifs ne sont pas présents simultanément dans un ARNm. L'ARNpré-m de la β-tropomyosine possède dans sa région centrale l'exon 6A qui est utilisé dans les cellules non musculaires, tandis que l'exon 6B est inclus dans l'ARNm des cellules musculaires (Libri *et al.*, 1989). Le gène de l'α-tropomyosine posséde également ce type d'exons alternatifs. Ainsi, l'exon 2 est inclus spécifiquement dans les muscles lisses alors que l'exon 3 est utilisé dans les ARNm de tous les autres tissus (Mullen *et al.*, 1991).

- **Les exons définis par des sites 5' ou 3' d'épissage alternatifs.** Les exons peuvent posséder deux sites 5' ou deux sites 3' d'épissage qui sont en compétition pour être reconnus. Ainsi, selon le site d'épissage sélectionné par la machinerie d'épissage, l'exon pourra être allongé ou raccourci. Ce type de combinatoire entre des sites 5' et 3' alternatif est fréquemment retrouvé chez les virus (Caputi *et al.*, 1999; Ropers *et al.*, 2004). Il est également présent chez l'homme comme dans l'ARNpré-m de Bcl-x où deux sites 5' d'épissage sont présents (Boise *et al.*, 1993) et dans l'exon 9 de l'ARNpré-m de la PTB qui contient deux sites 3' d'épissage alternatifs (Wollerton *et al.*, 2001).

- **La rétention d'intron.** Certains sites d'épissage peuvent être « ignorés » par la machinerie d'épissage, conduisant ainsi à la conservation d'un intron dans l'ARNm. Ce type d'épissage est observé dans l'ARNpré-m de l'élément P chez la drosophile où l'intron 3 est retenu dans l'ARNm dans les cellules somatiques alors qu'il est exclu dans les cellules germinales (Rio *et al.*, 1986).

- **L'utilisation de promoteurs alternatifs.** Un cas « extrême » d'utilisation de promoteur alternatif est observé pour les gènes de la neurexine où une combinaison d'utilisation alternative de promoteurs et d'évènements d'épissage alternatifs pourrait conduire à la formation de 3000 ARNm (Grabowski et Black, 2001; Rowen *et al.*, 2002). Le gène de l'α-tropomyosine possède deux promoteurs alternatifs localisés sur les exons 1A et 1B.

- **Les sites de polyadénylation alternatifs.** Un ARNm peut contenir plusieurs exons terminaux qui par définition seront mutuellement exclusifs. Les ARNm ainsi produits différeront par leur région 3' terminale pouvant conduire à des différences de stabilité ou de localisation de l'ARNm. L'ARNpré-m de CT/CGRP (*Calcitonin/Calcitonin-Gene-Related Peptide*) possède un site de polyadénylation dans l'exon 4 qui est spécifiquement utilisé dans les cellules thyroïdiennes alors que le site de polyadénylation présent dans l'exon 6 est sélectionné dans les cellules neuronales (Lou *et al.*, 1994). Les exons 9B et 9D du gène de l'α-tropomyosine répondent également à ce type de sélection.

- **Les exons composites internes/terminaux.** Ces exons contiennent un site 5' d'épissage interne suivi d'un site de polyadénylation ce qui leur permet d'être utilisé soit comme exon interne lorsque le site 5' d'épissage est reconnu, soit comme exon 3' terminal quand le signal de polyadénylation est sélectionné. Ce profil d'épissage est observé pour l'exon 9A9' de l'ARNpré-m de l'α-tropomyosine, mais aussi dans l'ARNpré-m des chaînes lourdes des immunoglobulines (Tsurushita *et al.*, 1987).

L'analyse des différents types d'épissage alternatif montre que les exons cassettes sont présents dans 49% des ARNpré-m humains. Les exons présentant des sites 5' ou 3' d'épissage alternatifs sont présents respectivement dans 26% et 25% des transcrits (de la Grange *et al.*, 2005). Les sites de polyadénylation alternatifs sont présents dans 20% des ARNpré-m (Tian *et al.*, 2007) et les introns sont retenus dans 15% des ARNpré-m (de la Grange *et al.*, 2005). 25% des gènes contiennent plus d'une région épissée de manière alternative (Fededa *et al.*, 2005), et en moyenne il est estimé qu'un gène humain code pour deux à trois ARNm différents. Un même ARNpré-m peut donc présenter plusieurs types d'épissage alternatifs, c'est le cas du gène de l'α-tropomyosine qui présente des promoteurs alternatifs, un couple d'exons mutuellement exclusifs et des sites de polyadénylation alternatifs.

2.2. Analyse globale des évènements d'épissage alternatif

Des approches permettant l'analyse des évènements d'épissage alternatif à grande échelle ont été développées afin de comprendre le rôle biologique de l'épissage alternatif. Ces études permettent de caractériser des séquences ou des facteurs régulateurs de certains évènements d'épissage, de mettre en évidence des corégulations d'épissages alternatifs et d'intégrer les différentes étapes de la maturation des ARNpré-m.

Des puces à ADN permettant l'étude à grande échelle des évènements d'épissage reposent sur l'utilisation d'oligonucléotides ciblant les exons et les jonctions exon-exon (Pan *et al.*, 2004; Fehlbaum *et al.*, 2005; Pan *et al.*, 2006; Ben-Dov *et al.*, 2008). La comparaison simultanée des profils d'épissage de plusieurs milliers d'exons dans différents tissus ou lignées cellulaires a permis de dégager des caractéristiques générales de l'épissage alternatif. Ainsi les gènes dont l'expression est régulée par épissage alternatif sont majoritairement impliqués dans la communication cellulaire et la régulation d'enzymes (Johnson *et al.*, 2003). Des puces dédiées à l'épissage alternatif peuvent également être utilisées pour étudier l'implication physiologique d'un facteur particulier dans la régulation de l'épissage alternatif. Ce type d'analyse a été effectué avec les protéines régulatrices ASF/SF2, hrp48, PSI, SRp55/B52 dans des cellules de drosophile (Blanchette *et al.*, 2005). Une approche identique a été réalisée afin de déterminer les cibles et les divers rôles de la protéine hnRNP L (Hung *et al.*, 2008).

Des approches in vitro de type SELEX (*Systematic Evolution of Ligand by Exponential Enrichment*) ont été réalisées avec diverses protéines SR et hnRNP. Ces expériences ont permis de caractériser des sites optimaux de fixation de ces protéines, qui seront présentés dans la suite de cette introduction. Des approches de SELEX fonctionnel dans lesquelles les ESE ont été sélectionnés dans des extraits cytoplasmiques de S100 en présence de protéines SR individuelles ont permis la recherche de motifs ESE plus spécifiques de chaque protéine SR. Ce type d'expérience a permis de déterminer des motifs de fixation indépendants pour les protéines ASF/SF2, SC35, SRp40, et SRp55 (Liu *et al.*, 1998). Ces travaux ont permis le développement du programme ESEfinder (http://rulai.cshl.edu/tools/ESE/) (Cartegni *et al.*, 2003) qui permet la recherche de motifs ESE putatifs de ces protéines au sein de séquences d'intérêt.

Des approches fonctionnelles basées sur la transfection de cellules en culture ont été développées afin de déterminer la séquence d'ESE actifs. Suite à la transfection d'un minigène dans lequel un ESE nécessaire à l'épissage d'un exon interne est remplacé par une séquence aléatoire de treize nucléotides, les ESE actifs sont caractérisés par RT-PCR et séquençage de l'exon inclus (Coulter *et al.*, 1997). Par ailleurs une autre méthode, également basée sur la transfection de cellules en culture, a été développée pour identifier des séquences exoniques inhibitrices (ESS) (Wang *et al.*, 2004). Cette approche est basée sur la transfection d'un minigène contenant trois exons, l'exclusion de l'exon deux permettant la synthèse de la protéine eGFP (*enhanced Grenn Fluorescent Protein*). Les cellules produisant de la eGFP fonctionnelle sont triées par FACS (*Fluorescence-Activated Cell Sorting*), puis leur exon deux est amplifié par PCR et séquencé. Certains des ESS putatifs ainsi caractériés sont similaires aux sites de liaison définis pour les protéines hnRNP A1 et H.

Une méthode nommée CLIP (*Cross-Linking and ImmunoPrecipitation*) a été mise au point par Ule et collaborateurs pour identifier les ARNpré-m cibles de facteurs régulateurs sans *a priori* de séquence (Ule *et al.*, 2003). Cette méthodologie repose sur un pontage covalent *in vivo* aux UV de la protéine sur ses ARN cibles suivi d'une immunoprécipitation des complexes ARN-protéine. Les ARN sont isolés, partiellement dégradés, puis les fragments d'ARN liés à la protéine sont extraits et rétro-transcrits. Les ADNc ainsi obtenus sont amplifiés par PCR puis séquencés afin d'identifier les ARN liés à la protéine. Cette stratégie a permis d'obtenir 340 séquences ARN fixées par la protéine NOVA-1 dans le cerveau de souris correspondant pour certaines à des ARN impliqués dans le fonctionnement synaptique (Ule *et al.*, 2003; Ule *et al.*, 2005). Cette stratégie permet donc d'identifier de nouveaux ARNpré-m cibles d'un facteur régulateur donné, mais également de déterminer les fonctions cellulaires régulées par la protéine d'intérêt.

Ces approches permettent de déterminer les ARN associés à une ou plusieurs protéines régulatrices, et de mettre en évidence les réseaux d'ARN sous le contrôle d'une même protéine. D'autres stratégies ont été développées pour déterminer les protéines recrutées par des fragments d'ARN. Les différentes méthodes ainsi développées seront présentées dans le deuxième chapitre des résultats.

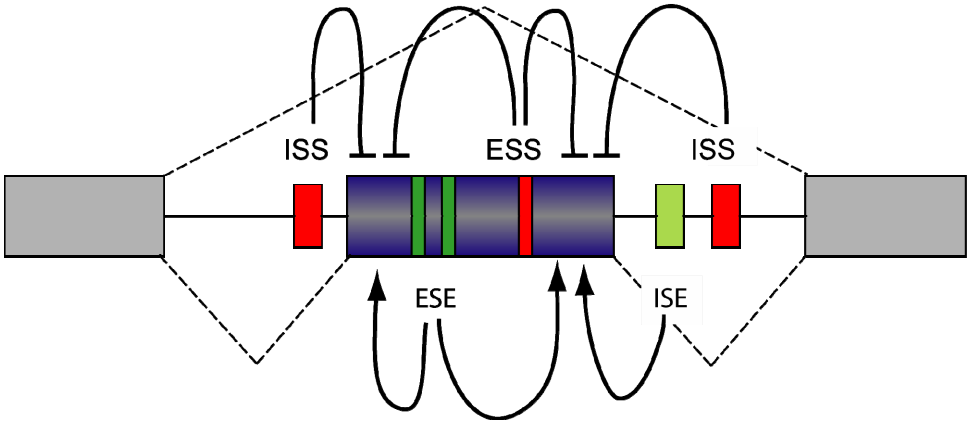


Figure 10- Séquences *cis* régulatrices impliquées dans la régulation de l'épissage alternatif. D'après (Matlin *et al.*, 2005). En plus de séquences d'épissage consensuelles des sites d'épissage, des séquences régulatrices sont présentes dans les introns ou dans les exons, et sont activatrices ou inhibitrices. ESE séquence exonique activatrice, ESS séquence exonique inhibitrice, ISE séquence intronique activatrice, ISS séquence intronique inhibitrice.

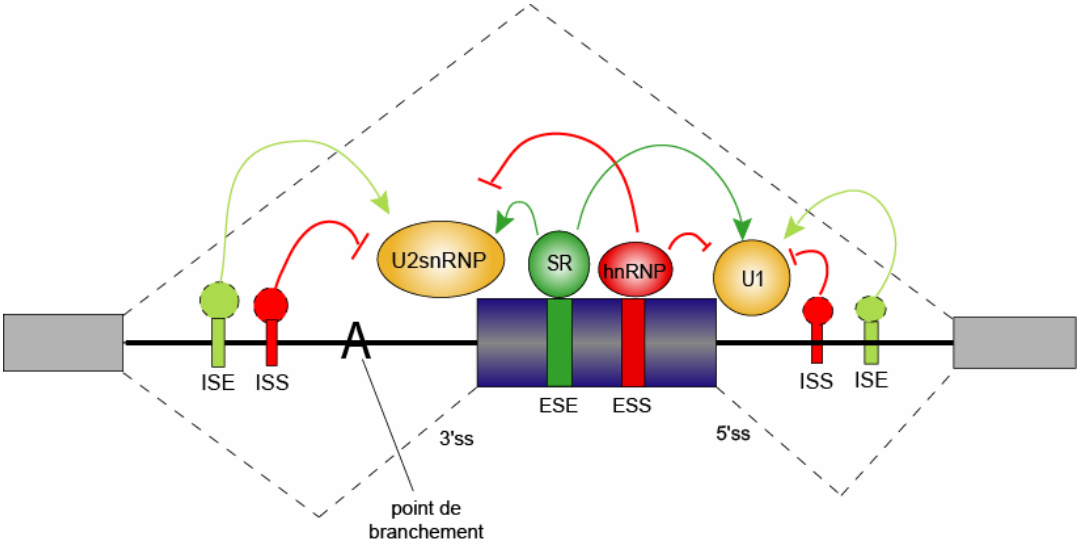


Figure 11- Séquences *cis* régulatrices impliquées dans la régulation de l'épissage alternatif. D'après (Wang et Burge, 2008). Le contrôle combinatoire des séquences *cis* régulatrices et des facteurs régulateurs permet la modulation de la reconnaissance des exons par les facteurs généraux d'épissage. Le point de branchement (A), les sites 5' et 3' d'épissage sont indiqués.

2.3. Les régulations de l'épissage alternatif

Comme de nombreux processus de régulation de l'expression génique, le contrôle de l'épissage alternatif est multifactoriel. En effet, la régulation de l'épissage alternatif est le fruit de l'interaction entre des facteurs régulateurs agissant en *trans* et des séquences régulatrices agissant en *cis* présentes sur l'ARNpré-m (Smith et Valcarcel, 2000; Hertel, 2008; House et Lynch, 2008; Wang et Burge, 2008).

2.3.1. Les éléments régulateurs agissant en *cis*

Les exons alternatifs sont le plus souvent délimités par des sites d'épissage et de polyadénylation qui différent quelque peu des séquences consensuelles caractérisées pour les exons constitutifs. Les sites d'épissage des exons alternatifs sont donc considérés comme faibles ou sous optimaux. Il y a également une différence de longueur entre les exons constitutifs et alternatifs. En effet, certains exons alternatifs sont courts par rapport à la taille moyenne des exons constitutifs (145 nt). Par exemple, l'exon N1 du gène c-src de souris n'est constitué que de 18 nucléotides. Si l'on considère une reconnaissance des sites d'épissage basée sur le modèle de la définition de l'exon, l'exon N1 ne sera pas défini. Une conséquence de la dégénérescence des sites d'épissage est la présence de nombreux sites cryptiques d'épissage particulièrement présents dans les introns. Au regard de ces caractéristiques, comment le spliceosome choisit-il les bons sites d'épissage ? Des séquences régulatrices additionnelles présentes dans l'ARNpré-m et permettant de moduler la sélection des sites d'épissage des exons alternatifs ont été caractérisées. Globalement ces séquences agissent précocement lors de la mise en place du complexe d'épissage, en favorisant ou en inhibant le recrutement des facteurs généraux d'épissage sur l'ARNpré-m. Celles-ci sont répertoriées selon leur localisation intronique ou exonique et selon leur fonction activatrice ou inhibitrice. Ces séquences sont réparties en quatre catégories (figure 10) : les séquences exoniques activatrices ou ESE, les séquences exoniques inhibitrices ou ESS, les séquences introniques activatrices ou ISE et enfin les séquences introniques inhibitrices ou ISS. Ces séquences régulatrices sont variables au niveau de leur taille et du contenu de leur séquence (Thanaraj et Stamm, 2003). L'étude des éléments exoniques a permis de montrer que les ESE sont généralement des sites de liaison des protéines SR (Graveley, 2000) et les ESS sont souvent reconnus par des protéines de la famille des hnRNP (Pozzoli et Sironi, 2005). Les éléments introniques ont été peu étudiés par des approches systématiques, les données actuellement

disponibles sont donc dues à l'analyse spécifique de différentes séquences ou de différentes protéines. Les ISS sont reconnus par des protéines de la famille des hnRNP, comme les protéines hnRNP I (ou PTB), hnRNP A1 ou hnRNP L (Hui *et al.*, 2005; Matlin *et al.*, 2005). Tout comme les ISS, les ISE sont reconnus par la protéines hnRNP L (Hung *et al.*, 2008), et par d'autres protéines de la familles des hnRNP. D'autres facteurs sont également recrutés au niveau des séquences introniques activatrices comme les facteurs Fox et Nova qui seront présentés ultérieurement (Ule *et al.*, 2003; Nakahata et Kawamoto, 2005; Underwood *et al.*, 2005).

La majorité des exons alternatifs possède plusieurs séquences régulatrices, activatrices ou inhibitrices, celles-ci étant souvent chevauchantes. Cette organisation complexe présentée dans la figure 10, influe fortement sur la régulation de l'épissage alternatif en mettant en compétition les séquences régulatrices entre elles. La fixation sur ces différents motifs de facteurs régulateurs va permettre de moduler la reconnaissance des exons. Cette combinatoire implique donc des complexes RNP régulateurs dont la composition oriente la maturation des ARNpré-m, on parle alors de code RNP (Wang et Burge, 2008)

La structure de l'ARNpré-m est également un facteur qui peut influencer le choix des sites d'épissage car elles peuvent interférer directement avec la reconnaissance des exons par la machinerie d'épissage. En effet, la formation d'une structure en tige-boucle dans l'ARNpré-m peut favoriser le rapprochement entre le point de branchement et le site 3' d'épissage comme c'est le cas dans l'ARNpré-m de l'adénovirus E1A (Chebli *et al.*, 1989) ou pour l'intron AG2 de l'actine chez la levure (Deshler et Rossi, 1991). Ces structures peuvent empêcher la reconnaissance du site 5' d'épissage de l'exon 7B de l'ARNpré-m de hnRNP A1 (Blanchette et Chabot, 1997), ou de l'exon 10 de l'ARNpré-m de Tau (Grover *et al.*, 1999; Varani *et al.*, 1999). Les structures double brin peuvent également réguler indirectement l'épissage alternatif en prévenant la fixation des facteurs régulateurs sur les séquences ESE, ESS, ISE, ISS. Par exemple, l'exon EDA de la fibronectine humaine est régulé par un ESE et un ESS. La mutation de l'ESS conduit à la formation d'une structure en tige-boucle emprisonnant l'ESE, empêchant ainsi son interaction avec les protéines SR (Buratti *et al.*, 2004). Une étude récente combinant l'analyse structurale des séquences régulatrices ESE, ESS, ISE, ISS validées expérimentalement et des transfections dans des cellules en culture, montre que les éléments régulateurs sont plus structurés en simple brin que les séquences contrôle (Hiller *et al.*, 2007), favorisant ainsi leur reconnaissance par des facteurs *trans*

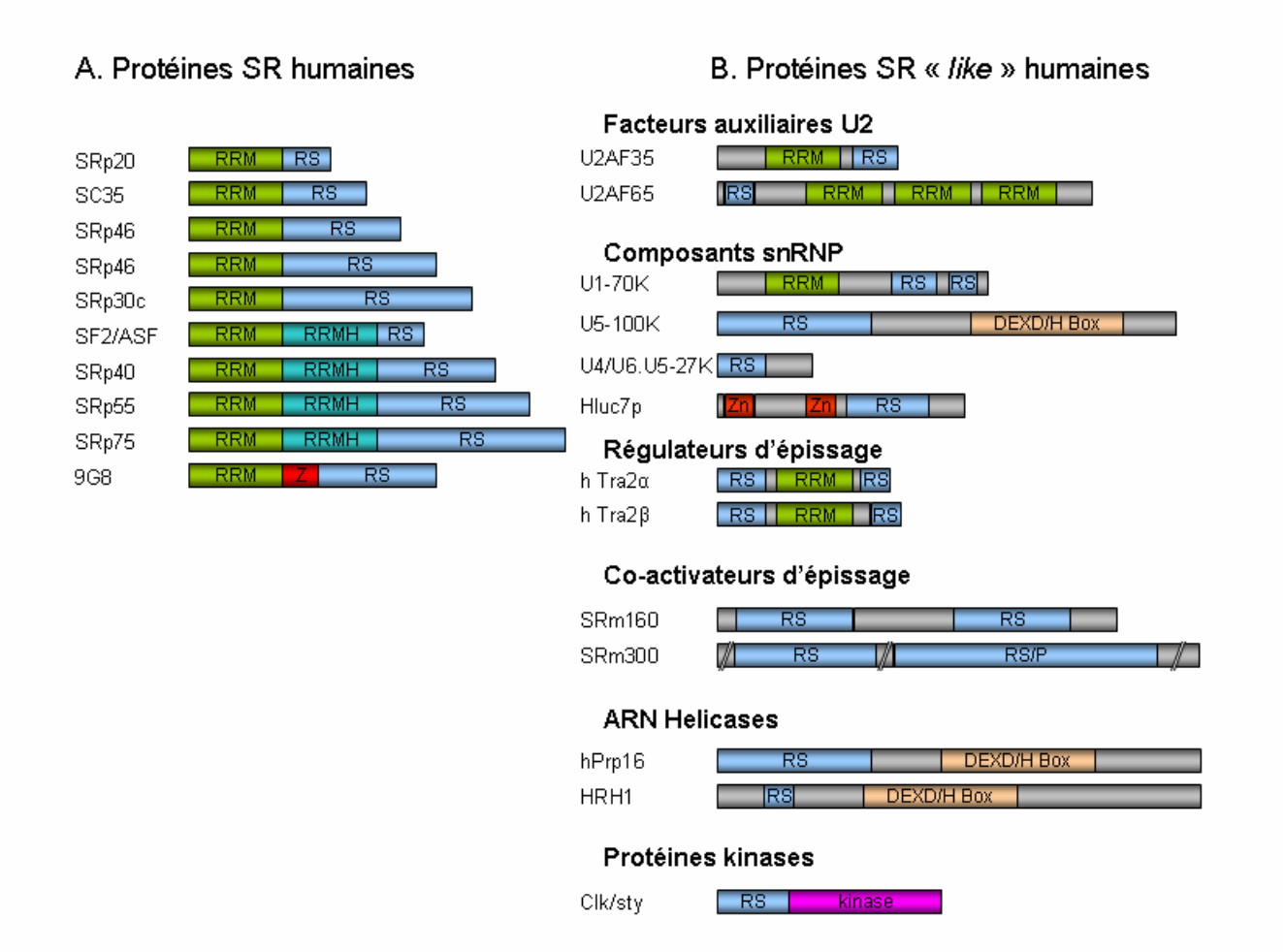


Figure 12- Représentation schématique des protéines SR humaines et des protéines apparentées. Selon (Graveley, 2000). **A.** Domaines structuraux des protéines SR. Motifs de liaison aux ARN (RRM) ou homologues (RRMH), domaines en doigt de zinc (Z) et domaines riches en arginine et en sérine (RS). **B.** Présentation des protéines apparentées aux protéines SR. Motif caractéristique des ARN hélicases (DEX/H Box).

régulateurs. Ces données mettent en avant l'importance de la structure des ARNpré-m dans la régulation des évènements de maturation alternatifs.

2.3.2. Les éléments régulateurs agissant en *trans*

L'interaction entre des facteurs protéiques et les séquences *cis* régulatrices permet un contrôle combinatoire des évènements d'épissage alternatif (figure 11). Deux grandes familles de protéines sont impliquées dans la régulation de l'épissage alternatif : la famille des protéines SR et la famille des protéines hnRNP.

2.3.2.1. Les protéines SR

Les protéines SR précédemment introduites du fait de leur implication dans l'épissage constitutif, sont également impliquées dans la régulation de l'épissage alternatif. Celles-ci sont considérées essentiellement comme des facteurs activateurs qui interagissent principalement avec les ESE. Cette famille est constitué de dix membres, présentant une structure modulaire : un ou deux RRM dans la région amino terminale et un domaine RS dans la région carboxy terminale (figure 12 ; tableau 1). La phosphorylation des sérines du domaine RS permet de moduler la localisation et la fonction des protéines SR. Les protéines SR sont localisées avec d'autres facteurs d'épissage dans des structures nucléaires appelées « speckles ». Quand la transcription est active ces protéines migrent des foyers de transcription (Caceres *et al.*, 1997; Misteli *et al.*, 1997; Misteli et Spector, 1999). Certaines protéines SR sont capables de faire la navette entre le noyau et le cytoplasme (Caceres *et al.*, 1998), c'est le cas de la protéine ASF/SF2 qui coordonnerait les étapes nucléaires et cytoplasmiques de la régulation de l'expression des gènes (Sanford *et al.*, 2008). En effet, la fixation de ASF/SF2 sur différents ARNm va permettre leur adressage dans les polysomes augmentant ainsi leur traduction.

D'autres protéines distinctes des protéines SR contiennent des domaines RS et sont impliquées dans diverses étapes de la maturation des ARNpré-m. Ces protéines dites SR « like » peuvent être différenciées des protéines SR car elles ne possèdent pas strictement un domaine RRM en position amino terminale et un domaine RS en position carboxy terminale, elles sont moins riches en dipeptide sérine/arginine dans leur domaine RS, elles présentent une quantité supérieure de dipeptides glutamate/aspartate et lysine/arginine, et enfin elles ne

hnRNP	Domaines	Séquences	Méthodes
SF2/ASF	2xRRM, RS	RGAAGAAC	SELEX
		AGGACRRAGC	SELEX
		SRSASGA	SELEX fonctionnel
		UGRWG	CLIP
SC35	1xRRM, RS	AGSAGAGUA	SELEX
		GUUCGAGUA	SELEX
		GRYYCSYR	SELEX fonctionnel
		UGUUCSAGWU	SELEX
		GWUWCCUGCUA	SELEX
		GGGUAUGCUG	SELEX
		GAGCAGUAGKS	SELEX
		AGGAGAU	SELEX
		UGCNGYY	SELEX fonctionnel
SRp20	1xRRM, RS	GGUCCUCUUC	Gel Shift
		WCWWC	Test dépissage
		CUCKUCY	Affinité à l'ARN
SRp30c	2xRRM, RS	GACGAC	SELEX fonctionnel
		AAAGAGCUCGG	SELEX fonctionnel
		CUGGAUU	Gel Shift
SRp40	2xRRM, RS	GAGCAGUCGGCUC	SELEX
		ACDGS	SELEX fonctionnel
SRp55	2xRRM, RS	USCGKM	SELEX fonctionnel
		UCAACCAGGCGAC	SELEX
SRp75	2xRRM, RS	GAAGGA	UV cross-linking
9G8	1xRRM, RS	UCAACA	UV cross-linking
		ACGAGAGAY	SELEX
		GGACGACGAG	SELEX fonctionnel
SRm160	2xRS	purines riche (GAA)n	Test dépissage

Tableau 1- Domaines et séquences de liaison des protéines SR. D'après (Long et Caceres, 2009). R :purines A ou G. Y : pyrimidines C ou U. S : G ou C. M : A ou C. K : G ou U. W : A ou U. D : A, G, U.N : A, C, G, U. Quand une séquence est répétée plusieurs fois, elle est suivie de n. RRM : *RNA Recognition Motif*, RS : domaine riche en dipeptide arginine-sérine. Les méthodes ayant permis l'identification des séquences sont indiquées. « Gel shift » correspond à des expériences de retard de migration sur gel, et « UV cross-linking » signifie transfert de marquage aux UV.

sont pas capables de complémenter un extrait S100 afin de le rendre compétent pour l'épissage (Bourgeois *et al.*, 2004). Parmi ces protéines apparentées aux protéines SR, on retrouve les deux sous-unités de l'hétérodimère U2AF qui est impliqué dans la reconnaissance du point de branchement, également la protéine SRm160 qui active l'épissage et qui favorise l'utilisation des exons 3' terminaux (McCracken *et al.*, 2002; Fortes *et al.*, 2007).

Fonctions des protéines SR dans l'épissage alternatif

Contrairement à leur redondance fonctionnelle décrite pour l'épissage constitutif, les protéines SR peuvent réguler individuellement des évènements d'épissage alternatif. Ainsi, la protéine ASF/SF2 est essentielle à la viabilité des cellules de poulet en culture, sa déplétion ne peut pas être restaurée par les protéines SC35 ou SRp40 (Wang *et al.*, 1996). La spécificité d'action des protéines SR a également été étudiée *in vivo* chez la souris (Moroy et Heyd, 2007). En effet, l'extinction de l'expression de SC35 ou de ASF/SF2 spécifiquement dans le cœur, ne conduit pas à une modification générale de l'épissage alternatif, mais à un changement de maturation d'ARNpré-m particuliers (Ding *et al.*, 2004; Xu *et al.*, 2005). D'un point de vue physiologique, la protéine ASF/SF2 est nécessaire au développement post-natal du cœur chez la souris (Xu *et al.*, 2005) et SC35 à la contraction cardiaque (Ding *et al.*, 2004). SC35 a également un effet spécifique dans le thymus, où sa délétion conduit à des défauts de maturation des cellules T et à une modification de l'épissage alternatif de l'ARNpré-m de CD45 (Wang *et al.*, 2001). La protéine CD45 est un récepteur tyrosine phosphatase connu pour sa régulation par épissage différentiel durant le développement et l'activation des thymocytes.

Trois grands modèles permettent d'expliquer les mécanismes permettant aux protéines SR de réguler l'inclusion des exons :

- Le modèle de « recrutement » montre que la fixation des protéines SR sur les ESE favorise la reconnaissance des sites 5' et 3' d'épissage, participant ainsi à la définition de l'exon (Graveley *et al.*, 2001). Par exemple, la fixation de protéines SR sur les ESE permet l'activation des sites 3' d'épissage faibles en favorisant le recrutement de U2AF65 dans l'intron situé en amont, dans ce cas, U2AF35 est requise pour une activation optimum (Zuo et Maniatis, 1996; Graveley *et al.*, 2001; Guth *et al.*, 2001). Le mécanisme exact du recrutement du facteur U2AF et par conséquent celui de la snRNP U2, par les protéines SR fixées sur l'ESE n'a pas été clairement caractérisé. En effet, il pourrait s'agir d'une interaction directe ou il pourrait y avoir intervention d'autres facteurs comme le co-activateur SRm160/300

(Eldridge *et al.*, 1999). Les protéines SR sont également capables de favoriser la sélection des sites 5' d'épissage. C'est le cas de la protéine ASF/SF2 (Krainer *et al.*, 1990) qui en interagissant avec le facteur U1-70K stimulerait l'association de la snRNP U1 avec le site 5' d'épissage (Eperon *et al.*, 2000).

- Dans le modèle de « l'inhibition », la fixation des protéines SR sur des ESE permet d'empêcher la répression due au recrutement de protéines hnRNP sur des ESS (Zhu *et al.*, 2001). Un antagonisme entre les protéines SR et hnRNP a été montré dans le dernier exemple cité précédemment. En effet, l'utilisation des deux sites 5' d'épissage de l'ARNpré-m modèle, dérivé de l'adénovirus, est favorisée par la liaison de la protéine ASF/SF2. A l'inverse, elle est inhibée par la protéine hnRNP A1 qui va alors empêcher le recrutement de la snRNP U1 (Eperon *et al.*, 2000). Le site 5' d'épissage le plus fort sera alors sélectionné. Ce type de compétition existe également dans le cas de la sélection de l'exon 3 de l'ARNpré-m *tat* de HIV1. La fixation de la protéine hnRNP A1 sur un ESS entraîne une liaison additionnelle et une polymérisation de hnRNP A1 vers le site 3' d'épissage, ce qui provoque l'exclusion de l'exon 3. La liaison de la protéine ASF/SF2 sur un ESE, situé entre l'ESS et le site 3' d'épissage, empêche la propagation de liaison du facteur hnRNP A1 permettant ainsi la sélection de l'exon 3 (Zhu *et al.*, 2001; Marchand *et al.*, 2002).

- Le modèle du « réseau d'interactions » explique que les protéines SR sont capables de former des complexes *via* des liaisons protéine-protéine à travers l'intron, favorisant ainsi le rapprochement des sites 5' et 3' d'épissage (Wu et Maniatis, 1993). En effet l'interaction entre les facteurs ASF/SF2 et SC35 et le facteur U170K au niveau du site 5' d'épissage et U2AF35 au niveau du site 3' d'épissage est dépendante des domaines RS de ces deux protéines SR (Wu et Maniatis, 1993). De plus, il a été montré que des protéines SR favorisent l'assemblage du spliceosome en recrutant la tri-snRNP U4/U6.U5 *via* des interactions avec les domaines RS des protéines SR like SRp55 et SRp110 (Makarova *et al.*, 2001).

Ces trois modèles ne sont pas exclusifs et pourraient agir de concert pour que les protéines SR favorisent l'inclusion des exons qu'elles régulent.

Autres fonctions des protéines SR

Les protéines SR ne sont pas seulement impliquées dans la régulation de l'épissage alternatif, elles interviennent également dans la transcription, la traduction et la stabilité des ARNm.

Le rôle de ces protéines dans l'élongation de la transcription a été montré par des expériences réalisées sur des cellules en cultures qui indiquent que la protéine SC35 permet

de favoriser l'élongation en recrutant le facteur P-TEFb (Lin *et al.*, 2008). D'autres expériences effectuées *in vitro* et *in vivo* ont permis de montrer que certaines protéines SR seraient également impliquées dans le couplage entre la transcription et l'épissage en interagissant avec le domaine carboxy terminal de l'ARNpol II (Das *et al.*, 2007; Sapra *et al.*, 2009).

Les protéines SR SRp20, 9G8 et ASF/SF2 sont également impliquées dans l'export nucléaire des ARNm en interagissant avec le récepteur d'export nucléaire TAP/NFX1 (Huang et Steitz, 2001; Huang *et al.*, 2003).

La surexpression des protéines ASF/SF2, SC35, SRp40 et SRp55 active la voie du NMD de manière dépendante de leur domaine RS (Zhang et Krainer, 2004). Cet effet des protéines SR sur les mécanismes de surveillance des ARNm est nucléaire et se produirait avant l'export des ARNm (Sato *et al.*, 2008).

Les protéines SR sont capables d'activer la traduction (Sanford *et al.*, 2004). En effet, la protéine ASF/SF2 favorise la phosphorylation du facteur 4E-BP, inhibiteur compétitif de la traduction dépendante de la coiffe. Cette phosphorylation va libérer le facteur eiF4E, le rendant ainsi accessible à la coiffe et permettant l'initiation de la traduction (Michlewski *et al.*, 2008).

Les protéines SR sont elles-mêmes régulées

La localisation et/ou l'activité de certaines protéines SR est modulée selon le type cellulaire, ou en réponse à des signaux extracellulaires spécifiques, et l'expression individuelle des protéines SR varie selon le contexte spatio-temporel. L'expression des protéines SRp20, ASF/SF2, SC35, 9G8, SRp30C, SRp40 et SRp55 peut être modulée au niveau post-transcriptionnel par épissage alternatif ou polyadénylation alternative (Bourgeois *et al.*, 2004). Du fait de leur fonction les protéines SR sont capables d'autoréguler la maturation de leur propre ARNpré-m. Une forte expression du facteur SC35 et pas des protéines ASF/SF2 et 9G8 active deux évènements d'épissage alternatif distincts dans la 3' UTR de son ARNpré-m (Sureau *et al.*, 2001). Les ARNm ainsi produits sont ciblés par le mécanisme de NMD et rapidement dégradés, ce qui conduit à une baisse globale du niveau de l'ARNm SC35. Le facteur SRp20 régulerait sa propre expression en se fixant sur l'exon 4 de son ARNm. L'inclusion de cet exon aboutit à la production d'une protéine non fonctionnelle sans domaine RS (Jumaa *et al.*, 1997).

Un autre mode de régulation des protéines SR est le contrôle de l'état de phosphorylation de leur domaine RS qui permet de moduler leur activité ou leur localisation

hnRNP	Domaines	Séquences
A1	2xRRM, RGG	UAGRG(A/U) UAGG GGGU UGGGGU
A2/B1	2xRRM, RGG	(UUAGGG)n UAGRG(A/U) UAGAGA UAGG GGGU UGGGGU
C1/C2	1xRRM	UUUUUU
D	2xRRM, RGG	AU riche AUUUA UUAG
E1/E2	3xKH	C riche
H/H'/F	3xRRM	GGGA GGGG
I (PTB)	4xRRM	UCUUC CUCUCU CUUCUCUCU
K	3xKH, RGG	C riche $UC_{3/4}$(U/A)(A/U)
L	4xRRM	CA répété ACAC CACA UACA UCAU CACC
M	3xRRM	G ou U riche RRGGAGGRR $(UA)_6$

Tableau 2- Domaines et séquences de liaison des protéines hnRNP. D'après (Singh et Valcarcel, 2005) et (Venables *et al.*, 2008). R :purines A ou G. Quand une séquence est répétée plusieurs fois, elle est suivie de n. RRM : *RNA Recognition Motif*, RGG : Arginine-Glycine-Glycine, KH : *hnRNP K Homology domain*.

(Graveley, 2000; Stamm, 2008). Un niveau précis de phosphorylation des protéines SR est requis pour qu'elles interviennent dans l'épissage car l'hypo- ou l'hyper-phosphorylation des protéines SR inhibe l'épissage (Prasad *et al.*, 1999). La phosphorylation du domaine RS de ASF/SF2 favorise les interactions protéine-protéine avec les domaines RS d'autres facteurs d'épissage (Xiao et Manley, 1997). Plusieurs kinases permettant de phosphoryler le domaine RS des protéines SR ont été caractérisées : la famille SRPK (*SR Protein Kinase*), la famille Clk/Sty et la topoisomérase I (Long et Caceres, 2009). SRPK1 est kinase spécifique des sérines et permet la phosphorylation uniquement de la région amino terminale du domaine RS de la protéine ASF/SF2 (Ngo *et al.*, 2008). La phosphorylation par Clk/Sty est beaucoup moins fine et concerne l'intégralité du domaine RS, aboutissant ainsi à un état hyperphosphorylé de ASF/SF2 (Ngo *et al.*, 2005). D'autres protéines kinases et phosphatases, répondant aux voies de transduction du signal sont également impliquées dans la régulation de l'état de phosphorylation des protéines SR, ou SR *like*. Par exemple, la protéine kinase Akt, impliquée dans la voie de transduction du signal dépendante de PI3K (phosphatidylinositol kinase 3), peut phosphoryler SRp40, ASF/SF2 et 9G8, leur permettant ainsi de réguler l'épissage de la protéine kinase C β II et de la fibronectine (Blaustein *et al.*, 2004; Patel *et al.*, 2005). La déphosphorylation de la protéine SR *like* SRp38 par la protéine phosphatase PP1 suite à un choc thermique va conduire à une répression de l'épissage par SRp38 (Shin *et al.*, 2004; Shi et Manley, 2007; Feng *et al.*, 2008).

2.3.2.2. Les protéines hnRNP

Les protéines hnRNP (*heterogeneous nuclear RiboNucleoProtein*) constituent une famille d'une vingtaine de membres et sont très conservées au cours de l'évolution. Contrairement aux protéines SR, les protéines hnRNP sont localisées de manière diffuse dans le noyau et certaines d'entre elles sont également présentes dans le cytoplasme. Certaines de ces protéines existent sous plusieurs isoformes produites par épissage alternatif et peuvent subir des modifications post-traductionnelles. Les protéines hnRNP possèdent en majorité des domaines de liaison aux ARN de type RRM (tableau 2). Les hnRNP E et K présentent des domaines de liaison aux ARN de type KH. Certaines d'entre elles possèdent un second domaine de liaison aux ARN de type Arginine-Glycine-Glycine (RGG). Les protéines hnRNP s'assemblent sur les transcrits naissants issus de l'ARNpol II, et ont été décrites initialement comme des protéines d'empaquetage de l'ARN en s'y fixant de manière non spécifique sur des sites de faible affinité. Ces protéines peuvent également se lier de manière spécifique sur

l'ARN au niveau de sites de haute affinité spécifiques à chacune d'entre elles (tableau 2). Ces interactions spécifiques participent aux diverses fonctions des hnRNP comme le transport, la localisation cytoplasmique, la stabilité, la traduction des ARNm. Ces protéines ont également un rôle majeur dans la régulation de l'épissage alternatif et la polyadénylation des ARNpré-m.

Les protéines hnRNP, décrites comme inhibitrices de l'épissage alternatif sont pour certaines, également capables d'activer l'épissage alternatif. C'est le cas de la protéine hnRNP H qui active l'épissage d'un ARNpré-m *tat* de HIV-1 (Jacquenet *et al.*, 2001; Schaub *et al.*, 2007), des protéines hnRNP F et H qui appartiennent à un complexe activateur intronique qui favorise la reconnaissance de l'exon N1 du gène c-src de souris (Markovtsov *et al.*, 2000), et des protéines hnRNP K, L et M qui seront présentées dans le second chapitre des résultats. Les deux protéines les mieux caractérisées pour leur implication dans la régulation de la maturation alternative des ARNpré-m sont la hnRNP A1 et la hnRNP I ou PTB (*Polypyrimidine Tract Binding protein*).

La protéine hnRNP A1

La hnRNP A1 possède deux domaines de liaison à l'ARN de type RRM en position amino terminale et un domaine de type RGG en position carboxy terminale (Dreyfuss *et al.*, 1993; Mayeda *et al.*, 1994). Celle-ci agit principalement comme répresseur de l'épissage alternatif en interagissant avec des séquences UAGRG(A/U) (Burd et Dreyfuss, 1994).

La protéine hnRNP A1 intervient dans la répression de son propre ARNpré-m (Blanchette et Chabot, 1999). La fixation de cette protéine sur deux séquences introniques inhibitrices, CE1a et CE4, situées de part et d'autre de l'exon 7B va permettre la formation d'une boucle d'exclusion qui empêcherait la reconnaissance de l'exon 7B (Blanchette et Chabot, 1999). La formation de cette structure particulière est possible du fait de la dimérisation de hnRNP A1 *via* son domaine carboxy terminal riche en glycine. La hnRNP A1 est également capable de réprimer l'utilisation de l'exon 7 de l'ARNpré-m de SMN2 (*Survival of Motor Neuron* 2) en se fixant sur un ESS situé en 5' de cet exon (Kashima 2007). Ce même type de régulation est observé pour la répression de l'épissage de l'exon K-SAM de l'ARNpré-m du récepteur 2 du FGF humain (FGFR2) (Del Gatto-Konczak *et al.*, 1999). Dans ce cas, la fixation de la hnRNP A1 sur l'ESS est assurée uniquement par le domaine de type RGG. La répression de l'épissage par la hnRNP A1 peut être due à une compétition de fixation avec les protéines SR ASF/SF2 et SC35 comme dans l'ARNpré-m de la

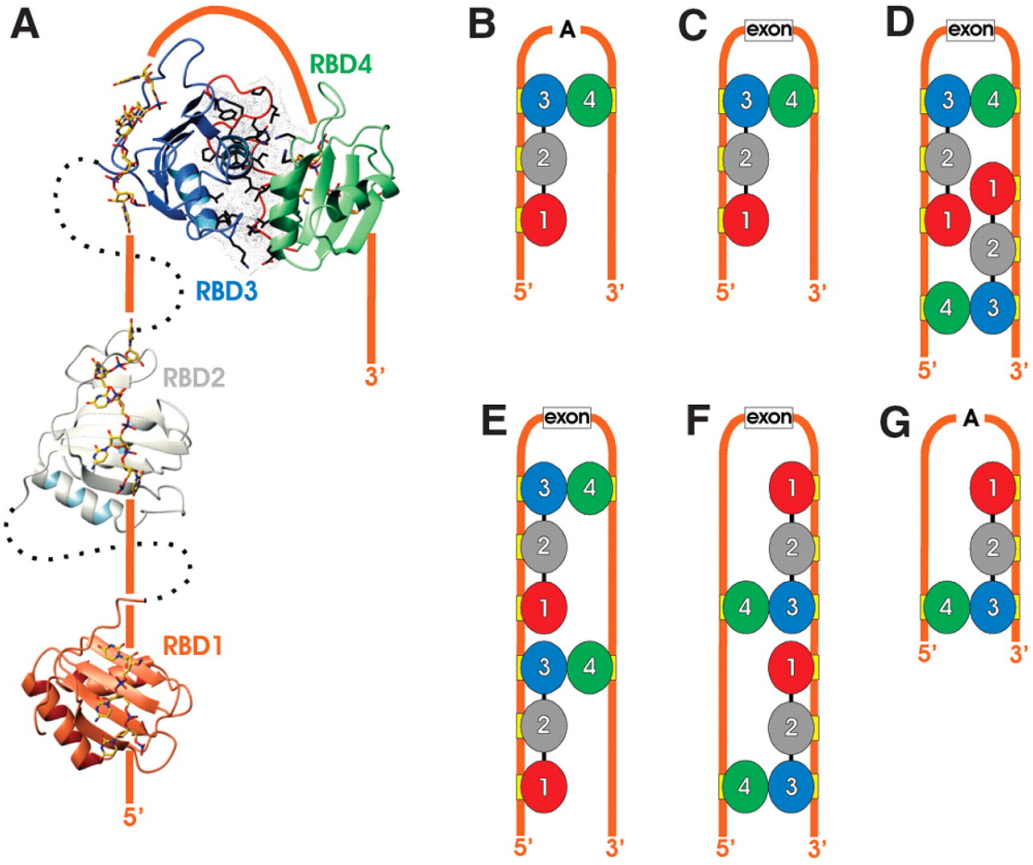


Figure 13- Modèles structuraux de la PTB liée à un ARN. D'après (Oberstrass *et al.*, 2005). A. Modèle de la structure de la PTB avec un ARN contenant quatre régions riches en pyrimidines. Les lignes pointillées représentent les domaines charnières situés entre les RBD (domaines de liaison à l'ARN ou RRM) et la ligne orange représente l'ARN. **B.** Modèle basé sur l'exemple de la répression de l'exon 9 de l'ARNpré-m GABA-γ2. L'adénosine du point de branchement (A) est exclue par la formation d'une boucle. **C.** Modèle permettant d'expliquer comment une seule molécule de PTB peut réprimer un exon alternatif court, comme dans le cas de l'exon 3' terminal de l'ARNpré-m de CT/CGRP. **D à F** Modèles impliquant au moins deux molécules de PTB et basés sur les données obtenues pour l'exon 3 de l'ARNpré-m de l'α-tropomyosine, l'exon N1 de l'ARNpré-m de c-src et l'exon IIIB de l'ARNpré-m de FGFR2. **G.** Modèle hypothétique illustrant comment la PTB peut former une boucle d'exclusion contenant le point de branchement (A), en interagissant avec le site 3' d'épissage et une courte région riche en pyrimidines située en amont du point de branchement. Les cercles de couleur correspondent aux différents RRM.

β-tropomyosine aviaire (Expert-Bezancon *et al.*, 2004). L'exon 6B sera donc inclus ou exclus en fonction du ratio des protéines SR et hnRNP A1. Ce dernier exemple montre l'importance des quantités relatives des protéines activatrices ou inhibitrices dans la régulation de l'épissage alternatif. Les variations de ratios des protéines ASF/SF2 et hnRNP A1 pourraient influencer de nombreux évènements d'épissage alternatif (Caceres *et al.*, 1994).

La protéine hnRNP I ou PTB

La protéine PTB, d'une taille de 57 kDa, a été initialement caractérisée par des expériences de transfert de marquage utilisant des extraits nucléaires de cellules HeLa, par sa capacité à se fixer sur la région riche en pyrimidines située en amont du site 3' d'épissage d'un ARNpré-m (Garcia-Blanco *et al.*, 1989). La séquence reconnue par la PTB est constituée par le motif UCUU présent dans un contexte riche en pyrimidines (Perez *et al.*, 1997). La PTB est un facteur régulateur qui contrôle l'épissage alternatif et la polyadénylation des ARNpré-m. Cette protéine est également impliquée dans la localisation cellulaire des ARNm Vg1 et α-actinine (Cote *et al.*, 1999 ; Ma, Liu, Sun, Xie 07), dans la stabilité des ARNm CD154, insuline et VEGF (Hamilton *et al.*, 2003; Coles *et al.*, 2004; Knoch *et al.*, 2004), ou encore dans la traduction des ARN viraux dépendante des IRES (Niepmann *et al.*, 1997; Gosert *et al.*, 2000; Balvay *et al.*, 2007).

La protéine PTB est généralement considérée comme une protéine abondante dont l'expression est ubiquitaire. Cependant chez la drosophile, l'expression de la PTB est restreinte, chez l'adulte, à la lignée germinale mâle (Robida et Singh, 2003).

a- Structure de la protéine PTB

La PTB est une protéine conservée au cours de l'évolution (Wagner et Garcia-Blanco, 2001). Ainsi il existe 85% d'identité entre la protéine de xénope et la protéine humaine. La PTB possède quatre domaines de liaison à l'ARN de type RRM. Les RRM 2 et 3, séparés par une région centrale riche en alanines, diffèrent de la séquence consensuelle des RRM du fait de l'absence d'acides aminés aromatiques. Toutefois, la structure de ces domaines en solution est similaire à celle des RRM classiques (Perez *et al.*, 1997; Conte *et al.*, 2000). Des études utilisant des expériences d'ultracentrifugation analytique et de diffraction aux rayons X ont montré que la PTB en solution n'existe que sous une forme monomérique (Simpson *et al.*, 2004; Amir-Ahmady *et al.*, 2005).

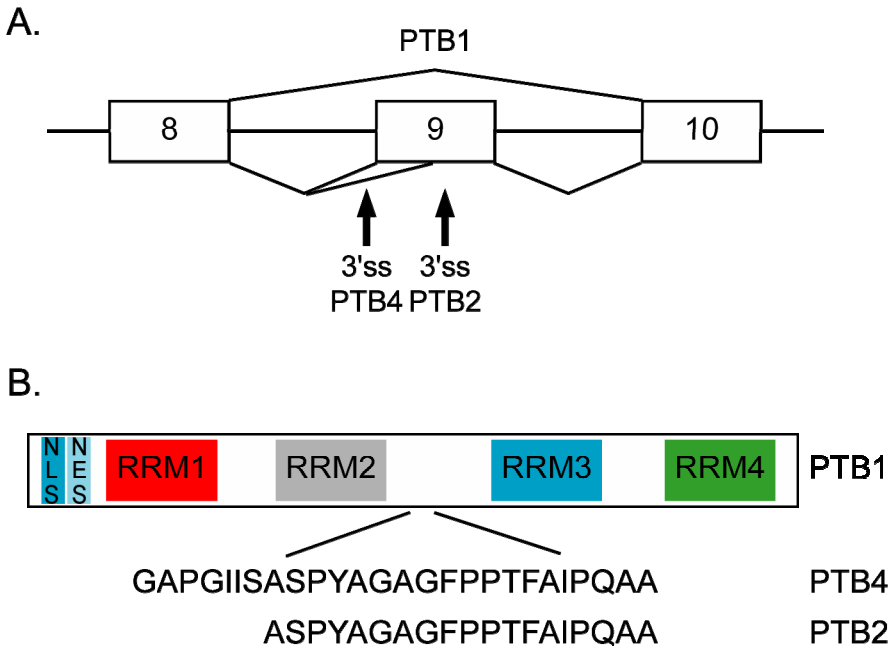


Figure 14- Les différentes isoformes de la protéine PTB. D'après (Wollerton *et al.*, 2001). **A.** Représentation des évènements d'épissage permettant la production des trois isoformes de la PTB suite à la sélection différentielle de sites 3'd'épissage (3'ss). La PTB1 ne contient pas l'exon 9. **B.** Isoformes protéiques de la PTB. Les acides aminés codés par l'exon 9 sont indiqués. NLS : signal de localisation nucléaire ; NES : signal d'export nucléaire ; RRM : domaine de liaison à l'ARN.

Les quatre RRM sont nécessaires pour une fixation efficace de la PTB sur l'ARN (Perez *et al.*, 1997; Yuan *et al.*, 2002; Simpson *et al.*, 2004). Une étude structurale de la PTB en solution utilisant des fragments de la protéine contenant les RRM1 et 2 ou les RRM3 et 4, en interaction avec un ARN constitué de répétitions du dinucléotide CU, montre que chaque domaine est capable de se fixer sur l'ARN et que les domaines 3 et 4 sont capables d'interagir entre eux, ce qui conduit à la structuration de l'ARN sous forme de boucle (figure 13) (Oberstrass *et al.*, 2005). Les changements de structure locale de l'ARN suite à la fixation de la PTB participeraient à la répression des exons régulés. En effet, l'utilisation d'une PTB de rat tronquée au niveau du RRM4 permet la fixation de la protéine en amont de l'exon neuronal de l'ARNpré-m de la sous-unité γ2 du récepteur au $GABA_A$ mais à une localisation différente. Cette modification est concomitante à la perte de la répression exercée sur cet exon (Liu *et al.*, 2002). D'autre part un mutant dominant négatif de la PTB tronquée au niveau des RRM3 et 4 n'est plus capable de fixer les séquences régulatrices MSE qui encadrent l'exon 5 de la troponine T cardiaque. Cet effet pourrait être dû à la dimérisation de ce mutant avec la PTB pleine longueur empêchant ainsi l'action de cette dernière (Charlet *et al.*, 2002; Ladd *et al.*, 2005). Malgré ces observations (Auweter et Allain, 2008), il est possible que le rôle de la PTB soit indépendant de son mode liaison à l'ARN. Il a en effet été montré que l'adressage du RRM2 associé à la région charnière le suivant, sur une région située en amont de l'exon 3 de l'α-tropomyosine est suffisante pour entraîner la répression de cet exon dans des cellules en culture (Robinson et Smith, 2005).

b- Expression et isoformes apparentées à la protéine PTB

La PTB est présente à la fois dans le noyau et dans le cytoplasme et fait la navette entre ces deux compartiments (Kamath *et al.*, 2001). Elle possède un signal de localisation nucléaire (NLS) en deux parties à l'extrémité amino terminale ainsi qu'un signal d'export (NES) imbriqué dans le NLS (Perez *et al.*, 1997). Comme pour les protéines SR la localisation de la PTB est en partie due à sont état de phosphorylation. Ainsi, dans les ovocytes de xénope la phosphorylation de la sérine en position 16 conduit à une localisation cytoplasmique, alors qu'une déphosphorylation permet une localisation nucléaire (Xie *et al.*, 2003).

Trois isoformes de la PTB résultant d'un épissage alternatif de la région centrale charnière ont été caractérisées chez les mammifères (figure 14). Les PTB 2 et 4 possèdent ainsi respectivement une insertion de 19 et 26 acides aminés entre les RRM2 et 3 par rapport à la PTB 1. Ces isoformes ont des degrés de répression différents (Wollerton *et al.*, 2001). La

PTB 4 réprimerait plus fortement l'épissage de l'exon 3 de l'α-tropomyosine de rat que les PTB 1 et 2, alors qu'aucune différence n'a été observée sur la régulation de l'exon SM de l'α-actinine (Wollerton *et al.*, 2001). La plupart des travaux s'intéressant à la fonction de la PTB porte sur la PTB 1.

Tout comme certaines protéines SR, la PTB est capable de réguler sa propre expression par épissage alternatif. La présence d'une quantité importante de PTB conduit à l'exclusion de l'exon 11 *via* sa fixation sur une séquence intronique localisée en amont (Wollerton *et al.*, 2004). L'épissage de l'exon 10 à l'exon 12 conduit à un décalage du cadre de lecture et à l'apparition d'un codon stop précoce dans l'exon 12. L'ARNm ainsi produit est dégradé par le mécanisme de NMD, conduisant à une diminution de la quantité de protéine PTB (Wollerton *et al.*, 2004). Ce mécanisme permettrait non seulement d'éviter une surexpression de la PTB, mais aussi d'activer la synthèse de PTB dans le noyau quand celle-ci est mobilisée dans le cytoplasme (Wollerton *et al.*, 2004).

En plus de ces trois isoformes de la PTB il existe trois protéines apparentées. Les protéines nPTB, smPTB et Rod1 présentent 70% à 80% d'identité avec la PTB. Elles présentent une expression spécifique à certains tissus. La protéine nPTB est exprimée préférentiellement dans les cellules neuronales, celle-ci est fortement exprimée dans le cerveau adulte de souris au contraire de la PTB (Lillevali *et al.*, 2001). La nPTB est impliquée dans la régulation de l'épissage alternatif et permet la régulation cellulaire de l'exon N1 de l'ARNpré-m c-src. La protéine smPTB est exprimée essentiellement dans les muscles lisses, elle est présente dans l'aorte, les poumons, l'utérus et dans la peau (Gooding *et al.*, 2003). Des expériences réalisées in vitro montrent qu'elle se fixe avec une meilleure affinité que la PTB sur des séquences régulant l'épissage de l'exon 3 de l'α-tropomyosine de rat, mais son rôle dans la régulation de l'épissage alternatif n'a pas été démontré (Gooding *et al.*, 2003). La protéine Rod1 est exprimée spécifiquement dans les cellules hématopoïétiques (Yamamoto *et al.*, 1999). A ma connaissance aucun rôle de la protéine Rod1 dans la régulation de l'épissage alternatif n'a encore été décrit. La PTB régule l'expression des protéines Rod 1 et nPTB par épissage alternatif en conduisant respectivement à l'exclusion des exons 2 et 10, ciblant ainsi ces ARNm vers le mécanisme de NMD (Spellman *et al.*, 2007).

c- Modes de régulation de l'épissage alternatif par la protéine PTB

La PTB est impliquée dans la répression d'un grand nombre d'exons alternatifs dont l'exon 6 de l'ARNpré-m de Fas (Izquierdo *et al.*, 2005), l'exon N1 de l'ARNpré-m de c-src (Sharma *et al.*, 2005), l'exon 6B de l'ARNpré-m de la β-tropomyosine (Sauliere *et al.*, 2006),

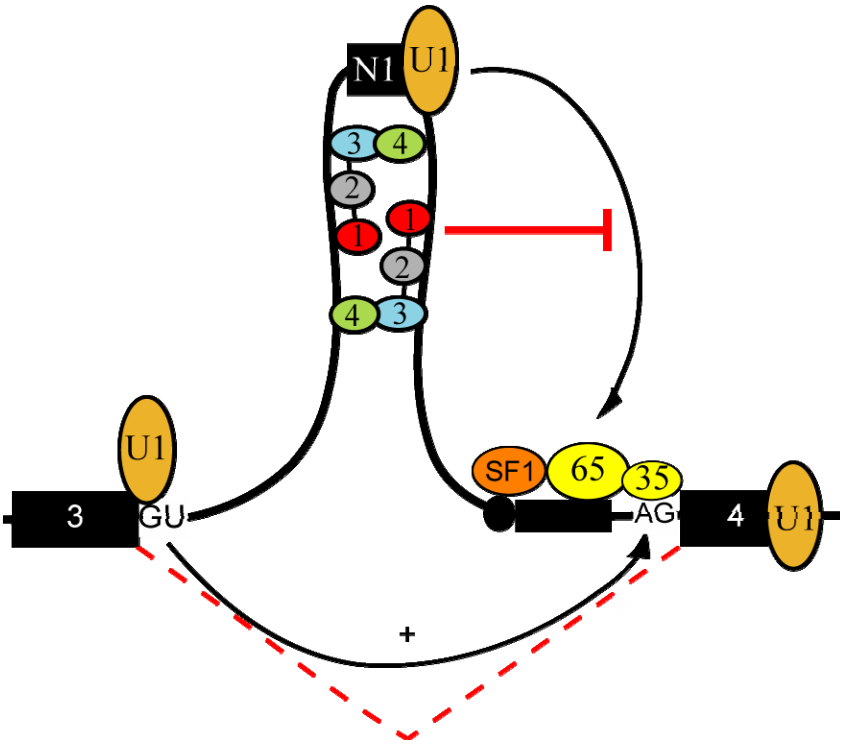


Figure 15- Modèle de la répression de l'exon N1 de l'ARNpré-m de c-src par plusieurs molécules de PTB. D'après (Spellman et Smith, 2006). La liaison de la PTB sur l'ARNpré-m n'inhibe pas la fixation de la snRNP U1 sur le site 5' d'épissage de l'exon N1. La PTB empêche les interactions à travers l'intron entre la snRNP U1 (U1) et le facteur U2AF65 (65). Cependant, la snRNP U1 fixée sur le site 5' d'épissage de l'exon 3 reste accessible et peut interagir à travers l'intron avec U2AF65. Ainsi, l'exon N1 est exclu de l'ARNm. Les motifs de liaison à l'ARN de type RRM de la PTB sont représentés par des cercles numérotés de 1 à 4.

l'exon 5 de l'ARNpré-m de la troponine T cardiaque (Charlet *et al.*, 2002), l'exon SM de l'ARNpré-m de l'α-actinine (Southby *et al.*, 1999), l'exon IIIb de l'ARNpré-m de FGFR2 (Wagner *et al.*, 2005), l'exon 9A9' de l'ARNpré-m de l'α-tropomyosine (Hamon *et al.*, 2004), l'exon IIIb de l'ARNpré-m de la fibronectine (Norton, 1994). Son rôle dans la répression de l'épissage alternatif est claire, et plusieurs mécanismes d'action ont été caractérisés (Sawicka *et al.*, 2008).

Des sites de liaison de la PTB sont souvent présents en plusieurs copies dans les introns encadrant l'exon régulé. Cette disposition particulière des sites de fixation a conduit à proposer un premier modèle qui suggère que la liaison de la PTB sur l'ensemble des sites présents sur l'ARNpré-m rendrait l'exon inaccessible aux facteurs d'épissage et/ou aux facteurs régulateurs (Wagner et Garcia-Blanco, 2001). Dans certains ARNpré-m, tels que celui du récepteur γ2 au $GABA_A$, les sites de liaison de la PTB sont localisés à proximité du point de branchement dans la région riches en pyrimidines (Ashiya et Grabowski, 1997; Zhang *et al.*, 1999). Le second modèle de répression est basé sur une compétition de fixation entre les facteurs PTB et U2AF65 au niveau de la région riche en pyrimidines. Ainsi, la fixation de PTB empêcherait la mise en place du spliceosome. Cet antagonisme de fixation serait impliqué dans la régulation de l'exon EIIIB de l'ARNpré-m de la fibronectine (Norton, 1994), des exons 2 et 3 de l'ARNpré-m de l'α-tropomyosine (Lin et Patton, 1995), ou de l'exon 6B de l'ARNpré-m de la β-tropomyosine (Sauliere *et al.*, 2006). Le troisième modèle repose sur la formation d'une boucle d'exclusion qui a pour effet d'isoler du reste de l'ARNpré-m, l'exon ou les éléments qui favorisent la reconnaissance de cet exon. Un tel mécanisme d'action a été proposé pour la régulation de l'exon N1 de l'ARNpré-m de c-src (Chou *et al.*, 2000).

L'utilisation de cet exon est réprimée dans les cellules non neuronales du fait de la présence de deux séquences situées de part et d'autre de l'exon et qui coopèrent pour une fixation stable de la PTB (Chan et Black, 1997; Chou *et al.*, 2000). Pour cet exemple précis, la fixation de la PTB permet la définition du site 5'd'épissage de l'exon N1, mais empêche l'interaction au travers de l'intron entre la snRNP U1 et le complexe U2AF mis en place au niveau de l'exon aval (figure 15) (Sharma *et al.*, 2005; Sharma *et al.*, 2008). Un complexe d'épissage appelé E', différent du complexe E est formé. Ce complexe E' conduit à la mise en place d'un complexe appelé A' qui ne permet pas l'assemblage d'un complexe d'épissage actif. Il y a donc exclusion de l'exon N1 (Sharma *et al.*, 2008).

L'épissage de l'exon 6 de l'ARNpré-m de Fas est également régulé par la PTB (Izquierdo *et al.*, 2005). La protéine Fas est impliquée dans l'apoptose (Krammer, 2000). L'inclusion de l'exon 6 conduit à une forme membranaire qui favorise l'apoptose. A l'inverse, son exclusion permet la synthèse d'une forme soluble de la protéine qui prévient la mort cellulaire (Cascino *et al.*, 1995). L'interaction de la PTB avec un ESS localisé dans l'exon 6 entraîne une diminution de la fixation du facteur U2AF65 au site 3' d'épissage (Izquierdo *et al.*, 2005). Cette répression est dépendante de la présence d'un site 5' d'épissage de l'exon 6 fonctionnel, et n'est pas le fait d'une compétition directe pour la liaison à l'ARN. La PTB interagirait donc en empêchant l'interaction entre la protéine U2AF65 et la snRNP U2 à travers l'exon (Izquierdo *et al.*, 2005).

En fonction de la localisation de ses sites de fixation, la PTB pourrait donc empêcher la communication à travers l'exon ou l'intron lors de la mise en place du spliceosome. Cette répression serait due au blocage des interactions entre la snRNPU1 et le facteur U2AF.

La PTB contrôle également la réaction de clivage/polyadénylation. Ces régulations seront abordées dans la partie traitant de la polyadénylation alternative.

2.3.2.4. Les autres protéines régulatrices

D'autres protéines que les protéines SR et hnRNP sont impliquées dans la régulation de l'épissage alternatif.

Les protéines de la famille CELF

Les protéines CELF constituent une famille composée de six membres : CUG-BP1, ETR-3 (ou CUG-BP2), et les CELF 3 à 6, qui présentent une structure identique. Elles possèdent deux RRM situés dans la région amino terminale, une région charnière centrale et un troisième RRM dans la région carboxy terminale (Barreau *et al.*, 2006). L'expression des protéines CELF est restreinte à certains tissus. Ainsi, les protéines ETR-3 et CUG-BP1 sont majoritairement exprimées dans le cerveau et le muscle strié (Ladd *et al.*, 2001; Faustino et Cooper, 2005). Les ARNm codant pour les protéines CELF3 et CELF5 sont exprimés uniquement dans le cerveau où l'ARNm codant pour la protéine CELF6 est enrichi (Ladd *et al.*, 2001; Ladd *et al.*, 2004).

Ces protéines sont impliquées dans l'activation de l'épissage de l'exon 5 de l'ARNpré-m de la troponine T cardiaque (cTNT) de poulet. Cet exon est inclus dans le cœur embryonnaire mais exclu chez l'adulte. Cette sélection spécifique de l'exon 5 est contrôlée par des séquences introniques activatrices qui encadrent l'exon. Celles-ci sont reconnues *in vitro* par les protéines CUG-BP1, ETR3, et CELF4. La surexpression de toutes les protéines CELF favorise l'inclusion de l'exon 5 dans un ARNm issu d'un minigène cTNT (Ladd *et al.*, 2001; Charlet *et al.*, 2002; Ladd *et al.*, 2004).

Tout comme les protéines SR, les protéines CELF peuvent avoir des fonctions essentielles dans certains organes (Ladd *et al.*, 2005). Ainsi, des souris transgéniques qui surexpriment un dominant négatif de l'ensemble des protéines CELF, présentent des défauts d'épissage et une cardiomyopathie (Ladd *et al.*, 2005). Les protéines CELF peuvent également inhiber l'épissage alternatif. La surexpression de la protéine CELF4 conduit à la répression des exons SM et NM de l'ARNpré-m de l'α-actinine de rat (Gromak *et al.*, 2003).

La régulation de l'épissage alternatif par les protéines CELF serait en partie basée sur un antagonisme avec la PTB. La protéine ETR-3 entre en compétition avec la PTB pour la fixation sur l'ARNpré-m et permet l'inclusion de l'exon 5 de l'ARNpré-m de cTNT dans les cellules musculaires (Charlet *et al.*, 2002; Faustino et Cooper, 2005). La PTB entre également en compétition avec la protéine CUG-BP1. Cette dernière permet une régulation de l'exon SM de l'ARNpré-m de l'α-actinine de rat. La surexpression des protéines CUG-BP1 et ETR-3 active l'inclusion de l'exon SM dans des cellules où celui-ci est normalement réprimé par la PTB (Gromak *et al.*, 2003).

Les protéines CELF ont également des fonctions cytoplasmiques. La protéine CUG-BP1 joue un rôle dans le contrôle de la traduction et dans la stabilité des ARNm. La fixation de CUG-BP1 sur des séquences riches en UGU (Marquis *et al.*, 2006) situées dans la région 3' UTR des ARNm régule leur stabilité (Barreau *et al.*, 2006) notamment en contrôlant l'état d'adénylation des ces ARNm (Paillard *et al.*, 2003). D'autre part la protéine CUG-BP1 stimule la traduction de l'ARNm codant pour la protéine p21, inhibiteur des kinases cycline-dépendantes (Iakova *et al.*, 2004).

Les protéines MBNL (muscleblind)

Les protéines MBNL1, MBNL2, MBNL3 sont codées par trois gènes distincts chez l'homme. Ces protéines se lient à l'ARN *via* un motif de type doigt de zinc composé de trois résidus cystéine et d'un résidu histidine (CCCH) (Pascual *et al.*, 2006). La protéine MNBL1

se lie sur les séquences YGCU(U/G)Y de l'ARN. Des expériences de surexpression ou d'inhibition réalisées dans des cellules indiquent que MBNL1 est impliquée dans la répression de l'exon 5 de l'ARNpré-m de cTNT et dans l'activation de l'exon 11 de l'ARNpré-m du récepteur à l'insuline (Ho *et al.*, 2004). Dans le cas de l'ARNpré-m de cTNT, il a été montré que MBNL se fixe au niveau des nucléotides mésappariés de l'hélice d'une structure en tige-boucle (Warf et Berglund, 2007), et qu'une seule protéine serait suffisante pour fixer les deux sites de liaison situés dans l'intron en amont de l'exon 5, favorisant ainsi l'apparition de cette boucle dans l'ARNpré-m responsable de l'exclusion de l'exon 5 (Teplova et Patel, 2008). Le rôle de MBNL1 dans la régulation des exons 5 de l'ARNpré-m cTNT et 11 de l'ARNpré-m du récepteur à l'insuline est antagoniste de celui des protéines CELF CUB-BP et ETR3 (Ho *et al.*, 2004), mais ne résulte pas d'une compétition entre ces facteurs pour la liaison à l'ARN. Des expériences effectuées chez la souris ont permis de montrer qu'il existait une régulation de ces protéines au cours du développement cardiaque (Kalsotra *et al.*, 2008). Les protéines CELF sont sous représentées au cours du développement cardiaque à l'inverse de la protéine MBNL1. Le profil d'expression de ces protéines permet une régulation de l'épissage au cours du développement (Kalsotra *et al.*, 2008).

Les protéines NOVA

Les protéines NOVA1 et NOVA sont exprimées spécifiquement dans les neurones du système nerveux central (Yang *et al.*, 1998), et interagissent *via* leur domaines KH avec l'ARN au niveau de séquences YCUAY (Jensen *et al.*, 2000; Dredge et Darnell, 2003; Dredge *et al.*, 2005). NOVA1 régule l'épissage de son propre ARNpré-m en agissant comme un inhibiteur de l'épissage de l'exon 4 *via* son interaction sur un ESS (Dredge *et al.*, 2005). Une analyse par puce à ADN des ARN issus de cerveau de souris invalidées pour le gène Nova1 ou Nova2 a révélé que ces protéines régulent 7% des évènements d'épissage spécifiques du cerveau (Ule *et al.*, 2005). La régulation des exons alternatifs dans le cerveau par NOVA est due à la présence des éléments de séquence YCAY dans l'ARNpré-m que se soit chez l'homme, le poulet, la souris ou le poisson zèbre (Jelen *et al.*, 2007).

Les protéines RAVER

La protéine RAVER1 est une protéine apparentée aux protéines hnRNP du fait de l'homologie de séquence au niveau de ses trois domaines de liaison à l'ARN de type RRM en position amino terminale. Cependant, aucun motif de liaison de cette protéine à l'ARN n'a encore été caractérisé. La protéine RAVER1 est exprimée de manière ubiquitaire et est

localisée dans le noyau et le cytoplasme (Huttelmaier *et al.*, 2001). A l'inverse, une protéine apparentée à RAVER1, nommée RAVER2, a un profil d'expression dépendant du contexte spatio-temporel (Kleinhenz *et al.*, 2005). Celle-ci interagit avec la protéine PTB et intervient dans la répression des certains évènements d'épissage alternatifs (Huttelmaier *et al.*, 2001). En effet, la surexpression de RAVER1 dans des cellules en culture conduit à l'exclusion de l'exon 3 de l'ARNpré-m de l'α-tropomyosine (Gromak *et al.*, 2003). De même, l'adressage de RAVER1 au niveau de l'exon 6 de l'ARNpré-m de Fas induit son exclusion (Izquierdo *et al.*, 2005). La protéine RAVER1 interagit avec le RRM2 de la PTB, cette interaction permettrait de potentialiser l'effet répresseur exercé par ces deux protéines, faisant de RAVER1 un cofacteur de la PTB (Rideau *et al.*, 2006). Tout comme la protéine RAVER1, la protéine RAVER2 peut interagir avec la PTB (Kleinhenz *et al.*, 2005).

La famille des protéines TIA

Les protéines TIA1 et TIAR possèdent trois domaines de liaison à l'ARN de type RRM en position amino terminale (Dember *et al.*, 1996), et sont exprimées dans de nombreux tissus. Généralement la fixation de la protéine TIA1 sur des séquences introniques riches en uridines proches du site 5' d'épissage favorise sa reconnaissance et donc l'utilisation de l'exon en amont. Ainsi, la protéine TIA1 active l'utilisation de l'exon 6 de l'ARNpré-m de Fas en favorisant le recrutement de la snRNP U1 au niveau du site 5' d'épissage de cet exon (Forch *et al.*, 2002; Izquierdo *et al.*, 2005). Des expériences effectuées sur des lignées cellulaires ont permis de montrer que le facteur TIA1 favorise le recrutement de la snRNP U1 au niveau du site 5' d'épissage de l'exon K-SAM conduisant ainsi à son activation (Gesnel *et al.*, 2007). Un antagonisme fonctionnel entre la PTB et les protéines de la famille TIA a été décrit pour la régulation de l'ARNpré-m de MYPT1 (*myosin phosphatase targeting subunit-1*), ou encore de l'exon 6 de l'ARNpré-m de Fas où la fixation de la PTB sur un ESS conduit à la répression de cet exon (Izquierdo *et al.*, 2005). Cet exon de l'ARNpré-m de Fas peut également être réprimé par la fixation du facteur HuR (*HuAntigen R*) sur un ESS présent dans l'exon 6 (Izquierdo, 2008). Dans ce cas, HuR empêche la définition du site 3' d'épissage de l'exon 6, en bloquant la fixation du facteurs U2AF65, sans inhiber la définition du site 3' d'épissage favorisée par TIA1 (Izquierdo, 2008).

La famille des protéines FOX

Les protéines FOX constituent une famille complexe issue de la transcription de plusieurs gènes qui sont soumis à des régulations par épissage alternatif (Underwood *et al.*,

2005). Les protéines FOX1 et 2 sont essentiellement exprimées dans le cerveau, le cœur et le muscle (Nakahata et Kawamoto, 2005). Elles reconnaissent des motifs ARN de type UGCAUG (Jin *et al.*, 2003; Auweter *et al.*, 2006). FOX1 est essentiellement un activateur de l'épissage alternatif. Sa surexpression conduit à l'activation des exons 4' de l'ARNpré-m de EWS, N1 de l'ARNpré-m de c-src, EIIIB de l'ARNpré-m de la fibronectine ou encore SM de l'ARNpré-m de la fibronectine (Jin *et al.*, 2003; Underwood *et al.*, 2005). FOX1 a un rôle antagoniste à la PTB dans la régulation de l'épissage de l'ARNpré-m de l'α-actinine (Jin *et al.*, 2003). Les protéines FOX1 et 2 activent l'épissage de l'exon E16 de l'ARNpré-m de 4.1R *via* leur interaction avec un ISE (Ponthier *et al.*, 2006). La protéine FOX1 peut également réprimer l'épissage de l'exon NM de l'ARNpré-m de l'α-actinine (Jin *et al.*, 2003).

2.3.3. Les différents modèles de régulation de l'épissage alternatif

L'épissage alternatif est régulé par la formation sur l'ARNpré-m de complexes ribonucléoprotéiques qui vont être à l'origine de la sélection ou de l'exclusion de différents exons. Plusieurs modèles ont été proposés pour expliquer la spécificité cellulaire de l'épissage alternatif.

Le premier suggère que la sélection différentielle des sites d'épissage serait sous le contrôle de facteurs dont l'expression est elle-même restreinte à certaines cellules. Ces facteurs régulateurs ont été identifiés chez la drosophile.

Chez les vertébrés peu de facteurs tissu-spécifiques permettant d'expliquer les régulations de l'épissage alternatif ont été caractérisés. De plus, les facteurs impliqués dans ces régulations présentent des sites de liaison sur l'ARN relativement courts et peuvent donc être rencontrés fréquemment dans le génome. Le second modèle propose que les régulations de l'épissage alternatif sont le fruit d'une combinaison de facteurs exprimés de manière ubiquitaire, et que la spécificité viendrait de la composition de ces complexes qui varierait d'un environnement cellulaire à un autre en fonction de la concentration nucléaire des protéines régulatrices (Singh et Valcarcel, 2005).

Un troisième modèle est basé sur la régulation de l'activité des protéines régulatrices. En effet, la localisation et l'activité des facteurs régulateurs peuvent être contrôlées par des stimuli intra- et extra- cellulaires qui vont conduire à la régulation de l'épissage de certains ARNpré-m (Shin et Manley, 2004).

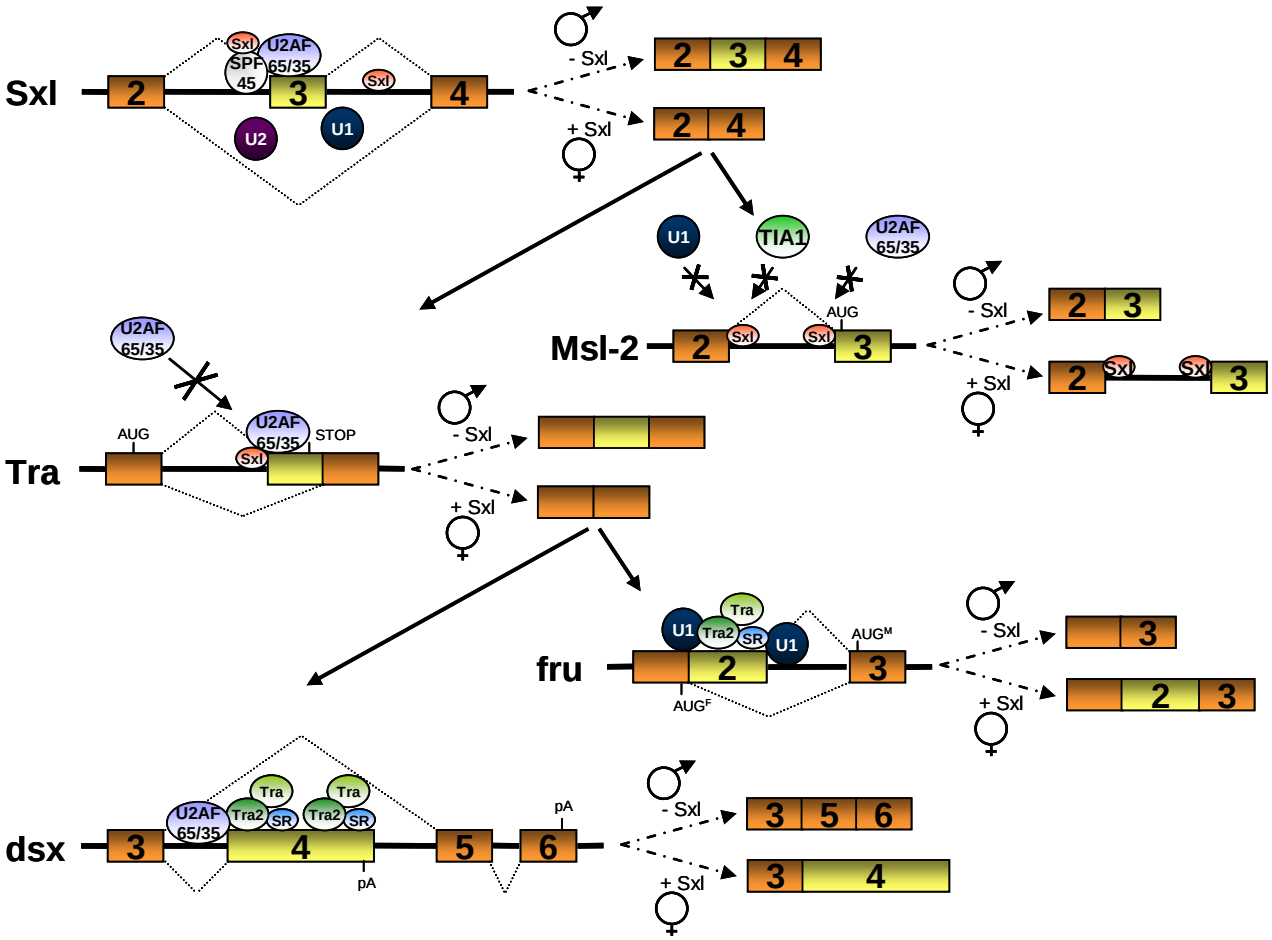


Figure 16- Cascade de régulations de l'épissage alternatif lors de la détermination du sexe chez la drosophile. D'après (Black, 2003) et (Forch et Valcarcel, 2003). La protéine SXL favorise le saut de l'exon 3 de son propre ARNpré-m supprimant le codon stop présent dans cet exon. Cette régulation favorise la production de la protéine SXL. SXL entraîne la rétention d'un intron dans la 5' UTR de l'ARNm de Msl-2, inhibant ainsi la traduction de cet ARNm chez la femelle. SXL réprime l'utilisation d'un site 3' d'épissage alternatif dans l'ARNpré-m de Tra. L'ARNm ainsi produit ne contient pas de codon stop précoce et permet la synthèse de la protéine Tra chez la femelle. Tra agit en combinaison avec Tra-2 qui est exprimée dans les deux sexes. Ces deux protéines favorisent l'utilisation d'un site 5' d'épissage alternatif dans l'exon 2 de l'ARNpré-m de Fru. Elles permettent également la sélection d'un site 3' d'épissage et d'un signal de polyadénylation spécifiques de la femelle dans l'ARNpré-m de dsx.

Ces trois grands modèles de régulations vont être présentés. La régulation de l'épissage alternatif fait également intervenir d'autres évènements cellulaires tels que la transcription. Les mécanismes de couplage entre la transcription et l'épissage alternatif seront traités dans la dernière partie de l'introduction.

2.3.3.1. Régulation de l'épissage alternatif par des protéines spécifiques d'un environnement cellulaire : voie du déterminisme sexuel chez la drosophile

Un des modèles permettant d'expliquer la spécificité de l'épissage alternatif repose sur l'existence de facteurs dont l'expression est elle-même spécifique d'un environnement cellulaire donné. Il en existe chez les vertébrés (NOVA1, nPTB, smPTB, ou les protéines de la famille CELF), mais les facteurs les mieux caractérisés sont ceux identifiés chez la drosophile (TRA, SXL et dmPTB). Chez la drosophile, certains évènements d'épissage alternatif sont régulés par l'expression restreinte de ces facteurs. Un exemple maintenant bien caractérisé est celui de la détermination du sexe, qui repose sur une cascade de régulations spécifiques de l'épissage alternatif par l'épissage alternatif (Black, 2003; Forch et Valcarcel, 2003). La protéine SXL est exprimée spécifiquement chez la femelle où elle réprime l'épissage des gènes impliqués dans le déterminisme sexuel. La spécificité de son expression résulte d'un contrôle transcriptionnel, puis est maintenue par une régulation de l'ARNpré-m Sxl par la protéine SXL. Ainsi, chez la femelle, la protéine SXL réprime l'utilisation de l'exon 3 qui introduit un codon stop en phase avec le codon initiateur dans l'exon 2. A l'inverse, chez le mâle, l'inclusion de l'exon 3 conduit à la synthèse d'une protéine tronquée et inactive. SXL régule également les ARNpré-m Msl2 (*Male-specific lethal 2*) et Tra (*Transformer*).

