



**HAL**  
open science

# Étude de l'action sur l'épissage de protéines nucléaires se liant à la région de l'ARN du virus VIH-1 contenant le site d'épissage A7 et role de ces protéines sur d'autres sites accepteurs d'épissage de VIH-1

Maryline Santerre

► **To cite this version:**

Maryline Santerre. Étude de l'action sur l'épissage de protéines nucléaires se liant à la région de l'ARN du virus VIH-1 contenant le site d'épissage A7 et role de ces protéines sur d'autres sites accepteurs d'épissage de VIH-1. Médecine humaine et pathologie. Université Henri Poincaré - Nancy 1, 2010. Français. NNT : 2010NAN10115 . tel-01746355

**HAL Id: tel-01746355**

**<https://hal.univ-lorraine.fr/tel-01746355>**

Submitted on 29 Mar 2018

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

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



## 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](mailto: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.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

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

**Thèse**

Présentée et soutenue publiquement pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE HENRI POINCARÉ**

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

par **Maryline SANTERRE**

**Identification des protéines capables de s'associer avec la région de l'ARN du virus de l'immunodéficience humaine (VIH-1) contenant le site d'épissage A7 et découverte d'un nouveau régulateur général de l'épissage alternatif de l'ARN du VIH-1, la protéine hnRNP K.**

**Le 10 Novembre 2010**

**Membres du jury :**

Rapporteurs :

Dr James STEVENIN

Directeur de recherche INSERM IGBMC, Illkirch

Dr Marie-Line ANDREOLA

Directeur de recherche CNRS, Bordeaux

Examineurs :

Dr Philippe FOSSE

Chargé de recherche CNRS, (LBPA) Cachan

Dr Christiane BRANLANT

Directeur de recherche CNRS, Nancy

Pr Iouri MOTORINE

Professeur Nancy Université, Nancy

Directeur de thèse



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

**Thèse**

Présentée et soutenue publiquement pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE HENRI POINCARÉ**

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

par **Maryline SANTERRE**

**Identification des protéines capables de s'associer avec la région de l'ARN du virus de l'immunodéficience humaine (VIH-1) contenant le site d'épissage A7 et découverte d'un nouveau régulateur général de l'épissage alternatif de l'ARN du VIH-1, la protéine hnRNP K.**

**Le 10 Novembre 2010**

**Membres du jury :**

Rapporteurs :

Dr James STEVENIN

Directeur de recherche INSERM IGBMC, Illkirch

Dr Marie-Line ANDREOLA

Directeur de recherche CNRS, Bordeaux

Examineurs :

Dr Philippe FOSSE

Chargé de recherche CNRS, (LBPA) Cachan

Dr Christiane BRANLANT

Directeur de recherche CNRS, Nancy

Pr Iouri MOTORINE

Professeur Nancy Université, Nancy

Directeur de thèse



# **SOMMAIRE**



# SOMMAIRE

## REMERCIEMENTS

## ABREVIATIONS

## SOMMAIRE

AVANT-PROPOS .....	1
INTRODUCTION.....	2
CHAPITRE I : Epissage des ARN pré-messagers nucléaires chez les mammifères.....	2
I.1 La réaction biochimique de l'épissage.....	2
I.2 Les séquences définissant les introns.....	3
I.2.1 Le site 5' donneur d'épissage.....	4
I.2.2 Le site 3' d'épissage .....	4
I.2.3 La définition des exons et la définition des introns.....	5
I.3 Les composants et assemblage du spliceosome .....	5
I.3.1 Les particules UsnRNP.....	6
I.3.2 Assemblage du spliceosome .....	16
I.3.3 Les facteurs d'épissage autres que les protéines des UsnRNP .....	17
I.4 L'épissage : assemblage du spliceosome et catalyse de la réaction.....	27
I.4.1 Le complexe H.....	27
I.4.2 Le complexe E ou complexe d'engagement .....	27
I.4.3 Le complexe A ou pré-spliceosome.....	28
I.4.4 Le complexe B .....	29
I.4.5 Le complexe C .....	29
I.5 Lien entre épissage des pré-ARNm nucléaires, transport et système NMD.....	30
I.5.1 Lien entre transcription et épissage.....	30
I.5.2 Lien entre export, NMD et traduction : rôle de l'EJC .....	37
CHAPITRE II : Régulation de l'épissage des introns des ARN pré-messagers nucléaires.....	41
II.1 Généralités sur l'épissage alternatif.....	41

II.1.1 Les différents modes d'épissage alternatif .....	41
II.1.2 Les sites d'épissage alternatifs sont souvent des sites sous optimaux.....	44
II.2 Protéines nucléaires jouant un rôle majeur dans la régulation de l'épissage.....	46
II.2.1 Les différents types de domaines de fixation à l'ARN.....	46
II.2.2 Les protéines hnRNP .....	52
II.2.3 La régulation de l'épissage alternatif par les protéines SR.....	68
II.2.4 Les co-activateurs SRm300 et 160 et la protéine Tra-2 .....	71
II.3 Autres protéines et paramètres intervenant dans la régulation de l'épissage.....	72
II.3.1 Implication d'autres protéines régulatrices de l'épissage.....	72
II.3.2 Implication de la structure secondaire de l'ARN dans la régulation de l'épissage.	74
CHAPITRE III : Régulation de l'épissage du transcrit primaire du virus de l'immunodéficience humaine de type I.....	76
III.1 Description du virus VIH-1 .....	76
III.1.1 Evolution phylogénétique du virus VIH-1 .....	76
III.1.2 Le virion : génome et structure .....	79
III.2 Cycle de multiplication du virus VIH-1 et fonction des protéines dans le cycle viral.	85
III.2.1 La phase pré-intégrative .....	85
III.2.2 La phase post-intégrative .....	89
III.3 Effets de la machinerie d'épissage sur la dimérisation de l'ARN.....	99
III.4 Régulation de l'épissage de l'ARN viral.....	101
III.4.1. Production des ARNm viraux .....	101
III.4.2. Mécanisme d'épissage de l'ARN viral.....	104
OBJECTIFS .....	116
RESULTATS : .....	118
CHAPITRE I Recherche de nouvelles protéines régulant l'épissage alternatif au site accepteur d'épissage A7. ....	118
I. Stratégie utilisée pour l'étude .....	118

II. Purification de complexes ARN SLS2-A7 / protéines d'extrait nucléaire de cellules HeLa.....	119
II.1 Composition de l'ARN hybride utilisé pour la chromatographie d'affinité.....	119
II.2 Formation des complexes ribonucléoprotéiques (RNP) sur l'ARN fixé à la matrice et élution.....	121
II.3 Analyse en spectrométrie de masse des complexes ARN SLS2-A7/protéines d'extrait nucléaire de cellules HeLa.....	122
III. DISCUSSION .....	125
CHAPITRE II Etude de l'interaction des protéines identifiées par spectrométrie de masse avec la tige-boucle contenant le site d'épissage A7 du virus VIH-1. ....	126
I. Objectifs de l'étude .....	126
II. Au moins quatre des protéines identifiées se lient directement à l'ARN SLS2-A7.....	126
III Sites potentiels de fixation des protéines nucléoline et hnRNP K identifiées sur l'ARN SLS2-A7.....	127
III.1 La nucléoline .....	127
III.2 La protéine hnRNP K.....	132
IV. Etude de l'efficacité de fixation des protéines nucléoline et hnRNP K sur l'ARN SLS2-A7 en retard sur gel. ....	132
IV.1 Etude de l'affinité de la nucléoline sur l'ARN SLS2-A7. ....	133
IV.2 Etude de l'affinité de la protéine hnRNP K pour l'ARN SLS2-A7.....	137
V. La protéine hnRNP K présente plusieurs sites de fixation sur l'ARN autour du site A7. ....	139
VI. Fonctions possibles de la fixation des protéines hnRNP K et nucléoline près du site A7 accepteur d'épissage.....	142
VII. Discussion.....	142
CHAPITRE III Effets des protéines nucléoline, hnRNP H et hnRNP K sur le taux d'utilisation de sites accepteurs d'épissage du virus VIH-1.....	146
I. Etude de l'effet de la nucléoline et des protéines hnRNP H et hnRNP K sur l'épissage <i>in vitro</i> de l'intron entre les sites d'épissage D4-A7 de VIH-1.....	146

I.1. Effet sur l'épissage de la protéine hnRNP H. ....	148
I.2 Effet sur l'épissage de la nucléoline. ....	148
I.3 Effet sur l'épissage de la protéine hnRNP K. ....	149
I.3.1 La protéine hnRNP K recombinante active l'épissage d'ARN pré-messagers témoins. ....	149
I.3.2 La protéine hnRNP K recombinante diminue l'efficacité d'épissage alternatif du site A7 de l'ARN pré-messager L47 WT.....	153
I.3.3 L'effet inhibiteur de hnRNP K est maintenu indépendamment des modifications de séquence de l'ARN.....	155
I.3.4 La régulation de l'épissage alternatif par la protéine hnRNP K est bien liée au site 3' accepteur A7. ....	164
II. Etude de l'effet de la protéine hnRNP K recombinante sur l'efficacité d'utilisation des sites accepteurs d'épissage des sites A1, A2, A3 et A5 de VIH-1 <i>in vitro</i> . ....	166
II.1 Etude de l'effet de la protéine recombinante hnRNP K sur l'épissage <i>in vitro</i> du site accepteur A1 de VIH-1. ....	166
II.2 Etude de l'effet de la protéine recombinante hnRNP K sur l'utilisation <i>in vitro</i> du site accepteur A2 de VIH-1. ....	169
II.2.1 La protéine hnRNP K active l'utilisation du site accepteur A2 <i>in vitro</i> et un effet synergique est observé lorsqu'elle est associée à la protéine ASF/SF2. ....	170
II.2.2 Etude du mécanisme d'action de la protéine hnRNP K sur l'utilisation du site accepteur A2.....	172
II.3 Etude de l'effet de la protéine recombinante hnRNP K <i>in vitro</i> sur l'utilisation des sites accepteurs A3 à A5 de VIH-1. ....	175
II.4 Les effets activateurs de la protéine hnRNP K observés sur l'épissage des ARN C1, C2 et C3 s'exercent sur les sites accepteurs. ....	177
III. Discussion et perspectives.....	179
III.1 Discussion .....	179
III.2 Perspectives : .....	182

CHAPITRE IV Etude du mécanisme d'inhibition de l'épissage au site accepteur A7 par la protéine hnRNP K et identification de nouvelles protéines potentiellement impliquées dans cette régulation. ....	183
I. L'addition de protéine recombinante hnRNP K change le profil de fixation des protéines sur l'ARN SLS2 lors de pontage aux UV. ....	183
II. La protéine hnRNP K intervient-elle dans la fonction inhibitrice de hnRNP A1 sur l'épissage <i>in vitro</i> de l'intron D4-A7 par hnRNP K ?.....	184
II.1 L'addition de protéine hnRNP K à l'extrait nucléaire renforce la fixation de la protéine hnRNP A1 sur l'ARN SLS2. ....	184
II.2 L'effet inhibiteur de la protéine hnRNP K sur l'efficacité d'épissage au site A7 peut être le résultat d'une fixation coopérative des deux protéines, hnRNP K et hnRNP A1, sur l'ARN. ....	186
II.3 Les effets de hnRNP A1 et hnRNP K sont cumulatifs lors de l'épissage <i>in vitro</i> au site A7. ....	187
III. Identification de protéines interagissant à la fois avec l'ARN SLS2-MS2 et la protéine hnRNP K.....	188
III.1 Objectifs de l'étude .....	188
III.2 Principe de la double purification permettant l'isolement de protéines interagissant avec l'ARN SLS2-MS2 et la protéine hnRNP K.....	189
III.3 Identification en spectrométrie de masse des protéines de l'extrait nucléaire de cellules HeLa interagissant avec la protéine His-hnRNP K et l'ARN SLS2-MS2:.....	191
III.3.1 Conditions d'analyse. ....	191
III.3.2 Identification des protéines. ....	191
III.3.3 Conclusion.....	194
CHAPITRE V Effets <i>in cellulo</i> de la protéine hnRNP K sur l'épissage de VIH-1. ....	197
I. Etat des connaissances au départ et stratégie utilisée pour l'étude.....	197
II. La diminution du taux de la protéine hnRNP K dans les cellules HeLa affecte peu le profil d'épissage <i>in cellulo</i> de l'ARN $\Delta$ PSP. ....	198
III. La surexpression de la protéine hnRNP K dans les cellules HeLa a un effet important sur le profil d'épissage <i>in cellulo</i> . ....	201

IV. Résultats et discussions.....	205
CHAPITRE VI Relation entre épissage et dimérisation de l'ARN de VIH-1. <i>Collaboration avec le Dr. Jean-Christophe Paillart, Laboratoire Architecture et Réactivité de l'ARN CNRS-UPR 9002, Strasbourg.</i> .....	209
I. Etat des connaissances au départ. ....	209
II. Stratégie utilisée pour l'étude.....	212
III. Résultats et discussion.....	212
CONCLUSION GENERALE ET PERSPECTIVES .....	222
MATERIELS ET METHODES .....	228
A. MATERIEL.....	228
CHAPITRE I Souches utilisées.....	228
I.1. Souches d' <i>E. coli</i> utilisées .....	228
I.2 Les cellules d'organismes eucaryotes .....	228
I.3 Milieux de culture, tampons et solution d'usage courant .....	228
CHAPITRE II Constructions génétiques .....	230
II.1. Constructions renfermant des séquences codant des protéines pour l'expression en <i>E. Coli</i> .....	230
II.2. Constructions pour l'épissage utilisées au cours de ce travail.....	230
II.3 Construction pΔPSP pour l'expression de VIH-1 <i>in vivo</i> .....	231
II.4 Oligonucléotides utilisés.....	231
B. METHODES .....	234
CHAPITRE I : Production et purification des protéines recombinantes dans <i>E. coli</i> .....	234
I.1 Protéine hnRNP A1.....	234
I.1.1 Sonication .....	234
I.1.2 Préparation de l'échantillon pour la purification .....	234
I.1.3 Etapes de purification .....	234
I.2 Protéine MS2-MBP.....	237
I.2.1 Sonication .....	237

I.2.2 Etapes de purification .....	237
I.3 Protéine hnRNP K.....	239
I.3.1 Sonication .....	239
I.3.2 Etapes de purification .....	239
I.4 Protéine hnRNP H.....	241
I.5 Protéine Nucléoline.....	241
CHAPITRE II : Techniques d'analyse des interactions ARN/protéines.....	242
II.1. Technique d'empreinte sur l'ARN en présence de protéine purifiée .....	242
II.1.1. Amplification par PCR des ADN matrices.....	242
II.1.2. Synthèse de transcrits froids courts .....	243
II.1.3. Principe de la technique d'empreinte des ARN.....	243
II.1.4 Conditions expérimentales de l'empreinte d'un ARN en présence de protéine purifiée .....	244
II.1.5 Séquençage de l'ARN .....	246
II.2. Technique de retard sur gel .....	247
II.2.1 Synthèse de transcrits radioactifs courts.....	247
II.2.2 Conditions expérimentales des retards sur gel .....	248
II.3. Pontages aux UV suivis d'immunoprécipitation .....	249
II.3.1. Pontages aux UV .....	249
II.3.2. Immunoprécipitation .....	250
CHAPITRE III : Méthodes spécifiques à l'étude de l'épissage <i>in vitro</i> .....	251
III.1 Synthèse de transcrits radioactifs <i>in vitro</i> .....	251
III.1.1 Préparation du fragment d'ADN à transcrire .....	251
III.1.2 Transcription <i>in vitro</i> .....	251
III.2 Epissage <i>in vitro</i> .....	252
III.2.1 Réaction d'épissage.....	253
III.2.2. Fractionnement sur gel et fluorographie .....	254

III.3 Détermination de l'efficacité d'épissage.....	254
CHAPITRE IV : Epissage in vivo.....	255
IV.1. Transfection et extraction ARN .....	255
IV.2. Analyse RT-PCR.....	255
CHAPITRE V : Dimérisation .....	257
CHAPITRE VI : Identification de nouvelles protéines en spectrométrie de masse.....	258
VI.1 Principe de la méthode de chromatographie d'affinité basée sur l'interaction ARN MS2-protéine MBP .....	258
VI.2 Paramètres critiques .....	258
VI.3 Liaison de l'ARN-MS2 avec la protéine MBP-MS2 .....	259
VI.4 Fixation sur les billes amylose .....	259
VI.5 Fixation des protéines nucléaires .....	259
VI.6 Elution.....	259
VI.7 Purification sur colonne Ni-NTA pour la double purification.....	260
VI.8 Analyse en spectrométrie de masse. ....	260
PUBLICATION: « Identification of protein partners of the human immunodeficiency virus 1 <i>tat/rev</i> exon 3 leads to the discovery of a new HIV-1 splicing regulator, protein hnRNP K” .....	
Soumis à RNA Biology.....	262
BIBLIOGRAPHIE .....	326





### REMERCIEMENTS

En premier lieu, je tiens à remercier la directrice du laboratoire, Madame Christiane Branlant pour son accueil dans le laboratoire et pour m'avoir fait confiance en me recrutant. Ses conseils, sa rigueur et son orientation constante vers des pistes qui nous ont permis d'aboutir à des résultats originaux, m'ont aidée tout au long du travail expérimental. Je la remercie vivement pour la formation scientifique qu'elle m'a apportée.

Je remercie Monsieur Iouri Motorine pour la direction de ma thèse. Par ses conseils et son attention continue sur l'évolution de notre travail, il a su me transmettre son enthousiasme pour la recherche et m'aider à trouver des solutions à tous les problèmes.

J'adresse toute ma gratitude à Madame Marie-Line Andreola et Monsieur James Stévenin d'avoir accepté d'être rapporteurs de ce mémoire, ainsi qu'à Monsieur Philippe Fossé pour m'avoir fait l'honneur de juger ce travail.

Je voudrais aussi remercier vivement Alain Van Dorsselaer, Sarah Sanglier-Cianférani, Laetitia Fouillen et Jean-Michel Saliou avec qui nous avons réalisé les analyses protéomiques. Un grand merci à Jean-Christophe Paillart pour m'avoir accueillie dans son laboratoire et Lucile Sinck de m'avoir initiée aux techniques de dimérisation.

Je remercie le SIDACTION d'avoir non seulement financé ma 4<sup>ème</sup> année de thèse mais aussi pour les journées scientifiques très enrichissantes qu'ils ont organisées et auxquelles ils m'ont conviée, ainsi que pour le travail immense qu'ils mènent tous les jours.

J'adresse ma profonde reconnaissance à Lilia pour m'avoir transmis tout son savoir faire et pour m'avoir aidée et conseillée durant toute ma thèse jusque dans la rédaction de ce manuscrit. Son soutien, sa disponibilité et son amitié ont toujours été un précieux réconfort.

Merci également aux autres occupants du Box 5, qui m'ont supportée pendant 4 ans et 10 mois. Merci à Christelle Aigueperse, pour sa présence et son aide pendant toute la durée de ma thèse. Merci à Gabrielle Bourgeois (Gaby) pour ta bonne humeur constante et tes tee-shirts roses qui mettaient de la joie dans le box.

Je veux également adresser un énorme merci à mon second box, le box 2 : à Lilia (promis je ferai des macarons) ; à Georges, grâce à qui je pouvais discuter VIH, échanger des publis et des avis sur les manip, ainsi que de longues discussions scientifiques fructueuses ; à la présidente Flo et sa répartie digne d'une brune dont je suis jalouse ; à Isa pour les discussions scientifiques, ses corrections, sa bonne humeur et son écoute ; à Valentin, pour ses blagues qui malgré tout me font rire ; à Aileen d'avoir encouragé maintes fois le petit scarabée que je suis et pour nos discussions très enrichissantes et enfin à Audrey, pour beaucoup et bien plus encore.

Un grand merci aussi à mon 3<sup>ème</sup> box, Anne et Alex pour votre écoute et vos conseils et Alex pour tes câlins.

Je remercie vivement l'équipe des correcteurs qui ont traqué sans relâche les fautes d'orthographe et les non-sens, et qui ont hautement contribué à améliorer le manuscrit (par ordre alphabétique) : Aileen, Alex, Anne, Audrey, Georges et Lilia bien sûr.

## REMERCIEMENTS

---

Il est dur de n'oublier personne, merci à : Maxime, Flo box 4, Sylvain, Fabrice, Xavier, Karine, Rémi, je remercie tous les membres du laboratoire AREMS qui ont participé de près ou de loin à l'élaboration de ce travail.

Un grand merci au banc de *Gadus morhua*, Aileen, Anne, Alex et Audrey. Leur amitié a été le meilleur des réconforts et exutoire lorsque la fatigue et la lassitude rendaient toujours plus improbable ma soutenance.

Je remercie aussi mes amis bruxellois, sud-africains, parisiens, isariens et lyonnais pour leur soutien inconditionnel, même dans les pires moments.

J'ai une pensée tendre pour celui qui supporte mes humeurs dans la vie civile et je le remercie très fort, pour des raisons que la raison ignore.

Je remercie mes parents pour leur soutien inconditionnel au cours de ces longues années d'études et d'avoir attendu si longtemps avant de pouvoir dire "ma fille a fini ses études" en me poussant chaque jour un peu plus davantage.

# **LISTE DES ABREVIATIONS**

# Abréviations

A adénine

ADN acide désoxyribonucléique

ADNc acide désoxyribonucléique complémentaire

AMV « avian myeloblastosis virus »

ARN acide ribonucléique

ARNm acide ribonucléique messenger

ARNt acide ribonucléique de transfert

ASF/SF2 « alternative splicing factor/splicing factor 2 »

ATP adénosine triphosphate

BSA « bovin serum albumin »

C cytosine

Ci/mCi curie/millicurie

cpm coups par minute

Da/kDa dalton/kilodalton

dATP désoxyadénosine triphosphate

dCTP désoxycytosine triphosphate

ddNTP didésoxyribonucléotide

dGTP désoxyguanosine triphosphate

DMS diméthylsulfate

DNase désoxyribonucléase

dNTP désoxyribonucléotide triphosphate

DTT dithiothréitol

dTTP désoxythymidine triphosphate

EDTA éthylène diamine tétra-acétate

EN extrait nucléaire

G guanine

HeLa Henrietta Lacks

HEPES acide (N-[2-hydroxyéthyl] piperazine-N'-[2-éthanesulfonique])

hnRNP « heterogeneous ribonucleoprotein »

VIH « Virus de l'Immunodéficience Humaine »

kb kilobase

m<sub>2,2,7</sub>G (2, 2, 7) tri-méthyl guanosine

m<sub>7</sub>G 7-méthyl guanosine

nt(s) nucléotide(s)

ORF « open reading frame » (phase ouverte de lecture)

pb paire de bases

PCR « polymerase chain reaction »

PPT « polypyrimidine tract »

p/v poids/volume

qsp quantité suffisante pour

R purine

RNase ribonucléase

rNTP ribonucléotide triphosphate

rpm rotation par minute

RT transcriptase inverse

S unité de sédimentation Svedberg

RS arginine-sérine

sb simple-brin

SDS sodium dodécylsulfate  
UTP uridine triphosphate  
SA site accepteur  
SD site donneur  
SIV « simian immunodeficiency virus »  
T thymine  
TBE tampon tris-borate-EDTA  
TE tampon tris-EDTA  
Tris tris (hydroxyméthyl) aminométhane  
U uracile  
UsnRNA « uridine rich small nuclear ribonucleic acid »  
UsnRNP « uridine rich small nuclear ribonucleoprotein particle »  
UV ultra-violet  
*vs versus*  
v/v volume/volume  
Y pyrimidine  
3'ss « 3' splice site » (site 3' d'épissage)  
5'ss « 5' splice site » (site 5' d'épissage)



# **INTRODUCTION**



### AVANT-PROPOS

Le virus d'immunodéficience humaine (VIH-1) est responsable du Syndrome d'Immunodéficience Acquis (SIDA). Ce virus entraîne l'apoptose des lymphocytes T, abolit progressivement la capacité du corps à combattre les infections et peut conduire à l'apparition de cancers. Le stade SIDA est la dernière étape suivant l'infection et peut conduire à la mort de l'organisme, des suites d'infection par des agents pathogènes.

Dans les pays les plus gravement affectés, le virus du VIH-1 a réduit l'espérance de vie de plus de 20 ans, ralenti la croissance économique et augmenté la pauvreté des foyers touchés. Dans l'Afrique Sub-saharienne, l'épidémie a rendu orphelins 12 millions d'enfants de moins de 18 ans. La pyramide des âges dans différentes populations de l'Afrique sub-saharienne a été déformée par le VIH-1, avec des conséquences potentiellement périlleuses pour le transfert de savoir et des valeurs d'une génération à la prochaine. En Asie, bien que les taux d'infection soient plus faibles qu'en Afrique, le VIH-1 cause une plus grande perte de productivité que n'importe quelle autre maladie, et va probablement pousser 6 millions de foyer supplémentaires dans la pauvreté avant 2015, à moins que des réponses nationales soient renforcées (Commission sur le SIDA en Asie, 2008). Le VIH reste un problème de santé global de dimension sans précédent. Inconnu il y a 27 ans, le VIH a déjà causé la mort de 25 millions de personnes à travers le monde. Globalement, il y avait environ 33,3 millions de personnes [31,4 millions–35,3 millions] vivant avec le VIH en 2009. Le nombre annuel de nouvelles infections a diminué, de 3 millions en 2001 à 2,6 millions en 2009. En tout, 1,8 millions de personnes sont mortes à cause du SIDA en 2009, comparé avec un nombre estimé de 1,7 million de décès en 2001 (Sources rapport UNAIDS 2010).

## INTRODUCTION

### CHAPITRE I : Epissage des ARN pré-messagers nucléaires chez les mammifères.

L'expression génétique chez les organismes eucaryotiques requiert le retrait de séquences introniques de transcrits primaires. Ces introns présents dans les transcrits primaires (pré-ARNm) sont excisés dans le noyau avant que l'ARNm mature ne soit exporté vers le cytoplasme. Les gènes eucaryotes peuvent contenir des dizaines d'introns dont la taille varie de plusieurs dizaines de nucléotides à des dizaines de milliers.

Tous les processus d'élimination des séquences excédentaires appelées introns sont rassemblés sous le terme épissage. Ainsi, au cours de la réaction d'épissage, les séquences introniques des ARN pré-messagers sont éliminées et les séquences codantes, dénommées exons, sont jointes les unes aux autres pour former l'ARN messenger.

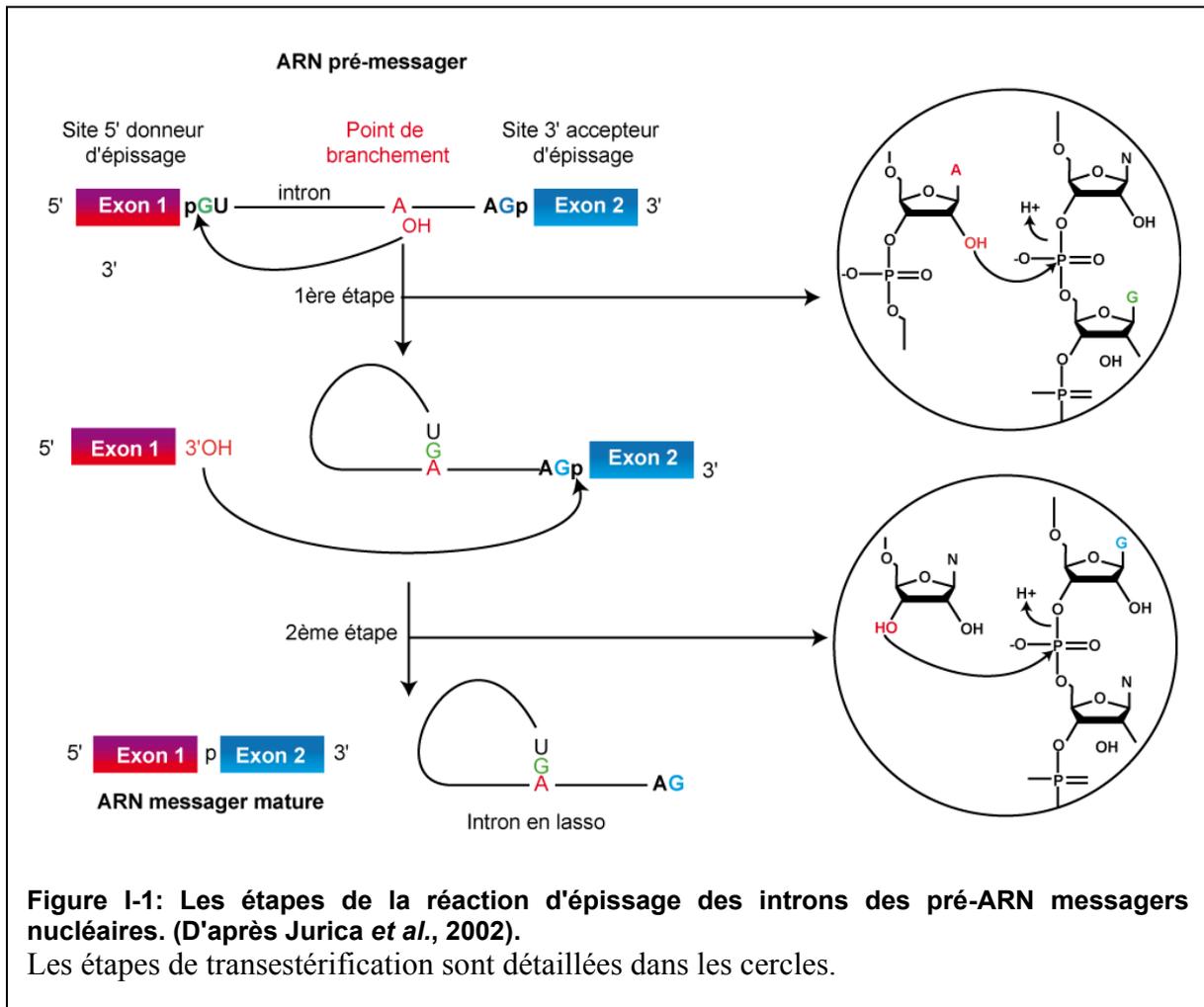
La production de différents ARNm à partir d'un même transcrit permet de produire différentes protéines. Ainsi l'épissage alternatif des transcrits des gènes nucléaires eucaryotes augmente la diversité des protéines qui peuvent être codées par le génome. Il existe de nombreux types d'épissage alternatifs observés, le plus commun étant l'« exon skipping » ou « saut d'exon ».

L'épissage alternatif résulte de la modulation du choix des sites d'épissage des pré-ARNm par des facteurs trans-activateurs et des éléments cis-activateurs. En effet, peu de séquences communes définissent les sites d'épissage. La reconnaissance de ces sites dépend largement de la fixation de facteurs d'épissage en *trans*.

L'un des principaux sujets d'étude actuel sur l'épissage consiste à comprendre comment les séquences introniques des pré-ARNm sont détectées et comment les variations de concentrations de facteurs nucléaires peuvent favoriser l'utilisation de l'un ou l'autre des sites alternatifs.

#### I.1 La réaction biochimique de l'épissage

L'épissage des pré-ARNm nucléaires implique deux étapes séquentielles de transestérification qui permettent d'éliminer les introns (Figure I-1).

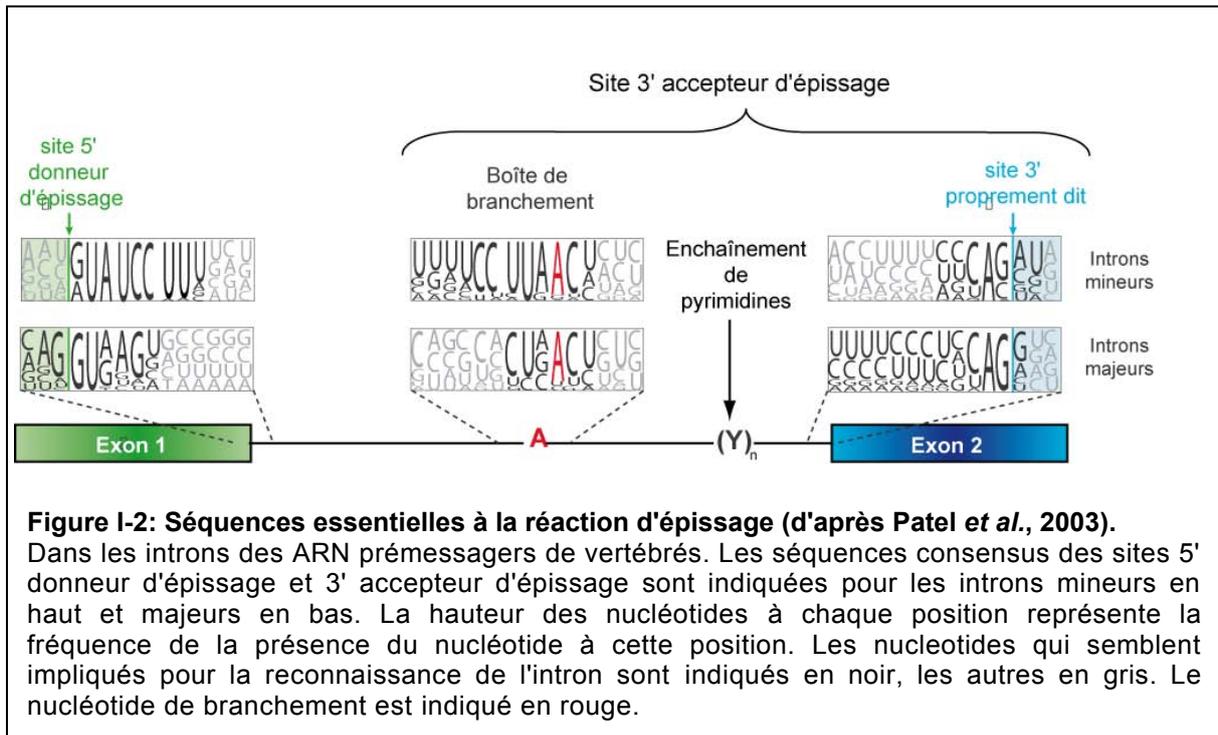


La première étape commence par l'attaque nucléophile du phosphate en 5' du premier exon par l'hydroxyle en 2' d'un résidu adénosine appartenant à la boîte de branchement de l'intron. Cette réaction génère deux intermédiaires : l'exon avec un 3' hydroxyle libre et une molécule intron-exon2 dans laquelle le nucléotide du point de branchement est impliqué dans 3 liaisons phosphodiester, ce qui forme une structure dite en « lasso ».

Lors de la seconde étape, l'hydroxyle en 3' de l'exon attaque le phosphate en 5' du second exon ce qui conduit à la ligation des 2 exons et à la libération de l'intron sous forme de lasso.

## I.2 Les séquences définissant les introns

Peu après la découverte de l'épissage en 1977, la présence d'une séquence caractéristique conservée aux extrémités des introns des ARNm cellulaires et viraux a été observée (Breathnach *et al.*, 1978). La présence de nucléotides GT aux sites 5' d'épissage et celle de nucléotides AT aux sites 3' a donné naissance à la règle GT-AG (Figure I-2).



Une étape critique pour le pré-ARNm est la reconnaissance des sites d'épissage 5' et 3'.

### 1.2.1 Le site 5' donneur d'épissage

Il est défini par un élément simple de 8 nucléotides AG/GURAGU (R = purine et / = jonction exon/intron), dans lequel le dinucléotide **GU** en 5' de l'intron est le motif le plus conservé. L'appariement de bases entre cet élément et le snRNA U1 joue un rôle clé dans l'épissage. Le site donneur n'est pas toujours conforme au consensus car il peut avoir une séquence dégénérée. De plus, d'autres éléments additionnels comme des inhibiteurs d'épissage et des complexes activateurs participent à la sélection du site d'épissage. Ainsi, la reconnaissance d'un site 5' dégénéré par le spliceosome n'est pas dépendante simplement de sa séquence mais dépend aussi des séquences environnant le site.

### 1.2.2 Le site 3' d'épissage

Il est défini par 2 séquences situées dans les 40 nucléotides en amont de la liaison intron/exon 3' (le point de branchement et la séquence polypyrimidine) ainsi que par la séquence à la jonction intron/exon 3'.

La boîte de branchement est située 18 à 40 nt en amont du site 3' d'épissage. Sa séquence consensus établie pour les introns de vertébrés, est très dégénérée : YNYURAC (Y = pyrimidine). Le résidu adénosine (en gras) correspondant au point de branchement est

responsable de la première réaction de transestérification. Dans quelques cas, ce résidu peut être une cytosine ou une uridine.

La séquence polypyrimidine correspond à un enchaînement de 10 à 20 résidus pyrimidines, les résidus U étant plus favorables à sa fonction. Elle est localisée immédiatement en amont du site 3' accepteur d'épissage.

Enfin, le site 3' accepteur d'épissage, situé à la bordure intron/exon 3', est caractérisé par une courte séquence conservée XAG/G ou X=U/G.

### **1.2.3 La définition des exons et la définition des introns**

La taille des introns a augmenté au cours de l'évolution de manière parallèle à l'augmentation de la taille des génomes (Marais *et al.*, 2005). En effet, la taille des introns est influencée par : l'insertion d'éléments transposables (Bartolome *et al.*, 2002), la présence d'éléments régulateurs contrôlant l'expression des gènes (Bergman *et al.*, 2001) et la présence de séquences non codantes (Mattick, 2001). La taille des introns a aussi une influence sur leur mode de reconnaissance (Berget, 1995). Dans les génomes ayant de petits introns comme la levure ou la drosophile, c'est l'intron qui est l'élément de reconnaissance. Tandis que chez les vertébrés où les introns sont grands, l'élément initial de reconnaissance est l'exon (Berget, 1995).

La reconnaissance des sites d'épissage par le mécanisme de définition des introns cesse quand la taille de l'intron est entre 200 et 250 nucléotides et cette reconnaissance se fait alors par le mécanisme de définition des exons (Fox-Walsh *et al.*, 2005).

La reconnaissance initiale des jonctions exon/intron est basée sur l'interaction directe entre la snRNP U1 avec le site 5', de la protéine U2AF avec la séquence polypyrimidine et de la snRNP U2 avec le point de branchement. La spécificité de ces séquences dépend de la qualité des interactions pré-ARNm/snRNA et de l'affinité de U2AF pour la séquence polypyrimidine. Les sites d'épissage sont classés selon leur complémentarité avec le snRNA U1 au niveau du site 5' et du snRNA U2 au niveau de la boîte de branchement. Une plus forte complémentarité avec le snRNA U1 et une séquence polypyrimidine longue sont des facteurs favorables à l'efficacité d'épissage (pour revue, Reed, 2000).

### **1.3 Les composants et assemblage du spliceosome**

L'assemblage des complexes spliceosomaux est un processus dynamique mettant en œuvre de nombreux réarrangements des interactions ARN-ARN, ARN-protéine et protéine-protéine.

Les UsnRNP (U1, U2, U4/U6 et U5) et environ 300 facteurs nucléaires s'assemblent de manière dynamique sur le pré-ARNm pour former les complexes spliceosomaux successifs qui catalysent les deux réactions de transestérification. Avant de participer à l'épissage, les UsnRNP doivent elles-mêmes être assemblées (pour revue, Wahl *et al.*, 2009).

### **I.3.1 Les particules UsnRNP.**

L'élimination des introns requiert les particules UsnRNP (small nuclear ribonucleoprotein particles), U1, U2, U4/U6 et U5 qui sont des éléments majeurs pour la formation des complexes spliceosomaux. Chaque particule UsnRNP est composée d'un ou plusieurs snRNA (small nuclear RNA) (U1, U2, U4, U5 et U6 snRNA) riche en résidus uridines, d'un ensemble de 7 protéines communes, les protéines Sm et de plusieurs protéines associées spécifiques à chaque UsnRNP. L'assemblage de toutes les UsnRNP commence par la transcription du UsnRNA (pour revue, Patel *et al.*, 2008).

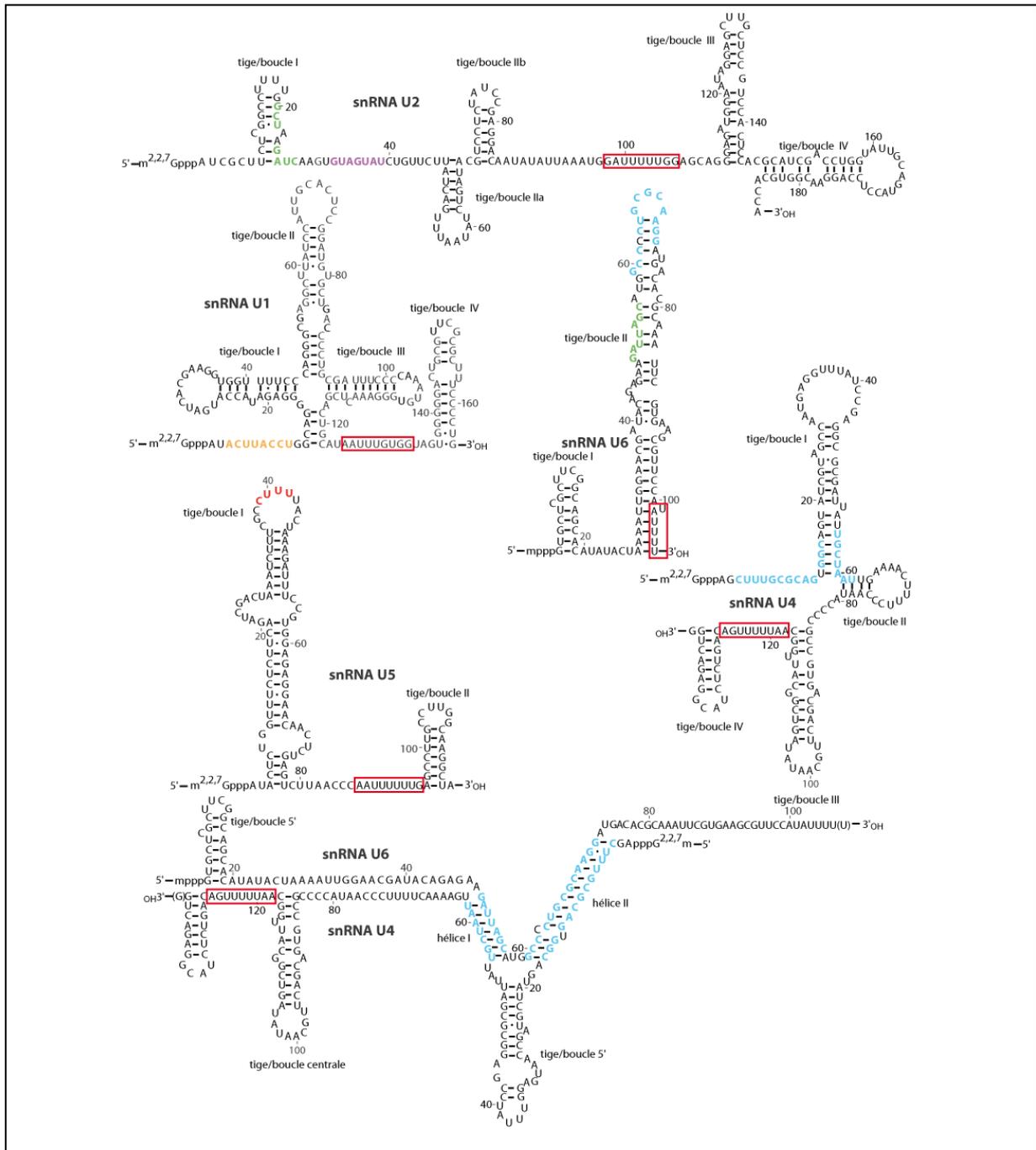
#### **I.3.1.1 Biogénèse des UsnRNA chez les mammifères.**

##### ***a. Cas des snRNA U1, U2, U4 et U5***

Les snRNA U1, U2, U4 et U5 sont transcrits dans le noyau par l'ARN polymérase II et possèdent une coiffe monométhylée ( $m^7G$ ) (Figure I-3) (Branlant *et al.*, 1980; Krol *et al.*, 1981; Branlant *et al.*, 1982; Branlant *et al.*, 1983; pour revue, Wahl *et al.*, 2009; Mougin *et al.*, 2002). Le pré-snRNA est assemblé en macro complexe permettant l'export dans le cytoplasme. Ce macro complexe comprend le complexe nucléaire de liaison à la coiffe (CBC) composé des protéines CBP20 et CBP80.

Les protéines Sm se lient au site de liaison Sm conservé des snRNA spliceosomaux (séquence consensus RAUUU/GUUGR) (Branlant *et al.*, 1982). La liaison des protéines Sm est un pré-requis pour l'hyperméthylation de la coiffe monométhylée  $m^7G$  en coiffe triméthylée (2,2,7-triméthylguanosine ou  $m^3G$ ) par la Tgs méthyltransférase, une méthylase associée au complexe SMN (Mattaj, 1986; Will *et al.*, 2001). La coiffe TMG qui vient d'être synthétisée et les protéines associées Sm constituent le signal permettant l'adressage nucléaire des pré-UsnRNP formés dans le cytoplasme. Ceci permet l'interaction des pré-UsnRNP avec la protéine de transport snurportin-1, qui permet leur passage au travers des pores nucléaires. Ces particules sont alors dirigées dans les corps de Cajal où les UsnRNA subissent des modifications et où les protéines Sm sont absentes. Les snRNP spliceosomales matures sont soit immédiatement assemblées en spliceosomes, soit accumulées dans la région d'interchromatine dans des particules appelées "speckles", qui sont des structures

généralement localisées près des gènes transcrits. Elles constituent probablement des sites de stockage pour des snRNP et/ou fonctionnent dans le recyclage des snRNP (Lamond *et al.*, 2003).



**Figure I-3: Séquences et structures secondaires des UsnRNA humains majeurs.**

Les structures secondaires représentées ci-dessus ont été établies à partir de celles déterminées expérimentalement chez le rat par : (Branlant *et al.*, 1980; Mount *et al.*, 1981) pour le snRNA U1, (Branlant *et al.*, 1982) pour le snRNA U2, (Krol *et al.*, 1981) pour le snRNA U4 et (Branlant *et al.*, 1983) pour le snRNA U5. La structure secondaire du snRNA U6 humain a été établie par (Mougin *et al.*, 2002). Enfin, la structure secondaire du duplex U4/U6 a été établie par (Brow *et al.*, 1988; Mougin *et al.*, 2002). Le site Sm permettant la fixation des protéines Sm (Branlant *et al.*, 1982) ou LSm (Achsel *et al.*, 1999) est encadrée en rouge dans les différents UsnRNA. Les nucléotides en orange sont nécessaires lors de la réaction d'épissage car impliqués dans l'appariement avec le site 5' (snRNA U1), ceux en violet s'apparient à la boîte de branchement (snRNA U2), ceux en rouge s'associent avec les deux exons. Les nucléotides en vert sont impliqués dans les appariements entre les snRNA U2-U6, ceux en bleu sont impliqués dans les appariements entre les snRNA U4-U6. Les modifications post-transcriptionnelles des UsnRNA ne sont pas indiquées sur cette figure.

### ***b. Cas du snRNA U6***

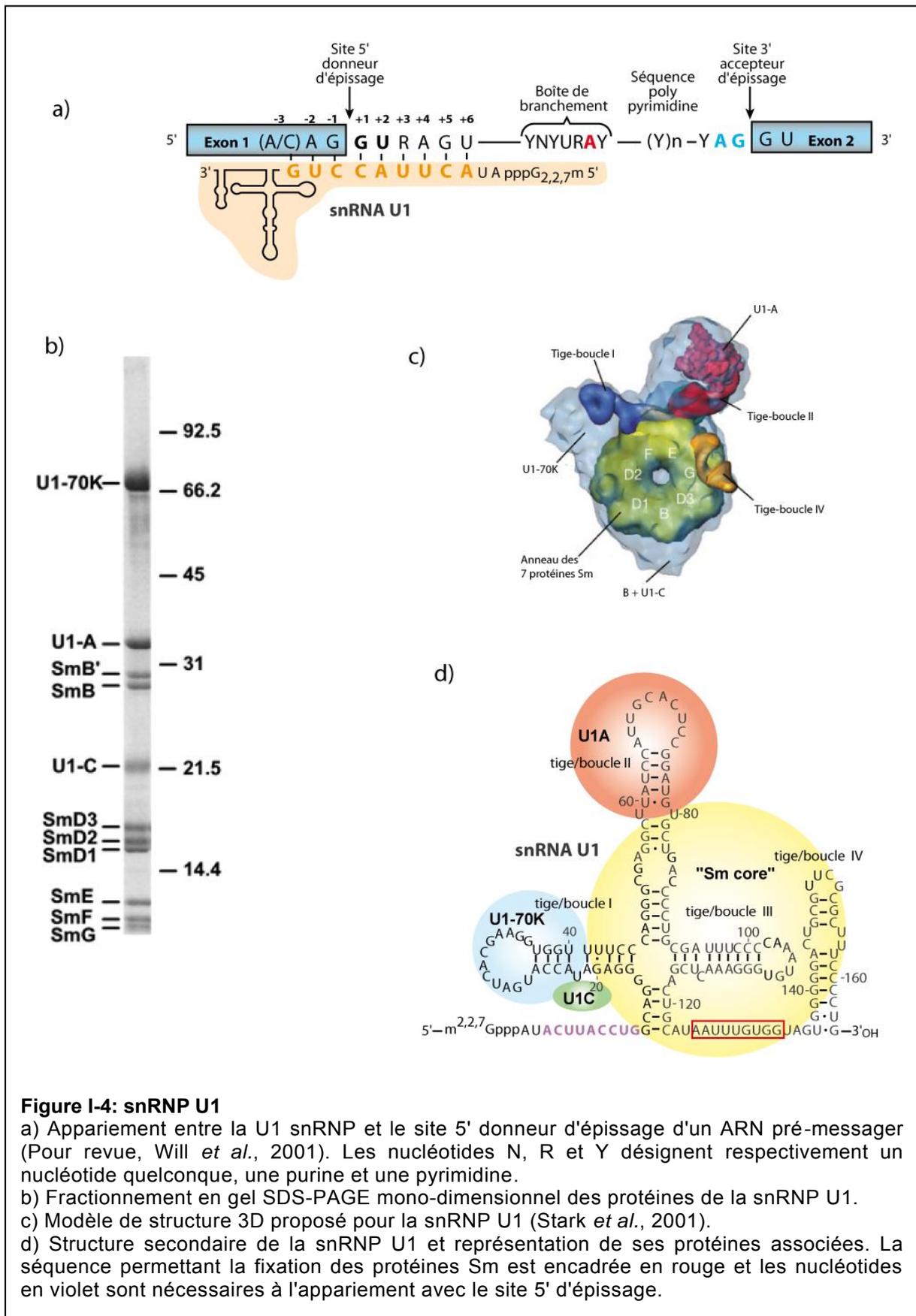
Contrairement aux autres snRNA, le snRNA U6 est transcrit par l'ARN polymérase III et sa coiffe a une structure mono  $\gamma$ -méthylée. La protéine La se lie à l'extrémité 3' des transcrits U6 nouvellement synthétisés, sur une séquence riche en résidus U, et les stabilise en contribuant ainsi à l'assemblage de la snRNP U6 (Xue *et al.*, 2000). Le snRNA U6 est modifié par pseudouridylation et 2'-O-méthylation dans le nucléole, grâce à des snoRNA guides de modification (Tycowski *et al.*, 1998; Ganot *et al.*, 1999).

### **I.3.1.2 La snRNP U1**

La particule snRNP U1 humaine (small nuclear ribonucleoprotein), qui est un composant du spliceosome, consiste en un snRNA U1 et 10 protéines différentes : 7 protéines Sm (B/B', D1, D2, D3, E, F et G et les 3 protéines spécifiques de U1, U1-70 kDa, U1-A et U1-C (Figure I-4).

Le snRNA U1 est un snRNA abondant de 164 nucléotides, observé pour la première fois dans des noyaux de cellules de mammifères (Hodnett *et al.*, 1968; Weinberg *et al.*, 1968) . L'extrémité 5' du snRNA U1 contient une séquence très conservée (ACUUACCUG) de la drosophile aux vertébrés (Mount *et al.*, 1981) et va être cappée, méthylée et pseudouridylée : m<sup>2</sup>, 2', 7Gppp AmUmAC $\psi$  $\psi$ ACCUG. Cette extrémité est complémentaire au site 5' donneur d'épissage des introns épissés par les spliceosomes majeurs (Branlant *et al.*, 1981; Mount *et al.*, 1983). Le snRNA U1 se lie généralement à la séquence allant des positions -3 à +6 du site 5' donneur d'épissage. L'appariement peut être légèrement étendu lorsque la séquence en 5' de l'intron n'est pas optimale (Cohen *et al.*, 1994).

La snRNP U1 contient 3 protéines spécifiques nommées 70 kDa, A et C, qui s'associent avec le snRNA U1. Les protéines U1-70 kDa et U1-A, qui contiennent des domaines RRM classiques, se lient directement aux tiges-boucles I et II du snRNA U1, respectivement, tandis que la protéine U1-C s'associe via des interactions protéine-protéine avec U1-70 kDa et/ou une ou plusieurs protéine Sm (Du *et al.*, 2002). La protéine U1-70 kDa contient en plus un domaine RS qui lui permet d'interagir avec la famille des protéines SR (Query *et al.*, 1989; Eperon *et al.*, 1993).



**Figure I-4: snRNP U1**

a) Appariement entre la U1 snRNP et le site 5' donneur d'épissage d'un ARN pré-messager (Pour revue, Will *et al.*, 2001). Les nucléotides N, R et Y désignent respectivement un nucléotide quelconque, une purine et une pyrimidine.

b) Fractionnement en gel SDS-PAGE mono-dimensionnel des protéines de la snRNP U1.

c) Modèle de structure 3D proposé pour la snRNP U1 (Stark *et al.*, 2001).

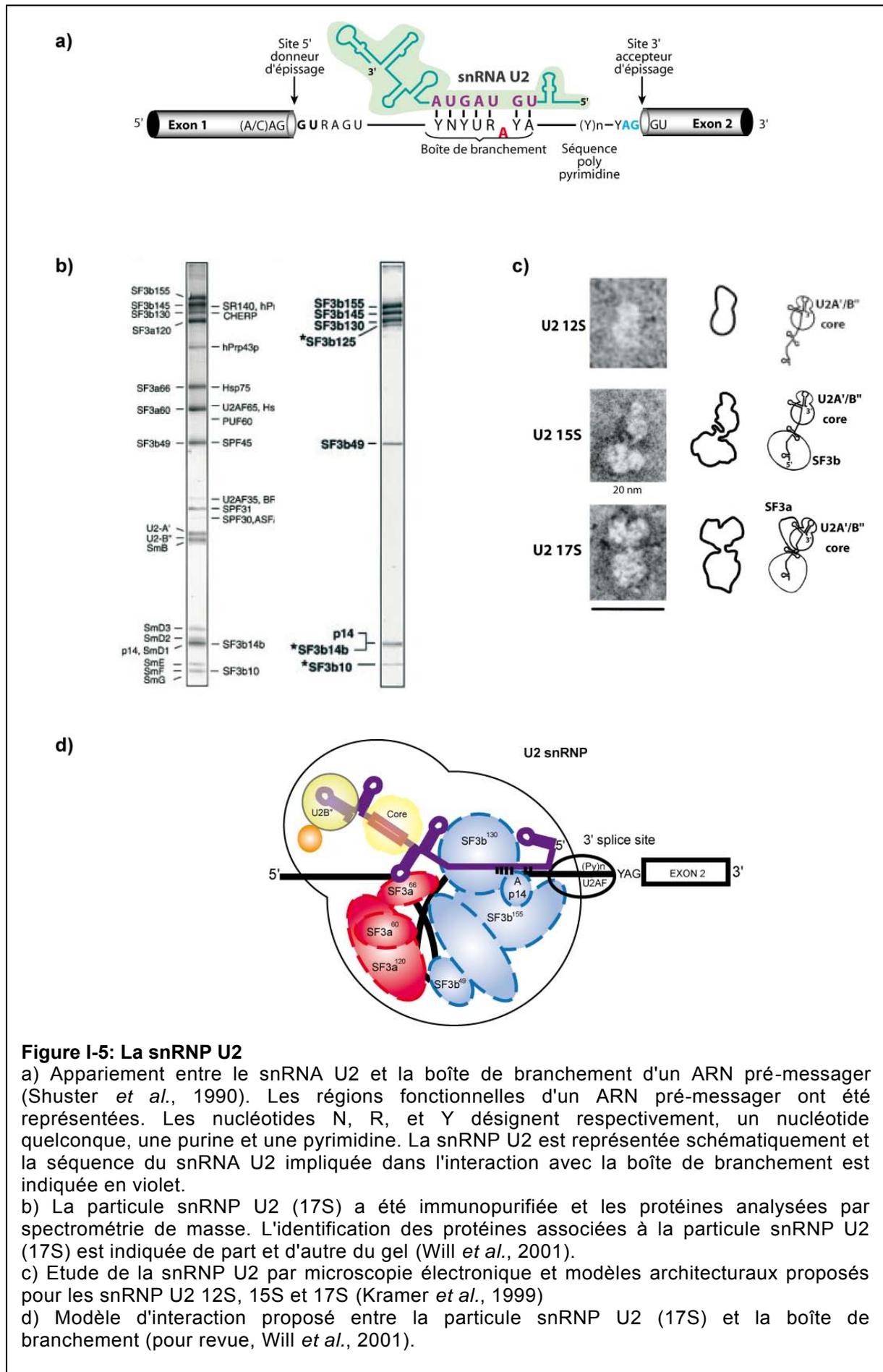
d) Structure secondaire de la snRNP U1 et représentation de ses protéines associées. La séquence permettant la fixation des protéines Sm est encadrée en rouge et les nucléotides en violet sont nécessaires à l'appariement avec le site 5' d'épissage.

### I.3.1.3 La snRNP U2

Le rôle majeur de la snRNP U2 (Figure I-5) dans l'épissage est de reconnaître le point de branchement dans l'intron du pré-ARNm non épissé. Quand des complexes d'épissage de cellules HeLa sont centrifugés sur gradient de glycérol, la snRNP U2 sédimente à 12S si la concentration de sels est élevée ou à 17S si la concentration de sels est faible. La particule 12S comprend le snRNA U2, les 7 protéines Sm et les protéines spécifiques du U2A' et U2B'' (Will *et al.*, 2001). L'équipe de A. Krämer a identifié les 2 complexes protéiques, SF3a et SF3b, requis pour cibler la snRNP U2 12S vers le pré-ARNm (Brosi *et al.*, 1993). SF3a correspond à 3 des protéines spécifiques de la snRNP U2 17S (les sous-unités de 60, 66 et 120 kDa), et SF3b contient quelques unes, voire toutes les protéines restantes spécifiques de la snRNP U2. SF3b peut se lier en premier à la snRNP 12S pour former un complexe intermédiaire, et cela facilite l'addition de SF3a pour former la particule 17S complète.

La majorité des particules snRNP U2 présentes dans un extrait nucléaire sont de type 17S et contiennent les protéines Sm, les protéines U2A' et U2B'' et les facteurs SF3b puis SF3a (Will *et al.*, 2002).

Le facteur U2AF se fixe sur la séquence polypyrimidine et le site 3' accepteur d'épissage dans les étapes précoces d'assemblage du spliceosome, de manière indépendante à la snRNP U2 (Will *et al.*, 2002).



### I.3.1.4 La tri-snRNP U4/U6.U5

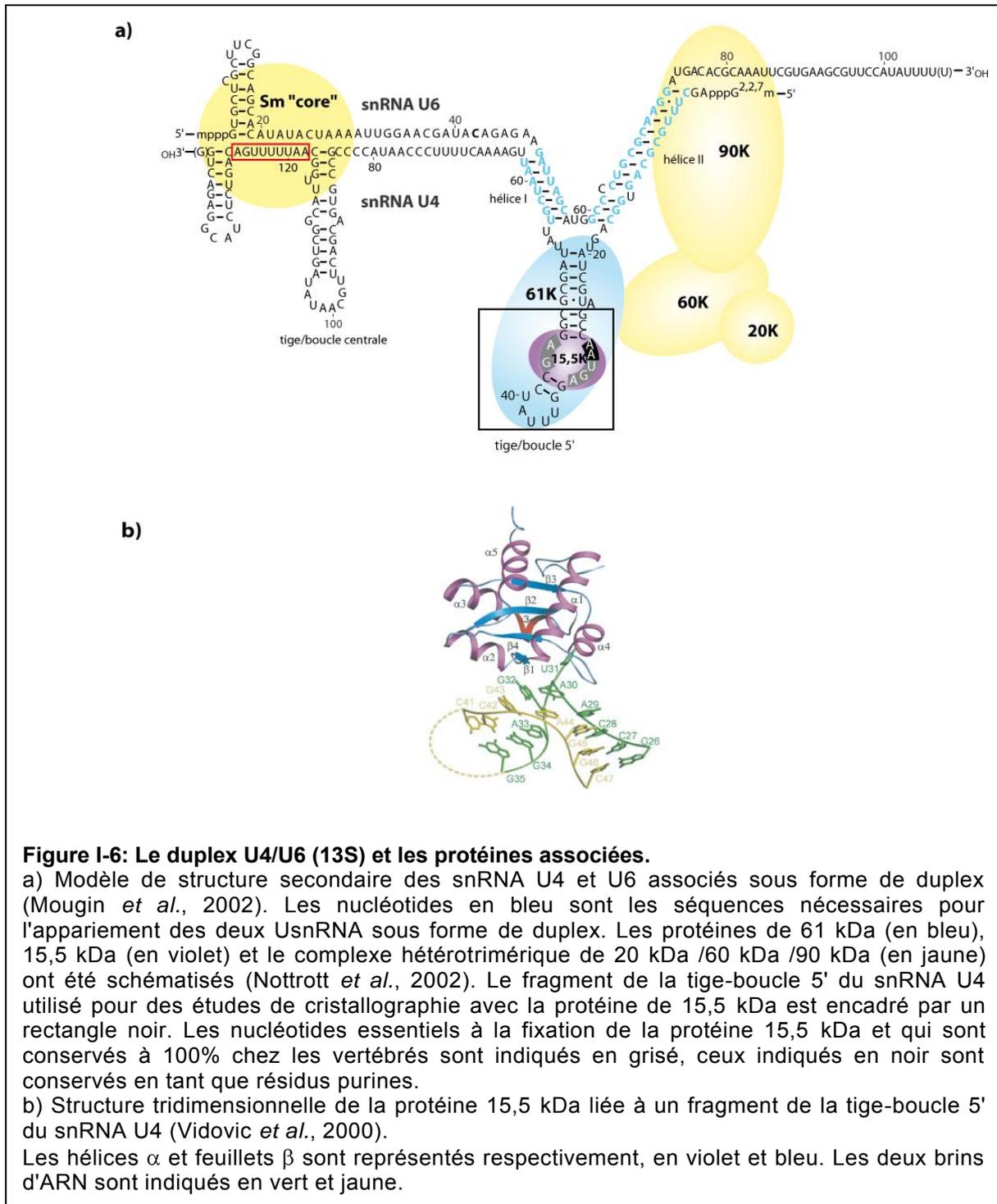
La tri-snRNP U4/U6.U5 est un complexe formé par l'association de la di-particule snRNP U4/6 avec la particule snRNP U5 (Behrens *et al.*, 1991).

#### **-Les snRNP U4 et U6**

Les snRNA U4 et U6 sont appariés dans la particule snRNP U4/U6 contenant les protéines Sm et LSm (Figure I-6). Au moins 5 protéines additionnelles peuvent être associées à la snRNP U4/U6, dont la protéine 15.5 kDa, et des polypeptides de 20, 60 et 90 kDa qui forment un complexe avec une protéine de 61 kDa (l'homologue de la protéine Prp31p de *S. cerevisiae*) (pour revue, Will *et al.*, 2001). La protéine 15.5 kDa se lie à la tige-boucle 5' du snRNA U4 sur une structure particulière de l'ARN dite en « K-turn » et pourrait servir de plateforme pour une association ultérieure des protéines spécifiques de la snRNP U4/U6 (Nottrott *et al.*, 1999; Watkins *et al.*, 2000).

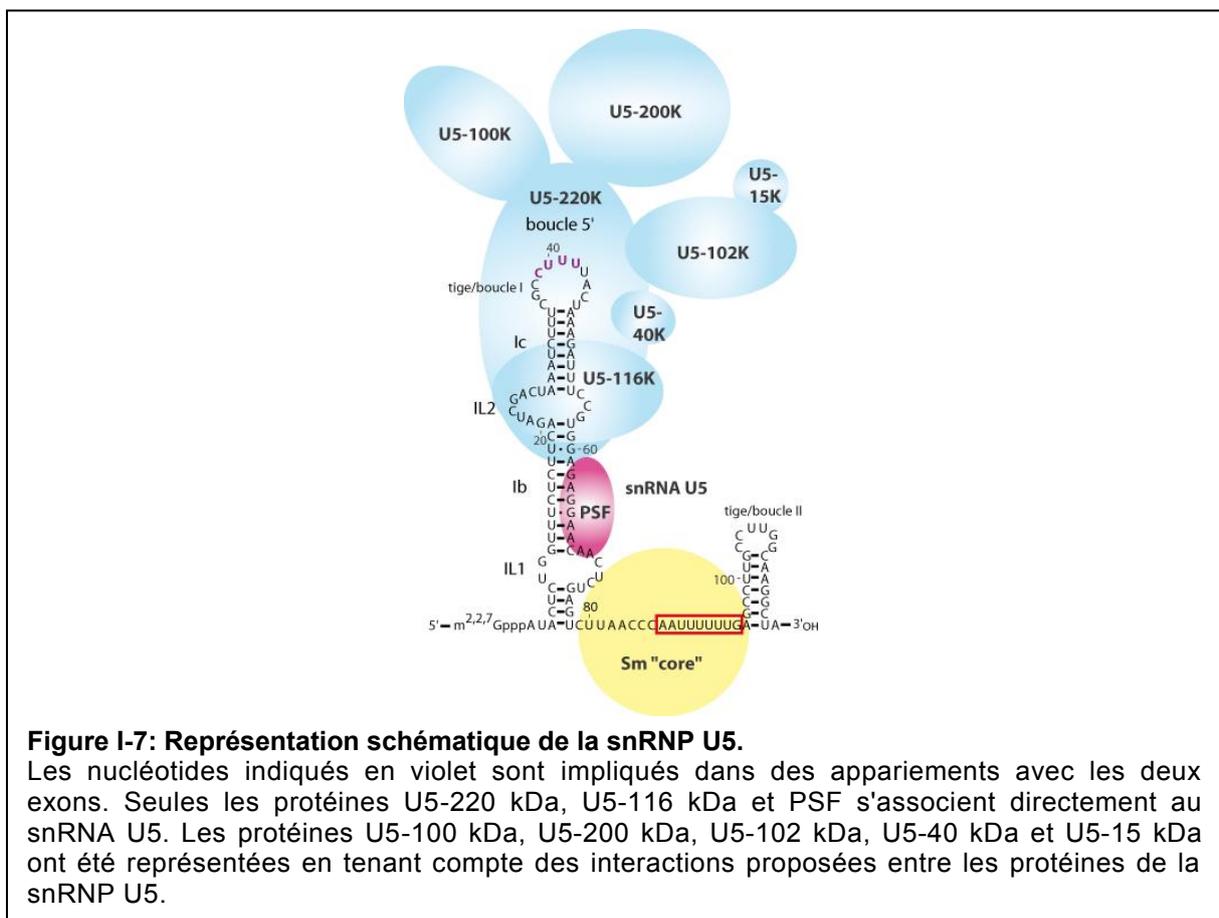
La snRNP U6 comporte 7 protéines LSm de la snRNP U6 appartenant à la même famille que les protéines Sm (Hermann *et al.*, 1995; Seraphin, 1995) qui possèdent toutes un domaine Sm. Les 7 protéines LSm et les 7 protéines canoniques Sm forment un complexe en forme d'anneau de 7-8 nm de diamètre et 3 nm d'épaisseur, comme cela a été montré par cristallographie aux rayons X et microscopie électronique (Achsel *et al.*, 1999; Kambach *et al.*, 1999; Mura *et al.*, 2001; Zaric *et al.*, 2005).

Certaines protéines de la snRNP U6 facilitent l'hybridation des snRNA U4 et U6 durant la formation de la di-snRNP U4/U6. Le snRNA U6 comporte une séquence très conservée (nucléotides 38-57) nécessaire à la fixation d'un facteur de recyclage SART3/p110 (ou Prp24 chez la levure) (Raghunathan et Guthrie, 1998; Bell *et al.*, 2002; Raghunathan *et al.*, 1998; Bell *et al.*, 2002; Rader *et al.*, 2002). Ce facteur est nécessaire à l'appariement du snRNA U4 au snRNA U6 dans le noyau pour former le duplex d'ARN U4/U6. Les snRNA U4 et U6 s'apparient via deux hélices intermoléculaires et adoptent alors une structure particulière en forme de Y (Bringmann *et al.*, 1984; Hashimoto et Steitz, 1984; Rinke *et al.*, 1985; Brow and Guthrie, 1988; Mougin *et al.*, 2002).



**-La snRNP U5**

La snRNP U5 complète est la plus grosse des UsnRNP spliceosomales (20S) (Bach *et al.*, 1990) (Figure I-7). A forte concentration saline une particule de 10S ne contenant que certaines des protéines est obtenue. La particule U5 facilite la ligation des exons en se liant à la bordure 3' du premier exon et la bordure 5' du second exon afin d'aligner les exons lors la seconde étape de transestérification (Patterson *et al.*, 1987; Winkelmann *et al.*, 1989). Les composants de la particule snRNP U5 (20S) sont le snRNA U5 associé aux 7 protéines Sm (B/B', D1, D2, D3, E, F et G) et au moins 9 protéines spécifiques ayant des poids moléculaires de 15, 40, 52, 100, 102, 110, 116, 200 et 220 kDa (Bach *et al.*, 1989).



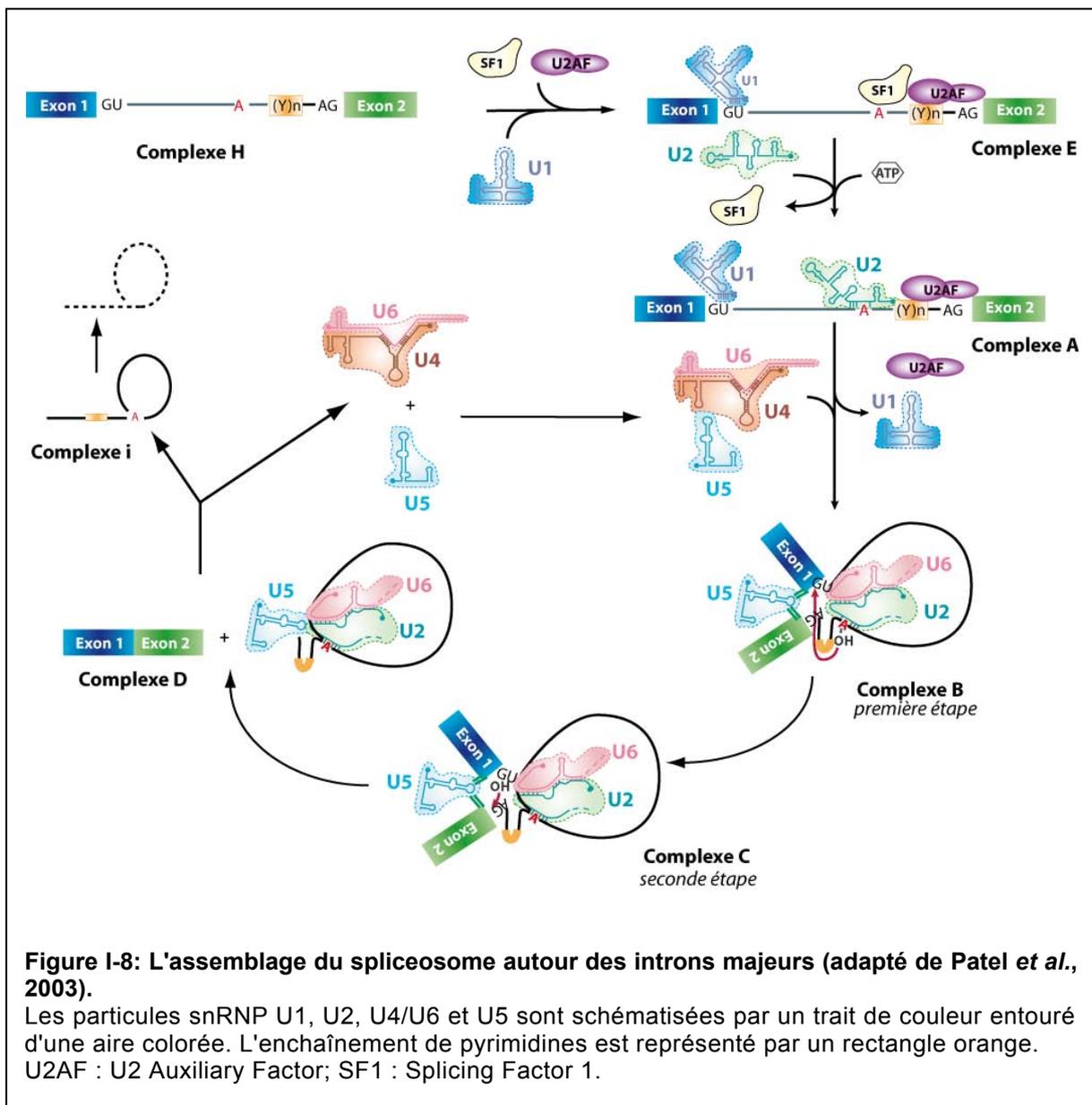
**-La tri-particule snRNP U4/U6.U5**

Avant d'interagir avec le pré-ARNm, la snRNP U4 interagit avec la snRNP U6 pour former la di-snRNP U4/U6 (Rinke *et al.*, 1985). Ensuite ce complexe se lie à la snRNP U5 pour former la tri-snRNP U4/U6.U5 et cette tri-particule de 25S s'associe aux pré-complexes spliceosomiaux et avec le pré-ARNm. Seules trois protéines (27, 65 et 110 kDa) ne sont

retrouvées que dans la tri-snRNP. Ces trois protéines spécifiques contiennent un domaine RS et sont considérées comme des protéines SR-like (Fetzer *et al.*, 1997), suggérant une participation directe de ces 3 protéines dans le recrutement de la tri-snRNP au sein du spliceosome (Roscigno *et al.*, 1995; Makarova *et al.*, 2001). Deux des protéines de la tri-snRNP, U5-102 kDa et U4/U6-61 kDa, forment un hétérodimère et sont essentielles à la formation de la tri-snRNP, car elles jouent un rôle de « lien » entre la partie U4/U6 et la partie snRNP U5 (Makarov *et al.*, 2002).

### I.3.2 Assemblage du spliceosome

Les complexes d'épissage formés durant les réactions d'épissage *in vitro* ont été abondamment analysés chez les levures et les extraits de cellules HeLa, la plupart du temps en utilisant des séparations électrophorétiques des complexes sous des conditions non dénaturantes. Un modèle de l'assemblage en différentes étapes d'un complexe mature d'épissage, appelé le spliceosome, a été établi (Figure I-8). La molécule pré-ARNm s'associe tout d'abord avec les protéines hnRNP (heterogeneous nuclear ribonucleoproteins) de façon co-transcriptionnelle. Cette interaction ne requiert pas d'énergie et peut être détectée pour n'importe quelle molécule d'ARN à 4°C. Des études montrent que la composition en hnRNP de ces complexes peut varier selon le substrat d'épissage, suggérant une spécificité d'interaction pouvant permettre une régulation fine. La snRNP U1 se lie ensuite au pré-ARNm grâce à des paires de bases entre l'extrémité 5' de l'ARN U1 et la jonction 5' du site d'épissage. Cette interaction est stabilisée par une boîte de branchement fonctionnelle, suggérant une interaction additionnelle entre le pré-ARNm et la snRNP U1 via des protéines associées. Ces complexes sont appelés « complexes d'engagement », depuis l'assemblage de pré-ARNm en spliceosome mature qui va pouvoir être épissé. La snRNP U2 va ensuite se lier au complexe d'engagement via une interaction à la région de branchement qui conduit à la formation d'un pré-spliceosome. Cette interaction implique l'appariement des bases entre le pré-ARNm et une séquence interne du snRNA U2. Les snRNP U4/U6 et U5 se pré-assemblent ensemble en particule triple, la tri-snRNP U4/U6.U5, qui s'associe avec le pré-spliceosome pour former le spliceosome. Les changements conformationnels de ce complexe peuvent être détectés par électrophorèse dans des conditions non dénaturantes. Le relargage du snRNA U4 du complexe va générer le spliceosome mature dans lequel la réaction d'épissage va pouvoir prendre place. Par la suite, le spliceosome se désassemble et l'ARNm mature va être libéré.



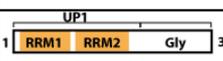
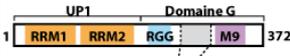
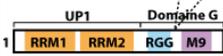
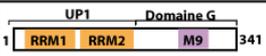
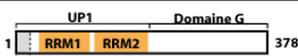
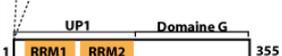
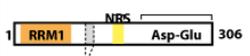
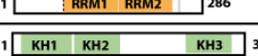
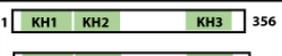
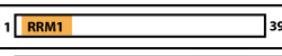
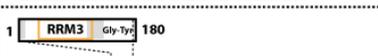
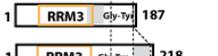
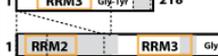
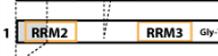
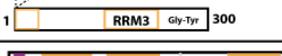
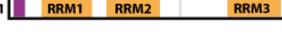
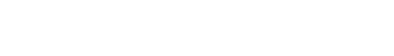
### I.3.3 Les facteurs d'épissage autres que les protéines des UsnRNP

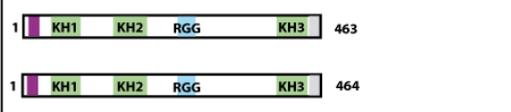
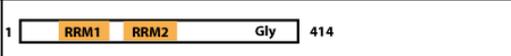
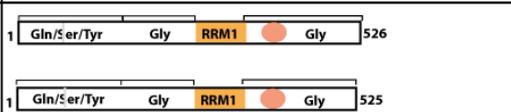
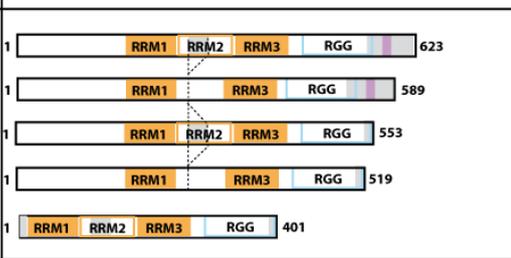
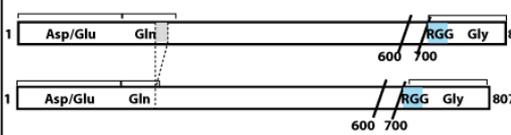
#### I.3.3.1 Les protéines hnRNP

Les ribonucléoprotéines nucléaires hétérogènes (hnRNP) forment une famille de protéines se liant à l'ARN en général sur les pré-ARNm naissants. Elles sont les protéines les plus abondantes liées au pré-ARNm et plus de 20 d'entre elles ont été caractérisées et sont identifiées par des lettres classées par ordre alphabétique, du plus petit poids moléculaire au plus grand, soit de hnRNP A1 à hnRNP U (Figure I-9). Ces protéines sont impliquées dans une grande variété de processus biologiques incluant biogénèse des télomères, traduction et stabilité de l'ARN. Plusieurs de ces protéines participent au contrôle de l'épissage (hnRNPA1, A2, F, H, I [PTB], K, G et L) (Venables *et al.*, 2008). L'étendue et la spécificité

de l'action des hnRNP restent peu décrites, même pour celles affectant la sélection du site d'épissage.

Comme les protéines hnRNP lient l'ARN, il est probable que les effets de la sélection du site d'épissage résultent d'une interaction directe avec l'ARN messenger cible. Il est toutefois possible qu'un changement dans l'abondance relative des isoformes d'épissage soit la conséquence de la stabilité différentielle des ARNm, et aussi par le fait que plusieurs protéines hnRNP font la navette vers le cytoplasme où elles peuvent affecter le turn-over des ARNm. De plus, étant donné que de nombreuses protéines liant l'ARN subissent elles-mêmes un épissage alternatif, changer l'expression d'une protéine hnRNP peut affecter l'épissage alternatif et ainsi l'activité d'autres protéines liant l'ARN dans le contrôle de l'épissage.

hnRNP	organisation structurale	PM (kDa)	L	séquence ARN	rôles cellulaires	
A0		31,7	N	G <sub>4</sub> (AU <sub>3</sub> ) <sub>4</sub> A	Contrôle de la traduction	
A1B		34	N,C	UAGGG(A/U) (Py)UAG	Epissage, export des ARN Maintenance des télomères	
A1		320				
A2		36	N,C	GCCAAGGAGCC	Epissage, localisation des ARNm, maintenance des télomères	
B1		38				
A3(1)		41	NC	GCCAAGGAGCC	Localisation des ARNm	
A3(2)		39				
C2		41	N	riche en U, AGUA(U) <sub>5</sub> GUGGA	Epissage, contrôle de la tra- duction, contrôle de la loca- lisation des pré-ARNm, maintenance des télomères	
C1		43				
D(1)		45	N	riche en A/U, UUAG/G	Maintenance des télomères, stabilisation ou destabilisation des ARNm, recombinaison	
D(2)		42				
D(3)		40				
D(4)		37				
E1		37,5	N,C	Poly(C), Poly(G)	Stabilité des ARNm, contrôle traduction	
E2						360
G		42	N	AAGU	Epissage	
F		45,5	N	Poly(G)	Epissage	
H		56	N	GCGA/C, poly(G)	Epissage, polyadénylation	
H'		56	N			
2H9(6)		15,5	N	GGGA	Epissage	
2H9(5)		16				
2H9(4)		22				
2H9(1)		37				
2H9(2)		35				
2H9(3)		31,5				
PTB(1)		59	N	riche en C/U, UCUU, UCUCU, CUUCUCUCU	Epissage, localisation des ARNm, polyadénylation	
PTB(2)						550
PTB(4)						556

hnRNP	organisation structurale	PM (kDa)	L	séquence ARN	rôles cellulaires
<b>K(1)</b> <b>K(2)</b>		62	N,C	poly(C), AU(C) <sub>3/4</sub> (U/A) <sub>2</sub>	Epissage, régulation de la transcription, contrôle de la traduction
<b>L</b>		60	N	répétitions CA	Epissage, export ARNm, stabilité ARNm
<b>M(1)</b> <b>M(2)</b>		74 72	N	GAGGAGG	Epissage
<b>TDP-43</b>		53	N	répétitions UG	Epissage, represser transcriptionnel
<b>TLS/ FUS</b>		53,5	N	GGUG	Epissage, oncogenèse
<b>Q3(1)</b> <b>Q2(2)</b> <b>Q1(3)</b> <b>Q(4)</b> <b>Q(5)</b>		69,6	N,C		Epissage
<b>R</b>		70,9	N		Edition, transport des ARNm, polyadénylation
<b>U(1)</b> <b>U(2)</b>		120	N	poly(G), riche en A/U	Régulation de la transcription, rétention nucléaire

-  domaine RRM
-  domaine KH
-  domaine RGG
-  domaine à doigt de zinc
-  domaine NRS ("nuclear retention signal")
-  domaine M9
-  domaine NLS ("nuclear localisation signal")
-  mutations ou délétions entre les isoformes
-  Gly domaine riche en résidus glycine
-  Asp/Glu domaine riche en résidus acide aspartique et glutamique

**Figure I-9: Les protéines hnRNP: structures, rôles, et séquences ARN reconnues (d'après Krecic *et al.*, 1999).**

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.

### I.3.3.2 Les protéines SR

#### ***-La famille des protéines SR***

La famille des protéines SR est composée de protéines phylogénétiquement et structurellement conservées, caractéristiques par leur domaine riche en résidus arginine (R) et sérine (S), connu sous le nom de domaine RS (Figure I-10). Les protéines SR (Fu, 1995) ont été découvertes au début des années 1990 en tant que facteurs régulant l'épissage. Le domaine RS a été observé à l'origine dans des régulateurs de l'épissage chez la drosophile, SWAP (suppressor of white apricot) (Chou *et al.*, 1987), Tra (transformer) (Boggs *et al.*, 1987) et Tra-2 (transformer-2) (Amrein *et al.*, 1988). Les protéines ASF/SF2 (Ge *et al.*, 1991; Krainer *et al.*, 1991) et SC35 (Fu *et al.*, 1990) ont ensuite été identifiées comme contenant un domaine RS.

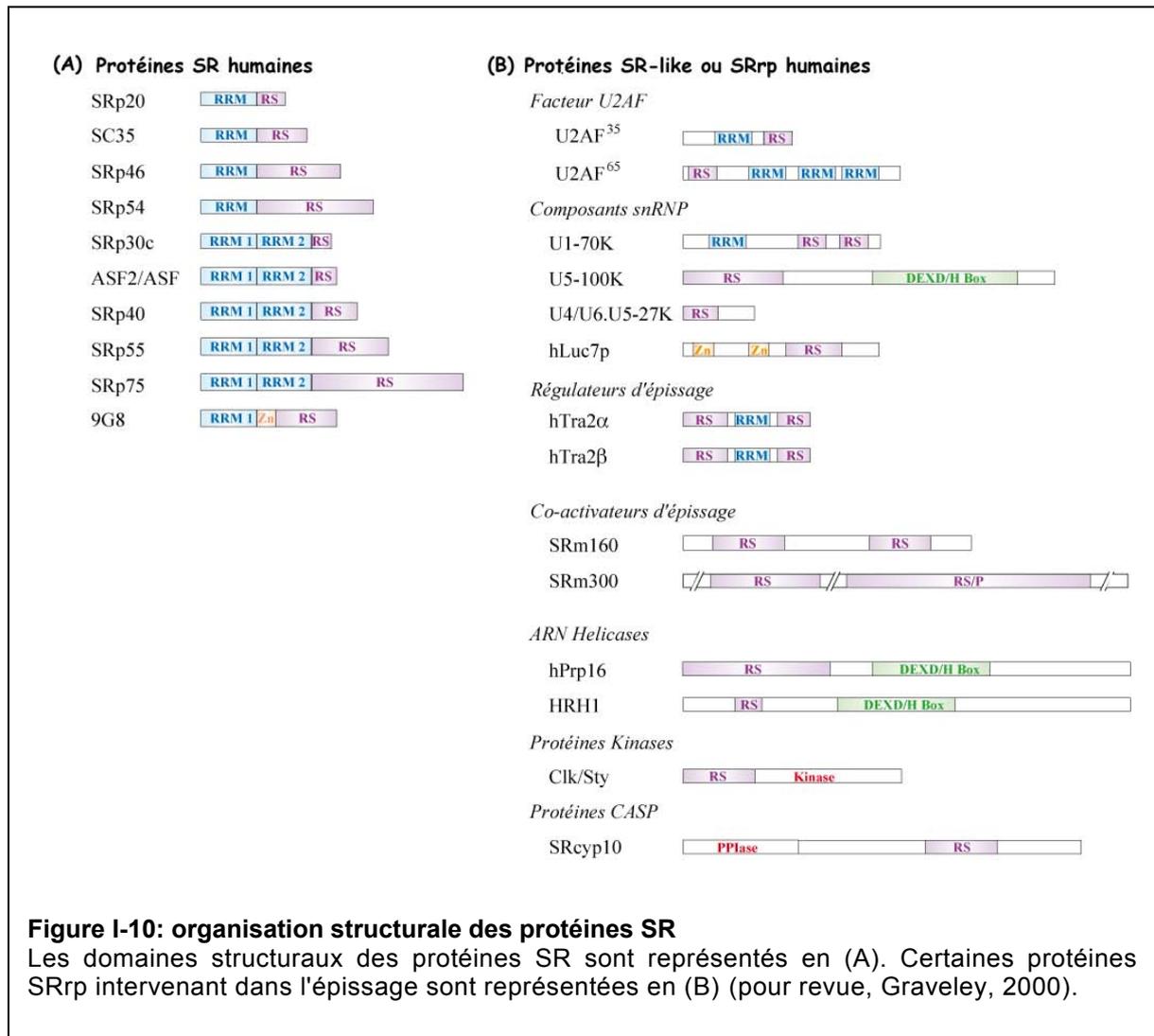
Les critères utilisés pour définir les membres de la famille des protéines SR dites « classiques » sont :

- une similarité structurale : une structure modulaire contenant une à deux copies de RRM (RNA recognition motif) situé dans la partie N-terminale et qui fournit une spécificité de liaison à l'ARN. Dans la partie C-terminale de la protéine, le domaine RS permet le contact direct avec le pré-messager et les interactions protéine-protéine qui facilitent le recrutement du spliceosome,
- une fonction double dans l'épissage constitutif et alternatif,
- la présence d'un phospho-épitope reconnu par l'anticorps mAb104 (Roth *et al.*, 1990),
- tous les membres de la famille des protéines SR sont hautement solubles dans le sulfate d'ammonium et insoluble dans le chlorure de magnésium 20 mM. Leur capacité à précipiter dans le MgCl<sub>2</sub> est une propriété intrinsèque de ces phosphoprotéines (Roth *et al.*, 1991).

#### ***-organisation structurale (Figure I-10).***

##### RRM

Les différences de composition des RRM des protéines SR ont conduit à distinguer deux sous-groupes (Mayeda *et al.*, 1999). Le premier regroupe les protéines ayant un motif RRM classique et un second plus atypique et incluant un heptapeptide caractéristique (ASF/SF2 et SRp40 par exemple). Le second groupe ne contient que les protéines à un seul domaine RRM comme SC35 ou 9G8 parmi d'autres.



**Figure I-10: organisation structurale des protéines SR**

Les domaines structuraux des protéines SR sont représentés en (A). Certaines protéines SRrp intervenant dans l'épissage sont représentées en (B) (pour revue, Graveley, 2000).

Les protéines SR reconnaissent un vaste profil de séquences ARN. Les consensus des sites de fixation sont souvent dégénérés et dans beaucoup de cas, le motif reconnu par une protéine SR donnée est également reconnu par d'autres protéines SR (Tacke *et al.*, 1995; Liu *et al.*, 1998). Une nouvelle nomenclature pour ces protéines a été établie (Manley *et al.*, 2010).

### Domaine RS

Une étude de tout le génome des métazoaires permet d'identifier un grand nombre de protéines contenant un domaine RS. Ce domaine peut se trouver au sein d'interactions protéine-protéine lors de processus divers comme l'épissage, le remodelage de la chromatine, la transcription, le cycle cellulaire ainsi que dans la communication entre différents processus (Boucher *et al.*, 2001). Un des modèles proposé pour le mécanisme d'action des protéines SR est le modèle d'interaction protéine-protéine. Dans ce modèle, le domaine RS d'une protéine liée à un activateur va interagir directement avec d'autres facteurs d'épissage qui contiennent un domaine RS et ainsi faciliter le recrutement du spliceosome (Graveley, 2000; Shen *et al.*,

2004). Wu et Maniatis ont montré par exemple que SC35 interagit avec lui-même et d'autres composants spliceosomaux, dont U1-70 kDa et la sous-unité de 35 kDa de U2AF (Wu *et al.*, 1993).

En 2004, Shen, Kan et Green montrent que le domaine RS fixe par intermittence le pré-ARNm, ce qui élargit le nombre de cibles potentielles des domaines RS (Shen *et al.*, 2004). Cependant, le débat reste ouvert pour déterminer si les forces qui dirigent l'assemblage du spliceosome sont dues aux contacts entre les domaines RS et le pré-ARNm ou avec d'autres facteurs d'épissage.

Le domaine RS permet dans certaines situations aux protéines SR de faire la navette entre le noyau et le cytoplasme. Chez ASF/SF2, une séquence de dipeptides du domaine RS, en conjonction avec le RRM, est nécessaire et suffisante pour lui permettre de faire la navette. Le second RRM d'ASF/SF2, et en particulier un hexapeptide phylogénétiquement conservé SWQDLKD, localisé dans la première hélice  $\alpha$  de ce domaine (Birney *et al.*, 1993) est essentiel pour l'activité de sélection du site d'épissage de la protéine ASF/SF2 (Dauksaite *et al.*, 2002; Dauksaite *et al.*, 2004). La structure de ce domaine révèle un repliement atypique du RRM qui lie l'ARN de façon originale (Tintaru *et al.*, 2007). La protéine SC35 quant à elle, a deux longues séquences de répétitions RS et n'est pas une protéine navette car elle détient un signal de rétention nucléaire dominant dans son domaine RS (Cazalla *et al.*, 2002). L'import nucléaire de plusieurs protéines SR, dont ASF/SF2 et SC35, se fait par l'intermédiaire d'un récepteur spécifique d'import, la transportine-TRN-SR). TRN-SR est un membre de la famille des importines  $\beta$  /transportines et se lie spécifiquement et directement aux domaines RS de ASF/SF2 et SC35 ainsi qu'à d'autres protéines SR additionnelles (Kataoka *et al.*, 1999).

### ***-Les séquences reconnues par les protéines SR (Figure I-11).***

Il existe différentes approches pour identifier les cibles ARN physiologiques des protéines SR. L'une d'elle est appelée SELEX (Selected Evolution of Ligands through EXponential enrichment) et implique la sélection de sites de liaison de haute affinité dans un groupe aléatoire de séquences ARN (Tuerk *et al.*, 1990). Cette technique a permis l'identification de sites de liaison de haute affinité pour ASF/SF2 et SC35 (Tacke *et al.*, 1995), SRp40 (Tacke *et al.*, 1997), 9G8 et SRp20 (Cavaloc *et al.*, 1999). Ces sites de liaison sont des séquences riches en purines qui ressemblent à des sites 5' d'épissage ou à des séquences exoniques connues pour être activatrices d'épissage. Le SELEX peut être utilisé en conjonction avec un criblage bioinformatique à grande échelle.

Une alternative à cette approche de SELEX basé sur la fixation à l'ARN est le développement d'un SELEX fonctionnel qui implique la sélection d'une séquence qui stimule l'épissage. Liu a employé l'approche par SELEX fonctionnel basé sur la complémentation d'extraits déplétés en protéines SR avec des protéines SR recombinantes pour identifier les activateurs d'épissage spécifiques d'ASF/SF2, de SRp40 et de SRp55 (Liu *et al.*, 1998).

Les motifs obtenus différaient significativement des séquences consensus déterminées par le SELEX conventionnel.

Protéine SR	Site de fixation	Méthode d'identification
ASF/SF2	RGAAGAAC AGGACRRAGC SRSASGA UGRWG	SELEX SELEX SELEX fonctionnel CLIP
SC35	AGSAGAGUA GUUCGAGUA GRYYCSYR UGUUCSAGWU GWUWCCUGCUA GGGUAUGCUG GAGCAGUAGKS AGGAGAU UGCNGYY	SELEX SELEX SELEX fonctionnel SELEX SELEX SELEX SELEX SELEX SELEX fonctionnel
SRp20	GGUCCUCUUC WCWWC CUCKUCY	Gel shift Test d'épissage Affinité à l'ARN
SRp75	GAAGGA	UV cross-linking
SRp40	GAGCAGUCGGCUC ACDGS	SELEX SELEX fonctionnel
SRp55/B52	USCGKM UCAACCAGGCGAC	SELEX fonctionnel SELEX
9G8	UCAACA ACGAGAGAY GGACGACGAG	UV cross-linking SELEX SELEX fonctionnel
p54	Riche en C	UV cross-linking
SRp30C	GACGAC AAAGAGCUCGG CUGGAUU	SELEX fonctionnel SELEX fonctionnel Gel shift
hTra2β	(GAA) <i>n</i>	SELEX
SRm160	Riche en purine (GAA) <i>n</i>	Test d'épissage

N: tout nucléotide; R: purine; Y: pyrimidine; D: A, G ou U; K: U ou G; M: A ou C; S: G ou C; W: A ou U

**Figure I-11: Séquences ARN identifiées comme sites de fixation des protéines SR (D'après Long *et al.*, 2009).**

### ***-Les protéines SR-like***

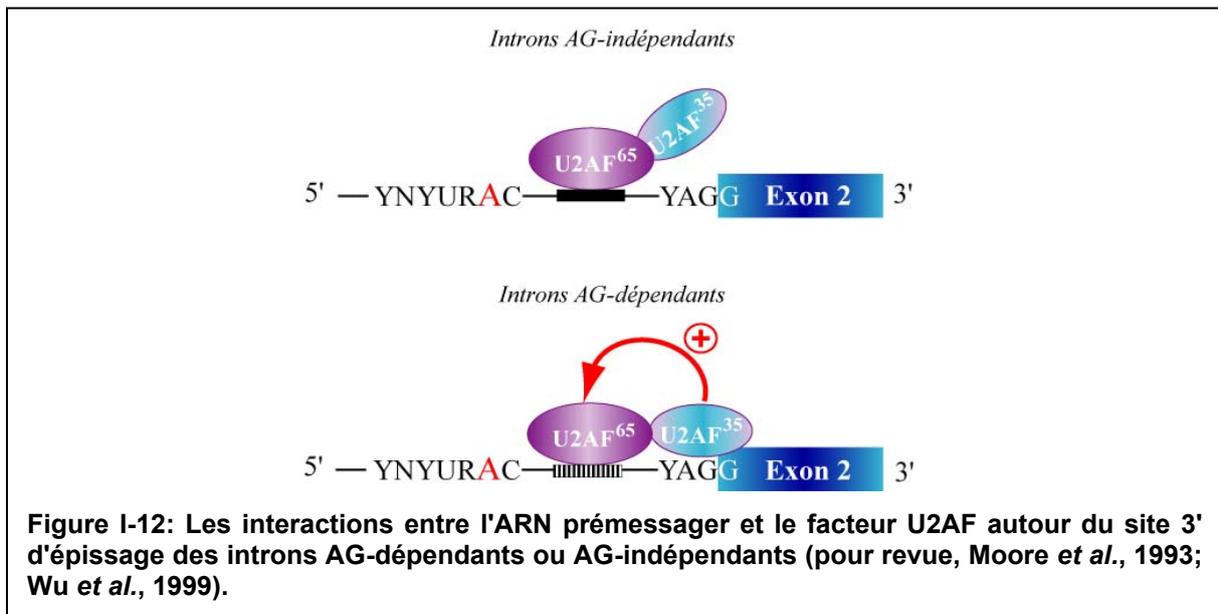
Il existe dans le génome des métazoaires un grand nombre de protéines contenant un domaine RS et qui ont un rôle dans l'épissage mais aussi dans d'autres processus cellulaires comme le remodelage de la chromatine, la transcription et la progression du cycle cellulaire (Boucher *et*

*al.*, 2001). Ces protéines contiennent un domaine RS mais n'ont pas toutes un domaine RRM. Toutefois, certaines protéines peuvent lier l'ARN à travers d'autres domaines comme le motif PWI trouvé dans l'activateur d'épissage SRm160 (Blencowe *et al.*, 1998). Ces facteurs sont connus sous le nom de protéines « SR-like » ou « SR-related » et ces groupes incluent les deux sous-unités de l'hétérodimère U2AF, la protéine U1-70 kDa et les co-activateurs d'épissage SRm160/300 (Blencowe *et al.*, 1999) parmi d'autres. Il a récemment été proposé que les protéines SR pourraient être redéfinies selon leurs caractéristiques structurales communes et leur rôle dans l'épissage du pré-ARNm (Lin *et al.*, 2007). Basé sur cela, une protéine SR « bona fide » doit contenir au moins un domaine RRM et un domaine RS (quelque soit sa position dans la protéine) et avoir un rôle dans l'épissage constitutif ou alternatif, selon le test de complémentation en extrait cytoplasmique S100 de cellules HeLa déficient pour l'épissage, ou en test d'épissage alternatif. Les homologues humains des régulateurs de l'épissage de la drosophile Tra2 $\alpha$  (Dauwalder *et al.*, 1996) et Tra2 $\beta$  (Beil *et al.*, 1997) contiennent un RRM flanqué par deux domaines RS, ce qui n'est pas la structure classique des domaines classique chez les protéines SR. Ces deux protéines sont des activateurs spécifiques de séquence de l'épissage, donc sont classifiées comme protéines SR. D'autres protéines peuvent contenir un domaine SR mais ont d'autres domaines requis pour leurs activités enzymatiques, comme c'est le cas pour les hélicases ARN HRH1 et hPRP16 qui contiennent un domaine DEAD box (Ohno *et al.*, 1996; Zhou *et al.*, 1998).

### I.3.3.3 Les facteurs U2AF

L'assemblage du spliceosome est une suite ordonnée d'événements qui commence avec la reconnaissance du site 5' d'épissage par la snRNP U1 et la liaison de U2AF (U2 small nuclear ribonucleoprotein auxiliary factor) à la séquence polypyrimidine et au site 3' d'épissage. Le facteur U2AF humain est un hétérodimère composé d'une sous-unité de 65 kDa (U2AF<sub>65</sub>) qui se lie à la séquence polypyrimidine (Ruskin *et al.*, 1988; Zamore *et al.*, 1989) et une sous-unité de 35 kDa (U2AF<sub>35</sub>) qui interagit avec les nucléotides AG du site 3' d'épissage (Wu *et al.*, 1999).

On distingue ainsi 2 types d'exons, dits AG-dépendants ou AG- indépendants, suivant que la séquence polypyrimidine est sous optimale ou non (Reed, 1989) (Figure I-12).



La protéine U2AF<sub>65</sub> contient trois motifs de reconnaissance de l'ARN RRM. Les 2 RRM centraux, RRM1 et RRM2, sont des domaines RRM canoniques responsables de la reconnaissance de la séquence polypyrimidine du pré-ARNm, tandis que le troisième RRM a des caractéristiques inhabituelles et est spécialisé dans les interactions protéine-protéine. Ce domaine appelé UHM (pour U2AF Homology Motif), est présent dans plusieurs autres facteurs d'épissage (Kielkopf *et al.*, 2004). Le UHM de U2AF<sub>65</sub> reconnaît le facteur d'épissage SF1 et cette interaction coopérative protéine-protéine renforce la liaison à la séquence polypyrimidine. Les acides aminés (85-112) de l'extrémité N-terminale de U2AF<sub>65</sub> interagissent avec U2AF<sub>35</sub> et cette association renforce encore plus la liaison avec la séquence polypyrimidine (Kielkopf *et al.*, 2004). Bien qu'elle ne soit pas un membre de la famille des protéines SR, la protéine U2AF<sub>65</sub> contient un domaine RS qui est requis pour l'assemblage du spliceosome *in vitro* (Valcarcel *et al.*, 1996). La protéine U2AF<sub>35</sub> contient un domaine central UHM impliqué dans l'interaction avec U2AF<sub>65</sub>, flanqué par deux motifs se liant au Zn<sup>2+</sup> et un domaine RS C-terminal.

Une interaction entre U2AF<sub>65</sub> et SAP 155, une protéine sous-unité de SF3b, peut fournir un mécanisme additionnel de recrutement du U2 (Gozani *et al.*, 1998). U2AF<sub>65</sub> interagit aussi avec UAP56, une hélicase DEAH box qui est importante pour la liaison de la snRNP U2 (Fleckner *et al.*, 1997). SF1/mBBP et U2AF<sub>65</sub> interagissent ensemble et cette interaction peut faciliter la reconnaissance du point de branchement et de la séquence polypyrimidine qui sont adjacents dans le pré-ARNm (Berghlund *et al.*, 1998).

### **I.4 L'épissage : assemblage du spliceosome et catalyse de la réaction**

Les spliceosomes contiennent 5 petites particules nucléaires ribonucléoprotéiques (snRNP) en plus de plus de 100 autres facteurs d'épissage (Smith *et al.*, 2008). Durant chaque cycle d'épissage, ces facteurs s'assemblent *de novo* sur le substrat qu'est le pré-ARNm.

