AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Nancy-Université
Université Henri Poincaré

cnrs
dépasser les frontières

**Ecole Doctorale BioSE (Biologie-Santé-Environnement)**

# Thèse

**Présentée et soutenue publiquement pour l'obtention du titre de**

## DOCTEUR DE L'UNIVERSITE HENRI POINCARE

## Mention : « Sciences de la Vie et de la Santé »

par **Aileen BAR**

# ETUDE STRUCTURALE ET FONCTIONNELLE DE L'ELEMENT NRS REGULATEUR NEGATIF DE L'EPISSAGE DU VIRUS DU SARCOME DE ROUS

**le 17 novembre 2011**

**Membres du jury :**

| | | |
|---|---|---|
| Rapporteurs : | Mme Joëlle MARIE | Directeur de recherche INSERM, Gif-sur-Yvette |
| | Mr Roland MARQUET | Directeur de recherche INSERM, Strasbourg |
| Examinateurs : | Mr James STEVENIN | Directeur de recherche INSERM, Illkirch |
| | Mr Alain MILLER | Professeur, PDG de la société CILBiotech, Mons |
| | Mr Athanase VISVIKIS | Professeur UHP, Nancy |
| | Mme Christiane BRANLANT | Directeur de recherche CNRS, Nancy<br>Directeur de thèse |

---

Laboratoire ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale
UMR 7214 CNRS-UHP Nancy I
Bâtiment Biopôle, Faculté de médecine de Nancy, Vandœuvre les Nancy, France

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Nancy-Université
Université Henri Poincaré

cnrs
dépasser les frontières

**Ecole Doctorale BioSE (Biologie-Santé-Environnement)**

# Thèse

**Présentée et soutenue publiquement pour l'obtention du titre de**

## DOCTEUR DE L'UNIVERSITE HENRI POINCARE

## Mention : « Sciences de la Vie et de la Santé »

par **Aileen BAR**

# ETUDE STRUCTURALE ET FONCTIONNELLE DE L'ELEMENT NRS REGULATEUR NEGATIF DE L'EPISSAGE DU VIRUS DU SARCOME DE ROUS

**le 17 novembre 2011**

**Membres du jury :**

| | | |
|---|---|---|
| Rapporteurs : | Mme Joëlle MARIE | Directeur de recherche INSERM, Gif-sur-Yvette |
| | Mr Roland MARQUET | Directeur de recherche INSERM, Strasbourg |
| Examinateurs : | Mr James STEVENIN | Directeur de recherche INSERM, Illkirch |
| | Mr Alain MILLER | Professeur, PDG de la société CILBiotech, Mons |
| | Mr Athanase VISVIKIS | Professeur UHP, Nancy |
| | Mme Christiane BRANLANT | Directeur de recherche CNRS, Nancy<br>Directeur de thèse |

---

Laboratoire ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale
UMR 7214 CNRS-UHP Nancy I
Bâtiment Biopôle, Faculté de médecine de Nancy, Vandœuvre les Nancy, France

A mon père,
bien au-delà des mots

# Remerciements

Je tiens tout d'abord à exprimer toute ma reconnaissance à ma directrice de thèse, Mme Christiane Branlant. Ses conseils, sa rigueur et son enthousiasme m'ont été d'une aide précieuse tout au long de mon travail de thèse. Je la remercie vivement pour la formation scientifique qu'elle m'a apportée et les nombreux échanges scientifiques que nous avons eus. J'ai été également très sensible à la grande compréhension dont elle a fait preuve durant les multiples épreuves que j'ai traversées.

Je remercie vivement Madame Jöelle Marie et Monsieur Roland Marquet qui me font l'honneur d'avoir accepté d'être les rapporteurs de ce manuscrit. J'exprime toute ma gratitude à Messieurs James Stévenin et Alain Miller d'avoir accepté de juger ce travail.

J'adresse ma profonde reconnaissance au Pr Rheinard Lührmann pour m'avoir accueillie dans son laboratoire et à Mr Klaus Hartmuth et Jorgen Decker pour leur encadrement lors de ce séjour.

J'exprime toute ma gratitude à Mr James Stévenin et Mme Natacha Dreumont pour leur collaboration fructueuse pour les expériences de SiRNA et les discussions scientifiques que nous avons eues. Je les remercie également de m'avoir fourni la protéine 9G8 utilisée au cours de ce travail.

Je tiens à remercier l'équipe d'Alain Van Dorsselaer, en particulier Mme Sarah Sanglier et Mlle Laëtitia Fouillen pour avoir réalisé les analyses par spectrométrie de masse.

Je tiens à remercier l'ensemble des équipes pédagogiques avec lesquelles j'ai travaillé durant mon monitorat à l'IUT de Thionville et mes ATER à l'IUT du Montet, en particulier Mr Pascal Picquot, Mr Michael Schneider, Mr Pascal Chappe et Mme Patricia Guinot-Thomas.

J'exprime ma profonde reconnaissance à Nathalie Robas qui m'a mis le pied à l'étrier de la recherche lors de mon premier stage et qui m'a encadrée durant mon M2R et au début de ma thèse. Je tiens à remercier également vivement Athanase Visvikis d'avoir poursuivi cet encadrement et de s'être investi pleinement dans cette thématique.

Je remercie également Annie Mougin pour les travaux antérieurs réalisés sur la thématique et pour me les avoir transmis dans les conditions les plus optimales. Je tenais également à remercier Virginie Marchand de m'avoir initiée aux techniques d'étude de la structure

secondaire des ARN et d’avoir toujours pris le soin d’inscrire nos échanges scientifiques dans un cercle vertueux.

Un grand merci à Lilia, Georges et Florence pour leur implication dans les expériences d’empreinte avec la protéine 9G8.

Je tenais aussi à remercier particulièrement Lilia et Maryline pour avoir relu mon manuscrit.

Je voulais remercier les personnes du laboratoire qui ont contribué à créer une ambiance de travail agréable. Je pense en particulier aux personnes du box 2, à Fabrice, à l’équipe enzymologiste de la pause café de 10h et « aux anciens » Nathalie, Benoit et Houda.

Merci à Maryline pour son soutien ainsi que nos nombreuses discussions scientifiques trépidantes.

Je terminerai par remercier les personnes qui me sont chères. J’adresse ma profonde reconnaissance à Athanase qui m’a beaucoup apporté, et ce, en toutes circonstances, et dont je loue l’optimisme, la patience et le dévouement qui ont été maintes fois mobilisés dans les moments difficiles. Je remercie infiniment mes amis, Fred et Arnaud dont le soutien sans faille a été bien plus que salutaire, surtout ces derniers temps. Quelle chance de vous avoir à mes côtés. Enfin, je remercie tendrement ma mère, pour ses encouragements inconditionnels durant toutes ces longues années d’étude et pour être la femme qu’elle est.

# Sommaire

**CHAPITRE I : EPISSAGE DES ARN PRE-MESSAGERS NUCLEAIRES CHEZ LES EUCARYOTES SUPERIEURS ..... 1**

I. LA REACTION MOLECULAIRE DE L'EPISSAGE ..... 1
II. LES SEQUENCES IMPORTANTES POUR LA RECONNAISSANCE DES INTRONS ..... 3
II.1. Le site donneur d'épissage ou site 5' ..... 3
II.2. Le site accepteur d'épissage ou site 3' ..... 4
II.2.1. La boîte de branchement ..... 4
II.2.2. La séquence polypyrimidine ..... 4
II.2.3. Le site 3' proprement dit ..... 4
III. LES COMPOSANTS ET L'ASSEMBLAGE DU SPLICEOSOME ..... 6
III.1. Les particules snRNP du spliceosome majeur ..... 6
III.1.1. La particule snRNP U1 ..... 7
III.1.2. La particule snRNP U2 ..... 10
III.1.3. La tri-snRNP [U4/U6.U5] ..... 12
III.1.3.1. La di-particule U4/U6 13S ..... 12
III.1.3.2. La particule snRNP U5 ..... 13
III.1.3.3. La tri-snRNP [U4/U6.U5] ..... 16
III.2. Les particules snRNP du spliceosome mineur ..... 18
III.2.1. La di-snRNP U11/U12 (18S) ..... 18
III.2.2. La tri-snRNP [U4atac/U6atac.U5] ..... 20
III.2.2.1. La di-snRNP U4atac/U6atac ..... 20
III.3. Assemblage des complexes spliceosomaux ..... 22
III.3.1. Généralité (Will et Luhrmann, 2011) ..... 22
III.3.2. Définition des exons et des introns et facteurs protéiques participant à cette définition ..... 24
III.3.2.2. Les protéines SR et leur rôle dans la reconnaissance du site donneur ..... 26
III.3.2.3. Les séquences reconnues par les protéines SR ..... 31
III.3.2.4. Les protéines SR like ..... 33
III.3.3. Le complexe H ..... 36
III.3.4. Les complexes E' et E ou complexes d'engagement ..... 37
III.3.5. Le complexe A ou préspliceosome ..... 40
III.3.6. Le complexes B et l'activation catalytique du spliceosome (Figure I-24 et I-25) ..... 43
III.3.6.5. Incorporation de la tri-snRNP [U4/U6.U5] dans le complexe B (Figure I-25, B) ..... 45
III.3.6.6. Le départ de la snRNP U1 et le complexe BΔU1 (Figure I-25, C) ..... 47
III.3.6.7. Le complexe Bact (Figure I-25, D) ..... 47
III.3.6.8. Le complexe B* (Figure I-26, E) ..... 50
III.3.7. Le complexe C et la seconde réaction de transestérification ..... 50
IV. INTEGRATION DE L'EPISSAGE DANS LE PROCESSUS D'EXPRESSION DES GENES ..... 52
IV.1. Lien entre la transcription et l'épissage ..... 52
IV.2. Lien entre la polyadénylation et l'épissage ..... 56
IV.3. Lien entre l'épissage, l'export et le système NMD, le rôle central du complexe EJC ..... 57

**CHAPITRE II : REGULATION DE L'EPISSAGE ALTERNATIF DES ARN PRE-MESSAGERS ..... 60**

I. GENERALITES SUR L'EPISSAGE ALTERNATIF ..... 60
I.1. Les différents types d'épissage alternatifs ..... 60
I.2. Les sites d'épissage alternatif sont souvent sous-optimaux ..... 62
I.3. L'utilisation des sites d'épissage faibles est souvent régulée par la combinaison de séquences *cis* auxquelles se fixent des facteurs *trans* ..... 62
II. LES PROTEINES HNRNP ET LEUR ROLE DANS L'EPISSAGE ..... 64
II.1. Organisation structurale des protéines hnRNP ..... 64
II.1.1. Les domaines de liaison à l'ARN ..... 66
II.1.1.1. Le domaine RRM ..... 66
II.1.1.2. Le domaine KH ..... 66
II.1.1.3. Le domaine RGG ou GAR ..... 68
II.1.2. Le domaine d'import (M9) et de rétention dans le noyau (NRS : Nuclear Retention Signal) ..... 68
II.1.3. Le domaine d'oligomérisation ou domaine Gly et son rôle dans la fonction des protéines des familles hnRNP A/B et H/F ..... 69
II.2. Les séquences ARN reconnues par les protéines des familles hnRNP A/B et hnRNP H/F ..... 70
II.3. Rôles et mécanismes généraux des protéines hnRNP dans la régulation de l'épissage alternatif ..... 72
III. ROLES ET MECANISMES GENERAUX DES PROTEINES SR DANS L'EPISSAGE ALTERNATIF ..... 74
IV. REGULATION DE L'EPISSAGE AU SITE ACCEPTEUR ..... 76
IV.1. Les sites accepteurs A3 et A7 du virus VIH-1 ..... 76
IV.2. Le site accepteur de l'exon 2a du pré-ARNm de l'α-tropomyosine de rat ..... 78
V. REGULATION DE L'EPISSAGE LORS DE LA DEFINITION DE L'INTRON ET DE L'EXON ..... 81
V.1. Rôles des protéines hnRNP ..... 81
V.2. Rôles des protéines « SR like » et SR ..... 83
VI. REGULATION DU RECRUTEMENT DE LA SNRNP U1 SUR LE SITE DONNEUR ..... 84
VI.1. Les différents cas de figure ..... 84
VI.2. Rôles des séquences ISE ..... 85
VI.3. Régulation de la compétition entre plusieurs sites donneurs ..... 86
VI.3.1. Rôles des séquences ESE et des protéines SR ..... 86
VI.3.2. Antagonisme entre les protéines SR et les protéines hnRNP ..... 88
VI.3.3. Architecture des exons ..... 90
VI.3.4. Régulation de l'épissage par les pseudo sites et les pseudoexons ..... 92
VI.4. Importance de la structure secondaire ..... 94

**CHAPITRE III : LE RETROVIRUS RSV ..... 97**

**AVANT PROPOS ..... 97**

I. CARACTERISTIQUES GENERALES DES RETROVIRUS ..... 98
II. LE RSV ET LES DIFFERENTS RETROVIRUS AVIAIRES ..... 99

III. PHYLOGENIE DES RETROVIRUS AVIAIRES ..... 101
IV. PATHOLOGIES ET IMPLICATIONS ECONOMIQUES LIEES AUX RETROVIRUS AVIAIRES ..... 102
V. STRUCTURE DU RETROVIRUS RSV (FIGURE III-4) ..... 104
VI. L'ARN GENOMIQUE ..... 105
VI.1. Les régions non traduites ..... 105
VI.1.1. La région 5' non traduite ..... 105
VI.1.2. La région 3' non traduite ..... 107
VI.2. Les phases ouvertes de lecture ..... 107
VII. CYCLE DE MULTIPLICATION DU VIRUS ..... 109
VII.1. La phase pré-intégrative ..... 109
VII.1.1. L'entrée du virus ..... 109
VII.1.2. La transcription inverse de l'ARN génomique (Telesnitsky et Goff, 1997) ..... 113
VII.1.3. Importation de l'ADN viral dans le noyau ..... 113
VII.1.4. Intégration de l'ADN proviral dans le génome de la cellule hôte ..... 115
VII.2. La phase post-intégrative ..... 117
VII.2.1. La transcription de l'ADN proviral ..... 117
VII.2.2. L'épissage du pré-ARNm ..... 120
VII.2.3. La polyadénylation ..... 121
VII.2.4. Export des ARN non épissés vers le cytoplasme ..... 123
VII.2.5. Traduction des ARN viraux ..... 127
VII.2.6. NMD et traduction des ARN non épissés ..... 129
VII.2.7. Le cas particulier de la protéine v-Src ..... 131
VII.2.8. Assemblage des nouveaux virions ..... 134
VII.2.8.1. Le rôle du précurseur Pr76 Gag-Pro lors de l'assemblage ..... 134
VII.2.8.2. Le recrutement de l'ARN génomique lors de l'assemblage ..... 135
VII.2.8.3. Le recrutement des protéines virales lors de l'assemblage ..... 137
VII.2.9. La maturation des nouveaux virions (Figure III-29) ..... 141

**CHAPITRE IV: LA REGULATION DE L'EPISSAGE DU RSV ET L'ELEMENT NRS ..... 143**

I. DESCRIPTION DES SITES D'EPISSAGE ET DES ARNM VIRAUX PRODUITS ..... 143
II. L'UNIQUE SITE DONNEUR EST UN SITE FORT ..... 145
III. LES SITES ACCEPTEURS 3' ENV ET 3' SRC SONT FAIBLES ..... 146
III.1. Le site 3'env a une boîte de branchement sous-optimale ..... 146
III.2. Le site 3'env est activé par une séquence ESE ..... 148
III.3. Le site 3'src ..... 149
III.3.1. Le site 3'src est d'origine cellulaire ..... 149
III.3.2. Le site 3'src a une séquence polypyrimidique sous-optimale ..... 151
III.4. Le site 3'src est inhibé par l'élément SSS (Src Suppressor of Splicing) ..... 153
III.5. Le site 3'src est inhibé par les éléments CTE (Constitutive Transport Element) ..... 155
IV. L'ELEMENT NRS EST UNE SEQUENCE INHIBITRICE DE L'EPISSAGE DU PRE-ARNM DU VIRUS RSV ..... 156
IV.1. L'élément NRS inhibe l'épissage aux sites 3'src et 3'env ..... 159
V. LE NRS EST UN ELEMENT BIPARTITE ..... 159

VI. PRINCIPE DE L'INHIBITION DE L'EPISSAGE DU PRE-ARNM DU VIRUS RSV PAR L'ELEMENT NRS ..... 160
VII. DESCRIPTION DU PSEUDO SITE DE L'ELEMENT NRS ..... 162
VII.1. Le pseudo site contient deux sites donneurs chevauchants qui fixent respectivement les snRNP U1 et U11 ..... 162
VII.2. Le pseudo site s'apparie avec le snRNA U1 ..... 162
VII.3. Le pseudo site s'apparie avec le snRNA U11 ..... 164
VII.4. Compétition des snRNP U11 et U1 pour la fixation au pseudo site de l'élément NRS ..... 165
VII.5. La snRNP U11 est recrutée par la hnRNP H ..... 165
VII.6. La fixation de la snRNP U1 sur le pseudo site est activée par la séquence ESE de la partie NRS 5' ..... 167
VII.6.1. La partie NRS 5' est une séquence ESE ..... 167
VII.6.1.1. La partie NRS 5' fixe les protéines SR et la protéine hnRNP H ..... 167
VII.6.1.2. L'activité ESE de la partie NRS 5' est essentielle pour l'inhibition de l'épissage par l'élément NRS ..... 169
VII.6.1.3. L'activité ESE de la partie NRS 5' favorise la fixation de la snRNP U1 au pseudo site ..... 171
VIII. L'ELEMENT NRS WT FORME UN COMPLEXE RIBONUCLEOPROTEIQUE NRS-C AVEC LES PROTEINES DE L'EXTRAIT NUCLEAIRE DE CELLULES HELA ..... 172
IX. L'ELEMENT NRS FORME DES COMPLEXES SPLICEOSOMAUX ABORTIFS LORSQU'IL EST ASSOCIE A UN SITE ACCEPTEUR ..... 172
X. LES COMPLEXES SPLICEOSOMAUX FORMES AVEC L'ELEMENT NRS ACTIVENT LA POLYADENYLATION ..... 174
XI. LE NRS FAVORISE LA MISE EN PLACE DES PROCESSUS D'ONCOGENESE CHEZ LES VIRUS ALV ..... 175
XI.1. Mécanismes d'oncogenèse provoqués par l'insertion des virus ALV ..... 175
XI.2. La troncation de l'élément NRS chez les virus ALV augmente leur caractère oncogénique ..... 177
XI.3. L'élément NRS régule l'épissage et la polyadénylation chez les virus ALV LR9 et ΔLR9 ..... 181
XII. IMPORTANCE DE LA STRUCTURE SECONDAIRE DE L'ELEMENT NRS ..... 183

**OBJECTIFS ..... 187**

## MATERIEL ET METHODES

**PARTIE MATERIEL ..... 189**

I. SOUCHES D'E. *COLI* UTILISEES ..... 189
II. LIGNEES DE CELLULES EUCARYOTES ..... 189
III. LES OLIGONUCLEOTIDES ..... 189
IV. LES SIRNA ..... 189
V. LES ANTICORPS ..... 194
VI. CONSTRUCTIONS GENETIQUES ..... 194
VI.1. Constructions renfermant les séquences de l'élément NRS du virus RSV de la souche Prague C ..... 194
VI.2. Constructions renfermant les séquences de l'élément NRS du virus ALV de la souche LR9 ..... 195

VI.3. Constructions réalisées pour la chromatographie d'affinité avec la technique tobramycine ..... 195
VI.4. Constructions réalisées pour la chromatographie d'affinité avec la technique MS2 ..... 198
VI.4.1. Réalisation de constructions génétiques par mutagenèse dirigée ..... 201
VI.5. Construction réalisées pour l'épissage *in vitro* ..... 202
VI.6. Construction réalisées pour l'épissage *in cellulo* ..... 204
VI.7. Constructions renfermant les séquences codant la protéine MPB-MS2 ..... 206

**PARTIE : METHODES ..... 207**

VII. TECHNIQUE D'ANALYSE DE LA STRUCTURE SECONDAIRE DES ARN ..... 207
VII.1. Préparation de la matrice ADN à transcrire ..... 207
VII.2. Transcription *in vitro* ..... 209
VII.3. Analyse de la structure secondaire des ARN par des sondes chimiques et enzymatiques (Ehresmann *et al*., 1987 ; Qu *et al*., 1983) ..... 209
VII.3.1. Principe ..... 209
VII.3.2. Conditions expérimentales ..... 210
VII.3.3. Modifications chimiques ..... 210
VII.3.4. Hydrolyses enzymatiques ..... 212
VII.4. Analyse par transcription inverse ..... 212
VII.4.1. Principe (Figure V-14) ..... 212
VII.4.2. Marquage radioactif des oligonucléotides en 5' ..... 213
VII.4.3. Réaction de transcription inverse (Qu *et al*., 1983) ..... 215
VII.4.3.1. Hybridation de l'amorce oligonucléotidique ..... 215
VII.4.3.2. Réaction d'élongation ..... 215
VII.4.4. Fractionnement sur gel ..... 216
VII.5. Prédiction in silico de la structure secondaire ..... 216
VIII. TECHNIQUES D'ANALYSE DES INTERACTIONS ARN/PROTEINES ..... 216
VIII.1. Technique d'empreinte des ARN en extrait nucléaire ou en présence de protéine purifiée ..... 216
VIII.1.1. Principe ..... 216
VIII.1.2. Assemblage des complexes ARN/protéines ..... 217
VIII.1.3. Hydrolyses enzymatiques ..... 217
VIII.2. Technique de retard sur gel ..... 218
VIII.2.1. Synthèse de transcrits radioactifs uniformément marqués ..... 218
VIII.2.2. Retard sur gel ..... 218
IX. METHODES SPECIFIQUES A L'ETUDE DE L'EPISSAGE *IN VITRO* ..... 220
IX.1. Epissage *in vitro* avec des transcrits uniformément marqués ..... 220
IX.1.1. Préparation de la matrice de transcription ..... 220
IX.1.2. Synthèse de transcrits radioactifs *in vitro* ..... 220
IX.1.3. Réaction d'épissage ..... 221
IX.1.4. Fractionnement sur gel et fluorographie ..... 222
IX.1.5. Détermination de l'efficacité d'épissage ..... 222
IX.2. Epissage *in vitro* avec des transcrits non radioactifs ..... 223
IX.2.1. Préparation de la matrice de transcription ..... 223
IX.2.2. Synthèse de transcrits non radioactifs *in vitro* ..... 223

IX.2.3. Réaction d'épissage ..... 223
IX.2.4. Analyse par RT-PCR des produits d'épissage ..... 224
IX.2.4.1. Principe ..... 224
IX.2.4.2. Réaction de transcription inverse sur les produits d'épissage ..... 225
IX.2.4.3. Amplification par PCR des produits de transcription inverse ..... 225
IX.2.4.4. Fractionnement en gel d'agarose des produits de RT-PCR et quantification ..... 229
X. METHODES SPECIFIQUES A LA PRODUCTION, LA PURIFICATION ET L'ANALYSE DES PROTEINES ..... 229
X.1. Analyse par électrophorèse en gel SDS-PAGE (Laemmli, 1970) ..... 229
X.2. Analyse par Western blot ..... 229
X.3. Production et purification de protéines recombinantes dans E. *coli* ..... 230
X.3.1. La protéine MS2-MBP ..... 230
X.3.1.1. Production ..... 230
X.3.1.2. Extraction des protéines cellulaires ..... 231
X.3.1.3. Chromatographie d'affinité sur des billes d'amylose ..... 231
X.3.1.4. Chromatographie d'affinité sur colonne Héparine ..... 231
XI. TECHNIQUES DE PURIFICATION DES COMPLEXES RIBONUCLEOPROTEIQUES PAR CHROMATOGRAPHIE D'AFFINITE ..... 231
XI.1. Chromatographie d'affinité reposant sur l'emploi de la tobramycine ..... 231
XI.1.1. Préparation de la matrice : fixation de la tobramycine sur la matrice Sepharose ..... 231
XI.1.2. Test de la matrice Tobra-Sepharose ..... 235
XI.1.2.1. Transcription *in vitro* des ARN utilisés dans la chromatographie d'affinité avec la technique tobramycine ..... 235
XI.1.2.2. Mesure de la capacité de la matrice Tobra-Sepharose à fixer l'ARN NRS-Tob ..... 236
XI.1.3. Purification des complexes ribonucléoprotéiques sur l'ARN NRS-Tob ..... 238
XI.1.4. Analyse des ARN contenus dans les complexes ribonucléoprotéiques ..... 241
XI.1.5. Analyse du contenu en protéines des complexes ribonucléoprotéiques ..... 241
XI.1.5.1. Analyse par SDS-PAGE ..... 241
XI.1.5.2. Analyse par Western Blot. ..... 242
XI.2. Purification des complexes ribonucléoprotéiques par chromatographie d'affinité reposant sur l'emploi de la protéine de l'enveloppe du phage MS2 ..... 242
XI.2.1. Analyse du contenu en ARN des complexes ribonucléoprotéiques par Northern Blot ..... 244
XI.2.2. Analyse du contenu en protéine des complexes ribonucléoprotéiques ..... 244
XI.2.2.3. Analyse par SDS-PAGE ..... 244
XI.2.2.4. Analyse par Western Blot ..... 245
XI.2.2.5. Analyse par spectrométrie de masse ..... 245
XII. METHODES UTILISEES POUR LA CULTURE CELLULAIRE ..... 245
XII.1. Lignée de cellules HeLa ..... 245
XII.2. Entretien des cellules ..... 245
XII.3. Congélation-décongélation ..... 246
XII.4. Conditions des transfections transitoires lors des surexpression des protéines SR ..... 246
XII.5. Conditions des transfections transitoires lors des expériences de siRNA ..... 247
XII.6. Extraction des ARN et analyse des produits d'épissage par transcription inverse ..... 247
XII.7. Extraction des protéines et analyse par Western blot ..... 248

# RESULTATS

**CHAPITRE VI : ANALYSE STRUCTURALE DE L'ELEMENT NRS DES VIRUS DE LA FAMILLE DES ASLV ..... 249**

I. ETAT DES TRAVAUX LORSQUE JE SUIS ARRIVEE AU LABORATOIRE ..... 249
II. STRATEGIE ADOPTEE POUR COMPLETER L'ANALYSE STRUCTURALE DE L'ELEMENT NRS DU VIRUS RSV .. 252
II.1. Choix de la séquence utilisée pour l'analyse structurale de l'élément NRS WT et son variant Δ76 . 252
II.2. Principes de l'emploi de sondes enzymatiques pour l'analyse de la structure des ARN en solution . 253
II.2.1. Identification des positions de clivage par transcription inverse et choix des oligonucléotides amorces ..... 254
II.3. Résultats obtenus ..... 254
II.4. Etudes comparatives des éléments NRS d'autres virus ..... 255

**CHAPITRE VII : ETUDE DES SITES DE FIXATION DES PROTEINES D'UN EXTRAIT NUCLEAIRE DE CELLULES HELA SUR LES ELEMENTS NRS DES VIRUS RSV ET ALV ..... 261**

I. OBJECTIFS ..... 261
II. STRATEGIE ..... 261
II.1. Principe ..... 261
II.2. Choix de l'extrait nucléaire et de la sonde enzymatique ..... 262
II.3. Mise au point des conditions d'empreinte en présence d'un extrait nucléaire de cellules HeLa ..... 262
III. IDENTIFICATION DES REGIONS PROTEGEES ..... 263
III.1.1. Les empreintes en extrait nucléaire recoupent le caractère bipartite de l'élément NRS ..... 264

**CHAPITRE VIII : RECHERCHE D'UNE METHODE DE CHROMATOGRAPHIE D'AFFINITE ADAPTEE A LA PURIFICATION DES COMPLEXES RIBONUCLEOPROTEIQUES FORMES SUR L'ELEMENT NRS ..... 269**

I. PRESENTATION DES TECHNIQUES DE PURIFICATION DES PARTICULES RIBONUCLEOPROTEIQUES PAR CHROMATOGRAPHIE D'AFFINITE ..... 269
II. ESSAI DE DEVELOPPEMENT D'UNE CHROMATOGRAPHIE D'AFFINITE BASEE SUR LE MOTIF ARN EN K-LOOP ET LA PROTEINE L7AE ..... 272
III. CHROMATOGRAPHIE D'AFFINITE BASEE SUR L'UTILISATION DE LA TOBRAMYCINE ..... 272
III.1. Les différents acteurs intervenant dans la chromatographie d'affinité basée sur l'emploi de la tobramycine ..... 272
III.2. Réalisation des constructions génétiques ..... 274
III.3. Evaluation de l'efficacité de rétention de l'ARN NRS-Tob sur la matrice Tobra-Sepharose ..... 274
III.3.1. Amélioration des conditions de rétention de l'ARN sur la matrice Tobra-Sepharose ..... 279
III.4. Purifications des complexes ribonucléoprotéiques ..... 279
III.5. Analyse des protéines contenues dans les complexes ribonucléoprotéiques assemblés sur l'ARN NRS-Tob ..... 283

III.6. Analyse des ARN contenus dans les complexes ribonucléoprotéiques assemblés sur l'ARN NRS-Tob ..... 286
III.7. Conclusion ..... 287
IV. CHROMATOGRAPHIE D'AFFINITE BASEE SUR LA PROTEINE DE L'ENVELOPPE DU PHAGE MS2 ..... 288
IV.1. Production des ARN hybrides ..... 288
IV.1.1. Le motif ARN reconnu par la protéine de l'enveloppe du phage MS2 ..... 288
IV.1.2. Réalisation des constructions génétiques ..... 288
IV.2. Production de la protéine de fusion ..... 290
IV.2.1. La protéine de fusion utilisée ..... 290
IV.2.2. Production et purification de la protéine de fusion ..... 292
IV.3. Adaptation de la chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2 à la purification des complexes ribonucléoprotéiques formés sur l'élément NRS ..... 294
IV.3.1. Test de la fixation de la protéine MS2-MBP sur les ARN MS2 et NRS-MS2 par expériences de retard sur gel ..... 294
IV.3.2. Définition des quantités d'extrait nucléaire à utiliser par expériences de retard sur gel ..... 295
IV.3.3. Etude des interactions aspécifiques de la résine amylose avec les composants de l'extrait nucléaire de cellules HeLa ..... 297
IV.4. Expérience test réalisée avec 500 pmol d'ARN et un rapport molaire [protéine MS2 MBP]/[ARN] de 20 ..... 299
IV.4.1. Purification des complexes ribonucléoprotéiques ..... 299
IV.4.2. Analyse par spectrométrie de masse des protéines contenues dans les complexes ribonucléoprotéiques purifiés ..... 303
IV.5. Expérience réalisée avec 250 pmol d'ARN et un rapport molaire [protéine MS2 MBP]/[ARN] de 5 ..... 309
IV.6. Purification des complexes ribonucléoprotéiques formés sur les ARN NRS tronqués ..... 317
IV.6.1. Test de la fixation de la protéine MBP-MS2 sur les différentes formes tronquées de l'ARN NRS fusionnées à la séquence MS2 par expériences de retard sur gel ..... 317
IV.7. Des complexes ribonucléoprotéiques ayant des compositions différentes sont obtenus en utilisant les différentes formes tronquées de l'ARN NRS ..... 318
IV.7.1. Comparaison des protéines identifiées dans les complexes assemblés sur l'ARN NRS-MS2 au cours des 3 expériences de chromatographie d'affinité ..... 325
IV.7.2. Comparaison des protéines identifiées dans les complexes assemblés sur l'ARN NRS-MS2 WT et sur les différents ARN NRS tronqués ..... 328
IV.7.2.6. Les protéines des particules snRNP ..... 328
IV.7.3. Etude de la présence des snRNA dans les complexes ribonucléoprotéiques par Northern blot ..... 345
IV.7.4. Etude de la présence des protéines SR et hnRNP dans les complexes ribonucléoprotéiques par Western blot ..... 347

**CHAPITRE IX : ETUDE DES REGIONS DE L'ELEMENT NRS PROTEGEES PAR LA PROTEINE HNRNP A1 RECOMBINANTE ..... 349**

I. OBJECTIF ..... 349
II. STRATEGIE ..... 349
II.1. Choix de l'ARN et de la sonde enzymatique ..... 349

II.2. Détermination de la quantité de protéine hnRNP A1 utilisée pour les expériences d'empreinte ..... 349

**CHAPITRE X : ETUDE FONCTIONNELLE DE L'ELEMENT NRS PAR DES EXPERIENCES D'EPISSAGE *IN VITRO* ET *IN CELLULO* ..... 355**

I. OBJECTIF ..... 355

II. ETUDE DE L'ACTIVITE DES ELEMENTS NRS SAUVAGE ET TRONQUES PAR DES EXPERIENCES D'EPISSAGE *IN VITRO* ..... 355

II.1. Design et production de constructions d'épissage permettant de tester *in vitro* l'activité de l'élément NRS ..... 355

II.2. Expériences d'épissage réalisées avec des pré-ARN uniformément marqués ..... 357

II.3. Expériences d'épissage réalisées avec des pré-ARNm non radioactifs et mise en évidence des produits d'épissage par RT-PCR ..... 359

II.3.1. Mise au point des conditions de RT-PCR ..... 359

II.3.1.7. Choix des constructions ..... 359

II.3.1.8. Conditions d'épissage ..... 359

II.3.1.9. Choix de l'extrait nucléaire et des oligonucléotides de la réaction de PCR ..... 359

II.3.1.10. Linéarité des amplifications par RT-PCR (Figure X-7) ..... 366

II.3.2. Etude fonctionnelle de l'élément NRS par expériences d'épissage *in vitro* ..... 368

II.3.3. Reproductibilité des expériences de RT-PCR ..... 370

III. ETUDE FONCTIONNELLE DE L'ELEMENT NRS PAR DES EXPERIENCES D'EPISSAGE *IN CELLULO* ..... 372

III.1. Choix de la lignée cellulaire, des constructions d'épissage et des plasmides utilisés pour l'étude *in cellulo* ..... 372

III.1.1. La lignée cellulaire ..... 372

III.1.2. Les constructions d'épissage et les plasmides ..... 372

III.1.3. Linéarité des amplifications PCR ..... 374

III.1.4. Résultats et reproductibilité des amplifications par RT-PCR ..... 376

III.2. Reproductibilité des résultats d'épissage *in cellulo* ..... 378

**ARTICLE SCIENTIFIQUE ..... 379**

**DISCUSSION ET PERSPECTIVES**

I. LES ELEMENTS NRS DES DIFFERENTES ESPECES DE VIRUS ADOPTENT UNE STRUCTURE SECONDAIRE SIMILAIRE ..... 430

II. LA STRUCTURE SECONDAIRE DE L'ELEMENT NRS EST CONSERVEE CHEZ LES VIRUS ASLV ..... 432

II.1. La conservation de la structure secondaire de l'élément NRS chez les virus ASLV ne repose pas uniquement sur la capacité codante de la séquence ..... 433

II.2. Les éléments NRS des virus ALV Strain Tym_S90 et Y53 Sarcoma virus ne seraient pas fonctionnels ..... 436

II.3. Lien entre l'élément NRS et l'oncogénicité des virus ..... 438

III. L'ELEMENT NRS 22 CONTENANT L'HELICE H1 ET LES TIGES BOUCLES SLS1 ET SLS4 EST LE NOUVEL ELEMENT NRS MINIMAL FONCTIONNEL POUR L'INHIBITION DE L'EPISSAGE *IN VITRO* ..... 441
III.1. Relation structure-fonction de l'élément NRS 22 ..... 444
III.2. Fonctionnement de l'élément NRS 22 lors de l'inhibition de l'épissage ..... 446
IV. LES TIGES BOUCLES SLS1 ET SLS4 DE L'ELEMENT NRS SONT ESSENTIELLES A SA FONCTION ..... 449
IV.1. La présence du pseudo site et la structure secondaire tige boucle SLS4 qu'il adopte sont essentielles mais pas suffisantes pour la fonction de l'élément NRS ..... 451
IV.1.1. La fixation de la snRNP U1 sur le pseudo site dépend de sa séquence et de la structure secondaire qu'il adopte ..... 451
IV.1.2. Le recrutement de la snRNP U1 sur le pseudo site lors de l'épissage dépend aussi de facteurs supplémentaires ..... 452
IV.2. Les tiges boucles SLS1 et SLS4 sont nécessaires à la fixation des protéines SR au sein des complexes ribonucléoprotéiques assemblés sur l'élément NRS ..... 452
IV.3. La séquence de la tige boucle SLS1 qui fixe les protéines ASF/SF2 et 9G8 est essentielle à l'inhibition de l'épissage par l'élément NRS ..... 454
V. L'INHIBITION DE L'EPISSAGE PAR L'ELEMENT NRS EST ACTIVEE PAR LA FIXATION DE LA PROTEINE 9G8 SUR LA SEQUENCE DE L'HELICE H2 ..... 455

## BIBLIOGRAPHIE

# Liste des Figures et des Tableaux :

## Introduction


Figure I-1 : Représentation moléculaire des deux étapes de la réaction d'épissage des introns des ARN pré-messagers nucléaires (d'après Jura et Moore, 2002) ..... **2**
Figure I-2 : Séquences des sites 5' et 3' des introns U2-dépendants (A, B) et U12-dépendants (C, D) essentielles à la réaction d'épissage. ..... **5**
Figure I-3 : Structure de la particule snRNP U1 (Pomeranz Krummel *et al*., 2009). ..... **8**
Figure I-4 : Les différents appariements entre la snRNP U1 et le site donneur du pré-ARNm. ..... **9**
Figure I-5 : Représentation de la structure de la particule snRNP U2 et de son appariement avec la boîte de branchement. **11**
Figure I-6 : Représentation de la structure secondaire des snRNA U4/U6 de la di-particule snRNP U4/U6 et de leurs protéines associées. ..... **13**
Figure I-7 : Représentation de la structure secondaire du snRNA U5 et de ses protéines associées. ..... **15**
Figure I-8 : Représentation de la tri-snRNP [U4/U6.U5] et de ses protéines associées. ..... **17**
Figure I-9 : Représentation de la structure secondaire des snRNA U11 et U12 et de la reconnaissance des sites donneur et accepteur mineurs par la di-snRNP U11/U12. ..... **19**
Figure I-10 : Représentation de la structure secondaire des snRNA U4atac/U6atac de la di-particule snRNP U4atac/U6atac et de leurs protéines associées. ..... **21**
Figure I-11 : Assemblage des spliceosomes majeurs et mineurs (d'après Will et Lührmann, 2010 ; Patel et Steitz, 2003). ..... **23**
Figure I-12 : Définition de l'intron et de l'exon d'un pré-ARNm au cours des étapes précoces de l'assemblage du spliceosome. ..... **25**
Figure I-13 : Organisation structurale des protéines SR humaines (Manley et Krainer, 2010 ; Graveley, 2000). ..... **26**
Figure I-14 : Interactions entre les domaines RS et le pré-ARNm au cours de l'épissage (d'après Shen et Green, 2004, Shen et Green, 2007). ..... **28**
Figure I-15 : Interaction entre la protéine U1-70K et la protéine ASF/SF2 phosphorylée lors du recrutement de la snRNP U1 sur le site donneur (Cho *et al*., 2011b). ..... **30**
Figure I-16 : Cycles de phosphorylations et navettes noyau-cytoplasme des protéines SR (Huang *et al*., 2004). ..... **31**
Figure I-17 : Séquences ARN des sites de fixation des protéines SR identifiées expérimentalement (A) ou *in silico* par le programme ESE finder (B). ..... **32**
Figure I-18 : Organisation structurale des protéines « SR like » humaines et identification des séquences de fixation de certaines d'entre elles (Graveley, 2000). ..... **33**
Figure I-19 : Reconnaissance du site accepteur par les facteurs U2AF et la protéine SF1/mBBP (Adapté de Mackereth *et al*., 2005 ; Moore, 2000). ..... **35**
Tableau I-20 : Les protéines du complexe H (Adapté de Jurica *et al*., 2002 ; Zhou *et al*., 2002). ..... **37**
Figure I-21 : Modèle d'assemblage des complexes E' et E (Adapté de Kent *et al*., 2005). ..... **39**
Figure I-22 : Modèle d'assemblage du complexe A (Adapté de Das *et al*., 2000 ; Perriman et Ares, 2010). ..... **41**
Tableau I-23 : Les protéines du complexe spliceosomal A (Behzadnia *et al*., 2007 ; Hartmuth *et al*., 2002). ..... **42**
Figure I-24 : Changements de la composition des complexes spliceosomaux A (A), B (B) et Bact (C) lors de l'épissage (Adapté de Behzadnia *et al*., 2007, Bessonov *et al*., 2010, Wahl *et al*., 2009, Agafonov *et al*., 2011). ..... **44**
Figure I-25 : Modèle des réarrangements des snRNA lors de l'activation du spliceosome (adapté de Wahl *et al*., 2009, Will et Lührmann, 2010, Schneider *et al*., 2010). ..... **46**
Figure I-26 : La protéine hPrp8 est au cœur catalytique du spliceosome et coordonne l'activation des hélicases permettant la transition du complexe B vers le complexe Bact (d'après Grainger et Beggs, 2005, Pena *et al*., 2007, Pena *et al*., 2008). ..... **49**
Figure I-27 : Changements de la composition des complexes spliceosomaux Bact (A) et C (B) lors de l'épissage (adapté de Bessonov *et al*., 2008, Wahl *et al*., 2009, Golas *et al*., 2010, Agafonov *et al*., 2011). ..... **51**
Figure I-28 : Modèle des réarrangements des snRNA lors de l'activation du spliceosome (adapté de Mayas *et al*., 2006). ..... **53**
Figure I-29 : Influence du taux d'élongation sur l'épissage alternatif (De la Mata *et al*., 2010). ..... **55**
Figure I-30 : Couplage de l'épissage et de la polyadénylation des pré-ARNm (d'après Martinson, 2010). ..... **57**
Figure I-31 : Mise en place du système NMD (Köhler et Hurt, 2007, Isken *et al*., 2008). ..... **59**

Figure II-1 : Les différents mécanismes d'épissage alternatif (Ladd et Cooper, 2002). ..... **61**
Figure II-2 : Les sites d'épissage alternatifs sont souvent faibles et leur utilisation est régulée par des séquences *cis*. ..... **63**
Figure II-3 : Structure, rôle et séquences ARN reconnues par les protéines de la famille hnRNP (d'après Han *et al*., 2010, Swanson et Dreyfuss, 1988 ; Dreyfuss *et al*., 1993 ; Honoré *et al.*, 1999 ; Caputi et Zahler, 2001 ; Lerga *et al*., 2001 ; Kafasla *et al*., 2002 ; Krécic et Swanstrom, 1999 ; Dreyfuss *et al*., 2002). ..... **65**
Figure II-4 : Structures du domaine RRM2 de la protéine hnRNP A1 et du domaine qRRM2 de la protéine hnRNP F (Dominguez *et al*., 2010). ..... **67**
Figure II-5 : Structure du troisième domaine KH de la protéine hnRNP K en interaction avec un ADN simple brin (Valverde *et al*., 2008). ..... **68**
Figure II-6 : Structures et importances fonctionnelles des différents domaines des protéines hnRNP A1, UP1 (A) et des protéines de la famille des hnRNP H/F (B). ..... **70**
Figure II-7 : Liste non exhaustive des sites de fixation reconnus par la protéine hnRNP A1 (A) et de la protéine hnRNP H (B). ..... **71**
Figure II-8 : Inhibition de l'épissage par les protéines hnRNP lors de la définition de l'exon (A) ou de la définition de l'intron (B). ..... **73**

Figure II-9 : Activation de l'épissage par les protéines SR lors de la définition de l'exon (A) ou de la définition de l'intron (B). ..... **75**
Figure II-10 : Régulation de l'épissage au site accepteur A3 du pré-ARNm du virus VIH-1 (Jacquenet *et al*., 2001 ; Hallay *et al*., 2006 ; Zahler *et al*., 2004). ..... **77**
Figure II-11 : Régulation de l'épissage au site accepteur A7 du pré-ARNm du virus VIH-1 (Marchand *et al*., 2002). ..... **78**
Figure II-12 : Régulation de l'épissage des exons 2a et 2b du pré-ARNm de l'α-tropomyosine de rat (Crawford et Patton, 2006 ; Gooding et Smith, 2008). ..... **80**
Figure II-13 : La protéine hnRNP A1 régule l'épissage alternatif de son propre pré-ARNm en séquestrant l'exon 7B dans une boucle (Chabot *et al*., 2003). ..... **81**
Figure II-14 : Modèle d'inhibition de l' épissage de l'exon N1 du pré-ARNm c-*src* par la protéine PTB (Shama *et al*., 2011). ..... **82**
Figure II-15 : Activation de la définition des introns par les co-activateurs SRm160/SRm300 (A) et la protéine SRp38 phosphorylée (B) (Eldridge *et al*., 1999 ; Feng *et al*., 2008). ..... **83**
Figure II-16 : Exemples de régulation du recrutement de la snRNP U1 sur le site donneur (Cho *et al*., 2011 ; Shin *et al*., 2004 ; Buratti *et al*., 2004). ..... **85**
Figure II-17 : Régulation du recrutement de la snRNP U1 sur le site donneur par la fixation de la protéine TIA-1 sur les séquences ISE (Forch *et al*., 2001 ; Izquierdo *et al*., 2005). ..... **86**
Figure II-18 : Régulation de l'épissage du pré-ARNm de l'unité de transcription E1A par une ESE bi-directionnelle liant différentes protéines SR (Bourgeois *et al*., 1999). ..... **87**
Figure II-19 : Exemple de mécanisme d'antagonisme entre les protéines hnRNP A1 et ASF/SF2 pour la sélection d'un site 5' d'épissage (Eperon *et al*., 2000). ..... **89**
Figure II-20 : Exemples de régulation de l'épissage entre deux sites donneurs en compétition. ..... **91**
Figure II-21 : Inhibition de l'épissage par la fixation de la snRNP U1 sur des sites donneurs (Dhir et Buratti, 2010 ; Smith et Valcarcel, 2000). ..... **93**
Figure II-22 : Rôle de la structure secondaire dans la régulation de l'épissage (Buratti et Baralle, 2004). ..... **95**

Figure III-1 : Infection et transformation des cellules par différents rétrovirus aviaires (d'après Martin, 2004). ..... **100**
Tableau III-2 : Classification des différents sous-groupes de rétrovirus du genre *Alpharétrovirus* selon la séquence de la protéine de l'enveloppe gp 85 (SU) et les espèces aviaires hôtes (d'après Venugopal, 1999). ..... **102**
Figure III-3 : Dendogramme représentant les différents sous-groupes de rétrovirus du genre *Alpharétrovirus* infectant le poulet (d'après Bai *et al*., 1995). ..... **103**
Figure III-4 : Structure du virion mature du virus RSV (adapté de Coffin *et al*., 1997). ..... **104**
Figure III-5 : Représentation des régions non traduites de l'ARN du virus RSV. ..... **106**
Figure III-6 : Traduction et maturation des protéines virales du virus RSV (d'après Coffin *et al*., 1997). ..... **108**
Tableau III-7 : Les rôles des différentes protéines virales. ..... **109**
Figure III-8 : La phase pré-intégrative du cycle viral. ..... **110**
Figure III-9 : Les gènes des récepteurs cellulaires responsables de la susceptibilité génétique des cellules de poulet à l'infection par les virus ASLV (d'après Fadly, 2008). ..... **111**
Figure III-10 : Processus de fusion entre la membrane du virion et la membrane de la cellule hôte (d'après White *et al*., 2008). ..... **112**
Figure III-11 : La transcription inverse de l'ARN génomique du RSV (d'après Telesnitsky et Goff, 1997). ..... **114**
Figure III-12 : Intégration de l'ADN viral dans le génome de la cellule hôte. ..... **116**
Figure III-13 : La phase post-intégrative du cycle viral. ..... **118**
Figure III-14 : Représentation du promoteur du RSV (d'après Ryden *et al*., 1195 ; Ruddell, 1995). ..... **119**
Figure III-15 : Représentation de l'ARN génomique et des différents ARNm du virus RSV. ..... **120**
Figure III-16 : Représentation des séquences importantes pour la polyadénylation des transcrits du virus RSV (d'après McNally, 2008). ..... **122**
Figure III-17 : Représentation des séquences DR dans l'ARN génomique des virus ALV et RSV (d'après Yang et Cullen, 1999 ; Ogert *et al*., 1996). ..... **124**
Figure III-18 : Modèle hypothétique d'export des ARN viraux non épissés par les éléments CTE localisés dans les séquences DR du virus RSV (d'après LeBlanc *et al*., 2007). ..... **126**
Figure III-19 : Représentation de la séquence IRES et des séquences régulant l'initiation de la traduction des ORF du virus RSV (d'après Deffaud et Darlix, 2000). ..... **128**
Figure III-20 : Traduction des protéines à partir des différents ARN viraux du virus RSV (d'après Coffin *et al*., 1997). ..... **130**
Figure III-21 : L'élément RSE protège l'ARNm *gag-pro* de l'action du NMD (d'après Withers et Beemon, 2010). ..... **132**
Figure III-22 : Activation de la protéine c-Src (d'après Martin, 2001). ..... **133**
Figure III-23 : Comparaison des gènes et des protéines codées par le proto-oncogène cellulaire c-*src* et l'oncogène v-*src* du RSV (d'après Martin, 2001). ..... **134**
Figure III-24 : Les différents domaines protéiques du précurseur Pr76 Gag-Pro importants pour le bourgeonnement. ..... **135**
Figure III-25 : Dimérisation de l'ARN génomique du virus RSV (Ben Ali *et al*., 2007). ..... **136**
Figure III-26 : Modèle de recrutement de l'ARN génomique lors de l'assemblage des particules virales (Gudleski *et al*., 2010). ..... **138**
Figure III-27 : Implication des corps multivesiculaires (MVB) dans le trafic endosomal de la cellule (d'après Raiborg et Stenmark, 2009). ..... **139**
Figure III-28 : Ubiquitinylation du précurseur Pr76 Gag-Pro et recrutement du complexe ESCRT lors du bourgeonnement (d'après Pincetic et Leis, 2009). ..... **140**
Figure III-29 : Bourgeonnement et maturation du virion du virus RSV (d'après Coffin *et al*., 1997). ..... **142**

Figure IV-1 : Transcrit primaire et ARNm du virus RSV (McNally, 2008). ....144
Figure IV-2 : Comparaison de la séquence du site donneur d'épissage de l'ARN du virus RSV avec la séquence consensus des sites donneurs des pré-ARNm des eucaryotes supérieurs. ....145
Figure IV-3 : Description du site accepteur 3'*env* au sein du transcrit primaire du virus RSV et comparaison de sa séquence avec la séquence consensus des sites accepteurs des pré-ARNm des eucaryotes supérieurs. ....147
Figure IV-4 : Le site 3'*env* possède une séquence ESE. ....149
Figure IV-5 : Description du site accepteur 3'*src* au sein du transcrit primaire du virus RSV et comparaison de sa séquence avec la séquence consensus des sites accepteurs des pré-ARNm des eucaryotes supérieurs. ....151
Figure IV-6 : L'épissage au site 3'*src* est inhibé par l'élément SSS. ....153
Figure IV-7 : L'épissage au site 3'*src* est inhibé par les éléments CTE. ....155
Figure IV-8 : Localisation de l'élément NRS au sein du transcrit primaire du virus RSV. ....157
Figure IV-9 : Définition de l'élément NRS WT et de l'élément NRS minimal fonctionnel NRS Δ76 en contexte hétérologue (Arrigo et Beemon, 1988 ; McNally *et al*., 1991). ....158
Figure IV-10 : L'élément NRS est un élément bipartite (McNally *et al*., 1991 ; McNally, 2008). ....160
Figure IV-11 : Modèle d'inhibition de l'épissage d'un pré-ARNm du RSV par l'élément NRS (d'après McNally, 2008). ....161
Figure IV-12 : Le pseudo site de l'élément NRS contient deux sites donneurs chevauchants (d'après Hibbert *et al*., 1999). ....162
Figure IV-13 : Appariement du snRNA U1 sur le pseudo site 5' de l'élément NRS (Hibbert *et al*., 1999 ; McNally *et al*., 1999). ....163
Figure IV-14 : Appariement du snRNA U11 sur le pseudo site 5' de l'élément NRS (McNally *et al*., 1999). ....165
Figure IV-15 : La fixation de la snRNP U11 sur le pseudo site de l'élément NRS est activée par la protéine hnRNP H (McNally *et al*., 2004 ; McNally *et al*., 2006). ....166
Figure IV-16 : La partie NRS 5' contient plusieurs séquences ESE qui fixent les protéines SR et la protéine hnRNP H (McNally *et al*., 1996 ; McNally *et al*., 1998 ; Fogel *et al*., 2000 ; Fogel *et al*., 2002). ....168
Figure IV-17 : L'activité ESE de la partie NRS 5' ainsi que des ESE hétérologues sont essentielles à l'inhibition de l'épissage par l'élément NRS et la partie NRS 5' favorise la fixation de la snRNP U1 sur le pseudo site de l'élément NRS (McNally et McNally, 1996 ; McNally et McNally, 1998 ; McNally et McNally, 1999). ....170
Figure IV-18 : Le NRS forme avec le site 3'*env* un complexe spliceosomal B abortif (Giles et Beemon, 2005). ....173
Figure IV-19 : Modèle d'activation de la polyadénylation du pré-ARNm viral par l'élément NRS (Maciolek et McNally, 2007 ; Fogel *et al*., 2002 ; Wilusz et Beemon, 2006). ....175
Figure IV-20 : Les insertions des provirus ALV provoquent des lymphomes à temps de latence court et long. ....176
Figure IV-21 : La séquence du virus EU8 qui induit des lymphomes à courte durée de latence est tronquée de 42 nucléotides dans l'élément NRS (Polony *et al*., 2003 ; Smith *et al*., 1997). ....178
Figure IV-22 : La troncation de 42 nucléotides de la séquence de l'élément NRS provoque une augmentation de la pathogenicité du virus ALV LR9 (Polony *et al*., 2003 ; Smith *et al*., 1997). ....179
Figure IV-23 : Alignement des séquences correspondant à l'élément NRS (nt 703-930) des virus RSV Prague C et ALV LR9 et EU8. ....180
Figure IV-24 : Répercussions de l'élément NRS ΔLR9 non fonctionnel sur l'épissage et la polyadénylation du pré-ARNm viral (Polony *et al*., 1997 ; Smith *et al*., 1997). ....182
Figure IV-25 : Structure secondaire théorique de l'élément NRS WT du virus RSV prédite par le logiciel Mfold (McNally *et al*., 1991). ....184
Figure IV-26 : Structure de la séquence du pseudo site de l'élément NRS WT du virus RSV (NRS 23) résolue par RMN et relation avec sa fonction (Cabello-Villegas *et al*., 2004 ; Paca *et al*., 2001). ....185

## Matériel et Méthodes

Tableau V-1 : Oligonucléotides utilisés au cours de ce travail. ....189
Tableau V-2 : Si RNA utilisés au cours de ce travail. ....193
Tableau V-3 : Liste des anticorps utilisés au cours de ce travail. ....194
Figure V-4 : Constructions contenant l'ADNc des virus RSV et ALV. ....196
Figure V-5 : Constructions réalisées pour produire les ARN utilisés dans les expériences de chromatographie d'affinité basée sur l'emploi de la tobramycine. ....197
Figure V-6 : Constructions réalisées pour produire l'ARN MS2. ....198
Figure V-7 : Constructions réalisées pour produire les ARN de fusion utilisés dans les expériences de chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2. ....200
Figure V-8 : Constructions utilisées pour les expériences d'épissage *in vitro*. ....202
Figure V-9 : Constructions réalisées pour les expériences d'épissage *in vitro*. ....203
Figure V-10 : Stratégie de clonage des constructions réalisées pour les expériences d'épissage *in cellulo*. ....205
Figure V-11 : Constructions réalisées pour les expériences d'épissage *in cellulo*. ....206
Figure V-12 : Constructions utilisées dans le cadre de l'étude structurale de l'élément NRS. ....208
Figure V-13 : Sondes enzymatiques et chimiques utilisées pour l'étude de la structure secondaire des ARN en solution. ....211
Figure V-14 : Principe de l' identification des coupures d'un ARN par une sonde enzymatique par la transcription inverse. ....214
Figure V-15 : Représentation de la réaction d'épissage *in vitro* suivie d'une analyse par RT-PCR. ....227
Figure V-16 : Exemple de quantification des produits de RT-PCR après électrophorèse en gel d'agarose en présence de bromure d'éthidium. ....228
Figure V-17 : Préparation de la matrice. ....232

Figure V-18 : Tampons utilisés pour la préparation de la matrice. ....233
Figure V-19 : Fixation de la tobramycine sur la matrice. ....234
Figure V-20 : Tampons utilisés pour le test de la matrice. ....237
Figure V-21 : Test de fixation de l'ARN NRS-Tob sur la matrice Tobra-Sepharose. ....237
Figure V-22 : Tampons utilisés pour la purification des complexes ribonucléoprotéiques. ....239
Figure V-23 : Purification des complexes ribonucléoprotéiques sur l'ARN NRS-Tob par chromatographie d'affinité sur la matrice Tobra-Sepharose. ....240
Figure V-24 : Purification des complexes ribonucléoprotéiques par chromatographie d'affinité reposant sur l'emploi de la protéine de l'enveloppe du phage MS2. ....243

## Résultats

Figure VI-1 : Construction génétique utilisée et ARN produits par A. Mougin pour l'étude structurale de l'élément NRS. ....250
Figure VI-2 : Comparaison des structures secondaires alternatives des régions 5' et 3' de l'élément NRS WT en l'absence (A) et en présence de 20 mM de $MgCl_2$ (B) lors d'expériences de cartographie de la structure de l'ARN en solution avec la structure tertiaire du segment 910-926 de l'élément NRS résolue par RMN (Cabello-Villegas *et al*., 2004). ....251
Figure VI-3 : Représentation des ARN contenant l'élément NRS choisis pour l'étude structurale. ....253
Figure VI-4 : Comparaison de la structure secondaire prédite (A) et des structures établies expérimentalement (B) pour l'élément NRS WT et l'élément NRS Δ76. ....256
Figure VI-5 : Comparaison des structures secondaires établies expérimentalement pour les éléments NRS WT (A), NRS LR9 (B) et NRS ΔLR9 (C). ....257
Figure VI-6 : Alignement des séquences correspondant à l'élément NRS de différents virus de la famille ASLV. ....259
Figure VI-7 : Conservation de la structure secondaire de l'élément NRS chez différents virus de la famille ASLV. ....260

Figure VII-1 : Analyse par expériences de retard sur gel des complexes ribonucléoprotéiques formés après incubation des ARN WT du virus RSV (A), NRS LR9 du virus ALV-LR9 (B) et NRS Δ76 du virus RSV (C) en présence d'extrait nucléaire de cellules HeLa. ....263
Figure VII-2 : Analyse des protections contre l'action de la RNase T1 générées sur l'ARN NRS par les composants d'un extrait nucléaire de cellules HeLa. ....265
Figure VII-3 : Analyse des protections contre l'action de la RNase T1 générées sur l'élément NRS Δ76 par les composants d'un extrait nucléaire de cellules HeLa. ....266
Figure VII-4 : Analyse des protections contre l'action de la RNase T1 générées sur l'élément NRS LR9 par les composants d'un extrait nucléaire de cellules HeLa. ....267
Figure VII-5 : Comparaison des protections obtenues sur l'ARN NRS du virus RSV lors d'expériences d'empreinte en présence d'extrait nucléaire de cellules HeLa avec les sites de fixation des protéines et des particules connus de cet ARN. ....268

Figure VIII-1 : Représentation des différentes séquences pouvant être fusionnées à l'ARN d'intérêt, leurs ligands et les résines associées dont nous avons testé l'efficacité pour la purification des complexes ribonucléoprotéiques formés avec l'ARN NRS. ....271
Figure VIII-2 : Description des différents éléments intervenant dans la chromatographie d'affinité basée sur l'emploi de la tobramycine. ....273
Figure VIII-3 : Constructions réalisées pour produire les ARN utilisés dans les expériences de chromatographie d'affinité basée sur l'emploi de la tobramycine. ....275
Figure VIII-4 : Test de fixation de l'ARN NRS-Tob sur la matrice Tobra-Sepharose. ....276
Figure VIII-5 : Evaluation et amélioration des étapes de rétention et d'élution de l'ARN NRS-Tob sur la matrice Tobra-Sepharose. ....278
Figure VIII-6 : Analyse par expériences de retard sur gel des complexes ribonucléoprotéiques formés après incubation des transcrits Tob (A), NRS-Tob (B) et NRS (C) en présence d'extrait nucléaire de cellules HeLa. ....280
Figure VIII-7 : Formation et purification des complexes ribonucléoprotéiques formés sur l'ARN NRS-Tob par chromatographie d'affinité sur la matrice Tobra-Sepharose. ....281
Figure VIII-8 : Formation et purification des complexes ribonucléoprotéiques sur l'ARN NRS-Tob. ....283
Figure VIII-9 : Analyse des protéines contenues dans les échantillons récupérés après l'ajout de l'extrait nucléaire sur la matrice. ....284
Figure VIII-10 : Analyse par Western blot de la présence des protéines ASF et 9G8 dans les échantillons collectés après l'ajout de l'extrait nucléaire sur la matrice. ....285
Figure VIII-11 : Analyse par expérience de retard sur gel des complexes formés après incubation des transcrits NRS-Tob (A), NRS (B) en présence de protéine ASF/SF2 recombinante purifiée. ....286
Figure VIII-12 : Analyse des ARN contenus dans les échantillons récupérés après l'ajout de l'extrait nucléaire sur la matrice. ....287
Figure VIII-13 : Séquences et structures des différents ARN fixant la protéine de l'enveloppe du phage MS2 (Valegard *et al*., 1997 ; Zhou *et al*., 2002). ....289
Figure VIII-14 : Constructions réalisées pour produire les ARN utilisés dans les expériences de chromatographie d'affinité basées sur l'emploi de la protéine de l'enveloppe du phage MS2. ....290
Figure VIII-15 : Structures cristallographiques des complexes formés entre l'homodimère de la protéine de l'enveloppe du phage MS2 et l'ARN TR (A), la protéine de l'enveloppe du phage MS2 et l'ARN aptamérique obtenu par SELEX (B) et la protéine MBP et le maltose (C) (Valegard *et al*., 1997 ; Convery *et al*., 1998 ; Quiocho *et al*., 1997). ....291

Figure VIII-16 : Production et purification de la protéine MS2-MBP. ....293
Figure VIII-17 : Analyse par expérience de retard sur gel des complexes ribonucléoprotéiques formés après incubation des ARN NRS-MS2 (A), NRS (B) et MS2 (C) en présence de protéine MS2-MBP purifiée. ....295
Figure VIII-18 : Analyse par expériences de retard sur gel des complexes ribonucléoprotéiques formés après incubation des ARN NRS-MS2 (A), NRS (B) et MS2 (C) en présence d'extrait nucléaire de cellules HeLa. ....296
Figure VIII-19 : Etude de la fixation aspécifique possible des protéines de l'extrait nucléaire de cellules HeLa sur la résine amylose. ....299
Figure VIII-20 : Principe de la purification des complexes ribonucléoprotéiques par chromatographie d'affinité reposant sur l'emploi de la protéine de l'enveloppe du phage MS2. ....301
Figure VIII-21 : Purification des complexes ribonucléoprotéiques formés sur les ARN MS2 (B) et NRS-MS2 (C) à un rapport molaire [protéine MS2-MBP]/[ARN] de 20. ....302
Figure VIII-22 : Analyse par SDS-PAGE et spectrométrie de masse des protéines contenues dans les complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 à un rapport molaire [protéine MS2-MBP]/[ARN] de 20. ....304
Figure VIII-23 : Protéines des complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 lors de la première expérience de chromatographie d'affinité. ....305
Tableau VIII-1 : Liste des protéines identifiées par spectrométrie de masse dans les complexes ribonucléoprotéiques assemblés sur les ARN MS2 et NRS-MS2 lors de la première expérience de chromatographie d'affinité. ....308
Figure VIII-24 : Purification des complexes ribonucléoprotéiques formés sur les ARN MS2 (B) et NRS-MS2 à un rapport molaire [protéine MS2-MBP]/[ARN] de 5. ....310
Figure VIII-25 : Analyse par SDS-PAGE suivie d'une analyse par spectrométrie de masse des protéines contenues dans les complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 à un rapport molaire [protéine MS2-MBP]/[ARN] de 5. ....311
Figure VIII-26 : Protéines des complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 lors de la deuxième expérience de chromatographie d'affinité. ....312
Tableau VIII-2 : Liste des protéines identifiées par spectrométrie de masse dans les complexes ribonucléoprotéiques assemblés sur les ARN MS2 et NRS-MS2 lors de la deuxième expérience de chromatographie d'affinité. ....314
Figure VIII-27 : Protéines des complexes ribonucléoprotéiques formés sur l'ARN NRS-MS2 retrouvées lors des deux expériences de chromatographie d'affinité. ....316
Figure VIII-28 : représentation de l'ARN MS2 (A), de l'ARN NRS-MS2 (B) ainsi que des différents éléments NRS tronqués fusionnés à la séquence MS2 (C) NRS 22-MS2, NRS 24-MS2, NRS 26-MS2, NRS Δ76-MS2 et NRS 28-MS2 dans la troisième expérience de chromatographie d'affinité. ....317
Figure VIII-29 : Analyse par expériences de retard sur gel des complexes formés après incubation des différents ARN NRS tronqués fusionnés (A1 à E1) ou non fusionnés (A2 à E2) en présence de quantités croissante de protéine MS2-MBP purifiée. ....319
Figure VIII-30 : Purification des complexes ribonucléoprotéiques formés sur les ARN MS2 (B) et NRS-MS2 (C). ....320
Figure VIII-31 : Purification des complexes ribonucléoprotéiques formés sur les ARN NRS Δ76-MS2 (B) et NRS 22-MS2 (C). ....321
Figure VIII-32 : Purification des complexes ribonucléoprotéiques formés sur les ARN NRS 24-MS2 (B) et NRS 26-MS2 (C). ....322
Figure VIII-33 : Purification des complexes ribonucléoprotéiques formés sur les ARN NRS 28-MS2. ....323
Figure VIII-34 : Fractionnement en gel SDS-PAGE des protéines contenues dans les complexes ribonucléoprotéiques formés sur l'ARN MS2, l'ARN NRS WT-MS2 et les différents ARN NRS-S2 tronqués par chromatographie d'affinité. ....324
Figure VIII-35 : Comparaison des protéines identifiées dans les complexes ribonucléoprotéiques formés sur l'ARN NRS-MS2 dans les trois expériences de chromatographie d'affinité. ....326
Tableau VIII-3 : Identification des protéines par spectrométrie de masse dans les complexes ribonucléoprotéiques assemblés sur l'ARN NRS-MS2 lors des 3 expériences de chromatographie d'affinité . ....334
Tableau VIII-4 : Liste des protéines identifiées par spectrométrie de masse dans les complexes ribonucléoprotéiques assemblés sur l'ARN NRS WT-MS2 et les différents ARN NRS tronqués-MS2 lors de l'expérience de chromatographie d'affinité. La présence de ces protéines dans les complexes spliceosomaux est indiquée dans la partie inférieure du tableau. ....344
Figure VIII-36 : Analyse par Northern blot des snRNA U1, U2 (A), U12, U11 (B) et U4, U6 (C) contenus dans les complexes ribonucléoprotéiques formés sur les différents éléments NRS-MS2 et purifiés par chromatographie d'affinité. ....346
Figure VIII-37 : Analyse par Western blot de certaines protéines SR et hnRNP contenues dans les complexes ribonucléoprotéiques formés sur l'ARN MS2 et les différents ARN NRS-MS2 qui ont été purifiés par chromatographie d'affinité. ....348

Figure IX-1 : Analyse des protections générées sur les séquences de l'élément NRS par la protéine hnRNP A1 en présence de la RNase T1. ....351
Figure IX-2 : Comparaison des sites de fixation de la protéine hnRNP A1 identifiés dans l'élément NRS avec les séquences connues pour fixer cette protéine. ....353
Figure IX-3 : Comparaison des protections obtenues sur l'ARN NRS du virus RSV lors d'expériences d'empreinte en présence de protéine hnRNP A1 avec les sites de fixation des protéines et particules connus de cet ARN. ....354

Figure X-1 : Constructions d'épissage utilisées pour l'étude de l'inhibition de l'épissage par l'élément NRS. ....356
Figure X-2 : Expériences d'épissage *in vitro* avec des pré-ARNm uniformément marqués. ....358
Figure X-3 : Expériences d'épissage *in vitro* et identification des produits d'épissage par RT-PCR. ....360

Figure X-4 : Expériences d'épissage in vitro avec les pré-ARNm C2, C2-NRS WT et C2-NRS 26 réalisées avec les extraits nucléaires N°2626 (A) et N°3783 (B) et analyse des produits d'épissage obtenus par RT-PCR. ....**362**
Figure X-5 : Comparaison par Western blot des quantités de certaines protéines SR et hnRNP contenues dans les 2 lots d'extrait nucléaire N°2626 et N°3783 utilisés lors des tests d'épissage *in vitro*. ....**364**
Figure X-6 : Amplification par RT-PCR des espèces épissées (Sp) d'une part, et des espèces épissées et non épissées (Sp+UnSp) d'autre part, avec deux couples d'oligonucléotides distincts. ....**365**
Figure X-7 : Linéarité de l'amplification par PCR des produits d'épissage. ....**367**
Figure X-8 : Analyse des produits d'épissage du pré-ARNm C2 et des différents pré-ARNm C2-NRS par RT-PCR. ....**369**
Figure X-9 : Reproduction de l'analyse des produits d'épissage des pré-ARNm C2-NRS 26 et C2-NRS 22 par RT-PCR. ....**371**
Figure X-10 : Amplification par RT-PCR des espèces épissées (Sp) et des espèces épissées plus non épissées (Sp+UnSp) lors des expériences d'épissage *in cellulo*. ....**373**
Figure X-11 : L'amplification des produits d'épissage par RT-PCR est linéaire. ....**375**
Figure X-12 : Analyse par RT-PCR des produits d'épissage *in cellulo* des pré-ARNm C2 et C2-NRS Inv et C2-NRS WT et reproduction de l'amplification par RT-PCR. ....**377**
Figure X-13 :Reproductibilité de l'analyse par RT-PCR des produits d'épissage obtenus pour les pré-ARNm C2, C2-NRS WT et C2-NRS Inv, lors de deux transfections. ....**378**

## Discussion et perspectives

Figure XII-1 : Comparaison des structures secondaires théoriques et expérimentales de l'élément NRS du virus RSV et report sur ces structures des sites de fixation des protéines connues. ....**431**
Figure XII-2 :Alignement des séquences protéiques du précurseur Pr76 Gag-Pro codées par la séquence de l'élément NRS des différents virus choisis pour l'étude phylogénétique. ....**434**
Figure XII-3 : Les domaines L du précurseur Pr76 Gag-Pro sont codés par l'élément NRS (Dilley *et al*., 2010). ....**435**
Figure XII-4 : Représentation des structures secondaires théoriques des éléments NRS des virus ALV Strain Tym_S90 RSV (A) et Y53 Sarcoma virus (B) ainsi que de leurs séquences du pseudo site appariées avec le snRNA U1 (Bar *et al*., 2011). ....**437**
Figure XII-5 : Génomes et traduction des protéines à partir des différents ARN viraux des virus ALV Strain Tym_S90 (A) et Y53 Sarcoma virus (B) (d'après Hatai *et al*., 2008 ; Kitamura *et al*., 1982). ....**439**
Figure XII-6 : Représentation de la structure secondaire théorique de l'élément NRS du virus Fujinami Sarcoma virus, de son génome et de la protéine de fusion P90 Gag-fps (d'après Coffin *et al*., 1997). ....**441**
Figure XII-7 : Représentation des éléments NRS tronqués fonctionnels en inhibition de l'épissage *in vitro*. ....**443**
Figure XII-8 : La fixation de la snRNP U1 sur le pseudoexon du gène ATM inhibe son épissage (Dhir *et al*., 2010). ....**445**
Figure XII-9 : Modèle hypothétique du fonctionnement de l'élément NRS 22. ....**448**
Figure XII-10 : Représentation des éléments NRS tronqués non fonctionnels en inhibition de l'épissage *in vitro*. ....**450**
Figure XII-11 : Modèle proposé pour le fonctionnement de l'élément NRS WT. ....**457**

# Abréviations

ψ : Pseudouridine

AEV : Avian Erythroblastosis Virus

ALV : Avian Leukosis Virus

AMV : Avian Myeloblastosis Virus

ARNt : acide ribonucléique de transfert

ASLV : Avian Sarcoma Leukosis Virus

ATP : adénosine triphosphate

BB : Boîte de Branchement

CA: Capside

CEF: Chicken Embryonic Fibroblasts

Cellules HFF : Cellules fibroblastiques humaines

Cellules NIH3T3 : Cellules fibroblastiques de souris

CPSF: Cleavage and Polyadenylation Specificity Factor

CStF: Cleavage Stimulation Factor

CTE: Constitutive Transport Element

DDR: Downstream Direct Repeat

DLS: Dimer Linkage Structure

DR: Direct Repeat

DSE: Downstream Sequence Element

Dsx : Doublesex protein

DUB : enzymes de déubiquitination

EF: Enhancer Factor

EJC: Exon Junction Complex

ELISA: Enzyme-Linked Immunosorbent Assay

Env: Enveloppe

ESCRT: Endosomal Sorting Complex Required for Transport

ESE: Exonic Splicing Enhancer

Facteur C/EBP: CCAAT/Enhancer Binding Protein

Facteur SRF: Serum Response Factor

FGV: Flowl Glioma Virus

Gag: Group specific antigen

Gly: Glycine

HIV: Human immunodeficiency virus

HPRS: Houghton Poultry Research Station

ILV: Intraluminal Vesicles

IN: intégrase

IRES: Internal Ribosome Entry site

ISPE: Intronic Splicing Processing Element

Kb: kilobase

LTR : Long Terminal Repeat

MA : Protéine de Matrice virale

MVB: Multivesicular Body

Myr: myristoylation modification site

NC: Nucléoprotéine

Neo: neomycine

NES: Nuclear Export Signal

NLS: Nuclear Localization Virus

NMD: Nonsense-Mediated Decay

nt(s): nucléotide(s)

ORF: Open Reading Frame

P : Proline

Pb : paire de bases

PBS: Primer Binding Site

PIC: Pre-Integration Complex

Pol : Polymérase

PPT : polypyrimidine tract

PR : Protéase virale

Protéine PAP : Polyadenylate Polymerase

Protéine VBP : Vitellogenin-Binding Protein

R: purine

RAV: Rous Associated Virus

RE: Reticulum Endoplasmique

RPV: Ring-necked Pheasant Virus

RRE: Rev Responsive Element

RSE: Rous Sarcoma Virus Stability Element

RSV, RSV PrC: Rous Sarcoma Virus souche Prague C

RSV-SR: Rous SarcomaVirus Schmidt-Ruppin

RT: Transcriptase inverse

RT-PCR: Reverse-Transcription-Polymerase Chain Reaction

S: Svedberg

S100: Extrait cytoplasmique

Séquence att: attachment site

Séquence R : Séquence répétée

Séquence U3 : Séquence unique en 3'

Séquence U5 : Séquence unique en 5'

SH (domain): Src Homology

Sss: suppressor of src splicing

SU : Glycoprotéine de surface de l'enveloppe

TM: Polyprotéine transmembranaire

TV: Tumor Virus (loci)

TVxR: Tumor Virus Resistant (loci)

TVxS: Tumor Virus Susceptibility (loci)

Ub: Ubiquitine

UDR: Upstream Direct Repeat

uORF: uptream Open Reading Frame

URZAV: Avian Sarcoma Virus URZ

USE: Upstream Sequence Element

UTR: Untranslated Region

VIH : « Virus de l'Immunodéficience Humaine»

VLP : Virus-Like Particles

Y : Tyrosine

*Y* : pyrimidin*e*

# Introduction

# Chapitre I : Epissage des ARN pré-messagers nucléaires chez les eucaryotes supérieurs

L'expression des gènes chez les eucaryotes supérieurs repose sur plusieurs mécanismes moléculaires complexes qui agissent de façon coordonnée, depuis la transcription, la maturation des ARN pré-messagers (pré-ARNm), jusqu'à leur traduction dans le cytoplasme. La réaction d'épissage est un processus de maturation des ARN pré-messagers qui représente une des étapes clés dans la régulation de l'expression des gènes des eucaryotes supérieurs.

L'épissage a lieu dans le noyau après la transcription de l'ARN pré-messager et consiste à en éliminer les séquences non codantes, appelées introns, et joindre entre elles des séquences codantes, appelées exons, pour produire un ARN mature (ARNm). Chez les eucaryotes supérieurs, la majorité des gènes nucléaires codant des protéines sont épissés, ce qui permet de produire plusieurs ARNm à partir d'un même pré-ARNm et donc d'avoir une grande diversité parmi les protéines codées par un même gène (Stamm *et al.*, 2005 ; Xing et Lee, 2006). Cela permet notamment de produire des isoformes protéiques spécifiques d'un tissu ou d'un stade de développement donné.

On peut distinguer deux grandes catégories de réactions d'épissage. La première catégorie est appelée épissage constitutif et concerne les exons qui sont reconnus et inclus systématiquement dans l'ARNm. La seconde catégorie est appelée épissage alternatif et s'applique aux exons qui sont reconnus et inclus de façon variable dans le pré-ARNm. Cette variabilité n'est pas aléatoire et repose sur la combinaison de plusieurs facteurs comme la reconnaissance des sites d'épissage, les séquences *cis*, les facteurs d'épissage agissant en *trans*, la vitesse de transcription ou la nature du tissu (Chen et Manley, 2009). L'étude de l'épissage alternatif consiste à comprendre comment les séquences introniques et exoniques sont reconnues et quels sont les paramètres qui gouvernent la modulation dans les choix des sites d'épissage alternatifs.

## I. La réaction moléculaire de l'épissage

Il existe plusieurs mécanismes différents qui conduisent à l'élimination des introns. En dépit de leurs différences, ces mécanismes sont regroupés sous le terme d'épissage. Dans cette introduction, nous nous intéresserons exclusivement à l'épissage des gènes nucléaires codant les protéines qui sont éliminés par un macro-complexe ribonucléoprotéique appelé spliceosome.

La réaction d'épissage se compose de deux étapes successives de transestérifications qui nécessitent la reconnaissance des sites d'épissage par le spliceosome catalysent la réaction. Ces sites délimitent les exons et des introns : le site donneur ou site 5' se situe à la bordure exon/intron tandis que le site accepteur ou site 3' se situe à la bordure intron/exon (**Figure I-1**). Le point de branchement situé dans l'intron contribue à la reconnaissance du site accepteur en aval.

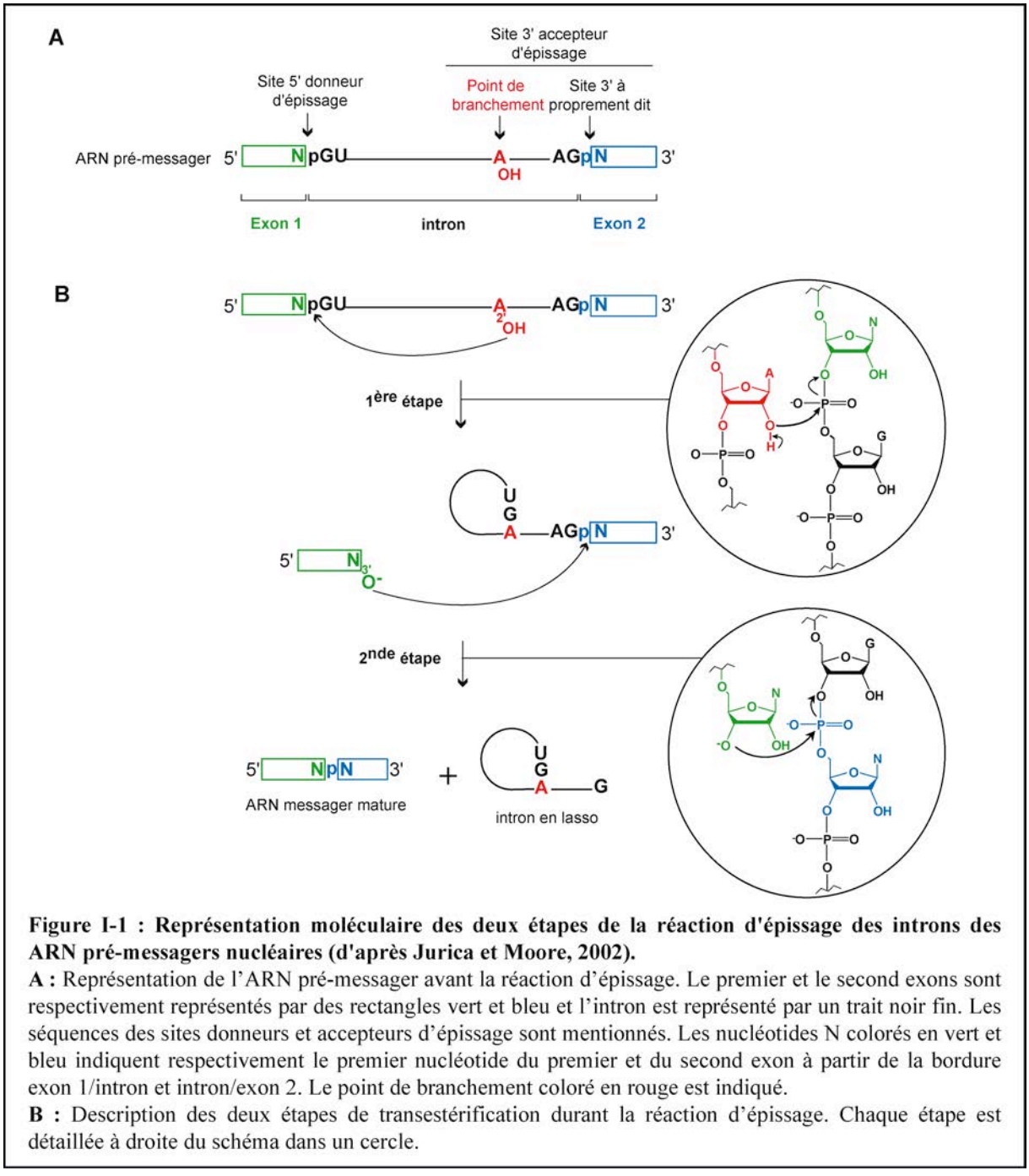



**Figure I-1 : Représentation moléculaire des deux étapes de la réaction d'épissage des introns des ARN pré-messagers nucléaires (d'après Jurica et Moore, 2002).**
**A :** Représentation de l'ARN pré-messager avant la réaction d'épissage. Le premier et le second exons sont respectivement représentés par des rectangles vert et bleu et l'intron est représenté par un trait noir fin. Les séquences des sites donneurs et accepteurs d'épissage sont mentionnés. Les nucléotides N colorés en vert et bleu indiquent respectivement le premier nucléotide du premier et du second exon à partir de la bordure exon 1/intron et intron/exon 2. Le point de branchement coloré en rouge est indiqué.
**B :** Description des deux étapes de transestérification durant la réaction d'épissage. Chaque étape est détaillée à droite du schéma dans un cercle.

La première étape commence par l'attaque nucléophile de l'hydroxyle en 2' du point de branchement sur le phosphate du nucléotide se situant à l'extrémité 5' de l'intron qui fait partie du site donneur d'épissage. Cette réaction produit deux intermédiaires : l'exon 1 avec un groupement hydroxyle libre et une molécule qui forme une structure dite en « lasso » contenant l'exon 2 et l'intron dans laquelle le point de branchement forme 3 liaisons phosphodiesters. Lors de la seconde étape, le groupement hydroxyle en 3' de l'exon 1 attaque le phosphate du nucléotide à l'extrémité 5' de l'exon 2 qui fait partie du site accepteur d'épissage. Cela conduit à la ligation des deux exons et à la libération de l'intron sous forme de lasso (**Figure I-1**).

## II. Les séquences importantes pour la reconnaissance des introns

Chez les eucaryotes supérieurs, les introns sont répertoriés en 2 grandes catégories en fonction du spliceosome qui catalyse leur excision. Les introns dits « majeurs » sont éliminés par le spliceosome majeur ou spliceosome U2-dépendant tandis que les introns dits « mineurs » car numériquement moins nombreux dans le génome, sont éliminés par le spliceosome mineur également appelé spliceosome U12-dépendant.

La comparaison des séquences des sites d'épissage de nombreux introns mineurs et majeurs a permis de définir des règles générales en fonction de la conservation des dinucléotides aux sites 5' et 3'. Les introns majeurs obéissent à la règle GT-AT qui reflète la conservation des dinucléotides GT aux sites 5' et celle de AT aux sites 3' de ces introns. La majorité des introns mineurs, ou U12-dépendants, possèdent les dinucléotides GT-AT mais ont néanmoins conservé pour des raisons historiques l'appellation d'introns AT-AC, bien que cette règle ne s'applique plus à la majorité d'entre eux (Sheth *et al.*, 2006 ; Tarn et Steitz, 1996). Néanmoins, il existe de nombreuses exceptions dans chacune des catégories qui montrent que des sites d'épissage peuvent avoir une séquence dégénérée tout en étant tout de même reconnus par le spliceosome lors de l'épissage.

### II.1. Le site donneur d'épissage ou site 5'

La séquence consensus du site donneur d'épissage majeur est composée des 8 nucléotides CAG/GURAGU (R= purine, / = jonction exon/intron) parmi lesquels le dinucléotide GU en 5' de l'intron est le plus conservé (Sheth *et al.*, 2006) (**Figure I-2**). Cependant, le dinucléotide GU peut être remplacé par le dinucléotide non canonique GC. Chez les eucaryotes supérieurs, la majorité des sites d'épissage donneurs divergent de la séquence consensus et possèdent des

séquences dégénérées. La reconnaissance du site donneur par le spliceosome se traduit par son appariement avec le snRNA U1 (Mount *et al.*, 1983).

Il a longtemps été suggéré que l'utilisation du site donneur lors de l'épissage dépendait entièrement du nombre d'appariements qu'il pouvait établir avec le snRNA U1. Bien que ce critère soit important, la reconnaissance d'un site donneur, et à fortiori d'un site donneur dégénéré, est plus complexe et repose la combinaison de plusieurs paramètres comme les facteurs *trans* et les séquences *cis* régulatrices inhibitrices et activatrices d'épissage ou la structure secondaire de l'ARN (Buratti et Baralle, 2010). Ainsi, l'utilisation d'un site donneur n'est pas seulement dépendante de sa séquence mais aussi de son contexte global, ce qui permet d'avoir une grande flexibilité dans le choix des sites donneurs lors de l'épissage.

Dans le cas des introns mineurs, la séquence des sites 5' définie par les nucléotides ((A/G)UAUCCU) est très conservée (**Figure I-2**).

### II.2. Le site accepteur d'épissage ou site 3'

Les sites accepteurs majeurs sont caractérisés par 3 séquences : le site accepteur à proprement dit, la boîte de branchement qui contient le point de branchement suivie de la séquence polypyrimidique (**Figure I-2**). Cette dernière est en revanche absente des sites accepteurs mineurs.

#### II.2.1. La boîte de branchement

La boîte de branchement est située dans l'intron 18 à 40 nucléotides en amont du site 3' d'épissage (**Figure I-2**). Sa séquence consensus de 7 nucléotides est très dégénérée YNYUR**A**C (Y = pyrimidine) et comporte le point de branchement (adénosine en gras) qui elle, est très conservée mais qui peut être une cytosine ou une uridine dans certains cas. Dans le cas des introns mineurs, cette séquence comporte 8 nucléotides très conservés (UCCUUR**A**C) dont le point de branchement est également une adénosine.

#### II.2.2. La séquence polypyrimidine

La séquence polypyrimidine correspond à un enchaînement de 10 à 20 pyrimidines, qui se situe immédiatement en amont du site 3' (**Figure I-2**). La présence de nombreux résidus uridine par rapport aux cytidines contribue davantage à sa fonction.

#### II.2.3. Le site 3' proprement dit

Le site 3' des introns majeurs est caractérisé par une séquence consensus courte qui est conservée YAG/G (Y = pyrimidine, / = bordure intron/exon) (**Figure I-2**).

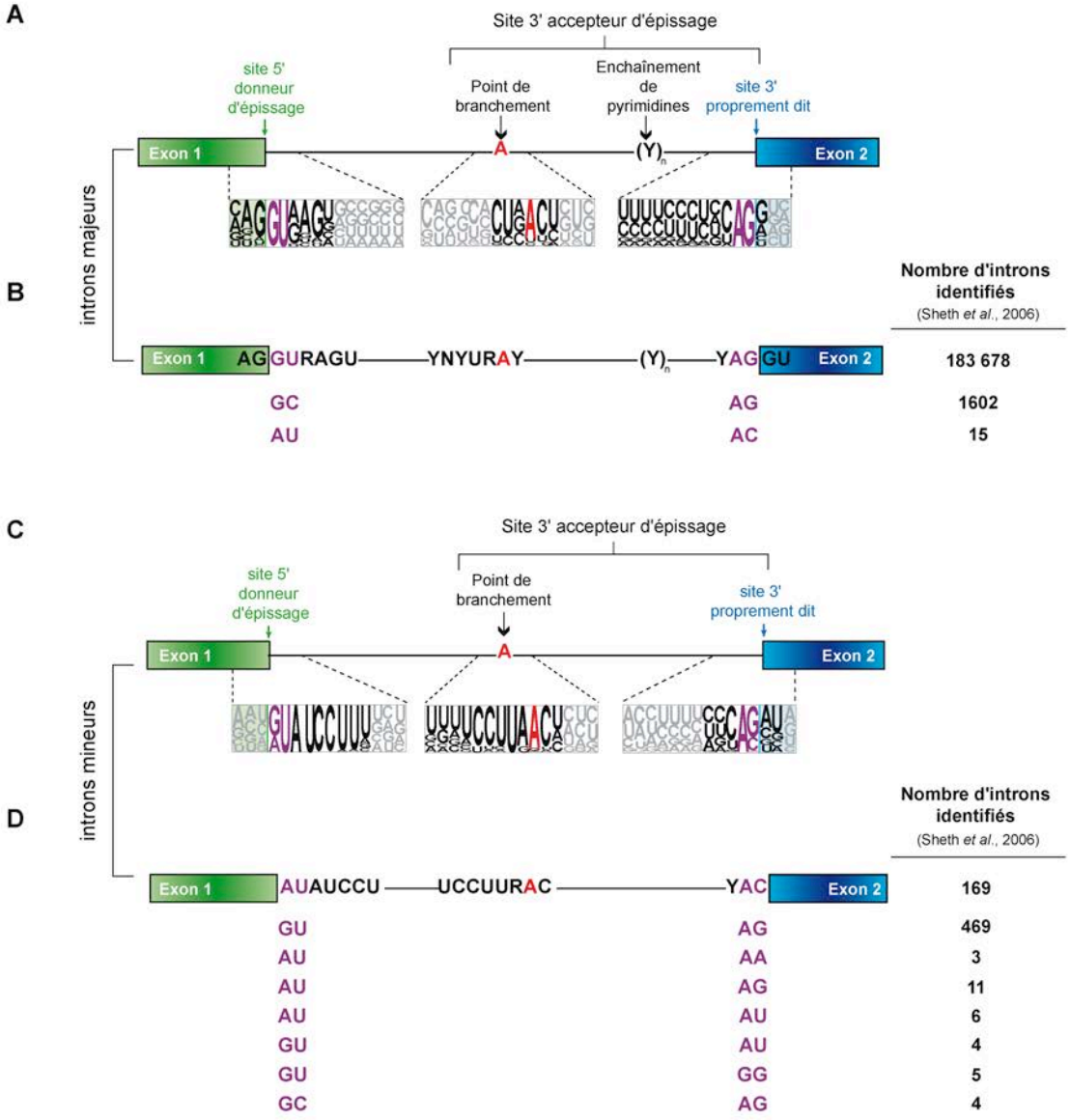



**Figure I-2 : Séquences des sites 5' et 3' des introns U2-dépendants (A, B) et U12-dépendants (C,D) essentielles à la réaction d'épissage.**

**A :** Séquences consensus des introns U2-dépendants chez les eucaryotes supérieurs (d'après Patel et Steitz, 2003).

**B :** Les différents types d'introns U2-dépendants humains. Le nombre de chaque type d'introns identifiés est reporté à droite des différents pré-ARNm (Sheth *et al*., 2006).

**C :** Séquences consensus des introns U12-dépendants chez les eucaryotes supérieurs (d'après Patel et Steitz, 2003).

**D :** Les différents types d'introns U12-dépendants humains. Le nombre de chaque type d'introns identifiés est reporté à droite des différents pré-ARNm (Sheth *et al*., 2006).

Les sites 5' donneur d'épissage, 3' accepteur d'épissage, la séquence polypyrimidique, la boîte de branchement ainsi que le point de branchement sont indiqués. La hauteur des nucléotides des séquences consensus (A, C) représente la fréquence de la présence de ce nucléotide à cette position. Les dinucléotides aux extrémités 5' et 3' de l'intron sont en violet, les nucélotides importants pour la reconnaissance de l'intron sont en noir et les autres nucléotides sont en gris. Le point de branchement est en rouge.

N : nucléotide quelconque, Y : pyrimidine, R : purine.

Les sites 3' des introns mineurs ont des séquences plus variables et leur séquence consensus est YA($^{C}/_{G}$)/N (/ = bordure intron/exon).

## III. Les composants et l'assemblage du spliceosome

L'épissage des introns par le spliceosome est un processus dynamique et séquentiel au cours duquel plusieurs complexes spliceosomaux vont s'assembler successivement autour du pré-ARNm pour catalyser les deux réactions de transestérification en mettant en oeuvre de nombreux réarrangements conformationnels tant au niveau des interactions ARN-ARN, ARN-protéines que protéines-protéines.

Les introns majeurs et mineurs sont respectivement épissés par le spliceosome majeur et le spliceosome mineur qui sont tous les deux des macro-complexes ribonucléoprotéiques. Ils sont chacun constitués de 5 particules appelées snRNP (small nuclear RiboNucleoParticle). Le spliceosome majeur est constitué des snRNP U1, U2, U4/U6 et U5 tandis que le spliceosome mineur est composé des snRNP U11, U12, U4atac/U6atac et U5. En plus de ces particules, le spliceosome comprend un grand nombre de protéines (environ 300) qui ont été identifiées grâce aux récentes approches protéomiques qui ont permis la purification des complexes (Will et Luhrmann, 2011).

### III.1. Les particules snRNP du spliceosome majeur

Chaque particule snRNP est composée d'un snRNA (small nuclear RNA) spécifique sur lequel s'assemblent 7 protéines Sm (B/B', D3, D2, D1, E, F et G) qui sont communes à toutes les particules ainsi qu'un nombre variable de protéines qui lui sont spécifiques.

Après leur transcription dans le noyau, les UsnRNA sont exportés dans le cytoplasme où ils sont pris en charge par le complexe SMN (Survival of Moto Neuron) qui permet l'assemblage des protéines Sm sur le site Sm présent dans chacun des UsnRNA. Après l'ajout de la coiffe tri-méthylée $m^{2,2,7}G$, les UsnRNA sont réimportés dans le noyau où ils séjournent transitoirement dans les corps de Cajal pour y subir les modifications post-transcriptionnelles (pseudouridylations et/ou 2'-O méthylations) et s'associer avec leur protéines spécifiques (Patel et Bellini, 2008). La biogenèse du UsnRNA U6 et U6 atac diffèrent des autres UsnRNA. Leur assemblage a lieu uniquement dans le noyau où leur coiffe est γ-méthylée et leurs modifications post transcriptionnelles ont lieu dans nucléole. Ces UsnRNA ne possèdent pas de site Sm mais une séquence riche en U à l'extrémité 3' sur laquelle s'associent des protéines Sm-like (Patel et Bellini, 2008).

Une fois matures, les snRNP participent aux réactions d'épissage ou bien sont accumulées dans les particules appelées « speckles » qui se situent généralement à proximité des gènes transcrits où elles constituent probablement des sites de stockage pour les snRNP et/ou le fonctionnement des snRNP (Spector et Lamond, 2011).

### III.1.1. La particule snRNP U1

La particule snRNP U1 se compose du snRNA U1 de 165 nucléotides et de 10 protéines : 7 protéines Sm et 3 protéines spécifiques, U1-70K, U1-A et U1-C. L'ARN snRNA U1 forme 4 tiges boucles (SLS 1 à 4) et possède un site Sm (AAUUUGUGG) entre les tiges boucles SLS3 et SLS4 sur lequel s'assemblent les 7 protéines Sm pour former le cœur de la particule (Weber *et al.*, 2010) (**Figure I-3**).

Les protéines U1-70k et U1-A se fixent respectivement sur les tiges boucles SLS1 et SLS2 grâce à leur domaine RRM (RNA Recognition motif) tandis que la protéine U1-C interagit avec les 67 premiers résidus de la protéine U1-70k (Nelissen *et al.*, 1994 ; Newman et Nagai, 2010 ; Urlaub *et al.*, 2000) (**Figure I-3**).

La snRNP U1 fait partie du spliceosome majeur et son rôle est de reconnaître le site donneur d'épissage majeur. Le snRNA U1 possède à son extrémité 5' une séquence très conservée (ACUUACCUG) qui s'apparie avec le site donneur d'épissage (Mount *et al.*, 1983) (**Figure I-3 et I-4**). Le degré de complémentarité entre les deux séquences est l'un des paramètres importants pour la reconnaissance du site donneur (Roca *et al.*, 2005). Le snRNA U1 s'apparie le plus souvent avec la séquence du site donneur allant des positions -3 à +6 mais il peut parfois être étendu aux positions +7 et +8 ce qui favorise la reconnaissance du site donneur chez les eucaryotes supérieurs (Freund *et al.*, 2005 ; Paca *et al.*, 2001).

Il existe également des sites donneurs atypiques appelés sites donneurs « shiftés » avec lesquels le snRNA U1 s'apparie en étant décalé d'un nucléotide en aval du site donneur défini d'après la séquence consensus (**Figure I-4**). Cela a notamment permis d'identifier l'existence de quelques sites donneurs que l'on pensait non fonctionnels ou très faiblement utilisés en épissage en raison du faible nombre d'appariements potentiellement établis de façon canonique avec le snRNA U1 (Roca et Krainer, 2009).

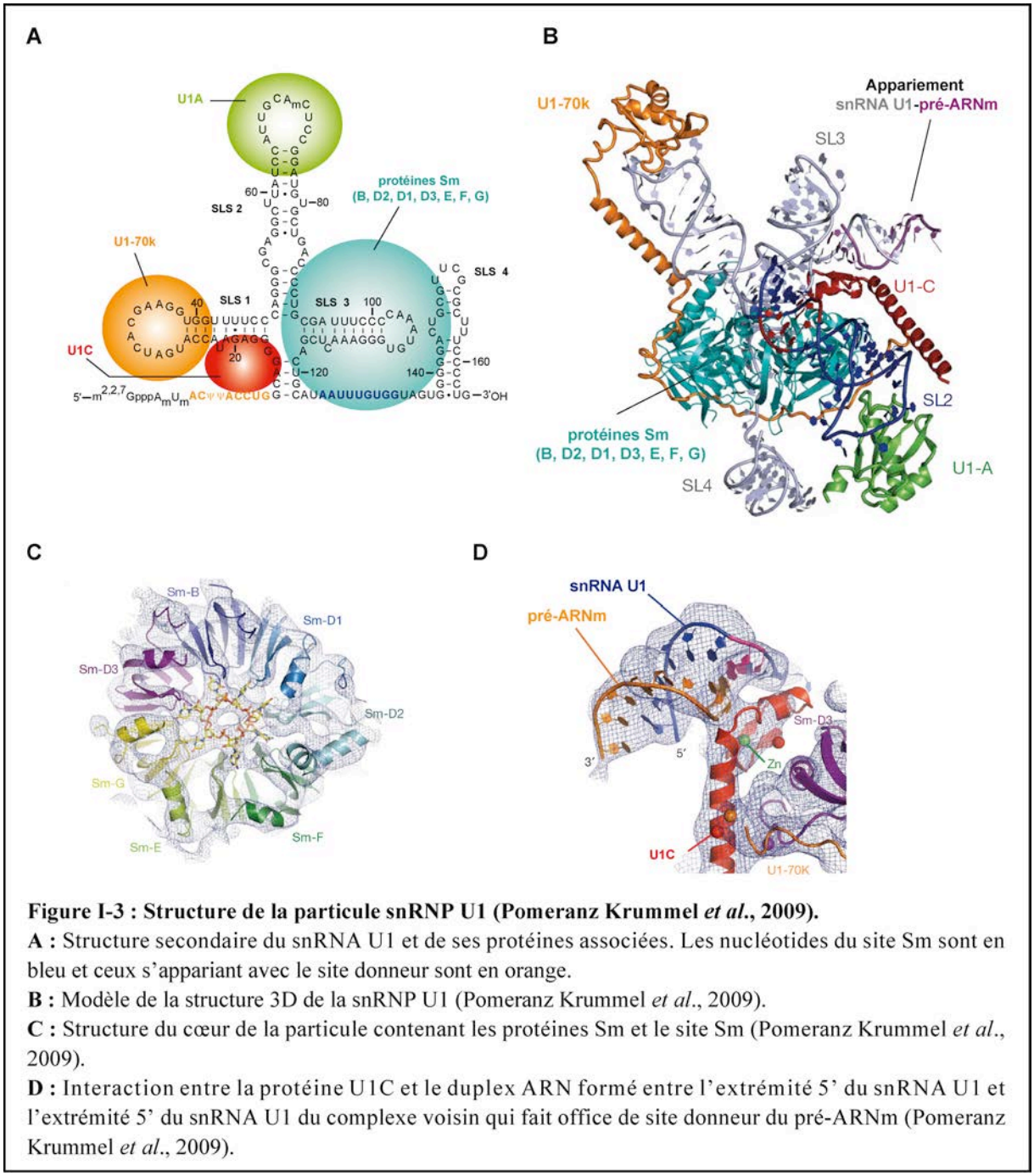



**Figure I-3 : Structure de la particule snRNP U1 (Pomeranz Krummel *et al.*, 2009).**
**A :** Structure secondaire du snRNA U1 et de ses protéines associées. Les nucléotides du site Sm sont en bleu et ceux s'appariant avec le site donneur sont en orange.
**B :** Modèle de la structure 3D de la snRNP U1 (Pomeranz Krummel *et al.*, 2009).
**C :** Structure du cœur de la particule contenant les protéines Sm et le site Sm (Pomeranz Krummel *et al.*, 2009).
**D :** Interaction entre la protéine U1C et le duplex ARN formé entre l'extrémité 5' du snRNA U1 et l'extrémité 5' du snRNA U1 du complexe voisin qui fait office de site donneur du pré-ARNm (Pomeranz Krummel *et al.*, 2009).

Les protéines spécifiques de la particule snRNP U1 jouent également un rôle dans la reconnaissance du site donneur. Ainsi, la snRNP U1 peut être recrutée sur un site donneur grâce à l'interaction de sa protéine U1-70K dont le domaine RRM interagit avec le domaine RRM de la protéine SR ASF/SF2 phosphorylée, préalablement fixée sur une séquence

activatrice à proximité du site donneur (Cho *et al.*, 2011b). La protéine U1-C interagit avec le site donneur et stabilise son appariement avec le snRNA U1 (Du et Rosbash, 2002 ; Pomeranz Krummel *et al.*, 2009). Sa présence est essentielle pour l'assemblage des complexes spliceosomaux lors des étapes précoces de l'épissage (Heinrichs *et al.*, 1990 ; Will *et al.*, 1996). Chez la levure, cette protéine peut fonctionnellement remplacer l'extrémité du snRNA U1 en se fixant directement sur le site donneur (Du et Rosbash, 2002).

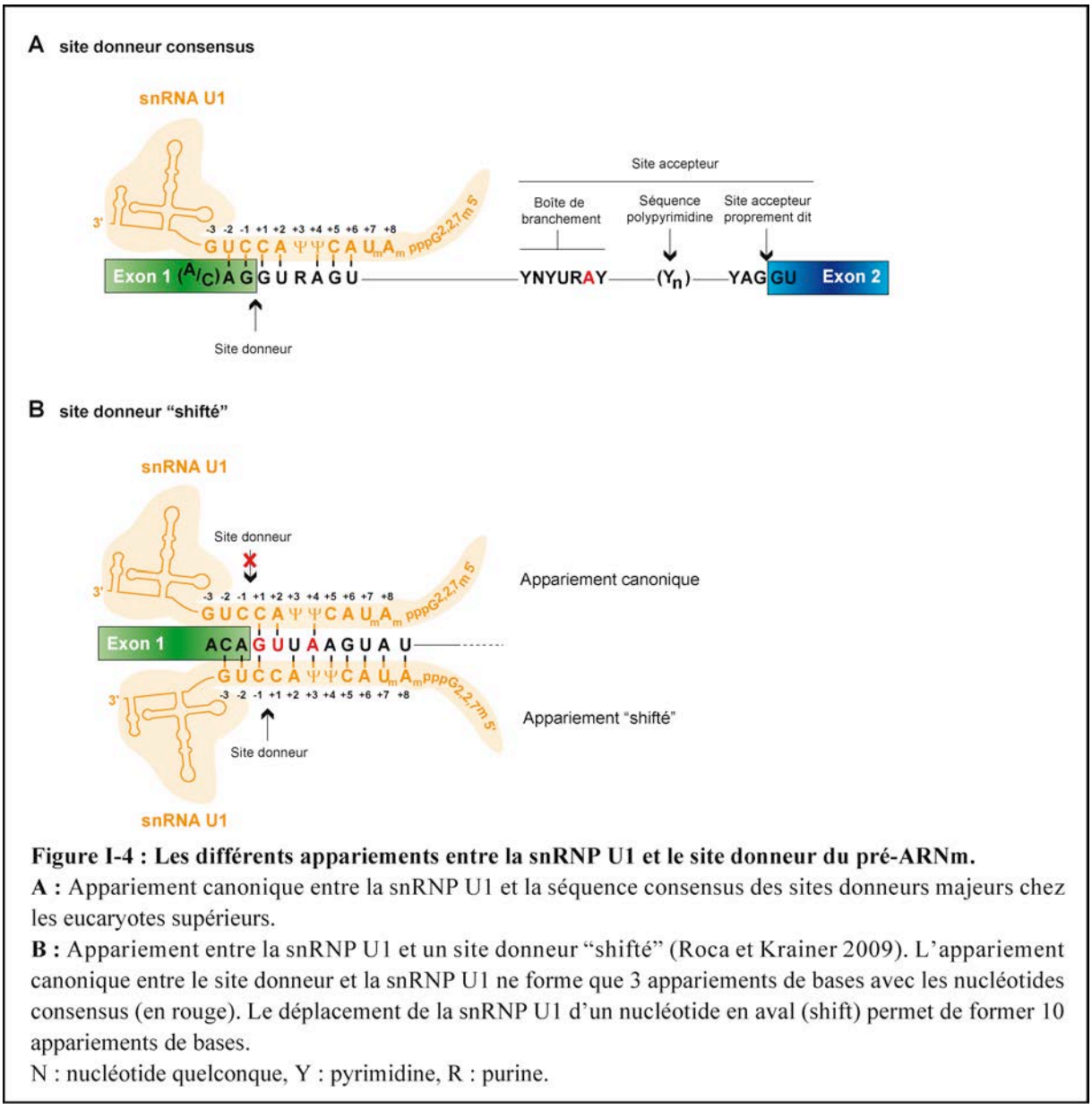



**Figure I-4 : Les différents appariements entre la snRNP U1 et le site donneur du pré-ARNm.**
**A :** Appariement canonique entre la snRNP U1 et la séquence consensus des sites donneurs majeurs chez les eucaryotes supérieurs.
**B :** Appariement entre la snRNP U1 et un site donneur "shifté" (Roca et Krainer 2009). L'appariement canonique entre le site donneur et la snRNP U1 ne forme que 3 appariements de bases avec les nucléotides consensus (en rouge). Le déplacement de la snRNP U1 d'un nucléotide en aval (shift) permet de former 10 appariements de bases.
N : nucléotide quelconque, Y : pyrimidine, R : purine.

### III.1.2. La particule snRNP U2

La particule snRNP U2 se compose du snRNA U2, des 7 protéines Sm et des protéines spécifiques U2A', U2B", et deux sous-unités multi protéiques, SF3a et SF3b (Splicing Factor 3a et 3b). Le snRNA U2 est un ARN de 187 nucléotides qui se structure en 5 tiges boucles annotées I, IIa, IIb, III et IV. L'assemblage des protéines sur cet ARN pour former la particule snRNP U2 17S active en épissage est séquentiel. Les protéines Sm s'associent tout d'abord sur le site Sm du snRNA U2 ainsi que les protéines U2A' et U2B" qui interagissent avec la tige boucle IV pour former la particule 12S (**Figure I-5**). La sous-unité SF3b vient ensuite s'associer à la région 5' du snRNA U2 pour former la particule 15S et recruter ensuite la sous-unité SF3a, pour former la particule 17S (Behrens *et al.*, 1993 ; Brosi *et al.*, 1993).

La sous-unité SF3b est composée de 7 protéines : SF3b155, SF3b145, SF3b130, SF3b49, p14, SF3b14b et SF3b125. La sous-unité SF3a comporte 3 protéines : SF3a60, SF3a66 et SF3a120 (Brosi *et al.*, 1993 ; Will *et al.*, 2001 ; Will *et al.*, 2002).

Lors de l'épissage, le rôle de la snRNP U2 est dans un premier temps de reconnaître la boîte de branchement puis dans un second temps de former le cœur catalytique du spliceosome en s'appariant avec le snRNA U6. Le snRNA U2 comporte dans sa région 5' une séquence conservée, UGUAGUA, qui s'apparie avec la boîte de branchement, à l'exception du point de branchement qui forme alors un bulge, ce qui lui permet de réaliser la première attaque nucléophile lors de l'épissage (Query *et al.*, 1994) (**Figure I-5**). Lors de la formation de cet appariement, les protéines SF3b14a (ou p14) et SF3b155 interagissent et forment une interface contenant une poche dans laquelle se loge le point de branchement (Schellenberg *et al.*, 2011 ; Will *et al.*, 2001) (**Figure I-5**). L'interaction du snRNA U2 avec la boîte de branchement est également stabilisée par l'ensemble des sous-unités SF3a et SF3b qui interagissent avec le pré-ARNm à proximité du point de branchement (Gozani *et al.*, 1996).

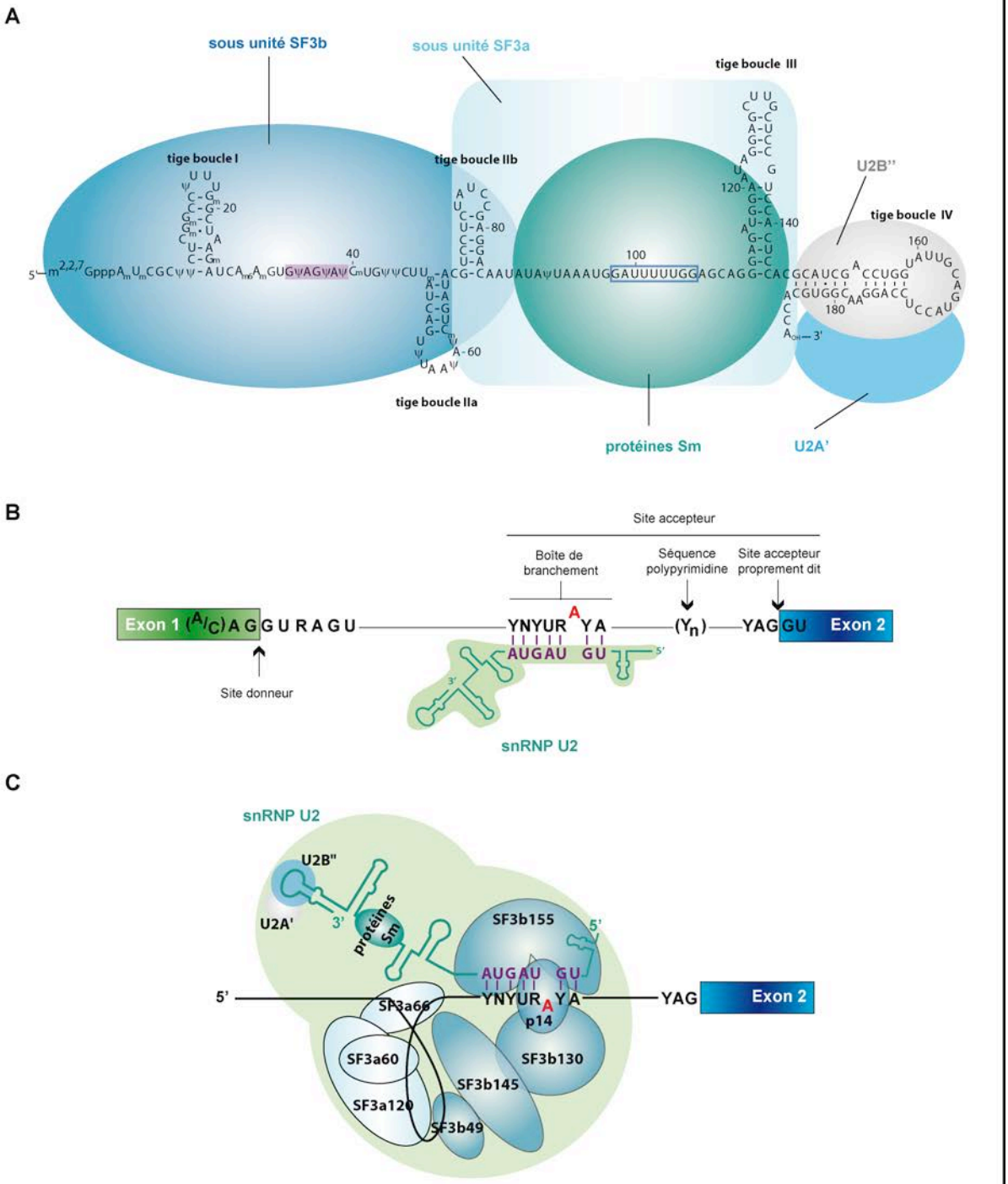



**Figure I-5 : Représentation de la structure de la particule snRNP U2 et de son appariement avec la boîte de branchement.**

**A :** Représentation de la structure secondaire du snRNA U2 et de ses protéines associées. Les nucléotides du site Sm sont encadrés en bleu, ceux s'appariant avec la boîte de branchement sont en violet.

**B :** Appariement entre le snRNA U2 et la séquence consensus de la boîte de branchement d'un pré-ARNm. Le point de branchement est coloré en rouge. N : nucléotide quelconque, Y : pyrimidine, R : purine.

**C :** Modèle proposé pour l'interaction de la particule snRNP U2 avec la boîte de branchement (Wahl *et al.*, 2009, Will *et al.*, 2001).

### III.1.3. La tri-snRNP [U4/U6.U5]

La tri-snRNP U4/U6.U5 est une tri-particule formée par l'association de la di-particule snRNP U4/U6 avec la particule snRNP U5. C'est sous la forme de cette tri-particule que les particules snRNP U4, U5 et U6 rejoignent le spliceosome.

#### III.1.3.1. La di-particule U4/U6 13S

La di-particule peut se former à partir de snRNP U4 et U6 néo-synthétisées ou libres qui viennent d'être recyclées après avoir servi lors d'une réaction d'épissage. Dans les deux cas, les snRNA U4 et U6 sont appariés par deux hélices intermoléculaires qui forment une structure en « Y » et la formation du duplex est catalysée par la protéine hPrp24 (ou p110) qui est associée aux protéines LSm du snRNA U6 (Achsel *et al.*, 1999 ; Mougin *et al.*, 2002 ; Rader et Guthrie, 2002 ; Raghunathan et Guthrie, 1998) (**Figure I-6**).

Ces deux snRNA interagissent ensuite avec leurs protéines spécifiques dont l'assemblage suit une hiérarchie bien précise. La protéine 15,5 K interagit spécifiquement en premier avec le motif en « K-turn » se situant à l'extrémité de sa tige boucle 5' du snRNA U4 (Nottrott *et al.*, 1999 ; Nottrott *et al.*, 2002 ; Schultz *et al.*, 2006b). Cette protéine sert ensuite de plateforme pour recruter la protéine hPrp31 (61K) ainsi que le complexe hétérotrimérique composé des protéines CypH (20K), hPrp4 (60K) et hPrp3 (90K) sur l'hélice II du duplex U4/U6 (Nottrott *et al.*, 2002 ; Teigelkamp *et al.*, 1998) (**Figure I-6**). Le snRNA U4 possède également un site Sm sur lequel viennent s'assembler les 7 protéines Sm tandis que les 7 protéines LSm du snRNA U6 sont regroupées au niveau de l'extrémité 3' riche en U du snRNA U6 (Achsel *et al.*, 1999 ; Sander *et al.*, 2006) (**Figure I-6**).

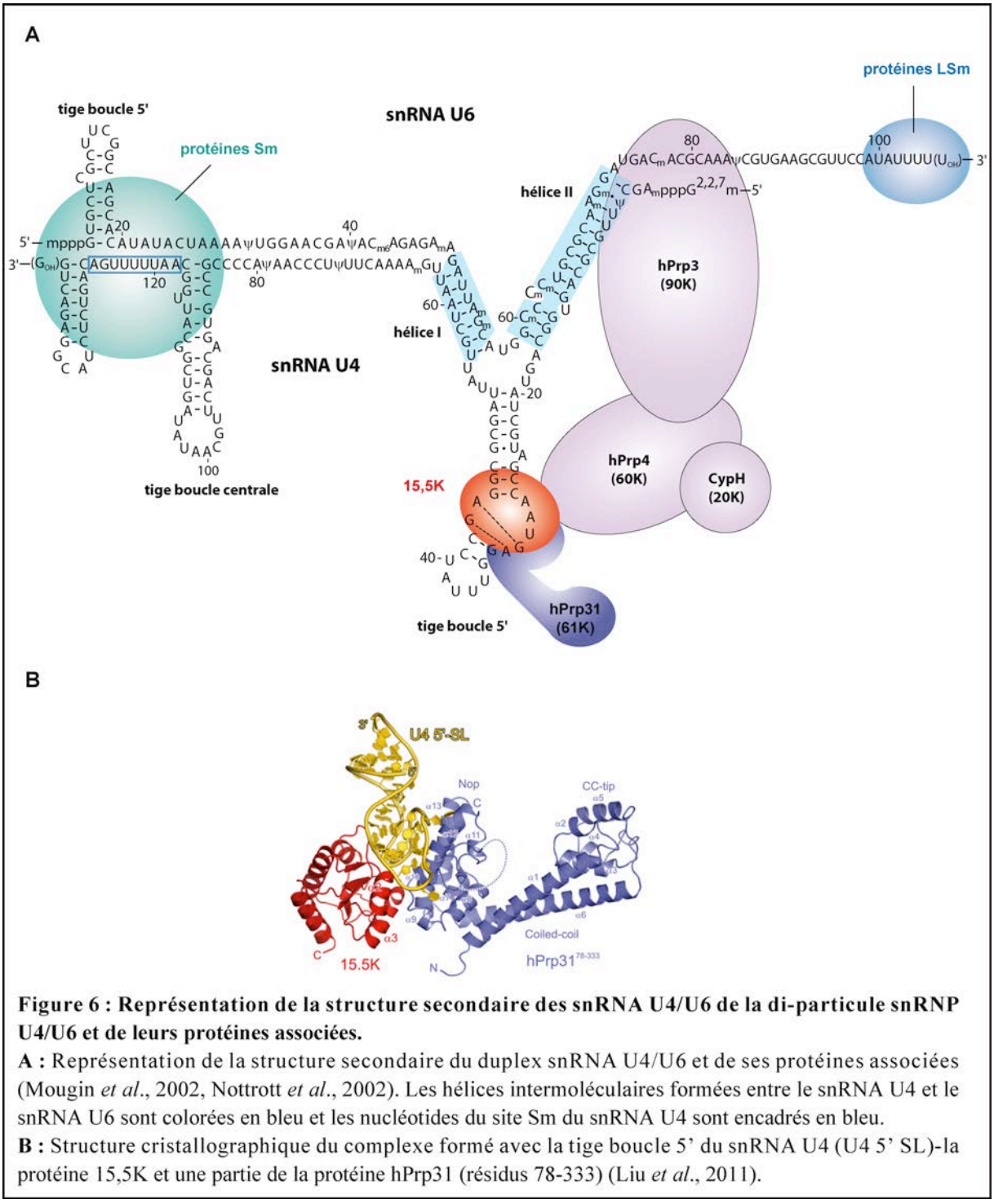



**Figure 6 : Représentation de la structure secondaire des snRNA U4/U6 de la di-particule snRNP U4/U6 et de leurs protéines associées.**

**A :** Représentation de la structure secondaire du duplex snRNA U4/U6 et de ses protéines associées (Mougin *et al*., 2002, Nottrott *et al*., 2002). Les hélices intermoléculaires formées entre le snRNA U4 et le snRNA U6 sont colorées en bleu et les nucléotides du site Sm du snRNA U4 sont encadrés en bleu.

**B :** Structure cristallographique du complexe formé avec la tige boucle 5' du snRNA U4 (U4 5' SL)-la protéine 15,5K et une partie de la protéine hPrp31 (résidus 78-333) (Liu *et al*., 2011).

### III.1.3.2. La particule snRNP U5

La snRNP U5 est la plus volumineuse des particules snRNP et joue un rôle très important lors de l'épissage. Elle facilite la ligation des exons en s'appariant à la bordure 3' du premier exon et la bordure 5' du second exon afin d'aligner les exons lors de la seconde étape de

transestérification (Newman et Norman, 1992). Le snRNA U5 forme deux tiges boucles (I et II) dont la boucle apicale de la tige boucle I, qui a une séquence très conservée, s'apparie avec les séquences exoniques (Sontheimer et Steitz, 1993 ; Wyatt *et al.*, 1992) (**Figure I-7**). La particule snRNP U5 comporte également les 7 protéines Sm et 8 protéines spécifiques : U5-15K (hDib1), U5-40K, U5-52K (CD2BP2), U5-100K (hPrp28), U5-102K (hPrp6), U5-116K (Snu114), U5-200K (hBrr2), U5-220K (hPrp8). Le processus d'assemblage de ces protéines sur le snRNA U5 pour former la particule mature qui pourra intégrer la tri-particule [U4/U6.U5] n'est pas connu en détail. Il semblerait que cet assemblage commence par la fixation des protéines Sm sur le site Sm et l'intégration du complexe composé des protéines Aar2, Snu114 et hPrp8 (Boon *et al.*, 2007). Une des grandes étapes de maturation de la particule consiste à remplacer la protéine hBrr2 à la place de la protéine Aar2 (Boon *et al.*, 2007 ; Gottschalk *et al.*, 2001 ; Weber *et al.*, 2011). Le départ de la protéine Aar2 est provoqué par sa phosphorylation qui rompt son interaction avec la protéine hPrp8 qui interagit ensuite avec la protéine hBrr2 (Weber *et al.*, 2011). Les autres protéines de la snRNP U5 seraient alors intégrées suite à ce recrutement (Boon *et al.*, 2007).

Au sein de la particule mature, la protéine Snu114 interagit avec la région IL2 du snRNA U5 et la protéine hPrp8 se fixe sur la tige boucle I et plus particulièrement la boucle apicale qu'elle peut fonctionnement remplacer lorsque celle-ci est tronquée (Dix *et al.*, 1998 ; O'Keefe *et al.*, 1996 ; Segault *et al.*, 1999 ; Urlaub *et al.*, 2000) (Frazer *et al.*, 2009) (**Figure I-7**).

La protéine hPrp8 est la plus volumineuse des protéines spliceosomales et également la plus conservée au cours de l'évolution (65 % d'identité entre l'Homme et la levure) (Grainger et Beggs, 2005). Cette protéine est essentielle lors l'épissage car elle se trouve au cœur catalytique du spliceosome où elle interagit avec les acteurs majeurs du spliceosome (Grainger et Beggs, 2005 ; Will et Luhrmann, 2011). Ces interactions sont établies principalement avec l'extrémité C-terminale de la protéine hPrp8 qui est composé d'un domaine RNase H like suivi d'un domaine Jab1/MNP (Boon *et al.*, 2006). Il a été suggéré que le domaine RNase-H like interagisse avec les séquences ARN comme le site donneur, et il a été montré que le domaine Jab1/MNP interagit principalement avec les protéines du spliceosome (Pena *et al.*, 2007 ; Pena *et al.*, 2008 ; Reyes *et al.*, 1999 ; Ritchie *et al.*, 2008). Le domaine Jab1/MNP est notamment très important car il permet à la protéine hPrp8 d'interagir au sein de la particule snRNP U5 avec la protéine hBrr2 et la protéine Snu114 pour

former un complexe très stable (Achsel *et al.*, 1998 ; Liu *et al.*, 2006 ; van Nues et Beggs, 2001).

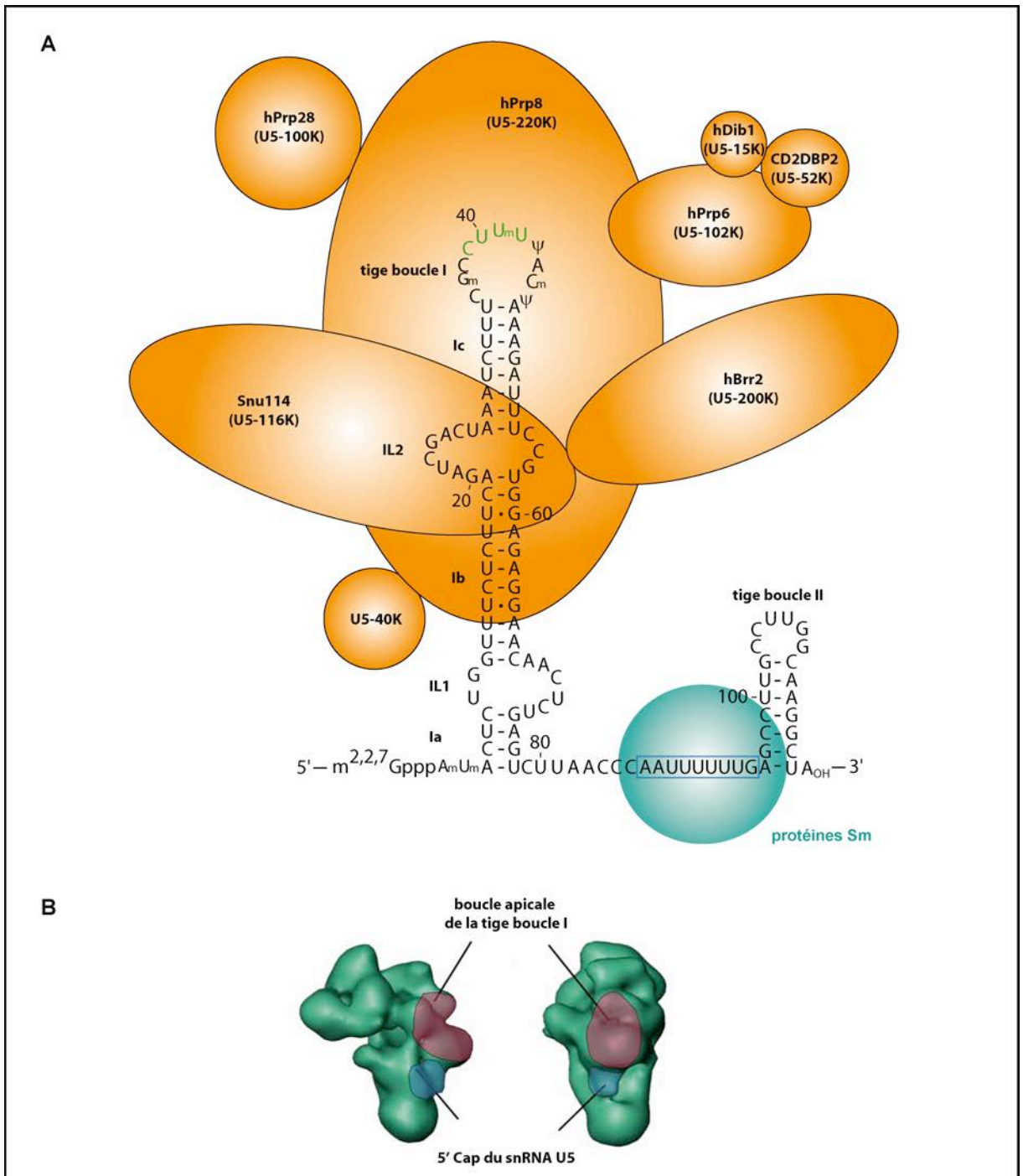



**Figure I-7 : Représentation de la structure secondaire du snRNA U5 et de ses protéines associées.**
**A :** Représentation de la structure secondaire du snRNA U5 (Krol *et al.*, 1981) et de ses protéines associées. Les nucélotides en vert s'apparient avec les bordures exoniques lors de l'épissage. Seules les protéines hPrp8 (U5-220K), Snu114 (U5-116K) ont été démontrées comme interagissant directement avec le snRNA U5. Les autres protéines ont été placées en fonction des interactions établies entre chacune d'entre elles.
**B :** Structure de la particule snRNP U5 résolue en microscopie électronique (Sander *et al.*, 2006). Les positions de l'extrémité 5' ainsi que de la boucle apicale de la tige boucle 1 du snRNA U5 sont indiquées.

Les interactions établies entre ces 3 protéines ont été largement étudiées en raison de leur rôle essentiel dans l'assemblage, l'activation et le remodelage du cœur catalytique du spliceosome tout au long de l'épissage (Achsel *et al.*, 1998 ; Will et Luhrmann, 2011). La protéine hBrr2 est l'hélicase qui désapparie le duplex formé entre les snRNA U4 et U6 qui est nécessaire à la mise en place du spliceosome catalytiquement actif (Raghunathan et Guthrie, 1998). L'activité de cette hélicase est régulée et dépend des interactions établies avec les protéines Snu114 et hPrp8 (Maeder *et al.*, 2009 ; Small *et al.*, 2006). Ces interactions sont abolies lorsque certains des résidus du domaine Jab1/MNP de la protéine hPrp8 sont mutés (Pena *et al.*, 2007). Dans ce cas, cela induit un défaut d'épissage certainement dû à un défaut de maturation/du recyclage de la particule snRNP U5 et est à l'origine de la maladie *retinitis pigmentosa* chez l'homme qui est une dégénérescence de la rétine (Grainger et Beggs, 2005 ; Weber *et al.*, 2011).

### III.1.3.3. La tri-snRNP [U4/U6.U5]

La tri-snRNP [U4/U6.U5] est formée par l'assemblage de la di-snRNP U4/U6 et de la snRNP U5. Cet assemblage s'accompagne du départ de la protéine CD2BP2 (U5-52K) de la snRNP U5 et de l'incorporation de 3 protéines spécifiques de la tri-particule : les protéines hSnu66 (110K), hSad1 (65K) et 27K (Laggerbauer *et al.*, 2005 ; Makarova *et al.*, 2002 ; Makarova *et al.*, 2001). (Fetzer *et al.*, 1997) (**Figure I-8**). Ces protéines sont qualifiées de protéines SR-like car elles possèdent toutes un domaine RS caractéristique des protéines SR. Par ce domaine, elles peuvent établir des interactions protéine-protéine avec les protéines SR qui permettent le recrutement de la tri-snRNP au sein du spliceosome (Makarova *et al.*, 2001). Pour maintenir la cohésion de la tri-particule [U4/U6.U5], de nombreuses interactions protéine-protéine sont établies entre la snRNP U4/U6 et la snRNP U5, comme la protéine hPrp6 (U5-102K) qui interagit avec une protéine spécifique de la di-particule U4/U6, la protéine hPrp31 (61K) (Liu *et al.*, 2006 ; Makarova *et al.*, 2002). La structure de la tri-snRNP humaine et celle de la levure *S. Cerevisiae* ont été résolues par microscopie électronique (Hacker *et al.*, 2008 ; Sander *et al.*, 2006) (**Figure I-8**). La tri-snRNP adopte une structure en forme d'«Y», constituée d'un corps principal, qui contient la snRNP U5 et d'un bras, contenant la snRNP U4/U6, qui est relié au corps par une région linker (Hacker *et al.*, 2008) (**Figure I-8**). Deux conformations de la tri-snRNP ont été identifiées : l'une fermée, avec le bras collé au corps de la particule et une autre ouverte, avec le bras espacé (Hacker *et al.*, 2008) (**Figure I-8**). Il a été suggéré que ces deux conformations puissent réguler l'activité de

l'hélicase hBrr2 qui ne peut être en contact avec le duplex U4/U6, qu'elle doit désapparier lors de l'épissage, uniquement dans la conformation fermée.

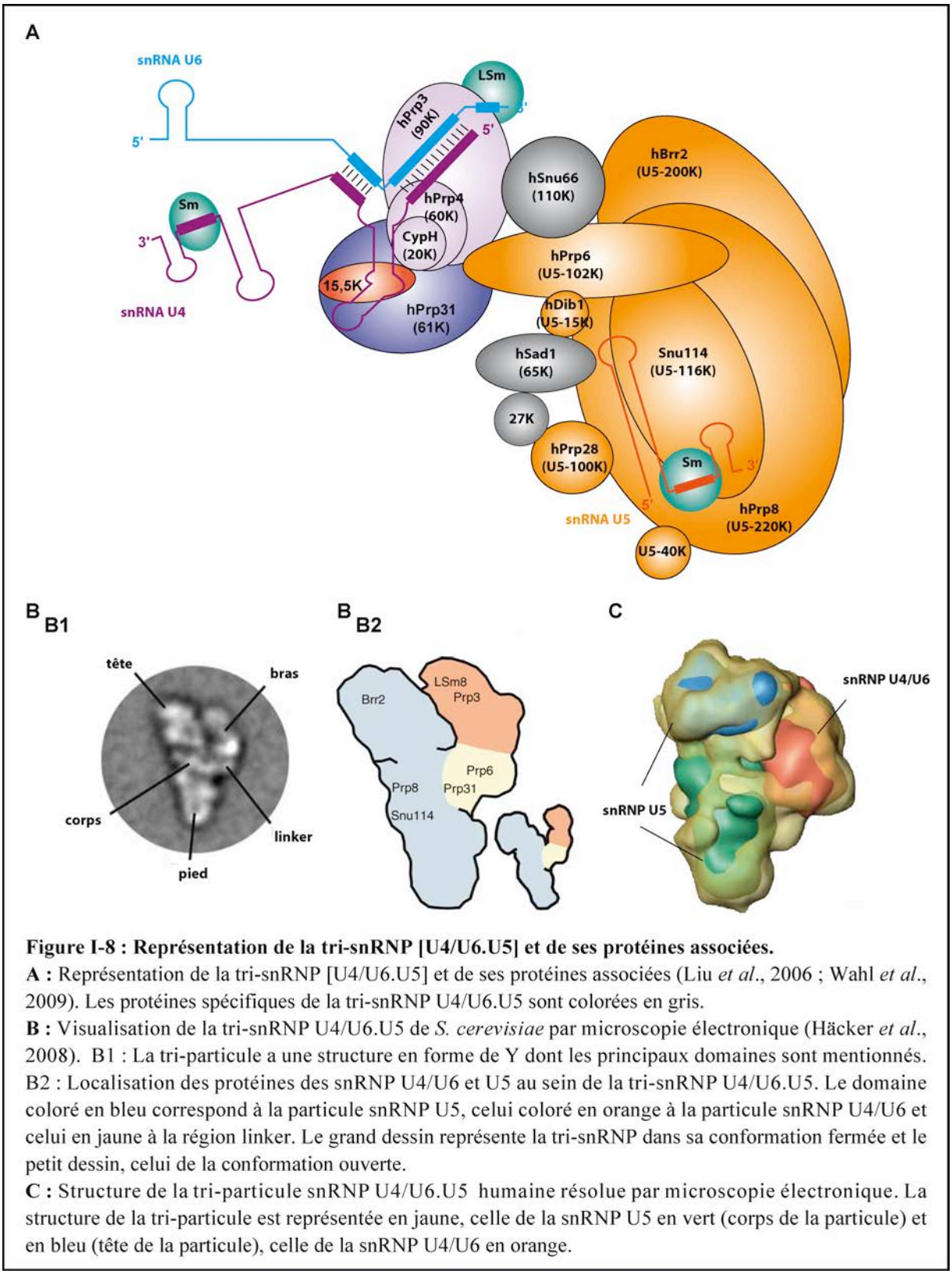



**Figure I-8 : Représentation de la tri-snRNP [U4/U6.U5] et de ses protéines associées.**
**A :** Représentation de la tri-snRNP [U4/U6.U5] et de ses protéines associées (Liu *et al*., 2006 ; Wahl *et al*., 2009). Les protéines spécifiques de la tri-snRNP U4/U6.U5 sont colorées en gris.
**B :** Visualisation de la tri-snRNP U4/U6.U5 de *S. cerevisiae* par microscopie électronique (Häcker *et al*., 2008). B1 : La tri-particule a une structure en forme de Y dont les principaux domaines sont mentionnés. B2 : Localisation des protéines des snRNP U4/U6 et U5 au sein de la tri-snRNP U4/U6.U5. Le domaine coloré en bleu correspond à la particule snRNP U5, celui coloré en orange à la particule snRNP U4/U6 et celui en jaune à la région linker. Le grand dessin représente la tri-snRNP dans sa conformation fermée et le petit dessin, celui de la conformation ouverte.
**C :** Structure de la tri-particule snRNP U4/U6.U5 humaine résolue par microscopie électronique. La structure de la tri-particule est représentée en jaune, celle de la snRNP U5 en vert (corps de la particule) et en bleu (tête de la particule), celle de la snRNP U4/U6 en orange.

### III.2. Les particules snRNP du spliceosome mineur

#### III.2.1. La di-snRNP U11/U12 (18S)

La snRNP U11 et U12 sont les particules du spliceosome mineur fonctionnellement équivalentes aux particules snRNP U1 et U2 du spliceosome majeur. Dans la cellule, et contrairement aux snRNP U1 et U2, les snRNP U11 et U12 sont majoritairement associées l'une à l'autre au sein de la di-snRNP U11/U12 (18S) et ne sont que minoritairement retrouvées sous forme de particules distinctes (Wassarman et Steitz, 1992). Lors de l'épissage, le site donneur et le point de branchement sont respectivement reconnus par les snRNP U11 et U12 qui s'y apparient de façon concomitante sous la forme de la di-particule (Frilander et Steitz, 1999).

Les snRNA U11 et U12 divergent des séquences des snRNA U1 et U2 mais adoptent en revanche une structure secondaire similaire (Schneider *et al.*, 2004 ; Sikand et Shukla, 2011 ; Tarn *et al.*, 1995) (**Figure I-9**). Ils comportent tous les deux des sites Sm sur lesquels les 7 protéines Sm identiques à celles du spliceosome majeur s'assemblent (Will *et al.*, 1999) (**Figure I-9**). En plus des protéines Sm, la snRNP U11 comporte quartes autres protéines, 25K, 35K, 48K et 59K, qui lui sont spécifiques et n'ont pas d'équivalents dans la snRNP U1 (Will *et al.*, 2004 ; Will *et al.*, 1999). La snRNP U12 est composée de 3 protéines spécifiques, 65K, 31K et 20K et du facteur SF3b (SF3b155, SF3b145, SF3b130, SF3b49, p14, SF3b14b et SF3b10) qui est également présent dans la snRNP U2 du spliceosome majeur (Will *et al.*, 2004 ; Will *et al.*, 2001 ; Will *et al.*, 1999) (**Figure I-9**). Il est supposé que le facteur SF3b remplisse la même fonction dans le spliceosome mineur et majeur, à savoir la reconnaissance et l'interaction avec le point de branchement (Golas *et al.*, 2005 ; Will *et al.*, 2001).

L'organisation structurale de chacune des snRNP U11 et U12 et de la di-snRNP U11-U12 et le rôle de chacune des protéines spécifiques de cette particule ne sont pas bien connus. La protéine U11-48K est nécessaire à la reconnaissance du site donneur avec lequel elle interagit directement (Turunen *et al.*, 2008). Il a été suggéré que cette interaction implique le domaine en doigt de zinc de la protéine U11-48K et stabilise l'appariement du snRNA U11 au site donneur, comme cela a été démontré dans le cas de la protéine U1-C de la snRNP U1 qui possède également un domaine en doigt de zinc CHHC (Muto *et al.*, 2004 ; Tidow *et al.*, 2009). La protéine U11-48K établit également des interactions protéine-protéine via son domaine N-terminal avec la protéine U11-59K (Turunen *et al.*, 2008) (**Figure I-9**).

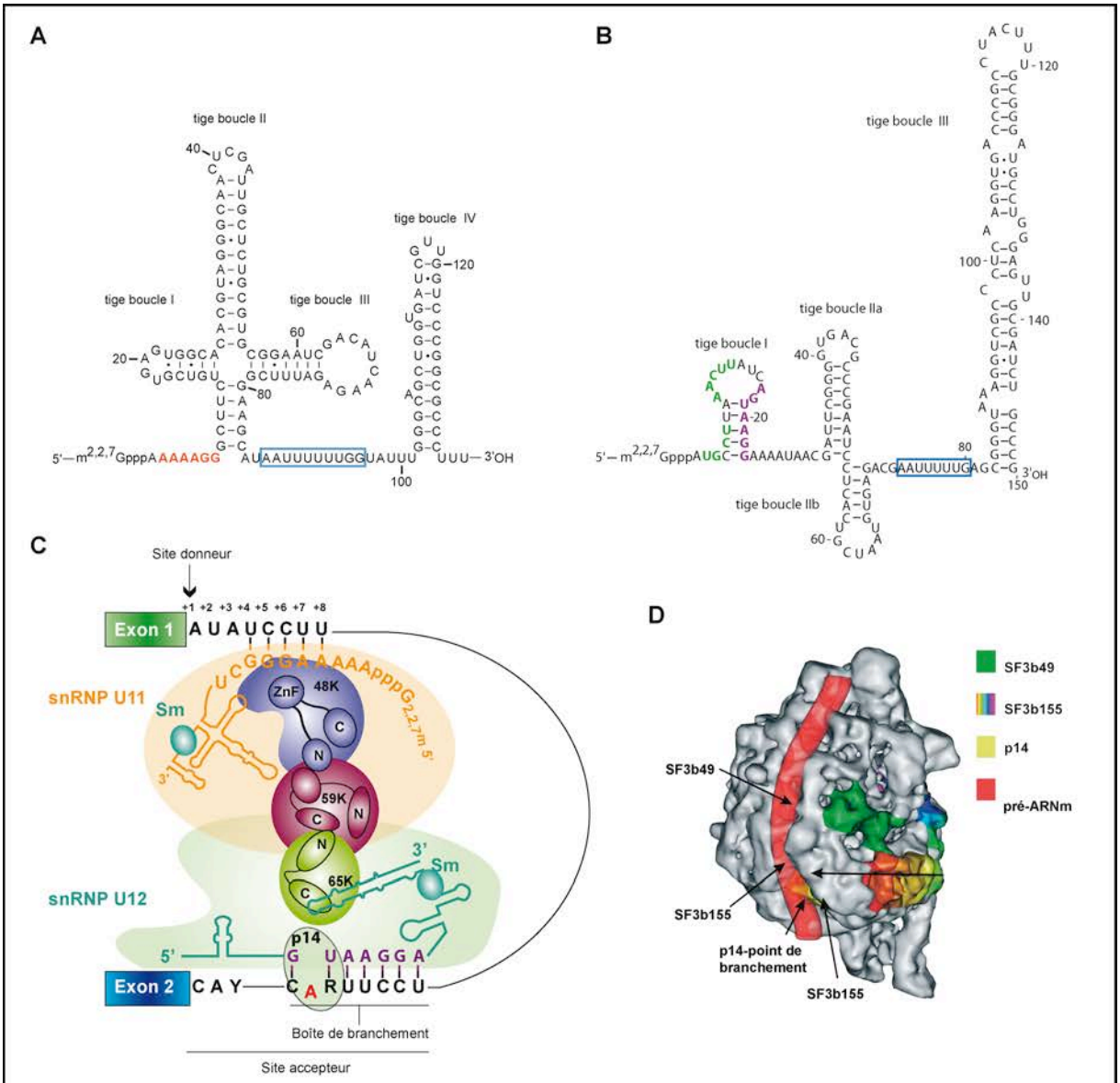



**Figure I-9 : Représentation de la structure secondaire des snRNA U11 et U12 et de la reconnaissance des sites donneur et accepteur mineurs par la di-snRNP U11/U12.**

**A :** Structure secondaire du snRNA U11 (Patel et Steitz, 2003). Le site Sm est encadré en bleu, les nucléotides s'appariant avec la séquence du site donneur sont en orange.

**B :** Structure secondaire du snRNA U12 (Sikand et Shukla, 2011). Le site Sm est encadré en bleu, les nucléotides s'appariant avec la séquence de la boîte de branchement sont en violet, ceux impliqués dans les appariements avec le snRNA U6 atac sont en vert.

**C :** Reconnaissance du site donneur et du site accepteur par la di-snRNP U11/U12 (Tidow *et al.*, 2009). Le snRNA U11 s'apparie avec le site donneur et cet appariement serait stabilisé par le domaine en doigt de zinc de la protéine U11-48K (Tidow *et al.*, 2009). Le snRNA U12 s'apparie avec la séquence de la boîte de branchement (Frilander et Steitz, 1999) sur laquelle se fixerait également la protéine p14 du facteur SF3b (Will et al., 2001 ; Golas *et al.*, 2005). Au sein de la di-snRNP U11/U12, l'extrémité C-terminale et le domaine central riche en arginines de la protéine U11-59K interagissent respectivement avec l'extrémité N-terminale de la protéine U12-65K et l'extrémité N-terminale de la protéine U11-48K (Benecke *et al.*, 2005 ; Turunen *et al.*, 2008).

**D :** Résolution par microscopie électronique de la structure de la di-snRNP U11/U12 et modélisation du pré-ARNm sur cette structure (Golas *et al.*, 2005). Les positions des protéines du facteur SF3b et du pré-ARNm sont colorées comme indiqué sur la figure. Les positions des pontages UV obtenus avec les différentes protéines ont été reportées sur la structure et sont indiquées par des flèches.

La snRNP U11 se compose également de la protéine U11-35K qui comporte un domaine RRM et une région riche en glycines qui sont similaires à ceux de la protéine U1-70K (Will *et al.*, 1999). Cela suggère que les deux protéines soient fonctionnellement homologues, bien que cela n'ait pas été démontré (Will *et al.*, 1999).

La protéine U12-65K possède deux domaines RRM homologues à ceux des protéines U1A et U2B" et interagit avec la tige boucle SLIII du snRNA U12 (Benecke *et al.*, 2005) (**Figure I-9**). Cette protéine interagit également avec la protéine U11-59K, ce qui suggère qu'elle participe à la cohésion des deux particules snRNP U11 et U12 au sein de la di-snRNP U11/U12 (Benecke *et al.*, 2005) (**Figure I-9**).

### III.2.2. La tri-snRNP [U4atac/U6atac.U5]

Dans le spliceosome mineur, les snRNA U4atac et snRNA U6atac sont les équivalents des snRNA U4 et U6 du spliceosome majeur tandis que la snRNP U5 est en revanche commune aux deux spliceosomes (Patel et Steitz, 2003). Lors de l'épissage, la tri-snRNP [U4atac/U6atac.U5] rempli les mêmes fonctions que la tri-snRNP [U4/U6.U5].

#### III.2.2.1. La di-snRNP U4atac/U6atac

La di-snRNP U4atac/U6atac est composée des snRNA U4atac et U6atac qui forment un duplex dont la structure est similaire à celle du duplex formé entre les snRNA U4 et U6 (Padgett et Shukla, 2002) (**Figure I-10**). Le snRNA U6atac possède à son extrémité 3' une séquence poly(U) qui fixe les protéines LSm ainsi qu'une région composée de deux tiges boucles qui sont essentielles à son adressage dans la tri-snRNP [U4atac/U6atac.U5] (Dietrich *et al.*, 2009). Le snRNA U6atac interagit de façon transitoire avec le facteur hPrp24 (p110) qui catalyse la formation du duplex avec le snRNA U4atac et qui sert également à son recyclage après la réaction d'épissage (Damianov *et al.*, 2004). Les protéines s'associant avec le duplex snRNA U4atac/U6atac sont les mêmes que celles de la di-snRNP U4/U6 et leur assemblage répond à la même hiérarchie (Nottrott *et al.*, 2002 ; Schneider *et al.*, 2002). Comme le snRNA U4, le snRNA U4atac possède dans sa tige boucle 5' un motif en « K-turn » sur lequel se fixe la protéine 15,5K (Nottrott *et al.*, 2002). Cette fixation permet ensuite le recrutement de la protéine hPrp31 (61K) puis du complexe CypH (20K), hPrp4 (60K) et hPrp3 (90K) (Nottrott *et al.*, 2002 ; Schneider *et al.*, 2002 ; Schultz *et al.*, 2006a) (**Figure I-10**). La fixation de la protéine 15,5K sur le snRNA U4atac dépend du motif en « K-turn » et représente une étape essentielle pour la formation de la di-snRNP U4atac/U6atac et ultérieurement de la tri-snRNP [U4atac/U6atac.U5] (Shukla *et al.*, 2002). Chez l'homme, il a

récemment été démontré que les mutations des nucléotides très conservés aux positions 50, 51 et 53 dans la tige boucle 5' du snRNA U4atac sont à l'origine de la maladie du syndrome de Taybi-Linder (TALS) (Edery *et al.*, 2011) (**Figure I-10**). Il a été suggéré que ces mutations provoquent une perte de la structure en « K turn », ce qui engendre un défaut d'assemblage de la di-snRNP U4atac/U6atac et par conséquent de la [U4atac/U6atac.U5]. Cela se traduit par une rétention des introns mineurs qui ne sont pas ou peu épissés (Edery *et al.*, 2011). Ce défaut d'épissage peut être en revanche restauré *in cellulo* en présence de snRNA U4atac sauvage (He *et al.*, 2011).

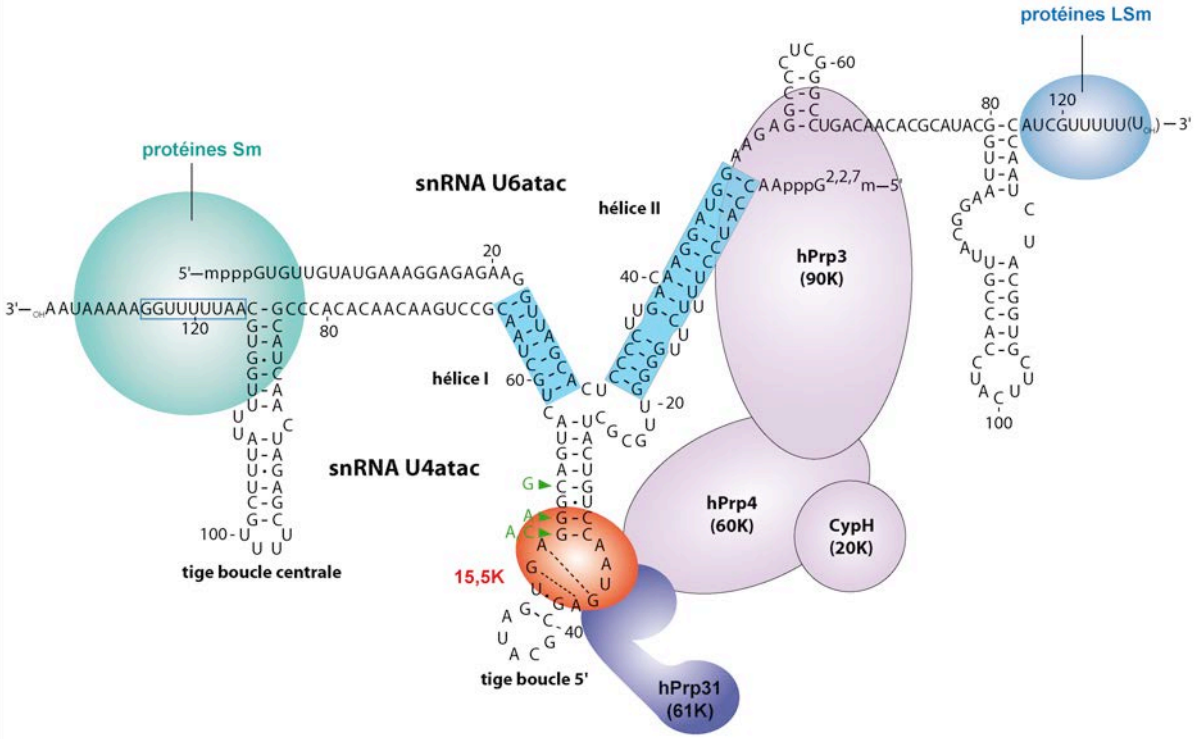



**Figure I-10 : Représentation de la structure secondaire des snRNA U4atac/U6atac de la di-particule snRNP U4atac/U6atac et de leurs protéines associées.**
Représentation de la structure secondaire du duplex snRNA U4atac/U6atac et de ses protéines associées (Padgett et Shukla, 2002 ; Nottrott *et al.*, 2002). Les hélices intermoléculaires formées entre le snRNA U4atac et le snRNA U6atac sont colorées en bleu et les nucléotides du site Sm du snRNA U4atac sont encadrés en bleu. Les mutations de la tige boucle 5' du snRNA U4atac à l'origine du syndrome de Taybi-Linder sont indiquées en vert (Edery *et al.*, 2011).

### III.3. Assemblage des complexes spliceosomaux

#### III.3.1. Généralité (Will et Luhrmann, 2011)

L'assemblage du spliceosome autour de l'intron du pré-ARNm se fait de manière séquentielle. Les études de l'assemblage des complexes spliceosomaux sur le pré-ARNm réalisées *in vitro* ont permis de mettre en évidence la formation successive de 6 complexes : les complexes H, E, A, B, B* et C. La formation séquentielle de ces complexes suit une dynamique selon laquelle chaque étape est accompagnée du départ ou de l'arrivée de particules snRNP et de protéines associées de manière stable ou transitoire durant toute la réaction d'épissage.

L'assemblage du spliceosome débute par la reconnaissance de l'intron qui comprend la reconnaissance du site donneur par la snRNP U1 et celle du site accepteur par la snRNP U2 (**Figure I-11**). Ces deux étapes ont lieu au sein de complexes spliceosomaux appelés E et A, respectivement. La tri-snRNP U4/U6.U5 intègre ensuite le complexe spliceosomal pour former le complexe B (**Figure I-11**). A cette étape, les snRNP vont suivre de nombreux réarrangements conformationnels sous l'action des hélicases du spliceosomes. Cela va conduire successivement au départ des snRNP U1 et U4 du spliceosome, aux appariements intermoléculaires des snRNA U2 et U6 pour former le complexe Bact puis le complexe B* au sein duquel la première étape de transestérification a lieu (**Figure I-11**). Le complexe spliceosomal C alors obtenu catalyse ensuite la deuxième réaction de transestérification, ce qui conduit au relargage de l'ARNm et des particules snRNP qui sont alors recyclées. La réaction d'épissage des introns mineurs par le spliceosome mineur est similaire à celle des introns majeurs, à l'exception des étapes d'assemblage précoces où les sites donneur et accepteur sont simultanément reconnus par la di-particule snRNP U11-U12 (**Figure I-11**).

Les étapes précoces d'assemblage du spliceosome autour de l'intron sont des étapes critiques de l'épissage car c'est d'elles dont dépendent la reconnaissance et le choix des sites qui vont être épissés ultérieurement. Comme nous le verrons dans la seconde partie de cette introduction, ces étapes sont soumises à de nombreuses régulations. Nous allons dans un premier temps décrire ces étapes ainsi que les protéines autres que les snRNP qui y sont impliquées.

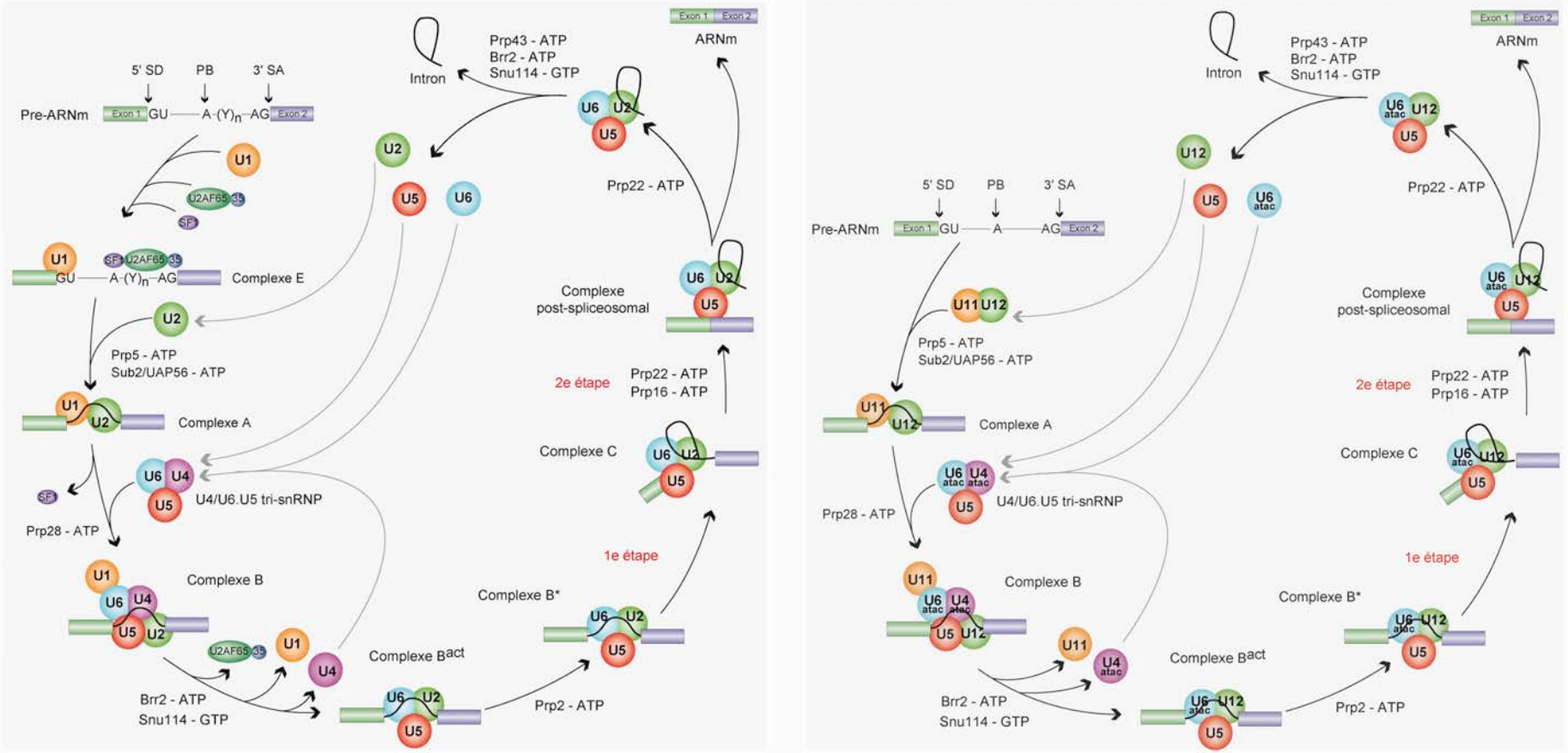



**Figure I-11 : Assemblage des spliceosomes majeurs et mineurs (D'après Will et Lührmann, 2010, Patel et Steitz, 2003).**
**A :** Assemblage du spliceosome majeur. Les exons sont représentés par des rectanges et l'intron par une ligne. Le site donneur (5' SD), le site accepetur (3' SA), le point de branchement (A) et la séquence polypyrimidine ($Y_n$) sont représentés. Les particules snRNP sont représentées par des sphères. Le nom de chaque complexe spliceosomal ainsi que le nom des hélicases importantes pour les changements conformationels impliqués dans chaque étape sont précisés. La première étape et la deuxième étape de transestérification sont colorées en rouge.
**B :** Assemblage du spliceosome mineur. Les snRNP U11, U12, U4atac et U6atac sont représentées selon le même code couleur que leur homologue fonctionnel U1, U2, U4 et U6 du spliceosome majeur. La légende est identique à la partie A).

### III.3.2. Définition des exons et des introns et facteurs protéiques participant à cette définition

La reconnaissance des sites d'épissage par le spliceosome dépend de la force intrinsèque de ces sites mais aussi de l'architecture des introns et des exons et plus précisément de leur taille (Fox-Walsh *et al.*, 2005). Lorsque la taille de l'intron est comprise entre 200-250 nucléotides, le spliceosome s'assemble sur les sites d'épissage situés aux extrémités de l'intron, ce mode de reconnaissance s'appelle la définition de l'intron. En revanche, lorsque la taille de l'intron est supérieure à 250 nucléotides, le spliceosome s'assemble sur les sites d'épissage situés aux extrémités de l'exon, on parle alors de définition de l'exon (Berget, 1995) (**Figure I-12**). Ce mode de reconnaissance est principalement utilisé chez les mammifères dont les pré-ARNm contiennent majoritairement des exons courts, de 120 nucléotides en moyenne, qui sont séparés par des introns dont la taille est supérieure à 200 nucléotides (Deutsch et Long, 1999).

Lors de la définition de l'intron, l'assemblage du spliceosome débute par la reconnaissance du site donneur situé à la bordure exon/intron sur lequel se fixe la snRNP U1. Cette étape est suivie de la reconnaissance du site accepteur situé à la bordure intron/exon ainsi que la séquence polypyrimidine auxquels se fixent respectivement les sous-unités U2AF35 et U2AF65 du facteur d'épissage U2AF (U2 snRNP Auxiliary Factor). La snRNP U2 reconnaît ensuite la boîte de branchement avec laquelle elle s'apparie (**Figure I-12**). La même hiérarchie d'assemblage du spliceosome est mise en place lors de la définition de l'exon. Souvent, cet assemblage nécessite, en plus, l'intervention d'autres protéines associées au spliceosome, les protéines SR, qui favorisent et stabilisent la fixation des facteurs sur le site donneur et le site accepteur de l'exon (Boukis *et al.*, 2004 ; Bourgeois *et al.*, 1999 ; Wu et Maniatis, 1993). Ces protéines se fixent sur des séquences exoniques appelées ESE (Exonic Splicing Enhancer) et établissent des interactions protéines-protéines de façon à former un pont de part et d'autre de l'exon qui contribue à la cohésion du complexe. Ces interactions sont établies entre les protéines SR et aussi des protéines appelées SR like comme la protéine U1-70K de la snRNP U1 et la sous-unité U2AF35 du facteur U2AF (Cho *et al.*, 2011b ; Graveley *et al.*, 2001) (**Figure I-12**). Ce pont peut également être établi de la même façon lors de la définition de l'intron avec les mêmes facteurs fixés sur les sites donneurs et accepteurs situés de part et d'autre de l'intron (Hertel et Maniatis, 1999 ; Wu et Maniatis, 1993) (**Figure I-12**).

L'assemblage du spliceosome selon les définitions des introns ou des exons ne s'applique pas pour le premier ou le dernier exon du pré-ARNm. Dans le cas du premier exon, la

reconnaissance de l'exon est réalisée par l'intermédiaire des protéines CBP20 et CBP80 du complexe CBC (Cap Binding Complex) associé à la coiffe qui interagissent et favorisent la fixation de la snRNP U1 sur le site donneur (Lewis *et al.*, 1996) (**Figure I-12**). Dans le cas du dernier exon, la reconnaissance de l'exon est facilitée par le complexe protéique réalisant la polyadénylation et notamment la protéine PAP (PolyA Polymérase) (Vagner *et al.*, 2000) (**Figure I-12**). Cette protéine interagit avec le signal de polyadénylation et recrute le facteur U2AF65 sur le site accepteur de l'exon en amont (Vagner *et al.*, 2000).

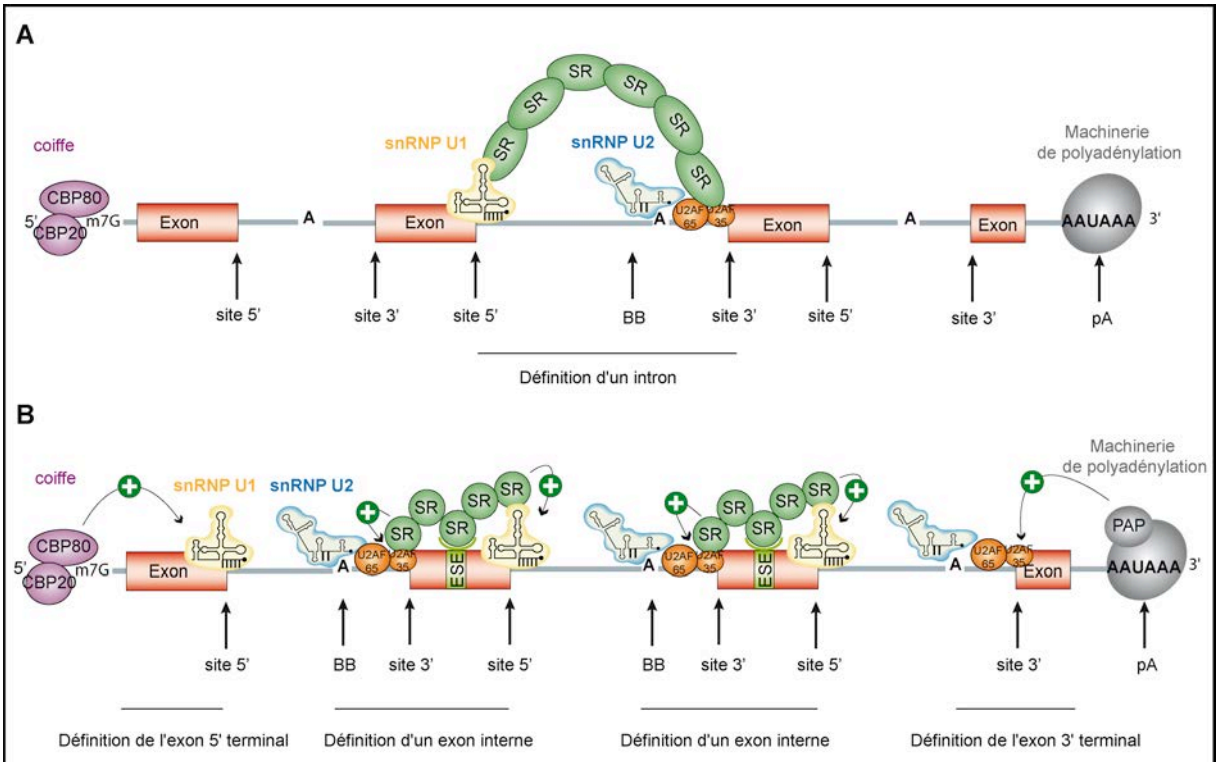



**Figure I-12 : Définition de l'intron et de l'exon d'un ARN pré-messager au cours des étapes précoces de l'assemblage du spliceosome.**
**A :** Définition d'un intron. Les snRNP U1 et U2 ainsi que le facteur U2AF se fixent respectivement sur le dite donneur et le site accepteur d'un même intron. Ces fixations peuvent être favorisées par les protéines SR qui interagissent entre elles et avec la protéine SR like U1-70K et les facteurs U2AF présents à chacune des extrémités de l'intron pour former un pont.
**B :** Définition d'un exon. Les protéines SR se fixent à la séquence exonique ESE et recrutent sur le même exon la snRNP U1 sur le site donneur en aval et les facteurs U2AF sur le site accepteur et la séquence polypyrimidine en amont.

La définition des exons et des introns est une étape essentielle de l'épissage. Les protéines SR ou SR like comme le facteur U2AF y jouent un rôle essentiel car ils contribuent à la reconnaissance des sites donneur et accepteur par le spliceosome. Comme nous le verrons dans le second chapitre, ces protéines exercent, en plus de leur rôle dans l'assemblage du spliceosome à cette étape, un rôle de régulation dans le choix des sites qui seront utilisé lors de l'épissage alternatif.

### III.3.2.2. Les protéines SR et leur rôle dans la reconnaissance du site donneur

Comme nous venons de l'aborder, les protéines SR sont très importantes lors des étapes précoces d'assemblage du spliceosome. Ces protéines portent le nom éponyme du domaine protéique les caractérisant qui est un domaine RS riche en enchaînement Sérine (S)-Arginine (R). Ces protéines sont structurellement très conservées et ont été classées selon ce critère en deux grandes familles : les protéines SR (dites « classiques ») et les protéines SR like. Les protéines SR regroupent les protéines possédant à leur extrémité N-terminale un ou deux domaines de liaison à l'ARN de type RRM (RNA Recognition Motif) suivis d'un domaine RS d'au moins 50 acides aminés composé de plus de 40% d'enchaînements de résidus arginine-sérine ou sérine-arginine (Manley et Krainer, 2010 ; Zahler *et al.*, 1992). Chez l'Homme, 12 protéines SR ont été identifiées et ont été renommées d'après une nouvelle nomenclature sous le nom SRSF (SR Splicing Factor) numérotées de 1 à 12 (**Figure I-13**). Les protéines SR like regroupent les protéines qui possèdent un domaine RS mais dont l'organisation structurale diffère des protéines SR (**Figure I-13**).

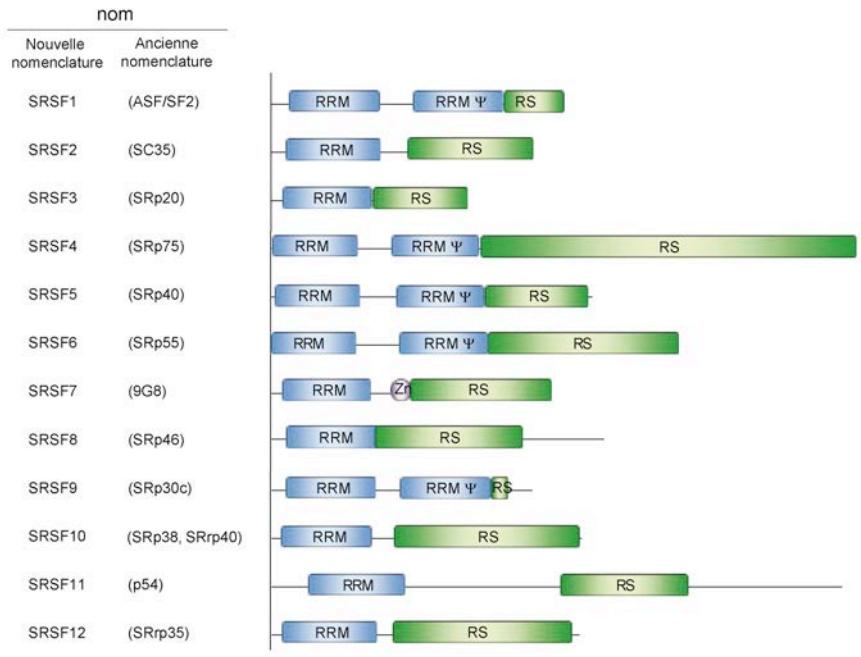



**Figure I-13 : Organisation structurale des protéines SR humaines (Manley et Krainer, 2010 ; Graveley, 2000).**
RRM : RNA Recognition Motif classique, RRM Ψ : RRM de séquence dégénérée, RS : domaine riche en résidus arginines (R) et sérines (S), Zn : doigt de zinc, DEXD/H : motif ARN hélicase, PPI : peptidyl-prolyl *cis*/*trans* isomérase.

#### III.3.2.2.1. Le domaine RRM des protéines SR

La famille des protéines SR est divisée en deux sous-groupes selon le nombre et la séquence du domaine RRM présent (Manley et Krainer, 2010). Le premier regroupe les protéines SR

ayant un domaine RRM classique qui possèdent les deux séquences signatures RNP très conservées caractéristiques du domaine RRM (**Figure I-13**). Le second regroupe les protéines SR ayant un premier RRM classique suivi d'un second domaine appelé RRM-Ψ dont les séquences sont très dégénérées par rapport aux séquences signatures et qui contient en plus un heptapeptide très conservé (**Figure I-13**). La protéine 9G8 possède en plus de son domaine RRM, un motif en doigt de zinc de type CCH qui contribue à la reconnaissance de l'ARN (Cavaloc *et al.*, 1999 ; Cavaloc *et al.*, 1994).

Le domaine RRM permet aux protéines SR de se fixer à l'ARN mais aussi dans certains cas d'interagir avec d'autres domaines RRM comme cela a été montré pour la protéine ASF/SF2 et la protéine SR like U1-70K (Cho *et al.*, 2011b). D'un point de vue structural, seulement 3 résidus du domaine RRM de la protéine SRp20 contactent directement l'ARN sur une longueur de 4 nucléotides ce qui est insuffisant pour avoir une reconnaissance spécifique de la séquence ARN (Hargous *et al.*, 2006). Cela explique pourquoi la séquence consensus de fixation définie pour une protéine SR donnée est souvent dégénérée et également reconnue par d'autres protéines SR (Liu *et al.*, 1998 ; Tacke *et al.*, 1997). Néanmoins, les protéines SR n'ont pas la même spécificité de fixation à l'ARN et exercent des activités différentes lors de l'épissage (Tacke et Manley, 1995, 1999).

La longueur de la séquence ARN en contact avec la protéine peut être plus longue et atteindre 10 nucléotides lorsque la protéine contient 2 domaines RRM, comme la protéine ASF/SF2 (Cho *et al.*, 2011a). Cette fixation nécessite la présence de la région flexible riche en résidus glycine séparant les 2 RRM (Cho *et al.*, 2011a). Après la fixation du premier RRM sur l'ARN, cette région permet la fixation du second domaine RRM sur l'ARN de façon coopérative (Cho *et al.*, 2011a).

#### III.3.2.2.2. Le domaine RS des protéines SR

Les protéines SR interagissent entre elles grâce aux interactions établies entre chacun de leur domaine RS (Wu et Maniatis, 1993). Ce domaine leur permet également d'interagir avec d'autres protéines du spliceosome qui possèdent également un domaine RS, les protéines SR-like, comme la sous-unité 35 kDa de U2AF (Wu et Maniatis, 1993) (**Figure I-12**). Les domaines RS des protéines ASF/SF2, SC35, SRp20 et SRp40 sont fonctionnellement équivalents lorsqu'ils sont interchangés entre chacune des protéines en présence de leur domaine RRM respectif (Chandler *et al.*, 1997 ; van Der Houven Van Oordt *et al.*, 2000). Cependant, la contribution du domaine RS de nombreuses protéines SR lors de l'activation de l'épissage n'est pas équivalent et dépend du nombre d'enchaînements de résidus RS qu'il

comporte (Graveley *et al.*, 1998). Contrairement au domaine RRM, le domaine RS d'une protéine SR comme la protéine ASF/SF2 peut être tronqué sans que cela n'affecte le taux d'épissage *in vitro* d'un pré-ARNm à condition que celui-ci possède des sites donneur et accepteur forts (Zhu et Krainer, 2000).

Le domaine RS de certaines protéines interagit également avec le pré-ARNm lors des différentes étapes de la réaction d'épissage (Shen et Green, 2007) (**Figure I-14**).

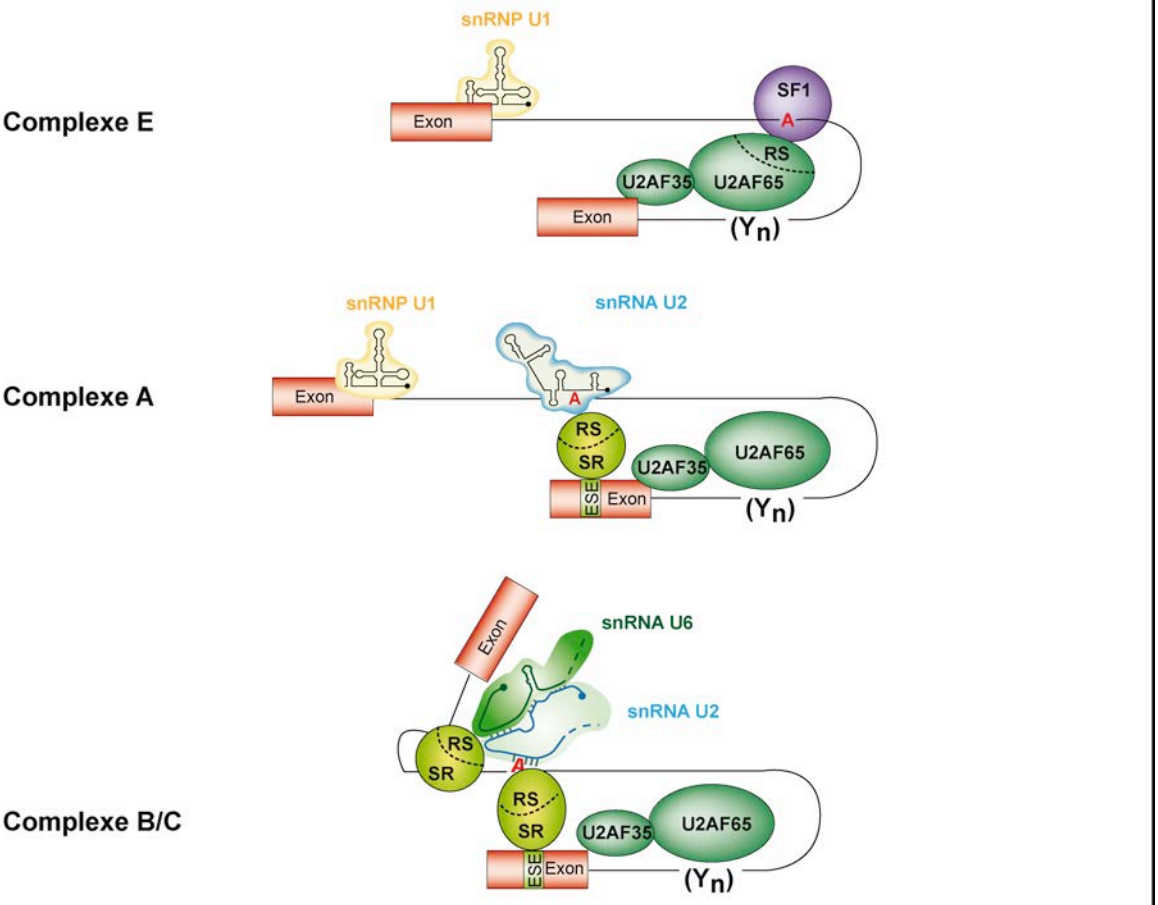



**Figure I-14 : Interactions entre les domaines RS et l'ARN pré-messager au cours de l'épissage (D'après Shen et Green, 2004, 2007).**
Le domaine RS du facteur U2AF65 interagit avec le point de branchement dans le complexe E. Il est ensuite remplacé par le domaine RS d'une protéine liée à une séquence ESE dans le complexe A. Le snRNA U6 est stabilisé sur le site donneur en intergissant avec une protéine SR.

Lors de l'assemblage du spliceosome, le point de branchement est tout d'abord en contact avec le domaine RS du facteur U2AF65 fixé préalablement sur la séquence polypyrimidine se trouvant en aval (Shen et Green, 2004 ; Shen *et al.*, 2004). L'appariement de la séquence du point de branchement avec la snRNP U2 rompt cette interaction au profit de la formation d'une seconde, avec cette fois le domaine RS d'une protéine SR fixée sur une séquence ESE

de l'exon (Shen et Green, 2004 ; Shen *et al.*, 2004) (**Figure I-14**). Les séquences en aval et en amont du site 5' sont respectivement en contact avec le domaine RS d'une protéine SR au cours de la première et de la seconde étape d'épissage (Shen et Green, 2004, 2007). Il a été suggéré que ces interactions stabilisent la fixation des snRNP U6 et U5 à ces séquences (Shen et Green, 2007).

**La phosphorylation des domaines RS :**

Les interactions que le domaine RS établit avec l'ARN ou les domaines RS d'autres protéines impliquent des résidus sérines phosphorylés (Cho *et al.*, 2011b ; Shen *et al.*, 2004 ; Xiao et Manley, 1997). Cette phosphorylation régule l'activité des protéines SR lors de l'épissage ce qui implique une forte régulation de l'activité des kinases et des phosphatases responsables des cycles de la phosphorylation/déphosphorylation de ce domaine (Shi et Manley, 2007 ; Shin et Manley, 2004). La protéine SRp38 est l'exemple type de la régulation exercée par la phosphorylation sur l'activité d'une protéine SR. En condition normale, cette protéine est phosphorylée et favorise la fixation des snRNP U1 et U2 sur le site donneur et le point de branchement, respectivement (Feng *et al.*, 2008). Lorsque la cellule entre en mitose ou suite à un choc thermique, cette protéine est déphosphorylée et inhibe la réaction d'épissage en empêchant la fixation de la snRNP U1 sur le site donneur (Feng *et al.*, 2008 ; Shin *et al.*, 2004).

La phosphorylation des protéines SR est essentielle aux étapes précoces de l'assemblage du spliceosome (Mermoud *et al.*, 1994). Par exemple, la phosphorylation du domaine RS de la protéine ASF/SF2 est nécessaire pour que la snRNP U1 soit recrutée sur le site donneur (Cao et Garcia-Blanco, 1998 ; Kohtz *et al.*, 1994). Il a longtemps été supposé que cette interaction implique le domaine RS phosphorylé de la protéine ASF/SF2 et le domaine RS de la protéine SR like U1-70K (Xiao et Manley, 1997 ; Yeakley *et al.*, 1999). Il a récemment été démontré que cette interaction implique en réalité les deux RRM de chacune des deux protéines (Cho *et al.*, 2011b) (**Figure I-16**). La phosphorylation du domaine RS de la protéine ASF/SF2 est essentielle car elle permet de rompre son interaction avec le domaine RRM qui peut ainsi interagir avec le domaine RRM de la protéine U1-70K (Cho *et al.*, 2011b).

A l'inverse, la déphosphorylation de certaines protéines SR est nécessaire pour la réaction d'épissage après l'assemblage du spliceosome (Mermoud *et al.*, 1992). Par exemple, la déphosphorylation des protéines U1-70K et ASF/SF2 est nécessaire pour que la première étape d'épissage puisse avoir lieu (Cao *et al.*, 1997 ; Tazi *et al.*, 1993).

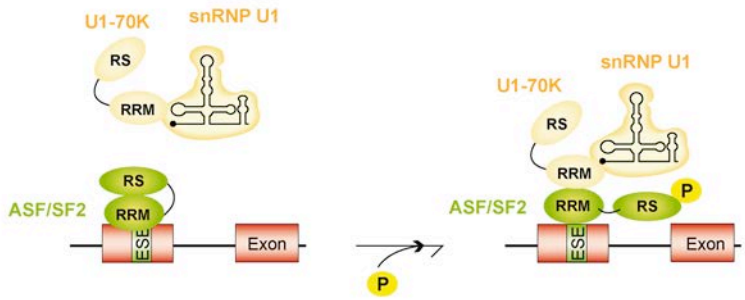



**Figure I-15 : Interaction entre la protéine U1-70K et la protéine ASF/SF2 phosphorylée lors du recrutement de la snRNP U1 sur le site donneur (Cho et al., 2011).**
Lorsque le domaine RS de la protéine ASF/SF2 est non phosphorylé, il interagit avec le domaine RRM de la protéine ASF/SF2 qui, lui même est en contact avec l'ARN. La phosphorylation du domaine RS rompt son interaction avec le domaine RRM de la protéine ASF/SF2 et peut alors interagir avec le domaine RRM de la protéine U1-70K.

En plus de leur rôle dans la modulation de la fonction lors de l'épissage, la phosphorylation des domaines RS des protéines SR régulent également leur localisation subcellulaire et leur capacité à faire la navette noyau-cytoplasme. Dans le noyau, les protéines SR phosphorylées sont concentrées dans les speckles qui sont des foyers nucléaires où se concentrent de nombreux facteurs impliqués dans l'épissage et qui sont situés dans la périphérie des sites actifs de transcription appelée paraspeckle (Misteli *et al.*, 1998 ; Spector et Lamond, 2011). Une fois la réaction d'épissage terminée, les protéines SR restent associées avec l'ARNm dont elles favorisent l'export vers le cytoplasme en interagissant avec le facteur d'export TAP/NFX1 après avoir été déphosphorylées (Huang *et al.*, 2003 ; Huang *et al.*, 2004) (**Figure I-16**). Les protéines SR y sont ensuite re-phosphorylées puis réimportées dans le noyau via leur domaine RS qui interagit avec le facteur d'import hMtr10/transportine-SR (Kataoka *et al.*, 1999 ; Lai *et al.*, 2001). Les protéines ASF/SF2 SRp20 et 9G8 effectuent cette navette noyau-cytoplasme tandis que la protéine SC35 possède un signal de rétention nucléaire dans son domaine RS qui la maintient dans le noyau (Cazalla *et al.*, 2002).

Trois famille de protéines kinases qui phosphorylent les protéines SR ont été identifiés (Bourgeois *et al.*, 2004) : la famille SRPK (SR Protein Kinase) composée de SRPK1 (Gui *et al.*, 1994) et de SRPK2 (Wang *et al.*, 1998), la famille Clk/Sty (Colwill *et al.*, 1996) et la DNA topoisomérase I (Rossi *et al.*, 1996). Trois phosphatases déphosphorylant les protéines SR ont été également identifiées : ΠΠ1, 2Xγ et PP2A.

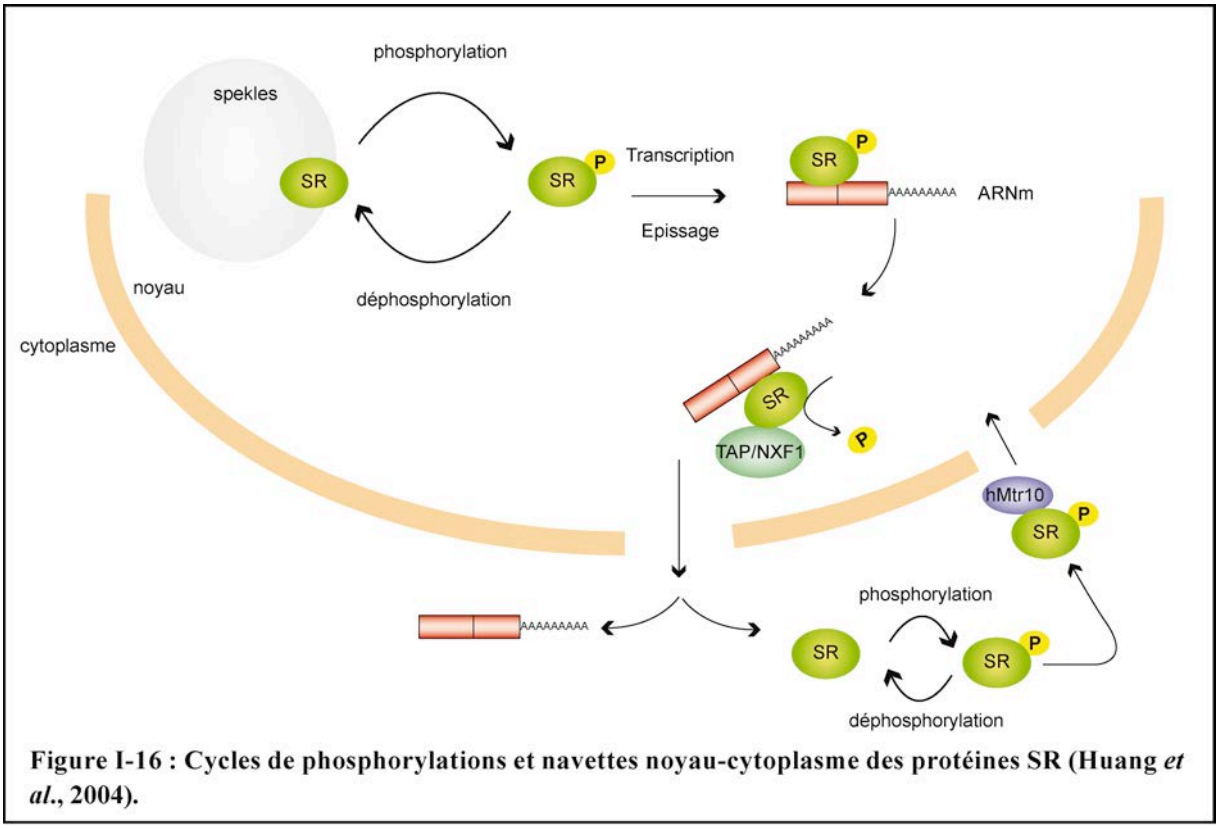



**Figure I-16 : Cycles de phosphorylations et navettes noyau-cytoplasme des protéines SR (Huang *et al.*, 2004).**

### III.3.2.3. Les séquences reconnues par les protéines SR

Comme nous l'avons précédemment abordé, lors de la définition de l'exon, la fixation des protéines SR sur les séquences exoniques ESE favorise la reconnaissance des sites d'épissage. Des expériences reposant sur l'emploi de la méthode SELEX (Systematic Evolution of Ligands throught EXponential enrichment) ont été menées pour définir les séquences consensus reconnues spécifiquement par ces protéines. Le principe de cette technique est de sélectionner parmi un ensemble de séquences aléatoires celles qui auront la meilleure affinité pour une protéine SR donnée. Dans les expériences SELEX dites classiques, les séquences ARN sont sélectionnées sur la capacité de la protéine SR à s'y fixer. Après plusieurs cycles de sélection, les séquences ARN formant un complexe ARN/protéine sont isolées puis séquencées. Dans les expériences de SELEX fonctionnels, la sélection est réalisée en fonction de la capacité des séquences ARN à activer l'épissage (Bourgeois *et al.*, 2004). Dans ce cas, les séquences aléatoires sont insérées dans des constructions d'épissage à la place de la séquence ESE naturelle (Cartegni *et al.*, 2002). Pour chaque protéine SR, les séquences obtenues par ces deux approches sont nombreuses et dégénérées (Bourgeois *et al.*, 2004 ; Long et Caceres, 2009) (**Figure I-17**).

Deux outils informatiques ont été créés dans le but de rechercher des séquences ESE putatives. Le programme « ESE finder » recherche des séquences ESE fixant les protéines

ASF/SF2, SC35, SRp40 et SRp55 d'après les séquences ESE obtenus par SELEX fonctionnel pour ces protéines (Cartegni *et al.*, 2003). Le programme « ESE RESCUE » (Relative Enhancer and Silencer Classification by Unanimous Enrichment) prend en compte l'architecture des exons et des introns et repose sur l'analyse statistique de la surreprésentation d'hexanucléotides se situant dans les exons à proximité des sites d'épissage (Fairbrother *et al.*, 2004).

**A**

| Protéine SR | | Site de fixation | Méthode | Référence |
|---|---|---|---|---|
| SRSF1 | ASF/SF2 | RGAAGAAC | SELEX | Tacke et Manley, 1995 |
| | | AGGACRRAGC | SELEX | Tacke et Manley, 1995 |
| | | SRSASGA | SELEX fonctionnel | Liu *et al.*, 1998 |
| | | UGRWG | CLIP | Sanford *et al.*, 2008 |
| SRSF2 | SC35 | AGSAGAGUA | SELEX | Tacke et Manley, 1995 |
| | | GUUCGAGUA | SELEX | Tacke et Manley, 1995 |
| | | GRYYCSYR | SELEX fonctionnel | Liu *et al.*, 2000 |
| | | UGUUCSAGWU | SELEX | Cavaloc *et al.*, 1999 |
| | | GWUWCCUGCUA | SELEX | Cavaloc *et al.*, 1999 |
| | | GGGUAUGCUG | SELEX | Cavaloc *et al.*, 1999 |
| | | GAGCAGUAGKS | SELEX | Cavaloc *et al.*, 1999 |
| | | AGGAGAU | SELEX | Cavaloc *et al.*, 1999 |
| | | UGCNGYY | SELEX fonctionnel | Schaal *et al.*, 1999 |
| SRSF3 | SRp20 | GGUCCUCUUC | Retard sur gel | Lou *et al.*, 1998 |
| | | WCWWC | Epissage | Cavaloc *et al.*, 1999 |
| | | CUCKUCY | Affinité à l'ARN | Schaal *et al.*, 1999 |
| SRSF4 | SRp75 | GAAGGA | Pontage aux UV | Exline *et al.*, 2008 |
| SRSF5 | SRp40 | GAGCAGUCGGCUC | SELEX | Tacke *et al.*, 1997 |
| | | ACDGS | SELEX fonctionnel | Liu *et al.*, 1998 |
| SRSF6 | SRp55/B52 | USCGKM | SELEX fonctionnel | Liu *et al.*, 1998b |
| | | UCAACCAGGCGAC | SELEX | Shi *et al.*, 1997 |
| SRSF7 | 9G8 | UCAACA | Pontage aux UV | Lynch et Maniatis 1996 |
| | | ACGAGAGAY | SELEX | Cavaloc *et al.*, 1999 |
| | | GGACGACGAG | SELEX fonctionnel | Schaal *et al.*, 1999 |
| SRSF9 | SRp30c | GACGAC | SELEX fonctionnel | Tian et Kole, 2001 |
| | | AAAGAGCUCGG | SELEX fonctionnel | Tian et Kole, 2001 |
| | | CUGGAUU | Retard sur gel | Simard et Chabot, 2002 |
| SRSF11 | p54 | Riche en C | Pontage aux UV | Kennedy *et al.*, 1998 |

**B**




**Figure I-17 : Séquences ARN des sites de fixation des protéines SR identifiées expérimentalement (A) ou *in silico* par le programme ESE finder (B).**

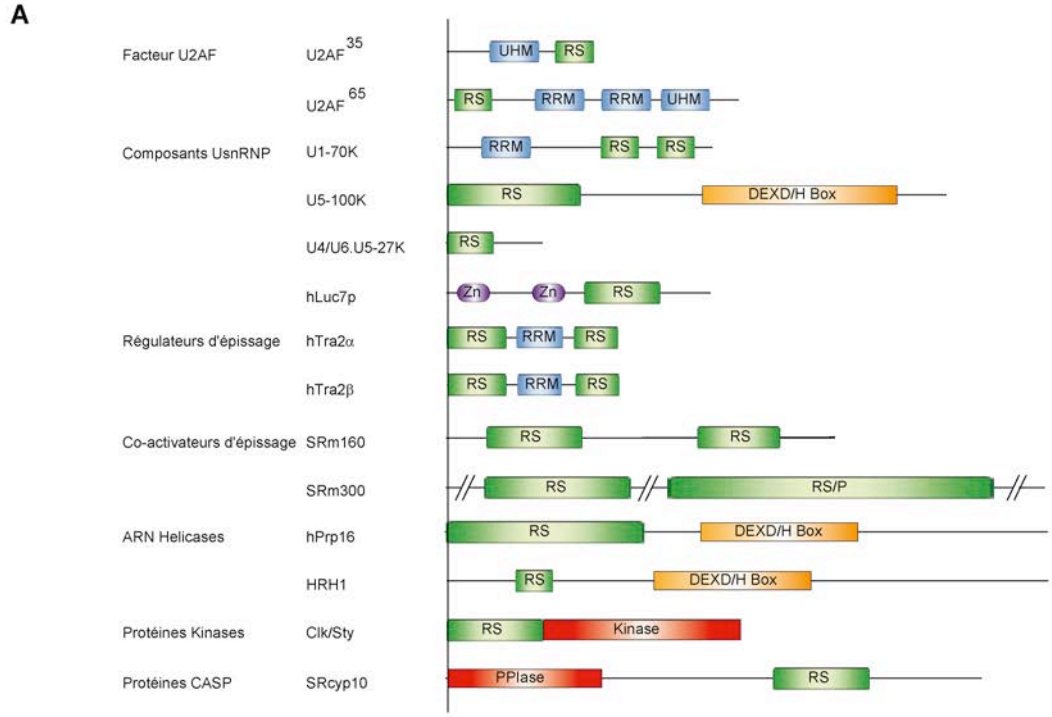
**A :** Liste des séquences ARN de fixation des différentes protéines SR (Long et Caceres, 2009). Pour chaque séquence, la méthode utilisée et la référence bibliographique sont précisées.

**B :** Séquences consensus de fixation de protéines SR déterminées par le programme ESE finder (Cartegni *et al.*, 2002). La hauteur des nucléotides reflète la fréquence de sa présence à cette position.

R : A / G ; Y : C / U ; S : C / G ; K : G / U ; W : A / U ; M : A / C ; D : A / G / U ; N : nucléotide quelconque.

### III.3.2.4. Les protéines SR like

Le terme protéine SR like (ou SRrp, pour SR protéine related protein) regroupe les protéines qui ont un domaine RS et qui sont impliquées dans le métabolisme du pré-ARNm mais qui n'ont pas l'organisation structurale d'une protéine SR (Manley et Krainer, 2010) (**Figure I-18**).




B

| Protéine SR | Site de fixation | Méthode | Référence |
|---|---|---|---|
| hTra2 β | (GAA)n | SELEX | Tacke *et al.*, 1998 |
| SRm160 | Riche en purine (GAA)n | Epissage | Eldridge *et al.*, 1999 |
| U2AF65 | $U_6YCCYU_8CC$ | SELEX | Singh *et al.*, 1995 |
| U2AF65-U2AF35 | UUUYYYYUNYAGGU | SELEX | Wu *et al.*, 1999 |

**Figure I-18 : Organisation structurale des protéines "SR like" humaines et identification des séquences de fixation de certaines d'entre elles (Graveley, 2000).**
**A :** Organisation structurale des protéines "SR like". Ces protéines contiennent toutes au moins un domaine RS mais ne possèdent pas forcément un domaine RRM.
**B :** Séquences ARN consensus de fixation de certaines protéines SR like.
RRM : RNA Recognition Motif classique, UHM : U2AF Homology Domain, RS : domaine riche en résidus arginines (R) et sérines (S), Zn : doigt de zinc, DEXD/H : motif ARN hélicase, PPI : peptidyl-prolyl *cis*/*trans* isomérase, Y : pyrimidine, N : nucléotide quelconque.

Elles comprennent des composants des snRNP comme la protéine U1-70K qui fait partie de la snRNP U1, la protéine hPrp28 (U5-100K) de la snRNP U5 et la protéine 27K de la tri-

snRNP U4/U6.U5 (Fetzer *et al.*, 1997 ; Query *et al.*, 1989 ; Teigelkamp *et al.*, 1997). Ce groupe comporte également les deux sous-unités du facteur U2AF, U2AF35 et U2AF65, des co-activateurs de l'épissage, SRm 160 et 300, des régulateurs de l'épissage, Traα et Traβ et des ARN hélicases comme hPrp16 (Amrein *et al.*, 1988 ; Boggs *et al.*, 1987 ; Zhou et Reed, 1998). Cette liste n'est pas exhaustive car les analyses du contenu protéique des complexes spliceosomaux par spectrométrie de masse de nombreuses ont permis d'identifier de nombreuses protéines contenant un domaine RS associées au pré-ARNm (Long et Caceres, 2009).

#### III.3.2.4.1. Les facteurs U2AF et la reconnaissance du site accepteur

Le facteur U2AF est un facteur d'épissage essentiel qui est impliqué dans la reconnaissance du site accepteur par la snRNP U2 lors des étapes précoces de l'assemblage du spliceosome. C'est un hétérodimère composé d'une sous-unité de 65 kDa (U2AF65) et d'une sous-unité de 35 kDa (U2AF35) (Zamore et Green, 1989, 1991) (**Figure I-19**).

La sous-unité U2AF65 interagit avec la séquence polypyrimidine par l'intermédiaire de ses 2 domaines RRM avec une affinité proportionnelle à la longueur et à la richesse en uridines de cette séquence (Sickmier *et al.*, 2006 ; Wu *et al.*, 1999 ; Zamore *et al.*, 1992). Lorsque cette séquence est optimale, la sous-unité U2AF65 s'y fixe seule et contacte le point de branchement via son domaine RS (Shen et Green, 2006 ; Valcarcel *et al.*, 1996). Elle interagit de façon coopérative via son 3$^{e}$ domaine RRM non canonique appelé UHM (U2AF Homology Motif) avec l'extrémité N-terminale de la protéine SF1/mBBP (Splicing Factor 1/mammalian Branch Point Binding Protein) qui vient se fixer sur le point de branchement (Berglund *et al.*, 1998 ; Berglund *et al.*, 1997 ; Liu *et al.*, 2001 ; Selenko *et al.*, 2003) (**Figure I-19**). La fixation de la protéine SF1 dépend du degré de similitude avec la séquence consensus et est transitoire car elle est déplacée par l'arrivée de la snRNP U2 (Berglund *et al.*, 1998 ; Corioni *et al.*, 2011 ; Query *et al.*, 1994). La protéine SF1 contribue à la cinétique optimale de l'assemblage du spliceosome et influence l'épissage alternatif de certains pré-ARNm (Corioni *et al.*, 2011).

En revanche, lorsque la séquence polypyrimidine est sous-optimale, c'est-à-dire que l'enchaînement de pyrimidines est entrecoupé de purines, la fixation de la sous-unité U2AF65 nécessite d'y être stabilisée par son interaction par son interaction avec la sous-unité U2AF35 qui vient se fixer sur les dinucléotides conservés AG du site accepteur (Guth *et al.*, 1999 ; Merendino *et al.*, 1999 ; Wu *et al.*, 1999 ; Zorio et Blumenthal, 1999) (**Figure I-19**). Dans ce

cas, on parle d'exon AG-dépendants, par opposition aux exons AG-indépendants (Moore, 2000).

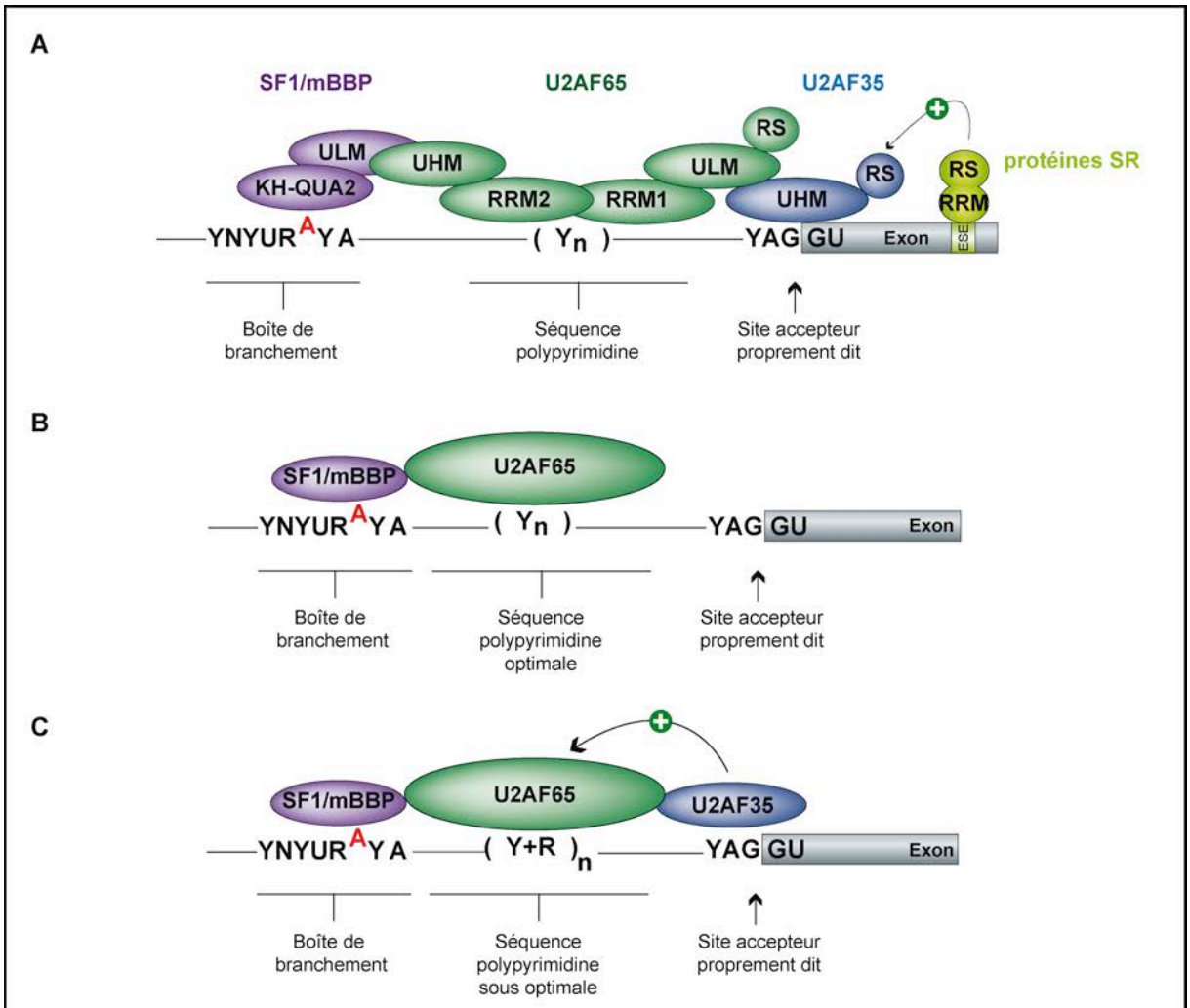



**Figure I-19 : Reconnaissance du site accepteur par les facteurs U2AF et la protéine SF1/mBBP (Adapté de Mackereth *et al.*, 2005 ; Moore, 2000).**

**A :** Interactions protéine-protéine et protéine-ARN lors de la reconnaissance du site accepteur par le facteur U2AF et la protéine SF1. La sous-unité U2AF65 interagit avec la séquence polypyrimidine par l'intermédiaire de ses deux domaines RRM et avec la protéine SF1/mBBP avec son 3e domaine RRM non canonique appelé UHM. La protéine SF1 interagit avec le point de branchement grâce à son domaine KH-QUA2. La sous unité U2AF35 contacte le dinucléotide AG via son domaine RRM non canonique UHM et interagit avec la sous-unité U2AF65 par son domaine de liaison ULM. Elle peut également être recrutée par des protéines SR avec lesquelles elle interagit par son domaine RS.
Le point de branchement est coloré en rouge.
**B :** Exon AG-indépendant. Lorsque la séquence polypyrimidine est optimale, la sous-unité U2AF65 s'y fixe seule et de façon coopérative avec la protéine SF1/mBBP qui se fixe sur le point de branchement.
**C :** Exon AG-dépendant. Lorsque la séquence polypyrimidine est sous-optimale, la fixation du facteur U2AF65 doit y être stabilisée par son interaction avec la sous-unité U2AF35 qui se fixe sur le dinucélotide AG du site accepteur.
RRM : RNA Recognition Motif, UHM : U2AF Homology Domain, RS : domaine riche en résidus arginines (R) et sérines (S), ULM : U2AF Ligand Motif, KH : K-homology, QUA2 : Quaking homology-2, Y : pyrimidine, R : purine, N : nucléotide quelconque.

La sous-unité U2AF35 interagit à la fois avec l'ARN et avec la sous-unité U2AF65 par un domaine UHM (Kielkopf *et al.*, 2001 ; Selenko *et al.*, 2003 ; Zorio et Blumenthal, 1999). Lors de la définition des exons, sa fixation sur le site accepteur peut être favorisée par la fixation antérieure de protéines SR sur une séquence ESE exonique avec lesquelles elle interagit par son domaine RS (Graveley *et al.*, 2001 ; Wu et Maniatis, 1993) (**Figure I-19**).

### III.3.3. Le complexe H

Le complexe H (pour Heterogeneous) est un complexe aspécifique qui s'assemble sur l'ARN neo-synthétisé dès sa transcription indépendamment de la présence d'intron. *In vitro*, ce complexe se forme dès qu'un ARN est incubé en présence d'extrait nucléaire en l'absence d'ATP. L'analyse de sa composition par spectrométrie de masse indique qu'il est principalement composé de protéines hnRNP (heterogeneous nuclear RNP) dont certaines s'associent quasiment de manière ubiquitaire au pré-ARNm (Bennett *et al.*, 1992 ; Jurica et Moore, 2003) (**Tableau I-20**).

Toutes les protéines hnRNP ne sont pas présentes en quantité équivalente dans le complexe H et la fixation de certaines dépend de la séquence du pré-ARNm (Bennett *et al.*, 1992). Certaines protéines hnRNP restent fixées sur le pré-ARNm tout au long de la réaction d'épissage et participent à sa régulation (Bennett *et al.*, 1992b). Nous aborderons cette fonction des protéines hnRNP plus en détail dans la seconde partie de l'introduction. Le complexe H comporte également des facteurs associés à l'épissage et des protéines de la coiffe (Jurica et Moore, 2002, 2003 ; Zhou *et al.*, 2002) (**Tableau I-20**).

| Protéines | N°d'accès | Jurica *et al.*, 2002 | Zhou *et al.*, 2002 |
|---|---|---|---|
| Proteines de la coiffe et hnRNP | | | |
| CBP80 | Q01081 | H | |
| hnRNP A0 | Q13151 | | H |
| hnRNP A1 | P09651 | H | H |
| hnRNP A2/B2 | P22626 | H | H |
| hnRNP A3 | P51991 | H | H |
| hnRNP C | P07910 | H | H |
| hnRNP D | Q14103 | H | H |
| hnRNP F | P52597 | H | |
| hnRNP G | P38159 | H | |
| hnRNP H | P31943 | H | |
| hnRNP I/PTB | P26599 | H | H |
| hnRNP K | Q07244 | H | H |
| hnRNP L | P14866 | H | H |
| hnRNP M | P52272 | H | |
| hnRNP Q | O60506 | H | H |
| hnRNP R | O43390 | H | H |
| hnRNP U | Q00839 | H | |
| hnRNP RALY | Q9BQX6 | H | H |
| PAB2 | Q15097 | H | |
| FUS | P35637 | H | H |

| Protéines | N°d'accès | Jurica *et al.*, 2002 | Zhou *et al.*, 2002 |
|---|---|---|---|
| Facteurs associés à l'épissage | | | |
| CDC5L | Q99459 | H | |
| SKIP | Q13573 | H | |
| PRL1 | 043660 | H | |
| Dbp1 | P42285 | H | |
| ISY1 | Q9BT05 | H | |
| ECM2 | Q9NW64 | H | |
| p68 | P17844 | H | |
| Ref | O43672 | H | |
| Ku70 | P12956 | H | |
| TopIIa | P11388 | H | |
| hUR | Q15717 | H | H |
| NF45 | A54857 | H | |
| NF90 | B54857 | H | H |
| Matrin 3 | P43243 | H | H |
| YB-1 | P16991 | H | H |
| TLS ip | | H | H |
| HSP70-2 | P08107 | H | H |
| HSP71 | P11142 | H | H |
| ILF3 | AJ271745 | | H |
| hBUB3 | O43684 | | H |
| TAFII68 | Q92804 | | H |
| DBPA | P16989 | | H |
| GRP78 | P11021 | | H |
| KSRP | Q3U0V1 | Rhode *et al.*, 2003 | |

**Tableau I-20 : Les protéines du complexe H (Adapté de Jurica *et al.*, 2002 ; Zhou *et al.*, 2002).**
Liste des protéines du complexe H associées au pré-ARNm en l'absence d'épissage. Les protéines des complexes H purifiés par Jurica *et al.*, 2002 et Zhou *et al.*, 2002 ont été identifiées par spectrométrie de masse et leur présence au sein des complexes est indiquée par un "H". Le numéro d'accès de chaque protéine dans la banque de données UniProt est indiqué.

### III.3.4. Les complexes E' et E ou complexes d'engagement

*In vitro*, et après incubation à 30°C, le complexe H assemblé sur le pré-ARNm est rapidement converti en complexe E également appelé complexe d'engagement car sa formation implique la reconnaissance des sites donneur et accepteur d'épissage (Das *et al.*, 2000 ; Michaud et Reed, 1993). L'assemblage du complexe E ne nécessite pas d'ATP et comprend la reconnaissance du site donneur par la snRNP U1, de la séquence polypyrimidine par la sous-unité U2AF65, le dinucléotide AG du site accepteur par la sous-unité U2AF35 et du point de branchement par le facteur mBBP/SF1 (Abovich et Rosbash, 1997 ; Reed, 2000 ; Staknis et Reed, 1994) (**Figure I-21**). L'assemblage du complexe E est séquentiel et est

précédé par la formation du complexe E' où les fixations préalables de la snRNP U1 et du facteur mBBP/SF1 permettent le recrutement du facteur U2AF (Cote *et al.*, 1995 ; Das *et al.*, 2000 ; Kent *et al.*, 2005) (**Figure I-21**). Le réseau d'interaction établi entre ces protéines auquel s'ajoute celui des protéines SR contribue à la formation du complexe E ainsi qu'au rapprochement des sites donneur et accepteur d'épissage (Cote *et al.*, 1995 ; Fu et Maniatis, 1992 ; Kent et MacMillan, 2002 ; Wu et Maniatis, 1993) (**Figure I-21**).

Le complexe E comprend également la snRNP U2 dont la présence est nécessaire à sa formation et qui, contrairement aux autres facteurs, contacte faiblement ou de manière indirecte le pré-ARNm (Das *et al.*, 2000 ; Donmez *et al.*, 2007 ; Donmez *et al.*, 2004 ; Gozani *et al.*, 1998 ; Hong *et al.*, 1997). La configuration spatiale de la snRNP U2 au sein du complexe E n'est pas bien connue. Néanmoins, des expériences d'empreintes indiquent que la snRNP U2 occupe une place centrale au sein du complexe E qui est importante pour le rapprochement dans l'espace les sites donneur et accepteur d'épissage (Donmez *et al.*, 2007 ; Kent et MacMillan, 2002). L'extrémité 5' du snRNA U2 est à la fois à proximité du site donneur, du point de branchement et du site 3'. Cette organisation spatiale est établie grâce à un réseau d'interactions protéine-protéine où la snRNP U2 interagit par l'intermédiaire de la protéine SF3b155 avec la sous-unité U2AF65 tandis que l'hélicase hPrp5 qui lui est associée ou la protéine SR SC35 permet de la ponter à la snRNP U1 (Fu et Maniatis, 1992 ; Gozani *et al.*, 1998 ; Xu *et al.*, 2004) (**Figure I-21**). La conformation de la snRNP U2 vis-à-vis de la snRNP U1 dans le complexe E ne semble pas changer de façon drastique lors de la formation du complexe spliceosomal suivant, le complexe A. Cela suggère que les séquences et les acteurs importants pour la formation du complexe spliceosomal ultérieur semblent déjà être spatialement pré-organisés au sein du complexe E (Donmez *et al.*, 2007).

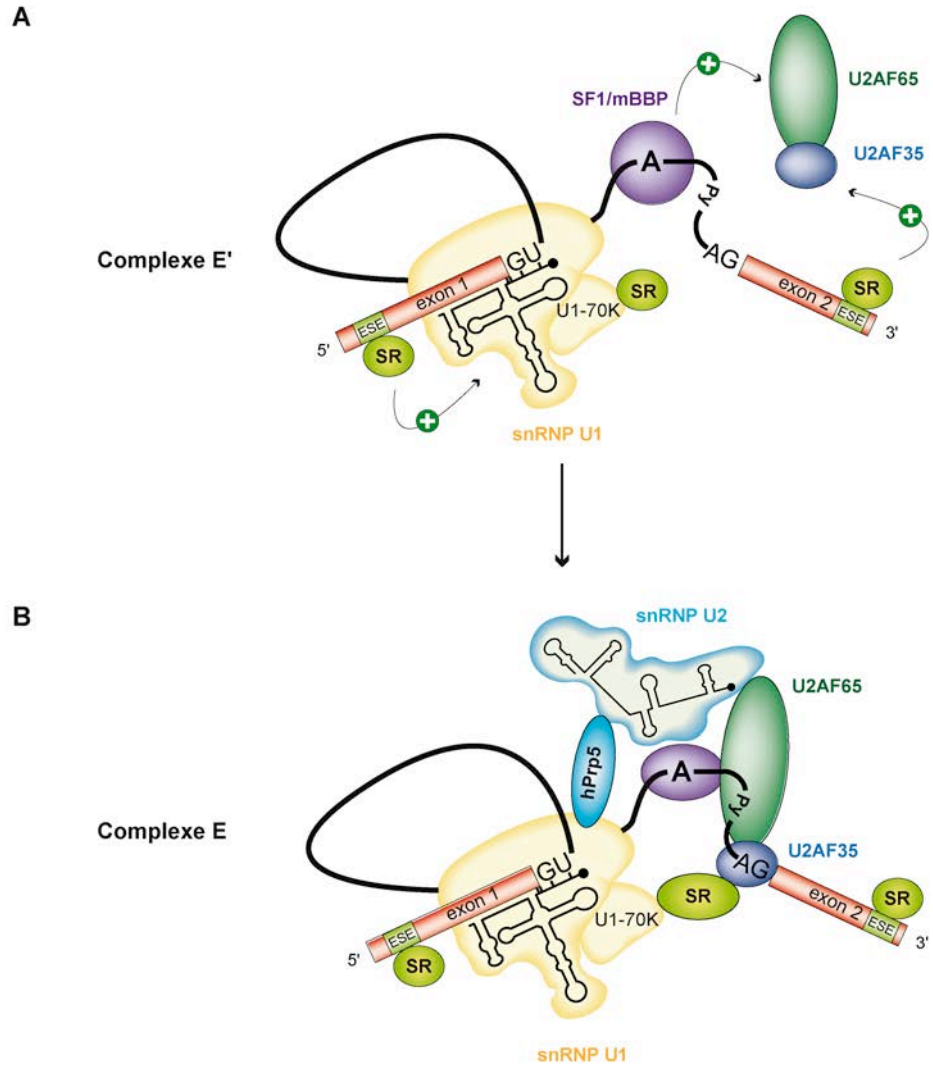



**Figure I-21 : Modèle d'assemblage des complexes E' et E (Adapté de Kent *et al.*, 2005).**
**A :** Lors de la formation du complexe E', le site donneur et le point de branchement sont respectivement associés avec la snRNP U1 et le facteur SF1/mBBP. Les protéines SR favorisent le recrutement de la snRNP U1 ainsi que celui de l'hétérodimère U2AF en collaboration avec le facteur SF1/mBBP.
**B :** Lors de la formation du complexe E, la sous-unité U2AF35 interagit avec le dinucélotide AG du site donneur et la sous-unité U2AF65 avec la séquence polypyrimidine et la facteur SF1/mBBP. Les protéines SR associées avec la protéine U1-70K contactent la sous-unité U2AF35, ce qui permet de ponter et de rapprocher les sites donneur et accepteur. La snRNP U2 n'est pas associée au complexe E de façon stable et ne contacte pas directement le pré-ARNm. Cette particule interagit directement avec la sous-unité U2AF65 et interagit avec la snRNP U1 par l'intermédiaire de l'hélicase hPrp5.

### III.3.5. Le complexe A ou préspliceosome

La formation du complexe A se caractérise par l'appariement du snRNA U2 avec la séquence de branchement, ce qui provoque le départ du facteur mBBP/SF1 qui y était préalablement fixé (Rutz et Seraphin, 1999) (**Figure I-22**).

Cet appariement nécessite l'hydrolyse de l'ATP par les hélicases hPrp5 et hUAP56 (U2AF65 Associated Protein) qui permettent la transition du complexe E vers le complexe A (Fleckner *et al.*, 1997 ; Shen *et al.*, 2008) (**Figure I-22**).

L'hélicase hPrp5 est associée à la particule snRNP U2 et catalyse les changements conformationnels du snRNA U2 lors de l'épissage (Kosowski *et al.*, 2009 ; Will *et al.*, 2002 ; Xu *et al.*, 2004). Lors de la transition du complexe E vers le complexe A, son activité consiste à remanier la structure BSL (Branchpoint interaction Stem loop) du snRNA U2 complémentaire à la boîte de branchement au profit d'une hélice avec la boîte de branchement dans lequel l'adénosine servant de point de branchement est en bulge (Perriman et Ares, 2010 ; Query *et al.*, 1994) (**Figure I-22**). L'hélicase hUAP56 contribue également à la formation de cet appariement mais son mécanisme d'action est inconnu (Shen *et al.*, 2008). Cette hélicase intègre le complexe A où elle y interagit avec la sous-unité U2AF65 après avoir hydrolysé l'ATP (Shen *et al.*, 2008) (**Figure I-22**).

Les protéines de la particule snRNP U2 participent également à son ancrage sur la boîte de branchement. Au cœur de la particule, la protéine SF3b155 interagit avec le domaine UHM de la protéine U2AF65 qui stabilise également par son domaine RS la fixation de la protéine SF3b14a sur le point de branchement (Gozani *et al.*, 1998 ; Valcarcel *et al.*, 1996 ; Will *et al.*, 2001) (**Figure I-22**).

L'intégration stable de la snRNP U2 au sein du complexe A s'accompagne aussi du rapprochement dans l'espace des sites donneurs et accepteurs qui avait préalablement été entrepris dans le complexe E (Donmez *et al.*, 2007). Cela met en jeu de nombreuses protéines comme les protéines SR et les protéines FBP21 et FBP11 (Formin Binding Protein 21 et 11) qui établissent des interactions entre les snRNP U1 et U2 (Bedford *et al.*, 1998 ; Fu et Maniatis, 1992 ; Lin *et al.*, 2004 ; Tarn et Steitz, 1995).

Les complexes A ont été purifiés par une double chromatographie d'affinité et leur composition a été analysée par spectrométrie de masse (Behzadnia *et al.*, 2007 ; Hartmuth *et al.*, 2002) (**Tableau I-23**).