Chez la femelle, SXL interagit avec deux sites présents dans le premier intron de l'ARNpré-m Msl2. L'un est situé dans la région riche en pyrimidines, bloquant ainsi l'interaction de U2AF. L'autre est localisé au niveau du site 5' d'épissage, et inhibe la fixation de TIA1 et de la snRNP U1 (Forch et Valcarcel, 2001; Black, 2003). Cette régulation conduit à la rétention du premier intron dans l'ARNm Msl2. Le cadre de lecture n'est pas modifié mais la région 5' UTR est différente, ce qui permet la fixation de SXL et l'inhibition de la traduction de l'ARNm de Msl2 (Gebauer *et al.*, 1998; Beckmann *et al.*, 2005).

La fixation de SXL sur la région riche en pyrimidines située en amont du site 3' d'épissage proximal de l'ARNpré-m de Tra, va empêcher la fixation de U2AF sur ce site et conduire à l'utilisation du site 3' d'épissage distal. Cette régulation permet l'exclusion d'un

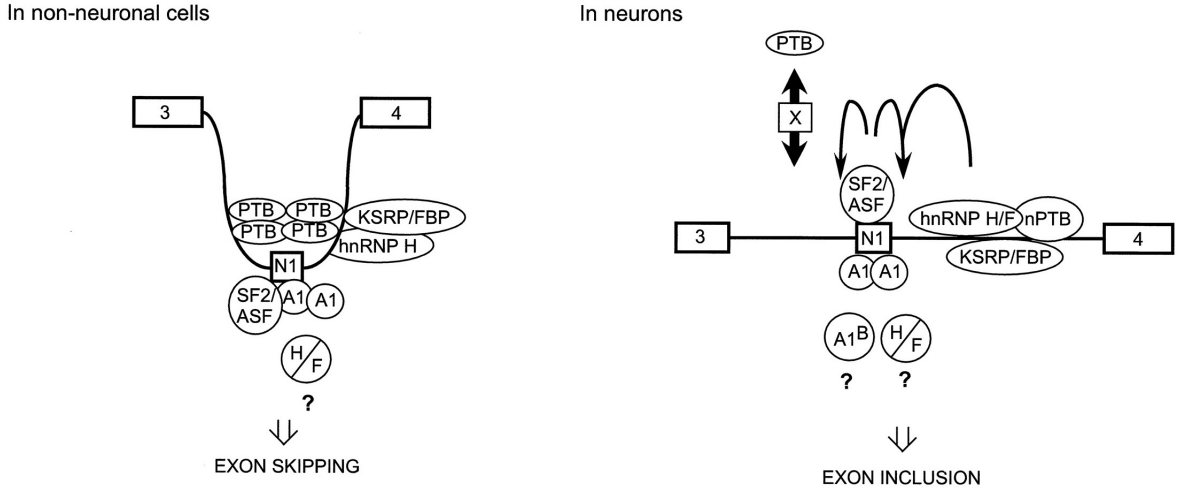


Figure 17- Modèle de contrôle combinatoire de l'exon N1 de l'ARNpré-m de c-src. D'après (Rooke *et al.*, 2003). La protéine PTB se fixe sur les répétitions CU situées de part et d'autre de l'exon N1, réprimant ainsi l'épissage de cet exon dans les cellules non neuronales. Dans les cellules neuronales, la PTB est déplacée et l'élément activateur est alors fonctionnel. La protéine hnRNP A1 (A1) serait également remplacée par la protéine hnRNP $A1^B$ ($A1^B$) ce qui favoriserait l'activation de l'exon N1. Le facteur ASF/SF2 fixé sur l'exon N1 activerait la reconnaissance des deux sites d'épissage en coopérant avec les protéines fixées sur la séquence activatrice.

codon stop de l'ARNm et la synthèse d'une protéine TRA fonctionnelle. Cette protéine associée à la protéine TRA-2, qui est surexprimée dans les deux sexes, régule les ARNpré-m Fru (*Fruitless*) et Dsx (*Doublesex*). L'exon 2 de l'ARNpré-m de Fru possède deux sites 5' d'épissage alternatifs dont la sélection va être régulée par les protéines TRA et TRA-2 (Lam *et al.*, 2003). Ces deux protéines activent, chez la femelle, l'utilisation du site 5' d'épissage aval, alors que chez le mâle, l'utilisation du site amont permet la synthèse d'un facteur de transcription impliqué dans le contrôle de certains aspects comportementaux et du développement (Demir et Dickson, 2005). La fixation des protéines TRA et TRA-2 sur un ESE situé dans l'exon 4 de l'ARNpré-m de Dsx favorise l'utilisation de son site 3' d'épissage faible et conduit à la synthèse d'un facteur de transcription activant la voie de différenciation femelle. Cette cascade de régulations est présentée dans la figure 16.

2.3.3.2. Régulation de l'épissage alternatif par des complexes de protéines : l'exon N1 de l'ARNpré-m de c-src

Deux exemples bien caractérisés peuvent être donnés pour illustrer la régulation de l'épissage alternatif par de multiples facteurs protéiques : les exons IIIb et IIIc de l'ARNpré-m de FGFR2 et l'exon N1 de c-src. La spécificité cellulaire de l'exon N1 serait due à un réarrangement des différents facteurs régulateurs au sein des complexes formés sur l'ARNpré-m de c-src. Ainsi, les complexes sont différents dans les cellules neuronales, dans lesquelles l'exon N1 est utilisé, par rapport aux autres types cellulaires, où l'exon N1 est sauté (figure 17).

L'exon N1 est régulé par des protéines recrutées sur deux séquences introniques inhibitrices situées de part et d'autre de l'exon, et deux séquences activatrices, l'une est localisée dans l'intron en aval et encadre l'élément inhibiteur, l'autre est présente dans l'exon. Dans les cellules non neuronales, les séquences inhibitrices permettent la mise en place d'un complexe intronique constitué des protéines PTB, KSRP/FBP, et hnRNP H. Dans les cellules neuronales, les protéines nPTB et hnRNP F sont présentes dans les complexes recrutés sur la séquence intronique activatrice (Black, 2003; Rooke *et al.*, 2003). La protéine nPTB est majoritairement exprimée dans les cellules neuronales et se lie sur les mêmes séquences que la protéine PTB. Des expériences réalisées in vitro ont montré que l'affinité de la nPTB est plus forte que celle de la PTB pour ces séquences de fixation, et que son activité inhibitrice est moins importante que celle de la PTB (Markovtsov *et al.*, 2000). Ainsi l'association de la nPTB sur les sites d'interaction de la PTB dans les neurones permettrait la déstabilisation de

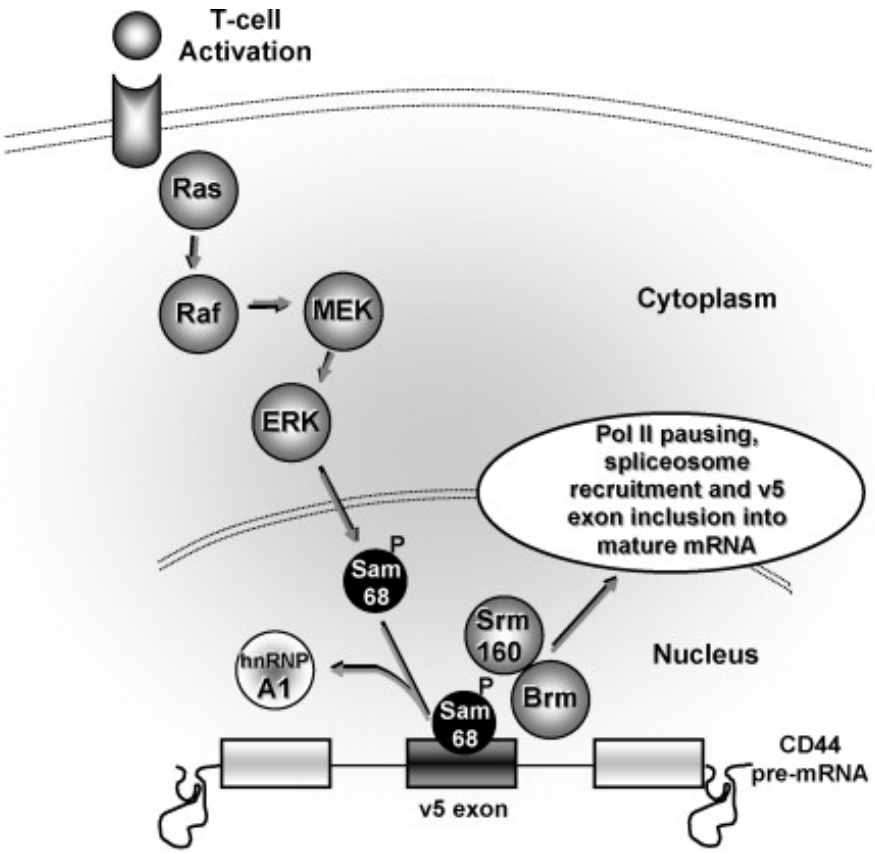


Figure 18- Exemple de régulation de l'épissage alternatif par une voie de transduction du signal cellulaire. D'après (Blaustein *et al.*, 2007). La stimulation des cellules T active la cascade Ras/Raf/MEK/ERK qui va permettre la phosphorylation de la protéine Sam68. Celle-ci interagit avec une séquence située dans l'exon v5 de l'ARNpré-m de CD44 et permet le recrutement des protéines SRm160 puis Brm (*Brahma*). Brm est capable d'interagir avec l'ARNpol II. L'ensemble de ces interactions conduit à l'activation de l'exon v5. La fixation de Sam68 sur l'ARNpré-m empêche l'inhibition médiée par la protéine hnRNP A1.

la PTB et serait à l'origine de l'association du complexe activateur sur la séquence intronique (Markovtsov *et al.*, 2000). Les protéines hnRNP A1 et ASF/SF2 sont recrutées sur l'exon N1, la première conduit à une répression, et la seconde à une activation de l'utilisation de l'exon N1 (Rooke *et al.*, 2003).

La régulation de l'exon N1 est donc basée sur une compétition de fixation entre les facteurs PTB et nPTB selon le contexte tissulaire. Un second type de régulation implique l'action des protéines ASF/SF2 et hnRNP A1 dans la reconnaissance de l'exon N1. De ce fait, les concentrations relatives de ces différents facteurs régulateurs sont à la base de la spécificité de l'épissage.

2.3.3.3. Régulation de l'épissage alternatif par des voies de signalisation cellulaires

La régulation de l'épissage alternatif peut également être dépendante des signaux externes reçus par la cellule (Stamm, 2002; Blaustein *et al.*, 2007). De nombreux stimuli extérieurs, tels que la stimulation d'un récepteur par une hormone ou un facteur de croissance, des voies de réponses au stress, ou encore la dépolarisation dans l'activité neuronale, peuvent entraîner des changements de régulation de l'épissage alternatif. Pour certains systèmes une partie des protéines impliquées est connue, mais dans de nombreux cas, pour lesquels des voies de transduction sont impliquées dans la régulation de l'épissage alternatif, la liste des régulateurs reste incomplète. En effet, ces voies sont complexes et impliquent de nombreuses cascades d'activation ou d'inhibition de protéines intracellulaires. Quatre mécanismes majeurs induisant des modifications de la sélection des sites d'épissage alternatifs en réponse à un stimulus externe sont connus : la synthèse (Goncalves *et al.*, 2008), la phosphorylation (Zhong *et al.*, 2009), le changement de localisation (Buxade *et al.*, 2005) et la dégradation (Katzenberger *et al.*, 2009) de protéines régulant l'épissage.

Un exemple bien caractérisé de régulation d'un gène par une voie de signalisation est l'exon v5 de l'ARNpré-m de CD44 (figure 18). La protéine transmembranaire CD44 est issue d'un seul gène dont l'expression est régulée par épissage alternatif contrôlant l'inclusion d'une dizaine d'exons. Certaines des isoformes ainsi produites sont présentes spécifiquement dans des tumeurs (Blaustein *et al.*, 2007). La stimulation des cellules T active la cascade Ras/Raf/MEK/ERK qui conduit à la phosphorylation du facteur Sam68. Cette protéine participe alors à la régulation de l'épissage de l'exon v5 de l'ARNpré-m de CD44 en se fixant

sur une séquence localisée dans l'exon (Matter *et al.*, 2002). Sam 68 permet l'inclusion de l'exon v5 au contraire de la protéine hnRNP A1 qui en se fixant sur le même élément de séquence exclut cet exon de l'ARNm CD44 (Matter *et al.*, 2000). Une autre protéine activatrice permet de favoriser l'inclusion de l'exon v5 quand la voie Ras est activée. En effet, la déplétion de la protéine SRm160 diminue l'utilisation de l'exon v5 dans des cellules en culture, et la surexpression de Sam68 dans ces cellules n'est pas suffisante pour permettre à nouveau l'inclusion de cet exon (Cheng et Sharp, 2006). Des expériences de co-immunoprécipitation ont montré une interaction entre les protéines SRm160 et Sam68 (Cheng et Sharp, 2006). La régulation de l'épissage alternatif de l'exon v5 fait également intervenir la transcription *via* la protéine Brm (*Brahma*). Celle-ci est associée à un complexe de remodelage de la chromatine et est impliquée dans le couplage entre la transcription et l'épissage (Batsche *et al.*, 2006). L'interaction entre Sam68 et Brm permettrait de modifier l'état de phosphorylation du domaine carboxy terminal de l'ARNpol II, induisant une pause dans la transcription. Ces évènements permettraient de faciliter la sélection de l'exon v5 de l'ARNpré-m de CD44 (Blaustein *et al.*, 2007).

Cet exemple est le mieux caractérisé de tous les mécanismes de régulation de l'épissage alternatif par les voies de transduction de signal. Les mécanismes de couplage entre la transcription et l'épissage seront présentés dans la dernière partie de l'introduction.

2.3.3.4. Régulation non physiologique de l'épissage alternatif

Comme présenté en introduction, un nombre croissant de maladies humaines est associé à des défauts d'épissage alternatif (Tazi *et al.*, 2009). Ce mécanisme est donc apparu, depuis quelques années, comme une cible privilégiée de thérapies médicamenteuses (Bracco et Kearsey, 2003; Hagiwara, 2005). La modification de la sélection des sites d'épissage pourrait également permettre d'expliquer les effets secondaires de certains traitements déjà utilisés (Sumanasekera *et al.*, 2008). Etant donné l'importance de l'épissage alternatif en santé humaine, de nombreux travaux réalisés ces six dernières années ont consisté à identifier des composés de faible masse moléculaire permettant de rétablir une sélection correcte des sites d'épissage, le plus spécifiquement possible. Plusieurs composés ciblant différents mécanismes et des exons spécifiques sont répertoriés dans (Sumanasekera 2008). Par exemple l'exon 7 du gène « *survival of motor neurons 2* » (SMN2), dont l'inclusion est diminuée dans l'amyotrophie spinale (SMA) peut voir son épissage modifié par différents composés

chimiques. Ainsi, le traitement de cultures de cellules lymphoïdes par le butyrate de sodium, favorise l'inclusion de l'exon 7 dans la protéine SMN (Chang *et al.*, 2001). De plus, l'administration par voie orale de ce même composé, à des souris présentant un phénotype SMA, améliore leur taux de survie et diminue les symptômes associés (Chang *et al.*, 2001). Dans des cultures de fibroblastes issus de patients traités par des doses physiologiques d'acide valproïque (VPA) la proportion d'ARNm et de protéine SMN2 contenant l'exon 7 est augmentée (Brichta *et al.*, 2003). Le VPA augmente l'expression des protéines SR, et de la protéine Htraβ-1 qui facilite non seulement la transcription du gène SMN, mais aussi l'épissage de l'ARNpré-m SMN2 (Brichta *et al.*, 2003). Les résultats récents d'une étude menée sur 42 patients atteints de SMA et traités au VPA, montrent que la quantité de protéine SMN tronquée est diminuée alors que la quantité de protéine entière reste inchangée (Swoboda *et al.*, 2009). D'autres substances comme le dérivé indole IDC16 peuvent moduler la fonction des protéines SR. L'IDC16 bloque l'épissage des exons dépendant d'un ESE reconnu par ASF/SF2. Cette protéine SR est nécessaire à la production d'ARNm de HIV-1 (Bakkour *et al.*, 2007). L'IDC16 permet de diminuer la production de particules virales sans modifier l'épissage endogène de gènes du cycle cellulaire et de l'apoptose (Bakkour *et al.*, 2007). L'épissage alternatif peut donc être modulé de manière non physiologique par des traitements pharmacologiques. Celui-ci peut également être modifié suite à des cassures de l'ADN provoquées par des radiations utilisées en radiothérapie, ou comme celles induites par les ultraviolets (Munoz *et al.*, 2009). L'irradiation de cellules en culture par des UV provoque une modification globale de l'épissage alternatif de manière co-transcriptionnelle (Munoz *et al.*, 2009). Cet effet est dû à un changement de l'état de phosphorylation de l'ARNpol II (Marengo et Garcia-Blanco, 2009; Munoz *et al.*, 2009) dont le lien avec l'épissage sera présenté dans la dernière partie de l'introduction.

3. La maturation des exons 3' terminaux

Les exons 3' terminaux sont délimités par un site 3' d'épissage et un signal de clivage/polyadénylation. L'utilisation de ces exons dans l'ARNpré-m résulte à la fois de l'épissage de l'intron en amont, et du processus de clivage/polyadénylation. Leur maturation fait donc intervenir le spliceosome et la machinerie de clivage/polyadénylation. Des liens physiques et fonctionnels permettent l'assemblage des machineries d'épissage et de

clivage/polyadénylation en un même macrocomplexe et permettraient d'augmenter l'efficacité des deux réactions et donc de favoriser la définition des exons terminaux (Bauren *et al.*, 1998; Cooke et Alwine, 2002). Les régulations contrôlant l'inclusion des exons 3' terminaux apparaissent donc complexes puisqu'elles peuvent cibler deux mécanismes qui mettent en jeu trois réactions catalytiques.

Cette partie de l'introduction détaillera les mécanismes impliqués dans les interactions entre les facteurs des machineries d'épissage et de clivage/polyadénylation au cours de la sélection des exons 3' terminaux. Les régulations des exons 3' terminaux alternatifs seront ensuite décrites.

3.1. Définition des exons 3' terminaux

Les exons 3' terminaux répondent au modèle de la définition de l'exon. En effet, l'utilisation d'ARN chimériques dérivés de virus a permis de montrer *in vitro* que la présence d'un signal de polyadénylation active l'épissage de l'intron localisé en amont, et que la présence d'un site 3' d'épissage en amont d'un signal (Niwa et Berget, 1991; Cooke *et al.*, 1999). D'autres expériences effectuées à l'aide de cellules en culture ont montré que l'efficacité de la maturation de l'extrémité 3' d'un ARNpré-m est corrélée à celle de l'épissage du dernier intron (Nesic et Maquat, 1994). Ces résultats montrent le couplage entre les réactions d'épissage et de clivage/polyadénylation. Le couplage entre les deux mécanismes semble d'autant plus important que le site 3' d'épissage et le signal de polyadénylation sont faibles, puisqu'un signal de polyadénylation fort va être reconnu même si le site 3' d'épissage faible n'a pas permis un épissage efficace de l'intron situé en amont.

Les bases moléculaires de ces interactions entre les complexes d'épissage et de clivage/polyadénylation ont été déterminées par des expériences réalisées *in vitro*. Ainsi plusieurs expériences indiquent que le facteur d'épissage U2AF65 joue un rôle majeur dans ce couplage. En effet, l'ajout dans des extraits nucléaires d'un peptide, nommé PRD (*PAP Regulatory Domain*), correspondant à la région de U1A qui interagit avec la partie carboxy terminale de la PAP, dissocie les réactions d'épissage et de polyadénylation (Gunderson *et al.*, 1997). De plus, l'adressage de la région carboxy terminale de la PAP sur un exon conduit *in vitro* à la stimulation de l'épissage de l'intron situé en amont (Vagner *et al.*, 2000). Cette

stimulation serait due à une interaction entre le domaine RS situé dans la région amino terminale de U2AF65 et le domaine carboxy terminal de la PAP (Vagner *et al.*, 2000). D'autres expériences ont permis de montrer que des mutations de la région riche en pyrimidines du site d'épissage qui diminuent la fixation de U2AF65 réduisent l'efficacité du clivage, à l'inverse l'adressage de U2AF65 en amont d'un signal de polyadénylation favorise le clivage d'ARNpré-m modèles dans des cellules en culture et *in vitro* (Millevoi *et al.*, 2002). Un lien physique entre U2AF65 et la sous-unité de 59 kDa du facteur de clivage CF Im a également été montré *in vitro*. Cette interaction est due aux domaines RS de U2AF65 et de CF Im 59 (Millevoi *et al.*, 2006). L'adressage de U2AF65 en amont d'un signal de clivage/polyadénylation facilite le recrutement de CF Im conduisant ainsi au clivage et à la polyadénylation des ARNpré-m modèles (Millevoi *et al.*, 2006). Il a été proposé que l'interaction entre U2AF65 et la PAP ait lieu quand le signal de clivage/polyadénylation est fort aidant ainsi à la reconnaissance d'un site 3' d'épissage faible. L'interaction entre U2AF65 et CF Im permettrait quand à elle de favoriser la reconnaissance d'un signal de clivage/polyadénylation faible par un site 3' d'épissage fort (Millevoi *et al.*, 2006). Ces deux interactions permettraient d'expliquer que U2AF65 inhibe *in vitro* l'activité de polyadénylation de la PAP dans certaines expériences (Ko et Gunderson, 2002), alors que pour d'autres expériences U2AF65 active à la fois le clivage et la polyadénylation (Millevoi *et al.*, 2006).

Une autre interaction directe entre les complexes d'épissage et de clivage/polyadénylation a également été montrée *via* la snRNP U2 et le complexe CPSF (Kyburz *et al.*, 2006). En effet, la purification du complexe CPSF à partir de thymus de veau a permis d'identifier plusieurs sous-unités de la snRNP U2 et le snRNA U2 (Kyburz *et al.*, 2006). *In vitro*, la présence de CPSF est nécessaire pour un épissage efficace d'exon 3' terminal, et la mutation du site de fixation de la snRNP U2 dans l'ARNpré-m diminue l'efficacité de clivage (Kyburz *et al.*, 2006). Ces expériences montrent également que la digestion du snRNA U2 conduit à une diminution du clivage d'un ARNpré-m modèle (Kyburz *et al.*, 2006). L'interaction entre CPSF et la snRNP U2 permet donc l'utilisation des exons 3' terminaux (Kyburz *et al.*, 2006).

D'autres éléments comme la snRNP U1 sont également impliqués dans le couplage entre l'épissage et le clivage/polyadénylation. En effet le blocage de la snRNP U1 à l'aide d'anticorps dans des extraits nucléaires de cellules HeLa bloquent la réaction de

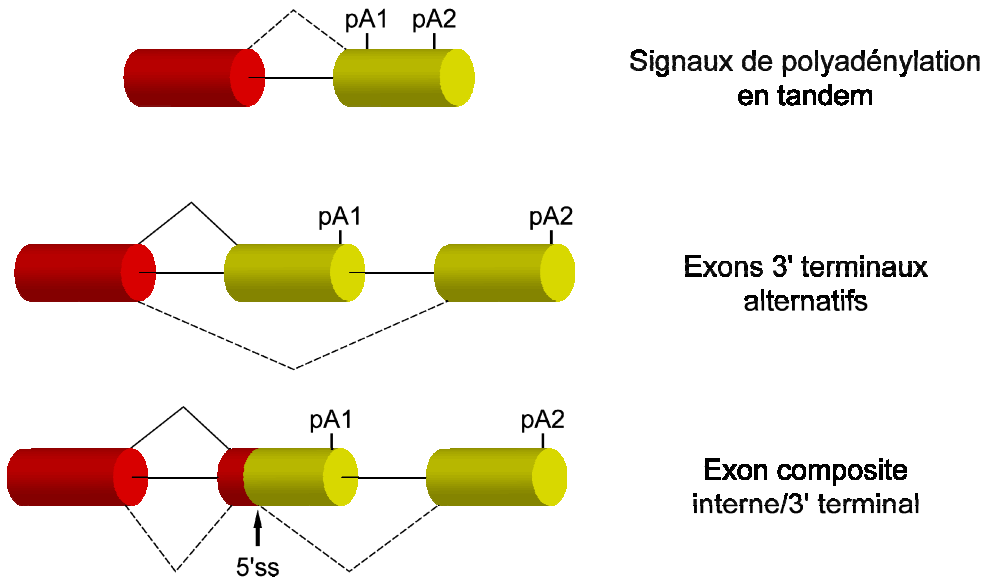


Figure 19- Différents types de polyadénylation alternative. Les exons constitutifs sont représentés en rouge et les exons alternatifs en jaune. Les traits pleins et pointillés indiquent les différents évènements d'épissage. pA : signal de polyadénylation ; 5'ss : site 5' d'épissage.

polyadénylation (Moore et Sharp, 1985). De plus, le snRNA U1 est coimmunoprécipité avec la PAP (Raju et Jacob, 1988). Les protéines U1A et U170K, composants de la snRNP U1, interagissent avec le domaine carboxy terminal de la PAP (Gunderson *et al.*, 1997; Gunderson *et al.*, 1998). La fixation de ces facteurs *via* leur domaine PRD inhibe l'activité de polyadénylation de la PAP (Ko et Gunderson, 2002). La protéine U170K est également coimmunoprécipitée avec le facteur CF Im (Awasthi et Alwine, 2003).

Des facteurs ne participant pas directement aux deux réactions peuvent également intervenir dans leur couplage. Le facteur SRm160/300 est coactivateur de l'épissage constitutif et de l'épissage dépendant des ESE en interagissant au travers de l'intron et/ou de l'exon avec les facteurs d'épissage directement recrutés sur l'ARNpré-m (Blencowe *et al.*, 1998; Eldridge *et al.*, 1999; Blencowe *et al.*, 2000). SRm160/300 active également le clivage des ARNpré-m *in vitro* et dans des cellules en culture *via* son domaine amino terminal PWI (proline tryptophane isoleucine) (McCracken *et al.*, 2002; Szymczyna *et al.*, 2003). Une interaction entre CPSF et SRm160 a été démontrée *in vitro*, et une interaction génétique entre SRm160 et Cstf50 a été mise en évidence chez *C. elegans* (McCracken *et al.*, 2002; McCracken *et al.*, 2003). SRm160 active plus efficacement le clivage d'un ARNpré-m capable de recruter un complexe d'épissage fonctionnel (McCracken *et al.*, 2002). Il a été proposé que SRm160 établisse un lien entre les facteurs d'épissage et de clivage/polyadénylation, permettant ainsi de coordonner les deux machineries afin de sélectionner correctement les exon 3' terminaux (McCracken *et al.*, 2002). Une étude réalisée chez *C. elegans* montre que l'orthologue la protéine humaine hRED120 permettait également de coupler physiquement les complexes d'épissage et de clivage/polyadénylation (Fortes *et al.*, 2007). La protéine hRDE120 interagit avec SRm160 et est capable d'interagir avec l'ARN *via* un domaine PWI. Ce facteur serait impliqué dans l'épissage et le processus de clivage/polyadénylation d'ARNpré-m spécifiques qui ne sont pas encore connus (Fortes *et al.*, 2007).

3.2. Régulation des exons 3' terminaux alternatifs

Les exons 3' terminaux alternatifs présentent des sites 3' d'épissage alternatifs et des signaux de clivage/polyadénylation alternatifs. Comme présenté précédemment, les exons 3' terminaux répondent au modèle de la définition de l'exon, ainsi la sélection d'un site 3'

d'épissage peut entraîner la sélection du signal de polyadénylation qui lui est associé, et inversement la sélection d'un signal de polyadénylation peut conduire à la sélection du site 3' d'épissage situé en amont.

Il existe trois types d'exons 3' terminaux alternatifs qui sont définis en fonction de l'organisation des sites d'épissage et des signaux de clivage/polyadénylation (figure 19). La première catégorie d'exons 3' terminaux alternatifs correspond à des exons présentant plusieurs signaux de polyadénylation disposés en tandem. Dans ce cas, l'utilisation de ces signaux est indépendante de la sélection du site 3' d'épissage en amont. La seconde classe correspond à des exons 3' terminaux alternatifs mutuellement exclusifs qui sont en compétition pour leur utilisation dans un ARNm. Leur sélection peut être régulée par la reconnaissance des sites 3' d'épissage et des signaux de clivage/polyadénylation. La troisième catégorie correspond également à des signaux de clivage/polyadénylation mutuellement exclusifs, mais un des exons possède un site 5' d'épissage interne. Cet exon appelé exon composite interne/3' terminal peut être, comme son nom l'indique, utilisé en tant qu'exon interne quand son site 5' d'épissage est reconnu, ou comme exon terminal quand son signal de polyadénylation est sélectionné. La régulation de l'utilisation de ces exons fait donc intervenir une compétition entre le site 5' d'épissage interne et le signal de polyadénylation.

La force des signaux de polyadénylation joue un rôle important dans la production de l'extrémité 3' des ARNm. En effet, la sélection de l'exon le plus proximal va conduire au clivage de l'ARNpré-m et donc prévenir l'utilisation des exons situés en aval. Il a été montré par analyses bioinformatiques d'EST (Tian *et al.*, 2005), confirmées expérimentalement (Moucadel *et al.*, 2007), que les signaux de polyadénylation distaux sont plus forts que les signaux les plus proximaux, les signaux distaux devraient donc être reconnus préférentiellement. Cette organisation particulière implique que l'utilisation des exons 3' terminaux alternatifs soit contrôlée par des régulations à la fois de leur site 3' d'épissage et de leur signal de polyadénylation.

3.2.1. Les régulations de la polyadénylation

Les exemples les mieux caractérisés dans la régulation du clivage/polyadénylation illustrent le rôle des facteurs d'épissage comme les facteurs U1A et U170K composant la

snRNP U1. D'autres protéines comme la protéine PTB ou le facteur CstF64 influencent également le processus de clivage/polyadénylation.

3.2.1.1. La protéine UIA

La protéine U1A régule la maturation de l'extrémité 3' terminale de son propre ARNpré-m. Des expériences réalisées *in vitro* ont montré que la protéine U1A se fixe spécifiquement sur une séquence régulatrice présenté dans la 3' UTR de son ARNpré-m. Cette séquence a une structure secondaire particulière qui juxtapose deux régions de sept nucléotides fixant chacune une molécule de U1A de manière coopérative. L'interaction entre U1A et son ARNpré-m inhibe la maturation de l'extrémité 3' de ce dernier, et diminue la quantité d'ARNm mature (Boelens *et al.*, 1993; van Gelder *et al.*, 1993). La protéine U1A inhiberait uniquement la polyadénylation *via* une interaction entre le domaine central du dimère de U1A et l'extrémité carboxy terminale de la PAP (Gunderson *et al.*, 1997; Klein Gunnewiek *et al.*, 2000).

La protéine U1A peut également activer la polyadénylation. Cette protéine se fixe sur une séquence activatrice, située en amont du signal de polyadénylation de l'ARNpré-m tardif du virus SV40 (Lutz et Alwine, 1994). La protéine U1A interagirait alors avec la sous-unité de 160 kDa du facteur CPSF, stabilisant ainsi la liaison de ce facteur sur l'hexanucléotide AAUAAA (Lutz *et al.*, 1996). De plus, l'ajout d'une protéine U1A recombinante au cours de ces expériences augmente la polyadénylation dépendante de CPSF.

Les différents types de régulation exercés par la protéine U1A pourraient faire intervenir des mécanismes moléculaires distincts. La répression de la polyadénylation de l'ARNpré-m du virus SV40 fait intervenir deux molécules de U1A libre. A l'inverse l'activation de la polyadénylation de l'ARNpré-m SV40 fait intervenir U1A présente dans la snRNP U1 (Lutz et Alwine, 1994). De plus, le snRNA U1 se fixe en amont du signal de polyadénylation AAUAAA de l'ARNpré-m tardif du virus SV40 (Wassarman et Steitz, 1993). La régulation par la protéine U1A dépendrait donc de sa participation ou non au complexe snRNP U1.

3.2.1.2. La protéine UI-70K

La protéine U1-70K, présente au sein de la snRNP U1, peut réguler le processus de clivage/polyadénylation de nombreux ARNpré-m viraux.

Le signal de polyadénylation tardif de l'ARNpré-m du papillomavirus bovin (BPV) est contrôlé par une séquence USE qui contient un pseudo-site 5' d'épissage. La fixation *in vitro* de la snRNP U1 sur le pseudo-site inhibe la réaction de polyadénylation. La répression de la polyadénylation est due à l'interaction entre la snRNP U1 et la PAP *via* la protéine U1-70K (Furth *et al.*, 1994; Gunderson *et al.*, 1998).

La snRNP U1 est également responsable de la répression du signal de polyadénylation du virus HIV1 contenu dans la 5' LTR (*long terminal repeat*). Cette répression faisant intervenir la protéine U1-70K est due à la fixation de la snRNP U1 sur un site 5' d'épissage situé en aval du signal de polyadénylation (Ashe *et al.*, 1997; Ashe *et al.*, 2000). Pour cet ARNpré-m, c'est l'étape de clivage qui serait réprimée par la protéine U1-70K. En effet, le site 5' d'épissage serait éliminé de l'ARNpré-m avant que le site de clivage situé en aval ne puisse être utilisé (Vagner *et al.*, 2000).

3.2.1.3. La protéine PTB

La PTB permet d'activer ou de réprimer le processus de clivage/polyadénylation des ARNpré-m.

La surexpression de la PTB dans des cellules en culture conduit à une diminution du clivage des ARNpré-m issus de minigènes possédant des sites de polyadénylation de forces différentes. Des expériences effectuées *in vitro* montrent que cet effet résulterait d'une compétition entre la PTB et le facteur CstF64 pour la fixation sur la séquence DSE du signal de polyadénylation (Castelo-Branco *et al.*, 2004). Cependant, dans cette même étude, la déplétion de la PTB dans les cellules en culture n'a pas conduit à une augmentation de l'efficacité de clivage des mêmes ARN (Castelo-Branco *et al.*, 2004).

La PTB peut également activer l'utilisation des signaux de clivage/polyadénylation. Par exemple, la PTB régule le signal de polyadénylation de l'exon 7a du gène GFAP (*Glial Fibrillary Acidic Protein*). En effet, la co-injection dans des cellules en culture d'un minigène contenant la région 3' terminale de ce gène et d'un vecteur permettant l'expression de la PTB, conduit à une augmentation de l'utilisation du signal de polyadénylation de l'exon 7a

(Blechingberg *et al.*, 2007). Ces mêmes expériences ont montré que les protéines SR permettent également d'activer ce signal de polyadénylation. Le signal de polyadénylation de l'enzyme COX2 (*cyclooxygenase*) est également régulé par la PTB. L'adressage de la PTB en amont d'un élément USE favorise l'utilisation du signal de polyadénylation proximal (Hall-Pogar *et al.*, 2007). Cette activation est dépendante de l'USE et peut également être observée suite à l'adressage des protéines U1A, PSF et p54. De plus, ces protéines sont fixées *in vitro* sur un ARN contenant la séquence USE (Hall-Pogar *et al.*, 2007). La fixation des protéines PTB, U1A, PSF et p54 sur l'USE favoriserait l'utilisation du signal de polyadénylation proximal de COX2 (Hall-Pogar et al., 2007). Le même type de régulation a été observé in vitro en utilisant l'ARNpré-m de HBB (Human β-globin) (Millevoi et al., 2009). Dans ce cas, la fixation de la PTB sur un USE favoriserait également le recrutement de la hnRNP H. Ces deux protéines participeraient à l'activation des réactions de clivage et de polyadénylation de l'ARNpré-m (Millevoi et al., 2009).

3.2.1.4. La protéine CstF64

La protéine CstF64 participe à la mise en place du complexe de clivage en interagissant avec les régions DSE présentes en aval des signaux de polyadénylation. Celle-ci permettrait également de réguler l'utilisation des signaux de polyadénylation alternatifs.

Dans les cellules de poulet, CstF64 est un facteur limitant de la formation de l'hétérodimère CstF et pourrait donc constituer un facteur régulateur (Takagaki et Manley, 1998). Il a été proposé que la fixation de CstF64 sur une séquence DSE résulte d'un équilibre entre la concentration nucléaire de CstF64 et son affinité pour la séquence DSE. Ainsi, une forte concentration de CstF64 serait nécessaire à la reconnaissance du signal de polyadénylation, appelé μs, de l'ARNpré-m des chaînes lourdes des IgM (Takagaki *et al.*, 1996; Takagaki et Manley, 1998). Le recrutement de CstF64 peut être modulé par des séquences accessoires. Par exemple, la séquence présente en amont du signal de polyadénylation de l'ARNpré-m du complément C2, favorise la liaison de CstF64 sur l'ARNpré-m (Moreira *et al.*, 1998). D'autre part, l'injection d'un minigène contenant deux signaux de polyadénylation dans des macrophages murins RAW 264.7 surexprimant la protéine CstF64, favorise l'utilisation du signal de polyadénylation le plus proximal du minigène (Shell *et al.*, 2005).

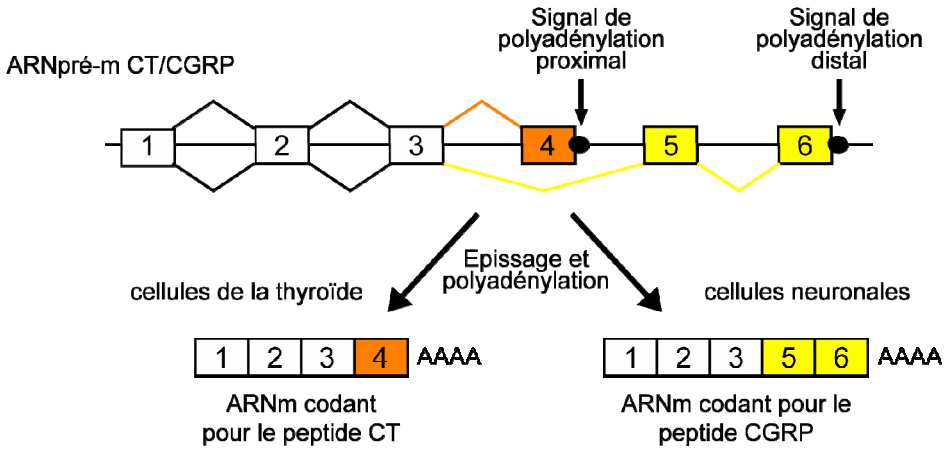


Figure 20- Représentation de la maturation de l'extrémité 3' de l'ARNpré-m « *calcitonin/calcitonin gene-related peptide* » (CT/CGRP). Les évènements d'épissage sont indiqués par des traits plein.

3.2.2. Modèles de la régulations des exons 3' terminaux alternatifs

3.2.2.1. Le modèle du gène calcitonin/calcitonin gene-related peptide

L'ARNpré-m « *calcitonin/calcitonin gene-related peptide* » (CT/CGRP) contient deux exons 3' terminaux qui sont différentiellement utilisés selon le type cellulaire. Dans les cellules T de la thyroïde, l'exon 4 est utilisé en tant qu'exon 3' terminal dans l'ARNm qui code pour le peptide CT. Dans les cellules neuronales cet exon est exclu et l'épissage des exons 3, 5 et 6 produit un ARNm codant pour le peptide CGRP (figure 20). L'exon 4 est délimité par un site 3' d'épissage et un signal de polyadénylation qui sont faibles. L'inclusion et la polyadénylation de cet exon sont dépendantes de la reconnaissance d'un élément intronique activateur. Cet ISE contient un pseudo-site 5' d'épissage encadré par une séquence riche en pyrimidines et une répétition de U (Lou *et al.*, 1995). La mutation d'un de ces trois éléments de l'ISE conduit à l'exclusion de l'exon 4 dans des cellules en culture. Des expériences de clivage réalisées *in vitro*, montrent que le pseudo-site 5' d'épissage et la région riche en pyrimidines sont requis pour un clivage efficace au niveau du signal de polyadénylation de l'exon 4. D'autre part, la répétition de U contrôlerait la reconnaissance du site 3' d'épissage de l'exon 4 (Lou *et al.*, 1996; Zhu *et al.*, 2003). Une autre séquence régulatrice permettant l'utilisation du site 3' d'épissage a été caractérisée dans l'exon 4. Ces séquences activatrices sont reconnues par un grand nombre de facteurs.

La protéine SRp20 interagit *in vitro* avec la séquence riche en pyrimidines. Sa surexpression dans des cellules en culture favorise l'utilisation de l'exon 4. Celle-ci participerait à la mise en place du complexe de clivage/polyadénylation. En effet, l'expression dans des cellules en culture, d'une protéine SRp20 ne possédant plus de domaine RS est corrélée à une baisse du recrutement du facteur CstF64 au niveau du signal de polyadénylation de l'exon 4 (Lou *et al.*, 1998).

In vitro, la PTB interagit à la fois avec la séquence localisée directement en aval du signal de polyadénylation de l'exon 4 et celle riche en pyrimidines. Dans des cellules en culture, la surexpression de la PTB favorise l'utilisation de l'exon 4 (Lou *et al.*, 1999). Suite à des expériences de titration effectuées *in vitro*, il a été proposé que la PTB activerait l'étape de clivage (Lou *et al.*, 1996).

La protéine TIAR interagit *in vitro* avec la séquence riche en U, et la snRNP U6 se fixe sur le pseudo-site 5' d'épissage. Ces interactions sont dépendantes l'une de l'autre. Ces

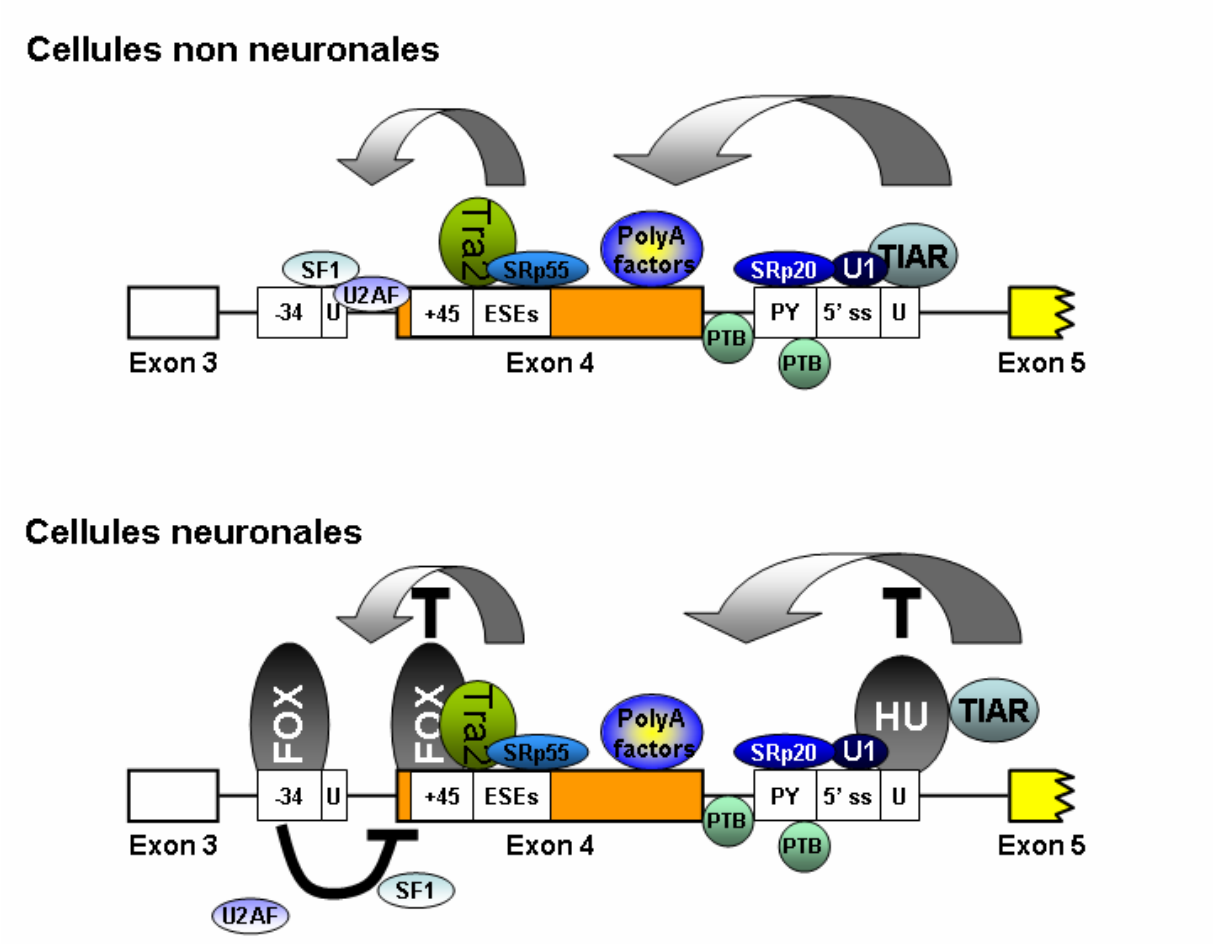


Figure 21- Régulation cellulaire de l'exon 4 du gène CT/CGRP. D'après (Zhou *et al.*, 2007). Dans les cellules non neuronales l'exon 4 est utilisé. Les différents facteurs activateurs ainsi que leurs sites de fixation sont représentés. Dans les cellules neuronales, la répression de l'utilisation de l'exon 4, médiée par les protéines FOX et HU est représentée. « PY » correspond à la région riche en pyrimidines, « U1 » à la snRNP U1, et « *PolyA factors* » aux facteurs de clivage/polyadénylation.

deux facteurs activeraient la reconnaissance du site 3' d'épissage de l'exon 4 (Zhu *et al.*, 2003). La snRNP U1 et le facteur ASF/SF2 se fixent également sur le pseudo-site 5' d'épissage. La snRNP U1 participerait à l'activation du clivage au niveau du signal de polyadénylation de l'exon 4 (Lou *et al.*, 1996), et le rôle de ASF/SF2 n'est pas encore caractérisé.

Les protéines hTraβ2 et SRp55 interagissent *in vitro* avec la séquence activatrice localisée dans l'exon 4 et favorisent la reconnaissance du site 3' d'épissage (Tran *et al.*, 2003; Tran et Roesser, 2003). Le recrutement de ces facteurs permet l'utilisation de l'exon 4 dans les cellules non neuronales. Des expériences récentes ont également permis de déterminer quels sont les facteurs impliqués dans la répression de l'utilisation de cet exon dans les cellules neuronales. L'ensemble de ces interactions est représenté dans la figure 21.

La déplétion des protéines Fox 1 et 2 dans des cellules en culture, favorise l'utilisation de l'exon 4. Ces protéines interagissent *in vitro* avec les motifs UGCAUG présents dans l'intron et dans l'exon, autour du site 3' d'épissage de l'exon 4. Enfin, la fixation des protéines Fox sur cet élément régulateur bipartite, empêche le recrutement du facteur U2AF 65 au niveau du site 3' d'épissage (Zhou *et al.*, 2007; Zhou et Lou, 2008). Les protéines Hu, spécifiques des cellules neuronales, entrent en compétition avec la protéine TIAR pour la fixation sur la séquence intronique riche en U (Zhu *et al.*, 2006). L'expression de dominants négatifs des protéines Hu dans des cellules en culture, conduit à l'inclusion de l'exon 4. Le saut de l'exon 4 dans les cellules neuronales est donc dû à la compétition de fixation entre les facteurs activateurs et les protéines FOX et Hu.

3.2.2.2. Le modèle des chaînes lourdes des immunoglobulines

L'ARNpré-m des chaînes lourdes des IgM contient deux exons 3' terminaux alternatifs. L'exon proximal, appelé C4μs, est un exon composite interne 3'/terminal. La maturation 3' terminale cet ARNpré-m varie au cours de la différenciation des lymphocytes B (figure 22). Dans les lymphocytes immatures, l'exon Cμ4 est utilisé en tant qu'exon interne et épissé aux exons M1 et M2. L'ARNm ainsi formé code pour un récepteur membranaire. Dans les cellules B matures, l'exon Cμ4S est utilisé comme exon 3' terminal pour produire un ARNm qui code pour la forme sécrétée des IgM.

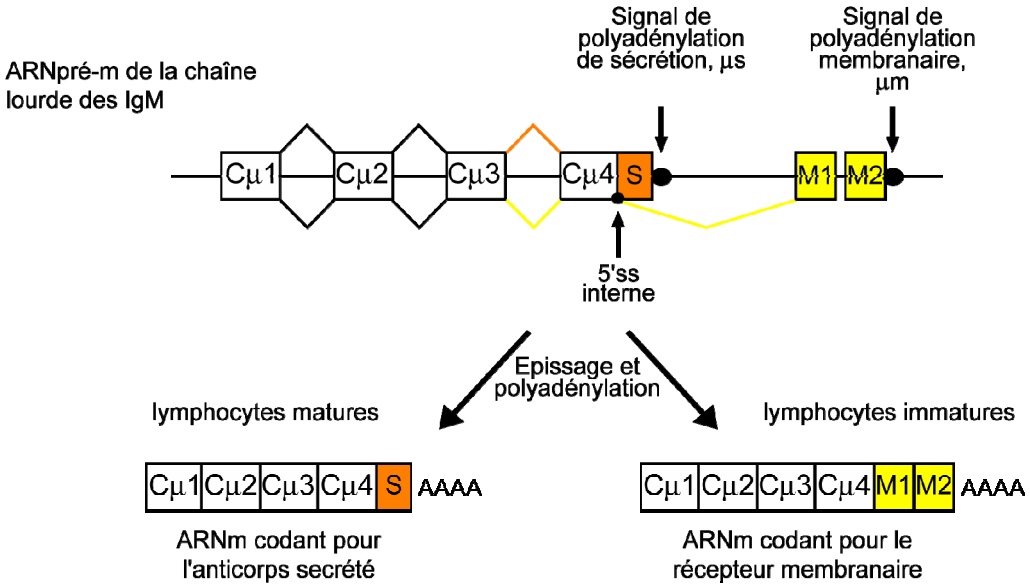


Figure 22- Maturation de l'extrémité 3' terminale de l'ARNpré-m de la chaîne lourde des IgM. D'après (Phillips *et al.*, 2001). Les évènements d'épissage sont représentés par des traits, et le site 5' d'épissage par 5'ss.

Il a été décrit que la présence d'un site de fixation de la snRNP U1 entre un site 3' d'épissage et un signal de polyadénylation, diminue à la fois l'épissage et la polyadénylation (Cooke *et al.*, 1999). Cette organisation particulière est retrouvée dans les exons composites. La reconnaissance du site d'épissage interne pourrait donc être un évènement important dans l'utilisation différentielle d'un tel exon, et serait la cible de régulations. Dans le cas des chaînes lourdes des IgM, les résultats obtenus semblent indiquer que la majorité des régulations s'exercent sur le signal de polyadénylation µs de l'exon Cµ4S.

Le signal de polyadénylation µs est faible et très sensible à la concentration du facteur CstF64. En effet, seule la surexpression de ce facteur peut activer l'utilisation de µs dans des cellules en culture (Takagaki *et al.*, 1996; Takagaki et Manley, 1998). La concentration de CstF64 augmente au cours de la différenciation, mais cette variation n'est pas corrélée à la modification de la maturation de l'ARNpré-m codant pour les IgM (Martincic *et al.*, 1998). L'utilisation du signal µs résulterait d'une variation de l'affinité de liaison du facteur CstF64 plus que d'une modification de la concentration intracellulaire de ce facteur (Edwalds-Gilbert et Milcarek, 1995). Des expériences effectuées *in vitro*, et à l'aide de cellules en culture dans lesquelles des vecteurs contenant des signaux de polyadénylation µs ont été insérés en aval d'un gène rapporteur luciférase, ont montré que la reconnaissance du signal µs serait controlée par la protéine U1A. Cette protéine se fixe sur trois séquences régulatrices situées en amont du signal µs. De plus, U1A inhibe la réaction de polyadénylation d'un ARNpré-m clivé au site µs, probablement en affectant l'activité de la PAP (Phillips *et al.*, 2001). Deux autres sites de liaison ont été identifiés en aval du signal µs, entre les séquences riches en dinucléotides GU qui constituent l'élément DSE. L'interaction de la protéine U1A avec ces sites, diminue la liaison du facteur CstF64 sur les séquences riches en GU, et inhibe la réaction de clivage au niveau du site de polyadénylation µs (Phillips *et al.*, 2004). D'autres expériences utilisant un signal de polyadénylation µs entier dans un vecteur contenant l'extrémité 3' terminale du gène µ, ont conduit à des observations différentes. De nombreuses mutations des séquences entourant l'hexanucléotide AAUAAA du signal de polyadénylation µs, ne modifient pas les régulations observées au cours de la maturation de cellules B (Peterson *et al.*, 2006). De plus, la régulation de l'utilisation du signal µs n'est pas modifiée dans des cellules surexprimant le facteur U1A. Les mutations des séquences de fixation de cette protéine ne modifient pas non plus ces régulations (Peterson *et al.*, 2006). Ces expériences de mutagenèse montrent également qu'il est difficile d'inactiver le signal de polyadénylation µs. Les séquences régulatrices participeraient donc à la reconnaissance du

signal μs par les facteurs généraux de la machinerie de clivage/polyadénylation (Peterson *et al.*, 2006).

4. Couplage épissage clivage/polyadénylation et transcription

Les trois réactions permettant la maturation des ARNpré-m que sont l'ajout de la coiffe, l'épissage et le clivage/polyadénylation peuvent être réalisées individuellement sur des substrats préalablement transcrits *in vitro*. Cependant de nombreuses études ont permis d'établir que ces différents processus sont étroitement liés à la transcription *in vivo* (McCracken *et al.*, 1997; Proudfoot *et al.*, 2002; Zorio et Bentley, 2004; Dye *et al.*, 2006; West *et al.*, 2008; Moore et Proudfoot, 2009). Le couplage entre ces différents mécanismes est aussi bien spatio-temporel que fonctionnel. Les évènements de maturation des ARNpré-m sont co-transcriptionnels plus que post-transcriptionnels.

4.1. Couplage spatial

La transcription et la maturation des ARNpré-m sont des mécanismes concertés dans l'espace. Des expériences d'immunofluorescence sur des cellules en culture ont révélé l'existence, dans le noyau, de structures particulières appelées « *speckles* » qui seraient des sites de stockage et/ou d'assemblage de facteurs d'épissage (Spector, 1993; Lamond et Spector, 2003; Handwerger et Gall, 2006; Sanchez-Alvarez *et al.*, 2006). Il existe un lien entre la forme de ces structures et la transcription puisque suite au blocage de la transcription celles-ci s'arrondissent et augmentent de taille (Carmo-Fonseca *et al.*, 1992). De plus, dans les cellules HeLa, quand la transcription d'un gène exogène est activée, le facteur ASF/SF2 est relocalisé des « *speckles* » vers des foyers de transcription (Misteli *et al.*, 1997). Des expériences de microscopie ont également montré que des protéines SR et le facteur de clivage CstF64 étaient présents dans les foyers de transcription (Neugebauer et Roth, 1997; Zeng *et al.*, 1997). Leur co-localisation varie en fonction du niveau d'activité de la transcription (Zeng *et al.*, 1997) et l'expression d'une ARNpol II ayant un CTD tronqué inhibe le ciblage des protéines d'épissage aux sites de transcription (Misteli et Spector, 1999). Les ARNpré-m seraient donc maturés au niveau « d'*usines* » de transcription qui regroupent l'ARNpol II et les facteurs nécessaires à la maturation (Misteli *et al.*, 1997; Bentley, 2005; Hagiwara et Nojima, 2007).

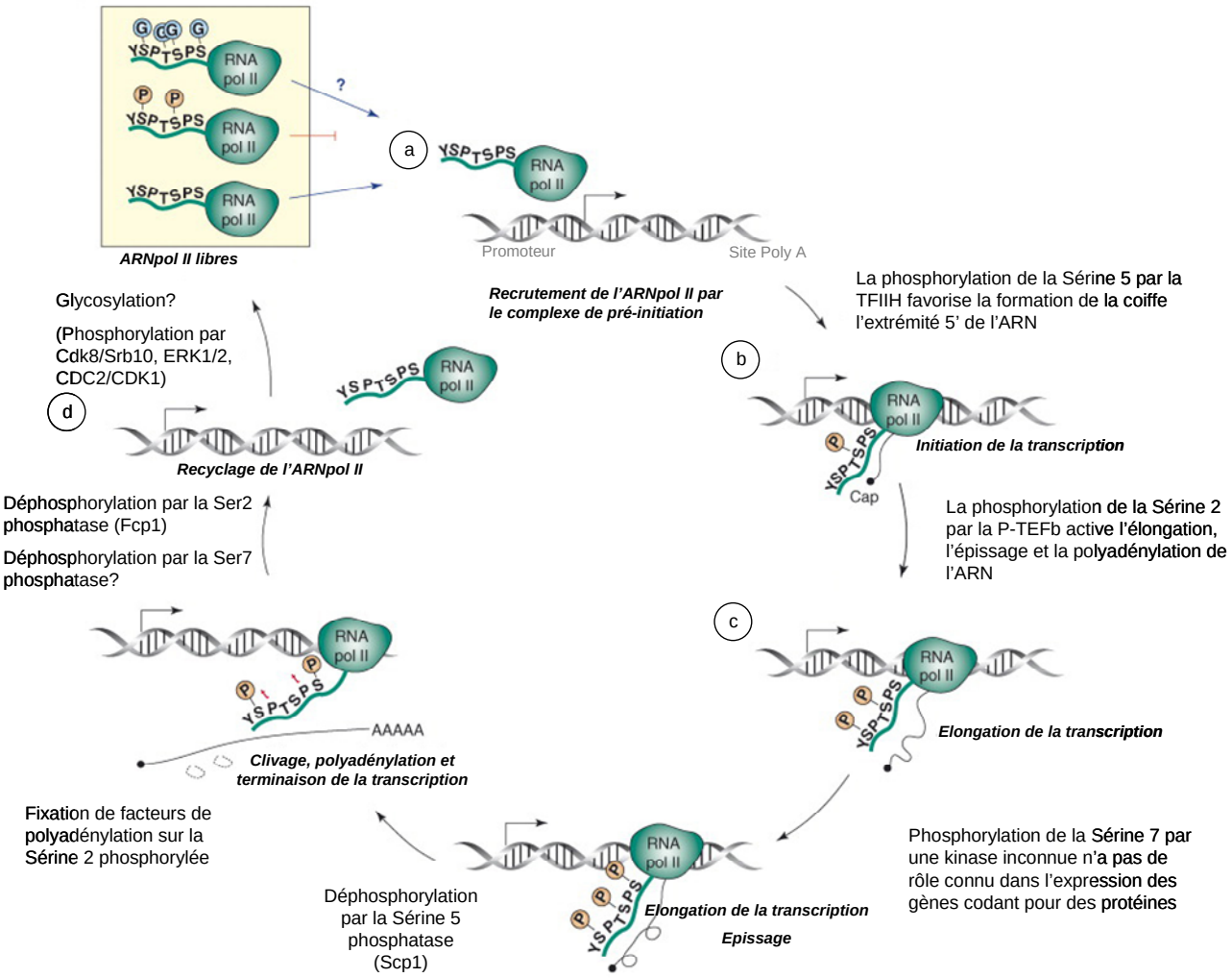


Figure 23- Etat de phosphorylation du CTD de l'ARNpol II au cours de la transcription. D'après (Egloff et Murphy, 2008). Très tôt après l'initiation de la transcription le CTD est phosphorylé sur la sérine en position 5 ce qui permet le recrutement des facteurs de la coiffe. Suite à l'initiation de l'élongation, la sérine en position 2 est phosphorylée ce qui permet le recrutement de facteurs d'épissage. Durant l'élongation, la sérine 5 subit une déphosphorylation. La phosphorylation unique de la sérine 2 favoriserait le recrutement de facteurs de clivage/polyadénylation. Enfin, les sérines 2 et 7 sont déphosphorylées et l'ARNpol II est recyclée. Le rôle de la phosphorylation de la sérine en position 7 dans la régulation des gènes codant pour des protéines n'a toujours pas été caractérisé.

4.2. Couplage temporel

Dès sa transcription, l'ARNpré-m est « décoré » de nombreuses protéines (Bergkessel *et al.*, 2009). Les complexes RNP formés sont remodelés au cours de la maturation du transcrit néo-synthétisé. Différentes sous-unités de ces complexes sont impliquées dans les trois processus de maturation et font le lien avec la transcription. De nombreuses expériences ont permis de montrer que le point central du couplage entre les machineries de maturation des ARNpré-m et la transcription serait le domaine carboxy terminal (CTD) de la grosse sous-unité de l'ARNpol II. Ainsi, l'utilisation dans des cellules en culture d'une ARNpol II tronquée au niveau du CTD conduit à un blocage de l'épissage et du processus de clivage/polyadénylation des ARNpré-m rapporteurs sans affecter la transcription (McCracken *et al.*, 1997).

Le CTD se compose d'un motif YSPTSPS très conservé au cours de l'évolution et dont le nombre de répétitions chez l'homme est de 52. Les sérines en positions 2, 5 et 7 de ce motif sont phosphorylées différentiellement au cours de la transcription (Phatnani et Greenleaf, 2006; Chapman *et al.*, 2007; Moore et Proudfoot, 2009). L'état de phosphorylation des sérines 2 et 5 est impliqué dans la transcription des gènes codant pour des protéines, et celui de la sérine 7 serait impliqué uniquement dans la transcription des snRNA (Egloff *et al.*, 2007; Egloff et Murphy, 2008). Seule l'implication du niveau de phosphorylation des sérines 2 et 5 dans la transcription sera abordée. Lors du recrutement de l'ARNpol II sur le promoteur, le CTD est déphosphorylé. Très tôt après l'initiation de la transcription le CTD est phosphorylé sur la sérine en position 5 ce qui permet le recrutement des facteurs de la coiffe (Cho *et al.*, 1997; McCracken *et al.*, 1997). Une fois que l'élongation de la transcription est initiée, la sérine en position 2 est phosphorylée à son tour. Cet état semble favoriser la fixation de facteurs d'épissage (Kameoka *et al.*, 2004), notamment de la snRNP U1 (Das *et al.*, 2007), et des protéines SR ASF/SF2, SC35, et SRp40 (Yuryev *et al.*, 1996; Kim *et al.*, 1997; Misteli et Spector, 1999; Das *et al.*, 2007). Durant l'élongation, la sérine 5 subit une déphosphorylation. La phosphorylation unique de la sérine 2 serait alors favorable au recrutement de facteurs de clivage/polyadénylation (figure 23) (Ahn *et al.*, 2004; Gudipati *et al.*, 2008).

Ces différentes données montrent l'importance de la phosphorylation du CTD de l'ARNpol II dans un couplage temporel entre la transcription et le recrutement des facteurs participant à la maturation des ARNpré-m. Le CTD pourrait être utilisé comme plateforme au

recrutement de différents composants des machineries de maturation impliqués dans le couplage, les ciblant de ce fait à proximité d'un transcrit naissant (Listerman *et al.*, 2006; Phatnani et Greenleaf, 2006).

4.3. Rôle de la transcription dans la régulation de l'épissage alternatif

L'observation que la transcription et la maturation des ARNpré-m sont couplés tant au niveau spatio-temporel que fonctionnel, suggère que l'épissage alternatif pourrait être régulé par des mécanismes intégrant ces différents processus.

L'influence directe de l'ARNpol II sur la régulation de l'épissage alternatif a été mise en évidence grâce à des expériences basées sur des échanges de promoteurs contrôlant l'expression de minigènes rapporteurs incluant l'exon alternatif. Ainsi, l'exon EDI de la fibronectine est inclus plus efficacement dans l'ARNm lorsque le minigène est placé sous le contrôle de son propre promoteur ou du promoteur CMV que sous le contrôle du promoteur α-globine (Cramer *et al.*, 1997; Cramer *et al.*, 1999). De même, des minigènes dérivés des gènes CD44 et CT/CGRP placés sous le contrôle de promoteurs viraux ou hormono-dépendants, sont à l'origine d'ARNm distincts (Auboeuf *et al.*, 2002). Des régulations différentes sont également observées selon la nature du promoteur hormono-dépendant (Auboeuf *et al.*, 2004). Chez la levure *S. pombe*, l'ARNpré-m rem1 est épissé différentiellement au cours de la méiose quand le gène est placé sous le contrôle de son propre promoteur. Le remplacement du promoteur du gène ubiquitaire cdc2 par le promoteur de rem1 permet l'épissage différentiel de cdc2 au cours de la méiose (Moldon *et al.*, 2008).

Des études effectuées sur des promoteurs hormono-dépendants ont montré que le recrutement de co-régulateurs transcriptionnels sur les promoteurs participe à la régulation de l'épissage alternatif. Des combinaisons différentes de co-régulateurs permettraient la spécificité de l'épissage alternatif. Ainsi, les co-régulateurs CoAA et CAPER α et β sont impliqués dans les régulations de la transcription et de l'épissage de manière dépendante du promoteur (Iwasaki *et al.*, 2001; Auboeuf *et al.*, 2004; Dowhan *et al.*, 2005; Meng *et al.*, 2007). CoAA agirait en recrutant des combinaisons de facteurs d'épissage différentes alors que les protéines CAPER auraient un rôle plus direct dans la régulation de l'épissage. Chaque

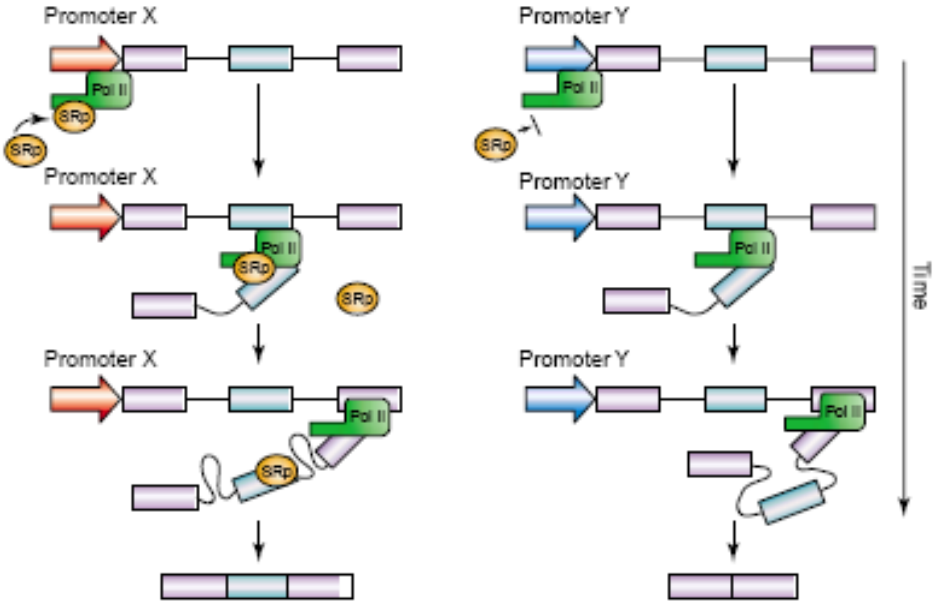


Figure 24- La structure du promoteur influence l'épissage alternatif. D'après (Caceres et Kornblihtt, 2002). Le recrutement des facteurs d'épissage au niveau du CTD de l'ARNpol II est différent selon la structure du promoteur. L'utilisation d'un promoteur (à gauche) peut conduire à l'inclusion d'un exon alternatif (en bleu), au contraire, celui-ci sera exclu suite à l'utilisation d'un autre promoteur (à droite).

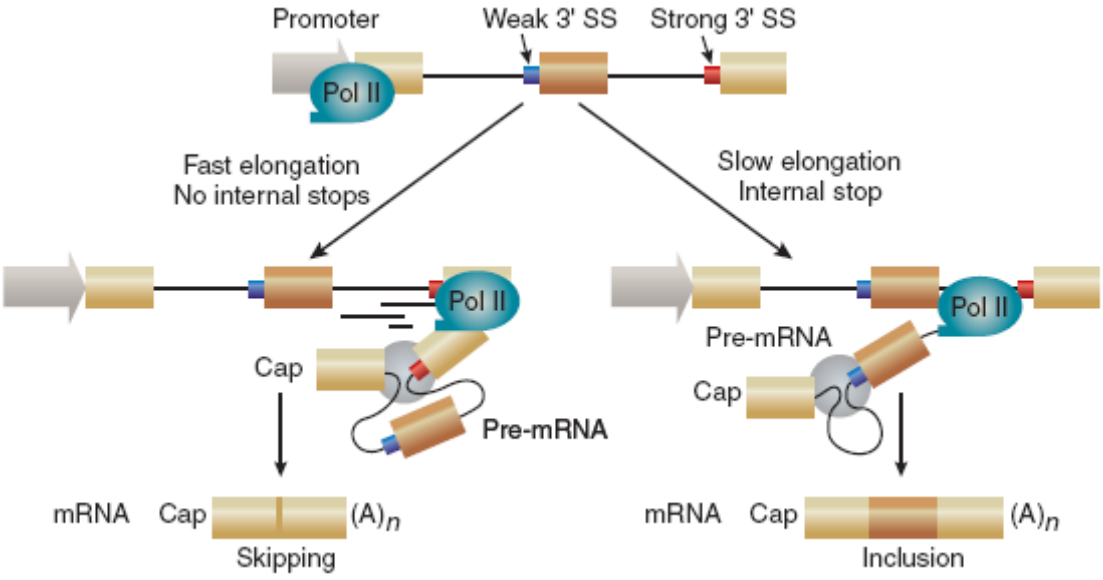


Figure 25- Modèle de régulation de l'épissage alternatif par la vitesse d'élongation de l'ARN polymérase II. D'après (Kornblihtt, 2006). Les exons alternatifs sont représentés en foncé, les exons constitutifs en clair et le spliceosome par un rond gris. Une vitesse d'élongation importante conduit au saut de l'exon alternatif (à gauche), alors qu'une faible vitesse d'élongation de l'ARNpol II permet sa sélection (à droite).

promoteur recruterait donc des combinaisons particulières de co-régulateurs transcriptionnels et de facteurs contrôlant l'épissage alternatif (figure 24) (Auboeuf *et al.*, 2005).

L'épissage alternatif d'un ARNpré-m résulte d'une compétition de sélection entre des sites d'épissage de force différente dont la reconnaissance est modulée par des séquences *cis* (Yu *et al.*, 2008). La cinétique et/ou la processivité de l'ARNpol II au cours de l'élongation de la transcription peut donc également réguler ce processus. De ce fait, en fonction de la processivité de l'ARNpol II, la cinétique d'apparition des sites d'épissage peut être modulée. Si l'ARNpol II est peu processive ou si elle effectue une pause dans l'élongation, l'élimination de l'intron situé en amont d'un exon alternatif sera favorisée bien que son site 3' d'épissage soit faible. Quand l'élongation reprend, la machinerie d'épissage n'a pas d'autre option que d'éliminer l'intron situé en aval, conduisant à l'inclusion de l'exon alternatif. A l'inverse, si l'ARNpol II est processive ou s'il n'y a pas de pause dans l'élongation, les deux exons seront présentés simultanément à la machinerie d'épissage, il y aura alors compétition entre les deux sites 3' d'épissage, et l'exon alternatif, plus faible, sera sauté (figure 25) (Nogues *et al.*, 2003). C'est le cas de l'exon alternatif EDI du gène de la fibronectine. Des expériences réalisées dans des cellules en culture ont permis de montrer que la surexpression du facteur VP16, connu pour activer l'élongation de la transcription, conduit au saut de l'exon EDI (Kadener *et al.*, 2001; Nogues *et al.*, 2002). A l'inverse, l'antigène T qui diminue l'élongation de la transcription, favorise l'inclusion de l'exon EDI.

La régulation de l'épissage alternatif n'est pas uniquement corrélée à l'activité transcriptionnelle des promoteurs (Gendra *et al.*, 2007). En effet, la mutation de séquences nécessaires au recrutement de facteurs d'élongation au niveau du promoteur SV40, induit une augmentation de l'inclusion de l'exon EDI sans altérer la transcription (Kadener *et al.*, 2002). L'influence de la processivité de l'ARNpol II sur l'épissage alternatif a été confirmée *in vivo* chez la drosophile où l'utilisation d'une ARNpol II modifiée de manière à diminuer sa vitesse d'élongation conduit à une modification de l'épissage de l'ARNpré-m endogène de *Ultrabithorax*. Ce promoteur favorise l'utilisation des sites d'épissage proximaux et de ce fait la synthèse d'un ARNm court (de la Mata *et al.*, 2003). Ces dernières expériences ont également permis de montrer que la processivité de l'ARNpol II est dépendante du promoteur utilisé. Chaque promoteur serait donc capable de recruter des complexes d'élongation différents, influençant ainsi le choix des sites d'épissage et donc la régulation des exons alternatifs. La composition de ce complexe d'élongation de la transcription peut également

être modifiée à posteriori. En effet, l'insertion de séquences particulières dans un minigène dérivé du gène *FGFR2* conduit à la régulation co-transcriptionnelle de l'épissage alternatif. L'insertion de l'élément MAZ4, connu pour provoquer des pauses de l'ARNpol II, entraîne une modification du profil d'épissage de l'ARNpré-m (Robson-Dixon et Garcia-Blanco, 2004). Ces variations d'épissage résultent non seulement d'un délai dans l'apparition des séquences *cis*, mais aussi de la modification de la composition du complexe d'élongation qui dépend de la nature du promoteur.

D'autres éléments permettant une pause de l'ARNpol II ont également été caractérisés. En effet, des études récentes menées par l'équipe de Proodfoot montrent que la terminaison de la transcription par la RNApol II est consécutive au clivage des ARN transcrits (West *et al.*, 2008). Ce clivage nécessite la reconnaissance d'un signal de polyadénylation. Dans le cas de l'ARNpré-m de la β-globine, une séquence additionnelle appelée « séquence terminatrice » située en aval du site de polyadénylation est nécessaire au clivage (Proudfoot *et al.*, 2002). Cette séquence permet également le clivage et l'utilisation d'un site de polyadénylation faible même s'il est en compétition avec un site distal plus fort (West et Proudfoot, 2009).

OBLECTIFS DE L'ETUDE

1. L'épissage d'un exon 3' terminal *in vivo*

L'épissage des exons 3' terminaux joue un rôle majeur dans le contrôle de l'expression des gènes. Il permet la production d'isoformes protéiques aux extrémités terminales spécifiques mais également d'ARNm avec des séquences 3'UTR distinctes qui peuvent être régulés différentiellement. D'un point de vue mécanistique, la régulation des exons 3' terminaux est particulièrement complexe puisqu'elle nécessite le contrôle coordonné des réactions d'épissage et de clivage/polyadénylation, cette dernière étant elle-même couplée à la fin de la transcription. Du fait de cette complexité peu de modèles d'études de la régulation des exons 3' terminaux ont été développés.

Le premier modèle décrit correspond l'ARNpré-m de la calcitonine/CGRP. Les études centrées sur la régulation de l'exon 3' terminal proximal ont permis d'identifier une séquence activatrice intronique complexe qui active *in vitro* les réactions d'épissage et de clivage/polyadénylation (Lou *et al.*, 1995). Comme cela vient d'être détaillé dans l'introduction, plusieurs facteurs interagissant avec cette séquence régulatrice ont été identifiés mais leurs mécanismes d'action n'ont toujours pas été caractérisés.

Le deuxième modèle qui a été extensivement étudié est l'ARNpré-m des immunoglobulines. Celui-ci possède un exon composite 3' terminal/interne appelé Cμ4S qui est différentiellement utilisé selon le stade de maturation des lymphocytes B. Des études réalisées à la fin des années 90 indiquent que le signal de polyadénylation μs de cet exon joue un rôle important dans la régulation (Takagaki *et al.*, 1996; Takagaki et Manley, 1998). Dans des cellules en culture, la surexpression du facteur CstF64 entraîne l'activation de ce signal de polyadénylation suggérant un modèle de régulation simple basé sur une variation de l'activité de ce facteur au cours de la différenciation des lymphocytes. Cependant une analyse récente par mutation du signal de polyadénylation μs a montré que celui-ci est difficile à inactiver, suggérant ainsi que des séquences régulatrices participeraient à sa reconnaissance par la machinerie de polyadénylation.

Pour déterminer les bases moléculaires de la régulation des exons 3' terminaux il manquait un système d'étude *in vitro*, permettant de décrire l'épissage, le processus clivage/polyadénylation et la transcription. A ces fins, des extraits nucléaires de cellules en culture, principalement de cellules HeLa, ont été développés pour étudier la réaction

d'épissage. Ils ont permis entre autre d'accéder aux intermédiaires de la réaction d'épissage (Padgett *et al.*, 1984), de détailler la composition du spliceosome (Jurica *et al.*, 2002; Zhou *et al.*, 2002) et de décrire les réarrangements moléculaires (Makarov *et al.*, 2002) et structuraux (Deckert *et al.*, 2006; Behzadnia *et al.*, 2007; Cohen-Krausz *et al.*, 2007) des complexes d'épissage. L'utilisation des extraits nucléaires de cellules HeLa a également été adaptée afin d'étudier la réaction de clivage/polyadénylation (Moore et Sharp, 1985). Toutefois, ces systèmes d'étude *in vitro* présentent un certain nombre de limites : l'efficacité de l'épissage est moins bonne *in vitro* que *in vivo* (Beyer et Osheim, 1988; Audibert *et al.*, 2002); la taille de l'intron doit être inférieure à 600 nt pour que l'efficacité d'épissage ne soit pas trop diminuée dans des extraits nucléaires, (Hernandez et Keller, 1983; Klinz et Gallwitz, 1985); la taille de l'ARNpré-m étudié est limitée à moins de 2000 nt, principalement dû aux limites de la transcription *in vitro*; la concentration en extrait nucléaire ainsi que la force ionique peuvent influencer le choix des sites d'épissage alternatif (Helfman *et al.*, 1988; Mayeda et Ohshima, 1988). Comme présenté en introduction, la limitation de la taille des introns épissés *in vitro* suggère que les sites d'épissage sont reconnus dans les extraits nucléaires selon le modèle de la définition de l'intron alors que *in vivo*, chez les vertébrés, ils sont majoritairement reconnus selon le modèle de la définition d'exon. Enfin, les systèmes d'étude *in vitro* ne permettent pas le couplage fonctionnel entre la transcription et les réactions de maturation des ARNpré-m d'une part, et entre les réactions d'épissage et de clivage/polyadénylation d'autre part. Ceci est une limitation importante puisqu'il est maintenant clairement établi que ces couplages peuvent jouer un rôle important dans la régulation de certains exons alternatifs. Un certain nombre d'études ont fait part de la mise au point d'extraits nucléaires dans lesquels un couplage fonctionnel entre la transcription et l'épissage (Natalizio et Garcia-Blanco, 2005), ou entre l'épissage et la réaction de clivage/polyadénylation (Helfman *et al.*, 1988; Mayeda et Ohshima, 1988) serait opérationnel. Cependant une autre étude a démontré que l'efficacité de l'épissage n'est pas améliorée dans de tels extraits et que si les deux réactions sont bien fonctionnelles, elles ne sont pas couplées (Lazarev et Manley, 2007).

L'étude des mécanismes de contrôle de l'épissage alternatif nécessite *in fine* de combiner des approches *in vivo* et *in vitro*. D'une part il est nécessaire de disposer d'un modèle d'étude *in vivo* qui reproduise l'épissage endogène de l'ARNpré-m modèle choisi afin d'identifier les éléments impliqués dans la sélection différentielle des sites d'épissage et de

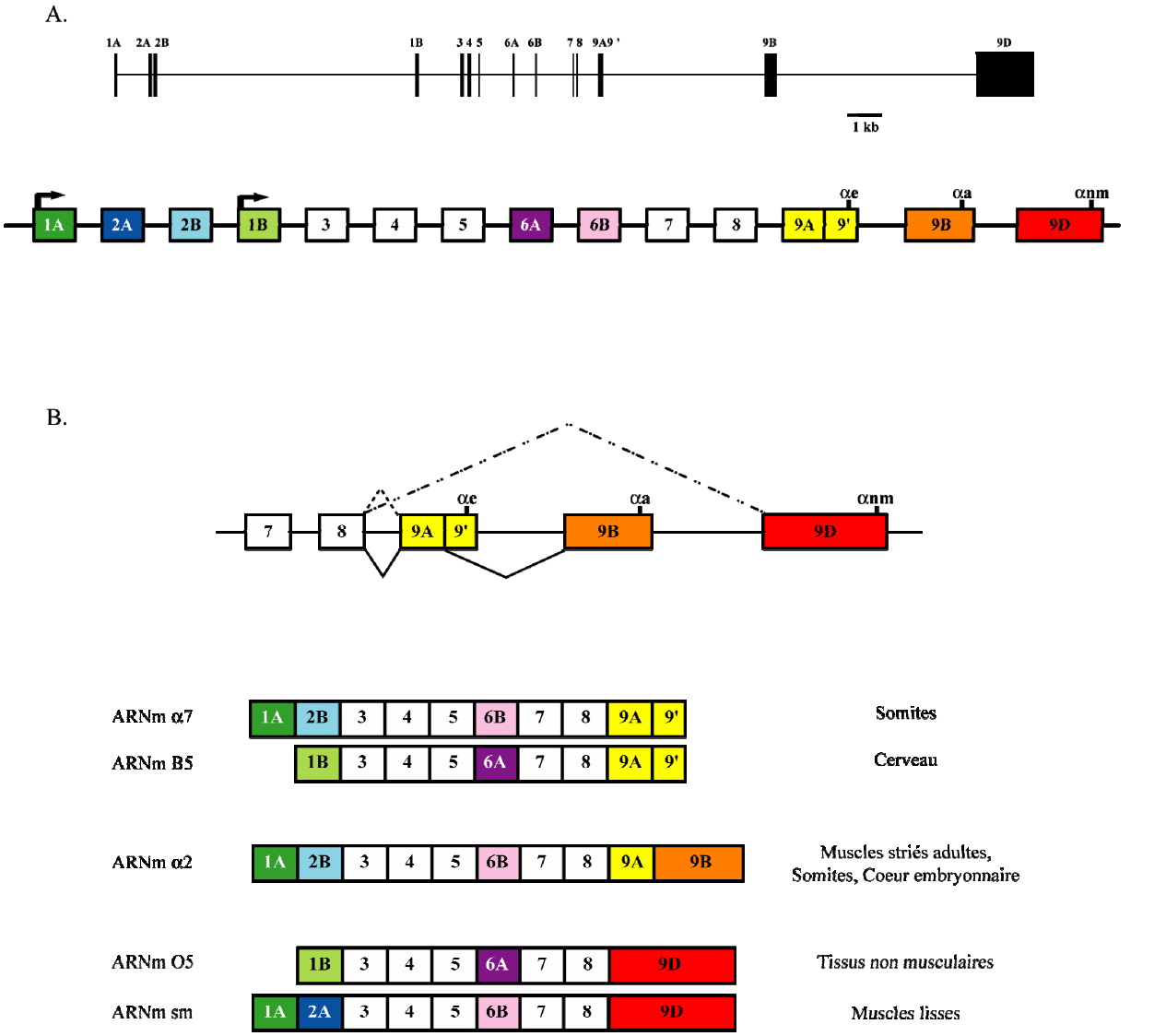


Figure 26- Organisation génomique du gène de l'α-tropomyosine de xénope et profil d'épissage de la région 3' terminale de l'ARNpré-m. Schémas d'après (Le Sommer, 2006). **A.** Présentation à l'échelle puis schématique du gène de l'α-tropomyosine. Les exons épissés de manière alternative sont représentés en couleur et les régions promotrices par des flèches. Les signaux de polyadénylation sont également indiqués. **B.** Epissages alternatifs de la région 3' terminale du gène de l'α-tropomyosine. Les ARNm obtenus ainsi que leur expression tissulaire sont indiqués.

clivage/polyadénylation. D'autre part, les systèmes d'études *in vitro* permettent de disséquer les mécanismes d'action au niveau moléculaire en étudiant la relation entre ces éléments et les machineries d'épissage et de clivage/polyadénylation.