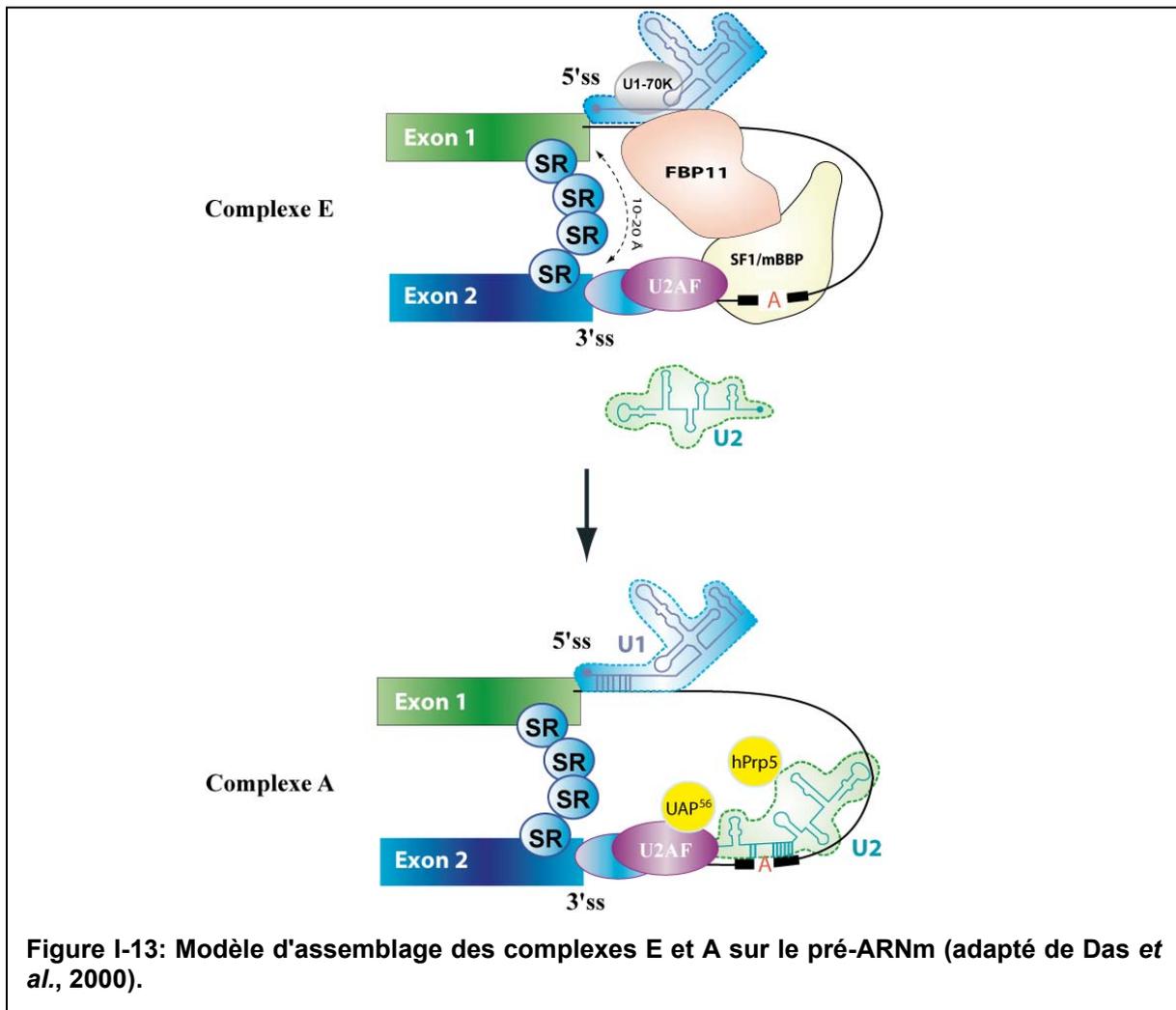
#### **I.4.1 Le complexe H**

Le complexe H, ou complexe 30S, est un complexe non spécifique qui ne nécessite pas d'ATP ni de site d'épissage 5' ou 3' fonctionnel pour son assemblage, il s'associe aussi bien sur un ARN en absence de site d'épissage. La composition du complexe H est la même quand le pré-ARNm est incubé en présence d'ATP (Reed, 1990). Le complexe H s'associe immédiatement dès l'addition de l'ARN dans la réaction d'épissage *in vitro* avec de nombreux types de protéines se liant à l'ARN, comme par exemple les protéines hnRNP, les protéines se liant à la coiffe, et les protéines SR. Le complexe H n'est pas un intermédiaire fonctionnel de l'assemblage du spliceosome et contient les mêmes protéines hnRNP que celles qui s'associent avec le pré-ARNm *in vivo* (Bennett *et al.*, 1992).

#### **I.4.2 Le complexe E ou complexe d'engagement**

Après incubation à 30°C, le complexe H est rapidement converti en complexe E. une fois inclus dans un complexe E, l'ARN pré-messager est irrémédiablement entraîné dans la voie d'épissage, d'où le nom de complexe d'engagement donné à ce premier intermédiaire de l'assemblage des spliceosomes. Cette liaison de protéines à l'ARN ne requiert pas de sites d'épissage. Le complexe spliceosomal E est le premier intermédiaire fonctionnel détectable dans l'assemblage du spliceosome *in vitro*.

Il renferme les facteurs U2AF et SF1, des protéines SR, la snRNP U1 et 5 protéines SAP de 115, 92, 88, 72 et 42 kDa (Bennett *et al.*, 1992; Chiara *et al.*, 1996; Abovich *et al.*, 1997). En présence d'ATP, le complexe E va devenir le premier complexe spliceosomal dépendant de l'ATP, nommé le complexe A (Figure I-13).



### I.4.3 Le complexe A ou pré-spliceosome

Le complexe A est le premier complexe dépendant de l'ATP dans l'assemblage du spliceosome. U2AF recrute la particule snRNP U2 pour lier le site de branchement au sein du complexe E lors d'une réaction nécessitant de l'ATP, pour former le complexe A (Figure I-13). Les bases de la snRNP U2 s'associent au point de branchement, causant la saillie de l'adénosine du site de branchement, dans une position qui va permettre l'attaque nucléophile du site d'épissage 5'. Le complexe A sert de substrat pour la formation du complexe B. La formation du complexe B est également dépendante de l'ATP, et la présence des sites 5' et 3' sur le pré-ARNm sont essentiels tous deux pour l'assemblage du complexe B. Les snRNP U4 et U6 ont des bases appariées et ce complexe U4/U6 s'associe avec la snRNP U5 pour former une particule tri-snRNP. Cette particule pré-assemblée tri snRNP U4/U6.U5 va ensuite se lier au complexe A pour former le complexe spliceosomal B.

### I.4.4 Le complexe B

Même si tous les snRNP sont présents dans le complexe B, il est catalytiquement inactif. Des réarrangements majeurs sont nécessaires, tant au niveau de la conformation que de la composition pour l'activation catalytique et pour devenir compétent pour faciliter la première étape de transestérification. Durant l'activation du spliceosome, U1 et U4 sont déstabilisés ou relâchés, donnant naissance au spliceosome activé, le complexe B\*. Le spliceosome activé va ensuite subir la première étape catalytique d'épissage, générant le complexe C.

A ce jour, les complexes spliceosomaux humains B\* n'ont pas été isolés sous des conditions physiologiques (Luhrmann *et al.*, 2009). Toutefois, les analyses de spectrométrie de masse des complexes B comparées aux complexes B\* isolés sous des conditions stringentes non physiologiques indiquent que la perte des protéines associées aux U4/U6 ont lieu durant la transition du complexe B au complexe B\* (Makarova *et al.*, 2004). De plus, les sous-unités des complexes Prp19/CDC5 et ses protéines rattachées deviennent plus stablement intégrées dans le complexe B\*, indiquant que des événements majeurs du remodelage des RNP impliquant ces protéines ont lieu durant l'activation (Makarov *et al.*, 2002). La transition du complexe B au complexe B\*, essentielle pour l'activation catalytique du spliceosome, requiert l'action de 3 protéines DExD/H-box additionnelles et de la GTPase Snu114. La protéine Prp28 est impliquée dans la médiation du transfert du site 5' d'épissage à l'extrémité 5' du snRNA U1 au motif ACAGAG dans le snRNA U6. Cet événement initie l'activation catalytique du spliceosome.

### I.4.5 Le complexe C

Le complexe spliceosomal C est un intermédiaire qui a une durée de vie très courte; les intermédiaires d'épissage sont rapidement convertis en produits d'épissage. Ainsi, les produits épissés sont relâchés très rapidement et aucun complexe contenant les différents produits d'épissage, lasso et ARNm, n'a jamais été isolé.

La conversion du complexe spliceosomal B en complexe C requiert de l'ATP. Les nombreux appariements des bases entre le snRNA U4 et snRNA U6 sont rompus durant la formation du complexe C. Il semblerait que ceci nécessite une activité de type hélicase associée avec les protéines DEAD box. Le complexe actif C est formé grâce à un changement conformationnel de la structure du complexe C. Après la formation du complexe actif C, la réaction d'épissage a lieu très rapidement.

### **La dynamique des spliceosomes**

Des analyses par spectrométrie de masse des différents intermédiaires d'assemblage du spliceosome indiquent qu'il y a un changement conséquent de protéines durant l'assemblage et l'activation. Une comparaison de la composition protéique des complexes humains A, B et C purifiés par affinité dans des conditions physiologiques identiques permet de classer les protéines en 3 groupes selon la composition des spliceosomes au cours des différentes étapes d'épissage.

Dans le premier groupe, différentes classes de protéines spliceosomales sont présentes dans le spliceosome tout au long de l'épissage, incluant les protéines associées au U2 et les membres des familles de protéines SR et hnRNP qui sont présents en quantité variable.

Dans le deuxième groupe, durant la transition du complexe A vers le complexe B et du complexe B vers le complexe C, un grand nombre de protéines spliceosomales additionnelles sont recrutées.

Dans le troisième groupe, de nombreuses protéines spliceosomales sont relâchées ou déstabilisées à chaque étape durant le processus d'épissage. Ainsi, plusieurs protéines non-snRNP du complexe A ne sont pas retrouvées dans le complexe B. Ces constatations sont la preuve d'un grand turn-over des protéines au sein des spliceosomes. La composition des spliceosomes au cours du processus d'épissage montre une grande dynamique.

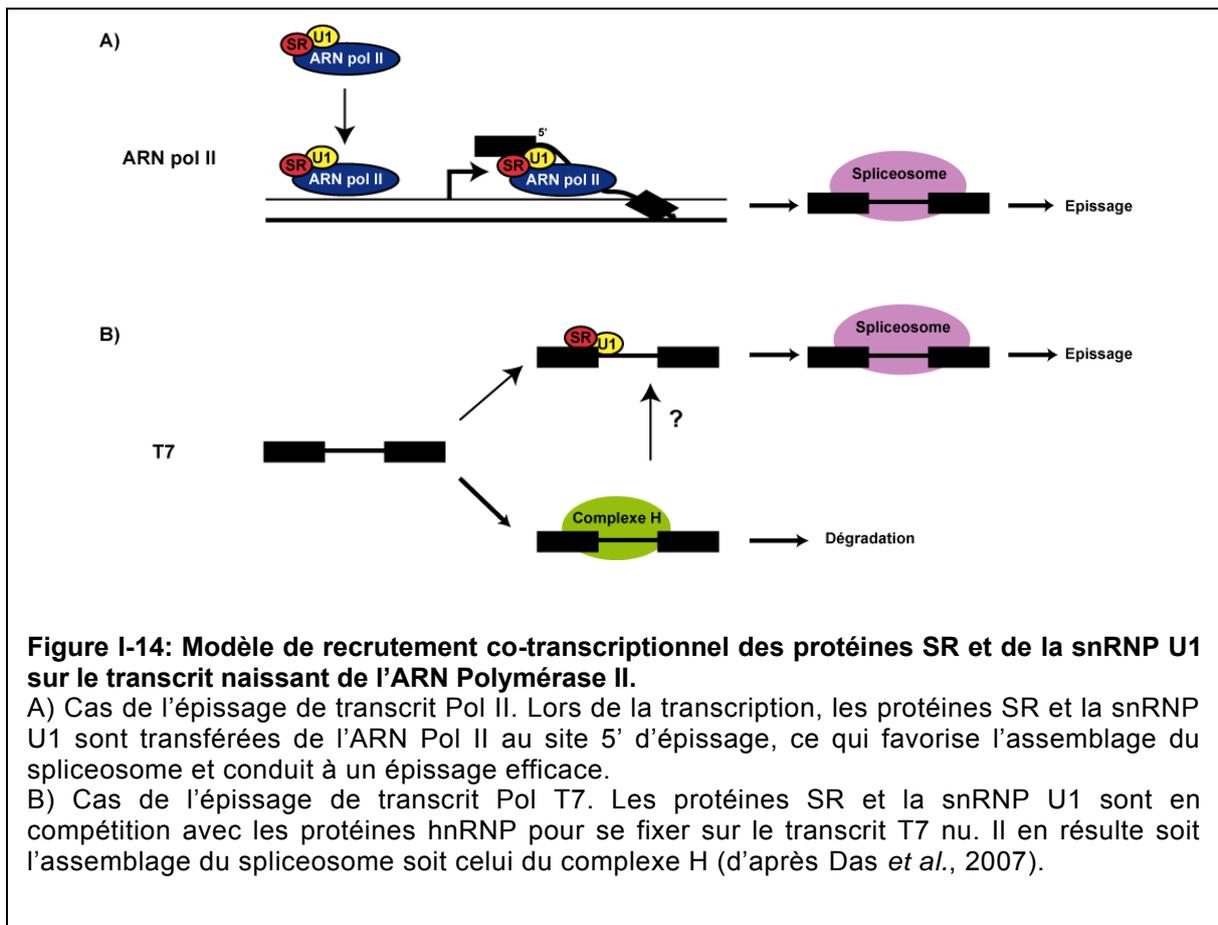
### **I.5 Lien entre épissage des pré-ARNm nucléaires, transport et système NMD**

Les transcrits primaires synthétisés par l'ARN polymérase II (pré-ARNm) subissent plusieurs étapes de maturation avant de devenir un ARN messager fonctionnel qui quitte le noyau pour être traduit en protéine dans le cytoplasme. Les étapes de maturation du pré-ARNm incluent l'addition d'une coiffe 7-méthyl guanosine (m7G) à l'extrémité 5', l'épissage des introns et la formation de l'extrémité 3' par clivage et polyadénylation.

#### **I.5.1 Lien entre transcription et épissage**

Transcription et épissage du pré-ARNm sont des processus extrêmement complexes qui impliquent des interactions protéine-protéine, ADN-protéine, et ARN-protéine. Un des plus grands gènes humain (2400kb) code la dystrophine. Ce gène nécessiterait environ 16 heures pour être transcrit et il a été montré qu'il était épissé co-transcriptionnellement (Tennyson *et*

*al.*, 1995). Il était en effet difficile à concevoir que l'épissage de dizaines d'introns de la dystrophine doive « attendre » que la synthèse du pré-ARNm soit finie. Cette observation a permis de montrer que l'épissage a lieu co-transcriptionnellement et que la sélection des sites d'épissage précède la polyadénylation (Beyer *et al.*, 1988). Le terme « co-transcriptionnel » implique un couplage fonctionnel significatif entre transcription et événement de maturation des ARN (Figure I-14).



Toutefois, des réactions peuvent avoir lieu pendant la transcription simplement car elles sont relativement rapides comparées avec le temps nécessaire pour transcrire le gène en entier. L'épissage du pré-ARNm commence co-transcriptionnellement et est souvent continué post-transcriptionnellement, comme le montre l'exemple du gène de 40 kb BR1 (Balbiani ring gene) de *Chironomus tentans* dans lequel l'intron 3 à 5 kb de l'extrémité 5' du pré-ARNm est excisé simultanément avec la transcription, tandis que l'intron 4 à 600 paires de bases du site de polyadénylation n'est épissé que de 5 à 10 % pendant la transcription et le reste post-transcriptionnellement (Bauren *et al.*, 1994). Il semble donc plus judicieux de parler de couplage entre transcription et épissage plutôt que d'épissage co-transcriptionnel (Neugebauer, 2002).

### I.5.1.1 Rôle du CTD de l'ARN Pol II dans le couplage

Le couplage entre la transcription et la maturation des pré-ARNm peut être en partie dû à la capacité de l'ARN polymérase II à lier certains des facteurs de la machinerie de maturation, « l'usine à ARNm » selon Bentley (Bentley, 2002). Le domaine C-terminal (CTD) de l'ARN polymérase II joue un rôle central dans le processus de couplage : le raccourcissement du CTD cause des déficiences dans l'ajout de la coiffe, le processus de clivage/polyadénylation et l'épissage (McCracken *et al.*, 1997). L'ARN polymérase II purifiée et phosphorylée d'extraits nucléaires de cellules HeLa va en effet être capable d'activer l'épissage *in vitro* (Hirose *et al.*, 1999), grâce au CTD de l'ARN polymérase II et seulement à la condition que le substrat contienne un exon interne (Zeng *et al.*, 2000). Ces conclusions suggèrent un rôle direct pour le CTD dans la reconnaissance des exons et conduisent à la spéculation que le CTD ne serait plus seulement une zone de fixation pour les facteurs d'épissage mais aussi dans le rapprochement des exons et l'assemblage du spliceosome en regroupant différents facteurs d'épissage à proximité de l'ARN sortant du site catalytique de l'ARN Pol II. Le CTD de l'ARN Pol II pourrait en particulier avoir un rôle très important en recrutant des facteurs d'épissage aux abords des sites de transcription (Misteli *et al.*, 1999).

### I.5.1.2 Les protéines en lien avec transcription et épissage

Une analyse protéomique du spliceosome humain (Zhou *et al.*, 2002; pour revue, Jurica *et al.*, 2003) montre qu'au moins 30 des 145 protéines spliceosomales identifiées dans ce travail sont des participants connus ou supposés pour le couplage entre l'épissage et d'autres étapes de l'expression des gènes.

Par exemple, le co-facteur transcriptionnel TAT-SF1 et les facteurs de transcription CA150, XAB2 et SKIP, sont présents dans le spliceosome. A l'inverse, plusieurs facteurs d'épissage, dont les protéines SR, ont été retrouvés associés aux promoteurs ou aux diverses séquences activatrices de transcription. Certaines protéines ont même des fonctions connues dans les deux processus, dont le facteur p54nrb, qui se lie au site d'épissage 5' au sein de grands complexes dans les cellules HeLa, puis s'associe avec la forme hyperphosphorylée de l'ARN polymérase II et aux snRNP U1 ou U2. Ces complexes macromoléculaires contiennent aussi d'autres facteurs couplant transcription et épissage comme PSF et TLS (Kameoka *et al.*, 2004).

### I.5.1.3 Importance du promoteur vis à vis de l'épissage

Des études menées avec des promoteurs différents ont montré que le changement de séquence du promoteur affecte fortement la sélection du site d'épissage. L'exon EDI code une répétition facultative de type III de la fibronectine et son inclusion est régulée durant le développement et dans des processus prolifératifs. Ce phénomène est totalement indépendant du niveau de production d'ARN (autrement dit de la force du promoteur), mais fait plutôt appel aux caractéristiques intrinsèques d'initiation du promoteur (Cramer *et al.*, 1999). L'ARN pré-messager synthétisé par l'ARN polymérase II est immédiatement dirigé dans la voie d'assemblage du spliceosome. Au contraire, l'ARN pré-messager synthétisé par l'ARN polymérase T7 est assemblé en complexe H non spécifique.

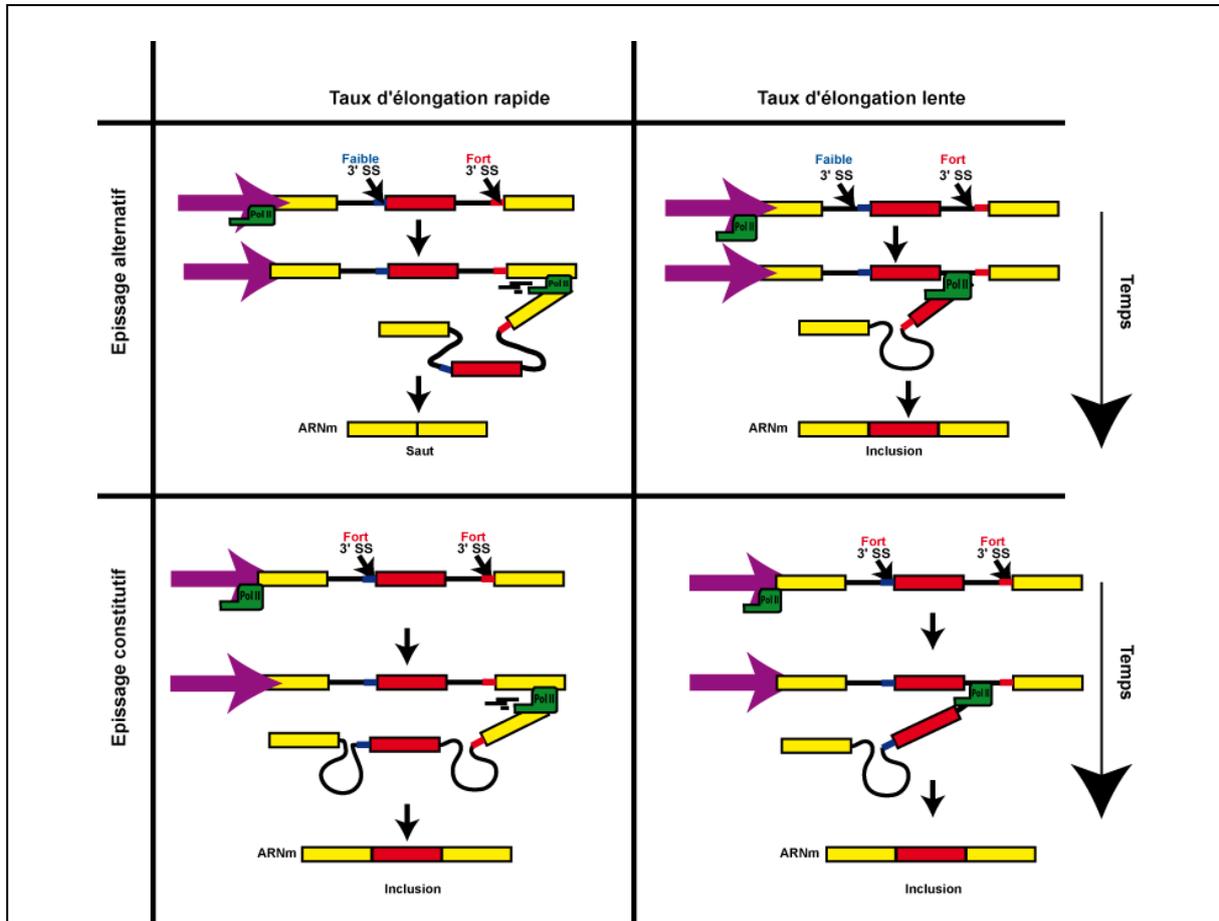
La découverte que la structure du promoteur est importante pour l'épissage alternatif suggère que les facteurs qui régulent cet épissage et qui peuvent agir par le promoteur ne résultent pas simplement d'une abondance variable de protéines SR mais d'un processus plus complexe impliquant une occupation du promoteur spécifique de la cellule (Das *et al.*, 2006; Das *et al.*, 2007). Toutefois, étant donné que les promoteurs ne sont pas échangés dans la nature comme on peut le faire *in vitro* et la plupart des gènes n'ayant un seul promoteur, la seule voie concevable par laquelle l'architecture du promoteur peut contrôler l'épissage alternatif *in vivo* est l'occupation différentielle des promoteurs par des facteurs de transcription avec des types différents de domaines d'activation et/ou des propriétés différentes. Il a ainsi été montré que les activateurs transcriptionnels avec des actions différentes sur l'initiation par l'ARN polymérase II et l'élongation affectent l'épissage alternatif différemment (Nogues *et al.*, 2002).

Ainsi, il a été démontré que bon nombre d'activateurs transcriptionnels influencent l'efficacité d'épissage (pour revue, Kornblihtt, 2005).

### I.5.1.4 Rôle de la vitesse d'élongation de l'ARN polymérase II

Un des modèles de régulation propose que les promoteurs puissent aussi moduler l'épissage alternatif via la régulation de l'élongation par l'ARN polymérase II ou la processivité (Figure I-15). Une faible vitesse ou des pauses internes dans l'élongation favoriseraient l'inclusion des exons alternatifs gouvernés par un mécanisme de saut d'exon, tandis qu'une vitesse d'élongation rapide de l'ARN polymérase II, en absence de pauses internes, favoriserait l'exclusion de ce type d'exons (Nogues *et al.*, 2002). Le saut d'exon de EDI se produit en effet car le site 3' d'épissage de l'intron en amont (intron 1) est sous-optimal comparé au site

3' d'épissage de l'intron en aval (intron 2). Si la polymérase pause n'importe où entre ces deux sites, seule l'élimination de l'intron 2 peut avoir lieu.



**Figure I-15: Influence de la vitesse d'élongation de l'ARN polymérase II sur l'épissage alternatif par "exon skipping".**

Epissage alternatif: Quand le site d'épissage 3' (SS) d'un exon alternatif est plus faible que le site 3' d'épissage de l'intron en amont, une faible vitesse d'élongation de la transcription favorise l'inclusion de l'exon, tandis qu'un taux d'élongation rapide favorise le saut.

Epissage constitutif: quand les 2 sites d'épissage 3' sont forts, l'exon est inclus constitutivement indépendamment de la vitesse d'élongation.

Une fois que la pause est passée et que la polymérase repart, la machinerie d'épissage est obligée d'éliminer l'intron 2, ce qui conduit à l'inclusion de cet exon. Une ARN polymérase II hautement processive ou l'absence de pauses internes favoriseront la présentation simultanée des deux introns de la machinerie d'épissage. Dans cette situation, le site 3' d'épissage le plus fort de l'intron 2 rentre en compétition avec le site 3' d'épissage le plus faible de l'intron 1, résultant en un saut d'exon. Dans ces conditions, la vitesse d'élongation de l'ARN polymérase affecte les quantités relatives d'isoformes d'épissage.

Ce modèle a pu être vérifié grâce à l'utilisation d'inhibiteur de l'élongation tel que DRB, qui est un inhibiteur des kinases responsables de la phosphorylation du CTD. Son utilisation lors

de la transcription du reporter d'épissage EDI augmente de trois fois l'inclusion de l'exon interne (Nogues *et al.*, 2002).

### **I.5.1.5 Influence de l'épissage sur la transcription**

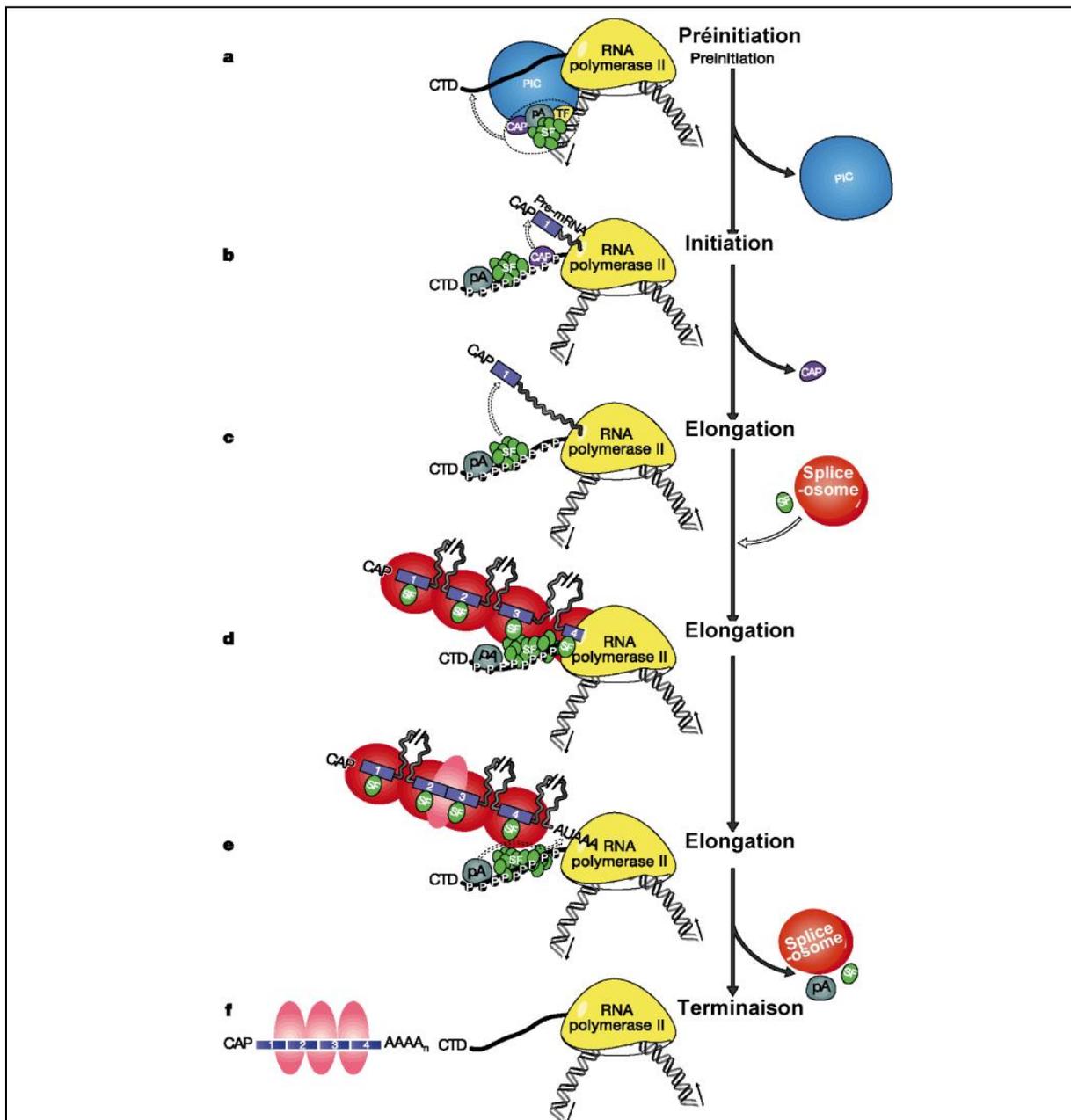
L'épissage alternatif peut donc être régulé par le processus de transcription, mais la machinerie d'épissage peut elle aussi influencer la transcription de différentes façons (Figure I-16).

- Fong et Zhou ont montré que les UsnRNP spliceosomaux interagissent avec le facteur humain d'élongation de la transcription p-TEFb et stimulent fortement l'élongation par l'ARN polymérase II probablement par la liaison de TAT-SF1 à p-TEFb. Ce complexe TAT-SF1-UsnRNP stimule aussi l'épissage *in vitro*, ce qui montre que des facteurs d'épissage peuvent stimuler directement l'élongation transcriptionnelle et que ces interactions réciproques existent dans le processus de couplage (Fong *et al.*, 2001).

- La présence d'un intron ou simplement d'un site 5' d'épissage immédiatement en aval d'un promoteur améliore nettement la transcription, à la fois chez les mammifères et les levures (Furger *et al.*, 2002), ce qui indique que les facteurs contrôlant l'excision des introns sont importants pour obtenir des niveaux normaux de transcription.

- Le snRNA U1 s'associe avec le facteur de transcription générale TFIIH, ce qui stimule *in vitro* les étapes d'initiation et de ré-initiation de la transcription (Kwek *et al.*, 2002).

- La protéine SKIP a une double fonction, elle régule indépendamment l'élongation de la transcription *in vitro* et *in vivo* quand elle est associée à p-TEFb et l'épissage du pré-ARNm (Bres *et al.*, 2005).



**Figure I-16: La transcription et la maturation des ARN pré-messager sont des processus liés (pour revue, Maniatis *et al.*, 2002).**

Les machineries de la transcription, de la formation de la coiffe, de l'épissage et de la polyadénylation sont représentées schématiquement. PIC : Complexe de pré-initiation de la transcription. TF : Facteurs de transcription. CTD : Domaine carboxy-terminal de l'ARN polymérase II. CAP : Facteur de la coiffe. SF : Facteur d'épissage. pA : Facteurs de polyadénylation. P : Domaine CTD phosphorylé. Les ovales roses représentent les complexes EJC se trouvant en amont des jonctions exon-exon après la réaction d'épissage.

- Ujvari et Luse ont montré quant à eux que la présence de U2AF65 permet d'assister l'ARN polymérase et donc d'éviter le phénomène de pause au moment de la transcription de la séquence polypyrimidine (Ujvari *et al.*, 2004).

### I.5.2 Lien entre export, NMD et traduction : rôle de l'EJC

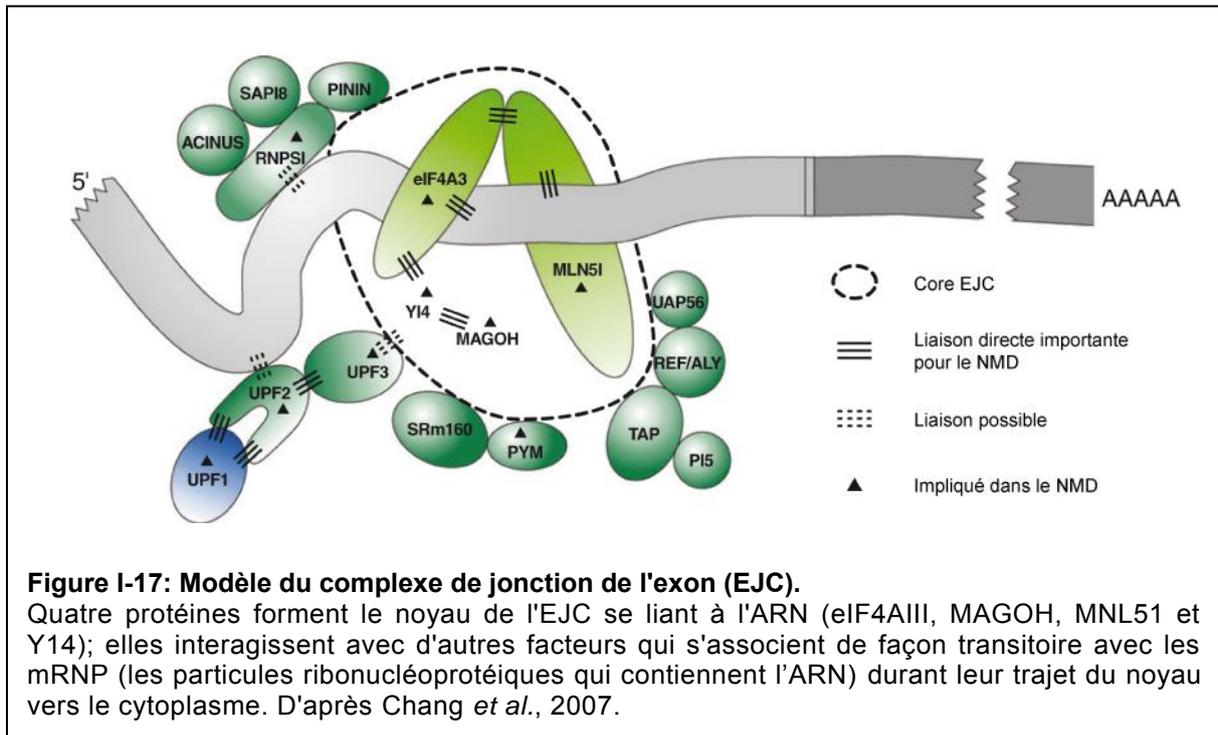
Chez les cellules de mammifères, l'EJC (Exon Junction Complex) s'assemble simultanément avec le processus d'épissage et avec l'ARN nouvellement formé dans une région à environ 24 nucléotides en amont de la jonction d'épissage. L'EJC est transporté avec l'ARNm mature vers le cytoplasme et reste associé avec le complexe d'épissage ARN-protéine jusqu'à ce que l'ARNm soit traduit. L'EJC est impliqué dans de nombreux processus cellulaires dont l'export nucléocytoplasmique des ARNm, la localisation subcellulaire et la traduction (Tange *et al.*, 2004; Giorgi *et al.*, 2007). L'EJC et ses protéines associées pourraient aussi être impliqués dans des mécanismes qui permettraient d'activer ou de réprimer la traduction (Ma *et al.*, 2008; Isken *et al.*, 2008).

#### I.5.2.1 Composants et assemblage du complexe EJC

Plus d'une douzaine de protéines ont été identifiées dans les complexes EJC (Figure I-17). Parmi elles, on trouve :

- des co-activateurs d'épissage : SRm160 (Blencowe *et al.*, 1998) et RNPS1 (Mayeda *et al.*, 1999),
- un facteur impliqué dans l'épissage alternatif : la pinine (Li *et al.*, 2003),
- des facteurs d'export : UAP56, REF/Aly (Zhou *et al.*, 2000), NXF1/TAP : p15,
- les protéines du domaine core : Y14 (Kataoka *et al.*, 2000), Magoh (Kataoka *et al.*, 2001); (Le Hir *et al.*, 2001), MLN51 (Degot *et al.*, 2004) et eIF4AIII (Chan *et al.*, 2004; Ferraiuolo *et al.*, 2004; Shibuya *et al.*, 2004),
- des facteurs du NMD : Upf3a/b, Upf2 et Upf1 (Leeds *et al.*, 1991; Cui *et al.*, 1995; Lee *et al.*, 1995),
- la protéine Acinus (Tange *et al.*, 2005),
- la protéine SAP 18 (Tange *et al.*, 2005).

Les études sur la structure du complexe EJC/ARN ont montré l'existence d'un noyau interne tétramérique qui est constitué des protéines eIF4A3, de MLN51 et de Y14/Magoh (Ballut *et al.*, 2005 ; Tange *et al.*, 2005 ; Le Hir *et al.*, 2008). Le noyau interne est couvert par RNPS1 et REF/Aly qui forment avec d'autres protéines la coque externe. UAP56 et le dimère d'export NXF1/p15 sont considérés comme des protéines navettes de l'EJC qui interagissent avec le complexe EJC mRNP d'une façon plus transitoire (Tange *et al.*, 2005).



L'EJC est une structure dynamique dont la composition en protéines change lors des différentes étapes de maturation de l'ARNm auquel il est associé d'abord dans le noyau puis dans le cytoplasme vers lequel il est exporté puis traduit (Reed, 2003; Dreyfuss *et al.*, 2002; Jurica *et al.*, 2003; Maquat, 2004).

### I.5.2.2 Rôle du complexe EJC :

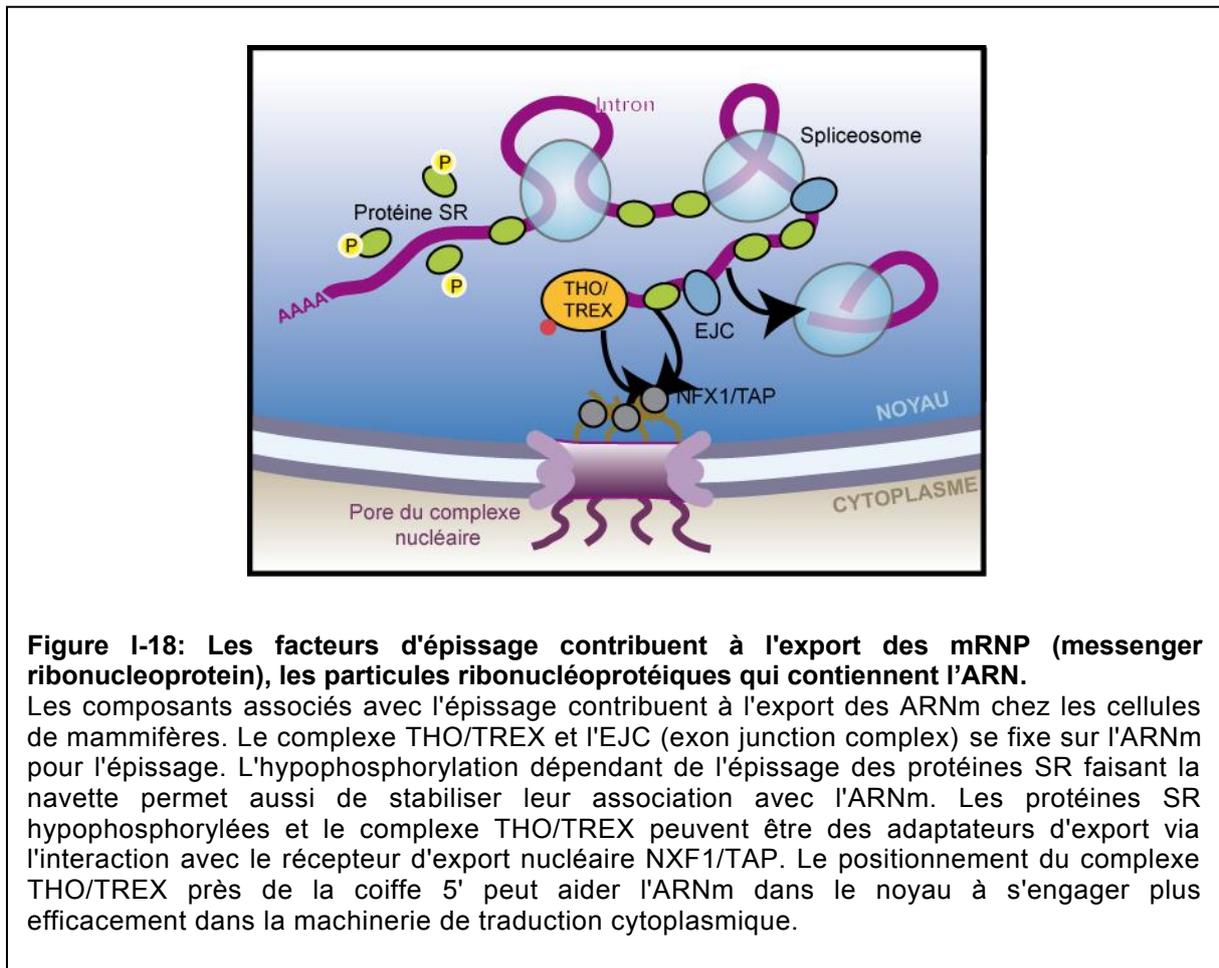
#### *-dans l'export des ARNm dans le cytoplasme*

Le complexe EJC jouerait un premier rôle de contrôle qualité des ARNm en facilitant l'export des ARNm épissés (Figure I-18). Des études décrivent un modèle dans lequel UAP56 et Aly/REF sont associés avec un ARNm épissé non pas avec l'EJC mais en tant que part du complexe TREX qui est recruté par l'extrémité 5' de l'ARNm (Cheng *et al.*, 2006). Dans ce modèle, c'est le complexe TREX qui est requis pour un export efficace de l'ARNm et qui peut interagir avec NFX1. Au moins 4 hélicases avec une activité ATPase peuvent jouer un rôle dans la libération de l'ARNm vers le cytoplasme.

#### *-dans la traduction des ARNm*

L'EJC va permettre aussi d'accroître l'efficacité de la traduction des ARNm épissés. Récemment il a été montré que PYM, un composant de l'EJC, interagit avec le complexe Magoh/Y14, le ribosome 40S et le complexe de pré-initiation 48S (Diem *et al.*, 2007) et que

leur interaction avec l'ARNm épissé va améliorer sa traduction. De plus, l'EJC lie le substrat Aly/REF-like (SKAR), un composant récemment identifié de l'EJC, et active les premiers cycles de traduction des ARNm épissés (Ma *et al.*, 2008). Toutefois, le mécanisme moléculaire précis d'activation de la traduction reste à élucider. Les facteurs Y14 et eIF4AIII sont quant à eux capables d'accroître la traduction des ARNm épissés avant et après la formation du complexe du ribosome 80S, respectivement. Ces résultats suggèrent que l'EJC module la traduction des ARNm épissés à de multiples étapes (Lee *et al.*, 2009).



### I.5.2.3 Complexe EJC et système NMD

Le NMD (Nonsense-mediated mRNA Decay) est un mécanisme couplé à la traduction qui élimine les ARNm contenant un codon prématuré de terminaison de la traduction (PTC). Le NMD est probablement un mécanisme de surveillance des ARNm pour empêcher la synthèse de protéines tronquées qui pourraient avoir des effets toxiques, mais son importance physiologique n'est pas encore très claire (Peltz *et al.*, 1993; Maquat, 1995).

Le modèle courant de NMD chez les mammifères est l'association de UPF3 avec l'EJC dans le noyau et le recrutement de UPF2 pendant l'export de l'ARNm vers le cytoplasme. Quand le ribosome en train de traduire va rencontrer un PTC en amont d'un EJC, le recrutement de UPF1 par la traduction relâche des facteurs et conduit à l'interaction avec les protéines UPF2 et UPF3 liées à l'EJC en aval et ainsi mène à l'assemblage du complexe de dégradation de l'ARNm (Lejeune *et al.*, 2005).

Chez les cellules de mammifères, le NMD est aussi lié à l'épissage des pré-ARNm et dans bien des cas une forte réduction des ARNm a lieu quand le PTC se trouve en amont d'un intron. Dans ce système, l'EJC sert d'intermédiaire entre l'épissage et le NMD (Nagy *et al.*, 1998; Thermann *et al.*, 1998; Zhang *et al.*, 1998).

## **CHAPITRE II : Régulation de l'épissage des introns des ARN pré-messagers nucléaires.**

### **II.1 Généralités sur l'épissage alternatif.**

L'épissage des pré-ARNm est un processus fondamental requis pour l'expression de la plupart des gènes des métazoaires. Il est accompli par le spliceosome qui catalyse l'excision des séquences introniques pour assembler les exons en ARNm matures avant son export et sa traduction. Etant donné la complexité des gènes des eucaryotes supérieurs et le niveau relativement bas de conservation des sites d'épissage, la précision de la machinerie d'épissage dans la reconnaissance des sites d'épissage est impressionnante. Des introns de taille allant de 100 à 100 000 bases sont retirés efficacement. Dans le même temps, un grand nombre d'événements d'épissage alternatif sont observés dans différents types cellulaires, durant le développement ou durant différents processus biologiques. Ceci implique une grande flexibilité du spliceosome qui doit être capable d'identifier les exons du pré-ARNm quelque soit le processus dans lequel l'ARNm est impliqué. Pour atteindre cette flexibilité, la sélection des sites d'épissage chez les eucaryotes supérieurs a évolué et dépend de multiples paramètres comme la force du site d'épissage, la présence ou l'absence de régulateurs d'épissage, la structure secondaire de l'ARN, l'architecture exon/intron et le processus de synthèse du pré-ARNm. La participation relative de chacun de ces paramètres contrôle l'efficacité de la reconnaissance des sites d'épissage et le retrait des introns flanquant.

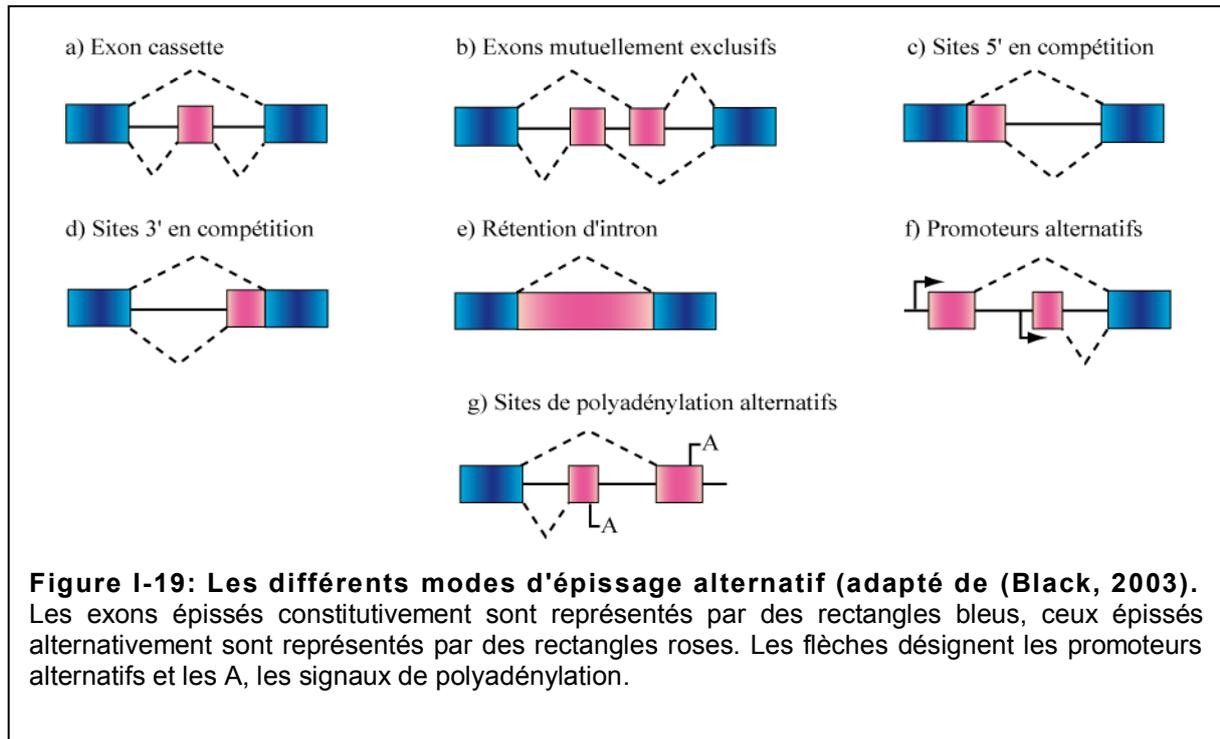
#### **II.1.1 Les différents modes d'épissage alternatif**

Le processus d'épissage alternatif peut se dérouler selon différents schémas (Figure I-19). Différentes combinaisons vont être utilisées pour générer des événements d'épissage alternatifs complexes (Black, 2003).

*-les exons dits «cassette».*

Un exon régulé qui est quelquefois inclus et quelquefois exclu de l'ARNm est appelé un exon cassette. Un exemple est la protéine hnRNP A1 qui autorégule l'épissage de son propre transcrit. Un exon cassette, 7b, est exclu dans la production de hnRNP A1, mais inclus dans les ARNm de hnRNP A1b. Cet exon est régulé à travers des sites de fixation de hnRNP A1 dans les introns l'entourant. Ceci conduit à un modèle dans lequel hnRNP A1 utilise sa

capacité de dimérisation pour interagir avec les deux éléments simultanément et bloquer la région contenant l'exon réprimé (Simard *et al.*, 2000).



-les exons «*mutuellement exclusifs*».

Des exons peuvent être dits mutuellement exclusifs quand 2 exons adjacents ou plus sont épissés de façon à ce que seulement un exon du groupe soit inclus à la fois. Dans ces systèmes, des mécanismes spéciaux peuvent renforcer les choix exclusifs. L'exemple de l'analyse de la structure du gène de l'échangeur  $\text{Na}^+/\text{Ca}^{2+}$  chez le lapin (Kofuji *et al.*, 1994) indique que le gène qui code cet échangeur est épissé alternativement pour donner naissance à 5 isoformes. L'analyse des séquences des liaisons intron/exon révèle la présence de 2 exons «*mutuellement exclusifs*» en conjonction avec 4 exons cassettes dans la région du gène qui code l'extrémité carboxyl-terminale de la région prédite comme une boucle intracellulaire. Ce réarrangement inhabituel d'exons du gène de l'échangeur  $\text{Na}/\text{Ca}$  peut permettre de générer plus de 32 ARNm différents.

-des exons de tailles variables du fait de sites 5' et 3' alternatifs.

Plusieurs sites d'épissage 5' ou 3' alternatifs en compétition les uns avec les autres peuvent être présents dans une même séquence exonique. La sélection d'un de ces sites conduit à l'expression d'une des isoformes de l'ARNm. Ce système est souvent utilisé par les virus, ce qui leur permet de produire de nombreux ARNm à partir d'un seul pré-ARNm. Par exemple,

l'unité E1A de l'adénovirus contient 3 sites 5' en compétition les uns avec les autres, et un seul site 3'. Leur utilisation permet la production de trois ARNm 9S, 12S et 13S. L'unité L1 de ce même virus contient, quant à elle, 2 sites 3' en compétition, les sites 55,52 kDa et IIIa (pour revue, Akusjarvi et Stevenin, 2003). Le pré-ARNm du virus VIH-1 contient un grand nombre de sites d'épissage en compétition, 4 sites donneurs et 8 sites accepteurs, ce qui permet la production d'une quarantaine d'ARNm distincts (Purcell et Martin, 1993).

*-Des exons terminaux alternatifs du fait de la présence de promoteurs ou de sites de polyadénylation alternatifs.*

Le choix des promoteurs transcriptionnels et des sites de polyadénylation influence l'identité des exons terminaux d'un ARNm. C'est le cas du transcrit primaire CT/CGRP commun à la calcitonine et au neuropeptide CGRP. Dans la thyroïde, l'utilisation du site de polyadénylation de l'exon 4 conduit à la production de la calcitonine, une hormone impliquée dans l'homéostasie du calcium, tandis que dans les neurones, le neuropeptide CGRP est produit par utilisation du site de polyadénylation de l'exon 6 (pour revue, Lou et Gagel, 1998). Plusieurs mécanismes d'épissage alternatif peuvent être combinés dans un même pré-ARNm. La production des isoformes d'ARNm est régulée selon la nature des sites d'épissage et par la présence de séquences régulatrices. Le choix des sites d'épissage peut dépendre du type cellulaire, de l'état de différenciation de la cellule, de l'activation des voies de signalisation ou encore, du sexe de l'individu. Tous les pré-ARNm pouvant être épissés de manière alternative et dont l'épissage est régulé, ont en commun la caractéristique d'avoir des sites d'épissage faibles.

*-la rétention d'intron.*

Un intron peut être inclus dans l'ARNm. Dans ce cas, on parle de rétention d'intron. Cette rétention peut être tissu-spécifique. Chez la drosophile, par exemple, l'épissage du 3<sup>ème</sup> intron de l'élément P produit une transposase dans les cellules germinales, alors que sa rétention dans les cellules somatiques conduit à la production d'une protéine qui réprime la transposition (Rio *et al.*, 1986).

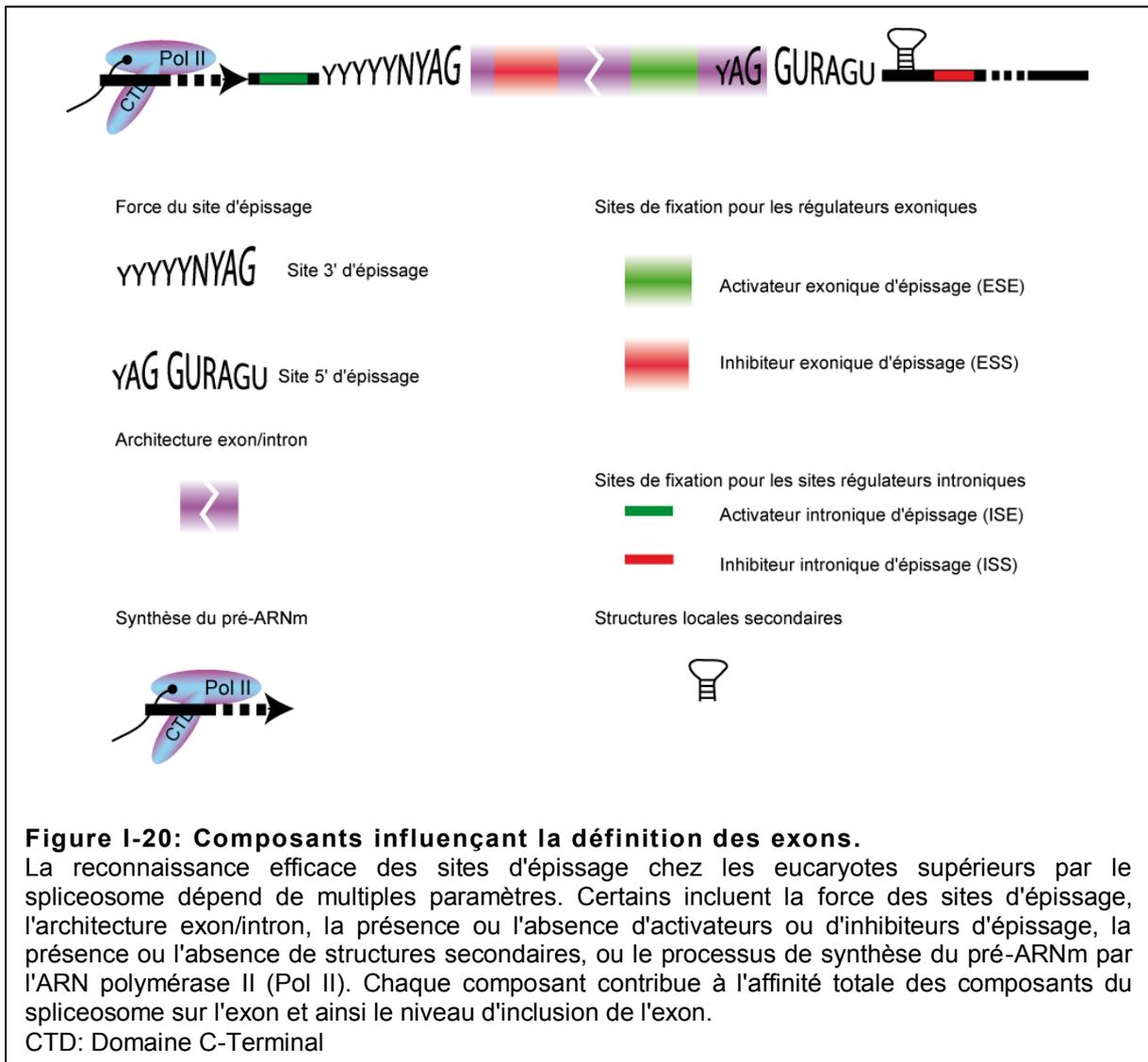
### II.1.2 Les sites d'épissage alternatifs sont souvent des sites sous optimaux

#### II.1.2.1 Caractéristiques de ces sites

Alors que les sites d'épissage forts utilisés constitutivement possèdent des séquences fortement similaires aux séquences consensus établies pour les sites 5' et 3' d'épissage, les sites d'épissage alternatifs ont souvent des séquences dégénérées par rapport aux séquences consensus. La force des sites 3' d'épissage dépend, en plus de la séquence du site 3' proprement dite, 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 non canonique, et parfois le nucléotide utilisé pour le branchement n'est même pas une adénosine. Ces caractéristiques rendent moins efficace l'interaction des composants spliceosomaux avec les séquences aux jonctions exon-intron et intron-exon (pour revue, (Blencowe, 2006).

#### II.1.2.2 L'utilisation des sites d'épissages faibles est régulée par des éléments en cis.

Dans les cas classiques d'épissage alternatif, les séquences *cis* de l'ARN augmentent l'inclusion de l'exon en servant de site de fixation pour l'assemblage de complexes activateurs d'épissage multi-composants. Comme ces éléments de séquence se trouvent dans l'exon régulé, on les nomme ESE pour activateurs exoniques d'épissage (exonic splicing enhancers) (Figure I-20). Les ESE sont généralement reconnus par au moins un membre de la famille des protéines SR qui recrute la machinerie d'épissage sur l'intron adjacent. Mais la régulation de l'épissage des pré-ARNm est plus complexe qu'un simple modèle de recrutement grâce aux ESE. Il existe d'autres séquences régulatrices exoniques ou introniques, activatrices ou inhibitrices (exonic splicing silencers: ESS, séquence inhibitrice exonique; intronic sequence silencer: ISS, séquence intronique d'épissage), voire avec une action bidirectionnelle activatrice ou inhibitrice selon le contexte.



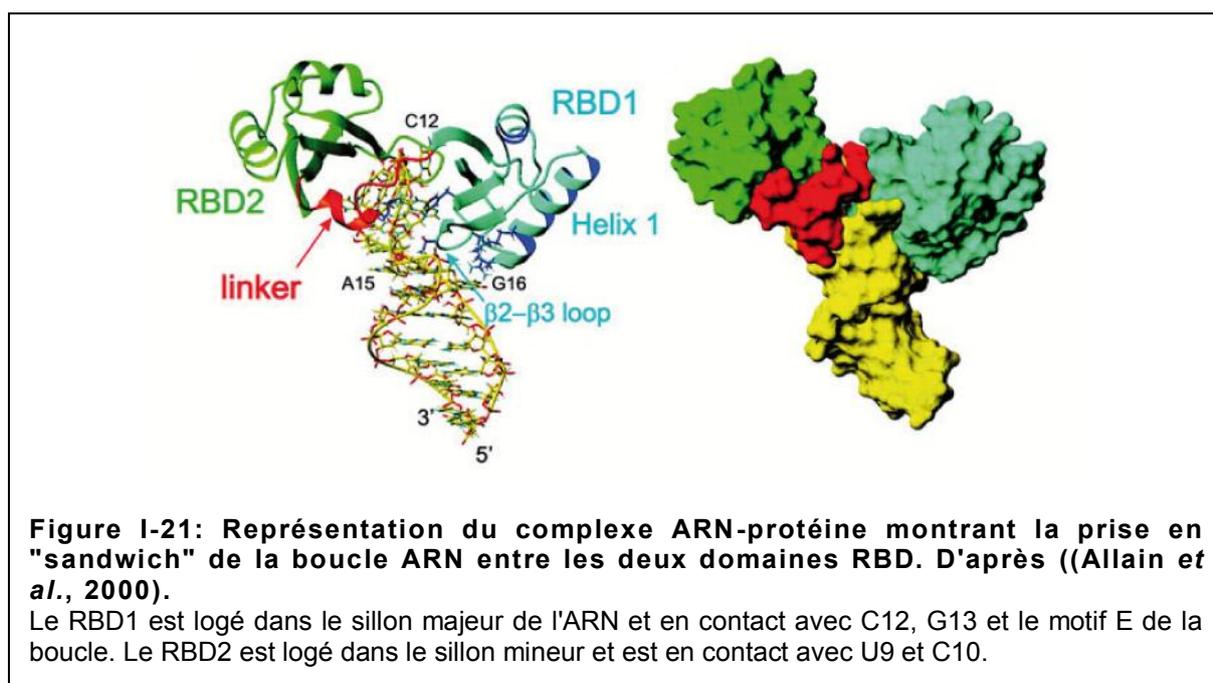
Les ESS sont reconnus le plus souvent par les membres de la famille des hnRNP, tandis que les ISS lient en général les protéines fixées à la séquence polypyrimidine. Les protéines hnRNP peuvent empêcher l'assemblage des spliceosomes en multimérisant sur les exons, bloquant ainsi le recrutement d'autres protéines (pour revue, Blencowe, 2006).

## II.2 Protéines nucléaires jouant un rôle majeur dans la régulation de l'épissage

### II.2.1 Les différents types de domaines de fixation à l'ARN

#### -Les RRM

Le motif de reconnaissance de l'ARN (RRM : RNA Recognition Motif), aussi connu comme domaine RBD (RNA Binding Domain) ou domaine RNP (RiboNucleoprotein Domain) est l'un des domaines de liaison à l'ARN les plus couramment rencontrés (Dreyfuss *et al.*, 1988) (Figure I-21).



Chez l'humain, 497 protéines contenant au moins un RRM ont été identifiées. Ce domaine d'environ 90 acides aminés contient une séquence centrale de 8 résidus conservés, principalement aromatiques et chargés positivement (Adam *et al.*, 1986; Swanson *et al.*, 1987). Cette séquence, appelée la séquence consensus RNP1 est définie par les acides aminés (Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe/Tyr-Val/Ile/Leu-X-Phe/Tyr) ou X est n'importe quel acide aminé. Une seconde séquence consensus moins conservée nommée RNP2, et contenant dans sa séquence 6 résidus dans le domaine N-terminal, a été identifiée plus tard (Ile/Val/Leu-Phe/Tyr-Ile/Val/Leu-XAsn-Leu) (Bandziulis *et al.*, 1989; Kenan *et al.*, 1991; Birney *et al.*, 1993).

Ce domaine protéique abondant est biologiquement important et impliqué dans de nombreuses fonctions cellulaires. Les RRM eucaryotiques sont présents dans tous les événements post-transcriptionnels: la maturation des pré-ARNm avec les protéines Cst-64, LA ou UPF3, l'épissage (U2B', U2AF35, U2AF65, hnRNP A1 ou Y14), l'épissage alternatif (hnRNP A1, PTB, sex-lethal, protéines SR), la stabilité des ARNm (CBP20, PABP ou HuD), l'export des ARNm (TLS) ou la formation des complexes pré-ARNr (nucléoline) par exemple.

Le RRM est replié en structure sandwich  $\alpha\beta$  avec une topologie  $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$  composée de 4 feuillets  $\beta$  antiparallèles arrangés spatialement dans l'ordre  $\beta_4 \beta_1 \beta_3 \beta_2$  et deux hélices  $\alpha$  ( $\alpha_1$  et  $\alpha_2$ ) (Figure I-22). Les motifs RNP1 et RNP2 sont localisés dans les brins centraux des feuillets  $\beta$  hautement conservés,  $\beta_3$  et  $\beta_1$  respectivement (pour revue, Clery *et al.*, 2008).

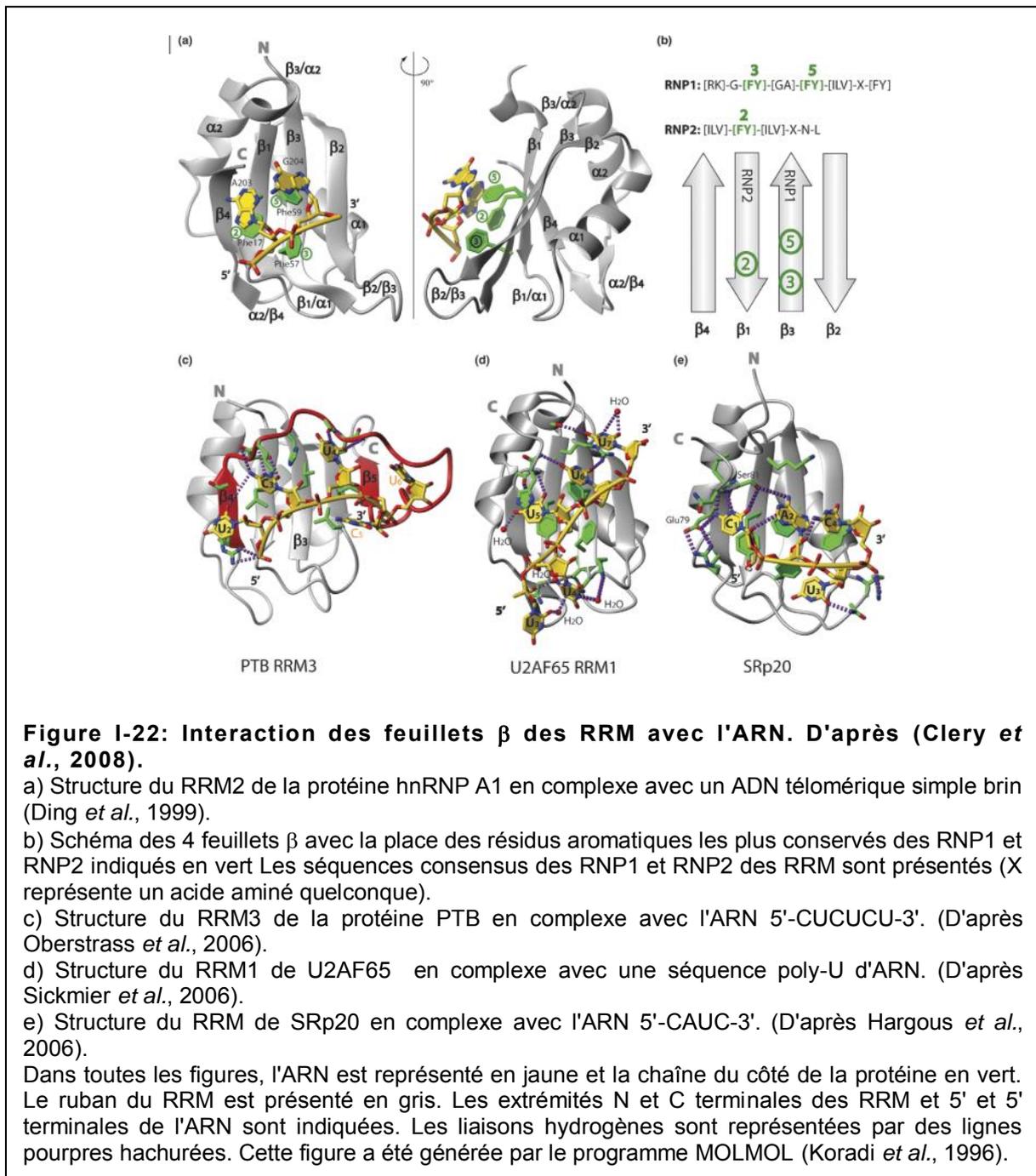
Plus de 30 structures de RRM ont été déterminées à ce jour, par RMN et cristallographie aux rayons X et des variations dans la structure des RRM ont pu être rencontrées. Le N-terminal du RRM de CstF-64 présente en effet une courte hélice  $\alpha$  supplémentaire (Perez Canadillas *et al.*, 2003), et un brin  $\beta$  peut être rajouté à un feuillet comme dans les RRM 2 et 3 de PTB (Conte *et al.*, 2000; Simpson *et al.*, 2004). Un autre exemple est rencontré avec la structure secondaire dans le RRM de U2AF35, prolongée par une hélice  $\alpha$  additionnelle. Elle est trois fois plus longue que dans un RRM classique et impliquée dans les interactions protéine-protéine (Kielkopf *et al.*, 2004).

De nombreux RRM lient les ARN avec une haute affinité, avec un Kd de l'ordre du nM, et une haute spécificité. Cependant, la spécificité de séquence n'implique pas nécessairement une haute affinité (Maris *et al.*, 2005). Pour avoir une affinité plus forte, certains RRM utilisent les 2 brins  $\beta_4$  et  $\beta_2$  externes, tandis que d'autres utilisent les boucles 1, 3 ou 5 des extrémités C ou N-terminales (Varani *et al.*, 1998).

Les domaines RRM sont souvent présents en de multiples copies dans les protéines (de 2 à 6 pour 44% des protéines), ce qui augmente l'affinité de fixation à l'ARN mais également la spécificité de fixation (pour revue, (Maris *et al.*, 2005). Les RRM sont souvent associés au sein des protéines avec un autre domaine, le doigt de zinc. Par exemple, la protéine U2AF35 renferme un domaine RRM et deux doigts de zinc (Selenko *et al.*, 2003).

Les études biochimiques et structurales ont montré que les RRM étaient aussi impliqués dans les interactions protéine-protéine selon 3 modes :

-les interactions protéiques entre 2 RRM: la première structure montrant une interaction entre deux RRM est la région N-terminale de hnRNP A1 (UP1) dans une forme libre qui contient deux RRM séparés par un linker court (Xu *et al.*, 1997; Shamoo *et al.*, 1997). Les deux RRM forment un repliement compact et interagissent l'un avec l'autre via leur hélice  $\alpha 2$ . Ces interactions RRM-RRM ne sont pas caractéristiques de tous les RRM des protéines. Dans le cas des protéines sex-lethal et nucléoline, dans les protéines libres, le linker est flexible et les deux RRM sont indépendants (Allain *et al.*, 2000; Crowder *et al.*, 1999).

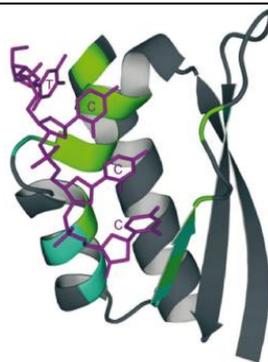


-les interactions entre un domaine RRM pouvant lier l'ARN et une protéine non-RRM : dans certains cas les protéines contenant des RRM ne peuvent s'associer avec l'ARN qu'en présence d'une autre protéine qui agit comme co-facteur. Les protéines U2B' et CBP20 ont besoin toutes deux d'un co-facteur, U2A' et CBP80, respectivement, pour reconnaître l'ARN. Les domaines RRM de U2B' et CBP20 interagissent avec le motif LRR (Leucine Rich Repeat) ou le domaine MIF4G à travers leurs hélices  $\alpha$  et leur boucle 4, gardant le feuillet  $\beta$  accessible pour la liaison à l'ARN (Price *et al.*, 1998; Mazza *et al.*, 2001; Mazza *et al.*, 2002; Calero *et al.*, 2002).

-les interactions entre un RRM qui ne fixe pas l'ARN et une autre protéine: certaines protéines contenant des domaines RRM sont impliquées dans des interactions protéines-protéines mais pas dans des interactions ARN-protéines. Deux différentes situations peuvent exister. Dans un cas, l'interaction protéique implique le feuillet  $\beta$  du domaine RRM, empêchant ainsi la liaison à l'ARN comme dans le complexe Y14-Magoh (Selenko *et al.*, 2003; Fribourg *et al.*, 2003; Lau *et al.*, 2003; Bono *et al.*, 2004) ou dans le complexe UPF2-UPF3 (Kadlec *et al.*, 2004). Dans le second cas, l'interaction se fait par l'intermédiaire de l'hélice  $\alpha$ , laissant le feuillet  $\beta$  exposé aux solvants et par conséquent capable de lier l'ARN, comme dans le cas de U2AF35-U2AF65 (Kielkopf *et al.*, 2001) et dans les complexes U2AF65-SF1 (Selenko *et al.*, 2003).

### ***-Le domaine KH***

Le domaine KH est nommé ainsi car il a été identifié pour la première fois au sein de la protéine hnRNP K (Siomi *et al.*, 1993) (Figure I-23).



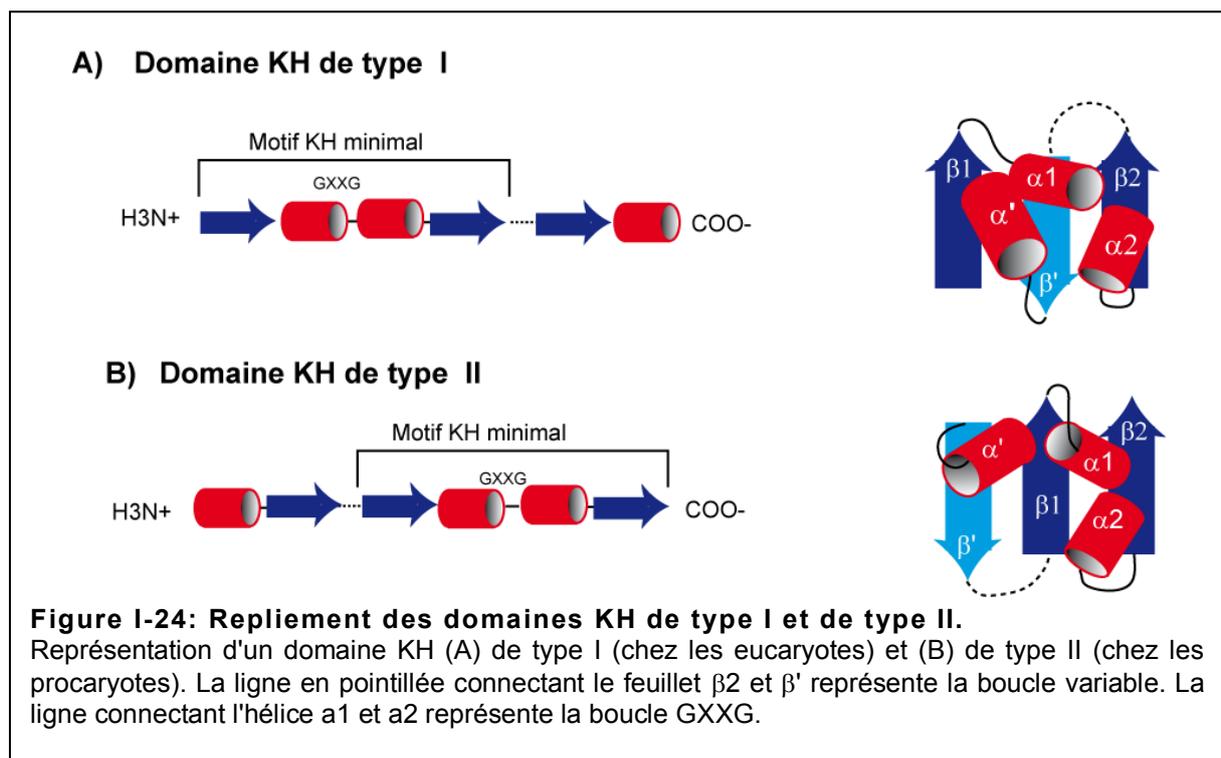
**Figure I-23: Structure en solution du domaine KH3 de hnRNP K lié à un ADN simple brin.**

Ce troisième domaine KH reconnaît une séquence 5'-dTCCC (en violet). Les régions de la protéine qui sont en contact avec l'acide nucléique ligand sont en vert (hydrophile) et en cyan (polaire).

Ce motif comprend environ 70 acides aminés dont l'octapeptide conservé Val-Ile-Gly-X2-Gly-X2-Ile (où X correspond à n'importe quel résidu) qui est situé environ au centre du motif et est retrouvé dans une grande diversité de protéines chez les archées, les bactéries et les eucaryotes (Grishin, 2001). Les domaines KH se retrouvent souvent en copies multiples, deux par exemple dans la protéine FMRP (Fragile X Mental Retardation Protein) (Siomi *et al.*, 1993; Ashley *et al.*, 1993; O'Donnell *et al.*, 2002), trois dans hnRNP K (Siomi *et al.*, 1993) (Ostareck-Lederer *et al.*, 2004) et 14 dans la vigiline (McKnight *et al.*, 1992; Currie *et al.*, 1999).

Les domaines KH peuvent reconnaître des ARN et des ADN simple brin et lorsqu'ils sont présents en plusieurs copies, ils peuvent fonctionner indépendamment ou en coopération. On retrouve ces domaines dans différents types de protéines qui sont impliquées dans de nombreux processus biologiques différents, dont l'épissage, la régulation transcriptionnelle et le contrôle traductionnel.

Grishin décrit deux versions différentes des motifs KH, de type I et de type II (Figure I-24) (Grishin, 2001). Le repliement de type I est retrouvé typiquement dans les protéines eucaryotiques tandis que le type II est retrouvé dans les protéines procaryotiques.



Les deux types partagent le même motif KH minimal ainsi qu'une boucle GXXG quelquefois altérée ou interrompue, mais l'arrangement tridimensionnel de la structure secondaire est différent. Dans sa forme minimale, le motif KH est replié dans une conformation  $\beta\alpha\alpha\beta$ . Les motifs KH de type I présentent en plus une extension C-terminale  $\beta\alpha$  (comme c'est le cas pour la protéine hnRNP K). Les motifs KH de type II présentent une extension N-terminale  $\alpha\beta$  (par exemple pour la protéine ribosomale S3) (pour revue, (Grishin, 2001).

La protéine hnRNP K sera décrite plus longuement dans un chapitre suivant.

### **-Les domaines RGG ou GAR**

Les domaines riches en arginine (R) et en glycine (G) ont initialement été identifiés dans la protéine hnRNP U (Kiledjian *et al.*, 1992). Ils sont trouvés dans une grande variété de protéines impliquées dans la maturation des ARN et dans le trafic intracellulaire des ARN comme les snRNP, les hnRNP et les snoRNP (small nucleolar RNP) et sont souvent des substrats pour la méthylation des arginines. Les domaines RGG contribuent eux-mêmes à la liaison à l'ARN à travers des interactions non spécifiques et semblent participer à l'assemblage et la maturation des complexes RNP (Kiledjian *et al.*, 1992; Ghisolfi *et al.*, 1992; Nadler *et al.*, 1991; Burd *et al.*, 1994). Les domaines RGG contribuent à la régulation et au couplage des événements de maturation des ARN. Les capacités de fixation à l'ARN de ces domaines sont régulées par des enzymes qui les modifient post-traductionnellement (kinases, phosphatases et méthylases) (Huang *et al.*, 2004; Lee *et al.*, 2005; McBride *et al.*, 2001).

Beaucoup d'autres protéines contiennent des domaines de type RRM ou RS associés à des domaines RGG. Certaines hnRNP qui font la navette entre le cytoplasme et le noyau et participent à l'export des ARN ont à la fois des domaines RS et RGG, comme la protéine Y14, membre de l'EJC. La nucléoline quant à elle contient 4 domaines RRM et un motif RGG situé à l'extrémité C-terminale (Ghisolfi *et al.*, 1992). Les protéines hnRNP, et en particulier hnRNP A1 et hnRNP H, impliquées dans la régulation de l'épissage contiennent aussi à la fois des domaines RRM et RGG (pour revue, (Krecic *et al.*, 1999). Nous disposons de peu d'indications sur la structure tridimensionnelle des motifs RGG, ceci sans doute du fait de la flexibilité de ces domaines et de leur faible solubilité qui rendent difficiles les études par radiocristallographie. La spécificité d'interaction de ces motifs avec l'ARN semble très faible. Néanmoins, il a été montré que le motif RGG situé en N-terminal de la protéine FMRP

présente une préférence pour les séquences riches en G, structurées en G-quartet (Ramos *et al.*, 2003).

### II.2.2 Les protéines hnRNP

Les protéines hnRNP (heterogeneous nuclear ribonucleoproteins) sont extrêmement abondantes dans les cellules et leur caractéristique principale est qu'elles se lient aux ARN, en particulier aux transcrits ARN naissants issus de l'ARN polymérase II. Plus de 20 hnRNP ont été identifiées et nommées par ordre alphabétique en fonction de leur taille, de hnRNP A1 à hnRNP U (Dreyfuss *et al.*, 1993). Ces protéines sont impliquées dans un grand nombre de processus biologiques dont la biogénèse des télomères, la traduction, la stabilité des ARN et pour certaines d'entre elles (hnRNP A1, A2, F, H, I [PTB], E, K et L) dans l'épissage alternatif.

Les complexes hnRNP des mammifères sont des mono particules de 20 nm (Beyer *et al.*, 1977) sédimentant à 40 S et contenant des ARN d'environ 700 nucléotides. Certaines protéines hnRNP sont très abondantes (environ 100 millions de copies par noyau pour les hnRNP A/B) tandis que d'autres sont seulement présentes dans de faibles taux (Kamma *et al.*, 1995; Hanamura *et al.*, 1998; Markovtsov *et al.*, 2000). Les hnRNP contiennent des motifs pour se lier à l'ARN (RRM, RGG ou KH) de manière plus ou moins spécifique, ainsi que des domaines auxiliaires composés d'acides aminés de distribution assez inhabituelle, qui pourraient servir pour des interactions protéine-protéine. Les hnRNP sont majoritairement localisées dans le nucléoplasme et plusieurs sont capables de faire la navette entre le noyau et le cytoplasme (Dreyfuss *et al.*, 1993; Nakielny *et al.*, 1997) et d'interagir avec elles-mêmes et d'autres facteurs.

Je détaillerai dans la suite de l'introduction le rôle des protéines hnRNP lors de l'épissage, et plus particulièrement celui des protéines hnRNP A/B et hnRNP K.

#### I.2.2.1 Les protéines hnRNP A/B :

Cette sous-famille de hnRNP a tout d'abord été décrite comme un groupe de protéines de faible poids moléculaire isolé du « noyau » 40S des particules hnRNP de cellules HeLa (Beyer *et al.*, 1977). La famille des gènes hnRNP A/B consiste en 4 gènes A0, A1, A2 et A3, tous connus pour être épissés alternativement afin de générer différentes isoformes de protéines. Les gènes HNRNPA0 et HNRNPA3 codent respectivement les protéines hnRNP

A0 et A3 et leurs isoformes (Makeyev *et al.*, 2005). Le gène HNRNPA2 (ou HNRNPA2B1) code les protéines hnRNP A2, B0, B1 et B2. Enfin, le gène HNRNPA1 code la protéine hnRNP A1 et également la protéine hnRNP A1<sup>B</sup> dont l'ARNm présente un exon supplémentaire (pour revue, (Chabot *et al.*, 2003).

### **a) Organisation structurale des hnRNP A/B**

Les protéines de la famille hnRNP A/B sont caractérisées par 2 RRM très conservés en N-terminal, liés par une région riche en glycine contenant une boîte RGG en C-terminal appelé M9 qui est un signal d'import/export nucléocytoplasmique (Dreyfuss *et al.*, 1993). Les protéines hnRNP A/B s'associent de préférence avec des ARN plutôt qu'avec des ADN (Nadler *et al.*, 1991). L'interaction de hnRNP A/B avec des polynucléotides a été observé pour la première fois pour UP1, un produit protéolytique de hnRNP A1 généré par une protéase de type trypsine (Pandolfo *et al.*, 1985). UP1-B est le produit de protéolyse de hnRNP A2 (Kumar *et al.*, 1986).

La spécificité d'interaction des protéines hnRNP A/B avec l'ARN a été particulièrement établie pour hnRNP A1 et hnRNP A2. hnRNP A1 s'associe préférentiellement avec la séquence UAUGAUAGGGACUUAGGGUG, dans laquelle la séquence courte la plus indispensable est UAGGGA(U) (Burd *et al.*, 1994). La hnRNP A1 recombinante se lie aussi à l'ARN contenant des séquences riches en séquences AUUUA *in vitro* (Hamilton *et al.*, 1993). Par leur fixation à des séquences cibles sur ces ARN (Ainger *et al.*, 1997), telles que les séquences A2RE1 de l'ARN du virus HIV-1 (Mouland *et al.*, 2001), les protéines hnRNP A2 et B1 permettent la localisation de certains ARNm dans des zones définies du cytoplasme.

### **b) Rôles des protéines de la famille hnRNP A/B**

*Dans la maintenance des chromosomes, la réplication et la réparation de l'ADN.*

Lors de la réplication, UP1 stimule l'activité de l'ADN polymérase  $\alpha$  (Kumar *et al.*, 1986) (Williams *et al.*, 1985), une enzyme qui synthétise une matrice ARN-ADN (le segment  $\alpha$ ) et initie la formation des fragments d'Okazaki durant la synthèse du brin d'ADN retardé (Chai *et al.*, 2003).

Les rôles de hnRNP A/B dans le métabolisme de l'ADN incluent aussi la maintenance des télomères, complexes protéine-ADN qui coiffent l'extrémité du chromosome, empêchant la

machinerie de réparation des cassures d'ADN double-brin de les fusionner ensemble (Slijepcevic *et al.*, 2005).

hnRNP B1, surexprimée dans les stades précoces de cancers de poumons, peut jouer un rôle dans la réparation de l'ADN (Iwanaga *et al.*, 2005). Cette protéine s'associe avec le complexe protéine kinase dépendant de l'ADN (DNA-PK) qui va permettre la réparation de l'ADN double-brin (Smith *et al.*, 1999).

### Dans la transcription des gènes.

Certaines protéines hnRNP A/B s'associent avec de multiples séquences promotrices et participent ainsi à la régulation de la transcription en tant qu'activateur ou répresseur. hnRNP A1 et hnRNP A2/B1 lient la région promotrice de c-myc (Takimoto *et al.*, 1993) et de APOE (Campillos *et al.*, 2003). hnRNP A3 agit aussi en facteur de transcription en se fixant à la région régulatrice du gène *Hoxc8* (Makeyev *et al.*, 2005).

### Dans l'export nucléaire de l'ARNm.

Les transcrits ARNm matures sont exportés du noyau accompagnés par un complexe hnRNP (Mili *et al.*, 2001). Deux récepteurs de transport de la famille de la karyophérine- $\beta$ , la transportine 1 (Trn1) et la transportine 2 (Trn2) ont été identifiés comme régulateurs pour l'import nucléaire des protéines hnRNP A/B (Pollard *et al.*, 1996; Fridell *et al.*, 1997; Rebane *et al.*, 2004). La transportine est capable de lier la nucléoporine et place la hnRNP A1 au cœur du complexe du pore nucléaire durant l'import nucléaire (Bonifaci *et al.*, 1997).

### Dans l'épissage.

Les protéines hnRNP A/B sont des composants essentiels du spliceosome et participent à l'épissage constitutif et alternatif (pour revue, Martinez-Contreras *et al.*, 2007). Les protéines majeures hnRNP A1, A2 et A3 sont parmi les protéines hnRNP qui sont présentes dans les spliceosomes à toutes les étapes majeures d'épissage (Hartmuth *et al.*, 2002; Jurica *et al.*, 2002; Makarov *et al.*, 2002; Zhou *et al.*, 2002; Rappsilber *et al.*, 2002; Jurica *et al.*, 2003; Makarova *et al.*, 2004; Deckert *et al.*, 2006).

Nous allons détailler plus particulièrement le rôle la protéine hnRNP A1 dans l'épissage. La hnRNP A1 cytoplasmique et la hnRNP A1 nucléaire ont des profils de fixation à l'ARN différents. La hnRNP A1 cytoplasmique a une haute affinité pour les éléments riches en AU

(Hamilton *et al.*, 1993; Henics *et al.*, 1994), tandis que la hnRNP A1 nucléaire a une haute affinité pour une séquence polypyrimidine bordée par AG à l'extrémité 3' des introns (Mayrand *et al.*, 1990; Swanson *et al.*, 1988). Cette différence de liaison à l'ARN est due à une déphosphorylation de sérines-thréonines qui peut être bloquée par des inhibiteurs des sérines-thréonines phosphatases *in vivo* (Hamilton *et al.*, 1997).

hnRNP A1 agit comme un facteur *trans* de la régulation de l'épissage *in vivo*. Les deux RRM, et plus particulièrement le résidu Phe dans le sous-motif RNP-1, sont essentiels pour la liaison spécifique de hnRNP A1 au pré-ARNm et la régulation de l'épissage alternatif. Pour certains transcrits comme celui du récepteur de type 2 du facteur de croissance humain, le domaine RGG à lui seul permet la régulation de l'épissage alternatif (Del Gatto *et al.*, 1995).

La protéine hnRNP A1 se fixe sur des séquences inhibitrices exoniques contenant des triplets UAG et réprime ainsi l'utilisation du site 3' en amont, en empêchant l'assemblage des complexes spliceosomaux autour de ce site d'épissage (Caputi *et al.*, 1999; Del Gatto-Konczak *et al.*, 1999). hnRNP A1 peut aussi stimuler l'utilisation des sites 5' distaux (pour revue, Martinez-Contreras *et al.*, 2007).

De nombreuses études ont montré que les protéines hnRNP sont antagonistes de l'action des protéines SR sur l'épissage (Mayeda *et al.*, 1992; Mayeda *et al.*, 1993; Caceres *et al.*, 1994; Yang *et al.*, 1994). En utilisant un pré-ARNm contenant 2 sites 5' d'épissage identiques en compétition l'un avec l'autre, il a été montré que la protéine ASF/SF2 induit l'utilisation du site 5' proximal, alors que la protéine hnRNP A1 induit l'utilisation du site distal (Mayeda *et al.*, 1992). La compétition entre le taux de hnRNP A1 fixé sur l'ARN et celui de ASF/SF2 influence la sélection du site d'épissage.

La fixation de hnRNP A1 aux éléments ESS bloque la reconnaissance des exons proximaux (Zhu *et al.*, 2001). Il arrive que des éléments ESS et ESE se recouvrent, les protéines SR et hnRNP A/B sont alors en compétition pour la fixation sur les éléments régulateurs, ce qui va déterminer la sélection du site d'épissage. La fixation d'une molécule de hnRNP A1 est suffisante pour éliminer l'association des protéines SR avec l'ESE. Quand ces sites régulateurs ne sont pas chevauchants, les protéines hnRNP A/B peuvent se lier coopérativement le long de l'exon, favorisant la répression de l'épissage (Mayeda *et al.*, 1992; Martinez-Contreras *et al.*, 2007). A forte concentration, la protéine hnRNP A1 multimérise le long du pré-ARNm, empêchant ainsi l'accès des facteurs d'épissage, en particulier de la snRNP U1. Un article très récent de l'équipe de A. Krainer montre que la fixation de la

protéine hnRNPA1 et la répression de son épissage peuvent avoir lieu sur un ARN non structuré (Okunola *et al.*, 2009). La fixation coopérative de hnRNP A1 se fait préférentiellement de 3' en 5' de l'exon et inhibe l'épissage en bloquant l'ESE, mais la fixation coopérative de hnRNP A1 peut aussi s'étendre de 5' en 3' et inhiber l'épissage de cette façon.

Des sites spécifiques de liaison de hnRNP A1 existent aussi dans les introns. L'ISS de l'exon 3 de la protéine Tat du virus VIH-1 chevauche le point de branchement, un site spécifique de liaison de la snRNP U2 et inhibe l'assemblage du spliceosome (Tange *et al.*, 2001).

V. Marchand a montré au laboratoire que l'élément ESE3 est ambivalent et qu'il joue un rôle clé dans la régulation du site A7. En effet, la séquence (GAA)<sub>3</sub> de l'élément ESE3 active l'épissage en fixant la protéine ASF/SF2 (Staffa *et al.*, 1995; Marchand *et al.*, 2002), tandis que les nucléotides situés en amont de cette séquence forment un élément AUA(GAA)<sub>3</sub> qui répriment l'épissage en fixant la protéine hnRNP A1.

Le pré-ARNm de la protéine hnRNP A1 elle-même est régulé par épissage alternatif. L'exon 7b est flanqué de multiples ISS qui lient les protéines hnRNP A/B et hnRNP H (Nasim *et al.*, 2002; Hutchison *et al.*, 2002; Chabot *et al.*, 1997; Blanchette *et al.*, 1999; Simard *et al.*, 2000).

### **c) Rôles des protéines hnRNP A2 et hnRNP A3**

Les hnRNP A1 et A2 humains ont 80% et 58% d'identité dans les régions RRM et riches en glycine, respectivement (Burd *et al.*, 1989; Nichols *et al.*, 2000). Malgré leur forte homologie, ces deux protéines vont avoir des rôles bien distincts au sein de la cellule. La séquence en acides aminés des RRM de hnRNP A3 a quant à elle une grande identité de séquence avec hnRNP A1, alors que son domaine RGG est plus proche de celui de hnRNP A2 (Ma *et al.*, 2002). Une des séquences de fixation de hnRNP A2 la mieux identifiée est la séquence de 21 nucléotides A2RE (hnRNP A2 response element). Elle est essentielle pour le rôle de transport cytoplasmique de hnRNP A2 de plusieurs ARNm dans les oligodendrocytes et les neurones (Munro *et al.*, 1999). hnRNP A2 peut interagir avec les séquences A2RE-1 et A2RE-2 qui sont trouvées dans une région de chevauchement entre les gènes *vpr* et *tat* du virus VIH-1 *in vitro* (Mouland *et al.*, 2001). hnRNP A2 lie aussi l'oncoprotéine SET, un régulateur-clé de la réplication de l'ADN, du remodelage de la chromatine et de la transcription des gènes (Vera *et al.*, 2006).

Des expériences *in vitro* ont montré que la protéine hnRNP A3 s'associe avec A2RE et AuRE (AU-rich elements) dans le 3'UTR de l'ARNm de COX-2 (Cok *et al.*, 2004). hnRNP A3 s'associe avec des répétitions télomériques simple brin d(TTAGGG), les protège des attaques des nucléases et inhibe à la fois la réaction télomérique et la synthèse de l'ADN *in vitro* (Tanaka *et al.*, 2007; Huang *et al.*, 2008). hnRNP A3 agit aussi comme facteur de transcription, se liant à la région régulatrice du gène Hoxc8 (Makeyev *et al.*, 2005).

### I.2.2.2 Rôles des protéines hnRNP H/H' et F

Cette famille de protéines regroupe les protéines hnRNP H (ou H1), H' (ou H2), 2H9 (ou H3), et F. Ces différentes protéines sont codées par 4 gènes différents (Honore *et al.*, 1995). La protéine 2H9 est produite sous la forme d'au moins 6 isoformes distinctes. Ces isoformes sont le résultat d'un épissage alternatif (Honore *et al.*, 1995). Les protéines hnRNP H et H' présentent 96 % d'identité, alors que la protéine hnRNP F présente, respectivement, 72 et 70 % d'identité avec les protéines hnRNP H et H' respectivement (Honore *et al.*, 1995).

Ces protéines possèdent deux domaines RRM (pour 2H9) ou trois (pour H, H' et F) non canoniques et deux domaines riches en glycines. Leurs domaines RRM sont également appelés quasi-RRM (qRRM) du fait de l'absence des motifs RNP consensus (Honore *et al.*, 1995). Les protéines hnRNP H/H'/F et 2H9 se fixent sur des séquences riches en G, avec une préférence pour les séquences GGGA (Honore *et al.*, 1995; Caputi *et al.*, 2001). Récemment, l'équipe de F.H. Allain a déterminé la structure des 3 RRM de hnRNP F libres puis liés à une séquence régulatrice de l'épissage du pré-ARNm *bcl-x* (CGGGAUGGGGUA). Bien que les 2 séquences consensus RNP-1 et RNP-2 ne soient pas conservées dans hnRNP F, les trois domaines RRM adoptent la structure consensus classique.

### *Rôles dans la régulation de l'épissage*

Les protéines hnRNP H/H'/F/2H9 peuvent aussi bien être impliquées dans l'activation que dans l'inhibition de l'épissage. Les protéines H/H'/F sont, en plus, impliquées dans la régulation de la polyadénylation des ARNm tandis que le rôle de hnRNP 2H9 dans ce processus n'est pas élucidé. La protéine 2H9 est impliquée dans l'épissage, notamment dans l'arrêt précoce de l'épissage induit par un choc thermique (Bagga *et al.*, 1998).

Les protéines hnRNP F/H modulent l'épissage alternatif du pré-ARNm du gène Bcl-x, un membre de la grande famille des gènes apoptotiques Bcl-2. Dans le pré-ARNm *Bcl-x*, 2 sites 5' d'épissage XS et XL sont en compétition, ce qui conduit à l'expression de 2 isoformes, la

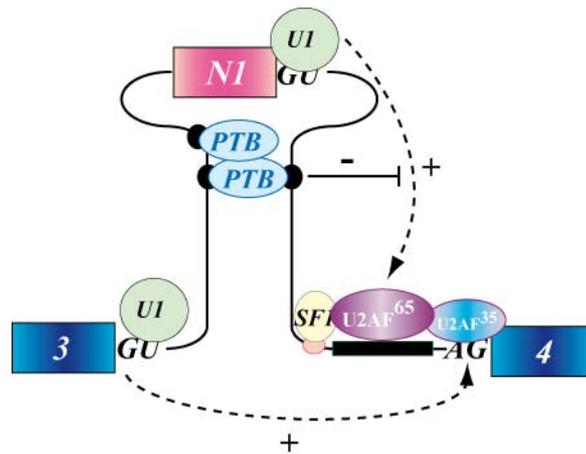
protéine anti-apoptotique Bcl-xL, et la protéine pro-apoptotique Bcl-xS (Garneau *et al.*, 2005). Le mécanisme d'action des protéines hnRNP H et F une fois fixées, n'est pas connu mais elles pourraient faciliter le recrutement de la snRNP U1 au niveau du site donneur d'épissage.

Dans les cellules non musculaires, la protéine hnRNP H empêche l'inclusion de l'exon 7 de l'ARNm de la  $\beta$ -tropomyosine de rat en se fixant à une séquence ESS11 située au voisinage du site 3' dans l'exon (Chen *et al.*, 1999). Un autre cas de régulation similaire a été décrit au laboratoire pour l'ARN du virus VIH-1: l'utilisation du site accepteur A3 est inhibée par la fixation de la protéine hnRNP H à une séquence exonique pESS2 située dans l'exon en aval au voisinage de ce site (Jacquet *et al.*, 2001). Dans ces 2 cas, la protéine pourrait agir en empêchant, par encombrement stérique, la fixation de U2AF35 sur la bordure intron-exon (Jacquet *et al.*, 2001).

### **I.2.2.3 La protéine hnRNP I (PTB)**

La protéine hnRNP I est une des isoformes codées par le gène PTB (Polypyrimidine Tract Binding protein) (Ghetti *et al.*, 1992). Dans les cellules de mammifères, il existe au moins 2 paralogues du gène *PTB*: *ROD1* exprimé principalement dans les cellules hématopoïétiques (Yamamoto *et al.*, 1999) et *nPTB* exprimé spécifiquement dans les lignées neuronales et les cellules du cerveau (Markovtsov *et al.*, 2000). Les protéines PTB contiennent 4 domaines RRM non canoniques précédés d'un élément NLS en N-terminal (position 1-50). La structure 3D de chaque domaine suit l'organisation générale d'un domaine RRM, cependant les résidus aromatiques conservés des motifs RNP-1 et RNP-2 ne sont pas présents (Ghetti *et al.*, 1992). Chacun des ses quatre domaines RRM va fixer une séquence nucléotidique CUCUCU (Oberstrass *et al.*, 2005). Cette protéine de 58kDa se liant à l'ARN est impliquée dans différents aspects du métabolisme de l'ARNm, dont la répression de l'épissage (pour revue, (Sawicka *et al.*, 2008).

Un exemple d'intervention de PTB dans la régulation de l'épissage est celui de la régulation de l'utilisation de l'exon N1 du gène c-src. (Figure I-25).



**Figure I-25: Modèle d'inhibition de l'épissage par la protéine PTB (pour revue, (Spellman *et al.*, 2006)).**

Dans les cellules non-neuronales, la protéine PTB inhibe la définition de l'intron reliant l'exon N1 à l'exon 4 du pré-ARNm *c-src*. La fixation coopérative de la protéine PTB au niveau des éléments inhibiteurs de part et d'autre de l'exon N1 n'empêche pas la fixation de la snRNP U1 au site 5' de l'exon N1 mais l'empêche de créer des interactions avec U2AF fixé à la séquence polypyrimidine en amont de l'exon 4 constitutif ce qui induit une exclusion de l'exon N1.

L'exon N1 est inclus dans le pré-ARNm dans les neurones et exclu dans les autres types cellulaires (Levy *et al.*, 1987). La protéine PTB se fixe aux éléments riches en séquences CU situés autour de l'exon N1 et forme un complexe de répression.

Dans les neurones, nPTB, plus abondante que PTB, permet l'assemblage d'un complexe renfermant, entre autres, la protéine hnRNP H, ce qui permet à l'inverse l'inclusion de l'exon N1 (Min *et al.*, 1995; Min *et al.*, 1997; Chou *et al.*, 1999; Markovtsov *et al.*, 2000).

hnRNP A1 et PTB sont deux répresseurs de l'épissage mais il a été montré pour l'épissage de l'exon N1 de *c-src* que leur mode d'inhibition se fait selon deux mécanismes différents (Rooke *et al.*, 2003).

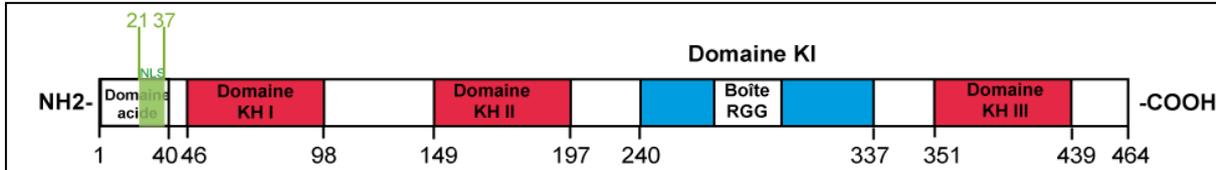
#### I.2.2.4 La protéine hnRNP K

La protéine hnRNP K est un facteur abondant trouvé dans le noyau (Swanson *et al.*, 1988), le cytoplasme (Ostrowski *et al.*, 1994) et les mitochondries (Ostrowski *et al.*, 2002). hnRNP K n'est pas exprimée dans tous les tissus, ni avec la même intensité dans tous les types cellulaires (Blanchette *et al.*, 2006).

##### Les domaines de hnRNP K

La protéine hnRNP K comprend 3 domaines KH liant les acides nucléiques et structurellement similaires et le domaine interactif KI qui contient 5 boîtes RGG (Bomsztyk *et al.*, 2004; Bomsztyk *et al.*, 1997) (Figure I-26). Le domaine KI permet l'interaction de

hnRNP K avec un grand nombre de partenaires protéiques, dont PKC $\delta$ , les kinases de la famille Src, Vav, Zik1, Eed et YB-1. Par exemple, la région KI contient les sites d'amarrage riches en proline, comme RXXPXXP et PXXPXR, qui interagissent avec les domaines SH3 et notamment avec les SH3 des kinases de la famille des Src (Bomsztyk *et al.*, 2004).



**Figure I-26: Structure modulaire de la protéine humaine hnRNP K.**

Le domaine NLS est représenté en vert, les domaines KH en rouge et le domaine KI contenant la boîte RGG en bleu.