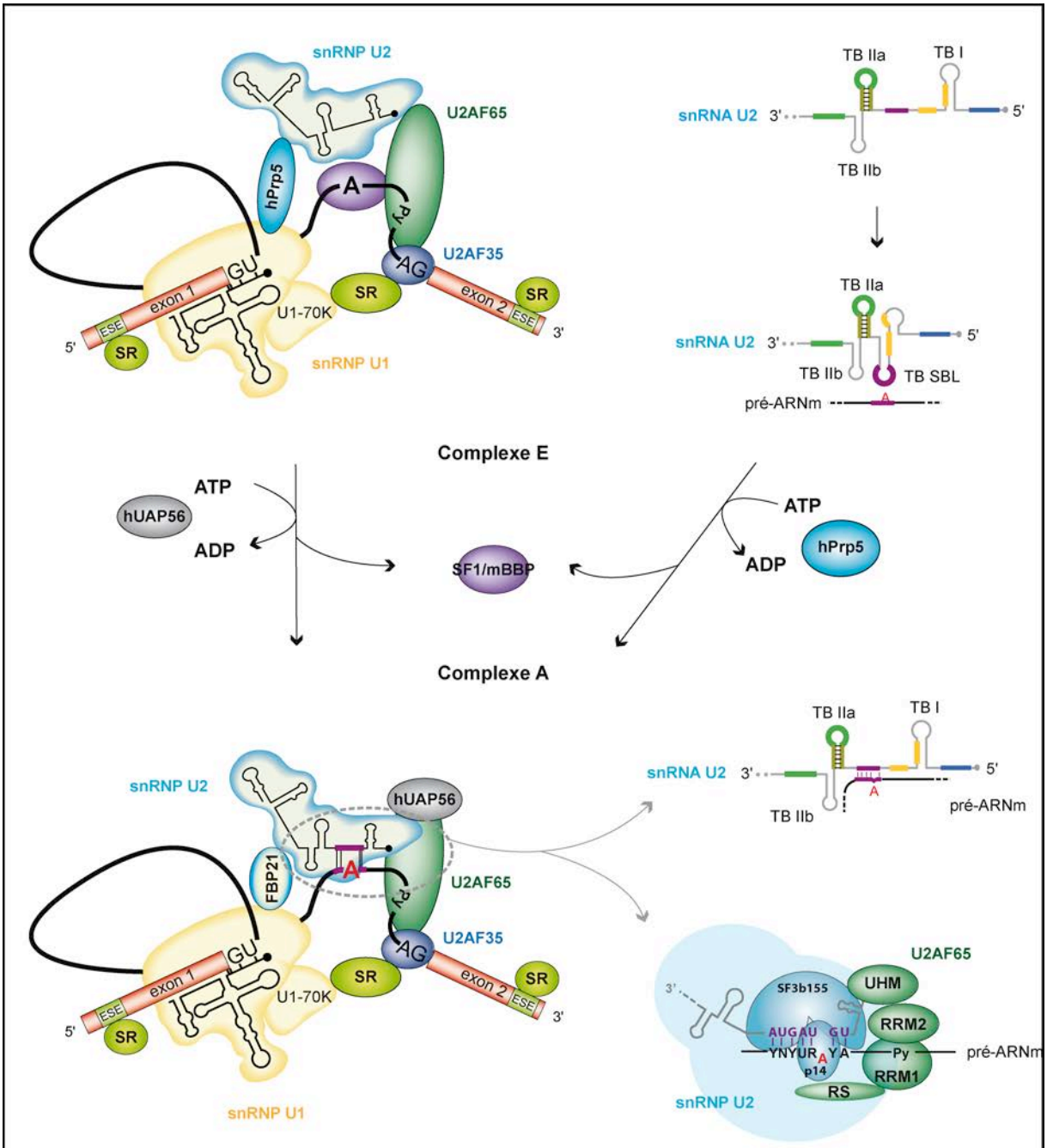



**Figure I-22 : Modèle d'assemblage du complexe A (Adapté de Das *et al.*, 2000 et Perriman et Ares, 2010).**

La formation du complexe A nécessite l'hydrolyse de l'ATP par les hélicases hUAP56 et hPrp5 qui permettent la fixation stable de la snRNP U2 sur la boîte de branchement. Les changements conformationnels du snRNA U2 lors de la transitition du complexe E vers le complexe A sont schématisés sur la droite. La séquence du snRNA U2 complémentaire de la boîte de branchement est colorée en violet, celles s'appariant ultérieurement avec le snRNA U6 sont en jaune et bleu, celles impliquées dans d'autres changements conformationnels sont en vert. Le détail moléculaire de l'appariement de la boîte de branchement du pré-ARNm avec le snRNA U2 ainsi que ses interactions avec les protéines du complexe A sont représentés à droite de ce complexe.

| Protéines | N° accès |
|---|---|
| **Protéines Sm** | |
| B | gi\|4507125 |
| D1 | gi\|5902102 |
| D2 | gi\|29294624 |
| D3 | gi\|4759160 |
| E | gi\|4507129 |
| F | gi\|4507131 |
| G | gi\|4507133 |
| **snRNP U1** | |
| U1-70K | gi\|29568103 |
| U1-A | gi\|4759156 |
| U1-C | gi\|4507127 |
| **Protéines associées à la snRNP U1** | |
| FBP11 | gi\|88953744 |
| S164 (fSAP94) | gi\|4050087 |
| **snRNP U2** | |
| U2A' | gi\|50593002 |
| U2B'' | gi\|4507123 |
| SF3a120 | gi\|5032087 |
| SF3a66 | gi\|21361376 |
| SF3a60 | gi\|5803167 |
| SF3b155 | gi\|54112117 |
| SF3b145 | gi\|55749531 |
| SF3b130 | gi\|54112121 |
| SF3b49 | gi\|5032069 |
| SF3b14a/p14 | gi\|7706326 |
| SF3b14b | gi\|14249398 |
| SF3b10 | gi\|13775200 |
| **Protéines associées à la snRNP U2** | |
| hPRP43 | gi\|68509926 |
| SPF45 | gi\|14249678 |
| SPF30 | gi\|5032113 |
| U2AF65 | gi\|6005926 |
| U2AF35 | gi\|5803207 |
| SPF31 | gi\|7657611 |
| hPRP5 | gi\|41327773 |
| SR140 (fSAPa) | gi\|51492636 |
| CHERP | gi\|21359884 |
| PUF60 | gi\|17298690 |
| SF3b125 | gi\|45446743 |
| **snRNP U5** | |
| 220K | gi\|3661610 |
| 200K | gi\|45861372 |
| 116K | gi\|41152056 |
| 102K | gi\|40807485 |
| 100K | gi\|41327771 |
| 52K | gi\|5174409 |
| **snRNP U4/U6** | |
| 90K | gi\|4758556 |
| **snRNP U4/U6.U5** | |
| 110K | gi\|13926068 |
| **Protéines SR** | |
| SF2/ASF | gi\|5902076 |
| 9G8 | gi\|72534660 |
| SRp20 | gi\|4506901 |
| SRp30c | gi\|4506903 |
| SRp40 | gi\|3929378 |
| SRp55 | gi\|20127499 |
| SRp75 | gi\|21361282 |
| SC35 (SFRS2) | gi\|47271443 |
| hTra-2 beta/SFRS10 | gi\|4759098 |

| Protéines | N° accès |
|---|---|
| **protéines apparentées aux protéines SR** | |
| FLJ10154 | gi\|48675817 |
| SRm160 | gi\|3005587 |
| SRm300 | gi\|19923466 |
| **hnRNP** | |
| hnRNP A1 | gi\|4504445 |
| hnRNP A3 | gi\|34740329 |
| hnRNP AB | gi\|12803583 |
| hnRNP A2/B1 | gi\|14043072 |
| hnRNP C1/C2 | gi\|4758544 |
| hnRNP G | gi\|56699409 |
| hnRNP K | gi\|14165435 |
| hnRNP Q | gi\|23397427 |
| hnRNP R | gi\|5031755 |
| hnRNP U | gi\|14141161 |
| PCBP1 | gi\|5453854 |
| E1B-AP5 | gi\|21536326 |
| **Protéines de la coiffe** | |
| CBP20 | gi\|19923387 |
| CBP80 | gi\|4505343 |
| **Complexe hPRP19/CDC5L** | |
| hPRP19 | gi\|7657381 |
| CDC5 (CDC5L) | gi\|11067747 |
| SPF27 | gi\|5031653 |
| PRL1 (PRLG1) | gi\|4505895 |
| CCAP1 (hsp73) | gi\|5729877 |
| CCAP2 (hspc148, AD-002) | gi\|7705475 |
| Catenin β-like 1 (CTNNBL1, NAP) | gi\|18644734 |
| Npw38BP | gi\|7706501 |
| Npw38 | gi\|5031957 |
| **Protéine associée au complexe hPRP19/CDC5L** | |
| PRCC | gi\|40807447 |
| **Autres protéines** | |
| RBM5/LUCA15 | gi\|5032031 |
| FLJ10839 | gi\|46852388 |
| ASR2B | gi\|58331218 |
| p68 (DDX5) | gi\|4758138 |
| SF4 (F23858) | gi\|33469964 |
| RNPC2 (CC1.3, CAPER, fSAP59) | gi\|4757926 |
| TLS/FUS | gi\|544357 |
| ELAV (HuR) | gi\|38201714 |
| YB-1 | gi\|34098946 |
| NFAR | gi\|24234750 |
| DDX9 | gi\|4503297 |
| SF1 | gi\|42544130 |
| tat SF1 | gi\|21361437 |
| CDC2L2 | gi\|16357494 |
| MGC2803 (fSAP18) | gi\|13128992 |
| RBM10 | gi\|12644371 |
| ZNF207 | gi\|4508017 |
| BUB3 | gi\|4757880 |
| TCERG1 (CA150) | gi\|21327715 |
| NRIP2 | gi\|13899327 |
| LOC124245 | gi\|31377595 |
| NIPP1 | gi\|13699256 |
| HCNGP | gi\|9994179 |
| PPP1CA | gi\|4506003 |
| NF45 | gi\|24234747 |
| p72/DDX17 | gi\|3122595 |
| Acinus (fSAP152) | gi\|7662238 |
| Pinin | gi\|33356174 |
| RNPS1 | gi\|6857826 |
| p30 DBC | gi\|24432106 |
| SEC31-like 2 | gi\|14149696 |
| AGGF1 | gi\|39725952 |
| CIRP | gi\|4502847 |
| TAFIIB | gi\|8392875 |

**Tableau I-23 : Les protéines du complexe spliceosomal A (Behzadnia *et al.*, 2007, Hartmuth *et al.*, 2002).**

Les complexes spliceosomaux A ont été purifiés par une première chromatographie d'affinité reposant sur l'emploi de la tobramycine suivie d'une seconde par immunoaffinité avec un anticorps spécifique de la protéine SF3a66 de la particule snRNP U2. Les complexes ont été isolés par sédimentation en gradient de glycérol et leurs protéines identifiées par spectrométrie de masse. Le numéro d'accès de chaque protéine dans la GenbanK est indiqué.

### III.3.6. Le complexes B et l'activation catalytique du spliceosome (Figure I-24 et I-25)

La transition du complexe A en complexe B est une étape ATP-dépendante qui repose sur l'incorporation de la tri-snRNP [U4/U6.U5]. Ce complexe n'est pas catalytiquement actif et pour le devenir, il va subir une cascade ordonnée de réarrangements structuraux majeurs qui vont modifier de façon coordonnée les appariements snRNA-snRNA au sein de la tri-snRNP [U4/U6.U5] et avec le snRNA U2 ainsi que les appariements snRNA-pré-ARNm.

Ces réarrangements sont séquentiels et vont aboutir à la formation d'un complexe spliceosomal catalytiquement actif, appelé complexe B*, au sein duquel la première étape de transestérification a lieu. Les grandes étapes de ces réarrangements comprennent :

- le départ du snRNA U1 du site donneur
- la dissociation du duplex snRNA U4/U6 et le départ du snRNA U4
- l'appariement du snRNA U6 avec le site donneur
- la formation d'un duplex snRNA U2/U6
- le positionnement du snRNA U5 à proximité des sites donneur et accepteur

Ces réarrangements sont séquentiels et sont spécifiquement catalysés par des hélicases qui sont majoritairement recrutées de façon transitoire avec un grand nombre d'autres protéines au sein du spliceosome lors de chacune de ses transitions structurales. Quatre complexes spliceosomaux intermédiaires successifs, les complexes B, BΔU1, Bact et B*, correspondant chacun à une étape du remodelage du spliceosome, ont été isolés (Bessonov *et al.*, 2010 ; Deckert *et al.*, 2006 ; Makarov *et al.*, 2002 ; Makarova *et al.*, 2004). Après l'incorporation de la tri-snRNP [U4/U6.U5] dans le complexe B, le départ de la snRNP U1 (complexe BΔU1) puis le départ de la snRNP U4 et la formation du duplex U2/U6 aboutissent à la formation du complexe Bact qui va ensuite donner le complexe B*. Les complexes B et Bact sont les seuls complexes ayant été isolés et purifiés dans des conditions physiologiques et leurs compositions très différentes reflètent bien la dynamique d'assemblage des différents facteurs lors de l'activation du spliceosome (Agafonov *et al.*, 2011 ; Behzadnia *et al.*, 2007 ; Bessonov *et al.*, 2010 ; Wahl *et al.*, 2009)(**Figure I-24**). Les complexes BΔU1 et B* ont quant à eux été purifiés dans des conditions plus stringentes et par immuno-affinité qui repose sur l'utilisation d'un anticorps dirigé spécifiquement contre une protéine associée de façon transitoire et caractéristique à chacun de ces 2 complexes (Makarov *et al.*, 2002 ; Makarova *et al.*, 2004).

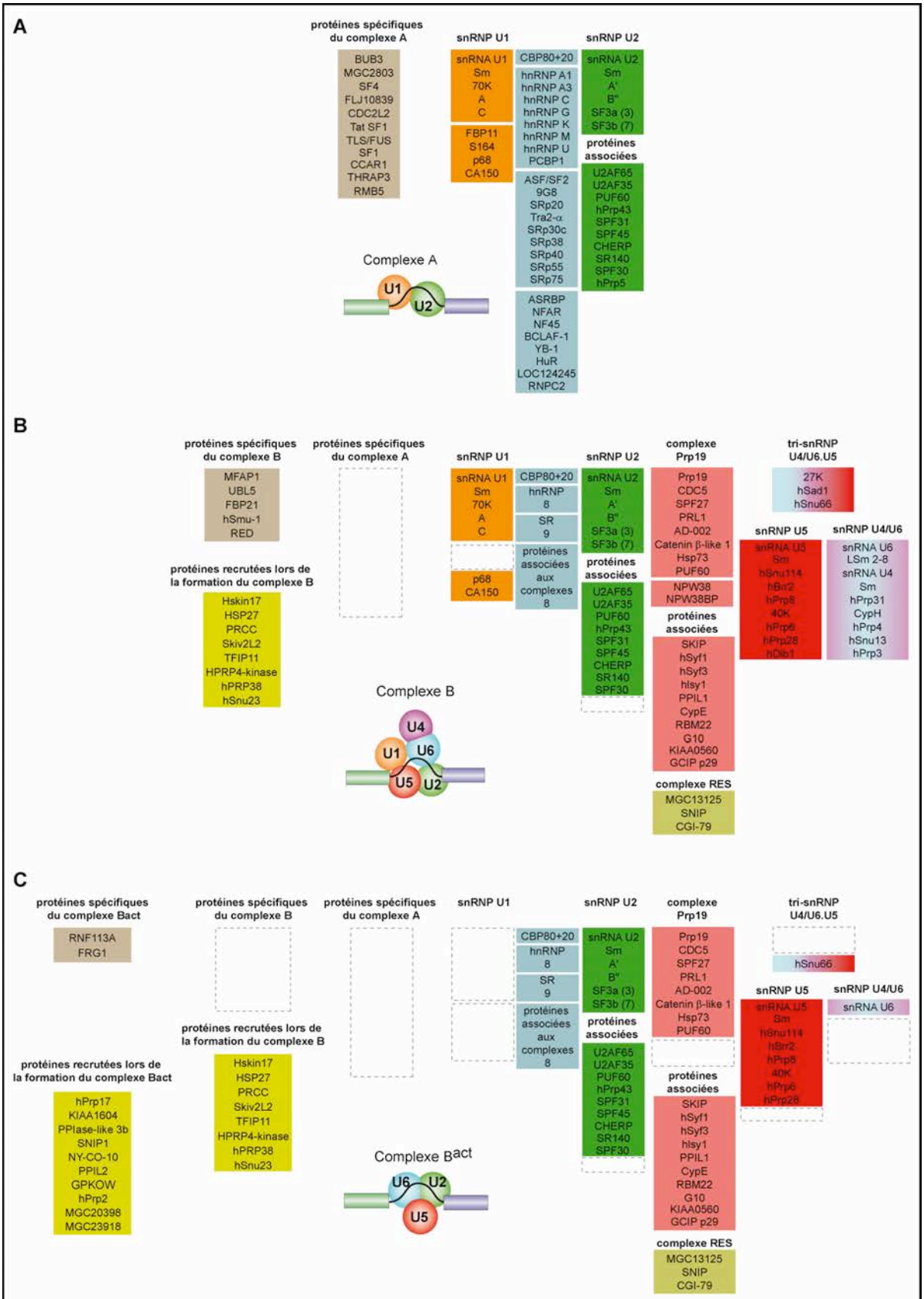



**Figure I-24 : Changements de la composition des complexes spliceosomaux A (A), B (B) et Bact (C) lors de l'épissage (Adapté de Behzadnia *et al.*, 2007, Bessonov *et al.*, 2010, Wahl *et al.*, 2009, Agafonov *et al.*, 2011).**

Les protéines sont regroupées selon leur présence dans les particules snRNP, dans les complexes hétéromériques ou leur fonction. La liste des protéines n'est pas exhaustive.

### III.3.6.5. Incorporation de la tri-snRNP [U4/U6.U5] dans le complexe B (Figure I-25, B)

Le recrutement de la tri-snRNP [U4/U6.U5] est favorisé par les protéines SR préalablement présentes dans le complexe spliceosomal A qui semblent faciliter son ancrage au sein du spliceosome B (Roscigno et Garcia-Blanco, 1995). Pour que l'intégration puisse avoir lieu, la protéine hPrp28 doit être préalablement phosphorylée par la kinase SRPK2 et les protéines hSnu66 et hSad1 doivent être présentes au sein de la tri-particule (Makarova *et al.*, 2001 ; Mathew *et al.*, 2008). Après l'intégration, les protéines hPrp6 et hPrp31 sont phosphorylées par la protéine kinase hPrp4 qui est recrutée de façon concomitante lors de la formation du complexe B (Schneider *et al.*, 2010a). Il a été suggéré que ces phosphorylations favorisent la stabilité de la tri-particule au sein du complexe B, certainement grâce à des interactions protéine-protéine entre les domaines RS des protéines SR et les protéines SR-like ainsi qu'avec les autres protéines du complexe spliceosomal (Schneider *et al.*, 2010a).

La protéine SPF30 est recrutée avec la tri-snRNP de façon concomitante et est essentielle pour la transition du complexe A vers le complexe B (Meister *et al.*, 2001 ; Rappsilber *et al.*, 2001). Cette protéine favorise l'incorporation de la tri-snRNP en interagissant simultanément avec la protéine hPrp3 de la tri-particule et la sous-unité U2AF35 préalablement fixée sur le site donneur (Little et Jurica, 2008). Cela facilite l'ancrage de la tri-snRNP au niveau du site accepteur, et donc de la snRNP U2 avec laquelle elle initie des réarrangements structuraux. A ce stade, le duplex U4/U6 au sein de la particule est encore intact mais permet à la séquence libre à l'extrémité 3' du snRNA U6 de s'apparier avec la séquence du snRNA U2 pour former l'hélice h II (Makarova *et al.*, 2004 ; Schneider *et al.*, 2010) (**Figure I-25, B**).

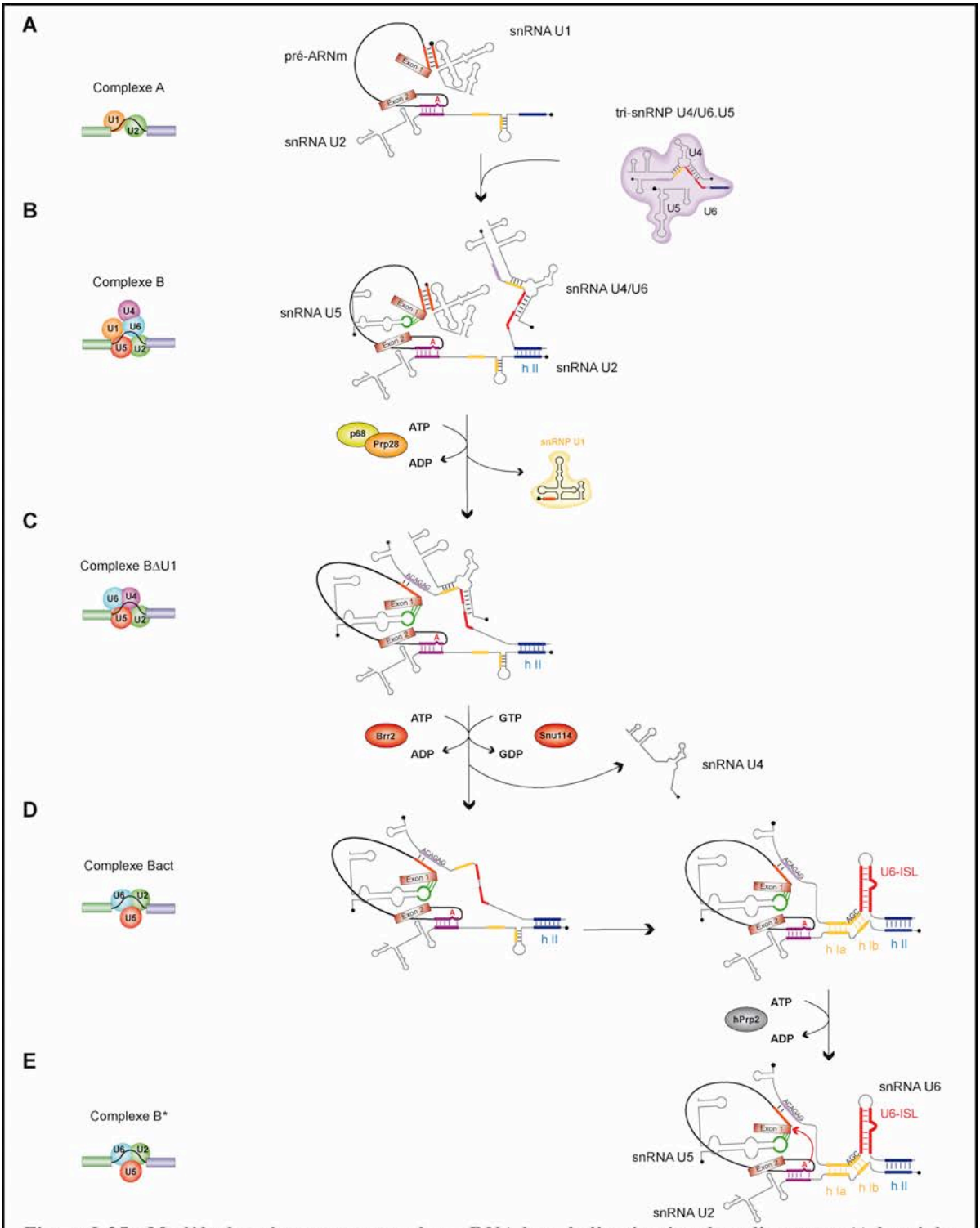



**Figure I-25 : Modèle des réarrangements des snRNA lors de l'activation du spliceosome (Adapté de Wahl *et al.*, 2009, Will et Lührmann, 2010, Schneider *et al.*, 2010).**

Les exons et l'intron du pré-ARNm sont respectivement représentés par des rectangles rouges et un trait noir. Les snRNA sont représentés en gris, la séquence du site donneur et la séquence du snRNA U1 qui s'y apparie sont représentées par des rectangles oranges. La séquence de la boîte de branchement et la séquence du snRNA U2 qui s'apparie avec elle sont représentées par des rectangles violets. Les séquences impliquées dans la formation du duplex U2/U6 sont représentées par des rectangles jaunes dans chacun des snRNA U2 et U6. La séquence du snRNA U6 s'appariant avec le site donneur est en violet et la séquence formant la tige boucle ISL est en rouge. La séquence du snRNA U5 s'appariant avec le site donneur est en vert. Les hélicases catalysant chacune des transitions structurales sont mentionnées. L'attaque nucléophile du point de branchement est représentée par une flèche rouge.

### III.3.6.6. Le départ de la snRNP U1 et le complexe BΔU1 (Figure I-25, C)

La transition du complexe B vers le complexe Bact est caractérisée par le départ des snRNP U1 et U4 des complexes spliceosomaux (Bessonov *et al.*, 2010). Un complexe intermédiaire, BΔU1, qui correspond au complexe spliceosomal obtenu après le départ de la snRNP U1 a été isolé (Makarova *et al.*, 2004). Ce complexe a été purifié en présence d'héparine par une chromatographie d'affinité avec un anticorps dirigé contre la protéine hPrp31 spécifique de la tri-snRNP (Makarova *et al.*, 2004). La légitimité de ce complexe a été discutée car la dissociation de la snRNP U1 pouvait être due aux conditions stringentes de la purification plutôt qu'être assimilé à un intermédiaire d'assemblage. Néanmoins, ce complexe est capable de catalyser la réaction d'épissage en présence d'un extrait nucléaire dont toutes les particules snRNP ont été enlevées, ce qui indique qu'il est fonctionnel (Makarova *et al.*, 2004).

Le départ du snRNA U1 du site donneur est catalysé par l'hélicase hPrp28 qui est une protéine de la particule snRNP U5, ainsi que par l'hélicase p68 qui est une protéine associée à la snRNP U1 (Ismaili *et al.*, 2001 ; Liu, 2002). L'hélicase hPrp28 déstabilise l'interaction entre la protéine U1-C qui ancre la particule au niveau du site donneur (Chen *et al.*, 2001). En plus de son activité en tant qu'hélicase, la protéine p68 semble également jouer un rôle dans l'intégration de la tri-snRNP en coordonnant le départ du snRNA U1 et son remplacement par le snRNA U6 au niveau du site donneur (Lin *et al.*, 2005). Les appariements du snRNA U1 et U6 sont mutuellement exclusifs car ils s'apparient tous les deux avec la même séquence du site donneur. Après le départ du snRNA U1, le snRNA U6 s'apparie via sa séquence très conservée 5'-ACAGAG-3' aux positions +4 à +6 du site donneur et cet appariement semble s'établir au sein du complexe BΔU1 (Makarova *et al.*, 2004 ; Sawa et Shimura, 1992 ; Wassarman et Steitz, 1992a). Cet appariement a lieu alors que le snRNA U6 et toujours apparié avec le snRNA U4 ainsi qu'avec le snRNA U2 avec lequel il forme l'hélice h II précédemment formée lors du complexe B (Makarova *et al.*, 2004) (**Figure I-25, C**).

### III.3.6.7. Le complexe Bact (Figure I-25, D)

Le complexe Bact est le complexe dans lequel le duplex entre les snRNA U4/U6 est détruit au profit de la formation du duplex entre les snRNA U6/U2. Cela induit le départ du snRNA U4 qui marque le tournant décisif de l'activation catalytique du spliceosome. Les réarrangements structuraux requis pour la formation du nouveau duplex sont catalysés par l'hélicase hBrr2 et la GTPase Snu114 (Bartels *et al.*, 2002 ; Laggerbauer *et al.*, 1998 ; Raghunathan et Guthrie, 1998 ; Small *et al.*, 2006). Afin que l'hélicase hBrr2 rompe le

duplexe U4/U6, il est nécessaire qu'elle soit activée par la GTPase Snu114 liée au GTP (Small *et al.*, 2006).

Le duplex étendu U6/U2 comprend la formation de 3 hélices intermoléculaires, h Ia, h Ib et h III, ainsi que celle d'une hélice intramoléculaire au sein du snRNA U6, U6 ISL (Intramolecular Stem Loop) à laquelle se fixe un ion $Mg^{2+}$ qui est essentiel à la catalyse (Madhani et Guthrie, 1992) (**Figure I-25, D**). La formation de ces hélices inter et intramoléculaires ainsi que la hiérarchie de leur formation n'est pas clairement définie et plusieurs modèles co-existent. Néanmoins, le duplex U4/U6 doit être préalablement rompu pour former le duplex étendu U2/U6 car les hélices intermoléculaires formées dans chacun d'entre eux impliquent les mêmes séquences du snRNA U6. Ainsi, le snRNA U4 joue un rôle d'ARN chaperon qui séquestre les séquences du snRNA U6 qui sont catalytiquement essentielles lors de l'activation du spliceosome.

Le départ du snRNA U4 s'accompagne du départ de nombreuses protéines comme les protéines spécifiques de la di-particule U4/U6, celles de la tri-particule [U4/U6.U5], ainsi que les protéines LSm du snRNA U6 (Bessonov *et al.*, 2010) (**Figure I-24**).

A l'inverse, d'autres protéines sont recrutées ou associées de façon plus stable lors de la formation du complexe Bact, comme le complexe hPrp19/CDC5L (Agafonov *et al.*, 2011 ; Bessonov *et al.*, 2010 ; Makarova *et al.*, 2004) (**Figure I-24**). Ce complexe, préalablement recruté lors de la formation du complexe B, s'associe de façon stable au complexe Bact au sein duquel il facilite les réarrangements et stabilise les appariements du snRNA U6 (Chan et Cheng, 2005 ; Chan *et al.*, 2003). Les protéines de ce complexe s'associent également avec les protéines de la particule snRNP U5, induisant son remodelage pour former une particule U5 35 S (Chan et Cheng, 2005 ; Chan *et al.*, 2003 ; Makarov *et al.*, 2002). Cela renforce aussi les appariements de la boucle I du snRNA U5 avec le site donneur préalablement stabilisés par la protéine hPrp8 (Reyes *et al.*, 1999 ; Sontheimer et Steitz, 1993 ; Teigelkamp *et al.*, 1995 ; Wyatt *et al.*, 1992).

#### III.3.6.7.1. Coordination des activités des hélicases lors de la transition du complexe B vers le complexe Bact (Figure I-26)

Le remplacement du snRNA U1 par le snRNA U6 au site donneur et la rupture du duplex des snRNA U4/U6 sont des évènements qui doivent être coordonnés et qui sont catalysés par les hélicases hBrr2 et hPrp28 dont les activités doivent être régulées afin de ne pas catalyser les réarrangements des snRNA précocement. Ces régulations sont exercées par la protéine hPrp8 qui se situe au cœur catalytique du spliceosome (Grainger et Beggs, 2005). La protéine

hPrp8 possède un domaine RNase-H like qui interagit avec le site donneur et avec l'hélicase hPrp28 afin de favoriser le départ du snRNA U1 et son remplacement par le snRNA U6 (Pena *et al.*, 2008). L'hélicase hBrr2 et sa protéine régulatrice Snu114 interagissent également avec le domaine RNAse-H like et le domaine Jab1/MNP de la protéine hPrp8 (Pena *et al.*, 2007). Cette interaction stimule l'activité de l'hélicase hBrr2 et semble également coordonner la régulation de l'hélicase hBrr2 par la GTPase Snu114 (Maeder *et al.*, 2009).

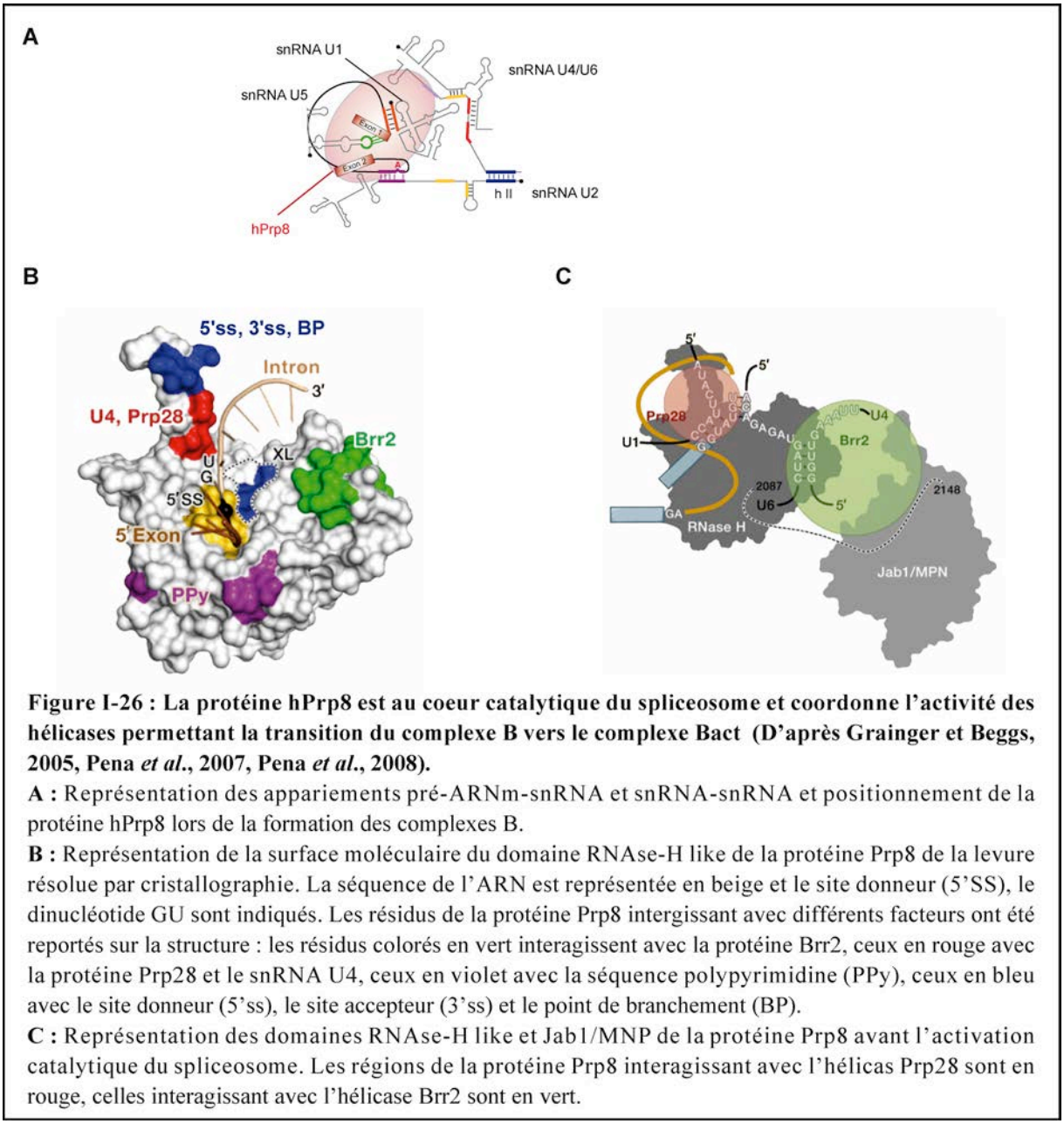



**Figure I-26 : La protéine hPrp8 est au coeur catalytique du spliceosome et coordonne l'activité des hélicases permettant la transition du complexe B vers le complexe Bact (D'après Grainger et Beggs, 2005, Pena *et al.*, 2007, Pena *et al.*, 2008).**

**A :** Représentation des appariements pré-ARNm-snRNA et snRNA-snRNA et positionnement de la protéine hPrp8 lors de la formation des complexes B.

**B :** Représentation de la surface moléculaire du domaine RNAse-H like de la protéine Prp8 de la levure résolue par cristallographie. La séquence de l'ARN est représentée en beige et le site donneur (5'SS), le dinucléotide GU sont indiqués. Les résidus de la protéine Prp8 intergissant avec différents facteurs ont été reportés sur la structure : les résidus colorés en vert interagissent avec la protéine Brr2, ceux en rouge avec la protéine Prp28 et le snRNA U4, ceux en violet avec la séquence polypyrimidine (PPy), ceux en bleu avec le site donneur (5'ss), le site accepteur (3'ss) et le point de branchement (BP).

**C :** Représentation des domaines RNAse-H like et Jab1/MNP de la protéine Prp8 avant l'activation catalytique du spliceosome. Les régions de la protéine Prp8 interagissant avec l'hélicas Prp28 sont en rouge, celles interagissant avec l'hélicase Brr2 sont en vert.

### III.3.6.8. Le complexe B* (Figure I-26, E)

Le complexe B* est le complexe catalytiquement actif obtenu après l'action de l'hélicase hPrp2 (Kim et Lin, 1996). La transition du complexe Bact vers le complexe B* ne provoque pas un grand remaniement dans la composition du spliceosome (Bessonov *et al.*, 2010). L'hélicase hPrp2 déstabilise les complexes hétéromériques SF3a et SF3b de la particule snRNP U2 qui étaient préalablement ancrés au niveau de la boîte de branchement de façon stable (Lardelli *et al.*, 2010 ; Warkocki *et al.*, 2009). Cela permet de libérer le point de branchement qui peut alors réaliser la première attaque nucléophile.

### III.3.7. Le complexe C et la seconde réaction de transestérification

Le complexe C est le complexe obtenu après la première étape de transestérification. Le cœur du complexe contient des quantités équimolaires de snRNA U2, U6 ainsi que de snRNA U5 qui forme avec les protéines du complexe hPrp19/CDC5L la particule U5 35 S (Agafonov *et al.*, 2011 ; Bessonov *et al.*, 2010 ; Golas *et al.*, 2010). Les autres protéines sont des facteurs d'épissage essentiels à la seconde étape de transestérification, des protéines du complexe EJC (Exon Junction Complex) ou du complexe TREX (TRanscription/EXport) (Golas *et al.*, 2010) (**Figure I-27**). Au sein de ce complexe, le snRNA U5, déjà préalablement apparié avec le site donneur (**Figure I-28, A**), s'apparie avec les nucléotides +1 et +2 du site accepteur de façon à aligner les 2 exons lors de leur ligation. La protéine hPrp8, qui interagit également avec les sites d'épissage, ainsi que la protéine hPrp18 contribuent à la stabilité de cet appariement (Chiara *et al.*, 1997 ; Kershaw *et al.*, 2009) (**Figure I-28, B**)

Afin de réaliser la seconde réaction d'épissage, le spliceosome subit un réarrangement conformationnel qui est catalysé par l'hélicase hPrp16 (Burgess *et al.*, 1990 ; Schwer et Guthrie, 1992). Chez la levure, il a été suggéré que ce remodelage enfouisse le site accepteur au coeur du spliceosome et nécessite les facteurs Slu7, prp18, et l'hélicase Prp22 (Schwer et Gross, 1998 ; Schwer et Guthrie, 1992). Cette étape sert de contrôle-qualité et le taux d'hydrolyse de l'ATP par cette hélicase va déterminer si le spliceosome peut s'engager dans la seconde réaction d'épissage ou si la réaction d'épissage va s'arrêter, entraînant la dissociation des composants du spliceosome. Cette hélicase permet ainsi de réguler la fidélité de la réaction d'épissage au niveau du point de branchement et permet d'écarter les substrats possédant un intermédiaire en lariat aberrant (Burgess et Guthrie, 1993, 1993b) (**Figure I-28, B**).

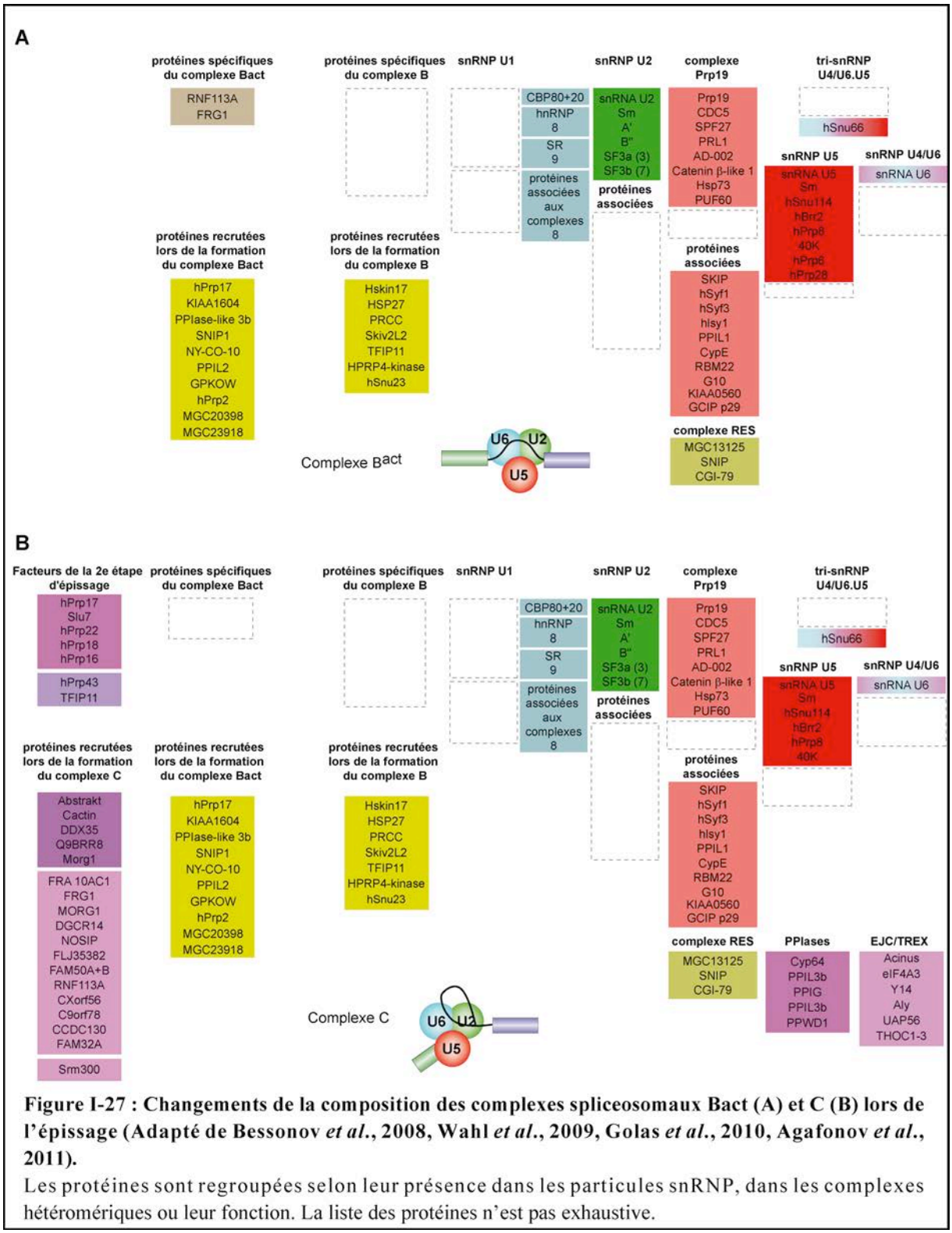



**Figure I-27 : Changements de la composition des complexes spliceosomaux Bact (A) et C (B) lors de l'épissage (Adapté de Bessonov *et al.*, 2008, Wahl *et al.*, 2009, Golas *et al.*, 2010, Agafonov *et al.*, 2011).**

Les protéines sont regroupées selon leur présence dans les particules snRNP, dans les complexes hétéromériques ou leur fonction. La liste des protéines n'est pas exhaustive.

Après la ligation des deux exons lors de la seconde étape de transestérification, l'hélicase hPrp22 provoque le départ de la snRNP U5 de l'ARNm en rompant les interactions de la protéine hPrp8 et les appariements du snRNA U5 avec les séquences exoniques (Schwer, 2008).

L'hydrolyse de l'ATP par l'hélicase hPrp22 provoque le relargage de l'ARNm (Company *et al.*, 1991 ; Schwer, 2008). L'hélicase hPrp22 exerce un contrôle sur la fidélité de la seconde réaction d'épissage. Son hydrolyse de l'ATP, qui provoque le relargage de l'ARNm et du lariat, a lieu uniquement lorsque la seconde réaction d'épissage a eu lieu correctement (Mayas *et al.*, 2006). Enfin, le relargage de l'intron et des particules snRNP U2, U6 et de la particule U5 35 S est catalysée par l'hélicase hPrp43 en collaboration avec la protéine TFIP11 (Martin *et al.*, 2002 ; Yoshimoto *et al.*, 2009). Cette hélicase intervient également dans le processus de fidélité de l'épissage en collaboration avec l'hélicase hPrp16 en assurant le désassemblage des pré-ARNm dont la première étape d'épissage est sous-optimal (Mayas *et al.*, 2010) (**Figure I-28**).

## IV. Intégration de l'épissage dans le processus d'expression des gènes

L'épissage d'un pré-ARNm est une étape qui s'inscrit dans le processus de l'expression d'un gène. Cette étape peut être liée ou couplée aux autres étapes la précédant, comme la transcription, l'ajout de la coiffe, ou lui succédant, comme le clivage à l'extrémité 3', la polyadénylation, l'export, la traduction et la dégradation des ARNm.

### IV.1. Lien entre la transcription et l'épissage

L'épissage des introns de façon co-transcriptionnelle est une règle qui s'applique à la plupart des introns, bien qu'il existe certaines exceptions où l'épissage a lieu après que le gène ait été totalement transcrit. *In vitro*, l'épissage peut avoir lieu indépendamment de la transcription mais son efficacité est plus importante lorsqu'il est précédé d'une étape de transcription, que ce soit *in vitro* ou *in cellulo*, ce qui indique que les deux processus sont couplés (Bird *et al.*, 2004 ; Das *et al.*, 2006). Le domaine C-terminal (CTD) de la large sous-unité de l'ARN polymérase II (Pol II) joue un rôle central dans ce couplage.

Le domaine CTD a la particularité de posséder 52 répétitions d'un héptapeptide, $Y_1S_2P_3T_4S_5P_6S_7$, dont la phosphorylation des sérines affecte les propriétés transcriptionnelles de l'ARN Pol II. Les sérines en positions 2 sont majoritairement phosphorylées lors de l'initiation de la transcription, permettant le recrutement des protéines chargées d'ajouter la coiffe à l'extrémité 5' m7G du pré-ARNm (Buratowski, 2009 ; McCracken *et al.*, 1997).

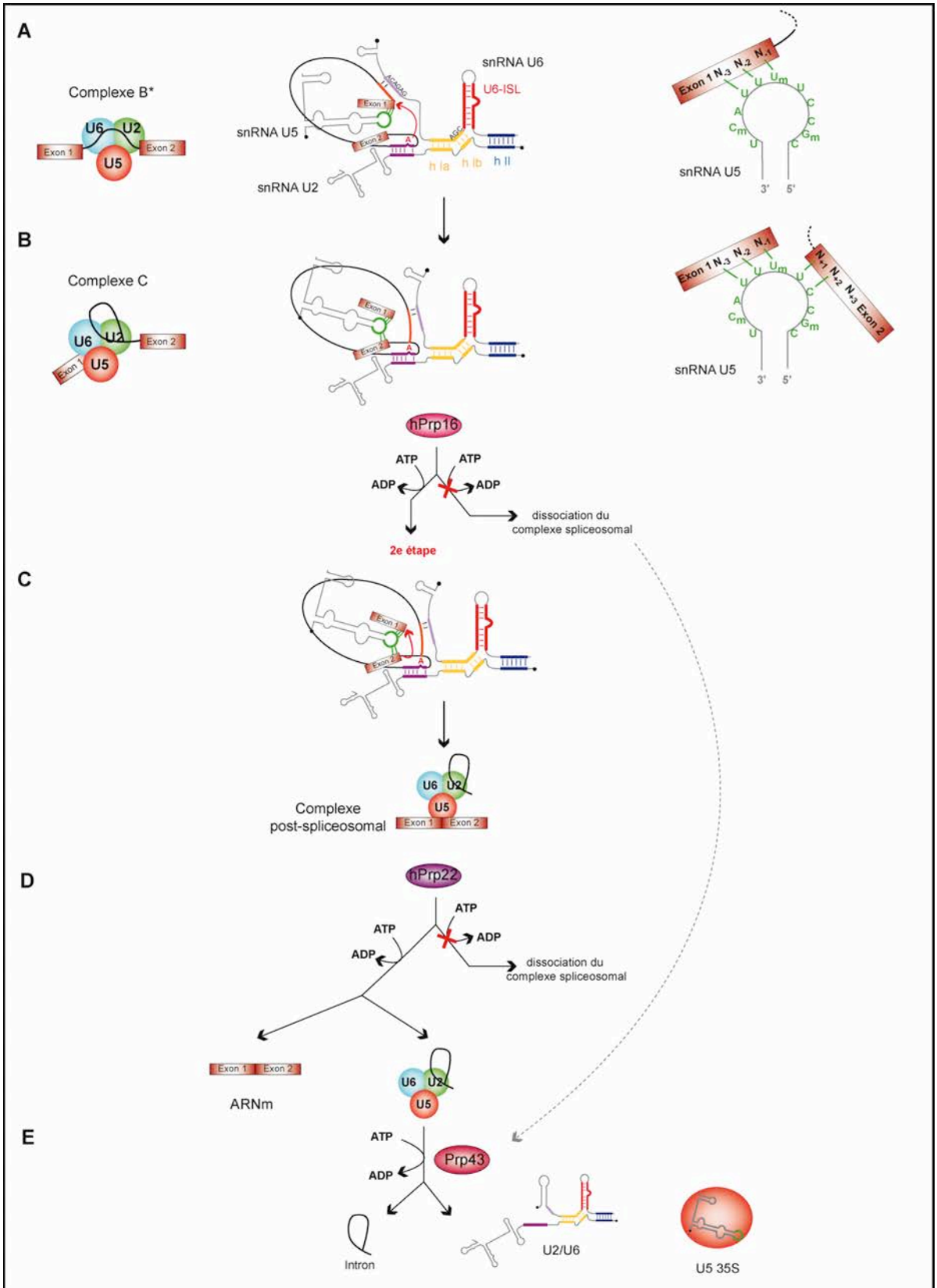



**Figure I-28 : Modèle des réarrangements des snRNA lors de l'activation du spliceosome (Adapté de Mayas *et al.*, 2006).**

Le pré-ARNm et les snRNA sont schématisés selon la même légende que la figure I-25. Les appariements du snRNA U5 avec le pré-ARNm au sein des complexes B* et C sont détaillés à droite des complexes. Les hélicases catalysant chacune des transitions structurales sont mentionnées. La première et la seconde réaction d'épissage sont représentées par une flèche rouge.

La présence de la coiffe augmente l'efficacité d'épissage et, comme nous l'avons vu précédemment, les protéines de la coiffe CBC20 et CBC80 permettent la définition du premier exon lors de l'assemblage des complexes spliceosomaux (Lewis *et al.*, 1996) (**Figure I-12**). Les sérines en position 5 sont majoritairement phosphorylées durant l'élongation de la transcription et sont nécessaires pour activer l'épissage (Buratowski, 2009 ; Hirose *et al.*, 1999). Deux modèles ont été établis pour expliquer le lien entre la transcription et l'épissage (Pandya-Jones, 2011).

D'après le premier modèle, la phosphorylation du domaine CTD permet de recruter des facteurs d'épissage, comme les protéines SR, la snRNP U1 (Das *et al.*, 2007) ou la sous-unité U2AF65 (David *et al.*, 2011). Dans le cas du pré-ARNm IgMA3, la fixation du domaine CTD phosphorylé en aval du site accepteur permet d'activer l'épissage *in vitro* en recrutant directement la sous-unité U2AF65 elle-même en interaction avec le complexe hPrp19 (David *et al.*, 2011). A l'inverse, l'épissage de l'exon EDI (Extra Domain I) du pré-ARNm du gène de la fibronectine est inhibé par la protéine SRp20 qui est recrutée par le domaine CTD phosphorylé (de la Mata et Kornblihtt, 2006).

Le recrutement des différents facteurs d'épissage n'est pas encore bien compris. Il a été suggéré que des protéines jouant un rôle de plateforme se lient directement au CTD phosphorylé pour ensuite recruter les facteurs d'épissage. C'est notamment le cas de la protéine CA150 qui interagit simultanément avec le domaine CTD phosphorylé et le facteur d'épissage SF1/mBBP grâce à son domaine WW (riche en prolines) et FF (riche en phénylalanines), respectivement (Carty *et al.*, 2000 ; Goldstrohm *et al.*, 2001). Le facteur d'élongation de la transcription p-TEFb, interagit avec la protéine TAT-SF1 ainsi que la protéine SKIP qui sont toutes deux impliquées à la fois dans la transcription et l'épissage (Bres *et al.*, 2005 ; Fong et Zhou, 2001 ; Makarova *et al.*, 2004). Les protéines PSF et p54nrb s'associent directement avec les sites 5' et avec le domaine CTD de l'ARN Pol II hyperphosphorylé au sein de larges complexes ribonucléoprotéiques qui contiennent également la snRNP U1 et les facteurs d'élongation de la transcription (Emili *et al.*, 2002 ; Kameoka *et al.*, 2004). Le recrutement des protéines SR n'est pas clairement établi et ne semble pas impliquer une interaction directe ou indirecte avec l'ARN Pol II mais avoir lieu directement au niveau du pré-ARNm naissant (Sapra *et al.*, 2009).

D'après le deuxième modèle, le lien entre l'épissage et la transcription dépend de la vitesse d'élongation de l'ARN Pol II (**Figure I-219**). Ce paramètre n'a pas de répercussion sur

l'épissage des exons constitutifs mais influence l'épissage des exons alternatifs, comme dans le cas où un site accepteur faible est en amont du second site accepteur qui est fort avec lequel il est en compétition. Lorsque l'élongation de la transcription par l'ARN Pol II est rapide, les deux sites accepteurs sont transcrits simultanément et sont en compétition pour être reconnus par le spliceosome. Les complexes spliceosomaux vont alors s'assembler avec le site accepteur le plus fort, ce qui provoque le saut de l'exon ayant le site accepteur faible (**Figure I-219, A**). Lorsque l'élongation par l'ARN Pol II est lente, les deux sites accepteurs ne sont pas en compétition car le site accepteur faible est transcrit et s'assemble avec les complexes spliceosomaux avant que ne soit transcrit le second site accepteur fort situé en aval (**Figure I-29, B**). La vitesse d'assemblage des complexes spliceosomaux sur ces sites accepteurs n'est pas forcément corrélée avec la vitesse d'épissage (de la Mata *et al.*, 2010).

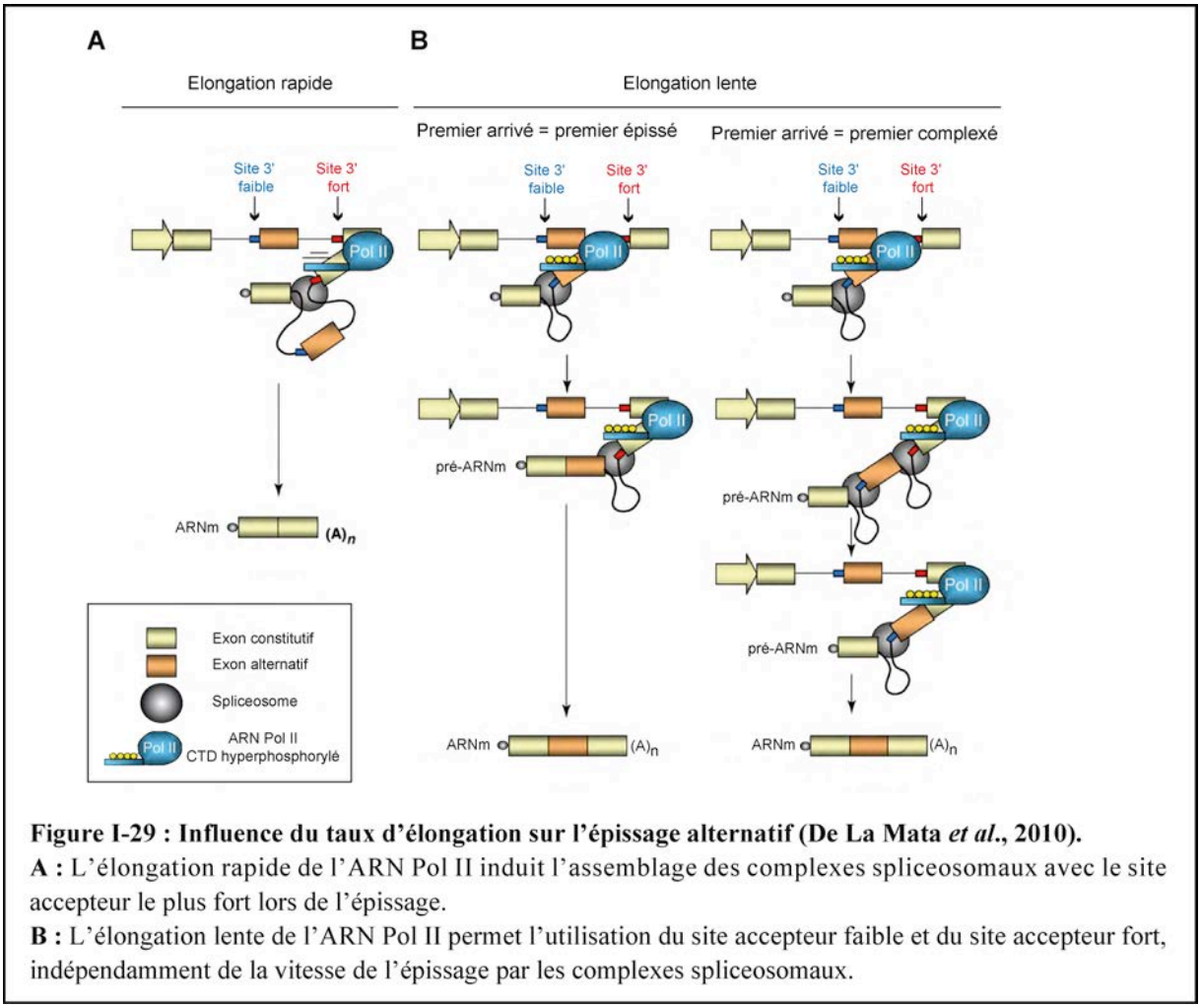



**Figure I-29 : Influence du taux d'élongation sur l'épissage alternatif (De La Mata *et al.*, 2010).**
**A :** L'élongation rapide de l'ARN Pol II induit l'assemblage des complexes spliceosomaux avec le site accepteur le plus fort lors de l'épissage.
**B :** L'élongation lente de l'ARN Pol II permet l'utilisation du site accepteur faible et du site accepteur fort, indépendamment de la vitesse de l'épissage par les complexes spliceosomaux.

### IV.2. Lien entre la polyadénylation et l'épissage

La polyadénylation et l'épissage sont deux processus couplés qui sont mutuellement bénéfiques. Dans un pré-ARNm, l'épissage au site accepteur du dernier exon favorise l'utilisation du site de polyadénylation situé en aval et inversement, la fixation des facteurs de la machinerie de polyadénylation sur ce site de polyadénylation permet la définition et la reconnaissance du dernier exon par la machinerie d'épissage (Cooke *et al.*, 1999 ; Niwa et Berget, 1991) (**Figure I-12**). De nombreuses interactions entre les machineries d'épissage et de polyadénylation ont lieu lors de la définition du dernier exon (**Figure I-30**). Comme l'assemblage du spliceosome, l'assemblage du complexe de polyadénylation/clivage est séquentiel (Millevoi et Vagner, 2010). La fixation de la sous-unité U2AF65 sur la séquence polypyrimidine recrute tout d'abord les facteurs CFIm (Cleavage Factor I) et plus précisément le facteur CFIm59, sur la séquence UGUA en amont du site de polyadénylation (Millevoi *et al.*, 2006). Lorsque le dernier exon possède une séquence ESE, les protéines SR s'y fixant (9G8, SRp20, hTraβ) renforcent ce recrutement en interagissant avec le domaine RS du facteur CFIm68 (Dettwiler *et al.*, 2004 ; Maciolek et McNally, 2007) (**Figure I-30**). La sous-unité U2AF65 interagit ensuite avec la protéine PAP (PolyA Polymérase) qui est recrutée de façon concomitante avec le complexe multiprotéique CPSF (Cleavage Polyadenylation Specificity Factor) qui se fixe sur le signal de polyadénylation (**Figure I-30**). Inversement, lorsque le complexe de polyadénylation/clivage est mis en place avant la définition de l'exon terminal, la protéine PAP favorise le recrutement de la sous-unité U2AF65 sur la séquence polypyrimidine du dernier exon (Vagner *et al.*, 2000). Ce recrutement du complexe CPSF est également facilité par son interaction avec la snRNP U2 lorsque le site accepteur est engagé dans le complexe spliceosomal (Kyburz *et al.*, 2006) (**Figure I-30**). C'est notamment la raison pour laquelle la mutation du point de branchement du dernier exon abolit la polyadénylation (Kyburz *et al.*, 2006).

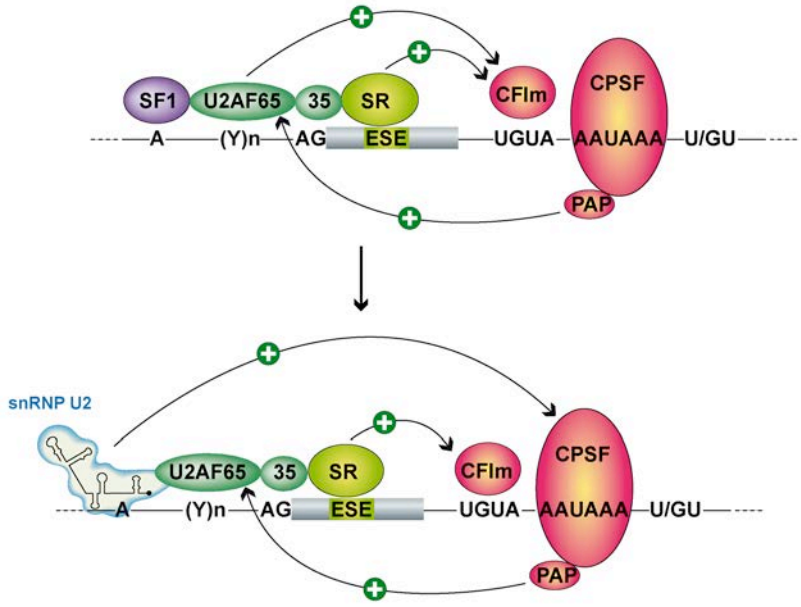



**Figure I-30 : Couplage de l'épissage et de la polyadénylation des pré-ARNm (d'après Martinson, 2010).**
Les facteurs d'épissage activent la polyadénylation en favorisant le recrutement des complexes CFIm et CPSF. Inversement, la protéine PAP aide à la définition de l'exon en favorisant la fixation de la sous-unité U2AF65. La fixation de la snRNP U2 favorise le recrutement du complexe CPSF lorsque le site accepteur est engagé dans le complexe spliceosomal.
L'exon terminal est représenté par un réctangle gris, la lettre A désigne le point de branchement, le dinucléotide AG, le site accepteur, la séquence AAUAAA, le signal de polyadénylation.

En plus de ce couplage, l'utilisation des sites de polyadénylation est régulée par la snRNP U1. *In cellulo*, l'inactivation de la snRNP U1 provoque le clivage et la polyadénylation des pré-ARNm à des sites de polyadénylation cryptiques qui se trouvent souvent dans un intron à moins de 5 kb du site d'initiation de la transcription (Kaida *et al.*, 2010). Le mécanisme de cette régulation n'est pas connu. Néanmoins, il semble que la fixation de la snRNP U1, sur des sites donneurs naturels mais aussi sur les sites donneurs cryptiques, inhibe le clivage et la polyadénylation du pré-ARNm aux signaux de polyadénylation cryptiques (Kaida *et al.*, 2010).

### IV.3. Lien entre l'épissage, l'export et le système NMD, le rôle central du complexe EJC

Le complexe EJC (Exon Jonction Complex) est un complexe protéique déposé 20 à 24 nucléotides en amont de chaque jonction exon-exon après l'épissage et qui reste fixé sur l'ARNm jusqu'au passage des ribosomes lors de la traduction (Moore et Proudfoot, 2009). Le

complexe EJC participe au contrôle qualité de l'ARNm qui est réalisé par le mécanisme NMD (Nonsens-mediated mRNA Decay). Son activité sert à éliminer les ARNm qui possèdent des codons de terminaison de la traduction prématurés (PTC) qui se situent à plus de 50-55 nucléotides en amont de la jonction exon-exon. Le complexe EJC contribue également à l'export des ARNm, et leur traduction (Moore et Proudfoot, 2009).

Le complexe EJC est déposé sur l'ARNm durant l'épissage, lors de la formation du complexe C (Golas *et al.*, 2010 ; Merz *et al.*, 2007) de façon concomitante avec un autre complexe, le complexe TREX qui est recruté sur l'ARNm par la protéine de la coiffe CBC80 (Cheng *et al.*, 2006 ; Masuda *et al.*, 2005) (**Figure I-31**). L'une des protéines du complexe TREX, Aly, recrute ensuite l'hétérodimère TAP-p15 qui est chargé d'exporter le pré-ARNm du noyau vers le cytoplasme (Kohler et Hurt, 2007). Ce rôle peut également être assuré par les protéines SR préalablement présentes sur l'ARNm (Huang *et al.*, 2003). L'hélicase Dbp5 est ensuite recrutée sur le pré-ARNm lors de son passage au travers du pore nucléaire et son hydrolyse de l'ATP dans le cytoplasme provoque le relargage des protéines de l'export (Kohler et Hurt, 2007).

L'ARNm est ensuite traduit par les ribosomes dont le passage provoque le décrochement des complexes EJC. En présence d'un codon stop prématuré, le complexe EJC sert de plateforme pour recruter les facteurs Upf2 et Upf3 du système NMD (**Figure I-31**). Ceux-ci recrutent à leur tour le complexe SURF, entraînant la phosphorylation de l'hélicase Upf1 qui inhibe la traduction et recrute la machinerie de dégradation de l'ARN (Isken *et al.*, 2008 ; Kashima *et al.*, 2006) (**Figure I-31**).

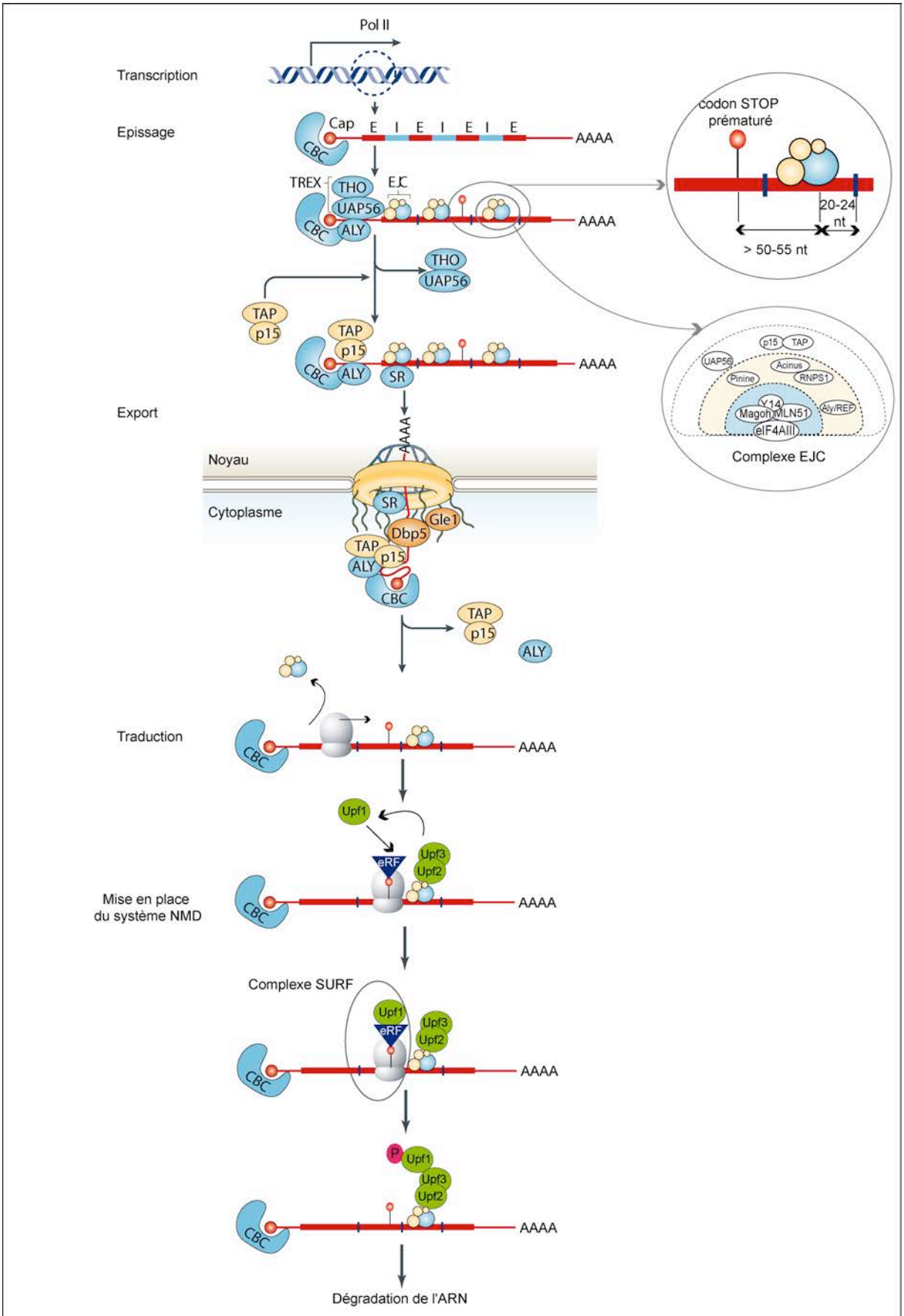



**Figure I-31 : Mise en place du système NMD (Köhler et Hurt, 2007, Isken *et al.*, 2008).**

# Chapitre II : Régulation de l'épissage alternatif des ARN pré-messagers

## I. Généralités sur l'épissage alternatif

Chez les eucaryotes supérieurs, la grande majorité des gènes possèdent des introns, faisant de l'épissage un processus majeur de leur expression. Ces évènements d'épissage sont le plus souvent alternatifs car les sites d'épissage ne sont pas systématiquement utilisés lors de la réaction d'épissage en raison de leurs séquences dégénérées par rapport aux séquences consensus. Les épissages alternatifs ont lieu le plus souvent dans les régions codantes, ce qui permet d'augmenter de façon considérable la capacité codante du génome. D'après les récentes estimations basées sur l'analyse du transcriptome humain, au moins une isoforme de protéine serait produite par plus de 90 % de la totalité des gènes épissés alternativement (Wang *et al.*, 2008). L'épissage alternatif est réalisé selon plusieurs mécanismes basiques qui sont présentés dans la **Figure II-1**.

### I.1. Les différents types d'épissage alternatifs

Un exon « cassette » peut être soit inclus soit exclu de l'ARNm. Ce type d'exon est en général de petite taille et son insertion dans l'ARNm est souvent tissu spécifique, ce qui permet de produire une isoforme de la protéine codée (**Figure II-1, A**).

Des exons d'un même pré-ARNm sont dits mutuellement exclusifs lorsqu'ils ne sont jamais présents simultanément dans le même ARNm (**Figure II-1, B**). Les deux couples de site 5'-3' sont en général tous les deux fonctionnels et sont en compétition lors de l'épissage où l'utilisation d'un couple exclu l'utilisation de l'autre couple.

Un pré-ARNm peut contenir plusieurs sites 5' alternatifs qui sont en compétition les uns avec les autres lors de l'épissage (**Figure II-1, C**) et cela peut également s'appliquer aux sites 3' (**Figure II-1, D**). Ces deux cas de figures sont souvent retrouvés chez les virus afin de produire de nombreux ARNm différents à partir d'un seul pré-ARNm.

L'absence d'épissage entre un site 5' et un site 3' aboutit à la rétention de l'intron qui se retrouve dans l'ARNm (**Figure II-1, E**). Chez l'Homme, ce phénomène est souvent à l'origine d'une pathologie liée à des mutations des sites d'épissage ou un défaut de la machinerie d'épissage.

L'utilisation alternative de promoteurs transcriptionnels et des sites de polyadénylation influence le choix l'épissage des exons terminaux de l'ARNm (**Figure II-1, F, G**).

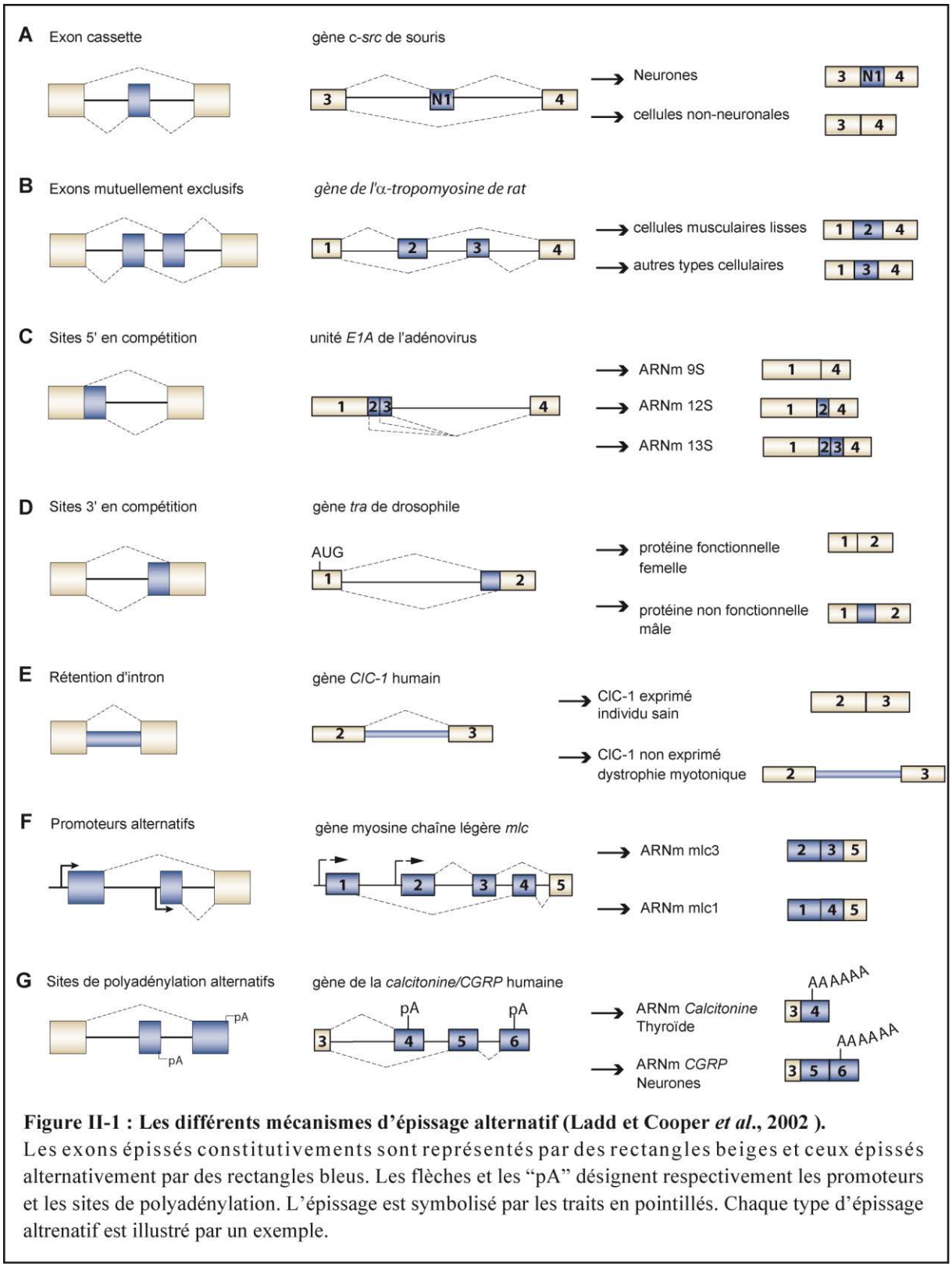



**Figure II-1 : Les différents mécanismes d'épissage alternatif (Ladd et Cooper *et al.*, 2002 ).**
Les exons épissés constitutivements sont représentés par des rectangles beiges et ceux épissés alternativement par des rectangles bleus. Les flèches et les "pA" désignent respectivement les promoteurs et les sites de polyadénylation. L'épissage est symbolisé par les traits en pointillés. Chaque type d'épissage altrenatif est illustré par un exemple.

L'épissage alternatif d'un pré-ARNm peut combiner plusieurs de ces mécanismes. Ce choix des sites d'épissage est par ailleurs régulé par de nombreux paramètres comme le type cellulaire, l'état de différenciation de la cellule ou l'activation des voies de signalisation. L'épissage alternatif constitue donc pour la cellule un excellent mécanisme flexible permettant de contrôler et adapter l'expression des gènes en réponse à un contexte donné. Cette régulation est à la base permise par le caractère sous-optimal des sites d'épissage dont l'utilisation lors de l'épissage est régulée par des éléments de séquences *cis* régulatrices sur lesquelles viennent se fixer des facteurs de régulation en *trans*.

### I.2. Les sites d'épissage alternatif sont souvent sous-optimaux

Les sites d'épissage forts utilisés constitutivement possèdent des séquences très similaires aux séquences consensus établies pour les sites 3' et 5' d'épissage tandis que les sites alternatifs sont souvent sous-optimaux car ils possèdent des séquences qui sont dégénérées par rapport à ces séquences consensus.

La force d'un site 5' est définie principalement par le degré de complémentarité de sa séquence avec celle du snRNA U1 avec lequel elle s'apparie (Roca *et al.*, 2005) (**Figure II-2**). La force d'un site 3' dépend, en plus de la séquence du site 3' proprement dite, située à la bordure intron/exon, de la séquence polypyrimidine et de la séquence de branchement. Les sites 3' faibles ont souvent un enchaînement de pyrimidines entrecoupé de purines, et/ou une séquence de branchement sous-optimale où le point de branchement n'est pas le résidu Adénosine canonique (**Figure II-2**). Cela défavorise leur reconnaissance respective par les facteurs U2AF65 et SF1 puis la snRNP U2 qui s'y fixent lors de la reconnaissance du site 3' (Corioni *et al.*, 2011 ; Zamore *et al.*, 1992).

### I.3. L'utilisation des sites d'épissage faibles est souvent régulée par la combinaison de séquences *cis* auxquelles se fixent des facteurs *trans*

Lorsque les séquences des sites d'épissage alternatifs sont dégénérées, l'assemblage des complexes spliceosomaux est guidé et facilité par des éléments de séquences *cis* sur lesquelles viennent se fixer des facteurs *trans*. Ces séquences activatrices d'épissage peuvent être exoniques, ce sont les séquences ESE (Exonic Splicing Enhancer), ou introniques, appelées séquences ISE (Intron Splicing Element) (**Figure II-2**). Ces séquences sont reconnues par des facteurs agissant en *trans* qui sont en général des protéines SR (Chen et Manley, 2009) (**Figure II-2**).

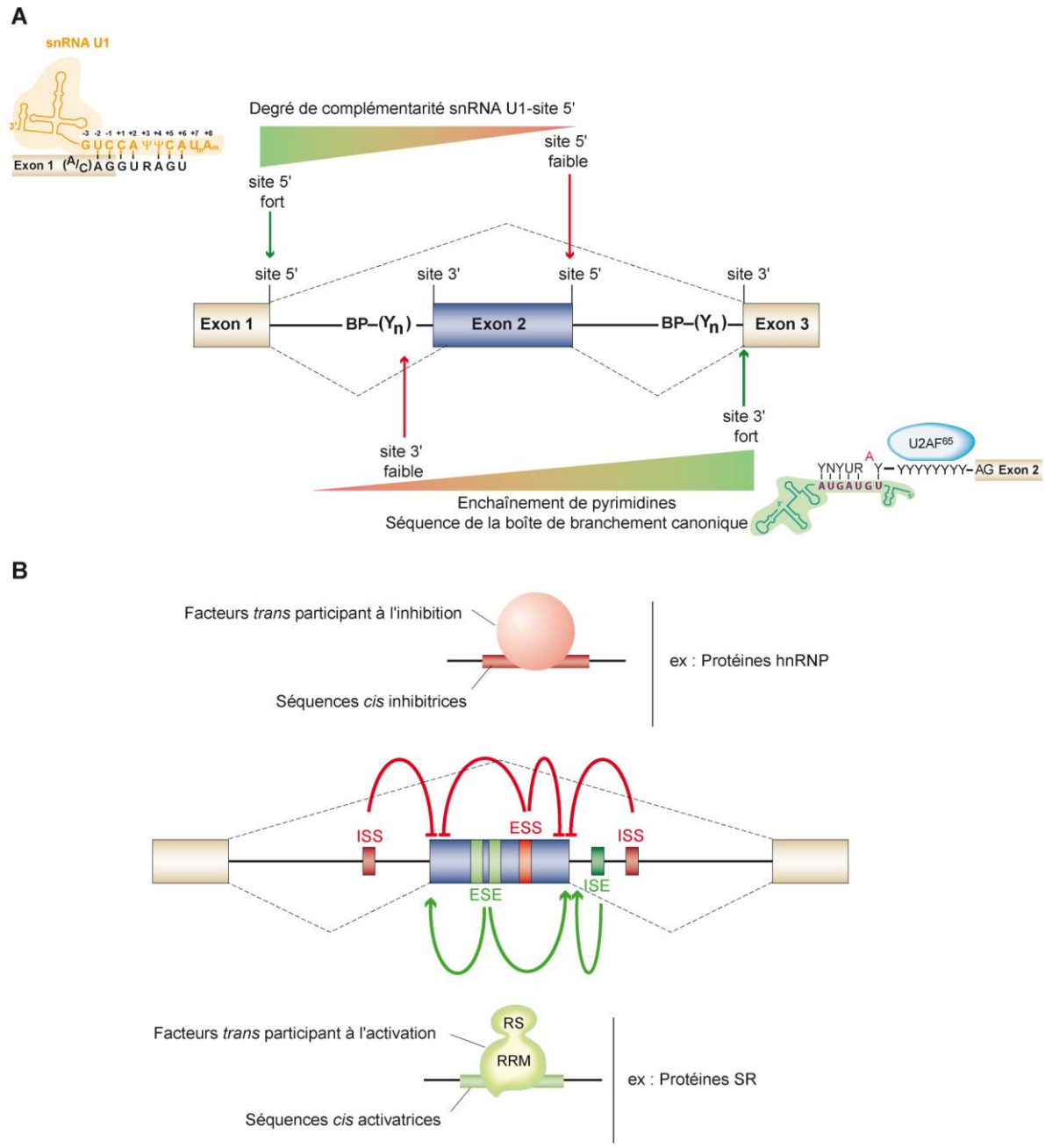



**Figure II-2 : Les sites d'épissage alternatifs sont souvent faibles et leur utilisation est régulée par des séquences *cis*.**

**A :** Caractéristiques des sites d'épissage constitutifs et alternatifs. Les exons épissés constitutivement (rectangles beiges) et alternativement (rectangles bleus) diffèrent par le degré de complémentarité entre le site 5' et la séquence du snRNA U1 ainsi que par la richesse en enchaînement de pyrimidines et la séquence de la boîte de branchement du site 3'.

**B :** Le choix des sites d'épissage alternatifs est régulé par des séquences *cis* sur lesquelles se fixent des facteurs *trans*. Les séquences *cis* activatrices (rectangles verts), fixent le plus souvent des protéines SR tandis que les séquences inhibitrices (rectangles rouges), fixent en général des protéines hnRNP.

A l'inverse, l'utilisation des sites d'épissage peut être réprimé par des séquences *cis* inhibitrices, se situant dans l'exon, désignées séquence ESS (Exonic Splicing Silencer) ou dans l'intron, appelées séquences ISS (Intronic Splicing Silencer) (**Figure II-2**). Les facteurs *trans* se fixant sur ces séquences appartiennent en général à la famille des protéines hnRNP (Chen et Manley, 2009) (**Figure II-2**).

Ces séquences *cis* ne sont pas spécifiquement associées à des sites alternatifs et leur distribution varie en fonction des exons constitutifs, alternatifs et des introns. Les introns contiennent en majorité des séquences inhibitrices, les exons constitutifs renferment une plus forte proportion de séquences activatrices tandis que les exons alternatifs possèdent autant de séquences activatrices qu'inhibitrices (Blencowe, 2006).

Nous décrirons l'action des protéines hnRNP et des protéines SR lors de l'épissage alternatif puis nous aborderons le rôle joué par d'autres régulateurs de l'épissage ainsi que d'autres paramètres qui modulent également le choix des sites d'épissage.

## II. Les protéines hnRNP et leur rôle dans l'épissage

Les protéines hnRNP (heterogeneous RiboNucleoprotein Particle) sont extrêmement abondantes dans les cellules et leur caractéristique principale est de se lier aux ARN, et ce, dès leur transcription par l'ARN Polymérase II. Plus de 20 protéines hnRNP ont été identifiées et classées par ordre alphabétique de A à U (Han *et al.*, 2010) (**Figure II-3**). Ces protéines interviennent dans les nombreux processus cellulaires impliquant l'ARN, comme la transcription, l'export, la stabilité, la traduction, ou la polyadénylation. Nous nous intéresserons plus particulièrement aux protéines de la famille des protéines hnRNP A/B et H/F qui sont impliquées dans la régulation de l'épissage alternatif.

### II.1. Organisation structurale des protéines hnRNP

Les protéines hnRNP sont des protéines modulaires qui possèdent au moins un domaine de liaison à l'ARN (domaines RRM, KH, RGG) ainsi que des domaines auxiliaires qui diffèrent d'une protéine à l'autre, comme les domaines riches en glycines. Certaines protéines existent sous différentes isoformes qui ont des propriétés distinctes et qui sont le résultat d'un épissage alternatif, comme les protéines de la famille hnRNP A/B.

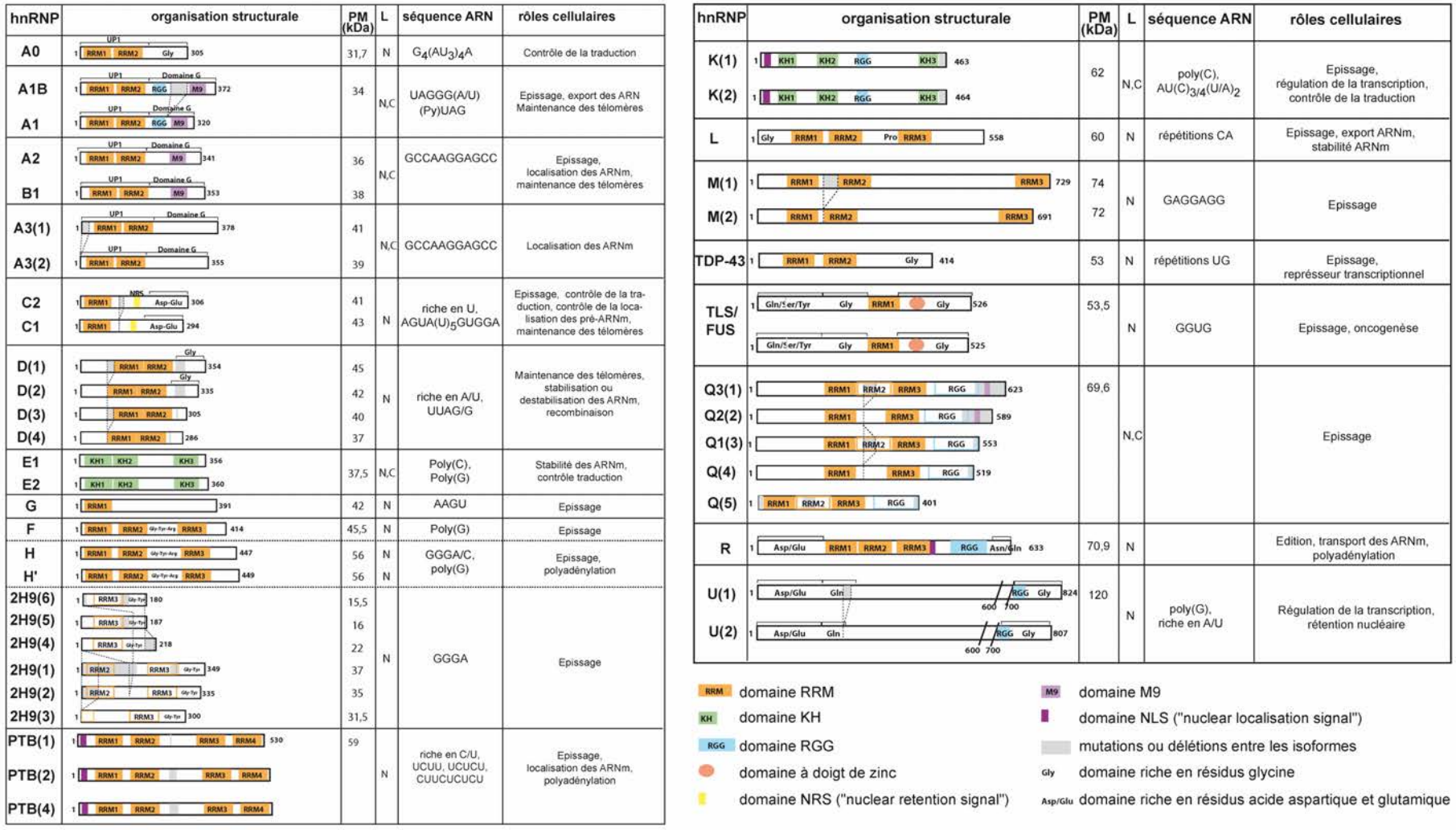



**Figure II-3 : Structure, rôle et séquences ARN reconnues par les protéines de la famille hnRNP (d'après Han *et al.*, 2010 ; Swanson et Dreyfuss, 1988 ; Dreyfuss *et al.*, 1993 ; Honoré *et al.*, 1999 ; Caputi et Zahler, 2001 ; Lerga *et al.*, 2001 ; Kafasla *et al.*, 2002 ; Krécic et Swanson, 1999; Drefuss *et al.*, 2002).**

Les différentes protéines hnRNP sont représentées. Chaque isoforme connue pour chaque protéine hnRNP est donnée, ainsi que son organisation structurale à l'échelle, son poids moléculaire prédictif (PM), sa localisation (L) nucléaire (N) et/ou cytoplasmique (C), la séquence ARN reconnue et sa fonction.

### II.1.1. Les domaines de liaison à l'ARN

#### II.1.1.1. Le domaine RRM

Le domaine RRM est le plus commun et le mieux caractérisé des domaines de liaison à l'ARN (Lunde *et al.*, 2007). Ce domaine se compose de 80-90 résidus qui se structurent en 4 brins β antiparallèles sur lesquels viennent se placer 2 hélices α selon une topologie de type $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ (Oubridge *et al.*, 1994) (**Figure II-4**). Les résidus interagissant avec l'ARN font partie de séquences signatures très conservées appelées RNP1 et RNP2 qui se situent respectivement dans les brins centraux β3 et β1 (Oubridge *et al.*, 1994) (**Figure II-4**). Cette interaction s'établie avec 4 nucléotides et peut s'étendre à 8 nucléotides lorsque les boucles reliant les brins β et d'autres éléments structuraux secondaires non canoniques sont impliqués dans l'interaction (Lunde *et al.*, 2007).

Contrairement à la famille des protéines hnRNP A/B, les protéines de la famille hnRNP H/F ne possèdent pas des domaines RRM canoniques. Leurs domaines RRM, désignés qRRM (quasi RRM) adoptent des structures identiques aux RRM canoniques mais ne possèdent pas les séquences signatures RNP1 et RNP2 (Dominguez *et al.*, 2010). Cela change considérablement le mode d'interaction du domaine avec l'ARN. Ainsi, contrairement aux domaines RRM canoniques où l'interaction avec l'ARN a lieu au niveau des brins β, celles des domaines qRRM ont lieu au niveau des 3 boucles très conservées (Dominguez *et al.*, 2010) (**Figure II-4**).

#### II.1.1.2. Le domaine KH

Seules les protéines hnRNP K et E1, E2, qui ne possèdent pas de domaine RRM, mais possèdent un domaine KH (K Homology). Ce domaine d'environ 70 résidus se compose de 3 brins antiparallèles sur lesquels viennent se placer 3 hélices α selon une topologie de type $\beta_1\alpha_1\alpha_2\beta_2\beta'\alpha'$ (Valverde *et al.*, 2008). Il comporte également une séquence signature de 8 nucléotides située au centre du domaine et contacte l'ARN sur une longueur de 4 nucléotides qui viennent s'intercaler dans l'espace défini entre la boucle GXXG, les hélices α1, α 2 et le brin β2 (Valverde *et al.*, 2008) (**Figure II-5**).

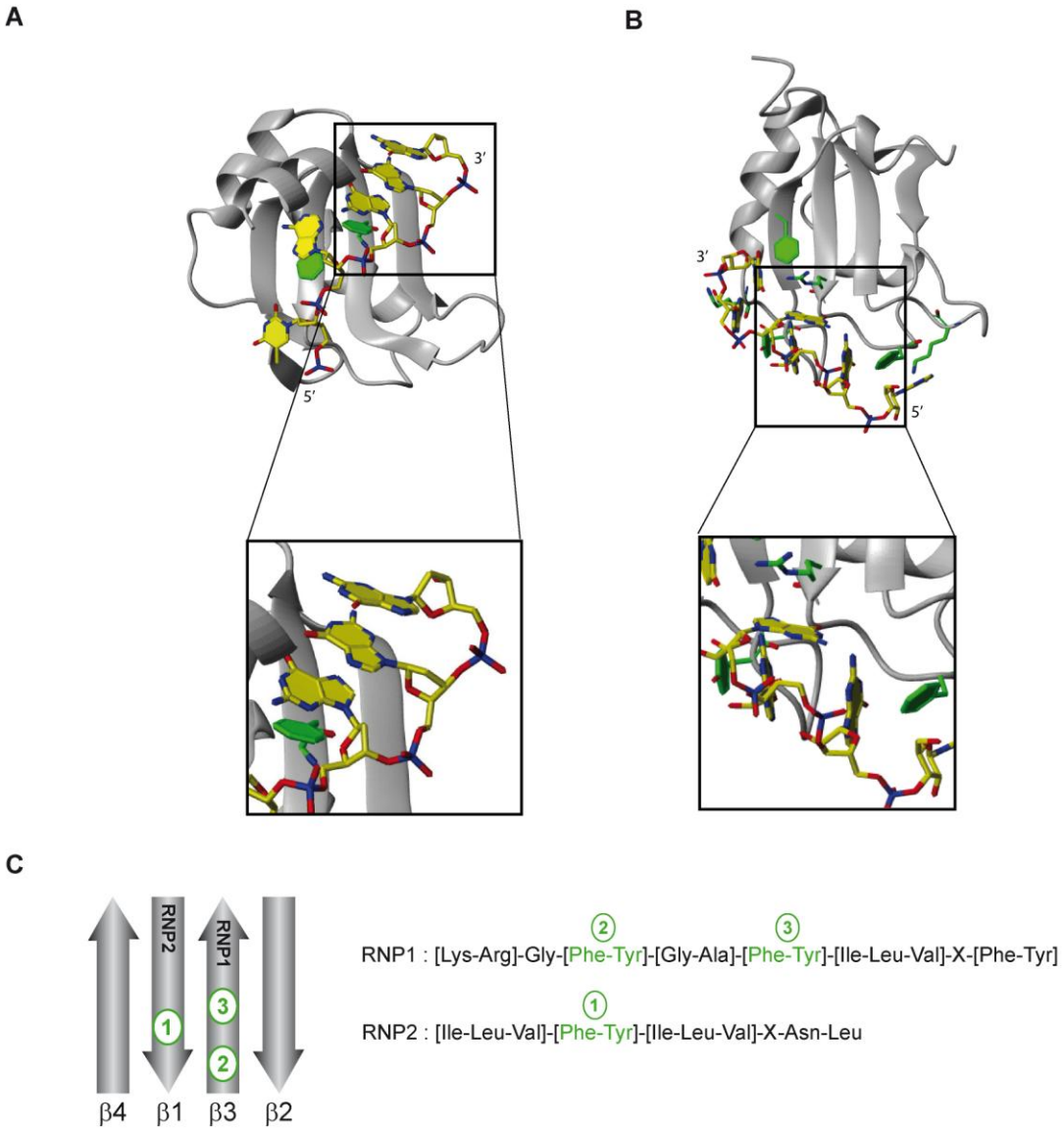



**Figure II-4 : Structures du domaine RRM2 de la protéine hnRNP A1 et du domaine qRRM2 de la protéine hnRNP F (Dominguez *et al.*, 2010).**

**A :** Structure du domaine RRM2 de la protéine hnRNP A1 en interaction avec la séquence AGGGAU résolue par RMN.

**B :** Structure du domaine qRRM2 de la protéine hnRNP F en interaction avec la séquence TAGGG résolue par RMN.

**C :** Représentation des 4 brins β antiparallèles du domaine RRM de la protéine hnRNP A1. Les séquences signatures conservées RNP1 et RNP2 y sont reportées ainsi que les résidus aromatiques impliqués dans l'interaction avec l'ARN symbolisés par des cercles notés 1, 2, 3. Leurs séquences sont détaillées sur la droite.

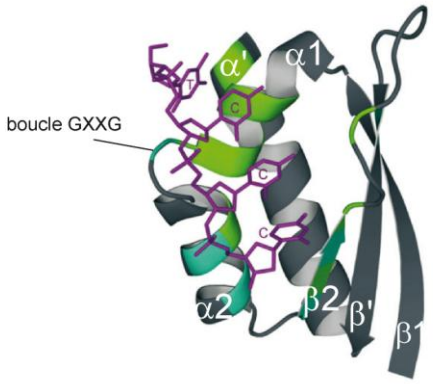



**Figure II-5 : Structure du troisième domaine KH de la protéine hnRNP K en interaction avec un ADN simple brin (Valverde *et al.*, 2008).**
La séquence ADN dTCCC est représentée en violet, les régions de la protéine en contact avec l'ARN sont colorées en vert et en bleu. La boucle GXXG, les hélices et les brins du domaine KH sont mentionnés.

### II.1.1.3. Le domaine RGG ou GAR

Ce domaine est composé d'environ 36 résidus et se caractérise par la présence de plusieurs répétitions Arg-Gly-Gly (RGG). Ce domaine permet à la protéine hnRNP U qui ne possède pas de domaine RRM, d'interagir avec l'ARN (Kiledjian et Dreyfuss, 1992). Le domaine RGG fait par ailleurs l'objet de nombreuses modifications post-traductionnelles, comme la méthylation des résidus arginines.

### II.1.2. Le domaine d'import (M9) et de rétention dans le noyau (NRS : Nuclear Retention Signal)

Bien que les protéines hnRNP soient localisées principalement dans le nucléoplasme des cellules, certaines d'entre elles transitent du noyau vers le cytoplasme puis réintègrent ensuite le noyau (Pinol-Roma et Dreyfuss, 1992). Dans le noyau, ces protéines sont associées aux pré-ARNm et dans le cytoplasme, aux ARNm. Les protéines hnRNP réintègrent ensuite le noyau grâce à leur domaine M9 qui leur permet d'interagir avec des facteurs d'imports, les transportines 1 et 2 (Rebane *et al.*, 2004). L'import de la protéine hnRNP A1 est par ailleurs régulé par une phosphorylation d'un peptide, le peptide F, situé juste en aval de la séquence M9. Cette phosphorylation gêne l'interaction de la protéine hnRNP A1 avec la transportine 1 et provoque une diminution de son import nucléaire (Allemand *et al.*, 2005). L'export des protéines hnRNP H et F du noyau vers le cytoplasme est lié à la transcription des ARN (Van Dusen *et al.*, 2010). Ces protéines retournent ensuite dans le noyau grâce à leur domaine Gly central qui interagit également avec la transportine 1 (Van Dusen *et al.*, 2010). A l'inverse, d'autres protéines ne sortent pas du noyau, comme les protéines hnRNP C et U, qui possèdent

un signal de rétention nucléaire (NRS Nuclear Retentions Signal) (Pinol-Roma et Dreyfuss, 1992).

### II.1.3. Le domaine d'oligomérisation ou domaine Gly et son rôle dans la fonction des protéines des familles hnRNP A/B et H/F

Le domaine Gly est présent dans de nombreuses protéines hnRNP (F, G, Q, D, U et R) et son rôle a été principalement étudié chez les protéines des familles hnRNP A/B et H/F. Le domaine C-terminal des protéines de la famille hnRNP A/B renferme une forte concentration en résidus glycines (40 %). Ce domaine, très flexible, est responsable de la fixation coopérative de plusieurs molécules de protéines hnRNP A1 sur les ARN (Cartegni *et al.*, 1996 ; Okunola et Krainer, 2009). Cette fixation coopérative serait réalisée par l'intermédiaire d'interactions hydrophobes établies entre les nombreux résidus phénylalanines et tyrosines intercalés entre les résidus glycines (Cartegni *et al.*, 1996). Ce domaine est également présent dans les protéines de la famille hnRNP H/F et est nécessaire à l'interaction entre deux protéines hnRNP H (Fisette *et al.*, 2010). L'implication du domaine Gly dans la fixation des protéines hnRNP sur l'ARN semble plus important pour les protéines de la famille hnRNP A/B que celles de la famille hnRNP H/F. La protéine UP1, qui contient les 2 RRM de la protéine hnRNP A1 mais pas le domaine Gly, a une affinité beaucoup plus faible pour l'ARN que celle de la protéine hnRNP A1 entière (Mayeda *et al.*, 1994 ; Mayeda *et al.*, 1998) (**Figure II-6**). En revanche, la région Gly-qRRM3-Gly peut être tronquée de la protéine hnRNP H sans affecter sa capacité de fixation à l'ARN et un seul domaine qRRM (qRRM1 ou qRRM3) de la protéine hnRNP F est suffisant pour fixer et réguler l'épissage du pré-ARNm *Bcl-x* (Dominguez *et al.*, 2010 ; Fisette *et al.*, 2010) (**Figure II-6**). Ce domaine permet également aux protéines hnRNP d'interagir avec les autres protéines hnRNP possédant un domaine Gly, comme c'est le cas des protéines hnRNP A1 et H dont les domaines Gly sont interchangeables (Fisette *et al.*, 2010).

L'ensemble des propriétés conférées aux protéines hnRNP par le domaine Gly sont essentielles pour la mise en place de certaines régulations de l'épissage par les protéines hnRNP A1, H et F qui nécessitent une oligomérisation ou une interaction entre ces protéines et que nous illustrerons par la suite (Fisette *et al.*, 2010 ; Martinez-Contreras *et al.*, 2007 ; Mayeda *et al.*, 1994 ; Okunola et Krainer, 2009).

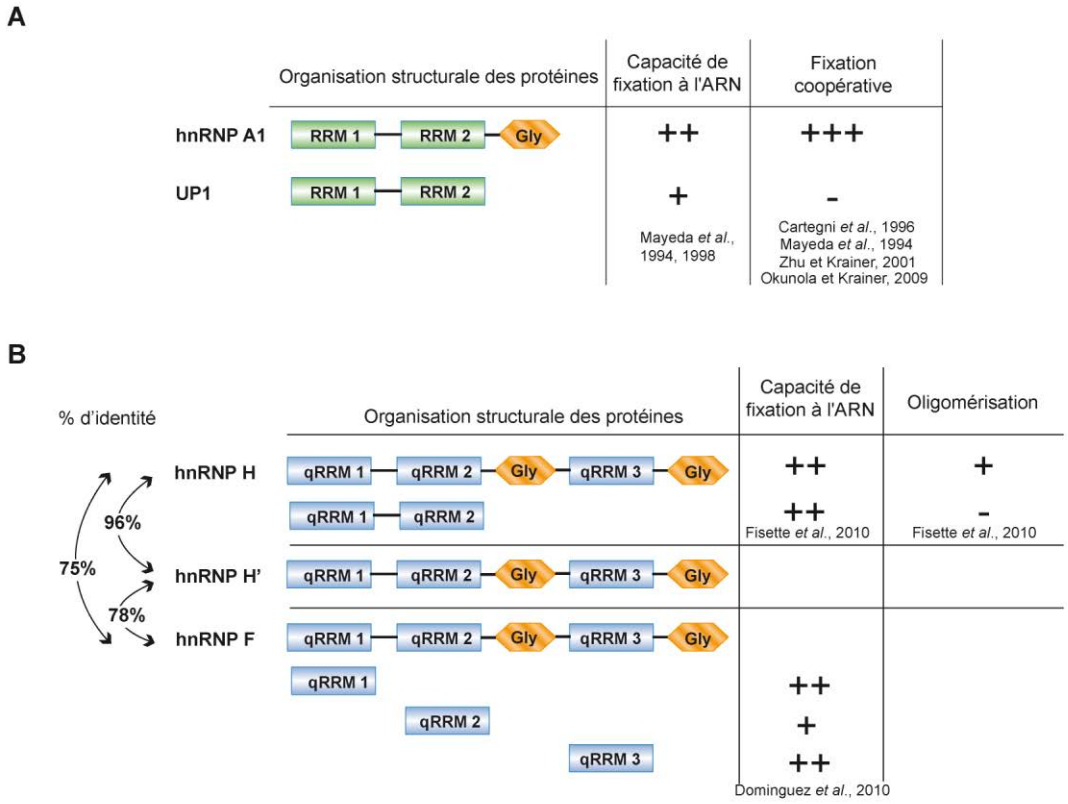



**Figure II-6 : Structures et importances fonctionnelles des différents domaines des protéines hnRNP A1, UP1 (A) et des protéines de la famille des hnRNP H/F (B).**

## II.2. Les séquences ARN reconnues par les protéines des familles hnRNP A/B et hnRNP H/F

L'existence de différents domaines de liaison à l'ARN ajoutée aux nombreuses modulations des domaines des protéines hnRNP changent considérablement leur capacité à reconnaître une séquence ARN. De manière générale, les protéines hnRNP ont toutes une affinité pour l'ARN avec néanmoins une affinité plus forte pour certains type de séquences : séquences riches en C pour les protéines hnRNP E, I et K, séquences riches en G pour les protéines hnRNP A/B, E, H, H', 2H9, M et U, séquences riches en U pour les protéines hnRNP C et PTB.

La protéine hnRNP A1 se fixe préférentiellement à des séquences ARN simple brin de type UAGRG(U/A) (R : purine), présentant des similitudes avec certains sites 5' et 3' d'épissage (Burd et Dreyfuss, 1994). Il existe de nombreux autres sites de fixation se rapprochant de cette séquence, comme les séquences régulatrices de l'épissage contenant un motif PyUAGPu ainsi que des séquences plus divergentes contenant des triplets de G (Chabot *et al.*, 1997 ; Han *et al.*, 2005 ; Hutchison *et al.*, 2002) (**Figure II-7, A**).

De manière générale, toutes les protéines de la famille hnRNP H/F se fixent à des séquences ARN simple brin riches en G contenant au moins un triplet de G. Toutes les protéines reconnaissent la séquence de type GGGA et les protéines hnRNP H/H' ont la particularité de se fixer aux séquences GGGC (Caputi et Zahler, 2001 ; Dominguez et Allain, 2006 ; Dominguez *et al.*, 2010 ; Han *et al.*, 2005) (**Figure II-7, B**).

| Protéine SR | Site de fixation | pré-ARNm | Référence |
|---|---|---|---|
| hnRNP A1  | UAGACA | SMN2 exon 7 | Kashima et Manley, 2003 |
| | UAGACU | VIH Tat exon 2 | Si *et al.*, 1997 ; Bilodeau *et al.*, 2001 |
| | UAGACA | CD44 exon v5 | Matter *et al.*, 2000 |
| | UAG:GAGGAAGGU | c-src exon N1 | Rooke *et al.*, 2003 |
| | UAG | hnRNP A1 exon 7b | Chabot *et al.*, 2003, 1997 |
| | UAGACU | hnRNP A1 exon 7b | Chabot *et al.*, 2003, 1997 |
| | AGGGA | c-H-ras exon IDX | Guil *et al.*, 2003 |
| | UAGGGCAGGC | FGFR2 exon K-SAM | Del Gatto *et al.*, 1996 |
| | CAGGGUUGAGGGGAGCAGGGU | β Tropomyosine de poulet exon 6/7 | Expert Bezançon *et al.*, 2004 |

| Protéine SR | Site de fixation | pré-ARNm | Référence |
|---|---|---|---|
| hnRNP H  | UGGGU | VIH Tat exon 2 | DelGatto-Konczak *et al.*, 1999<br>Jacquenet *et al.*, 2001<br>Caputi *et al.*, 1999 |
| | GGGGU | NF-1 exon 3 | Buratti *et al.*, 2009 |
| | GGGGGCUG | c-src exon N1 | Min *et al.*, 1995 |
| | GUGGG | β Tropomyosine de rat exon 6/7 | Chen *et al.*, 1999 |
| | GGGC | | Caputi et Zahler, 2001 |
| | GGGA<br>GGGU<br>GGGC | NRS | McNally *et al.*, 2006 |

**Figure II-7 : Liste non exhaustives des sites de fixation reconnus par la protéine hnRNP A1 (A) et la protéine hnRNP H (B).**

## II.3. Rôles et mécanismes généraux des protéines hnRNP dans la régulation de l'épissage alternatif

Les protéines hnRNP, et plus particulièrement les protéines des familles hnRNP A/B et H/F, jouent un rôle majeur dans la régulation de l'épissage alternatif. Environ 10% des évènements d'épissage alternatifs seraient contrôlés par la protéine hnRNP A1 et la protéine hnRNP H (Venables *et al.*, 2008). D'autres études abondent dans ce sens en démontrant qu'il existe une corrélation entre la force du site 5' et l'enrichissement 70 nt en aval dans l'intron de triplets de G, qui sont des sites de fixation potentiels des protéines hnRNP A1 et hnRNP H/F (Han *et al.*, 2005 ; Xiao *et al.*, 2009 ; Yeo *et al.*, 2004). Ces séquences sont également retrouvées en aval de 17 % des sites donneurs mineurs (McNally *et al.*, 2006). Lors de l'épissage alternatif, les protéines hnRNP sont le plus souvent des inhibiteurs de l'épissage mais elles peuvent également l'activer.

Comme nous l'avons vu dans la première partie de l'introduction, la définition des exons ou des introns est une étape clé de la réaction d'épissage lors de laquelle les sites 5' et 3' sont reconnus par les composants du spliceosome (**Chapitre I, III.3.2, Figure I-12**). Cette étape est fortement régulée et peut être inhibée par les protéines hnRNP (**Figure II-8**).

Les protéines hnRNP peuvent inhiber la définition de l'exon quand celui possède des séquences *cis* ESS sur lesquelles elles se fixent (Martinez-Contreras *et al.*, 2007) (**Figure II-8, A**). Placée à proximité du site donneur, la fixation des protéines hnRNP sur la séquence ESSS empêche la fixation du facteur U2AF sur le site 3' par encombrement stérique. Placée à proximité du site 5', cela empêche la fixation de la snRNP U1. Enfin, si la séquence ESS chevauche ou se situe à proximité d'une séquence ESE, les protéines hnRNP empêchent la fixation de facteurs *trans* activateurs d'épissage qui sont souvent des protéines SR. Dans ce cas, il y a une compétition entre les facteurs *trans* activateurs et inhibiteurs. Par ailleurs, ces différents modes d'inhibition peuvent être exercés soit par une fixation directe et isolée de la protéine hnRNP sur la séquence ESS, soit par son oligomérisation à partir du site initial de fixation qui déplace les autres facteurs préalablement fixés sur l'exon (**Figure II-8, A**).

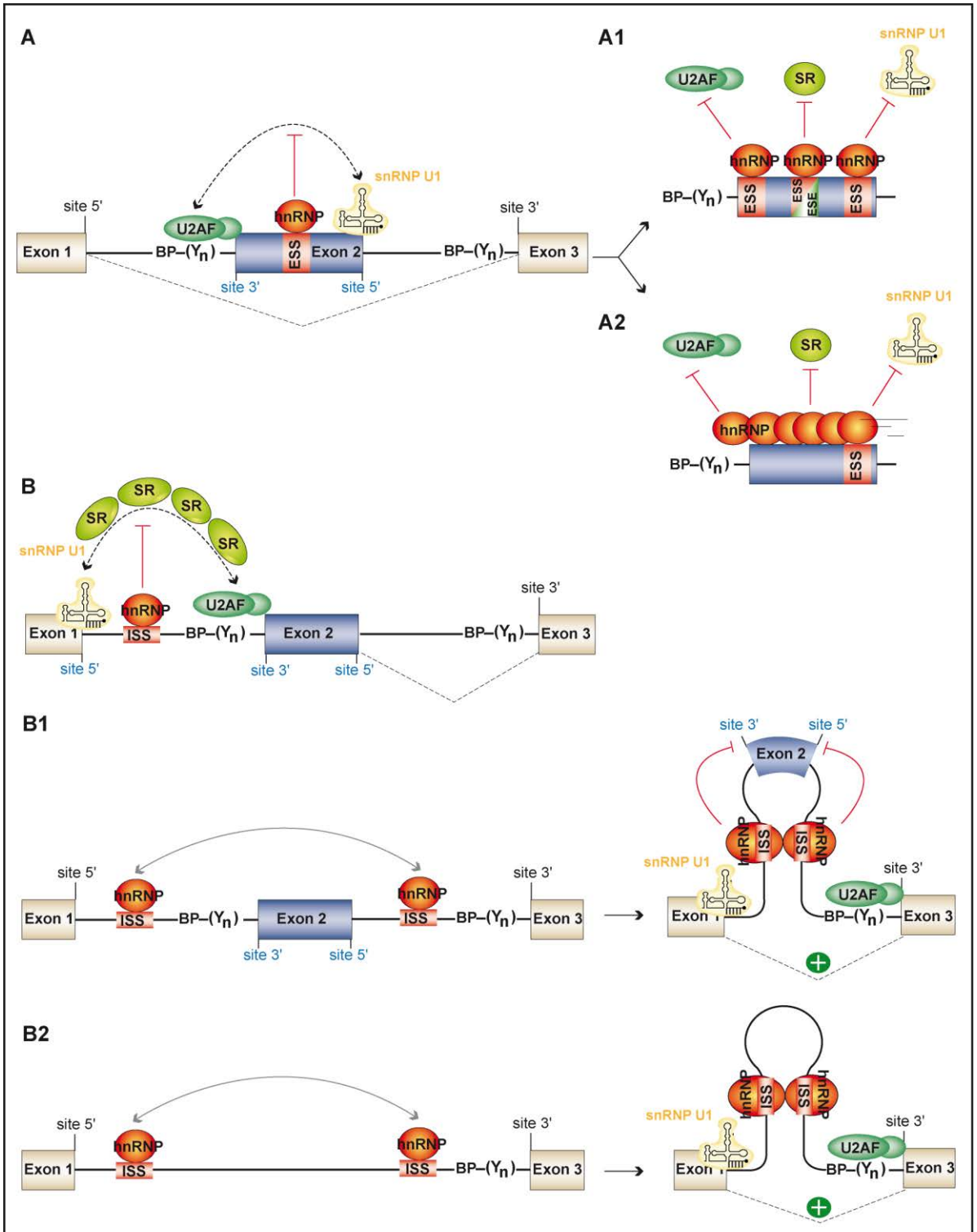



**Figure II-8 : Inhibition de l'épissage par les protéines hnRNP lors de la définition de l'exon (A) ou de la définition de l'intron (B).**

**A :** La fixation des protéines hnRNP sur une séquence ESS inhibe la définition de l'exon. Selon la position de la séquence ESS, les protéines hnRNP empêchent par encombrement stérique la fixation du facteur U2AF sur le site 3', de la snRNP U1 sur le site 5'et des protéines SR la séquence ESE chevauchant la séquence ESS (A1). Ces facteurs peuvent être aussi déplacés par l'oligomérisation des protéines hnRNP le long de l'exon (A2).

**B :** La fixation des protéines hnRNP sur une séquence ISS inhibe la définition de l'intron. L'interaction des protéines hnRNP fixées sur deux séquences ISS induit la formation d'une boucle qui réprime l'épissage de l'exon central (B1) et active l'épissage des exons terminaux (B1 et B2).

Les fixation des protéines hnRNP peuvent également inhiber simultanément la reconnaissance des sites 3' et 5' d'épissage en empêchant les facteurs *trans*, souvent des protéines SR, d'établir un réseau d'interaction protéines-protéines qui ponte les composants du spliceosomes, le facteur U2AF et la snRNP U1, respectivement fixés surs les sites 3' et 5' (**Figure II-8, A**).

Les protéines hnRNP peuvent également inhiber la définition de l'intron lorsque celui possède des séquences ISS (Martinez-Contreras *et al.*, 2007) (**Figure II-8, B**). Lorsque deux séquences ISS sont présentes dans l'intron de part et d'autre d'un exon, les protéines hnRNP fixées sur chacune de ces séquences peuvent interagir et former une boucle qui séquestre et empêche l'épissage de l'exon central (**Figure II-8, B1**). A l'inverse, la formation de cette boucle provoque un rapprochement dans l'espace des sites d'épissage des deux exons terminaux et favorise leur épissage (**Figure II-8, B1**). Cette activation de l'épissage existe également lorsque deux séquences ISS sont présentes dans un seul intron (**Figure II-8, B2**). Nous illustrerons par la suite ces modes de régulation de l'épissage avec quelques exemples représentatifs impliquant les protéines hnRNP A1, H et la protéine PTB.

## III. Rôles et mécanismes généraux des protéines SR dans l'épissage alternatif

La structure et le rôle des protéines SR dans l'épissage constitutif des pré-ARNm ont été précédemment abordés dans du chapitre I de cette introduction (**III.3.2.2**) et nous allons à présent nous intéresser aux rôles de ces protéines dans la régulation de l'épissage alternatif. La fixation des protéines SR sur leurs séquences ESE joue un rôle prépondérant lors de la définition des exons (**Figure II-9, A**). Elles favorisent la fixation de la snRNP U1 sur le site 5' ainsi que la fixation du facteur U2AF sur le site 3'. La fixation des protéines SR est souvent le résultat d'une compétition avec les protéines hnRNP se fixant sur une séquence ESS voisine à la séquence ESE (**Figure II-9, A**). Les protéines SR contribuent également à la définition de l'intron (**Figure II-9, B**). Les protéines SR fixées sur les séquences ESE ou ISE favorisent les recrutements de la snRNP U1 et le facteur U2AF sur les sites 5' et 3', respectivement. Ces derniers sont également stabilisés et pontés indirectement l'un à l'autre par les protéines SR qui établissent entre elles un réseau d'interaction protéines-protéines. Les protéines SR interviennent également plus tardivement dans l'épissage en stabilisant la snRNP U2 sur le point de branchement et en favorisant le recrutement de la tri-snRNP au sein des complexes spliceosomaux. Nous présenterons différents exemples qui illustrent ces régulations.

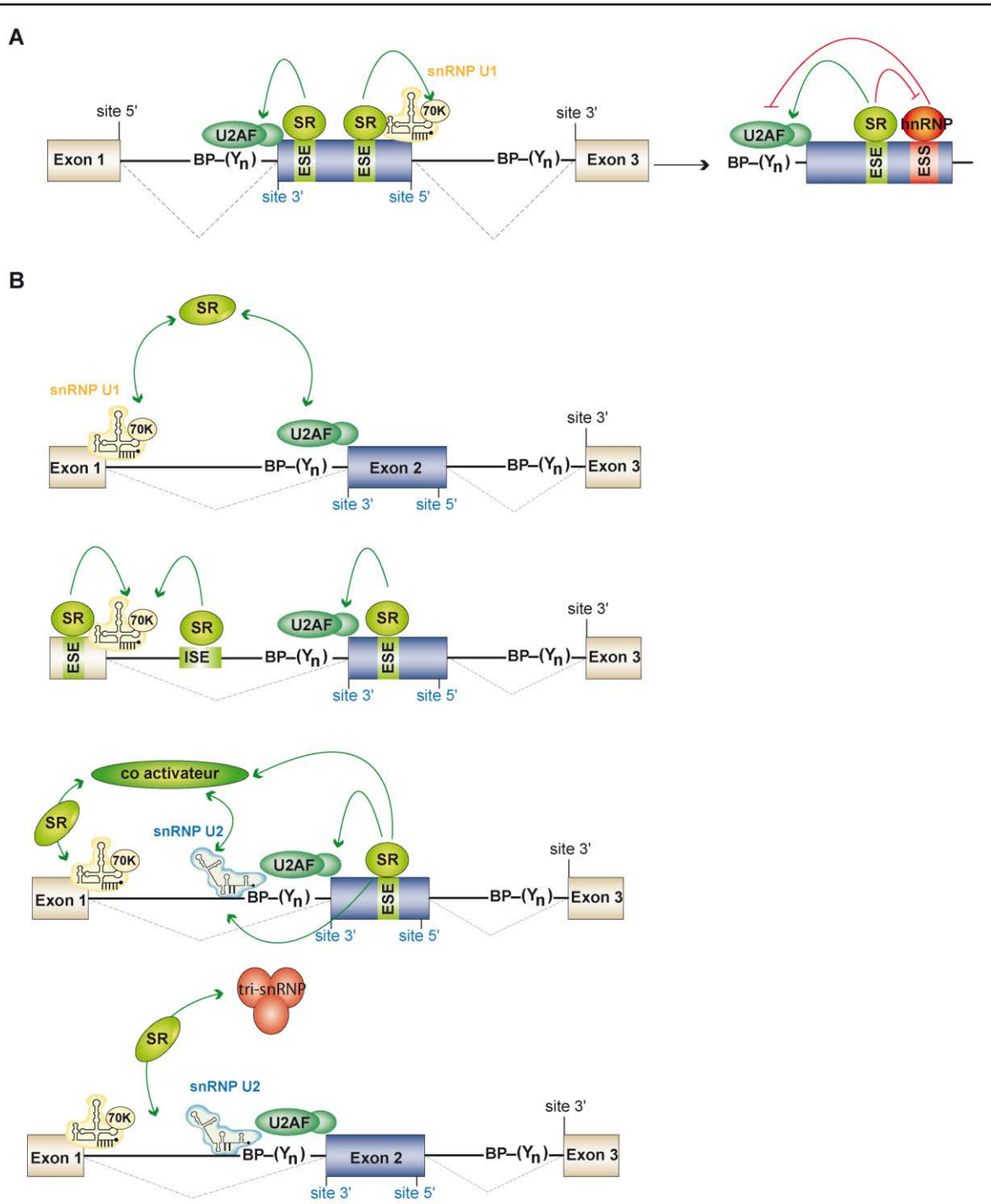



**Figure II-9 : Activation de l'épissage par les protéines SR lors de la définition de l'exon (A) ou de la définition de l'intron (B).**

**A :** La fixation des protéines SR sur une séquence ESE active la définition de l'exon en favorisant la fixation du facteur U2AF sur le site 3', de la snRNP U1 sur le site 5'. Les protéines SR sont souvent en compétition avec les protéines hnRNP pour réguler la reconnaissance du site donneur situé en amont de leurs sites de fixations respectifs.

**B :** Les protéines SR favorisent la définition de l'intron. Les recrutements de la snRNP U1 et du facteur U2AF sont activés par les protéines SR fixées sur les séquences ESE ou ISE ainsi que par le réseau d'interaction protéines-protéines qu'elles établissent entre eux. Les protéines SR favorisent également le recrutement de la snRNP U2 en collaboration avec des co-activateurs de l'épissage ainsi que le recrutement de la tri-snRNP.

## IV. Régulation de l'épissage au site accepteur

Les sites A3 et A7 du pré-ARNm du virus VIH-1 sont des sites accepteurs alternatifs faibles dont les régulations de l'épissage illustrent bien les différents cas de figures que l'on peut rencontrer pour un site accepteur régulé.

### IV.1. Les sites accepteurs A3 et A7 du virus VIH-1

Chez le virus VIH-1, le site A3 est exclusivement utilisé pour produire l'ARNm *tat* en combinaison avec les sites donneurs viraux D1, D2 et D3 (Saliou *et al.*, 2009). L'utilisation du site A3 est régulée négativement par les séquences exoniques, ESS2 et ESS2p (Amendt *et al.*, 1995 ; Jacquenet *et al.*, 2001b). L'inhibition de l'utilisation du site A3 par la séquence ESS2p est due à la fixation protéine hnRNP H qui empêche par encombrement stérique la fixation de la sous-unité U2AF35 sur la bordure intron/exon située à proximité, 8 nucléotides en amont (Jacquenet *et al.*, 2001b) (**Figure II-10, A**). La séquence ESS2 inhibe également l'épissage au site A3 en fixant la protéine hnRNP A1 (Caputi *et al.*, 1999).

A l'inverse, l'épissage au site A3 est activé par la séquence ESE2 qui fixe la protéine SC35 (Ropers *et al.*, 2004 ; Zahler *et al.*, 2004). La séquence ESE2 chevauche partiellement la séquence ESS2, impliquant une fixation exclusive des protéines SC35 et hnRNP A1 qui sont en compétition pour la fixation sur leurs séquences respectives chevauchantes (Hallay *et al.*, 2006 ; Zahler *et al.*, 2004) (**Figure II-10, B**). En se fixant sur la séquence ESE2 grâce à son domaine RRM, la protéine SC35 empêcherait l'oligomérisation de la protéine hnRNP A1 le long de l'exon (Hallay *et al.*, 2006) (**Figure II-10, B1**). Inversement, la fixation de la protéine hnRNP A1 sur la séquence ESS2 empêcherait la fixation de la protéine SC35 sur la séquence ESE2 (Hallay *et al.*, 2006 ; Zahler *et al.*, 2004) (**Figure II-10, B2**).

Le pré-ARNm du virus VIH-1 comporte un site accepteur A7 qui est utilisé alternativement pour la production des ARNm *tat* et *rev* (Saliou *et al.*, 2009). La régulation de son épissage est complexe et implique principalement 2 séquences inhibitrices, ISS et ESS3, et une séquence ESE3, également appelée élément Janus, qui est à la fois activatrice et inhibitrice, suivant les protéines qui s'y fixent (Amendt *et al.*, 1995 ; Marchand *et al.*, 2002 ; Staffa et Cochrane, 1995) (**Figure II-11, A**). L'inhibition de l'épissage au site A7 est due à la fixation de la protéine hnRNP A1 sur ces 3 séquences (Marchand *et al.*, 2002 ; Tange *et al.*, 2001 ; Zhu *et al.*, 2001). Il a été suggéré que l'élément Janus constitue le site primaire de fixation de la protéine hnRNP A1 à partir duquel la protéine s'oligomérise bidirectionnellement jusqu'à

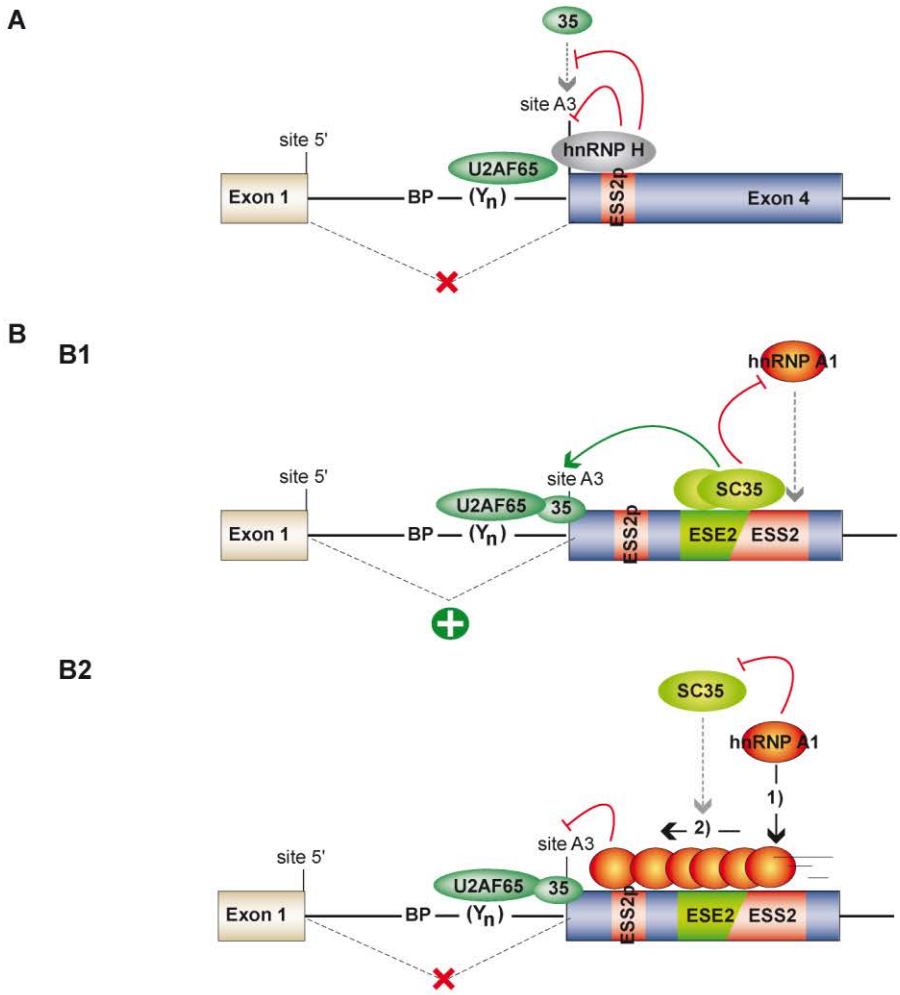



**Figure II-10 : Régulation de l'épissage au site accepteur A3 du pré-ARNm du virus VIH-1 (Jacquenet *et al.*, 2001, Hallay *et al.*, 2006, Zahler *et al.*, 2004)**

**A :** La fixation de la protéine hnRNP H sur la séquence ESS2p empêche par encombrement stérique la fixation de la sous-unité U2AF35, ce qui inhibe l'épissage au site A3.

**B1 :** La fixation de la protéine SC35 sur la séquence ESE2 active l'épissage au site A3 et empêche la fixation de la protéine hnRNP A1 sur la séquence ESS2 chevauchante.

**B2** : La fixation de la protéine hnRNP A1 sur la séquence ESS2 (1) puis son oligomérisation le long de l'exon (2) inhibe l'épissage au site A3 et empêche la fixation de la protéine SC35 sur la séquence ESE2 chevauchante.

atteindre ses seconds sites de fixation que sont les séquences ISS et ESS3a, ESS3b situées respectivement en amont et en aval. Cette fixation coopérative recouvre au passage le site accepteur et le point de branchement qui ne sont plus accessibles pour interagir avec le facteur U2AF (Marchand *et al.*, 2002) (**Figure II-10, A**). En revanche, la fixation de la protéine ASF/SF2 sur l'élément Janus empêche cette oligomérisation et active l'épissage au site A7. Cet effet activateur est obtenu à la fois avec une protéine ASF/SF2 entière ou dépourvue de son domaine RS, suggérant que la fixation de cette protéine à l'ARN n'est pas déplacée par la protéine hnRNP A1 (Damgaard *et al.*, 2002 ; Marchand *et al.*, 2002 ; Zhu *et al.*, 2001)

(**Figure II-10, B**). L'ambivalence de l'élément Janus serait dépendante des concentrations relatives des protéines qui s'y fixent. Il serait ainsi activateur en présence d'une forte concentration en protéine ASF/SF2 et inhibiteur en présence d'une forte concentration en protéine hnRNP A1 (Marchand *et al.*, 2002).

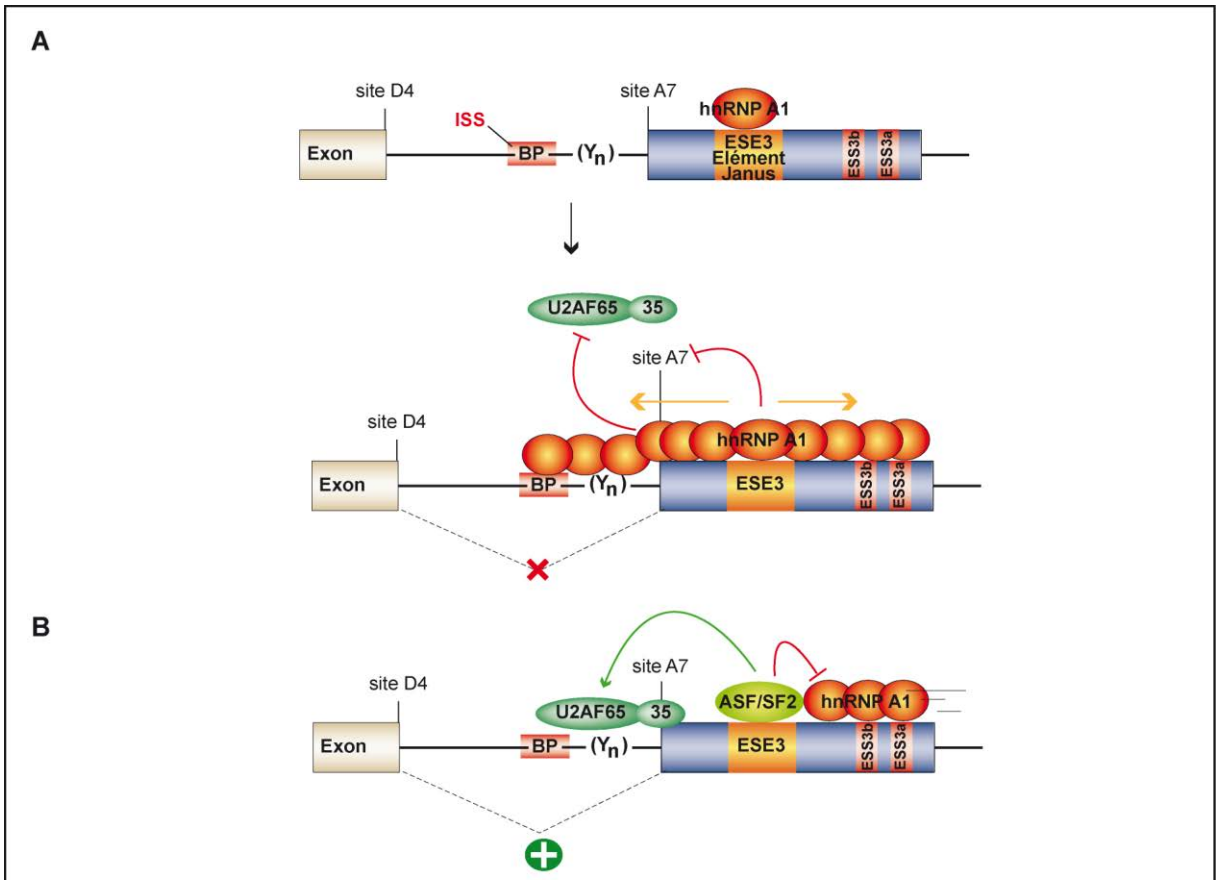



**Figure II-11 : Régulation de l'épissage au site accepteur A7 du pré-ARNm du virus VIH-1 (Marchand *et al.*, 2002).**

**A :** Inhibition de l'épissage au site A7 par la protéine hnRNP A1. La protéine hnRNP A1 se fixe tout d'abord sur la séquence ESE3, ou élément Janus, puis oligomérise bidirectionnellement vers les séquences ISS et ESS3, ce qui empêche l'accès du facteur U2AF au site donneur.

**B :** Activation de l'épissage au site A7 par la protéine ASF/SF2. La protéine ASF/SF2 active le recrutement du facteur U2AF et empêche l'oligomérisation de la protéine hnRNP A1 depuis les séquences ESS3a et ESS3b situées en amont.

### IV.2. Le site accepteur de l'exon 2a du pré-ARNm de l'*α-tropomyosine* de rat

Le pré-ARNm de l'*α-tropomyosine* de rat possède deux exons, 2a et 2b, séparés par un intron qui sont épissés alternativement de façon exclusive et tissu-dépendant. La proximité entre le site donneur de l'exon 2a et le point de branchement du site accepteur de l'exon 2b empêche, pour des raisons stériques, l'assemblage du spliceosome autour de l'intron séparant les deux exons (Smith et Nadal-Ginard, 1989) (**Figure II-12, A**). Le site accepteur de l'exon 2b est beaucoup plus fort que celui de l'exon 2a, ce qui entraîne son épissage « par défaut »

dans tous les types cellulaires, à l'exception des cellules musculaires lisses où seul l'exon 2a est épissé (Ellis *et al.*, 2004 ; Mullen *et al.*, 1991) (**Figure II-12, B, C**). Dans ce cas, l'épissage au site accepteur de l'exon 2b est réprimé et celui de l'exon 2a est activé.

L'exon 2a possède deux séquences riches en purines qui activent ou inhibent l'épissage du site accepteur suivant le facteur qui s'y fixe (Crawford et Patton, 2006 ; Dye *et al.*, 1998 ; Lin et Patton, 1995). *In vitro*, la fixation de la protéine 9G8 sur ces séquences active l'épissage tandis celle de la protéine hnRNP H ou F l'inhibe (Crawford et Patton, 2006). Les protéines hnRNP H, F et la protéine 9G8 sont en compétition pour la fixation sur ces séquences, ce qui est singulier car les protéines hnRNP H/F se fixent habituellement sur des séquences riches en G (Crawford et Patton, 2006). *In vitro*, la fixation de ces différentes protéines sur la séquence ESE dépend de leur concentration relative (Crawford et Patton, 2006). Il semblerait que cela soit également le cas *in cellulo* comme l'indique l'abondance des protéines hnRNP H et F dans les cellules HeLa qui sont supérieures à celles des cellules musculaires lisses. Cela suggère qu'elles participent à l'inhibition de l'épissage de l'exon 2a dans les types cellulaires autres que les cellules musculaires lisses (Crawford et Patton, 2006) (**Figure II-12, B**).

Dans les cellules des muscles lisses, l'inhibition de l'épissage de l'exon 2b nécessite la présence de part et d'autre de l'exon d'une séquence polypyrimidine et d'une séquence répétée UGC/CUG (Gooding *et al.*, 1998 ; Gromak et Smith, 2002 ; Perez *et al.*, 1997) (**Figure II-12, C**). La protéine PTB (hnRNP I) inhibe l'épissage en se fixant sur les séquences polypyrimidines et son effet inhibiteur est renforcé par un autre co-répresseur, la protéine Raver-1 qui ponte les protéines PTB fixées de part et d'autre de l'exon (Gooding *et al.*, 1998 ; Gromak *et al.*, 2003 ; Perez *et al.*, 1997 ; Spellman *et al.*, 2007). Cette inhibition de l'épissage par la protéine PTB est un exemple représentatif de l'activité de cette protéine dans la régulation de l'épissage alternatif.

L'inhibition de l'épissage de l'exon 2b nécessite aussi la présence de séquences répétées UGC/CUG situées également de chaque côté de l'exon 2b (**Figure II-12, C**). Deux familles de protéines, les protéines CELF et les protéines MBNL, sont des protéines régulant l'épissage alternatif et qui se fixent spécifiquement à ce type de séquence (Barreau *et al.*, 2006 ; Gallo et Spickett, 2010). L'épissage de l'exon 2b n'est pas inhibé par les protéines de

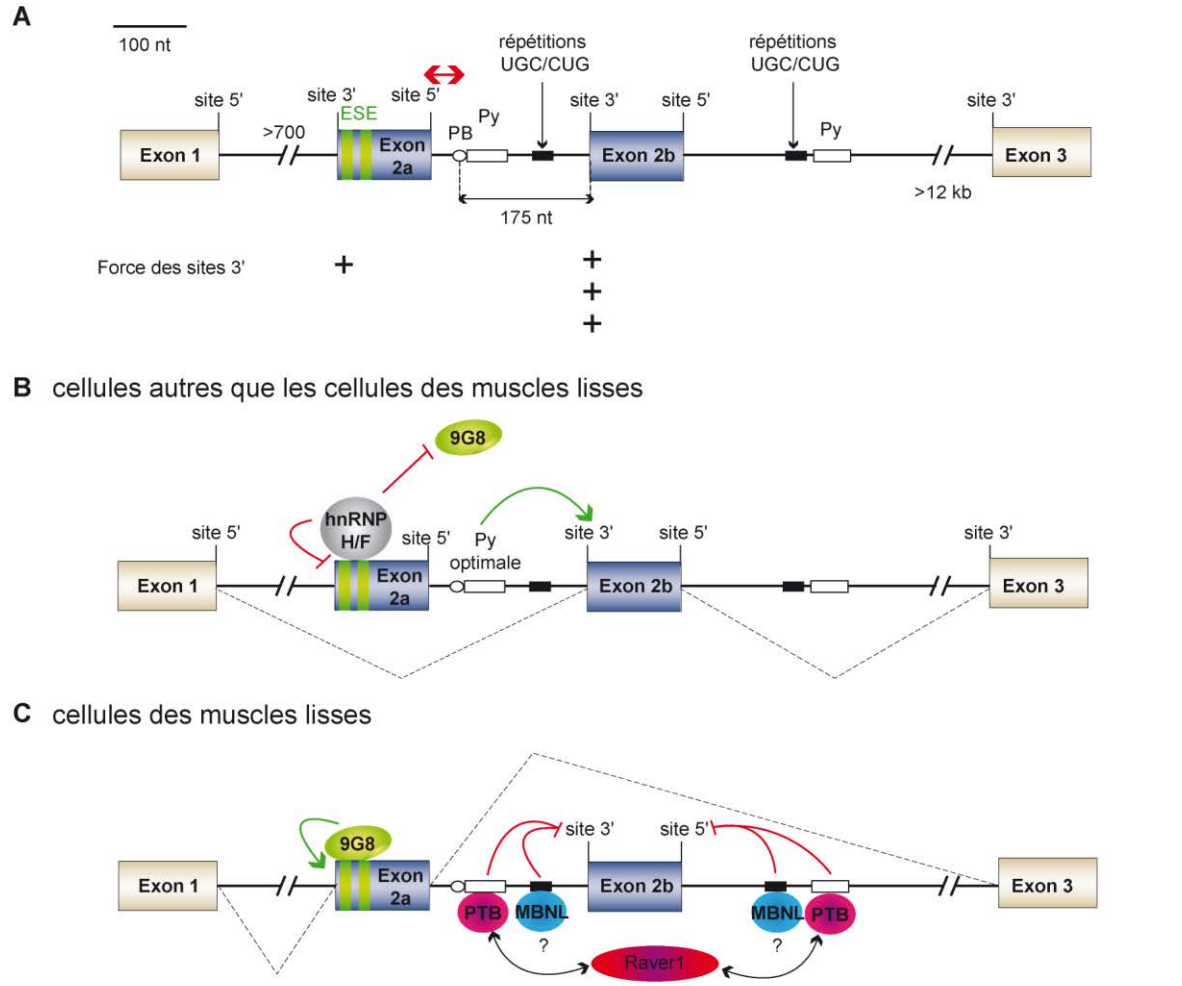



**Figure II-12 : Régulation de l'épissage des exons 2a et 2b du pré-ARNm de l'α tropomyosine de rat (Crawford et Patton, 2006, Gooding et Smith, 2008).**

**A :** Représentation du pré-ARNm de l'α tropomyosine de rat. Les exons 2a et 2b dont l'épissage est mutuellement exclusif sont représentés par des rectangles bleus. La séquence polypyrimidine (Py) et le point de branchement (BP) et la distance que le sépare du site 3' de l'exon 2b sont mentionnés. La double flèche rouge représente la proximité du site donneur de l'exon 2a et du point de branchement du site 3' de l'exon 2b qui empêche l'épissage de l'intron. Les séquences UGC/CUG répétées sont représentées par des rectangles noirs. La force relative des sites 3' des exons 2a et 2b est reportée en dessous du pré-ARNm.

**B :** L'exon 2b est épissé dans toutes les cellules qui ne sont pas des cellules des muscles lisses. Le site 3' de l'exon 2b est exclusivement épissé au détriment du site 3' de l'exon 2a grâce à sa séquence polypyrimidine optimale. Le site 3' de l'exon 2a est réprimé par la protéine hnRNP H qui se fixe sur la séquence ESE et empêche la fixation de la protéine 9G8.

**C :** L'exon 2a est épissé dans les cellules des muscles lisses. L'épissage de l'exon 2b est inhibé par les protéines PTB et la protéine Raver1 ainsi que par un autre facteur qui pourrait être les protéines MBNL. L'épissage de l'exon 2a est activé par la fixation de la protéine 9G8 sur la séquence ESE

la famille CELF et semble donc impliquer théoriquement les protéines de la famille MBNL (Gromak et Smith, 2002). L'épissage mutuellement exclusif des deux exons 2a et 2b selon le type cellulaire n'est pas encore totalement compris car les facteurs *en trans* y intervenant sont exprimés de façon ubiquitaire dans toutes les cellules.

# V. Régulation de l'épissage lors de la définition de l'intron et de l'exon

## V.1. Rôles des protéines hnRNP

Certains mécanismes de régulation permettent d'inhiber simultanément l'épissage au site 5' et au site 3' d'un même exon. Dans un pré-ARNm comprenant au moins 3 exons, un de ces mécanismes repose sur la séquestration de l'exon central au sein d'une boucle qui est formée grâce à l'interaction des protéines hnRNP fixées sur des séquences ISS situées de part et d'autre de cet exon central. L'exon central séquestré peut être reconnu par les composants du spliceosome (snRNP U1, facteur U2AF) mais ne peut pas s'engager dans un complexe spliceosomal. Cela provoque par ailleurs un rapprochement dans l'espace des deux exons terminaux, ce qui active leur épissage. Ce mécanisme est utilisé par la protéine hnRNP A1 pour réguler l'épissage alternatif de son propre pré-ARNm (Chabot *et al.*, 1997) (**Figure II-13**).

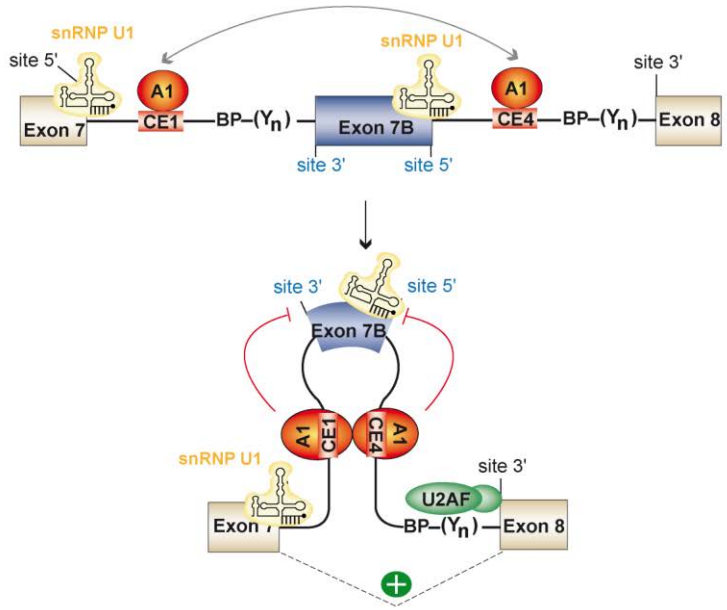



**Figure II-13 : La protéine hnRNP A1 régule l'épissage alternatif de son propre pré-ARNm en séquestrant l'exon 7B dans une boucle (Chabot *et al.*, 2003)**
L'interaction des protéines hnRNP A1 (A1) fixées sur les séquences ISS CE1 et CE4 bordant l'exon 7B provoque l'exclusion de cet exon.

Les fixations des protéines hnRNP A1 sur les séquences ISS CE1 et CE4 situées respectivement en amont et en aval de l'exon 7B, provoquent l'exclusion de cet exon et activent l'épissage des exons 7 et 8 (Blanchette et Chabot, 1999 ; Chabot *et al.*, 1997 ; Chabot *et al.*, 2003) (**Figure II-13**).

Un mécanisme similaire faisant intervenir la protéine PTB a été décrit pour l'exon N1 du pré-ARNm c-*src*. L'épissage de cet exon est inhibé dans tous les types de cellules par la protéine PTB à l'exception des cellules neuronales où la protéine PTB est moins abondante et où une autre isoforme, nPTB (ou PTB2, **Figure II-3**) est exprimée (Chan et Black, 1997 ; Martinez *et al.*, 1987). Lors de l'inhibition, les protéines PTB se fixent à des séquences introniques riches en CU en amont et en aval de l'exon N1 puis interagissent pour former une boucle qui séquestre l'exon N1 (Amir-Ahmady *et al.*, 2005 ; Lamichhane *et al.*, 2010 ; Spellman *et al.*, 2007) (**Figure II-14**).

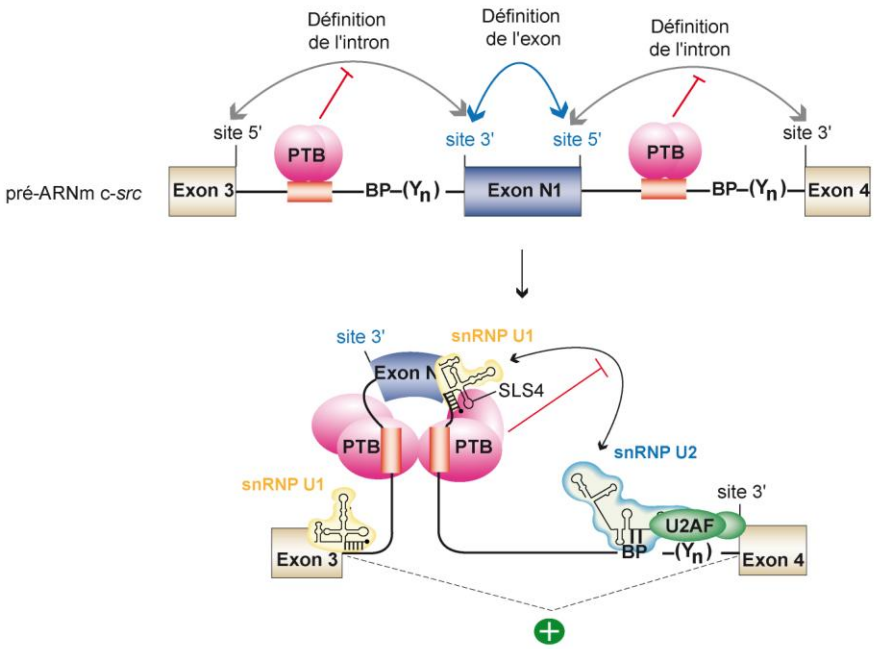



**Figure II-14 : Modèle d'inhibition de l'épissage de l'exon N1 du pré-ARNm de c-*src* par la protéine PTB (Shama *et al.*, 2011).**
Dans les cellules non neuronales, l'interaction des protéines PTB fixées sur les séquences introniques riches en CU (rectangles rouges) bordant l'exon N1 forme une boucle qui provoque l'exclusion de cet exon. L'interaction de la protéine PTB avec la séquence de la tige boucle 4 du snRNA U1 (SLS4) serait responsable de l'incapacité de la snRNP U1 à interagir avec la snRNP U2 fixée sur le point de branchement en aval.

Cela n'affecte pas la fixation de la snRNP U1 sur le site donneur de l'exon N1 (Sharma *et al.*, 2005 ; Sharma *et al.*, 2008). En revanche, les protéines PTB empêchent la définition de l'intron séparant l'exon N1 et l'exon 4 dont le site donneur et accepteur ne peuvent pas s'engager dans un complexe spliceosomal A (Sharma *et al.*, 2008). L'interaction entre la tige boucle SLS4 du snRNA U1 et les domaines RR1 et RRM2 de la protéine PTB serait en partie responsable de l'incapacité de la snRNP U1 à interagir avec la snRNP U2 au sein de ces complexes (Sharma *et al.*, 2010). Par ailleurs, l'analyse des complexes spliceosomaux assemblés sur un ARN comprenant l'exon D1, l'intron et l'exon 4 indique que la protéine

PTB empêche l'association de nombreux facteurs d'épissage qui sont importants pour le pontage des snRNP U1 et U2, comme le complexe SRm160/SRm300 (Sharma *et al.*, 2008).

### V.2. Rôles des protéines « SR like » et SR

Les protéines SRm160/SRm300 sont des co-activateurs de l'épissage qui favorisent la reconnaissance des sites 3' et 5' lors de la définition de l'intron. Ces protéines sont des « SR like » car elles ne possèdent pas de domaine RRM mais des domaines RS qui leur permettent d'interagir avec les protéines SR (Eldridge *et al.*, 1999 ; Longman *et al.*, 2001) (**Chapitre I, Figure I-18**). Elles participent ainsi avec les protéines SR à la formation du réseau d'interaction établi entre les snRNP U1 et U2 fixées sur chacune des bordures de l'intron, comme c'est le cas du pré-ARNm *doublesex* (*dsx*) de drosophile (Blencowe *et al.*, 2000 ; Eldridge *et al.*, 1999) (**Figure II-15, A**).

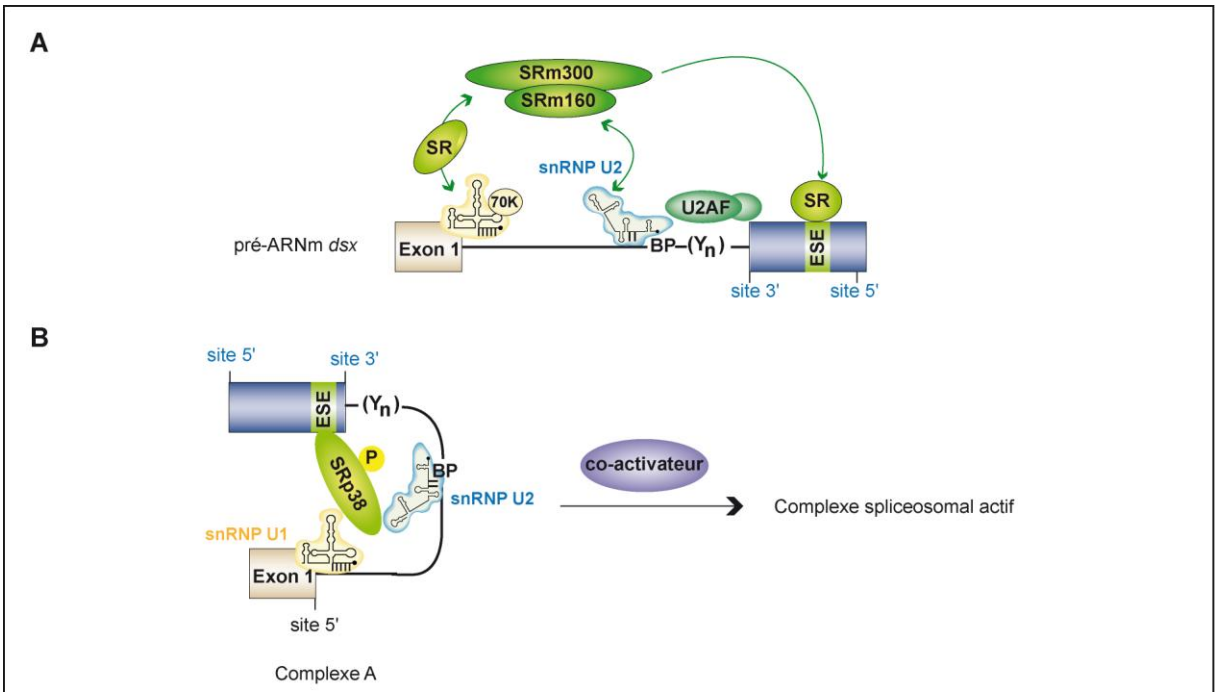



**Figure II-15 : Activation de la définition des introns par les co-activateurs SRm160/SRm300 (A) et la protéine SRp38 phosphorylée (B) (Eldridge *et al.*, 1999, Feng *et al.*, 2008).**
**A :** Le complexe SRm160/SRm300 active l'épissage en interagissant avec les particules snRNP U1 et U2 ainsi que les protéines SR fixées sur les séquences ESE.
**B :** La protéine SRp38 phosphorylée active l'épissage en se fixant à sa séquence ESE et facilite l'association des snRNP U1 et U2 avec le site 5' et le point de branchement. Le complexe A ainsi formé nécessite la présence d'un cofacteur pour la transition vers un complexe spliceosomal actif.

L'état de phosphorylation de ces co-activateurs ou des protéines SR joue également un rôle important dans l'activation de l'épissage. La protéine SRp38 en est un parfait exemple car cette protéine SR active ou inhibe l'épissage suivant son état de phosphorylation. Comme nous l'avons précédemment abordé (**Chapitre I**, **I.3.2.2.2**), le cycle de phosphorylation/déphosphorylation de cette protéine est régulé par le cycle cellulaire ou les conditions de stress cellulaire. La protéine SRp38 phosphorylée active l'épissage de certains pré-ARNm comportant une séquence ESE sur laquelle elle se fixe (Feng *et al.*, 2008). Contrairement aux autres protéines SR, cette protéine ne peut pas activer l'épissage *in vitro* lorsqu'elle est ajoutée à un extrait cytoplasmique S100 et nécessite pour cela la présence d'un cofacteur encore non identifié (Feng *et al.*, 2008). *In vitro*, la protéine SRp38 phosphorylée active la définition de l'intron en favorisant la fixation de la snRNP U1 sur le site donneur et la snRNP U2 sur le point de branchement, induisant la formation du complexe spliceosomal A (Feng *et al.*, 2008) (**Figure II-15, B**). *In cellulo*, il a été démontré que cette protéine régule l'épissage alternatif du pré-ARNm du récepteur au glutamate GRIA2 (Feng *et al.*, 2008).

## VI. Régulation du recrutement de la snRNP U1 sur le site donneur

### VI.1. Les différents cas de figure

Le recrutement de la snRNP U1 sur le site donneur est souvent activé par l'interaction entre une protéine SR, comme la protéine ASF/SF2, fixée sur la séquence ESE située en amont, et la protéine U1-70K de la particule (**Figure II-16, A**). La phosphorylation du domaine RS des protéines SR est importante pour que cette interaction puisse avoir lieu (Cho *et al.*, 2011). La protéine SRp38 est un cas particulier car sa phosphorylation active le recrutement de la snRNP U1 sur le site donneur tandis que sa déphosphorylation l'inhibe (Shin *et al.*, 2004) (**Figure II-16, B**). Dans ce cas, la protéine SRp38 interagit avec la protéine U1-70K, et empêche la fixation de la snRNP U1 sur le site donneur (Shin *et al.*, 2004) (**Figure II-16, B**).

L'accès au site donneur peut aussi être restreint par la fixation d'une protéine qui empêche la fixation de la snRNP U1 par encombrement stérique. Dans le pré-ARNm NF-1, la protéine hnRNP H se fixe avec affinité sur la séquence de la bordure exon/intron UGGGU du site donneur et y limite la fixation de la snRNP U1 (Buratti *et al.*, 2004) (**Figure II-16, C**).

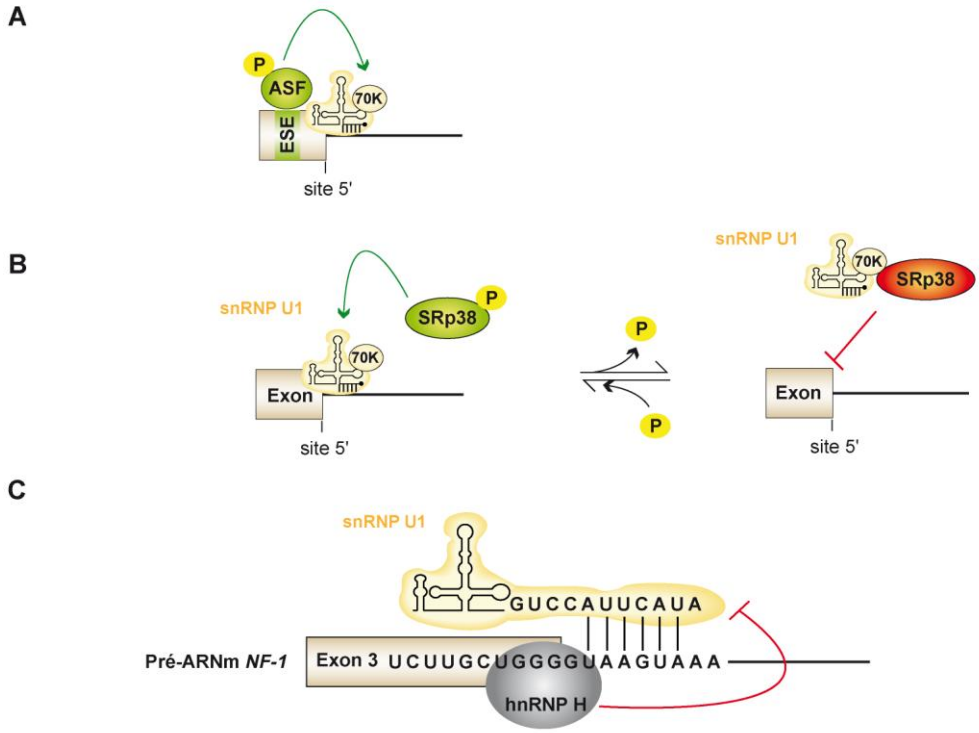



**Figure II-16 : Exemples de régulation du recrutement de la snRNP U1 sur le site donneur (Cho *et al.*, 2011, Shin *et al.*, 2004, Buratti *et al.*, 2004).**
**A :** Le recrutement de la snRNP U1 sur le site donneur est activé par la fixation de la protéine ASF/SF2 phosphorylée sur une séquence ESE située en amont.
**B :** La protéine SRp38 phosphorylée active le recrutement de la snRNP U1 sur le site donneur tandis que sa forme déphosphorylée l'inhibe en interagissant avec la protéine U1-70K de la particule.
**C :** Le recrutement de la snRNP U1 sur le site donneur de l'exon 3 du gène NF-1 est limité par la fixation de la protéine hnRNP H.

### VI.2. Rôles des séquences ISE

Le recrutement de la snRNP U1 sur le site donneur peut être activé par les séquences ISE situées en aval. Parmi les séquences ISE caractérisées, celles riches en U sont en général situées 6 à 30 nucléotides en aval du site donneur et réguleraient environ 15 % des exons cassettes (Aznarez *et al.*, 2008). La protéine TIA-1 (T cell Intracytoplasmic Antigen 1) et son homologue humain TIAR (TIA-1 Related) possèdent un domaine riche en glutamine et 3 domaines RRM et se fixent à ces séquences riches en U (Dember *et al.*, 1996). Ces protéines favorisent le recrutement de la snRNP U1 en interagissant avec la protéine U1-C de la particule (Del Gatto-Konczak *et al.*, 2000 ; Forch *et al.*, 2002).

Chez les drosophiles mâles, la fixation de la protéine TIA-1 sur la séquence ISE active l'épissage au site donneur du premier exon du pré-ARNm *mls-2* (Forch *et al.*, 2001) (**Figure II-17, A**). En revanche, chez les femelles, la protéine SXL (Sex Lethal), qui n'est pas exprimée chez les mâles, est un antagoniste de la protéine TIA-1 et se fixe sur la séquence

ISE, empêchant la fixation de la protéine TIA-1 et par conséquent le recrutement de la snRNP U1 sur le site donneur (**Figure II-17, A**).

Dans l'exon 6 du pré-ARNm *Fas*, la fixation de la protéine TIA-1 sur la séquence ISE favorise le recrutement de la snRNP U1 sur le site donneur. Celle-ci stabilise à son tour le facteur U2AF sur la séquence polypyrimidine, favorisant ainsi la définition de l'exon (Izquierdo *et al.*, 2005) (**Figure II-17, B**). La fixation de la protéine PTB sur une séquence ESS empêche la fixation du facteur U2AF, ce qui conduit à l'exclusion de l'exon (**Figure II-17, B**).

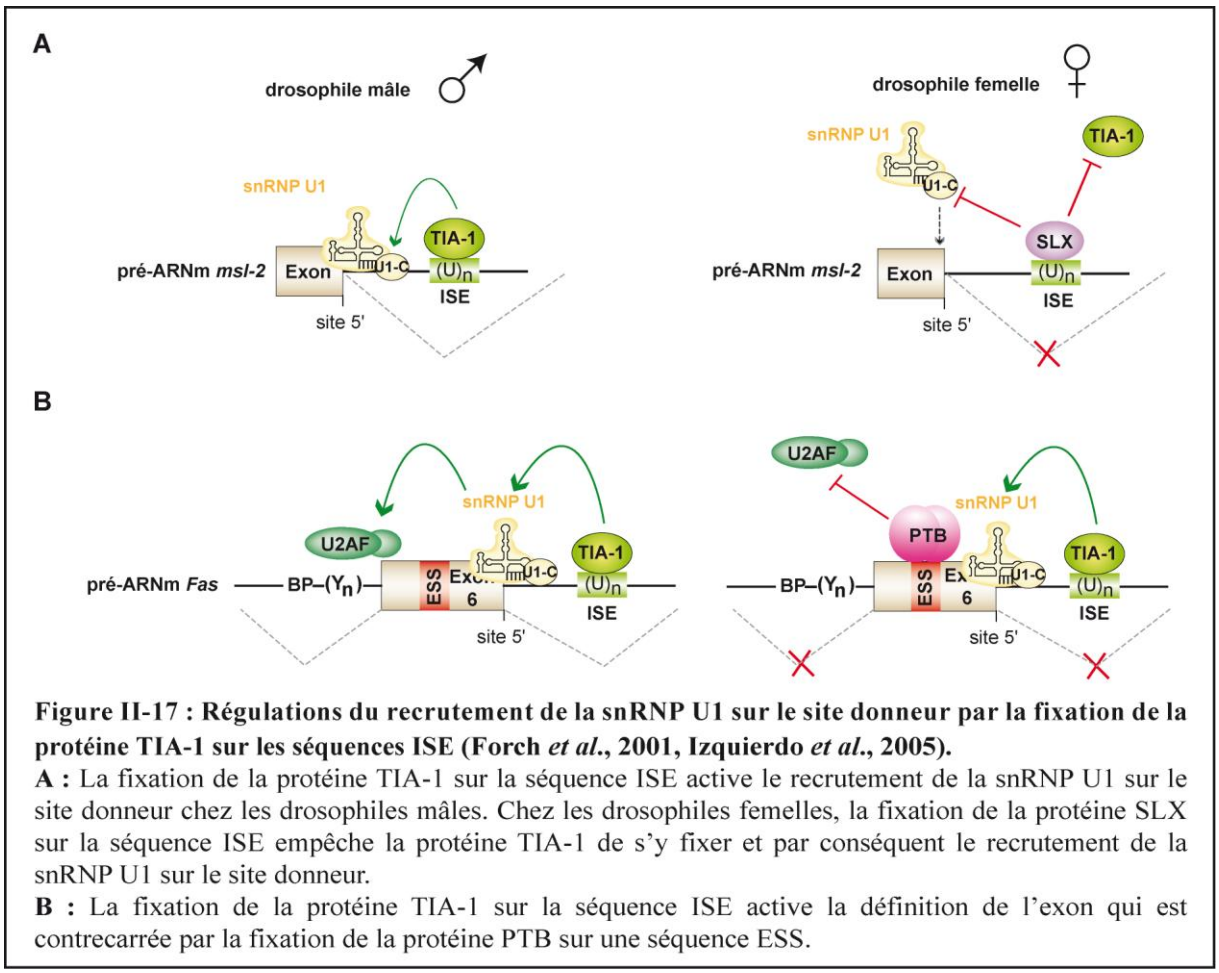



**Figure II-17 : Régulations du recrutement de la snRNP U1 sur le site donneur par la fixation de la protéine TIA-1 sur les séquences ISE (Forch *et al.*, 2001, Izquierdo *et al.*, 2005).**
**A :** La fixation de la protéine TIA-1 sur la séquence ISE active le recrutement de la snRNP U1 sur le site donneur chez les drosophiles mâles. Chez les drosophiles femelles, la fixation de la protéine SLX sur la séquence ISE empêche la protéine TIA-1 de s'y fixer et par conséquent le recrutement de la snRNP U1 sur le site donneur.
**B :** La fixation de la protéine TIA-1 sur la séquence ISE active la définition de l'exon qui est contrecarrée par la fixation de la protéine PTB sur une séquence ESS.

## VI.3. Régulation de la compétition entre plusieurs sites donneurs

### VI.3.1. Rôles des séquences ESE et des protéines SR

Certaines séquences ESE sont capables de fixer différentes protéines SR et ainsi, suivant la protéine SR fixée, de favoriser le choix d'un site par rapport à un autre. C'est le cas de l'unité de transcription E1A de l'adénovirus qui renferme 3 sites 5' et 2 sites 3' d'épissage (Bourgeois *et al.*, 1999) (**Figure II-18, A**). Les 3 sites 5' (9S, 12S et 13S) sont en compétition les uns avec les autres car ils peuvent être chacun utilisés en combinaison avec le site 3' distal.

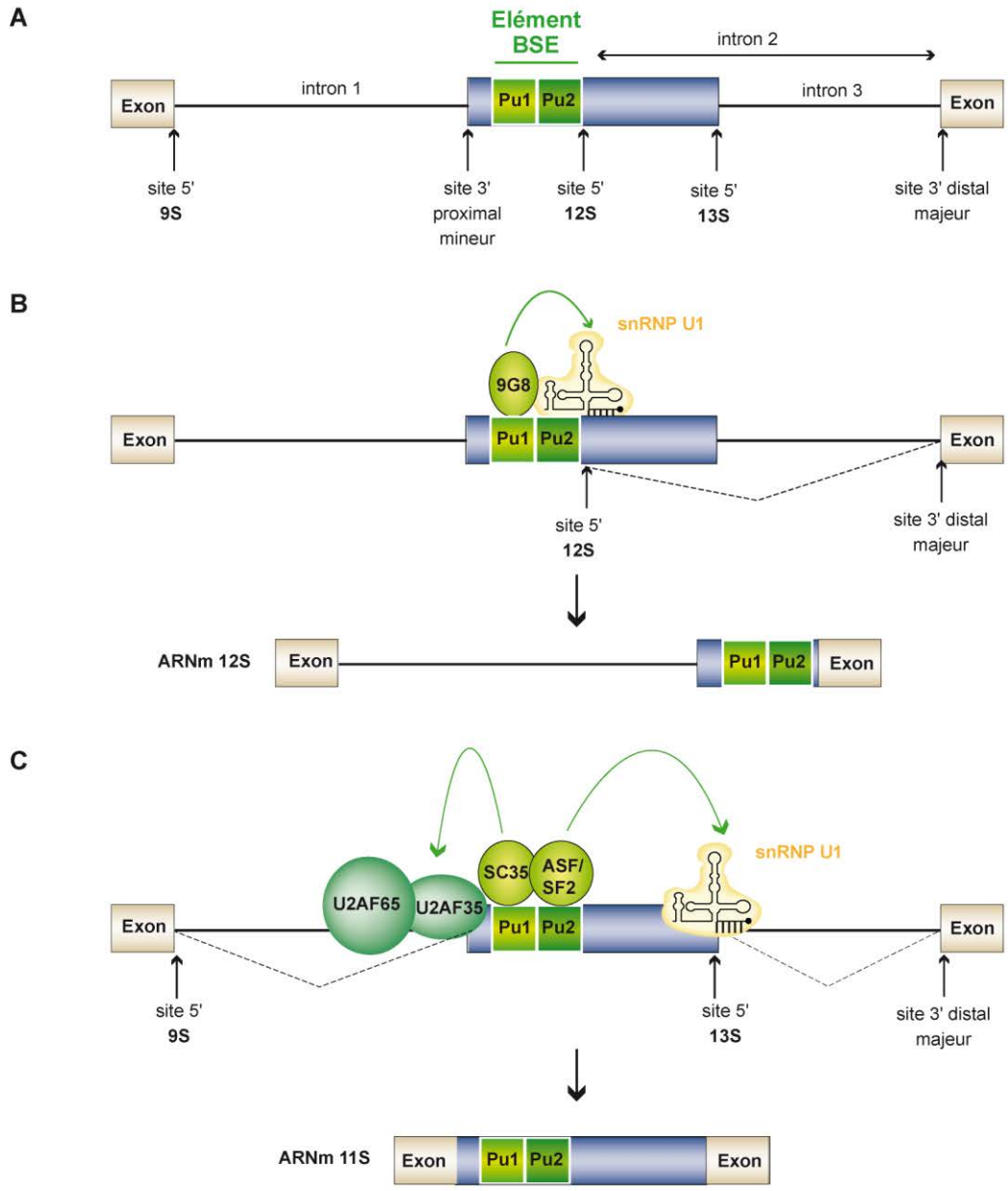



**Figure II-18 : Régulation de l'épissage du pré-ARNm de l'unité de transcription E1A par une ESE bi-directionnel liant différentes protéines SR (Bourgeois *et al.*, 1999).**
**A :** Représentation de l'unité de transcription E1A de l'adénovirus. Les différents sites d'épissage sont indiqués. Les deux motifs appelés Pu1 et Pu2 de l'élément BSE sont indiqués.
**B :** Modèle d'activation du site 5' 12S par la fixation de la protéine 9G8 sur le motif Pu1.
**C :** Modèle d'activation du site 3' proximal mineur et du site 5' 13S par la fixation des protéines ASF et SC35 sur l'élément BSE.

Le produit d'épissage majoritaire est l'ARNm 13S issu de l'utilisation du site 5' 13S avec le site 3' distal.

L'élément BSE (Bidirectional Splicing Enhancer) se situe directement en amont du site donneur 12S et se compose de 2 séquences riches en purines, Pu1 et Pu2 (**Figure II-18, A**). La fixation de la protéine 9G8 sur l'élément Pu1 active le site 5' 12S, ce qui conduit à

l'élimination de l'intron 2 et la production de l'ARNm 12S (**Figure II-18, B**). Les fixations des protéines ASF/SF2 et SC35 sur les séquences Pu1 et Pu2, respectivement, activent la définition de l'exon en favorisant la fixation du facteur U2AF sur le site 3' proximal et la fixation de la snRNP U1 sur le site 5' 12S (**Figure II-18, C**). Cela conduit à l'élimination du premier et du 3$^{e}$ intron et à la production de l'ARNm 11S. Les deux voies d'activation de l'épissage induites par l'élément BSE sont mutuellement exclusives (Bourgeois *et al.*, 1999).

### VI.3.2. Antagonisme entre les protéines SR et les protéines hnRNP

En règle générale, les protéines SR et les protéines hnRNP sont des facteurs de régulation de l'épissage qui sont antagonistes (Busch et Hertel, 2011 ; Martinez-Contreras *et al.*, 2007). Leurs sites de fixation sont souvent juxtaposés ou chevauchants, ce qui rend leur fixation exclusive pour des raisons d'encombrement stérique. L'épissage de certains pré-ARNm sont dépendants du taux relatif entre les protéines hnRNP A1 et ASF/SF2 et ces taux varient selon les tissus (Caceres *et al.*, 1994 ; Hanamura *et al.*, 1998). La fixation d'une protéine SR à un élément ESE peut parfois empêcher la fixation de la protéine hnRNP A1 sur l'ensemble d'une région ARN en empêchant, non seulement sa fixation au site de haute affinité, mais aussi sa multimérisation sur l'ARN (Caceres et Krainer, 1993 ; Caceres *et al.*, 1994 ; Cavaloc *et al.*, 1994 ; Eperon *et al.*, 1993 ; Krainer *et al.*, 1990). A l'inverse, la fixation de la protéine hnRNP A1 sur un élément ESS peut empêcher la fixation des protéines SR sur des séquences ESE voisines ou les déplacer grâce à sa capacité d'oligomérisation (Okunola et Krainer, 2009).

L'influence des sites de différentes affinités pour les protéines hnRNP A1 et ASF/SF2 sur le choix des sites donneurs en compétition lors de l'épissage a été étudiée *in vitro* (Eperon *et al.*, 2000) (**Figure II-19, A**). Lorsqu'elle est présente en forte concentration, la protéine hnRNP A1 occupe des sites de fixation à partir desquels elle oligomérise le long de l'ARN. Dans ce cas, la snRNP U1 ne peut s'associer qu'à un site 5' fort ou à un site 5' possédant dans son voisinage, un site de haute affinité pour la protéine ASF/SF2 (**Figure II-19, B**). Cette dernière est alors capable de déplacer efficacement la protéine hnRNP A1. Dans le pré-ARNm modèle étudié, l'oligomérisation de la protéine hnRNP A1 empêcherait l'utilisation du site proximal. La snRNP U1 est alors fixée de manière stable sur le site distal et induit son utilisation lors de l'épissage. Lorsque la concentration en protéine ASF/SF2 est élevée, la protéine SR stabilise la fixation de la snRNP U1 sur tous les sites 5' d'épissage et le site donneur proximal est alors préférentiellement choisi (**Figure II-19, C**).

Cette étude montre bien que la compétition entre deux facteurs antagonistes est multiparamétrique et dépend à la fois de leur concentration relative et de la position de leur site de fixation par rapport aux sites donneurs qui sont en compétition.

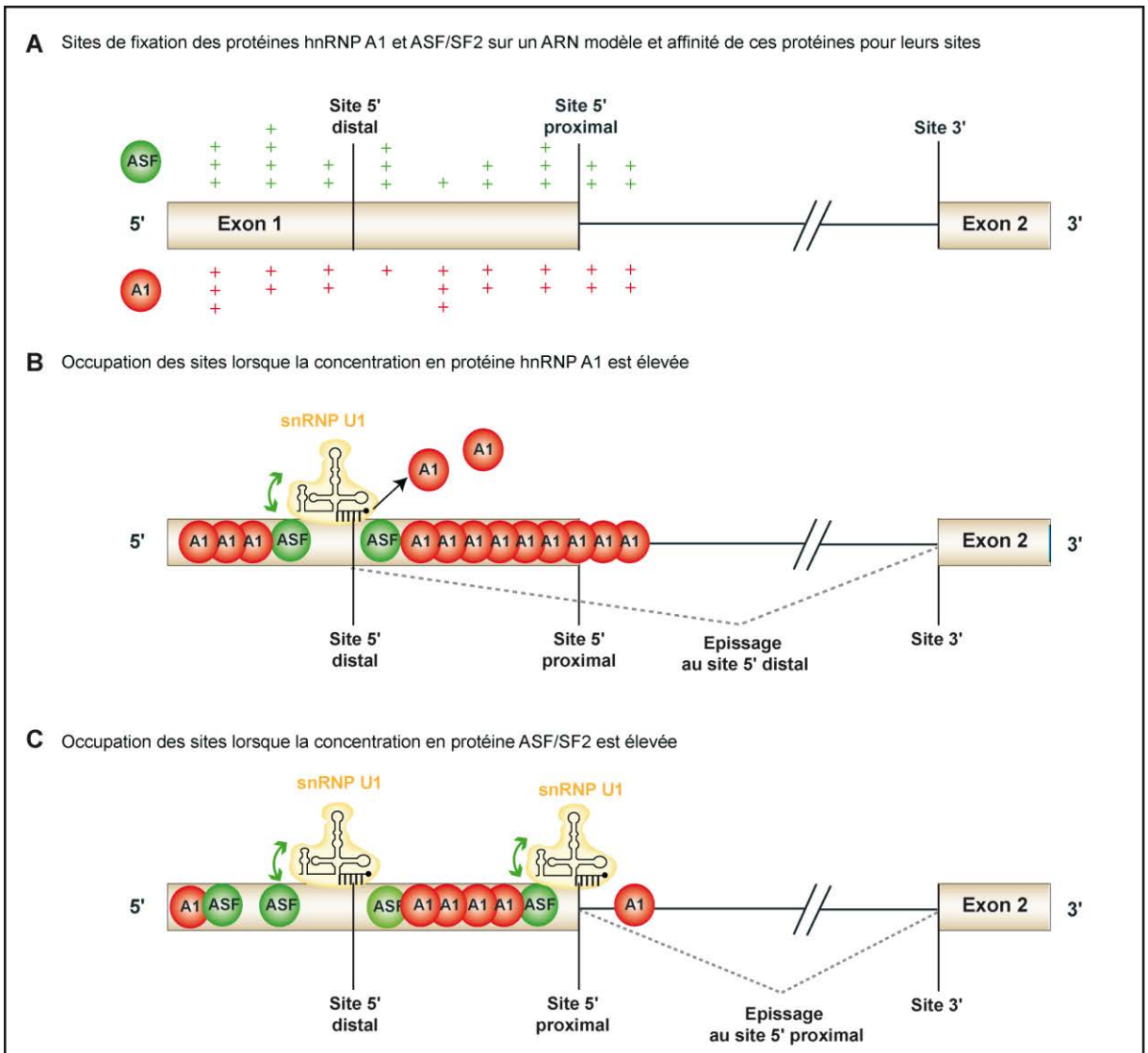



**Figure II-19 : Exemple de mécanisme d'antagonisme entre les protéines hnRNP A1 et ASF/SF2 pour la sélection d'un site 5' d'épissage (Eperon *et al.*, 2000)**

**A :** Sites de fixation de différentes affinités pour les protéines hnRNP A1 et ASF/SF2 sur un pré-ARNm contenant deux sites 5' d'épissage alternatifs. L'affinité de chaque site varie de + à ++++ (faible à très élevée).

**B :** Lorsque la concentration en protéine hnRNP A1 est plus élevée que celle en protéine ASF/SF2 dans un extrait nucléaire, seuls les sites de haute affinité peuvent être occupés par la protéine ASF/SF2. La liaison de la snRNP U1 aux sites 5' est réduite par l'occupation des sites par la protéine hnRNP A1. Dans ce cas, le site 5' distal est utilisé du fait de sa proximité avec un site de haute affinité pour ASF/SF2.

**C :** Lorsque la concentration en protéine ASF/SF2 est plus élevée que celle en protéine hnRNP A1 dans un extrait nucléaire, tous les sites de faible, moyenne, et haute affinité sont occupés par la protéine ASF/SF2. La snRNP U1 est fixée sur chacun des sites 5' et dans ce cas, le site 5' proximal est utilisé.

### VI.3.3. Architecture des exons

La taille des exons ainsi que celle des introns sont des paramètres qui peuvent réguler le choix des sites donneurs lors de l'épissage. Comme nous l'avons vu dans le chapitre I, la reconnaissance des sites d'épissage varie en fonction la taille de l'intron suivant deux modèles : la définition de l'exon et la définition de l'intron (**Chapitre I, III.3.2**). Les sites d'épissage sont plus efficacement reconnus lors de la définition de l'intron que lors de la définition de l'exon ce qui, par conséquent, régule en partie souvent le choix des sites donneurs en compétition (Fox-Walsh *et al.*, 2005). Lorsque deux sites donneurs de forces identiques sont en compétition pour l'association avec un même site accepteur, le site donneur proximal sera préférentiellement choisi (Reed et Maniatis, 1986).

La distance séparant les deux sites donneurs en compétition influence également le choix du site donneur. Lorsque les deux sites donneurs sont proches, la séquence ayant la plus grande complémentarité de séquence avec celle du snRNA U1 est préférentiellement choisi (Roca *et al.*, 2005) (**Figure II-20, A**). Lorsque les sites donneurs sont distants d'environ 180 nucléotides, il a été suggéré que le site donneur fort situé en amont active l'épissage du site proximal situé en aval, et ce indépendamment de la force de ce dernier et de la taille de l'intron (Hicks *et al.*, 2010) (**Figure II-20, B**). Le mécanisme de cette activation n'est pas connu mais serait dépendant de la distance séparant les deux sites.

La compétition entre deux sites donneurs peut également être régulée par la présence d'une séquence ESE. Son effet varie suivant sa position relative aux sites donneurs. Il a récemment été suggéré que la présence d'une séquence ESE placée entre les deux sites donneurs ait une activité bi-directionnelle, activant le site proximal et réprimant le site distal (Hicks *et al.*, 2010) (**Figure II-20, C**).

Il existe aussi des régulations permettant d'activer l'épissage au site donneur distal. Comme nous l'avons déjà abordé, ce mécanisme repose sur la formation d'une boucle de l'intron impliquant deux séquences introniques ISS sur lesquelles sont fixées les protéines hnRNP A1 qui interagissent entre elles (Blanchette et Chabot, 1999 ; Nasim *et al.*, 2002) (**Figure II-13**). Cette régulation existe également lorsque l'intron possède un site de fixation pour la protéine hnRNP A1 et un second site de fixation pour la protéine hnRNP H (Fisette *et al.*, 2010 ; Martinez-Contreras *et al.*, 2006).

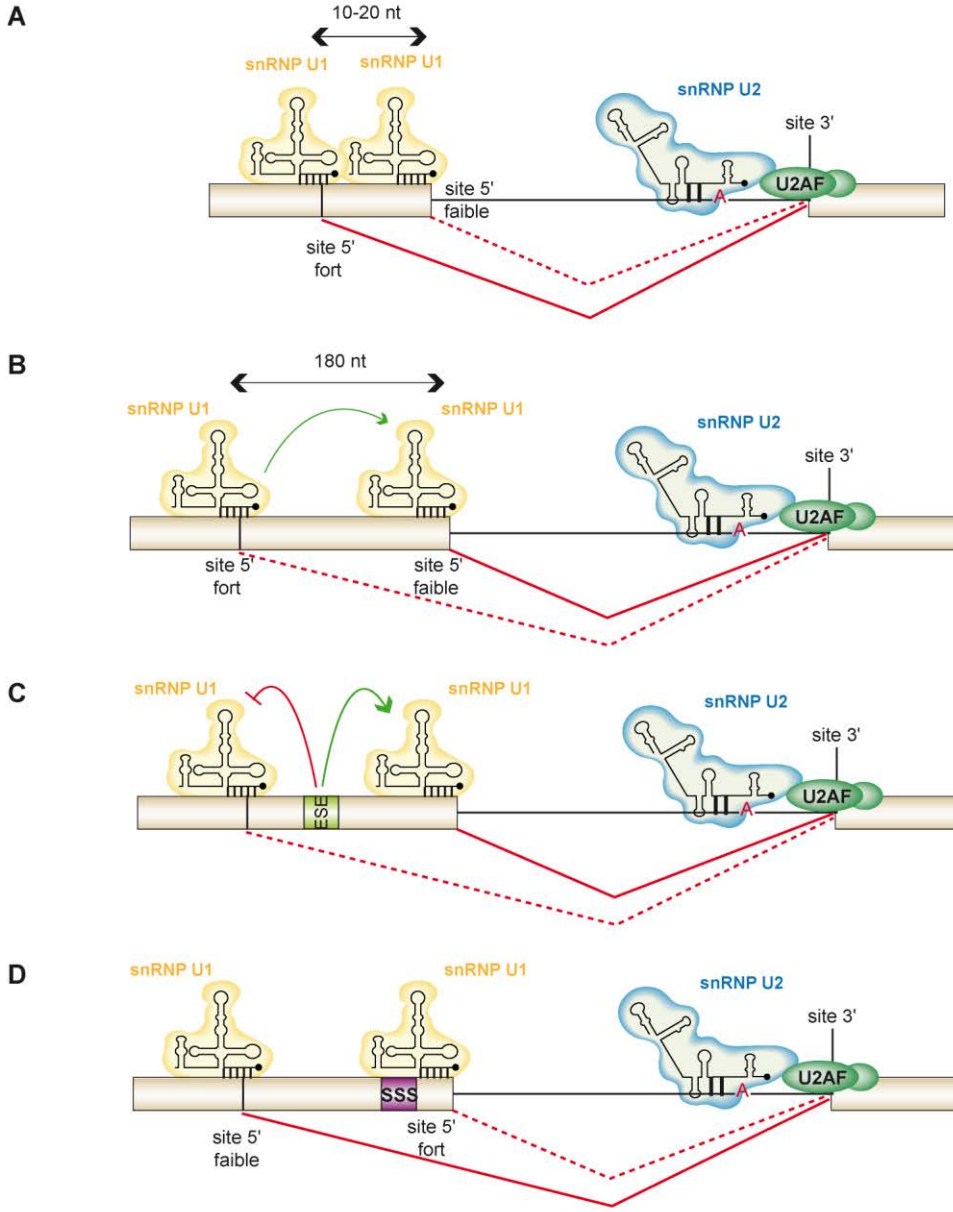



**Figure II-20 : Exemples de régulation de l'épissage entre deux sites donneurs en compétition.**
**A :** Lorsque les deux sites donneurs sont très proches, le site le plus fort est choisi lors de l'épissage.
**B :** Lorsque les deux sites donneurs sont éloignés, le site donneur fort active le site proximal.
**C :** La présence d'une séquence ESE entre les deux sites donneurs favorise l'utilisation du site proximal.
**D :** La séquence SSS associée au site proximal active l'utilisation du site distal, même si celui-ci est faible.
L'épissage du site donneur majoritairement utilisé est schématisé par un trait rouge, celui du site donneur minoritairement utilisé est schématisé par des pointillés rouges.

Une autre classe de séquences *cis* régulant la compétition de deux sites donneurs, appelée SSS (Strong Splicing Silencer) a été identifiée. Ces séquences sont spécifiquement associée avec un site donneur proximal en compétition avec un site donneur distal situé en amont (Yu *et al.*, 2008). Les séquences SSS répriment l'utilisation d'un site donneur proximal fort au profit de l'activation du site donneur distal faible (**Figure II-20, D**). Les séquences SSS

n'affectent pas la fixation de la snRNP U1 sur le site donneur qu'elles inhibent mais l'empêchent d'avoir une conformation optimale qui lui permette d'entrer en compétition avec l'autre site donneur pour s'engager dans un complexe spliceosomal avec le site accepteur (Yu *et al.*, 2008). Ces séquences ont également la particularité d'être autonomes et de n'avoir besoin d'aucun partenaire protéique pour exercer cette inhibition.

### VI.3.4. Régulation de l'épissage par les pseudo sites et les pseudoexons

Les pseudoexons sont des séquences de 50 à 2-300 nucléotides de long qui sont présentes dans la plupart des introns et qui possèdent un site donneur et un site accepteur théoriquement fonctionnels à leurs extrémités (Dhir et Buratti, 2010 ; Sun et Chasin, 2000). Cependant, ces sites ne sont jamais épissés, souvent à cause de la présence de séquences inhibitrices à leur proximité (Fairbrother et Chasin, 2000 ; Sironi *et al.*, 2004 ; Zhang et Chasin, 2004). Certaines de ces séquences sont des pseudo sites donneurs d'épissage car leurs séquences s'apparentent à celle du site donneur consensus (/GTRAGT) (Zhang et Chasin, 2004). La fixation de la snRNP U1 à certaines de ces séquences constitue un mécanisme d'inhibition du pseudoexon, comme c'est le cas des pseudoexons des gènes ATM et GHR (Akker *et al.*, 2007 ; Dhir *et al.*, 2010) (**Figure II-21**).

Dans le cas du gène ATM, la fixation de la snRNP U1 sur un site donneur situé au centre du pseudoexon (séquence ISPE) empêche la reconnaissance des sites donneur et accepteur bordants le pseudo exon et induit l'exclusion du pseudo exon de l'ARNm (Dhir *et al.*, 2010) (**Figure II-21, A**). L'absence de la séquence ISPE provoque la perte de cette inhibition, l'inclusion du pseudo exon dans l'ARNm qui est responsable de la maladie ataxia-telegiectasia (Pagani *et al.*, 2002).

Dans le cas du gène GHR, la fixation de la snRNP U1 sur le site donneur du pseudo exon plus celle de la hnRNP E1 sur une séquence ESS, inhibent l'épissage du site selon un mécanisme qui n'est pas connu (Akker *et al.*, 2007) (**Figure II-21, B**). La mutation A/G du nucléotide à la position -1 du site donneur empêche l'inhibition d'avoir lieu, ce qui induit l'inclusion du pseudo exon dans l'ARNm qui est responsable du syndrome de Laron (Maamra *et al.*, 2006).

Les pseudo sites donneurs peuvent également inhiber l'épissage, comme c'est le cas de l'élément P de la drosophile (Rio, 1991 ; Siebel *et al.*, 1992). Le 3$^{e}$ exon codant la transposase comprend un pseudo site donneur qui se situe 30 nucléotides en amont du site donneur

fonctionnel. La snRNP U1 est recrutée sur le pseudo site donneur par la protéine PSI (P-Element Somatic Inhibitor) qui interagit directement avec la protéine U1-70K (Labourier *et al.*, 2001 ; Labourier *et al.*, 2002). Avec la protéine hrp48, qui est l'homologue de la protéine hnRNP A1 humaine, ces protéines forment un complexe qui inhibe l'épissage au site donneur fonctionnel (**Figure II-21, C**). La protéine PSI est exclusivement somatique et son absence des cellules germinales permet l'épissage de l'exon 3 de l'élément P (Siebel *et al.*, 1995 ; Siebel *et al.*, 1994).

Il existe un autre exemple de pseudo site donneur, l'élément NRS du virus RSV, qui a fait l'objet de mon travail de thèse, et que nous aborderons plus en détail dans le IVe chapitre de cette introduction.

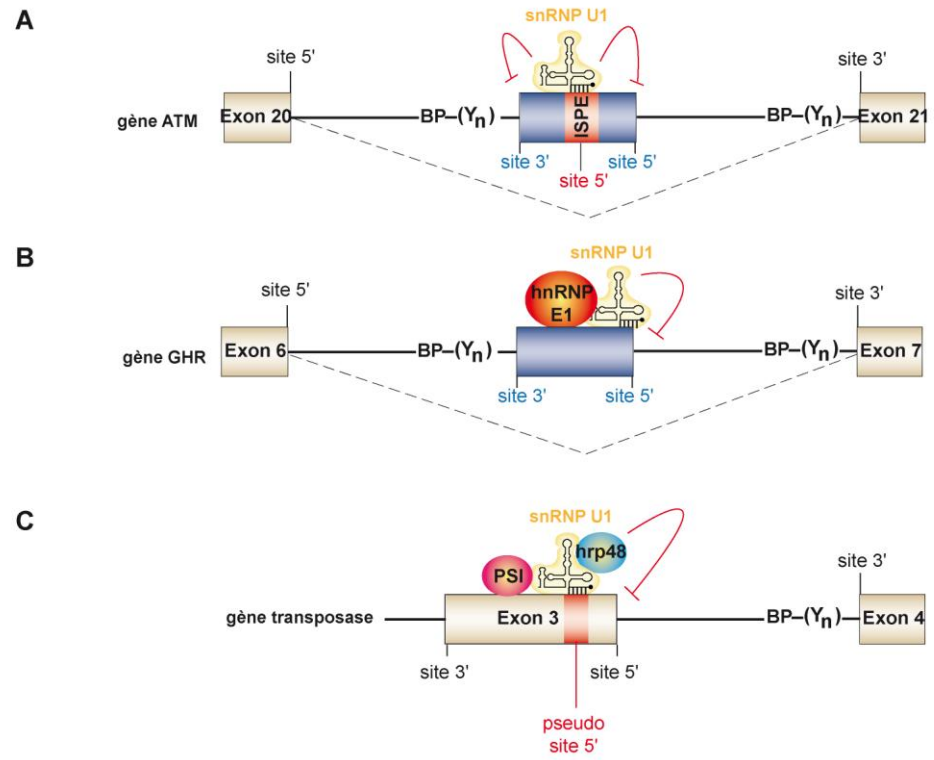



**Figure II-21 : Inhibition de l'épissage par la fixation de la snRNP U1 sur des sites donneurs (Dhir et Buratti., 2010, Smith et Valcarcel, 2000)**

**A :** La fixation de la snRNP U1 sur le site donneur au centre du pseudo exon (séquence ISPE) inhibe l'épissage de ce pseudoexon.

**B :** La fixation de la snRNP U1 sur le site donneur du pseudo exon plus celle de la protéien hnRNP E1 sur une séquence ESS inhibe l'épissage du pseudo exon.

**C :** La fixation des protéines hrp48 et PSI sur l'exon 3 ainsi que la fixation de la snRNP U1 sur un pseudo site donneur inhibent l'épissage de l'exon 3.

Les pseudo exons sont représentés par des rectangles bleus.

### VI.4. Importance de la structure secondaire

La structure secondaire est un paramètre qui peut réguler le choix des sites d'épissage, mais aussi l'épissage des pseudo exons ou la fonction des séquences *cis*.

La plupart des protéines impliquées dans la régulation de l'épissage interagissent avec des séquences d'ARN simple brin, ce qui implique que la majorité des séquences *cis* régulatrices doivent être en simple brin pour être fonctionnelles (Auweter *et al.*, 2006 ; Hiller *et al.*, 2007).

Le pré-ARNm du gène de la fibronectine possède un exon EDA dont l'épissage est régulé par une séquence ESE et une séquence ESS. La séquence ESE doit être en simple brin dans une boucle apicale pour que la protéine ASF/SF2 puisse s'y fixer et induire l'inclusion de l'exon EDA (Lavigueur *et al.*, 1993 ; Muro *et al.*, 1999) (**Figure II-22, A**). La troncation de la séquence ESS qui devrait théoriquement activer l'épissage, inhibe l'épissage de l'exon en raison du remaniement de la structure secondaire de l'élément ESE qu'elle induit (Muro *et al.*, 1999). La séquence ESE se retrouve alors dans une structure double brin qui l'empêche d'interagir avec la protéine ASF/SF2 et donc d'être fonctionnelle (Muro *et al.*, 1999). Le rôle de la séquence ESS serait donc de permettre à la séquence ESE d'être structurée en simple brin au sein du pré-ARNm. Cela montre que contexte est aussi un paramètre important pour la fonction de certaines séquences *cis*.

Il existe également de nombreux exemples montrant que la structure secondaire de la séquence d'un site donneur est importante pour son appariement avec le snRNA U1.

Le pré-ARNm de H-*ras* possède un exon épissé alternativement, IDX, dont l'épissage est activé par les protéines SC35 et SRp40 qui s'y fixent et réprimé par la protéine hnRNP A1 qui se fixe à la séquence ISS rasISS1 située en aval (Guil *et al.*, 2003). L'épissage de l'exon IDX serait aussi régulé par son appariement au sein d'une longue structure tige boucle avec la séquence rasISS1. La fixation de la protéine hnRNP H sur cette structure activerait l'épissage de l'exon IDX (Camats *et al.*, 2008). Cela serait contrecarré par l'hélicase p68 qui viendrait déstructurer la structure tige boucle, ce qui provoquerait le départ de la protéine hnRNP H et la diminution de l'épissage de l'exon IDX (Camats *et al.*, 2008). Le mécanisme exact de l'activation de l'épissage par la protéine hnRNP H dans ce contexte n'est pas connu.

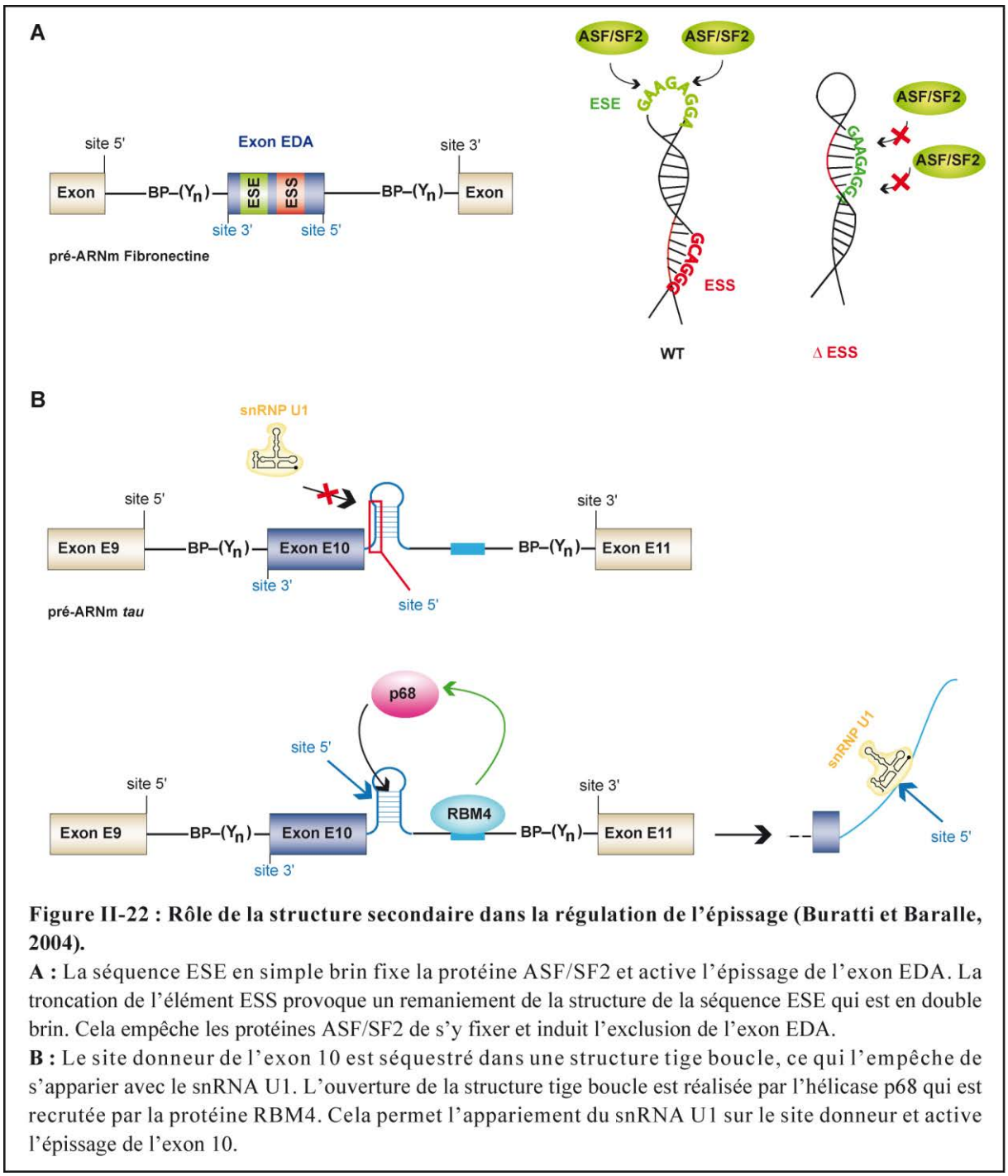



**Figure II-22 : Rôle de la structure secondaire dans la régulation de l'épissage (Buratti et Baralle, 2004).**

**A :** La séquence ESE en simple brin fixe la protéine ASF/SF2 et active l'épissage de l'exon EDA. La troncation de l'élément ESS provoque un remaniement de la structure de la séquence ESE qui est en double brin. Cela empêche les protéines ASF/SF2 de s'y fixer et induit l'exclusion de l'exon EDA.

**B :** Le site donneur de l'exon 10 est séquestré dans une structure tige boucle, ce qui l'empêche de s'apparier avec le snRNA U1. L'ouverture de la structure tige boucle est réalisée par l'hélicase p68 qui est recrutée par la protéine RBM4. Cela permet l'appariement du snRNA U1 sur le site donneur et active l'épissage de l'exon 10.

L'épissage au site donneur de l'exon 10 du pré-ARNm *tau* est régulé par la structure secondaire qu'il adopte en plus des nombreuses régulations faisant intervenir d'autres séquences *cis* régulatrices. Le site donneur se situe dans une structure tige boucle qui le rend peu accessible pour s'apparier avec le snRNA U1 (Jiang *et al.*, 2003 ; Varani *et al.*, 1999) (**Figure II-22, B**). Les mutations qui déstabilisent la structure tige boucle provoquent une augmentation de l'épissage de l'exon 10, ce qui est à l'origine de taupathies (Liu et Gong, 2008) (**Figure II-22, B**). Il a récemment été démontré que l'hélicase p68 favorisait

l'appariement du snRNA U1 sur le site donneur en déstabilisant la structure tige boucle qui le séquestre (Kar *et al.*, 2011). Il semblerait qu'elle puisse être recrutée sur le site donneur avant la snRNP U1 par la protéine RBM4 qui se fixe sur une séquence ISE en aval du site donneur (Kar *et al.*, 2011) (**Figure II-22, B**).

Ces exemples illustrent bien le fait que la régulation exercée par la structure secondaire sur l'efficacité d'épissage d'un site donneur est multi-paramétrique.

# Chapitre III : Le rétrovirus RSV

## Avant propos

Le virus du sarcome de Rous ou RSV a été découvert il y a un siècle par le scientifique américain éponyme Peyton Rous (Rous, 1911). L'expérience qui a permis cette découverte a consisté à inoculer un filtrat acellulaire provenant d'une tumeur d'un poulet à un poulet sain qui, à son tour, a développé la même tumeur, un fibrosarcome. Cette découverte démontrait que la formation d'une tumeur pouvait être provoquée par un virus (ou agent filtrant), ce qui constituait un paradigme très novateur dans le contexte scientifique de l'époque. Celui-ci fut accueilli avec le plus grand scepticisme par la communauté scientifique qui considérait alors que la formation des tumeurs ne pouvait avoir qu'une origine endogène à la cellule et non une origine infectieuse. Par ailleurs, les échecs des tentatives d'inoculation de ce cancer à d'autres espèces animales comme la souris ou le fait que le poulet soit considéré comme un modèle d'étude trop distant par rapport aux maladies humaines, détournèrent l'attention des scientifiques de cette découverte. Il fallut attendre les années 50 et l'avancée des techniques de culture cellulaire pour que les doutes quant à la théorie de Rous se dissipent. En 1955, H. Rubin démontra que chaque cellule issue d'une tumeur induite par une infection du RSV était capable de relarguer des virions infectieux sans que cela provoque la mort cellulaire (Rubin, 1955). En 1958, l'étude de la formation des foci viraux permit d'observer le changement morphologique des cellules embryonnaires de poulet transformées (ou CEF, Chicken Embryo Fibroblast) après leur infection par le RSV (Temin et Rubin, 1958). Ainsi, le virus de Rous fut clairement identifié comme l'agent cancérigène et qualifié d'oncovirus. P. Rous recevra à cet égard, 55 ans après sa découverte, le prix Nobel de médecine qui récompensa ses travaux précurseurs. Les années 60 et 70 constituèrent ensuite l'âge d'or de la biologie du RSV qui, de concert avec l'explosion des avancées en biologie moléculaire, ont abouti aux deux des plus grandes découvertes de la virologie moderne. En premier lieu, les travaux portant sur le RSV permirent en 1970 aux chercheurs H. Témin et D. Baltimore de découvrir la transcriptase inverse (Baltimore, 1970 ; Temin et Mizutani, 1970). Cette découverte bouleversa les dogmes de la biologie moléculaire et valut aux deux scientifiques le prix Nobel de médecine en 1975. Par la suite, le processus de tumorisation induit par le RSV fut abordé à travers les études de différentes souches du virus qui démontrèrent que le processus de cancérogenèse était d'une part indépendant de la réplication virale et d'autre part que le matériel génétique viral était suffisant pour l'induire (Hanafusa *et al.*, 1963 ; Martin, 1970 ; Rubin et Vogt, 1962). L'étude

du génome viral du virus RSV mena ainsi à l'identification du gène viral responsable de la formation des sarcomes : le gène v-*src* qui est une forme tronquée du gène originel cellulaire c-*src* (Stehelin *et al.*, 1976). Cette découverte permit de comprendre comment certains gènes cellulaires, appelés « proto oncogènes » pouvaient acquérir des propriétés oncogéniques et devenir des gènes « oncogènes » par mutation ou perte d'une partie de leur séquence. Cette compréhension de l'origine cellulaire du développement du cancer marqua un tournant dans le domaine de la cancérologie moderne et valu aux scientifiques JM. Bishop et H. Varmus de recevoir le prix Nobel de médecine en 1989.

Bien que le virus RSV fut le premier rétrovirus oncogénique découvert, ce virus reste encore aujourd'hui largement étudié, que ce soit au niveau de l'expression des gènes viraux, de l'assemblage des virions *in vitro* ou de l'oncogenèse. Le virus RSV est un rétrovirus simple par comparaison avec les rétrovirus dits complexes comme le virus de l'immunodéficience acquise humaine (VIH) qui contient davantage de gènes. Cependant, de nombreuses régulations virales sont communes aux deux types de virus ce qui a permis par exemple à de nombreuses études initiées chez le virus RSV d'être étendues par la suite au HIV. Par ailleurs, le fait qu'il soit facilement cultivable dans les conditions de laboratoire et non pathogène pour l'homme font de lui un bon modèle d'étude. En dépit des connaissances accumulées depuis sa découverte, certains aspects du cycle viral de ce virus sont encore mal compris. Ainsi, comme nous le verrons ultérieurement, nous avons focalisé nos études sur la compréhension du mécanisme de régulation de l'épissage chez ce virus.

## I. Caractéristiques générales des rétrovirus

Les rétrovirus sont des virus enveloppés d'environ 100 nm de diamètre possédant une géométrie icosaédrique. Le cœur de la particule virale contient la capside qui renferme le génome constitué de deux molécules d'ARN simple brin de polarité positive étroitement associées aux protéines de la nucléocapside. La particule virale est tapissée sur sa face interne de protéines de la matrice et est recouverte sur sa face externe par une bicouche lipidique dans laquelle sont enchassées les protéines de l'enveloppe virale responsables de la reconnaissance des récepteurs cellulaires de la cellule hôte qui sera infectée. Les particules rétrovirales contiennent également la transcriptase inverse, une ADN polymérase ARN et ADN dépendante, qui permet la synthèse de l'ADNc viral à partir de l'ARN génomique. Une fois rétrotranscrit, cet ADNc est intégré dans le génome de la cellule hôte grâce à l'enzyme virale intégrase pour former l'ADN proviral. A ce stade, toutes les étapes nécessaires à l'expression des gènes viraux seront assurées par la machinerie cellulaire de la cellule hôte. Après leur

synthèse, toutes les protéines et les ARN viraux s'assemblent pour former les virions qui, une fois libérés de la cellule hôte par bourgeonnement, vont infecter d'autres cellules.

## II. Le RSV et les différents rétrovirus aviaires

Le virus RSV appartient à la famille des *Retroviridae* et à la sous-famille des *Orthoretroviridae*. Cette sous-famille est elle-même divisée en 6 genres qui sont classés selon la morphologie de la particule virale et les propriétés infectieuses des virus. Le RSV appartient au genre des *Alpharétrovirus*, qui comprend des rétrovirus dits simples, ne possédant que les gènes *gag*, *pol* et *env* communs à tous les rétrovirus, par opposition aux rétrovirus dits complexes comme le HIV-1, du genre des lentivirus, qui possèdent en plus de ces gènes de nombreux gènes codant des protéines auxiliaires et régulatrices. Les espèces virales du genre des *Alpharétrovirus* sont relativement proches les unes des autres et sont responsables de la plupart des néoplasmes aviaires comme les sarcomes, les leucémies lymphoblastiques, érythroblastiques et myeloblastiques, qui sont respectivement induites par les virus RSV, ALV (Avian Leukosis Virus), AEV (Avian Erythroblastosis Virus) et AMV (Avian Myeloblastosis Virus) (Payne, 1998). C'est notamment pour cette raison que ces virus dans leur ensemble sont également regroupés sous l'appellation d'ASLV pour Avian Sarcoma Leukosis Virus. Tous ces néoplasmes sont induits à 100 % par les rétrovirus, soit en raison de la présence d'un oncogène dans le génome viral (**Figure III-1, A et B**), soit en raison de l'activation d'un proto-oncogène après l'intégration de l'ADN proviral dans le génome de la cellule hôte (**Figure III-1, C**).

Le virus RSV possède dans son génome les gènes *gag*, *pol* et *env* qui lui permettent de se répliquer ainsi que l'oncogène v-*src* qui provoque la transformation des cellules infectées et la formation de sarcome, ce virus est ainsi qualifié de virus transformant à réplication autonome (**Figure III-1, A**). D'autres virus possèdent également dans leur génome des oncogènes mais des troncations dans les gènes viraux essentiels ne leur permettent pas de se répliquer de façon autonome, ces virus sont qualifiés de virus transformant à la réplication déficiente (**Figure III-1, B**). Pour ces virus, il est nécessaire que la cellule soit simultanément infectée par un autre virus qui fournit au virus déficient la protéine manquante, ces derniers sont souvent des virus ALV et sont appelés virus helper (**Figure III-1, B**). Enfin, certains virus possèdent les gènes *gag*, *pol* et *env* qui leur permettent d'être autonomes dans leur réplication et peuvent induire des processus de tumorigenèse en s'insérant dans ou en amont/aval d'un proto-oncogène cellulaire (**Figure III-1, C**). C'est notamment le cas du virus ALV LR9 dont

les différentes insertions provoquent l'activation des gènes c-*myc* et c-*myb* (Hayward *et al.*, 1981 ; Smith *et al.*, 1997).

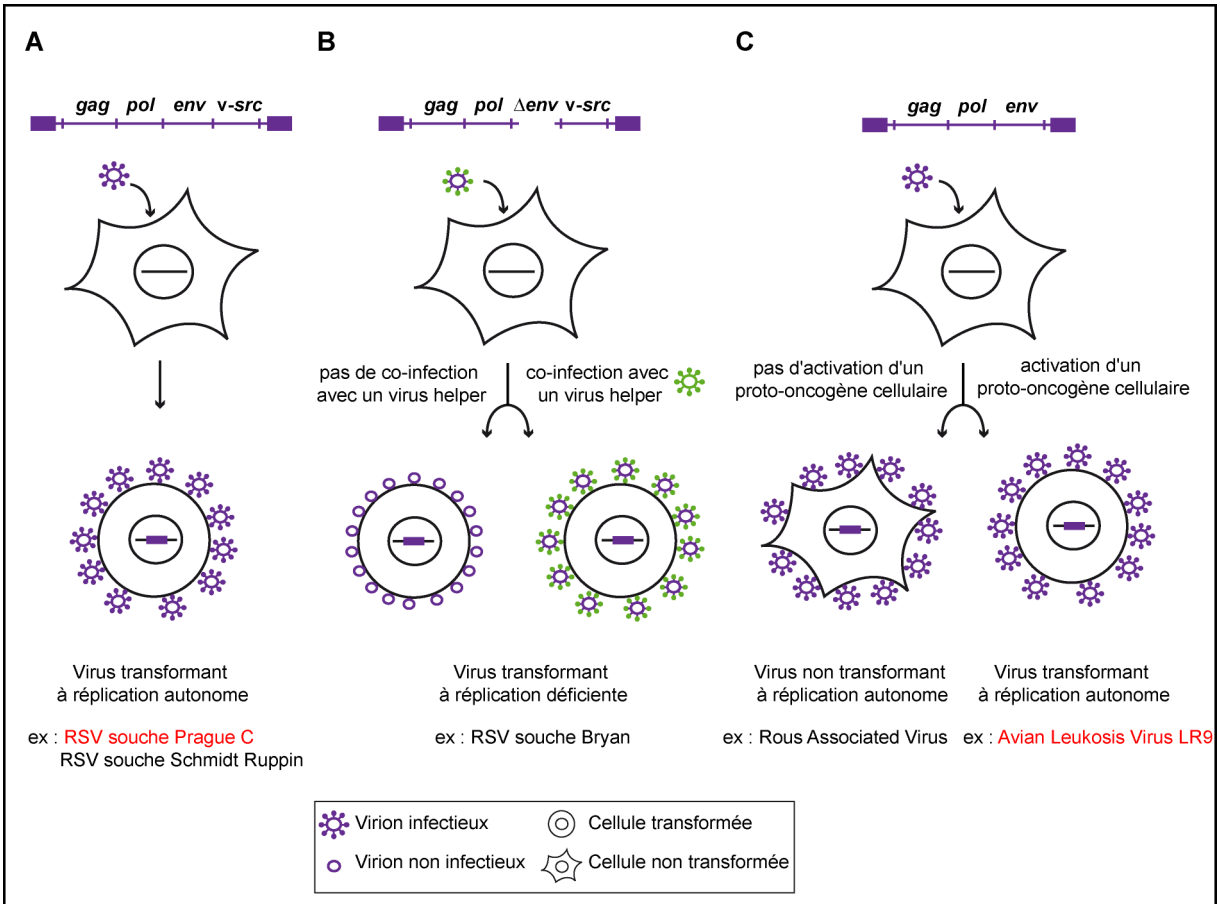



**Figure III-1 : Infection et transformation des cellules par différents rétrovirus aviaires (d'après Martin, 2004).**
**A** : Exemple de transformation d'une cellule par un virus transformant à réplication autonome. Le virus possède dans son génome tous les gènes (*gag*, *pol*, *env*) nécessaires à sa réplication ainsi que l'oncogène v-*src* responsable de la transformation de la cellule. La cellule infectée par ce virus est transformée et relargue des virions infectieux.
**B** : Exemple de la transformation d'une cellule par un virus transformant à la réplication déficiente. La cellule infectée par ce virus est transformée en raison de la présence de l'oncogène v-*src* dans le génome du virus. En revanche, l'absence d'un des gènes viraux essentiels à la réplication virale (ici le gène *env*) ne permet pas aux virions relargués d'être infectieux en raison du défaut de protéine de l'enveloppe à la surface des particules virales. Seule une co-infection de la cellule par un virus helper (représenté en C) permet d'apporter la protéine virale manquante et le relargage de virions infectieux.
**C** : Exemple d'infection d'une cellule par un virus à la réplication autonome. Le virus possède tous les gènes viraux nécessaires à sa réplication et ne possède pas d'oncogène dans son génome. Ces virus sont souvent des virus helper. Nénamoins, la cellule infectée peut être transformée lorsque l'insertion de l'ADN proviral active un proto-oncogène cellulaire. Dans les deux cas, la cellule infectée relargue des virions infectieux.
Les souches virales avec lesquelles nous avons travaillé sont indiquées en rouge.

Naturellement, les rétrovirus aviaires existent sous formes exogènes et endogènes c'est-à-dire soit sous forme de particule virale, soit sous forme d'ADN proviral intégré au génome de la cellule hôte. Les virus endogènes et exogènes se transmettent majoritairement par voie verticale, congénitalement de la poule à l'œuf et à la descendance ou génétiquement par les gamètes après l'accouplement. Les virus exogènes peuvent aussi se transmettre horizontalement par contact direct entre deux oiseaux mais cette voie de contamination reste largement minoritaire en raison de la courte durée de vie des virus en dehors des organismes (Weiss, 2006).

## III. Phylogénie des rétrovirus aviaires

Les virus du genre des *Alpharétrovirus* sont répertoriés en souches selon les caractéristiques de l'hôte naturel infecté qui sont définies par les protéines de l'enveloppe virale et plus particulièrement par la protéine de l'enveloppe, SU ou gp 85, qui interagit avec les récepteurs cellulaires de la cellule (Venugopal, 1999). C'est notamment pour cette raison que la phylogénie de ces virus est basée sur les comparaisons des séquences du gène viral *env*. Les souches sont classées en 9 groupes allant de A à J. Les groupes A à E et J infectent spécifiquement les poulets, tandis que les groupes F, G, H, I infectent d'autres espèces aviaires à savoir les faisans, les perdrix et les cailles, respectivement (Venugopal, 1999) (**Tableau III-2**).

Les groupes A à E sont relativement proches (77 à 85 % d'identité du gène *env*), tandis que le groupe J est plus distant (36 à 45 % d'identité du gène *env*) (**Figure III-3**) (Venugopal, 1999). Les groupes A à E et J sont les groupes les plus étudiés en raison des problèmes économiques importants qu'ils engendrent au niveau de l'industrie agroalimentaire.

Le virus RSV de la souche Prague C que nous avons étudié, appartient au groupe C qui est, pour des raisons historiques, le groupe le plus étudié au niveau du cycle viral dans sa globalité. Nous avons également étudié la souche ALV LR9 qui est un rétrovirus chimérique créé artificiellement et qui comprend les séquences LTR du virus RPV (Ring-necked Pheasant Virus) du groupe J et les gènes *gag*, *pol* et *env* du virus UR2AV (Avian sarcoma Virus UR2) du groupe A (Simon *et al.*, 1987).

| sous groupe | souche de virus type | Espèce aviaire hôte |
|---|---|---|
| A | ALV-RSA<br>RAV-1 | poulet |
| B | RAV-2 | poulet |
| C | RSV-PR-C<br>RAV-49 | poulet |
| D | RSV-SR-D<br>RAV-50 | poulet |
| E | RAV-0 | poulet |
| F | RAV-61 | faisan de Colchide |
| G | Golden pheasant virus | faisan doré |
| H | RAV-62 | perdrix hongroise |
| I | Gambel's quail virus | caille de Gambel |
| J | HPRS-103 | poulet |

**Tableau III-2 : Classification des différents sous-groupes de rétrovirus du genre *Alpharétrovirus* selon la séquence de la protéine de l'enveloppe gp 85 (SU) et les espèces aviaires hôtes (d'après Venugopal, 1999).**
ALV-RSA : Avian Leukosis virus de la souche RSA ; RAV : Rous-associated virus ; RSV-PR : Rous Sarcoma Virus de la souche Prague ; RSV-SR : Rous Sarcoma Virus de la souche Schmidt-Rupin ; HPRS : Houghton Poultry Research Station.
La souche RSV Prague-C (RSV-PR-C) utilisée au cours de ce travail est indiquée en rouge.

## IV. Pathologies et implications économiques liées aux rétrovirus aviaires

Les rétrovirus aviaires RSV et ALV sont responsables de la majorité des infections des espèces aviaires et plus particulièrement des espèces domestiques de l'industrie avicole (Payne, 1998). Ces rétrovirus provoquent la formation de néoplasmes dont la forme la plus répandue est la leucémie lymphoblastique à cellules B. Cette leucémie débute par la transformation des cellules de la bourse de Fabricius, qui est un organe lymphoïde primaire, pour ensuite former des métastases qui s'étendent au foie, à la rate et à l'ensemble des organes des viscères et qui entraînent la mort de l'animal (Fadly, 2008). Des leucémies érythroblastiques et myeloblastiques sont également induites par l'infection de ces rétrovirus mais sont moins répandues. Hormis les leucémies, ces virus provoquent également la

formation de tumeurs solides, majoritairement des sarcomes, qui résultent de l'infection des cellules des tissus conjonctifs.

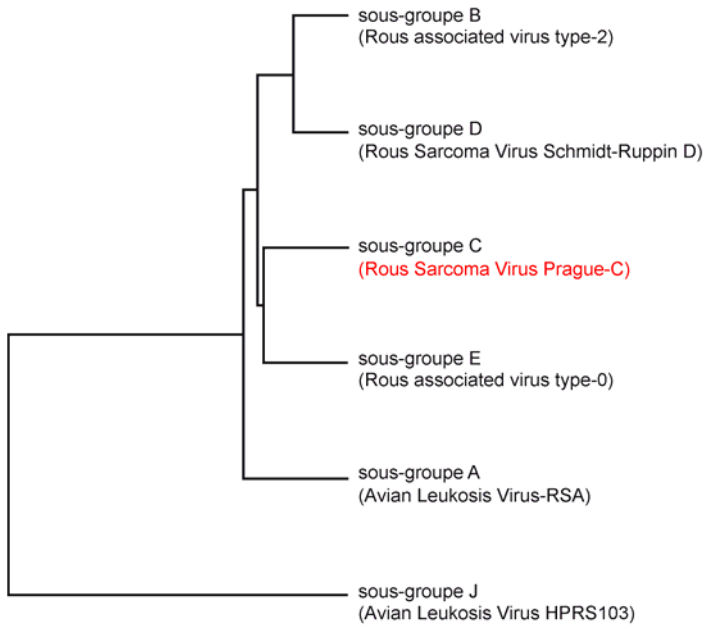



**Figure III-3 : Dendogramme représentant les différents sous-groupes de rétrovirus du genre *Alpharétrovirus* infectant le poulet (d'après Bai *et al.*, 1995).**
Les liens phylogénétiques ont été établis par le programme PILEUP avec les alignements des séquences de la protéine gp 85 (SU) des différents virus (indiqués entre parenthèse). La souche RSV Prague-C utilisée au cours de ce travail est indiquée en rouge.

Pour chacune de ces pathologies, le temps d'incubation varie suivant la dose virale, l'âge et le génome de l'animal et surtout de la souche virale. Ainsi, une leucémie lymphoblastique provoquera la mort de l'animal dans les 4 à 5 semaines après l'infection si le virus est un virus transformant qui possède un oncogène dans son génome viral. En revanche, ce délai est de 4 à 12 mois dans le cas d'un virus ne possédant aucun oncogène mais dont l'insertion de l'ADN proviral dans le génome de la cellule hôte active un proto-oncogène cellulaire (Fadly, 2008).

D'un point de vue industriel, l'infection des espèces aviaires par les rétrovirus n'a que très peu d'incidence sur le taux de mortalité (perte de 1 à 2 % du cheptel avicole) et se répercute essentiellement au niveau de la diminution de la productivité (fréquence de la ponte, qualité des œufs, taux de reproduction, poids des animaux) (Payne, 1998). En d'autres termes, seuls les virus qui provoquent des tumeurs avec un temps de latence important et les virus endogènes (rétrovirus du sous-groupe E principalement) qui ne sont pas ou peu oncogéniques sont responsables des pertes économiques du secteur avicole (Payne, 1998). A l'heure actuelle,

aucun vaccin n'existe et la stratégie adoptée afin d'éradiquer les rétrovirus a été de sélectionner génétiquement des lignées d'espèces aviaires ne possédant aucun virus endogène dans leur génome (Payne, 1998). La plupart des tests de dépistage des rétrovirus chez les espèces aviaires reposent sur la méthode ELISA, ainsi que sur des techniques de RT-PCR (Fadly, 2000 ; Kreager, 1998).

## V. Structure du rétrovirus RSV (Figure III-4)

Le virus du RSV est un virus enveloppé de structure sphérique, d'environ 100 nm de diamètre. Sa membrane externe est composée d'une bicouche lipidique provenant de la cellule dont il a bourgeonné. Cette bicouche est tapissée sur sa face interne des protéines de la matrice, MA (ou p19) tandis que dans sa face externe sont enchassées les protéines de l'enveloppe. Une de ces protéines est exposée en surface, la protéine SU (ou gp85) et interagit avec la seconde protéine de l'enveloppe, la protéine TM (ou gp37) qui, elle, est transmembranaire.

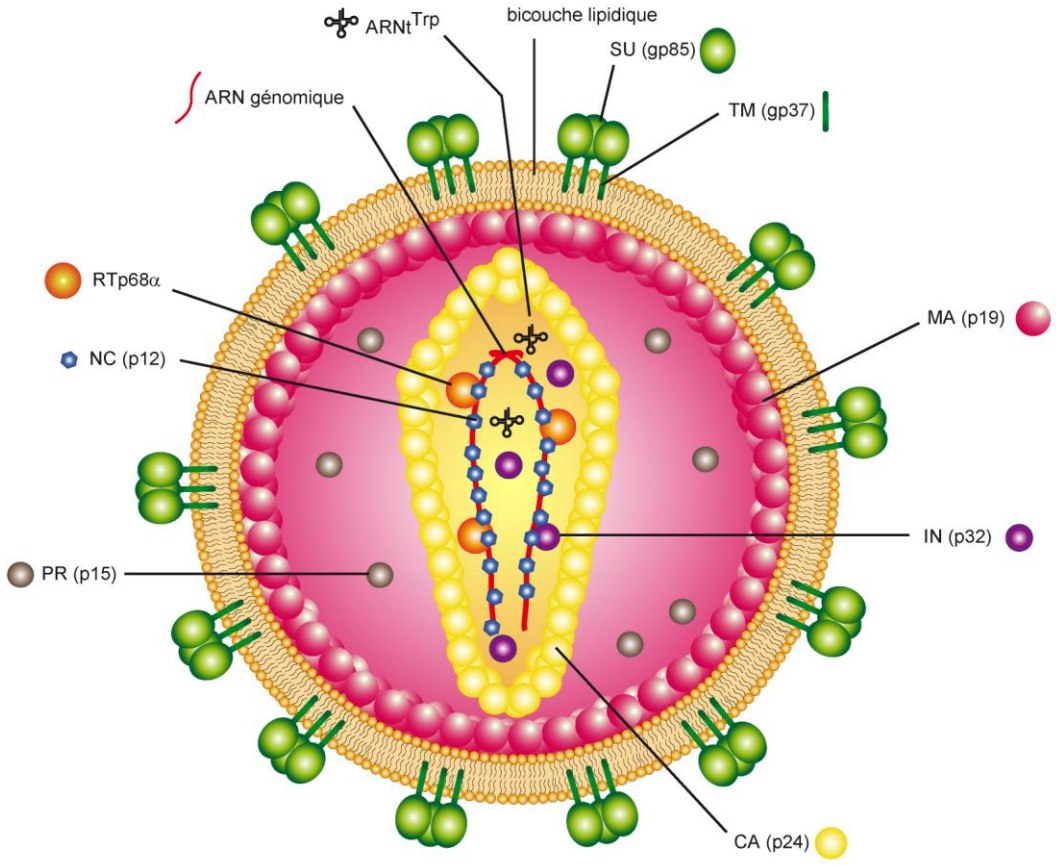



**Figure III-4 : Structure du virion mature du virus RSV (adapté de Coffin *et al.*, 1997)**
SU (gp85) : protéine de surface ; TM (gp37) : protéine transmembranaire ; CA (p24) : protéine de la capside ; MA (p19) : protéine de la matrice ; NC (p12): protéine de la nucléocapside ; RT (p68α) : transcriptase inverse ; PR (p15) : protéase ; IN (p32) : intégrase.