2. Le modèle d'étude

2.1. L'ARNpré-m de l'α-tropomyosine

Nous utilisons le gène de l'α-tropomyosine (α-TM) de xénope pour étudier les maturations des exons 3' terminaux. Ce gène d'une taille de 24 Kb possède 14 exons. Il présente deux promoteurs et trois groupes d'exons épissés de manière alternative dont l'utilisation est à l'origine de 5 ARNm (figure 26). La région 3' terminale de l'ARNpré-m contient trois exons 3'-terminaux alternatifs dont le plus proximal, l'exon 9A9', est un exon composite interne/3' terminal. Son utilisation différentielle en fonction de l'environnement tissulaire et du stade de développement conduit à la synthèse d'ARNm distincts. Suite à l'utilisation du signal de polyadénylation proximal associé à l'exon 9' (appelé signal αe), l'exon 9A9' est utilisé comme exon 3'-terminal dans les somites et le cerveau adulte produisant respectivement les ARNm α7 et B5 (Gaillard *et al.*, 1998; Hardy *et al.*, 1999). Dans les muscles striés adultes, le cœur embryonnaire et en faible proportion dans les somites, l'exon 9A est utilisé en tant qu'exon interne et le signal de polyadénylation αa associé à l'exon 9B est utilisé pour produire l'ARN α2 (Hardy *et al.*, 1991). Dans les tissus non musculaires et les muscles lisses adultes, l'exon 9A9' est sauté, l'exon 8 est épissé à l'exon 9D, il y a alors respectivement production des ARNm O5 et sm (Hardy *et al.*, 1991).

Les ARNm α2 et α7, qui différent uniquement par leur région 3' UTR, sont détectés dans les embryons à partir du stade *neurula* et codent pour une même isoforme musculaire de 284 aa (excepté le dernier acide aminé). Dans les embryons de stade 42, l'ARNm α7 représente 80% des transcrits de l'α-tropomyosine, puis dans les muscles striés adultes seul l'ARNm α2 est exprimé (Hardy *et al.*, 1999). L'ARNm O5 est exprimé en faible quantité dans tous les tissus testés et code pour une isoforme non musculaire de 248 aa (Hardy *et al.*, 1991).

L'utilisation de l'exon composite 9A9' dans les différents ARNm α-TM nécessite un contrôle coordonné de la reconnaissance des sites 3' et 5' d'épissage et du signal de clivage/polyadénylation αe. Ceci en fait un modèle particulièrement intéressant pour étudier,

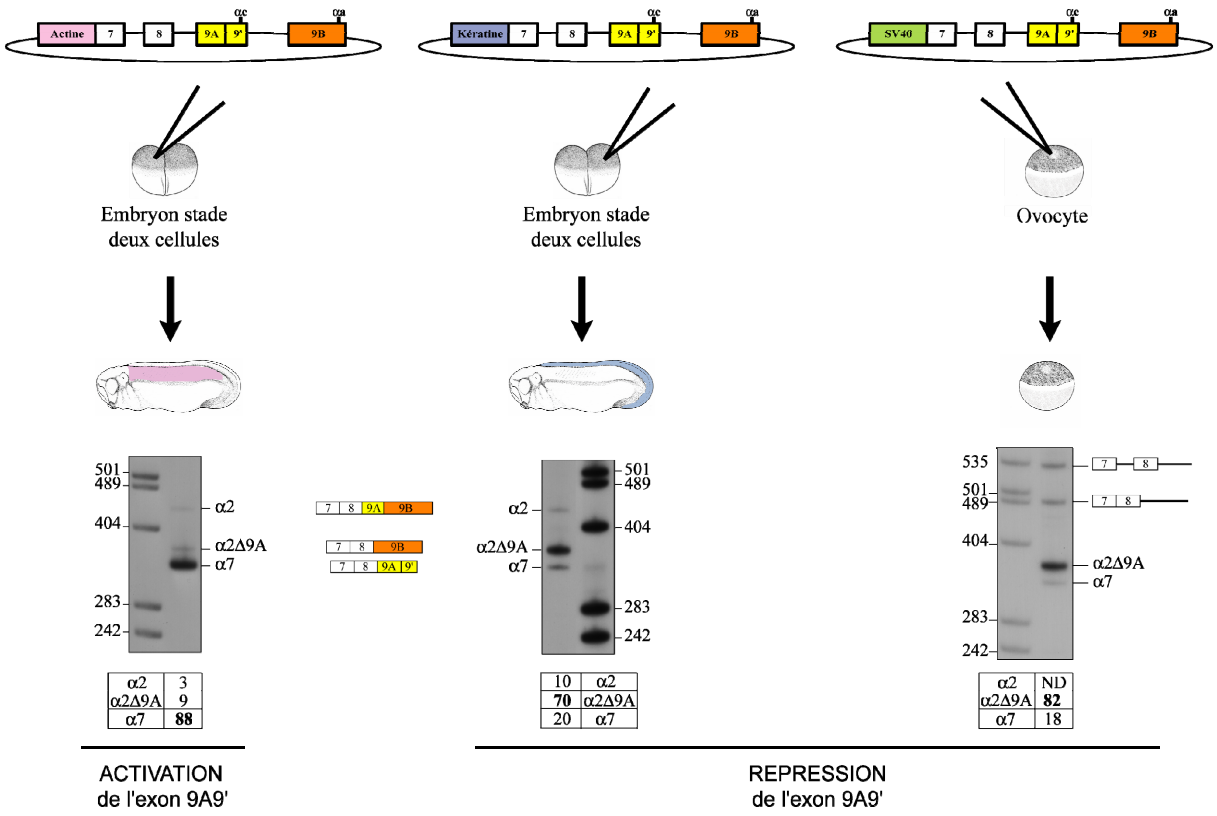


Figure 27- Représentation des trois minigènes utilisés et proportions des ARNm produits. D'après (Le Sommer, 2006). Trois minigènes couvrant la région 7 à 9B placée sous le contrôle de promoteurs spécifiques ont été construits au laboratoire. Le promoteur actine cardiaque cible l'expression dans les somites (représenté en rose dans les embryons de stade 27), le promoteur kératine cible l'expression dans l'épiderme (représenté en bleu sur l'embryon de stade 27) et le promoteur SV40 permet une expression dans les ovocytes. Ces minigènes injectés dans les embryons au stade deux cellules ou dans les ovocytes produisent différents ARNm détectés par RT-PCR radioactive, les produits sont déposés sur gel 4% polyacrylamide puis les différences espèces sont quantifiées à l'aide d'un PhosphorImager. Les ARNpré-m issus de ces minigènes ont un profil d'épissage similaire aux ARNpré-m endogènes : l'exon 9A9' est activé dans les tissus musculaires (somites) et réprimé dans les tissus non musculaires (épiderme, ovocyte).

de manière intégrée, les mécanismes moléculaires régissant la maturation différentielle des exons 3'-terminaux.

2.2. Le xénope

Xenopus laevis (*X. Laevis*) est un amphibien anoure utilisé depuis le début du vingtième siècle pour étudier la biologie cellulaire et le développement des vertébrés. Cette espèce présente plusieurs avantages : ses œufs, aisés à obtenir, sont de taille suffisamment importante (1,5 mm) pour faciliter l'injection de molécules ; ils sont produits en grande quantité, permettant les approches expérimentales de biochimie.

X. laevis est pseudo-tétraploïde, la plupart des gènes sont présents en deux copies qui présentent en moyenne 5 % de divergence nucléotidique, ce qui complique énormément les approches d'analyse génétique. Cependant, *X. laevis* a été largement utilisé pour observer l'expression des gènes, plus particulièrement pour étudier l'utilisation de promoteurs tissu-spécifiques (Koide *et al.*, 2005) et le contrôle de la traduction des ARNm maternels (de Moor et Richter, 2001; Paillard et Osborne, 2003). De plus, un grand nombre d'EST et plusieurs banques d'ADNc sont disponibles pour *X. laevis*.

Toutes ces caractéristiques en font un bon modèle pour étudier l'expression tissulaire des gènes soit en utilisant des minigènes s'exprimant de manière tissu-spécifique, soit en inactivant spécifiquement l'expression d'ARNm. Ces différentes stratégies complémentaires peuvent être utilisées pour caractériser les régulations de l'épissage alternatif *in vivo*.

2.2.1. Utilisation de minigènes pour étudier les régulations de l'épissage

Au laboratoire, nous avons développé plusieurs minigènes s'exprimant de manière tissu-spécifique et contenant la région génomique allant des exons 7 à 9B de l'α-tropomyosine (figure 27). Quand celle-ci est placée sous le contrôle du promoteur actine cardiaque ou kératine, l'expression est spécifiquement ciblée dans les somites ou l'épiderme embryonnaire respectivement. Ces minigènes sont injectés dans les embryons de xénope au stade deux cellules sous la forme d'un ADN plasmidique super-enroulé qui se propage sous forme épisomale. Ces minigènes reproduisent le profil d'épissage de l'ARNpré-m endogène : l'exon 9A9' est utilisé comme exon terminal dans les somites et sauté dans l'épiderme. Un minigène sous le contrôle du promoteur SV40 a également été développé pour pouvoir être exprimé

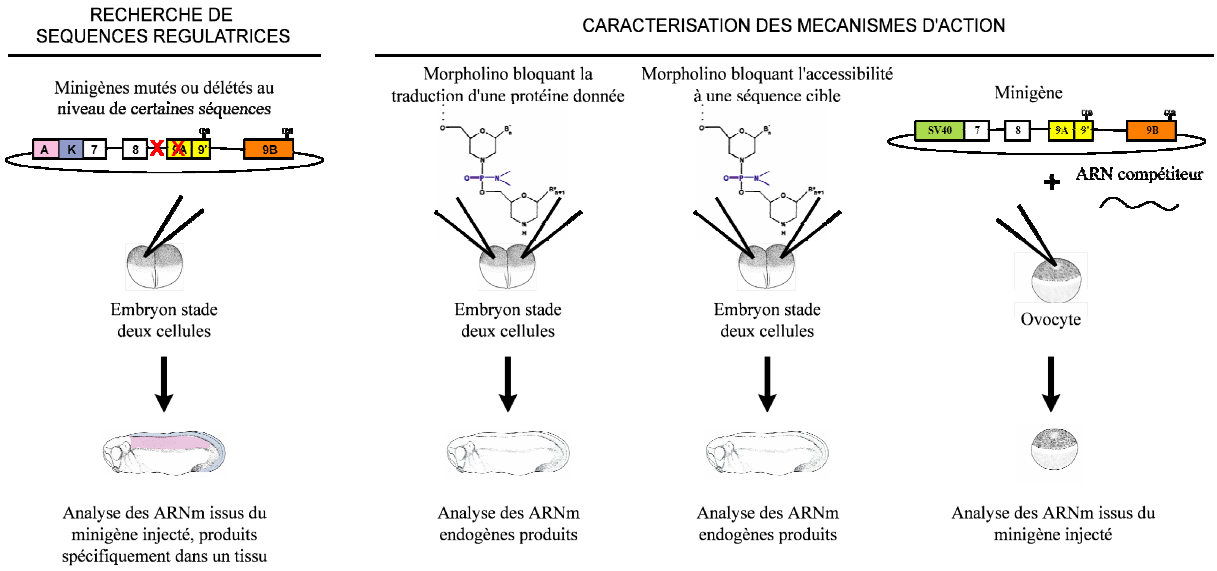


Figure 28- Représentation schématique des différentes stratégies utilisées pour étudier les régulations de la maturation des ARN pré-m *in vivo*. La recherche de séquences régulatrices peut être effectuée en injectant des minigènes placés sous le contrôle de promoteurs tissu-spécifiques (représentés en rose et en bleu) dans des embryons de xénope au stade deux cellules. Des expériences de mutagenèse dirigée (représentées avec des croix rouges) permettent d'observer les effets de la délétion ou de la mutation de séquences spécifiques sur la maturation des ARNpré-m. Les mécanismes d'actions d'une séquence régulatrice peuvent être caractérisés plus précisément en injectant dans les deux blastomères des embryons ou dans l'ovocyte respectivement un morpholino bloquant l'accessibilité d'une séquence cible ou un ARN compétiteur de la séquence d'intérêt. Les mécanismes d'action d'une protéine donnée peuvent être caractérisés en injectant dans les deux blastomères des embryons un morpholino bloquant sa traduction.

dans les ovocytes. Les ARNpré-m issus de ce minigène, injecté dans le noyau des ovocytes, reproduisent majoritairement le schéma de maturation endogène observé dans cette cellule, à savoir le saut de l'exon 9A9' (Hamon *et al.*, 2004).

2.2.2. Utilisation des morpholinos oligonucléotides antisens

Les morpholinos sont des oligonucléotides modifiés qui peuvent s'hybrider de manière stable avec des ARN. Ils possèdent les bases azotées classiques fixées sur des cycles morpholines au lieu du désoxyribose et qui sont liés entre-eux par des groupements phosphorodiamidates au lieu de groupements phosphates. Les hybrides formés avec l'ARN sont stables et ne sont pas ciblés par la RNase H. De ce fait, les morpholinos ne conduisent pas à la dégradation de l'ARN ciblé mais agissent de manière stérique. Les morpholinos oligonucléotides antisens sont résistants aux nucléases et peu toxiques quand ils sont employés *in vivo*. Ainsi ils peuvent être utilisés :

- pour bloquer la traduction, dans ce cas le morpholino choisi ciblera une région située entre 80 nucléotides avant et 30 nucléotides après le codon d'initiation de la traduction de l'ARNm. L'injection d'un morpholino complémentaire de la région -10 à +12 nt autour du codon d'initiation de la traduction de la PTB dans les embryons de xénope a permis de montrer que cette protéine est nécessaire à la répression de l'exon 9A9' dans un contexte minigène (Hamon *et al.*, 2004).

- pour inhiber l'épissage, en ciblant les sites donneurs ou les sites accepteurs d'épissage (Draper *et al.*, 2001).

- pour bloquer stériquement la fixation de microARN ou de protéines régulatrices sur leur séquence ARN cible. Ce blocage de l'accessibilité d'une séquence aux facteurs agissant en *trans*, sans qu'elle ne soit mutée ou supprimée, permet d'avoir accès aux régulations qui ont lieu sur l'ARNpré-m ou l'ARNm endogènes. Cette technique a déjà été utilisée avec succès *in vivo* dans des souris transgéniques pour masquer l'accès à un élément intronique inhibiteur (Hua *et al.*, 2008).

Ces approches, ont été utilisées pour caractériser les mécanismes de maturation de l'exon 9A9' du gène α-TM chez le xénope, et sont récapitulées dans la figure 28.

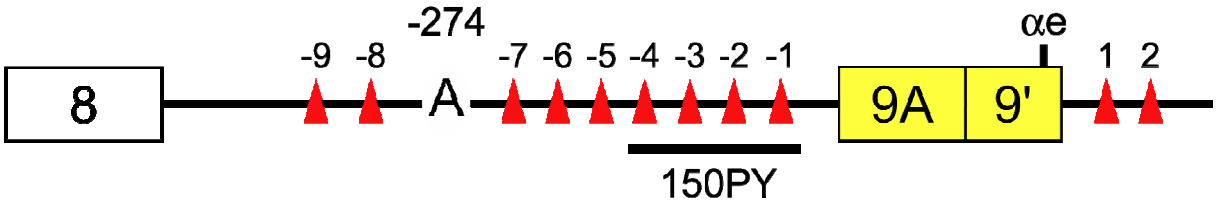


Figure 29- Représentation schématique de l'intron 8-9A9'. Le point de branchement est localisé 274 nucléotides en amont du site 3' d'épissage. Les sites de fixation de la protéine PTB sont représentés en rouge. La séquence inhibitrice 150 PY est indiquée.

3. Résultats précédemment obtenus au laboratoire

L'analyse par mutagenèse des minigènes α-TM a permis d'établir que l'exon 9A9' est défini par un signal de polyadénylation (αe) et des sites 5' et 3' d'épissage faibles. L'intron en amont de l'exon 9A9' se caractérise par un point de branchement très éloigné, situé à 274 nt en amont du dinucléotide AG du site 3' d'épissage et qui est spécifiquement utilisé *in vivo* pour l'épissage de l'exon 9A9' en tant qu'exon 3' terminal (Hamon *et al.*, 2004). Des approches biochimiques et *in vivo* ont conduit à l'identification de la protéine PTB comme facteur clé de la répression de l'exon 9A9' dans les tissus non musculaires de l'embryon. Celle-ci agit en se fixant à des motifs de haute affinité présents entre le point de branchement et le dinucléotide AG du site 3' d'épissage de l'exon 9A9' (Hamon *et al.*, 2004). L'utilisation de minigènes permettant d'analyser indépendamment les réactions d'épissage et de clivage/polyadénylation a permis de préciser que la PTB agit en réprimant la réaction d'épissage et la réaction de clivage/polyadénylation (Le Sommer *et al.*, 2005).

Des études *in vivo* d'inactivation ou de surexpression de PTB ont également montré que sa quantité relative est importante pour assurer une maturation tissu-spécifique de l'extrémité 3' terminale de l'ARNpré-m de l'α-TM. L'inactivation de la PTB par injection de morpholinos oligonucléotides antisens produit un profil d'épissage somitique dans la peau et dans l'endoderme, alors que sa surexpression dans les somites par transgénèse stable entraîne un profil d'épissage spécifique des cellules non musculaires (Le Sommer *et al.*, 2005).

A l'issue de cette étude, deux modèles non exclusifs ont été proposés pour expliquer son activité répressive. Le premier modèle propose que la PTB réprimerait activement l'utilisation de l'exon 9A9' en empêchant sa reconnaissance par des facteurs généraux d'épissage et de clivage/polyadénylation. Le deuxième modèle suggère que la PTB réprimerait indirectement l'exon 9A9' en empêchant la liaison de facteurs activateurs nécessaires à la reconnaissance des sites faibles définissant cet exon. La structure particulière de l'intron 8-9A9' avec un point de branchement particulièrement éloigné du site 3' d'épissage (figure 29) indique que des séquences activatrices pourraient être présentes dans cette région et une approche par mutagenèse de minigène s'exprimant dans les somites a été entreprise pour rechercher de telles séquences. Cette stratégie a permis d'identifier une séquence régulatrice

qui active spécifiquement l'utilisation de l'exon 9A9' en tant qu'exon 3' terminal dans les somites. Cette séquence a été appelée UTE pour *Upstream 3' Terminal exon Enhancer*.

4. Projet de recherche

Mes travaux de thèse ont eu pour objectif de caractériser la fonction moléculaire de la séquence UTE. Ils ont été articulés autour de deux axes complémentaires permettant d'apprécier la fonction de la séquence activatrice *in vivo*, et de définir ses mécanismes moléculaires *in vitro*.

- Le premier axe avait pour objectif d'étudier le fonctionnement de la séquence UTE *in vivo* dans le contexte d'ARNpré-m endogène. De nouvelles approches méthodologiques basées sur la microinjection de morpholinos oligonucléotides antisens ciblant la séquence UTE et les sites d'épissage et de polyadénylation de l'exon 9A9' ont été développées. Les résultats obtenus ont permis de préciser la fonction de la séquence UTE sur l'activation des réactions d'épissage et de clivage/polyadénylation et de définir la relation entre la séquence UTE et le facteur répresseur PTB. Une partie de ces travaux ont fait l'objet d'un article qui est actuellement soumis pour publication à *Journal of Biological Chemistry* et est présenté dans la première partie du chapitre I. La suite de ces travaux qui concerne plus précisément la relation entre l'élément UTE et les signaux qui définissent l'exon 9A9', à savoir le point de branchement, le site 3' d'épissage et le signal de polyadénylation αe, est présentée dans la deuxième partie du premier chapitre.

- Le deuxième axe avait pour objectif d'identifier des facteurs protéiques qui s'assemblent *in vitro* sur la séquence UTE. Nous avons développé des extraits nucléaires à partir d'un système hétérologue dans lequel la régulation de l'ARNpré-m α-TM est conservée et dépendante de la protéine PTB. La mise en place d'une stratégie de purification, choisie suite à une étude bibliographique, et l'utilisation d'extraits nucléaires contenant ou non de la PTB, nous ont permis d'identifier par protéomique des protéines appartenant aux complexes assemblés autour de la séquence UTE. Les différentes étapes ayant permis l'assemblage puis la purification des complexes protéiques recrutés sur l'élément UTE ainsi que les résultats de l'identification des protéines auxquels ces travaux ont aboutis, sont exposés dans le deuxième chapitre des résultats.

RESULTATS

Chapitre 1. Caractérisation fonctionnelle d'un élément intronique nécessaire à la reconnaissance de l'exon 9A9' de l'ARNpré-m de l'α-tropomyosine de xénope en tant qu'exon 3' terminal

1.1. Article 1- La PTB empêche l'activation d'un élément intronique régulateur qui favorise l'utilisation d'un exon 3' terminal composite

Résumé

L'épissage alternatif des exons 3' terminaux joue un rôle majeur dans l'expression des gènes en produisant des ARNm possédant différentes régions 3' non traduites (UTR) qui régulent leur devenir et leur expression. L'ARNpré-m de l'α-tropomyosine de xénope possède un exon composite interne/3' terminal, l'exon 9A9', dont l'expression dans l'embryon est dépendante de l'environnement tissulaire. L'exon 9A9' est réprimé par la PTB dans les tissus non musculaires, alors qu'il est utilisé comme exon interne dans le muscle strié adulte ou comme exon 3' terminal cellules myotomales. Nous avons identifié un élément intronique régulateur, appelé UTE (*Upstream Terminal exon Enhancer*), qui est nécessaire à l'utilisation de l'exon 9A9' en tant qu'exon 3' terminal dans le myotome. Nous avons montré que la PTB prévient l'utilisation de l'UTE dans les cellules non musculaires, au contraire de certaines protéines SR qui favorisent l'inclusion de l'exon 9A9' *via* l'UTE. Le blocage de l'UTE sur l'ARN endogène par l'injection de morpholinos oligonucléotides antisens spécifiques dans les embryons de xénope, inhibe fortement l'utilisation de l'exon 9A9' en tant qu'exon 3' terminal. Cette stratégie qui permet d'étudier sa fonction dans un contexte physiologique, nous a également permis de révéler une nouvelle forme d'épissage. L'ARNm ainsi produit ne possède pas de codon STOP en phase et présente un niveau de stabilité dépendant de la traduction. Ces résultats suggèrent que le mécanisme de nonstop-decay (NSD) pourrait participer au contrôle de la maturation des extrémités 3' des ARNpré-m de l'α-tropomyosine.

The polypyrimidine tract binding protein (PTB) prevents the activity of an intronic regulatory element that promotes the usage of a composite 3' terminal exon.

Vincent Anquetil, Caroline Le Sommer[1], Agnès Méreau, Sandra Hamon, Hubert Lerivray and Serge Hardy*

From CNRS-Université de Rennes1 UMR 6061, Institut de Génétique et Développement de Rennes, IFR 140, Faculté de Médecine, CS 34317, 35043 Rennes cedex, France
[1]present address : Department of Molecular Genetics and Microbiology and Center for RNA Biology, Duke University Medical Center, Durham, North Carolina, 27710

Running title : UTE : an Upstream Terminal exon Enhancer

*Correspondence to Serge Hardy : Tel : + 33 2 23 23 44 66; Fax : + 33 2 23 23 44 78; E mail : serge.hardy@univ-rennes1.fr

Alternative splicing of 3' terminal exons plays a critical role in gene expression by producing mRNA with distinct 3' untranslated regions (UTR) that regulate their fate and their expression. The *Xenopus* α-tropomyosin pre-mRNA possesses a composite internal/3' terminal exon (exon 9A9') that is differentially processed depending on the embryonic tissue. Exon 9A9' is repressed in non muscle tissue by the polypyrimidine tract binding protein (PTB) whereas it is selected as a 3' terminal or internal exon in myotomal cells and adult striated muscles respectively. We report here the identification of an intronic regulatory element, designated UTE (Upstream Terminal exon Enhancer), that is required for the specific usage of exon 9A9' as a 3' terminal exon in the myotome. We demonstrate that PTB prevents the activity of UTE in non muscle cells while a sub-class of SR proteins promotes the selection of exon 9A9' in a UTE dependant way. Morpholino targeted blocking of UTE in the embryo, strongly reduced the inclusion of exon 9A9' as a 3' terminal exon in the endogenous mRNA, demonstrating the function of UTE under physiological circumstances. This strategy allowed us to reveal a splicing pathway that generates a mRNA with no in frame stop codon and whose steady state level is translation dependant. This result suggests that a nonstop decay (NSD) mechanism participates to the strict control of the 3' end processing of the α-tropomyosin pre-mRNA.

Alternative splicing of 3'-terminal exons is a widespread mechanism in metazoans. Global in silico analysis showed that ~20 % of human transcript units contain alternative 3'-terminal exons (Tian *et al.*, 2007), (Yan et Marr, 2005)). This process contributes to the complexity of gene expression by not only expanding the proteome through the generation of protein isoforms with distinct carboxy terminal domains but also producing mRNAs that differ in their 3' UTR. These sequences are pivotal for determining the behaviour of mRNA since they are known to regulate translation, stability and localization. The importance of the 3' UTR in mRNA physiology was recently underscored by the identification of miRNAs that control the translation or half life of mRNAs through interactions with sequences present within the 3' UTR (Bartel, 2009). Mechanistically, the regulation of 3' terminal exon processing is very complex as it requires a coordinated control of the splicing and cleavage/polyadenylation reactions, the later being tightly interconnected to transcription termination (Proudfoot *et al.*, 2002). At present the factors and the mechanisms involved in the coupling between cleavage/polyadenylation and splicing remain largely unknown.

Two classes of alternative 3' terminal exons can be described bases on the organization of the splice sites and the poly(A) site within the exon. The first class comprises exons that are delimited by a 3' splice site and a poly(A) site. Competition for inclusion between two such exons results in

the production of mature mRNA with one of two

possible 3' end termini.. This class is described as skipped exons because the proximal exon of the pair is used as a terminal exon or entirely skipped (Edwalds-Gilbert *et al.*, 1997). The second class includes exons with 3' and 5' splices sites followed closely by a poly(A) site. These exons, designated composite exon (Edwalds-Gilbert *et al.*, 1997), can be used as terminal or internal exon, depending on the cellular context. *In silico* analysis showed that 10% of the trancription units within the human genome possess such composite exons (Yan et Marr, 2005). Bioinformatic studies also indicate that these exons are characterized by suboptimal 5' splice sites and weak poly(A) sites (Tian *et al.*, 2007), (Tian *et al.*, 2005). Such an organization suggests that the control of the 5' splice site or poly(A) site recognition is pivotal in regulating the processing of the composite exon as an internal or terminal exon respectively.

The alternative processing of the immunoglobulin (Ig) M pre-mRNA is the best documented example of a composite exon regulation reviewed in (Peterson, 2007). In matured B cells (plasma cells), selection of the μs poly(A) site present within the composite Cμ4 exon produces mRNAs encoding the secreted Ig proteins while non-matured B-cells activate the use of the 5' splice site associated with the splicing to dowstream exons M1 and M2 generating mRNAs encoding the membrane bound Ig proteins. It has been shown that the μs poly(A) site plays a key role in regulating the alternative processing of Ig M pre-mRNA (Peterson *et al.*, 1991). The fact that mutated Ig pre-mRNA as well as non Ig pre-mRNA modified to contain a poly(A) site that compete with a 5' splice are accurately regulated during B-cell maturation (Peterson, 1994), (Seipelt *et al.*, 1998) strongly indicated that regulatory *cis*-acting sequences are not required to control the usage of μs poly(A) site and suggests that the regulation is based on changes in general polyadenylation factors. Accordingly, it was shown that an artificial increase in the cleavage stimulatory factor 64 (CstF64) in non-maturated B cells was sufficient to increase the use of the proximal poly(A) site (μs) (Takagaki *et al.*, 1996) suggesting that the binding of CstF64 is crucial in the activation of the μs poly(A) site. However, the similar amount of CstF64 observed in established B cell and plasma cell lines that recapitulate the IgM splicing regulation (Edwalds-Gilbert et Milcarek, 1995) indicates that additional factors are required to regulate the binding of CstF64 during B-cell maturation. Furthermore, extensive mutations of sequences surrounding the hexanucleotide AAUAAA of the μs poly(A) site do not impair regulation during B cell differentiation implying they are not contributing to the developmentally regulated recognition of the μs poly(A) site. More importantly these mutation experiments also showed that it was very difficult to inactivate the μs poly(A) site suggesting that additional regulatory sequences could influence recognition by the basic polyadenylation machinery (Peterson *et al.*, 2006).

We have developed the *Xenopus leavis* embryo as a true *in vivo* model to study the basis of tissue specific splicing regulation (reviewed in ref (Mereau *et al.*, 2007). We are using this model and the α-fast tropomyosin (α-TM) pre-mRNA to study the regulation of alternative splicing of 3' terminal exons *in vivo*. This α-TM pre-mRNA contains a composite exon designated exon 9A9' that is differentially processed during *Xenopus* development. In embryonic myotomal cells, specific splicing of exon 9A9' as a 3' terminal exon by the selection of the αe poly(A) site present within this exon produces the α7 mRNA that encodes muscle α-TM (Hardy *et al.*, 1999) (see Figure 1). In embryonic heart and adult striated muscle cells α-TM is generated by the α2 mRNA isoform, which results from the activation of the internal 5' splice site associated with the splicing to the downstream terminal exon 9B (Hardy *et al.*, 1991). α2 and α7 mRNAs differ only by their 3' UTR and the physiological function of this differential processing is moment unknown. In non muscle cells exon 9A9' is skipped and exon 9D is used as a 3' terminal exon (Hardy *et al.*, 1991). Transient transgenesis of minigenes driven by tissue specific promoters that recapitulate the specific use of exon 9A9' in embryonic myotomal cells and its repression in embryonic epidermal can be used to study the function and interplay of *cis* sequences elements within the αTM pre-mRNA (Hamon *et al.*, 2004). Loss or gain of function of *trans*-acting factors and the consequences on the processing of pre-mRNA derived from minigenes or from the endogenous gene can also be apprehended (Hamon *et al.*, 2004), (Le Sommer *et al.*, 2005). Using this model we showed in previous studies

that xPTB, the *Xenopus* orthologue of the mammalian RNA binding protein PTB1, represses the usage of the composite exon 9A9' as a terminal exon in embryonic non muscle cells. This repression requires several high affinity PTB binding sites that are present in the upstream intron between the distant branch site and the 3' splice site. We also demonstrated that this effect is mediated by repressing both the splicing and cleavage/polyadenylation of exon 9A9'. In this study, we identify an intronic enhancer designated UTE, that is required for the specific use of the composite exon 9A9' as a 3' terminal exon in myotomal cells. Injection in the *Xenopus* embryos of antisense morpholino oligonucleotides that target this intronic element demonstrated its physiological relevance in the activation of the composite exon 9A9' in the myotome and also revealed a splicing pathway producing a mRNA with no in frame stop codon. Targeted translation inhibition of this mRNA increases its steady-state level suggesting that a non stop decay mechanism (NSD) participates in the strict control of 3'-end processing of the endogenous α-TM pre-mRNA. We demonstrate *in vivo* that PTB prevents the activity of the UTE in non muscle cells while a subclass of SR proteins promotes the usage of exon 9A9' through UTE.

EXPERIMENTAL PROCEDURES

Plasmid constructs -The α-TM minigenes pBS-SV40.7-9B, pBSActin.7-9B and pBSKeratin.7-9B have been described previously (Hamon *et al.*, 2004). The mutated minigenes were modified by PCR mediated site-directed mutagenesis and cassette substitutions as described in (Hamon *et al.*, 2004). The different oligonucleotides are given in the supplementary data (Table S1). For SR proteins overexpression in oocytes, the open reading frames of *Xenopus laevis* ASF/SF2, SC35, SRp20, 9G8, and human SRp30c, SRp40, SRp55 were PCR using specific upstream primers containing a *Bam*HI (xASF/SF2, hSRp30c, hSRp55) or *Eco*RV (all the others) restriction sites and downstream primers containing a *Not*I restriction site (Table S1). The clone IMAGE p998E0212316Q3 and the plasmids pCGSRp30c, pcGSRp40 and pcGSRp55 (Screaton *et al.*, 1995) were used as templates for the amplification of xASF/SF2, SRp30c, SRp40 and SRp55 respectively. *Xenopus* tailbud RNA was used as template for RT:PCR amplification of xSRp20 and x9G8. Each amplimer was cloned into pGEM-T easy vector (Promega) and verified by DNA sequencing. The *Bam*HI/*Not*I or *Eco*RV/*Not*I fragments were then cloned into pT7TS-V5 plasmid in frame with the carboxyl V5 epitope tag (Hamon *et al.*, 2004).

Xenopus embryos culture and injection - *Xenopus laevis* embryos were obtained by artificial fertilization of eggs from laboratory-reared females. Injections were carried out as previously described (Hamon *et al.*, 2004). For minigene injection, 250 pg of supercoiled DNA were injected in one blastomere at the two-cell stage and the embryos were cultured until stage 26 according to the table of Nieuwkoop and Faber (Nieuwkoop et Faber, 1956). Morpholino oligonucleotides (Mo) obtained from Gene Tools, LLC were injected into both blastomeres of two-cell-stage *Xenopus* embryos. For the translational inhibition of xPTB, xSC35 and muscle αTM, 20 ng per blastomere of xPTB Mo (Hamon *et al.*, 2004) , xSC35 Mo (5'-GAGGCCGACCGTAGCTCATGGAATC-3') and TM Mo (5'-TTCTTGATGGCGTCCATGGCTGCTG-3') were respectively injected. To target UTE, 10 ng per blastomere of UTE-1 Mo (5'-GATGAGGAATCAGACAGCGTGGAAG-3'), and UTE-2 Mo (5'-CATTGGCAAGAAACCACCATCCAGG-3') were injected. A control Mo (5'-CCTCTTACCTCAGTTACAATTTATA-3') was used in all experiments.

Oocyte culture and injection- Injection and culture of oocytes were carried out as previously described (Hamon *et al.*, 2004). 1 ng of pBSV40.7-9B or pSV40.7-9BΔUTE minigenes was co-injected into the nucleus with different amount of 150PY competitor RNA (Hamon *et al.*, 2004). For SR protein overexpression, different amount of *in vitro* transcribed capped mRNA encoding V5-tagged SR proteins were injected in the cytoplasm twelve hours before injection of the wild-type or ΔUTE minigenes into the nucleus.

In vitro transcription- 150PY competitor RNA was transcribed from pGEM vector, linearized at a *Sal*I site, using mMessage mMachine T7 kit (Ambion). The different recombinant V5-tagged SR proteins-capped mRNA were transcribed from pT7TS-V5 vector linearized at *Bam*HI or

*Sma*I, using the mMessage mMachine T7 kit (Ambion).

Antibodies and Western blot analysis - Western blot analysis was carried out as described previously (Hamon *et al.*, 2004). Muscle and non muscle α TM were visualized by the Odyssey infrared imaging system (Li-COR Biosciences) using as primary antibodies the monoclonal anti-tropomyosin antibody clone TM311 (Sigma T2780), and the anti-tropomyosin antibody (Sigma T3651) diluted 1:500 and 1:200 respectively. xSC35, xPTB and V5-tagged SR proteins were detected with the enhanced chemifluorescence (ECF) technique (Amersham Biosciences) using xPTB antibodies (1:600) (Hamon *et al.*, 2004) and monoclonal V5 antibodies (Invitrogen).

Isolation of RNA, Analysis by RT-PCR and RNase protection assay - RNA was isolated from embryos or oocytes as previously described (Le Sommer *et al.*, 2005). RNAs derived from minigenes were analysed by a 3' RACE/PCR assay (Hamon *et al.*, 2004). Briefly, 5µg of RNA was used for the reverse transcription (RT) reaction using an oligo(dT) anchor as a primer and one-sixth of the RT reaction was then used for the PCR amplification that was performed with ^{32}P-end-labeled forward primers specific of each minigene (pBSSV40, pBSActin and pBSKeratin) and a reverse PCR-anchor primer. Endogenous transcripts were detected after RT using random primers and SuperScript II RT (Invitrogen) according to the manufacturer's instructions. One seventh of the RT reaction was then used for the PCR amplification of the different αTM transcripts using the ^{32}P-labeled forward primer derived from exon 8 of αTM gene (5'-GAGTTTGCAGAGAGGACAGTA-3') and reverse primers that specifically hybridized to exon 9' (5'-TGGAAAGGGTACGGAGGTAAGC-3'), 9B (5'-AGGTTCCCAAGAAGAGACTG-3') and 9D (5'-CGGAATTCCTGGCACTCAAGA GCAAG-3'). PCR amplification of EF1α transcript was done with a ^{32}P-labeled EF1αss primer (5'-GAGAGGGAAGCTGCTGAGATGG-3') and EF1αas primer (5'-CCACAGGGAGATGTCA ATGGTA-3'). Following an initial denaturation at 94°C for 3 min, DNA was amplified by 22 cycles (94°C for 30 s, 55°C for 30 s, and 72°C for 45 s). Amplimers were separated on 4% acrylamide gel and quantified by PhosphorImager analysis (Amersham Biosciences). The RNase protection probes used for the detection of endogenous α-TM, EF1-α and MLC1f transcripts and analyzing αTM cleaved and uncleaved transcripts were described previously (Hardy *et al.*, 1999), (Le Sommer *et al.*, 2005). The RNA probe used to follow the splicing of exon 9A was PCR amplified from the pSV40.7-9B construct using the forward (5'-CGGAATTCATTGAAGTCATGTCATTGAG-3') and reverse (5'-CGGGATCCTCCTCTTTTGTGCCACC-3') primers. The RNA probe for analyzing splicing and cleavage of exon 9A9' was PCR amplified using the forward (5'-CGGGATCCTACCTTGCCTTCCCTACTC-3') and reverse (5'-CGGGATCCTCCTCTTTTGTGCCACC-3') primers. The PCR products were digested with BamHI and EcoRI and cloned into pBS vector. The resulting plasmids were linearized with *Xba*I, and transcription was carried out with T7 RNA polymerase (Promega). Different amount of total RNA were analyzed using the RNase protection assay previously described (Hardy *et al.*, 1999). The protected products were quantified with a Phosphorimager (GE Healthcare).

RESULTS

UTE, an intronic regulatory element required for the splicing of exon 9A9' as a 3' terminal exon.- We previously demonstrated that the major branch point used in the myotome to splice exon 9A9' as a 3' terminal exon is localized 274 nucleotides upstream of the AG dinucleotide. Mutation of this branch site (mutBS ; UUCUAAU to UUCUAUU) or of the associated 21 nucleotide long pyrimidine tract resulted in a strong skipping of exon 9A9' in a minigene derived transcript expressed in the myotome (Hamon *et al.*, 2004). Several alternatively spliced exons associated with distal branch sites have been described and in all cases, regulatory elements present between the branch site and the AG dinucleotide have been identified. Accordingly, we showed that several high affinity PTB binding sites distributed between the branch site and the dinucleotide AG were involved in PTB mediated repression of exon 9A9' usage in non muscle cells (Hamon *et al.*, 2004), (Le Sommer *et al.*, 2005). To determine whether activating sequences were also present

in this intronic region mutant minigenes encompassing exon 7 to exon 9B (Fig. 1B) and with a series of 40 nucleotide deletions spanning the region between the distal branch site and the 3' AG border were generated. The wild-type and mutant constructs driven by the cardiac actin promoter that targets expression in the embryonic myotomal cells, were injected into two cell-stage embryos (Fig. 1B). The embryos were then cultured until tailbud stage (stage 26) and extracted mRNA was assayed by 3' RACE/PCR as previously described (Fig. 1C) (Hamon *et al.*, 2004). One of the mutant (Δ-204-165) dramatically changed the splicing pattern from exon 9A9' inclusion to exon 9A9' skipping producing an mRNA designated α2Δ9A, (Fig. 2B, compare lanes 1 and 2). No change of the splicing pattern was observed with minigenes driven by a keratin promoter that target expression in non muscle cells in which exon 9A9' is already strongly skipped (data not shown). An increase of α2 mRNA was also noticed indicating that the deleted intronic sequence is required specifically for the splicing of exon 9A9' as a 3' terminal exon to produce α7 mRNA but not as an internal exon. As intronic regulatory elements are often conserved between species we undertook a comparative analysis of this 40-nucleotide element with the orthologuous mammalian and avian intronic sequences upstream of exon 9A. A 7-nucleotide motif (UGGAUGG) located 178 nucleotides upstream of exon 9A in *Xenopus* was identified (Fig. 2A). This motif is present 94 nucleotides and 153 nucleotides upstream of exon 9A in chicken and human respectively. Mutation of this motif (UGGAUGG to CUUAUAC) resulted in a 2.6 fold decrease in terminal exon 9A9' inclusion (α7 mRNA) relative to the wild-type construct (Fig. 2B, lane 3) and a concomitant increase in internal exon 9A splicing (α2 mRNA). However, the effect on exon 9A9' inclusion was less marked than for the 40-nucleotide deletion suggesting that additional motifs present within the deleted region were involved in the activation of exon 9A9'. Indeed during the construction of the 7-nucleotide motif mutant, one of the clones contained a point mutation fortuitously introduced at position -196 (A→T). This double mutant behaved like the Δ-204-165 mutant (compare lanes 2 and 4). We conclude that an intronic sequence containing at least two motifs is required for the activation of exon 9A9' as a 3' terminal exon in muscle cells. This sequence was designated UTE for Upstream Terminal exon Enhancer.

We previously showed that exon 9A9' is defined by a suboptimal 3' splice site and a weak polyadenylation site. Accordingly Hamon *et al* (17) showed that the reinforcement of the 3' splice site by the introduction of 7 consecutive pyrimidines nearby the AG dinucleotide (mutant 3' cons) or the substitution of the weak αe poly(A) site by a strong synthetic poly(A) site (mutant SPA) resulted in the constitutive usage of exon 9A9' as a 3' terminal exon in non muscle cells. To study whether these mutations activate exon 9A9' splicing through UTE and the distal -274 nt branch site, several additional mutants were constructed in which the 3' cons and SPA mutations were combined with the ΔUTE or branch site (mutBS) mutations. The different mutants were placed under the control of the cardiac actin promoter and injected in two cell stage embryos. As already described in Hamon *et al* (Hamon *et al.*, 2004) the single 3' cons and SPA mutations only weakly changed the splicing pattern (Fig. 2C, compare lanes 2 and 5 with lane 1). A 2.75 fold increase in the production of α2 mRNA was observed with the 3' cons mutant (lane 2) while the amount of α7 mRNA was slightly increased with the SPA mutant. As already described above, the single mutants mutBS and ΔUTE resulted in a strong skipping of exon 9A9' (lanes 3 and 4). When the 3' cons mutant was associated with the deletion of UTE or the mutation of the distal branch site, exon 9A9' was used as a 3' terminal exon or an internal exon to produce α7 and α2 mRNAs respectively (lanes 6 and 8). Hence the reinforcement of the 3' splice site of exon 9A9' is sufficient to define and splice this exon independently of UTE and the distal branch site. The association combination of ΔUTE or mutBS mutations with a strong poly(A) site mainly led to the inclusion of exon 9A9' as a 3' terminal exon (α7 mRNA) (lanes 7 and 9) demonstrating that the reinforcement of the poly(A) site is sufficient to define and splice exon 9A9' as a 3' terminal exon independently of UTE and the distal branch site (compare lanes 3 and 4, 6 and 8 and 7 and 9). Altogether these results show that UTE is required to define exon 9A9' as a 3' terminal exon in the context of the wild-type pre-mRNA. It is also important to notice that the branch site and UTE mutants behave similarly when they are present alone or in combination with the 3' cons or SPA mutations, strongly

suggesting that UTE is required for the efficient usage of the distal branch site.

Blocking the access to UTE in vivo, strongly inhibits the endogenous α7 mRNA production - Mutational analysis of minigene reporters is a powerful approach to identify important *cis*-regulatory sequences. However it is important to keep in mind that deletion or point mutation can also alter splicing by modifying the structure of the RNA or creating binding sites that change the composition of the RNP. Consequently, we decided to study the functional relevance of UTE by targeting it with antisense morpholino oligomers, the rationale being that such oligomers will block the binding of *trans*-acting factors that assemble onto UTE. In *Xenopus* embryo, the α-tropomyosin (TM) gene is expressed more strongly in the myotome than in the non muscle tissues (Hardy *et al.*, 1991). Consequently a whole embryo analysis reflects essentially the myotomal α-TM gene expression pattern. We took advantage of this feature to directly assay the relevance of UTE in the regulation of the endogenous α-TM pre-mRNA. Two contiguous 25 mer antisense morpholino (UTE-1 Mo and UTE-2 Mo) complementary to UTE were designed (Fig. 3A). 10ng/blastomere of UTE-1 Mo or UTE-2 Mo were injected into both blastomeres of two cell stage *Xenopus* embryos. Embryos developed with normal morphology but presented altered myotomal contractions (data not shown). RNA isolated from tailbud stage embryos was assayed by an RNase protection assay using a probe spanning exons 7-8-9A9' that can distinguish between the three possible fates of exon 9A9' : use as a 3' terminal exon (α7 mRNA), as an internal exon (α2 mRNA) and skipping. Two probes directed to EF1-α and MLC1f/3f were included in the assay as controls for RNA recovery and embryo staging (Fig. 3B). As already described, α7 and α2 mRNAs are strongly expressed in *Xenopus* embryos representing ~75 % and 20 % of the α-TM transcripts respectively, whereas the non muscle isoform O5 is hardly detectable and account for less than 5% of the α-TM mRNA (lane 3). Injection of UTE-1 Mo or UTE-2 Mo resulted in a 35 and 15 fold decrease of α7 mRNA respectively whereas a slight increase in the steady-state level of α2 mRNA is observed. A 2 to 3 fold increase in the partially protected fragment corresponding to RNA in which exon 9A9' is skipped was also observed. This result confirms the functional relevance of UTE as a myotomal specific enhancer of exon 9A9' usage as a 3' terminal exon. It also validates our *in vivo* minigene strategy to identify *cis*-regulatory elements.

UTE repression revealed a splicing pathway producing a mRNA with no stop codon in frame - The quantification of the α-TM protected fragments normalized to EF1-α transcripts and uridine content indicated a 4.5 fold reduction of the α-TM transcripts in the UTE-1 Mo-injected embryos (Fig. 3B). Indeed, the decrease in α7 mRNA was not compensated by a symmetric increase in other mRNAs. The partially protected fragment corresponding to mRNA in which exon 9A9' is skipped could potentially be generated from O5 mRNA but also mRNA in which exon 8 is spliced directly to exon 9B (α2Δ9A mRNA). To address this issue we performed RT/PCR assays with a forward primer specific for exon 8 and reverse primers specific for exon 9', 9B and 9D (see Fig. 1). This analysis confirmed that the amount of α7 mRNA was strongly reduced in Mo-injected embryos (Fig. 4A top panel) while the levels of α2 mRNA was slightly increased (middle panel). On the other hand it showed that the steady-state level of O5 mRNA was not modified (bottom panel). With the primer specific for exon 9B, an additional PCR product, present in very low amount in uninjected embryos, was increased 43 fold and 17 fold in UTE-1 and UTE-2 Mo-injected embryos respectively (middle panel, compare lanes 3 and 4 with lane 1 and see the bar diagram). Cloning and sequencing of the corresponding fragment showed that it corresponds to α2Δ9A mRNA in which exon 8 is spliced to the terminal exon 9B. This mRNA is characterized by the absence of an in frame stop codon (supplementary data Figure S2), suggesting that it could be a substrate for degradation *via* nonstop mRNA decay (NSD), a mechanism that requires the translation of the mRNA (Frischmeyer *et al.*, 2002). Such a feature could explain the overall decrease of α-TM mRNAs in Mo-injected embryos. To determine whether α2Δ9A mRNA is effectively a substrate for NSD, we analyzed its steady-state level when its translation was specifically inhibited (Fig. 4B). A morpholino oligomer (TM Mo) directed against the 5' region encompassing the initiation codon of muscle α-TM mRNAs (α2, α7 and α2Δ9A mRNAs share the same initiation codon) was coinjected alone or along with UTE-1 Mo in *Xenopus* embryos. Embryos appeared to be morphologically normal until the

tailbud stage, excepted they presented neither myotomal contractions nor heart beating (data not shown). As expected, TM Mo resulted in a strong decrease of muscle α-TM protein while the amount of non muscle α-TM protein is unaffected (compare lanes 1 and 2), demonstrating that TM Mo efficiently inhibits the translation of muscle α-TM. This inhibition of translation was associated with a 3.8 fold increase in α2Δ9A mRNA indicating that a low amount of α2Δ9A mRNA is physiologically produced during the processing of the α-TM pre-mRNA. UTE-1 Mo also generated a reduction in muscle α-TM protein which is probably the consequence of the dramatic decrease of α7 mRNA and as already described generated a 37 fold increase in α2Δ9A mRNA. Co-injection of UTE-1 Mo and TM Mo resulted in an additional 2 fold stabilization. These results demonstrate that the degradation of α2Δ9A RNA is translation dependant suggesting that NSD contributes to the precise control of α-TM pre-mRNA processing.

UTE promotes the splicing and 3' end cleavage at the proximal αe poly(A) site - Our results so far showed that the decrease in the accumulation of α-TM transcripts within the embryo after UTE inactivation, is the consequence at least, of splicing deregulation leading to the non productive α2Δ9A mRNA. To study whether additional processes were affected by the inactivation of UTE we next monitored the splicing as well as the 3' end cleavage of exon 9A9' at the αe poly(A) site. Cleavage was assayed by RNase protection using a probe that covers exon 9A9' and intronic sequences downstream of the αe poly(A) site (Fig. 5A, top). Because of 3 nucleotides differences between the two *Xenopus* α-TM genes within the sequence of exon 9', the uncleaved and cleaved RNAs distributed in four protected fragments. As expected from the previous RNase protection assay (see Fig. 3B), UTE-Mo mediated inactivation resulted in a strong decrease in cleaved RNA (Fig. 5A, compare lane 3 with lanes 5 and 6). The amount of cleaved RNA produced from the gene copy which is fully complementary to the probe (gene 1) and normalized to EF1-α was reduced 47 fold and 26 fold in UTE-1 Mo and UTE-2 Mo injected embryos respectively, which is consistent with the decrease of the α7 mRNA reported above (see Figure 3B). More importantly a 2 to 3 fold increase in uncleaved RNA derived from gene 1 was also observed. These results show that UTE is required for an efficient 3'-end cleavage of exon 9A9'. We next determined the consequence of UTE inactivation upon the splicing of exon 9A9'. A probe spanning the 3' splice site junction of exon 9A was used in a RNase protection assay (Fig. 5B top). This probe can identify spliced RNA derived from the two α-TM genes while, because of several nucleotides differences within the intronic sequences upstream of exon 9A9', only unspliced RNA derived from gene 1 can be visualized. UTE-1 Mo and UTE-2 Mo produced a 4 fold decrease in spliced RNA (Fig 5B compare lane 3 with lanes 5 and 6). Most of the spliced RNA corresponds likely to the 20 % of α-TM transcripts in which exon 9A is spliced as an internal exon (α2 mRNA) (see Fig. 3B). A 2 fold increase in unspliced RNA was also observed indicating that UTE is required to splice exon 9A9' as a 3' terminal exon.

Our results are consistent with a role of UTE in promoting 3'-end cleavage of exon 9A9' associated, according to exon definition, with the reinforcement of the 3' splice site that lead to the definition of exon 9A9' as a 3' terminal exon and splicing. However, our RNase protection assays can not rule out the possibility that uncleaved and unspliced RNA correspond to different molecules. To determine whether uncleaved and unspliced RNA molecules exist we performed an RNase protection assay with a probe spanning the 3' splice junction and the cleavage site of exon 9A9' (Fig. 5C). Analysis of cleaved and spliced RNA products was again quite complex because of the small differences in the sequences of the two α-TM genes that produced additional partially protected fragments. However it was possible to determine whether a full length protected fragment is produced from gene 1 which is fully complementary to the probe. Such a fragment was observed in non injected embryo and a 4 and 3 fold increase was quantified in UTE-1 Mo and UTE-2 Mo injected embryos respectively (Fig 5C compare lane 3 with lanes 5 and 6).

Altogether these results are in accordance with a requirement of UTE to define exon 9A9' as a 3' terminal exon. In addition, the increase in pre-mRNA in which exon 9A9' is still present indicates that the splicing of exon 8 to exon 9B in myotomal cells is less efficient that the splicing of exon 8 to exon 9A9'. This rise in unprocessed α-TM transcripts participates in the overall decrease in α-TM mRNAs.

xPTB prevents the activity of UTE in non muscle cells - We previously demonstrated that xPTB was a key determinant in repressing exon 9A9' in non muscle cells. It was therefore important to study the relationship between xPTB and UTE. For that purpose we used a strategy previously described (Hamon *et al.*, 2004) combining morpholino-mediated specific knockdown of xPTB and minigene injection into embryos. Morpholino antisense oligomers (xPTB Mo) directed against the 5' region of xPTB mRNA or control non-specific morpholino oligomers (Co Mo) were microinjected into both blastomers of *Xenopus* embryos at the two-cell stage along with wild-type or ΔUTE mutant minigenes driven by the *Xenopus* keratin promoter that targets expression to epidermis (Fig. 6A top). The xPTB Mo specificity was previously demonstrated by rescue experiments with a xPTB mRNA that was not targeted by the morpholino (Hamon *et al.*, 2004). Embryos were fixed at the tailbud stage and extracted RNA and proteins were analyzed by RT-PCR analysis and western blot respectively. As expected xPTB Mo microinjection resulted in a strong decrease of the amount of endogenous xPTB (Fig. 6A top panel, lanes 3 and 6). The knockdown of xPTB resulted, as previously reported, in the derepression of exon 9A9' in the embryos injected with the wild-type minigene (bottom panel, lane 3). In contrast, no change in the splicing pattern was observed in the embryos injected with the ΔUTE mutant minigenes (lane 6) demonstrating that xPTB prevents the activity of UTE in epidermis. We previously showed that the repression of exon 9A9' was similarly controlled in *Xenopus* oocytes and embryonic epidermal cells and that the xPTB mediated repression can be relieved in oocytes by microinjecting RNA competitors containing four high affinity PTB binding sites (150 PY RNAs) (Hamon *et al.*, 2004). As a supplementary evidence of the relationship between UTE and PTB, we therefore tested the effects of such competitors upon the splicing pattern of RNA derived from wild type and ΔUTE mutant minigenes. Minigenes driven by the viral SV40 promoter that is functional in *Xenopus* oocytes were injected into oocyte nuclei along with increasing amount of competitors (Figure 6B). As previously described the competitors caused a complete derepression of exon 9A9' in the context of the wild type minigene pre-mRNA (compare lanes 1 and 5). However, only a partial derepression was observed with the mutant (compare lanes 2 and 6). These results demonstrate that xPTB represses exon 9A9' usage in oocytes, in part, by antagonizing UTE activity as happens in embryonic epidermal cells. Altogether several important conclusions can be drawn from these experiments: i) xPTB prevents the activity of UTE in non muscle cells; ii) UTE is functional in non muscle cells when the PTB repression is relieved iii) UTE may interact with ubiquitously expressed factors when the amount of xPTB is low.

The SR protein SC35 activates exon 9A9' usage through UTE - In view of the previous data and the established role of the ubiquitously expressed SR proteins in splicing regulation we decided to study whether this class of RNA binding protein might participate in exon 9A9' activation through UTE. We first used the ESE finder program (Cartegni *et al.*, 2003) to examine the presence of potential SR proteins binding sites within UTE. Using the threshold values considered as being significant within the context of an ESE, ESE finder predicted several binding sites for SC35, SRp40, and SRp55 (supplementary Figure S1). The density of predicted binding sites was higher for SC35 and more importantly five and two SC35 binding sites out of nine were absent in ΔUTE and mutUTE mutants respectively. Since UTE is functional in *Xenopus* oocytes when repression exerted by xPTB is relieved, we decided to test whether SC35 activates the splicing of exon 9A9' as a 3' terminal exon in these cells. 4 ng, 8 ng or 16 ng of *in vitro* transcribed mRNA encoding xSC35 tagged with a V5 epitope were injected into the cytoplasm of *Xenopus* oocytes and 24 hours later the wild type minigene was injected into the nucleus. Overexpression of xSC35 was obtained in oocytes as determined by western blot with V5 antibody (Figure 7A top panel). As described above, in the absence of xSC35 mRNA injection, processing of the wild type minigene produced predominantly exon skipping (α2Δ9A mRNA) and a low level of exon 9A9' usage (α7 mRNA) (Figure 7A bottom panel, lane 1). When xSC35 was overexpressed exon 9A9' usage was increased 4.6 fold (compare lane 1 with lane 3). To analyze whether this activation was dependant on UTE, the effects of xSC35 overexpression was also assayed using the ΔUTE minigene (Figure 7B). Whereas the wild-type minigene showed a 7 fold activation of exon 9A9' usage with xSC35 RNA minjection

(compare lane 1 with lanes 3) no activation was observed with the ΔUTE construct. We therefore concluded that xSC35 activates the splicing of exon 9A9' and that this activation is UTE dependant. Additionally, the dose dependent effects of xSC35 overexpression upon exon 9A9' splicing strongly suggest a direct role.

In this study we demonstrated that the repression activity of xPTB upon exon 9A9' processing was mediated in part by UTE inactivation. This effect is probably direct as UTE contains two high affinity PTB binding sites and recombinant PTB cross-linked to UTE (data not shown). Therefore a possibility is that the use of exon 9A9' as a 3' terminal exon is regulated by a mechanism based on antagonistic binding of xSC35 and xPTB to UTE.

A sub class of SR proteins may activate exon 9A9' processing in a UTE dependent way.- To test the physiological relevance of xSC35 in the activation of exon 9A9' usage as a 3' terminal exon in embryonic myotomal cells, we examined the effects of xSC35 depletion upon the processing of the endogenous αTM pre-mRNA or of the pre-mRNA derived from the wild type minigene whose expression is targeted to myotomal cells. Morpholino mediated translation inhibition of xSC35 within the embryo failed to modify the splicing pattern of both endogenous and minigene αTM mRNA (data not shown). We therefore concluded that xSC35 inactivation within the embryo was not sufficient to prevent exon 9A9' splicing as a 3' terminal exon.

This result was not so surprising since it was reported that individual SR protein may have redundant function *in vivo*. To test whether additional SR proteins may activate α7 mRNA production through UTE we used the *Xenopus* oocyte model described in the previous paragraph. *In vitro* transcribed mRNA encoding homologous or orthologous SR proteins tagged with a V5 epitope were injected in oocyte cytoplasm and 24 hours later minigenes were injected into the nucleus. The accumulation of the V5 tagged protein in the nucleus and cytoplasm was checked by western-blot (Fig. 8A) and the splicing pattern was assayed by RT/PCR as already described (Fig. 8B). One of the SR protein tested, SRp30c did not activate exon 9A9' usage (compare lane 1 and 3) indicating that the switch to α7 mRNA is not the consequence of a general activation of exon 9A9' splicing by SR proteins. The remaining SR proteins activated exon 9A9' usage and can be divided into two classes. The first class corresponds to SR protein that activate exon 9A9' usage through UTE and comprises SRp40 and SRp55 (compares lanes 5 and 6, 7 and 8, 11 and 12). The second class corresponds to SR protein that activate exon 9A9' usage independently of USE and comprises xASF/SF2, x9G8 and xSRp20 (compares lanes 9 and 10, 11 and 12, 13 and 14). These results indicate that a subclass of SR proteins comprising SRp40, SRp55 and xSC35 is able to activate exon 9A9' usage through UTE suggesting that an activator complex comprising SR proteins is assembled onto UTE.

DISCUSSION

PTB in repression of the composite exon 9A9' - We previously showed that PTB was involved in the repression of exon 9A9' usage as a 3' terminal exon in non muscle cells, by inhibiting both the splicing and cleavage/polyadenylation reactions (Le Sommer *et al.*, 2005). Our present data demonstrate that PTB is acting in part by preventing the activity of the intronic element UTE required to define the weak exon 9A9' as a 3' terminal exon. Several high affinity PTB binding sites are present within UTE and UV-cross linking demonstrated that PTB bind to this element. This result suggests that PTB may antagonize the binding of activator factors onto UTE as previously demonstrated for the α-actinin and cardiac troponin pre-mRNA (24,25). The observation that in absence of UTE a 2 fold increase in exon 9A9' usage was still observed when PTB was down regulated strongly suggests that in addition to precluding the assembly of an activator complex onto UTE, PTB directly inhibits the 3' splice site and αe poly(A) site selection by the basal splicing and cleavage/polyadenylation machineries. Since numerous high affinity PTB binding sites are present between the -274 distal branch site and exon 9A9' it is likely that PTB inhibits the splicing of exon 9A9' by directly preventing the binding of U2AF65 as described for the mutually exclusive exon 6B of the β-tropomyosin pre-mRNA (Sauliere *et al.*, 2006). It is also conceivable that PTB blocks the transition from an exon 9A9' defined complex to a functional spliceosome as demonstrated for the regulation of the alternative exon N1 of the c-src pre-mRNA (Sharma *et al.*, 2008). Regarding the cleavage/polyadenylation reaction, it could be envisioned that PTB directly inhibits this process

by competing with CstF64 binding to the downstream uridine-rich region of αe poly(A), since it was demonstrated that PTB modulates the efficiency of polyadenylation by such a competition (Castelo-Branco *et al.*, 2004). However, the finding that a chimeric reporter β-globin minigene with an αe poly(A) site is not responsive to PTB repression is not in agreement with this hypothesis (C. Le Sommer and S. Hardy unpublished data). Moreover, overexpression of CstF64 in *Xenopus* oocytes failed to activate exon 9A9' usage indicating that an increase in CstF64 is not sufficient to relieve the repression exerted on the αe poly(A) site (C. Le Sommer and S. Hardy unpublished data). The morpholino mediated inhibition of UTE in myotomal cells resulted in a strong decrease in exon 9A9' usage associated with a direct splicing of exon 8 to exon 9B (α2Δ9A) RNA. This result differs from the splicing of exon 8 to exon 9D (O5 RNA) that occurs when PTB was overexpressed in myotomal cells (Le Sommer *et al.*, 2005). This discrepancy indicates that in addition to inhibiting of exon 9A9' definition PTB is either involved in the repression of exon 9B so that splicing of the distal 3'-terminal exon 9D is favored or usage of exon 9D is activated by PTB.

Our data demonstrate that PTB mediated repression of the proximal alternative 3' exon 9A9', is an important mechanism to ensure its complete skipping during 3' end processing of the αTM pre-mRNA. This observation led us to predict that PTB might be a major actor in the repression of proximal 3' terminal exons in general. Interestingly, the study of the microRNA miR-124 down regulation of PTB during neuronal differentiation uncovered several pre-mRNAs whose neural specific proximal 3' terminal exon is repressed in non neuronal cells by PTB (Makeyev *et al.*, 2007). The switch to nPTB during neuronal differentiation alleviates this repression allowing the use of these proximal 3' terminal exons. It would be of interest to study whether the activation of the proximal exon is dependant on intronic regulatory sequences and whether PTB prevents the assembly of an activating complex onto these sequences.

UTE function in the activation of a 3' terminal exon -The composite internal/3' terminal exon 9A9' is defined by suboptimal 3' and 5' splices sites and a weak poly(A) site (Hamon *et al.*, 2004). Informatic studies showed that such suboptimal signals are a general feature of this class of 3' terminal exon and it was proposed that it allows a dynamic competition between splicing and polyadenylation (Tian *et al.*, 2007). The binding of CstF64 to the downstream uridine-rich element of a poly(A) site is a direct way to regulate a poly(A) site usage. For example in the drosophila female germline, the RNA binding protein sex-lethal (SXL) competes with CstF64 for binding to the uridine-rich element of the enhancer of rudimentary pre-mRNA and represses poyladenylation at the proximal poly(A) site (Gawande *et al.*, 2006). CstF64 binding is also pivotal in the choice of the proximal μS poly(A) site of the immunoglobulin pre-mRNA during B-cells differentiation (Takagaki *et al.*, 1996). However, the observation that the amounts of CstF64 are similar in established B cell and plasma cells lines, which recapitulate this splicing regulation (Edwalds-Gilbert et Milcarek, 1995), suggested that additional factor and auxiliary *cis*-sequences participate to the regulation of CstF64 binding (Peterson *et al.*, 2006). The intronic regulatory element (UTE), we identified upstream of exon 9A9', defines a new class of intronic splicing enhancer (ISE) that is specifically involved in the splicing activation of a composite exon as a 3' terminal exon. The molecular mechanisms by which this element activates the splicing of exon 9A9' as a terminal exon remain to be determined. This task is arduous using *in vivo* approaches because of the functional coupling between the cleavage/polyadenylation and splicing reactions (Nesic et Maquat, 1994) and the exon definition in which activation of the 3' splice site favors the recognition of the poly(A) site and conversely (Niwa et Berget, 1991). *In vitro* systems should be appropriate to study both reactions independently however all approaches to develop such assays were unsuccessful (A. Méreau, unpublished data) suggesting that coupling of both reactions is required.

Our present data, showing that UTE and the -274 nt distal branch site are functionally linked, suggest a model in which UTE promotes the binding of U2snRNP to the distal branch site which in turn might stabilize the binding of CPSF by a direct interaction (Kyburz *et al.*, 2006). It is also possible that UTE controls the binding of U2AF65 to the polypyrimidine tract which in turn recruits the cleavage factors CFIm through their respective RS domains (Millevoi *et al.*, 2006). However the fact that the mutation of UTE or its inhibition *in vivo* result in a weak but

reproducible increase in the splicing of exon 9A as an internal exon strongly suggests that its primary function is to promote the assembly of an activating complex that increases the recognition of the weak αe poly(A) site. Then, in accordance with exon definition this complex may stabilize proteins involved in the recognition of the 3' splice site such as U2AF65 and U2AF35 leading to the definition of exon 9A9' as a 3' terminal exon. Overexpressions studies in the oocyte demonstrated that a subclass of SR proteins activate exon 9A9' usage through UTE. Only a few studies have reported a relationship between SR proteins and the cleavage/polyadenylation machinery. It was shown that SRp20 participates to the definition of the proximal terminal exon of the CT/CGRP pre-mRNA through binding to a downstream intronic enhancer (Lou *et al.*, 1998). More recently it was shown that the binding of SR proteins to the negative regulator of splicing (NRS) of the Rous sarcoma virus (RSV) stimulates polyadenylation (Maciolek et McNally, 2007). Interestingly, this effect was dependant of an additional site that bring the NRS and the poly(A) site closer together, indicating that the distance between the bound SR protein and the poly(A) site is important. The authors anticipated that SR-protein binding sites present upstream of a suboptimal or alternative poly(A) sites might regulate their recognition. This is in accordance with the position of UTE within the α-TM pre-mRNA since it is only 300 nt upstream of αe poly(A) site. At present it is not known how SR proteins might activate the cleavage/polyadenylation processes. It is known that SR protein can interact with U1snRNP (Kohtz *et al.*, 1994) which in turn associates to the cleavage factor CFI (Awasthi et Alwine, 2003). The SRm160 splicing coactivator is also a good candidate to mediate SR protein activity upon the cleavage/polyadenylation process. This factor that promotes transcript 3' end cleavage (McCracken *et al.*, 2002) physically interacts with SR family proteins *in vitro* through their respective RS domains. Moreover, functional inactivation of SRm160 and individual SR protein in *C. elegans* resulted in specific phenotypes demonstrating functional interactions between these factors (Longman *et al.*, 2001). To our knowledge UTE is the first intronic enhancer present in the intron upstream of a regulated composite exon. This element defines therefore, a new class of intronic enhancers and it will be interesting to study whether such elements are present upstream of other regulated composite exons.

NSD and alternative splicing of 3' terminal exons - In this study the morpholino-mediated inhibition of UTE revealed that an endogenous α-TM mRNA (α2Δ9A) corresponding to the direct splicing of exon 8 to exon 9B was produced at a low level in *Xenopus* embryos. This result was not surprising since the generation of α2Δ9A RNA from α-TM minigenes showed that there is no intrinsic blocage of the splicing of both exons. The resulting endogenous mRNA does not possess an in frame stop codon making it a potential substrate for Nonstop Decay (NSD). Stabilization of this mRNA after translation inhibition is in agreement with such a mechanism. There appears therefore to be no strict mechanism to preclude splicing of exon 8 to exon 9B, therefore NSD participates in the strict control of the 3' end processing of the endogenous α-TM pre-mRNA by degrading a non productive mRNA species. NSD was characterized in yeast (van Hoof *et al.*, 2002) and demonstrated to be present in mammalian cells (Frischmeyer *et al.*, 2002). It was proposed that use of 3' end processing signals present in the coding region may be the basis of NSD selection. Our data add a new potential role for NSD, that of monitoring the accuracy of alternative 3' end processing. Our data also showed that such a mechanism is active within the tissue of a living organism and contributes to the regulation of a complex gene regulation process. It is interesting to note that this process in not conserved through vertebrates phylogeny since in rat, splicing of exon 8 to exon 9B produces a mRNA that encode a tropomyosin isoform (TMBr2) specifically expressed in the brain (Lees-Miller *et al.*, 1990). This observation is in agreement with the fact that distinct alternative mRNA splicing participates in the speciation of organisms. *In silico* analyses showed that 20 % of the composite exons do not contain stop codons in frame with upstream exons (Yan et Marr, 2005) indicating that these mRNAs are potential NSD substrates. Hence NSD may participate in gene expression regulation in the same way as NMD (McGlincy et Smith, 2008) by down regulating genes in specific tissues through the splicing of terminal exon devoided of in frame stop codon.

Studying in vivo the mechanisms of alternative splicing and polyadenylation - Since the pioneering work of Kole (Kole, 1998) showing that antisense oligonucleotides targeting the 3' or 5' splice sites can modify the splicing pathway of alternative exons, antisense oligonucleotides have become a promising therapeutic tool to redirect the splicing of aberrant mRNA by targeting splices sites and *cis*-regulatory elements present within exons (Sazani et Kole, 2003). Such a strategy based on the skipping of the defective exons within the dystrophin pre-mRNA is currently in development as a treatment for Duschene muscular dystrophy (Sazani et Kole, 2003). The feasibility of modulating the splicing pattern of the fibroblast growth factor receptor 1, by targeting intronic regulatory elements was also reported (Bruno *et al.*, 2004). The present data demonstrate for the first time that the splicing of a 3' terminal exon can also be redirected by antisense oligonucleotides targeted at a *cis*-regulatory element. In addition to the potential of modifying splicing, antisense oligonucleotides might also be a powerful tool to study the complex interplay of regulatory sequences within an endogenous pre-mRNA *in vivo*. Indeed, simultaneously targeting regulatory elements upstream and downstream of the α-exon of the fibroblast growth factor receptor 1 demonstrated that these sequences act through a shared mechanism (Bruno *et al.*, 2004). In order to understand how UTE activates the usage of exon 9A9', simultaneous injection of antisense oligonucleotides targeting UTE and the 3' splice site or the αe poly(A) site of exon 9A9' may provide useful information on the mechanisms involved and of their relationship with the basal splicing and cleavage/polyadenylation machineries. This approach, associated with a morpholino mediated knockdown of the 5' → 3' exonuclease Xrn2 or the core component EXOSC9 of the exosome, should give access to intermediate RNA species. It would also be possible to identify additional *cis*-regulatory elements present within the endogenous α-tropomyosin pre-mRNA, using a systematic antisense oligonucleotides tiling strategy similar to that was successfully used to identify intronic *cis*-regulatory elements controlling the splicing of SMN2 exon 7 (Hua *et al.*, 2008). The *Xenopus* embryo model is fully adapted for a true *in vivo* approach and should give new insights into the difficult task of deciphering the mechanisms that control the differential processing of 3' terminal exons.

Acknowledgments - We thank Joelle Marie for the recombinant PTB protein and Adrian Krainer for the plasmids pCGSRp30c, pCGSRp40 and pCGSRp55. We thank Beverley Osborne for critically reading the manuscript and all the members of the lab for helpful comments and suggestions. This work was supported by grants to SH from the Association Française contre les myopathies (AFM). VA was supported by doctoral fellowhips from the French government and the Association pour la recherche contre le cancer (ARC).

REFERENCES

1. Tian, B., Pan, Z., and Lee, J. Y. (2007) *Genome Res* 17, 156-165.
2. Yan, J., and Marr, T. G. (2005) *Genome Res* 15, 369-375.
3. Bartel, D. P. (2009) *Cell* 136, 215-233.
4. Proudfoot, N. J., Furger, A., and Dye, M. J. (2002) *Cell* 108, 501-512.
5. Edwalds-Gilbert, G., Veraldi, K. L., and Milcarek, C. (1997) *Nucleic Acids Res.* 25, 2547-2561.
6. Tian, B., Hu, J., Zhang, H., and Lutz, C. S. (2005) *Nucleic Acids Res* 33, 201-212.
7. Peterson, M. L. (2007) *Immunol Res* 37, 33-46.
8. Peterson, M. L., Gimmi, E. R., and Perry, R. P. (1991) *Mol Cell Biol* 11, 2324-2327.
9. Peterson, M. L. (1994) *Mol Cell Biol* 14, 7891-7898.
10. Seipelt, R. L., Spear, B. T., Snow, E. C., and Peterson, M. L. (1998) *Mol Cell Biol* 18, 1042-1048.
11. Takagaki, Y., Seipelt, R. L., Peterson, M. L., and Manley, J. L. (1996) *Cell* 87, 941-952.
12. Edwalds-Gilbert, G., and Milcarek, C. (1995) *Mol Cell Biol* 15, 6420-6429.
13. Peterson, M. L., Bingham, G. L., and Cowan, C. (2006) *Mol Cell Biol* 26, 6762-6771.
14. Mereau, A., Le Sommer, C., Lerivray, H., Lesimple, M., and Hardy, S. (2007) *Biol Cell* 99, 55-65.
15. Hardy, S., Hamon, S., Cooper, B., Mohun, T., and Thiébaud, P. (1999) *Mech Dev* 87, 199-202.
16. Hardy, S., Fiszman, M. Y., Osborne, H. B., and Thiebaud, P. (1991) *Eur J Biochem* 202, 431-440.
17. Hamon, S., Le Sommer, C., Mereau, A., Allo, M. R., and Hardy, S. (2004) *J Biol Chem* 279, 22166-22175.
18. Le Sommer, C., Lesimple, M., Mereau, A., Menoret, S., Allo, M. R., and Hardy, S. (2005) *Mol Cell Biol* 25, 9595-9607.
19. Screaton, G. R., Caceres, J. F., Mayeda, A., Bell, M. V., Plebanski, M., Jackson, D. G., Bell, J. I., and Krainer, A. R. (1995) *EMBO J* 14, 4336-4349.
20. Nieuwkoop, P., and Faber, J. (1956) *Daudin, North-Holland, Amsterdam, The Netherlands*.
21. Sauliere, J., Sureau, A., Expert-Bezancon, A., and Marie, J. (2006) *Mol Cell Biol* 26, 8755-8769.
22. Frischmeyer, P. A., van Hoof, A., O'Donnell, K., Guerrerio, A. L., Parker, R., and Dietz, H. C. (2002) *Science* 295, 2258-2261.
23. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. (2003) *Nucleic Acids Res* 31, 3568-3571.
24. Charlet, B. N., Logan, P., Singh, G., and Cooper, T. A. (2002) *Mol Cell* 9, 649-658.
25. Gromak, N., Matlin, A. J., Cooper, T. A., and Smith, C. W. (2003) *RNA* 9, 443-456.
26. Sharma, S., Kohlstaedt, L. A., Damianov, A., Rio, D. C., and Black, D. L. (2008) *Nat Struct Mol Biol* 15, 183-191.
27. Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A., and Proudfoot, N. (2004) *Mol Cell Biol* 24, 4174-4183.
28. Makeyev, E. V., Zhang, J., Carrasco, M. A., and Maniatis, T. (2007) *Mol Cell* 27, 435-448.
29. Gawande, B., Robida, M. D., Rahn, A., and Singh, R. (2006) *EMBO J* 25, 1263-1272.
30. Nesic, D., and Maquat, L. E. (1994) *Genes Dev.* 8, 363-375.
31. Niwa, M., and Berget, S. M. (1991) *Genes Dev* 5, 2086-2095.

32. Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006) *Mol Cell* 23, 195-205.
33. Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S. Z., Keller, W., Antoniou, M., and Vagner, S. (2006) *EMBO J* 25, 4854-4864.
34. Lou, H., Neugebauer, K. M., Gagel, R. F., and Berget, S. M. (1998) *Mol. Cell. Biol.* 18, 4977-4985.
35. Maciolek, N. L., and McNally, M. T. (2007) *J Virol* 81, 11208-11217.
36. Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., and Manley, J. L. (1994) *Nature* 368, 119-124.
37. Awasthi, S., and Alwine, J. C. (2003) *RNA* 9, 1400-1409.
38. McCracken, S., Lambermon, M., and Blencowe, B. J. (2002) *Mol Cell Biol* 22, 148-160.
39. Longman, D., McGarvey, T., McCracken, S., Johnstone, I. L., Blencowe, B. J., and Caceres, J. F. (2001) *Curr Biol* 11, 1923-1933.
40. van Hoof, A., Frischmeyer, P. A., Dietz, H. C., and Parker, R. (2002) *Science* 295, 2262-2264.
41. Lees-Miller, J. P., Goodwin, L. O., and Helfman, D. M. (1990) *Mol Cell Biol* 10, 1729-1742.
42. McGlincy, N. J., and Smith, C. W. (2008) *Trends Biochem Sci* 33, 385-393.
43. Kole, R. (1998) in *Applied Antisense Oligonucleotides Technology* (Stein, C. A., and Krieg, A. M., eds), pp. 451-469, Wiley-Liss, Inc.
44. Sazani, P., and Kole, R. (2003) *J Clin Invest* 112, 481-486.
45. Bruno, I. G., Jin, W., and Cote, G. J. (2004) *Hum Mol Genet* 13, 2409-2420.
46. Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F., and Krainer, A. R. (2008) *Am J Hum Genet* 82, 834-848.

FIGURES

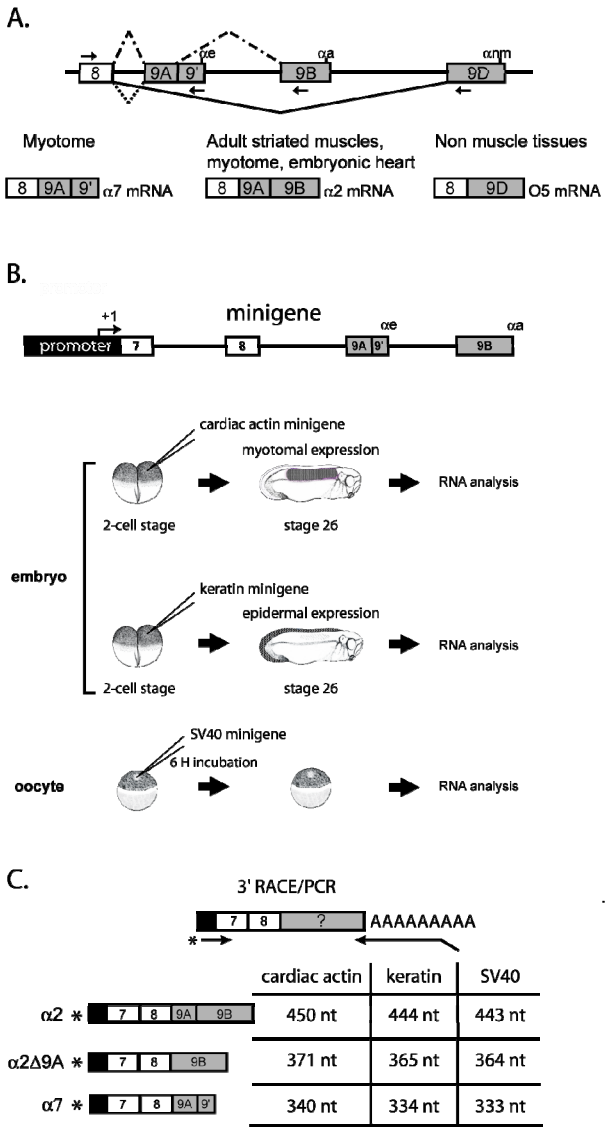


Figure 1. Differential processing of the 3'-end of the Xenopus α-TM pre-mRNA. A. Schematic diagram of the 3'-terminal region of the *Xenopus* α-TM pre-mRNA and of the different tissue-specific RNA processing patterns. Exons and introns are represented by boxes and horizontal lines respectively. αe, αa and αnm represent the poly(A) sites present within exons 9A9', 9B and 9D respectively. Arrows represent the forward and reverse primers used to specifically RT/PCR amplify the different RNA isoforms. The different RNA products in myotome, embryonic heart, adult striated muscle and non muscle tissues are indicated. **B.** Diagrammed representation of the α-tropomyosin minigene and of the different *in vivo* expression assays using the *Xenopus* embryo and oocyte **C.** Schematic representation of the 3' RACE/PCR assay. RNA was reverse transcribed with an oligod(T) anchor primer. cDNA was further amplified with a reverse anchor primer and a radiolabeled forward primer that hybridize at the junction of the promoter and exon 7. Primers are indicated by arrows and asterisk marks the radiolabeling. The structure and size of all splicing products produced by the different minigenes are given.