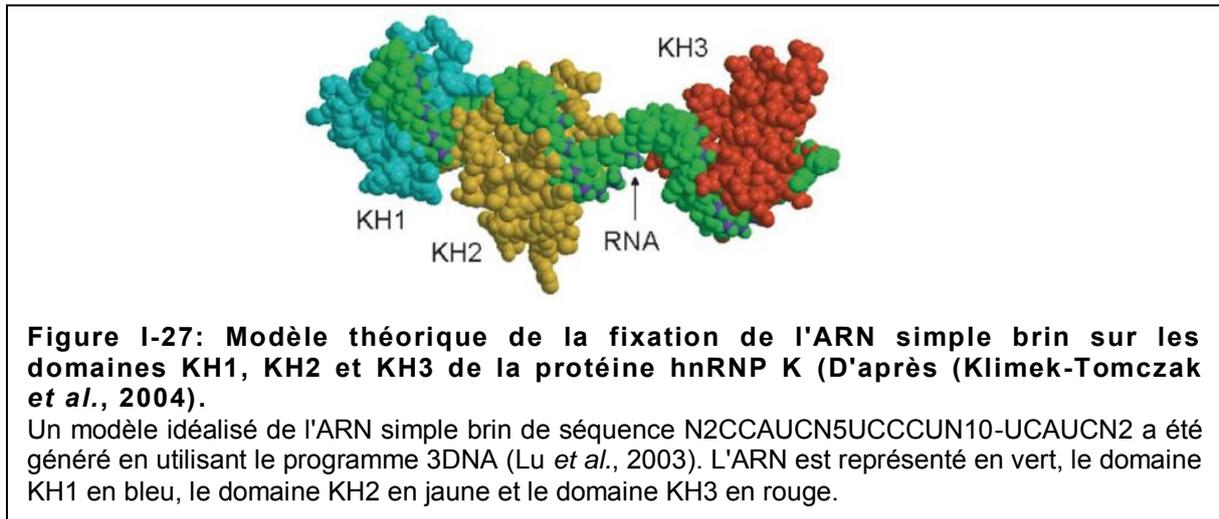
hnRNP K appartient à un groupe de protéines liant les ARN, caractérisé par une liaison préférentielle aux sites riches en C. La fixation aux poly (rC) de hnRNP K se fait par l'intermédiaire du troisième domaine KH (KH3). Le domaine KH1 lie plus particulièrement les cCAUc via ses résidus GKGG et le domaine KH2 lie les wCCCw (w : adénosine ou uracile) via les résidus GVKG (Klimek-Tomczak *et al.*, 2004) (Tableau 1). Les domaines KH pris individuellement montrent de faibles niveaux de fixation à l'ARN.

**Table 1 : Score de fixation de la protéine hnRNP K à la matrice nucléotidique, d'après (Klimek-Tomczak *et al.*, 2004).**

	Premier patch: cCAUc					$n(2-7)$	Deuxième patch: wCCCw					$n(7-18)$	Troisième patch: UCAYc				
	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
A	0	-3	5	-3	0	0.5	5	-3	-5	-3	5	0.5	-3	-5	5	-1	0
C	5	5	-1	-3	5	0.5	3	5	5	5	3	0.5	-1	5	-1	5	3
G	-5	-7	-7	-7	-7	-2	-3	-7	-10	-7	-3	-2	-7	-7	-7	-7	-5
U	0	-3	-3	5	0	0.5	3	-3	-5	-3	3	0.5	5	-5	-3	5	0

Chaque patch correspond à la séquence nucléotidique ARN susceptible de se fixer aux différents domaines KH. Le premier patch est prédit pour fixer préférentiellement KH1, le deuxième patch pour fixer KH2 et le troisième patch pour fixer KH3.  
W: adénosine ou uracile  
Y: pyrimidine, cytosine ou uracile

L'activité de fixation de la protéine hnRNP K entière est plus grande que la somme des activités individuelles générées par la fixation des domaines de liaisons, indiquant une synergie entre les 3 domaines KH quand ils se fixent à leur cible ARN (Figure I-27).



Les deux tiers N-terminaux de hnRNP K, dont KH1 et KH2, et la jonction entre KH2 et KH3, sont requis pour les interactions avec hnRNP E2, I, K et L. Grâce à la méthode triple hybride les cibles de hnRNP K ont été identifiées (Tableau 2) (Paziewska *et al.*, 2004).



Les séquences identifiées en triple hybride sont presque identiques aux séquences consensus liant hnRNP K identifiées en SAGE (Serial Analysis of Gene Expression) (Tableau 3) : CCAUCN2–7(A/U)CCC(A/U)N7–18UCA(C/U)C (Ostrowski *et al.*, 2002) (Klimek-Tomczak *et al.*, 2004).

**Table 3: Liste des étiquettes SAGE correspondant aux ARNm dans le noyau de cellules d'hépatomes de rat qui interagissent avec la protéine hnRNP K, d'après (Klimek-Tomczak *et al.*, 2004).**

Etiquette	N° accession du gène	Nom du gène	Fonction	Etiquettes comptées
GAAGCAGGAC	BC059143.1	Cofiline 1	Cytosquelette, liaison à l'actine	4
GGATGGGACA	BC059143.1	Cofiline 1 (antisens)	Cytosquelette, liaison à l'actine	4
GCTTGCTTCC	BC072470.1	RAB7	Liaison à l'ATP, transporteur protéique	8
ATGCAGGGCA	BC072470.1	RAB7 (antisens)	Liaison à l'ATP, transporteur protéique	13
TCTTCTCACA	XM_342342.1	Protéine ribosomale L39	Traduction, biogénèse et structure du ribosome	12
GCGAGGATCT	BC058489.1	Protéine ribosomale L39 (antisens)	Traduction, biogénèse et structure du ribosome	19
AAAACAGTGG	XM_343587.1	Protéine ribosomale L37a	Signal, transmembrane	5
AAACATTGGG	XM_220013.2	Membre 3 des protéines de la superfamille des transmembranes 9	Transporteur de la membrane interne de la mitochondrie	5
AGGTTGGCTT	NM_019354.1	UCP2	Fonction inconnue	28
CCGAAGTTGA	AA892143.1	RN.92843	Traduction, biogénèse et structure du ribosome	4
GAAATGTAAG	–	Gène inconnu	Fonction inconnue	5
GCCGCGGCC	BC061561.1	Protéine ribosomale S26	Traduction, biogénèse et structure du ribosome	6
GGCGACAGTC	–	Gène inconnu	Fonction inconnue	9
TCTTCAGAGT	NM_022333	TIAR nucleolysine	Lysosome	10
TGCGCCAAGT	XM_342342.1	Protéine ribosomale L34	Traduction, biogénèse et structure du ribosome	6
TTGAACACTT	XM_340841.1	Protéine d'interaction TAX 1	Cascade de signalisation intracellulaire, liaison de protéine	6

Quatre des cibles sont identiques ou presque à la séquence consensus identifiée par SELEX (Systematic Evolution of Ligands by eXponential enrichment) : UC3–4(A/U) (U/A) (Thisted *et al.*, 2001).

La protéine hnRNP K possède un domaine KNS (hnRNP K nuclear shuttling domain) qui partage de nombreuses caractéristiques avec le domaine M9 de hnRNP A1 et qui lui confère la possibilité d'un transport bidirectionnel à travers l'enveloppe nucléaire. Cet import nucléaire par l'intermédiaire de KNS peut être outrepassé par un NLS classique. Ce type de signal nucléaire de hnRNP K définit une troisième voie d'import protéique (Michael *et al.*, 1997).

#### Les rôles de hnRNP K au sein de la cellule

hnRNP K interagit avec une grande diversité de molécules impliquées dans l'expression des gènes et la transduction du signal (Bomszyk *et al.*, 2004; Makeyev *et al.*, 2002). Elle interagit aussi avec des ARN et des ADN (Ostrowski *et al.*, 1994) impliqués dans le remodelage de la

chromatine (de Alava *et al.*, 2000; Denisenko *et al.*, 1997; Shnyreva *et al.*, 2000), la transcription (Du *et al.*, 1998), l'épissage des ARN (Shnyreva *et al.*, 2000) et la traduction (Ostareck *et al.*, 2001; Evans *et al.*, 2003). De plus, des études ont montré que hnRNP K augmente l'activité transcriptionnelle des oncogènes c-myc et c-src (Takimoto *et al.*, 1993; Ostareck-Lederer *et al.*, 2002).

La transmission d'un signal soutenu extra-cellulaire régulé par la kinase ERK joue un rôle critique dans la production d'interleukine 2 (IL-2) des lymphocytes T. Cette étude montre que hnRNP K est l'une des cibles ERK essentielles pour la production d'IL-2 (Chang *et al.*, 2009). hnRNP K avait été précédemment décrit comme un substrat direct de ERK et qui forme des complexes avec de multiples protéines de signalisation ainsi qu'avec des ADN et ARN. Les résultats montrent une augmentation de hnRNP K dépendante de ERK après la stimulation des récepteurs des lymphocytes T. Un siRNA contre hnRNP K abroge la production d'IL-2 par les lymphocytes T. De plus, une diminution de l'expression de hnRNP K cause une augmentation de la protéolyse de Vav1, une cible de fixation de la hnRNP K. Comme Vav1 est une cible essentielle de l'activation des cellules T, les données suggèrent que la signalisation par ERK est requise pour l'activation des cellules en inhibant partiellement la protéolyse de Vav1 induite par l'activation.

hnRNP K s'associe avec la nucléoline pour jouer un rôle fonctionnel dans la maintenance de la stabilité des ARNm. La liaison de hnRNP K et PCBP1 (Poly(rC) Binding Protein 1) aux régions riches en C du 3' UTR de l'ARNm de la gastrine facilite l'association de la nucléoline à cet ARNm (Lee *et al.*, 2007).

La protéine hnRNP K a plusieurs rôles décrits lors d'infection virales. Lors de l'infection avec le virus Sindbis (*Alphavirus*, famille des *Togaviridae*; virion enveloppé, icosaédrique, ARN monocaténaire, de polarité positive) de cellules de fibroblastes de reins de hamster, la protéine hnRNP K est relocalisée du noyau vers le cytoplasme (Burnham *et al.*, 2007). De même, la protéine hnRNP K se lie à l'ADN du virus de l'hépatite B HBV (virus à ADN appartenant à la famille des *Hepadnaviridae*) et dans les études de transfections cellulaires, la surexpression de hnRNP K augmente la réplication du HBV (Ng *et al.*, 2005).

### La protéine hnRNP K et l'épissage

Plusieurs études ont décrit le rôle de la protéine hnRNP K dans l'épissage alternatif. Des analyses fonctionnelles ont montré que la hnRNP K est un composant du complexe activateur

qui stimule l'épissage de l'exon 6A du pré-ARNm de la  $\beta$ -tropomyosine à travers la séquence S4 sauvage (Expert-Bezancon *et al.*, 2002). La protéine hnRNP K pourrait aussi être impliquée dans la répression de l'épissage de l'exon 6B qui est réprimé par les protéines PTB et hnRNP A1 et activé par les protéines ASF/SF2 et SC35 (Sauliere *et al.*, 2006). Enfin, la protéine Nova peut être en compétition ou en synergie avec d'autres facteurs comme hnRNP K ou brPTB/nPTB sur un cas de contrôle de l'épissage alternatif *in vivo* (Ule *et al.*, 2006).

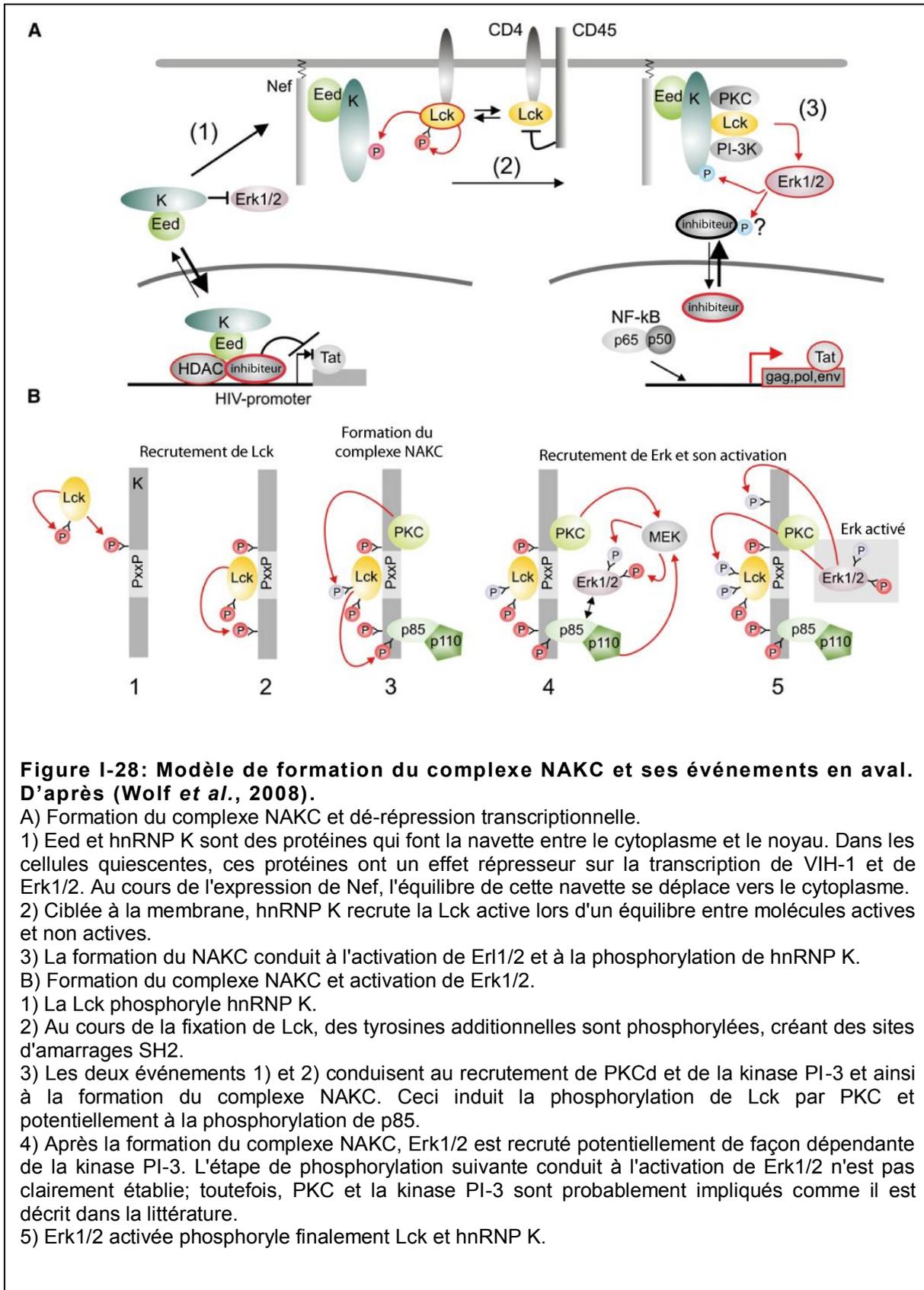
Pour évaluer la gamme d'action des principales hnRNP sur la production d'isoformes d'épissage d'un sous-ensemble de gènes apoptotiques, une approche d'interférence de l'ARN a été menée. Une plateforme de criblage de RT-PCR en haut débit a permis d'analyser l'effet de la diminution cellulaire de 14 hnRNP sur 56 événements d'épissage alternatif (ASE). Ce travail s'est fait sur des gènes apoptotiques exprimés dans des lignées humaines de cellules HeLa, PC-3 et BJT (Venables *et al.*, 2008). 50% des ASE sont affectés quand hnRNP K est déplété. hnRNP K provoque alors un ensemble d'événements de saut d'exons et d'inclusion d'exons. La moitié des unités réactives à la déplétion de hnRNP K contiennent des motifs de liaison optimale pour la protéine : AUC3/4(A/U)<sub>2</sub> (Thisted *et al.*, 2001).

### La protéine hnRNP K et le virus VIH-1

Sam68 est une protéine cible de la tyrosine kinase c-src. Sa surexpression remplace fonctionnellement la protéine Rev du VIH-1 dans l'expression des gènes régulés par l'intermédiaire de la séquence RRE (Rev-Response Element) et dans la réplication virale (Yang *et al.*, 2002). hnRNP K est une protéine qui interagit spécifiquement avec Sam68 *in vitro* et *in vivo*. Elle ne lie pas l'ARN RRE directement, mais forme un super complexe avec Sam68 et RRE *in vitro*. hnRNP K inhibe significativement l'expression de gènes dépendant de RRE mais pas par l'intermédiaire de la protéine virale Rev de VIH-1. De plus, Sam68 inhibe l'activation transcriptionnelle d'un élément CT par hnRNP K. L'élément CT (CCCTCCCCA) est la séquence du promoteur du gène humain c-myc qui fixe la protéine hnRNP K, ce qui va activer la transcription du gène (Michelotti *et al.*, 1996; Tomonaga *et al.*, 1995). L'interaction directe entre Sam68 et hnRNP K affecte l'activité de ces deux protéines dans la voie de transduction du signal des événements à la fois transcriptionnels et post-transcriptionnels (Yang *et al.*, 2002).

hnRNP K possède un autre rôle dans la régulation de VIH-1. En effet, la protéine DDX1 est une hélicase qui va se lier à la protéine Rev du virus VIH-1 (Fang et al., 2004) ainsi qu'à la protéine hnRNP K (Chen et al., 2002). La région de DDX1 où hnRNP K se lie chevauche la région interagissant avec Rev, ce qui affecte la libération de Rev du RRE dans le cytoplasme (Fang et al., 2004).

La protéine hnRNP K joue un rôle central dans un processus dans lequel la protéine Nef du virus VIH-1 coopère fonctionnellement avec la protéine transactivatrice Tat en dé-réprimant le promoteur viral via un complexe kinase NAKC (Nef associated kinase complexe) (Wolf *et al.*, 2008) (Figure I-28). Le complexe NAKC groupe des kinases interagissant avec Nef, dont Lck (une tyrosine kinase associée avec CD4 et CD8), PKC $\delta$  et la kinase PI-3. hnRNP K va être dans ce processus la plateforme d'amarrage de plusieurs kinases dans les cellules T. La formation du complexe NAKC autour de hnRNP K va mener à l'activation de Erk1/2 qui va ensuite phosphoryler Lck. Le complexe NAKC/Erk 1/2 ainsi activé va cibler des facteurs inhibiteurs encore inconnus qui occupent le site de reconnaissance de NF-kB sur le promoteur du virus VIH-1 ce qui inhibe la transactivation par Tat. Cette voie de signalisation conduit à la dé-répression du promoteur de VIH-1, et va permettre à des taux sous-optimaux de Tat et à des facteurs de transcription comme NF-kB d'initier la transcription virale (Wolf *et al.*, 2008).



**Figure I-28: Modèle de formation du complexe NAKC et ses événements en aval. D'après (Wolf et al., 2008).**

A) Formation du complexe NAKC et dé-répression transcriptionnelle.

1) Eed et hnRNP K sont des protéines qui font la navette entre le cytoplasme et le noyau. Dans les cellules quiescentes, ces protéines ont un effet répressif sur la transcription de VIH-1 et de Erk1/2. Au cours de l'expression de Nef, l'équilibre de cette navette se déplace vers le cytoplasme.

2) Ciblée à la membrane, hnRNP K recrute la Lck active lors d'un équilibre entre molécules actives et non actives.

3) La formation du NAKC conduit à l'activation de Erk1/2 et à la phosphorylation de hnRNP K.

B) Formation du complexe NAKC et activation de Erk1/2.

1) La Lck phosphoryle hnRNP K.

2) Au cours de la fixation de Lck, des tyrosines additionnelles sont phosphorylées, créant des sites d'amarrages SH2.

3) Les deux événements 1) et 2) conduisent au recrutement de PKC et de la kinase PI-3 et ainsi à la formation du complexe NAKC. Ceci induit la phosphorylation de Lck par PKC et potentiellement à la phosphorylation de p85.

4) Après la formation du complexe NAKC, Erk1/2 est recruté potentiellement de façon dépendante de la kinase PI-3. L'étape de phosphorylation suivante conduit à l'activation de Erk1/2 n'est pas clairement établie; toutefois, PKC et la kinase PI-3 sont probablement impliqués comme il est décrit dans la littérature.

5) Erk1/2 activée phosphoryle finalement Lck et hnRNP K.

### I.2.2.5 Rôle des protéines hnRNP G, L, M, Q, R et C.

#### **hnRNP G**

La protéine hnRNP G (RBM-X) a été impliquée dans le contrôle de l'épissage de plusieurs pré-ARNm incluant SMN (Sawicka *et al.*, 2008) et la  $\beta$ -tropomyosine (Nasim *et al.*, 2003).

#### **hnRNP L**

La protéine humaine hnRNP L stimule l'inclusion de l'exon dans le pré-ARNm eNOS et l'exclusion de l'exon variable dans CD45 (Hung *et al.*, 2008).

#### **hnRNP M**

La surexpression de hnRNP M peut stimuler à la fois l'inclusion de l'exon et les événements de saut d'exon dans quelques exons alternatifs des cellules 293T (Hovhannisyan *et al.*, 2007).

#### **hnRNP Q**

La protéine hnRNP Q (Syncrip) est requise pour un épissage efficace *in vitro* du pré-messager ARN (Chen *et al.*, 2008).

#### **hnRNP C**

Peu de données sont disponibles sur le rôle de hnRNP C dans l'épissage. Une étude récente montre cependant son implication dans l'épissage du gène SMN1 (Irimura *et al.*, 2009).

### II.2.3 La régulation de l'épissage alternatif par les protéines SR

Différents modèles ont été proposés pour expliquer le mécanisme par lequel les protéines SR régulent l'inclusion des exons. Le premier modèle de recrutement proposé se base sur la capacité des protéines SR liées aux ESE à recruter et à stabiliser les interactions entre la particule snRNP U1 au site 5' d'épissage et la U2AF<sub>65</sub> sur une séquence polypyrimidine faible du site 3' (Graveley *et al.*, 2001) dans un processus connu comme la définition des exons (Robberson *et al.*, 1990). Dans le deuxième modèle, les domaines RS des protéines SR liés aux activateurs interagissent directement avec l'ARN au point de branchement pour promouvoir l'assemblage du pré-spliceosome (Shen *et al.*, 2004; Shen *et al.*, 2004). Le troisième modèle fait appel à la compétition des protéines SR avec la famille des hnRNP où les protéines SR peuvent agir en antagonistes de l'activité négative sur l'épissage des protéines hnRNP reconnaissant les ESS.

### II.2.3.1 Activation de l'épissage par les protéines SR fixées à des séquences exoniques.

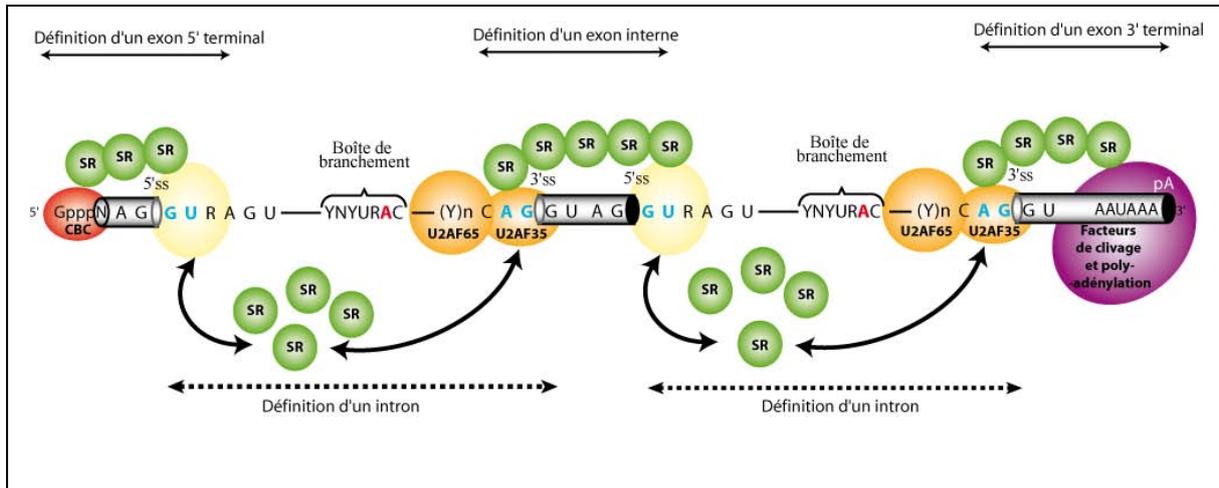
#### Le modèle d'activation de l'épissage par la fixation des protéines SR aux ESE.

Les protéines SR ont des rôles divers dans l'épissage alternatif. Un de ces rôles est la reconnaissance des activateurs exoniques d'épissage (ESE). Les protéines SR agissent de façon substrat-spécifique en liant les ESE, et reconnaissent des séquences dégénérées consensus de 5 à 7 nucléotides. Cette dégénérescence des ESE autorise probablement des liaisons chevauchantes entre les différentes protéines SR. Ces motifs ont évolué avec une extrême diversité de séquences et sont reconnus par un petit nombre de protéines SR qui lient l'ARN avec une spécificité de séquence limitée.

Les séquences consensus des sites d'épissages ne sont généralement pas suffisantes pour l'assemblage d'un spliceosome fonctionnel à elles seules et ont besoin des éléments auxiliaires ESE pour l'épissage constitutif et alternatif. La liaison des protéines SR aux ESE permet la reconnaissance par la machinerie d'épissage et ainsi le couplage du site 5' au site 3' correspondant en évitant le saut d'exon (Ibrahim *et al.*, 2005).

#### Régulation de l'épissage grâce à l'interaction protéine-protéine des protéines SR.

Les protéines SR peuvent former un réseau d'interaction entre protéines le long des introns pour rapprocher de façon précoce les sites 5' et 3' d'épissage dans l'assemblage des spliceosomes (Figure I-29). Par exemple, les protéines ASF/SF2 et SC35 interagissent avec U1-70 kDa au site 5' d'épissage et avec U2AF35 au site 3' lors d'une liaison dépendant du domaine RS (Wu *et al.*, 1993). Les protéines SR peuvent aussi faciliter le recrutement des tri.snRNP U4/U6.U5 au pré-spliceosome (Roscigno *et al.*, 1995) via des interactions régulées par les domaines RS avec les protéines SR-related SRrp65 et SRrp110 (Makarova *et al.*, 2001).



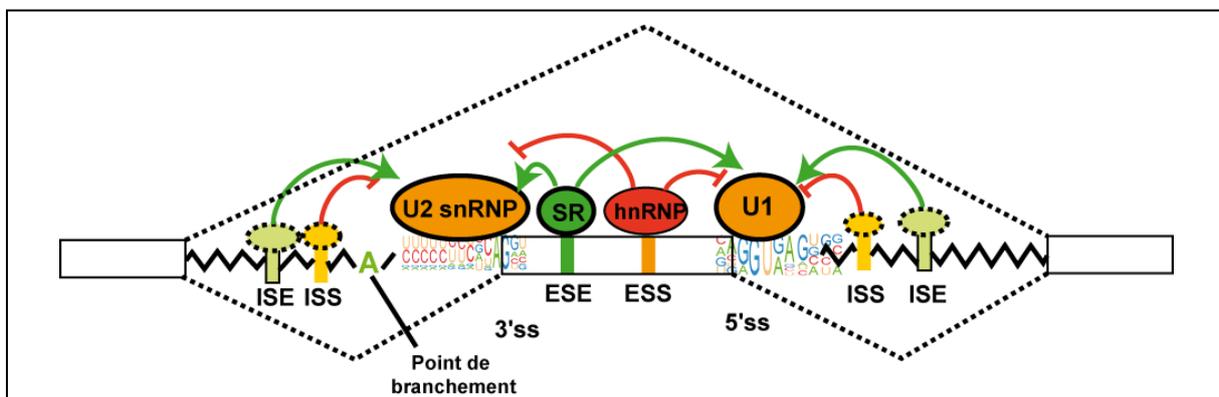
**Figure I-29: La définition des introns et des exons d'un ARN pré-messager au cours de l'assemblage du spliceosome (adapté de (Kohtz *et al.*, 1994).**

Les modes de définition des exons ou des introns sont représentés de manière schématique. 5'ss et 3'ss désignent respectivement les sites donneurs et accepteurs d'épissage. Sur ce schéma, le nombre de protéines SR nécessaires pour créer un lien entre les composants liés aux sites 5' et les composants liés au site 3' est arbitraire.

Antagonisme entre les protéines SR et les protéines hnRNP lors de l'épissage.

Des effets combinés des séquences *cis* et des facteurs *trans* modulent l'épissage des exons (Smith *et al.*, 2000). L'activité des protéines SR sur l'épissage est antagoniste à l'activité des protéines hnRNP A/B pour la sélection du site d'épissage (Figure I-30).

Le ratio hnRNP A1/protéine SR dans le noyau a une grande importance dans la régulation de l'épissage alternatif et peut avoir un rôle crucial dans la régulation de l'épissage qu'il soit spécifique du tissu ou lors du contrôle développemental.



**Figure I-30: Représentation schématique de la régulation de l'épissage alternatif (d'après (Wang *et al.*, 2008).**

Les exons sont représentés par les rectangles, les introns par des traits en dents de scie. Les motifs consensus des sites d'épissage sont représentés par des pictogrammes et l'adénosine du point de branchement est indiquée en vert. Deux voies d'épissage alternatif sont représentées en pointillé, avec l'exon central inclus ou exclu. L'épissage est régulé par des éléments *cis* (ESE, ESS, ISS et ISE) et des facteurs d'épissage *trans*-activateurs (protéines SR, hnRNP et facteurs inconnus).

Les protéines ASF/SF2 et hnRNP A1 ont été décrites comme ayant un rôle antagoniste dans la régulation de l'exon N1 spécifique des neurones du gène *c-src* (Rooke *et al.*, 2003). L'antagonisme entre les protéines SR et les protéines hnRNP est au sein d'un modèle de régulation hautement complexe d'exons mutuellement exclusifs chez le gène *Dscam* de la drosophile (Down's syndrome cell adhesion molecule) (Olson *et al.*, 2007).

### II.2.3.2 Inhibition de l'épissage alternatif par les protéines SR

Il est intéressant de noter que les protéines SR peuvent quelquefois avoir des effets antagonistes sur la sélection du site d'épissage, comme c'est le cas pour SRp20 et ASF/SF2 dans la régulation du pré-ARNm de SRp20 (Jumaa *et al.*, 1997), de ASF/SF2 et de SC35 dans la régulation de la  $\beta$ -tropomyosine (Gallego *et al.*, 1997) et dans l'épissage alternatif du pré-ARNm de l'hormone de croissance humaine (Solis *et al.*, 2008). Le rôle de ASF/SF2 dans l'épissage du pré-ARNm dépend du contexte de la séquence du pré-ARNm à laquelle la protéine s'est liée, comme le montre l'action inhibitrice de ASF/SF2 sur l'épissage du pré-ARNm de l'adénovirus IIIa. ASF/SF2 s'est lié à un élément intronique répresseur 3RE qui contient trois séquences identiques riches en purines et cette fixation conduit à l'inhibition du recrutement de la snRNP U2 au site 3' d'épissage du pré-ARNm IIIa (Kanopka *et al.*, 1996).

D'autres protéines SR, dites SR-like, dont p54, SRp38 et SRp86 fonctionnent seulement comme régulateurs négatifs de l'épissage alternatif, en étant antagonistes des protéines SR « classiques » et stimulant le saut d'exon (Zhang *et al.*, 1996; Cowper *et al.*, 2001; Barnard *et al.*, 2002).

### II.2.4 Les co-activateurs SRm300 et 160 et la protéine Tra-2

#### Les co-activateurs SRm300 et SRm160

Les protéines SRm160 (SR nuclear matrix, protéine de 160 kDa) et SRm300 (antigène de 300 kDa de la matrice nucléaire) sont des co-activateurs de l'épissage (Eldridge *et al.*, 1999). Ces deux protéines contiennent toutes deux des domaines RS mais n'ont pas de domaine RRM et ne se lient pas normalement au pré-ARNm en absence d'autres facteurs d'épissage. Elles ne sont pas capables de compléter un extrait cytoplasmique S100 pour l'activité d'épissage. La déplétion de SRm160/300 mais pas de SRm300 empêche l'épissage constitutif d'un pré-ARNm ainsi que de l'épissage alternatif dépendant de l'ESE d'un pré-ARNm *dsx* (Blencowe, 2000).

### La protéine Tra 2

Les protéines Tra et Tra2 ont été initialement identifiées chez la drosophile. Elles sont produites seulement chez les femelles, pas chez les mâles. Ce sont des protéines impliquées dans la détermination du sexe chez la drosophile au travers de leur activité de régulation de l'épissage. Les protéines Tra sont des facteurs d'épissage de la famille des protéines SR qui agissent directement en se fixant sur les transcrits cibles. Tra possède un domaine RS tandis que Tra2 en possède deux et possède aussi un domaine RRM. Tra et Tra2 coopèrent chez les femelles pour modifier l'épissage du gène *dsx*. Chez le mâle en absence de la protéine Tra les ARN de *dsx* sont épissés d'une manière différente.

Chez les mammifères, 2 protéines homologues à la protéine Tra2, hTra2 $\alpha$  et hTra2 $\beta$ , ont été identifiées. Comme les protéines Tra et Tra2, elles sont capables de réguler l'épissage en interagissant avec les protéines SR. C'est le cas de l'exon 7 du pré-ARNm *SMN2*, dont l'inclusion est activée par une séquence riche en purines de type GAA sur laquelle hTra2 $\beta$  se fixe et recrute la protéine SRp30c (Young *et al.*, 2002).

### **II.3 Autres protéines et paramètres intervenant dans la régulation de l'épissage.**

#### **II.3.1 Implication d'autres protéines régulatrices de l'épissage.**

##### **-YB-1**

La protéine multifonctionnelle YB-1 de mammifère comprend trois domaines très conservés : un domaine N-terminal riche en Alanine/Proline, un domaine sensible aux chocs thermiques froids CSD (Cold-Shock Domain) qui permet l'interaction avec l'ARN et l'ADN et un domaine C-terminal riche en Arginine/Lysine impliqué dans les interactions protéine/protéine et ARN/protéine.

YB-1 est aussi connue pour son rôle dans la régulation de la transcription et comme composant de complexes ribonucléoprotéiques se liant aux ARN avant la traduction (Wolffe *et al.*, 1996). Le domaine N-terminal interactif (NIII) de YB-1 est responsable de la fixation de la protéine hnRNP K. Comme ces deux protéines sont impliquées dans différentes étapes de l'expression des gènes comme la transcription, la traduction et le transport nucléaire, il est

possible que ces mécanismes soient régulées par une interaction entre YB-1 et hnRNP K (Shnyreva *et al.*, 2000).

### **-les protéines CELF : le gène cTnT**

Les protéines CELF (CUG-BP and ETR-3 Like Factors) constituent une famille de protéines relativement conservées (Lu *et al.*, 1999); (Ladd *et al.*, 2001) initialement nommées BRUNO-like pour leur homologie avec le régulateur de la traduction BRUNO décrit chez la drosophile (Good *et al.*, 2000). Les membres de la famille CELF jouent un rôle dans la régulation post-transcriptionnelle des ARN et sont des facteurs régulateurs de l'épissage alternatif (Ladd *et al.*, 2001; Ladd *et al.*, 2004; Philips *et al.*, 2000).

La famille CELF comprend 6 membres ayant des homologies de séquences entre eux et qui possèdent tous un équivalent chez la drosophile (Ladd *et al.*, 2001; Ladd *et al.*, 2004): CUGBP1 ou CUG-binding protein 1, aussi appelée BRUNO-like 2; ETR-3 appelée aussi BRUNO-like 3, CUGBP2 ou NAPOR; CELF3 ou BRUNO-like 1; CELF4 ou BRUNO-like 4; CELF5 ou BRUNO-like 5; CELF6 ou BRUNO-like 6.

Les protéines CELF sont caractérisées par la présence de trois domaines de liaison à l'ARN de type RRM avec deux domaines situés en N-terminal suivis par un domaine de liaison qui les sépare du troisième RRM situé en C-terminal. Ce domaine de liaison entre les RRM 2 et 3 est variable (Good *et al.*, 2000; Ladd *et al.*, 2001), comprend de 160 à 230 acides aminés et apparaît comme unique chez les protéines CELF. Sa fonction exacte est inconnue mais il est suggéré qu'il soit impliqué dans l'épissage par des interactions protéine/protéine (Singh *et al.*, 2004).

### **-fox**

Les protéines de la famille de protéines se liant à l'ARN Fox-1 ont été conservées lors de l'évolution. La famille Fox-1 est caractérisée par son domaine conservé de liaison à l'ARN de type RRM. Il existe trois membres de cette famille chez les mammifères : Fox-1, Fox-2 et Fox-3. Ce sont des régulateurs de l'épissage alternatif chez les métazoaires, spécifiques d'un tissu. Les protéines Fox-1 reconnaissent spécifiquement les séquences (U)GCAUG dans les exons régulés par l'épissage ou dans leurs introns flanquants, et stimule ou réprime l'inclusion des exons cibles. L'identification globale des cibles ARN de Fox-2 dans les cellules vivantes

révèle que la plupart des gènes cibles de Fox-2 codent eux-mêmes des régulateurs d'épissage (pour revue, (Kuroyanagi, 2009)).

### -Nova

Nova-1 et la protéine apparentée Nova-2 sont des protéines se fixant à l'ARN qui ont tout d'abord été identifiées comme des antigènes auto-immuns chez des patients avec des désordres neurologiques (Darnell, 2006). Les protéines Nova sont trouvées exclusivement dans les cellules neuronales où elles sont localisées dans le noyau. Nova-1 est importante pour la survie des cellules neuronales à un stade de développement post-natal. La fonction de Nova-1 n'est pas essentielle mais fait partie de facteurs indispensables pour l'épissage. Les protéines Nova contiennent 3 domaines de liaison à l'ARN de type KH. Le domaine KH3 se lie préférentiellement au motif ARN UCAY (Y:pyrimidine) et est suffisant pour la fixation spécifique à l'ARN. Les protéines Nova sont impliquées dans la régulation de l'épissage alternatif de pré-ARNm codant des protéines synaptiques comme par exemple les pré-ARNm de la sous-unité 2 du récepteur GABA et de la sous-unité  $\alpha 2$  du récepteur de la glycine (Dredge *et al.*, 2003; Ule *et al.*, 2003; Ule *et al.*, 2005).

### **II.3.2 Implication de la structure secondaire de l'ARN dans la régulation de l'épissage.**

Les pré-ARNm sont souvent représentés pour des raisons pratiques sous forme linéaire alors qu'il est plus probable que les ARN se replient en une structure secondaire qui va ensuite former des interactions tertiaires qui peuvent impliquer jusqu'à des centaines de nucléotides. Les pré-ARNm messagers sont donc en partie double brins. Cette conformation va influencer la reconnaissance des sites d'épissage, les structures adoptées pouvant inhiber ou activer l'assemblage du spliceosome. La reconnaissance des sites d'épissage et des séquences régulatrices va se faire différemment avec un ARN en parti double-brin qu'avec un ARN simple brin, les facteurs protéiques se liant différemment. Les structures d'ARN peuvent dissimuler des sites d'épissage et empêcher l'assemblage du spliceosome ou bien l'activer en masquant des sites fixant les inhibiteurs.

### Le rôle de la structure secondaire dans l'activation de l'épissage

Un exemple bien caractérisé de la façon dont la structure secondaire de l'ARN affecte l'épissage est l'inclusion de l'exon alternatif de la fibronectine EDA. Dans ce cas, une séquence en aval stabilise une séquence ESE en amont dans la boucle d'une structure tige-boucle qui est accessible aux protéines SR. Ceci va permettre la reconnaissance de l'exon EDA. Les mutations de la séquence en aval causent un changement conformationnel tel que l'ESE se trouve dans la tige et est relativement inaccessible aux protéines SR. L'exon EDA ne va donc pas pouvoir être reconnu (Buratti *et al.*, 2004).

### Le rôle de la structure secondaire dans l'inhibition de l'épissage

L'exon 10 épissé alternativement de Tau dans lequel le site 5' est séquestré dans un élément tige-boucle. La perturbation de cette tige-boucle par des mutations provoque une augmentation de la fixation de la snRNP U1 au site 5' et augmente l'inclusion de l'exon 10 (Varani *et al.*, 1999).



## **CHAPITRE III : Régulation de l'épissage du transcrit primaire du virus de l'immunodéficience humaine de type I.**

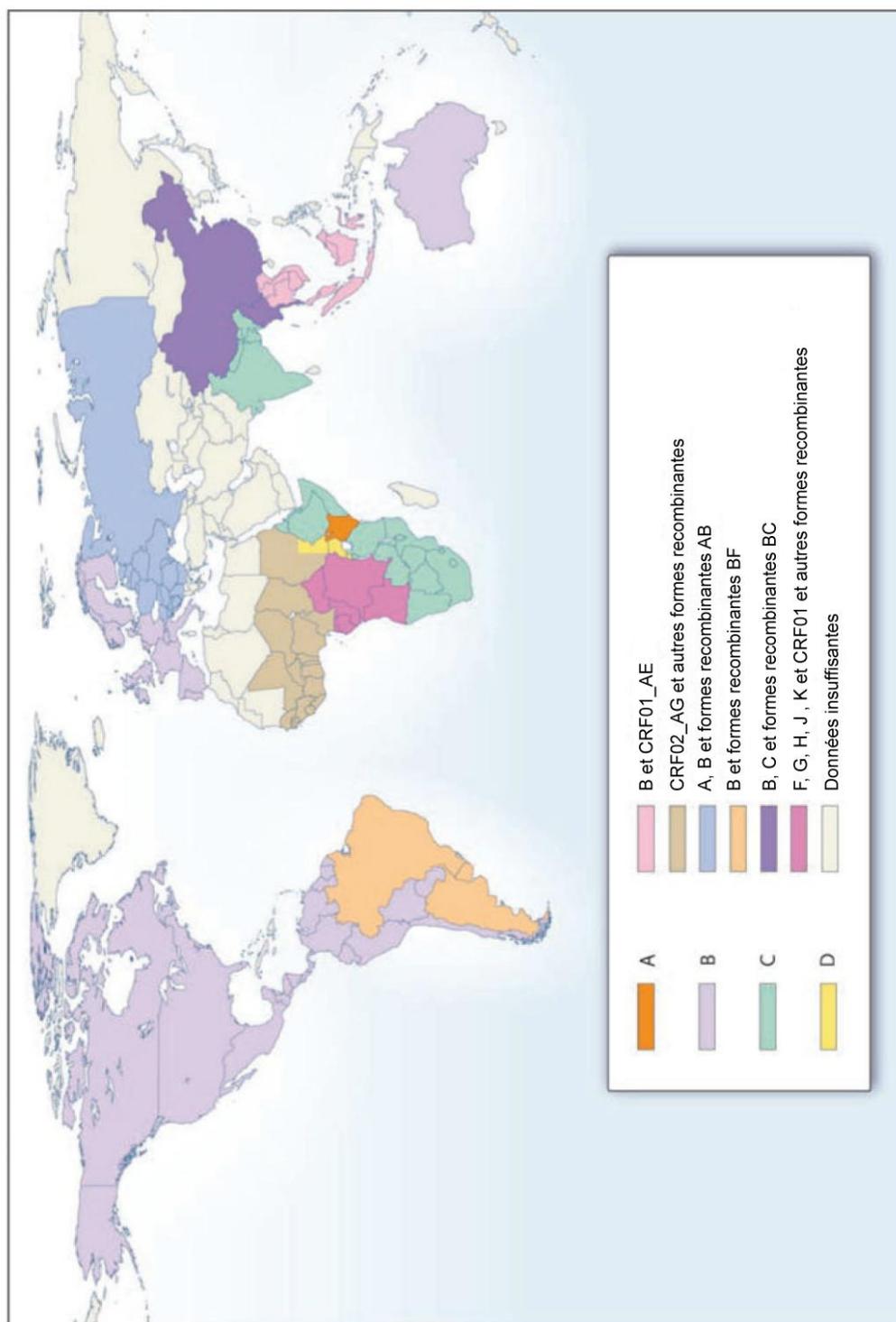
Le virus de l'immunodéficience acquise humaine (VIH-1) est un lentivirus de la famille des rétrovirus qui cause le syndrome d'immunodéficience acquise (SIDA), menant à une grave déficience du système immunitaire qui permet le développement d'infections opportunistes. Les travaux de R. Gallo et L. Montagnier (Gallo *et al.*, 2002) ont permis l'identification du rétrovirus en 1983 par l'équipe du Professeur Montagnier (Barre-Sinoussi *et al.*, 1983). Ces travaux ont été récompensés en 2008 par le prix Nobel de médecine décerné à Luc Montagnier et Françoise Barré-Sinoussi.

### **III.1 Description du virus VIH-1**

Les rétrovirus sont des particules infectieuses consistant en un génome ARN empaqueté dans une capsidie protéique, entourée d'une enveloppe lipidique. Cette enveloppe lipidique contient des chaînes polypeptidiques incluant des récepteurs viraux qui lient les récepteurs de membrane de la cellule hôte, initiant le processus d'infection. Les particules rétrovirales contiennent aussi l'enzyme transcriptase inverse qui permet la synthèse d'un ADN complémentaire (ADNc) en utilisant l'ARN viral comme matrice. Quand le rétrovirus infecte une cellule, il introduit son ARN dans le cytoplasme avec la transcriptase inverse. L'ADNc produit de la matrice ARN contient les instructions génétiques du virus et permet l'infection d'autres cellules.

#### **III.1.1 Evolution phylogénétique du virus VIH-1**

Le SIVcpz, qui infecte plusieurs communautés de chimpanzés isolées géographiquement dans le sud du Cameroun, serait à l'origine du VIH-1. Cet ancêtre est probablement passé des chimpanzés aux humains à travers une transmission par le sang, comme par exemple lors de la chasse et de la préparation de viande de primate. Des analyses phylogénétiques du VIH-1 et de virus apparentés de primates non-humains suggèrent qu'il a existé au moins quatre événements de transmission au début du 20<sup>ème</sup> siècle engendrant 4 groupes de VIH-1 : le groupe M (pour Majeur), le groupe O ( pour Outlier) et le groupe N (pour New, Non-majeur, Non-outlier) et le groupe P identifié récemment (Plantier *et al.*, 2009) (Figure I-31).

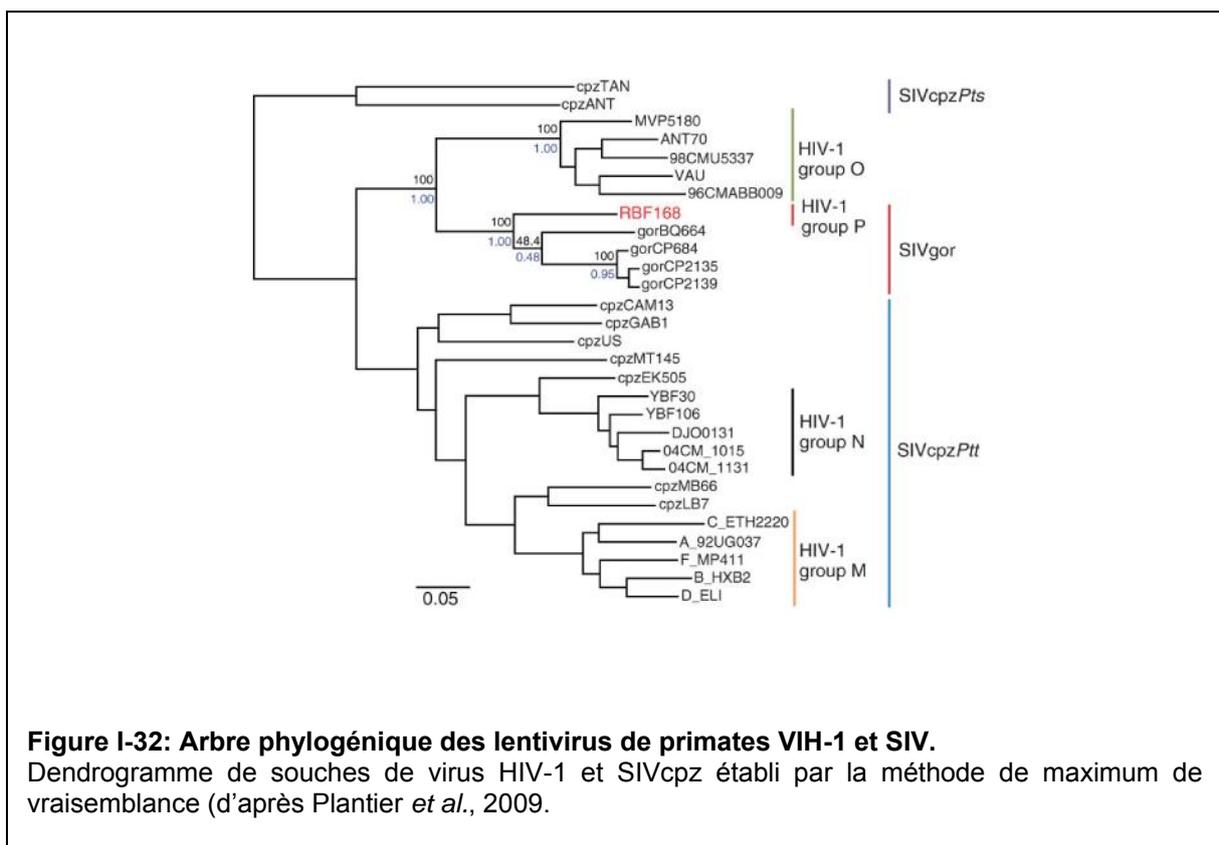


**Figure I-31: Distribution globale des sous-types de VIH-1 et des formes recombinantes. D'après Taylor *et al.*, 2008.**

Le groupe M est la cause de 95% des infections sur la planète alors que les trois autres groupes sont plus rares, voire ponctuel pour le groupe P (Figure I-32). Les études estiment que l'ancêtre commun de tous les VIH-1 groupe M date de l'année 1930 [IC : 1915-1941] (Korber *et al.*, 2000; Salemi *et al.*, 2001). Une période similaire est également estimée pour

l'origine du VIH-1 du groupe O: 1920 [IC : 1891-1940] (Lemey *et al.*, 2004) et aucune période n'a été pour l'instant définie pour les groupe N et O. Tandis que les souches apparentées aux groupes M et N ont été trouvées chez les chimpanzés, des preuves récentes montrent que le groupe O pourrait provenir des gorilles, chez lesquels les virus les plus proches de ce groupe ont été identifiés. Le virus s'est dispersé parmi les humains le long de la rivière Congo dans Kinshasa (Zaïre), où ont été décrits les plus anciens cas d'infection par le VIH-1 (avec la souche M) chez les humains avec une identification dans un échantillon sanguin de 1959.

Les groupes M, N, O et P sont eux-mêmes divisés en sous-types, sous-sous-types et formes recombinantes circulantes (CRF) (<http://www.hiv.lanl.gov/>). Le groupe M peut être ensuite divisé en plusieurs sous-types (A-K). Le sous-type B, qui englobe toutes les souches décrites originellement en Amérique du Nord et en Europe, est très rare en Afrique (Sharp *et al.*, 2008).



Le groupe O semble être endémique au Cameroun et les pays voisins dans l'ouest de l'Afrique Centrale, mais ces virus représentent une minorité des souches de VIH-1 : leur plus

haute prévalence est de 2 à 5% des échantillons positifs au VIH-1 (Peeters *et al.*, 1997; Zekeng *et al.*, 1994).

Le groupe N a été seulement récemment identifié, et est représenté par un nombre limité d'isolats de patients camerounais (Ayoub *et al.*, 2000; Simon *et al.*, 1998).

Très récemment, un nouveau groupe, nommé groupe P, a été isolé chez une femme camerounaise. Il est très proche du virus simien d'immunodéficience du gorille (SIVgor) et ne montre pas de preuves de recombinaison avec d'autres lignées de VIH-1. Ce nouveau virus semble être le prototype d'une nouvelle lignée de VIH-1 qui est distincte des groupes M, N et O (Plantier *et al.*, 2009).

### III.1.2 Le virion : génome et structure

#### III.1.2.1 Structure du virus VIH-1

Le VIH-1 a une structure à peu près sphérique avec un diamètre d'environ 120 nm. Le génome est composé de deux copies d'ARN simple brin positif qui codent les 9 gènes du virus. La membrane externe du virus est composée de lipides issus des cellules humaines et de protéines virales (par exemple gp41) qui traversent cette membrane lipidique. Le virus induit la synthèse d'un précurseur gp160 codé par le gène *env*, de poids moléculaire de 160kDa qui est clivé en 2 protéines plus petites, une glycoprotéine de 110/120 kDa qui est la protéine de surface gp120 et une de 41 kDa qui est la protéine transmembranaire gp41 (Figure I-33).

Une coiffe faite de 3 molécules glycoprotéiques gp120 est ancrée à une tige composée de 3 molécules gp41 liées par des interactions non-covalentes et associées en trimère à la surface virale pour ancrer la structure dans l'enveloppe virale. La protéine gp120 de surface (SU) se fixe spécifiquement sur les protéines CD4 des lymphocytes et des macrophages. Ces protéines gp120 correspondent aux spicules d'attachement (ce sont les "piquants" souvent représentés sur les schémas du virus). La protéine gp 41 transmembranaire (TM) participe à la fusion entre l'enveloppe virale et la membrane phospholipidique de la cellule. Ces protéines de surface et plus particulièrement gp120, sont considérées comme cibles de futurs traitements ou vaccins contre le VIH (Durova *et al.*, 2009).

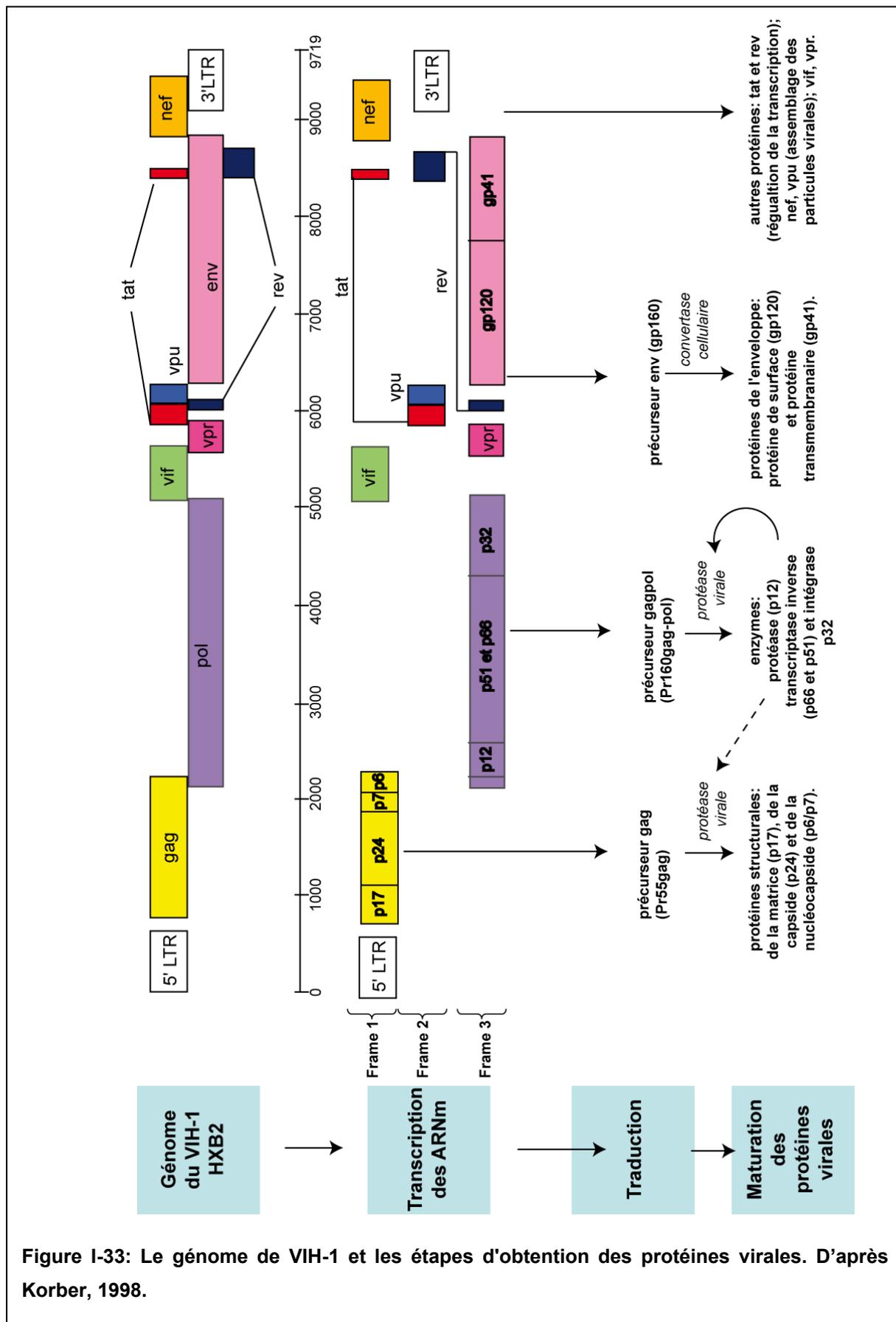


Figure I-33: Le génome de VIH-1 et les étapes d'obtention des protéines virales. D'après Korber, 1998.

La particule virale contient en plus les protéines virales Nef, Vif, Vpr et p6, ainsi que plusieurs facteurs cellulaires (Frankel *et al.*, 1998; Ott *et al.*, 2000). Parmi eux se trouve l'ARNt<sub>3</sub><sup>lys</sup> qui est utilisé comme amorce pour l'initiation de la transcription inverse.

### III.1.2.2 L'ARN génomique

Le génome de VIH-1 est un ARN d'environ 9200 pb qui contient les gènes *gag*, *pol* et *env*, ce qui est typique de tous les rétrovirus. Ces gènes codent les polyprotéines qui sont clivées par protéolyse en protéines individuelles. La polyprotéine Gag est clivée pour donner les protéines de matrice MA (p17), de capsid CA (p24) et de nucléocapsid NC (p7) qui vont former le cœur du virion, de même que la protéine p6 trouvée dans le virion. Le gène *Env* code les protéines de membrane SU (gp120) et TM (gp41) et le gène *Pol* code les enzymes PR (p12), RT (p66/p51) et IN (p32).

Le génome de VIH-1 contient de plus les gènes *tat*, *rev*, *nef*, *vif*, *vpr* et *vpu*, codant les protéines Tat et Rev, ainsi que les protéines accessoires, Nef, Vif, Vpr et Vpu. Ces protéines additionnelles sont codées par les ORF séparées et chevauchantes (Figure I-34).

Les protéines du VIH-1 peuvent être ordonnées en trois classes :

- Les protéines structurales ou enzymes virales : les produits des gènes *gag*, *pol* et *env* sont des composants essentiels de la particule rétrovirale.
- Les protéines régulatrices : les protéines Tat et Rev de VIH-1 modulent transcriptionnellement et post-transcriptionnellement les étapes de l'expression des gènes du virus et sont essentielles pour la propagation du virus.
- Les protéines accessoires ou auxiliaires : les protéines additionnelles du virion et non associées au virion produites par le rétrovirus VIH : Vif, Vpr, Vpu, et Nef. Tandis que les protéines accessoires ne sont en général pas nécessaires pour la propagation virale des tissus en culture, elles ont été conservées dans les différents isolats; cette conservation et les observations expérimentales suggèrent que leur rôle *in vivo* est très important.

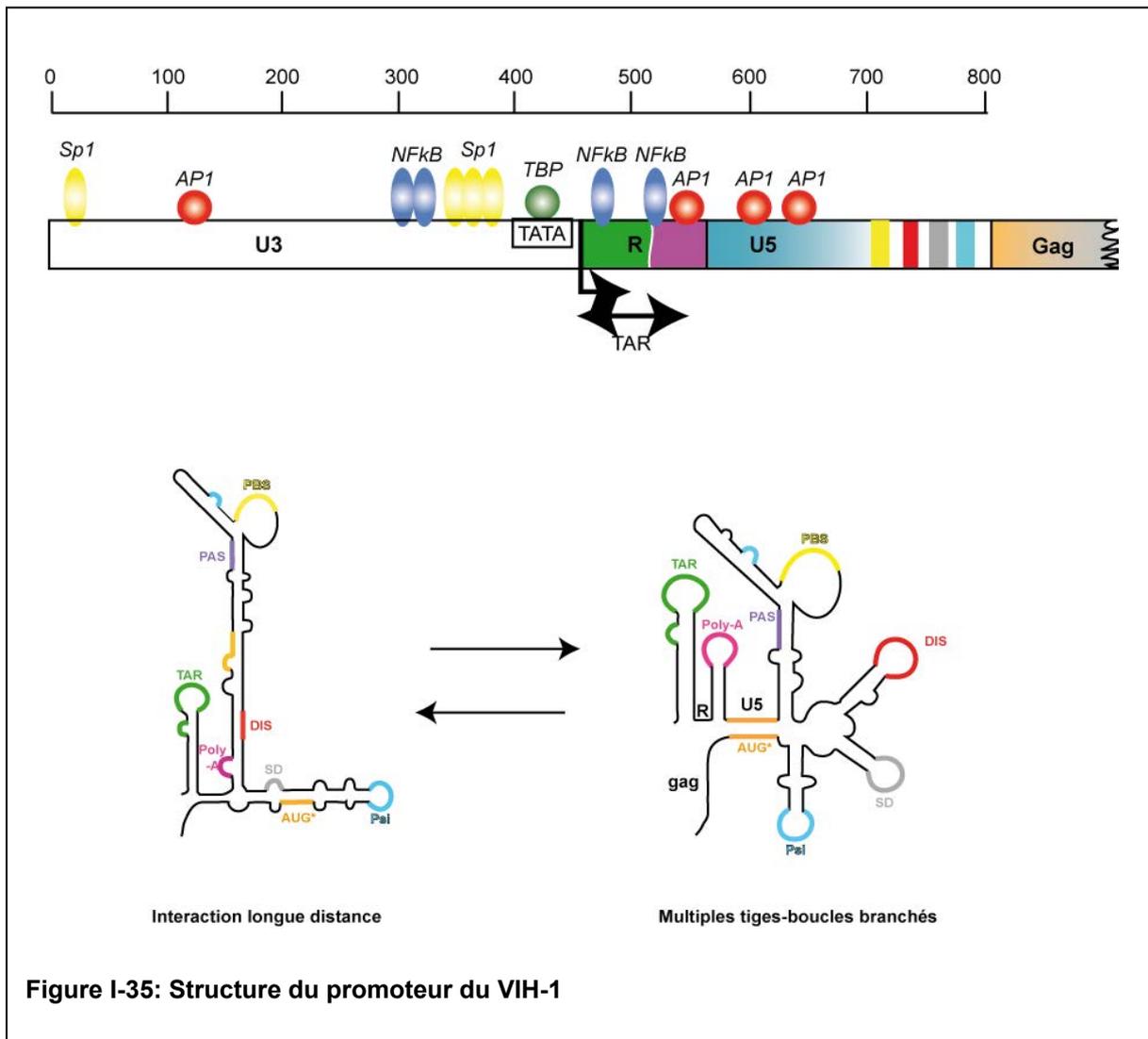
Protéine	Taille	Fonction	Localisation
<b>Gag</b>			
MA	p17	ancrage à la membrane; interaction avec env; transport nucléaire du core viral (pour les protéines myristoylées)	Virion
CA	p24	capside	Virion
NC	p7	nucléocapside, lie l'ARN	Virion
	p6	lie Vpr	Virion
<b>Pol</b>			
Protéase (PR)	p15	clivage et maturation de Gag/Pol	Virion
Transcriptase inverse (RT)	p66, p51	Transcription inverse	Virion
RNase H	p15	Activité RNase H	Virion
Intégrase (IN)	p31	Intégration de l'ADN proviral dans la cellule hôte	Virion
Env	gp120/gp41	Glycoprotéines virales externes se liant aux récepteurs CD4 et aux co-récepteurs	Membrane plasmique Enveloppe du virion
Tat	p16/p14	Transactivateur de la transcription virale	Nucléole/noyau
Rev	p19	Transport et stabilité des ARN viraux (phosphoprotéine)	Nucléole/noyau Navette entre le nucléole et le cytoplasme
Vif	p23	Maturation du virion et infectivité	Cytoplasme/virion
Vpr	p10-15	Permet la localisation nucléaire du complexe de préintégration, inhibe la division cellulaire, arrête les cellules infectées en G2/M	Noyau, virion (membrane nucléaire ?)
Vpu	p16	Favorise le relargage extra-cellulaire des particules virales; dégrade les CD4 dans le réticulum endoplasmique	Membranes
Nef	p27-p25	Dégradation des récepteurs CD4; inhibition de la traduction; altération des voies d'activation des lymphocytes T.	Cytoplasme Membrane plasmique (Virion?)
Tev	p28	Protéine tri-partite Tat-Env-Rev	Nucléole/noyau

Figure I-34: Fonctions et localisations cellulaires des protéines de VIH-1.

### III.1.2.3 Les éléments structuraux du génome de VIH

L'ARN génomique comporte plusieurs points de repère structuraux dont:

- Le LTR (Long Terminal Repeat) est la séquence ADN flanquant le génome des provirus intégrés. Elle contient d'importantes régions régulatrices, spécialement celles pour l'initiation de la transcription et la polyadénylation. Les régions LTR agissent comme des commutateurs pour contrôler la production de nouveaux virus et peuvent être déclenchées par des protéines de VIH ou de la cellule hôte.
- La région 5' non traduite : l'extrémité 5' de l'ARN génomique renferme les séquences R (redondant), U5 (Unique 5') et PBS (Primer binding Site) nécessaires pour l'étape de transcription inverse (Figure I-35).
- La séquence R contient la séquence TAR qui est la séquence cible pour la transactivation virale, elle est le site de fixation pour la protéine Tat et pour les protéines cellulaires et contient approximativement les 45 premiers nucléotides de l'ARNm viral de VIH-1 (ou les 100 premiers nucléotides dans VIH-2 et SIV). L'ARN TAR forme une structure de type « hairpin » tige-boucle avec une structure de type « bulge » sur le côté qui est nécessaire pour la fixation de Tat et sa fonction.
- Les séquences polypurines PPT 3' (PolyPurine Track) sont importantes pour l'étape de transcription inverse. Elles servent de base de départ pour la synthèse d'ADNc (Charneau *et al.*, 1992).
- L'élément RRE (Rev Responsive Element) est un élément ARN de réponse à Rev encodé dans la région *env* de VIH-1. Il comprend approximativement 200 nucléotides, enjambant la bordure de gp120 et gp41 (Mishra *et al.*, 2006).
- Les sites d'épissage : dans le VIH-1 il existe 4 sites donneurs (D1, D2, D3 et D4) et 8 sites accepteurs d'épissage (A1, A2, A3, A4a, A4b, A4c, A5, et A7) utilisés en combinaison pour générer les différents ARNm. L'ARN de certaines souches virales peut contenir un ou deux autres sites fonctionnels A6 et D5.



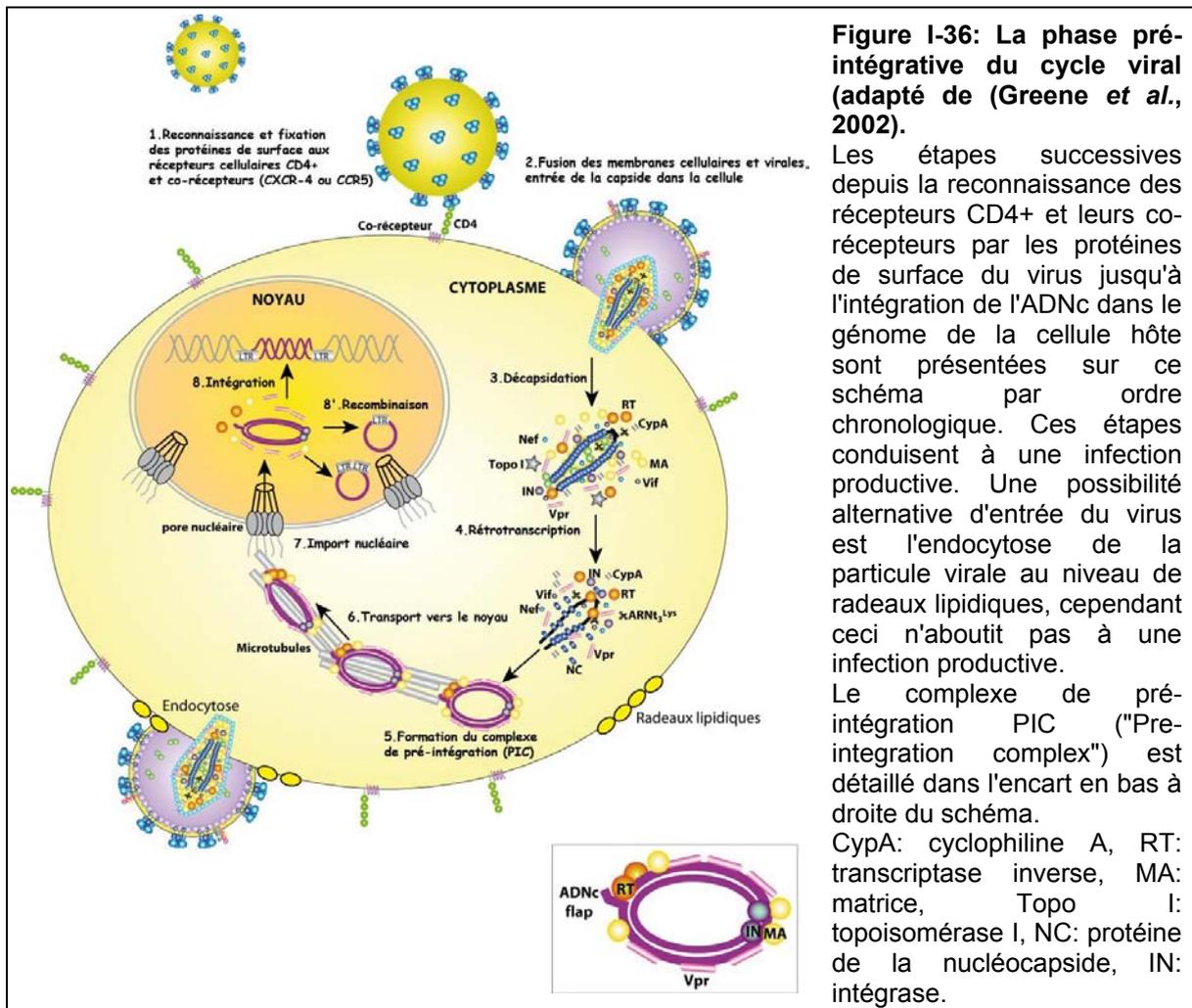
- Les phases ouvertes de lecture : le génome viral de VIH-1 contient 9 phases ouvertes de lecture (ORF : Open Reading Frame) *gag*, *pol*, *env*, *tat*, *rev*, *nef*, *vif*, *vpr* et *vpu*. Trois sont communes à l'ensemble des rétrovirus car indispensables pour la multiplication virale; il s'agit de *gag*, *pol* et *env* qui codent les précurseurs protéiques du même nom.

## III.2 Cycle de multiplication du virus VIH-1 et fonction des protéines dans le cycle viral.

### III.2.1 La phase pré-intégrative

#### III.2.1.1 Entrée du virion dans la cellule infectée

Le virus VIH-1 entre dans les cellules à récepteur CD4, comme les lymphocytes T auxiliaires, les cellules de la lignée monocyte-macrophage, les cellules dendritiques et les cellules neuronales. La fixation de gp120 sur CD4 induit un remaniement conformationnel aboutissant à l'exposition d'un nouveau site (appelée site CD4 induit) qui est impliqué dans la reconnaissance d'une deuxième molécule cellulaire (CCR5 ou CXCR4) agissant comme corécepteur. Ceci conduit au rapprochement entre cellule et virus puis à la fusion de l'enveloppe virale avec la membrane cellulaire et la libération de la capsid du VIH-1 dans la cellule (Figure I-36).



Le processus d'entrée peut être divisé en 3 étapes : l'attachement viral, la fixation du co-récepteur et la fusion. La première étape obligatoire dans le processus d'entrée du VIH-1 est la liaison spécifique de gp120 au CD4, le récepteur primaire pour le VIH-1 (Dalglish *et al.*, 1984; Klatzmann *et al.*, 1984; Maddon *et al.*, 1986). La fixation de gp120 au CD4 cause une reconfiguration des boucles variables V1/V2 et V3 de gp120 pour exposer le site faisant la liaison et former un site de fixation du co-récepteur (Kwong *et al.*, 1998; Rizzuto *et al.*, 1998; Wyatt *et al.*, 1998). La fixation au récepteur déclenche des changements conformationnels dans gp120 et gp41, ce qui conduit à l'exposition du peptide fusion (FP) et à son insertion dans la membrane de la cellule cible.

Des réarrangements ultérieurs dans le gp41 amènent les membranes du virus et de la cellule cible proches l'une de l'autre, créant ensuite un pore de fusion à travers lequel le cœur viral passe dans la cellule (Chan *et al.*, 1998). Plus d'une douzaine de co-récepteurs au VIH ont été identifiés, dont BOB et CXCR6 (Mueller *et al.*, 2007), mais tous les types de VIH utilisent au moins CCR5 (Deng *et al.*, 1996) ou CXCR4 (Feng *et al.*, 1996) pour l'entrée dans la cellule *in vivo*. Des inhibiteurs de chacune de ces étapes ont été identifiés et certains sont utilisés pour inhiber la fusion du virion avec la cellule comme l'enfuvirtide ou le maraviroc (Makinson *et al.*, 2009).

### III.2.1.2 Transcription inverse

La transcription inverse est l'étape de réplication du virus qui convertit le génome ARN en une copie ADN provirale, un mécanisme qui est partagé par les rétrovirus, les rétrotransposons et les hépadnavirus. La transcriptase inverse (RT) est obtenue après maturation du précurseur Gag-Pol et a une activité ADN polymérase, ARN et ADN dépendante. Elle fonctionne comme un hétérodimère avec des sous-unités de 66 et 51 kDa. La sous-unité p51 comprend seulement le domaine polymérase (Goff, 1990) et résulte d'un clivage de p66, qui est composée d'un domaine pour la polymérase RT et d'un domaine RNase H, par la protéase (PR) (di Marzo Veronese *et al.*, 1986; Lightfoote *et al.*, 1986). L'activité polymérase a été attribuée au composant p66 de l'hétérodimère (Richter-Cook *et al.*, 1992). Le processus de transcription inverse est composé de plusieurs étapes. La matrice est le génome d'ARN viral qui est flanqué de séquences répétées (LTR) aux extrémités 5' et 3'. Comme pour toutes les ADN polymérases, la transcriptase inverse a besoin d'une amorce avec un groupe 3'OH libre pour initier la synthèse de l'ADNc. La synthèse du brin d'ADN de polarité négative ou ADN (-) est initiée par l'hybridation des 18 nucléotides 3' terminaux de

l'ARNt3lys sur les 18 nucléotides complémentaires de la région PBS située dans la région 5' non codante du génome viral (Ratner *et al.*, 1985). La transcriptase inverse assure alors la synthèse des régions U5 puis R. Le brin d'ADN ainsi néoformé est appelé « ADN (-) strong stop ». La transcriptase inverse du VIH-1 est un hétérodimère composé de deux sous-unités, p66 (560 acides aminés) et p51 (440 acides aminés). Ces sous-unités sont obtenues après clivage de la polyprotéine Gag-Pol par la protéase virale. P66 et p51 partagent une séquence N-terminale commune. Les deux fonctions enzymatiques de la transcriptase inverse, ADN polymérase et RNase H, sont essentielles pour copier l'ARN simple brin en ADN double brin. La plus grande sous-unité p66 contient les sites actifs pour les deux activités enzymatiques tandis que la plus petite sous-unité a un rôle structural.

Cette polymérase a un grand taux d'erreur estimé à  $3 \times 10^{-5}$  par base par cycle de réplication (Mansky *et al.*, 1995). Par ailleurs, des erreurs peuvent être générées par une cytidine déaminase cellulaire, APOBEC3G (APOB mRNA-Editing enzyme catalytic polypeptide 3G ou CEM15) reconnaissant les ADNc étrangers et produite par la cellule hôte. APOBEC3G est empaquetée dans le virion, et conduit à la désamination des résidus déoxycytosine sur le premier brin d'ADN viral synthétisé durant la transcription inverse en déoxyuracile. Ainsi, APOBEC3G a une fonction antivirale pour l'hôte en produisant un grand nombre de mutations G en A dans le génome de VIH-1 le rendant non fonctionnel.

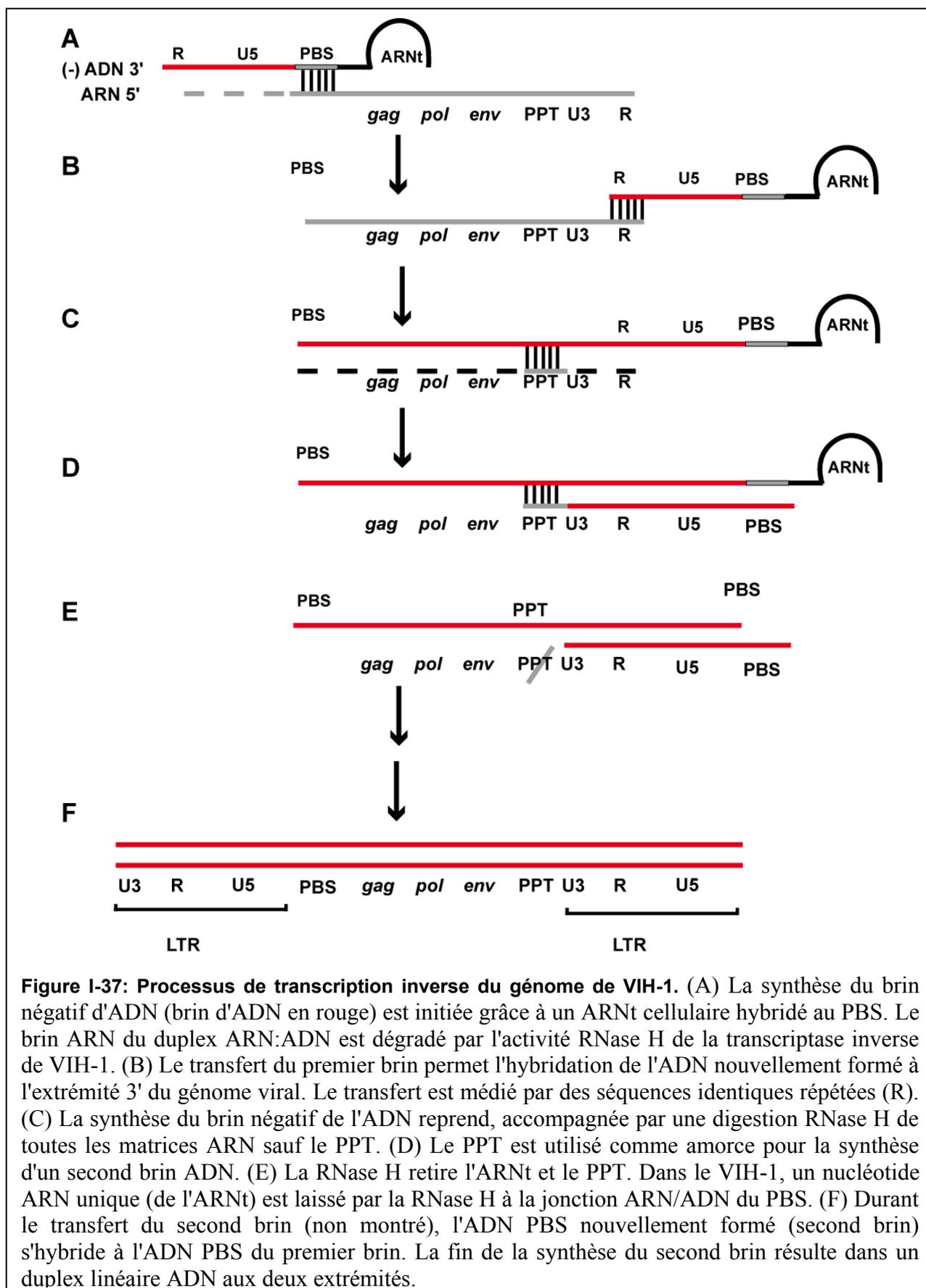
Des changements dans l'ORF de Vpr de VIH-1 Bru, G-en-A (position 5301) et C-en-U (position 5396) génèrent respectivement un codon stop et un codon Arg (AGG) au lieu de Gly (GGG). Cette même protéine Vpr a la capacité de recruter UNG, une uracile ADN glycosylase, ce qui semble avoir un effet sur le taux de mutation. UNG est une enzyme de réparation de l'ADN qui retire spécifiquement de l'ADN les bases uraciles présentes dans l'ADNc proviral (Mansky *et al.*, 2000).

### III.2.1.3 Transport et intégration de l'ADN proviral dans le génome

A la différence de la plupart des rétrovirus qui profitent de la dissolution de la membrane nucléaire pendant la mitose pour accéder au génome de la cellule hôte, le complexe de pré-intégration du VIH peut entrer dans le noyau pendant l'interphase (Figure I-37).

Le complexe de transcription inverse s'associe avec la tubuline puis est transporté vers le noyau le long des microtubules. Durant cette translocation, l'ARN génomique viral est converti en ADN double brin et s'assemble avec des récepteurs d'imports (importines) et des

protéines cellulaires qui vont protéger l'ADN de la dégradation et faciliter l'intégration (Turelli *et al.*, 2001). Cet ensemble est le complexe de pré-intégration (PIC). Il interagit avec des protéines de la matrice virale, l'intégrase et la protéine virale R (Vpr).



**Figure I-37: Processus de transcription inverse du génome de VIH-1.** (A) La synthèse du brin négatif d'ADN (brin d'ADN en rouge) est initiée grâce à un ARNt cellulaire hybridé au PBS. Le brin ARN du duplex ARN:ADN est dégradé par l'activité RNase H de la transcriptase inverse de VIH-1. (B) Le transfert du premier brin permet l'hybridation de l'ADN nouvellement formé à l'extrémité 3' du génome viral. Le transfert est médié par des séquences identiques répétées (R). (C) La synthèse du brin négatif de l'ADN reprend, accompagnée par une digestion RNase H de toutes les matrices ARN sauf le PPT. (D) Le PPT est utilisé comme amorce pour la synthèse d'un second brin ADN. (E) La RNase H retire l'ARNt et le PPT. Dans le VIH-1, un nucléotide ARN unique (de l'ARNt) est laissé par la RNase H à la jonction ARN/ADN du PBS. (F) Durant le transfert du second brin (non montré), l'ADN PBS nouvellement formé (second brin) s'hybride à l'ADN PBS du premier brin. La fin de la synthèse du second brin résulte dans un duplex linéaire ADN aux deux extrémités.

Vpr contribue à la liaison du PIC aux nucléoporines. La présence de Vpr est indispensable pour le transport du PIC vers le noyau (Popov *et al.*, 1998). La présence d'une séquence nommée « DNA flap central » située au milieu du génome viral semble déterminante pour l'entrée du PIC par les pores nucléaires, elle assiste l'import nucléaire en fournissant une conformation optimale au PIC pour transloquer à travers le pore nucléaire (Zennou *et al.*, 2000). Après l'import, l'ADNc proviral est intégré dans le génome de la cellule hôte par l'intégrase virale (pour revue, Delelis *et al.*, 2008). Cette enzyme est codée par l'extrémité 3' du gène *pol* et résulte de la maturation du précurseur Gag-Pol. L'intégrase a une fonction dans la stabilisation du PIC lors de son transport dans le noyau et une fonction catalytique pour intégrer l'ADNc dans l'ADN génomique de la cellule hôte. Elle se lie à une courte séquence sur chaque LTR et catalyse un clivage endonucléotidique et l'élimination de deux résidus sur le brin 5' à chaque extrémité de l'ADNc proviral, ce qui conduit à des extrémités 3' protubérantes. L'étape de clivage est effectuée dans le cytoplasme cellulaire. Le transfert de brins est ensuite réalisé par attaque nucléophile des extrémités 3'OH de l'ADN viral sur une liaison phosphodiester de l'ADN de la cellule hôte, entraînant la ligation des extrémités virales 3' avec les extrémités 5' de l'ADN cellulaire au sein du noyau. Le processus d'intégration est achevé par des enzymes cellulaires permettant la réparation des brèches formées par l'insertion de l'ADN viral (Skalka *et al.*, 2005). Ces brèches sont comblées par la duplication de quelques nucléotides de part et d'autre du provirus. L'ADN proviral n'a pas de sites d'intégration définis, mais il existe des sites préférentiels d'intégration au niveau des sites actifs de transcription (Marshall *et al.*, 2007).

### III.2.2 La phase post-intégrative

#### III.2.2.1 Expression des gènes viraux

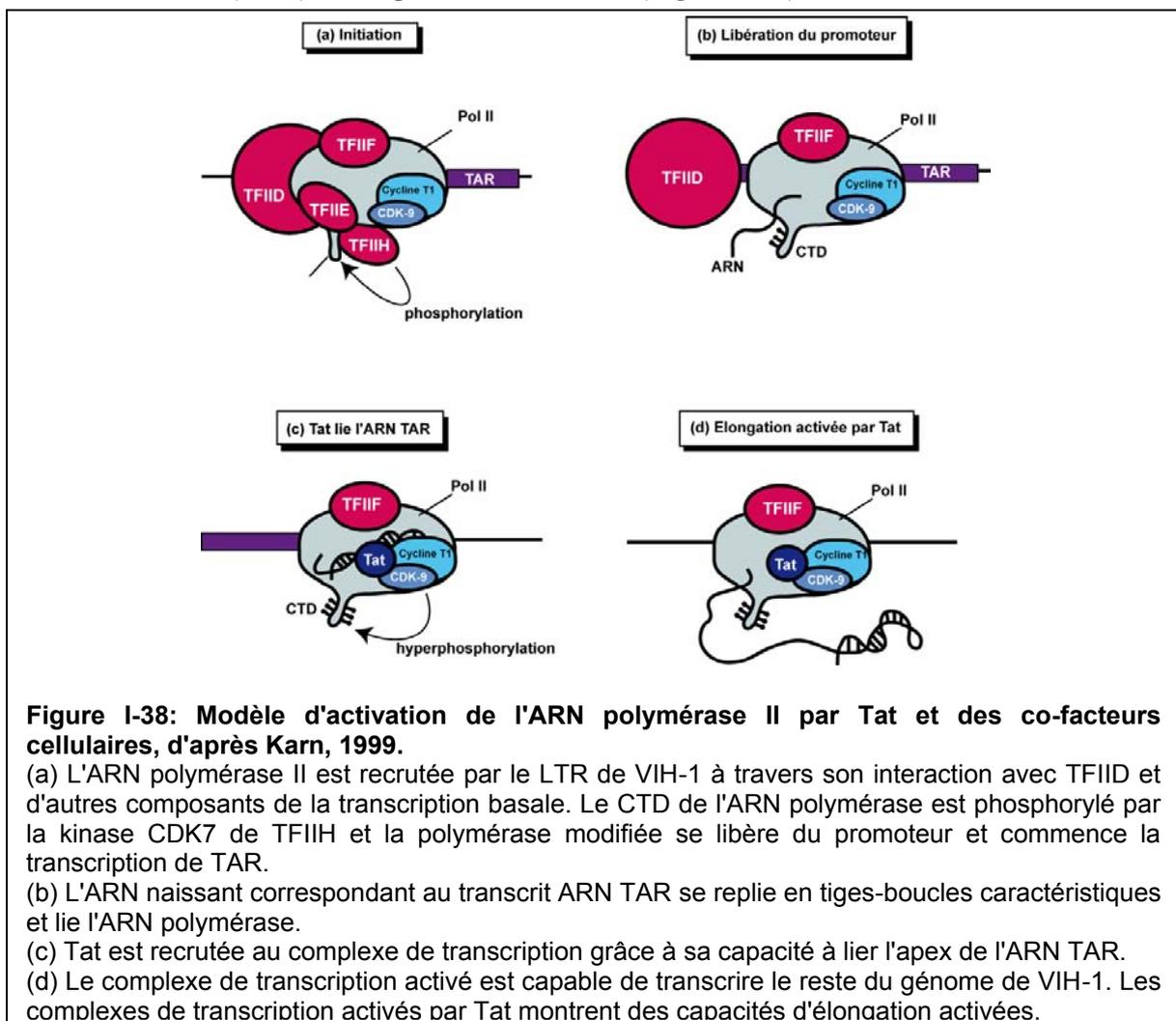
##### *a) Transcription de l'ADN proviral*

La synthèse d'un ARNm mature et fonctionnel de VIH lors de la transcription est un processus complexe en plusieurs étapes qui requiert l'action coopérative de protéines virales et cellulaires (Cullen, 1991; Kingsman *et al.*, 1996). Après l'intégration de l'ADNc viral dans la chromatine, le provirus s'intègre au sein du génome de la cellule hôte en adoptant le même type de condensation que la cellule qu'il infecte : la structure nucléosomale. De plus, le positionnement des nucléosomes au sein de l'ADN proviral n'est pas aléatoire. Les nucléosomes sont des barrières à la transcription et *nuc 1* (nucleosome 1) et empêchent

l'assemblage du complexe de transcription (Widlak *et al.*, 1998). Le remodelage des nucléosomes par des complexes d'acétylation des histones ATP-dépendants déclenche la transcription basale à partir du promoteur de VIH-1 (Pumfery *et al.*, 2003; Quivy *et al.*, 2002; Van Lint, 2000).

Le LTR 5' de VIH-1 se comporte comme un promoteur inducible lors de la transcription. De nombreux facteurs de transcription (NF- $\kappa$ B, Sp1, AP-1, NFAT) interagissent avec le LTR à des sites de fixations définis (Gaynor, 1992; Pereira *et al.*, 2000) et permettent la fixation de l'ARN polymérase II. Cependant, en absence de protéine Tat transactivatrice, le taux d'élongation par l'ARN polymérase est faible et n'entraîne principalement la synthèse de transcrits courts non poly-adenylés. La protéine Tat est aussi capable de stimuler la traduction de son propre ARNm (Charnay *et al.*, 2009).

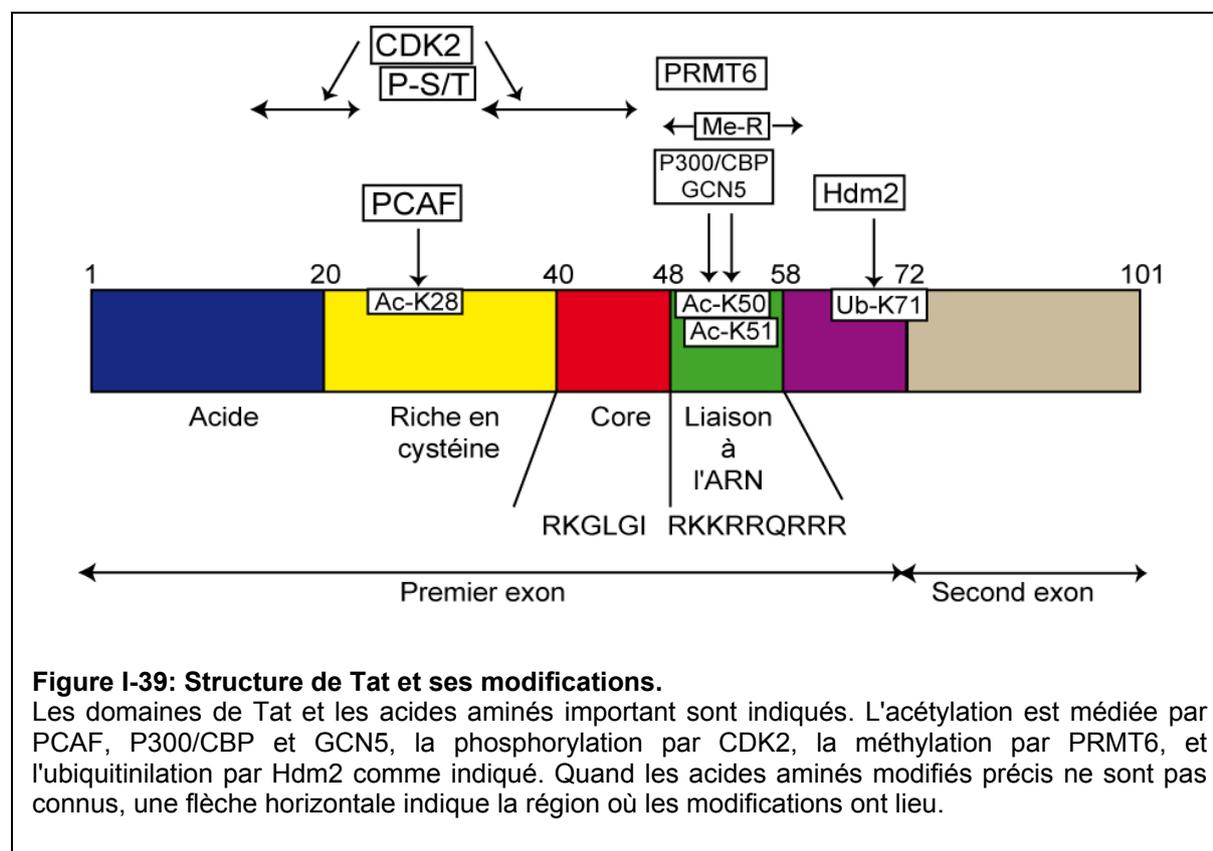
La protéine Tat requiert au minimum la transcription préalable du domaine ARN de réponse à la transactivation (TAR) sur lequel elle va se fixer (Figure I-38).



Le domaine d'activation de Tat se lie ensuite à la cycline T1 du complexe pTEFb, dont le partenaire est CDK9 (Wei *et al.*, 1998). pTEFb et Tat lient TAR coopérativement. Tat recrute la TBP, TFIIB et p-TEFb au promoteur pour former un complexe actif de pré-initiation dans lequel CDK9 hyperphosphoryle le domaine C-terminal (CTD) de l'ARN polymérase II (Garber *et al.*, 2000). Ce PIC est capable d'activer le départ de la polymérase et l'élongation transcriptionnelle.

*La protéine Tat.*

La protéine Tat transactivatrice (Figure I-39) de VIH-1 est une protéine d'environ 11 kDa, (apparaissant à 14 kDa en SDS-PAGE) codée par un ARNm contenant 2 exons (pour revue, Gatignol, 2007). L'exon 1 code les acides aminés de 1 à 72 et l'exon 2 de 73 à 86 ou de 73 à 101 selon la souche du virus.



Tandis que la forme entière de Tat de 101 acides aminés est la plus fréquente dans les cas cliniques, les variants tronqués de 72 et 86 acides aminés sont souvent utilisés pour les études *in vitro*. L'importance fonctionnelle des 14 derniers acides aminés a été étudiée dans les cellules T, dans lesquelles Tat 101 a une capacité pour induire l'apoptose réduite par rapport à

Tat86 (Campbell *et al.*, 2005). L'extrémité C-Terminale de 14 acides aminés a aussi un rôle dans le processus de neurotoxicité dans lequel Tat est impliquée (Passiatore *et al.*, 2009). L'équipe de KT. Jeang (Bethesda, USA) a montré que la substitution d'un résidu A (position 8016) par un résidu C a une forte influence sur la pathogénicité du virus (Smith *et al.*, 2003). La substitution de ce résidu A en C conduit à la substitution du codon STOP prématuré dans l'ORF de Tat par un codon sérine.

La protéine Tat peut être divisée en 5 domaines :

- Le domaine I N-terminal (acides aminés de 1 à 20). Les mutations de cette région ne modifient pas la transactivation de façon importante.
- Le domaine II contient 7 cystéines très conservées qui sont importantes pour la formation de ponts intramoléculaires disulfures.
- Le domaine III (acides aminés de 40 à 48) est essentiel pour la transactivation. Dans cette région, un simple changement de la Lysine 41 abolit l'activité de Tat.

Les trois premiers domaines (acides aminés de 1 à 48) représentent le domaine d'activation de Tat qui fonctionne comme un transactivateur quand il est lié à la région IV ou à un domaine de liaison hétérologue ARN/ADN. Toutefois, les 56 premiers acides aminés de la protéine Tat sont suffisants pour permettre une transactivation complète (Kuppuswamy *et al.*, 1989).

- Le domaine IV (acides aminés de 49 à 72) contient une séquence riche en arginine qui joue un rôle dans la fixation à l'ARN et la localisation nucléaire.
- Le domaine V, correspondant aux acides aminés 73 à 86 (ou 101 selon les souches), correspond au second exon. Ce domaine contient un motif RGD, qui n'est pas conservé dans les souches de HIV-2 et de SIV. Il est nécessaire pour la liaison de Tat à des intégrines cellulaires. Il contient aussi un motif conservé ESKKKVE, impliqué dans le transport des ARNm. Cet élément n'est pas nécessaire au mécanisme de transactivation de la transcription, cependant il contribue à la pathogénicité du virus. En effet, des souches ayant une protéine Tat86 sont moins virulentes que des souches possédant une protéine Tat101 (Neuveut *et al.*, 2003; Smith *et al.*, 2003).

*Autres activités de la protéine Tat.*

•**dans la traduction** : En plus de son activité de transactivation, Tat peut aussi réguler la traduction (Rosen *et al.*, 1986; Cullen, 1986). L'ARN TAR peut inhiber *in trans* la traduction de l'ARNm de VIH-1 (Parkin *et al.*, 1988; SenGupta *et al.*, 1989), probablement à travers

l'activation de la protéine PKR et l'addition de Tat peut inverser cette inhibition traductionnelle (Edery *et al.*, 1989; SenGupta *et al.*, 1990).

•**effets sur la fonction cellulaire** : Tat a des effets pléiotropiques sur les gènes cellulaires et le métabolisme de la cellule hôte (pour revue, Chang *et al.*, 1995). Tat fonctionne comme un facteur de croissance sécrété dans les cellules Kaposi-like (Ensoli *et al.*, 1990; Ensoli *et al.*, 1993; Barillari *et al.*, 1993) et stimule l'angiogénèse (Albini *et al.*, 1996). L'expression de plusieurs cytokines est modulée par Tat (Rautonen *et al.*, 1992) dont TNF $\alpha$  et  $\beta$  (Buonaguro *et al.*, 1994), TGF  $\alpha$  et  $\beta$  (Lotz *et al.*, 1994); (Nabell *et al.*, 1994), IL-2 (Purvis *et al.*, 1992) et IL-6 (Scala *et al.*, 1994). Tat active des voies de transduction du signal qui impliquent la phosphatidylinositol-3-kinase (Milani *et al.*, 1996) et NF- $\kappa$ B (Biswas *et al.*, 1995). Dans le cerveau, Tat induit une toxicité/dysfonction neuronale, ce qui résulte en une pathologie du système nerveux central, comme la démence et l'encéphalite associées à un neurosida (King *et al.*, 2006).

### *Implication de Nef dans l'activation de la transcription.*

La protéine Nef (p27) régule négativement le récepteur majeur viral, le CD4, ainsi que les molécules CMH de classe I et II (Piguet *et al.*, 1999). Bien que non indispensable *in vitro*, la protéine Nef permet d'atteindre un haut niveau de réplication virale dans les cellules infectées par un mécanisme encore mal connu. Nef coopère fonctionnellement avec le transactivateur viral Tat en dé-réprimant le promoteur viral grâce à un complexe kinase associé à Nef (NAKC). HnRNP K va interagir avec Nef dans le contexte du complexe NAKC et recrute Lck, PKC $\delta$  et la kinase PI-3. La formation de ce complexe est suffisante pour activer Lck et Erk1/2, lequel, dans les cellules T, va causer une augmentation de la transcription dépendant de Tat (Wolf *et al.*, 2008).

### ***b) Epissage des ARN pré-messagers***

L'épissage des ARN pré-messagers est indispensable dans la production des protéines rétrovirales de VIH-1. Les transcrits primaires sont gardés dans une forme non épissée pour servir d'ARNm pour la production du précurseur Gag-Pol ou être transportés dans le cytoplasme pour servir de génome pour les nouveaux virions. Ils peuvent aussi subir le processus d'épissage alternatif pour obtenir les différentes espèces d'ARNm qui permettront l'expression des protéines virales. Dans le virus VIH-1, l'épissage d'un transcrit primaire de

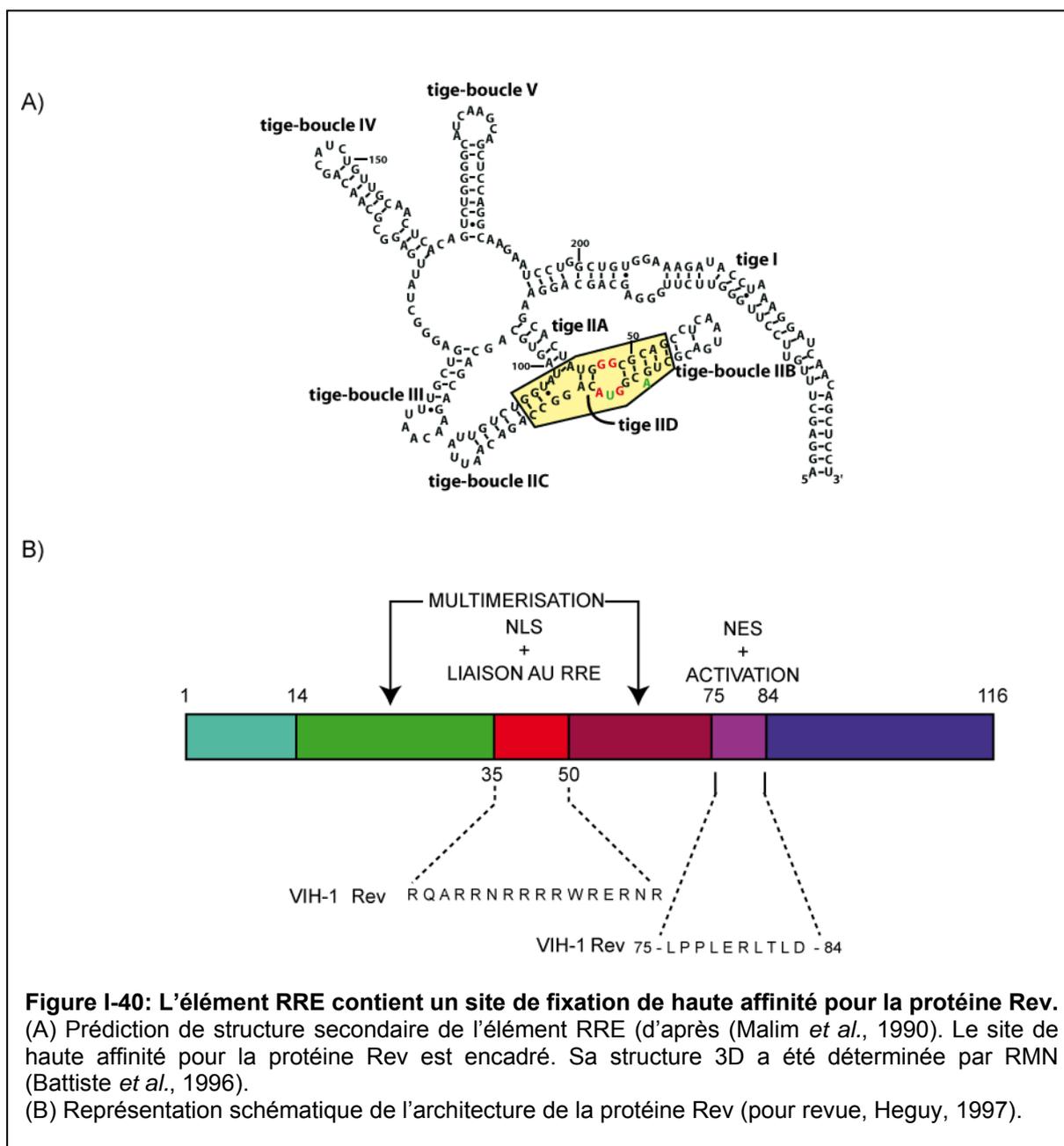
9.2 kb produit plus de 30 espèces d'ARNm (Purcell *et al.*, 1993) qui peuvent être divisés en 3 classes principales. La classe des transcrits épissés de façon multiple de 2 kb code les protéines Rev, Tat et Nef. La classe des transcrits simplement épissés de 4 kb code les protéines Env, Vpu, Vpr et Vif.