Les protéines de l'enveloppe forment des trimères à la surface de la particule virale et leur rôle est de reconnaître les récepteurs cellulaires de la cellule hôte. Le cœur de la particule virale contient la capside (protéine CA ou p27), qui renferme le génome constitué de deux molécules d'ARN simple brin de polarité positive étroitement associées aux protéines de la nucléocapside (NC ou p12). La capside renferme également les enzymes virales telles que la transcriptase inverse (RT ou p68α), l'intégrase (IN ou p32) et aussi l'ARNt$^{Trp}$. La protéase (PR ou p15) est quand à elle présente uniquement dans l'espace de la matrice.

## VI. L'ARN génomique

La longueur de l'ARN génomique varie selon les souches et est en moyenne de 7,5 kb. Pour les virus transformants dont le génome comporte un oncogène, cette longueur peut varier de 3,2 kb (UR2 sarcoma virus) à 9,3 kb (RSV Prague C) suivant les délétions plus ou moins importantes des gènes *gag*, *pol* et *env*. L'ARN génomique est produit et maturé par la machinerie cellulaire de la cellule hôte. Il est produit par transcription par l'ARN polymérase II cellulaire et comporte à son extrémité 5' la coiffe m7GpppN et à son extrémité 3' une queue polyA.

### VI.1. Les régions non traduites

Les régions non traduites (UTR, UnTranslated Region) se situent aux extrémités 5' et 3' de l'ARN génomique. La région 5' UTR comprend la séquence en amont du site d'initiation de la traduction du gène *gag*, les séquences R (Redondant), U5 (Unique en 5') et la région 5' leader. La région 3' UTR comprend la séquence en aval du site de terminaison du gène v-*src*, la séquence U3 (Unique en 3') et R. Après l'étape de transcription inverse, les séquences U3 et U5 sont dupliquées aux extrémités 5' et 3' de l'ADNc pour former les séquences LTR 5' et 3' (Long Terminal Repeat) ainsi composées respectivement des séquences U3-R-U5 et U5-R-U3.

#### VI.1.1. La région 5' non traduite

Les séquences essentielles à la transcription se situent dans la partie 5' LTR et sont composées des séquences « enhancer », du promoteur dans la partie U3 et du site d'initiation de la transcription dans la région R (Cullen *et al.*, 1985a ; Ruddell, 1995). La région 5' « leader » comporte quant à elle la séquence PBS (Primer Binding Site), la séquence Ψ, et la

séquence L3, qui sont respectivement essentielles aux étapes de transcription inverse, d'encapsidation et de dimérisation de l'ARN génomique (Banks et Linial, 2000 ; Morris et Leis, 1999 ; Polge *et al.*, 2000)(**Figure III-5**).

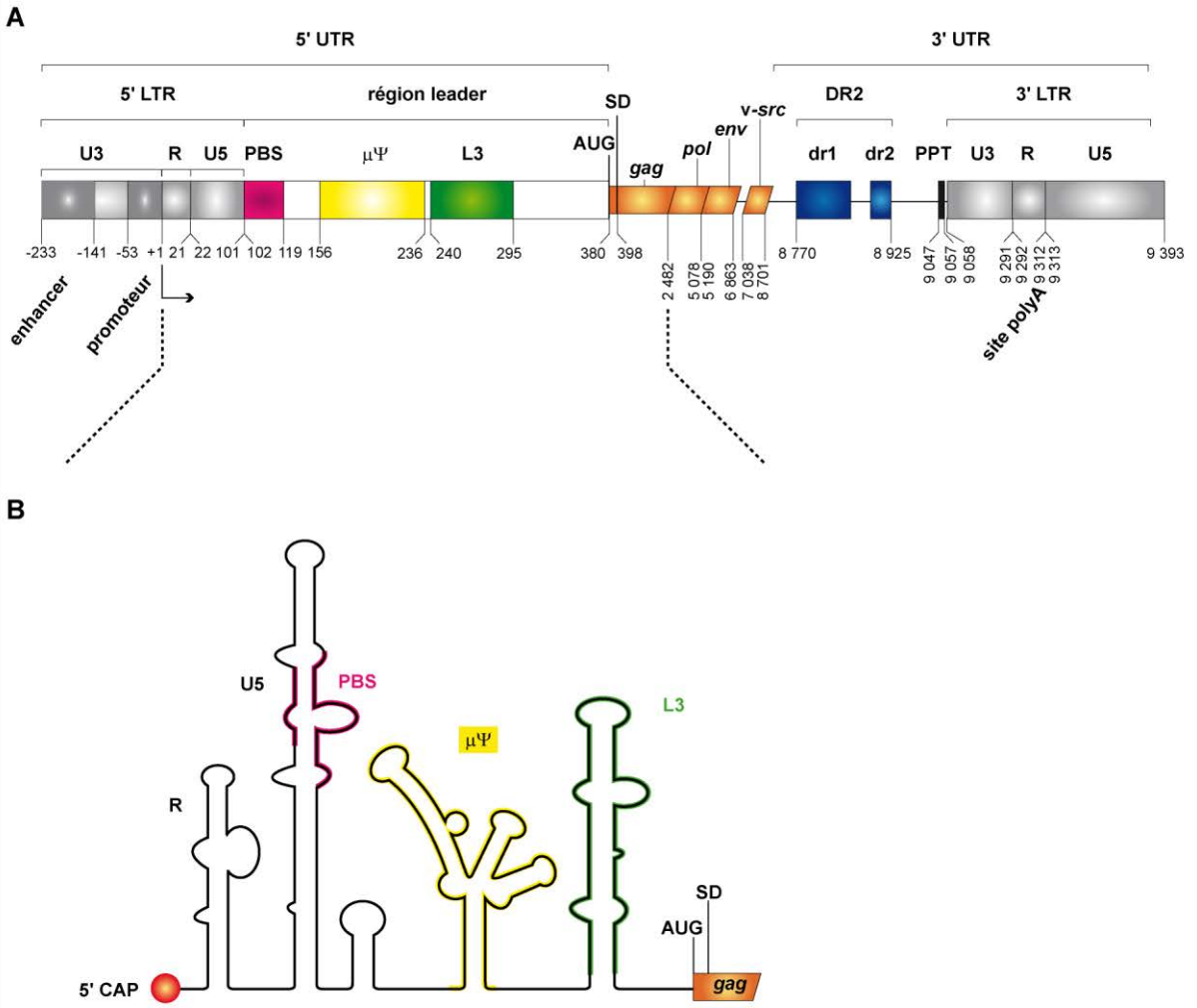



**Figure III-5 : Représentation des régions non traduites de l'ARN du virus RSV.**
**A** : Représentation des régions 5' UTR et 3' UTR au sein de l'ADN proviral. Le nom des séquences et les positions des nucléotides correspondant à leurs extrémités sont respectivement indiqués au-dessus et en-dessous de l'ADN proviral. La transcription de l'ARN génomique débute au niveau du promoteur dans la région 5' LTR et se termine au signal de polyadénylation qui se trouve dans la région 3' LTR. La région 5' UTR comporte également la séquence PBS à laquelle s'apparie l'ARNt Trp cellulaire, lors de l'initiation de la transcription inverse. Les séquences μΨ et L3 servent respectivement à l'encapsidation et à la dimérisation des molécules d'ARN génomiques. Dans la partie 3' UTR, la séquence DR2 sert à l'export des ARN du noyau vers le cytoplasme et la séquence PPT sert d'amorce lors de la transcription inverse.
**B** : Structure secondaire prédite pour la région 5' UTR de l'ARN génomique (Kanevsky *et al.*, 2003 ; Banks et Linial, 2000). Les séquences, indiquées par leur nom, sont reportées sur la structure secondaire selon le même code couleur que celui utilisé en (A).

### VI.1.2. La région 3' non traduite

La séquence comprise entre la fin du gène v-*src* et la séquence 3' LTR comporte la séquence DR2 qui elle-même se compose de deux séquences répétées, dr1 et dr2. Dans le génome du virus RSV, cette séquence est répétée de part et d'autre du gène v-*src*, alors qu'elle n'est présente que dans la partie 3' UTR dans celui des virus ALV. Cette séquence est assimilée à un élément CTE (Constitutive Transport Element) et est impliquée dans l'export du noyau vers le cytoplasme, la stabilité et l'encapsidation des ARN non épissés (Ogert *et al.*, 1996 ; Paca *et al.*, 2000). La région 3' non traduite comprend également la séquence PPT (Poly Purine Track), importante pour l'étape de transcription inverse, ainsi que le signal de polyadénylation dans la partie R de la région 3' LTR (Maciolek et McNally, 2008) (**Figure III-5**). Les parties U3 et R de la région 3' LTR peuvent également servir de promoteur permettant de transcrire les gènes cellulaires qui sont en aval de l'ADN proviral.

## VI.2. Les phases ouvertes de lecture

Le génome viral contient 4 ORF (Open Reading Frame) qui sont traduites selon 2 phases de lecture. Trois de ces ORF, les ORF *gag*, *pol*, *env*, sont communes à tous les rétrovirus et sont essentielles au cycle viral. L'ORF v-*src* est spécifique du virus RSV et code pour un proto-oncogène, la tyrosine kinase v-Src, responsable de la transformation des cellules et de la formation des sarcomes. Cette protéine n'est pas essentielle au cycle viral.

Les ORF *gag* et *pol* sont traduites à partir de l'ARN génomique qui est un ARN non épissé. Les ORF *env* et v-*src* sont en revanche traduites à partir des ARNm env et src, qui sont épissés entre l'unique site 5' et les sites 3' *env* et 3' *src*, respectivement (**Figure III-6**).

Les protéines codées par les ORF *gag*, *pol* et *env* sont tout d'abord traduites sous forme de précurseurs protéiques qui sont ensuite maturés par coupures protéolytiques pour donner les protéines virales fonctionnelles. Selon leur rôle, les protéines sont classées parmi les protéines structurales ou enzymatiques, exceptée la protéine Src qui n'a pas de rôle dans la constitution du virus (**Tableau III-7**).

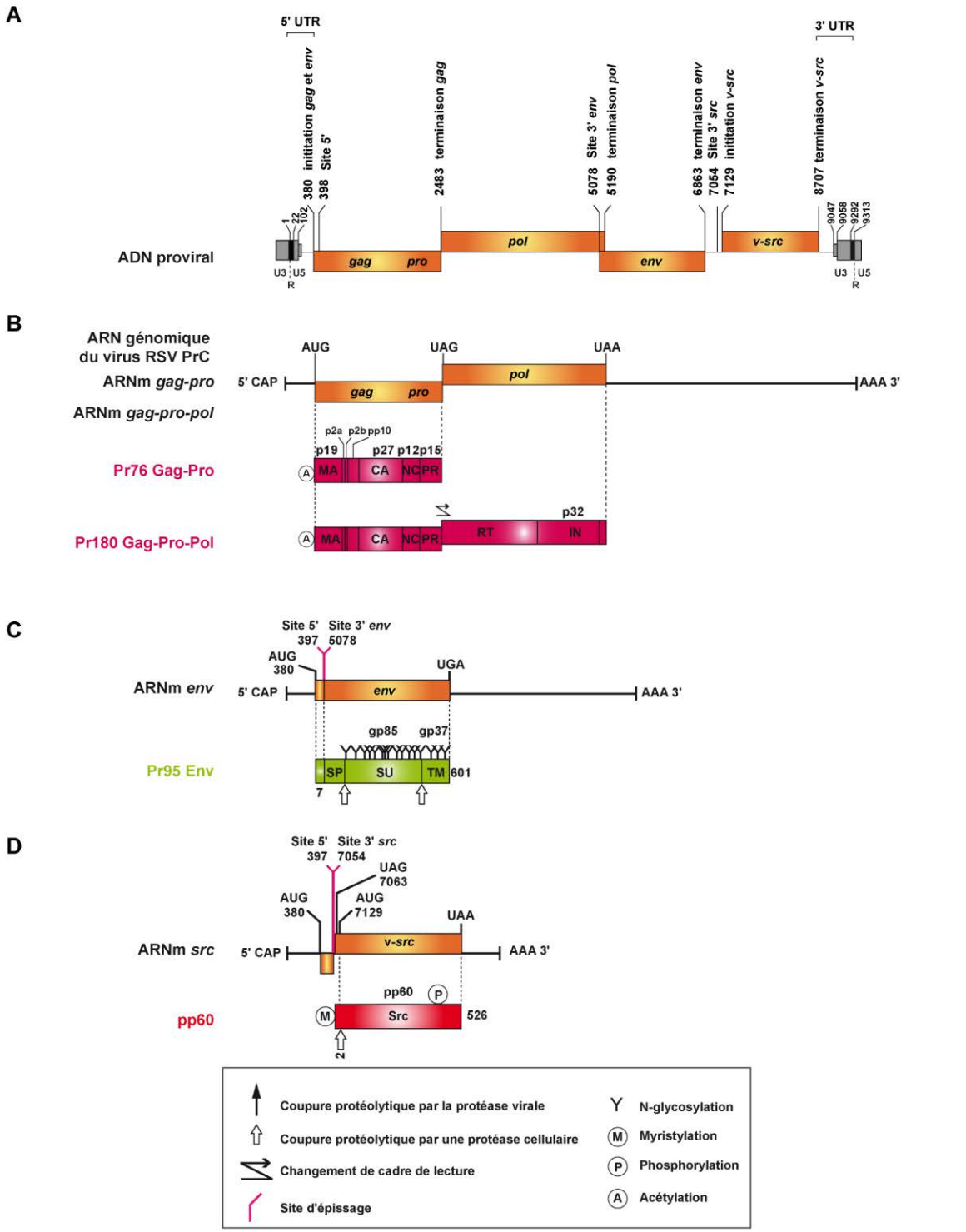



**Figure III-6 : Traduction et maturation des protéines virales du virus RSV (d'après Coffin *et al.* 1997).** L'ADN proviral possède 4 ORF (colorées en orange) codant pour 9 protéines. Les précurseurs Gag-Pro et Gag-Pro-Pol (colorés en rose) sont traduits à partir de l'ARN génomique et sont maturés pour donner les protéines structurales et enzymatiques. Le précurseur Pr95 Env (coloré en vert) est traduit à partir de l'ARNm *env* et sa maturation produit les protéines de l'enveloppe. La protéine Src (colorée en rouge) est traduite à partir de l'ARNm *src*. Les modifications post-traductionnelles des protéines sont illustrées dans la légende.
**A :** Les différentes ORF présentes dans l'ADN proviral du RSV.
**B :** Traduction des ARNm *gag-pro* et *gag-pro-pol.*
**C :** Traduction de l'ARNm *env.*
**D :** Traduction de l'ARNm *src.*

| ORF | nom de la protéine | Fonction |
|---|---|---|
| *gag* | MA | Interaction avec la bicouche lipidique |
| | p2a<br>p2b<br>p10 | protéine du précurseur Pr76 Gag-Pro<br>adressage du précurseur Pr76 Gag-Pro<br>bourgeonnement des virions immatures |
| | CA | constitution de la capside |
| | NC | constitution de la nucléocapside |
| *pol* | PR | protéase<br>Maturation des précurseurs protéiques |
| | RT | transcriptase inverse |
| | IN | Intégration de l'ADN proviral dans le génome de la cellule hôte |
| *env* | SU | protéine externe de l'enveloppe<br>reconnaissance des récepteurs cellulaires |
| | TM | protéine transmembranaire de l'enveloppe<br>point d'ancrage de la protéine SU au niveau de la bicouche lipidique |
| v-*src* | Src | tyrosine kinase<br>induit la transformation de la cellule hôte |

**Tableau III-7 : Les rôles des différentes protéines virales.**

# VII. Cycle de multiplication du virus

## VII.1. La phase pré-intégrative

La phase pré-intégrative comprend l'entrée du virion dans la cellule hôte, la transcription inverse de l'ARN génomique, puis le transport et l'intégration de l'ADN proviral dans le génome de la cellule hôte (**Figure III-8**).

### VII.1.1. L'entrée du virus

L'entrée du virus dans la cellule dépend de la présence dans la membrane cellulaire de récepteurs spécifiquement reconnus par les protéines de l'enveloppe de ce virus. Dans les cellules de poulet, il existe 3 récepteurs qui sont reconnus par différents sous-groupes viraux (**Tableau III-9**) (Fadly, 2008).

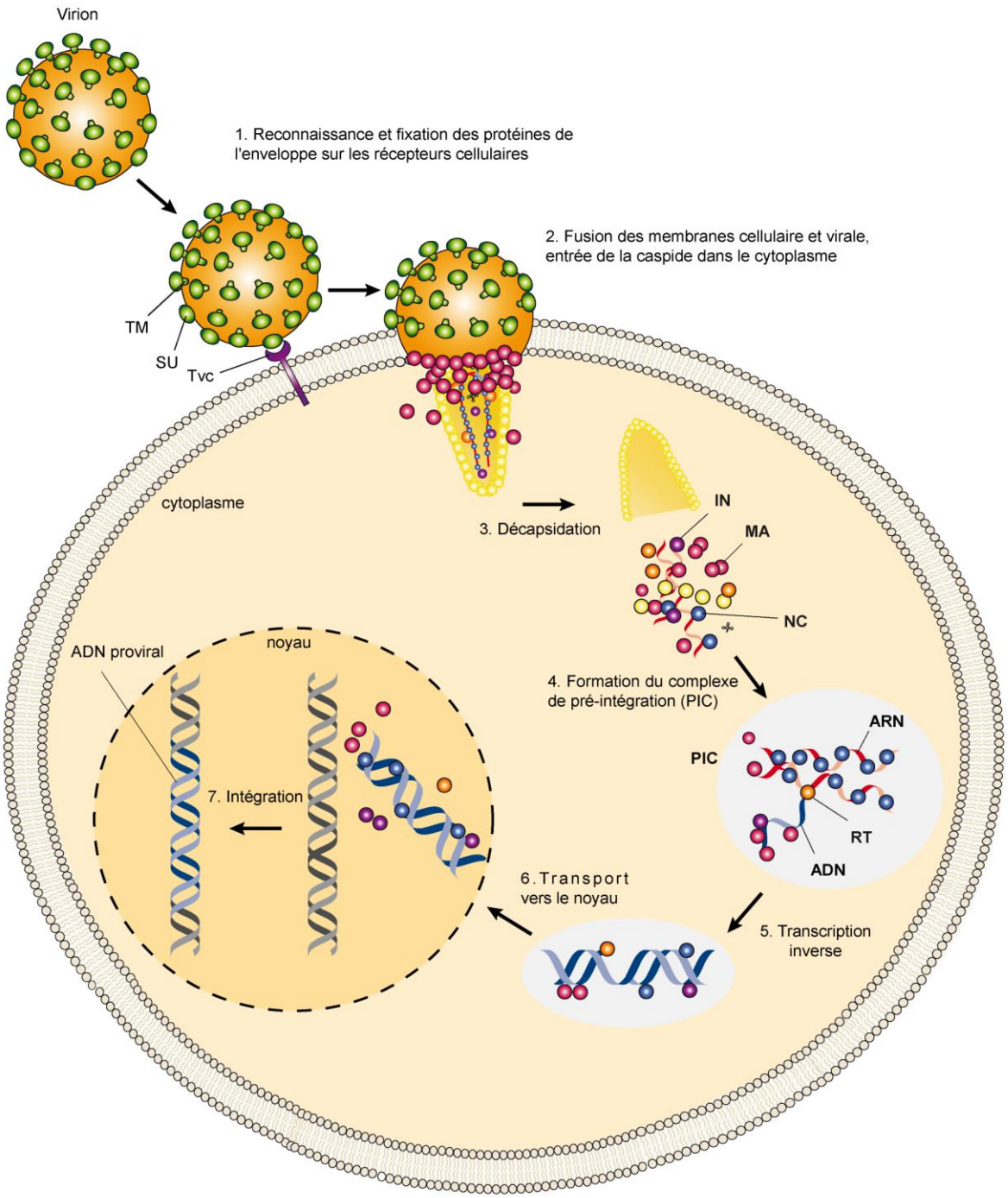



**Figure III-8 : La phase pré-intégrative du cycle viral.**
La phase pré-intégrative comprend 7 étapes. L'ARN génomique et l'ADN viral sont respectivement représentés en rouge et bleu, l'ADN de la cellule hôte est coloré en gris. Le complexe PIC est délimité par une sphère bleue. Le nom des protéines virales est indiqué : TM, protéine transmembranaire de l'enveloppe, SU : protéine de surface de l'enveloppe, RT : transcriptase inverse, MA : Matrice, NC, Nucléocapside, IN : Intégrase.

| Sous-groupe viral | Récepteur cellulaire | Locus | Allèle |
|---|---|---|---|
| A | Récepteur LDL (Low Density Lipoprotein) | TVA | TVA*S, TVA*R |
| B, D, E | Récepteur TNF (Tumor Necrosis Factor) | TVB | TVB*S1, TVB*S3, TVB*R |
| C | Immunoglobuline | TVC | TVC*S, TVC*R |

**Tableau III-9 : Les gènes des récepteurs cellulaires responsables de la susceptibilité génétique des cellules de poulet à l'infection par les virus ASLV (D'après Fadly, 2008).**
Les rétrovirus ASLV infectant les cellules de poulet sont classés en sous-groupes suivant le récepteur cellulaire qu'ils reconnaissent spécifiquement et qui leur permet d'entrer dans la cellule. Dans le génome du poulet, il existe trois récepteurs cellulaires différents codés par 3 *loci* autosomaux. Les récepteurs cellulaires, leur locus et leurs différents allèles spécifiques de chacun des sous-groupes viraux sont présentés.

Les *loci* de chacun des gènes de ces récepteurs ont été appelés *loci* TV (Tumor Virus) et leurs allèles TV*S ou TV*R (Susceptibility et Resistant). Les allèles TV*S codent pour des récepteurs cellulaires fonctionnels qui sont spécifiquement reconnus par un sous-groupe de rétrovirus. Les allèles TV*R en revanche, codent des récepteurs cellulaires non fonctionnels qui ne sont pas reconnus par les rétrovirus, soit parce qu'ils possèdent des mutations, soit parce qu'ils sont tronqués ou non exprimés. Les allèles TV*R sont récessifs, tandis que les allèles TV*S sont dominants. Ainsi, une cellule possédant deux allèles TV*R sera génétiquement résistante à l'infection par un sous-groupe viral donné tandis qu'une cellule possédant au moins un allèle TV*S sera génétiquement susceptible d'être infectée.

Le virus RSV appartient au sous-groupe viral C et reconnaît spécifiquement un récepteur appartenant à la famille des immunoglobulines (Elleder *et al.*, 2005). Ce récepteur est codé par un gène se trouvant au locus TVC. L'allèle TVC*S de ce locus code un récepteur fonctionnel et l'allèle TVC*R code pour un récepteur non fonctionnel en raison de la présence d'un codon stop prématuré (Elleder *et al.*, 2005).

L'infection de la cellule par le virus RSV débute par la reconnaissance spécifique de la protéine SU de l'enveloppe virale avec le récepteur TVC. Ensuite, la protéine TM qui est la protéine transmembranaire de l'enveloppe virale change de conformation et s'ancre dans la membrane de la cellule cible (Barnard *et al.*, 2006) (**Figure III-10**). Les membranes virales et cellulaires se rapprochent alors, grâce au repliement d'une partie de la protéine TM appelée « peptide fusion », composé de deux répétions de séquences structurées en hélice α qui s'assemblent en fagot (Barnard *et al.*, 2006 ; White *et al.*, 2008). S'en suivent les étapes d'hémifusion et de fusion des deux membranes qui aboutissent à la formation d'un pore à travers lequel le contenu du virion passera dans le cytoplasme de la cellule infectée (**Figure III-10**).

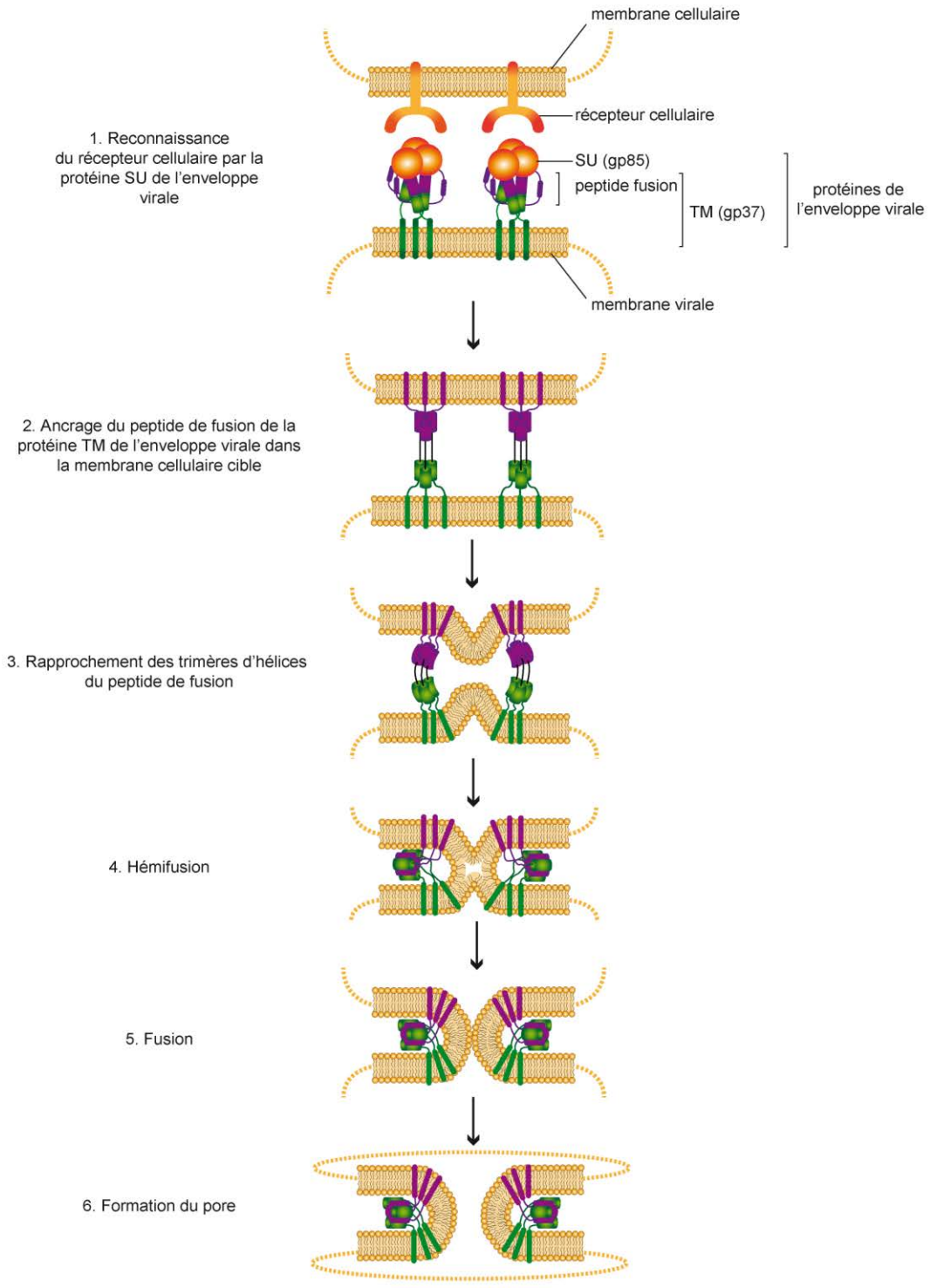



**Figure III-10 : Processus de fusion entre la membrance du virion et la membrane de la cellule hôte (D'après White *et al.*, 2008).**

Les protéines de l'enveloppe virale comprennent la protéine SU, exposée à la surface et la protéine TM, enchassée dans la membrane et sont associées en trimère. La reconnaissance des récepteurs cellulaires par la protéine SU (Etape 1) entraîne un réarrangement d'une partie de la protéine TM appelée "peptide fusion". Ce peptide contient une séquence qui va s'ancrer dans la membrane cellulaire cible (Etape 2) et deux séquences répétées structurées en hélice α (représentées par des cylindres verts et violets) qui vont se rapprocher (Etape 3) et s'assembler en fagot et provoquer l'hémifusion (Etape 4). Les membranes vont ensuite fusionner (Etape 5) et s'éloigner pour former un pore, permettant ainsi au rétrovirus d'introduire le contenu du virion dans le cytoplasme cellulaire (Etape 6).

### VII.1.2. La transcription inverse de l'ARN génomique (Telesnitsky et Goff, 1997)

La transcription inverse est l'étape du cycle viral au cours de laquelle l'ARN génomique est rétro-transcrit en ADN double brin. Cette étape est catalysée par le Transcriptase inverse (RT), une enzyme virale qui possède les fonctions d'ADN polymérase ARN et ADN dépendante et de RNase H, qui lui permet de dégrader les séquences ARN des duplex ADN-ARN.

La transcription inverse débute par l'hybridation de l'ARNt$^{Trp}$ cellulaire à la séquence PBS (primer Binding Site) de l'ARN génomique (**Figure III-11, Etape 1**). L'ARNt sert alors d'amorce pour la synthèse d'un court brin d'ADN (-) complémentaire aux séquences R et U5 de l'extrémité 5' de l'ARN génomique (**Figure III-11, Etape 2**). La séquence d'ARN ayant servi de matrice est ensuite dégradée par l'activité RNase H et la séquence d'ADN néo-synthétisée réalise un saut de brin pour s'apparier à la séquence R à l'extrémité 3' de l'ARN génomique (**Figure III-11, Etapes 2 et 3**). Cet appariement permet de créer une amorce qui permet la synthèse du brin d'ADN (-), qui se poursuit jusqu'à l'extrémité 5' de l'ARN génomique (**Figure III-11, Etape 3**). Suite à cette synthèse, l'activité RNase H dégrade la totalité de l'ARN matrice à l'exception d'une séquence PPT (PolyPurine Tract) qui reste appariée au brin d'ADN (-). Cette séquence sert d'amorce à la synthèse du brin d'ADN (+) jusqu'à l'extrémité 3' de l'ADN (-). Après la dégradation de la séquence PPT et de l'ARNt, le brin d'ADN (+) réalise un saut de brin et sa séquence PBS s'apparie à la séquence complémentaire présente à l'extrémité 5' du brin d'ADN (-) (**Figure III-11, Etapes 6 et 7**). L'élongation de ce brin achève la synthèse de l'ADN viral double brin (**Figure III-11, Etape 8**).

La protéine virale NC joue également un rôle important dans la transcription inverse et favorise le premier saut de brin et l'hybridation du brin d'ADN (-) (**Figure III-11**) (Werner *et al.*, 2001).

### VII.1.3. Importation de l'ADN viral dans le noyau

La transcription inverse a lieu dans le cytoplasme puis l'ADN viral est importé dans le noyau pour y être intégré au génome cellulaire. Cependant, la transcription inverse semble seulement être initiée dans le cytoplasme car la majorité des molécules d'ADN viral qui s'y trouvent ne sont pas double-brin sur toute leur longueur. Il semblerait que cette étape soit achevée dans le noyau après que l'ADN viral y soit importé (Lee et Coffin, 1991).

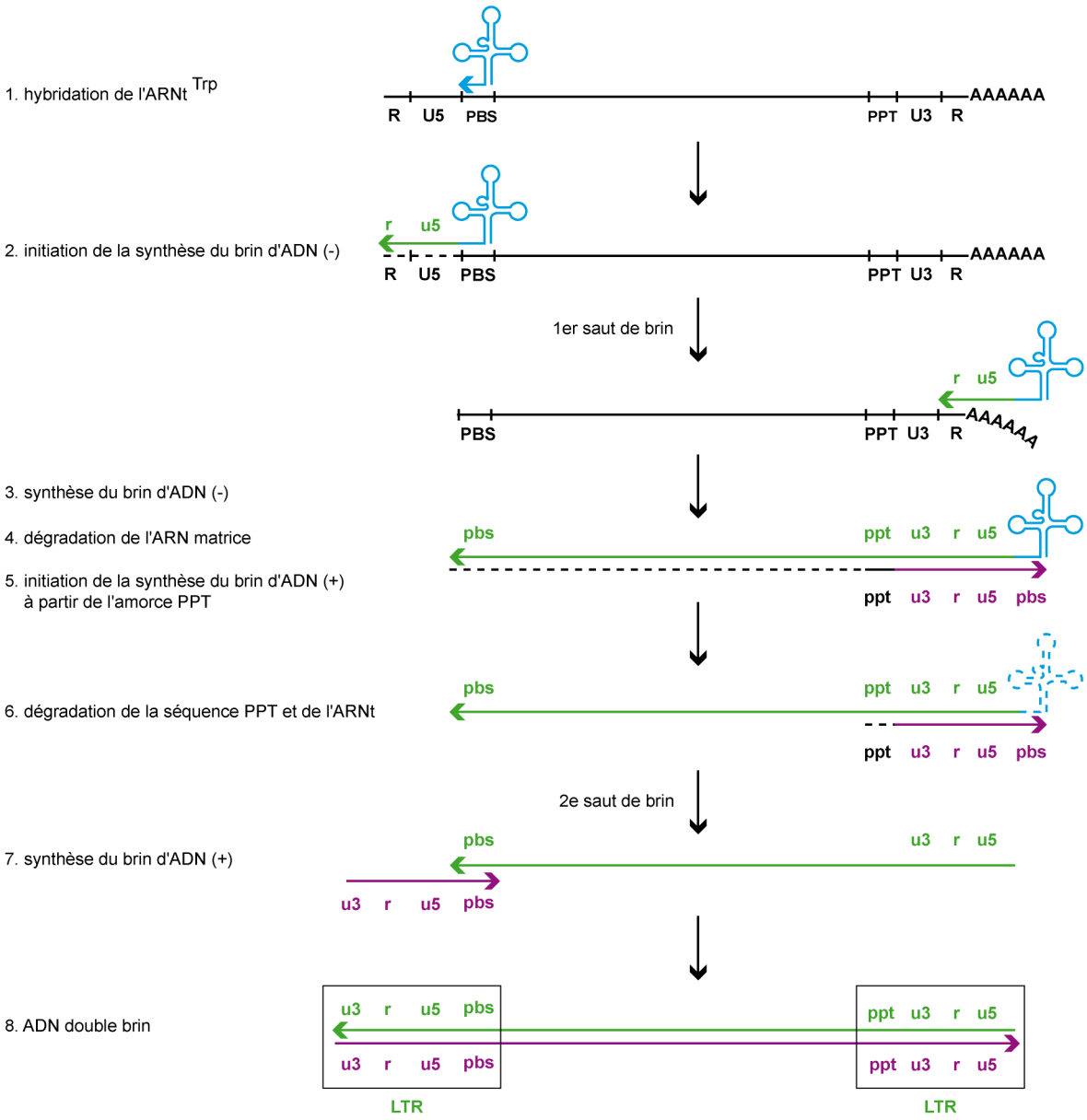



**Figure III-11 : La transcription inverse de l'ARN génomique du RSV (D'après Telesnitsky et Goff, 1997).**

La transcription inverse consiste à synthétiser un ADN proviral double brin à partir de l'ARN génomique. L'ARN génomique est représenté par un trait noir et ses séquences importantes pour la transcription inverse sont indiquées. L'ARNt cellulaire, le brin d'ADN (-) et (+) sont respectivement colorés en bleu, vert et violet. L'ARN dégradé par l'activité RNase H de la transcriptase inverse est représenté en pointillés. Le sens de la synthèse de brins d'ADN est symbolisé par un flèche.

Dans le cytoplasme, la transcription inverse a lieu au sein d'un complexe appelé PIC (PreIntegration Complex). La composition exacte de ce complexe n'a pas été caractérisée mais il est supposé comprendre les protéines virales NC, IN, MA et RT (Butterfield-Gerson *et al.*, 2006 ; Garbitt *et al.*, 2004). Après la transcription inverse, l'ADN viral au sein de ce complexe doit être importé dans le noyau afin de s'intégrer au génome cellulaire. La plupart du temps, il est nécessaire que la cellule soit en division pour que le complexe puisse entrer dans le noyau (Humphries *et al.*, 1981). Cela s'explique par le fait que le complexe PIC possède une taille qui l'empêche de diffuser passivement à travers les pores nucléaires. Néanmoins, le virus RSV peut également infecter des cellules en interphase mais avec une efficacité plus faible (Greger *et al.*, 2004 ; Hatziioannou et Goff, 2001), ce qui laisse penser qu'il existe une voie de transport du complexe PIC vers le noyau n'impliquant pas la déstructuration de l'enveloppe nucléaire. Cette idée est renforcée par la présence de séquences NLS (Nuclear Localization Signal) dans les protéines IN, MA et NC du complexe PIC qui interagissent avec les différentes protéines de la super famille des karyophérines (Andrake *et al.*, 2008 ; Butterfield-Gerson *et al.*, 2006).

### VII.1.4. Intégration de l'ADN proviral dans le génome de la cellule hôte

Après son entrée dans le noyau, l'ADN viral est intégré dans le génome de la cellule hôte par l'intégrase. L'intégration comprend 3 étapes. Lors de la première étape, l'intégrase virale reconnaît les séquences att (attachment site) aux extrémités de l'ADN viral et clive l'ADN au niveau des deux nucléotides adjacents aux dinucléotides très conservés CA de façon à les libérer à chacune des extrémités 3' (**Figure III-12, Etape 1**) (Roth *et al.*, 1989). Dans la seconde étape, l'ADN viral s'apparie à l'ADN chromosomique de la cellule hôte qui a été clivé de façon dissymétrique au niveau du site d'intégration (**Figure III-12, Etape 2**). Dans la troisième étape, les dinucléotides des extrémités 5' sortantes de l'ADN viral sont éliminées et les séquences manquantes de part et d'autre de l'ADN proviral sont synthétisées par les enzymes cellulaires de réparation de l'ADN, ce qui créé une duplication de 6 paires de bases (**Figure III-12, Etape 2**) (Hughes *et al.*, 1981 ; Smith et Daniel, 2006). Les sites d'intégration utilisés ne sont pas définis et leur choix ne semble pas dépendre de leur activité transcriptionnelle (Weidhaas *et al.*, 2000) (Barr *et al.*, 2005). En revanche, dans 85 % des cas de lymphomes induits par un virus ALV, l'ADN proviral est intégré dans le locus du gène cellulaire Myc (Hayward *et al.*, 1981).

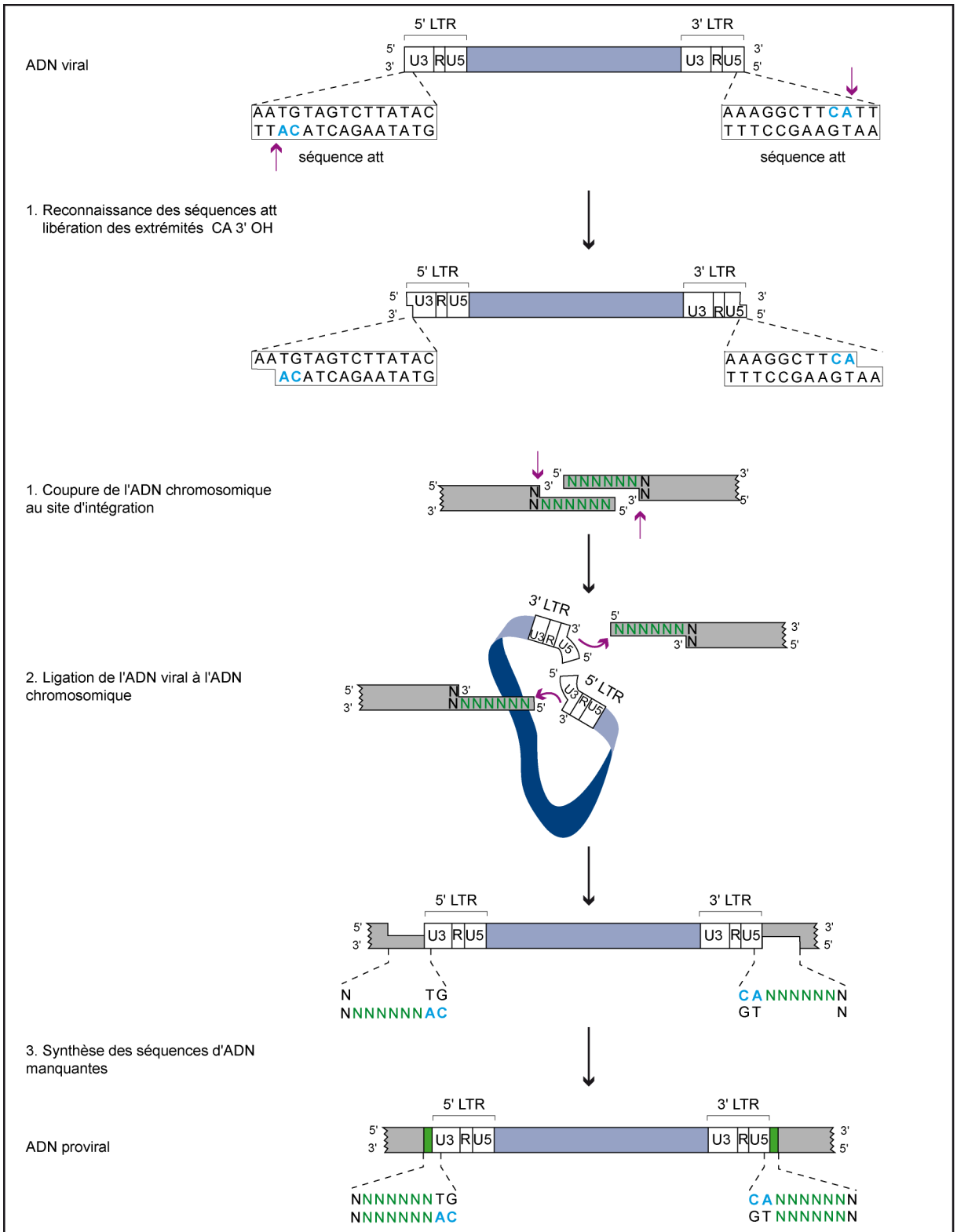



**Figure III-12 : Intégration de l'ADN viral dans le génome de la cellule hôte.**
L'intégration de l'ADN viral comprend 3 étapes. L'ADN viral est représenté en bleu et les séquences LTR en blanc. Les dinucléotides CA qui sont conservés dans les séquences att sont indiqués en bleu. L'ADN chromosomique de la cellule hôte est coloré en gris et le site d'intégration de l'ADN proviral dont la séquence n'est pas définie est représenté par les lettres N vertes. Les étapes catalysées par l'intégrase virale sont représentées par des flèches violettes.

### VII.2. La phase post-intégrative

La phase post-intégrative est la phase dans laquelle les gènes viraux sont exprimés. Après la transcription, une partie des pré-ARN messagers sont épissés en ARN messagers puis l'ensemble des ARN viraux sont exportés vers le cytoplasme (**Figure III-13**). Les ARN viraux sont ensuite traduits pour produire les protéines virales qui s'assemblent avec l'ARN génomique au niveau de la membrane plasmique ce qui aboutit au bourgeonnement du virion (**Figure III-13**). Le clivage des précurseurs protéiques à l'intérieur du virion par la protéase virale induit une réorganisation structurale du virion conduisant à la production de virions matures et infectieux (**Figure III-13**).

#### VII.2.1. La transcription de l'ADN proviral

Le promoteur du virus se situe dans la partie U3 et le site d'initiation de la transcription se situe à la jonction des séquences U3-R du LTR. Le promoteur du virus RSV est un promoteur fort qui contient une séquence « enhancer » située en amont de la séquence du promoteur minimal qui renferme une boîte TATA (**Figure III-14**) (Cullen *et al.*, 1985a). Chacun des LTR possède un promoteur mais celui situé dans le LTR en 5' est majoritairement utilisé. Le promoteur est reconnu par l'ARN polymérase II cellulaire et fonctionne aussi bien dans les cellules aviaires que dans les cellules de mammifères (Gorman *et al.*, 1982). La séquence « enhancer » contient de nombreux sites de fixation auxquels se fixent différents facteurs régulant la transcription cellulaire (**Figure III-14**).

Il existe 3 types de séquences enhancers : EF I, EF II et EF III répertoriées suivant les facteurs de transcription qui s'y fixent (EF : Enhancer Factor). Les séquences EF I également appelées « Y box » ou « CCAAT inverted box », fixent les protéines de la famille Y-Box, les séquences EF II ou séquences CArG fixent les facteurs SRF (Serum Related Factor) et les séquences EF III ou séquences CCAAT/« enhancer element » fixent les facteurs éponymes C/EBP (CCAAT/Enhancer Binding Protein) (**Figure III-14**) (Ruddell, 1995). La transcription est également activée par un autre facteur appelé EFIV qui interagit avec l'ensemble de la séquence « enhancer » (Houtz et Conklin, 1996). Tous ces facteurs interagissent directement avec l'ADN. D'autres protéines activent la transcription en interagissant avec les facteurs déjà fixés à l'ADN, comme c'est le cas pour la protéine RelA de la famille des facteurs de transcription NF-κB, qui interagit avec la protéine VBP (Vitogellin Binding Protein) de la famille des facteurs C/EBP (Bowers *et al.*, 1996).

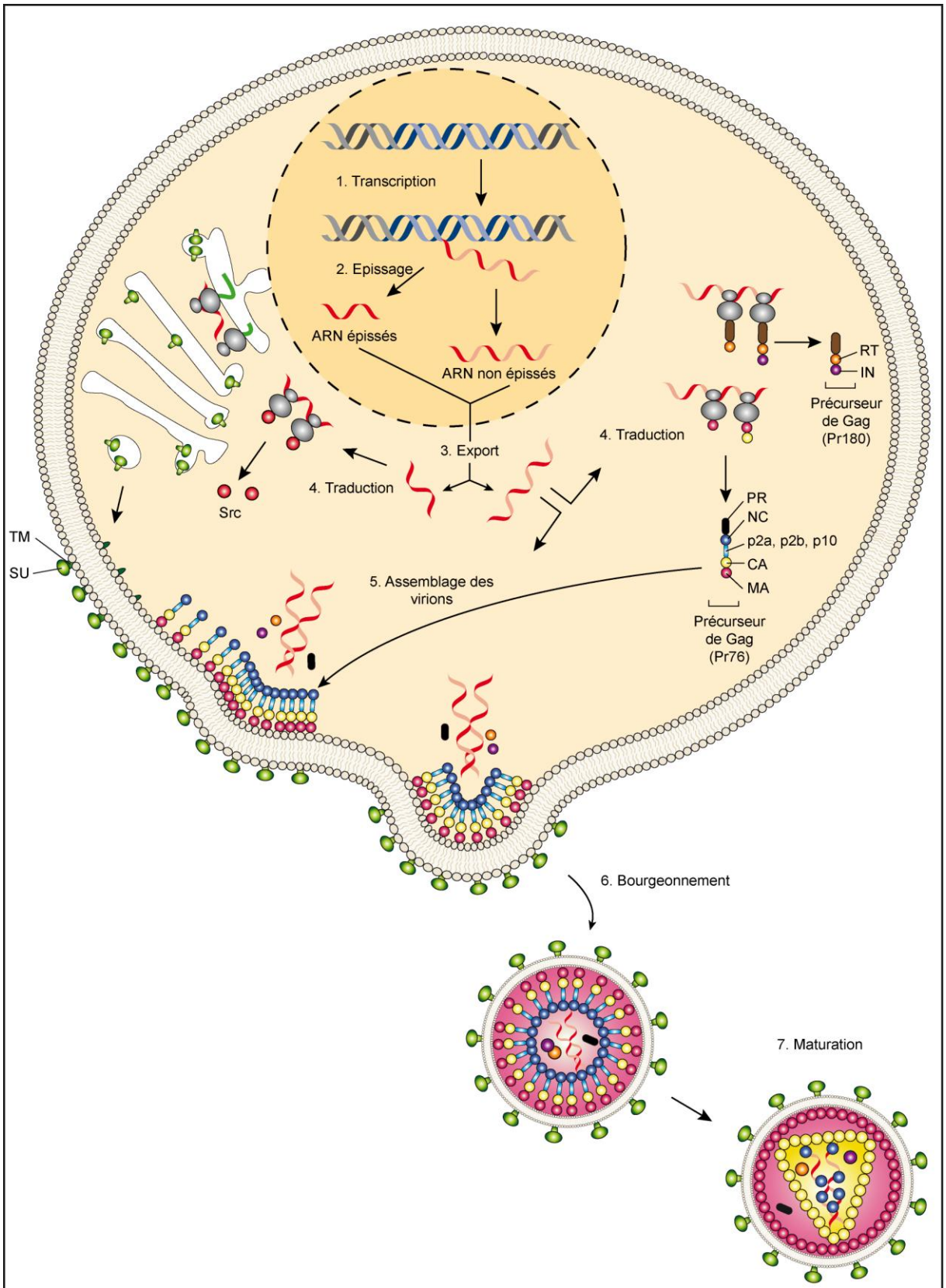



**Figure III-13 : La phase post-intégrative du cycle viral.**
La phase post-intégrative comprend 7 étapes. L'ADN proviral est représenté en bleu, l'ARN viral est coloré en rouge. Le nom des protéines virales est indiqué : TM, protéine transmembranaire de l'enveloppe, SU : protéine de surface de l'enveloppe, RT : transcriptase inverse, MA : matrice, NC, nucléocapside, IN : intégrase, PR : protéase, Src : tyrosine kinase Src.

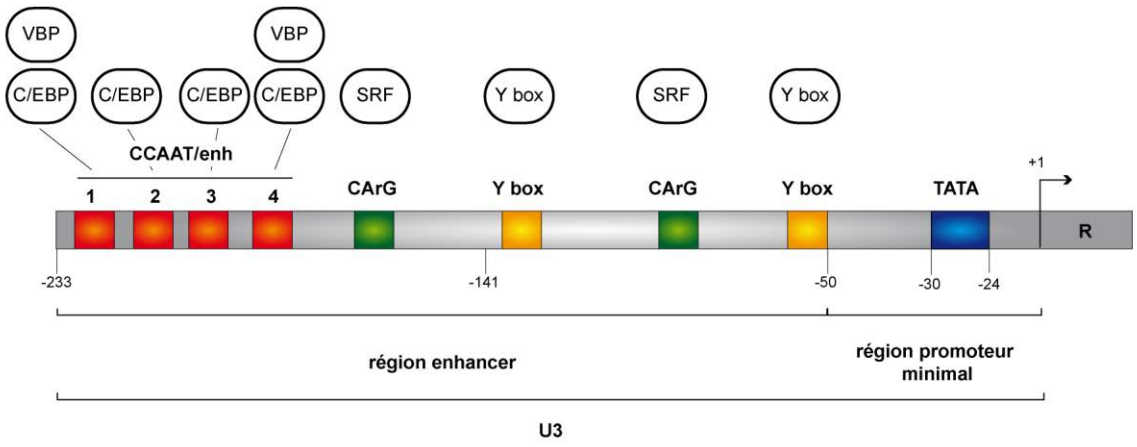



**Figure III-14 : Représentation du promoteur du RSV (d'après Ryden *et al.*, 1993, Ruddell, 1995).**
Le promoteur se situe dans la partie U3 du LTR et se compose d'une partie enhancer et d'une partie promoteur à proprement dit. Les séquences sont numérotées par rapport au site d'initiation de la transcription (+1) qui se situe à la jonction des séquences U3/R. Les séquences enhancer CCAAT/enh, CArG et Ybox sont respectivement colorées en rouge, vert et jaune et les facteurs de transcription s'y fixant sont représentés au-dessus. La TATA box est représentée en bleu.

La force du promoteur dépend de la présence des séquences « enhancer ». La délétion, même partielle, de ces séquences provoque une baisse de 90 % de la transcription (Cullen *et al.*, 1985a, b). Par ailleurs, il existe une corrélation directe entre la force du promoteur viral et la pathogenicité du virus (Robinson *et al.*, 1982). Pour les virus provoquant un néoplasme, les promoteurs forts induisent une transcription importante de l'oncogène viral (exemple : gène *v-src* du RSV) ou du proto-oncogène se situant en aval de l'ADN proviral (exemple : gène *c-myc* ou *c-myb* après intégration du virus ALV LR9) (Hayward *et al.*, 1981 ; Jiang *et al.*, 1997). En revanche, les virus non pathogènes, comme la plupart des virus endogènes (exemple : RAV-0) ont des promoteurs faibles qui sont dépourvus de séquences « enhancer » (Motta *et al.*, 1975).

La force du promoteur dépend également de la concentration des facteurs de transcription dans la cellule infectée. Suivant le type cellulaire, la concentration de ces facteurs varie et influe sur le degré d'activation de la transcription virale. Par exemple, il existe 4 isoformes du facteur VBP mais seule l'isoforme αβ est exprimée dans les cellules lymphoïdes de la bourse de Fabricius qui sont la cible des virus ASLV. Dans ces cellules, la transcription est principalement activée par la fixation de ce facteur sur la troisième séquence CCAAT/enhancer, tandis que cela n'est pas observé dans les cellules CEF (Curristin *et al.*, 1997).

### VII.2.2. L'épissage du pré-ARNm

La transcription de l'ADN proviral produit un pré-ARNm ou transcrit primaire qui commence au site d'initiation de la transcription et se termine au site de polyadénylation. Cet ARN peut être conservé intact ou bien être épissé. Lorsqu'il n'est pas épissé, il sert soit d'ARN génomique pour les futurs virions, soit d'ARNm pour la traduction des précurseurs protéiques Pr76 Gag-Pro et Pr180 Gag-Pro-Pol (**Figure III-15**). L'épissage de cet ARN produit deux ARNm, les ARNm *env* et *src* qui sont le résultat d'un épissage qui a lieu entre le site donneur unique et l'un des deux sites accepteurs alternatifs, 3' *env* et 3' *src* (**Figure III-15**). L'ARNm *env* sert à la traduction des protéines de l'enveloppe, TM et SU, et l'ARNm *src* sert à la traduction de la protéine v-Src (**Figure III-15**).

Les ARNm *env* et *src* représentent respectivement 13 et 12 % de l'ARN viral total, ce qui signifie qu'environ 25 % des ARN sont épissés et que l'ARN génomique représente environ 75 % de l'ARN viral total (Berberich et Stoltzfus, 1991). Le maintien de ces pourcentages entre ARN épissés et non épissés est primordial pour la multiplication du virus. Nous décrirons ultérieurement comment la régulation de l'épissage alternatif du virus RSV permet le maintien de cet équilibre.

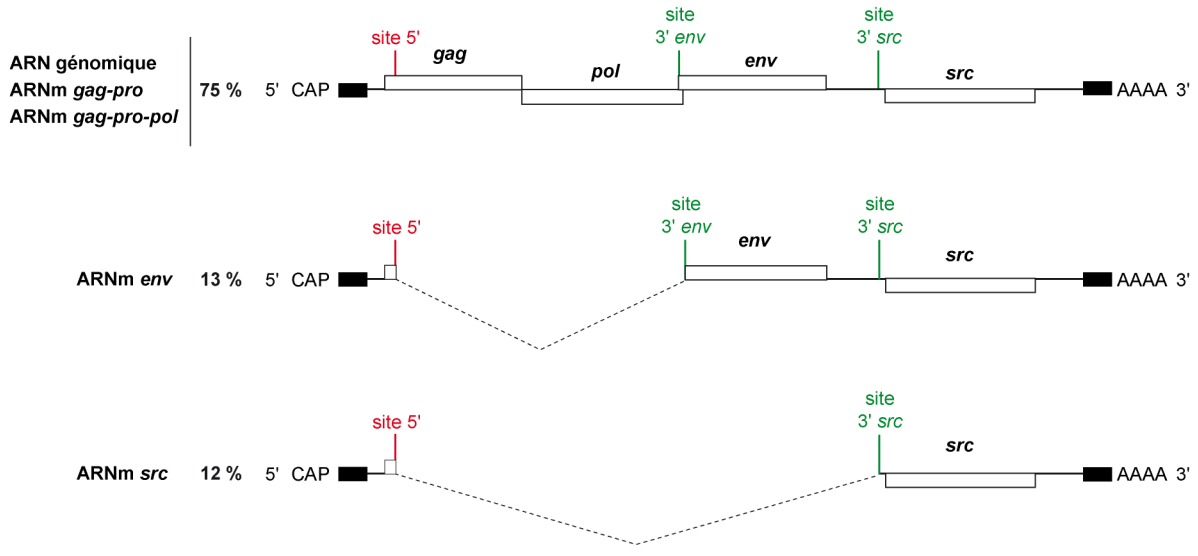



**Figure III-15 : Représentation de l'ARN génomique et des différents ARNm du virus RSV.**
L'ARN génomique est issu de la transcription de l'ADN proviral et n'est pas épissé. Cet ARN porte une coiffe et est polyadénylé et représente environ 75 % de l'ARN viral total. L'épissage de cet ARN produit 2 ARNm, l'ARNm *env* et l'ARNm *src*. Les ARNm *env* et *src* sont le résultat de l'épissage qui a lieu entre le site donneur 5' et les sites accepteurs en compétition, 3' *env* et 3' *src*, respectivement. Les ARNm *env* et *src* représentent respectivement 13 et 12 % de l'ARN viral total.
Le nom des gènes *gag*, *pol*, *env* et *src* sont reportés au-dessus des rectangles blancs qui les représentent. Les sites donneur et accepteurs d'épissage sont respectivement colorés en rouge et en vert. Le nom des ARNm, ainsi leur taux exprimés en pourcentage d'ARN total sont mentionnés. L'épissage entre le site donneur et les sites accepteurs est représenté par les lignes en pointillés.

### VII.2.3. La polyadénylation

Le processus de polyadénylation est couplé au processus de transcription et d'épissage et consiste à ajouter une queue polyA à l'extrémité 3' du transcrit primaire. Cette réaction comprend deux étapes qui sont le clivage endonucléolytique et la synthèse de la queue polyA. Pour que cette réaction ait lieu, il est nécessaire que le pré-ARNm possède des séquences *cis* qui soient reconnues par les complexes protéiques cellulaires qui catalysent la réaction.

Trois séquences agissant en *cis* sont essentielles : le signal de polyadénylation, le site de clivage et une séquence DSE (DownStream Element) (Mandel *et al.*, 2008)(**Figure III-16, A**). Le signal de polyadénylation, AAUAAA, est une séquence très conservée qui se situe dans la région U3 de chacun des LTR. Seul le signal de polyadénylation situé dans la partie 3'UTR est utilisé pour la polyadénylation, car celui situé dans la partie 5'UTR est localisé en amont du site d'initiation de la transcription de l'ARN viral (Guntaka, 1993). Le site de clivage se situe à la position 9312 au sein de l'ARN viral, soit 23 nucléotides en aval du signal de polyadénylation. Il se compose du dinucléotide conservé CA en 3' duquel a lieu le clivage. La séquence DSE est une séquence moyennement riche en G/U qui se trouve 30 nucléotides en aval du site de polyadénylation (Maciolek et McNally, 2008). Le virus RSV possède également une séquence USE (UpStream Element) qui fonctionnellement moins importante que les séquences citées précédemment et qui recrute des facteurs auxiliaires permettant d'ancrer la machinerie de polyadénylation sur l'ARN. Cette séquence se situe en amont du signal de polyadénylation (Maciolek et McNally, 2008).

Lors de la polyadénylation, le signal de polyadénylation est reconnu par le complexe multi-protéique CPSF (Cleavage and Polyadenylation Specificity Factor), qui se compose de 5 protéines (CPSF 160, 100, 73, 30 et hFip1). Ce complexe se fixe de façon coopérative avec le complexe hétérotrimérique CstF (Cleavage stimulation Factor) composé des protéines CstF 50, 77 et 64-kDa qui interagit avec l'élément DSE (**Figure III-, B**). Après la fixation des ces complexes, la protéine CSPF 73 kDa clive l'ARN au niveau du signal de polyadénylation, puis la protéine PAP (Poly(A)Polymérase) ajoute environ 250 nucléotides A en 3' de l'ARN.

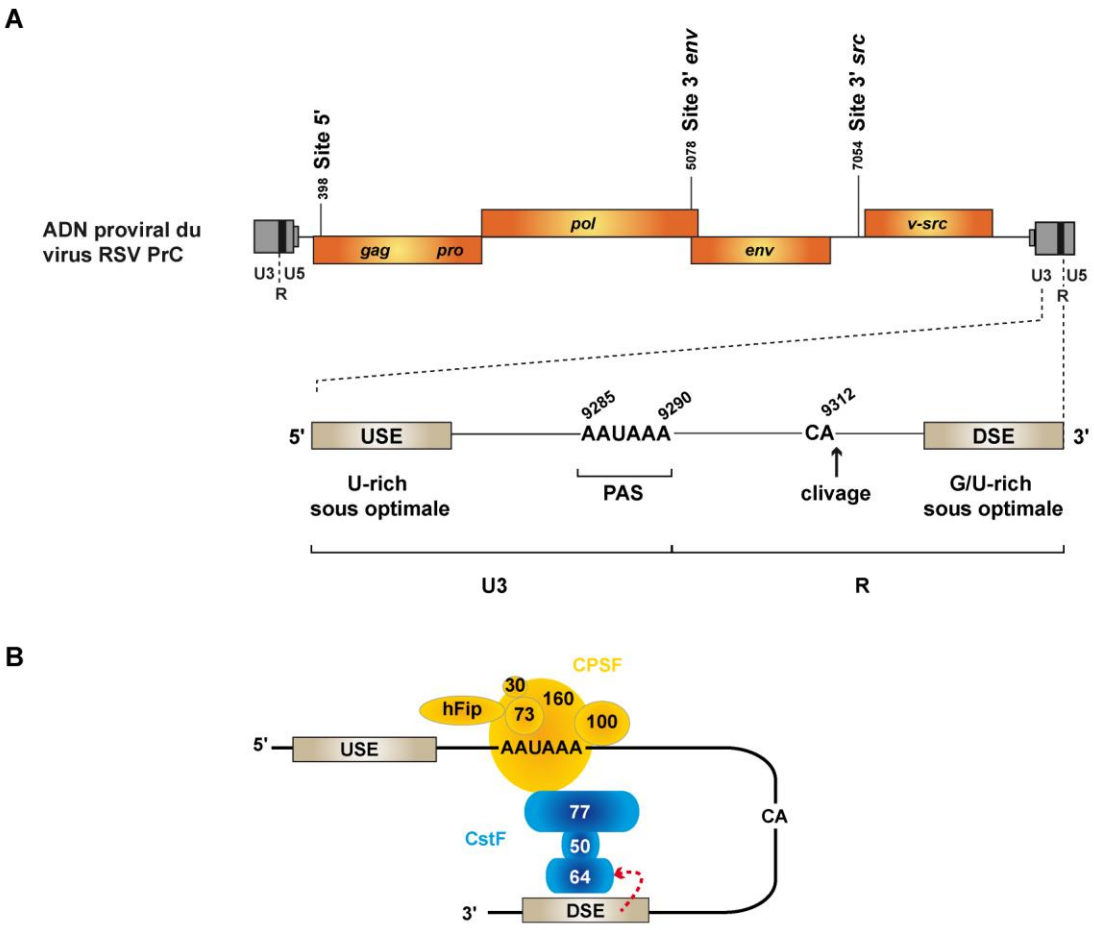



**Figure III-16 : Représentation des séquences importantes pour la polyadénylation des transcrits du virus RSV (d'après McNally, 2008).**
**A :** Représentation des séquences virales impliquées dans le processus de polyadénylation des ARN viraux au sein de l'ADN proviral. Le signal de polyadénylation (PAS), le site de clivage et les séquences USE et DSE (Upstream et Downstream Sequence Element) sont indiqués.
**B :** Assemblage des complexes protéiques lors de la reconnaissance du signal de polyadénylation. L'étape limitante de la polyadénylation se situe au niveau du recrutement de la protéine CstF-64kDa sur la séquence DSE qui est sous-optimale. Cette étape est représentée par une flèche rouge en pointillés.

La polyadénylation des transcrits du RSV n'est pas efficace car dans 15 % des cas, le site de polyadénylation utilisé n'est pas celui du virus mais celui qui se trouve dans le gène du génome de la cellule hôte en aval du site d'insertion de l'ADN proviral (Herman et Coffin, 1986 ; Miller et Stoltzfus, 1992). Dans les processus de tumorigenèse, cela provoque la transcription du proto-oncogène situé en aval de l'ADN proviral, comme c'est le cas pour le virus ALV LR-9 qui s'insère dans l'intron 1 du gène c-*myb* (Polony *et al.*, 2003 ; Smith *et al.*, 1997).

La faible utilisation du site de polyadénylation du virus est due au caractère sous-optimal des séquences USE et DSE (Maciolek et McNally, 2008). La séquence USE ne possède pas les séquences de fixation consensus pour les facteurs qui stabilisent la fixation du complexe

CPSF, et la séquence DSE n'est pas suffisamment riche en G/U pour permettre la bonne fixation de la protéine CstF 73 (Maciolek et McNally, 2008).

Néanmoins, la polyadénylation est activée par d'autres séquences virales qui sont utilisées de concert lors de l'épissage qui sont la séquence NRS (Negative Regulator of Splicing) et le site 3' *env* et dont nous aborderons ultérieurement en détails le mode d'action.

### VII.2.4. Export des ARN non épissés vers le cytoplasme

Dans les cellules des eucaryotes supérieurs, les pré-ARNm contenant des introns sont maintenus dans le noyau jusqu'à ce qu'ils soient épissés en ARNm, puis sont exportés vers le cytoplasme (Erkmann et Kutay, 2004). Contrairement aux ARNm cellulaires, les ARN viraux qui sont exportés vers le cytoplasme doivent être majoritairement non épissés afin que le virus puisse se multiplier. Chez les rétrovirus complexes comme le VIH-1, cette fonction est assurée par une protéine régulatrice, la protéine Rev. Celle-ci se fixe spécifiquement sur une séquence *cis* virale, la séquence RRE (Rev Responsive Element) du pré-ARNm ou des ARNm mono-épissés et interagit ensuite avec la protéine cargo Crm1 qui se charge de l'export du complexe à travers le pore nucléaire (Hope, 1999).

Le virus RSV est un rétrovirus simple qui ne code pas de protéine virale fonctionnellement équivalente à la protéine Rev et dépend donc des facteurs cellulaires impliqués dans les voies d'export cellulaires. L'export des ARN non épissés est assuré par des séquences *cis* virales appelées CTE (Constitutive Transport Element). Chez le RSV, il existe deux éléments CTE qui sont chacun de part et d'autre du gène *src* et qui correspondent à des séquences répétées appelées DR (Direct Repeat) (Ogert *et al.*, 1996). Chacune des séquences DR est composée d'une séquence répétée dr1 d'une centaine de nucléotides et d'une seconde séquence répétée appelée dr2 de 36 nucléotides (**Figure III-17, B**).

La première séquence DR, appelée UDR (Upstream Direct Repeat) se situe entre les gènes *env* et *src*, tandis que la seconde, appelée DDR (Downstream Direct Repeat) se situe en aval du gène *src* dans la partie 3' UTR (Ogert *et al.*, 1996). Il est supposé que les deux séquences DR soient le produit de la duplication de la séquence virale initialement présente en aval du gène *env* lors de la recombinaison qui a permis l'acquisition de l'oncogène v-*src* par le virus RSV (Sugden, 1993). Les autres virus ASLV ne possèdent pas l'oncogène v-*src* et ne possèdent donc qu'un seul exemplaire de la séquence DR qui se situe en aval du gène *env* dans la partie 3' UTR, où elle y remplit la même fonction d'élément CTE (Yang et Cullen, 1999)(**Figure III-17, A**). Ainsi, que ce soit chez les virus RSV ou ALV, la ou les séquences DR sont présentes dans tous les ARN viraux qu'ils soient épissés ou non.

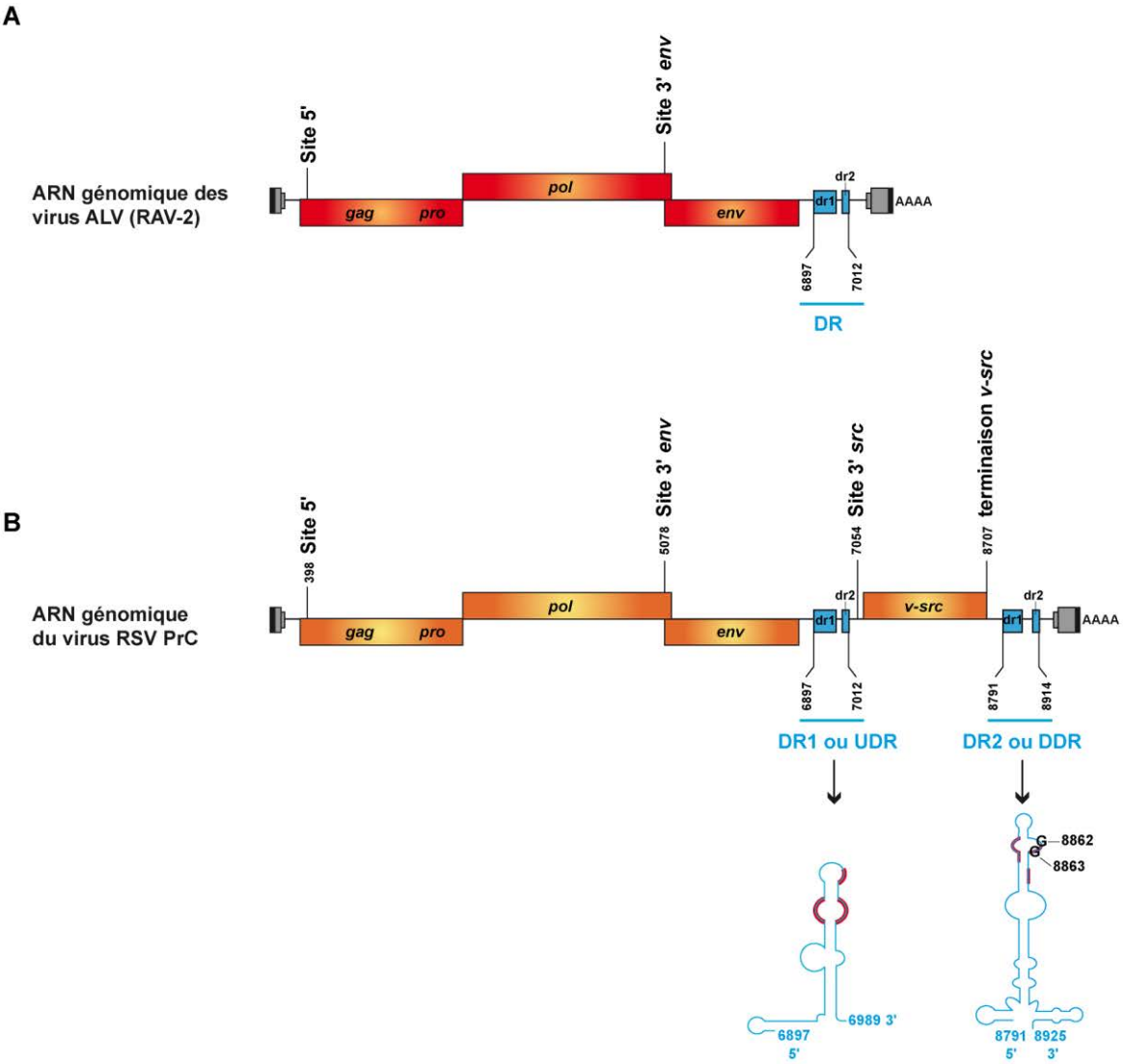



**Figure III-17 : Représentation des séquences DR dans l'ARN génomique des virus ALV et RSV (d'après Yang et Cullen, 1999 ; Ogert *et al.*, 1996).**
Les élément CTE sont localisés dans les séquences DR des rétrovirus aviaires. Chaque séquence DR est composée d'une séquence dr1 et dr2 qui sont représentées par des rectangles bleus.
**A :** Les virus ALV comme le virus RAV-2, ne possèdent qu'un seul exemplaire de la séquence DR en aval du gène *env* dans la partie 3' UTR.
**B :** Les virus RSV (ici, la souche PrC) possèdent deux séquences DR (DR1 et DR2) de part et d'autre du gène v-*src*. Les structures secondaires des séquences DR1 et DR2 sont représentées. Les séquences fonctionnellement importantes sont colorées en rouge. Les mutations ponctuelles qui inhibent totalement les fonctions CTE de ces éléments sont égalements mentionnées.

En plus de leur fonction dans l'export des ARN non épissés du noyau vers le cytoplasme, les éléments CTE pourraient être impliqués dans la stabilité des ARN non épissés (Ogert et Beemon, 1998 ; Ogert *et al.*, 1996 ; Simpson *et al.*, 1998), l'assemblage de Gag (Simpson *et al.*, 1998 ; Simpson *et al.*, 1997), et l'empaquetage de l'ARN génomique (Aschoff *et al.*, 1999 ; Sorge *et al.*, 1983). Parmi toutes ces fonctions, seul leur rôle dans l'export a été clairement caractérisé (LeBlanc *et al.*, 2007).

La troncation des éléments CTE provoque un défaut d'export de l'ARN non épissé vers le cytoplasme et donc un retard dans la réplication virale. Il n'est pas encore clairement établi si le défaut d'assemblage de la protéine Gag, la diminution de la stabilité et de l'empaquetage des ARN non épissés sont des conséquences directes ou indirectes de la troncation des éléments CTE (LeBlanc *et al.*, 2007).

Le virus RSV possède deux éléments CTE mais une seule copie est suffisante pour assurer l'export des ARN (Ogert *et al.*, 1996 ; Simpson *et al.*, 1998). Dans la souche RSV PrC, les deux séquences DR partagent 80 % d'homologie de séquence et sont fonctionnellement équivalentes (Ogert *et al.*, 1996). En revanche, dans la souche PrA, bien que les séquences DR aient chacune des activités CTE, seule la séquence DDR permet un assemblage de la protéine Gag conduisant à la formation des virions (Simpson *et al.*, 1998). Les éléments DDR de la souches PrA et PrC diffèrent par une substitution de base U en C au nucléotide 6957 (Stoltzfus *et al.*, 1987). La génération de cette mutation dans la séquence DDR de la souche PrA restaure l'assemblage de la protéine Gag et le relargage des virions (Simpson *et al.*, 1998).

Les séquences DR sont relativement bien conservées parmi les virus ASLV et la mutation des séquences conservées engendre dans la majorité des cas une perte de l'activité CTE (Aschoff *et al.*, 1999 ; Ogert et Beemon, 1998). Par ailleurs, la détermination de la structure secondaire des éléments CTE a montré que les régions fonctionnellement importantes de ces éléments se situent dans la partie apicale de la tige boucle (Aschoff *et al.*, 1999 ; Banks *et al.*, 1997 ; Ogert et Beemon, 1998 ; Yang et Cullen, 1999) (**Figure III-17, B**). Les mutations G8863C et G8862C abolissent totalement la fonction CTE et provoquent l'accumulation de l'ARN non épissé dans le noyau (LeBlanc *et al.*, 2007 ; Ogert *et al.*, 1996) (**Figure III-17, B**).

La voie d'export empruntée par les ARN non épissés du virus RSV grâce aux éléments CTE n'est pas encore totalement caractérisée. Cet export n'utilise pas la voie Crm1 mais une voie impliquant les facteurs Tap et Dbp5 (LeBlanc *et al.*, 2007 ; Paca *et al.*, 2000).

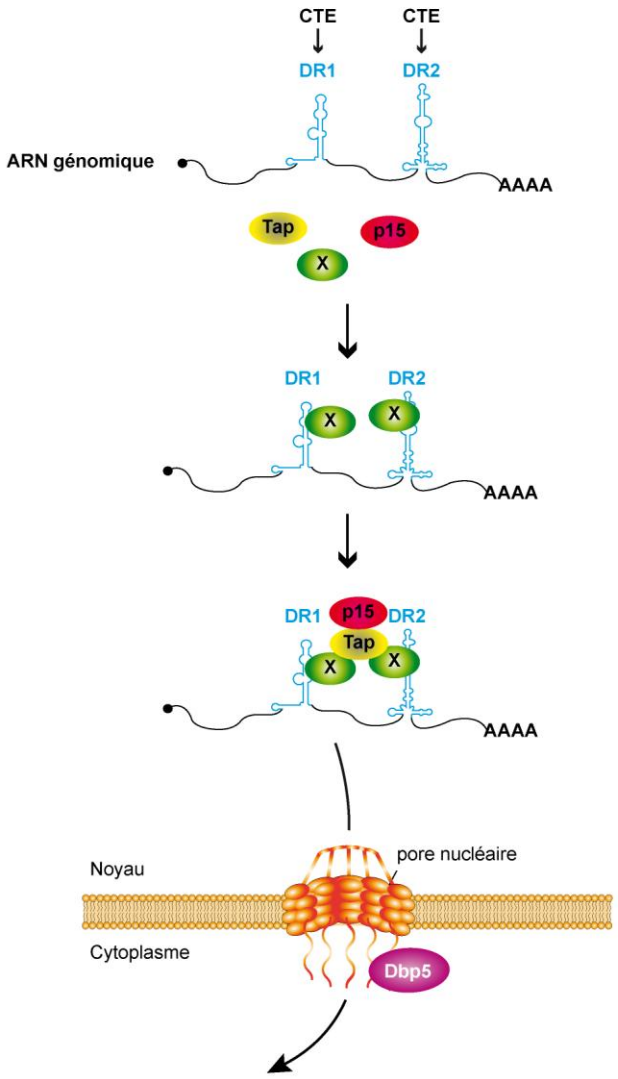



**Figure III-18 : Modèle hypothétique d'export des ARN viraux non épissés par les éléments CTE localisés dans les séquences DR du virus RSV (d'après LeBlanc *et al.* 2007).**
L'ARN génomique est représenté par un trait noir et les structures secondaires des séquences DR y sont représentées en bleu. Les séquences DR recrutent une protéine adaptatrice "X" qui, à son tour, recrute les protéines assurant l'export, Tap et son co-facteur p15. Le complexe ainsi formé est acheminé jusqu'au pore nucléaire où son export est pris en charge par la protéine Dbp5.

En revanche, aucune interaction entre la protéine Tap et les éléments CTE n'a pu être observée (Paca *et al.*, 2000). Il est supposé que les éléments CTE interagissent avec une protéine adaptatrice qui recrute ensuite la protéine Tap et son co-facteur p15 (LeBlanc *et al.*, 2007)(**Figure III-18**). L'identité de cette protéine reste à déterminer mais il est supposé qu'elle est spécifiquement présente dans les cellules aviaires, expliquant ainsi pourquoi l'export des ARN non épissés du virus RSV ne peut pas être observée dans les cellules de mammifères (Ogert et Beemon, 1998 ; Ogert *et al.*, 1996).

Hormis leur fonction dans l'export des ARN non épissés, les séquences DR participent également à la régulation de l'épissage du gène v-*src*. Ce rôle sera abordé plus en détail ultérieurement.

Une fois exportés dans le cytoplasme, les ARN viraux servent de matrice pour la traduction des protéines virales par les ribosomes.

### VII.2.5. Traduction des ARN viraux

La synthèse des protéines virales est réalisée par la machinerie de traduction de la cellule hôte. Le génome du virus RSV comporte 4 ORF, les ORF *gag-pro*, *gag-pro-pol*, *env* et *src* qui sont traduites selon deux phases de lecture (**Figure III-19, A**). Le premier codon initiateur AUG se situe dans l'ORF *gag* en amont de l'unique site donneur d'épissage. Par conséquent, le pré-ARNm et les ARNm *env* et *src* possèdent tous la même région 5'UTR qui régule l'initiation de leur traduction à un même codon d'initiation.

La région 5'UTR possède de nombreuses séquences fonctionnelles (PBS, Ψ), ainsi que 3 uORF (upstream Open Reading Frame) (Hackett *et al.*, 1991) (**Figure III-19**). Les uORF 1 et 3 régulent l'initiation de la traduction au niveau de l'AUG de l'ORF de *gag* et l'encapsidation des ARN génomiques (Donze *et al.*, 1995 ; Donze et Spahr, 1992). Selon le modèle de régulation proposé, le ribosome se fixerait sur l'uORF 1, puis se décrocherait en raison de la forte structure secondaire en aval. La traduction serait alors réinitiée au niveau de l'AUG de l'ORF *gag* par un mécanisme IRES dépendant (Deffaud et Darlix, 2000) (**Figure III-19**). En l'absence de traduction, la séquence μΨ rendue accessible pourrait alors interagir avec le précurseur Pr76 Gag-Pro, afin de permettre l'encapsidation de l'ARN viral (Sonstegard et Hackett, 1996 ; Zhou *et al.*, 2005).

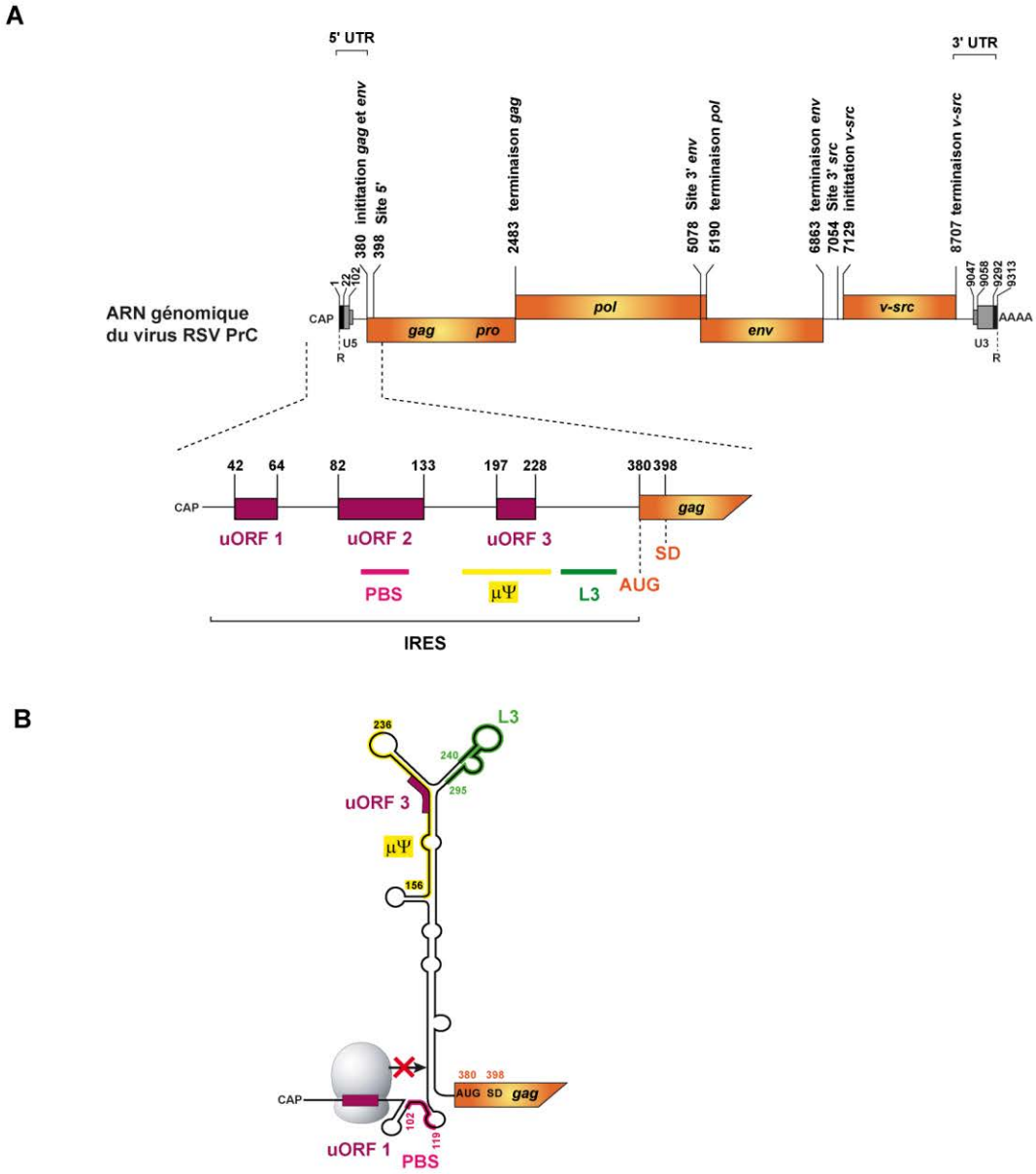



**Figure III-19 : Représentation de la séquence IRES et des séquences régulant l'initiation de la traduction des ORF du virus RSV (d'après Deffaud et Darlix 2000).**

**A :** Les 4 ORF (rectangles oranges), les sites d'épissages et les sites d'initiation et de terminaison de la traduction sont représentés sur l'ARN génomique. Les 2 cadres de lecture sont symbolisés par le décalage vertical des rectangles. Les séquences fonctionnelles de la séquence 5' UTR et le début de l'ORF *gag* sont détaillées en dessous de l'ARN génomique. Les uORF sont représentées par des rectangles violets, les autres séquences sont symbolisées par des traits de couleurs. PBS : Primer Binding Site, mY : signal d'encapsidation, L3 : séquence de dimérisation de l'ARN génomique, SD : site donneur.

**B :** Structure prédictive de la séquence IRES lorsque le ribosome est fixé sur l'uORF 1. La forte structuration de la séquence 5' UTR empêche le ribosome d'avancer plus en aval sur l'ARN. Les séquences fonctionnelles détaillées en A sont reportées sur la structure 2D de la séquence IRES.

Une fois initiée à l'AUG de l'ORF de *gag*, la traduction permet de produire les protéines virales. L'ARN génomique est un ARN non épissé qui contient les ORF *gag-pro* et *gag-pro-pol* et dont la traduction produit respectivement les précurseurs Pr76 Gag-Pro et Pr180 Gag-Pro-Pol (**Figure III-20, B**). Ces 2 ORF ne correspondent pas la même phase de lecture et après la traduction de l'ORF *gag-pro*, un glissement du ribosome est nécessaire pour passer dans la seconde phase de lecture pour traduire l'ORF *pol* (Swanstrom et Wills, 1997).

La traduction de l'ARNm *env* produit le précurseur Pr95Env qui code les protéines de l'enveloppe (**Figure III-20, C**). L'ORF *env* est traduite dans la même phase de lecture que celle de l'ORF *gag* et sa traduction est initiée au même codon AUG. Les 6 premiers acides aminés du précurseur Pr95Env sont codés par la séquence de l'ORF *gag* entre les nucléotides 380 et 397 qui se situent entre le site d'initiation de la traduction et le site donneur.

La traduction de l'ARNm *src* produit la protéine v-Src (**Figure III-20, D**). Cette protéine est codée dans une phase de lecture différente de celle de l'ORF *gag*. La traduction de cet ARNm produit tout d'abord un peptide de 9 acides aminés qui correspond à la traduction de la séquence comprise entre le codon initiateur de l'ORF *gag* et le codon stop en phase qui est dans l'ORF de *src* (Hughes *et al.*, 1984). La traduction est réinitiée au niveau du codon initiateur de l'ORF *src* (Deffaud et Darlix, 2000).

### VII.2.6. NMD et traduction des ARN non épissés

Le système NMD est un processus de contrôle-qualité qui a lieu lors de la traduction et qui reconnaît et provoque la dégradation des ARN contenant des codons stops prématurés (Chang *et al.*, 2007). Comme les ARN cellulaires épissés, les ARN viraux épissés possèdent les marques des complexes EJC déposés 20 à 24 nucléotides en amont des jonctions exon-exon qui permettent au système NMD de distinguer les codons stop naturels des codons stops prématurés (Chang *et al.*, 2007).

Le mécanisme NMD qui s'applique au transcrit primaire du virus est différent car cet ARN n'est pas épissé et ne présente donc pas les marques déposées par les complexes EJC. Cet ARN comporte l'ORF *gag* et sert d'ARNm pour la traduction des précurseurs protéiques Pr76 Gag et Pr180 Gag-Pol-Pro. Cet ARN possède une séquence *cis* appelée RSE (Rous sarcoma virus Stability Element) qui se situe en aval du codon stop naturel de l'ORF de *gag* qui le soustrait au contrôle du système NMD (Weil et Beemon, 2006 ; Withers et Beemon, 2010) (**Figure III-21, A**).

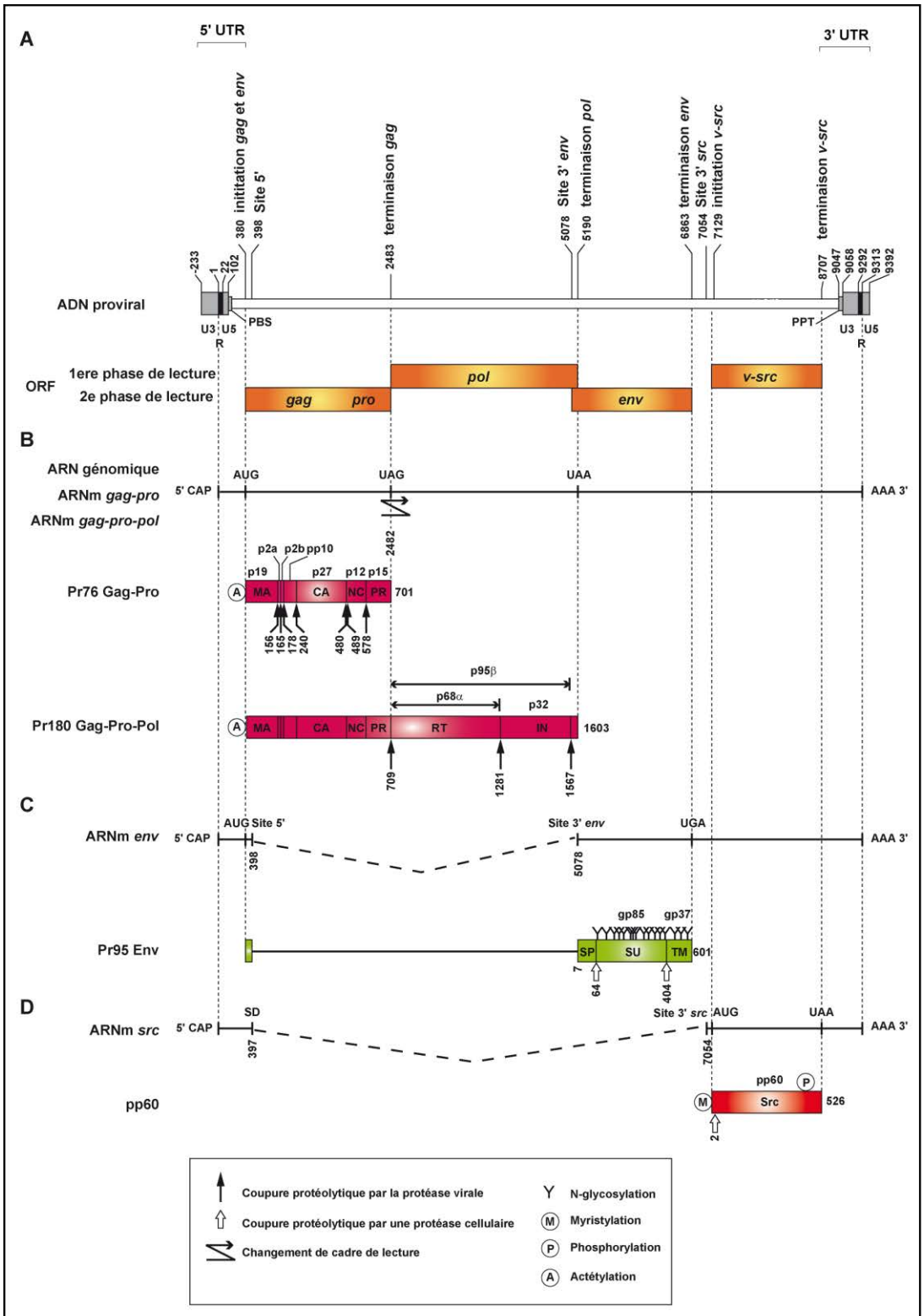



**Figure III-20 : Traduction des protéines à partir des différents ARN viraux du virus RSV (d'après Coffin *et al.* 1997).**

Les protéines synthétisées sont représentées en-dessous de l'ARN ayant servi de matrice. Les séquences non traduites sont représentées par un trait noir, les ORF par des rectangles oranges et les phases de lecture sont matérialisées par un décalage vertical. Les modifications post-traductionnelles des protéines sont représentées comme détaillé dans l'encart. Les précurseurs protéiques traduits à partir de l'ORF *gag-pro*, *gag-pro-pol* sont colorés en rose, ceux de l'ORF *env* en vert et la protéine Src est colorée en rouge.

**A :** Les différentes ORF présentes dans l'ADN proviral du RSV.

**B :** Traduction des ARNm *gag-pro* et *gag-pro-pol.*

**C :** Traduction de l'ARNm *env.*

**D :** Traduction de l'ARNm *src.*

Néanmoins, cet ARN est dégradé par le système NMD lorsqu'un codon stop, naturel ou prématuré, se situe environ à plus de 150 nucléotides en amont de l'élément RSE (LeBlanc et Beemon, 2004 ; Weil et Beemon, 2006 ; Withers et Beemon, 2010). Par ailleurs, l'élimination de cet élément provoque la dégradation de l'ARN par le système NMD, indépendamment de la présence ou de l'absence d'un codon stop prématuré (Weil et Beemon, 2006 ; Withers et Beemon, 2010).

L'élément RSE est une séquence de 296 nucléotides située 110 nucléotides en aval du codon stop naturel de l'ORF *gag* (**Figure III-21, A**). La séquence minimale fonctionnelle de cet élément se situe entre les nucléotides 2577 et 2732 (Withers et Beemon, 2010). Cette séquence est bien conservée dans les virus ASLV et adopte une structure secondaire qui semble être importante pour sa fonction (Weil *et al.*, 2009 ; Withers et Beemon, 2010). Son mécanisme d'action n'est pas encore bien connu. Au cours de la traduction, il est supposé qu'une/des protéines recrutées par l'élément RSE empêche l'interaction entre les protéines associées au ribosome et les protéines Upf de la machinerie NMD (Withers et Beemon, 2010) (**Figure III-21, B**).

### VII.2.7. Le cas particulier de la protéine v-Src

La protéine v-Src est une protéine oncogénique responsable de la formation du sarcome suite à l'infection des cellules par le virus RSV. Contrairement aux autres protéines du virus, cette protéine n'est pas essentielle au cycle viral mais est indispensable pour la transformation des cellules infectées (Martin, 2004). Cette protéine est codée par le gène v-*src* qui correspond au proto-oncogène cellulaire c-*src*. Le gène v-*src* n'est donc pas d'origine virale et a été acquis par le virus RSV au cours d'une recombinaison entre son génome et le génome de la cellule hôte (Swanstrom *et al.*, 1983 ; Takeya et Hanafusa, 1983). Dans la cellule, la protéine c-Src est impliquée dans les mécanismes cellulaires de prolifération, d'adhésion, de mobilité et d'angiogenèse (Yeatman, 2004). Cette protéine est une tyrosine kinase qui est impliquée dans la transduction du signal. Elle est constituée de 4 domaines SH (Src Homology) : SH1, qui est le domaine catalytique portant l'activité kinase, SH2 et SH3, qui modulent les conformations de la protéine, et SH4, à l'extrémité N-terminale, qui permet l'ancrage de la protéine à la membrane plasmique (Martin, 2001).

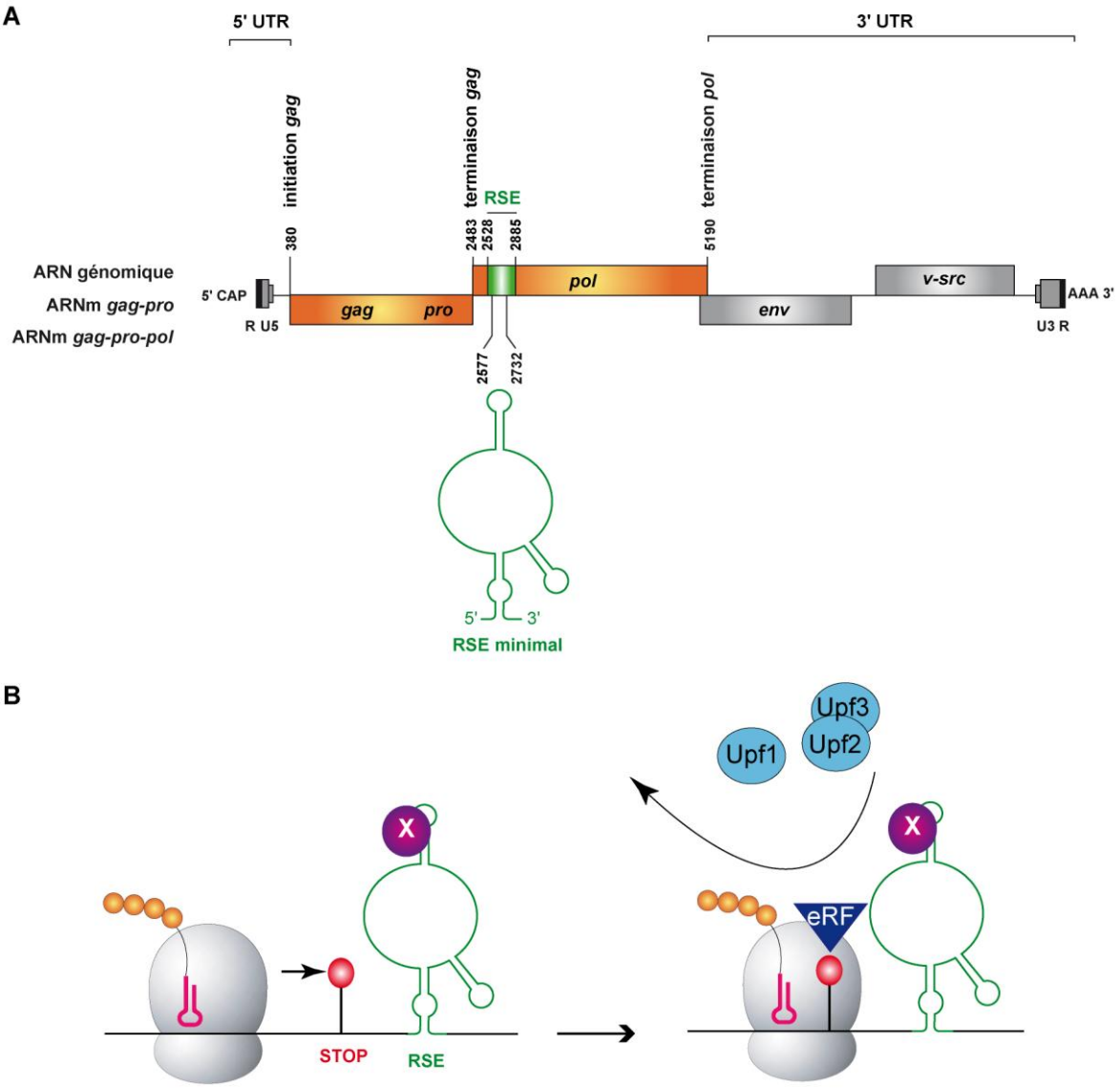



**Figure III-21 : L'élément RSE protège l'ARNm *gag-pro* de l'action du NMD (d'après Withers et Beemon 2010).**

**A :** L'ARNm *gag-pro* contient les ORF *gag* et *pol* (colorées en orange) codant pour les précurseurs Gag-Pro et Gag-Pro-Pol et une partie 3' UTR de 7 kb qui n'est pas traduite qui recouvre les gènes *env* et v-*src* (colorés en gris). L'élément RSE est coloré en vert et se situe 110 nucléotides en aval du codon de terminaison de la traduction de l'ORF *gag*. Cet élément se situe entre les nucléotides 2528-2885 et la structure de sa séquence minimale fonctionnelle (nt 2577-2732) est représentée en vert en-dessous de l'ARNm.

**B :** Implication de l'élément RSE lors de la mise en place du système NMD lors de la traduction de l'ARNm. La terminaison de la traduction se déclenche lorsque les facteurs du ribosomes eRF (eucaryotic Release Facor) reconnaissent le codon stop naturel de l'ORF *gag*. L'élément RSE est supposé fixer une protéine "X" qui empêche le recrutement des facteurs Upf du système NMD par les facteurs eRF.

Son activité est régulée par un cycle de phosphorylation/déphosphorylation d'un résidu tyrosine en position 527. La protéine est inactive lorsqu'elle est en conformation fermée (**Figure III-22, A**). Dans ce cas, la tyrosine 527 est phosphorylée et interagit avec le domaine SH2. La déphosphorylation de ce résidu entraîne la phosphorylation par la protéine de la tyrosine en position 416 et le dépliement de la protéine dans sa conformation ouverte qui est alors active (Cooper *et al.*, 1986) (**Figure III-22, C**).

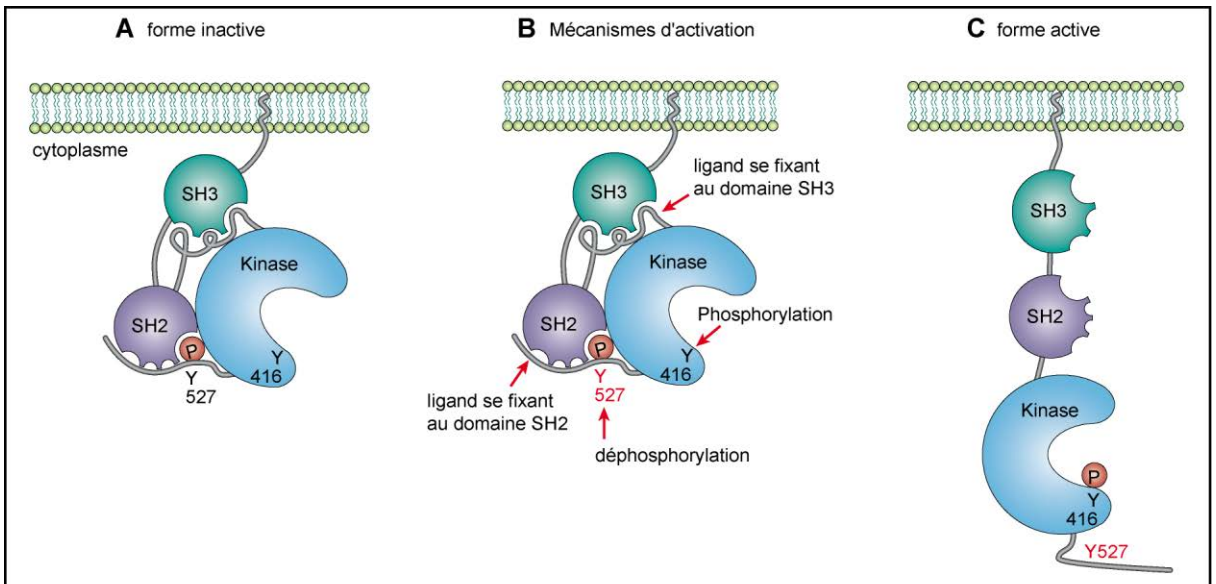



**Figure III-22 : Activation de la protéine c-Src (d'après Martin 2001).**
La protéine c-Src est une tyrosine kinase dont l'activité est régulée par la phosphorylation de la tyrosine 527 chez le poulet (530 chez l'Homme). Les différents domaines de la protéine, SH1 (kinase), SH2, et SH3 sont respectivement représentés en bleu, violet et vert. Le domaine SH4, enchassé dans la membrane plasmique est représenté par un segment gris. La tyrosine 527 qui est absente dans la protéine v-Src est colorée en rouge.
**A :** Conformation incative (fermée) de la protéine c-Src. La tyrosine 527 est phosphorylée et interagit avec le domaine SH2, induisant l'interaction du domaine SH3 avec la séquence située entre le domaine SH2 et le domaine à activité kinase.
**B :** Les différents mécanismes activant la protéine c-Src.
**C :** Conformation active (ouverte) de la protéine c-Src. La tyrosine 527 déphosphorylée provoque la phosphorylation de la tyrosine 416 et le dépliement des domaines SH2 et SH3.

Par rapport à la protéine c-Src, la protéine v-Src est tronquée de 19 résidus en C-terminal ce qui provoque la perte du résidu tyrosine en position 527 qui permet de réguler l'activité kinase. Par conséquent, la protéine v-Src est constamment maintenue dans sa conformation active, ce qui lui confère ses propriétés oncogéniques. L'extrémité C-terminale de la protéine c-Src est codée par le douzième et dernier exon du gène c-*src*. Dans le gène v-*src*, la partie 3' de cet exon est remplacée par une séquence cellulaire se situant en aval du gène c-*src* (Takeya et Hanafusa, 1983)(**Figure III-23**). Hormis cette substitution et quelques mutations ponctuelles, les séquences codantes des deux gènes sont identiques. Le gène v-*src* n'a conservé du gène c-*src* qu'uniquement les 12 exons ainsi qu'une séquence intronique de 17 nucléotides située en amont du premier exon du gène cellulaire (**Figure III-23**).

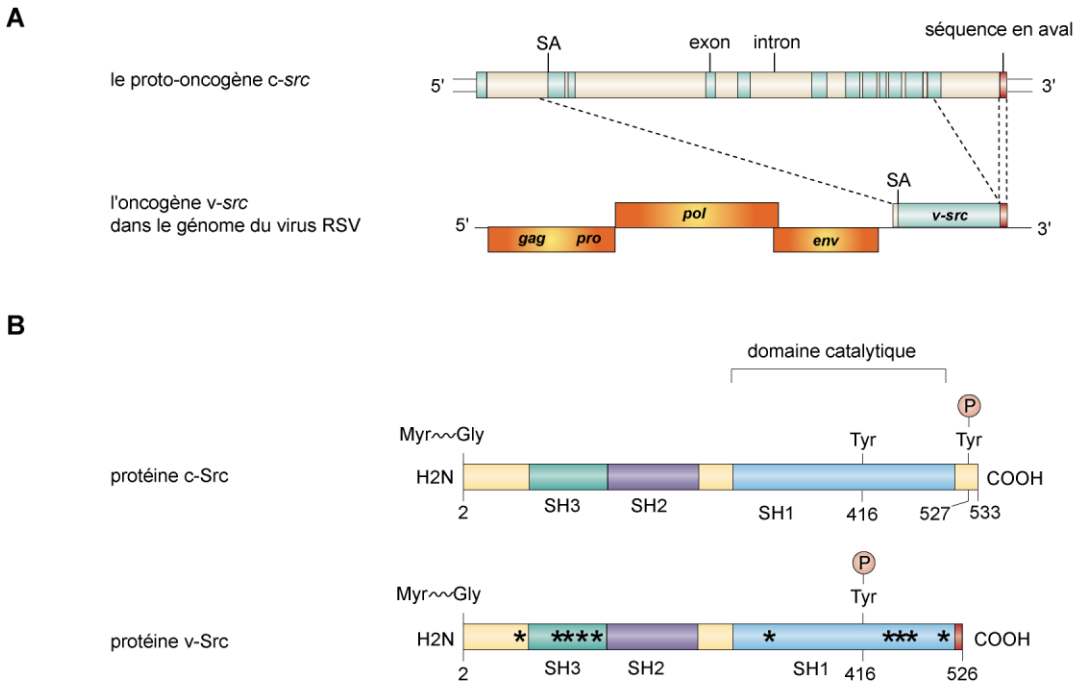



**Figure III-23 : Comparaison des gènes et des protéines codées par le pro-oncogène cellulaire c-*src* et l'oncogène v-*src* du RSV (d'après Martin 2001).**
**A :** Comparaison des gènes c-*src* cellulaire et v-*src* du RSV. Les exons et les introns sont respectivement colorés en vert et en marron. Le site accepteur (SA) commun au gène c-*src* et v-*src* est mentionné.
**B :** Comparaison des protéines c-Src et v-Src. Les domaines SH1, SH2 et SH3 sont respectivement colorés en bleu, violet et vert. Les extrémités N-terminales des deux protéines sont myristoylées afin d'ancrer les protéines dans la membrane plasmique. L'extrémité C-terminale de la protéine v-Src qui est différente de celle de la protéine c-Src est colorée en rouge. Les tyrosines 416 et 527 sont reportées et leur phosphorylation est schématisée par un P. Les astérisques indiquent les substitutions de résidus existant entre la protéine v-Src et la protéine c-Src.

### VII.2.8. Assemblage des nouveaux virions

Lors de l'étape d'assemblage, l'ARN génomique et les précurseurs protéiques viraux se concentrent au niveau de la membrane plasmique pour former les nouveaux virions. Le précurseur Gag Pr76 dont l'assemblage permet la structuration du futur virion, joue un rôle primordial à cette étape. Une fois assemblés, les nouveaux virions sont libérés par bourgeonnement. Ces virions sont dans un premier temps immatures et subissent une étape de maturation protéolytique qui réorganise leur architecture interne et leur permet d'être infectieux.

#### VII.2.8.1. Le rôle du précurseur Pr76 Gag-Pro lors de l'assemblage

Le précurseur Gag Pr76 orchestre l'assemblage des futurs virions. Sa seule expression dans les cellules suffit à produire des pseudos particules virales (VLP, Virus Like Particle). Il

contient plusieurs domaines fonctionnels qui interviennent à différentes étapes lors de l'assemblage des particules virales (**Figure III-24**).

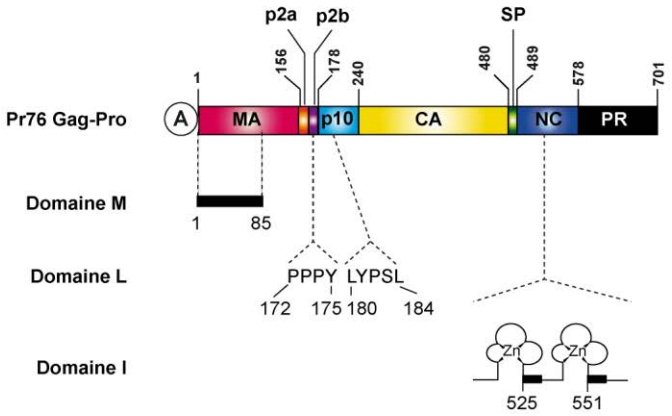



**Figure III-24 : Les différents domaines protéiques du précurseur Pr76 Gag-Pro importants pour le bourgeonnement.**
Le précurseur Pr76 Gag-Pro est composé des protéines suivantes : MA (matrice), p2a, p2b, p10, CA (capside), SP (peptide "spacer"), NC (nucléocapside), PR (protéase). L'acétylation du précurseur Pr76 à son extrémité N-terminale est indiquée par un A et les résidus sont numérotés à partir de cette extrémité. Les différents domaines importants au cours du bourgeonnement sont reportés en-dessous du précurseur. Les domaines M et I sont représentés par des rectangles noirs, et la séquence peptidique du domaine L est indiquée. Les doigts de zinc de la nucléocapside sont représentés par des ions zinc chélatés.

Le domaine M (Membrane binding) est impliqué dans l'interaction du précurseur Gag avec la membrane cellulaire. Il se compose des 85 premiers résidus de la protéine MA, à l'extrémité N-terminale du précurseur Gag (Nelle et Wills, 1996). Il possède 11 résidus basiques qui sont exposés à la surface et qui établissent des interactions électrostatiques avec les phospholipides de la membrane (Callahan et Wills, 2003).

Le domaine I (Interaction) permet la multimérisation des précurseurs Gag via des interactions intermoléculaires. Il est également nécessaire pour que les virions aient une densité protéique appropriée (Bennett *et al.*, 1993 ; Weldon et Wills, 1993). Ce domaine est riche en résidus basiques et est présent en deux copies qui sont juxtaposées aux motifs doigts de zinc dans la protéine NC.

Le domaine L (Late) joue un rôle essentiel à la fin de l'assemblage, au moment de la séparation de la membrane virale et de la membrane plasmique. Cette fonction est assurée par deux motifs peptidiques consensus PPPY et LYPSL qui se situent respectivement dans les peptides p2b et p10 (Dilley *et al.*, 2010 ; Parent *et al.*, 1995 ; Xiang *et al.*, 1996).

### VII.2.8.2. Le recrutement de l'ARN génomique lors de l'assemblage

L'encapsidation et la dimérisation de l'ARN génomique du virus sont deux processus liés dont l'interdépendance a toujours été une question très discutée. Cependant, la nécessité

d'avoir une dimérisation de l'ARN génomique avant son encapsidation est le modèle le plus répandu (Johnson et Telesnitsky, 2010). La séquence essentielle à l'encapsidation de l'ARN, DLS (Dimer Linkage Structure), est une séquence palindromique qui se situe dans la boucle apicale d'une structure tige boucle appelée L3 dans la partie 5'UTR (**Figure III-25**).

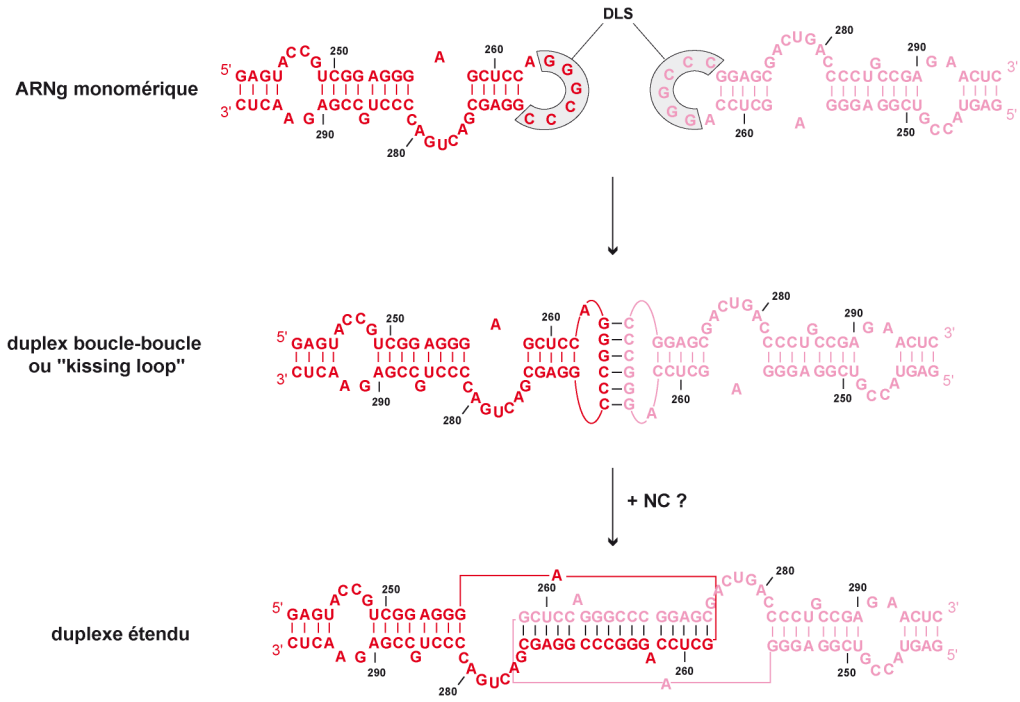



**Figure III-25 : Dimérisation de l'ARN génomique du virus ALV (Ben Ali *et al.*, 2007).**
Le site d'initiation de la dimérisation (DLS) se situe dans la partie 5' UTR de l'ARN génomique dans une structure tige boucle appelée L3. L'initiation de la dimérisation commence par la formation d'une "kissing loop" dans laquelle les séquences palindromiques de chacune des boucles s'apparient pour ensuite former un duplexe étendu lors de la maturation du duplex suite à l'intercation avec la protéine de la nucléocapside dans le virion.

Cette séquence va s'apparier avec la même séquence du second ARN génomique pour former une « kissing loop » (Polge *et al.*, 2000). Une fois encapsidés dans les virions, la protéine de la nucléocapside va permettre à l'appariement entre les deux ARN génomiques de s'étendre pour former un complexe étendu (Ali *et al.*, 2007).

La séquence Ψ est essentielle à l'encapsidation et se situe dans la partie 5' UTR en amont de l'unique site donneur. Par conséquent, tous les ARN viraux, qu'ils soient non épissés comme l'ARN génomique ou épissés comme les ARN *env* et *src*, possèdent cette séquence fonctionnelle. Cette séquence interagit spécifiquement avec les doigts de zinc de la protéine de la nucléocapside qui fait partie du précurseur Pr76 Gag-Pro (Zhou *et al.*, 2007). Le mécanisme qui permet de sélectionner spécifiquement l'ARN génomique des autres ARN viraux n'est pas encore bien compris. Une hypothèse est que l'ARN génomique soit

spécifiquement sélectionné parmi les autres ARN viraux, grâce à son interaction avec le précurseur de Pr76 Gag dans le noyau (Garbitt-Hirst *et al.*, 2009 ; Gudleski *et al.*, 2010) (**Figure III-25**). Après sa traduction dans le cytoplasme, ce précurseur est importé dans le noyau *via* deux séquences NLS présentes dans les parties MA et NC qui interagissent respectivement avec les importines imp α et imp 11 (Butterfield-Gerson *et al.*, 2006 ; Gudleski *et al.*, 2010). Une fois dans le noyau, le précurseur se dissocie des importines pour s'associer par sa partie NC à la séquence Ψ de l'ARN génomique. Cette liaison à l'ARN est importante car elle favorise la dimérisation des monomères de Pr76-Gag (Kenney *et al.*, 2008 ; Ma et Vogt, 2004). En l'absence d'ARN génomique disponible, la protéine Pr76-Gag peut recruter des ARN nucléaires comme le snRNA U6 (Giles *et al.*, 2004). Ce complexe est ensuite exporté vers le cytoplasme par le système d'export CRM1-RanGTP qui s'associe avec le signal NES présent dans la partie p10 du précurseur (Scheifele *et al.*, 2002 ; Scheifele *et al.*, 2005). Une fois dans le cytoplasme, le complexe ARN génomique-Pr76 Gag est adressé au site de bourgeonnement du virus.

#### VII.2.8.3. Le recrutement des protéines virales lors de l'assemblage

Lors du bourgeonnement, le précurseur Gag est dirigé vers un site de la membrane plasmique cellulaire dans laquelle les protéines de l'enveloppe virale sont déjà préalablement enchassées. Le précurseur Gag interagit avec la membrane grâce aux acides aminés basiques de la partie MA (Callahan et Wills, 2000 ; Dalton *et al.*, 2005). Une invagination de la membrane cellulaire nécessaire au bourgeonnement est ensuite réalisée en collaboration avec la machinerie cellulaire responsable de la formation des corps multivésiculaires (MVB).

Ce processus met en jeu des protéines impliquées dans la voie de recyclage des protéines membranaires (Raiborg et Stenmark, 2009) (**Figure III-27**).

Lors du bourgeonnement, il est supposé que le précurseur Gag soit mono-ubiquitinylé. Les enzymes E1, E2 et E3 seraient recrutées par l'un des domaines L, le motif PPPY, qui interagit avec l'ubiquitine ligase (E3) Nedd4 (**Figure III-28**) (Vana *et al.*, 2004). Ni l'ubiquitination du précurseur, ni les résidus mono-ubiquitinylés n'ont été mis en évidence expérimentalement. Cependant, l'inhibition des réactions d'ubiquitinylation dans la cellule empêche tout bourgeonnement viral (Kikonyogo *et al.*, 2001 ; Patnaik *et al.*, 2000 ; Spidel *et al.*, 2004).

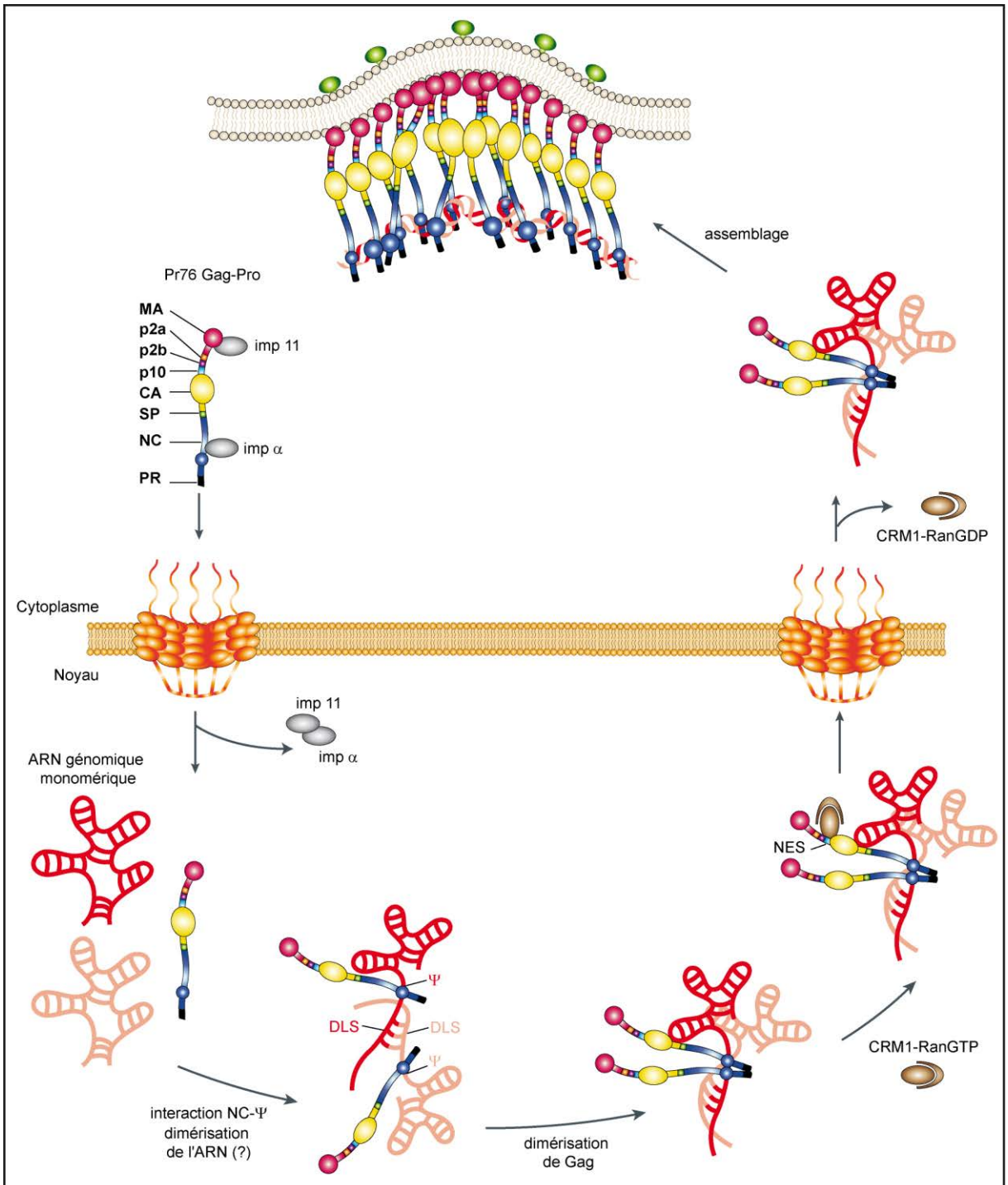



**Figure III-26 : Modèle de recrutement de l'ARN génomique lors de l'assemblage des particules virales (Gudleski *et al.*, 2010).**

Le précurseur Pr76 Gag-Pro est importé dans le noyau par les importines imp α et imp 11. Une fois dans le noyau, le précurseur s'associe aux ARN génomiques grâce à l'interaction de la partie NC avec la séquence Ψ. Ce complexe est exporté dans le noyau par la voie CRM1-RanGTP pour ensuite être adressé vers le site de bourgeonnement viral. La dimérisation des ARN génomiques dans le noyau lors du recrutement n'a pas été démontrée.

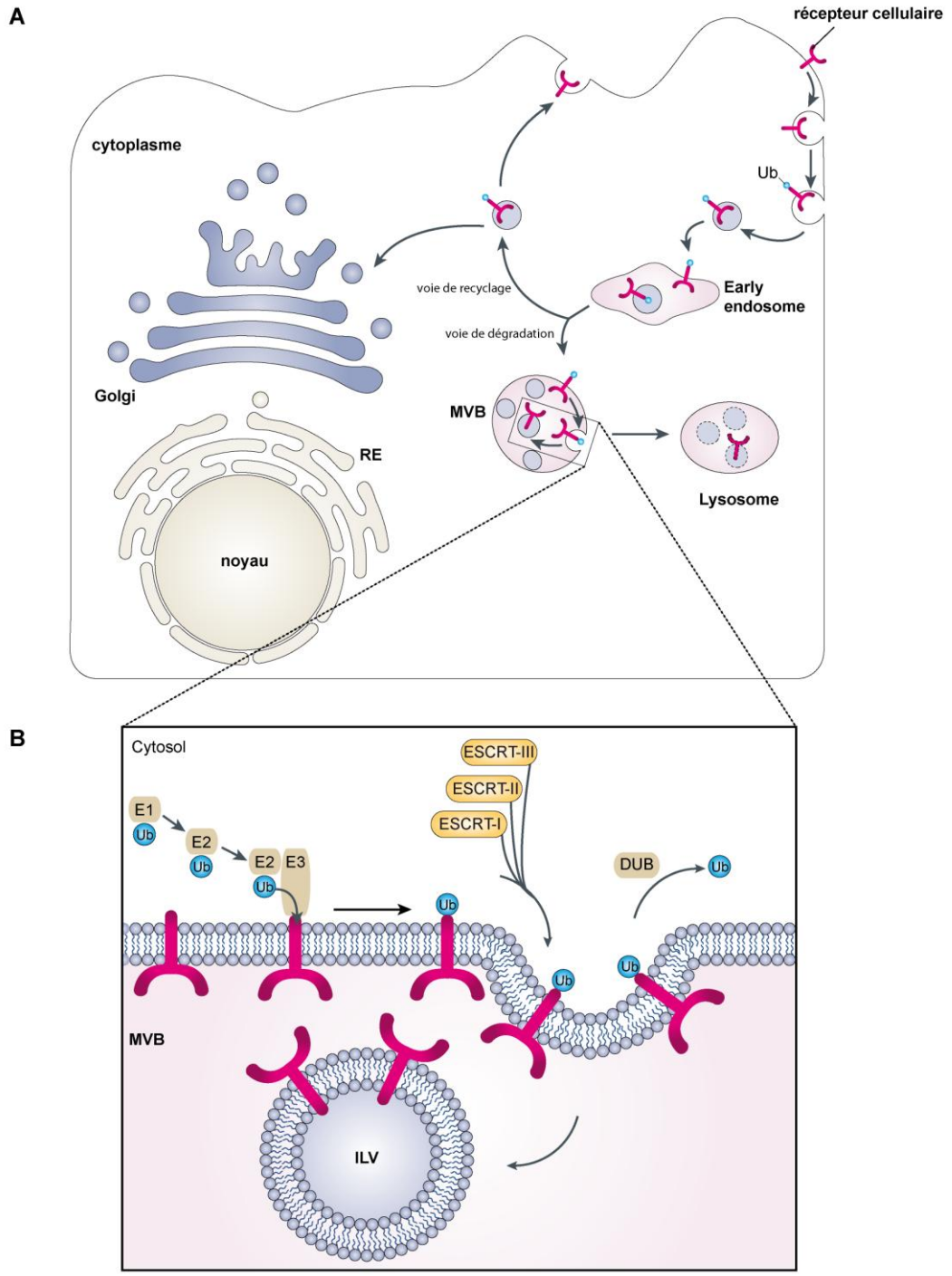



**Figure III-27 : Implication des corps multivésiculaires (MVB) dans le trafic endosomal de la cellule (D'après Raiborg et Stenmark, 2009).**

**A** : Les récepteurs cellulaires qui sont mono-ubiquitinylés et réintègrent le cytoplasme par endocytose en étant enchassées dans des endosomes puis fusionnent avec un endosome appelé « early endosome ». Elles sont ensuite triées soit vers des voies de recyclage puis réintégrées dans la membrane plasmique, soit vers une voie de dégradation lysosomale. Cette dernière nécessite leur transit par les corps multivésiculaires dont la fusion avec le lysosome provoque leur dégradation.

**B :** Détail du processus de formation des vésicules intraluminales (ILV) dans les corps multivésiculaires. La protéine est mono-ubiquitinylée (Ub) par les enzymes E1, E2 et E3 puis sont prises en charge par le complexe ESCRT (Endosomal Sorting Complex Required for Transport). Le complexe ESCRT est composé de 3 complexes (I, II, III) qui agissent séquentiellement pour provoquer l'invagination de la membrane du corps multivésiculaire. La protéine est désubiquitinylée par une désubiquitinylase (DUB) et intègre ensuite les vésicules intraluminales.

Cette ubiquitinylation permettrait dans un deuxième temps de recruter les protéines du complexe ESCRT (Medina *et al.*, 2008) pour les adresser vers le site de bourgeonnement au niveau de la membrane plasmique (**Figure III-28**). Le complexe ESCRT III y serait alors recruté par l'intermédiaire d'une protéine adaptatrice, Alix, dont la présence est essentielle au bourgeonnement du virus ne possédant que le domaine L LYPSL (Dilley *et al.*, 2010). Bien que l'ubiquitinylation et le complexe ESCRT soient essentiels au bourgeonnement viral, le recrutement et l'implication de ces différentes protéines et de ces différents complexes ne sont pas encore connus en détails.

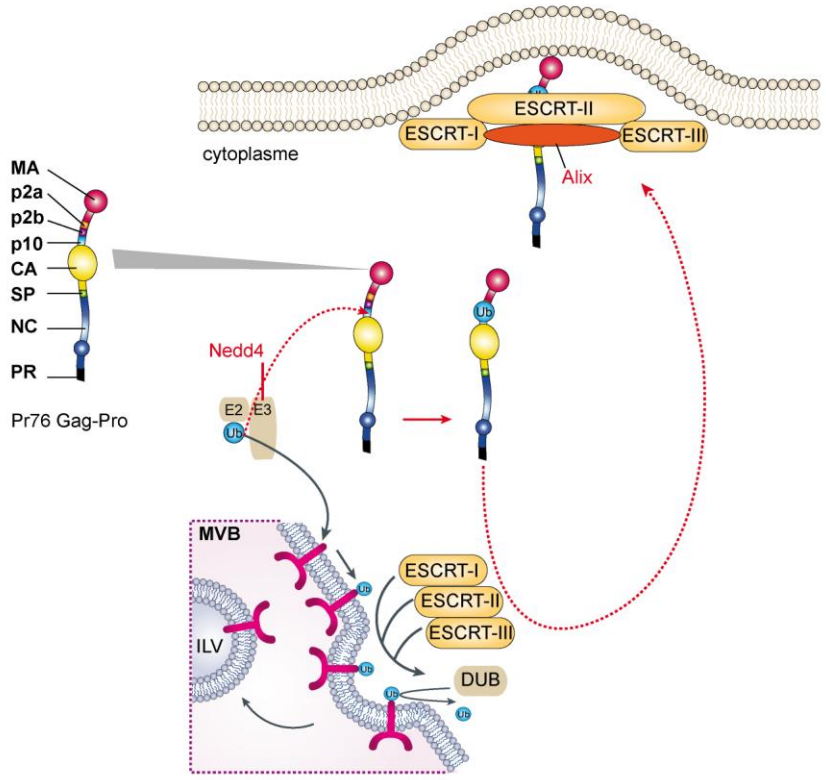



**Figure III-28 : Ubiquitinylation du précurseur Pr76 Gag-Pro et recrutement du complexe ESCRT lors du bourgeonnement (d'après Pincetic et Leis 2009).**
Lors du bourgeonnement, le précurseur Pr76 Gag-Pro ubiquitinylé détourne la machinerie cellulaire ESCRT pour la rediriger vers le site d'assemblage du virion au niveau de la membrane plasmique. Le recrutement des différents complexes et protéines par le précurseur viral sont représentés par des flèches en pointillés rouges. Les protéines Nedd4 et Alix ayant été démontrées comme essentielles dans le processus de bourgeonnement viral sont en rouge. Les différentes protéines composant le précurseur Pr76 Gag-Pro sont précisées.

### VII.2.9. La maturation des nouveaux virions (Figure III-29)

Lors de la maturation, la protéase clive les précurseurs Pr76 Gag et Pr180 Gag-Pro-Pol et libère les protéines structurales (MA, CA, NC) et les protéines enzymatiques (PR, NT, IN). Après ces clivages, la protéine MA reste associée à la membrane, tandis que la protéine de la nucléocapside va se fixer sur l'ARN génomique dont les séquences de dimérisation forment alors un duplex étendu. Les protéines de la capside s'associent en hexamère pour former la capside virale. Les peptides p2a, p2b, p10 et SP sont également libérés mais n'ont aucun rôle dans la structure du virion mature. En revanche, leur présence est essentielle lors du bourgeonnement et dans l'organisation structurale des virions immatures (Campbell et Vogt, 1997 ; Keller *et al.*, 2008).

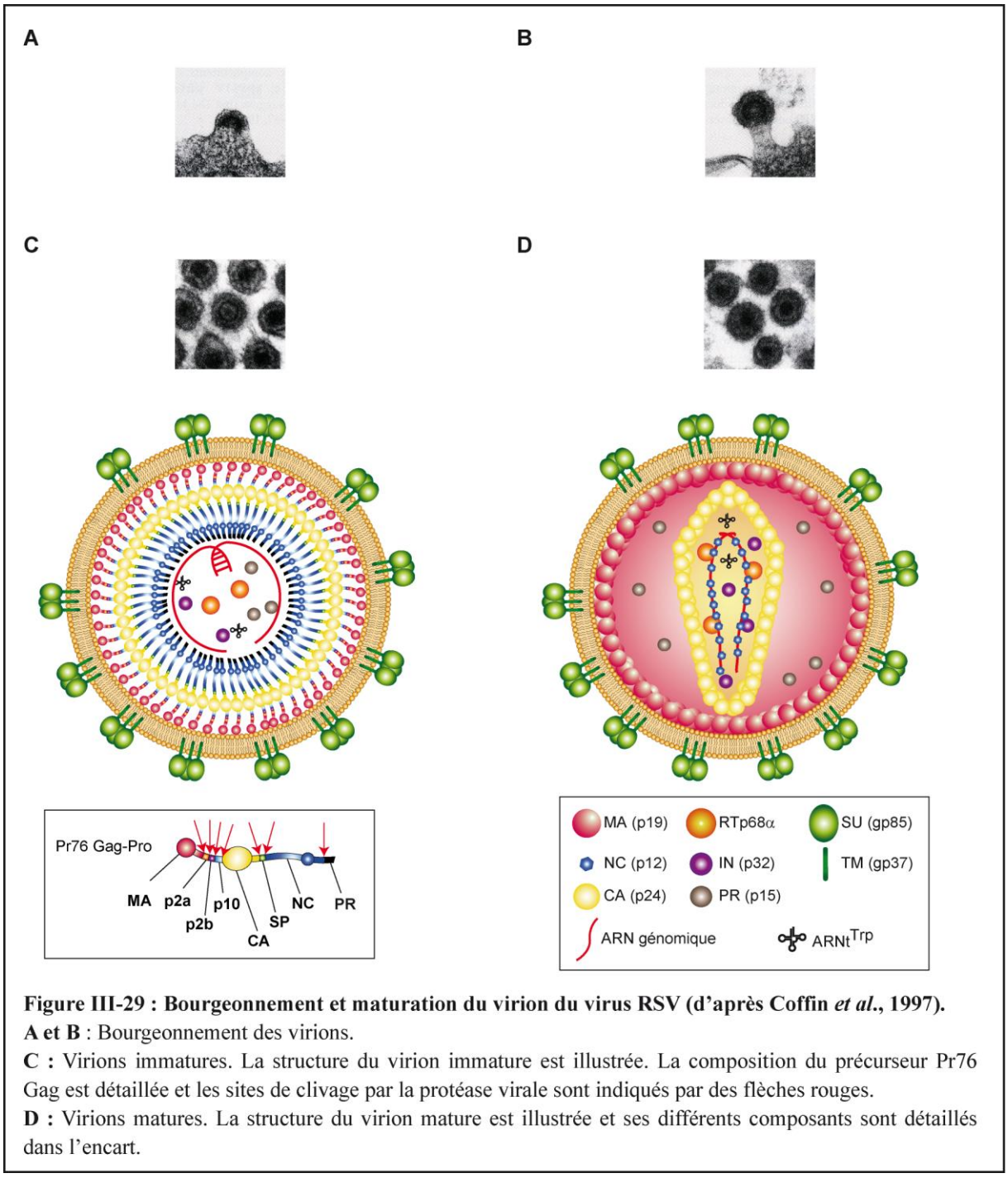



**Figure III-29 : Bourgeonnement et maturation du virion du virus RSV (d'après Coffin *et al.*, 1997).**
**A et B** : Bourgeonnement des virions.
**C :** Virions immatures. La structure du virion immature est illustrée. La composition du précurseur Pr76 Gag est détaillée et les sites de clivage par la protéase virale sont indiqués par des flèches rouges.
**D :** Virions matures. La structure du virion mature est illustrée et ses différents composants sont détaillés dans l'encart.

# Chapitre IV: La régulation de l'épissage du RSV et l'élément NRS

## I. Description des sites d'épissage et des ARNm viraux produits

Comme nous l'avons déjà mentionné, l'épissage joue un rôle central lors du cycle viral car c'est de cette étape de maturation dont va dépendre le devenir du transcrit primaire qui est à l'origine de l'expression de la totalité des gènes viraux. Le maintien d'un équilibre entre les taux relatifs d'ARN épissés et non épissés est primordial pour la multiplication du virus. De ce fait, l'étape d'épissage est fortement régulée afin qu'une partie importante du transcrit primaire échappe aux réactions d'épissage. Suivant la complexité des rétrovirus, les stratégies élaborées pour établir cet équilibre entre les ARN épissés et non épissés diffèrent. Le RSV est un rétrovirus simple dont l'épissage du pré-ARNm se distingue en plusieurs points de celui des rétrovirus complexes comme HIV-1.

Contrairement à l'ARN de HIV-1 dont l'utilisation alternative des 4 sites donneurs et des 8 sites accepteurs d'épissage produit plus de 40 ARNm différents (Purcell et Martin, 1993), l'épissage de l'ARN du RSV ne met en jeu qu'un site donneur et deux sites accepteurs (Chang et Stoltzfus, 1985) (**Figure IV-1**). Le site donneur est localisé au début du gène *gag* (nt 398), tandis que les deux sites accepteurs, 3'*env* (nt 5078) et 3'*src* (nt 7054), sont respectivement situés au début des gènes *env* et *src*. En combinaison avec le seul site donneur, l'utilisation alternative des sites 3'*env* et 3'*src* produit deux ARNm mono-épissés, les ARNm *env* et *src*. L'ARNm *env* code un polypeptide dont la protéolyse permet la production des protéines de l'enveloppe virale qui sont essentielles à la multiplication virale. L'ARNm *src* code la protéine oncogénique v-Src qui est responsable de la transformation des cellules hôtes mais qui n'est pas essentielle à la multiplication virale.

Lors de l'infection des cellules aviaires, le taux relatif d'épissage alternatif de ces deux ARNm est équivalent, à savoir que l'ARNm *env* représente 13% des ARN totaux viraux et l'ARNm *src* 12% (Berberich et Stoltzfus, 1991a). Ainsi, dès l'infection des cellules, les processus de réplication virale et de transformation sont mis en place simultanément. Le reste de l'ARN viral, à savoir environ 75% de l'ARN viral total, n'est constitué que du transcrit primaire qui n'est pas épissé. Il est soit incorporé dans les futurs virions pour constituer l'ARN génomique, soit traduit pour produire les autres protéines structurales et enzymatiques. Cette balance entre les ARN épissés et non épissés est maintenue durant tout le cycle viral, ce

qui signifie qu'il existe, dès le départ de l'infection, une régulation permettant de limiter l'épissage de l'ARN viral et de maintenir un taux élevé d'ARN non épissé.

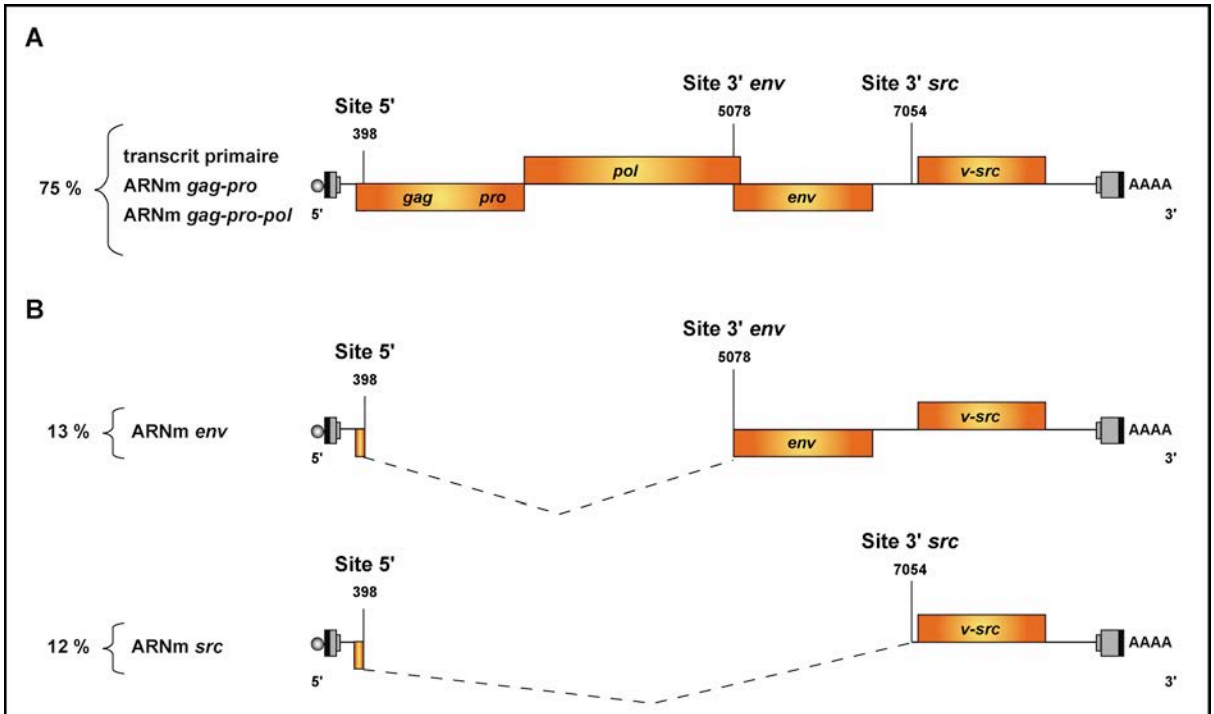



**Figure IV-1 : Transcrit primaire et ARNm du virus RSV (McNally, 2008).**
**A :** Représentation du transcrit primaire du RSV.
**B :** Représentation des ARNm *env* et *src* du virus RSV.
Les gènes sont représentés par des rectangles orangés et leur nom est indiqué. La coiffe est représentée par une sphère grise, les séquences des LTR par des rectangles gris et noirs. Les sites 5' et 3' ainsi que leur positions sont mentionnés. L'épissage est symbolisé par un trait en pointillé. Le pourcentage qui représente la quantité de chacun des ARN par rapport à l'ARN viral total est indiqué.

Le virus HIV-1 en revanche a, lui, un épissage chronologiquement différencié de ses ARNm, Au cours de son cycle, on peut distinguer deux phases bien distinctes. Au cours de la première phase, ou phase précoce, le transcrit primaire est multiplement épissé, ce qui conduit à la production des ARNm de la classe 2 kb codant les protéines régulatrices. La phase tardive est marquée par la production d'ARNm peu ou non épissés appartenant à la classe de 4 kb codant les protéines auxiliaires, structurales et enzymatiques et à la classe de 9 kb correspondant à l'ARN génomique (Purcell et Martin, 1993). Cette différentiation chronologique de l'épissage est notamment réalisée grâce à la régulation exercée en *trans* par une des protéines virales du HIV-1 comme la protéine Rev qui, lorsqu'elle atteint un taux suffisant en phase précoce, permet le passage en phase tardive. Le RSV, qui est un rétrovirus simple, ne code pas de protéines régulatrices (Schwartz *et al.*, 1983). Pour pallier leur absence, le RSV utilise des séquences ARN régulatrices agissant en *cis* et l'interaction de celles-ci avec des facteurs

cellulaires pour réguler son épissage. Sachant qu'il est nécessaire que le transcrit primaire soit majoritairement non épissé pour que le virus puisse se multiplier, la régulation de l'épissage exercée par les séquences en *cis* doit être globalement négative. Cette limitation de l'épissage repose à la fois sur le fait que les séquences intrinsèques des sites 3' d'épissage sont sous-optimales et sur l'action de séquences régulatrices.

## II. L'unique site donneur est un site fort

L'unique site donneur du transcrit primaire du RSV est localisé dans le gène *gag* (nt 398) (Ficht *et al.*, 1984 ; Hackett *et al.*, 1982 ; Swanstrom *et al.*, 1982). Ce site est un site donneur fort qui est efficacement reconnu par les composants du spliceosome lors de l'épissage. Sept des huit nucléotides de sa séquence sont identiques à ceux de la séquence consensus des sites donneurs définis pour les pré-ARNm des eucaryotes supérieurs (**Figure IV-2**). Comme tous les sites optimaux, ce site 5' possède une forte complémentarité avec le snRNA U1 et forme 2 des 3 appariements possibles avec le snRNA U6.

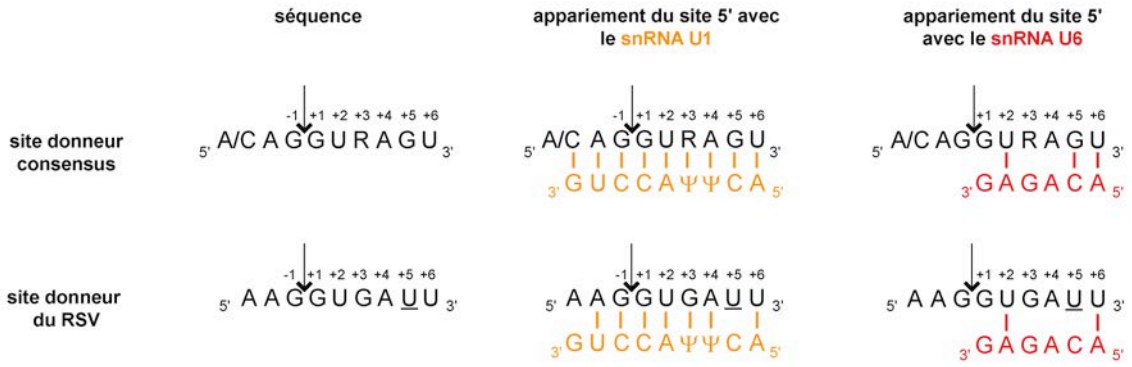



**Figure IV-2 : Comparaison de la séquence du site donneur d'épissage de l'ARN virus RSV avec la séquence consensus des sites donneurs des pré-ARNm des eucaryotes supérieurs.**
Le nucléotide du site donneur du RSV qui diffère de la séquence consensus est souligné. La flèche verticale représente la bordure exon-intron. Les appariements de séquences entre les sites donneurs et les séquences des snRNA U1 et U6 sont représentés. La séquence du snRNA U1 est colorée en orange et la séquence du snRNA U6 est colorée en rouge. R : purine.

Que ce soit dans le contexte viral ou dans un contexte hétérologue, le site 5' de l'ARN du RSV peut fonctionnellement être remplacé ou remplacer d'autres sites 5' forts sans que cela n'affecte le taux d'épissage du pré-ARNm et ce, indépendamment de la force du site 3' auquel il est couplé (Katz et Skalka, 1990 ; McNally et Beemon, 1992). Dans le contexte viral, l'utilisation du site 5' n'est pas activée et sa force qui lui est intrinsèque ne dépend que de sa

séquence. En revanche, son utilisation est réprimée par une séquence régulatrice appelée NRS qui a fait l'objet de nos études et dont nous détaillerons ultérieurement le mécanisme d'action.

## III. Les sites accepteurs 3' *env* et 3' *src* sont faibles

### III.1. Le site 3'*env* a une boîte de branchement sous-optimale

Le site 3'*env* se situe au début du gène *env* (nt 5078) et son utilisation en combinaison avec le site donneur conduit à la production de l'ARNm *env* de 4,7 kb (Chang et Stoltzfus, 1985 ; Ficht *et al.*, 1984). L'ARNm *env* ne contient donc que 2 exons qui codent pour le précurseur polypeptidique Pr95Env (**Figure IV-3**). Le site 3'*env* se situe dans une région commune aux 2 ORF Gag-Pro-Pol et Env qui se chevauchent sur une longueur de 112 nucléotides (**Figure IV-3**) (Schwartz *et al.*, 1983). Cependant, la séquence du site 3'*env* est plus importante pour l'épissage que pour la traduction (Katz et Skalka, 1988). Le site 3'*env* est un site faiblement épissé à cause de sa séquence de boîte de branchement qui est sous-optimale. Celle-ci ne peut établir que 2 appariements avec le snRNA U2 au lieu des 6 établis par la séquence consensus (**Figure IV-3**). Son point de branchement, bien que canonique, est aussi sous-optimal car il se situe à 16 nt en amont de la bordure intron/exon alors que ceux des sites accepteurs forts se situent dans une fenêtre de 18 à 40 nt (Fu *et al.*, 1991). La séquence polypyrimidique quant à elle est optimale et n'est pas responsable de la faible efficacité d'utilisation du site 3'*env* (Bouck *et al.*, 1995).

La faible efficacité du site 3'*env* a été démontrée *in cellulo* dans le contexte hétérologue d'un mini-gène d'épissage comprenant l'exon 2 et l'exon 3 séparés par l'intron 2 du gène c-*myc* de poulet. Le remplacement du site 3' fort de l'exon 3 de c-*myc* par le site 3'*env* provoque une diminution de l'épissage (McNally et Beemon, 1992). Le fait que le même résultat soit obtenu lorsque le site 3'*env* est couplé au site 5' du RSV ou au site 5' de l'exon 2 du gène *c-myc*, qui sont tous les deux des sites forts, indique que le site 3'*env* est un site accepteur faible et que ce caractère sous-optimal est inhérent à sa propre séquence (McNally et Beemon, 1992).

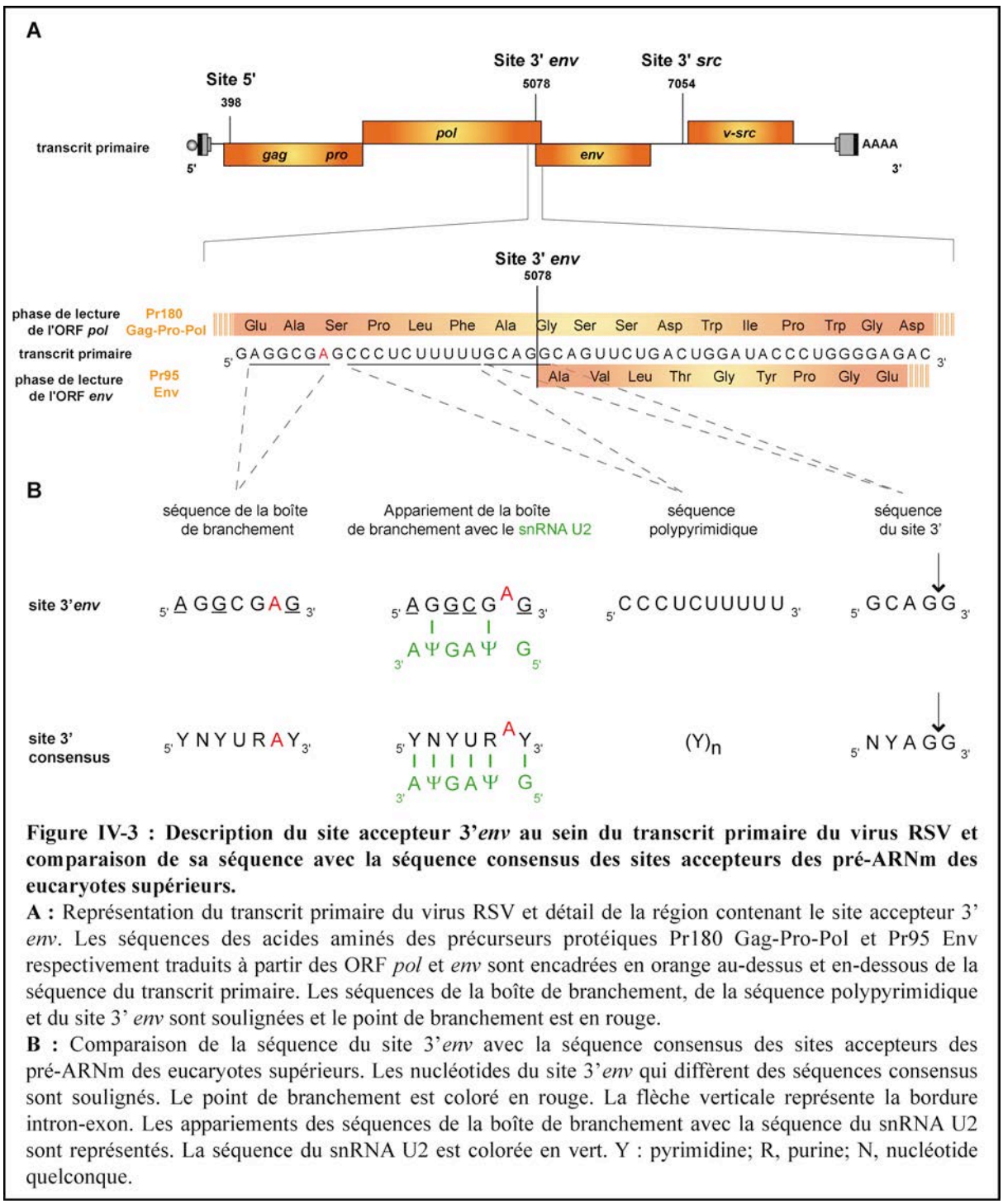



**Figure IV-3 : Description du site accepteur 3'*env* au sein du transcrit primaire du virus RSV et comparaison de sa séquence avec la séquence consensus des sites accepteurs des pré-ARNm des eucaryotes supérieurs.**

**A :** Représentation du transcrit primaire du virus RSV et détail de la région contenant le site accepteur 3' *env*. Les séquences des acides aminés des précurseurs protéiques Pr180 Gag-Pro-Pol et Pr95 Env respectivement traduits à partir des ORF *pol* et *env* sont encadrées en orange au-dessus et en-dessous de la séquence du transcrit primaire. Les séquences de la boîte de branchement, de la séquence polypyrimidique et du site 3' *env* sont soulignées et le point de branchement est en rouge.

**B :** Comparaison de la séquence du site 3'*env* avec la séquence consensus des sites accepteurs des pré-ARNm des eucaryotes supérieurs. Les nucléotides du site 3'*env* qui diffèrent des séquences consensus sont soulignés. Le point de branchement est coloré en rouge. La flèche verticale représente la bordure intron-exon. Les appariements des séquences de la boîte de branchement avec la séquence du snRNA U2 sont représentés. La séquence du snRNA U2 est colorée en vert. Y : pyrimidine; R, purine; N, nucléotide quelconque.

Le caractère sous-optimal du site 3'*env* est très important pour la multiplication du virus. Dans le contexte viral, toute optimisation de la séquence de la boîte de branchement du site 3'*env* provoque dans un premier temps un défaut de multiplication et dans un second temps l'apparition de révertants qui rétablissent le phénotype sauvage (Katz *et al.*, 1988 ; Katz et Skalka, 1990). Cette optimisation provoque une augmentation de l'épissage au site 3'*env* qui déséquilibre la balance entre le taux de pré-ARNm et d'ARNm viraux ce qui se traduit ensuite

par une réplication virale plus lente. Cette balance est rééquilibrée dans le cas des virus révertants qui possèdent des mutations dans la séquence de la boîte de branchement ou du point de branchement ou encore des troncations au niveau d'une séquence ESE située en aval du site accepteur qui ont pour effet de diminuer l'utilisation du site 3'*env* (Katz *et al.*, 1988 ; Katz et Skalka, 1990). L'apparition rapide des révertants illustre le fait que le site 3'*env* est soumis à une forte pression de sélection afin de conserver une faible efficacité. Le caractère sous-optimal de la séquence de sa boîte de branchement constitue un élément de régulation essentiel de l'épissage au site 3'*env*.

### III.2. Le site 3'*env* est activé par une séquence ESE

L'épissage au site 3'*env* est activé spécifiquement par une séquence ESE de 50 nt de long, riche en purines et qui se situe 31 nt en aval de ce site (**Figure IV-4**) (Fu *et al.*, 1991). L'effet positif exercé par cette séquence compense le caractère sous-optimal du site 3'*env* lié à sa boîte de branchement non canonique. Cependant, même en présence de cette séquence ESE, le site 3'*env* reste un site accepteur faible car son utilisation n'est pas suffisamment activée pour en faire un site accepteur fort (Katz *et al.*, 1988). Cette séquence est essentielle à la régulation de l'épissage au site 3'*env* car sa troncation du génome viral entraîne un défaut de réplication du virus qui est dû à une utilisation trop faible du site 3'*env*, et ce même quand la boîte de branchement de ce site est optimisée (Fu *et al.*, 1991 ; Katz *et al.*, 1988). Par ailleurs, des expériences d'épissage et de pontages aux UV réalisées *in vitro* avec un mini-gène contenant le site 5' fort du RSV et le site 3'*env* indiquent que la troncation de la séquence ESE provoque une diminution de l'assemblage des complexes spliceosomaux ainsi qu'une diminution de la fixation du facteur U2AF65 (Bouck *et al.*, 1998 ; Fu *et al.*, 1991). Ce facteur reconnaît spécifiquement, lorsque la séquence polypyrimidique est optimale, l'enchaînement de pyrimidines adjacent au site 3' d'épissage et facilite le recrutement de la snRNP U2 au niveau de la boîte de branchement. Lorsque la séquence polypyrimidine n'est pas optimale, la fixation de ce facteur sur les sites accepteurs peut être facilitée par la fixation de protéines SR sur des séquences ESE situées en amont dans l'exon. Il a été proposé que la séquence ESE associée au site 3'*env* favorise la fixation de U2AF65 sur la séquence polypyrimidique de ce site (**Figure IV-4**). Par ailleurs, l'activation de l'épissage par cette séquence ESE a été démontrée *in vitro* dans un contexte hétérologue (Tanaka *et al.*, 1994). Une partie de la séquence ESE (nt +43 à +78 par rapport au site 3'*env*) a été fusionnée à l'extrémité 5' du 4e exon du gène *dsx* de drosophile. L'épissage de cet exon est dépendant de la présence d'une séquence activatrice en amont de son site accepteur et les résultats d'épissage indiquent que la

séquence de l'ESE du gène *env* est capable de remplir cette fonction (Tanaka *et al.*, 1994). Cette séquence ESE fixe préférentiellement la protéine SR SRp40 et peut également fixer ASF/SF2 (Staknis et Reed, 1994).

Ainsi, l'élément ESE permet d'activer le site 3'*env* sous-optimal, et il est essentiel pour la multiplication du virus car en son absence, le site 3'*env* ne serait pratiquement pas utilisé.

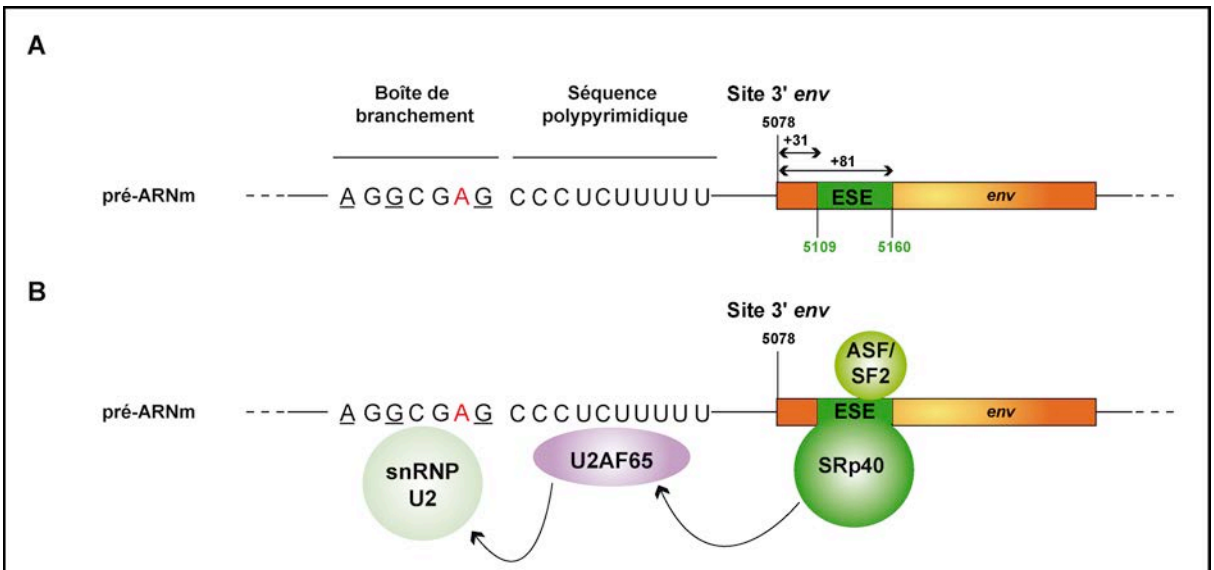



**Figure IV-4 : Le site 3'*env* possède une séquence ESE.**
**A :** Représentation de la région du site 3'*env* du pré-ARNm du RSV. Le site 3'*env*, la boîte de branchement et la séquence polypyrimidique sont indiqués, le point de branchement est en rouge, les nucléotides de la boîte de branchement non consensus sont soulignés. Le gène *env* et la séquence ESE sont respectivement représentés par des rectangles orange et verts.
**B :** Proposition du mécanisme d'activation de l'épissage au site 3'*env* par la séquence ESE. La protéine SR SRp40 et/ou en moindre quantité la protéine ASF/SF2, se fixeraient sur la séquence ESE pour recruter le facteur U2AF65 sur la séquence polypyrimidique qui, à son tour, favoriserait la fixation de la snRNP U2 sur la boîte de branchement.

## III.3. Le site 3'*src*

### III.3.1. Le site 3'*src* est d'origine cellulaire

Le site 3'*src* est au début du gène *src* (nt 7054) et son utilisation en combinaison avec le site 5' unique conduit à la production de l'ARNm *src* de 2,7 kb (Chang et Stoltzfus, 1985 ; Swanstrom *et al.*, 1983). Contrairement au site 3'*env*, le site 3'*src* ainsi que la totalité de la séquence du gène *src* sont d'origine cellulaire et correspondent à la séquence de l'ADNc du gène *c-src* (**Chapitre III, Figure III-23**) (Dorai *et al.*, 1991 ; Swanstrom *et al.*, 1983).

Dans le contexte cellulaire, le site 3'*src* définit la bordure en 5' du premier exon qui est épissé de façon constitutive avec différents sites donneurs alternatifs situés plus en amont

dans le gène *c-src*. Cet épissage alternatif confère à l'ARNm *c-src* une hétérogénéité de séquence en 5' dont la fonction reste à déterminer (Dorai *et al.*, 1991).

Dans le contexte viral, aucun de ces sites donneurs alternatifs n'est présent dans le gène *v-src* qui ne possède en amont du site 3'*src* qu'une séquence de 17 nt correspondant à une séquence intronique du gène *c-src* (Dorai *et al.*, 1991 ; Swanstrom *et al.*, 1983). Cette séquence de 17 nt comporte la boîte de branchement putative, ainsi que la séquence polypyrimidique du site 3'*src* (**Figure IV-5**) (Zhang et Stoltzfus, 1995). Cette séquence se situe dans la région entre les gènes *env* et *v-src* et ne fait partie d'aucune ORF (**Figure IV-5**) (Schwartz *et al.*, 1983). De ce fait, contrairement au site 3'*env*, la séquence du site 3'*src* est dépourvue de fonction codante.

L'ARNm *src* est un ARNm mono-épissé qui comporte deux exons dont seul le deuxième exon code la protéine v-Src (Hughes *et al.*, 1984), la protéine oncogénique responsable de l'apparition de sarcome dans les cellules aviaires (Brugge et Erikson, 1977). Comme nous l'avons dit dans le chapitre précédent, le gène *v-src* comporte les 12 séquences exoniques du gène cellulaire c-*src*, à l'exception de la séquence de l'exon 12 qui est partiellement tronquée à son extrémité 3' (Takeya et Hanafusa, 1983). Cette troncation induit la perte de la séquence codant les 19 acides aminés en C-terminal de la protéine c-Src qui contient la Tyrosine en position 527, qui permet de maintenir la protéine dans sa conformation inactive (Cooper *et al.*, 1986). Par conséquent, dans le contexte viral, la protéine v-Src traduite est constamment maintenue dans une conformation active, ce qui lui confère ses propriétés oncogéniques.

Contrairement aux autres gènes viraux, le gène *v-src* n'est pas indispensable à la réplication virale mais il est en revanche essentiel à la transformation des cellules (Duesberg et Vogt, 1970). Sachant que les processus de transformation et de réplication sont indissociés durant le cycle viral du RSV, l'utilisation du site 3'*src* lors de l'épissage doit donc être régulé.

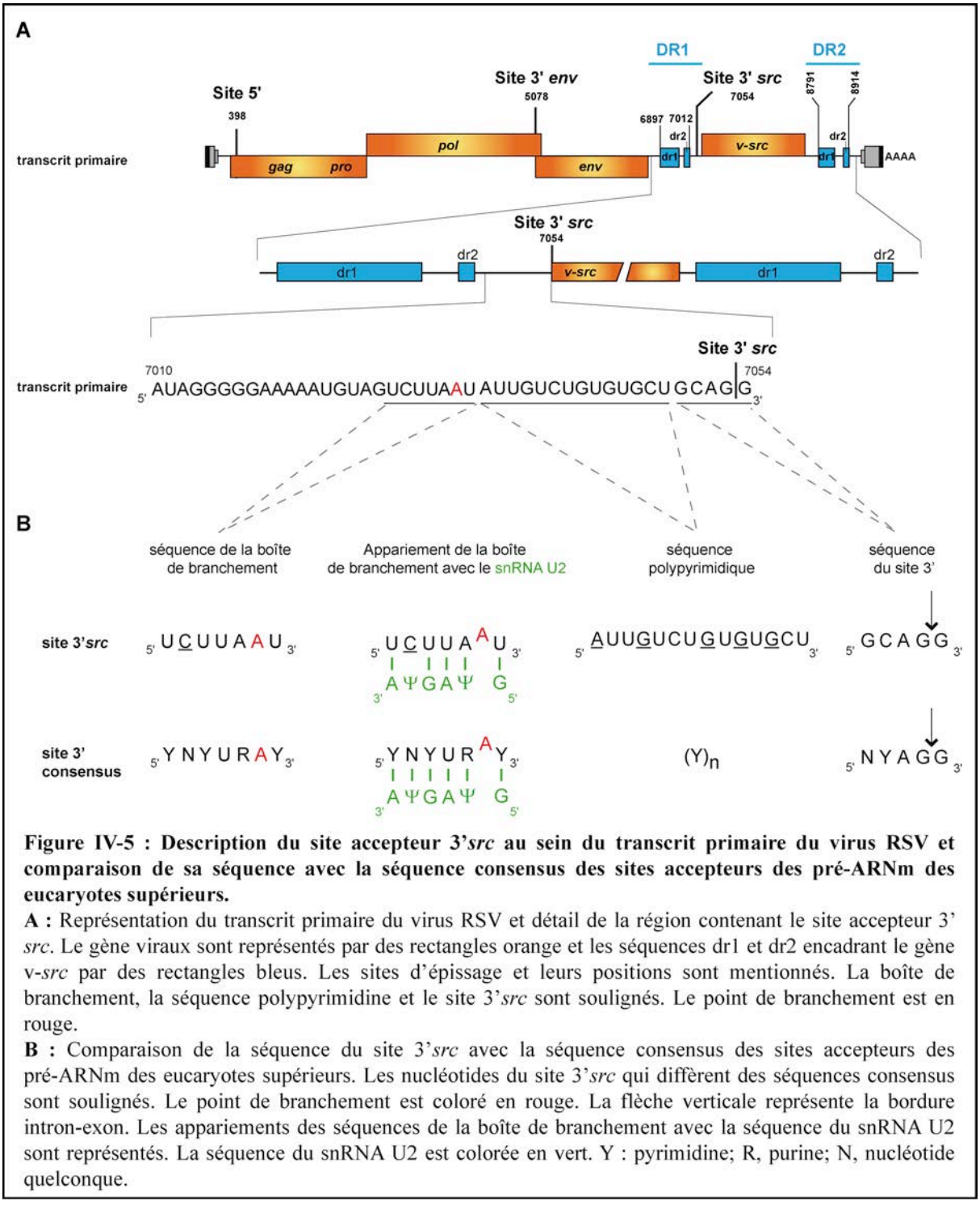



**Figure IV-5 : Description du site accepteur 3'*src* au sein du transcrit primaire du virus RSV et comparaison de sa séquence avec la séquence consensus des sites accepteurs des pré-ARNm des eucaryotes supérieurs.**

**A :** Représentation du transcrit primaire du virus RSV et détail de la région contenant le site accepteur 3' *src*. Le gène viraux sont représentés par des rectangles orange et les séquences dr1 et dr2 encadrant le gène v-*src* par des rectangles bleus. Les sites d'épissage et leurs positions sont mentionnés. La boîte de branchement, la séquence polypyrimidine et le site 3'*src* sont soulignés. Le point de branchement est en rouge.

**B :** Comparaison de la séquence du site 3'*src* avec la séquence consensus des sites accepteurs des pré-ARNm des eucaryotes supérieurs. Les nucléotides du site 3'*src* qui diffèrent des séquences consensus sont soulignés. Le point de branchement est coloré en rouge. La flèche verticale représente la bordure intron-exon. Les appariements des séquences de la boîte de branchement avec la séquence du snRNA U2 sont représentés. La séquence du snRNA U2 est colorée en vert. Y : pyrimidine; R, purine; N, nucléotide quelconque.

### III.3.2. Le site 3'*src* a une séquence polypyrimidique sous-optimale

Le site 3'*src* possède une séquence polypyrimidique sous-optimale qui est responsable de la faible utilisation de ce site accepteur (Zhang et Stoltzfus, 1995). Cette séquence se situe 4 nt en amont du site 3'*src* et se compose de 9 pyrimidines non contiguës entrecoupées par 5 purines (**Figure IV-5**). La boîte de branchement du site 3'*src* est optimale et se situe 18 nt en

amont de la bordure intron/exon. Cette séquence possède 6 des 7 nucléotides consensus qui permettent d'établir 5 appariements avec le snRNA U2 (**Figure IV-5**).

Le caractère sous-optimal du site 3'*src* a été démontré *in cellulo* dans un contexte hétérologue à l'aide du mini-gène similaire à celui ayant servi à étudier le site 5' et le site 3'*env* (McNally et Beemon, 1992). Après transfection des cellules CEF, l'épissage au site 3'*src* produit peu d'ARNm, qu'il soit couplé au site 5' du RSV ou au site 5' fort de l'exon 2 du gène *c-myc*. En revanche, le remplacement du site 3'*src* par le site 3' fort du troisième exon de c-myc provoque une augmentation de l'épissage. Le site 3'*src* est donc un site accepteur faible et ce caractère sous-optimal est inhérent à sa propre séquence, tout comme le site 3'*env* (McNally et Beemon, 1992).

Comme pour le site 3'*env*, le caractère sous-optimal du site 3'*src* est très important pour la multiplication du virus. Dans le contexte viral, l'optimisation de la séquence polypyrimidique du site 3'*src* provoque dans un premier temps un défaut de multiplication virale et dans un second temps l'apparition de virus révertants qui rétablissent le phénotype sauvage.

L'optimisation de la séquence polypyrimidique engendre une augmentation de l'épissage au site 3'*src* qui se traduit par une surproduction de l'ARNm *src*. De ce fait, la quantité d'ARNm *src* est quatre fois plus importante que la normale et elle s'accompagne d'une diminution des quantités de pré-ARNm et d'ARNm *env* (respectivement quatre et deux fois moins importantes). Ces défauts d'épissage déséquilibrent la balance en faveur des ARN épissés au détriment des ARN non épissés, ce qui engendre un défaut de multiplication du virus, qui se réplique deux fois plus lentement que le virus sauvage (Zhang et Stoltzfus, 1995).

Chez les virus révertants obtenus après plusieurs cycles d'infections des cellules CEF, le phénotype sauvage est restauré par un rétablissement de la balance entre les ARN épissés et non épissés (Zhang et Stoltzfus, 1995). Cette réversion phénotypique est due à des délétions qui diminuent le taux d'utilisation du site 3'*src*. Ces délétions concernent la séquence de la boîte de branchement, une partie de la séquence polypyrimidique ou la totalité du gène *src* et le site 3'*src* qui sont délétés par recombinaison homologue entre les séquences répétées appelées séquences dr1 qui se situent de part et d'autre du gène *src* (Zhang *et al.*, 1996 ; Zhang et Stoltzfus, 1995).

Comme cela a été observé pour l'étude du site 3'*env*, l'apparition rapide des révertants illustre, là aussi, le fait que le site 3'*src* est soumis à une forte pression de sélection afin de conserver une faible efficacité. Le caractère sous-optimal de la séquence polypyrimidique constitue donc un élément de régulation essentiel de l'épissage au site 3'*src*. Comme nous le

verrons ci-dessous, l'utilisation du site 3'*src* est également soumise à des régulations par d'autres séquences en *cis*, dont les effets s'ajoutent au caractère intrinsèquement faible de ce site accepteur.

### III.4. Le site 3'*src* est inhibé par l'élément SSS (Src Suppressor of Splicing)

L'élément SSS (pour Suppressor of src Splicing) est une séquence de 23 nt de long (nt 6865-6887), située dans la région intergénique séparant les gènes *env* et *src* (**Figure IV-6**) (Amendt *et al.*, 1995 ; McNally et Beemon, 1992). Dans le contexte viral, cet élément inhibe spécifiquement l'épissage au site 3'*src* et sa troncation totale ou partielle provoque une augmentation de l'épissage à ce site ainsi qu'une diminution de la quantité du pré-ARNm viral (Berberich et Stoltzfus, 1991b). La régulation négative exercée par la séquence SSS s'ajoute au caractère sous-optimal du site 3'*src* lié à sa séquence polypyrimidique non canonique (McNally et Beemon, 1992). La fonction de l'élément SSS est indépendante du contexte viral ou de la force du site 3' dont il inhibe l'épissage. En contexte hétérologue, lorsqu'il est placé dans l'intron d'un mini-gène en amont du site accepteur, l'élément SSS inhibe aussi bien l'épissage du site 3' fort du troisième exon du gène c-*myc* que celui du site 3'*src* sauvage ou optimisé par une séquence polypyrimidique canonique (McNally et Beemon, 1992 ; Zhang *et al.*, 1996).

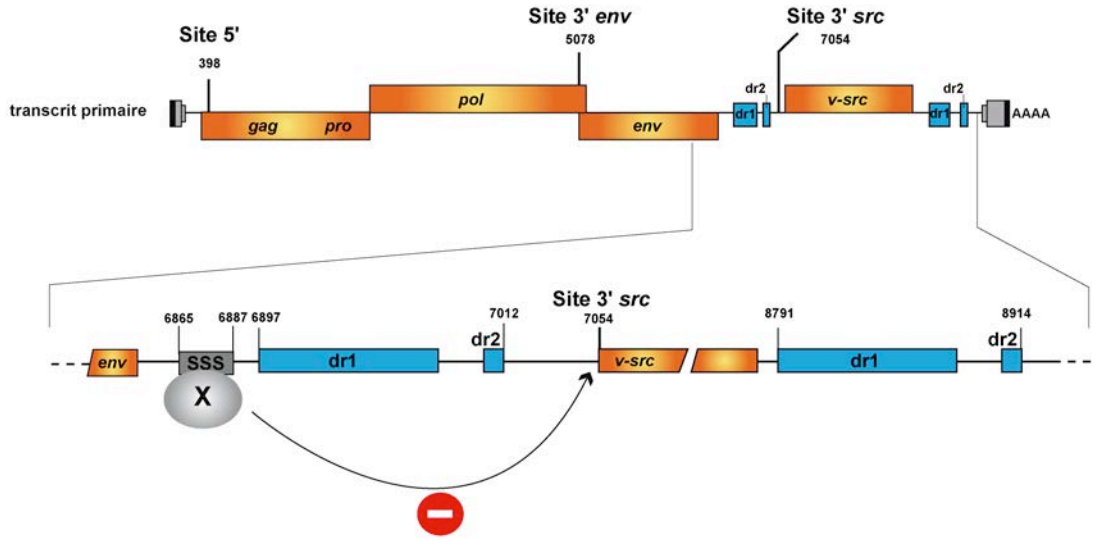



**Figure IV-6 : L'épissage au site 3'*src* est inhibé par l'élément SSS.**
Modèle proposé pour l'inhibition de l'épissage au site 3'*src* par l'élément SSS. Un ou des facteurs (X) spécifiquement présents dans les noyaux des cellules aviaires interagiraient avec l'élément SSS pour inhiber l'épissage au site 3'*src*.
Le transcrit primaire du virus RSV et un détail de la région contenant le site accepteur 3' *src* sont représentés. Les gènes viraux sont représentés par des rectangles orange et les séquences dr1 et dr2 encadrant le gène v-*src* par des rectangles bleus. Les sites d'épissage et leurs positions sont mentionnés.

Le mécanisme d'inhibition de l'épissage par l'élément SSS n'est pas connu mais il n'a lieu que lorsque les expériences *in cellulo*/*in vitro* sont réalisées dans des cellules aviaires ou avec des extraits nucléaires obtenus à partir de ces cellules (Amendt *et al.*, 1995). En effet, en présence ou en absence de l'élément SSS le même taux d'épissage au site 3'*src* est obtenu lorsque les transfections sont réalisées avec des fibroblastes humains (Amendt *et al.*, 1995). Dans les deux cas, l'épissage au site 3'*src* est trois fois plus important que celui mesuré lors des mêmes expériences réalisées avec des cellules CEF (Amendt *et al.*, 1995). Les mêmes résultats sont observés *in vitro* avec un mini-gène. L'absence d'inhibition au site 3'*src* par l'élément SSS lorsque l'épissage est réalisé avec un extrait nucléaire de cellules HeLa est rétabli lorsque cet extrait est complémenté avec un extrait nucléaire de fibroblastes de poulet (Amendt *et al.*, 1995). Il a été suggéré que l'élément SSS inhibe l'épissage au site 3'*src* en interagissant avec un ou plusieurs facteurs cellulaires spécifiques des noyaux de cellules de poulet, qui sont absents dans des fibroblastes humains. L'identité de ce (ces) facteur(s) et son (leur) mécanisme d'action reste à être déterminés.

Cette hypothèse pourrait expliquer la raison pour laquelle différents profils d'épissage du virus RSV sont observés lorsqu'il est présent dans des cellules de mammifères ou des cellules de poulet. En effet, dans des cellules CEF, les ARNm *env* et *src* sont produits en quantité équimolaires et représentent respectivement environ 13 et 12% de la totalité des ARN viraux (Berberich *et al.*, 1990), tandis que dans les cellules NIH3T3 ou HFF (fibroblastes murins et humains), l'ARNm *src* représente 50 à 60% de la totalité des ARN viraux et très peu d'ARNm *env* est produit (Amendt *et al.*, 1995 ; Berberich *et al.*, 1990). La perte de la régulation de l'épissage au site 3'*src* dans les cellules humaines pourrait s'expliquer par l'absence du ou des facteurs se fixant sur l'élément SSS en *trans*. Ceci pourrait aussi expliquer en partie le caractère non permissif des cellules humaines pour le virus RSV.

Jusqu'à présent, l'élément SSS a été identifié dans les souches Prague A et Prague C du virus RSV. Bien que les deux éléments SSS de chacune de ces souches ne diffèrent que par une substitution de base, l'élément SSS du virus de la souche Prague C réprime davantage l'épissage au site 3'*src* que celui de la souche Prague A (Stoltzfus *et al.*, 1987). Lors de l'infection des cellules CEF, le ratio entre l'ARNm *src* et le pré-ARNm de la souche Prague C est deux fois moins important que celui mesuré pour la souche Prague A. Aucune donnée n'est disponible sur la répercussion de cette substitution de base sur l'activité de l'élément SSS.

### III.5. Le site 3'*src* est inhibé par les éléments CTE (Constitutive Transport Element)

Le virus RSV comporte deux séquences répétées DR situées de part et d'autre du gène *src* (Ogert *et al.*, 1996) (**Figure IV-7**). La première répétition DR1 est localisée dans la région intergénique des gènes *env* et *src*, en aval de l'élément SSS, tandis que la seconde, DR2, se situe en aval du gène *src* dans la partie 3'UTR (Ogert *et al.*, 1996 ; Schwartz *et al.*, 1983) (**Figure IV-7**). Chacune des séquences DR est composée d'une séquence répétée dr1 d'une centaine de nucléotides et d'une seconde séquence répétée appelées dr2 de 36 nucléotides (**Figure IV-7**). Comme nous l'avons développé dans la partie introduction, ces séquences DR sont essentielles au bon déroulement du cycle viral (**Chapitre III, VI.2.4**). Elles jouent le rôle d'éléments CTE dont la fonction consiste à exporter les pré-ARNm viraux du noyau vers le cytoplasme (Ogert *et al.*, 1996 ; Simpson *et al.*, 1998 ; Yang et Cullen, 1999).

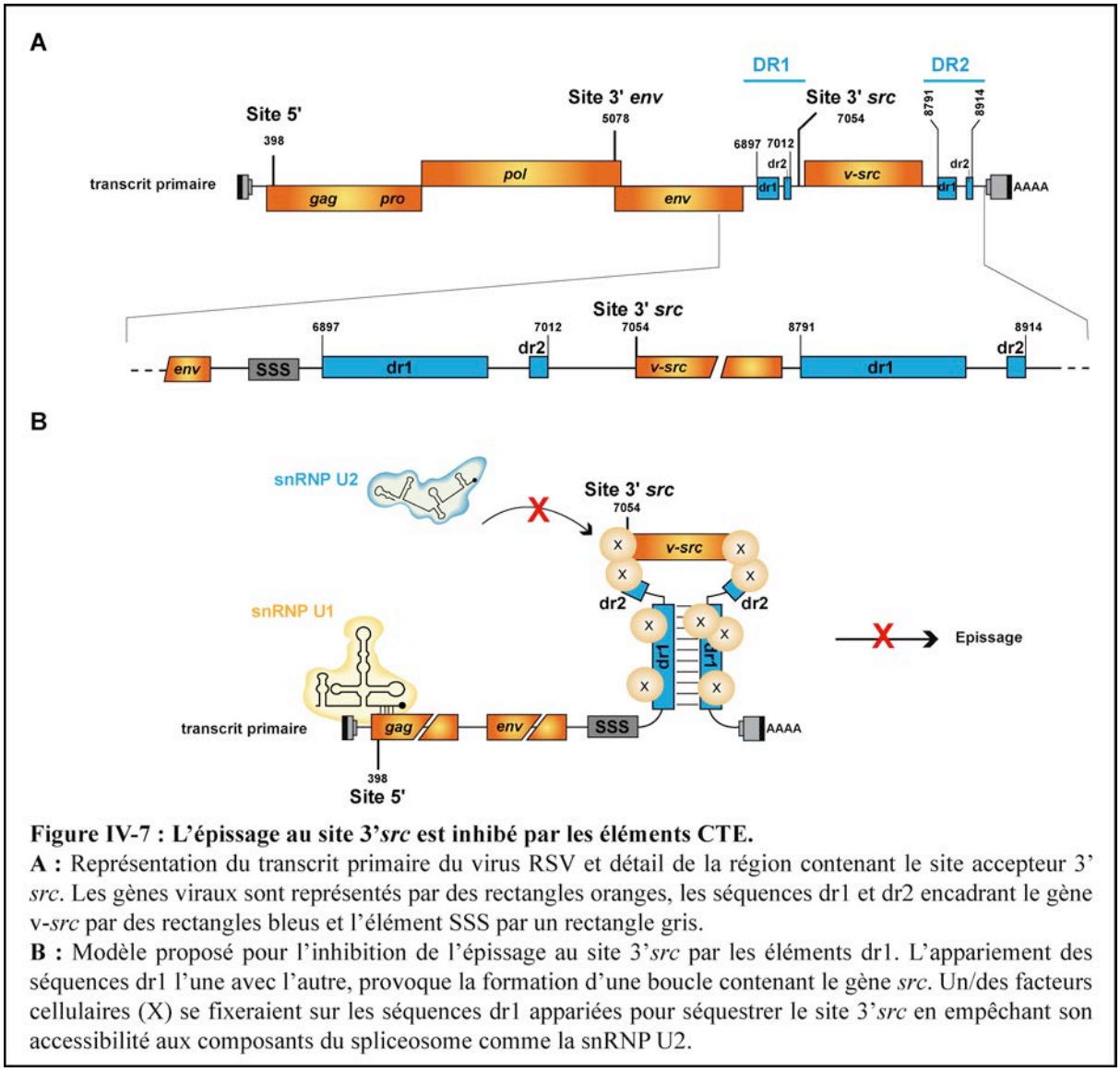



**Figure IV-7 : L'épissage au site 3'*src* est inhibé par les éléments CTE.**
**A :** Représentation du transcrit primaire du virus RSV et détail de la région contenant le site accepteur 3' *src*. Les gènes viraux sont représentés par des rectangles oranges, les séquences dr1 et dr2 encadrant le gène v-*src* par des rectangles bleus et l'élément SSS par un rectangle gris.
**B :** Modèle proposé pour l'inhibition de l'épissage au site 3'*src* par les éléments dr1. L'appariement des séquences dr1 l'une avec l'autre, provoque la formation d'une boucle contenant le gène *src*. Un/des facteurs cellulaires (X) se fixeraient sur les séquences dr1 appariées pour séquestrer le site 3'*src* en empêchant son accessibilité aux composants du spliceosome comme la snRNP U2.

En plus de leur fonction dans le transport des ARNm viraux, les éléments CTE inhibent l'épissage au site 3'*src* (Guo *et al.*, 2000 ; Simpson *et al.*, 1997). Cette inhibition est indépendante, spécifique et s'ajoute à celle exercée par l'élément SSS (Guo *et al.*, 2000). Pour cela, il est nécessaire que les deux éléments CTE soient fonctionnels et se situent de part et d'autre du gène *src*. En effet, les délétions ou les mutations d'un seul ou des deux éléments CTE provoquent une diminution de l'export des pré-ARNm, une augmentation spécifique de l'épissage du site 3'*src* ainsi qu'une diminution concomitante de la quantité de pré-ARNm (Guo *et al.*, 2000 ; Simpson *et al.*, 1997). L'activation de l'épissage au site 3'*src* est proportionnelle au degré d'inactivation des éléments CTE, ce qui montre qu'il existe un lien de corrélation entre la capacité des éléments CTE à exporter les pré-ARNm et à inhiber l'épissage au site 3'*src* (Guo *et al.*, 2000). L'analyse de l'ensemble des mutants des éléments CTE montre par ailleurs que ce sont les mutations de la séquence dr1 qui affectent le plus l'inhibition de l'épissage au site 3'src (Guo *et al.*, 2000).

Le mécanisme d'inhibition de l'épissage au site 3'*src* par les deux éléments CTE n'est pas encore compris. Il a été suggéré que les éléments CTE situés de part et d'autre du gène *src* s'apparient ensemble afin de confiner la séquence du gène *src* dans une boucle (Guo *et al.*, 2000) (**Figure IV-7**). Dans cette configuration, le ou les facteurs cellulaires interagissant en *trans* avec chacun des éléments CTE pourraient créer une gêne stérique limitant l'accessibilité des facteurs du spliceosome au site 3'*src* (**Figure IV-7**).

Jusqu'ici, nous avons décrit les sites d'épissage du virus RSV ainsi que la façon dont leur utilisation est régulée de façon à maintenir l'équilibre entre les ARN épissés et les ARN non épissés. Nous allons à présent décrire la séquence régulatrice qui a fait l'objet de notre étude, l'élément NRS, qui inhibe l'épissage du pré-ARNm viral.

## IV. L'élément NRS est une séquence inhibitrice de l'épissage du pré-ARNm du virus RSV

L'élément NRS est une séquence inhibitrice de l'épissage de 227 nt de long (nt 703-930) qui fait partie du gène *gag*. Il se situe 305 nt en aval du site 5' d'épissage (nt 397) et à plus de 4 000 et 6 000 nt en amont des sites accepteurs 3'*env* et 3'*src*, respectivement (Arrigo et Beemon, 1988 ; Schwartz *et al.*, 1983) (**Figure IV-8**). Cet élément, du fait de sa position intronique, n'est contenu que dans le pré-ARNm viral. Au sein de l'ORF de *gag*, l'élément NRS est situé à 400 nt en aval du codon initiateur de la traduction du pré-ARNm. Sa traduction correspond au sein du précurseur Pr76 Gag-Pro à un fragment de la protéine de la matrice, le peptide p2a et le peptide p2b plus une partie du peptide p10 qui contiennent le

domaine L qui est important pour le bourgeonnement (**Figure IV-8**). La séquence du NRS est donc à la fois une séquence importante pour la régulation de l'épissage et pour la traduction des protéines virales.

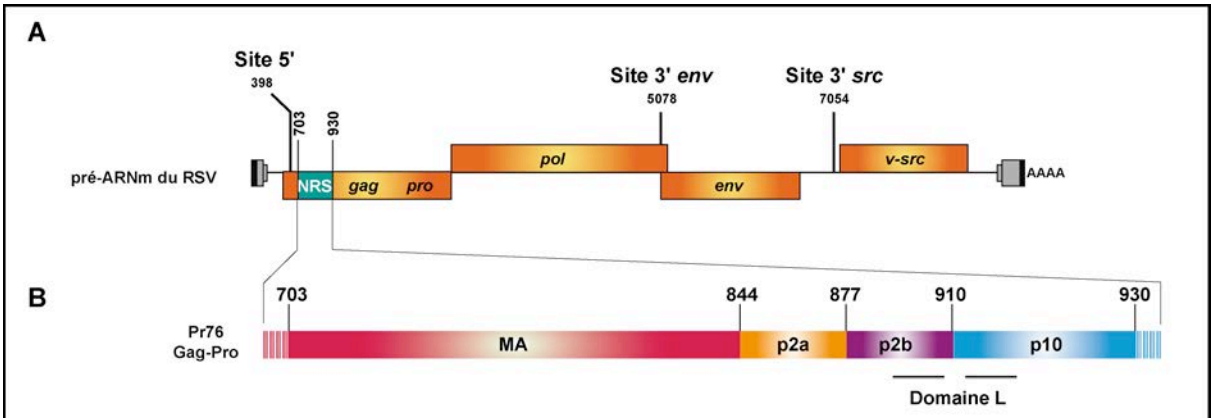



**Figure IV-8 : Localisation de l'élément NRS au sein du transcrit primaire du RSV.**
**A :** Représentation de l'élément NRS au sein du transcrit primaire du virus RSV. Les gènes sont représentés par des rectangles oranges, l'élément NRS par un rectange vert. Les sites d'épissage et le nom des gènes sont mentionnés.
**B** : Représentation de la partie du précurseur Pr76 Gag-Pro codée par la séquence de l'élément NRS. Les séquences des protéines de la matrice (MA), et des peptides p2a, p2b et p10 du précurseur Pr76 Gag-Pro sont représentés par des rectangles respectivement colorés en rose, orange violet et bleu. Les positions des nucléotides de la séquence codant ces peptides sont indiquées. Les séquences du domaine L importantes pour le bourgeonnement viral sont indiquées.

L'élément NRS a été découvert par l'équipe du Pr. K. Beemon (John Hopkins, Baltimore) qui a été la première à le décrire comme une séquence *cis* inhibitrice de l'épissage (Arrigo et Beemon, 1988). Cet élément a été découvert lorsqu'il a été observé que la troncation d'une partie du gène *gag* (nt 635-1010, fragment contenant l'élément NRS) du génome viral provoquait une augmentation des ARNm viraux et une diminution concomitante des pré-ARNm dans les cellules CEF transfectées (Arrigo et Beemon, 1988). D'autres expériences ont démontré que ce déséquilibre de la balance entre ARN épissés et ARN non épissés était indépendant des protéines virales car il était maintenu lorsque les protéines codées par les séquences qui étaient tronquées étaient apportées en *trans* par complémentation avec un virus helper (Stoltzfus et Fogarty, 1989).

L'inhibition de l'épissage par le NRS a été confirmée lors d'expériences d'épissage utilisant la démarche inverse qui consistait à démontrer que la présence de cette séquence issue du gène *gag* était suffisante pour inhiber l'épissage du pré-ARNm (Arrigo et Beemon, 1988). Pour cela, cette séquence a été insérée dans l'intron d'un mini-gène hétérologue (pRSV-Neo-Int) contenant le premier et le deuxième exon du gène cellulaire c-*myc* de poulet, tous deux partiellement tronqués et séparés par la séquence du deuxième intron de ce même gène,

également tronqué (Arrigo et Beemon, 1988) (**Figure IV-9**). La transfection transitoire de cette construction dans des cellules CEF a permis de démontrer que la séquence issue du gène *gag* inhibait l'épissage d'un pré-ARNm et ce, indépendamment du contexte viral et uniquement dans son orientation sens (Arrigo et Beemon, 1988).

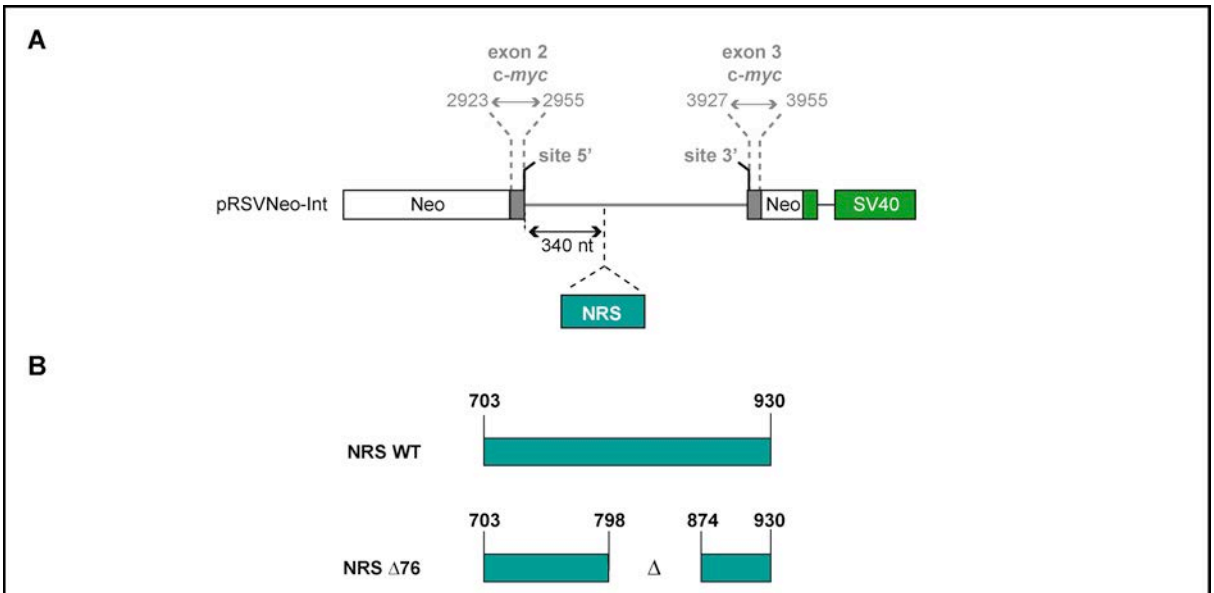



**Figure IV-9 : Définition de l'élément NRS WT et de l'élément NRS minimal fonctionnel NRS Δ76 en contexte hétérologue (Arrigo et Beemon, 1988 ; McNally *et al.*, 1991)**

**A :** Représentation de la construction d'épissage pRSVNeo-Int utilisée lors des transfections de cellules CEF pour étudier l'inhibition de l'épissage par l'élément NRS. Un mini-gène a été inséré dans le gène de la néomycine (Neo, rectangle blanc). Ce mini-gène comporte l'exon 2 et l'exon 3 partiellement tronqués du gène c-*myc* de poulet séparés par l'intron 2 également tronqué de ce même gène. Les exons sont représentés par des rectangles gris et l'intron par une ligne grise et les positions des sites donneur et accepteur sont mentionnés. Les différentes séquences étudiées de l'élément NRS (rectangle bleu) ont été insérées 340 nt en aval du site donneur dans l'intron. Les rectangles verts séparés par une ligne noire représentent la séquence contenant le signal de polyadénylation du virus SV40.

**B :** Représentation de l'élément NRS WT et de l'élément NRS minimal fonctionnel, NRS Δ76. Les positions des nucléotides sont indiquées, la troncation dans l'élément NRS Δ76 est symbolisée par un Δ.

Par la suite, des délétions moins importantes ont été réalisées dans le gène *gag* afin de délimiter de façon plus précise la séquence responsable de l'inhibition de l'épissage. L'élément NRS, appelé NRS WT, a ainsi été délimité à la séquence comprise entre les nucléotides 703-930 (McNally *et al.*, 1991) (**Figure IV-9**). Des troncations graduelles de cet élément NRS WT ont été réalisées et leur répercussion sur l'inhibition de l'épissage du pré-ARNm pRSVNeo-Int ont été mesurées pour définir les séquences minimales nécessaires à la conservation de l'activité inhibitrice de l'épissage par l'élément NRS (McNally *et al.*, 1991). Par rapport au NRS WT, la plus petite séquence NRS fonctionnelle est tronquée de 76 nt et comporte une séquence de 152 nt non contigus, constituée par les segments 703-798 et 875-930 (**Figure IV-9**). Cette séquence qui représente l'élément minimal fonctionnel a été appelé NRS Δ76 (McNally *et al.*, 1991)

Que ce soit dans le contexte viral ou le contexte hétérologue d'un mini-gène, les éléments NRS WT et Δ76 inhibent l'épissage uniquement lorsqu'ils sont présents dans leur orientation sens au sein du pré-ARNm (Arrigo et Beemon, 1988 ; McNally *et al.*, 1991). Leur activité inhibitrice est optimale lorsqu'ils sont situés à proximité du site 5', dans une fenêtre de 167 à 340 nt en aval du site 5' dont ils inhibent l'épissage (McNally *et al.*, 1991 ; Wang et McNally, 2003).

### IV.1. L'élément NRS inhibe l'épissage aux sites 3'*src* et 3'*env*

L'élément NRS inhibe l'épissage à chacun des sites 3'*env* et 3'*src*. Dans le contexte viral, la délétion 544-1806, qui comprend l'élément NRS WT, provoque une augmentation de l'épissage de deux fois à chacun des sites 3'*env* et 3'*src* (Stoltzfus et Fogarty, 1989). D'autres études montrent que des mutations inactivant le NRS provoquent une activation du site 3'*src* supérieure à celle du site 3'*env* (O'Sullivan *et al.*, 2002). Cela indique que dans le contexte viral, l'inhibition de l'épissage par le NRS semble davantage affecter le site 3'*src* que le site 3'*env*. Par ailleurs, l'inhibition de l'épissage au site 3'*src* est indépendante et s'ajoute à celle exercée par l'élément SSS (Stoltzfus et Fogarty, 1989). Tous ces résultats obtenus dans le contexte viral ont été confirmés par des expériences d'épissage *in cellulo* en contexte hétérologue avec un mini-gène (McNally et Beemon, 1992).

Ainsi, l'élément NRS peut inhiber l'épissage de n'importe quel site 3', qu'il soit fort, comme c'est le cas pour le site 3' de l'exon 2 du gène c-*myc* ou qu'il soit faible comme c'est le cas des sites 3'*env* et 3'*src*, et ce quelles que soient les causes de la faible efficacité du site 3' (boîte de branchement sous-optimale ou une séquence polypyrimidique sous-optimale).

## V. Le NRS est un élément bipartite

L'élément NRS WT (nt 703-930) peut être découpé en trois régions : une région riche en purine, une région centrale non essentielle à l'activité inhibitrice de l'épissage et une région 3' riche en pyrimidines (McNally *et al.*, 1991) (**Figure IV-10**). Les régions riches en purine (nt 703-798) et en pyrimidines (nt 874-930) sont respectivement appelées NRS 5' et NRS 3' (McNally *et al.*, 1991) (**Figure IV-10**). Ces deux séquences sont essentielles à l'activité inhibitrice de l'élément NRS et doivent être toutes les deux présentes pour conférer à l'élément NRS son activité inhibitrice (McNally *et al.*, 1991). Elles sont séparées par une séquence non essentielle qui peut être tronquée sans que cela n'affecte la capacité d'inhibition de l'épissage comme l'atteste l'activité de l'élément NRS Δ76 qui est l'élément NRS minimal fonctionnel dans lequel cette séquence a été tronquée (McNally *et al.*, 1991). La partie NRS

3' contient la séquence d'un pseudo site donneur d'épissage qui fixe la snRNP U1 et confère l'activité inhibitrice de l'épissage au NRS (**Figure IV-10**). La partie NRS 5' contient une séquence ESE dont le rôle est de favoriser la fixation de la snRNP U1 sur la partie NRS 3') (McNally, 2008) (**Figure IV-10**.

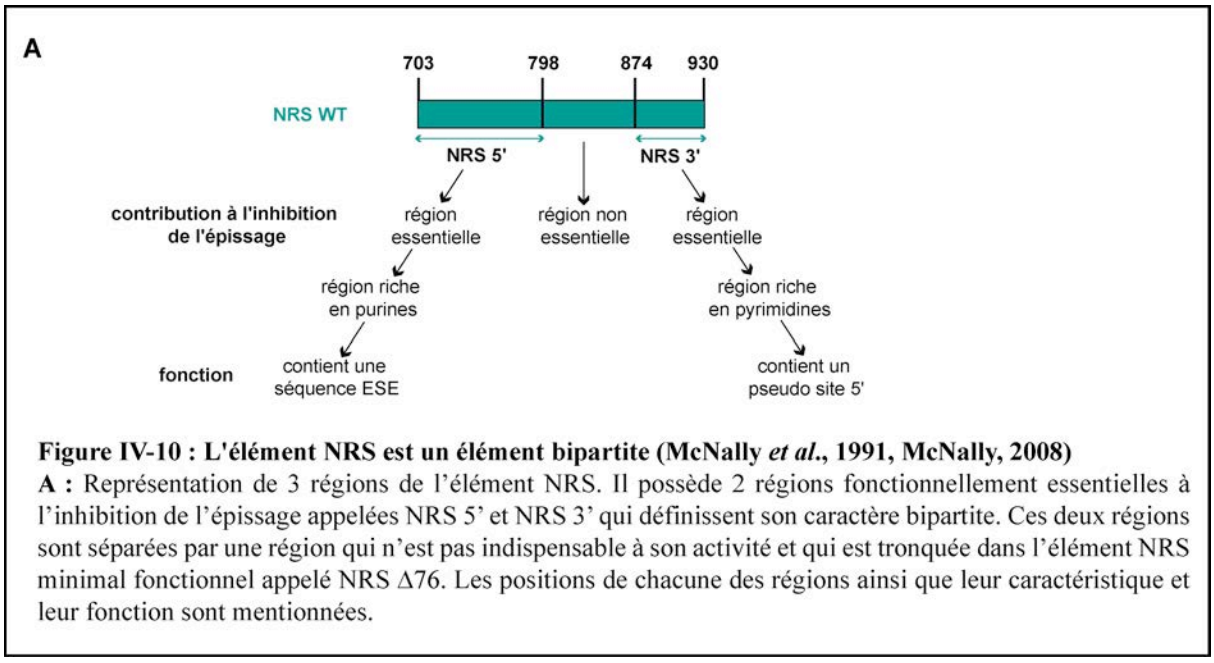



**Figure IV-10 : L'élément NRS est un élément bipartite (McNally *et al.*, 1991, McNally, 2008)**
**A :** Représentation de 3 régions de l'élément NRS. Il possède 2 régions fonctionnellement essentielles à l'inhibition de l'épissage appelées NRS 5' et NRS 3' qui définissent son caractère bipartite. Ces deux régions sont séparées par une région qui n'est pas indispensable à son activité et qui est tronquée dans l'élément NRS minimal fonctionnel appelé NRS Δ76. Les positions de chacune des régions ainsi que leur caractéristique et leur fonction sont mentionnées.

## VI. Principe de l'inhibition de l'épissage du pré-ARNm du virus RSV par l'élément NRS

Lors de l'épissage du pré-ARNm du RSV, l'unique site donneur fixe la snRNP U1 et s'associe avec l'un des sites donneurs 3'env ou 3'src pour former des complexes spliceosomaux fonctionnels dont la catalyse produit les ARNm *env* ou *src* (**Figure IV-11, A**). L'élément NRS contient un pseudo site donneur qui n'est pas fonctionnel en épissage et qui joue un rôle de leurre (Hibbert *et al.*, 1999 ; Paca *et al.*, 2001). Pour qu'il y ait inhibition de l'épissage, le pseudo site de cet élément doit entrer en compétition avec le site donneur fonctionnel pour la fixation de la snRNP U1. Sa similarité de séquence avec un site donneur fonctionnel lui permet d'être reconnu comme tel par le spliceosome et par conséquent de fixer la snRNP U1. Après cela, il va ensuite s'associer avec un des sites 3' pour former un complexe spliceosomal abortif qui ne peut pas catalyser la réaction d'épissage (**Figure IV-11, B**) (McNally, 2008).

Une des clés de la mise en place de l'inhibition de l'épissage par le NRS est donc la fixation de la snRNP U1 sur le pseudo site. Cette fixation est soumise à des régulations positives et négatives qui mettent en jeu les parties NRS 5' et NRS 3' que nous allons à présent aborder.

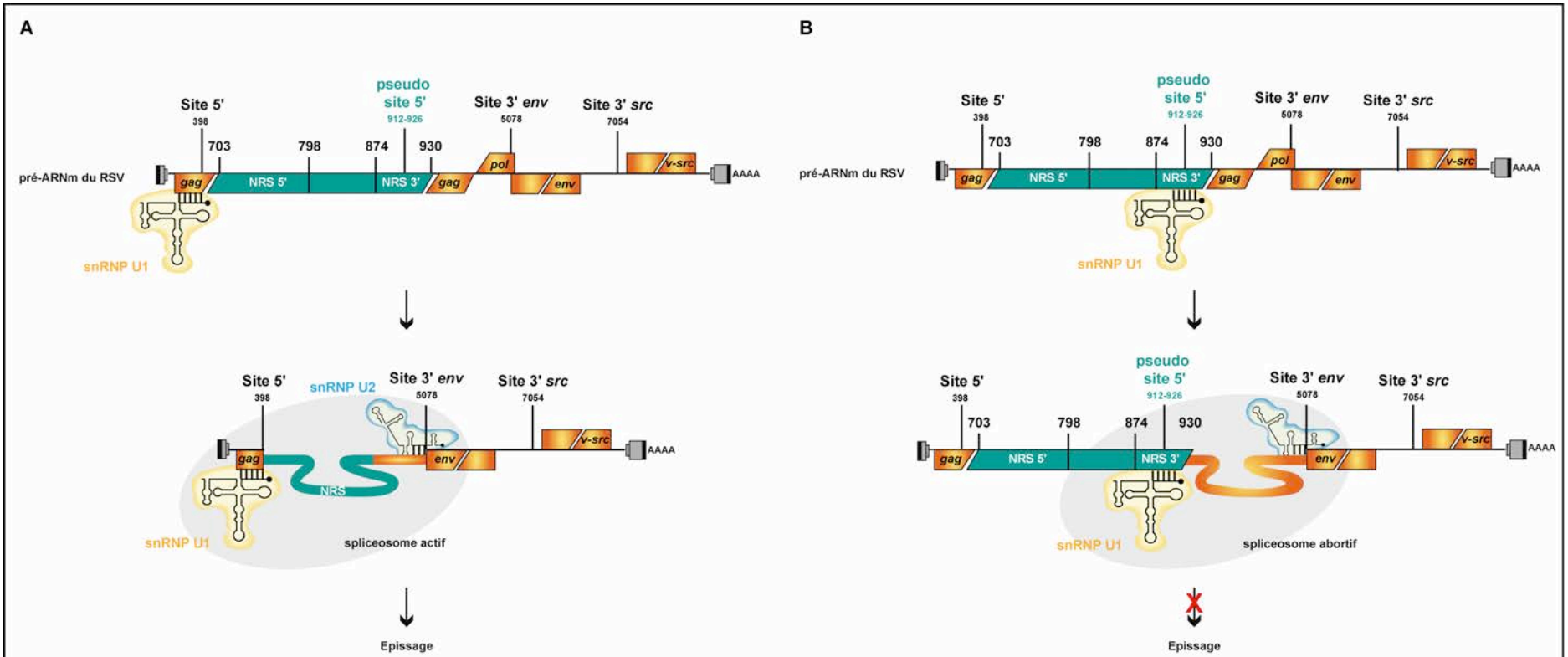



**Figure IV-11 : Modèle d'inhibition de l'épissage du pré-ARNm du RSV par l'élément NRS (d'après McNally, 2008).**
**A :** Le site donneur du pré-ARNm du virus RSV fixe la snRNP U1 et s'engage dans un complexe spliceosomal fonctionnel avec le site 3'*env*, ce qui conduit à l'épissage.
**B :** Le pseudo site donneur du NRS qui lui, n'est pas fonctionnel fixe la snRNP U1 et s'engage dans un complexe spliceosomal abortif avec le site 3'*env* qui ne peut pas catalyser l'épissage.
Le même modèle d'inhibition de l'épissage par le NRS s'applique également au site 3'*src*.

# VII. Description du pseudo site de l'élément NRS

## VII.1. Le pseudo site contient deux sites donneurs chevauchants qui fixent respectivement les snRNP U1 et U11

Le pseudo site donneur est une séquence de 14 nucléotides de long (nt 912-926) située dans la partie NRS 3' et qui comporte deux sites donneurs chevauchants (Gontarek *et al.*, 1993 ; Hibbert *et al.*, 1999 ; McNally et McNally, 1999) (**Figure IV-12**). L'un des sites fixe la snRNP U1 et l'autre la snRNP U11. Aucun de ces sites n'est fonctionnel en épissage lorsqu'ils sont couplés à un site accepteur U2-dépendant (Hibbert *et al.*, 1999 ; Paca *et al.*, 2001). En raison du chevauchement de leur site de fixation, la fixation des snRNP U1 et U11 au pseudo site est exclusive. Seule la fixation de la snRNP U1 permet la mise en place de l'inhibition de l'épissage par le NRS (Hibbert *et al.*, 1999 ; McNally et McNally, 1999). Par conséquent, la fixation en premier de la snRNP U11 empêche par la suite la fixation de la snRNP U1 et donc la mise en place de l'inhibition. Nous commencerons par décrire précisément chacun de ces sites puis nous aborderons les facteurs régulant la fixation des snRNP à ces sites.

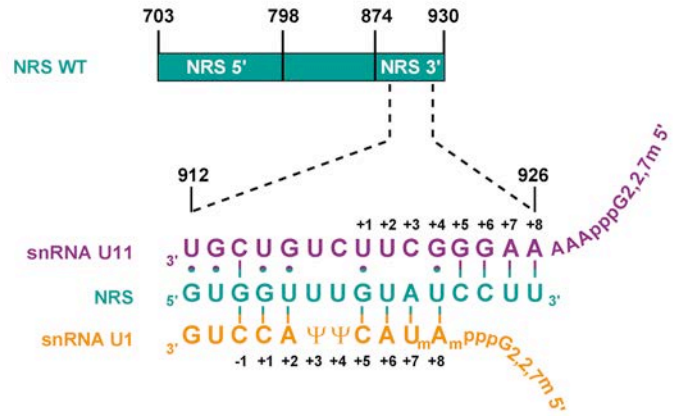



**Figure IV-12 : Le pseudo site de l'élément NRS contient deux sites donneurs chevauchants (d'après Hibbert *et al.*, 1999).**
Représentation des deux sites donneurs chevauchants du pseudo site d'épissage de l'élément NRS et de leurs appariements avec le snRNA U1 et le snRNA U11. La séquence du pseudo site est en bleu, celle du snRNA U1 en orange et celle du snRNA U11 en violet. Les positions des nucléotides de chacun des snRNA par rapport au site donneur auquel ils sont appariés sont mentionnés.

## VII.2. Le pseudo site s'apparie avec le snRNA U1

La séquence du site de fixation de la snRNP U1 du pseudo site 5' de l'élément NRS est similaire à celle d'un site donneur fonctionnel. La séquence du pseudo site forme 5 appariements de base sur les 8 appariements possibles établis par la séquence consensus du

site donneur U2-dépendant (**Figure IV-13**). Ses nucléotides aux positions -1, +1, +2, +5 et +6 sont canoniques tandis que les uridines aux positions -2, +3 et +4 ne le sont pas et ne peuvent pas former d'appariement avec le snRNA U1 (**Figure IV-13**). En revanche, les nucléotides aux positions +7 et +8, habituellement non impliqués, s'apparient également avec le snRNA U1, ce qui porte le nombre d'appariement du pseudo site avec le snRNA U1 à un total de 7 (**Figure IV-13**).

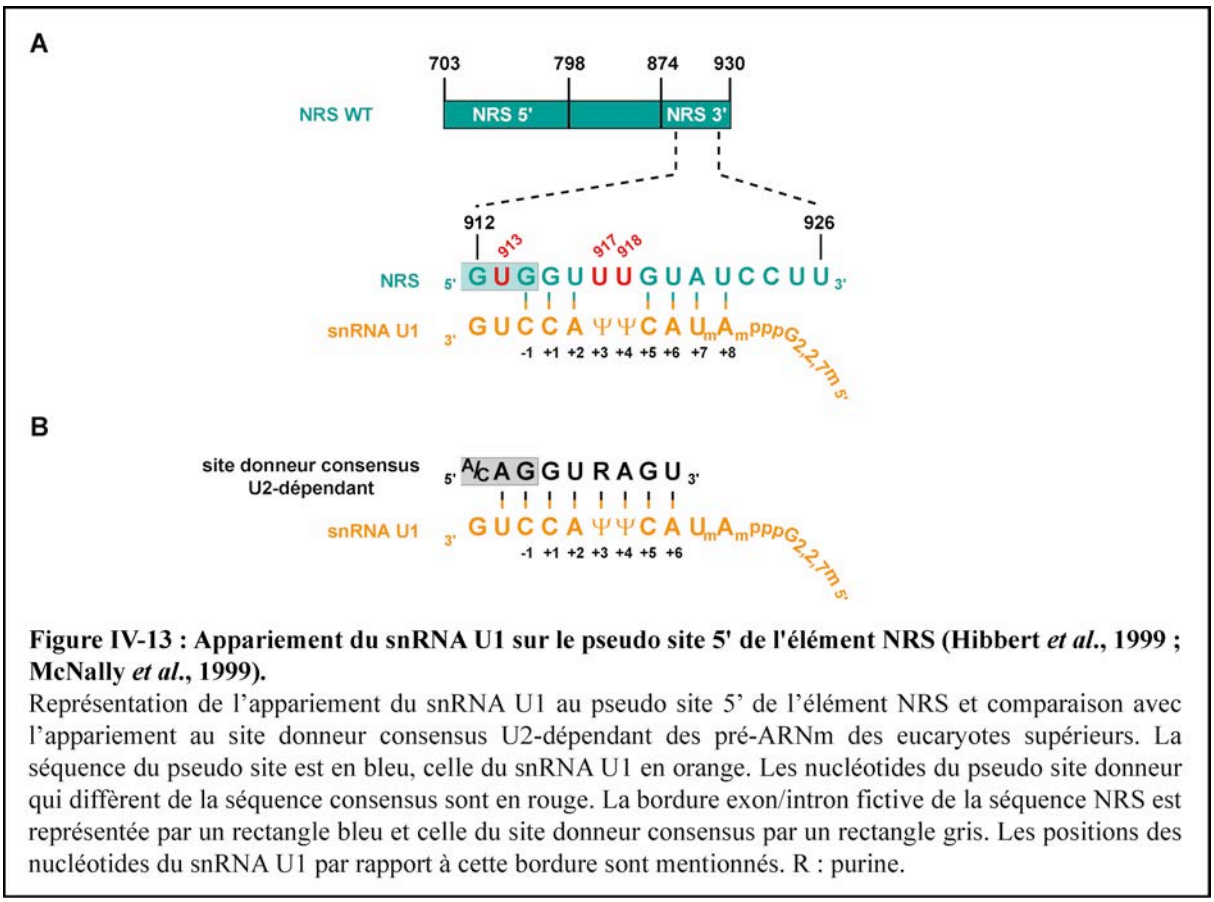



**Figure IV-13 : Appariement du snRNA U1 sur le pseudo site 5' de l'élément NRS (Hibbert *et al.*, 1999 ; McNally *et al.*, 1999).**
Représentation de l'appariement du snRNA U1 au pseudo site 5' de l'élément NRS et comparaison avec l'appariement au site donneur consensus U2-dépendant des pré-ARNm des eucaryotes supérieurs. La séquence du pseudo site est en bleu, celle du snRNA U1 en orange. Les nucléotides du pseudo site donneur qui diffèrent de la séquence consensus sont en rouge. La bordure exon/intron fictive de la séquence NRS est représentée par un rectangle bleu et celle du site donneur consensus par un rectangle gris. Les positions des nucléotides du snRNA U1 par rapport à cette bordure sont mentionnés. R : purine.

L'inhibition de l'épissage exercée par l'élément NRS repose sur la reconnaissance de la séquence du pseudo site par la snRNP U1 (Hibbert *et al.*, 1999 ; McNally et McNally, 1999). Aussi de nombreuses études ont été menées afin d'identifier les nucléotides du pseudo site 5' qui sont importants pour la fixation de la snRNP U1. La conclusion de ces études démontre que la complémentarité de séquence du pseudo site 5' avec le snRNA U1 est essentielle à l'activité de l'élément. Il existe une corrélation directe entre le nombre d'appariement établi entre le snRNA U1 et le pseudo site 5' et l'activité inhibitrice de l'élément NRS (McNally et McNally, 1999 ; Polony *et al.*, 2003).

Ainsi, l'abolition des appariements avec snRNA U1 par la mutation individuelle des nucléotides canoniques du pseudo site aux positions -1, +1, +2, +5 ainsi que ceux aux positions +7 et +8 provoquent une diminution de moitié de l'inhibition de l'épissage par le

NRS (Hibbert *et al.*, 1999 ; Polony *et al.*, 2003). Cette perte d'activité est compensée lorsqu'on génère dans le snRNA U1 une mutation compensatoire à celle de la mutation du pseudo site 5' à la position +5 (Hibbert *et al.*, 1999). Des résultats similaires sont obtenus dans le contexte viral où la diminution du nombre d'appariement du pseudo site 5' avec le snRNA U1 provoque une perte de l'activité inhibitrice de l'élément NRS qui se traduit ensuite par un retard de réplication du virus (Hibbert *et al.*, 1999 ; O'Sullivan *et al.*, 2002).

Par ailleurs, la mutation en adénosines de chacune des trois uridines non canoniques du pseudo site 5' aux positions -2, +3 et +4 convertissent le pseudo site 5' en un site fonctionnel (McNally et McNally, 1999 ; Paca *et al.*, 2001).

Les propriétés du pseudo site 5' sont donc liées à sa différence de séquence par rapport aux sites 5' canoniques.

### VII.3. Le pseudo site s'apparie avec le snRNA U11

La séquence du site de fixation de la snRNP U11 du pseudo site 5' de l'élément NRS est identique à la séquence consensus du site donneur U12-dépendant (**Figure IV-14**). Les nucléotides aux positions 919 et 922 à 926 peuvent former 6 appariements avec le snRNA U11 aux positions +1, +4, +5, +6, +7 et +8 (**Figure IV-14**). Cet appariement est également étendu aux positions -3 à -7 qui ne sont pas habituellement requises pour la reconnaissance du site donneur par le snRNA U12 (**Figure IV-14**). Ce site donneur est fonctionnel en épissage lorsqu'il est couplé à un site accepteur U12-dépendant et il répond à toutes les caractéristiques d'un site donneur U12-dépendant (McNally et McNally, 1999). Ainsi, ce site n'est plus fonctionnel lorsqu'il est muté aux positions qui sont connues pour abolir l'épissage à un site U12-dépendant (mutations des positions +1 et +2, +6 et +7, +7 et +8) (McNally et McNally, 1999). La perte de l'épissage due aux mutations peut être restauré en partie lorsque le snRNA U11 comporte dans sa séquence les mutations compensatoires correspondantes (Kolossova et Padgett, 1997 ; McNally et McNally, 1999). Par ailleurs, les mutations qui abolissent l'utilisation du site en épissage entraînent également la perte totale de l'appariement du snRNA U11 avec le NRS (McNally et McNally, 1999 ; McNally *et al.*, 2004).

Cependant, le site donneur U12-dépendant du pseudo site du NRS n'a pas besoin d'être utilisé en épissage pour que le snRNA U11 s'y apparie. Seule la séquence du site, qui est fonctionnelle lorsqu'elle est transposée dans un contexte U12-dépendant, est suffisante pour la fixation de la snRNP U11 (Gontarek *et al.*, 1993 ; McNally et McNally, 1999 ; McNally *et al.*, 2004).

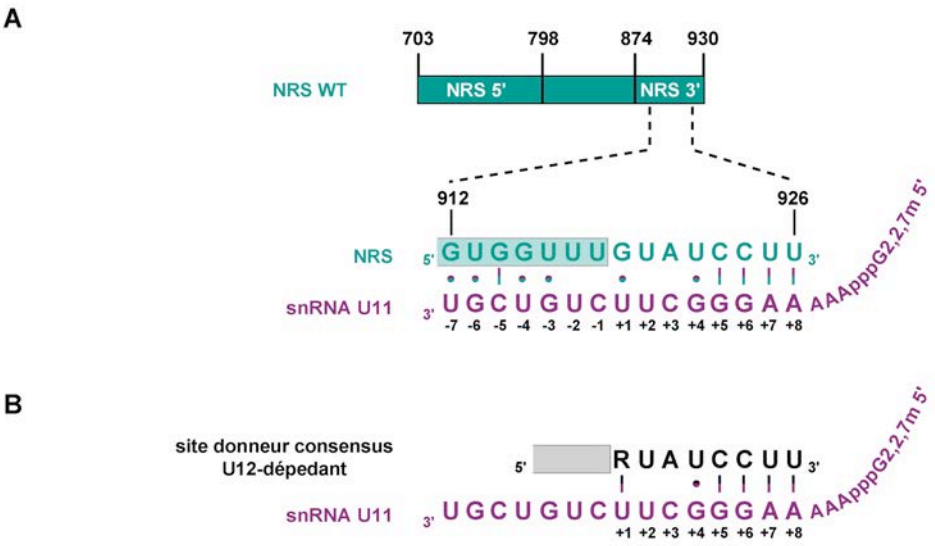



**Figure IV-14 : Appariement du snRNA U1 sur le pseudo site 5' de l'élément NRS (McNally *et al.*, 1999).** Représentation de l'appariement du snRNA U11 au pseudo site 5' de l'élément NRS et comparaison avec l'appariement au site donneur consensus U12-dépendant des pré-ARNm des eucaryotes supérieurs. La séquence du pseudo site est en bleu, celle du snRNA U11 en violet. La bordure exon/intron fictive de la séquence NRS est représentée par un rectangle bleu et celle du site donneur consensus par un rectangle gris. Les positions des nucléotides du snRNA U11 par rapport à cette bordure sont mentionnés. R : purine.