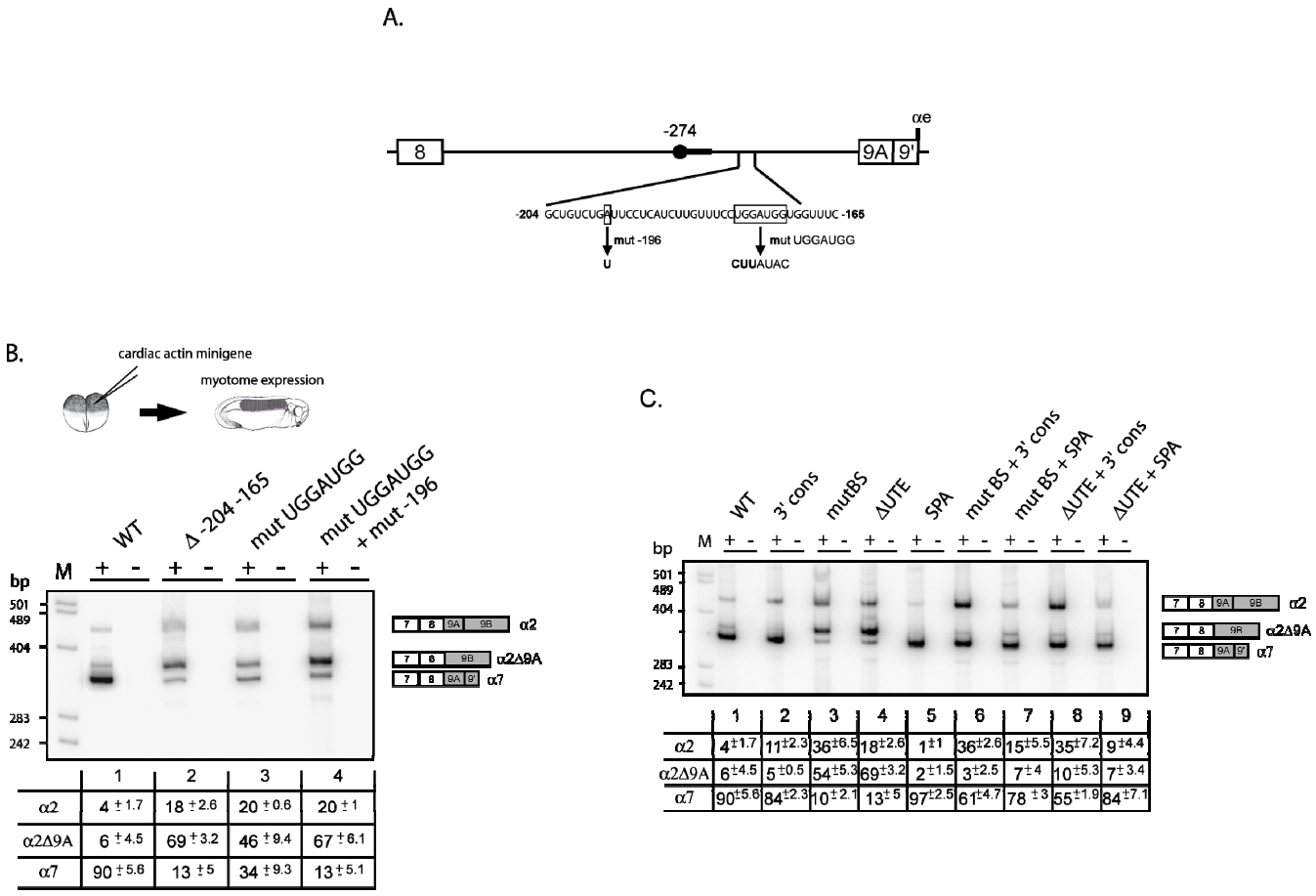


Figure 2. An intronic sequence upstream of exon 9A9' contributes to its splicing as a 3' terminal exon in muscle cells. A. Schematic map of the intronic region upstream of exon 9A9'. The distant -274 branch site and its associated polypyrimidine tract are represented by a black circle and a black rectangle, respectively. The intronic sequence from -204 to -165 upstream of exon 9A9' is written. The generated mutations are shown below the wild-type sequence. **B.** The wild-type (WT) and the indicated mutant minigenes under the control of the cardiac actin promoter were injected into *Xenopus* embryos. The mRNA splicing pattern was assayed by 3' RACE-PCR using a ^{32}P oligonucleotide specific of the cardiac actin minigene constructs in the presence (+) or absence (-) of reverse transcriptase. The structure of the amplified fragments is indicated on the right. Quantification was performed on a Phosphorimager to yield the proportion of each mRNA. The percentage of each mRNA is given in the table below a representative experiment as mean +/- SD (n = 20 for the wild type construct and n = 3 for the mutants) **C.** The distant -274 branch site and UTE are functionally coupled. The Δ -204-165 mutation (ΔUTE) and the mutation of the distant -274 nt branch site (mutBS) were associated to the 3'cons mutant in which seven consecutive pyrimidine nearby the dinucleotide AG were introduced or to the SPA mutant in which a strong synthetic poly(A) site was substituted to the αe poly(A) site. The wild-type (WT) and the indicated mutant minigenes under the control of the cardiac actin promoter were injected into *Xenopus* embryos. Samples were processed as for panel B. The percentage of each mRNA is given in the table below a representative experiment as mean +/- SD (n = 20 for the wild type construct and n = 3 for the mutants).

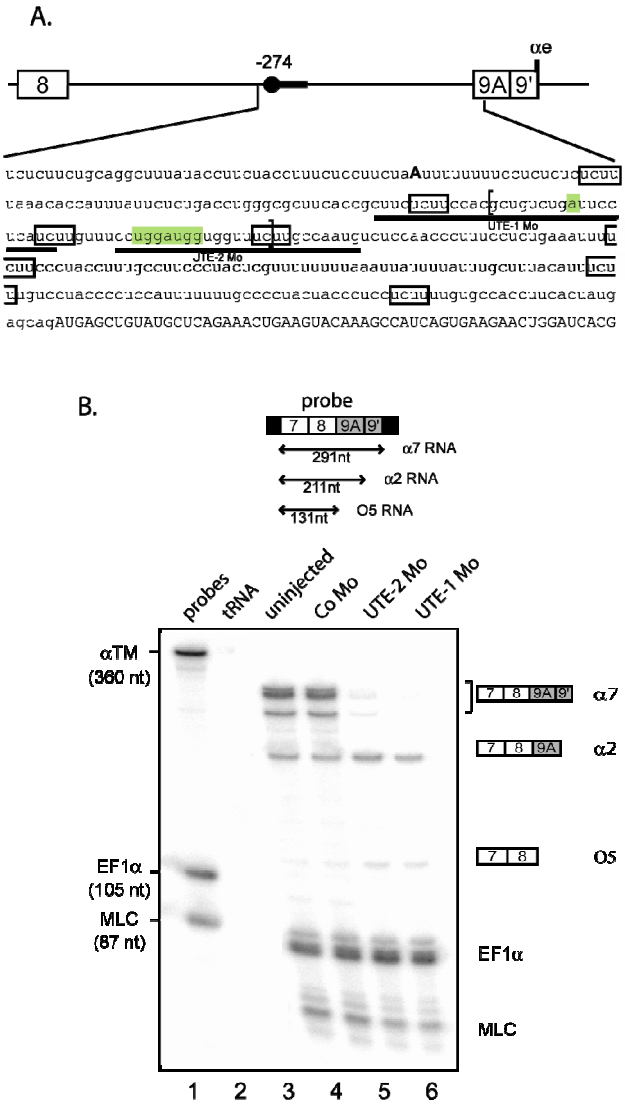


Figure 3. Morpholino targeted inhibition of the endogenous UTE modify the 3' end processing of the α-TM pre-mRNA. A. Partial sequence of the α-TM pre-mRNA upstream of exon 9A9' . Exon sequences are indicated in uppercase letters, and intron sequences are in lowercase letters. The distal -274 branch point is written in bold characters and the high affinity PTB binding sites are framed. UTE is delimited by brackets and the targeting positions of UTE-1 and UTE-2 Mo are underlined. **B.** Uninjected embryos or embryos injected with Co Mo or UTE-1 Mo or UTE-2 Mo were analyzed by a RNase protection assay using the probe illustrated at the top. Sequences derived from vector are represented by black rectangles. EF1-α and MLC1/3 probes were included in the assay as control for RNA recovery and embryo staging respectively. The identity or composition of each protected product is indicated on the right. α7 RNA produced two protected fragments because of three nucleotides differences within exon 9A9' between the two α-tropomyosin genes.

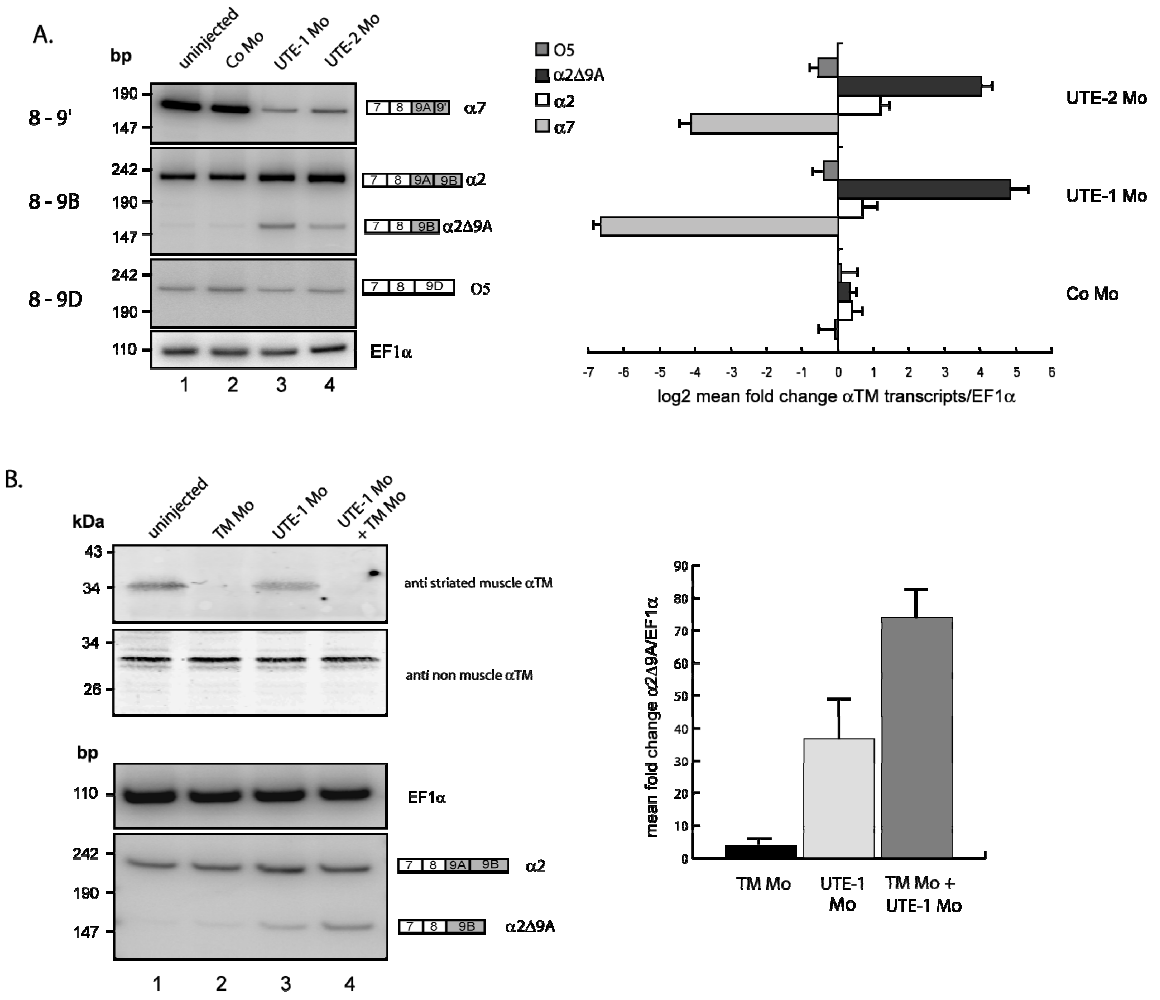


Figure 4. A. Targeting the endogenous UTE generates α2Δ9A mRNA, a potential non stop decay substrate. Endogenous α7, α2 and 05 RNAs were analyzed by RT-PCR using a ^{32}P labeled sense primer specific for exon 8 and reverse primers specifics for exon 9', 9B and 9D respectively. The PCR products were quantified by phosphorimager analysis and normalized to EF1α. The bar diagram shows the log2 mean fold change +/- SD of the different α tropomyosin isoforms in morpholino injected embryos relative to uninjected embryos (n = 3 independent experiments). **B.** Reduction of the α2Δ9A mRNA is translation dependant. A morpholino complementary to the region encompassing the start codon of the muscle α-tropomosin mRNAs (TM Mo) was injected alone or in combination with UTE-1 Mo. The steady state level of muscle α-tropomyosin was studied by western blot with the monoclonal anti-tropomyosin antibody (clone TM311). Non muscle tropomyosin was used as a loading control. Endogenous α2 and α2Δ9A RNAs were analyzed by RT-PCR using primers specific for exon 7 and 9B. The PCR products were quantified by phosphorimager analysis and normalized to EF1α. Data were presented as mean fold change +/- SD of α2Δ9A RNA in morpholino injected embryos relative to uninjected embryos (n = 3 independent experiments).

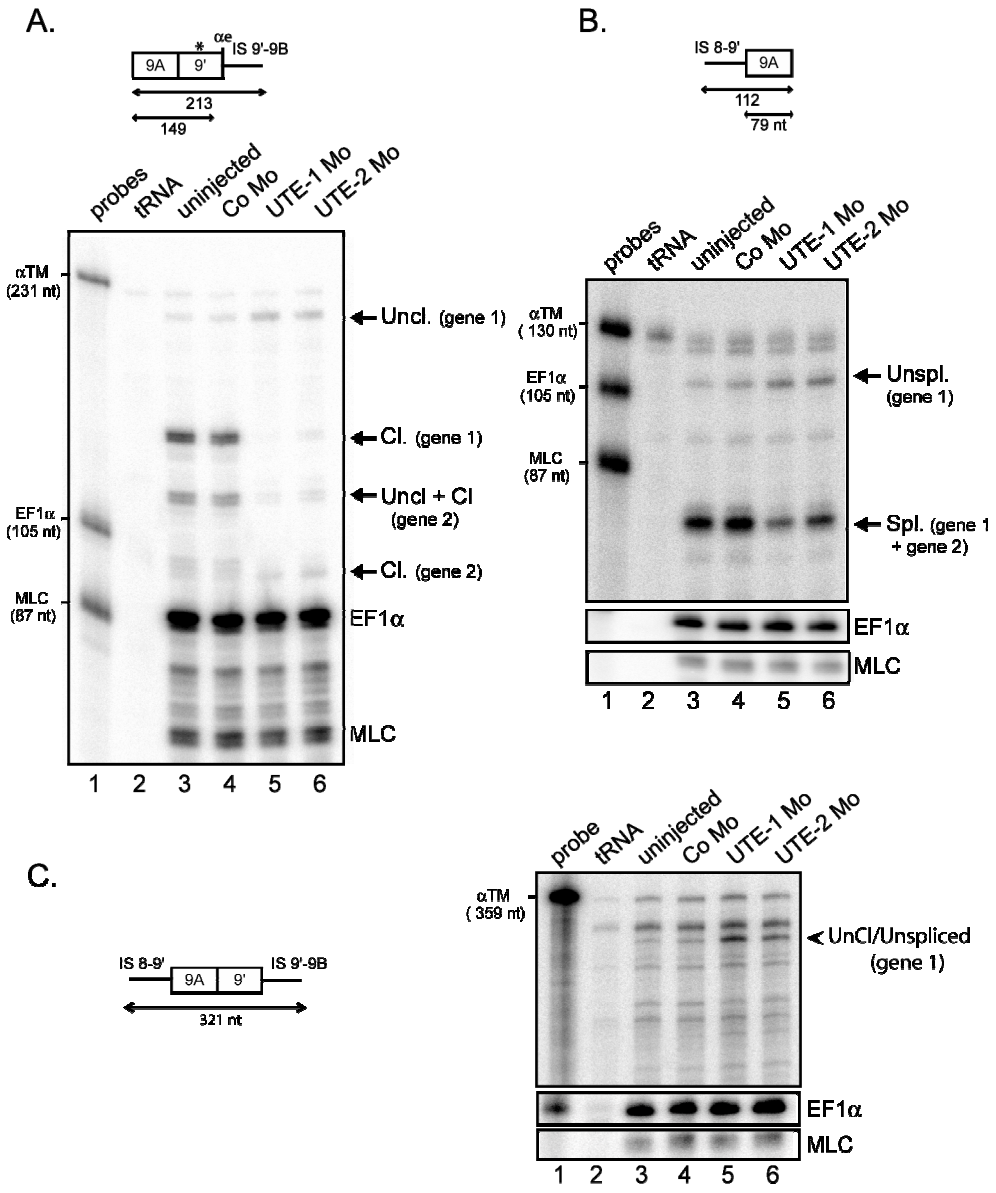
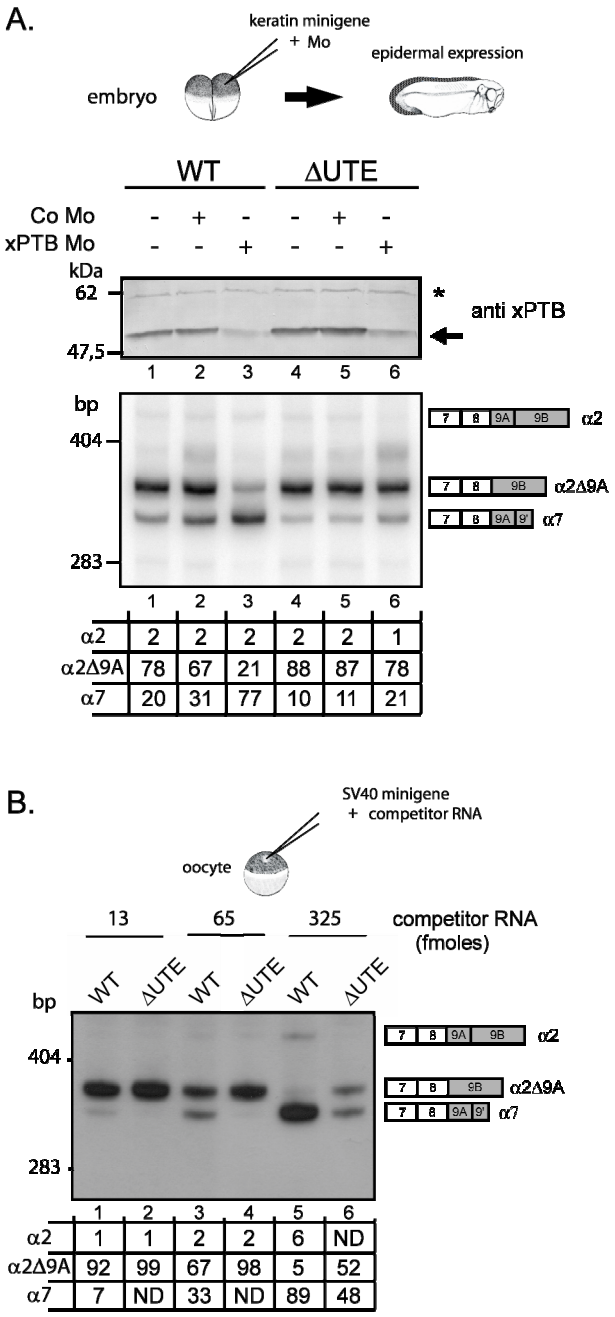


Figure 5. UTE activates the splicing and cleavage/polyadenylation of exon 9A9'. RNase protection analysis of RNA extracted from uninjected embryos or embryos injected with UTE-1 Mo, UTE-2 Mo, control morpholino (Co Mo). The RNase protection probes are presented above. **A.** UTE activates the cleavage/polyadenylation of exon 9A9'. RNase protection was realized with a RNA probe extending beyond αe poly(A) site as presented at the top. The 3 nucleotides difference within exon 9' between the two αTM genes is indicated by a star. The identity of uncleaved (Uncl.) and cleaved (Cl.) protected fragments is indicated on the right. **B.** UTE activates the splicing of exon 9A9'. RNase protection was realized with a RNA probe that cover the 3' splice site of exon 9A9' as presented at the top. The identity of unspliced (Unspl.) and spliced (Spl.) fragments is written on the right. **C.** Unprocessed RNAs accumulate in UTE Mo injected embryos. RNase protection was realized with a probe spanning the 3' splice site and extending beyond the cleavage site of exon 9A9' as described on the left. The protected fragment corresponding to uncleaved/unspliced (Uncl./Unspl) molecules was indicated on the right. **A-C** The protected fragments were quantified by phosphorimager analysis and normalized to uridine content and EF1-α.



	1	2	3	4	5	6
α2	2	2	2	2	2	1
α2Δ9A	78	67	21	88	87	78
α7	20	31	77	10	11	21

	1	2	3	4	5	6
α2	1	1	2	2	6	ND
α2Δ9A	92	99	67	98	5	52
α7	7	ND	33	ND	89	48

Figure 6. xPTB prevents the activity of UTE in non muscle cells. A. Mutant minigenes deleted for UTE (ΔUTE) and wild-type (WT) minigenes, driven by the keratin promoter, were injected in *Xenopus* embryo alone (-) or together with control morpholino (Co-Mo) or xPTB-Mo. The steady state level of xPTB was checked by western-blot using a xPTB antiserum. xPTB was indicated by an arrow and the nonspecific cross-reacting band indicated by a star was used as a loading control. The mRNA splicing pattern was assayed by 3' RACE-PCR using a ^{32}P oligonucleotide specific of the keratin minigene constructs. **B.** The wild-type (WT) and ΔUTE minigenes driven by the SV40 promoter were injected into oocyte nuclei alone or with increasing amount of 150PY competitor RNA. The mRNA splicing pattern was assayed by 3' race-PCR using a ^{32}P oligonucleotide specific of SV40 minigene constructs. **A, B** Quantification was performed on a Phosphorimager to yield the proportion of each product given in the table below. ND : Not detected.

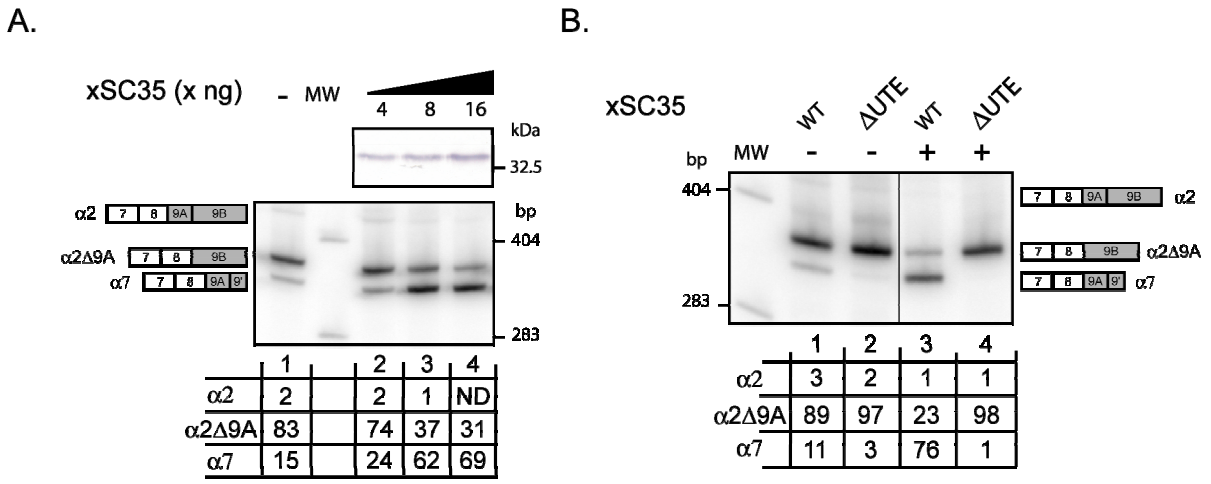


Figure 7. SC35 activates the usage of exon 9A9' as a 3' terminal exon through UTE. **A.** The wild-type minigene driven by the SV40 early promoter was injected into the nucleus of stage VI oocytes into which *in vitro* transcribed V5-tagged xSC35 RNA (4, 8 and 16 ng) had been previously injected into cytoplasm. V5-tagged xSC35 expression was monitored by western blotting using a V5 antibody. **B.** The wild-type (WT) and ΔUTE mutant minigenes were injected into the nucleus of stage VI oocytes into which *in vitro* transcribed V5-tagged xSC35 RNA had been previously injected into cytoplasm (+) or not (-). **A., B**. Total RNA was subjected to 3' RACE-PCR using a ^{32}P oligonucleotide specific of the SV40 minigene reporter. Quantification was performed on a Phosphorimager to yield the proportion of each product given in the table below.

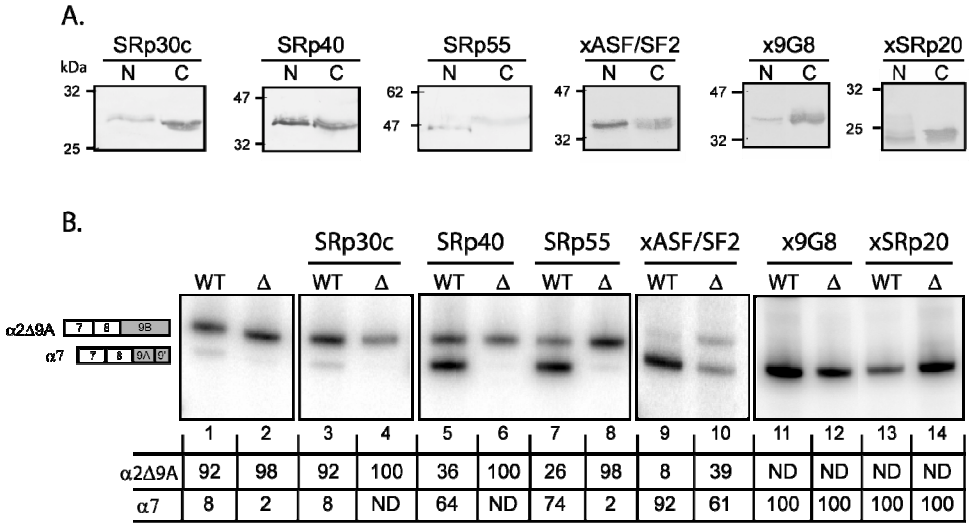


Figure 8. A subclass of SR proteins activate exon 9A9' splicing through UTE. The wild-type (WT) or ΔUTE mutant constructions were injected into the nucleus of stage VI oocyte alone or in presence of *in vitro* transcribed RNA encoding V5-tagged SR proteins. **A.** V5-tagged SR proteins expression in cytoplasmic and nuclear fractions was monitored by western blotting using a V5 antibody. **B.** Total RNA was subjected to 3' RACE-PCR using a ^{32}P oligonucleotide specific of the SV40 minigene reporter. The structure of the amplified fragments is indicated on the right. Quantification was performed on a Phosphorimager to yield the proportion of each mRNA given in the table below. ND : Not detected.

SUPPLEMENTAL DATA

name		sequence 5' - 3'
pActin.7-9B Δ -204-165	forward	GTCCATTGATGATTTAGAAGG
	reverse	AGGGTTGGAGACATTGGCAAGTGGAAGAGAAGCGGTGAAG
	forward	TTGCCAATGTCTCCAACCCT
	reverse	TGGAAAGGGTACGGAGGTAAGC
pActin.7-9B mut UGGAUGG	forward	GTCCATTGATGATTTAGAAGG
	reverse	GAAACCAGTATAAGGGAAACAAGATGAGGATCAG
	forward	CTTGTTTCCCTTATACTGGTTTCTTGCCAATGTCTC
	reverse	TGGAAAGGGTACGGAGGTAAGC
pActin.7-9B mut UGGAUGG +mut-196	forward	GTCCATTGATGATTTAGAAGG
	reverse	GATGAGGAAACAGACAGCGAGGAAGAGAAG
	forward	CGCTGTCTGTTTCCTCATCTTGTTTCCCT
	reverse	TGGAAAGGGTACGGAGGTAAGC
pActin.7-9B mut BS + 3'cons	forward	GTCCATTGATGATTTAGAAGG
	reverse	GGAAAAAAAAAATAGAAGGAGAAAGGTAGAAG
	forward	CTCCTTCTATTTTTTTTTCCTCTCTCTCTT
	reverse	CATGTGGTACCAGACTAGAG
pActin.7-9B mut BS	forward	GTCCATTGATGATTTAGAAGG
	reverse	GGAAAAAAAAAATAGAAGGAGAAAGGTAGAAG
	forward	CTCCTTCTATTTTTTTTTCCTCTCTCTCTT
	reverse	GGAGGGGTAGGACAAAGAAATG
pActin.7-9B SPA	forward	GCAGCACCCAGCCAATTC
	reverse	CCAACACACAGATCTAATGAAAATAAAGATCTTTTATTAAGTGGAAAGGGTACGG
	forward	TTCATTAGATCTGTGTGTTGGTTTTTTGTGTGCTTTTGGGTGTTTTTTACCCCTG
	reverse	CATGTGGTACCAGACTAGAG
xASF/SF2	forward	GGATCCATGGATACGTCAGCGGGC
	reverse	GCGGCCGCTGTACGAGAGCGAGATCTG
xSC35	forward	GATATCTGAGCTACGGTCGGCCT
	reverse	GCGGCCGCTGAAGACACTGCTCCCTC
xSRp20	forward	GATATCATGCATCGTGACTCCTGT
	reverse	GCGGCCGCCTTTCGCTCATTGGACC
x9G8	forward	GATATCATGTCGCGTTACGGGCGA
	reverse	GCGGCCGCCATTTCTTTCTGGGC
hSRp30c	forward	GGATCCACCATGTCGGGCTGGGCGGACG
	reverse	GCGGCCGCGTAGGGCCTGAAAGGAGAG
hSRp40	forward	GATATCCACCATGAGTGGCTGTCGGGTAT
	reverse	GCGGCCGCATTGCCACTGTCAACTGATCT
hSRp55	forward	GGATCCACCATGCCGCGCGTCTACATAG
	reverse	GCGGCCGCATCTCTGGAACTCGACCTG

Table S1 : Sequences of the forward and reverse oligonucleotides used to generate the different minigene mutants and PCR amplify the open reading frame of the different SR proteins.

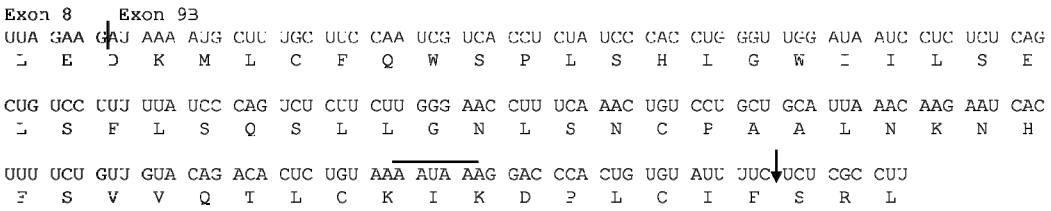


Figure S1. The splicing product α2Δ9A is a potential substrate for non sense decay (NSD). Nucleotide and derived amino acid sequences of the open reading frame encompassing exon 8 to exon 9B. The end of exon 8 and the beginning of exon 9B are written and the exon junction is indicated by the symbol (/). The hexanucleotide and the cleavage site of the αe poly(A) site are underlined and indicated by an arrow respectively.

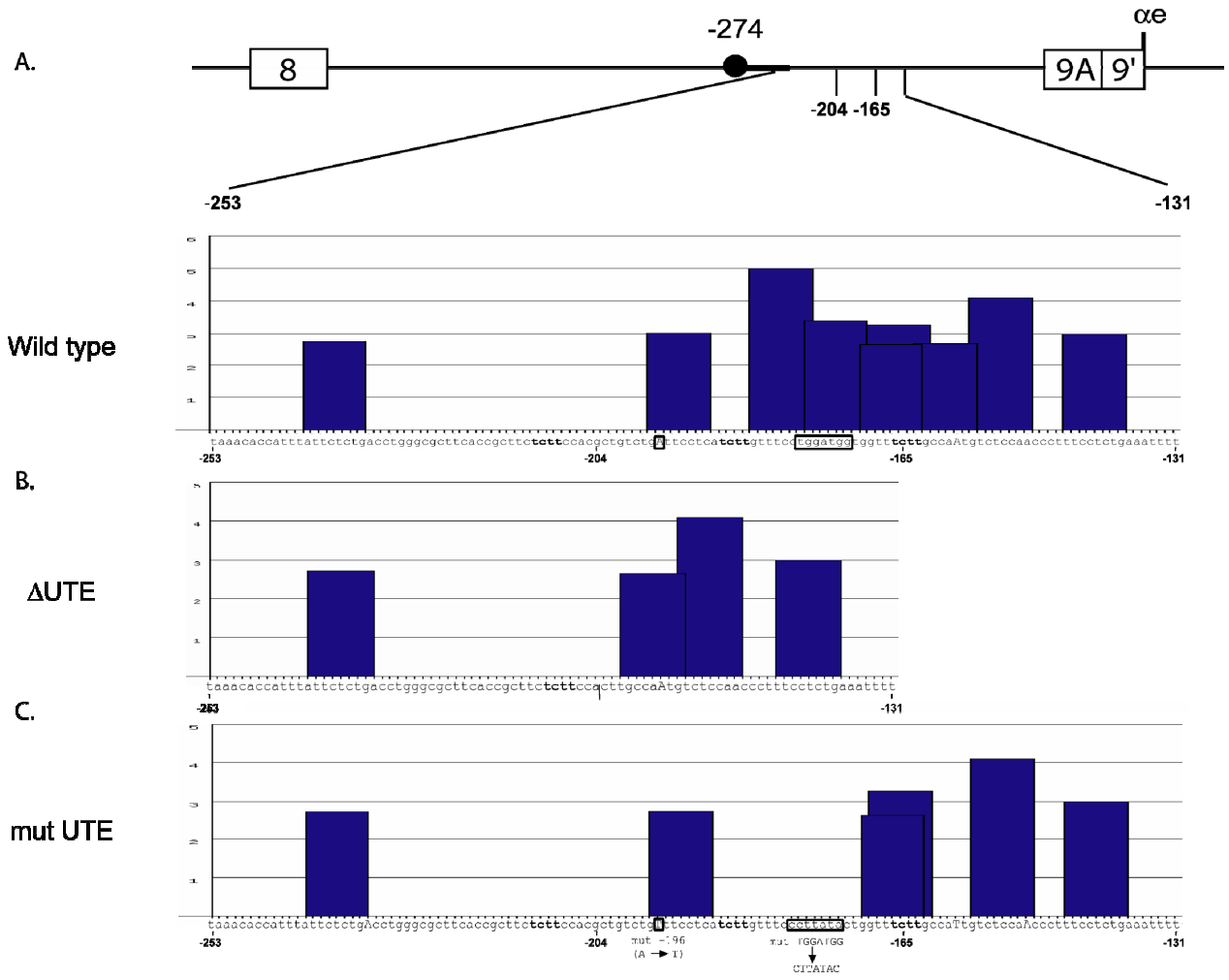
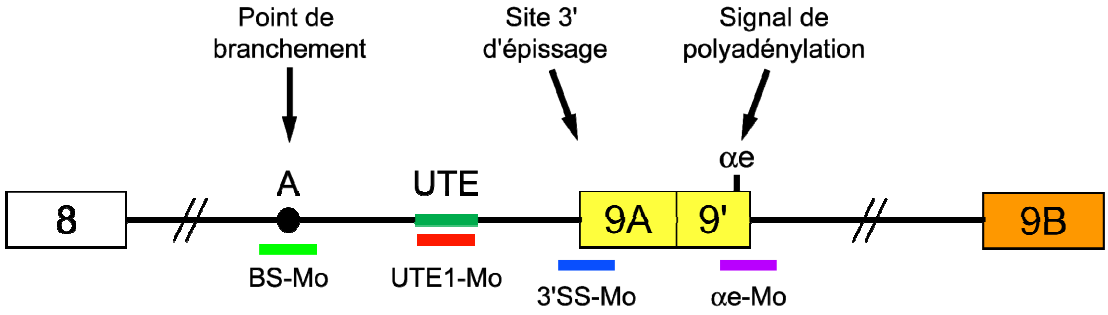


Figure S2. Localization of candidate SC35 binding sites in the intronic region encompassing UTE using ESE finder 3. **A.** The organisation of the wild-type pre-mRNA region spanning exon 8 to exon 9A9' is diagrammed at the top. The position of UTE from -204 to -165 is given and the distant -274 branch site and its associated polypyrimidine tract are represented by a black circle and a black rectangle respectively. The intronic sequence from -253 to -131 upstream of exon 9A is written below and the predicted SC35 binding sites are diagrammed with bars. The height of the bars indicates the strength of the potential SC35 binding sites as calculated by the ESE finder software. **B.** Effects of UTE deletion upon candidate SC35 binding sites. **C.** Effects of UTE mutation upon candidate SC35 binding sites. The mutations are indicated related to the wild type sequence.

A.



B.

BS-Mo	5'-TTAGAAGGAGAAAGGTAGAAGGTAT-3'
UTE1-Mo	5'-TTAGAAGGAGAAAGGTAGAAGGTAT-3'
3'SS-Mo	5'-GCATACAGCTCATCTGCTCATATGTG-3'
αe-Mo	5'-AGGGTAGGACGTGAGTTTTATTAAG-3'

Figure 30- A. Représentation schématique de l'extrémité 3' de l'ARNpré-m α-TM. A. Le point de branchement (A), la séquence régulatrice UTE et le signal de polyadénylation αe sont représentés. Les régions où s'hybrident les morpholinos oligonucléotides antisens sont indiquées par un trait. **B. Séquences des morpholinos oligonucléotides antisens.** Les nucléotides complémentaires du point de branchement, du dinucléotide AG du site 3' d'épissage et de l'hexanucléotide du signal de polyadénylation sont soulignés.

Nous avons donc montré que l'inactivation de la séquence UTE par injection d'un morpholino oligonucléotide antisens dérégule fortement l'épissage de l'ARNpré-m α-TM endogène. On observe en effet une diminution de la synthèse de l'ARNm α7 en prévenant l'utilisation des sites d'épissage et de clivage de l'exon 9A/9'. De plus, nous avons observé l'apparition d'une nouvelle isoforme ARNm qui résulte de l'épissage de l'exon 8 à l'exon 9B (l'ARNm α2Δ9A). Cette dérégulation entraîne une accumulation des ARNpré-m, et les ARNm α-TM sont globalement diminués.

1.2. Relation entre l'élément UTE et les signaux qui définissent l'exon 9A9' : point de branchement, site 3' d'épissage et signal de polyadénylation αe

Afin de préciser le mode d'action de la séquence UTE *in vivo* et d'étudier sa relation avec les machineries d'épissage et de clivage/polyadénylation, nous avons entrepris de perturber la reconnaissance du site 3' d'épissage et du signal polyadénylation αe de l'exon 9A9' avec des morpholinos oligonucléotides antisens. Les effets observés seront comparés à ceux obtenus avec le morpholino oligonucléotide antisens ciblant la séquence UTE précédemment utilisé et ayant l'effet le plus important (UTE1-Mo ; Anquetil *et al.*). Trois morpholinos oligonucléotides antisens ciblant respectivement le point de branchement -274 (Bs-Mo), le site 3' d'épissage (3'SS-Mo) et l'hexanucléotide AAUAAA du signal de polyadénylation αe (αe-Mo) de l'exon 9A9' ont été dessinés (figure 30).

Quinze nanogrammes des morpholinos 3'SS-Mo, αe-Mo et UTE1-Mo ont été injectés dans les deux blastomères d'embryons au stade deux cellules. Pour BS-Mo, la quantité injectée a été limitée à 5 ng, limite au-delà de laquelle les embryons injectés meurent au stade *neurula.* Après 36h de développement à 22°C, les embryons ont été fixés au stade bourgeon caudal (stade 26/27 d'après la classification de Nieuwkoop et Faber (Nieuwkoop et Faber, 1956)) et les ARN totaux ont été extraits et analysés.

1.2.1. Les morpholinos dirigés contre le point de branchement et le site 3' d'épissage bloquent l'épissage de l'exon 9A9'

Afin de vérifier que les morpholinos dirigés contre le point de branchement et le site 3' d'épissage bloquent effectivement l'épissage de l'exon 9A9', les ARN ont été analysés par

protection à la ribonucléase (RPA) à l'aide d'une sonde ARN antisens couvrant l'exon 9A et 33 nucléotides de l'intron en amont (figure 31). Les résultats présentés dans la figure 31 montrent que, comme attendu, le blocage du site 3' d'épissage par 3'SS-Mo provoque une augmentation importante (7,3 fois) de la forme non épissée, associée à une forte diminution de la forme épissée (comparer les pistes 3 et 6). De même pour BS-Mo (piste 5), on observe la disparition *quasi* complète de la forme épissée et une augmentation d'un facteur 3,1 de la forme non épissée. Les morpholinos 3'SS-Mo et BS-Mo sont donc fonctionnels et bloquent fortement la réaction d'épissage. Il est à noter que αe-Mo produit également une augmentation des ARN non épissés. Toutefois, la quantité des ARN épissés n'est pas modifiée par rapport aux embryons témoins non injectés (comparer les pistes 3 et 7). Ces résultats indiquent que le blocage du signal de polyadénylation n'affecte pas directement la réaction d'épissage. Comme nous l'avons décrit précédemment dans la publication, UTE1-Mo provoque également une augmentation des ARN non épissés (piste 4).

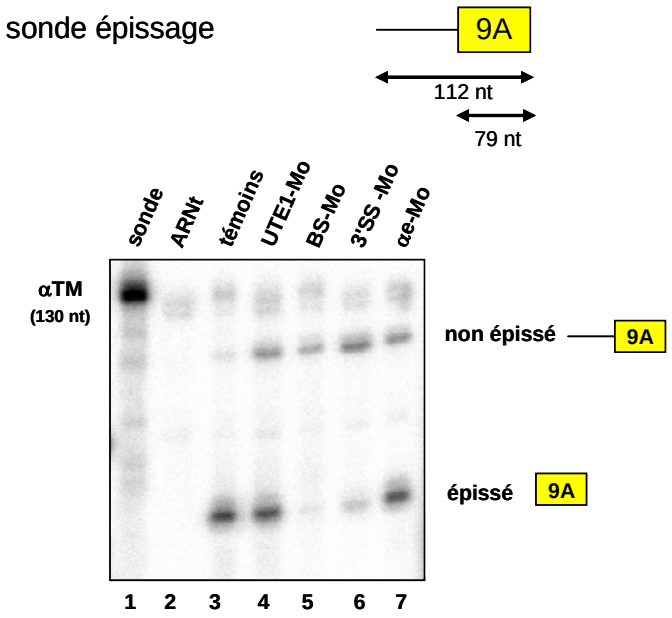


Figure 31- Analyse de la réaction d'épissage par protection à la ribonucléase des embryons injectés par différents morpholinos. Les morpholinos ciblant la séquence UTE (UTE1-Mo), le point de branchement -274 (Bs-Mo), le site 3' d'épissage (3'SS-Mo) et l'hexanucléotide AAUAAA du signal de polyadénylation αe (αe-Mo) de l'exon 9A9' ont été injectés dans les embryons de xénope. La sonde ARN antisens "épissage" et la taille attendue des fragments protégés après action de la RNase A sont indiqués. Les fragments ARN protégés ont été analysés par électrophorèse sur un gel d'acrylamide 5% en conditions dénaturantes. 10 µg d'ARN total d'embryons injectés ont été utilisés. Sur la droite, la structure des fragments protégés épissés et non épissés est indiquée.

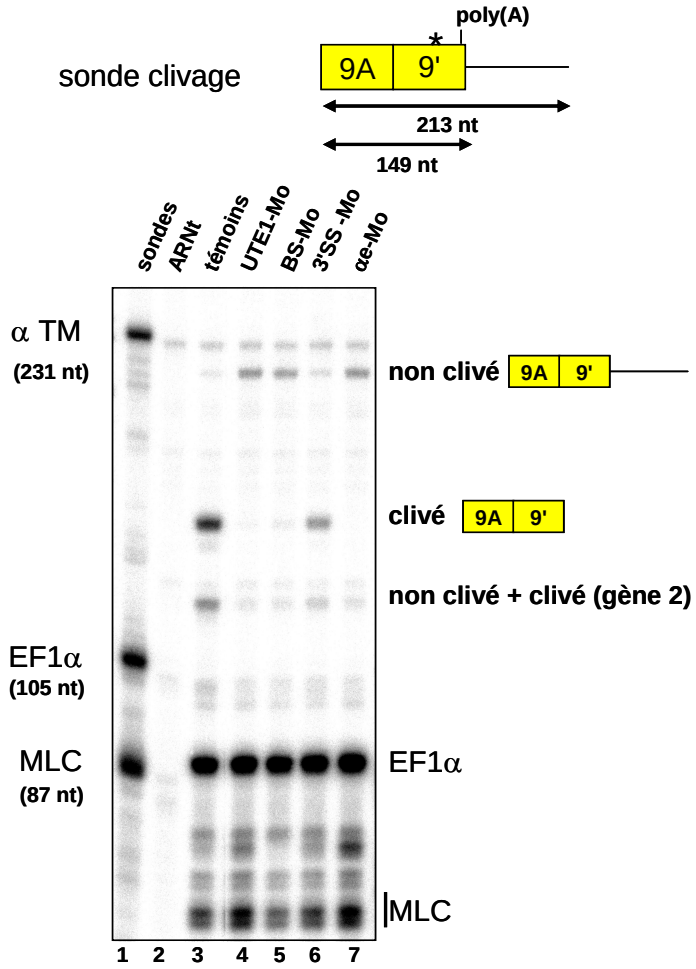


Figure 32- Analyse de la réaction de clivage par protection à la ribonucléase des embryons injectés par différents morpholinos. Les mêmes morpholinos que ceux utilisés dans la figure 31 ont été injectés dans des embryons de xénope. Une représentation schématique de la sonde antisens "clivage" et de la taille attendue des fragments protégés après action de la RNase A est donnée en haut de la figure. La différence de séquence entre les deux gènes α-TM présente dans l'exon 9' est indiquée par une étoile. Les fragments ARN protégés ont été analysés sur gel de polyacrylamide 5% en conditions dénaturantes. 10 µg d'ARN total d'embryons injectés ont été utilisés. Sur la droite sont indiquées schématiquement les structures des fragments protégés clivés et non clivés.

1.2.2. Le morpholino dirigé contre le signal de polyadénylation αe bloque le clivage de l'ARNpré-m

Les effets des différents morpholinos sur la réaction de clivage ont été analysés par RPA à l'aide d'une sonde couvrant l'exon 9A9' et 64 nt de l'intron en aval (figure 32). Comme attendu, l'injection de αe-Mo (piste 7) produit une forte augmentation (4,8 fois) de la forme non clivée, associée à la disparition de la forme clivée. Le blocage du signal de polyadénylation est donc effectif. Comme il l'a été décrit précédemment, l'injection de UTE1-Mo (piste 4) provoque une augmentation de la forme non clivée d'un facteur 5,7 et la disparition de la forme clivée. L'injection de BS-Mo (piste 5) induit le même effet avec une augmentation de la forme non clivée d'un facteur 5,7 et la disparition de la forme clivée. Ces résultats suggèrent que le point de branchement ainsi que la séquence UTE sont nécessaires à l'activation de la réaction de clivage. L'injection de 3'SS-Mo ne provoque qu'une légère augmentation des ARN non clivés (piste 6).

Du fait de sa pseudotétraploïdie, deux gènes α-TM sont présents dans le génome de *X. laevis*. La sonde utilisée est parfaitement complémentaire du gène 1 et permet donc de discriminer les formes clivées et non clivées des ARN produits uniquement par ce gène.

1.2.3. Conséquences du blocage du signal de polyadénylation et du site 3' d'épissage de l'exon 9A9' sur le profil d'épissage de l'ARNpré-m α-TM

Les effets des différents morpholinos sur les profils d'épissage du pre-ARNm α-TM ont été analysés par RPA à l'aide d'une sonde ARN antisens couvrant les exons 7-8-9A9' qui permet de discriminer les ARN α7 et α2 ainsi que les ARN O5 et/ou α2Δ9A (voir figure 33A). Une étude complémentaire par RT/PCR a également été réalisée pour discriminer les ARN05 et α2Δ9A. Les résultats de la RPA sont présentés dans la figure 33A. La normalisation des dépôts a été réalisée avec une sonde EF1α tandis qu'une sonde MLC1 permet de vérifier le stade de développement. Comme cela a été décrit précédemment les ARN α7 et α2 représentent respectivement 80 % et 20 % des ARNm α-TM présents dans les embryons témoins (piste 3). Comme décrit dans la publication, l'injection de UTE1-Mo (piste 4), provoque une diminution de près de 10 fois des ARN α7 et une augmentation de 2,8 fois des ARN α2. On observe une augmentation de 3 fois des fragments protégés exons 7-8 correspondant soit à des ARN O5, soit à des ARN α2Δ9A.

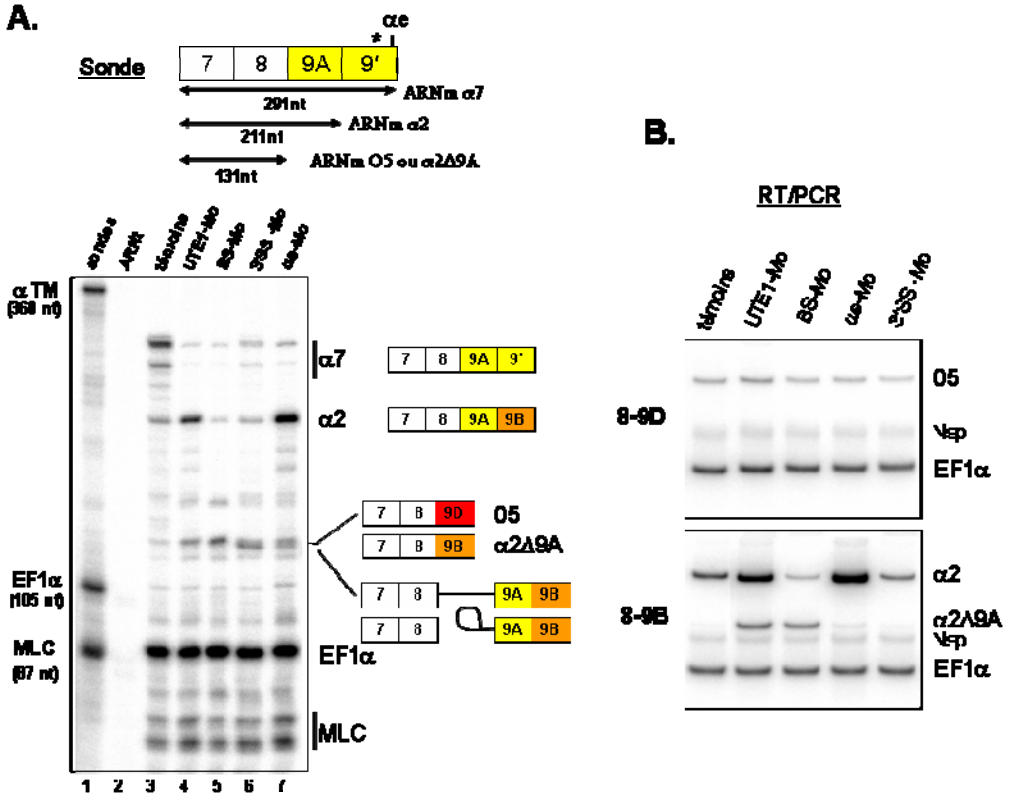


Figure 33- Analyse des profils d'épissage de l'extrémité 3' terminale de l'ARNpré-m α-TM. A. Les embryons ont été injectés avec les mêmes morpholinos que ceux utilisés pour les deux expériences précédentes puis les ARN totaux ont été analysés par protection à la ribonucléase à l'aide d'une sonde ARN antisens couvrant les exons 7-8-9A-9'. La différence de séquence entre les deux gènes α-TM présente dans l'exon 9' est indiquée par une étoile. La sonde antisens et la taille attendue des fragments protégés après action de la RNase sont représentées au-dessus du gel. Des sondes EF1α et MLC1/3 ont été incluses dans l'essai pour contrôler les dépôts et le développement des embryons. Les fragments ARN protégés ont été analysés sur un gel de polyacrylamide 5% en conditions dénaturantes. 10 µg d'ARN total d'embryons injectés ont été utilisés. Sur la droite sont indiquées schématiquement les structures des fragments protégés. Les séquences ARN protégées sont indiquées en grisé. **B.** Les ARN ont été analysés par RT/PCR avec des amorces spécifiques des exons 8 et 9D (panel du haut) et des exons 8 et 9B (panel du bas). Les dépôts ont été normalisés par amplification de l'ARNm EF1α. Un amplimère non spécifique amplifié avec les amorces EF1α est indiqué par Nsp.

ARN O5, soit à des ARN α2Δ9A. Une très légère diminution (d'un facteur 1,2) de la quantité globale de transcrits α-TM est également observée. L'injection de BS-Mo (piste 5) entraîne une très forte diminution (21 fois) des ARN α7 ainsi qu'une augmentation des fragments protégés 7-8. La quantité des ARN α2 reste sensiblement équivalente et la quantité totale des transcrits α-TM diminue de 5,5 fois. Le blocage du site d'épissage par 3'SS-Mo (piste 7) produit une forte diminution des ARN α7 et une légère diminution des ARN α2. Deux bandes de tailles très proches, correspondant aux fragments protégés 7-8, sont également observées. Comme précédemment décrit, la bande supérieure correspond aux ARN α2Δ9A ou O5. La bande inférieure correspond à des ARN dans lesquels l'exon 8 n'est pas suivi par l'exon 9B ou l'exon 9D. La différence de taille entre les deux bandes est due au fait que les exons 9B et 9D ont leurs deux premiers nucléotides en commun avec l'exon 9A de la sonde ADNc. La bande inférieure pourrait correspondre à des ARN où l'intron entre les exons 8 et 9A n'est pas épissé, ou à des ARN dans lesquels l'intron entre l'exon 8 et 9A est engagé dans la première réaction de transestérification. Le blocage du signal αe entraîne une diminution des ARN α7 d'un facteur 7, ainsi qu'une forte augmentation (4,7 fois) des ARN α2 (piste 8). A nouveau, deux bandes correspondant aux fragments protégés 7-8 sont observées. Le blocage du signal de polyadénylation αe semble donc favoriser la réaction d'épissage de l'exon 9A à l'exon 9B.

Afin d'identifier les ARNm correspondant aux fragments protégés 7-8, une analyse par RT-PCR a été réalisée avec une amorce sens spécifique de l'exon 8 et des amorces antisens spécifiques des exons 9B et 9D. Les résultats sont présentés dans la figure 33B. L'amplification des ARN EF1α est utilisée pour normaliser les quantifications. Les résultats obtenus avec les amorces 8 et 9D, présentés dans le panneau du haut, montrent que la quantité des ARN O5 ne varie pas quels que soient les morpholinos injectés. Nous pouvons donc exclure que l'augmentation des fragments 7-8, observée en RPA avec les différents morpholinos, soit la conséquence d'une augmentation des ARN O5. L'amplification réalisée avec les amorces 8 et 9B est présentée dans le panneau du bas. Les ARN α2 sont présents dans les embryons témoins (piste 1). L'injection de UTE1-Mo entraîne une augmentation des ARN α2 d'un facteur 2,8 et l'apparition des ARN α2Δ9A (piste 2) comme décrit précédemment dans l'article. L'augmentation de la quantité de fragments protégés 7-8 observée lors de l'analyse par RPA, résulte donc d'une augmentation de la quantité des ARN α2Δ9A. Pour BS-Mo (piste 3), on observe une très forte diminution des ARN α2 (5,7 fois) ainsi qu'une forte augmentation de la bande correspondant aux ARN α2Δ9A. Ces résultats

indiquent que le blocage de la séquence UTE et du point de branchement provoque le saut de l'exon 9A/9' et l'épissage de l'exon 8 à l'exon 9B. Comme attendu, le blocage du site 3' d'épissage par 3'SS-Mo diminue de 2 fois les ARN α2. L'épissage de l'exon 8 à l'exon 9A devient impossible, et donc empêche la formation des ARN α2. Par contre de manière surprenante, aucune augmentation des ARN α2Δ9A n'est observée. L'injection de αe-Mo provoque une très forte augmentation (4,2 fois) des ARN α2 mais n'entraîne pas la formation des ARN α2Δ9A. Comme observé par RPA, ce résultat indique que le blocage du signal de polyadénylation αe favorise la réaction d'épissage de l'exon 9A à l'exon 9B.

1.3. Discussion et perspectives

Ces travaux ont conduit à l'identification et à la caractérisation d'une séquence activatrice, nommée UTE, nécessaire à l'utilisation de l'exon 9A9' en tant qu'exon 3' terminal dans les cellules musculaires. La mutation de cette séquence abolit l'épissage de l'exon 9A9' en tant qu'exon 3' terminal (traduit par une baisse de la production des ARN α7) et provoque de manière concomitante une augmentation du saut de cet exon (traduit par une augmentation de la production des ARN α2Δ9A) et de son épissage en tant qu'exon interne (traduit par une augmentation de la production des ARN α2). Le renforcement du site 3' d'épissage de l'exon 9A9' et de son signal de polyadénylation αe permet un épissage de l'exon 9A9' en tant qu'exon terminal indépendamment de la séquence UTE, montrant que cette dernière est requise pour favoriser la reconnaissance de ces signaux faibles. De manière intéressante, nous avons également observé que la mutation du point de branchement distal -274 produit des effets similaires, suggérant que la séquence UTE agit en partie en favorisant la reconnaissance de ce point de branchement. Cette hypothèse est confortée par l'étude *in vivo* réalisée avec les morpholinos oligonucléotides ciblant la séquence UTE et le point de branchement. En effet, l'inactivation du point de branchement et de la séquence UTE induit le saut de l'exon 9A9' pour produire l'ARN α2Δ9A alors que le blocage du site 3' d'épissage ne produit pas un tel effet. La séquence UTE semble également importante pour l'activation du signal de polyadénylation αe. En effet, son inactivation provoque l'accumulation d'ARNpré-m non clivés de manière comparable à l'inactivation du signal de polyadénylation. Ces résultats suggèrent un modèle de régulation dans lequel la séquence UTE favoriserait la liaison de la

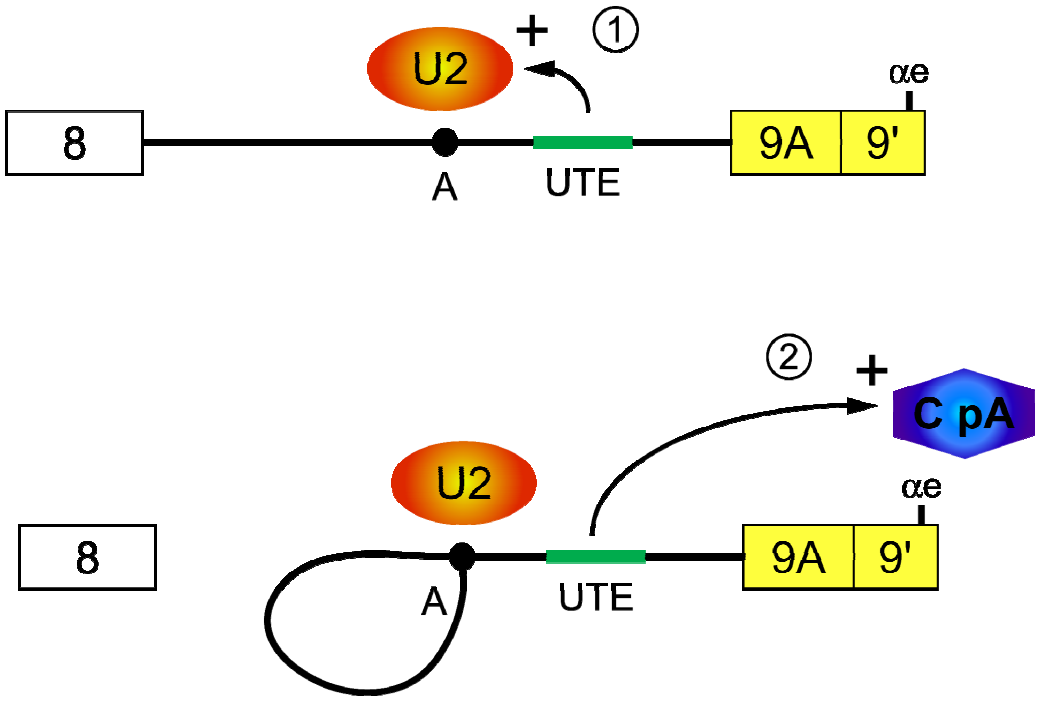


Figure 34- Modèle d'activation exercée par la séquence UTE sur l'épissage alternatif de l'exon terminal de l'ARNpré-m α-TM. La séquence UTE stabiliserait la liaison de la snRNP U2 sur le point de branchement distal et activerait la première réaction de transestérification. La séquence UTE activerait dans un second temps le signal de polyadénylation αe en facilitant le recrutement du complexe de clivage/polyadénylation (C pA).

U2snRNP sur le point de branchement distal qui à son tour stabiliserait, par une interaction directe, la liaison de facteurs de polyadénylation tels que CPSF au signal αe (figure 34). Par la suite, en accord avec le modèle de la définition de l'exon le complexe mis en jeu pourrait stabiliser les facteurs impliqués dans la reconnaissance du site 3' d'épissage tels que U2AF65 et U2AF35 permettant la définition de l'exon 9A9' en tant qu'exon terminal.

Les résultats obtenus suite à l'injection de 3'SS-Mo sont assez intrigants. En effet, celui-ci inhibe l'épissage de l'exon 9A/9', mais sans provoquer d'augmentation des ARN O5 et α2Δ9A qui résulteraient du saut de cet exon. Une interprétation serait que la première réaction de trans-estérification puisse être initiée indépendamment de la reconnaissance du site 3' d'épissage. La deuxième réaction de trans-estérification ne pouvant avoir lieu, les intermédiaires d'épissage seraient ensuite dégradés. L'apparition spécifique de fragments protégés 7-8 provenant d'ARN dans lesquels l'intron en aval de l'exon 8 n'a pas été excisé ou d'ARN finissant par l'exon 8 serait en accord avec une telle hypothèse. L'activation de la première réaction de transestérification indépendamment du site 3' a été démontrée *in vitro* avec un pré-ARNm épissé selon le modèle de l'intron définition (Smith et Nadal-Ginard, 1989). Cependant, cette hypothèse est en désaccord avec le modèle de la définition de l'exon qui présuppose que la reconnaissance de sites encadrant l'exon précède la mise en place d'un complexe d'épissage catalytiquement actif. Ici, la séquence UTE activerait le point de branchement distal -274 indépendamment de la définition de l'exon 9A9' et celui-ci serait défini secondairement par une interaction entre le complexe assemblé sur l'UTE et les facteurs généraux de clivage/polyadénylation et d'épissage. Ce modèle introduirait un nouveau mode de régulation de l'épissage alternatif, en amont de la définition d'un exon. Afin de vérifier cette hypothèse, deux approches sont envisageables. D'une part, les ARN extraits d'embryons injectés avec le morpholino 3'SS-Mo seront analysés par PCR quantitative avec plusieurs couples d'amorces permettant d'amplifier les jonctions exons/ introns des exons 8 et 9A et différentes régions de l'intron 8-9A. D'autre part, il est également envisageable de quantifier l'accumulation du lasso dans cette population d'ARN par une approche de RT/PCR basée sur la capacité de la transcriptase inverse à « traverser » une liaison phosphodiester 2'-5' (Vogel *et al.*, 1997). Cette seconde approche, semi-quantitative, présente en plus l'avantage de permettre de déterminer quel nucléotide de l'ARN est « branché ». En effet, il existe quatre autres points de branchement potentiels (Hamon Gouault, 2002) qui pourraient être utilisés. Les premiers résultats obtenus avec cette méthode montrent que le blocage de l'UTE à l'aide del'UTE1-Mo empêche la formation d'intermédiaire en lasso au niveau du point de branchement en -274 (Agnès Méreau, résultats non publiés). Il est possible que

l'intermédiaire « bloqué » après la première étape de transestérification soit dégradé. Ainsi, un moyen pour observer son accumulation serait d'inhiber également la machinerie de dégradation des ARN à l'aide de morpholinos ciblant une protéine de l'exosome telle que ExoSC9 ou l'exoribonucléase Xrn1.

Nous avons également observé que l'inactivation du signal de polyadénylation αe par le morpholino αe-Mo provoque une forte augmentation de la production des ARN α2. Ce résultat qui est en accord avec le modèle de la définition de l'exon, dans lequel l'activation d'un signal de polyadénylation en aval d'un site d'épissage 5' prévient l'utilisation de ce dernier, suggère que la synthèse exclusive de l'ARNm α2 dans les muscles striés adultes pourrait résulter de l'absence de reconnaissance du signal de polyadénylation αe. Afin de préciser le mécanisme de synthèse de l'ARN α2 dans des conditions physiologiques, il sera important de vérifier si l'épissage de l'exon 9A en tant qu'exon interne, après blocage du signal de polyadénylation αe, est dépendant de la séquence UTE et du point de branchement distal -274. Ceci pourra être réalisé par co-injection des morpholino UTE-1-Mo et αe-Mo d'une part et Bs-Mo et αe-Mo d'autre part.

Ces études ont également permis de préciser que la PTB réprime l'épissage de l'exon 9A9' dans les tissus non musculaires en prévenant l'utilisation de UTE. Plusieurs sites de haute affinité pour la PTB sont présents dans la séquence UTE suggérant que sa liaison dans les tissus non musculaires pourrait prévenir la liaison de facteurs activateurs comme cela a été montré pour les ARNpré-m de l'alpha actinine et de la TnT cardiaque (Charlet *et al.*, 2002; Ladd *et al.*, 2005). Les résultats de surexpression ou d'inactivation de la PTB, précédemment obtenus par l'équipe (Le Sommer *et al.*, 2005) ont démontré que la quantité relative de la PTB jouait un rôle important dans la maturation différentielle de l'extrémité 3' du pré-ARNm α-TM. Nous avons montré que trois protéines SR, SC35, SRp40 et SRp55, activent l'utilisation de l'exon 9A9' dans l'ovocyte, un contexte fortement répresseur, *via* la séquence activatrice UTE. Des études préliminaires, également réalisées dans l'ovocyte, ont permis de montrer qu'il existait un antagonisme fonctionnel entre la protéine SC35 et la PTB (Le Sommer, 2006). Ces données suggèrent que ces trois protéines SR pourraient participer à l'assemblage d'un complexe activateur sur la séquence UTE et contrarier la liaison de la PTB sur cette région intronique. Bien que l'étude *in silico* de la séquence UTE, à l'aide du logiciel ESEfinder (Cartegni *et al.*, 2003), ait révélé plusieurs motifs de liaison à ces trois protéines SR, nous n'avons pas mis en évidence de liaison de ces facteurs sur l'ARN *in vitro*. Cette

absence de liaison peut être expliquée par la nécessité d'une modification post-traductionnelle des protéines ou d'une liaison coopérative entre elles ou avec d'autres facteurs (Tenenbaum et Aguirre-Ghiso, 2005). Il est possible de vérifier cette dernière hypothèse *in vitro* en incubant un ARN dans un extrait nucléaire dépourvu de protéines SR, puis complémenté par des combinaisons des protéines SC35, SRp40 et SRp55.

Une augmentation de l'inclusion de l'exon 9A9' lorsque l'expression de la PTB est fortement diminuée est également observée en absence de l'élément UTE, démontrant que cet exon demeure faiblement défini en absence de l'élément UTE. Cette observation suggère que la PTB doit également inhiber la reconnaissance de l'exon 9A9' par les machineries d'épissage et de clivage/polyadénylation. De nombreux motifs de haute affinité pour la PTB sont présents entre le point de branchement et le site 3' d'épissage, il est donc possible que la PTB prévienne la liaison de U2AF65 comme cela a déjà été décrit pour l'exon 6B de l'ARNpré-m de la β-tropomyosine (Sauliere *et al.*, 2006). La fixation de la PTB inhibe alors le recrutement de la snRNP U2 et donc la mise en place du spliceosome. ll est également possible que la PTB empêche la définition de l'exon en inhibant les interactions au travers de l'exon (Izquierdo *et al.*, 2005) ou qu'elle bloque la transition d'un exon défini vers la mise en place d'un spliceosome actif (Sharma *et al.*, 2008). Cette dernière hypothèse observée dans le cas de l'exon N1 de l'ARNpré-m de c-src, signifierait que l'exon 9A9' serait correctement défini mais que la PTB inhiberait les interactions au travers de l'intron entre les snRNPU1 et U2. Il a également été montré que la PTB pouvait inhiber la réaction de clivage/polyadénylation en contrariant la liaison du facteur de polyadénylation CstF64 sur la séquence DSE (Castelo-Branco *et al.*, 2004). Cependant une activité directe de la PTB sur le signal de polyadénylation αe semble improbable car des rapporteur chimériques comportant le dernier exon du pré-ARNm de la β-globine et le signal de polyadénylation αe ne sont pas sensibles à la répression de la PTB. De plus, la surexpression de CstF64 dans l'ovocyte, un contexte répresseur, n'active pas l'utilisation de l'exon 9A9' démontrant qu'une augmentation de CStF64 n'est pas suffisante pour contrarier la répression exercée par la PTB. Il est également intéressant de noter que l'inactivation de UTE ne provoque aucune augmentation de l'épissage de l'exon 8 à l'exon 9D contrairement à ce qui est observé lorsque la PTB est spécifiquement surexprimée dans le myotome (Le Sommer *et al.*, 2005). Cette observation suggère que la PTB pourrait également réprimer la reconnaissance de l'exon 9B pour favoriser l'utilisation de l'exon 9D ou bien activer directement la reconnaissance de l'exon 9D. Ils serait intéressant de tester cette dernière hypothèse puisque plusieurs études ont démontré que la PTB pouvait

activer *in vitro* les réactions de clivage et de polyadénylation en se liant à des séquences USE (Moreira *et al.*, 1998 ; Hall Pogar *et al.*, 2007 ; Danckwardt *et al.*, 2007).

L'inactivation de la séquence UTE au sein de l'ARNpré-m α-TM endogène a permis de mettre en évidence un nouveau profil d'épissage de cet ARNpré-m, l'épissage direct de l'exon 8 à l'exon 9B pour produire l'ARNm α2Δ9A. Ce résultat n'est pas surprenant puisque les études réalisées avec des minigènes couvrant les exons 7 à 9B ont montré que l'épissage entre l'exon 8 et l'exon 9B est possible. L'ARN α2Δ9A présente cependant la particularité de ne posséder aucun codon stop en phase avec le cadre de lecture ouvert, suggérant qu'il pourrait être rapidement dégradé par le mécanisme de surveillance NSD (*Non stop Decay*) (Frischmeyer *et al.*, 2002). Cette dégradation spécifique expliquerait pourquoi cet ARNm non productif n'avait jamais été visualisé auparavant. Le mécanisme NSD étant dépendant de la traduction nous avons inhibé spécifiquement la traduction des ARNm α-TM musculaires. De manière assez surprenante, les embryons dépourvus d'α-tropomyosine musculaire se développent normalement jusqu'au stade bourgeon caudal tardif, sans présenter pour autant de battement cardiaque et de mouvement musculaire. La présence de l'α-tropomyosine musculaire ne semble donc pas nécessaire à l'assemblage du myotome. Il pourrait cependant être intéressant de vérifier par des études d'imagerie si le myotome est désorganisé dans ces embryons.

Nous avons montré que la quantité à l'équilibre d'ARNm α2Δ9A augmente suite à l'inhibition spécifique de la traduction ce qui est en accord avec une dégradation ciblée de ces ARN par le mécanisme NSD. Ce résultat suggère que le mécanisme NSD participe au contrôle strict de la maturation de l'extrémité 3' terminale de l'ARNpré-m α-TM. A notre connaissance, c'est la première fois qu'un rôle physiologique du NSD est évoqué chez les vertébrés dans le contrôle des ARNm. Chez la levure *S. cerevisiae*, modèle dans lequel le mécanisme NSD a été identifié et caractérisé, la dégradation de l'ARNm est médiée par le facteur Ski7p un paralogue du facteur de terminaison de la traduction eRF3. Ski7p se positionnerait dans le site A du ribosome bloqué à l'extrémité 3' de l'ARNm et recruterait l'exosome, le complexe de dégradation des ARN dans le sens 5' → 3'. Chez les vertébrés, des études préliminaires ont démontré la persistance de ce mécanisme de surveillance mais les processus mis en œuvre et les facteurs impliqués n'ont pas été caractérisés. Une étude phylogénique a révélé que le facteur Ski7p n'est présent que dans une sous famille des levures

ascomycètes (Atkinson *et al.*, 2008). La fonction de Ski7p chez les eucaryotes supérieurs pourrait être assurée par le facteur Hbsp1, un deuxième paralogue du facteur eRF3. Ce facteur est également impliqué dans la voie de dégradation NGD (*No Go Decay*) qui permet l'élimination des ARNm sur lesquels les ribosomes sont bloqués (Doma *et al*, 2006). L'implication du facteur Hbsp1 dans le processus de NSD et la dégradation ciblée de l'ARNm α2Δ9A pourra être vérifiée par l'inactivation de ce facteur dans l'embryon. Des études d'inactivation de constituants de l'exosome tels que ExoSC9 et de l'exonucléase 5' → 3' Xrn1 permettraient de préciser la voie de dégradation cytoplasmique impliquée dans le processus NSD chez les vertébrés. Un morpholino ciblant un des facteurs du noyau de l'exosome, ExoSC9, est actuellement en cours de validation dans l'équipe.

Chapitre 2. Identification de facteurs impliqués dans la régulation de l'épissage de l'exon 9A9'

2.1. Introduction

Les résultats présentés dans le chapitre précédent décrivent un élément activateur (UTE) nécessaire à la reconnaissance de l'exon 9A9' en tant qu'exon terminal dans les somites de l'embryon de xénope. Nous avons également montré *in vivo* que la protéine PTB contrecarre la fonction de l'élément activateur UTE. D'autre part, les résultats de la surexpression dans l'ovocyte des protéines SR SC35, SRp40 et SRp55, suggèrent un antagonisme fonctionnel, dépendant de l'UTE, entre ces protéines et la PTB. Ayant observé la présence de sites putatifs de fixation de protéines de protéines SR au sein de cette région ainsi de deux sites de fixation de la protéine PTB, nous avons émis l'hypothèse que la fixation de la PTB sur cette région pourrait inhiber l'interaction de facteurs activateurs, comme certaines protéines SR, et donc modifier la composition du complexe protéique assemblé autour de l'UTE.

Afin de tester cette hypothèse, et de déterminer quelles sont les protéines recrutées sur la séquence UTE, nous avons isolé les complexes assemblés autour la région activatrice dans des extraits nucléaires dépourvus ou non de PTB puis analysé leur contenu. Dans un premier temps, les stratégies déjà éprouvées pour purifier des complexes RNP ont été étudiées afin de déterminer celle qui est la plus adaptée à notre problématique. La description des différentes méthodologies et la justification de celle que nous avons choisi, font l'objet de la première partie de ce chapitre. Dans un second temps, nous avons déterminé la séquence de l'ARN à utiliser et choisi un extrait nucléaire adapté cette étude *in vitro*. Les critères qui ont guidé nos choix sont présentés dans la deuxième partie de ce chapitre. Enfin, la troisième partie décrira les résultats que nous avons obtenus, et l'implication des protéines que nous avons spécifiquement identifiées sera discutée.

2.2. Stratégies de purifications et de caractérisation des complexes RNP

Plusieurs approches peuvent être utilisées pour purifier des complexes RNP. Généralement, l'assemblage des complexes est effectué sur une séquence d'ARN qui a été

sélectionnée sur la base de son implication dans la régulation d'un évènement moléculaire (stabilité, épissage, clivage/polyadénylation, export, etc.). L'objectif étant de rechercher ses partenaires protéiques susceptibles d'être impliqués dans sa fonction régulatrice. Si une protéine du complexe est connue, il est possible d'immunoprécipiter le complexe et d'identifier son contenu en ARN ainsi que les protéines qui y sont associées (Keene *et al.*, 2006). Dans notre cas, en l'absence de connaissance préalable de protéines appartenant à ce complexe, cette stratégie ne semble pas appropriée. La seule chance d'isoler les complexes est donc d'utiliser l'ARN comme « appât ».

Dans la grande majorité des études, les complexes sont assemblés *in vitro* dans des extraits nucléaires (Grabowski et Sharp, 1986; Bennett *et al.*, 1992; Caputi *et al.*, 1999; Hartmuth *et al.*, 2002), et plus rarement *in vivo* en cellules ou dans des ovocytes de xénope (Srisawat et Engelke, 2001; Masuyama *et al.*, 2007). Les complexes peuvent être pré-formés autour de l'ARN, *in vitro* dans l'extrait nucléaire, ou *in vivo* dans les cellules, puis isolés en utilisant une propriété particulière de l'ARN et une matrice adéquate. Cette approche « plus physiologique » permet d'avoir accès aux complexes fixés sur les séquences régulatrices ainsi qu'aux différents facteurs de couplage entre deux mécanismes moléculaires, comme par exemple l'épissage et le processus de clivage/polyadénylation.

Il est également possible de reconstituer les complexes autour d'un ARN fixé sur un support. Cette stratégie implique une première étape de production de l'ARN suivie de son immobilisation sur une matrice et de l'incubation avec des extraits nucléaires. Cette procédure permet d'isoler les protéines qui s'assemblent sur l'ARN mais également d'étudier les mécanismes et la dynamique de l'assemblage de ces protéines sur les séquences régulatrices. Les protéines peuvent être éluées sélectivement ou non de cette colonne puis séparées sur gel d'électrophorèse avant d'être identifiées par western-blot ou spectrométrie de masse. L'immobilisation des ARN sur la matrice impose des contraintes physiques susceptibles de modifier l'assemblage des protéines autour de la séquence d'intérêt.

2.2.1. Utilisation de l'ARN comme « appât »

Les approches présentées précédemment nécessitent de transformer l'ARN d'intérêt en « appât » grâce à une « étiquette » qui va permettre de l'isoler dans un extrait cellulaire. La nature de ces étiquettes peut être diverse.

Synthèse de l'ARN modifié	Taille de l'ARN (nt)	Résultats obtenus	Extraits utilisés	Chromato/RNP	Références
Incorporation aléatoire d'un nucléotide biotinylé au cours d'une transcription *in vitro*	465	Purification de U2, U5, et U4/U6 snRNP	EN de cellules HeLa	RNP	(Grabowski et Sharp, 1986)
	465	Composition du complexe H	EN de cellules HeLa	RNP	(Reed, 1990)
	460	Purification des protéines des complexes A, B et C	EN de cellules HeLa	RNP	(Neubauer *et al.*, 1998)
	200	Identification des facteurs associés aux ESE riches en purines	ovocytes de xénope	RNP	(Masuyama *et al.*, 2007)
	213 - 547	Purification de protéines du spliceosome	EN de cellules HeLa	RNP	(Bennett *et al.*, 1992)
Biotinylation d'un ARN transcrit *in vitro*	50 - 71	Identification de hnRNP K dans la liaison à un ISE de l'exon 6A de la β–tropomyosine de poulet	EN de cellules HeLa	Chromato	(Expert-Bezancon *et al.*, 2002)
	39 - 45	Purification de protéines assemblées sur le BP, PY, et le site 3' d'épissage de l'exon 6B	EN de myoblastes QM7	Chromato	(Sauliere *et al.*, 2006)
Un ARN transcrit *in vitro* est oxydé à l'aide de sodium m-prériodate puis fixé sur une matrice	60	hnRNP A/B se fixe sur l'exon 2 de tat de HIV-1	EN de cellules HeLa	Chromato	(Caputi *et al.*, 1999)
	37	Purification de hnRNP se liant à l'exon N1 de c-src	EN de cellules HeLa et WERI1	Chromato	(Markovtsov *et al.*, 2000)
	36	Protéines de la famille hnRNP H activent l'épissage de de *tat* de HIV-1	EN de cellules HeLa	Chromato	(Schaub *et al.*, 2007)

Tableau 3- Ajout d'une structure artificielle sur un ARN par modifications chimiques. Le mode de synthèse des ARN est indiqué, ainsi que la taille des ARN obtenus. Les résultats obtenus en fixant l'ARN sur un support, le type d'extrait nucléaire (EN) utilisé, et l'ordre dans lequel la purification a été effectuée sont également précisés. « RNP » correspond à la purification de complexes déjà formés et « chromato » à l'assemblage des complexes RNP sur l'ARN préalablement immobilisé sur une matrice.

2.2.1.1. Une modification naturelle dans l'ARN

L'étiquette peut correspondre à une modification naturellement présente dans l'ARN. C'est le cas de la coiffe tri-méthyl-guanosine présente à l'extrémité 5' des UsnARN. Cette structure est reconnue par des anticorps spécifiques, et a permis la purification des snRNP par immunoprécipitation (Kastner *et al.*, 1990; Will *et al.*, 1993; Fabrizio *et al.*, 1994).

2.2.1.2. Une modification artificielle dans l'ARN

Il est également possible d'introduire des modifications artificielles au sein de l'ARN. Les modifications les plus couramment utilisées sont récapitulées dans le tableau 3. Une biotine peut être introduite dans l'ARN lors de la transcription *in vitro* (Grabowski et Sharp, 1986). L'ARN ainsi transcrit est biotinylé de manière aléatoire tout au long de sa séquence et peut être fixé sur une matrice de streptavidine. La biotine peut également être ajoutée à l'extrémité 3' de l'ARN par modification chimique en utilisant le métapériodate de sodium. Ce réactif va oxyder les groupements 2' et 3' OH libres du ribose du dernier nucléotide de l'ARN, qui pourra ensuite former une liaison stable avec les groupements hydrazides présents sur un dérivé de la biotine (biotine-hydrazide). L'ARN ainsi biotinylé à son extrémité 3' (Gallego *et al.*, 1997) peut ensuite être fixé à des billes d'agarose couplées à de la streptavidine (Expert-Bezancon *et al.*, 2002). L'ARN traité avec le métapériodate de sodium peut également être directement immobilisé après réaction avec les groupements hydrazides présents sur une matrice (matrice *adipic acid dihydrazide agarose*) (Wilchek et Lamed, 1974; Langland *et al.*, 1995). La liaison entre la streptavidine et la biotine est forte (Kd $\approx 10^{-15}$ M (Green, 1975)) et la liaison hydrazone formée suite à l'oxydation par le métapériodate de sodium est stable. Ces deux types de modifications chimiques permettent donc d'utiliser une large gamme de tampons (plus ou moins salins et contenant ou non du détergent) pour réaliser les purifications tout en limitant les fixations non spécifiques. Le manque de réversibilité de la liaison des ARN sur la matrice peut cependant être un inconvénient puisqu'elle impose une élution des protéiques en conditions dénaturantes, le plus souvent par ajout de tampon de Laemmli (Laemmli, 1970). Préalablement à cette élution, il est donc nécessaire d'effectuer des lavages suffisamment stringents pour éliminer les protéines fixées de manière non spécifiques sur la matrice. Des lavages trop stringents pouvant conduire au décrochage de certaines protéines spécifiques du complexe RNP, il est possible de limiter ce problème en éluant spécifiquement les protéines fixées sur l'ARN à l'aide de RNases.

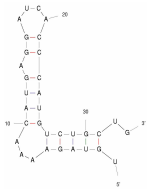
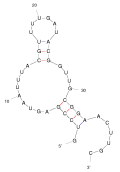
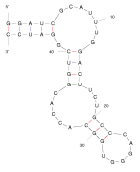
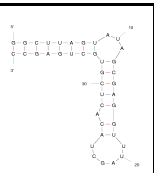
Nom	Synthèse de l'ARN étiqueté	Structure	Taille de l'ARN (nt)	Molécules ciblées	Résultats obtenus	Extraits utilisés	Chromato/RNP	Références
MS2 (3 X 33nt)	ARN d'intérêts synthétisés *in vitro* et contenant en général 3 tag MS2		213	protéine de fusion MS2-MBP	Association entre snRNP U2 et le complexe E	EN de cellules HeLa	RNP	(Das *et al.*, 2000)
			200		Purification de spliceosome natif	EN de cellules HeLa	RNP	(Jurica *et al.*, 2002)
			200		Identification par protéomique des composants d'un spliceosome actif	EN de cellules HeLa	RNP	(Zhou *et al.*, 2002)
Sephadex (40nt)	Transcription *in vitro*		-	matrice sephadex G100 ou G200	Méthode de purification	EN de cellules HeLa	RNP	(Srisawat *et al.*, 2001)
Streptavidine (44nt)	Tag ajouté par clonage, transformation de levures avec cette construction		369	résidus biotine	Etude de l'organisation de la RNase P	Extraits de *Sacaromyces cerevisiae*	RNP	(Srisawat et Engelke, 2001)
Streptomycine (46 nt)	Transcription *in vitro*		23 - 250	ATB spreptomycine	Purification de U1A	Extraits de *Sacaromyces cerevisiae*	RNP	(Bachler *et al.*, 1999)
Tobramycine (40 nt)	Transcription *in vitro*		267	ATB tobramycine	Analyse du pré-spliceosome humain	EN de cellules HeLa	Chromato	(Hartmuth *et al.*, 2002)
			178		Analyse de facteurs interagissant avec la 5' UTR de rpr	EN de cellules HeLa	RNP	(Vazquez-Pianzola *et al.*, 2005)

Tableau 4- Ajout d'une séquence spécifique dans les ARN permettant leur fixation sur une matrice. Le nom de l'étiquette utilisée, la méthode pour l'ajouter à l'ARN d'intérêt ainsi que les tailles des étiquettes et des ARN obtenus, sont indiqués. Les résultats obtenus en fixant l'ARN sur un support, le type d'extrait nucléaire (EN) utilisé, et l'ordre dans lequel la purification a été effectuée sont également précisés. « RNP » correspond à la purification de complexes déjà formés et « chromato » à l'assemblage des complexes RNP sur l'ARN préalablement immobilisé sur une matrice (nt correspond à nucléotides et ATB à antibiotique).