La production des ARNm codant les protéines VIH-1 dépend de l'utilisation alternative en combinaison de 4 sites 5' sites d'épissage D1 à D4 et de 8 sites 3' d'épissage (A1, A2, A3, A4a, -b, -c, A5 et A7).

### **c) Transport des ARN messagers et de l'ARN génomique**

La stratégie d'expression génomique du virus VIH-1 nécessite l'export d'ARN génomique non épissé. Rev est une phosphoprotéine de 116 acides aminés (19 kDa) qui possède un domaine N-terminal riche en arginine qui lui permet d'être importée vers le noyau (NLS), un domaine de liaison à la séquence ARN RRE (Rev Responsive Element) et deux domaines de multimérisation flanquant le domaine de liaison à l'ARN (Figure I-40). Le domaine C-terminal riche en leucine, connu comme signal d'activation, contient le signal d'export nucléaire de 10 acides aminés riche en leucines (NES) (Fischer *et al.*, 1995).

Les pré-ARNm cellulaires contenant des introns encore non épissés sont maintenus dans le noyau par interaction avec les facteurs d'épissage, jusqu'à ce qu'ils soient complètement épissés ou dégradés. La protéine Rev se fixe à l'élément RRE des ARNm non épissés ou épissés une seule fois et leur permet d'échapper au phénomène de rétention nucléaire, en favorisant leur transport du noyau vers le cytoplasme (pour revues, Stutz *et al.*, 1998; Hope, 1999). Une protéine de la famille des importines, CRM1 (Chromosome Maintenance Region 1), est le transporteur qui reconnaît le NES (Nuclear Exportation Signal; signal d'export nucléaire) et adresse la protéine Rev et les ARN viraux au pore nucléaire (Fornerod *et al.*, 1997). Rev agit à travers la voie CRM1. Le mécanisme d'export requiert l'hydrolyse du GTP par la GTPase Ran. Dans le noyau, en présence de la forme RanGTP, le récepteur CRM1 fixe son substrat. Le complexe traverse le pore nucléaire, grâce à l'interaction spécifique de CRM1 avec les nucléoporines, puis va dans le cytoplasme.



Lorsque la concentration en protéine Rev augmente dans la cellule infectée, la protéine se fixe à son site de haute affinité, puis multimérisse sur l'élément RRE par fixation coopérative des monomères le long de la structure tige-boucle I du RRE ((Malim *et al.*, 1991; Mann *et al.*, 1994). La phase tardive de réplication du VIH-1 se produit quand le niveau de protéine Rev (un des premiers gènes traduits) atteint une concentration assez élevée pour permettre l'exportation de l'ARN du VIH-1 non épissée ou partiellement épissée dans le cytoplasme. La protéine Rev joue un rôle primordial pour l'expression des gènes du virus VIH-1, car elle permet le passage de la phase précoce d'infection, caractérisée par la production des ARNm multiplement épissés codant les protéines Tat, Rev, Nef, à la phase tardive d'infection, au

cours de laquelle les ARNm sont peu épissés et permettent la production des protéines structurales et auxiliaires.

### III.2.2.2 Assemblage, bourgeonnement et maturation des nouveaux virions

L'assemblage, la libération et la maturation de la protéine virale sont des processus indispensables pour la production d'une nouvelle particule virale mature capable d'infecter une nouvelle cellule cible. La protéine structurale Gag joue un rôle central dans la voie de production de nouvelles particules virales et dirige le processus d'obtention d'une particule infectieuse mature à travers des interactions protéine-protéine, protéine-ARN et protéine-lipides. Ces interactions facilitent la multimérisation de Gag pour former la coque de la particule, l'encapsidation du génome ARN et le trafic des composants du virion et des glycoprotéines associés à l'enveloppe, détournant la machinerie de la cellule hôte pour faciliter la libération du virus et la maturation protéolytique du virion naissant.

#### *a) Recrutement des protéines virales*

La polyprotéine Gag, connue comme Pr55Gag, est traduite du transcrit ARNm viral non épissé par les ribosomes libres du cytoplasme. Le précurseur Pr55Gag est clivé en différentes protéines : MA, CA, NC, p2 et p6. Pr55Gag est la seule molécule codée par le virion requise pour l'assemblage des pseudo-particules virales immatures VLP (Virus-like particles) (Gheysen *et al.*, 1989). Elle peut s'assembler elle-même en VLP par une multimérisation ordonnée d'environ 5000 monomères de Pr55Gag, ce qui va former l'architecture primaire de la particule virale (Briggs *et al.*, 2004).

Plusieurs composants structuraux du précurseur Gag après protéolyse sont impliqués dans l'assemblage des particules virales :

- MA joue un rôle structural dans l'assemblage mais qui n'est pas déterminant dans la multimérisation de Gag tandis qu'elle a des fonctions importantes dans le trafic de Gag et dans l'incorporation de la protéine Env dans les virions.
- Le CTD joue un rôle majeur dans la multimérisation de Pr55Gag. CA est composée de deux domaines indépendants NTD et CTD (N-terminal domain et C-terminal domain) qui sont séparés par un domaine de liaison court et flexible (Berthet-Colominas *et al.*, 1999).

- Le domaine d'interaction (I) de Pr55Gag, localisé dans NC, joue un rôle majeur dans la multimérisation de Gag et dans l'assemblage. NC contient 2 motifs en doigts de zinc qui contribuent à la spécificité de l'encapsidation de l'ARN génomique viral mais n'est pas requis pour l'assemblage.

- Le domaine p6 contient des séquences qui interagissent directement avec des composants de la machinerie MVB. Des compartiments cellulaires, appelés corps multivésiculaires (multivesicular bodies, MVB), contenant des vésicules internes peuvent fusionner avec la membrane plasmique de la même manière que pour une exocytose. Dans certains types cellulaires, particulièrement dans les macrophages dérivés des monocytes primaires, l'assemblage et le bourgeonnement ont lieu dans le corps multivésiculaire (MVB). La protéine Vpr interagit avec la p6 après la maturation du virion.

-La protéine Vif agit durant l'assemblage viral pour assurer l'infectivité dans la cellule cible (Cullen, 1998), Vif interagit avec Gag (Simon *et al.*, 1999; Zhang *et al.*, 2000). L'interaction avec la protéine virale et l'ARN est importante pour l'assemblage correct de la particule et la stabilité du complexe de transcription inverse (Ohagen *et al.*, 2000).

### ***b) Recrutement des ARN viraux***

#### *Recrutement de l'ARN viral*

Le domaine NC de Gag a pour rôle d'encapsider et de protéger l'ARN viral. Deux copies d'ARN sont contenues dans chaque particule et le domaine NC lie ces ARN grâce à ses doigts de zinc. L'encapsidation du génome de VIH-1 implique une séquence de 110 nucléotides connue sous le nom de site Psi, qui contient 4 tiges-boucles requises pour l'emballage du génome. Deux de ces tiges-boucles lient le NC et l'une d'entre elle stabilise la structure de ce site (Amarasinghe *et al.*, 2001). C'est la liaison à l'ARN qui aide dans la localisation et la concentration des monomères de Gag pour permettre l'assemblage de Gag. Le rôle de l'ARN dans assemblage est si important que le VIH-1 recrutera d'autres ARN cellulaires si l'ARN viral n'est pas disponible (Khorchid *et al.*, 2002).

#### *Le recrutement de l'ARNt<sup>Lys</sup>.*

Durant l'assemblage du virus, les isoaccepteurs ARNt<sup>Lys</sup> humains majeurs ARNt<sub>1,2</sub><sup>Lys</sup> et ARNt<sub>3</sub><sup>Lys</sup> sont sélectivement empaquetés dans le VIH-1 durant l'assemblage (environ 20 molécules par virion) (Huang *et al.*, 1994), où l'ARNt<sub>3</sub><sup>Lys</sup> agit comme amorce pour l'initiation

de la transcription inverse. Leur concentration relative, par rapport à la population d'ARN de faible poids moléculaire, passe de 5-6 % dans la cellule à 50-60 % dans le virus (Kleiman *et al.*, 2004). La sélection des ARNt<sup>Lys</sup> est indépendante de l'incorporation de l'ARN viral et de la maturation des précurseurs protéiques (Kleiman *et al.*, 2004). Elle dépend de l'incorporation des précurseurs Gag-Pol et plus précisément de l'interaction entre la RT et l'ARNt<sub>3</sub><sup>Lys</sup> (Khorchid *et al.*, 2000). La fonction des ARNt<sub>1,2</sub><sup>Lys</sup> chez le VIH-1 n'est pas connue.

Le complexe d'empaquetage de l'ARNt<sub>3</sub><sup>Lys</sup> est formé quand un complexe Gag/Gag-Pol/ARN viral interagit avec un complexe ARNt<sup>Lys</sup>/LysRS, avec Gag interagissant avec lysRS et Gag-Pol avec l'ARNt<sup>Lys</sup>. L'incorporation de Gag-Pol dans VIH-1 requiert son interaction avec des multimères de Gag dont la polymérisation est stimulée par l'ARN (pour revue, Kleiman *et al.*, 2004).

### ***C) Traduction des ARN messagers***

La traduction des protéines virales est effectuée par la machinerie de la cellule hôte. Ce processus permet la synthèse des protéines lors de la lecture par les ribosomes de l'information contenue dans les ARN messagers. Le changement du cadre de lecture en -1 (ou décalage de phase en -1) est une anomalie de la traduction qui est utilisée par le virus de l'immunodéficience humaine de type 1 pour synthétiser ses enzymes lors de la traduction de l'un de ses ARN messagers. Le signal stimulateur, une structure particulière dans cet ARN, contrôle l'incidence de cette anomalie. Ce signal stimulateur est une tige-boucle irrégulière où une boucle interne faite de purines sépare la tige en deux portions hélicoïdales. Ce signal pourrait favoriser le décalage de phase en interagissant spécifiquement avec le ribosome ou bien avec un facteur protéique hypothétique. Seul un petit nombre de ribosomes effectuent le décalage de phase, et l'utilisation de cette stratégie permet d'obtenir une quantité de Gag-Pol par rapport à Gag qui est important pour la propagation du virus (Karacostas *et al.*, 1993).

La synthèse des protéines Env et Vpu ainsi que des précurseurs Gag et Gag-pol se fait à partir d'ARNm bi-cistroniques; l'ARNm bi-cistronique *env/vpu* contient deux ORF chevauchantes (Vpu et Env). A la différence des autres protéines, Vpu et Env sont traduites à partir du même ARNm avec une initiation ayant lieu à des codons d'initiation AUG séparés. L'initiation au AUG en aval pour produire Env est dépendant du balayage défailant (ou « leaky scanning ») du AUG du *vpu* en amont par la sous-unité ribosomale 43S. Dans ce mécanisme, le premier

codon initiateur n'est pas toujours celui utilisé, celui de l'ORF de Vpu, celui de Env est alors utilisé (Schwartz *et al.*, 1990; Schwartz *et al.*, 1992).

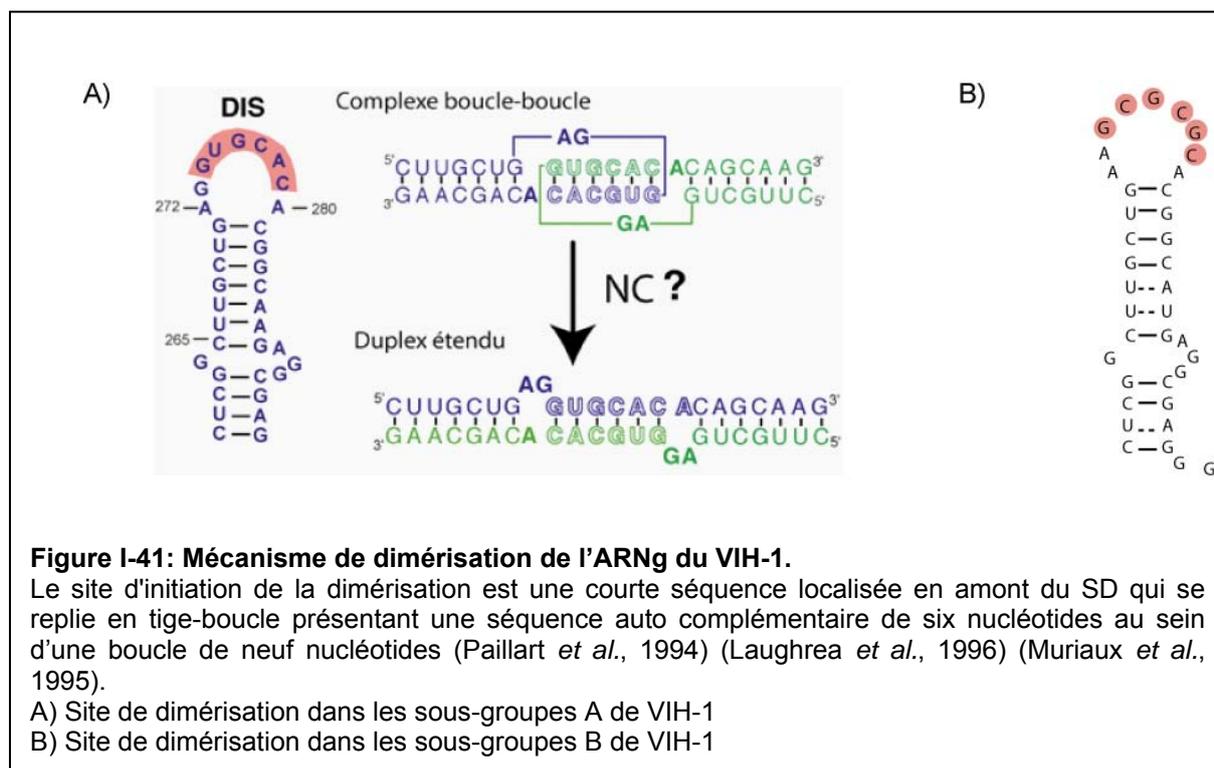
### ***d) Bourgeoisement et maturation du virion***

Lors du bourgeoisement, pour la libération du virus de la membrane plasmique, le domaine C-terminal p6 de VIH-1 Gag se lie directement à deux composants de cette machinerie de bourgeoisement cellulaire à travers les domaines L (Late-assembly). Ils lient le composant Tsg101 de ESCRT-1 et permettent la maturation protéolytique correcte de Gag pour la libération de la particule virale. Les protéines ESCRT sont des composants du système de trafic endosomal et sont essentiels pour le bourgeoisement de VIH-1 en coupant le lien membranaire qui maintient le virion à la membrane plasmique, libérant le virion de la cellule hôte (pour revue, McDonald *et al.*, 2009).

Durant ou pendant le processus de bourgeoisement, le processus de maturation est initié par le clivage de Pr55Gag et de Pr160Gag-Pol par la protéase virale. Ceci va induire des changements conformationnels dans le domaine CA et résulte en l'assemblage des différentes capsides coniques. La protéase de VIH-1 et les autres protéases virales codées par la protéine de fusion Gag-Pol sont générées quand 5% des ribosomes se décale en -1 tandis les protéines Gag-Pol qui en résultent sont empaquetées dans les virions via les interactions avec le grillage Gag-Gag. La protéase virale est libérée par auto protéolyse à un stade tardif de l'assemblage viral, ce qui permet de s'assurer que les protéines Gag ne sont pas maturées avant leur assemblage (Pettit *et al.*, 1998).

### **III.3 Effets de la machinerie d'épissage sur la dimérisation de l'ARN**

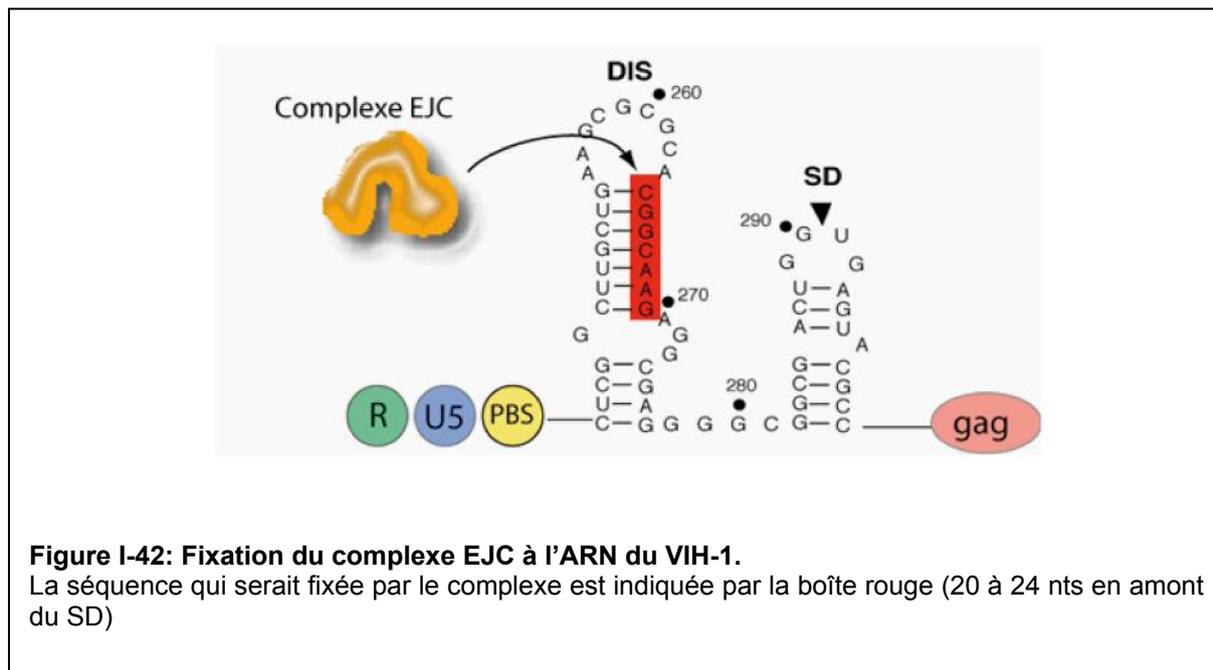
La sélection positive de l'ARN génomique au moment de l'assemblage de la particule virale est pour le moment un mécanisme mal connu. Si les ARN épissés du VIH-1 sont capables de dimériser dans la cellule, on peut imaginer que la sélection de l'ARN génomique puisse se faire dans le noyau, dès sa transcription (Figure I-41).



En effet, l'histoire nucléaire d'un ARNm peut influencer son devenir dans le cytoplasme, principalement à cause des protéines qui s'assemblent sur l'ARN (mRNP). Un exemple bien connu est la dépose du complexe EJC (Exon-Junction Complex) au moment de l'épissage et ses conséquences sur des événements tels que l'export nucléaire, la traduction ou encore la dégradation des ARNm par le NMD (Nonsense Mediated mRNA Decay) (Tange *et al.*, 2004). Le cœur de ce complexe est composé de 4 protéines: (1) eIF4AIII, une hélicase d'ARN appartenant à la famille des protéines à boîte DEAD; (2) Y14; (3) MAGOH et (4) MLN51 (ou Barentsz, Btz). Pendant l'épissage, l'EJC est déposé 20 à 24 nucléotides en amont de la jonction exonique sans spécificité de séquence et protège 8 à 9 nucléotides (Le Hir *et al.*, 2000) (Figure I-42).

La résolution récente de la structure cristallographique de ce complexe a montré que seules eIF4AIII et Btz sont en contact direct avec l'ARN (Bono *et al.*, 2006). A ce stade, il est intéressant à noter qu'au moment de l'épissage de l'ARN du VIH-1, le complexe EJC serait positionné dans la partie apicale du DIS. Ainsi, suivant l'ordre des événements, dimérisation et épissage pourraient réguler le devenir de l'ARN: (1) si l'épissage a lieu avant la dimérisation, l'EJC se dépose sur le DIS et pourrait inhiber la dimérisation des ARNm

épissés; (2) dans le cas contraire, l'épissage d'un brin d'ARN pourrait favoriser celui de la seconde molécule du dimère ou inhiber totalement l'épissage.



### III.4 Régulation de l'épissage de l'ARN viral

#### III.4.1. Production des ARNm viraux

L'expression différentielle des nombreux gènes des lentivirus, tels que VIH-1 a la particularité de reposer uniquement sur les étapes post-transcriptionnelles : épissage, traduction et stabilité relative des ARNm. L'épissage de l'ARN du VIH-1 est d'autant plus complexe qu'il met en jeu une compétition entre au moins 4 sites donneurs (D1, D2, D3 et D4) et 8 sites accepteurs (A1, A2, A3, A4a-c, A5 et A7) (Figure I-43), qui sont utilisés en de multiples combinaisons pour générer les différents ARNm (Figure (Figure I-44). Certaines souches de VIH-1, renferment même un autre couple de sites fonctionnels D5 et A6 (pour revue, (Saliou *et al.*, 2009).

Le site donneur D1 est utilisé dans tous les ARNm, et ceci quelle que soit la phase du cycle. Il peut être couplé aux sites A1, A2, A3, A4a, A4b, A4c ou A5 et plus rarement A7. Dans la phase précoce du cycle de multiplication, l'intron D4-A7 est presque toujours éliminé dans les ARNm. La production des ARNm *tat*, *rev* et *nef* nécessite respectivement l'utilisation des

sites A3 et A7 (ARNm *tat*), A4a, A4b ou A4c et obligatoirement A7 (ARNm *rev*), A5 (ARNm *nef*) (Purcell *et al.*, 1993) (Figure I-43).

Site donneur 5' d'épissage	Site accepteur 3' d'épissage	Point de branchement	Séquence polypyrimidine	Site d'épissage
Consensus $\begin{array}{c} \text{A} \\ \text{C} \end{array} \text{AG} \mid \begin{array}{c} \text{G} \\ \text{G} \end{array} \text{AGU}$	Consensus	YNYURAY	(Y)n	YAG   G
D1 (95) $\text{CUG} \mid \text{GUGAGU}$	A1 (26)	nd	UUUUCGGGUUUUUACAG	G
D1A (nd) $\text{CAG} \mid \text{GUAAGA}$	A1A (nd)	nd	UACUCCUCUUAAAAUUAG	C
D2 (38) $\text{AAG} \mid \text{GUGAAG}$	A2 (60)	UAGCAGA	UAUUUUGAUUGUUUUUCAG	A
D3 (45) $\text{AAG} \mid \text{GUAAGG}$	A3 (55)	nd	CUGCUGUUUAUCCAUUUCAG	A
	A4c	UUGC <u>U</u> UUGUAA	UGUUGCUUUC <u>U</u> CAUUGCCAAG	U
	A4a (48)	CUUU <u>C</u> AUUGCC <u>A</u> AG	UUUGUUUCAUGACAAAAG	C
	A4b	CUUU <u>C</u> AUUGCC <u>A</u> AG	UUUGUUUCAUGACAAAAGCUUAG	G
	A5	GUUUC <u>A</u> UG <u>A</u> CAAA <u>A</u> GCCU <u>U</u> AG	UCUCCU <u>A</u> UGGCAG	G
D4 (83) $\text{GCA} \mid \text{GUAAGU}$	A6 (nd)	nd	UGUGUUAGUUUAAAG	U
D5 (nd) $\text{AAG} \mid \text{GUGCAG}$	A7 (41)	UACUU <u>U</u> C...UAGGC <u>A</u> GGGAU <u>A</u> U	UU <u>A</u> UCGUUUUCAG	A

**Figure I-43: Séquences des sites 5' donneurs d'épissage et 3' accepteurs.**  
 Les nucléotides des sites 5' d'épissage qui ne s'apparient pas avec le U1 snRNA sont indiqués en rouge, et ceux qui sont supposés s'apparier avec le U6 snRNA sont marqués par un astérisque. Les jonctions exon-intron et intron-exon sont représentées par des barres verticales. Les numéros entre parenthèses indiquent l'efficacité relative de l'utilisation (en %) de sites d'épissage *in vivo*. Les nucléotides soulignés représentent les points de branchement identifiés ou putatifs. N=A, C, G ou U; Y=C ou U; R=A ou G

Comme pour tous les rétrovirus, une fraction de transcrits de VIH-1 restent intacts pour servir d'ARN génomique et pour coder les précurseurs protéiques Gag et Gag-Pol. L'épissage est aussi important pour contrôler la production de protéines virales qui pourraient avoir un effet négatif sur les cellules infectées (Figure I-44).

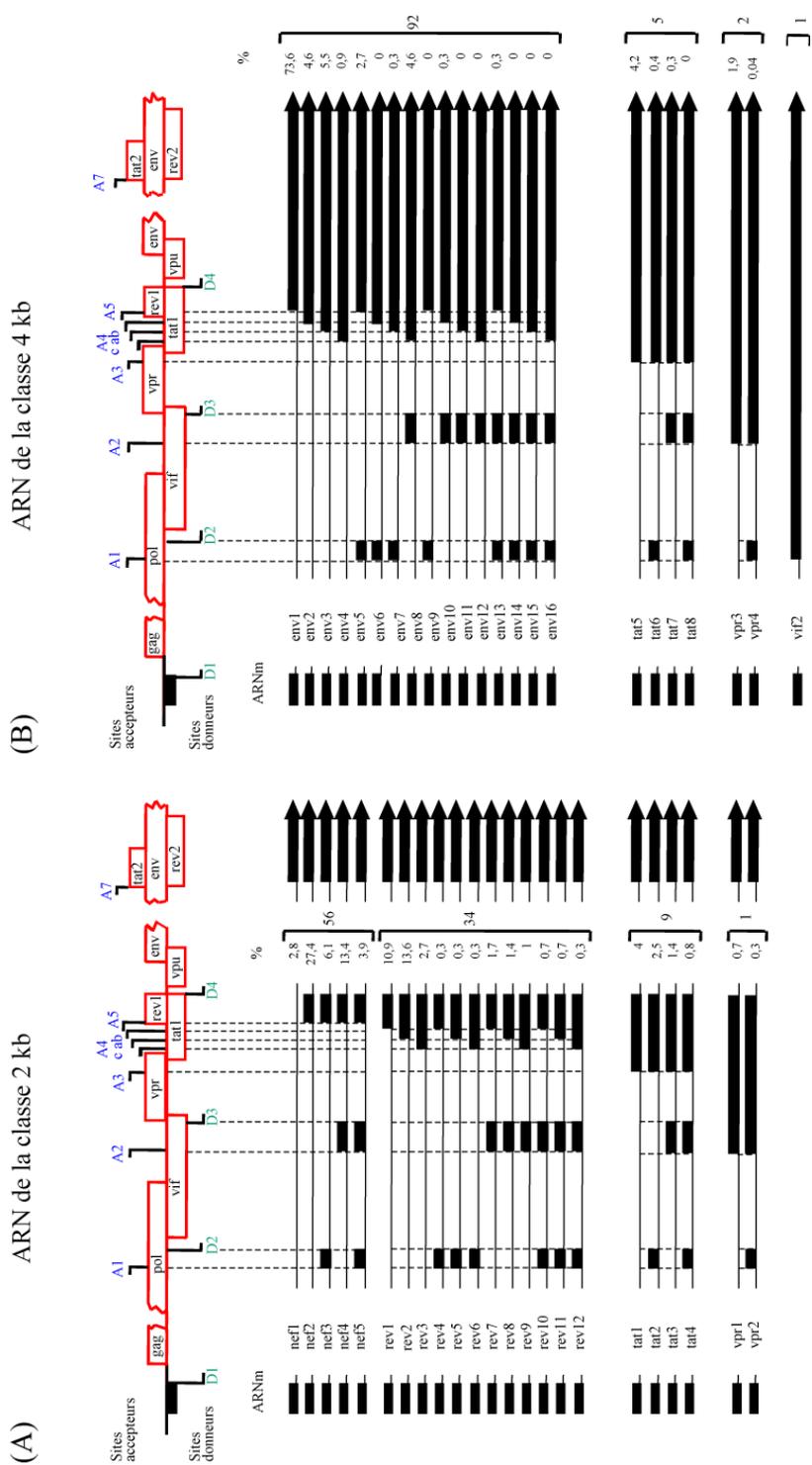


Figure I-44 : Composition en exons et abondance relative des ARNm des classes 2 kb (A) et 4 kb (B) du virus VIH-1. D'après Purcell *et al.*, 1993.

### III.4.2. Mécanisme d'épissage de l'ARN viral

#### III.4.2.1. Les sites donneurs d'épissage

Bien que les sites 5' donneurs d'épissages ne soient pas optimaux, la plupart d'entre eux ont une bonne complémentarité avec le snRNA U1, en particulier D1 et D4. La force relative des sites donneurs a été évaluée *in vitro* et *in vivo* (O'Reilly *et al.*, 1995; Damier, 1997) et est bien corrélée avec le niveau de complémentarité avec le snRNA U1. Cela suggère que la régulation au site 5' d'épissage est limitée, et aujourd'hui, seul quelques activateurs d'épissages ont été décrits pour agir aux sites 5'. On trouve par exemple l'élément GAR situé en amont du site D4, liant les protéines SR ASF/SF2 et SRp40. Cet élément agit de façon bidirectionnelle et active à la fois l'épissage au site A5 et le recrutement de la snRNP U1 au site D4. En effet, cet élément GAR augmente la complémentarité entre le snRNA U1 et le site D4 (Kammler *et al.*, 2001; Caputi *et al.*, 2004).

Le site D1 est le plus efficace de tous les sites 5' de VIH-1. Tous les événements d'épissage en aval dépendent de sa présence car aucun épissage n'aura lieu si il est muté (Baudin *et al.*, 1993). Une étude montre que si le site 5' donneur d'épissage D1 et son site cryptique sont mutés dans pNL4.3, seuls des ARN de VIH-1 non épissés s'accumulent dans les cellules transfectées. Ce résultat conforte l'hypothèse que l'épissage en aval du site donneur D4 avec le site accepteur A7 sont dépendants de l'épissage préalable de l'intron en amont du site D1 avec un site accepteur d'épissage en aval du site D1. Une des explications pourrait être que les ARN épissés de façon incomplète seraient dégradés par le NMD (Bohne *et al.*, 2005). Le site D1 est situé dans une région qui contient plusieurs éléments fonctionnels et une structure bidimensionnelle très conservée chez tous les types de VIH-1 (Abbink *et al.*, 2005; Abbink *et al.*, 2008). La faible force des sites D2 et D3 est importante pour limiter l'épissage des sites 3' d'épissage en amont A1 et A2 respectivement (Madsen *et al.*, 2005; Madsen *et al.*, 2006).

Les sites donneurs d'épissage peuvent aussi avoir des fonctions autres que l'épissage. Le site D4 va par exemple être important dans la stabilisation de l'ARNm *env* (Kammler *et al.*, 2001).

### III.4.2.2 Les sites accepteurs d'épissage du virus VIH-1 sont faibles

Tous les sites 3' accepteurs d'épissage de VIH-1 sont sous-optimaux. Ils possèdent des séquences polypyrimidines sous-optimales (Staffa *et al.*, 1994; Si *et al.*, 1997; Jacquenet *et al.*, 2001) car l'enchaînement de pyrimidines est entrecoupé de purines et ils possèdent des séquences de branchement non canoniques (Damier *et al.*, 1997; Swanson *et al.*, 1998; Tange *et al.*, 2001). Pour certains sites 3', deux séquences de branchement (ou plus) peuvent être utilisées, incluant des nucléotides autres que des résidus adénosines et présentant une faible efficacité intrinsèque (Si *et al.*, 1997; Damier *et al.*, 1997; Dyhr-Mikkelsen *et al.*, 1995; Kammler *et al.*, 2006). De plus, les sites 3' sont fréquemment inclus au sein d'ARN structurés très stables, comme pour les sites A3, A4c,a,c, A5, et A7 (Jacquenet *et al.*, 2001; Marchand *et al.*, 2002; Damgaard *et al.*, 2002). Leur utilisation est régulée négativement par la présence de nombreux éléments inhibiteurs d'épissage en *cis* (ESS) qui lient les protéines hnRNP, plus fréquemment les hnRNP A1/B (Marchand *et al.*, 2002; Caputi *et al.*, 1999; Bilodeau *et al.*, 2001; Jacquenet *et al.*, 2001; Tange *et al.*, 2001; Zhu *et al.*, 2001; Asai *et al.*, 2003).

### III.4.2.3. Epissage aux sites accepteurs A1 à A7 de l'ARN du VIH-1.

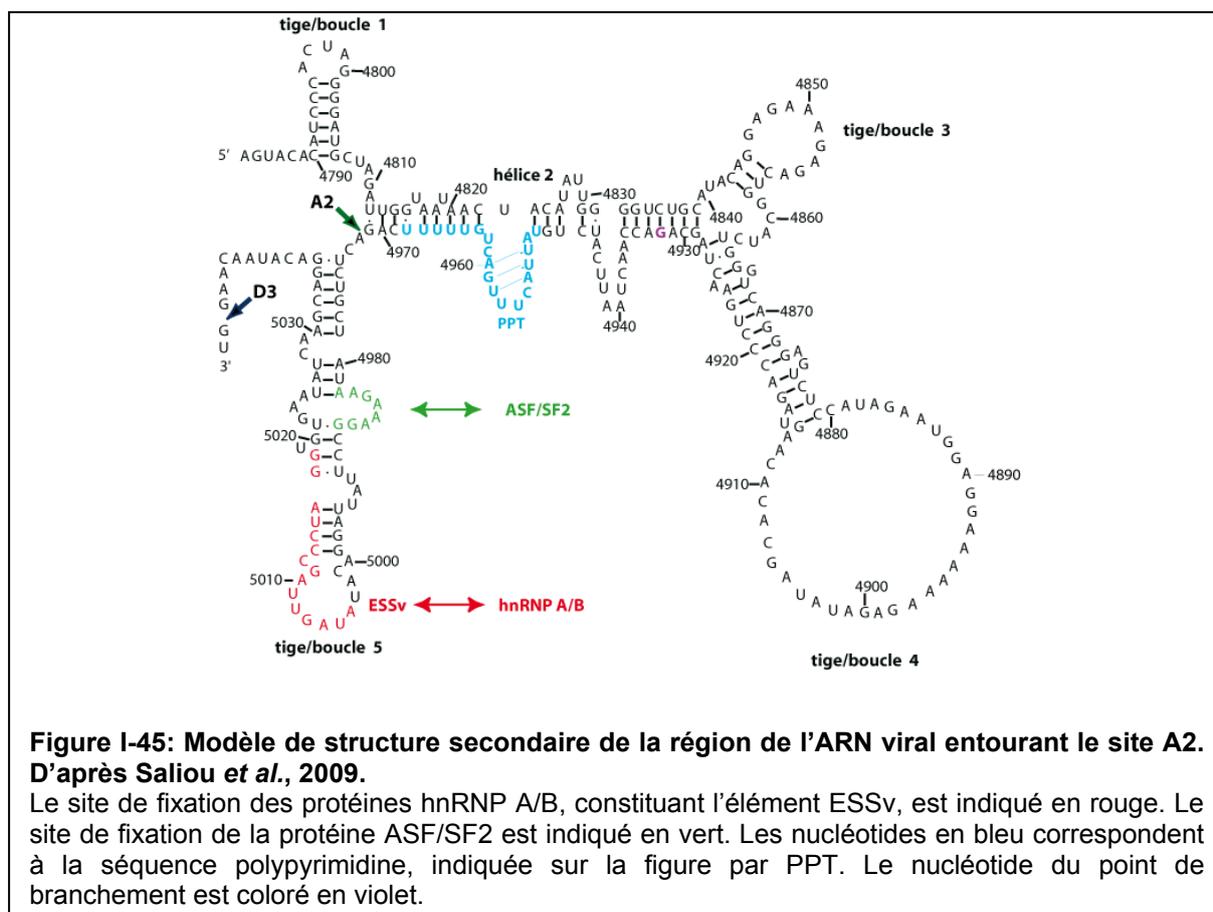
#### *Le site 3' accepteur d'épissage A1*

L'augmentation de l'épissage au site 3' d'épissage A1 conduit à l'accumulation de l'ARNm *vif* et augmente l'inclusion de l'exon 2 dans les espèces d'ARNm virales épissées. Un signal sous-optimal du site 5' d'épissage D2 en aval est nécessaire pour une utilisation adéquate du site 3' d'épissage, l'accumulation d'ARNm viral non épissé, l'expression de la protéine Gag et une production de virus efficace. De plus, une augmentation des niveaux d'ARNm *tat* est observée quand l'épissage est augmenté au site A1. La diminution de l'épissage au site A1 et la diminution de l'expression de Gag qui en découle est la cause primaire de déficience dans la production du virion chez les mutants du site A1 (Madsen *et al.*, 2006). En augmentant l'homologie du site 5' d'épissage D2 avec le signal consensus métazoaire 5', cela augmente considérablement l'efficacité d'épissage du site A1 (Madsen *et al.*, 2006).

**Le site 3' accepteur d'épissage A2**

Les ARNm *Vpr* sont peu abondants et incomplètement épissés et ils sont formés essentiellement par l'épissage du site 5' donneur D1 en combinaison avec le site accepteur 3' A2 (minoritairement D1-A1/D1-A2).

L'élément cis régulateur dominant au site 3' A2 est un ESS dans l'exon 3 appelé ESSv. L'ESSv est une séquence de 16 nucléotides située en aval du site A2 contenant des motifs PyUAG et un élément important AUAG. Ce motif AUAG se lie préférentiellement aux protéines membres de la famille des hnRNP A1 (Bilodeau *et al.*, 2001). Les protéines hnRNP A/B se lient à l'ESSv, oligomérisent le long de l'ARN et provoquent l'inhibition de la fixation de U2AF<sub>65</sub> à la séquence polypyrimidine du site (Domsic *et al.*, 2003). Au laboratoire, D. Ropers et L. Ayadi ont montré que la protéine SR ASF/SF2 active *in vitro* et *in vivo* l'épissage au site A2, tandis que les protéines SC35, SRp40 et 9G8 n'ont pas un effet aussi significatif (Ropers *et al.*, 2004). JM Saliou a réalisé la structure secondaire du site A2 (Figure I-45)



Cette structure montre que l'exon A2-D3 se situe dans une structure tige-boucle très conservée dans différentes souches de VIH-1 et le site A2 à la jonction entre 2 hélices (Saliou *et al.*, 2009).

L'équipe de J Tazi à Montpellier a identifié, par criblage d'une banque de molécules chimiques, des dérivés d'indole capables de bloquer spécifiquement l'action d'ASF/SF2. Certaines d'entre elles bloquent la multiplication du VIH-1 dans différents types cellulaires. Le dérivé indol IDC16 supprime spécifiquement l'activation par l'épissage par ASF/SF2 *in vivo* et *in vitro* et seul un large excès d'ASF/SF2 recombinante permet de retrouver l'activation du site A2 en présence d'IDC16 (Bakkour *et al.*, 2007).

Quand des mutations dans l'élément ESSv sont générées dans le virus entier pNL4-3, la réplication est réduite de 95% et l'inclusion de l'exon 3 est augmentée comme attendu après l'augmentation de l'activité de A2 (Madsen *et al.*, 2005).

Une diminution des taux d'ARNm de *tat* est observée quand le taux d'épissage au site A2 augmente. Ces observations suggèrent que la différence dans l'accumulation de l'ARNm *tat* et l'accumulation totale de l'ARNm viral chez les mutants, qui ont un site D2 optimisé et ainsi des sites accepteurs d'épissage A1 et A2 plus efficaces, peut être la conséquence d'un épissage plus efficace des sites 5' D2 et A3 que D3 et A3 (Madsen *et al.*, 2006). L'augmentation de l'épissage de l'ARNm viral chez ces mutants peut résulter dans l'augmentation de la biogénèse des miRNA codés par le virus et dérivés de séquences introniques virales épissées (Madsen *et al.*, 2006).

### ***Le site 3' accepteur d'épissage A3***

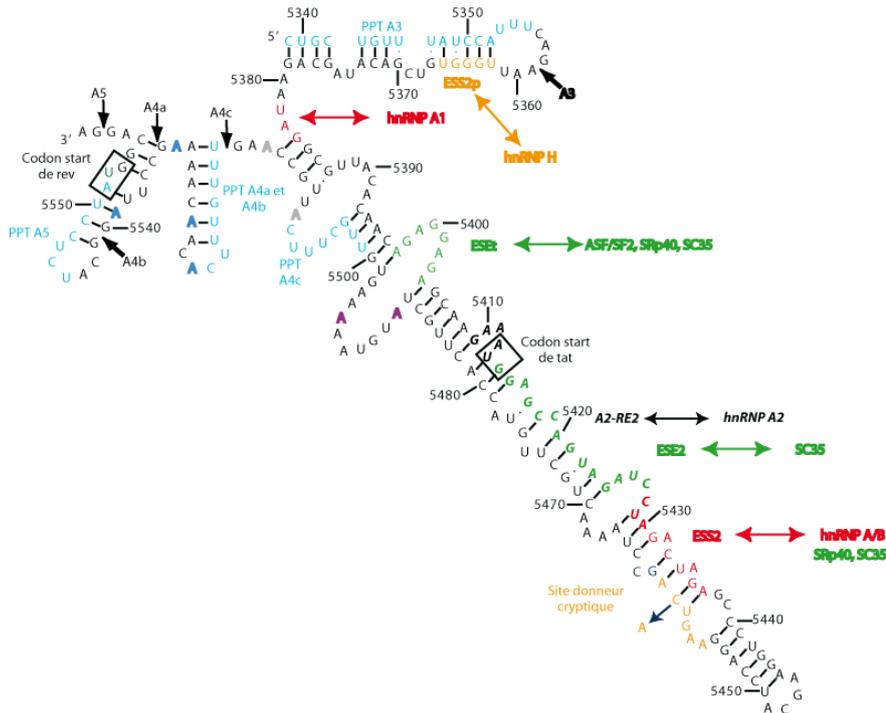
Le site A3 est exclusivement utilisé pour la production de l'ARNm *tat* en combinaison avec les sites donneurs 5' D1, D2, ou D3. Deux éléments inhibiteurs d'épissage limitent l'utilisation du site A3 d'épissage, les éléments ESS2 et ESS2p. L'élément ESS2 est situé 70 nucléotides en amont du site A3 et fixe hnRNP A1 (Caputi *et al.*, 1999; Amendt *et al.*, 1995). L'élément ESS2p, de séquence UUGGGU a été découvert au laboratoire par S. Jacquenet et A. Méreau; il est situé très près du site A3 qui fixe la protéine hnRNP H (Jacquenot *et al.*, 2001).

Le site A3 est aussi régulé par un élément ESE activateur d'épissage, l'élément ESE2, qui est activé par la fixation de la protéine SC35. ESE2 active l'épissage *in vitro* au site accepteur 3' A3 (Zahler *et al.*, 2004).

Les études structurales montrent que le site A3 est situé dans la boucle terminale d'une structure tige-boucle (Saliou *et al.*, 2009). Au laboratoire, S. Jacquenet a réalisé la structure secondaire de cette tige-boucle (Figure I-46) et montré que la déstabilisation de cette tige-boucle, qui rend la séquence polypyrimidine plus accessible aux facteurs spliceosomaux, augmente l'efficacité d'épissage au site A3 *in vitro* et *ex vivo* (Jacquenot *et al.*, 2001). Comme pour le site A2, l'efficacité de fixation de U2AF<sub>65</sub> à la séquence polypyrimidine du site A3 doit dépendre de l'association de U2AF<sub>35</sub> avec la jonction exon-intron. La fixation de hnRNP H à cet élément peut limiter l'association U2AF<sub>35</sub> au site A3 par encombrement stérique, expliquant comment ESS2p décroît l'activité du site A3 (Jacquenot *et al.*, 2001).

Des études moléculaires associées à des analyses RMN ont permis d'expliquer que la liaison de hnRNP A1 à l'ESS2 inhibe fortement l'activité du site A3, malgré sa localisation éloignée en aval. L'ESS2 est situé dans la partie apicale d'une longue structure irrégulière tige-boucle et joue un rôle dans l'association de hnRNP A1 avec la région SLS3-A3 (Hallay *et al.*, 2006).

Le site A3 est utilisé dans les cellules infectées (Purcell *et al.*, 1993). Son activité est fortement augmentée *in vivo* et *in vitro* par les protéines SR SC35 et SRp40 (Ropers *et al.*, 2004) qui entrent en compétition avec la protéine hnRNP A1 pour la fixation sur les sites ESE2 et ESS2 (Hallay *et al.*, 2006).



**Figure I-46: Modèle de structure secondaire de la région de l'ARN viral entourant le site A3 (Jacquet et al., 2001).**

Les 2 éléments régulateurs ESS2 et ESS2p sont indiqués respectivement, en rouge et en orange (Amendt et al., 1994; Amendt et al., 1995; Si et al., 1997; Del Gatto-Konczak et al., 1999; Caputi et al., 1999). Les éléments activateurs ESEt et ESE2 sont indiqués en vert. L'élément ESEt fixe les protéines SC35, SRp40 et ASF/SF2 tandis que l'élément ESE2 fixe la protéine SC35 (Zahler et al., 2004). Les nucléotides en bleu correspondent aux séquences polypyrimidines indiquées sur la figure par PPT. Les points de branchements sont représentés par des nucléotides colorés en violet pour A4c, en gris pour A4a et b et en bleu, pour A5 (Swanson et al., 1998). Les nucléotides en orange correspondent au site donneur cryptique (Verhoef et al., 2001) et les nucléotides noirs indiqués en italique et en gras constituent l'élément A2RE-2 reconnu par la protéine hnRNP A2 (Mouland et al., 2001).

### **Les sites 3' accepteurs d'épissage A4c, a, b et A5**

Les quatre sites 3' accepteurs A4c, A4a, A4b et A5 se trouvent au milieu du génome VIH-1 et sont utilisés pour produire les ARNm *rev* complètement épissés (A4c, a, b), les ARNm *nef* complètement épissés (A5) et les ARNm *env/vpu* incomplètement épissés.

L'épissage dans la région de 40 nucléotides contenant les sites 3' A4c, a, b et A5 implique l'utilisation de différents points de branchement pour l'épissage de chacun des 4 sites accepteurs d'épissage alternatifs. On observe un chevauchement entre les éléments d'épissage

*cis*; les points de branchement utilisés pour l'épissage au site A5 chevauchent les nucléotides AG utilisés pour l'épissage aux sites A4a et A4b (Swanson *et al.*, 1998). La compétition des sites chevauchants pour les différents facteurs d'épissage peut jouer un rôle dans la maintenance d'un équilibre idéal des quantités d'ARNm dans les cellules infectées par VIH-1. En effet, les mutations du dinucléotide AG au site *rev* A4b augmentent fortement l'épissage du site A5 *env/nef* (Swanson *et al.*, 1998; Purcell *et al.*, 1993; Riggs *et al.*, 1994). La liaison des facteurs au ou près du site A4b pour *rev* et la reconnaissance de la séquence dinucléotidique AG peut en effet interférer stériquement avec la formation des spliceosomes au site A5 en bloquant l'accès de la séquence de branchement (Swanson *et al.*, 1998).

### ***Le site 3' accepteur d'épissage A6***

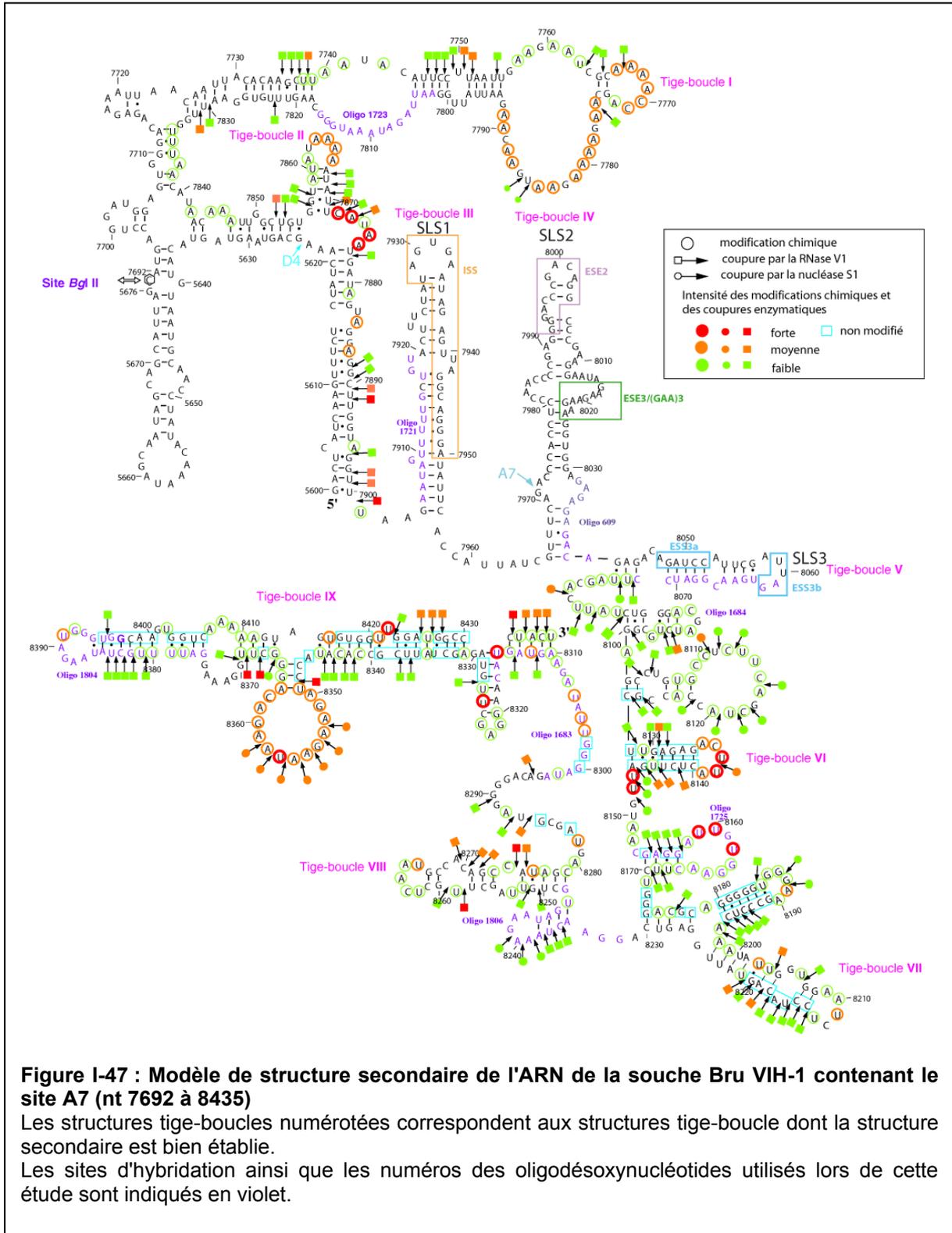
Certaines souches virales de VIH-1 comme le variant HXB2 possèdent des sites d'épissage supplémentaires, les sites d'épissage A6 et D5. Le site A6 est localisé dans l'exon *env* et constitue, avec le site D5, l'exon 6D (Wentz *et al.*, 1997). L'exon 6D est incorporé dans les ARNm aberrants *env*, *tat*, *rev* et *nef*, que l'on ne retrouve pas dans les souches Bru ou pNL4-3. La présence de l'exon 6D chez le virus sauvage montre la production de faible taux de protéine chimère des protéines Tat et Rev, la protéine Tev (Benko *et al.*, 1990). Dans l'isolat mutant, l'ARNm qui inclut l'exon 6D domine la population d'ARNm avec la déplétion concomitante de l'ARN normal, provoquant l'arrêt de la production du virus. L'incorporation de l'exon 6D dans l'ARN messager *rev* multiplement épissé provoque une diminution significative de l'expression de la protéine Rev fonctionnelle, ce qui peut être partiellement responsable de la réduction de la réplication du virus mutant. L'exon 6D contient un élément de régulation similaire à celui observé dans l'exon 3 de *tat*, un ESS potentiel juxtaposé sur une séquence riche en purine similaire à des ESE déjà décrits. En l'absence d'un site 5' flanquant, une mutation dans l'ESS dans l'exon 6D active fortement l'épissage engendrant *env*, suggérant que cet ESS putatif joue un rôle dans la limitation de cet épissage.

Une mutation ponctuelle U en C de la souche de VIH-1 HXB2 dans l'exon 6D provoque une importante augmentation de l'inclusion de cet exon (Wentz *et al.*, 1997). Caputi et Zahler montrent que des mutations U en C, U en A et U en G augmentent toutes la fixation de protéines SR à l'exon 6D. Ces séquences mutantes fonctionnent comme un ESE dépendant de SC35 quand elles sont testées en épissage *in vitro* sur des substrats hétérologues *dsx*. De plus, dans le contexte de mutation U en C, les membres de la famille hnRNP H se lient mieux à une

séquence GGGA trois nucléotides en aval de la mutation. Cette fixation est essentielle pour l'inclusion de l'exon 6D dans les ARNm VIH-1. Un troisième élément situé davantage en aval de l'exon 6D et riche en polypurines affecte l'inclusion de l'exon 6D quand ce U est muté en C. Paradoxalement, cet élément stimule la fixation des protéines SR dans un extrait nucléaire de cellules HeLa mais est aussi comme un ESS quand il est présent dans un substrat d'épissage *in vitro dsx* (Caputi *et al.*, 2002; pour revue, Martin Stoltzfus, 2009).

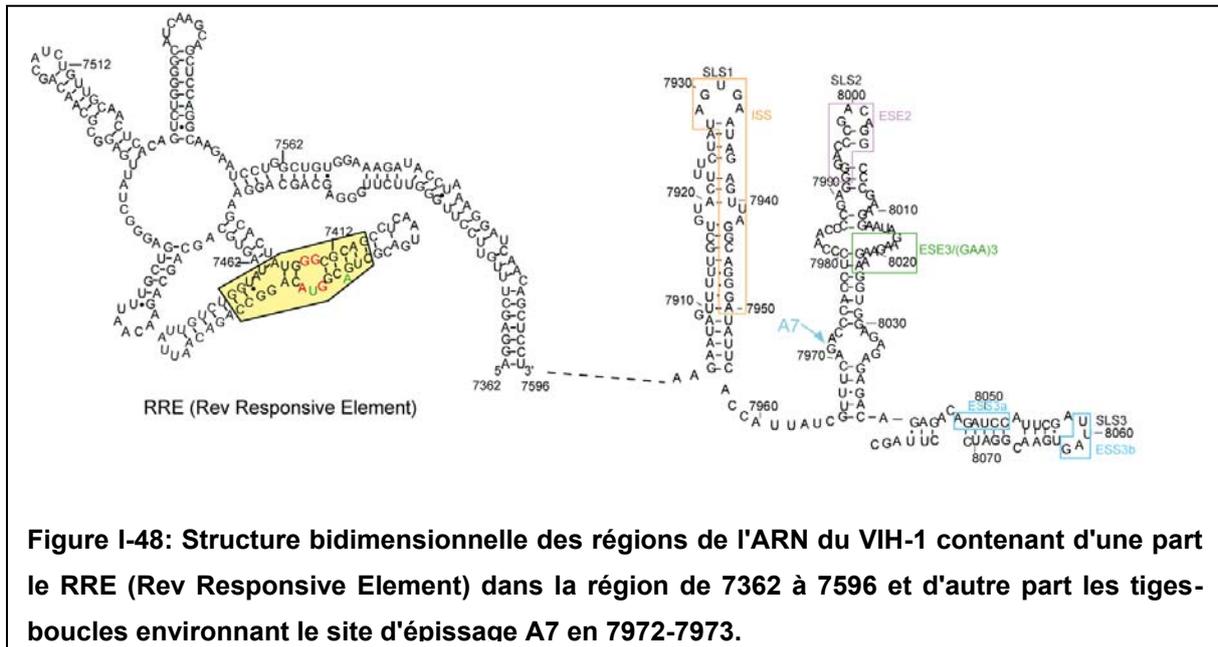
### ***Le site 3' accepteur d'épissage A7***

Plusieurs éléments régulant l'utilisation du site A7 ont été mis en évidence par d'autres équipes lorsque l'étude de ce site a été abordée au laboratoire. Un élément inhibiteur fort ESS3 est situé en aval du site A7 (Amendt *et al.*, 1995; Staffa *et al.*, 1995) et l'équipe de Krainer avait montré que son action est régulée par hnRNP A1 (Zhu *et al.*, 2001). L'équipe de Kjems avait mis en évidence un second élément inhibiteur ISS, localisé dans l'intron en amont du site A7 et montré que dans ce cas aussi hnRNP A1 est responsable de son effet inhibiteur (Tange *et al.*, 2001; Tange *et al.*, 2001). Enfin, l'équipe de Cochrane avait proposé que la séquence GAAGAAGAA (ESE3) située en aval du site A7 fixe ASF/SF2 et soit un élément activateur de ce site (Staffa *et al.*, 1995). Néanmoins, il était difficile de comprendre quel était le mécanisme global de régulation. Dans un premier temps, la structure 2D de la région de l'ARN viral contenant le site A7 a été déterminée (Marchand *et al.*, 2002) (Figure I-47). Chacun des trois éléments régulateurs (ISS, ESE et ESS) est présent dans une structure tige boucle distincte. Par ailleurs, la séquence ESE3 supposée activatrice fixe hnRNP A1 et elle est nécessaire à la fixation de cette protéine sur les éléments inhibiteurs ESS3 et ISS. La séquence UGAAGAAGAA constitue le site fort de fixation de hnRNP A1, ce qui permet par un phénomène de coopérativité sa fixation aux éléments ISS et ESS3. HnRNP A1 ne se lie pas du tout à la structure tige boucle SLS3-A7 prise individuellement. La fixation de hnRNP A1 à la séquence UGAAGAAGAA permet aussi sa liaison à la séquence polypurine qui relie ESE3 à ESS3. L'ESE3 est donc un élément régulateur ambivalent suivant la concentration relative en protéine hnRNP A1 et en protéine ASF/SF2, il peut être activateur en présence d'une forte concentration en ASF/SF2, ou être un inhibiteur extrêmement



puissant bloquant l'accès à tout facteur d'épissage lorsque la concentration en protéine hnRNP A1 est élevée. Il s'agissait d'un nouveau concept dans le domaine de la régulation de l'épissage (Marchand *et al.*, 2002).

La production des ARNm *tat* et *rev* nécessite respectivement l'utilisation des sites A3 et A7 (ARNm *tat*), A4a, A4b ou A4c et obligatoirement A7 (ARNm *rev*) (Purcell *et al.*, 1993). Lorsque la concentration nucléaire en protéine Rev devient suffisante, elle se lie et multimérise sur le segment de l'ARN viral RRE (pour « Rev Response Element ») (Malim *et al.*, 1990). Cet élément est situé dans l'intron D4-A7, 392 nucléotides en amont du site A7 (Fig. I-48).



La protéine Rev recrute le complexe contenant l'exportine, CRM1, ce qui permet un transport efficace des ARN non épissés entre D4 et A7 et aussi des ARN intacts dans le cytoplasme (Bogerd *et al.*, 1998). Les précurseurs protéiques Gag, Gag-Pol, Env et les protéines auxiliaires Vpu, Vpr et Vif sont alors produites. En parallèle, la protéine Tat stimule la transcription de l'ADNc proviral intégré dans le génome de la cellule hôte, en se fixant à l'extrémité 5' de l'ARN viral en cours de transcription (ARN TAR), ce qui permet le recrutement du facteur pTefb qui phosphoryle le domaine CTD de la grande sous unité de l'ARN Pol II (Garber *et al.*, 1998). Dans cette deuxième phase du cycle de multiplication virale, tous les composants requis pour la formation de nouveaux virions sont synthétisés et les virions sont produits par bourgeonnement. Les protéines Tat et Rev jouent donc un rôle

clef dans le cycle viral. D'où en conséquence, l'importance des sites accepteurs d'épissage A3 et A7, qui sont requis pour la production des ARNm *tat* et des sites A4a, A4b ou A4c et A7 utilisés pour la production des ARNm *rev*. La région de l'ARN viral qui contient le site A7 semble à la fois intervenir dans la régulation de l'épissage du site A7, mais aussi dans l'équilibre entre épissage et export de l'ARN non épissé (Staffa *et al.*, 1994). En effet, elle est située seulement 392 nucléotides en aval de l'élément RRE (Fig. I-48) et les facteurs nucléaires se liant à la région de l'ARN viral contenant le site A7 pourraient avoir des rôles modulateurs, positifs ou négatifs suivant leur identité, sur l'activation du transport de l'ARN viral par la protéine Rev.

L'équipe de Weeks (2009) a réalisé la structure de l'ARN entier de VIH-1 extrait de cellules infectées grâce à la technique de SHAPE (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension) (Watts *et al.*, 2009). Ce modèle est un point de départ pour l'étude des interactions longues distances des brins ARN. Les structures entières de la région 3' avec le site d'épissage A7 indiqué en rouge et les débuts et fin de traduction sont représentés dans la figure I-49.

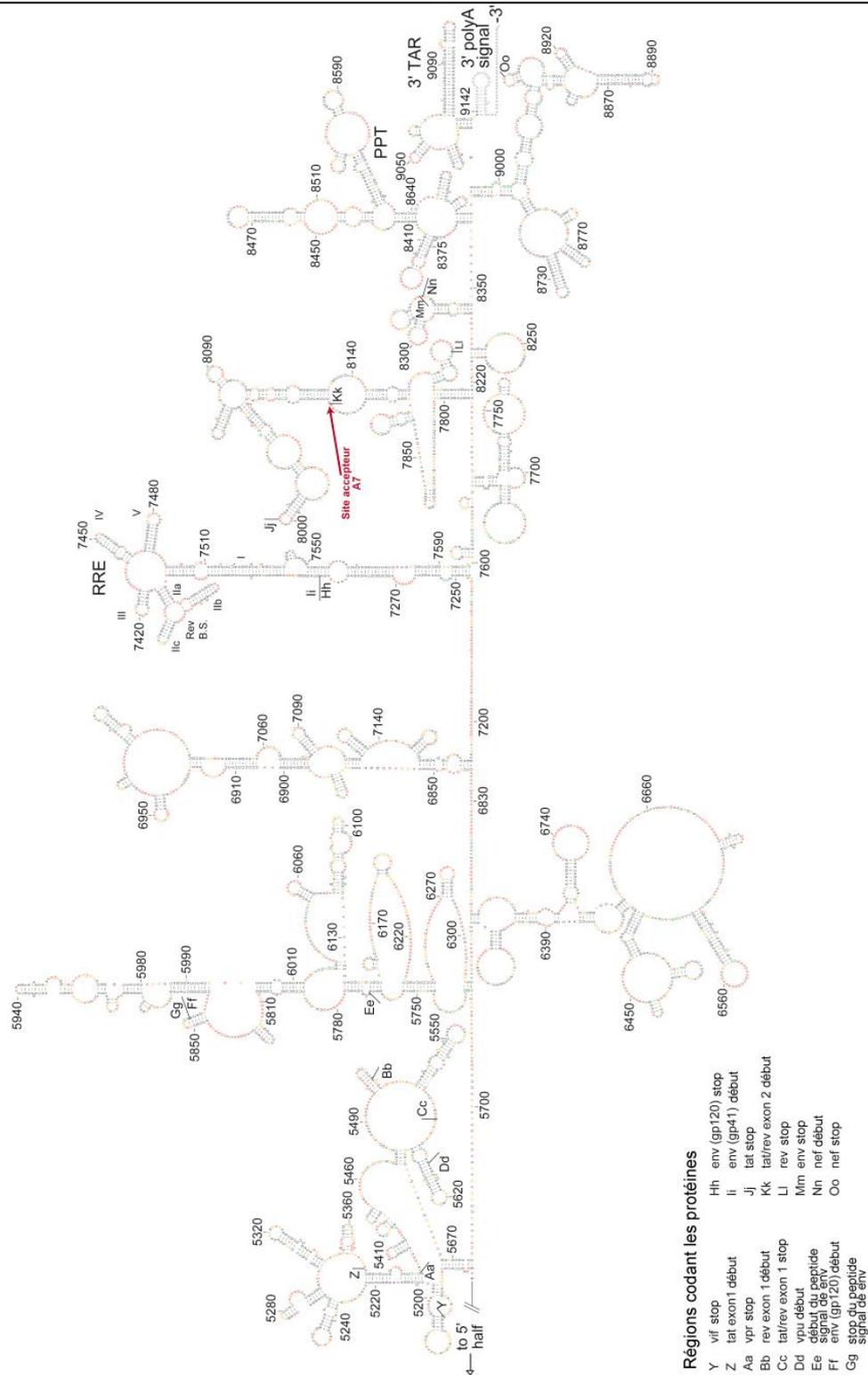


Figure I-49: Structure du génome ARN de pNL4.3 (Extrémité 3'). D'après Watts *et al.*, 2009.

# **OBJECTIFS**



## OBJECTIFS

Comme nous l'avons vu dans la partie Introduction, la régulation de l'épissage alternatif de l'ARN du virus VIH-1 est essentielle pour une multiplication efficace du virus. La compréhension des mécanismes de régulation devrait permettre de maîtriser l'étape d'épissage du virus et de pouvoir ainsi développer des approches thérapeutiques ayant pour cibles l'épissage. Comme nous venons de le décrire, seules les protéines SR ASF/SF2 activatrice de l'épissage et hnRNPA1 inhibitrice de l'épissage ont été montrées intervenant dans la régulation du site A7. Cependant, il était probable que d'autres protéines encore non identifiées qui jouent un rôle essentiel pour la production des protéines Tat et Rev interviennent dans la régulation de son utilisation. Par ailleurs comme nous l'avons vu, cette région de l'ARN du VIH-1 peut jouer d'autres rôles en particulier au niveau de la régulation de l'activité de Rev.

Afin d'identifier de nouvelles protéines pouvant se fixer à cette région de l'ARN viral, notre équipe a collaboré avec celle du Pr Lührmann du Max Planck Institut de Goettingen. Ce groupe avait mis au point une technique de purification et d'analyse par spectrométrie de masse de complexes spliceosomaux formés sur des mini ARN pré-messagers contenant un intron épissé de manière constitutive. Avec leurs conseils, nous avons entrepris de développer une approche permettant d'identifier l'ensemble des partenaires protéiques potentiels de la région de VIH-1 contenant le site d'épissage A7 et ses éléments de régulation.

**Le premier objectif de la thèse** était donc dans un **premier temps**, avec V Marchand, C Aigueperse et I Motorine qui avaient initié les travaux, de participer à la mise au point de cette approche et à son application à l'étude du site A7. Une centaine de protéines ont été détectées par spectrométrie de masse en collaboration avec S. Sanglier du laboratoire d'A. Vandorsselaer. V. Marchand avait pu proposer d'après des données de la littérature que certaines d'entre elles puissent interagir avec l'ARN. J'ai testé expérimentalement ces hypothèses par différentes approches, ce qui m'a permis de montrer une interaction directe entre l'ARN SLS2 et les protéines nucléoline, hnRNP H et hnRNP K. Ceci a nécessité le clonage et la production de certaines de ces protéines sous forme recombinante. De nombreux variants de l'ARN SLS2 disponibles au laboratoire et des expériences d'empreintes nous ont permis de caractériser plus précisément les sites de fixation de ces protéines sur l'ARN SLS2-A7.

Dans un **dernier temps** nous avons testé l'effet potentiel sur l'épissage au site A7 de ces protéines. J'ai mis en évidence un effet négatif de la protéine hnRNP K au site A7 en testant l'effet de la protéine hnRNP K recombinante sur l'épissage *in vitro* d'un mini ARN contenant

l'intron D4-A7 dans un extrait nucléaire de cellules HeLa. Pour comprendre comment hnRNP K pouvait avoir cet effet, nous avons testé si la fixation de hnRNP A1 et de hnRNP K au site A7 pouvait se faire de manière synergique.

Afin de confirmer la signification biologique des effets trouvés *in vitro*, le **second objectif de la thèse** a été de valider ces résultats par des expériences d'épissage *in cellulo* en collaboration avec C. Aigueperse. Pour cela j'ai testé l'effet de la surexpression et de la diminution du taux cellulaire de protéine hnRNP K dans les cellules HeLa sur l'épissage de l'ARN  $\Delta$ PSP contenant un ARN viral VIH-1 tronqué possédant tous les sites d'épissage. Ayant observé un effet très marqué de la protéine hnRNP K sur le profil d'épissage de l'ARN  $\Delta$ PSP, j'ai évalué l'effet de la surexpression de la protéine hnRNP K dans les cellules HeLa sur la concentration cellulaire de la protéine virale Nef.

J'ai alors testé en collaboration avec L. Ayadi l'effet de l'ajout de protéine hnRNP K recombinante dans un extrait nucléaire de cellules HeLa sur l'épissage de différents mini pré-ARNm du VIH-1 contenant le site donneur D1 et l'un des trois sites accepteurs A1, A2 ou A3. J'ai ainsi pu mettre en évidence en effet activateur de hnRNP K sur chacun de ces sites et tout particulièrement pour le site A2. On savait que ce site est activé par ASF/SF2. J'ai montré un effet activateur synergique des protéines ASF/SF2 et hnRNP K au site A2.

Dans le **dernier volet** de mon travail de thèse, j'ai participé à une collaboration avec Jean-Christophe Paillard et Lucile Sinck du Laboratoire Architecture et Réactivité de l'ARN CNRS-UPR 9002 à Strasbourg. Le génome du virus de l'immunodéficience humaine de type 1 (VIH-1) est constitué de deux molécules d'ARN homologues de polarité positive. La nature dimérique de l'ARN génomique (ARNg) est conservée à l'intérieur de la famille des Retroviridae, indiquant un rôle crucial de la dimérisation dans la réplication virale. En effet, la dimérisation de l'ARNg est un pré-requis à son encapsidation et offre au virus une opportunité unique de recombinaison lors de l'étape de rétrotranscription, lui permettant ainsi d'accroître sa variabilité. Certains de ces signaux d'encapsidation, et en particulier le site d'initiation de la dimérisation (DIS), sont présents aussi bien sur les ARNg que sur les ARN viraux épissés du VIH-1, ce qui suggère que le virus a dû développer des mécanismes précis afin de sélectionner spécifiquement l'ARNg aux dépens des ARN épissés.

Notre étude avait pour but de déterminer si la dimérisation de l'ARN inhibe l'épissage expliquant pourquoi les ARN encapsidés sont majoritairement non épissés.

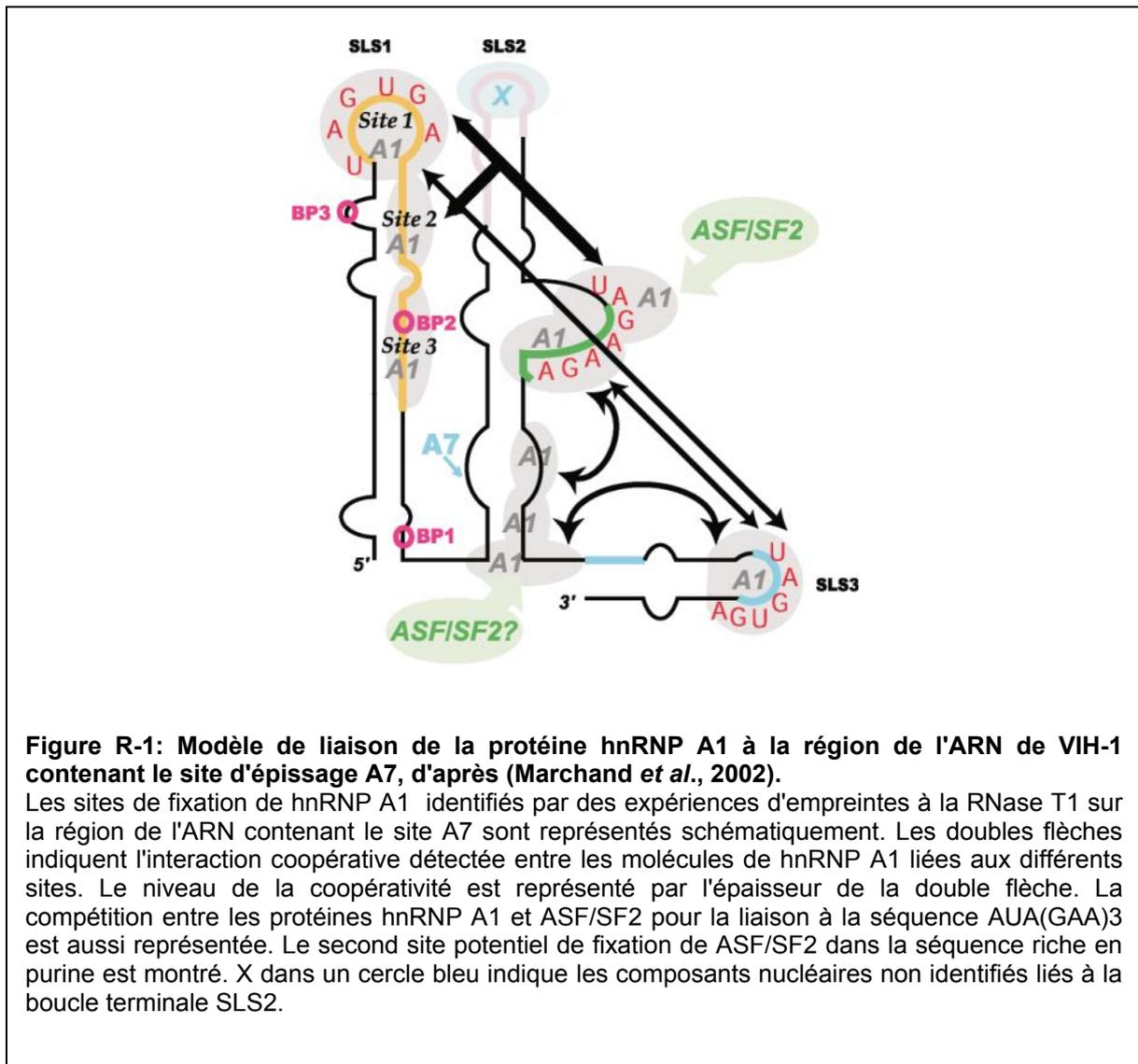
# **RESULTATS ET DISCUSSION**



**RESULTATS :**  
**CHAPITRE I Recherche de nouvelles protéines régulant l'épissage alternatif au site accepteur d'épissage A7.**

**I Stratégie utilisée pour l'étude**

Le modèle de régulation du site A7 par les protéines ASF/SF2 et hnRNP A1 connu aujourd'hui est certainement incomplet (Figure R-1) (Marchand *et al.*, 2002). En effet, d'autres protéines qui se lient à cette région pourraient également influencer l'efficacité d'utilisation de ce site.



Jusque là, la majorité des études réalisées sur les régulations des sites d'épissage étaient largement basées sur la recherche d'un effet possible de protéines connues pour influencer l'épissage alternatif, telles que les protéines hnRNP A1 et H et les protéines SR. Les

recherches des partenaires protéiques d'éléments inhibiteurs ou activateurs ont été essentiellement réalisées avec de très petits fragments d'ARN.

Pour essayer d'identifier l'ensemble des protéines d'un extrait nucléaire de cellules HeLa capables de se lier directement ou indirectement (par interactions protéine-protéine) à la structure tige-boucle 2, I. Motorine et V. Marchand ont décidé d'appliquer une technique qui venait d'être mise au point par K. Hartmuth dans le laboratoire de R. Lührmann (Max Planck Institute, Göttingen). J'ai participé à la mise au point de cette approche et à son application à l'étude du site A7 avec V. Marchand, C. Aigueperse et I. Motorine.

Le laboratoire du Pr R. Lührmann étudie les intermédiaires de la voie d'assemblage des spliceosomes. Leur équipe a mis au point une technique de purification et d'analyse par spectrométrie de masse de complexes spliceosomaux formés sur des mini ARN pré-messagers contenant un intron épissé de manière constitutive. Cette approche utilisant le couple « protéine de l'enveloppe du phage MS2-site de liaison de cette protéine sur l'ARN de ce phage » a été développée au laboratoire. Le principe de cette méthode est de greffer lors de la transcription une séquence additionnelle à l'ARN d'intérêt dont on recherche les partenaires. Cette séquence ARN MS2 contient plusieurs sites de fixation de la protéine de l'enveloppe du phage MS2, ce qui permet son immobilisation sur une résine. La protéine de l'enveloppe du phage MS2 est produite sous forme de fusion avec une protéine de liaison du maltose (MBP ou « Maltose Binding Protein »), formant la protéine MS2-MBP. Elle est fixée sur des billes d'amylose (Jurica *et al.*, 2002). L'affinité entre les deux partenaires ARN MS2/protéine MS2-MBP est très forte (Kd de l'ordre du nM) (Jurica *et al.*, 2002). Il est alors possible de former des complexes entre cet ARN et ses partenaires présents dans un extrait cellulaire, ceci en se plaçant dans des conditions de tampon identiques à celle utilisées pour la réalisation d'épissages *in vitro*.

## **II. Purification de complexes ARN SLS2-A7 / protéines d'extrait nucléaire de cellules HeLa.**

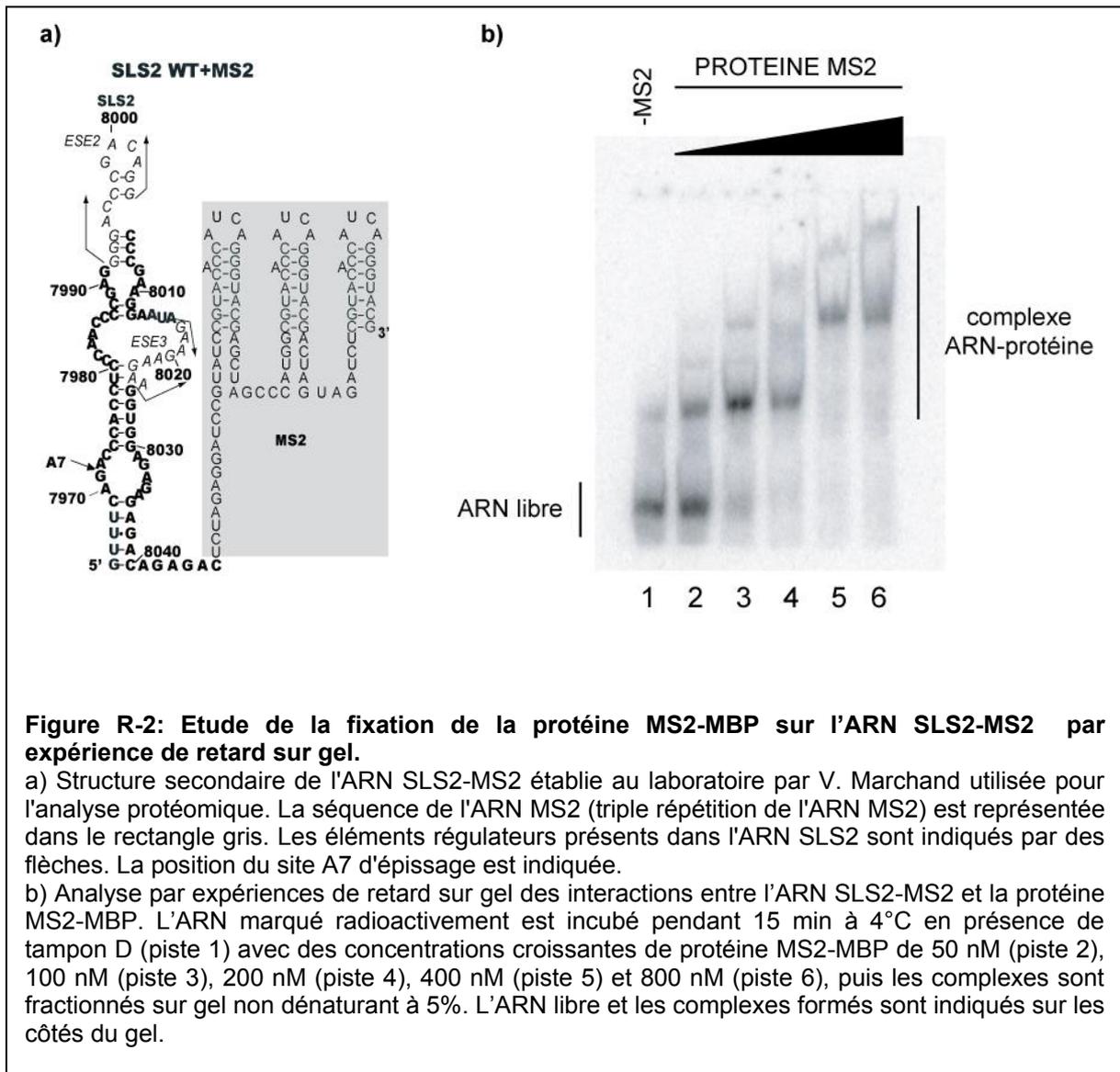
### **II.1 Composition de l'ARN hybride utilisé pour la chromatographie d'affinité.**

Notre but était d'identifier les protéines liées à la structure tige-boucle de l'ARN de VIH-1 Bru qui contient le site d'épissage A7 et les éléments régulateurs d'épissage ESE2 et ESE3.

Pour la méthode protéomique basée sur la chromatographie d'affinité avec la protéine MS2 de l'enveloppe de phage, les trois tiges-boucles ARN sites de liaison de la protéine MS2 de

l'enveloppe de phage sont fusionnées à l'extrémité 3' de l'ARN SLS2-A7 (ARN SLS2-A7-MS2). L'ARN MS2 forme une structure stable (Figure R-2) qui ne doit pas interférer avec la structure 2D du fragment d'ARN viral.

Pour vérifier la bonne fixation de la protéine MS2-MBP sur l'ARN SLS2-A7-MS2, une expérience de retard sur gel est réalisée et présentée en Figure R-2. Chaque palier observé au niveau des complexes ARN-protéine est représentatif de la fixation sur une tige-boucle de la protéine MS2-MBP. La fixation de la protéine sur l'ARN est totale dès 200 nM de protéine. On peut observer trois bandes qui correspondent à trois sites de fixation de la protéine MS2-MBP sur l'ARN MS2.



**Figure R-2: Etude de la fixation de la protéine MS2-MBP sur l'ARN SLS2-MS2 par expérience de retard sur gel.**

a) Structure secondaire de l'ARN SLS2-MS2 établie au laboratoire par V. Marchand utilisée pour l'analyse protéomique. La séquence de l'ARN MS2 (triple répétition de l'ARN MS2) est représentée dans le rectangle gris. Les éléments régulateurs présents dans l'ARN SLS2 sont indiqués par des flèches. La position du site A7 d'épissage est indiquée.

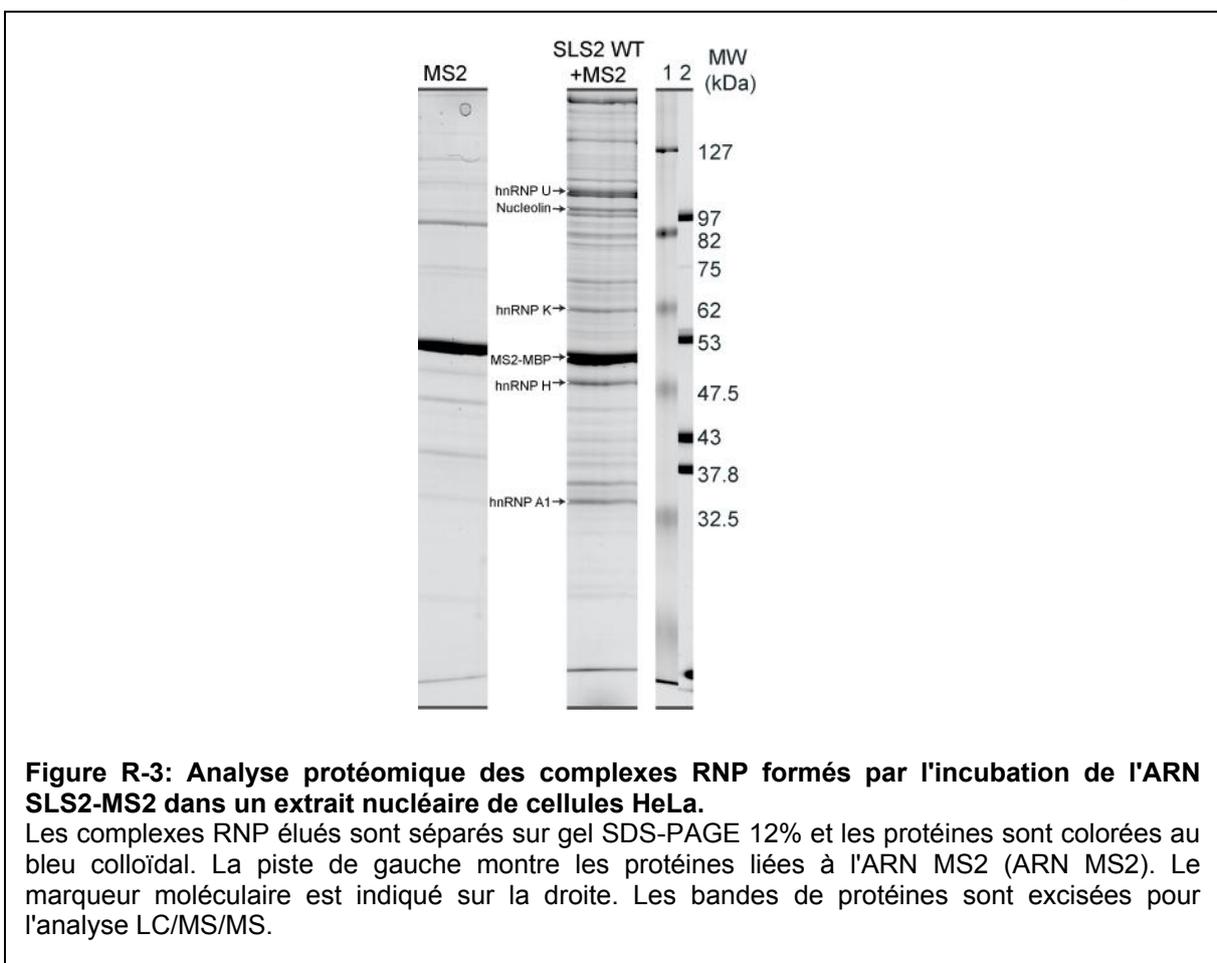
b) Analyse par expériences de retard sur gel des interactions entre l'ARN SLS2-MS2 et la protéine MS2-MBP. L'ARN marqué radioactivement est incubé pendant 15 min à 4°C en présence de tampon D (piste 1) avec des concentrations croissantes de protéine MS2-MBP de 50 nM (piste 2), 100 nM (piste 3), 200 nM (piste 4), 400 nM (piste 5) et 800 nM (piste 6), puis les complexes sont fractionnés sur gel non dénaturant à 5%. L'ARN libre et les complexes formés sont indiqués sur les côtés du gel.

## **II.2 Formation des complexes ribonucléoprotéiques (RNP) sur l'ARN fixé à la matrice et élution.**

Les complexes ARN-protéines sont formés par incubation avec les ARN hybrides dans un extrait nucléaire de cellules HeLa pendant 30 minutes. La protéine de fusion MS2-MBP liée aux complexes RNP permet la purification de ces complexes sur des billes d'amylose à travers l'interaction MBP-amylose. Les ratios protéine de fusion MS2-MBP/extrait nucléaire ont été optimisés. Comme la partie MBP assure la fixation de l'ARN sur les billes, après lavage, les complexes formés peuvent être élués par incubation avec du maltose. Un excès de maltose (12 mM) est utilisé pour l'élution.

De nombreuses expériences de contrôle ont permis de déterminer lesquelles des protéines retenues sur les billes sont spécifiquement associées avec la partie SLS2-A7 de l'ARN. La rétention possible des protéines sur la matrice d'amylose en absence d'ARN a été évaluée par incubation avec un extrait nucléaire en conditions utilisées pour la formation des complexes RNP mais sans association préalable avec un ARN MS2. Comme la rétention des protéines nucléaires était assez faible au pH utilisé pour les tests d'épissage *in vitro* (pH 7,9), les complexes SLS2-A7-MS2/MBP-MS2 ont été formés dans un extrait nucléaire de cellules HeLa dialysés contre le tampon D au pH 7,9. Les conditions d'incubation conduisant à une fixation efficace de l'ARN SLS2-A7-MS2 et une fixation limitée de l'ARN SLS2-A7 à la matrice d'amylose ont aussi été définies.

Enfin, il était nécessaire d'identifier les protéines nucléaires capables de lier les trois tiges-boucles de l'ARN MS2 seul. Ces protéines ont donc été fractionnées sur gel d'électrophorèse (Figure R-3) en parallèle avec les protéines retenues sur l'ARN de fusion SLS2-A7-MS2.



### II.3 Analyse en spectrométrie de masse des complexes ARN SLS2-A7/protéines d'extrait nucléaire de cellules HeLa.

Ces analyses par spectrométrie de masse ont été réalisées par S. Sanglier et L. Fouillen de l'équipe d'A. Van Dorsselaer à Strasbourg.

Les complexes formés avec les ARN hybrides contiennent un grand nombre de protéines par rapport aux complexes du contrôle avec l'ARN MS2. Au moins 50 bandes de protéines d'intensité variable sont détectées pour l'ARN hybride SLS2-A7-MS2 (Figure R-3), suggérant la formation de multiples complexes de différentes compositions. Dans une première série d'expériences, toutes les protéines clairement visibles dans le test et leurs pendants dans le contrôle (ARN MS2) sont numérotées, excisées du gel, digérées par la trypsine et analysées par spectrométrie de masse. Les protéines présentes dans ces bandes sont identifiées en combinant le « fingerprinting » de la masse des peptides par MALDI-TOF et confirmées par analyses nano LC-MS-MS. Les protéines identifiées sont comparées en Table 4.

Vingt protéines ont été trouvées dans les complexes formés avec l'ARN contrôle MS2. Certains de ces facteurs peuvent se lier non spécifiquement à la matrice amylose utilisée pour la purification des complexes ou montrer des capacités de liaison à l'ARN non spécifiques. Le profil des protéines fixées de façon non-spécifique à l'ARN MS2 est différent de celui observé avec l'ARN SLS2-A7-MS2.

Les protéines identifiées sur l'ARN SLS2-A7-MS2 peuvent être groupées en plusieurs grandes catégories. Comme attendu, nous avons trouvé un grand nombre de protéines hnRNP dans les complexes avec l'ARN SLS2-A7 (Table 4). D'autres protéines en lien avec l'épissage ont aussi été détectées dans les complexes avec l'ARN MS2. En plus des protéines liées à l'épissage, nous avons trouvé d'autres protéines liant l'ARN comme la nucléoline et des facteurs impliqués dans la transcription et la traduction. Un autre groupe est représenté par les protéines associées à la chromatine et des hélicases ADN/ARN. Certaines des protéines identifiées ne présentent pas de propriétés de fixation à l'ADN ou à l'ARN. Leur présence dans les complexes isolés peut être due aux interactions protéines-protéines avec des partenaires étroitement liés à l'ARN comme l'importine-5 ou la protéine Hsp 70.

	Protein name	Alternative Name	Accession numbers	Molecular weight (Da)	Unique peptide	
					MS2	SLS2-MS2
<b>hnRNP proteins</b>	Heterogeneous nuclear ribonucleoprotein A1		P09651	34,194	2	8
	Heterogeneous nuclear ribonucleoprotein A3		P51991	39,634		3
	Heterogeneous nuclear ribonucleoprotein D0		Q14103	38,337		4
	Heterogeneous nuclear ribonucleoprotein H		P31943	49,182		3
	Heterogeneous nuclear ribonucleoprotein K		P61978	50,961		8
	Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	O60506	69,616		2
	Heterogeneous nuclear ribonucleoprotein R		O43390	70,926		5
	Heterogeneous nuclear ribonucleoprotein U		Q00839	90,567		7
	Heterogeneous nuclear ribonucleoproteins A2/B1		P22626	37,460		7
<b>Nucleolin partners</b>	Nuclease-sensitive element-binding protein 1	Ybx1	P67809	35,712		2
	Heat shock 70 kDa protein 1	Hsp70	P08107	70,169		3
	Nucleolin		P19338	76,842		9
	Poly [ADP-ribose] polymerase 1	PARP1	P09874	113,070	3	7
<b>Splicing factors</b>	Splicing factor U2AF 35 kDa subunit		Q01081	27,797		2
	Splicing factor, proline- and glutamine-rich	PSF	P23246	75,424		2
<b>RNA helicases</b>	ATP-dependent RNA helicase A	DHX9, RHA	Q08211	140,903	6	7
	ATP-dependent RNA helicase DDX1		Q92499	82,415		6
	Probable ATP-dependent RNA helicase DDX5	p68	P17844	69,304		6
<b>Transcription factors</b>	Bromodomain adjacent to zinc finger domain protein 1A		Q9NRL2	178,689		3
	General transcription factor 3C polypeptide 3		Q9Y5Q9	101,258		2
	Probable global transcription activator SNF2L4		P51532	184,632		4
	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5		O60264	121,893		6
	Transcription intermediary factor 1-beta		Q13263	88,531		2
<b>Translation factors</b>	Elongation factor 1-alpha 1		P68104	50,167	5	5
	Selenocysteine-specific elongation factor		P57772	65,288		2
<b>DNA binding proteins</b>	ATP-dependent DNA helicase 2 subunit 1	XRCC6, Ku70	P12956	69,828		3
	ATP-dependent DNA helicase 2 subunit 2	XRCC5, Ku86	P13010	82,689		5
	DNA topoisomerase 2-alpha		P11388	174,373		2
	DNA-dependent protein kinase catalytic subunit	PRKDC	P78527	469,078		17
<b>Cell division proteins</b>	Anaphase-promoting complex subunit 7		Q9UJX3	63,117		2
	Sister chromatid cohesion protein PDS5 homolog A		Q29RF7	150,818		2
<b>Interleukins</b>	Interleukin enhancer-binding factor 2	ILF2, NF45	Q12905	43,045	11	9
	Interleukin enhancer-binding factor 3	ILF3, NF90	Q12906	95,321	7	7
<b>Other proteins</b>	E3 ubiquitin/ISG15 ligase TRIM25		Q14258	70,971		2
	E3 ubiquitin-protein ligase UHRF1		Q96T88	89,796		2
	Glycogen phosphorylase, liver form		P06737	97,134		3
	GTP-binding nuclear protein Ran		P62826	24,405		4
	Importin-5		O00410	123,614	3	
	Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	P19525	62,079	4	3
	Nucleophosmin		P06748	32,542	3	
	Ras GTPase-activating-like protein IQGAP1		P46940	189,241		3
UPF0568 protein C14orf166 homolog		Q9Y224	28,051		6	

Tableau 4: Protéines d'extraits nucléaires de cellules HeLa interagissant avec l'ARN SLS2-A7 isolées après purification MS2-MBP en spectrométrie de masse.

### III. DISCUSSION

#### **Purification de complexes ARN SLS2-A7/ protéines d'extrait nucléaire de cellules HeLa.**

Pour purifier les complexes RNP formés avec l'ARN de VIH-1 contenant le site 3' d'épissage A7 et deux éléments régulateurs d'épissage (ESE2 et ESE3), nous avons employé une méthode de purification sous des conditions faiblement stringentes. Cette chromatographie d'affinité en une étape ne permet pas d'éliminer des multiples protéines présentes dans les extraits nucléaires et pouvant se fixer de manière non spécifique. De telles contaminations peuvent résulter d'une haute affinité inattendue de certaines protéines cellulaires à l'amylose, ou encore plus probablement de leur affinité en général pour n'importe quel ARN. Ces deux propriétés leur permettent d'être co-purifiées avec les complexes spécifiques et d'être identifiées par spectrométrie de masse même si elles sont présentes en faibles quantités. De plus, des liaisons non spécifiques aux ARNm de nombreuses protéines cellulaires (dont les protéines hnRNP), en général très représentées dans les extraits nucléaires ont été démontrées. Ces protéines sont connues pour être des composants du complexe H assemblé autour de chaque séquence ARN de façon indépendante de l'épissage ou de la présence d'ATP (Bennett *et al.*, 1992; Jurica *et al.*, 2002).

Certaines protéines sont difficiles à mettre en évidence par la spectrométrie de masse LC.MS-MS, la protéine ASF/SF2 en est un exemple. La digestion par la trypsine n'est pas très adaptée pour l'identification d'ASF/SF2 car cette protéine contient peu de sites de digestion par cette enzyme ou des sites mal placés qui génèrent des peptides trop petits, difficilement identifiables en spectrométrie de masse. L'utilisation d'une enzyme de digestion plus appropriée permettrait de générer plus de peptides compatibles avec l'analyse de masse.

## **CHAPITRE II Etude de l'interaction des protéines identifiées par spectrométrie de masse avec la tige-boucle contenant le site d'épissage A7 du virus VIH-1.**

### **I. Objectifs de l'étude**

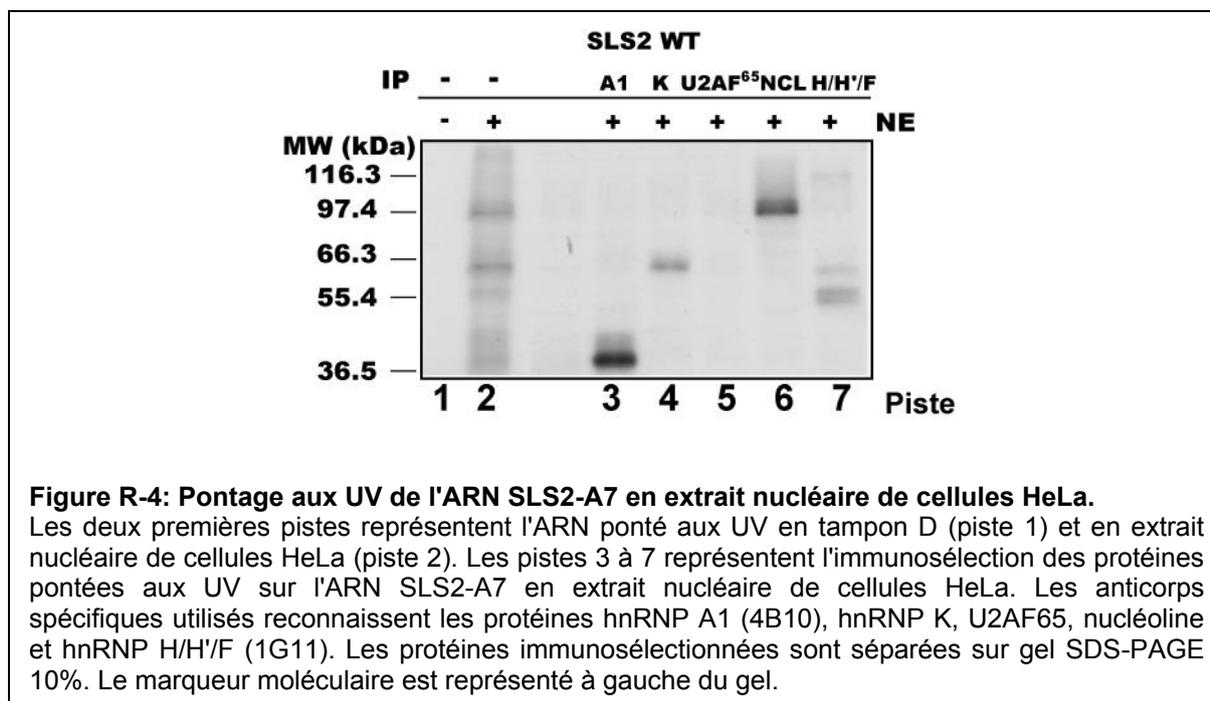
Un nombre important de protéines a été identifié en spectrométrie de masse dans les complexes protéiques interagissant avec l'ARN SLS2-A7. Cette population abondante peut s'expliquer par des associations indirectes de plusieurs de ces protéines entre elles à travers des interactions protéines-protéines. J'ai débuté les travaux visant à déterminer si certaines des protéines identifiées se liaient bien directement à la structure tige boucle SLS2-A7 grâce à une approche de pontage aux UV suivi d'immunosélection et à l'utilisation de la technique de retard sur gel.

La détermination des sites de fixation de ces protéines sur l'ARN a pour but de mieux comprendre leur rôle et notamment de comprendre comment elles peuvent influencer l'efficacité d'utilisation du site d'épissage.

### **II. Au moins quatre des protéines identifiées se lient directement à l'ARN SLS2-A7.**

Afin d'estimer le nombre de protéines directement liées à l'ARN SLS2-A7, j'ai transcrit radioactivement des ARN SLS2-A7 WT au  $[\alpha\text{-}^{32}\text{P}]$  CTP et au  $[\alpha\text{-}^{32}\text{P}]$  GTP puis utilisé ces ARN pour des expériences de pontages covalents aux UV après incubation dans un extrait nucléaire de cellules HeLa en tampon D pH 7,9. Cette technique permet de travailler dans un contexte où l'ARN est en contact avec toutes les protéines de l'extrait nucléaire en compétition les unes avec les autres. Une liaison covalente va se créer sous l'action des UV lorsque l'ARN marqué radioactivement est en contact direct avec la protéine. Les complexes ARN/extrait nucléaire sont exposés aux UV à une longueur d'onde de 254 nm (2 joules/cm<sup>2</sup>) puis dégradés par un mélange de RNase A et RNase T1 et enfin séparés sur un gel d'électrophorèse SDS-PAGE (Figure R-4, piste 1 et 2). J'ai ensuite immunosélectionné les protéines pontées covalamment à l'ARN SLS2-A7 (Figure R-4 pistes 3 à 7), en utilisant des anticorps dirigés contre les protéines hnRNP A1, hnRNP K/J, hnRNP H/H'/F ou la nucléoline. Ces protéines sont les plus représentées dans les fractionnements sur gel SDS-PAGE et des recherches sur leurs sites préférentiels de fixation sur les ARN ont montré qu'elles pouvaient potentiellement se fixer sur l'ARN SLS2-A7. J'ai aussi utilisé comme

contrôle négatif des anticorps contre le facteur d'épissage U2AF<sup>65</sup> qui lie la séquence polypyrimidine, qui n'est pas présente dans notre ARN SLS2-A7.



En accord avec les données précédentes (Marchand *et al.*, 2002; Damgaard *et al.*, 2002), hnRNP A1 a été identifiée comme l'une des protéines pontées covalamment (protéine de 38.5 kDa). Les immunosélections avec les autres anticorps montrent que les protéines pontées covalamment avec des poids moléculaires d'environ 42 et 52 kDa correspondent aux protéines hnRNP H/H' et F respectivement et les protéines de 62 et 100 kDa sont les protéines hnRNP K et nucléoline respectivement.

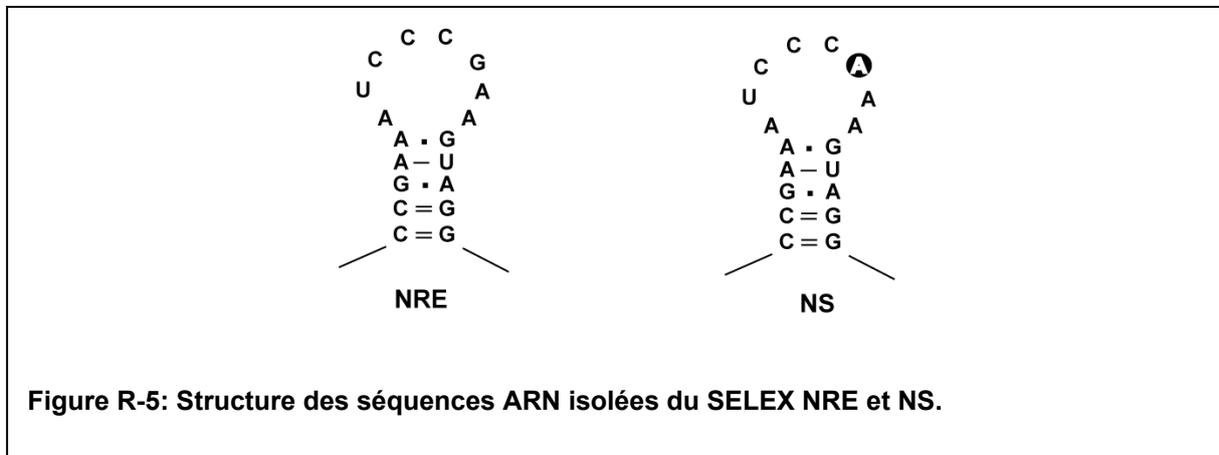
### III Sites potentiels de fixation des protéines nucléoline et hnRNP K identifiées sur l'ARN SLS2-A7.