### VII.4. Compétition des snRNP U11 et U1 pour la fixation au pseudo site de l'élément NRS

Contrairement à la fixation de la snRNP U1, la fixation de la snRNP U11 n'est pas nécessaire pour la mise en place de l'inhibition de l'épissage par l'élément NRS. En effet, dans le pseudo site où leur site de fixation sont chevauchants, les mutations qui diminuent spécifiquement le nombre d'appariements au snRNA U11 tout en préservant l'intégrité de la séquence d'appariement du snRNA U1 induisent une augmentation de l'inhibition de l'épissage par le NRS (Hibbert *et al.*, 1999 ; McNally et McNally, 1999). Cette augmentation est inversement proportionnelle au nombre d'appariements du pseudo site formés avec le snRNA U11 (Hibbert *et al.*, 1999 ; McNally et McNally, 1999). Au contraire, les mutations spécifiques qui diminuent le nombre d'appariements du snRNA U1 tout en préservant l'intégrité de la séquence d'appariement du snRNA U11 induisent une diminution de l'inhibition de l'épissage (McNally et McNally, 1999).

Les snRNP U1 et U11 entrent donc en compétition pour la fixation à leur site dans le pseudo site du NRS et la snRNP U11 est un antagoniste de la snRNP U1 (McNally et McNally, 1999).

### VII.5. La snRNP U11 est recrutée par la hnRNP H

La fixation de la snRNP U11 sur le pseudo site de l'élément NRS est favorisée par la protéine hnRNP H qui se fixe à des séquences riches en résidus guanosines (McNally *et al.*,

2004, 2006). Ces séquences peuvent être regroupées en 4 éléments appelés G1, G2, G3 et G4 qui sont non contigus et se situent juste en aval de l'élément NRS WT (**Figure IV-15**).

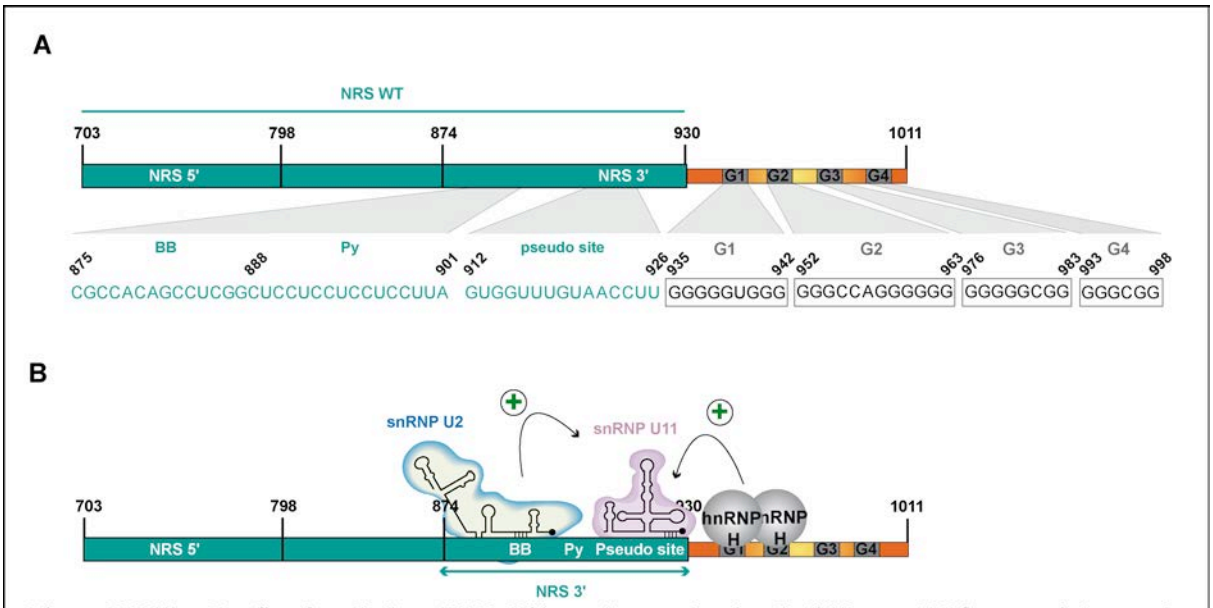



**Figure IV-15 : La fixation de la snRNP U11 sur le pseudo site de l'élément NRS est activée par la protéine hnRNP H (McNally *et al.*, 2004, McNally *et al.*, 2006).**
**A :** Représentation de l'élément NRS et des éléments G1, G2, G3, G4 riches en guanosines ainsi que la séquence riche en pyrimidines (Py) et la séquence ressemblant à la séquence de la boîte de branchement du gène de la calcitonine de rat (BB). L'élément NRS est représenté par un rectangle bleu, les séquences riches en guanosines par des rectangles gris et leurs positions sont indiquées. Les séquences Py, BB et des éléments G1, G2, G3 et G4 sont détaillées.
**B :** La fixation de la snRNP U11 sur le pseudo site de l'élément NRS est favorisée par la fixation de la protéine hnRNP H sur les éléments G1 et G2 ainsi que par celle de la snRNP U2 fixée sur la séquence BB.

Les délétions ou les mutations individuelles de chacun des éléments démontrent que les éléments G1 et G2 qui sont le plus proches du pseudo site du NRS fixent davantage la protéine hnRNP H et contribuent le plus à la fixation de la snRNP U11 sur le pseudo site (McNally *et al.*, 2004, 2006). La présence des éléments G1 et G2 provoquent également une augmentation de la fixation de la snRNP U11 sur le site U12-dépendant du pseudo site du NRS ainsi que son utilisation lors de l'épissage lorsque celui-ci est couplé dans un mini-gène à un site accepteur U12-dépendant (McNally *et al.*, 2004, 2006). Par ailleurs, la fixation de la snRNP U11 sur l'ARN de l'élément NRS seul est favorisée par la fixation de la snRNP U2 sur une séquence ressemblant à la séquence de boîte de branchement similaire à celle du gène de la calcitonine de rat suivie d'une séquence polypyrimidique située en amont du pseudo site (**Figure IV-15**) (Emeson *et al.*, 1989 ; McNally *et al.*, 2004). En revanche, sa troncation ne modifie pas l'utilisation du site U12-dépendant lors de l'épissage lorsque celui-ci est couplé dans un mini-gène à un site accepteur U12-dépendant (McNally *et al.*, 2004). Son site de fixation au sein de cette séquence n'est pas connu mais son interaction avait déjà été mise en

évidence auparavant par immunoprécipitation (Gontarek *et al.*, 1993 ; McNally *et al.*, 2004). Son rôle dans le recrutement de la snRNP U11 n'est pas connu et nécessite d'autres investigations.

Le recrutement de la snRNP U11 sur le pseudo site par la protéine hnRNP H et les éléments G1 et G2 n'a que très peu de répercussion sur le taux de fixation de la snRNP U1 sur le pseudo site. Dans ce contexte, les éléments G1 et G2 semblent plutôt jouer un rôle direct dans le recrutement de la snRNP U11 plutôt qu'un rôle indirect dans la compétition entre les snRNP U1 et U11 pour la fixation au pseudo site (McNally *et al.*, 2004).

Par ailleurs, la répercussion du recrutement de la snRNP U11 sur le pseudo site par la protéine hnRNP H et les éléments G1 et G2 n'a jamais été mesurée sur l'inhibition de l'épissage par le NRS. Le recrutement de la snRNP U11 par la protéine hnRNP H est un bon modèle pour expliquer comment la snRNP U11 entre en compétition avec la snRNP U1 alors que sa quantité dans le noyau est environ 100 fois moins importante que celle de la snRNP U1 (Pessa *et al.*, 2008 ; Tarn et Steitz, 1996). Aucune donnée n'existe cependant pour mesurer la contribution de ce recrutement de la snRNP U11 dans cette compétition.

## VII.6. La fixation de la snRNP U1 sur le pseudo site est activée par la séquence ESE de la partie NRS 5'

### VII.6.1. La partie NRS 5' est une séquence ESE

La partie NRS 5' située entre les nucléotides 703-797 est une séquence ESE. Cette fonction a été démontrée en utilisant un mini-gène hétérologue comprenant le 3$^{e}$ exon, le 3$^{e}$ intron et le 4$^{e}$ exon, tous partiellement tronqués, du gène *dsx* de drosophile. Dans cette construction, l'utilisation du site accepteur, qui est sous-optimal, est dépendant de la présence d'une séquence ESE (Tian et Maniatis, 1993). La partie NRS 5' placée en aval de ce site accepteur permet d'obtenir des produits d'épissage qui ne sont pas observés lorsqu'elle est remplacée par la partie NRS 3' (McNally et McNally, 1998). Les mêmes expériences réalisées avec différentes troncations de la partie NRS 5' ont permis de limiter les régions ESE de la partie NRS 5' aux séquences situées entre les nucléotides 715-748, 720-744 et 747-777 (**Figure IV-16**) (Fogel *et al.*, 2002 ; McNally et McNally, 1998). Comme toutes les séquences ESE, ces séquences ont la particularité de fixer les protéines SR.

#### VII.6.1.1. La partie NRS 5' fixe les protéines SR et la protéine hnRNP H

Les séquences ESE de la partie NRS 5' fixent plusieurs protéines de la famille des protéines SR ainsi que la protéine hnRNP H (Fogel *et al.*, 2002 ; Fogel et McNally, 2000 ; McNally et

McNally, 1996). L'ensemble des travaux montrent que la partie 5' du NRS fixe les protéines SR ASF/SF2, 9G8, SC35, SRp20, SRp40 et la protéine hnRNP H (Fogel *et al.*, 2002 ; Fogel et McNally, 2000 ; McNally et McNally, 1996). Les sites de fixation de ces protéines ont été déterminés par des expériences de pontage aux UV réalisées sur l'élément NRS entier, ou certains fragments de sa séquence, avec un extrait nucléaire de cellules HeLa et des protéines recombinantes purifiées, mais aussi par expérience de triple hybride dans la levure *S. cerevisiae* (Fogel et McNally, 2000 ; McNally et McNally, 1996) (**Figure IV-16**).

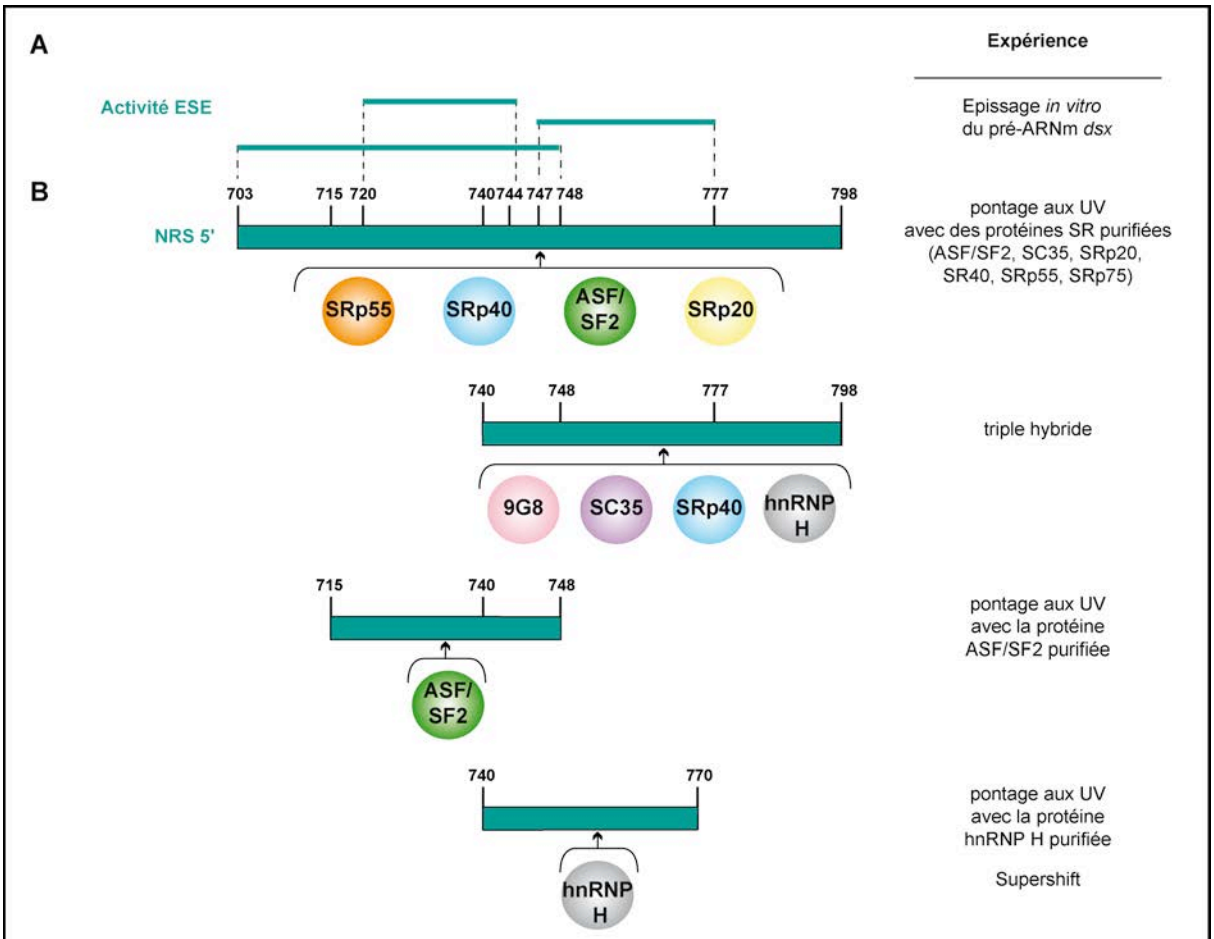



**Figure IV-16 : La partie NRS 5' contient plusieurs séquences ESE qui fixent les protéines SR et la protéine hnRNP H (McNally *et al.*, 1996, McNally *et al.*, 1998, Fogel *et al.*, 2000, Fogel *et al.*, 2002).**
**A :** Représentation schématique des séquences ESE contenues dans la partie NRS 5'. Les séquences de la partie NRS 5' ayant une activité ESE sont représentées par des traits verts, les positions des nucléotides sont reportées sur la séquence de la partie NRS 5'.
**B :** Représentation des différents fragments de l'élément NRS pour lesquels des interactions spécifiques avec des protéines SR ou la protéine hnRNP H ont été identifiées. La position des nucléotides et le nom des protéines sont reportés.
Les expériences utilisées dans chacun des cas sont mentionnées sur la droite.

Parmi les protéines identifiées, les protéines ASF/SF2 et hnRNP H ont été celles dont les sites de fixation ont été le plus étudiés. Ils correspondent respectivement aux séquences 715-748 et 740-770 (Fogel et McNally, 2000 ; McNally et McNally, 1996). La protéine SR SRp40 se fixe au fragment 740-798 et se fixerait également au fragment 703-798 lors de pontages

aux UV réalisés avec un mélange de protéines SR purifiées mais sa présence n'a pas été confirmée par une immunosélection des complexes ribonucléoprotéiques formés (Fogel *et al.*, 2002 ; McNally et McNally, 1996). La fixation de la protéine SC35 au fragment 740-798 a été démontrée par une expérience de triple hybride mais ce résultat est en contradiction avec les expériences de pontages aux UV réalisées sur le fragment 703-798 dont l'immunoprécipitation des complexes formés avec un anticorps spécifique contre cette protéine montrent qu'elle n'est pas pontée au NRS (Fogel *et al.*, 2002 ; McNally et McNally, 1996). La protéine SRp20 se fixerait au fragment 703-798 mais sa présence n'a pas été vérifiée par immunosélection (McNally et McNally, 1996). L'expérience de triple hybride a montré que la protéine 9G8 se fixe au fragment 740-798 et aucune autre donnée concernant cette protéine n'existe car elle n'a pas été utilisée lors des pontages aux UV réalisés avec le mélange de protéines SR purifiées (Fogel *et al.*, 2002 ; McNally et McNally, 1996).

Les expériences réalisées pour connaître les sites de fixation des protéines ne permettent pas de connaître le site précis de la fixation des protéines identifiées. Ces techniques présentent également des biais qui peuvent empêcher la mise en évidence d'une interaction, comme la capacité d'une protéine à être pontée aux UV. Parmi toutes les protéines identifiées, les protéines ASF/SF2 et hnRNP H sont celles dont les interactions directes sont formelles.

#### VII.6.1.2. L'activité ESE de la partie NRS 5' est essentielle pour l'inhibition de l'épissage par l'élément NRS

L'inhibition de l'épissage par le NRS nécessite la présence d'une séquence ESE dans la partie NRS 5'. Cette importance fonctionnelle de la partie NRS 5' a été étudiée *in cellulo* en contexte hétérologue dans le mini-gène du plasmide pRSV-NeoInt (**Figure IV-17**). Dans ce contexte, les troncations des séquences ESE 701-748 et 720-744 provoquent une diminution de l'inhibition de l'épissage par l'élément NRS (McNally et McNally, 1998). En revanche, la troncation de la séquence 701-715 n'induit aucune perte de l'activité de l'élément NRS (McNally et McNally, 1996). Ces données montrent qu'il existe une corrélation directe entre l'activité ESE de la région NRS 5' et l'efficacité d'inhibition de l'épissage du NRS et que la troncation de la séquence ESE à laquelle se fixe la protéine ASF/SF2, définie entre les nucléotides 715 et 748, inactive l'élément NRS. Cette séquence est donc essentielle à l'inhibition de l'épissage par l'élément NRS et suggère que la protéine ASF/SF2 est l'un des effecteurs principaux de l'activité ESE de la partie NRS 5' lors de l'inhibition de l'épissage par l'élément NRS (McNally et McNally, 1998).

Par ailleurs, la séquence NRS 5' peut être fonctionnellement remplacée par des séquences ESE hétérologues. Les séquences qui ont été testées sont la séquence ESE du gène de l'hormone de croissance bovine (FP) à laquelle se fixe la protéine ASF/SF2 et la séquence ESE de l'ASLV (ASLV) à laquelle se fixe majoritairement la protéine SRp40 et minoritairement la protéine ASF/SF2 (**Figure IV-17**) (Staknis et Reed, 1994 ; Sun *et al.*, 1993). L'activité inhibitrice du NRS est conservée lorsque chacune de ces séquences ESE est fusionnée à la partie du NRS 3' (McNally et McNally, 1998). Par conséquent, l'activité ESE de la partie 5' du NRS n'est pas uniquement dépendante de la fixation de la protéine ASF/SF2 pour activer l'inhibition de l'épissage par l'élément NRS (McNally et McNally, 1998). En revanche, l'utilisation d'une séquence ESE FP mutée qui ne fixe plus la protéine ASF/SF2 entraîne une perte de l'activité de l'élément NRS (McNally et McNally, 1998).

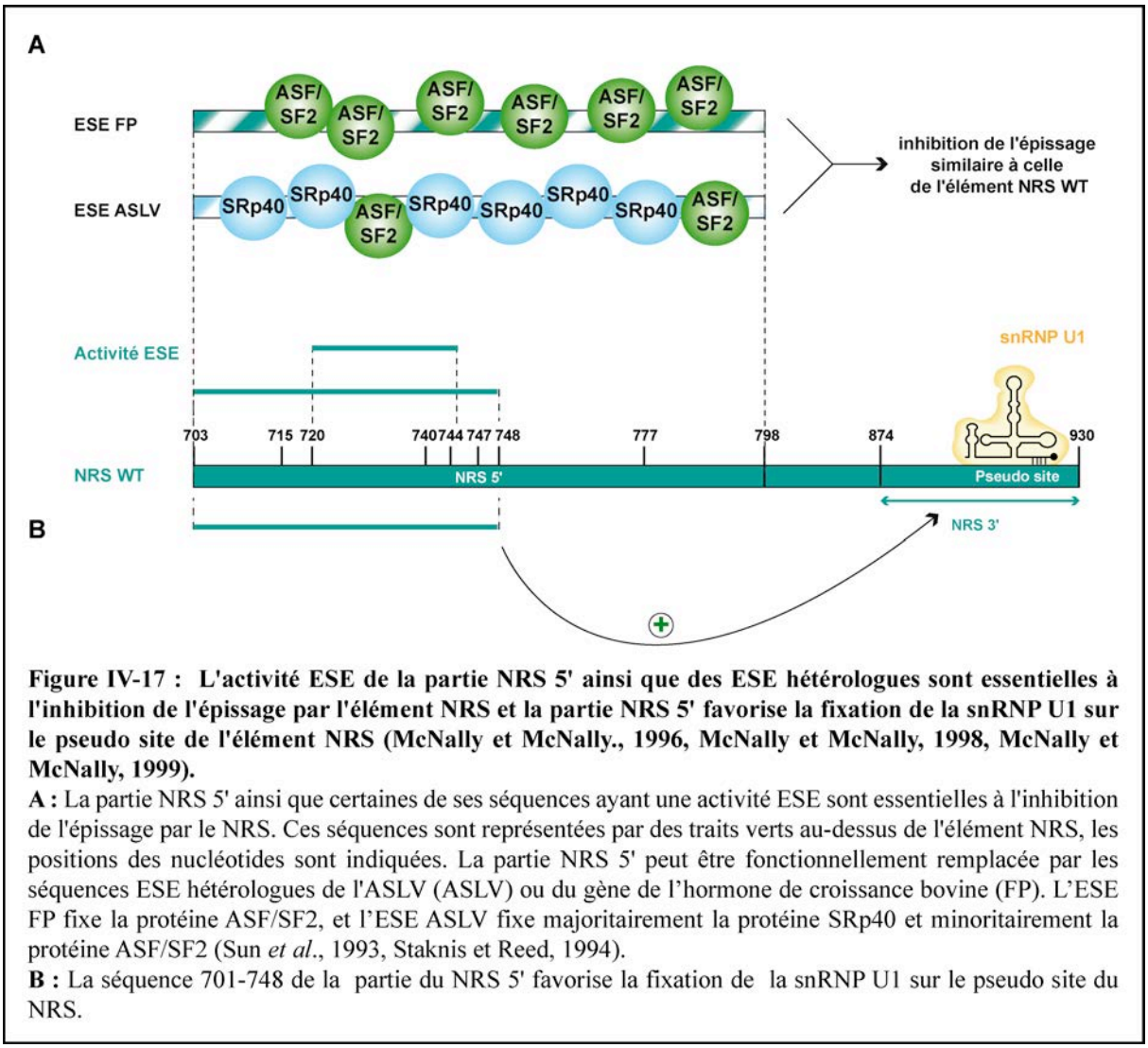



**Figure IV-17 : L'activité ESE de la partie NRS 5' ainsi que des ESE hétérologues sont essentielles à l'inhibition de l'épissage par l'élément NRS et la partie NRS 5' favorise la fixation de la snRNP U1 sur le pseudo site de l'élément NRS (McNally et McNally., 1996, McNally et McNally, 1998, McNally et McNally, 1999).**

**A :** La partie NRS 5' ainsi que certaines de ses séquences ayant une activité ESE sont essentielles à l'inhibition de l'épissage par le NRS. Ces séquences sont représentées par des traits verts au-dessus de l'élément NRS, les positions des nucléotides sont indiquées. La partie NRS 5' peut être fonctionnellement remplacée par les séquences ESE hétérologues de l'ASLV (ASLV) ou du gène de l'hormone de croissance bovine (FP). L'ESE FP fixe la protéine ASF/SF2, et l'ESE ASLV fixe majoritairement la protéine SRp40 et minoritairement la protéine ASF/SF2 (Sun *et al.*, 1993, Staknis et Reed, 1994).

**B :** La séquence 701-748 de la partie du NRS 5' favorise la fixation de la snRNP U1 sur le pseudo site du NRS.

Contrairement aux autres séquences ESE, la troncation de la séquence ESE 747-777 de l'élément NRS WT n'induit pas de perte d'activité de cet élément lors de l'inhibition de l'épissage (McNally et McNally, 1998). Cette séquence est inclue dans les fragments auxquels se fixent la protéine hnRNP H (nt 740-770) et les protéines SR 9G8, SRp40 et SC35 (nt 740-798) (Figure IV-). Cependant, la protéine hnRNP H ne semble pas être un antagoniste des protéines SR car les mutations de séquences qui l'empêchent de s'y fixer n'ont pas de répercussion sur sa fonction ESE (Fogel *et al.*, 2002). En revanche, les mutations qui diminuent le nombre de purines et qui théoriquement devraient empêcher la fixation des protéines SR à cette séquence inactivent sa fonction ESE. Par conséquent, la séquence 747-777 est une séquence ESE à laquelle se fixeraient des protéines SR qui contribuent peu à l'inhibition de l'épissage par l'élément NRS (Fogel *et al.*, 2002).

### VII.6.1.3. L'activité ESE de la partie NRS 5' favorise la fixation de la snRNP U1 au pseudo site

Une des fonctions de l'activité ESE de la partie 5' du NRS est de faciliter le recrutement de la snRNP U1 sur le pseudo site de l'élément NRS (**Figure IV-17**) (Fogel *et al.*, 2002 ; McNally et McNally, 1999). Les snRNA U1 et U11 se fixent à l'élément NRS WT lorsque celui-ci est incubé avec un extrait nucléaire de cellules HeLa (Gontarek *et al.*, 1993 ; McNally et McNally, 1999). La même expérience réalisée avec un extrait S100, qui ne contient pas de protéines SR, provoque une diminution de la fixation des snRNA U1 qui est restaurée lorsque des protéines SR purifiées sont ajoutées à l'extrait S100 (McNally et McNally, 1999). Cette restauration n'est possible que lorsque l'élément NRS comporte la région ESE 701-748, qui contient le site de fixation de la protéine ASF/SF2 (McNally et McNally, 1999). Le taux de fixation du snRNA U11 n'est en revanche ni affecté par la complémentation en protéines SR de l'extrait S100 ni par la troncation de la séquence 701-748. Seule la mutation spécifique de la séquence à laquelle s'apparie le snRNA U1 dans le pseudo site du NRS provoque une augmentation de sa fixation (McNally et McNally, 1999).

La perte de l'activité ESE de la séquence 747-777 provoque également une diminution de la fixation du snRNA U1 sur l'ARN NRS 740-1011 lorsqu'il est incubé en présence d'extrait nucléaire de cellules HeLa (Fogel *et al.*, 2002).

Ces données montrent qu'il existe une corrélation entre l'activité ESE de la partie NRS 5' et le taux de fixation du snRNA U1 sur le pseudo site. La fixation des protéines SR sur les différentes séquences ESE de la partie 5' du NRS favoriserait ainsi le recrutement du snRNA U1 plutôt que celui du snRNA U11 (Fogel *et al.*, 2002 ; McNally et McNally, 1999).

## VIII. L'élément NRS WT forme un complexe ribonucléoprotéique NRS-C avec les protéines de l'extrait nucléaire de cellules HeLa

L'élément NRS WT forme avec les facteurs de l'extrait nucléaire de cellules HeLa un complexe ribonucléoprotéique spécifique appelé complexe NRS-C (Cook et McNally, 1996). La formation de ce complexe est ATP indépendante et, contrairement aux complexes H aspécifiques, est limitée par la disponibilité des facteurs de l'extrait nucléaire (Cook et McNally, 1996). Il existe une corrélation entre la formation du complexe NRS-C et la capacité de l'élément NRS à inhiber l'épissage (Cook et McNally, 1996, 1998). Ainsi, la troncation de la séquence ESE 701-748 ou la mutation du site d'appariement du snRNA U1 qui inactivent l'élément NRS, empêchent la formation du complexe NRS-C (Cook et McNally, 1996). Par ailleurs, la protéine ASF/SF2 et le snRNA U1 sont des composants essentiels du complexe NRS-C. Aucun complexe NRS-C n'est formé lorsque l'élément NRS WT est incubé avec un extrait S100. L'ajout de la protéine ASF/SF2 purifiée à cet extrait suffit pour permettre la formation de ce complexe, ce qui n'est pas observé lors de l'ajout des protéines SC35 ou SRp40 (Cook et McNally, 1998). Par ailleurs, l'absence de la particule snRNP U1, mais pas celle de la snRNP U2, empêche la formation du complexe NRS-C (Cook et McNally, 1998).

Bien que les facteurs composants le complexe NRS-C ne soit pas tous connus, ce complexe s'apparente à un complexe d'engagement qui se forme au niveau d'un site donneur isolé (Cook et McNally, 1998 ; Michaud et Reed, 1993 ; Staknis et Reed, 1994).

## IX. L'élément NRS forme des complexes spliceosomaux abortifs lorsqu'il est associé à un site accepteur

Lors de l'inhibition de l'épissage, le pseudo site de l'élément NRS, assimilé à un site donneur fonctionnel, fixe la snRNP U1 puis s'associe avec un site accepteur avec lequel il forme un complexe spliceosomal abortif qui ne peut pas catalyser la réaction d'épissage (**Figure IV-18**) (Cook et McNally, 1999 ; Giles et Beemon, 2005). Cette inhibition a lieu avant la première attaque nucléophile et la présence de l'élément NRS comme site donneur au sein du complexe spliceosomal empêche la formation de complexe B catalytiquement actif (complexe B*).

La formation de ce complexe abortif a été étudiée *in vitro* en contexte hétérologue dans un pré-ARNm contenant l'élément NRS WT qui fait office de site donneur, le premier intron et le deuxième exon partiellement tronqués de l'unité tardive de l'adénovirus (Giles et Beemon,

2005). Après incubation avec un extrait nucléaire de cellules HeLa dans des conditions d'épissage, les complexes spliceosomaux ont été analysés par une ultracentrifugation en gradient de sucrose, électrophorèse en gel d'agarose et leur composition en snRNA a été analysée par Northern blot. L'élément NRS couplé au site 3' de l'adénovirus forme un complexe d'une taille de 50 S qui est instable et se dissocie en présence d'héparine en complexes de 25 S et 15 S qui s'apparentent respectivement aux complexes spliceosomaux A et E/H (Giles et Beemon, 2005). En revanche, aucun complexe de 35 S caractéristique des complexes spliceosomaux B/C n'est obtenu, ce qui est par ailleurs confirmé par l'analyse de ces complexes en gel d'agarose. Ces complexes contiennent les 5 snRNP majeures, snRNP U1, U2, U4, U5 et U6 (Giles et Beemon, 2005) (**Figure IV-18**).

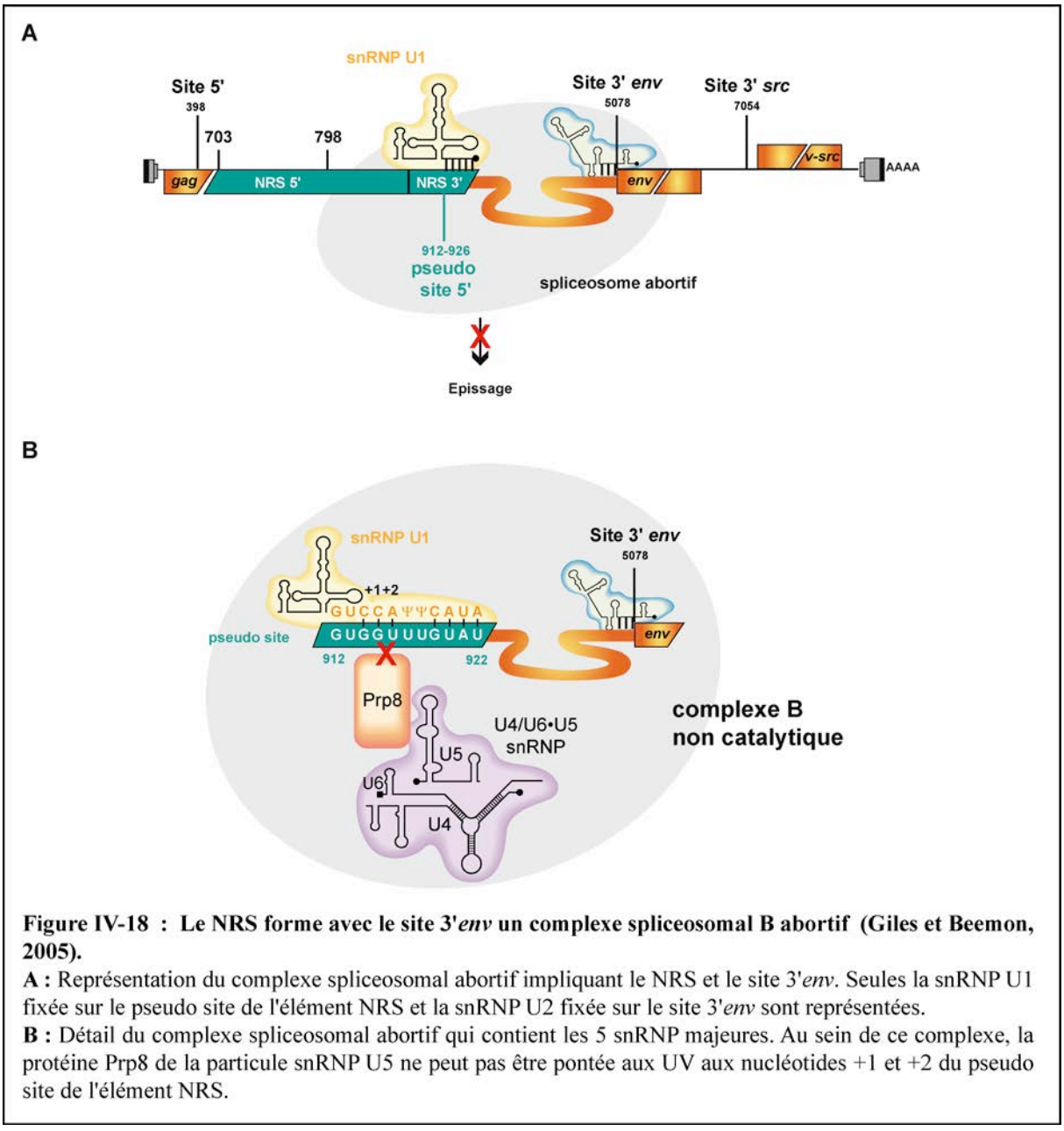



**Figure IV-18 : Le NRS forme avec le site 3'*env* un complexe spliceosomal B abortif (Giles et Beemon, 2005).**
**A :** Représentation du complexe spliceosomal abortif impliquant le NRS et le site 3'*env*. Seules la snRNP U1 fixée sur le pseudo site de l'élément NRS et la snRNP U2 fixée sur le site 3'*env* sont représentées.
**B :** Détail du complexe spliceosomal abortif qui contient les 5 snRNP majeures. Au sein de ce complexe, la protéine Prp8 de la particule snRNP U5 ne peut pas être pontée aux UV aux nucléotides +1 et +2 du pseudo site de l'élément NRS.

Ces complexes contiennent également la snRNP U11 mais sa présence n'est pas responsable de leur non-fonctionnalité car les complexes formés avec un extrait nucléaire ne contenant pas de snRNP U11 restent inactifs en épissage (Giles et Beemon, 2005). La présence de 5 snRNP indique que les évènements de remodelage des complexes spliceosomaux permettant de passer du complexe B au complexe B* n'ont pas lieu. Des expériences de pontage aux UV ont démontré qu'au sein de ces complexes, la protéine hPrp8 n'était pas pontée aux positions +1 et +2 du pseudo site d'épissage du NRS alors qu'un pontage de cette protéine est obtenu lorsque le complexe spliceosomal est fonctionnel (**Figure IV-18**). Cette donnée suggère que la tri-snRNP n'est pas correctement incorporée au complexe et que cela empêche la formation ultérieure des complexes BΔU1 et B* qui est le complexe catalytiquement actif.

## X. Les complexes spliceosomaux formés avec l'élément NRS activent la polyadénylation

En plus d'inhiber l'épissage, l'élément NRS fonctionnel active la polyadénylation du pré-ARNm. Pour que cette activation ait lieu dans le contexte viral, il est nécessaire que l'élément NRS soit fonctionnel et que le site accepteur 3' *env* soit présent (Miller et Stoltzfus, 1992 ; O'Sullivan *et al.*, 2002). La troncation du site 3'*src* n'a en revanche aucun effet sur le taux d'utilisation du site de polyadénylation (Miller et Stoltzfus, 1992). Comme nous l'avons vu dans l'introduction, le site de polyadénylation viral est sous-optimal et seulement 85 % des pré-ARNm sont polyadénylés au site de polyadénylation viral (Fogel *et al.*, 2002). Le modèle de régulation de la polyadénylation par l'élément NRS propose que l'engagement de l'élément NRS avec le site 3'*env* dans un complexe spliceosomal abortif favorise le recrutement de facteurs de polyadénylation (**Figure IV-19**). Cette activation est attribuée à la présence des séquences ESE dans la partie NRS 5' qui fixent les protéines SR (Maciolek et McNally, 2007). *In vitro*, la seule présence des séquences ESE de la partie NRS 5' située en amont de la séquence LTR du RSV suffit pour augmenter l'utilisation du signal de polyadénylation (Maciolek et McNally, 2007 ; Wilusz et Beemon, 2006). Dans le contexte viral, la troncation de ces séquences ESE comprises entre les nucléotides 712-744 ou 745-798 ou les mutations qui inactivent la séquence ESE comprises entre les nucléotides 740-770 provoquent de façon concomitante les diminutions de l'inhibition de l'épissage par l'élément NRS et de la polyadénylation (Fogel *et al.*, 2002). Par ailleurs, les séquences SELEX des protéines SR 9G8, ASF/SF2 et SRp20 peuvent fonctionnellement remplacer la partie NRS 5' lorsqu'elles sont fusionnées à la partie NRS 3' (Maciolek et McNally, 2007). Il est suggéré que les facteurs de polyadénylation soient recrutés par les protéines SR fixées sur l'élément

par des interactions avec leurs domaines RS ou par les autres facteurs présents sur le site 3'*env* comme le facteur U2AF65 (Maciolek et McNally, 2007) (**Figure IV-19**).

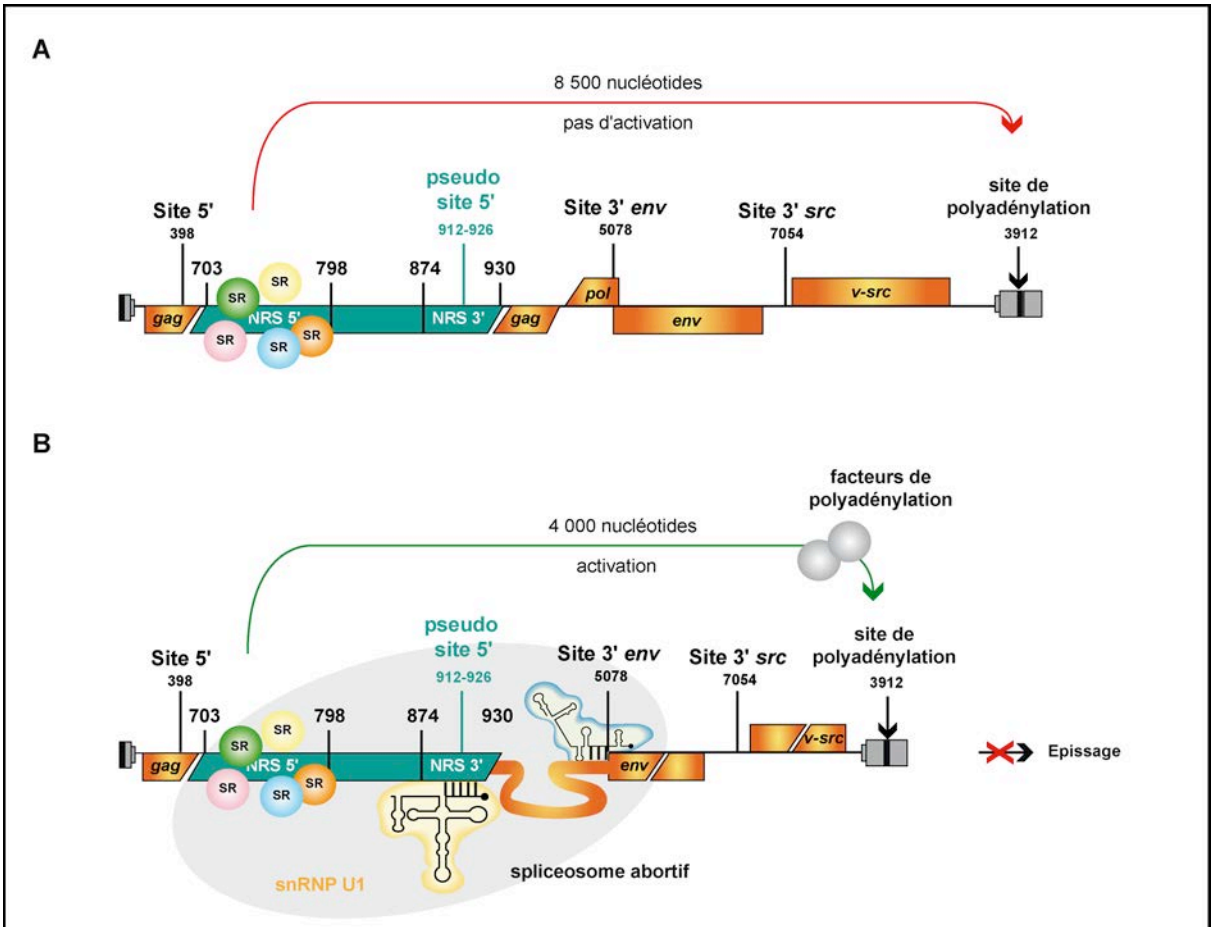



**Figure IV-19 : Modèle d'activation de la polyadénylation du pré-ARNm viral l'élément NRS (Maciolek et McNally, 2007, Fogel *et al.*, 2002, Wilusz et Beemon, 2006).**
**A :** L'élément NRS n'active pas la polyadénylation de l'ARN lorsqu'il n'est pas engagé dans un complexe spliceosomal abortif.
**B :** L'élément NRS engagé dans un complexe spliceosomal abortif avec le site 3'*env* inhibe l'épissage et stimule la polyadénylation de l'ARN. Les protéines SR fixées sur sa partie NRS 5' permettent le recrutement des facteurs de polyadénylation.

# XI. Le NRS favorise la mise en place des processus d'oncogenèse chez les virus ALV

## XI.1. Mécanismes d'oncogenèse provoqués par l'insertion des virus ALV

En plus de la régulation de l'épissage et de la polyadénylation, l'élément NRS a de fortes répercussions sur la mise en place du processus d'oncogenèse par les rétrovirus de la famille ASLV. L'infection des poulets par les virus ALV provoque des lymphomes qui ont des temps de latence courts (10 semaines) ou longs (4 à 10 mois). Ces temps de latence diffèrent suivant

le site d'insertion du provirus dans le génome, la souche du virus et l'âge de l'animal lors de l'infection. Les lymphomes à temps de latence long sont habituellement provoqués par l'insertion du provirus ALV dans le premier intron du gène c-*myc* (**Figure IV-20**).

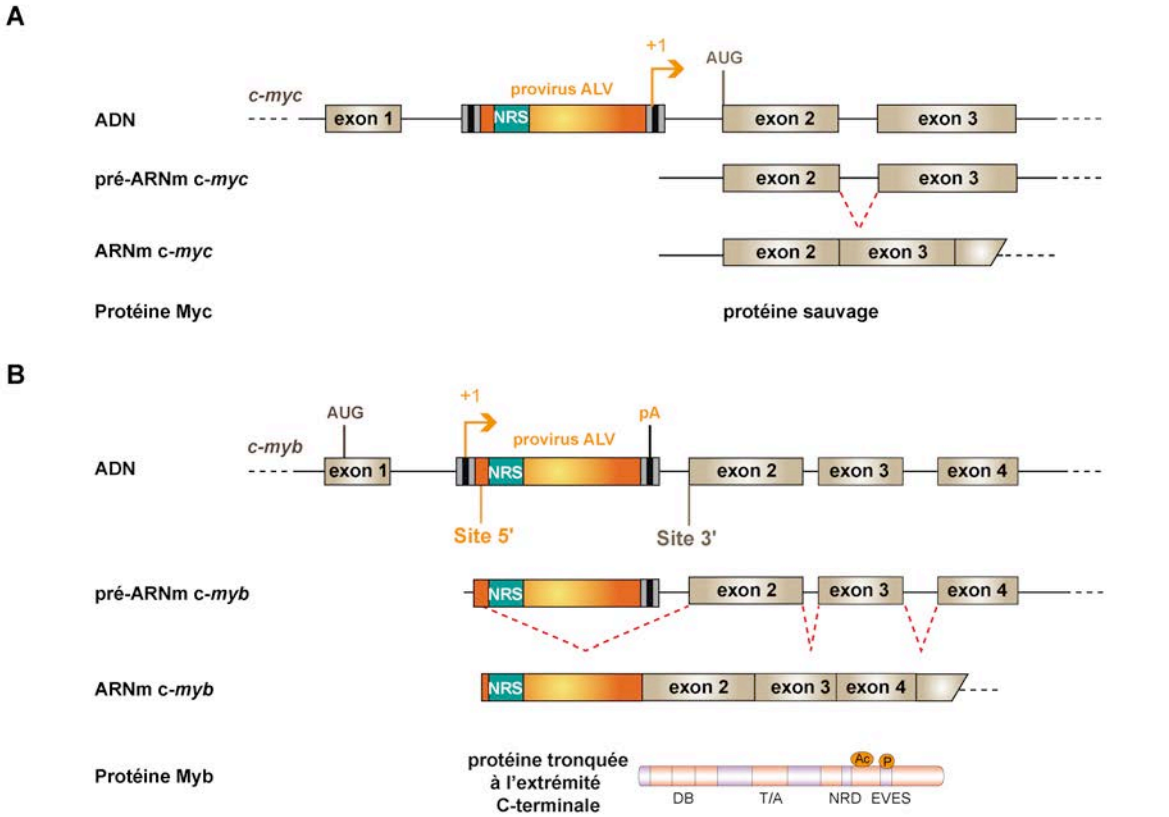



**Figure IV-20 : L'insertion des provirus ALV provoquent des lymphomes à temps de latence court et long.**

**A :** L'insertion du provirus ALV dans le premier intron du gène c-*myc* provoque une augmentation de l'expression de la protéine Myc qui induit des lymphomes à temps de latence long.

**B** : L'insertion du provirus ALV dans le premier intron du gène c-*myb* provoque la traduction d'une protéine Myb tronquée à l'extrémité N-terminale et provoque des lymphomes à temps de latence court.

Pour chaque exemple, le pré-ARNm, l'ARNm et la protéine obtenue sont illustrés. La séquence du provirus est représentée par un rectangle orange, ses séquences LTR par des rectangles gris et noirs. L'initiation de la transcription est représentée par une flèche orange, le site de polyadénylation viral (pA) et le site donneur sont mentionnés. Les exons des gènes cellulaires sont représentés par des rectangles bruns, l'épissage par des traits pointillés rouges. Les domaines fonctionnels de la protéine Myb sont indiqués : DB, domaine de liaison à l'ADN, T/A : domaines de transactivation, NRD : domaine de régulation négative, EVES : domaine d'interaction inter et intramoléculaires.

La forte transcription par le promoteur proviral de la région 3' LTR se situant juste en amont du codon initiateur de la traduction provoque l'augmentation de l'expression du gène c-*myc* (Hayward *et al.*, 1981). Cette expression, qui est 50 fois plus importante que celle des cellules non infectées, est responsable de la pathologie. L'épissage du pré-ARNm c-*myc*, la traduction, ainsi que la protéine Myc obtenue sont identiques à ceux des cellules non infectées.

Les lymphomes à temps de latence court sont provoqués par l'insertion du provirus dans le premier intron du gène c-*myb* en aval du premier exon qui contient le codon initiateur de la traduction (**Figure IV-20**) (Jiang *et al.*, 1997). La transcription des gènes viraux est initiée au niveau de la région 5' LTR et passe outre le signal de polyadénylation viral pour se poursuivre dans la partie du gène c-*myb* se situant en aval du site d'insertion du provirus (Jiang *et al.*, 1997) (**Figure IV-20**). Le 1$^{er}$ évènement d'épissage du pré-ARNm chimérique alors obtenu a lieu entre l'unique site donneur viral et le site accepteur du second exon du gène c-*myb*. Les 13 autres exons du gène c-*myb* qui sont en aval de l'exon 2 sont épissés normalement comme dans le pré-ARNm c-*myb* sauvage. La protéine traduite à partir de cet ARNm est tronquée de 20 premiers acides aminés situés à l'extrémité N-terminale. Cette troncation a lieu dans le domaine de fixation à l'ADN et confère à la protéine des propriétés très oncogéniques (Jiang *et al.*, 1997) (**Figure IV-20**).

### XI.2. La troncation de l'élément NRS chez les virus ALV augmente leur caractère oncogénique

Les lymphomes à temps de latence court ont lieu en général dans 10 à 30 % des infections des animaux (Kanter *et al.*, 1988). Un virus ALV recombinant, EU8, a la particularité de provoquer ces lymphomes à une fréquence de 60 à 70% (Kanter *et al.*, 1988 ; Simon *et al.*, 1987 ; Smith *et al.*, 1997). Ce virus comporte les séquences LTR et le gène *gag-pol* du virus RPV (Ring Necked Pheasant) et le gène *env* du virus UR2AV (Simon *et al.*, 1987) (**Figure IV-21**).

Un autre virus recombinant appelé ALV LR9 qui est identique au virus EU8 à l'exception de la séquence du gène *gag-pol* qui provient du virus UR2AV, provoque en revanche majoritairement des lymphomes à temps de latence long et minoritairement des lymphomes à temps de latence court (Simon *et al.*, 1987 ; Smith *et al.*, 1997) (**Figure IV-22**). La pathogenicité du virus EU8 est causée par une troncation d'une séquence de 42 nucléotides (nt 735-776) située dans gène *gag* et aussi dans l'élément NRS (Smith *et al.*, 1997) (**Figure IV-21**).

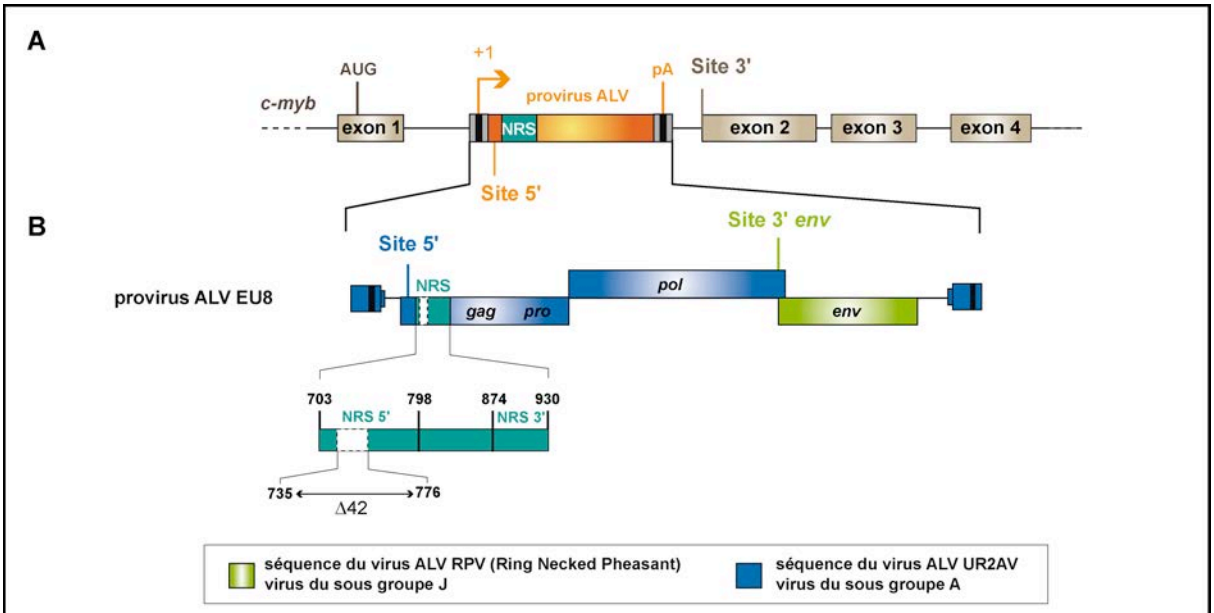



**Figure IV-21 : La séquence du virus EU8 qui induit des lymphomes à courte durée de latence est tronquée de 42 nucléotides dans l'élément NRS (Polony *et al.*, 2003, Smith *et al.*, 1997).**
**A :** Représentation générale de l'insertion d'un provirus ALV dans le premier intron du gène c-*myb* provoquant des lymphomes à temps de latence court. La séquence du provirus est représentée par un rectangle orange, les exons du gène c-*myb* par des rectangles bruns. L'initiation de la transcription est représentée par une flèche orange, le site de polyadénylation viral (pA) et le site donneur sont mentionnés.
**B :** Représentation du provirus ALV EU8. Ce virus possède les séquences LTR et les gènes *gag-pol* du virus RPV et le gène *env* du virus ALV UR2AV. Ces séquences sont colorées respectivement en vert et en bleu. Ce virus induit des lymphomes à courte durée de latence à une très haute fréquence en raison de la troncation de 42 nucléotides de l'élément NRS dans le gène *gag*. Les positions des nucléotides de l'élément NRS sont mentionnées.

La reproduction de cette troncation dans le génome du virus LR9, générant le virus ΔLR9, suffit à induire des lymphomes à temps de latence court à une fréquence aussi importante que celle observée pour le virus EU8 (Smith *et al.*, 1997) (**Figure IV-22**). L'augmentation de la pathogenicité du virus ΔLR9 est due à la perte de l'activité de l'élément NRS qui n'est plus fonctionnel lorsqu'il est tronqué de la séquence comprise entre les nucléotides 735-776 (Polony *et al.*, 2003 ; Smith *et al.*, 1997). Dans le virus LR9, l'inactivation de l'élément NRS par une mutation qui diminue la fixation de la snRNP U1 au pseudo site tout en préservant la capacité codante de la séquence (mutant G919A) provoque le même phénotype que celui provoqué par la troncation des 42 nucléotides chez le virus ΔLR9 (Neiman *et al.*, 2003 ; Polony *et al.*, 2003). Placé en contexte hétérologue dans l'intron d'un mini-gène, l'élément NRS du virus EU8 ainsi que l'élément NRS ΔLR9 n'inhibent pas l'épissage, contrairement aux éléments NRS du virus RSV et ALV qui sont fonctionnels (Smith *et al.*, 1997).

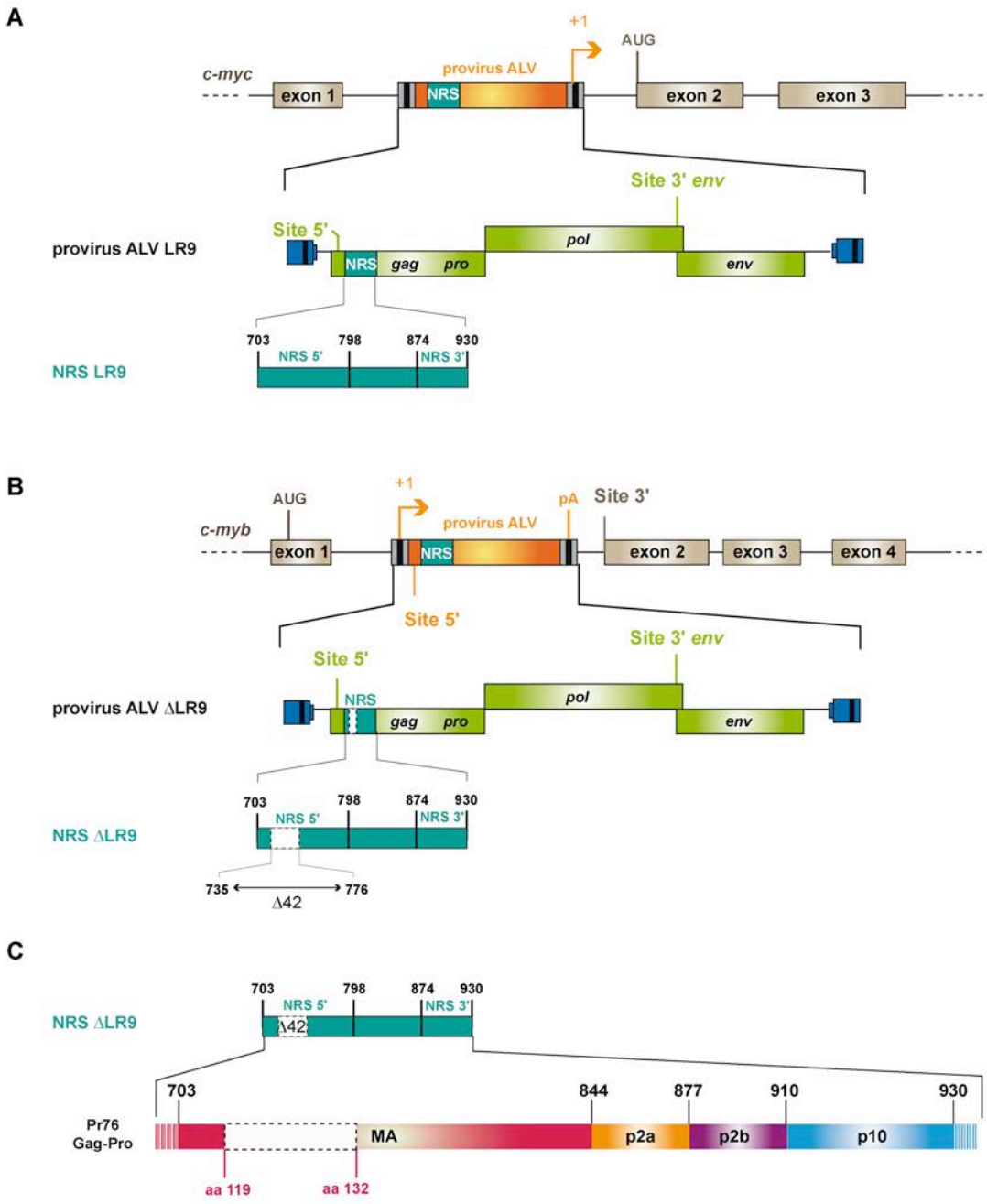



**Figure IV-22 : La troncation de 42 nucléotides de la séquence de l'élément NRS provoque une augmentation de la pathogenicité du virus ALV LR9 (Polony *et al.*, 2003, Smith *et al.*, 1997).**

**A :** Le virus ALV LR9 induit majoritairement des lymphomes à temps de latence long qui sont provoqués par son insertion dans le premier intron du gène c-*myc*. Les séquences du virus ALV LR9 qui proviennent du virus RPV (gènes *gag-pol*, *env*) sont en vert et celles qui proviennent du virus UR2AV (régions LTR) sont en bleu. Le virus LR9 contient dans son gène *gag* l'élément NRS entier (NRS LR9).

**B :** Le virus ALV LR9 dont le gène *gag* est tronqué des 42 nucléotides (virus ALV ΔLR9) qui contient une partie de l'élément NRS induit majoritairement des lymphomes à courte durée de latence à une très haute fréquence en s'insérant dans le premier intron du gène c-*myb*. Les positions de la troncation dans l'élément NRS ΔLR9 sont mentionnées.

**C :** Représentation de l'élément NRS ΔLR9 et de la partie du précurseur Pr76 Gag-Pro qu'il code. La troncation de 42 nucléotides du gène *gag* provoque la perte de la séquence codant 13 acides aminés de la protéine de la matrice mais n'a pas d'incidence sur la réplication virale.

Par ailleurs, cette troncation est en phase avec le cadre de lecture de l'ORF de *gag* et provoque la perte de la séquence codant les acides aminés 119 à 132 de la protéine de la matrice (**Figure IV-22**) mais cela n'a pas de répercussion sur le bourgeonnement, le caractère infectieux et la réplication virale (Nelle et Wills, 1996 ; Polony *et al.*, 2003 ; Smith *et al.*, 1997).

La perte de la fonction de l'élément NRS a donc une grande influence sur le caractère oncogénique des virus ALV. Cela implique que l'élément NRS, comme chez le virus RSV, régule également la polyadénylation et l'épissage des virus ALV, deux processus qui doivent être dérégulés lors de la mise en place de l'oncogenèse (Polony *et al.*, 2003 ; Smith *et al.*, 1997). Les séquences des gènes *gag* du virus EU8 et LR9 qui proviennent respectivement des virus RPV et UR2AV sont homologues et la séquence de leur élément NRS ne diffèrent de celle de l'élément NRS WT du virus RSV que par 4 et 10 substitutions de bases, respectivement (**Figure IV-23**) (Smith *et al.*, 1997).

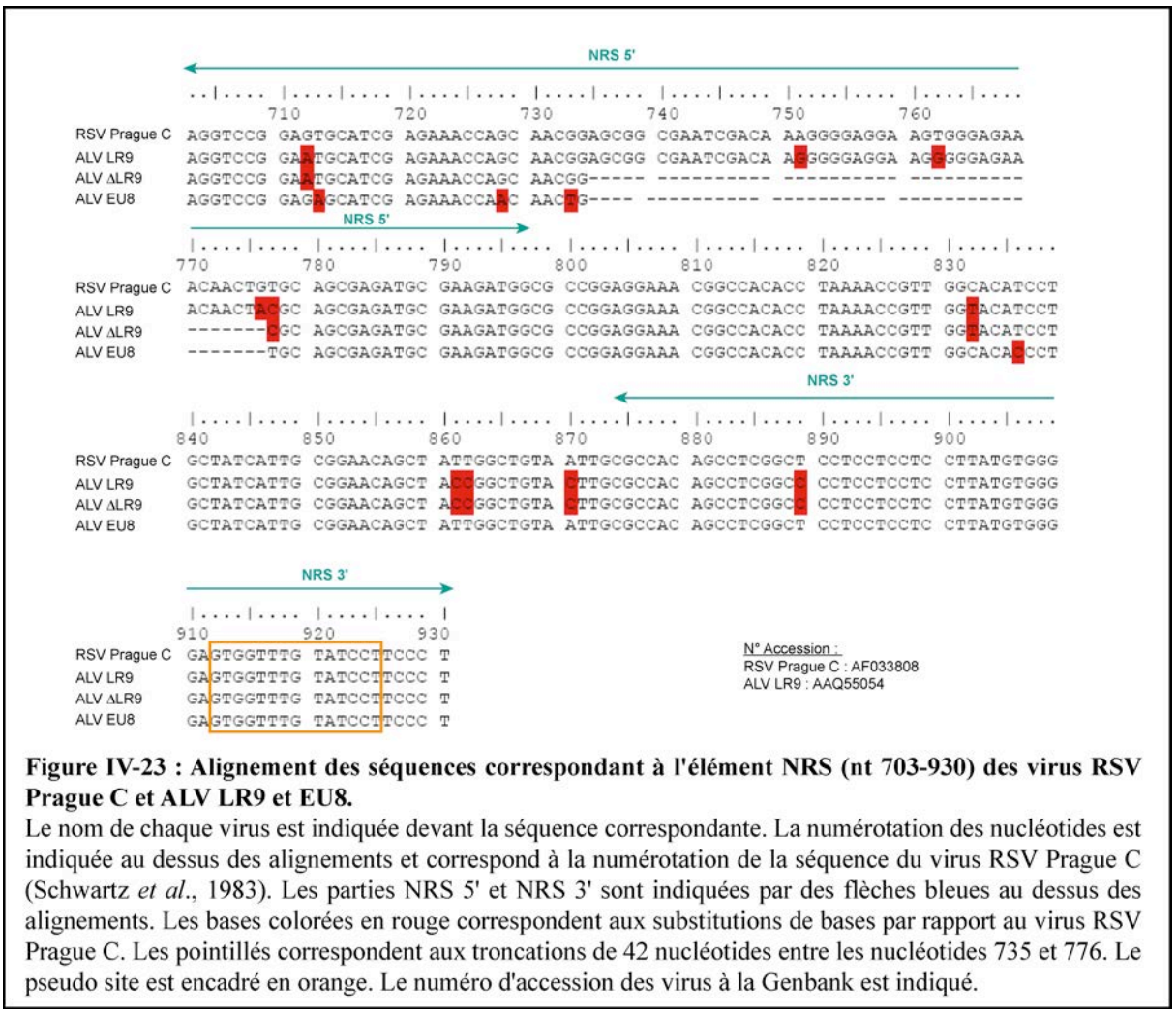



**Figure IV-23 : Alignement des séquences correspondant à l'élément NRS (nt 703-930) des virus RSV Prague C et ALV LR9 et EU8.**
Le nom de chaque virus est indiquée devant la séquence correspondante. La numérotation des nucléotides est indiquée au dessus des alignements et correspond à la numérotation de la séquence du virus RSV Prague C (Schwartz *et al.*, 1983). Les parties NRS 5' et NRS 3' sont indiquées par des flèches bleues au dessus des alignements. Les bases colorées en rouge correspondent aux substitutions de bases par rapport au virus RSV Prague C. Les pointillés correspondent aux troncations de 42 nucléotides entre les nucléotides 735 et 776. Le pseudo site est encadré en orange. Le numéro d'accession des virus à la Genbank est indiqué.

### XI.3. L'élément NRS régule l'épissage et la polyadénylation chez les virus ALV LR9 et ΔLR9

La troncation dans l'élément NRS des 42 nucléotides situés entre les positions 735 et 776 abolissent ses fonctions d'inhibiteur de l'épissage et d'activateur de la polyadénylation (Polony *et al.*, 2003 ; Smith *et al.*, 1997). Comme nous l'avons abordé précédemment, ces deux fonctions sont liées car l'élément NRS doit être fonctionnel pour former des complexes spliceosomaux abortifs avec le site 3'*env* qui à leur tour vont stimuler l'utilisation du site de polyadénylation viral (Giles et Beemon, 2005 ; Maciolek et McNally, 2007).

Dans le contexte du virus ΔLR9, l'absence d'inhibition de l'épissage par l'élément NRS ΔLR9 a deux répercussions majeures. Premièrement, cela induit une diminution de 60 % de l'utilisation du site de polyadénylation viral ce qui provoque la transcription de la partie du gène c-*myb* se situant en aval (Polony *et al.*, 2003). Deuxièmement, cela provoque au niveau du pré-ARNm chimérique alors obtenu, l'épissage entre le site donneur unique viral et le site accepteur du second exon du gène c-*myb* qui est davantage utilisé que le site viral 3'*env* faible (Polony *et al.*, 2003). Ces deux dérégulations sont à l'origine de la mise en place du processus d'oncogenèse chez le virus ALV ΔLR9. Cependant, l'élément NRS ΔLR9 n'a pas d'incidence sur la multiplication du virus qui se réplique à la même vitesse que le virus possédant l'élément NRS LR9 WT (Polony *et al.*, 2003).

Par ailleurs, les fonctions d'inhibition de l'épissage et d'activation de la polyadénylation par l'élément NRS dépendent en grande partie de la présence de la région NRS 5' qui a une activité ESE et qui est partiellement tronquée dans l'élément NRS ΔLR9 (**Figure IV-24**) (Maciolek et McNally, 2007 ; McNally et McNally, 1998). Par rapport à l'élément NRS LR9 entier ou au NRS WT du virus RSV, cette troncation provoque une diminution de la fixation des protéines ASF/SF2 et hnRNP H sur l'élément NRS ΔLR9 dans les complexes ribonucléoprotéiques assemblés *in vitro* avec un extrait nucléaire de cellules HeLa (Smith *et al.*, 1997). Néanmoins, dans ces conditions, la troncation n'a pas d'effet sur le taux de fixation de la snRNP U1 sur le pseudo site (Smith *et al.*, 1997). Cela souligne le fait que la fixation de la snRNP U1, bien qu'essentielle aux régulations exercées par l'élément NRS, n'est pas suffisante pour leur établissement.

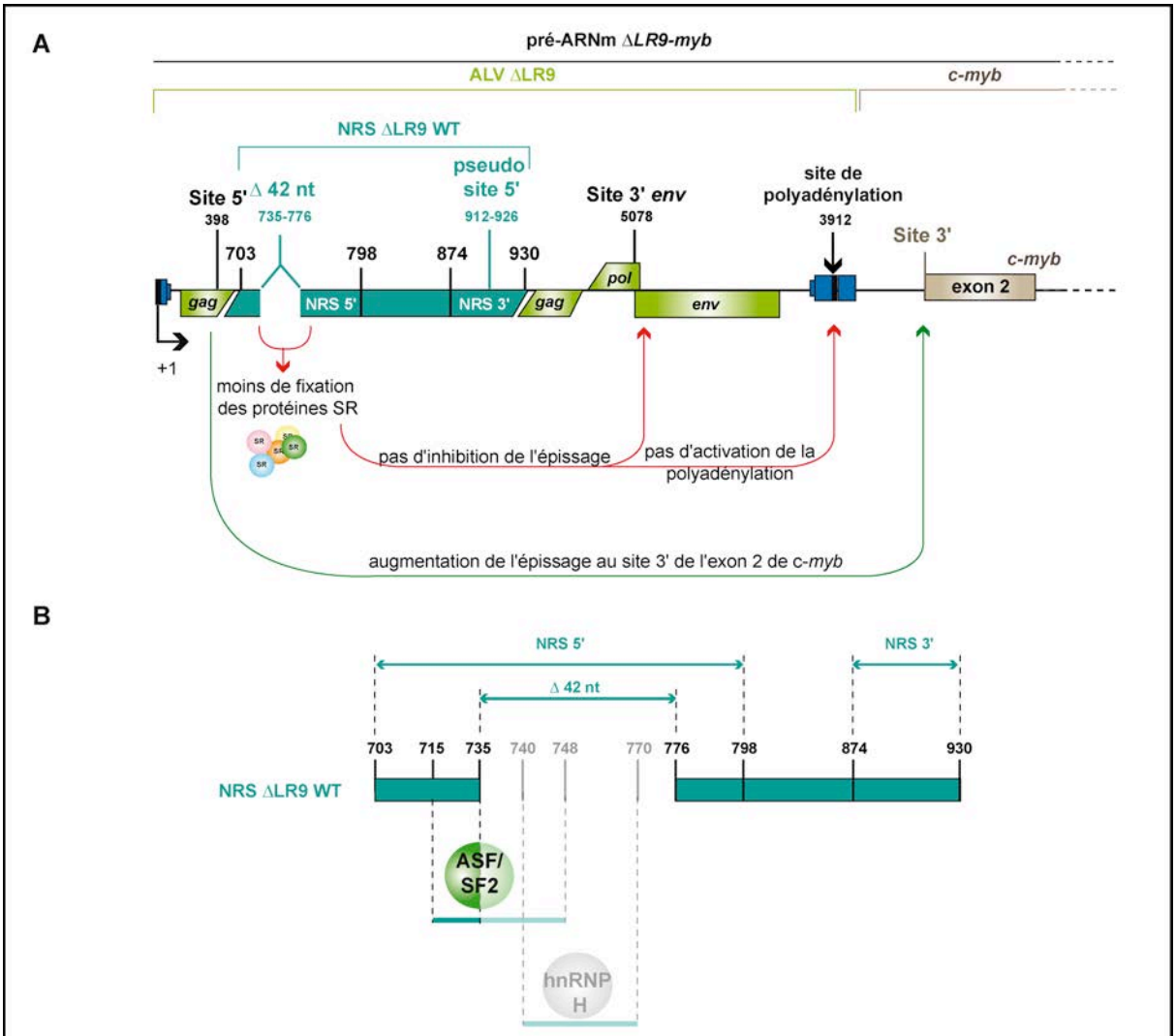



**Figure IV-24 : Répercussions de l'élément NRS ΔLR9 non fonctionnel sur l'épissage et la polyadénylation du pré-ARNm viral (Polony *et al.*, 1997, Smith *et al.*, 1997).**
**A :** La troncation de 42 nucléotides de l'élément NRS (NRS ΔLR9) inactive ses fonctions d'inhibiteur de l'épissage et d'activateur de la polyadénylation. Cette perte de fonction pourrait être due à la diminution de la fixation des protéines SR sur la partie NRS 5'. Cela provoque une augmentation de la transcription de gènes chimériques ALV DLR9-myb et de l'épissage au site 3' de l'exon 2 de c-*myb*.
**B :** Représentation de l'élément NRS ΔLR9. Les sites de fixation de la protéine SR ASF/SF2 et de la protéine hnRNP H sont reportées. Les parties de ces sites de fixation qui sont tronquées sont représentés en transparence.

L'ensemble des résultats obtenus pour les virus ALV LR9 et ΔLR9 rejoint ceux obtenus pour l'élément NRS du virus RSV de la souche Prague C. Que ce soit chez le virus RSV ou le virus ALV LR9, l'inactivation de l'élément NRS se traduit par une augmentation du caractère oncogénique. Dans le virus RSV, certaines des mutations qui altèrent l'activité de l'élément NRS provoquent une augmentation de l'épissage au site 3'*src* ainsi qu'une diminution de l'utilisation du site de polyadénylation viral ce qui se traduit par une transformation plus rapide des cellules infectées (O'Sullivan *et al.*, 2002) (**Figure IV-24**). Par ailleurs, la mutation de la séquence du NRS 5' comprise entre les nucléotides 740-770 qui diminue son

activité ESE provoque une diminution de l'utilisation du site de polyadénylation viral (Fogel *et al.*, 2002).

L'élément NRS est donc un élément de régulation important, tant pour le virus RSV que pour le virus ALV LR9. Dans les deux cas, la régulation de l'épissage par cet élément semble davantage inhiber l'épissage au site accepteur distal, qui est soit le site 3'*src* pour le virus RSV, soit le site 3' du gène en aval du site d'insertion comme le gène c-*myb* du virus ALV LR9. Dans les deux cas, la formation des complexes spliceosomaux abortifs impliquant le pseudo site de l'élément NRS est nécessaire pour promouvoir la polyadénylation des pré-ARNm viraux ou chimériques.

## XII. Importance de la structure secondaire de l'élément NRS

Lorsque la taille de l'élément NRS fonctionnel a été limitée aux nucléotides 703-930, un modèle de sa structure secondaire théorique prédite à l'aide du logiciel Mfold a été proposé (McNally *et al.*, 1991) (**Figure IV-25**). Dans cette structure, les séquences riches en purines et en pyrimidines des parties NRS 5' et NRS 3' étaient extensivement appariées et formaient une longue hélice irrégulière. Le site de fixation de la protéine ASF/SF2 était majoritairement dans une structure tige boucle tandis que la séquence du pseudo site était dans une hélice comprenant deux bulges (**Figure IV-25**).

La structure tridimensionnelle du pseudo site isolé a ensuite été établie par RMN (Cabello-Villegas *et al.*, 2004). La séquence étudiée, NRS 23, a été délimitée aux nucléotides 907-929 et comporte la séquence du pseudo site qui est compris entre les nucléotides 912-926 (Cabello-Villegas *et al.*, 2004). Dans ce contexte, la séquence du pseudo site adopte une structure tige boucle qui est différente de celle précédemment proposée qui avait été prédite par informatique avec la totalité de la séquence de l'élément NRS (Cabello-Villegas *et al.*, 2004 ; McNally *et al.*, 1991). Cette structure comprend une hélice de 6 pb, un nucléotide en bulge, une hélice de 3 pb et une tétra-boucle (**Figure IV-26**) (Cabello-Villegas *et al.*, 2004).

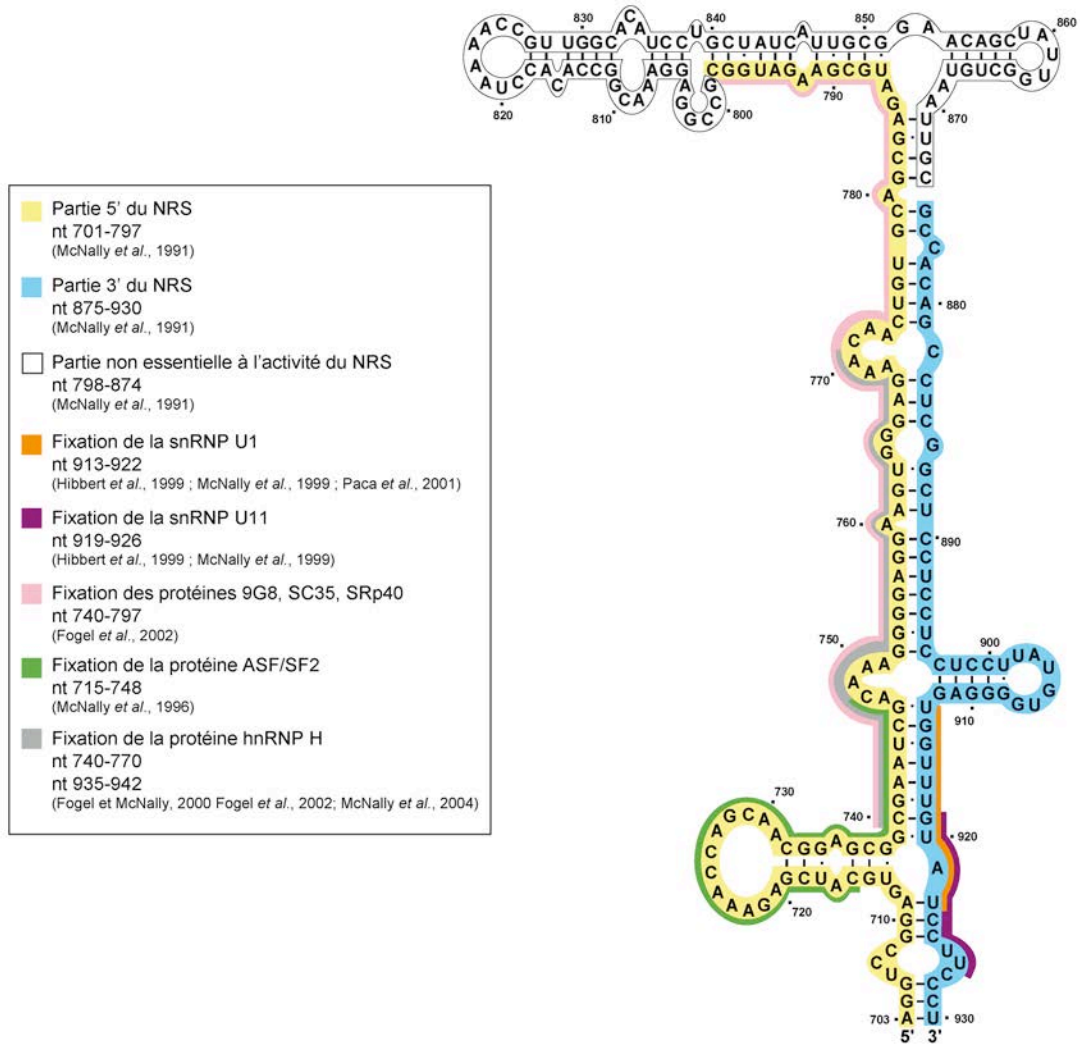



**Figure IV-25 : Structure secondaire théorique de l'élément NRS WT du virus RSV prédite par le logiciel Mfold (McNally *et al.*, 1991).**
Les différentes parties fonctionnelles ainsi que les sites de fixation des différents partenaires ont été reportés sur le modèle de structure théorique de l'élément NRS WT. Ces séquences sont encadrées selon le code couleur présenté dans l'encart.

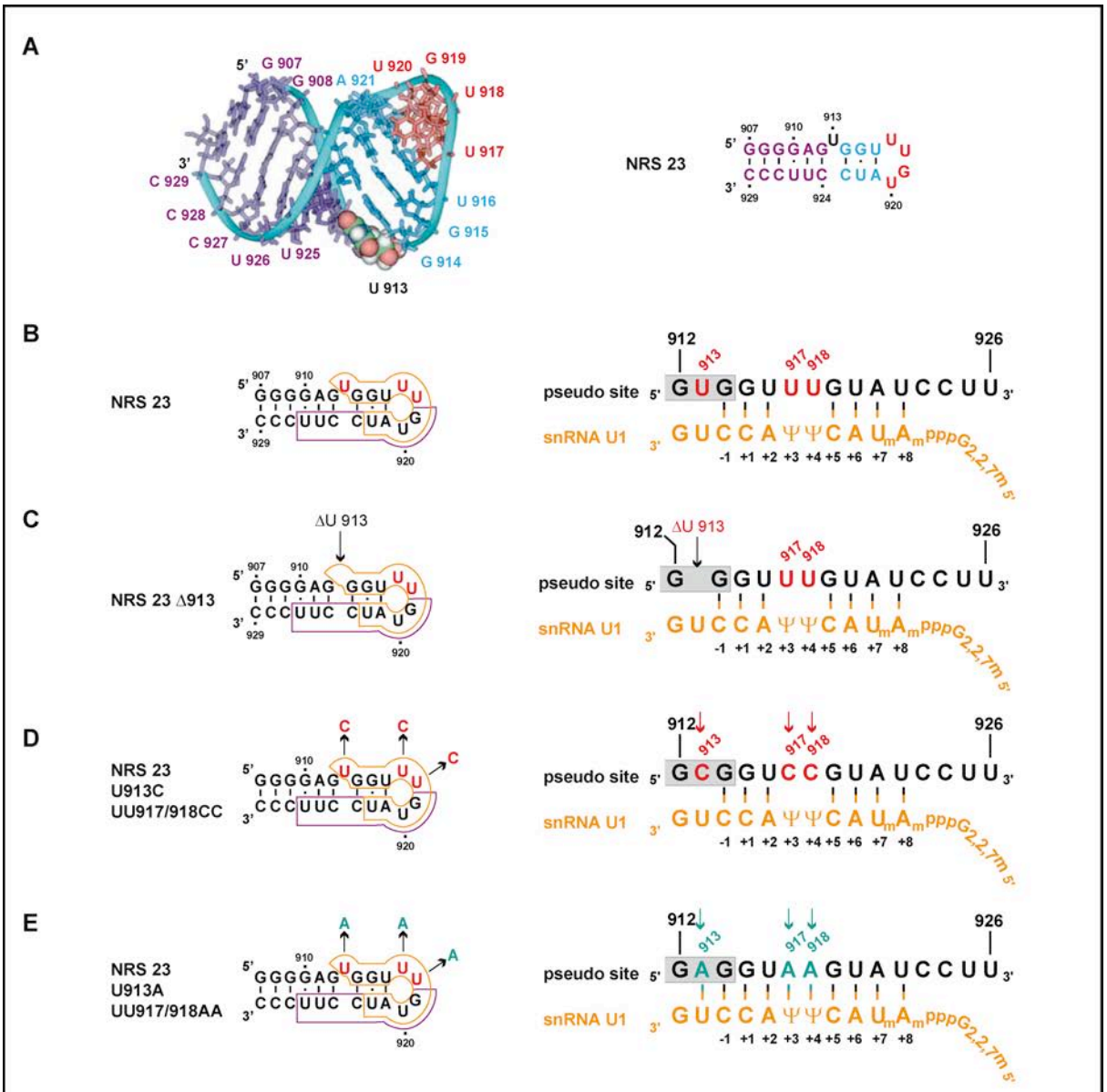



**Figure IV-26 : Structure de la séquence du pseudo site de l'élément NRS WT du virus RSV (NRS 23) résolue par RMN et relation avec sa fonction (Cabello-Villegas *et al.*, 2004, Paca *et al.*, 2001).**

**A :** Représentation de la structure résolue par RMN de la séquence du pseudo site comprise entre les nucléotides 907 et 929 (NRS 23). La structure secondaire correspondant est représentée en dessous de la struture résolue par RMN. Les nucléotides de la tige en amont et en aval du bulge sont respectivement colorés en violet et bleu, le nucléotide en bulge en noir, et les nucléotides de la boucle en rouge. Les positions des nucléotides sont indiquées.

**B :** Report sur la structure secondaire du NRS 23 des séquences d'appariement des snRNA U1 et U11. Ces séquences sont respectivement encadrées en orange et en violet. Les appariements des nucléotides du pseudo site avec le snRNA U1 correspondant sont représentés à droite de la structure. Les nucléotides du pseudo site différents de la séquence consensus du site donneur sont colorés en rouge.

**C :** Représentation du mutant NRS 23 Δ913. La troncation du nucléotide U913 entraîne une diminution de la fixation du snRNA U1 sur la séquence du pseudo site.

**D :** Représentation des mutants U913C et UU917/918CC. Ces mutants inactivent l'inhibition de l'épissage par le NRS et le mutant UU917/918CC diminue la fixation du snRNA U1 sur la séquence du pseudo site.

**E :** Représentation des mutants U913A et UU917/918AA. Ces mutants rendent le pseudo site fonctionnel en épissage. Le mutant UU917/918CC augmente la fixation du snRNA U1 sur la séquence du pseudo site.

Cette structure est importante pour la fonction de l'élément NRS car la troncation du nucléotide U913 qui est en bulge dans la structure, provoque une diminution de 33 % de la fixation de la snRNP U1 sur le pseudo site par rapport à la tige boucle NRS 23 (**Figure IV-26**) (Cabello-Villegas *et al.*, 2004). Le nucléotide U913 n'est pas un nucléotide canonique et sa troncation n'entraîne pas une diminution du nombre d'appariement de la séquence du pseudo site avec le snRNA U1. En revanche, sa troncation entraîne une hyperstabilisation de la structure du pseudo site qui forme hélice parfaite de 9 pb (Cabello-Villegas *et al.*, 2004). Par conséquent, la présence de ce nucléotide en bulge dans la structure du pseudo site est nécessaire pour déstabiliser la structure de la tige afin de faciliter les désappariements de ses nucléotides pour pouvoir former des appariements avec le snRNA U1 (Cabello-Villegas *et al.*, 2004).

En plus de sa structure, la séquence du pseudo site est importante pour sa fonction. Le pseudo site comporte 3 uridines non canoniques aux positions 913, 917 et 918 qui correspondent respectivement aux positions -2, +3 et +4 dans la séquence d'appariement avec le snRNA U1 (**Figure IV-26**) (Hibbert *et al.*, 1999 ; Paca *et al.*, 2001). Dans la structure du pseudo site NRS 23, ces nucléotides sont non appariés, en bulge ou dans la boucle. Leurs mutations en cytosine, qui est aussi non canonique à ces positions, ne devraient avoir aucune incidence sur le nombre d'appariement du pseudo site avec le snRNA U1 et la structure du pseudo site. Pourtant, ces mutations entraînent l'inactivation de la fonction inhibitrice de l'épissage par le NRS (Paca *et al.*, 2001). La mutation UU 917/918 CC provoque une diminution de la fixation du snRNA U1 du même ordre que celui provoqué par la troncation U913 (Cabello-Villegas *et al.*, 2004).

D'un autre côté, les mutations de ces nucléotides en adénosines qui sont des nucléotides canoniques qui ne devraient provoquer aucun remaniement de structure rendent le pseudo site fonctionnel en épissage (**Figure IV-26**) (Paca *et al.*, 2001). La mutation UU 917-918 AA provoque une augmentation de la fixation du snRNA U1 sur le pseudo site supérieure à celle observée pour l'élément NRS 23 (Cabello-Villegas *et al.*, 2004).

Par conséquent, tant au niveau de la séquence que de la structure, les nucléotides non consensus du pseudo site de l'élément NRS jouent un rôle fonctionnel important dans la fixation du snRNA U1.

# Objectifs

Comme nous l'avons vu dans l'introduction, la régulation de l'épissage alternatif du transcrit primaire du virus RSV est essentielle au bon déroulement du cycle viral. L'élément NRS est un des acteurs majeur de cette régulation et contribue à inhiber l'épissage du transcrit primaire de façon à maintenir un pourcentage d'ARN non épissé correspondant à environ 70 à 80% de l'ARN viral total. Cependant, le mécanisme par lequel cet élément inhibe l'épissage n'est pas encore clairement compris à l'échelle moléculaire. Bien que l'élément NRS ait été clairement identifié entre les positions 703-930, aucune étude structurale de cette séquence ARN n'avait été jusqu'ici entreprise. Seule une prédiction informatique de la structure 2D de l'élément NRS et l'étude par RMN d'un petit fragment (positions 910-926) avaient été réalisée (Cabello-Villegas *et al.*, 2004 ; McNally *et al.*, 1991). Ces résultats théoriques et expérimentaux étaient contradictoires et la structure du fragment 910-926 établie expérimentalement n'était pas retrouvée dans la structure prédictive pour l'élément entier. Etant donné que la structure 2D des ARN joue un rôle important dans la régulation de l'épissage et que le laboratoire est spécialisé dans l'étude expérimentale de la structure des ARN en solution, nous avons décidé d'entreprendre l'étude de la région de l'ARN viral, ce qui devait apporter des informations permettant de mieux comprendre son fonctionnement. Cette étude avait été débutée au laboratoire par A. Mougin et je l'ai poursuivie au cours de ce travail de thèse.

Mon premier objectif, en collaboration avec V. Marchand, a été de réaliser l'analyse expérimentale de la structure de l'élément NRS entier du virus RSV, ainsi que d'une forme tronquée définie comme étant l'élément minimal conservant la capacité d'inhiber l'épissage *in cellulo* (élément NRS Δ76). J'ai complété ces travaux par l'analyse structurale d'un autre élément NRS, celui du virus ALV LR-9, ainsi que de sa forme tronquée appelée ΔLR9. Cette forme tronquée favorise la mise en place du processus de tumorigenèse chez le poulet (Polony *et al.*, 2003 ; Smith *et al.*, 1997). L'analyse de ces régions d'ARN a impliqué la mise en œuvre d'expériences de cartographie de la structure 2D en solution et des études phylogénétiques.

Après avoir caractérisé la structure de l'élément NRS, mon second objectif était d'essayer de définir le rôle possible des éléments structuraux ainsi définis, dans l'inhibition de l'épissage. Pour cela, une série de mutants portant des troncations internes respectant la structure secondaire ont été réalisés et les éléments NRS mutés ont été insérés dans l'intron d'un pré-ARNm afin de mesurer leur efficacité d'inhibition de l'épissage. Ceci a été réalisé en collaboration avec A. Visvikis, à la fois *in vitro* dans un extrait nucléaire de cellules HeLa et *in cellulo* avec des constructions permettant de transfecter des cellules HeLa.

Mon troisième objectif a été d'identifier les partenaires protéiques de l'élément NRS et pour certains d'entre eux, de rechercher leur fonction. Bien que l'on sache comme on l'a vu dans l'introduction, que l'élément NRS fixe les particules snRNP U1 et U11, ainsi que de nombreuses protéines SR (ASF/SF2, 9G8, SRp40, SRp20, SC35), et la protéine hnRNP H, les sites de fixation de ces protéines SR n'avaient jamais été clairement définis. De ce fait, le mécanisme d'inhibition de l'épissage par l'élément NRS est encore mal compris. En plus, on pouvait supposer que les connaissances sur les facteurs protéiques interagissant avec l'élément NRS étaient loin d'être exhaustives. Afin d'étudier ses partenaires potentiels, nous avons décidé de former des complexes ribonucléoprotéiques en incubant l'ARN NRS (nt 703-930) dans un extrait nucléaire de cellules HeLa, puis de les purifier par chromatographie d'affinité et d'en analyser le contenu en protéines par spectrométrie de masse.

Ceci impliquait l'application d'une méthode de chromatographie d'affinité et en collaboration avec A. Visvikis, j'en ai testé plusieurs. La plus performante s'est avérée être la technique basée sur la fusion à l'ARN étudié à 3 segments du phage MS2 fixant la protéine de l'enveloppe et l'utilisation de la protéine MS2.

Les résultats de ces chromatographies d'affinité ont permis de mettre en évidence le rôle des protéines hnRNP A1 et 9G8 dans l'inhibition de l'épissage par l'élément NRS. Jusqu'à présent, aucun rôle fonctionnel de ces protéines n'avait été décrit dans l'inhibition de l'épissage par l'élément NRS. Par conséquent, nous avons défini les sites de fixation de ces protéines par des expériences d'empreinte. Les résultats montrent que les sites de fixation de chacune des protéines se chevauchent ce qui suppose qu'elles entrent en compétition pour la fixation sur l'élément NRS. Cette hypothèse a été vérifiée par G. Khoury par des expériences d'immunosélection indiquant que la protéine hnRNP A1 chasse la protéine 9G8.

Par ailleurs, avec A. Visvikis, nous avons mis en évidence le rôle de la protéine 9G8 dans l'activité du NRS par des expériences d'épissage *in vitro* réalisées avec des extraits nucléaires de cellules HeLa auquel a été ajouté de la protéine 9G8 recombinante. Ces résultats ont été confirmés avec A. Visvikis et G. Khoury par des expériences d'épissage *in cellulo*, au cours desquelles la protéine 9G8 était soit surexprimée, soit au contraire son expression était diminuée par l'action d'un siRNA dirigé contre son ARNm. L'ensemble de ces données montre que la protéine 9G8 est un activateur de l'activité inhibitrice de l'élément NRS.

# Matériel et Méthodes

# Partie Matériel

## I. Souches d'*E. coli* utilisées

**Souche DH5α :** *deo*R, *sup*E44, Δ(*laci*ZYA-argF)U169, Φ80*lac*ZΔM15, *hsd*R17 ($r_K^-$ $m_K^+$), *rec*A1, *end*A1, *gyr*A96, *thi*-1, *rel*A1, $F^-$ (Hanahan, 1983) a été utilisée pour la production et la purification des plasmides.

**Souche BL21-CodonPlus-RIL** : B F-, *ompT, hsdS* ($r_B$– $m_B$–), *dcm*+, $Tet^r$, *gal, endA,* Hte *[argU ileY leuW* $Cam^r$] (Studier et Moffatt, 1986) a été utilisée pour la production des protéines recombinantes.

**Souche XL1-blue** : *rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17, *sup*E44, *rel*A1, *lac*F', *pro*AB, $lacI^qZ\Delta M15$, Tn10 ($tet^r$) (Bullock *et al.*, 1987) a été utilisée pour les mutagenèses dirigées.

## II. Lignées de cellules eucaryotes

**Cellules HeLa :** N°ATCC® : CCL-2TM (Scherer, 1953) lignée cellulaire provenant d'un prélèvement effectué sur une patiente atteinte d'un cancer du col de l'utérus.

## III. Les oligonucléotides

L'ensemble des oligonucléotides utilisés pour les clonages, les amplifications par PCR, les extensions d'amorces et les Northern blot est détaillé dans le **Tableau V-1**.

## IV. Les siRNA

Les siRNA utilisés dans le cadre des expériences d'épissage *in cellulo* nous ont été fourni par J. Stévenin et sont répertoriés ci-dessous dans le **Tableau V-2**.

| Nom | Séquence (5'-3') | cible | publication |
|---|---|---|---|
| **si9G8** | AGGAGAGUUAGAAAGGGCU (dT) | ARNm de la protéine 9G8 | (Dreumont *et al.*, 2010) |
| **siLuc** | GUAGCGCGGUGUAUUAUAC | ARNm de la protéine Luciférase Renilla | (Blaustein *et al.*, 2005) |

**Tableau V-2** : siRNA utilisé au cours de ce travail.

| Numéro | Séquence (5'-3') | site de restriction/ promoteur | séquence virale ciblée | nucléotides de la séquence ciblée | utilisation | Construction ou ARN produit/amplifié |
|---|---|---|---|---|---|---|
| **2954** | CG**GAATTC**AAATTTAATACGACTCACTATAGGGT CTCTCCCCCAGGTCCGGA | *Eco*RI | RSV, ALV | 691-711 | PCR | NRS, Δ76, LR9, ΔLR9 |
| **2956** | CATCTCGCTGCACAGTTG | | RSV, ALV | 771-788 | extension d'amorce | Δ76 |
| **2955** | AATTTTGCTGCTCTCCCA | | RSV, ALV | 939-951 | extension d'amorce, PCR | NRS, Δ76, LR9 |
| **1618** | TGCTGCTCTCCCACCCCCG | | RSV | 933-951 | extension d'amorce, PCR | NRS |
| **1619** | AGGGAAGGATACAAACCACTCC | | RSV, ALV | 909-930 | extension d'amorce | NRS, Δ76, LR9, ΔLR9 |
| **1617** | GGATGTGCCAACGGTTTTAG | | RSV | 838-819 | extension d'amorce | NRS, Δ76 |
| **2858** | GGATGTACCAACGGTTTTAG | | ALV | 818-838 | extension d'amorce | LR9, ΔLR9 |
| **3704** | GCCGAGGCTGTGGCGCAA | | ALV | 871-888 | extension d'amorce | ΔLR9 |
| **3826** | CATCTTCGCATCTCGCTG | | RSV | 779-796 | extension d'amorce | RSV |
| | | | | | | |
| **3096** | CCC**AAGCTT**GCGATTTAGGTGACACTATAGAAG AGGTCTCTCCCCCAGG | *Hind*III, promoteur SP6 | RSV | 691-705 | clonage | pAPTO NRS |
| **3097** | GC**TCTAGA**TGCTGCTCTCCCACC | *Xba*I | RSV | 937-951 | clonage | pAPTO NRS |
| **3098** | **GATC**CGGCTTAGTATAGCGAGGTTTAGCTACAC TCGTGCTGAGCC**G** | *Bam*HI | | | clonage | pAPTO |
| **3099** | **GATC**CGGCTCAGCACGAGTGTAGCTAAACCTCG CTATACTAAGCC**G** | *Bam*HI | | | clonage | pAPTO |
| **3173** | **AGCTT**GCGATTTAGGTGACACTATAG**T** | *Xba*I, promoteur SP6, *Hind*III | | | clonage | pPAPTO |
| **3196** | **CTAGA**CTATAGTGTCACCTAAATCGC**A** | *Xba*I, promoteur SP6, *Hind*III | | | clonage | pPAPTO |
| | | | | | | |

| Numéro | Séquence (5'-3') | site de restriction/ promoteur | séquence virale ciblée | nucléotides de la séquence ciblée | utilisation | Construction ou ARN produit/amplifié |
|---|---|---|---|---|---|---|
| **4311** | **GATC**CGGCCCATAAGATCATCGCTCACTCTGGG CC**CTGCA** | *Bam*HI, *Pst*I | | | clonage | pKt |
| **4336** | **G**GGCCCAGAGTGAGCGATGATCTTATGGGCC**G** | *Bam*HI, *Pst*I | | | clonage | pKt |
| **2953** | GC**TCTAGA**AATTTTGCTGCTCTCCCACCC | *Xba*I | RSV | 946-951 | clonage | pKt NRS, pKt Δ76 |
| **4312** | **AAGCTT**AAATTTAATACGACTCACTATAGGGCCC AGAGTGAGCGATG | *Hind*III, promoteur T7 | | | clonage | pKt NRS, pKt Δ76 |
| | | | | | | |
| **4672** | **GATCC**GGGATATCACTCAGCATAATTTAAA**GAG CT** | *Bam*HI, *Sac*I, promoteur T7 | | | clonage | pMS2 |
| **4673** | **C**TTTAAATTATGCTGAGTGATATCCC**G** | *Bam*HI, *Sac*I, promoteur T7 | | | clonage | pMS2 |
| **4666** | **GAGCTC**AAATTTAATACGACTCACTATAGGGTCT CTCCCCCAGGTCCGGA | *Sac*I | RSV | 691-711 | clonage | pNRS-MS2, pΔ76-MS2 |
| **4667** | **GGATCC**AATTTTGCTGCTCTCCCA | *Bam*HI | RSV | 939-956 | clonage | pNRS-MS2, pΔ76-MS2 |
| **5422** | GGAGCGGCGAATCGACA / TATGTGGGGAGTGGTTTGTATCCT | | RSV | 733-749/902-925 | Mutagenèse | p22NRS-MS2, pC2-22NRS |
| **5423** | AAACCACTCCCCACATA / TGTCGATTCGCCGCTCCGTTGCT | | RSV | 727-749/902-918 | Mutagenèse | p22NRS-MS2, pC2-22NRS |
| **5424** | CCTCCTCCTCCTTATGT / TGGCGGGGGTGGGAGAGCAGCAAA | | RSV | 890-906/930-953 | Mutagenèse | p24NRS-MS2, pC2-24NRS |
| **5425** | CTCTCCCACCCCCGCCA / ACATAAGGAGGAGGAGGAGCCGAG | | RSV | 883-906/930-946 | Mutagenèse | p24NRS-MS2, pC2-24NRS |
| **5426** | CCCAGGTCCGGAGTGCA / CAAAGGGGAGGAAGTGGGAGAAAC | | RSV | 700-716/748-771 | Mutagenèse | p26NRS-MS2, pC2-26NRS |
| **5427** | CCACTTCCTCCCCTTTG / TGCACTCCGGACCTGGGGGAGAGA | | RSV | 693-716/748-764 | Mutagenèse | p26NRS-MS2, pC2-26NRS |
| **5428** | GGAAGTGGGAGAAACAA / CCTCGGCTCCTCCTCCTCCTTATG | | RSV | 757-773/882-905 | Mutagenèse | p28NRS-MS2, pC2-28NRS |
| **5429** | AGGAGGAGGAGCCGAGG / TTGTTTCTCCCACTTCCTCCCCTT | | RSV | 750-773/882-898 | Mutagenèse | p28NRS-MS2, pC2-28NRS |
| | | | | | | |

| Numéro | Séquence (5'-3') | site de restriction/ promoteur | séquence virale ciblée | nucléotides de la séquence ciblée | utilisation | Construction ou ARN produit/amplifié |
|---|---|---|---|---|---|---|
| **5418** | CCC**AAGCTT**GGGAAATTTAATACGACTCACTATAGGGGTCTCTCTGGTTAGACCAG | *Hind*III, promoteur T7 | HIV-1 | 1-20 | clonage | pC2, pC1 |
| **5112** | **AGATCT**CCCATCTCTCTCCTTCTAGC | *Bgl*II | HIV-1 | 321-340 | clonage | pC2, pC1 |
| **5117** | **CTGCAG**GACAGCAGTACAAATGGC | *Pst*I | HIV-1 | 4330-4347 | clonage | pC1 |
| **5118** | **GAATTC**CCAGAGGAGCTTTGCTGGTC | *Eco*RI | HIV-1 | 4521-4540 | clonage | pC1 |
| **5140** | **CTGCAG**GGGAGTCTCCATAGAATGGA | *Pst*I | HIV-1 | 4871-4890 | clonage | pC2 |
| **5116** | **GAATTC**GATCCTACCTTGTTATGTCCTGC | *Eco*RI | HIV-1 | 5031-5053 | clonage | pC2 |
| **5113** | **AGATCT**GGTCTCTCCCCCAGGTCCGG | *Bgl*II | RSV | 934-956 | clonage | pC2-NRS, pC2-Δ76, pC1-NRS |
| **5114** | **CTGCAG**AATTTTGCTGCTCTCCCACCCCC | *Pst*I | RSV | 691-710 | clonage | pC2-NRS, pC2-Δ76, pC1-NRS |
| **5114** | **CTGCAG**AATTTTGCTGCTCTCCCACCCCC | *Pst*I | RSV | 934-956 | clonage | pC2-NRS, pC2-Δ76 |
| **5141** | **CTGCAG**GGTCTCTCCCCCAGGTCCGG | *Pst*I | RSV | 691-710 | clonage | pC2-NRSrev |
| **5142** | **AGATCT**AATTTTGCTGCTCTCCCACCCCC | *Bgl*II | RSV | 934-956 | clonage | pC2-NRSrev |
| **5824** | CATG**CCATGG**CATGGGTCTCTCTGGTTAGACCAG | *Nco*I | HIV-1 | 1-20 | clonage | pGL-C2, pGL-C2NRS, pGL-NRSrev, |
| **5825** | GC**TCTAGA**GCGATCCTACCTTGTTATGTCCTGC | *Xba*I | HIV-1 | 5031-5053 | clonage | pGL-C2, pGL-C2NRS, pGL-NRSrev, |
| | | | | | | |
| **5825** | GC**TCTAGA**GCGATCCTACCTTGTTATGTCCTGC | *Xba*I | HIV-1 | 4031-4053 | RT | |
| **6246** | TATAGCAGAGT // CAGTCGC | | HIV-1 | 283-289//4472-4482 | RT-PCR | |
| **6519** | GGTCTCTCTGGTTAGACCAG | | HIV-1 | 1-20 | RT-PCR | |
| **6187** | CGCTTTCAAGTCCCTGTT | | HIV-1 | 191-208 | RT-PCR | |
| | | | | | | |

| Numéro | Séquence (5'-3') | site de restriction/ promoteur | séquence virale ciblée | nucléotides de la séquence ciblée | utilisation | Construction ou ARN produit/amplifié |
|---|---|---|---|---|---|---|
| **6534** | TACGAGATCTACTAGTCGTACC | | ARN MS2 | 60-81 | Northern blot | |
| | | | | | | |
| **6524** | ACATCCGGAGTGCAATGGAT | | snRNA U1 humain | 61-80 | Northern blot | |
| **6525** | **TCTCTTGATGTCGATTCCGC** | | snRNA U11 Humain | 52-71 | Northern blot | |
| **6526** | **GTTCCTGGAGGTACTGCAAT** | | snRNA U2 Humain | 160-179 | Northern blot | |
| **6527** | **CGACCTTTACCCGCTCAAAA** | | snRNA U12 Humain | 75-96 | Northern blot | |

**Tableau V-1 : Oligonucléotides utilisés au cours de ce travail.**
La séquence des oligonucléotides est précisée, ainsi que la séquence à laquelle ils s'hybrident et leur utilisation. Les séquences des sites de restriction sont en caractère gras, les séquences des promoteurs sont soulignées et les oligonucléotides additionnels introduits à l'extrémité 3' de l'élément NRS sont doublement soulignés. Les bornes des troncations générées lors des mutagenèses sont indiquées par une barre oblique, la jointure des deux exons de l'ARNm amplifié par RT-PCR est indiquée par une double barre oblique.

## V. Les anticorps

Les anticorps utilisés pour les expériences de Western blot sont répertoriés dans le **Tableau V-3**.

| Nom | Clonalité | Utilisation | Dilution | Origine |
|---|---|---|---|---|
| anti-9G8 (1C6) | monoclonal de souris | Western blot | 1/15 000e | Euromedex |
| anti-ASF/SF2 (1D7) | monoclonal de souris | Western blot | 1/12 500e | |
| anti-SC35 (1SC-4F11) | monoclonal de souris | Western blot | 1/2 000 e | Millipore |
| anti-β Tubuline I (4G5) | monoclonal de souris | Western blot | 1/30 000 e | Sigma-Aldrich |
| anti-hnRNP A1 (4B10) | monoclonal de souris | Western blot | 1/40 000 e | Pierce |
| anti-hnRNP H (1G11) | monoclonal de souris | Western blot | 1/5 000 e | Sigma-Aldrich |
| anti-IgG souris | mouton | Western blot | 1/30 000 e | Jackson Immunoresearch |

**Tableau V-3** : Liste des anticorps utilisés au cours de ce travail.

## VI. Constructions génétiques

L'ADNc de plusieurs souches virales comportant l'élément NRS a été utilisé au cours de ce travail. Nous avons ainsi travaillé avec l'ADNc provenant de la souche Prague C du virus RSV ainsi que l'ADNc provenant de la souche LR-9 du virus ALV. Les séquences de ces virus sont respectivement accessibles à la GenBank sous les numéros d'accession AF033808 et AY350569. L'ensemble des constructions contenant les différents éléments NRS nous a été fourni par le Dr K. Beemon (**Figure V-4**).

### VI.1. Constructions renfermant les séquences de l'élément NRS du virus RSV de la souche Prague C

Les constructions contenant les séquences de l'élément NRS du virus RSV de la souche Prague C sont schématisées dans la **Figure V-4, A**. Les fragments XS et XS Δ76 du virus RSV PrC ont été insérés dans le plasmide pRSVNeo-Int (McNally *et al.*, 1991). Ce plasmide contient un mini gène composé des exons 2 et 3 et de l'intron 2, tous partiellement tronqués,

du gène c-*myc* de poulet. L'élément NRS est inséré dans l'intron à 340 nt en aval du site donneur. Ce plasmide est habituellement utilisé lors de transfection de cellules CEF ou HEK 293 afin de mesurer le taux d'inhibition de l'épissage par le NRS. Le fragment XS (nt 631-7006) contient l'élément NRS sauvage (nt 703-1006) et le fragment XSΔ76 contient l'élément NRS tronqué appelé NRSΔ76 (nt 703-930, Δ nt 799–874) défini comme étant l'élément NRS minimal fonctionnel (McNally *et al.*, 1991).

### VI.2. Constructions renfermant les séquences de l'élément NRS du virus ALV de la souche LR9

Les constructions contenant les séquences de l'élément NRS du virus ALV de la souche LR9 sont représentées dans la **Figure V-4, B**. La totalité du génome du virus ALV LR9 (nt 1-7547) ainsi que sa forme tronquée appelée ALV ΔLR9 (nt 1-7547, Δ nt 735-776) ont été clonées dans le vecteur pUC13 pour respectivement obtenir les plasmides pLR9 et pΔLR9 (Polony *et al.*, 2003). Le virus ALV LR9 est un virus recombinant qui comporte les séquences LTR du virus Ring Necked Pheasant Virus (RPV) et les gènes *gag*, *pol* et *env* du virus UR2 Associated Virus (UR2AV). Le plasmide pLR9 comporte l'élément NRS sauvage, appelé NRS LR9 (nt 703-930) et le plasmide pΔLR9 comporte l'élément NRS tronqué non fonctionnel appelé NRS ΔLR9 (nt 703-930, Δ nt 735-776).

### VI.3. Constructions réalisées pour la chromatographie d'affinité avec la technique tobramycine

Il a été nécessaire de réaliser une construction permettant de transcrire un ARN comportant la séquence de l'aptamère J6f1 fusionnée à l'élément NRS.

Dans un premier temps, j'ai construit un plasmide d'intérêt général dans lequel la séquence ADN de l'aptamère J6f1 a été clonée. Pour cela, j'ai utilisé deux oligonucléotides complémentaires chevauchants recouvrant la totalité de la séquence ADN correspondant à l'aptamère et portant à chacune de leurs extrémités des sites de restriction clivés par l'enzyme *Bam*HI (3098/3099, **Tableau V-1**). Après hybridation, le fragment a été directement inséré au site *Bam*HI du plasmide pUC18. Les plasmides obtenus pAPTO et pAPTOrev, correspondent, respectivement, à l'insertion dans les deux orientations de la séquence ADN de l'aptamère dans le plasmide pUC18 (**Figure V-5**). Dans ces plasmides, de nombreux sites de restriction uniques de la cassette de clonage sont présents de part et d'autre de la séquence ADN de l'aptamère. Ces sites peuvent alors être utilisés pour le clonage des séquences ADN correspondant aux ARN d'intérêt, que l'on veut fusionner à la séquence de l'aptamère.

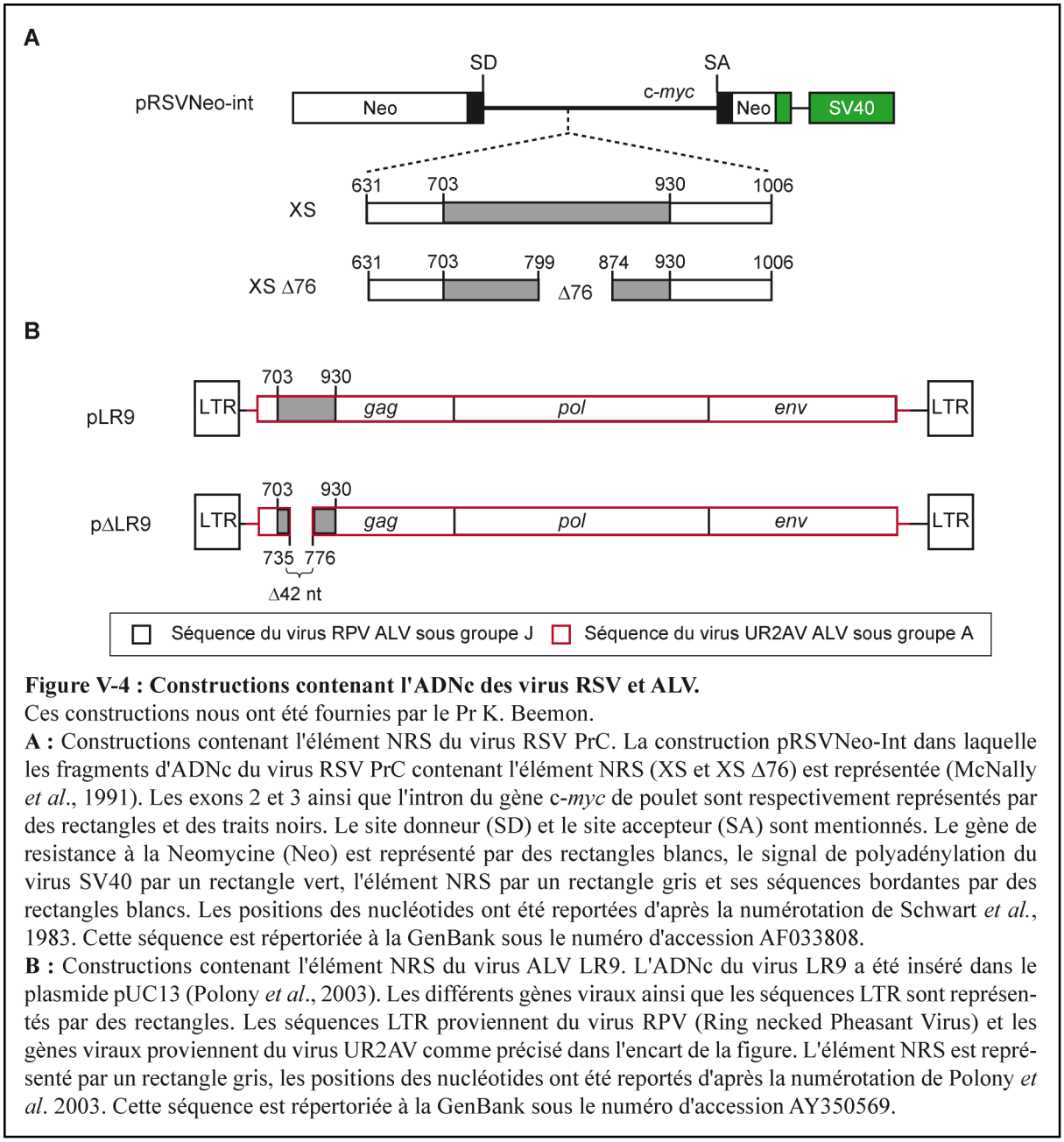



**Figure V-4 : Constructions contenant l'ADNc des virus RSV et ALV.**
Ces constructions nous ont été fournies par le Pr K. Beemon.
**A :** Constructions contenant l'élément NRS du virus RSV PrC. La construction pRSVNeo-Int dans laquelle les fragments d'ADNc du virus RSV PrC contenant l'élément NRS (XS et XS Δ76) est représentée (McNally *et al*., 1991). Les exons 2 et 3 ainsi que l'intron du gène c-*myc* de poulet sont respectivement représentés par des rectangles et des traits noirs. Le site donneur (SD) et le site accepteur (SA) sont mentionnés. Le gène de resistance à la Neomycine (Neo) est représenté par des rectangles blancs, le signal de polyadénylation du virus SV40 par un rectangle vert, l'élément NRS par un rectangle gris et ses séquences bordantes par des rectangles blancs. Les positions des nucléotides ont été reportées d'après la numérotation de Schwart *et al.*, 1983. Cette séquence est répertoriée à la GenBank sous le numéro d'accession AF033808.
**B :** Constructions contenant l'élément NRS du virus ALV LR9. L'ADNc du virus LR9 a été inséré dans le plasmide pUC13 (Polony *et al*., 2003). Les différents gènes viraux ainsi que les séquences LTR sont représentés par des rectangles. Les séquences LTR proviennent du virus RPV (Ring necked Pheasant Virus) et les gènes viraux proviennent du virus UR2AV comme précisé dans l'encart de la figure. L'élément NRS est représenté par un rectangle gris, les positions des nucléotides ont été reportés d'après la numérotation de Polony *et al*. 2003. Cette séquence est répertoriée à la GenBank sous le numéro d'accession AY350569.

Dans un deuxième temps, j'ai construit le plasmide pPAPTO qui permet de produire par transcription *in vitro* l'ARN aptamérique. La construction de ce plasmide a consisté à insérer la séquence du promoteur du phage SP6 en amont de la séquence ADN de l'aptamère contenue dans le plasmide pAPTO. L'insertion du promoteur a été réalisée en hybridant deux oligonucléotides complémentaires chevauchants correspondants à la totalité de la séquence du promoteur et portant à chacune de leurs extrémités des sites de restriction clivés par les enzymes *Hind*III et *Xba*I (oligonucléotides 3173/3196, **Tableau V-1**). Les oligonucléotides hybridés ont ensuite été directement insérés entre les sites *Hind*III et *Xba*I du plasmide

pAPTO. Le plasmide pPAPTO ainsi obtenu a été linéarisé par l'enzyme de restriction *Sma*I qui se situe en aval de la séquence ADN de l'aptamère pour produire par transcription *in vitro* l'ARN Tob (**Figure V-5**).

Dans un troisième temps, j'ai construit le plasmide pAPTO NRS qui permet de produire par transcription *in vitro* l'ARN NRS fusionné à l'aptamère Tob. Cette construction a consisté à insérer dans le plasmide pAPTO la séquence ADN correspondant à l'élément NRS précédé de la séquence du promoteur du phage SP6. Cette séquence a été amplifiée par PCR avec un couple d'oligonucléotides générant un site de restriction *Hind*III et un site *Xba*I (oligonucléotides 2955/3096, **Tableau V-1**). Le fragment ADN amplifié a été hydrolysé par ces deux enzymes et inséré entre les sites correspondants du plasmide pAPTO. Le plasmide pAPTO NRS ainsi obtenu permet de produire par transcription *in vitro* soit l'ARN de fusion NRS-Tob si le plasmide est linéarisé par l'enzyme *Sma*I, soit l'ARN NRS seul si le plasmide est linéarisé par l'enzyme *Bam*HI.

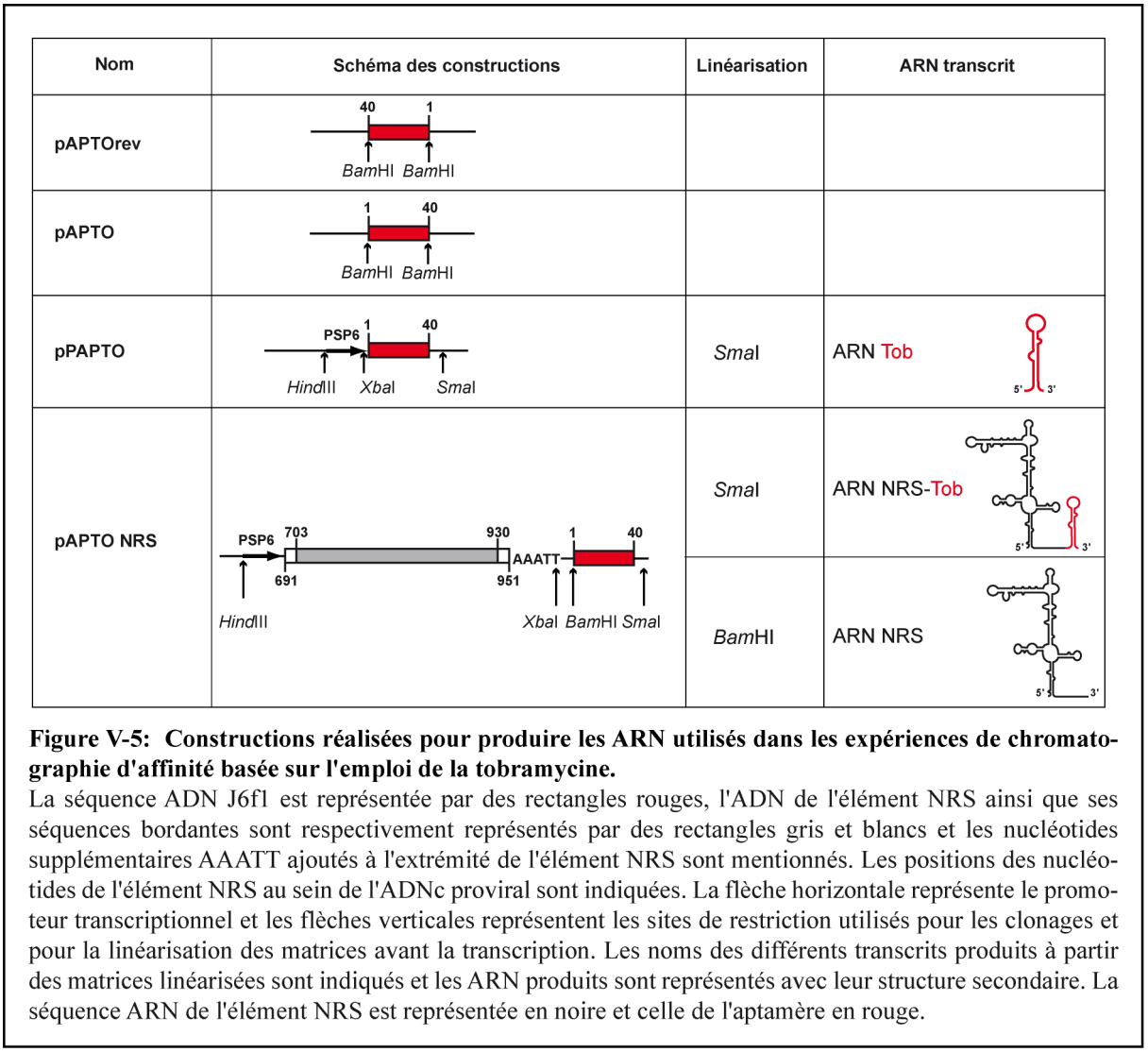



**Figure V-5: Constructions réalisées pour produire les ARN utilisés dans les expériences de chromatographie d'affinité basée sur l'emploi de la tobramycine.**
La séquence ADN J6f1 est représentée par des rectangles rouges, l'ADN de l'élément NRS ainsi que ses séquences bordantes sont respectivement représentés par des rectangles gris et blancs et les nucléotides supplémentaires AAATT ajoutés à l'extrémité de l'élément NRS sont mentionnés. Les positions des nucléotides de l'élément NRS au sein de l'ADNc proviral sont indiquées. La flèche horizontale représente le promoteur transcriptionnel et les flèches verticales représentent les sites de restriction utilisés pour les clonages et pour la linéarisation des matrices avant la transcription. Les noms des différents transcrits produits à partir des matrices linéarisées sont indiqués et les ARN produits sont représentés avec leur structure secondaire. La séquence ARN de l'élément NRS est représentée en noire et celle de l'aptamère en rouge.

### VI.4. Constructions réalisées pour la chromatographie d'affinité avec la technique MS2

Les constructions réalisées pour les expériences de chromatographie d'affinité reposant sur l'emploi de la protéine MS2 est représenté dans la **Figure V-6 et V-7**.

L'ARN MS2 qui a une grande affinité pour la protéine de l'enveloppe du phage MS2, est une séquence de 110 nucléotides qui est composée de trois séquences aptamériques identiques (Zhou *et al.*, 2002). Nous avons dans un premier temps construit un plasmide permettant de transcrire sous le contrôle du promoteur du bactériophage T7 cet ARN. Cette construction a été réalisée à partir du plasmide pADML-M3 qui est une construction d'épissage réalisée par l'équipe de R. Reed et qui nous a été fournie par l'équipe du Pr R. Lührmann. Dans cette construction, le fragment d'ADN correspondant à l'ARN MS2 est inséré entre les sites de restriction *Bam*HI et *Xba*I et fusionnée à l'extrémité 3' du mini-gène ADML (Zhou *et al.*, 2002). Après hydrolyse du plasmide pADML-M3 par les enzymes *Bam*HI et *Xba*I, le fragment d'ADN correspondant à l'ARN MS2 a été inséré dans le plasmide pUC18 préalablement linéarisé par les mêmes enzymes de restriction pour obtenir le plasmide pMS2.

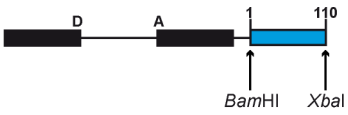
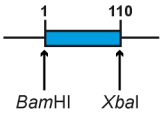
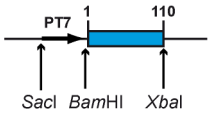
| Nom | Schéma des constructions | Linéarisation | ARN transcrit |
|---|---|---|---|
| pAdML-M3<br>Zhou *et al.*, 2002 |  | | |
| pMS2 |  | | |
| pT7MS2 |  | *Xba*I |  MS2 |

**Figure V-6: Constructions réalisées pour produire l'ARN MS2 servant de témoin.**
Les séquences exoniques et intronique du plasmide pADML-M3 sont respectivement représentées par des rectangles et des traits noirs et le site donneur (D) et accepteur (A) sont précisés. Les trois séquences répétées de la séquence aptamérique de la protéine de la capside du phage MS2 sont représentées par des rectangles bleus. La flèche horizontale représente le promoteur transcriptionnel et les flèches verticales représentent les sites de restriction utilisés pour les clonages et la linéarisation des matrices avant la transcription. Le nom du transcrit produit à partir de la matrice linéarisée est indiqué et l'ARN aptamérique MS2 produit est représenté selon sa structure secondaire en bleu.

J'ai ensuite inséré dans le plasmide pMS2 le promoteur du bactériophage T7 en amont de la séquence ADN correspondant à l'ARN MS2 (plasmide pT7MS2) (**Figure V-6**). Le clonage du promoteur a été réalisé en hybridant deux oligonucléotides complémentaires chevauchants et générant des sites de restriction *Sac*I et *Bam*HI (oligonucléotides 4676/4673, **Tableau V-1**) ce qui a permis son insertion entre ces sites dans le plasmide pMS2. La transcription du plasmide pT7MS2 linéarisé par *Xba*I par l'ARN polymérase de bactériophage T7 permet de produire l'ARN MS2.

Dans un deuxième temps, j'ai construit une série de plasmides permettant de produire par transcription *in vitro* les différents ARN NRS entier ou tronqués dont l'extrémité 3' était fusionnée à l'ARN MS2. Dans un premier temps, j'ai construit le plasmide pNRS WT-MS2 qui permet de transcrire l'ARN de fusion NRS WT-MS2. Pour cela, j'ai amplifié par PCR un fragment d'ADN correspondant au promoteur du bactériophage T7 suivi de la séquence du NRS. Le couple d'oligonucléotides utilisé pour l'amplification a permis d'ajouter respectivement en amont et en aval de la séquence NRS le sites *Sac*I suivi du promoteur T7 et le site *Bam*HI (oligonucléotides 4666 et 4667, **Tableau V-1**). Après hydrolyse par ces enzymes, le produit d'amplification a été inséré entre les sites correspondant du plasmide pMS2. Le plasmide pNRS WT-MS2 ainsi obtenu permet de produire par transcription *in vitro* l'ARN de fusion NRS WT-MS2 ou l'ARN NRS suivant s'il est linéarisé par l'enzyme *Xba*I ou l'enzyme *Bam*HI (**Figure V-7**).

Les plasmides permettant de transcrire les ARN NRS tronqués fusionnés à l'ARN MS2 (NRS Δ76-MS2, NRS 22-MS2, NRS 24-MS2, NRS 26-MS2, NRS 28-MS2) ont été produits par mutagenèse dirigée à partir du plasmide pNRS-MS2 en utilisant les oligonucléotides (**Tableau V-1**) et le protocole détaillé dans le **paragraphe VI.4.1**. Les troncations de l'élément NRS ont été délimitées à partir du modèle de structure secondaire établi pour l'élément NRS WT afin de conserver les différents éléments de structure secondaire comme nous l'expliquerons dans le chapitre I des Résultats. Comme pour le plasmide pNRS-MS2, chacune des constructions (pNRS Δ76-MS2, pNRS 22-MS2, pNRS 24-MS2, pNRS 26-MS2, pNRS 28-MS2) permet de produire par transcription *in vitro* l'ARN fusionné aux motifs MS2 ou l'ARN seul lorsque le site de restriction *Xba*I ou *Bam*HI est respectivement choisi pour la linéarisation.

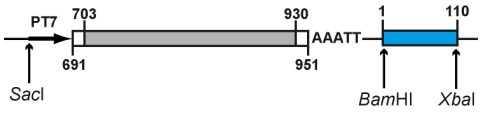
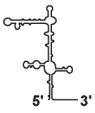
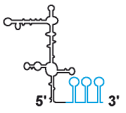
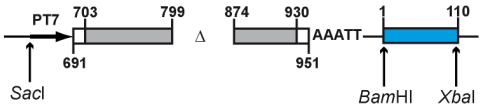
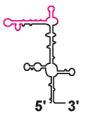
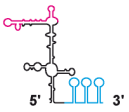
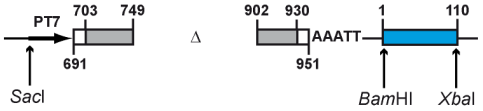
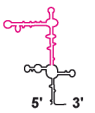
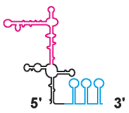
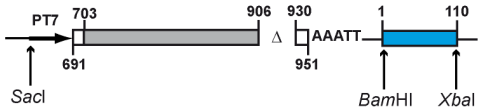
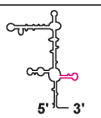
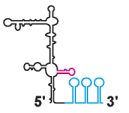
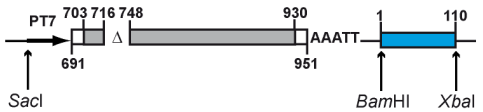
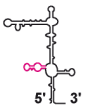
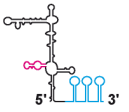
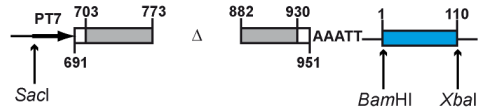
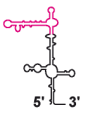
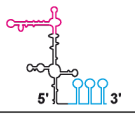
| Nom | Schéma des constructions | Linéarisation | ARN transcrit |
|---|---|---|---|
| pNRS-MS2 |  | *Bam*HI |  NRS |
| | | *Xba*I |  NRS-MS2 |
| pNRS Δ76-MS2 |  | *Bam*HI |  NRS Δ76 |
| | | *Xba*I |  NRS Δ76-MS2 |
| pNRS 22-MS2 |  | *Bam*HI |  NRS 22 |
| | | *Xba*I |  NRS 22-MS2 |
| pNRS 24-MS2 |  | *Bam*HI |  NRS 24 |
| | | *Xba*I |  NRS 24-MS2 |
| pNRS 26-MS2 |  | *Bam*HI |  NRS 26 |
| | | *Xba*I |  NRS 26-MS2 |
| pNRS 28-MS2 |  | *Bam*HI |  NRS 28 |
| | | *Xba*I |  NRS 28-MS2 |

**Figure V-7: Constructions réalisées pour produire les ARN de fusion utilisés dans les expériences de chromatographie d'affinité basée sur l'emploi de la protéine de la capside du phage MS2.**
L'ADN correspondant à l'ARN MS2 est représenté par des rectangles bleus, l'ADN de l'élément NRS ainsi que ses séquences bordantes sont respectivement représentés par des rectangles gris et blancs et les nucléotides supplémentaires AAATT ajoutés à l'extrémité du NRS sont mentionnés. Les positions des nucléotides de l'élément NRS au sein de l'ADNc proviral sont indiqués. La flèche horizontale représente le promoteur transcriptionnel et les flèches verticales représentent les sites de restriction utilisés pour les clonages et la linéarisation des matrices avant la transcription. Les noms des différents transcrits produits à partir des matrices linéarisées sont indiqués et les ARN produits sont représentés selon leur structure secondaire. La séquence ARN de l'élément NRS est colorée en noire, les séquences tronquées en rose et la séquence de l'ARN MS2 en bleu.

### VI.4.1. Réalisation de constructions génétiques par mutagenèse dirigée

Les constructions pNRS 22-MS2, pNRS 24-MS2, pNRS 26-MS2 et pNRS 28-MS2 comportent des éléments NRS tronqués qui ont été produits à partir de la construction pNRS WT-MS2 par la technique de mutagenèse dirigée par PCR. Le principe de cette technique repose sur le fait que l'ADN plasmidique sauvage, servant de matrice pour la PCR, a été produit chez *E. coli* XL1-blue et porte, par conséquent des méthylations. Ce plasmide est amplifié par l'ADN polymérase thermostable de l'Archae *Pyrococcus furiosus* (Pfu) avec deux oligonucléotides complémentaires qui s'hybrident de part et d'autre de la région à tronquer. Après amplification par PCR, le mélange réactionnel est incubé avec l'enzyme de restriction *Dpn*I qui reconnaît les sites méthylés et donc dégrade spécifiquement le plasmide sauvage ayant servi de matrice. Ensuite les bactéries sont transformées par les produits PCR restés intacts afin de produire de nouveaux plasmides.

Le mélange réactionnel est le suivant :

| | |
|---|---|
| Plasmide à muter | 10 ng |
| Tampon 10X | 5 µl |
| (Tris-HCl 200mM pH=8.0 ; $MgCl_2$ 20 mM, KCl 100mM ; $(NH4)_2SO_4$ 60mM ; Triton X100 1% ; BSA 100 mg/ml) | |
| Oligonucléotide 5' | 125 ng |
| Oligonucléotide 3' | 125 ng |
| Mélange des 4 dNTP (1,25 mM chacun) | 8 µl |
| Pfu (2,5 U/µl, Stratagène) | 1µl |
| $H_2O$ | qsp 50 µl |

Après un cycle de dénaturation de 5 minutes à 95°C, l'amplification se déroule en 18 cycles chacun composé de paliers de 30 secondes à 95°C, 1 minute à 54-58°C (dépendant du couple d'oligonucléotides), 8 minutes à 68°C. Les produits PCR sont visualisés par une électrophorèse sur gel d'agarose à 0,8 % (p/v). La matrice est ensuite digérée par 10 U d'enzyme *Dpn*I pendant 1 heure à 37°C. Enfin, les bactéries *E. coli* XL1-Blue, rendues compétentes par traitement au $CaCl_2$ (Sambrook *et al.*, 1989), sont transformées avec 1 µl de produit PCR traité par la *Dpn*I. Les oligonucléotides utilisés pour tronquer les différentes séquences de l'élément NRS sont récapitulés dans le **Tableau V-1** et les constructions contenant les troncations sont présentées dans la **Figure V-7**. La séquence du plasmide recombinant a été vérifiée par séquençage de l'ADN double brin (GexByWeb, Genome Express).

### VI.5. Construction réalisées pour l'épissage *in vitro*

Ces constructions ont été réalisées afin d'étudier la fonctionnalité de l'élément NRS sauvage ainsi que de ses différentes formes tronquées lors d'expérience d'épissage *in vitro*. Pour ce faire, nous avons inséré les différents éléments NRS du virus RSV dans les constructions C1 et C2. Ces dernières ont été réalisées dans le plasmide pBluescript KSII+ par L. Damier au cours de sa thèse au laboratoire (Damier *et al.*, 1997 ; Ropers *et al.*, 2004). Ces constructions sont des mini-gènes comportant des séquences d'ADNc du virus HIV-1 de la souche BRU/LAI. Les constructions C1 et C2 contiennent respectivement les couples de sites donneurs et accepteurs D1-A1 et D1-A2 et chacun de ces couples est respectivement séparé par un intron de 283 et 398 nucléotides composé de séquences d'ADNc non contiguës du virus HIV-1 qui sont jointes par le site de restriction *Bgl*II (**Figure V-8**). Après linéarisation par l'enzyme de restriction *Eco*RI, ces constructions permettent de produire par transcription *in vitro* les ARN C1 et C2.

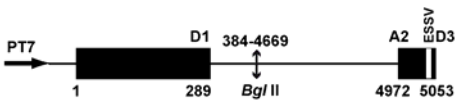
| Nom | vecteur | Schéma des constructions | Linéarisation |
|---|---|---|---|
| C1 (Damier *et al.*, 1997) | pBlueScript KSII+ |   | *Eco*RI |
| C2 (Damier *et al.*, 1997) | pBlueScript KSII+ |   | *Eco*RI |

**Figure V-8 : Constructions utilisées pour les expériences d'épissage *in vitro*.**
Constructions C1 et C2 qui ont été antérieurement réalisées au laboratoire (Damier *et al*., 1997). Ces constructions comprennent des fragments d'ADNc du virus HIV-1 de l'isolat BRU. Les exons et les introns sont respectivement représentés par des rectangles et des lignes noires, les sites donneurs et accepteurs sont mentionnés et leur position sont indiquées. La double flèche verticale représente le site *Bgl* II utilisé pour réaliser la jonction des fragments d'ADNc et la flèche horizontale représente le promoteur transcriptionnel du bactériophage T7. Le site de restriction utilisé pour linéariser la matrice avant la transcription est indiqué dans la colonne linéarisation.

Les constructions chimériques C1 NRS et C2 NRS consistent à insérer l'élément NRS dans la séquence intronique des constructions C1 et C2 en respectant les contraintes de distance compatibles d'une part avec la fonctionnalité de l'élément NRS lors de l'inhibition de l'épissage et d'autre part avec la taille des pré-ARNm utilisés lors des réactions d'épissage *in vitro*. Pour cela, l'élément NRS sauvage a été inséré de façon à ce que la séquence du pseudo site se situe à une distance de 278 nucléotides en aval du site donneur D1 et les tailles des introns des constructions C1 et C2 ont été respectivement raccourcies de 60 et 246 nucléotides.

L'ensemble des constructions chimériques réalisées pour les expériences d'épissage *in vitro* est illustré dans la **Figure V-9**.

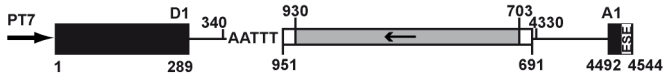
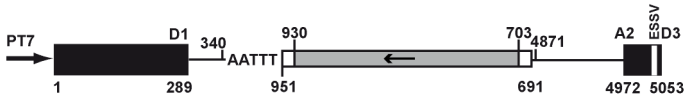
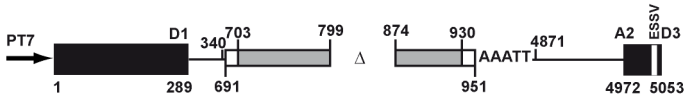
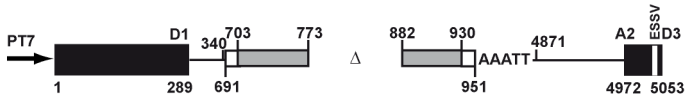
| Nom | vecteur | Schéma des constructions | Linéarisation |
|---|---|---|---|
| pC1-NRS | pUC 18 |  | *Eco*RI |
| pC1-NRS Inv | pUC 18 |  | *Eco*RI |
| pC2-NRS WT | pUC 18 |  | *Eco*RI |
| pC2-NRS Inv | pUC 18 |  | *Eco*RI |
| pC2-NRS Δ76 | pUC 18 |  | *Eco*RI |
| pC2-NRS 22 | pUC 18 |  | *Eco*RI |
| pC2-NRS 24 | pUC 18 |  | *Eco*RI |
| pC2-NRS 26 | pUC 18 |  | *Eco*RI |
| pC2-NRS 28 | pUC 18 |  | *Eco*RI |

**Figure V-9 : Constructions réalisées pour les expériences d'épissage *in vitro*.**
Les schémas représentent les constructions C1 et C2 dans lesquelles les différents éléments NRS ont été insérés dans l'intron. Les exons et les introns sont respectivement représentés par des rectangles et des lignes noires, les sites donneurs et accepteurs sont mentionnés. L'élément NRS est représenté par des rectangles gris, ses séquences bordantes par des rectangles blancs et les nucléotides supplémentaires AAATT ajoutés à l'extrémité du NRS sont mentionnés. Les séquences tronquées et les positions des nucléotides des séquences bordantes sont précisées. La flèche horizontale représente le promoteur transcriptionnel du bactériophage T7. Le site de restriction utilisé pour linéariser la matrice avant la transcription *in vitro* est indiqué dans la colonne linéarisa-

Les constructions chimériques C1 NRS et C2 NRS ont été réalisées en deux temps. Dans un premier temps, la séquence d'ADNc du virus HIV-1 correspondant à la seconde partie de l'intron en aval du site d'insertion de l'élément NRS (nt 4330-4491 pour la construction C1 et nt 4871-4971 pour la construction C2) ainsi que le second exon (nt 4492-4544 pour la construction C1 et nt 4972-5053 pour la construction C2) ont respectivement été amplifiés par PCR avec les oligonucléotides 5117-5118 et 5140-5116 portant les sites *Pst*I et *Eco*RI à leur

extrémité. Les deux produits d'amplifications ont été digérés par les enzymes de restriction *Pst*I et *Eco*RI et clonés dans le plasmide pUC18 pour respectivement obtenir les plasmides pA1 et pA2. Dans un deuxième temps, les séquences d'ADNc du virus HIV-1 correspondant à l'exon 1 et la première partie de l'intron ont été amplifiés par PCR avec des oligonucléotides 5418 et 5122 comportant respectivement le promoteur T7, le site *Hind*III et le site *Bgl*II à leur extrémité. Parallèlement, le fragment NRS sauvage a été amplifié par PCR avec le couple d'oligonucléotides 5113 et 5114 comportant respectivement les sites de restriction *Bgl*II et *Pst*I à leur extrémité. Ces deux produits PCR ont été respectivement digérés par les enzymes *Hind*III-*Bgl*II et *Bgl*II-*Pst*I et respectivement clonés dans le plasmide pA1 et pA2 préalablement digérés avec les enzymes de restriction *Hind*III et *Pst*I pour respectivement obtenir les plasmides pC1-NRS WT et pC2-NRS WT. La même stratégie a été adoptée pour créer les plasmides pC1-NRS rev, pC2-NRS rev et pC2-NRS Δ76. Les autres constructions chimériques comprenant les autres formes tronquées de l'élément NRS, à savoir les constructions pC2-NRS 22, pC2-NRS 24, pC2-NRS 26 et pC2-NRS 28 ont été obtenues en réalisant des mutagenèses dirigées sur le plasmide pC2-NRS WT avec les mêmes couples d'oligonucléotides ayant précédemment servi à tronquer les mêmes segments de l'élément NRS pour les fusionner à la séquence MS2. L'ensemble des oligonucléotides ayant servi à réaliser ces clonages est détaillé dans le **Tableau V-1**.

### VI.6. Construction réalisées pour l'épissage *in cellulo*

La stratégie de clonage utilisée pour construire les mini gènes chimériques pour l'étude de l'inhibition de l'épissage par l'élément NRS *in cellulo* est illustrée dans la **Figure V-10** et les différentes constructions réalisées sont détaillées dans la **Figure V-11**. La construction C2 ainsi que les constructions chimériques où les différents éléments NRS ont été insérés dans l'intron de la construction C2 (à savoir, pC2-NRS WT, pC2-NRS rev, pC2-NRS Δ76, pC2-NRS 22, pC2-NRS 24, pC2-NRS 26, pC2-NRS 28) ont été isolés à partir du plasmide pUC18 et insérés dans le plasmide pGL3c pour obtenir respectivement les construction pGL-C2, pGL-C2 NRS WT, pGL-C2 NRS rev, pGL-C2 NRS Δ76, pGL-C2 NRS 22, pGL-C2 NRS 24, pGL-C2 NRS 26 et pGL-C2 NRS 28 (**Figure V-11**). Pour ce faire, chacun des minis gènes a été amplifié par PCR avec le couple d'oligonucléotides 5824 et 5825 comportant respectivement les sites de restriction *Nco*I et *Xba*I à leur extrémité. Les produits PCR ont ensuite été digérés par les enzymes de restriction *Nco*I et *Xba*I et clonés dans le plasmide pGL3c préalablement digéré par les mêmes enzymes de restriction. La transcription et la

polyadénylation du pré-ARNm du mini gène chimérique sont placés respectivement sous le contrôle du promoteur et du signal de polyadénylation du virus SV40.

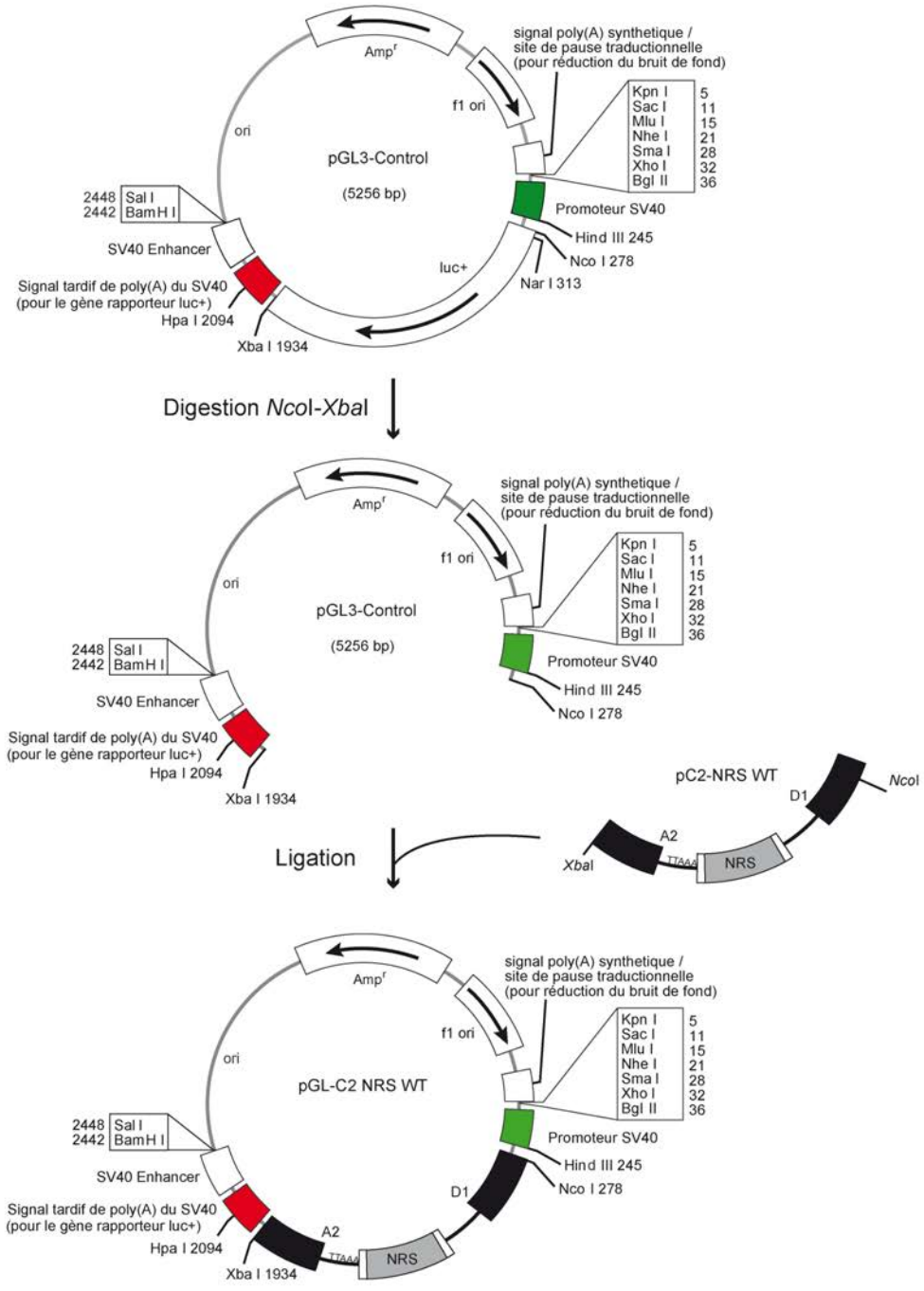



**Figure V-10: Stratégie de clonage des constructions réalisées pour les expériences d'épissage *in cellulo***
Stratégie de clonage utilisée pour réaliser le vecteur pGL-C2 NRS WT. La construction pC2-NRS WT a été digérée par les enzymes de restriction *Nco*I et *Xba*I et insérée dans le vecteur pGL3c préalablement digéré par les mêmes enzymes pour réaliser la construction pGL-C2 NRS WT. Les autres constructions comprenant les différents éléments NRS ont été réalisées suivant la même stratégie. Les rectangles et les lignes noires représentent respectivement les exons et les introns, le site accepteur et le site donneur sont précisés. L'élément NRS ainsi que ses séquences bordantes sont respectivement représentés par des rectangles gris et blancs et les nucléotides supplémentaires AAATT ajoutés à l'extrémité du NRS sont mentionnés. Le mini-gène C2-NRS WT est sous le contrôle du promoteur SV40 qui est représenté par un rectangle vert et la polyadénylation est dirigée par le signal de polyadénylation du SV40 qui est représenté par un rectangle rouge.

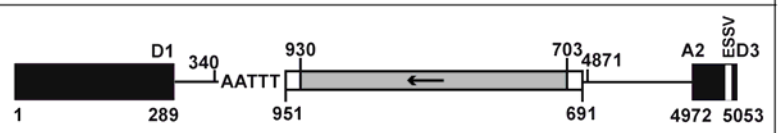
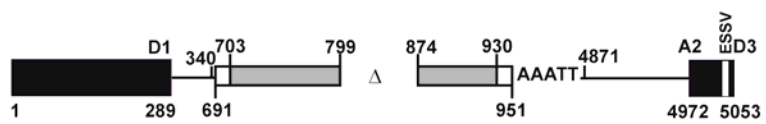
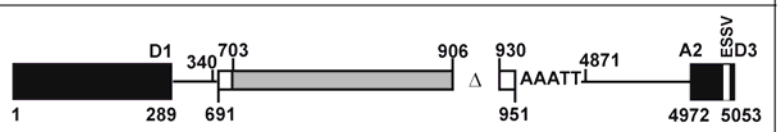
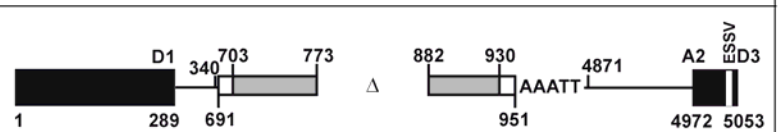
| Nom | vecteur | Schéma des constructions |
|---|---|---|
| pGL-C2 | pGL3c |  |
| pGL-C2 NRS WT | pGL3c |  |
| pGL-C2 NRS Inv | pGL3c |  |
| pGL-C2 NRS Δ76 | pGL3c |  |
| pGL-C2 NRS 22 | pGL3c |  |
| pGL-C2 NRS 24 | pGL3c |  |
| pGL-C2 NRS 26 | pGL3c |  |
| pGL-C2 NRS 28 | pGL3c |  |

**Figure V-11: Constructions réalisées pour les expériences d'épissage *in cellulo***
Les schémas représentent la construction C2 ainsi que les constructions C2-NRS qui ont été insérées dans le vecteur pGL3c entre les sites *Nco*I et *Xba*I. Les exons et les introns sont respectivement représentés par des rectangles et des lignes noires, les sites donneurs et accepteurs sont mentionnés et leur position sont indiquées. Dans la construction pGL-C2, la double flèche verticale représente le site *Bgl* II utilisé pour réaliser la jonction des fragments d'ADNc. L'élément NRS ainsi que ses séquences bordantes sont respectivement représentés par des rectangles gris et blancs et les nucléotides supplémentaires AAATT ajoutés à l'extrémité du NRS sont mentionnés. Les séquences tronquées de l'élément NRS et les positions des nucléotides séquences bordantes sont précisées.

## VI.7. Constructions renfermant les séquences codant la protéinet MPB-MS2

Le plasmide pMAL comporte la séquence de la protéine MS2 de l'enveloppe du phage MS2 qui a été fusionnée traductionnellement à la séquence de l'extrémité N-terminale de la protéine MBP (Zhou *et al.*, 2002). Cette construction nous a été fournie par le Pr R. Lührmann.

# Partie : Méthodes

## VII. Technique d'analyse de la structure secondaire des ARN

### VII.1. Préparation de la matrice ADN à transcrire

Dans le cadre de l'étude de la structure secondaire des différents éléments NRS, les matrices ADN utilisées pour la transcription ont été obtenues après une amplification par PCR avec le couple d'oligonucléotides 2954 et 2955 (**Tableau V-1**). L'oligonucléotide 2954 comporte à son extrémité 5' la séquence du promoteur du phage T7 qui permet la transcription des matrices ADN. Les éléments NRS amplifiés par PCR sont compris entre les nucléotides 691 et 951 (**Figure V-12**).

Le mélange réactionnel est le suivant :

| | |
|---|---|
| Plasmide | 10 ng |
| Oligonucléotide 5' (100 µM) | 0,5 µl |
| Oligonucléotide 3' (100 µM) | 0,5 µl |
| Mélange de dNTP (1, 25 mM de chaque) | 8 µl |
| Tampon Go Taq® 5X (Promega) | 10 µl |
| Go Taq polymérase® (5 U/µl, Promega) | 0,25 µl |
| $H_2O$ | qsp 50µl |

Après un cycle de dénaturation de 5 minutes à 95°C, l'amplification se déroule en 30 cycles qui sont chacun composés de paliers de 30 secondes à 95°C, 30 secondes à 55°C et 1 minute à 72°C et se termine par une incubation de 10 minutes à 72°C. Les produits PCR sont ensuite soumis à une extraction au phénol-chloroforme (1/1) puis à une extraction par un mélange chloroforme-alcool isaomylique (24/1). L'ADN est ensuite précipité par addition d'1/10$^{e}$ de volume d'acétate de sodium (0,3 M, concentration finale) et 2 volumes d'éthanol 96% puis est lavé avec de l'éthanol 70%. Le culot est ensuite séché et repris dans l'$H_2O$ et le volume ajusté de façon à ce que la concentration soit de 1,5 pmol/µl.

| Nom | Publication | vecteur | Virus | Oligonucléotides utilisés pour l'amplification par PCR | Schématisation des éléments NRS amplifiés par PCR | Matrice obtenue |
|---|---|---|---|---|---|---|
| pXS Neo-Int | McNally *et al.*, 1991 | pNeo-Int | RSV PrC | 2954-2955 |   | NRS WT |
| pXSΔ76 Neo-Int | McNally *et al.*, 1991 | pNeo-Int | RSV PrC | 2954-2955 |   | NRS Δ 76 |
| pLR9 | Polony *et al.*, 2003 | pUC 13 | ALV LR9 | 2954-2955 |   | LR9 |
| pΔLR9 | Polony *et al.*, 2003 | pUC 13 | ALV LR9 | 2954-2955 |   | Δ LR9 |

**Figure V-12 : Constructions utilisées dans la cadre de l'étude structurale de l'élément NRS**

Toutes ces constructions nous ont été fournies par le Dr K. Beemon. Les publications dans lesquelles ces constructions ont été utilisées sont mentionnées.

Le nom des constructions, le vecteur de clonage, l'origine de l'ADN proviral contenu dans les constructions sont précisés. Les différents éléments NRS ont été amplifiés par PCR avec les couples d'oligonucléotides indiqués. Les schémas représentent les matrices PCR utilisées pour la transcription *in vitro*. La flèche horizontale indique le promoteur transcriptionnel, les positions des nucléotides au sein de l'ADNc proviral sont indiquées ainsi que les sites de délétions. Les rectangles gris et noirs délimitent respectivement les éléments NRS des virus RSV et ALV, les rectangles blancs représentent les séquences bordantes de l'élément NRS. Les nucléotides supplémentaires AAATT qui ont été ajoutés à l'extrémité 3' de la matrice sont mentionnés.

### VII.2. Transcription *in vitro*

L'ADN est transcrit à partir du promoteur de l'ARN polymérase du phage T7 selon les conditions suivantes :

| | |
|---|---|
| Produit PCR | 1,5 pmol |
| Tampon de transcription 3X<br>(60 mM $MgCl_2$ ; 50 mM NaCl ; 15 mM DTT ; 3 mM spermidine ; 240 mg/ml PEG 8000 ; 0.03% Triton 100X (v/v) ; 120 mM Tris-HCl pH 8.0) | 10 µl |
| Mélange des 4 rNTP (12,5 mM chacun) | 10 µl |
| BSA (2mg/ml) | 1µl |
| RNasin (20 U/µl, Fermentas) | 1 µl |
| ARN T7 polymérase (produite au laboratoire) | 3 µl |
| $H_2O$ | qsp 30 µl |

Le mélange est incubé 2h à 37°C puis la matrice ADN est dégradée par l'ajout de 5 U de TURBO$^{TM}$ DNase exempte de RNase (Ambion) pendant 30 minutes à 37°C. Les ARN sont ensuite purifiés par électrophorèse en gel dénaturant de polyacrylamide 5%-urée 8 M. Après la migration, les transcrits sont visualisés sous UV ($\lambda$ = 254 nm), la bande du gel le contenant est découpée et le transcrit est élué dans le tampon d'élution (NaCl 250 mM ; EDTA 10 mM ; Tris-HCl 20 mM pH 7,5) à 4°C pendant 12h. Le transcrit élué est ensuite précipité avec de l'éthanol 96 %, et lavé avec de l'éthanol 70%. Le culot est ensuite séché et repris dans de l'$H_2O$, la concentration est calculée à partir de l'absorbance à 260 nm et les transcrits sont conservés à -80°C.

### VII.3. Analyse de la structure secondaire des ARN par des sondes chimiques et enzymatiques (Ehresmann *et al.*, 1987 ; Qu *et al.*, 1983)

#### VII.3.1. Principe

Pour identifier les régions simple brin ou double brin de l'ARN, nous avons réalisé des modifications chimiques (par le diméthylsulfate ou DMS et le kéthoxal) et des digestions enzymatiques (par les RNases T1, T2, et V1) en conditions statistiques. Les positions des bases auxquelles les sondes enzymatiques et chimiques clivent ou modifient l'ARN sont illustrées dans la **Figure V-13**.

Les hydrolyses enzymatiques et les modifications chimiques sont réalisées dans des conditions ménagées en présence d'ARNt, de manière à avoir en moyenne une coupure ou une modification par molécule d'ARN. Les liaisons phosphodiesters hydrolysées par les enzymes et les bases modifiées par les sondes chimiques au niveau des bases impliquées dans

les appariements Waston-Crick sont mises en évidence par l'extension d'une amorce oligonucléotidique par la transcriptase inverse. L'oligonucléotide est marqué en 5' au $^{32}P$ suivant le protocole décrit dans le **paragraphe VII.4.2**. En parallèle, une extension d'amorce oligonucléotidique par la transcriptase inverse est effectuée sur l'ARN n'ayant subi aucune modification ou clivage.

Cela permet de différencier les arrêts d'élongation dus aux modifications chimiques et coupures enzymatiques, des pauses de la transcriptase inverse en cours d'élongation. Les produits d'élongation sont fractionnés par une électrophorèse en gel de polyacrylamide dénaturant, en parallèle avec les produits de la réaction de séquençage réalisée avec la transcriptase inverse et le même oligonucléotide amorce. La position des bases modifiées et des coupures enzymatiques est déduite par comparaison avec le séquençage.

### VII.3.2. Conditions expérimentales

Les modifications chimiques et enzymatiques sont effectuées sur 3 pmol de transcrit synthétisé *in vitro*. La qualité du transcrit est préalablement vérifiée par un séquençage avec la transcriptase inverse. Les expériences ont été répétées au moins deux fois afin de vérifier leur reproductibilité.

### VII.3.3. Modifications chimiques

Avant de réaliser les modifications chimiques au DMS, 3 pmol d'ARN sont dilués dans 50 µl de tampon cacodylate (KCl 300 mM ; $MgCl_2$ 10 mM ; cacodylate de sodium 50 mM, pH 7,5) dans lequel la réaction sera réalisée, et sont pré-incubés 5 minutes à 65°C puis 10 minutes à 20°C. Les modifications au DMS (Aldrich) sont réalisées de façon ménagée par ajout de 0,5 ou 1 µl d'une solution de DMS 100% diluée au ¼ dans de l'éthanol absolu (concentration finale 0,03 M) par essai. Après une incubation de 10 minutes à 20°C, les réactions sont arrêtées par l'ajout de 10 µl de solution G-stop (Acétate de sodium 1,5 M, pH 5,3 ; β-mercaptoéthanol 1 M ; Tris-acétate 1 M, pH 7,5) suivi d'une précipitation à l'éthanol 96 %. L'ARN est ensuite lavé à l'éthanol 70%, séché puis dissous dans 4 µl d'$H_2O$ stérile. Un µl de cette solution d'ARN est utilisé pour une réaction d'élongation avec la transcriptase inverse, dans les conditions décrites au paragraphe **VII.4.3**.

**A**

RNase spécifique des régions en double brin ou des régions où les bases sont empilées :

RNase V1

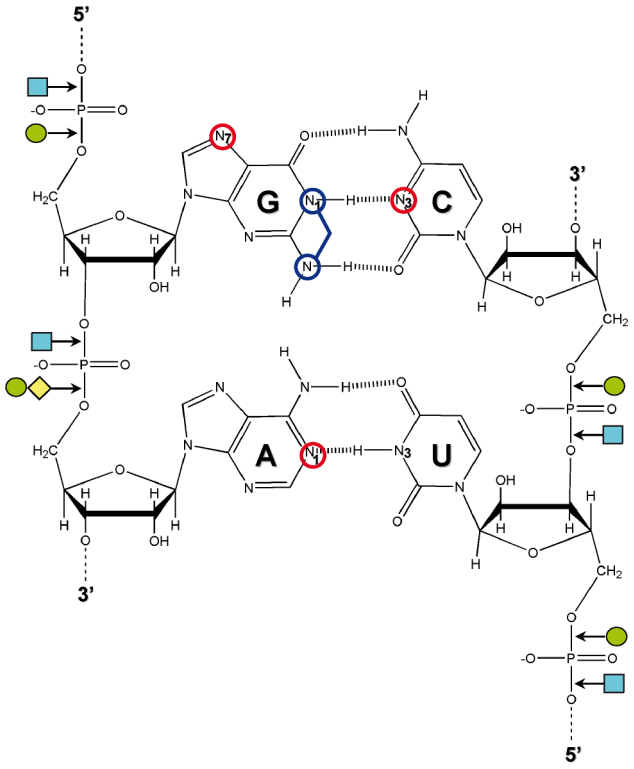
RNase et sondes chimiques spécifiques des régions simple brin :

RNase T1

RNase T2

DMS

Kéthoxal




**B**

| Sondes enzymatiques | Cibles | Produit généré |
|---|---|---|
| RNAse T1 | G non apparié | ...Gp (3'p) |
| RNAse T2 | N non apparié | ...Np (3'p) |
| RNAse V1 | N apparié | pN... (5'p) |
| **Sondes chimiques** | **Cibles** | **Produit généré** |
| DMS | $N^1A$<br>$N^3C$<br>$N^7G$ | $N^1$-$CH_3$<br>$N^3$-$CH_3$<br>$N^7$-$CH_3$ |
| Kéthoxal | $N^1G$<br>$N^2G$ | $N^1$-CHOH<br>$N^2$-CROH |

**Figure V-13: Sondes enzymatiques et chimiques utilisées pour l'étude de la structure secondaire des ARN en solution**

**A :** Schématisation des positions de clivages par les sondes enzymatiques ou de modifications par les sondes chimiques. Les sites de coupures enzymatiques et de modifications chimiques sont respectivement représentés par des flèches et des cercles.Les sondes enzymatiques et chimiques s'y rapportant sont schématisées comme mentionnée dans la légende de la figure.

**B :** Caractéristiques fonctionnelles des sondes enzymatiques et chimiques. Les cibles des sondes enzymatiques et chimiques et les produits générés sont mentionnés.

Les ARN traités par le kéthoxal sont incubés dans le même mélange que les ARN traités au DMS. Les ARN sont également pré-incubés 5 minutes à 65°C puis 10 minutes à 20°C. Les modifications au kéthoxal sont réalisées par l'ajout de 1 ou 2 µl de kéthoxal à 37 mg/ml (GE Healthcare), dilué au ¼ dans de l' $H_2O$. Après 10 minutes d'incubation à 20°C, les réactions sont arrêtées par ajout de 10 µl de borate de potassium 0,5 M, pH 7,0. Un µl est utilisé par réaction d'élongation avec la transcriptase inverse, dans les conditions décrites au paragraphe **VII.4.3**.

### VII.3.4. Hydrolyses enzymatiques

Avant l'ajout des RNases, l'ARN transcrit *in vitro* est incubé dans le mélange réactionnel suivant :

| | |
|---|---|
| ARN transcrit *in vitro* | 3 pmol |
| Tampon D<br>(EDTA 0,2 mM ; KCl 0,1 M ; DTT 0,5 mM ; PMSF 0,25 mM ; HEPES-KOH 20 mM pH 7,9 ; glycérol 20 %) | 3,6 µl |
| $MgCl_2$ (62,5 mM) | 1 µl |
| ARNt (5 µg/µl) | 1 µl |
| $H_2O$ | qsp 14 µl |

Le mélange est incubé 10 minutes à 65°C puis 10 minutes à température ambiante pour permettre la renaturation de l'ARN qui est ensuite figée par un court refroidissement dans la glace. L'ARN est ensuite hydrolysé enzymatiquement par incubation 5 minutes à 30°C en présence de $1.10^{-4}$ U de RNase V1 (Kemothex), de 1 U de RNase T1 (Roche Diagnostics) ou de 0,5 U de RNase T2 (Invitrogen). Les digestions par les RNase T1, T2 et V1 sont arrêtées par l'ajout de 20 µg d'ARNt de levure auxquels est ajouté 1 µl d'EDTA 100 mM pour arrêter les digestions par la RNase V1. Les ARN sont ensuite soumis à une extraction au phénol-chloroforme (24/1) puis au chloroforme-alcool isoamylique (24/1), précipités à l'éthanol absolu en présence de 0,3 M final d'acétate de sodium, lavés à l'éthanol 70 % et repris dans 4 µl d' $H_2O$. Les analyses par élongation d'amorce par la transcriptase inverse sont réalisées avec 1 µl de chacune des préparations d'ARN obtenues après hydrolyse enzymatique.

## VII.4. Analyse par transcription inverse

### VII.4.1. Principe (Figure V-14)

Les coupures des sondes enzymatiques et les modifications des sondes chimiques sur l'ARN sont identifiées par une réaction de transcription inverse. Pour cela, un oligonucléotide marqué en 5' au $P^{32}$ est hybridé à une région déterminée de l'ARN. Il sert d'amorce à la

transcriptase inverse qui synthétise, en présence de 4 désoxynucléotides (dNTP), un brin d'ADN complémentaire à l'ARN. Une seconde réaction de transcription inverse est réalisée avec un ARN n'ayant subi aucune coupure enzymatique ou modification chimique pour séquencer l'ARN. Dans ce cas, des didésoxynucléotides (ddNTP) sont ajoutés au milieu réactionnel à une concentration donnée pour que la synthèse du brin d'ADN soit arrêtée statistiquement à chaque base. Une autre réaction de transcription inverse réalisée avec un ARN n'ayant subi aucune coupure enzymatique ou modification chimique en l'absence de ddNTP et sert de témoin permettant d'identifier les pauses naturelles de la transcriptase inverse.

### VII.4.2. Marquage radioactif des oligonucléotides en 5'

Les oligonucléotides sont marqués radioactivement en 5' par la polynucléotide kinase du bactériophage T4 qui catalyse le transfert du phosphate en position γ de l'ATP [γ-$^{32}$P] sur l'extrémité 5'OH de l'oligonucléotide. Le mélange réactionnel utilisé pour la réaction de phosphorylation est le suivant :

| | |
|---|---|
| [γ-$^{32}$P] ATP (10 mCi/ml, 3000 Ci/mmol, GE Healthcare) | 1µl |
| Oligonucléotide (100 µM) | 15 pmol |
| Tampon T4 10X (Fermentas) | 1 µl |
| ($MgCl_2$ 0,1 M ; DTT 50 mM ; Tris-HCl 0,5 M pH 7,6 ; spermidine 1 mM ; EDTA 1 mM) | |
| T4 Polynucléotide kinase (T4 PNK, 10 U/µl, Fermentas) | 1 µl |
| $H_2O$ | qsp 10 µl |

Après une incubation de 45 minutes à 37°C, le mélange est chauffé pendant 10 minutes à 65°C pour inactiver la kinase, puis 10 µl de bleu de formamide (EDTA 125 mM ; bleu de bromophénol 0,03% ; bleu de xylène cyanol 0,03 %, le tout dans 80% de formamide désionisée) sont ajoutés et le mélange est déposé sur un gel de polyacrylamide 10%-urée 8M dénaturant. L'électrophorèse est réalisée à 17 V/cm dans du tampon TBE 1X. A la fin de l'électrophorèse, un film d'autoradiographie est exposé sur le gel et la bande contenant l'oligonucléotide est découpée et incubée dans du tampon d'élution (NaCl 250 mM ; EDTA 10 mM ; Tris-HCl 20 mM pH 7,5) à 37°C durant 12h. L'oligonucléotide élué est ensuite précipité en présence d'éthanol 96% et de 10 µg de glycogène, le culot est ensuite lavé avec de l'éthanol 70%, et séché. La radioactivité contenue dans le culot est mesurée au compteur à scintillation en mode Cerenkov et le culot est repris dans de l'$H_2O$ à raison de 100 cps/µl.

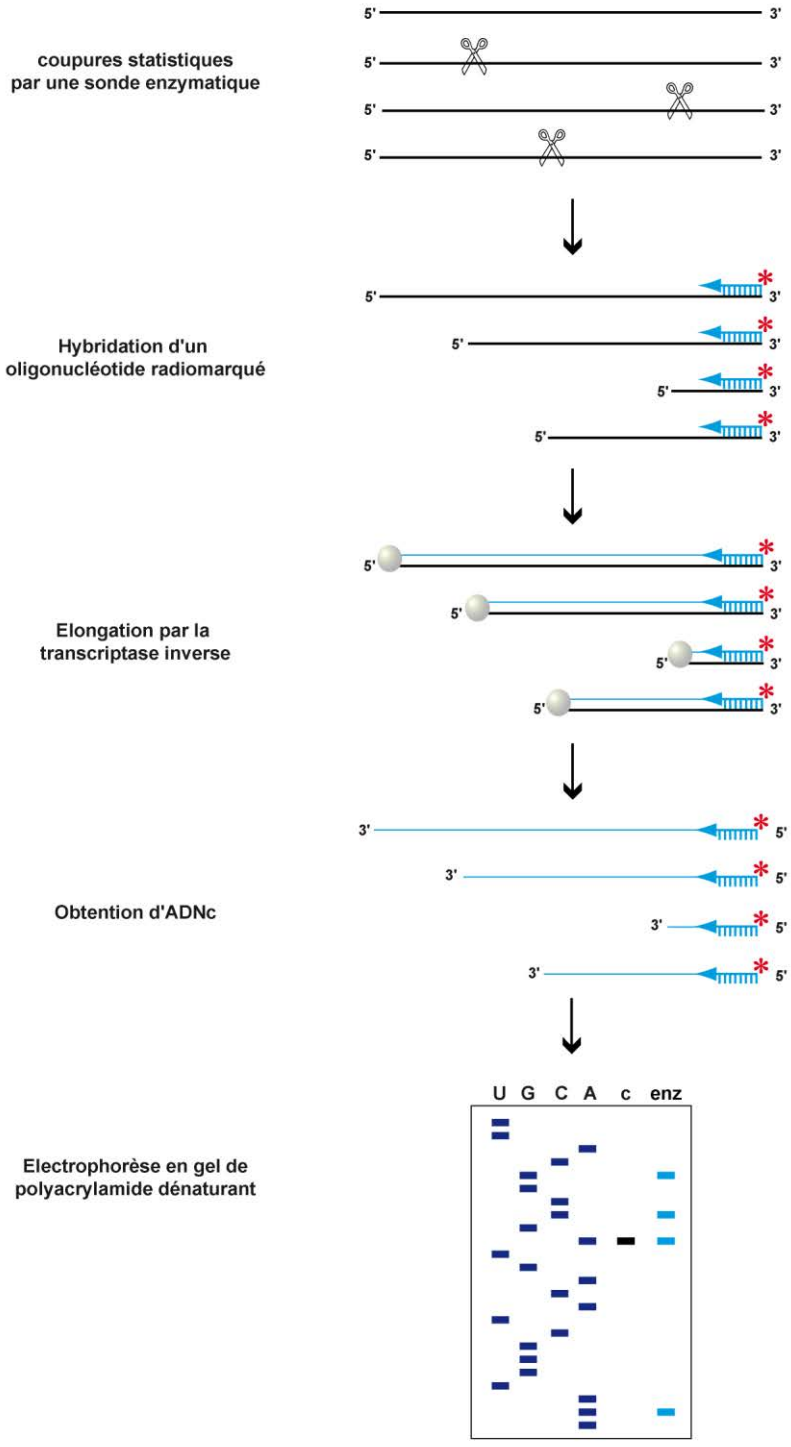



**Figure V-14: Principe de l'identification des coupures d'un ARN par une sonde enzymatique par la transcription inverse.**

Les paires de ciseaux représentent les coupures statistiques de l'ARN par une sonde enzymatique. L'ARN est représenté par un trait noir, l'ADNc par un trait bleu. Le marquage radioactif en 5' de l'oligonucléotide utilisé pour la réaction de transcription inverse est symbolisé par une étoile rouge. La transcriptase inverse est représentée par une sphère grise. Les pistes U,G,C,A correspondent au séquençage de l'ARN, la piste c correspond à la piste contrôle, la piste enz correspond à la digestion de l'ARN par une sonde enzymatique.

### VII.4.3. Réaction de transcription inverse (Qu *et al.*, 1983)

#### VII.4.3.1. Hybridation de l'amorce oligonucléotidique

L'oligonucléotide radiomarqué ainsi que l'ARN sont incubés ensemble dans le mélange réactionnel suivant :

| | |
|---|---|
| Oligonucléotide marqué en 5' à l'ATP [$\gamma$-$^{32}$P] (100 cps/µl) | 1 µl |
| ARN coupé par une sonde enzymatique | 1 µl |
| Tampon AMV 10X (Q-BIOgene) ($MgCl_2$ 60 mM, KCl 0,4 M, Tris-HCl 0,5 M pH 8,3) | 0,25 µl |
| $H_2O$ | qsp 2,5 µl |

Le mélange est incubé à 65°C durant 10 minutes puis laissé 10 minutes à température ambiante. Cette lente descente en température du mélange réactionnel permet à l'oligonucléotide de s'hybrider à sa séquence complémentaire sur l'ARN. Dans le cas d'un séquençage, on utilise 1,5 pmol d'un transcrit *in vitro* non radioactif n'ayant subi aucune modification.

#### VII.4.3.2. Réaction d'élongation

Suite à l'hybridation de l'oligonucléotide sur l'ARN, le mélange réactionnel suivant est ajouté afin que l'élongation puisse avoir lieu :

| | |
|---|---|
| Mélange des 4 dNTP (5 mM chacun) | 0,1µl |
| Tampon AMV 10X (Q-BIOgene) ($MgCl_2$ 60 mM, KCl 0,4 M, Tris-HCl 0,5 M pH 8,3) | 0,25 µl |
| Transcriptase inverse AMV(2 U/µl, Q-BIOgene) | 0,25 µl |
| $H_2O$ | qsp 2,5µl |

L'ensemble est incubé 45 minutes à 42°C puis plongé dans la glace et la réaction d'élongation est arrêtée par l'ajout de 4 µl de bleu de dépôt de formamide.

Dans le cas d'un séquençage, les mélanges réactionnels d'hybridation et d'élongation sont réalisés en quatre exemplaires. Dans ces tubes notés A, C, G et T, un volume de 0,5 µl de ddATP, ddCTP, ddGTP et ddTTP est respectivement ajouté au mélange d'élongation pour avoir une concentration finale du ddNTP de 0,5 mM afin que l'incorporation de chacun des didésoxynucléotides soit statistique.

### VII.4.4. Fractionnement sur gel

Les ADNc sont dénaturés 2 minutes à 96°C, puis placés dans la glace. Les produits de séquençage et d'élongation sont fractionnés par une électrophorèse en gel de polyacrylamide dénaturant à 7% dont la composition est la suivante : acrylamide 7% (p/v) et N,N'-méthylène bisacrylamide dans un rapport de 20/1, tampon TBE 1X (acide borique 89 mM ; EDTA 2 mM ; Tris borate 89 mM, pH 8,3), urée 8 M, TEMED 0,08% (v/v) et persulfate d'ammonium (APS) 0,01% (p/v). Le gel est pré-chauffé à une température de 48°C par une pré-électrophorèse à 45V/cm, puis les échantillons sont déposés sur le gel. L'électrophorèse est réalisée à 45V/cm en tampon TBE 1X, la migration des bleus de bromophénol et de xylène cyanol est utilisée comme référence. Le gel est ensuite transféré sur papier Whatmann 3MM, recouvert d'un film plastique puis séché par chauffage sous vide pendant 30 minutes à 80°C. Le gel est ensuite placé en autoradiographie à -80°C au contact d'un film Fuji X-ray, en présence d'écrans amplificateurs. Après le temps d'exposition désiré, l'autoradiogramme est obtenu suite au développement du film.

### VII.5. Prédiction *in silico* de la structure secondaire

Les prédictions de structure secondaire ont été réalisées par le programme Mfold version 3.2 qui est accessible à l'adresse suivante : http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi (Mathews *et al.*, 1999 ; Zuker, 2003). Pour une séquence d'ARN donnée, ce programme propose les structures secondaires les plus stables d'après le calcul de la stabilité thermodynamique des appariements des bases de l'ARN. Ce calcul est effectué pour un ARN dans une solution de NaCl 1M à 37°C. Ce programme permet également d'imposer des contraintes de structuration en simple et double brin ce qui permet d'obtenir des modèles de structure secondaire qui sont en accord avec les données expérimentales obtenues lors de l'hydrolyse ménagée de l'ARN par les sondes enzymatiques et chimiques.

## VIII. Techniques d'analyse des interactions ARN/protéines

### VIII.1. Technique d'empreinte des ARN en extrait nucléaire ou en présence de protéine purifiée

#### VIII.1.1. Principe

La fixation d'une protéine sur une séquence d'ARN protège cette séquence de l'action des sondes enzymatiques utilisées lors des analyses de structure secondaire des ARN. Lorsque l'ARN est incubé en présence d'un extrait nucléaire ou de protéine purifiée, des complexes

ribonucléoprotéiques se forment et sont ensuite soumis aux hydrolyses enzymatiques d'une RNase ajoutée au mélange réactionnel. La même expérience est réalisée en parallèle avec l'ARN seul afin de connaître quelles séquences sont accessibles à la RNase en l'absence de tout facteur protéique. Les sites de fixations des facteurs protéiques sont ensuite identifiés par la comparaison des résultats de transcriptions inverses effectuées sur l'ARN seul et l'ARN impliqué dans les complexes ribonucléoprotéiques. Ces sites sont identifiés lorsqu'un certain nombre de coupures, présentes sur l'ARN seul, disparaissent lorsque l'ARN est incubé avec les facteurs protéiques.

### VIII.1.2. Assemblage des complexes ARN/protéines

L'ARN dont on veut identifier les sites de fixation de facteurs protéiques est incubé dans le mélange suivant :

| | |
|---|---|
| ARN Transcrit *in vitro* | 3 pmol |
| ARNt (1 µg/µl) | 1,25 µl |
| Tampon D | 3,6 µl |
| (EDTA 0,2 mM ; KCl 0,1 M ; DTT 0,5 mM ; PMSF 0,25 mM ; HEPES-KOH 20 mM pH 7,9 ; glycérol 20 %) | |
| $H_2O$ | qsp 14 µl |

Le mélange est incubé 10 minutes à 65°C puis 10 minutes à température ambiante pour permettre à l'ARN de se renaturer puis cette conformation est figée par un court refroidissement dans la glace. Un volume de 8 µl d'extrait nucléaire ou de protéine purifiée dilués ou non dans le tampon D sont ensuite ajoutés et le mélange est incubé 5 minutes à 30°C.

### VIII.1.3. Hydrolyses enzymatiques

Une fois les complexes ribonucléoprotéiques formés, 1,5 U de RNase T1 sont ajoutés durant 5 minutes à 30°C. Les digestions enzymatiques sont arrêtées par l'ajout de 20 µg d'ARNt de levure et les ARN sont extraits au phénol-chloroforme (24/1) puis au chloroforme-alcool isoamylique (24/1). Les ARN sont ensuite précipités à l'éthanol absolu en présence de 0,3 M d'acétate de sodium, les culots sont lavés avec de l'éthanol 70 %, séchés et repris dans 4 µl d'$H_2O$. Les ARN sont ensuite analysés par transcription inverse dans les conditions décrites au **paragraphe VII.4.3**.

## VIII.2. Technique de retard sur gel

### VIII.2.1. Synthèse de transcrits radioactifs uniformément marqués

Des transcrits uniformément marqués ont été produits à partir des matrices PCR produites dans les conditions décrites dans le paragraphe **1.1**. Le mélange réactionnel de transcription est le suivant :

| | |
|---|---|
| Matrice PCR | 1,5 pmol |
| Tampon de transcription 5X : (Tris-HCl 200 mM, pH 7,5 ; $MgCl_2$ 30 mM ; spermidine 5mM) | 3 µl |
| RNasin (20 U/µl, Fermentas) | 1 µl |
| Mélange de rNTP : (ATP 20 mM, CTP 20 mM, GTP 20 mM, UTP 1 mM) | 1 µl |
| [$\alpha$-$^{32}$P] UTP (800 Ci/mmol, 20 mCi/ml, GE Healthcare) | 1 µl |
| ARN polymérase T7 (produite au laboratoire) | 1,5 µl |
| $MgCl_2$ (0,1 M) | 2 µl |
| $H_2O$ | qsp 15 µl |

Le mélange est incubé 2h à 37°C. La matrice ADN est ensuite dégradée pendant 30 minutes à 37°C en présence de 5 U de DNaseI, exempte de RNase (Fermentas). Un volume de 15 µl de bleu de formamide est ajouté et les ARN sont ensuite purifiés par électrophorèse en gel dénaturant de polyacrylamide 5%-urée 8 M. L'électrophorèse est réalisée à 17 V/cm dans du tampon TBE 0,5 X (acide borique 44,5 mM ; EDTA 1 mM ; Tris-borate 44,5 mM, pH 8,3). A la fin de l'électrophorèse, le gel est autoradiographié et la bande contenant les transcrits est découpée et incubée dans du tampon d'élution (NaCl 250 mM ; EDTA 10 mM ; Tris-HCl 20 mM pH 7,5) à 4°C durant 12h. Les ARN élués sont ensuite précipités en présence d'éthanol 96% et de 10 µg de glycogène. Le culot est lavé avec de l'éthanol 70%, et séché. La radioactivité contenue dans le culot est mesurée au compteur à scintillation en mode Cerenkov et le culot est repris dans de l' $H_2O$ à raison de 100 fmol/µl.

### VIII.2.2. Retard sur gel

Cette technique est aussi appelée EMSA pour « Electrophoretic Mobility Shift Assay ». Le principe de cette technique repose sur le fait qu'un ARN n'a pas la même mobilité électrophorétique suivant qu'il est libre ou lié à une protéine. En absence de protéine, l'ARN supposé porter des sites de fixation pour des protéines migre à une position donnée sur un gel non dénaturant qui ne pourra être atteinte lorsque le même ARN est complexé avec une ou

plusieurs protéines. Le retard de migration engendré dans ce dernier cas de figure sera dû à la diminution de la vitesse électrophorétique de l'ARN complexé.

Les expériences de retard sur gel ont été effectuées selon deux conditions : soit avec de l'ARN uniformément marqué (10 à 100 fmol), soit avec un mélange d'ARN non radioactif (0,5 à 1,5 pmol) et uniformément marqué (10 fmol). Dans ce dernier cas, l'ARN radioactif sert de traceur et la quantité utilisée est largement inférieure à celle de l'ARN non radioactif utilisé. Dans chaque réaction de retard sur gel, des ARNt de levure sont ajoutés afin de jouer le rôle des compétiteurs aspécifiques. Leur quantité est ajustée par rapport à la quantité de l'ARN dont on veut observer le retard électrophorétique de façon à ce qu'ils soient présents à un excès molaire de 10 à 100 fois.

Préalablement à l'ajout des protéines ou de l'extrait nucléaire, l'ARN est incubé dans le mélange réactionnel suivant :

| | |
|---|---|
| Tampon D<br>(EDTA 0,2 mM ; KCl 0,1 M ; DTT 0,5 mM ; PMSF 0,25 mM ; HEPES-KOH 20 mM pH 7,9 ; glycérol 20 %) | 1,8 µl |
| $MgCl_2$ (62,5 mM) | 0,5 µl |
| $H_2O$ | qsp 7µl |

Le mélange est incubé 10 minutes à 65°C puis 10 minutes à température ambiante pour permettre à l'ARN de se renaturer puis placé dans la glace. Un volume de 4 µl d'extrait nucléaire ou de protéine purifiée dilués ou non dans le tampon D est ensuite ajouté et le mélange est incubé 5 minutes à 30°C afin de permettre la formation des complexes ribonucléoprotéiques. Sept µl de bleu de dépôt non dénaturant (HEPES-KOH 200 mM pH 7,9 ; glycérol 40 %, bleu de bromophénol 0,03 % (p/v) ; xylène cyanol 0,03% (p/v)) sont ensuite ajoutés. Lorsque les complexes sont formés avec un extrait nucléaire, 10 µg d'héparine sont ajoutés en même temps que le bleu de dépôt afin de d'éviter la rétention des complexes dans les puits du gel. Les échantillons sont incubés 10 minutes à température ambiante et sont déposés sur un gel d'acrylamide/bisacrylamide (38 :2) de 5% contenant 5% de glycérol. Le pourcentage de ce gel peut varier de 5 à 8 % en acrylamide/bisacrylamide suivant la taille des ARN utilisés lors de l'expérience. La migration est réalisée à 4°C dans un tampon TBE 0,5 X (acide borique 44,5 mM ; EDTA 1 mM ; Tris-borate 44,5 mM, pH 8,3), glycérol 0,5 %, et la durée de l'électrophorèse est estimée d'après les migrations

électrophorétiques des bleus de xylène cyanol et de bromophénol. Le gel est ensuite séché et exposé en autoradiographie durant une nuit.

## IX. Méthodes spécifiques à l'étude de l'épissage *in vitro*

### IX.1. Epissage *in vitro* avec des transcrits uniformément marqués

#### IX.1.1. Préparation de la matrice de transcription

Les plasmides servant de matrice pour la transcription *in vitro* sont linéarisés par l'enzyme de restriction adéquate (**Figure V-9**). Les linéarisations sont vérifiées par le dépôt d'une fraction aliquote du mélange de digestion sur un gel d'agarose 0,8%. L'ADN est ensuite soumis à une extraction avec du phénol-chloroforme (1/1) puis à une extraction avec un mélange chloroforme-alcool isoamylique (24/1). L'ADN est ensuite précipité à l'éthanol en présence d'acétate de sodium 0,3 M et repris dans de l' $H_2O$ à une concentration de 250 ng/µl.

#### IX.1.2. Synthèse de transcrits radioactifs *in vitro*

Les transcrits sont uniformément marqués grâce à l'incorporation d'un ribonucléotide marqué au $^{32}P$ lors de la transcription par l'ARN polymérase T7. Le plasmide linéarisé servant de matrice est transcrit dans le milieu réactionnel suivant :

| | |
|---|---|
| Plasmide linéarisé (500 ng/µl) | 0,5 µl |
| Tp transcription US 5X<br>($MgCl_2$ 30 mM ; Tris-HCl 200 mM, pH 7,5; spermidine 5 mM) | 2 µl |
| DTT (0,2 M) | 0,5 µl |
| RNasin (20 U/µl, Fermentas) | 0,5 µl |
| Mélange de rNTP (ATP 5 mM, CTP 5 mM, GTP 2 mM) | 1 µl |
| $m^7$GpppG (10 mM, NEB) | 1,25 µl |
| UTP (0,5 mM) | 1 µl |
| [α-$^{32}$P]UTP (800 Ci/mmol, 20 mCi/ml, GE Healthcare) | 0,9 µl |
| ARN polymérase T7 (produite au laboratoire) | 0,5 µl |
| BSA (2 mg/ml) | 0,5 µl |
| $H_2O$ | qsp 10 µl |

Par rapport aux autres rNTP, le ribonucléotide rGTP est en plus faible concentration dans le mélange réactionnel afin de favoriser l'incorporation de la coiffe $m^7$GpppG à l'extrémité 5' du transcrit qui le protège de la dégradation. Dans le mélange transcriptionnel, la

concentration en ribonucléotide rUTP est 22 fois moins importante que son homologue radioactif [$\alpha$-$^{32}$P]UTP de façon à permettre l'incorporation importante de ce dernier dans les ARN au cours de la transcription.

La transcription est réalisée durant 1 heure à 37°C puis la matrice ADN est dégradée par l'ajout de 2 U de TURBO$^{TM}$ DNase exempte de RNases (Ambion), pendant 30 minutes à 37°C. Après ajout de 20 µg d'ARNt de levure servant d'entraîneur de précipitation et de 190 µl d' $H_2O$, les transcrits sont extraits avec un volume de phénol-chloroforme (1/1), puis avec un volume de chloroforme/alcool isoamylique (24/1). Les ARN sont ensuite soumis à deux précipitations successives dans de l'éthanol absolu, la première se faisant en présence d'acétate d'ammonium 0,25 M et la seconde, en présence de NaCl 0,2 M. Ces précipitations permettent d'éliminer les ribonucléotides non incorporés. Le culot est lavé deux fois dans de l'éthanol à 80%, puis séché et repris dans 10 µl d' $H_2O$. La quantité de transcrits radioactifs obtenue est déterminée par mesure de la radioactivité contenue dans 1 µl de transcrit au compteur à scintillation en mode Cerenkov. L'ARN est ensuite dilué dans de l' $H_2O$ à la concentration de 10 fmol/µl. La qualité de la transcription est vérifiée par dépôt d'une fraction (10 fmol) sur gel de polyacrylamide 5% dénaturant (polyacrylamide 5%, urée 8M, TBE 0,5X) et autoradiographie.

### IX.1.3. Réaction d'épissage

Pour réaliser les expériences d'épissage *in vitro* avec des transcrits uniformément marqués, nous avons utilisé des extraits nucléaires de cellules HeLa provenant de la société CilBiotech. L'extrait nucléaire peut être décongelé et recongelé 1 à 2 fois sans perte notable d'activité.

Chaque réaction d'épissage *in vitro* est réalisée avec 1 µl de transcrit radioactif, soit 10 fmol. Le transcrit est ajouté au mélange réactionnel suivant :

| | |
|---|---|
| ARN uniformément marqué | 10 fmol |
| $MgCl_2$ (62,5 mM) | 1 µl |
| ATP (25 mM) | 1 µl |
| Créatine phosphate (625 mM) | 1 µl |
| RNasin (20 U/µl, Fermentas) | 0,4 µl |
| Alcool polyvinylique 13% (p/v) | 6 µl |
| Tampon D | 3,6 µl |
| (EDTA 0,2 mM; KCl 0,1 M; DTT 0,5 mM; PMSF 0,25 mM; HEPES 20 mM pH 7,9 ; glycérol 20%) | |
| $H_2O$ | qsp 14 µl |

Huit µl d'extrait nucléaire sont préalablement incubés 10 minutes à 30°C afin d'éliminer l'ATP endogène puis ajoutés au mélange réactionnel. Le mélange est placé 5 minutes dans la

glace, puis incubé 2 heures à 30°C. La réaction d'épissage est arrêtée par l'ajout d'un volume de 180 µl contenant 20 µg de protéinase K (Roche Diagnostics) diluée dans son tampon (EDTA 12,5 mM, pH 7,4 ; NaCl 150 mM; SDS 1% ; Tris-HCl 100 mM, pH 7,5). Après une incubation de 30 minutes à 30°C, les ARN sont extraits au phénol-chloroforme, puis au chloroforme-alcool isoamylique (24/1) et ensuite précipités en présence d'éthanol absolu sans ajout de sels. Le précipité est lavé puis séché et repris dans 5 µl d'une solution de bleu de formamide (EDTA 20 mM, bleu de bromophénol 0,05%, bleu de xylène cyanol 0,05%, dans de la formamide désionisée).

### IX.1.4. Fractionnement sur gel et fluorographie

Les produits d'épissage sont dénaturés pendant 2 minutes à 96°C puis plongés dans la glace, avant d'être analysés par électrophorèse sur gel de polyacrylamide dénaturant dont la composition est la suivante : acrylamide 5% (p/v) et N, N'-méthylène bisacrylamide dans un rapport 20/1 (p/p), tampon TBE 1X (acide borique 89 mM, EDTA 2 mM, Tris borate 89 mM, pH 8,3), urée 8 M, TEMED 0,08% (v/v) et persulfate d'ammonium 0,1% (p/v)). Après une pré-électrophorèse à 45 V/cm, destinée à préchauffer le gel à 48°C, 1500 cpm Cerenkov sont déposés sur le gel. L'électrophorèse est poursuivie à 45 V/cm pendant 4 heures à une température d'environ 48°C, le bleu de xylène cyanol étant utilisé comme référence de migration. Le gel est transféré sur du papier Whatmann 3MM, recouvert d'un film plastique, puis séché par chauffage sous vide à 80°C. Le gel est alors placé en autoradiographie à -80°C au contact d'un film Fuji X-Ray, en présence d'écrans amplificateurs.

### IX.1.5. Détermination de l'efficacité d'épissage

Le gel est scanné par l'appareil Typhoon (GelLifeSciences) et la radioactivité contenue dans les bandes d'intérêt est quantifiée à l'aide du logiciel Image Quant, version 5.2. Les quantités de molécules de précurseurs et de produits d'épissage sont calculées sur la base du contenu en résidus uridine de chaque type d'ARN. L'efficacité d'épissage est exprimée par le rapport de la radioactivité contenue dans la bande d'ARN messager (M) et celle contenue dans la bande de précurseur restant à la fin de la réaction (P). Les expériences ont été répétées au moins deux fois afin de vérifier leur reproductibilité.

## IX.2. Epissage *in vitro* avec des transcrits non radioactifs

### IX.2.1. Préparation de la matrice de transcription

Les plasmides servant de matrice pour la transcription *in vitro* des ARN non radioactifs sont les mêmes que ceux utilisés pour la transcription *in vitro* des ARN uniformément marqués (**Figure V-9**). Les plasmides sont linéarisés avec l'enzyme de restriction adéquate.

### IX.2.2. Synthèse de transcrits non radioactifs *in vitro*

Le plasmide linéarisé servant de matrice est transcrit dans le milieu réactionnel suivant :

| | |
|---|---|
| Plasmide linéarisé (500 ng/µl) | 10 µl |
| Tampon de transcription 5X (200mM Tris-HCl, pH 7,5 ; $MgCl_2$ 30 mM ; spermidine 5 mM) | 8 µl |
| DTT (0,2 M) | 2 µl |
| RNasin (20 U/µl, Fermentas) | 2 µl |
| Mélange de rNTP (ATP 5 mM, CTP 5 mM, GTP 2 mM, UTP 5 mM) | 1 µl |
| $m^7$GpppG (10 mM, NEB) | 4 µl |
| ARN polymérase T7 (produite au laboratoire) | 3 µl |
| BSA (2 mg/ml) | 2 µl |
| $H_2O$ | qsp 40 µl |

Par rapport aux autres rNTP, le ribonucléotide rGTP est en plus faible concentration dans le mélange réactionnel afin de favoriser l'incorporation de la coiffe $m^7$GpppG à l'extrémité 5' du transcrit qui le protège de la dégradation. La transcription est réalisée durant 2 heures à 37°C puis la matrice est dégradée par l'ajout de 10 U de TURBO™ DNase (Ambion) durant 30 minutes à 37°C. Le mélange est ensuite incubé 10 minutes à 65°C pour inactiver thermiquement la DNase et un aliquot de la réaction de transcription est prélevé et déposé sur gel d'agarose 1% afin de vérifier la qualité des transcrits. Les transcrits sont ensuite purifiés avec le kit GenElute™ Mammalian Total RNA (Sigma-Aldrich) selon les recommandations du fabriquant.

### IX.2.3. Réaction d'épissage

Pour réaliser les expériences d'épissage *in vitro* avec des transcrits non radioactifs, nous avons utilisé des extraits nucléaires et des extraits cytoplasmiques de cellules HeLa provenant de la société CilBiotech (Mons, Belgique). L'extrait nucléaire peut être décongelé et recongelé 1 à 2 fois sans perte notable d'activité. Les protéines SR recombinantes ont été

produites grâce au système baculovirus et nous ont été fournies par J. Stévenin (IGBMC, Strasbourg) (Cavaloc *et al.*, 1999).

Chaque réaction d'épissage *in vitro* est réalisée avec 20 fmol de transcrit non radioactif. Le protocole est identique à celui utilisé pour les réactions d'épissage réalisées avec un ARN uniformément marqué (**paragraphe VII.2.1**).

La réaction d'épissage est arrêtée par l'ajout d'un volume de 180 µl contenant 20 µg de protéinase K (Roche Diagnostics) diluée dans son tampon (EDTA 12,5 mM, pH 7,4 ; NaCl 150 mM; SDS 1% ; Tris-HCl 100 mM, pH 7,5). Après une incubation de 30 minutes à 30°C, les ARN sont extraits au phénol-chloroforme, puis au chloroforme-alcool isoamylique (24/1) et ensuite précipités en présence d'éthanol absolu sans ajout de sels. Le précipité est lavé puis séché et repris dans 20 µl d'$H_2O$.

Lors de l'étude de l'effet des protéines 9G8 et ASF/SF2 sur l'inhibition de l'épissage par le NRS *in vitro*, un volume de 8 µl composé de 4 µl d'extrait nucléaire et de 4 µl d'extrait cytoplasmique auquel est ajoutée une quantité donnée de protéines ASF/SF2 ou 9G8 recombinantes sont ajoutés au mélange réactionnel. Les échantillons sont ensuite traités de façon identique qu'au cours des expériences réalisées en l'absence de protéines SR.

### IX.2.4. Analyse par RT-PCR des produits d'épissage

#### IX.2.4.1. Principe

A la fin de la réaction d'épissage réalisée avec de l'ARN non radioactif, les quantités des ARN présents sont trop faibles pour que les produits d'épissage soient visualisés directement par une électrophorèse. Il est donc nécessaire d'utiliser une réaction de RT-PCR afin d'amplifier les espèces épissées et non épissées.

Après la réaction d'épissage, les produits d'épissage épissés (ARNm) et non épissés (pré-ARNm) sont tout d'abord converti en ADNc par une réaction de transcription inverse. Pour ce faire, on utilise un oligonucléotide qui s'hybride à la séquence se situant à l'extrémité 3' de tous les ARN qui est commune aussi bien aux produits épissés que non épissés. Une fois les ARN rétro-transcrits en ADNc, une amplification par PCR est réalisée à l'aide de deux couples d'oligonucléotides qui amplifient spécifiquement les espèces épissées et les espèces épissées et non épissées. Les produits d'amplification sont ensuite fractionnés sur un gel d'agarose 1,2 % contenant du bromure d'éthidium (0,5 µg/ml), le gel est photographié à l'aide d'une caméra digitale ccd et les produits d'épissage sont quantifiés avec un logiciel de traitement d'image.

### IX.2.4.2. Réaction de transcription inverse sur les produits d'épissage

Les produits d'épissage sont rétro transcrits par la transcriptase inverse MuMLV en utilisant pour chaque construction un oligonucléotide qui s'hybride à l'extrémité 3' du second exon qui est une séquence commune à la fois aux ARNm et aux pré-ARNm. Les transcriptions inverses des produits d'épissage de la construction C1 et C2 ont été réalisées respectivement avec l'oligonucléotide 6039 (nt 4511-4528) et l'oligonucléotide 5825 (nt 5031-5053). La réaction de transcription inverse est schématisée dans la **Figure V-15** et les séquences des oligonucléotides sont détaillées dans le **Tableau V-1**. Deux µl de la préparation de produits d'épissage sont incubés dans le mélange réactionnel suivant :

| | |
|---|---|
| Produits d'épissage | 2 µl |
| Tampon MuMLV 5X (Promega) (Tris-HCl 250mM, pH 8.3, KCl 375mM, $MgCl_2$15mM, DTT 50mM) | 4 µl |
| Mélange des 4 dNTP (10 mM) | 2 µl |
| RNasin (20 U/µl, Fermentas) | 1µl |
| Oligonucléotide 5825 ou 6039 (5 µM) | 1 µl |
| Transcriptase inverse MuMLV (200 U/µl, Promega) | 1µl |
| $H_2O$ | qsp 20 µl |

La réaction de transcription inverse a lieu durant 1 heure à 37°C puis le mélange est incubé 10 minutes à 65°C pour inactiver thermiquement la transcriptase inverse.

### IX.2.4.3. Amplification par PCR des produits de transcription inverse

Les produits de transcription inverse sont ensuite amplifiés par une réaction de PCR et deux réactions de PCR sont réalisées par échantillon. Une première réaction de PCR est réalisée afin d'amplifier spécifiquement les produits épissés. Les produits d'épissage de la construction C1 sont amplifiés lors de la PCR avec le couple 6519/6245 qui s'hybrident respectivement aux séquences du premier exon (nt 1-20) et au niveau de la jonction exon1/exon 2 (nt 281-289/4492-4503) des produits épissés. Les produits d'épissage de la construction C2 sont amplifiés lors de la PCR avec le couple d'oligonucléotides 6519/6246 qui s'hybrident respectivement aux séquences du premier exon (nt 1-20) et au niveau de la jonction exon 1/exon 2 (nt 283-289/4472-4482) des produits épissés. La seconde réaction de PCR est réalisée afin d'amplifier les produits épissés et non épissés, ce qui est équivalent à amplifier la totalité des produits présents à la fin de la réaction d'épissage. Pour cela, on utilise un seul couple d'oligonucléotides car la séquence du premier exon contenant le site D1 est commune aux constructions C1 et C2. Les oligonucléotides 6519/6187 s'hybrident sur la séquence de l'exon 1 (nt 1-20 et 191-208, respectivement) qui est présente à la fois chez les

espèces épissées et non épissées. Les sites d'hybridation de ces différents couples d'oligonucléotides sont schématisés dans la **Figure V-15** et les séquences des oligonucléotides sont détaillées dans le **Tableau V-1**. Le mélange réactionnel de la réaction de PCR est le suivant :

| | |
|---|---|
| Produit de la transcriptase inverse | 1 µl |
| Mélange des 4 dNTP (10 mM) | 1 µl |
| Tampon Go Taq® 5X (Promega) | 10 µl |
| Oligonucléotide 6519 (50 µM) | 1µl |
| Oligonucléotide 6245 ou 6246 (50 µM) | 1µl |
| OU | |
| Oligonucléotide 6519 (50 µM) | 1µl |
| Oligonucléotide 6187 (50 µM) | 1µl |
| Go Taq polymérase® (5 U/µl, Promega) | 0,25 µl |
| $H_2O$ | qsp 50 µl |

Un mélange réactionnel identique mais ne contenant pas d'ADNc est également réalisé. Cela constitue un témoin négatif qui permet de vérifier que les différents composants utilisés lors de la réaction de PCR ne sont pas contaminés.

Après un cycle de dénaturation de 2 minutes à 94°C, les produits épissés plus non épissés et les produits épissés sont respectivement amplifiés lors de 30 et 35 cycles d'amplification chacun composé de paliers de 30 secondes à 94°C, 45 secondes à 46°C et 45 secondes à 72°C pour ensuite terminer par une incubation de 10 minutes à 72°C.

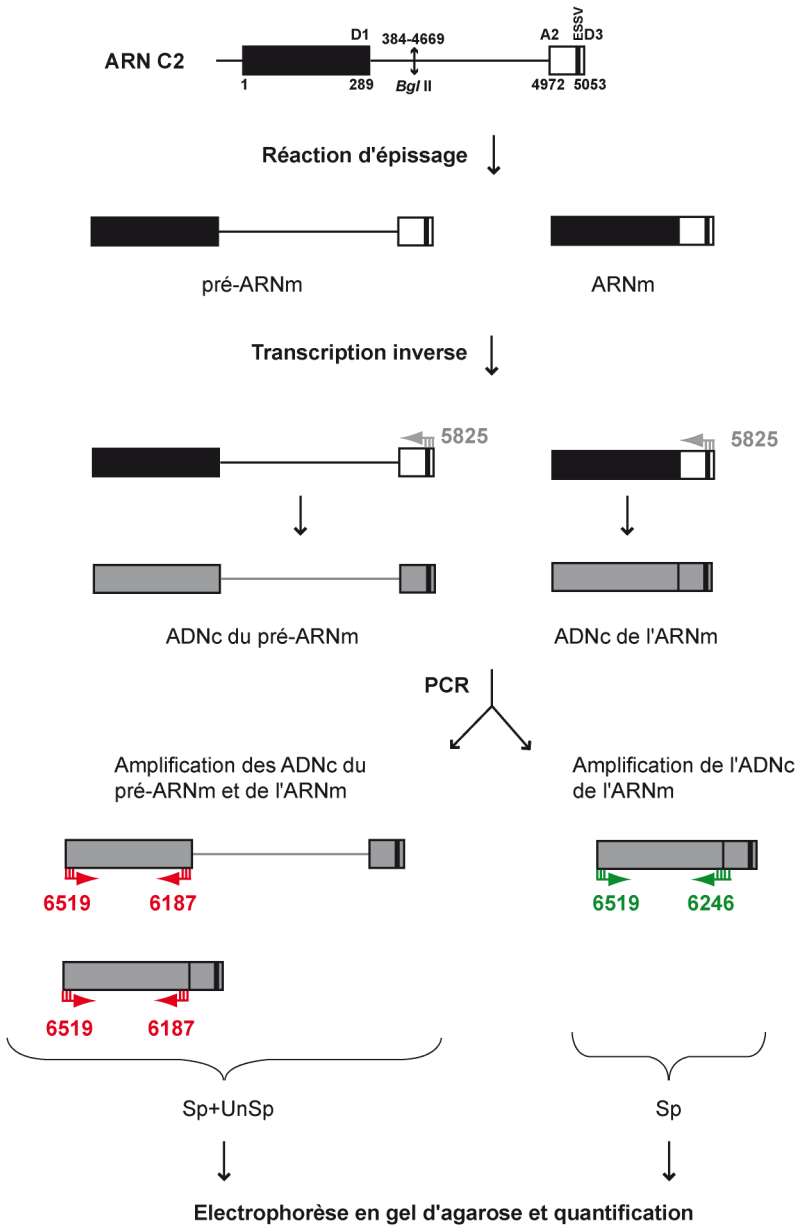
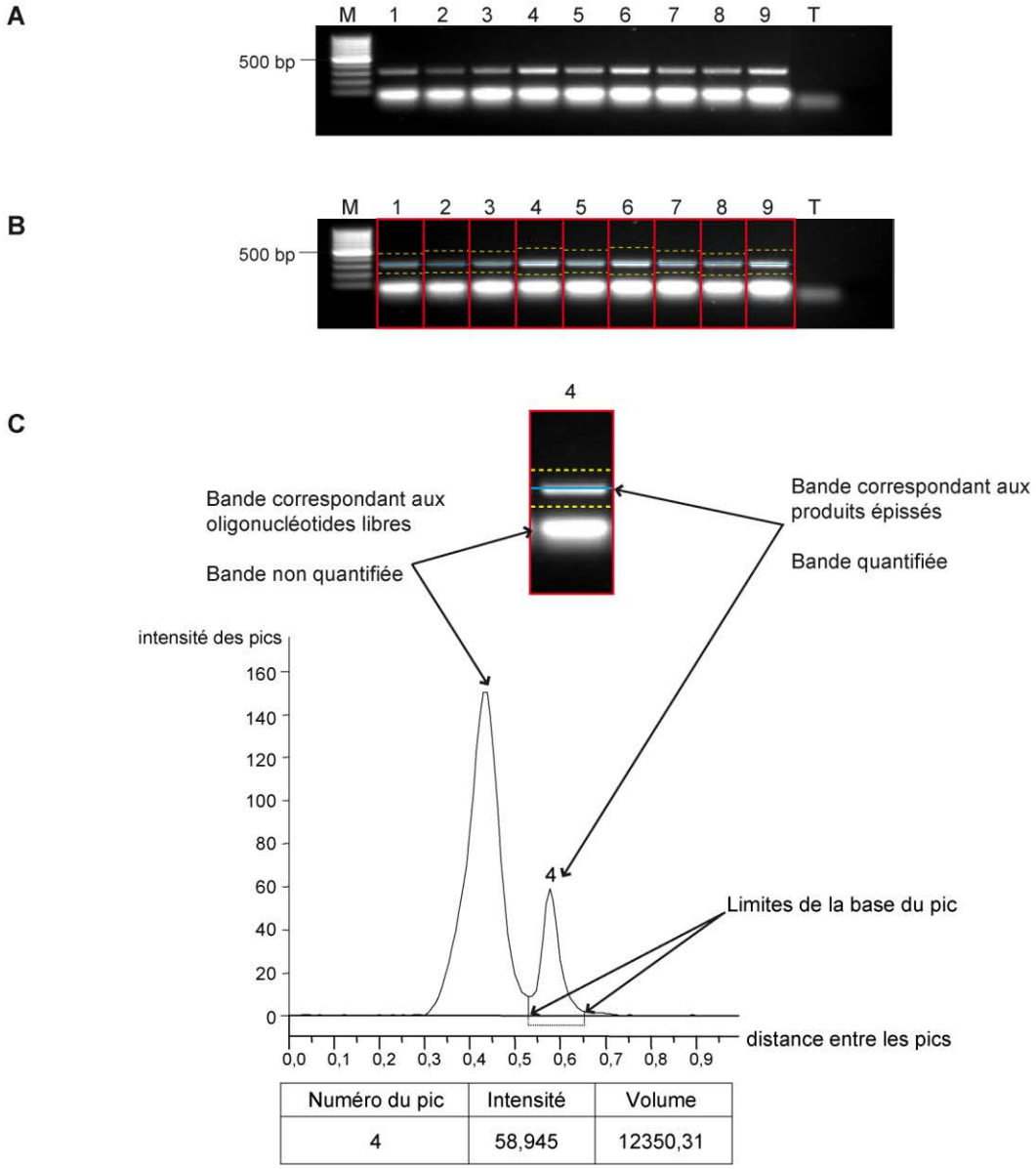



**Figure V-15: Représentation de la réaction d'épissage *in vitro* suivie d'une analyse par RT-PCR.**
L'analyse par RT-PCR des produits d'épissage est illsutrée avec l'ARN C2 mais la démarche est identique pour les autres constructions utilisées en épissage. Le premier et le deuxième exon sont respectivement représentés par des rectangles noir et blanc et l'élément ESSV est représenté par un trait noir vertical. La séquence ARN de l'intron est représentée par une ligne noire et les noms des sites donneur et accepteur sont précisés. Les oligonucléotides sont représentés par des flèches et leur numéro est indiqué. L'oligonucléotide utilisé pour la réaction de transcription inverse et les séquences des ARN rétro-transcrits en ADNc sont colorés en gris. Le couple d'oligonucléotides colorés en rouge et en vert correspondent aux couples utilisés lors de la PCR pour amplifier respectivement les ADNc correspondant aux pré-ARNm plus l'ARNm et l'ARNm seul. Les séquences des oligonucléotides sont détaillées dans le Tableau X.




| Numéro du pic | Intensité | Volume |
|---|---|---|
| 4 | 58,945 | 12350,31 |

**Figure V-16 : Exemple de quantification des produits de RT-PCR après électrophorèse en gel d'agarose en présence de bromure d'éthidium.**

Les images ont été aquises grâce à l'appareil GeneGenius (SynGene, caméra CCD) piloté par le logiciel GeneSnap et la quantification a été réalisée à l'aide du logiciel SynGeneTools version 3.08.

**A :** Exemple de fractionnement d'échantillons de RT-PCR par électrophorèse en gel d'agarose. Les pistes 1 à 9 correspondent à l'amplification des ADNc des ARNm obtenus après l'épissage de différentes pré-ARNm. La piste M correspond au marqueur de taille, la piste témoin (T) correspond à la réaction de PCR réalisée sans produit de transcription inverse.

**B :** Traitement de l'image du gel pour la quantification par le logiciel SynGeneTools. Chaque piste est délimitée par des rectangles rouges et les bandes à quantifier sont délimitées par des lignes jaunes et bleues qui délimitent respectivement la largeur de l'image quantifiée et sa région de plus forte intensité.

**C :** Exemple de valeur obtenue après la quantification de la bande dans la piste numéro 4. Les valeurs des intensités des bandes sont mesurées et reportées sur un chromatogramme. La valeur utilisée est celle du volume qui correspond à l'intégration de l'aire du pic contenue dans les limites du pic du chromatogramme.

La même démarche est empruntée pour quantifier les intensités des produits de RT-PCR correspondant à l'amplification des produits épissés et non épissés.

#### IX.2.4.4. Fractionnement en gel d'agarose des produits de RT-PCR et quantification

Un volume de 10 µl de la réaction de RT-PCR est ensuite déposé sur un gel d'agarose 1,2 % contenant 0,5 µg/ml de bromure d'éthidium. L'électrophorèse est réalisée à 100 V/ cm dans du tampon TAE 1X (40 mM Tris-acétate ; 1 mM EDTA, pH 7.6) durant 1 heure. Les produits PCR sont ensuite photographiés par une caméra digitale ccd (GeneGenius, SynGene) et quantifiés avec le logiciel de traitement d'image SynGeneTools version 3.08. Les réactions de RT-PCR ont été reproduites en triplicat et l'écart type des valeurs obtenues a été calculé. La procédure de quantification des produits de RT-PCR est illustrée dans la **Figure V-16**. Pour chaque réaction d'épissage l'intensité correspondant aux ARN épissés (Sp) et celle correspondant aux ARN épissés+non épissés (Sp+UnSp) sont mesurées. L'efficacité d'épissage est calculée par la formule Sp/(Sp+UnSp).

## X. Méthodes spécifiques à la production, la purification et l'analyse des protéines

### X.1. Analyse par électrophorèse en gel SDS-PAGE (Laemmli, 1970)

Les protéines sont diluées dans du bleu de dépôt SBL 1X (Tris 250 mM, pH 6,8 ; SDS 8 % ; β-mercaptoéthanol 20 % ; bleu de bromophénol 0,04 %, glycérol 10 %), chauffées 5 min à 96°C puis déposées sur un gel SDS-PAGE (gel de séparation à 5 %, gel de concentration à 10 %). La migration est effectuée dans du tampon de migration (Tris 25 mM ; glycine 192 mM ; SDS 0,1 %) à ampérage constant (15 mA pour le gel de concentration, 30 mA pour le gel de séparation). A la fin de la migration, le gel est coloré dans une solution de bleu de coomassie durant 30 min et décoloré dans un bain d'éthanol (20%)/acide acétique (10 %). Le gel peut également être coloré au bleu de coomassie colloïdal. Pour cela, le gel est tout d'abord lavé 3 fois durant 10 minutes sous agitation dans une solution de fixation (éthanol 20%, acide acétique 10%) pour fixer les protéines et éliminer le SDS du gel. Puis le gel est coloré pendant une nuit sous agitation par la solution de bleu colloïdal ($(NH_4)_2SO_4$ 10 % ; $H_3PO_4$ 1,6 % ; bleu de coomassie G250 0,1% ; éthanol 20%). Le gel est ensuite décoloré par lavages avec de l'$H_2O$ sous agitation durant quelques heures.

### X.2. Analyse par Western blot

Les protéines sont tout d'abord fractionnées par électrophorèse en gel SDS-PAGE selon le protocole décrit précédemment. A la fin de l'électrophorèse, les protéines sont transférées sur une membrane en nitrocellulose (0,45 µm) sous l'action d'un champ électrique durant 1h à

100V dans du tampon de transfert (Tris 25 mM ; glycine 192 mM, Ethanol 20 % (v/v)). La membrane de nitrocellulose est ensuite traitée selon le protocole recommandé par le kit ECL™ (GE Healthcare). Après avoir été saturée avec une solution de PBS 1X Tween 0,1%, lait écrémé 5% ($Na_2HPO_4$ $2H_2O$ 0,81 mM ; $KH_2PO_4$ 0,147mM ; NaCl 13,7 mM ; KCl 0,27 mM, pH 7,4, Tween 0,1 %, lait écrémé 5% (p/v)), la membrane est mise en présence de l'anticorps primaire dirigé contre la protéine recherchée durant 1h sous agitation. La membrane est ensuite lavée 3 fois durant 10 min avec une solution de PBS 1X-Tween 0,1% ($Na_2HPO_4$ $2H_2O$ 0,81 mM ; $KH_2PO_4$ 0,147mM ; NaCl 13,7 mM ; KCl 0,27 mM, pH 7,4, Tween 0,1 %). L'anticorps secondaire conjugué à la peroxydase, dilué dans une solution de PBS 1X-Tween 0,1%, est ensuite ajouté sur la membrane pendant 45 min sous agitation. La membrane est ensuite lavée 3 fois avec une solution de PBS 1X-Tween 0,1% puis égouttée. La membrane est recouverte d'un solution de luminol/peroxyde (1/1, v/v) durant 5 min, égouttée et les protéines réactives à l'anticorps sont révélées par une autoradiographie.

## X.3. Production et purification de protéines recombinantes dans *E. coli*

### X.3.1. La protéine MS2-MBP

#### X.3.1.1. Production

La séquence correspondant à l'extrémité C-terminale de la protéine MBP (Maltose Binding Protein) a été fusionnée traductionnellement à l'extrémité N-terminale de la séquence correspondant de la protéine de l'enveloppe MS2 du phage MS2 dans le vecteur pMAL (vecteur pMAL-MS2) (Jurica et Moore, 2002). La transcription de l'ARN codant cette protéine de fusion est placée sous le contrôle du promoteur P*tac* qui est inductible par l'ajout d'IPTG. Les bactéries compétentes *E. coli* de la souche DH5α sont transformées par le plasmide pMAL-MS2, puis cultivées durant 1 heure à 37°C sous agitation avec 800 µl de milieu LB supplémenté avec 100 µg/ml d'ampicilline. Un volume de 400 µl de ce cette culture est ensuite prélevé pour inoculer 25 ml de milieu LB/A qui est ensuite incubée toute une nuit à 37°C sous agitation. Un volume de 20 ml de cette pré-culture est ensuite prélevé pour ensemencer 1 L de milieu LB/A qui est ensuite incubée à 37°C sous agitation jusqu'à ce que l'absorbance de la culture atteigne 0,8 U $DO_{600}$/ml. La production de la protéine MS2-MBP est alors induite durant 2h30 par l'ajout d'IPTG à une concentration finale de 0,5 mM. Les cellules sont ensuite récoltées par centrifugation à 6 000 g à 4°C pendant 15 minutes et le culot cellulaire est lavé avec 10 ml de tampon de sonication (Tris-HCl 20 mM, pH 7,6 ; NaCl 200 mM ; EDTA 1 mM) puis centrifugé de nouveau à 6 000 g à 4°C pendant 15 minutes. Le culot cellulaire est ensuite pesé et resuspendu dans six volumes de tampon de sonication.

#### X.3.1.2. Extraction des protéines cellulaires

Les cellules sont plongées dans de la glace et sont traitées aux ultrasons sous une puissance de 40 W avec des impulsions séparées d'une seconde durant 5x2 min, chaque étape de sonication étant séparée d'une période d'incubation de 2 minutes dans la glace. Après sonication, le lysat bactérien est centrifugé pendant 15 min à 4°C à 30 000 g et seul le surnageant qui contient la protéine soluble est utilisé pour la suite de la purification.

#### X.3.1.3. Chromatographie d'affinité sur des billes d'amylose

Le surnageant de sonication est ajouté sur 1,8 ml de billes d'amylose (NEB) préalablement lavées 3 fois avec du tampon de sonication. L'ensemble est incubé durant 1 heure à 4°C sur un agitateur rotatif. Les billes sont ensuite centrifugées 1 minute à 2 000 g à 4°C puis lavées 3 fois avec 10 ml de tampon de sonication. Les protéines MS2-MBP sont ensuite éluées par deux ajouts successifs de 1,5 ml de tampon d'élution contenant du maltose (Tris-HCl 20 mM, pH 7,6 ; maltose 15 mM).

#### X.3.1.4. Chromatographie d'affinité sur colonne Héparine

Cette chromatographie d'affinité est réalisée pour d'éliminer les acides nucléiques liés à la protéine MBP-MS2. L'éluat de la première chromatographie d'affinité est déposé par volume de 1 ml sur la colonne Héparine-Sepharose (HiTrap Heparin HP, GE Healthcare) préalablement équilibrée avec au moins 12 volumes de colonne de tampon Tris-HCl 20 mM, pH 7,6. Une fois la protéine MBP-MS2 retenue sur le gel, un gradient de concentration en NaCl allant de 0 à 1M NaCl est appliqué. La protéine MBP-MS2 est éluée à une concentration d'environ 200 mM NaCl et les fractions contenant la protéine sont rassemblées et dialysées dans 1 L de tampon D durant une nuit sous agitation à 4°C. La concentration de la protéine est mesurée en utilisant le réactif de Bradford et son degré de pureté est estimé en déposant 0,5 µg de la protéine sur un gel SDS-PAGE.

## XI. Techniques de purification des complexes ribonucléoprotéiques par chromatographie d'affinité

### XI.1. Chromatographie d'affinité reposant sur l'emploi de la tobramycine

#### XI.1.1. Préparation de la matrice : fixation de la tobramycine sur la matrice Sepharose

La matrice NHS-Sepharose (NHS-activated Sepharose$^{TM}$ 4 Fast Flow, GE Healthcare) (NHS pour N-hydroxysuccinimide) est une matrice couramment utilisée pour fixer covalemment des ligands possédant des groupements amine, comme c'est le cas de la

tobramycine qui possède 5 groupements amines primaires. Les billes de Sepharose portent un bras espaceur de 8 atomes à l'extrémité duquel se situe un groupement acide aminocaproïque dont la fonction carboxylique est protégée par un groupement succinimide. La préparation de la matrice consiste à remplacer le groupement NHS par la tobramycine pour donner la matrice Tobra-Sepharose (**Figure V-17**) (Hartmuth *et al.*, 2004 ; Wang et Rando, 1995).

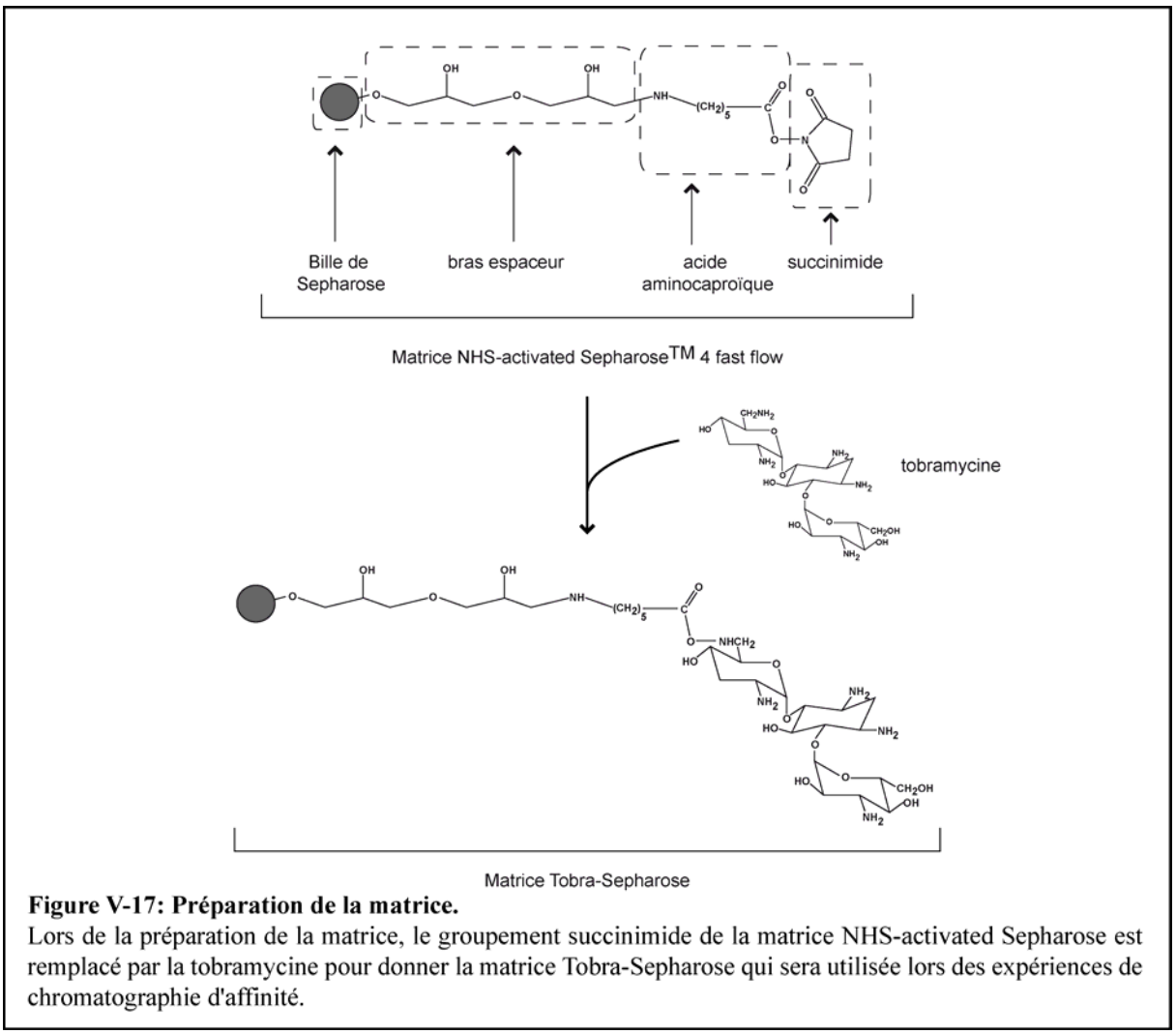



**Figure V-17: Préparation de la matrice.**
Lors de la préparation de la matrice, le groupement succinimide de la matrice NHS-activated Sepharose est remplacé par la tobramycine pour donner la matrice Tobra-Sepharose qui sera utilisée lors des expériences de chromatographie d'affinité.