2.2.1.3. Un aptamère ou étiquette dans l'ARN

Les aptamères sont des séquences très structurées présentant une forte affinité pour un ligand immobilisé sur une matrice, et sont utilisés pour étiqueter un ARN. Les différentes étiquettes connues, présentées dans le tableau 4, ont pour la plupart été développées grâce à la méthode SELEX. Les ligands sont de nature très diverse comprenant des antibiotiques, comme la streptomycine (Wallace et Schroeder, 1998) ou la tobramycine (Wang et Rando, 1995), des polymères comme le sephadex G100 ou G200 (Srisawat *et al.*, 2001) ou des protéines telle que la protéine de bactériophage R17 de *E. coli* (Bernardi et Spahr, 1972; Carey *et al.*, 1983; Lowary et Uhlenbeck, 1987; Bardwell et Wickens, 1990). Ce dernier aptamère appelé MS2 n'a pas été sélectionné par SELEX, mais par affinité de liaison à la protéine R17 puis séquençage. La séquence utilisée comme étiquette correspond à trois répétitions de l'aptamère MS2 ce qui augmente l'affinité de l'ARN pour la protéine R17 qui est le plus souvent fusionnée à la protéine de liaison au maltose. Un aptamère de la biotine, appelé « étiquette streptavidine » (Srisawat et Engelke, 2001) a également été décrit. L'affinité des ces étiquettes pour leur cible (entre 1 µM et 1 pM) est plus faible que celle observée pour les modifications artificielles, le choix des tampons permettant l'assemblage et la fixation des complexes sera donc plus limité. Il est également important de s'assurer que la présence de l'étiquette dans l'ARN n'interfère pas avec la mise en place et la stabilité des complexes RNP. L'introduction des séquences aptamères dans les ARN d'intérêts est réalisée lors de la synthèse de matrices ADN utilisées pour les transcriptions *in vitro*. Ces matrices peuvent être produites par PCR en utilisant un oligonucléotide 5' contenant la séquence de fixation de l'ARN polymérase T7 et un oligonucléotide 3' contenant la séquence du tag, ou l'inverse (figure 35).

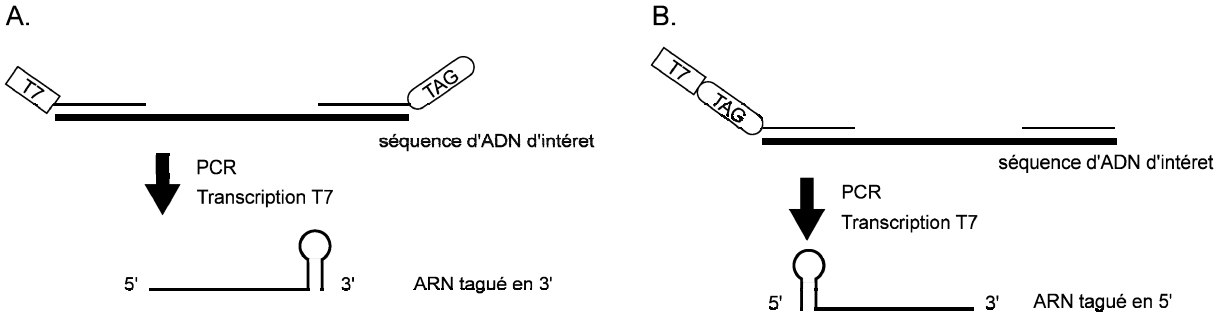


Figure 35- Stratégie d'insertion d'une étiquette ARN par PCR puis transcription *in vitro*. A. L'oligonucléotide sens contient une séquence permettant la transcription par la T7 ARN polymérase ainsi qu'une séquence spécifique de l'ADN considéré. L'oligonucléotide antisens contient une séquence spécifique de l'ADN considéré et une séquence codant pour le tag désiré. **B.** L'oligonucléotide sens contient une séquence permettant la transcription par la T7 ARN polymérase, une séquence codant pour l'étiquette désirée, ainsi qu'une séquence spécifique de l'ADN considéré. L'oligonucléotide antisens contient une séquence spécifique de l'ADN considéré. A. et B. Après la réaction de PCR puis la transcription *in vitro*, on obtient respectivement un ARN étiqueté en 3' et en 5'.

Stratégies	Taille de l'ARN (nt)	Résultats obtenus	Extraits utilisés	Purification de RNP	Références
Purification *via* un oligonucléotide biotinylé	190	Cartographie des interactions entre ARNpré-m et U2snRNP	EN de cellules HeLa	oligo dans EN puis matrice	(Barabino *et al.*, 1989)
	150	Purification de U2 et U4/U6 snRNP	EN de *Trypanosoma brucei*	RNP sur matrice plus oligo	(Palfi *et al.*, 1991)
	62	Purification de U7 snRNP	EN de TB21 mastocytes de souris	oligo dans EN puis matrice	(Smith *et al.*, 1991)
Incorporation d'un oligonucléotide ARN biotinylé dans un ARN	222	Purification des protéines des complexes H et A	EN de cellules HeLa	ARN biotinylé dans EN puis matrice	(Rhode *et al.*, 2003)

Tableau 5- Fixation d'un ARN sur une matrice *via* un oligonucléotide biotinylé. La technique utilisée pour fixer l'ARN est indiquée, ainsi que la taille des ARN utilisés. Les résultats obtenus, en fixant l'ARN sur un support, le type d'extrait nucléaire utilisé, et l'ordre dans lequel la purification a été effectuée sont également précisés. Oligo dans EN puis matrice signifie que l'oligonucléotide antisens a été incubé en même temps que l'ARN en présence d'extrait nucléaire, les complexes alors formés ont été incubés sur une matrice. RNP sur matrice plus oligo signifie que les complexes RNP ont été formés puis incubés sur une matrice sur laquelle est fixée l'oligonucléotide antisens. ARN biotinylé dans EN puis matrice signifie que l'oligonucléotide ARN biotinylé à été incorporé dans l'ARN, les complexes RNP ont ensuite été formés en présence d'extrait nucléaire avant d'être incubés sur la matrice permettant la purification.

L'étiquette peut également être clonée plus classiquement dans un vecteur comportant la séquence d'intérêt. Ce vecteur peut ensuite être injecté dans des cellules en culture où il produira l'ARN étiqueté, ou bien être utilisé dans des réactions de transcription *in vitro*. L'insertion de l'aptamère dans un ARN d'intérêt ne doit pas modifier les structures de ces deux séquences ainsi que l'affinité de l'aptamère pour son ligand. Le premier point peut être vérifié avec des logiciels de prédiction de structure des ARN tels que Mfold (Zuker, 2003). Les complexes doivent rester stables en présence des ligands. Ce deuxième point peut être testé *in vitro* à l'aide d'expériences de retard de migration sur gel. Il est également nécessaire de s'assurer que l'extrait nucléaire utilisé ne contient pas naturellement des facteurs susceptibles d'interagir avec l'aptamère. Enfin, pour toutes les étiquettes, il est important de vérifier que seul l'ARN étiqueté est retenu par la matrice d'intérêt. Quel que soit l'étiquette utilisée, il est possible de purifier des complexes pré-formés ou bien assemblés sur l'ARN immobilisé sur un support contenant un ligand greffé. L'utilisation de ces étiquettes permet de réaliser une élution des complexes en condition native en utilisant le ligand en solution (antibiotique, dextran, maltose, biotine, etc.).

2.2.1.4. Fixation de l'ARN via un oligonucléotide biotinylé

Il est également possible de purifier des complexes RNP *via* l'ARN sans modifier ce dernier, ni chimiquement ni par ajout d'un aptamère en 5' ou en 3'. Cette stratégie basée sur l'utilisation d'oligonucléotides antisens modifiés et biotinylés a surtout été employée pour la purification de complexes RNP stables comme les snRNP (Barabino *et al.*, 1989 ; Smith *et al.*, 1991 ; Palfi *et al.*, 1991). Ces différentes expériences sont présentées dans le tableau 5. Les oligonucléotides utilisés sont généralement 2'O-méthylés afin que l'hétérodimère ARN/ADN ne soit pas dégradé par la RNase H qui est souvent présente dans les extraits nucléaires (Barabino *et al.*, 1989). La mise en place de cette stratégie nécessite l'existence dans l'ARN d'une région qui reste accessible à l'oligonucléotide en présence des protéines de l'extrait nucléaire. Cependant, il est indispensable que la fixation de l'oligonucléotide ne change pas la stabilité du complexe ni la structure de l'ARN.

2.2.2. Choix de la stratégie utilisée

Notre objectif est d'identifier les protéines assemblées autour de la séquence UTE. Il est théoriquement possible de purifier des complexes RNP dans des conditions « physiologiques ». Dans notre cas, une telle stratégie nécessite de placer l'ARN étiqueté dans un contexte cellulaire activateur ou répresseur de l'épissage de l'exon 9A9' puis de purifier les complexes préformés sans modifier leur composition. Réaliser une telle purification *in vivo* semble ardue. En effet, de nombreux travaux mettent en évidence la dynamique très complexe de la réaction d'épissage et de sa régulation. Pour purifier les complexes mis en place au cours de l'une des étapes d'épissage, il faudrait pouvoir bloquer la réaction à l'étape choisie afin d'accumuler les complexes dans un état équivalent. Une autre limite à la purification des complexes *in vivo* est due au fait que nous nous intéressons aux facteurs assemblés sur des séquences introniques dont la durée de vie est faible. Devant ces difficultés nous avons aussi la possibilité d'effectuer les purifications *in vitro*, avec des extraits nucléaires qui reproduisent la répression ou l'activation de l'épissage de l'exon 9A9'. Les protéines identifiées par une telle approche *in vitro*, devront ensuite être validées fonctionnellement dans l'embryon de xénope.

Nous avons tout d'abord pensé utiliser un ARN étiqueté avec un aptamère pour purifier les complexes préformés dans l'extrait nucléaire. Afin de déterminer quel est en théorie le « meilleur » aptamère à utiliser, il est nécessaire de prendre en compte la taille et la structure aptamères eux-mêmes, ainsi que celle des ARN auxquels ils sont ajoutés. Plus la séquence de l'étiquette est longue, plus celle-ci a des chances de former une structure secondaire avec l'ARN dans lequel elle est ajoutée. Les étiquettes tobramycine, sephadex, streptavidine et streptomycine ont une longueur comprise entre 40 et 46 nucléotides. L'étiquette MS2 comprend quant à elle, trois répétitions de l'aptamère de 33 nucléotides. L'étude bibliographique sur l'utilisation des différentes étiquettes met en évidence un nombre plus important d'analyses réalisées avec l'étiquette MS2 pour purifier des complexes RNP. Cette étiquette est intéressante puisqu'elle permet l'assemblage puis la fixation des complexes à pH neutre. Mais l'analyse de la structure secondaire de différentes séquences d'ARN comprenant les éléments régulateurs et étiquetés en 5' ou en 3' à l'aide du logiciel Mfold (Zuker, 2003) montre que seuls les étiquettes tobramycine et sephadex se replient correctement quand elles sont localisées à l'extrémité 3' des ARN que nous souhaitons utiliser. Lorsqu'elle est placée en 5' de ces ARN, aucune des étiquettes testées n'adopte de

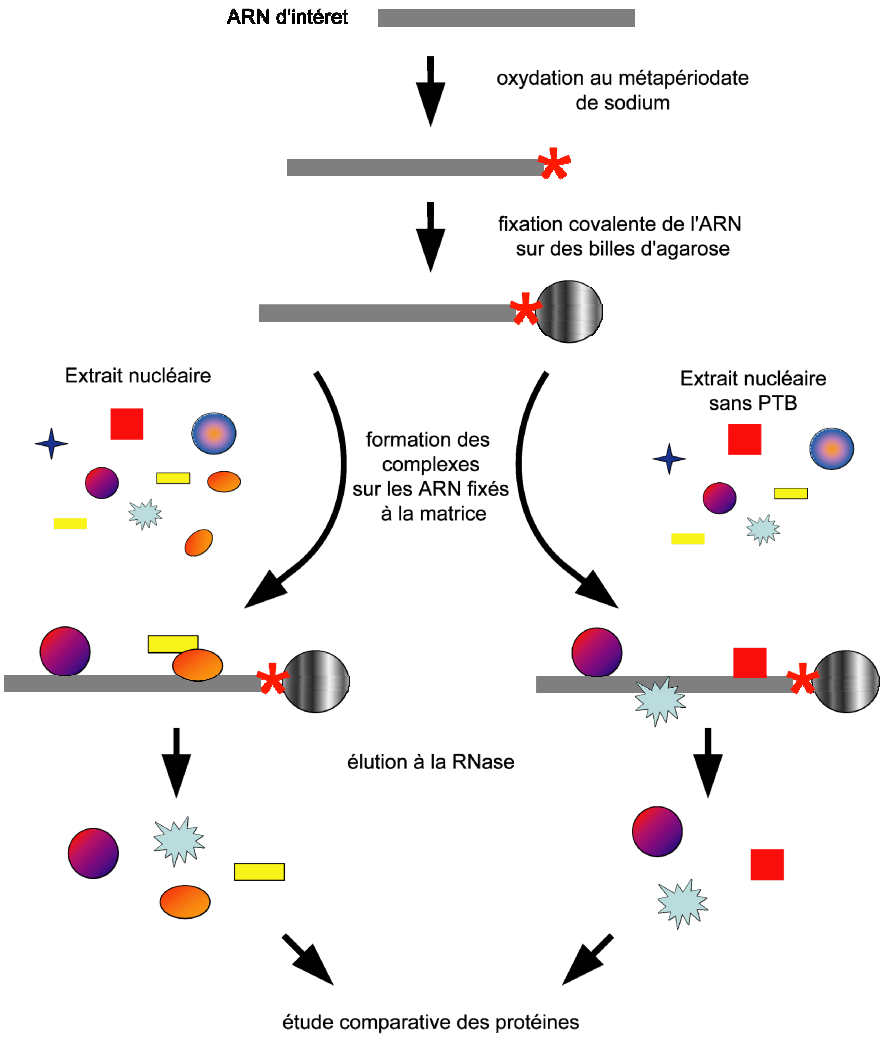


Figure 36- Purification des complexes suite à la fixation d'un ARN sur une matrice à l'aide du métapériodate de sodium. Une fois la production d'extraits nucléaires fonctionnels pour l'épissage et le choix des ARN effectués, les ARN d'intérêts sont liés de manière covalente sur une matrice d'agarose puis incubés en présence des différents extraits. Les protéines des complexes alors formés sont éluées par traitement à la RNase puis analysées.

structure secondaire correcte. Pour choisir entre les étiquettes tobramycine et sephadex nous nous sommes basés sur une étude comparative de l'utilisation des étiquettes MS2 et tobramycine pour purifier le complexe B du spliceosome qui montre qu'il n'y a pas de différence significative entre les protéines ainsi identifiées (Deckert *et al.*, 2006). Etant donné que peu de purifications de complexes à l'aide de l'étiquette sephadex ont été décrites (Srisawat et Engelke, 2002; Walker *et al.*, 2008), l'étiquette tobramycine semblait donc être la plus adaptée à notre problématique. Nous avons testé son utilisation pour purifier des complexes assemblés sur un ARN dans différents extraits nucléaires. Dans nos mains, l'efficacité de fixation des ARN étiquetés tobramycine sur une matrice activée est inférieure à 50 %. De plus, l'interaction entre l'étiquette et l'aminoglycoside a lieu à pH basique (8,1), affectant ainsi la stabilité de notre ARN d'intérêt.

Les résultats de ces expériences nous ont conduit à nous intéresser à une stratégie consistant à fixer l'ARN d'intérêt sur une matrice puis à l'incuber en présence d'extraits nucléaires contenant ou non la PTB. Ces conditions permettent de déterminer de manière simple quelles sont les protéines recrutées sur un fragment d'ARN. Plusieurs modifications de l'ARN sont possibles. Pour choisir parmi les techniques déjà utilisées, nous nous sommes basés sur le fait que moins l'ARN est modifié, plus le complexe assemblé sera proche de celui qui se forme réellement autour de la séquence d'intérêt. La méthode la plus simple, ne modifiant l'ARN qu'à son extrémité 3', comportant le moins d'étapes, et la moins coûteuse, est celle utilisant le métapériodate de sodium pour permettre la fixation de l'ARN sur une matrice (Caputi *et al.*, 1999). Les différentes étapes de cette stratégie sont présentées dans la figure 36.

2.3. Mise en œuvre de la purification

2.3.1. Choix de l'ARN utilisé pour la purification des complexes

L'élément UTE identifié par mutagenèse (Anquetil *et al.*) s'étend des nucléotides -204 à -165 dans l'intron en amont de l'exon 9A9'. Il comporte un motif UGGAUGG conservé entre le xénope, le poulet, le rat et l'humain dont la mutation provoque une diminution de 2,6 fois de la quantité d'ARN contenant l'exon terminal 9A9' (ARN α7). L'analyse de la séquence UTE par le logiciel ESEfinder (Cartegni *et al.*, 2003) révèle trois motifs de liaison

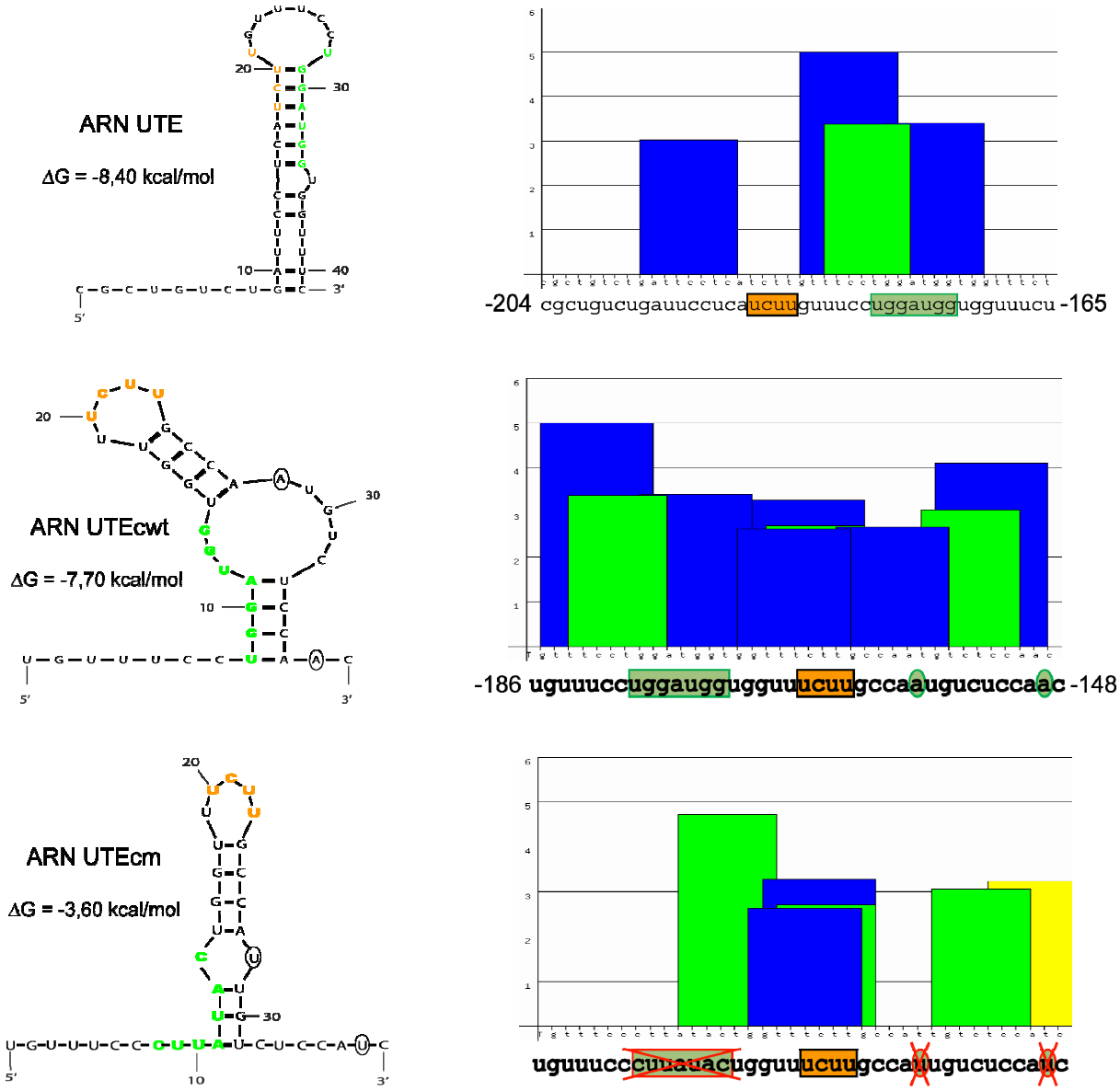


Figure 37- Prédictions de structure et des sites de liaison de protéines SR des ARN UTEc. La structure des ARN a été prédite à l'aide du logiciel Mfold, et des sites de liaisons potentiels des protéines SR ASF/SF2 (en rouge), SC35 (en bleu), SRp40 (en vert) et SRp55 (en jaune) ont été recherchés à l'aide du programme ESEfinder dans les UTE court (UTEc) mutés (m) ou non (wt) au niveau de la séquence UTE. La séquence de chaque ARN est indiquée en abscisse. Les scores obtenus suite à la comparaison de la séquence avec les sites de liaison consensuels définis pour chacune des protéines SR sont représentés en ordonnée.

de la protéine SC35 et un motif de liaison de la protéine SRp40 centrés autour de la séquence UGGAUGG (figure 37). Une nouvelle étude réalisée sur une région intronique plus large a décelé une région de 37 nucléotides présentant une forte densité de motifs de liaison des protéines SR SC35, SRp40 et SRp55 (figure 37). Cette région s'étend des nucléotides -186 à -128 et encadre de nouveau le motif UGGAUGG. Elle a été appelée UTEc (court). L'analyse de la structure de cet ARN à l'aide du logiciel Mfold montre que la séquence UGGAUGG est plus accessible dans cet ARN que dans l'ARN correspondant à la séquence UTE initiale (figure 37). Nous avons donc choisi de rechercher les protéines qui s'assemblent autour du motif UGGAUGG en utilisant l'ARN UTEc comme appât. Afin d'identifier les protéines qui s'assemblent spécifiquement sur le motif UGGAUGG, nous avons également produit un ARN UTEc dans lequel ce motif est muté. Cet ARN appelé UTEcm comporte également les mutations A -158 et A -149 en T, qui ont été introduites lors de la mutagenèse des minigènes α tropomyosine. Ces deux mutations font disparaître deux motifs de liaison à SC35 (figure 37) mais n'affectent pas l'épissage des ARNpré-m issus des minigènes (Hamon Gouault, 2002).

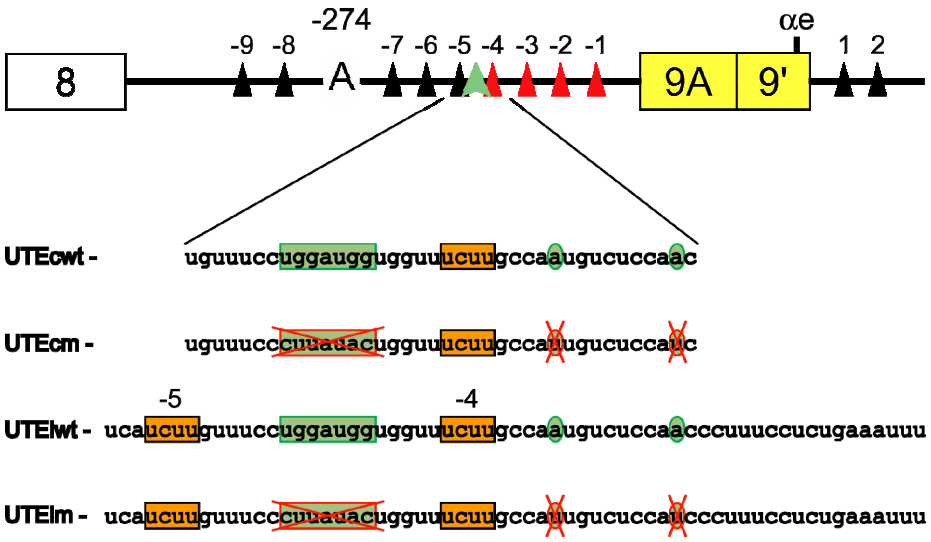


Figure 38- Des sites de liaison de PTB sont présents autour de l'élément activateur UTE. Haut, représentation schématique du fragment d'ARNpré-m couvrant la séquence allant de l'exon 8 au début de l'intron 9A9'-9B. Les sites PTB sont indiqués par des triangles rouges s'ils font partie de la séquence 150PY ou noir. La séquence UTE est indiquée par un triangle vert et le point de branchement en -274 est indiqué par un 'A'. Bas, les séquences des quatre ARN UTE sont indiquées. Les sites de fixation de la PTB sont indiqués en orange. Les éléments de séquence qui constituent l'UTE sont indiqués en vert, quand ces séquences sont mutées elles sont barrées.

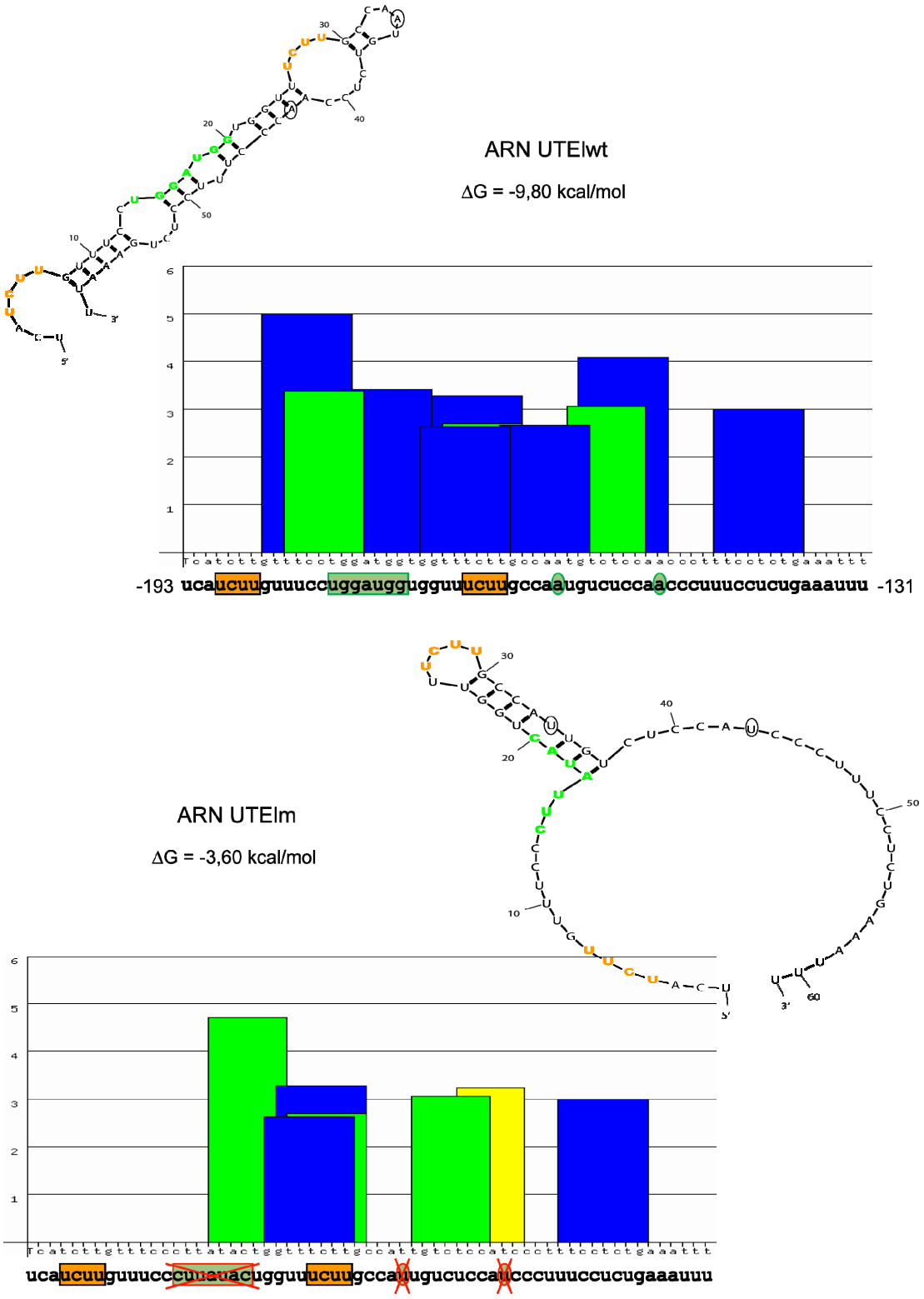


Figure 39- Prédictions de structure et des sites de liaison de protéines SR des ARN UTEl. La structure des ARN a été prédite à l'aide du logiciel Mfold, et des sites de liaisons potentiels des protéines SR ASF/SF2 (en rouge), SC35 (en bleu), SRp40 (en vert) et SRp55 (en jaune) ont été recherchés à l'aide du programme ESEfinder dans les UTE long (UTEl) mutés (m) ou non (wt) au niveau de la séquence UTE. La séquence de chaque ARN est indiquée en abscisse. Les scores obtenus suite à la comparaison de la séquence avec les sites de liaison consensuels définis pour chacune des protéines SR sont représentés en ordonnée.

Pour étudier l'influence de la fixation de la PTB sur l'assemblage des complexes, nous avons produit un second ARN incluant un site de fixation à la PTB avoisinant. Cet ARN devait avoir un nombre de site de liaison à la PTB augmenté par rapport à l'ARN UTEcwt tout en conservant un nombre similaire de sites putatifs de liaison à SC35. Nous avons pris soin de ne pas inclure le site A-196 qui correspond à un motif de liaison à SC35 et dont la mutation semble renforcer l'effet de la mutation du motif UGGAUGG (Anquetil *et al.*). Nous avons donc choisi d'inclure le site PTB situé en 5' du motif conservé (figure 38) et d'étendre la séquence en 3' jusqu'au nucléotide T-131 afin que la structure du motif UGGAUGG soit conservée dans ces ARN plus longs. La structure de ce nouvel ARN appelé UTElwt est présentée dans la figure 39. Cet ARN est structuré en tige boucle avec une stabilité équivalente à celle de l'ARN UTEcwt et contient les deux sites d'interaction avec la PTB dans des régions accessibles. Il contient un motif de liaison à SC35 supplémentaire mais les délétions de cette région dans l'intron sont sans effet sur l'utilisation de l'exon (Hamon Gouault, 2002). La même extension de 17 nucléotides en 3' nous a permis de produire une version plus longue de l'ARN UTEcm que nous avons appelé UTElm. Les prédictions des structures de ces deux derniers ARN sont équivalentes (figure 39).

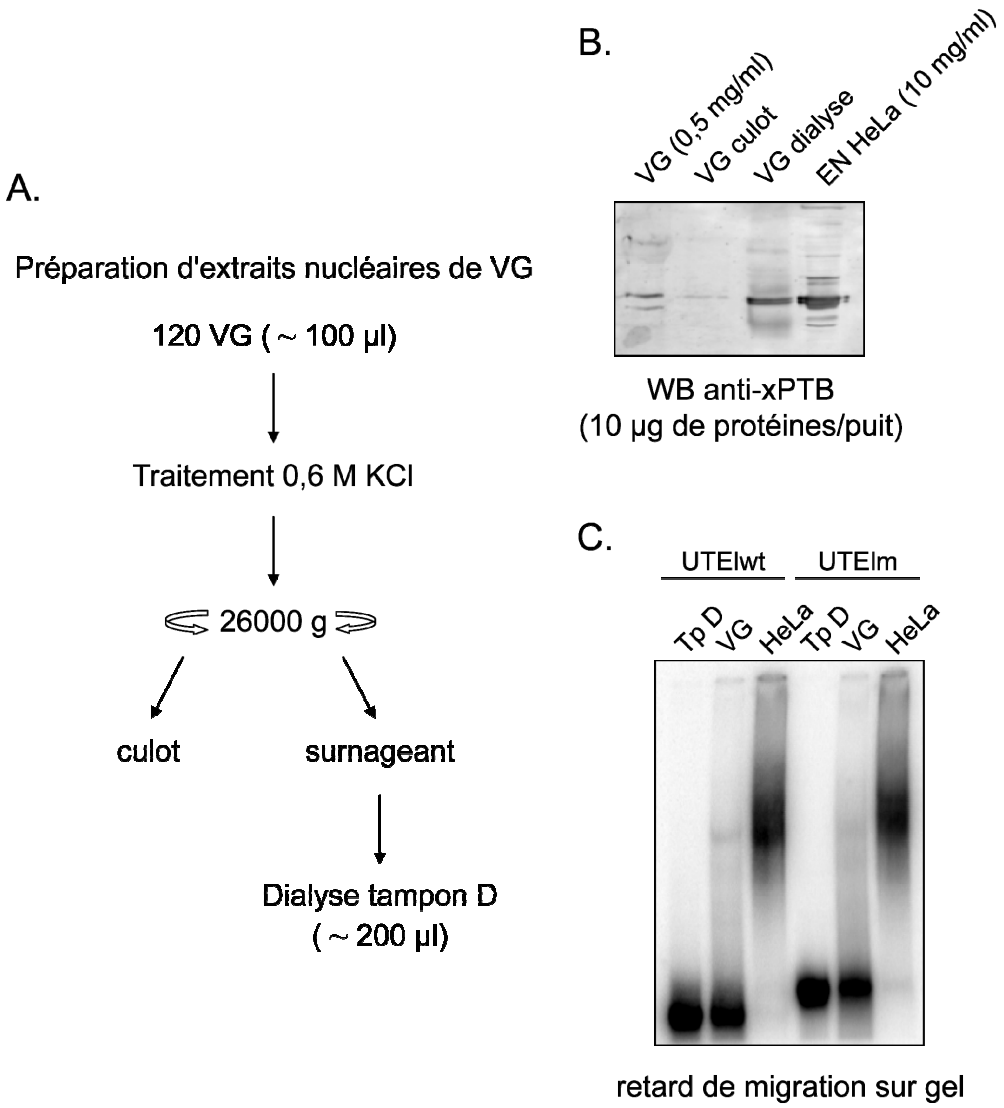
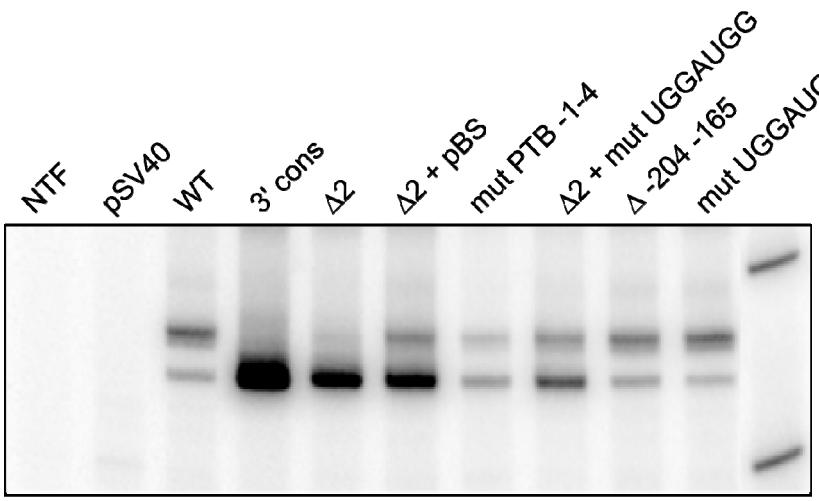


Figure 40- Préparation d'extrait nucléaire à partir de vésicules germinatives d'ovocytes de xénope. A. Principe expérimental correspondant au protocole de préparation d'extrait nucléaire décrite par Dignam et collaborateurs (Dignam *et al.*, 1983). **B.** Le contenu en PTB des échantillons produits au cours de la préparation est comparé à celui d'extraits nucléaires de cellules HeLa. **C.** Comparaison de la formation de complexes sur les ARN UTElwt et UTElm radiomarqués à partir d'extraits nucléaires de VG (10 µl à 1 µg/µl) ou de cellules HeLa (1 µl à 10 µg/µl). Les complexes sont séparés sur gel natif 8 % polyacrylamide/5% glycérol/0,5X TBE, puis révélés au Storm.

2.3.2. Choix des extraits nucléaires

2.3.2.1. Les extraits nucléaires d'ovocytes

L'utilisation de l'exon 9A9' est réprimée dans les ovocytes de xénope. Cependant, la diminution du taux de PTB libre dans l'ovocyte par la présence d'ARN compétiteurs sur lesquels la protéine PTB se lie avec une forte affinité, active l'utilisation de l'exon 9A9' (Hamon *et al.*, 2004). Cette activation est dépendante de la présence de la séquence UTE (Anquetil *et al.*). Le noyau des ovocytes contient donc les facteurs nécessaires à l'activité de l'élément UTE lorsque le taux de protéine PTB libre est diminué. Les extraits nucléaires d'ovocytes sont donc apparus comme un matériel de choix pour notre étude à condition de mettre au point une méthode de préparation des extraits qui permette de conserver tous les facteurs requis pour les réactions d'épissage sous une forme active. Le premier protocole que nous avons utilisé est directement adapté de la procédure décrite par Dignam et collaborateurs (Dignam *et al.*, 1983) pour la préparation d'extraits nucléaires à partir de cellules HeLa. Ce protocole est largement utilisé pour préparer des extraits nucléaires permettant de réaliser des réactions d'épissage *in vitro*. A partir d'environ 120 vésicules germinatives (VG) isolées manuellement, nous avons obtenu 200 µl d'extrait nucléaire dont la concentration protéique était 20 fois inférieure à celle des extraits nucléaires de cellules HeLa préparés dans les mêmes conditions à partir de 10^7 cellules (figure 40 A). Une expérience de western-blot montre que la quantité de protéine PTB détectée dans les extraits issus d'ovocytes semble être moins importante que dans les cellules HeLa (figure 40 B.). Nous avons ensuite testé si les protéines présentes dans les extraits nucléaires étaient capables d'interagir avec un fragment d'ARN contenant la séquence UTE sauvage ou mutée. Une nouvelle préparation d'extrait nucléaire, réalisée à partir de 200 VG nous a permis d'obtenir environ 200 µl d'extrait nucléaire ayant une concentration protéique de 1 mg/ml (soit une concentration 10 fois inférieure à celle des extraits nucléaires de cellules HeLa). Les expériences de retard de migration sur gel présentées dans la figure 40 C ont été réalisées avec 10 µl d'extrait nucléaire d'ovocytes ou 1 µl d'extrait nucléaire de cellules HeLa. Cette expérience montre très clairement que l'assemblage des complexes autour des ARN est très peu efficace dans les extraits nucléaires d'ovocyte. Les extraits nucléaires d'ovocytes dont nous disposons ne sont donc pas exploitables pour purifier les complexes assemblés autour des séquences UTE en raison de leur faible concentration en facteurs protéiques.



%	1	2	3	4	5	6	7	8	9	10
α2Δ9A	0	0	74	ND	5	10	37	43	69	80
α7	0	0	26	99	95	90	63	57	31	20

Promoteur SV40
Cellules HeLa

%	1	2	3	4	5	6	7	8	9	10
α2Δ9A	0	0	90	-	9	-	42	51	-	-
α7	0	0	10	-	91	-	58	49	-	-

Promoteur SV40
ovocytes de xénope

%	1	2	3	4	5	6	7	8	9	10
α2Δ9A	0	0	80	ND	ND	23	34	-	80	85
α7	0	0	20	100	100	77	66	-	20	15

Promoteur kératine
embryons de xénope

Figure 41- L'épissage des minigènes est conservé dans les cellules HeLa. Différents minigènes sous le contrôle du promoteur SV40 ont été transfectés dans Les cellules HeLa. Les transcrits produits ont été analysés par RT-PCR radioactive. Les ADNc sont déposés sur un gel 4% polyacrylamide, visualisés et quantifié à l'aide d'un PhosphorImager Les valeurs qui ne sont pas quantifiables sont notées ND, quand l'expérience n'a pas été effectuée aucune valeur n'est indiqué par. Les pourcentages donnés pour les minigènes injectés dans les ovocytes et dans les embryons de xénope proviennent d'expériences précédemment réalisées au laboratoire. Les cellules non transfectées sont notées NTF, pBS correspond à l'ajout d'un fragment de 150 pb (paires de bases) provenant du vecteur pBS, 3' cons correspond au renforcement du site 3' d'épissage de l'exon 9A9', et Δ2 à l'élément répresseur 150PY.

2.3.2.2. Les extraits nucléaires de cellules HeLa

Un des systèmes de choix pour étudier la régulation de l'épissage *in vitro* est celui des extraits nucléaires de cellules HeLa dans lesquels il est possible d'effectuer l'épissage d'ARNpré-m d'origine diverse. Ces extraits nucléaires pourront être utilisés pour notre étude si la régulation de l'épissage de l'ARNpré-m α-TM est conservée dans ce système hétérologue. La nature des cellules HeLa laisse supposer que dans ces cellules, la régulation de l'épissage de l'exon 9A9' serait identique à celle observée dans les tissus non musculaires de l'embryon, avec une répression dépendante de l'interaction de la protéine PTB avec la séquence intronique 150PY, riche en motifs de haute affinité pour la PTB (Hamon *et al.*, 2004), et une activation dépendante de la séquence UTE.

Afin de caractériser la régulation de l'épissage de l'ARNpré-m α-TM dans les cellules HeLa, j'ai tiré profit du fait que les minigènes SV40-α-TM (>dans le chapitre 1<) pouvaient être utilisés pour la réalisation de transfection dans les cellules HeLa. L'analyse par RT-PCR des produits d'épissage dérivés des minigènes dans les cellules HeLa peut donc être directement comparée à celle réalisée dans l'ovocyte et dans les tissus non musculaires de l'embryon. J'ai utilisé différents minigènes modifiés ou non au niveau de plusieurs sites (site 3' d'épissage, élément inhibiteur 150PY, point de branchement, élément UTE) et analysé par RT-PCR les produits d'épissage issus de ces minigènes (figure 41). L'analyse de nos résultats nous a permis de répondre à plusieurs questions montrant ainsi que les cellules HeLa sont un système d'étude adapté à notre problématique.

Premièrement, le profil d'épissage du minigène sauvage dans les cellules HeLa est identique à celui observé dans les tissus non musculaires de l'embryon de Xénope (piste 3). Les cellules HeLa constituent donc un environnement cellulaire où l'utilisation de l'exon 9A9' est réprimée. Le renforcement du site 3' d'épissage de cet exon conduit à l'expression très majoritaire de l'ARNm α7 conformément à ce qui est observé dans l'ovocyte et les tissus non musculaires de l'embryon (piste 4).

Deuxièmement, nous pouvons conclure que la répression de l'exon 9A9' est dépendante de l'interaction de la protéine PTB avec l'intron 8. En effet, la délétion de l'élément inhibiteur 150PY (noté Δ2 dans la figure 41) entraîne une importante modification de la maturation des ARN issus du minigène vers la production majoritaire de l'ARNm α7 (piste 5). Cet effet n'est pas dû au raccourcissement de l'intron suite à la délétion puisque le remplacement de cette délétion par une séquence hétérologue provenant du vecteur pBS

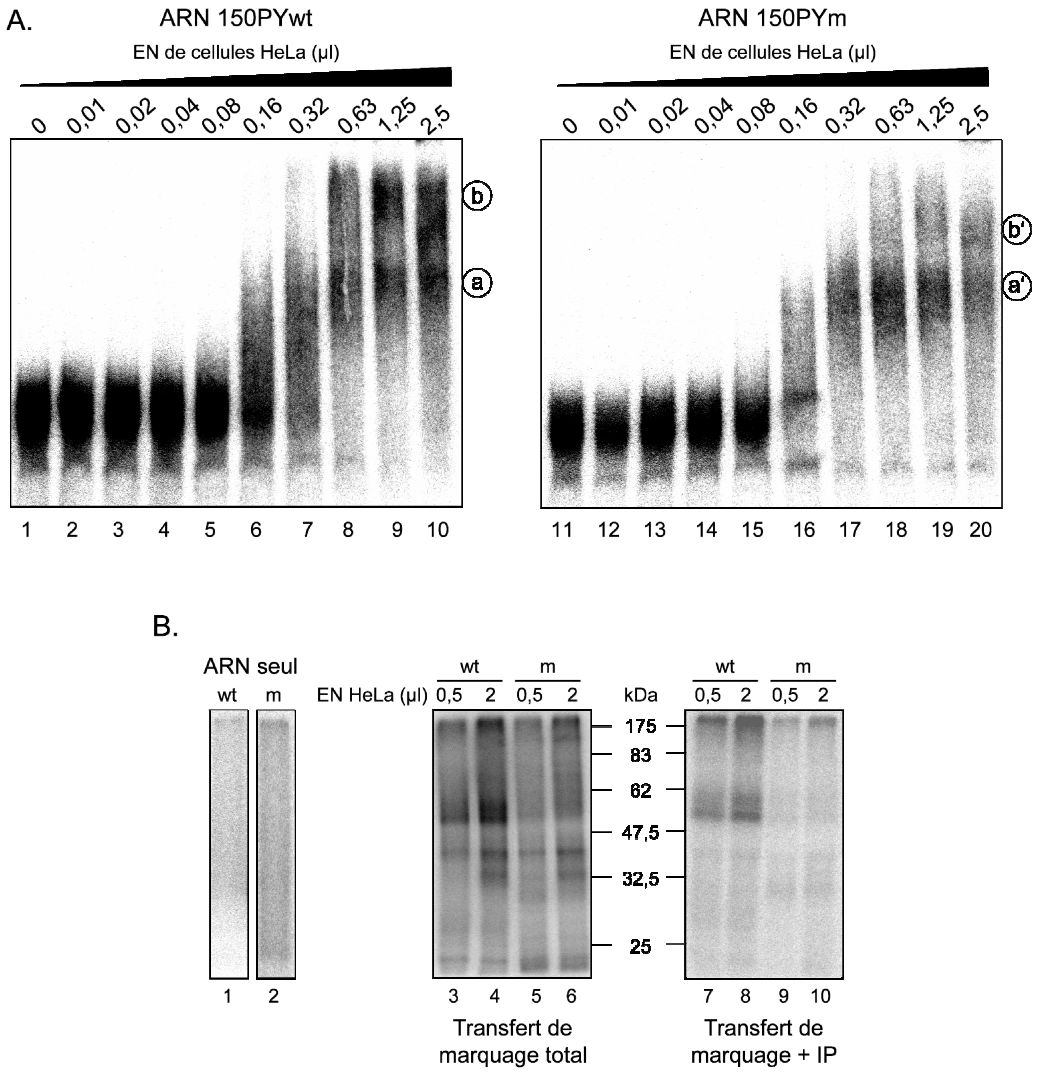


Figure 42- La fixation de la PTB sur la séquence 150PY est dépendante des quatre sites de fixation. A. Les ARN radiomarqués 150PY muté (m) ou non (wt) au niveau des quatre sites de liaison de la protéine PTB sont incubés en présence de quantité croissante d'extrait nucléaire de cellule HeLa. Les complexes sont séparés sur gel natif 8 % polyacrylamide/5% glycérol/0,5X TBE, puis révélés au Storm. B. 25fmol d'ARN 150PYwt (wt) ou muté au niveau des quatre sites de liaison à la PTB (m) et radiomarqués, sont incubés en présence de 0,5 ou 2 µl d'extrait nucléaire de cellules HeLa et irradiés aux UV (0,8J/cm2) à une distance de 10cm pendant 10min. Après traitement à la RNase T1, une fraction des protéines de l'extrait nucléaire est séparée sur gel SDS-PAGE 10% et les protéines radioactives sont visualisées au Storm. L'autre fraction des protéines est soumise à une immunoprécipitation (IP) par l'anticorps anti-PTB. Les protéines immunoprécipitées et radioactives sont visualisées au Storm.

(150pBS) entraîne également une forte utilisation de l'exon 9A9' (piste 6). Des résultats identiques ont été obtenus précédemment en utilisant ces minigènes exprimés dans les tissus non musculaires de l'embryon (Hamon *et al.*, 2004). D'autre part, nous observons que la mutation des quatre motifs de liaison à la PTB, présents dans la région 150PY, entraine une forte augmentation de la production des ARN de type α7 (piste 7). La même observation a été faite à partir du minigène exprimé dans les tissus non musculaires de l'embryon (Hamon *et al.*, 2004). Ces résultats indiquent que les quatre motifs de liaison à la PTB présents dans l'élément répresseur participent également à la répression de l'exon 9A9' dans les cellules HeLa, suggérant que la perte d'interaction de la protéine PTB avec l'intron permet l'activation de l'exon 9A9' et que tous les facteurs nécessaires cette activation sont présents dans les cellules HeLa.

Nous avons vérifié si l'interaction de la protéine PTB avec l'élément 150PY dans les extraits nucléaires de cellules HeLa était dépendante de la présence des quatre motifs de liaison. La formation de complexes ribonucléoprotéiques entre les protéines de l'extrait nucléaire de cellules HeLa et les ARN 150PYwt et 150PYm a donc été étudiée à l'aide d'expériences de retard de migration sur gel, suivie d'expériences de transfert de marquage et d'immunoprécipitation avec des anticorps dirigés contre la protéine PTB. Les ARN ont été incubés en présence de quantités croissantes d'extrait nucléaire et les complexes formés ont été analysés sur gel d'acrylamide en conditions natives (figure 42 A). Les deux ARN forment un complexe présentant une migration identique sur gel (« a » et « a' »), mais seul l'ARN 150PYwt forme un complexe de plus haut poids moléculaire (« b »). Les complexes assemblés autour des deux ARN ont été irradiés aux UV (256nm) pendant 10 min puis hydrolysés par la RNase T1 et immunoprécipités à l'aide d'anticorps anti-PTB (figure 42 B). Cette expérience montre que seuls les ARN sauvages sont capables de fixer la PTB contenue dans l'extrait nucléaire et que les quatre motifs UCUU sont nécessaires à la fixation de la PTB. Une étude de cartographie d'ARN en milieu liquide (Anquetil, 2005) a montré que la mutation des sites UCUU en CCCC n'affecte pas la structure secondaire de l'ARN. Le changement de séquence UCUU en CCCC est donc bien à l'origine de la disparition des sites de liaison de la PTB.

Troisièmement nous avons pu observer que l'activation est dépendante de l'UTE. Comme nous venons de le décrire, l'activation de l'exon 9A9' peut être obtenu par délétion (Δ2) de l'élément inhibiteur (figure 41, piste 5). L'addition de la mutation de l'UTE à cette délétion provoque une baisse de cette activation (piste 8) ce qui suggère que cette activation est en partie dépendante de la présence de l'UTE. La mutation de la séquence UTE (piste 10

et 11) ou sa délétion (piste 9) n'ont qu'un effet très modéré sur la maturation de l'ARNpré-m. Une diminution du taux endogène de protéine PTB par l'utilisation de siARN aurait probablement pu nous permettre d'obtenir des effets plus importants de ces mutations sur la maturation du minigène. Les mêmes observations ont été faites lorsque l'effet de ces mutations a été analysé dans les tissus non musculaires de l'embryon (Anquetil *et al.*) (Hamon Gouault, 2002). Dans un contexte cellulaire riche en protéine PTB, l'épissage est fortement réprimé et de nombreux sites de liaison de la PTB sont présents dans l'intron 8 au voisinage de l'UTE, du site 3' d'épissage et du point de branchement. La fonction de l'UTE n'a pu être analysée que dans un environnement cellulaire où l'exon 9A9' est activé, comme les somites de l'embryon. Dans un contexte non musculaire, son étude nécessite une diminution de la fixation de PTB sur l'intron 8. Elle a pu être obtenue dans les ovocytes en utilisant un ARN compétiteur qui séquestre la protéine (Anquetil *et al.*). Dans les cellules non musculaires de l'embryon, la fonction de l'UTE a pu être étudiée en utilisant des morpholinos oligonucléotides bloquant la traduction de la protéine (Anquetil *et al.*).

L'ensemble de ces résultats nous permet de conclure que les extraits nucléaires de cellules HeLa constituent un environnement cellulaire où l'exon 9A9' est réprimé, que la modulation de l'interaction de la protéine PTB avec l'intron 8 conduit à une levée partielle de cette répression qui dépend en partie de l'UTE. La déplétion de la protéine PTB des extraits nucléaires de cellules HeLa devrait être suffisante pour lever l'inhibition de l'exon 9A9'. Nous avons donc mis en place deux stratégies permettant la déplétion de cette protéine sans affecter les concentrations des autres protéines et les fonctions d'épissage des extraits nucléaires.

2.3.3. Obtention d'extraits nucléaires de cellules HeLa dépourvus de PTB

La première stratégie utilisée est une déplétion de la protéine après passage de l'extrait sur une colonne d'affinité sur laquelle est greffé un ARN qui interagit très fortement avec la protéine (Wollerton *et al.*, 2001; Sauliere *et al.*, 2006). La deuxième stratégie consiste à réaliser une immunodéplétion de l'extrait avec des anticorps commerciaux dirigés contre la protéine.

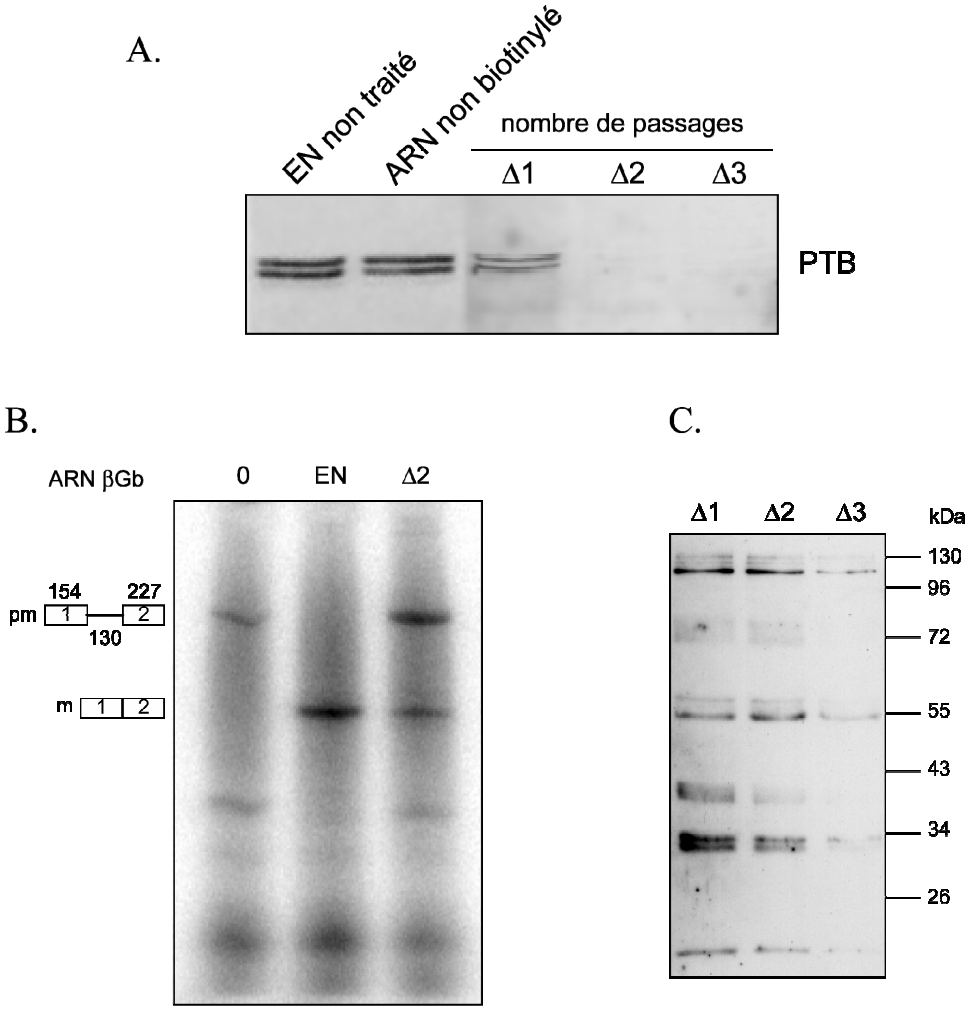


Figure 43- La déplétion en PTB des extraits nucléaires de cellules HeLa par chromatographie d'affinité à l'ARN conduit à une diminution de l'épissage de l'ARNpré-m rapporteur de la β-globine. Les EN de cellules HeLa ont été incubés en présence d'une matrice de streptavidine couplée à l'ARN $(CUCUU)_8$ biotinylés. Le numéro correspond au nombre de passages sur la matrice. La piste ARN non biotinylé correspond à une matrice de streptavidine incubée en présence d'ARN $(CUCUU)_8$ non biotinylé. **A.** L'analyse par western blot de PTB (anticorps α-PTB) montre une déplétion de 50% pour Δ1 et de 100% pour Δ2 de la PTB (quantification réalisée sur les signaux de fluorescence des anticorps secondaires révélés avec le système Osyssey). NB- Le western-blot présenté ici est issu de la même membrane de nitrocellulose, les pistes ont été rapprochées afin de ne présenter que les échantillons d'intérêt. **B.** L'ARNpré-m radiomarqué de la β-globine a été incubé deux heures à 30°C dans solution d'épissage contenant 40% d'extrait nucléaire de cellules HeLa soit non traité (EN), soit déplété en PTB (Δ2). Les signaux correspondant aux ARNpré-m et aux ARNm ont été quantifiés au PhosphorImager. Une diminution de 65% de l'efficacité d'épissage est observée entre les extraits EN et Δ2. La taille des exons et de l'intron de la β-globine est indiquée. **C.** Une diminution proportionnelle au nombre de passages sur les billes est observée pour les protéines SR, suite à la révélation de l'anticorps 10H3 par l'ECL (Pierce).

2.3.3.1. Déplétion de la PTB par chromatographie sur ARN immobilisés

L'ARN $(CUCUU)_8$ interagit très fortement avec la protéine PTB (Southby *et al.*, 1999). Cet ARN a été biotinylé puis fixé sur des billes magnétiques de streptavidine selon le protocole décrit par Saulière et collaborateurs (Sauliere *et al.*, 2006). Les extraits nucléaires de cellules HeLa ont été incubés avec ces billes puis récupérés, et leur contenu en PTB a été analysé par western-blot. L'efficacité de déplétion est de 50 % après un premier passage sur les billes, et de 100% suite au deuxième passage (figure 43 A. Δ1 et Δ2). Afin de s'assurer que la déplétion ne touchait que la PTB et pas d'autres facteurs nucléaires, nous avons comparé les capacités des extraits traités ou non avec les billes à réaliser l'épissage de l'ARNpré-m de la β-globine. Comme présenté dans la figure 43 B, la déplétion de PTB par cette technique diminue de trois fois la quantité d'ARNm β-globine produit. Cet effet pourrait être du à la déplétion concomitante d'autres protéines impliquées dans l'épissage comme les protéines U2AF65 qui ont une affinité pour les séquences riches en pyrimidines ou la co-déplétion non spécifique de protéines SR. Etant donné que les résultats que nous avons obtenus dans les ovocytes de xénope montrent que les protéines SR sont impliquées dans l'activation de l'exon9A9', il nous a semblé indispensable de vérifier que cette technique de déplétion n'affectait pas le taux de ces protéines dans les extraits nucléaires. Les western-blot réalisées sur les trois extraits déplétés en PTB montrent une diminution globale de la quantité de protéines SR. Cette baisse est proportionnelle au nombre de passages sur les billes fixant l'ARN $(CUCUU)_8$ (figure 43 C. pistes Δ1 à Δ3). Cette stratégie permet donc de diminuer la concentration de PTB dans les extraits nucléaires, mais cette déplétion n'est pas complètement spécifique de la PTB puisque d'autres facteurs impliqués dans la régulation de l'épissage sont également partiellement déplétés. Cette stratégie semble donc peu adaptée pour notre étude.

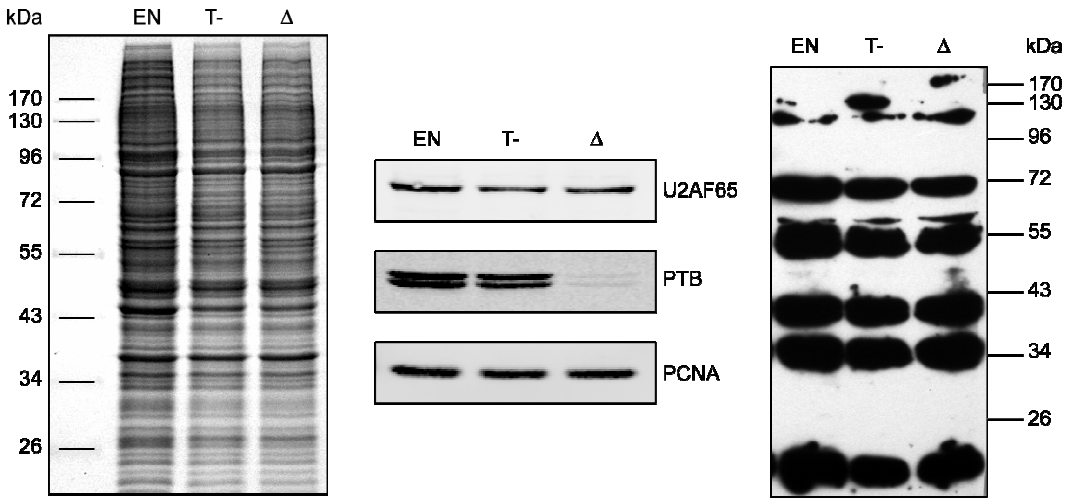


Figure 44- L'immunodéplétion de la PTB des extraits nucléaires de cellules HeLa est spécifique de cette protéine. A. Les extraits nucléaires de cellules HeLa non traité (EN), immunodéplétés en PTB (Δ) ou non traités avec l'anticorps (T-) ont été déposés sur un gel 10% polyacrylamide/SDS. La coloration des protéines au bleu de coomassie ne révèle pas de différence entre les trois échantillons. **B.** L'analyse par western blot de PTB (anticorps α-PTB) montre une déplétion de 98% de la protéine (quantification réalisée sur les signaux de fluorescence des anticorps secondaires révélés avec le système Osyssey). La quantité de U2AF65 et de PCNA est similaire dans les trois échantillons. **C.** Aucune différence significative en protéines SR n'est observée entre les trois échantillons suite à la révélation de l'anticorps 10H3 par l'ECL (Pierce).

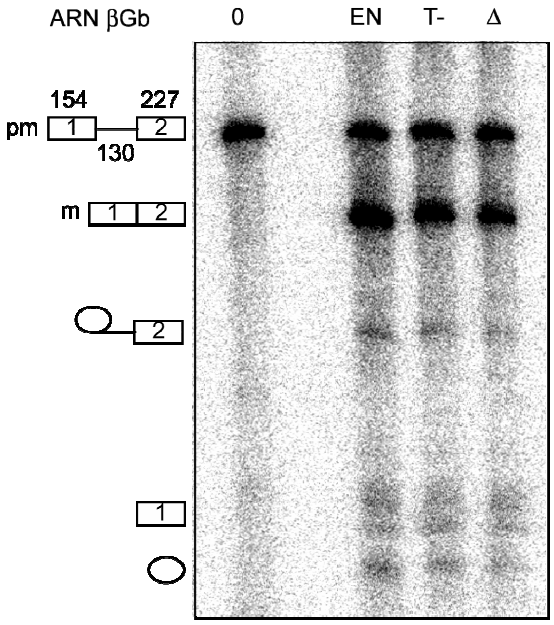


Figure 45- Les extraits nucléaires de cellules HeLa immunodéplétés en PTB sont fonctionnels pour l'épissage *in vitro*. L'ARNpré-m radiomarqué de la β-globine a été incubé deux heures à 30°C dans solution d'épissage contenant 40% d'extrait nucléaire de cellules HeLa soit non traité (EN), soit immunodéplété en PTB (Δ), soit non traité avec l'anticorps (T-). Les signaux correspondant aux ARNpré-m et aux ARNm ont été quantifiés au PhosphorImager. Une diminution de moins de 8% de l'efficacité d'épissage est observée entre les extraits T et Δ PTB. La taille des exons et de l'intron de la β-globine est indiquée.

2.3.3.2. Imunodéplétion en PTB

La deuxième stratégie que nous avons testée est une immunodéplétion de la PTB dans les extraits nucléaires à l'aide d'un anticorps monoclonal anti-PTB commercial ((Ab-1) Mouse mAb (SH54) (CALBIOCHEM)) (Sauliere *et al.*, 2006). L'anticorps anti-PTB est couplé à des billes magnétiques protéine G. L'extrait nucléaire est ensuite traité avec ces billes et le contenu de PTB est analysé par western-blot. Cette stratégie conduit à une déplétion en PTB d'environ 95% (figure 44 B. panel du milieu). Aucune différence du profil général des protéines n'est observée entre les extraits initiaux, traités en l'absence d'anticorps anti-PTB, et déplétés en PTB (figure 44 A.). Une étude plus fine, réalisée par western-blot, de U2AF65 (figure 44 B. panel du haut) et des protéines SR (figure 44 C.) ne montre pas de différence entre les trois extraits. Enfin, l'épissage de l'ARN β-globine dans ces trois extraits a la même efficacité (figure 45).

Cette deuxième approche nous permet donc d'obtenir des extraits nucléaires de cellules HeLa dont le contenu en PTB est très nettement inférieur à celui des extraits d'origine. Ces extraits nucléaires peuvent être utilisés pour étudier l'influence de la PTB sur la composition des complexes assemblés autour des séquences activatrices de l'exon 9A9'.

2.3.4. Purification des complexes RNP assemblés sur la séquence UTE *in vitro* à l'aide d'ARN modifiés avec le métapériodate de sodium

2.3.4.1. Formation des complexes RNP in vitro

Nous avons effectué des expériences de retard de migration sur gel afin d'analyser la formation des complexes sur les quatre ARN UTEcwt, UTEcm, UTElwt et UTElm (figure 38) dans les extraits nucléaires de cellules HeLa contenant ou non la protéine PTB. Ces expériences nous permettent également de tester différentes conditions d'assemblage des protéines sur les ARN. Afin de limiter les fixations non spécifiques des protéines de l'extrait sur les ARN, nous avons testé l'effet de l'héparine sur la formation des complexes. L'addition d'héparine permet en effet de diminuer les fixations non spécifiques des protéines (Markovtsov *et al.*, 2000; Jurica *et al.*, 2002). Cependant, il a également été décrit dans la littérature qu'un traitement à l'héparine peut déstabiliser les complexes H et E du spliceosome et pourrait conduire à la fixation de nouveaux facteurs protéiques (Sharma *et al.*, 2005).

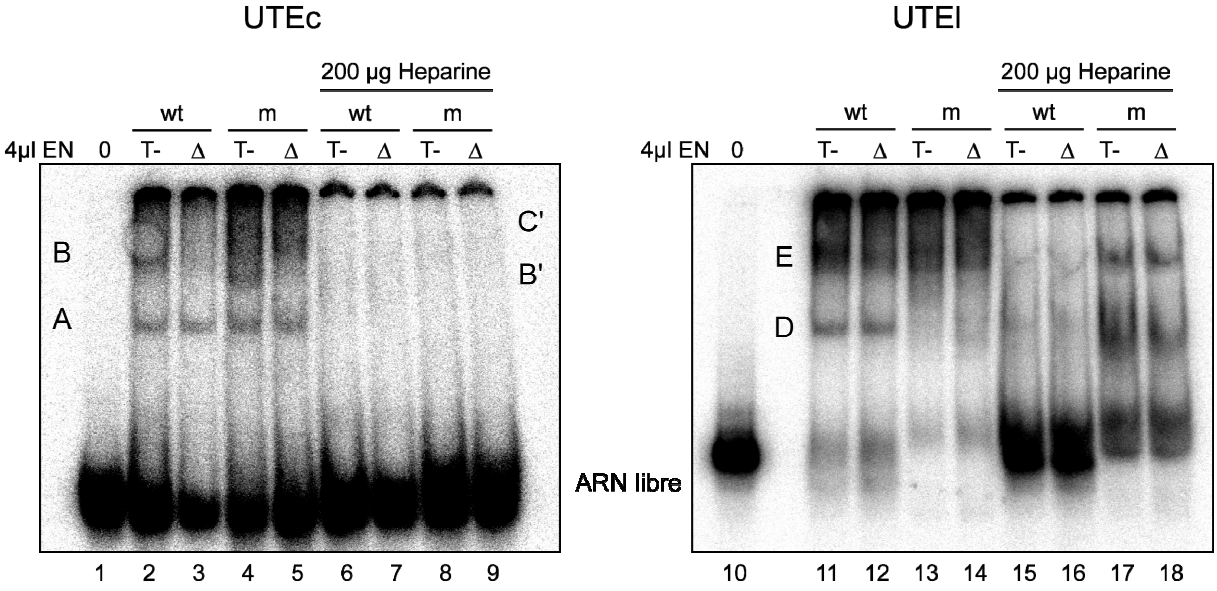


Figure 46- Les complexes RNP recrutés sur les ARN UTEc et UTEl sont différents. Les ARN UTEcwt/m et UTElwt/m radiomarqués ont été incubés en présence d'extraits nucléaires de cellules HeLa déplétés (Δ) ou non (T-) en PTB. Les complexes ainsi formés ont été séparés sur gel 8% polyacrylamide/5% glycérol/ 0,5X TBE, puis révélés au Storm.

Pour former les complexes nous avons employé le milieu réactionnel utilisé lors des réactions d'épissage *in vitro* additionné ou non d'héparine. Dans chaque expérience, une même quantité d'ARN (5 pmol) a été incubée avec les extraits nucléaires de cellule HeLa préalablement traités (Δ) ou non (T-) avec les anticorps anti-PTB. Les résultats de nos expériences présentés dans la figure 46, indiquent que seule une partie des complexes assemblés autour des ARN ont été séparés sur le gel après électrophorèse, l'analyse a donc été réalisée à partir de ces résultats.

Dans les conditions expérimentales utilisées, l'allongement des ARN de 17 nt augmente l'efficacité de formation de complexes comme en témoigne la comparaison des pistes (1-5) avec les pistes (10-14) de la figure 46. En effet, une quantité importante d'ARN UTEcwt et UTEcm reste sous forme libre alors que la majorité des ARN UTElwt et UTElm forment des complexes.

L'analyse des complexes assemblés autour des ARN courts met en évidence qu'ils sont sensibles à la présence d'héparine dans le milieu réactionnel (figure 46 comparer les pistes 2 à 5 aux pistes 6 à 9). En absence d'héparine, nous observons la formation d'un complexe noté A qui est indépendante de la quantité de PTB contenue dans les extraits nucléaires (comparaison des pistes 2 et 4 avec respectivement les pistes 3 et 5) et indépendante du motif UTE (comparaison des pistes 2 et 3 avec respectivement les pistes 4 et 5). Un second complexe appelé B est formé sur l'ARN UTEcwt uniquement en présence de PTB dans l'extrait nucléaire (figure 46, comparer les pistes 2 et 3). Deux autres complexes notés B' et C', sont assemblés autour de l'ARN UTEcm dans l'extrait nucléaire témoin (piste 4). Le complexe C' persiste dans les extraits déplétés de PTB par contre la formation du complexe B' semble être dépendante de la présence de PTB dans l'extrait nucléaire (comparer les pistes 4 et 5).

Deux complexes notés D et E sont assemblés sur l'ARN UTElwt (figure 46, pistes 11 et 12). Etant donné que les ARN UTEl contiennent deux sites de fixation à la protéine PTB, nous nous attendions à une variation du profil de migration de ces complexes en fonction de la quantité de PTB contenue dans les extraits nucléaires. Les résultats présentés dans la figure 46 ne nous permettent pas de faire ces observations (comparer les pistes 11 et 12). Toutefois, nous ne pouvons pas exclure que ces variations existent car une importante proportion des complexes n'a pas pu être analysée comme l'indique la forte radioactivité détectée au niveau des puits des gels d'électrophorèse (figure 46). La présence d'héparine dans l'extrait nucléaire

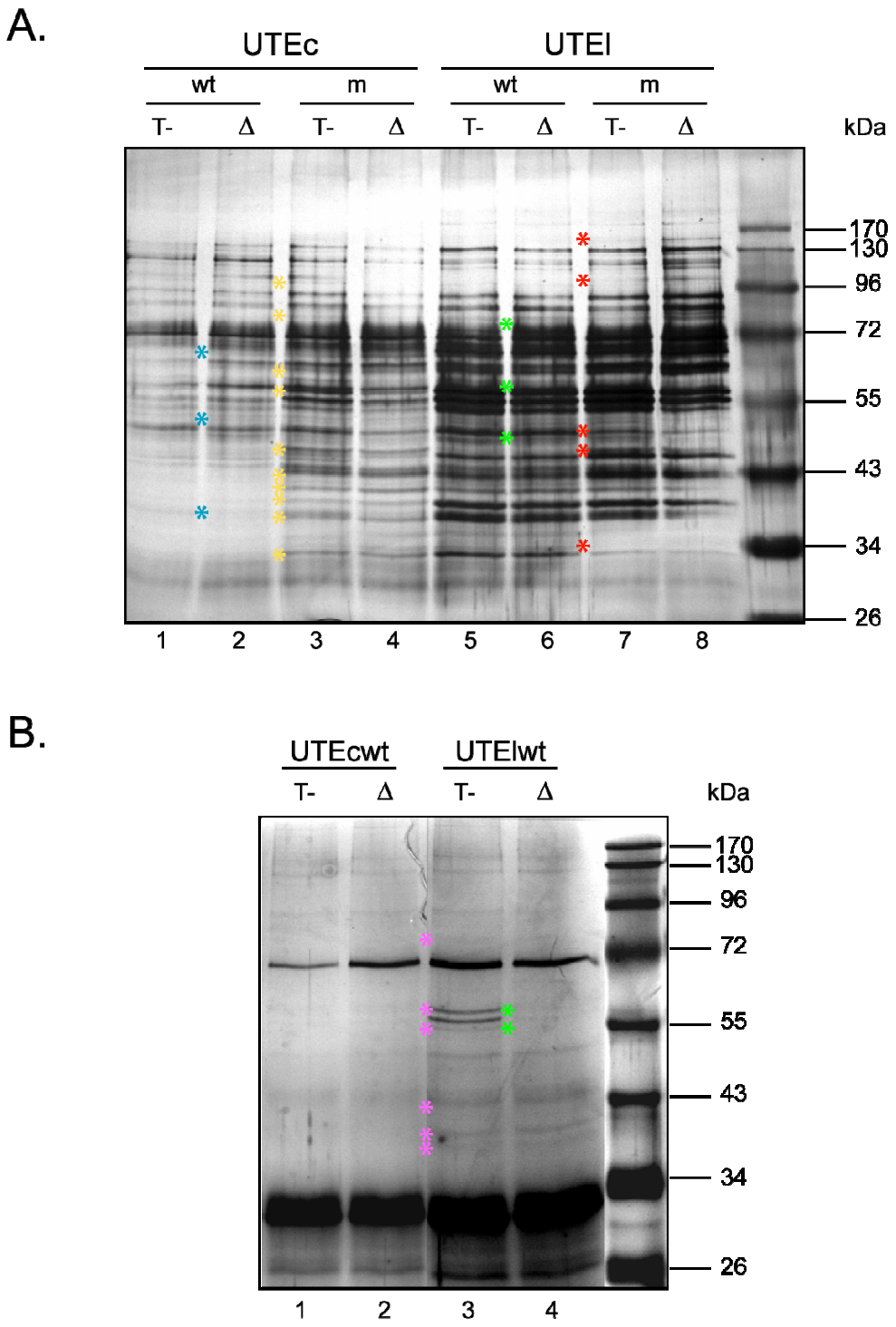


Figure 47- Purification des complexes sans ou avec héparine. Les ARN UTEcwt/m et UTElwt/m sont incubés dans des extraits nucléaires de cellules HeLa déplété (Δ) ou non (T-) en PTB en **absence A**. ou en **présence B.** d'héparine. Les complexes RNP sont isolés par chromatographie d'affinité à l'ARN et les protéines purifiées sont séparées sur gel 10% polyacrylamide/SDS puis visualisées par coloration au nitrate d'argent. Les protéines présentes uniquement dans un des deux types d'extrait nucléaire et complexées aux ARN UTEcwt et UTElwt sont indiquées respectivement en bleu et en vert. **A.** Les protéines de l'extrait nucléaire T- différentiellement fixées sur les ARN UTEcwt et m ou UTElwt et m, sont signalées par des étoiles respectivement jaunes ou rouges. **B.** Les protéines présentes uniquement sur les ARN UTEl sont indiquées par des étoiles roses.

provoque la déstabilisation des complexes assemblés autour de l'UTElwt (pistes 11-12 comparées aux pistes 13-14).

Un seul complexe ayant une migration similaire à celle du complexe E s'assemble autour de l'ARN dont le motif UGGAUGG est muté (UTElm ; figure 46 pistes 13 et14). La formation de ce complexe ne semble pas dépendre de la quantité de protéine PTB contenue dans l'extrait nucléaire (comparer les pistes 13 et 14). En présence d'héparine, deux complexes de plus grande mobilité électrophorétique restent assemblés autour de cet ARN muté (pistes 13-14 comparées aux pistes 17-18), alors que tous les complexes assemblés autour des ARN UTElwt sont déstabilisés.

Ces expériences montrent que la quantité de PTB contenue dans l'extrait nucléaire ainsi que la présence du motif UGGAUGG vont guider la formation de différents complexes autour des ARN. Nous avons donc poursuivi la mise en place de la stratégie de purification des complexes assemblés autour des quatre ARN UTEcwt, UTEcm, UTElwt et UTElm.

2.3.4.2. Purification des complexes RNP par chromatographie d'affinité à l'ARN

Les quatre ARN UTEcwt, UTEcm, UTElwt et UTElm (100 pmol de chaque ARN) ont été traités par le métapériodate de sodium afin d'être immobilisés sur une matrice d'agarose selon le protocole décrit dans la partie matériels et méthodes. L'efficacité de fixation des ARN sur les billes d'agarose a été calculée en utilisant des traces d'ARN radiomarqués. Pour tous les ARN utilisés, le rendement de fixation sur la matrice est supérieur à 95%. Les ARN fixés sur la matrice ont été incubés en présence de 100 µl d'extrait nucléaire de cellules HeLa préalablement traité ou non avec les anticorps anti-PTB. Après quatre lavages avec un tampon contenant 0,1 % de NP40 puis trois lavages sans détergent, les protéines retenues sur les colonnes sont éluées par traitement à la RNase puis séparées par migration sur gel SDS-PAGE 10%. Les résultats de cette première expérience sont présentés dans la figure 47A. Nous observons un certain nombre de protéines spécifiquement fixées sur les ARN en fonction du type d'extrait nucléaire de cellules HeLa utilisé (figure 47A comparer les pistes 1 et 2 ; 5 et 6 ; 1 et 3 ; 5 et 7). Cependant, la majeure partie des protéines est commune à toutes les conditions, suggérant qu'elles sont retenues de manière non spécifique sur tous les ARN et de manière indépendante de la quantité de PTB présente dans l'extrait nucléaire. Afin de diminuer ces fixations non spécifiques, nous avons réalisé la même expérience en présence d'héparine (0,25 mg/ml). Le nombre de protéines retenues sur les différents ARN chute considérablement comme l'indique la figure 47B, ce qui est en accord avec les analyses

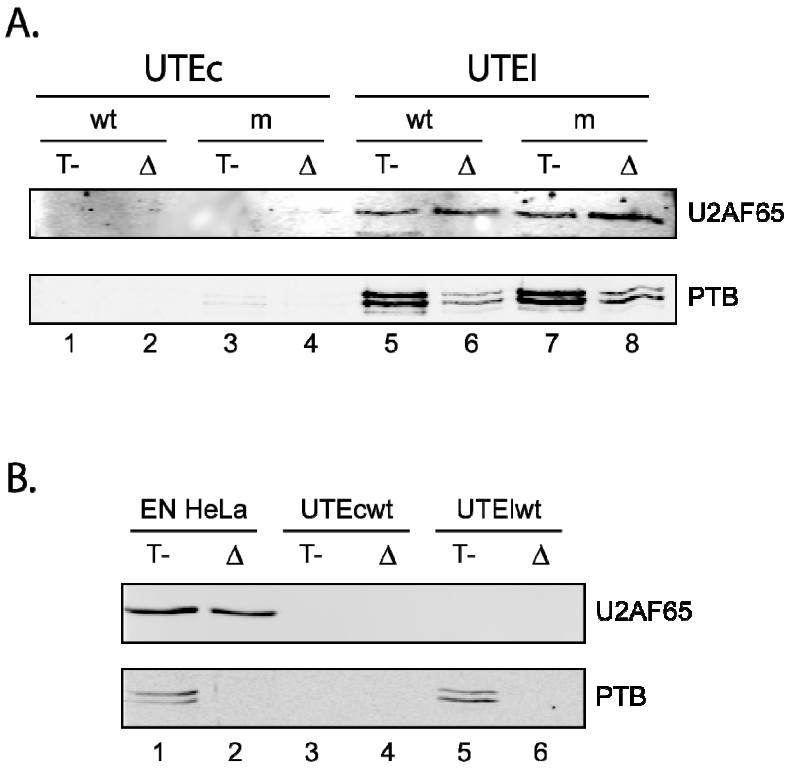


Figure 48- La composition des complexes en PTB et U2AF65 varie en fonction de l'ajout d'héparine. Les extraits nucléaires de cellules HeLa (EN) déplétés (Δ) ou non (T-) en PTB ont été incubés **A. sans héparine** en présence des ARN UTEcwt/m et UTElwt/m fixés à une matrice Adipic acid dihydrazide-Agarose ; ou **B. en présence d'héparine** et des ARN UTEcwt et UTElwt. Les complexes sont purifiés puis les protéines sont séparées sur gel 10% polyacrylamide/SDS (dépôt d'une quantité équivalente de protéines totales par puit). L'analyse par western blot est réalisée avec un anticorps dirigé soit contre la protéine PTB, soit contre U2AF65. Les signaux ont été révélés avec le système Odyssey. Les pistes EN HeLa correspondent à 1/125ème de la fraction de départ utilisée pour la formation des complexes.

préliminaires de formation des complexes par des expériences de retard de migration sur gel. De manière intéressante, certaines protéines apparaissent spécifiquement éluées des ARN UTEl (figure 47B. comparer les pistes 1-2 et 3-4) en fonction de la quantité de PTB présente dans l'extrait nucléaire (comparer les pistes 3 et 4).

L'ensemble de ces expériences montre qu'il existe des différences dans le recrutement des protéines autour des ARN utilisés, en particulier celles dont la masse moléculaire est aux alentours de 72 kDa (figure 47A. comparer les pistes 5 et 6 ; 7 et 8 ; 5 et 7).

Nous avons ensuite recherché la présence des protéines SR et PTB dans ces différentes fractions protéiques par des expériences de western-blot. La protéine U2AF65 a également été analysée puisqu'une compétition de fixation entre la PTB et le facteur U2AF65 est impliquée dans la régulation de l'épissage de l'exon 6B de la β-tropomyosine (Sauliere *et al.*, 2006). Les résultats sont présentés dans la figure 48 (A. complexes assemblés sans héparine ; B. complexes assemblés en présence de 0,25 mg/ml d'héparine). En absence d'héparine, la protéine U2AF65 est présente dans les complexes purifiés à partir des ARN UTEl indépendamment de la quantité de PTB présente dans l'extrait nucléaire (figure 48A. pistes 5 à 8). Elle est également détectée en très faible quantité dans les complexes recrutés autour de l'ARN UTEcm en absence de PTB (piste 4).