#### III.1 La nucléoline :

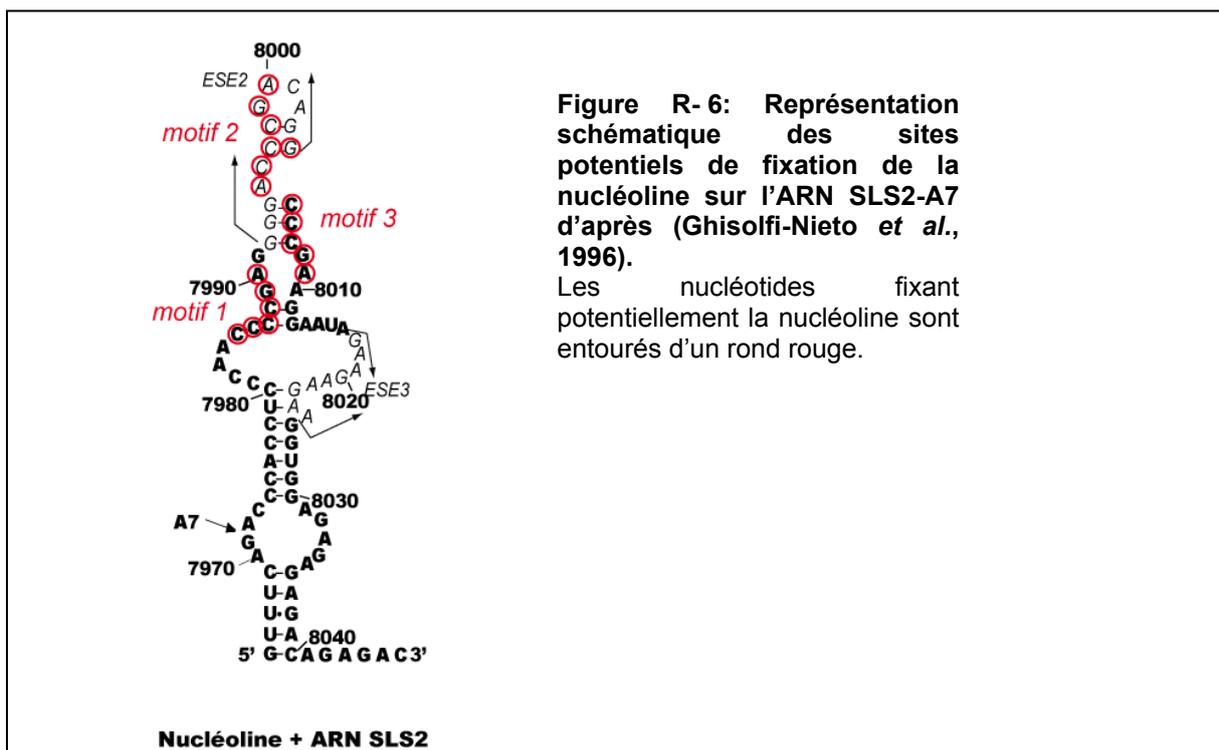
La nucléoline contient quatre motifs RRM de liaison à l'ARN et un domaine GAR qui peut faciliter la fixation des RRM à l'ARN. Deux différents types spécifiques de séquences ARN (séquences NRE pour Nucleolin Recognition Element et ECM pour Evolutionary Conserved Motif) ont été décrits comme substrat ARN pour la nucléoline (Ginisty *et al.*, 2001) (Figure R-5). Les deux premiers domaines RRM de la protéine sont suffisants pour permettre la fixation spécifique de la séquence NRE (5'-(U/G)CCCGA-3') (K<sub>D</sub> de 20 à 50 nM) (Bouvet *et*

*al.*, 2001; Ginisty *et al.*, 2001), tandis que la fixation de la séquence ECM (5'-UCGA-3') requiert la présence des quatre domaines RRM de la protéine.

La mutation d'un nucléotide G en A donnant l'ARN NS réduit drastiquement l'affinité de la nucléoline pour l'ARN (Serin *et al.*, 1997) (Figure R-5).

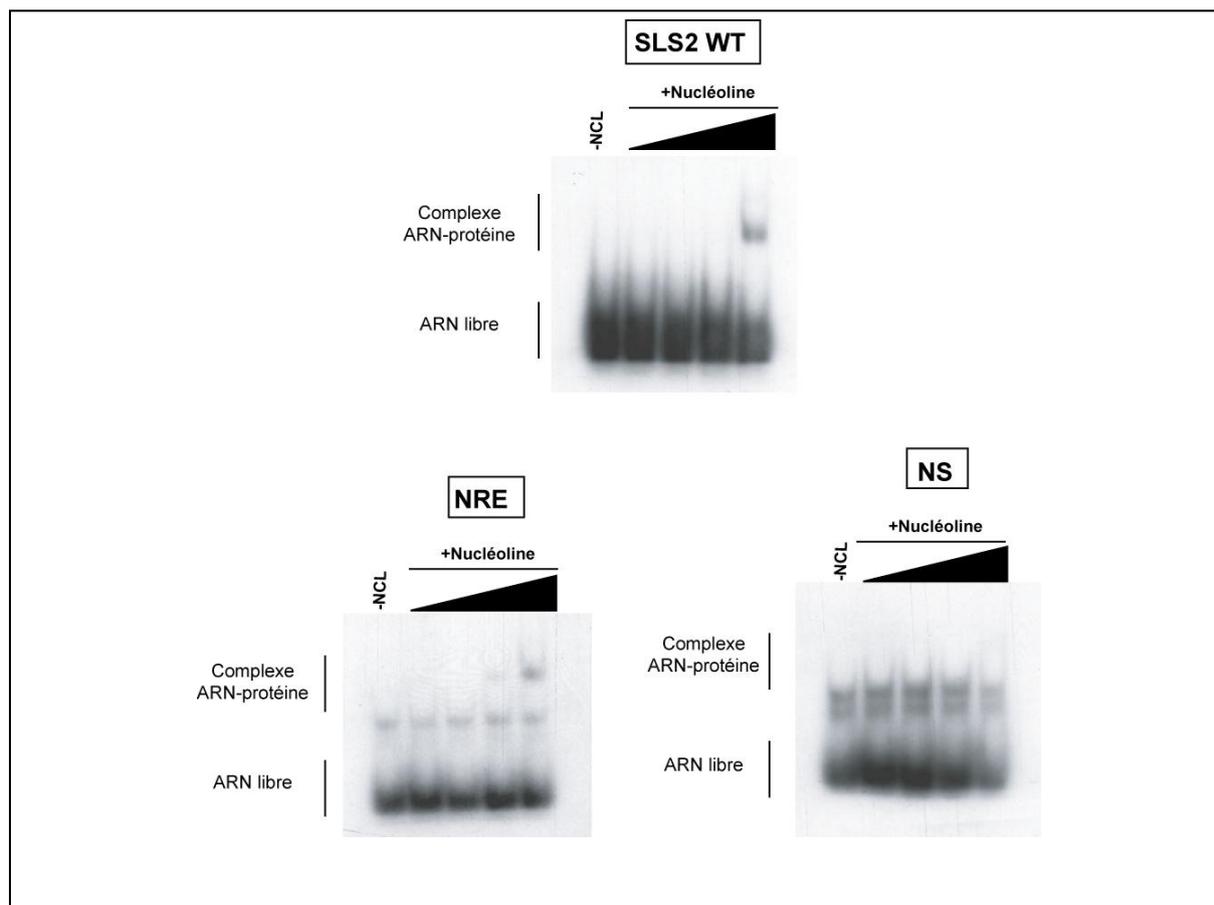


L'examen de la séquence SLS2-A7 montre la présence d'au moins 3 éléments distincts consécutifs ayant une homologie significative avec la séquence NRE conservée. Ces 3 motifs (5'-CCCCGA-3', 5'-ACCCGA-3', 5'-GCCCGA-3') sont situés entre les positions 7985 et 8009 dans la partie apicale de la structure tige-boucle SLS2-A7 (Figure R-6). Le motif central adopte une structure 2D similaire à celle définie pour une liaison efficace du domaine N-terminal de la nucléoline. Il est intéressant de noter que cet élément central englobe l'activateur ESE2.



De plus, 13 des 18 nucléotides situés entre les positions 7997 et 8015 (5'-CCGACAGGCCCGAAGGAAU-3') dans SLS2-A7 sont presque identiques à la séquence identifiée par l'expérience de SELEX menée avec la nucléoline (Bouvet *et al.*, 2001). Ainsi, les comparaisons de séquences suggèrent la présence d'un ou plusieurs sites de liaisons de la nucléoline dans la partie apicale de l'ARN SLS2-A7, qui sont représentés Figure R-6.

Des expériences de retard sur gel permettent de mettre en évidence la différence de fixation de la nucléoline sur ces différents ARN (Figure R-7).



**Figure R-7: Etude de la fixation de la protéine p40 nucléoline sur l'ARN SLS2-A7 WT et les ARN NRE et NS par expérience de retard sur gel.**

Analyse par expériences de retard sur gel des interactions entre l'ARN SLS2-A7 WT et les ARN NRE et NS et la protéine p40. L'ARN marqué radioactivement uniformément est incubé pendant 15 min à 4°C en présence de concentrations croissantes de protéine p40 nucléoline (0.5375, 0.1075, 0.15 et 0.43  $\mu$ M), puis les complexes sont fractionnés sur gel non dénaturant à 5%. L'ARN libre et les complexes formés sont indiqués sur les côtés.

De façon étonnante, les motifs GCCCGA et CCCGA sont retrouvés respectivement à 3 endroits seulement dans le génome de VIH-1 souche BRU pour GCCCGA, et à 2 endroits pour les motifs CCCGA (Figure R-8). Le premier emplacement se trouve dans la région 5' LTR U5, plus précisément dans la séquence PBS (« PrimBinding Site») à proximité de l'élément Psi du LTR requis pour l'encapsidation contenant lui-même le codon ATG d'initiation de la traduction de Gag. La séquence de 18 nucléotides PBS est complémentaire

aux 18 nucléotides de l'extrémité 3' de l'ARNt cellulaire Lys(3) qui est utilisé pour initier la transcription inverse de VIH-1. Le second emplacement est au sein de la tige-boucle SLS2-A7 contenant le site A7, cette tige-boucle comprenant aussi les 2 motifs CCCGA. Le troisième emplacement du motif GCCCGA se trouve peu avant le codon STOP de Nef. La séquence ECM (pour Evolutionary Conserved Motif) (5'UAGAUCGAUGU 3') a une forte affinité pour la nucléoline (Ginisty *et al.*, 2001). Cette séquence contient le motif de fixation à la nucléoline 5'-UCGA-3', motif qui requiert la présence des quatre domaines RRM pour la fixation à l'ARN. Le motif UCGA est aussi peu fréquent dans le génome de VIH-1 BRU.

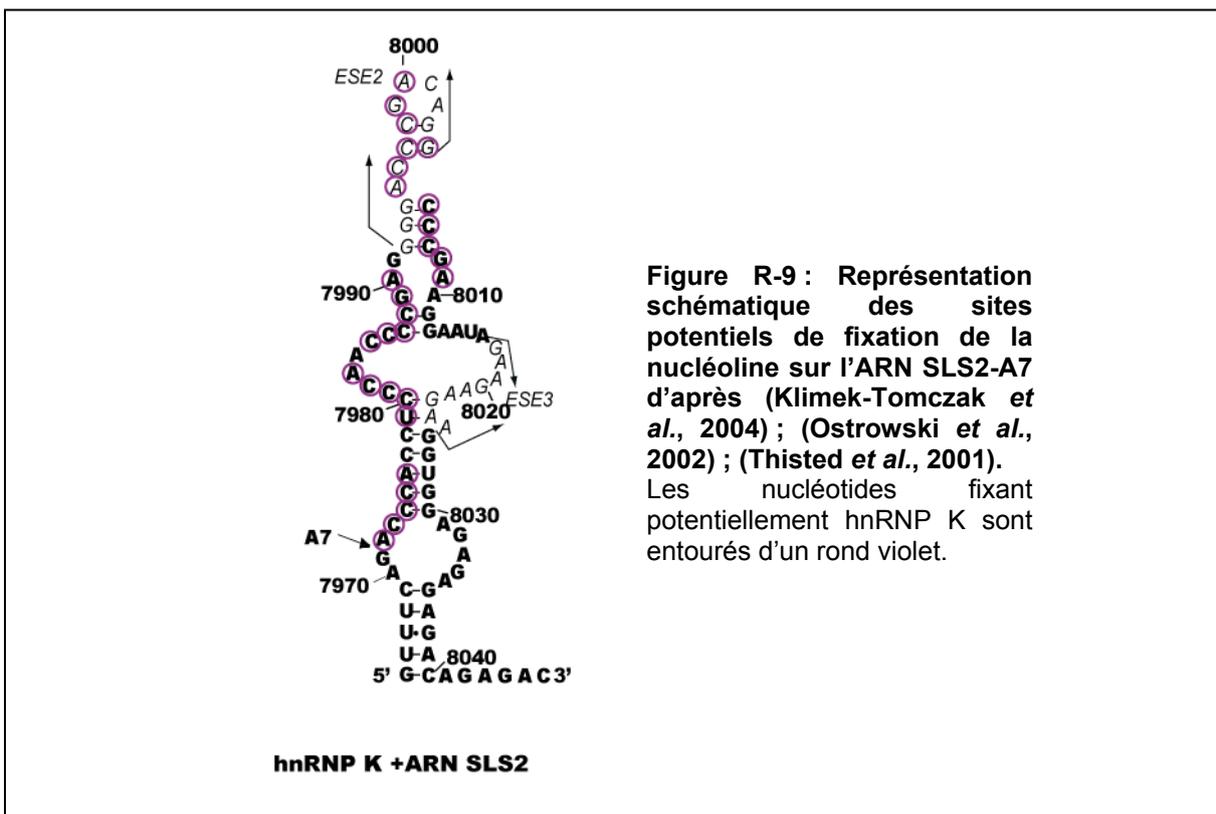
En lien avec cette observation, il est intéressant de noter que la nucléoline est incorporée dans les virions du VIH-1 et ceux du MoMLV et que son incorporation est diminuée lorsque la séquence Psi d'encapsidation, localisée juste en aval de l'élément PBS est éliminée (Ueno *et al.*, 2004). Par ailleurs, le domaine C-terminal de la nucléoline interagit avec la partie NC (nucléocapside) du précurseur Gag. Des mutations dans Gag, qui abolissent l'interaction avec la nucléoline, diminuent l'efficacité de relargage des virions néoformés (Ueno *et al.*, 2004). Si la structure tige-boucle SLS2-A7 fixe bien la nucléoline, une mutation de son site de fixation pourrait aussi avoir un effet sur la maturation ou le relargage des virions. Enfin, étant donné que la nucléoline augmente la demi-vie de certains ARNm tels que les ARNm *bcl-2*, on peut aussi se demander si la nucléoline ne peut pas augmenter la demi-vie de certains ARNm du VIH contenant la structure tige-boucle SLS2-A7.



### III.2 La protéine hnRNP K :

La protéine hnRNP K est connue pour se lier à des motifs riches en cytidines simple brin et sa séquence consensus établie par SELEX est 5'-AU(C)<sub>3/4</sub> (U/A)<sub>2</sub>-3' (Thisted *et al.*, 2001). La séquence nucléotidique située entre 7972 et 7989, immédiatement en aval du site A7 dans l'ARN SLS2-A7, est une séquence riche en résidus C (12 résidus C sur 16 nucléotides) (Figure R-9). Cette région peut par conséquent contenir plusieurs sites de fixation pour hnRNP K. Ce segment de 16 nucléotides dans l'ARN SLS2-A7, la séquence 5'-CUCCCAA-3', qui est très similaire à la séquence consensus obtenue à partir d'expériences de SELEX, est partiellement simple brin. Ces analyses permettent donc de penser que la protéine hnRNP K peut avoir des sites de fixation sur l'ARN SLS2-A7.

Comme hnRNP K et hnRNP H sont capables d'interagir l'une avec l'autre (Mikula *et al.*, 2006), il est de plus possible que la présence de hnRNP K liée à l'ARN SLS2-A7 puisse stabiliser la fixation de hnRNP H dans la même région.

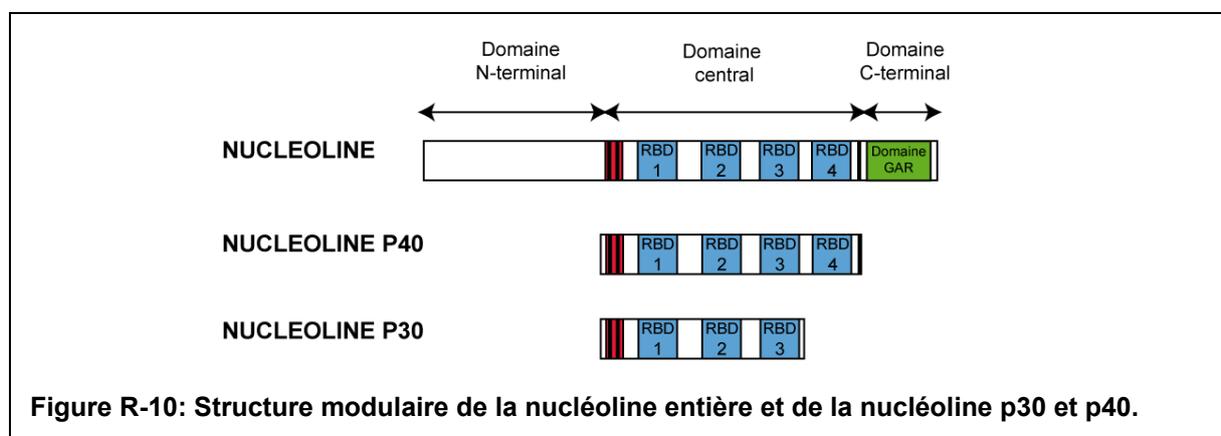


### IV. Etude de l'efficacité de fixation des protéines nucléoline et hnRNP K sur l'ARN SLS2-A7 en retard sur gel.

Des expériences de retard sur gel constituent une autre approche pour vérifier que la nucléoline et hnRNP K se fixent bien en plus de hnRNP A1 sur la structure tige-boucle SLS2-A7, et pour confirmer leur liaison directe à l'ARN SLS2-A7.

#### IV.1 Etude de l'affinité de la nucléoline sur l'ARN SLS2-A7.

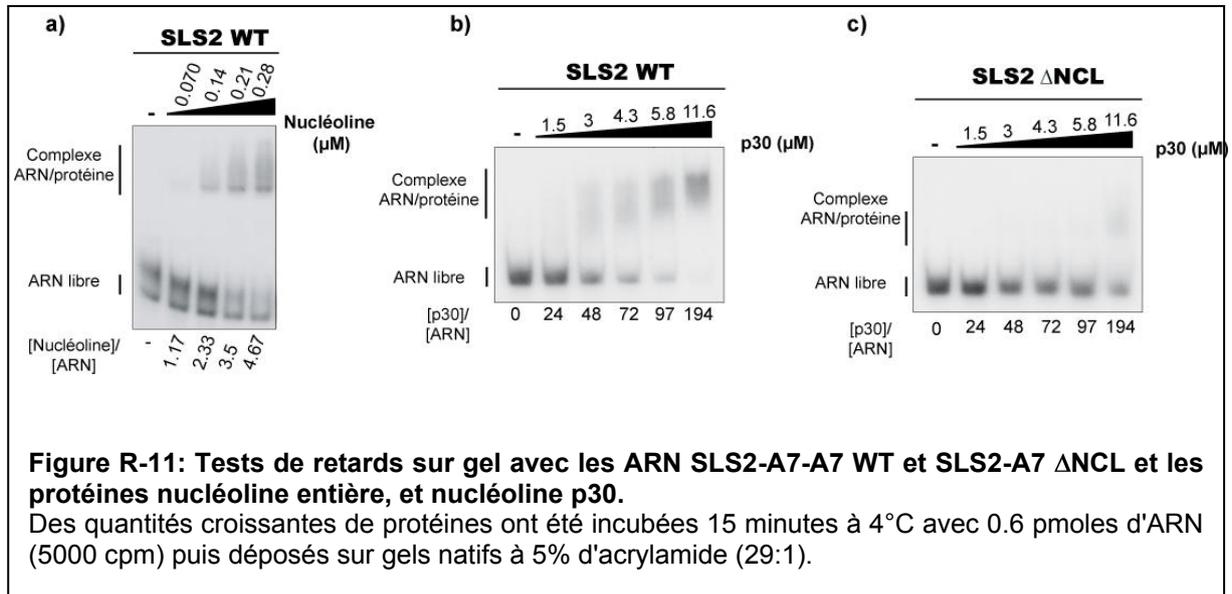
La nucléoline entière n'a pas pu être produite dans *E. coli*, nous l'avons donc purifiée à partir d'un extrait nucléaire de cellules HeLa (Matériel et Méthodes). La protéine nucléoline p40 ne contenant que les 4 RRM nous a été fournie par P. Bouvet (ENS Lyon). Cette protéine comprend les quatre RRM de la nucléoline qui sont suffisants pour la fixation à l'ARN (Serin *et al.*, 1997). Nous disposons du plasmide pET15b-p30 et nous avons produit et purifié la protéine nucléoline p30 comportant les trois premiers RRM (Figure R-10). Seuls les deux premiers RRM de la nucléoline sont indispensables pour la fixation à l'ARN (Serin *et al.*, 1997).



La partie apicale de la tige-boucle SLS2-A7 contient l'élément activateur d'épissage ESE2 lie la protéine hnRNP A1 (Marchand *et al.*, 2002) et le haut de la tige-boucle a aussi été décrit comme pouvant fixer ASF/SF2 (Mayeda *et al.*, 1999).

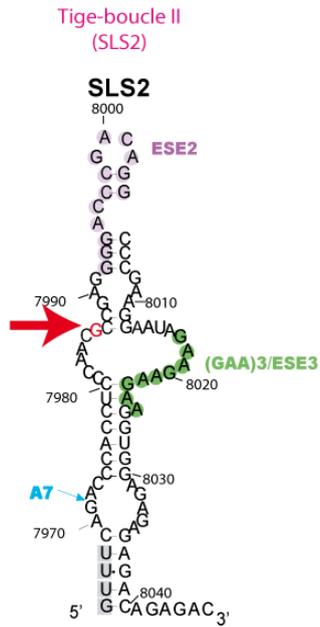
Pour vérifier la fixation de la nucléoline sur la partie apicale de SLS2-A7, nous avons utilisé un fragment de la nucléoline contenant l'extrémité N-terminale avec les trois premiers RRM (Figure R-10), qui a été ainsi nommé p30. Les retards sur gels, présentés en Figure R-11, montrent que la fixation de la nucléoline entière purifiée à partir d'extraits nucléaires de cellules HeLa a une meilleure affinité pour l'ARN que la protéine recombinante p30. Cela correspond à ce qui est publié : si les deux premiers RRM sont indispensables, l'affinité est cependant meilleure avec les quatre (Serin *et al.*, 1997). La nucléoline entière se lie à l'ARN SLS2-A7 avec un  $K_D$  apparent de 200 nM, ce qui est proche du  $K_d$  trouvé pour les pré-ARNr (Allain *et al.*, 2000). Trois sites potentiels de fixation de la nucléoline se trouvent dans la partie apicale de SLS2-A7 (Figure R-6). Une diminution importante de la fixation de p30 est observée (Figure R-11) lors de la délétion du haut de la tige-boucle (mutant  $\Delta$ NCL, Figure R-12a). Le haut de la tige-boucle est donc important pour la fixation de la nucléoline sur l'ARN SLS2-A7, seulement une faible bande de retard sur gel est observée pour une concentration de

p30 de 11,6  $\mu\text{M}$  avec l'ARN tronqué tandis que le Kd estimé pour l'ARN SLS2-A7 WT avec la protéine p30 est d'environ 3,5  $\mu\text{M}$ . Ces données montrent que les sites de fixation de la p30 les plus efficaces pour l'ARN SLS2-A7 semblent être dans sa partie apicale.

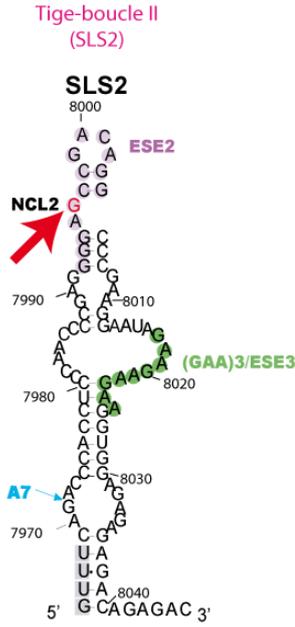


Différents ARN mutants de la tige-boucle SLS2-A7 sont disponibles au laboratoire (Figure R-12a et b) et la fixation de nos protéines recombinantes a été testée sur chacun.

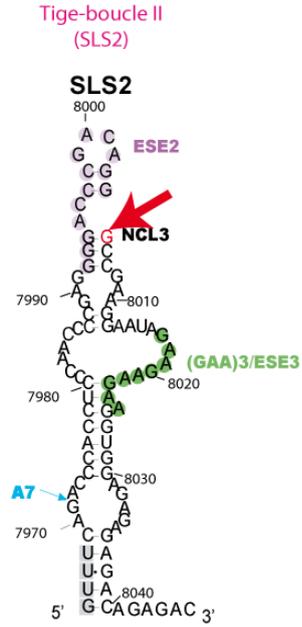
**SLS2 NCL1**



**SLS2 NCL2**



**SLS2 NCL3**



**SLS2 ΔNCL**

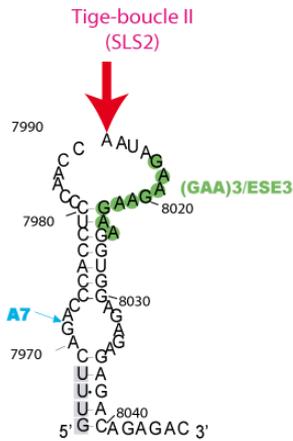


Figure R-12 a: Structures et séquences des différents mutants souhaités de la construction SLS2 WT.

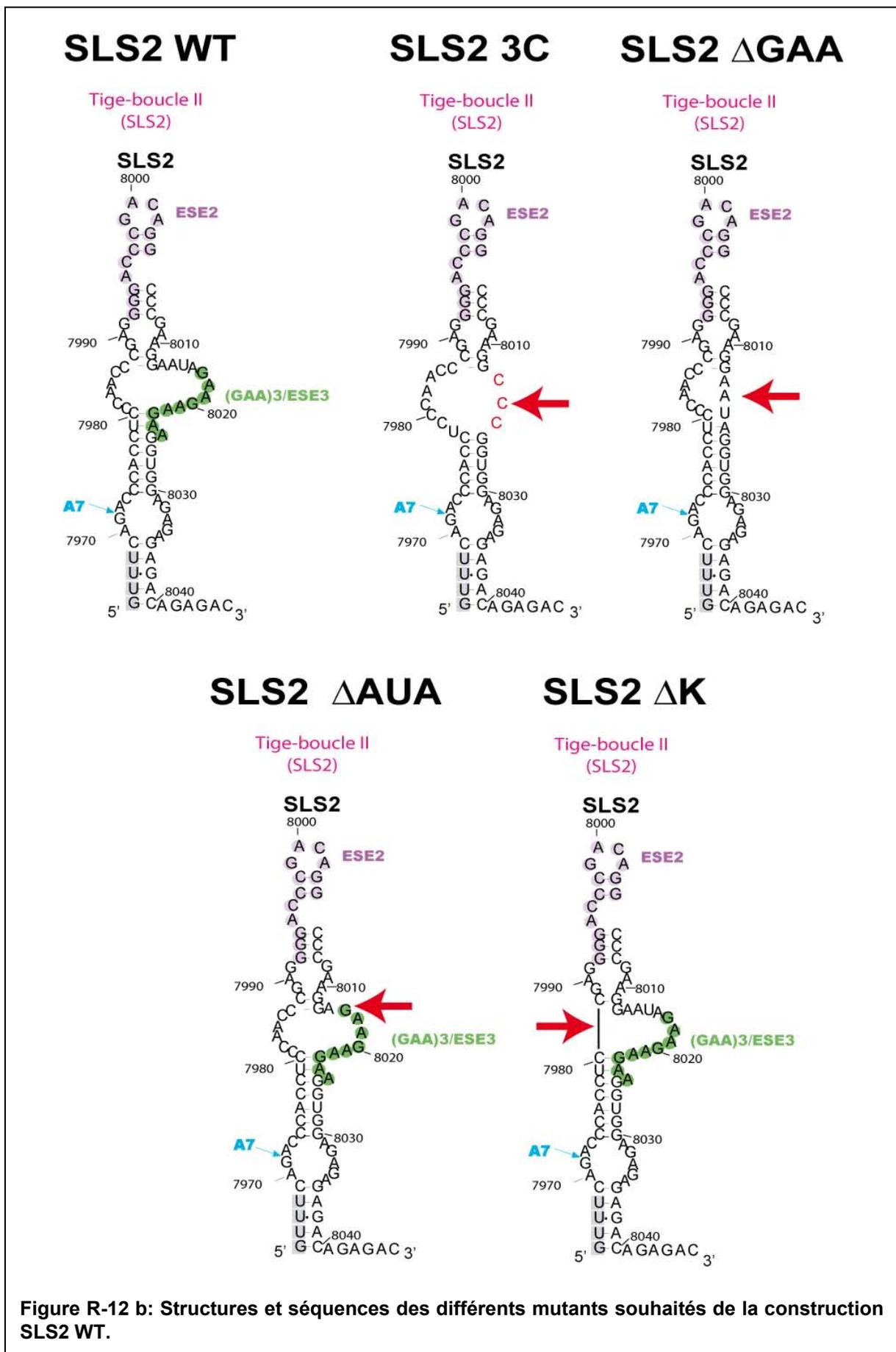
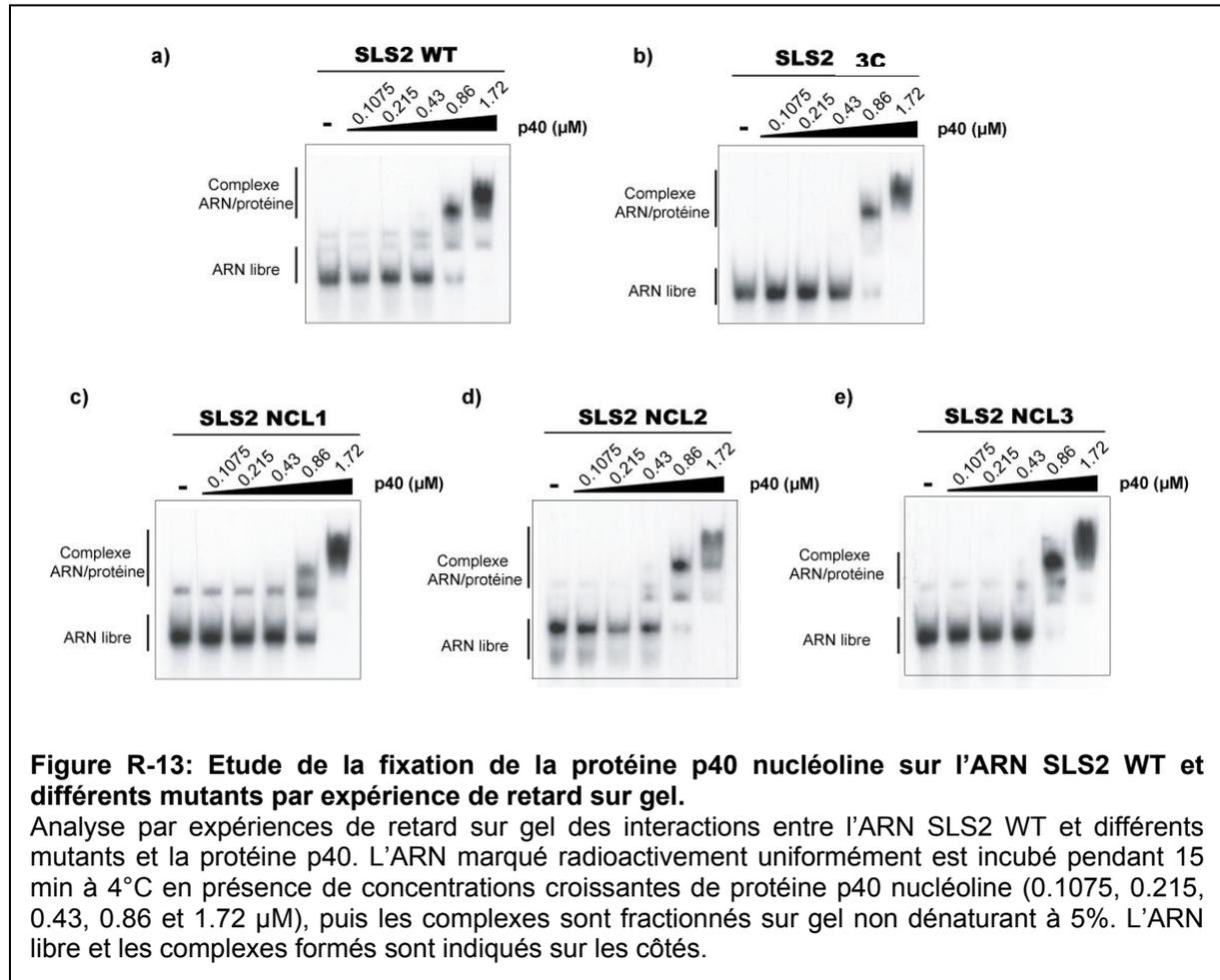


Figure R-12 b: Structures et séquences des différents mutants souhaités de la construction SLS2 WT.

Comme les études précédentes ont montré qu'une substitution de C en G à la position en 5' de la séquence CCCGA conduit à une diminution significative de l'affinité pour la nucléoline (Bouvet *et al.*, 2001), des substitutions C en G ont été introduites dans chacun des trois sites potentiels dans l'ARN SLS2-A7 et l'affinité de la protéine p30 pour ces ARN testée en retard sur gel (variants SLS2-A7 mNCL1, mNCL2 et mNCL3, Figure R-12a et 13).



Aucun changement important d'affinité n'est observé avec ces mutations, à part une légère diminution avec l'ARN SLS2-A7 NCL2. Ces mutations n'affectent donc pas la fixation de p30 contrairement à ce qui était prévu.

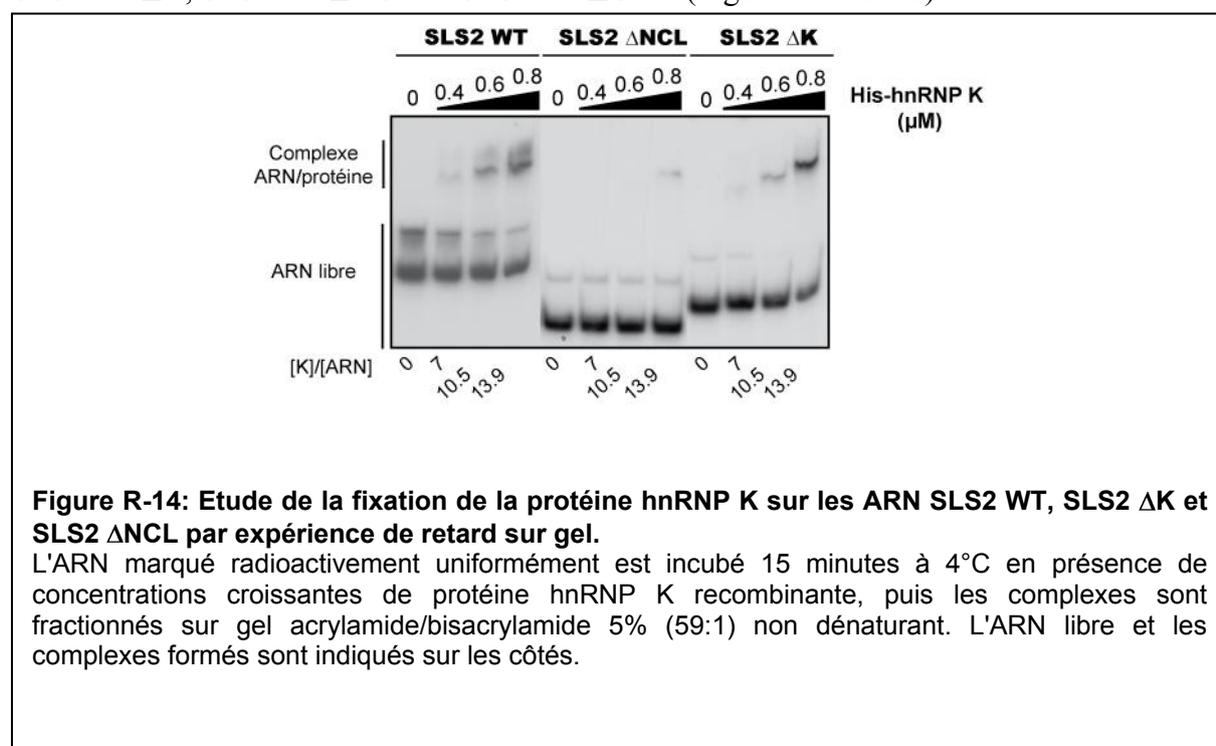
#### IV.2 Etude de l'affinité de la protéine hnRNP K pour l'ARN SLS2-A7.

Pour cette étude, j'ai cloné, produit et purifié la protéine recombinante hnRNP K entière avec une étiquette histidine en conditions natives (Matériel et Méthodes). Sur la base des sites de fixation que l'on pouvait prédire pour hnRNP K sur SLS2-A7 d'après les données de la littérature (Figure R-9), j'ai utilisé des variants de la structure tige-boucle SLS2-A7 disponibles au laboratoire (Figure R-12 a et b),

Nous avons vu que de nombreux sites potentiels de fixation sont possibles pour hnRNP K sur la tige-boucle SLS2-A7. Etant donné la préférence de fixation de cette protéine pour les séquences poly-C simple brin, la délétion des bases 7980-7987 dans le mutant  $\Delta K$  devrait modifier la fixation de hnRNP K à l'ARN.

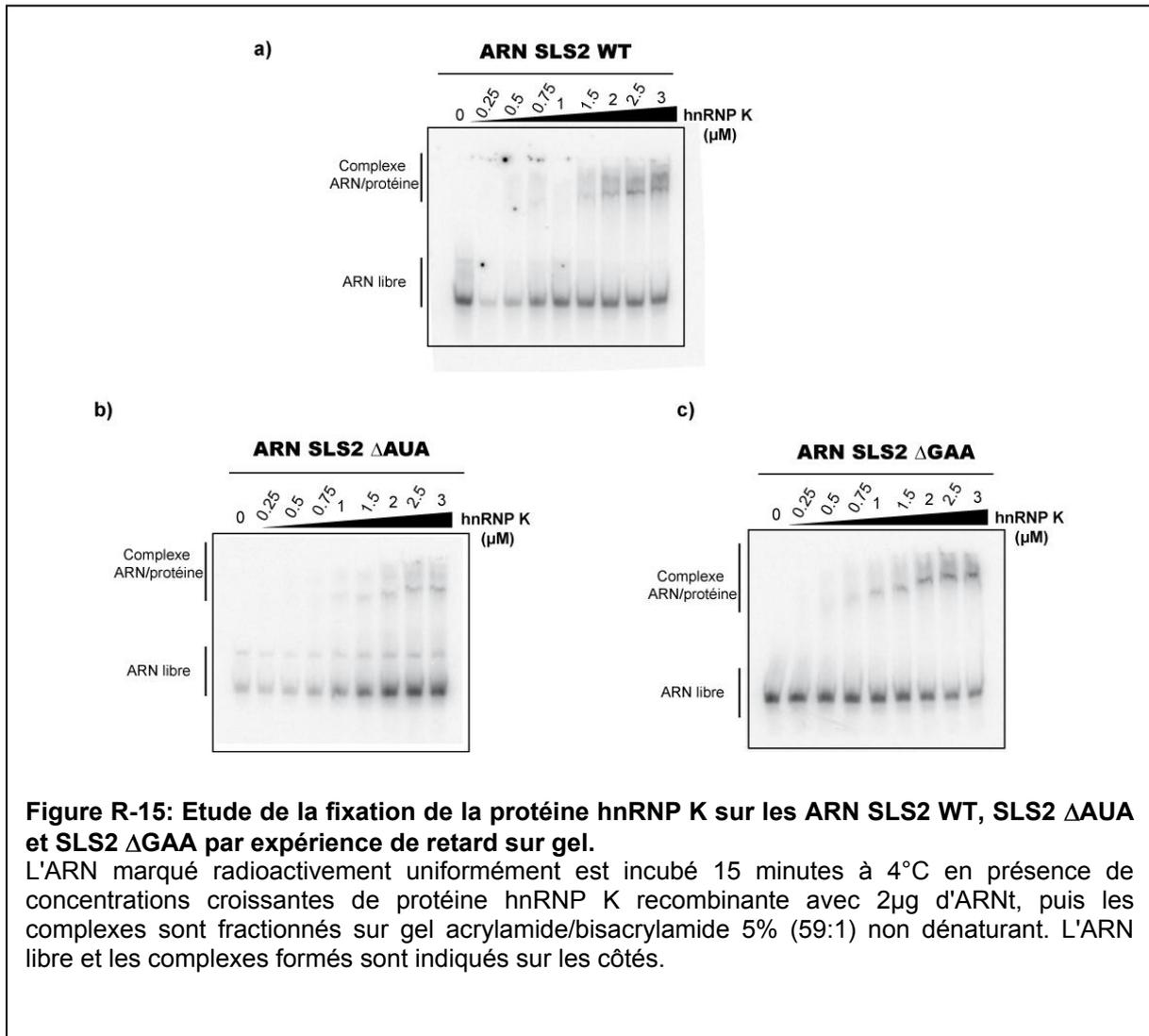
Une autre délétion a été réalisée au niveau de l'élément bivalent Janus [AAUA(GAA)<sub>3</sub>] de la tige-boucle 2 et le fragment d'ARN délété a été remplacé par trois cytidines. Il était intéressant d'observer l'effet de la délétion de ce fragment sur la fixation de hnRNP K même si l'ajout de ce poly-C peut provoquer une fixation de hnRNP K de façon artificielle.

Nous avons comparé l'affinité de hnRNP K pour les ARN SLS2-A7 WT, SLS2-A7 $\Delta$ NCL, SLS2-A7  $\Delta$ K, SLS2-A7  $\Delta$ AUA et SLS2-A7  $\Delta$ GAA (Figure R-14 et 15).



Le profil de migration en escalier du complexe hnRNP K/ARN montre que la protéine hnRNP K pourrait avoir plusieurs sites de fixation coopératifs sur cet ARN. La fixation de hnRNP K sur le mutant SLS2-A7  $\Delta$ NCL est très nettement inférieure et montre que le haut de la tige-boucle est important pour la fixation de hnRNP K sur SLS2-A7.

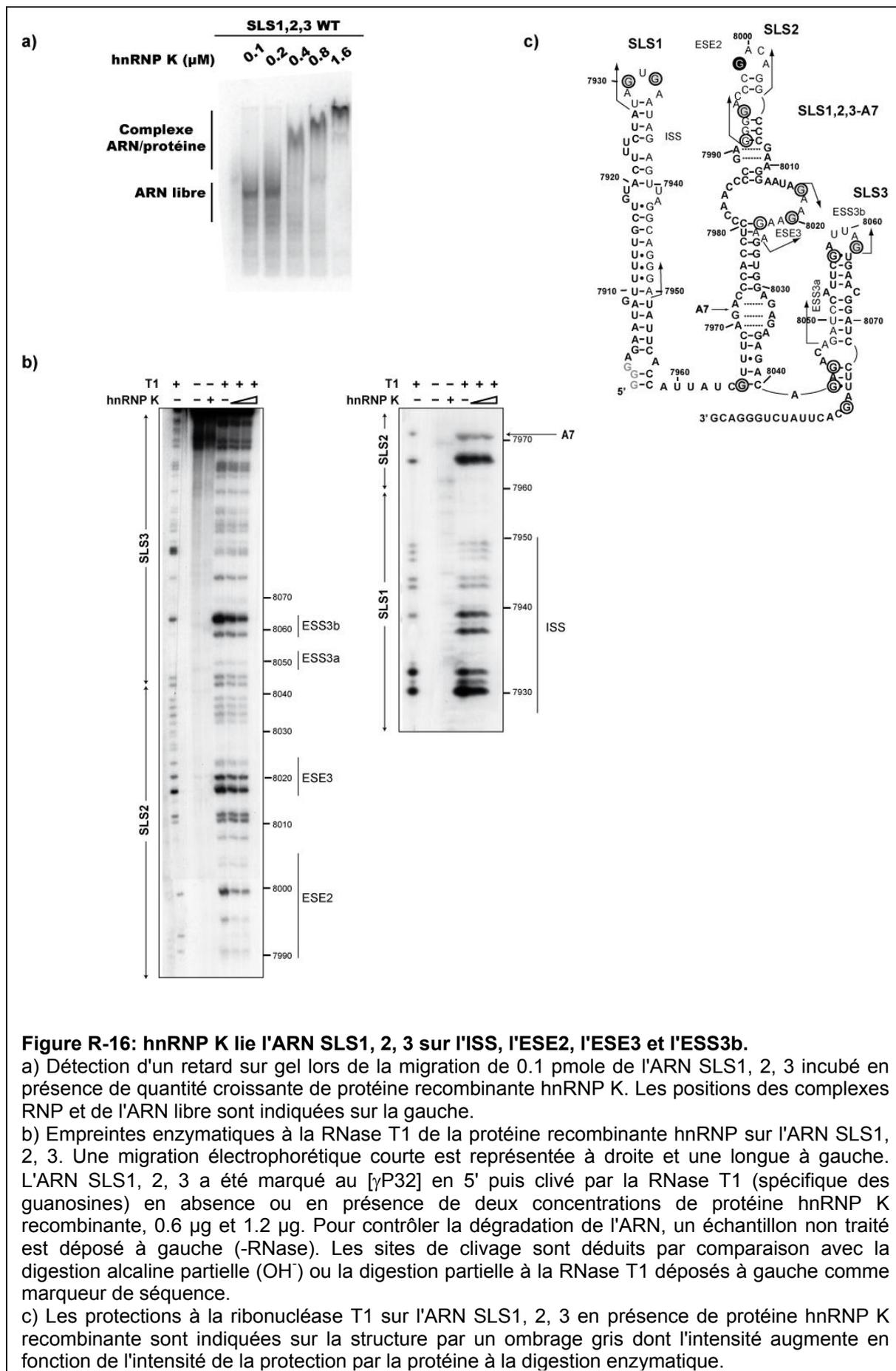
Il est surprenant de noter que le mutant  $\Delta$ K, dont le site de fixation le plus probable de hnRNP K était délété, ne montre pas de différence notable de fixation de hnRNP K. Par contre, la délétion des trois nucléotides AUA dans l'élément Janus en amont de l'ESE3 provoque une légère diminution de la fixation de hnRNP K sur l'ARN  $\Delta$ AUA (Figure R-15).



## V. La protéine hnRNP K présente plusieurs sites de fixation sur l'ARN autour du site A7.

Pour déterminer les sites de fixation de hnRNP K à proximité du site accepteur d'épissage A7, j'ai réalisé des expériences d'empreintes enzymatiques avec un ARN contenant les structures tiges-boucles SLS1, 2, 3 (Figure R-16). Des retards sur gel ont permis d'estimer la concentration de protéines requises dans les empreintes avec la protéine hnRNP K (Figure R-16 a). La position et l'intensité du clivage sont estimées après électrophorèse sur gel dénaturant. Comme représenté schématiquement en Figure R-16 b et c, les protections les plus fortes sont trouvées sur plusieurs sites. La protéine hnRNP K couvre certains nucléotides dans la boucle terminale de SLS1 (ISS, élément intronique inhibiteur), dans les éléments ESE2 et ESE3 considérés comme des éléments activateurs, et dans la boucle terminale SLS3, dans l'élément ESS3b. Ces données montrent que les quatre éléments régulateurs de l'épissage au site A7 (ISS, ESE2, ESE3 et ESS3b) constituent des sites forts de fixation pour

la protéine hnRNP K. L'empreinte confirme la fixation présumée de hnRNP K sur la partie apicale de SLS2-A7. hnRNP K se fixe aussi sur la séquence AAUAGAA de l'ESE3. Cependant, les retards sur gel montrent que la fixation de hnRNP K est moins affectée par ces mutations, la fixation de la protéine semble donc être plus forte sur la partie apicale de l'ARN, dans l'ESE2 plutôt que dans l'ESE3.



**Figure R-16: hnRNP K lie l'ARN SLS1, 2, 3 sur l'ISS, l'ESE2, l'ESE3 et l'ESS3b.**

a) Détection d'un retard sur gel lors de la migration de 0.1 pmole de l'ARN SLS1, 2, 3 incubé en présence de quantité croissante de protéine recombinante hnRNP K. Les positions des complexes RNP et de l'ARN libre sont indiquées sur la gauche.

b) Empreintes enzymatiques à la RNase T1 de la protéine recombinante hnRNP sur l'ARN SLS1, 2, 3. Une migration électrophorétique courte est représentée à droite et une longue à gauche. L'ARN SLS1, 2, 3 a été marqué au [γP32] en 5' puis clivé par la RNase T1 (spécifique des guanosines) en absence ou en présence de deux concentrations de protéine hnRNP K recombinante, 0.6 μg et 1.2 μg. Pour contrôler la dégradation de l'ARN, un échantillon non traité est déposé à gauche (-RNase). Les sites de clivage sont déduits par comparaison avec la digestion alcaline partielle (OH<sup>-</sup>) ou la digestion partielle à la RNase T1 déposés à gauche comme marqueur de séquence.

c) Les protections à la ribonucléase T1 sur l'ARN SLS1, 2, 3 en présence de protéine hnRNP K recombinante sont indiquées sur la structure par un ombrage gris dont l'intensité augmente en fonction de l'intensité de la protection par la protéine à la digestion enzymatique.

## **VI. Fonctions possibles de la fixation des protéines hnRNP K et nucléoline près du site A7 accepteur d'épissage.**

### *Nucléoline :*

Parmi les protéines identifiées dans les complexes ARN SLS2-A7/RNP, l'identification de la nucléoline comme protéine interagissant directement avec l'ARN SLS2-A7 est surprenante. Jusqu'à présent aucun lien n'a été observé entre la nucléoline et l'épissage. La nucléoline est en effet décrite comme étant principalement impliquée dans la plupart des étapes de production des ribosomes : la régulation de la transcription des ARN ribosomiques, mais aussi leur processus de maturation, l'assemblage des sous-unités ribosomiques et le transport nucléocytoplasmique de ces sous-unités. Par des expériences de Western blot quantitatif, nous avons montré que cette protéine est en effet très abondante dans l'extrait nucléaire de cellules HeLa (environ 100 ng/ml dans l'extrait nucléaire de cellules HeLa).

La présence de sites forts de fixation de la nucléoline à proximité du site accepteur d'épissage A7 pourrait indiquer son implication dans la régulation de l'épissage de VIH-1.

### *hnRNP K :*

A l'époque où nous avons commencé l'étude de la protéine hnRNP K sur l'épissage du VIH-1, la seule description de l'effet de cette protéine sur l'épissage était faite par A. Expert-Bezançon qui montrait qu'elle avait un effet activateur sur l'épissage de l'exon 6A du pré-ARNm de la  $\beta$ -tropomyosine de poulet (Expert-Bezançon *et al.*, 2002).

## **VII. Discussion**

Nous avons pu isoler trois nouvelles protéines interagissant avec l'ARN contenant le site A7 d'épissage. Nous avons purifié ces protéines sous forme entière ou tronquée ce qui nous a permis d'étudier leurs propriétés de fixation *in vitro* sur notre ARN. Les sites de fixation sur l'ARN ont été déterminés précisément pour hnRNP K et situés principalement dans le haut de la tige-boucle SLS2-A7 pour la nucléoline.

### **De multiples complexes ARN/protéines sont formés sur un ARN SLS2-A7 incubé dans un extrait nucléaire de cellules HeLa.**

En prenant en compte les résultats de pontage covalent aux UV et les masses moléculaires des complexes RNP isolés, nous pouvons conclure que la majorité des protéines trouvées dans les

échantillons sont retenues par des interactions protéines-protéines et non par une liaison directe à l'ARN. Ce résultat était attendu, étant donné que cette organisation semble être commune lors de l'assemblage de grands complexes RNP.

Toutefois, les études de pontage aux UV permettant de mettre en évidence la liaison directe des protéines à l'ARN peuvent sous-estimer le nombre de partenaires directs liés à l'ARN. En effet, il a été précédemment montré que d'autres protéines peuvent lier l'ARN SLS2-A7 WT, comme la protéine SR ASF/SF2 qui lie l'ARN SLS2-A7 en absence de la protéine hnRNP A1. De plus, le ratio ARN/extrait nucléaire (100 pmole/6-7 mg de protéines totales) utilisé dans nos expériences peut favoriser même de faibles affinités de fixation de protéines, vu que l'ARN est présent en large excès. Ceci expliquerait pourquoi la plupart des protéines trouvées dans les complexes RNP purifiés sont présents dans des quantités non-stœchiométriques.

### **Sites de fixation de la protéine nucléoline p30.**

Nous nous attendions à ce que les mutations ponctuelles de C en G dans les sites de fixation potentiels de la nucléoline change les affinités pour la protéine p30, contrairement à ce qui est observé. V. Marchand a réalisé des empreintes enzymatiques à la RNase T1 avec la protéine p30 sur l'ARN SLS1, 2, 3. Elle a observé que la p30 se fixe sur la partie apicale de la tige-boucle 1, dans l'ISS (nucléotides G 7930 et 7932), dans l'ESE2 (nucléotides G 7993-7994 et CGA 7998 à 8000) et dans l'ESS3b (nucléotide G 8062). La nucléoline ne se fixerait donc pas au niveau des sites consensus connus. Cette différence entre la prédiction et l'observation ne semble pas être due à l'utilisation d'une protéine tronquée car la fixation simultanée des 2 premiers RRM de la nucléoline à l'ARN est requis et suffisant pour une fixation de haute affinité et une interaction spécifique.

La nucléoline a été trouvée dans l'une des bandes les plus intenses de l'analyse SDS-PAGE des RNP SLS2-A7-A7-MS2. La présence des 4 RRM de la nucléoline est requise pour la fixation de la séquence ECM (5'UAGAUCGAUGU 3') tandis que les 2 premiers domaines RRM de la protéine recombinante p30 que nous avons produite sont suffisants pour se fixer spécifiquement au motif NRE (Ginisty *et al.*, 2000). Deux des trois sites de fixation de la protéine p30 trouvés dans la partie apicale de la tige-boucle SLS2-A7 correspondent à la séquence NRE consensus (5'-U/GCCCCGA-3'). Le troisième site diffère seulement en position 5' par la présence d'un résidu A. Dans la structure 3D qui a été établie pour l'interaction p30-NRE (Finger *et al.*, 2004), le NRE est situé dans la boucle terminale d'une structure tige-boucle et la boucle ARN est prise en sandwich entre les deux RRM et leur domaine de liaison. L'observation que la protéine p30 se lie aussi aux éléments ESE3, ESS3

et ISS peut être expliquée par la présence de séquences UAG, plus particulièrement sur la séquence UUAG de l'ESS3. La nucléoline a été trouvée associée avec des séquences UUAG d'ARNm cellulaires (Ishikawa *et al.*, 1993).

La nucléoline est impliquée dans plusieurs aspects du cycle de multiplication de VIH-1. Les nouveaux sites de fixation de la nucléoline mis en évidence ici pourraient jouer un rôle dans certaines de ses fonctions déjà établies ou bien impliquées dans d'autres fonctions encore inconnues de la nucléoline dans le cycle de multiplication du VIH-1.

Bien que située de façon prédominante dans le nucléole, la nucléoline est une protéine navette. Elle est aussi trouvée dans le cytoplasme et la membrane plasmique (Srivastava *et al.*, 1999; Ginisty *et al.*, 1999). Elle peut être située sur la surface des cellules en étroite association avec le cytosquelette intra-cellulaire d'actine. Cette nucléoline de surface peut permettre l'import intra-cellulaire de ligands spécifiques (Dumler *et al.*, 1999; Larrucea *et al.*, 1999). Ainsi, la nucléoline pourrait servir de facilitateur pour l'internalisation des particules de VIH-1. L'ancrage des virions VIH-1 sur les cellules permissives conduit à la co-agrégation des particules virales avec la nucléoline de surface sur des microdomaines de la membrane (Nisole *et al.*, 2002).

Lors de l'infection par VIH-1, une partie de la nucléoline relocate dans le cytoplasme (Marasco *et al.*, 1994) et peut participer avec la protéine virale Gag à la production du virion (Bacharach *et al.*, 2000). Gag reconnaît le signal de packaging (Psi) de l'ARN génomique de VIH-1 par interaction avec ses motifs en doigts de zinc et cette interaction joue un rôle pivot dans le processus de packaging de l'ARN (Ueno *et al.*, 2004). Comme Gag et la nucléoline sont supposés se lier avec la structure tige-boucle contenant Psi et qu'ils sont incorporés ensemble dans les virions (Ueno *et al.*, 2004; Bacharach *et al.*, 2000), la nucléoline semble impliquée dans l'assemblage du virion. De plus, elle a montré un rôle positif sur le bourgeonnement viral (Ueno *et al.*, 2004). D'après ces différentes propriétés, les nouveaux sites de fixation de la nucléoline découverts sur l'ARN SLS2-A7 pourraient jouer un rôle dans le trafic de VIH-1, l'assemblage des virions et le bourgeonnement.

### **Fixation mutuellement exclusive et coopérativité des protéines régulatrices sur l'ARN SLS2-A7.**

En examinant la taille de l'ARN SLS2-A7 et le nombre de sites potentiels de fixation de protéines dans cette région, il paraît inconcevable que toutes ces protéines interagissant avec l'ARN en même temps. Les masses moléculaires observées et l'hétérogénéité des complexes RNP observée par gradient de glycérol accèdent aussi la compétition possible entre nucléoline, hnRNP H, hnRNP K et hnRNP A1.

V. Marchand a étudié cette compétition entre les différentes protéines recombinantes par des expériences de gel retard. Nos données montrent que la fixation de la nucléoline à l'ARN SLS2-A7 semble être indépendante et mutuellement exclusive de la fixation des protéines hnRNP H et hnRNP K. D'autre part, notre étude révèle que la nucléoline n'est pas capable de déplacer la protéine hnRNP A1 liée à l'ARN SLS2-A7. Ces résultats sont en accord avec nos données précédentes montrant que hnRNP A1 ne se lie pas seulement fortement avec l'ARN mais est aussi capable de multimériser le long de la molécule d'ARN (Okunola *et al.*, 2009; Zahler *et al.*, 2004). Cette fixation coopérative de hnRNP A1 à l'ARN SLS2-A7 couvre entièrement l'ARN et, en conséquence, masque les sites de fixation pour d'autres protéines. Ceci pourrait expliquer l'incapacité de la protéine p30 à perturber le complexe hnRNP A1/ARN SLS2-A7 même à concentrations de p30 assez élevées.

Par contre, la protéine p30 est capable de rentrer en compétition pour sa fixation à l'ARN avec hnRNP H et hnRNP K. Les sites de fixation pour p30 et hnRNP K se chevauchant partiellement, ces deux protéines sont par conséquent en compétition pour leur fixation à l'ARN SLS2-A7. Toutefois la protéine hnRNP K est incapable de déplacer la p30 déjà fixée à l'ARN.

Ces résultats sont obtenus avec des protéines recombinantes, entières pour hnRNP A1 et hnRNP K mais tronquées pour les protéines p30 et hnRNP H RRM1, 2. Il faut donc prendre en compte le fait que leur affinité pour l'ARN et les interactions potentielles entre elles peuvent différer légèrement de ce qui pourrait être observé avec des protéines purifiées à partir d'extrait cellulaire.

## **CHAPITRE III Effets des protéines nucléoline, hnRNP H et hnRNP K sur le taux d'utilisation de sites accepteurs d'épissage du virus VIH-1.**

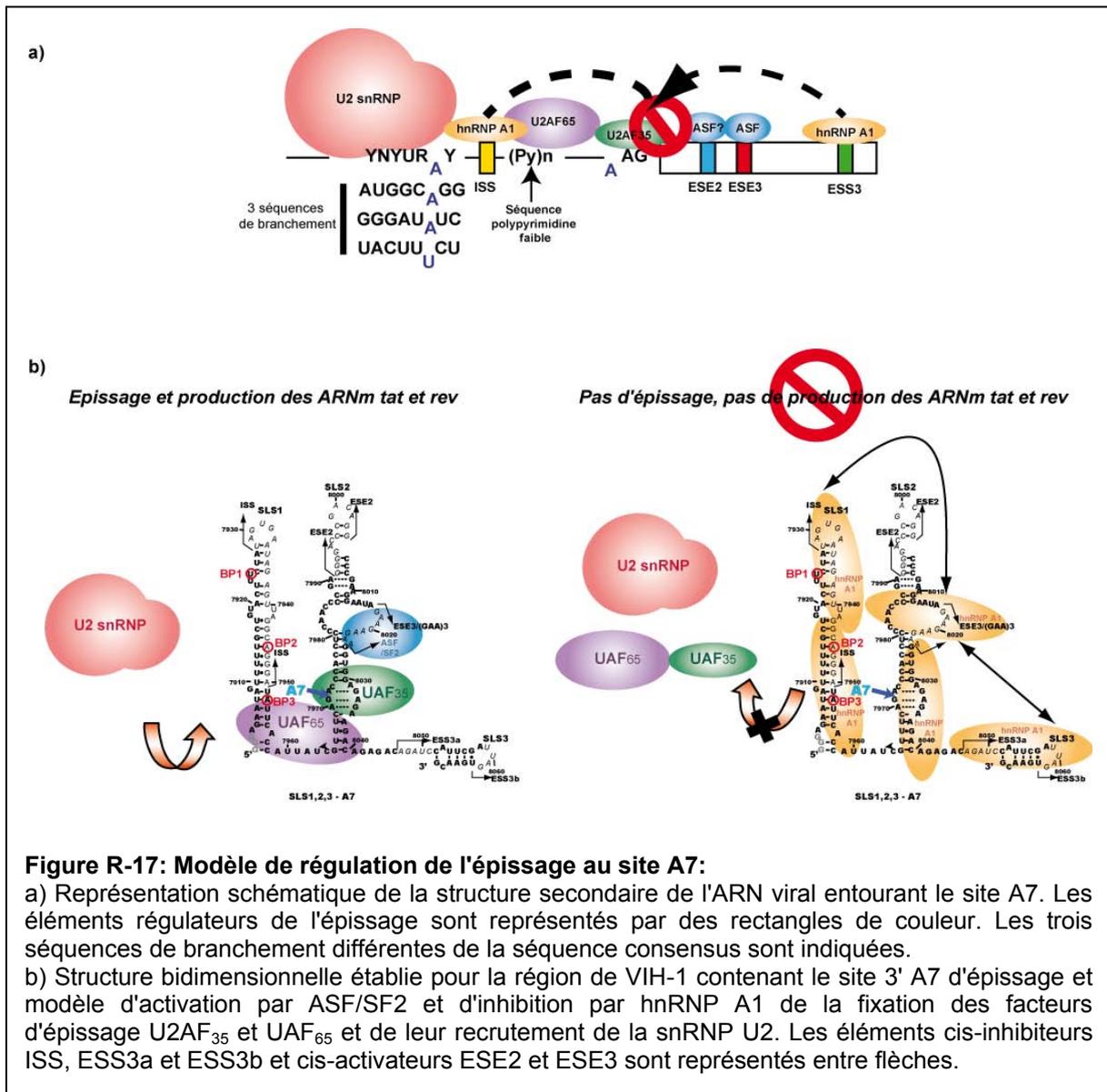
Avant mon arrivée au laboratoire, L. Damier et V. Marchand ont produit une série de constructions permettant d'obtenir des ARN renfermant les différents couples de sites d'épissage de l'ARN du virus VIH-1 (souche BRU) et L. Damier a défini l'efficacité relative de ces différents couples de sites *in vitro*. J'ai pu utiliser ces nombreuses constructions pour tester l'effet des protéines identifiées par spectrométrie de masse sur l'épissage *in vitro*.

### **I. Etude de l'effet de la nucléoline et des protéines hnRNP H et hnRNP K sur l'épissage *in vitro* de l'intron entre les sites d'épissage D4-A7 de VIH-1.**

Les résultats obtenus au laboratoire ainsi que et de nombreuses données de la littérature indiquent que les régulations de l'ARN du virus VIH-1 dépendent largement de régulations s'appliquant au niveau des sites accepteurs. Le mécanisme décrit pour l'épissage au site A7 met en jeu un équilibre entre la fixation de la protéine SR ASF/SF2 qui favorise et stabilise la fixation de U2AF<sub>35</sub> à la jonction 3', et la fixation de la protéine inhibitrice hnRNPA1 (Figure R-17). Le facteur général d'épissage U2AF<sub>65</sub> reconnaît spécifiquement l'enchaînement de pyrimidines adjacent au site 3' d'épissage et facilite ainsi le recrutement de la snRNP U2 au niveau de la boîte de branchement du pré-ARNm. La séquence polypyrimidine du site A7 étant sous-optimale, la fixation de UAF<sub>35</sub> au site 3' est indispensable pour stabiliser la fixation de U2AF<sub>65</sub> sur l'enchaînement de pyrimidines. L'ensemble de ces interactions favoriserait l'interaction de la snRNP U2 avec le pré-ARNm et induirait une conformation particulière de l'intron visant certainement à rapprocher le site 5' du site 3' d'épissage.

Les protéines de la famille des hnRNP sont connues pour être impliquées dans l'épissage des pré-ARNm. Elles peuvent réprimer l'épissage en bloquant la reconnaissance des sites d'épissage ou interférer avec la fixation des protéines qui doivent se lier aux activateurs (Figure R-17). Récemment, les protéines hnRNP ont été décrites comme empêchant la communication entre des facteurs liés à différents sites d'épissage mais aussi comme ayant un rôle positif dans l'épissage des pré-ARNm. De plus, des interactions coopératives entre les protéines hnRNP liées peuvent stimuler l'épissage entre les paires spécifiques de sites d'épissage tandis qu'elles gênent simultanément d'autres combinaisons (Martinez-Contreras

*et al.*, 2007). Ainsi, les protéines hnRNP utilisent une grande variété de stratégies pour contrôler la sélection d'un site d'épissage.



**Figure R-17: Modèle de régulation de l'épissage au site A7:**

a) Représentation schématique de la structure secondaire de l'ARN viral entourant le site A7. Les éléments régulateurs de l'épissage sont représentés par des rectangles de couleur. Les trois séquences de branchement différentes de la séquence consensus sont indiquées.

b) Structure bidimensionnelle établie pour la région de VIH-1 contenant le site 3' A7 d'épissage et modèle d'activation par ASF/SF2 et d'inhibition par hnRNP A1 de la fixation des facteurs d'épissage U2AF<sub>35</sub> et UAF<sub>65</sub> et de leur recrutement de la snRNP U2. Les éléments cis-inhibiteurs ISS, ESS3a et ESS3b et cis-activateurs ESE2 et ESE3 sont représentés entre flèches.

La forte concentration d'éléments NRE (Nucleolin Recognition Element) (Ginisty *et al.*, 2000) dans la partie apicale de l'ARN SLS2-A7 contenant le site d'épissage A7 et le fait que seulement un nombre limité de motifs de type NRE existe dans le génome de VIH-1 sont des indications pour une fonction importante de la fixation de la nucléoline dans la région de l'ARN SLS2-A7. La nucléoline se lie à des sites identifiés auparavant comme des sites de fixation de la protéine hnRNP A1 (Marchand *et al.*, 2002; Damgaard *et al.*, 2002; Zhu *et al.*, 2001). Selon la concentration cellulaire de hnRNP A1 et de nucléoline, l'une ou l'autre protéine pourrait se fixer sur l'ARN et influencer la régulation de l'épissage au site A7.

Pour analyser l'effet de l'ajout de protéines recombinantes sur l'épissage *in vitro* au site A7, j'ai mené des tests d'épissage avec le pré-ARNm L47 WT. La construction permettant de produire cet ARN avait été réalisée par L. Damier. Elle contient les éléments régulateurs ESE3 et ESS3 et possède deux exons et un mini-intron délimités par les sites d'épissage D4 et A7. En effet, dans les étapes précoces de l'infection cellulaire, le site 5' donneur D4 est toujours couplé au site 3' accepteur A7 et l'élimination de l'intron D4-A7 permet la production des ARNm *tat*, *rev* et *nef*.

### **I.1. Effet sur l'épissage de la protéine hnRNP H.**

Les deux motifs RRM N-terminaux sont suffisants pour la fixation de hnRNP H à un ARN de VIH-1 contenant le site A3 accepteur d'épissage, mais cette protéine tronquée ne montre pas d'effet sur l'épissage *in vitro* au site A3 (Thèse H. Hallay, 2006). Nous avons testé l'effet de la protéine hnRNP H RRM 1, 2 sur l'efficacité d'épissage au site A7 *in vitro* et nous avons pu montrer que celle-ci n'est pas affectée, ce malgré la présence d'un site de fixation à proximité du site accepteur d'épissage. Des données récentes (Fisette *et al.*, 2009) montrent que la partie N-terminale de la protéine hnRNP H est nécessaire et suffisante pour la fixation de la protéine à l'ARN, mais la partie C-terminale riche en glycines est indispensable pour l'activité de la hnRNP H lors de l'épissage. Ces travaux n'ont pas été effectués sur VIH-1 mais peuvent expliquer en partie l'absence d'effet lors de nos tests sur l'épissage avec la protéine hnRNP H RRM1, 2, avec le domaine C-terminal tronqué.

### **I.2 Effet sur l'épissage de la nucléoline.**

Nous avons essayé de purifier la nucléoline à partir d'un extrait nucléaire de cellules HeLa, mais les concentrations obtenues étaient très faibles. Nous n'avons pas observé d'effet de la nucléoline purifiée sur l'épissage lors des tests *in vitro* réalisés par C. Aigueperse. La première hypothèse pour expliquer l'absence de résultat pourrait être les trop faibles concentrations utilisées. La nucléoline est présente en grande quantité dans l'extrait nucléaire et pour observer une variation de l'épissage, il est probablement nécessaire d'en rajouter une quantité supérieure à ce que nous avons pu obtenir.

La seconde hypothèse est que même si la nucléoline p30 est capable de se fixer sur l'ARN, un effet sur le métabolisme de l'ARN peut nécessiter la protéine entière. La nucléoline fixée sur l'ARN SLS2-A7 pourrait également être impliquée dans d'autres processus de métabolisme de l'ARN. En effet, elle est engagée dans différents mécanismes lors de l'infection de cellules

par des virus. Il est intéressant de noter que durant l'infection par le poliovirus, la nucléoline relocalise dans le cytoplasme des cellules infectées (Waggoner *et al.*, 1998). De plus, chez les patients infectés par le VIH-1, une sensibilité accrue des cellules T à l'apoptose est associée au phénomène complexe qu'est la dysrégulation du cycle cellulaire (CCD) (Paiardini *et al.*, 2004) qui implique une augmentation de l'activation du complexe cycline B/p34-cdc2 et conduit à une structure nucléolaire anormale avec une dérégulation du turn-over de la nucléoline (Galati *et al.*, 2003). Lors de l'infection par le VIH-1, il est possible que la nucléoline soit relocalisée vers le cytoplasme et ainsi puisse être impliquée dans l'export nucléaire des ARN de VIH-1. D'autre part, la fixation de la nucléoline à l'ARN de VIH-1 peut être importante pour son transit nucléolaire. En effet, il a été démontré qu'un ribozyme localisé dans le nucléole inhibe la réplication de VIH-1, suggérant un trafic nucléolaire de l'ARN VIH-1 (Michienzi *et al.*, 2000).

De plus, l'implication de la nucléoline dans le processus de réplication et de bourgeonnement de VIH-1 est démontrée. La nucléoline a été identifiée comme protéine cellulaire qui lie les protéines Gag du virus MuLV (Bacharach *et al.*, 2000). Les protéines Gag de VIH-1 jouent un rôle essentiel dans le bourgeonnement du virion, dans lequel les motifs en doigt de zinc des protéines Gag reconnaissent le signal de packaging de l'ARN génomique (Psi) (Ueno *et al.*, 2004). La nucléoline active le relargage des virions de VIH-1 qui contiennent Psi. De plus, la nucléoline et les protéines Gag forment un complexe incorporé dans les virions (Ueno *et al.*, 2004). La nucléoline stimule ainsi l'infectivité de VIH-1.

### **I.3 Effet sur l'épissage de la protéine hnRNP K.**

#### **I.3.1 La protéine hnRNP K recombinante active l'épissage d'ARN pré-messagers témoins.**

Afin de confirmer que la protéine hnRNP K recombinante que j'ai purifiée est fonctionnelle, A. Expert-Bezançon nous a fourni différents ARN de  $\beta$ -tropomyosine (Expert-Bezançon *et al.*, 2002).

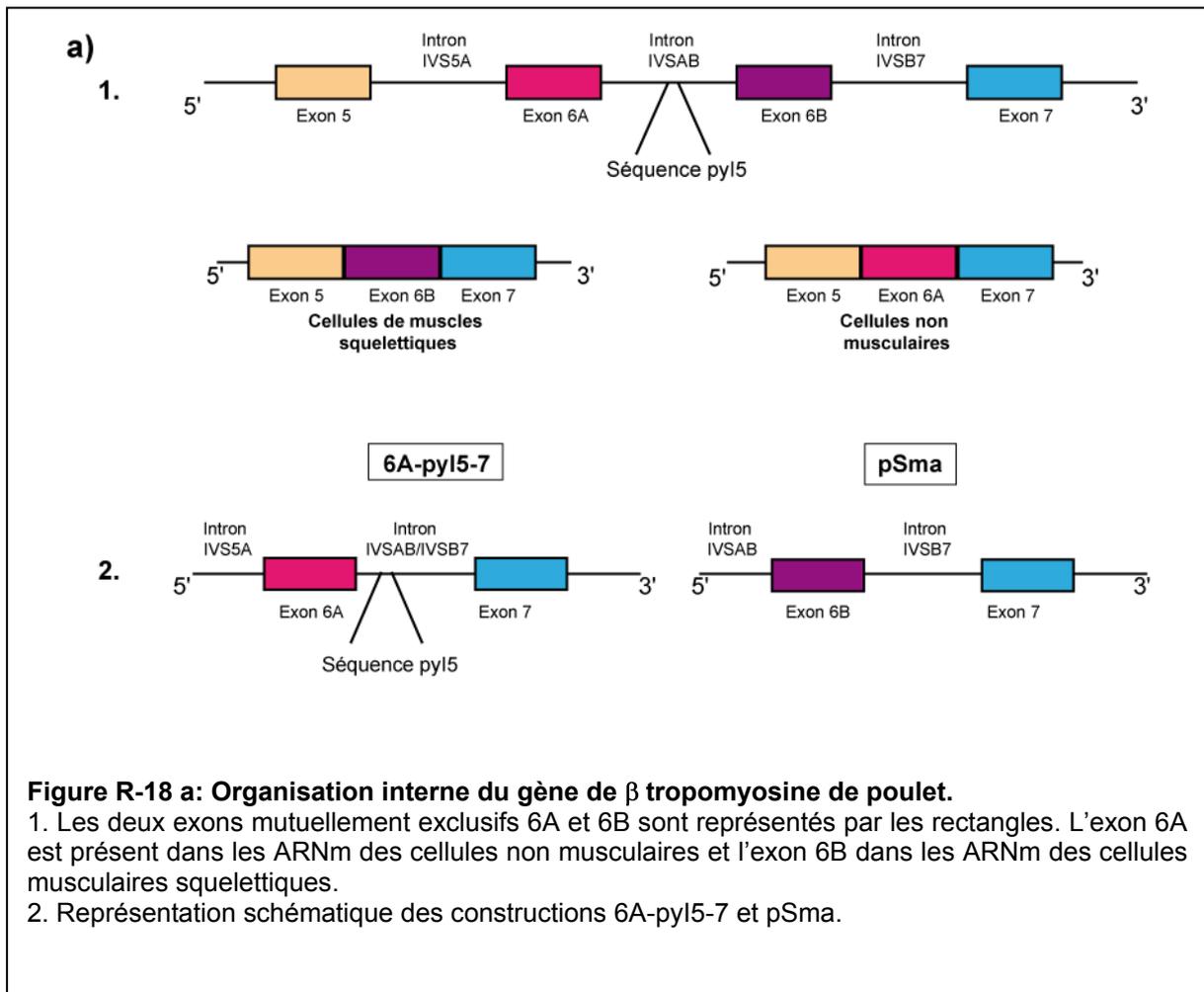
Le pré-ARNm de la  $\beta$ -tropomyosine (6A-pyI5-7) de poulet fournit un exemple de la complexité de la régulation de l'épissage alternatif. Ce transcrit contient 2 exons mutuellement exclusifs (6A et 6B) qui sont sélectionnées de façon tissu-spécifique (Libri *et al.*, 1991) (Figure R-18a). En épissage *in vitro* dans un extrait nucléaire de cellules HeLa, l'exon 6A est inclus et l'exon 6B est exclu (Goux-Pelletan *et al.*, 1990). La reconnaissance de l'exon 6A requiert un élément intronique activateur riche en pyrimidines nommé S4 (Del

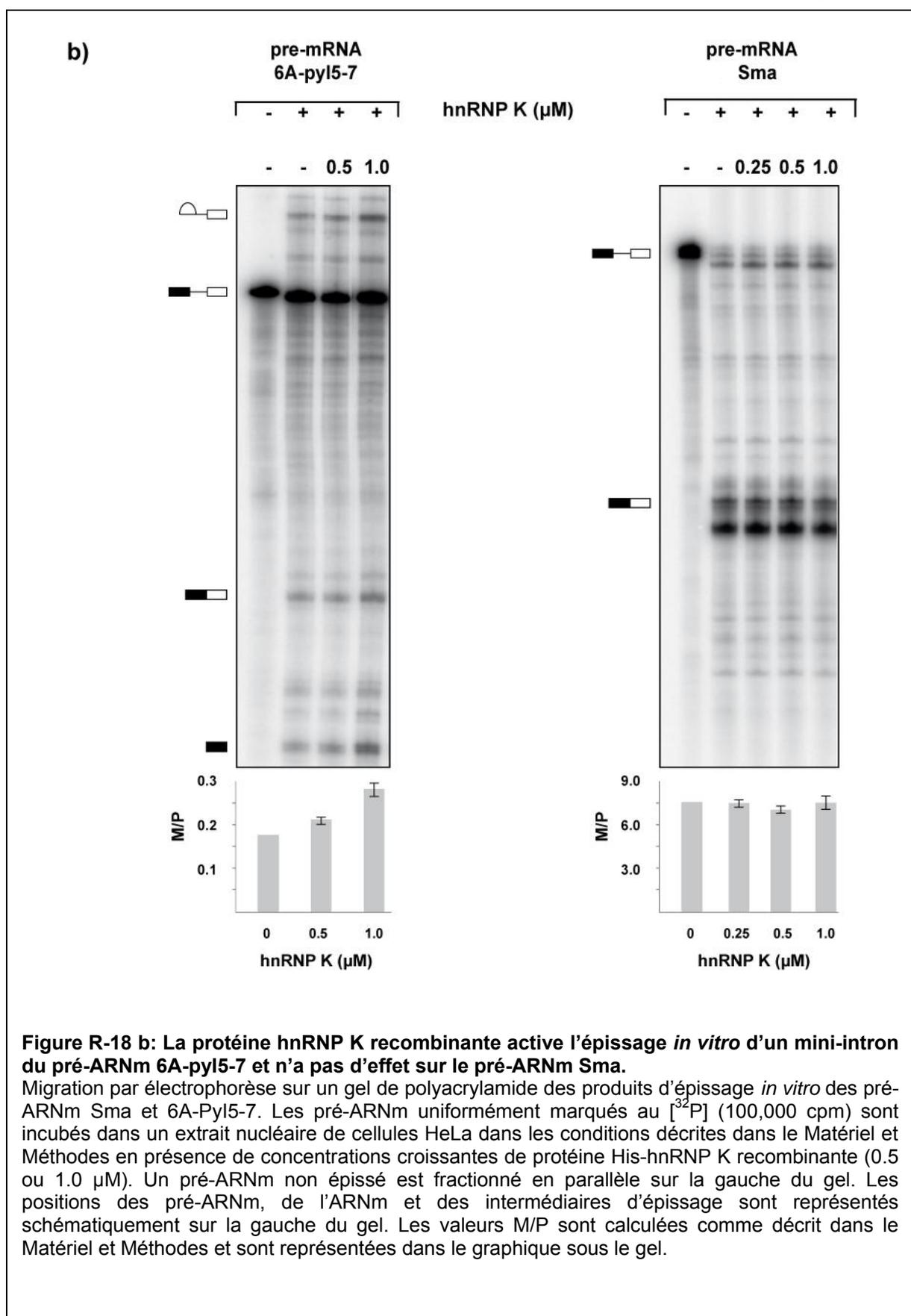
Gatto-Konczak *et al.*, 2000) placé 37 nucléotides en aval du site 5' de l'exon 6A (Balvay *et al.*, 1992; Gallego *et al.*, 1992). Cette séquence activatrice exerce aussi un contrôle négatif sur la reconnaissance de l'exon 6B par un mécanisme pour l'instant inconnu (Balvay *et al.*, 1992).

L'ajout de protéine hnRNP K recombinante à un extrait nucléaire de cellules HeLa augmente l'efficacité de l'épissage du pré-ARNm 6A-pyI5-7 *in vitro*. Cet ARN pré-messager nous sert de témoin positif pour vérifier l'activité de la protéine hnRNP K en épissage *in vitro*.

L'épissage de l'ARN produit à partir du plasmide pSma est insensible à l'ajout de la protéine hnRNP K. Cet ARN pré-messager nous sert de témoin négatif pour vérifier que l'effet observé de la protéine hnRNP K n'est pas dû à un contaminant dans le tampon dans lequel elle se trouve.

Différentes quantités testées de la protéine hnRNP K recombinante produite au laboratoire ne présentent pas d'effet sur l'épissage *in vitro* de la construction pSma. Au contraire, l'épissage de l'ARN 6A-pyI5-7 est activé par la protéine hnRNP K comme décrit par A. Expert-Bezançon (Expert-Bezançon *et al.*, 2002). Les résultats obtenus (Figure R-18b) montrent ainsi que la protéine recombinante hnRNP K produite en bactéries *E. Coli* et purifiée est bien active en épissage *in vitro*.

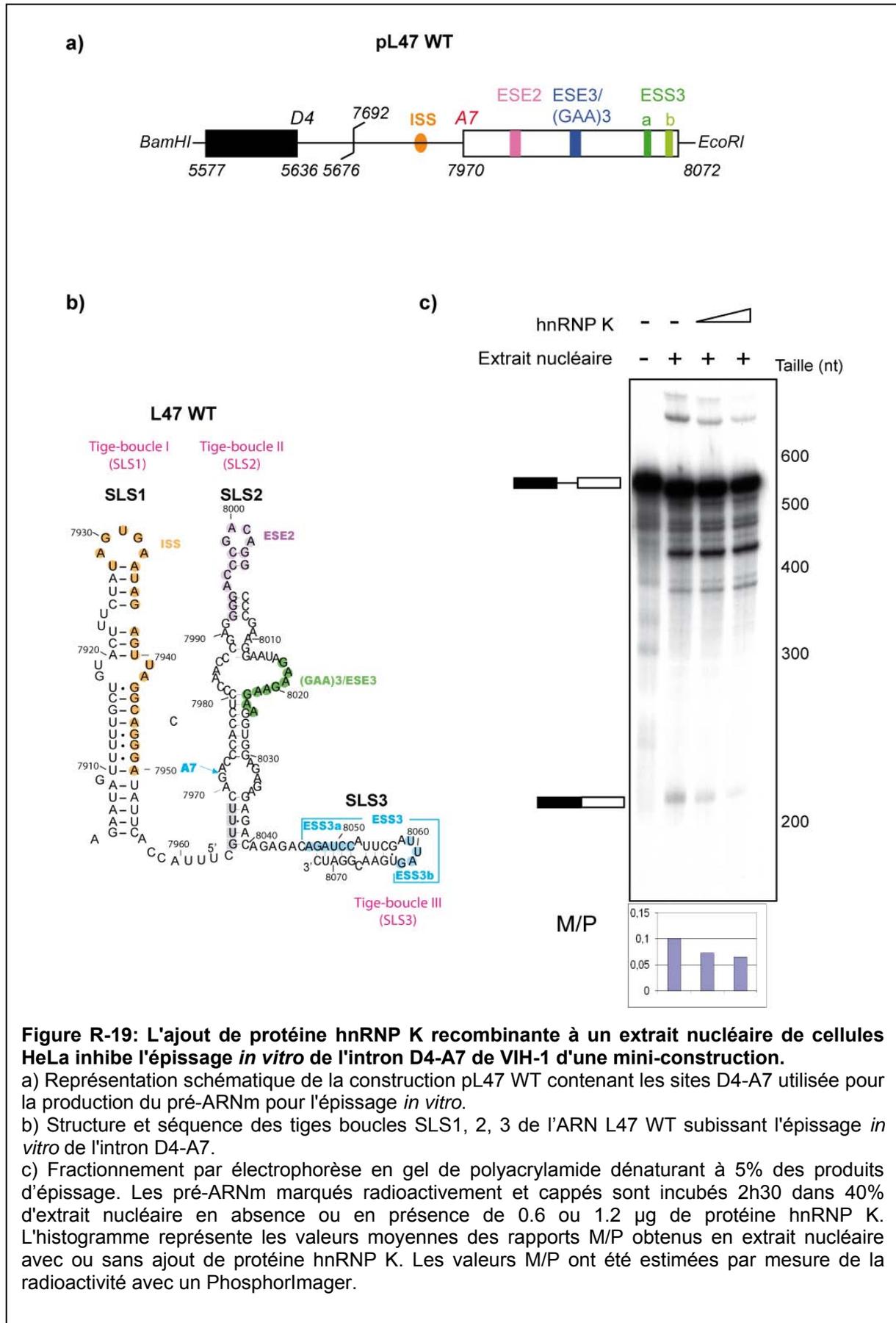




### **I.3.2 La protéine hnRNP K recombinante diminue l'efficacité d'épissage alternatif du site A7 de l'ARN pré-messager L47 WT.**

Le travail sur l'identification des sites de fixation de la protéine hnRNP K sur l'ARN SLS1, 2, 3 a montré que la protéine se fixe sur des éléments régulateurs importants de l'épissage (Chapitre II Résultats, Figure R-12). Pour analyser l'effet de l'ajout de la protéine hnRNP K sur l'épissage *in vitro* au site A7, j'ai réalisé des expériences d'épissage *in vitro* sur l'ARN L47 WT en rajoutant des quantités croissantes de hnRNP K à l'extrait nucléaire. Les résultats montrent que le taux relatif d'épissage diminue d'environ 20% avec 600 ng de protéine et de 40% avec 1.2 µg (Figure R-19).

La quantité de protéine hnRNP K ajoutée peut sembler importante par rapport à la quantité de protéines SR habituellement ajoutée, mais il faut prendre en compte le fait que la protéine hnRNP K a un rôle dans de nombreux mécanismes cellulaires. En raison des propriétés de plateforme d'amarrage de la protéine, une fraction de hnRNP K ajoutée à un extrait nucléaire de cellules HeLa va être impliquée dans des complexes autres que ceux d'épissage et ne sera ainsi pas accessible pour la régulation de l'épissage (Mikula *et al.*, 2006). Les expériences de retard sur gel présentés dans les figures R-10 à 12 montrent que l'affinité de la protéine hnRNP K est très bonne pour l'ARN SLS2-A7 contenant le site A7 d'épissage. Cette protéine a été purifiée sur colonne d'héparine, ce qui va sélectionner des protéines ayant des qualités de fixation à l'ARN, en éliminant les protéines mal repliées qui n'en seraient pas capables. En effet, la purification de la protéine ayant été réalisée sur colonne d'héparine, seules les protéines ayant des qualités de fixation à l'ARN seront sélectionnées. Les protéines hnRNP K mal repliées sont quant à elles éliminées par ce mode de purification.



**Figure R-19: L'ajout de protéine hnRNP K recombinante à un extrait nucléaire de cellules HeLa inhibe l'épissage *in vitro* de l'intron D4-A7 de VIH-1 d'une mini-construction.**

a) Représentation schématique de la construction pL47 WT contenant les sites D4-A7 utilisée pour la production du pré-ARNm pour l'épissage *in vitro*.

b) Structure et séquence des tiges boucles SLS1, 2, 3 de l'ARN L47 WT subissant l'épissage *in vitro* de l'intron D4-A7.

c) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage. Les pré-ARNm marqués radioactivement et cappés sont incubés 2h30 dans 40% d'extrait nucléaire en absence ou en présence de 0.6 ou 1.2  $\mu\text{g}$  de protéine hnRNP K. L'histogramme représente les valeurs moyennes des rapports M/P obtenus en extrait nucléaire avec ou sans ajout de protéine hnRNP K. Les valeurs M/P ont été estimées par mesure de la radioactivité avec un PhosphorImager.

### **I.3.3 L'effet inhibiteur de hnRNP K est maintenu indépendamment des modifications de séquence de l'ARN.**

Par leur fixation à proximité ou sur les sites d'épissage, les protéines hnRNP peuvent par encombrement stérique, limiter la fixation des facteurs généraux ou des facteurs activateurs d'épissage. Leurs séquences de fixation constituent alors des séquences régulatrices de l'épissage. Des mutations dans les différentes séquences régulatrices connues de l'ARN L47 WT permettent d'étudier le rôle de chacune d'entre elles dans l'inhibition de l'épissage par la protéine hnRNP K. Chaque ARN testé présente une mutation d'une séquence régulatrice mutée à la fois.

- Dans la tige-boucle 1, l'ISS est un site connu de fixation de la protéine hnRNP A1. En se fixant à ce site, la protéine hnRNP A1 va créer un encombrement stérique près des points de branchement du site A7 sur lesquels se fixe la snRNP U2, et ainsi inhiber l'épissage (Tange *et al.*, 2001). Il a été décrit que la délétion complète de l'ISS diminue l'inhibition de l'épissage par la protéine hnRNP A1. Les expériences d'empreintes ont montré que la protéine hnRNP K a un site de fixation sur les nucléotides GUGA (de 7930 à 7933) de l'élément ISS (Figure R-12). Nous disposions au laboratoire d'un ARN avec les bases de 7929 à 7932 AGUG mutées en GCGC (Figure R-20), donnant l'ARN L47-mISS. Le taux basal d'épissage de l'intron D4-A7 dans ce contexte est plus élevé d'environ 50 % par rapport à la construction sauvage, confirmant la perte d'un élément inhibiteur de l'épissage. Cependant, l'ajout de protéine hnRNP K recombinante conduit à la même diminution de l'efficacité d'épissage que la séquence ISS soit mutée ou non. Il est étonnant de constater que la mutation de cet élément inhibiteur fixe la protéine hnRNP K mais que sa délétion n'a pas d'effet sur l'efficacité d'inhibition de la protéine (Figure R-21).

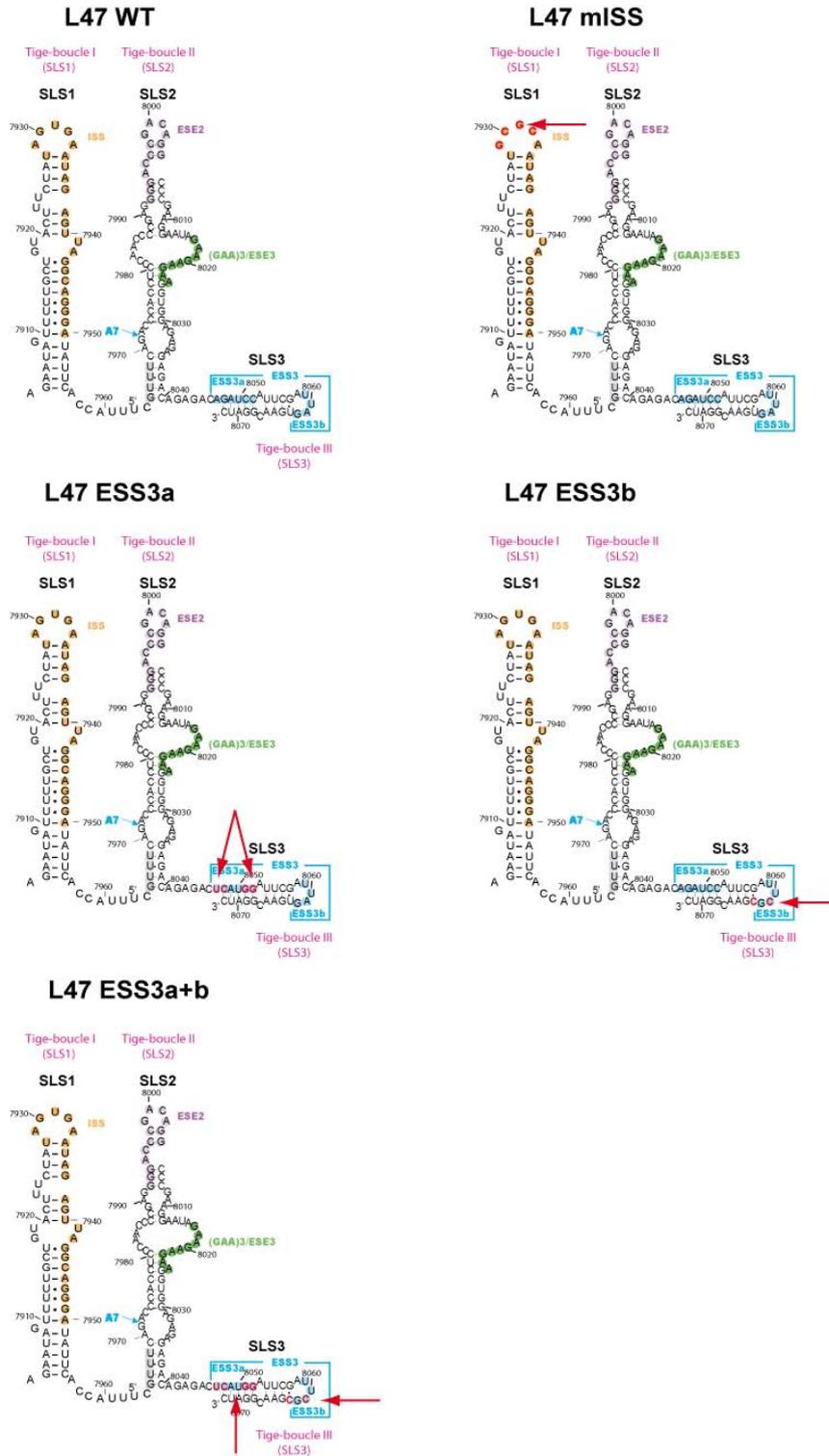
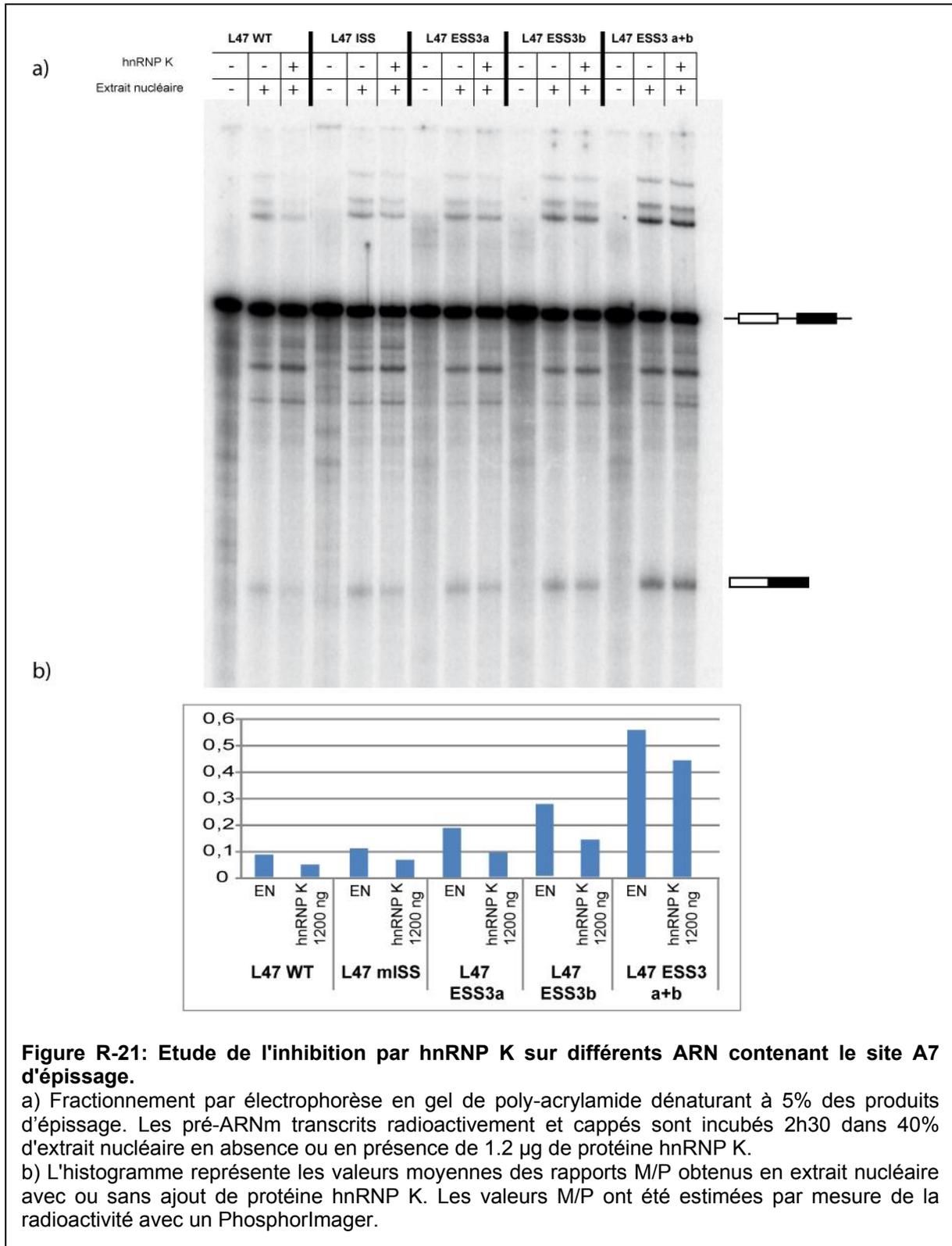


Figure R-20 : Représentation schématique de la structure 2D des ARN L47 sauvage et mutés utilisés.

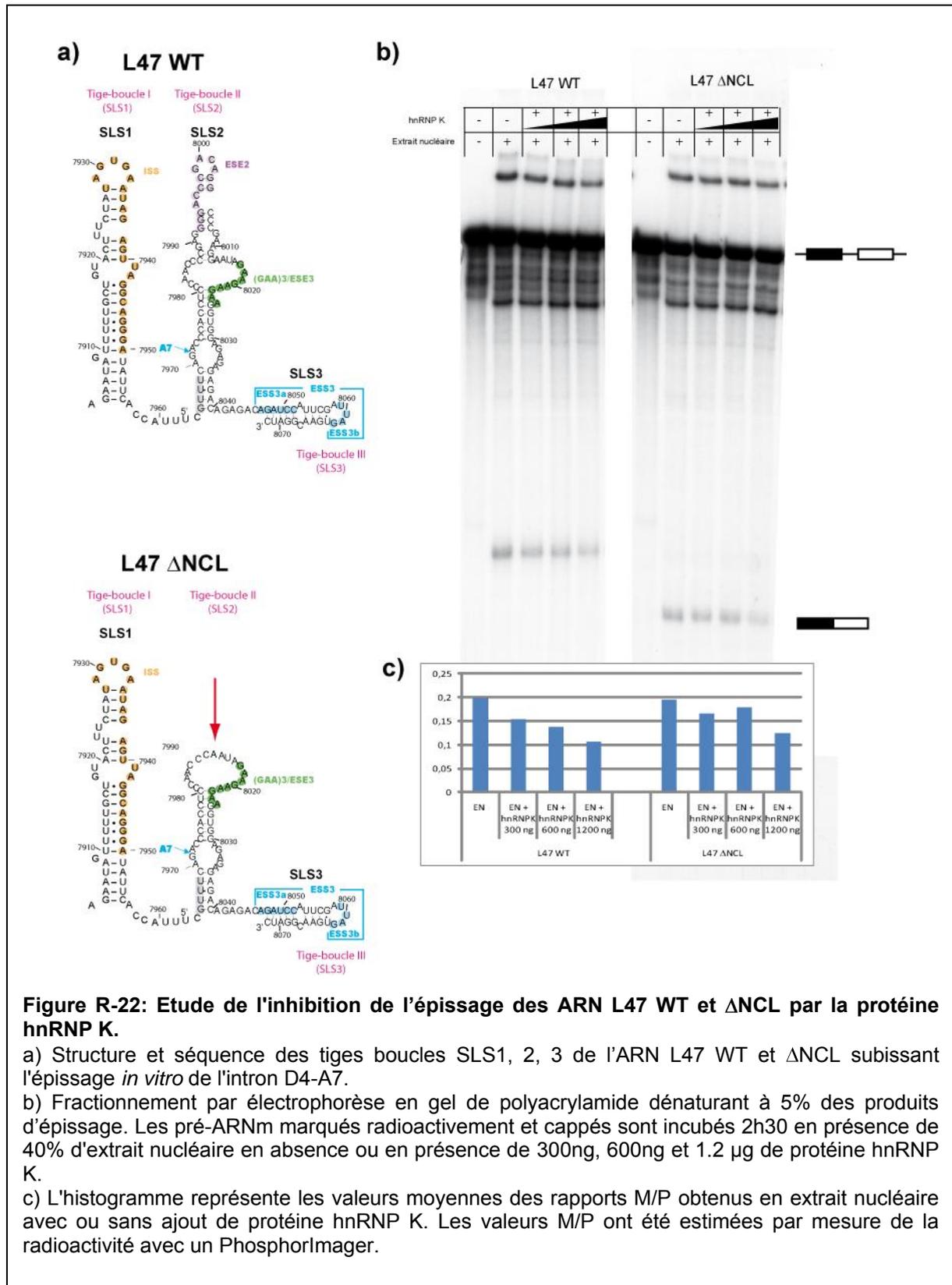


- L'élément ESS3 (Figure R-20) est composé de deux séquences (AGAUCC : ESS3a) et (UUAG : ESS3b) qui peuvent inhiber l'épissage indépendamment l'une de l'autre (Si *et al.*, 1998) et leurs effets sont cumulatifs. Les protéines hnRNP A/B liées à l'élément ESS3 sont capables d'oligomériser et de se propager en amont de l'élément ESS3 (Zhu *et al.*, 2001; Okunola *et al.*, 2009).

Lorsque les deux éléments ESS3a et l'ESS3b sont mutés indépendamment l'un de l'autre ou bien lorsqu'ils sont mutés simultanément, on observe une importante augmentation du taux basal de l'épissage engendrée par la perte de leur capacité à fixer des protéines inhibitrices de l'épissage. Les taux d'inhibition de l'épissage par la protéine hnRNP K des ARN mutants L47-ESS3a, L47-ESS3b et L47-ESS3a+b sont identiques à ce qui est observé pour l'ARN sauvage L47-WT. La délétion du site qui fixe pourtant la protéine hnRNP K (nucléotides 8057 et 8062 de l'ESS3b) n'affecte pas l'inhibition de l'épissage par hnRNP K (Figure R-21).

- Dans la tige-boucle 2 contenant le site A7 d'épissage, le haut de la tige-boucle a été délété (L47- $\Delta$ NCL) des nucléotides 7992 à 8006 inclus. Le taux basal d'épissage observé pour le mutant L47- $\Delta$ NCL reste inchangé par rapport à celui du L47-WT et l'inhibition de l'épissage est globalement très légèrement diminuée par rapport à l'inhibition obtenue avec l'ARN sauvage (Figure R-22). Nous nous attendions à une diminution plus franche de l'inhibition par hnRNP K. En effet, nous avons montré par empreintes que la protéine hnRNP K se fixe à l'extrémité de la tige-boucle 2 et protégeait en particulier le nucléotide 8000 de l'ARN (Figure R-12).

- Une délétion des nucléotides 7980 à 7987 (L47- $\Delta$ K) contenant deux séquences poly-C en simple brin a été effectuée. Malgré la présence de ces deux séquences poly-C, les empreintes enzymatiques ont montré que cette séquence n'était pas un site de fixation de la protéine hnRNP K (Figure R-23).