Etape de déprotection :
HCl 1 mM

Etape de couplage :
Tampon de couplage : $NaCO_3$ 0,2 M ; NaCl 0,5 M ; tobramycine 5 mM ; pH ajusté à 8,3

Etape de neutralisation :
Tampon de neutralisation : $NaCO_3$ 0,2 M ; NaCl 0,1 M ; éthalonamine 1M ; pH ajusté à 8,0

Etape de lavage :
Tampon de lavage : PBS 1X : NaCl 130 mM ; $KH_2PO4$ 20 mM ; pH ajusté à 8,0
PBS-NaN3 : PBS 1X ; $NaN_3$ 0,02 % (p/v) ; pH ajusté à 8,0
Conservation :
PBS-NaN3 : PBS 1X ; $NaN_3$ 0,02 % (p/v) ; pH ajusté à 8,0

**Figure V-18: Tampons utilisés pour la préparation de la matrice.**

Cette réaction se compose de trois étapes (**Figure V-19**).

La première étape est une étape de déprotection lors de laquelle le groupement protecteur (hydroxy-succinimide) est hydrolysé. Pour cela, 2 ml de billes de Sepharose sont successivement lavés 4 fois avec 9 ml d'HCl 1 mM, chaque lavage étant suivi d'une centrifugation à 4°C de 1 minute à 4 000 g.

La seconde étape est une étape de couplage lors de laquelle la fonction carboxylique de l'acide aminocaproïque va réagir avec la fonction amine de la tobramycine. Les billes de Sepharose sont incubées avec 1 ml de tampon de couplage ($NaCO_3$ 0,2 M ; NaCl 0,5 M, pH 8,3) contenant 5 mM de tobramycine durant une nuit sous agitation rotative à 4°C. La tobramycine n'ayant pas été couplé aux billes de Sepharose est éliminée par une centrifugation de 5 minutes à 4 000 g à 4°C. Cette réaction de couplage n'est pas efficace à 100 % et de nombreux groupements carboxyliques de l'acide aminocaproïque ne réagissent pas avec les fonctions amines de la tobramycine.

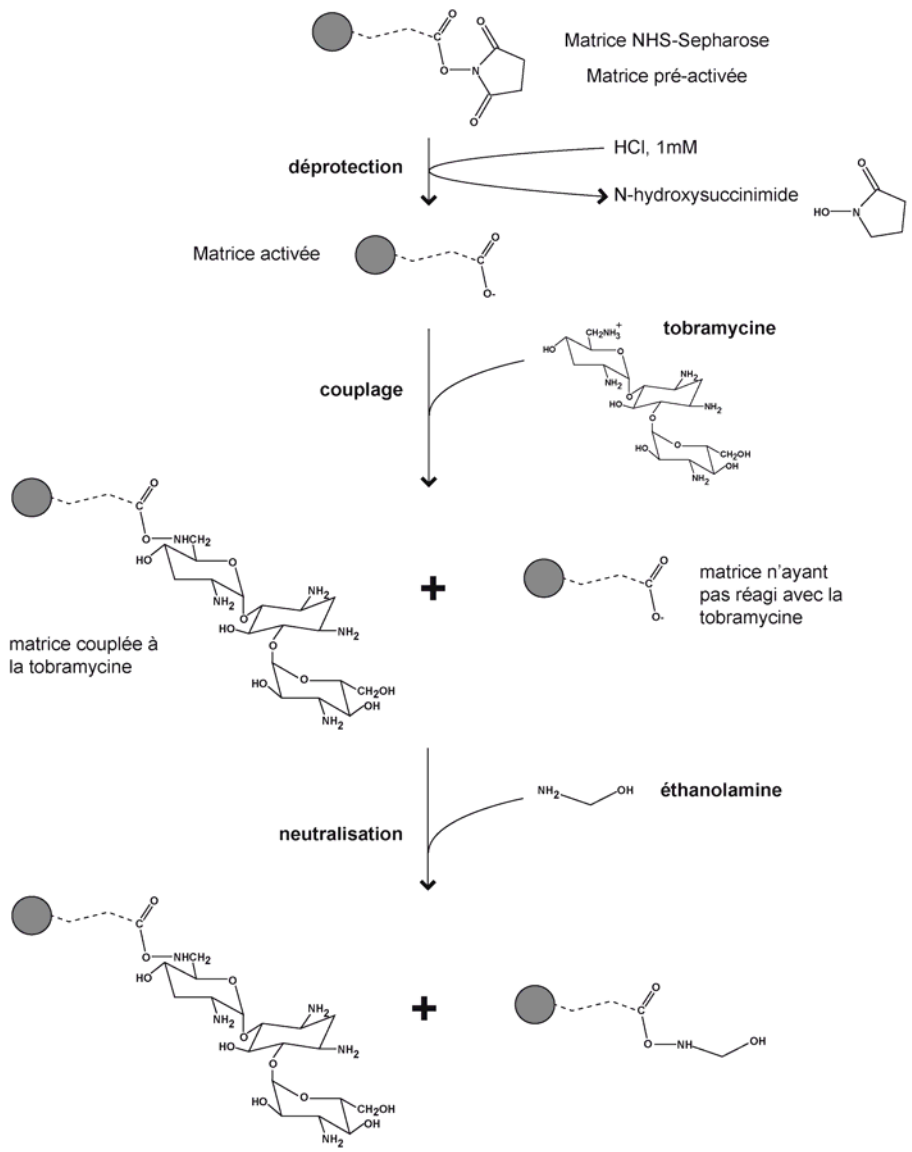



**Figure V-19: Fixation de la tobramycine sur la matrice**
Description des différentes étapes de fixation de la tobramycine sur la matrice. Une fois déprotégée par une hydrolyse acide, la fonction carboxylique de l'acide aminocaproïque réagit avec une des fonctions amines de la tobramycine lors d'une réaction de couplage. Lorsque la tobramycine est covalemment liée à la matrice, les fonctions carboxyliques qui n'ont pas réagi avec la tobramycine sont neutralisées par ajout d'éthanolamine. Afin de simplifier le schéma, seuls la fonction carboxylique de l'acide aminocaproïque et le groupement succinimide des billes de Sepharose sont représentés.

Ces fonctions carboxyliques sont très réactives et sont neutralisées lors de la troisième étape qui consiste à remettre un groupement protecteur sur ces fonctions. Cette réaction consiste à réaliser une seconde étape de couplage avec de l'éthanolamine qui porte une fonction amine primaire. Pour cela, 8 ml de tampon de neutralisation ($NaCO_3$ 0,2 M ; NaCl 0,1 M, pH 8,0) contenant 1 M d'éthanolamine sont ajoutés aux billes de Sepharose durant 2 h à température ambiante. Les billes sont ensuite lavées 3 fois à 4°C avec 9 ml de tampon PBS 1X (NaCl 130 mM ; $K_2HPO_4$ 20 mM, pH 8,0), chaque lavage étant suivi d'une centrifugation d'une minute à 4°C à 4000 g. Les billes sont ensuite lavées 2 fois à 4°C avec 9 ml de tampon PBS 1X-azide (PBS 1X, 0,02% azide de sodium (p/v), pH 8,0), chaque lavage étant suivi d'une centrifugation à 4°C durant 1 minute à 4 000 g. Les billes de Sepharose sont ensuite resuspendues dans 2 ml de tampon PBS 1X-azide et conservées à 4°C.

### XI.1.2. Test de la matrice Tobra-Sepharose

Cette expérience est réalisée pour vérifier que la tobramycine a bien été fixée sur la matrice Sepharose.

#### XI.1.2.1. Transcription *in vitro* des ARN utilisés dans la chromatographie d'affinité avec la technique tobramycine

Les plasmides servant de matrice ADN pour la transcription sont préalablement linéarisés par les enzymes de restriction adéquates. Les ARN non radioactifs et uniformément marqués sont transcrits *in vitro* à partir du promoteur SP6 avec le kit SP6 MEGAscript (Ambion) selon les protocoles suivants :

Protocole de transcription des ARN non radioactifs :

| | |
|---|---|
| Plasmide linéarisé (1 µg/µl) | 1 µl |
| Tampon SP6 MEGAscript 10 X (Ambion) | 2 µl |
| Mélange des 4 rNTP (10 mM chacun) | 4µl |
| SP6 (20 U/µl, Ambion) | 2 µl |
| $H_2O$ | qsp 20 µl |

Protocole de transcription des ARN uniformément marqués :

| | |
|---|---|
| Plasmide linéarisé (1 µg/µl) | 1 µl |
| Tampon SP6 MEGAscript 10 X (Ambion) | 2 µl |
| Mélange des 4 rNTP (10 mM chacun) | 4µl |

| | |
|---|---|
| [α-$^{32}$P] UTP (3 000 Ci/mmol, 10 mCi/ml, GE Healthcare) | 1 µl |
| SP6 (20 U/µl, Ambion) | 2 µl |
| $H_2O$ | qsp 20 µl |

Les mélanges réactionnels sont incubés 1h à 37°C puis la matrice ADN est dégradée par l'ajout de 5 U de TURBO$^{TM}$ DNase exempte de RNase (Ambion) pendant 30 minutes à 37°C. Un volume de 20 µl de bleu de formamide est ajouté et les transcrits non radioactifs et radioactifs sont ensuite déposés séparément et purifiés par électrophorèse en gel dénaturant de polyacrylamide 5%-urée 8 M. L'électrophorèse est réalisée à 17 V/cm dans du tampon TBE 0,5 X (acide borique 44,5 mM ; EDTA 1 mM ; Tris-borate 44,5 mM, pH 8,3). Après la migration, les transcrits non radioactifs et radioactifs sont respectivement visualisés sous UV (λ = 254 nm) et par autoradiographie, les bandes du gel contenant les transcrits sont découpées et les transcrits sont élués dans le tampon d'élution (NaCl 250 mM ; EDTA 10 mM ; Tris-HCl 20 mM pH 7,5) à 4°C pendant 12h. Les transcrits élués sont ensuite précipités avec de l'éthanol 96 %, et lavés avec de l'éthanol 70%. Le culot est ensuite séché et repris dans de l'$H_2O$. Les concentrations des transcrits non radioactifs sont mesurées par l'absorbance à 260 nm et les concentrations des transcrits radioactifs sont mesurées au compteur à scintillation en Cerenkov.

### XI.1.2.2. Mesure de la capacité de la matrice Tobra-Sepharose à fixer l'ARN NRS-Tob

Le test de la matrice Tobra-Sepharose est réalisé avec l'ARN NRS-Tob et se compose de 5 étapes: I) lavage de la matrice II) saturation de la matrice III) Fixation de l'ARN sur la matrice IV) lavage de la matrice, V) élution des ARN par ajout de tobramycine (**Figure V-20 et Figure V-21**). Une seconde expérience est réalisée avec l'ARN NRS qui sert de témoin négatif.

**Etape I et II :**

Vingt-cinq µl de la matrice Tobra-Sepharose sont lavés 3 fois avec 500 µl de tampon de saturation (Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 301 mM ; $CaCl_2$ 1 mM ; KCl 300 mM ; ARNt 0,1 mg/ml ; BSA 0,5 mg/ml ; NP40 0,01 %), chaque lavage étant suivi d'une centrifugation à 4 000g durant 1 minute et à 4°C. Les sites potentiels de fixation aspécifiques des acides nucléiques ou des protéines de la matrice sont saturés par une incubation dans 500 µl de ce même tampon durant toute une nuit sous agitation rotative.

Etape I et II :
Tampon de saturation : Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 1 mM ; $CaCl_2$ 1 mM ; KCl 300 mM ; ARNt 0,1 mg/ml ; BSA 0,5 mg/ml ; NP40 0,01 %

Etape III :
Tampon de fixation : Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 1 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; ARNt 0,1 mg/ml ; DTT 0,8 mM

Etape IV :
Tampon de lavage : Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 1 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; NP40 0,1 %

Etape V :
Tampon d'élution : Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 3 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; DTT 0,8 mM ; tobramycine 5 mM ; glycérol 5 %

**Figure V-20: Tampons utilisés pour le test de la matrice.**

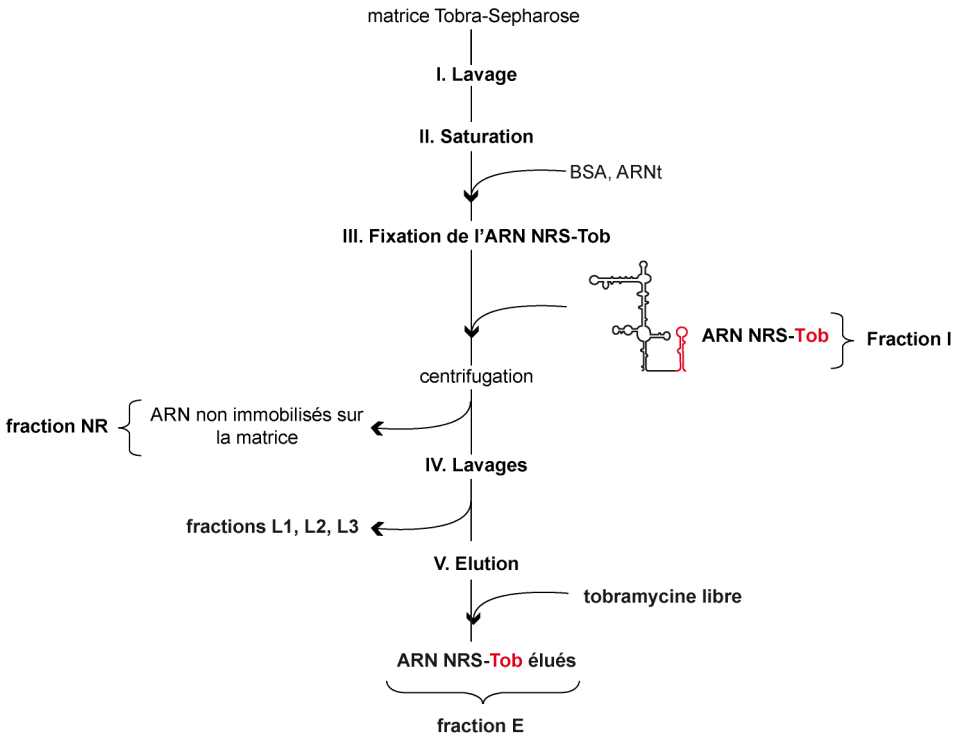



**Figure V-21: Test de fixation de l'ARN NRS-Tob sur la matrice Tobra-Sepharose.**
Cette expérience a servi à tester la fonctionnalité de la matriceTobra-Sepharose. Elle comprend 5 étapes : I) lavage, II) saturation, III) fixation de l'ARN NRS-Tob, IV) lavages et V) Elution. Les tests ont été réalisés avec un mélange d'ARN non radioactifs (60 pmol) et d'ARN uniformément marqués (20 fmol). La quantité d'ARN de chaque fraction a été évaluée en mesurant leur radioactivité au compteur à scintillation en Cerenkov. La fraction I correspond à l'Input (100 % de l'ARN au départ), NR, à l'ARN non retenu sur la matrice, L1, L2, L3, aux 3 lavages de la matrice et E, à l'éluat. L'ARN NRS-Tob est représenté par sa structure secondaire, l'élément NRS est en noir et l'aptamère Tob est en rouge. Le même protocole a été utilisé pour l'ARN NRS seul.

**Etape III :**

Avant d'être ajoutés à la matrice, un mélange de 100 pmol d'ARN NRS-Tob non radioactifs et 20 fmol d'ARN NRS-Tob uniformément marqués sont dénaturés 10 minutes à 65°C puis renaturés durant 10 min à température ambiante. Ils sont dilués dans un volume de 400 µl de tampon de fixation (Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 1 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; DTT 0,8 mM ; ARNt 0,1 mg/ml) puis sont ajoutés à la matrice. L'ensemble est incubé à 4°C durant 1h30 sous agitation rotative. Les ARN n'ayant pas été retenus sur la colonne sont éliminés après une centrifugation à 4°C à 4 000 g durant une minute (fraction NR).

**Etape IV :**

La matrice est lavée 3 fois avec 1 ml de tampon de lavage (Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 1 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; NP40 0,1 %), chaque lavage étant suivi d'une centrifugation à 4°C à 4 000 g durant une minute (fractions L1, L2, L3).

**Etape V :**

La matrice est incubée avec 250 µl de tampon d'élution (Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 3 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; DTT 0,8 mM ; tobramycine 5 mM ; glycérol 5 %) durant 10 minutes sous agitation à 4°C puis centrifugée à 4°C à 4 000 g durant une minute et le surnageant contenant les ARN NRS-Tob élués est récupéré (fraction E).

La quantité d'ARN présente dans les différentes fractions est calculée en mesurant leur radioactivité au compteur à scintillation en Cerenkov. La matrice Tobra-Sepharose est estimée être fonctionnelle lorsque 50-70% de l'ARN NRS-Tob et moins de 2 % de l'ARN NRS de départ sont respectivement fixés sur la matrice et que plus de 80 % de l'ARN NRS-Tob immobilisé sur la matrice est élué de la matrice (Hartmuth *et al.*, 2004).

### XI.1.3. Purification des complexes ribonucléoprotéiques sur l'ARN NRS-Tob

La purification des complexes ribonucléoprotéiques est divisée en 6 étapes: I) lavage de la matrice, II) saturation de la matrice, III) fixation de l'ARN NRS-Tob sur la matrice, IV) ajout de l'extrait nucléaire, V) lavages de la matrice, VI) élution des complexes ribonucléoprotéiques par ajout de tobramycine libre (**Figure V-22 et V-23**).

Les étapes I, II, III sont réalisées selon le même protocole décrit pour le test de la matrice Tobra-Sepharose (**paragraphe XI.1.2.2**). Une seconde expérience est réalisée en parallèle avec la matrice seule qui sert de témoin négatif.

Chaque expérience est réalisée au départ en doublon avec 25 µl de matrice. Les deux matrices sont traitées séparément de l'étape I jusqu'au premier lavage de l'étape V puis sont ensuite rassemblées à partir du second lavage de l'étape V jusqu'à l'élution des complexes ribonucléoprotéiques à l'étape VI.

Etape V :
Tampon d'assemblage : KCl 25 mM ; $MgCl_2$ 3 mM ; créatine phosphate 20 mM ; ATP 2 mM ; 35 % (v/v) d'extrait nucélaire de cellules HeLa

Etape VI :
Tampon de lavage : Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 1 mM ; $CaCl_2$ 1 mM ; KCl 75 mM ; ARNt 0,1 mg/ml ; DTT 0,2 mM ; NP40 0,1 %

Etape VII :
Tampon d'élution : Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 3 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; DTT 20 mM ; tobramycine 5 mM ; glycérol 5 %

**Figure V-22: Tampons utilisés pour la purification des complexes ribonucléoprotéiques.**

**Etape V :**

Un volume de 230 µl de tampon d'assemblage (KCl 25 mM ; $MgCl_2$ 3 mM ; Créatine phosphate 20 mM ; ATP 2 mM) contenant 35 % d'extrait nucléaire de cellules HeLa (soit 80 µl) est ajouté sur les billes. L'ensemble est incubé durant 30 minutes à 30°C afin que les complexes ribonucléoprotéiques puissent se former. L'excédent de protéines n'étant pas impliqué dans ces complexes est ensuite éliminé par une brève centrifugation à 4°C à 4 000 g (échantillon S).

**Etape VI :**

La matrice est lavée 1 fois avec 750 µl de tampon de lavage (Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 1 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; DTT 0,8 mM ; NP40 0,1 % ; glycérol 5%), suivi d'une brève centrifugation à 4°C à 4 000 g (échantillon L4). Lors du second lavage, chacune des matrices en doublon sont resuspendues dans 375 µl de tampon de lavage, rassemblées dans un même tube et centrifugées brièvement à 4°C à 4 000 g (échantillon W75 L5). Le troisième lavage est réalisé avec 750 µl de tampon puis la matrice est centrifugée brièvement à 4°C à 4 000 g (échantillon L6).

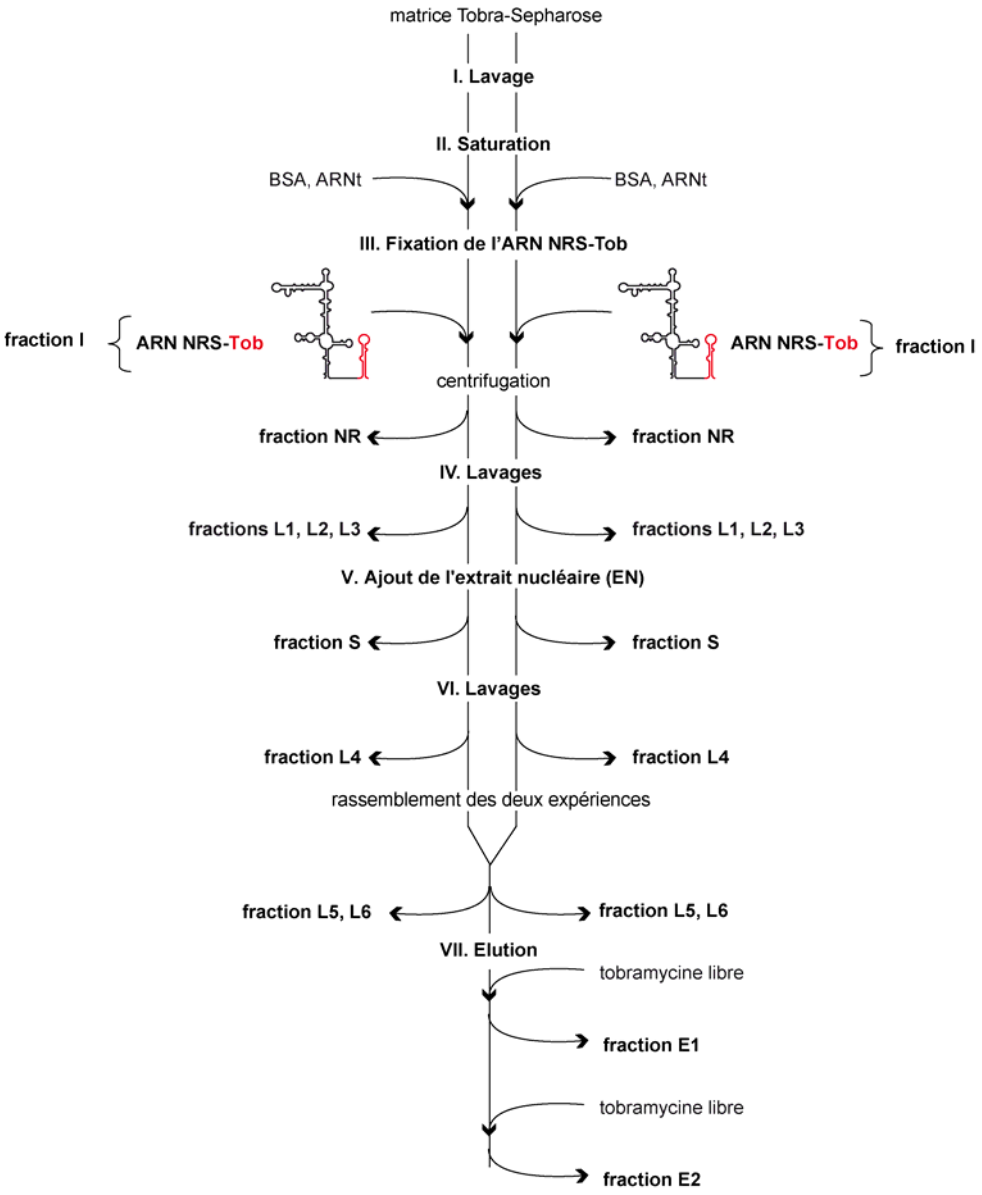



**Figure V-23: Purification des complexes ribonucéloprotéiques sur l'ARN NRS-Tob par chromatographie d'affinité sur la matrice Tobra-Sepharose.**
Cette expérience comprend 6 étapes : I) lavage, II) saturation, III) fixation de l'ARN NRS-Tob, IV) lavages V) ajout de l'extrait nucléaire, VI) lavages, VII) élution. Les étapes de l'expérience réalisées en doublon sont schématisées par deux flèches de part et d'autre du nom de l'étape concernée. L'expérience a été réalisée avec un mélange d'ARN non radioactifs (100 pmol) et d'ARN uniformément marqués (20 fmol). La quantité d'ARN dans les différentes fractions a été calculée en mesurant leur radioactivité en Cerenkov au compteur à scintillation. La fraction I correspond à l'Input, NR, à l'ARN non retenu sur la matrice, L, aux 3 lavages, S, au matériel non retenu sur la matrice, L4, L5, L6, aux 3 lavages, E1 et E2 aux deux élutions successives. L'ARN NRS-Tob est représenté par sa structure secondaire, l'élément NRS est en noir et l'aptamère Tob est en rouge. Le même protocole a été utilisé pour l'ARN NRS seul.

**Etape VII :**

Les complexes ribonucléoprotéiques sont successivement élués deux fois avec 250 µl de tampon d'élution (Tris-HCl 20 mM, pH 9,1 à 4°C ; $MgCl_2$ 3 mM ; $CaCl_2$ 1 mM ; KCl 145 mM ; DTT 0,8 mM ; tobramycine 5 mM ; glycérol 5 %), chaque élution étant suivie d'une brève centrifugation à 4°C à 4 000 g.

La radioactivité des différents échantillons prélevés au cours de l'expérience est mesurée au compteur à scintillation en Cerenkov afin de déterminer la quantité d'ARN NRS-Tob. Les acides nucléiques et les protéines contenus dans les échantillons sont séparés par une extraction phénol-chloroforme (1/1). Les acides nucléiques contenus dans la phase aqueuse sont ensuite précipités avec 2,5 volumes d'éthanol 96 % et 0,3 M d'acétate de sodium. Les protéines contenues dans la phase phénolique sont précipitées avec 4 volumes d'acétone. La précipitation des acides nucléiques et des protéines est réalisée durant 15 minutes et 1 heure à -80 °C respectivement.

### XI.1.4. Analyse des ARN contenus dans les complexes ribonucléoprotéiques

Après avoir été précipités, les culots d'acides nucléiques des différents échantillons collectés sont lavés avec 100 µl d'éthanol 70 %, séchés, puis repris dans un volume de 10 µl d' $H_2O$ auxquels sont ajoutés 10 µl de bleu de formamide (EDTA 125 mM ; bleu de bromophénol 0,03% ; bleu de xylène cyanol 0,03 %, le tout dans de la formamide désionisée). Le 1/10e de l'échantillon S, et la moitié des échantillons L4, L5, L6 et E (échantillons E1 et E2 rassemblés) sont fractionnés par une électrophorèse en gel dénaturant de polyacrylamide 10%-urée 8M dans du tampon TBE 0,5 X. A la fin de l'électrophorèse, les ARN sont révélés par une coloration au bromure d'éthidium (0,5 µg/ml) et sont visualisés sous UV.

### XI.1.5. Analyse du contenu en protéines des complexes ribonucléoprotéiques

#### XI.1.5.1. Analyse par SDS-PAGE

Après avoir été précipitées, les culots de protéines des différents échantillons collectés sont lavés avec de l'éthanol 70 %, séchés et repris dans 10 µl de tampon de dépôt (Tris-HCl 244 mM, pH 6,8 ; SDS 2,4% ; DTT 244 mM ; glycérol 21,2 % ; bleu de bromophénol 0,12 % (p/v)). L'échantillon correspondant aux protéines non retenues sur la matrice (échantillon S) est repris dans un volume de 120 µl de tampon de dépôt en raison de sa forte concentration en protéine. Le 1/10e de l'échantillon S, et la moitié des échantillons L4, L5, L6 et E

(échantillons E1 et E2 rassemblés) sont fractionnés par SDS-PAGE et le gel est ensuite coloré avec du bleu de coomassie colloïdal (**paragraphe X.1)**.

### XI.1.5.2. Analyse par Western Blot.

Les échantillons L4, L5, L6, E (fractions E1 et E2 rassemblées) et 1/10$^{e}$ de l'échantillon S sont fractionnés par SDS-PAGE puis analysés par Western blot avec des anticorps spécifiquement dirigés contre les protéines ASF/SF2 et 9G8 selon le protocole décrit dans le **paragraphe X.2**.

## XI.2. Purification des complexes ribonucléoprotéiques par chromatographie d'affinité reposant sur l'emploi de la protéine de l'enveloppe du phage MS2

La formation et la purification des complexes ribonucléoprotéiques par chromatographie d'affinité reposant sur l'emploi de la protéine de l'enveloppe du phage MS2 est illustrée dans la **Figure V-24**.

Une quantité de 250 pmol d'ARN fusionné à l'ARN MS2 non radioactif sont dénaturés 10 min à 65°C dans un volume de 25 µl de tampon D (EDTA 0,2 mM ; KCl 0,1 M; DTT 0,5 mM; PMSF 0,25 mM; HEPES 20 mM pH 7,9 ; glycérol 20%) en présence de 2,8 mM de $MgCl_2$ puis renaturés 10 min à température ambiante et placés dans la glace. L'ARN est ensuite incubé avec 1250 pmol de protéine MBP-MS2 diluée dans un volume de 100 µl de tampon D contenant 2,8 mM de $MgCl_2$ durant 30 minutes à 4°C sous agitation rotative. Lorsque les complexes ARN-MS2/MBP-MS2 sont formés, ceux-ci sont incubés durant 30 minutes à 4°C sous agitation rotative avec un volume de 1,23 ml de tampon d'assemblage (KCl 25 mM ; $MgCl_2$ 3 mM ; créatine phosphate 20 mM ; ATP 2 mM) contenant 30 % d'extrait nucléaire de cellules HeLa (v/v) (soit 4 mg de protéine) et un excès molaire d'ARNt de 100 fois. Ces complexes ribonucléoprotéiques sont ensuite ajoutés sur 200 µl de billes d'amylose (NEB) préalablement lavées 3 fois avec 1 ml de tampon D durant 1 heure à 4°C sous agitation rotative. Le matériel non retenu sur les billes est éliminé par une brève centrifugation à 2 000 g à 4°C (fraction NR). Les billes sont ensuite lavées 3 fois avec 2 ml de tampon D contenant 2,8 mM de $MgCl_2$, chaque lavage étant suivi d'une brève centrifugation à 4°C à 2 000 g (fractions L1, L2, L3). Les complexes ribonucléoprotéiques sont ensuite élués par l'ajout successif de 2x100 µl de tampon d'élution contenant 15 mM de maltose dans du tampon D complété avec 2,8 mM de $MgCl_2$ (fractions E1 et E2), chaque élution étant suivie d'une brève centrifugation à 4°C à 2 000 g.

Afin de suivre la formation des complexes au cours de l'expérience, 10 µl de chaque fraction prélevée au cours de l'expérience sont déposés sur un gel natif de polyacrylamide 5 % qui est ensuite coloré au bromure d'éthidium.

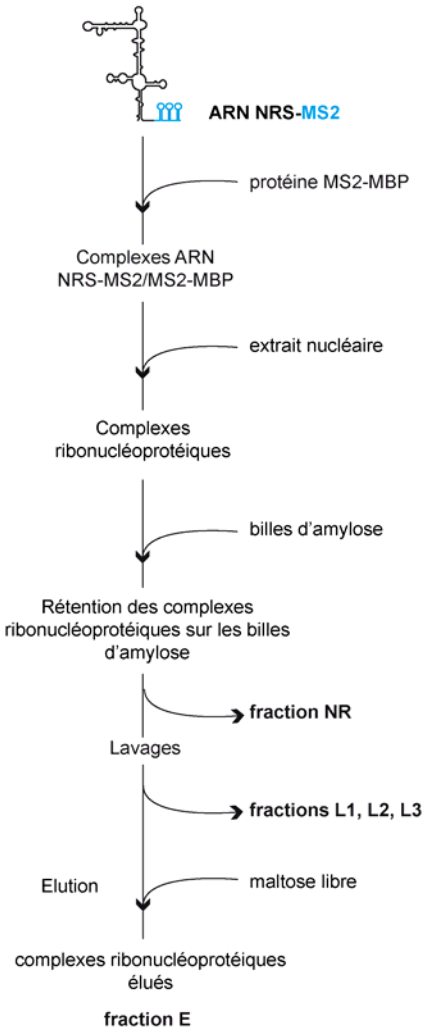



**Figure V-24: Purification des complexes ribonucléoprotéiques par chromatographie d'affinité reposant sur l'emploi de la protéine de l'enveloppe du phage MS2.**
L'expérience a été réalisée avec 250 pmol d'ARN NRS -MS2 et 1250 pmol de protéine MS2-MBP. Les fractions collectées durant l'expérience sont mentionnées, la fraction NR correspond au matériel non retenu sur les billes d'amylose, L1, L2, L3 aux 3 lavages des billes et E à l'élution. L'ARN NRS-MS2 est représenté par sa structure secondaire, l'ARN NRS est en noir et l'ARN MS2 est en bleu.
Le même protocole a été utilisé avec l'ARN MS2 et les autres éléments NRS fusionnés à l'ARN MS2.

### XI.2.1. Analyse du contenu en ARN des complexes ribonucléoprotéiques par Northern Blot

Les ARN contenus dans 140 µl d'élution (fractions E1 et E2 rassemblées) sont purifiés par une première extraction au phénol-chloroforme (1/1), une seconde extraction au chloroforme alcool isoamylique (1/24) et sont précipités avec 2,5 volumes d'éthanol 96 % et 0,3 M d'acétate de sodium. Les culots sont ensuite lavés avec 100 µl d'éthanol 70 %, séchés puis repris dans un volume de 50 µl d' $H_2O$. Un volume de 7,5 µl auquel est ajouté 10 µl de bleu de formamide sont déposés sur un gel dénaturant acrylamide 10 %-urée 8M. La migration est réalisée dans un tampon, TBE 0,5X à 30 mA. Après l'électrophorèse, les ARN sont transférés sur une membrane de nylon (0,45 µm, Hybond N+, GE Healthcare) par champ électrique à une intensité de 3 mA/cm2 (Biorad Trans Blot SD semi dry Transfer cell) durant 30 minutes. Les ARN sont ensuite fixés covalemment à la membrane par une exposition aux UV durant 5 minutes. La membrane est placée dans un tube à hybridation à 37°C en présence de 20 ml de tampon de pré-hybridation (SSPE 6X : NaCl 1 M ; $NaH_2PO_4$,$H_2O$ 66 mM ; EDTA 6 mM, pH 7,4) ; réactif de Denhardt 5X (Ficoll type 400 0,1% ; polyvinylpyrrolidon 0,1% ; BSA fraction V poudre 0,1%) ; SDS 0,5 %, préchauffé à 37°C) sous agitation rotative durant 1 heure. Les sondes utilisées sont des oligonucléotides (**Tableau V-1**) radioactivement marqués en 5' selon le protocole de marquage indiqué dans le **paragraphe VII.4.3.1**. Un volume de 1 ml de tampon de pré-hybridation contenant la totalité du mélange réactionnel ayant servi au marquage de la sonde est filtré à travers un filtre de 0,22 µm puis est ajouté au tampon dans le tube à hybridation. L'hybridation est réalisée toute la nuit sous agitation à 37°C dans le four à hybridation. La membrane est ensuite lavée 4 fois avec 50 ml de tampon SSPE 6X-SDS 0,1 % à 42°C. La membrane est ensuite placée en fluorographie en présence d'écran amplificateur et les bandes sont quantifiées à l'aide du PhophorImager.

### XI.2.2. Analyse du contenu en protéine des complexes ribonucléoprotéiques

#### XI.2.2.3. Analyse par SDS-PAGE

Le contenu en protéines des différents échantillons collectés au cours de la chromatographie d'affinité après l'ajout de l'extrait nucléaire a été analysé par SDS-PAGE. Un volume de 10 µl de chaque échantillon (soit 0,5 % du volume des fractions NR et L1, L2, L3 et 10 % des élutions E1 et E2) auquel est ajouté 10 µl de bleu SBL (Tris 250 mM, pH 6,8 ; SDS 8 % ; β-mercaptoéthanol 20 % ; bleu de bromophénol 0,04 %, glycérol 10 %) sont chauffés 2 minutes à 96°C puis déposés sur le gel. Les protéines sont révélées avec une coloration du gel au bleu de coomassie colloïdal selon le protocole détaillé au **paragraphe X.1**.

#### XI.2.2.4. Analyse par Western Blot

Un volume de 5 µl de l'élution (soit 1,25 % du volume des élutions E1 et E2 rassemblées) est déposé sur un gel SDS-PAGE ainsi que 2 µg de protéines totales d'extrait nucléaire de cellules HeLa qui sert de témoin positif. L'expérience de Western blot est réalisée selon les conditions décrites dans le **paragraphe X.2** avec des anticorps spécifiquement dirigés contre les protéines ASF/SF2, 9G8, SC35, hnRNP A1 et hnRNP H. Afin de s'assurer que toutes les pistes soient équilibrées, la membrane est colorée au Rouge Ponceau avant l'étape de saturation pour visualiser la quantité de protéine MBP-MS2 présente dans chaque piste.

#### XI.2.2.5. Analyse par spectrométrie de masse

Un volume de 60 µl de l'élution (soit 30 % de du volume des élutions E1 et E2 rassemblées) auquel est ajouté 20 µl de bleu de dépôt SBL 4X (Tris 1 M, pH 6,8 ; SDS 32 % ; β-mercaptoéthanol 80 % ; bleu de bromophénol 0,16%, glycérol 40 %) sont chauffés 2 minutes à 96 °C puis déposés sur un gel SDS-PAGE. L'électrophorèse est réalisée dans du tampon de migration (Tris 25 mM ; glycine 192 mM ; SDS 0,1 %) à 25 mA. Le gel est ensuite coloré avec du bleu de coomassie colloïdal selon le protocole détaillé dans le **paragraphe X.1**. Après la coloration, le gel est découpé tous les 2 mm et les bandes sont envoyées au laboratoire de spectrométrie de masse bio-organique ECPM pour analyse. Les protéines contenues dans les bandes de gel subissent une digestion trypsique puis les peptides sont ionisés selon la technique d'electro-spray et analysés par nanoLC-MS/MS selon le protocole détaillé par Miguet *et al.* (Miguet *et al.*, 2006).

## XII. Méthodes utilisées pour la culture cellulaire

### XII.1. Lignée de cellules HeLa

La lignée cellulaire HeLa est cultivée dans un milieu DMEM (Dulbecco's Modified Eagle Medium) complémenté avec 10 % de SVF (sérum de veau fœtal, FBS), 10 U/ml de pénicilline/streptomycine et 2 mM de glutamine (Gibco). Les cellules sont cultivées en flasques de 75 $cm^2$ à 37°C sous une atmosphère saturée en humidité et à 5% de $CO_2$.

### XII.2. Entretien des cellules

Les cellules sont cultivées dans des flasques de 75 $cm^2$ et lorsque la culture arrive à confluence, le milieu de culture est retiré et les cellules sont décollées par l'ajout de 4 ml de trypsine (0,5 g/l ; EDTA 0,2 g/l) durant 5 minutes à 37°C. Un volume de 8 ml de milieu est

ajouté pour remettre les cellules en suspension puis le tout est centrifugé 5 minutes à 800 g. Le culot de cellules est ensuite lavé avec 10 ml de milieu, remis en suspension et 1 ml est utilisé pour ensemencer 15 ml de milieu disposé dans une nouvelle flasque.

### XII.3. Congélation-décongélation

Les stocks de cellules congelées sont préparés dans du SVF contenant 10 % de DMSO. Les cellules sont congelées lentement jusqu'à - 80 °C, puis conservées dans l'azote liquide.

La décongélation des cellules se fait en les réchauffant rapidement à 37 °C dans leur milieu de culture usuel. Les cellules sont utilisées pendant un nombre de passage limité (environ 20) avant la décongélation d'un nouveau stock.

### XII.4. Conditions des transfections transitoires lors des surexpression des protéines SR

Vingt-quatre heures avant la transfection, une plaque de 6 puits est ensemencée à raison de 200 000 cellules/puits afin d'obtenir des cellules en phase exponentielle de croissance (environ 50 à 70 % de confluence).

Les cellules contenues dans un puits sont transfectées avec 2 µg d'ADN plasmidique total. Pour les expériences d'épissage *in cellulo*, l'ADN plasmidique utilisé pour la transfection est composé de 0,5 µg de plasmide de la construction d'épissage (**Figure V-11**), 0,1 µg de plasmide phRGB (plasmide codant la Luciférase Renilla) et de 1,4 µg de plasmide pXJ42 (plasmide non recombinant). Pour les expériences d'épissage *in cellulo* avec une surexpression des protéines SR, l'ADN plasmidique utilisé pour la transfection est composé de 0,5 µg de plasmide de la construction d'épissage, 0,1 µg de plasmide phRGB, 0,5 µg de plasmide pXJ42-SR (construction contenant la séquence codante des protéines SR) et 0,9 µg de plasmide pXJ42 (plasmide non recombinant). Lors de la transfection, l'ADN plasmidique est dilué dans un volume final de 100 µl de NaCl 9% puis mélangé avec 4 µl de réactif de transfection JetPEI dilué aussi dans un volume final de 100 µl de NaCl 9% (PolyPlus Transfection) de façon à obtenir un ratio ADN/JetPEI de 1:2 selon les recommandations du fabricant. Les complexes ADN/JetPEI sont formés pendant 20 minutes à température ambiante, puis 200 µl de ce mélange de transfection est ajouté par puits. Les cellules sont ensuite incubées 48 heures et lavées 2 fois avec du PBS 1X stérile (KCl 2,7 mM ; $KH_2PO4$ 1,8 mM ; NaCl 137 mM ; $Na_2HPO_4$ 10,1 mM). Les cellules sont ensuite décollées par l'ajout de 100 µl de trypsine par puits, lavée avec 1 ml de PBS 1X et le culot cellulaire est resuspendu dans 1 ml de PBS 1X.

### XII.5. Conditions des transfections transitoires lors des expériences de siRNA

Les cellules HeLa ont été cultivées dans un milieu Opti-MEM (Invitrogen) complémenté avec 5 % de SVF et sans antibiotique. Vingt-quatre heures avant la première transfection, une plaque de 6 puits est ensemencée à raison de 200 000 cellules/puits. Les cellule ont été transfectées deux fois à 24 h d'intervalle avec un siRNA dirigé soit contre l'ARNm codant la protéine 9G8 (si9G8, **Tableau V-1**) (Blaustein *et al.*, 2005), soit contre l'ARNm codant la protéine Luciférase Renilla (siLuc, **Tableau V-1**) (Dreumont *et al.*, 2010). Pour cela, les siRNA ont été dilués dans un volume final de 250 µl de milieu Opti-MEM et ont été mélangés avec 5 µl de réactif de transfection lipofectamine 2000 (Invitrogen) dilués dans un volume final de 250 µl selon les recommandations du fournisseur. Les complexes siRNA/lipofectamine sont formés pendant 20 minutes à température ambiante puis 500 µl de ce mélange de transfection sont ajoutés par puits.

Vingt-quatre heures après, les cellules sont transfectées avec l'ADN plasmidique : 0,5 µg de plasmide de la construction d'épissage et 3,5 µg de plasmide pBSK (plasmide non recombinant) sont dilués dans un volume final de 250 µl de milieu Opti-MEM puis mélangés avec 10 µl de lipofectamine diluée dans un volume final de 250 µl. Les complexes ADN plasmidique/lipofectamine sont formés durant 20 min à température ambiante puis 500 µl de ce mélange de transfection sont ajoutés par puits. Les cellules sont ensuite incubées 24 heures puis lavées 2 fois avec du PBS 1X stérile (KCl 2,7 mM ; $KH_2PO4$ 1,8 mM ; NaCl 137 mM ; $Na_2HPO_4$ 10,1 mM). Les cellules sont ensuite décollées par l'ajout de 100 µl de trypsine par puits puis lavée avec 1 ml de PBS 1X et le culot cellulaire est resuspendu dans 1 ml de PBS 1X.

### XII.6. Extraction des ARN et analyse des produits d'épissage par transcription inverse

Les ARN totaux sont extraits avec le kit GenElute$^{TM}$ Mammalian Total RNA (Sigma-Aldrich) selon les recommandations du fournisseur. La concentration des ARN est déterminée en mesurant l'absorbance à 260 nm et 4 µg d'ARN totaux sont dilués dans le tampon TURBO DNase 1X (Ambion) et traités avec 4 U de TURBO DNase (Ambion) durant 30 min à 37°C. La réaction est arrêtée par l'ajout de 20 µl d'EDTA 30 mM suivi d'une incubation de 15 min à 75°C. Une quantité d'1 µg d'ARN est utilisé pour la réaction de transcription inverse. Le protocole utilisé est identique à celui décrit pour les réactions d'épissage *in vitro* (**paragraphe IX.2.4.2**), excepté l'oligonucléotide utilisé qui est dans ce cas un oligo$(dT)_{18}$. L'amplification des produits de transcription inverse par PCR, l'électrophorèse et la quantification est

identique au protocole décrit pour les réactions d'épissage *in vitro* (**paragraphes IX.2.4.2 et IX.2.4.3**).

### XII.7. Extraction des protéines et analyse par Western blot

Le culot cellulaire est resuspendu dans 150 µl de tampon Gerdy (50 mM Tris-HCl, pH 8,0 ; 20 mM EDTA ; 5% SDS) puis les cellules sont traitées aux ultrasons sous une puissance de 10 W avec des impulsions séparées d'une seconde durant 10 sec. Le lysat est centrifugé à 4°C durant 10 min à 13 000g, le surnageant est prélevé et mélangé à du bleu de SBL 1X final. Les protéines contenues dans un volume de 10 µl sont fractionnées par gel SDS-PAGE puis analysées par expérience de Western blot (**paragraphe X.2**).

# Résultats

# Chapitre VI : Analyse structurale de l'élément NRS des virus de la famille des ASLV

## I. Etat des travaux lorsque je suis arrivée au laboratoire

Comme nous l'avons montré dans l'introduction, la structure 2D des ARN peut jouer un rôle important dans la régulation de l'épissage. D'où l'intérêt d'étudier comme nous l'avons déjà dit la structure 2D de l'élément du virus RSV. A. Mougin (CR1, CNRS) avait entrepris une étude expérimentale de l'élément NRS du virus RSV. Après son départ à Toulouse, j'ai poursuivi ce travail dans le cadre d'abord de mon stage de M2, puis de mon travail de thèse. J'étais encadrée par N. Robas (MC, UHP) et V. Marchand, une étudiante en thèse, qui avait déjà acquis une expérience en analyse de la structure 2D des ARN, m'a conseillée. Je résumerai brièvement les données qu'avait obtenues A. Mougin avant mon arrivée.

K. Beemon (Pr, Johns Hopkins University) avait fourni à l'équipe de C. Branlant des clones contenant des fragments NRS qui correspondaient aux éléments NRS entiers du virus RSV de la souche Prague C et du virus ALV de la souche LR9, ainsi que leurs formes tronquées, respectivement appelées, Δ76 et ΔLR9. A. Mougin avait alors cloné sous contrôle d'un promoteur T7 l'ADNc de l'élément NRS, c'est-à-dire le segment compris entre les positions 703 à 930. Elle avait réalisé une analyse de structure 2D par hydrolyses enzymatiques en utilisant des ARN marqués au $^{32}P$ soit à leur extrémité 3', soit à leur extrémité 5'. D'autre part, elle avait produit un ARN NRS comportant une extension en 3' de façon à pouvoir après hydrolyse enzymatique ou modification chimique par le DMS ou la CMCT réaliser une analyse des positions de coupures et de modifications par extension d'amorce (**Figure VI-1**). Dans les deux cas, des hydrolyses enzymatiques avaient été réalisées à différentes concentrations en $Mg^{2+}$, que l'on sait stabiliser la structure des ARN. Les données d'A. Mougin obtenues à forte concentration en $Mg^{2+}$ étaient compatibles avec la structure 2D proposée par Mfold mais un nombre faible de clivages étaient détectés à plus faible concentration en $Mg^{2+}$. Une analyse plus détaillée avait pu être obtenue et les données n'étaient pas en accord avec le modèle prédit par Mfold. A Mougin avait alors retenu une structure compatible avec les données obtenues à faible concentration en $Mg^{2+}$ (**Figure VI-2**). Dans ce modèle, le pseudo-site était localisé dans une courte structure tige-boucle et la région fixant ASF/SF2 dans une autre structure tige-boucle. On pouvait constater que les coupures enzymatiques obtenues à plus forte concentration en $Mg^{2+}$ étaient aussi compatibles avec ce

nouveau modèle structural. Néanmoins, la question se posait de savoir si l'élément NRS pouvait adopter deux structures 2D alternatives ou si le nouveau modèle de structure 2D représentait la structure authentique de cet élément. Par ailleurs, A. Mougin n'avait analysé la structure 2D que des 2 segments 715 à 773 et 894 à 930 de l'élément NRS et pas la totalité de la structure.

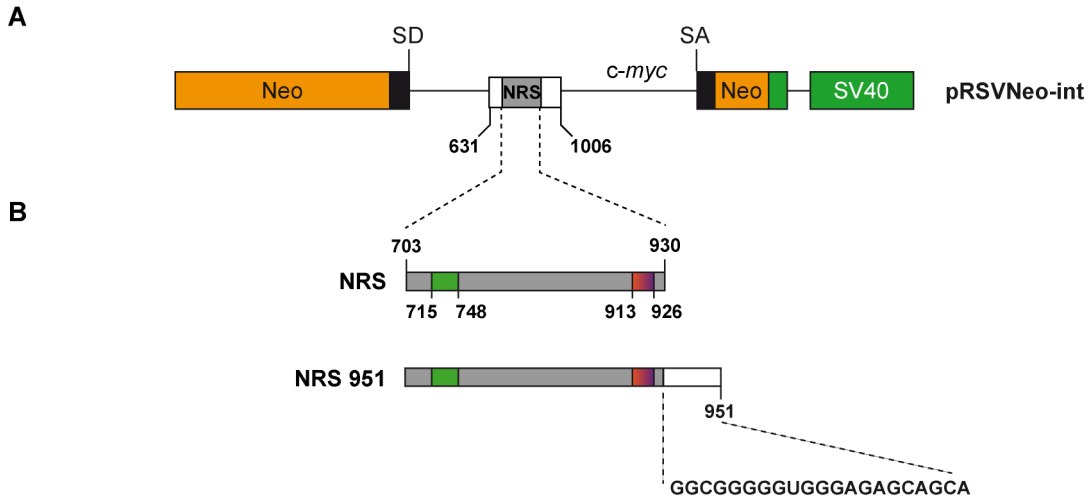



**Figure VI-1: Construction génétique utilisée et ARN produits par A. Mougin pour l'étude structurale de l'élément NRS.**
**A :** Construction pRSVNeo-int à partir de laquelle les matrices PCR du NRS ont été amplifiées. Cette construction est utilisée pour des expériences d'épissage *in cellulo* et nous a été donnée par K. Beemon. La séquence d'ADNc du virus RSV PrC (rectangle blanc) renfermant l'élément NRS (rectangle gris) a été insérée dans l'intron 2 du gène c-*myc* de poulet (trait noir). Les exons 2 et 3 du gène c-*myc,* les séquences du gène de résistance à la Neomycine (Neo) et le signal de polyadénylation du SV40 sont respectivement représentés par des rectangles noirs, oranges et verts.
**B :** ARN utilisés pour l'étude structurale de l'élément NRS. L'élément NRS est représenté par un rectangle gris, le site de fixation de la protéine ASF/SF2 et le pseudo site d'épissage sont respectivement représentés par des rectangles verts et orange-violets. Les positions des nucléotides sont précisées. L'ARN NRS 951 correspond à l'ARN NRS auquel a été rajouté en 3' la séquence bordante comprise entre les nucléotides 930-951. La séquence de cette extension est indiquée.

Sur la base des données d'A. Mougin, une analyse de la structure 3D du fragment 910 à 926 avait alors été réalisée par RMN par Cabello-Villegas (Cabello-Villegas *et al.*, 2004) en collaboration avec l'équipe de K. Beemon et A. Mougin de notre laboratoire ce qui avait permis de vérifier les capacités de ce segment à former une structure tige-boucle.

Il était nécessaire de compléter cette étude débutée par A. Mougin. C'est ce que j'ai entrepris lors de mon stage de M2 et le début de ma thèse.

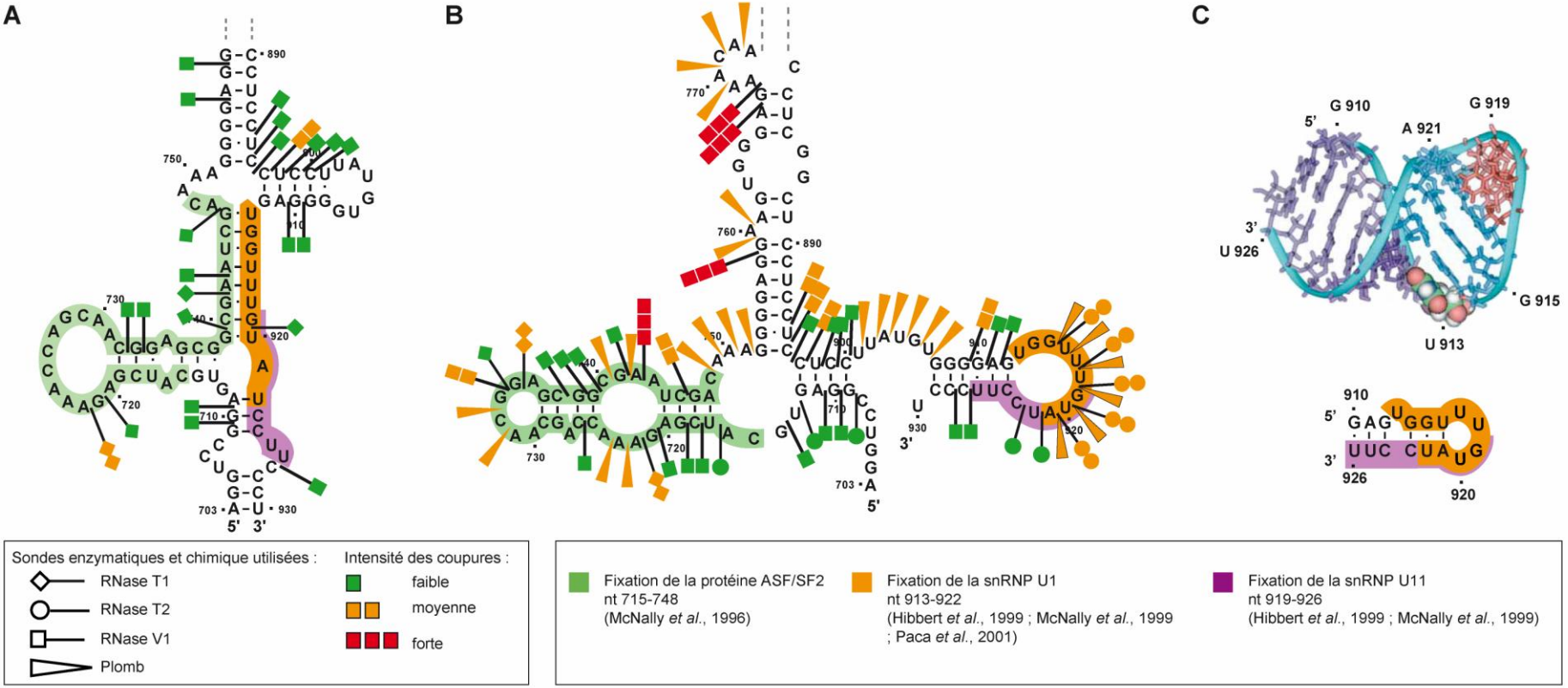



**Figure VI-2 : Comparaison des structures secondaires alternatives des régions 5' et 3' de l'élément NRS WT en l'absence (A) et en présence de 20 mM $MgCl_2$ (B) lors d'expériences de cartographie de la structure de l'ARN en solution avec la structure tertiaire du segment 910-926 de l'élément NRS obtenue par RMN (Cabello-Villegas *et al.*, 2004).**

Les positions des nucléotides sont indiquées et les clivages de l'ARN par les sondes enzymatiques et chimiques sont reportées sur l'ARN (A et B) selon la légende donnée dans l'encart de gauche. Les séquences fonctionnelles se trouvant aux extrémités 5' et 3' du NRS sont colorées selon la légende donnée dans l'encart de droite. Seuls les nucléotides dont la structure a été établie expérimentalement sont représentés, les séquences pour lesquelles aucune structure n'avait été déterminée sont symbolisées par des pointillés.

**A :** Structure secondaire proposée pour les régions 5' et 3' de l'élément NRS WT en l'absence de $MgCl_2$. La structure proposée est identique à celle prédite par informatique par McNally *et al.*, 1991.

**B :** Les données obtenues pour les régions 5' et 3' de l'élément NRS WT en présence de 20 mM de $MgCl_2$.

**C :** Représentation des structures tertiaire et secondaire du segment 910-926 de l'élément NRS. Les paires de bases appariées situées en amont et en aval du nucléotide U 913 sont respectivement colorées en violet et en bleu. Les atomes du nucléotide U 913 qui est en bulge sont dessinés et les nucléotides situés dans la boucle sont colorés en rouge. La structure secondaire du segment 910-926 est reportée en-dessous de la structure tertiaire.

# II. Stratégie adoptée pour compléter l'analyse structurale de l'élément NRS du virus RSV

## II.1. Choix de la séquence utilisée pour l'analyse structurale de l'élément NRS WT et son variant Δ76

Sachant que le NRS contient un pseudo-site d'épissage et que nous envisagions par la suite une étude de relation structure-fonction au niveau de l'épissage, nous avons choisi de réaliser les analyses structurales en tampon d'épissage, comme cela avait déjà été fait au laboratoire pour les études réalisées sur le VIH-1 (Jacquenet *et al.*, 2001). Ainsi, nous avons réalisé les expériences de cartographie de la structure de l'ARN en solution en tampon D contenant du $MgCl_2$ à une concentration de 4,4 mM.

Nous avons observé une forte conservation des séquences bordantes de l'élément NRS dans différents virus de la famille ASLV. De plus, les séquences présentes en amont étaient riches en résidus C et elles présentaient une forte complémentarité avec celles présentes en aval. Ce qui permettait d'envisager la formation d'une hélice fermant la structure de l'élément NRS. La formation de cette hélice était d'ailleurs prédite par utilisation du logiciel Mfold. D'où l'idée que ces séquences bordantes pouvaient stabiliser la structure adoptée par les séquences fonctionnelles de l'élément NRS. Nous avons donc réalisé l'étude structurale de l'élément NRS en incluant respectivement aux extrémités 5' et 3' de l'ARN produit par transcription *in vitro*, les séquences comprises entre les nucléotides 691-702 et 930-951 (**Figure VI-3**). Nous avons encore en plus rajouté à l'extrémité 3' de l'ARN, 5 nucléotides AAAUU non présents dans la séquence du virus RSV ceci afin d'avoir une séquence en simple brin à l'extrémité 3' de l'ARN sur laquelle pouvait se fixer l'oligonucléotide permettant la réalisation des transcriptions inverses destinées à identifier les positions des coupures enzymatiques.

La même stratégie a été utilisée pour l'analyse de l'élément NRS Δ76 qui avait été décrit comme l'élément minimal ayant une activité (**Figure VI-3**).

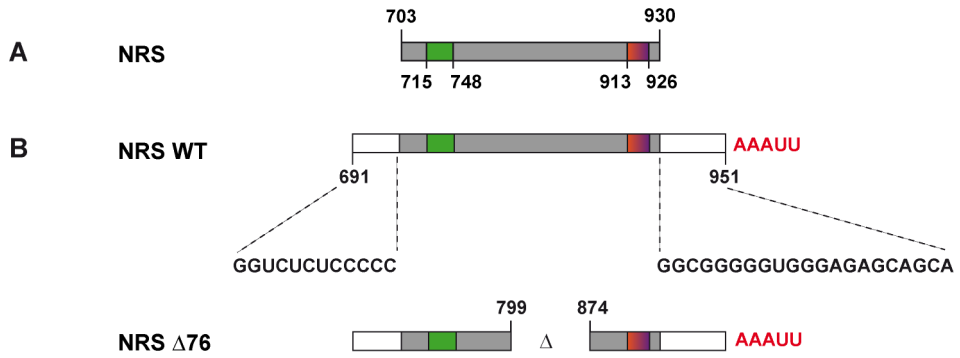



**Figure VI-3 : Représentation des ARN contenant l'élément NRS choisis pour l'étude structurale.**
L'élément NRS est représenté par un rectangle gris et les séquences le bordant sont représentées par des rectangles blancs. Le site de fixation de la protéine ASF/SF2 et le pseudo site d'épissage sont respectivement représentés par des rectangles vert et orange-violet. Les positions des nucléotides sont précisées.
**A :** Représentation de l'ARN de l'élément NRS fonctionnel.
**B :** Représentation des ARN utilisés pour l'étude structurale. L'ARN NRS WT représente l'ARN NRS auquel ont été ajouté en 5' la séquence 691-702 et en 3' la séquence 931-951. La séquence AAAUU est une séquence non naturelle qui a été ajoutée à l'extrémité 3' des ARN afin d'améliorer l'efficacité des transcriptions inverses. Les mêmes extensions sont présentes sur l'ARN NRS Δ76.

## II.2. Principes de l'emploi de sondes enzymatiques pour l'analyse de la structure des ARN en solution

Afin d'identifier les régions en simple et en double brin des ARN, nous avons utilisé 3 sondes enzymatiques :

- La RNase T1 qui coupe spécifiquement les liaisons phosphodiesters des ARN en 3' des résidus guanines non appariés, lorsqu'elle est utilisée en conditions limitantes. Cette enzyme libère des ARN avec des extrémités 3' phosphate et 5'-OH.
- La RNase T2 qui coupe également les régions en simple brin, lorsqu'elle est utilisée en conditions limitantes. Cette enzyme coupe les liaisons phosphodiesters de la même façon que la RNase T1, mais n'a pas de spécificité de reconnaissance vis-à-vis du résidu en 3' duquel elle clive la liaison phosphodiester.
- La RNase V1 qui permet d'identifier les résidus présents dans les régions en double brin ou les régions structurées par empilement de bases, sans spécificité vis-à-vis de l'identité des résidus. Cette enzyme clive les liaisons phosphodiesters en 5' des résidus et libère des fragments avec des extrémités 3'-OH et 5' phosphate.

Les conditions dans lesquelles ces sondes enzymatiques ont été utilisées sont décrites dans la partie Matériel et Méthodes.

### II.2.1. Identification des positions de clivage par transcription inverse et choix des oligonucléotides amorces

Les positions de coupures par les enzymes ont été identifiées par extension d'amorce par la transcriptase inverse. Lorsque l'ARN est clivé par la RNase V1, la présence du phosphate en 5' sur l'ARN coupé permet l'élongation de l'amorce par la transcriptase inverse jusqu'au dernier résidu. En revanche, l'absence de ce phosphate lorsque l'ARN est clivé par les RNases T1 ou T2 provoque l'arrêt de l'élongation au nucléotide précédent.

Les analyses par extension d'amorce nécessitent l'emploi d'oligonucléotides s'hybridant de façon spécifique à l'ARN. Nous avons défini 3 oligonucléotides amorce de façon à analyser l'ensemble de la séquence de l'élément NRS WT qui comporte 250 nucléotides. Les mêmes oligonucléotides ont servi aux extensions d'amorces de l'ARN Δ76, à l'exception de l'amorce O-1617 dont la séquence complémentaire se situe dans la séquence qui est tronquée de cet ARN. A la place, nous avons utilisé l'oligonucléotide O-2956 qui s'hybride en amont du site de troncation.

## II.3. Résultats obtenus

Les données obtenues nous ont permis de montrer que grâce à l'hélice formée par les séquences bordantes, l'élément NRS du RSV se replie selon une structure 2D très stable dans laquelle le pseudo-site 5' est bien situé dans une structure tige-boucle SLS4 et la région permettant la fixation de la protéine SR ASF/SF2 dans une autre structure tige-boucle SLS1 et ceci quelle que soit la concentration en $Mg^{2+}$ utilisée (**Figure VI-4**). Le reste de la séquence forme une grande tige-boucle irrégulière SLS2 et une petite tige-boucle SLS3 reliées aux tiges-boucles SLS1 et SLS4 par une longue hélice H2. La troncation réalisée dans l'élément NRS Δ76 ne touche que ces 2 structures tiges-boucles SLS2 et SLS3 (**Figure VI-4**). Elles sont alors remplacées par une petite tige-boucle qui prolonge l'hélice H2. Tout le reste de la structure en particulier les structures tiges-boucles SLS1 et SLS4 est conservé dans l'élément NRS Δ76, ce qui est en accord avec sa capacité à inhiber l'épissage. Nous aboutissions donc à une structure 2D dans laquelle le pseudo-site 5' appartient à un motif structural dont la capacité de formation avait été vérifiée par RMN (Cabello-Villegas *et al.*, 2004). L'autre élément de séquence important, celui fixant ASF/SF2, forme un motif structural indépendant selon le modèle établi. Néanmoins, le fait que ces 2 structures tiges-boucles soient reliées entre elles et avec les hélices H1 et H2 par une grande boucle interne laisse présager une flexibilité de l'ensemble de la structure grâce à cette grande boucle. On peut donc imaginer que les partenaires de chacune de ces structures tiges-boucles puissent interagir entre eux.

### II.4. Etudes comparatives des éléments NRS d'autres virus

Lorsque la structure 2D d'un ARN joue un rôle important sur le plan fonctionnel, elle est en général conservée par mutations compensatoires au cours de l'évolution et la structure de cet ARN est retrouvée dans plusieurs organismes ou virus. Si la structure que nous avions établie pour le virus RSV était importante sur le plan fonctionnel, on pouvait supposer que des mutations compensatoires devaient permettre de la conserver dans d'autres espèces de la même famille de virus. La mise en évidence de telles conservations était un argument en faveur de la validité de la structure proposée. Nous avons décidé d'analyser expérimentalement la structure 2D de l'élément NRS d'un autre virus de cette famille, le virus ALV LR9 dont la séquence diffère de celle de l'élément NRS de la souche PrC par 10 substitutions de bases, ceci afin d'apporter des arguments en faveur de la structure établie. L'élément NRS du virus ALV LR9 ainsi qu'un variant tronqué de cet élément avaient déjà été caractérisés. Cet élément tronqué était décrit comme ayant perdu sa fonction d'élément inhibiteur de l'épissage (Smith *et al.*, 1997).

En utilisant la même stratégie que celle décrite ci-dessus pour l'élément NRS de RSV, j'ai pu déterminer expérimentalement la structure 2D de l'élément NRS entier du virus LR9 (**Figure VI-5**). En dépit de la présence des 10 substitutions de bases, nous avons observé qu'il adopte en solution la même structure 2D que celle établie pour l'élément NRS du virus RSV. Les substitutions des nucléotides sont soit localisées dans des simples brins, soit renforcent la stabilité de l'hélice H2. Ces données apportaient un argument fort en faveur de la validité de la structure 2D établie pour RSV. Par ailleurs, la structure établie pour l'élément ΔLR9 non actif était fortement altérée par rapport à celle établie pour l'élément entier. Seules les structures tige-boucle SLS4, SLS3 et une partie de la structure tige-boucle SLS3 étaient conservées, tandis que la structure tige-boucle SLS1 était absente et l'hélice H2 totalement remodelée (**Figure VI-5**). Ce qui explique sans doute pourquoi cet élément n'est pas fonctionnel.

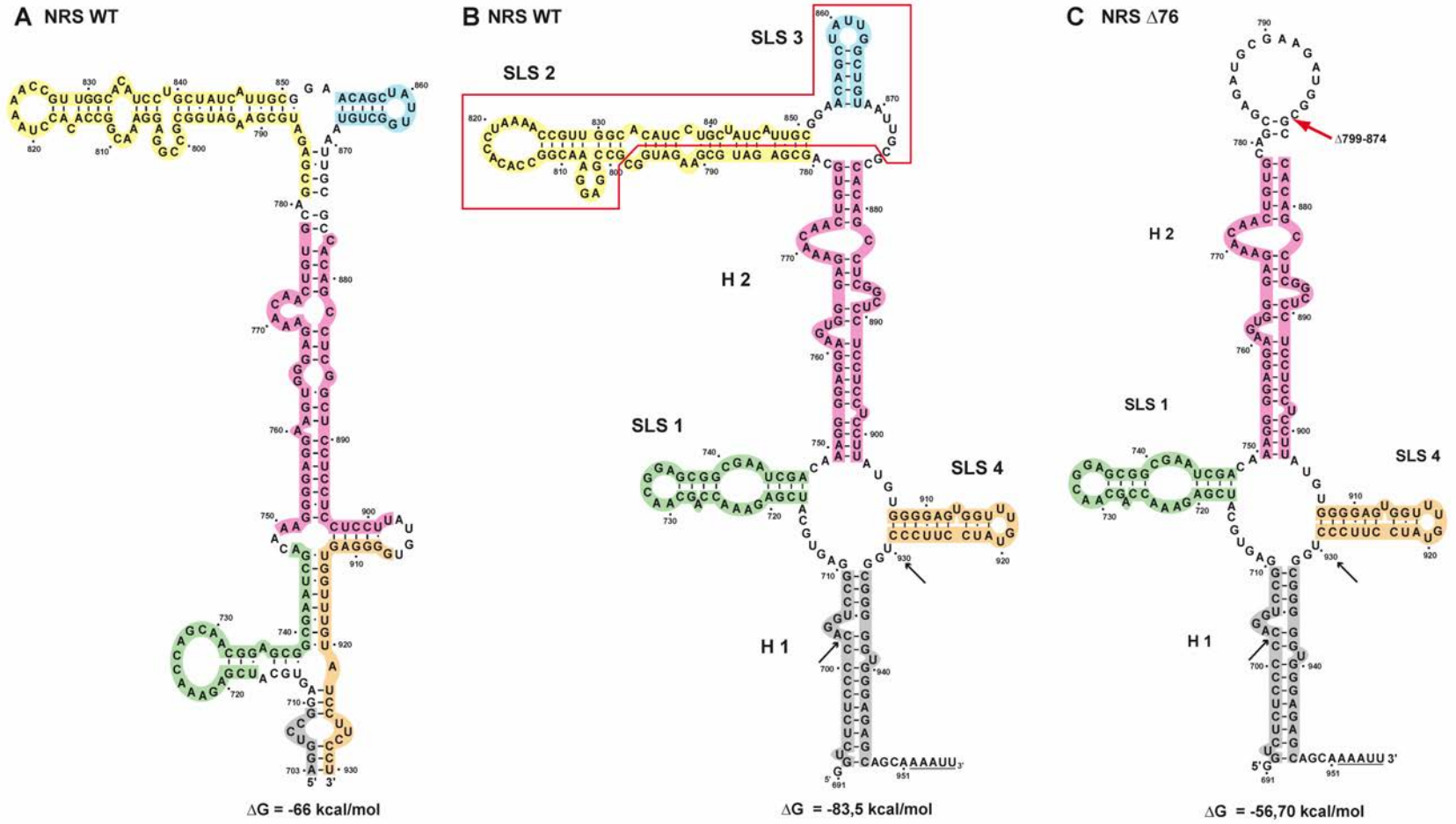



**Figure VI-4 : Comparaison de la structure secondaire prédite (A) et des structures établies expérimentalement (B) pour l'élément NRS WT et de l'élément NRS Δ76.**
**A :** Structure secondaire prédite pour l'élément NRS WT par le logiciel Mfold (McNally *et al*., 1991).
**B :** Structure secondaire établie expérimentalement pour l'élément NRS WT. La séquence qui est délétée dans le mutant NRS Δ76 est indiquée par un cadre rouge.
**C :** Structure secondaire établie expérimentalement pour l'élément NRS Δ76. Le site de délétion est indiqué par une flèche rouge.
Les noms de chacune des structures tibe boucles (SLS), ainsi que des hélices (H) sont reportées sur chacune des structures. L'énergie libre calculée par le logiciel Mfold pour chacune des structures et indiquée en-dessous de la structure. Les séquences colorées correspondent aux séquences des différents éléments structuraux déterminés pour le modèle de structure secondaire de l'ARN NRS WT établi expérimentalement (B) et ont été reportées sur la structure prédite de l'élément NRS WT (A) et la structure établie expérimentalement pour l'élément NRS Δ76 (C). Les séquences de l'hélice 1 (H1) sont colorées en gris, celles de la structure tige-boucle 1 (SLS1) en vert, celles de l'hélice 2 (H2) en rose, celles de la tige-boucle 2 (SLS2) en jaune, celles de la tige-boucle 4 (SLS4) en orange. Les extrémités 5' et 3' de l'élément NRS WT utilisé en (A) sont indiquées par des flèches noires sur les modèles de structure secondaire en (B) et en (C). Les nucléotides non naturels ajoutés à l'extrémité 3' de l'élément NRS sont soulignés.

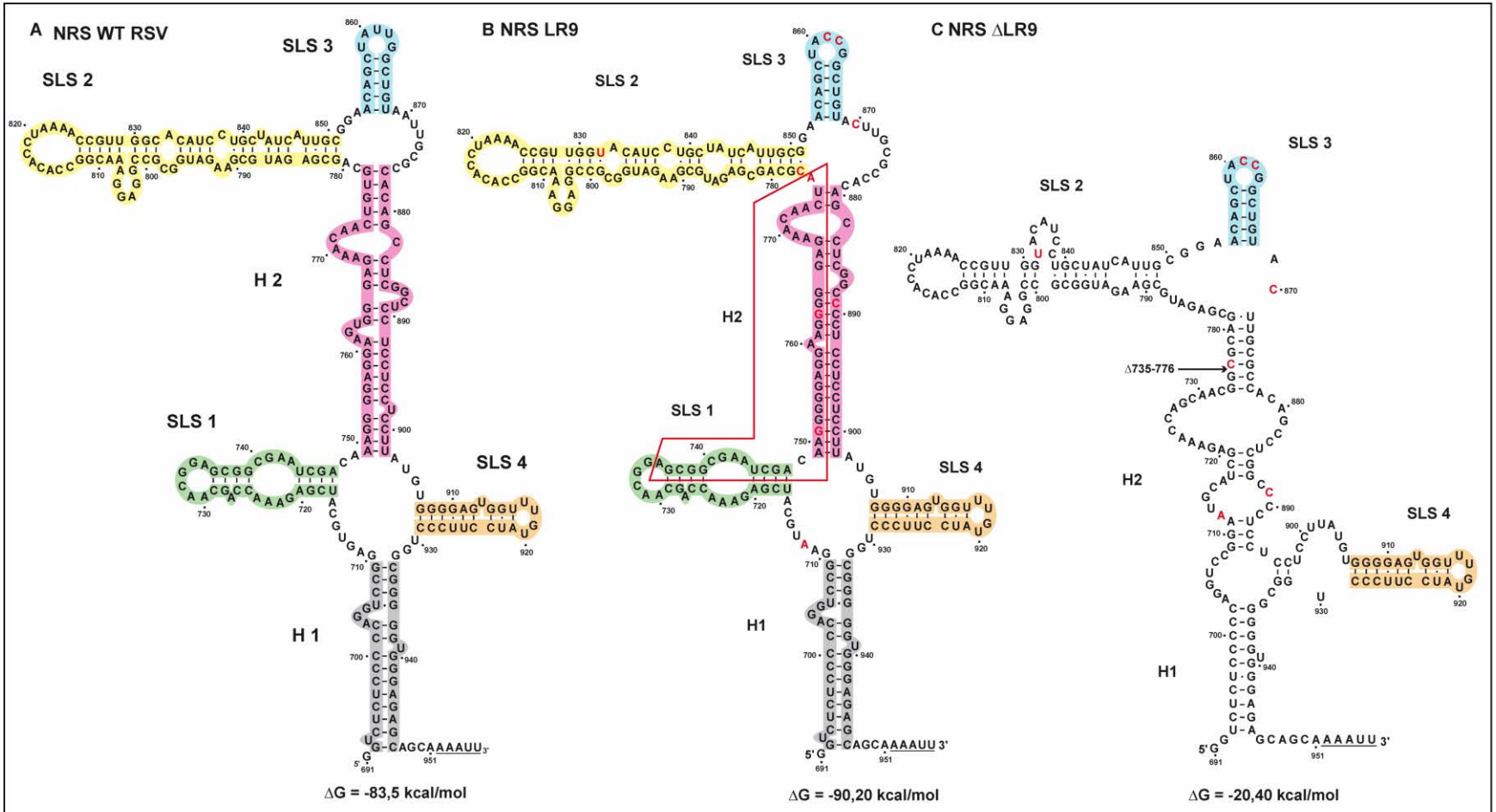



**Figure VI-5 : Comparaison des structures secondaires établies expérimentalement pour les éléments NRS WT (A) NRS LR9 (B) et NRS ΔLR9 (C).**
**A :** Structure secondaire établie expérimentalement pour l'élément NRS WT de la souche RSV PrC.
**B :** Structure secondaire établie expérimentalement pour l'élément NRS LR9. Les nucléotides correspondant aux substitutions de bases par rapport à l'élément NRS WT de la souche PrC sont colorés en rouge. La séquence qui est éliminée dans le mutant NRS ΔLR9 (C) est encadrée en rouge.
**C :** Structure secondaire établie expérimentalement pour l'élément NRS ΔLR9. Le site de délétion est indiqué par une flèche noire.
Les noms de chacune des structures tibe boucles (SLS) ainsi que des hélices (H) sont reportées sur chacune des structures expérimentales. L'énergie libre calculée par le logiciel Mfold pour chacune des structures et reportée en-dessous de la structure. Les séquences colorées correspondent aux séquences des différents éléments structuraux déterminés pour le modèle de structure secondaire de l'ARN NRS WT établi expérimentalement (A) et ont été reportées sur les structures expérimentales de l'élément NRS LR9 (B) et de l'élément NRS ΔLR9 (C). Les séquences de l'hélice 1 (H1) sont colorées en gris, celles de la structure tige-boucle 1 (SLS1) en vert, celles de l'hélice 2 (H2) en rose, celles de la tige-boucle 2 (SLS2) en jaune, celles de la tige-boucle 4 (SLS4) en orange. Les nucléotides non naturels ajoutés à l'extrémité 3' des éléments NRS sont soulignés.

Afin d'élargir la comparaison à un plus grand nombre de virus de la famille ASLV, j'ai alors entrepris une étude prédictive de la structure 2D des éléments NRS potentiels de 7 virus de la famille ASLV dont la séquence avait été établie. Pour cela, j'ai réalisé des alignements de séquences me permettant d'identifier les éléments NRS potentiels et ensuite pour chacun d'eux, j'ai testé s'il pouvait se replier selon le modèle que nous avons établi pour le RSV (**Figure VI-6**).

Cela s'est avéré être le cas pour tous les virus étudiés en dépit de variations de séquence importante au niveau de l'hélice H2 et des structures tige-boucle SLS2 et SLS3. Quelques exemples de mutations compensatoires sont représentés dans la **Figure VI-7**.

On peut remarquer que les substitutions de bases et courtes répétions de motifs présentes dans les différents éléments NRS potentiels étudiés n'ont pas un effet très déstructurant sur la capacité de structuration. Par exemple, la duplication de 7 nucléotides dans le virus ALV MQNCSU (A) située dans la partie apicale de l'hélice H2 ne perturbe pas la structure de cette hélice mais la renforce en la prolongeant de 4 appariements de bases. Le virus ALV ADOL-7501 comporte 6 substitutions de bases dans l'hélice H2 qui n'altère pas fortement cette hélice. Par ailleurs, bien que les séquences des structures tige-boucle fonctionnelles SLS 1 et SLS4 soient très conservées, on note la présence de mutations semi compensatoires dans la SLS1 chez le virus Schmidt Ruppin A et B et compensatoire dans la structure SLS4 du virus Y53 Sarcoma virus. Ces observations étaient en faveur d'une importance fonctionnelle de la structure 2D que nous avion établie pour l'élément NRS, soit dans un mécanisme d'inhibition de l'épissage, soit pour un autre mécanisme intervenant dans la multiplication du virus.

De ce fait, nous avons abordé l'étude du rôle fonctionnel possible des différents motifs structuraux. L'analyse des structures que nous venons de résumer est présentée dans l'article (**Chapitre VIII**).

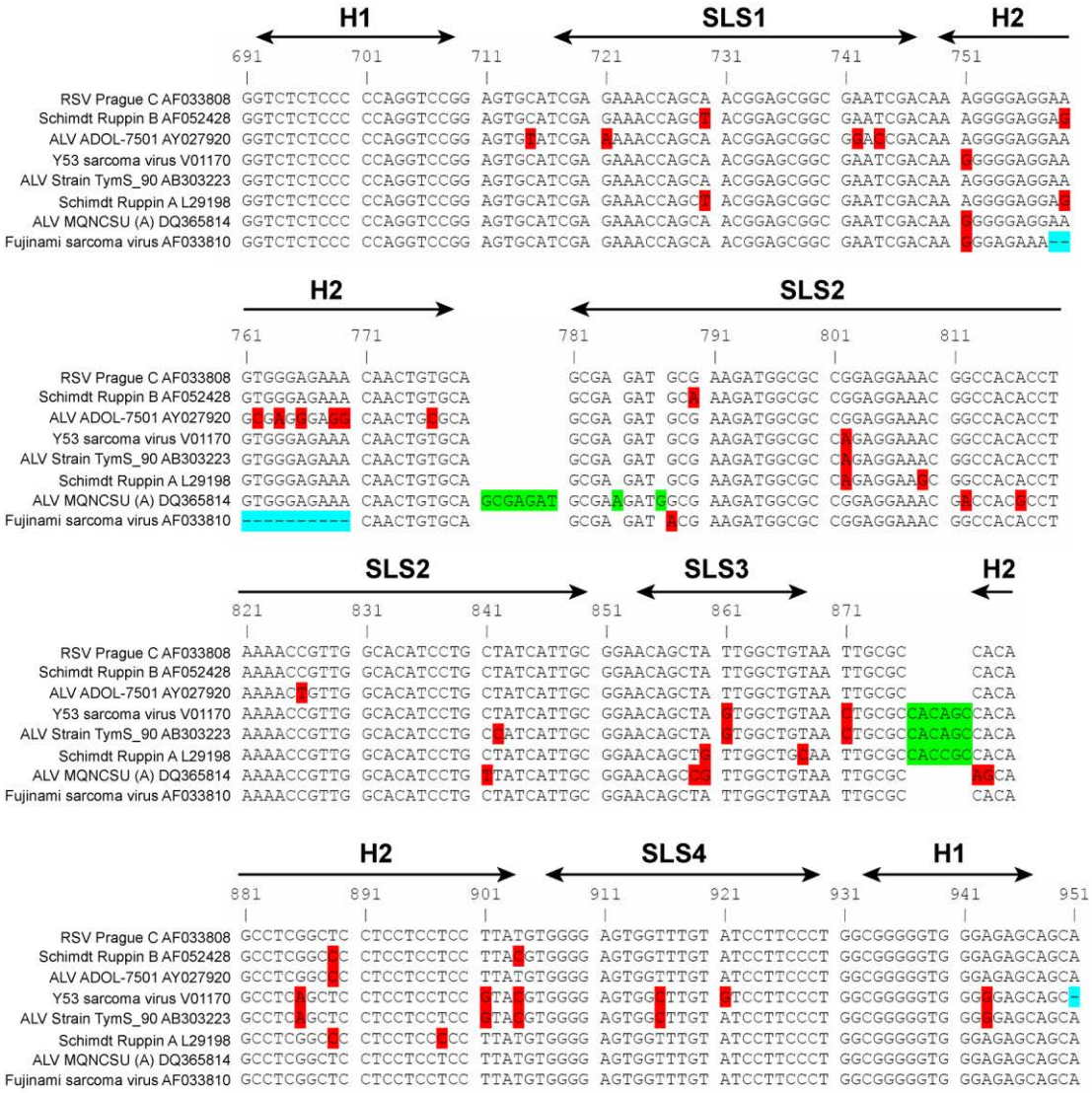



**Figure VI-6 : Alignement des séquences correspondant à l'élément NRS de différents virus de la famille ASLV.**

Le nom et le numéro d'accès à la Genbank des différents virus est indiquée devant chaque séquence. Les structures tige boucle (SLS) et les hélices (H) établies pour l'élément NRS du virus Prague C sont indiquées au dessus des alignements. Les bases colorées en rouge correspondent aux substitutions de bases par rapport au virus RSV Prague C. Les bases colorées en vert et bleu correspondent respectivement à des insertions et des délétions.

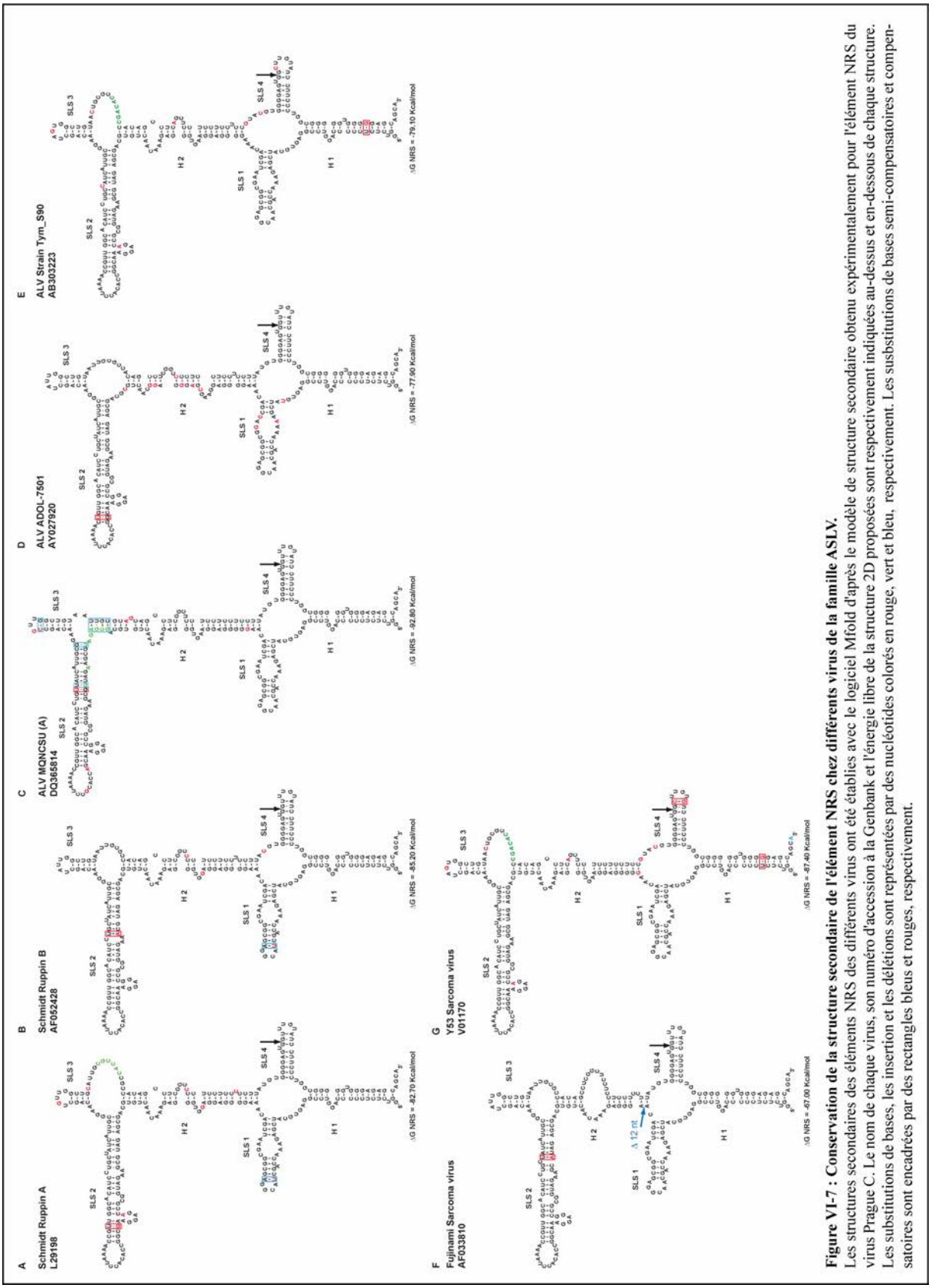



**Figure VI-7 : Conservation de la structure secondaire de l'élément NRS chez différents virus de la famille ASLV.**
Les structures secondaires des éléments NRS des différents virus ont été établies avec le logiciel Mfold d'après le modèle de structure secondaire obtenu expérimentalement pour l'élément NRS du virus Prague C. Le nom de chaque virus, son numéro d'accession à la Genbank et l'énergie libre de la structure 2D proposées sont respectivement indiquées au-dessus et en-dessous de chaque structure. Les substitutions de bases, les insertion et les délétions sont représentées par des nucléotides colorés en rouge, vert et bleu, respectivement. Les susbstitutions de bases semi-compensatoires et compensatoires sont encadrées par des rectangles bleus et rouges, respectivement.

# Chapitre VII : Etude des sites de fixation des protéines d'un extrait nucléaire de cellules HeLa sur les éléments NRS des virus RSV et ALV

## I. Objectifs

L'élément NRS contient un pseudo-site donneur d'épissage qui inhibe l'épissage du pré-ARNm du virus RSV. Son activité est en partie régulée par des facteurs *trans* qui se fixent sur certaines de ses séquences. En présence d'extrait nucléaire, le NRS forme un complexe ribonucléoprotéique qui a été purifié par chromatographie d'exclusion de taille et qui a été dénommé complexe NRS-C (Cook et McNally, 1996, 1998). Il existe une corrélation entre la formation du complexe NRS-C et la capacité de l'élément NRS à inhiber l'épissage. Bien qu'un certain nombre de facteurs présents au sein de ce complexe aient été identifiés, leurs sites de fixation n'ont pas été définis précisément, et la liste n'est pas exhaustive.

Le premier objectif de mon travail a consisté à définir les sites de fixation des facteurs de l'extrait nucléaire sur l'élément NRS et d'envisager l'identification de ces facteurs. Pour cela, j'ai utilisé la technique d'empreinte qui est la technique la plus adaptée pour définir précisément le site de fixation d'un facteur sur un ARN et qui n'avait jamais été utilisée auparavant sur l'ARN NRS. Cette technique avait également l'avantage d'être compatible avec les conditions d'épissage *in vitro* et les conditions dans lesquelles la structure secondaire de l'élément NRS avait été définie.

Nous avons réalisé ces expériences sur les éléments NRS WT et Δ76 du virus RSV et l'élément NRS du virus LR9 dont nous avons établi la structure secondaire. Nous avons montré que ces éléments ont des structures similaires et inhibent de façon équivalente l'épissage d'un pré-ARNm lorsqu'ils sont insérés dans l'intron (McNally *et al.*, 1991 ; Smith *et al.*, 1997).

## II. Stratégie

### II.1. Principe

Les expériences d'empreinte ont consisté à former des complexes ribonucléoprotéiques entre l'ARN NRS produit par transcription *in vitro* et un extrait nucléaire et à traiter les complexes formés par des ribonucléases. Des hydrolyses étaient réalisées dans les mêmes

conditions sur l'ARN libre afin d'identifier par comparaison les régions de l'ARN protégées par les composants de l'extrait nucléaire présents dans les complexes ribonucléoprotéiques.

### II.2. Choix de l'extrait nucléaire et de la sonde enzymatique

Dans ces expériences, nous avons utilisé des extraits nucléaires de cellules HeLa fonctionnels en épissage. Sachant que le NRS est assimilable à un site donneur isolé, il était important que l'extrait nucléaire comporte tous les composants du spliceosome afin de former sur le NRS des complexes similaires à ceux formés *in vivo*. Le fait d'utiliser cet extrait nucléaire humain nous plaçait dans des conditions hétérologues puisque les virus ASLV infectent les espèces aviaires. Néanmoins, l'élément NRS s'avère actif en extrait nucléaire de cellules HeLa lorsqu'on l'intègre dans un pré-ARNm utilisable en épissage *in vitro*. Il a aussi été montré qu'il est actif aussi bien dans des cellules humaines (HEK 293) que dans des cellules d'embryon de poulet (CEF) (Hibbert *et al.*, 1999). Par ailleurs, le séquençage de la totalité du génome du poulet (*Gallus*, *gallus*) et la recherche des gènes codant les composants du spliceosome et les protéines de régulation de l'épissage a montré que ceux-ci étaient très conservés entre le poulet et l'Homme.

Les expériences d'empreinte ont été réalisées en utilisant la RNase T1 comme sonde enzymatique. Lorsqu'elle est utilisée en conditions ménagées, cette enzyme clive les liaisons phosphodiesters en 3' des G en simple brin. Lors de l'étude de la structure 2D de l'élément NRS (**Chapitre VI**), nous avions constaté qu'elle permet d'obtenir de nombreux clivages sur l'ensemble de la séquence de l'élément NRS.

### II.3. Mise au point des conditions d'empreinte en présence d'un extrait nucléaire de cellules HeLa

Avant de réaliser les expériences d'empreinte, nous avons réalisé des expériences de retard sur gel pour définir les conditions dans lesquelles les ARN étaient totalement complexés avec les composants de l'extrait nucléaire. Pour cela, 1,5 pmol d'ARN non radioactif et 10 fmol d'ARN uniformément marqué ont été incubés en présence d'une quantité croissante d'extrait nucléaire selon le protocole décrit dans la partie Matériel et Méthodes. La totalité de l'ARN NRS WT était complexée en présence d'une quantité d'extrait nucléaire correspondant à 19,5 µg de protéines (**Figure VII-1, A**). Des résultats similaires ont été obtenus avec les ARN NRS Δ76 et NRS LR9 (**Figure VII-1, C, B**). Nous avons donc réalisé les expériences d'empreinte en utilisant une quantité d'extrait nucléaire correspondant à 39 µg de protéines totales et 3 pmol d'ARN froid.

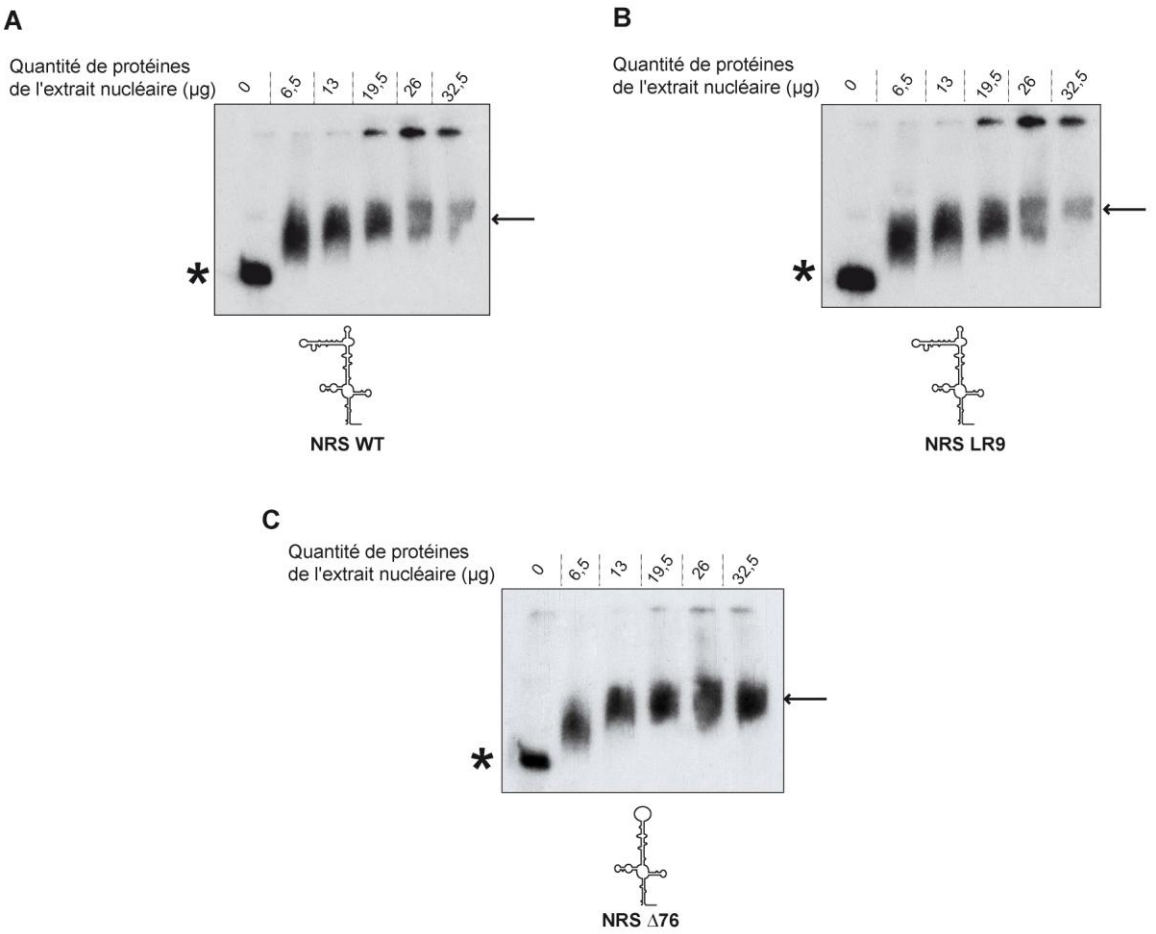



**Figure VII-1 : Analyse par expériences de retard sur gel des complexes ribonucléoprotéiques formés après incubation des ARN NRS WT du virus RSV (A), NRS LR9 du virus ALV-LR9 (B) et NRS Δ76 du virus RSV (C) en présence d'extrait nucléaire de cellules HeLa.**
Pour chacun des ARN, un mélange de transcrits non radioactifs (1,5 pmol) et uniformément marqués (10 fmol) a été incubé en présence d'une quantité croissante d'extrait nucléaire. L'ensemble a été fractionné par électrophorèse en gel de polyacrylamide 5% non dénaturant. La position des ARN libres est indiquée par une étoile et celle des ARN complexés est indiquée par une flèche. Les ARN sont schématisés en-dessous des autoradiogrammes et leur nom est indiqué.

## III. Identification des régions protégées

Les complexes ribonucléoprotéiques assemblés sur les ARN NRS WT, NRS Δ76 et NRS LR9 ont été soumis à une hydrolyse par la RNase T1 pendant 5 min à 30°C et les positions des clivages ont été identifiées par extension d'amorce par la transcriptase inverse en utilisant les mêmes amorces que pour les analyses de structure 2D (**Tableau X, Matériel et Méthode**). Les ADNc produits ont été fractionnés en gel de polyacrylamide dénaturant 7%. Chacune des expériences d'empreintes a été répétée au moins deux fois avec des lots d'ARN distincts, afin de s'assurer de la reproductibilité des résultats.

Des exemples de fractionnement des ADNc obtenus pour les ARN NRS WT, NRS-Δ76 et NRS LR9 sont respectivement représentés dans les **Figures VII-2, VII-3 et VII-4**.

Pour les 3 ARN, le nombre et la position des hydrolyses des ARN par la RNase T1 en présence ou en absence d'extrait nucléaire sont quasiment identiques, ce qui indique que la fixation des composants de l'extrait nucléaire ne provoque pas de remaniement marqué de la structure secondaire. Par conséquent, la diminution de l'intensité des coupures par la RNase T1 observées en présence d'extrait nucléaire est due à la fixation des facteurs de l'extrait nucléaire sur l'ARN. Cette diminution n'est pas observée sur certains G qui sont invariablement hydrolysés en présence ou en absence de l'extrait nucléaire. Cela renforçait l'idée que les diminutions de taux d'hydrolyse observées correspondaient à des protections.

### III.1.1. Les empreintes en extrait nucléaire recoupent le caractère bipartite de l'élément NRS

Pour les 3 ARN étudiés, nous observons des protections dans les structures tige-boucles SLS1 et SLS4 qui ont été respectivement proposées fixer des protéines SR (McNally et McNally, 1996) et les snRNP U1 et U11 (Hibbert *et al.*, 1999 ; McNally et McNally, 1999). De manière intéressante, dans les éléments entiers NRS WT des virus RSV et LR9, nous observons aussi une forte protection de la structure tige-boucle SLS2 pour laquelle aucun partenaire protéique n'a encore été défini. La partie de la séquence SLS2 présente dans l'élément NRS Δ76 (nt 781-796) est aussi fortement protégée en extrait nucléaire, ce qui suggère que cette séquence a une forte affinité pour des composants de l'extrait nucléaire. Dans la **Figure VII-5**, nous avons reporté les sites de fixation connus des snRNP U1 et U11 et des protéines SR sur l'élément NRS du virus RSV sur un schéma de la structure 2D où les positions de clivage protégées en extrait nucléaire sont indiquées. Nos données de protections sont, comme on peut le voir, compatibles avec les données de la littérature obtenues par d'autres approches puisque certaines des protections observées en extrait nucléaire peuvent s'expliquer par la fixation de ces facteurs. En ce qui concerne les autres régions protégées et en particulier la tige-boucle SLS2, les protections observées, ne permettent pas de savoir quel(s) composant(s) de l'extrait nucléaire est(sont) impliqué(s). C'est pourquoi nous avons aussi ensuite réalisé les mêmes expériences sur des complexes formés avec des protéines SR et de la protéine hnRNP A1 recombinantes et développé des approches en vue de purifier les complexes ribonucléoprotéiques formés en extrait nucléaire et d'identifier leur contenu.

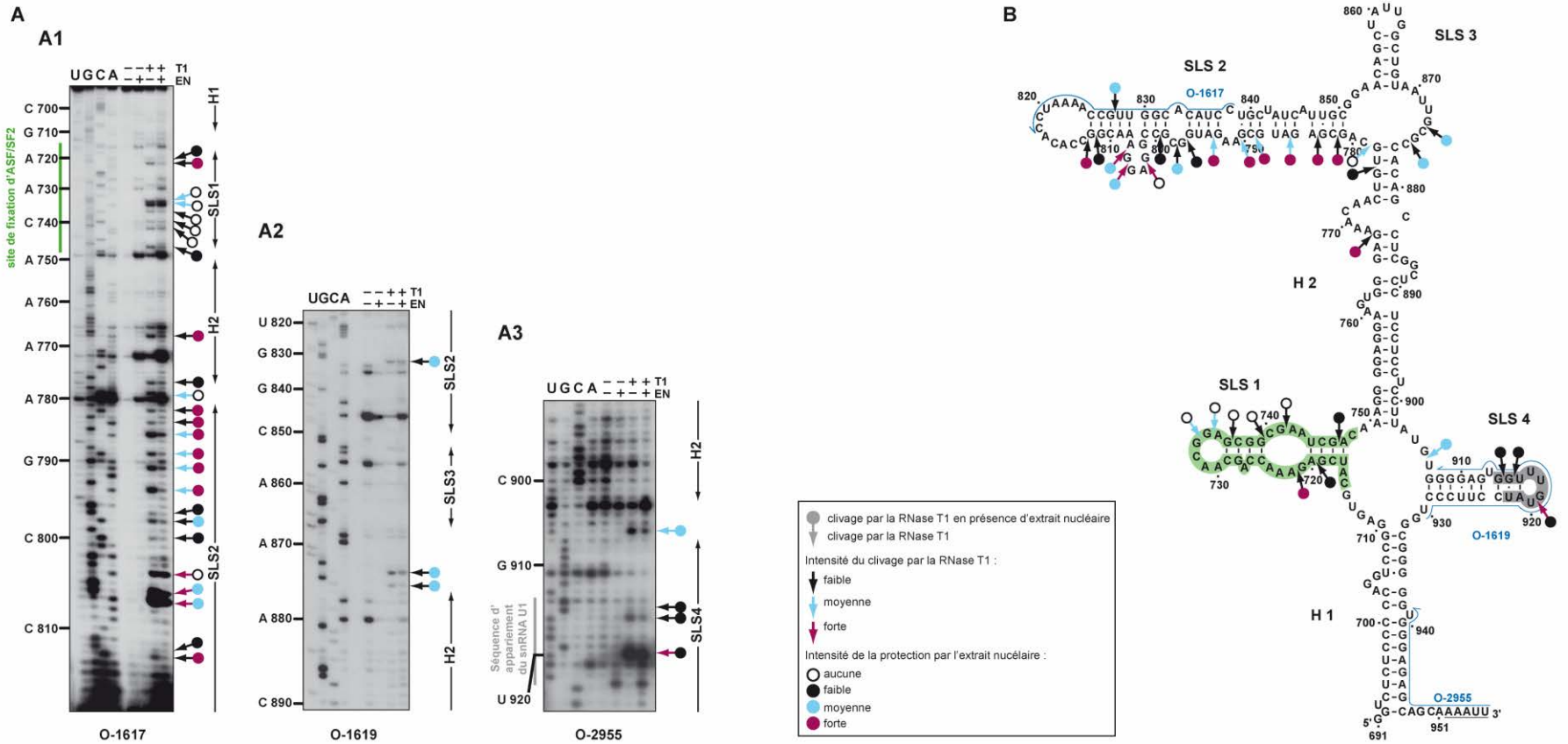



**Figure VII-2 : Analyse des protections contre l'action de la RNase T1 générées sur l'élément NRS par les composants d'un extrait nucléaire de cellules HeLa.**
L'ARN a été soumis à une hydrolyse ménagée par la RNase T1 (+)(10 U/µg d'ARN) en présence (+) ou en absence (-) d'extrait nucléaire de cellules HeLa (EN), selon les conditions décrites dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension d'amorce par la transcriptase inverse avec les oligonucléotides 1617 (A1), 1619 (A2) et 2955 (A3) radio-marqués en 5'. Les pistes U, G, C, A correspondent au séquençage de l'ARN. Les positions des nucléotides, ainsi que le site de fixation de la protéine ASF/SF2 (McNally *et al*., 1996) et la séquence d'appariement du snRNA U1 (Hibbert *et al*., 1999) sont indiqués à gauche des autoradiogrammes. Les structures secondaires (tige boucle : SLS, hélice : H) délimitées par des doubles flèches et les coupures par la RNase T1, ainsi que les protections des nucléotides représentées selon la légende précisée dans l'encart, sont reportées à droite des autoradiogrammes.
**B :** Représentation des protections contre l'action de la RNase T1 générées par les composants de l'extrait nucléaire sur le modèle de structure secondaire de l'ARN NRS WT du virus RSV. Les coupures par la RNase T1 et les protections des nucléotides sont reportées selon la légende décrite en (A). La numérotation des nucléotides, les hélices (H) et les structures tige boucle (SLS) sont mentionnées. Les nucléotides non naturels ajoutés à l'extrémité 3' de l'ARN sont soulignés. Les oligonucléotides sont représentés par des flèches bleues et les séquences de fixation de la protéine ASF/SF2 (McNally *et al*., 1996) et d'appariement du snRNA U1 (Hibbert *et al*., 1999) sont respectivement colorées en vert et gris.

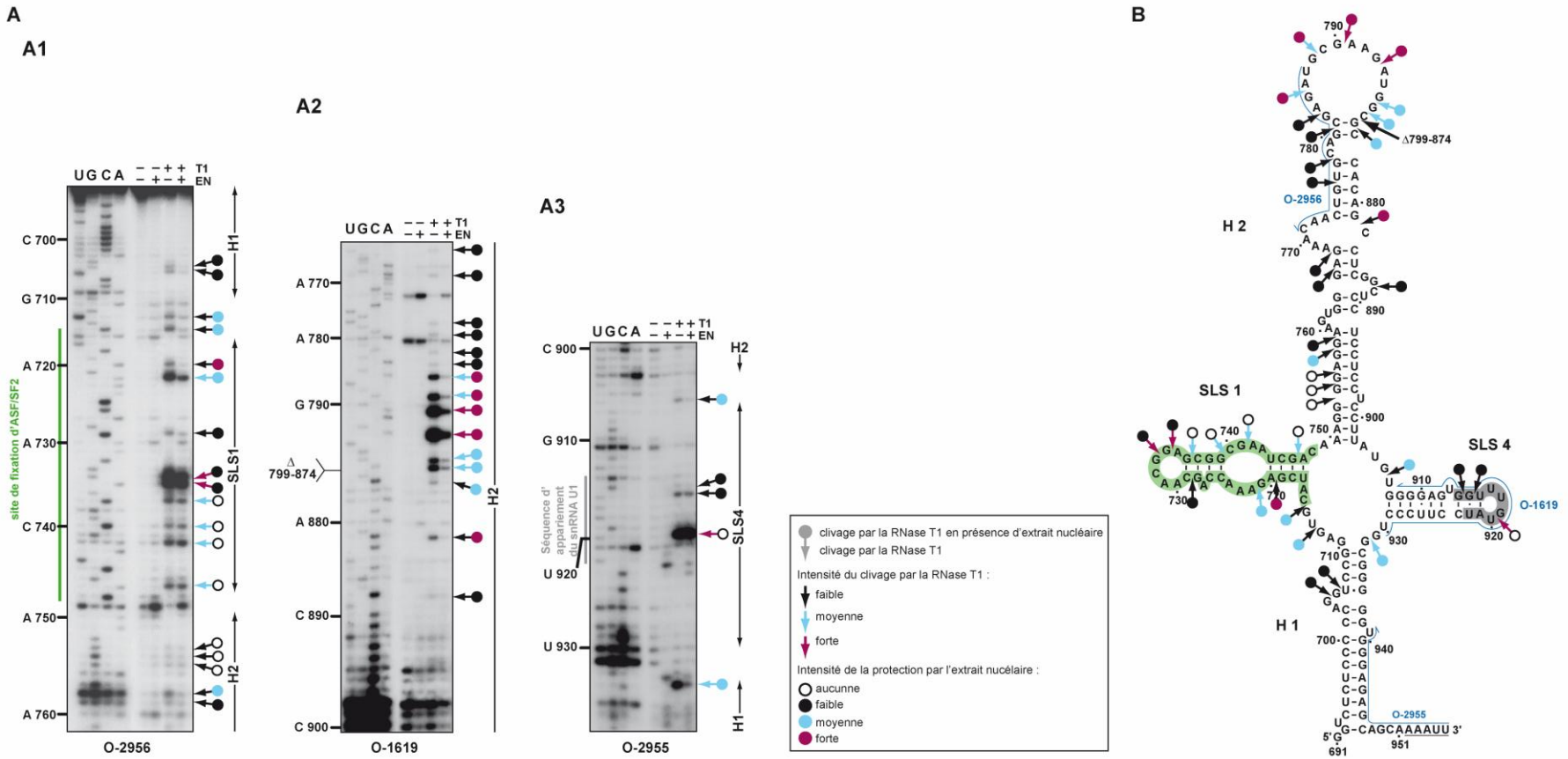



**Figure VII-3 : Analyse des protections contre l'action de la RNase T1 générées sur l'élément NRS Δ76 par les composants d'un extrait nucléaire de cellules HeLa.**
L'ARN a été soumis à une hydrolyse ménagée par la RNase T1 (+)(10 U/µg d'ARN) en présence (+) ou en absence (-) d'extrait nucléaire de cellules HeLa (EN), selon les conditions décrites dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension d'amorce par la transcriptase inverse avec les oligonucléotides 2956 (A1), 1619 (A2) et 2955 (A3) radio-marqués en 5'. Les pistes U, G, C, A correspondent au séquençage de l'ARN. Les positions des nucléotides, ainsi que le site de fixation de la protéine ASF/SF2 (McNally *et al*., 1996) et la séquence d'appariement du snRNA U1 (Hibbert *et al*., 1999) sont indiqués à gauche des autoradiogrammes. Les structures secondaires (tige boucle : SLS, hélice : H) délimitées par des doubles flèches et les coupures par la RNase T1, ainsi que les protections des nucléotides représentées selon la légende précisée dans l'encart, sont reportées à droite des autoradiogrammes.
**B :** Représentation des protections contre l'action de la RNase T1 générées par les composants de l'extrait nucléaire sur le modèle de structure secondaire de l'ARN NRS Δ76 du virus RSV. Les coupures par la RNase T1 et les protections des nucléotides sont reportées selon la légende décrite en (A). La numérotation des nucléotides, les hélices (H) et les structures tige boucle (SLS) sont mentionnées. Les nucléotides non naturels ajoutés à l'extrémité 3' de l'ARN sont soulignés. Les oligonucléotides sont représentés par des flèches bleues et les séquences de fixation de la protéine ASF/SF2 (McNally *et al*., 1996) et d'appariement du snRNA U1 (Hibbert *et al*., 1999) sont respectivement colorées en vert et gris.

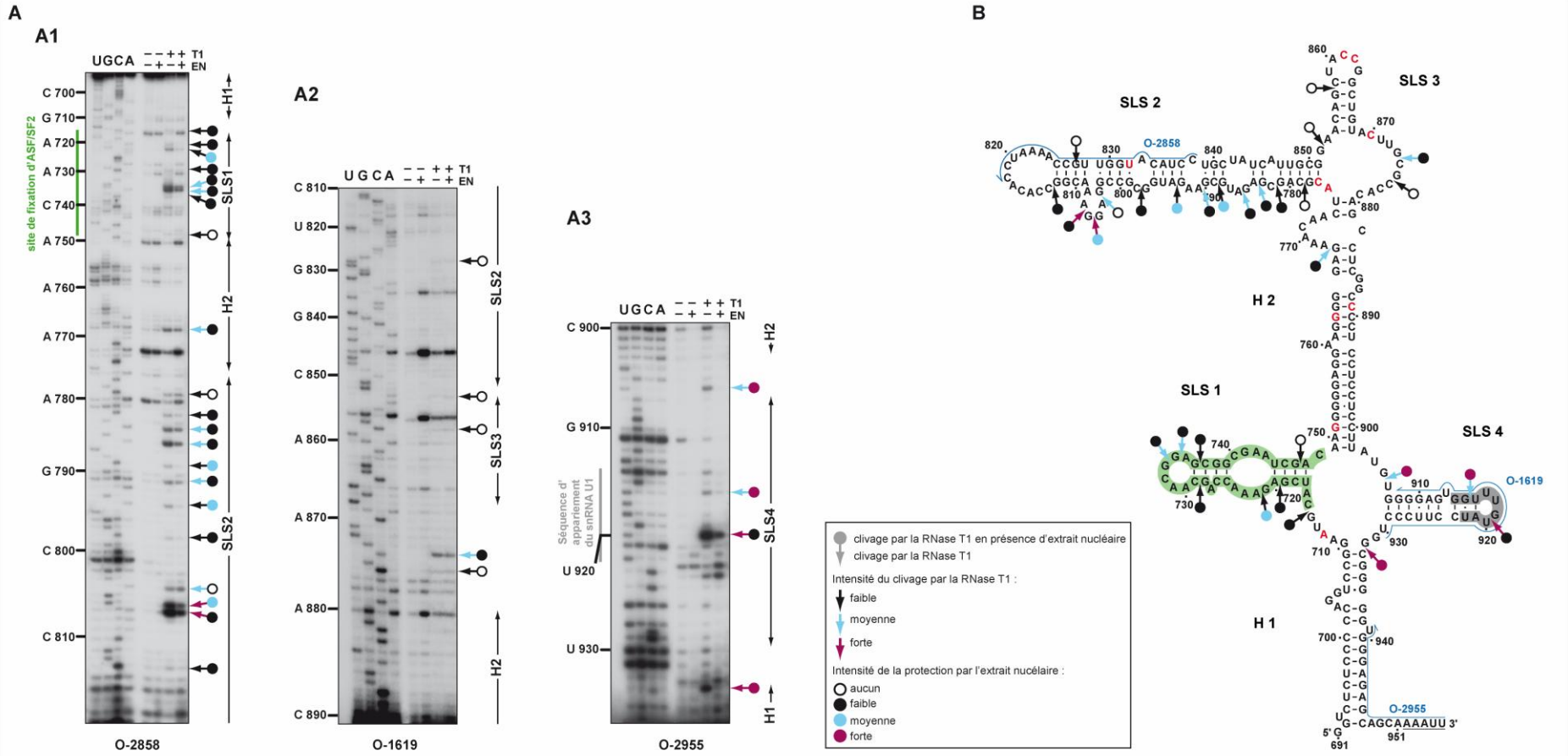



**Figure VII-4 : Analyse des protections contre l'action de la RNase T1 générées sur l'élément NRS LR9 du virus ALV par les composants d'un extrait nucléaire de cellules HeLa.**
L'ARN a été soumis à une hydrolyse ménagée par la RNase T1 (+)(10 U/µg d'ARN) en présence (+) ou en absence (-) d'extrait nucléaire de cellules HeLa (EN), selon les conditions décrites dans la partie Matériel et Méthodes. Les positions des coupures ont été identifiées par extension d'amorce par la transcriptase inverse avec les oligonucléotides 2858 (A1), 1619 (A2) et 2955 (A3) radio-marqués en 5'. Les pistes U, G, C, A correspondent au séquençage de l'ARN. Les positions des nucléotides, ainsi que le site de fixation de la protéine ASF/SF2 (McNally *et al*., 1996) et la séquence d'appariement du snRNA U1 (Hibbert *et al*., 1999) sont indiqués à gauche des autoradiogrammes. Les structures secondaires (tige boucle : SLS, hélice : H) délimitées par des doubles flèches et les coupures par la RNase T1, ainsi que les protections des nucléotides représentées selon la légende précisée dans l'encart, sont reportées à droite des autoradiogrammes.
**B :** Représentation des protections contre l'action de la RNase T1 générées par les composants de l'extrait nucléaire sur le modèle de structure secondaire de l'ARN NRS LR9 du virus ALV. Les coupures par la RNase T1 et les protections des nucléotides sont reportées selon la légende décrite en (A). La numérotation des nucléotides, les hélices (H) et les structures tige boucle (SLS) sont mentionnées. Les nucléotides non naturels ajoutés à l'extrémité 3' de l'ARN sont soulignés et les substitutions de bases par rapport à la séquence NRS WT du virus RSV sont colorés en rouge. Les oligonucléotides sont représentés par des flèches bleues et les séquences de fixation de la protéine ASF/SF2 (McNally *et al*., 1996) et d'appariement du snRNA U1 (Hibbert *et al*., 1999) sont respectivement colorées en vert et gris.

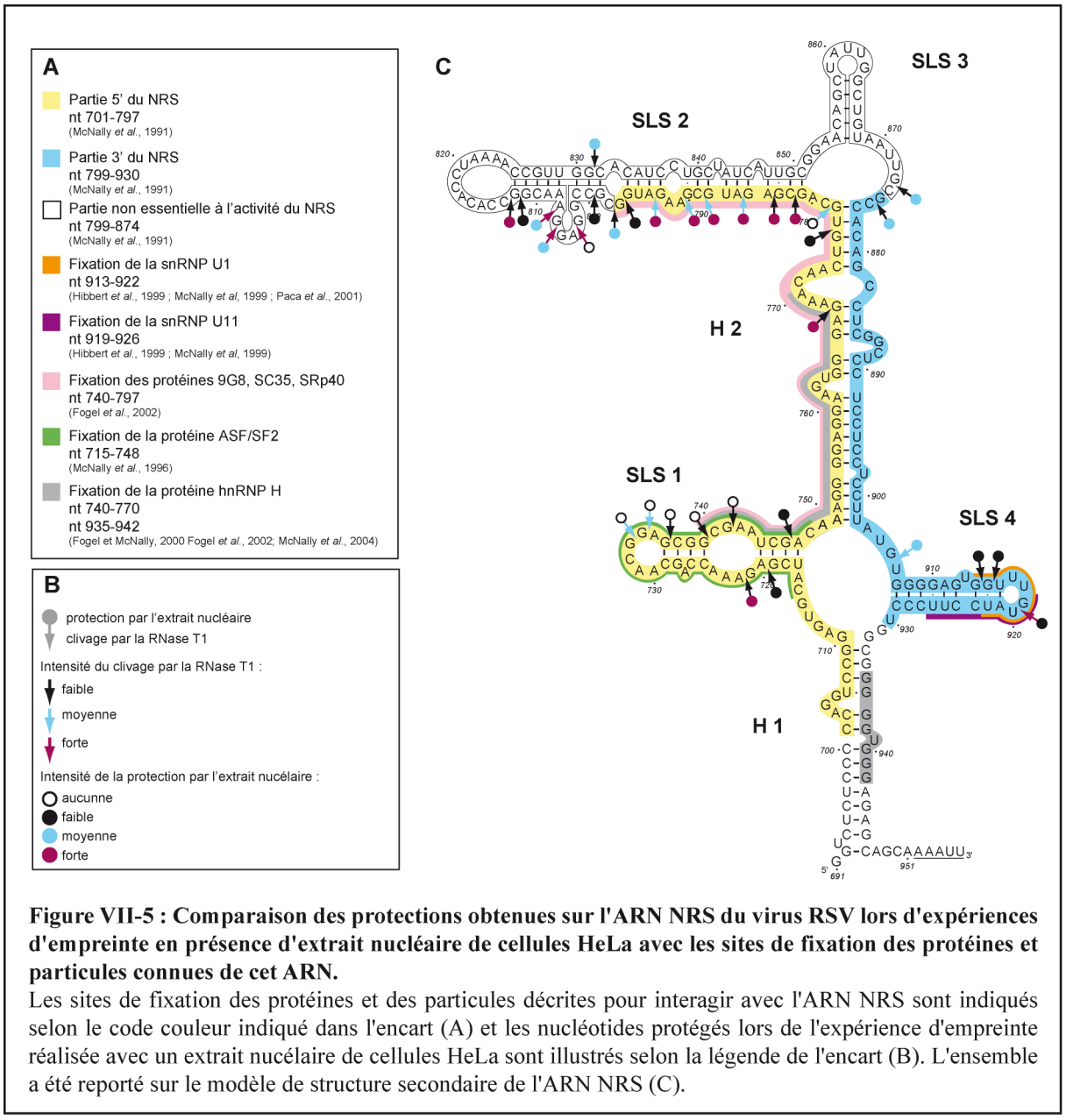



**Figure VII-5 : Comparaison des protections obtenues sur l'ARN NRS du virus RSV lors d'expériences d'empreinte en présence d'extrait nucléaire de cellules HeLa avec les sites de fixation des protéines et particules connues de cet ARN.**

Les sites de fixation des protéines et des particules décrites pour interagir avec l'ARN NRS sont indiqués selon le code couleur indiqué dans l'encart (A) et les nucléotides protégés lors de l'expérience d'empreinte réalisée avec un extrait nucélaire de cellules HeLa sont illustrés selon la légende de l'encart (B). L'ensemble a été reporté sur le modèle de structure secondaire de l'ARN NRS (C).

# Chapitre VIII : recherche d'une méthode de chromatographie d'affinité adaptée à la purification des complexes ribonucléoprotéiques formés sur l'élément NRS

Afin d'identifier les partenaires protéiques de l'élément NRS, nous avons d'abord testé les performances de différentes méthodes d'affinité pour purifier les complexes ribonucléoprotéiques formés en extrait nucléaire avec l'élément NRS du virus RSV.

Notre objectif était d'analyser l'ensemble des composants d'un extrait nucléaire capables de se lier à cet élément. Pour cela, nous devions développer une méthode de purification efficace permettant d'isoler les complexes ribonucléoprotéiques formés avec l'ARN NRS en conditions non dénaturantes, de fractionner les protéines contenues dans ces complexes par électrophorèse en gel SDS-PAGE et de les identifier par spectrométrie de masse.

## I. Présentation des techniques de purification des particules ribonucléoprotéiques par chromatographie d'affinité

Lorsque nous avons débuté ce travail, trois méthodes étaient disponibles pour purifier les complexes ribonucléoprotéiques.

La première méthode reposait sur l'incorporation dans l'ARN d'intérêt au cours de la transcription *in vitro* de rNTP modifiés portant des groupements chimiques ayant une forte affinité pour un ligand qui, lui, était fixé covalemment à une résine. L'ARN était ensuite incubé avec un extrait nucléaire et les complexes ribonucléoprotéiques formés étaient spécifiquement retenus sur la résine. Un des couples groupement chimique/ligand parmi les plus utilisés dans ce type d'expérience est celui composé de la biotine et de la streptavidine. Ce couple a notamment permis de purifier des complexes spliceosomaux (Grabowski et Sharp, 1986). Nous n'avons pas retenu cette méthode car les groupements chimiques comme la biotine sont stériquement encombrants et leur incorporation au sein de l'ARN risquait de perturber la structuration de l'ARN et la fixation des protéines. Par ailleurs, l'interaction de haute affinité entre les deux partenaires (Kd = $10^{-15}$M) empêchait d'éluer les complexes ribonucléoprotéiques dans des conditions natives.

La seconde méthode employée consistait à hybrider sur une région simple brin de l'ARN d'intérêt un oligonucléotide portant à son extrémité 5' un groupement biotine (Blencowe *et al.*,

1989). Les complexes ribonucléoprotéiques formés sur l'ARN étaient ensuite purifiés selon le même principe que celui de la première technique. Nous n'avons pas envisagé d'utiliser cette technique car d'après le modèle de structure secondaire de l'élément NRS que nous avions établi expérimentalement, il n'existe pas de séquence simple brin suffisamment longue pour y hybrider un oligonucléotide.

La troisième méthode consistait à fusionner co-transcriptionnellement la séquence de l'ARN d'intérêt à une séquence d'ARN ayant une forte affinité pour un ligand. Suivant sa nature, le ligand utilisé peut être immobilisé sur une résine par une liaison covalente ou non covalente. Suite à l'incubation de cet ARN de fusion avec un extrait nucléaire, les complexes ribonucléoprotéiques formés sont retenus spécifiquement sur la résine grâce à l'interaction entre la séquence ARN ajoutée et son ligand. Cette technique présente l'avantage de permettre d'éluer les complexes dans des conditions natives, grâce à des élutions par compétition.

Nous avons alors décidé d'utiliser cette dernière approche pour purifier les complexes assemblés sur l'élément NRS. Pour ceci, différents couples ARN/ligand ont été envisagés sur la base de trois exigences (**Figure VIII-1**).

- Premièrement, la séquence ajoutée à l'ARN d'intérêt devait être courte et devait adopter une structure secondaire stable se formant indépendamment du reste de l'ARN. Comme nous avions préalablement caractérisé la structure secondaire de l'élément NRS, nous avons testé à l'aide du logiciel Mfold si cette structure n'était pas altérée au sein de l'ARN fusionné.
- Deuxièmement, la séquence additionnelle devait avoir une forte affinité pour son ligand immobilisé sur une résine et cette interaction devait être maintenue dans les conditions expérimentales de la chromatographie d'affinité.
- Troisièmement, la rétention de l'ARN de fusion sur la résine devait être réversible afin de permettre l'élution des complexes ribonucléoprotéiques dans des conditions préservant leur intégrité.

Suivant la nature des couples de séquence ARN/ligand utilisés, deux cas peuvent être envisagés. Certains ligands peuvent être immobilisés directement sur la résine par une liaison covalente et donc irréversible (chromatographie basée sur l'emploi de la tobramycine). Dans ce cas, c'est l'interaction entre la séquence aptamérique de l'ARN et le ligand fixé sur la résine qui doit être rompue pour permettre l'élution des complexes ribonucléoprotéiques.

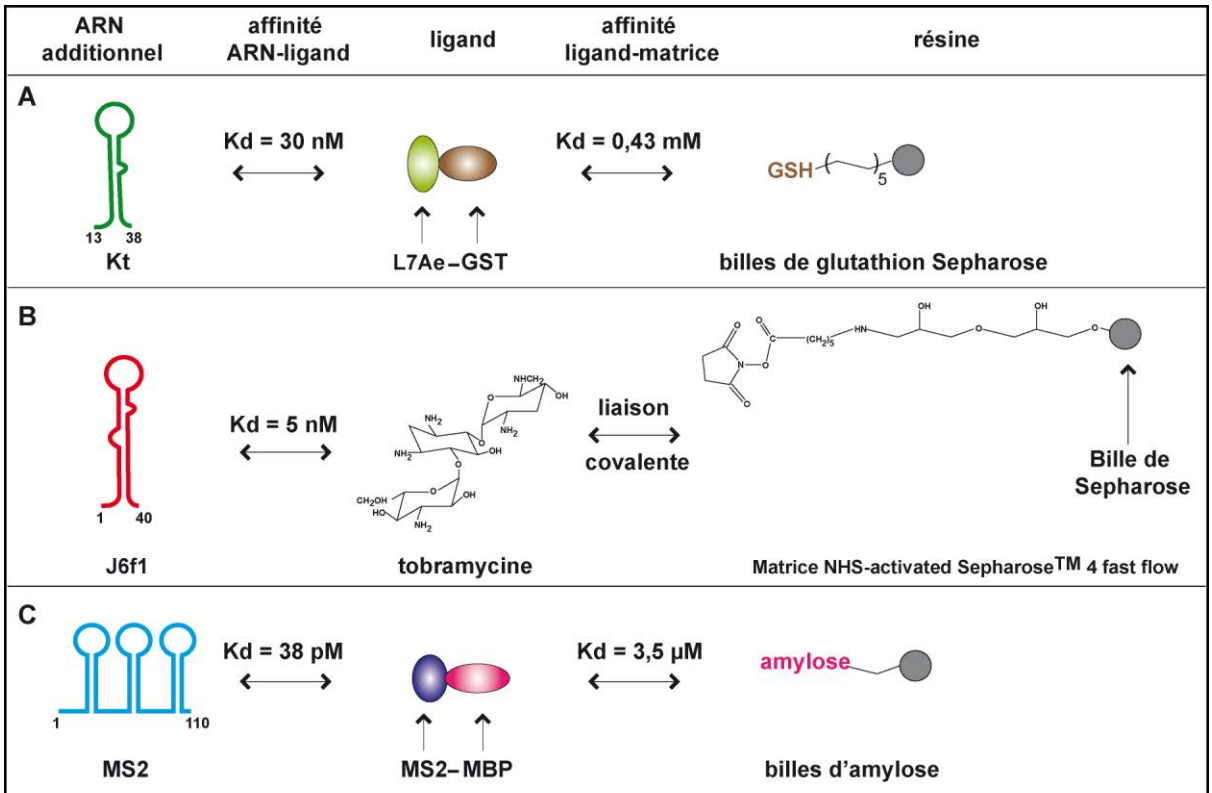



**Figure VIII-1 : Représentation des différentes séquences pouvant être fusionnées à l'ARN d'intérêt, leurs ligands et les résines associées dont nous avons testé l'efficacité pour la purification des complexes ribonucléoprotéiques formés avec l'ARN NRS.**
**A :** Partenaires utilisés dans la chromatographie d'affinité basée sur l'emploi de la protéine L7Ae.
**B :** Partenaires utilisés dans la chromatographie d'affinité basée sur l'emploi de la tobramycine.
**C :** Partenaires utilisés dans la chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2.
Les noms des séquences ajoutées, de leur ligand, ainsi que de la matrice utilisée sont indiqués. La structure 2D des régions additionnelles représentée schématiquement et les positions des nucléotides à leur extrémité sont indiquées. Les interactions entre les différents partenaires sont représentées par des flèches doubles. GSH : glutathion, MBP : Maltose Binding Protein.

D'autres ligands peuvent être immobilisés sur la résine par une liaison non covalente comme des interactions par affinité (chromatographies basées sur l'emploi de la protéine L7Ae et MS2). Le plus souvent, ces ligands sont des protéines de fusion qui comportent une partie protéique ayant une affinité pour la séquence ARN aptamérique et une autre ayant une affinité pour la résine. Dans ce cas, le ligand se dissocie de la résine lors de l'élution des complexes. L'interaction ligand-résine est souvent moins forte (Kd de l'ordre du µM) que celle entre une séquence aptamérique et son ligand (Kd de l'ordre du nM). Les Kd des couples ARN/ligand et ligand-résine que nous avons utilisés sont suffisamment élevés pour envisager de purifier de complexes ribonucléoprotéiques par des chromatographies d'affinité.

## II. Essai de développement d'une chromatographie d'affinité basée sur le motif ARN en K-loop et la protéine L7AE

Cette approche reposait sur l'identification au laboratoire d'un motif ARN fréquent chez les archae qui fixe avec une grande affinité la protéine d'archae L7Ae. L'idée était de l'ajouter à l'ARN d'intérêt et de retenir l'ARN de fusion avec une protéine de fusion L7Ae-GST, elle même retenue sur des billes glutathion-Sepharose. Cette méthode s'est avérée manquer de spécificité et sa mise au point ne sera pas discutée.

## III. Chromatographie d'affinité basée sur l'utilisation de la tobramycine

Cette chromatographie d'affinité avait été mise au point par K. Hartmuth de l'équipe du Pr. R. Lührmann (Max Planck Institute, Göettingen) (Hartmuth *et al.*, 2004). J'ai effectué un stage dans le laboratoire du Pr. Lührmann pour apprendre à utiliser cette technique qui repose sur l'ajout à l'ARN d'intérêt d'un segment d'ARN ayant une forte affinité pour la tobramycine qui est un antibiotique de la famille des aminoglycosides. Lorsque l'équipe de R. Lührmann utilisait cette approche pour purifier des complexes spliceosomaux, l'étape de chromatographie d'affinité était suivie d'une étape supplémentaire de purification, soit une ultracentrifugation en gradient de glycérol, soit une immunosélection avec un anticorps dirigé contre une protéine présente dans le complexe spliceosomal à purifier (Deckert *et al.*, 2006 ; Hartmuth *et al.*, 2002).

Notre objectif était de purifier des complexes ribonucléoprotéiques que nous supposions beaucoup plus hétérogènes que les complexes spliceosomaux. Nous avons donc essayé d'utiliser des conditions assez stringentes pour obtenir une spécificité suffisante en une seule étape de purification.

### III.1. Les différents acteurs intervenant dans la chromatographie d'affinité basée sur l'emploi de la tobramycine

La séquence aptamérique J6f1, sélectionnée pour son affinité pour la tobramycine par l'approche SELEX (pour Systematic Evolution of Ligands by Exponential Enrichment) (Wang et Rando, 1995), possède une forte affinité pour la tobramycine (Kd = 5 nM) (Hamasaki *et al.*, 1998). Cet aptamère adopte une structure secondaire en tige-boucle très stable (**Figure VIII-2**) (Cho *et al.*, 1998). Les nucléotides importants pour l'interaction avec la tobramycine se situent dans la partie apicale de l'hélice ainsi que dans la boucle terminale.

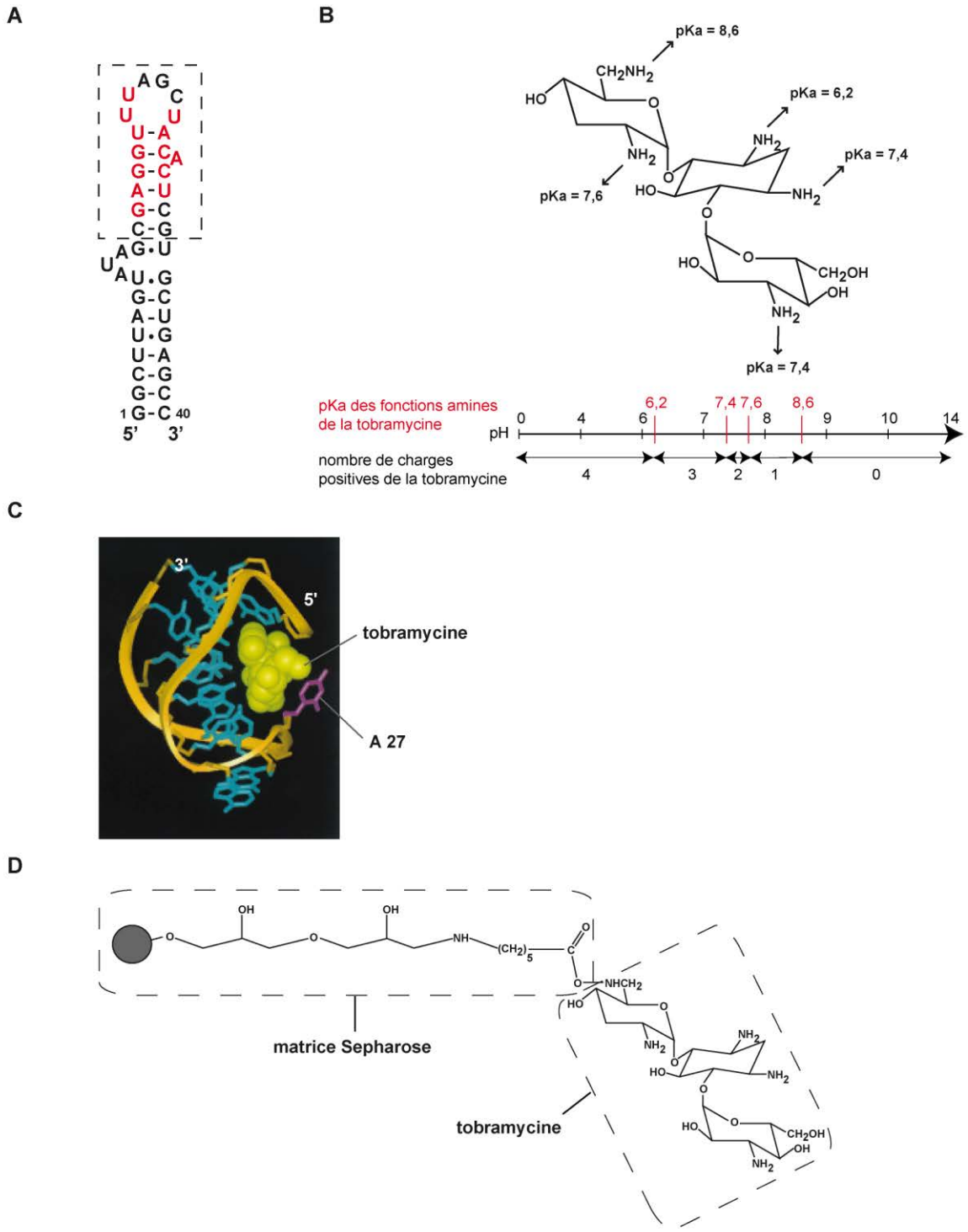



**Figure VIII-2 : Description des différents acteurs intervenant dans la chromatographie d'affinité basée sur l'emploi de la tobramycine.**

**A :** Structure 2D de l'aptamère J6f1 (Hamasaki *et al*., 1998, Cho *et al*., 1998). L'aptamère J6f1 forme une tige boucle stable comportant une boucle latérale. Les nucléotides qui sont importants pour la reconnaissance et l'affinité pour la tobramycine sont colorés en rouge. La partie de cet ARN dont la structure 3D a été établie par RMN (C) est encadrée.

**B :** Structure de la tobramycine. La tobramycine est un aminoglycoside comportant 5 fonctions amines primaires. Les différentes fonctions amines, leur pKa et leur ionisation en fonction du pH sont indiqués.

**C :** Structure du complexe aptamère/tobramycine établie par RMN (Patel et Suri, 2000). Le squelette phosphodiester de l'ARN est représenté par un ruban orange, les bases sont colorées en bleu à l'exception de la base en bulge A27 qui est colorée en violet. La tobramycine(ligand) est représentée en jaune.

**D :** Matrice Tobra-Sepharose utilisée pour la chromatographie d'affinité.

La tobramycine (**Figure VIII-2**) possède cinq groupements amines primaires dont les pKa s'échelonnent entre 6,2 et 8,6. Ces groupements amines lui permettent d'être fixée de manière covalente à la matrice NHS-Sepharose pour donner la matrice Tobra-Sepharose (**Figure X, cf Matériel et Méthodes**). Cette fixation a lieu au cours de la réaction de couplage de matrice qui est réalisée à un pH de 8,3 (**cf Matériel et Méthodes**). A ce pH, seul le groupement $NH_2$ possédant un pKa de 8,6 est ionisé et donc davantage susceptible de réagir avec la matrice que les autres fonctions amines dont le pKa est inférieur (**Figure VIII-2**). La majorité des molécules de tobramycine seraient donc covalemment liées à la matrice par un seule amine et ne possèderait donc plus que 4 groupements amines primaires libres. Compte tenu des capacités d'ionisation de ces groupements, lors de la chromatographie d'affinité, il est préférable de travailler à un pH de 9,1 pour que la tobramycine ne soit pas chargée positivement. L'absence de ces charges limite les interactions électrostatiques aspécifiques avec les ARN et les protéines contenus dans l'extrait nucléaire.

### III.2. Réalisation des constructions génétiques

Pour réaliser la chromatographie d'affinité, j'ai construit les plasmides pPAPTO et pAPTO NRS. Après linéarisation par l'enzyme de restriction *Sma*I, le plasmide pPAPTO permet de transcrire *in vitro* l'ARN aptamérique J6f1 que nous appellerons ARN Tob (**Figure VIII-3**). Le plasmide pAPTO NRS, après linéarisation par l'enzyme de restriction l'enzyme *Sma*I ou *Bam*HI permet de transcrire *in vitro* l'ARN l'ARN NRS-Tob, où le NRS est fusionné à l'aptamère Tob, ou l'ARN NRS seul, respectivement (**Figure VIII-3**).

### III.3. Evaluation de l'efficacité de rétention de l'ARN NRS-Tob sur la matrice Tobra-Sepharose

Pour préparer la matrice Tobra-Sepharose, 2 ml de matrice en suspension ont été incubés avec 5 µmol de tobramycine lors de la réaction de couplage (**cf Matériel et Méthodes**) (Hartmuth *et al.*, 2004). Si l'on considère que la réaction de fixation est efficace à 100 %, cela signifiait que dans ces conditions, nous avions 2,5 µmol de tobramycine fixée/ml de matrice. Nous envisagions de réaliser les chromatographies d'affinité avec 25 µl de matrice Tobra-Sepharose et 60 pmol d'ARN NRS-Tob. Sachant que la tobramycine interagit avec l'ARN Tob ou l'ARN NRS-Tob de manière stoechiométrique, nous devions pouvoir théoriquement

immobiliser $6{,}25.10^{-8}$ mol d'ARN NRS-Tob sur la matrice, soit une quantité d'ARN largement supérieure à celle que nous souhaitions utiliser. Cependant, le rendement de la réaction de couplage ne peut pas être mesuré directement. Par conséquent, nous avons dû, dans un premier temps, tester la capacité de rétention de l'ARN NRS-Tob sur la matrice Tobra-Sepharose. Ces tests ont été réalisés dans les mêmes conditions que celles que nous voulions utiliser ultérieurement pour purifier les complexes ribonucléoprotéiques. En parallèle, nous avons réalisé un autre test avec l'ARN NRS seul qui a servi de témoin négatif.

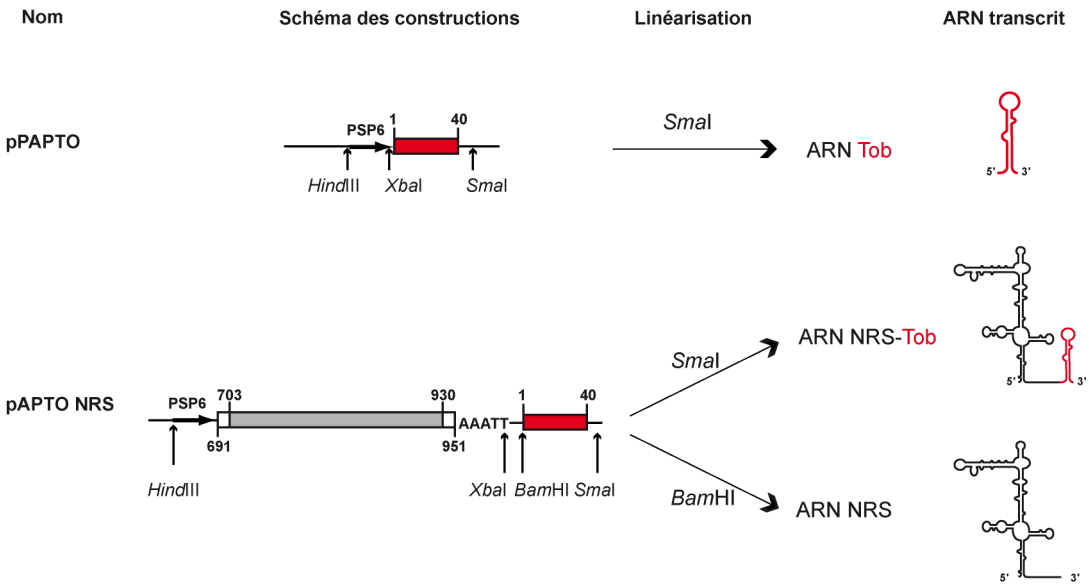



**Figure VIII-3 : Constructions réalisées pour produire les ARN utilisés dans les expériences de chromatographie d'affinité basée sur l'emploi de la tobramycine.**
La séquence ADN J6f1 est représentée par des rectangles rouges, l'ADN de l'élément NRS ainsi que ses séquences bordantes sont respectivement représentés par des rectangles gris et blancs et les nucléotides supplémentaires AAATT ajoutés à l'extrémité de l'élément NRS sont mentionnés. Les positions des nucléotides de l'élément NRS au sein de l'ADNc proviral sont indiquées. La flèche horizontale représente le promoteur transcriptionnel SP6 et les flèches verticales représentent les sites de restriction utilisés pour les clonages et pour la linéarisation des matrices avant la transcription. Le nom des différents transcrits produits à partir des matrices linéarisées sont indiqués et les ARN produits sont représentés avec leur structure secondaire. La séquence ARN de l'élément NRS est représentée en noire et celle de l'aptamère en rouge.

Le test de capacité de la matrice à fixer l'ARN comprend plusieurs étapes : I) lavage, II) saturation, III) Fixation de l'ARN, IV) lavage, V) Elution (**cf Matériel et Méthodes et Figure VIII-4**).

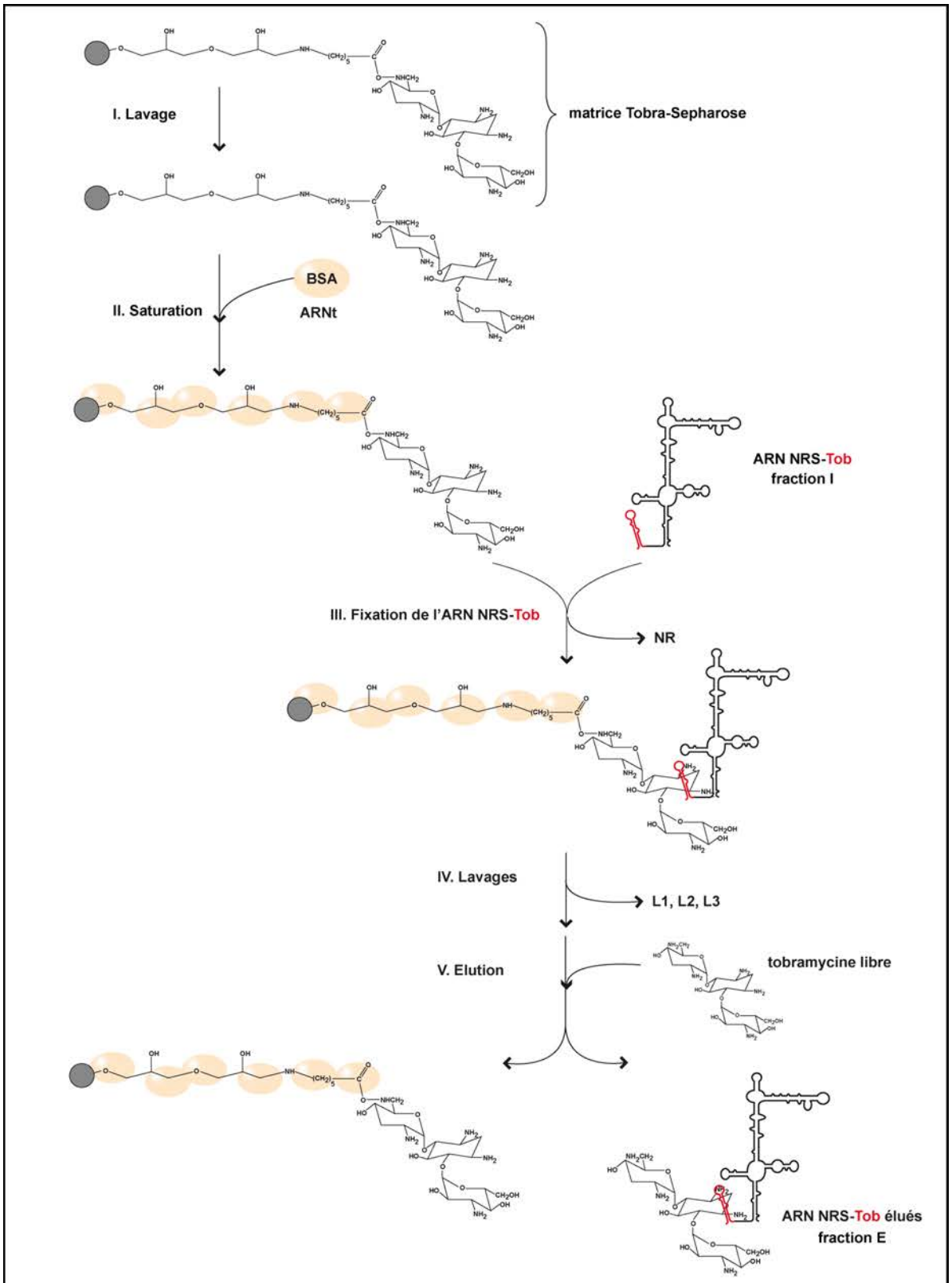



**Figure VIII-4 : Test de fixation de l'ARN NRS-Tob sur la matrice Tobra-Sepharose.**
Cette expérience a servi à tester la fonctionnalité de la matrice activée. Elle comprend 5 étapes : I) lavage, II) saturation, III) fixation de l'ARN NRS-Tob, IV) lavages et V) Elution. Le nom des différentes fractions collectées à chaque étape est mentionné : I, input ; NR, non retenu ; L1, L2, L3, lavages 1, 2 et 3 ; E, éluat. L'ARN NRS-Tob est représenté par sa structure secondaire, la séquence de l'élément NRS et celle de l'aptamère Tob sont respectivement en noir et en rouge. La saturation de la matrice par les molécules de BSA est schématisée par des sphères roses.
Le même protocole a été utilisé pour l'ARN NRS seul.

Sachant que les molécules de tobramycine fixées sur la matrice sont théoriquement présentes en excès par rapport aux molécules d'ARN qui y sont ajoutées, il est nécessaire de se placer dès le départ dans des conditions qui limitent au maximum les interactions aspécifiques de la tobramycine avec les ARN. L'étape de lavage sert à faire passer la matrice d'un tampon de pH 8,0 dans lequel elle est conservée à un pH de 9,1 de façon à limiter l'ionisation des groupements amines de la tobramycine. L'étape de saturation sert à masquer les sites d'interactions aspécifiques potentiels de la matrice avec de la BSA et des ARNt.

L'ARN (fraction I, Input) composé d'un mélange d'ARN non radioactif (60 pmol) et d'une quantité très faible d'ARN uniformément marqué qui servait de traceur (20 fmol) a ensuite été ajouté à la matrice (Etape III) (**Figure VIII-4**). Après 1h30 d'incubation, les ARN non retenus ont été éliminés par une centrifugation (fraction NR). Puis la matrice a été lavée trois fois de suite et les fractions contenant les ARN non retenus sur la matrice ont été colletés après chaque lavage (Etape IV, fraction L1, L2, L3). Lors de la dernière étape, les ARN retenus sur la matrice ont été élués (Etape V, fraction E) par incubation avec une solution contenant 5 mM de tobramycine libre.

La quantité d'ARN contenue dans chaque fraction a été déterminée en mesurant leur radioactivité au compteur à scintillation en Cerenkov. La radioactivité de l'ARN de départ ajouté sur la matrice (fraction Input) a été prise comme référence (100 %) et la radioactivité des autres fractions a été calculée en pourcentage par rapport à cette référence.

Lorsque les tests ont été réalisés selon les conditions établies par K. Hartmuth (Hartmuth *et al.*, 2004), les ARN NRS-Tob n'étaient ni efficacement immobilisés, ni efficacement élués de la matrice (**Figure VIII-5, B1**). Après les 3 lavages, seulement 30 % des ARN de départ sont toujours immobilisés sur la matrice et moins de la moitié de ces ARN (12 % de la radioactivité de départ) est éluée. En revanche, dans ces mêmes conditions, 98,4 % des ARN NRS ne sont pas retenus sur la matrice après les 3 lavages (**Figure VIII-5, B2**). Ce faible taux de rétention de l'ARN NRS indique que la matrice a été suffisamment bien saturée pour ne retenir spécifiquement que les ARN fusionnés à l'ARN Tob. Selon les critères de K. Hartmuth, les conditions utilisées pour saturer la matrice étaient donc satisfaisantes car moins de 2 % des ARN NRS non fusionnés à l'ARN Tob étaient retenus sur la matrice. En revanche, les conditions de rétention de l'ARN NRS-Tob n'étaient pas satisfaisantes car, toujours d'après les critères de K. Hartmuth, 50-70 % des ARN de départ auraient dû être immobilisés sur la matrice après les 3 lavages et plus de 80 % ces ARN auraient dû être élués.

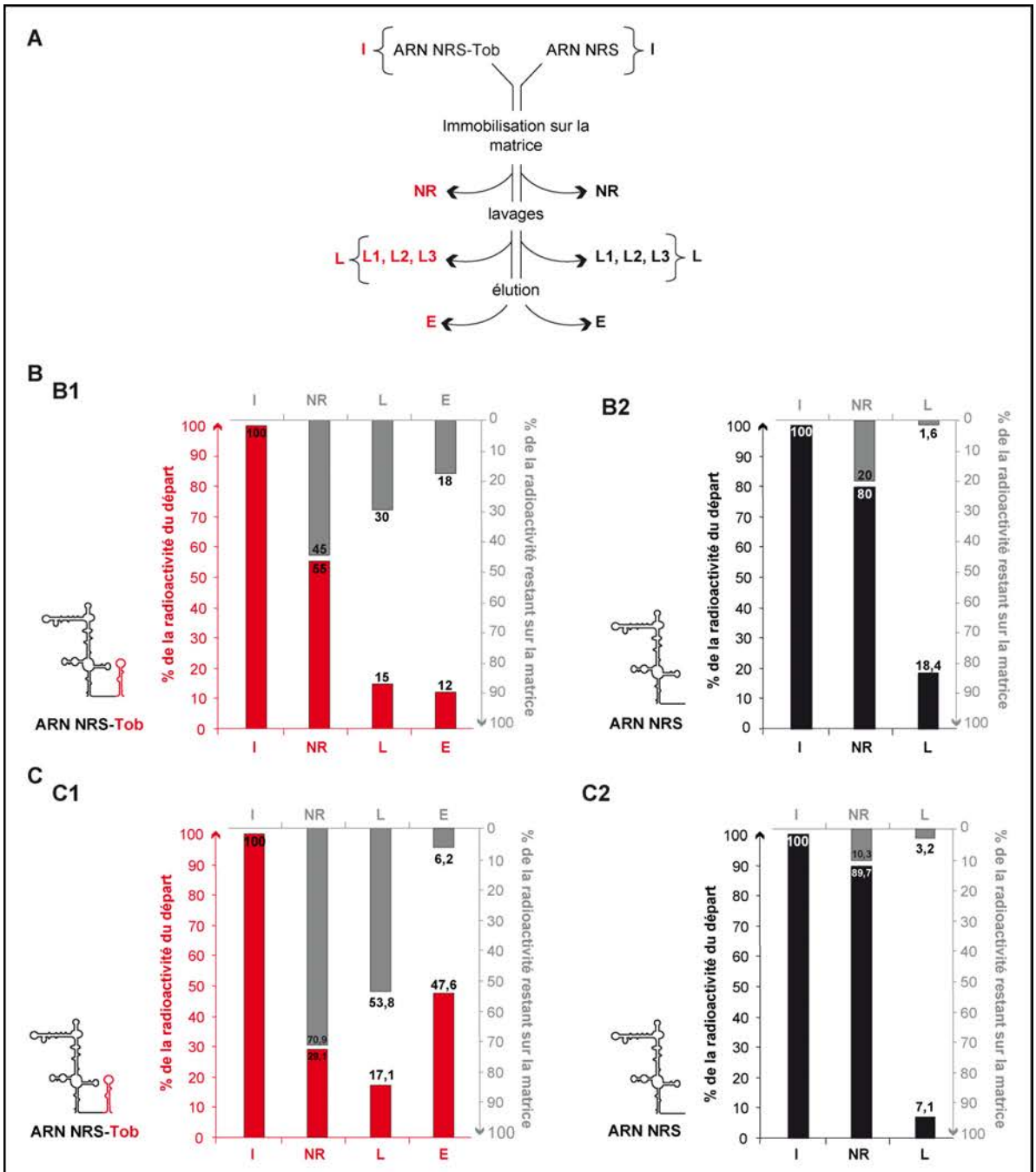



**Figure VIII-5 : Evaluation et amélioration des étapes de rétention et d'élution de l'ARN NRS-Tob sur la matrice Tobra-Sepharose.**

**A :** Récapitulatif des différentes étapes des tests de rétention des ARN sur la matrice Tobra-Sepharose. Les tests ont été réalisés avec un mélange d'ARN non radioactifs et d'ARN uniformément marqués. La quantité d'ARN de chaque fraction a été évaluée en mesurant leur radioactivité au compteur à scintillation en Cherenkov. La fraction I correspond à l'Input (100 % de l'ARN au départ), NR, à l'ARN non retenu sur la matrice, L, à l'ensemble des 3 lavages de la matrice et E, à l'éluat. Les histogrammes représentent la radioactivité contenue dans chaque fraction collectée (en rouge pour l'expérience réalisée avec l'ARN NRS-Tob, et en noir pour l'expérience réalisée avec l'ARN NRS) et la radioactivité restant sur la matrice après chaque étape (en gris). Les valeurs sont exprimées en pourcentage par rapport à la radioactivité de la fraction I.

**B :** Rétention et élution de l'ARN NRS-Tob (**B1**) et absence de rétention de l'ARN NRS (**B2**) sur la matrice selon les conditions de départ : 60 pmol d'ARN ont été ajoutés sur la matrice et les ARN NRS-Tob ont été élués avec 50 µl de tampon d'élution.

**C :** Rétention et élution de l'ARN NRS-Tob (**C1**) et absence de rétention de l'ARN NRS (**C2**) sur la matrice dans les conditions améliorées : 100 pmol d'ARN ont été ajoutés sur la matrice et les ARN NRS-Tob ont été élués avec 250 µl de tampon d'élution.

### III.3.1. Amélioration des conditions de rétention de l'ARN sur la matrice Tobra-Sepharose

Afin d'améliorer la fixation de l'ARN NRS-Tob sur la matrice, nous avons augmenté la quantité d'ARN NRS-Tob de départ (100 pmol au lieu de 60 pmol) et nous avons par ailleurs augmenté le volume de tampon d'élution (250 µl au lieu de 50 µl).

Dans ces nouvelles conditions, 54% de la totalité des ARN NRS-Tob ajoutés au départ ont été immobilisés sur la matrice après l'étape de lavage, et 90 % de ces ARN ont été élués (**Figure VIII-5, C1**). Par ailleurs, dans ces conditions, nous n'observions pas de rétention de l'ARN NRS (97% n'étaient pas retenus sur la matrice après l'étape de lavage) (**Figure VIII-5, C2**). Ces résultats étaient en accord avec les critères établis par K. Hartmuth et nous avons donc retenu ces conditions pour la suite des expériences.

La fixation et l'élution de l'ARN NRS-Tob ont été améliorées en augmentant respectivement la concentration de l'ARN de départ et le volume d'élution. Cela suggère que dans ces conditions, le Kd du complexe aptamère-tobramycine devait être supérieur à celui décrit dans la littérature (5 nM). Ce Kd avait été mesuré avec de la tobramycine libre par une expérience d'anisotropie de fluorescence dans un tampon de composition et de pH différents de celui utilisé lors de l'étape de rétention (pH 7,5 vs pH 9,1) (Hamasaki *et al.*, 1998). Les différences entre ces deux tampons pourraient être à l'origine de la plus faible affinité que nous avons observée. L'ionisation de la tobramycine à un pH proche de la neutralité permettrait la mise en place d'interactions électrostatiques supplémentaires avec l'aptamère et de ce fait contribuer à augmenter l'affinité (Jiang et Patel, 1998).

## III.4. Purifications des complexes ribonucléoprotéiques

Avant d'entreprendre les expériences de chromatographie d'affinité, il était important de vérifier que l'ARN Tob seul ne fixait pas les protéines de l'extrait nucléaire. Pour le vérifier, nous avons réalisé des expériences de retard sur gel avec les ARN NRS-Tob, NRS et Tob. Le lot d'extrait nucléaire utilisé lors du premier essai a été utilisé dans toute la suite des expériences de chromatographie d'affinité basées sur l'emploi de la tobramycine.

Comme l'indique la **Figure VIII-6, A**, aucun retard sur gel n'a été observé pour l'ARN Tob et ce, même lorsque d'importantes quantités d'extrait nucléaire ont été utilisées (32,5 µg pour la dernière piste). La séquence aptamérique ne fixait donc pas de protéines de l'extrait

nucléaire. En revanche, dans les mêmes conditions, des retards sur gel ont été obtenus pour les ARN NRS-Tob et NRS (**Figure VIII-6, A et B**). Les expériences de retard sur gel ayant donné des résultats satisfaisants, nous avons alors entrepris la purification des complexes ribonucléoprotéiques.

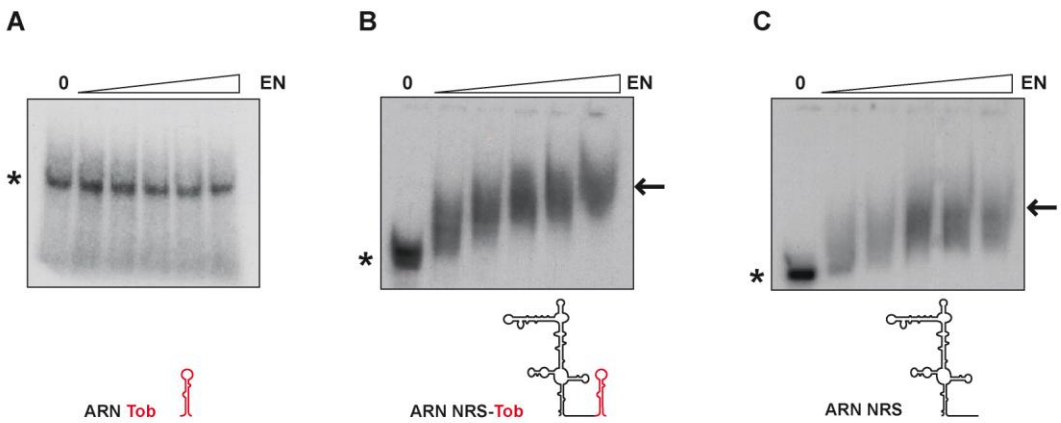



**Figure VIII-6 : Analyse par expérience de retard sur gel des complexes ribonucléoprotéiques formés après incubation des transcrits Tob (A), NRS-Tob (B) et NRS (C) en présence d'extrait nucléaire de cellules HeLa.**

Pour chacun des 3 ARN, un mélange de transcrits radioactifs (1,5 fmol) et non radioactifs (1,5 pmol) a été incubé en présence d'une quantité croissante d'extrait nucléaire de cellules HeLa (0 ; 6,5 ; 13 ; 19,5 ; 26 ; 32,5 µg de protéines totales) dans du tampon D (EDTA 0,2 mM ; KCl 0,1 M ; DTT 0,5 mM ; PMSF 0,25 mM ; HEPES-KOH 20 mM ; glycérol 20 %) supplémenté avec 2,8 mM de $MgCl_2$. Après électrophorèse en gel de polyacrylamide 5% non dénaturant en tampon TBE 0,5 X et glycérol 5%, les ARN sont révélés par autoradiographie. Les positions des ARN libres sont indiquées par des étoiles et celles des ARN complexés par des flèches.

La purification des complexes ribonucléoprotéiques est composée de 7 étapes (**Figure VIII-7**) : I) lavage de la matrice, II) saturation de la matrice III) fixation de l'ARN NRS-Tob, IV) lavage de la matrice, V) incubation avec l'extrait nucléaire VI) lavage de la matrice VII) élution des complexes ribonucléoprotéiques. Afin d'obtenir une quantité importante de complexes ribonucléoprotéiques tout en évitant d'avoir une concentration en extrait nucléaire trop élevée en contact avec la matrice, les expériences ont été réalisées au départ en doublon dans deux tubes distincts des étapes I à V et ont été ensuite réunies après le premier lavage de l'étape VI (**Figure VIII-7**).

Les étapes I à IV sont identiques à celles décrites pour le test de rétention de l'ARN NRS-Tob sur la matrice (**paragraphe III.3.1**). A la fin de l'étape IV, la radioactivité de chacune des fractions récoltées est mesurée et la quantité d'ARN retenu sur la matrice est calculée. L'expérience est poursuivie à condition que 50 à 70 % de l'ARN de départ soit retenu sur la matrice. Dans ce cas, l'extrait nucléaire est ajouté sur la matrice (étape V) et les complexes

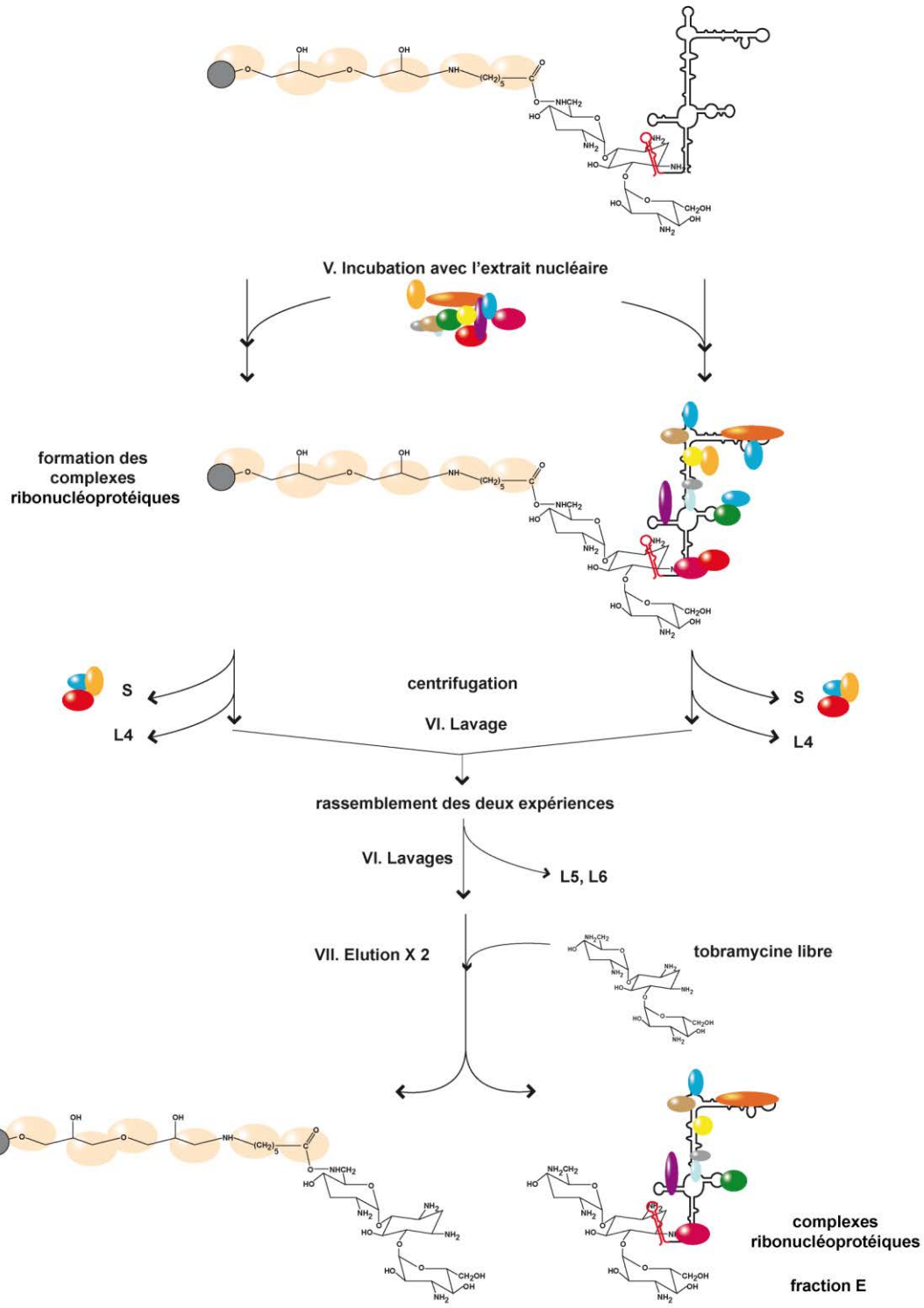



**Figure VIII-7 : Formation et purification des complexes ribonucléoprotéiques formés sur l'ARN NRS-Tob par chromatographie d'affinité sur la matrice Tobra-Sepharose.**
Le nom des différentes étapes ainsi que ceux des fractions collectées au cours de l'expérience sont mentionnés. Les étapes de l'expérience réalisées en doublon sont schématisées par deux flèches de part et d'autre du nom de l'étape concernée. L'ARN NRS-Tob est représenté par sa structure secondaire, la séquence de l'élément NRS et celle de l'aptamère Tob sont respectivement colorées en noir et rouge. Les étapes de saturation et de fixation de l'ARN NRS-Tob sur la matrice ne sont pas représentées. La saturation de la matrice par la BSA est schématisée par des sphères roses.
Le même protocole est utilisé pour l'expérience réalisée avec la matrice seule.

ribonucléoprotéiques sont formés durant 30 minutes à 30°C. Le matériel non retenu sur la matrice est éliminé par une brève centrifugation (fraction S) puis la matrice est lavée 3 fois de suite (étape VI, fractions L4, L5, L6). Les complexes ribonucléoprotéiques sont ensuite élués lors de la VIIe étape par deux ajouts successifs de 250 µl de tampon d'élution (fractions E1 et E2).

La même expérience est réalisée avec la matrice seule qui sert de témoin négatif.

La quantité d'extrait nucléaire ajouté sur de la matrice est un facteur important pour la formation des complexes ribonucléoprotéiques. Les conditions décrites par K. Hartmuth avaient été optimisées pour l'assemblage des complexes spliceosomaux sur un pré-ARNm (Hartmuth *et al.*, 2002). Dans ces expériences, un volume de 1,5 ml de tampon de fixation contenant 35 % d'extrait nucléaire, soit un volume de 525 µl, était ajouté sur une matrice où étaient préalablement fixés 33 à 38 pmol d'ARN.

Nous avons adapté cette étape à l'étude de l'élément NRS. La quantité d'extrait nucléaire ajoutée sur la matrice a été calculée d'après les résultats des expériences de retard sur gel (**paragraphe III.4**) et d'après la quantité d'ARN NRS-Tob retenu sur la matrice. Un volume de 230 µl de tampon de fixation contenant 35 %, soit 80 µl, d'extrait nucléaire, a été utilisé pour assembler les complexes ribonucléoprotéiques sur l'ARN NRS-Tob. Lors de cette étape, la concentration de l'ARN NRS-Tob est de $2{,}34.10^{-7}$ M soit 10 fois plus que celle de K. Hartmuth qui était de $2{,}2.10^{-8}$ M. En revanche, la quantité d'extrait nucléaire utilisée dans nos expériences était environ 10 fois moins importante par rapport à celle utilisée par K. Hartmuth. Dans ces conditions, nous nous attendions à ce que toute protéine ayant une affinité pour l'ARN soit retenue sur celui-ci.

La quantité d'ARN contenue dans chaque fraction a été déterminée en mesurant leur radioactivité au compteur à scintillation en Cerenkov. La radioactivité de l'ARN de départ ajouté sur la matrice (fraction Input) a été prise comme référence (100 %) et la radioactivité des autres fractions a été calculée en pourcentage par rapport à cette référence. Comme nous pouvons le constater, 54 % d'ARN NRS-Tob de départ ont été immobilisé sur la matrice après le premier lavage (**Figure VIII-8**). L'incubation de la matrice avec l'extrait nucléaire et les lavages qui s'en suivent ne provoquent qu'une faible dissociation de l'ARN de la matrice car environ 40 % de l'ARN de départ reste immobilisé sur la matrice après ces étapes. La quasi-totalité de ces ARN est ensuite élué au cours des deux élutions successives (**Figure VIII-8**).

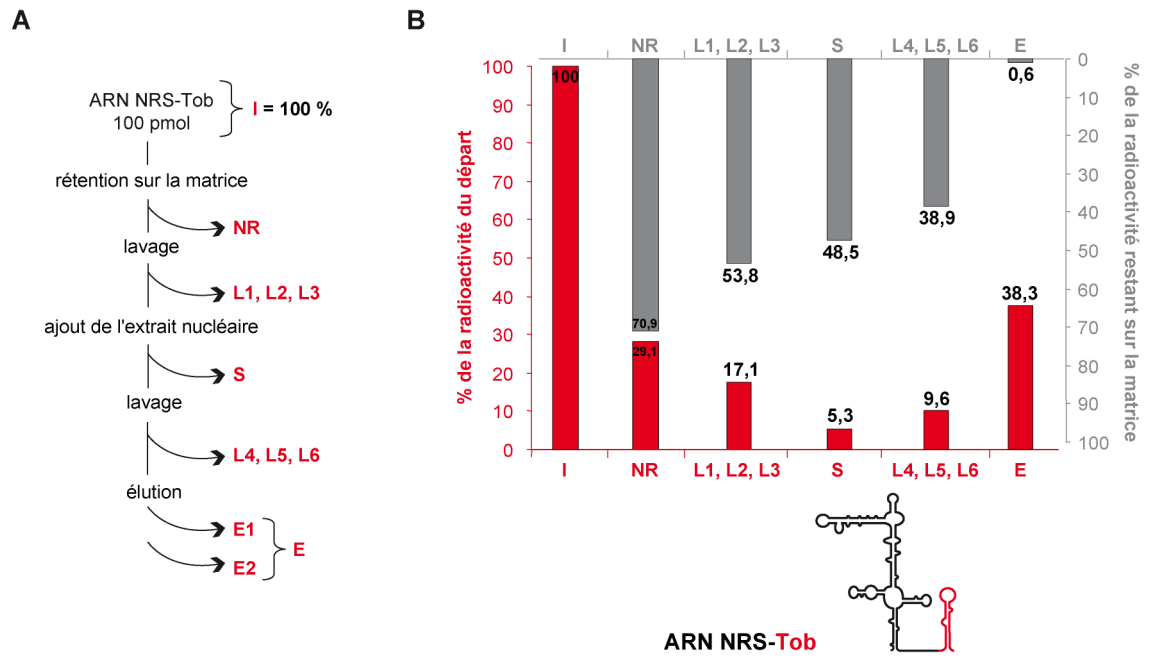



**Figure VIII-8 : Formation et purification des complexes ribonucléoprotéiques sur l'ARN NRS-Tob.**
**A :** Récapitulatif de l'expérience avec les différentes étapes et le nom des fractions collectées. L'assemblage des complexes a été réalisée avec un mélange d'ARN NRS-Tob non radioactifs (100 pmol) et uniformément marqués (20 fmol). La radioactivité contenue dans les fractions collectées a été mesurée en Cerenkov au compteur à scintillation. La fraction I correspond à l'Input (100 % de l'ARN au départ), NR, aux ARN non retenus sur la matrice, L1, L2, L3, à 3 lavages successifs de la matrice, S, à la fraction non retenue après l'ajout de l'extrait nucléaire, L4, L5, L6 à 3 lavages successifs de la matrice et E1, E2 aux 2 élutions successives.
**B :** Les histogrammes rouge et gris représentent respectivement la radioactivité contenue dans chaque fraction et la radioactivité restant sur la matrice après chaque étape. Les valeurs sont exprimées en pourcentage par rapport à la radioactivité de la fraction I (100 %).

Les fractions collectées après l'ajout de l'extrait nucléaire sont analysées. Les protéines et les ARN contenus dans ces fractions sont séparés par une extraction phénol-chloroforme. Les protéines contenues dans la phase phénolique ont été précipitées à l'acétone et analysés par SDS-PAGE et Western blot. Les ARN contenus dans la phase aqueuse ont été précipités à l'éthanol et analysés par électrophorèse en gel de polyacrylamide dénaturant suivie d'une coloration au bromure d'éthidium.

### III.5. Analyse des protéines contenues dans les complexes ribonucléoprotéiques assemblés sur l'ARN NRS-Tob

Les protéines contenues dans les fractions collectées ont été fractionnées sur gel SDS-PAGE et colorées au bleu colloïdal (**Figure VIII-9**). En premier lieu, nous avons constaté que les lavages servant à éliminer les protéines faiblement associées à la matrice étaient efficaces car très peu de protéines sont observées dans la fraction L6. Très peu de protéines sont retrouvées

dans la fraction d'élution lorsque la matrice seule était utilisée et en revanche, beaucoup de protéines sont éluées lorsque l'ARN NRS-Tob avait été immobilisé sur la matrice.

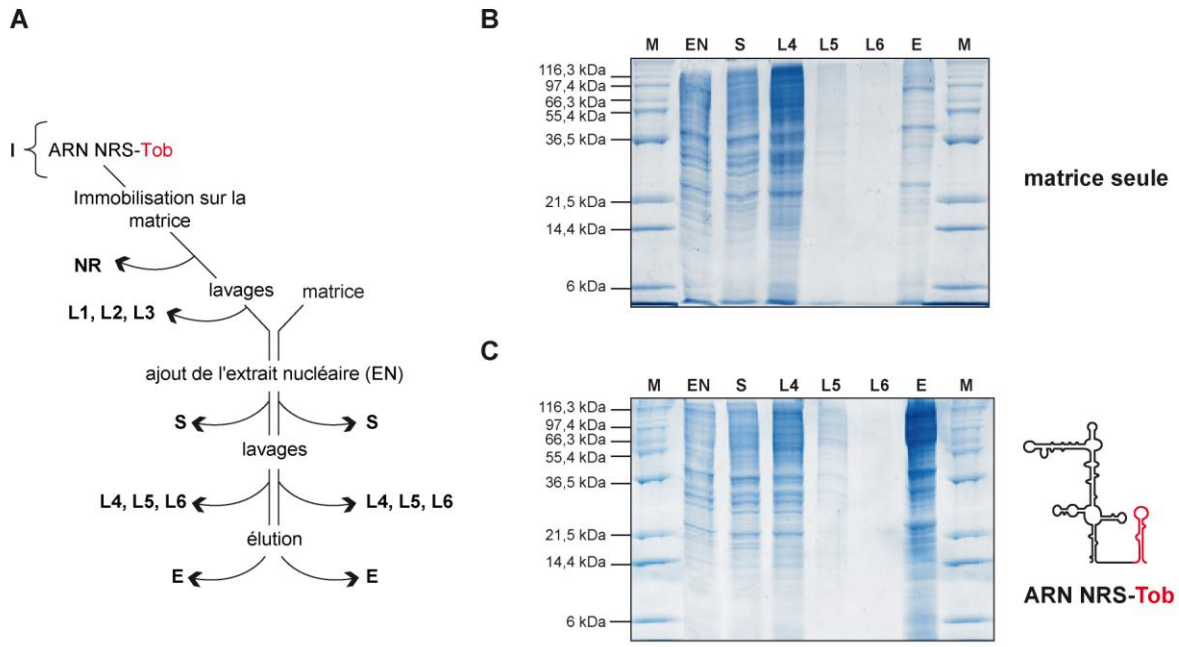



**Figure VIII-9 : Analyse des protéines contenues dans les échantillons récupérés après l'ajout de l'extrait nucléaire sur la matrice.**

**A :** Récapitulatif des expériences réalisées avec la matrice où l'ARN NRS-Tob a été préalablement fixé et avec la matrice seule. Les différentes étapes et le nom des fractions collectées sont mentionnés. La fraction I correspond à l'ARN de départ ou input, NR, aux ARN non retenus sur la matrice, L1, L2, L3, à 3 lavages de la matrice, S, à la fraction non retenue après l'ajout de l'extrait nucléaire, L4, L5, L6 à 3 lavages de la matrice et E à l'élution. Seules les protéines des fractions récupérées après l'ajout de l'extrait nucléaire ont été analysées par électrophorèse en gel SDS-PAGE.

**B :** Fractionnement par gel SDS-PAGE des protéines des fractions collectées après l'ajout de l'extrait nucléaire de cellules HeLa sur la matrice seule où aucun ARN n'a été préalablement immobilisé. Les protéines correspondant à 1/10e de la fraction S, la moitié des fractions L4, L5, L6 et la moitié des deux élutions rassemblées E ont été précipitées avec de l'acétone et déposées. La piste EN correspond au dépôt de 1 µl de l'extrait nucléaire brut utilisé durant la chromatographie d'affinité, la piste M correspond au marqueur de taille et les poids moléculaires des protéines sont reportés.

**C :** Fractionnement par gel SDS-PAGE des protéines des fractions collectées après l'ajout de l'extrait nucléaire de cellules HeLa sur la matrice où l'ARN NRS-Tob a été préalablement immobilisé. Les quantités de protéines déposées sont identiques à celles déposées sur le gel en B).

Il n'était pas possible de visualiser un potentiel enrichissement d'une protéine donnée dans les fractions d'élution analysées en gel SDS-PAGE. Afin de vérifier la spécificité des complexes ribonucléoprotéiques formés sur l'ARN NRS-Tob, nous avons voulu rechercher la présence des protéines SR ASF/SF2 et 9G8 qui avaient été décrites comme se fixant sur l'élément NRS (Fogel *et al.*, 2002 ; McNally et McNally, 1996). Pour cela, nous avons réalisé des expériences de Western blot avec des anticorps spécifiques dirigés contre ces protéines (**Figure VIII-10**).

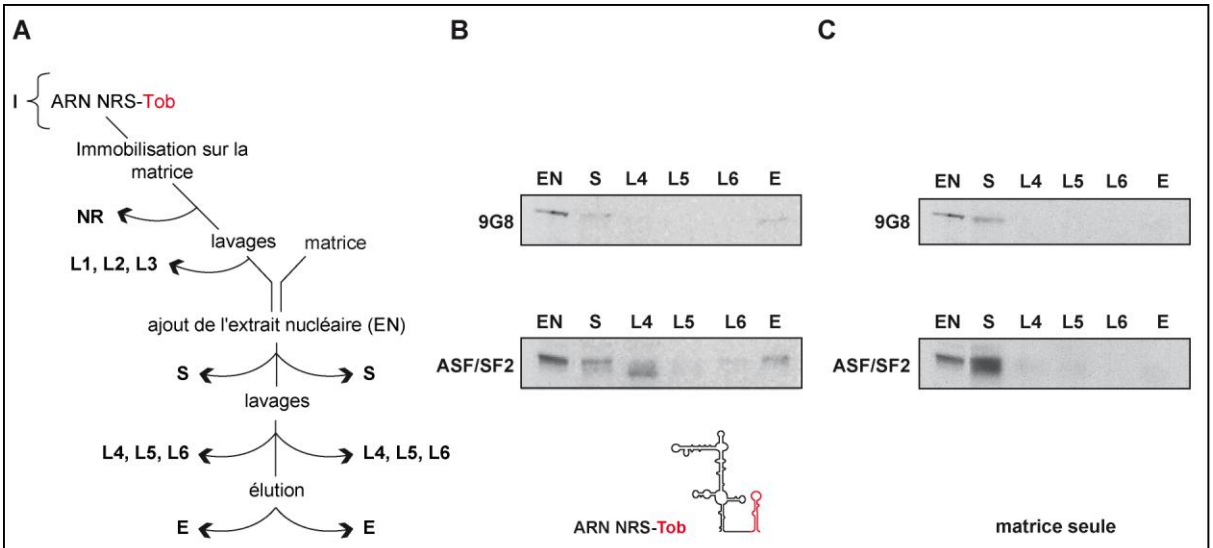



**Figure VIII-10 : Analyse par Western blot de la présence des protéines ASF et 9G8 dans les échantillons collectées après l'ajout de l'extrait nucléaire sur la matrice.**
**A :** Récapitulatif des expériences réalisées avec la matrice où l'ARN NRS-Tob a été préalablement fixé et avec la matrice seule. Les différentes étapes et le nom des fractions collectées sont mentionnés. La fraction I correspond à l'ARN de départ ou input, NR, aux ARN non retenus sur la matrice, L1, L2, L3, à 3 lavages successifs de la matrice, S, à la fraction non retenue après l'ajout de l'extrait nucléaire, L4, L5, L6 à 3 lavages successifs de la matrice et E à l'élution. Seules les protéines des fractions récupérées après l'ajout de l'extrait nucléaire ont été fractionnées par électrophorèse en gel SDS-PAGE. Les protéines ont été ensuite transférées sur une membrane de nitrocellulose pour l'expérience de Western blot. Les protéines ASF/SF2 et 9G8 ont été révélées avec des anticorps leur étant spécifiques. La piste EN correspond au dépôt de 5 µg de protéine totale d'extrait nucléaire de cellules HeLa.
**B :** Recherche par Western blot de la présence des protéines ASF/SF2 et 9G8 dans les différents échantillons collectés après l'ajout de l'extrait nucléaire sur la matrice où l'ARN NRS-Tob a été préalablement immobilisé. Le Western blot a été réalisé avec les quantités de protéines correspondant à 1/10e des fractions S, L4, L5, L6 et des deux élutions rassemblées, E.
**C :** Recherche par Western blot de la présence des protéines ASF/SF2 et 9G8 dans les différents échantillons collectés après l'ajout de l'extrait nucléaire sur la matrice seule.

Ces protéines SR étaient présentes dans les fractions d'élution uniquement lorsque l'ARN NRS-Tob était immobilisé sur la matrice, ce qui validait l'approche de chromatographie d'affinité utilisée. Nous avons vérifié la fixation d'ASF/SF2 sur les ARN NRS-Tob et NRS en réalisant des expériences de retard sur gel avec la protéine recombinante purifiée (**Figure VIII-11**). L'expérience était réalisée avec un mélange de 1,5 pmol d'ARN non radioactif et 1,5 fmol d'ARN uniformément marqué incubé en présence de quantités croissantes d'ASF/SF2. Comme il est montré dans la **Figure VIII-11**, les ARN NRS et NRS-Tob étaient entièrement complexés pour une concentration de protéine ASF/SF2 de 1,9 µM, ce qui représente une bonne affinité de la protéine pour l'élément NRS.

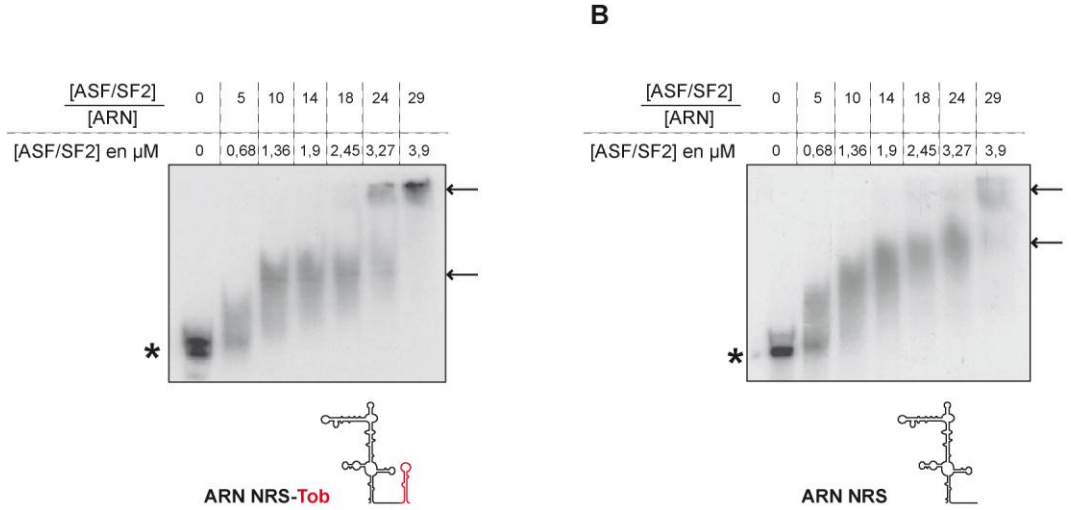



**Figure VIII-11 : Analyse par expérience de retard sur gel des complexes formés après incubation des transcrits NRS-Tob (A), NRS (B) en présence de protéine ASF/SF2 recombinante purifiée.**
Pour chacun des 2 ARN, un mélange de transcrits radioactifs (1,5 fmol) et non radioactifs (1,5 pmol) a été incubé en présence d'une quantité croissante de protéine ASF/SF2 (0 ; 200 ; 400 ; 600 ; 800 ; 1000 ; 1200 ng). Après électrophorèse en gel de polyacrylamide 5% non dénaturant en tampon TBE 0,5 X et glycérol 5%, les ARN sont révélés par autoradiographie. Les positions des ARN libres sont indiquées par des étoiles et celles des ARN complexés par des flèches.

### III.6. Analyse des ARN contenus dans les complexes ribonucléoprotéiques assemblés sur l'ARN NRS-Tob

Nous avons ensuite analysé le contenu en UsnRNA des fractions éluées de la colonne d'affinité par une électrophorèse en condition dénaturante suivie d'une coloration au bromure d'éthidium. Comme le montre la **Figure VIII-12**, aucun ARN n'est visible dans la fraction d'élution lorsque la matrice seule était utilisée. En revanche, la fraction d'élution E est enrichie en snRNA U1, U2 ainsi qu'un ARN migrant au niveau de l'ARN 5S lorsque l'ARN NRS-Tob était préalablement immobilisé sur la matrice. Les deux snRNP U1 et U2 ont été déjà décrites comme interagissant avec l'élément NRS : la snRNP U1 interagit avec le pseudo-site d'épissage et la snRNP U2 interagirait avec une séquence riche en polypyrimidine en amont du pseudo site (Cook et McNally, 1998 ; Hibbert *et al.*, 1999 ; McNally et McNally, 1999). Comme le montrent ces résultats, la chromatographie d'affinité permettrait d'assembler et de purifier des complexes ribonucléoprotéiques spécifiques sur l'ARN NRS-Tob.

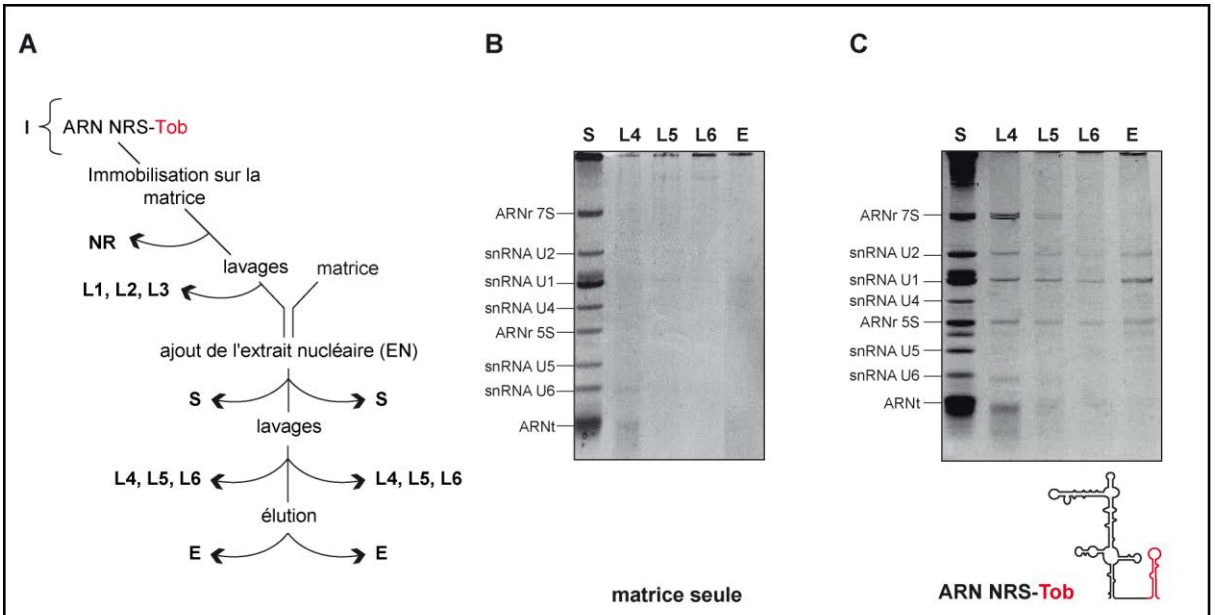



**Figure VIII-12 : Analyse des ARN contenus dans les échantillons récupérés après l'ajout de l'extrait nucléaire sur la matrice.**

**A :** Récapitulatif des expériences réalisées avec la matrice où l'ARN NRS-Tob a été préalablement fixé et avec la matrice seule. Les différentes étapes et le nom des fractions collectées sont mentionnées. La fraction I correspond à l'ARN de départ ou input, NR, aux ARN non retenus sur la matrice, L1, L2, L3, à 3 lavages successifs de la matrice, S, à la fraction non retenue après l'ajout de l'extrait nucléaire, L4, L5, L6 à 3 lavages successifs de la matrice et E à l'élution. Seuls les ARN des fractions récupérées après l'ajout de l'extrait nucléaire ont été analysés en gel de polyacrylamide dénaturant.

**B :** Fractionnement des ARN des fractions collectées après l'ajout de l'extrait nucléaire de cellules HeLa sur la matrice seule où aucun ARN n'a été préalablement immobilisé. Les ARN contenus dans 1/10e de la fraction S, et la moitié des fractions L4, L5, L6 et E ont été purifiés par une extraction au phénol-chloroforme, précipités et fractionnés par électrophorèse en gel de polyacrylamide dénaturant 10%-urée 8M. Les ARN ont été révélés par une coloration au bromure d'éthidium.

**C :** Fractionnement des ARN des fractions collectées après l'ajout de l'extrait nucléaire de cellules HeLa sur la matrice où l'ARN NRS-Tob a été préalablement immobilisé. Les quantités d'ARN déposées sont identiques à celles déposées sur le gel en B).

## III.7. Conclusion

La méthode de chromatographie d'affinité basée sur l'emploi de la tobramycine dont nous avions optimisé les conditions nous a permis d'obtenir des complexes formés sur l'ARN NRS, ceci avec un degré de pureté relativement satisfaisant. Cette méthode présente l'inconvénient d'être réalisée à un pH élevé de 9,1, ce qui pouvait altérer la composition en protéine des complexes. Nous avons donc décidé d'évaluer une deuxième technique, basée sur l'emploi de la protéine de l'enveloppe du phage MS2 et de son ARN partenaire. Cette technique venait d'être mise au point par l'équipe de R. Lührmann et dans le cadre d'une collaboration, nous avons alors testé cette méthode afin de la comparer aux données obtenues avec la technique basée sur l'emploi de la tobramycine.

# IV. Chromatographie d'affinité basée sur la protéine de l'enveloppe du phage MS2

Comme nous l'avons déjà expliqué, cette méthode repose sur le couple ARN-MS2/protéine MS2-MBP. Il était nécessaire de réaliser en premier lieu des constructions nous permettant de transcrire *in vitro* l'ARN NRS fusionné à l'ARN MS2 et d'autre part de produire la protéine de fusion MS2-MBP.

## IV.1. Production des ARN hybrides

### IV.1.1. Le motif ARN reconnu par la protéine de l'enveloppe du phage MS2

Le phage MS2 possède dans son ARN viral un élément inhibiteur de la traduction du gène de la réplicase. Lors de l'inhibition, l'AUG initiateur de la traduction se retrouve dans une structure tige boucle de 19 nucléotides appelée TR (Translation Repressor) qui interagit avec la protéine virale de l'enveloppe (**Figure VIII-13, A**). Cette interaction est spécifique et la séquence TR possède une forte affinité pour la protéine de la capside (Kd =3 nM) (Lim et Peabody, 1994). Ce complexe empêche le ribosome d'accéder à l'AUG et sert d'évènement déclencheur pour amorcer l'assemblage des nouveaux virions.

La séquence que nous avons utilisée pour la chromatographie d'affinité est un ARN aptamérique qui est dérivé de la séquence naturelle TR et qui a été sélectionné par la méthode SELEX pour sa forte affinité pour la protéine de l'enveloppe du phage MS2 (Valegard *et al.*, 1997). Cet aptamère se structure en une tige boucle stable de 7 nucléotides avec une boucle latérale et d'une boucle terminale de 4 nucléotides (**Figure VIII-13, C**). L'adénosine de la boucle latérale et 3 des 4 nucléotides situés dans la boucle terminale sont les déterminants de l'interaction avec la protéine de l'enveloppe. Pour réaliser la chromatographie d'affinité, trois exemplaires de cet aptamère, ayant un Kd de 38 pM pour la protéine de l'enveloppe du phage MS2 ont été fusionnés cotranscriptionnellement à l'extrémité 3' de l'ARN NRS (Jurica et Moore, 2002 ; Zhou *et al.*, 2002) (**Figure VIII-13, D**).

### IV.1.2. Réalisation des constructions génétiques

Pour réaliser la chromatographie d'affinité, j'ai d'abord construit le plasmide pT7MS2. Une fois linéarisé par *Xba*I, ce plasmide permet de transcrire *in vitro* l'ARN MS2 contenant les 3 séquences aptamériques, que nous avons utilisé comme contrôle (**Figure VIII-14, A**). Dans un second temps, j'ai construit le plasmide pNRS WT-MS2 qui permet de produire par

transcription *in vitro* soit l'ARN NRS WT fusionné à l'ARN MS2 soit l'ARN NRS seul, suivant le site de restriction (*Xba*I ou *Bam*HI) utilisé pour la linéarisation (**Figure VIII-14, A**). Les plasmides contenant les séquences des éléments NRS tronqués NRS Δ76, NRS 22, NRS 24, NRS 26 et NRS 28 ont été produits par mutagenèse dirigée à partir du plasmide pNRS-MS2, ceci en utilisant les oligonucléotides et le protocole détaillé dans la partie Matériel et Méthodes (**Figure VIII-14, B**). Les troncations de l'élément NRS ont été délimitées à partir du modèle de structure secondaire établi pour l'élément NRS sauvage, ceci afin de conserver les différents éléments de structure secondaire comme nous l'avons expliqué dans le chapitre I des Résultats. Comme pour le plasmide pNRS WT-MS2, chacune des constructions (pNRS Δ76-MS2, pNRS 22-MS2, pNRS 24-MS2, pNRS 26-MS2, pNRS 28-MS2) nous permettait de produire par transcription *in vitro* un ARN NRS fusionné à l'ARN MS2 ou un ARN NRS seul en fonction du choix du site de restriction *Xba*I ou *Bam*HI respectivement utilisé pour la linéarisation.

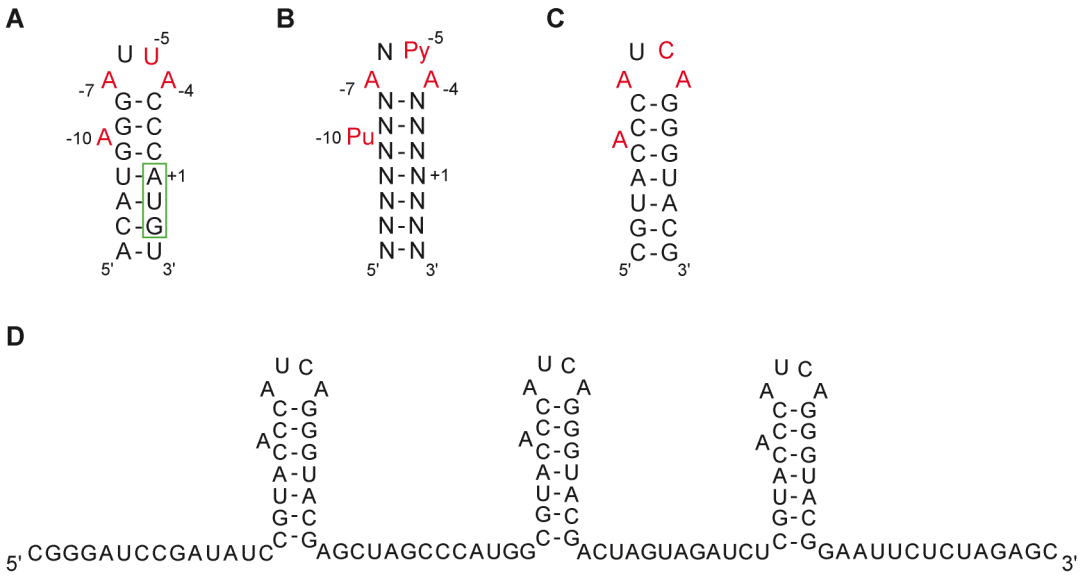



**Figure VIII-13 : Séquences et structures des différents ARN fixant la protéine de l'enveloppe du phage MS2 (Valegard *et al.*, 1997 ; Zhou *et al.*, 2002).**
**A :** Structure de l'élément TR du phage MS2 fixant la protéine de l'enveloppe. Les nucléotides sont numérotés par rapport au site d'initiation de la traduction du gène de la réplicase qui est encadré en vert. Les nucléotides importants pour l'interaction avec la protéine de l'enveloppe sont représentés en rouge.
**B :** Motif consensus des aptamères obtenus par l'approche SELEX avec la protéine de l'enveloppe. Les nucléotides importants pour l'interaction avec la protéine de l'enveloppe sont indiqués en rouge. Pu : purine ; Py : pyrimidine ; N : n'importe quel nucléotide.
**C :** Aptamère MS2 utilisé pour la chromatographie d'affinité. Les nucléotides importants pour la reconnaissance et l'affinité pour la protéine MS2 sont colorés en rouge.
**D :** Structure adoptée par l'ensemble des 3 séquences aptamériques qui sont ajoutées à l'ARN d'intérêt pour réaliser la chromatographie d'affinité.

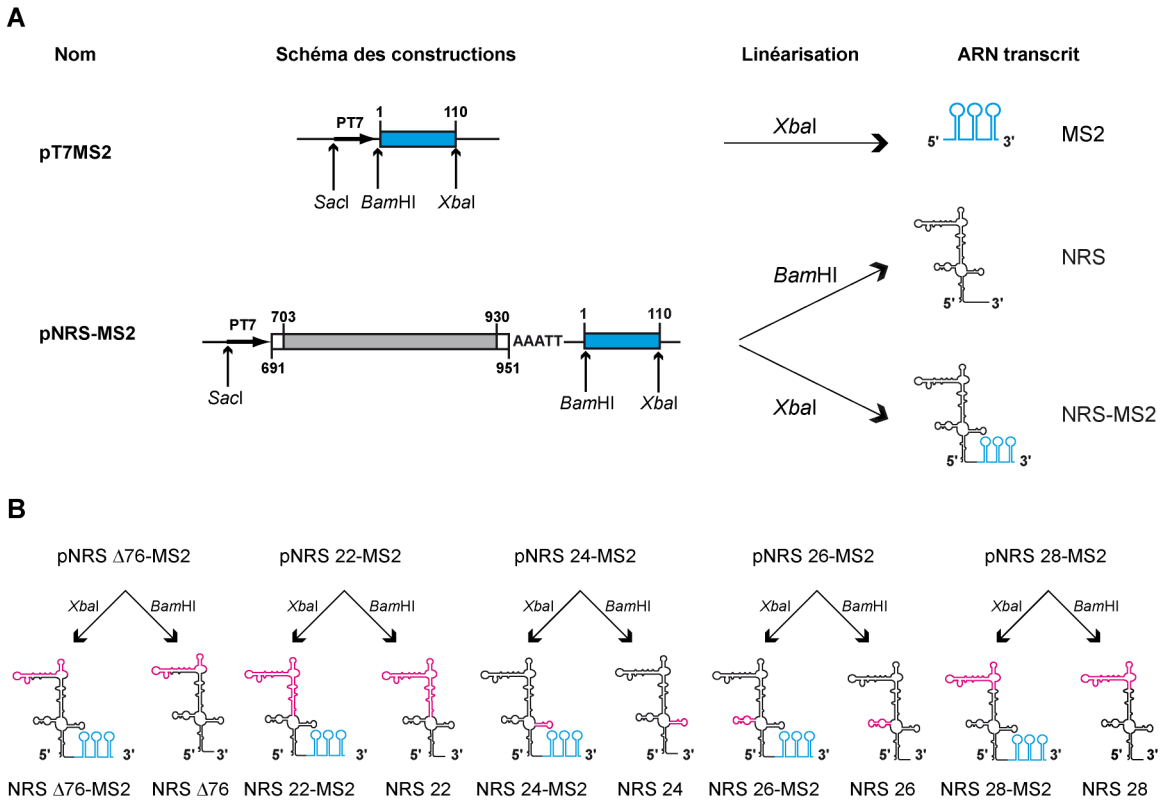



**Figure VIII-14 : Constructions réalisées pour produire les ARN utilisés dans les expériences de chromatographie d'affinité basées sur l'emploi de la protéine de l'enveloppe du phage MS2.**
**A :** Constructions réalisées pour transcrire *in vitro* les ARN MS2 et NRS WT-MS2.
**B :** Constructions réalisées pour transcrire *in vitro* les différents éléments NRS tronqués fusionnés à l'ARN MS2.
L'ADN correspondant à l'ARN MS2 est représenté par des rectangles bleus, l'ADN de l'élément NRS, ainsi que ses séquences bordantes sont respectivement représentés par des rectangles gris et blancs et les nucléotides supplémentaires AAATT ajoutés à l'extrémité du NRS sont indiqués. Les positions des nucléotides de l'élément NRS sont indiqués (Schwart *et al.*, 1983). La flèche horizontale représente le promoteur transcriptionnel du bactériophage T7 et les flèches verticales représentent les sites de restriction utilisés pour les clonages et la linéarisation des matrices avant la transcription. Le nom des différents transcrits produits à partir des matrices linéarisées sont indiqués et les ARN produits sont représentés selon leur structure secondaire. La séquence ARN de l'élément NRS est colorée en noir, les séquences tronquées en rose et la séquence de l'ARN MS2 en bleu.

## IV.2. Production de la protéine de fusion

### IV.2.1. La protéine de fusion utilisée

La protéine de l'enveloppe du phage MS2 comporte 129 résidus. Elle interagit avec sa séquence naturelle TR sous forme d'homodimère non covalent (Valegard *et al.*, 1994) (**Figure VIII-15, A**). Par rapport à la protéine sauvage, la protéine MS2 qui est utilisée pour la méthode de chromatographie d'affinité est mutée au niveau de deux résidus (V75Q et A81G) pour ne plus former d'homodimère et interagir en condition stoechiométrique avec l'ARN aptamérique (Jurica et Moore, 2002 ; LeCuyer *et al.*, 1995) (**Figure VIII-15, B**). Afin d'être utilisée pour la purification des complexes ribonucléoprotéiques, la protéine MS2 est

fusionnée à la protéine MBP (Maltose Binding Protein) (**Figure VIII-15, C**). Cette dernière a une très forte affinité pour l'amylose et le maltose (Kd amylose = 0,16 µM, Kd maltose = 3,5 µM). Ainsi, la protéine de fusion MS2-MBP peut être spécifiquement retenue sur des billes d'amylose et dissociée de ces billes par l'ajout de maltose.

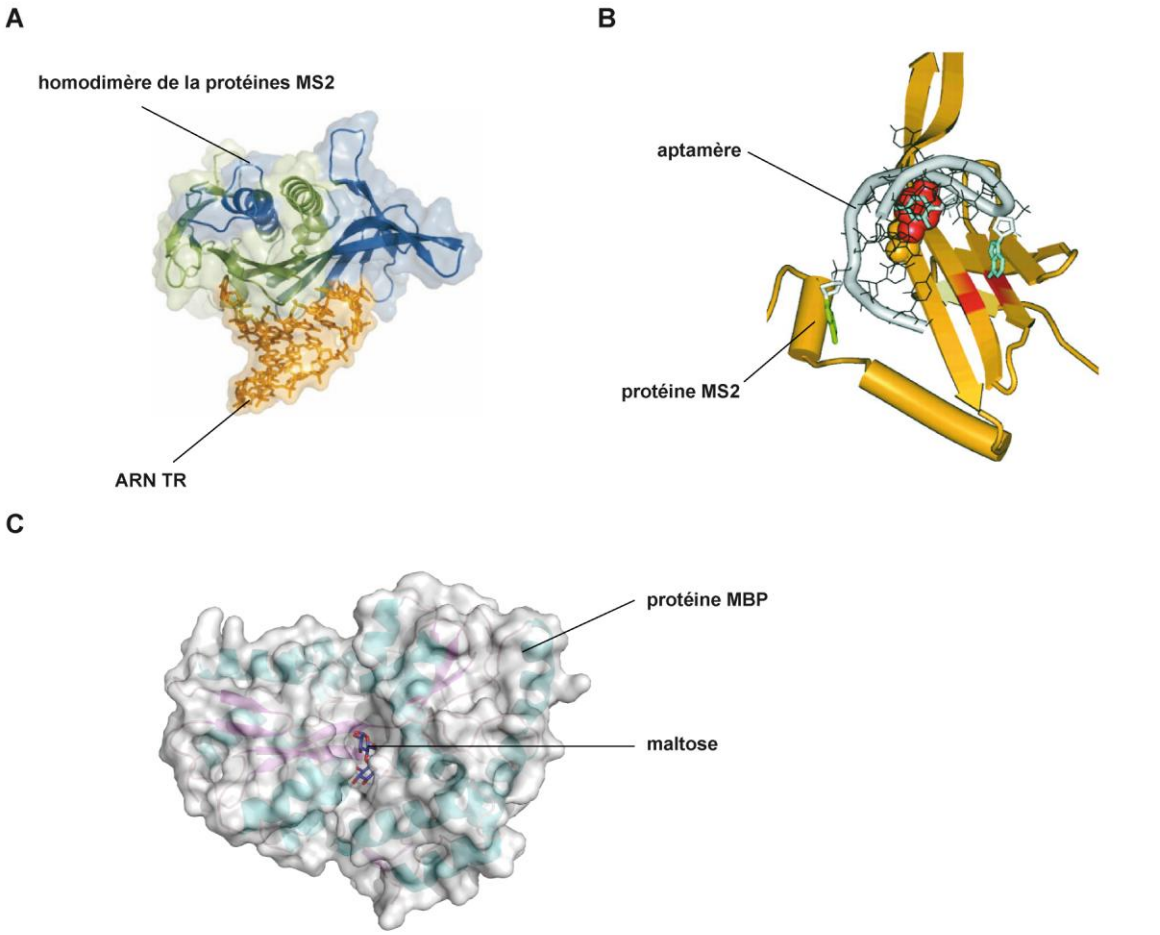



**Figure VIII-15 : Structures cristallographiques des complexes formés entre l'homodimère de la protéine de l'enveloppe du phage MS2 et l'ARN TR (A), la protéine de l'enveloppe du phage MS2 et l'ARN aptamérique obtenu par SELEX (B) et la protéine MBP et le maltose (C) (Valegard *et al.*, 1997 ; Convery *et al.*, 1998 ; Quiocho *et al.*, 1997).**

**A :** Structure cristallographique du complexe d'inhibition de la traduction du gène de la réplicase du phage MS2. Le complexe est formé par l'homodimère de la protéine de l'enveloppe et le fragment TR de l'ARN du phage MS2. L'ARN est représenté en jaune, la protéine MS2 en vert et bleu (Valegard *et al.*, 1997).

**B :** Structure cristallographique de la protéine de l'enveloppe du phage MS2 associée à l'ARN aptamérique obtenu par l'approche SELEX. Le squelette phosphodiester de l'ARN est représenté en gris, la protéine MS2 est représentée en jaune. Les sites d'interaction sont en rouge sur la protéine et en vert et bleu sur l'ARN (Convery *et al.*, 1998).

**C :** Structure cristallographique de la protéine MBP associée avec une molécule de maltose. La structure de la protéine est en bleu et violet (Quiocho *et al.*, 1997).

### IV.2.2. Production et purification de la protéine de fusion

Le vecteur pMAL contenant la séquence codante de la protéine MS2-MBP nous avait été fourni par l'équipe de R. Lührmann. Ce vecteur permet de produire la protéine MS2 fusionnée à son extrémité N-terminale avec la protéine MBP, ceci sous le contrôle d'un promoteur inductible P*tac*. Dans ce vecteur, la production de la protéine de fusion peut être induite par l'ajout d'IPTG dans le milieu de culture, et réprimée par le répresseur LacI codé par le vecteur. LacI empêche l'accès de l'ARN polymérase au promoteur, tandis que l'IPTG lève cette inhibition. Les bactéries de la souche d'*E. coli* DH5α transformées avec le plasmide d'expression pMAL-MS2 ont été cultivées dans un milieu LB contenant de l'ampicilline à 37°C. Lorsque la DO atteignait 0,8, la production de la protéine était induite par l'ajout d'IPTG à une concentration finale de 500 µM durant 3 h. Après vérification de l'induction de la production de la protéine MS2-MBP par SDS-PAGE (**Figure VIII-16, A**), les bactéries étaient lysées par sonication comme décrit dans la partie Matériel et Méthodes. La fraction soluble de l'homogénat était ensuite déposée sur une colonne amylose et les protéines MS2-MBP retenues sur la colonne étaient éluées par l'ajout de 15 mM de maltose. Ces protéines étaient ensuite purifiées sur une colonne héparine-Sepharose et éluées par un gradient de NaCl allant de 0 à 1M. Les protéines étaient éluées lorsque la concentration en NaCl était comprise entre 120 mM et 230 mM (**Figure VIII-16, D, fractions 3 à 7**). Après dialyse contre du tampon D, le degré de pureté de la protéine a été estimé par SDS PAGE et son bon repliement vérifié par une expérience de retard sur gel avec l'ARN MS2 (**Figure VIII-16, E**).

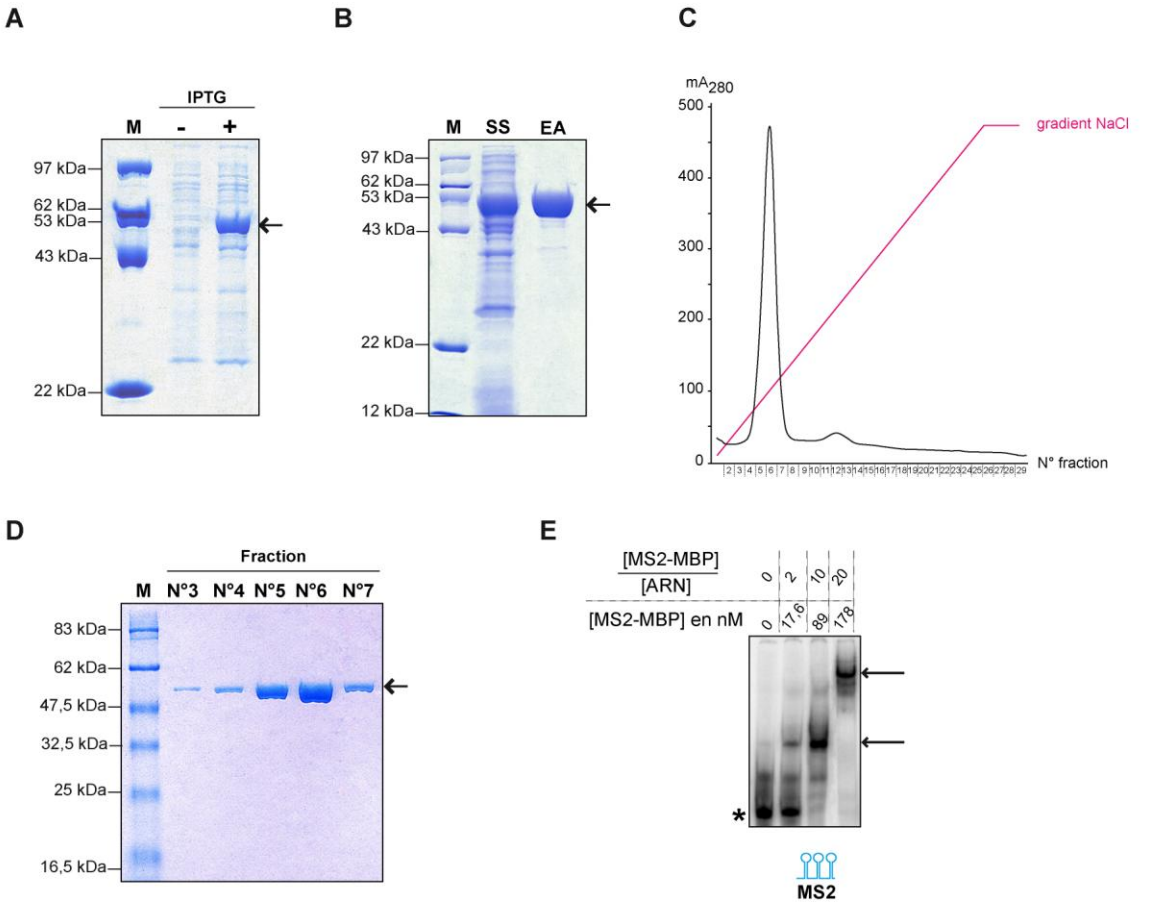



**Figure VIII-16 : Production et purification de la protéine MS2-MBP.**
**A :** Vérification de la surproduction de la protéine MS2-MBP après l'induction par l'IPTG. Des échantillons ont été prélevés avant (IPTG-) et 3 h après l'ajout de l'IPTG (IPTG+), soniqués et analysés par électrophorèse en gel SDS-PAGE 10 %. La protéine MS2-MBP est indiquée par une flèche, les poids moléculaires des protéines du marqueur de taille (M) sont indiqués.
**B :** Purification de la protéine MS2-MBP sur colonne amylose. Après l'induction, les bactéries ont été lysées par sonication et la fraction soluble du lysat bactérien obtenue après centrifugation (SS) a été purifiée sur une colonne amylose. La protéine MS2-MBP retenue sur la colonne a été décrochée par l'ajout de 15 mM de maltose (EA). La protéine MS2-MBP est indiquée par une flèche, les poids moléculaires des protéines du marqueur de taille (M) sont indiqués.
**C :** Chromatogramme de la purification de la protéine MS2-MBP sur colonne héparine-Sepharose. L'éluat de la colonne amylose est purifié lors d'une seconde étape sur une colonne héparine par un gradient en NaCl.
**D :** Purification de la protéine MS2-MBP sur colonne héparine. Les protéines contenues dans les fractions d'élution de la colonne (N°3 à 7) ont été analysées par électrophorèse en gel SDS PAGE 10 %. La protéine MS2-MBP est indiquée par une flèche, les poids moléculaires des protéines du marqueur de taille (M) sont indiqués.
**E :** Vérification du bon repliement de la protéine MS2-MBP par expérience de retard sur gel avec l'ARN MS2. Un mélange de transcrits non radioactifs (100 fmol) et uniformément marqués (1,5 fmol) a été incubé en présence de quantités croissantes de protéine MS2-MBP purifiée. La position des ARN libres est indiquée par une étoile et les ARN complexés sont indiqués par une flèche. L'ARN MS2 est schématisé en dessous de l'autoradiogramme.

## IV.3. Adaptation de la chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2 à la purification des complexes ribonucléoprotéiques formés sur l'élément NRS

### IV.3.1. Test de la fixation de la protéine MS2-MBP sur les ARN MS2 et NRS-MS2 par expériences de retard sur gel

Lorsque nous avons commencé à utiliser cette approche, I. Motorine et C. Aigueperse avaient déjà développé son utilisation au laboratoire pour la purification de complexes formés sur un fragment de l'ARN du virus VIH-1 et de ce fait, avaient déjà adapté les conditions mises au point par le laboratoire de R. Lührmann pour la purification des complexes spliceosomaux. Néanmoins, l'élément NRS sur lequel nous travaillons était plus grand que l'ARN du VIH-1 qu'ils étudiaient et en plus il pouvait fixer à la fois des ARN, des UsnRNP, et des protéines. Tout en bénéficiant des mises au point déjà faites, il nous a paru nécessaire de refaire une mise au point adaptée à l'élément NRS.

Dans un premier temps, nous devions définir les concentrations respectives d'ARN NRS-MS2 et de protéine MS2-MBP à utiliser pour avoir une immobilisation efficace et spécifique des complexes ARN NRS-MS2/MS2-MBP sur les billes d'amylose. Nous avons d'abord testé l'affinité de l'ARN NRS-MS2 pour la protéine de fusion MS2-MBP par des expériences de retard sur gel.

Les expériences ont été réalisées avec l'ARN NRS-MS2, l'ARN MS2 contrôle et l'ARN NRS non fusionné servant de témoin négatif. Un mélange de transcrits non radioactifs (100 fmol), de transrits uniformément marqués (1,5 fmol) et d'ARNt (10 pmol) a été incubé en présence de quantités croissantes de protéine MS2-MBP recombinante purifiée (2,7 ng à 0,1 µg). Les complexes formés ont été analysés par une électrophorèse en gel non dénaturant (**Figure VIII-17)**.

Lorsque la protéine MS2-MBP était présente en excès molaire de 10 fois, l'ARN MS2, ainsi que l'ARN NRS-MS2, étaient totalement complexés alors que l'ARN NRS était majoritairement présent sous forme libre. Un complexe n'apparaissait pour cet ARN qu'à un ratio protéine/ARN de 25 et ceci dans de très faibles proportions. L'affinité de la protéine de fusion pour l'ARN NRS était donc très faible, ce qui nous permettait en utilisant un ratio protéine/ARN de 20 dans les expériences ultérieures d'avoir une spécificité de rétention optimale des ARN fusionnés. Par ailleurs, cela nous permettait de vérifier que les motifs MS2 se repliaient correctement, même en présence de la séquence NRS. La présence de plusieurs

complexes lorsque nous augmentions la concentration de la protéine de fusion s'expliquait par la présence des 3 motifs MS2.

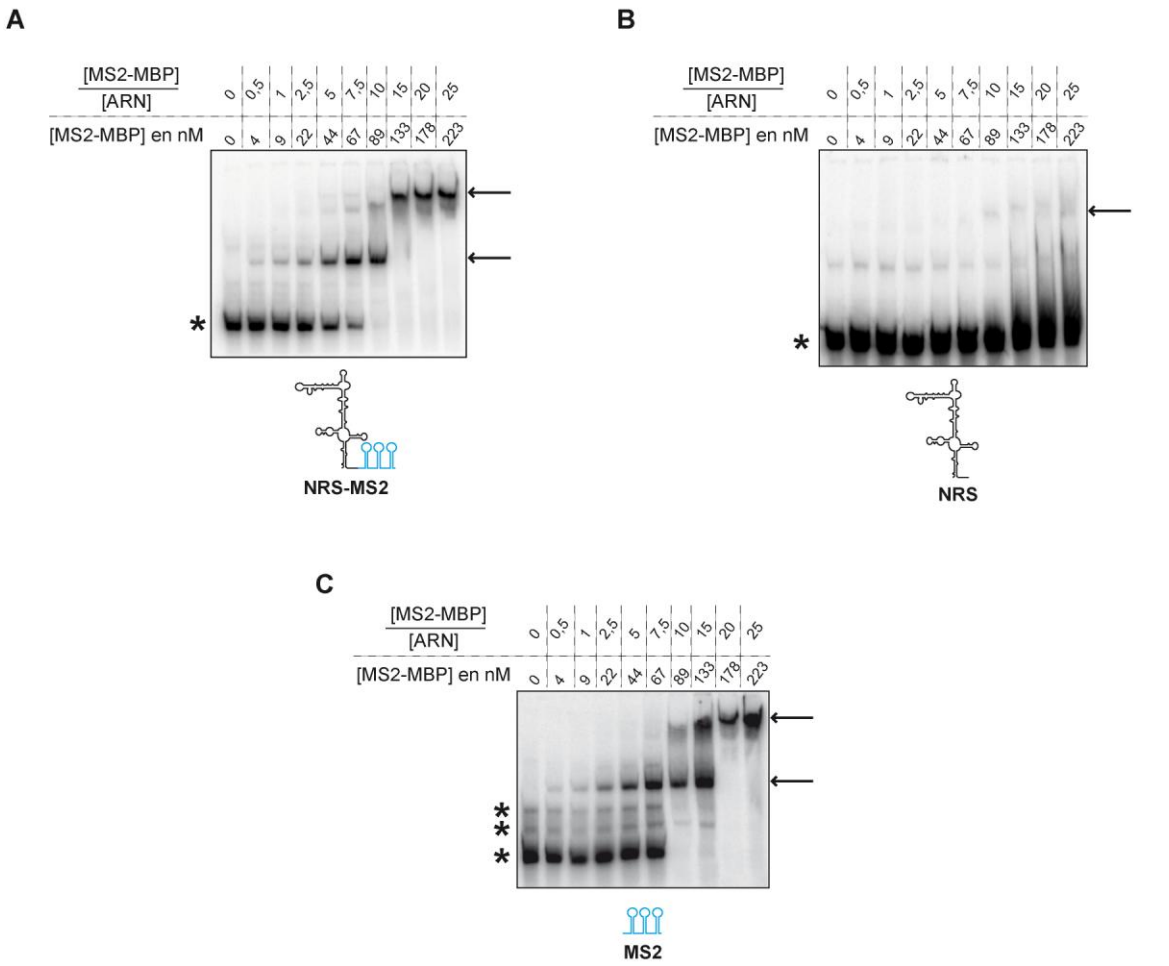



**Figure VIII-17 : Analyse par expériences de retard sur gel des complexes ribonucléoprotéiques formés après incubation des ARN NRS-MS2 (A), NRS (B) et MS2 (C) en présence de protéine MS2-MBP purifiée.**

Pour chacun des ARN, un mélange de transcrits non radioactifs (100 fmol) et uniformément marqués (1,5 fmol) a été incubé en présence de quantités croissantes de protéine MS2-MBP purifiée. Les positions des ARN libres sont indiqués par une étoile et les ARN complexés sont indiqués par une flèche. Les ARN sont schématisés en dessous des autoradiogrammes et leur nom est indiqué. La séquence NRS est représentée en noir, la séquence MS2 en bleu.