La révélation des western-blot à l'aide du système Odyssey permet de quantifier les protéines contenues dans les différentes fractions éluées. La PTB est présente dans les complexes assemblés autour de l'ARN UTEl même lorsque les extraits nucléaires où la quantité de PTB est diminuée de plus de 95% sont utilisés. Dans ces conditions, la quantité de PTB retrouvée dans ces complexes est six fois moins importante que dans ceux assemblés dans un extrait nucléaire non traité avec l'anti-PTB (figure 48A. comparer les pistes 5 et 6 ; 7 et 8). La PTB est également présente en faible quantité dans les complexes purifiés à partir de l'ARN UTEcm (figure 48A. piste 3). Ces expériences ne montrent pas de compétition de liaison entre la PTB et la protéine U2AF65 puisque cette dernière semble globalement présente en quantité constante quelle que soit la quantité de PTB observée dans les complexes purifiés. Cependant, lorsqu'on observe le comportement des protéines assemblées autour de l'ARN UTEcm, on remarque que la protéine U2AF65 est détectée uniquement lorsque la quantité de PTB contenue dans l'extrait nucléaire est diminuée (figure 48A., pistes 3 et 4). Cette protéine n'est pas observée dans les fractions issues de l'assemblage

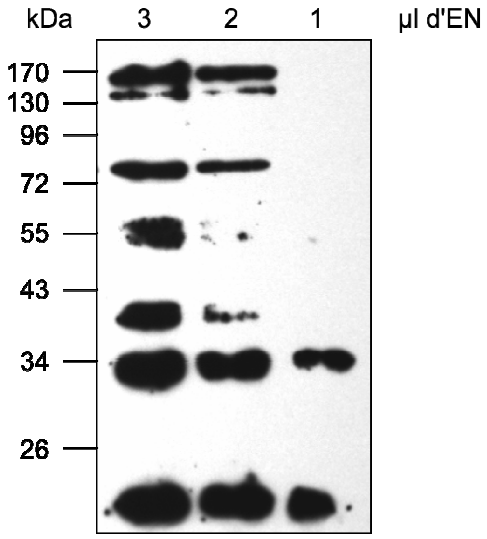


Figure 49- Trois µl d'extrait nucléaire de cellules HeLa sont nécessaires pour détecter les protéines SR par western blot. 1, 2 et 3 µl d'extrait nucléaire de cellules HeLa sont déposés sur un gel 10% polyacrylamide/SDS. L'analyse par western-blot est réalisée avec un anticorps dirigé contre un épitope phosphorylé commun aux protéines SR. Les signaux ont été révélés par ECL West Dura (Pierce).

de complexes autour de l'ARN UTEcwt. Ce résultat peut s'expliquer par l'analyse de la structure secondaire de ces ARN. Comme présenté précédemment dans la figure 37, la mutation de la séquence UTE de TGGATGG vers CTTATAC, modifie les régions appariées des ARN UTEc. Celle-ci favorise l'apparition de régions en simple brin, plus favorables à la fixation de facteurs protéiques que les zones en double brin de l'ARN. Ces régions non appariées sont riches en pyrimidines, particulièrement en uracile, constituant ainsi des sites de fixation potentiels pour PTB et U2AF65. La PTB pourrait donc se fixer sur l'ARN UTEcm, tout comme U2AF65 en l'absence de PTB. Le même type de modification structurale due à la mutation de la séquence UTE est observable pour les ARN UTEl. Toutefois, l'ARN UTElm ne présente qu'une seule région riche en pyrimidines structurée en simple brin de plus que l'ARN UTElwt (figure 39). Ceci pourrait expliquer la faible augmentation des signaux (1,5 fois) correspondant aux protéines PTB et U2AF65 issues des complexes recrutés sur l'ARN UTElm par rapport à ceux assemblés sur l'ARN UTElwt. Ces légères différences ne peuvent donc pas être considérées comme significatives.

En présence d'héparine, nous détectons uniquement la protéine PTB dans les fractions issues des complexes formés sur l'ARN UTElwt à partir d'extrait nucléaire non traité avec l'anticorps anti-PTB (figure 48B.). L'héparine ayant un effet beaucoup trop délétère sur la stabilité des complexes, nous avons décidé de ne plus l'utiliser dans nos expériences.

Aucune protéine SR n'a pu être détectée dans les différentes fractions protéiques analysées. Les expériences réalisées à partir d'extrait nucléaire de cellules HeLa montrent qu'il existe un seuil de détection des protéines SR en western-blot (figure 49). Ce seuil correspond à 1/33ème de la quantité d'extrait utilisé pour la formation des complexes. La quantité de protéines dans les complexes pourrait être insuffisante pour qu'elles puissent être détectées par Western-Blot à l'aide de notre anticorps. Nous ne pouvons donc pas conclure sur l'absence de protéines SR dans nos fractions.

Nos expériences d'assemblage de complexes réalisées avec les ARN UTEl ont révélées des différences dans le recrutement de protéines dont le poids moléculaire apparent se situe aux alentours de 72 kDa (figure 47 et figure 50). Nous avons donc analysé par chromatographie liquide couplée à la spectrométrie de masse en tandem (LC-MS/MS) les protéines des complexes RNP assemblés sur les ARN UTElwt et muté en présence ou en absence de PTB, migrant entre 68 et 80 kDa.

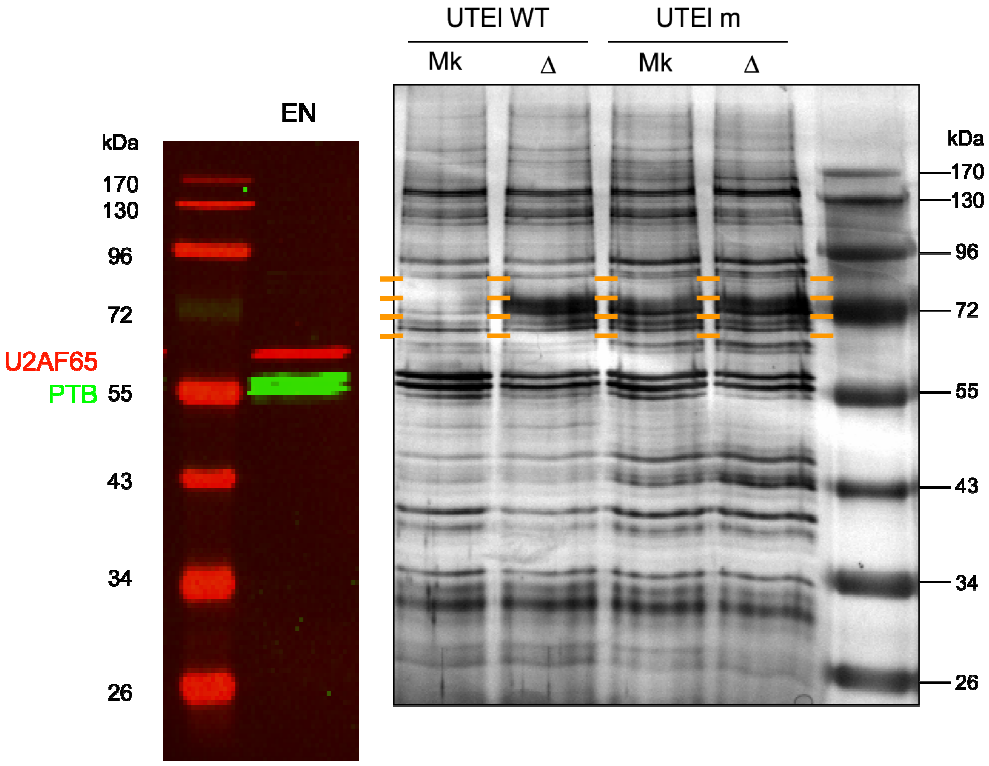


Figure 50- Fragments de gel utilisés pour les identifications par spectrométrie de masse. Les extraits nucléaires de cellules HeLa déplétés (Δ) ou non en PTB (T-) sont incubés en présence des ARN UTElwt ou UETl**.** Les complexes sont isolés par chromatographie d'affinité à l'ARN et les protéines purifiées sont séparées sur gel 10% polyacrylamide/SDS puis visualisées par coloration au nitrate d'argent. Chaque piste a été découpée selon les marques orange placées sur le gel. Le western-blot placé à gauche permet d'observer les protéines U2AF65 et PTB révélées par fluorescence avec le système Odyssey.

2.4. Analyse par spectrométrie de masse de la fraction 68 à 80 kDa des protéines recrutées sur les ARN UTEl

Les protéines présentes dans les complexes assemblés sur les ARN UTEl ont été fractionnées par électrophorèse en SDS-PAGE. Pour chaque piste, le gel a été découpé en trois fractions situées entre 68 et 80 kDa, comme présenté dans la figure 50. Ces régions ont volontairement été choisies éloignées de la zone susceptible de contenir la PTB, qui du fait de sa présence en grande quantité pourrait gêner l'identification d'autres protéines présentes en plus faible quantité. Les protéines contenues dans les différentes fractions ont été éluées du gel puis digérées par de la trypsine. Les peptides ainsi produits ont été séparés puis caractérisés par LC-MS/MS sur la plate-forme protéomique du Centre Génomique de Québec. Les produits ont été identifiés à l'aide du logiciel Mascot (Matrix Science) et les résultats visualisés avec le logiciel Scaffold (Proteome Software). Nous avons considéré qu'une protéine est présente dans l'échantillon quand au moins deux de ses peptides ont une probabilité d'identification supérieure à 95%. Les protéines ainsi identifiées ayant une masse moléculaire non compatible avec les échantillons sélectionnés à plus ou moins 15 kDa ont été éliminées. Cette marge d'erreur, tient compte du découpage manuel des fragments de gel, de la qualité des marqueurs de taille utilisés, des éventuelles modifications post-traductionnelles des protéines purifiées, et du fait que certaines protéines ont une masse moléculaire apparente différente de leur masse réelle. J'ai réalisé une deuxième étude d'identification manuelle de tous les peptides sur le serveur ExPASy afin de valider la première analyse. Finalement, cette dernière étude nous a permis de sélectionner 38 protéines retenues sur les ARN UTElwt et UTElm. Ces protéines sont présentées dans le tableau 6.

Les protéines obtenues ont été comparées à celles précédemment identifiées par Chen et collaborateurs (Chen *et al.*, 2007) qui comprend les données de trois autres études (Neubauer *et al.*, 1998; Rappsilber *et al.*, 2002; Zhou *et al.*, 2002). Ces quatre analyses ont consisté à identifier le plus exhaustivement possible les protéines composant le spliceosome. Dans la première étude, les complexes spliceosomaux ont été formés dans des cellules HeLa ou dans des cellules de poulet SmD3-TAP DT40. Ces complexes ont ensuite été isolés par ultracentrifugation puis chromatographie d'affinité et les protéines purifiées ont été identifiées par spectrométrie de masse. Les trois autres études, entièrement réalisées *in vitro*, ont consisté à former des complexes d'épissage à partir d'extrait nucléaire de cellules HeLa sur des ARN

Nom des protéines	N° d'accession ExPASy	Masse moléculaire (kDa)	Crible cellules HeLa	Identifiées à partir de	Interactions
Protéines connues pour être associées au spliceosome					
ATP dependent DNA helicase 2 subunit 1 (XRCC6)	P12956	69	X	WT- ; WΔ	D
NCBP1 HUMAN Nuclear cap-binding protein subunit 1	Q09161	91	X	WT- ; MΔ	R
PTBP1 Polypyrimidine tract-binding protein 1	P26599	57	X	WT- ; WΔ	R
HSP70.1 like Heat shock 70 kDa protein 1L	P34931	70	X	T	N (ATP)
HSP7C Heat shock cognate 71 kDa protein	P11142	70	X	T	P
ARN-binding protein 39 (Splicing factor HCC1)	Q14498	60	X	T	R
ARN-binding protein EWS	Q01844	68	X	T	P, R
ARN-binding protein FUS	P35637	53	X	T	D, P, R
SF01 Splicing factor 1	Q15636	68	X	WT- ; MΔ	R
ARN hélicases					
ATP-dependent ARN helicase DDX1	Q92499	82	X	T	R
ATP-dependent ARN helicase DDX17	Q92841	72	X	T	R
ATP-dependent ARN helicase DDX3X	O00571	73	X	T	D, R
ATP-dependent ARN helicase DDX5	P17844	69	X	T	R
Putative pre-mARN-splicing factor ATP-dependent ARN helicase DHX15	O43143	90	X	T	N (ATP)
Protéines participant à la maturation de l'extrémité 3' des ARN					
CPSF3 Cleavage and polyadenylation specificity factor subunit 3 (CPSF73)	Q9UKF6	77	X	WT-	P, R
CSTF2 Cleavage stimulation factor 64 kDa subunit	P33240	60	0	WT- ; WΔ	R
CSTF3 Cleavage stimulation factor 77 kDa subunit (CstF77)	Q12996	83	X	WT- ; MΔ	P, R
PABP1 Polyadenylate-binding protein 1	P11940	70	X	WT- ; WΔ	R
Protéines hnRNP					
hnRPK Heterogeneous nuclear ribonucleoprotein K	P61978	51	X	WΔ ; MT-	D, R
hnRPL Heterogeneous nuclear ribonucleoprotein L	P14866	64	X	WΔ ; MT-	R
hnRPM Heterogeneous nuclear ribonucleoprotein M	P52272	77	X	WΔ	R
hnRPQ Heterogeneous nuclear ribonucleoprotein Q	O60506	69	X	T	R
hnRPR Heterogeneous nuclear ribonucleoprotein R	O43390	71	X	T	R

Protéines de liaison aux ARN					
ARN binding motif protein 14 variant	Q59GV2	56	X	WT-	D, R
ILF3 Interleukin enhancer-binding factor 3	Q12906	95	X	T	R (X2)
Protéines associées aux snRNP					
U4/U6.U5 tri-snRNP-associated protein 2	Q53GS9	65	X	WΔ ; MT-	P, R
Protéines participant à la transcription					
FUBP1 Far upstream element binding protein 1	Q96AE4	67	-	T	D (X1)
Protéines participant à l'épissage					
ESRP2 Epithelial splicing regulatory protein 2	Q9H6T0	78	0	WT-	R
FUBP2 Far upstream element-binding protein 2	Q92945	73	0	T	D, R
Pas de fonction connue dans la maturation des ARNpré-m					
ATP dependent DNA helicase 2 subunit 2 (XRCC5)	P13010	82	X	WT- ; WΔ	D
HSP90 Heat shock protein HSP 90-beta	P08238	83	X	T	P
EIF2AK2 Interferon-induced double-stranded ARN-activated protein kinase	P19525	62	0	T	R (X2), P
hRED1 ARN editing deaminase 1	O43263	76	0	MT-	R (X2)
IF2B1 Insulin-like growth factor 2 mARN-binding protein 1	Q9NZI8	63	0	T	R
IF2B3 Insulin-like growth factor 2 mARN-binding protein 3	O00425	63	X	T	R
SYNCRIP Synaptotagmin binding cytoplasmic ARN interacting protein	Q59GL1	60	0	WΔ	D, R
TBL3 Transducin beta-like protein 3	Q12788	89	0	MT-	P
DNA repair protein XRCC1	P18887	70	0	MT-	D, P

Tableau 6- Identification par spectrométrie de masse des protéines assemblées sur les ARN UTEl. Les protéines identifiées par spectrométrie de masse (LC-MS/MS) à partir des quatre échantillons ont été comparées à celles précédemment identifiées par (Chen *et al.*, 2007). Quand celles-ci sont présentes dans les deux analyses, elles sont marquées par X, le symbole "-" signifie que la protéine a été identifiée lors des études réalisées par (Rappsilber *et al.*, 2002), 0 correspond aux protéines spécifiquement identifiées au cours de notre analyse. Le numéro d'accession sur la base de donnée ExPASy est indiqué, ainsi que la masse moléculaire et le type préférentiel de liaisons : D liaison à l'ADN (noté X1 si en simple brin), N liaison aux nucléotides avec entre parenthèses le nucléotide d'intérêt, P liaison aux protéines, R liaison à l'ARN (noté X2 si en double brin). Les protéines purifiées à partir des 4 conditions sont notées T, les protéines présentes sur l'ARN UTElwt en présence d'extrait nucléaire de cellules HeLa déplété ou non en PTB sont notées respectivement WΔ et WT-, et celles présentes sur l'ARN UTElm en présence d'extrait nucléaire de cellules HeLa déplété ou non en PTB sont notées respectivement MΔ et MT-.

modifiés de manière à interagir avec une matrice permettant la purification des complexes puis l'identification des protéines.

Sur les 38 protéines que nous avons identifié neuf d'entre elles n'ont été isolées dans aucune de ces quatre études. Ces protéines sont : CSTF2, IF2B1, ESRP2, FUBP2, EIF2AK2, hRED1, SYNCRIP, TBL3, XRCC1. Parmi celles-ci, CSTF2 et FUBP2 sont impliquées dans la maturation des ARNpré-m (tableau 7), tout comme ESRP2 qui sera présentée en détail ultérieurement. A l'inverse, les six autres protéines n'ont pas de fonction connue dans la maturation des ARNpré-m.

Les protéines IF2B1 et EIF2AK2, sont présentes dans les quatre conditions utilisées pour les purifications, suggérant que leur interaction avec les deux ARN que nous avons utilisés n'est liée ni au motif UGGAUGG, ni à la quantité de PTB présente dans l'extrait nucléaire. IF2B1 (ou IMP1) possède deux RRM et quatre domaines KH permettant sa fixation à l'ARN. Cette protéine est capable d'entrer en compétition avec la PTB pour la fixation à l'ARN, ceci de manière dépendante de la présence de $MgCl_2$. Nielsen et collaborateurs ont pu déterminer que la fixation de IF2B1 et PTB sur un ARN IGF-II était à l'équilibre pour une concentration en $MgCl_2$ de 1,5 mM (Nielsen *et al.*, 1999). L'assemblage des complexes *in vitro* étant réalisé à une concentration de 4 mM en $MgCl_2$, il favorise la fixation de IF2B1 sur les ARN UTEl. A ce jour, seule une fonction dans la répression traductionnelle de IGF-II et de PABP a été décrite pour IF2B1 (Nielsen *et al.*, 1999; Patel *et al.*, 2005). Tout comme IF2B1, EIF2AK2 (ou PKR) se fixe sur l'ARN. Sa transcription est induite par la présence d'interféron. Elle est ensuite activée en se fixant sur de l'ARN double brin d'origine viral. Son activation provoque une inhibition générale de la traduction, limitant ainsi la synthèse de protéines virales (Zhang *et al.*, 2001). Il est donc peu probable que la fixation de EIF2AK2 sur l'ARN UTEl soit en relation avec la fonction de la séquence régulatrice UTE sur la maturation de l'exon 9A9'.

Les protéines hRED1, TBL3, et XRCC1 sont présentes exclusivement dans les complexes assemblés autour de l'ARN UTElm quand celui-ci est incubé dans l'extrait nucléaire de cellules HeLa non déplété en PTB. hRED1 se fixe sur de l'ARN structuré en double brin grâce à deux domaines de liaison à l'ARN double brin (dsRBD). Cette protéine a

Protéines	Rôle, fonctions connues	Références
Rôle dans la maturation des ARNpré-m		
CSTF2	Mise en place des complexes de clivage/polyadénylation	(MacDonald *et al.*, 1994; Perez Canadillas et Varani, 2003)
DDX1	Interaction avec hnRNP K et CstF 64, modifie l'affinité du complexe formé pour l'ARN	(Bleoo *et al.*, 2001; Chen *et al.*, 2002)
DDX5 DDX17	Impliquées de l'EA. Interaction avec snRNP U1. Pourraient moduler la structure II des ARNpré-m	(Honig *et al.*, 2002; Lee, 2002; Liu, 2002; Lin *et al.*, 2005; Camats *et al.*, 2008)
DHX15	Interviendrait dans relargage de l'intron en lasso et désassemblage du spliceosome	(Jurica *et al.*, 2002; Wen *et al.*, 2008)
EWS FUS	Activation transcription, EA, formation d'oncoprotéines. Interaction avec SC35, Srp75 et SRm160	(Meissner *et al.*, 2003; Law *et al.*, 2006)
FUBP1	Active la transcription. Interaction avec TIA 1, régulation de l'EA ?	(Duncan *et al.*, 1994; Rothe *et al.*, 2006)
FUBP2	Activation de l'EA. Fixation sur un ISE	(Min *et al.*, 1997); (Markovtsov *et al.*, 2000)
hnRPQ hnRPR	90 % d'homolgie de séquence. Régulation de l'EA +/-. Interaction pyrimidines et poly(U)	(Mourelatos *et al.*, 2001; Rossoll *et al.*, 2002; Chen *et al.*, 2008)
HCC1	Famille des SR. Régulation transcription et EA de gènes répondant aux hormones stéroïdes	(Dowhan *et al.*, 2005; Ellis *et al.*, 2008)
PTBP1	Participe à la régulation de la transcription, de l'EA, du C/P, de l'export, de la stabilité, de la traduction	(Sawicka *et al.*, 2008)
Rôle dans l'export, la stabilité des ARNm ou dans la traduction		
DDX3X	Participe à l'export nucléaire des ARNm	(Merz *et al.*, 2007)
ILF3	Fixation 3' UTR ARNm MyoD et P21WAF1/CIP1. Stabilisation de ces ARNm.	(Shi *et al.*, 2005)
PABP1	Stabilisation des ARNm. Favorise initiation traduction. Export des ARNm et allongement de la queue poly(A)?	(Minvielle-Sebastia *et al.*, 1997; Afonina *et al.*, 1998; Khacho *et al.*, 2008)
Implication dans la transcription ou indirectement dans l'épissage		
HSP70.1 HSP7C	Repliement des protéines. Pourraient favoriser la reprise de l'épissage suite choc thermique (levure)	(Milner et Campbell, 1990; Ito *et al.*, 1998; Tsukahara *et al.*, 2000; Biamonti et Caceres, 2009)
XRCC5 XRCC6	Réparation cassures doubles brin de l'ADN. Rôle dans la régulation de la transcription	(Lieber *et al.*, 1997; Willis *et al.*, 2002; Wang *et al.*, 2004)
Pas de fonction connue dans la régulation de la maturation des ARNpré-m		
HSP90-beta	Fixation sur enveloppe nucléaire. Liens indirects avec maturation des ARN ?	(Freeman et Yamamoto, 2002; Schlatter *et al.*, 2002; Zhao *et al.*, 2008)
EIF2AK2	Inhibition de la traduction suite à infection virale	(Zhang *et al.*, 2001)
IF2B1	Compétition de fixation avec PTB dépendante de MgCl2. Export. Répression de la traduction	(Nielsen *et al.*, 1999; Patel *et al.*, 2005)
IF2B3	Fixation 3' UTR ARNm. Rôle dans stabilité, export des ARNm	(Vikesaa *et al.*, 2006)

Tableau 7- Protéines identifiées dans les complexes assemblés sur l'ARN UTElwt indépendamment de la présence de PTB. Les protéines sont regroupées par fonction. EA signifie épissage alternatif, C/P processus de clivage/polyadénylation et ISE (*Intronic Splicing Enhancer*) élément intronique activateur de l'épissage. Les fonctions hypothétiques des protéines sont représentées par « ? ».

également une affinité très importante *in vitro* pour les répétitions de U en simple brin. Cette fixation est telle, qu'il n'est pas possible de retrouver l'activité de la protéine suite à son élution d'une colonne de poly(U) par une forte concentration saline (O'Connell *et al.*, 1997). Comme présenté précédemment, la mutation de la séquence UTE entraîne une modification de la structure secondaire de l'ARN UTEl (figure 39). L'ARN UTElm possède sept U en simple brin de plus que l'ARN UTElwt. Ces résidus, plus accessibles aux protéines, pourraient permettre la fixation de hRED1 dont l'implication dans la maturation des ARNpré-m n'a encore jamais été démontrée. L'identification des protéines TBL3 et XRCC1 est surprenante. En effet, aucune de ces deux protéines n'est connue pour fixer l'ARN, même indirectement. La protéine XRCC1, qui est impliquée dans la réparation de l'ADN (Levy *et al.*, 2006), n'a pas été montrée comme interagissant, directement ou non, avec l'une des protéines identifiées. Le même constat est fait avec la protéine TBL3. Celle-ci peut se fixer sur d'autres protéines du fait de la présence de nombreuses répétitions tryptophane-aspartate. Son interaction avec le facteur TBL1 pourrait lui permettre de participer à la répression transcriptionnelle de certains récepteurs nucléaires (Choi *et al.*, 2008). De plus, selon les annotations de Gene Ontology, elle pourrait participer à la maturation d'ARN ribosomaux, mais aucune preuve expérimentale n'a encore été apportée.

Ces trois protéines, spécifiquement fixées sur l'ARN UTElm, soulèvent le problème de l'implication biologique des complexes formés *in vitro* sur des ARN dont la séquence a été mutée. Ces ARN sont relativement courts (60 nt), une modification de leur séquence peut non seulement empêcher la fixation spécifique de certaines protéines, mais aussi créer de nouveaux sites d'interaction. Ces changements d'affinité pour différentes protéines peuvent être dus à la séquence ou à la structure de l'ARN. C'est pourquoi, bien que les conditions de reconstitution des complexes *in vitro* semblent optimales, nous avons choisi dans la suite de cette étude de ne comparer que les protéines identifiées en présence et en absence de PTB à partir de l'ARN UTElwt.

Un premier groupe de 24 protéines recrutées sur l'ARN UTElwt indépendamment de la quantité de PTB présente dans les extraits a été mis en évidence. Ces protéines sont présentées dans le tableau 7.

La présence de la PTB parmi ces 24 protéines peut être explicitée. En effet, la déplétion de PTB dans les extraits nucléaires de cellules HeLa n'est que de 95 % (section 2.3.3.2) et il reste suffisamment de protéine pour la détecter en western-blot sur les ARN UTElwt (figure 48). L'affinité de cette protéine pour les régions riches en pyrimidines est importante. La majeure partie de la PTB restant dans les extraits nucléaires de cellules HeLa pourrait se fixer sur l'ARN que nous utilisons. Il est possible de comparer la quantité relative d'une protéine entre deux échantillons analysés par LC-MS/MS à condition qu'ils aient été préparés dans des conditions identiques. Pour une protéine donnée, le nombre de peptides pouvant être identifié suite à la digestion trypsique sera d'autant plus grand que la concentration de cette protéine sera importante. Ainsi le pourcentage de recouvrement de la séquence protéique sera plus important pour une protéine fortement présente dans l'échantillon analysé. Le recouvrement de séquence en acides aminés calculé à partir de la PTB issue des complexes formés à partir d'extrait nucléaire témoin est de 15%. Celui-ci descend à 5% quand les expériences de purifications sont réalisées à partir d'extraits déplétés en PTB. Ces résultats sont en accord avec ceux obtenus en western-blot (figure 48) et impliquent qu'il y ait dans les extraits déplétés en PTB, soit moins de PTB sur chaque molécule d'ARN, soit une augmentation des molécules d'ARN qui ne sont pas liées par la PTB.

Suite aux observations précédemment effectuées par western-blot, nous nous attendions à identifier la protéine U2AF65 parmi les 24 protéines communes aux deux échantillons. De manière surprenante, cela n'est pas le cas. Comme présenté dans la figure 50, celle-ci apparaît relativement proche de la PTB, l'explication la plus simple serait donc qu'elle ne soit pas présente dans les fragments de gel que nous avons utilisés pour effectuer la spectrométrie de masse. Une hypothèse alternative serait que la protéine U2AF65 soit présente en trop faible quantité dans les complexes protéiques assemblés sur les ARN UTEl pour être identifiée par spectrométrie de masse. Toutefois, les résultats obtenus nous permettent de discuter l'hypothèse selon laquelle une compétition de fixation entre la PTB et le facteur U2AF65 participerait à la régulation de l'exon 9A9'. Bien que ces expériences ne nous permettent pas de dire si la liaison de ces protéines sur l'ARN est directe, la présence de PTB dans les complexes n'influe pas sur celle de la protéine U2AF65. Pour écarter définitivement cette hypothèse des expériences de transfert de marquage dans les mêmes conditions de liaison peuvent être réalisées. La protéine U2AF65 étant présente quasi

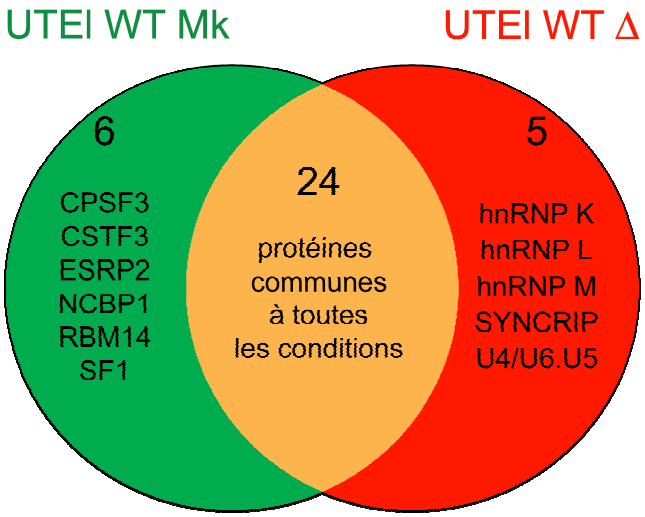


Figure 51- Comparaison des protéines assemblées sur l'ARN UTElwt à partir d'extraits nucléaires de cellules HeLa déplétés ou non en PTB. Les protéines assemblées en présence (T-) ou en absence de PTB (Δ) sont représentées dans le diagramme de Venn. Les 24 protéines communes aux deux conditions sont détaillées au centre, les protéines spécifiques sont présentées de part et d'autre du diagramme.

Protéines	Rôle, fonctions connues	Références
Protéines du complexe de clivage/polyadénylation		
CPSF3 (CPSF 73)	Interagit avec les autres sous-unités du complexe CPSF et avec l'ARN pour réaliser le clivage endonucléolytique de l'ARNpré-m.	(Ryan *et al.*, 2004; Mandel *et al.*, 2006)
CSTF3 (CstF 77)	Interagit sous forme d'homodimère avec les sous unités 50 et 64 de CSTF, et avec CPSF 160. Interagit avec les dinucléotides CA (présents 3 fois dans l'ARN UTElwt).	(Takagaki et Manley, 1994; Murthy et Manley, 1995; Legrand *et al.*, 2007)
Protéine de la coiffe		
NCBP1 (CBP 80)	Interaction avec CBP20 conduit à la formation d'un complexe au niveau de la coiffe des ARN (nécessaire à l'épissage de l'exon 5' terminal). CBP80 est impliquée dans l'export nucléaire (5' vers 3').	(Izaurralde *et al.*, 1994; Cheng *et al.*, 2006; Qiu *et al.*, 2007)
Protéine de régulation de la transcription		
RBM14 (CoAA)	hnRNP-like interagissant avec le corégulateur transcriptionnel TRBP. Elle est impliquée dans la régulation de la transcription et de l'épissage de manière dépendante du promoteur.	(Auboeuf *et al.*, 2004)
Pas de fonction connue dans la régulation de la maturation des ARNpré-m		
SYNCRIP	Majoritairement cytoplasmique. Issue de l'épissage alternatif de hnRNPQ. Séquence identique des aa 50 à 562 de hnRNPQ1, ses 22 premiers aa sont spécifiques. Impliquée dans la réplication de l'ARN viral.	(Mizutani *et al.*, 2000; Choi *et al.*, 2004; Liu *et al.*, 2009)

Tableau 8- Protéines identifiées dans les complexes assemblés sur l'ARN UTElwt en fonction de la présence de PTB. Ne sont présentées ici que les protéines dont la fonction décrite à ce jour ne permet pas d'expliquer le lien avec la séquence UTE quand les complexes sont assemblés *in vitro* sans couplage avec un autre mécanisme de maturation des ARN. Les protéines sont regroupées par fonction. « aa » signifie acides aminés.

exclusivement sur les ARN UTEl, il sera également intéressant de réaliser une cartographie de l'empreinte de cette protéine sur l'ARN le plus long. Un deuxième groupe de onze protéines, spécifiquement recrutées selon la quantité de PTB présente dans les extraits nucléaires de cellules HeLa, a été caractérisé (figure 51). Six de ces protéines sont recrutées autour de l'ARN en présence de PTB : CPSF 3, CSTF 3, ESRP 2, NCBP 1, RBM 14 et SF1. Lorsque la quantité de PTB est fortement diminuée, cinq de ces protéines sont assemblées autour de l'ARN : hnRNP K, hnRNP L, hnRNP M, SYNCRIP et la protéine 2 associée à la tri-snRNP U4/U6.U5.

Ces onze protéines différentiellement présentes sur les ARN UTEl en fonction de la concentration de PTB ne sont probablement pas les seules. En effet, nous n'avons analysé par LC-MS/MS que la fraction des protéines qui présentait le différentiel le plus flagrant. Cependant, il semble exister d'autres différences entre les protéines purifiées, notamment aux alentours de 50 et de 100 kDa (figure 50 comparer les pistes T- et Δ de l'ARN UTElwt). Il serait nécessaire d'identifier l'ensemble des protéines pour déterminer la composition des complexes recrutés autour de la séquence activatrice. Malgré tout, l'étude bibliographique de ces protéines met en avant leur rôle dans le métabolisme des ARN, aussi bien au niveau transcriptionnel que post-transcriptionnel. Toutefois, il semble difficile de justifier le recrutement sur l'ARN des protéines CPSF 3, CSTF 3, NCBP 1, RBM 14 et SYNCRIP par leur fonction biologique (présentée dans le tableau 8). En effet, l'ARN UTElwt ne possède ni coiffe ni signal de clivage/polyadénylation et les extraits nucléaires utilisés ne permettent ni la réplication ni la transcription. L'implication de ces protéines dans la régulation tissulaire de l'utilisation de l'exon 9A9' ne sera donc pas discutée contrairement aux protéines SF1, ESRP2, hnRNP K ;L ;M et la protéine 2 associée à la tri-snRNP U4/U6.U5, dont l'identification nous permet d'avancer dans certaines hypothèses et d'en poser de nouvelles.

La protéine SF1 pourrait être impliquée dans la régulation de la maturation de l'extrémité 3' terminale de l'ARNpré-m de l'α-tropomyosine. Il a été montré que ce facteur se fixe spécifiquement au niveau du point de branchement (Liu *et al.*, 2001) et que sa liaison coopérative avec le facteur U2AF65 favorise le recrutement de la snRNP U2 (Peled-Zehavi *et al.*, 2001). La déplétion de la protéine SF1 dans des cellules en culture n'inhibe pas l'épissage des ARNpré-m endogènes observés (Tanackovic et Kramer, 2005), tout en conduisant à une létalité cellulaire. D'autre part, le facteur SF1 régule la définition d'un micro exon du gène codant la troponine T cardiaque de poulet (Carlo *et al.*, 2000). La protéine SF1

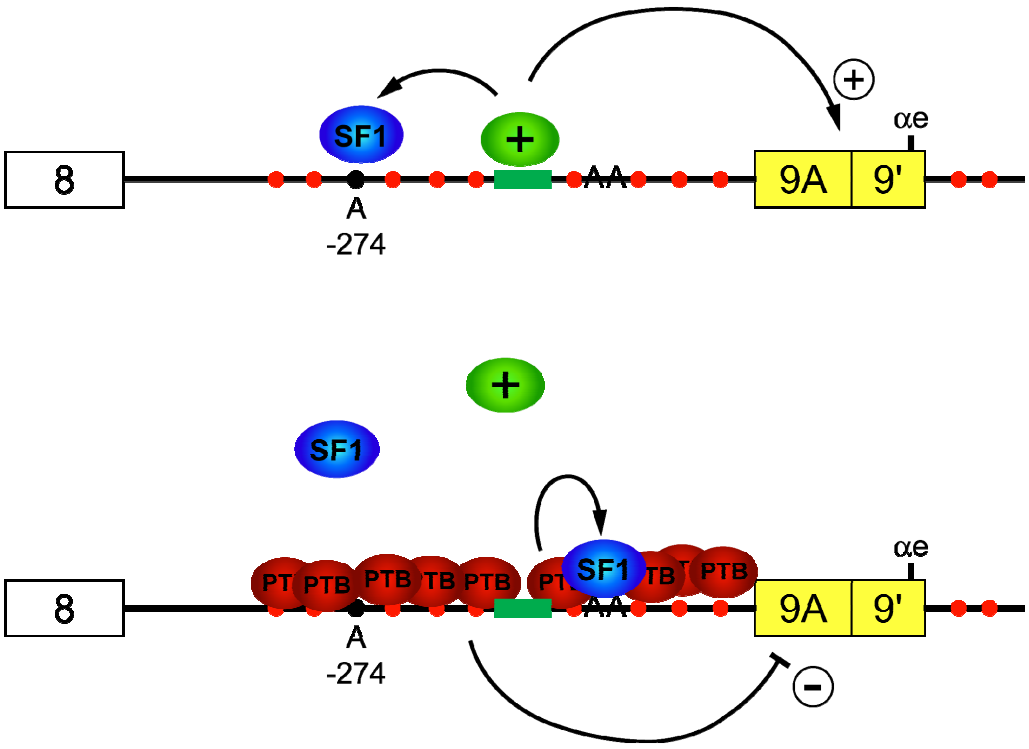


Figure 52- Modèle de recrutement de la protéine SF1 en présence de la PTB. La PTB pourrait empêcher la fixation de facteurs activateurs (représentés par +) sur la séquence UTE (trait vert) qui favoriseraient la fixation de la protéine SF1 sur le point de branchement localisé en -274. La fixation de la PTB sur ses sites de liaison (points rouges) pourrait indirectement participer au recrutement du facteur SF1 au niveau des A localisés en position -158 et -149 (représentés par AA). Le recrutement du facteur SF1 sur ces points de branchement putatifs pourrait conduire au blocage de la réaction d'épissage utilisant le point de branchement situé en -274.

participerait donc à la régulation de l'épissage alternatif en favorisant l'assemblage du spliceosome (Guth et Valcarcel, 2000). Le rôle du facteur SF1 dans la régulation de l'exon 9A9' a déjà été abordé au laboratoire en ciblant son expression dans les embryons de xénope à l'aide d'un morpholino (Le Sommer, 2006). La déplétion de ce facteur conduit à un important retard de développement embryonnaire, et la co-injection du morpholino et des minigènes placés sous le contrôle des promoteurs actine cardiaque et kératine diminue fortement la quantité d'ARNm issus de ces minigènes (Le Sommer, 2006). De plus, cette déplétion modifie la proportion des isoformes α7 et α2Δ9A : dans les somites, la quantité relative d'ARNm α2Δ9A augmente de cinq fois, tandis que dans l'épiderme l'ARNm α7 devient la forme majoritaire (Le Sommer, 2006). Ces résultats indiquent que le facteur SF1 pourrait intervenir dans la régulation de l'exon 9A9'. Bien que l'identification d'une protéine par LC-MS/MS ne permette pas de l'impliquer dans un mécanisme de régulation, les données que nous avons obtenues montrent un lien entre la protéine SF1 et la PTB. Le facteur SF1 interagit spécifiquement avec les sites de branchement. De plus, la fixation de la PTB masque le point de branchement. La PTB pourrait donc permettre le recrutement de la protéine SF1 au niveau des A localisés en position -158 et -149. L'analyse de ces deux A par bioinformatique les caractérise comme des sites de branchement potentiels. Le recrutement du facteur SF1 sur ces points de branchement putatifs pourrait conduire au blocage de l'utilisation du point de branchement situé en -274 sélectionné *in vivo* dans les somites par le spliceosome. A l'inverse, en absence de PTB, la fixation de facteurs activateurs sur la séquence UTE favoriserait la fixation de la protéine SF1 sur le point de branchement localisé en -274, favorisant ainsi l'activation de l'exon (figure 52). Il serait donc intéressant de réaliser les mêmes expériences de purification des complexes RNP en utilisant un ARN plus long comportant le point de branchement. Si cette hypothèse est juste, la protéine SF1 devrait être présente en absence de PTB uniquement quand la séquence UTE et le point de branchement en -274 sont accessibles. Nous envisageons également de réaliser des expériences de cross-link des complexes en utilisant l'AMT-psoralène afin de détecter si la snRNPU2 est associée au point de branchement A-274 de manière UTE dépendante.

Tout comme SF1, le facteur ESRP2 pourrait être impliqué dans la répression de l'exon 9A9'. Cette protéine, et son paralogue ESRP1 sont exprimés spécifiquement dans les cellules épithéliales. ESRP2 est impliqué dans la régulation différentielle des exons IIIb et IIIc de l'ARNpré-m de FGFR2 (Warzecha *et al.*, 2009). Dans les cellules mésenchymateuses, la

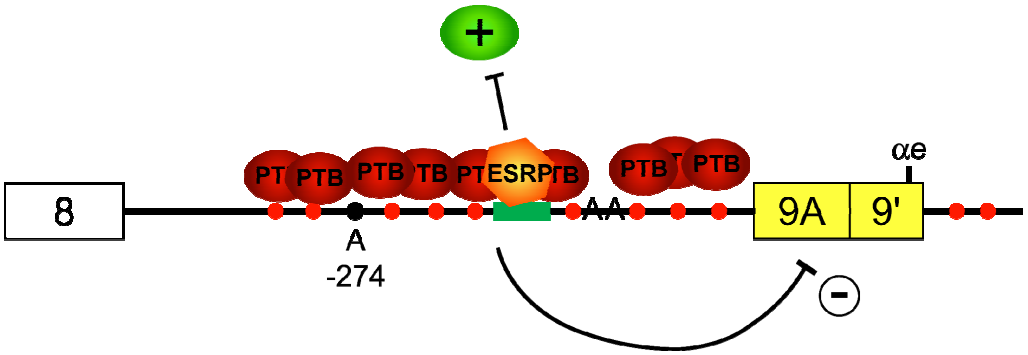


Figure 53- Modèle de blocage de l'exon 9A9' par les facteurs ESRP. Présents spécifiquement dans les cellules épithéliales, les facteurs ESRP1 et ESRP2 pourraient se fixer sur la séquence UTE (trait vert), empêchant ainsi avec la PTB le recrutement des facteurs activateurs (représentés par +), empêchant ainsi l'utilisation de l'exon 9A9'. Les sites de fixation de la PTB sont représentés par des points rouges et le point de branchement par « A-274 ».

fixation de la PTB sur des séquences introniques régulatrices localisées de part et d'autres de l'exon IIIb réprime l'utilisation de cet exon (Carstens *et al.*, 2000; Wagner et Garcia-Blanco, 2002). Dans les cellules épithéliales, l'exon IIIb est activé notamment par ESRP2 qui réprime également l'exon IIIc (Warzecha *et al.*, 2009). Les protéines ESRP2 et ESRP1 contrôlent également la maturation d'autres ARNpré-m régulés différentiellement selon leur expression dans des cellules épithéliales ou mésenchymateuses. Ainsi, elles favorisent l'inclusion d'exons alternatifs situés entre les exons 5 et 15 de CD44, ou l'inclusion de l'exon 11a de l'ARNpré-m hMena (Warzecha *et al.*, 2009). A l'inverse, elles réprimeraient l'utilisation d'exons alternatifs localisés entre les exons 1 et 5 de l'ARNpré-m de la p120-catenine (Warzecha *et al.*, 2009). Le mode d'action et les séquences de liaison des protéines ESRP sur l'ARN n'ont pas été précisément caractérisés. Toutefois dans l'ARN FGFR2, elles se lient à un élément intronique régulateur riche en dinucléotides GU et UG (Warzecha *et al.*, 2009). La séquence de l'élément UTE est UGGAUGG, directement suivie par les nucléotides UGGUU. De part sa fixation au niveau de la séquence UTE, la protéine ESRP2 pourrait, avec la PTB, participer au blocage de l'élément UTE réprimant l'utilisation de l'exon 9A9' dans l'épiderme (figure 53). ESRP2 n'étant pas présente dans le muscle, la séquence UTE serait alors en partie accessible et pourrait donc favoriser l'activation de l'exon 9A9'. Des expériences *in vitro* de transfert de marquage permettraient de déterminer les sites de fixation de ESRP2 sur l'ARN. Afin de tester l'implication des protéines ESRP *in vivo*, il est possible de cibler leur expression à l'aide de morpholinos. Dans un premier temps, nous pourrons tester les effets de leur déplétion à l'aide de morpholino sur la maturation de l'extrémité 3' terminal de l'ARNpré-m endogène de l'α-tropomyosine dans les tissus non musculaires. Selon les résultats obtenus, il sera possible de mettre en place une approche consistant à cibler l'expression de ces protéines dans les cellules musculaires à l'aide de techniques de transgénèse afin de déterminer si leur présence provoque une diminution de l'utilisation de l'exon 9A9'. L'étude *in vivo* des protéines ESRP semble particulièrement importante. A ce jour, ces protéines sont les seuls facteurs dont l'expression tissu-spécifique peut être reliée à la régulation différentielle de l'utilisation de l'exon 9A9' par l'intermédiaire d'une liaison à la séquence UTE.

La protéine hnRNP M pourrait entrer en compétition avec ESRP2 pour la fixation sur la séquence UTE. En effet, la protéine hnRNP M a de l'affinité pour les séquences poly(G), poly(U) (Datar *et al.*, 1993) et pour les séquences GU (Hovhannisyan et Carstens, 2007). Ces séquences de fixation sont donc compatibles avec une liaison sur la séquence UTE

(UGGAUGG). Le recrutement de cette protéine sur des séquences introniques conduit à la répression de l'épissage de l'exon IIIc de FGFR2, ou au contraire à l'inclusion de l'exon R35 de la troponine T cardiaque (Cooper, 1999). La protéine étant présente dans les complexes uniquement en absence de PTB, elle favoriserait le recrutement d'un complexe activateur au niveau de l'élément UTE. Son action pourrait être coopérative avec les protéines hnRNP K et L également identifiées dans les complexes en absence de PTB. Le site de fixation ($AUC_{3/4}$(U/A)(A/U) (Thisted *et al.*, 2001)) de la protéine hnRNP K n'est pas présent dans l'ARN UTElwt, par contre elle pourrait interagir avec la protéine hnRNP L (Kim *et al.*, 2000) qui se fixerait sur les deux répétitions CCAA espacées de cinq nucléotides présentes dans l'ARN UTEl (Hui *et al.*, 2005). Tout comme la protéine hnRNP M, ces deux hnRNP peuvent activer ou réprimer l'épissage alternatif. La fixation de la protéine hnRNP K sur un ISE favorise l'utilisation de l'exon 6A de la β-tropomyosine (Expert-Bezancon *et al.*, 2002), et celle de hnRNP L peut conduire à l'exclusion exonique, comme dans l'ARNpré-m de CD45 (House et Lynch, 2006), ou au contraire à l'inclusion d'un exon de l'ARNpré-m de eNOS (Hui *et al.*, 2003). Deux études récentes, consistant à inhiber l'expression de la protéine hnRNP L dans des cellules en culture, ont montré leur implication dans la régulation de l'épissage alternatif d'un ensemble de 18 gènes (Hung *et al.*, 2008; Venables *et al.*, 2008). Ces régulations concernent la sélection des sites 5' et 3' d'épissage et de signaux de clivage/polyadénylation (Hung *et al.*, 2008). Son mode d'action n'a toujours pas été clairement identifié. Cependant la présence de répétitions C/A, site de fixation de la protéine hnRNP L, à proximité d'un site d'épissage peut inhiber la définition de l'exon comme cela a été décrit dans l'ARNpré-m de CD45 où hnRNP L se fixe sur un ESS proche du site 3' d'épissage et inhibe la définition de l'exon 4 (House et Lynch, 2006), alors que la présence de sites de fixation éloignés du site 5' d'épissage favoriserait l'utilisation de l'exon aval (Hui *et al.*, 2005).

Selon le modèle d'activation proposé par Hui et collaborateurs (Hui *et al.*, 2005), on peut imaginer que la fixation de hnRNP L sur un élément de séquence (présent dans l'ARN UTElwt) éloigné du site 5' d'épissage de l'exon 8 puisse favoriser l'inclusion de l'exon 9A9'. Cette activation pourrait être renforcée par le recrutement de hnRNP K. En parallèle, il pourrait exister une compétition de fixation entre la protéine hnRNP M et les protéines ESRP

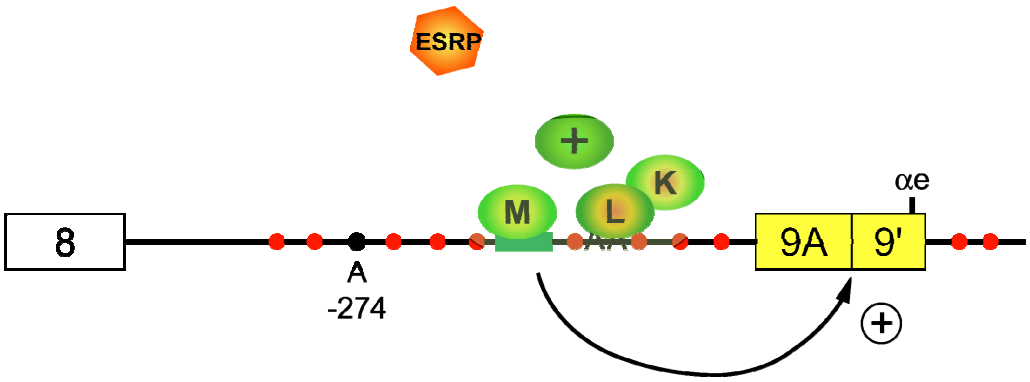


Figure 54- Modèle de l'activation de l'exon 9A9' par les protéines hnRNP K, L et M en absence de PTB. En absence de fixation de PTB sur ses sites (points rouges), la protéine hnRNP M (M) pourrait se fixer sur la séquence UTE (trait vert) empêchant ainsi le recrutement des facteurs ESRP et permettant le recrutement de facteurs activateurs (représentés par +). La fixation de la protéine hnRNP L (L) sur des séquences C/A, elle-même pouvant recruter la protéine hnRNP K (K), pourrait favoriser l'utilisation de l'exon 9A9'. La fixation de ces trois hnRNP pourrait favoriser la formation d'un complexe activateur (représenté par un cercle vert). Le point de branchement est indiqué par « A-274 ».

pour la reconnaissance de l'élément UTE. Ainsi, en absence de la PTB le facteur ESRP2 ne serait plus recruté sur la séquence UTE autorisant alors la liaison de hnRNP M. La fixation de hnRNP M sur l'UTE permettrait ou favoriserait directement l'assemblage d'un complexe activateur sur cet élément. Ces différentes hypothèses sont représentées dans la figure 54. Il serait intéressant de réaliser des expériences de transfert de marquage pour analyser les sites de fixation de ces trois protéines hnRNP. Nous pourrions également étudier l'effet de la déplétion de ces trois hnRNP sur la maturation de l'exon 9A9' dans les embryons de xénope.

Une dernière protéine appartient aux complexes assemblés sur l'ARN UTElwt en absence de PTB. La protéine 2 associée à U4/U6.U5 (ou 65 kDa SR-related protein) est nécessaire au recrutement de la tri-snRNP (Makarova *et al.*, 2001) et est présente dans le complexe B du spliceosome (Deckert *et al.*, 2006). Le recrutement de la tri-snRNP U4/U6.U5 est dépendant de la fixation au préalable de la snRNP U2 au niveau du point de branchement. L'ARN UTElwt ne contient ni le site 3' d'épissage, ni la séquence du point de branchement autour du A-274, le recrutement de la snRNPU2 par la séquence UTE est donc peu probable. L'analyse de l'intégralité des protéines purifiées à partir de l'ARN UTElwt en présence ou en absence de PTB permettrait de déterminer si d'autres facteurs associés à la snRNP U2 ou à la tri-snRNP sont également recrutés. Nous pourrions ainsi évaluer le rôle de la séquence UTE dans l'assemblage du complexe B du spliceosome mis en place autour de l'intron 8.

Nous avons donc mis au point une technique permettant de purifier et d'identifier les protéines des complexes formés sur différents ARN *in vitro*. Cette approche relativement simple à mettre en place présente toutefois un certain nombre de limites.

Tout d'abord, les ARN que nous avons utilisés sont relativement courts, ils sont donc d'autant plus sensibles à une modification de leur séquence. En effet, nos expériences montrent qu'une déstabilisation ou une modification de la structure secondaire peut conduire à un recrutement de protéines différentes. Il sera donc intéressant de voir s'il est possible d'utiliser des ARN de plus grande taille, pour lesquelles une mutation de séquence n'entraînerait pas de modification de la structure secondaire de l'ARN.

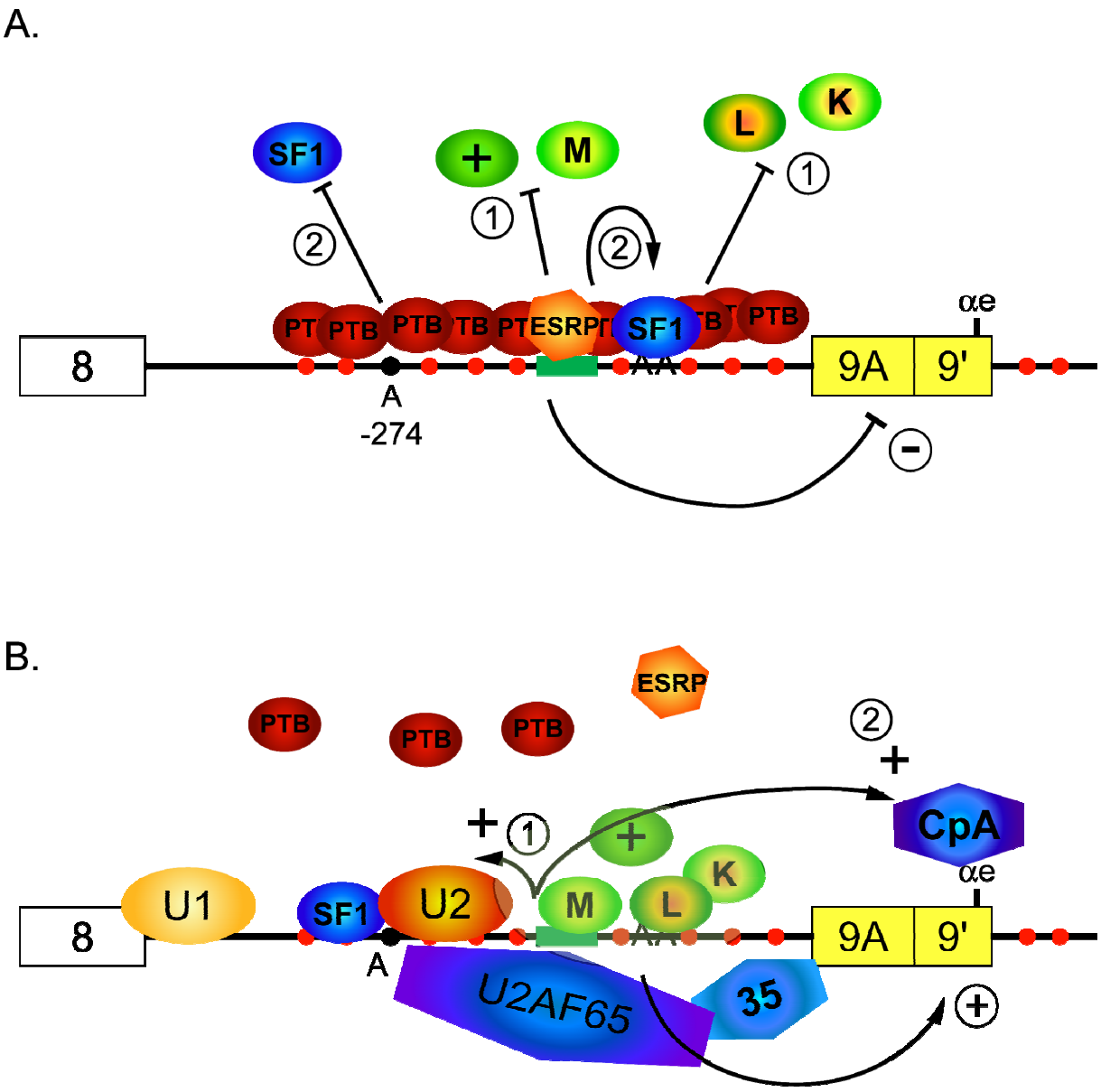


Figure 55- Modèle de régulation de l'exon 9A9', relation entre la PTB et l'UTE. A. en présence de PTB, les facteurs ESRP seraient recrutés sur l'UTE (trait vert), conduisant ainsi au masquage complet de l'UTE. ① Les facteurs activateurs ne pourraient donc pas être recrutés, il pourrait s'agir des protéines hnRNP M (M), hnRNP L (L) et hnRNP K (K), et d'autres protéines activatrices comme les protéines SR. ② le facteur SF1 ne serait pas adressé au point de branchement (A-274), mais serait retenu sur une autre région de l'ARN. L'ensemble de ces mécanismes conduit au blocage de la définition de l'exon 9A9' comme exon terminal. **B.** En absence de PTB la protéine hnRNP M est fixée au niveau de la séquence UTE, et la protéine hnRNP L sur les CA situés en aval de l'UTE. La fixation de ces protéines permet indirectement ou non la formation d'un complexe activateur, ① favorisant le recrutement de la snRNP U2 au niveau du point de branchement qui à son tour ② stabiliserait, par une interaction directe, la fixation de facteurs de polyadénylation (CpA). L'exon 9A9' serait alors défini comme exon 3' terminal.

Ensuite, les complexes que nous avons purifiés sont reconstitués autour de l'ARN fixé sur une matrice. Il est possible que l'assemblage des protéines sur un ARN immobilisé soit différent de celui qui serait obtenu autour du même ARN en solution.

Enfin, la reconstitution des complexes sur des fragments d'ARN de petite taille permet l'étude des facteurs assemblés précocement au cours de la mise en place du spliceosome. Pour analyser les complexes d'épissage il serait nécessaire de disposer d'ARN de plus grande taille possédant les signaux d'épissage et de clivage/polyadénylation.

2.5 Conclusions

Ces expériences nous ont permis d'identifier les protéines SF1, ESRP2, hnRNP K ; L ; M comme étant assemblés sur l'ARN UTElwt de manière sélective, avec SF1 et ESRP2 plutôt associées au complexe répresseur et les protéines hnRNP K, L et M plutôt associées au complexe activateur. Ces résultats ainsi que ceux obtenus *in vivo* suite au blocage de l'accessibilité de l'UTE et des séquences permettant la définition de l'exon 9A9' suggèrent que la régulation tissulaire de l'exon 9A9' est basée sur une compétition de fixation entre des facteurs activateurs et inhibiteurs sur la séquence régulatrice UTE (figure 55).

En présence de PTB (figure 55 A.), les facteurs ESRP seraient recrutés sur l'UTE (trait vert), conduisant ainsi au masquage complet de l'UTE. Les facteurs activateurs ne pourraient donc pas être recrutés. Le facteur SF1 ne serait pas adressé au point de branchement (A-274), mais serait retenu sur une autre région de l'ARN. L'ensemble de ces mécanismes conduit au blocage de la définition de l'exon 9A9' comme exon terminal.

En absence de PTB (figure 55 B), la protéine hnRNP M serait fixée au niveau de la séquence UTE, et la protéine hnRNP L sur les CA situés en aval de l'UTE. La fixation de ces protéines permettrait indirectement ou non la formation d'un complexe activateur, favorisant le recrutement de la snRNP U2 au niveau du point de branchement qui à son tour stabiliserait le complexe de clivage/polyadénylation sur le site de polyadénylation associé à l'exon 9A9'. Par la suite, le complexe mis en jeu pourrait stabiliser les facteurs impliqués dans la reconnaissance du site 3' d'épissage tels que U2AF65 et U2AF35. L'exon 9A9' serait alors défini en tant qu'exon 3' terminal. Il serait intéressant de tester l'implication de la protéine SRm160 pour laquelle un rôle dans le couplage entre les processus d'épissage et de clivage/polyadénylation a été décrit (McCracken *et al.*, 2002). Ces différents mécanismes

seront déterminés par les approches présentées précédemment combinant l'extinction de l'expression de ces protéines *in vivo* à l'aide de morpholinos, et *in vitro* par des approches biochimiques de purification de complexes. Afin de déterminer l'ordre d'assemblage de ces différents complexes, il sera intéressant d'analyser la composition des complexes assemblés dans l'extrait nucléaire dans des conditions où l'assemblage du spliceosome est bloqué au niveau du complexe E (en absence d'ATP) ou au niveau du complexe A (en absence de la snRNP U1).

Conclusion générale

Le choix d'un exon 3' terminal implique la sélection d'un site de polyadénylation couplée à celle d'un site 3' d'épissage. Il y a donc une régulation concertée des processus d'épissage et de clivage/polyadénylation de l'ARNpré-m en lien direct avec la transcription. Nous avons étudié *in vivo* la maturation de la région 3' terminale de l'ARNpré-m de l'α-tropomyosine. Cet ARNpré-m contient trois exons 3' terminaux alternatifs, les exons 9A9', 9B et 9D, qui sont utilisés différentiellement en fonction du contexte tissulaire. Ces trois exons possèdent chacun un signal de polyadénylation appelé respectivement αe, αa et αnm. Nous nous sommes intéressés plus particulièrement au contrôle de l'exon le plus proximal, l'exon composite interne/3' terminal 9A9' dont la sélection en tant qu'exon terminal empêche l'utilisation des deux exons localisés en aval.

Les études que nous avons réalisées *in vivo* dans l'embryon de xénope ont mis en évidence une séquence appelée UTE (*Upstream Terminal exon Enhancer*) impliquée dans la définition de l'exon 9A9' en tant qu'exon terminal dans les cellules musculaires. En bloquant l'accessibilité de l'élément UTE, nous avons démontré qu'il active la définition du site 3' d'épissage et du signal de polyadénylation αe, et favorise l'utilisation du point de branchement distal associé à l'exon 9A9' dans les cellules musculaires. Nous avons également montré que la PTB prévient l'utilisation de l'UTE dans les cellules non musculaires à l'inverse de certaines protéines SR qui favorisent la sélection de l'exon 9A9' en tant qu'exon 3' terminal de manière dépendante de cette séquence. De plus, les approches que nous avons développées *in vitro* ont permis d'identifier des protéines spécifiquement assemblées sur un ARN contenant la séquence UTE de manière dépendante de la présence de la PTB.

Premièrement, les protéines SR SC35, SRp40 et SRp55 activent l'utilisation de l'exon terminal 9A9' lorsqu'elles sont exprimées dans l'ovocyte, dans lequel cet exon est normalement exclu des ARNm α-tropomyosine. Nous avons montré que l'élément UTE est nécessaire à cette activation mais nous n'avons pas pu mettre en évidence d'interaction directe *in vitro* entre ces protéines et un ARN contenant l'élément UTE. Des expériences récentes effectuées à l'aide de cellules en culture montrent que les protéines SR interagissent avec les ARNpré-m en cours de synthèse, et que cette interaction permet leur recrutement sur la chromatine. De plus, chaque protéine SR est recrutée sur le gène modèle *FOS* de manière

dépendante d'une transcription active (Sapra *et al.*, 2009). La protéine SR SC35 coopérerait avec l'ARNpol II pour favoriser l'élongation de la transcription (Fededa et Kornblihtt, 2008; Lin *et al.*, 2008). Enfin, le CTD de l'ARNpolII peut fixer certaines protéines SR (Das *et al.*, 2007). Dans ces expériences réalisées *in vitro*, il a été montré que les protéines doivent être chargées sur l'ARNpol II pour qu'il y ait une activation de l'épissage (Das *et al.*, 2007). Les facteurs ainsi « portés » par l'ARNpol II seraient ensuite adressés à l'ARNpré-m pour effectuer leur régulation. Dans les cellules musculaires, l'UTE pourrait ainsi recruter certains facteurs « transportés » par l'ARNpol II activant alors l'utilisation de l'exon 9A9'. A l'inverse, dans les cellules non musculaires, l'UTE étant masqué par la PTB, ces facteurs ne pourraient pas se fixer, l'exon 9A9' ne serait donc pas utilisé. Cette hypothèse permettrait d'expliquer pourquoi les protéines SR SRp40 et Srp55, qui interagissent avec le CTD de l'ARNpol II (Das *et al.*, 2007), conduisent à l'utilisation de l'exon 9A9' *via* l'UTE, sans que nous ayons pu observer *in vitro* de fixation de ces protéines sur un ARN contenant la séquence UTE. « L'adressage » des protéines par le complexe de transcription permettrait également d'expliquer leur recrutement sur l'ARN malgré la présence des protéines de liaison aux ARN, qui sont très abondantes dans le noyau et qui se fixent sur les ARNpré-m naissants. Il serait donc intéressant de purifier les complexes contenant l'ARNpol II et l'ARN puis d'observer si ces protéines sont recrutées au cours de l'élongation de la transcription. Nous pourrions également déterminer quelles protéines SR sont recrutées sur l'ARNpol II dans un système *in vitro* permettant le couplage entre la transcription et l'épissage d'un ARN contenant l'exon 8, l'intron 8-9A et l'exon 9A9'.

Deuxièmement, nous avons utilisé les extraits nucléaires de cellules HeLa pour caractériser les partenaires protéiques qui interagissent avec l'élément UTE. Ces extraits protéiques constituent un environnement cellulaire comparable à celui du noyau des ovocytes. Nous avons observé que la quantité de protéine PTB, présente dans l'extrait nucléaire, influence la nature du complexe qui s'assemble autour de l'UTE. Nos résultats montrent que la protéine ESRP2 est recrutée sur cette séquence, et que suite à une diminution de la quantité de PTB, les protéines hnRNP K ; L et M s'assemblent autour de la séquence UTE. Ces trois protéines sont exprimées dans tous les tissus contrairement à la protéine ESRP2 qui est exprimée spécifiquement dans les cellules épithéliales. Cette observation soulève la question de l'implication de facteurs tissu-spécifiques dans la régulation de l'épissage alternatif, plus particulièrement dans la maturation de l'exon 9A9'. Quelques protéines spécifiques d'un tissu ont été décrites dans la régulation de l'épissage alternatif. Chez la drosophile, les protéines

TRA et SXL sont impliquées dans la détermination du sexe qui repose sur une cascade de régulations spécifiques de l'épissage alternatif (Black, 2003; Forch et Valcarcel, 2003). Chez les vertébrés les protéines NOVA-1, nPTB, smPTB, et plus récemment les protéines ESRP ont été montrées comme étant impliquées dans les régulations de l'épissage alternatif spécifique d'un tissu (Ule *et al.*, 2005; Coutinho-Mansfield *et al.*, 2007; Warzecha *et al.*, 2009). De plus, des modifications post-traductionnelles de certaines protéines sont spécifiques du développement ou de voies de signalisation et permettent de moduler l'action de facteurs régulateurs. Ainsi, des protéines ubiquitaires peuvent voir leur activité modulée en fonction de leur environnement. Le niveau phosphorylation du domaine RS des protéines SR, en réponse à des voies de signalisation, suite à un stress, ou au cours du développement, peut être à l'origine d'une régulation de l'épissage alternatif (Mermoud *et al.*, 1994; Prasad *et al.*, 1999; Shi et Manley, 2007; Zhong *et al.*, 2009). La protéine ESRP2 semble donc être un bon facteur candidat car elle pourrait participer à la régulation tissulaire de l'exon 9A9'. Afin de déterminer quels sont les différents facteurs impliqués dans la régulation tissu-spécifique de l'exon 9A9', il sera important de déterminer la fonction de ESRP2 *in vivo* dans les embryons de xénope, en inhibant sa production dans les tissus non musculaires et en favorisant son expression dans les somites. Il sera également intéressant d'étudier les modifications post-traductionnelles des protéines SR, dont une modification de l'état de phosphorylation pourrait également être à l'origine d'un recrutement sur la séquence UTE.

Ces hypothèses concernant le mode d'action des protéines SR sont essentiellement basées sur le couplage entre les processus de maturation des ARNpré-m et la transcription. Notre modèle d'étude *in vivo* nous permet d'avoir accès à un tel couplage, c'est pourquoi je propose les hypothèses suivantes :

Premièrement il pourrait y avoir un lien entre la séquence UTE et l'élongation de la transcription. En effet, une hypothèse simple serait que dans les somites, l'utilisation du signal proximal αe soit favorisée par l'apparition tardive des deux signaux distaux αa et αnm, éloignés respectivement de 4,5 kb et 9,5 kb du fait de la grande taille des introns 9A9'-9B et 9B-9D. D'autre part, la taille importante de l'intron 9A9'-9B suggère qu'il existerait des éléments permettant de ralentir l'élongation de la transcription, comme l'élément MAZ4 (Ashfield *et al.*, 1991; Roberts *et al.*, 1998; Robson-Dixon et Garcia-Blanco, 2004) situé en aval du signal de polyadénylation du complément C2. Une pause en aval de l'exon 9A9' éviterait, dans les somites, que sa reconnaissance soit en compétition avec celle des exons

suivants. Des expériences d'immunoprécipitation de la chromatine à l'aide d'anticorps dirigés contre l'ARNpol II permettraient de déterminer si elle effectue une pause en aval de l'exon 9A9'.

Deuxièmement, il existerait un lien entre la terminaison de la transcription et l'élément UTE. Cette hypothèse est d'autant plus intéressante que l'exon 9A9' est un exon 3' terminal et que la terminaison de la transcription par l'ARNpol II est consécutive au clivage des ARNpré-m qui nécessite la reconnaissance d'un signal de polyadénylation (West *et al.*, 2004; West *et al.*, 2008) .Trois mécanismes possibles sont discutés ici.

1- La séquence UTE favorise la reconnaissance du site 3' d'épissage et du signal αe dans les somites. Ainsi, en recrutant un complexe activateur, l'élément UTE pourrait favoriser la fixation de CPSF au niveau de l'hexanucléotide AAUAAA. L'hexanucléotide est responsable à lui seul de pauses transcriptionnelles précédant l'évènement de terminaison (Orozco *et al.*, 2002). Cette pause en indépendante de CstF (Nag *et al.*, 2006) mais nécessite une interaction entre CPSF et l'ARNpol II (Nag *et al.*, 2007). Dans ce cas, suite au clivage de l'ARNpré-m au niveau du signal αe, l'exonucléase 5' → 3' Xrn2 pourrait rapidement « torpiller » l'ARNpol II (Gromak *et al.*, 2006), induisant l'arrêt de la transcription avant même que les exons 9B et 9D ne soient transcrits. Cette hypothèse pourrait être testée par expériences d'immunoprécipitation de la chromatine à l'aide d'anticorps dirigés contre l'ARNpol II, couplées à des analyses PCR quantitative. Les quantités relatives des sites de polyadénylation αe et αa associés à l'ARNpol II seront comparées dans des conditions où l'UTE est accessible ou bloqué par la présence de morpholino.

2- Il a été montré dans le cas de l'ARNpré-m de la β-globine, qu'une séquence additionnelle appelée « séquence terminatrice », située en aval du site de polyadénylation, est nécessaire à cette étape de clivage (Dye et Proudfoot, 2001). Cette séquence permet le clivage et la terminaison de la transcription suite à l'utilisation d'un signal de polyadénylation faible même s'il est en compétition avec un site distal plus fort (West et Proudfoot, 2009). L'étude de l'épissage d'ARNpré-m β-globine dérivés de minigène contenant ou non la séquence « terminatrice » montre que des ARN non épissés s'accumulent lorsque la terminaison de la transcription est inefficace, ce qui suggère une influence de la terminaison de la transcription sur l'épissage (West et Proudfoot, 2009). Dans le contexte de l'ARNpré-m de l'α-tropomyosine, on peut faire l'hypothèse que le rôle de la séquence UTE est de renforcer le site

de polyadénylation αe afin que la terminaison de la transcription se fasse de manière efficace au niveau de ce site, permettant ainsi l'épissage de l'exon 8 à l'exon 9A9'. Le blocage de la fonction de l'UTE serait donc responsable d'une diminution de l'efficacité du clivage au niveau du signal αe, entraînant ainsi une terminaison inefficace de la transcription et l'accumulation des ARN au niveau des sites de transcription. Ces ARN seraient alors dégradés par l'exosome nucléaire (Bousquet-Antonelli *et al.*, 2000; Milligan *et al.*, 2005; Houseley *et al.*, 2006). Cette hypothèse permettrait d'expliquer que le blocage de la fonction de l'UTE s'accompagne d'une baisse générale des ARN produits à partir du gène de l'α-tropomyosine. Des expériences d'immunoprécipitation de la chromatine nous permettraient d'étudier l'effet du blocage de la fonction de l'UTE sur l'association de l'exon 9A9' à l'ARNpol II et sur la capacité de l'ARNpol II à prolonger l'étape d'élongation au-delà du site de polyadénylation αe. Nous pourrons ainsi évaluer l'implication de l'UTE dans la terminaison de la transcription.

3- Nous devons également considérer le fait que la terminaison de la transcription, ainsi que la maturation de l'extrémité 3' et l'épissage des ARNpré-m, sont liés à l'initiation de la transcription. Le domaine CTD de l'ARNpol II joue un rôle important dans le couplage entre ces différentes étapes. De nombreux facteurs impliqués dans le processus de clivage/polyadénylation sont associés aux régions promotrices des gènes (Calvo et Manley, 2003). Chez la levure, le facteur ssu72 a une double fonction. En effet, celui-ci fait partie du complexe de polyadénylation et a une activité phosphatase spécifique des sérines 5 du domaine CTD de l'ARNpol II (Reyes-Reyes et Hampsey, 2007). Ceci suggère qu'il existe une communication entre les complexes d'initiation et de terminaison de la transcription des gènes. Des expériences d'immunoprécipitation de la chromatine montrent que ces facteurs sont progressivement recrutés par le complexe d'élongation le long du gène en cours de transcription, et que leur concentration augmente dans la partie 3' des gènes (Kim *et al.*, 2004). Certains auteurs relient ce recrutement progressif de l'extrémité 5' vers l'extrémité 3' au changement d'état de phosphorylation de la partie carboxy terminale de l'ARNpolII (Kim *et al.*, 2004). De plus, chez la levure, il existe un lien physique entre les régions promotrice et terminale des gènes ce qui faciliterait la ré-initiation de la transcription par l'ARNpol II lorsqu'un cycle de transcription se termine (O'Sullivan *et al.*, 2004; Ansari et Hampsey, 2005; Singh et Hampsey, 2007). Il est possible que l'élément UTE serve de plateforme pour le chargement, sur l'ARNpol II, de facteurs permettant la ré-initiation de la transcription. Dans ce cas, le blocage de l'élément UTE conduirait à une diminution de la transcription. Des

expériences de co-immunoprécipitation de la chromatine couplées à celle de l'ARNpol II nous permettraient de voir quels sont les facteurs associés à la polymérase dans les somites et si cette association persiste lorsque l'accessibilité de l'élément UTE est bloquée.

La régulation de l'épissage alternatif de l'exon 3' terminal 9A9' est donc un processus complexe qui fait intervenir de multiples facteurs impliqués de la synthèse à la fin de la maturation de l'ARN de l'α-tropomyosine.

Ces différentes régulations conduisent à la synthèse des ARNm α2 et α7, qui diffèrent uniquement par leur 3' UTR et codent pour une même isoforme musculaire de 248 acides aminés. Dans l'embryon, l'ARNm α7 représente 80 % des transcrits alors que dans le muscle strié adulte seul l'ARNm α2 est exprimé. Quel est l'intérêt pour un organisme de sélectionner un mécanisme permettant de produire deux ARNm qui ne diffèrent que par leur 3' UTR ?

Les 3' UTR ont des fonctions connues dans la stabilité, la localisation et la traductibilité des ARNm (Moore, 2005). Ces différents rôles sont le fait d'interactions avec des protéines régulatrices et des miRNA. Les 3' UTR sont la cible de nombreuses protéines de liaison aux ARN comme les protéines de la famille CELF qui peuvent contrôler la traduction et la stabilité d'ARNm (Kuersten et Goodwin, 2003). La protéine CUGBP1 régule la stabilité des ARNm en se fixant sur des régions riches en GU (Barreau *et al.*, 2006; Marquis *et al.*, 2006; Vlasova *et al.*, 2008). L'exon 9' contenant huit répétitions de ce type, il serait intéressant de déterminer s'il est la cible de CUGBP1.

Ces régions 3' terminales sont également la cible de microARN. Des expériences réalisées sur des cellules T ont permis de montrer que l'activation de leur prolifération favorise l'utilisation des sites de clivage/polyadénylation proximaux des ARNpré-m présentant des sites de polyadénylation alternatifs (Sandberg *et al.*, 2008). L'analyse des banques de données a également établi une corrélation entre la prolifération et le raccourcissement de la 3' UTR dans différents types cellulaires et tissulaires chez l'homme (Sandberg *et al.*, 2008). Les auteurs ont prouvé expérimentalement que le raccourcissement de la 3' UTR permet aux ARNm d'échapper à la régulation par les miRNA. D'autres études montrent l'implication des miRNA miR-1 et miR-133 respectivement dans la différenciation et dans la prolifération du muscle strié squelettique (Chen *et al.*, 2006). Ces expériences suggèrent que les miRNA participent aux régulations post-transcriptionnelles qui contrôlent l'expression des gènes dans le muscle et au développement embryonnaire. Du fait de ces différentes données, un projet consistant à étudier les relations entre les miRNA et les 3' UTR

des ARNm α2 et α7 a été initié au laboratoire par Agnès Méreau et Hubert Lerivray. L'un des objectifs est de déterminer si la production de l'ARN α7, qui possède la 3' UTR la plus courte, permet d'échapper à une régulation par un ou plusieurs miRNA au cours de l'embryogenèse. Ce mécanisme permettrait, dans l'embryon, la production de l'α-tropomyosine, dont le rôle est de réguler en association avec le complexe troponine, la formation du complexe actomyosine impliqué dans le contrôle de la contraction musculaire, et qui est nécessaire au développement embryonnaire.

ANNEXE 1 - Matériels et méthodes

1. Matériels

1.1. Anticorps

Plusieurs anticorps ont été utilisés au cours de ces travaux :

- Les anticorps monoclonaux de souris anti-PTB (BB7) nous ont été fournis par le Pr C. Smith (Cambridge) et ont été utilisés dilués au 1/2000ème (Chou *et al.*, 2000).

- Les anticorps anti-xPTB, ont été produits par la société Eurogentec. Ces anticorps polyclonaux issus de lapin sont dirigés contre une région C terminale de la protéine PTB de Xénope. Ces anticorps ont été utilisés dilués au 1/500ème (Hamon *et al.*, 2004).

- Les anticorps monoclonaux de souris anti-SR (10H3) nous ont été fournis par le Dr J. Stevenin (IGBMC, Illkirch). Ils reconnaissent un épitope phosphorylé commun aux protéines SR, et ont été utilisés au 1/10ème (Roth *et al.*, 1990).

- Les anticorps monoclonaux de souris anti-U2AF65, fournis par le Pr J. Valcarcel (CRG, Barcelone) ont été utilisés au 1/100ème (Gama-Carvalho *et al.*, 1997).

1.2. Constructions disponibles

Le minigène pBSAct7-9Bwt a été réalisé au laboratoire. La région génomique couvrant les exons 7 à 9B de l'α-tropomyosine de xénope a été insérée en aval du promoteur actine cardiaque (Hamon Gouault, 2002). Ce minigène a été muté au niveau de la séquence UTE et du point de branchement en -274 et des A en -233 ; -196 ; -158 ; et -119 pour donner le minigène pBSAct7-9BmutBS (Hamon Gouault, 2002).

1.3. Extrait nucléaire et tampon

L'extrait nucléaire de cellules HeLa utilisé lors de nos expériences *in vitro* a été préparé par la société CILBIOTECH selon la méthode proposée par Dignam en 1983 (Dignam *et al.*, 1983).

Le tampon D, nécessaire à la grande majorité de nos expériences est composé de : EDTA 0,2 mM ; KCl 0,1 M ; Glycérol 20% ; DTT 0,5 mM ; Hepes/KOH 150 mM ; pH 7,2. Lorsqu'il entre dans la composition d'un milieu réactionnel, sa proportion est indiquée.

2. Méthodes

Une partie des méthodes utilisées est décrite dans la partie « *materials and methods* » de l'article Anquetil et al 2009. Il s'agit principalement des méthodes d'analyse des transcrits par protection à la ribonucléase (ou RPA), par transcription inverse et PCR (RT-PCR) et de l'analyse des protéines par western-blot. Ces techniques ne seront donc pas détaillées dans cette partie.

2.1. Matrices de transcription in vitro

Les matrices de transcription des ARN UTEc ont été préparées en hybridant un oligonucléotide antisens contenant la séquence du promoteur T7 suivie de la séquence UTE mutée ou non, avec un oligonucléotide sens correspondant à la séquence du promoteur T7 (tableau ci-dessous). Les deux oligonucléotides ont été placés à 10 mM dans un tampon TrisHCl 40 mM à pH 8, chauffés 5 min à 65°C et refroidis lentement jusqu'à retour à température ambiante. La matrice permettant de synthétiser l'ARN $(CUCUU)_8$ a été produite de la même manière.

Les matrices de transcription des ARN UTEl ont été préparées par PCR à partir des minigènes pBSAct7-9Bwt ou pBSAct7-9BmutBS, avec le couple d'oligonucléotides T7sens et UTEl anti-sens. Les séquences des différents oligonucléotides utilisés sont indiquées dans le tableau qui suivant :

Nom	Séquence 5' - 3'
UTEcwt T7	GTTGGAGACATTGGCAAGAAACCACCATCCAGGAAACACCCTATAGTGAGTCGTATTA
UTEcm T7	GTTGGAGACAATGGCAAGAAACCAGTATAAGGGAAACACCCTATAGTGAGTCGTATTA
T7 sens	TAATACGACTCACTATAG
UTEl sens	AAATTAATACGACTCACTATAGGGTCATCTTGTTTCC
UTEl antisens	AAATTTCAGAGGAAAGGG
T7 PTB	AAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGCCCTATAGTGAGTCGTATTA
Les séquences correspondant au promoteur T7 sont soulignées	

2.2. Transcription in vitro

10 pmol de matrice ADN ont été placés dans un mélange réactionnel : BSA 0,05 µg/µl ; ATP, CTP, GTP, UTP 4,8 mM chaque ; $MgCl_2$ 40 mM ; RNAsin (Promega) 10 U ; DTT 5 mM ; Spermidine 1 mM ; Triton X100 0,01% ; PEG 80 mg/ml ; TrisHCl 40 mM pH 8 ; T7 RNA polymérase (Promega) 160 U ; Pyrophosphatase 0,05 U/µl. Pour produire un transcrit radiomarqués 20 µCi d'UTP [α-^{32}P] ont été ajoutés.

Ce mélange a été incubé 12 heures à 37°C, traité à la DNase, les ARN ont été extraits au phénol/chloroforme avant d'être précipités puis purifiés sur un gel dénaturant de polyacrylamide 15%.

2.3. Déplétion en PTB des extraits nucléaires de cellules HeLa

Ces deux protocoles ont été adaptés des expériences décrites par Saulière et collaborateurs (Sauliere *et al.*, 2006).

2.3.1. Par chromatographie sur ARN immobilisés

50 pmol d'ARN $(CUCUU)_8$ ont été incubés dans 150 µl d'acétate de sodium 3,3 mM pH 5 et 1,5 mM métapériodate de sodium 30 min à 4°C dans l'obscurité, puis additionnés d'un volume de solution BACH 25 mM (EZ-Link Biotin hydrazine 25mM, dilué dans DMF/DMSO/eau (15,6% / 31,2% / 53,2%). Ce mélange a été incubé 30 min à température ambiante dans l'obscurité. Un volume d'une solution de sodium cyano-borohydride 100 mM et Phosphate de Sodium 0,2 M pH 8 a ensuite été ajouté et incubé 30 min à température ambiante. Les ARN biotinylés ont été purifiés sur colonne Miscrospin G25 (Roche) et précipités. Les billes de streptavidine-agarose ont été saturées avec un mélange BSA (500 µg/ml) et ARNt E.coli (100 µg/ml) dans le tampon WB50 (TrisHCl 50 mM pH 7.5 ; NaCl 50 mM ; NP40 0,01%) et incubées 15 min à 4°C sur roue. Les billes ont été lavées 1 fois avec le tampon WB100 (correspond au tampon WB50 à 100 mM NaCl). La fixation de l'ARN sur les billes (3 pmol ARN/µl de billes) a été effectuée dans le même tampon pendant 45 min à 4°C. Les billes ont ensuite été lavées trois fois avec du tampon D puis mises en présence d'extrait nucléaire (10µl billes/10µl d'extrait nucléaire) pendant 5 min à 4°C. L'extrait nucléaire est récupéré après centrifugation (5min, 1000g, 4°C) et peut être traité une ou deux fois supplémentaires avec des billes préalablement traitées de manière identique. La quantité de PTB présente dans les extraits nucléaires à la fin de la déplétion a été évaluée par western-blot.