**Figure R-22: Etude de l'inhibition de l'épissage des ARN L47 WT et ΔNCL par la protéine hnRNP K.**

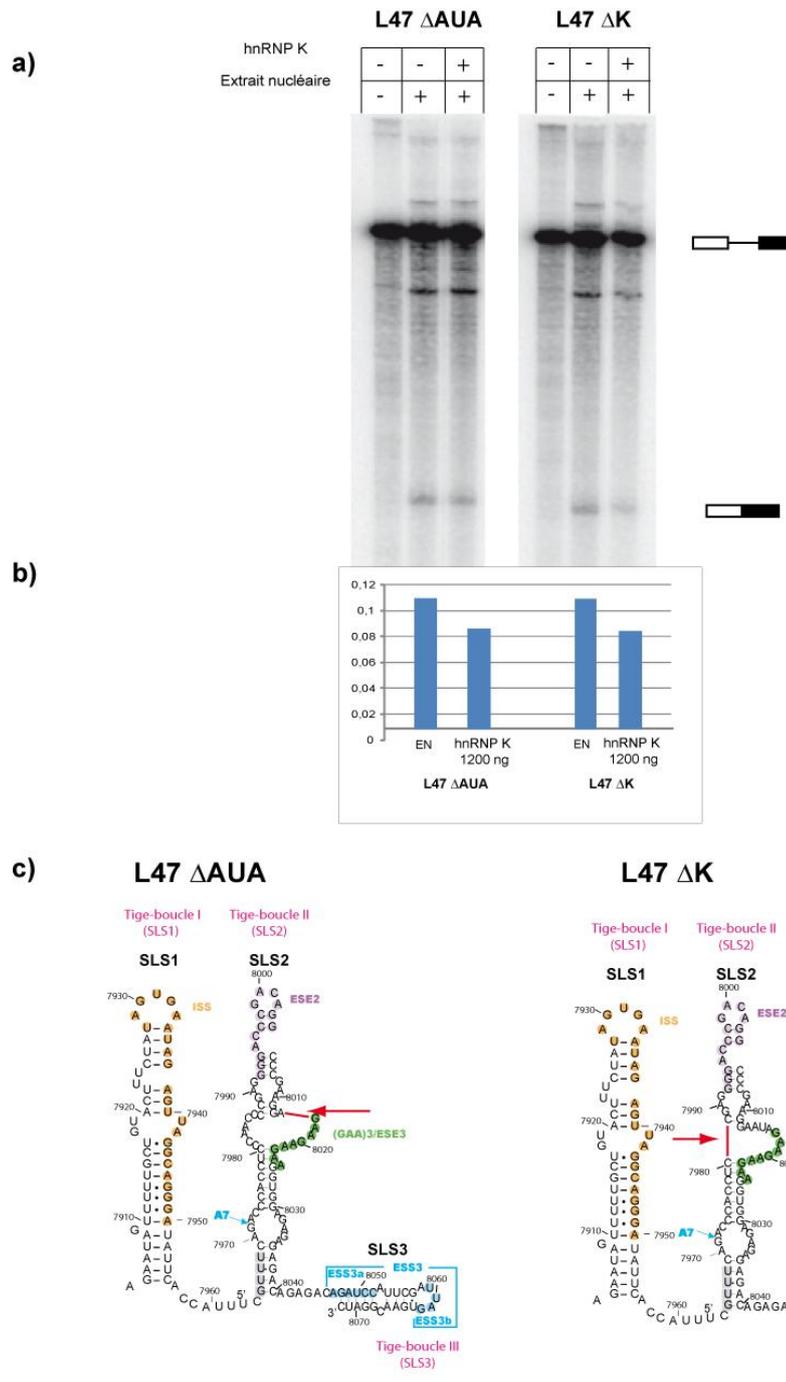
a) Structure et séquence des tiges boucles SLS1, 2, 3 de l'ARN L47 WT et ΔNCL subissant l'épissage *in vitro* de l'intron D4-A7.

b) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage. Les pré-ARNm marqués radioactivement et cappés sont incubés 2h30 en présence de 40% d'extrait nucléaire en absence ou en présence de 300ng, 600ng et 1.2 μg de protéine hnRNP K.

c) L'histogramme représente les valeurs moyennes des rapports M/P obtenus en extrait nucléaire avec ou sans ajout de protéine hnRNP K. Les valeurs M/P ont été estimées par mesure de la radioactivité avec un PhosphorImager.

Nous avons tout de même testé l'effet de la délétion de ces séquences sur l'épissage et là encore nous n'avons pas observé de différence d'efficacité d'épissage en extrait nucléaire ni de différence d'inhibition par hnRNP K.

- Le taux basal d'épissage du mutant L47- $\Delta$ AUA est augmenté d'environ 50 % par rapport à l'ARN L47 WT. Cependant, l'ajout de protéine hnRNP K recombinante conduit à la même diminution de l'efficacité d'épissage (30%) lorsque la séquence AUA est délétée (Figure R-23).
- Le taux basal d'épissage de l'ARN pré-messager  $\Delta$ GAA (Figure R-24) est un peu plus faible que celui de l'ARN sauvage. La séquence (GAA)<sub>3</sub> de fixation de la protéine ASF/SF2 permettant l'activation de l'épissage au site A7 n'étant plus là, cela peut expliquer cette diminution. On observe avec ce mutant une diminution de l'efficacité d'épissage lors de l'ajout de la protéine hnRNP K d'environ 30% (Figure R-25).
- La séquence appelée Janus [AUA(GAA)<sub>3</sub>] (Marchand *et al.*, 2002) a été délétée des nucléotides 8013 à 8025 et remplacée par 3 cytidines (construction L47-3C) (Figure R-24). L'ARN appelé 3C présente une efficacité d'épissage très supérieure à l'ARN non muté mais une inhibition de l'épissage par hnRNP K identique à l'ARN L47 WT (Figure R-25). C. Aigueperse a évalué le taux d'activation du mutant 3C par la protéine ASF/SF2 en comparaison avec celui de l'ARN sauvage (résultat non présenté). L'activation de l'épissage de l'ARN 3C par la protéine ASF/SF2 est meilleure. Ce résultat est surprenant car le site de fixation consensus (GAAGAAGAA) de la protéine ASF/SF2 a été retiré et les nucléotides CCC ne fixent pas mieux ASF/SF2. Le site de fixation d'ASF/SF2 dans la partie apicale de l'ARN semble donc être suffisant pour l'activation au site A7. Le mutant 3C lie beaucoup mieux la protéine hnRNP K, ce qui est attendu avec cette protéine liant les poly-C. On s'attendait donc à une augmentation de l'inhibition par la protéine hnRNP K de ce mutant qui fixe beaucoup mieux cette protéine et non pas à la même inhibition qu'avec l'ARN sauvage.



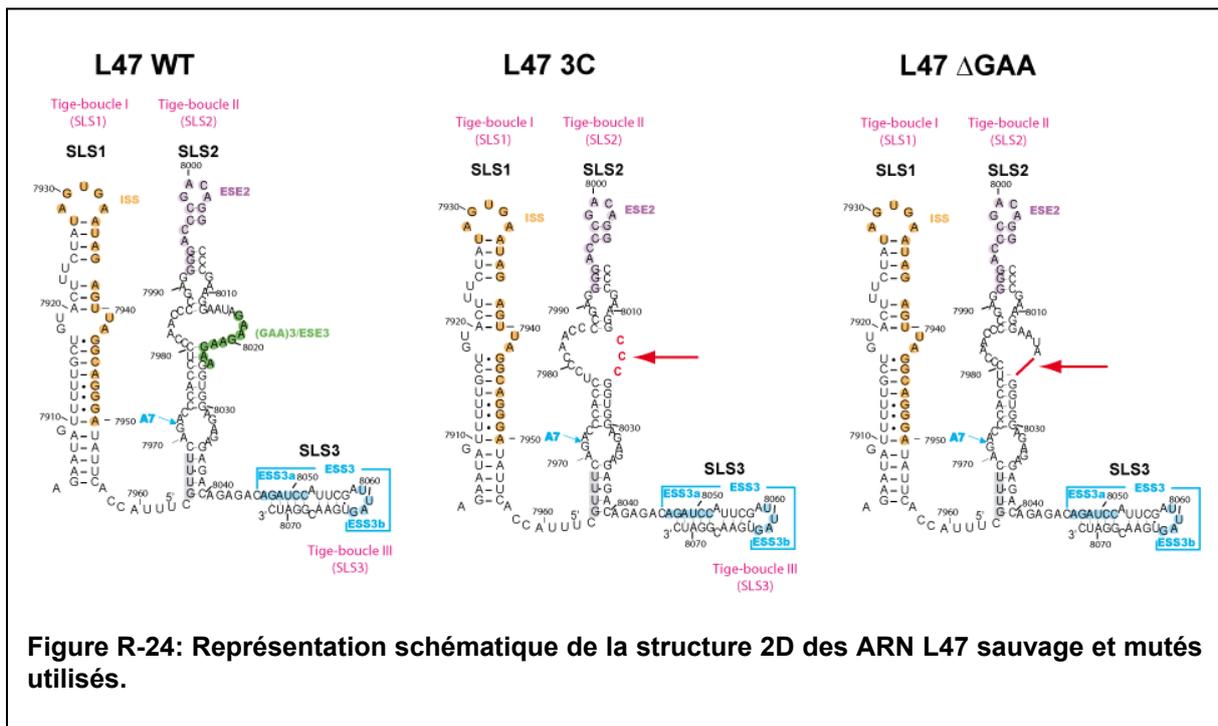
**Figure R-23: Etude de l'inhibition par hnRNP K sur les mutants ARN  $\Delta$ AUA et  $\Delta$ K autour du site A7 d'épissage.**

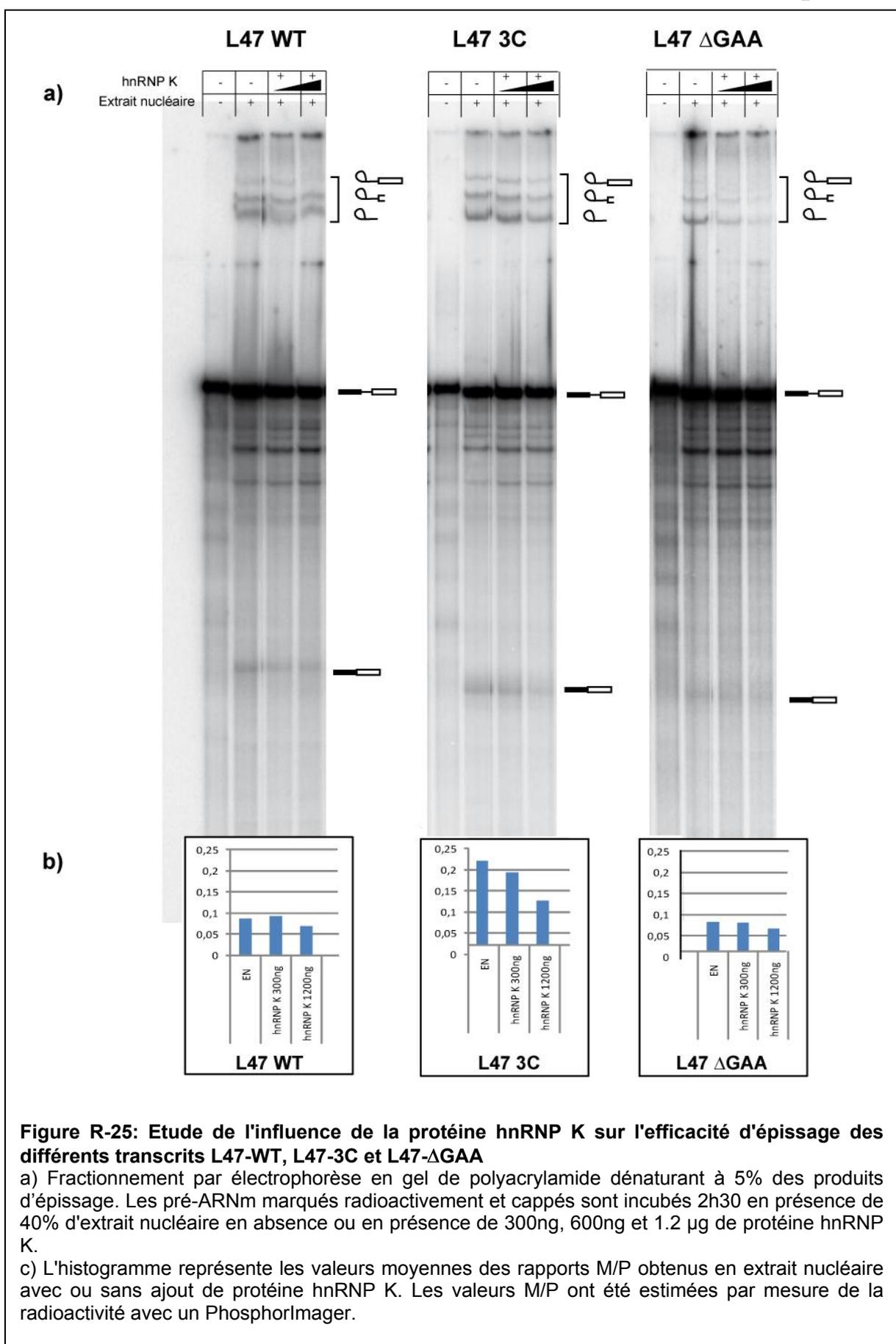
a) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage. Les pré-ARNm marqués radioactivement et cappés sont incubés 2h30 en présence de 40% d'extrait nucléaire en absence ou en présence de 300ng, 600ng et 1.2  $\mu$ g de protéine hnRNP K.

b) L'histogramme représente les valeurs moyennes des rapports M/P obtenus en extrait nucléaire avec ou sans ajout de protéine hnRNP K. Les valeurs M/P ont été estimées par mesure de la radioactivité avec un PhosphorImager.

c) Structure et séquence des tiges boucles SLS1, 2, 3 de l'ARN L47  $\Delta$ AUA et  $\Delta$ K subissant l'épissage *in vitro* de l'intron D4-A7.

Les expériences d’empreintes décrites précédemment ont montré que la protéine hnRNP K présente quatre sites de fixation sur les trois tige-boucles environnant le site A7. Ces sites se situent dans l’ISS de la tige-boucle 1, dans l’ESE2 et l’ESE3 de la tige-boucle 2 et dans l’ESS3b de la tige-boucle 3. Nos résultats d’expérience d’épissage *in vitro* montrent que le fait de déléter l’un de ces sites de fixation ne perturbe pas le mécanisme d’inhibition de l’épissage de la protéine hnRNP K ou trop peu pour être mis en évidence. Les capacités de fixation de la protéine hnRNP K à des séquences ARN variées, comme le montrent différentes expériences de SELEX et de SAGE (Klimek-Tomczak *et al.*, 2004), pourraient limiter l’effet des mutations individuelles. De plus, les facteurs régulateurs comme les protéines SR ou hnRNP ayant souvent des sites de fixation dégénérés, toutes les mutations ne vont pas altérer de la même façon l’activité d’un élément régulateur. Il faut noter aussi que le taux basal d’épissage du site A7 étant bas, de trop faibles variations peuvent être difficiles à mettre en évidence. Aucune des modifications de séquence de l’ARN testées n’a donc pu mettre en évidence une perte franche de l’effet inhibiteur de hnRNP K sur l’épissage.



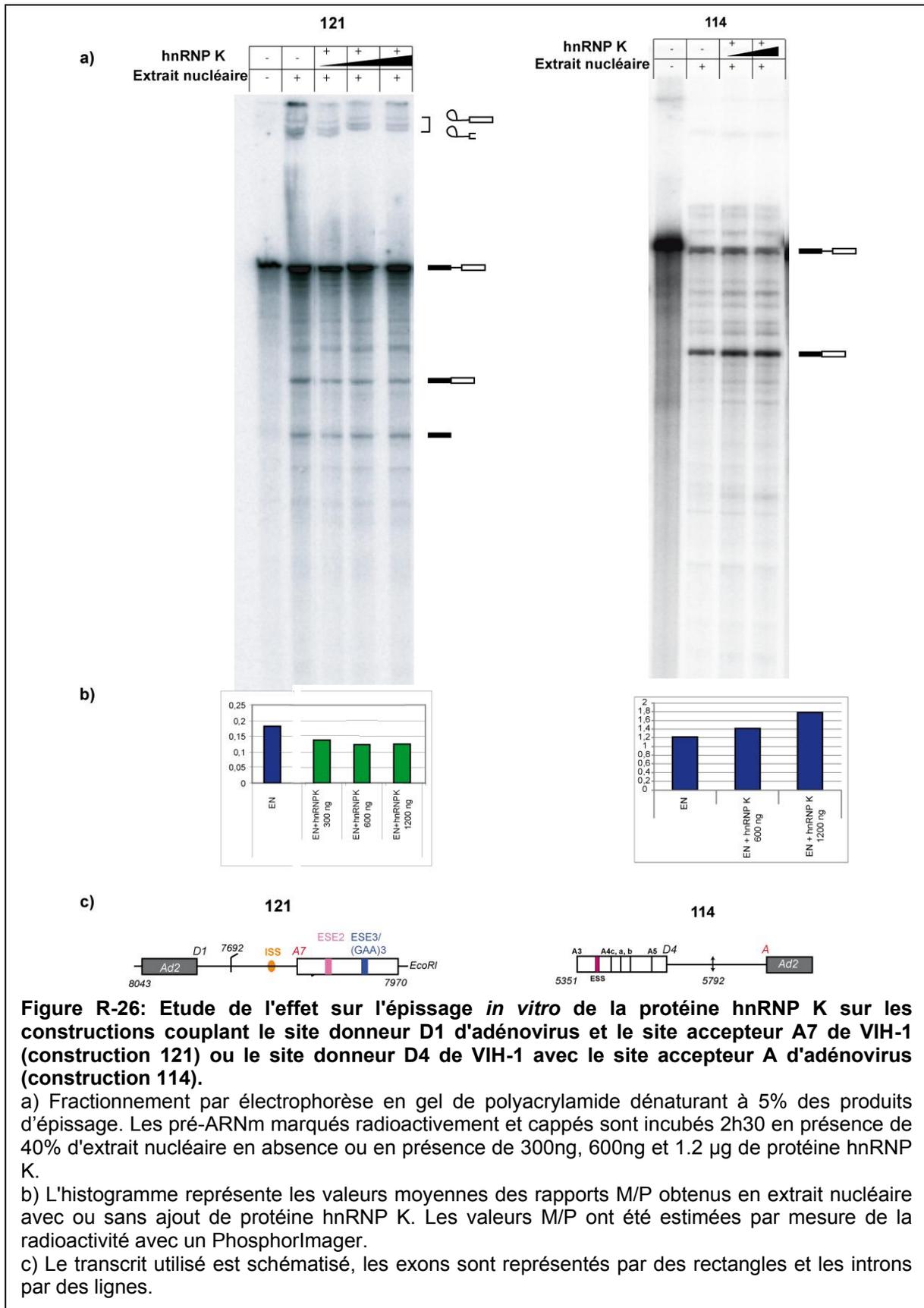


### **I.3.4 La régulation de l'épissage alternatif par la protéine hnRNP K est bien liée au site 3' accepteur A7.**

L'épissage alternatif peut être régulé par des fixations de protéines en *trans* sur des séquences régulatrices, tant au niveau du site donneur que du site accepteur qui vont influencer le mécanisme de définition des exons. Un moyen d'apporter des informations sur l'efficacité relative des sites donneurs d'une part et des sites accepteurs d'autre part était de coupler chacun des sites donneurs du VIH-1 à un même site accepteur fort et *vice versa* de coupler chacun des sites accepteurs de VIH-1 à un même site donneur fort. C'est dans ce but que des constructions hybrides adénovirus/VIH-1 avaient été réalisées au laboratoire par L. Damier (Figure R-26 c). Dans ces constructions, le site donneur fort Ad2D et le site accepteur fort Ad2A sont utilisés.

Pour m'assurer que l'effet observé de l'inhibition de l'épissage de l'intron D4-A7 par la protéine hnRNP K est lié à sa fixation autour du site A7 et n'a pas de lien avec le site donneur D4, j'ai utilisé la construction 121 dans laquelle l'exon 1 avec le site donneur D4 de VIH-1 est échangé contre l'exon 2 d'adénovirus avec son site donneur Ad2D (Figure R-26). L'inhibition de l'efficacité d'épissage provoquée par l'ajout de protéine hnRNP K sur cet ARN est la même que celle observée pour l'ARN L47-WT. Le changement de site donneur n'a donc pas eu d'effet sur l'inhibition de l'épissage par hnRNP K, ce qui confirme que son action inhibitrice s'exerce bien sur le site 3' accepteur A7.

Nous avons ensuite testé l'épissage *in vitro* de l'ARN issu de la construction 114 qui permet d'observer le couplage du site donneur D4 du VIH-1 avec le site accepteur de l'adénovirus Ad2A. On observe dans ce cas une augmentation de l'efficacité d'épissage *in vitro* lors de l'ajout de la protéine hnRNP K (Figure R-26). L'ensemble de ces résultats montrent donc que l'effet inhibiteur de la protéine hnRNP K sur l'épissage D4-A7 est lié spécifiquement au site accepteur d'épissage A7.



**Figure R-26: Etude de l'effet sur l'épissage *in vitro* de la protéine hnRNP K sur les constructions couplant le site donneur D1 d'adénovirus et le site accepteur A7 de VIH-1 (construction 121) ou le site donneur D4 de VIH-1 avec le site accepteur A d'adénovirus (construction 114).**

a) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage. Les pré-ARNm marqués radioactivement et cappés sont incubés 2h30 en présence de 40% d'extrait nucléaire en absence ou en présence de 300ng, 600ng et 1.2 µg de protéine hnRNP K.

b) L'histogramme représente les valeurs moyennes des rapports M/P obtenus en extrait nucléaire avec ou sans ajout de protéine hnRNP K. Les valeurs M/P ont été estimées par mesure de la radioactivité avec un PhosphorImager.

c) Le transcrit utilisé est schématisé, les exons sont représentés par des rectangles et les introns par des lignes.

## **II. Etude de l'effet de la protéine hnRNP K recombinante sur l'efficacité d'utilisation des sites accepteurs d'épissage des sites A1, A2, A3 et A5 de VIH-1 *in vitro*.**

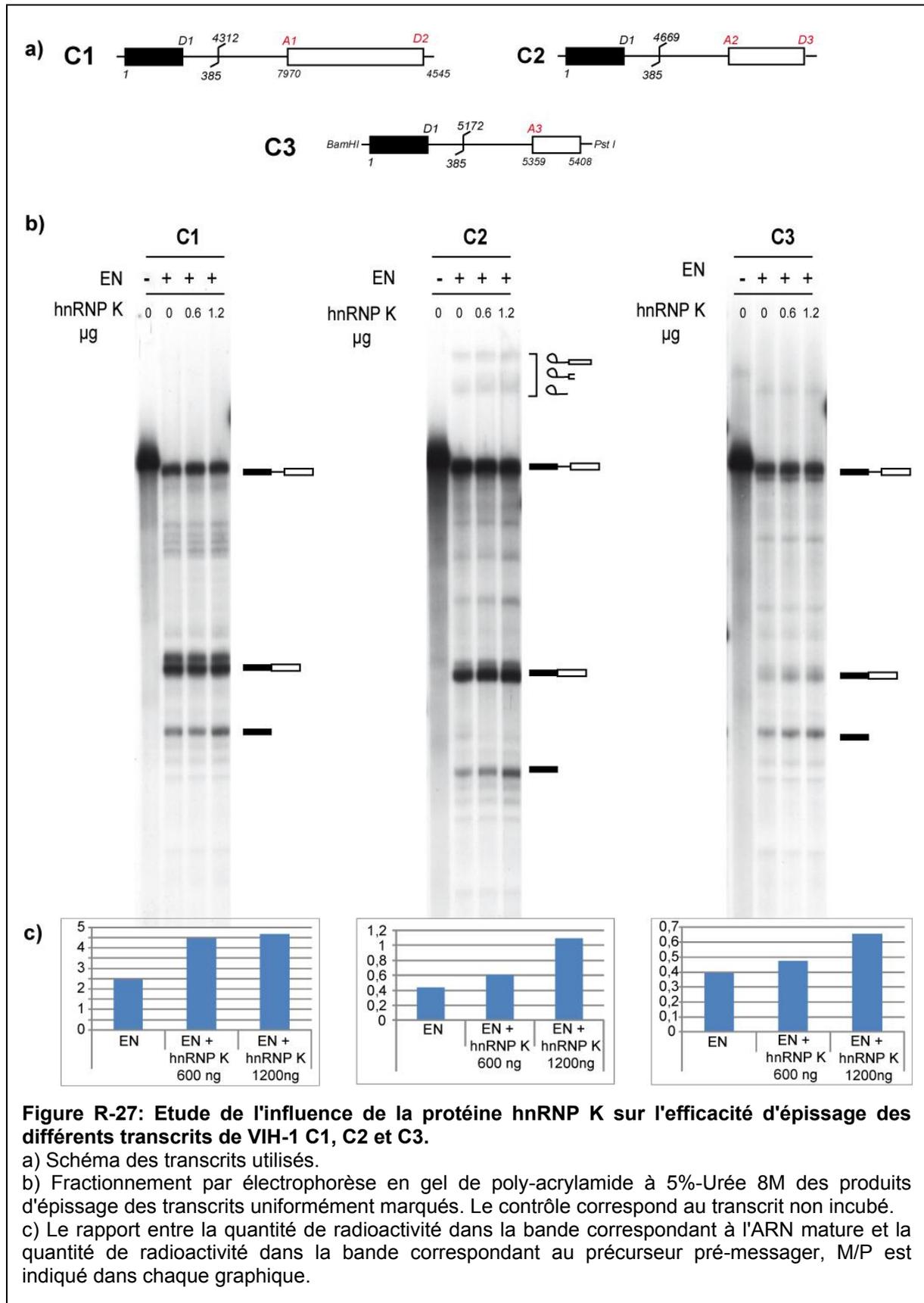
Après avoir montré l'effet inhibiteur de la protéine hnRNP K sur l'épissage au site A7, il était intéressant de regarder son effet sur les autres sites accepteurs d'épissage A1, A2, A3 et A5 du VIH-1. En effet, les protéines régulatrices de l'épissage du VIH-1 interviennent fréquemment dans la régulation de plusieurs sites et parfois de manière différente.

Pour cela, nous avons utilisé des constructions qui avaient été produites au laboratoire par L. Damier (Thèse Damier 1997) dans lesquelles le site donneur D1, le plus efficace des sites donneurs du VIH, est couplé à un ou plusieurs sites accepteurs. Dans le contexte naturel de l'ARN VIH-1, le site D1 est très distant des sites accepteurs. Par conséquent, seuls des mini-introns où une grande partie de la région interne de l'intron est délétée peuvent être utilisés pour tester l'efficacité des couples D1-A1, D1-A2, D1-A3 et D1-A5.

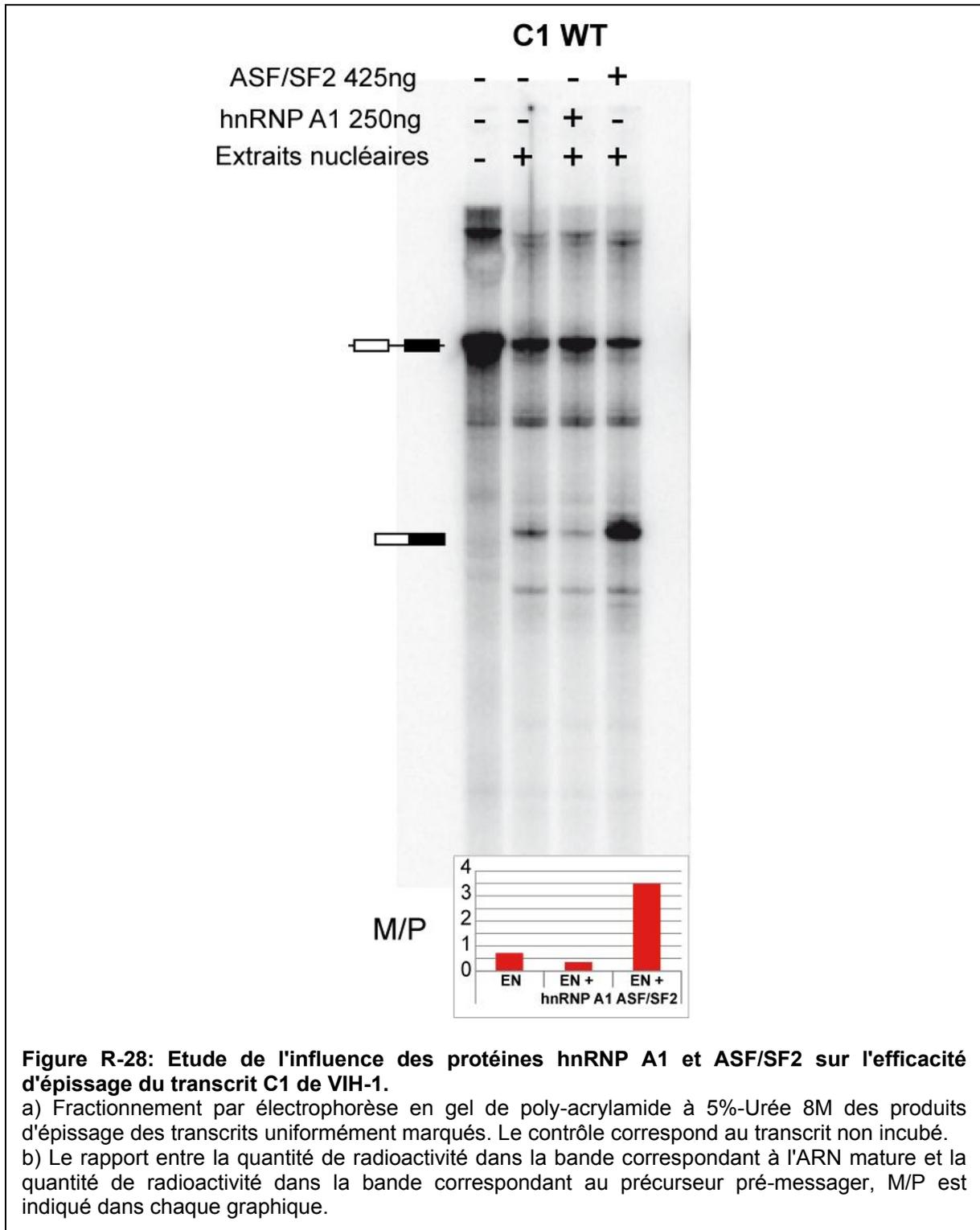
### **II.1 Etude de l'effet de la protéine recombinante hnRNP K sur l'épissage *in vitro* du site accepteur A1 de VIH-1.**

L'utilisation de ce site A1 en combinaison avec le site donneur D1 est nécessaire pour la production des ARNm codant la protéine Vif. Les protéines impliquées dans la régulation de l'épissage au site A1 ont été peu étudiées. La faible efficacité du site 5' donneur en aval D2 est impliquée dans la faible utilisation du site A1 (Stoltzfus, 2009). L'équipe de M. Stoltzfus (University of Iowa, USA) a montré qu'à la fois un élément inhibiteur GGGG situé immédiatement en aval du site D2 et un élément activateur trouvé en aval du site A1 qui lie la protéine SRp75, modulent l'utilisation du site D2 et en conséquence du site A1 (Exline *et al.*, 2008). Les partenaires de l'inhibiteur GGGG n'ont pas encore été identifiés.

L'ajout de 0.6 µg et de 1.2 µg de protéine hnRNP K ne montre pas de variation significative de l'efficacité d'épissage de l'intron D1-A1 de l'ARN C1 (Figure R-27).



J'ai testé en parallèle l'effet de la protéine hnRNP A1 sur ce site d'épissage, et elle présente un effet inhibiteur de l'épissage comme sur tous les autres sites d'épissage sur lesquelles elle a été testée (Figure R-28). Il est intéressant de noter aussi que l'ajout de protéine ASF/SF2 à l'extrait nucléaire provoque une nette augmentation de l'épissage *in vitro* de l'ARNm C1.



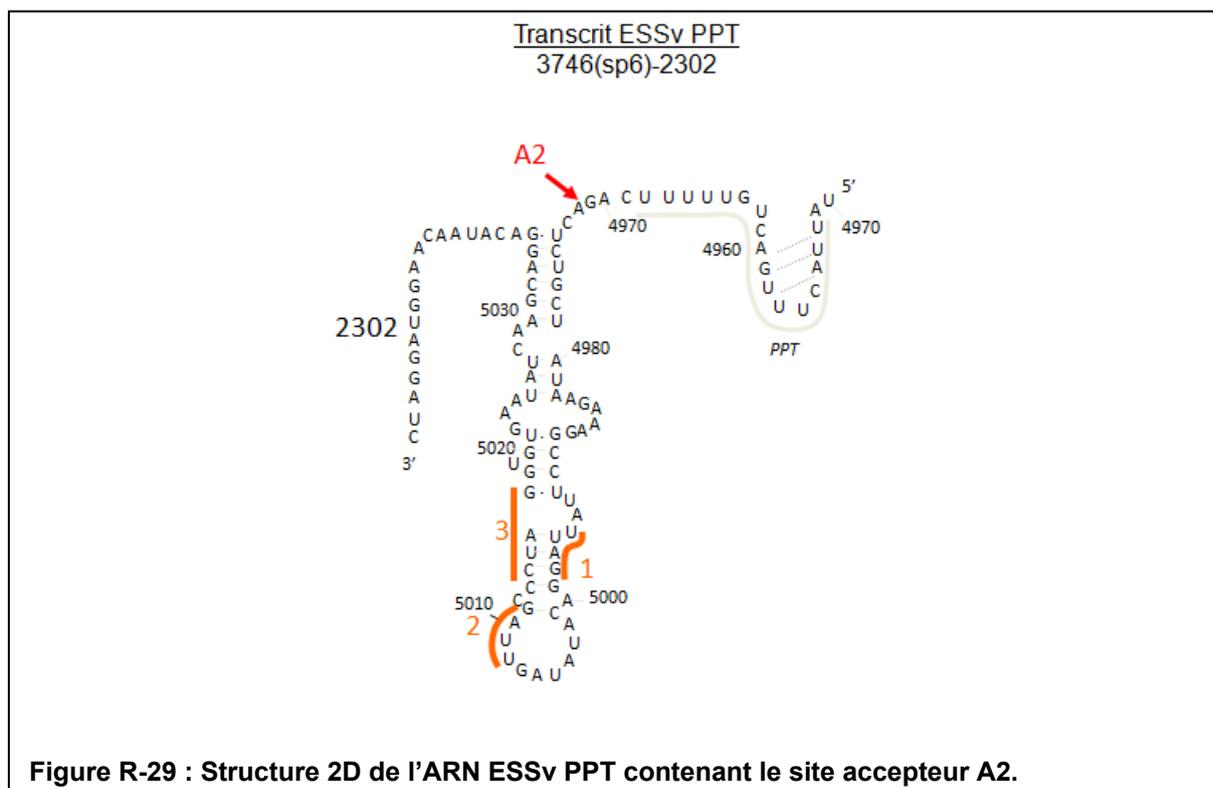
**Figure R-28: Etude de l'influence des protéines hnRNP A1 et ASF/SF2 sur l'efficacité d'épissage du transcrit C1 de VIH-1.**

a) Fractionnement par électrophorèse en gel de poly-acrylamide à 5%-Urée 8M des produits d'épissage des transcrits uniformément marqués. Le contrôle correspond au transcrit non incubé.

b) Le rapport entre la quantité de radioactivité dans la bande correspondant à l'ARN mature et la quantité de radioactivité dans la bande correspondant au précurseur pré-messager, M/P est indiqué dans chaque graphique.

## II.2 Etude de l'effet de la protéine recombinante hnRNP K sur l'utilisation *in vitro* du site accepteur A2 de VIH-1.

La structure secondaire de la région 4790 à 5035 de VIH-1/Bru renfermant le site A2 a été établie au laboratoire (Saliou *et al.*, 2009) (Figure R-29). Ce site se trouve à la base d'une structure tige-boucle très conservée dans différentes souches de VIH-1.



**Figure R-29 : Structure 2D de l'ARN ESSv PPT contenant le site accepteur A2.**

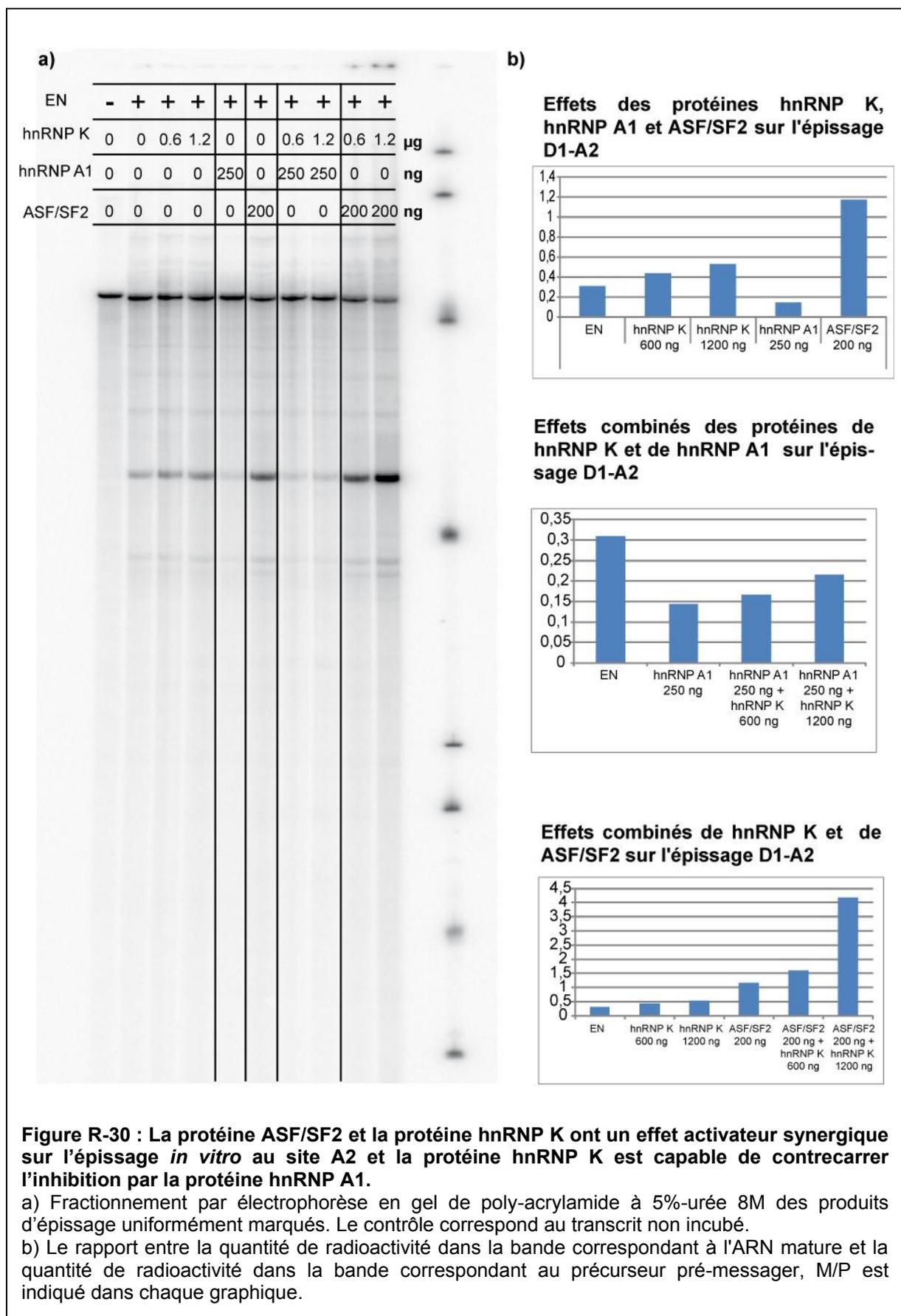
L'épissage au site accepteur A2 est inhibé par les protéines hnRNP A1/B qui se fixent sur les éléments de l'ESSv (Del Gatto-Konczak *et al.*, 1999; Caputi *et al.*, 1999; Bilodeau *et al.*, 2001) et activé par la protéine ASF/SF2 à la fois *in vitro* et *ex vivo* (Ropers *et al.*, 2004).

J'ai utilisé la construction C2 représentée dans la figure R-27 pour transcrire l'ARN pré-messager D1-A2 et réaliser les expériences d'épissage *in vitro* en présence de la protéine hnRNP K. Ces expériences nous ont permis de montrer des augmentations importantes de l'efficacité d'épissage de 40% et 150% environ après ajout de 600 ng et 1,2 µg de hnRNP K respectivement (Figure R-30).

Donc, contrairement au site A7 qui est plutôt négativement régulé par la protéine hnRNP K, le site accepteur A2 est régulé positivement par cette même protéine. Il faut donc désormais envisager que la protéine hnRNP K puisse avoir un rôle bivalent sur l'épissage *in vitro* de VIH-1.

### **II.2.1 La protéine hnRNP K active l'utilisation du site accepteur A2 *in vitro* et un effet synergique est observé lorsqu'elle est associée à la protéine ASF/SF2.**

J'ai ensuite testé l'effet de cette protéine en présence cette fois, soit de la protéine inhibitrice hnRNP A1, soit de la protéine activatrice ASF/SF2. Les résultats de cette étude nous ont permis de montrer que la protéine hnRNP K était capable non seulement de contrecarrer l'effet inhibiteur de hnRNP A1 mais aussi de conforter l'effet positif de la protéine ASF/SF2. Nous avons même noté un effet synergique des protéines ASF/SF2 et hnRNP K sur l'épissage *in vitro* de l'intron D1-A2 puisque l'augmentation de l'épissage atteint 1250% lorsqu'elles sont rajoutées ensemble dans la réaction d'épissage (Figure R-30). Ce résultat est d'autant plus important qu'aucune interaction ni effet associé entre hnRNP K et ASF/SF2 n'ont été décrits à ce jour.

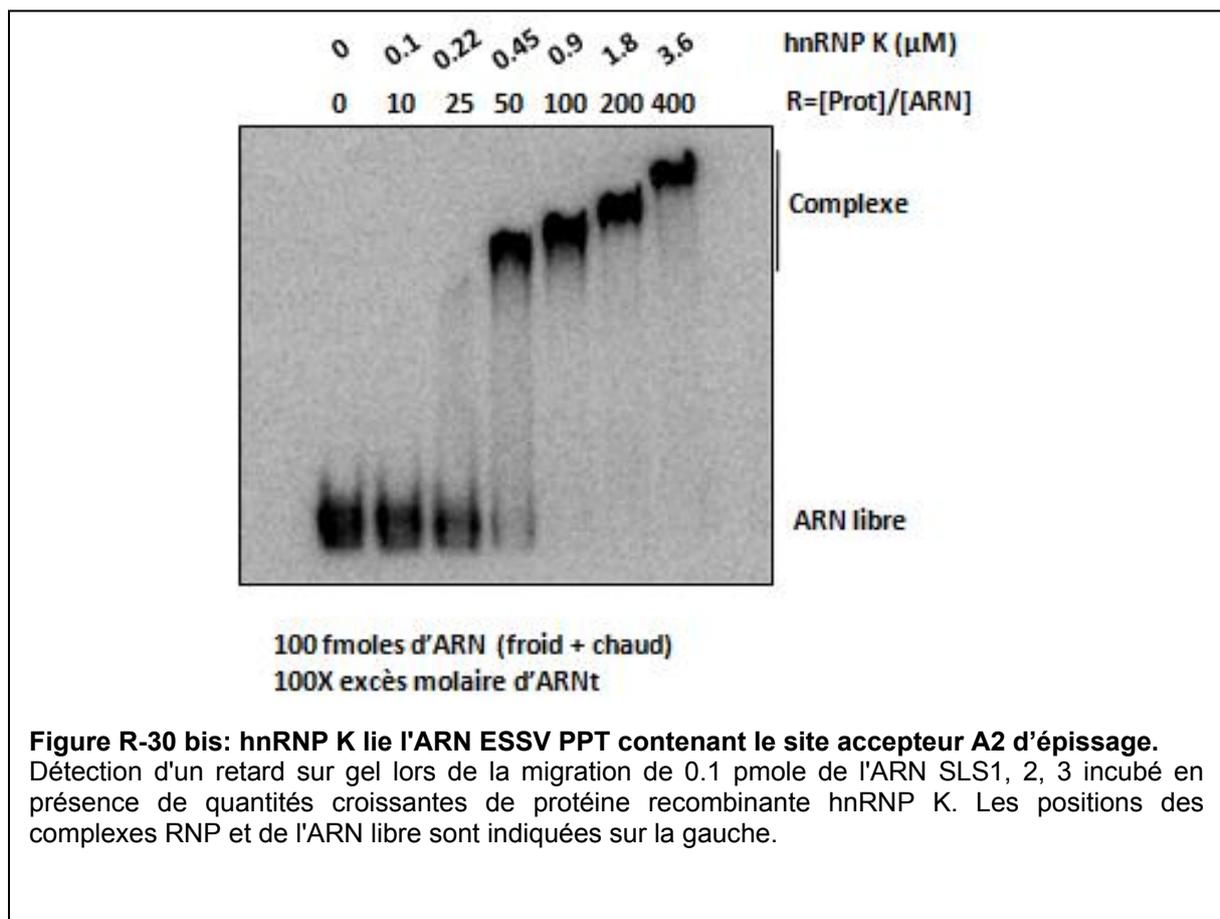


**II.2.2 Etude du mécanisme d'action de la protéine hnRNP K sur l'utilisation du site accepteur A2.**

**II.2.2.1 La protéine hnRNP K est capable de se fixer sur un fragment d'ARN du VIH-1 incluant le site A2.**

Afin de mieux comprendre le mécanisme d'action de la protéine hnRNP K, il était tout d'abord intéressant de voir si la protéine était capable de se fixer dans la région de l'ARN entourant le site accepteur A2 (Figure R-31).

Pour cela nous avons réalisé une expérience de retard sur gel en utilisant des quantités croissantes de protéine recombinante hnRNP K et un ARN appelé ESSv-PPT renfermant le site A2. Cet ARN correspond globalement à la séquence PPT (Polyprymidine Tract) suivi de l'exon organisé en tige-boucle et défini par les sites A2-D3. Nous avons ainsi montré que la protéine hnRNP K a une bonne affinité de fixation pour cet ARN (Figure R-30 bis).

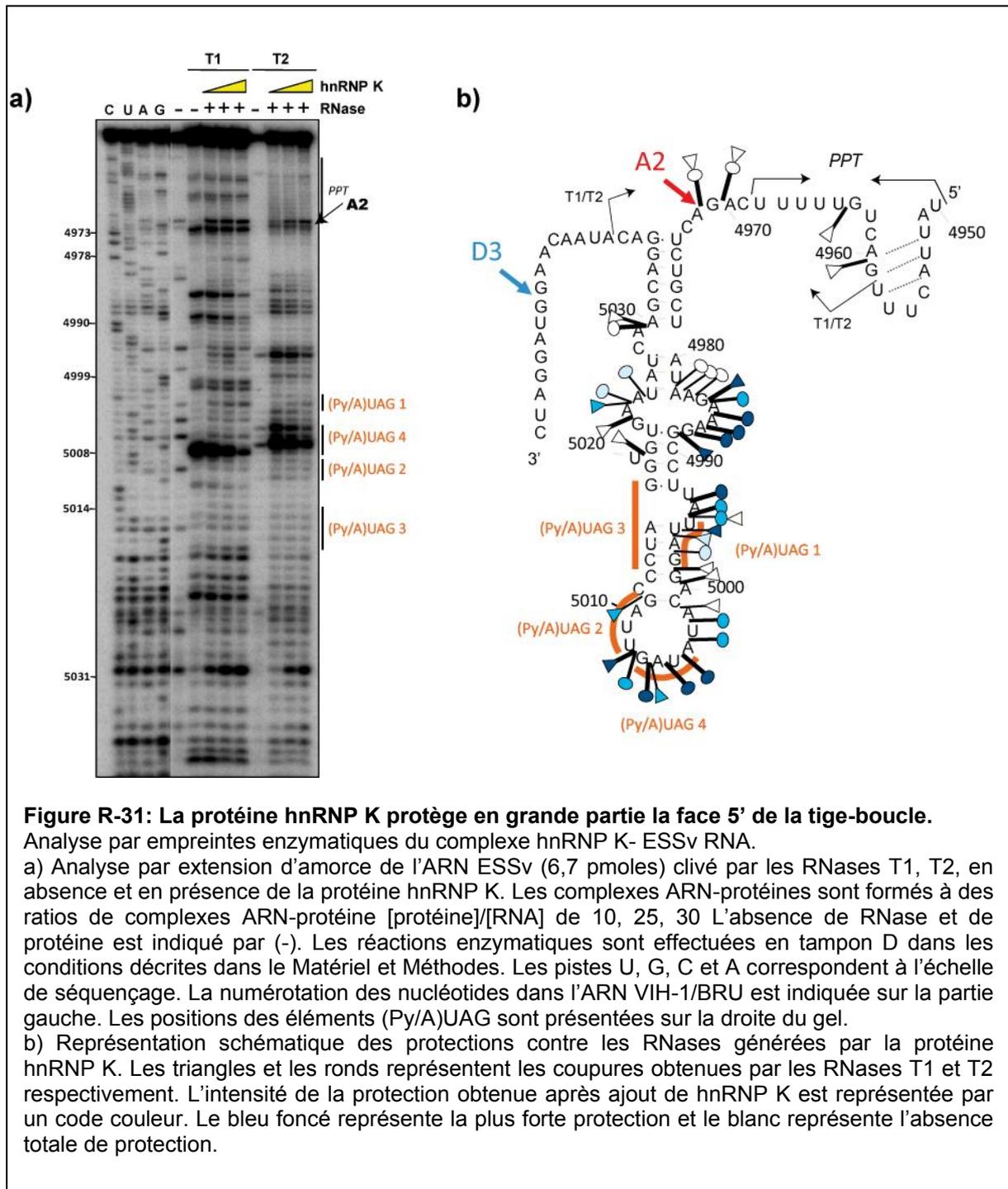


### **II.2.2.2 Les sites de fixation de la protéine hnRNP K sur l'ARN ESSv PPT chevauchent ceux des protéines hnRNP A1 et ASF/SF2.**

Afin de mieux préciser les sites de fixation de la protéine hnRNP K sur l'ARN, L. Ayadi a réalisé des expériences d'empreintes enzymatiques, en utilisant les ribonucléases T1 et T2 et des quantités croissantes de protéine hnRNP K (voir Matériel et Méthodes). Les résultats de cette expérience sont présentés dans la figure R-31.

Un changement de structure est observé en présence de la protéine hnRNP K au niveau des nucléotides 4978 et 5031 qui sont normalement appariés, mais semblent plus sensibles à la coupure enzymatique en présence de hnRNP K, ce qui indique une légère ouverture de la tige-boucle à ce niveau (Figure R-31).

Les nucléotides de 4982 à 4988, déjà décrits comme fixant faiblement les protéines hnRNP A1 et ASF/SF2 (résultats non publiés, thèse D. Ropers), sont fortement protégés de la coupure enzymatique par la fixation de la protéine hnRNP K. La protéine hnRNP K couvre aussi les nucléotides 4993 et 5007 situés en partie dans la boucle terminale de la tige-boucle et renfermant deux des quatre séquences (Py/A)UAG qui constituent l'élément ESSv impliqué dans la fixation de l'activité inhibitrice de la protéine hnRNP A1. Il a été montré au laboratoire que cette région fixait également la protéine ASF/SF2.

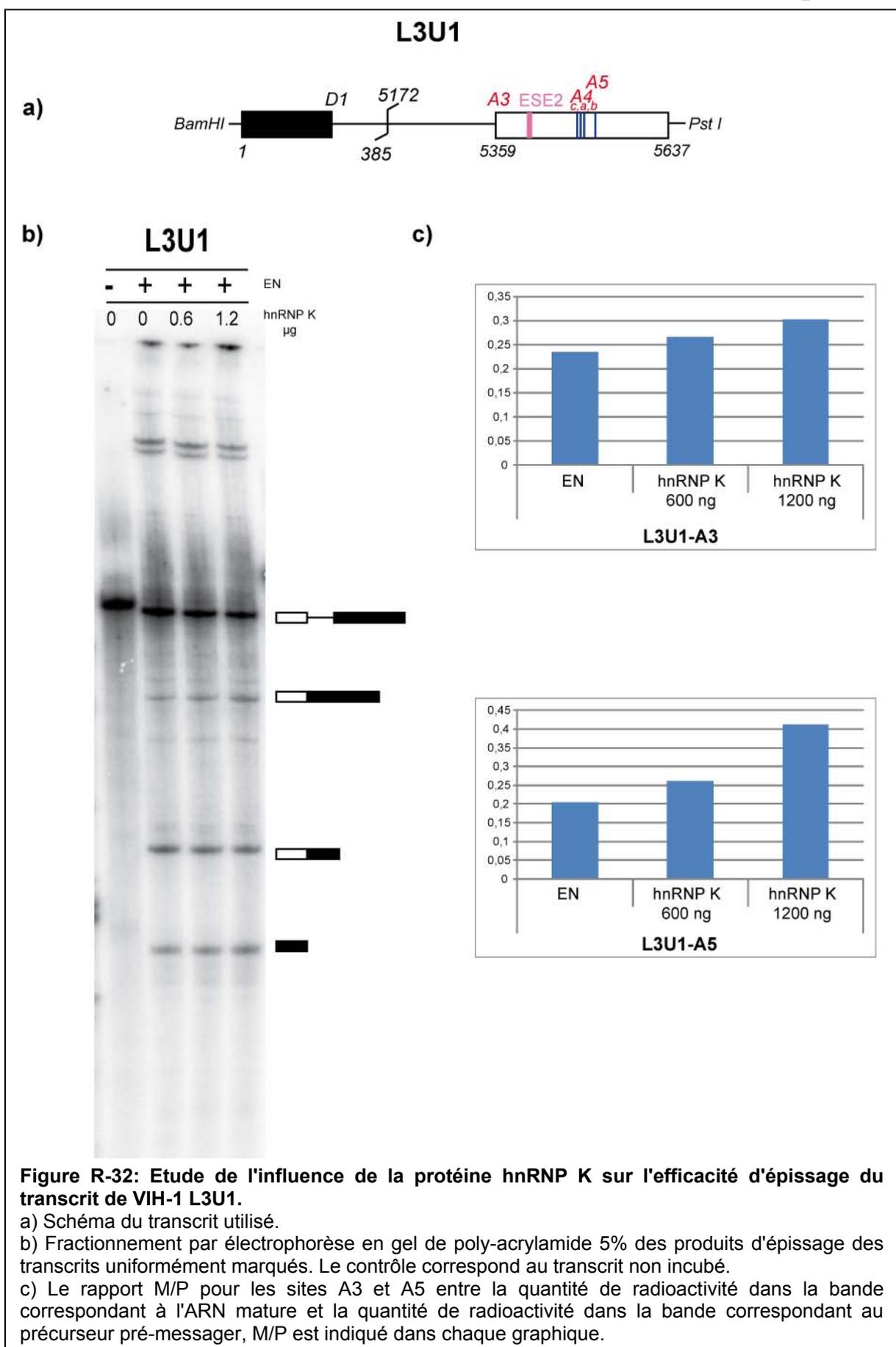


Les protéines hnRNP K et ASF/SF2 rentrent probablement en compétition avec la protéine hnRNP A1 pour la fixation sur cette tige-boucle lors de la régulation de l'épissage au site A2 puisque ces trois protéines utilisent des sites de fixations proches ou chevauchants (résultats non publiés). Ces résultats suggèrent que l'épissage au niveau de ce site est certainement dépendant du taux relatif de chacun des acteurs moléculaires dans la cellule.

### II.3 Etude de l'effet de la protéine recombinante hnRNP K *in vitro* sur l'utilisation des sites accepteurs A3 à A5 de VIH-1.

Le site accepteur A3 est situé dans une région de l'ARN viral riche en sites accepteurs puisque 5 sites accepteurs en compétition les uns avec les autres, A3, A4a, A4b, A4c et A5, sont situés dans un segment de 200 nucléotides. Ceci s'explique par le fait que cette portion centrale de l'ARN viral code les protéines Vpr, Tat et Rev. Le site A3 peut être utilisé en combinaison avec les sites donneurs D1, D2, D3. La jonction du site A3 à ces sites donneurs permet de produire les ARNm *tat*. En comparaison avec les autres sites accepteurs en compétition avec lui, la séquence polypyrimidine du site A3 apparaît plus efficace pour la fixation du facteur U2AF que celles des autres sites. Pourtant le site A3 est faiblement utilisé *in vivo*, aussi bien dans les cellules lymphoïdes (Guatelli *et al.*, 1990; Robert-Guroff *et al.*, 1990), que dans les cellules HeLa (Purcell *et al.*, 1993). Ceci s'explique par la présence de deux éléments inhibiteurs agissant en *cis*, l'élément ESS2p (Jacquetet *et al.*, 2001) et l'élément ESS2 (Amendt *et al.*, 1994; Si *et al.*, 1997). L'utilisation de ce site est permise grâce à son activation par les protéines SR, SC35 et SRp40 (Ropers *et al.*, 2004; Jacquetet *et al.*, 2005).

Pour étudier l'effet sur l'épissage de hnRNP K sur le site A3, nous avons utilisé la construction C3 qui permet de produire un pré-ARNm contenant un mini-intron D1-A3, sans les sites A4a, A4b, A4c et A5 et sans l'élément inhibiteur ESS2, ainsi que la construction L3-U1 (Figure R-32) qui permet de produire un ARN contenant le site A3, l'élément ESS2 et l'ensemble des sites A4a, A4b, A4c, A5. L'ajout de 0.6 et 1.2 µg de protéine hnRNP K à la réaction d'épissage *in vitro* conduit à une augmentation d'environ 20 et 100% respectivement de l'efficacité d'épissage pour l'ARN C3 (Figure R-27).



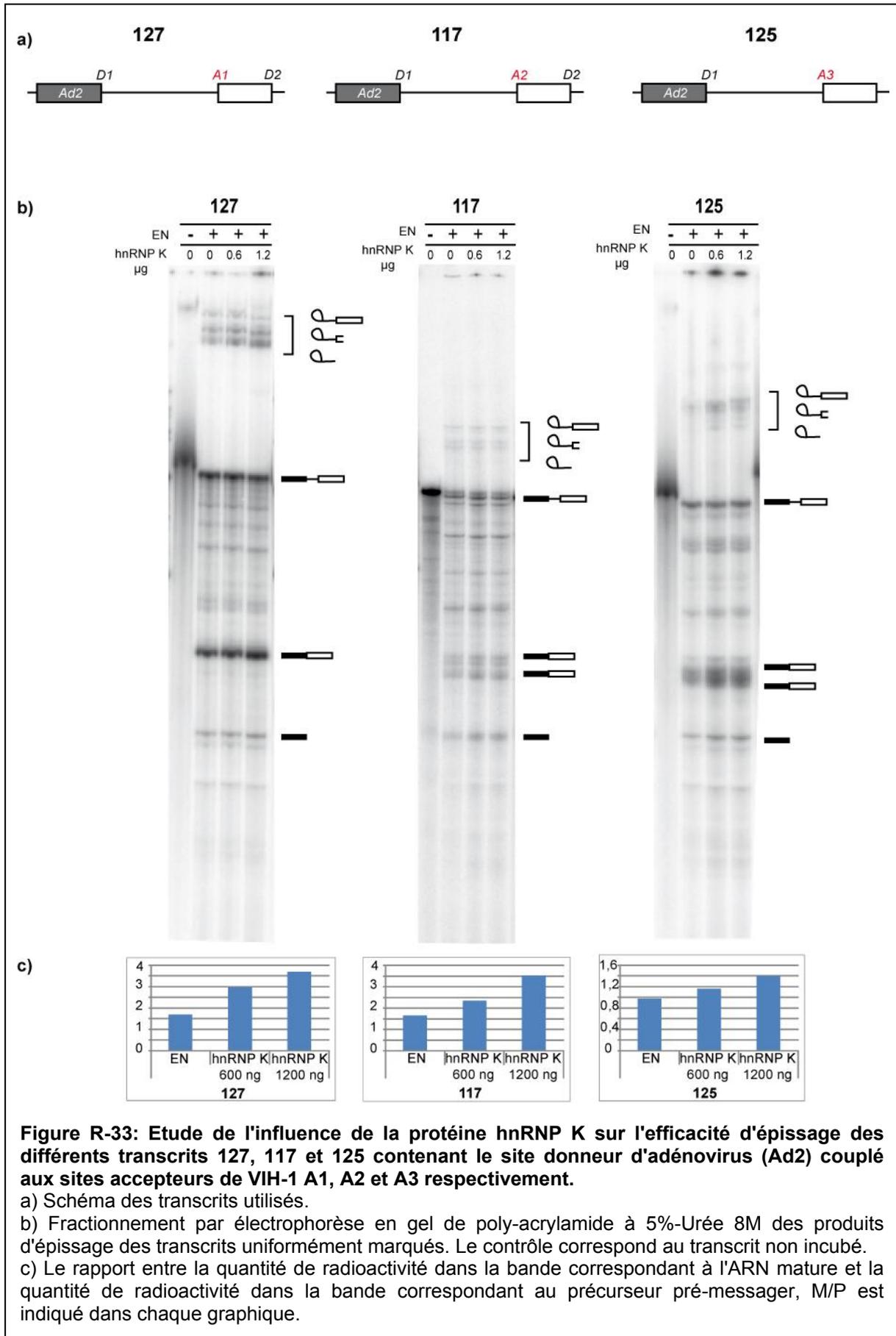
Dans le contexte de l'ARN L3-U1, contenant le site donneur D1 et les sites accepteurs A3 à A5 de VIH-1, la protéine hnRNP K est également capable d'induire une augmentation de l'efficacité d'épissage des sites accepteurs A3 et A5. L'épissage est activé de 12 et 28% au site A3 et de 28 et 101% au site A5 avec 600 ng et 1,2 µg de protéine hnRNP K respectivement (Figure R-33).

Cette étude de l'effet de la protéine hnRNP K sur l'épissage de différents sites accepteurs du VIH-1 met en évidence une fonction bivalente de cette protéine qui peut être activatrice (sites A2, A3 et A5) ou inhibitrice (site A7) de l'épissage selon le site accepteur. Cependant, ces expériences n'excluent pas un rôle de la protéine hnRNP K sur les sites donneurs D1 ou D4.

### **II.4 Les effets activateurs de la protéine hnRNP K observés sur l'épissage des ARN C1, C2 et C3 s'exercent sur les sites accepteurs.**

Les expériences précédentes réalisées avec des ARN utilisant le site D1 (pour A1, A2, A3 et A5) ou D4 (pour A7) n'excluent pas un effet possible sur les sites donneurs. Nous avons donc réalisé une série d'expériences avec des ARN couplant les sites accepteurs du virus VIH-1 à un site donneur d'adénovirus. Ainsi les constructions 127, 117 et 125 présentent un site donneur d'adénovirus couplé respectivement aux sites accepteurs A1, A2 et A3. L'étude de l'épissage de ces pré-ARNm permet de vérifier que les effets que nous avons observés lors de l'ajout de protéine recombinante en épissage *in vitro* sont bien liés à une régulation de l'épissage au niveau des sites accepteurs.

Un effet activateur important est obtenu pour les constructions 127, 117, 125 (Figure R-33).



L'activation de l'épissage de l'ARN 127 (AdD-A1) par la protéine hnRNP K est similaire à celle observée avec l'ARN C1 (D1-A1). L'activation de l'épissage de l'ARN 117 par l'ajout de 600ng de protéine hnRNP K est comparable à celle de la construction C2 (D1-A2) mais cet épissage est activé plus fortement que C2 avec 1,2 µg de hnRNP K. L'activation par la protéine hnRNP K de l'épissage de l'ARN 125 (AdD-A3) va être quant à elle plus importante que pour l'ARN (D1-A3).

Quel que soit le site donneur testé, la protéine hnRNP K va activer l'épissage *in vitro* en présence des sites accepteurs d'épissage A1, A2 ou A3, ce qui suggère un effet direct sur les sites accepteurs.

### III. Discussion et perspectives

#### III.1 Discussion

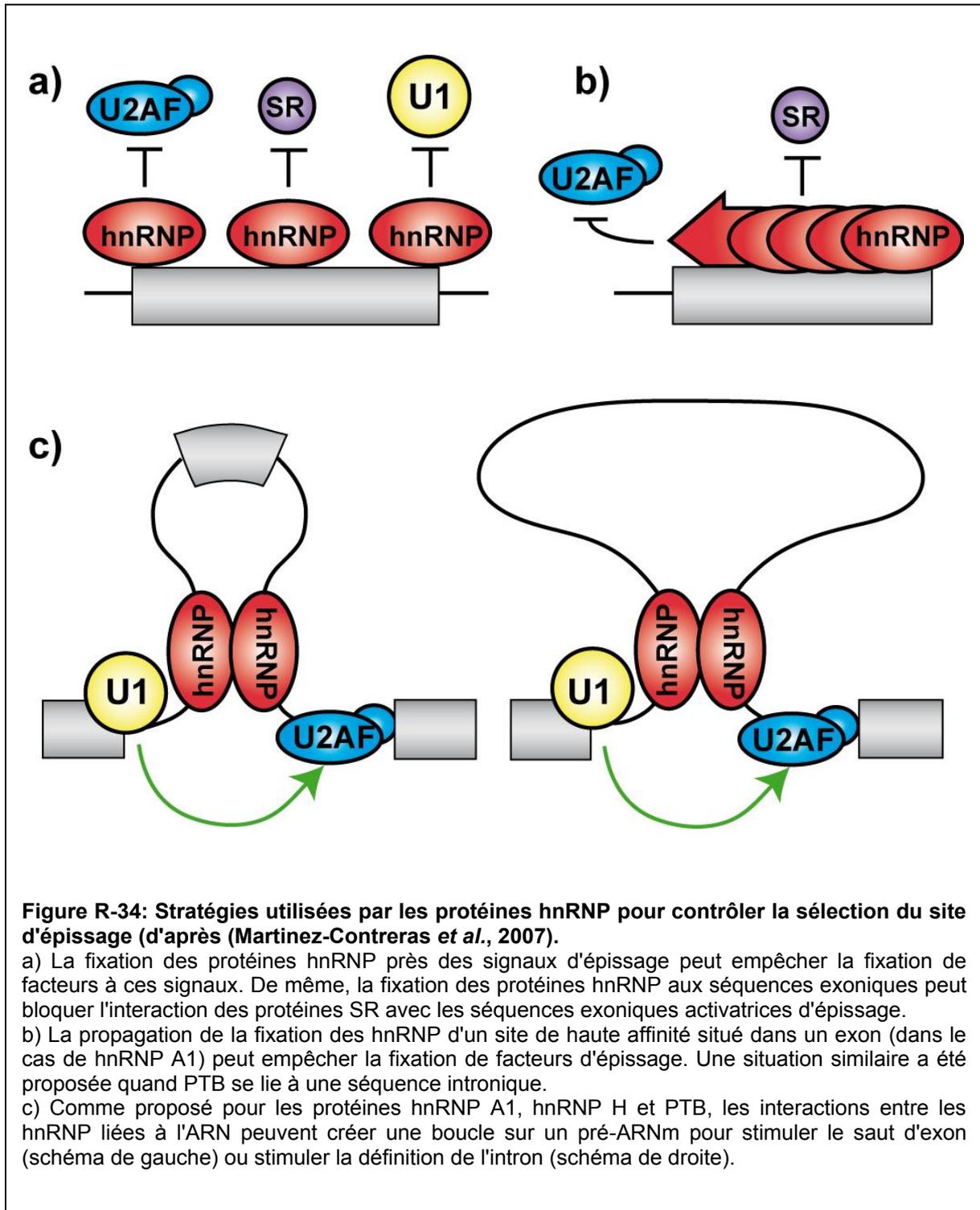
Les expériences ont été répétées avec différentes préparations d'extrait nucléaire de cellules HeLa. Le taux basal de l'épissage au site A7 est faible, par conséquent j'ai dû sélectionner un lot d'extrait nucléaire permettant un taux d'épissage satisfaisant pour cette expérience.

Les protéines hnRNP peuvent réprimer l'épissage en gênant la reconnaissance des sites d'épissage, ou en interférant avec la fixation des protéines liées aux activateurs. Récemment, il a été démontré que les hnRNP peuvent empêcher la communication entre facteurs liés à différents sites d'épissage (Martinez-Contreras *et al.*, 2007) (Figure R-34 a et b). Le plus souvent, la fixation de hnRNP bloque l'épissage en obstruant l'interaction de composants spliceosomaux à un site adjacent ou chevauchant. Par exemple, hnRNP A1, PTB et hnRNP H interfèrent avec la fixation des protéines SR, U2AF et U1snRNP, respectivement, au pré-ARNm. Une version modifiée du mécanisme d'occlusion implique la propagation par multimérisation des protéines hnRNP le long du pré-ARNm. Ceci crée un blocage quand un ou plusieurs sites de haute affinité sont flanqués par des sites adjacents de faible affinité (Figure R-34 b). La multimérisation suite à des interactions coopératives de la protéine hnRNP A1 va créer une zone de répression qui va empêcher l'interaction de composants spliceosomaux aux sites d'épissage situés pourtant loin des sites de fixation de haute affinité de la protéine hnRNP A1. La distance exacte de propagation à partir du site de haute affinité reste mal définie (Okunola *et al.*, 2009). Cependant, il est à noter que la plupart des sites de fixation fonctionnellement pertinents existent au moins en paires. La fixation coopérative d'une protéine hnRNP à des sites multiples augmenterait ainsi l'affinité et la spécificité de ces protéines qui ont des fixations moins fortes.

Inversement, il a été décrit dans la littérature plusieurs rôles positifs pour des protéines hnRNP dans l'épissage des pré-ARNm (Figure R-34 c).

De plus, des interactions coopératives entre les protéines hnRNP liées pourrait encourager l'épissage entre des paires spécifiques de sites d'épissage tout en empêchant d'autres combinaisons de sites d'épissage. Ainsi, les protéines hnRNP utilisent une grande variété de stratégies pour contrôler la sélection des sites d'épissage de façon importante.

Le rôle soit activateur, soit inhibiteur sur l'épissage de hnRNP K est décrit dans d'autres contextes. hnRNP K agit comme un répresseur de l'épissage dans le pré-ARNm de la déhydrogénase glucose-6-phosphate (Griffith *et al.*, 2006). hnRNP K peut aussi fixer le pré-ARNm de Nova1 et peut diminuer l'inclusion de l'exon 4 alternatif (Ule *et al.*, 2006). Récemment, il a été montré que hnRNP K joue un rôle prééminent dans le contrôle de l'épissage alternatif car presque la moitié des 56 événements d'épissage alternatif dans des gènes apoptotiques sont affectés par la déplétion de hnRNP K, activant ou inhibant l'inclusion de l'exon (Venables *et al.*, 2008). Ce rôle large dans la régulation de l'épissage peut être expliqué par sa capacité à interagir avec plusieurs protéines ayant un rôle dans l'épissage, dont d'autres protéines hnRNP (Bomszyk *et al.*, 2004; Mikula *et al.*, 2006). Dans l'article récent de l'équipe de B. Chabot sur l'épissage du pré-mRNA Bcl-x (Revil *et al.*, 2009) hnRNP K inhibe la production de l'isoforme d'épissage Bcl-x<sub>S</sub>. La fixation de hnRNP K sur l'élément B1 peut rentrer en compétition avec une protéine activatrice inconnue, qui n'est pas ASF/SF2, inhibant ainsi l'épissage au site Bcl-x<sub>S</sub>.



**Figure R-34: Stratégies utilisées par les protéines hnRNP pour contrôler la sélection du site d'épissage (d'après (Martinez-Contreras *et al.*, 2007)).**

a) La fixation des protéines hnRNP près des signaux d'épissage peut empêcher la fixation de facteurs à ces signaux. De même, la fixation des protéines hnRNP aux séquences exoniques peut bloquer l'interaction des protéines SR avec les séquences exoniques activatrices d'épissage.

b) La propagation de la fixation des hnRNP d'un site de haute affinité situé dans un exon (dans le cas de hnRNP A1) peut empêcher la fixation de facteurs d'épissage. Une situation similaire a été proposée quand PTB se lie à une séquence intronique.

c) Comme proposé pour les protéines hnRNP A1, hnRNP H et PTB, les interactions entre les hnRNP liées à l'ARN peuvent créer une boucle sur un pré-ARNm pour stimuler le saut d'exon (schéma de gauche) ou stimuler la définition de l'intron (schéma de droite).

### III.2 Perspectives :

Il serait intéressant de plus de déterminer si l'ajout de concentrations croissantes de hnRNP K permet d'augmenter la fixation d'ASF/SF2. Des expériences de pontage aux UV suivies d'immunosélection avec un ARN contenant le site A2 d'épissage vont être effectuées par L. Ayadi afin de vérifier cette hypothèse.

L'équipe de J Tazi à Montpellier a identifié, par criblage d'une banque de molécules chimiques, des dérivés d'indole capables de bloquer spécifiquement l'action d'ASF/SF2. Certaines de ces molécules bloquent la multiplication du VIH dans différents types cellulaires. L'un de ces dérivés bloque complètement la multiplication virale et a un effet inhibiteur différentiel sur les sites accepteurs d'épissage de l'ARN du virus HIV-1. Une étude récente de l'équipe de J. Tazi en collaboration avec notre laboratoire a en effet montré que le dérivé indol IDC16 inhibe l'épissage de l'intron D1-A2 au sein de la construction C2 de façon dose-dépendante (Bakkour *et al.*, 2007). IDC16 interfère avec l'activité activatrice de la protéine SR ASF/SF2 sur l'épissage dépendant des ESE, ce qui supprime la production de protéines virales clés en compromettant la synthèse de pré-ARNm de pleine longueur de VIH-1 et l'assemblage des particules infectieuses.

Au cours de l'épissage, la protéine hnRNP K pourrait jouer un rôle de plateforme pour les protéines régulatrices d'épissage en aidant leur recrutement aux sites d'épissages. Son effet activateur pourrait être la conséquence de son aide à la fixation sur les sites régulateurs d'épissage de protéines activatrices, plus qu'un rôle directement activateur comme celui des protéines SR. Un traitement par la molécule IDC16 lors de l'épissage *in vitro* bloquant l'effet de ASF/SF2 pourrait permettre d'observer si l'on conserve l'activation par la protéine hnRNP K. Si cette activation est diminuée par l'ajout d'IDC16, cela renforcerait l'idée que la protéine hnRNP K est une protéine plateforme pour des protéines régulant l'épissage.



---

## **CHAPITRE IV Etude du mécanisme d'inhibition de l'épissage au site accepteur A7 par la protéine hnRNP K et identification de nouvelles protéines potentiellement impliquées dans cette régulation.**

### **I. L'addition de protéine recombinante hnRNP K change le profil de fixation des protéines sur l'ARN SLS2-A7 lors de pontage aux UV.**

La protéine hnRNP K est une protéine plateforme pour beaucoup d'autres protéines (Mikula *et al.*, 2006). Elle est ainsi capable de coordonner les cascades de signalisation en facilitant la communication entre kinases et facteurs qui eux-mêmes interagissent au cours de divers processus avec les acides nucléiques (Bomsztyk *et al.*, 2004). Nous avons montré précédemment que hnRNP K interagit physiquement avec l'ARN SLS2-A7 contenant le site A7 accepteur d'épissage et que cette protéine a un effet inhibiteur sur l'épissage *in vitro* de l'intron D4-A7. Lors de sa fixation sur l'ARN autour du site A7, cette protéine pourrait recruter certaines protéines sur l'ARN et/ou en chasser d'autres. Notre but est d'identifier des protéines dont la liaison à l'ARN SLS2-A7 change en présence de hnRNP K recombinante, ce qui peut permettre de mieux comprendre le rôle de celle-ci.

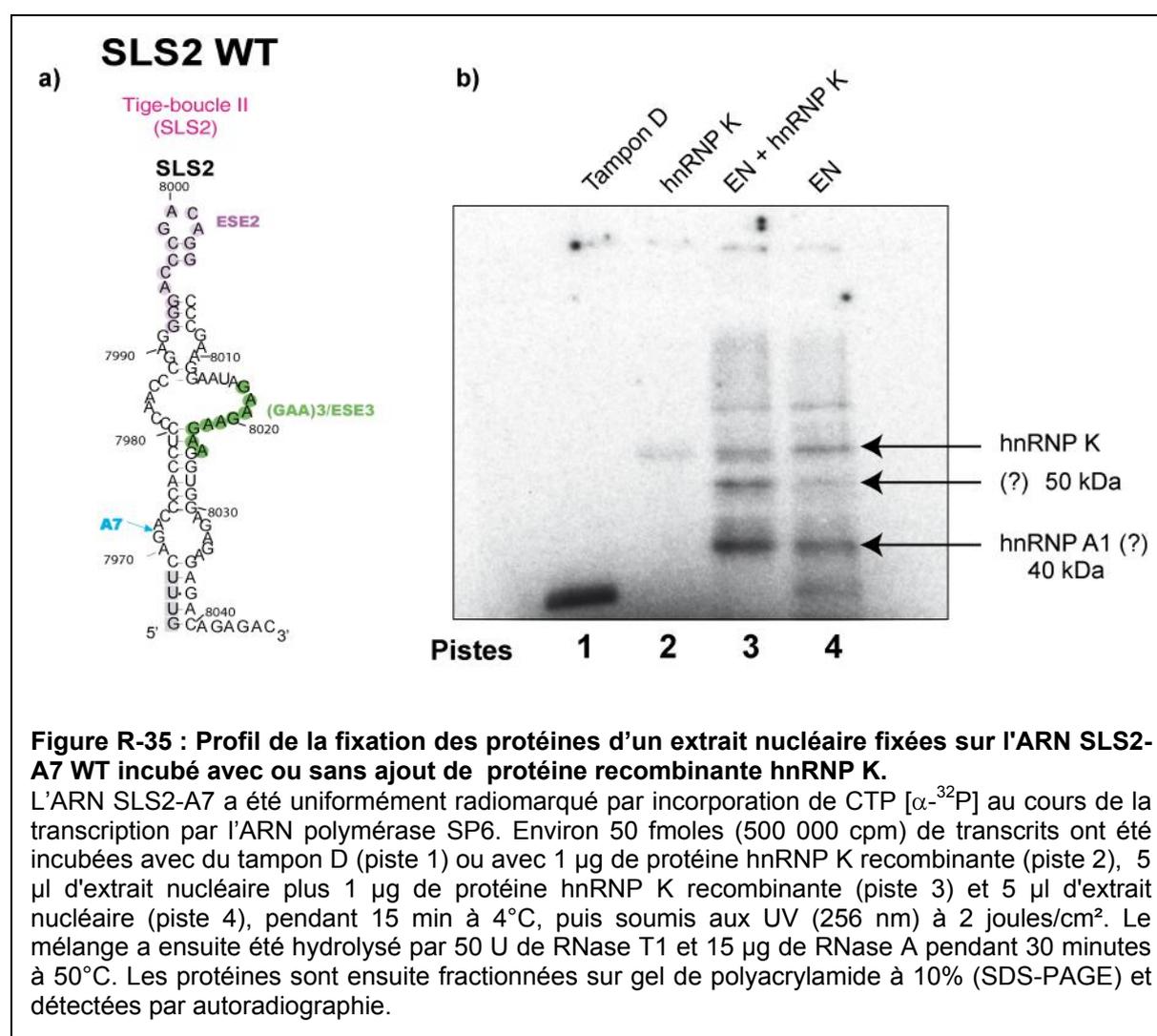
Le profil de fixation des protéines sur l'ARN lors du pontage aux UV en extrait nucléaire de cellules HeLa est observé sans et avec ajout de protéine hnRNP K recombinante. Bien que le taux de pontage soit faible avec cette technique, les résultats sont reproductibles. Lorsque l'on analyse les profils obtenus sur un gel SDS-PAGE, la fixation d'une protéine de poids moléculaire d'environ 40 kDa ainsi que d'une protéine d'environ 50 kDa semble améliorée lors de l'ajout de hnRNP K dans l'extrait nucléaire. Jusqu'à présent, les deux seules protéines connues pour agir sur la régulation de l'utilisation du site A7 sont la protéine hnRNP A1 et la protéine SR ASF/SF2 (Marchand *et al.*, 2002). La bande correspondant à une protéine de 40 kDa pourrait être la protéine hnRNP A1. Il sera intéressant d'identifier ces protéines pour comprendre le mode de régulation de hnRNP K sur l'épissage.

Il était tout d'abord intéressant d'étudier le lien potentiel entre les deux protéines hnRNP K et hnRNP A1 pour l'inhibition à ce site.

## II. La protéine hnRNP K intervient-elle dans la fonction inhibitrice de hnRNP A1 sur l'épissage *in vitro* de l'intron D4-A7 par hnRNP K ?

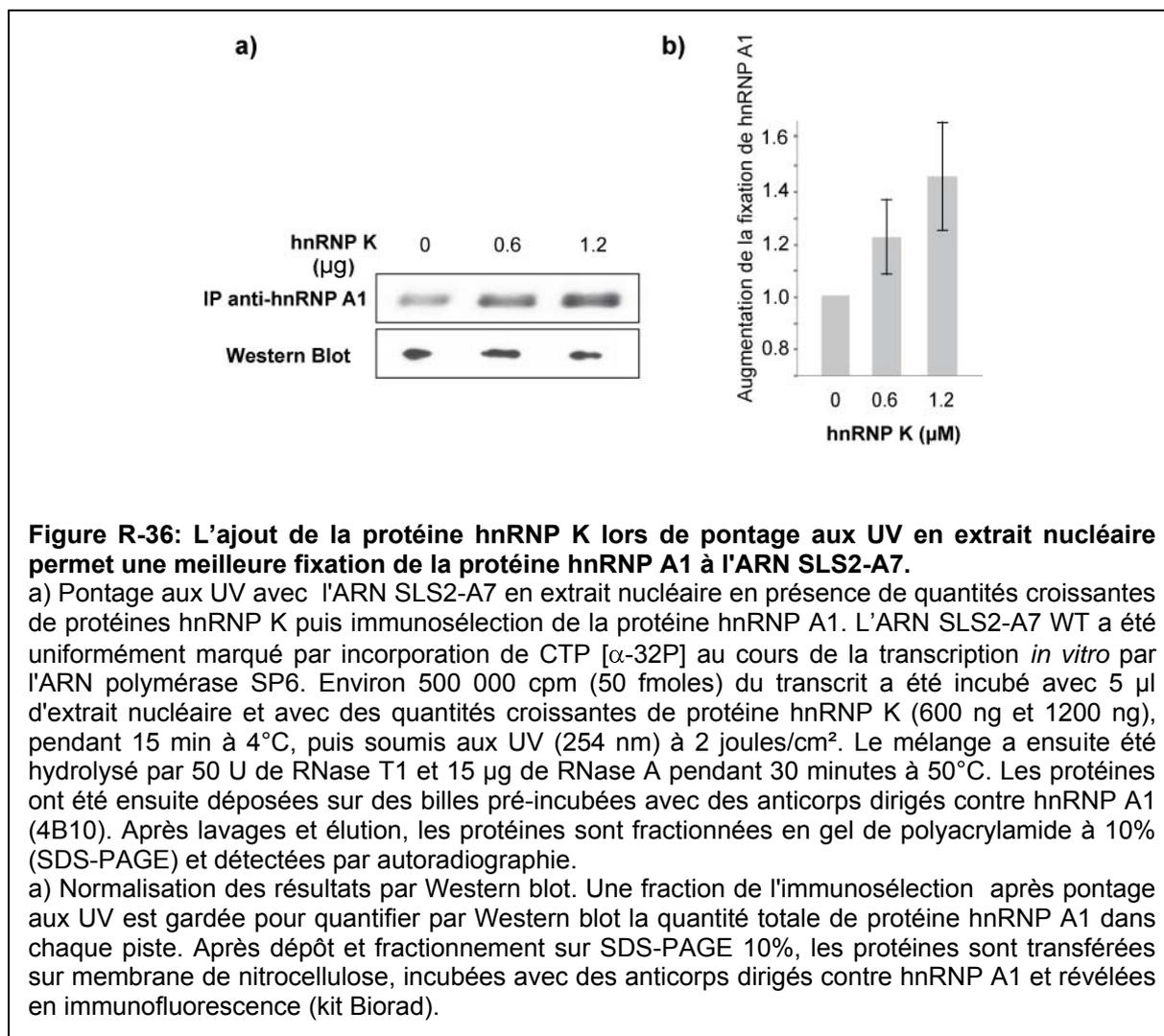
### II.1 L'addition de protéine hnRNP K à l'extrait nucléaire renforce la fixation de la protéine hnRNP A1 sur l'ARN SLS2-A7.

Le profil de fixation des protéines sur l'ARN SLS2-A7 après pontage aux UV montre une fixation différentielle des protéines en absence ou en présence de protéine hnRNP K. L'une des protéines dont l'ajout de la protéine hnRNP K semble augmenter sa fixation migre à 40 kDa, ce qui pourrait concorder avec la protéine hnRNP A1 qui a un poids moléculaire de 39 kDa (Figure R-35).



Pour vérifier cette hypothèse, une gamme croissante de protéine hnRNP K a été ajoutée à un extrait nucléaire lors d'un pontage aux UV avec l'ARN SLS2-A7 WT. Ce pontage a été suivi par une immunosélection avec un anticorps monoclonal fixé sur des billes protéine G-

sépharose et dirigé contre la protéine hnRNP A1. Les résultats montrent clairement que lorsque l'on augmente la quantité de protéine hnRNP K dans l'extrait nucléaire, le taux de protéine hnRNP A1 fixé sur l'ARN est augmenté (Figure R-36). Pour normaliser ce résultat, un western blot contre la protéine hnRNP A1 a été réalisé (Figure R-36).

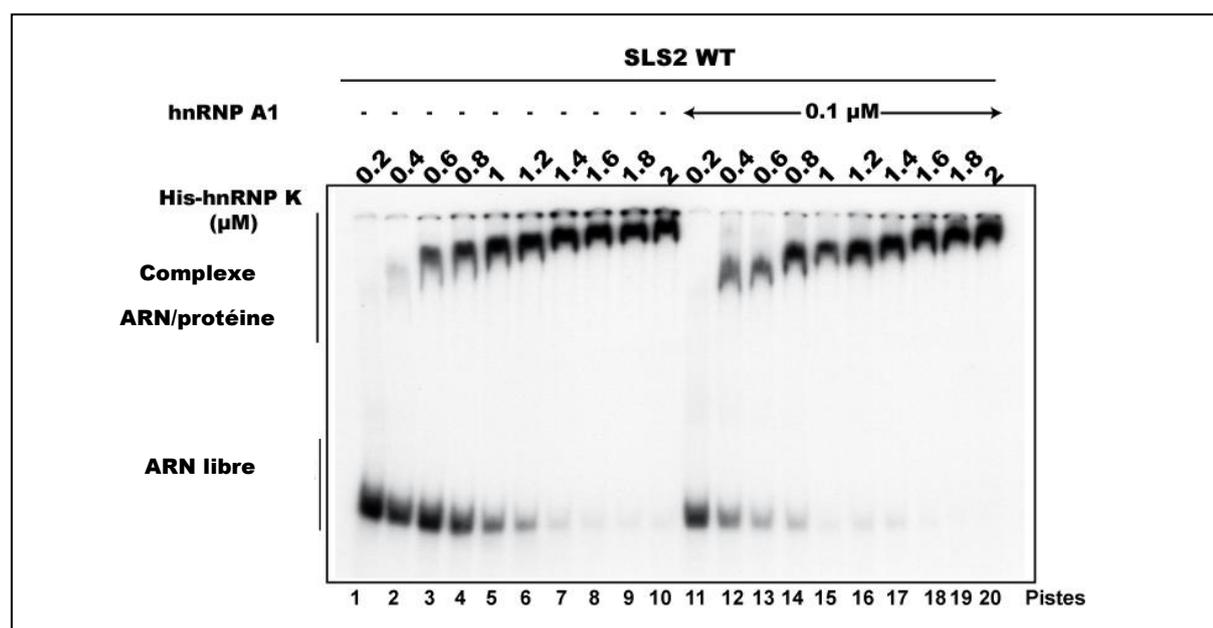


Ces résultats nous ont permis de démontrer que la bande observée à 40 kDa et dont l'intensité est augmentée lors de l'ajout de la protéine hnRNP K recombinante est bien celle de la protéine hnRNP A1.

**II.2 L'effet inhibiteur de la protéine hnRNP K sur l'efficacité d'épissage au site A7 peut être le résultat d'une fixation coopérative des deux protéines, hnRNP K et hnRNP A1, sur l'ARN.**

D'après les résultats des expériences de pontage aux UV précédemment décrites, le mécanisme d'inhibition de l'épissage de la protéine hnRNP K semble avoir un lien avec celui de la protéine hnRNP A1. En effet, la protéine hnRNP K permet d'améliorer la fixation de hnRNP A1 sur l'ARN en présence d'extrait nucléaire. De plus, j'ai pu montrer que ces deux protéines partagent trois sites de fixation sur des séquences régulatrices de l'épissage au site A7.

Une approche par retard sur gel nous a permis d'observer plus précisément les propriétés de fixation de la protéine hnRNP A1 sur l'ARN SLS2-A7 WT en présence de hnRNP K (Figure R-37). Une faible concentration de hnRNP A1 de 0,1  $\mu$ M a été utilisée. Cette concentration n'est pas suffisante pour observer un retard de migration de l'ARN radiomarqué sur gel après électrophorèse (Piste 11). L'ajout de hnRNP A1 montre une augmentation significative de la fixation de la protéine hnRNP K sur l'ARN SLS2-A7 malgré la faible quantité de protéine hnRNP A1 ajoutée.



**Figure R-37: Expérience de retards sur gel de l'ARN SLS2-A7 avec les protéines recombinantes hnRNP A1 et hnRNP K.**

0,1 pmole d'ARN SLS2-A7 est incubé avec 20 fmoles d'ARN SLS2-A7 radioactif avec les protéines recombinantes. Les pistes 1 à 10 et 11 à 20 contiennent des quantités croissantes de protéines hnRNP K. Les pistes 11 à 20 contiennent en plus une quantité fixe de 0,1  $\mu$ M de hnRNP A1. Les complexes sont déposés sur gel natif d'électrophorèse 5% acrylamide (59:1).

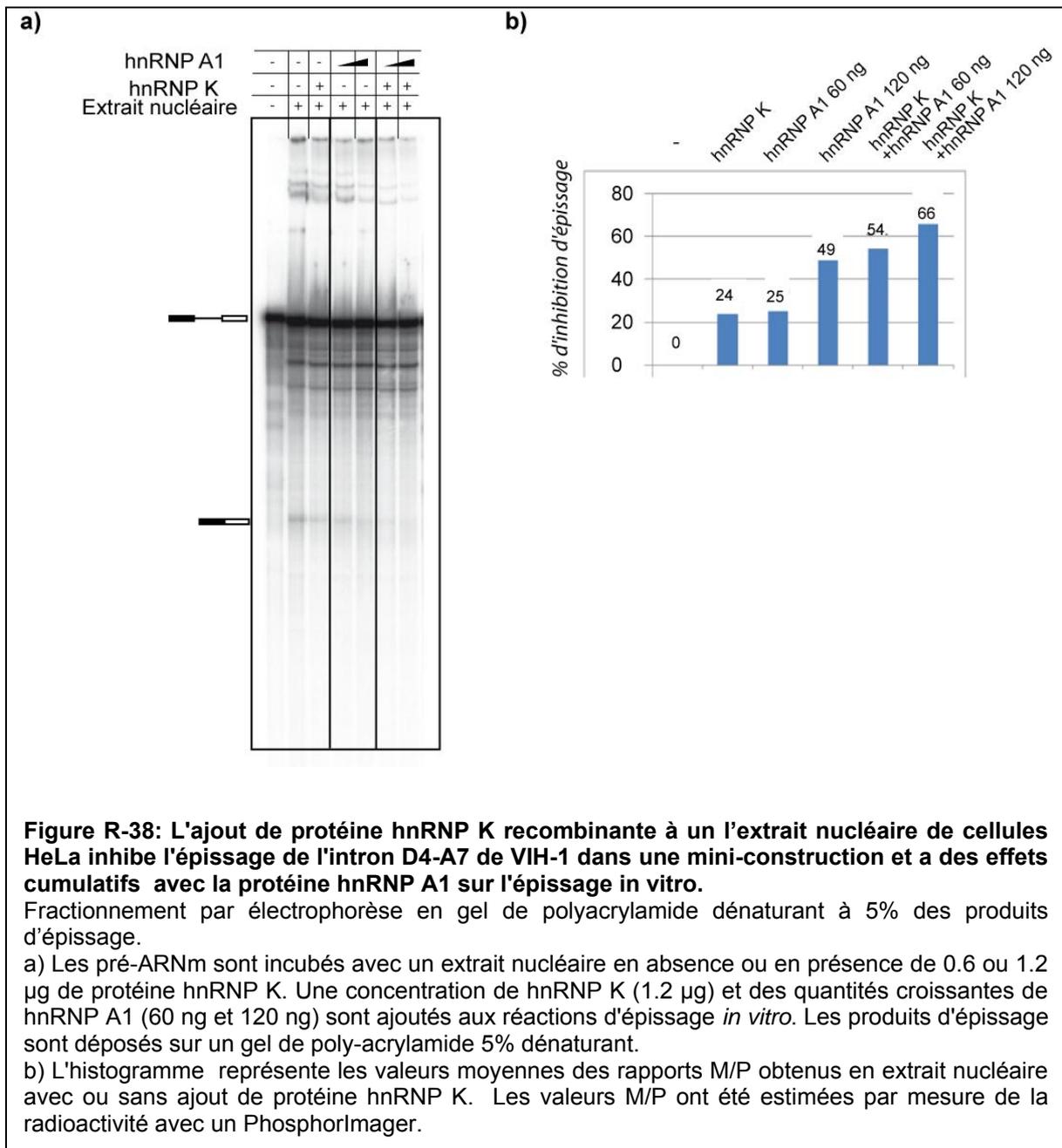
### **II.3 Les effets de hnRNP A1 et hnRNP K sont cumulatifs lors de l'épissage *in vitro* au site A7.**

Pour examiner les conséquences de cette fixation coopérative entre les protéines hnRNP K et A1 sur l'épissage, j'ai réalisé des expériences d'épissage en ajoutant différentes concentrations de ces deux protéines simultanément (Figure R-38). Les effets des protéines hnRNP A1 et hnRNP K se sont montrés cumulatifs mais ces résultats ne nous permettent pas de mettre en évidence un effet synergique. En effet, le taux basal d'épissage de l'intron D4-A7 empêche d'évaluer les effets réels des deux protéines ensemble sur l'épissage et rend difficile l'observation d'un potentiel effet synergique.

La fixation de la protéine hnRNP A1 à l'ISS et à l'ESS3 est un pré-requis pour l'inhibition de l'épissage au site A7. L'association de la protéine hnRNP A1 à l'ESS3 est quant à elle le point de départ de sa multimérisation vers l'ESE3 et son association à l'ISS vers le point de branchement du snRNP U2. La protéine hnRNP A1 régule ainsi l'épissage grâce à un blocage stérique (Tange *et al.*, 2001). La mutation d'un seul de ces sites régulateurs change le mode de régulation de l'épissage au site A7 car la protéine hnRNP A1 a besoin de se fixer à ces sites précis pour multimériser et bloquer l'accès à la machinerie spliceosomale.

Le mode d'inhibition de la protéine hnRNP K semble différent du mode d'action de la protéine hnRNP A1. Elle ne doit pas avoir de point de départ de multimérisation sur l'un de ses sites de fixation. Cependant, le fait que la hnRNP K partage certains sites de fixation avec la protéine hnRNP A1 amène à rechercher un processus de régulation partagé entre ces deux protéines.

Des études de spectrométrie de masse visant à déterminer la stœchiométrie de la protéine hnRNP K et de la protéine hnRNP A1 sur l'ARN SLS2-A7 ont été menées en collaboration avec l'équipe d'A. Vandorsselaer à Strasbourg, mais n'ont pas donné de résultats interprétables du fait de problèmes techniques. La détermination du nombre exact de molécules de ces protéines sur l'ARN SLS2-A7 aurait en effet pu aider à l'interprétation de leur mécanisme de régulation de l'épissage.



### III. Identification de protéines interagissant à la fois avec l'ARN SLS2-A7-MS2 et la protéine hnRNP K

#### III.1 Objectifs de l'étude

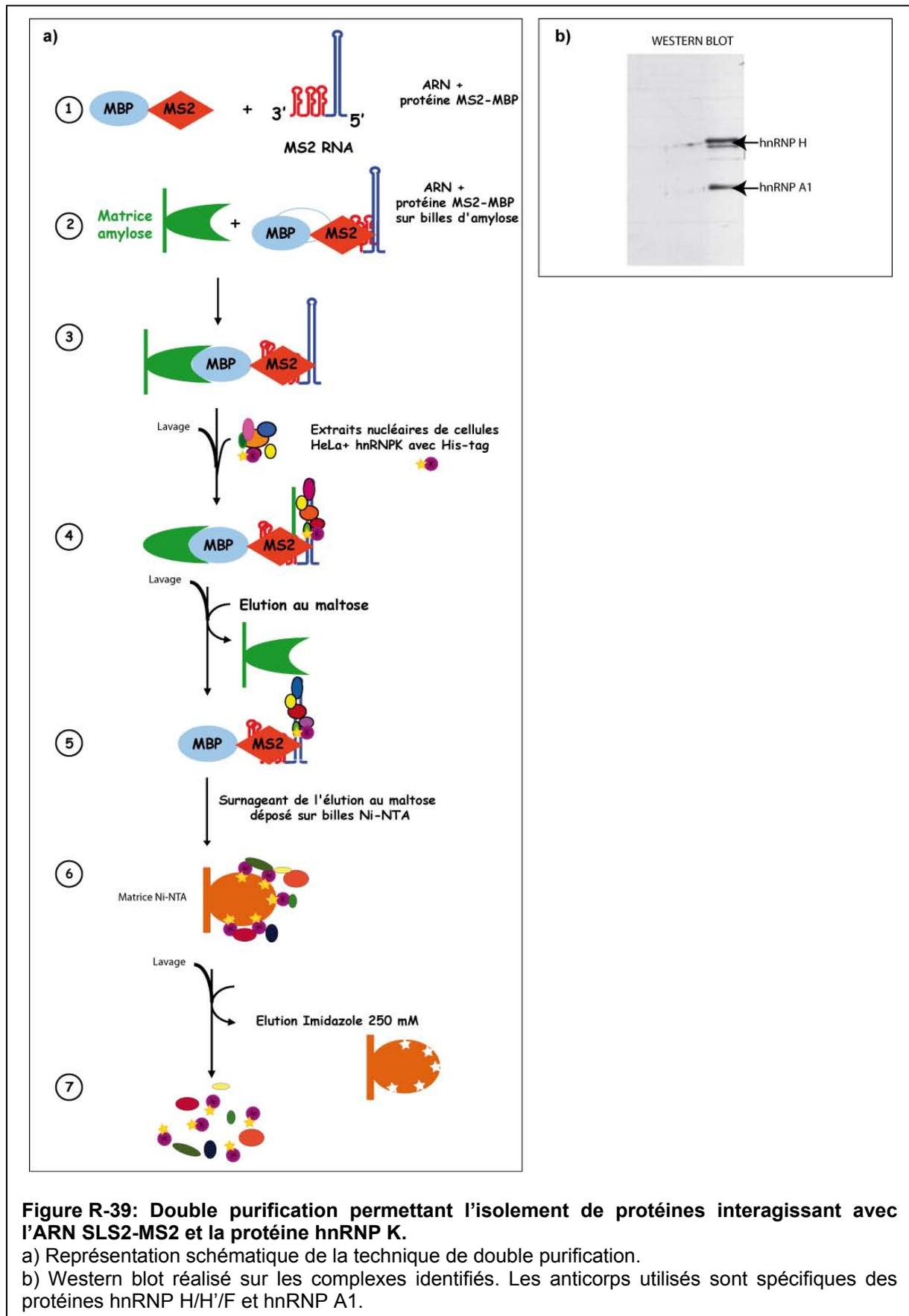
Afin de mieux comprendre le mode d'action de la protéine hnRNP K lors de l'épissage alternatif, il était intéressant de rechercher ses partenaires protéiques parmi les protéines identifiées comme ayant une interaction avec l'ARN SLS2-A7 contenant le site d'épissage A7. L'action pléiotropique de la protéine hnRNP K reflète sa capacité à interagir avec différentes classes de facteurs. Des travaux de l'équipe de Ostrowski (Mikula *et al.*, 2006),

ont permis de déterminer les partenaires protéiques de la protéine hnRNP K grâce à une purification d'affinité.

### **III.2 Principe de la double purification permettant l'isolement de protéines interagissant avec l'ARN SLS2-A7-MS2 et la protéine hnRNP K**

Afin de sélectionner les protéines de l'extrait nucléaire qui interviendraient dans la régulation de l'utilisation du site A7 en association avec la protéine hnRNP K, nous avons réalisé une purification de complexes associés à l'ARN SLS2-A7 WT, faisant intervenir deux étapes. La première étape de purification suit le protocole classique de purification (Figure R-39, étape 1 à 4). L'ARN SLS2-A7-MS2 reconnu par la protéine MS2-MBP fixée à la matrice amylose est incubé en extrait nucléaire auquel nous avons préalablement ajouté de la protéine hnRNP K recombinante avec une étiquette histidine. Après plusieurs lavages, les complexes ARN-protéines formés sont élués par ajout de maltose. Un contrôle négatif est réalisé en effectuant la même expérience en parallèle mais sans ajouter de protéine His-hnRNP K dans l'extrait nucléaire.

Une seconde étape de purification va permettre d'isoler uniquement les complexes ARN-protéines qui renferment la protéine hnRNP K. Pour cela, l'éluat issu de la première étape de purification est déposé sur des billes Ni-NTA qui permettent de retenir l'étiquette histidine de la protéine hnRNP K. Ces protéines étant dans du tampon D à base d'HEPES non dénaturant, les interactions protéine-protéine vont être conservées. Les protéines associées à l'ARN SLS2-A7 et à la protéine His-hnRNP K fixées aux billes Ni-NTA vont rester associées aux billes après lavage et pourront être éluées à l'imidazole. Les protéines ainsi sélectionnées sont déposées et séparées par électrophorèse sur gel SDS-PAGE. Le gel est ensuite coloré au bleu colloïdal et les bandes de gel contenant les protéines sont découpées. Les parties de gel correspondantes de la piste témoin dont aucune bande de protéine n'est visible après coloration au bleu colloïdal sont également découpées. L'identification de ces bandes est réalisée par spectrométrie de masse.



### **III.3 Identification en spectrométrie de masse des protéines de l'extrait nucléaire de cellules HeLa interagissant avec la protéine His-hnRNP K et l'ARN SLS2-A7-MS2:**

Travail réalisé avec Jean-Michel SALIOU

*Laboratoire de Spectrométrie de Masse Bio-Organique STRASBOURG*

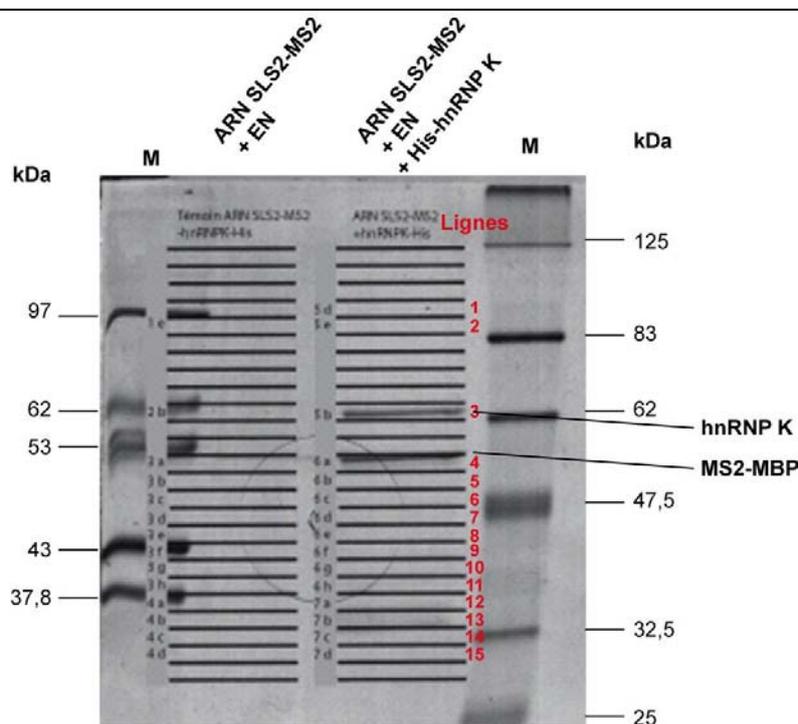
#### **III.3.1 Conditions d'analyse**

Une découpe des pistes du gel correspondant aux protéines issues des complexes isolés et visibles après coloration au bleu colloïdal a été effectuée (Figure R-40). Les protéines de chaque bande de gel ont été ensuite réduites-alkylées, puis digérées à la trypsine. Les peptides ont ensuite été extraits du gel par un solvant ACN/H<sub>2</sub>O/HCOOH (50/50/1), puis évaporé jusqu'à un volume d'environ 8 µl. L'échantillon (5 µl) est injecté sur nanoLC-Chip 150mm (Agilent) couplée au spectromètre de masse HCT Ultra (Bruker).