### IV.3.2. Définition des quantités d'extrait nucléaire à utiliser par expériences de retard sur gel

Nous avons ensuite réalisé des expériences de retard sur gel afin de définir les quantités d'extrait nucléaire à utiliser pour former des complexes avec l'ARN NRS-MS2. L'ARN NRS non fusionné a été utilisé comme témoin.

Comme précédemment, ces expériences ont été réalisées avec un mélange de transcrits non radioactifs (100 fmol) et uniformément marqués (1,5 fmol) et en présence d'ARNt compétiteurs (10 pmol). Les ARN ont été incubés avec de quantités croissantes d'extrait nucléaire de cellules HeLa (0,31 µg à 12,9 µg de protéine totale). Les mélanges réactionnels ont ensuite été analysés par une électrophorèse en gel non dénaturant (**Figure VIII-18**). Le même lot d'extrait nucléaire a été utilisé pour toutes ces expériences ainsi que celles réalisées ultérieurement.

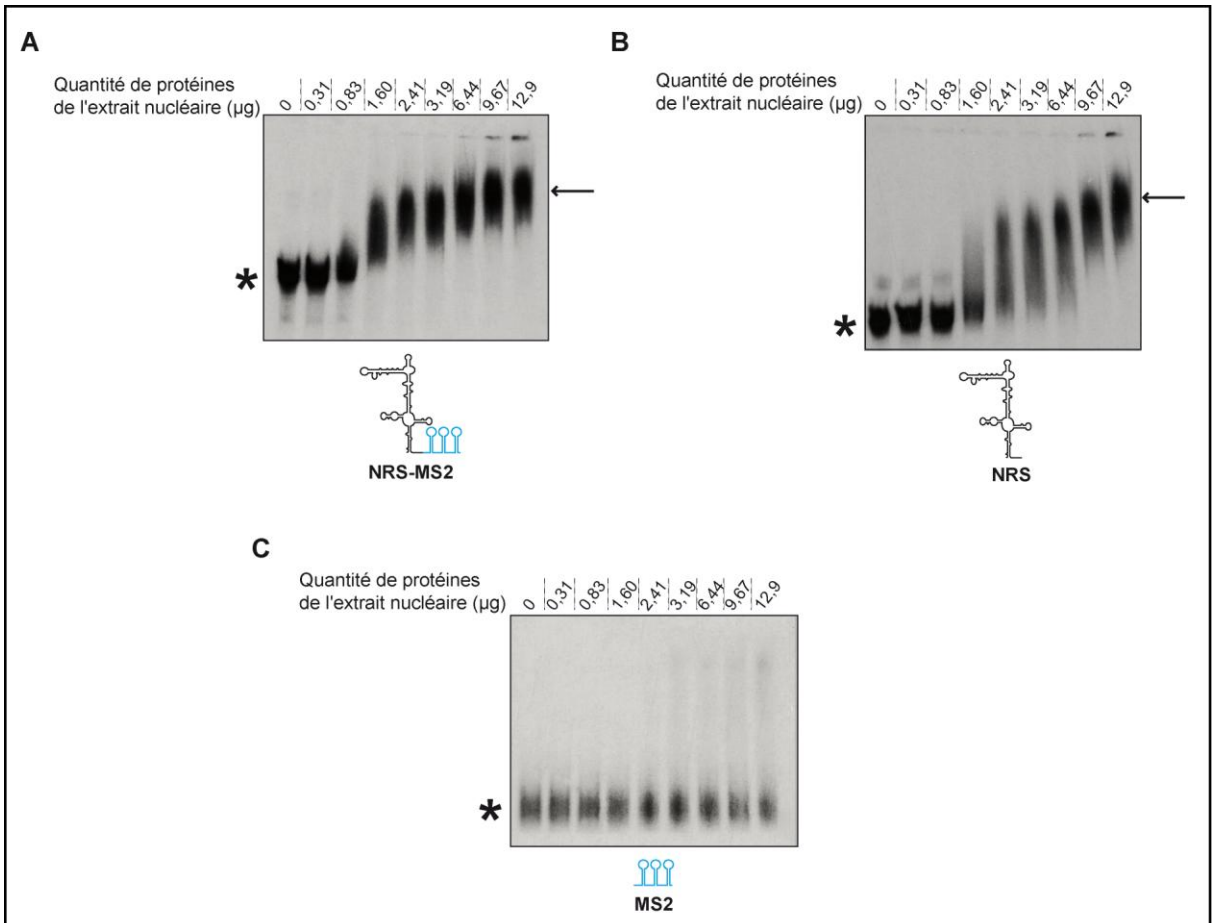



**Figure VIII-18 : Analyse par expériences de retard sur gel des complexes ribonucléoprotéiques formés après incubation des ARN NRS-MS2 (A), NRS (B) et MS2 (C) en présence d'extrait nucléaire de cellules HeLa.**
Pour chacun des ARN, un mélange de transcrits non radioactifs (100 fmol) et uniformément marqués (1,5 fmol) a été incubé en présence de quantités croissantes d'extrait nucléaire. Les positions des ARN libres sont indiquées par une étoile et les ARN complexés sont indiqués par une flèche. Les ARN sont schématisés en dessous des autoradiogrammes et leur nom est indiqué. La séquence NRS est représentée en noir, la séquence MS2 en bleu.

Comme il est montré dans la **Figure VIII-18**, quelque soit la quantité d'extrait nucléaire utilisé, aucun retard sur gel n'était observé pour l'ARN MS2, tandis que des complexes ribonucléoprotéiques étaient formés sur les ARN NRS MS2 et NRS dès l'ajout d'une quantité d'extrait nucléaire correspondant à environ 1,6 µg de protéines. Par ailleurs, les mobilités électrophorétiques des ARN NRS MS2 et NRS étaient comparables.

Dans son application de cette méthode de chromatographie d'affinité, l'équipe de R. Lührmann utilisait une quantité d'ARN pré-messager de 180 pmol pour 3,6 ml d'extrait nucléaire.

Nous avons testé si des quantités d'ARN de 250 et 500 pmol étaient satisfaisantes sachant que nous étions limités par la quantité d'extrait nucléaire disponible. Nous ne voulions pas utiliser de grandes quantités d'extrait nucléaire, afin de limiter la formation d'interactions aspécifiques des protéines à la fois avec l'ARN et avec la résine. Pour ces raisons, nous avons choisi les conditions où nous observions par retard sur gel un retard électrophorétique de l'ARN NRS MS2 avec la plus petite des quantités d'extrait nucléaire testées, c'est-à-dire lorsque 100 fmol d'ARN NRS MS2 étaient totalement complexés avec une quantité d'extrait nucléaire correspondant à environ 1,6 µg de protéines (**Figure VIII-18**). D'après ces conditions, nous devions donc utiliser 4 et 8 mg de protéine totale d'extrait nucléaire lors des chromatographies d'affinité réalisées avec 250 et 500 pmol d'ARN NRS MS2. Les mêmes conditions devaient également être appliquées à la chromatographie d'affinité témoin réalisée avec l'ARN MS2.

En ce qui concerne les conditions d'assemblage des complexes ribonucléoprotéiques, nous avons utilisé le même tampon que celui qu'utilisait par l'équipe de R. Lührmann pour l'assemblage des complexes spliceosomaux afin d'être dans des conditions utilisées pour réaliser des épissages *in vitro* (KCl 25 mM ; $MgCl_2$ 3 mM ; créatine phosphate 20 mM ; ATP 2 mM ; 30 % d'extrait nucléaire de cellules HeLa (v/v)). En revanche, afin d'optimiser la formation des complexes, nous avons ajusté le volume réactionnel de façon à nous placer à une concentration en ARN de 200 nM, soit une concentration en ARN 13 fois supérieure à celle utilisée par l'équipe de R. Lührmann (15 nM).

### IV.3.3. Etude des interactions aspécifiques de la résine amylose avec les composants de l'extrait nucléaire de cellules HeLa

Nous voulions vérifier que les conditions de mise en œuvre de la technique de chromatographie d'affinité étaient suffisamment stringentes pour limiter au maximum les interactions aspécifiques de la résine avec les composants de l'extrait nucléaire. Pour cela, nous avons réalisé un test en absence d'ARN. Nous nous somme placés dans les conditions que nous voulions utiliser lors de la purification des complexes ribonucléoprotéiques en omettant d'ajouter l'ARN au départ de l'expérience sur la base des données décrites ci-dessus. Pour la purification des complexes ribonucléoprotéiques, nous envisagions de former des

complexes ARN NRS MS2/protéine MS2-MBP avec 500 pmol d'ARN et 10 000 pmol de protéine MS2-MBP (ratio ARN/protéine de 20) puis d'ajouter un volume d'extrait nucléaire contenant 8 mg de protéines. Sachant que la capacité de rétention de la protéine MS2-MBP sur la résine d'amylose est d'environ 3 mg de protéine/ml de résine, soit $5{,}53.10^{-8}$ mol protéine/ml de résine, nous devions utiliser 180 µl de résine par test. Sur cette base, nous avons testé deux conditions : nous avons ajouté une quantité d'extrait nucléaire correspondant à 8 mg de protéines sur 180 µl de résine amylose seule ou sur 180 µl de résine où 10 000 pmol de la protéine MS2-MBP avait été préalablement immobilisée (**Figure VIII-19**).

La résine a tout d'abord été lavée avec du tampon D (EDTA 0,2 mM ; KCl 0,1 M; DTT 0,5 mM; PMSF 0,25 mM; HEPES 20 mM pH 7,9 ; glycérol 20%) en présence de 2,8 mM de $MgCl_2$ et d'ARNt afin de saturer les sites potentiels d'interaction aspécifiques avec les composants de l'extrait nucléaire. La protéine MS2-MBP diluée dans du tampon D a ensuite été incubée avec la résine pendant 1 h à 4°C (seul du tampon D a été ajouté à cette étape pour le test de chromatographie d'affinité réalisé sans la protéine MS2-MBP). Le matériel non retenu (fraction NR) a été éliminé par centrifugation, puis la matrice lavée et centrifugée 3 fois avec du tampon D et les fractions collectées après chaque lavage (fractions L1, L2, L3). La résine a ensuite été incubée durant 30 min avec 4,2 ml de tampon d'assemblage (KCl 25 mM ; $MgCl_2$ 3 mM ; créatine phosphate 20 mM ; ATP 2 mM) contenant 30 % d'extrait nucléaire de cellules HeLa (v/v). Le matériel non retenu sur la résine a été éliminé par une centrifugation (fraction S), la résine lavée et centrifugée 3 fois avec du tampon D, des fractions étant collectées après chaque lavage (fractions L4, L5, L6). La protéine MS2-MBP a ensuite été éluée de la résine par deux ajouts successifs de 100 µl de tampon d'élution contenant du tampon D incluant 2,8 mM de $MgCl_2$ et 15 mM de maltose (fractions E1 et E2). Un volume de 10 µl de chacune des fractions collectées au cours de l'expérience a été déposé sur gel SDS-PAGE et les résultats obtenus pour chacun des tests de chromatographie d'affinité sont présentés dans la **Figure VIII-19**.

Aucune protéine de l'extrait nucléaire n'était observée dans les fractions d'élution, que ce soit lors du test réalisé avec la résine seule ou avec la résine sur laquelle la protéine MS2-MBP avait été préalablement immobilisée. Cela indiquait que les conditions utilisées étaient satisfaisantes et permettaient de limiter les interactions aspécifiques de la résine avec les composants de l'extrait nucléaire de cellules HeLa. Ces conditions ont par conséquent été utilisées par la suite afin de purifier les complexes ribonucléoprotéiques formés sur l'ARN NRS MS2.

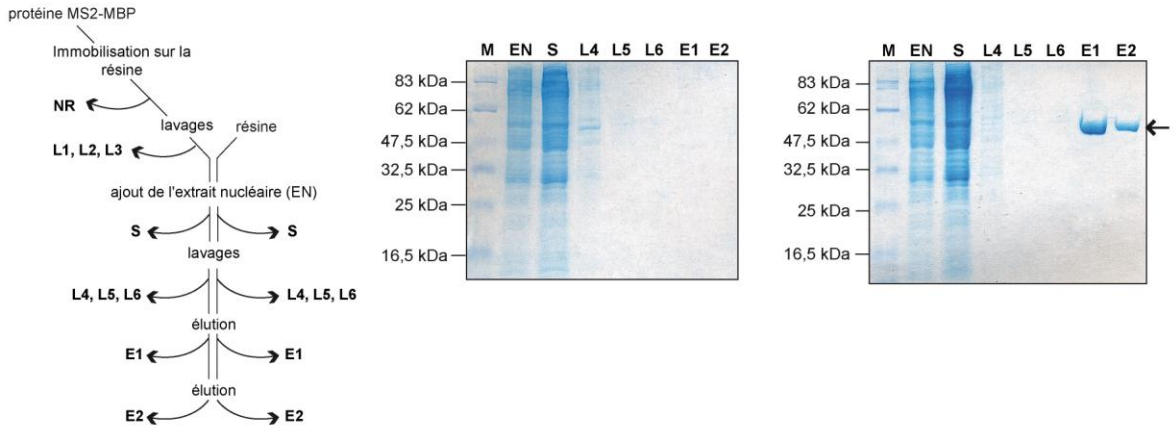



**Figure VIII-19 : Etude de la fixation aspecifique possible des protéines de l'extrait nucléaire de cellules HeLa sur la résine amylose.**

**A :** Etapes des tests réalisés pour tester la fixation de protéines lorsque l'extrait nucléaire était ajouté sur la résine seule (B) ou la résine sur laquelle la protéine MS2-MBP avait été préalablement fixée (C). Les différentes étapes et le nom des fractions collectées sont mentionnés. La fraction NR correspond à la protéine MS2-MBP non retenue sur la résine, L1, L2, L3, aux 3 lavages de la résine, S, au matériel non retenu après l'ajout de l'extrait nucléaire, L4, L5, L6 aux 3 lavages de la résine et E1 et E2 aux 2 élutions successives. Seules les protéines des fractions récupérées après l'ajout de l'extrait nucléaire ont été analysées par SDS-PAGE.

Le volume déposé des fractions S, L4, L5, L6 sur le gel correspond à 0,5 % de la totalité de chacun des échantillons. Le volume des fractions E1 et E2 déposé sur le gel correspond à 10 % de la totalité de chaque échantillon. La piste EN correspond à une quantité de 7 µg de protéines de l'extrait nucléaire de cellules HeLa utilisé lors des expériences. La protéine MS2-MBP est indiquée par une flèche.

**B :** Analyse par SDS-PAGE des différentes fractions collectées après l'ajout de l'extrait nucléaire de cellules HeLa sur la résine seule.

**C :** Analyse par SDS-PAGE des différentes fractions collectées après l'ajout de l'extrait nucléaire de cellules HeLa sur la résine où la protéine MS2-MBP a été préalablement immobilisée.

## IV.4. Expérience test réalisée avec 500 pmol d'ARN et un rapport molaire [protéine MS2 MBP]/[ARN] de 20

### IV.4.1. Purification des complexes ribonucléoprotéiques

La première purification de complexes ribonucléoprotéiques formés sur l'ARN NRS MS2 a été réalisée en utilisant 500 pmol d'ARN non radioactif et 10 pmol d'ARN uniformément marqué, ceci afin de suivre le devenir de l'ARN durant la chromatographie. Nous avons également utilisé 10 000 pmol de protéine MS2-MBP afin de nous placer à un rapport molaire [protéine MS2-MBP]/[ARN NRS MS2] de 20. Le volume d'extrait nucléaire utilisé correspondait à environ 8 mg de protéines. Une seconde chromatographie d'affinité servant de contrôle a été réalisée en parallèle avec l'ARN MS2.

Après une étape de dénaturation/renaturation dans du tampon D dans lequel nous avions ajouté 2,8 mM de $MgCl_2$, l'ARN a été incubé avec la protéine MS2-MBP durant 30 minutes à 4°C. Une fois les complexes ARN/protéine MS2-MBP formés, ceux-ci ont été incubés durant 30 min dans un volume de 2,7 ml de tampon d'assemblage (KCl 25 mM ; $MgCl_2$ 3 mM ; créatine phosphate 20 mM ; ATP 2 mM) contenant 30 % d'extrait nucléaire de cellules HeLa (v/v) ainsi que des ARNt (185 µM). Les complexes ribonucléoprotéiques ainsi formés ont ensuite été incubés pendant 1 h avec les billes d'amylose, préalablement lavées 3 fois avec du tampon D contenant 2,8 mM de $MgCl_2$. Le matériel non retenu sur les billes a été éliminé par une brève centrifugation (fraction S), puis les billes ont été lavées trois fois avec du tampon D (fractions L4, L5, L6). Les complexes ribonucléoprotéiques ont ensuite été élués par l'ajout du tampon d'élution contenant 15 mM de maltose. Les différentes étapes sont présentées dans la **Figure VIII-20**.

A chacune des étapes, un aliquot a été prélevé afin de suivre le devenir de l'ARN, la formation des complexes et leur rétention sur les billes d'amylose. Les ARN contenus dans les différents échantillons ont été analysés par électrophorèse en gel de polyacrylamide non dénaturant suivie d'une autoradiographie. Les protéines ont été analysées en parallèle par SDS-PAGE suivie d'une coloration du gel au bleu de coomassie. Les résultats obtenus sont présentés dans la **Figure VIII-21**.

Comme attendu, les ARN MS2 et NRS MS2 étaient totalement complexés avec la protéine MS2 MBP (**Figure VIII-21, B1 et C1**, respectivement) et après l'ajout de l'extrait nucléaire, la grande majorité des complexes ribonucléoprotéiques alors formés étaient retenus sur les billes (**Figure VIII-21, B2 et C2, piste S**). La comparaison du contenu en ARN et en protéines des fractions L4, L5 et L6 indique que les ARN avaient bien été immobilisés sur les billes pendant les lavages, tandis que les protéines interagissant avec peu ou pas de spécificité avec les billes ou les complexes ribonucléoprotéiques avaient été éliminées. Aucune protéine n'était observée dans la fraction L6 ce qui suggère que seuls des complexes spécifiques étaient immobilisés sur les billes avant l'étape d'élution.

Le profil électrophorétique de la fraction d'élution des complexes ribonucléoprotéiques formés sur l'ARN MS2 contrôle était différent, avec beaucoup moins de bandes de protéines que celui de l'ARN NRS MS2 (**Figure VIII-21, B3 et C3**). Nous en avons conclu que les complexes formés étaient spécifiques de chacun des ARN et donc que nous pouvions utiliser ces conditions pour rechercher les partenaires de l'élément NRS.

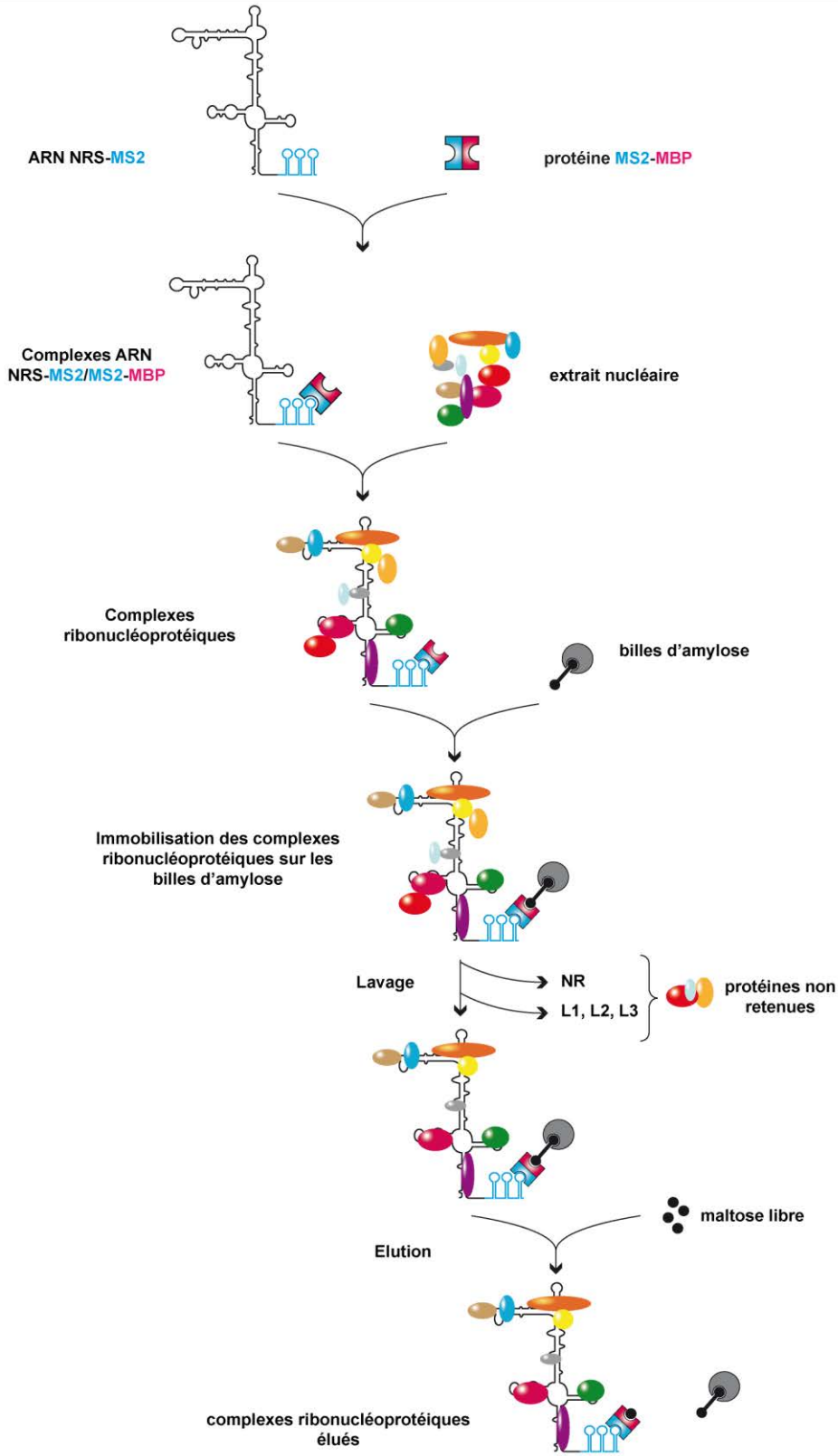



**Figure VIII-20 : Principe de la purification des complexes ribonucléoprotéiques par chromatographie d'affinité reposant sur l'emploi de la protéine de l'enveloppe du phage MS2.**
L'ARN NRS-MS2 est incubé avec la protéine de fusion MS2-MBP pour former le complexe ARN/protéine NRS-MS2/MS2-MBP. Ce complexe est ensuite incubé avec de l'extrait nucléaire pour former des complexes ribonucléoprotéiques qui sont ensuite spécifiquement retenus sur les billes d'amylose. Après plusieurs lavages, l'ajout de maltose libre permet l'élution des complexes. Le même protocole est appliqué lorsque l'ARN MS2 est utilisé.

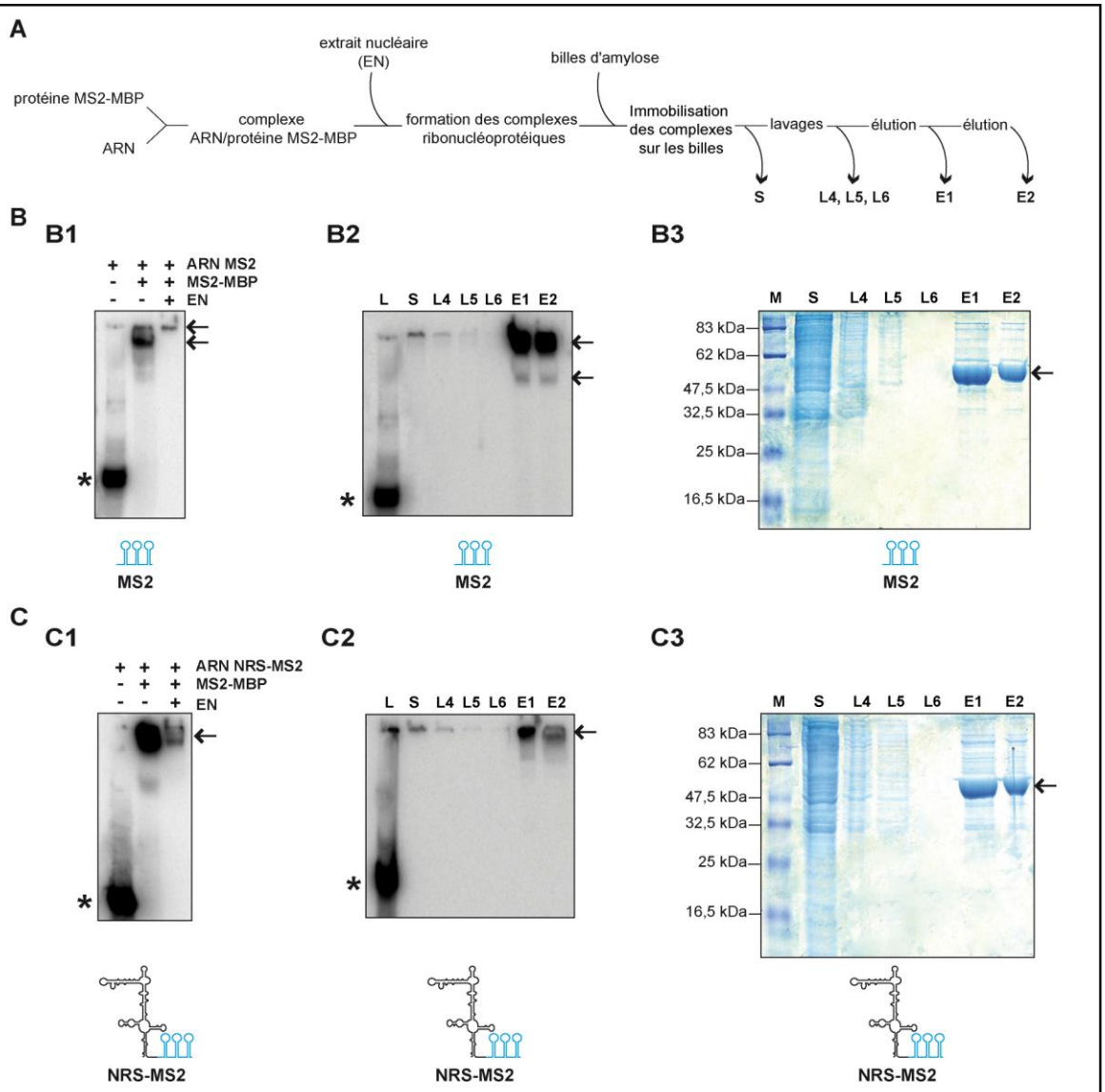



**Figure VIII-21 : Purification des complexes ribonucléoprotéiques formés sur les ARN MS2 (B) et NRS-MS2 (C) à un rapport molaire [protéine MS2-MBP]/[ARN] de 20.**

**A :** Récapitulatif des différentes étapes de la chromatographie d'affinité. Les expériences ont été réalisées avec un mélange de 500 pmol d'ARN non radioactif et de 10 pmol d'ARN uniformément marqué. Les ARN ont tout d'abord été incubés avec 10 000 pmol de protéine MS2-MBP de façon à avoir un rapport molaire [protéine]/[ARN] de 20, puis incubés dans 2,7 ml d'extrait nucléaire dilué 3 fois dans du tampon d'assemblage. L'ensemble a été incubé avec les billes d'amylose, puis après 3 lavages, les complexes ont été élués par ajout de maltose. La fraction S correspond au matériel non retenu sur les billes après l'ajout de l'extrait nucléaire, L4, L5, L6 aux 3 lavages des billes et E1 et E2 aux 2 élutions successives.

**B :** Purification des complexes ribonucléoprotéiques formés sur l'ARN MS2. Analyse par électrophorèse en gel de polyacrylamide non dénaturant de l'ARN MS2 dans les différents échantillons collectés avant (**B1**) et après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose (**B2**). Une quantité de 10 fmol d'ARN libre, ainsi qu'un volume correspondant à 1 % des volumes d'assemblage des différents complexes a été déposé sur un gel de polyacrylamide natif (**B1**). Une quantité de 10 fmol d'ARN libre (piste L), ainsi qu'un volume correspondant à 2,5% du volume des fractions S, L4, L5, L6 et à 5 % du volume des fraction E1 et E2 ont été déposés sur un gel de polyacrylamide non dénaturant (**B2**). La position des ARN a été révélée par autoradiographie, le retard sur gel des ARN complexés est indiqué par une flèche, la position de l'ARN libre est indiqué par une étoile. **B3** : Analyse par SDS-PAGE du contenu en protéines des différentes fractions collectées après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose. Un volume correspondant à 0,5% du volume des fractions S, L4, L5, L6 et à 5 % du volume des fractions E1 et E2 a été déposé. La protéine MS2-MBP est indiquée par une flèche.

**C :** Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS-MS2. La légende des encarts C1, C2 et C3 est respectivement identique à celle des encarts B1, B2, et B3.

### IV.4.2. Analyse par spectrométrie de masse des protéines contenues dans les complexes ribonucléoprotéiques purifiés

Nous avons ensuite cherché à identifier les protéines contenues dans les complexes ribonucléoprotéiques. Pour cela, nous avons utilisé la technique de spectrométrie de masse en collaboration avec S. Sanglier et L. Fouillen de l'équipe d'A. Van Dorsselaer (Laboratoire LSMBO, Strasbourg). Nous avons reproduit l'expérience exactement dans les mêmes conditions, mais sans utiliser d'ARN uniformément marqué et nous avons déposé un quart des éluats sur un gel SDS-PAGE qui a ensuite été coloré au bleu de coomassie colloïdal (**Figure VIII-22**). Les bandes de protéines visibles à la fois pour les complexes ribonucléoprotéiques assemblés sur l'ARN MS2 témoin et sur l'ARN NRS-MS2 ont été découpées. L'équipe d'A. Van Dorsselaer a réalisé l'hydrolyse des protéines par la trypsine, identifié la masse de certains des peptides libérés par MALDI-TOF et complété l'analyse par nanoLC/MS/MS.

Douze et onze bandes ont été respectivement découpées pour les complexes formés avec l'ARN NRS-MS2 et l'ARN MS2 (**Figure VIII-22**). Seize protéines ont été identifiées dans les complexes formés avec l'ARN contrôle MS2 tandis que 25 protéines ont été caractérisées dans les complexes formés avec l'ARN NRS-MS2 (**Figure VIII-22 et Tableau VIII-1**).

Dans cette expérience qui n'était pas une analyses systématique de l'ensemble des pistes du gel et donc dont les résultats dépendaient du découpage des bandes, 7 protéines n'étaient trouvées qu'avec l'ARN MS2, 9 étaient communes aux complexes formés avec chacun des deux ARN et 16 n'étaient retrouvées qu'avec l'ARN NRS MS2 (**Figure VIII-22, VIII-22 et Tableau VIII-1**).

Seize protéines ont été identifiées dans les complexes assemblés sur l'ARN NRS-MS2 mais pas détectées dans ceux formés avec l'ARN contrôle MS2 (**Figure VIII-23, Tableau VIII-1**). Toutes ces protéines possèdent un domaine de liaison aux acides nucléiques, ce qui suggère qu'elles sont toutes des partenaires potentiels de l'élément NRS. Parmi ces protéines, les protéines SR ASF et 9G8 étaient attendues car elles ont été décrites comme interagissant avec l'élément NRS (Fogel et McNally, 2000 ; McNally et McNally, 1996). Ces protéines sont difficiles à identifier par spectrométrie de masse nanoLC/MS/MS du fait de leur séquence, elles possèdent peu de sites de coupure par la trypsine ou possèdent des sites qui génèrent des peptides de petite masse qui sont difficilement détectés par spectrométrie de masse. Bien que

l'analyse nanoLC/MS/MS ne soit pas quantitative, leur identification était un argument en faveur de la spécificité des complexes ribonucléoprotéiques que nous avions purifiés.

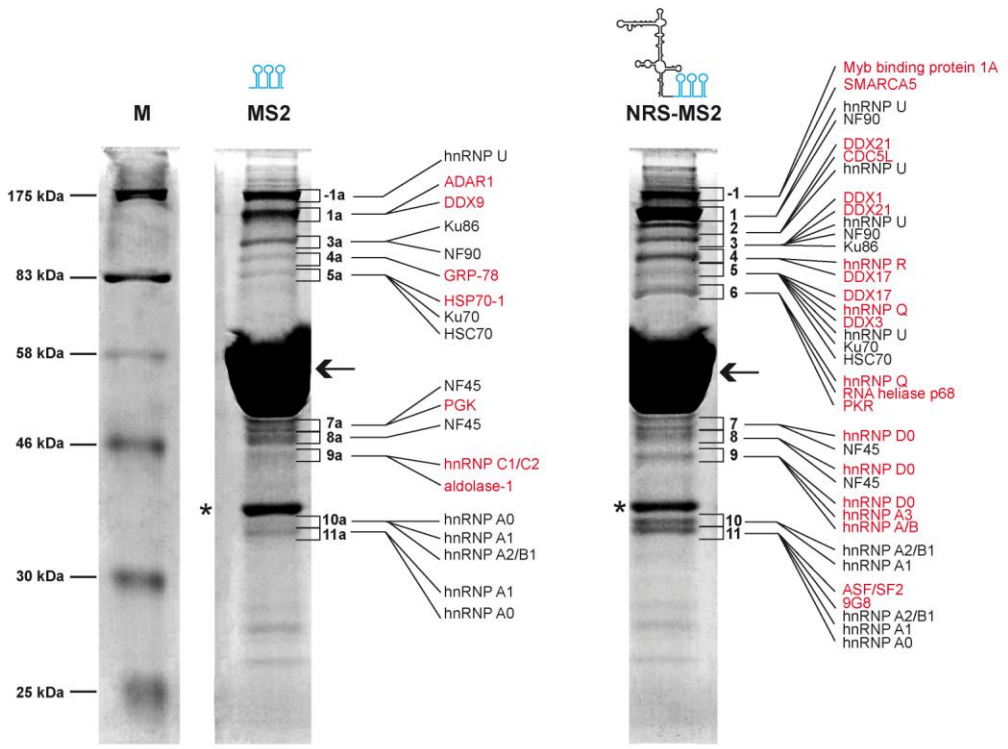
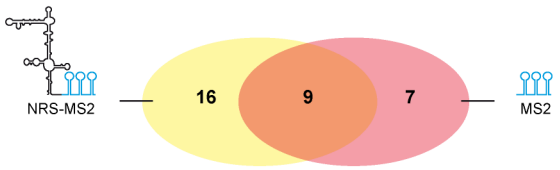



**Figure VIII-22 : Analyse par SDS-PAGE et spectrométrie de masse des protéines contenues dans les complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 à un rapport molaire [protéine MS2-MBP]/[ARN] de 20.**

Un quart de chacun des éluats des expériences de chromatographie d'affinité réalisées avec l'ARN MS2 et l'ARN NRS-MS2 a été déposé sur un gel SDS-PAGE 10%. Après l'électrophorèse, les protéines ont été révélées par une coloration avec du bleu de coomassie colloïdal et les bandes des protéines visibles ont été découpées et analysées par spectrométrie de masse. Les bandes découpées sont numérotées et indiquées par des crochets noirs. La piste de gauche désigne le marqueur de taille (M), la protéine MS2-MBP est indiquée par une flèche, une protéine contaminante co-purifiée avec la protéine MS2-MBP est indiquée par une étoile. Les protéines identifiées par spectrométrie de masse sont indiquées à droite des bandes qui ont été analysées. Les protéines identifiées uniquement dans les complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 sont indiquées en rouge et celles qui leur sont communes sont indiquées en noir.

**A**




| Protéines trouvées uniquement avec l'ARN NRS-MS2 | Protéines communes aux ARN NRS-MS2 et MS2 | Protéines trouvées uniquement avec l'ARN MS2 |
|---|---|---|
| ASF | hnRNP U | hnRNP C1/C2 |
| 9G8 | hnRNP A0 | ADAR1 |
| DDX17 | hnRNP A1 | DDX9 |
| RNA helicase p68 (DDX5) | hnRNP A2/B1 | GRP-78 |
| DDX21 | Ku86 | PGK |
| hnRNP A3 | NF90 | aldolase-1 |
| hnRNP D0 | Ku70 | HSP70-1 |
| hnRNP A/B | HSC70 | |
| hnRNP R | NF45 | |
| hnRNP Q | | |
| CDC5L | | |
| DDX3 | | |
| DDX1 | | |
| Myb binding protéine 1A | | |
| PKR | | |
| SMARCA5 | | |

**B**

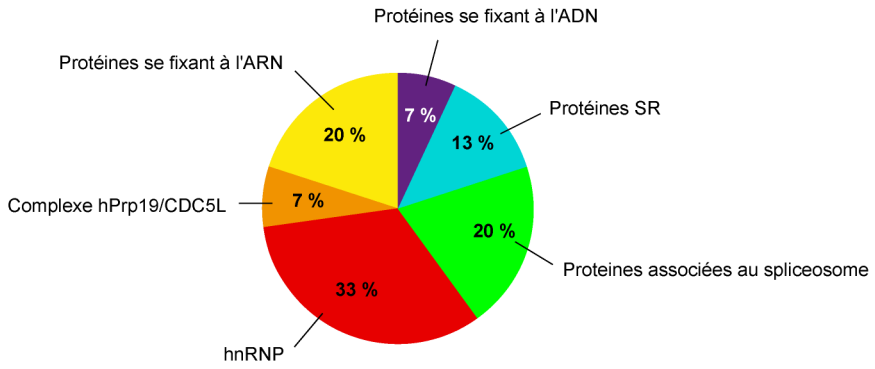



**Figure VIII-23 : Protéines des complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS MS2 lors de la première expérience de chromatographie d'affinité.**
**A :** Représentation schématique et liste des protéines identifiées pour chacun des ARN NRS MS2 et MS2 lors de l'expérience de chromatographie d'affinité. Le nombre de protéines totales identifiées pour chaque ARN est indiqué par un chiffre dans un cercle de couleur. Le nombre des protéines communes aux deux ARN est inscrit dans la zone de recouvrement des cercles.
**B :** Représentation en pourcentage des différentes fonctions des protéines des complexes ribonucléoprotéiques formés sur l'ARN NRS MS2.

Les complexes assemblés sur l'ARN NRS-MS2 contenaient également un grand nombre de protéines hnRNP ce qui était attendu car ces protéines sont abondantes dans les extraits nucléaires de cellules HeLa et se fixent sur la plupart des pré-ARNm (Bennett *et al.*, 1992 ; Jurica et Moore, 2002). Jusqu'à présent, la seule protéine hnRNP ayant été décrite comme interagissant avec l'élément NRS est la hnRNP H (Fogel et McNally, 2000). Les autres protéines hnRNP identifiées ici pouvaient donc être de nouveaux partenaires de l'élément NRS, voire potentiellement des antagonistes des protéines SR ASF et 9G8 comme c'est souvent le cas lors de la régulation de l'épissage alternatif (Crawford et Patton, 2006 ; Eperon *et al.*, 2000 ; Saliou *et al.*, 2009).

Nous avons noté la présence de deux protéines connues pour intérvenir dans l'assemblage du spliceosome, les hélicases DDX5 (ou RNA helicase p68) et DDX17. Ces deux protéines paralogues sont connues pour se fixer sur l'ARN simple ou double brin et sont essentielles à la réaction d'épissage (Janknecht, 2010 ; Liu, 2002). Elles participent à la dissociation de la snRNP U1 du site donneur (Lin *et al.*, 2005). Par ailleurs, récemment, l'implication de l'hélicase p68 dans la régulation de l'épissage de l'exon 10 du gène Tau a été démontrée. Le site donneur de ce gène se situe dans une structure tige boucle sur laquelle l'activité de l'hélicase p68 favoriserait l'appariement du snRNA U1 (Kar *et al.*, 2011). Les hélicases p68 et DXX17 sont donc de très bonnes candidates pour interagir avec le pseudo site de l'élément NRS et pourraient jouer un rôle dans l'appariement du snRNA U1 avec celui-ci.

Ces résultats préliminaires étaient donc encourageants, cependant, nous avons observé sur les profils électrophorétiques que la protéine MS2-MBP occupait beaucoup d'espace sur le gel SDS-PAGE (**Figure VIII-22**). Cela masquait la présence des protéines des complexes ayant des mobilités électrophorétiques voisines. C'est la raison pour laquelle nous avons décidé de refaire les expériences de chromatographie d'affinité en diminuant le rapport molaire ARN/MS2-MBP de 1/20 à 1/5.

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | N° bande | MS2 | | NRS-MS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Nbr peptides | Couv | Nbr peptides | Couv |
| **Proteine SR** | | | | | | | | | | | |
| | **ASF** | SRp30a, Splicing factor arginine/serine rich 1 | Q07955 | 27,73 | 2 RRM, RS | 10.37 | 11 | | | 6 | 27% |
| | **9G8** | Splicing factor arginine/serine rich 7 | Q16629 | 27,35 | RRM, ZnF | | 11 | | | 6 | 25% |
| **Protéines associées au spliceosome** | | | | | | | | | | | |
| | **DDX17** | DEAD box protein 17, DEAD box protein p72, P72, RH70, RNA-dependent helicase p72 | Q92841 | 72,33 | DEXDc, HELIC | 8.82 | 4, 5 | | | 6, 9 | 10%, 16% |
| | **RNA helicase p68** | ATP-dependent RNA helicase DDX5 | P17844 | 69,11 | DEXDc, HELIC | | 6 | | | 11 | 18% |
| | **DDX21** | Nucleolar RNA helicase 2, Gu-alpha, GURDB, Nucleolar RNA helicase Gu, Nucleolar RNA helicase II, RH II/Gu, RH-II/GU, RH-II/GuA | Q9NR30 | 87,29 | DEXDc, HELIC | 9.32 | 2, 3 | | | 25, 4 | 38%, 7% |
| **hnRNP** | | | | | | | | | | | |
| | **hnRNP A0** | | Q13151 | 30,82 | 2 RRM | | 10a, 11a, 11 | 4 | 18% | 6, 7 | 23%, 23% |
| | **hnRNP A1** | Helix destabilizing protein | P09651 | 38,69 | 2 RRM | | 10a, 11a,10, 11 | 8, 9 | 21%, 28% | 11, 12 | 22%, 22% |
| | **hnRNP A2/B1** | | P22626 | 37,41 | 2 RRM | | 10a,10, 11 | 10 | 25% | 14, 7 | 37%, 15% |
| | **hnRNP U** | SAF-A | Q00839 | 90,29 | SPRY, SAP | | 1a, 1, 2, 3, 5 | 5 | 7% | 18, 5, 3, 3 | 22%, 8%, 6%, 4% |
| | **hnRNP C1/C2** | hnRNP C | P07910 | 33,65 | RRM | 4.95 | 9a | 4 | 15% | | |
| | **hnRNP A3** | FBRNP | P51991 | 39,57 | 2 RRM | 9.10 | 9 | | | 5 | 15% |
| | **hnRNP D0** | AUF1, AUF1A, AU-rich element RNA-binding protein 1 | Q14103 | 38,41 | 2 RRM | 7.62 | 7, 8, 9 | | | 3, 4, 6 | 9%, 12%, 14% |
| | **hnRNP A/B** | | Q99729 | 36,59 | 2 RRM | | 9 | | | 7 | 21% |
| | **hnRNP R** | | O43390 | 70,90 | 3 RRM | | 4 | | | 25 | 33% |
| | **hnRNP Q** | SYNCRIP, NSAP1 | O60506 | 69,56 | 3 RRM | | 5, 6 | | | 14, 19 | 22%, 27% |
| **Complexe hPrp19/CDC5L** | | | | | | | | | | | |
| | **CDC5L** | KIAA0432, Cell division cycle 5-like protein | Q99459 | 92,19 | 2 SANT | | 2 | | | 6 | 9% |
| **EJC** | | | | | | | | | | | |
| | **DDX3** | DDX14, DDX3X, DBX, Helicase-like protein 2, HLP2 | O00571 | 73,20 | DEXDc, HELIC | 6.73 | 5 | | | 6 | 12% |
| **Protéines se fixant à l'ARN** | | | | | | | | | | | |
| | **NF45** | Interleukin enhancer-binding factor 2, ILF2 | Q12905 | 43,04 | | | 7a, 8a, 7, 8 | 19, 5 | 48%, 15% | 20, 4 | 61%, 14% |
| | **NF90** | Interleukin enhancer-binding factor 3, ILF3, DRBP76 | Q12906 | 95,28 | 2 DSRM, DZF | | 3a, 1, 3 | 24 | 27% | 6, 19 | 8%, 19% |
| | **ADAR1** | Double-stranded RNA-specific adenosine deaminase | P55265 | 135,91 | 3 DSRM | | 1a | 18 | 17% | | |
| | **DDX9** | ATP-dependent RNA helicase A | Q08211 | 140,79 | 2 DSRM, DEXDc, HELIC | | 1a | 8 | 8% | | |
| | **DDX1** | ATP-dependent RNA helicase DDX1, DEAD box protein retinoblastoma, DBP-RB | Q92499 | 82,38 | SPRY, DEXDc, HELIC | 6.81 | 3 | | | 6 | 12% |
| | **Myb binding protein 1A** | P160, PAP2 | Q9BQG0 | 148,76 | | 9.34 | -1 | | | 3 | 2% |
| | **PKR** | p68 kinase, Interferon-induced, double-stranded RNA-activated protein kinase | P19525 | 62,06 | 2 DSRM, STY kinase | | 6 | | | 3 | 6% |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | N° bande | MS2 | | NRS-MS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Nbr peptides | Couv | Nbr peptides | Couv |
| **Protéines se fixant à l'ADN** | | | | | | | | | | | |
| | **Ku 70** | ATP dependent DNA helicase 2 subunit 1, 70 kDa subunit of Ku antigen, CTC75, Lupus Ku autoantigen protein p70 | P12956 | 69,80 | SAP, Ku78 | 6.23 | 5a, 5 | 19 | 31% | 19 | 38% |
| | **Ku 86** | Ku 80, ATP dependent DNA helicase 2 subunit 2, 86 kDa subunit of Ku antigen, Lupus Ku autoantigen protein p86 | P13010 | 82,65 | Ku78, VWA | 5.55 | 3a, 3 | 20 | 32% | 11% | 20% |
| | **SMARCA5** | SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 5, hSNF2H, ISWI | O60264 | 121,83 | DEXDc, HELIC, 2 SANT | | -1 | | | 5 | 4% |
| **Autres** | | | | | | | | | | | |
| | **HSC70** | Heat shock cognate 71 kDa protein, HSPA8 | P11142 | 70,85 | | | 5a, 5 | 15 | 22% | 9 | 17% |
| | **GRP-78** | 78 kDa glucose-regulated protein precursor, Heat shock 70 kDa protein 5 | P11021 | 72,29 | | | 4a | 7 | 14% | | |
| | **PGK** | Phosphoglycerate kinase 1 | P00558 | 44,46 | PGK | | 7a | 10 | 24% | | |
| | **Aldolase 1** | Fructose-bisphosphate aldolase A | P04075 | 39,26 | | | 9a | 7 | 18% | | |
| | **HSP70-1** | Heat shock 70 kDa protein 1 | P08107 | 70,01 | | | 5a | 8 | 14% | | |

**Tableau VIII-1 : Liste des protéines identifiées par spectrométrie de masse dans les complexes ribonucléoprotéiques assemblés sur les ARN MS2 et NRS-MS2 lors de la première expérience de chromatographie d'affinité.**

Les protéines sont répertoriées selon leur fonction cellulaire. Le nom alternatif, le numéro d'accession (N°AC), le poids moléculaire (PM), le point isoélectrique (pI) et numéro de chaque bande du gel SDS-PAGE dans laquelle la protéine a été identifiée sont indiqués. Pour chaque protéine, le pourcentage de recouvrement de la séquence protéique (Couv) et le nombre de peptides identifiés sont reportés.

### IV.5. Expérience réalisée avec 250 pmol d'ARN et un rapport molaire [protéine MS2 MBP]/[ARN] de 5

Nous avons donc refait les expériences dans les mêmes conditions mais en se plaçant à un rapport molaire [protéine MS2 MBP]/[ARN] de 5 (250 pmol d'ARN et 1250 pmol de protéine MS2-MBP) au lieu de 20 comme précédemment et le volume réactionnel ainsi que la quantité d'extrait nucléaire ont été ajustés en conséquence (1,23 ml et une quantité d'extrait nucléaire correspondant à 4 mg de protéines). A chacune des étapes, un aliquot a été prélevé, les ARN et les protéines de ces échantillons ont été analysés et les résultats obtenus sont présentés dans la **Figure VIII-24**.

Les ARN MS2 et NRS MS2 étaient totalement complexés avec la protéine MS2 MBP (**Figure VIII-24, B1 et C1**, respectivement) mais tous les sites de fixation de la protéine MS2 MBP n'étaient pas occupés puisque deux complexes de mobilités électrophorétiques différentes étaient observés. Cependant, ces conditions nous ont parues suffisantes pour immobiliser la grande majorité des complexes ribonucléoprotéiques sur les billes (**Figure VIII-24, B2 et C2**). Durant les 3 lavages, les ARN étaient bien restés immobilisés sur les billes (**Figure VIII-24 B2 et C2, pistes L4, L5, L6**), tandis que les protéines interagissant avec peu ou pas de spécificité étaient éliminées (**Figure VIII-24, B3 et C3, pistes L4, L5, L6**). D'après les fractions d'élution, le nombre des protéines des complexes assemblés sur l'ARN MS2 était bien inférieur à celui observé pour l'ARN NRS MS2 (**Figure VIII-24, B3 et C3, pistes E1, E2**). En revanche, le profil électrophorétique et le nombre de protéines observées pour l'ARN NRS MS2 semblait être comparable à ceux obtenus précédemment avec un rapport molaire [protéine MS2 MBP]/[ARN] de 20. Sur la base de ces résultats, nous avons décidé d'identifier les protéines présentes dans ces complexes par spectrométrie de masse comme cela avait été réalisé pour l'expérience précédente.

Dans le cas de l'expérience contrôle, seules les quelques bandes de protéines visibles sur le gel SDS-PAGE ont été découpées. Par contre, pour l'expérience réalisée avec l'ARN NRS-MS2, vu le grand nombre de bandes visibles, l'analyse a presque été réalisée de façon systématique. Ainsi, 30 et 11 bandes ont été respectivement découpées dans les pistes correspondant aux complexes formés avec les ARN NRS-MS2 et MS2 (**Figure VIII-25**). Les analyses par spectrométrie de masse réalisées ont permis d'identifier 34 protéines différentes : 23 protéines étaient uniquement détectées dans les complexes assemblés sur l'ARN NRS-MS2, une seule était uniquement détectée dans ceux formés sur l'ARN MS2 et 11 étaient retrouvées dans les 2 types de complexes (**Figure VIII-26, Tableau VIII-2**).

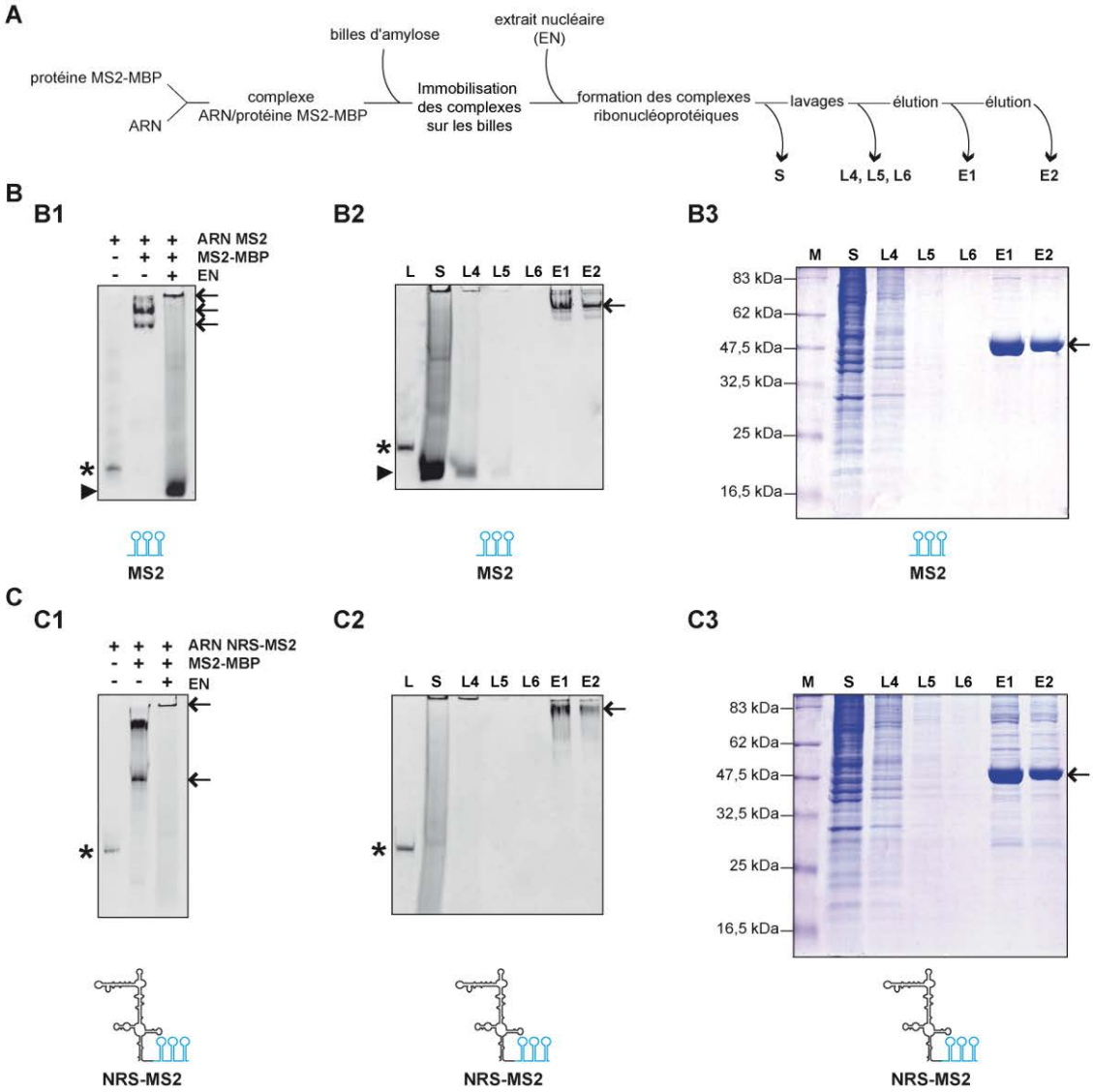



**Figure VIII-24 : Purification des complexes ribonucléoprotéiques formés sur les ARN MS2 (B) et NRS-MS2 (C) à un rapport molaire [protéine MS2-MBP]/[ARN] de 5.**

**A :** Récapitulatif des différentes étapes de la chromatographie d'affinité. Les expériences ont été réalisées avec 250 pmol d'ARN non radioactif qui ont été incubés avec 1250 pmol de protéine MS2-MBP de façon à se placer à un rapport molaire [protéine]/[ARN] de 5. Le mélange a été incubé avec 1,23 ml d'extrait nucléaire dilué 3 fois dans du tampon d'assemblage. L'ensemble a été incubé avec les billes d'amylose, puis après 3 lavages, les complexes ont été élués par ajout de maltose. La fraction S correspond au matériel non retenu sur les billes après l'ajout de l'extrait nucléaire, L4, L5, L6 aux 3 lavages des billes et E1 et E2 aux 2 élutions successives.

**B :** Purification des complexes ribonucléoprotéiques formés sur l'ARN MS2. Analyse par électrophorèse en gel de polyacrylamide non dénaturant de l'ARN MS2 des différents échantillons collectés avant (**B1**) et après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose (**B2**). Une quantité de 100 ng d'ARN libre, ainsi qu'1 % des volumes d'assemblage des différents complexes ont été déposés sur un gel de polyacrylamide non dénaturant (**B1**). Une quantité de 100 ng d'ARN libre (piste L), ainsi que 2,5% du volume des fractions S, L4, L5, L6 et 5 % du volume des fractions E1 et E2 ont été déposés sur un gel de polyacrylamide non dénaturant (**B2**). La position des ARN a été révélée par une coloration du gel au bromure d'éthidium, le retard sur gel des ARN complexés est indiqué par une flèche, la position de l'ARN libre est indiqué par une étoile, celle des ARNt par un triangle. **B3** : Analyse par gel SDS-PAGE du contenu en protéines des différentes fractions collectées après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose. Un volume correspondant à 0,5% du volume des fractions S, L4, L5, L6 et à 5 % du volume des fractions E1 et E2 ont été déposés. La piste M correspond au marqueur de taille, la protéine MS2-MBP est indiquée par une flèche.

**C :** Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS-MS2. La légende des encarts C1, C2 et C3 est respectivement identique à celle des encarts B1, B2, et B3.

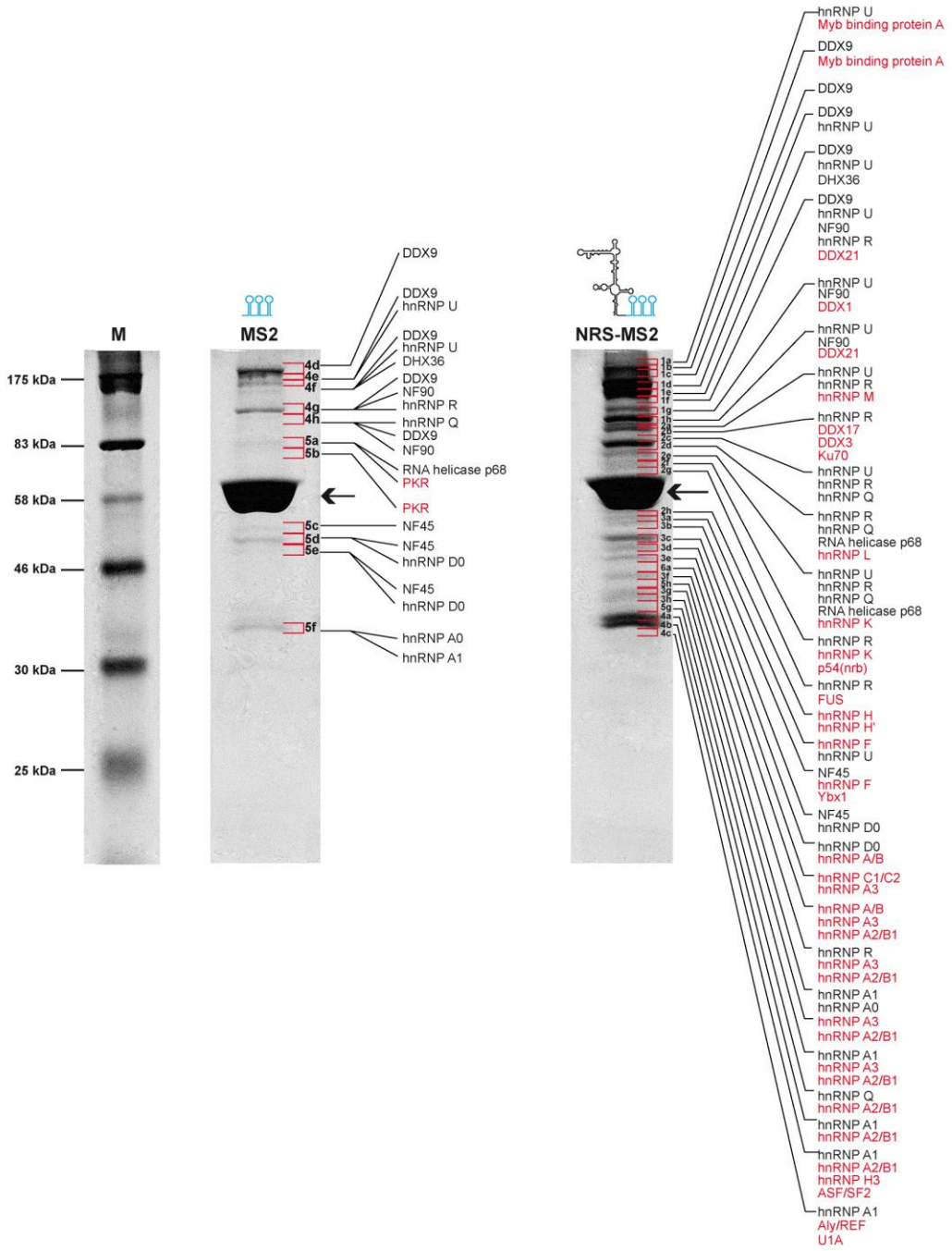



**Figure VIII-25 : Analyse par SDS-PAGE suivie d'une analyse par spectrométrie de masse des protéines contenues dans les complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 à rapport molaire [protéine MS2 MBP]/[ARN] de 5.**

Un quart de chacun des éluats des expériences de chromatographie d'affinité réalisées avec l'ARN MS2 et l'ARN NRS-MS2 a été déposé sur un gel SDS-PAGE 10%. Après l'électrophorèse, les protéines ont été révélées par une coloration avec du bleu de coomassie colloïdal et les bandes des protéines visibles ont été découpées et analysées par spectrométrie de masse. Les bandes découpées sont numérotées et indiquées par des crochets rouges. La piste de gauche désigne le marqueur de taille (M), la protéine MS2-MBP est indiquée par une flèche.

Les protéines identifiées par spectrométrie de masse sont indiquées à droite des bandes qui ont été analysées. Les protéines trouvées uniquement dans les complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 sont indiquées en rouge et celles qui leur sont communes sont indiquées en noir.

**A**




| Protéines trouvées uniquement avec l'ARN NRS-MS2 | | Protéines communes aux ARN NRS -MS2 et MS2 | Protéines trouvées uniquement avec l'ARN MS2 |
|---|---|---|---|
| U1 A | hnRNP C1/C2 | RNA helicase p68 | PKR |
| ASF | hnRNP H' | hnRNP D0 | |
| DDX21 | hnRNP A/B | hnRNP R | |
| p54(nrb) | hnRNP A2/B1 | hnRNP Q | |
| DDX17 | FUS | hnRNP A1 | |
| hnRNP A3 | Ybx1 | hnRNP A0 | |
| hnRNP M | Aly/REF | hnRNP U | |
| hnRNP F | DDX3 | DDX9 | |
| hnRNP L | DDX1 | NF90 | |
| hnRNP H3 | Myb binding protein A | NF45 | |
| hnRNP K | Ku70 | DHX36 | |
| hnRNP H | | | |

**B**

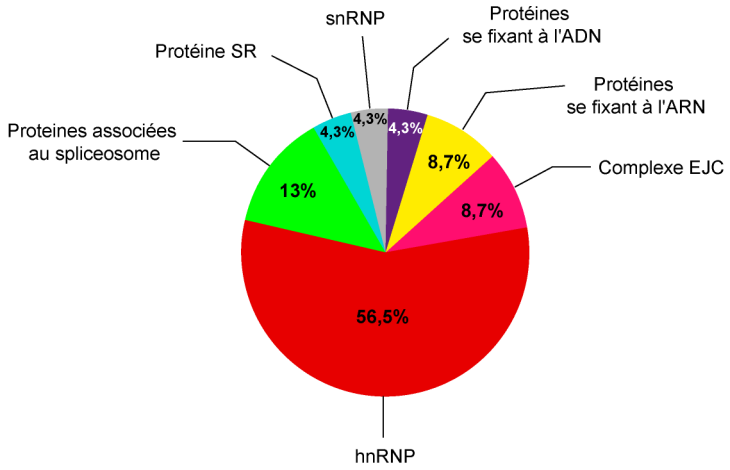



**Figure VIII-26 : Protéines des complexes ribonucléoprotéiques formés sur les ARN MS2 et NRS-MS2 lors de la deuxième expérience de chromatographie d'affinité.**

**A :** Représentation schématique et liste des protéines identifiées pour chacun des ARN NRS-MS2 et MS2 lors de l'expérience de chromatographie d'affinité. Le nombre de protéines totales identifiées pour chaque ARN est indiqué par un chiffre dans un cercle de couleur. Le nombre des protéines communes aux deux ARN est inscrit dans la zone de recouvrement des cercles.

**B :** Représentation en pourcentage des différentes fonctions des protéines des complexes ribonucléoprotéiques formés sur l'ARN NRS-MS2.

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | N°bande | MS2 Nbr peptides | MS2 Couv | NRS-MS2 Nbr peptides | NRS-MS2 Couv |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **snRNP** | | | | | | | | | | | |
| | **U1 A** | U1 small nuclear ribonucleoprotein A , U1 snRNP A | P09012 | 31,26 | 2 RRM | 9.83 | C4 | | | 5 | 11% |
| **Proteine SR** | | | | | | | | | | | |
| | **ASF** | SRp30a, Splicing factor arginine/serine rich 1 | Q07955 | 27,73 | 2 RRM, RS | 10.37 | B4, C4 | | | 4, 4 | 21%, 17% |
| **Protéines associées au spliceosome** | | | | | | | | | | | |
| | **RNA helicase p68** | ATP-dependent RNA helicase DDX5 | P17844 | 69,11 | DEXDc, HELIC | | C2, D2, E2, A5 | 3 | 5% | 10, 12, 4 | 17%, 21%, 6% |
| | **DDX21** | Nucleolar RNA helicase 2, Gu-alpha, GURDB, Nucleolar RNA helicase Gu, Nucleolar RNA helicase II, RH II/Gu, RH-II/GU, RH-II/GuA | Q9NR30 | 87,29 | DEXDc, HELIC | 9.32 | F1, G1, H1 | | | 15, 15, 5 | 26%, 27%, 8% |
| | **p54(nrb)** | NonO protein, Non-POU domain-containing octamer-binding protein, NRB54, P54, NMT55 | Q15233 | 54,20 | 2 RRM | 9.01 | F2 | | | 2 | 6% |
| | **DDX17** | DEAD box protein 17, DEAD box protein p72, P72, RH70, RNA-dependent helicase p72 | Q92841 | 72,33 | DEXDc, HELIC | 8.82 | H1, B2 | | | 11, 11 | 18%, 19% |
| **hnRNP** | | | | | | | | | | | |
| | **hnRNP D0** | AUF1, AUF1A, AU-rich element RNA-binding protein 1 | Q14103 | 38,41 | 2 RRM | 7.62 | C3, D3, E3, A6, D5, E5 | 4, ,3 | 14%, 9% | 8, 7, 6, 6 | 20%, 20%, 17%, 20% |
| | **hnRNP R** | | O43390 | 70,90 | 3 RRM | | H1, A2, B2, C2, D2, E2, F2, G2, H5, G4 | 2, 4 | 3%, 9% | 19, 17, 9, 7, 17, 6, 5, 6 | 3%, 27%, 16%, 12%, 30%, 10%, 9%, 12% |
| | **hnRNP Q** | SYNCRIP, NSAP1 | O60506 | 69,56 | 3 RRM | | B2, C2, D2, E2, H3, G5, H4 | 2, 4 | 4%, 9% | 23, 10, 28, 8, 3 | 37%, 18%, 36%, 15%, 6% |
| | **hnRNP A1** | Helix destabilizing protein | P09651 | 38,69 | 2 RRM | | H5, G3, H3, G5, A4, B4, C4, F5 | 9 | 16% | 5, 6, 6, 7, 10, 19, 6 | 13%, 15%, 19%, 15%, 26%, 34%, 13% |
| | **hnRNP A0** | | Q13151 | 30,82 | 2 RRM | | H3, G5, B4, F5 | 5 | 25% | 2, 4, 4 | 7%, 19%, 19% |
| | **hnRNP U** | SAF-A | Q00839 | 90,29 | SPRY, SAP | | A1, D1, E1, F1, G1, H1, A2, C2, E2, A3, E4, F4 | 6, 3 | 10%, 5% | 3, 6, 19, 8, 7, 3, 3, 3, 6, 3 | 6%, 10%, 22%, 12%, 12%, 6%, 6%, 5%, 8%, 5% |
| | **hnRNP A3** | FBRNP | P51991 | 39,57 | 2 RRM | 9.10 | A6, F3, H5, G3, H3 | | | 6, 7, 7, 7, 8 | 20%, 25%, 23%, 17%, 23% |
| | **hnRNP M** | NAGR1 | P52272 | 77,46 | 3 RRM | 8.84 | A2 | | | 9 | 15% |
| | **hnRNP F** | Nucleolin-like protein mcs94-1 | P52597 | 45,64 | 3 RRM, Znf CHHC | 5.38 | A3, B3 | | | 11, 3 | 21%, 9% |
| | **hnRNP L** | | P14866 | 60,15 | 3 RRM | 6.65 | E2 | | | 9 | 20% |
| | **hnRNP H3** | hnRNP 2H9 | P31942 | 36,90 | 2 RRM | 6.37 | G5, A4, B4 | | | 3, 4, 3 | 14%, 14%, 14% |
| | **hnRNP K** | CSBP | P61978 | 50,94 | 3 KH | 5.39 | E2, F2 | | | 7, 5 | 22%, 15% |
| | **hnRNP H** | hnRNP H1 | P31943 | 49,20 | 3 RRM, Znf CHHC | 5.89 | H2, A3 | | | 11, 7 | 32%, 18% |
| | **hnRNP C1/C2** | hnRNP C | P07910 | 33,65 | RRM | 4.95 | A6 | | | 6 | 16% |
| | **hnRNP H'** | hnRNP H2, FTP3 | P55795 | 49,23 | 3 RRM, Znf CHHC | 5.89 | H2 | | | 8 | 25% |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | N°bande | MS2 Nbr peptides | MS2 Couv | NRS-MS2 Nbr peptides | NRS-MS2 Couv |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **hnRNP (suite)** | | | | | | | | | | | |
| | **hnRNP A/B** | APOBEC1-binding protein 1 | Q99729 | 36,59 | 2 RRM | 9.04 | E3, F3 | | | 9, 6 | 18%, 19% |
| | **hnRNP A2/B1** | | P22626 | 37,41 | 2 RRM | | F3, H5, G3, H3, G5, A4, B4 | | | 4, 5, 9, 10, 6, 12, 5 | 14%, 14%, 21%, 30%, 16%, 32%, 12% |
| | **FUS** | Oncogene TLS, RNA-binding protein FUS | P35637 | 53,39 | RRM, ZnF | | G2 | | | 4 | 7% |
| | **Ybx1** | Nuclease sensitive element-binding protein 1 , CCAAT-binding transcription factor I subunit A, Y-box-binding protein 1 | P67809 | 35,77 | | | B3 | | | 4 | 21% |
| **EJC** | | | | | | | | | | | |
| | **Aly/REF** | THO complex subunit 4, Tho4, Transcriptional coactivator Aly/REF, ALY, BEF | Q86V81 | 26,87 | RRM | 11.15 | C4 | | | 6 | 36% |
| | **DDX3** | DDX14, DDX3X, DBX, Helicase-like protein 2, HLP2 | O00571 | 73,20 | DEXDc, HELIC | 6.73 | B2 | | | 5 | 10% |
| **Protéines se fixant à l'ARN** | | | | | | | | | | | |
| | **DDX9** | ATP-dependent RNA helicase A | Q08211 | 140,79 | 2 DSRM, DEXDc, HELIC | | B1, C1, D1, E1, F1, D4, E4, F4, G4, H4 | 35, 26, 8, 4, 5 | 28%, 25%, 8%, 4%, 6% | 24, 33, 26, 8, 4 | 20%, 29%, 24%, 7%, 4% |
| | **NF90** | Interleukin enhancer-binding factor 3, ILF3, DRBP76 | Q12906 | 95,28 | 2 DSRM, DZF | | F1, G1, H1, G4, H4 | 27, 6 | 33%, 8% | 14, 14, 17 | 19%, 18%, 17% |
| | **NF45** | Interleukin enhancer-binding factor 2, ILF2 | Q12905 | 43,04 | | | B3, C3, D3, C5, D5, E5 | 7, 24, 3 | 27%, 58%, 12% | 3, 25, 9 | 8%, 66%, 31% |
| | **DHX36** | RNA helicase associated with AU-rich eleme | Q9H2U1 | 114,70 | DEXDc, HELIC | | E1, F4 | 4 | 4% | 5 | 6% |
| | **PKR** | p68 kinase, Interferon-induced, double-stranded RNA-activated protein kinase | P19525 | 62,06 | 2 DSRM, STY kinase | | A5, B5 | 11, 4 | 22%, 9% | | |
| | **DDX1** | ATP-dependent RNA helicase DDX1, DEAD box protein retinoblastoma, DBP-RB | Q92499 | 82,38 | SPRY, DEXDc, HELIC | 6.81 | G1 | | | 6 | 11% |
| | **Myb binding protein 1A** | P160, PAP2 | Q9BQG0 | 148,76 | | 9.34 | A1, B1 | | | 25,11 | 25%, 10% |
| **Protéines se fixant à l'ADN** | | | | | | | | | | | |
| | **Ku 70** | ATP dependent DNA helicase 2 subunit 1, 70 kDa subunit of Ku antigen, CTC75, Lupus Ku autoantigen protein p70 | P12956 | 69,80 | SAP, Ku78 | 6.23 | B2 | | | 5 | 11% |

**Tableau VIII-2 : Liste des protéines identifiées par spectrométrie de masse dans les complexes ribonucléoprotéiques assemblés sur l'ARN MS2 et NRS-MS2 lors de la deuxième expérience de chromatographie d'affinité.**

Les protéines sont répertoriées selon leur fonction cellulaire. Le nom alternatif, le numéro d'accession (N°AC), le poids moléculaire (PM), le point isoélectrique (pI) et numéro de chaque bande du gel SDS-PAGE dans laquelle la protéine a été identifiée sont indiqués. Pour chaque protéine, le pourcentage de recouvrement de la séquence protéique (Couv) et le nombre de peptides identifiés sont reportés.

Parmi les protéines spécifiques de l'ARN NRS MS2, 3 protéines déjà connues pour se fixer sur l'élément NRS ont été identifiées : la protéine ASF (SFRS1), la protéine hnRNP H et la protéine U1A.

La protéine SR ASF a déjà été identifiée au cours de la première expérience de chromatographie d'affinité, ce qui confirme l'idée que c'est bien un partenaire de l'élément NRS. La protéine hnRNP H se fixe sur les séquences riches en G qui se situent dans les hélices HI et HII de l'élément NRS (Fogel *et al.*, 2002 ; Fogel et McNally, 2000). Il a été proposé qu'elle favorise le recrutement de la snRNP U11 sur le pseudo site et qu'elle ne soit pas un antagoniste des protéines SR (Fogel *et al.*, 2002 ; Fogel et McNally, 2000 ; McNally *et al.*, 2004). La protéine U1A fait partie de la particule snRNP U1 et se fixe avec une forte affinité sur la tige boucle II du snRNA U1 (Oubridge *et al.*, 1994). Son identification suggère que la particule snRNP U1 était présente dans les complexes assemblés sur l'ARN NRS-MS2, ce qui est en accord avec les données obtenues par la chromatographie d'affinité basée sur la tobramycine où nous avions mis en évidence la présence du snRNA U1.

Comme lors de la première analyse, les complexes formés avec chacun des 2 ARN contenaient de nombreuses protéines hnRNP. Néanmoins, les bandes correspondant à certaines d'entre elles comme la protéine hnRNP A1 étaient beaucoup plus intenses dans les complexes assemblés avec l'ARN NRS-MS2. Parmi les protéines associées à l'ARN NRS-MS2, nous avons noté la présence de la protéine p54(nrb) qui est connue pour se lier directement à l'ARN et possède de nombreuses fonctions (Shav-Tal et Zipori, 2002). Elle a été retrouvée associée à la fois à l'extrémité C-terminale de la large sous-unité de l'ARN polymérase II et au site donneur d'épissage au sein de complexes associant transcription et épissage (Kameoka *et al.*, 2004).

Nous avons ensuite comparé les résultats de la deuxième chromatographie d'affinité avec ceux de la première chromatographie d'affinité (**Figure VIII-27**).

La comparaison des résultats des deux expériences mettait en évidence 8 protéines uniquement identifiées dans les complexes assemblés sur l'ARN NRS MS2 qui étaient communes aux deux expériences. L'une d'elle était la protéine ASF que l'on sait se fixer sur l'élément NRS, ce qui était satisfaisant (McNally et McNally, 1996). Les 6 autres protéines n'étaient pas connues pour s'associer avec l'élément NRS. Toutes possèdent un domaine de liaison à l'ARN et sont impliquées dans le métabolisme des ARN. A l'exception de la protéine Myb binding protein A, elles ont déjà été trouvées associées dans des complexes spliceosomaux ou des complexes formés sur des éléments de régulation de l'épissage

(Behzadnia *et al.*, 2007 ; Bessonov *et al.*, 2010 ; Jurica et Moore, 2002 ; Rappsilber *et al.*, 2002 ; Zhou *et al.*, 2002).

L'avantage d'avoir uniquement analysé le contenu en protéines des bandes bien visibles sur le gel était de limiter le nombre d'analyses par spectrométrie de masse et de savoir qu'au moins une ou deux protéines étaient présentes en grande quantité. Néanmoins, il nous a paru important de réaliser une troisième expérience où les pistes du gel contenant les protéines des complexes assemblés sur l'ARN MS2 et l'ARN NRS-MS2 soient découpées systématiquement tous les 2 mm. En parallèle, nous avions décidé d'étudier l'effet de la troncation de différents éléments structuraux de la structure 2D de l'élément NRS sur le contenu en protéines des complexes formés.

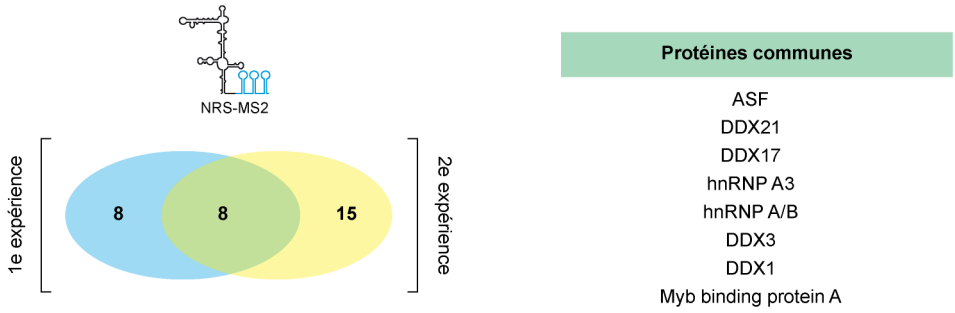



**Figure VIII-27 : Protéines des complexes ribonucléoprotéiques formés sur l'ARN NRS-MS2 retrouvées lors des deux expériences de chromatographie d'affinité.**
Représentation schématique du nombre de protéines identifiées pour l'ARN NRS-MS2 lors des deux expériences de chromatographie d'affinité. Le nombre de protéines distinctes identifiées est indiqué par un chiffre dans un cercle de couleur. Le nombre des protéines communes aux deux expériences est inscrit dans la zone de recouvrement des cercles et leur nom est indiqué à droite.

Nous avons donc répété les expériences avec l'ARN contrôle MS2 et l'ARN NRS-MS2 entier ainsi qu'avec des mutants de l'élément NRS ayant subi différentes troncations des éléments de structure secondaire, à savoir les ARN NRS 22-MS2, NRS 24-MS2, NRS 26-MS2, NRS Δ76-MS2 et NRS 28-MS2 (**Figure VIII-28**). L'élément NRS 22 est le plus court des éléments NRS étudiés, il est tronqué de l'hélice H2 et des tiges-boucles SLS2 et SLS3. Les éléments NRS 24 et NRS 26 sont respectivement tronqués des tige boucles SLS4 et SLS1 et l'élément NRS Δ76 est tronqué d'une partie de la tige boucle SLS2 et de la totalité de la tige boucle SLS3. Enfin, le mutant NRS 28 est tronqué des tiges-boucles SLS2 et SLS3 ainsi que de la partie apicale de l'hélice H2.

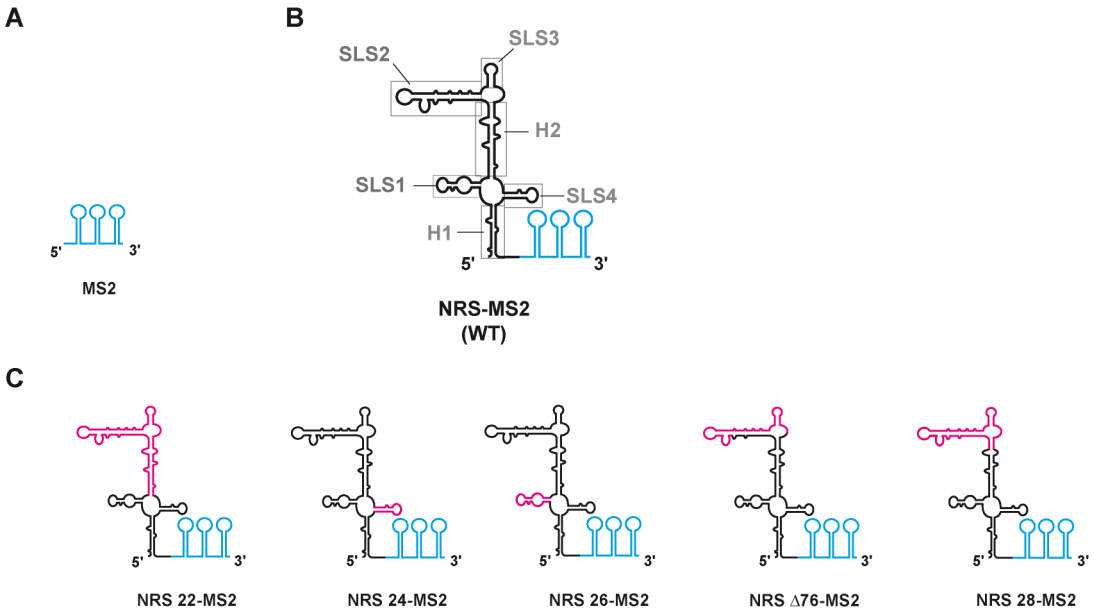



**Figure VIII-28 : Représentation de l'ARN MS2 (A), de l'ARN NRS-MS2 (B) ainsi que des différents éléments NRS tronqués fusionnés à la séquence MS2 (C) NRS 22-MS2, NRS 24-MS2, NRS 26-MS2, NRS Δ76-MS2 et NRS 28-MS2 utilisés dans la troisième expérience de chromatographie d'affinité.**
**A :** Représentation de l'ARN MS2 qui est coloré en bleu.
**B :** Représentation de l'ARN NRS-MS2. L'ARN MS2 est représenté en bleu, l'ARN NRS entier est représenté en noir. Les éléments de structure secondaire sont encadrés et leur nom est indiqué.
**C :** Représentation des éléments NRS tronqués fusionnés à la séquence MS2. La séquence MS2 est colorée en bleu, la séquence NRS présente dans l'ARN est en noir, la séquence NRS tronquée de l'ARN est en rose.

Nous avons utilisé les conditions sélectionnées pour la deuxième expérience de chromatographie d'affinité, c'est-à-dire 250 pmol d'ARN et un rapport molaire [protéine MS2 MBP]/[ARN] de 5.

## IV.6. Purification des complexes ribonucléoprotéiques formés sur les ARN NRS tronqués

### IV.6.1. Test de la fixation de la protéine MBP-MS2 sur les différentes formes tronquées de l'ARN NRS fusionnées à la séquence MS2 par expériences de retard sur gel

Des expériences de retard sur gel ont été réalisées pour vérifier que la protéine MS2 MBP interagissait bien avec les différents ARN NRS tronqués fusionnés à la séquence MS2. Pour cela, nous avions appliqué les mêmes conditions que précédemment. Un mélange de transcrits non radioactifs (100 fmol) et uniformément marqués (1,5 fmol), ainsi que des ARNt (excès molaire de 100 fois) ont été incubés en présence d'une quantité croissante de protéine MS2

MBP recombinante purifiée. Les complexes formés ont été analysés par électrophorèse en gel non dénaturant (**Figure VIII-29**).

Les différents ARN NRS tronqués (NRS 22, NRS Δ76, NRS 24, NRS 26 et NRS 28) fusionnés à la séquence MS2 étaient majoritairement complexés avec la protéine MS2 MBP à partir d'un ratio protéine/ARN de 10, alors que les mêmes ARN non fusionnés à la séquence MS2 restaient majoritairement sous forme libre (**Figure VIII-29**). Sur la base de ces résultats, nous pouvions donc entreprendre les expériences de chromatographie d'affinité avec l'ensemble de ces ARN.

### IV.7. Des complexes ribonucléoprotéiques ayant des compositions différentes sont obtenus en utilisant les différentes formes tronquées de l'ARN NRS

Ces expériences ont été réalisées avec le protocole mis au point précédemment. Comme précédemment, un échantillon a été prélevé à chaque étape. Les ARN ont été analysés par électrophorèse en gel de polyacrylamide non dénaturant coloré au bromure d'éthidium et les protéines par SDS-PAGE. Les résultats obtenus pour les différents ARN sont présentés dans les **Figure VIII-30 à 33**.

Quelque soit l'ARN utilisé lors de la chromatographie d'affinité, les complexes ARN/protéine MS2-MBP formés étaient bien immobilisés sur les billes.

Afin d'analyser les protéines contenues dans chacun des éluats par spectrométrie de masse, il était nécessaire de déposer une quantité suffisamment importante de matériel. Un volume correspondant à 60% de la première élution (fraction E1) des expériences réalisées sur chaque ARN a été fractionné en gel SDS-PAGE et coloré au bleu de coomassie colloïdal. Chaque piste a été découpée systématiquement tous les 2 mm et le contenu en protéines de chacune des bandes a été analysé par spectrométrie de masse (**Figure VIII-34**).

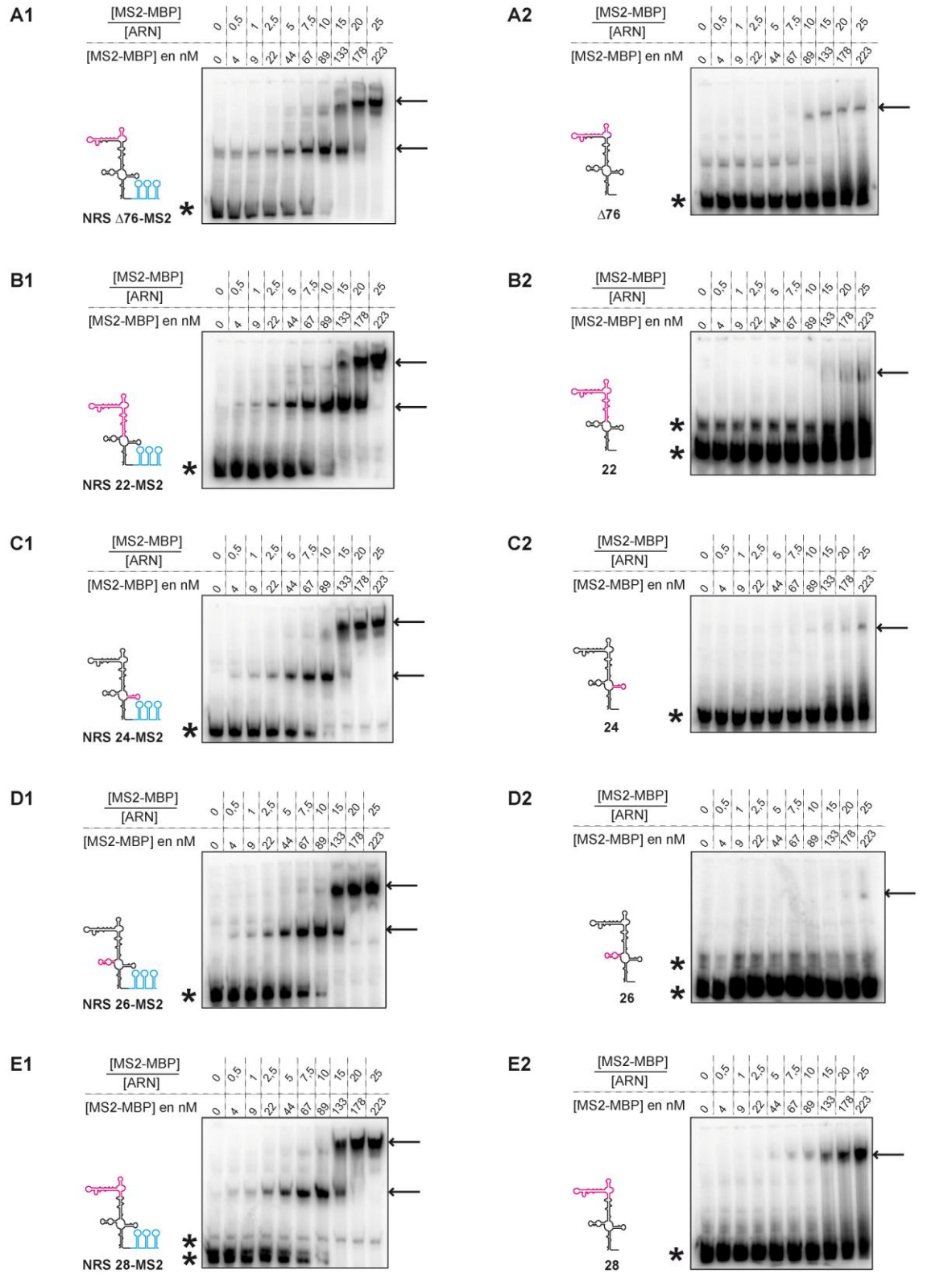



**Figure VIII-29 : Analyse par expériences de retard sur gel des complexes formés après incubation des différents ARN NRS tronqués fusionnés (A1 à E1) ou non fusionnés à la séquence MS2 (A2 à E2) en présence de quantités croissantes de protéine MS2-MBP purifiée.**

Pour chacun des ARN, un mélange de transcrits non radioactifs (100 fmol) et uniformément marqués (1,5 fmol) a été incubé en présence de quantités croissantes de protéine MS2-MBP purifiée. Les positions des ARN libres sont indiquées par une étoile et celles des ARN complexés par une flèche. Les différents ARN NRS fusionnés ou non à la séquence MS2 sont schématisés à gauche des autoradiogrammes et leur nom est indiqué. La séquence MS2 est colorée en bleu, la séquence NRS présente dans l'ARN est représentée en noir, la séquence NRS tronquée de l'ARN est représentée en rose.

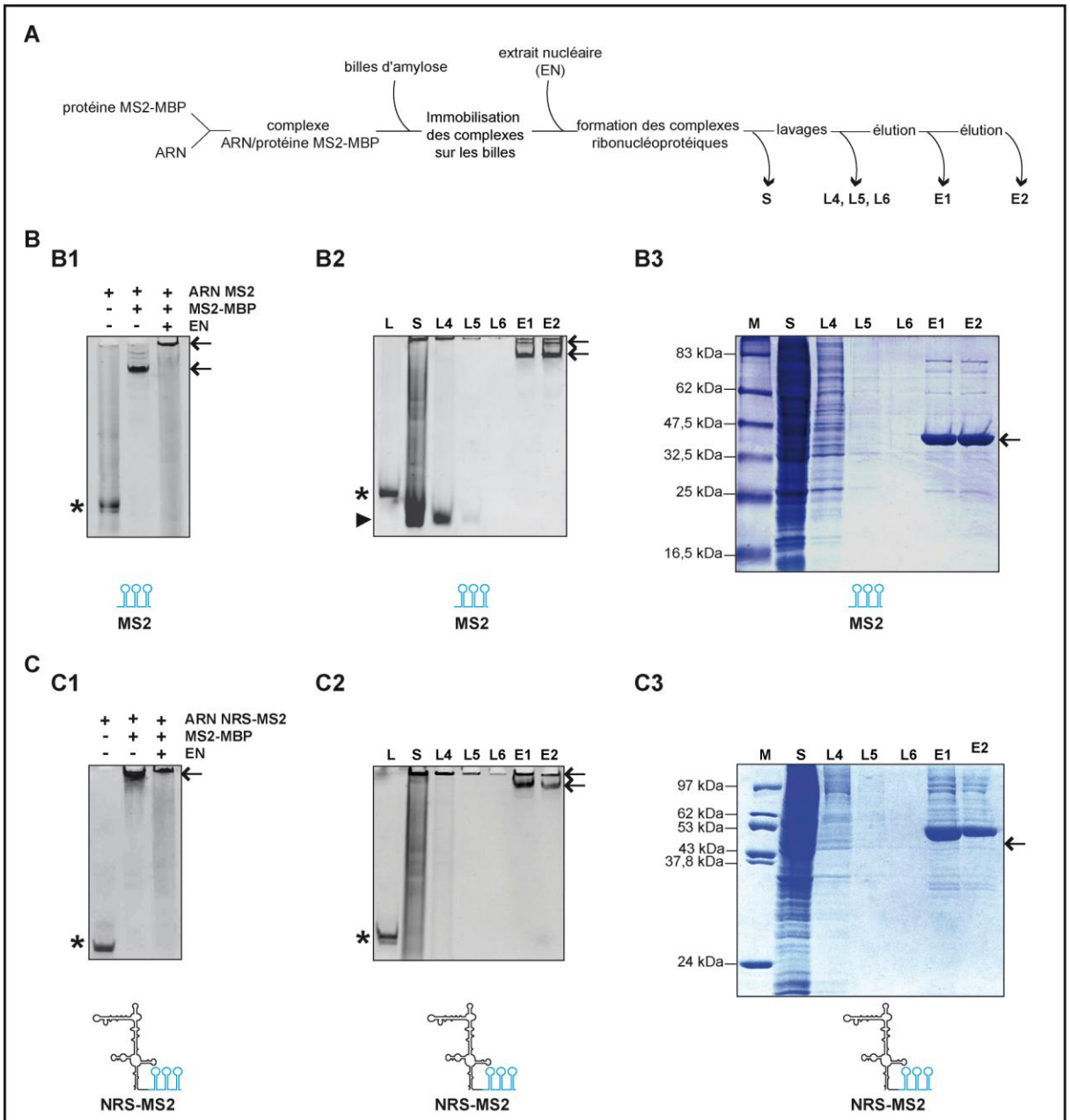



**Figure VIII-30 : Purification des complexes ribonucléoprotéiques formés sur les ARN MS2 (B) et NRS-MS2 (C).**

**A :** Récapitulatif des différentes étapes de la chromatographie d'affinité.

**B :** Purification des complexes ribonucléoprotéiques formés sur l'ARN MS2. Analyse par électrophorèse en gel de polyacrylamide non dénaturant de l'ARN MS2 dans les différents échantillons collectés avant (**B1**) et après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose (**B2**). Le retard sur gel des ARN complexés est indiqué par une flèche, la position de l'ARN libre est indiquée par une étoile, celle des ARNt par un triangle. **B3** : Analyse par gel SDS-PAGE du contenu en protéines des différentes fractions collectées après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose. La piste M correspond au marqueur de taille, la protéine MS2-MBP est indiquée par une flèche.

**C :** Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS-MS2. La légende des encarts C1, C2 et C3 est identique à celle des encarts B1, B2, et B3, respectivement.

Les conditions expérimentales sont identiques à celles décrites dans la Figure VIII-24.

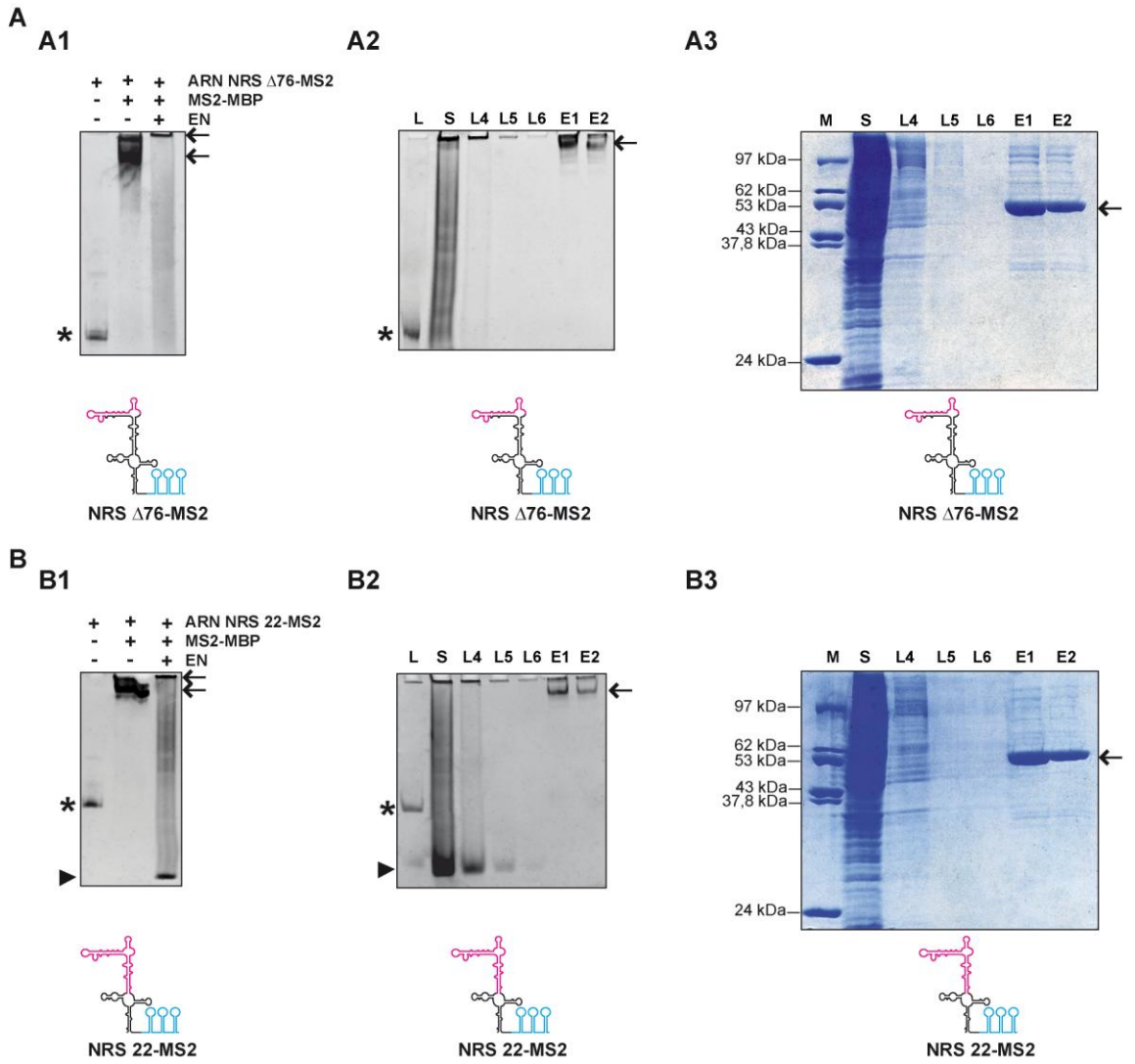



**Figure VIII-31 : Purification des complexes ribonucléoprotéiques formés sur les ARN NRS Δ76-MS2 (A) et NRS 22-MS2 (B).**

**A :** Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS Δ76-MS2. Analyse par électrophorèse en gel de polyacrylamide non dénaturant de l'ARN NRS Δ76-MS2 dans les différents échantillons collectés avant (**B1**) et après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose (**B2**). Le retard sur gel des ARN complexés est indiqué par une flèche, la position de l'ARN libre est indiquée par une étoile, celle des ARNt par un triangle. **B3** : Analyse par gel SDS-PAGE du contenu en protéines des différentes fractions collectées après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose. La piste M correspond au marqueur de taille, la protéine MS2-MBP est indiquée par une flèche.

**B :** Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS 22-MS2. La légende des encarts B1, B2 et B3 est identique à celle des encarts B1, B2, et B3, respectivement.

Les conditions expérimentales sont identiques à celles décrites dans la Figure VIII-24.

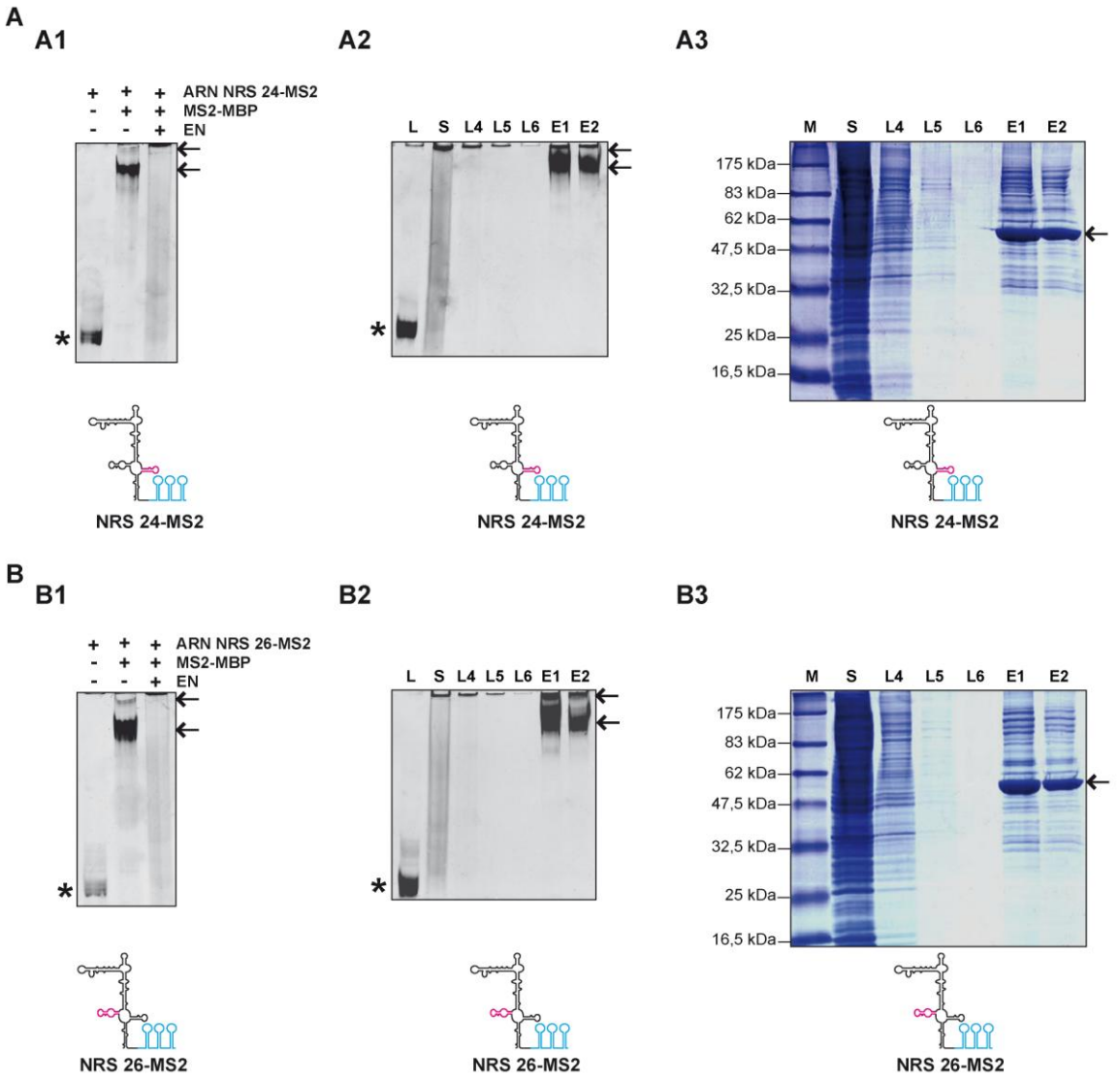



**Figure VIII-32 : Purification des complexes ribonucléoprotéiques formés sur les ARN NRS 24-MS2 (A) et NRS 26-MS2 (B).**

**A :** Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS 24-MS2. Analyse par électrophorèse en gel de polyacrylamide non dénaturant de l'ARN 24-MS2 dans les différents échantillons collectés avant (**B1**) et après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose (**B2**). Le retard sur gel des ARN complexés est indiqué par une flèche, la position de l'ARN libre est indiqué par une étoile. **B3** : Analyse par gel SDS-PAGE du contenu en protéines des différentes fractions collectées après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose. La piste M correspond au marqueur de taille, la protéine MS2-MBP est indiquée par une flèche.

**B :** Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS 26-MS2. La légende des encarts B1, B2 et B3 est identique à celle des encarts B1, B2, et B3, respectivement.

Les conditions expérimentales sont identiques à celles décrites dans la Figure VIII-24.

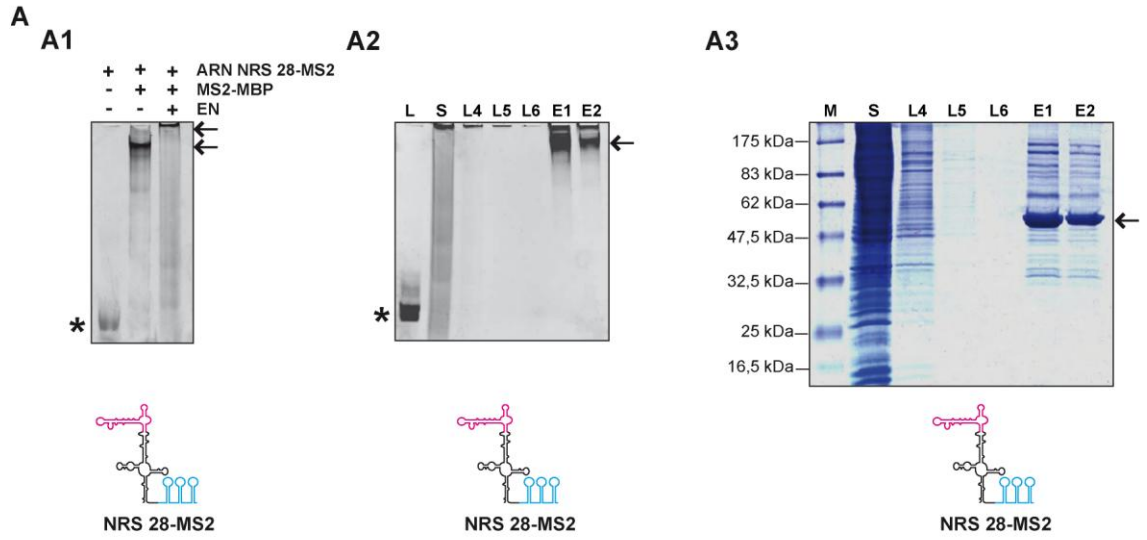



**Figure VIII-33 : Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS 28-MS2.**
**A :** Purification des complexes ribonucléoprotéiques formés sur l'ARN NRS 28-MS2. Analyse par électrophorèse en gel de polyacrylamide non dénaturant de l'ARN NRS 28-MS2 dans les différents échantillons collectés avant (**A1**) et après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose (**A2**). Le retard sur gel des ARN complexés est indiqué par une flèche, la position de l'ARN libre est indiqué par une étoile. **A3** : Analyse par gel SDS-PAGE du contenu en protéines des différentes fractions collectées après l'ajout des complexes ribonucléoprotéiques sur les billes d'amylose. La piste M correspond au marqueur de taille, la protéine MS2-MBP est indiquée par une flèche.

Les conditions expérimentales sont identiques à celles décrites dans la Figure VIII-24.

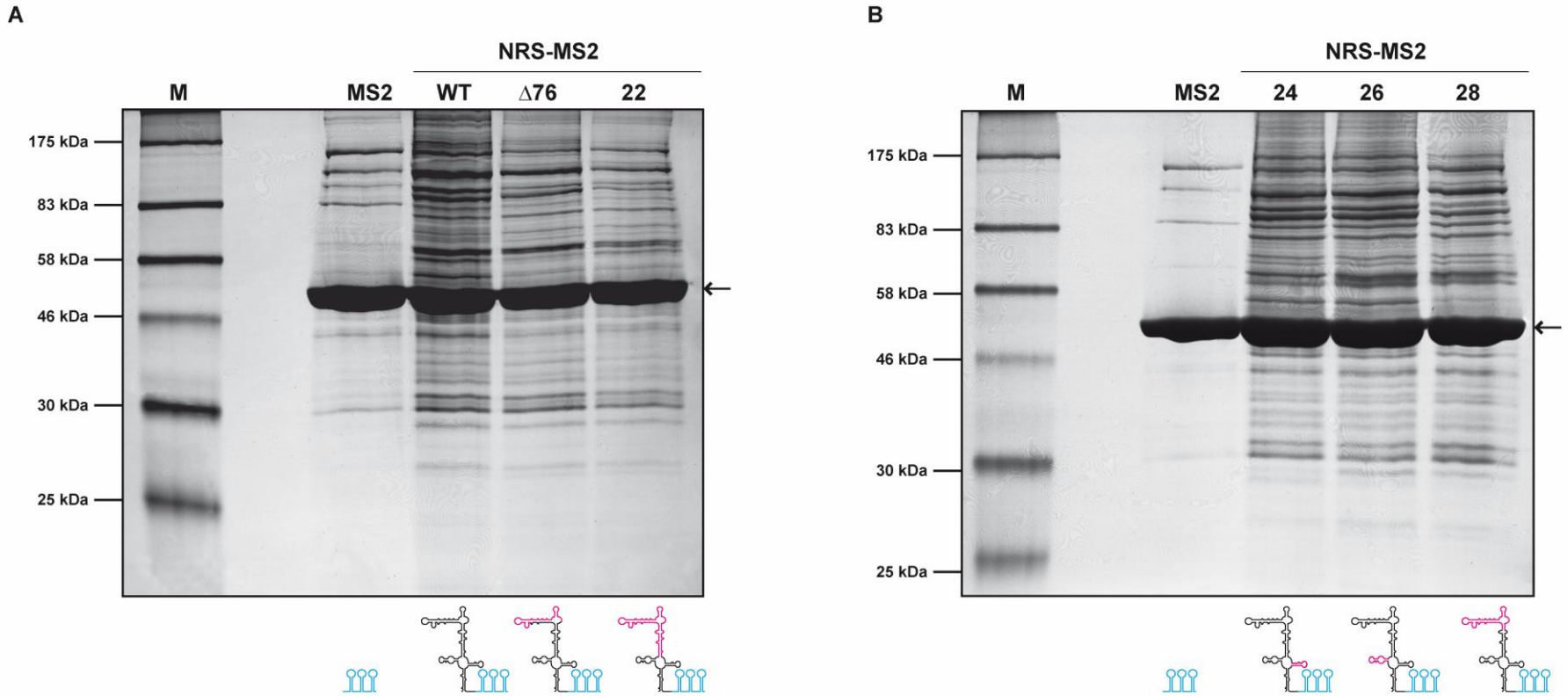



**Figure VIII-34 : Fractionnement en gel SDS-PAGE des protéines contenues dans les complexes ribonucléoprotéiques formés sur l'ARN MS2, l'ARN NRS WT-MS2 et les différents ARN NRS-MS2 tronqués et purifiés par chromatographie d'affinité.**

Un volume correspondant à 60 % de l'élution totale des chromatographies d'affinité réalisées avec l'ARN MS2 et les différents ARN NRS-MS2 a été fractionné par gel SDS-PAGE 10 %. Après l'électrophorèse, le gel a été coloré au bleu de coomassie colloïdal et chaque piste du gel a été découpée tous les 2 mm et les protéines présentes dans ces bandes ont été analysées par spectrométrie de masse. La piste M désigne le marqueur de taille et le poids moléculaire des protéines est indiqué. La protéine MBP-MS2 est indiquée par une flèche. Les ARN utilisés lors des chromatographies d'affinité sont indiqués au-dessus des pistes et sont représentés en-dessous des pistes : la séquence MS2 est colorée en bleu, la séquence NRS présente dans l'ARN est représentée en noir, la séquence NRS tronquée est représentée en rose.

**A :** Fractionnement par gel SDS-PAGE des protéines contenues dans les complexes ribonucléoprotéiques assemblés sur les ARN MS2, NRS WT-MS2, NRS Δ76-MS2 et NRS 22-MS2.

**B :** Fractionnement par gel SDS-PAGE des protéines contenues dans les complexes ribonucléoprotéiques assemblés sur les ARN MS2, NRS 24-MS2, NRS 26-MS2 et NRS 28-MS2.

Vu le très grand nombre de protéines identifiées, nous avons classé les protéines identifiées sur la base de plusieurs critères reposant sur le fait que la détection d'une protéine au sein d'un complexe peut avoir plusieurs causes :

- Cette protéine peut être associée directement à l'ARN
- Cette protéine peut ne pas établir de contact direct avec l'ARN mais y être associée par interaction protéine-protéine

Notre classement des protéines identifiées était donc basé sur les données structurales des protéines (domaine de liaison à l'ARN, domaine d'interaction protéine-protéine), ainsi que sur des données de la littérature.

Il faut souligner aussi qu'une analyse par spectrométrie de masse n'est jamais complètement exhaustive. Certaines protéines sont plus facilement identifiées en raison de coupure par la trypsine, du nombre et de la taille des peptides qui sont générés par cette coupure et leur capacité d'ionisation. De ce fait, les analyses par spectrométrie de masse, telles qu'elles ont été réalisées, ne sont pas quantitatives.

### IV.7.1. Comparaison des protéines identifiées dans les complexes assemblés sur l'ARN NRS-MS2 au cours des 3 expériences de chromatographie d'affinité

Sachant que les complexes assemblés sur l'ARN NRS-MS2 ont été analysés 3 fois, nous avons tout d'abord comparé les protéines détectées dans ces 3 expériences (**Figure VIII-35, Tableau VIII-3**).

Le nombre de protéines identifiées variait selon l'expérience. Les 8 protéines qui étaient communes à la 1$^{e}$ et la 2$^{e}$ expériences ont également été retrouvées dans la 3$^{e}$ expérience (**Figure VIII-35**).

Hormis la protéine ASF dont nous avons déjà évoqué la présence attendue au sein des complexes, les 7 autres protéines communes aux 3 expériences n'ont jamais été décrites comme interagissant avec l'élément NRS.

Quatre de ces protéines, DDX1, DDX3, DDX17, DDX21, sont des hélicases qui font partie de la famille des hélicases à boîte DEAD. Ces protéines sont impliquées dans de nombreux processus de métabolisme de l'ARN.

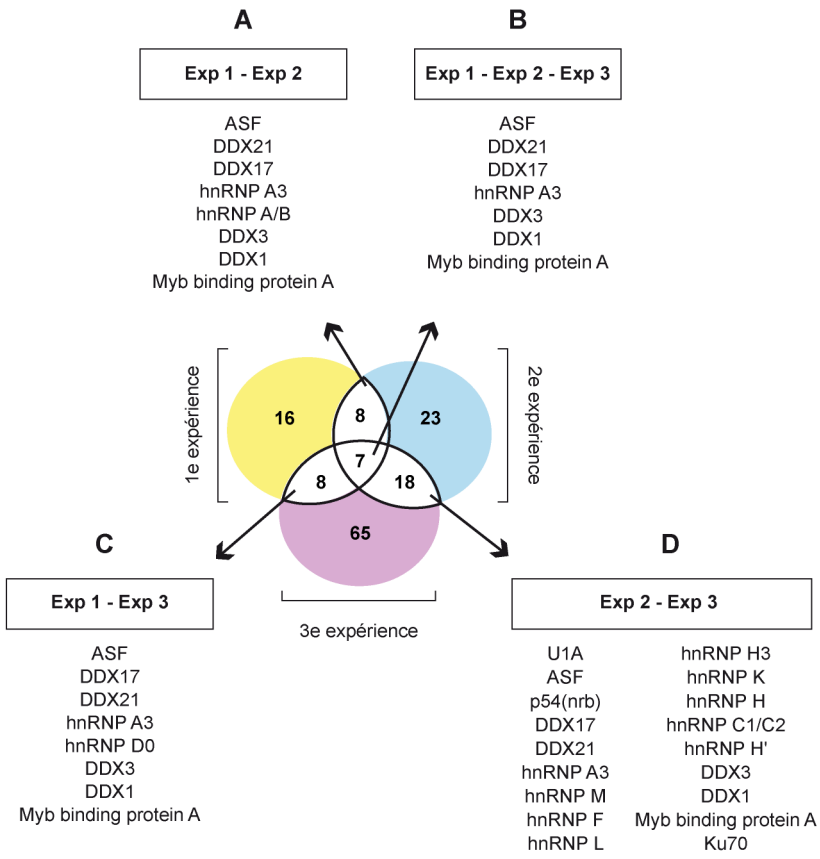



**Figure VIII-35 : Comparaison des protéines identifiées dans les complexes ribonucléoprotéiques formés sur l'ARN NRS-MS2 dans les trois expériences de chromatographie d'affinité.**
Le nombre total de protéines identifiées dans chaque expérience est indiqué par un chiffre dans un cercle de couleur. Le nombre des protéines communes à plusieurs expériences est inscrit dans la zone de recouvrement des cercles et leur nom est indiqué.
**A :** Protéines communes aux expériences 1 et 2.
**B :** Protéines communes aux expériences 1, 2 et 3.
**C :** Protéines communes aux expériences 1 et 3.
**D :** Protéines communes aux expériences 2 et 3.

- La protéine DDX17 ou p72 se fixe aux ARN en double brin avec une meilleure affinité que ceux en simple brin. Sa fixation à l'ARN stimule son activité ATPasique qui fournit l'énergie nécessaire au désappariement des duplex d'ARN. Comme nous l'avons déjà mentionné, la protéine DDX17, comme la protéine paralogue DDX5 (p68), participeraient à la dissociation de la snRNP U1 du site donneur (Lee, 2002). Comme nous l'avons aussi mentionné, elle peut réguler l'épissage alternatif en favorisant l'ouverture de la structure 2D dans les pré-ARNm (Honig *et al.*, 2002).

- La protéine DDX3 a été, comme nous l'avons déjà mentionné, trouvée associée à des complexes spliceosomaux. Son homologue chez la levure, Ded1, fait partie des

spliceosomes et est essentielle à l'épissage chez la levure (Burckin *et al.*, 2005 ; Stevens *et al.*, 2002). Ded1 interagit avec la protéine Prp8 de la particule snRNP U5 et contrecarre l'effet de la mutation portée par la protéine variante Prp8-1 qui entraîne un défaut d'épissage (Jamieson *et al.*, 1991). La protéine DDX3 joue sans doute un rôle dans le transport des ARN du noyau vers le cytoplasme comme le suggèrent ses interactions avec la protéine Aly du complexe EJC et la protéine TAP impliquée dans l'export des ARN vers le cytoplasme (Lai *et al.*, 2008 ; Masuda *et al.*, 2005). La protéine DDX3 semble s'associer aux complexes EJC déposés sur les ARN épissés mais a également été identifiée au sein du complexe spliceosomal B (Deckert *et al.*, 2006 ; Merz *et al.*, 2007). Chez le HIV, cette protéine favorise l'export des ARN non épissés par une voie CRM1 indépendamment de la voie d'export utilisant la protéine Rev (Yedavalli *et al.*, 2004).

- La protéine DDX1 n'a pas été décrite comme étant associée au spliceosome et ne semble avoir aucune implication dans l'épissage. Son affinité pour les séquences polyA et son interaction avec les protéines hnRNP K et CStf64 suggèrent une implication dans la maturation en 3' des pré-ARNm (Bleoo *et al.*, 2001 ; Chen *et al.*, 2002). Cette protéine est l'une des rares hélicases à posséder un domaine SPRY qui lui permet d'interagir avec d'autres protéines ce qui pourrait expliquer sa présence au sein des complexes.

- La protéine DDX21 est impliquée dans la transcription des ARN ribosomiques et joue le rôle de co-activateur de la transcription en interagissant avec le facteur de transcription c-Jun (Henning *et al.*, 2003 ; Holmstrom *et al.*, 2008). Cette protéine a été trouvée associée aux complexes spliceosomaux B et C mais la signification de sa présence n'est pas connue (Bessonov *et al.*, 2010). Cette protéine interagit avec Myb 1A, ce qui pourrait expliquer la présence de cette protéine dans nos complexes.

La protéine hnRNP A3 est également identifiée dans les complexes assemblés sur l'ARN NRS WT lors des 3 expériences de chromatographie d'affinité. Cette protéine appartient à la famille des hnRNP de type A/B et possède 2 domaines RRM et 1 domaine RGG. Elle a été découverte récemment, contrairement aux protéines hnRNP A1 et A2, ses fonctions ne sont pas encore clairement connues (He et Smith, 2009). Elle se fixe à des séquences de type A2RE GCCAAGGAGCC, ainsi qu'à des séquences AuRE dans les régions 3'-UTR des

ARNm COX-2 (Cok *et al.*, 2004) et TNFα (Ma *et al.*, 2002 ; Rousseau *et al.*, 2002). Cette protéine peut transiter vers le cytoplasme et semble être impliquée dans l'export des ARNm (Ma *et al.*, 2002). Cette protéine a été détectée dans les complexes spliceosomaux durant toutes les étapes d'assemblage mais sa fonction n'est pas connue (Behzadnia *et al.*, 2007 ; Bessonov *et al.*, 2010 ; Deckert *et al.*, 2006). Contrairement aux autres protéines hnRNP de type A/B, la protéine hnRNP A3 n'a pas encore été décrite comme régulant l'épissage alternatif d'un pré-ARNm (He et Smith, 2009).

La protéine Myb binding protein 1A est principalement décrite comme un co-facteur interagissant avec les facteurs de transcription comme c-Jun et RelA/p65 de la famille NF-κB dont elle régule l'activité (Owen *et al.*, 2007). Lors d'un stress nucléaire, cette protéine est exportée vers le cytoplasme, où elle facilite l'acétylation de p53 ainsi que son interaction avec la protéine p300 (possédant une activité Histone transférase), ce qui entraîne ensuite l'activation transcriptionnelle des gènes responsables de l'apoptose (Kuroda *et al.*, 2011). La protéine Myb binding protéin 1A ne possède pas de domaine d'interaction avec l'ARN et n'est pas décrite comme ayant un rôle dans l'épissage.

### IV.7.2. Comparaison des protéines identifiées dans les complexes assemblés sur l'ARN NRS-MS2 WT et sur les différents ARN NRS tronqués

Nous avons comparé les résultats de spectrométrie de masse obtenus sur les complexes ribonucléoprotéiques assemblés sur l'ARN NRS-WT et les différents ARN NRS tronqués en utilisant la même classification que ci-dessus (**Tableau VIII-4**).

#### IV.7.2.6. Les protéines des particules snRNP

##### IV.7.2.6.1. Les protéines de la particule snRNP U1

Les deux protéines U1-A et U1-70K ont été identifiées dans les complexes formés sur les ARN NRS-MS2 et NRS 28-MS2 (**Tableau VIII-4**) et la protéine U1-70K était identifiée dans les complexes assemblés sur les ARN NRS 26, 24 et 22-MS2. Il était par ailleurs surprenant de ne pas les détecter dans les complexes formés avec l'ARN NRS Δ76-MS2 sachant que cet élément fixe aussi efficacement la snRNP U1 que l'élément NRS (Gontarek *et al.*, 1993). Il

faut aussi noter que la protéine U1-A existe également sous forme libre en dehors de la particule snRNP U1 ce qui peut expliquer sa présence dans les complexes assemblés sur les ARN NRS-24 MS2. Par ailleurs, l'absence de détection de ces protéines dans certains complexes put être lié aux aléas des analyses par spectrométrie de masse. Il est également important de noter que la particule snRNP U1 comporte d'autres protéines (U1-C et les protéines Sm E, F, G, D1-D3, B) qui ne sont pas détectées ici et dont l'identification a déjà été décrite comme difficile en raison de leur petit poids moléculaire (8,5 à 24 kDa) (Behzadnia *et al.*, 2007).

#### IV.7.2.6.2. Les protéines de la particule snRNP U2

La particule 17S snRNP U2 comporte 19 protéines. Quatre d'entre elles ont été identifiées dans l'analyse des complexes assemblés sur les éléments NRS WT ou tronqués, à savoir les protéines A', B", SF3b155, SF3b130 et SF3b49. Cette détection suggère que la snRNP U2 s'associe à l'élément NRS (**Tableau VIII-4**). Néanmoins, parmi les ARN NRS tronqués, seuls les complexes formés avec l'ARN NRS 28-MS2 contenaient, d'après les analyses par spectrométrie de masse, le même ensemble de protéines de la snRNP U2 que l'ARN NRS WT-MS2. A ce stade de l'étude, il était difficile d'interpréter les résultats obtenus pour les autres ARN NRS tronqués.

La présence de la snRNP U2 et U1 associées à l'élément NRS est en accord avec les données antérieures obtenues sur les complexes formés sur l'élément NRS WT (Cook et McNally, 1998).

Lors de l'assemblage séquentiel des complexes spliceosomaux, l'intégration de la particule snRNP U2 peut avoir lieu soit au niveau du complexe spliceosomal E, soit du complexe spliceosomal A. Dans le complexe E, il a été démontré que la snRNP U2 peut s'associer avec la snRNP U1 qui, elle, interagit avec le site donneur, indépendamment de la présence du point de branchement et d'ATP (Das *et al.*, 2000 ; Hong *et al.*, 1997). Par ailleurs, les particules snRNP U1 et U2 peuvent être présentes au sein de complexes assemblés sur un exon ne possédant qu'un site donneur et sont connues pour interagir l'une avec l'autre (Kar *et al.*, 2011 ; Mattaj *et al.*, 1986 ; Schneider *et al.*, 2010).

Dans le complexe A, la snRNP U2 interagit avec le point de branchement et nécessite l'hydrolyse d'ATP (Xu *et al.*, 2004). Il est intéressant de noter que le NRS possède en aval de la tige boucle SLS3 et dans l'hélice H2 (nt 875-886) une séquence dont 10 des 12 nucléotides sont identiques à la boîte de branchement de l'exon 4 du pré-ARNm de la calcitonine de rat

sur laquelle la snRNP U2 pourrait se fixer (Adema et Baas, 1991 ; Emeson *et al.*, 1989 ; McNally *et al.*, 2004 ; McNally *et al.*, 1991).

#### IV.7.2.6.3. Les protéines de la particule snRNP U5

La particule snRNP U5 contient huit protéines spécifiques. Trois d'entre elles, U5-116k, U5-200k et U5-220k ont été identifiées dans les complexes assemblés sur les ARN NRS WT, 28 et 24 (**Tableau VIII-4**). Deux de ces protéines sont retrouvées dans les complexes assemblés avec les ARN NRS Δ76 et NRS 22 et une seule avec l'ARN NRS 26. Cela suggère que la snRNP U5 soit associée avec ces ARN.

Néanmoins, contrairement aux particules U1 et U2, la particule snRNP U5 n'a jamais été décrite comme interagissant avec l'élément NRS isolé. Lors de l'épissage, la particule snRNP U5 intègre le complexe spliceosomal B au sein de la tri-particule U4/U6.U5. Comme l'élément NRS est assimilable à un site donneur isolé, cela ne peut pas expliquer la présence de la particule snRNP U5 au sein des complexes ribonucléoprotéiques purifiés.

Dans les données de spectrométrie de masse, seule la protéine hPrp3 faisant partie de la di-particule U4/U6 a été identifiée, ce qui suggère que la particule snRNP U5 présente dans les complexes est à l'état libre et ne fait pas partie de la tri-particule U4/U6.U5. Par ailleurs, il a été montré que les protéines de la particule snRNP U5 libre interagissent avec celles de la particule U1 (Abovich et Rosbash, 1997 ; Awasthi *et al.*, 2001 ; Gottschalk *et al.*, 2001 ; van Nues et Beggs, 2001). Par conséquent, nous supposons que dans l'extrait, la particule snRNP U5 pourrait interagir avec la particule snRNP U1 et par son intermédiaire s'associer à l'élément NRS.

Afin de tester si seules certaines des protéines des snRNP s'étaient associées avec l'élément NRS ou s'il s'agissait des snRNP entières, nous avons décidé d'identifier les UsnRNA présents dans les complexes.

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | Exp 1 | | Exp 2 | | Exp 3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Nbr peptides | Couv | Nbr peptides | Couv | Nbr peptides | Couv |
| **snRNP** | | | | | | | | | | | | |
| | **U1-70K** | U1 snRNP 70 kDa, snRNP70, U1AP, U1AP1, U1 small nuclear ribonucleoprotein 70 kDa | P08621 | 51,53 | RRM, SR | 9.94 | | | | | 6 | 12% |
| | **U1 A** | U1 small nuclear ribonucleoprotein A , U1 snRNP A | P09012 | 31,26 | 2 RRM | 9.83 | | | 5 | 28% | 9 | 38% |
| | **U2 snRNP A'** | U2 small nuclear ribonucleoprotein A' | P09661 | 28,40 | LRR | 8.72 | | | | | 4, 6 | 20%, 30% |
| | **SF3b155** | SAP 155 | O75533 | 145,72 | 11 HEAT | 6.58 | | | | | 8 | 8% |
| | **SF3b130** | SAP 130 | Q15393 | 135,49 | | 5.13 | | | | | 13 | 15% |
| | **SmB/B'** | snRNP-B, SNRPB1 Small nuclear ribonucleoprotein associated proteins B and B' | P14678 | 24,59 | Sm | 11.20 | | | | | 4, 2 | 17%, 9% |
| | **hPrp43** | PRP43, DBP1, DDX15, HRH2 | O43143 | 90,88 | DEAH, HELIC, HA2 | 7.12 | | | | | 6 | 11% |
| | **U4/U6 snRNP 90 kDa** | hPrp3, Pre-mRNA-splicing factor 3 | O43395 | 77,48 | PWI | 9.50 | | | | | 2 | 3% |
| | **U5-116 kDa** | 116 kDa U5 small nuclear ribonucleoprotein component, Snrp116, Snu114, U5 snRNP-specific protein,116 kDa | Q15029 | 109,37 | GTP EFTU D2 | 4.84 | | | | | 5, 2 | 7%, 3% |
| | **U5 snRNP-specific 200 kDa protein** | hBrr2 | O75643 | 244,35 | 2 DEXDc, 2 HELIC | 5.73 | | | | | 22 | 14% |
| | **U5 snRNP-specific 220 kDa protein** | Pre-mRNA-processing-splicing factor 8, hPrp8 | Q6P2Q9 | 273,43 | JAB | 8.95 | | | | | 18 | 10% |
| **Proteine SR** | | | | | | | | | | | | |
| | **ASF** | SRp30a, Splicing factor arginine/serine rich 1 | Q07955 | 27,73 | 2 RRM, RS | 10.37 | 6 | 27% | 4, 4 | 21%, 17% | 7 | 32% |
| | **SC35** | SRp30b, Splicing factor arginine/serine rich 2 | Q01130 | 25,46 | RRM, RS | 11.86 | | | | | 2 | 11% |
| | **9G8** | Splicing factor arginine/serine rich 7 | Q16629 | 27,35 | RRM, ZnF | | 6 | 25% | | | | |
| **Protéines associées au spliceosome** | | | | | | | | | | | | |
| | **hPrp22** | PRP22, ATP-dependent RNA helicase DHX8, DDX8, HRH1 | Q14562 | 139,23 | S1, DEXDc, HELIC | 8.54 | | | | | 2 | 2% |
| | **SKIV2L2** | fSAP118, Dbp1, Superkiller viralicidic activity 2 like 2, KIAA0052, ATP-dependent helicase SKIV2L2, Dob1, Mtr4 | P42285 | 117,73 | DEXDc, HELIC | 6.12 | | | | | 10 | 14% |
| | **CC1.3** | UAF65-related factor HCC1, CAPER, fSAP59, HCC1, RNPC2, RNA-binding protein 39 | Q14498 | 59,34 | 3 RRM, RS | 10.10 | | | | | 3 | 8% |
| | **PSF** | PTB-associated-splicing factor, SFPQ, Splicing factor, proline- and glutamine-rich | P23246 | 76,10 | 2 RRM | 9.45 | | | | | 16 | 28% |
| | **p54(nrb)** | NonO protein, Non-POU domain-containing octamer-binding protein, NRB54, P54, NMT55 | Q15233 | 54,20 | 2 RRM | 9.01 | | | 2 | 6% | 13 | 32% |
| | **DDX17** | DEAD box protein 17, DEAD box protein p72, P72, RH70, RNA-dependent helicase p72 | Q92841 | 72,33 | DEXDc, HELIC | 8.82 | 6,9 | 10%,16% | 11, 11 | 18%, 19% | 11, 7 | 2%, 12% |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | Exp 1 | | Exp 2 | | Exp 3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Nbr peptides | Couv | Nbr peptides | Couv | Nbr peptides | Couv |
| **Protéines associées au spliceosome (suite)** | | | | | | | | | | | | |
| | **TOE-1** | Target of EGR1 protein 1 | Q96GM8 | 56,51 | Znf C3H1 | 6.76 | | | | | 8 | 23% |
| | **hsp 27** | heat shock 27kDa protein 1, Heat shock protein beta 1, 28 kDa heat shock protein, Hsp25, HSP28 | P04792 | 22,77 | | 5.98 | | | | | 2 | 13% |
| | **CIRP** | Cold inducible RNA binding protein | Q14011 | 18,64 | RRM | 9.51 | | | | | 3 | 20% |
| | **DNPK1** | PRKDC, DNA-PK catalytic subunit, HYRC1, XRCC7 | P78527 | 468,79 | 2 HEAT, FAT, PI3K/PI4K, FACT, Leu Zip | 6.75 | | | | | 30 | 9% |
| | **RNA helicase p68** | ATP-dependent RNA helicase DDX5 | P17844 | 69105,00 | DEXDc, HELIC | | 11 | 18% | | | | |
| | **DDX21** | Nucleolar RNA helicase 2, Gu-alpha, GURDB, Nucleolar RNA helicase Gu, Nucleolar RNA helicase II, RH II/Gu, RH-II/GU, RH-II/GuA | Q9NR30 | 87,29 | DEXDc, HELIC | 9.32 | 25,4 | 38%, 7% | 15, 2 | 27%, 8% | 3, 22, 14 | 5%, 37%, 24% |
| **hnRNP** | | | | | | | | | | | | |
| | **hnRNP A3** | FBRNP | P51991 | 39,57 | 2 RRM | 9.10 | 5 | 15% | 6, 7, 7, 7, 8 | 20%, 25%, 23%, 17%, 23% | 8, 6, 6 | 24%, 20%, 17% |
| | **hnRNP M** | NAGR1 | P52272 | 77,46 | 3 RRM | 8.84 | | | 9 | 15% | 10, 10 | 17%, 18% |
| | **hnRNP F** | Nucleolin-like protein mcs94-1 | P52597 | 45,64 | 3 RRM, Znf CHHC | 5.38 | | | 11, 3 | 21%, 9% | 8, 3 | 33%, 9% |
| | **hnRNP L** | | P14866 | 60,15 | 3 RRM | 6.65 | | | 9 | 20% | 12 | 32% |
| | **hnRNP G** | | P38159 | 42,31 | RRM | 10.06 | | | | | 8 | 21% |
| | **hnRNP H3** | hnRNP 2H9 | P31942 | 36,90 | 2 RRM | 6.37 | | | 3, 4, 3 | 14%, 14%, 14% | 4, 4 | 15%, 15% |
| | **hnRNP K** | CSBP | P61978 | 50,94 | 3 KH | 5.39 | | | 7, 5 | 22%, 15% | 12, 7 | 32%, 20% |
| | **hnRNP D0** | AUF1, AUF1A, AU-rich element RNA-binding protein 1 | Q14103 | 38,41 | 2 RRM | 7.62 | 3, 4, 6 | 9%, 12%, 14% | | | 4, 6 | 14%, 18% |
| | **hnRNP H** | hnRNP H1 | P31943 | 49,20 | 3 RRM, Znf CHHC | 5.89 | | | 11, 7 | 32%, 18% | 4 | 14% |
| | **hnRNP U-like protein 2** | hnRNP LU-2, Scaffold-attachment factor A2, SAF-A2 | Q1KMD3 | 85,05 | SAP, SPRY | 4.85 | | | | | 5 | 8% |
| | **hnRNP C1/C2** | hnRNP C | P07910 | 33,65 | RRM | 4.95 | | | 6 | 16% | 2, 5, 7 | 9%, 18%, 23% |
| | **TDP43** | TAR DNA binding protein 43, TARDBP | Q13148 | 44,71 | 2 RRM | 5.85 | | | | | 6 | 25% |
| | **hnRNP H'** | hnRNP H2, FTP3 | P55795 | 49,23 | 3 RRM, Znf CHHC | 5.89 | | | 8 | 25% | 3 | 11% |
| | **hnRNP A/B** | APOBEC1-binding protein 1 | Q99729 | 36,59 | 2 RRM | 9.04 | 7 | 21% | 9, 6 | 18%, 19% | | |
| | **hnRNP A2/B1** | | P22626 | 37,41 | 2 RRM | | | | 4, 5, 9, 10, 6, 12, 5, 4 | 14%, 14%, 21%, 30%, 16%, 32%, 12%, 14% | | |
| | **hnRNP R** | | O43390 | 70,90 | 3 RRM | | 25, 12, 16 | 33%, 19%, 22% | | | | |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | Exp 1 | | Exp 2 | | Exp 3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Nbr peptides | Couv | Nbr peptides | Couv | Nbr peptides | Couv |
| **hnRNP (suite)** | | | | | | | | | | | | |
| | **hnRNP Q** | SYNCRIP, NSAP1 | O60506 | 69,56 | 3 RRM | | 14, 19 | 22%, 27% | | | | |
| | **FUS** | Oncogene TLS, RNA-binding protein FUS | P35637 | 53,39 | RRM, ZnF | | | | 4 | 7% | | |
| | **Ybx1** | Nuclease sensitive element-binding protein 1 , CCAAT-binding transcription factor I subunit A, Y-box-binding protein 1 | P67809 | 35,77 | | | | | 4 | 21% | | |
| **Complexe hPrp19/CDC5L** | | | | | | | | | | | | |
| | **CDC5L** | KIAA0432, Cell division cycle 5-like protein | Q99459 | 92,19 | 2 SANT | | 6 | 9% | | | | |
| | **CTNNBL1/NAP** | Beta-catenin-like protein 1, C22orf28, HSPC117, UPF0027 protein, Testis development protein NYD-SP19 | Q9Y3I0 | 55,18 | | 6.77 | | | | | 6 | 13% |
| **EJC** | | | | | | | | | | | | |
| | **Aly/REF** | THO complex subunit 4, Tho4, Transcriptional coactivator Aly/REF, ALY, BEF | Q86V81 | 26,87 | RRM | 11.15 | | | 6 | 36% | | |
| | **DDX3** | DDX14, DDX3X, DBX, Helicase-like protein 2, HLP2 | O00571 | 73,20 | DEXDc, HELIC | 6.73 | 6 | 12% | 5 | 10% | 21, 4 | 39%, 7% |
| **Protéines de la polyadénylation** | | | | | | | | | | | | |
| | **CPSF 160** | Cleavage and polyadenylation specificity factor 160 kDa subunit, CPSF1, Cleavage and polyadenylation specificity factor subunit 1 | Q10570 | 160,78 | | 5.99 | | | | | 16 | 15% |
| | **CPSF 100** | Cleavage and polyadenylation specificity factor 100 kDa subunit, CPSF2, Cleavage and polyadenylation specificity factor subunit 2 | Q9P2I0 | 88,43 | β-CASP | 4.98 | | | | | 5 | 10% |
| | **Symplekin** | HYPI, SPK, SYM | Q92797 | 141,06 | | 5.82 | | | | | 3 | 4% |
| **Protéines se fixant à l'ARN** | | | | | | | | | | | | |
| | **DDX50** | Nucleolar protein Gu2, Gu-beta, GU2, GUB, Nucleolar protein Gu2, RH-II/GuB | Q9BQ39 | 82,51 | DEXDc, HELIC | 9.26 | | | | | 9 | 16% |
| | **DDX1** | ATP-dependent RNA helicase DDX1, DEAD box protein retinoblastoma, DBP RB | Q92499 | 82,38 | SPRY, DEXDc, HELIC | 6.81 | 6 | 12% | 6 | 11% | 12 | 23% |
| | **RNA-binding motif protein 19** | KIAA0682, RBM19 | Q9Y4C8 | 107,27 | 6 RRM | 6.11 | | | | | 9 | 13% |
| | **S1 RNA binding domain containing protein 1** | | Q8N5C6 | 111,71 | S1 | 8.94 | | | | | 3 | 3% |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | Exp 1 | | Exp 2 | | Exp 3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Nbr peptides | Couv | Nbr peptides | Couv | Nbr peptides | Couv |
| **Protéines se fixant à l'ARN (suite)** | | | | | | | | | | | | |
| | **Myb binding protein 1A** | P160, PAP2 | Q9BQG0 | 148,76 | | 9.34 | 3 | 2% | 25, 11 | 25%, 10% | 8, 27 | 9%, 27% |
| | **PKR** | p68 kinase, Interferon-induced, double-stranded RNA-activated protein kinase | P19525 | 62,06 | 2 DSRM, STY kinase | | 3 | 6% | | | | |
| | **Matrin 3** | KIAA0723, MAT3 | P43243 | 94,57 | 2 RRM | 5.87 | | | | | 8 | 13% |
| | **IGF2BP1** | IMP1, CRDBP, ZBP1, Insulin-like growth factor 2 mRNA-binding protein 1, IGF2 mRNA-binding protein 1 | Q9NZI8 | 63,42 | 4 KH, 2 RRM | 9.26 | | | | | 2 | 5% |
| **Protéines se fixant à l'ADN** | | | | | | | | | | | | |
| | **TOP2A** | Top IIa, DNA topoisomerase 2 alpha, TOP2 | P11388 | 174,28 | | 8.82 | | | | | 10 | 8% |
| | **Ku 70** | ATP dependent DNA helicase 2 subunit 1, 70 kDa subunit of Ku antigen, CTC75, Lupus Ku autoantigen protein p70 | P12956 | 69,80 | SAP, Ku78 | 6.23 | | | 5 | 11% | 6 | 14% |
| | **Ku 86** | Ku 80, ATP dependent DNA helicase 2 subunit 2, 86 kDa subunit of Ku antigen, Lupus Ku autoantigen protein p86 | P13010 | 82,65 | Ku78, VWA | 5.55 | | | | | 4 | 7% |
| | **TOP 1** | DNA topoisomerase 1 | P11387 | 90,67 | | 9.35 | | | | | 3 | 4% |
| | **p30 DBC** | DBC.1, Protein KIAA1967, NET35 | Q8N163 | 102,84 | | 5.14 | | | | | 11 | 15% |
| | **SMARCA5** | SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 5, hSNF2H, ISWI | O60264 | 121,83 | DEXDc, HELIC, 2 SANT | | 5 | 4% | | | | |
| **Autres** | | | | | | | | | | | | |
| | **UPF0568 protein C14orf166** | CGI-99, CLE, Chromosome 14 open reading frame 166, LCRP369 | Q9Y224 | 28,05 | | 6.19 | | | | | 3, 10 | 19%, 50% |
| | **Clathrin heavy chain 1** | CLH 17, CLTC/TFE3 Fusion gene, Clathrin heavy chain 1, Hc, KIAA0034 | Q00610 | 191,49 | | 5.48 | | | | | 10 | 8% |
| | **Coatomer subunit alpha** | Alpha COP, HEP-COP, xenin | P53621 | 138,24 | 6 WD40 | 7.70 | | | | | 3 | 3% |
| | **N acetyltransferase 10** | | Q9H0A0 | 115,63 | | 8.50 | | | | | 7 | 8% |
| | **GAPD** | Glyceraldehyde 3 phosphate dehydrogenase | P04406 | 36,03 | | 8.57 | | | | | 6 | 27% |
| | **TfR1** | Transferrin receptor protein 1, CD71, TRFR, p90, CD71 | P02786 | 84,82 | TM, PA | 6.18 | | | | | 3 | 5% |
| | **NEDD-5** | Septin 2, DIFF6, KIAA0158, NEDD5,Pnutl3, Septin-2 | Q15019 | 41,46 | | 6.15 | | | | | 3 | 12% |
| | **HSP 86** | Heat shock protein HSP 90 alpha, Heat shock 86 kDa, Hsp89, Hsp90 | P07900 | 84,61 | | 4.94 | | | | | 4 | 6% |
| | **Protein FAM98B** | FLJ38426 | Q52LJ0 | 37,17 | | 5.96 | | | | | 3 | 13% |
| | **SEC22** | Vesicle trafficking protein SEC22b, ERS-24, vesicle trafficking protein-like 1, SEC22L1 | O75396 | 24,73 | | 8.67 | | | | | 2 | 12% |

**Tableau VIII-3 : Identification des protéines par spectrométrie de masse dans les complexes ribonucléoprotéiques assemblés sur l'ARN NRS-MS2 lors de 3 expériences de chromatographie d'affinité.**

Les protéines sont répertoriées selon leur fonction cellulaire. Le nom alternatif, le numéro d'accession (N°AC), le poids moléculaire (PM), le point isoélectrique (pI) et numéro de chaque bande du gel SDS-PAGE dans laquelle la protéine a été identifiée sont indiqués. Pour chaque protéine, le pourcentage de recouvrement de la séquence protéique (Couv) et le nombre de peptides identifiés sont reportés.

| Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | NRS WT-MS2 | NRS Δ76-MS2 | NRS 28-MS2 | NRS 26-MS2 | NRS 24-MS2 | NRS 22-MS2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **small nuclear ribonucleoprotein** | | | | | | | | | | | |
| **U1-70K** | U1 snRNP 70 kDa, snRNP70, U1AP, U1AP1, U1 small nuclear ribonucleoprotein 70 kDa | P08621 | 51,53 | RRM, SR | 9.94 | 6 | | 9 | 4 | | |
| **U1 A** | U1 small nuclear ribonucleoprotein A , U1 snRNP A | P09012 | 31,26 | 2 RRM | 9.83 | 9 | | 8 | | 3 | 9 |
| **U2 snRNP A'** | U2 small nuclear ribonucleoprotein A' | P09661 | 28,40 | LRR | 8.72 | 4, 6 | | 5 | 3 | | 8 |
| **U2 snRNP B''** | U2 small nuclear ribonucleoprotein B'' | P08579 | 25,47 | 2 RRM | 9.72 | | | | | 2 | |
| **SF3b155** | SAP 155 | O75533 | 145,72 | 11 HEAT | 6.58 | 8 | 8 | 8 | | 2 | |
| **SF3b130** | SAP 130 | Q15393 | 135,49 | | 5.13 | 13 | 12 | 5 | | | |
| **SF3b49** | SAP 49 | Q15427 | 44,36 | 2 RRM | 8.54 | | | 3 | | | |
| **SmB/B'** | snRNP-B, SNRPB1 Small nuclear ribonucleoprotein associated proteins B and B' | P14678 | 24,59 | Sm | 11.20 | 4, 2 | 2 | 4 | 4, 3 | 3 | 4 |
| **hPrp43** | PRP43, DBP1, DDX15, HRH2 | O43143 | 90,88 | DEAH, HELIC, HA2 | 7.12 | 6 | 3 | | | | |
| **U4/U6 snRNP 90 kDa** | hPrp3, Pre-mRNA-splicing factor 3 | O43395 | 77,48 | PWI | 9.50 | 2 | | | | | |
| **U5-116 kDa** | 116 kDa U5 small nuclear ribonucleoprotein component, Snrp116, Snu114, U5 snRNP-specific protein,116 kDa | Q15029 | 109,37 | GTP EFTU D2 | 4.84 | 5, 2 | 5 | 3 | | 7 | |
| **U5 snRNP-specific 200 kDa protein** | hBrr2 | O75643 | 244,35 | 2 DEXDc, 2 HELIC | 5.73 | 22 | 14 | 6 | 16 | 12, 4 | 12 |
| **U5 snRNP-specific 220 kDa protein** | Pre-mRNA-processing-splicing factor 8, hPrp8 | Q6P2Q9 | 273,43 | JAB | 8.95 | 18 | | 5 | | 3, 11 | 9 |
| **Proteine SR** | | | | | | | | | | | |
| **ASF** | SRp30a, Splicing factor arginine/serine rich 1 | Q07955 | 27,73 | 2 RRM, RS | 10.37 | 7 | | 8, 5 | 5, 8 | 2, 4 | 5, 8 |
| **SC35** | SRp30b, Splicing factor arginine/serine rich 2 | Q01130 | 25,46 | RRM, RS | 11.86 | 2 | | | | 2 | |
| **Splicing associated proteins** | | | | | | | | | | | |
| **PSF** | PTB-associated-splicing factor, SFPQ, Splicing factor, proline- and glutamine-rich | P23246 | 76,10 | 2 RRM | 9.45 | 16 | 8 | 5 | 5, 9 | 13 | 11, 9 |
| **p54(nrb)** | NonO protein, Non-POU domain-containing octamer-binding protein, NRB54, P54, NMT55 | Q15233 | 54,20 | 2 RRM | 9.01 | 13 | 5 | 5 | 15, 8 | 5, 12, 10 | 7 |
| **DDX17** | DEAD box protein 17, DEAD box protein p72, P72, RH70, RNA-dependent helicase p72 | Q92841 | 72,33 | DEXDc, HELIC | 8.82 | 11, 7 | 7, 8 | 7,6 | 5, 7, 7 | 8, 7, 6 | 6 |
| **DDX21** | Nucleolar RNA helicase 2, Gu-alpha, GURDB, Nucleolar RNA helicase Gu, Nucleolar RNA helicase II, RH II/Gu, RH-II/GU, RH-II/GuA | Q9NR30 | 87,29 | DEXDc, HELIC | 9.32 | 3, 22, 14 | 16, 17, 15 | 20, 21, 7 | 10, 22, 17, 6, 4 | 8, 28, 17, 8 | 18, 7 |
| **SKIV2L2** | fSAP118, Dbp1, Superkiller viralicidic activity 2 like 2, KIAA0052, ATP-dependent helicase SKIV2L2, Dob1, Mtr4 | P42285 | 117,73 | DEXDc, HELIC | 6.12 | 10 | | | 13 | 17 | |

| Référence | Rappsilber *et al.*, 2002 | Zhou *et al.*, 2002 | Zhou *et al.*, 2002 | Jurica et Moore, 2002 | Merz *et al.*, 2007 | Das *et al.*, 2000 | Hartmuth *et al.*, 2002 | Behzadnia *et al.*, 2007 | Schneider *et al.*, 2010 | Schneider *et al.*, 2010 | Deckert *et al.*, 2006 | Bessonov *et al.*, 2008 | Bessonov *et al.*, 2010 | Makarova *et al.*, 2004 | Bessonov *et al.*, 2010 | Bessonov *et al.*, 2010 | Makarov *et al.*, 2002 | Bessonov *et al.*, 2010 | Bessonov *et al.*, 2010 | Jurica *et al.*, 2002 | Bessonov *et al.*, 2008 | Shi *et al.*, 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nature du complexe | S | S | H | H | AdML ΔI-M3 | E | A | A | cross exon | B-like | B | B | B | BΔU1 | Bact | Bact | B* | C | C | C | C | 3' processing cplx |
| ARN | AdML/B globin | AdML/Ftz | AdML/Ftz | AdML | AdML ΔI-M3 | AdML | MINX | MINX | MINX | MINX | MINX | PM5 (AdML) | PM5 (AdML) | MINX | PM5-10 | PM5-20 | MINX | MINX-GG | PM5 | AdML | PM5 (AdML) | SV40/Adeno |
| **small nuclear ribonucleoprotein** | | | | | | | | | | | | | | | | | | | | | | |
| U1-70K | S | S | | | | | A | A | Ex | B-like | B | B | B | | B act | | | C | C | | | 3' |
| U1 A | S | S | | H | | E | A | A | Ex | B-like | B | B | B | | B act | | | C | C | | | |
| U2 snRNP A' | S | S | | H | | | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | |
| U2 snRNP B'' | S | S | | | | E | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | |
| SF3b155 | S | S | | H | | | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | 3' |
| SF3b130 | S | S | | H | | E | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | |
| SF3b49 | S | S | | | | | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | 3' |
| SmB/B' | S | S | | H | | | A | | | | | | | | | | B* | | | C | | |
| hPrp43 | S | S | | H | | | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | 3' |
| U4/U6 snRNP 90 kDa | S | S | | | | | | A | Ex | B-like | B | B | B | BvU1 | B act | | | C | C | | C | |
| U5-116 kDa | S | S | | H | | | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | |
| U5 snRNP-specific 200 kDa protein | S | S | | H | AdMLΔI-M3 | | A | | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | |
| U5 snRNP-specific protein 220 kDa | S | S | | H | AdMLΔI-M3 | | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | |
| **Proteine SR** | | | | | | | | | | | | | | | | | | | | | | |
| ASF | S | S | | | | | A | A | Ex | B-like | B | B | B | | B act | B act | | C | C | C | C | 3' |
| SC35 | S | S | | | | | A | A | Ex | | B | | | | | B act | | | C | | C | |
| **Protéine associée au spliceosome** | | | | | | | | | | | | | | | | | | | | | | |
| PSF | S | | | H | | | | | | | | | | | | | B* | | | | | 3' |
| p54(nrb) | | | | | | | | | | | | | | | | | | | | | | 3' |
| DDX17 | S | S | | H | | | | A | Ex | B-like | B | B | B | | | B act | | C | C | C | C | 3' |
| DDX21 | S | | | | | | | | | | | B | B | | B act | B act | | | C | | C | 3' |
| SKIV2L2 | S | S | | H | | | | | Ex | B-like | | B | B | | B act | B act | | C | C | C | C | 3' |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | NRS WT-MS2 | NRS Δ76-MS2 | NRS 28-MS2 | NRS 26-MS2 | NRS 24-MS2 | NRS 22-MS2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Protéine associée au spliceosome (suite)** | | | | | | | | | | | | |
| | **DNPK1** | PRKDC, DNA-PK catalytic subunit, HYRC1, XRCC7 | P78527 | 468,79 | 2 HEAT, FAT, PI3K/PI4K, FACT, Leu Zip | 6.75 | 30 | 6 | | | 9 | |
| | **TOE-1** | Target of EGR1 protein 1 | Q96GM8 | 56,51 | Znf C3H1 | 6.76 | 8 | | | | | 4 |
| | **hPrp22** | PRP22, ATP-dependent RNA helicase DHX8, DDX8, HRH1 | Q14562 | 139,23 | S1, DEXDc, HELIC | 8.54 | 2 | | | | | |
| | **CC1.3** | UAF65-related factor HCC1, CAPER, fSAP59, HCC1, RNPC2, RNA-binding protein 39 | Q14498 | 59,34 | 3 RRM, RS | 10.10 | 3 | | | | | |
| | **hsp 27** | heat shock 27kDa protein 1, Heat shock protein beta 1, 28 kDa heat shock protein, Hsp25, HSP28 | P04792 | 22,77 | | 5.98 | 2 | | | | | |
| | **CIRP** | Cold inducible RNA binding protein | Q14011 | 18,64 | RRM | 9.51 | 3 | | | | | |
| | **RACK1** | Gnb2-rs1, GNB2L1, Protein kinase C receptor for activated 1, HLC7, PIG21 | P63244 | 35,06 | 7 WD40 | 7.60 | | | | | 4 | |
| | **RBM7** | RNA binding protein 7 | Q9Y580 | 30,49 | RRM | 9.56 | | | | | 3 | |
| | **IMP3** | IGF2 mRNA-binding protein 3, IGF2BP3, Koc1, VICKZ3 | O00425 | 63,68 | 4 KH, 2 RRM | 8.99 | | | | | 5 | |
| | **DBPA** | DNA-binding protein A, YB-2, dbpA/YB-2, Cold shock domain-containing protein A, ZONAB, , MSY3/4, 70-1-associated nucleic acid binding protein | P16989 | 40,07 | CSD | 9.77 | | | | 4 | | |
| **hnRNP** | | | | | | | | | | | | |
| | **hnRNP A3** | FBRNP | P51991 | 39,57 | 2 RRM | 9.10 | 8, 6, 6 | 6, 5 | 6, 8, 5 | 9, 12, 7, 4, 7 | 8 | 8 |
| | **hnRNP M** | NAGR1 | P52272 | 77,46 | 3 RRM | 8.84 | 10, 10 | 5, 9, 8 | 5, 9, 10 | 9, 7 | 13, 12 | 7, 6 |
| | **hnRNP F** | Nucleolin-like protein mcs94-1 | P52597 | 45,64 | 3 RRM, Znf CHHC | 5.38 | 8, 3 | 8 | 9, 8, 12 | 7, 5 | 9, 6 | 5, 7 |
| | **hnRNP L** | | P14866 | 60,15 | 3 RRM | 6.65 | 12 | 5 | 7 | 15, 6 | 11, 5 | 8 |
| | **hnRNP G** | | P38159 | 42,31 | RRM | 10.06 | 8 | 5 | 7, 7 | 2 | 9 | 3 |
| | **hnRNP H3** | hnRNP 2H9 | P31942 | 36,90 | 2 RRM | 6.37 | 4, 4 | | 3, 7, 5 | 3 | 4, 5 | 2, 4, 3 |
| | **hnRNP K** | CSBP | P61978 | 50,94 | 3 KH | 5.39 | 12, 7 | 9 | 8, 6 | 2 | | 9 |
| | **hnRNP D0** | AUF1, AUF1A, AU-rich element RNA-binding protein 1 | Q14103 | 38,41 | 2 RRM | 7.62 | 4, 6 | | 4, 3 | 4, 2 | 4 | 4 |
| | **hnRNP H** | hnRNP H1 | P31943 | 49,20 | 3 RRM, Znf CHHC | 5.89 | 4 | 8 | 9 | | 4 | 6 |
| | **hnRNP U-like protein 2** | hnRNP LU-2, Scaffold-attachment factor A2, SAF-A2 | Q1KMD3 | 85,05 | SAP, SPRY | 4.85 | 5 | 6 | 8 | 8 | | |
| | **hnRNP C1/C2** | hnRNP C | P07910 | 33,65 | RRM | 4.95 | 2, 5, 7 | | | | 2 | 4 |
| | **TDP43** | TAR DNA binding protein 43, TARDBP | Q13148 | 44,71 | 2 RRM | 5.85 | 6 | | | | | 3 |
| | **hnRNP H'** | hnRNP H2, FTP3 | P55795 | 49,23 | 3 RRM, Znf CHHC | 5.89 | 3 | | | | | |

| Référence | Rappsilber *et al.*, 2002 | Zhou *et al.*, 2002 | Zhou *et al.*, 2002 | Jurica et Moore, 2002 | Merz *et al.*, 2007 | Das *et al.*, 2000 | Hartmuth *et al.*, 2002 | Behzadnia *et al.*, 2007 | Schneider *et al.*, 2010 | Schneider *et al.*, 2010 | Deckert *et al.*, 2006 | Bessonov *et al.*, 2008 | Bessonov *et al.*, 2010 | Makarova *et al.*, 2004 | Bessonov *et al.*, 2010 | Bessonov *et al.*, 2010 | Makarov *et al.*, 2002 | Bessonov *et al.*, 2010 | Bessonov *et al.*, 2010 | Jurica *et al.*, 2002 | Bessonov *et al.*, 2008 | Shi *et al.*, 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nature du complexe | S | S | H | H | AdMLΔI-M3 | E | A | A | cross exon | B-like | B | B | B | BΔU1 | Bact | Bact | B* | C | C | C | C | 3 'processing cplx |
| ARN | AdML/B globin | AdML/Ftz | AdML/Ftz | AdML | AdMLΔI-M3 | AdML | MINX | MINX | MINX | MINX | MINX | PM5 (AdML) | PM5 (AdML) | MINX | PM5-10 | PM5-20 | MINX | MINX-GG | PM5 | AdML | PM5 (AdML) | SV40/Adeno |
| **Protéine associée au spliceosome (suite)** | | | | | | | | | | | | | | | | | | | | | | |
| DNPK1 | | | | | | | | | Ex | B-like | | | | | | | | | | | | 3' |
| TOE-1 | | | | | | | | | | | | | | | B act | | | C | C | | C | |
| hPrp22 | S | S | | | | | | | | | | B | B | | B act | B act | B* | C | C | C | C | |
| CC1.3 | S | S | | | | | A | A | | | B | B | B | | B act | B act | | | | | C | 3' |
| hsp 27 | | | | | | | | | Ex | B-like | | | B | | B act | B act | | C | C | | | |
| CIRP | | S | | | | | | A | Ex | | B | B | | | | | | | | | C | |
| RACK1 | | | | | | | | | Ex | B-like | | | | | | | B* | | | | C | 3' |
| RBM7 | S | | | | | | | | Ex | B-like | B | B | B | | B act | B act | | C | C | | C | 3' |
| IMP3 | S | S | | | | | | | Ex | B-like | | B | B | | B act | B act | | C | C | | C | |
| DBPA | S | | H | | | | | | Ex | B-like | B | B | B | | | B act | | C | C | | C | |
| **hnRNP** | | | | | | | | | | | | | | | | | | | | | | |
| hnRNP A3 | S | | H | H | | | A | A | Ex | B-like | B | B | B | | | B act | | C | C | C | C | 3' |
| hnRNP M | S | | | H | | | | | Ex | B-like | B | B | B | BΔU1 | B act | B act | | C | C | C | C | |
| hnRNP F | S | | | H | | | | | | B-like | B | B | B | | B act | | | C | C | C | C | |
| hnRNP L | S | | H | H | | | | | | | | | | | | | B* | | | | | |
| hnRNP G | S | | | H | AdMLΔI-M3 | | A | | Ex | B-like | B | B | B | | B act | B act | B* | C | C | C | C | |
| hnRNP H3 | S | | | | | | | | | | | B | | | | | | | | | | |
| hnRNP K | S | | H | H | | | A | | | | | B | B | | | B act | | | | C | C | |
| hnRNP D0 | S | | H | | | | | | Ex | B-like | | B | | | | | | | | | C | 3' |
| hnRNP H | S | | | H | | | | | Ex | B-like | | B | B | | B act | B act | | C | C | C | C | |
| hnRNP U-like protein 2 | | | | | | | | | | | | | | | | | | | | | | |
| hnRNP C1/C2 | S | | H | H | AdMLΔI-M3 | | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | |
| TDP 43 | | | | | | | | | | | | B | B | | B act | B act | | C | | | C | |
| hnRNP H' | | | | | | | | | | | B | | | | | | | | | | C | |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | NRS WT-MS2 | NRS Δ76-MS2 | NRS 28-MS2 | NRS 26-MS2 | NRS 24-MS2 | NRS 22-MS2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **hnRNP (suite)** | | | | | | | | | | | | |
| | **E1B-AP5** | hnRNP U-like protein 1, hnRNP LU-1, Adenovirus early region 1B-associated protein 5 | Q9BUJ2 | 95,68 | SAP, SPRY | 6.49 | | | 2,3 | | | 5 |
| | **hnRNP A/B** | APOBEC1-binding protein 1 | Q99729 | 36,59 | 2 RRM | 9.04 | | | 6 | | | |
| **Complexe hPrp19/CDC5L** | | | | | | | | | | | | |
| | **hPrp19** | Pre-mRNA-processing factor 19, PRP19, hPSO4, NMP200, SNEV, U4/U6.U5 tri-snRNP associated protein 1 | Q9UMS4 | 55,15 | UBOX, 7 WD40 | 6.14 | 8 | 4 | 3 | 4 | 7 | 3 |
| | **CTNNBL1/NAP** | Beta-catenin-like protein 1, C22orf28, HSPC117, UPF0027 protein, Testis development protein NYD-SP19 | Q9Y3I0 | 55,18 | | 6.77 | 6 | | | 3 | | |
| | **SPF 27** | Breast carcinoma-amplified sequence 2, BCAS2 | O75934 | 26,12 | | 5.48 | | | | 2 | | |
| **Protéine associée au complexe hPrp19/CDC5L** | | | | | | | | | | | | |
| | **hSYF1** | HCRN, HCNP, XPAB2 | Q9HCS7 | 99,95 | 11 HAT | 5.87 | | | | | 2 | |
| | **hSYF3** | CRNKL1, Crooked neck like protein 1, Clf1, hCrn | Q9BZJ0 | 100,39 | 13 HAT | 8.10 | | | 2 | 4 | | |
| **Complexe EJC** | | | | | | | | | | | | |
| | **Aly/REF** | THO complex subunit 4, Tho4, Transcriptional coactivator Aly/REF, ALY, BEF | Q86V81 | 26,87 | RRM | 11.15 | 7 | 3 | 3 | 3, 4 | 5 | |
| | **DDX3** | DDX14, DDX3X, DBX, Helicase-like protein 2, HLP2 | O00571 | 73,20 | DEXDc, HELIC | 6.73 | 21, 4 | 5, 7 | 4, 11 | 5, 7 | 9 | |
| **Complexe THO/TREX** | | | | | | | | | | | | |
| | **CIP29** | Nuclear protein Hcc-1, HSPC316, SAP domain-containing ribonucleoprotein | P82979 | 23,66 | SAP | 6.10 | | | | | 5 | |
| **Protéine de la polyadénylation** | | | | | | | | | | | | |
| | **CPSF 160** | Cleavage and polyadenylation specificity factor 160 kDa subunit, CPSF1, Cleavage and polyadenylation specificity factor subunit 1 | Q10570 | 160,78 | | 5.99 | 16 | | | 3 | 7 | |
| | **CPSF 100** | Cleavage and polyadenylation specificity factor 100 kDa subunit, CPSF2, Cleavage and polyadenylation specificity factor subunit 2 | Q9P2I0 | 88,43 | β-CASP | 4.98 | 5 | | 2 | | 5 | |
| | **CstF 77** | Cleavage stimulation factor subunit 3, Cleavage stimulation factor 77 kDa subunit | Q12996 | 82,87 | 10 HAT | 8.26 | | | | 3 | | |
| | **Symplekin** | HYPI, SPK, SYM | Q92797 | 141,06 | | 5.82 | 3 | | | | 2, 6 | |
| | **hFip1** | FIP1-like 1 protein, Pre-mRNA 3'-end-processing factor FIP1 | Q6UN15 | 66,49 | | 5.42 | | | | | 4 | |
| **Protéine fixant l'ARN** | | | | | | | | | | | | |
| | **Myb binding protein 1A** | P160, PAP2 | Q9BQG0 | 148,76 | | 9.34 | 8, 27 | 20 | 12, 24 | 37 | 30, 12 | 5 |
| | **S1 RNA binding domain containing protein 1** | | Q8N5C6 | 111,71 | S1 | 8.94 | 3 | | 9 | 7 | | 4 |

| Référence | Rappsilber *et al.*, 2002 | Zhou *et al.*, 2002 | Zhou *et al.*, 2002 | Jurica et Moore, 2002 | Merz *et al.*, 2007 | Das *et al.*, 2000 | Hartmuth *et al.*, 2002 | Behzadnia *et al.*, 2007 | Schneider *et al.*, 2010 | Schneider *et al.*, 2010 | Deckert *et al.*, 2006 | Bessonov *et al.*, 2008 | Bessonov *et al.*, 2010 | Makarova *et al.*, 2004 | Bessonov *et al.*, 2010 | Bessonov *et al.*, 2010 | Makarov *et al.*, 2002 | Bessonov *et al.*, 2010 | Bessonov *et al.*, 2010 | Jurica *et al.*, 2002 | Bessonov *et al.*, 2008 | Shi *et al.*, 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nature du complexe | S | S | H | H | AdMLΔI-M3 | E | A | A | cross exon | B-like | B | B | B | BΔU1 | Bact | Bact | B* | C | C | C | C | 3' processing cplx |
| ARN | AdML/B globin | AdML/Ftz | AdML/Ftz | AdML | AdMLΔI-M3 | AdML | MINX | MINX | MINX | MINX | MINX | PM5 (AdML) | PM5 (AdML) | MINX | PM5-10 | PM5-20 | MINX | MINX-GG | PM5 | AdML | PM5 (AdML) | SV40/Adeno |
| **hnRNP (suite)** | | | | | | | | | | | | | | | | | | | | | | |
| E1B-AP5 | S | | | | | | A | | Ex | B-like | | B | | | | | | | | | C | |
| hnRNP A/B | S | | | | | | | A | Ex | B-like | B | B | B | | | | | C | C | | | |
| **Complexe hPrp19/CDC5L** | | | | | | | | | | | | | | | | | | | | | | |
| hPrp19 | S | S | | H | | | A | A | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | 3' |
| CTNNBL1/NAP | S | | | | | | | A | Ex | B-like | B | B | B | | B act | B act | | C | C | | C | |
| SPF 27 | S | S | | | | | | A | Ex | B-like | B | B | B | Prp19/CDC5L cplx | B act | B act | B* | C | C | | C | |
| **Protéine associée au complexe hPrp19/CDC5L** | | | | | | | | | | | | | | | | | | | | | | |
| hSYF1 | S | S | | | | | | | Ex | B-like | B | B | B | | B act | B act | B* | C | C | C | C | |
| hSYF3 | S | S | | | | | | | Ex | B-like | B | B | B | BΔU1 | B act | B act | B* | C | C | C | C | |
| **Complexe EJC** | | | | | | | | | | | | | | | | | | | | | | |
| Aly/REF | S | S | | | AdMLΔI-M3 | | A | | Ex | B-like | B | B | B | | | B act | B* | C | C | | C | 3' |
| DDX3 | | S | | | | | | | Ex | B-like | B | B | B | | B act | B act | | C | C | | C | 3' |
| **Complexe THO/TREX** | | | | | | | | | | | | | | | | | | | | | | |
| CIP29 | | | | | | | | | | | | | | | | | | | | | | |
| **Protéine de la polyadénylation** | | | | | | | | | | | | | | | | | | | | | | |
| CPSF 160 | S | | | | | | | | Ex | B-like | | B | | | | B act | B* | | | | | 3' |
| CPSF 100 | S | | | | | | | | Ex | | | | | | B act | | | | | | | 3' |
| CstF 77 | S | | | | | | | | | | | | | | | | | | | | | 3' |
| Symplekin | | | | | | | | | | | | | | | | | | | | | | 3' |
| hFip1 | S | | | | | | | | | | | | | | | | | | | | | 3' |
| **Protéine fixant l'ARN** | | | | | | | | | | | | | | | | | | | | | | |
| Myb binding protein 1A | | | | | | | | | | | | | | | | | | | | | | |
| S1 RNA binding domain containing protein 1 | | | | | | | | | | | | | | | | | | | | | | |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | NRS WT-MS2 | NRS Δ76-MS2 | NRS 28-MS2 | NRS 26-MS2 | NRS 24-MS2 | NRS 22-MS2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Protéine fixant l'ARN (suite)** | | | | | | | | | | | | |
| | **RNA-binding motif protein 19** | KIAA0682, RBM19 | Q9Y4C8 | 107,27 | 6 RRM | 6.11 | 9 | | | 4 | 3 | |
| | **Gemin5** | Gem-associated protein 5 | Q8TEQ6 | 168,46 | 13 WD40 | 6.14 | | | | 7 | | |
| | **DDX50** | Nucleolar protein Gu2, Gu-beta, GU2, GUB, Nucleolar protein Gu2, RH-II/GuB | Q9BQ39 | 82,51 | DEXDc, HELIC | 9.26 | 9 | | | 8 | 9 | |
| | **DDX1** | ATP-dependent RNA helicase DDX1, DEAD box protein retinoblastoma, DBP-RB | Q92499 | 82,38 | SPRY, DEXDc, HELIC | 6.81 | 12 | | 6 | | | |
| | **Matrin 3** | KIAA0723, MAT3 | P43243 | 94,57 | 2 RRM | 5.87 | 8 | 3 | | | | |
| | **Int1** | Integrator complex subunit 1, KIAA1440, NET28 | Q8N201 | 244,14 | | 5.77 | | | | 4 | 7 | |
| | **IGF2BP1** | IMP1, CRDBP, ZBP1, Insulin-like growth factor 2 mRNA-binding protein 1, IGF2 mRNA-binding protein 1 | Q9NZI8 | 63,42 | 4 KH, 2 RRM | 9.26 | 2 | | | | | |
| | **La protein** | Lupus La protein, Sjogren syndrome antigen B, SSB, LARP3 | P05455 | 46,81 | RRM, La | 6.68 | | | 2 | | | |
| | **G3BP-1** | Ras GTPase activating protein binding protein 1 | Q13283 | 52,13 | RRM, NTF2 | 5.36 | | | | | 2 | |
| | **POLRA1** | DNA-directed RNA polymerase I subunit RPA1, RNA polymerase I 194 kDa subunit, RPA1, RPA194, RPO14 | O95602 | 194,07 | | 6.80 | | | | | 7 | |
| | **ZNF278** | ZNF278/EWS fusion gene, POZ, BTB POZ domain zinc finger transcription factor, MAZR, PATZ, POZ at HOOK and zinc finger containing protein | Q9HBE1 | 74,01 | BTB/POZ, AT hook, 7 Znf C2H2 | 8.71 | | | | | 2 | |
| **Protéine fixant l'ADN** | | | | | | | | | | | | |
| | **TOP 1** | DNA topoisomerase 1 | P11387 | 90,67 | | 9.35 | 3 | | 5 | 6 | | |
| | **TOP2A** | Top IIa, DNA topoisomerase 2 alpha, TOP2 | P11388 | 174,28 | | 8.82 | 10 | | | 4 | 14 | |
| | **Chromodomain helicase DNA binding protein 3** | Mi-2a, ATP-dependent helicase CHD3, CHD-3, hZFH, Mi-2 autoantigen 240 kDa protein, ZFH, Zinc finger helicase | Q12873 | 226,45 | HELIC, DEXDc, 2 CHROMO, 2 PHD | 6.92 | | | | 5 | 7 | |
| | **Ku 70** | ATP dependent DNA helicase 2 subunit 1, 70 kDa subunit of Ku antigen, CTC75, Lupus Ku autoantigen protein p70 | P12956 | 69,80 | SAP, Ku78 | 6.23 | 6 | | | | | |
| | **Ku 86** | Ku 80, ATP dependent DNA helicase 2 subunit 2, 86 kDa subunit of Ku antigen, Lupus Ku autoantigen protein p86 | P13010 | 82,65 | Ku78, VWA | 5.55 | 4 | | | | | |
| | **p30 DBC** | DBC.1, Protein KIAA1967, NET35 | Q8N163 | 102,84 | | 5.14 | 11 | | | | | |
| | **MTA2** | MTA1L1 protein, Metastasis associated 1 like 1, PID, p53 target protein in deacetylase complex | O94776 | 74,98 | BAH, ELM2, ZNF GATA | 9.70 | | | | | 4 | |
| | **BRG1/SMARCA4** | BAF 190, ATP-dependent helicase SMARCA4, BRM/SWI2 related gene 1, SNF2L4, SNF2LB, BRG1,SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 4, SWI2 | P51532 | 184,53 | HELIC, BROMO, DEXDc, HSA | 7.83 | | | | | 3 | |

| Référence | Rappsilber *et al.*, 2002 | Zhou *et al.*, 2002 | Zhou *et al.*, 2002 | Jurica et Moore, 2002 | Merz *et al.*, 2007 | Das *et al.*, 2000 | Hartmuth *et al.*, 2002 | Behzadnia *et al.*, 2007 | Schneider *et al.*, 2010 | Schneider *et al.*, 2010 | Deckert *et al.*, 2006 | Bessonov *et al.*, 2008 | Bessonov *et al.*, 2010 | Makarova *et al.*, 2004 | Bessonov *et al.*, 2010 | Bessonov *et al.*, 2010 | Makarov *et al.*, 2002 | Bessonov *et al.*, 2010 | Bessonov *et al.*, 2010 | Jurica *et al.*, 2002 | Bessonov *et al.*, 2008 | Shi *et al.*, 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nature du complexe | S | S | H | H | AdMLΔI-M3 | E | A | A | cross exon | B-like | B | B | B | BΔU1 | Bact | Bact | B* | C | C | C | C | 3' processing cplx |
| ARN | AdML/B globin | AdML/Ftz | AdML/Ftz | AdML | AdMLΔI-M3 | AdML | MINX | MINX | MINX | MINX | MINX | PM5 (AdML) | PM5 (AdML) | MINX | PM5-10 | PM5-20 | MINX | MINX-GG | PM5 | AdML | PM5 (AdML) | SV40/Adeno |
| **Protéine fixant l'ARN (suite)** | | | | | | | | | | | | | | | | | | | | | | |
| RNA-binding motif protein 19 | | | | | | | | | | | | | | | | | | | | | | |
| Gemin5 | | | | | | | | | Ex | B-like | | | | | | | | | | | C | |
| DDX50 | | | | | | | | | | | | | | | B act | B act | | | | | | |
| DDX1 | S | | | | | | | | | | | | | | | | | | | | | |
| Matrin 3 | S | | H | H | | | | | | B-like | B | B | B | | | | | C | C | C | C | |
| Int1 | S | | | | | | | | | | | | | | | | | | | | | |
| IGF2BP1 | S | | | | | | | | | | | | | | | | | | | | | |
| La protein | S | | | | | | | | | | | | | | | | | | | | | |
| G3BP-1 | | | | | | | | | | | | | | | | | | | | | | 3' |
| POLRA1 | | | | | | | | | | | | | | | | | | | | | | |
| ZNF278 | | | | | | | | | | | | | | | | | | | | | | |
| **Protéine fixant l'ADN** | | | | | | | | | | | | | | | | | | | | | | |
| TOP 1 | | | | | | | | | | | | | | | | | | | | | | |
| TOP2A | | | | H | | | | | Ex | | | | | | | | | | | | | 3' |
| Chromodomain helicase DNA binding protein 3 | | | | | | | | | | | | | | | | | | | | | | |
| Ku 70 | S | | | H | | | | | | | | | | | | | | | | | | 3' |
| Ku 86 | | | | | | | | | | | | | | | | | | | | | | 3' |
| p30 DBC | | | | | | | | | | B-like | | | | | | | | | | | | |
| MTA2 | | | | | | | | | | B-like | | | | | | | | | | | | |
| BRG1/SMARCA4 | | | | | | | | | | | | | | | | | | | | | | 3' |

| | Nom de la protéine | Nom alternatif | N° AC | PM (KDa) | motif | pI | NRS WT-MS2 | NRS Δ76-MS2 | NRS 28-MS2 | NRS 26-MS2 | NRS 24-MS2 | NRS 22-MS2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Autres** | | | | | | | | | | | | |
| | **N acetyltransferase 10** | | Q9H0A0 | 115,63 | | 8.50 | 7 | 10 | 11, 16 | 4, 8 | 8, 22 | 8 |
| | **UPF0568 protein C14orf166** | CGI-99, CLE, Chromosome 14 open reading frame 166, LCRP369 | Q9Y224 | 28,05 | | 6.19 | 3, 10 | | 4 | | 5, 3 | |
| | **PWP2** | Periodic tryptophan protein 2 homolog, EHOC-17 | Q15269 | 102,39 | 14 WD40 | 5.76 | | | 4 | 7 | 5, 5 | |
| | **Ribonucleases P/MRP protein subunit POP1** | KIAA0061 | Q99575 | 114,64 | | 9.36 | 6 | | 4 | 6 | | |
| | **WD repeat protein 36** | GLC1G, TAWDRP, TA-WDRP | Q8NI36 | 105,26 | 9 WD40 | 7.32 | | | | 8 | 7 | |
| | **Transducin beta-like 3 protein** | SAZD, WD repeat protein SAZD | Q12788 | 56,01 | 13 WD40 | 8.39 | | | | 3 | 4 | |
| | **Protein FAM98B** | FLJ38426 | Q52LJ0 | 37,17 | | 5.96 | 3 | | 3 | | | |
| | **GAPD** | Glyceraldehyde 3 phosphate dehydrogenase | P04406 | 36,03 | | 8.57 | 6 | | | | | 6 |
| | **NEDD-5** | Septin 2, DIFF6, KIAA0158, NEDD5,Pnutl3, Septin-2 | Q15019 | 41,46 | | 6.15 | 3 | | | | | 3 |
| | **Importin subunit alpha 4** | Importin Alpha 3, Qip1 Protein, SRP3 Importin alpha Q1, IPOA3, SRP3 | O00629 | 57,85 | IBB, 10 ARM | 4.80 | | 2 | | | | |
| | **TfR1** | Transferrin receptor protein 1, CD71, TRFR, p90, CD71 | P02786 | 84,82 | TM, PA | 6.18 | 3 | | | | | |
| | **HSP 86** | Heat shock protein HSP 90 alpha, Heat shock 86 kDa, Hsp89, Hsp90 | P07900 | 84,61 | | 4.94 | 4 | | | | | |
| | **SEC22** | Vesicle trafficking protein SEC22b, ERS-24, vesicle trafficking protein-like 1, SEC22L1 | O75396 | 24,73 | | 8.67 | 2 | | | | | |
| | **Clathrin heavy chain 1** | CLH 17, CLTC/TFE3 Fusion gene, Clathrin heavy chain 1, Hc, KIAA0034 | Q00610 | 191,49 | | 5.48 | 10 | | | | | |
| | **Coatomer subunit alpha** | Alpha COP, HEP-COP, xenin | P53621 | 138,24 | 6 WD40 | 7.70 | 3 | | | | | |
| | **ISG15 ligase TRIM25** | EFP, Estrogen-responsive finger protein, RING finger protein 147, RNF147, Tripartite motif-containing protein 25, Ubiquitin/ISG15-conjugating enzyme TRIM25, ZNF147 | Q14258 | 70,94 | Znf C3H4 RING, SPRY | 8.44 | | | | 2 | | |
| | **WD repeat protein 3** | | Q9UNX4 | 106,03 | 13 WD40 | 6.20 | | | | | 12 | |
| | **G3BP2** | Ras GTPase activating protein binding protein 2 | Q9UN86 | 54,09 | RRM | 5.41 | | | | | 4 | |
| | **HSP 84** | Heat shock protein HSP 90 beta, Heat shock 84 kDa, HSP90B | P08238 | 83,21 | | 4.97 | | | | | | 4 |

**Tableau VIII-4 : Liste des protéines identifiées par spectrométrie de masse dans les complexes ribonucléoprotéiques assemblés sur l'ARN NRS-WT-MS2 et les différents ARN NRS tronqués-MS2 lors de l'expériences de chromatographie d'affinité. La présence de ces protéines dans les complexes spliceosomaux est indiquée dans la seconde partie du tableau.**

Les protéines sont répertoriées selon leur fonction cellulaire. Le nom alternatif, le numéro d'accession (N°AC), le poids moléculaire (PM), le point isoélectrique (pI) et numéro de chaque bande du gel SDS-PAGE dans laquelle la protéine a été identifiée sont indiqués. Pour chaque protéine, le

### IV.7.3. Etude de la présence des snRNA dans complexes ribonucléoprotéiques par Northern blot

Pour réaliser les expériences de Northern blot, nous avons utilisé un oligonucléotide ARN radioactif complémentaire d'une séquence spécifique présente dans chacun des snRNA des particules snRNP U1, U2, U4, U5 et U6. Un second oligonucléotide complémentaire à une séquence unique présente dans l'ARN MS2 a été utilisé pour détecter les différents ARN NRS-MS2. La séquence et le site d'hybridation de chacun de ces oligonucléotides sont répertoriés dans le **Tableau X** de la partie Matériel et Méthodes. Les analyses ont été réalisées sur les ARN extraits des complexes ribonucléoprotéiques assemblés sur les différents éléments NRS (**Figure VIII-36**).

Les oligonucléotides radio-marqués utilisés ont permis de détecter spécifiquement la présence des snRNA U1, U2, U4 et U6 ciblés dans l'extrait nucléaire de cellules HeLa (**Figure VIII-36, A, B, D, piste EN**). En raison de leur faible quantité dans les extraits nucléaires, les snRNA U11 et U12 étaient difficilement détectés (**Figure VIII-36, C, piste EN**), et malheureusement, l'oligonucléotide utilisé pour le snRNA U5 n'était pas fonctionnel.

Le snRNA U1 était détecté dans tous les complexes ribonucléoprotéiques assemblés sur les différents ARN NRS hormis le mutant NRS 24-MS2 qui ne contient pas le pseudo site (**Figure VIII-36, A**). Cela suggère que la détection des 3 peptides de la protéine U1-A dans les complexes formés avec l'ARN NRS 24-MS2 ne correspond pas à une fixation de la snRNP U1 mais à une association de la protéine qui peut ne pas être significative. De même, comme le snRNA U2 était très faiblement présent dans les complexes à l'exception de ceux formés avec l'ARN Δ76-MS2 (**Figure VIII-36, B**), on peut se poser la question de la signification de la détection des protéines de la snRNP U2 par spectrométrie de masse dans les différents complexes (**Tableau VIII-4**). Cela suggère que la quantité de particule snRNP U2 associée aux complexes soit si faible que seules certaines de ces protéines sont détectées, suggérant que sa présence n'est pas significative.

L'absence de cohérence entre les résultats de spectrométrie de masse concernant les protéines des UsnRNP et l'analyse par Northern blot pour la présence des snRNA correspondant à ces snRNP nous a amené à focaliser nos études sur les protéines SR et hnRNP et à étudier leur taux relatifs dans les différents complexes étudiés par l'approche de Western blot.

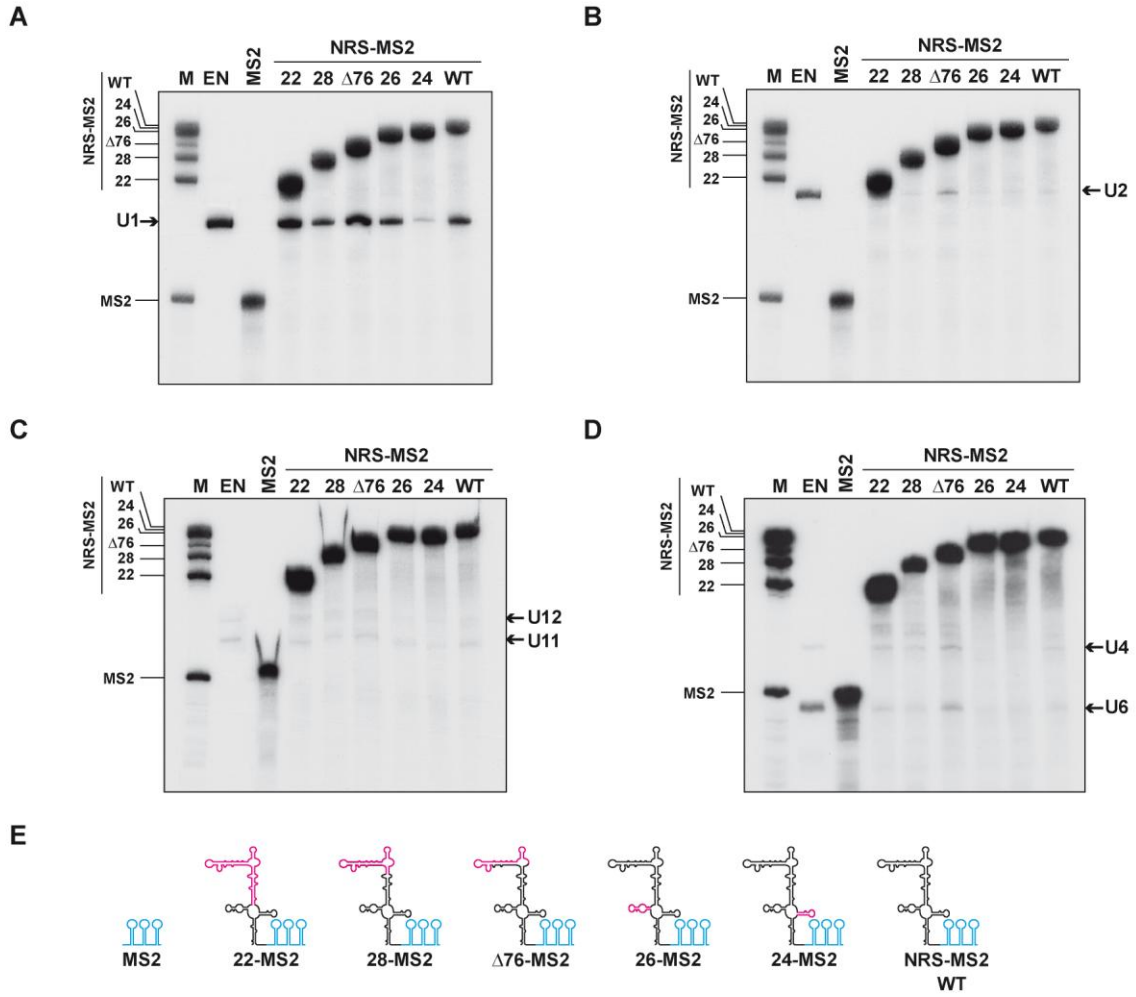



**Figure VIII-36 : Analyse par Northern blot des snRNA U1, U2 (A), U12, U11 (B) et U4, U6 (C) contenus dans les complexes ribonucléoprotéiques formés sur les différents éléments NRS-MS2 et purifiés par chromatographie d'affinité.**

Les ARN contenus dans les complexes ribonucléoprotéiques assemblés sur les différents ARN NRS-MS2 sont purifiés par une extraction au phénol-chloroforme, précipités puis repris dans de l'eau. Un volume correspondant à 15% de la totalité des ARN est ensuite fractionné par électrophorèse en gel acrylamide dénaturant 10 %. Les ARN sont ensuite révélés par une expérience de Northern blot avec des sondes radio-marquées à leur extrémité 5'.

**A :** Northern blot réalisé avec des sondes radio-marquées complémentaires du snRNA U1 et de l'ARN MS2.

**B :** Northern blot réalisé avec des sondes radio-marquées complémentaires du snRNA U2 et de l'ARN MS2.

**C :** Northern blot réalisé avec des sondes radio-marquées complémentaires du snRNA U11, U12 et de l'ARN MS2.

**D :** Northern blot réalisé avec des sondes radio-marquées complémentaires du snRNA U4, U6 et de l'ARN MS2.

La piste M correspond au marqueur de taille qui contient l'ensemble des ARN NRS-MS2, la piste EN correspond aux ARN totaux extraits d'un volume de 7 µl d'extrait nucléaire de cellules HeLa. Les positions des sNRNA sont indiqués par des flèches à gauche des autoradiographies.

**E :** Représentation des différents ARN utilisés lors des purifications des complexes ribonucléoprotéiques par la chromatographie d'affinité. La séquence MS2 et les séquences de l'élément NRS tronquées dans les mutants sont respectivement représentées en bleu et en rose.

### IV.7.4. Etude de la présence des protéines SR et hnRNP dans les complexes ribonucléoprotéiques par Western blot

Nous avons utilisé l'approche de Western blot afin de rechercher la présence de protéines SR et de protéines hnRNP dans les complexes assemblés sur le NRS et ses différents mutants. Contrairement à la technique de spectrométrie de masse, cette approche nous permettait d'avoir des résultats quantitatifs et d'une plus grande sensibilité de détection.. Nous avons choisi de rechercher la présence des protéines SR ASF/SF2, SC35 et 9G8 ainsi que les protéines hnRNP A1 et hnRNP H. Les protéines SR ainsi que la protéine hnRNP H ont précédemment été décrites comme interagissant avec l'élément NRS et ont été détectées par spectrométrie de masse dans nos expériences (Fogel *et al.*, 2002 ; Fogel et McNally, 2000 ; McNally et McNally, 1996). En revanche, la protéine hnRNP A1 n'a jamais été décrite comme interagissant avec l'élément NRS mais les analyses de spectrométrie de masse indiquaient qu'elle était présente en grande quantité dans les complexes formés sur l'ensemble des ARN NRS.

Un volume de 10 µl des différents complexes ribonucléoprotéiques purifiés a été fractionné sur gel SDS-PAGE puis les protéines ont été transférées sur une membrane de nitrocellulose. La quantité de protéine MS2-MBP, révélée après coloration de la membrane au rouge Ponceau, a servi de référence pour équilibrer les pistes (**Figure VIII-37, A**).

Comme attendu, les protéines ASF/SF2, SC35, 9G8 et hnRNP H ont été trouvées associées avec l'ARN NRS WT-MS2 et elles sont également identifiées dans les complexes assemblés sur les ARN NRS 22, 28 et Δ76-MS2 (**Figure VIII-37**). De façon inattendue, ces protéines étaient très peu présentes dans les complexes formés sur les ARN NRS 24-MS2 et NRS 26-MS2. Ce résultat était surprenant car le site de fixation de la protéine ASF/SF2 et une partie du site de fixation des protéines 9G8 et SC35 sont présents dans l'ARN NRS-24 alors qu'ils sont respectivement totalement et partiellement tronqués dans l'ARN NRS-26 (**Figure VIII-37**). Cela suggérait que les structures tige boucles SLS1 et SLS4, qui sont respectivement tronquées des éléments NRS 26 et NRS 24, sont toutes les deux nécessaires pour la fixation de ces protéines. Il est possible qu'en l'absence de ces séquences, la fixation des protéines SR ne soit pas suffisamment stable et induisent leur dissociation des complexes ribonucléoprotéiques.

Par ailleurs, il était surprenant d'observer que la protéine hnRNP A1 est fortement associée à chacun des ARN NRS, qu'il soit entier ou tronqué. Cela suggérait que cette protéine soit impliquée dans un mécanisme de régulation de l'activité du NRS. Sur la base de ces données,

nous avons ensuite réalisé des expériences d'empreinte avec la protéine hnRNP A1 recombinante afin de déterminer ses sites de fixation.

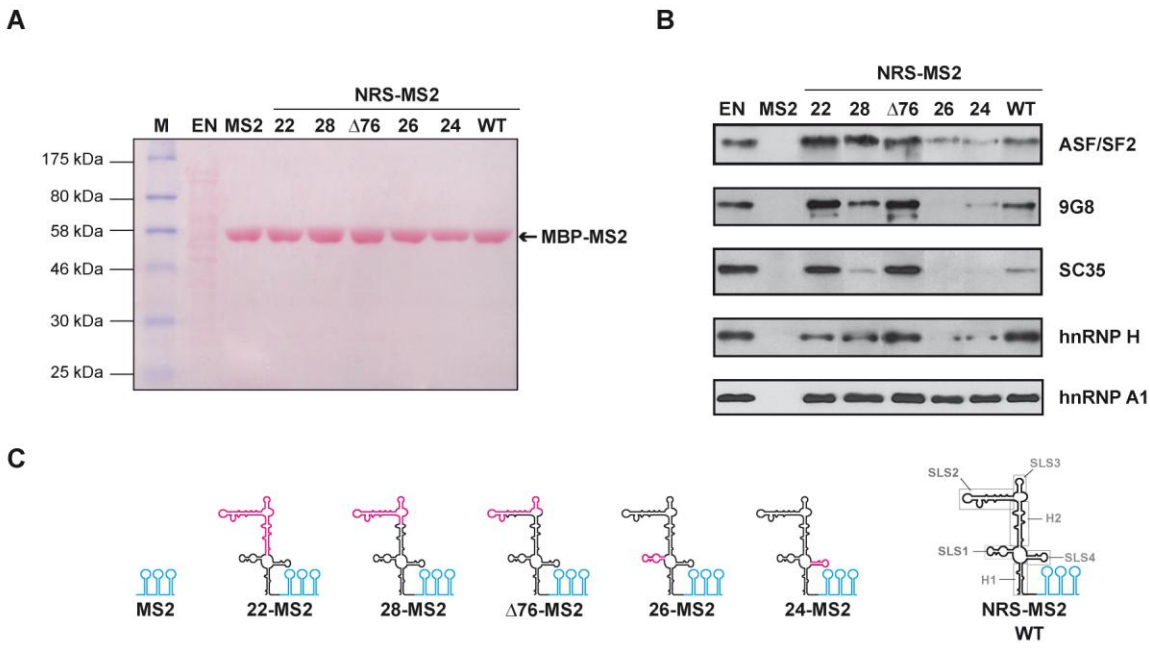



**Figure VIII-37 : Analyse par Western blot de certaines protéines SR et hnRNP contenues dans les complexes ribonucléoprotéiques formés sur l'ARN MS2 et les différents ARN NRS-MS2 qui ont été purifiés par chromatographie d'affinité.**

Un volume correspondant à 10% des complexes ribonucléoprotéiques assemblés sur les différents ARN et purifiés par chromatographie d'affinité, a été fractionné par SDS-PAGE 10%. Les protéines ont ensuite été transférées sur une membrane de nitrocellulose qui a été colorée au rouge Ponceau (**A**) puis ont été détectées avec des anticorps spécifiques (**B**) selon le protocole décrit dans la partie Matériel et Méthode.

**A :** Coloration au rouge Ponceau de la membrane en nitrocellulose avant l'ajout des anticorps. La piste M désigne le marqueur de taille, la piste EN correspond à 5µg de protéines de l'extrait nucléaire. Le nom des ARN sur lesquels les complexes ont été assemblés est indiqué au-dessus des pistes. La protéine MS2-MBP est indiquée par une flèche.

**B :** Western blots réalisés avec des anticorps spécifiques de chacune des protéines. Le nom de la protéine détectée est indiquée à droite des autoradiogrammes, le nom des ARN sur lesquels les complexes ont été assemblés est indiqué au-dessus des pistes.

**C :** Représentation des différents ARN utilisés lors des purifications des complexes ribonucléoprotéiques par chromatographie d'affinité. La séquence MS2 et les séquences de l'élément NRS tronquées dans les mutants sont respectivement représentées en bleu et en rose. Les éléments de structure secondaire de l'ARN NRS WT-MS2 sont encadrés et leur nom est indiqué.

# Chapitre IX : Etude des régions de l'élément NRS protégées par la protéine hnRNP A1 recombinante

## I. Objectif

L'identification de la protéine hnRNP A1 au sein des complexes assemblés sur les différents ARN NRS a retenu notre attention car cette protéine n'a jamais été décrite comme étant un partenaire de l'ARN NRS. Contrairement aux autres protéines identifiées lors de Western blot, la quantité de protéine hnRNP A1 était équivalente dans chacun des complexes assemblés sur les différents ARN NRS (**Chapitre VIII, Figure VIII-37**). Cette protéine qui est très abondante dans l'extrait nucléaire de cellules HeLa, s'associe de façon spécifique mais aussi aspécifique aux pré-ARNm (Martinez-Contreras *et al.*, 2007). Par conséquent, nous nous sommes demandés si sa présence au sein des complexes purifiés était bien spécifique ou aspécifique. La protéine hnRNP A1 a le plus souvent un rôle d'inhibiteur de l'épissage eten se fixant spécifiquement à des séquences de type ESS ou ISS (He et Smith, 2009). Nous avons donc recherché si de telles séquences existaient au sein de l'ARN NRS et si la protéine hnRNP A1 pourrait intéragir directement avec cet ARN.

## II. Stratégie

### II.1. Choix de l'ARN et de la sonde enzymatique

Comme nous voulions déterminer les sites de fixation de la protéine hnRNP A1 sur l'élément NRS entier, nous avons réalisé des expériences d'empreintes avec l'ARN NRS WT et la RNase T1 que nous avions déjà utilisée pour les expériences d'empreinte en extrait nucléaire de cellules HeLa (**Chapitre VII**). Afin de se placer dans des conditions similaires de celles des expériences d'épissage, nous avons réalisé les empreintes dans du tampon D avec une concentration en $MgCl_2$ de 2,8 mM.

### II.2. Détermination de la quantité de protéine hnRNP A1 utilisée pour les expériences d'empreinte

Les empreintes ont été réalisées avec la protéine hnRNP A1 entière recombinante qui avait été produite au laboratoire (Hallay *et al.*, 2006). Pour définir la quantité de protéine hnRNP A1 que nous allions utiliser lors des expériences d'empreinte, nous avons réalisé au préalable

des expériences de retard sur gel. Un mélange des transcrits de NRS WT non radioactifs (1,5 pmol) et uniformément marqués (1,5 fmol) ont été incubés en présence d'une quantité croissante de protéine hnRNP A1 recombinante purifiée. Les ratios [hnRNP A1/ARN NRS WT] utilisés s'échelonnaient entre 7 et 43.

Les résultats nous ont conduits à sélectionner des ratios [hnRNP A1/ARN NRS WT] de 7 et de 20. Dans le ratio de 20, tout l'ARN était complexé.

Une fois assemblés, les complexes ont été soumis à une hydrolyse par la RNase T1 et les positions des clivages ont été identifiées par extension d'amorce par la transcriptase inverse. Nous avons utilisé les mêmes oligonucléotides radio-marqués à leur extrémité 5' que ceux qui ont été utilisés pour les expériences d'empreinte avec la protéine 9G8 recombinante purifiée (O-1618, O-3704, O-3826, **Chapitre X**) pour les analyses de la structure secondaire en solution de cet ARN (O-1617, **Chapitre VI**). Les expériences d'empreintes ont été chacune réalisées trois fois, avec des lots d'ARN et de protéines différents afin de vérifier la reproductibilité des résultats. Des exemples illustrant des autoradiogrammes des fractionnements de produits de transcription inverse sont représentés dans la **Figure IX-1**.

Aucune hydrolyse à des positions non coupées dans l'ARN nu n'a été détectée en présence de protéine hnRNP A1, suggérant que sa fixation ne provoque pas de remaniement de la structure secondaire de l'ARN. Par conséquent, la diminution de l'intensité des coupures observée en présence de la protéine hnRNP A1 devait correspondre à la fixation de la protéine sur l'ARN. Les taux de coupure aux résidus G (G 905, G 849) étaient invariables et ils nous ont servi de référence.

Les empreintes observées sur l'ARN NRS WT ont été répertoriées en fonction des degrés de protection générés par les deux quantités de protéine hnRNP A1 utilisées lors des expériences (ratios [hnRNP A1/ARN NRS WT] de 7 et 20).

Les nucléotides fortement protégés à la plus faible concentration en protéine hnRNP A1 sont supposés correspondre à des sites primaires de fixation de haute affinité. Les nucléotides qui sont faiblement protégés avec la concentration la plus faible en protéine hnRNP A1 et dont le degré de protection est renforcé à la plus forte concentration peuvent correspondre soit à des sites de plus faible affinité, soit à une multimérisation de la protéine.

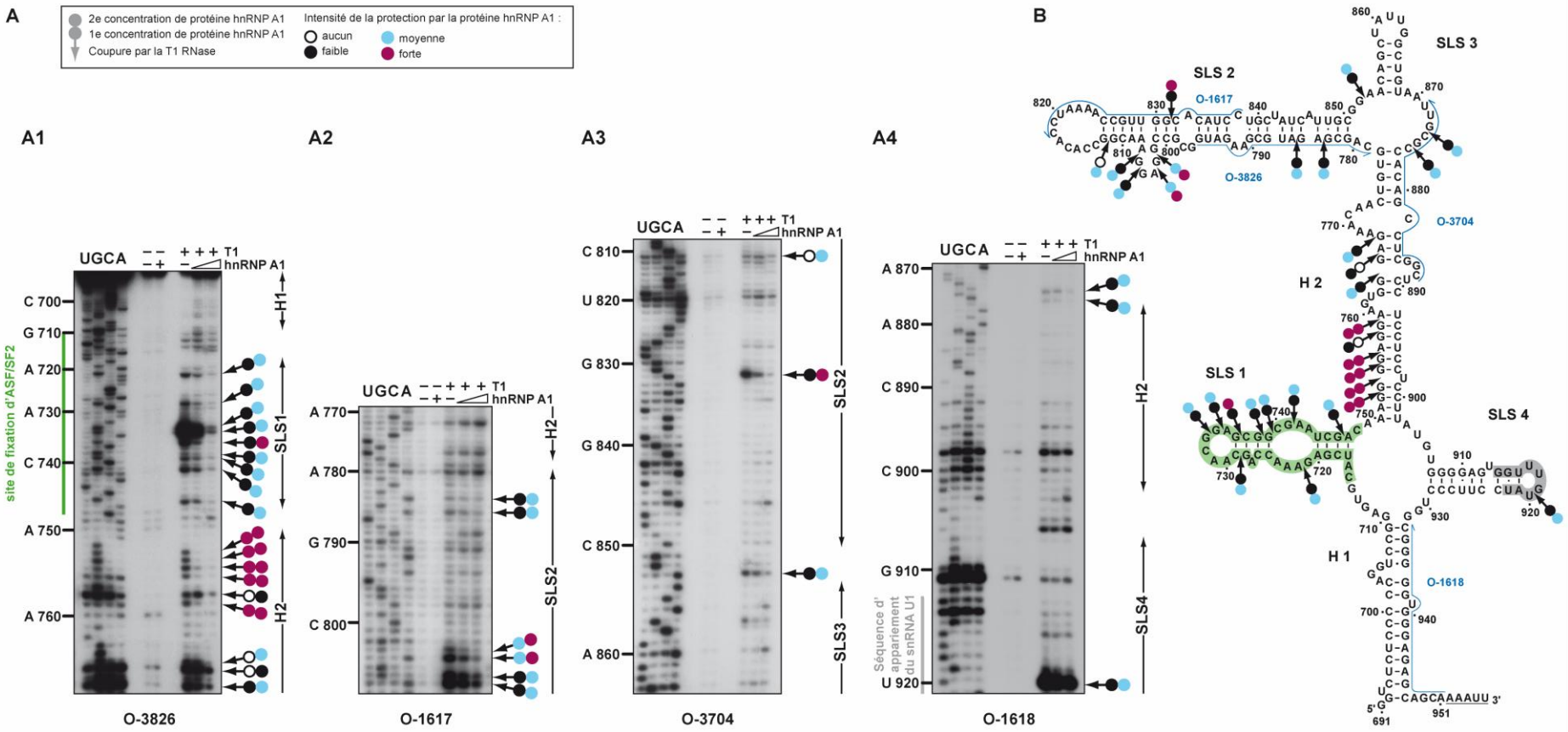



**Figure IX-1 : Analyse des protections générées sur les séquences de l'élément NRS par la protéien hnRNP A1 en présence de la RNase T1.**

**A :** L'ARN a été soumis à une hydrolyse ménagée par la RNase T1 (+)(75 U/µg d'ARN) en présence (+) ou en absence (-) de protéine hnRNP A1 selon les conditions décrites dans la partie Matériel et Méthodes. Les complexes ARN/protéine ont été formés à des ratios [hnRNP A1]/[ARN] de 7 et 20 illutrés par un triangle au dessus des pistes. Les positions des coupures ont été identifiées par extension d'amorce par la transcriptase inverse avec les oligonucléotides 3826 (A1), 1617 (A2) et 3704 (A3) radio-marqués en 5'. Les pistes U, G, C, A correspondent au séquençage de l'ARN. Les positions des nucléotides ainsi que le site de fixation de la protéine ASF/SF2 (McNally *et al*., 1996) et la séquence d'appariement du snRNA U1 (Hibbert *et al*., 1999) sont indiquées à gauche des autoradiogrammes. Les structures secondaires (tige boucle : SLS, hélice : H) délimitées par des doubles flèches et les coupures par la RNase T1 ainsi que les protections des nucléotides représentées selon la légende précisée dans l'encart, sont reportées à droite des autoradiogrammes.

**B :** Représentation des protections contre la RNase T1 générées par la protéine hnRNP A1 sur le modèle de structure secondaire de l'ARN NRS. Les coupures par la RNase T1 et les protections des nucléotides sont reportées selon la légende décrite en (A). La numérotation des nucléotides, les hélices (H) et les structures tige boucle (SLS) sont mentionnées. Les nucléotides non naturels ajoutés à l'extrémité 3' de l'ARN sont soulignés. Les oligonucléotides sont représentés par des flèches bleues et les séquences de fixation de la protéine ASF/SF2 (McNally *et al*., 1996) et d'appariement du snRNA U1 (Hibbert *et al*., 1999) sont respectivement colorées en vert et gris.

Deux sites forts de fixation de la protéine hnRNP A1 ont ainsi été identifiés dans l'ARN NRS WT :

- une région de l'hélice H2 (nt 752-758)
- une région de la tige boucle SLS2 (nt 802-807)

Quatre sites plus faibles dans l'ARN NRS WT ont été identifiés à une plus forte concentration :

- une région de la tige boucle SLS1 (nt 721-746)
- une région de l'hélice H2 (nt 764-767)
- une région de la tige boucle SLS2 (nt 783-785)
- une région simple brin liant la tige boucle SLS 3 et l'hélice H2 (nt 874-876)

Des protections ponctuelles ont également été identifiées aux positions G831 (tige boucle SLS2), G852 (région simple brin reliant les tige boucle SLS2 et SLS3) et G919 (tige boucle SLS4) (**Figure IX-1**).

Les régions protégées sont concentrées dans la région de l'élément NRS qui est riche en purines, ce qui est un contexte favorable pour la fixation de la protéine hnRNP A1 (He et Smith, 2009).

D'après ces données, les séquences AGGGGA (nt 751-756) et GGA (nt 802-803) de l'élément NRS semblent être des sites forts de fixation de la protéine hnRNP A1, bien qu'elles ne correspondent pas aux séquences consensus établies pour les sites de fixation de cette protéine (UAGGG(A/U) ou (Py)UAG )(Burd et Dreyfuss, 1994) (Hutchison *et al.*, 2002). Néanmoins, de nombreux sites hnRNP A1 correspondant à des enchainements de purines ont déjà été décrits, soit dans des pré-mRNA (Expert-Bezancon *et al.*, 2004 ; Guil *et al.*, 2003), soit dans l'ARN du virus HIV (**Figure IX-2**).

Par ailleurs, la protéine hnRNP A1 est aussi capable de se lier au segment de l'hélice H2 (nt 752-758) décrit comme fixant la protéine hnRNP H (Fogel *et al.*, 2002 ; Fogel et McNally, 2000) (**Figure IX-3**). La protéine hnRNP H interagit avec la séquence de l'élément NRS WT comprise entre les nucléotides 740-770 mais ses sites de fixation n'ont pas été précisément identifiés. Seuls les enchaînements de G aux positions 752-755 et 763-765 ont été démontrés comme étant essentiels à sa fixation (Fogel *et al.*, 2002). D'après ces données, les fixations des protéines hnRNP A1 et hnRNP H sur la séquence de l'hélice H2 comprise entre les nucléotides 752-755 semblent exclusives.

Nos données mettaient en évidence une forte fixation de la protéine hnRNP A1 sur les régions de l'élément NRS décrites comme fixant les protéines SR (**Figure IX-3**). D'où l'idée qu'elle pourrait entrer en compétition avec la fixation de ces protéines. Pour pouvoir le tester, nous avons développé *in vitro* et *in cellulo*, des systèmes permettant d'étudier l'activité de l'élément NRS sur l'épissage.

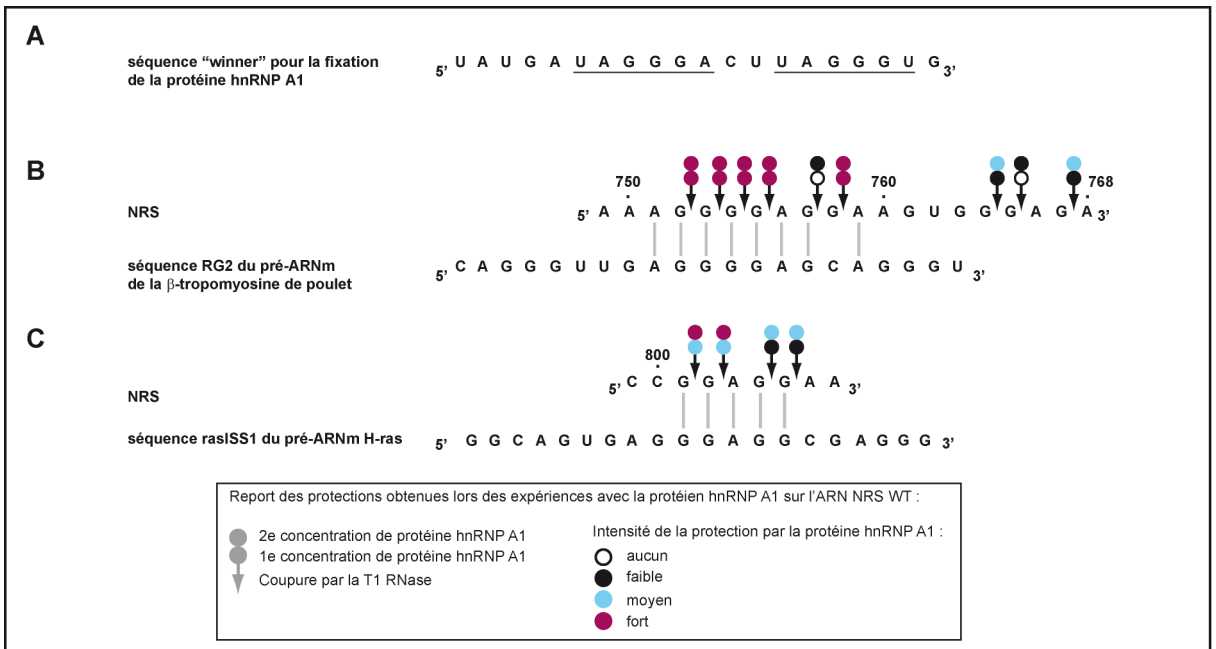



**Figure IX-2 : Comparaison des sites de fixation de la protéine hnRNP A1 identifiés dans l'élément NRS avec les séquences connues pour fixer cette protéine.**
**A :** Séquence de très haute affinité pour la protéine hnRNP A1 appelée séquence "winner" (Burd et Dreyfuss, 1994). Les séquences importantes pour la reconnaissance par la protéine hnRNP A1 sont soulignées.
**B :** Comparaison de la séquence du site primaire de fixation de la protéine hnRNP A1 dans l'hélice H2 de l'élément NRS avec la séquence RG2 de l'intron en aval de l'exon 6B dans le pré-ARNm de la β-tropomyosine de poulet (Expert-Bezançon *et al.*, 2004). Les protections des nucléotides de l'élément NRS par la protéine hnRNP A1 lors des expériences d'empreinte sont illustrées selon la légende précisée dans l'encart. Les traits gris indiquent les nucléotides communs aux 2 séquences.
**C :** Comparaison de la séquence du site primaire de fixation de la protéine hnRNP A1 dans la tige boucle SLS3 de l'élément NRS avec la séquence rasISS1 de l'intron en aval de l'exon IDX dans le pré-ARNm de c-H-ras humain (Guil *et al.*, 2003). Les protections des nucléotides de l'élément NRS par la protéine hnRNP A1 lors des expériences d'empreinte sont illustrées selon la légende précisée dans l'encart. Les traits gris indiquent les nucléotides communs aux 2 séquences.

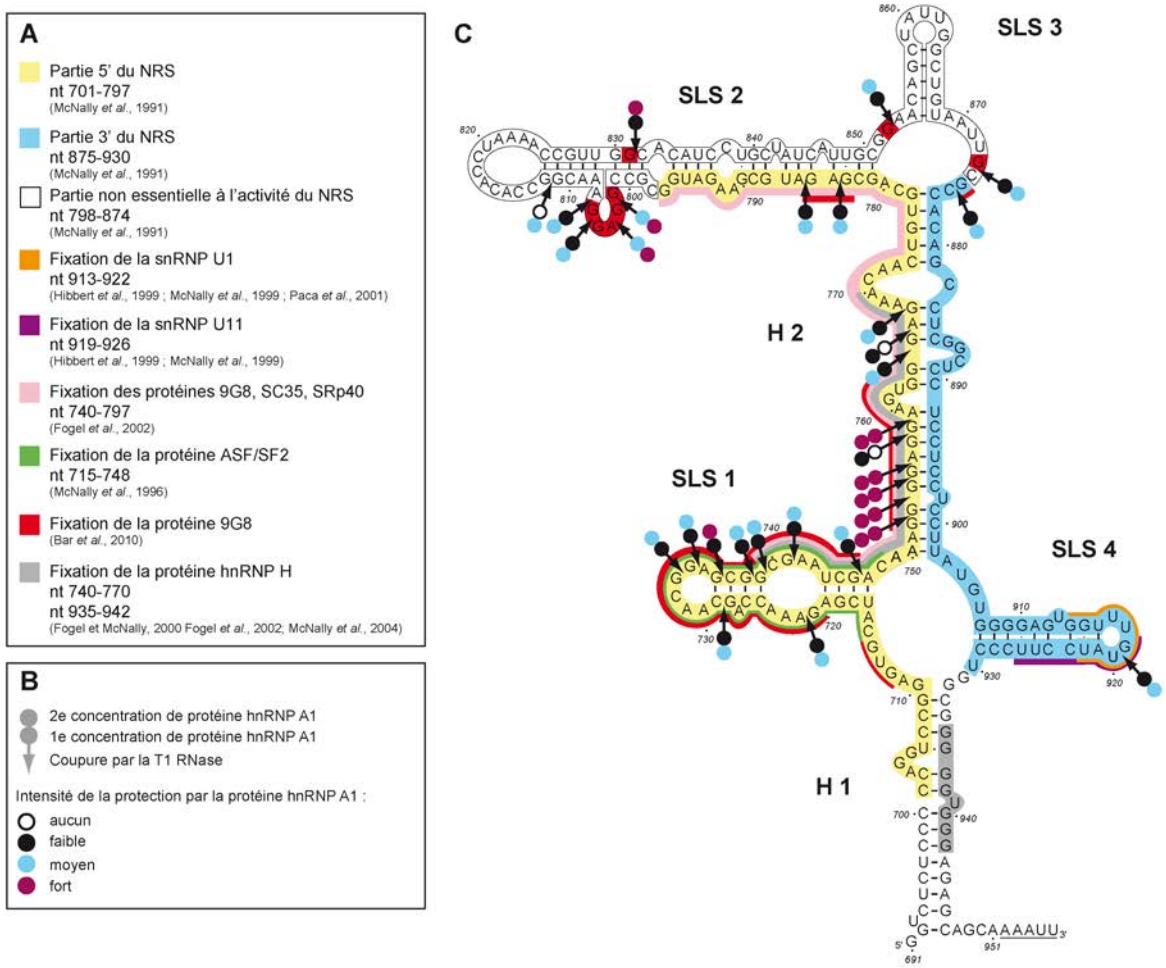



**Figure IX-3 : Comparaison des protections obtenues sur l'ARN NRS du virus RSV lors d'expériences d'empreinte en présence de protéine hnRNP A1 avec les sites de fixation des protéines et particules connues de cet ARN.**

Les sites de fixation des protéines et des particules décrites pour interagir avec l'ARN NRS sont indiqués selon le code couleur indiqué dans l'encart (A) et les nucléotides protégés lors de l'expérience d'empreinte réalisée avec la protéine hnRNP A1 sont illustrés selon la légende de l'encart (B). L'ensemble a été reporté sur le modèle de structure secondaire de l'ARN NRS (C).

# Chapitre X : Etude fonctionnelle de l'élément NRS par des expériences d'épissage *in vitro* et *in cellulo*

## I. Objectif

Notre objectif était double. Nous voulions d'une part identifier l'importance fonctionnelle des différents éléments structuraux que nous avions mis évidence dans l'élément NRS WT (**Chapitre VI**) en utilisant les différents mutants NRS Δ76, 22, 24, 26 et 28 dont nous avions recherché les partenaires par chromatographie d'affinité (**Chapitre VIII**). Nous voulions aussi définir quelle protéine SR pouvait avoir une influence sur l'activité de l'élément NRS en se fixant sur ces différents éléments structuraux. Pour cela, nous avons produit un grand ensemble de constructions où les différents éléments NRS ont été insérés dans l'intron d'un pré-ARNm et nous avons étudié leur activité en mesurant l'inhibition de l'épissage.

## II. Etude de l'activité des éléments NRS sauvage et tronqués par des expériences d'épissage *in vitro*

### II.1. Design et production de constructions d'épissage permettant de tester *in vitro* l'activité de l'élément NRS

Pour tester l'activité de l'élément NRS *in vitro*, nous avons choisi de l'insérer dans les constructions C1 et C2 contenant respectivement les couples de sites donneurs-accepteurs D1-A1 et D1-A2 du VIH-1 (Damier *et al.*, 1997) (**Figure X-1, A et B**). L'avantage des pré-ARNm C1 et C2 étaient qu'ils contenaient un site donneur (D1) fort, et un site accepteur (A1 ou A2) faible et pouvant être régulé, et qu'ils reproduisaient un contexte d'épissage qui se rapproche du contexte viral (McNally, 2008). Les sites accepteurs A1 et A2 sont tous les deux régulés négativement par la protéine hnRNP A1 (Bilodeau *et al.*, 2001 ; Saliou *et al.*, 2009) et le site A2 est activé par la protéine SR ASF/SF2 (Ropers *et al.*, 2004). Nous avons aussi testé la possibilité de mesurer l'activité de l'élément NRS en l'insérant dans le pré-ARNm Minx qui est très efficacement épissé *in vitro* et souvent utilisé pour l'étude de l'assemblage des spliceosomes (Deckert *et al.*, 2006 ; Makarova *et al.*, 2004) (**Figure X-1, C**). Cette construction nous avait été fournie par l'équipe de R. Lührmann et avait été choisie car elle permettait de reproduire les constructions d'épissage habituellement utilisées pour étudier l'inhibition de l'épissage par le NRS (McNally et Beemon, 1992 ; McNally *et al.*, 1991).

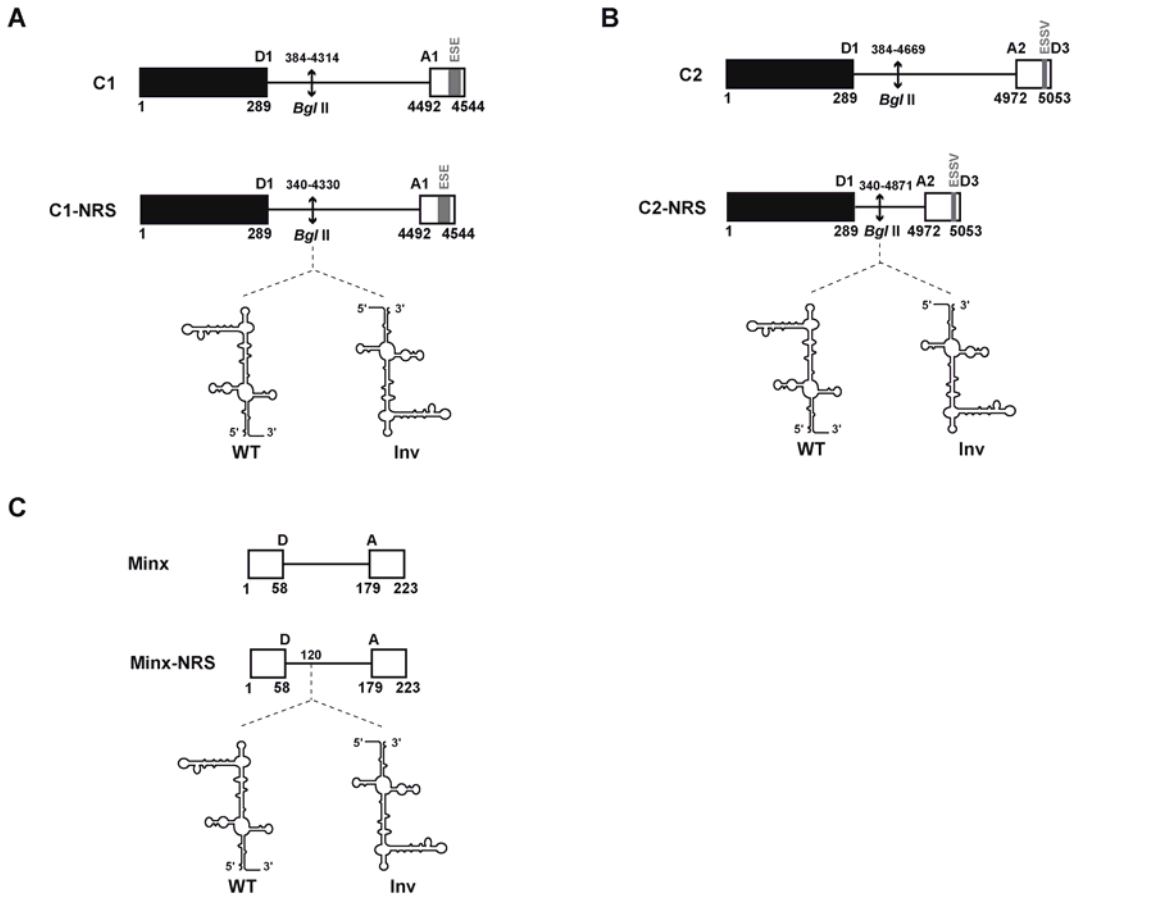



**Figure X-1 : Constructions d'épissage utilisées pour l'étude de l'inhibition de l'épissage par l'élément NRS.**

**A :** Représentation de la construction C1 du virus HIV-1 (Damier *et al*., 1997) dans laquelle l'élément NRS WT et NRS-Inv ont été insérés (C1-NRS).