2.3.2. Immunodéplétion

Les anticorps anti-PTB ((Ab-1) Mouse mAb (SH54) (CALBIOCHEM)) ont été fixés sur des billes magnétiques couplées à de la protéine G (Dynal). 25 µl de billes ont été lavées deux fois à l'aide d'un tampon citrate (Acide citrique 25 mM ; Na_2HPO_4 52 mM ; pH 5) puis incubées 12 heures à 4°C avec 5 µg d'anticorps. Deux lavages ont été effectués à l'aide du tampon citrate, puis les billes ont été saturées avec du tampon D additionné d'ARNt de levure

(0,1 mg/ml) et de BSA (0,5 mg/ml) 15 min à 4°C, lavées une fois avec du tampon D, et incubées 1 heure à 4°C sous agitation avec 50 µl d'extrait nucléaire de cellules HeLa. Cet extrait a été récupéré et placé sur des billes préalablement traitées de manière identique, puis incubé dans les mêmes conditions. L'extrait nucléaire alors récupéré a été analysé par western-blot afin d'évaluer l'efficacité de la déplétion.

2.4. Epissage in vitro

L'extrait nucléaire a été placé 10 min à 30°C afin d'épuiser l'ATP endogène. 2 pmoles d'ARN radiomarqué sont placées dans 13 µl de tampon d'épissage (ATP 2 mM ; Créatine Phosphate 48 mM ; $MgCl_2$ 4,8 mM ; RNAsin 10 U ; PVA 6% (v/v) ; tampon D 28%) additionné de 10 µl d'extrait nucléaire. Le mélange a été incubé 5 min à 4°C avant d'être placé 2 heures à 30°C. La réaction a été stoppée par ajout dans chaque tube de : 100 µl de tampon protéinase K 2X (Tris pH=7,4 0,2 M, EDTA 25mM, NaCl 0,1M, SDS 0,5%) ; 75 µl H_2O et 4 µl de protéinase K (5 mg/ml). Le mélange réactionnel a été incubé de 30 min à 30°C. Les ARN ont été récupérés par traitement au phénol/chloroforme et précipitation à l'acétate d'ammonium, puis séparés sur gel polyacrylamides 5% en conditions dénaturantes. Les ARN ont été visualisés à l'aide du Phosphorimager (Molecular Dynamics).

2.5. Retard de migration sur gel

5 pmoles d'ARN radiomarqués ont été incubés dans un tampon de dilution (tampon D 36% ; ARNt levure 0,6 mg/ml ; RNasin 3 U/µl ; DTT 5 mM). Pour obtenir une population d'ARN homogène au niveau structural, les ARN ont été dénaturés par chauffage 40 secondes à 96°C puis renaturés 10 min à température ambiante. Les 5 µl de ce mélange ont été incubés pendant 10 min à 30°C dans 10 µl de mélange protéique contenant l'extrait nucléaire de cellules HeLa (40%), la créatine phosphate (62,5 mM), l'ATP (2,5 mM), la RNasin (3 U/µl), DTT (3 mM) et $MgCl_2$ (6 mM). Les 200 µg d'héparine ont été ajoutés ou non avant dépôt des échantillons sur natif de polyacrylamide 5% ; glycérol 5% ; TBE 0,5X. Les gels ont été révélés à l'aide du système Phosphorimager.

2.6. Transfert de marquage

L'ARN radiomarqué complexé aux protéines, selon le protocole décrit précédemment, a été placé à 10 cm d'une source UV (256 nm à 0,8 $J.cm^{-2}$) pendant 10 min à 4°C. Les échantillons ont ensuite été traités à la RNAse T1 (100 U) pendant 1 heure à 37°C, dénaturés à 90°C dans du tampon de Laemmli (Tris 60 mM pH 6,8 ; SDS 2% ; Glycérol 10% ; DTT 100

mM ; bleu de bromophénol 0,03%), puis analysés par électrophorèse sur gel dénaturant SDS-PAGE. Les protéines radioactives ont été révélées à l'aide du système Phosphorimager. La protéine PTB a été immunoprécipitée à l'aide de billes Dynabeads-protéine A couplées à l'anticorps anti-PTB. Dans un premier temprs, l'anticorps a été couplé aux billes Dynabeads selon le protocole suivant : Les billes lavées dans un tampon Phosphate 0,1 M pH 8,1 ont été incubées en présence d'anticorps dans le même tampon contenant 0,1% de BSA (100 µg d'immunoglobulines G (IgG) pour 100 µl de billes). Les anticorps ont été fixés aux billes en présence de 0,2 M triéthanolamine pH 8,2 et de 20 mM DNP pendant 30 min à température ambiante. Les IgG non fixés covalemment ont été éliminés par des lavages successifs avec une solution de glycine 0,1 M à pH 3. Les billes couplées ont été resuspendues dans 100 µl de tampon Phosphate 0,1 M pH 8,1. Dans un second temps, les échantillons issus d'expériences de transfert de marquage ont été placés en présence de 20 µl de solution contenant les billes Dynabeads-protéine A couplé à l'anticorps anti-PTB, pendant 2 heures à 4°C. Les billes ont ensuite été lavées trois fois avec 1 ml de tampon L (Tris 50 mM pH 7,5 ; NP40 0,25%; NaCl 0,15 M) et les protéines retenues par les billes ont été éluées dans le tampon de Laemmli, dénaturées 5 min à 90°C puis analysés par SDS-PAGE. Les protéines radiomarquées et immunoprécipitées ont été révélées à l'aide du système Phosphorimager.

2.7. Purification de complexes à l'aide d'ARN oxydés par le métapériodate de sodium sur les billes adipic acid dihydrazide agarose

Le protocole que nous avons utilisé pour effectuer nos purifications a été adapté des expériences de Caputi et collaborateurs (Caputi *et al.*, 1999).

Dans un premier temps, la matrice a été préparée. 400 µl de solution de billes adipic acid dihydrazide-agarose à 50% (SIGMA A0802), ont été lavés avec 10 ml de tampon Acétate de sodium 0,1 M pH 5 puis centrifugé 3 min à 300 rpm (1000g) à 4°C. Cette étape de lavage a été répétée trois fois puis les billes ont été reprises dans 300 µl de tampon NaAc 0,1 M pH 5.

Les ARN ont été oxydés pour être fixés sur la matrice activée : 100 pmoles d'ARN ont été placées dans 80 µl de tampon acétate de sodium 0,1 M pH 5 contenant 5 mM métapériodate de Sodium (SIGMA S1878) puis incubées 1 heure à température ambiante à l'abri de la lumière. L'ARN a été précipité avec 3 volumes d'éthanol 100%. Le culot d'ARN a été récolté après une centrifugation de 30 min à 13000 rpm à 4°C puis lavé deux fois à l'éthanol 70%. Après élimination de l'alcool et séchage, les culots d'ARN ont été resuspendus dans 100 µl de tampon acétate de sodium 0,1 M pH 5 et incubés avec 140 µl du mélange

billes-tampon acétate de sodium 0,1 M pH 5 pendant une nuit à 4°C sous agitation. Les billes ont été récupérées par centrifugation 2 min à 3000 rpm à 4°C, lavées trois fois avec 1ml de NaCl 2 M, puis trois fois avec le tampon D.

Avant la formation des complexes RNP, les billes sur lesquelles sont fixées les ARN sont saturées à l'aide de 1,4 ml de tampon de blocage (ARNt de levure 0,1 mg/ml ; BSA 1 mg/ml ; +/- Héparine 0,25 mg/ml selon la méthode de purification). Ce mélange a été incubé 1h30 à 4°C, puis 30 min à 30°C sous agitation. Les billes sont ensuite récupérées par centrifugation 3 min à 3000 rpm à température ambiante après élimination du surnageant.

Parallèlement l'extrait nucléaire préalablement incubé 10 min à 30°C, a été placé à 40% final dans un tampon de dilution des protéines (ARNt de levure 1mg/ml ; $MgCl_2$ 3 mM ; tampon D 35% ; DTT 2 mM ; BSA 0,2 mg/ml ; RNAsin). Ce mélange a été incubé 20 min à 4°C. Selon la méthode de purification, de l'héparine à 0,25 mg/ml final, a été ajoutée et le mélange de nouveau incubé 5 min à 30°C.

Les complexes ont été formés en présence de 250 µl de ce mélange et 100 pmol d'ARN immobilisés sur les billes. Après 20 min d'incubation à 30°C, les billes sont lavées quatre fois avec 1 ml de tampon D additionné de $MgCl_2$ (4 mM) et NP40 (0,1%) puis trois fois avec 1ml de tampon D. Entre chaque lavage la matrice a été récupérée par centrifugation 2 min à 3000 rpm à 4°C.

Les protéines ont été éluées par ajout de 80 µl de tampon D contenant de la RNase A (0,5 µg/µl) et de la RNase T1 (1U/µl). Après incubation 30 min à 30°C les protéines éluées sont récupérées par centrifugation et concentrées sur colonnes Nanosep 10 K (Pall).

2.8. Coloration au nitrate d'argent

Les gels SDS-PAGE 10% polyacrylamide ont été fixés dans une solution éthanol 30% ; acide acétique 10% pendant 2 heures puis incubés 30 min dans un tampon thiosulfate (éthanol 30% ; Acétate de Sodium 10 mM ; pH 6 avec acide acétique glacial ; thiosulfate de Na à 0,1% final). Le gel a ensuite été lavé trois fois avec de l'eau pendant 10 min. La révélation a été effectuée à l'aide d'une solution de Nitrate d'Argent ($AgNO_3$ 0,1% dans de l'eau, et 0,1 volume de formaldéhyde 37% pour 400 volumes de solution). 1 rinçage rapide a été effectué à l'eau. Le gel a été développé par ajout d'une solution (Na_2CO_3 2,5% dans H_2O ; et 0,1 volume de formaldéhyde 37% pour 200 volumes de solution) et incubé jusqu'à l'apparition des premières bandes. La coloration a été arrêtée par ajout d'acide acétique 1% final puis lavage à l'eau.

ANNEXE 2 - Description d'une nouvelle méthode de détection de transcrits alternatifs par pyroséquençage : PASP (Pyrosequencing analysis of splicing pattern)

Le séquençage du génome humain a permis de montrer qu'un faible nombre de gènes est capable de générer un protéome complexe. La majeure partie des évènements moléculaires qui augmente la diversité protéique intervient post-transcriptionnellement au cours de la maturation des ARNpré-m en ARNm. L'épissage alternatif qui permet la synthèse de plusieurs ARNm à partir d'un seul ARNpré-m concerne plus de 70% des gènes humains et constitue un mécanisme important de régulation de l'expression génétique chez l'homme. Il n'est pas étonnant qu'un nombre croissant de maladies soit montré associé ou causé par des perturbations de la régulation de l'épissage. Chez l'homme, plus de la moitié des mutations induisant des maladies conduisent à des défauts de la réaction d'épissage (Wang et Cooper, 2007). Cette constatation souligne la nécessité de disposer de méthodes rapides et simples pour analyser les profils d'épissage d'un gène donné dans un grand nombre d'échantillon dans un but de diagnostic et de pronostic. Les méthodes les plus répandues sont les RT-PCR utilisant des amorces qui entourent l'exon épissé de manière alternative et permettent la détection des différents évènements d'épissage. Cette technique reste cependant assez lourde à mettre en place dès que le nombre d'échantillon à analyser devient important (coût élevé des réactifs, temps d'expérience long dont certaines étapes sont difficilement automatisables). Les analyses à haut débit font appel aux puces dédiées aux évènements d'épissage Les oligonucléotides fixés sur ces puces correspondent à des jonctions exon-exon. Un signal positif est donc obtenu si l'échantillon contient l'ARNm ayant subit un évènement d'épissage. D'autres puces ont ensuite été développées et permettent l'analyse simultanée de plus plusieurs évènements d'épissage qui ont lieu dans une population d'ARNm donnée. Cependant, ces technologies sont coûteuses et seul un nombre limité de population d'ARN peut être ainsi analysé. Une technologie qui permet de comparer des profils d'épissage entre des ARNm d'origine très diverse fait toujours défaut. Cette technologie permettrait de comparer plusieurs organes provenant de fond génétiques différents, faire une cinétique précise d'apparition d'un transcrit au cours du temps, ou bien rechercher de manière rapide la présence de transcrits anormalement épissés dans une population d'ARN provenant de patients ou de différents modèles animaux.

J'ai participé à la mise au point d'une méthode d'analyse de profils d'épissage que nous avons appelé PASP pour « *Pyrosequencing Analysis of Splicing Patterns* » qui combine une RT-PCR avec le pyroséquençage des produits PCR. La technique de pyroséquençage est basée sur la détection des pyrophosphates (PPi) relargués lors de la synthèse d'ADN à partir d'une amorce. Un mélange enzymatique va convertir le PPi en ATP qui, à son tour, va convertir la luciférine en oxyluciférine. Cette conversion s'accompagne de l'émission de photons qui sont alors détectés par l'appareil et traduits par un pic dont la hauteur est proportionnelle à la quantité de PPi initiale donc de nucléotides incorporés lors de la réaction de séquençage (Ronaghi *et al.*, 1996). Cette technologie a tout d'abord été développée pour détecter des SNP (Single Nucleotide Polymorphisms) et permet d'attribuer une fréquence allélique à un SNP dans une population d'ADN (Gruber *et al.*, 2002). En choisissant une amorce adéquate et un ordre d'arrivée des nucléotides approprié, il devait être possible de détecter le taux relatif de deux transcrits alternatifs qui possèdent une partie commune et qui diffèrent au niveau des exons alternatifs. Nous avons tout d'abord montré que le ratio de deux transcrits alternatifs, que nous avions produit par transcription *in vitro*, présents dans différents mélanges que nous avions composés, peut être mesuré de manière très précise par PASP. Nous avons ensuite montré que dans des extraits biologiques, le ratio de ces deux transcrits mesurés par PASP est corrélé à celui mesuré par la technique classique de RT-PCR radioactive. Enfin, nous avons montré que différents évènements d'épissage (épissage d'exons terminaux, exons cassettes, exons mutuellement exclusifs et rétentions d'introns) peuvent être détectés de manière très précise dans des populations d'ARN provenant de différents tissus et que les ratios entre les différents ARN provenant d'évènements d'épissage déterminés par PASP sont identiques à ceux déterminés par RT-PCR radioactive. Bien que basée sur des expériences de RT-PCR, cette nouvelle technique présente l'avantage d'être composée d'étapes pour la plupart automatisables rendant ainsi l'analyse de centaines d'évènements d'épissage peu coûteuse en temps et en argent. L'article qui suit décrit cette nouvelle technique.

Published online 11 August 2009

Nucleic Acids Research, 2009, Vol. 37, No. 19 ***e126***
doi:10.1093/nar/gkp626

Analysis of splicing patterns by pyrosequencing

Agnès Méreau[1,2,3], Vincent Anquetil[1,2,3], Marie Cibois[1,2,3], Maud Noiret[1,2,3], Aline Primot[1,3,4], Audrey Vallée[1,3,4] and Luc Paillard[1,2,3,*]

[1]Institut de Génétique et Développement de Rennes, Université de Rennes 1, IFR 140, [2]CNRS, UMR6061, équipe Expression Génétique et Développement, [3]Université Européenne de Bretagne and [4]CNRS, UMR6061, équipe Régulation Transcriptionnelle et Oncogenèse, F-35000 Rennes, France

Received June 11, 2009; Revised July 12, 2009; Accepted July 13, 2009

ABSTRACT

Several different mRNAs can be produced from a given pre-mRNA by regulated alternative splicing, or as the result of deregulations that may lead to pathological states. Analysing splicing patterns is therefore of importance to describe and understand developmental programs, cellular responses to internal or external cues, or human diseases. We describe here a method, Pyrosequencing Analysis of Splicing Patterns (PASP), that combines RT–PCR and pyrosequencing of PCR products. We demonstrated that: (i) Ratios of two pure RNAs mixed in various proportions were accurately measured by PASP; (ii) PASP can be adapted to virtually any splicing event, including mutually exclusive exons, complex patterns of exon skipping or inclusion, and alternative 3′ terminal exons; (iii) In extracts from different organs, the proportions of RNA isoforms measured by PASP reflected those measured by other methods. The PASP method is therefore reliable for analysing splicing patterns. All steps are done in 96-wells microplates, without gel electrophoresis, opening the way to high-throughput comparisons of RNA from several sources.

INTRODUCTION

Gene expression in eukaryotes consists of several coupled steps that take place in the nucleus and the cytoplasm. The nucleus is the place where pre-mRNA are transcribed from DNA matrices and matured to mRNA. Maturation consists of 5′ end capping, 3′ end cleavage and polyadenylation and removal of internal introns associated with exons splicing. Furthermore, different mRNAs can be obtained from the same pre-mRNA due to a process referred to as alternative splicing. In extreme cases, such as *Drosophila* Dscam, several thousands of different mRNAs may be generated from a given pre-mRNA (1). The term alternative splicing encompasses mutually exclusive exons (where splicing leads to the inclusion of either of two exons), exon skipping, intron retention, alternative 5′ or 3′ splice sites (leading to retention of all or only part of an exon or an intron), or alternative initial or terminal exons (2).

Alternative splicing is responsible for a considerable variety of mature, cytoplasmic mRNAs, greatly expanding the proteome. It is generally regulated in a cell or tissue-specific manner, often during developmental programmes. In addition to regulated alternative splicing, defects in constitutive splicing may also generate some diversity in mature mRNAs. These splicing defects may lead to deleterious effects. It is estimated that a large part (up to 50%) of the mutations causing human diseases alter mRNA splicing (3).

These observations highlight the need for methods to analyse mRNA-splicing patterns. A number of methods have been published, and all of them have advantages and limitations. Here, we sought to test if a quantification of alternative mRNA isoforms can be achieved by pyrosequencing. In contrast to dideoxynucleotide-based methods for DNA sequencing, pyrosequencing is a quantitative method, in as much as the heights of the peaks generated from pyrosequencing a mixture of several DNAs are strictly proportional to the molar ratios of these DNAs (4). These properties were used to develop quantitative analyses on DNA, such as characterization of methylation patterns (5,6) and determination of Single Nucleotide Polymorphism (SNP) allele frequencies in a population (7,8). Pyrosequencing also allowed measurements of mRNA ratios after reverse-transcription, for instance to demonstrate allele-specific transcript expression of imprinted genes (9,10) or to compare gene expression levels from different sources following sequence-tagged RT-PCR (11).

Pyrosequencing was used in a limited number of cases to analyse splicing patterns. Exon 3 of human stimulatory G protein Gαs can be spliced or skipped. Furthermore, there exists an alternative 3′ splice site in intron 3, leading to shortening of exon 4 by 3 nt. Consequently, two alternative 5′ ends of exon 4 can be spliced to either exon 2 or exon 3, generating four splice isoforms. Frey *et al.*

*To whom correspondence should be addressed. Tel: +33 223 23 44 73; Fax: +33 223 23 44 78; Email: luc.paillard@univ-rennes1.fr

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

quantified these isoforms in human tissues by pyrosequencing (12). Also in several tissues, Szafranski *et al*. measured the frequencies of alternative 3′ splice sites usage by pyrosequencing (13). However, these two articles did not examine other types of splicing events, like mutually exclusive exons, alternative terminal exons or consecutive alternative exons potentially leading to complex patterns. We develop here a RT–PCR-pyrosequencing based method that we termed Pyrosequencing Analysis of Splicing Patterns (PASP). PASP is applicable to analysing virtually any splicing pattern. Furthermore, the majority of the steps can be automated, allowing several hundreds of splicing events to be analysed within a limited time scale and in cost-effective conditions.

MATERIALS AND METHODS

General biological and biochemical methods

Anatomical pieces from euthanized mice or Xenopus were harvested in liquid nitrogen. Total RNA was extracted using Tri-Reagent (Euromedex). Animals were treated following European rules for animal welfare. Xenopus embryos were obtained by *in vitro* fertilization following standard procedures.

Pure RNA containing either the 6A or the 6B exon of the Xenopus α-tropomyosin gene were obtained by *in vitro* transcription (Promega) from the XTMα24 and XTMO54 plasmids (14) and mixed in various proportions. Pure RNA or RNA extracted from animal organs or embryos were reverse-transcribed using random primers (except 3′RACE), 0.5 mM dNTP and Superscript II Reverse Transcriptase (Invitrogen) following standard procedures. For 3′RACE, reverse transcriptions were made with using an anchored oligo(dT) primer (GGCCAGGGTTTTCCC AGTCACGACT$_{16}$V, V = A, C or G) as described (15).

For radioactive analysis of splicing patterns, the α-tropomyosin forward primer (sequence CAGAGGAAC GTGCTGAACTTT) was radiolabelled using T4 polynucleotide kinase (Promega) and γ^{32}P-ATP. cDNAs were amplified using that radiolabelled forward primer and a reverse primer (sequence ATTTGTCCTCCTTCT GGGAGTA) with the following program: 90 s at 94°C, then (1 min at 94°C, 30 s at 58°C, 30 s at 72°C) for 30 cycles. The resulting PCR mixtures were divided into three parts, one of which was directly electrophorized on 12% polyacrylamide gels, and the two other parts were digested with BstNI or AvaII before electrophoresis. Dried gels were submitted to STORM Phosphorimager (Molecular Dynamics) analysis and quantified with the Imagequant software.

Pyrosequencing

PCR and sequencing primers were designed using the PSQ assay design software and purchased from Sigma Proligo or Eurogentec. The PCR conditions were the same as above except for the number of cycles. The sequences of PCR and sequencing primers are given in Table 1. Pyrosequencing was done following the instructions of the manufacturer of the pyrosequencing device (Biotage). Briefly, the resulting amplimeres were bound

Table 1. Primer sequences and formulae

mRNA	PCR primers	Sequencing primer	Sequences of each isoform	Nucleotide dispensation order	Formulae
α-tropo, 6A/6B	b-fwd CAGAGGAACGTGCTGAACTTT	s TCCTTCTGGGAGTACTT	6A: CTCTTCTG	ACTCTCGTGC	(6B)%=100×G7/(G7+G9)
	rev ATTTGTCCTCCTTCTGGGAGTA		6B: CTCTGCC		(6A)%=100×G9/(G7+G9)
α-tropo, terminal exons	Fwd AGGCTGAAACACGTGCTGAG	s1 GTCCATTGATGATTTAGA	9A:AGATGA	CAGAGTGA	(O5)%=100×G5/(G5+G7)
	b-rev CGCCAGGGTTTTCCCAGTCACGAC		O5:AGAGAAA		(9A)%=100×G7/(G5+G7)
		s2 CTCTCAATGACATGACTTC	9′:AATGTAAATTC	GATGATATCG	(9′)%=100×G4/(G4+G10)
			9B:AATATAAAATG		(9B)%=100×G10/(G4+G10)
Tnnt2	fwd GGAGAGCCGAGAGCATGT	s1 GGAGTACGAGGAGGAAC	4 and 9: AGGAAGAAGCTG	GAGAGAGAGCTGAG	(4+9)%=100×G12/(G9+G12+G14)
	b-rev TCAGGACCAACCTCTTCTACG		5: AGGAAGAGGAAGA		(5)%=100×G9/(G9+G12+G14)
			ref: AGGAAGAGCAAG		(ref)%=100×G14/(G9+G12+G14)
		s2 AACAGGGAAGAAGCTGTG	4: GAAGAGCAAGA	TGAGAGCAGAC	(4)%=100×C7/(C7+C11)
			9: GAAGAGGAAGAC		(9)%=100×C11/(C7+C11)

b, biotinylated; fwd, forward primer; rev, reverse primer; s, sequencing primer.

to agarose–streptavidin beads (GE Healthcare) for 15 min under continuous agitation. Beads were loaded on the Biotage Workstation under vacuum for successive treatments with 70% ethanol, alkaline denaturation buffer and neutralization buffer. Finally, beads were resuspended in the pyrosequencing plate in wells containing 0.4 µM of sequencing primer and heat-denatured (2 min, 85°C). The content of the plate was pyrosequenced on a Biotage PSQ 96 MA device and analysed using the Pyromark software. The nucleotide dispensation order and the formulae used to calculate the percentages of the different isoforms are given in Table 1.

RESULTS

Principe of analysis of splicing pattern by pyrosequencing

The flow chart of the PASP method is shown in Figure 1. Total RNA is extracted and reverse-transcribed following standard procedures. The cDNAs are PCR-amplified using primers complementary to conserved regions flanking the alternative-splicing event to be analysed (indicated by alternative-splicing event 1 or 2 on Figure 1). This results in a mixture of DNAs whose relative proportions are related to those of the initial mRNA pool. These proportions can be measured by pyrosequencing as would be done in a SNP analysis by determining, for a population of molecules, the relative proportions of each nucleotide at a given genomic position. Briefly, correct amplification can be checked by native electrophoresis (but this is not always necessary, see below). Next, PCR products are bound to streptavidin beads via the biotin moiety that is present at the 5′ end of one of the PCR primers. Bound PCR products are denatured by alkaline treatment. A sequencing primer is then annealed to the resulting single-stranded DNA fragment bound to the beads. This primer must be designed to hybridize to a region that is conserved between all the PCR products, but should be adjacent (1–3 nt) to the alternative region. Finally, pyrosequencing is done using a nucleotide dispensation order permitting calculations of the relative proportions of the alternative PCR products (see examples below), using software dedicated to SNP analyses. Importantly, most of the PASP steps can be automated, greatly increasing throughput.

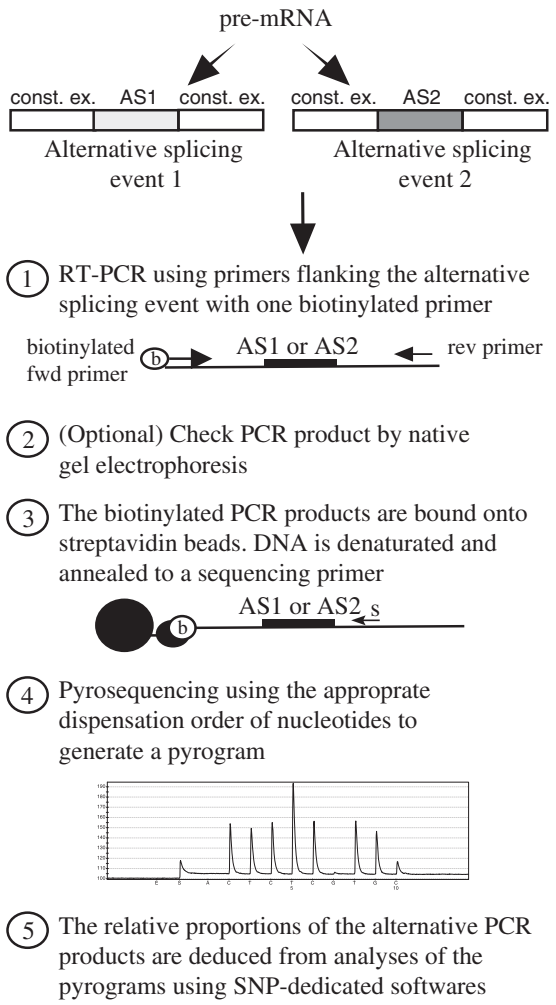


Figure 1. Flow-chart of the PASP method. See text for details.

Mutually exclusive exons

We decided to use the Xenopus α-tropomyosin mRNA to provide a proof-of-principle of the method. Both exons 5 and 7 are constitutive, and two mutually exclusive exons 6 (6A and 6B) contain the same number of nucleotides despite different sequences (Figure 2A) (14). A method to determine the relative proportions of these two exons within a population of mRNA is to run a PCR using primers hybridizing in the exons 5 and 7, one of which is radiolabelled, then to separate the PCR products by electrophoresis and to analyse them by autoradiography. One such analysis is shown in Figure 2A, upper panel. RNAs were obtained by *in vitro* transcription from α-tropomyosin templates containing either the 6A or the 6B exon. These RNAs were mixed in various proportions. After reverse transcription, they were PCR-amplified using a forward ^{32}P-labelled reverse primer. Due to the fact that exons 6A and 6B have the same size, this yielded a unique, 133-bp long, PCR product, irrespective of the proportions of exons 6A and 6B (Figure 2B, upper panel). To discriminate between the two PCR products, restriction with AvaII or BstNI was used. The BstNI site is only present in the 6B exon whereas the AvaII site is only present in the 6A exon. Hence the proportions of BstNI-restricted PCR product reflect those containing exon 6B (middle panel) and the proportions of AvaII-restricted PCR product reflect those containing exon 6A (lower panel). The ratios of exons 6A to 6B in the populations, calculated from the proportions of the differentially restricted cDNAs, were in close agreement with the expected ratios (Figure 2C). Very similar results were obtained when 10 times or 100 times more RNA were used as RT–PCR matrices (data not shown).

For the PASP method, cDNAs were PCR-amplified using a biotinylated forward primer complementary to

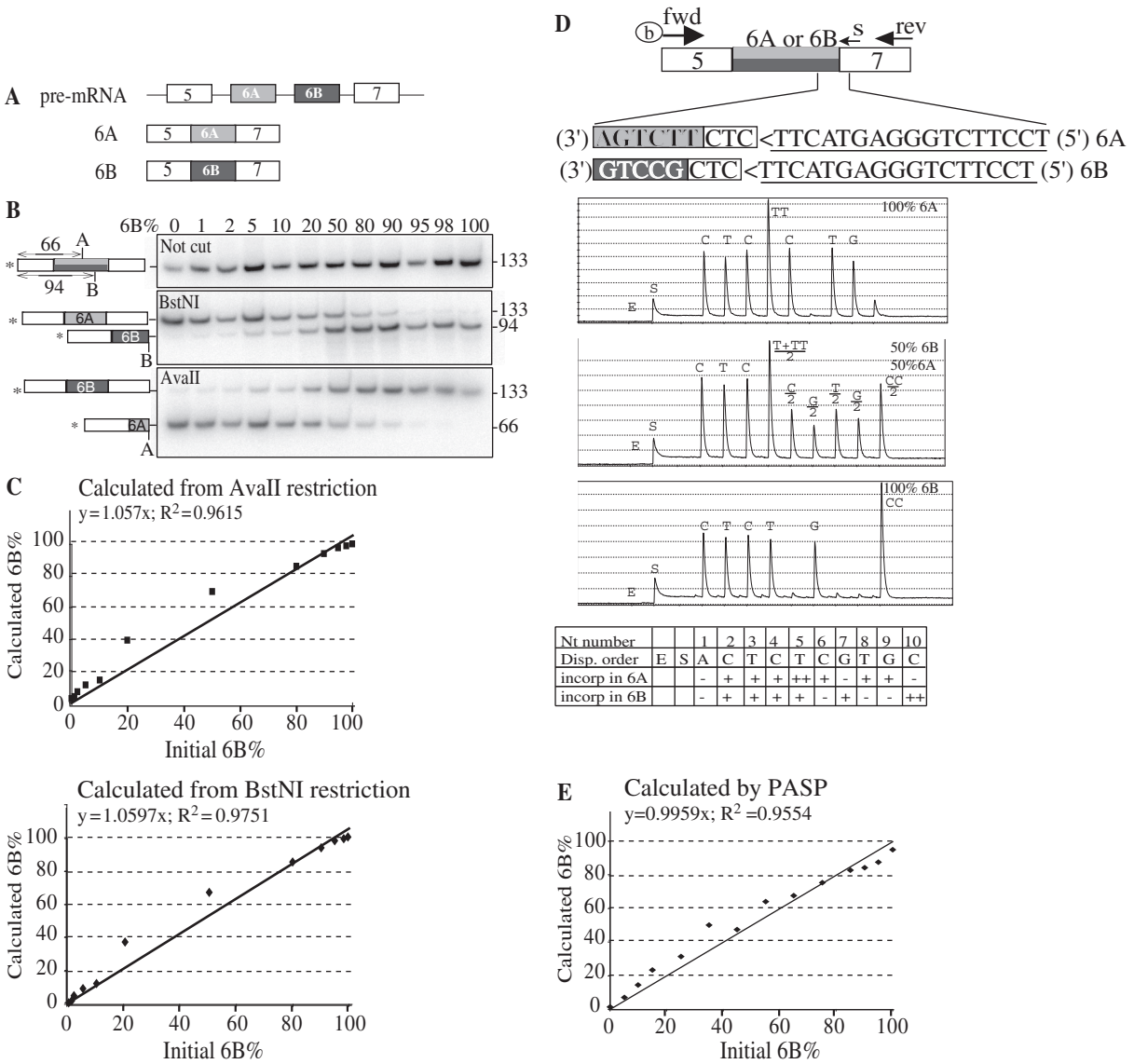


Figure 2. Comparison of radioactive and PASP methods to analyse splicing patterns of two mutually exclusive exons. (**A**) Schematic of Xenopus α-tropomyosin pre-mRNA and the two alternative mature isoforms (6A and 6B). Exons are represented by boxes and introns by lines. (**B**) RNAs containing either the exon 6A or the exon 6B of Xenopus α-tropomyosin were obtained by *in vitro* transcription and mixed in the indicated proportions (6B%). In all mixes, the total amount of RNA was 1 ng. RNAs were reverse-transcribed using random primers and PCR-amplified using a radiolabelled (as indicated by the star) forward primer complementary to exon 5 and a reverse primer complementary to exon 7. One-third of the PCR products was electrophoresed in a native acrylamide gel (upper panel), and the other two-thirds were digested by either BstNI or AvaII before electrophoresis (middle and lower panels, respectively). On the sides of the gels are indicated the positions of the restriction sites BstNI (B, in exon 6B) and AvaII (A, in exon 6A) as well as the sizes of the radiolabelled restriction fragments. (**C**) The percentages of exon 6B calculated from the proportion of DNA left intact after AvaII restriction (upper panel) or cut by BstNI (lower panel) were plotted versus the proportion of exon 6B in the initial RNA mixture. (**D**) Positions of the biotinylated forward (fwd), reverse (rev) and sequencing (s) primers used for PASP of α-tropomyosin. Exons 6A and 6B are in light and dark grey, respectively. Inverse complementary of the sequencing primer is underlined (note that 5′–3′ orientation is from right to left). Boxed CTC is a trinucleotide that is present in both 6A and 6B PCR products and is immediately downstream of the sequencing primer. Three representative pyrograms corresponding to the indicated proportions of 6A and 6B RNA are shown with the nucleotide (Nt) dispensation order. E and S stand for enzyme and substrate, respectively. (**E**) One nanogram of total RNA containing various proportions of 6A and 6B isoforms was reverse transcribed, PCR amplified for 35 cycles, and submitted to PASP analysis. The proportions of exon 6B calculated from the ratios of peak heights in pyrograms [$100 \times G7/(G7 + G9)$] were plotted versus the proportions of exon 6B in the initial RNA mixture.

exon 5 and a reverse primer complementary to exon 7 (Figure 2D). The PCR products were pyrosequenced using an antisense sequencing primer (s) complementary to the 5′ end of exon 7 (just 3′ of the 6A or 6B exon). With this primer, the beginning of the sequence read from a 6A matrix would be CTCTTCTGA, whereas that from a 6B matrix would be CTCTGCCTG. The dispensation order and typical pyrograms, corresponding to different percentages of exon 6B, are shown in Figure 2D. The first dispensed nucleotide (A1) is present in neither of the two sequences, and was used to estimate the background signal. The following three nucleotides (C2, T3 and C4) are present in both isoforms. As expected, they were detected at similar levels irrespective of the

percentage of 6B. Nucleotides 5–10 discriminate between the two isoforms (Figure 2D).

Peak heights of the pyrograms, as provided by the Pyromark software, were used to calculate the percentages of 6B mRNA. The yields of the different nucleotides were not exactly the same. For example, in Figure 2D, the peak for T3 is slightly smaller than the peaks for C2 and C4, although these three peaks correspond to constitutive nucleotides and should be similarly incorporated. Since these differences might introduce a bias if different nucleotides were used to calculate isoforms ratios, we used the same nucleotide that is present in both isoforms at different positions for the calculations. G7 and G9 correspond to the 6B and 6A isoforms respectively (Figure 2D), hence the percentage of 6B equals to $100\times$ G7/(G7 + G9). Figure 2E demonstrates a close agreement between the percentage of 6B calculated by this method and the percentage of 6B in the initial RNA mixture. Furthermore, the PASP results from 30 or 35 PCR cycles were indistinguishable, as well as the results obtained from various initial amounts of mixed mRNAs (data not shown), showing that the PASP method is largely insensitive to variations in these reaction conditions. Comparison of Figures 2C and 2E indicates that the radioactive and PASP methods are equally precise in quantifying the proportions of mutually exclusive exons in a population.

In order to reduce the time of the PASP analysis, we evaluated whether checking the PCR amplification by gel electrophoresis was mandatory. We observed that inefficient PCR amplifications, resulting in too low an amount of product to be sequenced, yielded low-quality pyrograms, and incorrect estimations of the 6B percentage (data not shown). To eliminate the electrophoresis step, we therefore developed quality controls of the pyrograms so as to be able to discard potentially non-reliable results due to bad amplifications. We examined the predictive values for reliable results of two putative quality controls (QC). As a first quality control (QC1), we examined the height of the first constitutive nucleotide (C2), reasoning that a high peak value should reveal a large amount of amplified DNA. Quality control 2 (QC2) was calculated as the ratio between C2 and A1 (respectively, first constitutive nucleotide and nucleotide absent from the two isoforms). This ratio was expected to correspond to a signal/noise ratio; hence it should be as high as possible. To determine the values of QC1 and QC2 relative to the quality of PCR amplification, PCR were performed on mixtures of 6A and 6B RNA either with (Figure 3, lanes 1–3, 5–7 and 9–11) or without (lanes 4, 8 and 12) previous reverse-transcription. As expected, correct PCR products were present only in the presence of reverse-transcriptase (Figure 3). When PCR products were present, the values for QC1 (strength of signal), were above 40, and the values for QC2 (signal/background) were above or close to 100. In contrast, when amplification failed, QC1 values were less than 1, and QC2 values were no more than 14. Several experiments correlated correct amplifications (as tested by gel electrophoresis) with high QC1 and QC2 values (data not shown). We found that values higher than 40 and 80 for QC1 and QC2, respectively, accurately reflected correct amplifications and PASP. Hence, checking PCR by gel electrophoresis can be replaced by adequate QCs of PASP.

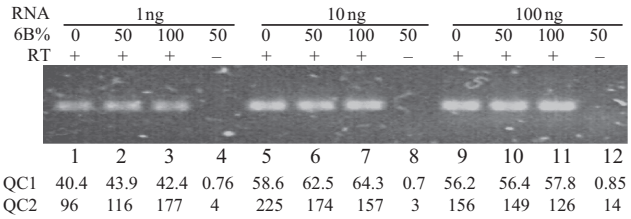


Figure 3. Quality controls of PASP. RNAs containing exons 6A or 6B were mixed in the indicated proportions (6B%) to a total amount of 1, 10 and 100 ng and reverse-transcribed, except for lanes 4, 8 and 12 where the reverse transcriptase was omitted. The resulting cDNAs were PCR-amplified for 25 cycles in a total volume of 50 µl. Five microlitres of the PCR products were electrophorised on a native agarose gel in the presence of Syber-Green and the gel was photographed. The remaining PCR products were submitted to PASP with the same dispensation order as in Figure 2D. QCs 1 and 2 were calculated as described in text and the values are given below each lane.

The above results were obtained from mixes of pure RNAs produced by *in vitro* transcription. It was important therefore to check that the PASP method also applies to complex RNA mixtures extracted from cells or tissues. It was shown previously that exon 6A is essentially retained in mRNA transcribed in oocytes, whereas exon 6B is essentially retained in embryonic somite (presumptive muscles) or adult skeletal and cardiac muscle (14). By PASP, oocytes and early embryos (Stage 8) were found mainly to express a 6A isoform, whereas adult heart, skeletal muscle or stomach, or whole late embryos (Stage 35) and dissected somites were found to express mainly a 6B isoform. An intermediate situation was found for oviducts that contain muscles but are frequently contaminated by oocytes (Figure 4, upper panel). These results were similar to those obtained by radioactive RT–PCR (Figure 4, lower panel) and to previous data (14).

Complex combination of exons skipping and inclusion

Mutually exclusive exons are only part of the alternative-splicing events that can occur on a pre-mRNA. One characteristic is that they may produce alternatively spliced products that have the same size (if the mutually exclusive exons have the same size, as for exons 6A and 6B of Xenopus α-tropomyosin). In contrast, splicing products of different sizes are obtained from other alternative-splicing events [exon skipping, intron retention, alternative 5′ or 3′ sites (2)]. In addition, more than two isoforms can be produced by alternative splicing. We evaluated therefore whether the PASP method could apply to these complex-splicing events. We choosed murine Tnnt2 (troponin T type 2, also named cTNT for cardiac troponin T) as a model to test this issue. Two constitutive exons (3 and 6) flank two alternative exons (4 and 5). In muscular cells, each alternative exon can be either included or skipped, so that four isoforms of different sizes can be obtained from the same pre-mRNA (Figure 5A) (16).

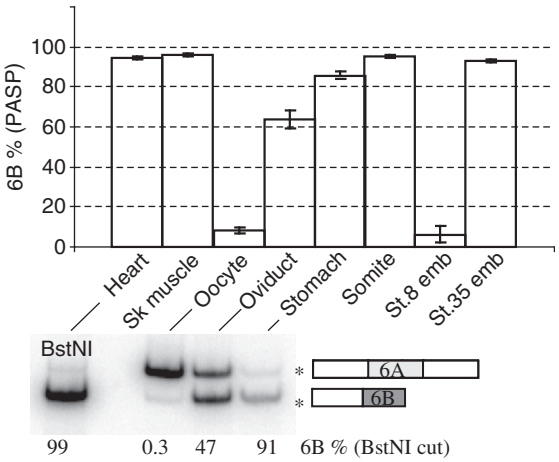
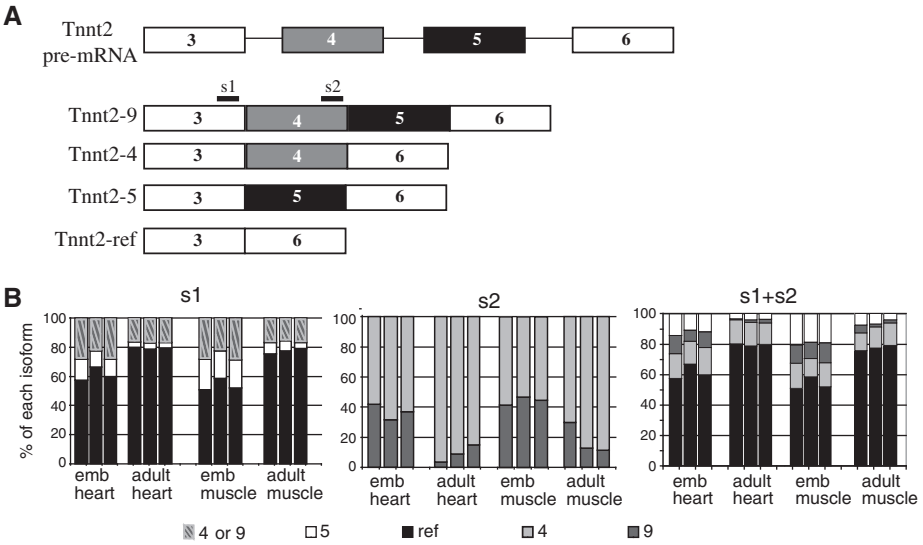


Figure 4. PASP of α-tropomyosin in Xenopus embryos and organs. RNAs were extracted from oocytes, stage 8 or 35 embryos, dissected embryonic somites or the indicated adult organs, and α-tropomyosin-splicing patterns were analysed by PASP. The percentages of exon 6B (mean ± SD of six independent RNA preparations) are shown (upper panel). Adult heart, oviduct stomach, and oocyte RNA were also analysed by radioactive RT–PCR as in Figure 2B (lower panel). The percentages of 6B exon (restricted by BstNI) are indicated under the gel.

For PASP analysis, total RNA extracted from mouse heart or skeletal muscle was reverse-transcribed, PCR-amplified using a biotinylated reverse primer, and pyrosequenced using two different sequencing primers (Figure 5A). Sequencing primer s1 hybridizes to exon 3. It allows sequencing of the exon spliced 3′ to exon 3 and discriminates between isoforms 4 + 9 (exon 4), 5 (exon 5) and ref (reference, exon 6). It does not discriminate between isoforms 4 and 9, which is achieved by sequencing primer s2. Organs from three different mice (adults or foetuses) were analysed. Results of pyrosequencing using primers s1 and s2 are shown in Figure 5B, left and middle panels, respectively. The percentages of each of the four isoforms were calculated from these series of data (right panel). The results obtained from the three different mice were very similar, and average values were calculated (Figure 5C). These results show that, whereas the splicing patterns in skeletal muscle and heart are similar, the percentages of alternative exons inclusion strongly decreased between embryos and adults. Inclusion of exon 5 decreased more than 3-fold (from 22.2 and 31.3% down to 5.3 and 9.1% in heart and muscle, respectively). A similar decrease in exon 5 inclusion has already been



C

	embryonic heart	adult heart	embryonic muscle	adult muscle
ref %	61.2 +/– 4.8	79.5 +/– 0.6	53.8 +/– 4.1	77.5 +/– 1.8
4 %	16.6 +/– 1.3	15.2 +/– 1.0	14.9 +/– 2.5	13.5 +/– 1.5
9 %	9.8 +/– 2.5	1.5 +/– 0.9	11.7 +/– 1.1	3.0 +/– 1.8
5 %	12.3 +/– 1.8	3.7 +/– 0.4	19.6 +/– 0.9	6.1 +/– 1.8
ex4 inclusion %	26.4 +/– 3.8	16.7 +/– 1.9	26.6 +/– 3.6	16.5 +/– 3.3
ex5 inclusion %	22.2 +/– 4.2	5.3 +/– 1.3	31.3 +/– 2.0	9.1 +/– 3.6

Figure 5. PASP analysis of Tnnt2 splicing in mice. (**A**) Schematic drawing of mouse Tnnt2 pre-mRNA and of the alternative isoforms of Tnnt2 mRNA. The positions of the primers for pyrosequencing are shown. Primer s1 hybridizes in the constitutive exon 3 and sequences either exon 4 (isoforms 4 and 9), exon 5 (isoform 5) or constitutive exon 6 (isoform ref, reference). Primer s2 hybridizes in exon 4 and sequences either exon 5 or exon 6, hence it discriminates between isoforms 4 and 9. (**B**) PASP of Tnnt2 in the hearts or the skeletal muscles of three different adult or foetal mice using primer s1 or s2 as indicated (two left panels). The right panel (s1 + s2) combines the data shown in the two left panels to calculate the percentages of the four Tnnt2 isoforms. (**C**) Mean ± SD values were calculated from the above data. Inclusion of exon 4 corresponds to isoforms 4 and 9, and inclusion of exon 5 corresponds to isoforms 5 and 9.

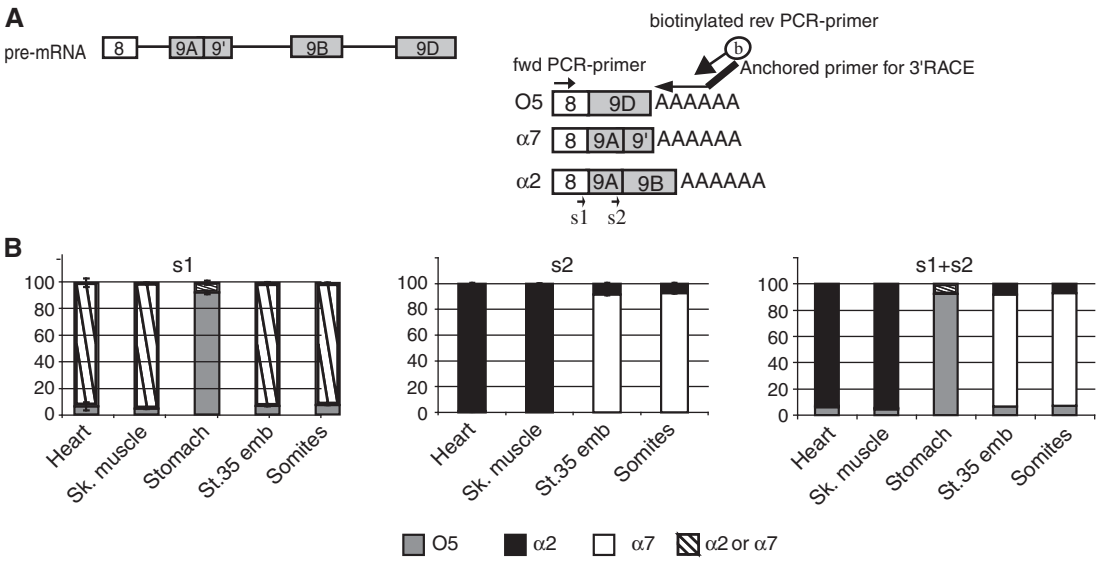


Figure 6. Combined PASP and 3′RACE to analyse splicing of alternative 3′ terminal exons. (**A**) Schematic drawing of the 3′ terminal part of Xenopus α-tropomyosin pre-mRNA. Constitutive exon 8 can be spliced to exon 9D to generate isoform O5. Alternatively, it can be spliced to exon 9A9′. Exon 9A9′ can be used as the terminal exon (yielding isoform α7) or can be spliced to exon 9B via an alternative internal 5′ splice site to create isoform α2. In 3′RACE-PASP, an anchored oligo(dT) primer is used for RT, and a biotinylated reverse oligonucleotide corresponding to the anchor is used for PCR together with a forward primer in constitutive exon 8. Sequencing primer s1 hybridizes within constitutive exon 8, to discriminate between exon 9D, isoform O5, and 9A, isoforms α2 plus α7. Sequencing primer s2 hybridizes within exon 9A, to discriminate between exon 9′, isoform α7, and 9B, isoform α2. (**B**) 3′RACE-PASP of α-tropomyosin in the indicated tissues. Shown are the results obtained using sequencing primers 1 and 2 (mean ± SD of three different animals), and the results calculated from combining these two sequencings.

observed (16). Consequently, PASP analysis successfully reproduced results obtained with other methods.

Alternative 3′ terminal exons

We next used PASP to analyse alternative 3′ terminal exons. In this case, there is no constitutive exon 3′ to the alternative-splicing event that can be targeted by a PCR primer. In 3′RACE, reverse transcription is made with an oligo(dT) primer containing an additional 5′ extension ('anchor'). Next, a primer corresponding to the anchor is used together with a gene-specific primer for PCR (15). To test 3′RACE-PASP on Xenopus α-tropomyosin mRNA, we made RTs using RNAs from several different Xenopus tissues with an oligo(dT)-anchor, then a PCR with a biotinylated anchor and a forward primer specific for α-tropomyosin exon 8 (Figure 6A). α-Tropomyosin mRNA may have three different terminal exons as depicted in Figure 6A (17). We used two sequencing primers to discriminate between the three different isoforms. Primer s1 hybridizes in exon 8 and discriminates O5 versus α2 and α7, s2 hybridizes in exon 9A and discriminates α2 versus α7.

Results obtained with sequencing primer s1 are shown in Figure 6B, left panel. Striated muscles, as well as embryos and embryonic somites, mainly spliced exon 9A9′ to exon 8. In contrast, in stomach, isoform O5 was clearly predominant. Results of sequencing with primer s2 (Figure 6B, middle) showed a difference between adult organs that produced mainly α2, and embryonic organs that produced α7. Due to the small amount of exon 9A (binding site for primer s2) in stomach, the sequencing using primer s2 failed and it was not possible to discriminate α2 from α7 isoforms in that organ. Combining the analyses with s1 and s2 (right panel) showed that the main isoforms were O5 in stomach, α2 in adult striated muscles and α7 in embryos, again consistent with previous results (17).

DISCUSSION

Several methods aimed at analysing splicing patterns have been published. Northern blotting may be considered if the isoforms to be discriminate have sufficiently different sizes to be resolved by electrophoresis. An advantage of northern blotting is that the RNA to be analysed are directly detected, avoiding amplification bias that may arise with PCR. However, this approach is virtually no longer used due to a very limited sensitivity (it requires ten to several tens μg of total RNA), and an incapacity to analyse splicing events that lead to isoforms of similar sizes. RNase protection assays (RPA) share with northern blotting the lack of requirement for amplification. The sensitivity of RPA is better than that of northern, although still low (it requires 1–10 μg of total RNA), and RPA is able to analyse more subtle splicing patterns [for instance (18)]. However, the method is time-consuming. Among non-automatable steps, it requires synthesis of labelled RNA probes, hybridization and RNase treatment, gel electrophoresis and drying, and quantification of radioactive signals by densitometry or from phosphorimager files. Furthermore, due to the extreme reactivity of RNase, this assay cannot be used to screen for splicing patterns in different samples with genetic variations including single nucleotide

polymorphism. Invader RNA assays rely on the detection of a cleaved DNA probe that occurs when hybridized to a complementary RNA. Consequently, using a set of DNA probe corresponding to the sequences of the different isoforms, it is possible to quantify these isoforms (19). However, this is not amenable to analyses of numerous splicing patterns.

The most widely used methods are real-time and end-point RT–PCR. Real-time RT–PCR is nowadays the best method to quantify RNA levels, due to its unequalled sensitivity and the facility to simultaneously analyse several targets. However, in its simplest (and cheapest) design, that is with a fluorescent DNA intercalating dye, only one amplification product can be quantified with a given pair of primers. Consequently, to analyse alternative-splicing patterns, it is necessary to design several primer pairs, each of them being specific for one mRNA isoform (20,21). To accurately measure the isoform ratios, it is therefore necessary to use primer pairs with the same amplification efficiencies, which is often difficult to achieve. In contrast, in end-point PCR, it is generally possible to use only one pair of PCR primers to amplify all the isoforms, and to individually quantify the isoforms. For example, if one of the PCR primers is radiolabelled, the PCR products are easy to detect and quantify after electrophoresis by autoradiography or preferably phosphorimager analysis. On the basis of differential electrophoretic mobilities, it may be possible to directly discriminate between the PCR products derived from the differentially spliced mRNAs. This method can be used for the analysis of splicing of both endogenous mRNAs and recombinant RNAs transcribed from a minigene [e.g. see (18)].

PASP shares with RT–PCR the advantage of being very sensitive. Furthermore, since only one pair of primers is used to amplify all the isoforms, its efficiency does not need to be carefully controlled. Finally, several kinds of splicing event can be detected by radioactive RT–PCR or PASP: the presence of either of two sequences (mutually exclusive exons), the presence or the absence of one sequence (exon skipping, intron retention, alternative 5′ or 3′ sites) or different 3′ mRNA extremities (alternative terminal exons). Pyrosequencing was previously used in a limited number of cases to analyse splicing patterns (12,13), but neither of these articles examined the capacity of RT–PCR-pyrosequencing to measure the ratios of isoforms that originate from this variety of splicing events. The possibility to use two sequencing primers to analyse complex-splicing patterns was also not investigated in these articles. However, PASP has several additional advantages over radioactive RT–PCR. First, no radioactive isotope is required. Second, it is much more rapid, since all steps are performed in 96-wells plates. Even gel electrophoresis is dispensable if the correct quality controls are used. Starting from cDNAs, the time required to generate a raw data file corresponding to one plate is about 4 h. Since three different devices are used (thermocycler, workstation for binding of amplimeres to streptavidin and treatments, pyrosequencer), it is possible to treat several plates within the same working day, allowing several hundreds of splicing events to be analysed, or to fractionate the experiments into different days. A considerable increase in throughput is possible by automating all of these steps.

A challenger to PASP is non-radioactive RT–PCR followed by detection of the amplimeres by capillary electrophoresis. This technology was used at a medium-throughput level to identify splicing signatures of ovarian cancer (22). However, as for radioactive RT–PCR, analysing splicing patterns by RT–PCR capillary electrophoresis requires that all the isoforms have different sizes, a requirement that does not hold for PASP. Furthermore, the isoforms are identified only by their electrophoretic mobilities in RT–PCR capillary electrophoresis, whereas they are identified by sequencing in PASP. This adds an additional level of specificity that is valuable if some aspecific PCR amplification occurred.

A high-throughput method to analyse splicing patterns is with splicing arrays. Initially, the oligonucleotides spotted on the chips were designed so as to hybridize to exon–exon junctions. Consequently, a positive signal was detected for a given spot only if the tested sample included mRNAs that were subject to the corresponding splicing event (23). The determination of probes was based on the characterization of gene exons from known mRNA. Hence, exon boundaries must be known initially and if a mRNA sequence has not been determined, exons will be missed. Furthermore, the differences in melting temperature between probes and the cross-hybridization that are likely to occur at junction probes have to be included in the analysis. Other array designs include the use of probes to target the internal part of constitutive and alternative exons (24–27). A very high-density human 'exon array' has been developed by Affymetrix and most global analyses rely on the estimation of the relative abundance of each exon (28). These chip-based technologies provide valuable information on the frequency of alternative-splicing event and exon usage in a quantitative manner and may also lead to the discovery of new splicing events. These approaches allow the simultaneous analysis of thousands of splicing events that may occur within a given mRNA population and that must subsequently be validated. An alternative to microarrays for genome-wide analyses of splicing patterns is deep-sequencing of mRNAs within a population (2,29).

However, due to the cost of either of these two methods, only a limited number of different RNA populations can be tested. In this respect, PASP is complementary to exon-junctions arrays or deep sequencing. These latter methods allow hundreds of thousands of splicing events to be simultaneously analysed in a RNA population, whereas PASP may compare the splicing patterns of several mRNAs from a large number of different origins. For example, with PASP one could compare several organs with different genetic backgrounds, make precise time-course experiments, or rapidly screen for potentially abnormally spliced RNAs in human patients or model animals. PASP provides detailed knowledge about the abundance of individual alternative-splicing transcripts and the possibility to combine this technology with micro-array analysis will be of great value in establishing variations in alternative-splicing among different individuals

that could have medical implications such as resistance to drug or disease susceptibility.

ACKNOWLEDGEMENTS

The authors thank H Beverley Osborne, Serge Hardy, Marie-Dominique Galibert-Anne and members of the EGD and RTO teams for discussions.

FUNDING

Association pour la Recherche contre le Cancer (ARC4003 to L.P., doctoral fellowship to V.A.); Agence Nationale de la Recherche (ANR-07-JCJC-0097-01 to L.P.); Région Bretagne (ARED, doctoral fellowships to A.V.). Undergraduate Research Fellowship to A.D. Funding for open access charge: ANR-07-JCJC-0097-01.

Conflict of interest statement. None declared.

REFERENCES

1. Schmucker,D., Clemens,J.C., Shu,H., Worby,C.A., Xiao,J., Muda,M., Dixon,J.E. and Zipursky,S.L. (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, **101**, 671–684.
2. Wang,Z. and Burge,C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802–813.
3. Wang,G.S. and Cooper,T.A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.*, **8**, 749–761.
4. Ahmadian,A., Ehn,M. and Hober,S. (2006) Pyrosequencing: history, biochemistry and future. *Clin. Chim. Acta*, **363**, 83–94.
5. Tost,J. and Gut,I.G. (2007) Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. *Methods Mol. Biol.*, **373**, 89–102.
6. Shaw,R.J., Liloglou,T., Rogers,S.N., Brown,J.S., Vaughan,E.D., Lowe,D., Field,J.K. and Risk,J.M. (2006) Promoter methylation of P16, RARbeta, E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing. *Br. J. Cancer*, **94**, 561–568.
7. Bhathena,A., Mueller,T., Grimm,D.R., Idler,K., Tsurutani,A., Spear,B.B. and Katz,D.A. (2007) Frequency of the frame-shifting CYP2D7 138delT polymorphism in a large, ethnically diverse sample population. *Drug Metab. Dispos.*, **35**, 1251–1253.
8. Alderborn,A., Kristofferson,A. and Hammerling,U. (2000) Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res.*, **10**, 1249–1258.
9. Sun,A., Ge,J., Siffert,W. and Frey,U.H. (2005) Quantification of allele-specific G-protein beta3 subunit mRNA transcripts in different human cells and tissues by Pyrosequencing. *Eur. J. Hum. Genet.*, **13**, 361–369.
10. Bentley,L., Nakabayashi,K., Monk,D., Beechey,C., Peters,J., Birjandi,Z., Khayat,F.E., Patel,M., Preece,M.A., Stanier,P. *et al.* (2003) The imprinted region on human chromosome 7q32 extends to the carboxypeptidase A gene cluster: an imprinted candidate for Silver-Russell syndrome. *J. Med. Genet.*, **40**, 249–256.
11. Zhang,X., Wu,H., Chen,Z., Zhou,G., Kajiyama,T. and Kambara,H. (2009) Dye-free gene expression detection by sequence-tagged reverse-transcription polymerase chain reaction coupled with pyrosequencing. *Anal. Chem.*, **81**, 273–281.
12. Frey,U.H., Nuckel,H., Dobrev,D., Manthey,I., Sandalcioglu,I.E., Eisenhardt,A., Worm,K., Hauner,H. and Siffert,W. (2005) Quantification of G protein Gaalphas subunit splice variants in different human tissues and cells using pyrosequencing. *Gene Expr.*, **12**, 69–81.
13. Szafranski,K., Schindler,S., Taudien,S., Hiller,M., Huse,K., Jahn,N., Schreiber,S., Backofen,R. and Platzer,M. (2007) Violating the splicing rules: TG dinucleotides function as alternative 3′ splice sites in U2-dependent introns. *Genome Biol.*, **8**, R154.
14. Duriez,P., Lesimple,M., Allo,M.R. and Hardy,S. (2000) Alternative splicing of Xenopus alphafast-tropomyosin pre-mRNA during development: identification of determining sequences. *DNA Cell Biol.*, **19**, 365–376.
15. Hamon,S., Le Sommer,C., Mereau,A., Allo,M.R. and Hardy,S. (2004) Polypyrimidine tract-binding protein is involved in vivo in repression of a composite internal/3′-terminal exon of the Xenopus alpha-tropomyosin Pre-mRNA. *J. Biol. Chem.*, **279**, 22166–22175.
16. Ho,T.H., Bundman,D., Armstrong,D.L. and Cooper,T.A. (2005) Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum. Mol. Genet.*, **14**, 1539–1547.
17. Hardy,S., Fiszman,M.Y., Osborne,H.B. and Thiebaud,P. (1991) Characterization of muscle and non muscle Xenopus laevis tropomyosin mRNAs transcribed from the same gene. *Eur. J. Biochem.*, **202**, 431–440.
18. Le Sommer,C., Lesimple,M., Mereau,A., Menoret,S., Allo,M.R. and Hardy,S. (2005) PTB regulates the processing of a 3′-terminal exon by repressing both splicing and polyadenylation. *Mol. Cell Biol.*, **25**, 9595–9607.
19. Wagner,E.J., Curtis,M.L., Robson,N.D., Baraniak,A.P., Eis,P.S. and Garcia-Blanco,M.A. (2003) Quantification of alternatively spliced FGFR2 RNAs using the RNA invasive cleavage assay. *RNA*, **9**, 1552–1561.
20. Bustin,S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.*, **25**, 169–193.
21. Kafert,S., Krauter,J., Ganser,A. and Eder,M. (1999) Differential quantitation of alternatively spliced messenger RNAs using isoform-specific real-time RT-PCR. *Anal. Biochem.*, **269**, 210–213.
22. Klinck,R., Bramard,A., Inkel,L., Dufresne-Martin,G., Gervais-Bird,J., Madden,R., Paquet,E.R., Koh,C., Venables,J.P., Prinos,P. *et al.* (2008) Multiple alternative splicing markers for ovarian cancer. *Cancer Res.*, **68**, 657–663.
23. Johnson,J.M., Castle,J., Garrett-Engele,P., Kan,Z., Loerch,P.M., Armour,C.D., Santos,R., Schadt,E.E., Stoughton,R. and Shoemaker,D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.
24. Pan,Q., Shai,O., Misquitta,C., Zhang,W., Saltzman,A.L., Mohammad,N., Babak,T., Siu,H., Hughes,T.R., Morris,Q.D. *et al.* (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell*, **16**, 929–941.
25. Sugnet,C.W., Srinivasan,K., Clark,T.A., O'Brien,G., Cline,M.S., Wang,H., Williams,A., Kulp,D., Blume,J.E., Haussler,D. *et al.* (2006) Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.*, **2**, e4.
26. Blanchette,M., Green,R.E., Brenner,S.E. and Rio,D.C. (2005) Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. *Genes Dev.*, **19**, 1306–1314.
27. Clark,T.A., Sugnet,C.W. and Ares,M. Jr. (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science*, **296**, 907–910.
28. Yeo,G.W., Xu,X., Liang,T.Y., Muotri,A.R., Carson,C.T., Coufal,N.G. and Gage,F.H. (2007) Alternative splicing events identified in human embryonic stem cells and neural progenitors. *PLoS Comput. Biol.*, **3**, 1951–1967.
29. Ben-Dov,C., Hartmann,B., Lundgren,J. and Valcarcel,J. (2008) Genome-wide analysis of alternative pre-mRNA splicing. *J. Biol. Chem.*, **283**, 1229–1233.

BIBLIOGRAPHIE

Abelson, J. (2008). "Is the spliceosome a ribonucleoprotein enzyme?" Nat Struct Mol Biol **15**(12): 1235-7.

Afonina, E., R. Stauber et G. N. Pavlakis (1998). "The human poly(A)-binding protein 1 shuttles between the nucleus and the cytoplasm." J Biol Chem **273**(21): 13015-21.

Ahn, S. H., M. Kim et S. Buratowski (2004). "Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing." Mol Cell **13**(1): 67-76.

Amir-Ahmady, B., P. L. Boutz, V. Markovtsov, M. L. Phillips et D. L. Black (2005). "Exon repression by polypyrimidine tract binding protein." Rna **11**(5): 699-716.

Anquetil, V. (2005). Contribution à l'étude du mécanisme moléculaire conduisant à la répression de l'épissage d'un exon 3' terminal du gène de la tropomyosine alpha de xénope. Rennes, Université de Rennes 1: 35.

Ansari, A. et M. Hampsey (2005). "A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping." Genes Dev **19**(24): 2969-78.

Arhin, G. K., M. Boots, P. S. Bagga, C. Milcarek et J. Wilusz (2002). "Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals." Nucleic Acids Res **30**(8): 1842-50.

Ashe, M. P., A. Furger et N. J. Proudfoot (2000). "Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation." Rna **6**(2): 170-7.

Ashe, M. P., L. H. Pearson et N. J. Proudfoot (1997). "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): 5752-63.

Ashfield, R., P. Enriquez-Harris et N. J. Proudfoot (1991). "Transcriptional termination between the closely linked human complement genes C2 and factor B: common termination factor for C2 and c-myc?" Embo J **10**(13): 4197-207.

Ashiya, M. et P. J. Grabowski (1997). "A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart." Rna **3**(9): 996-1015.

Auboeuf, D., D. H. Dowhan, M. Dutertre, N. Martin, S. M. Berget et B. W. O'Malley (2005). "A subset of nuclear receptor coregulators act as coupling proteins during synthesis and maturation of RNA transcripts." Mol Cell Biol **25**(13): 5307-16.

Auboeuf, D., D. H. Dowhan, Y. K. Kang, K. Larkin, J. W. Lee, S. M. Berget, *et al.* (2004). "Differential recruitment of nuclear receptor coactivators may determine alternative RNA splice site choice in target genes." Proc Natl Acad Sci U S A **101**(8): 2270-4.

Auboeuf, D., D. H. Dowhan, X. Li, K. Larkin, L. Ko, S. M. Berget, *et al.* (2004). "CoAA, a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing." Mol Cell Biol **24**(1): 442-53.

Auboeuf, D., A. Honig, S. M. Berget et B. W. O'Malley (2002). "Coordinate regulation of transcription and splicing by steroid receptor coregulators." Science **298**(5592): 416-9.

Audibert, A., D. Weil et F. Dautry (2002). "In vivo kinetics of mRNA splicing and transport in mammalian cells." Mol Cell Biol **22**(19): 6706-18.

Auweter, S. D. et F. H. Allain (2008). "Structure-function relationships of the polypyrimidine tract binding protein." Cell Mol Life Sci **65**(4): 516-27.

Auweter, S. D., R. Fasan, L. Reymond, J. G. Underwood, D. L. Black, S. Pitsch, *et al.* (2006). "Molecular basis of RNA recognition by the human alternative splicing factor Fox-1." Embo J **25**(1): 163-73.

Awasthi, S. et J. C. Alwine (2003). "Association of polyadenylation cleavage factor I with U1 snRNP." Rna **9**(11): 1400-9.

Azubel, M., N. Habib, R. Sperling et J. Sperling (2006). "Native spliceosomes assemble with pre-mRNA to form supraspliceosomes." J Mol Biol **356**(4): 955-66.

Bachler, M., R. Schroeder et U. von Ahsen (1999). "StreptoTag: a novel method for the isolation of RNA-binding proteins." Rna **5**(11): 1509-16.

Baek, D. et P. Green (2005). "Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing." Proc Natl Acad Sci U S A **102**(36): 12813-8.

Balvay, L., M. Lopez Lastra, B. Sargueil, J. L. Darlix et T. Ohlmann (2007). "Translational control of retroviruses." Nat Rev Microbiol **5**(2): 128-40.

Bakkour, N., Y. L. Lin, S. Maire, L. Ayadi, F. Mahuteau-Betzer, C. H. Nguyen, *et al.* (2007). "Small-molecule inhibition of HIV pre-mRNA splicing as a novel antiretroviral therapy to overcome drug resistance." PLoS Pathog **3**(10): 1530-9.

Barabino, S. M., B. S. Sproat, U. Ryder, B. J. Blencowe et A. I. Lamond (1989). "Mapping U2 snRNP--pre-mRNA interactions using biotinylated oligonucleotides made of 2'-OMe RNA." Embo J **8**(13): 4171-8.

Bardwell, V. J. et M. Wickens (1990). "Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method." Nucleic Acids Res **18**(22): 6587-94.

Barilla, D., B. A. Lee et N. J. Proudfoot (2001). "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): 445-50.

Barreau, C., L. Paillard, A. Mereau et H. B. Osborne (2006). "Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions." Biochimie **88**(5): 515-25.

Batsche, E., M. Yaniv et C. Muchardt (2006). "The human SWI/SNF subunit Brm is a regulator of alternative splicing." Nat Struct Mol Biol **13**(1): 22-9.

Bauren, G., S. Belikov et L. Wieslander (1998). "Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron." Genes Dev **12**(17): 2759-69.

Beaudoing, E., S. Freier, J. R. Wyatt, J. M. Claverie et D. Gautheret (2000). "Patterns of variant polyadenylation signal usage in human genes." Genome Res **10**(7): 1001-10.

Beckmann, K., M. Grskovic, F. Gebauer et M. W. Hentze (2005). "A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila." Cell **122**(4): 529-40.

Behzadnia, N., M. M. Golas, K. Hartmuth, B. Sander, B. Kastner, J. Deckert, *et al.* (2007). "Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes." Embo J **26**(6): 1737-48.

Belfort, M., M. E. Reaban, T. Coetzee et J. Z. Dalgaard (1995). "Prokaryotic introns and inteins: a panoply of form and function." J Bacteriol **177**(14): 3897-903.

Benchaouir, R., M. Meregalli, A. Farini, G. D'Antona, M. Belicchi, A. Goyenvalle, *et al.* (2007). "Restoration of human dystrophin following transplantation of exon-skipping-engineered DMD patient stem cells into dystrophic mice." Cell Stem Cell **1**(6): 646-57.

Ben-Dov, C., B. Hartmann, J. Lundgren et J. Valcarcel (2008). "Genome-wide analysis of alternative pre-mRNA splicing." J Biol Chem **283**(3): 1229-33.

Bennett, M., S. Michaud, J. Kingston et R. Reed (1992). "Protein components specifically associated with prespliceosome and spliceosome complexes." Genes Dev **6**(10): 1986-2000.

Bentley, D. L. (2005). "Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors." Curr Opin Cell Biol **17**(3): 251-6.
Berget, S. M. (1995). "Exon recognition in vertebrate splicing." J Biol Chem **270**(6): 2411-4.
Bergkessel, M., G. M. Wilmes et C. Guthrie (2009). "SnapShot: Formation of mRNPs." Cell **136**(4): 794, 794 e1.
Berglund, J. A., M. L. Fleming et M. Rosbash (1998). "The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence." Rna **4**(8): 998-1006.
Bernardi, A. et P. F. Spahr (1972). "Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17." Proc Natl Acad Sci U S A **69**(10): 3033-7.
Beyer, A. L. et Y. N. Osheim (1988). "Splice site selection, rate of splicing, and alternative splicing on nascent transcripts." Genes Dev **2**(6): 754-65.
Biamonti, G. et J. F. Caceres (2009). "Cellular stress and RNA splicing." Trends Biochem Sci **34**(3): 146-53.
Bienroth, S., W. Keller et E. Wahle (1993). "Assembly of a processive messenger RNA polyadenylation complex." Embo J **12**(2): 585-94.
Black, D. L. (1992). "Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro." Cell **69**(5): 795-807.
Black, D. L. (2003). "Mechanisms of alternative pre-mRNA splicing." Annu. Rev. Biochem. **72**: 291-336.
Blanchette, M. et B. Chabot (1997). "A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B." Rna **3**(4): 405-19.
Blanchette, M. et B. Chabot (1999). "Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization [In Process Citation]." Embo J **18**(7): 1939-52.
Blanchette, M., R. E. Green, S. E. Brenner et D. C. Rio (2005). "Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila." Genes Dev **19**(11): 1306-14.
Blaustein, M., F. Pelisch, O. A. Coso, M. J. Bissell, A. R. Kornblihtt et A. Srebrow (2004). "Mammary epithelial-mesenchymal interaction regulates fibronectin alternative splicing via phosphatidylinositol 3-kinase." J Biol Chem **279**(20): 21029-37.
Blaustein, M., F. Pelisch et A. Srebrow (2007). "Signals, pathways and splicing regulation." Int J Biochem Cell Biol **39**(11): 2031-48.
Blechingberg, J., S. Lykke-Andersen, T. H. Jensen, A. L. Jorgensen et A. L. Nielsen (2007). "Regulatory mechanisms for 3'-end alternative splicing and polyadenylation of the Glial Fibrillary Acidic Protein, GFAP, transcript." Nucleic Acids Res **35**(22): 7636-50.
Blencowe, B. J., G. Bauren, A. G. Eldridge, R. Issner, J. A. Nickerson, E. Rosonina*, et al.* (2000). "The SRm160/300 splicing coactivator subunits." Rna **6**(1): 111-20.
Blencowe, B. J., R. Issner, J. A. Nickerson et P. A. Sharp (1998). "A coactivator of pre-mRNA splicing." Genes Dev **12**(7): 996-1009.
Bleoo, S., X. Sun, M. J. Hendzel, J. M. Rowe, M. Packer et R. Godbout (2001). "Association of human DEAD box protein DDX1 with a cleavage stimulation factor involved in 3'-end processing of pre-MRNA." Mol Biol Cell **12**(10): 3046-59.
Boelens, W. C., E. J. Jansen, W. J. van Venrooij, R. Stripecke, I. W. Mattaj et S. I. Gunderson (1993). "The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA." Cell **72**(6): 881-92.
Boise, L. H., M. Gonzalez-Garcia, C. E. Postema, L. Ding, T. Lindsten, L. A. Turka*, et al.* (1993). "bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death." Cell **74**(4): 597-608.

Bon, E., S. Casaregola, G. Blandin, B. Llorente, C. Neuveglise, M. Munsterkotter, *et al.* (2003). "Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns." Nucleic Acids Res **31**(4): 1121-35.

Bourgeois, C. F., F. Lejeune et J. Stevenin, Eds. (2004). Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. Prog Nucleic Acid Res Mol Biol.

Bousquet-Antonelli, C., C. Presutti et D. Tollervey (2000). "Identification of a regulated pathway for nuclear pre-mRNA turnover." Cell **102**(6): 765-75.

Bracco, L. et J. Kearsey (2003). "The relevance of alternative RNA splicing to pharmacogenomics." Trends Biotechnol **21**(8): 346-53.

Brichta, L., Y. Hofmann, E. Hahnen, F. A. Siebzehnrubl, H. Raschke, I. Blumcke, *et al.* (2003). "Valproic acid increases the SMN2 protein level: a well-known drug as a potential therapy for spinal muscular atrophy." Hum Mol Genet **12**(19): 2481-9.

Brown, K. M. et G. M. Gilmartin (2003). "A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im." Mol Cell **12**(6): 1467-76.

Buratti, E., A. F. Muro, M. Giombi, D. Gherbassi, A. Iaconcig et F. E. Baralle (2004). "RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon." Mol Cell Biol **24**(3): 1387-400.

Burd, C. G. et G. Dreyfuss (1994). "RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing." Embo J **13**(5): 1197-204.

Burge, C. B., T. Tuschl et P. A. Sharp (1999). Splicing of Precursors to mRNAs by the Spliceosomes. The RNA World Second edition. S. R. Gesteland, T. Cech and J. Atkins. Cold Spring Harbor, New-York, Cold Spring Harbor Laboratory Press**:** 525-560.

Butcher, S. E. et D. A. Brow (2005). "Towards understanding the catalytic core structure of the spliceosome." Biochem Soc Trans **33**(Pt 3): 447-9.

Buxade, M., J. L. Parra, S. Rousseau, N. Shpiro, R. Marquez, N. Morrice, *et al.* (2005). "The Mnks are novel components in the control of TNF alpha biosynthesis and phosphorylate and regulate hnRNP A1." Immunity **23**(2): 177-89.

Caceres, J. F. et A. R. Kornblihtt (2002). "Alternative splicing: multiple control mechanisms and involvement in human disease." Trends Genet **18**(4): 186-93.

Caceres, J. F., T. Misteli, G. R. Screaton, D. L. Spector et A. R. Krainer (1997). "Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity [In Process Citation]." J Cell Biol **138**(2): 225-38.

Caceres, J. F., G. R. Screaton et A. R. Krainer (1998). "A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm." Genes Dev **12**(1): 55-66.

Caceres, J. F., S. Stamm, D. M. Helfman et A. R. Krainer (1994). "Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors." Science **265**(5179): 1706-9.

Calvo, O. et J. L. Manley (2003). "Strange bedfellows: polyadenylation factors at the promoter." Genes Dev **17**(11): 1321-7.

Camats, M., S. Guil, M. Kokolo et M. Bach-Elias (2008). "P68 RNA helicase (DDX5) alters activity of cis- and trans-acting factors of the alternative splicing of H-Ras." PLoS ONE **3**(8): e2926.

Caputi, M., A. Mayeda, A. R. Krainer et A. M. Zahler (1999). "hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing." Embo J **18**(14): 4060-7.

Carey, J., V. Cameron, P. L. de Haseth et O. C. Uhlenbeck (1983). "Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site." Biochemistry **22**(11): 2601-10.

Carlo, T., R. Sierra et S. M. Berget (2000). "A 5' splice site-proximal enhancer binds SF1 and activates exon bridging of a microexon." Mol Cell Biol **20**(11): 3988-95.

Carmo-Fonseca, M., R. Pepperkok, M. T. Carvalho et A. I. Lamond (1992). "Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies." J Cell Biol **117**(1): 1-14.

Carstens, R. P., E. J. Wagner et M. A. Garcia-Blanco (2000). "An intronic splicing silencer causes skipping of the IIIb exon of fibroblast growth factor receptor 2 through involvement of polypyrimidine tract binding protein [In Process Citation]." Mol Cell Biol **20**(19): 7388-400.

Cartegni, L. et A. R. Krainer (2003). "Correction of disease-associated exon skipping by synthetic exon-specific activators." Nat Struct Biol **10**(2): 120-5.

Cartegni, L., J. Wang, Z. Zhu, M. Q. Zhang et A. R. Krainer (2003). "ESEfinder: A web resource to identify exonic splicing enhancers." Nucleic Acids Res **31**(13): 3568-71.

Cascino, I., G. Fiucci, G. Papoff et G. Ruberti (1995). "Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing." J Immunol **154**(6): 2706-13.

Castelo-Branco, P., A. Furger, M. Wollerton, C. Smith, A. Moreira et N. Proudfoot (2004). "Polypyrimidine Tract Binding Protein Modulates Efficiency of Polyadenylation." Mol Cell Biol **24**(10): 4174-4183.

Chabot, B., C. LeBel, S. Hutchison, F. H. Nasim et M. J. Simard (2003). "Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA." Prog Mol Subcell Biol **31**: 59-88.

Chan, R. C. et D. L. Black (1997). "The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream." Mol Cell Biol **17**(8): 4667-76.

Chang, J. G., H. M. Hsieh-Li, Y. J. Jong, N. M. Wang, C. H. Tsai et H. Li (2001). "Treatment of spinal muscular atrophy by sodium butyrate." Proc Natl Acad Sci U S A **98**(17): 9808-13.

Chang, Y. F., J. S. Imam et M. F. Wilkinson (2007). "The nonsense-mediated decay RNA surveillance pathway." Annu Rev Biochem **76**: 51-74.

Chapman, R. D., M. Heidemann, T. K. Albert, R. Mailhammer, A. Flatley, M. Meisterernst*, et al.* (2007). "Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7." Science **318**(5857): 1780-2.

Charlet, B. N., P. Logan, G. Singh et T. A. Cooper (2002). "Dynamic Antagonism between ETR-3 and PTB Regulates Cell Type-Specific Alternative Splicing." Mol Cell **9**(3): 649-58.

Chebli, K., R. Gattoni, P. Schmitt, G. Hildwein et J. Stevenin (1989). "The 216-nucleotide intron of the E1A pre-mRNA contains a hairpin structure that permits utilization of unusually distant branch acceptors." Mol Cell Biol **9**(11): 4852-61.

Chen, H. C., W. C. Lin, Y. G. Tsay, S. C. Lee et C. J. Chang (2002). "An RNA helicase, DDX1, interacting with poly(A) RNA and heterogeneous nuclear ribonucleoprotein K." J Biol Chem **277**(43): 40403-9.

Chen, H. H., J. G. Chang, R. M. Lu, T. Y. Peng et W. Y. Tarn (2008). "The RNA binding protein hnRNP Q modulates the utilization of exon 7 in the survival motor neuron 2 (SMN2) gene." Mol Cell Biol **28**(22): 6929-38.

Chen, J. F., E. M. Mandel, J. M. Thomson, Q. Wu, T. E. Callis, S. M. Hammond, *et al.* (2006). "The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation." Nat Genet **38**(2): 228-33.

Chen, Y. I., R. E. Moore, H. Y. Ge, M. K. Young, T. D. Lee et S. W. Stevens (2007). "Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors." Nucleic Acids Res **35**(12): 3928-44.

Cheng, C. et P. A. Sharp (2006). "Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion." Mol Cell Biol **26**(1): 362-70.

Cheng, H., K. Dufu, C. S. Lee, J. L. Hsu, A. Dias et R. Reed (2006). "Human mRNA export machinery recruited to the 5' end of mRNA." Cell **127**(7): 1389-400.

Chiara, M. D., O. Gozani, M. Bennett, P. Champion-Arnaud, L. Palandjian et R. Reed (1996). "Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly." Mol Cell Biol **16**(7): 3317-26.

Cho, E. J., T. Takagi, C. R. Moore et S. Buratowski (1997). "mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain." Genes Dev **11**(24): 3319-26.