#### **III.3.2 Identification des protéines**

Les différents spectres de fragmentation des peptides obtenus sont ensuite comparés aux spectres de fragmentation théoriques de peptides des protéines issues de la banque SwissProt (version 57.4). Pour cette étude, la liste des protéines identifiées est établie de manière automatique en prenant comme critère de validation la présence de 2 peptides de la protéine identifiés avec un bon niveau de confiance.

L'analyse de cette liste permet d'identifier clairement les protéines majoritaires contenues dans les 2 bandes de gel fortement colorées (5b et 6a), il s'agit respectivement de la protéine hnRNP K (17 peptides identifiés) et de la protéine de fusion MS2-MBP (25 peptides identifiés). Afin de clarifier l'analyse, cette liste a été triée en retirant l'enzyme de digestion (trypsine), les contaminants (kératines et TruD issu du marqueur de taille) et les protéines provenant de la méthode de purification (Maltose-binding periplasmic protein et Coat protein), puis classée dans le tableau qui suit pour comparer les conditions sans et avec hnRNPK



**Figure R-40: Gel de purification des complexes protéiques comprenant l'ARN SLS2 et les protéines d'extrait nucléaire qui pourraient interagir à la fois avec l'ARN SLS2 et la protéine hnRNP K.**

Les complexes ont été fractionnés sur gel SDS-PAGE puis les bandes ont été découpées selon la grille d'identification des bandes présentée ici. Les marqueurs de taille moléculaire (M) sont déposés dans les pistes aux extrémités. La piste de gauche correspond au complexe témoin obtenu sans ajout de protéine hnRNP K. La piste de droite correspond au complexe formé après ajout de hnRNP K.

Le témoin (sans hnRNP K) est de bonne qualité puisque seuls 2 peptides de hnRNP A2/B1 sont retrouvés dans la bande 4b de la ligne 13 (Figure 40, Tableaux 5 et 6).

La protéine hnRNP K est bien présente dans la condition avec hnRNP K et se trouve essentiellement dans la bande 5b de la ligne 3. D'autres hnRNP semblent aussi fixées au complexe SLS2-A7/hnRNP K, il s'agit :

- de la protéine hnRNP H (bandes 6b et 6c, lignes 5 et 6),
- de la protéine hnRNP G (bande 6, ligne 8),
- de la protéine hnRNP A3 (bandes 6g, 6h et 7a, lignes 10, 11 et 12),
- de la protéine hnRNP A2/B1, présente dans le témoin mais apparemment plus importante dans la condition avec hnRNP K, au regard du nombre de peptides présents (bandes 7a et 7b, lignes 12 et 13),
- de la protéine hnRNP H3 (bandes 7a et 7b, lignes 12 et 13),

Lignes	Protein accession numbers	Protein name	Protein molecular weight (Da)	sans hnRNP K		avec hnRNP K	
				Number of unique peptides	Percentage sequence coverage	Number of unique peptides	Percentage sequence coverage
1	HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	51 012,4	pas de bande donnée		4	11,90%
2	SFPQ_HUMAN	Splicing factor, proline- and glutamine-rich	75 423,9			5	8,15%
	HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	51 012,4			3	8,62%
3	HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	51 012,4			17	43,30%
4	HNRH1_HUMAN	Heterogeneous nuclear ribonucleoprotein H	49 181,8			2	7,57%
5	HNRH1_HUMAN	Heterogeneous nuclear ribonucleoprotein H	49 181,8			10	35,60%
6	HNRH1_HUMAN	Heterogeneous nuclear ribonucleoprotein H	49 181,8			7	22,50%
7	HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	51 012,4			2	7,54%
8	HNRPG_HUMAN	Heterogeneous nuclear ribonucleoprotein G	42 187,7			2	6,70%
9							
10	ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	39 634,2			3	14,20%
11	ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	39 634,2			2	8,18%
12	HNRH3_HUMAN	Heterogeneous nuclear ribonucleoprotein H3	36 909,6			3	11,00%
	ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1	37 460,4			4	12,50%
	ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	39 634,2			2	10,00%
13	ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1	37 460,4	2	7,37%	7	28,90%
	HNRH3_HUMAN	Heterogeneous nuclear ribonucleoprotein H3	36 909,6			3	9,25%
	ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	34 194,4			5	24,10%
14	ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	34 194,4			2	10,60%
15	ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	38 828,1			3	9,41%

**Tableau 5 : Protéines majoritaires des complexes protéiques purifiés identifiées en spectrométrie de masse.**

- et de la protéine hnRNP A1, essentiellement dans la bande 7b (ligne 13), en accord avec la masse moléculaire théorique. Les différences de masse moléculaire dans le tableau sont dues aux différents isoformes de hnRNP A1.

Une autre protéine intéressante dans le cadre de l'étude est la protéine PSF (SFPQ\_HUMAN) (bande 5, ligne 2).

La protéine PSF (PTB associated Splicing Factor) lie les acides nucléiques ARN et ADN et est impliquée dans plusieurs processus nucléaires. Elle forme souvent un hétérodimère avec la protéine p54nrb pour former le complexe p54nrb/PSF (Peng *et al.*, 2002; Zhang *et al.*, 1993) et participe à la fois aux étapes précoces et tardives de l'épissage (Peng *et al.*, 2002). Elle s'associe aussi avec PTB (Patton *et al.*, 1993) ainsi qu'avec la forme libre de la protéine U1A des snRNP (Lutz *et al.*, 1998) et se lie spécifiquement au snRNA U5 *in vitro* (Peng *et al.*, 2002), mais son rôle exact dans l'épissage n'est pas connu. La protéine PSF est composée d'un domaine N-terminal riche en proline et glutamine, de deux motifs RRM en tandem et de deux domaines NLS dans la partie C-terminale. PSF se lie à l'ARN pré-messager dans le

complexe spliceosomal C et spécifiquement à la séquence polypyrimidine. Les ARN de VIH-1 *gag/pol* et *env* contiennent des éléments régulateurs agissant en *cis* (INS) qui diminuent la stabilité de ces ARN, leur transport nucléocytoplasmique et la traduction par des mécanismes inconnus. Cette régulation négative peut être neutralisée par la protéine virale Rev, résultant en un export efficace et l'expression de ces ARNm. La région INS de l'ARNm *gag* de VIH-1 est un ligand de haute affinité du complexe p54nrb/PSF. PSF affecte spécifiquement les transcrits Rev-dépendant codant Gag-Pol et Env et contenant INS.

PSF inhibe fortement l'expression de l'ARNm contenant l'INS quand elle est surexprimée dans des cellules 293. De plus, PSF est capable de neutraliser la régulation de Rev et agit pour cela de façon dominante sur Rev-CRM1. La protéine PSF exogène peut agir indépendamment de la régulation de Rev plutôt qu'en affectant la voie Rev-CRM1 elle-même. PSF exogène agit *via* l'INS pour inhiber l'expression des ARNm contenant l'INS dans les cellules humaines.

Comme indiqué dans les tableaux 5 et 6, aucune protéine SR n'a pu être identifiée avec les critères de validation précédemment cités. La recherche a donc été faite avec des critères beaucoup moins stringents (1 peptide avec un faible niveau de confiance), mais à nouveau aucun peptide de protéine SR n'a pu être identifié.

### III.3.3 Conclusion

L'analyse a permis de révéler la présence de nombreuses protéines hnRNP et de la protéine PSF liée au complexe SLS2-A7/hnRNP K. Il conviendra néanmoins de répéter l'expérience afin de valider ces identifications.

S'agissant de l'absence des protéines SR (en particulier d'ASF connue pour fixer cette région de l'ARN), il peut être proposé d'analyser une nouvelle purification en utilisant une enzyme de digestion plus appropriée, générant plus de peptides compatibles avec l'analyse de masse. Une autre expérience pourrait aussi être réalisée en ajoutant ASF fusionnée à l'étiquette histidine.

Une liste de protéines identifiées avec moins de confiance pourra aussi être établie afin de creuser plus l'étude des partenaires minoritaires purifiés.

## RESULTATS : Chapitre IV

MS/MS sample name	MS/MS sample name	Protein accession numbers	Protein name	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage
1E	E035712 (F048181)	K1C10_HUMAN	Keratin, type I cytoskeletal 10	59 494,3	3	7,25%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K22E_HUMAN	Keratin, type II cytoskeletal 2 epidermal	65 415,9	2	5,63%
		MALE_ECOLI	Maltose-binding periplasmic protein	43 371,9	2	8,84%
2B	E035713 (F048184)	K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	8	15,10%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K1C9_HUMAN	Keratin, type I cytoskeletal	62 047,8	3	8,67%
3A	E035714	K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	7	13,30%
		TRYP_PIG	Trypsin	24 391,3	3	16,50%
3B	E035715 (F048188)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		MALE_ECOLI	Maltose-binding periplasmic protein	43 371,9	2	6,82%
3C	E035716 (F048191)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	2	3,77%
3D	E035717 (F048193)	K1C10_HUMAN	Keratin, type I cytoskeletal 10	59 494,3	17	37,30%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K1C9_HUMAN	Keratin, type I cytoskeletal	62 047,8	5	17,30%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	12	21,40%
3E	E035718 (F048195)	K1C10_HUMAN	Keratin, type I cytoskeletal 10	59 494,3	2	5,23%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	3	6,28%
3F	E035719 (F048198)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		TRUD_ECOLI	tRNA pseudouridine synthase D	39 073,7	7	30,90%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	2	3,77%
3G	E035720 (F048200)	K1C10_HUMAN	Keratin, type I cytoskeletal 10	59 494,3	4	9,44%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	4	7,85%
3H	E035721	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K1C10_HUMAN	Keratin, type I cytoskeletal 10	59 494,3	6	11,30%
4A	E035722 (F048202)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K22E_HUMAN	Keratin, type II cytoskeletal 2 epidermal	65 415,9	2	6,42%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	4	7,06%
		ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1	37 460,4	2	7,37%
4B	E035723 (F048204)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
4C	E035724	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
4D	E035725	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
5B	E035726	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	51 012,4	17	43,30%
5D	E035727	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	51 012,4	4	11,90%
5E	E035728 (F048211)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		SFPQ_HUMAN	Splicing factor, proline- and glutamine-rich	75 423,9	5	8,15%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	5	9,58%
		HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	51 012,4	3	8,62%
6A	E035729 (F048212)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		HNRH1_HUMAN	Heterogeneous nuclear ribonucleoprotein H	49 181,8	2	7,57%
		MALE_ECOLI	Maltose-binding periplasmic protein	43 371,9	25	68,70%
		COAT_BPMS2	Coat protein	13 709,3	3	21,70%
6B	E035730 (F048213)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		HNRH1_HUMAN	Heterogeneous nuclear ribonucleoprotein H	49 181,8	10	35,60%
		MALE_ECOLI	Maltose-binding periplasmic protein	43 371,9	10	38,90%

6C	E035731 (F048258)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		HNRH1_HUMAN	Heterogeneous nuclear ribonucleoprotein H	49 181,8	7	22,50%
		MALE_ECOLI	Maltose-binding periplasmic protein	43 371,9	2	8,84%
6D	E035732 (F048259)	MALE_ECOLI	Maltose-binding periplasmic protein	43 371,9	2	7,07%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	2	3,77%
		HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	51 012,4	2	7,54%
		REF_HEVBR	Rubber elongation factor protein	14 703,9	2	15,20%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
6E	E035733 (F048260)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		MALE_ECOLI	Maltose-binding periplasmic protein	43 371,9	2	8,84%
		HNRPG_HUMAN	Heterogeneous nuclear ribonucleoprotein G	42 187,7	2	6,70%
6F	E035734 (F048261)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	2	3,77%
6G	E035735 (F048262)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	39 634,2	3	14,20%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	2	4,40%
6H	E035736 (F048263)	K1C10_HUMAN	Keratin, type I cytoskeletal 10	59 494,3	11	23,10%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
		K1C9_HUMAN	Keratin, type I cytoskeletal	62 047,8	7	22,20%
		K22E_HUMAN	Keratin, type II cytoskeletal 2 epidermal	65 415,9	7	14,10%
		ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	39 634,2	2	8,18%
		K2C1_HUMAN	Keratin, type II cytoskeletal 1	65 472,7	9	16,50%
7A	E035737 (F048264)	HNRH3_HUMAN	Heterogeneous nuclear ribonucleoprotein H3	36 909,6	3	11,00%
		ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1	37 460,4	4	12,50%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
		ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	39 634,2	2	10,00%
7B	E035738 (F048265)	HNRH3_HUMAN	Heterogeneous nuclear ribonucleoprotein H3	36 909,6	3	9,25%
		ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1	37 460,4	7	28,90%
		TRYP_PIG	Trypsin	24 391,3	4	25,10%
		ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	34 194,4	5	24,10%
7C	E035739 (F048266)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	34 194,4	2	10,60%
7D	E035740 (F048267)	TRYP_PIG	Trypsin	24 391,3	4	25,10%
		ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	38 828,1	3	9,41%

**Tableau 6 : Tableau complet des protéines identifiées en spectrométrie de masse issues du gel de purification des complexes protéiques.**

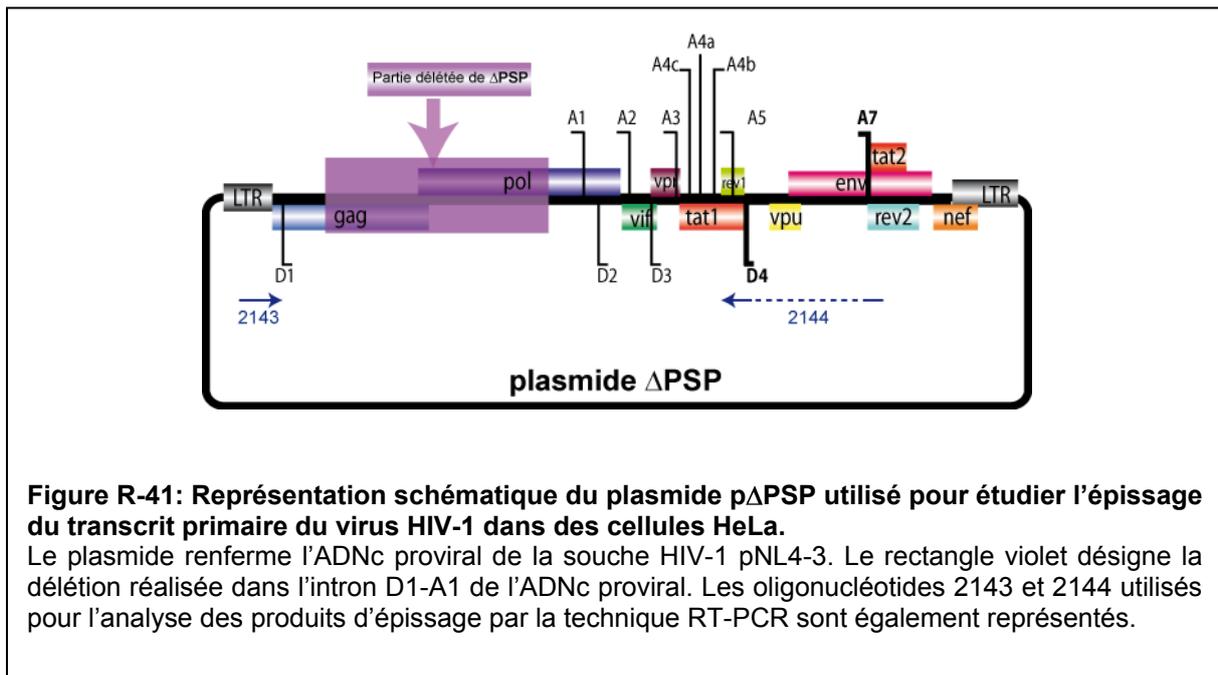
## CHAPITRE V Effets *in cellulo* de la protéine hnRNP K sur l'épissage de VIH-1.

### I. Etat des connaissances au départ et stratégie utilisée pour l'étude

Les facteurs cellulaires régulant l'épissage des ARNm viraux sont considérés comme des cibles thérapeutiques potentielles (Bakkour *et al.*, 2007), il est donc important d'évaluer l'effet que de telles protéines ont sur les différents aspects du métabolisme des ARNm viraux. Les interactions entre protéines cellulaires et séquences ARN agissant en *cis* ont tout d'abord été étudiées en utilisant des minigènes viraux qui représentent une partie du génome viral. Le travail sur des constructions contenant des mini-introns a permis d'étudier localement l'effet de la protéine hnRNP K sur l'épissage *in vitro* du VIH-1. L'épissage du VIH-1 dans son contexte naturel met en compétition tous les sites donneurs et accepteurs d'épissage d'une part, et tous les facteurs interagissant en *trans* sur l'ARN d'autre part. Il est ainsi difficile de prévoir comment les protéines peuvent réguler la balance entre les différents complexes régulateurs du génome viral complet. De plus, les protéines de la famille des hnRNP exercent des effets pléiotropiques sur l'expression des gènes cellulaires et viraux. Afin de valider fonctionnellement les résultats que nous avons obtenus *in vitro*, nous avons testé l'effet de la variation des taux cellulaires de la protéine hnRNP K dans les cellules HeLa sur la production des ARNm viraux. Nous disposons du plasmide pΔPSP fourni par C.M. Stoltzfus (Université d'Iowa, Etats-Unis) (Figure R-41). Ce plasmide renferme l'ADNc proviral de la souche pNL4.3, phylogénétiquement très proche de la souche VIH-1 BRU que nous utilisons *in vitro*. Cet ADNc est tronqué dans l'intron D1-A1, ce qui empêche la production de particules virales infectieuses. Après transfection de cellules HeLa par ce plasmide, le profil d'épissage du transcrit primaire est similaire à celui observé après infection de lymphocytes T par le virus VIH-1 (Purcell *et al.*, 1993; Jacquenet *et al.*, 2001).

Les tests ont été effectués avec deux lots de cellules HeLa, le premier provenant de notre laboratoire et le second provenant du laboratoire de N. Charlet (IGBMC, Illkirch) et présentant un phénotype et une vitesse de croissance différent. Les résultats ont été reproductibles avec ces deux lots de cellules.

Afin de faire varier le taux de protéine hnRNP K cellulaire, nous avons utilisé deux techniques : l'ARN interférence pour diminuer la quantité de hnRNP K dans les cellules, et la surexpression de la protéine hnRNP K.

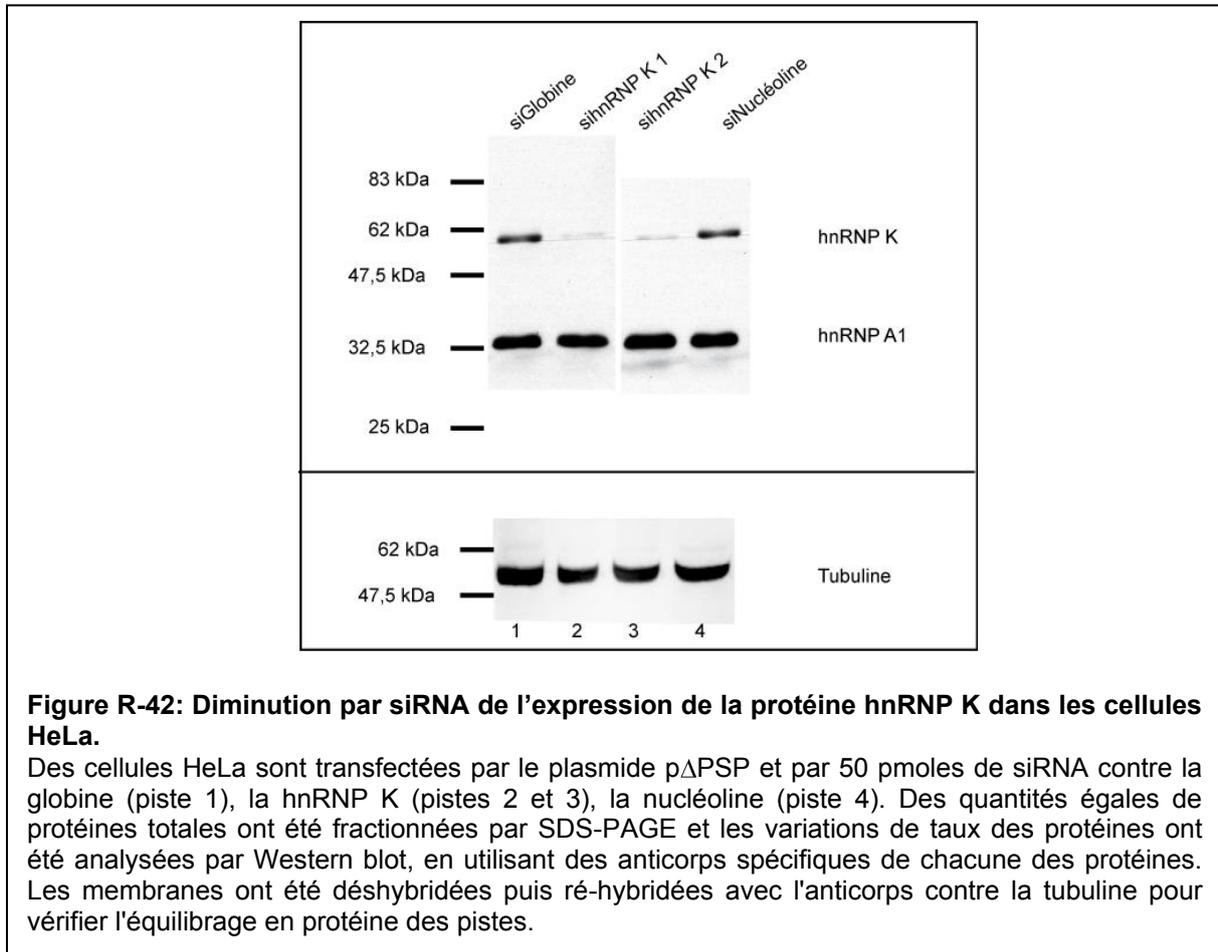


## II. La diminution du taux de la protéine hnRNP K dans les cellules HeLa affecte peu le profil d'épissage *in cellulo* de l'ARN $\Delta$ PSP.

Deux siRNA différents ont été utilisés pour diminuer le taux cellulaire de protéine hnRNP K. Le premier siRNA a été dessiné par C. Aigueperse. Les effets potentiels « off-target » ou faux-positifs ont été évalués lors de leur conception grâce au logiciel siRNA Design Software (<http://i.cs.hku.hk/~sirna/software/sirna.php>). Le second siRNA est le même que celui utilisé dans la publication de Lee (Lee *et al.*, 2007). Les cellules ont tout d'abord été transfectées par les siRNA grâce à la technique de précipité au phosphate de calcium. Après 24 heures, le plasmide p $\Delta$ PSP est introduit dans les cellules HeLa par transfection utilisant le réactif JetPEI. Après de nouveau 24 heures, soit 48 heures de siRNA, la quantité de protéine hnRNP K cellulaire est contrôlée par Western blot selon le protocole décrit dans la partie Matériel et Méthodes (Figure R-42). Ainsi, nous avons pu montrer une diminution de la quantité de protéine hnRNP K de 70% après utilisation des deux siRNA contre hnRNP K, alors que les taux de protéines hnRNP A1 et tubuline ne varient pas. Le Western blot contre la tubuline permet de normaliser la quantité de protéine totale dans chaque piste.

En parallèle, les cellules HeLa sont transfectées par un témoin siRNA dirigé contre la globine et un autre témoin contre la nucléoline. Comme attendu, le taux de protéine hnRNP K dans ces cellules n'est pas perturbé. Ces cellules serviront de témoins pour s'assurer que les effets potentiels observés sur l'épissage sont bien spécifiques de la modulation de taux cellulaire de hnRNP K.

Les cellules HeLa n'ont pas présenté de mortalité excessive en présence des siRNA ni de changement macroscopique de phénotype. Seul un léger retard de croissance a été observé mais spécifique au traitement de transfection, non à la diminution de hnRNP K.

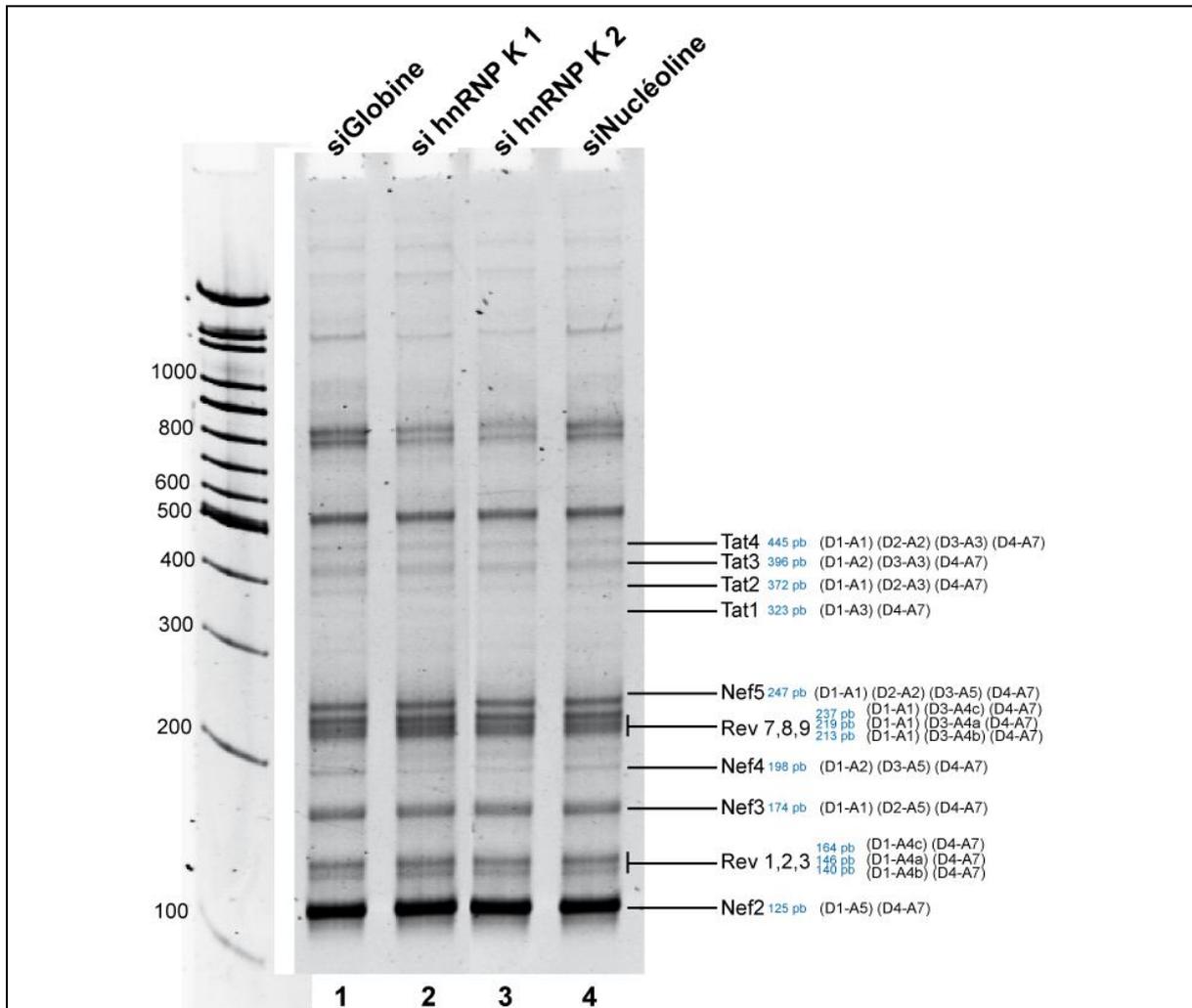


Les ARN totaux des cellules HeLa ont été extraits après récolte des cellules à 48 heures de traitement siRNA. Un traitement à la DNase I a permis d'éliminer toute trace d'ADN p $\Delta$ PSP transfecté qui fausserait les résultats de RT-PCR. Des témoins ont été réalisés afin de s'en assurer.

L'oligonucléotide utilisé pour la transcription inverse (2144) est complémentaire à la séquence générée après épissage entre les sites D4 et A7. Il est couplé pour la PCR avec l'oligonucléotide 2143 qui s'hybride en amont du site donneur D1 sur les positions 700 à 727 (souche pNL4.3). Ainsi, seuls les ARN épissés entre ces deux sites sont analysés par cette approche.

Le profil des ADNc obtenu par RT-PCR après diminution de la protéine hnRNP K de 70 % dans la cellule ne montre qu'une faible diminution de *nef4* par rapport aux témoins (Figure R-

43). Un siRNA dirigé contre la nucléoline nous sert aussi de témoin négatif et a montré le même profil des ADNc en RT-PCR que les témoins sous traitement siRNA globine. Les expériences sont menées en double et répétées au minimum chacune trois fois. Une bonne reproductibilité des résultats a été observée.

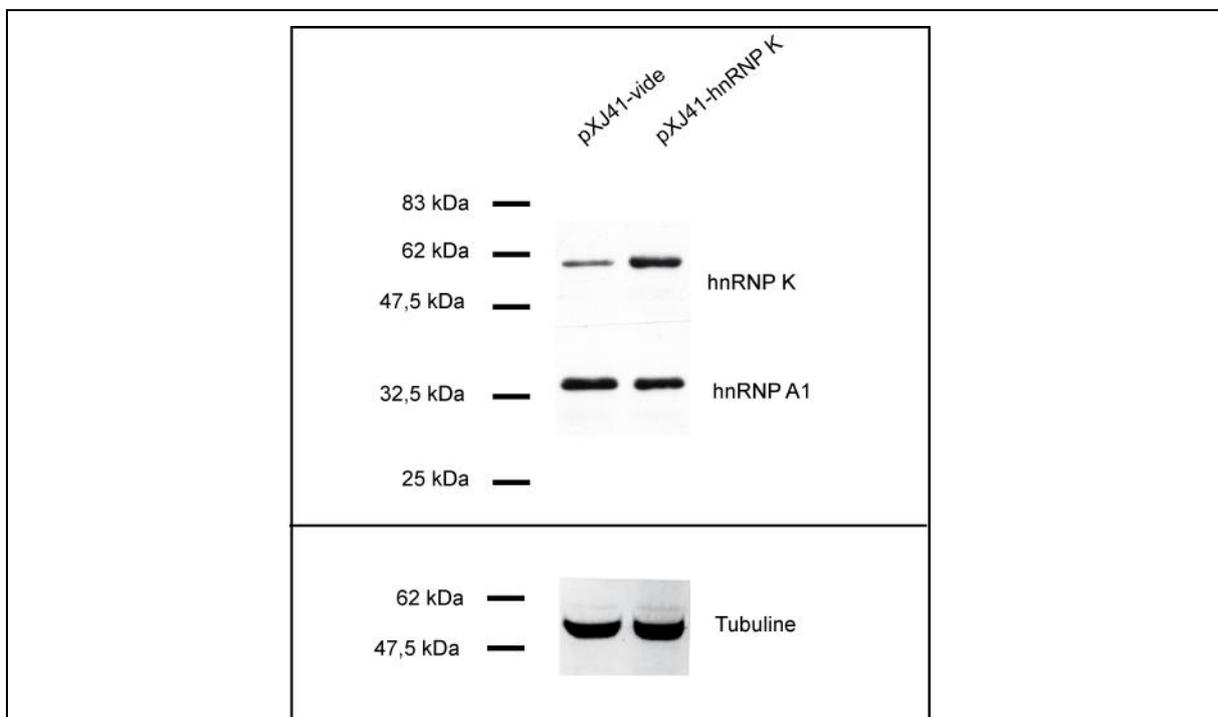


**Figure R-43: Influence de la diminution du taux cellulaire de la protéine hrRNP K par siRNA sur l'épissage alternatif de l'ARN du virus VIH-1.**

Analyse RT-PCR des produits de l'épissage de l'ARN produit par le plasmide p $\Delta$ PSP. Les cellules HeLa ont été transfectées avec le plasmide p $\Delta$ PSP après traitement par les siRNA dirigés contre la globine (piste 1), la protéine hrRNP K (siRNA n°1, piste 2 et siRNA n°2 piste 3) et la nucléoline (piste 4). Chaque analyse RT-PCR a été réalisée avec la même quantité d'ARN total et avec les oligonucléotides spécifiques de l'ARN du virus VIH-1 2143 et 2144. L'identité des ADNc fractionnés correspondant aux différents ARNm de HIV-1 est indiquée à droite selon la nomenclature de Purcell & Martin (Purcell and Martin, 1993).

### III. La surexpression de la protéine hnRNP K dans les cellules HeLa a un effet important sur le profil d'épissage *in cellulo*.

L'ORF codant la protéine hnRNP K a été cloné dans le plasmide pXJ41. Lors de la transfection, 1 µg de plasmide pΔPSP est mélangé à 2 µg de plasmide pXJ41-hnRNP K puis l'ensemble est transfecté dans les cellules HeLa grâce au JetPEI. L'augmentation du taux de la protéine hnRNP K dans les cellules HeLa observée en Western blot est d'environ 50% (Figure R-44). Les cellules ne présentent pas de changements d'aspects macroscopiques. La croissance est restée identique aux cellules non traitées ainsi qu'aux cellules transfectées avec le plasmide vide pXJ41 qui nous sert de témoin négatif.



**Figure R-44: Augmentation de l'expression de la protéine hnRNP K dans les cellules HeLa.**

Des extraits de cellules HeLa sont transfectées par le plasmide pΔPSP et par 2 µg de plasmide pXJ41 vide ou 2 µg de plasmide pXJ41 contenant l'ORF hnRNP K. Des quantités égales de protéines totales ont été fractionnées par SDS-PAGE. Les variations de taux des protéines ont été analysées par Western blot, en utilisant des anticorps spécifiques de chacune des protéines. Les membranes ont été déshybridées puis ré-hybridées avec l'anticorps contre la tubuline pour vérifier l'équilibrage en protéine des pistes.

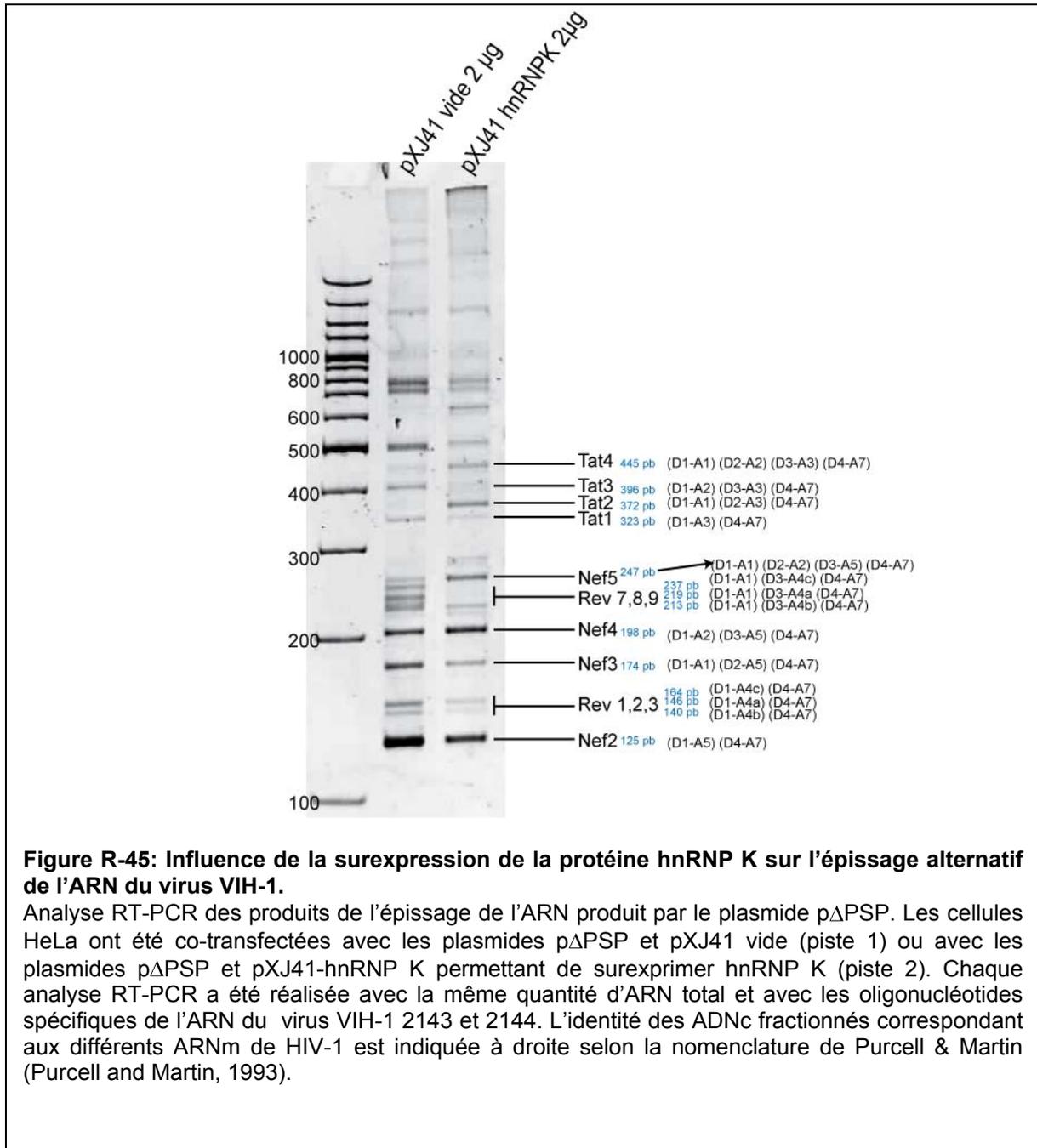
Le profil observé en RT-PCR change distinctement lors de la surexpression de la protéine hnRNP K (Figure R-45). Nous avons observé des variations marquées de la quantité de différents ADNc. Les ADNc *nef2* et *nef3* sont bien moins abondants lors de la surexpression de hnRNP K, tandis que les ADNc *nef4* et *nef5* présentent une nette augmentation. Sur la Figure I-44 sont représentés les différents ARNm épissés de ces ARNm selon Purcell et Martin (Purcell *et al.*, 1993). Les ARNm *nef2* et *nef3* représentent 33,5 % des ARNm de VIH-

1 tandis que les ARNm *nef4* et *nef5* ne représentent que 17,3 % (Purcell *et al.*, 1993). Leurs abondances relatives dans la cellule sont représentées. On peut s'attendre ainsi à une diminution de la protéine Nef totale après traduction, si les ARNm *nef* sont tous traduits de façon équivalente.

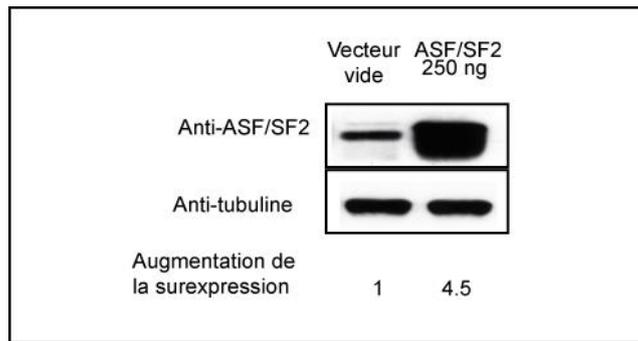
Les quantités des différents ADNc *rev* observables diminuent. Seuls les ADNc *rev1*, 2, 3, 7, 8 et 9 sont observables. Les autres formes sont en trop faibles quantités pour être visualisées. On peut s'attendre à ce que la quantité de protéine traduite diminue de la même façon que les ARNm.

Les taux observés des ADNc *tat* varient selon les sites épissés comme pour *nef*. Les formes *tat1* et *tat3* diminuent lors de la surexpression de hnRNP K tandis que les taux de *tat2* et *tat4* augmentent. *Tat1* et *tat3* représentent 5,4 % de la population des ARNm et *tat2* et *tat4* n'en représentent que 3,3 %. On peut donc s'attendre à une légère baisse de la quantité de protéine Tat traduite dans la cellule après la surexpression de hnRNP K.

Le décryptage des combinaisons des différents sites d'épissage affectés lors de l'augmentation de hnRNP K dans la cellule va aider à comprendre quels sont les sites affectés lors de l'épissage *in cellulo*. Les résultats obtenus *in vitro* vont aider à déchiffrer ces mécanismes.



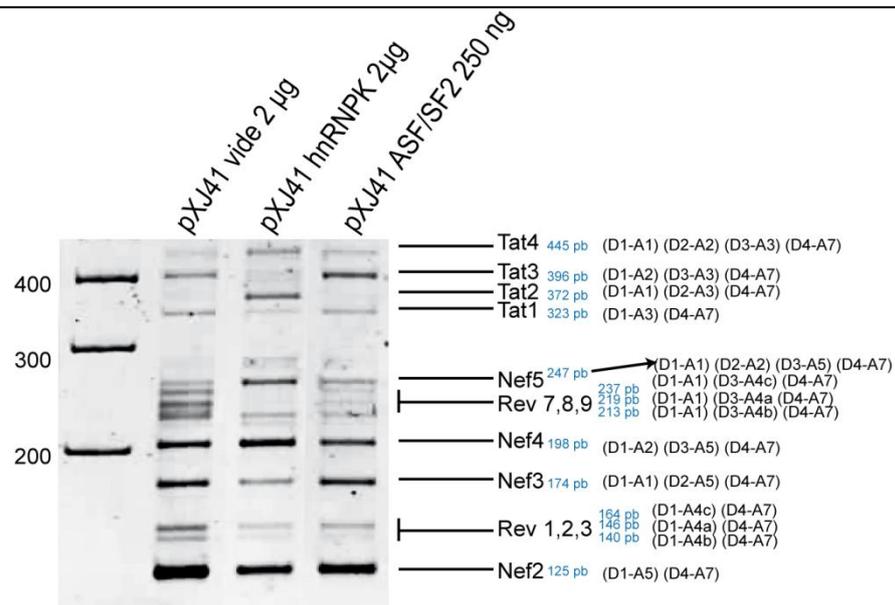
Les changements du profil d'épissage lors de la surexpression de la protéine ASF/SF2 a déjà été décrit (Ropers *et al.*, 2004). Le dépôt côte à côte d'échantillons après surexpression des protéines hnRNP K et ASF/SF2 peut permettre de comparer les effets des deux protéines. La surexpression de la protéine ASF/SF2 dans les cellules HeLa est augmentée d'un facteur d'environ 4.5 (Figure R-46).



**Figure R-46: Augmentation de l'expression de la protéine ASF/SF2 dans les cellules HeLa.**

Des extraits de cellules HeLa sont transfectés par le plasmide p $\Delta$ PSP et par 250 ng de plasmide pXJ41 vide ou 250 ng de plasmide pXJ41 contenant ASF/SF2. Des quantités égales de protéines totales ont été fractionnées par SDS-PAGE. Les variations de taux des protéines ont été analysées par Western blot, en utilisant des anticorps spécifiques de chacune des protéines. Les membranes ont été déshybridées puis ré-hybridées avec l'anticorps contre la tubuline pour vérifier l'équilibre en protéine des pistes.

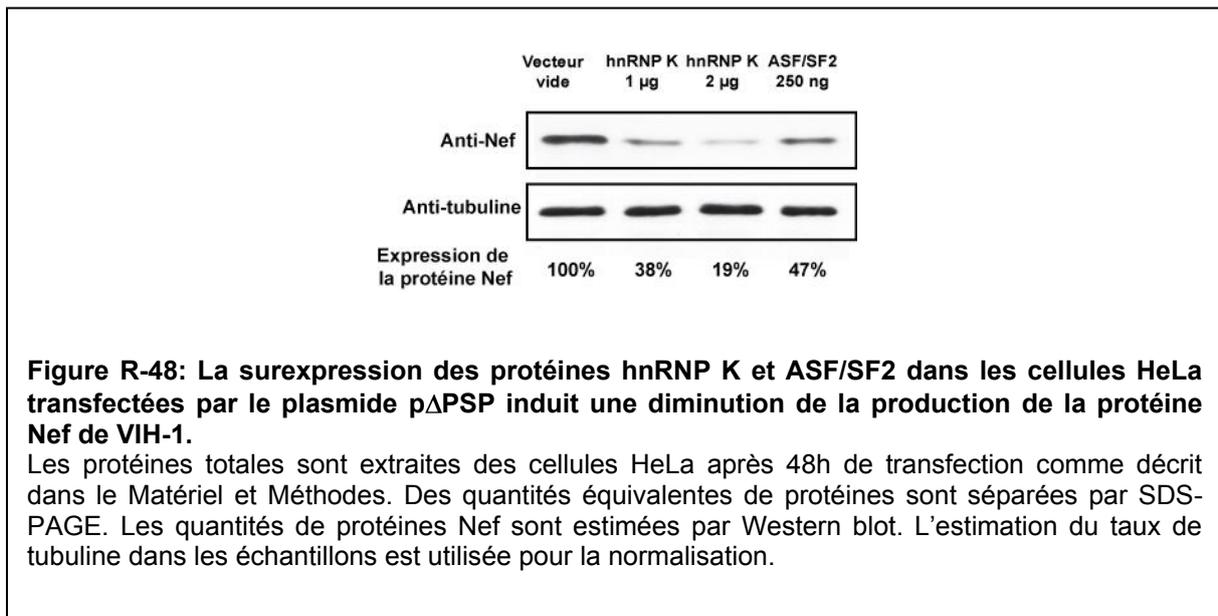
Les effets sur le profil d'épissage de la surexpression de ASF/SF2 montrent une légère augmentation des ARNm *nef3*, *nef5* et *tat4* et une diminution de l'ARNm *nef2*. L'effet sur le profil d'épissage est plus marqué par la surexpression de hnRNP K qu'avec la protéine ASF/SF2 (Figure R-47).



**Figure R-47: Comparaison de l'influence de la surexpression des protéines ASF/SF2 et hnRNP K sur l'épissage alternatif de l'ARN du virus VIH-1.**

Analyse RT-PCR des produits de l'épissage de l'ARN produit par le plasmide p $\Delta$ PSP. Les cellules HeLa ont été co-transfectées avec les plasmides p $\Delta$ PSP et pXJ41 vide ou avec les plasmides p $\Delta$ PSP et pXJ41-hnRNP K ou pXJ42-ASF/SF2. Chaque analyse RT-PCR a été réalisée avec la même quantité d'ARN total et avec les oligonucléotides spécifiques de l'ARN du virus VIH-1 2143 et 2144. L'identité des ADNc fractionnés correspondant aux différents ARNm de HIV-1 est indiquée à droite selon la nomenclature de Purcell & Martin (Purcell and Martin, 1993).

Un Western blot avec des anticorps ciblés contre la protéine Nef de VIH-1 a permis de confirmer la diminution du taux de la protéine Nef dans les extraits protéiques issus de cellules dans lesquelles la protéine hnRNP K a été surexprimée (Figure R-48). Cette baisse est plus importante que celle observée lors de la surexpression de la protéine ASF/SF2, comme ce que pouvait suggérer les taux d'ARNm observés.



#### IV. Résultats et discussions.

L'observation des combinaisons site donneur-site accepteur utilisés pour la production des différents ARNm viraux peut nous permettre de savoir quels couples de sites sont activés lors de l'épissage quand le taux cellulaire de protéine hnRNP K est augmenté. En raison du choix de l'oligonucléotide 2144 pour la PCR, la régulation D4-A7 ne peut pas être observée.

Les variations des taux d'ADNc observées après surexpression sont représentées dans la Figure R-49.

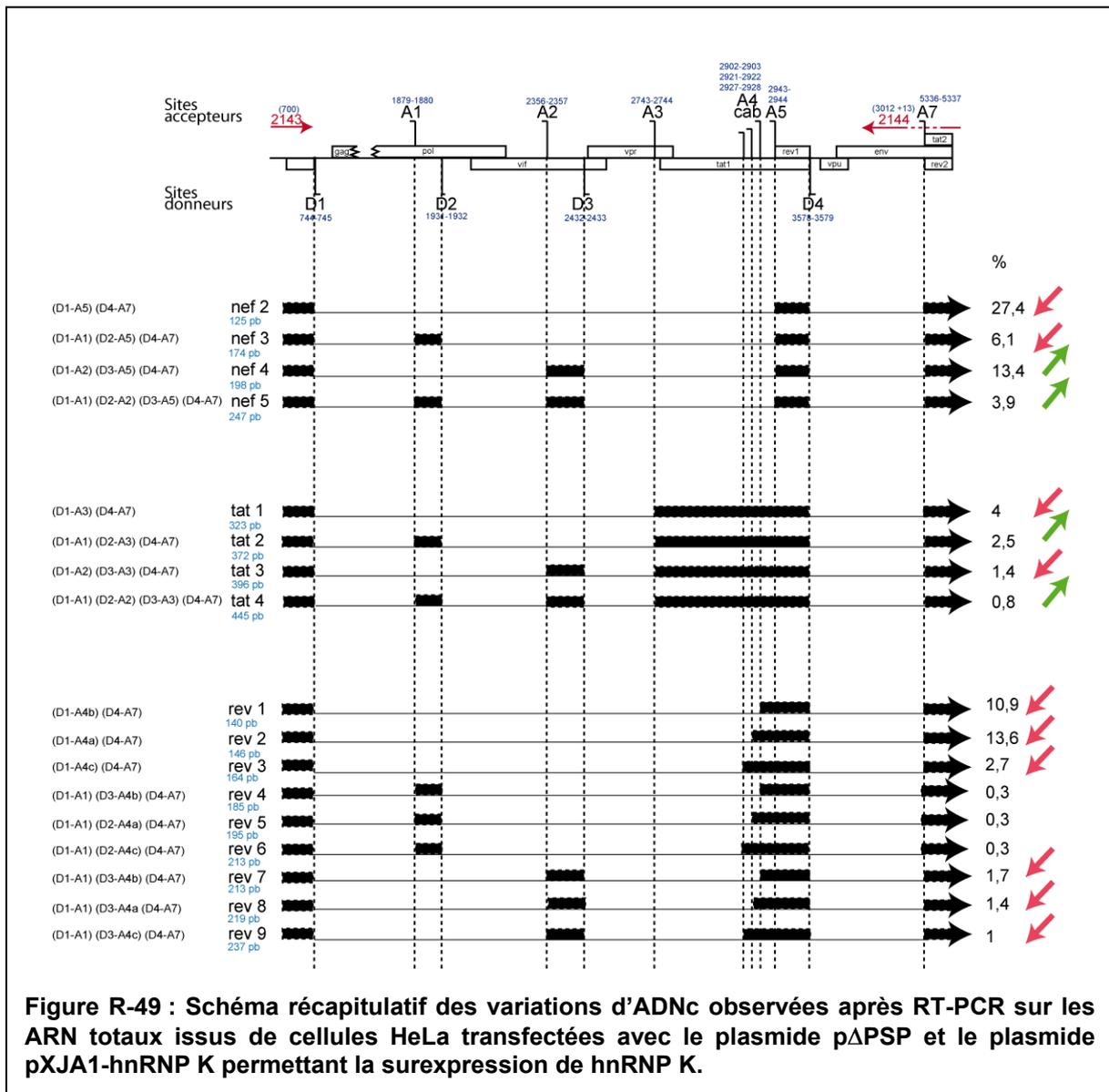
La recherche des sites potentiellement régulés par la protéine hnRNP K va se faire en éliminant les sites qui ne paraissent pas régulés. La seule combinaison donneur-accepteur commune aux ADNc *nef4* et *nef5* dont la quantité augmente est D3-A5. De plus, la combinaison D1-A1 commune à *nef3* et *nef5* ne peut pas être impliquée puisque *nef3* est diminué. Les combinaisons D1-A2 et D2-A2 pourraient également être activées. L'analyse des autres variations obtenues pour les autres ARNm permettra peut-être de le vérifier.

Les deux ADNc *tat2* et *tat4* dont le taux augmente après surexpression de la protéine hnRNP K dans les cellules utilisent un épissage D1-A1, dont on vient de montrer qu'il ne pouvait pas être impliqué dans l'activation. *Tat2* utilise également la combinaison D2-A3 qui, elle, pourrait être activée par hnRNP K. Pour *tat4*, l'activation s'exercerait probablement sur le couple D2-A2 puisque le couple D3-A3 est également utilisé pour l'ARNm *tat3* qui lui est régulé négativement.

L'augmentation de l'épissage au site A5 par la combinaison D3-A5 pourrait expliquer la chute des taux des ADNc *rev* qui sont tous épissés au site accepteur A4. Les ADNc de *rev4* à *rev9* utilisent tous la combinaison D1-A1, ce qui permet de nouveau d'éliminer une régulation à ce niveau.

Les protéines hnRNP sont largement exprimées dans les cellules de mammifères et participent à une large gamme de fonctions cellulaires. Ainsi, l'équipe de Caputi (Jablonski *et al.*, 2009) a montré que les effets de l'augmentation de leur taux cellulaire sur la maturation de l'ARNm viral de VIH-1 ne sont pas toujours corrélés à ceux causés par leur KO. Les analyses semi-quantitatives RT-PCR des ARNm 2kb et 4kb révèlent que le KO par siRNA des protéines hnRNP induit en général un changement moins drastique des isoformes virales d'ARNm que les tests de surexpression. Ceci est probablement dû au fait que le traitement siRNA réduit la plupart des gènes ciblés de moins de 80% tandis que la surexpression peut augmenter leur niveau de plus de 10 fois (Jablonski *et al.*, 2009).

Ces données expliqueraient pourquoi nous observons un si faible effet de nos siRNA alors que le taux de protéine hnRNP K est fortement baissé. Par exemple, le siRNA de la protéine hnRNP A1 augmente le niveau total d'ARN viral et une augmentation des niveaux relatifs des ARNm *gag/pol* et *env1*. Au contraire, l'augmentation de hnRNP A1 réduit drastiquement les taux de *nef2* et *rev2* (Jablonski *et al.*, 2009).



Les travaux de Hadian rapportent que le KO de hnRNP K diminue de plus de 20% les taux intra et extracellulaires de p24 et donc la production de VIH-1, de même que le font les hnRNP A1, Q et R. Ceci montre clairement le lien entre ces hnRNP et la réplication de VIH-1, suggérant que ces hnRNP peuvent fonctionner comme modulateurs positifs de la réplication de VIH-1 (Hadian *et al.*, 2009)

La forte activation de l'épissage des couples D2-A2 et D3-A5 diminue l'épissage aux autres sites. Ainsi, les espèces ARNm présentes majoritairement vont être celles qui seront épissées D2-A2 et D3-A5 au détriment des autres espèces dont le taux diminue lorsque l'on augmente le taux cellulaire de hnRNP K.

La déplétion de hnRNP K par siRNA réduit également la transcription de VIH-1 (Denisenko *et al.*, 1997; Denisenko *et al.*, 1996). Il a été montré que les protéines hnRNP K et Eed agissent coopérativement comme inhibiteur de transcription : hnRNP K lie des facteurs répresseurs comme Zik ou Eed en inhibant l'activité Erk1/2 (Wolf *et al.*, 2008).

Nous observons une forte réduction de *nef2* comme lors de la surexpression de ASF/SF2 (Ropers *et al.*, 2004). La protéine Nef est pour une large part impliquée dans la pathologie SIDA (Arora *et al.*, 2002). Nef coopère fonctionnellement avec le trans-activateur viral Tat en dérégulant le promoteur viral grâce à un complexe kinase associé à Nef (NAKC). hnRNP K va interagir avec Nef dans le contexte du complexe NAKC et recrute Lck, PKC $\delta$  et la kinase PI-3. La formation de ce complexe est suffisante pour activer Lck et Erk1/2, lequel, dans les cellules T, va causer une augmentation de la transcription dépendante de Tat (Wolf *et al.*, 2008). L'omission de hnRNP K dans le complexe NAKC (ajout seulement de Nef, PKC, Eed et Lck) réduit notablement l'effet d'activation de la transactivation dépendante de Tat par dérégulation transcriptionnelle. Son activité de plateforme d'amarrage est indispensable à ce mécanisme.

Probablement par translocation à la membrane plasmique et des modifications post-traductionnelles, hnRNP K a un rôle activateur dans la transcription en servant de plateforme d'amarrage pour les kinases cytoplasmiques activées. Cette conversion fonctionnelle peut être induite par la formation du complexe NAKC ou d'autres stimuli activateurs.

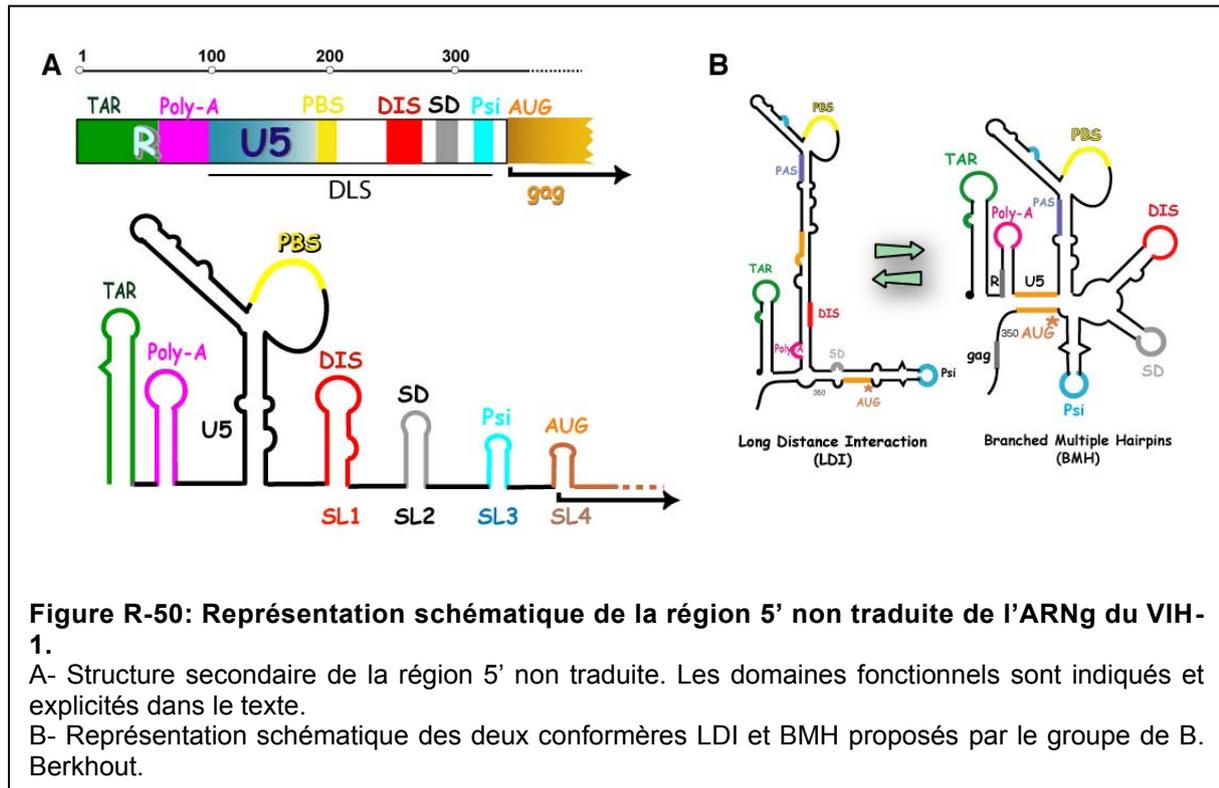
La présence de hnRNP K dans le complexe NAKC est indispensable pour lever complètement la répression et permettre l'activation de la transcription de VIH-1 par Tat. La variation des différentes espèces de *tat* observée lors de la surexpression de hnRNP K laisse envisager une diminution de la protéine totale. S'il y a effectivement une diminution de Tat, s'agirait-il d'un système de rétro-contrôle négatif ?

## CHAPITRE VI Relation entre épissage et dimérisation de l'ARN de VIH-1. Collaboration avec le Dr. Jean-Christophe Paillart, Laboratoire Architecture et Réactivité de l'ARN CNRS-UPR 9002, Strasbourg.

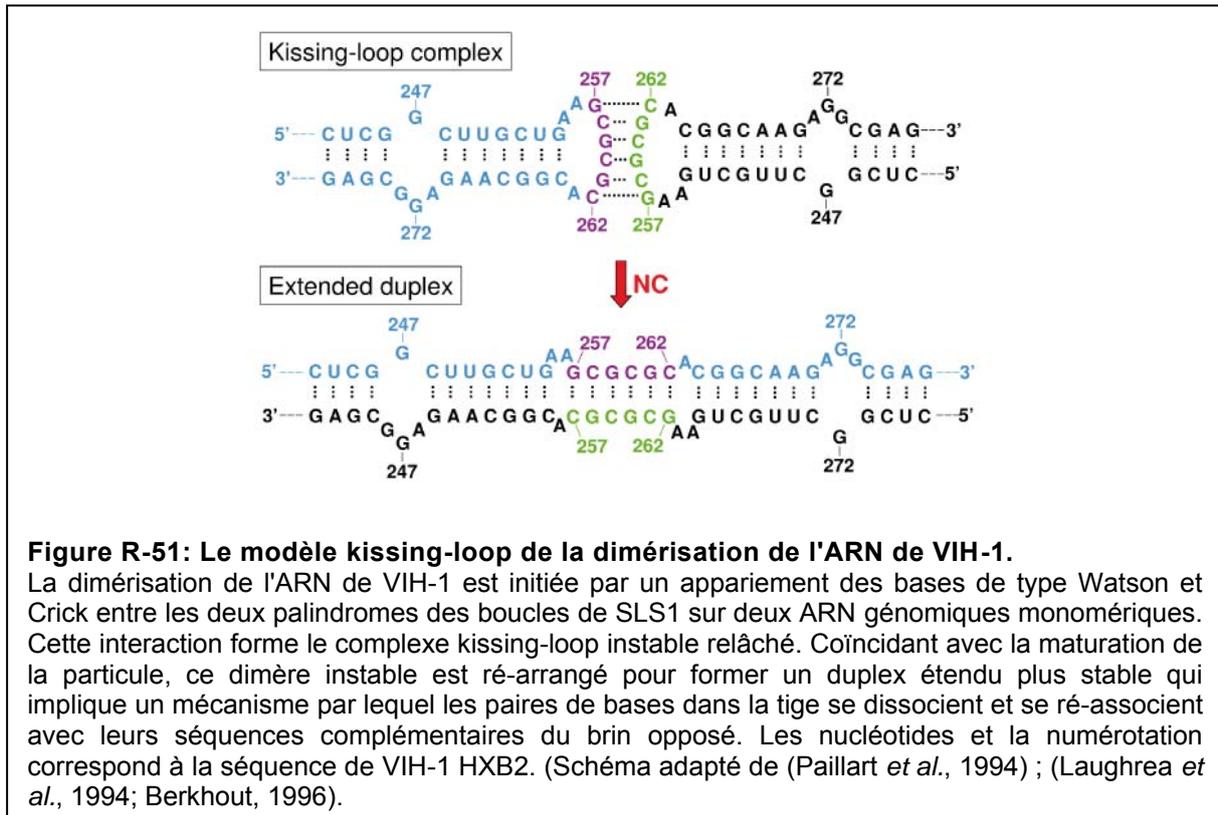
### I. Etat des connaissances au départ.

Le génome du virus de l'immunodéficience humaine de type 1 (VIH-1) est constitué de deux molécules d'ARN homologues de polarité positive. La nature dimérique de l'ARN génomique (ARNg) est conservée à l'intérieur de la famille des Retroviridae, indiquant un rôle crucial de la dimérisation dans la réplication virale. En effet, la dimérisation de l'ARNg est un pré-requis à son encapsidation et offre au virus une opportunité unique de recombinaison lors de l'étape de rétrotranscription, lui permettant ainsi d'accroître sa variabilité. Le site nécessaire à l'initiation de la dimérisation (DIS) est présent aussi bien sur les ARNg que sur les ARN viraux épissés du VIH-1. Ce site est situé dans la région 5' non traduite de l'ARNg du VIH-1, une courte région de 330-350 nucléotides largement étudiée et dont plusieurs modèles de structure secondaire ont été proposés (Baudin *et al.*, 1993; Damgaard *et al.*, 2004; Wilkinson *et al.*, 2008). Cette région, la plus conservée du génome, peut être décomposée en plusieurs domaines structurellement et fonctionnellement indépendants impliqués dans des étapes clés de la réplication virale (Figure R-50).

La région R comprend le signal de trans-activation de la transcription (TAR), site de fixation de la protéine Tat, et le signal de poly-adénylation. La région U5 contient le site de fixation de l'ARNt<sub>3</sub><sup>Lys</sup> amorce de la rétrotranscription (PBS). Plus en aval sont localisées des tiges-boucles qui correspondent au site d'initiation de la dimérisation (DIS), au site majeur donneur d'épissage (SD), au signal d'encapsidation de l'ARNg (Psi ou Ψ) et au site de démarrage de la traduction du gène *gag*.

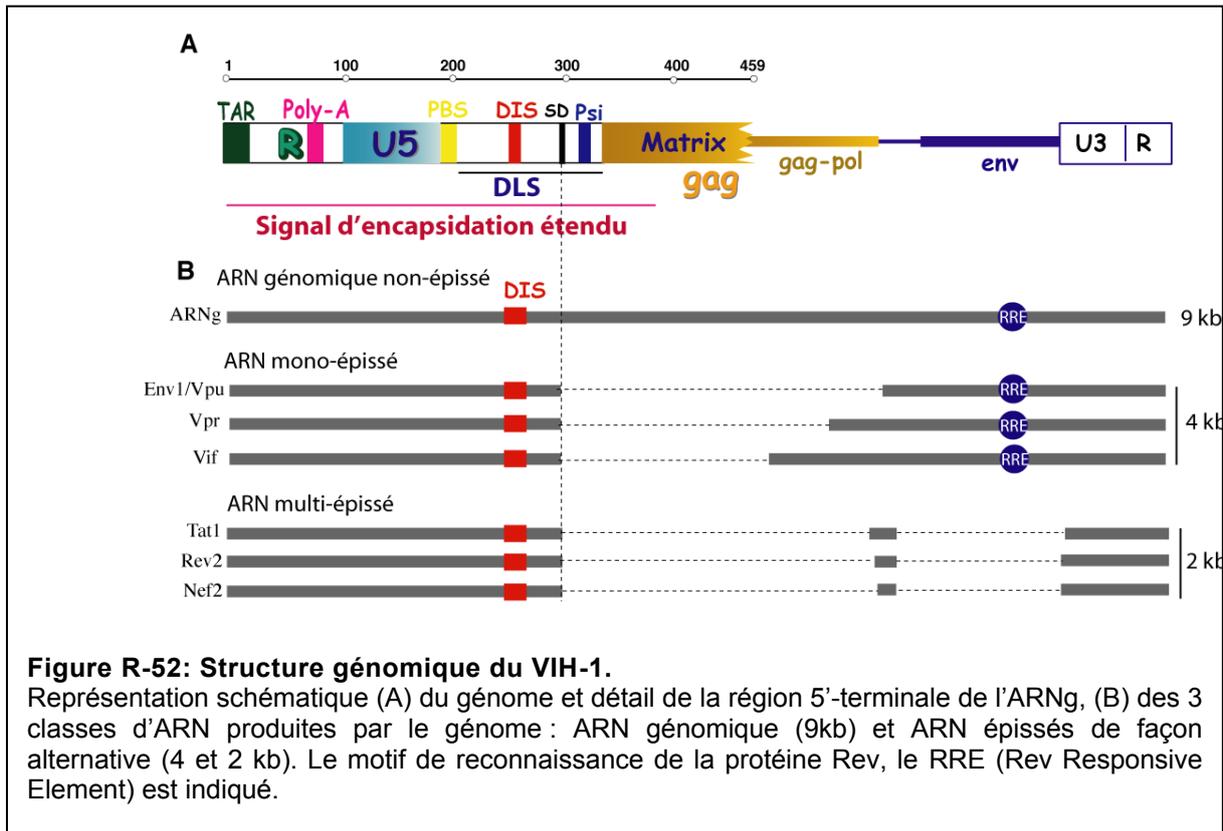


La dimérisation de l'ARNg du VIH-1 s'établit au niveau du site d'initiation de la dimérisation (DIS) de l'ARNg du VIH-1. Il s'agit d'une courte séquence localisée en amont du site donneur D1 qui se replie en tige-boucle présentant une séquence auto-complémentaire de 6 nucléotides au sein d'une boucle de 9 nucléotides (Figure R-51) (Paillart *et al.*, 1994; Laughrea *et al.*, 1996; Muriaux *et al.*, 1995). La dimérisation se fait par le biais d'interactions de type Watson-Crick entre les séquences auto-complémentaires de chacun des ARN monomères pour former un complexe boucle-boucle (ou « kissing complex ») qui est converti en un complexe plus stable (qui pourrait être un duplex étendu) par l'intermédiaire du domaine NC du précurseur Gag (Figure R-51) (Paillart *et al.*, 1996; Muriaux *et al.*, 1996). La nature exacte du dimère mature au sein des particules virales reste cependant à déterminer (Paillart *et al.*, 2004).



La localisation 5' terminale du DIS et sa proximité du site donneur (SD) majeur d'épissage suggère l'existence d'un lien entre les phénomènes de dimérisation et d'épissage, d'autant plus que le complexe EJC (Exon-Junction Complex) serait déposé au niveau du DIS lors de l'épissage. Ainsi, notre but était de tester l'hypothèse selon laquelle la dimérisation de l'ARN pourrait affecter l'épissage ou réciproquement, si l'épissage pourrait inhiber la dimérisation des ARNm viraux. La réaction d'épissage pourrait ainsi être un point de contrôle important pour la dimérisation et la localisation cellulaire du dimère d'ARN.

Les ARN épissés du VIH-1 possèdent aussi un DIS fonctionnel et peuvent théoriquement former des hétérodimères avec les ARNg (Sinck *et al.*, 2007) (Figure R-52). Des ARN viraux épissés sont également retrouvés dans les particules virales, mais en faible quantité (Russell *et al.*, 2003; Houzet *et al.*, 2007). Aucune étude n'a pour l'instant montré que le DIS des ARN épissés du VIH-1 est fonctionnel dans la cellule (noyau et cytoplasme) et si celui-ci permet la formation d'hétérodimères entre ces ARN.



**Figure R-52: Structure génomique du VIH-1.**

Représentation schématique (A) du génome et détail de la région 5'-terminale de l'ARNg, (B) des 3 classes d'ARN produits par le génome : ARN génomique (9kb) et ARN épissés de façon alternative (4 et 2 kb). Le motif de reconnaissance de la protéine Rev, le RRE (Rev Responsive Element) est indiqué.

## II. Stratégie utilisée pour l'étude.

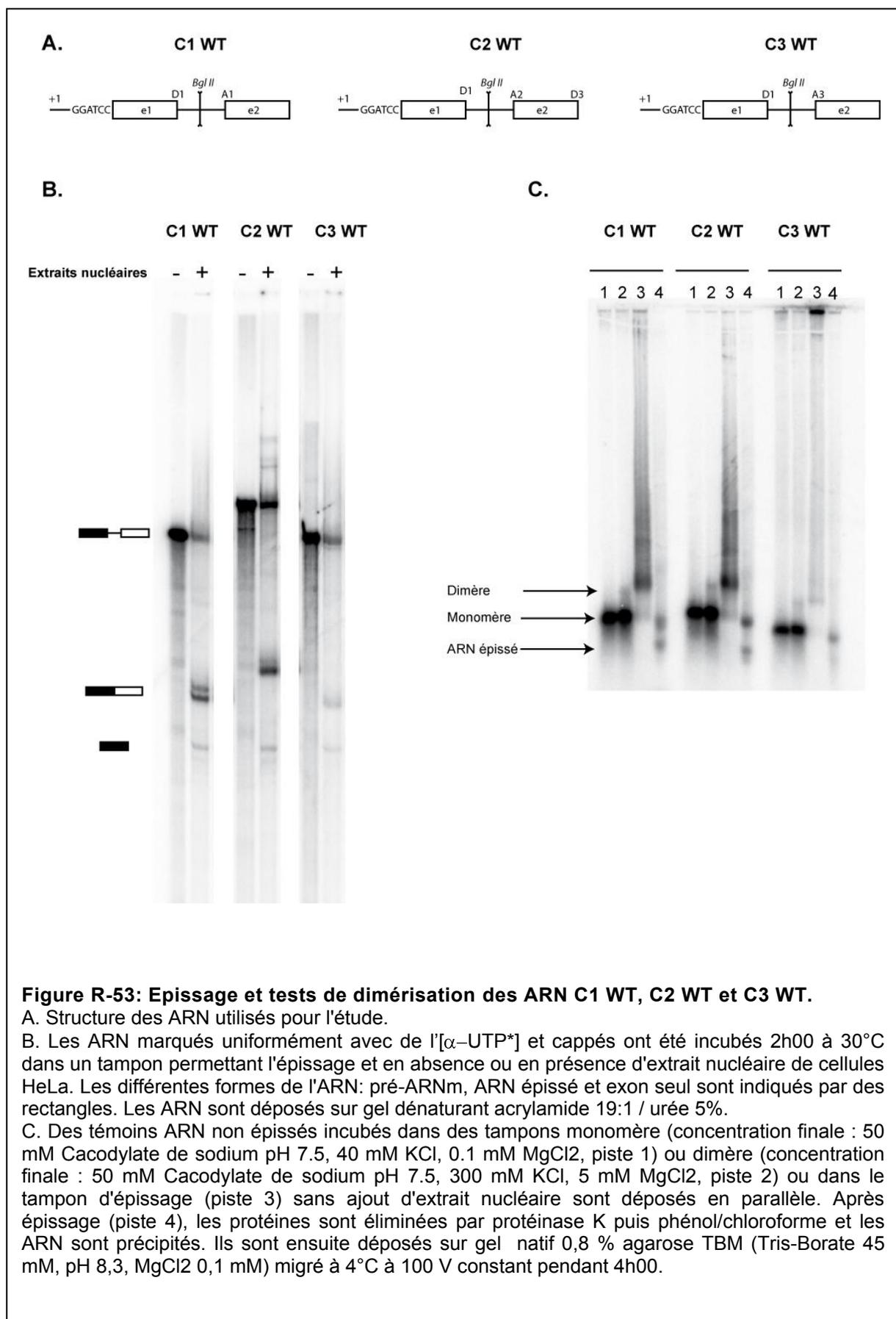
Nous avons réalisé des réactions d'épissage *in vitro* à partir d'un extrait nucléaire de cellules HeLa (Ropers *et al.*, 2004) et d'ARN pré-messagers (WT et mutants) du VIH-1 (Jacquenot *et al.*, 2001; Jacquenet *et al.*, 2001; Ropers *et al.*, 2004). La nature dimérique des produits de la réaction a été analysée sur gel d'agarose en condition native. En parallèle, nous avons vérifié sur gel dénaturant que les ARN ont effectivement subi une réaction d'épissage.

## III. Résultats et discussion.

Laurence Damier a débuté en 1996 une étude consistant à déterminer si la dimérisation de l'ARN rétroviral empêchait son épissage. Pour cela il était nécessaire de mettre au point des conditions où le dimère soit stable et qui soient compatibles avec les conditions requises pour qu'il y ait épissage *in vitro* dans un extrait nucléaire. Au cours d'un séjour à Strasbourg, elle a essayé en collaboration avec J.C. Paillart et R. Marquet de mettre au point de telles conditions de dimérisation. Ils ont pu obtenir des dimères stables à une concentration en cacodylate de sodium de 50 mM, en MgCl<sub>2</sub> de 5 mM et en KCl de 300 mM. Cependant ces conditions se sont révélées incompatibles avec les conditions requises pour qu'il y ait épissage *in vitro*. Ils ont donc dû renoncer à tester l'effet de la dimérisation sur l'utilisation du site donneur D1.

J'ai effectué mon séjour dans cette même équipe à Strasbourg pour apprendre la technique de dimérisation. Au laboratoire AREMS, j'ai tout d'abord travaillé avec les ARNm C1 WT, C2 WT et C3 WT (Figure R-53). Le rendement d'épissage *in vitro* de ces trois pré-ARNm est différent car chaque site accepteur a une efficacité différente. Il sera intéressant d'observer si cette efficacité d'épissage influe sur la dimérisation de ces pré-ARNm qui ont une partie 5' commune mais la partie 3' différente. Dans la construction C1, le motif de dimérisation **aagcgcgcac** sauvage a ensuite été remplacé par la séquence **acaaacacac** (construction C1 A mut), ce qui supprime la capacité de l'ARN à dimériser. Ce site de dimérisation se trouve dans le premier exon, à 28 nucléotides en amont du site donneur 5' d'épissage. Selon le logiciel MFOLD, cette mutation ne semble pas modifier la structure secondaire de l'ARN, mais des expériences de structure seraient nécessaires pour le vérifier.

Afin d'étudier en parallèle l'épissage et la dimérisation, l'ARNm a été transcrit radioactivement dans des conditions permettant l'ajout d'une coiffe.



**Figure R-53: Epissage et tests de dimérisation des ARN C1 WT, C2 WT et C3 WT.**

A. Structure des ARN utilisés pour l'étude.

B. Les ARN marqués uniformément avec de l' $[\alpha\text{-UTP}^*]$  et cappés ont été incubés 2h00 à 30°C dans un tampon permettant l'épissage et en absence ou en présence d'extrait nucléaire de cellules HeLa. Les différentes formes de l'ARN: pré-ARNm, ARN épissé et exon seul sont indiqués par des rectangles. Les ARN sont déposés sur gel dénaturant acrylamide 19:1 / urée 5%.

C. Des témoins ARN non épissés incubés dans des tampons monomère (concentration finale : 50 mM Cacodylate de sodium pH 7.5, 40 mM KCl, 0.1 mM MgCl<sub>2</sub>, piste 1) ou dimère (concentration finale : 50 mM Cacodylate de sodium pH 7.5, 300 mM KCl, 5 mM MgCl<sub>2</sub>, piste 2) ou dans le tampon d'épissage (piste 3) sans ajout d'extrait nucléaire sont déposés en parallèle. Après épissage (piste 4), les protéines sont éliminées par protéinase K puis phénol/chloroforme et les ARN sont précipités. Ils sont ensuite déposés sur gel natif 0,8 % agarose TBM (Tris-Borate 45 mM, pH 8,3, MgCl<sub>2</sub> 0,1 mM) migré à 4°C à 100 V constant pendant 4h00.

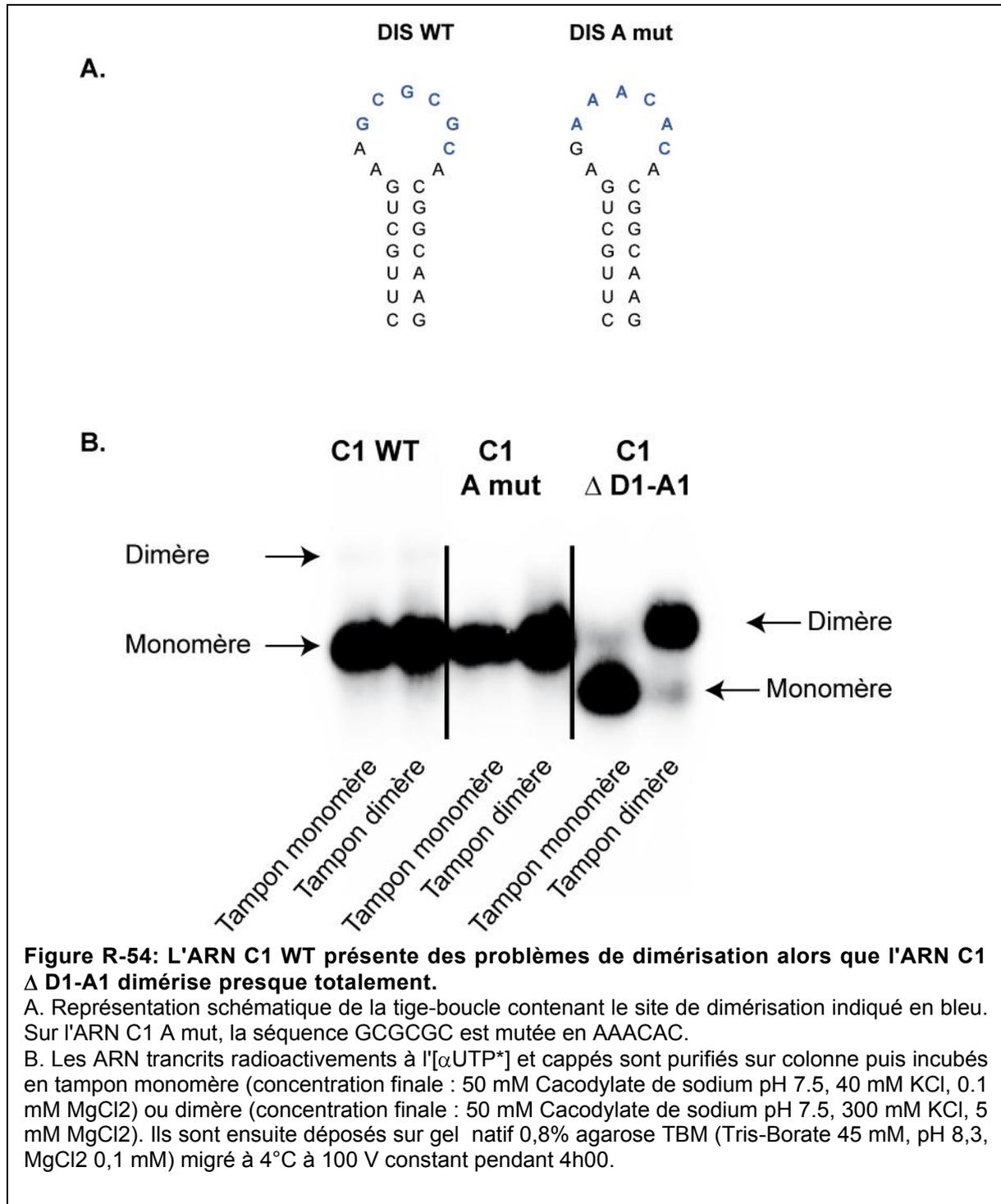
Les ARN transcrits uniformément avec de l' [ $\alpha$ UTP\*] et cappés ont été épissés dans 40% d'extrait nucléaire de cellules HeLa pendant 2h00 et déposés en parallèle sur gel dénaturant acrylamide 19 :1 / urée et sur gel natif TBM (Tris-Borate 45 mM, pH 8.3, MgCl<sub>2</sub> 0.1 mM, agarose 0.8%). Pour déterminer les formes monomères et dimères, des ARN incubés dans des tampons et des conditions permettant l'obtention de ces formes ont été déposés en parallèle sur le gel d'agarose. Pour connaître l'état de dimérisation de l'ARN dans le tampon d'épissage, il a été incubé dans les conditions d'épissage mais sans ajout d'extrait nucléaire. Le taux de dimérisation des ARN C1, C2 et C3 WT apparait très faible. Les expériences ont été refaites plusieurs fois sans succès. De façon surprenante, si la dimérisation est très mauvaise dans le tampon de dimérisation, l'ARN incubé dans le tampon d'épissage apparait en forme presque complètement dimère. Les formes épissées et non épissées apparaissent bien sur le gel d'agarose.

Il est nécessaire de pouvoir observer sur gel la migration du potentiel dimère du pré-ARNm ainsi que de l'ARNm. Avec l'ARNm C1 WT, l'ARNm de forme dimère migrerait au niveau du pré-ARNm, ce qui le rendra impossible à voir. Par conséquent, nous avons réalisé un ARN délété de l'intron D1-A1 nommé C1  $\Delta$  D1-A1 correspondant à l'ARNm pour observer sa dimérisation.

Plusieurs conditions de purification des transcrits ont été testés pour essayer de résoudre les problèmes de dimérisation. La première technique consiste à faire une double précipitation des ARN. Après transcription, les contaminants protéiques sont extraits au phénol/chloroforme puis l'ajout d'acétate d'ammonium (AcONH<sub>4</sub>) permet une première précipitation des grands fragments en éliminant les nucléotides radioactifs non incorporés. Le culot d'ARN est ensuite lavé au tris-NaCl puis précipité une seconde fois. Cette méthode est celle classiquement utilisée au laboratoire. La seconde technique est la purification des transcriptions sur gel de polyacrylamide suivi de l'éluion des fragments d'ARN du gel et leur précipitation à l'acétate de sodium. Cette méthode a été abandonnée car des bandes supplémentaires non spécifiques apparaissent sur gel lors de l'épissage sans que nous ne puissions l'expliquer. La troisième méthode utilisée est celle qui tend à remplacer la méthode de double précipitation. La qualité des transcriptions est vérifiée par le dépôt d'un aliquot sur gel puis les transcriptions sont purifiées grâce aux colonnes « Mini Quick Spin Columns » de Roche.

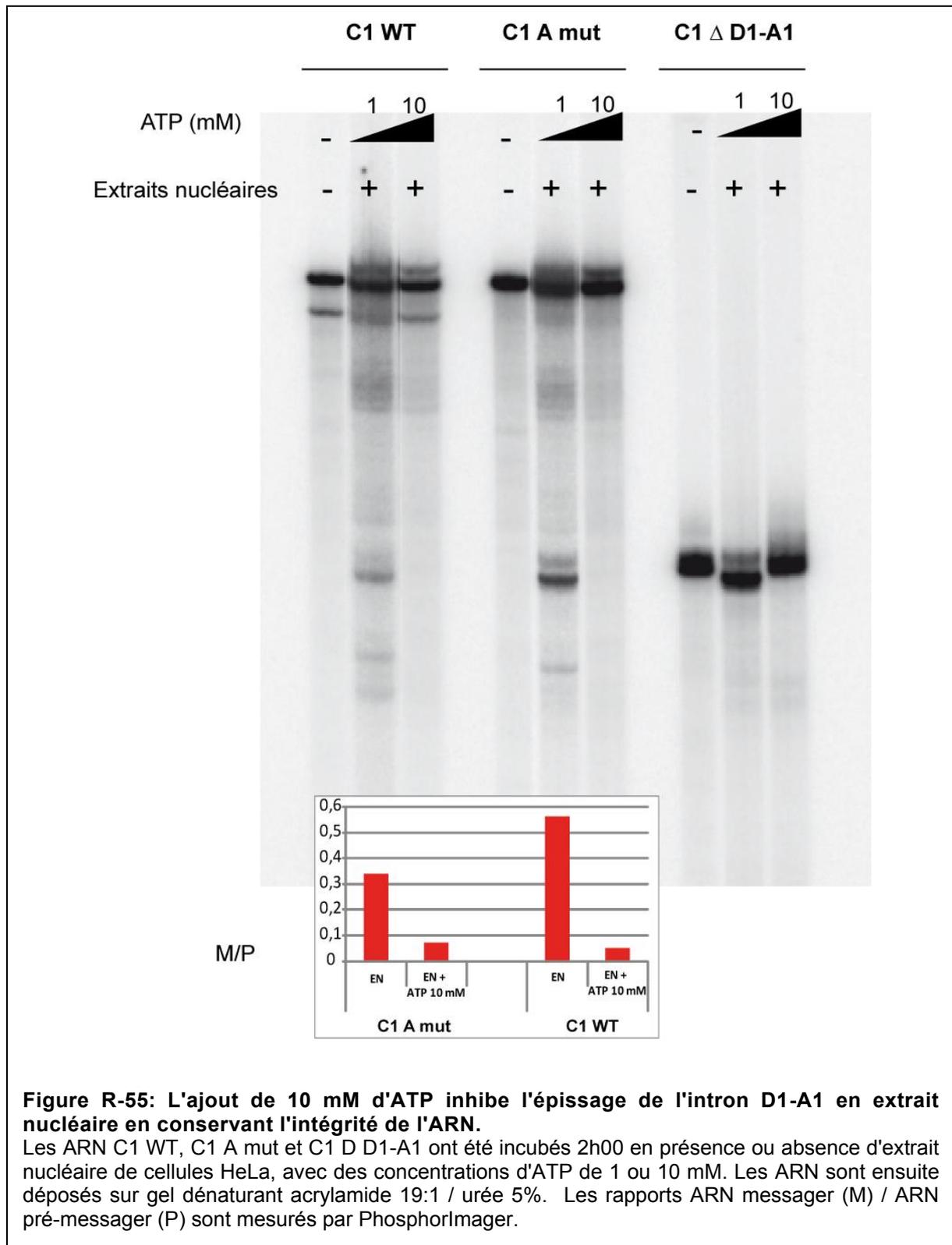
L'ARN C1  $\Delta$  D1-A1 est obtenu en éliminant l'intron, ce qui permet de produire un ARN équivalent à l'ARN épissé. Si les ARN C1 WT et C1 A mut ne montrent pas de dimérisation malgré toutes les tentatives, le mutant C1  $\Delta$  D1-A1 montre près de 100% de dimérisation. Il a

déjà été démontré que la structure ARN en amont du site donneur D1 est similaire dans l'ARN VIH-1 épissé (Sinck *et al.*, 2007). Les ARN épissés ont donc la capacité de dimériser. Ce témoin sans intron, et ne subissant donc pas de processus d'épissage, est important pour localiser le dimère de l'ARN épissé mais aussi pour déterminer si l'effet potentiellement observé dépend bien du phénomène d'épissage (Figure R-54).



Il était primordial aussi de pouvoir incuber les ARN dans un extrait nucléaire sans que le processus d'épissage n'ait lieu, pour observer les phénomènes spécifiques à l'épissage

influençant la dimérisation. L'ajout de 10 mM d'ATP permet d'incuber l'ARN dans un extrait nucléaire en inhibant presque totalement l'épissage, de 90 % avec l'ARN C1 WT et de 80 % avec l'ARN C1 A mut (Figure R-55).

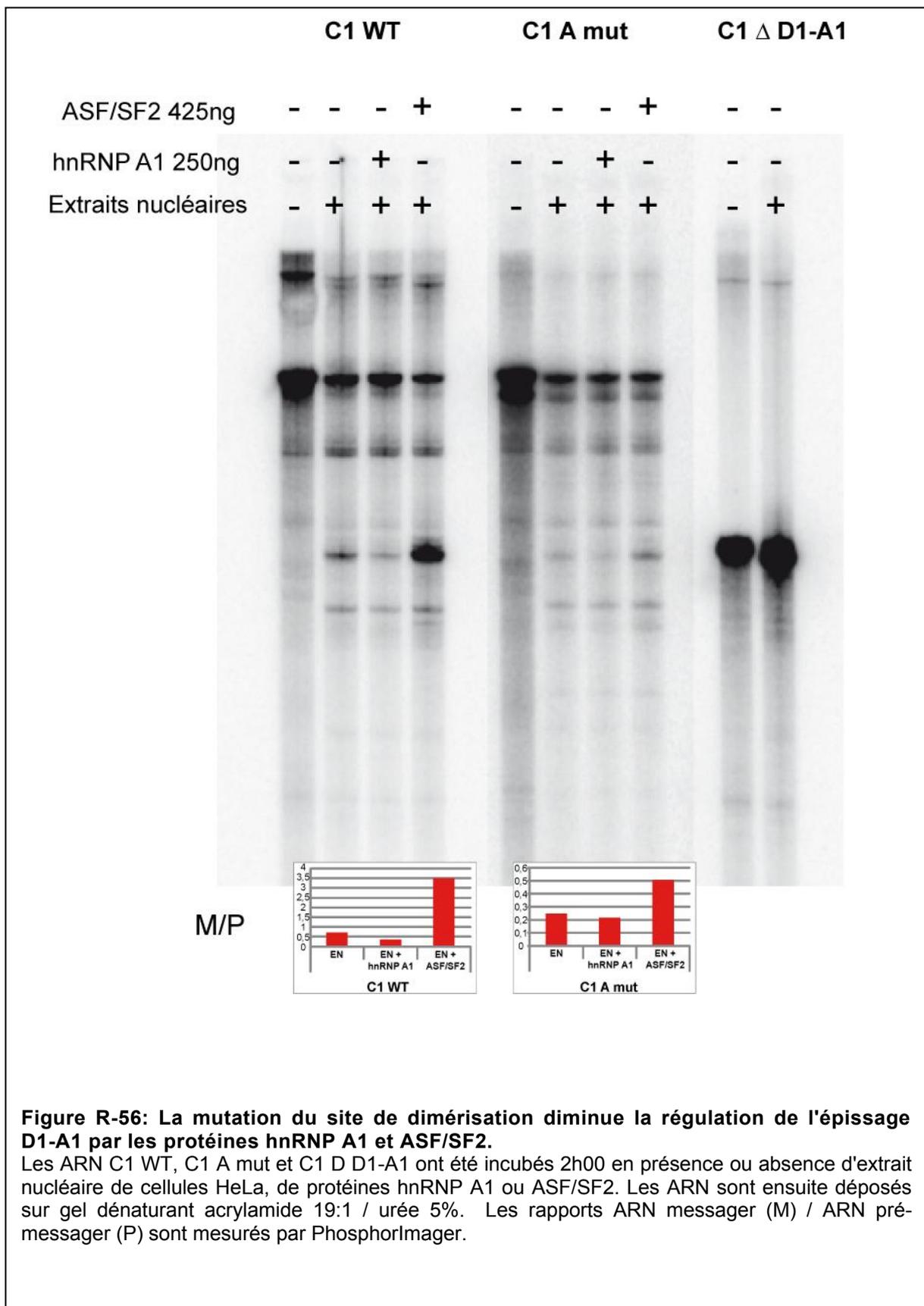


**Figure R-55: L'ajout de 10 mM d'ATP inhibe l'épissage de l'intron D1-A1 en extrait nucléaire en conservant l'intégrité de l'ARN.**

Les ARN C1 WT, C1 A mut et C1 D D1-A1 ont été incubés 2h00 en présence ou absence d'extrait nucléaire de cellules HeLa, avec des concentrations d'ATP de 1 ou 10 mM. Les ARN sont ensuite déposés sur gel dénaturant acrylamide 19:1 / urée 5%. Les rapports ARN messager (M) / ARN pré-messager (P) sont mesurés par PhosphorImager.

Nous avons réussi à trouver une condition qui permet d'incuber les ARN en extrait nucléaire tout en évitant tout phénomène d'épissage et sans dégrader l'ARN, ce qui sera très utile comme témoin pour des expériences ultérieures.

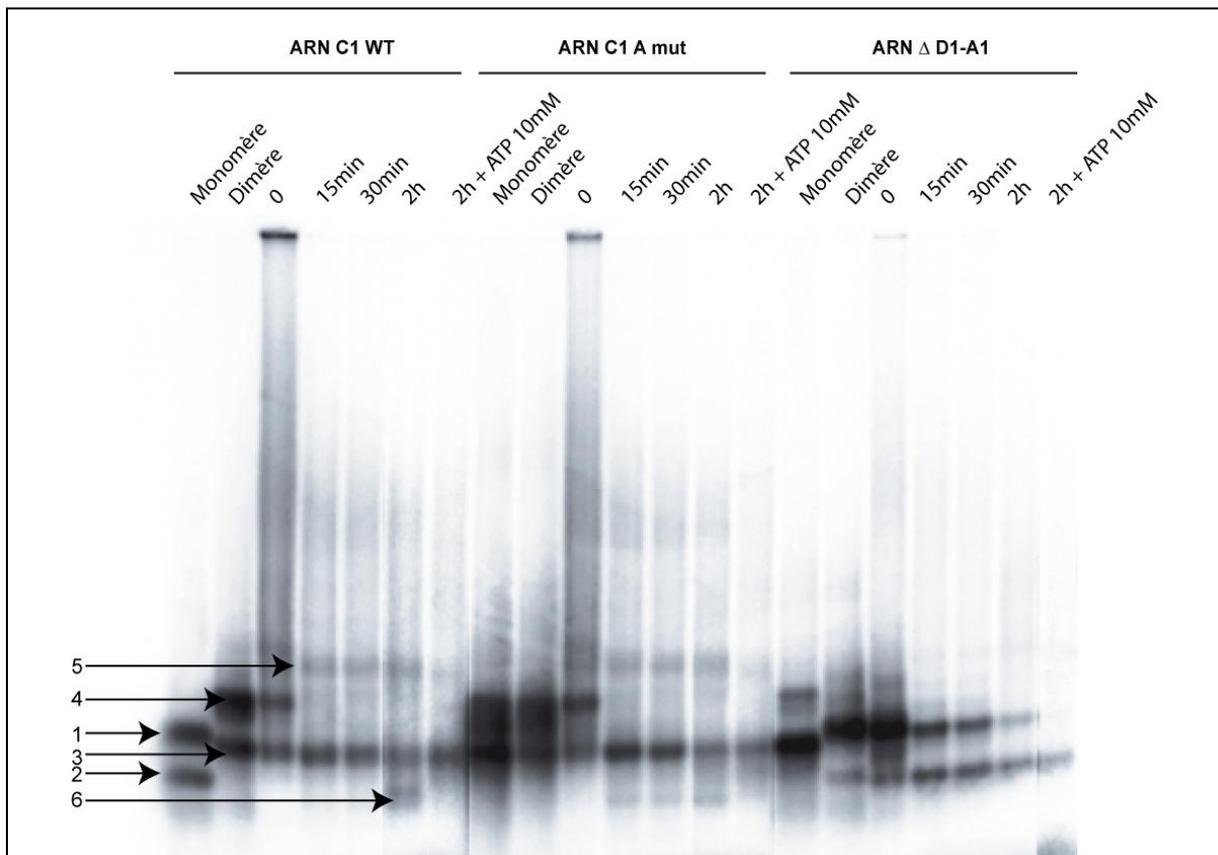
Une des observations que nous pouvons faire même en l'absence de dimérisation de nos ARN est que le choix de ces nucléotides (**aagcgcgcac** remplacé par **acaaacacac**) pour la mutation peut provoquer d'importants bouleversements sur l'épissage *in vitro*. Il a été démontré que des mutations près du site de dimérisation DIS modifient l'épissage au site D1. Il n'est donc pas étonnant d'observer un effet négatif de ces mutations sur l'épissage. De plus, la présence de répétitions CA peut réguler le choix du site d'épissage alternatif positivement ou négativement selon le contexte et la distance de la répétition avec le site d'épissage (Hui *et al.*, 2005). Nous avons voulu savoir si cette mutation affecte aussi l'effet des protéines hnRNP A1 et ASF/SF2 sur la régulation de l'épissage (Figure R-56). Le taux basal d'épissage du mutant C1 A mut est nettement inférieur à celui observé avec la séquence de dimérisation sauvage. L'effet des protéines régulatrices de l'épissage hnRNP A1 et ASF/SF2 va aussi être différent après la mutation. La protéine hnRNP A1 inhibe l'épissage d'environ 50 % sur l'ARN C1 WT et seulement de 15% sur l'ARN C1 A mut. La régulation de l'épissage est aussi modifiée lors de l'ajout d'ASF/SF2. Tandis que l'on observe une activation de l'épissage de 7 fois avec l'ARN C1 WT, l'épissage est activé seulement de 2 fois avec l'ARN C1 A mut. La régulation de l'épissage par les protéines hnRNP A1 et ASF/SF2 est donc aussi affectée par cette mutation.



**Figure R-56: La mutation du site de dimérisation diminue la régulation de l'épissage D1-A1 par les protéines hnRNP A1 et ASF/SF2.**

Les ARN C1 WT, C1 A mut et C1 D D1-A1 ont été incubés 2h00 en présence ou absence d'extrait nucléaire de cellules HeLa, de protéines hnRNP A1 ou ASF/SF2. Les ARN sont ensuite déposés sur gel dénaturant acrylamide 19:1 / urée 5%. Les rapports ARN messager (M) / ARN pré-messager (P) sont mesurés par PhosphorImager.

Malgré les problèmes de dimérisation, nous avons incubé les ARN transcrits radioactivement et cappés dans un extrait nucléaire en conditions d'épissage, puis après éliminations des protéines et dépôt sur gel natif, nous avons observé l'état de dimérisation des ARN (Figure R-57). Il est surprenant d'observer des doubles bandes qui ne migrent pas à la même taille dans les témoins dimère (bandes 1 et 2) et monomère (bandes 3 et 4), n'ayant pas subi de processus d'épissage. Une bande supérieure (bande 5) apparaît dès 15 minutes d'épissage mais sa nature n'est pas identifiée.



**Figure R-57: Dimérisation des ARN C1 WT, C1 A mut et C1 D D1-A1 après épissage.**

Les ARN marqués uniformément à l' $[\alpha\text{UTP}^*]$  et cappés ont été incubés 2h00 à 30°C dans un tampon permettant l'épissage en présence d'extrait nucléaire de cellules HeLa. Des témoins ARN non épissés incubés dans des tampons monomère (concentration finale : 50 mM Cacodylate de sodium pH 7.5, 40 mM KCl, 0.1 mM  $\text{MgCl}_2$ ) ou dimère (concentration finale : 50 mM Cacodylate de sodium pH 7.5, 300 mM KCl, 5 mM  $\text{MgCl}_2$ ). Les ARN sont récupérés avec différents temps d'épissage: 0, 15, 30 minutes, 2 heures. Un témoin à 2 heures d'incubation avec 10 mM d'ATP inhibant l'épissage est déposé en parallèle. Les protéines sont éliminées par protéinase K puis phénol/chloroforme et les ARN sont précipités. Ils sont ensuite déposés sur gel natif 0,8% agarose TBM (Tris-Borate 45 mM, pH 8,3,  $\text{MgCl}_2$  0,1 mM) migré à 4°C à 100 V constant pendant 4h00.

La bande 6 pourrait correspondre à la forme épissée pour l'ARN C1 WT, d'autant plus qu'on ne l'observe pas lorsque l'épissage est inhibé avec 10 mM d'ATP. De même, avec l'ARN C1 A mut, on peut observer que la forme non épissée disparaît au profit de la forme épissée que l'on ne retrouve pas avec 10 mM d'ATP. Il apparaît par contre difficile d'analyser l'identité des bandes avec l'ARN C1  $\Delta$  D1-A1. Les problèmes de dimérisation n'ont pas pu être résolus en changeant le mode de purification des ARN. Nos conditions de dimérisation et nos tampons permettent pourtant la dimérisation presque totale de l'ARN C1  $\Delta$  D1-A1. Ce profil observé paraît difficile à analyser. Si les obstacles rencontrés lors de la dimérisation sont la cause de ce profil complexe, la résolution de ces problèmes devrait faciliter l'interprétation des résultats. Les quantités d'ARN utilisées en conservant les ratios ARN/extrait nucléaire restent donc tout à fait raisonnables, les formes épissées et non épissées sont visibles sur gel d'agarose. S'il existe un changement de dimérisation lors de l'épissage, les quantités utilisées semblent adaptées pour l'observer. Une trainée noire persiste sur le gel d'agarose dans les pistes ayant subi l'épissage. Une seconde purification au phénol/chloroforme serait peut-être nécessaire pour l'éviter.

La délétion de l'intron devra être aussi faite pour les ARN C2 et C3 pour observer la dimérisation de la forme épissée.

Ces expériences n'ont pas permis d'avancer dans la recherche d'un lien entre dimérisation et épissage. Il est cependant intéressant de noter qu'une mutation dans le site de dimérisation affecte l'efficacité d'épissage ainsi que sa régulation par les protéines hnRNP A1 et ASF/SF2.