**B :** Représentation de la construction C2 du virus HIV-1 (Damier *et al*., 1997) dans laquelle l'élément NRS WT et NRS-Inv ont été insérés (C2-NRS).

**C :** Représentation de la construction Minx de l'adénovirus (Zillman *et al*., 1988) dans laquelle l'élément NRS WT et NRS-Inv ont été insérés (Minx-NRS).

Les exons sont représentés par des rectangles, les introns par des traits. Les sites donneurs (D), accepteurs (A) et leurs positions sont mentionnés. Le site d'insertion des éléments NRS WT et Inv est symbolisé par les traits en pointillés. L'éléments NRS WT est représenté par le dessin de sa structure secondaire qui a été retournée pour représenter l'élément NRS Inv. Le site *Bgl*II représente la jonction des fragments d'ADNc.

Nous avons inséré la séquence de l'ADNc de l'élément NRS WT (fournie par K. Beemon) dans la séquence correspondant à l'intron des constructions C1, C2 et Minx, ceci en tenant compte des contraintes de distance requises entre le site donneur d'épissage et l'élément NRS pour que celui-ci exerce un effet inhibiteur sur l'épissage (Wang et McNally, 2003). L'élément NRS a été inséré de façon à ce que la séquence de son pseudo site se situe 278 nucléotides en aval du site donneur (**Figure X-1**). L'insertion de l'élément NRS qui mesure 265 nucléotides provoquait une augmentation de la taille des introns des différents pré-ARNm. Or, on sait que l'efficacité d'épissage *in vitro* d'un pré-ARNm diminue si la taille de l'intron

augmente de manière importante. Afin de contourner ce biais, nous avons tronqué les introns des constructions C1 et C2 de part et d'autre du site d'insertion de l'élément NRS afin de pas augmenter de façon importante leur longueur initiale une fois l'élément NRS inséré. Nous avons également réalisé une construction servant de contrôle où nous avons inséré à la même position que l'ARN NRS WT sa séquence complémentaire, l'élément NRS Inv, qui n'est pas fonctionnel (**Figure X-1**) (Gontarek *et al.*, 1993 ; McNally *et al.*, 1991). La réalisation des différentes constructions pour la réalisation des tests d'épissage est décrite dans la partie Matériel et Méthodes (**Chapitre V**).

Toutes ces constructions d'épissage utilisées sont insérées dans des plasmides en aval d'un promoteur T7 et en amont d'un site de coupure unique par une enzyme de restriction. Après linéarisation du plasmide par cette enzyme, les pré-ARNm sont transcrits *in vitro* selon le protocole décrit dans la partie Matériel et Méthodes.

### II.2. Expériences d'épissage réalisées avec des pré-ARN uniformément marqués

Dans un premier temps, nous avons réalisé des expériences d'épissage *in vitro* avec des pré-ARNm uniformément marqués. Après la linéarisation des différentes constructions, les pré-ARNm ont été produits par transcription *in vitro* en présence d'UTP [$\alpha$-$^{32}$P] selon le protocole décrit dans la partie Matériel et Méthodes (**Chapitre V**). Dix fmol de chaque pré-ARNm ont été incubés en présence d'extrait nucléaire de cellules HeLa et l'épissage a été réalisé durant 2 h à 30°C dans les conditions détaillées dans la partie Matériel et Méthodes (**Chapitre V**). Les produits de la réaction d'épissage ont ensuite été fractionnés par une électrophorèse en gel de polyacrylamide à 5% dénaturant et les ARN ont été révélés par autoradiographie (**Figure X-2**).

Les pré-ARNm C1, C2 et Minx étaient épissés efficacement comme l'indiquent les intensités des ARNm correspondants (**Figure X-2, A pistes 1, 4 et B, piste 1, respectivement**). En revanche, aucun produit d'épissage n'était visible lorsque l'élément NRS était présent dans l'intron de ces pré-ARNm, et ce, indépendamment de son orientation (**Figure X-2, A pistes 2, 3, B, pistes 2, 3**). Ces expériences ont été répétées plusieurs fois avec des lots d'ARN et d'extrait nucléaire différents et des résultats identiques ont été obtenus. Dans le cas du pré-ARNm C1, cette absence de détection de produit d'épissage pouvait être due à l'augmentation de la taille de l'intron provoquée par l'insertion de l'élément NRS WT ou Inv qui pouvait être un facteur limitant de l'efficacité d'épissage *in vitro*. Gontarek *et al.*, (Gontarek *et al.*, 1993) avait rencontré le même problème en insérant l'élément NRS entier (nt 631-1006) dans un pré-ARNm dérivé de l'adénovirus dont l'épissage *in vitro* était inhibé

indépendamment de l'orientation de l'élément NRS. Par contre, ce pré-ARNm était bien épissé lorsque l'élément NRS entier était remplacé par l'élément NRS tronqué Δ76 dans son orientation antisens (Gontarek *et al.*, 1993). Par ailleurs, il est aussi possible que nous ayons éliminé des éléments activateurs de l'épissage en tronquant l'intron du pré-ARNm C2 dans le pré-ARNm C2-NRS afin de conserver sa taille initiale (**Figure X-1**).

Enfin, une autre possibilité était que le taux d'épissage des pré-ARNm contenant l'élément NRS soit trop faible pour que les produits d'épissage puissent être détectés directement par autoradiographie et qu'une méthode plus sensible telle que la RT-PCR nous permette de les mettre en évidence et de les quantifier. Nous avons donc décidé de mettre au point cette approche pour la détection.

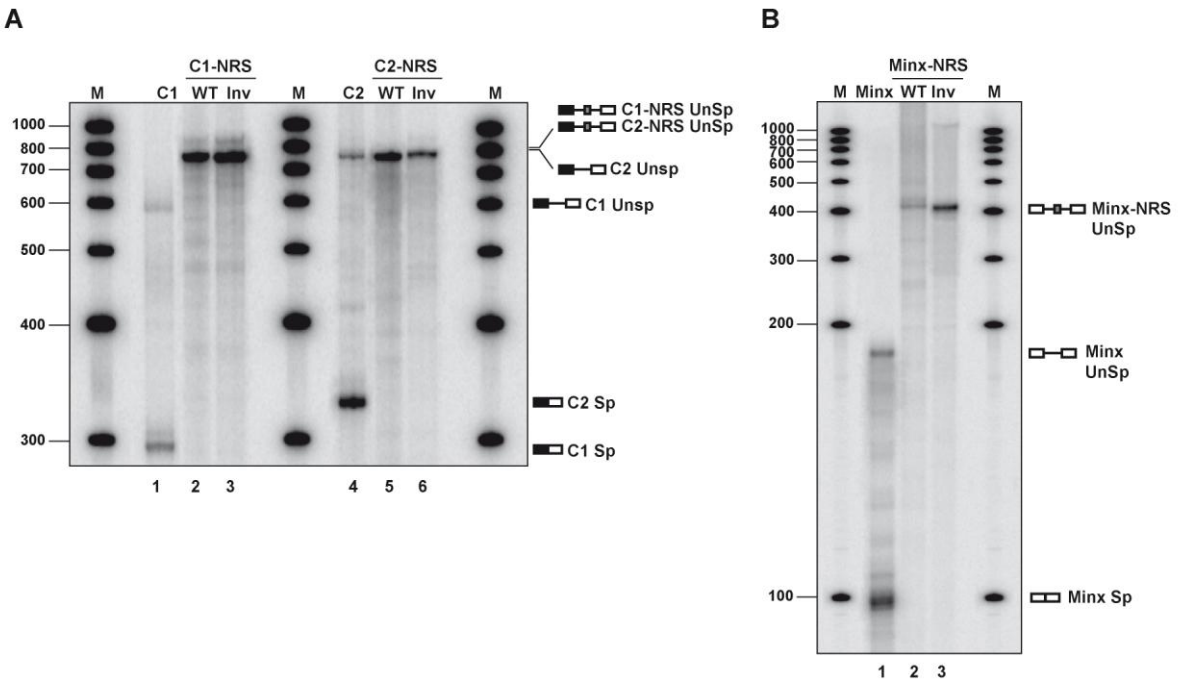



**Figure X-2 : Expérience d'épissage *in vitro* avec des pré-ARNm uniformément marqués.**
Les produits d'épissage ont été fractionnés sur un gel de polyacrylamide dénaturant 5% et révélés par autoradiographie.
**A :** Fractionnement des produits des réactions d'épissage réalisées avec les pré-ARNm C1 (piste 1) et C2 (piste 4) contenant l'élément NRS dans l'orientation sens (C1-NRS WT, piste 2, C2-NRS WT, piste 5) et antisens (C1-NRS Inv, piste 3, C2-NRS Inv, piste 6). Les positions des pré-ARNm (UnSp) et des ARNm (Sp) sont indiquées à droite des autoradiogrammes. Le premier et le deuxième exon des ARN C1 et C2 sont respectivement schématisés par des rectangles noirs et blancs, et le NRS par un rectangle gris. Le nombre de bases de chacun des fragments d'ADN du marqueur de taille (piste M) est indiqué à gauche de l'autoradiogramme.
**B :** Fractionnement des produits des réactions d'épissage réalisées avec le pré-ARNm Minx (piste 1) contenant l'élément NRS dans l'orientation sens (Minx-NRS WT, piste 2) et antisens (Minx-NRS Inv, piste 3). Les positions des pré-ARNm (UnSp) et des ARNm (Sp) sont indiquées à droite des autoradiogrammes. Le premier et le deuxième exon de l'ARN Minx sont chacun schématisés par un rectangle blanc et le NRS par un rectangle gris. Le nombre de bases de chacun des fragments d'ADN du marqueur de taille (piste M) est indiqué à gauche de l'autoradiogramme.

## II.3. Expériences d'épissage réalisées avec des pré-ARNm non radioactifs et mise en évidence des produits d'épissage par RT-PCR

### II.3.1. Mise au point des conditions de RT-PCR

#### II.3.1.7. Choix des constructions

Pour mettre au point les conditions de la réaction de RT-PCR, nous avons choisi les pré-ARNm C2, Minx et les pré-ARNm C2-NRS et Minx-NRS (**Figure X-3**). Nous avons également utilisé un pré-ARNm du gène de la GGT humaine dont la construction avait été réalisée au laboratoire par A. Visvikis (**Figure X-3, C**). Ce pré-ARNm nous a servi de témoin car son épissage avait déjà été analysé par RT-PCR.

#### II.3.1.8. Conditions d'épissage

Les réactions d'épissage ont été réalisées dans les mêmes conditions que précédemment, à l'exception des matrices ARN qui ont été transcrites en l'absence de nucléotide radiomarqué selon le protocole décrit dans la partie Matériel et Méthodes (**Chapitre V**). Les réactions d'épissage ont été effectuées en utilisant deux quantités différentes de pré-ARNm de départ. Nous avons choisi d'utiliser 10 fmol de pré-ARNm qui est la même quantité que celle utilisée pour les réactions d'épissage réalisées avec des pré-ARNm uniformément marqués, ainsi que 20 fmol afin de tester la sensibilité de la détection des produits d'épissage par RT-PCR. Un contrôle a été réalisé avec 10 fmol de pré-ARNm sans ajout d'extrait nucléaire.

#### II.3.1.9. Choix de l'extrait nucléaire et des oligonucléotides de la réaction de PCR

##### II.3.1.9.1. Oligonucléotides permettant l'amplification simultanée des espèces épissées (Sp) et non épissées (UnSp)

A la fin de la réaction d'épissage, un seul oligonucléotide qui s'hybride à l'extrémité 3' du deuxième exon a été utilisé pour convertir les ARN non épissés (pré-ARNm) et les produits épissés (ARNm) en ADNc (**Figure X-3**). Les ADNc obtenus ont ensuite été amplifiés au cours d'une réaction de PCR comprenant 35 cycles d'amplification.

Dans un premier temps, nous avons choisi d'utiliser un seul couple d'oligonucléotides de façon à amplifier simultanément les espèces épissées (Sp) et non épissées (UnSp) (**Figure X-3**). Les produits PCR correspondant aux ARN épissés et non épissés ont une longueur suffisamment différente pour être fractionnés distinctement par électrophorèse en gel d'agarose

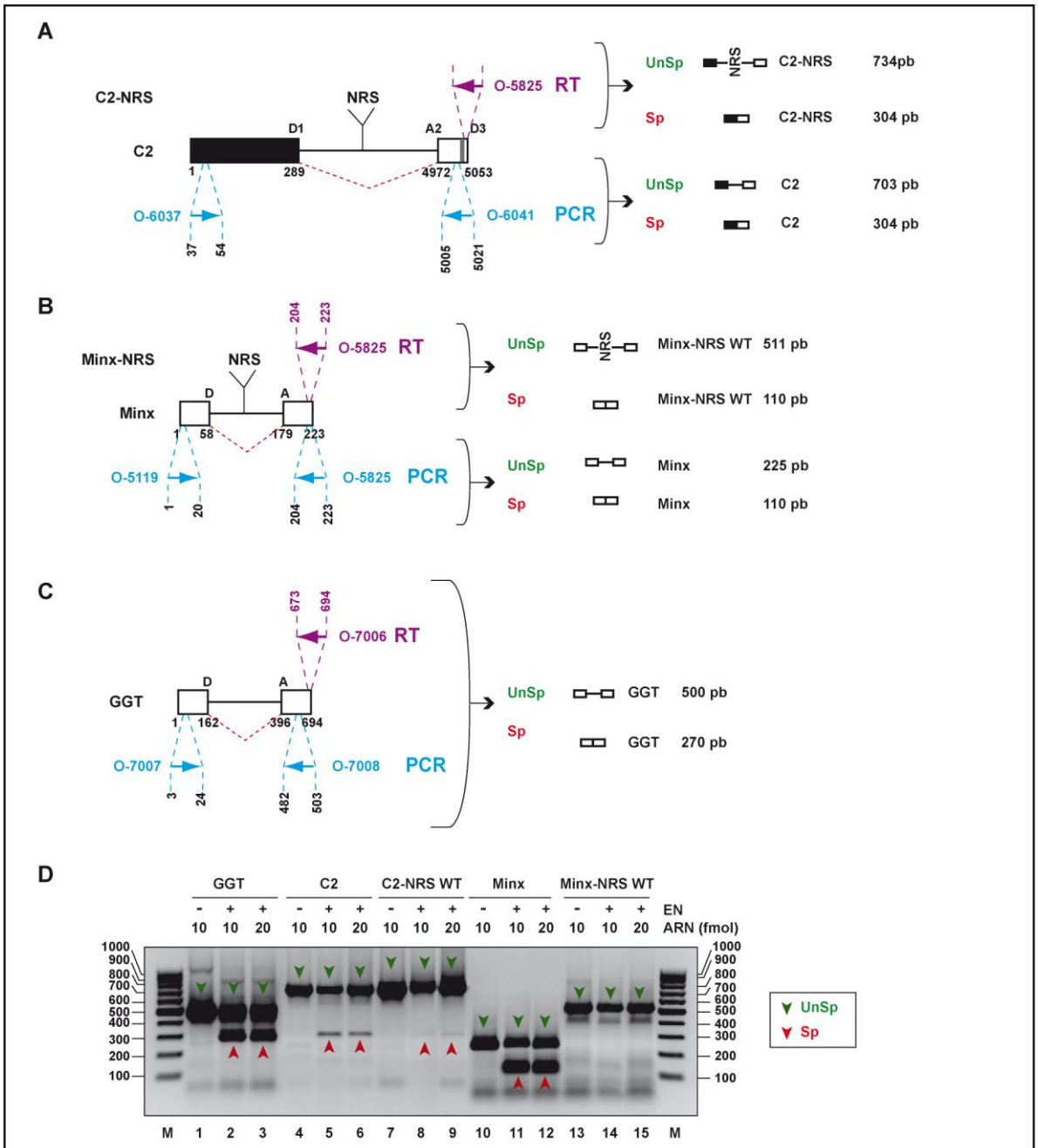



**Figure X-3 : Expérience d'épissage *in vitro* et identification des produits d'épissage par RT-PCR.**
**A-C :** Représentation des sites d'hybridation des oligonucléotides utilisés lors de la réaction de RT-PCR sur les différents pré-ARNm. L'oligonucléotide utilisé pour la réaction de transcription inverse (RT) est illustré par une flèche violette, le couple d'oligonucléotides utilisé pour la réaction de PCR par des flèches bleues. La séquence d'hybridation de chaque oligonucléotide est précisée. Les exons sont représentés par des rectangles, l'intron par une ligne noire. Les sites donneur (D) et accepteur (A) sont mentionnés et leur épissage est symbolisé par un trait rouge en pointillé. La présence de l'élément NRS dans les pré-ARNm C2-NRS et Minx-NRS est représenté au-dessus du pré-ARNm. La taille des produits non épissés (UnSp) et épissés (Sp) obtenus par RT-PCR est mentionnée à droite des pré-ARNm.
**D :** Analyse par électrophorèse en gel d'agarose des produits d'épissage amplifiés par RT-PCR. Les réactions d'épissage ont été réalisées en présence (+) et en absence d'extrait nucléaire (-). Le nom et la quantité des pré-ARNm utilisés lors des réactions d'épissage sont mentionnés au-dessus des pistes. Les produits PCR correspondant aux pré-ARNm (UnSp) et aux ARNm (Sp) sont respectivement indiqués par des flèches vertes et rouges. Le nombre de paires de bases des fragments du marqueur de taille (piste M) est indiqué à gauche et à droite du gel.

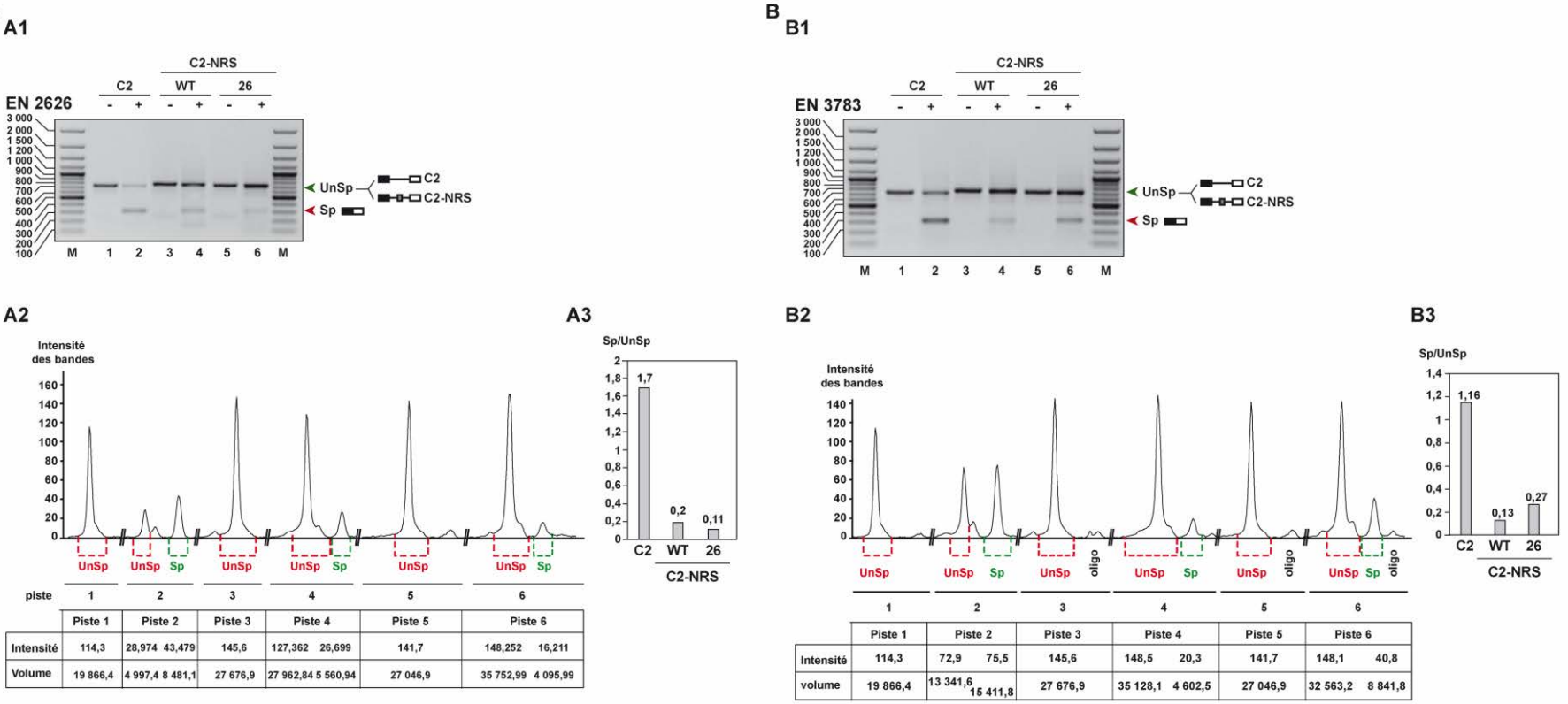
Les résultats préliminaires obtenus sont présentés dans la **Figure X-3**.

Pour l'ARN GGT nous observions la présence de deux bandes d'une taille de 500 et 270 pb environ qui correspondaient bien aux tailles attendues des fragments amplifiés à partir du pré-ARNm (UnSp) et de l'ARNm (Sp) (**Figure X-3, piste 2 et 3**). Ces résultats indiquaient que la réaction d'épissage avait bien fonctionné et que la réaction de RT-PCR avait permis de détecter les produits de l'épissage. Les mêmes résultats étaient obtenus pour les ARN C2 et Minx (**Figure X-3, pistes 5-6, 11-12, respectivement**). Néanmoins, les intensités des bandes détectées à la fois pour les expériences réalisées avec 10 et 20 fmol de pré-ARNm de départ variaient peu dans les deux cas, ce qui indiquait que nous n'étions pas dans des conditions de RT-PCR qui permettaient d'avoir une quantification fiable.

Pour l'ARN C2-NRS WT, deux bandes d'environ 700 et 300 pb qui correspondaient respectivement au fragment amplifié à partir du pré-ARNm (UnSp) et à celui amplifié à partir de l'ARNm (Sp) étaient détectées (**Figure X-3, pistes 8 et 9**). Cela démontrait qu'il y avait bien un épissage du pré-ARNm C2-NRS WT et que l'approche de RT-PCR permettait de visualiser les produits d'épissage. En comparant les intensités des bandes Sp/UnSp de l'ARN C2 avec celles de l'ARN C2-NRS WT, nous pouvions dire que l'insertion de l'élément NRS dans le pré-ARNm C2 provoquait une inhibition de son épissage (**Figure X-3, pistes 5, 6 vs 8, 9**). Par conséquent, l'ARN C2-NRS WT semblait être un bon modèle pour l'étude de l'inhibition de l'épissage par le NRS. Ceci n'était en revanche pas le cas de l'ARN Minx-NRS pour lequel aucune bande de la taille correspondant aux ARNm n'était observée (**Figure X-3, pistes 14 et 15**).

#### II.3.1.9.2. Choix de l'extrait nucléaire

Ces résultats préliminaires nous ont permis de montrer que l'élément NRS inhibait l'épissage lorsqu'il était inséré dans l'intron du pré-ARNm C2. Néanmoins, les signaux obtenus pour les ARN non épissés (UnSp) et les ARN épissés (Sp) étaient respectivement très forts et très faibles, ce qui empêchait toute quantification précise et fiable, ceci pouvant être dû en partie à l'extrait nucléaire utilisé. L'expérience de notre laboratoire dans le domaine de l'épissage montre que l'efficacité des réactions d'épissage *in vitro* est très différente suivant les lots d'extrait nucléaire utilisés. Nous avons donc testé plusieurs lots d'extrait nucléaire afin de choisir celui qui permettait d'obtenir la meilleure efficacité d'épissage. De plus, nous avons diminué le nombre des cycles des réactions de PCR à 30 afin d'obtenir des signaux quantifiables. Pour ces expériences, nous avons utilisé, en plus des pré-ARNm C2 et C2-




| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 |
|---|---|---|---|---|---|---|
| Intensité | 114,3 | 28,974 43,479 | 145,6 | 127,362 26,699 | 141,7 | 148,252 16,211 |
| Volume | 19 866,4 | 4 997,4 8 481,1 | 27 676,9 | 27 962,84 5 560,94 | 27 046,9 | 35 752,99 4 095,99 |

| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 |
|---|---|---|---|---|---|---|
| Intensité | 114,3 | 72,9 75,5 | 145,6 | 148,5 20,3 | 141,7 | 148,1 40,8 |
| volume | 19 866,4 | 13 341,6 15 411,8 | 27 676,9 | 35 128,1 4 602,5 | 27 046,9 | 32 563,2 8 841,8 |

**Figure X-4 : Expériences d'épissage *in vitro* avec les pré-ARNm C2, C2-NRS WT et C2-NRS 26 réalisées avec les extraits nucléaires N° 2626 (A) et N°3783 (B) et analyse des produits d'épissage obtenus par RT-PCR.**

**A1 et B1 :** Analyse par électrophorèse en gel d'agarose des produits d'épissage amplifiés par RT-PCR. Les réactions d'épissage ont été réalisées en présence (+) et en absence (-) d'extrait nucléaire N°2626 (A1) et d'extrait nucléaire N°3783 (B1). Les noms des pré-ARNm utilisés lors des réactions d'épissage sont mentionnés au-dessus des pistes. Les produits PCR correspondant aux pré-ARNm (UnSp) et aux ARNm (Sp) sont respectivement indiqués par des flèches vertes et rouges. Le nombre de paires de bases des fragments du marqueur de taille (piste M) est indiqué à gauche et à droite du gel.

**A2 et B2 :** Quantification de l'intensité des bandes du gel d'agarose présenté en A1) et B1). Les intensités des bandes sont mesurées et reportées sous forme de pic délimité par des pointillés sur le chromatogramme. Le tableau comprend, pour chaque pic, la valeur de l'intensité maximale et le volume correspondant à l'intégration de l'intensité de tous les pixels. Les produits de RT-PCR correspondant aux produits non épissés (UnSp) et épissés (Sp) et les oligonucléotides libres sont reportés en dessous des pics correspondants.

**A3 et B3 :** Représentation graphique des ratios Sp/UnSp calculés pour chacun des ARN à partir des valeurs des volumes des pics présentés en A2) et B2). Les valeurs des ratios Sp/UnSp sont indiquées au-dessus de chaque histogramme.

NRS-WT, le mutant C2-NRS-26 théoriquement non fonctionnel en inhibition d'épissage. Les expériences d'épissage, de RT-PCR et l'analyse des résultats ont été effectuées comme décrit auparavant.

Parmi les résultats obtenus avec les différents lots d'extrait nucléaire que nous avons testés, nous ne présenterons que ceux obtenus avec les lots N° 2626 (**Figure X-4, A**) et N°3783 (**Figure X-4, B**) qui sont représentatifs de l'ensemble des tests. Par rapport au pré-ARNm C2, l'insertion de l'élément NRS dans l'intron du pré-ARNm C2 inhibait très fortement l'épissage et ce, quelque soit l'extrait nucléaire utilisé (**Figure X-4, A1 et B1 piste 4 vs piste 2**). En revanche, le taux d'épissage du mutant NRS-26 différait selon l'extrait nucléaire utilisé. Ce mutant théoriquement non fonctionnel, aurait dû restaurer, au moins partiellement, le taux d'épissage et avoir un ratio Sp/UnSp supérieur à celui mesuré pour l'élément NRS WT. Or, cela n'était observé que lorsque l'extrait N°3783 était utilisé (**Figure X-4, A3 vs B3**). L'utilisation de l'extrait N°2626 (ainsi que plusieurs autres lots d'extrait nucléaire) ne nous permettait donc pas de voir les différences d'efficacité d'inhibition de l'épissage attendues entre l'élément NRS WT et le mutant NRS-26. L'extrait nucléaire N°3783 semblait donc être le plus approprié à utiliser dans notre étude.

Sachant que les protéines SR jouent un rôle important aussi bien dans l'épissage du pré-ARNm C2 que dans l'inhibition de l'épissage par l'élément NRS, les différences que nous avons observées en épissage pouvaient être dues à des variations de quantité de ces protéines dans les deux lots d'extrait nucléaire. Pour le vérifier, nous avons réalisé des Western blots sur une quantité d'extrait nucléaire correspondant à 2 µg de protéines totales des extraits nucléaires N°3783 et N°2626, avec les anticorps disponibles au laboratoire dirigés contre certaines protéines hnRNP et SR. Les résultats montraient que les deux extraits nucléaires contenaient des quantités équivalentes pour chacune des protéines hnRNP recherchées mais des quantités différentes pour chacune des protéines SR testées (**Figure X-5**). L'extrait nucléaire N°2626 contenait davantage de protéine ASF/SF2 et SC35 mais moins de protéine 9G8 que l'extrait N°3783. Les différences observées pouvaient expliquer les différences de réponses obtenues en épissage.

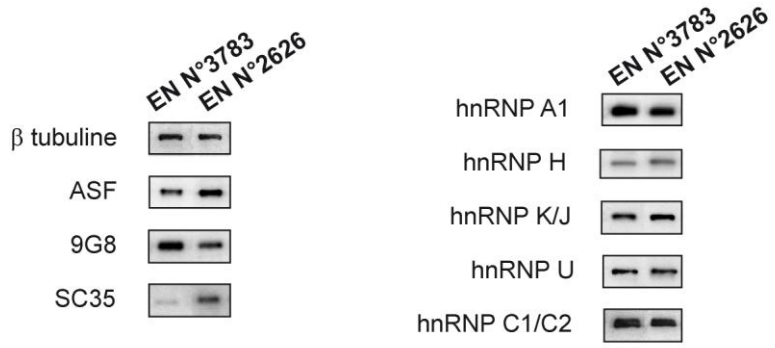



**Figure X-5 : Comparaison par Western blots des quantités de certaines protéines SR et hnRNP contenues dans les 2 lots d'extrait nucléaire N°2626 et N° 3783 utilisés lors des tests d'épissage *in vitro*.**
Les expériences de Western blots ont été réalisées avec une quantité d'extrait nucléaire correspondant à 2 µg de protéines totales des extraits nucléaires N°2626 et 3783. Le nom de la protéine reconnue spécifiquement par l'anticorps utilisé est indiqué à gauche des fluorographies.

L'ensemble des résultats décrits ci-dessus nous indiquait que l'inhibition de l'épissage par l'élément NRS était trop importante dans le contexte du pré-ARNm C2 pour que nos conditions de PCR puissent amplifier les fragments correspondant aux ARNm (Sp) à des quantités suffisantes pour permettre leur quantification sur un gel d'agarose. Nous supposions que dans ces conditions, les fragments correspondant aux pré-ARNm (UnSp) entraient en compétition avec les fragments correspondant aux produits Sp pour l'hybridation avec les amorces lors de la PCR. Le fait que les produits UnSp soient majoritaires en raison de l'inhibition exercée par l'élément NRS pouvait ainsi empêcher l'amplification des produits Sp. Nous avons donc décidé de mettre au point une autre stratégie permettant l'amplification et la détection individuelle des espèces Sp d'une part et des espèces Sp+UnSp d'autre part lors de deux réactions de PCR distinctes.

### II.3.1.9.3. Amplification des espèces épissées (Sp) d'une part, et des espèces épissées et non épissées (Sp+UnSp) d'autre part, avec deux couples d'oligonucléotides distincts

Pour les raisons mentionnées ci-dessus, nous avons fait le choix par la suite d'utiliser deux couples d'oligonucléotides, chaque couple étant utilisé lors d'une réaction de PCR :

- L'amplification des produits Sp : était réalisée avec un oligonucléotide qui chevauche la bordure entre le premier et second exon (O-6246, **Figure X-6**) et un deuxième oligonucléotide qui s'hybride à l'extrémité 5' du premier exon (O-6519, **Figure X-6**).

- L'amplification des produits Sp+UnSp : était réalisée avec un couple oligonucléotides qui s'hybride sur le premier exon (O-6519 et O-6187, **Figure X-6**), l'exon 1 étant présent à la fois dans le pré-ARNm et l'ARNm.

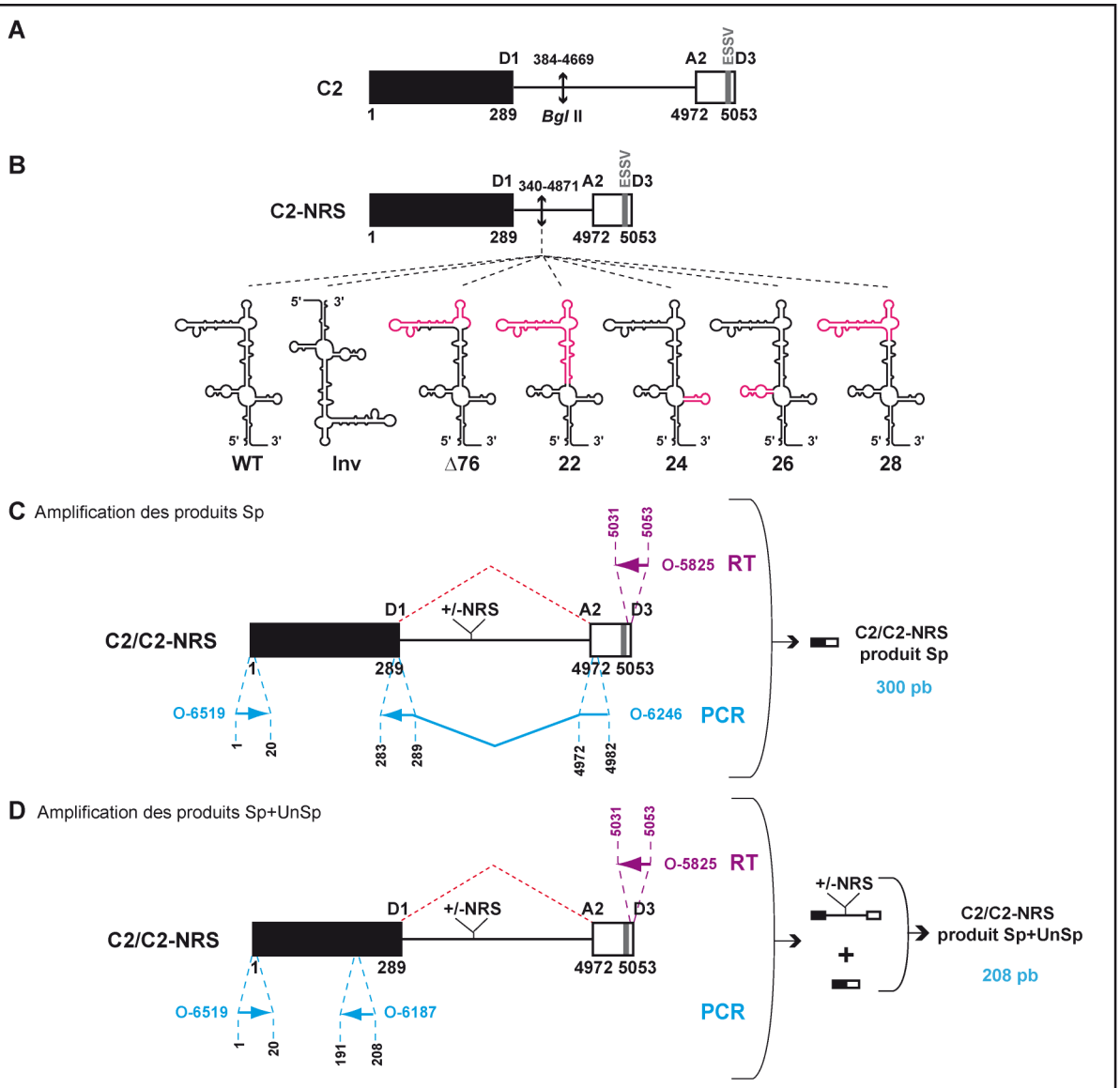



**Figure X-6 : Amplification par RT-PCR des espèces épissées (Sp) d'une part, et des espèces épissées et non épissées (Sp+UnSp) d'autre part, avec deux couples d'oligonucléotides distincts.**

**A et B :** Représentation des pré-ARNm C2 (A) et C2-NRS (B). Les exons sont représentés par des rectangles, l'intron par une ligne noire. Les sites donneur (D) et accepteur (A) sont mentionnés, le site d'insertion des différents éléments NRS dans le pré-ARNm C2 est représenté par une double flèche et les séquences tronquées dans les différents mutants de l'élément NRS sont en rose.

**C :** Représentation des sites d'hybridation des oligonucléotides utilisés lors de la réaction de RT-PCR pour amplifier les produits épissés (Sp). L'oligonucléotide utilisé pour la réaction de transcription inverse (RT) est illustré par une flèche violette, le couple d'oligonucléotides utilisé pour la réaction de PCR, par des flèches bleues. Les sites d'hybridation de chaque oligonucléotide sont précisés. Les exons sont représentés par des rectangles, l'intron par une ligne noire. Les sites donneur (D) et accepteur (A) sont mentionnés et leur utilisation lors de l'épissage est symbolisée par le trait rouge en pointillé. La taille des produits obtenus par RT-PCR (Sp) est mentionnée.

**D :** Représentation des sites d'hybridation des oligonucléotides utilisés lors de la réaction de RT-PCR pour amplifier les produits épissés et non épissés (Sp+UnSp). L'oligonucléotide utilisé pour la réaction de transcription inverse (RT) est illustré par une flèche violette, le couple d'oligonucléotides utilisé pour la réaction de PCR, par des flèches bleues. Les sites d'hybridation de chaque oligonucléotide sont précisés. Les exons sont représentés par des rectangles, l'intron par une ligne noire. Les sites donneur (D) et accepteur (A) sont mentionnés et leur utilisation lors de l'épissage est symbolisée par le trait rouge en pointillé. La taille des produits obtenus par RT-PCR (Sp+UnSp) est mentionnée.

La réaction de PCR des produits Sp comportait 35 cycles d'amplification et celle des produits Sp+UnSp en comportaient 30. Les températures des différents paliers étaient identiques dans les deux cas.

Afin de valider la méthode, il était important de s'assurer que les intensités des bandes correspondant aux produits Sp et Sp+UnSp obtenues par RT-PCR étaient proportionnelle aux quantités d'ARNm et de pré-ARNm présentes à la fin de la réaction d'épissage.

### II.3.1.10. Linéarité des amplifications par RT-PCR (Figure X-7)

Une première expérience a été réalisée pour démontrer la linéarité des amplifications par PCR des produits Sp correspondant aux ARNm obtenus après la réaction d'épissage. Après une réaction d'épissage *in vitro*, les ARN ont été rétrotranscrits lors d'une réaction de transcription inverse et une gamme de différents volumes de cette réaction allant de 0 à 1,5 µl a été utilisée pour des réactions de PCR avec le couple d'oligonucléotides amplifiant spécifiquement les produits épissés (Sp). Cette expérience a été réalisée avec les pré-ARNm C2-NRS WT et C2-NRS Inv car des résultats préliminaires indiquaient que l'amplification de leurs produits épissés (Sp) donnait des bandes de faible et forte intensités, respectivement (**Figure X-7**).

Une autre expérience a été réalisée pour démontrer que les amplifications par PCR de la somme des produits épissés et non épissés (Sp+UnSp) étaient linéaires. Pour cela, nous avons utilisé une gamme de différentes quantités de pré-ARNm C2 allant de 0 à 2 fmol qui ont été rétrotranscrites lors de réactions de transcription inverse. Un volume de 1 µl de cette réaction a ensuite été utilisé pour la réaction de PCR. Les quantités de pré-ARNm de cette gamme ont été choisies en fonction du protocole d'épissage. L'épissage *in vitro* est réalisé au départ avec 10 fmol d'ARN. Après la réaction d'épissage et après les étapes de déprotéinisation avec la protéinase K et l'extraction au phénol-chloroforme, ces ARN se retrouvent dans un volume de 200 µl dont 140 µl sont récupérés, ce qui correspond à environ 7 fmol d'ARN. Les ARN sont ensuite précipités et repris dans un volume de 20 µl et 2 µl de ce volume sont utilisés pour la réaction de transcription inverse, ce qui représente au final 0,7 fmol d'ARN, ARNm et pré-ARNm confondus. Par conséquent, nous avons choisi d'utiliser une gamme de 0,1 ; 0,5 ; 1 ; 2 fmol de pré-ARNm C2 pour démontrer la linéarité de l'amplification des produits Sp+UnSp par PCR (**Figure X-7, A3, B3, C3**).

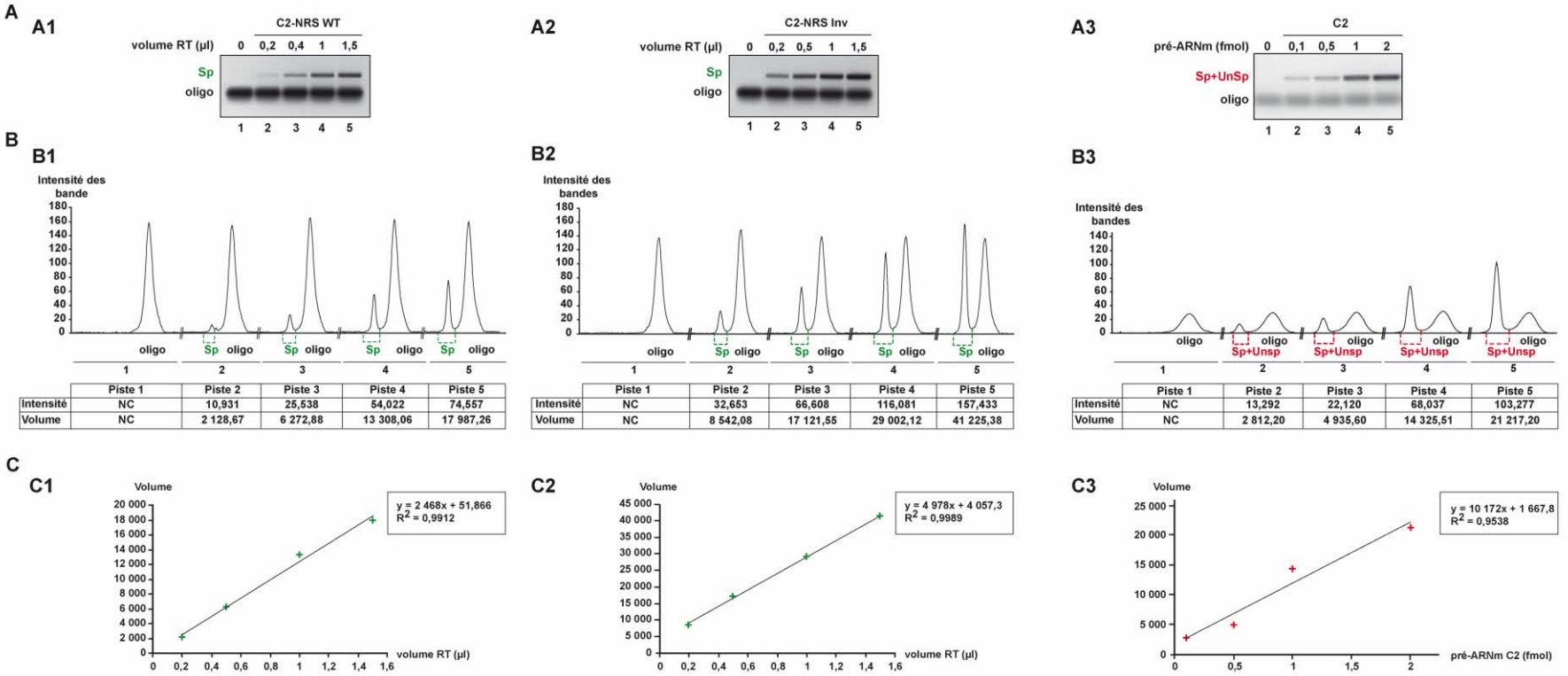



**Figure X-7 : Linéarité de l'amplification par PCR des produits d'épissage.**

Des réactions d'épissage ont été réalisées avec 10 fmol de pré-ARNm C2-NRS WT et C2-NRS Inv et les produits d'épissage ont été rétrotranscrits lors d'une expérience de transcription inverse. Une gamme de volumes allant de 0 à 1,5 µl du mélange réactionnel de cette transcription inverse a ensuite été utilisée pour des amplifications PCR avec le couple d'oligonucléotides amplifiant spécifiquement les produits correspondant aux ARNm (Sp). Une gamme de 0 à 2 fmol de pré-ARNm C2 qui n'a pas été épissé, a été utilisée pour les réactions de PCR afin d'amplifier les produits correspondant aux produits épissés et non épissés (Sp+UnSp).

**A :** Electrophorèses en gel d'agarose des produits de RT-PCR. Les fragments correspondent aux produits épissés (Sp) du pré-ARNm C2-NRS WT (A1), C2-NRS Inv (A2) ou aux produits épissés plus non épissés (Sp+UnSp) du pré-ARNm C2 (A3). Le volume de la transcription inverse (A1, A2), la quantité de pré-ARNm (A3) et les noms des pré-ARNm utilisés lors des expériences sont indiqués au dessus des pistes. Les positions des produits épissés (Sp) et épissés plus non épissés (Sp+UnSp) et des oligonucléotides libres sont mentionnés à gauche des gels.

**B :** Quantification de l'intensité des bandes correspondant aux produits épissés (Sp) et aux produits épissés plus non épissés (Sp+UnSp) des gel d'agarose présentés en A). Les chromatogrammes B1, B2, B3 correspondent aux quantifications des bandes des gels d'agarose A1, A2 et A3, respectivement. Les intensités des bandes sont mesurées et reportées sous forme de pic délimité par des pointillés sur le chromatogramme. Le tableau comprend pour chaque pic, la valeur de l'intensité maximale et le volume correspondant à l'intégration de l'aire du pic. Les produits de RT-PCR correspondant aux produits épissés (Sp), aux produits épissés plus non épissés (Sp+UnSp) et les oligonucléotides libres (oligo) sont reportés en dessous des pics correspondants.

**C :** Les volumes des pics des chromatogrammes présentés en B) ont été reportés sur un graphique en fonction du volume de RT utilisé (A1, A2) ou de la quantité de pré-ARNm utilisée (A3). Les graphiques C1, C2, C3 correspondent aux reports des valeurs des chrommatogrammes B1, B2 et B3, respectivement. L'équation de la droite ainsi que le coefficient de régression linéaire sont encadrés.

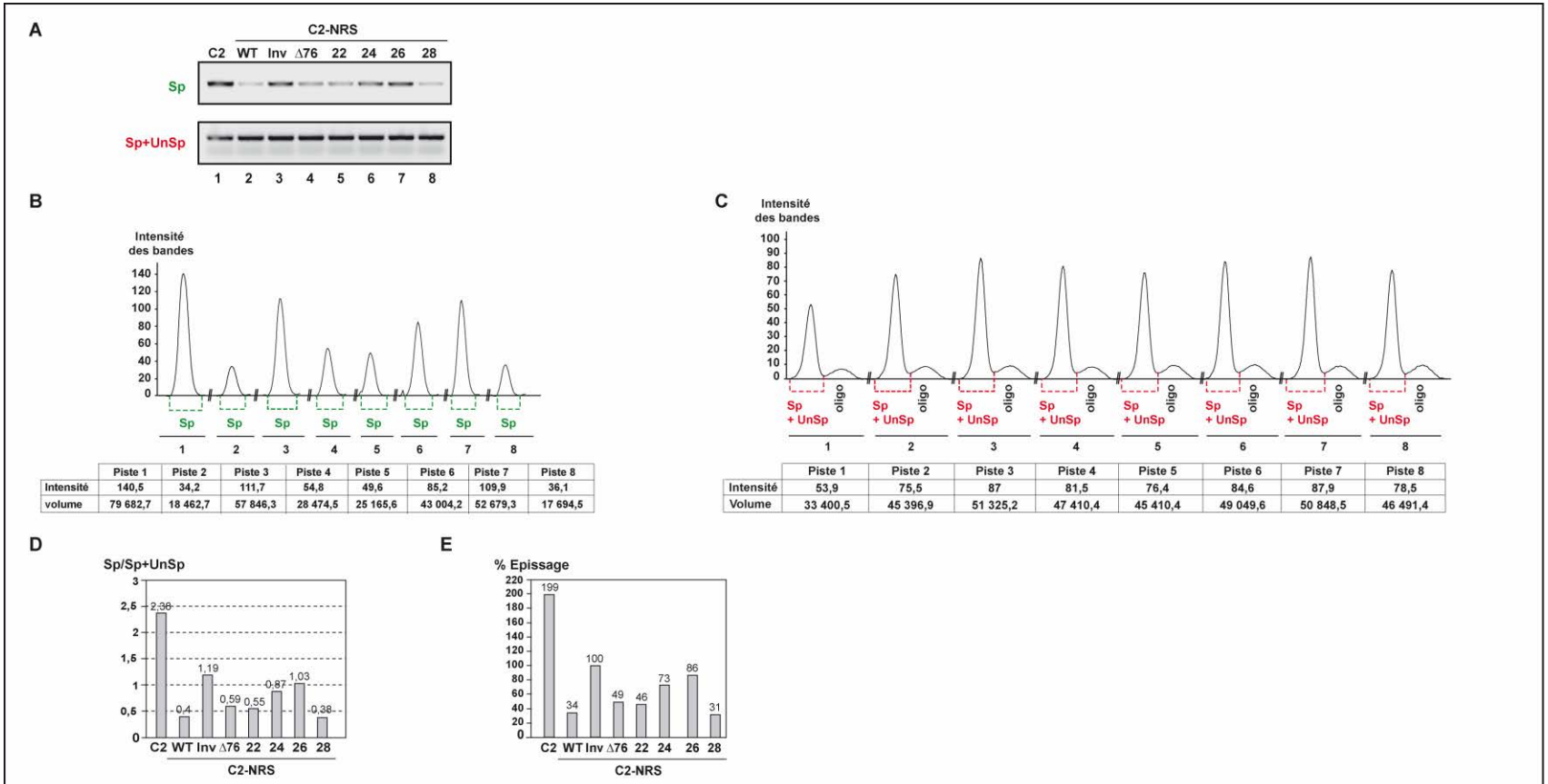
Les produits de RT-PCR ont été fractionnés par une électrophorèse en gel d'agarose puis quantifiés. Les valeurs des quantifications des bandes correspondant aux produits Sp ou Sp+UnSp ont ensuite été reportées sur un graphique en fonction du volume de réaction de transcription inverse et de quantité de pré-ARNm C2, respectivement (**Figure X-7**). Dans les 3 expériences, les valeurs des quantifications peuvent être alignées sur une droite dont le coefficient de corrélation est proche de 1 (**Figure X-7, C**). Cela signifie que les amplifications par PCR sont linéaires pour les gammes de volumes de la réaction de transcription inverse et les gammes de quantités de pré-ARNm utilisées. Sachant que nos conditions de RT-PCR se situent dans cette zone de linéarité, nous pouvons en conclure que les intensités des bandes Sp et Sp+UnSp que nous obtenons après les réactions de RT-PCR sont proportionnelles aux quantités d'ARNm et de pré-ARNm présentes à la fin de la réaction d'épissage.

L'ensemble des résultats de la mise au point des conditions d'épissage *in vitro* et de détection montre que la méthode de RT-PCR que nous avons développée est fiable et adaptée pour l'étude fonctionnelle de l'élément NRS.

### II.3.2. Etude fonctionnelle de l'élément NRS par expériences d'épissage *in vitro*

Les réactions d'épissage *in vitro* étaient réalisées avec le pré-ARNm C2 et les pré-ARNm C2-NRS comprenant tous les mutants de l'élément NRS (**Figure X-6**). Après les réactions de RT-PCR réalisées selon les conditions décrites ci-dessus, les produits d'amplification étaient fractionnés par électrophorèse en gel d'agarose en présence de bromure d'éthidium. Les intensités des bandes obtenues étaient quantifiées et les rapports Sp/Sp+UnSp calculés. Les résultats obtenus sont présentés dans la **Figure X-8**.

Les résultats montraient des différences significatives entre les différents mutants de l'élément NRS. Les pré-ARNm C2-NRS 24, C2-NRS 26 et C2-NRS Inv étaient épissés à des taux bien supérieurs à ceux des pré-ARNm C2-NRS Δ76, C2-NRS 22 et C2-NRS 28. En accord avec la littérature, le taux d'épissage du pré-ARNm C2-NRS Δ76 était similaire à celui du pré-ARNm C2-NRS WT. L'inhibition de l'épissage des mutants NRS 22 et NRS 28 n'avait jamais été examinée auparavant, ce qui nous a amené à proposer un nouvel élément NRS minimal fonctionnel *in vitro* comportant seulement l'hélice H1 et les tiges boucles SLS1 et SLS4. L'ensemble de ces résultats sont présentés et discutés plus en détail dans l'article scientifique joint (**Chapitre XI**).




| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 | Piste 7 | Piste 8 |
|---|---|---|---|---|---|---|---|---|
| Intensité | 140,5 | 34,2 | 111,7 | 54,8 | 49,6 | 85,2 | 109,9 | 36,1 |
| volume | 79 682,7 | 18 462,7 | 57 846,3 | 28 474,5 | 25 165,6 | 43 004,2 | 52 679,3 | 17 694,5 |

| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 | Piste 7 | Piste 8 |
|---|---|---|---|---|---|---|---|---|
| Intensité | 53,9 | 75,5 | 87 | 81,5 | 76,4 | 84,6 | 87,9 | 78,5 |
| Volume | 33 400,5 | 45 396,9 | 51 325,2 | 47 410,4 | 45 410,4 | 49 049,6 | 50 848,5 | 46 491,4 |

**Figure X-8 : Analyse des produits d'épissage du pré-ARNm C2 et des différents pré-ARNm C2-NRS par RT-PCR.**

**A :** Analyse par électrophorèse en gel d'agarose des produits d'épissage amplifiés par RT-PCR. Les noms des pré-ARNm utilisés lors des réactions d'épissage sont mentionnés au-dessus des pistes. Les produits PCR correspondant aux ARNm (Sp) et aux pré-ARNm plus les ARNm (Sp+UnSp) sont indiqués à gauche des gels.

**B :** Quantification de l'intensité des bandes correspondant aux produits épissés (Sp) du gel d'agarose présenté en A). Les intensités des bandes sont mesurées et reportées sous forme de pic délimité par des pointillés sur le chromatogramme. Le tableau comprend pour chaque pic la valeur de l'intensité maximale et le volume correspondant à l'intégration de l'aire du pic. Les produits de RT-PCR correspondant aux produits épissés (Sp) sont reportés en dessous des pics correspondants.

**C :** Quantification des bandes correspondant aux produits épissés et non épissés (Sp+UnSp) du gel d'agarose présenté en A). Les caractéristiques du chromatogramme sont identiques à celui décrit en B). Les produits de RT-PCR correspondant aux produits épissés et non épissés (Sp+UnSp) et les oligonucléotides libres (oligo) sont reportés en dessous des pics correspondants.

**D :** Représentation graphique des ratios Sp/UnSp qui ont été calculés pour chacun des ARN à partir des valeurs des volumes des pics présentés en B) et C).

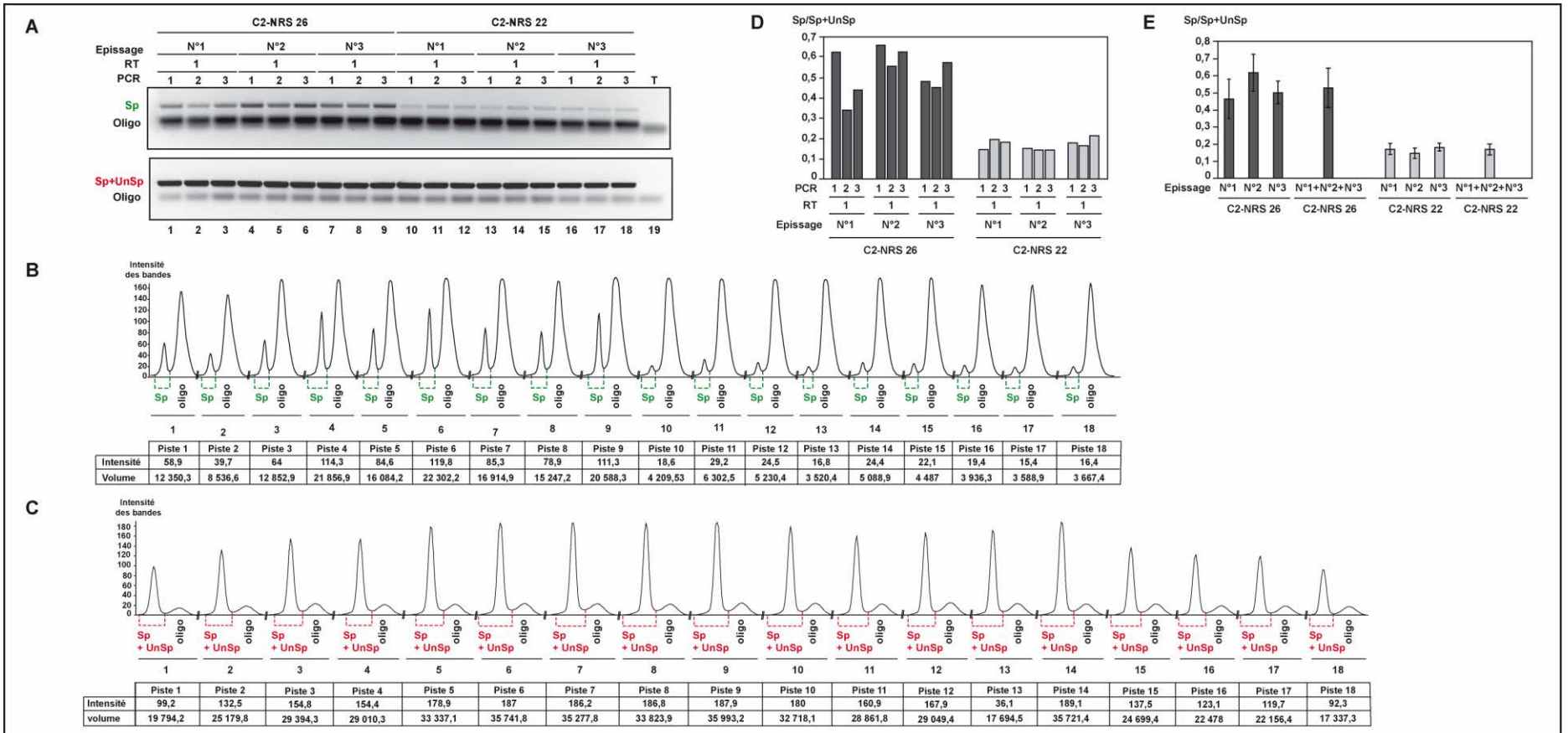
**E :** Représentation graphique des pourcentages d'épissage qui ont été calculés à partir du ratio Sp/UnSp obtenu pour l'ARN C2-NRS Inv qui a été pris comme référence (100 %.) Les valeurs des pourcentages sont indiquées au-dessus de chaque histogramme.

### II.3.3. Reproductibilité des expériences de RT-PCR

Pour vérifier la reproductibilité des expériences de RT-PCR, nous avons réalisé des expériences d'épissage *in vitro* en triplicat. Sur la base des résultats obtenus précédemment, nous avons choisi d'étudier l'épissage des pré-ARNm C2-NRS 22 et C2-NRS 26 dont les amplifications des produits d'épissage par RT-PCR donnent des bandes correspondant aux produits épissés (Sp) de faible et forte intensités, respectivement (**Figure X-8, A pistes 5 et 7**).

Pour chaque pré-ARNm, 3 réactions d'épissage suivies d'une réaction de transcription inverse ont été réalisées. Les ADNc obtenus ont ensuite été amplifiés lors de 3 réactions de PCR dont les produits ont été fractionnés par une électrophorèse en gel d'agarose puis quantifiés (**Figure X-9, A, B, C**). Pour chaque échantillon, les ratios Sp/Sp+UnSp ont été calculés (**Figure X-9, D**) ainsi que les écarts type et les moyennes correspondants aux triplicata (**Figure X-9, E**).

Pour chacune des 3 expériences d'épissage réalisées avec le pré-ARNm C2-NRS 22, les valeurs moyennes des 3 analyses PCR des ratios Sp/Sp+UnSp étaient de : 0,183±0,025, 0,148±0,033 et 0,172±0,034 (**Figure X-9, E**). Pour chacune des 3 expériences d'épissage réalisées avec le pré-ARNm C2-NRS 26, les valeurs moyennes des 3 analyses PCR des ratios Sp/Sp+UnSp étaient de : 0,466±0,118, 0,619±0,109 et 0,502±0,069 (**Figure X-9, E**). Pour un pré-ARNm donné, les moyennes des valeurs Sp/Sp+UnSp calculées pour les chacune des 3 expériences d'épissage étaient proches et se recoupaient avec les valeurs des écarts types (**Figure X-9, E**). Cela indiquait que les expériences de RT-PCR étaient réalisées dans des conditions semi-quantitatives. Quelque soit la valeur de la moyenne des ratios Sp/Sp+UnSp des expériences de PCR réalisées en triplicat, les valeurs obtenues pour l'épissage du pré-ARNm NRS-C2 22 étaient significativement différentes de celles obtenues pour l'épissage du pré-ARNm C2-NRS 26. Les valeurs des ratios Sp/Sp+UnSp de chacune des 3 expériences de PCR réalisées en triplicat étant proches, il était possible de calculer leur moyenne. Pour le pré-ARNm C2-NRS 22 la moyenne des 3 expériences d'épissage est de 0,17±0,03 et celle du pré-ARNm C2-NRS 26 est de 0,53±0,12 (N°1+N°2+N°3, **Figure X-9, E**).




| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 | Piste 7 | Piste 8 | Piste 9 | Piste 10 | Piste 11 | Piste 12 | Piste 13 | Piste 14 | Piste 15 | Piste 16 | Piste 17 | Piste 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intensité | 58,9 | 39,7 | 64 | 114,3 | 84,6 | 119,8 | 85,3 | 78,9 | 111,3 | 18,6 | 29,2 | 24,5 | 16,8 | 24,4 | 22,1 | 19,4 | 15,4 | 16,4 |
| Volume | 12 350,3 | 8 536,6 | 12 852,9 | 21 856,9 | 16 084,2 | 22 302,2 | 16 914,9 | 15 247,2 | 20 588,3 | 4 209,53 | 6 302,5 | 5 230,4 | 3 520,4 | 5 088,9 | 4 487 | 3 936,3 | 3 588,9 | 3 667,4 |

| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 | Piste 7 | Piste 8 | Piste 9 | Piste 10 | Piste 11 | Piste 12 | Piste 13 | Piste 14 | Piste 15 | Piste 16 | Piste 17 | Piste 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intensité | 99,2 | 132,5 | 154,8 | 154,4 | 178,9 | 187 | 186,2 | 186,8 | 187,9 | 180 | 160,9 | 167,9 | 36,1 | 189,1 | 137,5 | 123,1 | 119,7 | 92,3 |
| volume | 19 794,2 | 25 179,8 | 29 394,3 | 29 010,3 | 33 337,1 | 35 741,8 | 35 277,8 | 33 823,9 | 35 993,2 | 32 718,1 | 28 861,8 | 29 049,4 | 17 694,5 | 35 721,4 | 24 699,4 | 22 478 | 22 156,4 | 17 337,3 |

**Figure X-9 : Reproductibilité de l'analyse des produits d'épissage des pré-ARNm C2-NRS 26 et C2-NRS 22 par RT-PCR.**

Trois réactions d'épissage ont été réalisées avec les pré-ARNm C2-NRS 26 et C2-NRS 22 (Epissage N°1, 2 et 3). Les produits d'épissage obtenus ont été rétrotranscrits lors d'une réaction de transcription inverse (RT) et amplifiés par 3 réactions de PCR indépendantes (PCR 1, 2 et 3).

**A :** Analyse par électrophorèse en gel d'agarose des produits d'épissage amplifiés par RT-PCR. Les noms des pré-ARNm utilisés lors des réactions d'épissage sont mentionnés au-dessus des pistes. Les produits PCR correspondant aux ARNm (Sp) et aux pré-ARNm plus les ARNm (Sp+UnSp) et les oligonucléotides libres sont indiqués à gauche des gels. Une réaction de PCR réalisée sans matrice ADNc issue de la réaction de transcription inverse sert de témoin (T).

**B et C :** Quantification de l'intensité des bandes correspondant aux produits épissés (Sp) et aux produits épissés plus non épissés (Sp+UnSp) des gels d'agarose présentés en A). Les intensités des bandes sont mesurées et reportées sous forme de pic délimité par des pointillés sur le chromatogramme. Le tableau comprend pour chaque pic la valeur de l'intensité maximale et le volume correspondant à l'intégration de l'aire du pic. Les produits de RT-PCR correspondant aux produits épissés (Sp), aux produits épissés plus non épissés (Sp+UnSp) et les oligonucléotides libres (oligo) sont reportés en dessous des pics correspondants.

**D :** Représentation graphique des ratios Sp/Sp+UnSp qui ont été calculés pour chacun des ARN et pour chaque expérience à partir des valeurs des volumes des pics présentés en B) et C).

**E :** Représentation graphique des valeurs moyennes et des écrats types de chacune et du total des 3 réactions de PCR présentées en D).

## III. Etude fonctionnelle de l'élément NRS par des expériences d'épissage *in cellulo*

Nous avons également étudié l'inhibition de l'épissage par les différents éléments NRS *in cellulo*, ceci afin de confirmer les résultats obtenus lors des expériences d'épissage *in vitro*.

### III.1. Choix de la lignée cellulaire, des constructions d'épissage et des plasmides utilisés pour l'étude *in cellulo*

#### III.1.1. La lignée cellulaire

La lignée cellulaire HeLa est employée couramment pour l'étude de l'épissage *in cellulo* et a été utilisée pour purifier les extraits nucléaires que nous avons utilisés lors de l'étude de l'épissage *in vitro*. Bien que l'élément NRS provienne d'un virus infectant exclusivement les cellules aviaires, sa fonction en tant qu'élément inhibiteur d'épissage est identique dans les cellules aviaires et les cellules humaines (Hibbert *et al.*, 1999).

#### III.1.2. Les constructions d'épissage et les plasmides

Les constructions utilisées étaient destinées à produire *in cellulo* des ARN contenant les séquences des constructions C2, C2-NRS WT et C2-NRS Inv que nous avions précédemment utilisées lors des expériences d'épissage *in vitro*. Les séquences ADN correspondantes ont été insérées dans le vecteur d'expression pGL3-contrôle à la place du gène de la luciférase entre les sites de restriction *Nco*I et *Xba*I, selon le protocole décrit dans la partie Matériel et Méthodes (**Chapitre V**). Nous appelons ces constructions pGL-C2, pGLC2-NRS WT et pGLC2-NRS Inv. Leur transcription est placée sous le contrôle du promoteur fort du SV40 et leur signal de polyadénylation est également celui du SV40 (**Figure X-10, A et B**).

Nous avons également utilisé le plasmide phRGB qui est un vecteur d'expression eucaryote codant la luciférase Renilla. Lors de chaque expérience, ce plasmide a été co-transfecté avec un des plasmides pGL- afin de normaliser les efficacités de transfection.

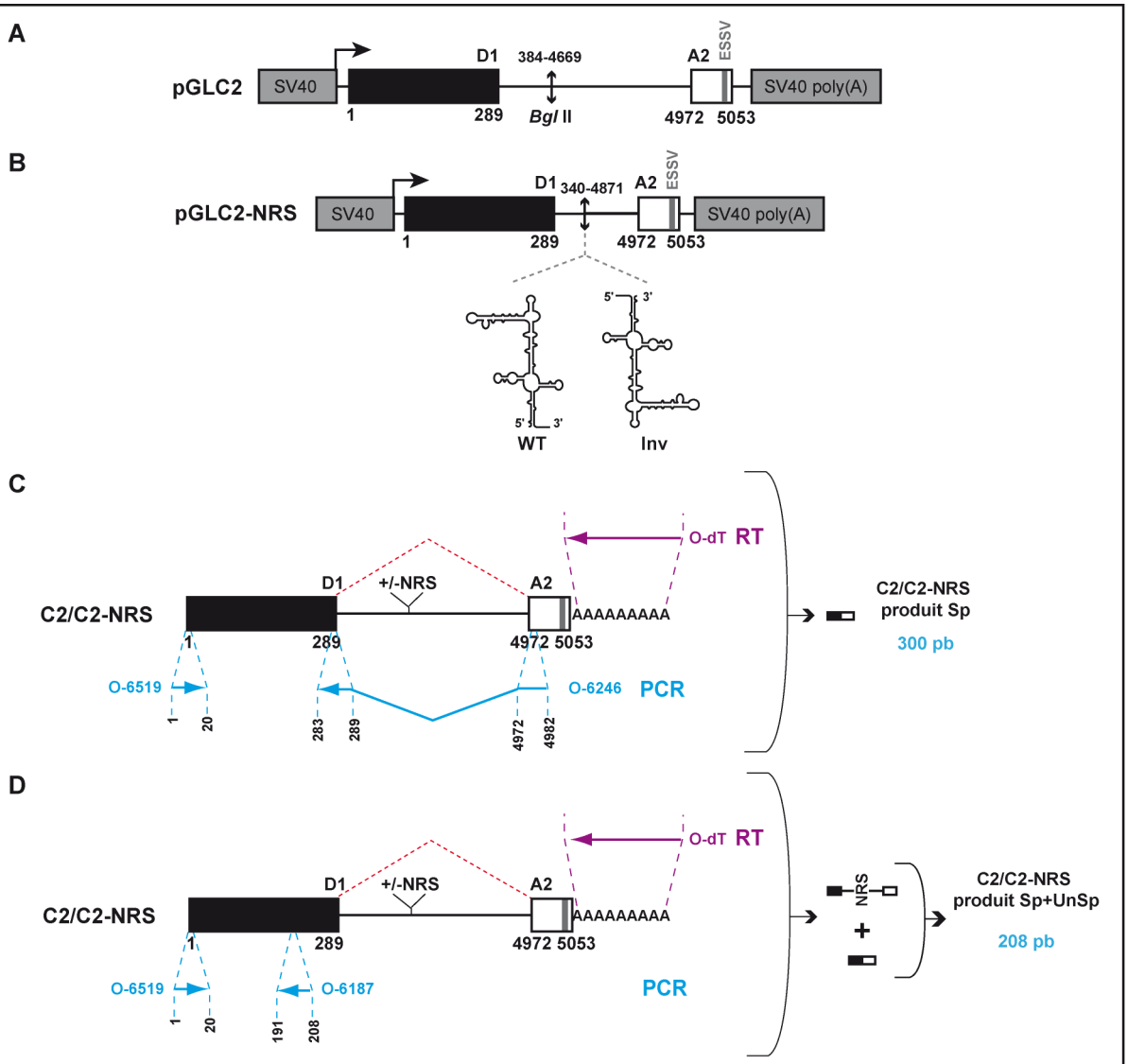



**Figure X-10 : Amplification par RT-PCR des espèces épissées (Sp) et des espèces épissées plus non épissées (Sp+UnSp) lors des expériences d'épissage *in cellulo*.**

**A et B :** Représentation des constructions pGLC2 (A) et pGLC2-NRS (B). Les exons sont représentés par des rectangles, l'intron par une ligne noire et les sites donneur D1 et accepteur A2 sont mentionnés. Le promoteur et le signal de polyadénylation du SV40 sont représentés par des rectangles gris. L'initiation de la transcription est symbolisée par une flèche noire. L'insertion de l'élément NRS dans la construction pGLC2 est représentée par des traits en pointillé gris.

**C :** Représentation des sites d'hybridation des oligonucléotides utilisés lors de la réaction de RT-PCR pour amplifier les produits épissés (Sp). L'oligonucléotide utilisé pour la réaction de transcription inverse (RT) est illustré par une flèche violette, le couple d'oligonucléotides utilisé pour la réaction de PCR, par des flèches bleues. Les sites d'hybridation de chaque oligonucléotide sont précisés. Les exons sont représentés par des rectangles, l'intron par une ligne noire. Les sites donneur D1 et accepteur A2 sont mentionnés, la réaction d'épissage est symbolisée par le trait rouge en pointillé. La taille des produits obtenus par RT-PCR (Sp) est mentionnée.

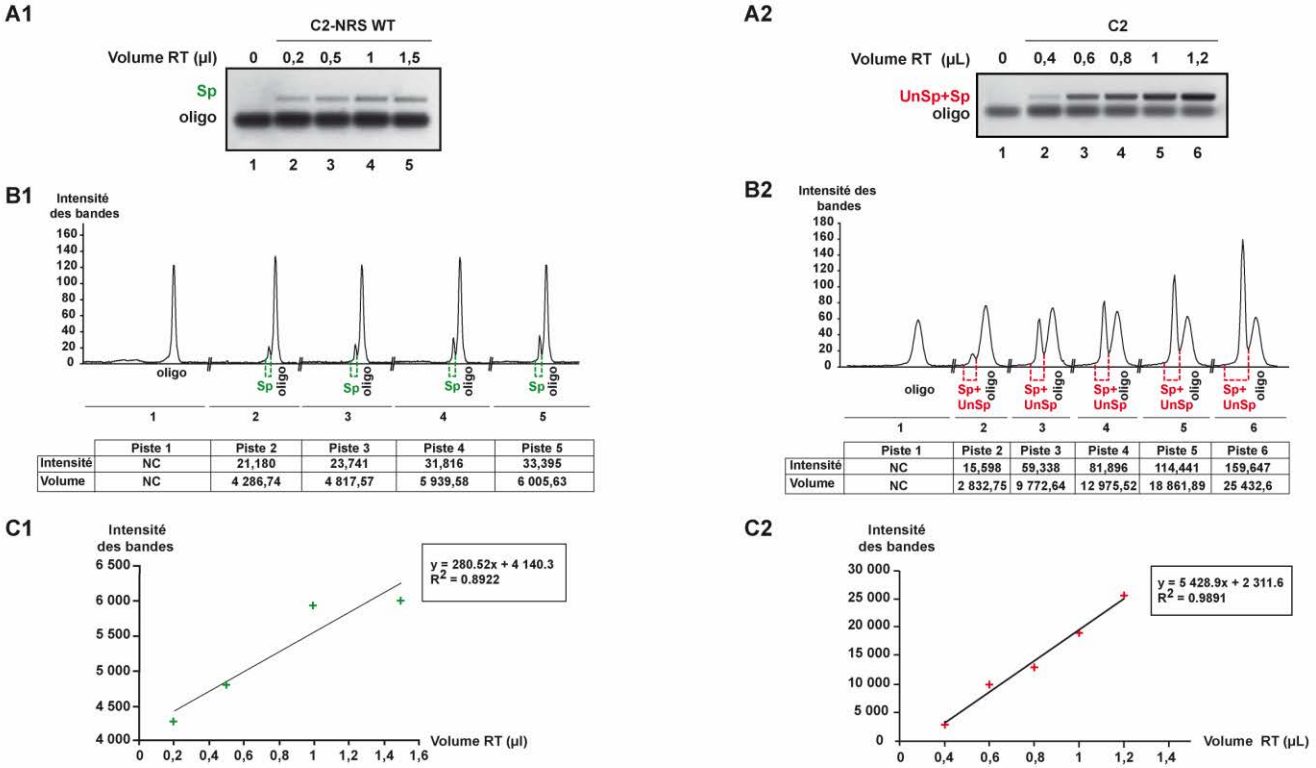
**D :** Représentation des sites d'hybridation des oligonucléotides utilisés lors de la réaction de RT-PCR pour amplifier les produits épissés et non épissés (Sp+UnSp). L'oligonucléotide utilisé pour la réaction de transcription inverse (RT) est illustré par une flèche violette, le couple d'oligonucléotides utilisé pour la réaction de PCR, par des flèches bleues. Les sites d'hybridation de chaque oligonucléotide sont précisés. Les exons sont représentés par des rectangles, l'intron par une ligne noire. Les sites donneur D1 et accepteur A2 sont mentionnés et la réaction d'épissage est symbolisée par le trait rouge en pointillé. La taille des produits obtenus par RT-PCR (Sp+UnSp) est mentionnée.

Les résultats d'épissage obtenus *in cellulo* sont en accord avec ceux obtenus *in vitro*

Les données ont été obtenues expérimentalement par A. Visvikis, j'ai participé à l'exploitation des résultats. Pour ces expériences d'épissage *in cellulo*, les cellules HeLa ont été transfectées avec les différentes constructions d'épissage d'intérêt. Après 48 h de culture, les ARN totaux ont été extraits et l'ADN co-purifié avec eux dégradé par une l'action d'une DNase. Les produits d'épissage ont ensuite été analysés par RT-PCR selon le protocole que nous avions mis au point pour les expériences d'épissage *in vitro*. Une quantité d'1 µg d'ARN total a été utilisée pour la réaction de transcription inverse avec un oligonucléotide dT (O-dT) qui s'hybridait sur la queue polyA des ARN (**Figure X-10**). Le même protocole de PCR que celui mis au point pour les épissages *in vitro*, ainsi que les mêmes oligonucléotides, ont été utilisés pour amplifier les produits épissés (Sp) et les produits épissés plus non épissés (Sp+UnSp) (**Figure X-10**).

### III.1.3. Linéarité des amplifications PCR

Nous nous sommes tout d'abord assurés que l'amplification par PCR des produits épissés (Sp) d'une part et épissés et non épissés (Sp+UnSp) d'autre part a lieu dans une zone de linéarité. Pour cela, nous avons transfecté les cellules avec les constructions pGLC2 et pGLC2-NRS WT puis nous avons extraits les ARN totaux. Nous avons ensuite réalisé une réaction de transcription inverse et utilisé une gamme de différents volumes de cette réaction pour amplifier par PCR les produits épissés plus non épissés (Sp+UnSp) d'une part et les produits épissés (Sp) d'autre part. Les valeurs correspondant aux intensités des produits Sp et Sp+UnSp ont été reportées sur un graphique en fonction du volume de RT et s'alignent sur une droite dont le coefficient de corrélation est proche de 1 (**Figure X-11**). Nous pouvons en conclure que nos conditions de PCR se situent dans la zone de linéarité d'amplification du signal et sont proportionnelles aux quantités d'ARNm et de pré-ARNm extraites des cellules.




| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 |
|---|---|---|---|---|---|
| Intensité | NC | 21,180 | 23,741 | 31,816 | 33,395 |
| Volume | NC | 4 286,74 | 4 817,57 | 5 939,58 | 6 005,63 |

| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 |
|---|---|---|---|---|---|---|
| Intensité | NC | 15,598 | 59,338 | 81,896 | 114,441 | 159,647 |
| Volume | NC | 2 832,75 | 9 772,64 | 12 975,52 | 18 861,89 | 25 432,6 |

**Figure X-11 : L'amplification des produits d'épissage par RT-PCR est linéaire.**

Après la transfection des cellules avec les constructions pGL C2-NRS WT et pGLC2 (codant respectivement pour les pré-ARNm C2 NRS-WT et C2), les ARN totaux ont été extraits et rétrotranscrits lors d'une expérience de transcription inverse avec un oligonucléotide dT. Une gamme de volumes allant de 0 à 1,5 µl du mélange réactionnel de la transcription inverse réalisée sur les ARN totaux des cellules transfectées avec la construction pGLC2-NRS WT a ensuite été utilisée pour des amplifications PCR avec le couple d'oligonucléotides amplifiant spécifiquement les produits épissés (Sp) (A1, B1, C1). Une gamme de volumes allant de 0 à 1,2 µl du mélange réactionnel de la transcription inverse réalisée sur les ARN totaux des cellules transfectées avec la construction pGLC2 a été utilisée pour les réactions de PCR afin d'amplifier les produits correspondant aux produits épissés plus non épissés (Sp+UnSp) (A2, B2, C2).

**A :** Electrophorèses en gel d'agarose des produits de RT-PCR. Les fragments correspondent aux produits épissés (Sp) du pré-ARNm C2-NRS WT (A1) ou aux produits épissés plus non épissés (Sp+UnSp) du pré-ARNm C2 (A2). Le volume de la transcription inverse et les noms des constructions utilisées lors des transfections sont indiqués au-dessus des pistes. Les positions des produits épissés (Sp) et épissés plus non épissés (Sp+UnSp) et des oligonucléotides libres sont mentionnés à gauche des gels.

**B :** Quantification de l'intensité des bandes correspondant aux produits épissés (Sp) et aux produits épissés plus non épissés (Sp+UnSp) des gels d'agarose présentés en A). Les chromatogrammes B1 et B2 correspondent aux quantifications des bandes des gels d'agarose A1 et A2, respectivement. Les intensités des bandes sont mesurées et reportées sous forme de pic délimité par des pointillés sur le chromatogramme. Le tableau comprend pour chaque pic, la valeur de l'intensité maximale et le volume correspondant à l'intégration de l'aire du pic. Les produits de RT-PCR correspondant aux produits épissés (Sp), aux produits épissés plus non épissés (Sp+UnSp) et les oligonucléotides libres (oligo) sont reportés en dessous des pics correspondants.

**C :** Les intensités des pic des chromatogrammes présentés en B) ont été reportées sur un graphique en fonction du volume de RT utilisé. Les graphiques C1 et C2 correspondent aux reports des valeurs des chrommatogrammes B1 et B2, respectivement. L'équation de la droite ainsi que le coefficient de régression linéaire sont encadrés.

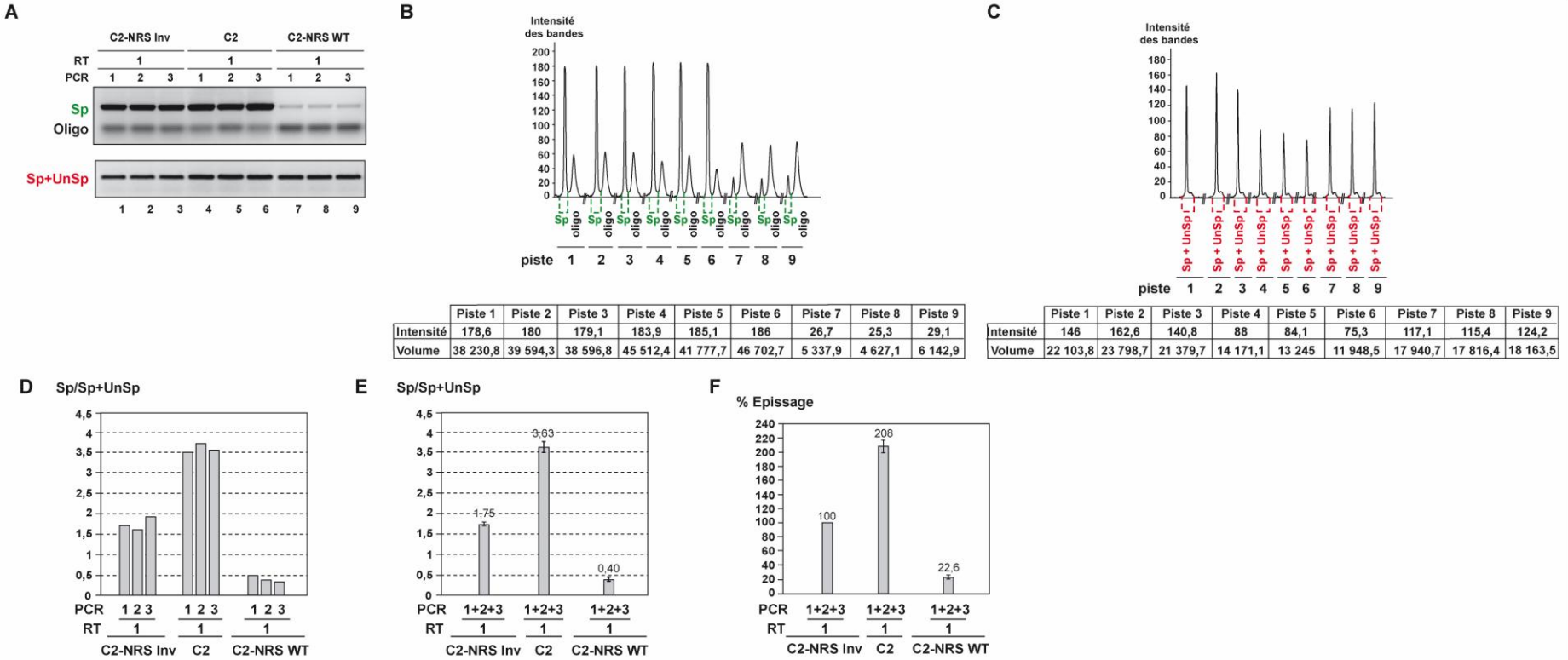
### III.1.4. Résultats et reproductibilité des amplifications par RT-PCR

Nous avons réalisé une première transfection des cellules HeLa avec les constructions pGLC2, pGLC2-NRS WT et pGLC2-NRS Inv. L'extraction des ARN totaux et l'amplification par RT-PCR des espèces épissées (Sp) et épissées plus non épissées (Sp+UnSp) ont été réalisées selon les conditions que nous avons décrites auparavant. Afin de vérifier la reproductibilité des résultats de RT-PCR, la réaction de transcription inverse a été réalisée une fois et les amplifications par PCR ont été réalisées en triplicat (PCR 1, 2, 3, **Figure X-12**). Les produits de RT-PCR obtenus ont été fractionnés par une électrophorèse en gel d'agarose puis quantifiés selon le même protocole que celui précédemment utilisé pour les épissages *in vitro*.

La comparaison des ratios Sp/Sp+UnSp des pré-ARNm C2 et C2-NRS Inv indiquait que l'insertion dans le pré-ARNm C2 de l'élément NRS dans son orientation antisens (C2-NRS Inv) provoquait une diminution de l'épissage (1, 75 vs 3, 63, respectivement **Figure X-12, E**). Cette diminution était encore plus marquée lorsque l'élément NRS était inséré dans son orientation sens (C2-NRS WT), ce qui indiquait que l'épissage était inhibé *in cellulo* par l'élément NRS fonctionnel.

Les ratios Sp/Sp+UnSp obtenus pour les 3 amplifications PCR d'un même échantillon présentaient peu de variations, ce qui indiquait que nos conditions de RT-PCR étaient reproductibles (**Figure X-12, D**).

Pour mieux comparer les données, le ratio Sp/Sp+UnSp obtenu pour le pré-ARNm C2-NRS Inv avait été pris pour le 100% d'épissage et les ratios obtenus pour les autres pré-ARNm avait été calculés en % par rapport à cette référence (**Figure X-12, F**). Le pourcentage d'épissage *in cellulo* du pré-ARNm C2-NRS WT était de 22,6 %, ce qui reflétait une inhibition de l'épissage plus importante que celle que nous avions observée *in vitro* où cette valeur était d'environ 40 % (**Figure X-7**).




| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 | Piste 7 | Piste 8 | Piste 9 |
|---|---|---|---|---|---|---|---|---|---|
| Intensité | 178,6 | 180 | 179,1 | 183,9 | 185,1 | 186 | 26,7 | 25,3 | 29,1 |
| Volume | 38 230,8 | 39 594,3 | 38 596,8 | 45 512,4 | 41 777,7 | 46 702,7 | 5 337,9 | 4 627,1 | 6 142,9 |

| | Piste 1 | Piste 2 | Piste 3 | Piste 4 | Piste 5 | Piste 6 | Piste 7 | Piste 8 | Piste 9 |
|---|---|---|---|---|---|---|---|---|---|
| Intensité | 146 | 162,6 | 140,8 | 88 | 84,1 | 75,3 | 117,1 | 115,4 | 124,2 |
| Volume | 22 103,8 | 23 798,7 | 21 379,7 | 14 171,1 | 13 245 | 11 948,5 | 17 940,7 | 17 816,4 | 18 163,5 |

**Figure X-12 : Analyse par RT-PCR des produits d'épissage *in cellulo* des pré-ARNm C2 et C2-NRS Inv et C2-NRS WT et reproductibilité de l'amplification par RT-PCR.**

Les cellules ont été transfectées avec les plasmides codant les pré-ARNm C2, C2-NRS Inv et C2-NRS WT. Une réaction de transcriptase inverse (RT) et 3 réactions de PCR (1, 2, 3) ont été réalisées sur chaque échantillon.

**A :** Analyse par électrophorèse en gel d'agarose des produits d'épissage amplifiés par RT-PCR. Les cellules ont été transfectées avec les plasmides codant les pré-ARNm qui sont mentionnés au-dessus des pistes. Les produits de RT-PCR correspondant aux ARNm (Sp), aux pré-ARNm plus les ARNm (Sp+UnSp) ainsi que les oligonucléotides libres (Oligo) sont indiqués à gauche des gels.

**B :** Quantification de l'intensité des bandes correspondant aux produits épissés (Sp) du gel d'agarose présenté en A). Les intensités des bandes sont mesurées et reportées sous forme de pic délimité par des pointillés sur le chromatogramme. Le tableau comprend pour chaque pic la valeur de l'intensité maximale et le volume correspondant à l'intégration de l'aire du pic. Les produits de RT-PCR correspondant aux produits épissés (Sp) ainsi que les oligonucléotides libres (Oligo) sont reportés en dessous des pics correspondants.

**C :** Quantification des bandes correspondant aux produits épissés plus non épissés (Sp+UnSp) du gel d'agarose présenté en A). Les caractéristiques du chromatogramme sont identiques à celui décrit en B). Les produits de RT-PCR correspondant aux produits épissés plus non épissés (Sp+UnSp) sont reportés en dessous des pics correspondants.

**D :** Représentation graphique des ratios Sp/Sp+UnSp qui ont été calculés pour chacun des ARN à partir des valeurs des volumes des pics présentés en B) et C).

**E :** Représentation graphique des valeurs moyennes et des écrats types des 3 réactions de PCR présentées en D).

**F :** Représentation graphique des pourcentages d'épissage qui ont été calculés à partir du ratio Sp/Sp+UnSp obtenu pour l'ARN C2-NRS Inv qui a été pris comme la référence de 100 %. Les valeurs des pourcentages sont indiquées au-dessus de chaque histogramme.

### III.2. Reproductibilité des résultats d'épissage *in cellulo*

Pour vérifier la reproductibilité de l'expérience, une seconde transfection a été réalisée. Les conditions de transfection et d'analyses des produits d'épissage sont identiques à celles utilisées pour la première transfection. Les pourcentages d'épissage des pré-ARNm de la seconde transfection sont présentés et comparés à ceux de la première transfection dans la **Figure X-13**. Les pourcentages d'épissage des pré-ARNm C2 et C2-NRS WT sont similaires à ceux obtenus pour la première transfection (**Figure X-13, B**). Cela indiquait que nos conditions de transfection et nos conditions de RT-PCR étaient reproductibles et nous permettaient d'étudier l'inhibition de l'épissage par l'élément NRS dans le contexte du pré-ARNm C2.

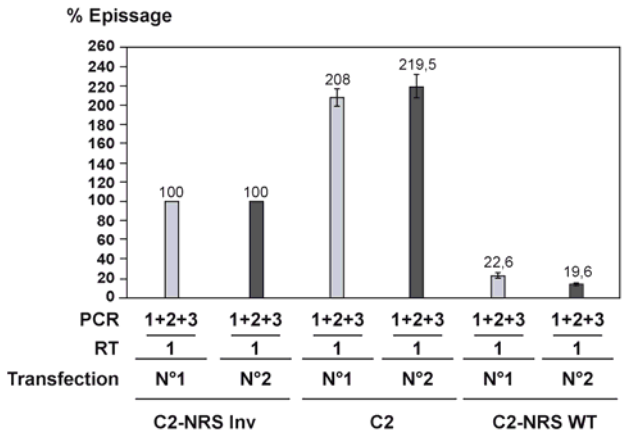



**Figure X-13 : Reproductibilité de l'analyse par RT-PCR des produits d'épissage obtenus pour les pré-ARNm C2, C2-NRS WT et C2-NRS Inv, lors de deux transfections.**
Représentation graphique des pourcentages d'épissage de chaque pré-ARNm obtenus lors de 2 transfections indépendantes (N°1, N°2). La valeur correspond à la moyenne des valeurs obtenues lors de 3 expériences de PCR et l'écart type à la variation des valeurs de ces 3 expériences. L'épissage du pré-ARNm C2-NRS Inv a été choisi comme la référence de 100 % d'épissage et les pourcentages des épissages des autres pré-ARNm ont été calculés à partir de cette référence. Le nom des pré-ARNm est indiqué en dessous des histogrammes.

# Article scientifique

# Structural and functional analysis of the Rous Sarcoma Virus Negative Regulator of Splicing and demonstration of its activation by the 9G8 SR protein

Aileen Bar[1], Virginie Marchand[1,+], Georges Khoury[1], Natacha Dreumont[2], Annie Mougin[1,•], Nathalie Robas[1], James Stévenin[2], Athanase Visvikis[1,*], Christiane Branlant[1,*]

[1]« ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale », Nancy Université-UMR 7214 CNRS-UHP, Faculté des Sciences et Techniques, BP 70239, 54506, Vandoeuvre-lès-Nancy cedex, France

[2]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Strasbourg Université-CNRS-INSERM, 1, rue Laurent Fries, BP 163, 67404, Illkirch, France

+ Present address :
Developmental Biology Unit, European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany

● Present address :
Laboratoire de Biologie Moléculaire Eucaryote; Université de Toulouse III; UMR 5099 CNRS-UPS; 118 route de Narbonne, 31062; Toulouse; France

Corresponding authors :
*Christiane BRANLANT
ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale (AREMS) UMR 7214 CNRS-UHP Nancy 1 Faculté des Sciences et Techniques, BP 70239, 54506 Vandoeuvre-lès-Nancy cedex, France
Tel : 33 3 83 68 43 03
Fax : 33 3 83 68 43 07
E-mail: christiane.branlant@maem.uhp-nancy.fr

*Athanase VISVIKIS
ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale (AREMS) UMR 7214 CNRS-UHP Nancy 1 Faculté des Sciences et Techniques, BP 70239, 54506 Vandoeuvre-lès-Nancy cedex, France
Tel : 33 3 83 68 42 80
Fax : 33 3 83 68 43 07
E-mail : athanase.visvikis@maem.uhp-nancy.fr

Key words: ASLV, NRS, RNA secondary structure, 9G8 SR protein, splicing inhibition

**ABSTRACT**

Retroviruses require both spliced and unspliced RNAs for replication. Accumulation of Rous Sarcoma Virus (RSV) unspliced RNA depends upon the Negative Regulator of Splicing (NRS). Its 5' part is considered as an ESE binding SR proteins. Its 3' part contains a decoy 5' splice site (ss), which inhibits splicing at the *bona fide* 5' ss. Only the 3D structure of a small NRS fragment had been experimentally studied. Here, by chemical and enzymatic probing, we determine the 2D structure of the entire RSV NRS. Structural analysis of other avian NRSs and comparison with all sequenced avian NRSs is in favour of a phylogenetic conservation of the NRS 2D structure. By combination of approaches : i) *in vitro* and *in cellulo* splicing assays, ii) footprinting assays, and iii) purification and analysis of reconstituted RNP complexe, we define a small NRS element retaining splicing inhibitory property. We also demonstrate the capability of the SR protein 9G8 to increase NRS activity *in vitro* and *in cellulo*. Altogether these data bring new insights on how NRS fine tune splicing activity.

Keywords : ASLV, NRS, RNA secondary structure, 9G8 SR protein, splicing inhibition

## INTRODUCTION

For all retroviruses, alternative splicing plays a key role in the modulation of gene expression. After integration within the host genome, the proviral genome is transcribed as a primary transcript by RNA polymerase II. A large number of the transcribed molecules remain unspliced. They are needed to serve as genomes for new virions and as messenger RNAs for translation of the Gag-Pol protein precursor. The remaining part of the primary transcript undergoes alternative splicing in the nucleus to generate mRNAs for production of other retroviral proteins. Rous Sarcoma Virus (RSV), which was the first discovered replication competent oncogenic retrovirus, exemplified this situation (1). This simple avian retrovirus belongs to the avian sarcoma/leucosis virus (ASLV) family. It contains the *gag*, *pol* and *env* genes common to all retroviruses and codes the oncogenic protein v-Src. ASLV viruses deserve interest because of the damage they generate in poultry breeding. In addition, RSV is an interesting, already well studied model, to investigate mechanisms involved in the regulation of retroviral alternative splicing (2). Its RNA contains a single 5' splice site (ss) (position 398 as referred to the transcription start site) and two 3' ss in competition one with the other, the *env* and *src* 3' ss (positions 5078 and 7054, respectively). Coupling of the *env* 3' or *src* 3' ss with the 5' ss leads to the production of the *env* and *src* mRNAs, respectively.

The low level of RSV RNA splicing (20 to 30 %) is due to the presence of suboptimal splicing sites and to the presence of strong splicing inhibitory elements. The branch point sequence of the *env* 3' ss is not canonical and the polypyrimidine tract of *src* 3' ss is suboptimal (3-6). Splicing at the *src* 3' ss is specifically down regulated by an upstream sequence named SSS for Suppressor of Src Splicing (7). SSS seems to bind an unknown factor specific to chicken cells and its mechanism of action is not elucidated. A dr1 direct repeat flanks the *src* gene. It was first identified as a constitutive transport element (CTE) promoting the accumulation of unspliced viral RNA into the cytoplasm (8,9). This element may in addition contain a second silencing element acting specifically on the *src* 3' ss (10,11). A third negative *cis* splicing regulatory sequence, called the Negative Regulator of Splicing (NRS), prevents splicing at both 3' ss (12-15). NRS (positions 703 to 930) is located in the *gag* gene approximatively 300 nt downstream from the 5' ss.

NRS was found to inhibit splicing of a heterologous *myc* intron as well *in vitro* as *in cellulo*. Deletion of an internal 76-nt long segment (nts 799 to 874) (NRS Δ76) had a limited effect on NRS activity (13). Thus, NRS Δ76 was defined as the minimal sequence required for splicing inhibition activity. Two segments of the NRS were described as having an important role for

splicing inhibition : the NRS 5' part (positions 700 to 800, NRS-5') was found to have splicing enhancer activity in an heterologous context, the NRS 3' terminal part (positions 801 to 930, NRS-3') contains two overlapping 5' ss sequences. One is capable of binding U1 snRNA (major site), the other one binds the U11 snRNP (minor site) (16-18). U1 snRNP binding does not induce splicing and is required for the inhibition activity, suggesting that the degenerated major 5' ss (5 out of 8 matches fit to the 5' ss consensus) is a decoy 5' ss (17). Unlike U1 snRNP, U11 snRNP binding to the minor 5' ss is dispensable for splicing inhibition and is just expected to have a moderate negative effect on U1 snRNP binding to the major site (17,18). NRS-5' had been shown to bind SR proteins and hnRNP H (19-21). Binding of SR proteins to the NRS-5' segment was proposed to enhance U1, but not U11 snRNP recruitment (18). The NRS appears to function as a non productive decoy 5' ss, which competes with the authentic viral 5' ss, and forms abortive spliceosomal complexes when coupled to one of the two 3' ss. *In vitro* analysis demonstrated that human U5-220kD protein (hPrp8) does not cross link to the NRS pseudo 5' ss, highlighting the fact that the abortive complex cannot switch into a catalytic complex (22).

A computer assisted 2D structure model of the NRS element was proposed long time ago (13). In this model, the NRS-5' and NRS-3' regions are paired together. Our previous chemical and enzymatic probing of part of the NRS structure in collaboration with the Beemon group showed that the NRS-3' region, encompassing nucleotides from positions G910 to U926, forms an individual stem-loop structure and this was the basis for an NMR-3D structure analysis of this NRS fragment (23). Formation of the identified stem-loop structure is incompatible with the previously proposed 2D structure of entire NRS. An experimental analysis of the 2D structure of the entire NRS element was therefore needed to get a clear answer on how NRS folds. As determination of RNA 3D structure by NMR can only be performed on short RNAs, only the RNA probing method based on utilization of chemical and enzymatic probes was adapted to this purpose. It should be pointed out that RNA 2D structure was already shown to play an important role in pre-mRNA splicing regulation (24-26). Therefore, determination of the NRS structure was expected to bring some insight on how NRS inhibits splicing.

In this work, we performed an experimental analysis of the entire RSV and ALV LR9 NRSs (27,28) 2D structures. The data provide evidence for a cruciform structure different from the previously proposed computer-based 2D structure. In the new cruciform structure, NRS-3' forms an individual stem-loop structure, which is in accordance with previous NMR data (23). The NRS-5' segment that binds ASF/SF2 also forms a stem-loop structure. Based on

these data and in order to test for phylogenetic conservation of the NRS 2D structure, we undertook a comparative analysis of the possible 2D structures of avian retrovirus NRSs using available avian retrovirus nucleotide sequences. In parallel, we tested the NRS activity of truncated forms of the RSV NRS by *in vitro* splicing assays using a hybrid construct derived from HIV-1. By this approach, we identified a small discontinuous 114-nt element as the minimum element required for splicing inhibition. Then, we used a new affinity chromatography procedure (29) to define which of the produced truncated RSV NRSs can recruit the SR proteins ASF/SF2, SC35 and 9G8. Our finding of a strong binding of the SR protein 9G8 on WT and some of the truncated NRSs stimulated us to test the activation role of this SR protein on RSV NRS activity. By *in vitro* and *in cellulo* assays, we demonstrated a strong stimulatory effect of protein 9G8 on RSV NRS activity and we identified its binding site on RSV NRS by footprinting assays.

## MATERIALS AND METHODS

### DNA constructs

The 2D structure of NRS WT, NRS Δ76, LR9 and ΔLR9 RNAs was analysed on transcripts produced from PCR products. PCR products were amplified from the previously described pRSVNeo-Int, pΔ76Neo-Int, pLR9 and pΔLR9 plasmids (13,28). Plasmids pRSVNeo-Int, pΔ76Neo-Int contain DNA sequences of the WT and Δ76 RSV NRSs from strain Prague C (PrC) (13), respectively. The sense primer (O-2954, Additional Table 1) contained a T7 RNA polymerase promoter and the antisense primer (O-2955, Additional Table 1) contained the unrelated AAATT sequence at its 5’ end (Figure 1, A).

To get constructs containing WT and Δ76 NRS coding sequences under the control of a T7 RNA polymerase promoter which are fused to a DNA sequence coding 3 binding sites for the coat protein of phage MS2, the MS2 DNA sequences were excised from plasmid pADML-M3 (30) by BamHI-HindIII digestions and were subcloned into plasmid pUC18 (pMS2 plasmid). To obtain the control pT7MS2 plasmid, a T7 promoter was generated by annealing primers O-4672 and O-4673 and the resulting DNA was inserted between the BamHI and SacI sites located upstream of the MS2 DNA in plasmid pMS2. To get pNRS-MS2 and pNRS Δ76-MS2 plasmids, the NRS WT and NRS Δ76 DNA fragments were PCR amplified with O-4666 (containing a T7 promoter sequence) and O-4667 and inserted into a BamHI/SacI digested pMS2 plasmid. Other NRS variant DNAs, NRS 22, NRS 24, NRS 26 and NRS 28 were

obtained by PCR mutagenesis using the primers pairs O-5422/O-5423, O-5424/O-5425, O-5426/O-5427 and O-5428/O-5429, respectively (Additional Table 1).

For *in vitro* splicing assays, plasmid pC2-NRS and its derivatives were produced by insertion of the NRS DNA sequences in the sense (pC2-NRS and derivatives) or antisense orientations (pC2-NRS Inv) into the intron of the previously described C2 HIV construct built with HIV-1 BRU cDNA (31). These constructions were made in two steps using PCR primers described in Additional Table 1. First, the HIV sequence from position 4871 to 5053 was PCR amplified and cloned between the PstI and EcoRI sites of plasmid pUC18. Secondly, a DNA fragment containing a T7 RNA polymerase promoter followed by the HIV sequence including exon 1 (nts 1 to 289) and the intron from position 290 to 340 and a second fragment corresponding to RSV NRS WT were PCR amplified, ligated together using BglII sites introduced at their extremities and cloned into the HindIII and PstI sites of plasmid pC2 to obtain pC2-NRS. The same strategy was used to insert the NRS sequence in the reverse orientation (NRS Inv) and the NRS Δ76 sequence. Other NRS variant DNAs, NRS 22, NRS 24, NRS 26 and NRS 28 were obtained by PCR mutagenesis as described above.

To produce plasmids used for transfection assays, fragments from pC2, pC2-NRS WT and pC2-NRS Inv constructs were amplified using O-5824 and O-5825 and cloned between the NcoI and XbaI sites of plasmid pGL3c (Promega), so that, C2-NRS DNA sequences replaced the luciferase gene and were under the control of the SV40 promoter (pGL3-C2, pGL3-C2 NRS WT and pGL3-C2 NRS Inv, respectively). The SV40 poly(A) signal sequence was located at the other extremity. Plasmid pXJ42-9G8 (32) was used for overexpression of the 9G8 SR protein. The empty pXJ42 plasmid was used in control experiments. Plasmid phRGB (Promega) was used to verify the efficiency of transfection during overexpression experiments. The pBluescript SK-II (Stratagene) was used in order to reach a final amount of 2 and 4 μg of plasmid during the overexpression and siRNA experiments, respectively.

**RNA transcription**

For RNA secondary structure analysis and RNP purification using MS2-RNAs, unlabeled RNAs were transcribed using 1.7 to 3 pmols of PCR products and T7 RNA polymerase in conditions previously described (33). Transcripts were purified by electrophoresis on 5 % polyacrylamide gel as previously described (33). For *in vitro* splicing assays, 5 μg of plasmid were linearized by EcoRI and used as the template for the synthesis of unlabeled capped RNAs as previously described, except that [$\alpha$-$^{32}$P] UTP was omitted (31). After digestion of

the DNA matrix with 5 µl of TURBO DNase (Ambion), RNAs were purified using the GenElute kit (Sigma-Aldrich) according to manufacturer's recomandations.

**RNA secondary structure analysis**

RNA secondary structure analyses were performed on 100 ng of RNA transcript, in the presence of 3.125 mM $MgCl_2$ in buffer D (73 mM KCl, 0.15 mM EDTA, 0.35 mM DTT, 0.18 mM PMSF, 15 % glycerol, 15 mM HEPES-KOH pH 7.9) and in the presence of 5 µg of yeast tRNA mixture. RNA was pre-incubated for 5 min at 65°C in 15 µl of buffer D and after slow cooling to 30°C, 0.5 U of RNase T1 (Roche), 0.25 U of RNase T2 (Invitrogen) or $5.10^{-5}$ U of RNase V1 (Kemotex) were added and incubation was continued for 10 min at 30°C. RNase V1 hydrolyses were stopped by addition of 1 µl of 100 mM EDTA and 20 µg of yeast tRNA. RNase T1 and T2 hydrolyses were stopped by addition of 20 µg of yeast tRNA, followed by phenol extraction and ethanol precipitation. Kethoxal and DMS modifications were performed as previously described (34). Positions of enzymatic and chemical cleavages were identified by primer extension analyses using Avian Myeloblastosis Virus reverse transcriptase (Q Biogene) and the 5'-end labeled primers described in Additional Table 1. Primer 5'-end-labeling, annealing and extensions were performed as previously described (34). The sequencing ladder was obtained as usual, using the RNA sequencing method (34) with the same primer as for identification of cleavage and modification positions.

**RNA secondary structure prediction and phylogenetic analysis**

Predictions of secondary structure were made with the Mfold software version 3.2 (available at http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi) (35,36). The Mfold calculated thermodynamic values correspond to RNA in 1 M NaCl at 37°C. Results of experimental analyses were introduced in the computer search, in order to define the most stable secondary structure which had the best fit with the experimental data.

The GenBank accession number of RSV and related ASLV sequences used in the comparative secondary structure analysis are : L29198 (Schmidt Ruppin A), AF052428 (Schmidt Ruppin B), DQ365814 (ALV MQNCSU (A)), AY027920 (ALV ADOL-7501), AB303223 (ALV Stain Tym_S90), AF033810 (Fujinami Sarcoma Virus), V01170 (Y53 Sarcoma Virus).

**Footprinting assays on complexes formed with recombinant 9G8 protein**

To form 9G8-NRS RNP complexes, 100 ng of cold NRS WT RNA transcript (1.17 pmol) and 5 µg of yeast tRNA were mixed in 15 µl of buffer D containing 3.125 mM $MgCl_2$. NRS WT RNA transcript was denatured for 5 min at 65°C and after slow cooling at 30°C, recombinant protein was added. Then, 0.7 or 1.5 µg (28.5 and 57 pmol, respectively) of purified recombinant 9G8 protein expressed in baculovirus and purified as previously described (37) were added. Complexes were formed by incubation for 10 min at 30°C and then, 0.75 U of T1 RNase was added for 10 min at 30°C. The reaction was stopped by addition of 20 µg of yeast tRNA mixture and by rapid cooling to 0°C, immediately followed by phenol-extraction. As a control, enzymatic digestion was performed with the same amount of NRS WT RNA transcript and yeast tRNA incubated in the same conditions except that 9G8 protein was missing. Positions of enzymatic cleavages were identified by extension of 5'-end-labeled primers (Additional Table 1), in conditions previously described (34). The sequencing ladder was obtained as usual, using the RNA sequencing method (34) with the same primer as for identification of cleavage and modification positions. Experiments were all repeated several times using different batches of RNA transcripts.

**MS2 affinity chromatography**

The procedure used was an adaptation of the one previously described (29). MS2-tagged purified WT or variant NRS RNAs (250 pmol) were denatured, renatured, and incubated with a 5 fold molar excess of purified MS2-MBP fusion protein at 4°C for 30 min. This protein was prepared as previously described (38). The RNA-MS2-MBP complexes formed were incubated for 30 min at 4°C, with 4.3 mg of HeLa nuclear extract in a 1.1 ml of buffer D containing 60 µg of yeast tRNA. The HeLa nuclear extract was purchased from CILBiotech. Then, 100 µl of amylose beads (GE Healthcare) equilibrated in buffer D were incubated with the complexes for 1 h at 4°C under constant agitation. Three successive washes with buffer D were performed, and RNP complexes were eluted by incubation with 100 µl of buffer D containing 15 mM maltose (30 min at 4°C). About 10% of the proteins contained in the eluted RNP complexes were fractionated by 10% SDS-PAGE for Western blot analysis. Another 10% of the eluted material was phenol extracted and RNAs were fractionated on a 10% denaturing polyacrylamide gel for Northern blot analysis. Finally, for WT NRS-MS2 and control MS2 RNP complexes, the remaining part of the eluate was fractionnated by SDS-PAGE and the gel was systematically sliced. Then, proteins extracted from the bands were analysed by Nano-liquid chromatography-mass spectrometry experiments, using a CapLC

capillary liquid chromatography system (Waters, Milford, MA, USA) coupled to a hybrid quadrupole time-of-flight mass spectrometer (Q-TOF II, Waters) according to standard protocols (39).

***In vitro* splicing assays**

Twenty fmol of purified pre-mRNA that contained the WT NRS sequence, its reverse sequence or variant NRS sequences inserted in the C2 HIV-1 pre-mRNA were incubated for 2 h at 30°C in a 22 µl standard reaction containing 8 µl of HeLa cell nuclear extract (31). To study the influence of protein 9G8 on NRS activity, splicing reactions were performed with a 1:1 mixture of HeLa cell cytoplasmic extract and HeLa cell nuclear extract and 180 or 270 ng of purified recombinant human 9G8 protein (produced in baculovirus) or 500 ng of commercial purified BSA used as a control. Splicing reactions were stopped by addition of 20 µg of Proteinase K, followed by phenol/chloroform extraction. RNAs were ethanol precipitated and dissolved in 20 µl of bi-distilled $H_2O$. Reverse transcriptions were performed by addition of 2 µl of this solution in 20 µl of 1x MMLV buffer (Promega) containing 5 pmol of O-5825 primer, 1 mM of each dNTP, 1 µl of RNAguard (GE Healthcare) and 200 U of MMLV reverse transcriptase (Promega). Incubation was for 1 h at 37°C. In order to detect spliced RNAs, PCR amplifications were performed with the O-6246/O-6519 primer pair. To detect both spliced and unspliced RNAs, the PCR amplification was performed with the O-6519/O-6187 primer pair. PCR reactions were performed in semi-quantitative conditions using 50 µl of 1x GoTaq buffer containing 1 µl of reverse transcriptase products, 200 µM of each dNTP, 1 µM of each primers and 1.25 U of GoTaq polymerase (Promega). The following protocol was used : initial denaturation for 5 min at 94°C, followed by 30 cycles of amplification for analysis of spliced plus unspliced RNAs (Sp+UnSp) or 35 cycles of amplification for analysis of spliced RNAs (Sp). Each cycle included incubations for 30 s at 94°C, 45 s at 46°C and 45 s at 72°C. After agarose gel electrophoresis, ethidium bromide fluorescence of RT-PCR products was quantified using a digital ccd camera (GeneGenius, SynGene) and the SynGene GeneTools version 3.08 software. Each *in vitro* splicing experiment was repeated 3 times using different batches of RNAs and for each assay, RT-PCR amplifications were performed in triplicate. The mean values of the three RT-PCR experiments and their standard deviations were established. This allowed us to estimate the mean values of splicing efficiencies and their standard deviations. Student's *t* paired test was used for statistical analysis.

***In cellulo* splicing assays**

HeLa cells were grown in Dulbecco modified Eagle's medium (DMEM) medium (Sigma-Aldrich) supplemented with 10 % of fetal calf serum and 2 mM glutamine. About 24 h before transfection, cells were seeded in six well plates at a density of 200,000 cells per well. To test for the effect of 9G8 expression, cells were cotransfected with 0.5 µg of plasmid pGL3-C2, pGL3-C2 NRS WT or pGL3-C2 NRS Inv expressing the hybrid pre-mRNAs, 0.5 µg of pXJ42-9G8 expressing 9G8 and 0.9 µg of empty pXJ42 plasmid, together with 0.1 µg of phRGB plasmid encoding the renilla luciferase, this plasmid was used to verify transfection efficiency. In control experiments the pXJ42-9G8 plasmid was omitted and 1.4 µg of empty pXJ42 plasmid was used in order to maintain the same total amount of plasmid. Transfections were performed with the JetPEI (Polyplus-transfection kit) in a DNA/JetPEI ratio of 1:2, according to manufacturer's recommendations.

For siRNA experiments, HeLa cells were grown in OptiMEM medium (Invitrogen), supplemented with 5 % fetal calf serum, in the absence of antibiotics. As above, 24 h before transfection, cells were seeded in six well plates at a density of 200,000 cells per well. Cells were transfected using lipofectamine 2000 (Invitrogen) according to the manufacturer's recommendations. Two successive transfections of 100 nM of siRNA directed either against 9G8 mRNA (si9G8, (40)) or against the Renilla luciferase transcript of plasmid phRGB (siLuc, (41)) were carried out at 24 h interval. Cells were transfected 24 h after the second siRNA treatment with the following plasmid mixture : 0.5 µg of either pGL3-C2, pGL3-NRS WT or pGL3-C2 NRS Inv and 3.5 µg of pBSK plasmid per well.

To estimate the *in cellulo* splicing efficiencies, total cellular RNA was extracted with the GenElute kit (Sigma-Aldrich), 48 h after the unique transfection step for overexpression experiments and 24 h after the third transfection step for siRNA experiments. Then, 4 µg of total cellular RNA were treated with 4 U of TURBO DNase (Ambion) in a 20 µl of TURBO DNase buffer 1X for 30 min at 37 °C. The reaction was stopped by adding 20 µl of 30 mM EDTA and heating at 75°C for 15 min. Finally, 1 µg of DNase treated RNA was subjected to reverse transcription, using 0.5 pmol of oligo$(dT)_{18}$ as primer and PCR amplification and quantification were performed as described above. Two independent experiments were performed for the 9G8 overexpression and siRNA experiments, respectively. In each experiment, total cellular proteins were extracted and analysed by SDS-PAGE followed by Western blot analysis, using antibodies specific for 9G8 and β tubulin (Sigma-Aldrich) as previously described (41) .

## RESULTS

### Experimental analysis of the RSV NRS 2D structure

To analyse experimentally the 2D structure of the entire RSV NRS in solution, we used RNases T1 and T2 in conditions such that they only cleaved single-stranded segments and V1 RNase which cleaves specifically double-stranded and stacked RNA regions. Inspection of the NRS bordering sequences in ASLV retroviruses revealed the conservation of a C rich segment 5' to NRS and of a G rich segment 3' to NRS. These two segments showed conserved complementarity in the ASLV family. We hypothesised that they may form a helix closing the NRS. Therefore, our structural analyses were performed on an *in vitro* transcript produced by T7 RNA polymerase that encompassed the RSV RNA segment from position 691 to 951 (NRS encompasses nts 703 to 930) (Figure 1, A). The cleaved NRS residues were identified by primer extension analyses. Efficient binding of RT primers requires their base pairing with single-stranded residues. Hence, an AAAUU sequence, which was not expected to interfere with the helix predicted to be formed by segments 691-700 and 940-947, was added at the 3' end of the transcript (Figure 1, A). The RT primer (O-2955) complementary to this additional sequence and segment 939-951 (Additional Table 1) was efficient to identify cleavages located between positions 901 and 936. We defined two other primers (O-1619 and O-1617), which allowed us to identify RNase cleavage positions in the entire NRS region (Additional Table 1). The 2D structure probing experiments were performed after RNA refolding in conditions described in Materials and Methods. Experiments were repeated several times and reproducible results were obtained. Examples of primer extension analyses are shown in Figure 1, B (B1 to B3). Enzymatic probing experiments were completed by chemical probing experiments. We used DMS, which methylates single-stranded adenosines at position N-1 and single-stranded cytosines at position N-3, and kethoxal that modifies unpaired guanine residues at position N-3. Location of modifications was also achieved by primer extension analyses, representative examples are shown in Figure 1, C (C1 and C2).

### Computer assisted determination of the RSV NRS structure

The previously proposed computer predicted structure for RSV NRS WT (13) could not explain the results of our chemical and enzymatic probing experiments (Figure 2, A). To search for the most stable 2D structure in agreement with our experimental data, we used the Mfold software (35,36) and introduced our data as constraints. The structure identified (Figure 2, B) has a very low free energy (- 83.5 kcal/mol). In agreement with our previous probing data, which has served as the basis for NMR analysis of the NRS-3' structure (23), segment 907 to 929 containing both

the U1 and U11 binding sites, forms an individual stem-loop structure (SLS4). In the established structure, segment 717-747, which was previously found to bind the ASF/SF2 SR protein (19), also folds into an individual stem-loop structure (SLS1). As expected, the 3' and 5' termini of the NRS form an irregular helix (H1) that is extended by the complementary sequences bordering the NRS. The segment from position 750 to 778, which was previously proposed to bind several SR proteins and hnRNP H, forms a very long irregular helix (H2) by interaction with segment 877-902. The sequence from position 781 to 850, which is largely deleted in the functional NRS Δ76 truncated element, is involved in formation of a long irregular SLS2 stem-loop structure, and fragment 854 to 868 which is also deleted in NRS Δ76, forms a small stem-loop structure (SLS3).

**The main structural features of NRS are conserved in the NRS Δ76 variant**

As the variant NRS Δ76 is functional, we hypothesized that its 2D structure should not be drastically modified. To verify this hypothesis, we probed its structure experimentally. The NRS Δ76 transcript contained the same additional sequences at the 5' and 3' extremities as the full length NRS transcript used above, and was analysed as the full length NRS. The data obtained are in perfect agreement with formation of helices H1 and H2 and stem-loops SLS1 and SLS4 in NRS Δ76 (Figure 2, C and Additional Figure 1). The parts of the SLS2 and SLS3 stem-loops, which are remaining in NRS Δ76, form two base pairs extending helix H2 and a large terminal loop. Therefore, the 2D structure elements, formed by sequences which are important for binding of U1 snRNA, U11 snRNA (SLS4), SR and hnRNP proteins (SLS1, H1 and H2, respectively) are conserved in the NRS Δ76.

**NRSs from viruses of the ASLV family can fold according to the structure established for RSV NRS**

LR9 is an RSV related recombinant ALV virus which contains the *gag pol* and *env* genes derived from the UR2-associated virus (UR2AV) and the LTR of ring-necked pheasant virus (RPV) (27). Its NRS element is highly similar to RSV NRS (10 base substitutions) and shows similar efficiency in an heterologous context (27). To test whether the presence of these 10 base substitutions are compatible with the 2D structure established for the RSV NRS, we produced an LR9 NRS transcript homologous to that used for RSV NRS and probed its structure with T1, T2 and V1 RNases. As shown in Additional Figure 2, base substitutions are either located in single strands, or reinforce the stability of helix H2. Therefore, the possibility to form the structure is

preserved. Note that when repeating the same experiment on the defective LR9 NRS deletion mutant (ΔLR9), which is missing the segment 735 to 776 (27), we observed strong 2D structure alterations (Additional Figure 3) : the SLS1 structure disappeared and, helix H2 and SLS2 were remodelled. Therefore, both the sequence and 2D structure of this defective NRS are profoundly altered.

Then, for a larger investigation on the possible phylogenetic conservation of the NRS 2D structure, we took advantage of the available sequences for several ASLVs genomic RNAs (42). As shown in the Additional Figure 4, A to G, the nucleotide sequences involved in formation of helix H1 and SLS1 and SLS4 stems are highly conserved. Only one half of helix 2 has a conserved sequence. The various base substitutions, insertions or deletions in the viruses that we compared have no expected detrimental effect on the RNA 2D structure, which is in favour of a phylogenetic conservation of the NRS 2D structure.

**A minimal NRS element including helix H1 and stem-loops SLS1 and SLS4 is active *in vitro***

Then, based on the established structure, we generated internal truncation in the NRS in order to define the minimal NRS sequence required for NRS activity. To this end, we used a heterologous construct derived from the HIV-1 C2 construct (Figure 3, A1) (31) which produces a pre-mRNA containing a short intron bordered by the HIV-1 strong D1 5’ ss and the A2 3’ ss. Site A2 has a limited efficiency due to the presence of the ESSV splicing silencer that binds hnRNP A1 (43), and up to now, only an activation by the SR protein ASF/SF2 was detected at this site (44). Therefore, we hypothesized, that the C2 construct might be a favourable model to compare activities of WT and variant NRSs. The NRS WT coding sequence was inserted in the middle of the C2 intron (Figure 3, A2). In these conditions, the NRS pseudo 5’ ss was located 278 nt downstream from the HIV-1 D1 5’ ss. This distance was selected as it was previously shown to be optimal for NRS activity (13). As *in vitro* splicing efficiency is dependent on the length of the intron to be spliced, we compared the splicing efficiency of C2-NRS WT RNA to a control C2 RNA containing the NRS sequence in the reverse orientation (C2-NRS Inv RNA). Transcripts were spliced in a Hela cell nuclear extract, as described in Materials and Methods. The ratio of spliced (Sp) *versus* unspliced plus spliced (Sp+UnSp) products was quantified by RT-PCR as described in Materials and Methods (Figure 3, A2, and Additional Table 1). For each transcript, splicing assays were performed in triplicate and highly reproducible results were obtained as evidenced by data in the Additional Figure 5, A. As expected, insertion of a non-related 266-nt long sequence in C2 RNA decreased the *in vitro* splicing efficiency by about 40% (C2-NRS Inv RNA compared to C2 RNA). However, the decrease observed for the C2-NRS WT RNA was

reproducibly more drastic (80%), demonstrating NRS activity in the C2 context. We thus considered that this model is valid and could be used to test the effects of deletions in NRS on splicing efficiency.

As a positive control, we used the functional NRS Δ76 mutant and we produced four other variants (Figure 3, A2). In variant 28 both stem-loops SLS2 and SLS3, as well as a very small portion of helix H2, were deleted. In variant 22, helix H2 and stem-loops SLS2 and SLS3 were deleted. In this variant, only helix H1 and the stem-loops SLS1 and SLS4 are present and they are joined by a 7-nt long single-stranded segment. Stem-loop SLS4 was deleted in variant 24 and SLS1 was absent in variant 26. As expected NRS Δ76 was as efficient in splicing inhibition as the NRS WT (Figure 3, B1 lane 4 and B2). Interestingly, NRS 28, which is smaller than NRS Δ76 however had a splicing inhibitory property superior to that of NRS Δ76 (Figure 3, B1 lane 8 and B2). No splicing inhibition was detected for NRS 24, which confirmed the requirement of the U1 snRNA binding site for NRS activity (Figure 3, B1 lane 6 and B2). The same absence of inhibition was observed for NRS 26, showing that deletion of stem-loop SLS1 (position 717 to 747) has the same negative effect on NRS activity as deletion of stem-loop SLS4 (Figure 3, B1 lane 7 and B2). Accordingly, previous data (20) showed a strong loss of NRS activity with a deletion which included part of the SLS1 sequence (deletion from position 720 to 744). Importantly, NRS 22, which only contained stem-loops SLS1 and SLS4 and helix H1, had a negative effect on splicing similar to that of NRS Δ76. Therefore, our data reveal that a very small truncated NRS variant containing only stem-loop structures SLS1 and SLS4 and helices H1 and H2 is sufficient to inhibit splicing *in vitro*.

**Sequences corresponding to SLS1 and SLS4 are required for recruitment of SR proteins**

As a complement to the *in vitro* splicing inhibition study of truncated NRSs described above, we attempted to identify the nuclear proteins recruited by these NRSs mutant. To this end, we used an RNP purification approach that is based on the fusion of MS2 sequences to the studied RNA (29,30). The WT, Δ76, 22, 24, 26 and 28 NRS variant RNAs, described above, were transcribed with three binding sites for the MS2 coat protein fused at their 3' ends (Figure 4, A). Each of these RNAs was bound to a recombinant fusion protein containing the MS2 coat protein RNA binding domain and the *E. coli* Maltose Binding Protein (MBP). The RNA-protein complexes obtained were then incubated with a Hela cell nuclear extract competent for *in vitro* splicing. The WT and variant NRS RNPs formed were bound to amylose beads, washed and eluted by addition of maltose, as described in Materials and Methods.

Both the RNA and the protein composition of the purified complexes were analysed. The presence of U1 snRNA and MS2 tagged RNAs in the purified complexes was evidenced by Northern blot analysis using specific oligonucleotides (Additional Table 1), and quantified as described in Figure 4. As expected, only the deletion of the SLS4 sequence had a marked negative effect on U1 snRNA association (Figure 4, B and C). Proteins present in the RNP complexes formed with the WT NRS were subjected to SDS-PAGE, the bands were cut and analysed by mass spectrometry. We detected proteins already found to bind the entire NRS element (ASF/SF2, hnRNP H) (19,21,45) among a large number of other proteins. For a comparative analysis of SR and hnRNP H and A1 proteins in RNP complexes formed with the WT and variant NRSs, we turned to the Western blot approach using specific antibodies (Figure 4, D). As expected the SR proteins ASF/SF2, 9G8 and SC35 and the hnRNP H proteins were detected in complexes formed with NRS WT. They were also detected with NRSs Δ76, 22 and 28. In contrast, weak amounts of ASF/SF2 and only traces of 9G8 and SC35 were detected in complexes formed with NRSs 24 and 26, suggesting that both the SLS1 and SLS4 sequences are required for efficient binding of SR proteins. The strong binding of hnRNP A1 to each of the truncated forms of NRSs (Figure 4, D) suggested a possible involvement of this protein in regulation of NRS activity.

**The SR protein 9G8 has a stimulatory effect on NRS activity**

Our observation of a strong binding of the SR protein 9G8 on NRS together with previous description of its association with NRS (21) encouraged us to test for a possible effect of this protein on NRS activity. First, we used the *in vitro* splicing assay based on C2 derived contructs. The C2-NRS WT RNA was spliced in HeLa cell nuclear extract in the absence or the presence of increasing amounts of recombinant 9G8 protein which had been produced in baculovirus. As illustrated in Figure 5, A and Additional Figure 5, B, the NRS inhibitory property was significantly and reproducibly increased by addition of the recombinant 9G8 protein, whereas addition of BSA had no significant effect. For the C2-NRS WT RNA, since only 50 % of the splicing activity was detected upon addition of 270 ng of 9G8.

To verify that 9G8 also stimulates the NRS inhibitory property *in cellulo*, the C2, C2-NRS WT and C2-NRS Inv DNA sequences from the *in vitro* constructs were inserted within plasmid pGL3c, in such a way that they replaced the luciferase gene. The resulting plasmids, pGL3-C2, pGL3-C2 NRS WT or pGL3-C2 NRS Inv were co-transfected in HeLa cells with either the empty plasmid pXJ42 or plasmid pXJ42-9G8 expressing 9G8 protein and 48 h after transfection and the amounts of spliced (Sp) and spliced plus unspliced (Sp+UnSp) transcripts were estimated

by RT-PCR as described in Figure 6. The over-expression of protein 9G8 by a factor of about 3 reproducibly decreased splicing of the C2-NRS WT RNA, but not of splicing of C2 and C2-NRS Inv transcripts (Figure 6, A and Additional Figure 5, C). Therefore, the stimulatory effect of 9G8 on the NRS activity that we discovered by *in vitro* assays also exists *in cellulo*.

To confirm the existence of this stimulatory effect *in cellulo*, we then diminished the level of 9G8 in HeLa cells by using a si9G8 directed against the endogeneous 9G8 mRNA (si9G8) (40) . As a control we used a siRNA directed against the Renilla luciferase (siLuc) (41). Cells were transfected twice with siRNAs and 48 h later were transfected by the C2, C2-NRS Inv or C2-NRS WT constructs for 24 h. Two independent experiments were performed and in our transfection conditions, the 9G8 concentration was decreased to a moderate extent (factor of about 2) (Figure 6, B, B1). In these conditions, C2 RNA splicing was slightly decreased and we reproducibly detected an increased splicing of C2-NRS WT RNA (factor of about 2.5) while no increase was detected in the control experiment performed with the siLuc siRNA) (Figure 6, B, B2 and B3). Therefore, we concluded that the strength of splicing regulatory activity of RSV NRS depends upon the yield of expression of protein 9G8.

**Helix H2 is required for stimulation of the NRS splicing inhibition by 9G8**

Then, we used again *in vitro* splicing assays in order to define which NRS regions are important to get the 9G8 stimulatory effect. Tests were performed three times on each of the NRS Δ76, 22 and 28 variants (Figure 5, B and Additional Figure 5, D). Strong increases of the Δ76 and 28 NRS inhibitory properties were reproducibly detected in the presence of recombinant 9G8. In contrast, addition of 9G8 protein did not stimulate the splicing inhibitory activity of the NRS 22 variant, showing that in spite of the strong association of 9G8 with SLS1, the helix H2 sequence which is missing in NRS 22 is required for stimulation of the RSV NRS activity by 9G8. This suggested a direct binding of protein 9G8 on sequences in helix H2.

Therefore, for a better delineation of the 9G8 binding sites on RSV NRS, we then performed footprinting assays on RNP complexes formed by incubation of RSV WT NRS transcripts with the recombinant 9G8 protein at two distinct [9G8]/[RNA] ratios 25 and 50, respectively, as described in Materials and Methods. Formation of RNP complexes was verified by EMSA experiments (data not shown). The patterns of T1 RNase digestion of free and complexed RNA were compared by primer extension analysis followed by gel electrophoresis. Examples of cDNA fractionation are shown in Figure 7, A, A1 to A3. Neither SLS3 nor SLS4 are protected. At the lowest [9G8]/[RNA] ratio, protections are mainly detected in SLS1 and its bordering sequences and in the purine rich large bulged loop in SLS2 (Figure 7, B). Upon increasing 9G8

concentration, protections in these regions were reinforced and new protections are detected in helix H2 (Figure 7, B). Interestingly, these data are in agreement with our finding of a requirement of the SLS1 sequence for efficient association of 9G8 to RSV NRS. They are also in agreement with previous data suggesting that there are 9G8 protein binding sites in the RSV NRS segment extending from positions 740-798 (21). However, our data reveals that 9G8 binding sites in RSV NRS are not restricted to this segment.

## DISCUSSION

Here we describe the first experimental analysis of the 2D structure of the entire RSV NRS and based on a proteomic analysis, followed by *in vitro* splicing assays, we show that the SR protein 9G8 stimulates the RSV NRS splicing inhibition property.

### Each of the NRS functional sequences belongs to a define 2D structure motif and the possibility to form these motifs is conserved in ASLVs

Retroviruses generally have highly structured genomic RNAs in order to resist to digestion by RNases of the host cells. The RSV NRS element does not escape to this rule. It folds into a tight 2D structure. Other ASLV NRSs can adopt this structure due to compensatory and semi-compensatory base substitutions in the stems and the location of several point mutations and insertion of nucleotide stretches in single-stranded segments.

The structural motif established by NMR for a small RSV NRS segment containing the pseudo 5' ss (23) is included in the established structure. This motif corresponds to the apical part of the SLS4 stem-loop structure. Importantly, formation of this apical part of SLS4 was previously shown to be needed for efficient binding of U1 snRNP to the pseudo 5' ss (23).

Our data change the conception of NRS architecture which was based on a previous computer prediction (13). We found that the ASF/SF2 binding region, which is also shown here to bind protein 9G8, folds into a stem-loop structure facing the stem-loop structure containing the U1 and U11 binding sites. In the previous computer model, these two sequences were in part base-paired together (13). Absence of protection by protein 9G8 in the SLS4 stem-loop, while sequences in SLS1 are protected, is in agreement with the formation of two distinct structural elements by these two NRS sequences. However, as the sequences bordering the SLS1 and SLS4 stem-loops form a large internal loop, one can imagine that the flexibility of this internal loop facilitates the establishment of interactions between partners of SLS1 and SLS4 : protein ASF/SF2 and/or 9G8 bound to SLS1 and the U1 snRNP bound to SLS4. The nucleotide sequences of these two stem-loop structures are highly conserved. The possibility

to form helix H1 was not taken into consideration in the previous model. Based on the present data, one of its functions may be to stabilize the NRS structure. In addition, helix H1, whose primary sequence is highly conserved, contains one of the G tracts (nts 935 to 942) which are needed to bind hnRNP H and this binding is important for U11 snRNP association (46). Conservation of this G tract in ASLVs is also in accordance with the presence of similar G tracts downstream from several of the minor 5' ss that bind U11 snRNP (46). Sequences involved in helix H2 formation were previously proposed to bind SR proteins (47). Here, we show that this helix is required for activation of NRS by protein 9G8 and delineated the H2 segment protected by this SR protein.

Noticeably, the strong sequence conservations of SLS1, SLS4, helix H1 and the bottom part of helix H2 can not be simply explained by specific needs for coding capacity, since all the NRS sequences code for a single polypeptide (the Gag protein). Consequently, the strong conservations of some parts of the NRS sequence and of its overall 2D structure altogether likely reflect specific needs for NRS activity.

**SLS1, SLS2 linked together by helix H1 are sufficient for splicing inhibition**

The HIV-1 derived construct that we designed for *in vitro* and *in cellulo* assay of the NRS activity turned to be efficient. It shows that utilization of the pair of HIV-1 D1 and A2 sites can be efficiently down regulated by the heterologous NRS sequence. Accordingly, like the RSV *env* and *src* 3' ss, the HIV-1 site A2 is poorly efficient due to the presence of : i) suboptimal polypyrimidine tract and branch site sequences and, ii) an ESSV inhibitory element (31,43). By using the HIV-1-NRS heterologous construct, we confirmed the requirement of sequences contained in SLS1 for NRS activity and showed that these sequences are as much important for NRS activity as sequences in SLS4. This strongly suggest that the previously identified ASF/SF2 binding sites in SLS1 (19,20) and/or the presently identified 9G8 binding sites in SLS1 is/are needed to get NRS activity. These data are consistent with the absence of activity of the NRS ΔLR9 which is missing the sequence from position 735 to 776 (27).

The NRS Δ76 variant was previously defined as the minimal functional NRS element (13). Interestingly, although shorter in length than NRS Δ76, variant NRS 28, designed on the basis of the RNA 2D structure, inhibited *in vitro* splicing as efficiently as the NRS Δ76. Most importantly, we could design a very short NRS variant (NRS 22), which is still able to inhibit

splicing *in vitro*. This is the first observation of the activity of a very small NRS variant which only contains SLS1 linked to SLS4 by helix H1.

**Increased protein 9G8 concentration reinforces the NRS inhibitory property**

Whereas the SR protein 9G8 was previously shown to bind NRS, its activity on this element had never been tested directly. Here, by *in vitro* and *in cellulo* splicing assays, we show the dependence of the NRS activity on the 9G8 protein concentration. The effect on NRS activity observed upon addition of 9G8 protein in nuclear extract was almost opposite to that observed upon decreasing the 9G8 concentration by siRNAs *in cellulo*. Importantly also, addition of extra 9G8 protein in nuclear extract increased the NRS WT, Δ76 and 28 activity, but had no effect on the NRS 22 inhibition property, in spite of the presence of 9G8 binding sites in SLS1. We propose the following interpretation of our splicing data. First, by binding to SLS1, protein 9G8 may have in combination with ASF/SF2 a positive basic effect on NRS activity. Secondly, at higher concentration, by binding to sequences in helix H2 which are present in NRS Δ76 and 28 but are missing in NRS 22, protein 9G8 may have a supplementary stimulatory effect on NRS activity.

Altogether, our data bring strong support to the implication of protein 9G8 in modulation of NRS activity. Obviously, these data were obtained by using human nuclear extracts and recombinant human SR proteins, which are more easily available than splicing competent avian nuclear extracts. However, based on the strong sequence conservation of SR proteins and components of the splicing machinery in mammalian and avian organisms, we propose that activation of NRS by the 9G8 avian counterpart protein also takes place in avian cells.

**Deletions of SLS1 or SLS4 both have a negative effect on SR binding to NRS in nuclear extract**

Our adaptation of the MS2 based RNP purification method to the purification of RNP complexes formed by WT or variant NRSs upon incubation in nuclear extract led to an unexpected result : binding of SR proteins to NRS is found to depend on both SLS1 and SLS4, while binding of U1 snRNP is strictly dependent on the presence of SLS4. The dependence of the binding of the SR proteins ASF/SF2 on the presence of SLS1 is in agreement with previous data showing its direct binding to a segment overlapping SLS1 (19,20). Finding of a requirement of SLS1 for recruitment of proteins 9G8 is in accord with our 9G8 footprinting assays (Figure 7, A and B).

Our finding of a strong negative effect of SLS4 deletion on SR protein binding is unexpected. Firstly, because we found no significant protection of SLS4 in RNP complexes formed with protein 9G8 (Figure 7, A and B), secondly because the NRS segment extending from position 715 to 748 binds alone the recombinant ASF/SF2 protein (19). One possible explanation is that, in nuclear extracts containing a high level of protein hnRNP A1, binding of the U1 snRNP or other splicing factors on SLS4 may help SR protein to compete with hnRNP A1 and to bind to SLS1. Accordingly, we detected a strong binding of hnRNP A1 on all the truncated NRSs that were tested, which was confirmed by footprinting assays (data not shown).

One model to explain the overall data presented in this paper is that the role of SR proteins in NRS activity is probably not to reinforce U1 snRNP binding on NRS but to couple the NRS pseudo 5' ss to the RSV 3' ss which is needed for splicing inhibition. SR proteins and hnRNP A1 likely compete for binding to NRS and in future studies the possible role of hnRNP A1 will have to be investigated.

**FUNDING**

This work was supported by the «Centre National de la Recherche Scientifique» (CNRS), the french Ministry «de l'Enseignement Supérieur et de la Recherche» ; by the French National Agency for Research (ANR) [ANR-05-BLAN-0261-03] ; by the European Alternative Splicing Network of excellence [LSHGCT 2005 518238] and by funds of «Région Lorraine» (Bioingénierie project). A. Bar, V. Marchand and G. Khoury were fellows from the French «Ministère de la Recherche et de l'Enseignement».

**ACKNOWLEDGEMENTS**

K. Beemon (John Hopkins, Baltimore) is strongly thanked for her generous gift of plasmids pRSVNeo-Int, pΔ76Neo-Int, pLR9 and pΔLR9 and helpful discussions. R. Lührmann is acknowledged for providing plasmid pADML-M3 containing the three MS2 coat protein binding sites and plasmid pMAL containing the MS2-MBP coding sequence. R. Lührmann and K. Hartmuth (Max Planck Institute, Göettingen) are strongly thanked for providing MS2 RNP chromatography affinity protocol and technical formation in their lab. L. Fouillen, S. Sanglier and A. Van Dorsselaer (Laboratoire de Spectrométrie de Masse Bio-Organique, Strasbourg) are thanked for mass spectrometry analyses. L. Ayadi and F. Oillo-Blanloeil are thanked for their participation to footprinting assays. J. Ugolini is thanked for her excellent technical assistance.

CILBiotech society (Mons, Belgium) is thanked for providing nuclear and cytoplasmic extracts form HeLa cells.

**REFERENCES**

1. Coffin, J.M., Hughes, S.H. and Varmus, H.E. (1997) Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, NY.
2. McNally, M.T. (2008) RNA processing control in avian retroviruses. *Front Biosci*, **13,** 3869-3883.
3. Katz, R.A. and Skalka, A.M. (1990) Control of retroviral RNA splicing through maintenance of suboptimal processing signals. *Mol Cell Biol*, **10,** 696-704.
4. Fu, X.D., Katz, R.A., Skalka, A.M. and Maniatis, T. (1991) The role of branchpoint and 3'-exon sequences in the control of balanced splicing of avian retrovirus RNA. *Genes Dev*, **5,** 211-220.
5. Zhang, L. and Stoltzfus, C.M. (1995) A suboptimal src 3' splice site is necessary for efficient replication of Rous sarcoma virus. *Virology*, **206,** 1099-1107.
6. Zhang, L., Simpson, S.B. and Stoltzfus, C.M. (1996) Selection and characterization of replication-competent revertants of a Rous sarcoma virus src gene oversplicing mutant. *J Virol*, **70,** 3636-3644.
7. Amendt, B.A., Simpson, S.B. and Stoltzfus, C.M. (1995) Inhibition of RNA splicing at the Rous sarcoma virus src 3' splice site is mediated by an interaction between a negative cis element and a chicken embryo fibroblast nuclear factor. *J Virol*, **69,** 5068-5076.
8. Ogert, R.A., Lee, L.H. and Beemon, K.L. (1996) Avian retroviral RNA element promotes unspliced RNA accumulation in the cytoplasm. *J Virol*, **70,** 3834-3843.
9. Simpson, S.B., Guo, W., Winistorfer, S.C., Craven, R.C. and Stoltzfus, C.M. (1998) The upstream, direct repeat sequence of Prague A Rous sarcoma virus is deficient in mediating efficient Gag assembly and particle release. *Virology*, **247,** 86-96.
10. Simpson, S.B., Zhang, L., Craven, R.C. and Stoltzfus, C.M. (1997) Rous sarcoma virus direct repeat cis elements exert effects at several points in the virus life cycle. *J Virol*, **71,** 9150-9156.
11. Guo, W., Winistorfer, S.C. and Stoltzfus, C.M. (2000) Selective inhibition of splicing at the avian sarcoma virus src 3' splice site by direct-repeat posttranscriptional cis elements. *J Virol*, **74,** 8513-8523.

12. Arrigo, S. and Beemon, K. (1988) Regulation of Rous sarcoma virus RNA splicing and stability. *Mol Cell Biol*, **8,** 4858-4867.
13. McNally, M.T., Gontarek, R.R. and Beemon, K. (1991) Characterization of Rous sarcoma virus intronic sequences that negatively regulate splicing. *Virology*, **185,** 99-108.
14. McNally, M.T. and Beemon, K. (1992) Intronic sequences and 3' splice sites control Rous sarcoma virus RNA splicing. *J Virol*, **66,** 6-11.
15. O'Sullivan, C.T., Polony, T.S., Paca, R.E. and Beemon, K.L. (2002) Rous sarcoma virus negative regulator of splicing selectively suppresses SRC mRNA splicing and promotes polyadenylation. *Virology*, **302,** 405-412.
16. Gontarek, R.R., McNally, M.T. and Beemon, K. (1993) Mutation of an RSV intronic element abolishes both U11/U12 snRNP binding and negative regulation of splicing. *Genes Dev*, **7,** 1926-1936.
17. Hibbert, C.S., Gontarek, R.R. and Beemon, K.L. (1999) The role of overlapping U1 and U11 5' splice site sequences in a negative regulator of splicing. *Rna*, **5,** 333-343.
18. McNally, L.M. and McNally, M.T. (1999) U1 small nuclear ribonucleoprotein and splicing inhibition by the rous sarcoma virus negative regulator of splicing element. *J Virol*, **73,** 2385-2393.
19. McNally, L.M. and McNally, M.T. (1996) SR protein splicing factors interact with the Rous sarcoma virus negative regulator of splicing element. *J Virol*, **70,** 1163-1172.
20. McNally, L.M. and McNally, M.T. (1998) An RNA splicing enhancer-like sequence is a component of a splicing inhibitor element from Rous sarcoma virus. *Mol Cell Biol*, **18,** 3103-3111.
21. Fogel, B.L. and McNally, M.T. (2000) A cellular protein, hnRNP H, binds to the negative regulator of splicing element from Rous sarcoma virus. *J Biol Chem*, **275,** 32371-32378.
22. Giles, K.E. and Beemon, K.L. (2005) Retroviral splicing suppressor sequesters a 3' splice site in a 50S aberrant splicing complex. *Mol Cell Biol*, **25,** 4397-4405.
23. Cabello-Villegas, J., Giles, K.E., Soto, A.M., Yu, P., Mougin, A., Beemon, K.L. and Wang, Y.X. (2004) Solution structure of the pseudo-5' splice site of a retroviral splicing suppressor. *Rna*, **10,** 1388-1398.
24. Varani, L., Hasegawa, M., Spillantini, M.G., Smith, M.J., Murrell, J.R., Ghetti, B., Klug, A., Goedert, M. and Varani, G. (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,** 8229-8234.

25. Donahue, C.P., Muratore, C., Wu, J.Y., Kosik, K.S. and Wolfe, M.S. (2006) Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. *J Biol Chem*, **281,** 23302-23306.
26. Jiang, Z., Cote, J., Kwon, J.M., Goate, A.M. and Wu, J.Y. (2000) Aberrant splicing of tau pre-mRNA caused by intronic mutations associated with the inherited dementia frontotemporal dementia with parkinsonism linked to chromosome 17. *Mol Cell Biol*, **20,** 4036-4048.
27. Smith, M.R., Smith, R.E., Dunkel, I., Hou, V., Beemon, K.L. and Hayward, W.S. (1997) Genetic determinant of rapid-onset B-cell lymphoma by avian leukosis virus. *J Virol*, **71,** 6534-6540.
28. Polony, T.S., Bowers, S.J., Neiman, P.E. and Beemon, K.L. (2003) Silent point mutation in an avian retrovirus RNA processing element promotes c-myb-associated short-latency lymphomas. *J Virol*, **77,** 9378-9387.
29. Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianferani, S., Van Dorsselaer, A., Clerc, P., Avner, P. *et al.* (2010) 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol*, **8,** e1000276.
30. Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H. and Luhrmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol*, **26,** 5528-5543.
31. Damier, L., Domenjoud, L. and Branlant, C. (1997) The D1-A2 and D2-A2 pairs of splice sites from human immunodeficiency virus type 1 are highly efficient in vitro, in spite of an unusual branch site. *Biochem Biophys Res Commun*, **237,** 182-187.
32. Bourgeois, C.F., Popielarz, M., Hildwein, G. and Stevenin, J. (1999) Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol Cell Biol*, **19,** 7347-7356.
33. Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M. and Branlant, C. (1996) Secondary structure of the yeast Saccharomyces cerevisiae pre-U3A snoRNA and its implication for splicing efficiency. *Rna*, **2,** 1079-1093.

34. Jacquenet, S., Ropers, D., Bilodeau, P.S., Damier, L., Mougin, A., Stoltzfus, C.M. and Branlant, C. (2001) Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing. *Nucleic Acids Res*, **29,** 464-478.
35. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31,** 3406-3415.
36. Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, **288,** 911-940.
37. Cavaloc, Y., Bourgeois, C.F., Kister, L. and Stevenin, J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA*, **5,** 468-483.
38. Jurica, M.S. and Moore, M.J. (2002) Capturing splicing complexes to study structure and mechanism. *Methods*, **28,** 336-345.
39. Miguet, L., Pacaud, K., Felden, C., Hugel, B., Martinez, M.C., Freyssinet, J.M., Herbrecht, R., Potier, N., van Dorsselaer, A. and Mauvieux, L. (2006) Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. *Proteomics*, **6,** 153-171.
40. Blaustein, M., Pelisch, F., Tanos, T., Munoz, M.J., Wengier, D., Quadrana, L., Sanford, J.R., Muschietti, J.P., Kornblihtt, A.R., Caceres, J.F. *et al.* (2005) Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol*, **12,** 1037-1044.
41. Dreumont, N., Hardy, S., Behm-Ansmant, I., Kister, L., Branlant, C., Stevenin, J. and Bourgeois, C.F. (2009) Antagonistic factors control the unproductive splicing of SC35 terminal intron. *Nucleic Acids Res*, **38,** 1353-1366.
42. Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*, **35,** D61-65.
43. Bilodeau, P.S., Domsic, J.K., Mayeda, A., Krainer, A.R. and Stoltzfus, C.M. (2001) RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element. *J Virol*, **75,** 8487-8497.
44. Ropers, D., Ayadi, L., Gattoni, R., Jacquenet, S., Damier, L., Branlant, C. and Stevenin, J. (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,** 29963-29973.

45. Cook, C.R. and McNally, M.T. (1998) SR protein and snRNP requirements for assembly of the Rous sarcoma virus negative regulator of splicing complex in vitro. *Virology*, **242,** 211-220.
46. McNally, L.M., Yee, L. and McNally, M.T. (2006) Heterogeneous nuclear ribonucleoprotein H is required for optimal U11 small nuclear ribonucleoprotein binding to a retroviral RNA-processing control element: implications for U12-dependent RNA splicing. *J Biol Chem*, **281,** 2478-2488.
47. Fogel, B.L., McNally, L.M. and McNally, M.T. (2002) Efficient polyadenylation of Rous sarcoma virus RNA requires the negative regulator of splicing element. *Nucleic Acids Res*, **30,** 810-817.
48. Schwartz, D.E., Tizard, R. and Gilbert, W. (1983) Nucleotide sequence of Rous sarcoma virus. *Cell*, **32,** 853-869.
49. Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K. *et al.* (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature*, **313,** 277-284.

**FIGURE LEGENDS**

**Figure 1. (A) Schematic representation of the RSV Prague C (PrC) genome and of the NRS RNAs used for secondary structure analyses.** The architecture of the RSV PrC genome is shown on top of the Figure. The 5' ss and 3' ss are indicated and boxes represent the open reading frames. The grey box depicts the NRS element (nt 703-930) located in the *gag* gene used for secondary structure experiments. Viral sequences belonging to the RSV segments located upstream and downstream from NRSs are in white. The non viral AAAUU sequence at the 3' end of the transcripts is shown. Delimitations of NRS are indicated according to NRS nucleotide numbering (48).
**(B and C) Primer extension analyses of enzymatic cleavages and chemical modifications of NRS WT.** (**B**) Primer extension analyses of enzymatic cleavages in NRS WT (position 691 to 951) (B1 to B3), (**C**) primer extension analyses of chemical modifications in NRS WT (position 710 to 890) (C1 and C2). T1, T2, V1, D and Ke above the lanes indicate T1, T2, V1 RNase digestions and DMS or Kethoxal modifications, respectively. Conditions for digestions, modifications and reverse transcription are given in Materials and Methods. Lanes c correspond to control experiments with untreated RNA ; lanes U, G, C, and A correspond to the sequencing ladder. Primers used for reverse transcription are indicated below each autoradiogram (see Additional Table 1 for their nucleotide sequences). Nucleotide numbering according to (48) is indicated on the left side of the autoradiograms. Positions of the stem-loop structures (SLS1, SLS2, SLS3 and SLS4) and the helices (H1 and H2) are shown on the right side of the autoradiograms.

**Figure 2. Schematic representation of the experimental data on the previously proposed structure for RSV NRS WT (13) in (A), and on the secondary structure models that we established for RSV WT and RSV Δ76 NRSs (B and C, respectively)**. Cleavages by enzymes and modifications by chemical reagents are depicted as shown in the inset (modified residues are circled, phosphodiester bonds cleavages by T1, T2 or V1 RNases are indicated by arrows surmounted by lozenges, circles or squares, respectively). The intensities of cleavages and modifications are indicated by colors (red, orange and green for strong, medium and low, respectively). Positions of nucleotides are according to (48). Stem-loop structures (SLS) and helices (H) are numbered. The free energies of the proposed structures as calculated by the Mfold software (37°C, in 1 M NaCl) (35,36) are given below each structure. The

hybridisation sites used for primer extensions are indicated by blue arrows. The underlined sequences at the 3' end of the RNAs correspond to the added unrelated nucleotides.

**Figure 3. Comparison of splicing inhibition properties of NRS WT and Δ76, 22, 24, 26 and 28 NRS variants in the C2 pre-mRNA context.**

(**A**) Schematic representation of the HIV C2 pre-mRNA (A1) (31) and C2-NRS heterologous pre-mRNAs (A2). Black boxes and thin lines represent exons and intron, respectively. Numbering of C2 RNA is according to (49). End joining of the two HIV-1 sequences in the C2 intron is depicted by an arrow. The D1 5' ss and A2 HIV-1 3' ss as well as the ESSV regulatory element are shown (43). Heterologous C2-NRS constructs, created by insertion of the WT and variant Δ76, 22, 24, 26 and 28 NRSs in the intron, are schematically represented below. Segments deleted in the variant NRS elements are in grey. The RT (in grey), and PCR (in black) primers used for evaluation of the yields of C2 and C2-NRS pre-mRNAs splicing efficiencies are represented by arrows above the C2-NRS construct. Yields of spliced RNAs (Sp) and of spliced plus unspliced products (Sp+UnSp) were analysed by PCR amplifications using the pairs of primers marked by Sp and Sp+UnSp, respectively.

(**B**) RT-PCR analysis of *in vitro* splicing products of the C2 and heterologous C2-NRS pre-mRNAs. *In vitro* splicing assays were performed as described in Materials and Methods. Experiments were repeated 3 times using different batches of RNA (Additional Figure 5, A). The relative splicing efficiencies of the various constructs were estimated by semi-quantitative RT-PCR performed in triplicate (see additional Figure 5, A). A representative example of cDNA fractionation obtained for each of the tested pre-mRNAs is shown (B1). Identities of RT-PCR products spliced (Sp) or spliced plus unspliced (Sp+UnSp) products are indicated. The identity of the pre-mRNA used in the assay is given above the lane. The Sp/Sp+UnSp pre-mRNA ratio obtained for the C2-NRS Inv construct was taken as the reference (100%). The Sp/Sp+UnSp ratios established for the other constructs are given as a percentage of this reference. The mean values of the Sp/Sp+UnSp percentages determined for three independent splicing reaction and their standard deviations are given (B2). * $P<0.05$, ** $P<0.01$, NS : not significant.

**Figure 4. Only functional NRS variants bind SR proteins.** (**A**) Schematic representation of NRS-MS2 and NRS-MS2 derived RNAs used in affinity selection experiments. The three binding sites for the MS2 coat protein (MS2 RNA) were fused to the 3' end of the NRS RNAs. Formation of RNP complexes and their purification is described in Materials and

Methods. (**B** and **C**) Northern blot analysis of RNA from purified RNP complexes, using 5'-end-labeled U1 snRNA (O-6524) and MS2 RNA (O-6534) specific probes (Additional Table 1). The RNAs used to form the RNP complexes are indicated above the lanes. M, is a mixture of MS2 and all NRS-MS2 RNAs used as size markers. NE corresponds to RNAs extracted from cell nuclear extract without addition of exogeneous RNA. Intensities of the bands were quantified and the ratios of U1 snRNA *versus* NRS-MS2 RNA are given in Panel C**.** The U1 snRNA/NRS-MS2 RNA ratio obtained for the WT NRS RNA is taken as the 100 % reference and ratios obtained for the other NRS-MS2 RNA are given as a percentage of this reference. The mean values and standard deviations of two experiments are shown. **(D**) Western blot analyses of SR and hnRNP proteins contained in the purified RNP complexes formed with the various NRS-MS2 RNAs. Proteins extracted from equal amounts of purified complexes were analysed by Western blot using antibodies specific for the SR and hnRNP proteins indicated on the right side of the blot. The identity of the NRS-MS2 RNA used in the assay is given above the gel. Two µg of total proteins of HeLa nuclear extract (NE) were loaded as a reference.

**Figure 5. Protein 9G8 enhances NRS splicing inhibition properties *in vitro*. (A**) WT NRS activity is strongly enhanced by 9G8 protein *in vitro*. C2 (depicted in Figure 3, A1), C2-NRS WT and C2-NRS Inv pre-mRNAs (depicted in Figure 3, A2) were incubated for 2 h in a 1:1 mixture of HeLa cell cytoplasmic and nuclear extracts in the absence or the presence of purified recombinant human 9G8 protein (180 and 270 ng) or BSA (500 ng). RNAs were subjected to RT-PCR analysis (in triplicate) as in Figure 3. One representative example of cDNA fractionation is shown (A1). Splicing experiments were repeated three times using different batches of RNA. For each tested pre-mRNA, the Sp/Sp+UnSp ratio obtained in the absence of 9G8 or BSA addition was taken as the 100 % reference and the Sp/Sp+UnSp ratios obtained in the presence of 180 or 270 ng of 9G8 or 500 ng of BSA were expressed as a percentage of this reference. The data for the 3 splicing experiments are shown in the Additional Figure 5, B. The mean values and the standard deviations of these experiments are given in A2. * $P<0.05$, ** $P<0.01$, NS : not significant.

(**B**) 9G8 increases the *in vitro* activity of the Δ76 and 28 variant NRSs but not of the 22 variant NRS. The C2-NRS Inv, C2-NRS Δ76, C2-NRS 22 and C2-NRS 28 pre-mRNAs were incubated in a 1:1 mixture of HeLa cell cytoplasmic and nuclear extract without (9G8 -) or with (9G8 +) 270 ng of 9G8 protein and were subjected to RT-PCR analysis as described above. Three independent experiments were performed and a representative example of

fractionation of the RT-PCR products is shown in B1. The Sp/Sp+UnSp ratios were quantified as in Figure 3 and the Sp/Sp+UnSp ratio obtained in the absence of 9G8 over-expression were taken as the 100% reference. The Sp/Sp+UnSp ratios obtained in the presence of 270 ng of 9G8 were expressed as a percentage of this reference. The percentages established for the three independent experiments are shown in Additional Figure 5, D. Their mean values and standard deviations are given in B2. * $P<0.05$, ** $P<0.01$, NS : not significant.

**Figure 6. NRS activity *in cellulo* depends on the level of protein 9G8 expression.** (**A**) Protein 9G8 also increases NRS activity *in cellulo.* DNAs of the C2, C2-NRS WT and C2-NRS Inv constructs described in Figure 3, A2 were cloned into plasmid pGL3c. HeLa cells were co-transfected with these constructs and a vector expressing the 9G8 SR protein (9G8 +) or an empty vector as the control (9G8 -) ; 48 h after transfection, total cellular proteins and RNAs were extracted as described in Materials and Methods and analysed by Western blot after SDS PAGE analysis (A1) and by RT-PCR (A2 and A3), respectively. The presence of equal amounts of extracted proteins in each assay was controlled by Western blots using β tubulin antibody and the level of 9G8 overexpression was evaluated by Western blot using a specific 9G8 antibody. For RT-PCR analysis of the extracted RNAs, the same primers as in Figure 3, A2 were used, except that the RT primer was an oligo(dT) (See Additional Table 1). Two independent experiments were performed and a representative example of cDNA fractionation is shown in A2. For each of the tested constructs, the Sp/Sp+UnSp ratios obtained in the absence of 9G8 overexpression was used as the 100 % reference and the Sp/Sp+UnSp ratios obtained in the presence of 9G8 overexpression are expressed as a percentage of this reference. The percentages established for the two independent experiments are shown in Additional Figure 5, C. Their mean values and standard deviations are shown in A3.

(**B**) NRS activity is strongly reduced upon decreasing the cellular 9G8 protein concentration. HeLa cells were treated with siRNA targeting the 9G8 mRNA (si9G8) or with control siRNA (targeting luciferase mRNA, siLuc) in conditions described in Materials and Methods. They were then transfected with plasmids pGL3-C2, pGL3-C2 NRS Inv or pGL3-C2 NRS WT. Total proteins and RNAs were extracted 24 hours after transfection and were analysed by Western blot (B1) and RT-PCR (B2 and B3), respectively, as described for the Panel A. Two independent experiments were performed and a representative exemple of cDNA fractionation is shown in B2. RT-PCR were performed in triplicate and quantified, as

described above. For each construct, the Sp/Sp+UnSp ratios obtained for the control siLuc siRNA was taken as the 100% reference and the Sp/Sp+UnSp ratios obtained for the si9G8 siRNA are expressed as a percentage of this reference. The mean values of the two independent experiments and their standard deviations are shown in panel B3.

**Figure 7. Enzymatic footprinting of 9G8 binding sites on RSV NRS RNA**

(**A1-A3**) Primer extension analyses of RSV NRS cleaved by T1 RNase in the absence or the presence of protein 9G8. RNA-protein complexes were formed at [9G8]/[RNA] molar ratios of 25 and 50, respectively. The absence of protein addition is indicated by (-). Enzymatic reactions were performed as described in Materials and Methods. Lanes U, G, C and A correspond to the sequencing ladder and primers used for reverse transcription are indicated below each autoradiogram. Nucleotide numbering in the RSV NRS RNA and positions of the ASF/SF2 (19) binding site are indicated on the left side of the autoradiogram. Positions of helices (H) and stem-loop structures (SLS) are shown on the right side of the autoradiograms. Cleavages by T1 RNase and protection generated by protein 9G8 are represented by arrows with two circles. Colours of the first and second circles represent the level of protection obtained at the 25 and 50 [protein]/[RNA] molar ratios, respectively (white, black, blue and purple for none, low, medium, and strong protection, respectively). (**B**) Schematic representation of the protections generated by protein 9G8 is depicted on the RSV NRS RNA 2D structure using the same rules as in panel A. Primers are indicated by blue arrows. The ASF/SF2 and snRNA U1 binding sites according to references (17,19) are highlighted in green and grey, respectively.

**Figure 1**

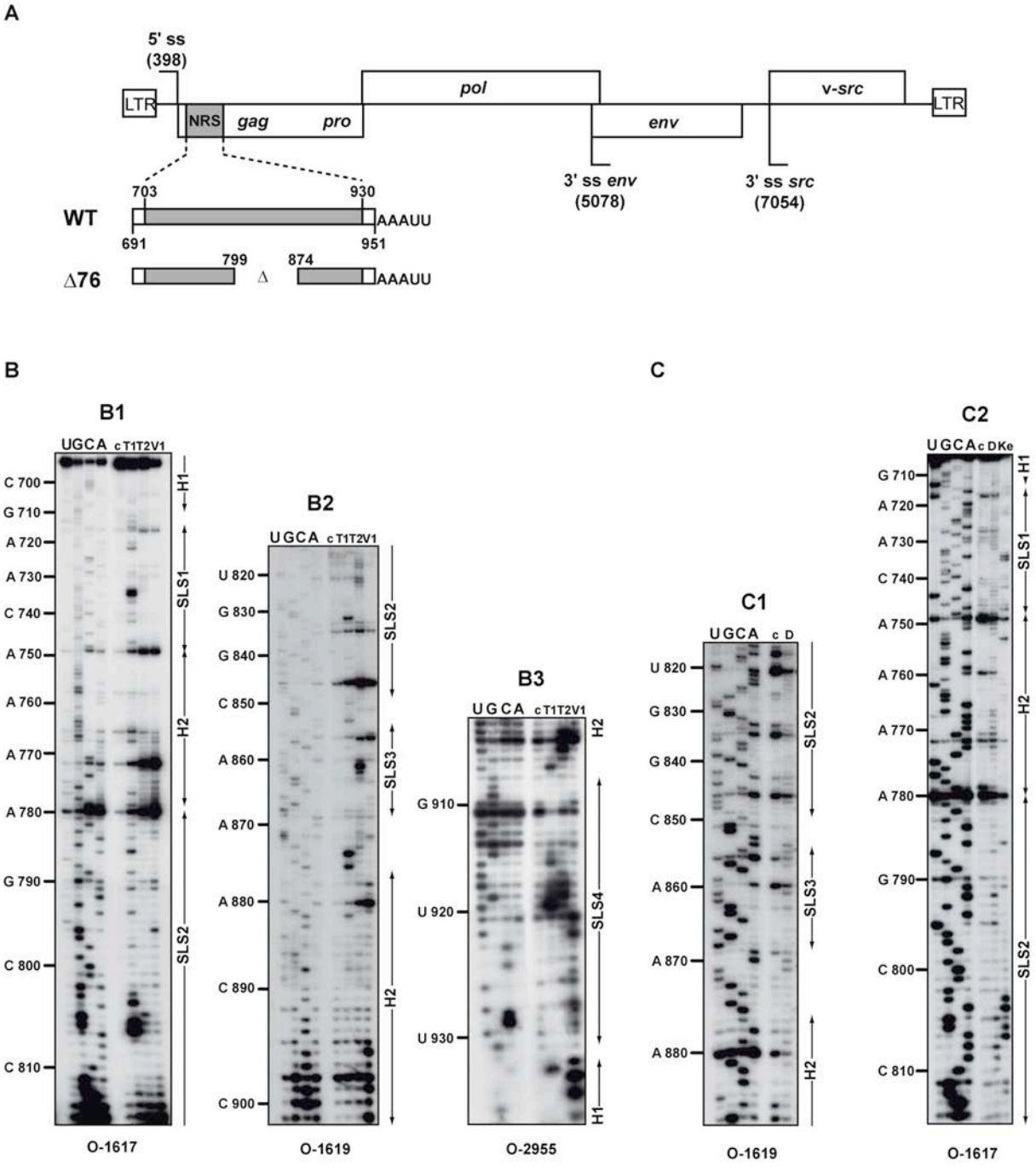

A
5' ss
(398)
LTR
NRS
gag
pro
pol
env
v-src
LTR
3' ss env
(5078)
3' ss src
(7054)
703
930
WT
AAAUU
691
951
799
874
Δ76
Δ
AAAUU
B
B1
UGCA cT1T2V1
C 700
G 710
A 720
A 730
C 740
A 750
A 760
A 770
A 780
G 790
C 800
C 810
H1
SLS1
H2
SLS2
O-1617
B2
UGCA cT1T2V1
U 820
G 830
G 840
C 850
A 860
A 870
A 880
C 890
C 900
SLS2
SLS3
H2
O-1619
B3
UGCA cT1T2V1
G 910
U 920
U 930
H2
SLS4
H1
O-2955
C
C1
UGCA c D
U 820
G 830
G 840
C 850
A 860
A 870
A 880
SLS2
SLS3
H2
O-1619
C2
UGCA c DKe
G 710
A 720
A 730
C 740
A 750
A 760
A 770
A 780
G 790
C 800
C 810
H1
SLS1
H2
SLS2
O-1617

**Figure 2**

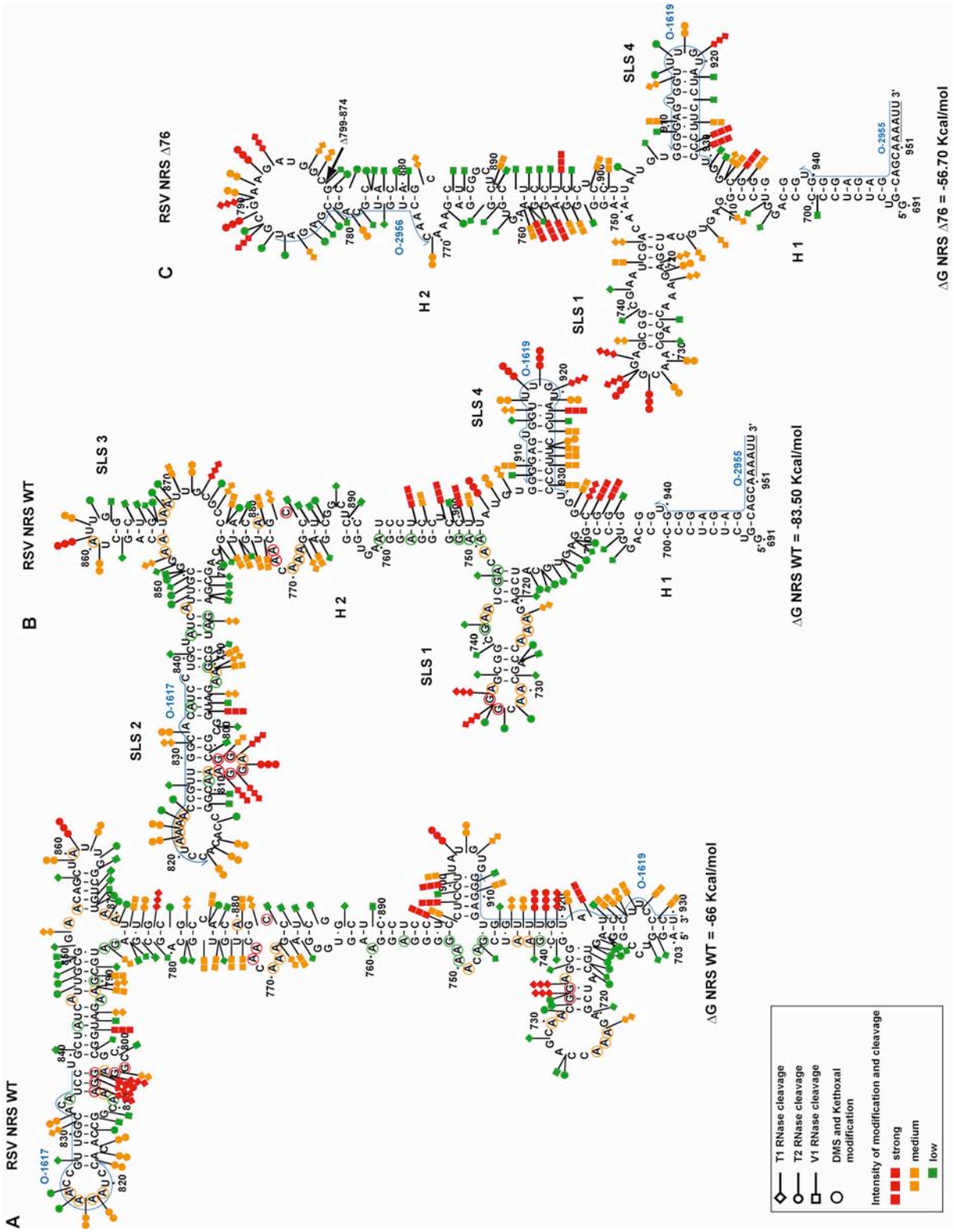

A
RSV NRS WT
ΔG NRS WT = -66 Kcal/mol
B
RSV NRS WT
SLS 2
SLS 3
SLS 1
SLS 4
H 2
H 1
ΔG NRS WT = -83.50 Kcal/mol
C
RSV NRS Δ76
SLS 1
SLS 4
H 2
H 1
ΔG NRS Δ76 = -56.70 Kcal/mol
T1 RNase cleavage
T2 RNase cleavage
V1 RNase cleavage
DMS and Kethoxal modification
Intensity of modification and cleavage
strong
medium
low

**Figure 3**

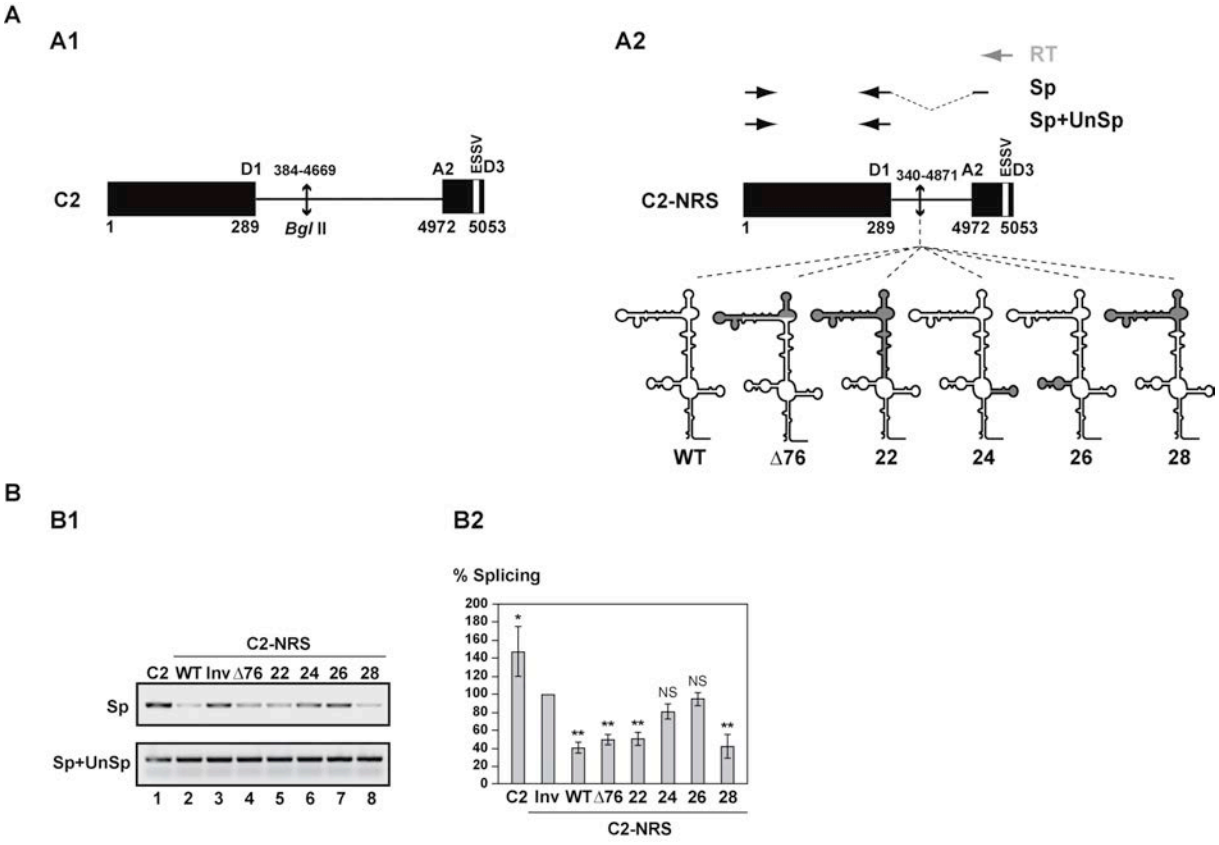

A
A1
C2
D1
384-4669
A2
ESSV
D3
1
289
Bgl II
4972
5053
A2
RT
Sp
Sp+UnSp
C2-NRS
D1
340-4871
A2
ESSV
D3
1
289
4972
5053
WT
Δ76
22
24
26
28
B
B1
C2-NRS
C2 WT Inv Δ76 22 24 26 28
Sp
Sp+UnSp
1 2 3 4 5 6 7 8
B2
% Splicing
200
180
160
140
120
100
80
60
40
20
0
*
**
**
**
NS
NS
**
C2 Inv WT Δ76 22 24 26 28
C2-NRS

**Figure 4**

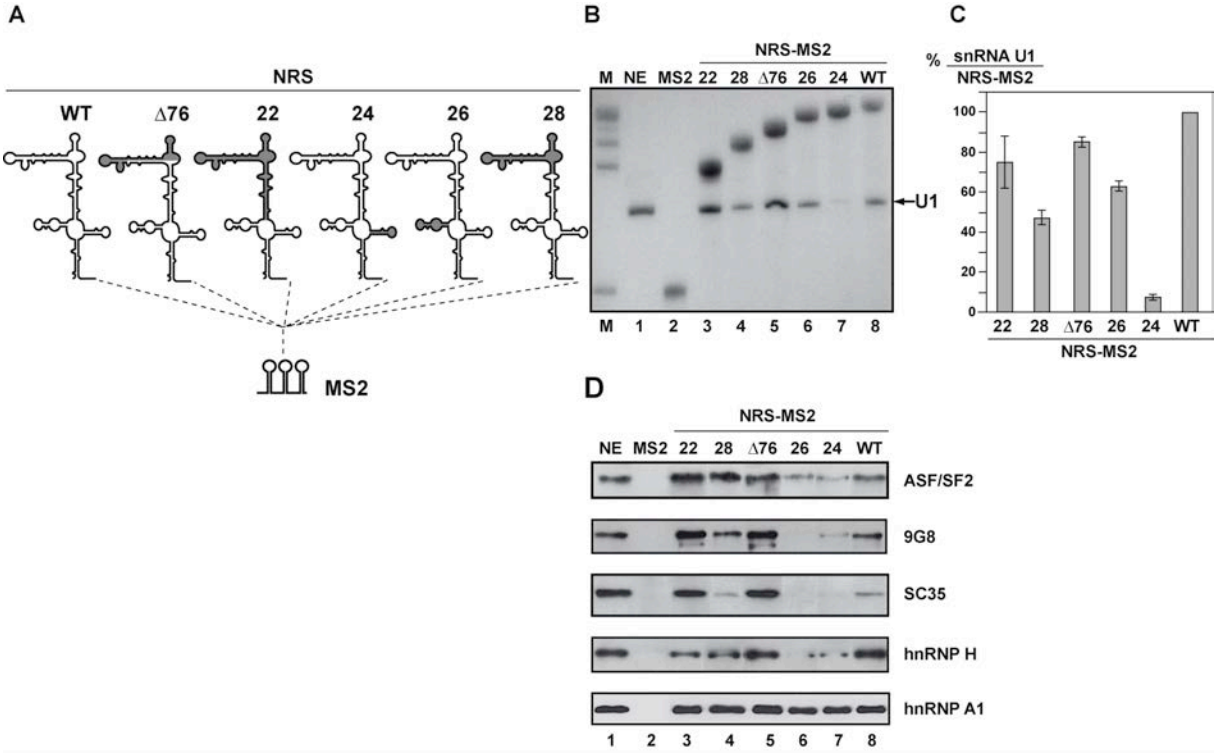

A
NRS
WT
Δ76
22
24
26
28
MS2
B
NRS-MS2
M NE MS2 22 28 Δ76 26 24 WT
U1
M 1 2 3 4 5 6 7 8
C
% snRNA U1
NRS-MS2
100
80
60
40
20
0
22 28 Δ76 26 24 WT
NRS-MS2
D
NRS-MS2
NE MS2 22 28 Δ76 26 24 WT
ASF/SF2
9G8
SC35
hnRNP H
hnRNP A1
1 2 3 4 5 6 7 8

**Figure 5**

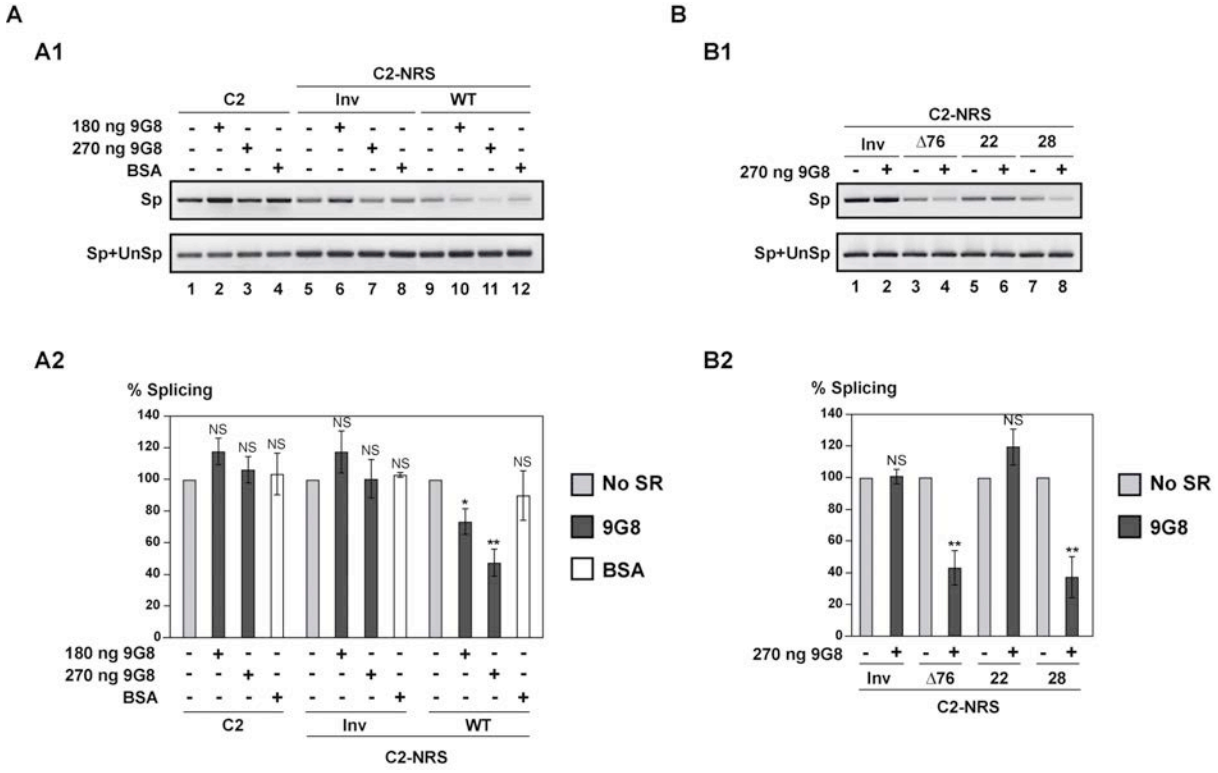

A
A1
C2
C2-NRS
Inv
WT
180 ng 9G8
270 ng 9G8
BSA
Sp
Sp+UnSp
1 2 3 4 5 6 7 8 9 10 11 12
A2
% Splicing
NS
*
**
No SR
9G8
BSA
C2
Inv
WT
C2-NRS
B
B1
C2-NRS
Inv
Δ76
22
28
270 ng 9G8
Sp
Sp+UnSp
1 2 3 4 5 6 7 8
B2
% Splicing
No SR
9G8
Inv
Δ76
22
28
C2-NRS

**Figure 6**

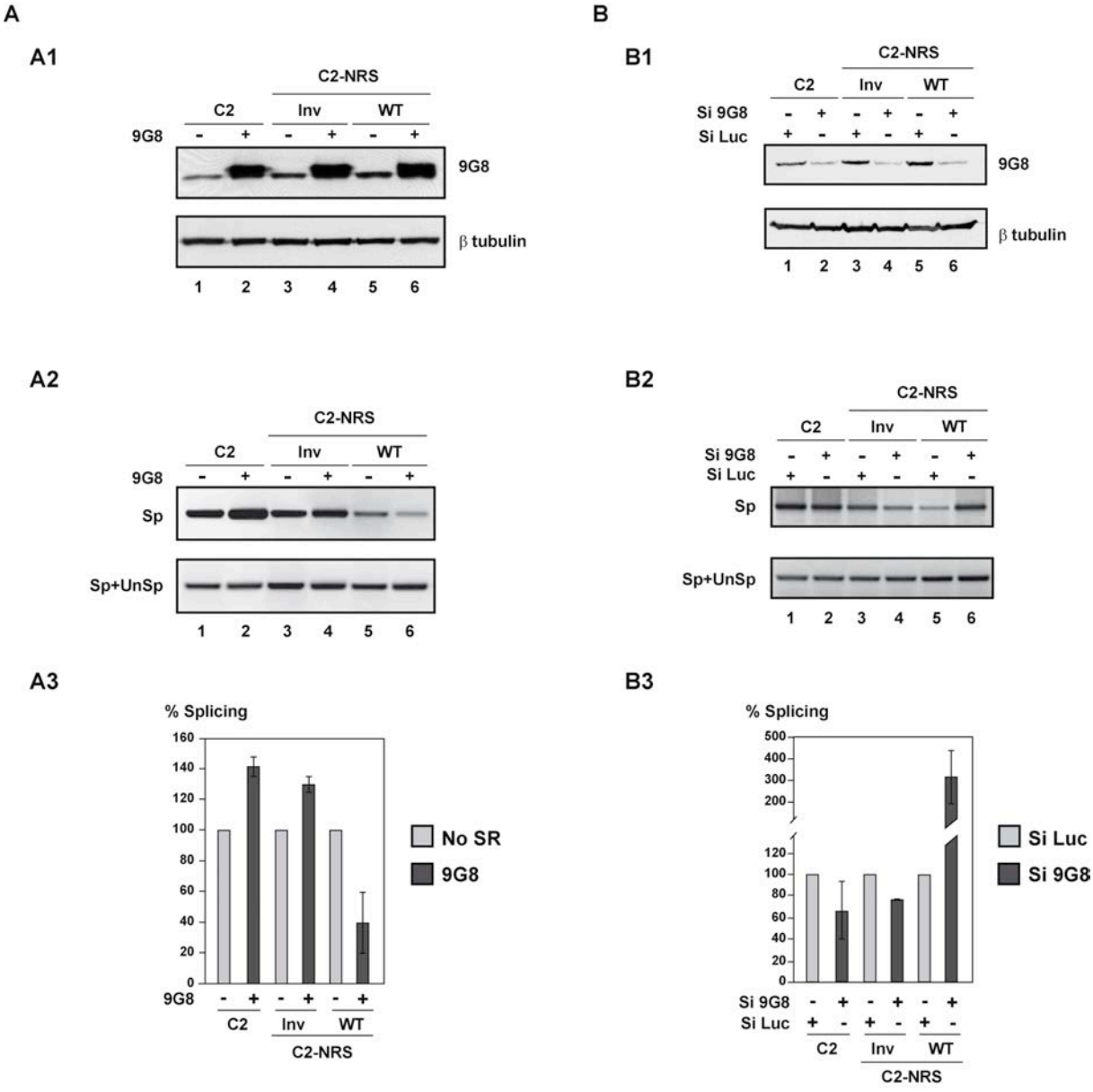
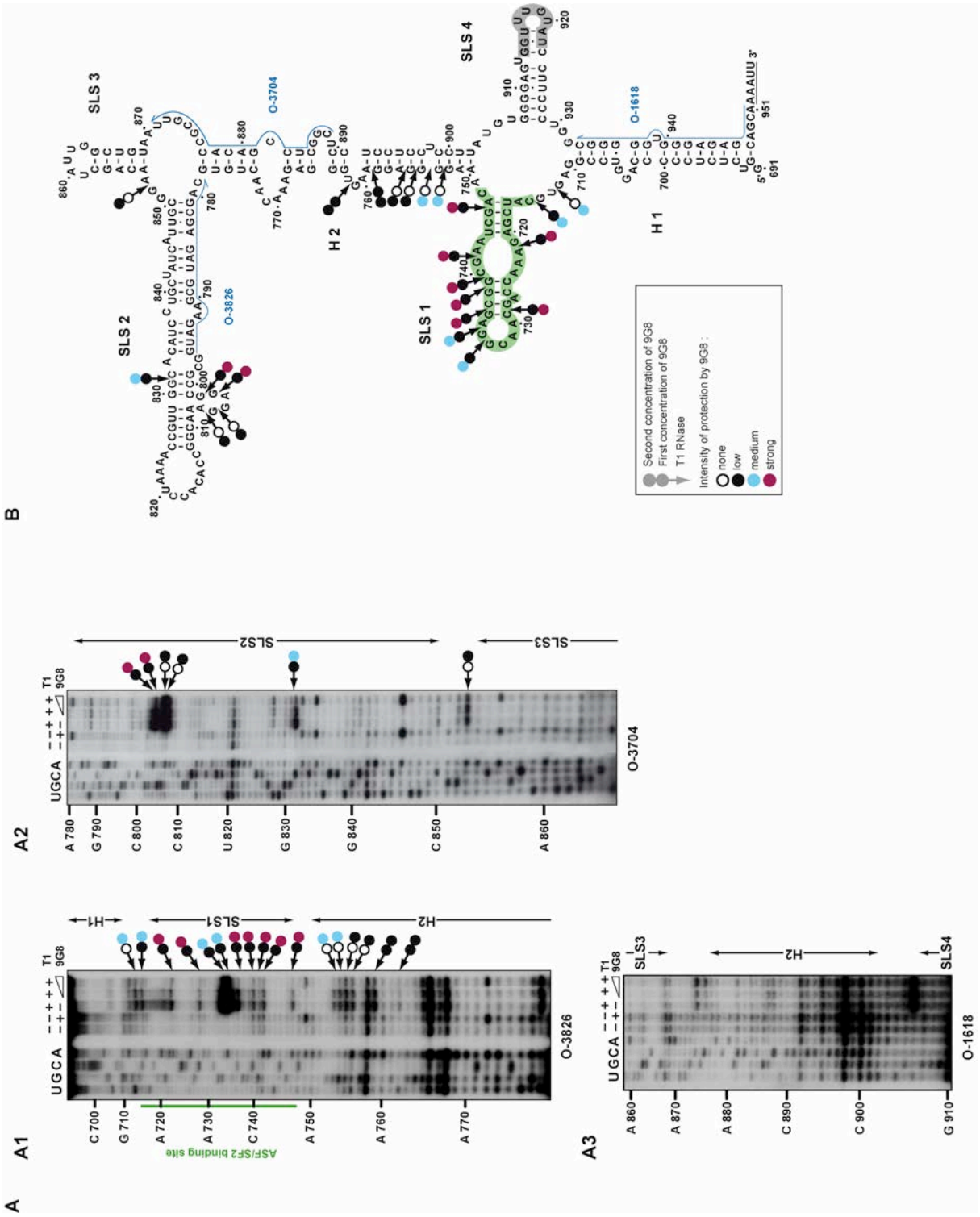

A
A1
C2-NRS
C2
Inv
WT
9G8
9G8
β tubulin
1 2 3 4 5 6
A2
C2-NRS
C2
Inv
WT
9G8
Sp
Sp+UnSp
1 2 3 4 5 6
A3
% Splicing
160
140
120
100
80
60
40
20
0
9G8
C2
Inv
WT
C2-NRS
No SR
9G8
B
B1
C2-NRS
C2
Inv
WT
Si 9G8
Si Luc
9G8
β tubulin
1 2 3 4 5 6
B2
C2-NRS
C2
Inv
WT
Si 9G8
Si Luc
Sp
Sp+UnSp
1 2 3 4 5 6
B3
% Splicing
500
400
300
200
120
100
80
60
40
20
0
Si 9G8
Si Luc
C2
Inv
WT
C2-NRS
Si Luc
Si 9G8

**Figure 7**


A
A1
UGCA
T1
9G8
H1
SLS1
H2
O-3826
ASF/SF2 binding site
A2
SLS2
SLS3
O-3704
B
SLS 3
SLS 2
SLS 4
SLS 1
H 2
H 1
O-3826
O-3704
O-1618
Second concentration of 9G8
First concentration of 9G8
T1 RNase
Intensity of protection by 9G8 :
none
low
medium
strong
A3
SLS3
H2
SLS4
O-1618

## SUPPLEMNETARY DATA

## FIGURE LEGENDS

**Additional Table 1.**

**Primers used in this study.** The number, sequence, and usage of each primer are given. Restriction sites are in boldfaces, T7 promoter sequences are underlined, unrelated nucleotides added at the 3' end of NRSs are underlined by a double line, (/) denotes extremities of deletions, (//) denotes the border between exon 1 and exon 2 in the RT-PCR primer.

**Additional Figure 1.**

**Secondary structure analysis and 2D structure model for the Δ76 NRS RNA, region from positions 691 to 951. (A)** Schematic representation of architecture of the RSV PrC genome genome and of the Δ76 NRS RNA used for secondary structure analyses. The 5' ss and 3' ss are indicated and boxes represent the open reading frames. The grey box depicts the Δ76 NRS element located in the *gag* gene. Viral sequences belonging to the RSV segments located upstream and downstream from NRS are in white. The non viral AAAUU sequence at the 3' end of the transcript is shown. Delimitations of NRS and of NRS deletions are indicated according to NRS nucleotide numbering (48).

**(B)** Primer extension analyses of enzymatic cleavages in Δ76 NRS (B1 to B3). Lanes c correspond to control experiments with untreated RNA ; lanes U, G, C, and A correspond to the sequencing ladder. Primers used for reverse transcription are indicated below each autoradiogram (see Additional Table 1 for their nucleotide sequences). Nucleotide numbering is indicated on the left side of the autoradiograms. Positions of the stem-loop structures (SLS) and the helices (H) are shown on the right side of the autoradiograms.

**(C)** Schematic representation of the experimental data on the secondary structure models that we established for RSV Δ76 NRS. Cleavages by enzymes are depicted as shown in the inset (phosphodiester bonds cleavages by T1, T2 or V1 RNases are indicated by arrows surmounted by lozenges, circles or squares, respectively). The intensities of cleavages are indicated by colors (red, orange and green for strong, medium and low, respectively). Positions of nucleotides are according to (48). Stem-loop structures (SLS) and helices (H) are numbered. The free energies of the proposed structures as calculated by the Mfold software (37°C, in 1 M NaCl) (35,36) are given below the structure. The hybridisation sites used for

primer extensions are indicated by blue arrows. The underlined sequences at the 3' end of the RNAs correspond to the added unrelated nucleotides.

**Additional Figure 2.**

**Secondary structure analysis and 2D structure model for the LR9 NRS, region from positions 691 to 951. (A)** Schematic representation of architecture of the RSV PrC genome genome and of the ALV LR9 NRS RNA used for secondary structure analyses. Indications are the same as in Additional Figure 1, A.

**(B)** Primer extension analyses of enzymatic cleavages in ALV LR9 NRS (B1 to B3). T1, T2, V1, above the lanes indicate T1, T2, V1 RNase digestions, respectively. Conditions for digestions and reverse transcription are given in Materials and Methods. Indications are the same as in Additional Figure 1, B.

**(C)** Schematic representation of the experimental data on the secondary structure model that we established for ALV LR9 NRS. Cleavages by enzymes are depicted as shown in the inset (phosphodiester bonds cleavages by T1, T2 or V1 RNases are indicated by arrows surmounted by lozenges, circles or squares, respectively). Indications are the same as in Additional Figure 1, C. Positions of nucleotides are according to (28) and the red nucleotides in LR9 NRS correspond to base substitutions in ALV compared to RSV.

**Additional Figure 3.**

**Secondary structure analysis and 2D structure model for the ΔLR9 NRS, region from positions 691 to 951. (A)** Schematic representation of architecture of the RSV PrC genome genome and of the ALV ΔLR9 NRS RNA used for secondary structure analyses. Indications are the same as in Additional Figure 1, A.

**(B)** Primer extension analyses of enzymatic cleavages in ALV ΔLR9 NRS (B1 to B3). T1, T2, V1, above the lanes indicate T1, T2, V1 RNase digestions, respectively. Conditions for digestions and reverse transcription are given in Materials and Methods. Indications are the same as in Additional Figure 1, B.

**(C)** Schematic representation of the experimental data on the secondary structure model that we established for ALV ΔLR9 NRS. Cleavages by enzymes are depicted as shown in the inset (phosphodiester bonds cleavages by T1, T2 or V1 RNases are indicated by arrows surmounted by lozenges, circles or squares, respectively). Indications are the same as in

Additional Figure1, C. Positions of nucleotides are according to (28) and the red nucleotides in LR9 NRS correspond to base substitutions in ALV compared to RSV.

**Additional Figure 4.**

**Conservation of the NRS RNA secondary structure in ASLVs. (A, B, C, D, E, F, G)**. Secondary structures have been drawn for the NRSs of various ASLVs based on the structure that we established experimentally for the RSV NRS of strain PrC. NCBI Accession numbers and the name of the viral strain are given above each structure. The RNA regions were identified by alignment of the PrC RSV NRS sequence with sequences from other ASLVs using the Boedit software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Base substitutions, insertions and deletions as compared to the PrC RSV NRS are indicated by red, green and blue nucleotides, respectively. The pseudo 5' ss which is conserved in the studied strains is indicated by an arrow. Red and blue rectangles indicate compensatory, semi-compensatory base substitutions and new base paired formation, respectively.

**Additional Figure 5.**

**(A) Comparison of splicing inhibition properties of NRS WT and Δ76, 22, 24, 26 and 28 NRS variants in the C2 pre-mRNA context.**

RT-PCR analysis of *in vitro* splicing products of the C2 and heterologous C2-NRS pre-mRNAs. *In vitro* splicing assays were performed in HeLa cell nuclear extract using conditions described in Materials and Methods. For each pre-mRNA the experiment was repeated three times using different batches of RNA. The relative splicing efficiencies of the various constructs were estimated by semi-quantitative RT-PCR performed in triplicate (see Material and Methods). The amplified cDNA products were fractionated by agarose gel electrophoresis, revealed by ethidium bromide staining and bands corresponding to spliced (Sp) and spliced plus unspliced (Sp+UnSp) products, respectively, were quantified using the GeneTools software. A representative example of cDNA fractionation obtained for each of the tested pre-mRNAs is shown in Figure 3, A1. The Sp/Sp+UnSp pre-mRNA ratio obtained for the C2-NRS Inv construct was taken as the reference (100%). The Sp/Sp+UnSp ratios established for the other constructs are given as a percentage of this reference. The Sp/Sp+UnSp percentages determined for three independent splicing reactions (N°1, 2, 3) and their standard deviations are given.

**(B**) **WT NRS activity is enhanced by 9G8 protein *in vitro***. C2 (depicted in Figure 3, A1), C2-NRS WT and C2-NRS Inv pre-mRNAs (depicted in Figure 3, A2) were incubated in a 1:1

mixture of HeLa cell cytoplasmic and nuclear extracts in the absence or the presence of purified recombinant human 9G8 protein (180 and 270 ng) or BSA (500 ng). After 2 h of incubation, RNAs were subjected to RT-PCR analysis (in triplicate) as in Figure 3. RT-PCR products were fractionated on agarose gel and bands corresponding to the spliced (Sp) or the spliced plus unspliced (Sp+UnSp) products were quantified. All the values for the 3 independent experiments are shown. For each tested pre-mRNA, the Sp/Sp+UnSp ratios obtained in the absence of 9G8 or BSA addition was taken as the 100 % reference and the Sp/Sp+UnSp ratios obtained in the presence of 180 or 270 ng of 9G8 or 500 ng of BSA were expressed as a percentage of this reference. The Sp/Sp+UnSp percentages determined for three independent splicing reactions (N°1, 2, 3) and their standard deviations are given.

(**C**) **Protein 9G8 also increases NRS activity *in cellulo***. The constructs containing the C2, C2-NRS WT and C2-NRS Inv sequences described in Figure 3, A2 were cloned into the pGL3c plamsid and were co-transfected in HeLa cells with a vector expressing the 9G8 SR protein (9G8 +) or an empty vector as the control (9G8 -). After 48 h of growth, extracted RNAs were analysed by RT-PCR using the same primers as in Figure 3, A2 (Additional Table 1) except that the RT primer was replaced by an oligo(dT) primer. Two independent *in cellulo* splicing assays were performed. For each of the tested constructs, the Sp/Sp+UnSp ratio obtained in the absence of 9G8 overexpression was used as the 100 % reference and the Sp/Sp+UnSp ratio obtained in the presence of 9G8 overexpression were expressed as a percentage of this reference. The Sp/Sp+UnSp percentages determined for two independent splicing reactions (N°1, 2) and their standard deviations are given.

(**D**) **9G8 increases the *in vitro* activity of the Δ76 and 28 variant NRSs but not of the 22 variant NRS**. The C2-NRS Δ76, C2-NRS 22 and C2-NRS 28 pre-mRNAs were incubated in a 1:1 mixture of HeLa cell cytoplasmic and nuclear extracts without (9G8 -) or with (9G8 +) 270 ng of purified recombinant human 9G8 protein. RNAs were subjected to RT-PCR and quantified as described in the panel B. The mean value and standard deviation of each of the three independent experiments splicing reactions (N°1, 2, 3) performed for each RNA is shown.

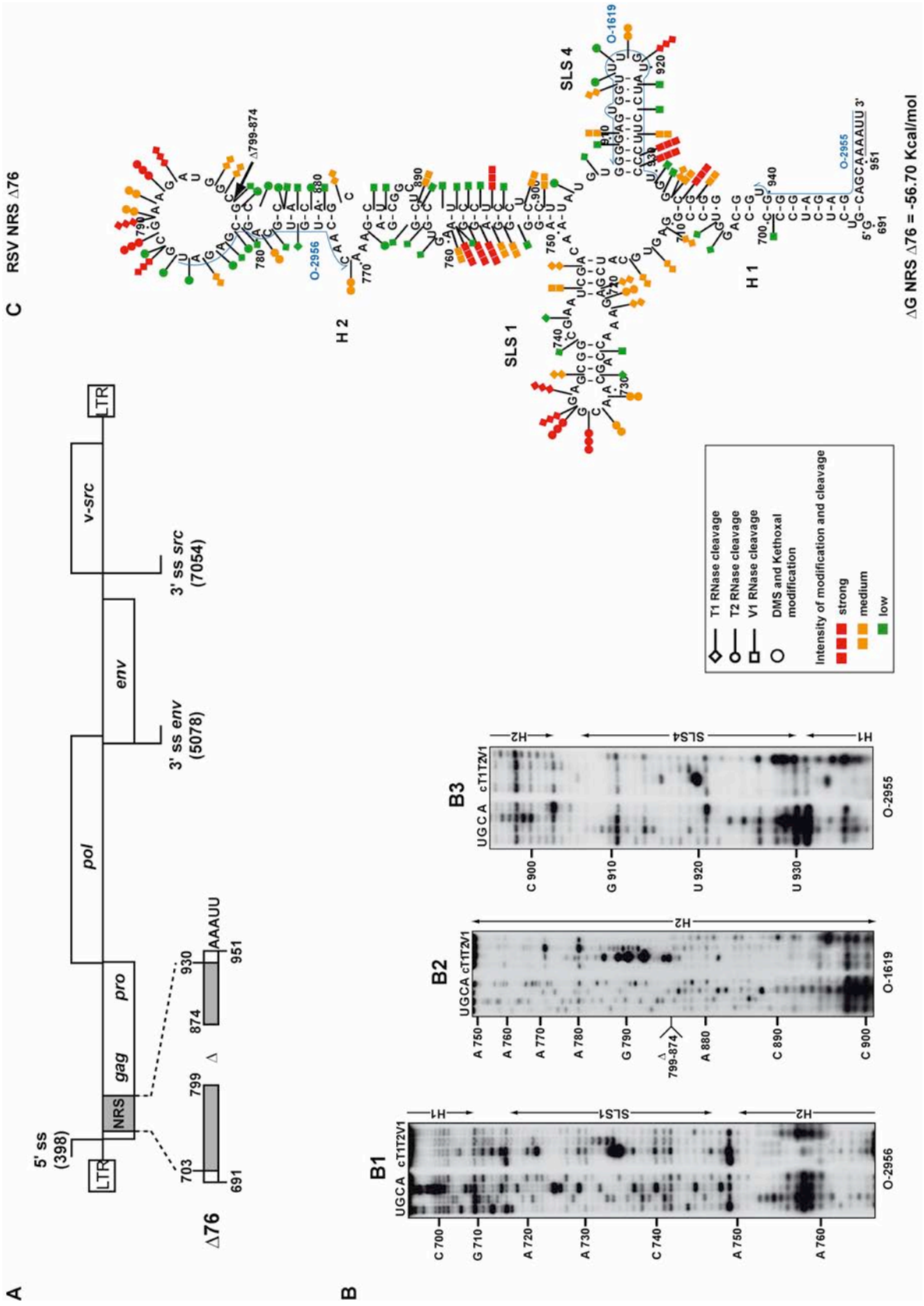
**Additional Table 1**

| Primer Number | Sequence (5'-3') | Restriction site | Targeted sequence | Location on target sequence | Utilization | Generated constructs or mutants |
|---|---|---|---|---|---|---|
| O-2954 | CG**GAATTC**AAATTTAATACGACTCACTATAGGGTCTCTCCCC CAGGTCCGGA | *Eco*RI | RSV, ALV | 691-711 | PCR | NRS WT, NRS Δ76, NRS LR9, NRS ΔLR9 |
| O-2956 | CATCTCGCTGCACAGTTG | | RSV, ALV | 771-788 | Primer extension | NRS Δ76 |
| O-2955 | AATTTTGCTGCTCTCCCA | | RSV, ALV | 939-951 | PCR, Primer extension | NRS WT, NRS Δ76, NRS LR9 |
| O-1619 | AGGGAAGGATACAAACCACTCC | | RSV, ALV | 909-930 | Primer extension | NRS WT, NRS Δ76, NRS LR9, NRS ΔLR9 |
| O-1617 | GGATGTGCCAACGGTTTTAG | | RSV | 838-819 | Primer extension | NRS WT, NRS Δ76 |
| O-1618 | TGCTGCTCTCCCACCCCCG | | RSV | 933-951 | Primer extension | NRS WT |
| O-2858 | GGATGTACCAACGGTTTTAG | | ALV | 818-838 | Primer extension | NRS LR9, NRS ΔLR9 |
| O-3704 | GCCGAGGCTGTGGCGCAA | | RSV, ALV | 871-888 | Primer extension | NRS WT, NRS ΔLR9 |
| O-3826 | CATCTTCGCATCTCGCTG | | RSV | 779-795 | Primer extension | NRS WT |
| O-5422 | GGAGCGGCGAATCGACA / TATGTGGGGAGTGGTTTGTATCCT | | RSV | 733-749/902-925 | Mutagenesis | pNRS 22-MS2, pC2-NRS 22 |
| O-5423 | AAACCACTCCCCACATA / TGTCGATTCGCCGCTCCGTTGCT | | RSV | 727-749/902-918 | Mutagenesis | pNRS 22-MS2, pC2-NRS 22 |
| O-5424 | CCTCCTCCTCCTTATGT / TGGCGGGGGTGGGAGAGCAGCAAA | | RSV | 890-906/930-951 | Mutagenesis | pNRS 24-MS2, pC2-NRS 24 |
| O-5425 | CTCTCCCACCCCCGCCA / ACATAAGGAGGAGGAGGAGCCGAG | | RSV | 883-906/930-946 | Mutagenesis | pNRS 24-MS2, pC2-NRS 24 |
| O-5426 | CCCAGGTCCGGAGTGCA / CAAAGGGGAGGAAGTGGGAGAAAC | | RSV | 700-716/748-771 | Mutagenesis | pNRS 26-MS2, pC2-NRS 26 |
| O-5427 | CCACTTCCTCCCCTTTG / TGCACTCCGGACCTGGGGGAGAGA | | RSV | 693-716/748-764 | Mutagenesis | pNRS 26-MS2, pC2-NRS 26 |
| O-5428 | GGAAGTGGGAGAAACAA / CCTCGGCTCCTCCTCCTCCTTATG | | RSV | 757-773/882-905 | Mutagenesis | pNRS 28-MS2, pC2-NRS 28 |
| O-5429 | AGGAGGAGGAGCCGAGG / TTGTTTCTCCCACTTCCTCCCCTT | | RSV | 750-773/882-898 | Mutagenesis | pNRS 28-MS2, pC2-NRS 28 |
| O-4672 | **GATCC**GGGATATCACTCAGCATAATTTAAA**GAGCT** | *Bam*HI, *Sac*I | | | T7 promoter cloning | pMS2 |
| O-4673 | **C**TTTAAATTATGCTGAGTGATATCCC**G** | *Bam*HI, *Sac*I | | | T7 promoter cloning | pMS2 |
| O-4666 | **GAGCTC**AAATTTAATACGACTCACTATAGGGTCTCTCCCCCA GGTCCGGA | *Sac*I | RSV | 691-711 | Cloning | pNRS WT-MS2, pNRS Δ76-MS2 |

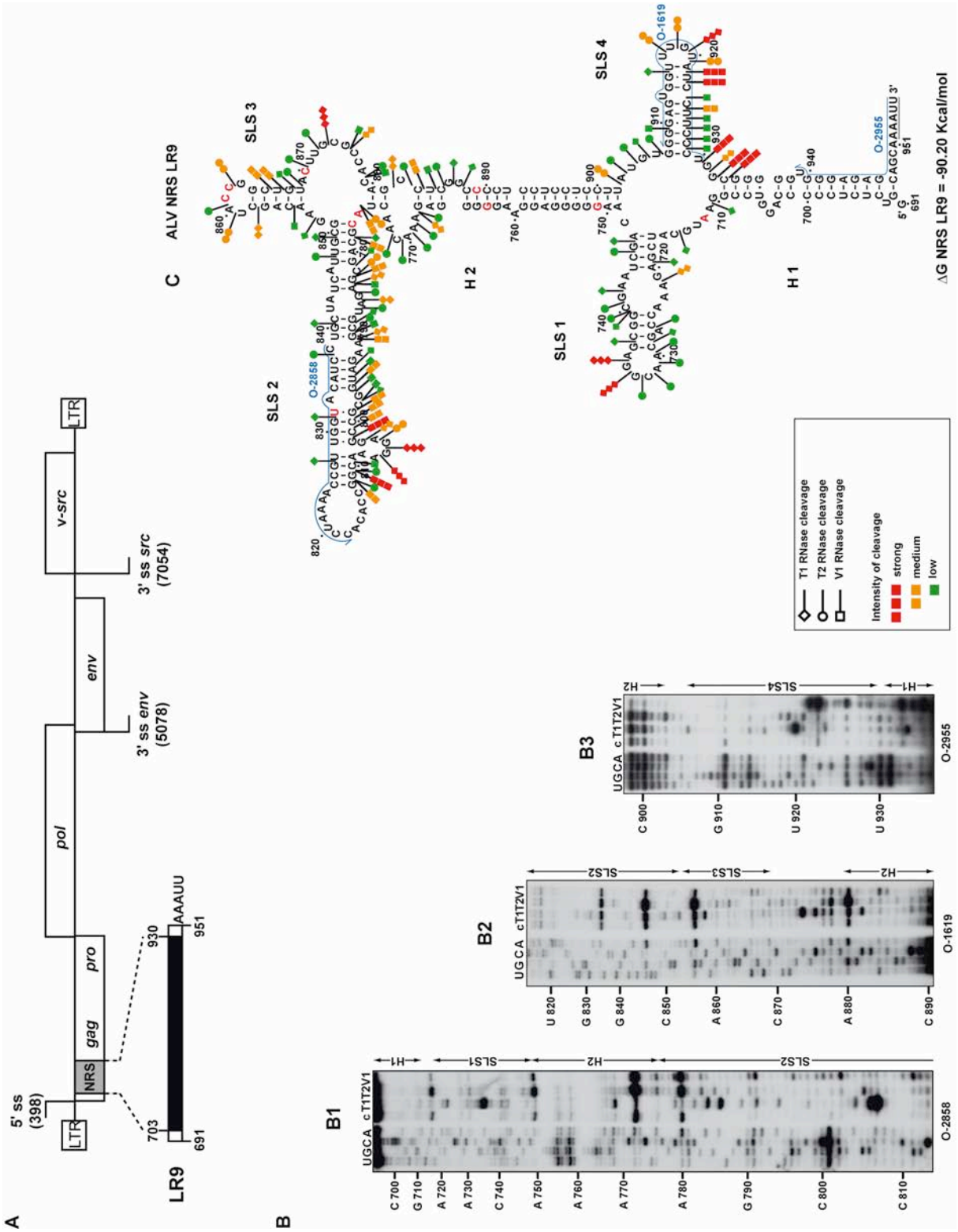
**Additional Table 1**

| Primer Number | Sequence (5'-3') | Restriction site | Targeted sequence | Location on target sequence | Utilization | Generated constructs or mutants |
|---|---|---|---|---|---|---|
| O-4667 | **GGATCC**AATTTTGCTGCTCTCCCA | *Bam*HI | RSV | 939-951 | Cloning | pNRS WT-MS2, pNRS Δ76-MS2 |
| O-5418 | CCC**AAGCTT**GGGAAATTTAATACGACTCACTATAGGGGTCTC TCTGGTTAGACCAG | *Hind*III | HIV-1 | 1-20 | Cloning | pC2 |
| O-5112 | **AGATCT**CCCATCTCTCTCCTTCTAGC | *Bgl*II | HIV-1 | 321-340 | Cloning | pC2 |
| O-5140 | **CTGCAG**GGGAGTCTCCATAGAATGGA | *Pst*I | HIV-1 | 4871-4890 | Cloning | pC2 |
| O-5116 | **GAATTC**GATCCTACCTTGTTATGTCCTGC | *Eco*RI | HIV-1 | 5031-5053 | Cloning | pC2 |
| O-5113 | **AGATCT**GGTCTCTCCCCCAGGTCCGG | *Bgl*II | RSV | 691-710 | Cloning | pC2-NRS WT, pC2-NRS Δ76 |
| O-5114 | **CTGCAG**AATTTTGCTGCTCTCCCACCCCC | *Pst*I | RSV | 934-951 | Cloning | pC2-NRS WT, pC2-NRS Δ76 |
| O-5141 | **CTGCAG**GGTCTCTCCCCCAGGTCCGG | *Pst*I | RSV | 691-710 | Cloning | pC2-NRS Inv |
| O-5142 | **AGATCT**AATTTTGCTGCTCTCCCACCCCC | *Bgl*II | RSV | 934-956 | Cloning | pC2-NRS Inv |
| O-5824 | CATG**CCATGG**CATGGGTCTCTCTGGTTAGACCAG | *Nco*I | HIV-1 | 1-20 | Cloning | pGL-C2, pGL-C2 NRS WT, pGL-C2 NRS Inv, |
| O-5825 | GC**TCTAGA**GCGATCCTACCTTGTTATGTCCTGC | *Xba*I | HIV-1 | 5031-5053 | Cloning | pGL-C2, pGL-C2 NRS WT, pGL-C2 NRS Inv |
| O-6534 | TACGAGATCTACTAGTCGTACC | | MS2 RNA | 60-81 | Northern blot | |
| O-6524 | ACATCCGGAGTGCAATGGAT | | Human snRNA U1 | 61-80 | Northern blot | |
| O-5825 | GC**TCTAGA**GCGATCCTACCTTGTTATGTCCTGC | *Xba*I | HIV-1 | 4031-4053 | RT | |
| O-6246 | TATAGCAGAGT // CAGTCGC | | HIV-1 | 283-289//4472-4482 | RT-PCR | |
| O-6519 | GGTCTCTCTGGTTAGACCAG | | HIV-1 | 1-20 | RT-PCR | |
| O-6187 | CGCTTTCAAGTCCCTGTT | | HIV-1 | 191-208 | RT-PCR | |

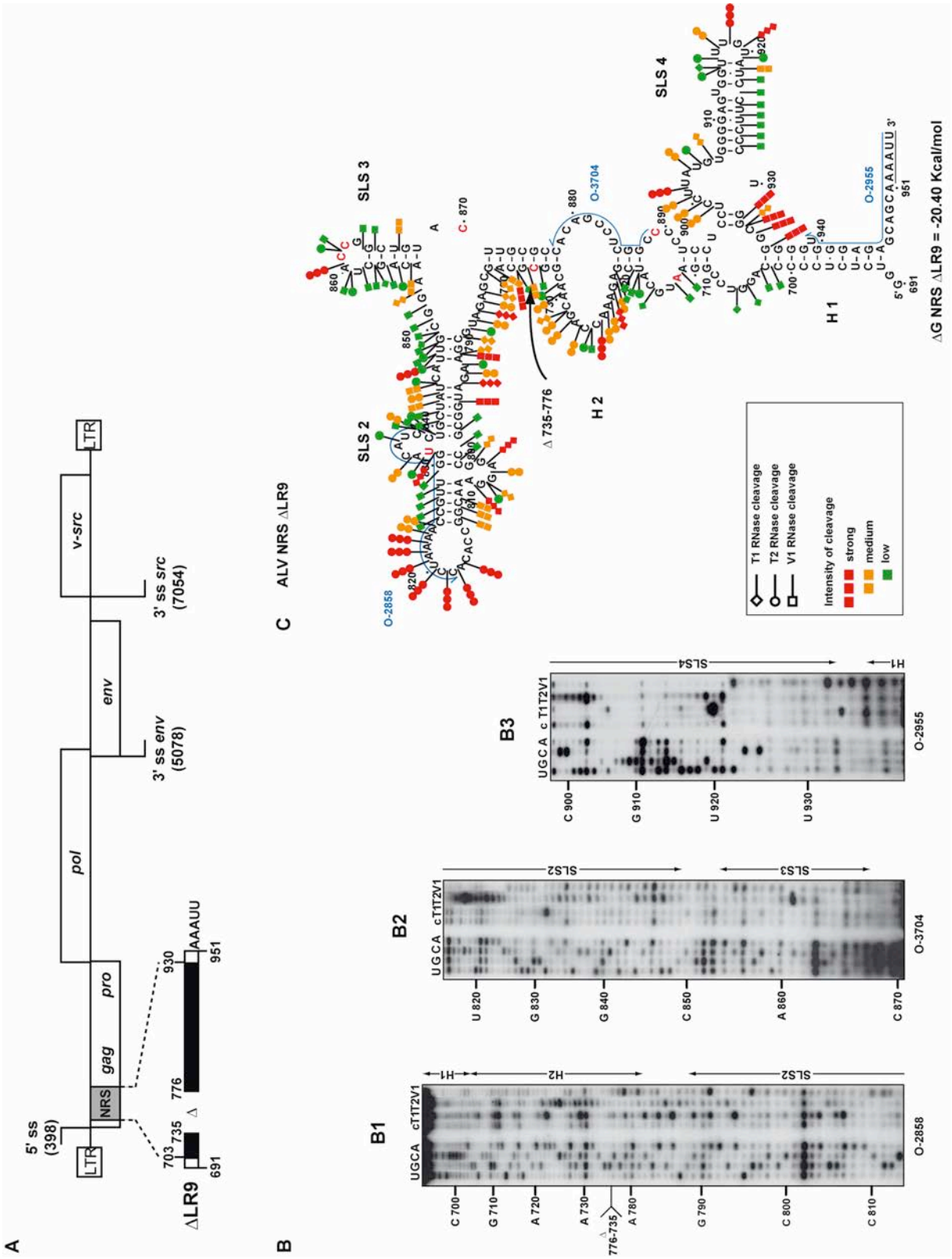
**Additional Figure 1**



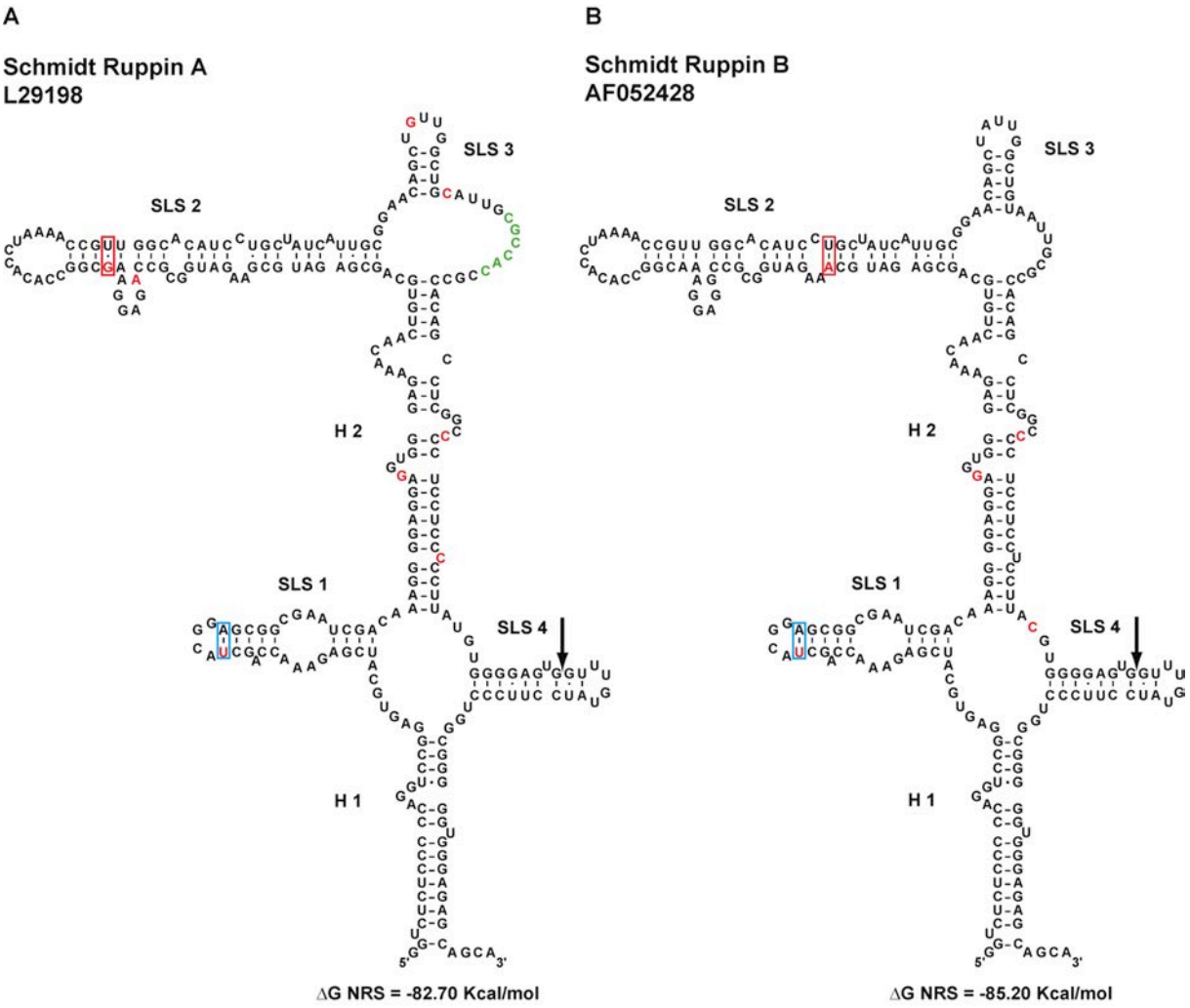
**Additional Figure 2**



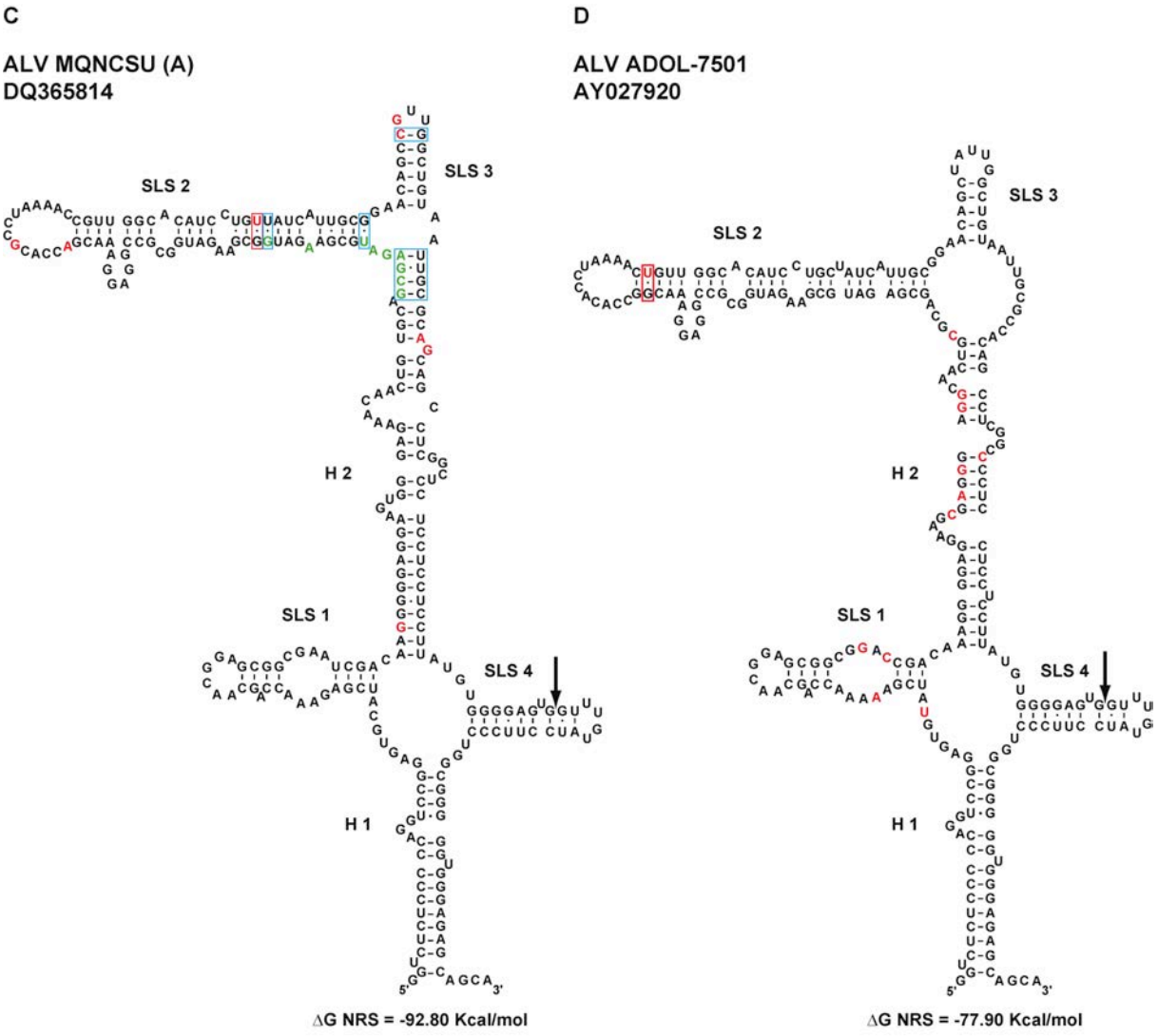
**Additional Figure 3**



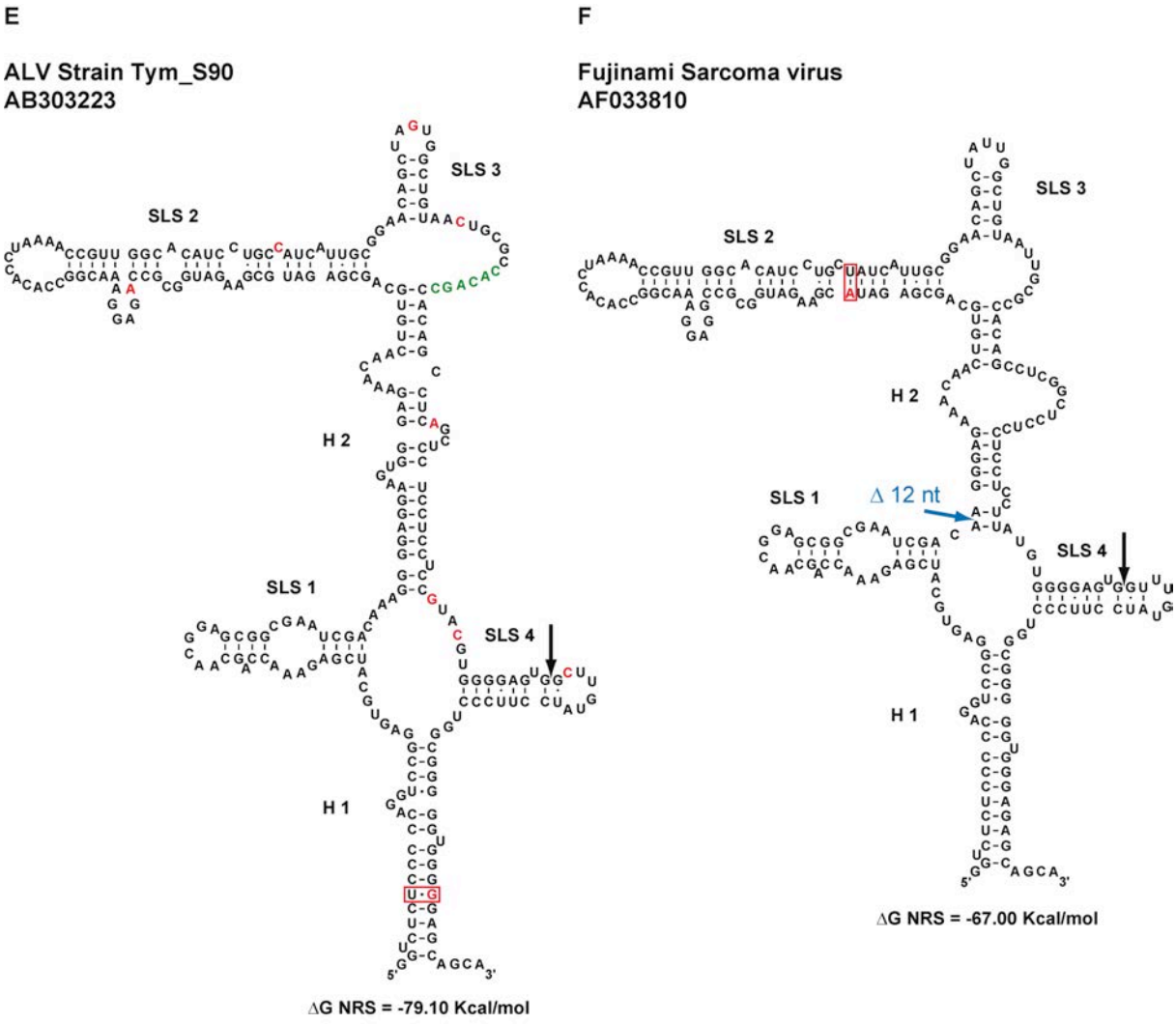
**Additional Figure 4**



**Additional Figure 4**



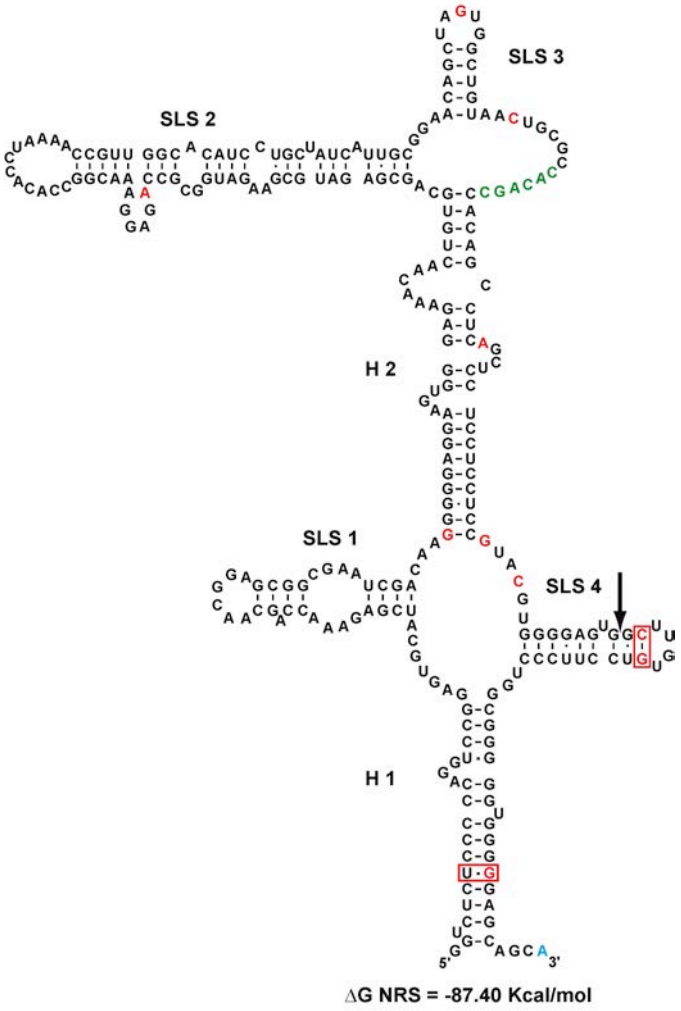
**Additional Figure 4**



**Additional Figure 4**

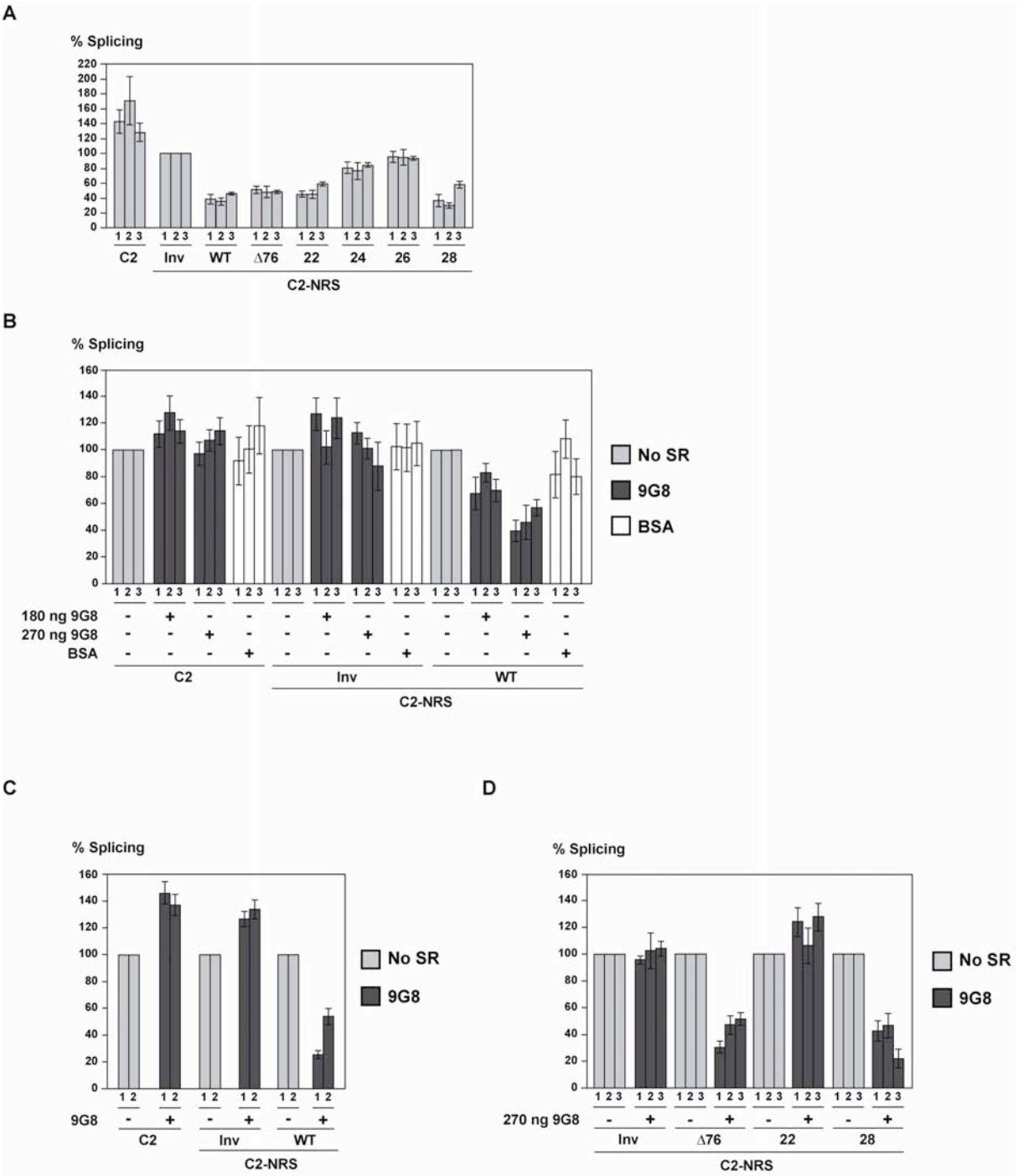
G

Y53 Sarcoma virus
V01170




ΔG NRS = -87.40 Kcal/mol

**Additional Figure 5**



# Discussion et perspectives

## I. Les éléments NRS des différentes espèces de virus adoptent une structure secondaire similaire

Au cours de cette étude, nous avons défini expérimentalement la structure secondaire de l'élément NRS du virus RSV. En solution, cet élément NRS adopte une structure cruciforme qui est très différente de la structure qui avait été prédite par informatique (**Figure XII-1**) (McNally *et al.*, 1991). Dans la structure que nous avons établie, la région ESE sur laquelle se fixe la protéine ASF/SF2 et que nous avons montré être un site de fixation de la protéine 9G8, adopte une structure en tige boucle (SLS1) qui, dans la structure 2D, fait face à la structure tige boucle (SLS 4) comportant le pseudo site (**Figure XII-1, B**). Ces deux structures tiges boucles sont chacune bordées de séquences en simple brin qui forment une grande boucle interne. Sachant que les séquences des tiges boucles SLS1 et SLS4 sont essentielles pour que l'inhibition de l'épissage par l'élément NRS puisse avoir lieu, il est possible que la boucle interne qui les sépare permette une grande flexibilité importante pour leurs fonctions. Une telle flexibilité pourrait faciliter l'interaction de leurs partenaires respectifs, à savoir les protéines SR ASF/SF2, 9G8 (SLS1) et la snRNP U1 (SLS4). Par ailleurs, nos données sont en accord avec les données d'une analyse de la structure 3D formée par une partie de la structure tige boucle SLS4 (Cabello-Villegas *et al.*, 2004).