Choi, H. K., K. C. Choi, H. B. Kang, H. C. Kim, Y. H. Lee, S. Haam, *et al.* (2008). "Function of multiple Lis-Homology domain/WD-40 repeat-containing proteins in feed-forward transcriptional repression by silencing mediator for retinoic and thyroid receptor/nuclear receptor corepressor complexes." Mol Endocrinol **22**(5): 1093-104.

Choi, K. S., A. Mizutani et M. M. Lai (2004). "SYNCRIP, a member of the heterogeneous nuclear ribonucleoprotein family, is involved in mouse hepatitis virus RNA synthesis." J Virol **78**(23): 13153-62.

Chou, M. Y., J. G. Underwood, J. Nikolic, M. H. Luu et D. L. Black (2000). "Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing." Mol Cell **5**(6): 949-57.

Clark, T. A., A. C. Schweitzer, T. X. Chen, M. K. Staples, G. Lu, H. Wang, *et al.* (2007). "Discovery of tissue-specific exons using comprehensive human exon microarrays." Genome Biol **8**(4): R64.

Cohen-Krausz, S., R. Sperling et J. Sperling (2007). "Exploring the architecture of the intact supraspliceosome using electron microscopy." J Mol Biol **368**(2): 319-27.

Coles, L. S., M. A. Bartley, A. Bert, J. Hunter, S. Polyak, P. Diamond, *et al.* (2004). "A multi-protein complex containing cold shock domain (Y-box) and polypyrimidine tract binding proteins forms on the vascular endothelial growth factor mRNA. Potential role in mRNA stabilization." Eur J Biochem **271**(3): 648-60.

Conte, M. R., T. Grune, J. Ghuman, G. Kelly, A. Ladas, S. Matthews, *et al.* (2000). "Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold." Embo J **19**(12): 3132-41.

Cooke, C. et J. C. Alwine (2002). "Characterization of specific protein-RNA complexes associated with the coupling of polyadenylation and last-intron removal." Mol Cell Biol **22**(13): 4579-86.

Cooke, C., H. Hans et J. C. Alwine (1999). "Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal." Mol Cell Biol **19**(7): 4971-9.

Cooper, T. A. (1999). "Defining pre-mRNA cis elements that regulate cell-specific splicing." Methods Mol Biol **118**: 391-403.

Corcos, L. et S. Solier (2005). "[Alternative mRNA splicing, pathology and molecular therapeutics]." Med Sci (Paris) **21**(3): 253-60.

Cote, C. A., D. Gautreau, J. M. Denegre, T. L. Kress, N. A. Terry et K. L. Mowry (1999). "A Xenopus protein related to hnRNP I has a role in cytoplasmic RNA localization." Mol Cell **4**(3): 431-7.

Coulter, L. R., M. A. Landree et T. A. Cooper (1997). "Identification of a new class of exonic splicing enhancers by in vivo selection." Mol Cell Biol **17**(4): 2143-50.

Coutinho-Mansfield, G. C., Y. Xue, Y. Zhang et X. D. Fu (2007). "PTB/nPTB switch: a post-transcriptional mechanism for programming neuronal differentiation." Genes Dev **21**(13): 1573-7.

Cramer, P., J. F. Caceres, D. Cazalla, S. Kadener, A. F. Muro, F. E. Baralle*, et al.* (1999). "Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer." Mol Cell **4**(2): 251-8.

Cramer, P., C. G. Pesce, F. E. Baralle et A. R. Kornblihtt (1997). "Functional association between promoter structure and transcript alternative splicing." Proc Natl Acad Sci U S A **94**(21): 11456-60.

Das, R. et R. Reed (1999). "Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels." Rna **5**(11): 1504-8.

Das, R., J. Yu, Z. Zhang, M. P. Gygi, A. R. Krainer, S. P. Gygi*, et al.* (2007). "SR proteins function in coupling RNAP II transcription to pre-mRNA splicing." Mol Cell **26**(6): 867-81.

Das, R., Z. Zhou et R. Reed (2000). "Functional association of U2 snRNP with the ATP-independent spliceosomal complex E." Mol Cell **5**(5): 779-87.

Datar, K. V., G. Dreyfuss et M. S. Swanson (1993). "The human hnRNP M proteins: identification of a methionine/arginine-rich repeat motif in ribonucleoproteins." Nucleic Acids Res **21**(3): 439-46.

de la Grange, P., M. Dutertre, N. Martin et D. Auboeuf (2005). "FAST DB: a website resource for the study of the expression regulation of human gene products." Nucleic Acids Res **33**(13): 4276-84.

de la Mata, M., C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein, F. Pelisch*, et al.* (2003). "A slow RNA polymerase II affects alternative splicing in vivo." Mol Cell **12**(2): 525-32.

de Moor, C. H. et J. D. Richter (2001). "Translational control in vertebrate development." Int Rev Cytol **203**: 567-608.

de Vries, H., U. Ruegsegger, W. Hubner, A. Friedlein, H. Langen et W. Keller (2000). "Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors." Embo J **19**(21): 5895-904.

Deckert, J., K. Hartmuth, D. Boehringer, N. Behzadnia, C. L. Will, B. Kastner*, et al.* (2006). "Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions." Mol Cell Biol **26**(14): 5528-43.

Del Gatto-Konczak, F., M. Olive, M. C. Gesnel et R. Breathnach (1999). "hnRNP A1 recruited to an exon In vivo can function as an exon splicing silencer [In Process Citation]." Mol Cell Biol **19**(1): 251-60.

Dember, L. M., N. D. Kim, K. Q. Liu et P. Anderson (1996). "Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities." J Biol Chem **271**(5): 2783-8.

Demir, E. et B. J. Dickson (2005). "fruitless splicing specifies male courtship behavior in Drosophila." Cell **121**(5): 785-94.

Deshler, J. O. et J. J. Rossi (1991). "Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron." Genes Dev **5**(7): 1252-63.

Dettwiler, S., C. Aringhieri, S. Cardinale, W. Keller et S. M. Barabino (2004). "Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization." J Biol Chem **279**(34): 35788-97.

Dignam, J. D., R. M. Lebovitz et R. G. Roeder (1983). "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei." Nucleic Acids Res **11**(5): 1475-89.

Ding, J. H., X. Xu, D. Yang, P. H. Chu, N. D. Dalton, Z. Ye, *et al.* (2004). "Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart." Embo J **23**(4): 885-96.

Dix, I., C. S. Russell, R. T. O'Keefe, A. J. Newman et J. D. Beggs (1998). "Protein-RNA interactions in the U5 snRNP of Saccharomyces cerevisiae." Rna **4**(10): 1239-50.

Dowhan, D. H., E. P. Hong, D. Auboeuf, A. P. Dennis, M. M. Wilson, S. M. Berget, *et al.* (2005). "Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta." Mol Cell **17**(3): 429-39.

Dredge, B. K. et R. B. Darnell (2003). "Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer." Mol Cell Biol **23**(13): 4687-700.

Dredge, B. K., G. Stefani, C. C. Engelhard et R. B. Darnell (2005). "Nova autoregulation reveals dual functions in neuronal splicing." Embo J **24**(8): 1608-20.

Dreyfuss, G., M. J. Matunis, S. Pinol-Roma et C. G. Burd (1993). "hnRNP proteins and the biogenesis of mRNA." Annu Rev Biochem **62**: 289-321.

Duncan, R., L. Bazar, G. Michelotti, T. Tomonaga, H. Krutzsch, M. Avigan, *et al.* (1994). "A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif." Genes Dev **8**(4): 465-80.

Dye, M. J., N. Gromak et N. J. Proudfoot (2006). "Exon tethering in transcription by RNA polymerase II." Mol Cell **21**(6): 849-59.

Dye, M. J. et N. J. Proudfoot (2001). "Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II." Cell **105**(5): 669-81.

Edwalds-Gilbert, G. et C. Milcarek (1995). "The binding of a subunit of the general polyadenylation factor cleavage- polyadenylation specificity factor (CPSF) to polyadenylation sites changes during B cell development." Nucleic Acids Symp Ser(33): 229-33.

Egloff, S. et S. Murphy (2008). "Cracking the RNA polymerase II CTD code." Trends Genet **24**(6): 280-8.

Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, *et al.* (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." Science **318**(5857): 1777-9.

Eldridge, A. G., Y. Li, P. A. Sharp et B. J. Blencowe (1999). "The SRm160/300 splicing coactivator is required for exon-enhancer function." Proc Natl Acad Sci U S A **96**(11): 6125-30.

Ellis, J. D., D. Lleres, M. Denegri, A. I. Lamond et J. F. Caceres (2008). "Spatial mapping of splicing factor complexes involved in exon and intron definition." J Cell Biol **181**(6): 921-34.

Eperon, I. C., O. V. Makarova, A. Mayeda, S. H. Munroe, J. F. Caceres, D. G. Hayward, *et al.* (2000). "Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1 [In Process Citation]." Mol Cell Biol **20**(22): 8303-18.

Expert-Bezancon, A., J. P. Le Caer et J. Marie (2002). "Heterogeneous Nuclear Ribonucleoprotein (hnRNP) K Is a Component of an Intronic Splicing Enhancer

Complex That Activates the Splicing of the Alternative Exon 6A from Chicken beta - Tropomyosin Pre-mRNA." J Biol Chem **277**(19): 16614-23.

Expert-Bezancon, A., A. Sureau, P. Durosay, R. Salesse, H. Groeneveld, J. P. Lecaer, *et al.* (2004). "hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B." J Biol Chem **279**(37): 38249-59.

Fabrizio, P., S. Esser, B. Kastner et R. Luhrmann (1994). "Isolation of S. cerevisiae snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts." Science **264**(5156): 261-5.

Faustino, N. A. et T. A. Cooper (2005). "Identification of putative new splicing targets for ETR-3 using sequences identified by systematic evolution of ligands by exponential enrichment." Mol Cell Biol **25**(3): 879-87.

Fededa, J. P. et A. R. Kornblihtt (2008). "A splicing regulator promotes transcriptional elongation." Nat Struct Mol Biol **15**(8): 779-81.

Fededa, J. P., E. Petrillo, M. S. Gelfand, A. D. Neverov, S. Kadener, G. Nogues, *et al.* (2005). "A polar mechanism coordinates different regions of alternative splicing within a single gene." Mol Cell **19**(3): 393-404.

Fehlbaum, P., C. Guihal, L. Bracco et O. Cochet (2005). "A microarray configuration to quantify expression levels and relative abundance of splice variants." Nucleic Acids Res **33**(5): e47.

Feng, Y., M. Chen et J. L. Manley (2008). "Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator." Nat Struct Mol Biol **15**(10): 1040-8.

Forch, P., O. Puig, C. Martinez, B. Seraphin et J. Valcarcel (2002). "The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites." Embo J **21**(24): 6882-92.

Forch, P. et J. Valcarcel (2001). "Molecular mechanisms of gene expression regulation by the apoptosis-promoting protein TIA-1." Apoptosis **6**(6): 463-8.

Forch, P. et J. Valcarcel (2003). "Splicing regulation in Drosophila sex determination." Prog Mol Subcell Biol **31**: 127-51.

Fortes, P., D. Longman, S. McCracken, J. Y. Ip, R. Poot, I. W. Mattaj, *et al.* (2007). "Identification and characterization of RED120: a conserved PWI domain protein with links to splicing and 3'-end formation." FEBS Lett **581**(16): 3087-97.

Freeman, B. C. et K. R. Yamamoto (2002). "Disassembly of transcriptional regulatory complexes by molecular chaperones." Science **296**(5576): 2232-5.

Frischmeyer, P. A., A. van Hoof, K. O'Donnell, A. L. Guerrerio, R. Parker et H. C. Dietz (2002). "An mRNA surveillance mechanism that eliminates transcripts lacking termination codons." Science **295**(5563): 2258-61.

Furth, P. A., W. T. Choe, J. H. Rex, J. C. Byrne et C. C. Baker (1994). "Sequences homologous to 5' splice sites are required for the inhibitory activity of papillomavirus late 3' untranslated regions." Mol Cell Biol **14**(8): 5278-89.

Gaillard, C., N. Theze, H. Lerivray, S. Hardy, D. Lepetit et P. Thiebaud (1998). "A novel tropomyosin isoform encoded by the Xenopus laevis alpha-TM gene is expressed in the brain." Gene **207**(2): 235-9.

Gallego, M. E., R. Gattoni, J. Stevenin, J. Marie et A. Expert-Bezancon (1997). "The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A." Embo J **16**(7): 1772-84.

Gama-Carvalho, M., R. D. Krauss, L. Chiang, J. Valcarcel, M. R. Green et M. Carmo-Fonseca (1997). "Targeting of U2AF65 to sites of active splicing in the nucleus." J Cell Biol **137**(5): 975-87.

Garcia-Blanco, M. A., S. F. Jamison et P. A. Sharp (1989). "Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns." Genes Dev **3**(12A): 1874-86.

Gebauer, F., L. Merendino, M. W. Hentze et J. Valcarcel (1998). "The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA." Rna **4**(2): 142-50.

Gendra, E., D. F. Colgan, B. Meany et M. M. Konarska (2007). "A sequence motif in the simian virus 40 (SV40) early core promoter affects alternative splicing of transcribed mRNA." J Biol Chem **282**(16): 11648-57.

Gesnel, M. C., S. Theoleyre, F. Del Gatto-Konczak et R. Breathnach (2007). "Cooperative binding of TIA-1 and U1 snRNP in K-SAM exon splicing activation." Biochem Biophys Res Commun **358**(4): 1065-70.

Gilbert, W. (1978). "Why genes in pieces?" Nature **271**(5645): 501.

Goncalves, V., P. Matos et P. Jordan (2008). "The beta-catenin/TCF4 pathway modifies alternative splicing through modulation of SRp20 expression." Rna **14**(12): 2538-49.

Gooding, C., P. Kemp et C. W. Smith (2003). "SmPTB: a novel polypyrimidine tract binding protein (PTB) paralog expressed in smooth muscle cells." J Biol Chem.

Gosert, R., K. H. Chang, R. Rijnbrand, M. Yi, D. V. Sangar et S. M. Lemon (2000). "Transient expression of cellular polypyrimidine-tract binding protein stimulates cap-independent translation directed by both picornaviral and flaviviral internal ribosome entry sites In vivo." Mol Cell Biol **20**(5): 1583-95.

Grabowski, P. J. et D. L. Black (2001). "Alternative RNA splicing in the nervous system." Prog Neurobiol **65**(3): 289-308.

Grabowski, P. J. et P. A. Sharp (1986). "Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome." Science **233**(4770): 1294-9.

Graveley, B. R. (2000). "Sorting out the complexity of SR protein functions." Rna **6**(9): 1197-211.

Graveley, B. R., K. J. Hertel et T. Maniatis (2001). "The role of U2AF35 and U2AF65 in enhancer-dependent splicing." Rna **7**(6): 806-18.

Green, N. M. (1975). "Avidin." Adv Protein Chem **29**: 85-133.

Gromak, N., A. J. Matlin, T. A. Cooper et C. W. Smith (2003). "Antagonistic regulation of alpha-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein." Rna **9**(4): 443-456.

Gromak, N., A. Rideau, J. Southby, A. D. Scadden, C. Gooding, S. Huttelmaier, *et al.* (2003). "The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing." Embo J **22**(23): 6356-6364.

Gromak, N., S. West et N. J. Proudfoot (2006). "Pause sites promote transcriptional termination of mammalian RNA polymerase II." Mol Cell Biol **26**(10): 3986-96.

Gross, S. et C. L. Moore (2001). "Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation." Mol Cell Biol **21**(23): 8045-55.

Grover, A., H. Houlden, M. Baker, J. Adamson, J. Lewis, G. Prihar, *et al.* (1999). "5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10." J Biol Chem **274**(21): 15134-43.

Gruber, J. D., P. B. Colligan et J. K. Wolford (2002). "Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing." Hum Genet **110**(5): 395-401.

Gudipati, R. K., T. Villa, J. Boulay et D. Libri (2008). "Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice." Nat Struct Mol Biol **15**(8): 786-94.

Gunderson, S. I., M. Polycarpou-Schwarz et I. W. Mattaj (1998). "UI snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase." Mol. Cell **1**: 255-264.

Gunderson, S. I., S. Vagner, M. Polycarpou-Schwarz et I. W. Mattaj (1997). "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): 761-73.

Guo, M., P. C. Lo et S. M. Mount (1993). "Species-specific signals for the splicing of a short Drosophila intron in vitro." Mol Cell Biol **13**(2): 1104-18.

Guth, S., T. O. Tange, E. Kellenberger et J. Valcarcel (2001). "Dual function for U2AF(35) in AG-dependent pre-mRNA splicing." Mol Cell Biol **21**(22): 7673-81.

Guth, S. et J. Valcarcel (2000). "Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF." J Biol Chem **275**(48): 38059-66.

Hagiwara, M. (2005). "Alternative splicing: a new drug target of the post-genome era." Biochim Biophys Acta **1754**(1-2): 324-31.

Hagiwara, M. et T. Nojima (2007). "Cross-talks between transcription and post-transcriptional events within a 'mRNA factory'." J Biochem **142**(1): 11-5.

Hall-Pogar, T., S. Liang, L. K. Hague et C. S. Lutz (2007). "Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR." Rna **13**(7): 1103-15.

Hamilton, B. J., A. Genin, R. Q. Cron et W. F. Rigby (2003). "Delineation of a novel pathway that regulates CD154 (CD40 ligand) expression." Mol Cell Biol **23**(2): 510-25.

Hamon Gouault, S. (2002). Etude de la maturation differentielle du gène α-fast tropomyosine chez Xenopus laevis : identification d'éléments en cis régulant l'utilisation d'un exon composite interne/3' terminal. Rennes, Université de RENNES 1: 188.

Hamon, S., C. Le Sommer, A. Mereau, M. R. Allo et S. Hardy (2004). "Polypyrimidine tract-binding protein is involved in vivo in repression of a composite internal/3' -terminal exon of the Xenopus alpha-tropomyosin Pre-mRNA." J Biol Chem **279**(21): 22166-75.

Handwerger, K. E. et J. G. Gall (2006). "Subnuclear organelles: new insights into form and function." Trends Cell Biol **16**(1): 19-26.

Hardy, S., M. Y. Fiszman, H. B. Osborne et P. Thiebaud (1991). "Characterization of muscle and non muscle Xenopus laevis tropomyosin mRNAs transcribed from the same gene. Developmental and tissue-specific expression." Eur J Biochem **202**(2): 431-40.

Hardy, S., S. Hamon, B. Cooper, T. Mohun et P. Thiebaud (1999). "Two skeletal alpha-tropomyosin transcripts with distinct 3'UTR have different temporal and spatial patterns of expression in the striated muscle lineages of Xenopus laevis." Mech Dev **87**(1-2): 199-202.

Hartmuth, K., H. Urlaub, H. P. Vornlocher, C. L. Will, M. Gentzel, M. Wilm, *et al.* (2002). "Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method." Proc Natl Acad Sci U S A **99**(26): 16719-24.

Helfman, D. M., W. M. Ricci et L. A. Finn (1988). "Alternative splicing of tropomyosin pre-mRNAs in vitro and in vivo." Genes Dev **2**(12A): 1627-38.

Hernandez, N. et W. Keller (1983). "Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts." Cell **35**(1): 89-99.

Hertel, K. J. (2008). "Combinatorial control of exon recognition." J Biol Chem **283**(3): 1211-5.

Hiller, M., Z. Zhang, R. Backofen et S. Stamm (2007). "Pre-mRNA secondary structures influence exon recognition." PLoS Genet **3**(11): e204.
Ho, T. H., B. N. Charlet, M. G. Poulos, G. Singh, M. S. Swanson et T. A. Cooper (2004). "Muscleblind proteins regulate alternative splicing." Embo J.
Honig, A., D. Auboeuf, M. M. Parker, B. W. O'Malley et S. M. Berget (2002). "Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72." Mol Cell Biol **22**(16): 5698-707.
House, A. E. et K. W. Lynch (2006). "An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition." Nat Struct Mol Biol **13**(10): 937-44.
House, A. E. et K. W. Lynch (2008). "Regulation of alternative splicing: more than just the ABCs." J Biol Chem **283**(3): 1217-21.
Houseley, J., J. LaCava et D. Tollervey (2006). "RNA-quality control by the exosome." Nat Rev Mol Cell Biol **7**(7): 529-39.
Hovhannisyan, R. H. et R. P. Carstens (2007). "Heterogeneous ribonucleoprotein m is a splicing regulatory protein that can enhance or silence splicing of alternatively spliced exons." J Biol Chem **282**(50): 36265-74.
Hua, Y., T. A. Vickers, H. L. Okunola, C. F. Bennett et A. R. Krainer (2008). "Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice." Am J Hum Genet **82**(4): 834-48.
Huang, Y., R. Gattoni, J. Stevenin et J. A. Steitz (2003). "SR Splicing Factors Serve as Adapter Proteins for TAP-Dependent mRNA Export." Mol Cell **11**(3): 837-43.
Huang, Y. et J. A. Steitz (2001). "Splicing Factors SRp20 and 9G8 Promote the Nucleocytoplasmic Export of mRNA." Mol Cell **7**(4): 899-905.
Hui, J., L. H. Hung, M. Heiner, S. Schreiner, N. Neumuller, G. Reither*, et al.* (2005). "Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing." Embo J **24**(11): 1988-98.
Hui, J., K. Stangl, W. S. Lane et A. Bindereif (2003). "HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats." Nat Struct Biol **10**(1): 33-7.
Hung, L. H., M. Heiner, J. Hui, S. Schreiner, V. Benes et A. Bindereif (2008). "Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis." Rna **14**(2): 284-96.
Huttelmaier, S., S. Illenberger, I. Grosheva, M. Rudiger, R. H. Singer et B. M. Jockusch (2001). "Raver1, a dual compartment protein, is a ligand for PTB/hnRNPI and microfilament attachment proteins." J Cell Biol **155**(5): 775-86.
Iakova, P., G. L. Wang, L. Timchenko, M. Michalak, O. M. Pereira-Smith, J. R. Smith*, et al.* (2004). "Competition of CUGBP1 and calreticulin for the regulation of p21 translation determines cell fate." Embo J **23**(2): 406-17.
Ito, Y., A. Ando, H. Ando, J. Ando, Y. Saijoh, H. Inoko*, et al.* (1998). "Genomic structure of the spermatid-specific hsp70 homolog gene located in the class III region of the major histocompatibility complex of mouse and man." J Biochem **124**(2): 347-53.
Iwasaki, T., W. W. Chin et L. Ko (2001). "Identification and characterization of RRM-containing coactivator activator (CoAA) as TRBP-interacting protein, and its splice variant as a coactivator modulator (CoAM)." J Biol Chem **276**(36): 33375-83.
Izaurralde, E., J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz et I. W. Mattaj (1994). "A nuclear cap binding protein complex involved in pre-mRNA splicing." Cell **78**(4): 657-68.
Izquierdo, J. M. (2008). "Hu antigen R (HuR) functions as an alternative pre-mRNA splicing regulator of Fas apoptosis-promoting receptor on exon definition." J Biol Chem **283**(27): 19077-84.

Izquierdo, J. M., N. Majos, S. Bonnal, C. Martinez, R. Castelo, R. Guigo, *et al.* (2005). "Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition." Mol Cell **19**(4): 475-84.

Jacquenet, S., A. Mereau, P. S. Bilodeau, L. Damier, C. M. Stoltzfus et C. Branlant (2001). "A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H." J Biol Chem **276**(44): 40464-75.

Jelen, N., J. Ule, M. Zivin et R. B. Darnell (2007). "Evolution of Nova-dependent splicing regulation in the brain." PLoS Genet **3**(10): 1838-47.

Jensen, K. B., B. K. Dredge, G. Stefani, R. Zhong, R. J. Buckanovich, H. J. Okano, *et al.* (2000). "Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability [see comments]." Neuron **25**(2): 359-71.

Jin, Y., H. Suzuki, S. Maegawa, H. Endo, S. Sugano, K. Hashimoto, *et al.* (2003). "A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG." Embo J **22**(4): 905-12.

Johnson, J. M., J. Castle, P. Garrett-Engele, Z. Kan, P. M. Loerch, C. D. Armour, *et al.* (2003). "Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays." Science **302**(5653): 2141-4.

Jumaa, H., J. L. Guenet et P. J. Nielsen (1997). "Regulated expression and RNA processing of transcripts from the Srp20 splicing factor gene during the cell cycle." Mol Cell Biol **17**(6): 3116-24.

Jurica, M. S., L. J. Licklider, S. R. Gygi, N. Grigorieff et M. J. Moore (2002). "Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis." Rna **8**(4): 426-39.

Jurica, M. S. et M. J. Moore (2003). "Pre-mRNA splicing: awash in a sea of proteins." Mol Cell **12**(1): 5-14.

Kadener, S., P. Cramer, G. Nogues, D. Cazalla, M. de la Mata, J. P. Fededa, *et al.* (2001). "Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing." Embo J **20**(20): 5759-68.

Kadener, S., J. P. Fededa, M. Rosbash et A. R. Kornblihtt (2002). "Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation." Proc Natl Acad Sci U S A **99**(12): 8185-90.

Kalsotra, A., X. Xiao, A. J. Ward, J. C. Castle, J. M. Johnson, C. B. Burge, *et al.* (2008). "A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart." Proc Natl Acad Sci U S A **105**(51): 20333-8.

Kamath, R. V., D. J. Leary et S. Huang (2001). "Nucleocytoplasmic shuttling of polypyrimidine tract-binding protein is uncoupled from RNA export." Mol Biol Cell **12**(12): 3808-20.

Kameoka, S., P. Duque et M. M. Konarska (2004). "p54(nrb) associates with the 5' splice site within large transcription/splicing complexes." Embo J **23**(8): 1782-91.

Kastner, B., M. Bach et R. Luhrmann (1990). "Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family." Proc Natl Acad Sci U S A **87**(5): 1710-4.

Katzenberger, R. J., M. S. Marengo et D. A. Wassarman (2009). "Control of alternative splicing by signal-dependent degradation of splicing-regulatory proteins." J Biol Chem **284**(16): 10737-46.

Kaufmann, I., G. Martin, A. Friedlein, H. Langen et W. Keller (2004). "Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase." Embo J **23**(3): 616-26.

Keene, J. D., J. M. Komisarow et M. B. Friedersdorf (2006). "RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts." Nat Protoc **1**(1): 302-7.

Kent, O. A., D. B. Ritchie et A. M. Macmillan (2005). "Characterization of a U2AF-independent commitment complex (E') in the mammalian spliceosome assembly pathway." Mol Cell Biol **25**(1): 233-40.

Kerwitz, Y., U. Kuhn, H. Lilie, A. Knoth, T. Scheuermann, H. Friedrich, *et al.* (2003). "Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA." Embo J **22**(14): 3705-14.

Khacho, M., K. Mekhail, K. Pilon-Larose, A. Pause, J. Cote et S. Lee (2008). "eEF1A is a novel component of the mammalian nuclear protein export machinery." Mol Biol Cell **19**(12): 5296-308.

Kielkopf, C. L., S. Lucke et M. R. Green (2004). "U2AF homology motifs: protein recognition in the RRM world." Genes Dev **18**(13): 1513-26.

Kim, E., L. Du, D. B. Bregman et S. L. Warren (1997). "Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA." J Cell Biol **136**(1): 19-28.

Kim, J. H., B. Hahm, Y. K. Kim, M. Choi et S. K. Jang (2000). "Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm." J Mol Biol **298**(3): 395-405.

Kim, M., S. H. Ahn, N. J. Krogan, J. F. Greenblatt et S. Buratowski (2004). "Transitions in RNA polymerase II elongation complexes at the 3' ends of genes." Embo J **23**(2): 354-64.

Kirschbaum-Slager, N., G. M. Lopes, P. A. Galante, G. J. Riggins et S. J. de Souza (2004). "Splicing factors are differentially expressed in tumors." Genet Mol Res **3**(4): 512-20.

Klein Gunnewiek, J. M., R. I. Hussein, Y. van Aarssen, D. Palacios, R. de Jong, W. J. van Venrooij, *et al.* (2000). "Fourteen residues of the U1 snRNP-specific U1A protein are required for homodimerization, cooperative RNA binding, and inhibition of polyadenylation." Mol Cell Biol **20**(6): 2209-17.

Kleinhenz, B., M. Fabienke, S. Swiniarski, N. Wittenmayer, J. Kirsch, B. M. Jockusch, *et al.* (2005). "Raver2, a new member of the hnRNP family." FEBS Lett **579**(20): 4254-8.

Klinz, F. J. et D. Gallwitz (1985). "Size and position of intervening sequences are critical for the splicing efficiency of pre-mRNA in the yeast Saccharomyces cerevisiae." Nucleic Acids Res **13**(11): 3791-804.

Knoch, K. P., H. Bergert, B. Borgonovo, H. D. Saeger, A. Altkruger, P. Verkade, *et al.* (2004). "Polypyrimidine tract-binding protein promotes insulin secretory granule biogenesis." Nat Cell Biol **6**(3): 207-14.

Ko, B. et S. I. Gunderson (2002). "Identification of new poly(A) polymerase-inhibitory proteins capable of regulating pre-mRNA polyadenylation." J Mol Biol **318**(5): 1189-206.

Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, *et al.* (1994). "Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors." Nature **368**(6467): 119-24.

Koide, T., T. Hayata et K. W. Cho (2005). "Xenopus as a model system to study transcriptional regulatory networks." Proc Natl Acad Sci U S A **102**(14): 4943-8.

Konarska, M. M. et P. A. Sharp (1988). "Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA." Proc Natl Acad Sci U S A **85**(15): 5459-62.

Kornblihtt, A. R. (2006). "Chromatin, transcript elongation and alternative splicing." Nat Struct Mol Biol **13**(1): 5-7.

Krainer, A. R., G. C. Conway et D. Kozak (1990). "The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites." Cell **62**(1): 35-42.
Kramer, A. et U. Utans (1991). "Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs." Embo J **10**(6): 1503-9.
Krammer, P. H. (2000). "CD95's deadly mission in the immune system." Nature **407**(6805): 789-95.
Kreivi, J. P. et A. I. Lamond (1996). "RNA splicing: unexpected spliceosome diversity." Curr Biol **6**(7): 802-5.
Kuehn, U., M. Guendel, A. Knoth, Y. Kerwitz, S. Ruedel et E. Wahle (2009). "Poly(A) tail length is controlled by the nuclear poly(A) binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor." J Biol Chem.
Kuersten, S. et E. B. Goodwin (2003). "The power of the 3' UTR: translational control and development." Nat Rev Genet **4**(8): 626-37.
Kyburz, A., A. Friedlein, H. Langen et W. Keller (2006). "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.
Ladd, A. N., B. N. Charlet et T. A. Cooper (2001). "The CELF Family of RNA Binding Proteins Is Implicated in Cell-Specific and Developmentally Regulated Alternative Splicing." Mol Cell Biol **21**(4): 1285-1296.
Ladd, A. N., N. H. Nguyen, K. Malhotra et T. A. Cooper (2004). "CELF6, a member of the CELF family of RNA binding proteins, regulates MSE-dependent alternative splicing." J Biol Chem.
Ladd, A. N., M. G. Stenberg, M. S. Swanson et T. A. Cooper (2005). "Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development." Dev Dyn **233**(3): 783-93.
Ladd, A. N., G. Taffet, C. Hartley, D. L. Kearney et T. A. Cooper (2005). "Cardiac tissue-specific repression of CELF activity disrupts alternative splicing and causes cardiomyopathy." Mol Cell Biol **25**(14): 6267-78.
Lam, B. J., A. Bakshi, F. Y. Ekinci, J. Webb, B. R. Graveley et K. J. Hertel (2003). "Enhancer-dependent 5'-splice site control of fruitless pre-mRNA splicing." J Biol Chem **278**(25): 22740-7.
Lamond, A. I. et D. L. Spector (2003). "Nuclear speckles: a model for nuclear organelles." Nat Rev Mol Cell Biol **4**(8): 605-12.
Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, *et al.* (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Langford, C. J. et D. Gallwitz (1983). "Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts." Cell **33**(2): 519-27.
Langland, J. O., S. M. Pettiford et B. L. Jacobs (1995). "Nucleic acid affinity chromatography: preparation and characterization of double-stranded RNA agarose." Protein Expr Purif **6**(1): 25-32.
Law, W. J., K. L. Cann et G. G. Hicks (2006). "TLS, EWS and TAF15: a model for transcriptional integration of gene expression." Brief Funct Genomic Proteomic **5**(1): 8-14.
Lazarev, D. et J. L. Manley (2007). "Concurrent splicing and transcription are not sufficient to enhance splicing efficiency." Rna **13**(9): 1546-57.
Le Hir, H., D. Gatfield, E. Izaurralde et M. J. Moore (2001). "The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay." Embo J **20**(17): 4987-97.

Le Hir, H., E. Izaurralde, L. E. Maquat et M. J. Moore (2000). "The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions." Embo J **19**(24): 6860-9.

Le Hir, H. et B. Seraphin (2008). "EJCs at the heart of translational control." Cell **133**(2): 213-6.

Le Sommer, C. (2006). Identification de facteurs régulant en trans la maturation différentielle de la région 3' terminale de l'ARN pré-messager tropomyosine alpha chez Xenopus laevis. Rennes, Université de Rennes 1: 187.

Le Sommer, C., M. Lesimple, A. Mereau, S. Menoret, M. R. Allo et S. Hardy (2005). "PTB regulates the processing of a 3'-terminal exon by repressing both splicing and polyadenylation." Mol Cell Biol **25**(21): 9595-607.

Lee, C. G. (2002). "RH70, a bidirectional RNA helicase, co-purifies with U1snRNP." J Biol Chem **277**(42): 39679-83.

Legrand, P., N. Pinaud, L. Minvielle-Sebastia et S. Fribourg (2007). "The structure of the CstF-77 homodimer provides insights into CstF assembly." Nucleic Acids Res **35**(13): 4515-22.

Lejeune, F. et L. E. Maquat (2005). "Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells." Curr Opin Cell Biol **17**(3): 309-15.

Levy, N., A. Martz, A. Bresson, C. Spenlehauer, G. de Murcia et J. Menissier-de Murcia (2006). "XRCC1 is phosphorylated by DNA-dependent protein kinase in response to DNA damage." Nucleic Acids Res **34**(1): 32-41.

Lewis, J. D. et E. Izaurralde (1997). "The role of the cap structure in RNA processing and nuclear export." Eur J Biochem **247**(2): 461-9.

Libri, D., J. Marie, E. Brody et M. Y. Fiszman (1989). "A subfragment of the beta tropomyosin gene is alternatively spliced when transfected into differentiating muscle cells." Nucleic Acids Res **17**(16): 6449-62.

Lieber, M. R., U. Grawunder, X. Wu et M. Yaneva (1997). "Tying loose ends: roles of Ku and DNA-dependent protein kinase in the repair of double-strand breaks." Curr Opin Genet Dev **7**(1): 99-104.

Lillevali, K., A. Kulla et T. Ord (2001). "Comparative expression analysis of the genes encoding polypyrimidine tract binding protein (PTB) and its neural homologue (brPTB) in prenatal and postnatal mouse brain." Mech Dev **101**(1-2): 217-20.

Lin, C., L. Yang, J. J. Yang, Y. Huang et Z. R. Liu (2005). "ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome." Mol Cell Biol **25**(17): 7484-93.

Lin, C. H. et J. G. Patton (1995). "Regulation of alternative 3' splice site selection by constitutive splicing factors." Rna **1**(3): 234-45.

Lin, S., G. Coutinho-Mansfield, D. Wang, S. Pandit et X. D. Fu (2008). "The splicing factor SC35 has an active role in transcriptional elongation." Nat Struct Mol Biol **15**(8): 819-26.

Listerman, I., A. K. Sapra et K. M. Neugebauer (2006). "Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells." Nat Struct Mol Biol **13**(9): 815-22.

Liu, H., W. Zhang, R. B. Reed, W. Liu et P. J. Grabowski (2002). "Mutations in RRM4 uncouple the splicing repression and RNA-binding activities of polypyrimidine tract binding protein." Rna **8**(2): 137-49.

Liu, H. M., H. Aizaki, K. S. Choi, K. Machida, J. J. Ou et M. M. Lai (2009). "SYNCRIP (synaptotagmin-binding, cytoplasmic RNA-interacting protein) is a host factor involved in hepatitis C virus RNA replication." Virology **386**(2): 249-56.

Liu, H. X., M. Zhang et A. R. Krainer (1998). "Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins." Genes Dev **12**(13): 1998-2012.

Liu, Z., I. Luyten, M. J. Bottomley, A. C. Messias, S. Houngninou-Molango, R. Sprangers*, et al.* (2001). "Structural basis for recognition of the intron branch site RNA by splicing factor 1." Science **294**(5544): 1098-102.

Liu, Z. R. (2002). "p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex." Mol Cell Biol **22**(15): 5443-50.

Long, J. C. et J. F. Caceres (2009). "The SR protein family of splicing factors: master regulators of gene expression." Biochem J **417**(1): 15-27.

Lopez, P. J. et B. Seraphin (1999). "Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition." Rna **5**(9): 1135-7.

Lou, H., G. J. Cote et R. F. Gagel (1994). "The calcitonin exon and its flanking intronic sequences are sufficient for the regulation of human calcitonin/calcitonin gene-related peptide alternative RNA splicing." Mol Endocrinol **8**(12): 1618-26.

Lou, H., R. F. Gagel et S. M. Berget (1996). "An intron enhancer recognized by splicing factors activates polyadenylation." Genes Dev. **10**(2): 208-19.

Lou, H., D. M. Helfman, R. F. Gagel et S. M. Berget (1999). "Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon." Mol Cell Biol **19**(1): 78-85.

Lou, H., K. M. Neugebauer, R. F. Gagel et S. M. Berget (1998). "Regulation of alternative polyadenylation by U1snRNPs and SRp20." Mol. Cell. Biol. **18**(9): 4977-4985.

Lou, H., Y. Yang, G. J. Cote, S. M. Berget et R. F. Gagel (1995). "An intron enhancer containing a 5' splice site sequence in the human calcitonin/calcitonin gene-related peptide gene." Mol Cell Biol **15**(12): 7135-42.

Lowary, P. T. et O. C. Uhlenbeck (1987). "An RNA mutation that increases the affinity of an RNA-protein interaction." Nucleic Acids Res **15**(24): 10483-93.

Lukong, K. E., K. W. Chang, E. W. Khandjian et S. Richard (2008). "RNA-binding proteins in human genetic disease." Trends Genet **24**(8): 416-25.

Lutz, C. S. et J. C. Alwine (1994). "Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal." Genes Dev **8**(5): 576-86.

Lutz, C. S., K. G. K. Murthy, N. Schek, J. P. O'Connor, J. L. Manley et J. C. Alwine (1996). "Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro." Genes Dev. **10**: 325-37.

MacDonald, C. C., J. Wilusz et T. Shenk (1994). "The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location." Mol Cell Biol **14**(10): 6647-54.

Makarov, E. M., O. V. Makarova, H. Urlaub, M. Gentzel, C. L. Will, M. Wilm*, et al.* (2002). "Small Nuclear Ribonucleoprotein Remodeling During Catalytic Activation of the Spliceosome." Science.

Makarova, O. V., E. M. Makarov et R. Luhrmann (2001). "The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes." Embo J **20**(10): 2553-63.

Mandel, C. R., Y. Bai et L. Tong (2008). "Protein factors in pre-mRNA 3'-end processing." Cell Mol Life Sci **65**(7-8): 1099-122.

Mandel, C. R., S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J. L. Manley*, et al.* (2006). "Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease." Nature **444**(7121): 953-6.

Marchand, V., A. Mereau, S. Jacquenet, D. Thomas, A. Mougin, R. Gattoni, *et al.* (2002). "A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding." J Mol Biol **323**(4): 629-52.

Marengo, M. S. et M. A. Garcia-Blanco (2009). "Shedding UV light on alternative splicing." Cell **137**(4): 600-2.

Markovtsov, V., J. M. Nikolic, J. A. Goldman, C. W. Turck, M. Y. Chou et D. L. Black (2000). "Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein [In Process Citation]." Mol Cell Biol **20**(20): 7463-79.

Marquis, J., L. Paillard, Y. Audic, B. Cosson, O. Danos, C. Le Bec, *et al.* (2006). "CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding." Biochem J **400**(2): 291-301.

Martincic, K., R. Campbell, G. Edwalds-Gilbert, L. Souan, M. T. Lotze et C. Milcarek (1998). "Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G0 to S phase transition." Proc Natl Acad Sci U S A **95**(19): 11095-100.

Masuyama, K., I. Taniguchi, K. Okawa et M. Ohno (2007). "Factors associated with a purine-rich exonic splicing enhancer sequence in Xenopus oocyte nucleus." Biochem Biophys Res Commun **359**(3): 580-5.

Matlin, A. J., F. Clark et C. W. Smith (2005). "Understanding alternative splicing: towards a cellular code." Nat Rev Mol Cell Biol **6**(5): 386-98.

Matter, N., P. Herrlich et H. Konig (2002). "Signal-dependent regulation of splicing via phosphorylation of Sam68." Nature **420**(6916): 691-5.

Matter, N., M. Marx, S. Weg-Remers, H. Ponta, P. Herrlich et H. Konig (2000). "Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways." J Biol Chem **275**(45): 35353-60.

Matunis, E. L., M. J. Matunis et G. Dreyfuss (1993). "Association of individual hnRNP proteins and snRNPs with nascent transcripts." J Cell Biol **121**(2): 219-28.

Mayeda, A., S. H. Munroe, J. F. Caceres et A. R. Krainer (1994). "Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins." Embo J **13**(22): 5483-95.

Mayeda, A. et Y. Ohshima (1988). "Short donor site sequences inserted within the intron of beta-globin pre-mRNA serve for splicing in vitro." Mol Cell Biol **8**(10): 4484-91.

McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, *et al.* (1997). "5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II." Genes Dev **11**(24): 3306-18.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, *et al.* (1997). "The C-terminal domain of RNA polymerase II couples mRNA processing to transcription." Nature **385**(6614): 357-61.

McCracken, S., M. Lambermon et B. J. Blencowe (2002). "SRm160 splicing coactivator promotes transcript 3'-end cleavage." Mol Cell Biol **22**(1): 148-60.

McCracken, S., D. Longman, I. L. Johnstone, J. F. Caceres et B. J. Blencowe (2003). "An evolutionarily conserved role for SRm160 in 3'-end processing that functions independently of exon junction complex formation." J Biol Chem **278**(45): 44153-60.

McCullough, A. J. et S. M. Berget (1997). "G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection." Mol Cell Biol **17**(8): 4562-71.

McGlincy, N. J. et C. W. Smith (2008). "Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?" Trends Biochem Sci **33**(8): 385-93.

Mefford, M. A. et J. P. Staley (2009). "Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps." Rna **15**(7): 1386-97.

Meissner, M., S. Lopato, J. Gotzmann, G. Sauermann et A. Barta (2003). "Proto-oncoprotein TLS/FUS is associated to the nuclear matrix and complexed with splicing factors PTB, SRm160, and SR proteins." Exp Cell Res **283**(2): 184-95.

Meng, Q., S. K. Rayala, A. E. Gururaj, A. H. Talukder, B. W. O'Malley et R. Kumar (2007). "Signaling-dependent and coordinated regulation of transcription, splicing, and translation resides in a single coregulator, PCBP1." Proc Natl Acad Sci U S A **104**(14): 5866-71.

Mermoud, J. E., P. T. Cohen et A. I. Lamond (1994). "Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism." Embo J **13**(23): 5679-88.

Merz, C., H. Urlaub, C. L. Will et R. Luhrmann (2007). "Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment." Rna **13**(1): 116-28.

Michlewski, G., J. R. Sanford et J. F. Caceres (2008). "The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1." Mol Cell **30**(2): 179-89.

Millevoi, S., A. Decorsiere, C. Loulergue, J. Iacovoni, S. Bernat, M. Antoniou, *et al.* (2009). "A physical and functional link between splicing factors promotes pre-mRNA 3' end processing." Nucleic Acids Res.

Millevoi, S., F. Geraghty, B. Idowu, J. L. Tam, M. Antoniou et S. Vagner (2002). "A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing." EMBO Rep **3**(9): 869-74.

Millevoi, S., C. Loulergue, S. Dettwiler, S. Z. Karaa, W. Keller, M. Antoniou, *et al.* (2006). "An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries." Embo J **25**(20): 4854-64.

Milligan, L., C. Torchet, C. Allmang, T. Shipman et D. Tollervey (2005). "A nuclear surveillance pathway for mRNAs with defective polyadenylation." Mol Cell Biol **25**(22): 9996-10004.

Milner, C. M. et R. D. Campbell (1990). "Structure and expression of the three MHC-linked HSP70 genes." Immunogenetics **32**(4): 242-51.

Min, H., C. W. Turck, J. M. Nikolic et D. L. Black (1997). "A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer." Genes Dev **11**(8): 1023-36.

Minvielle-Sebastia, L., P. J. Preker, T. Wiederkehr, Y. Strahm et W. Keller (1997). "The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation." Proc Natl Acad Sci U S A **94**(15): 7897-902.

Misteli, T. (2001). "The concept of self-organization in cellular architecture." J Cell Biol **155**(2): 181-5.

Misteli, T. (2005). "Concepts in nuclear architecture." Bioessays **27**(5): 477-87.

Misteli, T., J. F. Caceres et D. L. Spector (1997). "The dynamics of a pre-mRNA splicing factor in living cells." Nature **387**(6632): 523-7.

Misteli, T. et D. L. Spector (1999). "RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo." Mol Cell **3**(6): 697-705.

Mizutani, A., M. Fukuda, K. Ibata, Y. Shiraishi et K. Mikoshiba (2000). "SYNCRIP, a cytoplasmic counterpart of heterogeneous nuclear ribonucleoprotein R, interacts with ubiquitous synaptotagmin isoforms." J Biol Chem **275**(13): 9823-31.

Modrek, B. et C. Lee (2002). "A genomic view of alternative splicing." Nat Genet **30**(1): 13-9.
Moldon, A., J. Malapeira, N. Gabrielli, M. Gogol, B. Gomez-Escoda, T. Ivanova, *et al.* (2008). "Promoter-driven splicing regulation in fission yeast." Nature **455**(7215): 997-1000.
Moore, C. L. et P. A. Sharp (1985). "Accurate cleavage and polyadenylation of exogenous RNA substrate." Cell **41**(3): 845-55.
Moore, M. J. (2005). "From birth to death: the complex lives of eukaryotic mRNAs." Science **309**(5740): 1514-8.
Moore, M. J. et N. J. Proudfoot (2009). "Pre-mRNA processing reaches back to transcription and ahead to translation." Cell **136**(4): 688-700.
Moore, M. J. et P. A. Sharp (1993). "Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing." Nature **365**(6444): 364-8.
Moreira, A., Y. Takagaki, S. Brackenridge, M. Wollerton, J. L. Manley et N. J. Proudfoot (1998). "The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms." Genes Dev **12**(16): 2522-34.
Moroy, T. et F. Heyd (2007). "The impact of alternative splicing in vivo: mouse models show the way." Rna **13**(8): 1155-71.
Moucadel, V., F. Lopez, T. Ara, P. Benech et D. Gautheret (2007). "Beyond the 3' end: experimental validation of extended transcript isoforms." Nucleic Acids Res **35**(6): 1947-57.
Mourelatos, Z., L. Abel, J. Yong, N. Kataoka et G. Dreyfuss (2001). "SMN interacts with a novel family of hnRNP and spliceosomal proteins." Embo J **20**(19): 5443-52.
Mullen, M. P., C. W. Smith, J. G. Patton et B. Nadal-Ginard (1991). "Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice." Genes Dev **5**(4): 642-55.
Munoz, M. J., M. S. Perez Santangelo, M. P. Paronetto, M. de la Mata, F. Pelisch, S. Boireau, *et al.* (2009). "DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation." Cell **137**(4): 708-20.
Murthy, K. G. et J. L. Manley (1995). "The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation." Genes Dev **9**(21): 2672-83.
Nag, A., K. Narsinh, A. Kazerouninia et H. G. Martinson (2006). "The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity." Rna **12**(8): 1534-44.
Nag, A., K. Narsinh et H. G. Martinson (2007). "The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase." Nat Struct Mol Biol **14**(7): 662-9.
Nakahata, S. et S. Kawamoto (2005). "Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities." Nucleic Acids Res **33**(7): 2078-89.
Nesic, D. et L. E. Maquat (1994). "Upstream introns influence the efficiency of final intron removal and RNA 3'-end formation." Genes Dev. **8**(3): 363-75.
Neubauer, G., A. King, J. Rappsilber, C. Calvio, M. Watson, P. Ajuh, *et al.* (1998). "Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex." Nat Genet **20**(1): 46-50.
Neugebauer, K. M. et M. B. Roth (1997). "Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription." Genes Dev **11**(9): 1148-59.

Ngo, J. C., S. Chakrabarti, J. H. Ding, A. Velazquez-Dones, B. Nolen, B. E. Aubol, *et al.* (2005). "Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2." Mol Cell **20**(1): 77-89.

Ngo, J. C., K. Giang, S. Chakrabarti, C. T. Ma, N. Huynh, J. C. Hagopian, *et al.* (2008). "A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1." Mol Cell **29**(5): 563-76.

Nielsen, J., J. Christiansen, J. Lykke-Andersen, A. H. Johnsen, U. M. Wewer et F. C. Nielsen (1999). "A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development." Mol Cell Biol **19**(2): 1262-70.

Niepmann, M., A. Petersen, K. Meyer et E. Beck (1997). "Functional involvement of polypyrimidine tract-binding protein in translation initiation complexes with the internal ribosome entry site of foot-and-mouth disease virus." J Virol **71**(11): 8330-9.

Nilsen, T. W. (2002). "The spliceosome: no assembly required?" Mol Cell **9**(1): 8-9.

Niwa, M. et S. M. Berget (1991). "Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns." Genes Dev **5**(11): 2086-95.

Nogues, G., S. Kadener, P. Cramer, D. Bentley et A. R. Kornblihtt (2002). "Transcriptional activators differ in their abilities to control alternative splicing." J Biol Chem **277**(45): 43110-4.

Nogues, G., M. J. Munoz et A. R. Kornblihtt (2003). "Influence of polymerase II processivity on alternative splicing depends on splice site strength." J Biol Chem **278**(52): 52166-71.

Norton, P. A. (1994). "Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding." Nucleic Acids Res **22**(19): 3854-60.

Oberstrass, F. C., S. D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, *et al.* (2005). "Structure of PTB bound to RNA: specific binding and implications for splicing regulation." Science **309**(5743): 2054-7.

O'Connell, M. A., A. Gerber et W. Keller (1997). "Purification of human double-stranded RNA-specific editase 1 (hRED1) involved in editing of brain glutamate receptor B pre-mRNA." J Biol Chem **272**(1): 473-8.

Orozco, I. J., S. J. Kim et H. G. Martinson (2002). "The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template." J Biol Chem **277**(45): 42899-911.

O'Sullivan, J. M., S. M. Tan-Wong, A. Morillon, B. Lee, J. Coles, J. Mellor, *et al.* (2004). "Gene loops juxtapose promoters and terminators in yeast." Nat Genet **36**(9): 1014-8.

Padgett, R. A., M. M. Konarska, P. J. Grabowski, S. F. Hardy et P. A. Sharp (1984). "Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors." Science **225**(4665): 898-903.

Paillard, L., V. Legagneux et H. Beverley Osborne (2003). "A functional deadenylation assay identifies human CUG-BP as a deadenylation factor." Biol Cell **95**(2): 107-13.

Paillard, L. et H. B. Osborne (2003). "East of EDEN was a poly(A) tail." Biol Cell **95**(3-4): 211-9.

Palfi, Z., A. Gunzl, M. Cross et A. Bindereif (1991). "Affinity purification of Trypanosoma brucei small nuclear ribonucleoproteins reveals common and specific protein components." Proc Natl Acad Sci U S A **88**(20): 9097-101.

Pan, Q., A. L. Saltzman, Y. K. Kim, C. Misquitta, O. Shai, L. E. Maquat, *et al.* (2006). "Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression." Genes Dev **20**(2): 153-8.

Pan, Q., O. Shai, C. Misquitta, W. Zhang, A. L. Saltzman, N. Mohammad, *et al.* (2004). "Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform." Mol Cell **16**(6): 929-41.

Pascual, M., M. Vicente, L. Monferrer et R. Artero (2006). "The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing." Differentiation **74**(2-3): 65-80.

Patel, G. P., S. Ma et J. Bag (2005). "The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex." Nucleic Acids Res **33**(22): 7074-89.

Patel, N. A., S. Kaneko, H. S. Apostolatos, S. S. Bae, J. E. Watson, K. Davidowitz, *et al.* (2005). "Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase CbetaII alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40." J Biol Chem **280**(14): 14302-9.

Patel, S. B. et M. Bellini (2008). "The assembly of a spliceosomal small nuclear ribonucleoprotein particle." Nucleic Acids Res.

Peled-Zehavi, H., J. A. Berglund, M. Rosbash et A. D. Frankel (2001). "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): 5232-41.

Perez Canadillas, J. M. et G. Varani (2003). "Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein." Embo J **22**(11): 2821-30.

Perez, I., C. H. Lin, J. G. McAfee et J. G. Patton (1997). "Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo." Rna **3**(7): 764-78.

Perez, I., J. G. McAfee et J. G. Patton (1997). "Multiple RRMs contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein." Biochemistry **36**(39): 11881-90.

Peterson, M. L., G. L. Bingham et C. Cowan (2006). "Multiple features contribute to the use of the immunoglobulin M secretion-specific poly(A) signal but are not required for developmental regulation." Mol Cell Biol **26**(18): 6762-71.

Phatnani, H. P. et A. L. Greenleaf (2006). "Phosphorylation and functions of the RNA polymerase II CTD." Genes Dev **20**(21): 2922-36.

Phillips, C., S. Jung et S. I. Gunderson (2001). "Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA." Embo J **20**(22): 6443-52.

Phillips, C., N. Pachikara et S. I. Gunderson (2004). "U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions." Mol Cell Biol **24**(14): 6162-71.

Ponthier, J. L., C. Schluepen, W. Chen, R. A. Lersch, S. L. Gee, V. C. Hou, *et al.* (2006). "Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16." J Biol Chem **281**(18): 12468-74.

Pozzoli, U. et M. Sironi (2005). "Silencers regulate both constitutive and alternative splicing events in mammals." Cell Mol Life Sci **62**(14): 1579-604.

Prasad, J., K. Colwill, T. Pawson et J. L. Manley (1999). "The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing." Mol Cell Biol **19**(10): 6991-7000.

Proudfoot, N. J. et G. G. Brownlee (1976). "3' non-coding region sequences in eukaryotic messenger RNA." Nature **263**(5574): 211-4.

Proudfoot, N. J., A. Furger et M. J. Dye (2002). "Integrating mRNA processing with transcription." Cell **108**(4): 501-12.

Qiu, J., F. Cheng et D. Pintel (2007). "Distance-dependent processing of adeno-associated virus type 5 RNA is controlled by 5' exon definition." J Virol **81**(15): 7974-84.

Raju, V. S. et S. T. Jacob (1988). "Association of poly(A) polymerase with U1 RNA." J Biol Chem **263**(23): 11067-70.
Rappsilber, J., U. Ryder, A. I. Lamond et M. Mann (2002). "Large-scale proteomic analysis of the human spliceosome." Genome Res **12**(8): 1231-45.
Reed, R. (1990). "Protein composition of mammalian spliceosomes assembled in vitro." Proc Natl Acad Sci U S A **87**(20): 8031-5.
Reed, R. (2000). "Mechanisms of fidelity in pre-mRNA splicing." Curr Opin Cell Biol **12**(3): 340-345.
Reyes-Reyes, M. et M. Hampsey (2007). "Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation." Mol Cell Biol **27**(3): 926-36.
Rhode, B. M., K. Hartmuth, H. Urlaub et R. Luhrmann (2003). "Analysis of site-specific protein-RNA cross-links in isolated RNP complexes, combining affinity selection and mass spectrometry." Rna **9**(12): 1542-51.
Rideau, A. P., C. Gooding, P. J. Simpson, T. P. Monie, M. Lorenz, S. Huttelmaier, *et al.* (2006). "A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain." Nat Struct Mol Biol **13**(9): 839-48.
Rino, J. et M. Carmo-Fonseca (2009). "The spliceosome: a self-organized macromolecular machine in the nucleus?" Trends Cell Biol.
Rino, J., T. Carvalho, J. Braga, J. M. Desterro, R. Luhrmann et M. Carmo-Fonseca (2007). "A stochastic view of spliceosome assembly and recycling in the nucleus." PLoS Comput Biol **3**(10): 2019-31.
Rio, D. C., F. A. Laski et G. M. Rubin (1986). "Identification and immunochemical analysis of biologically active Drosophila P element transposase." Cell **44**(1): 21-32.
Ritchie, D. B., M. J. Schellenberg, E. M. Gesner, S. A. Raithatha, D. T. Stuart et A. M. Macmillan (2008). "Structural elucidation of a PRP8 core domain from the heart of the spliceosome." Nat Struct Mol Biol.
Robberson, B. L., G. J. Cote et S. M. Berget (1990). "Exon definition may facilitate splice site selection in RNAs with multiple exons." Mol Cell Biol **10**(1): 84-94.
Roberts, G. C., C. Gooding, H. Y. Mak, N. J. Proudfoot et C. W. Smith (1998). "Co-transcriptional commitment to alternative splice site selection." Nucleic Acids Res **26**(24): 5568-72.
Robida, M. D. et R. Singh (2003). "Drosophila polypyrimidine-tract binding protein (PTB) functions specifically in the male germline." Embo J **22**(12): 2924-33.
Robinson, F. et C. W. Smith (2005). "A splicing repressor domain in polypyrimidine tract binding protein." J Biol Chem.
Robson-Dixon, N. D. et M. A. Garcia-Blanco (2004). "MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb." J Biol Chem **279**(28): 29075-84.
Romfo, C. M., C. J. Alvarez, W. J. van Heeckeren, C. J. Webb et J. A. Wise (2000). "Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe." Mol Cell Biol **20**(21): 7955-70.
Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlen et P. Nyren (1996). "Real-time DNA sequencing using detection of pyrophosphate release." Anal Biochem **242**(1): 84-9.
Rooke, N., V. Markovtsov, E. Cagavi et D. L. Black (2003). "Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1." Mol Cell Biol **23**(6): 1874-84.
Ropers, D., L. Ayadi, R. Gattoni, S. Jacquenet, L. Damier, C. Branlant, *et al.* (2004). "Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA." J Biol Chem **279**(29): 29963-73.

Roscigno, R. F. et M. A. Garcia-Blanco (1995). "SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome." Rna **1**(7): 692-706.

Rossoll, W., A. K. Kroning, U. M. Ohndorf, C. Steegborn, S. Jablonka et M. Sendtner (2002). "Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons?" Hum Mol Genet **11**(1): 93-105.

Roth, M. B., C. Murphy et J. G. Gall (1990). "A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle." J Cell Biol **111**(6 Pt 1): 2217-23.

Rothe, F., C. Gueydan, E. Bellefroid, G. Huez et V. Kruys (2006). "Identification of FUSE-binding proteins as interacting partners of TIA proteins." Biochem Biophys Res Commun **343**(1): 57-68.

Rowen, L., J. Young, B. Birditt, A. Kaur, A. Madan, D. L. Philipps*, et al.* (2002). "Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity." Genomics **79**(4): 587-97.

Ruegsegger, U., K. Beyer et W. Keller (1996). "Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors." J Biol Chem **271**(11): 6107-13.

Ruegsegger, U., D. Blank et W. Keller (1998). "Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits." Mol Cell **1**(2): 243-53.

Ryan, K., O. Calvo et J. L. Manley (2004). "Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease." Rna **10**(4): 565-73.

Sanchez-Alvarez, M., A. C. Goldstrohm, M. A. Garcia-Blanco et C. Sune (2006). "Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions." Mol Cell Biol **26**(13): 4998-5014.

Sandberg, R., J. R. Neilson, A. Sarma, P. A. Sharp et C. B. Burge (2008). "Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites." Science **320**(5883): 1643-7.

Sanford, J. R., P. Coutinho, J. A. Hackett, X. Wang, W. Ranahan et J. F. Caceres (2008). "Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF." PLoS One **3**(10): e3369.

Sanford, J. R., N. K. Gray, K. Beckmann et J. F. Caceres (2004). "A novel role for shuttling SR proteins in mRNA translation." Genes Dev **18**(7): 755-68.

Sapra, A. K., M. L. Anko, I. Grishina, M. Lorenz, M. Pabis, I. Poser*, et al.* (2009). "SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo." Mol Cell **34**(2): 179-90.

Sato, H., N. Hosoda et L. E. Maquat (2008). "Efficiency of the pioneer round of translation affects the cellular site of nonsense-mediated mRNA decay." Mol Cell **29**(2): 255-62.

Sauliere, J., A. Sureau, A. Expert-Bezancon et J. Marie (2006). "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): 8755-69.

Sawicka, K., M. Bushell, K. A. Spriggs et A. E. Willis (2008). "Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein." Biochem Soc Trans **36**(Pt 4): 641-7.

Schaub, M. C., S. R. Lopez et M. Caputi (2007). "Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes." J Biol Chem **282**(18): 13617-26.

Schlatter, H., T. Langer, S. Rosmus, M. L. Onneken et H. Fasold (2002). "A novel function for the 90 kDa heat-shock protein (Hsp90): facilitating nuclear export of 60 S ribosomal subunits." Biochem J **362**(Pt 3): 675-84.

Seraphin, B. et M. Rosbash (1989). "Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing." Cell **59**(2): 349-58.

Sharma, S., A. M. Falick et D. L. Black (2005). "Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex." Mol Cell **19**(4): 485-96.

Sharma, S., L. A. Kohlstaedt, A. Damianov, D. C. Rio et D. L. Black (2008). "Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome." Nat Struct Mol Biol **15**(2): 183-91.

Sharp, P. A. (1994). "Split genes and RNA splicing." Cell **77**(6): 805-15.

Sheets, M. D., S. C. Ogg et M. P. Wickens (1990). "Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro." Nucleic Acids Res **18**(19): 5799-805.

Shell, S. A., C. Hesse, S. M. Morris, Jr. et C. Milcarek (2005). "Elevated Levels of the 64-kDa Cleavage Stimulatory Factor (CstF-64) in Lipopolysaccharide-stimulated Macrophages Influence Gene Expression and Induce Alternative Poly(A) Site Selection." J Biol Chem **280**(48): 39950-61.

Shi, L., G. Zhao, D. Qiu, W. R. Godfrey, H. Vogel, T. A. Rando, *et al.* (2005). "NF90 regulates cell cycle exit and terminal myogenic differentiation by direct binding to the 3'-untranslated region of MyoD and p21WAF1/CIP1 mRNAs." J Biol Chem **280**(19): 18981-9.

Shi, Y. et J. L. Manley (2007). "A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock." Mol Cell **28**(1): 79-90.

Shin, C., Y. Feng et J. L. Manley (2004). "Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock." Nature **427**(6974): 553-8.

Shin, C. et J. L. Manley (2004). "Cell signalling and the control of pre-mRNA splicing." Nat Rev Mol Cell Biol **5**(9): 727-38.

Shyu, A. B., M. F. Wilkinson et A. van Hoof (2008). "Messenger RNA regulation: to translate or to degrade." Embo J **27**(3): 471-81.

Simpson, P. J., T. P. Monie, A. Szendroi, N. Davydova, J. K. Tyzack, M. R. Conte, *et al.* (2004). "Structure and RNA interactions of the N-terminal RRM domains of PTB." Structure (Camb) **12**(9): 1631-43.

Singh, B. N. et M. Hampsey (2007). "A transcription-independent role for TFIIB in gene looping." Mol Cell **27**(5): 806-16.

Singh, R. et J. Valcarcel (2005). "Building specificity with nonspecific RNA-binding proteins." Nat Struct Mol Biol **12**(8): 645-53.

Smith, C. W. et J. Valcarcel (2000). "Alternative pre-mRNA splicing: the logic of combinatorial control." Trends Biochem Sci **25**(8): 381-8.

Smith, H. O., K. Tabiti, G. Schaffner, D. Soldati, U. Albrecht et M. L. Birnstiel (1991). "Two-step affinity purification of U7 small nuclear ribonucleoprotein particles using complementary biotinylated 2'-O-methyl oligoribonucleotides." Proc Natl Acad Sci U S A **88**(21): 9784-8.

Soret, J., N. Bakkour, S. Maire, S. Durand, L. Zekri, M. Gabut, *et al.* (2005). "Selective modification of alternative splicing by indole derivatives that target serine-arginine-rich protein splicing factors." Proc Natl Acad Sci U S A **102**(24): 8764-9.

Southby, J., C. Gooding et C. W. Smith (1999). "Polypyrimidine tract binding protein functions as a repressor To regulate alternative splicing of alpha-actinin mutally exclusive exons [In Process Citation]." Mol Cell Biol **19**(4): 2699-711.

Spadaccini, R., U. Reidt, O. Dybkov, C. Will, R. Frank, G. Stier*, et al.* (2006). "Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing." Rna **12**(3): 410-25.

Spector, D. L. (1993). "Nuclear organization of pre-mRNA processing." Curr Opin Cell Biol **5**(3): 442-7.

Spellman, R., M. Llorian et C. W. Smith (2007). "Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1." Mol Cell **27**(3): 420-34.

Spellman, R. et C. W. Smith (2006). "Novel modes of splicing repression by PTB." Trends Biochem Sci **31**(2): 73-6.

Sperling, J., M. Azubel et R. Sperling (2008). "Structure and function of the Pre-mRNA splicing machine." Structure **16**(11): 1605-15.

Srebrow, A. et A. R. Kornblihtt (2006). "The connection between splicing and cancer." J Cell Sci **119**(Pt 13): 2635-2641.

Srisawat, C. et D. R. Engelke (2001). "Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins." Rna **7**(4): 632-41.

Srisawat, C. et D. R. Engelke (2002). "RNA affinity tags for purification of RNAs and ribonucleoprotein complexes." Methods **26**(2): 156-61.

Srisawat, C., I. J. Goldstein et D. R. Engelke (2001). "Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures." Nucleic Acids Res **29**(2): E4.

Staknis, D. et R. Reed (1994). "SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex." Mol Cell Biol **14**(11): 7670-82.

Staley, J. P. et C. Guthrie (1998). "Mechanical devices of the spliceosome: motors, clocks, springs, and things." Cell **92**(3): 315-26.

Stamm, S. (2002). "Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome." Hum Mol Genet **11**(20): 2409-16.

Stamm, S. (2008). "Regulation of alternative splicing by reversible protein phosphorylation." J Biol Chem **283**(3): 1223-7.

Stamm, S., S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber*, et al.* (2005). "Function of alternative splicing." Gene **344**: 1-20.

Steitz, J. A., G. Dreyfuss, A. R. Krainer, A. I. Lamond, A. G. Matera et R. A. Padgett (2008). "Where in the cell is the minor spliceosome?" Proc Natl Acad Sci U S A **105**(25): 8485-6.

Sterner, D. A., T. Carlo et S. M. Berget (1996). "Architectural limits on split genes." Proc Natl Acad Sci U S A **93**(26): 15081-5.

Stevens, S. W., D. E. Ryan, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee*, et al.* (2002). "Composition and functional characterization of the yeast spliceosomal penta-snRNP." Mol Cell **9**(1): 31-44.

Sumanasekera, C., D. S. Watt et S. Stamm (2008). "Substances that can change alternative splice-site selection." Biochem Soc Trans **36**(Pt 3): 483-90.

Sureau, A., R. Gattoni, Y. Dooghe, J. Stevenin et J. Soret (2001). "SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs." Embo J **20**(7): 1785-96.

Swoboda, K. J., C. B. Scott, S. P. Reyna, T. W. Prior, B. LaSalle, S. L. Sorenson, *et al.* (2009). "Phase II open label study of valproic acid in spinal muscular atrophy." PLoS One **4**(5): e5268.

Szymczyna, B. R., J. Bowman, S. McCracken, A. Pineda-Lucena, Y. Lu, B. Cox, *et al.* (2003). "Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing." Genes Dev **17**(4): 461-75.

Takagaki, Y. et J. L. Manley (1994). "A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein." Nature **372**(6505): 471-4.

Takagaki, Y. et J. L. Manley (1998). "Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation." Mol Cell **2**(6): 761-71.

Takagaki, Y. et J. L. Manley (2000). "Complex protein interactions within the human polyadenylation machinery identify a novel component." Mol Cell Biol **20**(5): 1515-25.

Takagaki, Y., R. L. Seipelt, M. L. Peterson et J. L. Manley (1996). "The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation." Cell **87**(5): 941-52.

Talerico, M. et S. M. Berget (1994). "Intron definition in splicing of small Drosophila introns." Mol Cell Biol **14**(5): 3434-45.

Tanackovic, G. et A. Kramer (2005). "Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo." Mol Biol Cell **16**(3): 1366-77.

Tazi, J., N. Bakkour et S. Stamm (2009). "Alternative splicing and disease." Biochim Biophys Acta **1792**(1): 14-26.

Tazi, J., S. Durand et P. Jeanteur (2005). "The spliceosome: a novel multi-faceted target for therapy." Trends Biochem Sci **30**(8): 469-78.

Tenenbaum, S. A. et J. Aguirre-Ghiso (2005). "Dephosphorylation shows SR proteins the way out." Mol Cell **20**(4): 499-501.

Teplova, M. et D. J. Patel (2008). "Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1." Nat Struct Mol Biol **15**(12): 1343-51.

Thanaraj, T. A. et S. Stamm (2003). "Prediction and statistical analysis of alternatively spliced exons." Prog Mol Subcell Biol **31**: 1-31.

Thickman, K. R., M. C. Swenson, J. M. Kabogo, Z. Gryczynski et C. L. Kielkopf (2006). "Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors." J Mol Biol **356**(3): 664-83.

Thisted, T., D. L. Lyakhov et S. A. Liebhaber (2001). "Optimized RNA targets of two closely related triple KH domain proteins, heterogeneous nuclear ribonucleoprotein K and alphaCP-2KL, suggest Distinct modes of RNA recognition." J Biol Chem **276**(20): 17484-96.

Tian, B., J. Hu, H. Zhang et C. S. Lutz (2005). "A large-scale analysis of mRNA polyadenylation of human and mouse genes." Nucleic Acids Res **33**(1): 201-12.

Tian, B., Z. Pan et J. Y. Lee (2007). "Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing." Genome Res **17**(2): 156-65.

Tran, Q., T. P. Coleman et J. R. Roesser (2003). "Human transformer 2beta and SRp55 interact with a calcitonin-specific splice enhancer." Biochim Biophys Acta **1625**(2): 141-52.

Tran, Q. et J. R. Roesser (2003). "SRp55 is a regulator of calcitonin/CGRP alternative RNA splicing." Biochemistry **42**(4): 951-7.

Tsukahara, F., T. Yoshioka et T. Muraki (2000). "Molecular and functional characterization of HSC54, a novel variant of human heat-shock cognate protein 70." Mol Pharmacol **58**(6): 1257-63.

Tsurushita, N., N. M. Avdalovic et L. J. Korn (1987). "Regulation of differential processing of mouse immunoglobulin mu heavy- chain mRNA." Nucleic Acids Res **15**(11): 4603-15.

Turner, I. A., C. M. Norman, M. J. Churcher et A. J. Newman (2006). "Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome." Rna **12**(3): 375-86.

Ule, J., K. Jensen, A. Mele et R. B. Darnell (2005). "CLIP: a method for identifying protein-RNA interaction sites in living cells." Methods **37**(4): 376-86.

Ule, J., K. B. Jensen, M. Ruggiu, A. Mele, A. Ule et R. B. Darnell (2003). "CLIP identifies Nova-regulated RNA networks in the brain." Science **302**(5648): 1212-5.

Ule, J., A. Ule, J. Spencer, A. Williams, J. S. Hu, M. Cline*, et al.* (2005). "Nova regulates brain-specific splicing to shape the synapse." Nat Genet **37**(8): 844-52.

Underwood, J. G., P. L. Boutz, J. D. Dougherty, P. Stoilov et D. L. Black (2005). "Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals." Mol Cell Biol **25**(22): 10005-16.

Vagner, S., U. Ruegsegger, S. I. Gunderson, W. Keller et I. W. Mattaj (2000). "Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP." Rna **6**(2): 178-88.

Vagner, S., C. Vagner et I. W. Mattaj (2000). "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.

Valadkhan, S., A. Mohammadi, C. Wachtel et J. L. Manley (2007). "Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing." Rna **13**(12): 2300-11.

van Gelder, C. W., S. I. Gunderson, E. J. Jansen, W. C. Boelens, M. Polycarpou-Schwarz, I. W. Mattaj*, et al.* (1993). "A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation." Embo J **12**(13): 5191-200.

Varani, L., M. Hasegawa, M. G. Spillantini, M. J. Smith, J. R. Murrell, B. Ghetti*, et al.* (1999). "Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17." Proc Natl Acad Sci U S A **96**(14): 8229-34.

Vazquez-Pianzola, P., H. Urlaub et R. Rivera-Pomar (2005). "Proteomic analysis of reaper 5' untranslated region-interacting factors isolated by tobramycin affinity-selection reveals a role for La antigen in reaper mRNA translation." Proteomics **5**(6): 1645-55.

Venables, J. P., C. S. Koh, U. Froehlich, E. Lapointe, S. Couture, L. Inkel*, et al.* (2008). "Multiple and specific mRNA processing targets for the major human hnRNP proteins." Mol Cell Biol **28**(19): 6033-43.

Venkataraman, K., K. M. Brown et G. M. Gilmartin (2005). "Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition." Genes Dev **19**(11): 1315-27.

Vikesaa, J., T. V. Hansen, L. Jonson, R. Borup, U. M. Wewer, J. Christiansen*, et al.* (2006). "RNA-binding IMPs promote cell adhesion and invadopodia formation." Embo J **25**(7): 1456-68.

Villa, T., J. A. Pleiss et C. Guthrie (2002). "Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core?" Cell **109**(2): 149-52.

Vlasova, I. A., N. M. Tahoe, D. Fan, O. Larsson, B. Rattenbacher, J. R. Sternjohn, *et al.* (2008). "Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1." Mol Cell **29**(2): 263-70.

Vogel, J., W. R. Hess et T. Borner (1997). "Precise branch point mapping and quantification of splicing intermediates." Nucleic Acids Res **25**(10): 2030-1.

Wagner, E. J., A. P. Baraniak, O. M. Sessions, D. Mauger, E. Moskowitz et M. A. Garcia-Blanco (2005). "Characterization of the intronic splicing silencers flanking FGFR2 exon IIIb." J Biol Chem **280**(14): 14017-27.

Wagner, E. J. et M. A. Garcia-Blanco (2001). "Polypyrimidine tract binding protein antagonizes exon definition." Mol Cell Biol **21**(10): 3281-8.

Wagner, E. J. et M. A. Garcia-Blanco (2002). "RNAi-mediated PTB depletion leads to enhanced exon definition." Mol Cell **10**(4): 943-9.

Wahl, M. C., C. L. Will et R. Luhrmann (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell **136**(4): 701-18.

Walker, S. C., F. H. Scott, C. Srisawat et D. R. Engelke (2008). "RNA affinity tags for the rapid purification and investigation of RNAs and RNA-protein complexes." Methods Mol Biol **488**: 23-40.

Wallace, S. T. et R. Schroeder (1998). "In vitro selection and characterization of streptomycin-binding RNAs: recognition discrimination between antibiotics." Rna **4**(1): 112-23.

Wang, G. S. et T. A. Cooper (2007). "Splicing in disease: disruption of the splicing code and the decoding machinery." Nat Rev Genet **8**(10): 749-61.

Wang, H., R. Fang, J. Y. Cho, T. A. Libermann et P. Oettgen (2004). "Positive and negative modulation of the transcriptional activity of the ETS factor ESE-1 through interaction with p300, CREB-binding protein, and Ku 70/86." J Biol Chem **279**(24): 25241-50.

Wang, H. Y., X. Xu, J. H. Ding, J. R. Bermingham, Jr. et X. D. Fu (2001). "SC35 plays a role in T cell development and alternative splicing of CD45." Mol Cell **7**(2): 331-42.

Wang, J., Y. Takagaki et J. L. Manley (1996). "Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability." Genes Dev **10**(20): 2588-99.

Wang, X., S. Bruderer, Z. Rafi, J. Xue, P. J. Milburn, A. Kramer, *et al.* (1999). "Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly." Embo J **18**(16): 4549-59.

Wang, Y. et R. R. Rando (1995). "Specific binding of aminoglycoside antibiotics to RNA." Chem Biol **2**(5): 281-90.

Wang, Z. et C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." Rna **14**(5): 802-13.

Wang, Z., M. E. Rolish, G. Yeo, V. Tung, M. Mawson et C. B. Burge (2004). "Systematic identification and analysis of exonic splicing silencers." Cell **119**(6): 831-45.

Warf, M. B. et J. A. Berglund (2007). "MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T." Rna **13**(12): 2238-51.

Warzecha, C. C., T. K. Sato, B. Nabet, J. B. Hogenesch et R. P. Carstens (2009). "ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing." Mol Cell **33**(5): 591-601.

Wassarman, K. M. et J. A. Steitz (1993). "Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP?" Genes Dev **7**(4): 647-59.

Wen, X., S. Tannukit et M. L. Paine (2008). "TFIP11 Interacts with mDEAH9, an RNA Helicase Involved in Spliceosome Disassembly." Int J Mol Sci **9**(11): 2105-13.

West, S., N. Gromak et N. J. Proudfoot (2004). "Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites." Nature **432**(7016): 522-5.

West, S. et N. J. Proudfoot (2009). "Transcriptional termination enhances protein expression in human cells." Mol Cell **33**(3): 354-64.

West, S., N. J. Proudfoot et M. J. Dye (2008). "Molecular dissection of mammalian RNA polymerase II transcriptional termination." Mol Cell **29**(5): 600-10.

Wilchek, M. et R. Lamed (1974). "Immobilized nucleotides for affinity chromatography." Methods Enzymol **34**: 475-9.

Will, C. L., S. E. Behrens et R. Luhrmann (1993). "Protein composition of mammalian spliceosomal snRNPs." Mol Biol Rep **18**(2): 121-6.

Will, C. L. et R. Luhrmann (2001). "Spliceosomal UsnRNP biogenesis, structure and function." Curr Opin Cell Biol **13**(3): 290-301.

Willis, D. M., A. P. Loewy, N. Charlton-Kachigian, J. S. Shao, D. M. Ornitz et D. A. Towler (2002). "Regulation of osteocalcin gene expression by a novel Ku antigen transcription factor complex." J Biol Chem **277**(40): 37280-91.

Wollerton, M. C., C. Gooding, F. Robinson, E. C. Brown, R. J. Jackson et C. W. Smith (2001). "Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB)." Rna **7**(6): 819-32.

Wollerton, M. C., C. Gooding, E. J. Wagner, M. A. Garcia-Blanco et C. W. Smith (2004). "Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay." Mol Cell **13**(1): 91-100.

Wu, J. Y. et T. Maniatis (1993). "Specific interactions between proteins implicated in splice site selection and regulated alternative splicing." Cell **75**(6): 1061-70.

Wu, S., C. M. Romfo, T. W. Nilsen et M. R. Green (1999). "Functional recognition of the 3' splice site AG by the splicing factor U2AF35." Nature **402**(6763): 832-5.

Xiao, S. H. et J. L. Manley (1997). "Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing." Genes Dev **11**(3): 334-44.

Xie, J., J. A. Lee, T. L. Kress, K. L. Mowry et D. L. Black (2003). "Protein kinase A phosphorylation modulates transport of the polypyrimidine tract-binding protein." Proc Natl Acad Sci U S A **100**(15): 8776-81.

Xu, X., D. Yang, J. H. Ding, W. Wang, P. H. Chu, N. D. Dalton*, et al.* (2005). "ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle." Cell **120**(1): 59-72.

Yamamoto, H., K. Tsukahara, Y. Kanaoka, S. Jinno et H. Okayama (1999). "Isolation of a mammalian homologue of a fission yeast differentiation regulator." Mol Cell Biol **19**(5): 3829-41.

Yan, J. et T. G. Marr (2005). "Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat." Genome Res **15**(3): 369-75.

Yang, Y. Y., G. L. Yin et R. B. Darnell (1998). "The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia." Proc Natl Acad Sci U S A **95**(22): 13254-9.

Yu, Y., P. A. Maroney, J. A. Denker, X. H. Zhang, O. Dybkov, R. Luhrmann*, et al.* (2008). "Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition." Cell **135**(7): 1224-36.

Yuan, X., N. Davydova, M. R. Conte, S. Curry et S. Matthews (2002). "Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein." Nucleic Acids Res **30**(2): 456-62.

Yuryev, A., M. Patturajan, Y. Litingtung, R. V. Joshi, C. Gentile, M. Gebara*, et al.* (1996). "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): 6975-80.

Zamore, P. D. et M. R. Green (1989). "Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor." Proc Natl Acad Sci U S A **86**(23): 9243-7.

Zamore, P. D., J. G. Patton et M. R. Green (1992). "Cloning and domain structure of the mammalian splicing factor U2AF." Nature **355**(6361): 609-14.

Zeng, C., E. Kim, S. L. Warren et S. M. Berget (1997). "Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity." Embo J **16**(6): 1401-12.

Zhang, F., P. R. Romano, T. Nagamura-Inoue, B. Tian, T. E. Dever, M. B. Mathews*, et al.* (2001). "Binding of double-stranded RNA to protein kinase PKR is required for dimerization and promotes critical autophosphorylation events in the activation loop." J Biol Chem **276**(27): 24946-58.

Zhang, H., J. Hu, M. Recce et B. Tian (2005). "PolyA_DB: a database for mammalian mRNA polyadenylation." Nucleic Acids Res **33**(Database issue): D116-20.

Zhang, L., W. Liu et P. J. Grabowski (1999). "Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB." Rna **5**(1): 117-30.

Zhang, M. Q. et T. G. Marr (1994). "Fission yeast gene structure and recognition." Nucleic Acids Res **22**(9): 1750-9.

Zhang, Z. et A. R. Krainer (2004). "Involvement of SR proteins in mRNA surveillance." Mol Cell **16**(4): 597-607.

Zhao, J., L. Hyman et C. Moore (1999). "Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis." Microbiol Mol Biol Rev **63**(2): 405-45.

Zhao, R., Y. Kakihara, A. Gribun, J. Huen, G. Yang, M. Khanna*, et al.* (2008). "Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation." J Cell Biol **180**(3): 563-78.

Zhong, X. Y., J. H. Ding, J. A. Adams, G. Ghosh et X. D. Fu (2009). "Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones." Genes Dev **23**(4): 482-95.

Zhou, H. L., A. P. Baraniak et H. Lou (2007). "Role for Fox-1/Fox-2 in mediating the neuronal pathway of calcitonin/calcitonin gene-related peptide alternative RNA processing." Mol Cell Biol **27**(3): 830-41.

Zhou, H. L. et H. Lou (2008). "Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins." Mol Cell Biol **28**(17): 5507-16.

Zhou, Z., L. J. Licklider, S. P. Gygi et R. Reed (2002). "Comprehensive proteomic analysis of the human spliceosome." Nature **419**(6903): 182-5.

Zhu, H., R. A. Hasman, V. A. Barron, G. Luo et H. Lou (2006). "A nuclear function of Hu proteins as neuron-specific alternative RNA processing regulators." Mol Biol Cell **17**(12): 5105-14.

Zhu, H., R. A. Hasman, K. M. Young, N. L. Kedersha et H. Lou (2003). "U1 snRNP-dependent function of TIAR in the regulation of alternative RNA processing of the human calcitonin/CGRP pre-mRNA." Mol Cell Biol **23**(17): 5959-71.

Zhu, J., A. Mayeda et A. R. Krainer (2001). "Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins." Mol Cell **8**(6): 1351-61.

Zorio, D. A. et D. L. Bentley (2004). "The link between mRNA processing and transcription: communication works both ways." Exp Cell Res **296**(1): 91-7.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-15.

Zuo, P. et T. Maniatis (1996). "The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing." Genes Dev **10**(11): 1356-68.