La structure que nous avons établie diffère aussi de la structure prédite par la présence de l'hélice H1 correspondant à l'appariement des extrémités 3' et 5' de l'élément NRS et de leurs séquences bordantes (**Figure XII-1, B**). Cette hélice H1 pourrait faciliter la structuration de l'élément NRS et participer à sa stabilité. Il avait été montré que la séquence riche en G située sur l'un des brins de cette hélice (nt 935-942) fixe la hnRNP H qui à son tour faciliterait le recrutement de la particule snRNP U11 sur le pseudo site (McNally *et al.*, 2006). On sait que la protéine hnRNP H se lie à des séquences en simple brin ce qui soulève la question d'une ouverture de cette hélice.

Nous avons aussi mis en évidence une longue hélice centrale H2 (**Figure XII-1, B**) dont l'un des brins (nt 750-778) a été décrit comme fixant les protéines SR SRp40, SC35, 9G8 et la protéine hnRNP H (Fogel *et al.*, 2002 ; Fogel et McNally, 2000). Les chercheurs qui ont poursuivi les travaux au laboratoire ont en effet observé la fixation de la protéine 9G8 sur cette hélice par des expériences d'empreinte. J'ai personnellement montré que la protéine hnRNP A1 se fixe à cette hélice H2, ce qui suggère que les deux protéines entrent en compétition pour la fixation sur l'hélice H2.

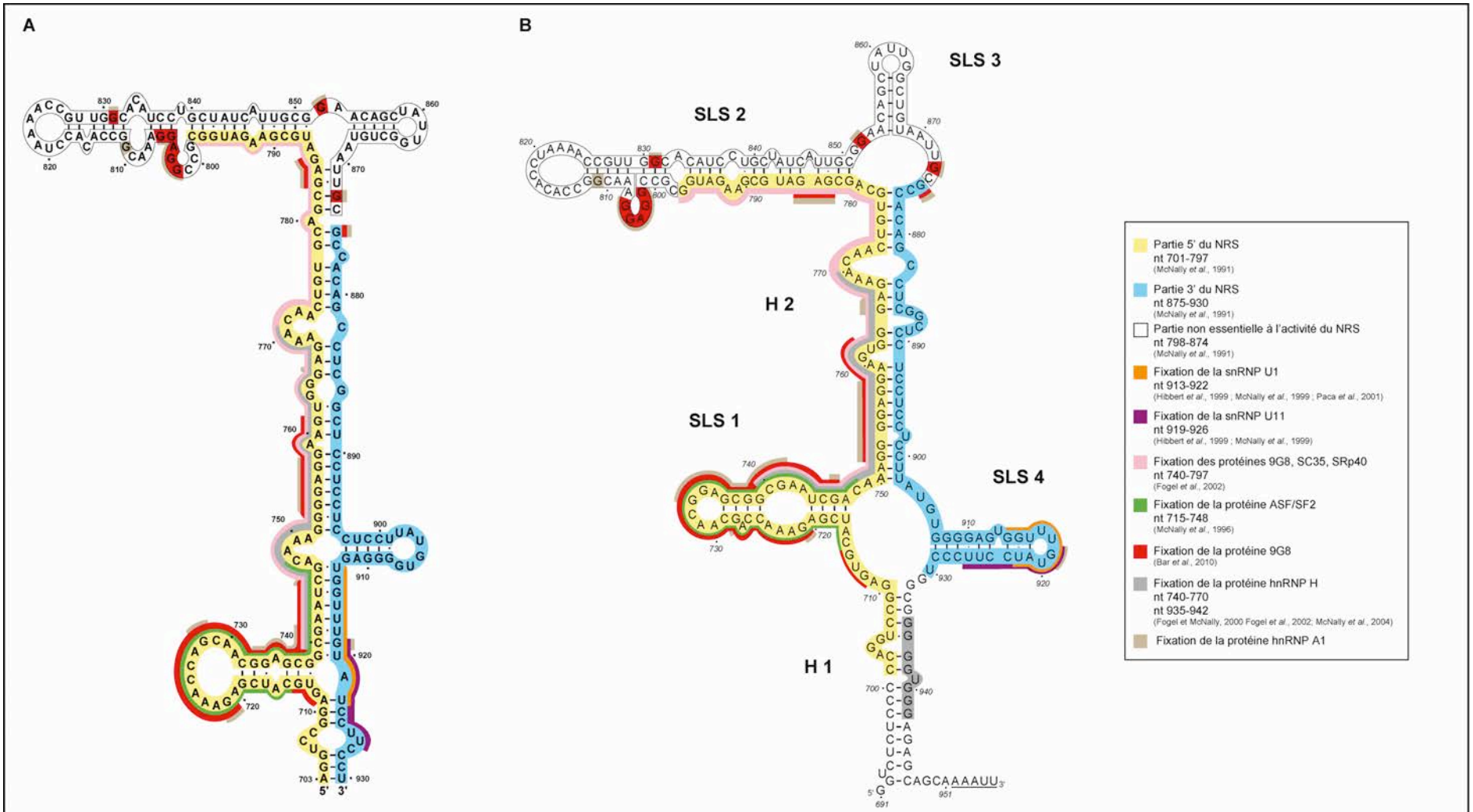



**Figure XII-1 : Comparaison des structures secondaires théorique et expériementale de l'élément NRS du virus RSV et report sur ces structures des sites de fixation des protéines connues.**
**A :** Structure secondaire théorique de l'élément NRS du virus RSV prédite par le logiciel Mfold (McNally *et al.*, 1991).
**B :** Structure secondaire de l'élément NRS du virsu RSV établie expérimentalement (Bar *et al.*, 2011).
Les sites de fixation des protéines connues pour interagir avec l'élément NRS sont reportées sur chacune des deux structures selon la légende précisée dans l'encart.

Le restant de la structure de l'élément NRS comporte les structures tiges boucles SLS2 et SLS3 dont les troncations totales ou partielles n'ont pas de répercussion sur l'activité de l'élément NRS (**Figure XII-1, B**).

Les structures 2D ont également été définies expérimentalement pour l'élément tronqué NRS Δ76 du virus RSV et l'élément NRS du virus ALV LR9. Ces deux éléments NRS sont fonctionnels en inhibition de l'épissage et adoptent des structures similaires à celle définie pour l'élément NRS que nous venons de décrire. La troncation des nucléotides 798-874 de l'élément NRS Δ76 ne provoque qu'un remaniement de structure local et les 10 substitutions de bases de l'élément NRS LR9 renforcent les éléments structuraux définis pour l'élément NRS WT du RSV. En revanche, la structure définie pour l'élément NRS ΔLR9 est très différente en raison de la troncation des nucléotides 735-776 qui remodèle entièrement la structure à l'exception de la tige boucle qui contient le pseudo site (SLS4). L'élément NRS ΔLR9, contrairement aux autres éléments NRS étudiés, n'est pas fonctionnel ce qui est un argument en faveur de l'existence d'une corrélation entre la structure secondaire de l'élément NRS et sa fonction.

## II. La structure secondaire de l'élément NRS est conservée chez les virus ASLV

Nous avons réalisé une étude phylogénétique de différents virus de la famille ASLV afin d'étudier la conservation de la structure secondaire de l'élément NRS et d'apporter des arguments en faveur d'un rôle fonctionnel de la structure secondaire de l'élément NRS. La séquence correspondant à l'élément NRS est présente chez tous les virus ASLV car elle se situe dans l'ORF *gag* du pré-ARNm et code pour le précurseur Pr76 Gag-Pro qui comprend chez la plupart des virus les protéines de la matrice (MA), et les peptides p2a, p2b et p10 (Schwartz *et al.*, 1983). Cela laisse supposer que la séquence correspondant à l'élément NRS puisse avoir les mêmes fonctions chez ces autres virus que celles qui ont été caractérisées chez les virus RSV PrC et LR9, à savoir inhiber l'épissage et activer la polyadénylation du pré-ARNm viral (Arrigo et Beemon, 1988 ; O'Sullivan *et al.*, 2002 ; Smith *et al.*, 1997).

Pour que cette étude soit biologiquement pertinente, nous avons étudié les séquences de virus qui ont été totalement séquencés et identifiés chez des espèces aviaires infectées. Ces virus appartiennent à des sous-groupes viraux différents. Certains sont réplicativement autonomes, d'autres non. Les structures secondaires des différentes séquences de ces virus correspondant à l'élément NRS ont été définies théoriquement à l'aide du logiciel Mfold

(Zuker, 2003). Les résultats obtenus montrent que ces séquences sont capables d'adopter une structure 2D similaire à celle du modèle structural de l'élément NRS que nous avions déterminé expérimentalement pour le virus RSV de la souche PrC. Les substitutions de bases ou les duplications de séquences des virus étudiés par rapport à la séquence du virus RSV, renforcent les structures secondaires lorsqu'elles sont dans des régions en double brin, ou bien se situent dans des régions simples brins comme des boucles et ne provoquent pas de remaniements structuraux. Ces résultats rejoignaient ceux obtenus expérimentalement pour la structure secondaire de l'élément NRS de la souche LR9 qui comporte 10 substitutions de bases par rapport à celui du virus RSV PrC. Cette étude a également mis en évidence l'existence d'un degré de conservation entre la séquence, sa structure secondaire et son importance dans la fonction de l'élément NRS. Les séquences fonctionnellement essentielles, à savoir celles des tiges boucles SLS1 et SLS4, de l'hélice H1 et du début de l'hélice H2, sont fortement conservées contrairement aux séquences des tiges boucles SLS2 et SLS3 qui, elles, ne sont pas indispensables à la fonction de l'élément NRS.

### II.1. La conservation de la structure secondaire de l'élément NRS chez les virus ASLV ne repose pas uniquement sur la capacité codante de la séquence

Chez l'ensemble des virus étudiés, le nombre de substitutions de bases ou de duplications de séquences, même dans les séquences les plus divergentes par rapport à celle du virus RSV PrC, est assez restreint. Le fait que les virus ASLV soient phylogénétiquement proches et que la séquence correspondant à l'élément NRS soit codante sont autant de biais qui pourraient expliquer nos résultats. Néanmoins, ces biais ne semblent pas suffisants pour expliquer la conservation de la structure secondaire correspondant à la séquence de l'élément NRS.

En effet, la pression de sélection exercée habituellement par la capacité codante d'une séquence ne semble pas pouvoir expliquer la conservation de la plupart de la séquence correspondant à l'élément NRS. D'après notre étude, il n'y a pas plus de divergences entre les séquences et les structures secondaires des virus réplicativement autonomes (ex : ALV strain Tym_S90) et celles des virus non autonomes (ex : Y53 Sarcoma virus) qui, en raison de troncation de séquence dans leur génome viral, ne produisent pas les protéines du précurseur Pr76 Gag-Pro qui sont apportées en *trans* par un virus helper (**Figure XII-2**). De plus, chez les virus réplicativement autonomes, la séquence de l'ORF *gag* comprenant la séquence de l'élément NRS tolère de nombreuses mutations et troncations de séquence qui peuvent inactiver l'élément NRS sans affecter le bourgeonnement et le caractère infectieux des virus (Gontarek *et al.*, 1993 ; Wills *et al.*, 1994). Cela est par exemple illustré par le virus ALV ΔLR9, où la troncation des acides aminés 114-132 de la protéine de la matrice (MA) n'a pas

de répercussion sur la synthèse protéique ou sur la réplication virale (Neiman *et al.*, 2003 ; Nelle et Wills, 1996 ; Polony *et al.*, 2003 ; Smith *et al.*, 1997) (**Figure XII-2**).

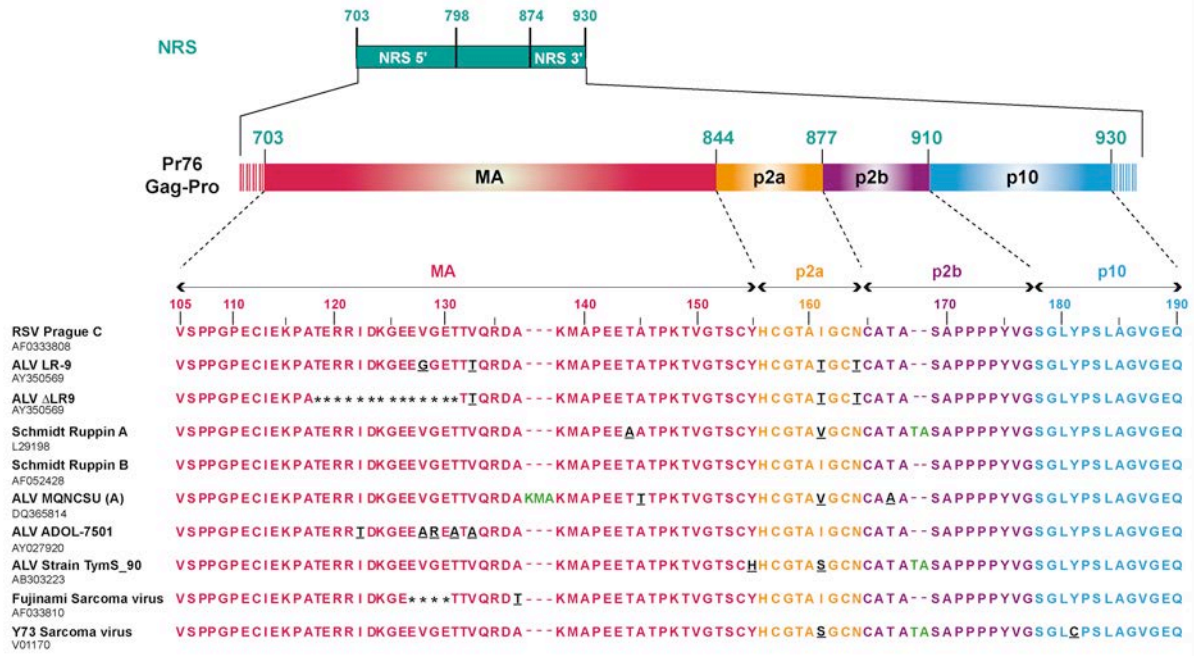



**Figure XII-2 : Alignement des séquences protéiques du précurseur Pr76 Gag-Pro codées par la séquence de l'élément NRS des différents virus choisis pour l'étude phylogénétique.**
L'élément NRS fait partie de l'ORF de *gag* et code pour les protéines de la matrice (MA), p2a, p2b, et p10 du précurseur Pr76 Gag-Pro. La séquence peptidique de la protéine de la matrice est colorée en rose, celle du peptide p2a en orange, celle du peptide p2b en violet et celle du peptide p10 en bleu. Les séquences d'acides aminés des différents virus sont alignés par rapport à celle de la souche RSV Prague C : les acides aminés différents sont soulignés et colorés en noir, les acides aminés supplémentaires sont colorés en vert, les acides aminés tronqués sont mentionnés par des astérisques. Le nom des virus et leur numéro d'accession à la GenBank sont mentionnés.

Les seules parties de l'élément NRS dont la séquence pourrait être soumise à une pression de conservation due à la synthèse protéique sont comprises entre les nucléotides 893-904 et 917-931. Ces séquences codent respectivement pour les acides aminés PPPY et LYPSL qui sont des domaines L très conservés (Dilley *et al.*, 2010) (**Figure XII-3**).

Ces domaines L sont essentiels lors du bourgeonnement du virus où le motif PPPY joue un rôle plus important que celui du motif LYPSL (Dilley *et al.*, 2010). D'après notre modèle de structure secondaire, la séquence codant le motif PPPY se situe sur l'un des brins de l'hélice H2 tandis que la séquence codant le motif LYSPL correspond à une partie de la tige boucle SLS4 et chevauche la séquence du pseudo site (**Figure XII-3**). Notre étude phylogénétique montre que la séquence nucléotidique correspondant au motif PPPY est moins conservée que celle du motif LYPSL, bien que la séquence en acides aminés soit identique dans tous les cas et que cela ne provoque pas de remaniements structuraux de l'hélice H2. Cela suggère que la séquence codant le motif LYPSL, malgré son rôle secondaire lors du bourgeonnement, est

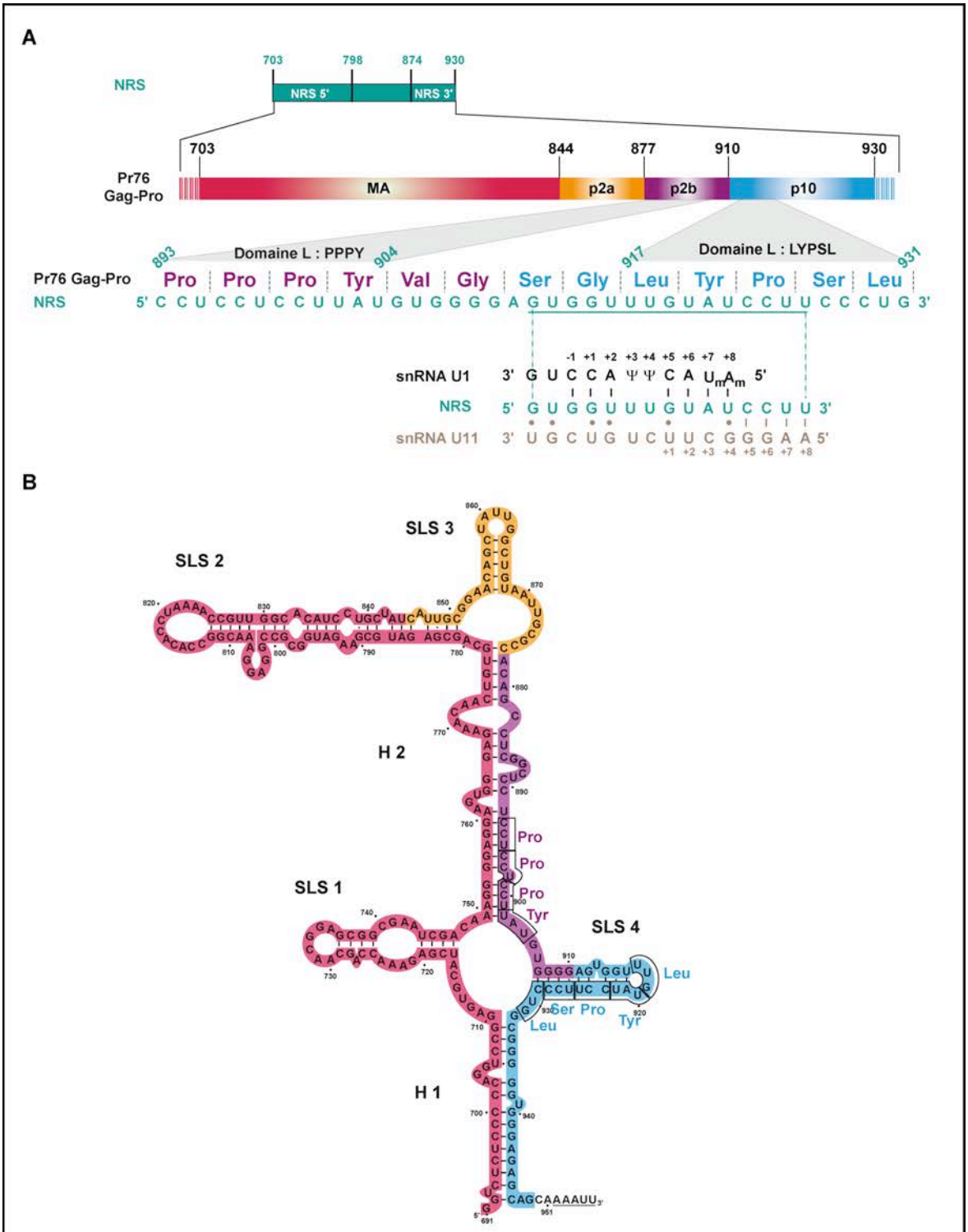



**Figure XII-3 : Les domaines L du précurseur Pr76 Gag-Pro sont codés par l'élément NRS (Dilley *et al.*, 2010).**

**A :** Représentation des protéines de la matrice (MA), p2a, p2b et p10 du précurseur Pr76 Gag-Pro codés par l'élément NRS ainsi que de ses domaines L nécessaires au bourgeonnement viral. La séquence de l'élément NRS est colorée en vert et les résidus des domaines L, PPPY et LYPSL, sont respectivement colorés en violet et en bleu. La séquence du pseudo site de l'élément NRS qui code en partie le domaine L LYPSL est souligné et ses appariements avec les snRNA U1 et U11 sont représentés.

**B :** Report sur la structure secondaire de l'élément NRS des séquences codant les différentes protéines du précurseur Pr76 Gag-Pro. Les séquences codant les protéines de la matrice, p2a, p2b et p10 sont respectivement colorées en rose, orange, violet et bleu. Les codons codant les domaines L PPPY et LYPSL sont encadrés en noir et le nom du résidu correspondant est indiqué.

davantage soumis à une forte pression de conservation. Cela pourrait s'expliquer par le fait qu'elle renferme une partie du pseudo site de l'élément NRS, dont la séquence et la structure sont capitales pour sa fonction inhibitrice de l'épissage (Cabello-Villegas *et al.*, 2004).

### II.2. Les éléments NRS des virus ALV Strain Tym_S90 et Y53 Sarcoma virus ne seraient pas fonctionnels

Parmi les virus que nous avons étudiés, seuls deux d'entre eux, ALV Strain Tym_S90 et Y53 Sarcoma virus possèdent des substitutions de bases dans la séquence du pseudo site par rapport à celle du virus RSV PrC. Le nucléotide, U916 (à la position +2 par rapport à l'appariement du pseudo site avec le snRNA U1), est substitué par un C dans la séquence du virus ALV Strain Tym_S90 ainsi que dans celle du virus Y53 Sarcoma virus qui possède en plus la substitution A921G (à la position +7) (**Figure XII-4**). La mutation U916C à la position +2 provoque la perte d'un appariement avec le snRNA U1 au profit d'un appariement avec le snRNA U11. Dans l'élément NRS du virus RSV PrC, cette mutation inactive l'élément NRS, ce qui suggère que les éléments NRS des virus ALV Strain Tym_S90 et Y53 Sarcoma virus ne sont pas fonctionnels (Hibbert *et al.*, 1999) (O'Sullivan *et al.*, 2002). La répercussion de la mutation A921G sur l'activité de l'élément NRS du virus RSV PrC n'a jamais été testée. Cependant, d'après les structures secondaires théoriques, cette substitution de base devrait former un appariement Wooble G.U à la place d'un appariement Watson Crick A-U entre le pseudo site et la snRNA U1 (**Figure XII-4, C**). Par ailleurs, d'après la structure secondaire théorique de l'élément NRS du virus Y53 Sarcoma virus, la substitution de base A921G devrait s'apparier avec la substitution U916C pour former un appariement G-C à la place de l'appariement A-U retrouvé dans l'hélice de la tige boucle SLS4 de l'élément NRS du virus PrC (**Figure XII-4, B**). Cet appariement, en plus de la substitution de base, pourrait inactiver l'appariement du snRNA U1 avec le pseudo site en renforçant la stabilité de l'hélice de la tige boucle SLS4 (ΔG SLS4 Y53= -11,40 Kcal/mol contre ΔG SLS4 RSV PrC= -9,10 Kcal/mol). Ce cas de figure a déjà été démontré avec le mutant ΔU913 de l'élément NRS RSV PrC, où l'absence de bulge dans l'hélice de la tige boucle SLS4 renforce la stabilité de la structure (ΔG ΔU913 = -12,90 Kcal/mol contre ΔG SLS4 RSV PrC= -9,10 Kcal/mol) et empêche la fixation du snRNA U1 (Cabello-Villegas *et al.*, 2004). Ces données sont concordantes avec le fait que la fixation du snRNA U1 sur le pseudo site de l'élément NRS dépend à la fois de sa séquence et de sa structure secondaire (Cabello-Villegas *et al.*, 2004).

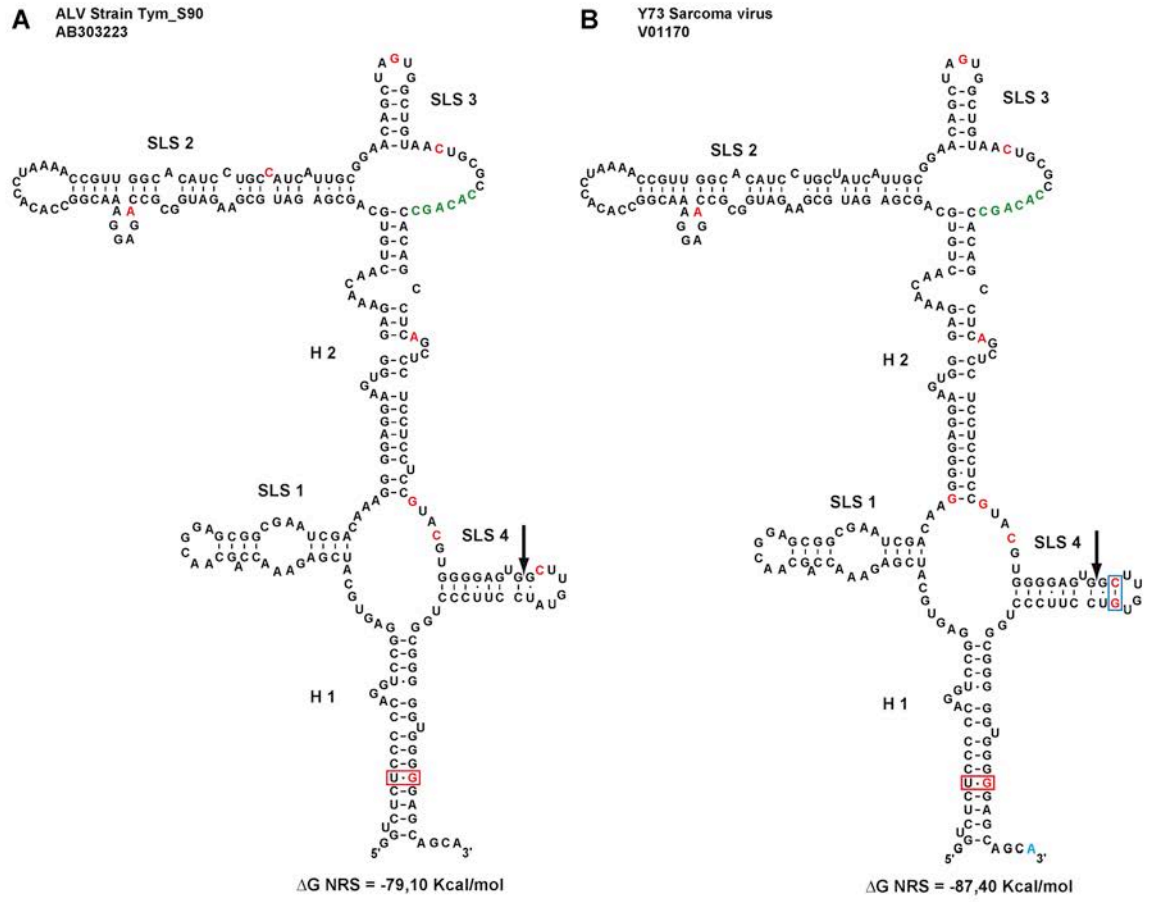



**Figure XII-4 : Représentation des structures secondaires théoriques des éléments NRS des virus ALV Strain Tym_S90 RSV (A) et Y73 Sarcoma virus (B) ainsi que de leurs séquences du pseudo site appariées avec le snRNA U1 (C) (Bar *et al.*, 2011).**

Les structures secondaires des éléments NRS des virus ALV Strain Tym_S90 RSV (A) et Y73 Sarcoma virus (B) ont été établies avec le logiciel Mfold d'après le modèle de structure secondaire obtenu expérimentalement pour l'élément NRS du virus RSV Prague C. Le nom de chaque virus, son numéro d'accession à la Genbank et l'énergie libre de la structure 2D proposée sont respectivement indiqués au-dessus et en-dessous de chaque structure. Les substitutions de bases, les insertions et les délétions sont représentées par des nucléotides colorés en rouge, vert et bleu, respectivement. Les susbstitutions de bases semi-compensatoires et compensatoires sont encadrées par des rectangles rouges et bleus, respectivement. Le pseudo site est indiqué par une flèche.

**C :** Comparaison des appariements formés entre la séquence du pseudo site de l'élément NRS et le snRNA U1 chez les virus RSV PrC, ALV Strain Tym_S90 et Y73 Sarcoma virus. Les substitutions de bases par rapport à la séquence du virus RSV PrC sont colorées en rouge.

### II.3. Lien entre l'élément NRS et l'oncogénicité des virus

Par ailleurs, si les éléments NRS des virus Y53 Sarcoma Virus et ALV Strain Tym_S90 sont non fonctionnels comme tend à le montrer notre étude, cela peut en partie contribuer aux pathologies qu'ils provoquent. Le virus ALV Strain Tym_S90 qui est réplicativement autonome et qui ne contient pas d'oncogène provoque chez les volailles du Japon qu'il infecte la formation d'un gliome et la prolifération de cellules neuronales immatures (Hatai *et al.*, 2008) (**Figure XII-5, A**). Ces dernières ne sont jamais observées chez cette espèce chez laquelle seuls des gliomes sont induits à la suite d'une infection par le virus FGV (Fowl Glioma Virus) (Hatai *et al.*, 2008). L'histologie de ces lésions et le caractère oncogénique du virus ALV Strain Tym_S90 ne sont pas connus. Cependant, il a été suggéré que la surexpression des proto-oncogènes c-Myc et K-Ras dans les cellules du gliome soit due à l'insertion provirale et l'activité promotrice des LTR du virus (Hatai *et al.*, 2008). L'inactivation de l'élément NRS chez le virus ALV Strain Tym_S90 pourrait contribuer aux surexpressions de ces proto-oncogènes.

Ce lien a déjà été démontré dans le cas du virus RSV et LR9. L'élément NRS régule l'oncogénicité du virus RSV en inhibant l'épissage au site 3'*src*, et celle du virus LR9 en activant l'utilisation du site de polyadénylation viral de façon à inhiber la transcription des proto-oncogènes situés en aval du provirus (Jiang *et al.*, 1997 ; O'Sullivan *et al.*, 2002 ; Polony *et al.*, 2003 ; Smith *et al.*, 1997). L'inactivation de l'élément NRS provoque chez le virus RSV une augmentation de la production de la protéine v-Src et chez le virus LR9 une diminution de l'utilisation du signal de polyadénylation viral, qui provoque la transcription du gène cellulaire en aval, le plus souvent le proto-oncogène c-*myb*. En revanche, pour les deux virus l'inactivation de l'élément NRS n'a que peu de répercussion sur la réplication virale car il ne change de façon importante le taux d'épissage au site 3'*env* nécessaire à la production de l'ARNm qui code les protéines de l'enveloppe (O'Sullivan *et al.*, 2002 ; Polony *et al.*, 2003). Le virus ALV Strain Tym_S90 pourrait s'apparenter à celui du virus LR9 où l'oncogénicité virale induite par l'inactivation de l'élément NRS serait due à la diminution de l'utilisation du site de polyadénylation viral et à la transcription des proto-oncogènes situés en aval du provirus.

Le virus Y73 Sarcoma virus possède l'oncogène v-*yes* et provoque la formation de sarcome. Comme la majorité des virus possédant un oncogène, il n'est pas réplicativement autonome et son génome est tronqué des gènes *pol* et *env* tandis que le gène *gag* est partiellement tronqué (Kitamura *et al.*, 1982) (**Figure XII-5, B**). Le pré-ARNm de ce virus n'est pas épissé et sa

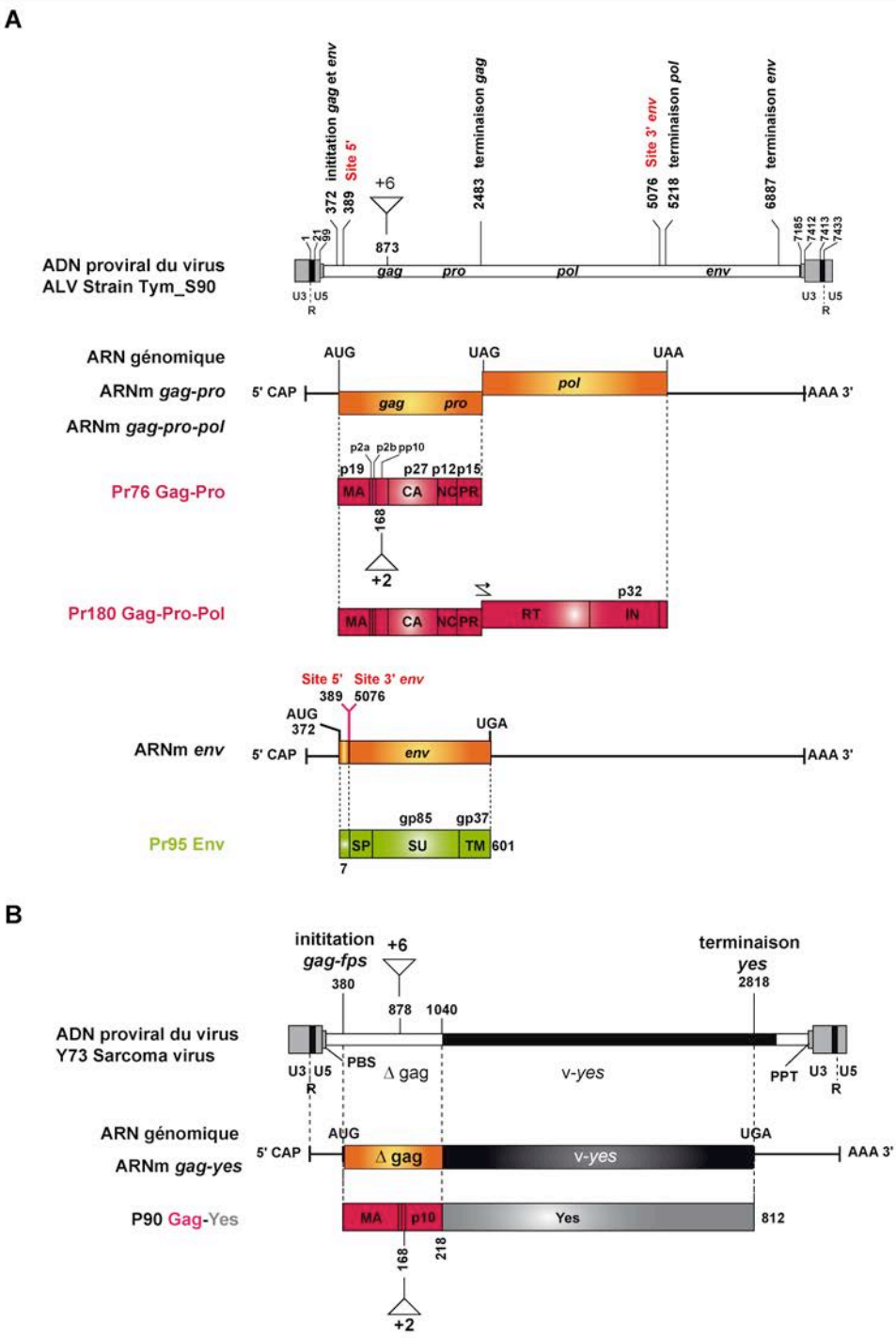



**Figure XII-5 : Génomes et traduction des protéines à partir des différents ARN viraux des virus ALV Strain Tym_S90 (A) et Y73 Sarcoma virus (B) (d'après Hatai *et al.* 2008 ; Kitamura *et al.*, 1982).**
Les protéines synthétisées sont représentées en-dessous de l'ARN ayant servi de matrice. Les séquences non traduites sont représentées par un trait noir, les phases de lecture sont matérialisées par le décalage vertical des ORF. Les sites d'épissage sont représentés par des traits roses et leur nom est indiqué en rouge. Les précurseurs protéiques traduits à partir de l'ORF *gag-pro*, *gag-pro-pol* sont colorés en rose, ceux de l'ORF *env* en vert et la protéine Yes est colorée en gris.
**A :** Traduction des ARNm *gag-pro*, *gag-pro-pol* et *env* du virus ALV Strain Tym_S90.
**B :** Traduction de l'ARNm *gag-yes* du virus Y73 Sarcoma virus.

traduction produit une protéine de fusion p90 Gag-Yes qui ne subit pas de maturation protéolytique (Coffin *et al.*, 1997). En l'absence d'épissage, l'élément NRS pourrait ne pas

être indispensable au cycle viral et son inactivation pourrait être sans conséquence. Dans ce contexte, le pseudo site de l'élément NRS ne serait pas soumis à la pression de conservation de sa séquence, ce qui expliquerait la présence des deux substitutions de bases dans la séquence du pseudo site de l'élément NRS qui théoriquement devraient le rendre non fonctionnel.

Le virus Fujinami Sarcoma virus est aussi un virus qui possède un oncogène, v-*fps*, et dont le génome est tronqué des gènes *pol* et *env*, ce qui le rend réplicativement non autonome (Kitamura *et al.*, 1982). Le pré-ARNm de ce virus n'est pas épissé et sa traduction produit une protéine de fusion p140 Gag-Fps qui ne subit pas de maturation protéolytique (Coffin *et al.*, 1997) (**Figure XII-6**). La séquence et la structure secondaire du pseudo site de l'élément NRS du virus FuSV sont identiques à celles du virus RSV PrC, ce qui suggère que la snRNP U1 peut s'y fixer (**Figure XII-6**). Cependant, l'élément NRS est tronqué de la séquence comprise entre les nucléotides 751-763 dans l'hélice H2 que nous avons identifié comme étant un site de fixation de la protéine 9G8 chez le virus RSV PrC (**Figure XII-6**). Sachant que cette séquence est nécessaire pour que cette protéine active l'inhibition de l'épissage *in vitro* par l'élément NRS (Bar *et al.*, 2011), il est possible que sa troncation puisse inactiver l'élément NRS. D'après notre modèle de structure secondaire théorique, cette troncation ne provoque pas de remaniements structuraux, à part localement dans l'hélice H2 (**Figure XII-6**). En l'absence de données structurales et fonctionnelles, nous ne pouvons pas conclure quant à la possible interdépendance de la structure secondaire et de la fonction de cet élément NRS.

Dans leur ensemble, ces résultats vont dans le sens d'un rôle de la structure secondaire dans la fonction de l'élément NRS. Pour le confirmer et afin d'étayer cette analyse phylogénétique, il serait nécessaire de définir expérimentalement les structures secondaires des séquences correspondant à l'élément NRS des différents virus. Parmi les éléments NRS étudiés, ceux des virus FuSV, ALV Tym_S90 et Y73 Sarcoma Virus sont les candidats les plus intéressants en raison de leur divergence de séquence avec celle du virus RSV PrC et de la répercussion potentielle de ces changements au niveau pathologique. Dans un deuxième temps, la capacité de ces séquences à inhiber l'épissage serait étudiée en utilisant la construction hétérologue C2-NRS et l'approche RT-PCR que nous avons développée lors notre étude.

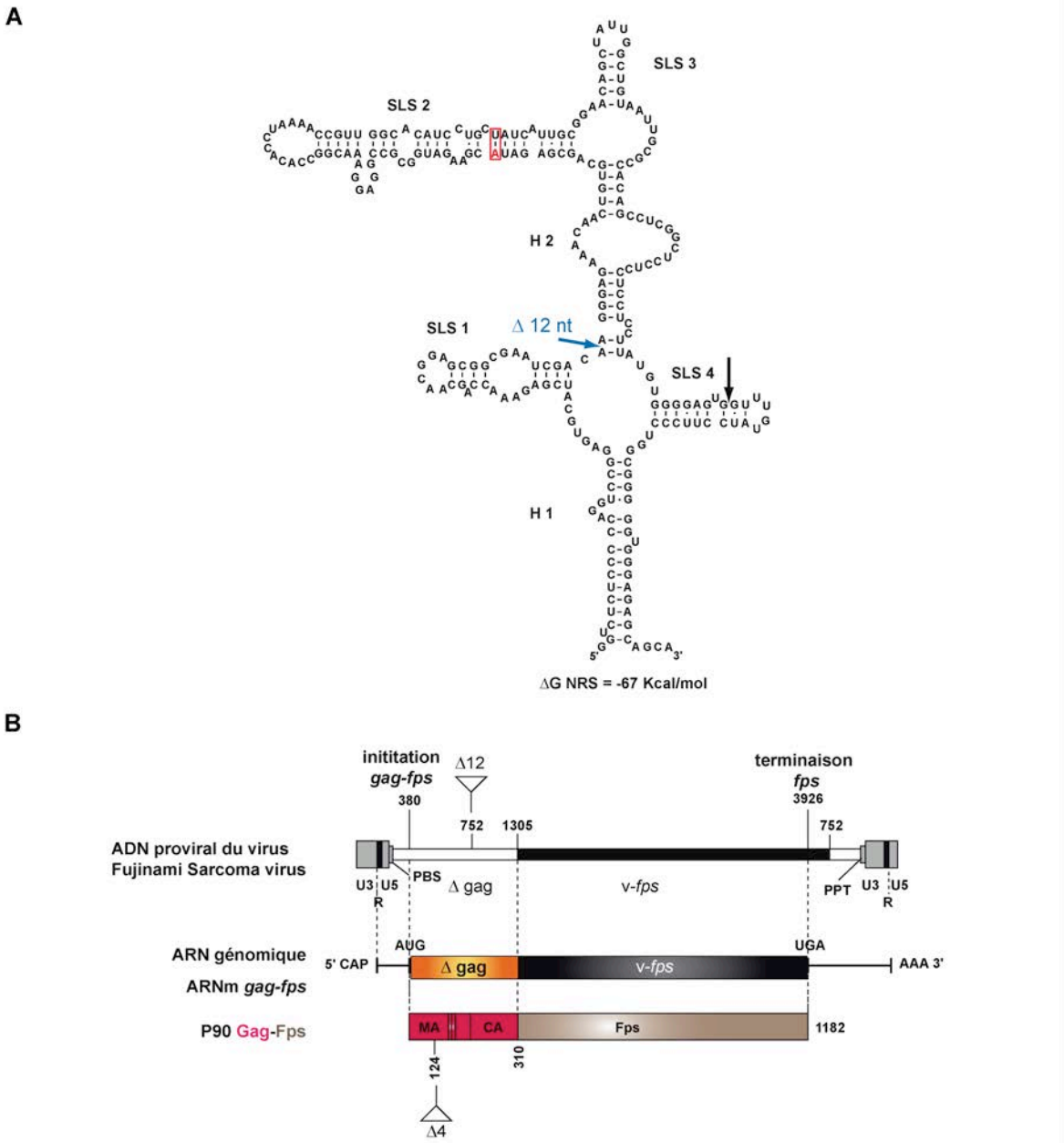



**Figure XII-6 : Représentation de la structure secondaire théorique de l'élément NRS du virus Fujinami Sarcoma virus, de son génome et de la protéine de fusion P90 Gag-fps (d'après Coffin *et al.* 1997).**
**A :** La structure secondaire de l'élément NRS du virus Fujinami Sarcoma virus a été établie avec le logiciel Mfold d'après le modèle de structure secondaire obtenu expérimentalement pour l'élément NRS du virus Prague C. La substitution de base semi-compensatoire est colorée en rouge et encadrée par un rectangle rouge. La troncation de 12 nucléotides est indiquée par une flèche bleue, le pseudo site par une flèche noire.
**B :** Représentation du génome du virus Fujinami Sarcoma virus et de la protéine de fusion P90 Gag-Fps.

## III. L'élément NRS 22 contenant l'hélice H1 et les tiges boucles SLS1 et SLS4 est le nouvel élément NRS minimal fonctionnel pour l'inhibition de l'épissage *in vitro*

Après avoir défini expérimentalement la structure secondaire de l'élément NRS, nous avons étudié l'importance fonctionnelle de chacun des éléments de structure le composant lors de

l'inhibition de l'épissage. La tige boucle SLS1 a été tronquée dans le mutant NRS 26, la tige boucle SLS4 dans le mutant NRS 24, l'hélice H2 a été respectivement partiellement et totalement tronquée dans les mutants NRS 28 et 22, en plus des troncations des tiges boucles SLS2 et SLS3 (**Figure XII-7**). Nous avons mesuré l'inhibition de l'épissage de ces différents mutants ainsi que des éléments NRS WT et Δ76 qui sont décrits comme des éléments NRS fonctionnels et qui nous ont servi de témoins (McNally *et al.*, 1991).

Les éléments NRS ont été insérés dans l'intron du mini-gène C2 à une distance optimale du site donneur permettant la mise en place de l'inhibition de l'épissage par l'élément NRS WT (Gontarek *et al.*, 1993 ; McNally *et al.*, 1991 ; Wang et McNally, 2003). L'insertion de l'élément NRS dans un mini-gène pouvait modifier le taux d'épissage du pré-ARNm dans lequel elle est insérée. Pour contourner ce biais, nous avons également inséré dans l'intron du mini-gène C2 l'élément NRS Inv qui correspond à l'élément NRS dans son orientation antisens et qui n'est pas fonctionnel en inhibition d'épissage. Le taux d'épissage du pré-ARNm C2-NRS Inv a servi de référence de 100% d'épissage et les pourcentages d'épissage des différents pré-ARNm C2-NRS ont été calculés à partir de cette référence.

Les éléments NRS WT et Δ76 précédemment décrits comme étant fonctionnels, inhibent l'épissage entre les sites D1-A2 du pré-ARNm C2 lorsqu'ils sont insérés dans l'intron. 40 et 50 % des pré-ARNm C2-NRS WT et C2-NRS Δ76 étaient respectivement épissés, ce qui montrait que l'utilisation de la construction C2 était adaptée pour étudier l'inhibition de l'épissage par l'élément NRS. Ceci rejoignait pas ailleurs les résultats de la littérature, indiquant que l'élément NRS Δ76 était aussi fonctionnel que l'élément NRS WT (McNally *et al.*, 1991). L'élément NRS Δ76 a longtemps été décrit comme étant l'élément minimal fonctionnel (McNally *et al.*, 1991). Cependant, l'étude des mutants NRS 22 et 28, qui sont plus courts que l'élément NRS Δ76 en raison de la troncation de la tige boucle SLS3 et de la troncation partielle (NRS 28) ou totale de l'hélice H2 (NRS 22), sont aussi fonctionnels en inhibition de l'épissage *in vitro* que l'élément NRS WT et NRS Δ76 (**Figure XII-7**). Nous avons ainsi défini un nouvel élément NRS minimal fonctionnel, le NRS 22 qui ne comporte que l'hélice H1 et les tiges boucles SLS1 et SLS4.

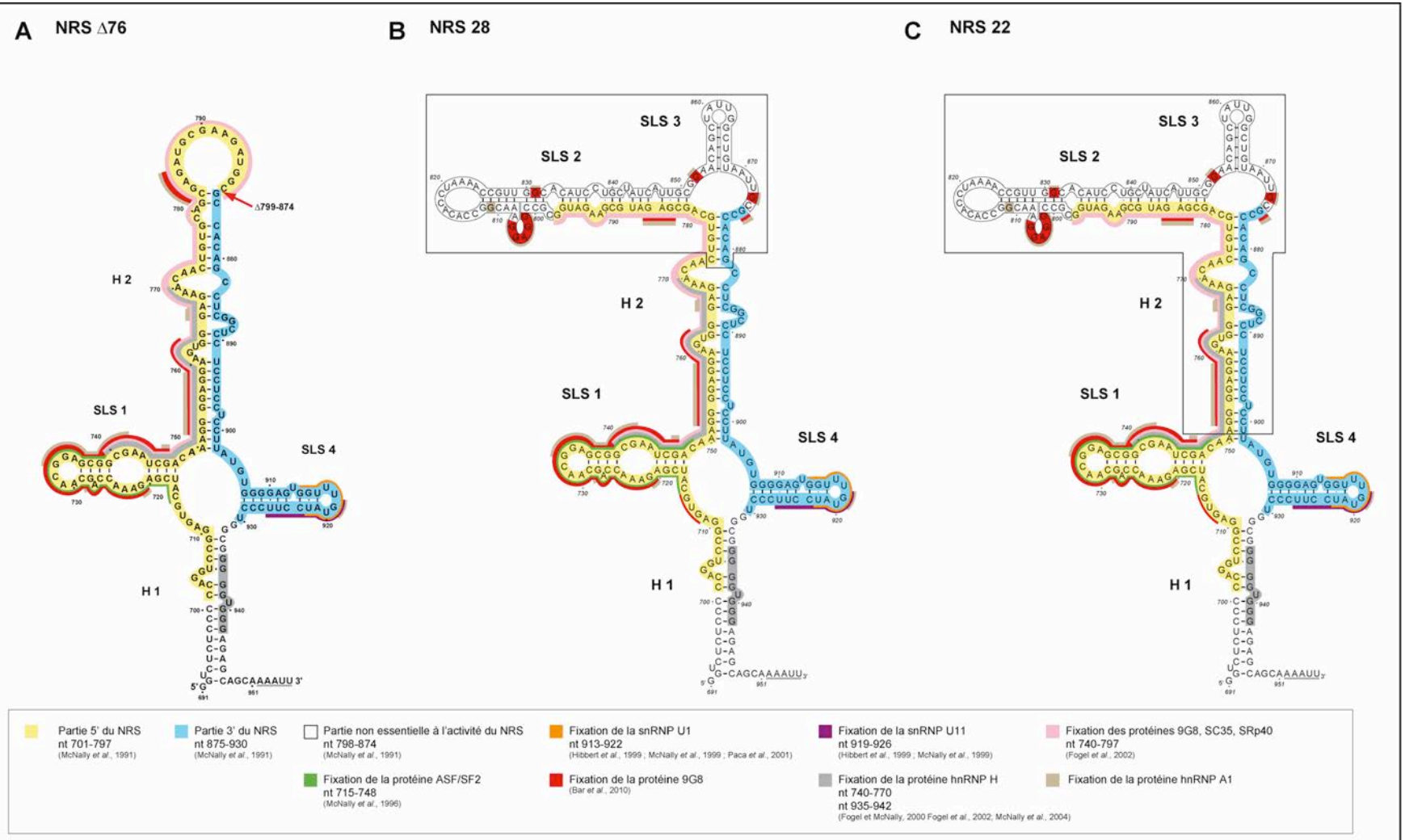



**Figure XII-7 : Représentations des éléments NRS tronqués fonctionnels en inhibition de l'épissage *in vitro*.**
**A :** Représentation de la structure secondaire de l'élément NRS Δ76 définie expérimentalement. La troncation de la séquence comprise entre les nucléotides 799-874 est indiquée par une flèche rouge.
**B :** Représentation de la structure secondaire de l'élément NRS 28, la séquence tronquée est encadrée en noir.
**C :** Représentation de la structure secondaire de l'élément NRS 22, la séquence tronquée est encadrée en noir.
Le nom des tiges boucles (SLS) et des hélices (H) est indiqué sur chacune des structures secondaires. Les sites de fixation des protéines sont reportés sur la structure secondaire et sont colorés selon la légende précisée dans l'encart.

### III.1. Relation structure-fonction de l'élément NRS 22

La structure secondaire de l'élément NRS 22 n'a pas été définie expérimentalement. La structure secondaire prédite théoriquement à l'aide du logiciel Mfold indique que les séquences de cet ARN adoptent les mêmes structures secondaires que celles qui sont définies expérimentalement pour l'élément NRS WT, à savoir l'hélice H1, et les tiges boucles SLS1 et SLS4.

D'un point de vue fonctionnel, la structure secondaire de chacune des tiges boucles SLS1 et SLS4 ainsi que la structure secondaire de l'ensemble de l'élément NRS 22 pourraient jouer un rôle important. La large boucle interne qui relie les tiges boucles SLS1 et SLS4 apporte une flexibilité à l'ensemble de la structure qui pourrait faciliter l'interaction avec les partenaires de chacune des tiges boucles. De plus, elle empêcherait un encombrement stérique qui entraînerait une fixation exclusive de chacun des partenaires, comme cela a été démontré dans le cas du pseudoexon du gène ATM (Dhir *et al.*, 2010). Ce pseudoexon se situe dans l'intron entre l'exon 20 et 21 du gène ATM (**Figure XII-8**). Il comprend à son extrémité 5' un site accepteur et à son extrémité 3' un site donneur. Il comporte en son centre un site donneur d'épissage appelé ISPE (Intronic Splicing Processing Element) qui est suivi d'une séquence fixant la protéine ASF/SF2. Le site donneur ISPE est un site fonctionnel mais il n'est jamais utilisé lors de l'épissage du pré-ARNm du gène ATM. La fixation de la snRNP U1 sur le site donneur ISPE inhibe l'épissage du pseudoexon (**Figure XII-8, A1**). En revanche, la troncation du site donneur ISPE provoque l'épissage du pseudoexon et est à l'origine de la maladie *ataxia telangiectasia* (Pagani *et al.*, 2002) (**Figure XII-8, B1**).

Comme dans le cas de l'élément NRS, la fixation de la snRNP U1 provoque une inhibition de l'épissage. Cependant, les mécanismes d'inhibition sont différents. En effet, contrairement à l'élément NRS, les fixations de la snRNP U1 et de la protéine ASF/SF2 sur le pseudoexon sont mutuellement exclusives et antagonistes (Dhir *et al.*, 2010). Cela est dû aux structures secondaires adoptées par les sites de fixation de ces deux facteurs. Contrairement à l'élément NRS 22, dans le pseudoexon, la structure secondaire des sites de fixation de la protéine ASF/SF2 et de la snRNP U1 se situent sur chacun des brins d'une boucle interne d'une hélice, ce qui empêche par encombrement stérique la fixation simultanée des deux facteurs (Dhir *et al.*, 2010) (**Figure XII-8, A2**). En l'absence de la séquence ISPE, la snRNP U1 ne se fixe plus, ce qui permet la fixation de la protéine ASF/SF2 et la reconnaissance des sites accepteur et donneur du pseudoexon entraînant ainsi son épissage (Dhir *et al.*, 2010)(**Figure XII-8, B2**).

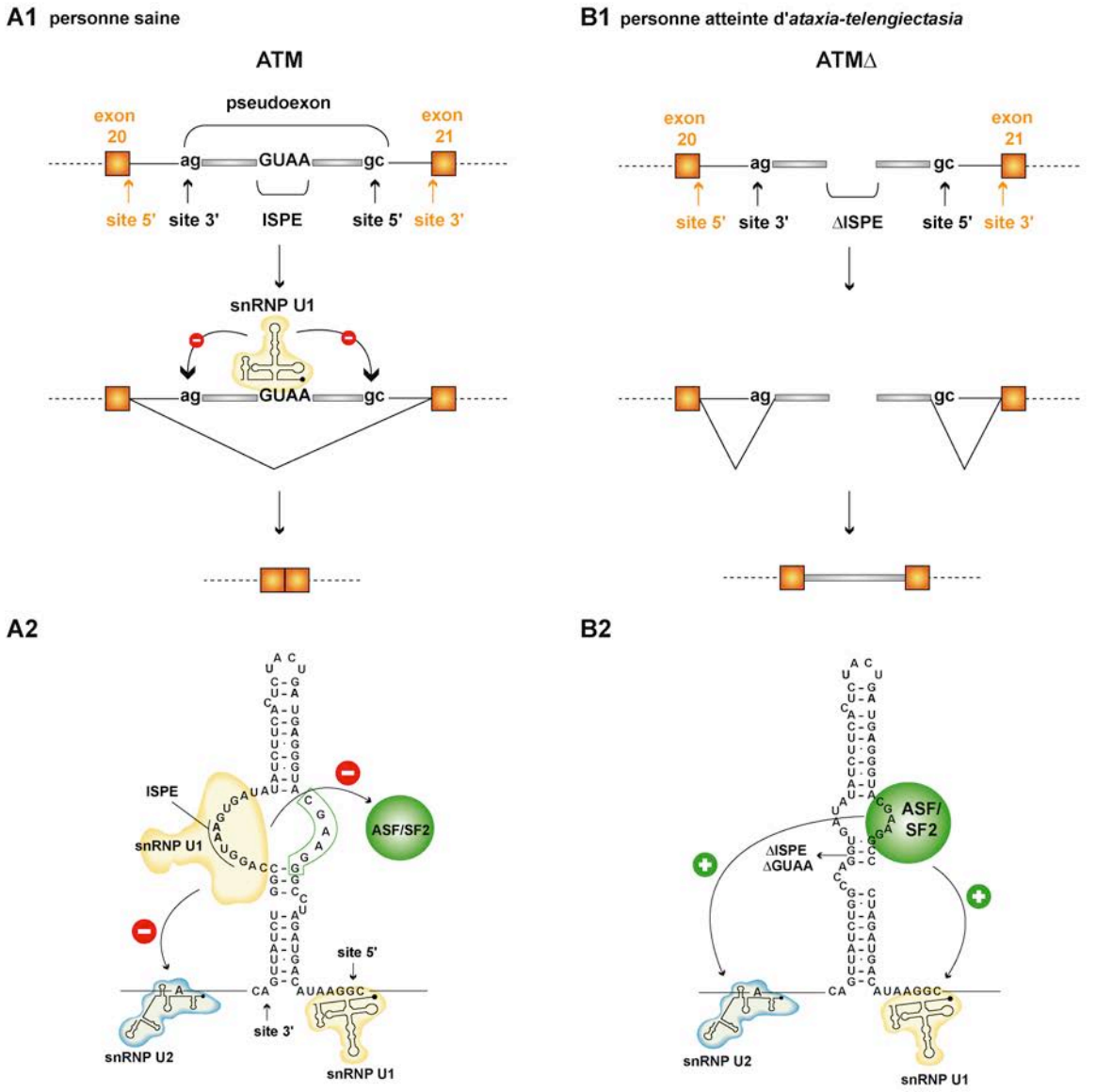



**Figure XII-8 : La fixation de la snRNP U1 sur le pseudoexon du gène ATM inhibent son épissage (Dhir *et al.*, 2010).**

**A1 :** Chez les personnes saines, la fixation de la snRNP U1 sur le site donneur de la séquence ISPE inhbibe l'épissage du pseudoexon. Les exons 20 et 21 et le pseudoexon sont respectivement représentés en orange et en gris. Le site 5' (gc) et 3' d'épissage (ag) bordant le pseudoexon et la séquence ISPE sont mentionnés.

**A2 :** La fixation de la snRNP U1 sur la séquence ISPE et la structure secondaire adoptée par le pseudoexon empêchent par encombrement stérique la fixation de la protéine ASF/SF2 sur le brin de la boucle interne faisant face à la séquence ISPE (séquence encadrée en vert). La fixation de la snRNP U1 empêche également le recrutement efficace de la snRNP U2 sur le point de branchement du site 3' du pseudoexon.

**B1 :** Chez les personnes atteintes d'*ataxia telengiectasia*, la troncation de la séquence ISPE (ΔISPE) du pré-ARNm ATM (ΔATM) entraîne l'épissage du pseudoexon.

**B2 :** La troncation de la séquence ISPE permet la fixation de la protéine ASF/SF2 qui favorise le recrutement de la snRNP U2 et de la snRNP U1 sur les sites 3' et 5' bordant le pseudoexon.

Bien que nous n'ayons pas démontré le rôle de la structure de la tige boucle SLS1 dans l'inhibition de l'épissage, la comparaison de la structure secondaire théorique de l'élément NRS 22 avec celle du pseudoexon du gène ATM sont deux exemples qui pourraient illustrer le rôle positif ou négatif que peut jouer la structure secondaire dans le recrutement des facteurs comme la protéine ASF/SF2 et la snRNP U1.

Pour le confirmer, la structure de l'élément NRS 22 devra dans un premier temps être définie expérimentalement en solution afin de s'assurer qu'elle est identique à la structure théorique. Dans un premier temps, nous étudierons si la boucle interne reliant les tiges boucles SLS1, SLS4 et l'hélice I apporte une flexibilité qui est nécessaire à la fonction de l'élément NRS. Pour cela, nous pourrions faire varier la taille de cette boucle interne et en mesurer l'effet sur la fixation de la protéine ASF/SF2 et de la snRNP U1 ainsi que sur l'inhibition de l'épissage.

Dans un second temps, nous étudierons l'importance de la structure de la tige boucle SLS1 pour la fixation de la protéine ASF/SF2. Pour cela, nous réaliserons tout d'abord des expériences d'empreinte avec la protéine ASF/SF2 recombinante afin de définir précisément ses sites de fixation sur la tige boucle SLS1. Nous réaliserons des mutations afin de déstructurer la tige boucle SLS1 tout en conservant les sites de fixations de cette protéine intacts et, à l'inverse, muter les sites de fixation de façon à conserver la structure tige boucle SLS1. Il existe de nombreux exemples démontrant que la structure secondaire des séquences ESE joue un rôle important dans leur fonction. Par exemple, l'ESE de l'exon EDA de la fibronectine de souris normalement en simple brin dans une boucle apicale du pré-ARNm n'active plus l'épissage de cet exon lorsqu'elle se situe dans une région plus structurée en double brin dans une hélice (Buratti *et al.*, 2004). Ce changement de conformation provoque également une perte de la fixation de son principal partenaire, la protéine ASF/SF2 (Buratti *et al.*, 2004). Ainsi, comme dans le cas de l'ESE de l'exon EDA, la fixation de la protéine ASF/SF2 sur la tige boucle SLS1 pourrait dépendre de la structure de cette dernière.

### III.2. Fonctionnement de l'élément NRS 22 lors de l'inhibition de l'épissage

D'après nos expériences d'empreinte, la protéine 9G8 se fixe sur la tige boucle SLS1 ce qui coïncide avec les sites de fixation de la protéine ASF/SF2 (**Figure XII-7, C**). Cependant, cette fixation ne semble pas suffisante car l'inhibition de l'épissage de l'élément NRS 22 n'est pas stimulée par la protéine 9G8 lors d'expériences d'épissage *in vitro* (Bar *et al.*, 2011 ; McNally et McNally, 1996). Dans le cas de l'élément NRS 22, la protéine ASF/SF2 serait le

principal activateur de l'inhibition de l'épissage (**Figure XII-9**). Pour le vérifier, il serait nécessaire de réaliser des expériences d'épissage *in vitro* avec un extrait nucléaire complémenté avec la protéine ASF/SF2 recombinante et de le confirmer par des expériences d'épissage *in cellulo* après surexpression de la protéine ASF/SF2 ou utilisation de siRNA dirigés contre l'ARNm codant cette protéine.

Une autre approche pour démontrer la nécessité d'un site de fixation de la protéine ASF/SF2 pour l'inhibition de l'épissage par le NRS, serait de remplacer la tige boucle SLS1 par une tige boucle de même stabilité comme la tige boucle aptamérique reconnaissant la protéine du phage MS2 de façon à ne pas remodeler la structure secondaire de l'ensemble de l'élément NRS 22. La protéine ASF/SF2 serait fusionnée à la protéine de l'enveloppe du phage MS2 qui se fixe avec une grande affinité à la séquence aptamérique. De cette façon, nous pourrions savoir si la proximité d'une structure tige boucle fixant la protéine ASF/SF2 favorise d'une part le recrutement de la snRNP U1 sur le pseudo site de la tige boucle SLS4 et d'autre part l'inhibition de l'épissage par l'élément NRS.

Les protéines hnRNP A1 et hnRNP H se fixent également sur l'élément NRS 22 (**Figure XII-9**). Ces deux protéines sont présentes au sein des complexes ribonucléoprotéiques assemblés sur cet ARN lors des chromatographie d'affinité (Bar *et al.*, 2011). La protéine hnRNP H se fixe sur la séquence riche en G dans l'hélice H1 et d'après nos expériences d'empreinte réalisées sur l'élément NRS WT, la protéine hnRNP A1 se fixe sur la tige boucle SLS1 (Bar *et al.*, 2011 ; McNally *et al.*, 2004). Dans le contexte de l'élément NRS 22, la protéine hnRNP A1 pourrait entrer en compétition avec la protéine ASF/SF2 pour la fixation sur la tige boucle SLS1. Il existe de nombreux exemples de régulation de l'épissage alternatif pour lesquels la protéine hnRNP A1 et la protéine ASF/SF2 ont des effets antagonistes et entrent en compétition pour la fixation à une même séquence ou à des séquences chevauchantes (Eperon *et al.*, 2000 ; Expert-Bezancon *et al.*, 2004 ; Rooke *et al.*, 2003). Il est possible qu'un tel scénario puisse exister au niveau de la tige boucle SLS1 de l'élément NRS 22 : le recrutement de la snRNP U1 sur le pseudo site situé sur la tige boucle SLS4 pourrait être activé par la fixation de la protéine ASF/SF2 tandis que la fixation de la protéine hnRNP A1 l'inhiberait (**Figure XII-9**). Pour le confirmer, nous pourrions réaliser des empreintes sur l'élément NRS 22 avec chacune des protéines purifiées afin d'identifier précisément leur sites de fixation et de savoir s'ils sont identiques. Pour montrer que ces protéines entrent en compétition, nous pourrions réaliser des expériences de pontages aux UV avec des gammes de concentration de chacune des protéines afin de voir si la fixation de l'une d'entre elle peut

déplacer celle de l'autre. Ces expériences pourront être réalisées avec l'élément NRS 22 mais aussi avec la tige boucle SLS1 seule.

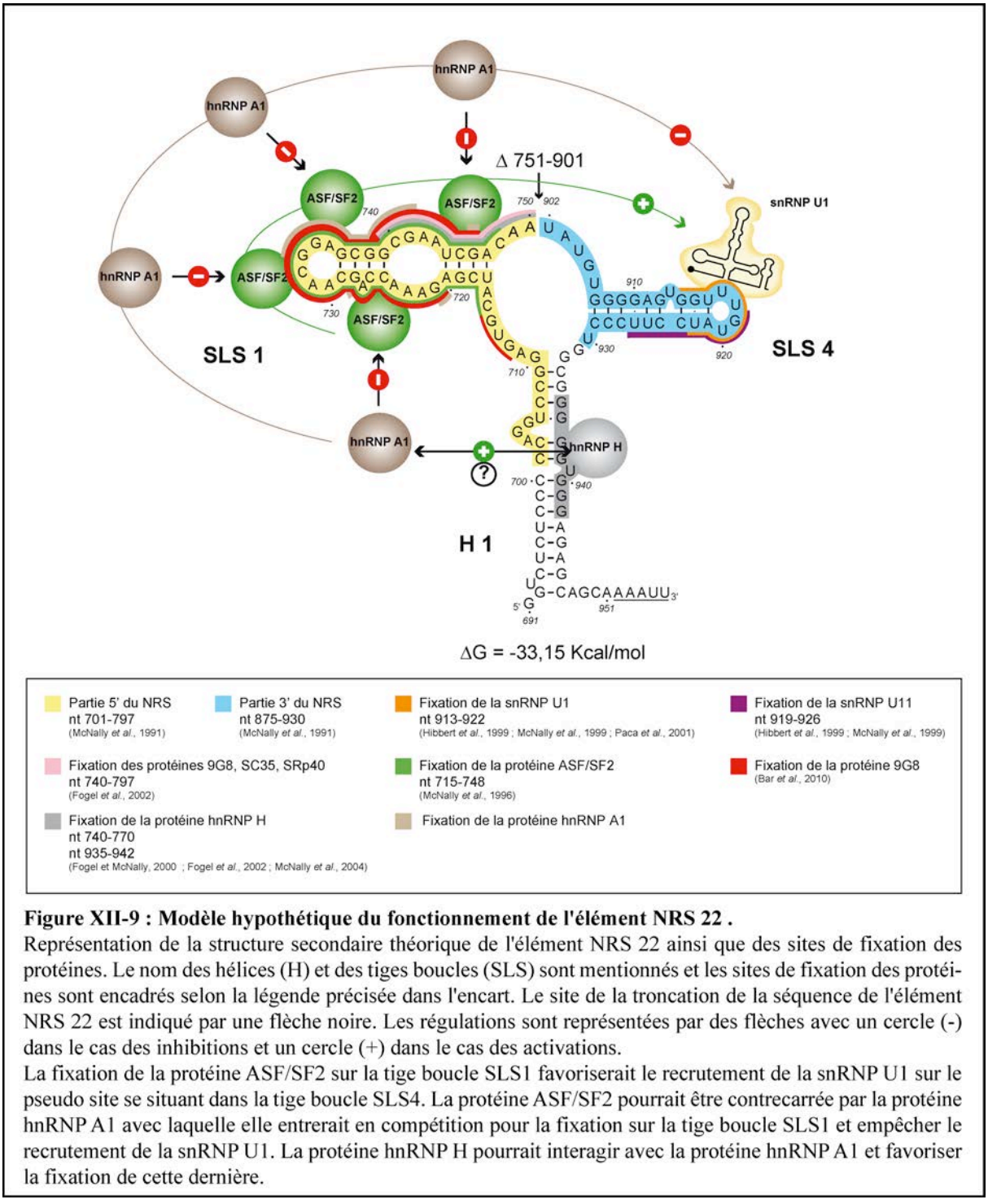



**Figure XII-9 : Modèle hypothétique du fonctionnement de l'élément NRS 22 .**
Représentation de la structure secondaire théorique de l'élément NRS 22 ainsi que des sites de fixation des protéines. Le nom des hélices (H) et des tiges boucles (SLS) sont mentionnés et les sites de fixation des protéines sont encadrés selon la légende précisée dans l'encart. Le site de la troncation de la séquence de l'élément NRS 22 est indiqué par une flèche noire. Les régulations sont représentées par des flèches avec un cercle (-) dans le cas des inhibitions et un cercle (+) dans le cas des activations.
La fixation de la protéine ASF/SF2 sur la tige boucle SLS1 favoriserait le recrutement de la snRNP U1 sur le pseudo site se situant dans la tige boucle SLS4. La protéine ASF/SF2 pourrait être contrecarrée par la protéine hnRNP A1 avec laquelle elle entrerait en compétition pour la fixation sur la tige boucle SLS1 et empêcher le recrutement de la snRNP U1. La protéine hnRNP H pourrait interagir avec la protéine hnRNP A1 et favoriser la fixation de cette dernière.

En se fixant sur les 4 séquences riches en G se situant en aval du pseudo site, la hnRNP H favorise le recrutement de la snRNP U11 sur le pseudo site dans la tige boucle SLS4 (McNally *et al.*, 2004, 2006). Dans le contexte de l'élément NRS 22, seule une de ces séquences riches en G est présente dans l'hélice H1, ce qui n'est pas suffisant pour que la protéine hnRNP H puisse recruter la snRNP U11 (McNally *et al.*, 2004, 2006). Aucune autre

fonction n'est connue pour ces séquences riches en G. Toutefois, sur la base de nos résultats d'empreinte avec la protéine hnRNP A1, il est possible que la protéine hnRNP H favorise la fixation de la protéine hnRNP A1. Il a récemment été démontré que ces deux protéines interagissent entre elles via leur domaine Gly (Fisette *et al.*, 2010). Afin d'étudier s'il existe une coopérativité mutuelle pour la fixation de ces deux protéines, nous pourrions réaliser des pontages aux UV avec les deux protéines recombinantes tronquées de leur domaine Gly ainsi que des expériences d'empreintes. Nous pourrions également mesurer l'inhibition de l'épissage par l'élément NRS lors expériences d'épissage *in vitro* réalisées avec extrait nucléaire déplété en chacune des deux protéines hnRNP H et A1. Nous pourrions également muter la séquence riche en G de l'hélice H1 de façon à ce que la protéine hnRNP H ne s'y fixe plus et en mesurer les répercussions sur la fixation de la protéine hnRNP A1 et l'inhibition de l'épissage.

## IV. Les tiges boucles SLS1 et SLS4 de l'élément NRS sont essentielles à sa fonction

Nous avons également confirmé le fait que les séquences des tige boucles SLS1 et SLS4 sont essentielles à l'inhibition de l'épissage par l'élément NRS, comme l'indiquent les mutants NRS 24 et 26 qui en sont respectivement délétés et qui ne sont pas fonctionnels (**Figure XII-10**). Dans le contexte du pré-ARNm C2 dans lequel les différents éléments NRS ont été insérés, cela se traduit par une augmentation de l'épissage qui est similaire à celle mesurée pour l'élément NRS Inv qui correspond à l'orientation antisens de l'élément NRS WT et qui est non fonctionnel. Le nouvel élément NRS 22 minimal fonctionnel que nous avons défini lors des expériences d'épissage *in vitro* et qui contient ces deux tiges boucles renforce également leur importance fonctionnelle dans la mise en place de l'inhibition de l'épissage.

La perte de l'inhibition de l'élément NRS 24 est due à la troncation de la séquence du pseudo site comprise dans la tige boucle SLS1, ce qui provoque la perte de la fixation de la snRNP U1 sur l'élément NRS. Contrairement aux autres éléments NRS qui possèdent la tige boucle SLS1, la snRNP U1 n'est pas présente au sein des complexes assemblés sur l'élément NRS 24 purifiés par la chromatographie d'affinité basée sur l'emploi de la protéine de l'enveloppe du phage MS2 (Bar *et al.*, 2011). Ces résultats rejoignent les données de la littérature démontrant que l'absence d'un pseudo site fonctionnel inactive l'élément NRS (Hibbert *et al.*, 1999 ; Paca *et al.*, 2001).

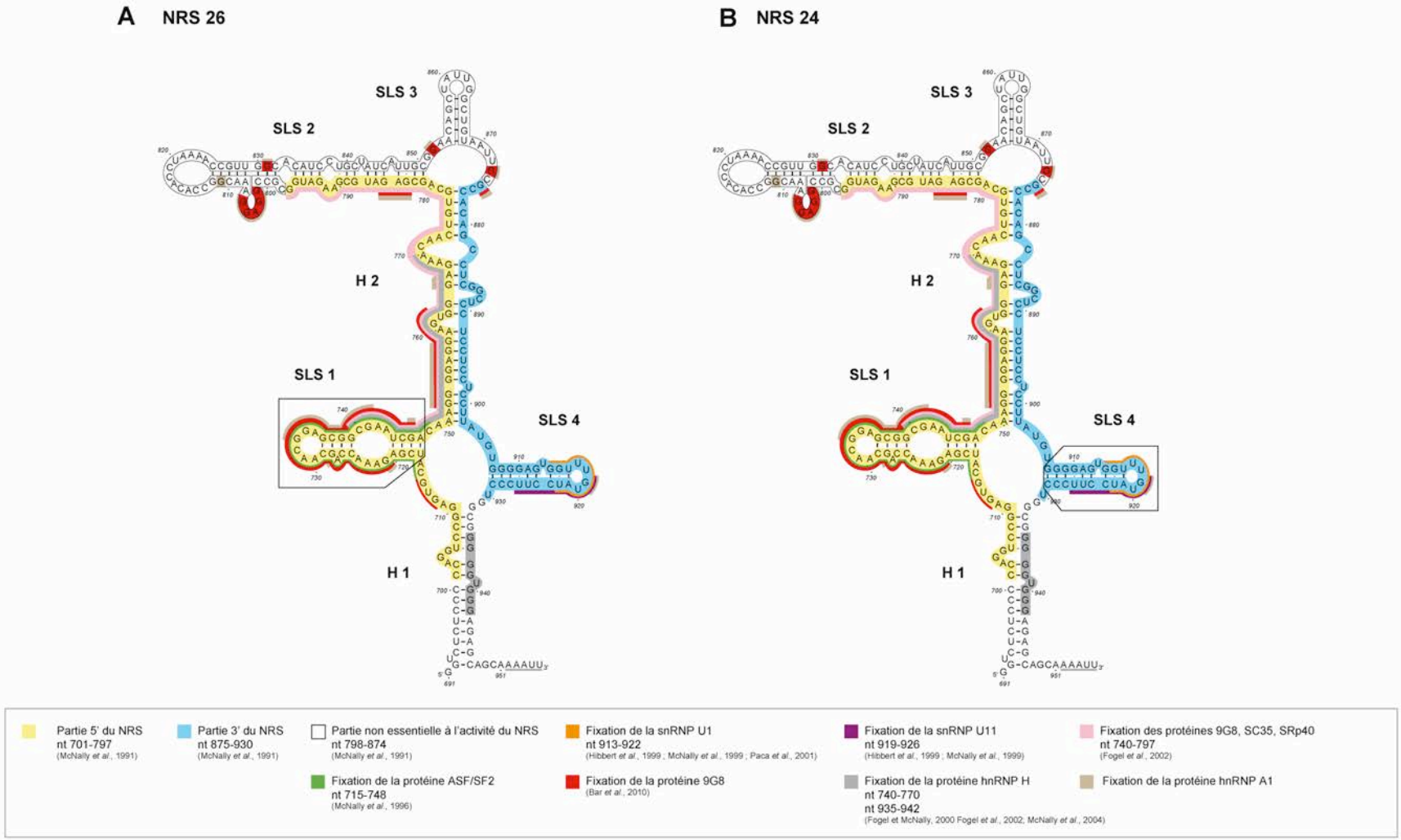



**Figure XII-10 : Représentations des éléments NRS tronqués non fonctionnels en inhibition de l'épissage *in vitro*.**
**A :** Représentation de la structure secondaire de l'élément NRS 26, la séquence tronquée est encadrée en noir.
**B :** Représentation de la structure secondaire de l'élément NRS 24, la séquence tronquée est encadrée en noir.
Le nom des tiges boucles (SLS) et des hélices (H) est indiqué sur chacune des structures secondaires. Les sites de fixation des protéines sont reportés sur la structure secondaire et sont colorés selon la légende précisée dans l'encart.

## IV.1. La présence du pseudo site et la structure secondaire tige boucle SLS4 qu'il adopte sont essentielles mais pas suffisantes pour la fonction de l'élément NRS

### IV.1.1. La fixation de la snRNP U1 sur le pseudo site dépend de sa séquence et de la structure secondaire qu'il adopte

Comme nous l'avons déjà mentionné, le taux de fixation de la snRNP U1 sur le pseudo site n'est pas seulement directement corrélé avec le nombre d'appariements que le snRNA U1 peut établir avec celui-ci mais dépend également de la stabilité de la tige boucle SLS1 (Cabello-Villegas *et al.,* 2004). Le snRNA U1 ne se fixe pas sur le pseudo site lorsque la structure secondaire adoptée par celui-ci est stabilisée par la troncation du nucléotide U913. Cette troncation permet la formation de 9 appariements de bases continus dans l'hélice de la tige boucle SLS4 sans changer le nombre d'appariements potentiels entre le pseudo site et le snRNA U1 car le nucléotide U913 est non canonique (Cabello-Villegas *et al.,* 2004). Dans ce cas, il semble que les appariements de l'hélice ne puissent pas se rompre pour permettre à la séquence du pseudo site de s'apparier avec le snRNA U1. La structure secondaire de la tige boucle SLS4 constitue donc un paramètre supplémentaire qui régule la disponibilité du pseudo site à la snRNP U1. Cette régulation est également retrouvée chez les virus VIH-1 et MLV dont l'efficacité d'utilisation en épissage du site donneur D1 et D, respectivement, est régulée par la stabilité de la structure secondaire en tige boucle qui le contient (Abbink et Berkhout, 2008 ; Zychlinski *et al.,* 2009).

En plus de la structure secondaire, l'appariement du snRNA U1 au pseudo site peut être régulé par d'autres facteurs comme l'hélicase p68 (ou DDX5). Il a été montré que cette hélicase défait les appariements entre le snRNA U1 et le site donneur, ce qui provoque le départ de la snRNP U1 (Lee, 2002 ; Lin *et al.,* 2005). Mais, il a aussi été montré que l'hélicase p68 pouvait au contraire favoriser l'appariement du snRNA U1 sur le site donneur de l'exon 10 du gène Tau en déstabilisant la structure tige boucle qui le séquestre (Kar *et al.,* 2011). L'hélicase p68 seule peut se fixer sur un site donneur, ce qui suggère qu'elle y serait recrutée avant la snRNP U1 (Kar *et al.,* 2011). Ce recrutement serait réalisé par la protéine RBM4 qui se fixe sur une séquence ISE en aval du site donneur (Kar *et al.*, 2011). La protéine p68 est présente dans les complexes ribonucléoprotéiques assemblés sur l'élément NRS WT que nous avons purifiés et caractérisés tandis que la protéine RBM4 n'y a pas été identifiée. Il est possible que cette hélicase puisse être recrutée sur le pseudo site par un autre facteur.

### IV.1.2. Le recrutement de la snRNP U1 sur le pseudo site lors de l'épissage dépend aussi de facteurs supplémentaires

Comme nous venons de l'aborder, la fixation de la snRNP U1 sur la séquence du pseudo site dépend de la séquence et de la structure de celui-ci. A ces deux paramètres vient s'ajouter un troisième qui est le recrutement de la snRNP U1 sur le pseudo site par les protéines associées au spliceosome lors de l'épissage.

*In vitro*, la snRNP U1 peut se fixer à tout élément NRS à partir du moment où celui-ci possède une séquence et une structure du pseudo site qui sont identiques à celles de l'élément NRS WT. En revanche, lors de l'inhibition de l'épissage, la snRNP U1 ne peut pas se fixer sur le pseudo site d'un élément NRS si celui-ci ne peut pas fixer des facteurs d'épissage qui eux, vont permettre le recrutement de la snRNP U1 sur le pseudo site lors de l'assemblage des complexes spliceosomaux. L'élément NRS ΔLR9 est un exemple qui illustre ce dernier point.

L'élément NRS ΔLR9 n'est pas fonctionnel mais son pseudo site a la même séquence et adopte la même structure que celles établies pour l'élément NRS WT, qui lui est fonctionnel (Bar *et al.*, 2011 ; Smith *et al.*, 1997). *In vitro*, le taux de fixation du snRNA U1 sur l'élément NRS ΔLR9 est identique à celui de l'élément NRS WT (Polony *et al.*, 2003). En revanche, la troncation de la séquence comprise entre nucléotides 735-776 de l'élément NRS ΔLR9 provoque un remaniement important de sa structure secondaire ainsi qu'une diminution de la fixation des protéines ASF/SF2 et hnRNP H, par rapport à celles observées sur l'élément NRS WT (Bar *et al.*, 2011 ; Polony *et al.*, 2003). Cela suggère que dans le cas de l'élément NRS ΔLR9, la diminution de la fixation de ces protéines et/ou le remaniement de la structure secondaire sont en partie responsables de la perte du recrutement de la snRNP U1 sur le pseudo site lors de l'épissage, qui se traduit par une absence d'inhibition de l'épissage (Bar *et al.*, 2011 ; Smith *et al.*, 1997).

D'une manière générale, cela suppose donc que des séquences et/ou structures autres que celles du pseudo site au sein de l'élément NRS sont déterminantes pour sa fonction et participent à la reconnaissance du pseudo site par le spliceosome lors de la mise en place de l'inhibition de l'épissage.

## IV.2. Les tiges boucles SLS1 et SLS4 sont nécessaires à la fixation des protéines SR au sein des complexes ribonucléoprotéiques assemblés sur l'élément NRS

Afin de connaître les facteurs responsables de la perte de fonction des éléments NRS 24 et 26, nous avons étudié la composition des complexes ribonucléoprotéiques assemblés sur ces éléments. Comme attendu, la troncation de la tige boucle SLS4 comprenant le pseudo site

dans l'élément NRS 24 provoque la perte de la fixation de la snRNP U1 tandis que la troncation dans l'élément NRS 26 de la tige boucle SLS1 comprenant les sites de fixation des protéines ASF/SF2 et 9G8 diminue de 40 % le taux de fixation de la snRNP U1 sur le pseudo site (Bar *et al.*, 2011).

Lors de l'assemblage des complexes, la fixation de la snRNP U1 sur l'élément NRS ne semble dépendre principalement de la présence du pseudo site, mais sa fixation est favorisée par la présence de la tige boucle SLS1. Ces résultats renforcent l'hypothèse que la fixation des protéines SR sur la tige boucle SLS1 favorise le recrutement de la snRNP U1 sur le pseudo site dans la tige boucle SLS4 (McNally et McNally, 1999). En revanche, d'après nos résultats d'épissage, la fixation de la snRNP U1 sur le pseudo site dépendrait de la présence de la tige boucle SLS1 et des protéines SR qui s'y fixent.

Nous avons également montré que les tiges boucles SLS1 et SLS4 sont interdépendantes pour recruter les protéines SR ASF/SF2, 9G8, SC35 et la protéine hnRNP H au sein complexes ribonucléoprotéiques assemblés sur l'élément NRS (Bar *et al.*, 2011).

L'absence des ces facteurs au sein des complexes assemblés sur l'élément NRS 26 rejoint en partie des résultats précédents obtenus par chromatographie d'exclusion de taille où il était démontré que les complexes ribonucléoprotéiques ne pouvaient s'assembler spécifiquement sur l'élément NRS qu'en présence de la protéine ASF/SF2 et de son site de fixation (Cook et McNally, 1996, 1998). Ce résultat rejoint également nos résultats d'empreinte réalisés avec la protéine 9G8 recombinante dont les sites de fixation de haute affinité se situent dans la tige boucle SLS1 (Bar *et al.*, 2011) (**Figure XII-11**). L'absence des protéines SC35 et hnRNP H des complexes assemblés sur l'élément NRS 26 peuvent quant à elles être dues à la troncation partielle de la séquence comprise entre les nucléotides 740-747 qui fait partie de leur site de fixation (Fogel *et al.*, 2002 ; Fogel et McNally, 2000) (**Figure XII-11**).

Contrairement à l'élément NRS 26, l'absence de ces facteurs au sein des complexes assemblés sur l'élément NRS 24 dont la tige boucle SLS4 qui contient le pseudo site est tronquée, n'était pas un résultat attendu (Bar *et al.*, 2011). D'après nos résultats et les données de la littérature, aucune des ces protéines n'interagit avec la tige boucle SLS4 (Fogel *et al.*, 2002 ; Fogel et McNally, 2000 ; McNally et McNally, 1996). Les répercussions de la mutation ou de la troncation du pseudo site sur la fixation des protéines SR ou hnRNP sur l'élément NRS n'ont jamais été étudiées. Il a seulement été observé que la mutation du pseudo site sur lequel ne se fixent plus les particules snRNP U1 et U11 diminue fortement

l'assemblage de complexes ribonucléoprotéiques sur l'élément NRS (Cook et McNally, 1996 ; Gontarek *et al.*, 1993). La composition de ces complexes n'a jamais été étudiée mais dans les deux cas il est possible que l'absence au sein des complexes ribonucléoprotéiques de la particule snRNP U1 induise la formation de complexes peu stables sur l'élément NRS. Il est possible que la particule snRNP U1 stabilise la fixation des protéines SR sur la tige boucle SLS1 en interagissant avec elles, et inversement. C'est notamment le cas de la protéine U1-70k de la particule snRNP U1 qui interagit avec les protéines SR fixées sur des séquences ESE en amont du site donneur (Kohtz *et al.*, 1994).

### IV.3. La séquence de la tige boucle SLS1 qui fixe les protéines ASF/SF2 et 9G8 est essentielle à l'inhibition de l'épissage par l'élément NRS

Nous supposons que l'inactivation de l'élément NRS 26 est due en partie à la perte de la fonction ESE de la séquence de la tige boucle SLS1 qui est tronquée. La séquence de cette tige boucle SLS1 comprise entre les nucléotides 717-747 n'a jamais été démontrée comme étant une séquence ESE mais de nombreuses données concordantes appuieraient cette fonction.

Comme nous l'avons observé pour la troncation de la tige boucle SLS1 de l'élément NRS 24, les troncations des séquences ESE comprises entre les nucléotides 701-748 et 720-744, provoque une forte diminution de l'inhibition de l'épissage par l'élément NRS (McNally et McNally, 1998). En revanche, la troncation de la séquence comprise entre les nucléotides 701-715 n'avait aucune répercussion sur l'inhibition de l'épissage par l'élément NRS, ce qui suggérait que cette séquence n'avait aucune activité ESE (McNally et McNally, 1996).

Le site de fixation pour la protéine ASF/SF2 a été défini entre les nucléotides 715 et 748 et correspond donc à la séquence ESE de la tige SLS1 (McNally et McNally, 1996). Il a longtemps été suggéré que la fixation de la protéine ASF/SF2 sur la tige boucle SLS1 était le seul facteur qui activait l'inhibition de l'épissage par l'élément NRS bien que cela n'ait jamais été démontré directement (Cook et McNally, 1998 ; McNally et McNally, 1996, 1998). Les données antérieures montrent seulement que la présence de sites de fixation pour cette protéine est nécessaire à l'inhibition de l'épissage par l'élément NRS. Ainsi, le remplacement de la partie NRS 5' (nt 701-798) par la séquence ESE de l'ASLV sur laquelle se fixe la protéine ASF/SF2 permet d'obtenir un taux d'inhibition de l'épissage identique à celui obtenu avec l'élément NRS WT. En revanche, la même expérience répétée avec une séquence ESE de l'ASLV mutée sur laquelle la protéine ASF/SF2 ne se fixe pas provoque une diminution de l'inhibition de l'épissage par l'élément NRS (McNally et McNally, 1998).

Nous avons démontré par des expériences d'empreinte que la séquence SLS1 est également un site de fixation de haute affinité pour la protéine 9G8 (Bar *et al.*, 2011). Cependant, aucune activation de l'inhibition de l'épissage par la protéine 9G8 n'est observée avec l'élément NRS 22 qui contient pourtant la tige boucle SLS1 (Bar *et al.*, 2011). Il est possible que la protéine ASF/SF2 et la protéine 9G8 soient en compétition pour la fixation sur la séquence de la tige boucle SLS1. Cette compétition peut dépendre des concentrations relatives de chacune des protéines dans la cellule et aussi de leurs affinités respectives pour cette séquence. Si la protéine ASF/SF2 a une meilleure affinité que la protéine 9G8, cela expliquerait pourquoi cette dernière n'a pas d'effet sur la capacité d'inhibition de l'élément NRS 22. Pour le démontrer, il serait nécessaire dans un premier temps de réaliser des expériences d'empreinte avec la protéine ASF/SF2 recombinante sur l'élément NRS WT. En effet, les sites de fixation de cette protéine ne sont pas connus car son site de fixation compris entre les nucléotides 715-748 a été défini lors d'expériences de pontage aux UV (McNally et McNally, 1996). De cette façon, nous pourrions identifier de manière précise les sites de fixation de cette protéine, observer si ceux-ci sont les mêmes que ceux de la protéine 9G8 et également savoir si d'autres sites de fixation existent dans la partie NRS 5'. Dans un deuxième temps, nous pourrions mesurer l'affinité de chacune des protéines pour la séquence de la tige boucle SLS1 en réalisant des expériences de retard sur gel avec chacune des protéines purifiées. Cela pourrait être complété par des expériences de pontages aux UV suivies d'immunoprécipitation. Enfin, pour connaître la contribution de chacune de ces protéines dans l'activité ESE de cette séquence et la mise en place de l'inhibition de l'épissage, nous pourrions entreprendre des expériences d'épissage *in vitro* avec un extrait nucléaire dépourvu de protéine ASF/SF2, en présence d'une quantité croissante de protéine 9G8 et inversement.

## V. L'inhibition de l'épissage par l'élément NRS est activée par la fixation de la protéine 9G8 sur la séquence de l'hélice H2

L'inhibition de l'épissage par l'élément NRS est activée par la protéine 9G8 et ce, uniquement lorsque la séquence de l'hélice H2 est présence en totalité ou en majorité, comme l'indiquent les résultats d'épissage *in vitro* obtenus respectivement avec les éléments NRS Δ76 et NRS 28 (Bar *et al.*, 2011). En l'absence de la séquence de l'hélice H2 (élément NRS 22), aucune activation de l'inhibition de l'épissage *in vitro* par la protéine 9G8 n'est observée (Bar *et al.*, 2011). D'après ces résultats, la protéine 9G8 semble donc activer l'inhibition de l'épissage en se fixant sur l'hélice H2. Cette hypothèse est confortée par nos résultats d'empreinte qui montrent que des sites de fixation de la protéine 9G8 existent sur cette

séquence (Bar *et al.*, 2011) (**Figure XII-11**). Par ailleurs, *in cellulo*, la protéine 9G8 est essentielle à l'inhibition de l'épissage par l'élément NRS WT car la diminution de sa concentration cellulaire provoque une perte de fonction de cet élément. Ces résultats suggèrent qu'*in cellulo*, la fonction de la protéine 9G8 ne peut pas être remplacée par la protéine ASF/SF2 ou que la protéine ASF/SF2 seule n'est pas suffisante pour supporter la fonction du NRS.

Nous avons aussi montré que la protéine hnRNP A1 se fixe aussi bien sur la tige boucle SLS1 que sur l'hélice H2 et que ces sites sont identiques à ceux de la protéine 9G8 (**Figure XII-11**). De plus, il a été montré que la protéine hnRNP H se fixe aussi sur l'hélice H2 (Fogel *et al.*, 2002 ; Fogel et McNally, 2000).

L'ensemble de nos résultats permet de proposer le modèle suivant. La protéine 9G8 en se fixant sur l'hélice H2 empêcherait la fixation de la protéine hnRNP A1 et/ou de la protéine hnRNP H et permettrait alors la fixation de la protéine ASF/SF2 sur la tige boucle SLS1. La présence de ces protéines SR permettrait la reconnaissance du pseudo site par le spliceosome et par conséquent la fixation de la snRNP U1. Par contre, la fixation de la protéine hnRNP A1 seule, ou en interaction avec la protéine hnRNP H sur l'hélice H2 pourrait alors y empêcher la fixation de la protéine 9G8. Une fois fixée, la protéine hnRNP A1 pourrait alors oligomériser le long de l'ARN jusqu'à atteindre la tige boucle SLS1 et déplacer/empêcher la protéine ASF/SF2 de s'y fixer. Dans ce cas, la snRNP U1 ne serait pas recrutée sur le pseudo site, et donc pas de formation de complexe spliceosomal abortif menant à l'inhibition de l'épissage.

L'ensemble de ces données suggère que les protéines SR activent l'inhibition de l'épissage au moins en favorisant le recrutement de la snRNP U1. Il serait intéressant de savoir si ces protéines peuvent jouer d'autres rôles dans cette inhibition. Par exemple, elles pourraient participer au recrutement d'une hélicase comme la protéine p68 qui permettrait l'ouverture de la structure tige boucle du pseudo site ou participer à la formation et à la stabilité des complexes spliceosomaux abortifs.

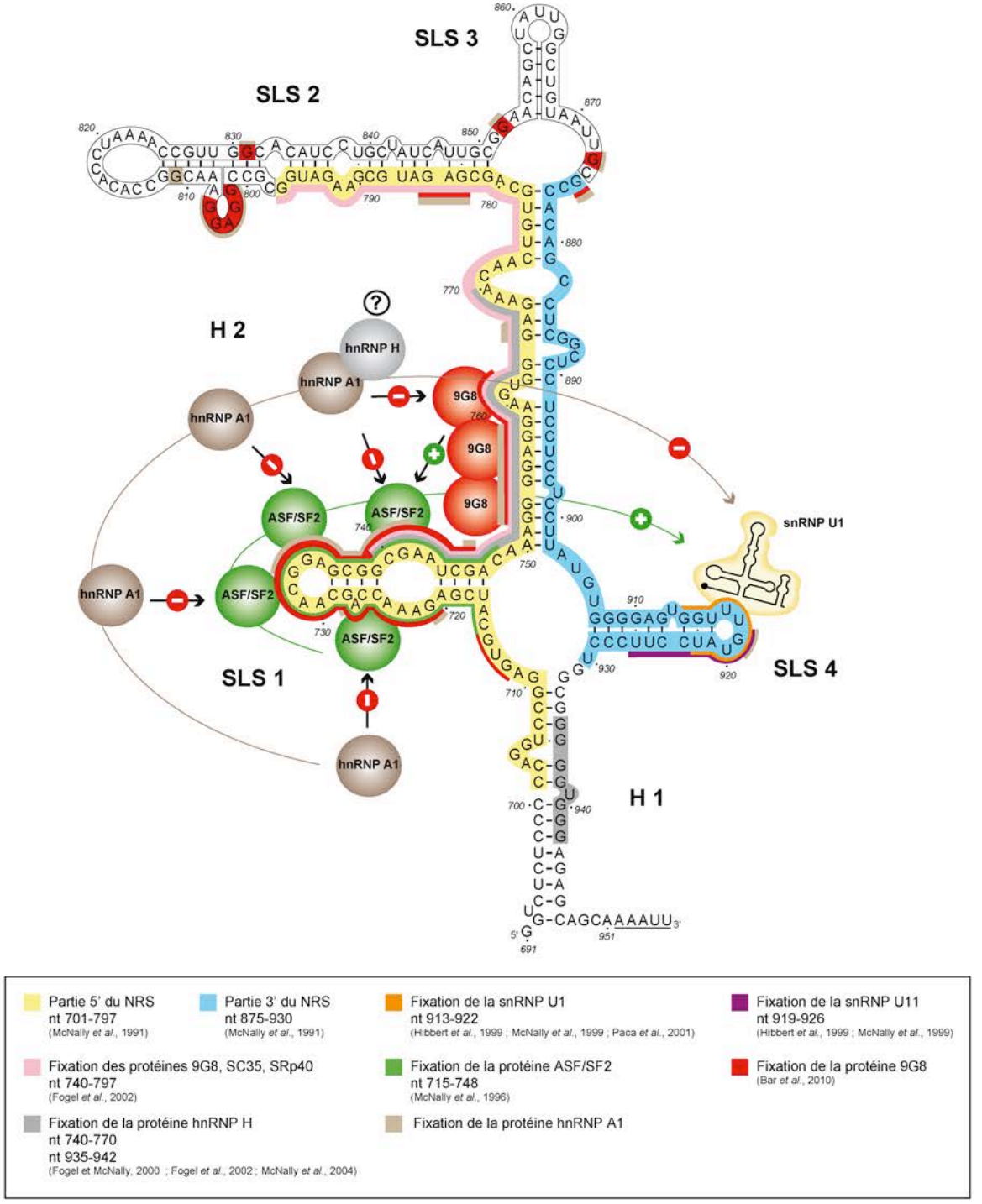



**Figure XII-11 : Modèle proposé pour le fonctionnement de l'élément NRS WT .**
Représentation de la structure secondaire expérimentale de l'élément NRS WT ainsi que des sites de fixation des protéines. Le nom des hélices (H) et des tiges boucles (SLS) sont mentionnés et les sites de fixation des protéines sont encadrés selon la légende précisée dans l'encart. Les régulations sont représentées par des flèches avec un cercle (-) dans le cas des inhibitions et un cercle (+) dans le cas des activations.
La fixation de la protéine 9G8 sur l'hélice H2 favoriserait la fixation de la protéine ASF/SF2 sur la tige boucle SLS1 qui favoriserait ensuite le recrutement de la snRNP U1 sur le pseudo site se situant dans la tige boucle SLS4. La protéine 9G8 pourrait être contrecarrée par la protéine hnRNP A1 et/ou la protéine hnRNP H avec laquelle elle entrerait en compétition pour la fixation sur l'hélice H2. Une fois fixée sur l'hélice H2, la protéine hnRNP A1 pourrait s'oligomériser et empêcher alors la fixation de la protéine ASF/SF2 sur la tige boucle SLS1, ce qui provoquerait alors la perte de la fixation de la snRNP U1.

# Bibliographie

# A


Abbink, T. E. and Berkhout, B. (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J Virol*, **82** (6), 3090-3098.

Abovich, N. and Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*, **89** (3), 403-412.

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S. and Luhrmann, R. (1998) The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol Cell Biol*, **18** (11), 6756-6766.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Luhrmann, R. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *Embo J*, **18** (20), 5789-5802.

Adema, G. J. and Baas, P. D. (1991) Deregulation of alternative processing of Calcitonin/CGRP-I pre-mRNA by a single point mutation. *Biochem Biophys Res Commun*, **178** (3), 985-992.

Agafonov, D. E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C. L., Urlaub, H. and Luhrmann, R. (2011) Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol Cell Biol*, **31** (13), 2667-2682.

Akker, S. A., Misra, S., Aslam, S., Morgan, E. L., Smith, P. J., Khoo, B. and Chew, S. L. (2007) Pre-spliceosomal binding of U1 small nuclear ribonucleoprotein (RNP) and heterogenous nuclear RNP E1 is associated with suppression of a growth hormone receptor pseudoexon. *Mol Endocrinol*, **21** (10), 2529-2540.

Alexander, F., Leis, J., Soltis, D. A., Crowl, R. M., Danho, W., Poonian, M. S., Pan, Y. C. and Skalka, A. M. (1987) Proteolytic processing of avian sarcoma and leukosis viruses pol-endo recombinant proteins reveals another pol gene domain. *J Virol*, **61** (2), 534-542.

Ali, M. B., Chaminade, F., Kanevsky, I., Ennifar, E., Josset, L., Ficheux, D., Darlix, J. L. and Fosse, P. (2007) Structural requirements for nucleocapsid protein-mediated dimerization of avian leukosis virus RNA. *J Mol Biol*, **372** (4), 1082-1096.

Allemand, E., Guil, S., Myers, M., Moscat, J., Caceres, J. F. and Krainer, A. R. (2005) Regulation of heterogenous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. *Proc Natl Acad Sci U S A*, **102** (10), 3605-3610.

Amendt, B. A., Si, Z. H. and Stoltzfus, C. M. (1995a) Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. *Mol Cell Biol*, **15** (11), 6480.

Amendt, B. A., Simpson, S. B. and Stoltzfus, C. M. (1995b) Inhibition of RNA splicing at the Rous sarcoma virus src 3' splice site is mediated by an interaction between a negative cis element and a chicken embryo fibroblast nuclear factor. *J Virol*, **69** (8), 5068-5076.

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

Amrein, H., Gorman, M. and Nothiger, R. (1988) The sex-determining gene tra-2 of Drosophila encodes a putative RNA binding protein. *Cell*, **55** (6), 1025-1035.

Andrake, M. D., Sauter, M. M., Boland, K., Goldstein, A. D., Hussein, M. and Skalka, A. M. (2008) Nuclear import of Avian Sarcoma Virus integrase is facilitated by host cell factors. *Retrovirology*, **5**, 73.

Arrigo, S. and Beemon, K. (1988) Regulation of Rous sarcoma virus RNA splicing and stability. *Mol Cell Biol*, **8** (11), 4858-4867.

Aschoff, J. M., Foster, D. and Coffin, J. M. (1999) Point mutations in the avian sarcoma/leukosis virus 3' untranslated region result in a packaging defect. *J Virol*, **73** (9), 7421-7429.

Ast, G. and Weiner, A. M. (1997) A novel U1/U5 interaction indicates proximity between U1 and U5 snRNAs during an early step of mRNA splicing. *RNA*, **3** (4), 371-381.

Auweter, S. D., Oberstrass, F. C. and Allain, F. H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res*, **34** (17), 4943-4959.

Awasthi, S., Palmer, R., Castro, M., Mobarak, C. D. and Ruby, S. W. (2001) New roles for the Snp1 and Exo84 proteins in yeast pre-mRNA splicing. *J Biol Chem*, **276** (33), 31004-31015.

Aznarez, I., Barash, Y., Shai, O., He, D., Zielenski, J., Tsui, L. C., Parkinson, J., Frey, B. J., Rommens, J. M. and Blencowe, B. J. (2008) A systematic analysis of intronic sequences downstream of 5' splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation. *Genome Res*, **18** (8), 1247-1258.

# B

Bachler, M., Schroeder, R. and von Ahsen, U. (1999) StreptoTag: a novel method for the isolation of RNA-binding proteins. *RNA*, **5** (11), 1509-1516.

Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, **226** (5252), 1209-1211.

Banks, J. D., Beemon, K. and Linial, M. L. (1997) RNA regulatory elements in the genomes of simple retroviruses. *Seminars in virology*, **8** (3), 194-204.

Banks, J. D. and Linial, M. L. (2000) Secondary structure analysis of a minimal avian leukosis-sarcoma virus packaging signal. *J Virol*, **74** (1), 456-464.

Bar, A., Marchand, V., Khoury, G., Dreumont, N., Mougin, A., Robas, N., Stevenin, J., Visvikis, A. and Branlant, C. (2011) Structural and functional analysis of the Rous Sarcoma virus negative regulator of splicing and demonstration of its activation by the 9G8 SR protein. *Nucleic Acids Res*, **39** (8), 3388-3403.

Barnard, R. J., Elleder, D. and Young, J. A. (2006) Avian sarcoma and leukosis virus-receptor interactions: from classical genetics to novel insights into virus-cell membrane fusion. *Virology*, **344** (1), 25-29.

Barr, S. D., Leipzig, J., Shinn, P., Ecker, J. R. and Bushman, F. D. (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J Virol*, **79** (18), 12035-12044.

Barreau, C., Paillard, L., Mereau, A. and Osborne, H. B. (2006) Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions.
*Biochimie*, **88** (5), 515-525.

Bartels, C., Klatt, C., Luhrmann, R. and Fabrizio, P. (2002) The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome. *EMBO Rep*, **3** (9), 875-880.

Bedford, M. T., Reed, R. and Leder, P. (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. *Proc Natl Acad Sci U S A*, **95** (18), 10602-10607.

Beemon, K. L. (2008) Retroviruses of birds. In The Encyclopedia of virology. 3rd edition. Elsevier, anonymous Oxford, 455-459.

Behrens, S. E., Galisson, F., Legrain, P. and Luhrmann, R. (1993) Evidence that the 60-kDa protein of 17S U2 small nuclear ribonucleoprotein is immunologically and functionally related to the yeast PRP9 splicing factor and is required for the efficient formation of prespliceosomes. *Proc Natl Acad Sci U S A*, **90** (17), 8229-8233.

Behzadnia, N., Golas, M. M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C. L., Urlaub, H., Stark, H., *et al.* (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *Embo J*, **26** (6), 1737-1748.

Behzadnia, N., Hartmuth, K., Will, C. L. and Luhrmann, R. (2006) Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. *RNA*, **12** (9), 1738-1746.

Benecke, H., Luhrmann, R. and Will, C. L. (2005) The U11/U12 snRNP 65K protein acts as a molecular bridge, binding the U12 snRNA and U11-59K protein. *Embo J*, **24** (17), 3057-3069.

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

Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G. and Reed, R. (1992) Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol Cell Biol*, **12** (7), 3165-3175.

Bennett, R. P., Nelle, T. D. and Wills, J. W. (1993) Functional chimeras of the Rous sarcoma virus and human immunodeficiency virus gag proteins. *J Virol*, **67** (11), 6487-6498.

Berberich, S. L., Macias, M., Zhang, L., Turek, L. P. and Stoltzfus, C. M. (1990) Comparison of Rous sarcoma virus RNA processing in chicken and mouse fibroblasts: evidence for double-spliced RNA in nonpermissive mouse cells. *J Virol*, **64** (9), 4313-4320.

Berberich, S. L. and Stoltzfus, C. M. (1991a) Analysis of spliced and unspliced Rous sarcoma virus RNAs early and late after infection of chicken embryo fibroblasts: effect of cell culture conditions. *Virology*, **182** (1), 135-144.

Berberich, S. L. and Stoltzfus, C. M. (1991b) Mutations in the regions of the Rous sarcoma virus 3' splice sites: implications for regulation of alternative splicing. *J Virol*, **65** (5), 2640-2646.

Berget, S. M. (1995) Exon recognition in vertebrate splicing. *J Biol Chem*, **270** (6), 2411-2414.

Berget, S. M., Moore, C. and Sharp, P. A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A*, **74** (8), 3171-3175.

Berglund, J. A., Abovich, N. and Rosbash, M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev*, **12** (6), 858-867.

Berglund, J. A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, **89** (5), 781-787.

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M. M., Sander, B., Will, C. L., Urlaub, H., Stark, H. and Luhrmann, R. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA*, **16** (12), 2384-2403.

Bilodeau, P. S., Domsic, J. K., Mayeda, A., Krainer, A. R. and Stoltzfus, C. M. (2001) RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by

binding of hnRNP A/B proteins to an exonic splicing silencer element. *J Virol*, **75** (18), 8487-8497.

Bird, G., Zorio, D. A. and Bentley, D. L. (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol*, **24** (20), 8963-8969.

Blanchette, M. and Chabot, B. (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *Embo J*, **18** (7), 1939-1952.

Blaustein, M., Pelisch, F., Tanos, T., Munoz, M. J., Wengier, D., Quadrana, L., Sanford, J. R., Muschietti, J. P., Kornblihtt, A. R., Caceres, J. F., *et al.* (2005) Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol*, **12** (12), 1037-1044.

Blencowe, B. J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci*, **25** (3), 106-110.

Blencowe, B. J. (2006) Alternative splicing: new insights from global analyses. *Cell*, **126** (1), 37-47.

Blencowe, B. J., Bauren, G., Eldridge, A. G., Issner, R., Nickerson, J. A., Rosonina, E. and Sharp, P. A. (2000) The SRm160/300 splicing coactivator subunits. *RNA*, **6** (1), 111-120.

Blencowe, B. J., Sproat, B. S., Ryder, U., Barabino, S. and Lamond, A. I. (1989) Antisense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. *Cell*, **59** (3), 531-539.

Bleoo, S., Sun, X., Hendzel, M. J., Rowe, J. M., Packer, M. and Godbout, R. (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), 3046-3059.

Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M. and McKeown, M. (1987) Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. *Cell*, **50** (5), 739-747.

Boon, K. L., Grainger, R. J., Ehsani, P., Barrass, J. D., Auchynnikava, T., Inglehearn, C. F. and Beggs, J. D. (2007) prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nat Struct Mol Biol*, **14** (11), 1077-1083.

Boon, K. L., Norman, C. M., Grainger, R. J., Newman, A. J. and Beggs, J. D. (2006) Prp8p dissection reveals domain structure and protein interaction sites. *RNA*, **12** (2), 198-205.

Bouck, J., Fu, X. D., Skalka, A. M. and Katz, R. A. (1995) Genetic selection for balanced retroviral splicing: novel regulation involving the second step can be mediated by transitions in the polypyrimidine tract. *Mol Cell Biol*, **15** (5), 2663-2671.

Bouck, J., Fu, X. D., Skalka, A. M. and Katz, R. A. (1998) Role of the constitutive splicing factors U2AF65 and SAP49 in suboptimal RNA splicing of novel retroviral mutants. *J Biol Chem*, **273** (24), 15169-15176.

Boukis, L. A., Liu, N., Furuyama, S. and Bruzik, J. P. (2004) Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. *J Biol Chem*, **279** (28), 29647-29653.

Bourgeois, C. F., Lejeune, F. and Stevenin, J. (2004) Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog Nucleic Acid Res Mol Biol*, **78**, 37-88.

Bourgeois, C. F., Popielarz, M., Hildwein, G. and Stevenin, J. (1999) Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol Cell Biol*, **19** (11), 7347-7356.

Bowers, W. J., Baglia, L. A. and Ruddel, A. (1996) Regulation of avian leukosis virus long terminal repeat-enhanced transcription by C/EBP-Rel interactions. *J Virol*, **70** (5), 3051-3059.

Bres, V., Gomes, N., Pickle, L. and Jones, K. A. (2005) A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. *Genes Dev*, **19** (10), 1211-1226.

Brosi, R., Groning, K., Behrens, S. E., Luhrmann, R. and Kramer, A. (1993) Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science*, **262** (5130), 102-105.

Brugge, J. S. and Erikson, R. L. (1977) Identification of a transformation-specific antigen induced by an avian sarcoma virus. *Nature*, **269** (5626), 346-348.

Bullock, W. O., Fernandez, J. M. and Short, J. M. (1987) XL1-blue: A high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *Bio. Techniques*, **5**, 376.

Buratowski, S. (2009) Progression through the RNA polymerase II CTD cycle. *Mol Cell*, **36** (4), 541-546.

Buratti, E. and Baralle, D. (2010) Novel roles of U1 snRNP in alternative splicing regulation. *RNA Biol*, **7** (4), 412-419.

Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y. M. and Baralle, F. E. (2004a) hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res*, **32** (14), 4224-4236.

Buratti, E., Dhir, A., Lewandowska, M. A. and Baralle, F. E. (2007) RNA structure is a key regulatory element in pathological ATM and CFTR pseudoexon inclusion events. *Nucleic Acids Res*, **35** (13), 4369-4383.

Buratti, E., Muro, A. F., Giombi, M., Gherbassi, D., Iaconcig, A. and Baralle, F. E. (2004b) 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), 1387-1400.

Burckin, T., Nagel, R., Mandel-Gutfreund, Y., Shiue, L., Clark, T. A., Chong, J. L., Chang, T. H., Squazzo, S., Hartzog, G. and Ares, M., Jr. (2005) Exploring functional relationships between components of the gene expression machinery. *Nat Struct Mol Biol*, **12** (2), 175-182.

Burd, C. G. and Dreyfuss, G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *Embo J*, **13** (5), 1197-1204.

Burge, C. B., Padgett, R. A. and Sharp, P. A. (1998) Evolutionary fates and origins of U12-type introns. *Mol Cell*, **2** (6), 773-785.

Burgess, S., Couto, J. R. and Guthrie, C. (1990) A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell*, **60** (5), 705-717.

Burgess, S. M. and Guthrie, C. (1993) Beat the clock: paradigms for NTPases in the maintenance of biological fidelity. *Trends Biochem Sci*, **18** (10), 381-384.

Burgess, S. M. and Guthrie, C. (1993b) A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell*, **73** (7), 1377-1391.

Busch, A. and Hertel, K. J. (2011) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA*.

Butterfield-Gerson, K. L., Scheifele, L. Z., Ryan, E. P., Hopper, A. K. and Parent, L. J. (2006) Importin-beta family members mediate alpharetrovirus gag nuclear entry via interactions with matrix and nucleocapsid. *J Virol*, **80** (4), 1798-1806.

# C


Cabello-Villegas, J., Giles, K. E., Soto, A. M., Yu, P., Mougin, A., Beemon, K. L. and Wang, Y. X. (2004) Solution structure of the pseudo-5' splice site of a retroviral splicing suppressor. *RNA*, **10** (9), 1388-1398.

Caceres, J. F. and Krainer, A. R. (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *Embo J*, **12** (12), 4715-4726.

Caceres, J. F., Stamm, S., Helfman, D. M. and Krainer, A. R. (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science*, **265** (5179), 1706-1709.

Callahan, E. M. and Wills, J. W. (2000) Repositioning basic residues in the M domain of the Rous sarcoma virus gag protein. *J Virol*, **74** (23), 11222-11229.

Callahan, E. M. and Wills, J. W. (2003) Link between genome packaging and rate of budding for Rous sarcoma virus. *J Virol*, **77** (17), 9388-9398.

Camats, M., Guil, S., Kokolo, M. and Bach-Elias, M. (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.

Campbell, S. and Vogt, V. M. (1997) In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. *J Virol*, **71** (6), 4425-4435.

Cao, W. and Garcia-Blanco, M. A. (1998) A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding. *J Biol Chem*, **273** (32), 20629-20635.

Cao, W., Jamison, S. F. and Garcia-Blanco, M. A. (1997) Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *RNA*, **3** (12), 1456-1467.

Caputi, M., Mayeda, A., Krainer, A. R. and Zahler, A. M. (1999) hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *Embo J*, **18** (14), 4060-4067.

Caputi, M. and Zahler, A. M. (2001) Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem*, **276** (47), 43850-43859.

Cartegni, L., Chew, S. L. and Krainer, A. R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet*, **3** (4), 285-298.

Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S. and Biamonti, G. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol*, **259** (3), 337-348.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. and Krainer, A. R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*, **31** (13), 3568-3571.

Carty, S. M., Goldstrohm, A. C., Sune, C., Garcia-Blanco, M. A. and Greenleaf, A. L. (2000) Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proc Natl Acad Sci U S A*, **97** (16), 9015-9020.

Cavaloc, Y., Bourgeois, C. F., Kister, L. and Stevenin, J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA*, **5** (3), 468-483.

Cavaloc, Y., Popielarz, M., Fuchs, J. P., Gattoni, R. and Stevenin, J. (1994) Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. *Embo J*, **13** (11), 2639-2649.

Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A. R. and Caceres, J. F. (2002) Nuclear export and retention signals in the RS domain of SR proteins. *Mol Cell Biol*, **22** (19), 6871-6882.

Chabot, B., Blanchette, M., Lapierre, I. and La Branche, H. (1997) An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol Cell Biol*, **17** (4), 1776-1786.

Chabot, B., LeBel, C., Hutchison, S., Nasim, F. H. and Simard, M. J. (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. and Black, D. L. (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), 4667-4676.

Chan, S. P. and Cheng, S. C. (2005) The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J Biol Chem*, **280** (35), 31190-31199.

Chan, S. P., Kao, D. I., Tsai, W. Y. and Cheng, S. C. (2003) The Prp19p-associated complex in spliceosome activation. *Science*, **302** (5643), 279-282.

Chandler, S. D., Mayeda, A., Yeakley, J. M., Krainer, A. R. and Fu, X. D. (1997) RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc Natl Acad Sci U S A*, **94** (8), 3596-3601.

Chang, L. J. and Stoltzfus, C. M. (1985) Cloning and nucleotide sequences of cDNAs spanning the splice junctions of Rous sarcoma virus mRNAs. *J Virol*, **53** (3), 969-972.

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

Charron, C., Manival, X., Charpentier, B., Branlant, C. and Aubry, A. (2004a) Purification, crystallization and preliminary X-ray diffraction data of L7Ae sRNP core protein from Pyrococcus abyssii. *Acta Crystallogr D Biol Crystallogr*, **60** (Pt 1), 122-124.

Charron, C., Manival, X., Clery, A., Senty-Segault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C. and Aubry, A. (2004b) The archaeal sRNA binding protein L7Ae has a 3D structure very similar to that of its eukaryal counterpart while having a broader RNA-binding specificity. *J Mol Biol*, **342** (3), 757-773.

Chen, C. D., Kobayashi, R. and Helfman, D. M. (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev*, **13** (5), 593-606.

Chen, H. C., Lin, W. C., Tsay, Y. G., Lee, S. C. and Chang, C. J. (2002) An RNA helicase, DDX1, interacting with poly(A) RNA and heterogeneous nuclear ribonucleoprotein K. *J Biol Chem*, **277** (43), 40403-40409.

Chen, J. Y., Stands, L., Staley, J. P., Jackups, R. R., Jr., Latus, L. J. and Chang, T. H. (2001) Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol Cell*, **7** (1), 227-232.

Chen, M. and Manley, J. L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*, **10** (11), 741-754.

Cheng, H., Dufu, K., Lee, C. S., Hsu, J. L., Dias, A. and Reed, R. (2006) Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, **127** (7), 1389-1400.

Chiara, M. D., Palandjian, L., Feld Kramer, R. and Reed, R. (1997) Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. *Embo J*, **16** (15), 4746-4759.

Cho, J., Hamasaki, K. and Rando, R. R. (1998) The binding site of a specific aminoglycoside binding RNA molecule. *Biochemistry*, **37** (14), 4985-4992.

Cho, S., Hoang, A., Chakrabarti, S., Huynh, N., Huang, D. B. and Ghosh, G. (2011a) The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRMs. *Nucleic Acids Res*.

Cho, S., Hoang, A., Sinha, R., Zhong, X. Y., Fu, X. D., Krainer, A. R. and Ghosh, G. (2011b) Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc Natl Acad Sci U S A*, **108** (20), 8233-8238.

Clery, A., Senty-Segault, V., Leclerc, F., Raue, H. A. and Branlant, C. (2007) Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as the recognition site for the Snu13p-Rrp9p protein pair. *Mol Cell Biol*, **27** (4), 1191-1206.

Coffin, J. M., Hughes, S. H. and Varmus, H. E. (1997) Retroviral virions and genomes. In Retroviruses. Cold Spring Harbor Laboratory Press, New-York, 27-69.

Cok, S. J., Acton, S. J., Sexton, A. E. and Morrison, A. R. (2004) Identification of RNA-binding proteins in RAW 264.7 cells that recognize a lipopolysaccharide-responsive element in the 3-untranslated region of the murine cyclooxygenase-2 mRNA. *J Biol Chem*, **279** (9), 8196-8205.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C. and Duncan, P. I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *Embo J*, **15** (2), 265-275.

Company, M., Arenas, J. and Abelson, J. (1991) Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature*, **349** (6309), 487-493.

consortium, I. C. G. S. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432** (7018), 695-716.

Convery, M. A., Rowsell, S., Stonehouse, N. J., Ellington, A. D., Hirao, I., Murray, J. B., Peabody, D. S., Phillips, S. E. and Stockley, P. G. (1998) Crystal structure of an RNA aptamer-protein complex at 2.8 A resolution. *Nat Struct Biol*, **5** (2), 133-139.

Cook, C. R. and McNally, M. T. (1996) Characterization of an RNP complex that assembles on the Rous sarcoma virus negative regulator of splicing element. *Nucleic Acids Res*, **24** (24), 4962-4968.

Cook, C. R. and McNally, M. T. (1998) SR protein and snRNP requirements for assembly of the Rous sarcoma virus negative regulator of splicing complex in vitro. *Virology*, **242** (1), 211-220.

Cook, C. R. and McNally, M. T. (1999) Interaction between the negative regulator of splicing element and a 3' splice site: requirement for U1 small nuclear ribonucleoprotein and the 3' splice site branch point/pyrimidine tract. *J Virol*, **73** (3), 2394-2400.

Cooke, C., Hans, H. and Alwine, J. C. (1999) Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol Cell Biol*, **19** (7), 4971-4979.

Cooper, J. A., Gould, K. L., Cartwright, C. A. and Hunter, T. (1986) Tyr527 is phosphorylated in pp60c-src: implications for regulation. *Science*, **231** (4744), 1431-1434.

Corioni, M., Antih, N., Tanackovic, G., Zavolan, M. and Kramer, A. (2011) Analysis of in situ pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Res*, **39** (5), 1868-1879.

Cote, J., Beaudoin, J., Tacke, R. and Chabot, B. (1995) The U1 small nuclear ribonucleoprotein/5' splice site interaction affects U2AF65 binding to the downstream 3' splice site. *J Biol Chem*, **270** (8), 4031-4036.

Crawford, J. B. and Patton, J. G. (2006) Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol Cell Biol*, **26** (23), 8791-8802.

Cullen, B. R., Raymond, K. and Ju, G. (1985a) Functional analysis of the transcription control region located within the avian retroviral long terminal repeat. *Mol Cell Biol*, **5** (3), 438-447.

Cullen, B. R., Raymond, K. and Ju, G. (1985b) Transcriptional activity of avian retroviral long terminal repeats directly correlates with enhancer activity. *J Virol*, **53** (2), 515-521.

Curristin, S. M., Bird, K. J., Tubbs, R. J. and Ruddell, A. (1997) VBP and RelA regulate avian leukosis virus long terminal repeat-enhanced transcription in B cells. *J Virol*, **71** (8), 5972-5981.

# D

Dalton, A. K., Murray, P. S., Murray, D. and Vogt, V. M. (2005) Biochemical characterization of rous sarcoma virus MA protein interaction with membranes. *J Virol*, **79** (10), 6227-6238.

Damgaard, C. K., Tange, T. O. and Kjems, J. (2002) hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA*, **8** (11), 1401-1415.

Damianov, A., Schreiner, S. and Bindereif, A. (2004) Recycling of the U12-type spliceosome requires p110, a component of the U6atac snRNP. *Mol Cell Biol*, **24** (4), 1700-1708.

Damier, L., Domenjoud, L. and Branlant, C. (1997) The D1-A2 and D2-A2 pairs of splice sites from human immunodeficiency virus type 1 are highly efficient in vitro, in spite of an unusual branch site. *Biochem Biophys Res Commun*, **237** (1), 182-187.

Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M. and Reed, R. (2006) Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev*, **20** (9), 1100-1109.

Das, R., Yu, J., Zhang, Z., Gygi, M. P., Krainer, A. R., Gygi, S. P. and Reed, R. (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell*, **26** (6), 867-881.

Das, R., Zhou, Z. and Reed, R. (2000) Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell*, **5** (5), 779-787.

David, C. J., Boyne, A. R., Millhouse, S. R. and Manley, J. L. (2011) The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev*, **25** (9), 972-983.

Davila Lopez, M., Rosenblad, M. A. and Samuelsson, T. (2008) Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. *Nucleic Acids Res*, **36** (9), 3001-3010.

de la Mata, M. and Kornblihtt, A. R. (2006) RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol*, **13** (11), 973-980.

de la Mata, M., Lafaille, C. and Kornblihtt, A. R. (2010) First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA*, **16** (5), 904-912.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., Stark, H., Urlaub, H. and Luhrmann, R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol*, **26** (14), 5528-5543.

Deffaud, C. and Darlix, J. L. (2000) Rous sarcoma virus translation revisited: characterization of an internal ribosome entry segment in the 5' leader of the genomic RNA. *J Virol*, **74** (24), 11581-11588.

Del Gatto, F., Gesnel, M. C. and Breathnach, R. (1996) The exon sequence TAGG can inhibit splicing. *Nucleic Acids Res*, **24** (11), 2017-2021.

Del Gatto-Konczak, F., Bourgeois, C. F., Le Guiner, C., Kister, L., Gesnel, M. C., Stevenin, J. and Breathnach, R. (2000) The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. *Mol Cell Biol*, **20** (17), 6287-6299.

Dember, L. M., Kim, N. D., Liu, K. Q. and Anderson, P. (1996) Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. *J Biol Chem*, **271** (5), 2783-2788.

Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W. and Barabino, S. M. (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), 35788-35797.

Deutsch, M. and Long, M. (1999) Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res*, **27** (15), 3219-3228.

Dhir, A. and Buratti, E. (2010) Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *Febs J*, **277** (4), 841-855.

Dhir, A., Buratti, E., van Santen, M. A., Luhrmann, R. and Baralle, F. E. (2010) The intronic splicing code: multiple factors involved in ATM pseudoexon definition. *Embo J*, **29** (4), 749-760.

Dietrich, R. C., Incorvaia, R. and Padgett, R. A. (1997) Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol Cell*, **1** (1), 151-160.

Dietrich, R. C., Padgett, R. A. and Shukla, G. C. (2009) The conserved 3' end domain of U6atac snRNA can direct U6 snRNA to the minor spliceosome. *RNA*, **15** (6), 1198-1207.

Dilley, K. A., Gregory, D., Johnson, M. C. and Vogt, V. M. (2010) An LYPSL late domain in the gag protein contributes to the efficient release and replication of Rous sarcoma virus. *J Virol*, **84** (13), 6276-6287.

Dix, I., Russell, C. S., O'Keefe, R. T., Newman, A. J. and Beggs, J. D. (1998) Protein-RNA interactions in the U5 snRNP of Saccharomyces cerevisiae. *RNA*, **4** (12), 1675-1686.

Dominguez, C. and Allain, F. H. (2006) NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Res*, **34** (13), 3634-3645.

Dominguez, C., Fisette, J. F., Chabot, B. and Allain, F. H. (2010) Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat Struct Mol Biol*, **17** (7), 853-861.

Domsic, J. K., Wang, Y., Mayeda, A., Krainer, A. R. and Stoltzfus, C. M. (2003) Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. *Mol Cell Biol*, **23** (23), 8762-8772.

Donmez, G., Hartmuth, K., Kastner, B., Will, C. L. and Luhrmann, R. (2007) The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol Cell*, **25** (3), 399-411.

Donmez, G., Hartmuth, K. and Luhrmann, R. (2004) Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA*, **10** (12), 1925-1933.

Donze, O., Damay, P. and Spahr, P. F. (1995) The first and third uORFs in RSV leader RNA are efficiently translated: implications for translational regulation and viral RNA packaging. *Nucleic Acids Res*, **23** (5), 861-868.

Donze, O. and Spahr, P. F. (1992) Role of the open reading frames of Rous sarcoma virus leader RNA in translation and genome packaging. *Embo J*, **11** (10), 3747-3757.

Dorai, T., Levy, J. B., Kang, L., Brugge, J. S. and Wang, L. H. (1991) Analysis of cDNAs of the proto-oncogene c-src: heterogeneity in 5' exons and possible mechanism for the genesis of the 3' end of v-src. *Mol Cell Biol*, **11** (8), 4165-4176.

Dreumont, N., Hardy, S., Behm-Ansmant, I., Kister, L., Branlant, C., Stevenin, J. and Bourgeois, C. F. (2010) Antagonistic factors control the unproductive splicing of SC35 terminal intron. *Nucleic Acids Res*, **38** (4), 1353-1366.

Du, H. and Rosbash, M. (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature*, **419** (6902), 86-90.

Duesberg, P. H. and Vogt, P. K. (1970) Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses. *Proc Natl Acad Sci U S A*, **67** (4), 1673-1680.

Dybkov, O., Will, C. L., Deckert, J., Behzadnia, N., Hartmuth, K. and Luhrmann, R. (2006) U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. *Mol Cell Biol*, **26** (7), 2803-2816.

Dye, B. T., Buvoli, M., Mayer, S. A., Lin, C. H. and Patton, J. G. (1998) Enhancer elements activate the weak 3' splice site of alpha-tropomyosin exon 2. *RNA*, **4** (12), 1523-1536.

# E

Edery, P., Marcaillou, C., Sahbatou, M., Labalme, A., Chastang, J., Touraine, R., Tubacher, E., Senni, F., Bober, M. B., Nampoothiri, S., *et al.* (2011) Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science*, **332** (6026), 240-243.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P. and Ehresmann, B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res*, **15** (22), 9109-9128.

Eldridge, A. G., Li, Y., Sharp, P. A. and Blencowe, B. J. (1999) The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc Natl Acad Sci U S A*, **96** (11), 6125-6130.

Elleder, D., Stepanets, V., Melder, D. C., Senigl, F., Geryk, J., Pajer, P., Plachy, J., Hejnar, J., Svoboda, J. and Federspiel, M. J. (2005) The receptor for the subgroup C avian sarcoma and leukosis viruses, Tvc, is related to mammalian butyrophilins, members of the immunoglobulin superfamily. *J Virol*, **79** (16), 10408-10419.

Ellis, P. D., Smith, C. W. and Kemp, P. (2004) Regulated tissue-specific alternative splicing of enhanced green fluorescent protein transgenes conferred by alpha-tropomyosin regulatory elements in transgenic mice. *J Biol Chem*, **279** (35), 36660-36669.

Emeson, R. B., Hedjran, F., Yeakley, J. M., Guise, J. W. and Rosenfeld, M. G. (1989) Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor. *Nature*, **341** (6237), 76-80.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P. W., Kobayashi, R., Blencowe, B. J. and Ingles, C. J. (2002) Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *RNA*, **8** (9), 1102-1111.

Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A. and Krainer, A. R. (1993) Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *Embo J*, **12** (9), 3607-3617.

Eperon, I. C., Makarova, O. V., Mayeda, A., Munroe, S. H., Caceres, J. F., Hayward, D. G. and Krainer, A. R. (2000) Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol*, **20** (22), 8303-8318.

Erkmann, J. A. and Kutay, U. (2004) Nuclear export of mRNA: from the site of transcription to the cytoplasm. *Exp Cell Res*, **296** (1), 12-20.

Exline, C. M., Feng, Z. and Stoltzfus, C. M. (2008) Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. *J Virol*, **82** (8), 3921-3931.

Expert-Bezancon, A., Sureau, A., Durosay, P., Salesse, R., Groeneveld, H., Lecaer, J. P. and Marie, J. (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), 38249-38259.

# F

Fadly, A. M. (2000) Isolation and identification of avian leukosis viruses: a review. *Avian Pathol*, **29** (6), 529-535.

Fadly, A. M. (2008) Neoplastic diseases. In Diseases of Poultry, 12th Edition. Wiley-Blackwell, 405-516.

Fairbrother, W. G. and Chasin, L. A. (2000) Human genomic sequences that inhibit splicing. *Mol Cell Biol*, **20** (18), 6816-6825.

Fairbrother, W. G., Yeo, G. W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P. A. and Burge, C. B. (2004) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res*, **32** (Web Server issue), W187-190.

Feng, Y., Chen, M. and Manley, J. L. (2008) Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nat Struct Mol Biol*, **15** (10), 1040-1048.

Fetzer, S., Lauber, J., Will, C. L. and Luhrmann, R. (1997) The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase. *RNA*, **3** (4), 344-355.

Ficht, T. A., Chang, L. J. and Stoltzfus, C. M. (1984) Avian sarcoma virus gag and env gene structural protein precursors contain a common amino-terminal sequence. *Proc Natl Acad Sci U S A*, **81** (2), 362-366.

Fisette, J. F., Toutant, J., Dugre-Brisson, S., Desgroseillers, L. and Chabot, B. (2010) hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection. *RNA*, **16** (1), 228-238.

Fleckner, J., Zhang, M., Valcarcel, J. and Green, M. R. (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev*, **11** (14), 1864-1872.

Fogel, B. L., McNally, L. M. and McNally, M. T. (2002) Efficient polyadenylation of Rous sarcoma virus RNA requires the negative regulator of splicing element. *Nucleic Acids Res*, **30** (3), 810-817.

Fogel, B. L. and McNally, M. T. (2000) A cellular protein, hnRNP H, binds to the negative regulator of splicing element from Rous sarcoma virus. *J Biol Chem*, **275** (41), 32371-32378.

Fong, Y. W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, **414** (6866), 929-933.

Forch, P., Merendino, L., Martinez, C. and Valcarcel, J. (2001) Modulation of msl-2 5' splice site recognition by Sex-lethal. *RNA*, **7** (9), 1185-1191.

Forch, P., Puig, O., Martinez, C., Seraphin, B. and Valcarcel, J. (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *Embo J*, **21** (24), 6882-6892.

Fox-Walsh, K. L., Dou, Y., Lam, B. J., Hung, S. P., Baldi, P. F. and Hertel, K. J. (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci U S A*, **102** (45), 16176-16181.

Frazer, L. N., Lovell, S. C. and O'Keefe, R. T. (2009) Analysis of synthetic lethality reveals genetic interactions between the GTPase Snu114p and snRNAs in the catalytic core of the Saccharomyces cerevisiae spliceosome. *Genetics*, **183** (2), 497-515-491SI-494SI.

Freund, M., Hicks, M. J., Konermann, C., Otte, M., Hertel, K. J. and Schaal, H. (2005) Extended base pair complementarity between U1 snRNA and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Res*, **33** (16), 5112-5119.

Frilander, M. J. and Steitz, J. A. (1999) Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. *Genes Dev*, **13** (7), 851-863.

Fu, X. D., Katz, R. A., Skalka, A. M. and Maniatis, T. (1991) The role of branchpoint and 3'-exon sequences in the control of balanced splicing of avian retrovirus RNA. *Genes Dev*, **5** (2), 211-220.

Fu, X. D. and Maniatis, T. (1992) The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc Natl Acad Sci U S A*, **89** (5), 1725-1729.

# G

Gallo, J. M. and Spickett, C. (2010) The role of CELF proteins in neurological disorders. *RNA Biol*, **7** (4), 474-479.

Garbitt, R. A., Bone, K. R. and Parent, L. J. (2004) Insertion of a classical nuclear import signal into the matrix domain of the Rous sarcoma virus Gag protein interferes with virus replication. *J Virol*, **78** (24), 13534-13542.

Garbitt-Hirst, R., Kenney, S. P. and Parent, L. J. (2009) Genetic evidence for a connection between Rous sarcoma virus gag nuclear trafficking and genomic RNA packaging. *J Virol*, **83** (13), 6790-6797.

Giles, K. E. and Beemon, K. L. (2005) Retroviral splicing suppressor sequesters a 3' splice site in a 50S aberrant splicing complex. *Mol Cell Biol*, **25** (11), 4397-4405.

Giles, K. E., Caputi, M. and Beemon, K. L. (2004) Packaging and reverse transcription of snRNAs by retroviruses may generate pseudogenes. *RNA*, **10** (2), 299-307.

Golas, M. M., Sander, B., Bessonov, S., Grote, M., Wolf, E., Kastner, B., Stark, H. and Luhrmann, R. (2010) 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Mol Cell*, **40** (6), 927-938.

Golas, M. M., Sander, B., Will, C. L., Luhrmann, R. and Stark, H. (2005) Major conformational change in the complex SF3b upon integration into the spliceosomal

U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, **17** (6), 869-883.

Goldstrohm, A. C., Albrecht, T. R., Sune, C., Bedford, M. T. and Garcia-Blanco, M. A. (2001) The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol Cell Biol*, **21** (22), 7617-7628.

Gontarek, R. R., McNally, M. T. and Beemon, K. (1993) Mutation of an RSV intronic element abolishes both U11/U12 snRNP binding and negative regulation of splicing. *Genes Dev*, **7** (10), 1926-1936.

Gooding, C., Roberts, G. C. and Smith, C. W. (1998) Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. *RNA*, **4** (1), 85-100.

Gooding, C. and Smith, C. W. (2008) Tropomyosin exons as models for alternative splicing. *Adv Exp Med Biol*, **644**, 27-42.

Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I. and Howard, B. H. (1982) The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. *Proc Natl Acad Sci U S A*, **79** (22), 6777-6781.

Gottschalk, A., Kastner, B., Luhrmann, R. and Fabrizio, P. (2001) The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. *RNA*, **7** (11), 1554-1565.

Gozani, O., Feld, R. and Reed, R. (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev*, **10** (2), 233-243.

Gozani, O., Potashkin, J. and Reed, R. (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol*, **18** (8), 4752-4760.

Grabowski, P. J. and Sharp, P. A. (1986) Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science*, **233** (4770), 1294-1299.

Grainger, R. J. and Beggs, J. D. (2005) Prp8 protein: at the heart of the spliceosome. *RNA*, **11** (5), 533-557.

Grandgenett, D., Quinn, T., Hippenmeyer, P. J. and Oroszlan, S. (1985) Structural characterization of the avian retrovirus reverse transcriptase and endonuclease domains. *J Biol Chem*, **260** (14), 8243-8249.

Graveley, B. R., Hertel, K. J. and Maniatis, T. (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *Embo J*, **17** (22), 6747-6756.

Graveley, B. R., Hertel, K. J. and Maniatis, T. (2001) The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA*, **7** (6), 806-818.

Greger, J. G., Katz, R. A., Taganov, K., Rall, G. F. and Skalka, A. M. (2004) Transduction of terminally differentiated neurons by avian sarcoma virus. *J Virol*, **78** (9), 4902-4906.

Gromak, N., Rideau, A., Southby, J., Scadden, A. D., Gooding, C., Huttelmaier, S., Singer, R. H. and Smith, C. W. (2003) The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing. *Embo J*, **22** (23), 6356-6364.

Gromak, N. and Smith, C. W. (2002) A splicing silencer that regulates smooth muscle specific alternative splicing is active in multiple cell types. *Nucleic Acids Res*, **30** (16), 3548-3557.

Gudleski, N., Flanagan, J. M., Ryan, E. P., Bewley, M. C. and Parent, L. J. (2010) Directionality of nucleocytoplasmic transport of the retroviral gag protein depends on sequential binding of karyopherins and viral RNA. *Proc Natl Acad Sci U S A*, **107** (20), 9358-9363.

Gui, J. F., Tronchere, H., Chandler, S. D. and Fu, X. D. (1994) Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc Natl Acad Sci U S A*, **91** (23), 10824-10828.

Guil, S., Gattoni, R., Carrascal, M., Abian, J., Stevenin, J. and Bach-Elias, M. (2003) Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol Cell Biol*, **23** (8), 2927-2941.

Guntaka, R. V. (1993) Transcription termination and polyadenylation in retroviruses. *Microbiol Rev*, **57** (3), 511-521.

Guo, W., Winistorfer, S. C. and Stoltzfus, C. M. (2000) Selective inhibition of splicing at the avian sarcoma virus src 3' splice site by direct-repeat posttranscriptional cis elements. *J Virol*, **74** (18), 8513-8523.

Guth, S., Martinez, C., Gaur, R. K. and Valcarcel, J. (1999) Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Mol Cell Biol*, **19** (12), 8263-8271.

# H

Hacker, I., Sander, B., Golas, M. M., Wolf, E., Karagoz, E., Kastner, B., Stark, H., Fabrizio, P. and Luhrmann, R. (2008) Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat Struct Mol Biol*, **15** (11), 1206-1212.

Hackett, P. B., Dalton, M. W., Johnson, D. P. and Petersen, R. B. (1991) Phylogenetic and physical analysis of the 5' leader RNA sequences of avian retroviruses. *Nucleic Acids Res*, **19** (24), 6929-6934.

Hackett, P. B., Swanstrom, R., Varmus, H. E. and Bishop, J. M. (1982) The leader sequence of the subgenomic mRNA's of Rous sarcoma virus is approximately 390 nucleotides. *J Virol*, **41** (2), 527-534.

Hallay, H., Locker, N., Ayadi, L., Ropers, D., Guittet, E. and Branlant, C. (2006) Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site. *J Biol Chem*, **281** (48), 37159-37174.

Hamasaki, K., Killian, J., Cho, J. and Rando, R. R. (1998) Minimal RNA constructs that specifically bind aminoglycoside antibiotics with high affinities. *Biochemistry*, **37** (2), 656-663.

Han, K., Yeo, G., An, P., Burge, C. B. and Grabowski, P. J. (2005) A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol*, **3** (5), e158.

Han, S. P., Tang, Y. H. and Smith, R. (2010) Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J*, **430** (3), 379-392.

Hanafusa, H., Hanafusa, T. and Rubin, H. (1963) The defectiveness of Rous sarcoma virus. *Proc Natl Acad Sci U S A*, **49**, 572-580.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, **166** (4), 557-580.

Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., Jr. and Krainer, A. R. (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA*, **4** (4), 430-444.

Hargous, Y., Hautbergue, G. M., Tintaru, A. M., Skrisovska, L., Golovanov, A. P., Stevenin, J., Lian, L. Y., Wilson, S. A. and Allain, F. H. (2006) Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *Embo J*, **25** (21), 5126-5137.

Hartmuth, K., Urlaub, H., Vornlocher, H. P., Will, C. L., Gentzel, M., Wilm, M. and Luhrmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A*, **99** (26), 16719-16724.

Hartmuth, K., Vornlocher, H. P. and Luhrmann, R. (2004) Tobramycin affinity tag purification of spliceosomes. *Methods Mol Biol*, **257**, 47-64.

Hatai, H., Ochiai, K., Nagakura, K., Imanishi, S., Ochi, A., Kozakura, R., Ono, M., Goryo, M., Ohashi, K. and Umemura, T. (2008) A recombinant avian leukosis virus associated with fowl glioma in layer chickens in Japan. *Avian Pathol*, **37** (2), 127-137.

Hatziioannou, T. and Goff, S. P. (2001) Infection of nondividing cells by Rous sarcoma virus. *J Virol*, **75** (19), 9526-9531.

Hayward, W. S., Neel, B. G. and Astrin, S. M. (1981) Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. *Nature*, **290** (5806), 475-480.

He, H., Liyanarachchi, S., Akagi, K., Nagy, R., Li, J., Dietrich, R. C., Li, W., Sebastian, N., Wen, B., Xin, B., *et al.* (2011) Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science*, **332** (6026), 238-240.

He, Y. and Smith, R. (2009) Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol Life Sci*, **66** (7), 1239-1256.

Heinrichs, V., Bach, M., Winkelmann, G. and Luhrmann, R. (1990) U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science*, **247** (4938), 69-72.

Henning, D., So, R. B., Jin, R., Lau, L. F. and Valdez, B. C. (2003) Silencing of RNA helicase II/Gualpha inhibits mammalian ribosomal RNA production. *J Biol Chem*, **278** (52), 52307-52314.

Herman, S. A. and Coffin, J. M. (1986) Differential transcription from the long terminal repeats of integrated avian leukosis virus DNA. *J Virol*, **60** (2), 497-505.

Hertel, K. J. and Maniatis, T. (1999) Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. *Proc Natl Acad Sci U S A*, **96** (6), 2651-2655.

Hibbert, C. S., Gontarek, R. R. and Beemon, K. L. (1999) The role of overlapping U1 and U11 5' splice site sequences in a negative regulator of splicing. *RNA*, **5** (3), 333-343.

Hicks, M. J., Mueller, W. F., Shepard, P. J. and Hertel, K. J. (2010) Competing upstream 5' splice sites enhance the rate of proximal splicing. *Mol Cell Biol*, **30** (8), 1878-1886.

Hiller, M., Zhang, Z., Backofen, R. and Stamm, S. (2007) Pre-mRNA secondary structures influence exon recognition. *PLoS Genet*, **3** (11), e204.

Hirose, Y., Tacke, R. and Manley, J. L. (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev*, **13** (10), 1234-1239.

Holmstrom, T. H., Mialon, A., Kallio, M., Nymalm, Y., Mannermaa, L., Holm, T., Johansson, H., Black, E., Gillespie, D., Salminen, T. A., *et al.* (2008) c-Jun supports ribosomal RNA processing and nucleolar localization of RNA helicase DDX21. *J Biol Chem*, **283** (11), 7046-7053.

Hong, W., Bennett, M., Xiao, Y., Feld Kramer, R., Wang, C. and Reed, R. (1997) Association of U2 snRNP with the spliceosomal complex E. *Nucleic Acids Res*, **25** (2), 354-361.

Honig, A., Auboeuf, D., Parker, M. M., O'Malley, B. W. and Berget, S. M. (2002) Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. *Mol Cell Biol*, **22** (16), 5698-5707.

Hope, T. J. (1999) The ins and outs of HIV Rev. *Arch Biochem Biophys*, **365** (2), 186-191.

Houtz, E. K. and Conklin, K. F. (1996) Identification of EFIV, a stable factor present in many avian cell types that transactivates sequences in the 5' portion of the Rous sarcoma virus long terminal repeat enhancer. *J Virol*, **70** (1), 393-401.

Huang, Y., Gattoni, R., Stevenin, J. and Steitz, J. A. (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell*, **11** (3), 837-843.

Huang, Y., Yario, T. A. and Steitz, J. A. (2004) A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci U S A*, **101** (26), 9666-9670.

Hughes, S., Mellstrom, K., Kosik, E., Tamanoi, F. and Brugge, J. (1984) Mutation of a termination codon affects src initiation. *Mol Cell Biol*, **4** (9), 1738-1746.

Hughes, S. H., Mutschler, A., Bishop, J. M. and Varmus, H. E. (1981) A Rous sarcoma virus provirus is flanked by short direct repeats of a cellular DNA sequence present in only one copy prior to integration. *Proc Natl Acad Sci U S A*, **78** (7), 4299-4303.

Humphries, E. H., Glover, C. and Reichmann, M. E. (1981) Rous sarcoma virus infection of synchronized cells establishes provirus integration during S-phase DNA synthesis prior to cellular division. *Proc Natl Acad Sci U S A*, **78** (4), 2601-2605.

Hutchison, S., LeBel, C., Blanchette, M. and Chabot, B. (2002) Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J Biol Chem*, **277** (33), 29745-29752.

# I

Isken, O., Kim, Y. K., Hosoda, N., Mayeur, G. L., Hershey, J. W. and Maquat, L. E. (2008) Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell*, **133** (2), 314-327.

Ismaili, N., Sha, M., Gustafson, E. H. and Konarska, M. M. (2001) The 100-kda U5 snRNP protein (hPrp28p) contacts the 5' splice site through its ATPase site. *RNA*, **7** (2), 182-193.

Izquierdo, J. M., Majos, N., Bonnal, S., Martinez, C., Castelo, R., Guigo, R., Bilbao, D. and Valcarcel, J. (2005) Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol Cell*, **19** (4), 475-484.

# J

Jackson, I. J. (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res*, **19** (14), 3795-3798.

Jacquenet, S., Mereau, A., Bilodeau, P. S., Damier, L., Stoltzfus, C. M. and Branlant, C. (2001b) 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), 40464-40475.

Jacquenet, S., Ropers, D., Bilodeau, P. S., Damier, L., Mougin, A., Stoltzfus, C. M. and Branlant, C. (2001) Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing. *Nucleic Acids Res*, **29** (2), 464-478.

Jamieson, D. J., Rahe, B., Pringle, J. and Beggs, J. D. (1991) A suppressor of a yeast splicing mutation (prp8-1) encodes a putative ATP-dependent RNA helicase. *Nature*, **349** (6311), 715-717.

Jamison, S. F., Pasman, Z., Wang, J., Will, C., Luhrmann, R., Manley, J. L. and Garcia-Blanco, M. A. (1995) U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. *Nucleic Acids Res*, **23** (16), 3260-3267.

Janknecht, R. (2010) Multi-talented DEAD-box proteins and potential tumor promoters: p68 RNA helicase (DDX5) and its paralog, p72 RNA helicase (DDX17). *Am J Transl Res*, **2** (3), 223-234.

Jiang, L. and Patel, D. J. (1998) Solution structure of the tobramycin-RNA aptamer complex. *Nat Struct Biol*, **5** (9), 769-774.

Jiang, L., Suri, A. K., Fiala, R. and Patel, D. J. (1997a) Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex. *Chem Biol*, **4** (1), 35-50.

Jiang, W., Kanter, M. R., Dunkel, I., Ramsay, R. G., Beemon, K. L. and Hayward, W. S. (1997b) Minimal truncation of the c-myb gene product in rapid-onset B-cell lymphoma. *J Virol*, **71** (9), 6526-6533.

Jiang, Z., Tang, H., Havlioglu, N., Zhang, X., Stamm, S., Yan, R. and Wu, J. Y. (2003) Mutations in tau gene exon 10 associated with FTDP-17 alter the activity of an exonic splicing enhancer to interact with Tra2 beta. *J Biol Chem*, **278** (21), 18997-19007.

Johnson, S. F. and Telesnitsky, A. (2010) Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS Pathog*, **6** (10).

Jurica, M. S. and Moore, M. J. (2002) Capturing splicing complexes to study structure and mechanism. *Methods*, **28** (3), 336-345.

Jurica, M. S. and Moore, M. J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, **12** (1), 5-14.

# K

Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L. and Dreyfuss, G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, **468** (7324), 664-668.

Kameoka, S., Duque, P. and Konarska, M. M. (2004) p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *Embo J*, **23** (8), 1782-1791.

Kammler, S., Otte, M., Hauber, I., Kjems, J., Hauber, J. and Schaal, H. (2006) The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology*, **3**, 89.

Kanter, M. R., Smith, R. E. and Hayward, W. S. (1988) Rapid induction of B-cell lymphomas: insertional activation of c-myb by avian leukosis virus. *J Virol*, **62** (4), 1423-1432.

Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S. and Wu, J. Y. (2011) RNA Helicase p68 (DDX5) Regulates tau Exon 10 Splicing by Modulating a Stem-Loop Structure at the 5' Splice Site. *Mol Cell Biol*, **31** (9), 1812-1821.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G. and Ohno, S. (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev*, **20** (3), 355-367.

Kashima, T. and Manley, J. L. (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat Genet*, **34** (4), 460-463.

Kataoka, N., Bachorik, J. L. and Dreyfuss, G. (1999) Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol*, **145** (6), 1145-1152.

Katz, R. A., Kotler, M. and Skalka, A. M. (1988) cis-acting intron mutations that affect the efficiency of avian retroviral RNA splicing: implication for mechanisms of control. *J Virol*, **62** (8), 2686-2695.

Katz, R. A. and Skalka, A. M. (1988) A C-terminal domain in the avian sarcoma-leukosis virus pol gene product is not essential for viral replication. *J Virol*, **62** (2), 528-533.

Katz, R. A. and Skalka, A. M. (1990) Control of retroviral RNA splicing through maintenance of suboptimal processing signals. *Mol Cell Biol*, **10** (2), 696-704.
Keller, P. W., Johnson, M. C. and Vogt, V. M. (2008) Mutations in the spacer peptide and adjoining sequences in Rous sarcoma virus Gag lead to tubular budding. *J Virol*, **82** (14), 6788-6797.
Kennedy, C. F., Kramer, A. and Berget, S. M. (1998) A role for SRp54 during intron bridging of small introns with pyrimidine tracts upstream of the branch point. *Mol Cell Biol*, **18** (9), 5425-5434.
Kenney, S. P., Lochmann, T. L., Schmid, C. L. and Parent, L. J. (2008) Intermolecular interactions between retroviral Gag proteins in the nucleus. *J Virol*, **82** (2), 683-691.
Kent, O. A. and MacMillan, A. M. (2002) Early organization of pre-mRNA during spliceosome assembly. *Nat Struct Biol*, **9** (8), 576-581.
Kent, O. A., Ritchie, D. B. and Macmillan, A. M. (2005) Characterization of a U2AF-independent commitment complex (E') in the mammalian spliceosome assembly pathway. *Mol Cell Biol*, **25** (1), 233-240.
Kershaw, C. J., Barrass, J. D., Beggs, J. D. and O'Keefe, R. T. (2009) Mutations in the U5 snRNA result in altered splicing of subsets of pre-mRNAs and reduced stability of Prp8. *RNA*, **15** (7), 1292-1304.
Kielkopf, C. L., Rodionova, N. A., Green, M. R. and Burley, S. K. (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell*, **106** (5), 595-605.
Kikonyogo, A., Bouamr, F., Vana, M. L., Xiang, Y., Aiyar, A., Carter, C. and Leis, J. (2001) Proteins related to the Nedd4 family of ubiquitin protein ligases interact with the L domain of Rous sarcoma virus and are required for gag budding from cells. *Proc Natl Acad Sci U S A*, **98** (20), 11199-11204.
Kiledjian, M. and Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *Embo J*, **11** (7), 2655-2664.
Kim, S. H. and Lin, R. J. (1996) Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol Cell Biol*, **16** (12), 6810-6819.
Kitamura, N., Kitamura, A., Toyoshima, K., Hirayama, Y. and Yoshida, M. (1982) Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. *Nature*, **297** (5863), 205-208.
Kohler, A. and Hurt, E. (2007) Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol*, **8** (10), 761-773.
Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A. and Manley, J. L. (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, **368** (6467), 119-124.
Kolossova, I. and Padgett, R. A. (1997) U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA*, **3** (3), 227-233.
Kosowski, T. R., Keys, H. R., Quan, T. K. and Ruby, S. W. (2009) DExD/H-box Prp5 protein is in the spliceosome during most of the splicing cycle. *RNA*, **15** (7), 1345-1362.
Krainer, A. R., Conway, G. C. and Kozak, D. (1990) The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell*, **62** (1), 35-42.
Kreager, K. S. (1998) Chicken industry strategies for control of tumor virus infections. *Poult Sci*, **77** (8), 1213-1216.
Krol, A., Gallinaro, H., Lazar, E., Jacob, M. and Branlant, C. (1981) The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes. *Nucleic Acids Res*, **9** (4), 769-787.

Kuroda, T., Murayama, A., Katagiri, N., Ohta, Y. M., Fujita, E., Masumoto, H., Ema, M., Takahashi, S., Kimura, K. and Yanagisawa, J. (2011) RNA content in the nucleolus alters p53 acetylation via MYBBP1A. *Embo J*, **30** (6), 1054-1066.

Kyburz, A., Friedlein, A., Langen, H. and Keller, W. (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.

# L

Labourier, E., Adams, M. D. and Rio, D. C. (2001) Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol Cell*, **8** (2), 363-373.

Labourier, E., Blanchette, M., Feiger, J. W., Adams, M. D. and Rio, D. C. (2002) The KH-type RNA-binding protein PSI is required for Drosophila viability, male fertility, and cellular mRNA processing. *Genes Dev*, **16** (1), 72-84.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227** (5259), 680-685.

Laggerbauer, B., Achsel, T. and Luhrmann, R. (1998) The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. *Proc Natl Acad Sci U S A*, **95** (8), 4188-4192.

Laggerbauer, B., Liu, S., Makarov, E., Vornlocher, H. P., Makarova, O., Ingelfinger, D., Achsel, T. and Luhrmann, R. (2005) The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation. *RNA*, **11** (5), 598-608.

Lai, M. C., Lee, Y. H. and Tarn, W. Y. (2008) The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control. *Mol Biol Cell*, **19** (9), 3847-3858.

Lai, M. C., Lin, R. I., Huang, S. Y., Tsai, C. W. and Tarn, W. Y. (2000) A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J Biol Chem*, **275** (11), 7950-7957.

Lai, M. C., Lin, R. I. and Tarn, W. Y. (2001) Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc Natl Acad Sci U S A*, **98** (18), 10154-10159.

Lamichhane, R., Daubner, G. M., Thomas-Crusells, J., Auweter, S. D., Manatschal, C., Austin, K. S., Valniuk, O., Allain, F. H. and Rueda, D. (2010) RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. *Proc Natl Acad Sci U S A*, **107** (9), 4105-4110.

Lardelli, R. M., Thompson, J. X., Yates, J. R., 3rd and Stevens, S. W. (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA*, **16** (3), 516-528.

Lavigueur, A., La Branche, H., Kornblihtt, A. R. and Chabot, B. (1993) A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes Dev*, **7** (12A), 2405-2417.

LeBlanc, J. J. and Beemon, K. L. (2004) Unspliced Rous sarcoma virus genomic RNAs are translated and subjected to nonsense-mediated mRNA decay before packaging. *J Virol*, **78** (10), 5139-5146.

LeBlanc, J. J., Uddowla, S., Abraham, B., Clatterbuck, S. and Beemon, K. L. (2007) Tap and Dbp5, but not Gag, are involved in DR-mediated nuclear export of unspliced Rous sarcoma virus RNA. *Virology*, **363** (2), 376-386.

LeCuyer, K. A., Behlen, L. S. and Uhlenbeck, O. C. (1995) Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. *Biochemistry*, **34** (33), 10600-10606.

Lee, C. G. (2002) RH70, a bidirectional RNA helicase, co-purifies with U1snRNP. *J Biol Chem*, **277** (42), 39679-39683.

Lee, Y. M. and Coffin, J. M. (1991) Relationship of avian retrovirus DNA synthesis to integration in vitro. *Mol Cell Biol*, **11** (3), 1419-1430.

Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. and Mattaj, I. W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev*, **10** (13), 1683-1698.

Lim, F. and Peabody, D. S. (1994) Mutations that increase the affinity of a translational repressor for RNA. *Nucleic Acids Res*, **22** (18), 3748-3752.

Lin, C., Yang, L., Yang, J. J., Huang, Y. and Liu, Z. R. (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), 7484-7493.

Lin, C. H. and Patton, J. G. (1995) Regulation of alternative 3' splice site selection by constitutive splicing factors. *RNA*, **1** (3), 234-245.

Lin, K. T., Lu, R. M. and Tarn, W. Y. (2004) The WW domain-containing proteins interact with the early spliceosome and participate in pre-mRNA splicing in vivo. *Mol Cell Biol*, **24** (20), 9176-9185.

Little, J. T. and Jurica, M. S. (2008) Splicing factor SPF30 bridges an interaction between the prespliceosome protein U2AF35 and tri-small nuclear ribonucleoprotein protein hPrp3. *J Biol Chem*, **283** (13), 8145-8152.

Liu, F. and Gong, C. X. (2008) Tau exon 10 alternative splicing and tauopathies. *Mol Neurodegener*, **3**, 8.

Liu, H. X., Chew, S. L., Cartegni, L., Zhang, M. Q. and Krainer, A. R. (2000) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol*, **20** (3), 1063-1071.

Liu, H. X., Zhang, M. and Krainer, A. R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev*, **12** (13), 1998-2012.

Liu, S., Rauhut, R., Vornlocher, H. P. and Luhrmann, R. (2006) The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA*, **12** (7), 1418-1430.

Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A. and Sattler, M. (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*, **294** (5544), 1098-1102.

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), 5443-5450.

Liu, Z. R., Laggerbauer, B., Luhrmann, R. and Smith, C. W. (1997) Crosslinking of the U5 snRNP-specific 116-kDa protein to RNA hairpins that block step 2 of splicing. *RNA*, **3** (11), 1207-1219.

Long, J. C. and Caceres, J. F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem J*, **417** (1), 15-27.

Longman, D., McGarvey, T., McCracken, S., Johnstone, I. L., Blencowe, B. J. and Caceres, J. F. (2001) Multiple interactions between SRm160 and SR family proteins in enhancer-dependent splicing and development of C. elegans. *Curr Biol*, **11** (24), 1923-1933.

Lou, H., Neugebauer, K. M., Gagel, R. F. and Berget, S. M. (1998) Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol Cell Biol*, **18** (9), 4977-4985.

Lunde, B. M., Moore, C. and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*, **8** (6), 479-490.

Lynch, K. W. and Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes Dev*, **10** (16), 2089-2101.

# M

Ma, A. S., Moran-Jones, K., Shan, J., Munro, T. P., Snee, M. J., Hoek, K. S. and Smith, R. (2002) Heterogeneous nuclear ribonucleoprotein A3, a novel RNA trafficking response element-binding protein. *J Biol Chem*, **277** (20), 18010-18020.

Ma, Y. M. and Vogt, V. M. (2004) Nucleic acid binding-induced Gag dimerization in the assembly of Rous sarcoma virus particles in vitro. *J Virol*, **78** (1), 52-60.

Maamra, M., Milward, A., Esfahani, H. Z., Abbott, L. P., Metherell, L. A., Savage, M. O., Clark, A. J. and Ross, R. J. (2006) A 36 residues insertion in the dimerization domain of the growth hormone receptor results in defective trafficking rather than impaired signaling. *J Endocrinol*, **188** (2), 251-261.

Maciolek, N. L. and McNally, M. T. (2007) Serine/arginine-rich proteins contribute to negative regulator of splicing element-stimulated polyadenylation in rous sarcoma virus. *J Virol*, **81** (20), 11208-11217.

Maciolek, N. L. and McNally, M. T. (2008) Characterization of Rous sarcoma virus polyadenylation site use in vitro. *Virology*, **374** (2), 468-476.

Madhani, H. D. and Guthrie, C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71** (5), 803-817.

Maeder, C., Kutach, A. K. and Guthrie, C. (2009) ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat Struct Mol Biol*, **16** (1), 42-48.

Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M. and Luhrmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, **298** (5601), 2205-2208.

Makarova, O. V., Makarov, E. M., Liu, S., Vornlocher, H. P. and Luhrmann, R. (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *Embo J*, **21** (5), 1148-1157.

Makarova, O. V., Makarov, E. M. and Luhrmann, R. (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), 2553-2563.

Makarova, O. V., Makarov, E. M., Urlaub, H., Will, C. L., Gentzel, M., Wilm, M. and Luhrmann, R. (2004) A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *Embo J*, **23** (12), 2381-2391.

Mandel, C. R., Bai, Y. and Tong, L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci*, **65** (7-8), 1099-1122.

Manley, J. L. and Krainer, A. R. (2010) A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev*, **24** (11), 1073-1074.

Marchand, V., Mereau, A., Jacquenet, S., Thomas, D., Mougin, A., Gattoni, R., Stevenin, J. and Branlant, C. (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), 629-652.

Maroney, P. A., Romfo, C. M. and Nilsen, T. W. (2000) Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol Cell*, **6** (2), 317-328.

Martin, A., Schneider, S. and Schwer, B. (2002) Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J Biol Chem*, **277** (20), 17743-17750.

Martin, G. S. (1970) Rous sarcoma virus: a function required for the maintenance of the transformed state. *Nature*, **227** (5262), 1021-1023.

Martin, G. S. (2001) The hunting of the Src. *Nat Rev Mol Cell Biol*, **2** (6), 467-475.

Martin, G. S. (2004) The road to Src. *Oncogene*, **23** (48), 7910-7917.

Martinez, R., Mathey-Prevot, B., Bernards, A. and Baltimore, D. (1987) Neuronal pp60c-src contains a six-amino acid insertion relative to its non-neuronal counterpart. *Science*, **237** (4813), 411-415.

Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J. F., Revil, T. and Chabot, B. (2007) hnRNP proteins and splicing control. *Adv Exp Med Biol*, **623**, 123-147.

Martinez-Contreras, R., Fisette, J. F., Nasim, F. U., Madden, R., Cordeau, M. and Chabot, B. (2006) Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol*, **4** (2), e21.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N. and Reed, R. (2005) Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev*, **19** (13), 1512-1517.

Mathew, R., Hartmuth, K., Mohlmann, S., Urlaub, H., Ficner, R. and Luhrmann, R. (2008) Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat Struct Mol Biol*, **15** (5), 435-443.

Mathews, D. H., Sabina, J., Zuker, M. and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, **288** (5), 911-940.

Matlin, A. J., Clark, F. and Smith, C. W. (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol*, **6** (5), 386-398.

Mattaj, I. W., Habets, W. J. and van Venrooij, W. J. (1986) Monospecific antibodies reveal details of U2 snRNP structure and interaction between U1 and U2 snRNPs. *Embo J*, **5** (5), 997-1002.

Matter, N., Marx, M., Weg-Remers, S., Ponta, H., Herrlich, P. and Konig, H. (2000) Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways. *J Biol Chem*, **275** (45), 35353-35360.

Mayas, R. M., Maita, H., Semlow, D. R. and Staley, J. P. (2010) Spliceosome discards intermediates via the DEAH box ATPase Prp43p. *Proc Natl Acad Sci U S A*, **107** (22), 10020-10025.

Mayas, R. M., Maita, H. and Staley, J. P. (2006) Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol*, **13** (6), 482-490.

Mayeda, A., Munroe, S. H., Caceres, J. F. and Krainer, A. R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *Embo J*, **13** (22), 5483-5495.

Mayeda, A., Munroe, S. H., Xu, R. M. and Krainer, A. R. (1998) Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. *RNA*, **4** (9), 1111-1123.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S. and Bentley, D. L. (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), 3306-3318.

McNally, L. M. and McNally, M. T. (1996) SR protein splicing factors interact with the Rous sarcoma virus negative regulator of splicing element. *J Virol*, **70** (2), 1163-1172.

McNally, L. M. and McNally, M. T. (1998) An RNA splicing enhancer-like sequence is a component of a splicing inhibitor element from Rous sarcoma virus. *Mol Cell Biol*, **18** (6), 3103-3111.

McNally, L. M. and McNally, M. T. (1999) U1 small nuclear ribonucleoprotein and splicing inhibition by the rous sarcoma virus negative regulator of splicing element. *J Virol*, **73** (3), 2385-2393.

McNally, L. M., Yee, L. and McNally, M. T. (2004) Two regions promote U11 small nuclear ribonucleoprotein particle binding to a retroviral splicing inhibitor element (negative regulator of splicing). *J Biol Chem*, **279** (37), 38201-38208.

McNally, L. M., Yee, L. and McNally, M. T. (2006) Heterogeneous nuclear ribonucleoprotein H is required for optimal U11 small nuclear ribonucleoprotein binding to a retroviral RNA-processing control element: implications for U12-dependent RNA splicing. *J Biol Chem*, **281** (5), 2478-2488.

McNally, M. T. (2008) RNA processing control in avian retroviruses. *Front Biosci*, **13**, 3869-3883.

McNally, M. T. and Beemon, K. (1992) Intronic sequences and 3' splice sites control Rous sarcoma virus RNA splicing. *J Virol*, **66** (1), 6-11.

McNally, M. T., Gontarek, R. R. and Beemon, K. (1991) Characterization of Rous sarcoma virus intronic sequences that negatively regulate splicing. *Virology*, **185** (1), 99-108.

Medina, G., Pincetic, A., Ehrlich, L. S., Zhang, Y., Tang, Y., Leis, J. and Carter, C. A. (2008) Tsg101 can replace Nedd4 function in ASV Gag release but not membrane targeting. *Virology*, **377** (1), 30-38.

Meister, G., Hannus, S., Plottner, O., Baars, T., Hartmann, E., Fakan, S., Laggerbauer, B. and Fischer, U. (2001) SMNrp is an essential pre-mRNA splicing factor required for the formation of the mature spliceosome. *Embo J*, **20** (9), 2304-2314.

Merendino, L., Guth, S., Bilbao, D., Martinez, C. and Valcarcel, J. (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature*, **402** (6763), 838-841.

Mermoud, J. E., Cohen, P. and Lamond, A. I. (1992) Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res*, **20** (20), 5263-5269.

Mermoud, J. E., Cohen, P. T. and Lamond, A. I. (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *Embo J*, **13** (23), 5679-5688.

Merz, C., Urlaub, H., Will, C. L. and Luhrmann, R. (2007) Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. *RNA*, **13** (1), 116-128.

Michaud, S. and Reed, R. (1993) A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev*, **7** (6), 1008-1020.

Miguet, L., Bogumil, R., Decloquement, P., Herbrecht, R., Potier, N., Mauvieux, L. and Van Dorsselaer, A. (2006) Discovery and identification of potential biomarkers in a prospective study of chronic lymphoid malignancies using SELDI-TOF-MS. *J Proteome Res*, **5** (9), 2258-2269.

Miller, J. T. and Stoltzfus, C. M. (1992) Two distant upstream regions containing cis-acting signals regulating splicing facilitate 3'-end processing of avian sarcoma virus RNA. *J Virol*, **66** (7), 4242-4251.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S. Z., Keller, W., Antoniou, M. and Vagner, S. (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *Embo J*, **25** (20), 4854-4864.

Millevoi, S. and Vagner, S. (2010) Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res*, **38** (9), 2757-2774.
Min, H., Chan, R. C. and Black, D. L. (1995) The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev*, **9** (21), 2659-2671.
Misteli, T., Caceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F. and Spector, D. L. (1998) Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J Cell Biol*, **143** (2), 297-307.
Moore, M. J. (2000) Intron recognition comes of AGe. *Nat Struct Biol*, **7** (1), 14-16.
Moore, M. J. and Proudfoot, N. J. (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*, **136** (4), 688-700.
Morris, S. and Leis, J. (1999) Changes in Rous sarcoma virus RNA secondary structure near the primer binding site upon tRNATrp primer annealing. *J Virol*, **73** (8), 6307-6318.
Motta, J. V., Crittenden, L. B., Purchase, H. G., Stone, H. A. and Witter, R. L. (1975) Low oncogenic potential of avian endogenous RNA tumor virus infection or expression. *J Natl Cancer Inst*, **55** (3), 685-689.
Mougin, A., Gottschalk, A., Fabrizio, P., Luhrmann, R. and Branlant, C. (2002) Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J Mol Biol*, **317** (5), 631-649.
Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J. A. (1983) The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell*, **33** (2), 509-518.
Mullen, M. P., Smith, C. W., Patton, J. G. and Nadal-Ginard, B. (1991) Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev*, **5** (4), 642-655.
Muro, A. F., Caputi, M., Pariyarath, R., Pagani, F., Buratti, E. and Baralle, F. E. (1999) Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol Cell Biol*, **19** (4), 2657-2671.
Muto, Y., Pomeranz Krummel, D., Oubridge, C., Hernandez, H., Robinson, C. V., Neuhaus, D. and Nagai, K. (2004) The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol*, **341** (1), 185-198.

# N

Nasim, F. U., Hutchison, S., Cordeau, M. and Chabot, B. (2002) High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA*, **8** (8), 1078-1089.
Neiman, P. E., Grbic, J. J., Polony, T. S., Kimmel, R., Bowers, S. J., Delrow, J. and Beemon, K. L. (2003) Functional genomic analysis reveals distinct neoplastic phenotypes associated with c-myb mutation in the bursa of Fabricius. *Oncogene*, **22** (7), 1073-1086.
Nelissen, R. L., Will, C. L., van Venrooij, W. J. and Luhrmann, R. (1994) The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *Embo J*, **13** (17), 4113-4125.
Nelle, T. D. and Wills, J. W. (1996) A large region within the Rous sarcoma virus matrix protein is dispensable for budding and infectivity. *J Virol*, **70** (4), 2269-2276.
Newman, A. J. and Nagai, K. (2010) Structural studies of the spliceosome: blind men and an elephant. *Curr Opin Struct Biol*, **20** (1), 82-89.

Newman, A. J. and Norman, C. (1992) U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell*, **68** (4), 743-754.

Niwa, M. and Berget, S. M. (1991) Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev*, **5** (11), 2086-2095.

Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R. and Luhrmann, R. (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *Embo J*, **18** (21), 6119-6133.

Nottrott, S., Urlaub, H. and Luhrmann, R. (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *Embo J*, **21** (20), 5527-5538.

# O

Ogert, R. A. and Beemon, K. L. (1998) Mutational analysis of the rous sarcoma virus DR posttranscriptional control element. *J Virol*, **72** (4), 3407-3411.

Ogert, R. A., Lee, L. H. and Beemon, K. L. (1996) Avian retroviral RNA element promotes unspliced RNA accumulation in the cytoplasm. *J Virol*, **70** (6), 3834-3843.

O'Keefe, R. T., Norman, C. and Newman, A. J. (1996) The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell*, **86** (4), 679-689.

Okunola, H. L. and Krainer, A. R. (2009) Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol*.

O'Reilly, M. M., McNally, M. T. and Beemon, K. L. (1995) Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology*, **213** (2), 373-385.

O'Sullivan, C. T., Polony, T. S., Paca, R. E. and Beemon, K. L. (2002) Rous sarcoma virus negative regulator of splicing selectively suppresses SRC mRNA splicing and promotes polyadenylation. *Virology*, **302** (2), 405-412.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372** (6505), 432-438.

Owen, H. R., Elser, M., Cheung, E., Gersbach, M., Kraus, W. L. and Hottiger, M. O. (2007) MYBBP1a is a novel repressor of NF-kappaB. *J Mol Biol*, **366** (3), 725-736.

# P

Paca, R. E., Hibbert, C. S., O'Sullivan, C. T. and Beemon, K. L. (2001) Retroviral splicing suppressor requires three nonconsensus uridines in a 5' splice site-like sequence. *J Virol*, **75** (16), 7763-7768.

Paca, R. E., Ogert, R. A., Hibbert, C. S., Izaurralde, E. and Beemon, K. L. (2000) Rous sarcoma virus DR posttranscriptional elements use a novel RNA export pathway. *J Virol*, **74** (20), 9507-9514.

Padgett, R. A. and Shukla, G. C. (2002) A revised model for U4atac/U6atac snRNA base pairing. *RNA*, **8** (2), 125-128.

Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dork, T. and Baralle, F. E. (2002) A new type of mutation causes a splicing defect in ATM. *Nat Genet*, **30** (4), 426-429.

Xiang, Y., Cameron, C. E., Wills, J. W. and Leis, J. (1996) Fine mapping and characterization of the Rous sarcoma virus Pr76gag late assembly domain. *J Virol*, **70** (8), 5695-5700.
Xiao, S. H. and Manley, J. L. (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), 334-344.
Xiao, S. H. and Manley, J. L. (1998) Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *Embo J*, **17** (21), 6359-6367.
Xiao, X., Wang, Z., Jang, M., Nutiu, R., Wang, E. T. and Burge, C. B. (2009) Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol*, **16** (10), 1094-1100.
Xing, Y. and Lee, C. (2006) Alternative splicing and RNA selection pressure--evolutionary consequences for eukaryotic genomes. *Nat Rev Genet*, **7** (7), 499-509.
Xu, Y. Z., Newnham, C. M., Kameoka, S., Huang, T., Konarska, M. M. and Query, C. C. (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *Embo J*, **23** (2), 376-385.

# Y

Yang, J. and Cullen, B. R. (1999) Structural and functional analysis of the avian leukemia virus constitutive transport element. *RNA*, **5** (12), 1645-1655.
Yeakley, J. M., Tronchere, H., Olesen, J., Dyck, J. A., Wang, H. Y. and Fu, X. D. (1999) Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *J Cell Biol*, **145** (3), 447-455.
Yeatman, T. J. (2004) A renaissance for SRC. *Nat Rev Cancer*, **4** (6), 470-480.
Yedavalli, V. S., Neuveut, C., Chi, Y. H., Kleiman, L. and Jeang, K. T. (2004) Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell*, **119** (3), 381-392.
Yeo, G., Hoon, S., Venkatesh, B. and Burge, C. B. (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc Natl Acad Sci U S A*, **101** (44), 15700-15705.
Yoshimoto, R., Kataoka, N., Okawa, K. and Ohno, M. (2009) Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Res*, **37** (3), 891-902.
Yu, Y., Maroney, P. A., Denker, J. A., Zhang, X. H., Dybkov, O., Luhrmann, R., Jankowsky, E., Chasin, L. A. and Nilsen, T. W. (2008) Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell*, **135** (7), 1224-1236.
Zahler, A. M., Damgaard, C. K., Kjems, J. and Caputi, M. (2004) SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing.
*J Biol Chem*, **279** (11), 10077-10084.

# Z

Zahler, A. M., Lane, W. S., Stolk, J. A. and Roth, M. B. (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev*, **6** (5), 837-847.
Zamore, P. D. and Green, M. R. (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A*, **86** (23), 9243-9247.

Pandya-Jones, A. (2011) Pre-mRNA splicing during transcription in the mammalian system. *Wiley Interdiscip Rev RNA*, **2** (5), 700-717.

Parent, L. J., Bennett, R. P., Craven, R. C., Nelle, T. D., Krishna, N. K., Bowzard, J. B., Wilson, C. B., Puffer, B. A., Montelaro, R. C. and Wills, J. W. (1995) Positionally independent and exchangeable late budding functions of the Rous sarcoma virus and human immunodeficiency virus Gag proteins. *J Virol*, **69** (9), 5455-5460.

Park, J. W. and Graveley, B. R. (2007) Complex alternative splicing. *Adv Exp Med Biol*, **623**, 50-63.

Patel, A. A. and Steitz, J. A. (2003) Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*, **4** (12), 960-970.

Patel, S. B. and Bellini, M. (2008) The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res*, **36** (20), 6482-6493.

Patnaik, A., Chau, V. and Wills, J. W. (2000) Ubiquitin is part of the retrovirus budding machinery. *Proc Natl Acad Sci U S A*, **97** (24), 13069-13074.

Payne, L. N. (1998) Retrovirus-induced disease in poultry. *Poult Sci*, **77** (8), 1204-1212.

Pena, V., Liu, S., Bujnicki, J. M., Luhrmann, R. and Wahl, M. C. (2007) Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. *Mol Cell*, **25** (4), 615-624.

Pena, V., Rozov, A., Fabrizio, P., Luhrmann, R. and Wahl, M. C. (2008) Structure and function of an RNase H domain at the heart of the spliceosome. *Embo J*, **27** (21), 2929-2940.

Perez, I., Lin, C. H., McAfee, J. G. and Patton, J. G. (1997) Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *RNA*, **3** (7), 764-778.

Perriman, R. and Ares, M., Jr. (2010) Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. *Mol Cell*, **38** (3), 416-427.

Pessa, H. K., Will, C. L., Meng, X., Schneider, C., Watkins, N. J., Perala, N., Nymark, M., Turunen, J. J., Luhrmann, R. and Frilander, M. J. (2008) Minor spliceosome components are predominantly localized in the nucleus. *Proc Natl Acad Sci U S A*, **105** (25), 8655-8660.

Pinol-Roma, S. and Dreyfuss, G. (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature*, **355** (6362), 730-732.

Polge, E., Darlix, J. L., Paoletti, J. and Fosse, P. (2000) Characterization of loose and tight dimer forms of avian leukosis virus RNA. *J Mol Biol*, **300** (1), 41-56.

Polony, T. S., Bowers, S. J., Neiman, P. E. and Beemon, K. L. (2003) Silent point mutation in an avian retrovirus RNA processing element promotes c-myb-associated short-latency lymphomas. *J Virol*, **77** (17), 9378-9387.

Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K., Li, J. and Nagai, K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature*, **458** (7237), 475-480.

Purcell, D. F. and Martin, M. A. (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol*, **67** (11), 6365-6378.

# Q

Qu, H. L., Michot, B. and Bachellerie, J. P. (1983) Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct

mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA. *Nucleic Acids Res*, **11** (17), 5903-5920.
Query, C. C., Bentley, R. C. and Keene, J. D. (1989) A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol Cell Biol*, **9** (11), 4872-4881.
Query, C. C., Moore, M. J. and Sharp, P. A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev*, **8** (5), 587-597.
Quiocho, F. A., Spurlino, J. C. and Rodseth, L. E. (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure*, **5** (8), 997-1015.

# R

Rader, S. D. and Guthrie, C. (2002) A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA*, **8** (11), 1378-1392.
Raghunathan, P. L. and Guthrie, C. (1998) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science*, **279** (5352), 857-860.
Raiborg, C. and Stenmark, H. (2009) The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature*, **458** (7237), 445-452.
Rappsilber, J., Ajuh, P., Lamond, A. I. and Mann, M. (2001) SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome. *J Biol Chem*, **276** (33), 31142-31150.
Rappsilber, J., Ryder, U., Lamond, A. I. and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res*, **12** (8), 1231-1245.
Rebane, A., Aab, A. and Steitz, J. A. (2004) Transportins 1 and 2 are redundant nuclear import factors for hnRNP A1 and HuR. *RNA*, **10** (4), 590-599.
Reed, R. (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr Opin Cell Biol*, **12** (3), 340-345.
Reed, R. and Maniatis, T. (1986) A role for exon sequences and splice-site proximity in splice-site selection. *Cell*, **46** (5), 681-690.
Reyes, J. L., Gustafson, E. H., Luo, H. R., Moore, M. J. and Konarska, M. M. (1999) The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA*, **5** (2), 167-179.
Reyes, J. L., Kois, P., Konforti, B. B. and Konarska, M. M. (1996) The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA*, **2** (3), 213-225.
Rio, D. C. (1991) Regulation of Drosophila P element transposition. *Trends Genet*, **7** (9), 282-287.
Ritchie, D. B., Schellenberg, M. J., Gesner, E. M., Raithatha, S. A., Stuart, D. T. and Macmillan, A. M. (2008) Structural elucidation of a PRP8 core domain from the heart of the spliceosome. *Nat Struct Mol Biol*, **15** (11), 1199-1205.
Robinson, H. L., Blais, B. M., Tsichlis, P. N. and Coffin, J. M. (1982) At least two regions of the viral genome determine the oncogenic potential of avian leukosis viruses. *Proc Natl Acad Sci U S A*, **79** (4), 1225-1229.
Roca, X. and Krainer, A. R. (2009) Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. *Nat Struct Mol Biol*, **16** (2), 176-182.
Roca, X., Sachidanandam, R. and Krainer, A. R. (2005) Determinants of the inherent strength of human 5' splice sites. *RNA*, **11** (5), 683-698.

Rooke, N., Markovtsov, V., Cagavi, E. and Black, D. L. (2003) Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. *Mol Cell Biol*, **23** (6), 1874-1884.

Ropers, D., Ayadi, L., Gattoni, R., Jacquenet, S., Damier, L., Branlant, C. and Stevenin, J. (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), 29963-29973.

Roscigno, R. F. and Garcia-Blanco, M. A. (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA*, **1** (7), 692-706.

Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C. and Tazi, J. (1996) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature*, **381** (6577), 80-82.

Roth, M. J., Schwartzberg, P. L. and Goff, S. P. (1989) Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell*, **58** (1), 47-54.

Rous, P. (1911) A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells. *J Exp Med*, **13** (4), 397-411.

Rousseau, S., Morrice, N., Peggie, M., Campbell, D. G., Gaestel, M. and Cohen, P. (2002) Inhibition of SAPK2a/p38 prevents hnRNP A0 phosphorylation by MAPKAP-K2 and its interaction with cytokine mRNAs. *Embo J*, **21** (23), 6505-6514.

Rubin, H. (1955) Quantitative relations between causative virus and cell in the Rous no. 1 chicken sarcoma. *Virology*, **1** (5), 445-473.

Rubin, H. and Vogt, P. K. (1962) An avian leukosis virus associated with stocks of Rous sarcoma virus. *Virology*, **17**, 184-194.

Ruddell, A. (1995) Transcription regulatory elements of the avian retroviral long terminal repeat. *Virology*, **206** (1), 1-7.

Rutz, B. and Seraphin, B. (1999) Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA*, **5** (6), 819-831.

Ryden, T. A., de Mars, M. and Beemon, K. (1993) Mutation of the C/EBP binding sites in the Rous sarcoma virus long terminal repeat and gag enhancers. *J Virol*, **67** (5), 2862-2870.

# S

Saliou, J. M., Bourgeois, C. F., Ayadi-Ben Mena, L., Ropers, D., Jacquenet, S., Marchand, V., Stevenin, J. and Branlant, C. (2009) Role of RNA structure and protein factors in the control of HIV-1 splicing. *Front Biosci*, **14**, 2714-2729.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, New York.

Sander, B., Golas, M. M., Makarov, E. M., Brahms, H., Kastner, B., Luhrmann, R. and Stark, H. (2006) Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, **24** (2), 267-278.

Sanford, J. R., Coutinho, P., Hackett, J. A., Wang, X., Ranahan, W. and Caceres, J. F. (2008) Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS One*, **3** (10), e3369.

Sapra, A. K., Anko, M. L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E. M. and Neugebauer, K. M. (2009) SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. *Mol Cell*, **34** (2), 179-190.

Sashital, D. G., Cornilescu, G., McManus, C. J., Brow, D. A. and Butcher, S. E. (2004) U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat Struct Mol Biol*, **11** (12), 1237-1242.

Sawa, H. and Shimura, Y. (1992) Association of U6 snRNA with the 5'-splice site region of pre-mRNA in the spliceosome. *Genes Dev*, **6** (2), 244-254.

Schaal, T. D. and Maniatis, T. (1999) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol*, **19** (3), 1705-1719.

Scheifele, L. Z., Garbitt, R. A., Rhoads, J. D. and Parent, L. J. (2002) Nuclear entry and CRM1-dependent nuclear export of the Rous sarcoma virus Gag polyprotein. *Proc Natl Acad Sci U S A*, **99** (6), 3944-3949.

Scheifele, L. Z., Ryan, E. P. and Parent, L. J. (2005) Detailed mapping of the nuclear export signal in the Rous sarcoma virus Gag protein. *J Virol*, **79** (14), 8732-8741.

Schellenberg, M. J., Dul, E. L. and MacMillan, A. M. (2011) Structural model of the p14/SF3b155 . branch duplex complex. *RNA*, **17** (1), 155-165.

Scherer, S. F. (1953) The utilization of a pure strain of mammalian cells (Earle) for the cultivation of viruses in vitro. I. Multiplication of pseudorabies and herpes simplex viruses. *Am J Pathol*, **29** (1), 113-137.

Schneider, C., Will, C. L., Brosius, J., Frilander, M. J. and Luhrmann, R. (2004) Identification of an evolutionarily divergent U11 small nuclear ribonucleoprotein particle in Drosophila. *Proc Natl Acad Sci U S A*, **101** (26), 9584-9589.

Schneider, C., Will, C. L., Makarova, O. V., Makarov, E. M. and Luhrmann, R. (2002) Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol Cell Biol*, **22** (10), 3219-3229.

Schneider, M., Hsiao, H. H., Will, C. L., Giet, R., Urlaub, H. and Luhrmann, R. (2010a) Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. *Nat Struct Mol Biol*, **17** (2), 216-221.

Schneider, M., Will, C. L., Anokhina, M., Tazi, J., Urlaub, H. and Luhrmann, R. (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol Cell*, **38** (2), 223-235.

Schultz, A., Nottrott, S., Hartmuth, K. and Luhrmann, R. (2006a) RNA structural requirements for the association of the spliceosomal hPrp31 protein with the U4 and U4atac small nuclear ribonucleoproteins. *J Biol Chem*, **281** (38), 28278-28286.

Schultz, A., Nottrott, S., Watkins, N. J. and Luhrmann, R. (2006b) Protein-protein and protein-RNA contacts both contribute to the 15.5K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs. *Mol Cell Biol*, **26** (13), 5146-5154.

Schwartz, D. E., Tizard, R. and Gilbert, W. (1983) Nucleotide sequence of Rous sarcoma virus. *Cell*, **32** (3), 853-869.

Schwartz, S. H., Silva, J., Burstein, D., Pupko, T., Eyras, E. and Ast, G. (2008) Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res*, **18** (1), 88-103.

Schwer, B. (2008) A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell*, **30** (6), 743-754.

Schwer, B. and Gross, C. H. (1998) Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *Embo J*, **17** (7), 2086-2094.

Schwer, B. and Guthrie, C. (1992) A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *Embo J*, **11** (13), 5033-5039.

Segault, V., Will, C. L., Polycarpou-Schwarz, M., Mattaj, I. W., Branlant, C. and Luhrmann, R. (1999) Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic

steps of pre-mRNA splicing in HeLa nuclear extracts. *Mol Cell Biol*, **19** (4), 2782-2790.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A. and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell*, **11** (4), 965-976.

Sharma, S., Falick, A. M. and Black, D. L. (2005) Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol Cell*, **19** (4), 485-496.

Sharma, S., Kohlstaedt, L. A., Damianov, A., Rio, D. C. and Black, D. L. (2008) Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol*, **15** (2), 183-191.

Sharma, S., Maris, C., Allain, F. H. and Black, D. L. (2010) U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. *Mol Cell*, **41** (5), 579-588.

Shav-Tal, Y. and Zipori, D. (2002) PSF and p54(nrb)/NonO--multi-functional nuclear proteins. *FEBS Lett*, **531** (2), 109-114.

Shen, H. and Green, M. R. (2004) A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell*, **16** (3), 363-373.

Shen, H. and Green, M. R. (2006) RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev*, **20** (13), 1755-1765.

Shen, H. and Green, M. R. (2007) RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nat Struct Mol Biol*, **14** (7), 597-603.

Shen, H., Kan, J. L. and Green, M. R. (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell*, **13** (3), 367-376.

Shen, H., Zheng, X., Luecke, S. and Green, M. R. (2010) The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev*, **24** (21), 2389-2394.

Shen, H., Zheng, X., Shen, J., Zhang, L., Zhao, R. and Green, M. R. (2008) Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes Dev*, **22** (13), 1796-1803.

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R. and Sachidanandam, R. (2006) Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res*, **34** (14), 3955-3967.

Shi, H., Hoffman, B. E. and Lis, J. T. (1997) A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. *Mol Cell Biol*, **17** (5), 2649-2657.

Shi, Y. and Manley, J. L. (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., Feng, Y. and Manley, J. L. (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature*, **427** (6974), 553-558.

Shin, C. and Manley, J. L. (2004) Cell signalling and the control of pre-mRNA splicing. *Nat Rev Mol Cell Biol*, **5** (9), 727-738.

Shukla, G. C., Cole, A. J., Dietrich, R. C. and Padgett, R. A. (2002) Domains of human U4atac snRNA required for U12-dependent splicing in vivo. *Nucleic Acids Res*, **30** (21), 4650-4657.

Si, Z., Amendt, B. A. and Stoltzfus, C. M. (1997) Splicing efficiency of human immunodeficiency virus type 1 tat RNA is determined by both a suboptimal 3' splice

site and a 10 nucleotide exon splicing silencer element located within tat exon 2. *Nucleic Acids Res*, **25** (4), 861-867.
Sickmier, E. A., Frato, K. E., Shen, H., Paranawithana, S. R., Green, M. R. and Kielkopf, C. L. (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell*, **23** (1), 49-59.
Siebel, C. W., Admon, A. and Rio, D. C. (1995) Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev*, **9** (3), 269-283.
Siebel, C. W., Fresco, L. D. and Rio, D. C. (1992) The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes Dev*, **6** (8), 1386-1401.
Siebel, C. W., Kanaar, R. and Rio, D. C. (1994) Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes Dev*, **8** (14), 1713-1725.
Sikand, K. and Shukla, G. C. (2011) Functionally important structural elements of U12 snRNA. *Nucleic Acids Res*.
Simard, M. J. and Chabot, B. (2002) SRp30c is a repressor of 3' splice site utilization. *Mol Cell Biol*, **22** (12), 4001-4010.
Simon, M. C., Neckameyer, W. S., Hayward, W. S. and Smith, R. E. (1987) Genetic determinants of neoplastic diseases induced by a subgroup F avian leukosis virus. *J Virol*, **61** (4), 1203-1212.
Simpson, S. B., Guo, W., Winistorfer, S. C., Craven, R. C. and Stoltzfus, C. M. (1998) The upstream, direct repeat sequence of Prague A Rous sarcoma virus is deficient in mediating efficient Gag assembly and particle release. *Virology*, **247** (1), 86-96.
Simpson, S. B., Zhang, L., Craven, R. C. and Stoltzfus, C. M. (1997) Rous sarcoma virus direct repeat cis elements exert effects at several points in the virus life cycle. *J Virol*, **71** (12), 9150-9156.
Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G. P., Bresolin, N., Giorda, R. and Pozzoli, U. (2004) Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res*, **32** (5), 1783-1791.
Small, E. C., Leggett, S. R., Winans, A. A. and Staley, J. P. (2006) The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol Cell*, **23** (3), 389-399.
Smith, C. W. and Nadal-Ginard, B. (1989) Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. *Cell*, **56** (5), 749-758.
Smith, C. W. and Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, **25** (8), 381-388.
Smith, J. A. and Daniel, R. (2006) Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses. *ACS Chem Biol*, **1** (4), 217-226.
Smith, M. R., Smith, R. E., Dunkel, I., Hou, V., Beemon, K. L. and Hayward, W. S. (1997) Genetic determinant of rapid-onset B-cell lymphoma by avian leukosis virus. *J Virol*, **71** (9), 6534-6540.
Sonstegard, T. S. and Hackett, P. B. (1996) Autogenous regulation of RNA translation and packaging by Rous sarcoma virus Pr76gag. *J Virol*, **70** (10), 6642-6652.
Sontheimer, E. J. and Steitz, J. A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262** (5142), 1989-1996.
Sorge, J., Ricci, W. and Hughes, S. H. (1983) cis-Acting RNA packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus. *J Virol*, **48** (3), 667-675.
Spector, D. L. and Lamond, A. I. (2011) Nuclear speckles. *Cold Spring Harb Perspect Biol*, **3** (2).

Spellman, R., Llorian, M. and Smith, C. W. (2007) Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol Cell*, **27** (3), 420-434.

Spidel, J. L., Craven, R. C., Wilson, C. B., Patnaik, A., Wang, H., Mansky, L. M. and Wills, J. W. (2004) Lysines close to the Rous sarcoma virus late domain critical for budding. *J Virol*, **78** (19), 10606-10616.

Staffa, A. and Cochrane, A. (1995) Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. *Mol Cell Biol*, **15** (8), 4597-4605.

Staknis, D. and Reed, R. (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), 7670-7682.

Stamm, S. (2008) Regulation of alternative splicing by reversible protein phosphorylation. *J Biol Chem*, **283** (3), 1223-1227.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T. A. and Soreq, H. (2005) Function of alternative splicing. *Gene*, **344**, 1-20.

Stehelin, D., Varmus, H. E., Bishop, J. M. and Vogt, P. K. (1976) DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature*, **260** (5547), 170-173.

Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D. and Abelson, J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell*, **9** (1), 31-44.

Stoltzfus, C. M. and Fogarty, S. J. (1989) Multiple regions in the Rous sarcoma virus src gene intron act in cis to affect the accumulation of unspliced RNA. *J Virol*, **63** (4), 1669-1676.

Stoltzfus, C. M., Lorenzen, S. K. and Berberich, S. L. (1987) Noncoding region between the env and src genes of Rous sarcoma virus influences splicing efficiency at the src gene 3' splice site. *J Virol*, **61** (1), 177-184.

Studier, F. W. and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, **189** (1), 113-130.

Sugden, B. (1993) How some retroviruses got their oncogenes. *Trends Biochem Sci*, **18** (7), 233-235.

Sun, H. and Chasin, L. A. (2000) Multiple splicing defects in an intronic false exon. *Mol Cell Biol*, **20** (17), 6414-6425.

Sun, Q., Mayeda, A., Hampson, R. K., Krainer, A. R. and Rottman, F. M. (1993) General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes Dev*, **7** (12B), 2598-2608.

Swanstrom, R., Parker, R. C., Varmus, H. E. and Bishop, J. M. (1983) Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. *Proc Natl Acad Sci U S A*, **80** (9), 2519-2523.

Swanstrom, R., Varmus, H. E. and Bishop, J. M. (1982) Nucleotide sequence of the 5' noncoding region and part of the gag gene of Rous sarcoma virus. *J Virol*, **41** (2), 535-541.

Swanstrom, R. and Wills, J. (1997) Synthesis, assembly and processing of viral proteins. In Retroviruses. Cold Spring Harbor Laboratory Press, New-York, 263-334.

# T

Tacke, R., Chen, Y. and Manley, J. L. (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc Natl Acad Sci U S A*, **94** (4), 1148-1153.

Tacke, R. and Manley, J. L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *Embo J*, **14** (14), 3540-3551.

Tacke, R. and Manley, J. L. (1999) Determinants of SR protein specificity. *Curr Opin Cell Biol*, **11** (3), 358-362.

Tacke, R., Tohyama, M., Ogawa, S. and Manley, J. L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93** (1), 139-148.

Takeya, T. and Hanafusa, H. (1983) Structure and sequence of the cellular gene homologous to the RSV src gene and the mechanism for generating the transforming virus. *Cell*, **32** (3), 881-890.

Tanaka, K., Watakabe, A. and Shimura, Y. (1994) Polypurine sequences within a downstream exon function as a splicing enhancer. *Mol Cell Biol*, **14** (2), 1347-1354.

Tange, T. O., Damgaard, C. K., Guth, S., Valcarcel, J. and Kjems, J. (2001) The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *Embo J*, **20** (20), 5748-5758.

Tarn, W. Y. and Steitz, J. A. (1995) Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. *Proc Natl Acad Sci U S A*, **92** (7), 2504-2508.

Tarn, W. Y. and Steitz, J. A. (1996a) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science*, **273** (5283), 1824-1832.

Tarn, W. Y. and Steitz, J. A. (1996b) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell*, **84** (5), 801-811.

Tarn, W. Y., Yario, T. A. and Steitz, J. A. (1995) U12 snRNA in vertebrates: evolutionary conservation of 5' sequences implicated in splicing of pre-mRNAs containing a minor class of introns. *RNA*, **1** (6), 644-656.

Tazi, J., Kornstadt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C. and Luhrmann, R. (1993) Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature*, **363** (6426), 283-286.

Teigelkamp, S., Achsel, T., Mundt, C., Gothel, S. F., Cronshagen, U., Lane, W. S., Marahiel, M. and Luhrmann, R. (1998) The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA*, **4** (2), 127-141.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C. L. and Luhrmann, R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *RNA*, **3** (11), 1313-1326.

Teigelkamp, S., Newman, A. J. and Beggs, J. D. (1995) Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *Embo J*, **14** (11), 2602-2612.

Telesnitsky, A. and Goff, S. (1997) Reverse Transcriptase and the generation of retroviral DNA. In Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 121-160.

Temin, H. M. and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, **226** (5252), 1211-1213.

Temin, H. M. and Rubin, H. (1958) Characteristics of an assay for Rous sarcoma virus and Rous sarcoma cells in tissue culture. *Virology*, **6** (3), 669-688.

Tian, H. and Kole, R. (2001) Strong RNA splicing enhancers identified by a modified method of cycled selection interact with SR protein. *J Biol Chem*, **276** (36), 33833-33839.

Tian, M. and Maniatis, T. (1993) A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell*, **74** (1), 105-114.

Tidow, H., Andreeva, A., Rutherford, T. J. and Fersht, A. R. (2009) Solution structure of the U11-48K CHHC zinc-finger domain that specifically binds the 5' splice site of U12-type introns. *Structure*, **17** (2), 294-302.

Turunen, J. J., Will, C. L., Grote, M., Luhrmann, R. and Frilander, M. J. (2008) The U11-48K protein contacts the 5' splice site of U12-type introns and the U11-59K protein. *Mol Cell Biol*, **28** (10), 3548-3560.

# U

Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G. and Luhrmann, R. (2000) A general approach for identification of RNA-protein cross-linking sites within native human spliceosomal small nuclear ribonucleoproteins (snRNPs). Analysis of RNA-protein contacts in native U1 and U4/U6.U5 snRNPs. *J Biol Chem*, **275** (52), 41458-41468.

# V

Vagner, S., Vagner, C. and Mattaj, I. W. (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.

Valcarcel, J., Gaur, R. K., Singh, R. and Green, M. R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science*, **273** (5282), 1706-1709.

Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. and Liljas, L. (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature*, **371** (6498), 623-626.

Valegard, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G. and Liljas, L. (1997) The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol*, **270** (5), 724-738.

Valverde, R., Edwards, L. and Regan, L. (2008) Structure and function of KH domains. *Febs J*, **275** (11), 2712-2726.

van Der Houven Van Oordt, W., Newton, K., Screaton, G. R. and Caceres, J. F. (2000) Role of SR protein modular domains in alternative splicing specificity in vivo. *Nucleic Acids Res*, **28** (24), 4822-4831.

Van Dusen, C. M., Yee, L., McNally, L. M. and McNally, M. T. (2010) A glycine-rich domain of hnRNP H/F promotes nucleocytoplasmic shuttling and nuclear import through an interaction with transportin 1. *Mol Cell Biol*, **30** (10), 2552-2562.

van Nues, R. W. and Beggs, J. D. (2001) Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of Saccharomyces cerevisiae. *Genetics*, **157** (4), 1451-1467.

Vana, M. L., Tang, Y., Chen, A., Medina, G., Carter, C. and Leis, J. (2004) Role of Nedd4 and ubiquitination of Rous sarcoma virus Gag in budding of virus-like particles from cells. *J Virol*, **78** (24), 13943-13953.

Varani, L., Hasegawa, M., Spillantini, M. G., Smith, M. J., Murrell, J. R., Ghetti, B., Klug, A., Goedert, M. and Varani, G. (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), 8229-8234.

Venables, J. P., Koh, C. S., Froehlich, U., Lapointe, E., Couture, S., Inkel, L., Bramard, A., Paquet, E. R., Watier, V., Durand, M., *et al.* (2008) Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol Cell Biol*, **28** (19), 6033-6043.

Venugopal, K. (1999) Avian leukosis virus subgroup J: a rapidly evolving group of oncogenic retroviruses. *Res Vet Sci*, **67** (2), 113-119.

# W

Wahl, M. C., Will, C. L. and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136** (4), 701-718.

Wallace, S. T. and Schroeder, R. (1998) In vitro selection and characterization of streptomycin-binding RNAs: recognition discrimination between antibiotics. *RNA*, **4** (1), 112-123.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. and Burge, C. B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456** (7221), 470-476.

Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C. and Fu, X. D. (1998) SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol*, **140** (4), 737-750.

Wang, Y. and McNally, M. T. (2003) Position dependence of the Rous sarcoma virus negative regulator of splicing element reflects proximity to a 5' splice site. *Virology*, **313** (2), 629-637.

Wang, Y. and Rando, R. R. (1995) Specific binding of aminoglycoside antibiotics to RNA. *Chem Biol*, **2** (5), 281-290.

Warkocki, Z., Odenwalder, P., Schmitzova, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P. and Luhrmann, R. (2009) Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nat Struct Mol Biol*, **16** (12), 1237-1243.

Wassarman, D. A. and Steitz, J. A. (1992a) Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. *Science*, **257** (5078), 1918-1925.

Wassarman, D. A. and Steitz, J. A. (1993) A base-pairing interaction between U2 and U6 small nuclear RNAs occurs in > 150S complexes in HeLa cell extracts: implications for the spliceosome assembly pathway. *Proc Natl Acad Sci U S A*, **90** (15), 7139-7143.

Wassarman, K. M. and Steitz, J. A. (1992) The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. *Mol Cell Biol*, **12** (3), 1276-1285.

Weber, G., Cristao, V. F., de, L. A. F., Santos, K. F., Holton, N., Rappsilber, J., Beggs, J. D. and Wahl, M. C. (2011) Mechanism for Aar2p function as a U5 snRNP assembly factor. *Genes Dev*, **25** (15), 1601-1612.

Weber, G., Trowitzsch, S., Kastner, B., Luhrmann, R. and Wahl, M. C. (2010) Functional organization of the Sm core in the crystal structure of human U1 snRNP. *Embo J*, **29** (24), 4172-4184.

Weidhaas, J. B., Angelichio, E. L., Fenner, S. and Coffin, J. M. (2000) Relationship between retroviral DNA integration and gene expression. *J Virol*, **74** (18), 8382-8389.

Weil, J. E. and Beemon, K. L. (2006) A 3' UTR sequence stabilizes termination codons in the unspliced RNA of Rous sarcoma virus. *RNA*, **12** (1), 102-110.
Weil, J. E., Hadjithomas, M. and Beemon, K. L. (2009) Structural characterization of the Rous sarcoma virus RNA stability element. *J Virol*, **83** (5), 2119-2129.
Weiss, R. A. (2006) The discovery of endogenous retroviruses. *Retrovirology*, **3**, 67.
Weldon, R. A., Jr. and Wills, J. W. (1993) Characterization of a small (25-kilodalton) derivative of the Rous sarcoma virus Gag protein competent for particle release. *J Virol*, **67** (9), 5550-5561.
Werner, S., Vogel-Bachmayr, K., Hollinderbaumer, B. and Wohrl, B. M. (2001) Requirements for minus-strand transfer catalyzed by Rous sarcoma virus reverse transcriptase. *J Virol*, **75** (21), 10132-10138.
White, J. M., Delos, S. E., Brecher, M. and Schornberg, K. (2008) Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol*, **43** (3), 189-219.
Will, C. L. and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol*, **3** (7).
Will, C. L., Rumpler, S., Klein Gunnewiek, J., van Venrooij, W. J. and Luhrmann, R. (1996) In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res*, **24** (23), 4614-4623.
Will, C. L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T. and Luhrmann, R. (2004) The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA*, **10** (6), 929-941.
Will, C. L., Schneider, C., MacMillan, A. M., Katopodis, N. F., Neubauer, G., Wilm, M., Luhrmann, R. and Query, C. C. (2001) A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *Embo J*, **20** (16), 4536-4546.
Will, C. L., Schneider, C., Reed, R. and Luhrmann, R. (1999) Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science*, **284** (5422), 2003-2005.
Will, C. L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M. and Luhrmann, R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *Embo J*, **21** (18), 4978-4988.
Wills, J. W., Cameron, C. E., Wilson, C. B., Xiang, Y., Bennett, R. P. and Leis, J. (1994) An assembly domain of the Rous sarcoma virus Gag protein required late in budding. *J Virol*, **68** (10), 6605-6618.
Wilusz, J. E. and Beemon, K. L. (2006) The negative regulator of splicing element of Rous sarcoma virus promotes polyadenylation. *J Virol*, **80** (19), 9634-9640.
Withers, J. B. and Beemon, K. L. (2010) Structural features in the Rous sarcoma virus RNA stability element are necessary for sensing the correct termination codon. *Retrovirology*, **7**, 65.
Wu, J. Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75** (6), 1061-1070.
Wu, S., Romfo, C. M., Nilsen, T. W. and Green, M. R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, **402** (6763), 832-835.
Wyatt, J. R., Sontheimer, E. J. and Steitz, J. A. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev*, **6** (12B), 2542-2553.

# X

Zamore, P. D. and Green, M. R. (1991) Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *Embo J*, **10** (1), 207-214.

Zamore, P. D., Patton, J. G. and Green, M. R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, **355** (6361), 609-614.

Zhang, L., Simpson, S. B. and Stoltzfus, C. M. (1996) Selection and characterization of replication-competent revertants of a Rous sarcoma virus src gene oversplicing mutant. *J Virol*, **70** (6), 3636-3644.

Zhang, L. and Stoltzfus, C. M. (1995) A suboptimal src 3' splice site is necessary for efficient replication of Rous sarcoma virus. *Virology*, **206** (2), 1099-1107.

Zhang, X. H. and Chasin, L. A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev*, **18** (11), 1241-1250.

Zhou, J., Bean, R. L., Vogt, V. M. and Summers, M. (2007) Solution structure of the Rous sarcoma virus nucleocapsid protein: muPsi RNA packaging signal complex. *J Mol Biol*, **365** (2), 453-467.

Zhou, J., McAllen, J. K., Tailor, Y. and Summers, M. F. (2005) High affinity nucleocapsid protein binding to the muPsi RNA packaging signal of Rous sarcoma virus. *J Mol Biol*, **349** (5), 976-988.

Zhou, Z., Licklider, L. J., Gygi, S. P. and Reed, R. (2002a) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419** (6903), 182-185.

Zhou, Z. and Reed, R. (1998) Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *Embo J*, **17** (7), 2095-2106.

Zhou, Z., Sim, J., Griffith, J. and Reed, R. (2002b) Purification and electron microscopic visualization of functional human spliceosomes. *Proc Natl Acad Sci U S A*, **99** (19), 12203-12207.

Zhu, J. and Krainer, A. R. (2000) Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev*, **14** (24), 3166-3178.

Zhu, J., Mayeda, A. and Krainer, A. R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell*, **8** (6), 1351-1361.

Zhu, L., Zhang, Y., Zhang, W., Yang, S., Chen, J. Q. and Tian, D. (2009) Patterns of exon-intron architecture variation of genes in eukaryotic genomes. *BMC Genomics*, **10**, 47.

Zillmann, M., Zapp, M. L. and Berget, S. M. (1988) Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol Cell Biol*, **8** (2), 814-821.

Zorio, D. A. and Blumenthal, T. (1999) Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature*, **402** (6763), 835-838.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31** (13), 3406-3415.

Zychlinski, D., Erkelenz, S., Melhorn, V., Baum, C., Schaal, H. and Bohne, J. (2009) Limited complementarity between U1 snRNA and a retroviral 5' splice site permits its attenuation via RNA secondary structure. *Nucleic Acids Res*, **37** (22), 7429-7440.

**RESUME**

Afin de se répliquer, les rétrovirus doivent disposer à la fois d'ARN épissés et non épissés. Chez le virus du Sarcome de Rous (RSV), l'accumulation de l'ARN non épissé dépend de l'élément NRS (Negative Regulator of Splicing). L'élément NRS est un élément bipartite. Sa région 5' est assimilée à une séquence ESE (Exon Splicing Enhancer) à laquelle se fixent de nombreuses protéines SR tandis que sa région 3' contient un pseudo-site 5' non fonctionnel qui constitue un leurre qui est responsable de l'inhibition de l'épissage à l'unique site 5' fonctionnel du virus. Seule la structure 3D de la partie 3' du NRS qui contient le pseudo-site a été expérimentalement établie. Dans ce travail, nous avons déterminé la structure 2D de la totalité de l'élément NRS à l'aide de sondes chimiques et enzymatiques. La comparaison de cette structure expérimentale à celles que nous avons établies pour d'autres éléments NRS mutants fonctionnels et non fonctionnel ainsi qu'à celles théoriques de la totalité des virus aviaires séquencés argumente en faveur de la forte signification biologique de notre modèle. Des expériences d'épissage *in vitro* réalisées sur l'élément NRS sauvage ainsi que ses formes tronquées ont permis de mettre en évidence le rôle crucial de deux structures tige-boucles dans la fonction du NRS. Les expériences de purification de complexes formés avec un extrait nucléaire de cellules HeLa sur ces différents éléments NRS par des techniques chromatographie d'affinité ont permis de démontrer l'importance de l'association de ces deux structures tige-boucles avec les protéines SR et la snRNP U1. Nous avons défini un nouvel élément NRS minimal fonctionnel capable d'inhiber l'épissage et nous avons démontré l'activation de l'inhibition de l'épissage de l'élément NRS par la protéine 9G8 *in vitro* et *in cellulo*.

**TITLE**

Structural and functional study of the Negative Regulator of Splicing from Rous Sarcoma Virus

**SUMMARY**

Retroviruses require both spliced and unspliced RNAs for productive replication. Accumulation of unspliced RNA in Rous Sarcoma Virus (RSV) depends on the NRS element, (Negative Regulator of Splicing). The NRS element is bipartite. Its 5' terminal part is considered as an ESE that binds SR proteins and its 3' part contains a decoy 5'-splice site (ss), which inhibits splicing at the *bona fide* 5' ss. Only the 3D structure of a small NRS fragment including the decoy 5' ss had been experimentally studied. Here, by chemical and enzymatic probing of entire RSV NRS, we determine its 2D structure. By comparative analysis of 2D structures of functional and non-functional avian NRS variants and of all sequenced avian NRSs, we bring strong arguments for a biological significance of the established structure. By *in vitro* splicing assays, we show a crucial role of two of the established stem-loop structures and by affinity purification of complexes formed by WT and truncated NRSs in HeLa cell nuclear extract, we demonstrate their importance for SR protein and U1 snRNP association. We define a new small NRS element retaining splicing inhibitory properties and finally demonstrate the capability of the SR protein 9G8 to increase NRS activity *in vitro* and *in cellulo.*

**MOTS CLÉS**

Virus RSV, élément NRS, Structure secondaire de l'ARN, inhibition de l'épissage, protéines SR, complexes ARN-protéines

**DISCIPLINE-SPECIALITE DOCTORALE**

ARN : Biogenèse et Structure-Fonction

**INTITULE ET ADRESSE DU LABORATOIRE**

Laboratoire d'ARN, RNP, structure-fonction-maturation, Enzymologie Moléculaire et Structurale (AREMS)
UMR 7214 CNRS-UHP Nancy 1
Bâtiment Biopôle, Faculté de médecine de Nancy, Vandœuvre les Nancy, France
BP 184
54506 Vandoeuvre-lès-Nancy Cedex, France