**CONCLUSION GENERALE**

**ET**

**PERSPECTIVES**



### CONCLUSION GENERALE ET PERSPECTIVES

L'épissage est une étape clé du cycle viral, la compréhension des mécanismes impliqués dans sa régulation devrait permettre de mieux connaître son fonctionnement. L'étape d'épissage représente de plus une cible intéressante pour le développement de nouvelles thérapies antivirales.

**Grâce à l'approche permettant d'identifier l'ensemble des partenaires protéiques potentiels de la région de VIH-1 contenant le site d'épissage A7 et ses éléments de régulation, nous avons démontré l'interaction directe entre l'ARN viral et 4 des protéines identifiées : nucléoline, hnRNP A1/B, hnRNP H et hnRNP K.**

Lors de nos analyses protéomiques, nous n'avons pas pu détecter certaines protéines dont Sam68. Cette dernière pourrait interagir avec hnRNP K et hnRNP H liés à l'ARN SLS2 pour recruter la machinerie nécessaire pour le clivage et la polyadénylation de l'ARN VIH-1.

La protéine hnRNP K peut interagir directement avec la protéine Sam68 (Yang *et al.*, 2002). Or la protéine Sam68 se lie à une séquence ARN consensus « UAAA ». Cette séquence fait partie de la séquence consensus AAUAAA du signal eucaryotique de polyadénylation (Lin *et al.*, 1997; Itoh *et al.*, 2002). L'interaction entre Sam68, hnRNP K et le cofacteur transcriptionnel CBP (Yang *et al.*, 2002) suggère donc qu'au sein de ce complexe Sam68 pourrait interagir directement avec le signal de polyadénylation pour recruter la machinerie nécessaire pour le clivage de l'ARN et la polyadénylation (McLaren *et al.*, 2004). La protéine hnRNP H a aussi été décrite comme activatrice de la polyadénylation (Arhin *et al.*, 2002 ; Fogel *et al.*, 2002; Wilusz *et al.*, 2006). Par conséquent, il est envisageable qu'un complexe formé par Sam68-hnRNP H-hnRNP K puisse jouer un rôle dans la maturation de l'extrémité 3' du VIH-1. Il a été également noté que la protéine Sam68 complète fonctionnellement et travaille en synergie avec la protéine virale Rev lors de l'expression de gènes et la production virale liées au RRE. La liaison de la protéine Rev à la protéine hnRNP K suggère que cette interaction pourrait être liée à de nouveaux mécanismes biologiques, rendant Rev capable d'influencer la réplication et la pathogénicité du VIH-1 par des mécanismes qui s'étendent au-delà de son rôle dans la régulation de VIH-1 (Hadian *et al.*, 2009).

Nous envisageons d'étendre ces études de protéomiques sur des mutants de la tige-boucle SLS2-A7, où des éléments régulateurs importants comme l'ESE seraient mutés ou délétés. L'objectif est de comparer les protéines fixées sur les différents mutants dans le but d'identifier de nouvelles protéines se fixant sur les éléments régulateurs de l'ARN.

## CONCLUSION GENERALE ET PERSPECTIVES

---

Cependant, de telles mutations peuvent affecter la structuration de l'ARN induisant un changement dans le profil des protéines fixées. En effet, la structure secondaire est importante pour la fixation de nombreuses protéines. D'autres méthodes pourraient permettre de déterminer à la fois les protéines se fixant à l'ARN et leurs sites précis de fixation. Une méthode d'immunosélection de protéines liant l'ARN appelée RIP (RNA binding protein ImmunoPrecipitation), permet d'identifier les composants de complexes ribonucléoprotéiques. Combinée avec l'amplification RT-PCR avec des amorces spécifiques, cette méthode est largement utilisée pour tester la fixation de protéines à l'ARN et identifier les ARNm fixées aux protéines d'intérêt (Gilbert *et al.*, 2006). L'équipe de R. Lürhmann a mis au point une autre méthode qui permet de déterminer le site spécifique de fixation des protéines sur l'ARN (Rhode *et al.*, 2003). Cette méthode combine (1) la préparation du pré-ARNm modifié à un site spécifique contenant la biotine (2) le pontage par photo-affinité (3) la purification des pontages ARN-protéine par purification d'affinité biotine-streptavidine et (4) leur analyse par spectrométrie de masse. Cette approche consiste à introduire le nucléotide biotine-dT à un site défini dans un oligonucléotide synthétique. Ainsi, le pontage ARN-protéine peut être isolé via la purification grâce à la biotine et purifié à haut rendement pour l'analyse par spectrométrie de masse (Rhode *et al.*, 2003). Mais ces techniques sont encore pour l'instant très lourdes à mettre en place.

**HnRNP K possède plusieurs sites de fixation dans la région du site A7 où les protéines hnRNP A1 et hnRNP K se lient de façon coopérative. De plus, nous avons montré un effet inhibiteur de hnRNP K sur l'épissage *in vitro* D4-A7 dans un extrait nucléaire de cellules HeLa.**

Les études effectuées au niveau du site A7 peuvent être complétées en observant la fixation des protéines hnRNP K et ASF/SF2 lorsqu'elles sont simultanément mises en contact avec l'ARN SLS2-A7. ASF/SF2 est un activateur fort de l'épissage au site A7 (Tange *et al.*, 2001), nous imaginons qu'il pourrait rentrer en compétition pour la fixation à ce site avec hnRNP K comme il le fait avec hnRNP A1. De plus, il est intéressant de suivre l'effet de la surexpression et la diminution des taux cellulaires de hnRNP K sur l'épissage *in cellulo* au site A7. Nous pouvons analyser le taux des messagers épissés et non épissés au niveau des sites D4-A7 par RT-PCR semi-quantitative voire quantitative.

D'autre part, il est apparu que la fixation de la protéine de hnRNP K et de la protéine hnRNP A1 était coopérative. Cette interrelation est peut-être la clé de l'inhibition de l'épissage au site

## CONCLUSION GENERALE ET PERSPECTIVES

---

A7 par la protéine hnRNP K et nous amène à nous demander si les fonctions de la protéine hnRNP A1 dans le mécanisme de VIH-1 peuvent être partagées par hnRNP K. HnRNP A1 présente différents rôles dans des mécanismes de régulation du cycle viral.

L'équipe de A. Mouland a montré que l'infection par le VIH-1 active l'expression de hnRNP A1 et contribue à sa relocalisation vers le cytoplasme. HnRNP A1 et l'ARN génomique viral non épissé (ARNv) restent colocalisés dans le cytoplasme laissant supposer une fonction post-nucléaire durant les étapes tardives de la réplication du VIH-1. La régulation positive et la rétention cytoplasmique de hnRNP A1 par VIH-1 assurerait l'expression de protéines structurales virales dans les cellules infectées par VIH-1 (Monette *et al.*, 2009). L'augmentation de hnRNP A1 inhibe drastiquement la réplication du virion et la prolifération cellulaire de 50%. En accord avec ces données, la diminution de hnRNP A1 augmente la production des virions (Jablonski *et al.*, 2009). Si hnRNP K n'a pas montré de rôle dans la relocalisation de l'ARN viral, la diminution de son taux cellulaire par siRNA diminue la réplication du VIH-1 (Hadian *et al.*, 2009).

De plus, hnRNP A1 agit comme facteur *trans*-activateur d'une séquence IRES (internal ribosomal entry site) régulant positivement l'initiation de la traduction dépendant de l'IRES de l'ARNv de VIH-1 (Monette *et al.*, 2009). Le rôle de la protéine hnRNP K dans la traduction a été décrit en particulier dans le cas du papillomavirus HPV-16 (Collier *et al.*, 1998). Il est donc envisageable que la protéine hnRNP K puisse avoir un rôle dans la traduction de VIH-1.

Par ailleurs, hnRNP A1 diminue la transcription virale en diminuant l'épissage des ARNm spécifiques de Tat, et ainsi la synthèse du *trans*-activateur viral. Il a été montré également que les protéines hnRNP affectent la transcription en interagissant fonctionnellement avec des portions distinctes du promoteur LTR du génome viral (Jablonski *et al.*, 2009). Par exemple, la déplétion de hnRNP K par siRNA réduit la transcription de VIH-1 (Denisenko *et al.*, 1996; Denisenko *et al.*, 1997).

En conclusion, vu les nombreux rôles de la protéine hnRNP A1 dans le cycle viral et l'interrelation hnRNP K-hnRNP A1, on est en droit de se demander si hnRNP K intervient dans l'un de ces mécanismes.

## CONCLUSION GENERALE ET PERSPECTIVES

---

**Nous avons montré que la protéine hnRNP K se lie à l'ESSv et active l'épissage au site A2, en synergie avec ASF/SF2.**

L'élément ESSv réprime spécifiquement l'épissage au site A2 et est dépendant des protéines hnRNP A/B. Il a été muté pour évaluer son importance lors de la réplication virale. Les résultats de cette expérience indiquent que l'ESSv est requis pour maintenir l'équilibre nécessaire entre les ARNm épissés et non épissés pour une réplication efficace (Madsen *et al.*, 2005). La protéine hnRNP K a donc potentiellement un rôle à jouer dans ce mécanisme de régulation de la réplication.

Il a déjà été montré un rôle de la protéine hnRNP K dans d'autres virus. En effet, la vimentine interagit avec les protéines hnRNP C1 /C2 et K et la protéine NS1 (NonStructural protein 1) de la dengue et est importante pour la réplication virale et le relargage du virus (Kanlaya *et al.*). De plus, hnRNP K est impliquée dans la réplication du virus de l'hépatite B (Ng *et al.*, 2005 ; Zhang *et al.*, 2008) et la transcription de VIH-1 (Wolf *et al.*, 2008). Quant aux protéines APOBEC3 (Apolipoprotein B mRNA Editing Catalytic subunit-3), elles ont montré une large activité antivirale contre les rétrovirus, dont le VIH-1 (Sheehy *et al.*, 2002 ; Bieniasz, 2004 ; Goff, 2004 ; Harris *et al.*, 2004 ; Navarro *et al.*, 2004 ; Rose *et al.*, 2004 ; Turelli *et al.*, 2004). L'équipe de Xu a montré que APOBEC3B agit comme un facteur antiviral inné puissant qui supprime la réplication du virus de l'hépatite B HBV dans les hépatocytes en influençant la transcription régulée par hnRNP K et l'expression des gènes du virus HBV (Zhang *et al.*, 2008).

Des données récentes montrent un rôle de la protéine hnRNP K dans l'infection par le virus de l'herpès HSV où elle est nécessaire pour le transport efficace de la particule virale hors du cytoplasme de l'hôte (Schmidt *et al.*). Des études futures auront pour but d'analyser si hnRNP K participe au processus de phosphorylation qui contribue au transport du virion. hnRNP K interagit aussi avec plusieurs familles de virus comme l'hépatite C ou le virus de la dengue (Hsieh *et al.*, 1998; Chang *et al.*, 2001 ), le virus de la fièvre africaine du cochon (Hernaiz *et al.*, 2008) et l'entérovirus 71 (Lin *et al.*, 2008).

A l'analyse de ces données provenant de différents groupes de virus, la protéine hnRNP K représente une cible potentielle pour une nouvelle stratégie antivirale applicable à une large variété de virus. Le laboratoire AREMS a d'ailleurs déjà caractérisé, en collaboration avec l'équipe de J Tazi de Montpellier, une molécule chimique capable de bloquer l'épissage de

## CONCLUSION GENERALE ET PERSPECTIVES

---

l'ARN du VIH-1. Cette molécule inhibe la multiplication virale *in cellulo* (Bakkour *et al.*, 2007).

### **Selon le contexte, la protéine hnRNP K peut être activatrice ou inhibitrice de l'épissage du VIH-1.**

Il a été déjà montré que la protéine hnRNP K est capable d'inhiber comme activer l'épissage. Par exemple, lors de la production d'isoformes d'épissage de gènes apoptotiques, hnRNP K provoque un ensemble d'événements de saut et d'inclusion d'exons, établissant alors hnRNP K comme l'effecteur d'épissage le plus général dans ce groupe de protéines hnRNP (Venables *et al.*, 2008). HnRNP K est un composant du complexe activateur qui stimule l'épissage de l'exon 6A de la  $\beta$ -tropomyosine (Expert-Bezançon *et al.*, 2002). hnRNP K pourrait également être impliquée dans la répression de l'exon 6B, qui est aussi réprimé par les protéines PTB et hnRNP A1 et activé par les protéines ASF/SF2 et SC35 (Sauliere *et al.*, 2006). Enfin, la protéine Nova peut être en compétition ou en synergie avec d'autres facteurs comme hnRNP K ou brPTB/nPTB sur un cas de contrôle de l'épissage alternatif *in vivo* (Ule *et al.*, 2006).

La protéine hnRNP K est aussi activatrice du site A3 lors d'expérience d'épissage *in vitro*. Il a été montré au laboratoire que ce site peut être activé par les protéines SR, ASF/SF2 et 9G8. Il serait intéressant d'étendre l'étude de hnRNP K à ce site pour rechercher si l'effet synergique observé au site A2 avec ASF/SF2 pourrait se retrouver au site A3 ou se produire aussi avec la protéine 9G8. Nous étudierons l'effet sur la fixation des protéines suite à l'ajout simultané des protéines hnRNP K et ASF/SF2, hnRNP K et 9G8 mais aussi de hnRNP K avec hnRNP A1.

### **La surexpression de la protéine hnRNP K modifie le profil d'épissage de l'ARN $\Delta$ PSP.**

Il serait aussi intéressant de regarder l'effet de la surexpression de hnRNP K sur les différentes espèces d'ARNm, cela nous apporterait beaucoup de détails pour comprendre plus finement l'effet sur l'épissage de hnRNP K. Une étude de la variation de l'épissage au niveau de chaque site accepteur permettrait d'établir efficacement un profil de l'action de hnRNP K sur les sites d'épissage.

Une étude très détaillée a été réalisée pour observer l'effet des variations cellulaires de différentes hnRNP et protéines SR sur la transcription et la réplication virale sur VIH-1 (Jablonski *et al.*, 2009). Une expérience de l'équipe de Brack-Werner a montré qu'une diminution de hnRNP K diminuait le taux de p24 intra et extra cellulaire, ce qui signifierait une diminution de l'expression virale. hnRNP K influencerait donc positivement l'expression

## CONCLUSION GENERALE ET PERSPECTIVES

---

virale dans la cellule. Ce résultat va à l'encontre de ce qui a été observé avec plusieurs hnRNP et protéines SR dont hnRNP A1 et ASF/SF2 qui diminuent la réplication virale, ainsi que la transcription, tout en diminuant le taux d'ARNm *tat1* (Jablonski *et al.*, 2009). Les auteurs de cette étude concluent que le fait de varier le taux cellulaire des protéines hnRNP et SR brise leur équilibre, ce qui provoque une diminution des espèces de 2kb et une augmentation des espèces de 4 et 9 kb. Cette conclusion est appuyée par le fait que les taux cellulaires de hnRNP et SC35 sont diminués lors du pic d'infection virale dans les macrophages dérivés de monocytes, là aussi corrélé avec une diminution des espèces de 2 kb et une augmentation des 4 et 9 kb (Dowling *et al.*, 2008).

Les protéines hnRNP et SR sont décrites comme des facteurs de restriction car une modification de leur expression cellulaire décroît la réplication virale et l'infectivité. Ces facteurs pourraient être ainsi impliqués dans différents mécanismes viraux.

### **L'augmentation de la concentration cellulaire de hnRNP K dans les cellules HeLa diminue l'expression de la protéine virale Nef.**

L'omission de hnRNP K dans le complexe NAKC (Nef-associated kinase complex) comprenant les protéines Nef, PKC, Eed, Lck et hnRNP K, réduit notablement l'effet d'activation de la transactivation dépendante de Tat par dé-répression transcriptionnelle. La présence de hnRNP K dans le complexe NAKC est indispensable pour lever complètement la répression et permettre l'activation de la transcription de VIH-1 par Tat. La variation des différentes espèces d'ARNm *tat* observée lors de la surexpression de hnRNP K laisse envisager une diminution de l'expression de la protéine Tat. Dans le cas d'une diminution de la protéine Tat, s'agirait-il d'un système de rétro-contrôle négatif? Cela pourrait être un mécanisme similaire au rétro-contrôle observé en phase tardive de l'infection par VIH-1, quand la protéine Tat inhibe son expression au profit des protéines de l'enveloppe. Une étude importante a montré l'implication de la protéine Tat sur la régulation de l'épissage indépendante de l'activité transcriptionnelle. La surexpression de Tat conduit à une diminution des pré-ARNm non épissés et de l'épissage au site A3 ainsi qu'à l'augmentation de l'épissage aux sites A4a-c et A5 (Jablonski *et al.*, 2010). Il serait intéressant d'étudier si la protéine hnRNP K pourrait intervenir dans ce mécanisme.

Il serait également intéressant de vérifier par Western blot l'expression des protéines Tat, Rev et même temps Vpr dont l'expression doit être modifiée lors de variation de l'épissage au site A2.

# **MATERIEL ET METHODES**



## MATERIELS ET METHODES

### A. MATERIEL

#### CHAPITRE I Souches utilisées

##### I.1. Souches d'*E. coli* utilisées

**Souche DH5 $\alpha$**  : *deoR*, *supE44*,  $\Delta(lacZYA-argF)U169$ ,  $\Phi80lacZ\Delta M15$ , *hsdR17* ( $r_K^- m_K^+$ ), *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*, F<sup>-</sup> (Hanahan, 1983) a été utilisé pour la production et la purification des plasmides.

**Souche BL21-CodonPlus-RIL** : B F<sup>-</sup>, *ompT*, *hsdS* ( $r_B^- m_B^-$ ), *dcm+*, Tet<sup>r</sup>, *gal*, *endA*, Hte [*argU ileY leuW Cam<sup>r</sup>*] (Studier *et al.*, 1986) a été utilisé pour la production des protéines recombinantes.

##### I.2 Les cellules d'organismes eucaryotes

**Cellules HeLa** : N°ATCC® : CCL-2™ (Scherer *et al.*, 1953); lignée cellulaire cancéreuse provenant d'un prélèvement effectué sur une patiente atteinte d'un cancer du col de l'utérus, et décédée en 1951, **Henrietta Lacks**.

##### I.3 Milieux de culture, tampons et solution d'usage courant

**Milieu LB (Luria Bertani)** : peptone pancréatique animale 5 g/l ; peptone tryptique de caséine 5 g/l ; extrait de levure 5 g/l ; NaCl 10 g/l.

Le pH est ajusté à 7,5 avec NaOH 10 N. Le milieu de culture est ensuite autoclavé 20 minutes à 121°C sous une pression de 1 bar.

**Milieu solide LB** : l'agar est ajouté à raison de 17 g par litre de milieu de culture LB avant stérilisation à l'autoclave.

**Milieu LBA** : l'ampicilline (100 mg/ml) est ajoutée après autoclave dans le milieu de culture LB à une concentration finale de 100  $\mu$ g/ml.

**Milieu LBK** : la kanamycine (100 mg/ml) est ajoutée après autoclave dans le milieu de culture LB à une concentration finale de 100  $\mu$ g/ml.

**TE** : Tris-HCl 10 mM, pH 7,4 ; EDTA 1 mM.

**TBE** : Tris borate 89 mM, pH 8,3 ; EDTA 2 mM.

**PBS 10X** : NaCl 1,37 M ; KCl 27 mM ; KH<sub>2</sub>PO<sub>4</sub> 11,5 mM ; Na<sub>2</sub>HPO<sub>4</sub> 64 mM.

**Tampon de transfert western blot 10X** : Tris 500 mM ; glycine 145 mM.

**Solution de transfert** : Blotting buffer 1X ; éthanol 20% (v/v).

**Solution de polyacrylamide dénaturante :** Acrylamide/Bisacrylamide (23,75/1,25) 25% (p/v) ; TBE 1X ; urée 8 M.

**Solution de polyacrylamide native :** Acrylamide/Bisacrylamide (38/2 et 59/1) 40% (p/v) ; TBE 1X.

**Solution de bleu de dépôt agarose (6X) :** EDTA 125 mM ; SDS 2% (v/v) ; glycérol 25% (v/v) ; bleu de bromophénol 0,12% (p/v).

**Solution de bleu de formamide :** EDTA 20 mM ; bleu de bromophénol 0,05 % (p/v) ; bleu de xylène cyanol 0,05% (p/v) dans de la formamide désionisée.

**Solution de bleu de dépôt (non dénaturant) pour les expériences de retard sur gel (bleu CSB) :** HEPES-KOH 20 mM, pH 7,9 ; glycérol 40% (v/v) ; bleu de bromophénol 0,05% (p/v); bleu de xylène cyanol 0,03% (p/v).

**Tampon d'élution :** Tris-HCl 10 mM, pH 7,5 ; NaCl 300 mM ; EDTA 1 mM ; SDS1%.

**Tampon D :** MgCl<sub>2</sub> 1,5 mM ; EDTA 0,2 mM ; KCl 150 mM ; glycérol 10% (p/v) ; HEPESKOH 20 mM pH 7,9.

**Tampon Protéinase K 1 X :** EDTA 25 mM pH 7,4 ; NaCl 0,3 M ; SDS 2% ; Tris-HCl 20 mM pH 7,5.

## CHAPITRE II Constructions génétiques

### II.1. Constructions renfermant des séquences codant des protéines pour l'expression en *E. Coli*.

- Protéines hnRNP K et hnRNP U entières : Les plasmides pOTB7-hnRNP K et pCMVSPORT6-hnRNP U proviennent du RZPD (Deutsches Ressourcenzentrum für Genomforschung GmbH). Des amplifications par PCR sur ces plasmides ont permis de rajouter des sites de restrictions autour des phases codantes de ces protéines pour sous-cloner dans un plasmide d'expression pET28b. Ainsi, les sites NheI et SalI ont été ajoutés pour hnRNP K et NheI et NotI pour hnRNP U. Après ligation des produits PCR dans le vecteur PCR 2.1 grâce au kit TOPO TA Cloning Kit (avec le vecteur pCR2.1-TOPO) de Invitrogen, les plasmides pCR2.1 ont été digérés par les enzymes adéquates puis sous-clonées dans le plasmide d'expression pET28b. Ce plasmide va permettre de produire ces protéines avec une étiquette histidine. Les constructions et l'absence de mutation ont été vérifiées par séquençage.
- Les fragments de nucléoline P30 contiennent les trois premiers RRM de la protéine nucléoline et P40 les quatre premiers RRM. Le plasmide pET15b-nucléoline RRM1, 2, 3 permettant la production de la protéine P30 a été fourni par P. Bouvet (Toulouse). P. Bouvet nous a fourni de plus la protéine recombinante P40 purifiée.
- hnRNP A1 : Nous disposons du plasmide pET11a-hnRNP A1, fourni par le Pr G. Dreyfuss renfermant la séquence codant la protéine hnRNP A1 entière sans étiquette.

### II.2. Constructions pour l'épissage utilisées au cours de ce travail.

Nos constructions dérivent du plasmide pBRU3, qui nous a été fourni par F. Clavel (Institut Pasteur). Il contient la séquence entière du génome du virus VIH-1 de la souche VIH-1/BRU, des nucléotides +1 à +9229, la position +1 correspondant au premier nucléotide de la séquence R située à l'extrémité 5' du génome (Charneau *et al.*, 1992). Ensuite par PCR, L. Damier, au laboratoire, a construit des plasmides pBluescript KSII+ recombinants renfermant des séquences non contiguës du génome HIV-1 BRU. Ces constructions permettent de générer, par transcription *in vitro*, des ARN contenant des couples de sites d'épissage, dont la taille est compatible avec des expériences d'épissage *in vitro*. Par mutagenèse dirigée par PCR, V. Marchand a généré de nouveaux plasmides pBluescript KSII+ recombinants renfermant des séquences non contiguës du génome du virus HIV-1 BRU. Les différentes constructions utilisées au cours de ce travail sont schématisées dans la figure R-9 pour les

mutants de la tige-boucle 2 contenant le site A7 et utilisés pour les retards sur gel, les pontages aux UV et les empreintes. Les différentes constructions dérivées du plasmide pL47 WT sont représentées au cours des résultats près des gel d'épissage correspondant.

### **II.3 Construction p $\Delta$ PSP pour l'expression de VIH-1 *in cellulo*.**

La construction p $\Delta$ PSP, fournie par CM. Stoltzfus utilisée en transfection, a été réalisée à partir du plasmide infectieux pNL4-3 (GenBank<sup>TM</sup> accession number M19921). La délétion réalisée entre les positions 1511 et 4551 et qui rend p $\Delta$ PSP non-infectieux a été obtenue par digestion de pNL4-3 par les nucléases *SpeI* et *BalI*, puis religation avec les oligonucléotides 5'-CTAGACGCGTTTGG-3' et 5'-CCAAACGCGT-3'.

### **II.4 Oligonucléotides utilisés**

Nom de l'oligonucléotide	Sequence (5'-3')	Positions	Utilisation
<b>Amplification pour la préparation de l'ARN</b>			
O-3858	TTCAGACCCACCTCGAGGGGACCCG	Bru 7967-7999	Mutagenèse pUC18-SLS2 ΔK
O-3859	CGGGTCCCCTCGAGGTGGGTCTGAA	Bru 7967-7999	Mutagenèse pUC18-SLS2 ΔK
O-4705	GGTTGGGAGGTGGGTCTG	Bru 7969-7986	Mutagenèse pUC18-SLS2 ΔNCL
O-4706	AATAGAAGAAGAAGGTGG	Bru 8013-8030	Mutagenèse pUC18-SLS2 ΔNCL
O-4707	AGCTTAATACGACTCACTATAGGG		pUC18-T7-MS2
O-4708	TCGACCCATAGTGAGTCGTATTA		pUC18-T7-MS2
O-3104	GCTCTAGAGTCTCTGTCTCTCTC		Amplification de l'ARN SLS2-A7
<b>Epissage <i>in cellulo</i></b>			
O-2143	GGCTTGCTGAAGCGCGCACGGCAAGAGG	pNL4-3 nts 700-727	Epissage <i>in cellulo</i>
O-2144	TTGGGAGGTGGGTTGCTTTGATAGAG	pNL4-3 nts 8381-8369 and 6041-6029	Epissage <i>in cellulo</i>
O-5751	CCGCAGATCGTCCAGATAAG		Odp32 RT-PCR oligo 3' dans l'exon 7 du ΔPSP
O-5752	ACAGGCCCGAAGGAATAGAA		RT-PCR oligo 5' dans l'exon 7 du ΔPSP en aval du site A7
<b>Séquençage du plasmide pUC18</b>			
O-67	CAGCACTGACCCTTTTG		Oligo universel, appelé primer -40, séquençage à l'opposé du 122
O-122	CGGGTCCCCTCGAGGTGGGTCTGAA		Oligo universel, hybride entre le primer -48 et -21, séquençage à l'opposé du 67
<b>Clonage d'ADNc dans des vecteurs de surexpression de protéines.</b>			
O-4693	GCTAGCATGAGTTCCTCGCCTG		Oligo sens pour cloner cDNA hnRNP U dans pET28b (site Nhe)
O-4694	GCGGCCGCTCAATAATATCCTTGGTG		Oligo sens pour cloner cDNA hnRNP U dans pET28b (site Not I)
O-4878	GGTTTGTCTTGATACTTA		Oligo pour séquencer hnRNPU
O-4695	<b>GCTAGC</b> ATGGAAACTGAACAGCC		hnRNP K amplification cDNA
O-4696	<b>GTCGAC</b> TTAGAATCCTCAACAT		hnRNP K amplification cDNA

O-3100	CCCAAGCTTGCATTTAGGTGACACTATAGAAGAGTTTCAGACCCACCTC		Promoteur SP6 pour clonage pUC18
<b>Structure 2D de la tige-boucle SLS1, 2, 3</b>			
O-2156	GGTGGGTCTGAAACG	Bru 7978-7965	ARN SLS2-A7
O-2065	CGGGATCCTCTAATACGACTCACTATAGGGAGAATAGTTTTGCTGTAC	Bru 7900-7921	ARN SLS1,2,3-A7
O-2382	CGTCCCAGATAAGTGC	Bru 8077-8092	ARN SLS1,2,3-A7
<b>Etudes de dimérisation</b>			
O-6814	GGACAGCAGAGATCCAGTTT		Délétion intron D1-A1 dans construction C1
O66815	CAGTCGCCGCCCTCGCCTCTT		Délétion intron D1-A1 dans construction C1

Tableau 7: Oligonucléotides utilisés dans cette étude, et leur position d'hybridation sur VIH-1 Bru ou pNL4.3.

## B. METHODES

### CHAPITRE I : Production et purification des protéines recombinantes dans *E. coli*.

#### I.1 Protéine hnRNP A1

La protéine hnRNP A1 est produite dans la souche *E. coli* BL21 DE3 transformée par le plasmide pET11a-hnRNP A1. Le gène de l'ARN polymérase T7 qui est intégré dans le chromosome de la bactérie est inductible par l'IPTG.

Un litre de milieu LB supplémenté en ampicilline (200 µg/ml) estensemencé à 0,15 unité de DO<sub>600</sub>/ml à partir d'une préculture d'une nuit issue d'une colonie, dans le même milieu. Cette culture est agitée à 37°C jusqu'à ce que l'absorbance à 600 nm atteigne 0,8-1 unité de DO<sub>600</sub>/ml, puis la production est induite par l'ajout d'IPTG à une concentration finale de 1 mM.

L'agitation se poursuit à 20°C pendant 20 h. Les cellules sont alors récoltées par centrifugation à 4000 g pendant 15 min, lavées avec du tampon PBS 1x et stockées à - 20°C.

##### I.1.1 Sonication

Les culots cellulaires sont repris dans 5 volumes/poids de tampon de lyse (20 mM Tris pH 7,6, 250 mM NaCl, 0,5 mM DTT) en présence d'un anti-protéase (Roche diagnostic). Les cellules sont ensuite traitées aux ultrasons sous une puissance de 40 W pendant 4 min, par période de 30 secondes. Après sonication, le lysat bactérien contenant les protéines, les acides nucléiques et les débris cellulaires est centrifugé pendant 20 min à 4°C à 30 000g. Seule la partie soluble de la protéine, située dans le surnageant, va être utilisée par la suite.

##### I.1.2 Préparation de l'échantillon pour la purification

Le surnageant de sonication est filtré avec un filtre 0,45 µm puis dilué pour être à 125 mM NaCl avec du tampon (1mM EDTA; 20 mM Tris pH 7,6). Cet échantillon est déposé sur colonne S-sépharose à 4°C.

##### I.1.3 Etapes de purification

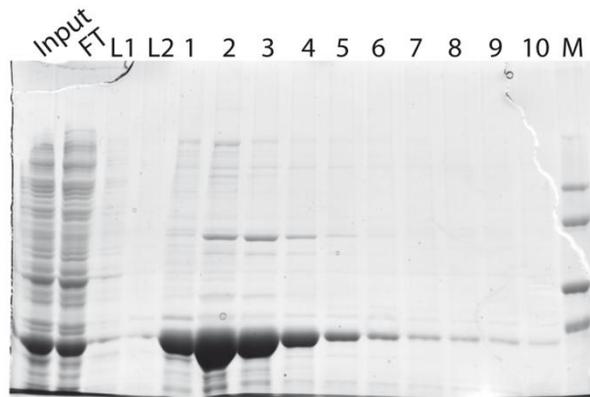
###### I.1.3.1 Colonne S-sépharose

Cette colonne (Amersham bioscience) est une échangeuse de cations. Elle est chargée négativement et fixe des protéines chargées positivement. La protéine hnRNP A1 que nous produisons dans *E. coli* n'a pas d'étiquette facilitant la purification. Nous allons donc utiliser

ses propriétés physico-chimiques. Elle a un pI théorique de 9.26, ce qui devrait lui conférer une bonne fixation sur la colonne S-sépharose. Après fixation des protéines, la colonne S-sépharose est lavée avec un tampon (1mM EDTA; 20 mM Tris pH 7,6 ; 100 mM NaCl) puis éluée avec du tampon (1mM EDTA; 20 mM Tris pH 7,6 ; NaCl 350 mM).

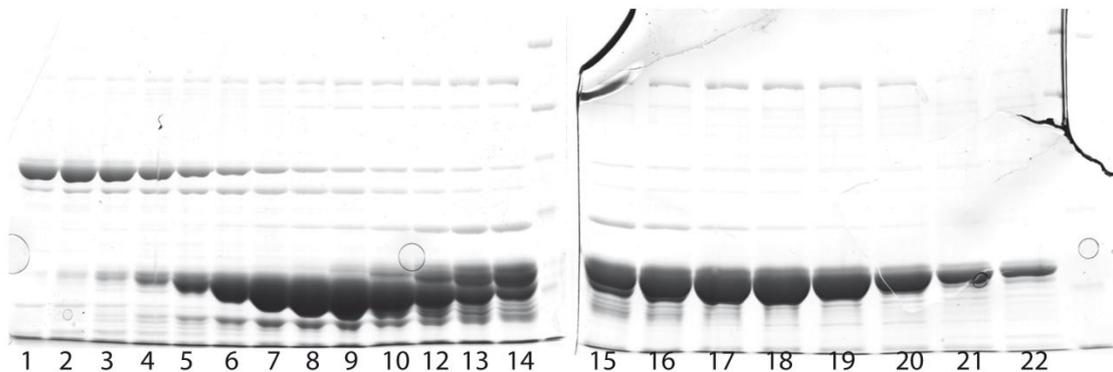
### **I.1.3.2 Colonne Héparine-sépharose**

Les fractions récupérées contenant la protéine hnRNP A1 sont déposées sur colonne héparine après dilution de l'échantillon qui est à 350 mM à la sortie de la colonne S-sépharose avec du tampon sans sels (1mM EDTA; 20 mM Tris pH 7,6) pour obtenir une concentration d'environ 115 mM. Cette dilution est injectée sur la colonne héparine. Après lavage de la colonne héparine avec un tampon (1mM EDTA; 20 mM Tris pH 7.6, NaCl 120 mM), un gradient croissant de sels va permettre d'éluer les protéines. Les protéines hnRNP A1 sont éluées de la colonne vers 350 à 400 mM de sels. On observe deux populations de protéines qui ont des affinités différentes pour l'héparine. Ce profil présentant différentes populations de protéine hnRNP A1 a déjà été décrit par Cobianchi (Cobianchi *et al.*, 1988). Son étude a montré que l'existence d'espèces protéiques plus petites est due à la troncation de l'extrémité COOH-terminale seulement. Seules les protéines de taille supérieure, ayant le plus d'affinité, sont conservées. La dernière étape de purification est un passage des protéines éluées sur une colonne Mono-S de l'AKTA Explorer HPLC.



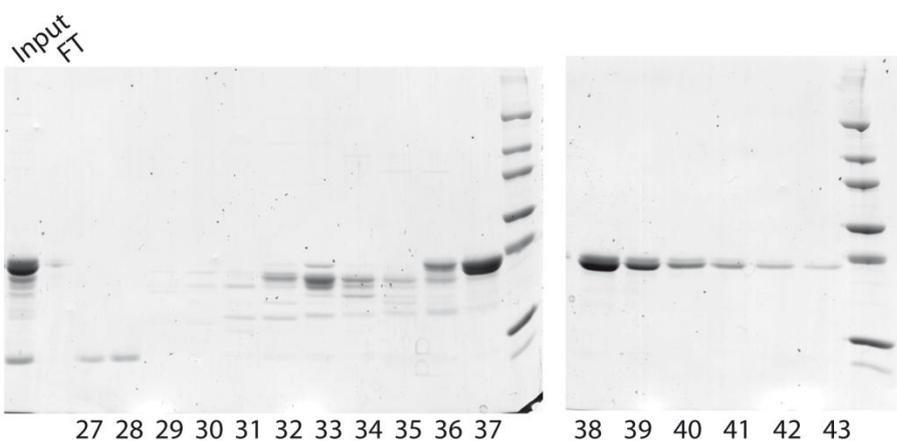
**Figure M-1 : Purification de la protéine hnRNP A1 sur colonne S-sépharose.**

La piste « Input » représente la fraction déposée sur la colonne. La piste « FT » est composée des protéines qui ne sont pas fixées sur la colonne. Les pistes L1 et L2 sont les dépôts des 2 étapes de lavage de la colonne. Les pistes de 1 à 10 sont les fractions d'élution.



**Figure M-2 : Purification de la protéine hnRNP A1 sur colonne héparine-sépharose.**

Les pistes de 1 à 4 de l'élution de la colonne S-sépharose ont été regroupées puis diluées dans du tampon sans sels (1mM EDTA, 20 mM Tris pH7.6) pour baisser la concentration saline de 350 mM NaCl à environ 115 mM NaCl pour l'injection sur la colonne héparine.



**Figure M-3 : Purification de la protéine hnRNP A1 sur colonne Mono-S.**

Les pistes de 15 à 20 de l'élution de la colonne héparine-sépharose ont été regroupées puis diluées dans du tampon sans sels (1mM EDTA, 20 mM Tris pH 7.5) pour baisser la concentration saline à environ 100 mM NaCl pour l'injection sur la colonne Mono-S.

### **I.1.3.3 Colonne Mono-S.**

Cette colonne est une échangeuse de cations. Le gradient de sel va être mis au point afin de bien séparer les différents pics de protéines et obtenir une protéine hnRNP A1 pure. Le poids moléculaire exact de la protéine purifiée a été vérifié en spectrométrie de masse. La protéine a été dialysée contre du tampon D avec 20 % de glycérol.

## **I.2 Protéine MS2-MBP**

Les bactéries BL21 Rosetta sont transformées avec le plasmide pMS2-MBP, portant une résistance à l'ampicilline, provenant du laboratoire du Dr Robin Reed. La culture se fait dans les mêmes conditions que la protéine hnRNP A1. L'induction se fera par contre 3 heures à 37 °C.

### **I.2.1 Sonication**

Les conditions de sonication puis de préparation sont les mêmes que pour la protéine hnRNP A1 mais se font dans le tampon (Tris-HCl 20 mM pH7.6 ; 200 mM NaCl ; 1mM EDTA ; inhibiteur de protéases ; 0.1 mM DTT).

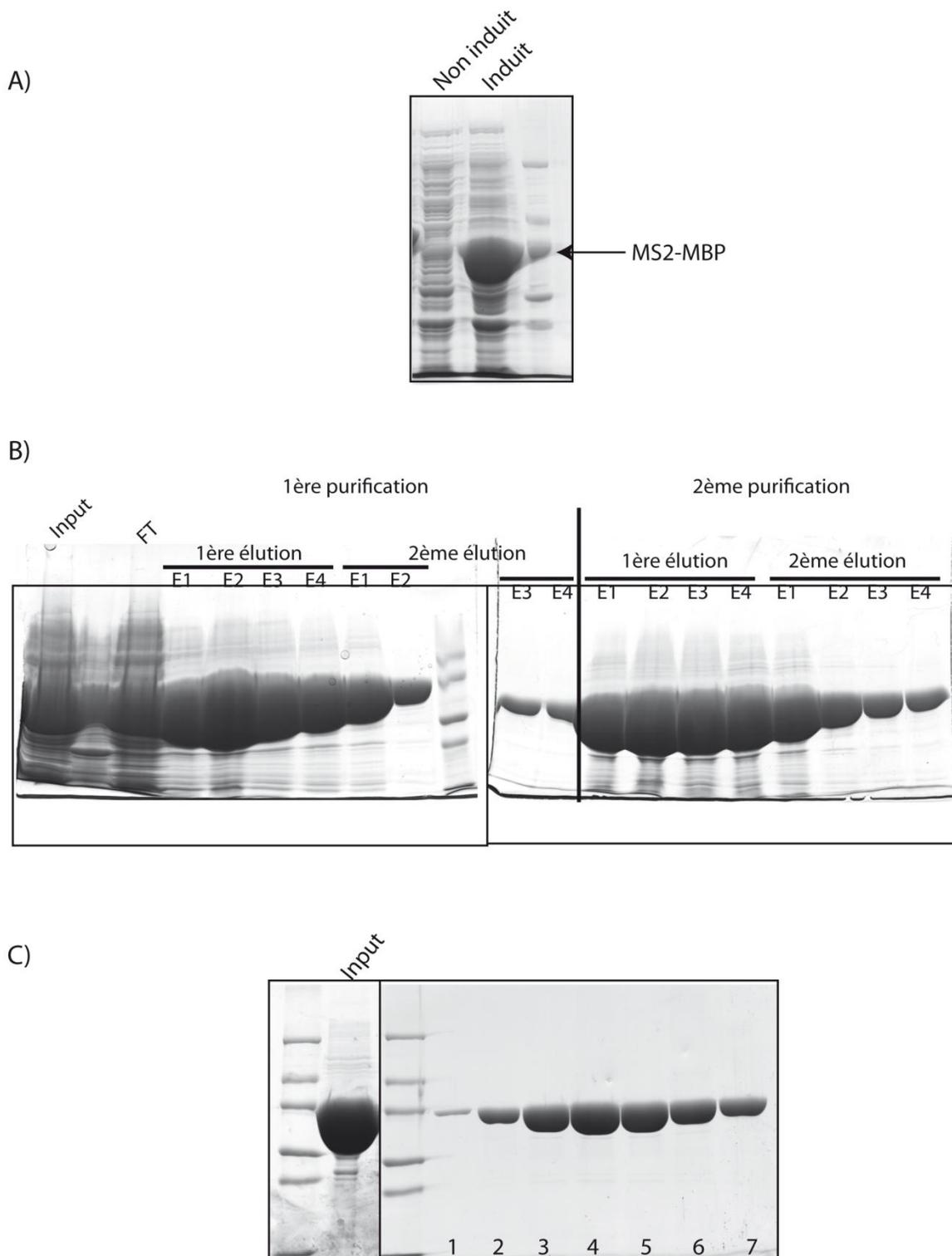
## **I.2.2 Etapes de purification**

### **I.2.2.1 Résine amylose**

La résine amylose (NEB, 1,8 ml) est lavée avec 10 ml de tampon de sonication puis ajoutée au surnageant de bactéries après centrifugation et filtration. Après incubation sur roue une heure à 4 °C, la résine est centrifugée à 1000 rpm. Le surnageant (FT) est conservé pour un second tour de purification sur la même résine après son lavage. La résine est saturée par la grande quantité de protéine lors de la première purification. La résine est lavée deux fois avec 2 ml de tampon A (20 mM Tris-HCl pH7.6 ; 0.1 mM DTT) puis éluée avec 2 ml de tampon maltose (20 mM Tris-HCl pH7.6, 15 mM maltose ; 0.1 mM DTT).

### **I.2.2.2 Colonne héparine-sépharose**

Après injection de l'élué maltose sur la colonne héparine, elle est lavée au tampon (20 mM Tris-HCl pH7.6 ; 0.1 mM DTT) puis éluée avec un gradient linéaire de sels [tampon A (20 mM Tris-HCl pH7.6 ; 0.1 mM DTT)-tampon B (20 mM Tris-HCl pH7.6 ; NaCl 1M ; 0.1 mM DTT)].



**Figure M-4 : Purification de la protéine MS2-MBP.**

A) Induction de la protéine dans les bactéries *E. coli* de souche Rosetta.

B) Première étape de purification sur résine amylose. La piste « Input » présente les protéines totales déposées sur la résine. La piste « FT » montre les protéines qui ne sont pas fixées sur la résine. Les pistes E1 à E4 sont les différentes fractions d'éluion. Deux étapes de purification ont eu lieu, la résine étant saturée au premier passage.

C) Seconde purification sur colonne héparine-sépharose. La piste « Input » présente les protéines totales déposées sur la résine. Les pistes de 1 à 7 sont les fractions d'éluion de la colonne héparine.

### **I.3 Protéine hnRNP K**

Les bactéries BL21 Rosetta sont transformées avec le plasmide pET28b-hnRNP K portant une résistance à la kanamycine. La culture se fait dans les mêmes conditions que la protéine hnRNP A1. L'induction se fera par contre 5 heures à 37 °C.

#### **I.3.1 Sonication**

Les conditions de sonication puis de préparation sont les mêmes que pour la protéine hnRNP A1 mais se font dans le tampon (Tris-HCl 20 mM pH7.6 ; 200 mM NaCl ; 1mM EDTA ; inhibiteur de protéases ; 0.1 mM DTT).

#### **I.3.2 Etapes de purification**

##### **I.3.2.1 Résine Ni-NTA agarose**

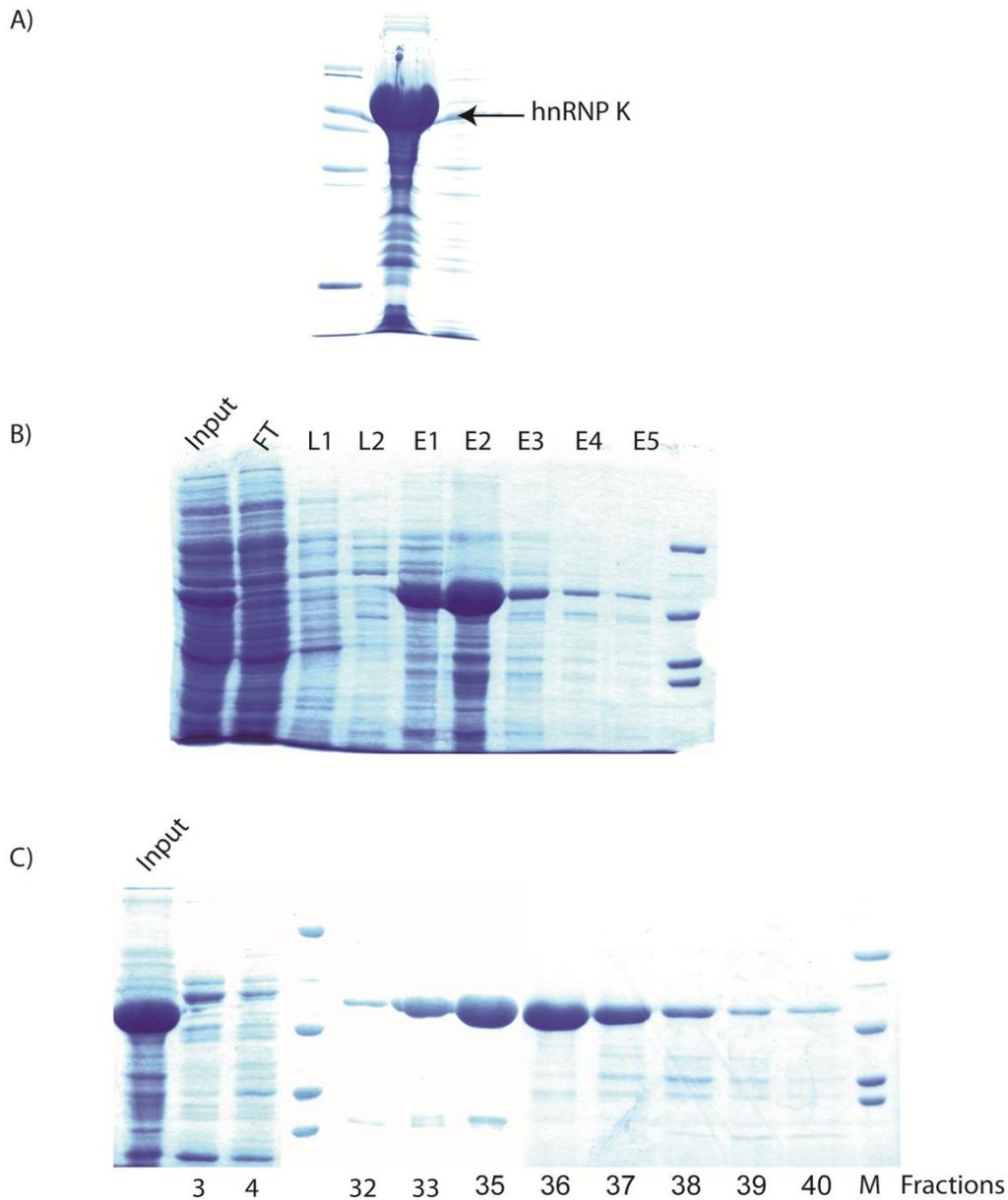
1 ml de résine Ni-NTA agarose va retenir spécifiquement la protéine hnRNP K par son étiquette histidine. La résine est lavée avec 10 ml de tampon de sonication puis ajoutée au surnageant de bactéries après centrifugation et filtration. Après incubation sur roue une heure à 4 °C, la résine est mise en colonne et le surnageant est éliminé par gravité. La résine est lavée trois fois avec 5 ml de tampon A (20 mM Tris-HCl pH7.6 ; 200 mM NaCl ; 0.1 mM DTT) puis éluée avec 2 ml de tampon d'éluion à l'imidazole (20 mM Tris-HCl pH7.6, 200mM imidazole ; antiprotéases ; 0.1 mM DTT).

##### **I.3.2.2 Colonne héparine-sépharose**

La protéine recombinante est purifiée sur colonne héparine, ce qui doit permettre de sélectionner les protéines correctement repliées ayant la capacité de se lier à l'ARN. On peut donc estimer que la majorité de notre protéine recombinante se lie bien à l'ARN, ce qui est vérifié par les expériences de retard sur gel ou l'affinité de la protéine recombinante pour l'ARN cible est très bonne.

Après injection de l'éluion Ni-NTA sur la colonne héparine, elle est lavée au tampon A (20 mM Tris-HCl pH7.6 ; 200 mM NaCl ; 0.1 mM DTT) puis éluée avec un gradient linéaire de sels [tampon A ((20 mM Tris-HCl pH7.6 ; 200 mM NaCl ; 0.1 mM DTT))-tampon B ((20 mM Tris-HCl pH7.6 ; 0.1 mM DTT ; NaCl 1M)]. La protéine est éluée à environ 350 mM de sels.

La protéine va être dialysée contre du tampon D avec 50 % de glycérol et conservée à -80°C.



**Figure M-5 : Purification de la protéine recombinante hnRNP K.**

A) Induction de la protéine hnRNP K dans les bactéries *E. coli* de souche Rosetta. Le milieu de culture LB supplémenté en antibiotique kanamycine (mg/ml) à 0.15 unité de DO600/ml à partir d'une pré-culture d'une nuit issue d'une colonie. Cette culture est agitée à 37 °C jusqu'à une DO600 de 0.8 à 1, puis la production est induite par l'ajout d'IPTG à une concentration finale de 1mM pendant 5 heures à 37°C.

B) Purification de la protéine hnRNP K sur résine Ni-NTA. La protéine hnRNP K recombinante est retenue sur la résine Ni-NTA grâce à son étiquette histidine. La piste « Input » est le sonicat bactérien filtré puis centrifugé pour éliminer les fragments de bactéries et déposé sur la colonne. La colonne est éluée (après lavage de la colonne L1 et L2) par 200 mM d'imidazole (E1 à E4) puis avec 1M d'imidazole (E5).

C) Purification de la protéine hnRNP K sur colonne héparine-sépharose. Les fractions d'éluion E1 et E2 de la résine Ni-NTA ont été regroupées puis injectées sur la colonne héparine. Elles vont être éluées par un gradient croissant de sels.

### **I.4 Protéine hnRNP H**

J'ai utilisé la protéine hnRNP H produite sous forme recombinante chez *E. Coli* avec une étiquette poly-histidine par Houda Hallay. A cause de sa localisation dans les corps d'inclusion, la purification de la hnRNP H entière a dû se faire dans des conditions dénaturantes, suivi par une renaturation sur colonne (Chou *et al.*, 1999).

I. Motorine l'a produite avec une étiquette de six histidines après induction de la protéine hnRNP H par ajout d'IPTG dans les cultures d'*E. coli* BL21 DE3 transformé par le plasmide pET15b-hnRNP H permettant la production de la protéine hnRNP H avec une étiquette de 6 histidines en N-terminal.

### **I.5 Protéine Nucléoline**

La nucléoline a été purifiée par I. Motorine et V. Marchand à partir d'extraits nucléaires de cellules HeLa grâce à 2 étapes successives de chromatographie (Belenguer *et al.*, 1990). Une protéine recombinante contenant les quatre premiers RRM de la nucléoline appelée P40 nous a été fournie par Philippe Bouvet (Laboratoire de Pharmacologie et de Biologie Structurale, Toulouse).

## CHAPITRE II : Techniques d'analyse des interactions ARN/protéines

### II.1. Technique d'empreinte sur l'ARN en présence de protéine purifiée

#### II.1.1. Amplification par PCR des ADN matrices

L'ARN SLS1, 2, 3 (180 nucléotides) est transcrit *in vitro* à partir d'une matrice PCR. Il est synthétisé en utilisant une amorce introduisant un promoteur pour l'ARN polymérase T7 (oligonucléotide 2065), dans les conditions suivantes :

Plasmide C51L 50 ng (50 ng/μl)	1 μl	}
Tampon 10X Pfu	5 μl	
Oligonucléotide 2065 100 ng (100 ng/μl)	1 μl	
Oligonucléotide 2382 100 ng (100 ng/μl)	1 μl	
Mélange des 4 dNTP (1,25 mM chacun)	8 μl	
Pfu ADN polymérase	1 μl	
H <sub>2</sub> O qsp	50 μl	

Après une dénaturation de 2 minutes à 95°C, l'ADN est amplifié pendant 25 cycles d'une minute à 94°C, une minute à 45°C, 1 minute à 72°C. Un cycle de 10 minutes à 72°C finira la réaction d'amplification. Le produit PCR est ensuite extrait au phénol/chloroforme, puis au chloroforme/alcool isoamylique (24/1) et précipité deux fois de suite en présence de 2,5 M d'acétate d'ammonium de façon à éliminer au maximum les nucléotides non incorporés. Le produit PCR, lavé et séché, est repris dans de l'eau stérile.

### II.1.2. Synthèse de transcrits froids courts

Les conditions de réaction sont les suivantes :

Matrice PCR ou plasmide linéarisé 500 ng à 1 µg	2 µl
Tampon de transcription 5x (Tris-HCl pH7.5 200 mM ; MgCl <sub>2</sub> 30 mM ; Spermidine 5 mM)	10 µl
RNasine (31 U/µl)(Amersham biosciences)	0,5 µl
Mélange de rNTP (12,5 mM chacun)	10 µl
ARN polymérase SP6 ou T7 (20 U/µl)(MBI Fermentas)	2 µl
H <sub>2</sub> O qsp 30 µl	

Le mélange est incubé 4h à 37°C, puis la matrice d'ADN est dégradée par incubation pendant 30 minutes à 37°C, en présence de 10 U de DNase I RQ1 (Amersham Biosciences), exempte de RNases. 30 µl de bleu de formamide sont ajoutés et le mélange est déposé sur le gel. L'électrophorèse est réalisée à 17 V/cm dans du tampon TBE 0.5x (acide borique 44.5 mM, EDTA 1 mM, Tris borate 44.5 mM, pH 8,3). Le gel est ensuite observé aux UV et la bande contenant l'ARN est excisée et incubée dans 400 µl de tampon d'élution (NaCl 0,25M; EDTA 0,01M; Tris-HCl 0,02 M pH 7,5) à 4°C sur la nuit.

### II.1.3. Principe de la technique d'empreinte des ARN

La fixation d'une protéine sur l'ARN protège cet ARN de l'action des sondes enzymatiques habituellement utilisées dans les analyses de structure secondaire des ARN. L'ARN est incubé en tampon D (EDTA 0,2 mM; KCl 0,1 M; DTT 0,5 mM; HEPES 20 mM pH 7,6 ; glycérol 20%), seul ou en présence de protéine purifiée, dans des conditions permettant l'épissage. Des hydrolyses enzymatiques sont ensuite réalisées. Les ARN peuvent être marqués radioactivement en 5' au [ $\gamma$ -<sup>32</sup>P] ATP (10 mCi/ml, 3000 Ci/mmol) (Amersham) et visualisés sur gel directement après la digestion enzymatique et une extraction au phénol/chloroforme 24 :1 ou analysés par transcription inverse. On observe au niveau de la zone d'interaction ARN/protéine, un certain nombre de coupures ou modifications présentes au niveau de l'ARN seul, mais absentes lorsque l'ARN est incubé en présence de protéine.

## II.1.4 Conditions expérimentales de l’empreinte d’un ARN en présence de protéine purifiée

### II.1.4.1 Assemblage des complexes ARN/protéines puis analyse après transcription inverse.

L’ARN dont on veut identifier les sites de fixation de facteurs protéiques est incubé dans le mélange suivant :

Transcrit froid 200 ng	}	
ARNt de levure qsp 5 µg d’ARN total		
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%)		3,6 µl
MgCl <sub>2</sub> 62,5 mM		1 µl
H <sub>2</sub> O qsp 18 µl		

Le mélange est ensuite incubé 10 min à 65°C, puis 10 min à température ambiante et enfin, 10 min dans la glace, afin d’obtenir une population d’ARN avec une structuration secondaire homogène. Quatre µl de protéine diluée dans le tampon D sont ensuite ajoutés et le mélange est incubé 10 min dans la glace.

### II.1.4.2 Assemblage des complexes ARN/protéines avec un transcrit ARN marqué radioactivement en 5’ au [ $\gamma$ -<sup>32</sup>P] ATP.

Transcrit ARN marqué en 5’ 100 cps	}	
ARNt de levure 2 µg/µl		
Tampon D 10X		0,5 µl
MgCl <sub>2</sub> 62,5 mM		1 µl
H <sub>2</sub> O qsp 5 µl		

Le mélange est ensuite incubé 10 min à 65°C, puis 10 min à température ambiante et enfin, 10 min dans la glace, afin d’obtenir une population d’ARN avec une structuration secondaire

homogène. Quatre µl de protéine diluée dans le tampon D sont ensuite ajoutés et le mélange est incubé 10 min dans la glace.

### II.1.4.3 Hydrolyses enzymatiques

On fait ensuite agir les sondes enzymatiques pendant 10 min à 30°C. Les quantités d'agent chimique ou enzymatique généralement ajoutées et les conditions d'arrêt des réactions sont les suivantes :

---

Agents	Quantités ajoutées	Arrêt de la réaction
RNase V1 (Kemothex)	5.10 <sup>-4</sup> U EDTA 100 mM,	20 µg ARNt de levure, incubation dans la glace
RNase T1	0,4 à 0,6 U	20 µg ARNt, incubation dans la glace
RNase T2 (Invitrogen)	0,5 U ou 0,6 U	20 µg ARNt, incubation dans la glace

---

Les échantillons sont aussitôt extraits au phénol/chloroforme (1/1) puis au chloroforme/alcool isoamylique (24/1) et précipités à l'éthanol en présence d'acétate de sodium 0,3 M. Les culots sont ensuite lavés, séchés et repris dans 3 à 4 µl d'eau. Les ARN sont ensuite analysés par transcription inverse ou déposés directement sur gel quand l'ARN est radiomarqué.

### II.1.4 Séquençage de l'ARN

#### II.1.4.1 Principe

Un oligonucléotide, marqué en 5' à l'ATP [ $\gamma$ -<sup>32</sup>], est hybridé à une région donnée de l'ARN à séquencer. Il sert d'amorce à la transcriptase inverse qui synthétise, en présence des 4 désoxynucléotides dNTP, un brin d'ADN complémentaire à l'ARN. Des didésoxy-nucléotides ddNTP sont ajoutés au milieu réactionnel en concentration telle que la synthèse du brin d'ADN est arrêtée statistiquement. Un témoin permettant d'identifier les pauses naturelles de la transcriptase inverse est effectué en absence de ddNTP.

#### II.1.4.2 Marquage de l'oligonucléotide en 5'

L'oligonucléotide est marqué en 5' en présence d'ATP [ $\gamma$ -<sup>32</sup>]P par action de la polynucléotide kinase du bactériophage T4. Le mélange réactionnel utilisé pour la réaction de phosphorylation est le suivant :

## MATERIELS ET METHODES

[ $\gamma$ - <sup>32</sup> P] ATP (10 mCi/ml, 3000 Ci/mmol) (Amersham)	1 $\mu$ l
Oligonucléotide (100 ng/ $\mu$ l)	1 $\mu$ l
Tampon kinase 10x (MgCl <sub>2</sub> 0,1 M; DTT 0,05 M; Tris-HCl 0,5 M pH 7,6)	1 $\mu$ l
Polynucléotide kinase (10 U/ $\mu$ l) (MBI Fermentas)	1 $\mu$ l
H <sub>2</sub> O qsp 10 $\mu$ l	

Après incubation 45 minutes à 37°C, le mélange est chauffé pendant 10 minutes à 65°C pour inactiver la kinase, puis l'oligonucléotide est dilué à 1 ng/ $\mu$ l dans de l'eau stérile. Certains oligonucléotides doivent être purifiés sur gel de polyacrylamide 10% dénaturant. Dans ce cas, à l'issue de la réaction de marquage, 5  $\mu$ l de bleu de formamide sont ajoutés et le mélange est déposé sur le gel. L'électrophorèse est réalisée à 17 V/cm dans du tampon TBE 0.5x (acide borique 44.5 mM, EDTA 1 mM, Tris borate 44.5 mM, pH 8,3). Le gel est ensuite autoradiographié et la bande contenant l'oligonucléotide est excisée et incubée dans 400  $\mu$ l de tampon d'élution (NaCl 0,25M; EDTA 0,01M; Tris-HCl 0,02 M pH 7,5) à 4°C sur la nuit. La phase aqueuse est ensuite prélevée et centrifugée 1 min à 12000 rpm pour éliminer la présence éventuelle de fins débris de gel. L'oligonucléotide contenu dans le surnageant est alors précipité en présence d'éthanol absolu et de 1  $\mu$ l de glycogène (10  $\mu$ g/ $\mu$ l), sans ajout de sels. Le culot est ensuite lavé, séché puis repris dans de l'eau à raison de 100 cps/ $\mu$ l.

### II.1.4.3 Réaction de séquençage (Qu *et al.* 1983)

#### a. Hybridation de l'amorce oligonucléotidique

Les conditions d'hybridation sont les suivantes :

Oligonucléotide marqué en 5' à l' ATP [ $\gamma$ - <sup>32</sup> P]	1 $\mu$ l	}
Transcrit froid (1 pmole)	1 $\mu$ l	
Tampon AMV RT 10X (MgCl <sub>2</sub> 60 mM; KCl 0,4 M;		
Tris-HCl 0,5 M, pH 8,3)	0,25 $\mu$ l	
H <sub>2</sub> O qsp 2,5 $\mu$ l		

Le mélange est incubé durant 40 secondes à 96°C, puis est immédiatement plongé dans la glace.

### b. Réaction d'élongation

Aux 2,5 µl du mélange d'hybridation sont ajoutés :

Mélange de dNTP (5 mM chacun)	0,1 µl
Tampon AMV RT 10X	0,25 µl
Transcriptase inverse AMV (2 U/µl) (Q-Biogen)	0,25 µl
H <sub>2</sub> O qsp 5 µl	

A cette étape, 0,5 µl de l'un des ddNTP 500 µM est ajouté dans chacun des 4 tubes A, C, G et T du séquençage. Le mélange réactionnel est incubé pendant 30 minutes à 45°C. Les réactions sont arrêtées par ajout de 4 µl de bleu de formamide.

### c. Fractionnement sur gel

Les ADNc sont dénaturés 2 minutes à 96°C, puis placés dans la glace. L'analyse des produits de séquençage et d'élongation se fait par séparation sur gel de polyacrylamide dénaturant à 7%. Les échantillons sont déposés sur le gel (2 µl), l'électrophorèse est réalisée à 45 V/cm en tampon TBE 1x (acide borique 89 mM, EDTA 2 mM, Tris borate 89 mM, pH 8,3).

Les produits de séquençage et d'élongation ont parfois été séparés sur gel de polyacrylamide dénaturant à 10%. Dans ce cas, suite à l'électrophorèse, les ARN sont fixés par immersion du gel pendant 5 minutes dans une solution d'éthanol 30% et acide acétique 10% (v/v). Le gel est transféré sur un film plastique puis recouvert de papier Whatmann 3MM avant d'être séché sous vide et mis en fluorographie.

## II.2. Technique de retard sur gel

### II.2.1 Synthèse de transcrits radioactifs courts

Des transcrits uniformément marqués ont été produits à partir de matrices PCR produites dans les conditions décrites dans le paragraphe II.1.1. Le mélange réactionnel de transcription est le suivant :

Matrice ADN ou PCR (1µg/µl)	1µl
Tampon de transcription US 5X	
(MgCl <sub>2</sub> 30 mM; Spermidine 5mM; Tris-HCl 0,2 M) 8 µl	
DTT 200 mM	1µl
RNasine (31 U/µl) (Amersham biosciences)	0,5 µl
Mélange de rNTP : ATP (5 mM) CTP (5 mM) GTP (5 mM)(MBI Fermentas)	1 µl
UTP 0,1 mM (MBI Fermentas)	1 µl
α- <sup>32</sup> P UTP (800 Ci/mmol ; 10 mCi/ml) (ICN)	1 µl
ARN polymérase SP6 (20 U/µl) (MBI Fermentas) ou T7	1 µl
H <sub>2</sub> O qsp 30 µl	

Le mélange est incubé 4 heures à 37°C. La matrice ADN est ensuite dégradée par incubation en présence de 10 U de DNase I RQ1 (Amersham biosciences) exempte de RNases, pendant 30 minutes à 37°C. A l'issue de la réaction de transcription, 15 µl de bleu de formamide sont ajoutés et le mélange est déposé sur gel de polyacrylamide dénaturant à 6%. L'électrophorèse est réalisée à 17 V/cm dans du tampon TBE 0,5X (acide borique 44,5 mM, EDTA 1 mM, Tris borate 44,5 mM, pH 8,3). Le gel est ensuite autoradiographié et la bande contenant l'ARN est excisée et incubée dans 180 µl de tampon d'élution (NaCl 0,25 M; EDTA 0,01 M; Tris-HCl 0,02 M pH 7,5) à température ambiante sur la nuit. La phase aqueuse est ensuite prélevée et centrifugée 1 min à 12000 rpm pour éliminer la présence éventuelle de fins débris de gel. L'ARN est extrait au phénol/chloroforme puis chloroforme/alcool isoamylique et précipité en présence d'éthanol absolu et de 1 µl de glycogène (10 µg/µl), sans ajout de sels. Le précipité d'ARN est lavé deux fois avec de l'éthanol à 80%, puis séché et repris dans de l'eau stérile.

### II.2.2 Conditions expérimentales des retards sur gel

Cette technique aussi appelée EMSA (Electrophoretic Mobility Shift Assay) est basée sur le changement de mobilité électrophorétique d'un ARN lorsqu'il est lié à une protéine. En absence de protéines, l'ARN supposé porter des sites de fixation pour des protéines migre à une position donnée dans un gel non dénaturant. S'il est fixé par une ou plusieurs protéines, sa migration sera retardée.

20 fmoles d'ARN marqués et 0.1 pmoles d'ARN froid sont incubés dans le mélange réactionnel utilisé pour assembler les complexes ARN/protéines (décrit dans le Paragraphe II.1.4.1), en présence de 0.2 µg d'ARNt de levure. Après 15 min d'incubation à 4°C, 6 µl de bleu de dépôt non dénaturant (glycérol 40% ; bleu de bromophénol 0,03% ; xylène cyanol 0,03%; HEPES 200 mM pH 7,9) sont additionnés et les échantillons, déposés sur un gel d'acrylamide/bisacrylamide (59 :1) à 5% contenant 5% de glycérol. La migration est réalisée en tampon TBE 0,5X (acide borique 44,5 mM, EDTA 1 mM, Tris borate 44,5 mM, pH 8,3), glycérol 0,5%. Le gel est ensuite séché et exposé en fluorographie pendant une nuit.

### II.3. Pontages aux UV suivis d'immunoprécipitation

#### II.3.1. Pontages aux UV

Lorsqu'un complexe est formé entre un ARN et une protéine, il est possible de fixer covalentement la protéine à l'ARN par irradiation aux UV à 254 nm. Un transcrit marqué est utilisé pour permettre la visualisation du complexe sur un film. Le protocole de marquage du transcrit est indiqué dans le paragraphe II.1.1.

Dans un premier temps, un complexe est formé dans le mélange réactionnel suivant:

MgCl <sub>2</sub> 62,5 mM	0,5 µl	}
ARNt à 2 µg/µl	1 µl	
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%)	1,8 µl	
ARN cible uniformément marqué 250 000 cpm	1 µl	
H <sub>2</sub> O qsp 5.5 µl		

Le mélange est ensuite placé 10 min à 65°C, puis 10 min à température ambiante et enfin, 10 min dans la glace afin d'obtenir une structuration secondaire homogène. 5 µl de protéine ou d'un mélange de protéine dilué dans le tampon sont ensuite ajoutés et le mélange est incubé 20 min dans la glace. Après incubation, le volume de réaction est transféré dans un puits de plaque de microtitration, préalablement recouvert de 20 µg de sérumalbumine bovine pour limiter les interactions non-spécifiques des protéines sur le support. La plaque placée sur la glace est ensuite irradiée par 2 Joules/cm<sup>2</sup> aux UV à 254 nm.

Après pontage, l'ARN est hydrolysé par 50 U de RNase T1(Boehringer) et 15 µg de RNase A pendant 30 minutes à 40°C. Puis 4 µl de bleu de dépôt de protéines sont ajoutés à la réaction, le tout est chauffé 5 minutes à 96°C et déposé sur gel SDS-PAGE à 10 %. Un marqueur de poids moléculaire est fractionné en parallèle de façon à évaluer la taille des protéines fixées (Biolabs, 6-175 kDa). Après migration, le gel est soumis à une fluorographie et les protéines liées de manière covalente à des nucléotides peuvent ainsi être détectées.

### **II.3.2. Immunoprécipitation**

Cette technique permet d'identifier une protéine complexée à un ARN grâce à l'emploi d'anticorps spécifiquement dirigés contre cette protéine. La protéine G ou la protéine A (selon la chaîne lourde l'anticorps) fixant spécifiquement les chaînes lourdes des Ig, est liée de façon covalente à des billes de sépharose. Ceci permet de retenir sur les billes, les complexes ARN/protéine liés à l'anticorps utilisé dans l'expérience. Les anticorps utilisés nous ont été fournis par G. Dreyfuss (Philadelphie, USA) pour hnRNP A1, D. Black () pour hnRNP H, J. Stévenin (IGBMC, Illkirch) pour ASF/SF2, et Sylviane Müller pour la nucléoline.

Les anticorps d'intérêt sont fixés sur 40 µl de billes et ajouté au mélange ARN/protéines pontés aux UV dans 300µl 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%), avec 0.2 mg/ml de BSA (Sigma).

L'ensemble est incubé sur un agitateur rotatif pendant 2 h à 4°C. Les billes sont ensuite lavées 3 fois avec 1 ml de tampon IPP 1 % (NaCl 150 mM, NP40 1 %; Tris-HCl 50 mM pH 7,5) et remises en suspension dans 20 µl de bleu de dépôt de protéines.

## CHAPITRE III : Méthodes spécifiques à l'étude de l'épissage *in vitro*

### III.1 Synthèse de transcrits radioactifs *in vitro*

#### III.1.1 Préparation du fragment d'ADN à transcrire

Afin de limiter la transcription à la région nous intéressant, l'ADN servant de matrice est préalablement hydrolysé par une enzyme de restriction, au niveau d'un site localisé juste en aval de la séquence clonée. Après hydrolyse par une enzyme de restriction, l'ADN matrice est soumis à une extraction au phénol/chloroforme (1/1) et à une extraction par un mélange chloroforme/alcool isoamylique (24/1). L'ADN est ensuite précipité à l'éthanol en présence d'acétate de sodium 0,3 M. L'ADN linéarisé est repris dans de l'eau et ajusté à la concentration de 500 ng/ $\mu$ l.

#### III.1.2 Transcription *in vitro*

La synthèse de transcrits uniformément marqués au  $^{32}\text{P}$  est réalisée par réaction de polymérisation par l'ARN polymérase du phage T7 en présence d'un ribonucléotide marqué au  $^{32}\text{P}$ . L'ADN linéarisé (250 ng) est transcrit dans le milieu réactionnel suivant :

Tampon de transcription US 5X ( $\text{MgCl}_2$  30 mM; Spermidine 5mM; Tris-HCl 0,2 M) 2  $\mu$ l

DTT 200 mM	0,5 $\mu$ l
RNasin 31 U/ $\mu$ L (Amersham Biosciences)	0,25 $\mu$ l
Mélange de rNTP : ATP (5 mM), CTP (5 mM), GTP (2 mM)	1 $\mu$ l
UTP 0,5 mM	1 $\mu$ l
$[\alpha\text{-}^{32}\text{P}]$ UTP (800 Ci/mmol ; 10 mCi/ml) (ICN)	0,9 $\mu$ l
Coiffe m <sup>7</sup> GpppG 10 mM (NEB)	1,25 $\mu$ l
ARN polymérase T7 (61 U/ $\mu$ l) (USB)	1 $\mu$ l
BSA (2 mg/ml)(Sigma)	0,5 $\mu$ l
H <sub>2</sub> O qsp 10 $\mu$ l	

On diminue la concentration en GTP par rapport aux autres rNTPs, afin de favoriser l'incorporation de la coiffe m<sup>7</sup>GpppG à l'extrémité 5' du transcrit, ce qui protège l'ARN de la dégradation. Le mélange est incubé 1 h à 37°C. La matrice ADN est ensuite dégradée par incubation en présence de 10 U de DNase I, exempte de RNases (Amersham), 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'eau stérile, 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 nucléotides non incorporés. Le précipité d'ARN est lavé deux fois dans de l'éthanol à 80%, puis séché et repris dans 10 µl d'eau stérile. La quantité de transcrits radioactifs obtenue est déterminée par mesure de la radioactivité contenue dans 1 µl d'ARN au compteur à scintillation (Beckman). L'ARN est ensuite dilué dans de l'eau stérile à la concentration de 100 000 cpm (mesurés en Cerenkov/µl soit environ à 10 fmol/µl). La qualité de la transcription est vérifiée par dépôt d'une fraction aliquote (100 000 cpm) sur gel de polyacrylamide dénaturant (polyacrylamide 5%, urée 8 M, TBE 0,5x).

### III.2 Epissage *in vitro*

Pour réaliser les expériences d'épissage *in vitro*, nous avons utilisé des extraits nucléaires de cellules HeLa provenant de la société CilBiotech (Belgique). L'extrait nucléaire peut être décongelé et recongelé 1 à 2 fois sans perte notable d'activité. Les protéines SR, également fournies par J. Stévenin, sont des protéines recombinantes exprimées dans le système baculovirus (Cavaloc *et al.*, 1999).

### III.2.1 Réaction d'épissage

Chaque réaction d'épissage *in vitro* est réalisée avec 1 µl de transcrit radioactif, soit environ 100 000 cpm. Le transcrit est ajouté au mélange réactionnel suivant :

MgCl <sub>2</sub> 62,5 mM	1 µl
ATP 25 mM 1 µl	
Créatine phosphate 625 mM (Fournisseur)	1 µl
RNasin (31 U/µl)(Amersham Biosciences)	0,2 µl
Tampon RNasin (KCl 50 mM; DTT 5 mM; glycérol 50%;	
HEPES-KOH 20 mM pH 7,6)	0,2 µl
Alcool polyvinylique 13% (p/v)	6 µl
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%)	3,6 µl

La réaction est initiée par ajout à ce mélange de 8 µl d'extrait nucléaire, préalablement traité 10 minutes à 30°C afin d'éliminer l'ATP endogène. Le mélange est placé 5 minutes sur glace, puis incubé 2h30 à 30°C. Dans le cas des ARN dérivés de l'adénovirus l'incubation est réalisée pendant 1h30. Lors des tests de l'effet des protéines SR sur l'efficacité d'épissage *in vitro*, 150 à 300 ng de protéines SR recombinantes purifiées sont ajoutés au milieu réactionnel, en même temps que l'extrait nucléaire. Les échantillons sont traités de la même façon qu'au cours des expériences réalisées en absence de protéines SR. La réaction d'épissage est arrêtée par ajout de 20 µg de protéinase K (Roche) dans un tampon contenant : EDTA 12,5 mM; NaCl 150 mM; SDS 1% ; Tris-HCl 100 mM, pH 7,4. Après une incubation de 30 minutes à 30°C, les protéines sont extraites au phénol/chloroforme, puis au chloroforme/alcool isoamylique (24/1) et les ARN sont 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).

### III.2.2. 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) (catalyseur de la polymérisation) et persulfate d'ammonium 0,1% (p/v) (initiateur de la polymérisation). 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. Dans le cas des ARN dérivés de l'adénovirus, l'analyse des produits d'épissage est réalisée sur un gel d'acrylamide 6%, pendant 1h15. Les ARN sont ensuite fixés par immersion du gel pendant 5 min dans une solution d'acide acétique 10 % (v/v). 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 fluorographie à -80°C au contact d'un film Fuji XRay, en présence d'écrans amplificateurs.

### III.3 Détermination de l'efficacité d'épissage

Le gel est scanné par un appareil PhosphorImager (Molecular Dynamics) 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 messenger (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.

## CHAPITRE IV : Epissage *in cellulo*

### IV.1. Transfection et extraction ARN

- Transfection aux Jet-PEI :

Les cellules HeLa (2X10<sup>5</sup> cellules par puits) sont transfectées dans des plaques 6 puits avec 1 µg de plasmide pΔPSP et un plasmide pXJ41 surexprimant les protéines ASF/SF2 et hnRNP K, en utilisant le réactif Jet-PEI selon les spécifications du fournisseur (PolyPlus Transfection). Les cellules sont ensuite incubées pendant 48h, puis lavées au PBS 1X (KCl 2,7mM ; KH<sub>2</sub>PO<sub>4</sub> 1,8mM ; NaCl 137mM ; Na<sub>2</sub>HPO<sub>4</sub> 10,1mM). Les ARN totaux cellulaires sont finalement extraits des cellules HeLa au Trizol selon les spécifications du fournisseur (Invitrogen).

- Transfection des siRNA au phosphate de calcium :

La veille, environ 200 000 cellules sontensemencées par boîte de 3,5 cm. Le milieu de culture est changé 1 à 5 h avant transfection. Pour une boîte 3,5 cm :

50 pmoles siRNA

qsp 36,8 µl avec TE 0,1X ( Tris 1mM, EDTA 0,1mM)

5,2 de CaCl<sub>2</sub> (2 M)

42 µl HBS 2X en faisant buller (entraîne la précipitation) (Hepes 50 mM, NaCl 0,28 M, KCl 10 mM, pH 7,05).

L'incubation se fait 20 min à température ambiante avant d'être déposé sur les cellules.

### IV.2. Analyse RT-PCR

Les ARN totaux sont rétro transcrits par la reverse transcriptase MuMLV en utilisant comme primer l'oligonucléotide 2144 (5'-TGGAGGTGGGTTGCTTTGATAGAG-3', complémentaire aux séquences 6029–6041 et 8369–8381 de l'ARN de la souche HIV-1 pNL4–3). Les conditions de la réaction sont les suivantes : 2µg d'ARN totaux préalablement repris dans 12,5µl d'H<sub>2</sub>O, puis dénaturés 5min à 65°C et renaturés dans la glace ; 0,5µl d'oligonucléotide (100µM) ; Tampon MuMLV 5X (Invitrogen) ; 0,25µl DTT (1M) ; 1,25µl

## MATERIELS ET METHODES

---

dNTP (10mM chacun) ; 0,5µl MuMLV-RT (200U/µl) (Invitrogen) ; H<sub>2</sub>O qsp 25 µl. La réaction a lieu pendant 45min à 42°C.

Les produits de la réaction de rétro transcription sont ensuite amplifiés par PCR en utilisant les oligonucléotides 2143 (5'-GGCTTGCTGAAGCGCGCACGGCAAGAGG-3', contenant la séquence 700–727 de l'ARN de la souche HIV-1 pNL4–3) et 2144, dans les conditions suivantes : 2µl de RT ; 0,5µl de chaque oligonucléotides (100µM) ; Tampon GoTaq Flexi Green 5X (Promega) ; 1,5µl MgCl<sub>2</sub> (25mM) ; 0,5µl dNTP (10 mM chacun) ; 0,125µl GoTaq Flexi ADN polymérase (5U/µl) (Promega) ; H<sub>2</sub>O qsp 25 µl. Après une dénaturation 5min à 96°C, l'ADN est amplifié pendant 30 cycles de 30sec à 96°C, 30sec à 60°C, 2min à 72°C. Les ADNc produits (10µl de la réaction) sont ensuite séparés par électrophorèse sur gel non dénaturant 6% polyacrylamide, puis le gel est coloré par le bromure d'éthidium et analysé par le système d'images Genegenius (Syngene). Les ADNc sont enfin identifiés par référence aux données précédemment obtenues (Ropers *et al.*, 2004).

### **CHAPITRE V : Dimérisation**

40 nM d'ARN non radiomarqué sont dilués dans 8 µl d'eau avec 5000 cpm soit 3 à 5 nM du même ARN radiomarqué. Les échantillons sont dénaturés deux minutes à 90 °C et refroidis rapidement pendant deux minutes dans la glace. La dimérisation est initiée par l'addition de tampon de dimérisation concentré 5X (concentration finale : 50 mM Tris pH 7.5, 300 mM KCl et 5 mM MgCl<sub>2</sub>) et les échantillons sont incubés pendant 30 minutes à 37 °C. En parallèle, des échantillons sont incubés avec du tampon monomère 5X (concentration finale : 50 mM Tris pH 7.5, 40 mM KCL et 0.1 mM MgCl<sub>2</sub>) pour induire le repliement de l'ARN monomérique. Les échantillons sont déposés sur un gel d'agarose dans du tampon TBM (0.53 mM Tris-Borate, 0.1 mM MgCl<sub>2</sub>) et migré à 4°C. Les gels sont séchés sous vide à température ambiante quinze minutes puis à 80 °C quinze minutes. Les bandes radioactives correspondant aux formes monomériques et dimériques sont visualisées et quantifiées au PhosphorImager.

### CHAPITRE VI : Identification de nouvelles protéines en spectrométrie de masse.

#### VI.1 Principe de la méthode de chromatographie d'affinité basée sur l'interaction ARN MS2-protéine MBP.

Cette méthode repose sur le couple protéine de l'enveloppe du phage MS2-site de liaison de cette protéine sur l'ARN de ce phage (Figure I-52). L'affinité entre les deux partenaires est forte (Kd de l'ordre du nM) (Jurica *et al.*, 2002). La protéine de l'enveloppe est produite sous forme de fusion avec une protéine de liaison du maltose (MBP ou « Maltose Binding Protein »). Elle est fixée sur des billes d'amylose (Jurica *et al.*, 2002). L'ARN dont on recherche les partenaires protéiques est fusionné à l'ARN MS2 contenant plusieurs sites de fixation de la protéine de l'enveloppe du phage MS2. Le principe de l'expérience est le même que celui des expériences basées sur l'emploi de la tobramycine. Cette méthode présente deux avantages : la formation des complexes peut être faite à pH physiologique 7,5, car les problèmes qui étaient liés à la présence de charges sur la tobramycine n'existent pas avec cette méthode. Par ailleurs, les billes d'amylose semblent avoir un plus faible pouvoir de rétention des protéines comparées aux billes de sépharose. Le ratio MS2-MBP/ARN doit être optimisé pour augmenter l'efficacité de la formation des complexes et leur purification (Zhou *et al.*, 2002).

#### VI.2 Paramètres critiques :

- L'efficacité de la formation des complexes ARN-protéine est un facteur majeur de détermination du rendement final des complexes purifiés. Des expériences de retard sur gel peuvent permettre d'évaluer la capacité de fixation ARN-protéine.
- La qualité des extraits nucléaires a une grande importance. Une haute qualité permet une meilleure capacité de l'assemblage des spliceosomes et une dégradation moindre de l'ARN. Un autre facteur important est la qualité de la protéine MS2-MBP. Sa purification sur une colonne héparine est importante pour assurer sa capacité de fixation à l'ARN et un rendement final correct des complexes. Le dernier paramètre critique est le contrôle négatif de l'expérience avec l'ARN MS2 seul pour identifier les protéines liées non spécifiquement.
- Les dégradations potentielles de l'ARN ou des protéines réduirait ou éliminerait le rendement de purification des complexes. Ainsi, à chaque étape, une fraction de la réaction est prélevée et déposée d'une part sur gel SDS-PAGE pour vérifier l'intégrité des protéines et d'autre part sur gel acrylamide/urée 19 :1 pour vérifier l'intégrité des ARN.

### **VI.3 Liaison de l'ARN-MS2 avec la protéine MBP-MS2**

100 pmol d'ARN sont dilués dans 200 µl d'eau et avec du MgCl<sub>2</sub> (4mM final) et ajoutés à 200 µl de tampon D. Le tampon D va être ainsi dilué de moitié. Les échantillons vont être dénaturés 10 minutes à 65 °C puis renaturés 10 minutes à température ambiante puis 10 minutes dans la glace.

La protéine MS2-MBP est ajoutée dans un ratio de 5x la concentration molaire d'ARN, soit 500 pmoles de protéines MS2-MBP (56 kDa, donc 28 µg) pour 100 pmoles d'ARN. Ce mélange est incubé 20 minutes dans la glace.

Une fraction de mélange est prélevée à chaque étape pour vérifier l'absence de dégradation de l'ARN.

### **VI.4 Fixation sur les billes amylose**

40 µl de suspension de résine amylose (Biolabs) sont utilisées soit 20 µl de billes par ARN. La résine est lavée deux fois avec 5 volumes de tampon D sans glycérol.

Les billes sont incubées avec le mélange d'ARN-MS2 et de la protéine MS2-MBP une heure sur roue à 4°C. Une fraction du surnageant est prélevée pour évaluer le taux de protéine MS2-MBP fixées ainsi que l'absence de dégradation de l'ARN.

### **VI.5 Fixation des protéines nucléaires**

Pour éliminer les complexes ARN/protéines qui ne se seraient pas fixées sur les billes, elles sont lavées trois fois avec 500 µl de tampon D.

100 µl d'EN dialysés contre le tampon D sont ajoutés ainsi que des ARNt à 5µM final. Ce mélange est incubé une heure sur roue à 4°C.

Dans le cas de l'expérience de double purification, l'extrait nucléaire est mélangé préalablement à 20 µg de protéine hnRNP K. Le mélange est ajouté sur la résine ainsi que des ARNt à 5µM final. Ce mélange est incubé une heure sur roue à 4°C.

Après élimination du surnageant, la résine amylose est lavée au moins trois fois avec 500 µl de tampon D pendant 10 minutes sur roue pour chaque lavage.

### **VI.6 Elution**

Les protéines sont éluées avec un mélange de tampon D avec 10 mM de maltose pendant 15 minutes sur roue ou vortex doux à température ambiante.

Pour la purification des complexes spliceosomaux présentés dans le premier chapitre, seules les étapes 1 à 4 de la figure M-6 ont été nécessaires. Les étapes 5 à 7 permettent la double purification présentée dans le chapitre IV et pour cette expérience, de la protéine hnRNP K recombinante a été ajoutée à l'extrait nucléaire.

### **VI.7 Purification sur colonne Ni-NTA pour la double purification.**

L'élution de la résine amylose est directement déposée sur des billes Ni-NTA (40 µl de slurry soit 20 µl de billes) lavées 3 fois au tampon D. Une fraction de l'élution est incubée 30 minutes à 37 °C avec 20 µg de RNase A pour digérer l'ARN. Une fraction supplémentaire est conservée pour des western blot.

Les mélanges (+/- RNase A) sur la résine Ni-NTA sont incubés une heure à 4°C sur roue.

Après élimination du surnageant, les billes sont lavées quatre fois avec 500 µl de tampon D sans glycérol pendant 10 minutes sur roue pour chaque lavage.

L'élution se fait à température ambiante sur roue ou vortex doux en ajoutant 80 µl de tampon Tris-Hcl 20mM, NaCl 250, 250 mM imidazole et 2 mM EDTA.

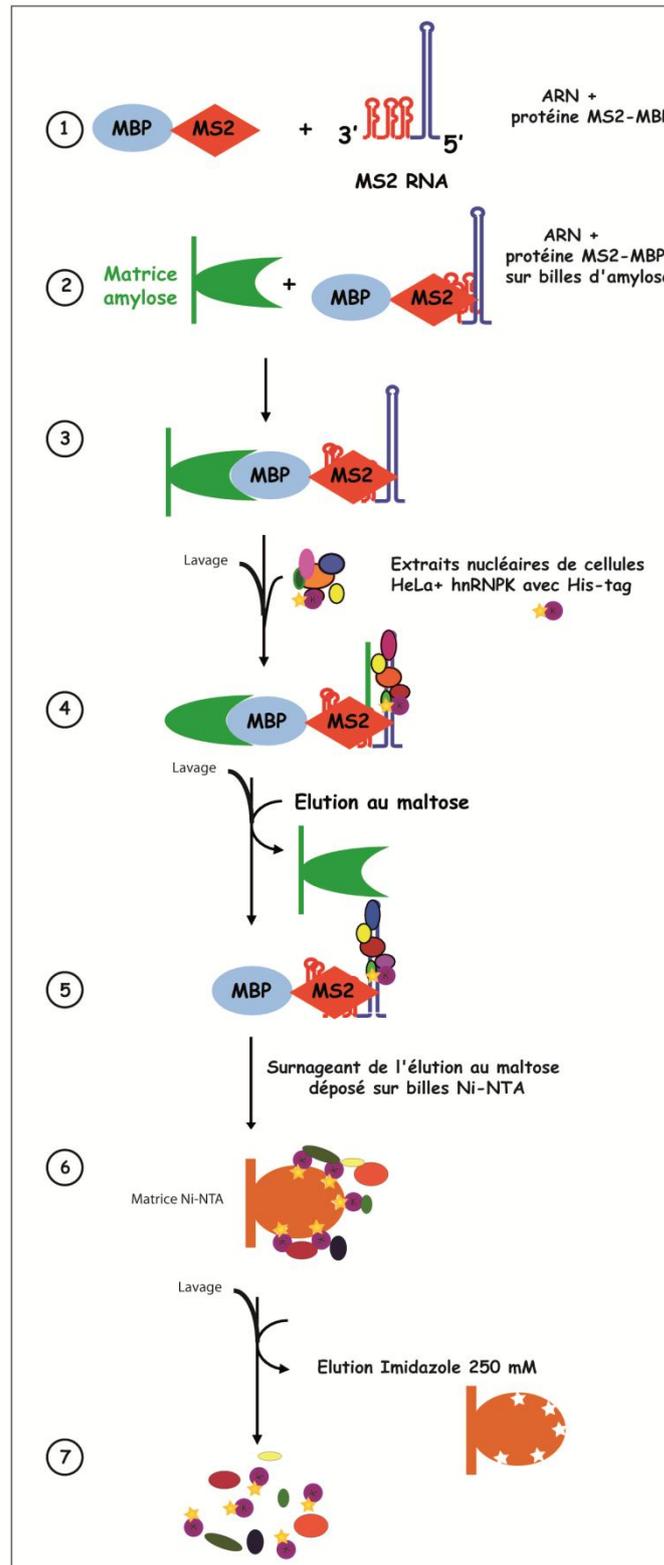
Une fraction d'élution est conservée pour des Western blot. Le reste est déposé sur gel SDS-PAGE 10 % puis coloré au bleu colloïdal.

Tous les tampons sont exempts de kératine qui pourrait brouiller le signal de spectrométrie de masse.

Une petite fraction de chaque étape est prélevée pour analyser l'absence de dégradation de l'ARN à chaque étape et la présence des protéines.

### **VI.8 Analyse en spectrométrie de masse.**

Les protéines contenues dans la bande de gel ont été réduit-alkylées, puis digérées à la trypsine. Les peptides ont ensuite été extraits du gel par un solvant ACN/H<sub>2</sub>O/HCOOH (50/50/1). L'échantillon est injecté sur nanoLC-Chip 150mm (Agilent) couplée au spectromètre de masse HCT Ultra (Brucker).



**Figure : Purification de protéines liées à la fois à l'ARN SLS2 et à la protéine hnRNP K.**

Représentation schématique de la technique de purification grâce à une première sélection des protéines d'un extrait nucléaire de cellules HeLa liant l'ARN SLS2 et à une seconde sélection des protéines parmi celles isolées qui lient hnRNP K.

# **ANNEXES**



**Identification of protein partners of the human immunodeficiency virus 1 *tat/rev* exon 3 leads to the discovery of a new HIV-1 splicing regulator, protein hnRNP K**

Virginie Marchand<sup>1 X O</sup>, Maryline Santerre<sup>1 X</sup>, Christelle Aigueperse<sup>1</sup>, Laetitia Fouillen<sup>2</sup>, Jean-Michel Saliou<sup>2</sup>, Alain Van Dorsselaer<sup>2</sup>, Sarah Sanglier-Cianfèrani<sup>2</sup>, Christiane Branlant<sup>1+</sup> and Yuri Motorin<sup>1</sup>

<sup>1</sup> Laboratoire ARN-RNP Maturation-Structure-Fonction, Enzymologie Moléculaire et Structurale (AREMS), UMR7214 CNRS-UHP, Faculté des Sciences et Technologies, Nancy Université, Bld des Aiguillettes, BP 70239, 54506 Vandoeuvre-les-Nancy, France

<sup>2</sup> Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Université de Strasbourg, IPHC, 25 rue Becquerel 67087 Strasbourg, France. CNRS, UMR7178, 67037 Strasbourg, France.

<sup>X</sup> The two first authors contributed equally to this work

<sup>+</sup> Corresponding author

Christiane BRANLANT

Laboratoire AREMS UMR7214 CNRS-UHP, Faculté des Sciences et Technologies, Nancy Université, Bld des Aiguillettes, BP 70239, 54506 Vandoeuvre-les-Nancy, France

Tel: +33 3 83 68 43 03, Fax: +33 3 83 68 43 07

Email: [Christiane.Branlant@maem.uhp-nancy.fr](mailto:Christiane.Branlant@maem.uhp-nancy.fr)

<sup>O</sup> Present address:

European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany

**Keywords:** virus HIV-1, RNP affinity chromatography, mass spectrometry, splicing regulation, Nucleolin, hnRNP K

## **Acknowledgements**

R Lührmann is thanked for his generous gift of plasmids MS2-MBP and pAdML-M3 and for the training in his laboratory of VM. D Black is acknowledged for his generous gift of plasmid pET15b-hnRNP H. J Stevenin is thanked for his generous gift of plasmids pXJ41 and pXJ41-ASF/SF2 and the ASF/SF2 antibodies. P Bouvet is thanked for his generous gift of plasmid pET15b-p30. G Dreyfuss is thanked for his generous gift of plasmid pET11a-hnRNPA1 and of hnRNP A1 specific antibodies. J Valcarcel is acknowledged for his generous gift of U2AF65 specific antibodies. AM Aubertin is thanked for giving us the MaTG0020 antibody directed against the Nef protein. S Muller is thanked for her generous gift of nucleolin antibodies. We thank C. M. Stoltzfus from Iowa University for the generous gift of plasmid  $\Delta$ PSP. VM and MS were fellows of the french « Ministère de la Recherche et de l'Enseignement Supérieur ». VM was a post-doctoral fellow of the EURASNET European network. MS was a fellow of the Sidaction foundation for her last year of PhD preparation. LF was supported by the Alsace Region and AliX compagny. JMS was a post-doctoral fellow of Sidaction foundation. This work was supported by the Centre National de la Recherche Scientifique (CNRS), the french « Ministère de la Recherche et de l'Enseignement Supérieur », the Université de Strasbourg (UdS), the Alsace Region, the french « Agence Nationale de la Recherche sur le SIDA » (ANRS) (Contracts numbers 05195-06194 and 08061-09069), the french « Agence Nationale pour la Recherche » (ANR) (contract number ANR-05-BLAN-0261-03), the European Community (EURASNET Rex, Contract number LSHG-CT-2005-518238) and the « Pôle de Recherche Scientifique et Technologique Bioingénierie » of Région Lorraine.

**Abbreviations:**

RRM, RNA recognition motif

EMSA, Electrophoretic Mobility Shift assays

hnRNP heterogeneous nuclear RiboNucleoProtein

RRE Rev Response Element

ESE Exonic Splicing Enhancer

ESS Exonic Splicing Silencer

ISS Intronic Splicing Silencer

RNP RiboNucleoprotein Particle

NRE Nucleolin Recognition Element

Nano LC-MS /MS Liquid Chromatography coupled to tandem Mass Spectrometry

SLS Stem Loop Structure

MS2-MBP MS2-Maltose Binding Protein

HOT Head Over Tail

SELEX Systematic Evolution of Ligands by EXponential enrichment

**Running title:** *Proteomic identifies a HIV-1 splicing regulator*

## Abstract

HIV-1 pre-mRNA splicing depends upon 4 donor and 8 acceptor sites, which are used in combination to produce more than 40 different mRNAs. Acceptor site A7 plays an essential role for *tat* and *rev* mRNA production. The SLS2-A7 stem-loop structure containing site A7 was also proposed to modulate HIV-1 RNA export by the Rev protein. To further characterize nuclear factors involved in these processes, we purified RNP complexes formed by incubation of SLS2-A7 transcripts in HeLa cell nuclear extracts by affinity chromatography and identified 33 associated proteins by nanoLC-MS/MS. By UV cross-linking, immunoselection and EMSA, we showed that, in addition to the well known hnRNP A1 inhibitor of site A7, nucleolin, hnRNP H and hnRNP K interact directly with SLS2-A7 RNA. Nucleolin binds to a cluster of successive canonical NRE motifs in SLS2-A7 RNA, which is unique in HIV-1 RNA. Proteins hnRNP A1 and hnRNP K bind synergistically to SLS2-A7 and both have a negative effect on site A7 activity. By the use of a plasmid expressing a truncated version of HIV-1 RNA, we showed a strong effect of the over-expression of hnRNP K in HeLa cells on HIV-1 alternative splicing. As a consequence, production of the Nef protein was strongly reduced. Interestingly also, many proteins identified in our proteomic analysis are known to modulate either the Rev activity or other mechanisms required for HIV-1 multiplication and several of them seems to be recruited by hnRNP K, suggesting that hnRNP K plays an important role for HIV-1 biology.

## Introduction

Splicing plays a key role for human immunodeficiency virus 1 (HIV-1) multiplication. The integrated DNA copy of the viral HIV-1 genome is transcribed as a long RNA that contains several splicing sites (4 donor and 8 acceptor sites) which can be used according to a complex pattern of combinations.<sup>1</sup> In an early phase of virus multiplication, the long transcript is multiply spliced to give rise to the *tat*, *rev* and *nef* mRNAs.<sup>2</sup> The Tat protein strongly increases production of full-length viral transcripts.<sup>3</sup> The Nef protein modifies cellular activities of host cells in order to stimulate virus multiplication.<sup>4</sup> When a sufficient amount of Rev protein is produced, it oligomerizes on its response element (RRE) in the viral RNA, which allows the transport of unspliced or singly spliced viral RNAs into the cytoplasm.<sup>5</sup> Unspliced RNA is used as genomic RNA for the production of new viruses and is required to serve as mRNA for synthesis of Gag-Pol precursor protein. The singly spliced mRNAs encode the Env precursor protein as well as the Vpr, Vpu and Vif proteins.

The *tat* and *rev* ORFs are interrupted by a common intron, which is delimited by the D4 donor site and the A7 acceptor site.<sup>1</sup> Although, Tat plays an essential role for virus multiplication, virus HIV-1 controls its production, because of the apoptotic properties of the Tat protein, on both infected and non-infected cells.<sup>6</sup> This control largely takes place at the splicing step. The two acceptor sites used to produce *tat* mRNAs (sites A3 and A7) are highly down-regulated by the presence of splicing silencers that bind hnRNP proteins (hnRNP H and hnRNP A1 at site A3 and hnRNP A1 at sites A1, A2 and A7).<sup>2, 7-17</sup> As the Tat and Rev proteins are required for HIV-1 multiplication, abolition of their production at the splicing step may be one way to limit HIV-1 progression. Therefore, it is of high importance to understand how splicing at sites A3 and A7 is regulated. Utilization of site A7 is regulated by a combination of *cis* regulatory sequences. Three silencer elements that bind protein hnRNP A1 in a cooperative manner were identified: ISS (Intronic Splicing Silencer) in the upstream intron, ESE3 (Exonic

Splicing Silencer 3) which is both a silencer and an enhancer element depending on the associated proteins and, ESS3, a bipartite Exonic Splicing Silencer composed of the ESS3a and ESS3b elements.<sup>8, 10-12, 18-20</sup> Two Exonic Splicing Enhancers (ESE2 and ESE3) were also detected.<sup>8, 10, 18, 19</sup> ESE3 was clearly shown to bind the SR protein ASF/SF2.<sup>19</sup> ESE2 was also proposed to bind this protein.<sup>18</sup> More recently, hnRNP E1 and hnRNP E2 proteins were found to associate with the SLS3 stem-loop structure of *tat/rev* exon 3 containing ESS3a. Each of these two proteins has a distinct role in modulating HIV-1 gene expression. Binding of hnRNP E1 to ESS3 reduces *rev* mRNA translation and HIV-1 RNA transport.<sup>21</sup>

The secondary structure of the HIV-1 RNA region containing site A7 was previously studied experimentally.<sup>8, 12</sup> The ESE2 and ESE3 splicing regulatory elements were found to be located in an irregular stem-loop structure called SLS2-A7 (positions 7965 to 8046 as referred to nucleotide positions in the HIV-1 BRU isolate). ISS was found in an upstream SLS1-A7 stem-loop structure and ESS3 in an SLS3-A7 downstream stem-loop structure.<sup>8, 12</sup> A 2D structure model for the entire HIV-1 RNA was recently proposed.<sup>22</sup> Some but not all of the previously described 2D structure motifs around site A7 are found in this model. However, it should be pointed out that in the present stage of the knowledge on the thermodynamic stability of RNA structures, it is very difficult to establish an accurate 2D structure of a 9,173-nt long RNA, using a single chemical reagent as the probe. Previous studies were performed using two chemical reagents and three enzymatic probes. In addition, data from our previous structural analysis were confirmed by the analysis of several fragments with variable lengths. In the present study we will use the well established and previously published 2D structure.<sup>8,</sup>

12

The SLS2-A7 stem-loop structure includes site A7 and the *tat* mRNA termination codon. Therefore, proteins bound to the SLS2-A7 region may be involved in regulation of splicing at site A7, genomic HIV-1 RNA nuclear export, *tat* mRNA transport to the cytoplasm and/or

translation regulation. Furthermore, as the RRE element is located at a relatively short distance (312 nts) upstream from stem-loop SLS2-A7, cross-talks between proteins bound to SLS2-A7 and RRE are not excluded.

For all these reasons, it was important to further characterize nuclear factors able to bind directly or through protein-protein interaction to the SLS2-A7 stem-loop structure. To this end, we used two distinct RNP affinity chromatography methods in order to purify the RNP complexes formed upon incubation of SLS2-A7 transcripts in HeLa nuclear extract in conditions used for *in vitro* splicing assays. These methods are based on the transcriptional fusion of a tag sequence (three MS2 coat protein binding sites or an aptamer with high affinity for tobramycin) at one extremity of the studied RNA.<sup>23-25</sup>

Hence, here we describe how by this proteomic-based approach coupled with mass spectrometry analysis, we identified novel protein partners of the SLS2-A7 HIV-1 RNA region. We focused further experimental studies on three of these proteins that interact directly with SLS2-A7 RNA: nucleolin, hnRNP H and hnRNP K. We demonstrated the presence of a cluster of three strong nucleolin binding sites in the SLS2-A7 sequence, which is unique in the entire HIV-1 RNA sequence. Furthermore, as by *in vitro* splicing assays, we found that hnRNP K can regulate site A7 utilization, we tested whether it can also regulate the *in cellulo* utilization of other HIV-1 splicing sites and this turned to be the case. Therefore, we looked for the effect of hnRNP K over-expression in HeLa cells on Nef protein production. Finally, on the basis of data in the literature, we evaluated the possible functional implication of the association with SLS2-A7 RNA of nucleolin and the other identified protein partners.

## Results

### Purification and proteomic analysis of SLS2-A7-MS2 RNP complexes

With the aim to identify new protein partners of the HIV-1 *tat/rev* exon 3 5'-terminal region, three binding sites for the MS2 coat protein were transcriptionally linked to the 3'-extremity of an HIV-1 BRU RNA segment extending from positions 7965 to 8046 (**Fig. 1**). This segment folds into a stem-loop structure called SLS2-A7. It includes site A7 and the downstream splicing regulatory elements ESE2 and ESE3 (**Fig. 2A**). Each of the three MS2 coat protein binding sites was able to fold independently into a stable stem-loop structure. We verified experimentally that their addition did not alter SLS2-A7 RNA folding (data not shown). The experimental protocol that we used to purify the SLS2-A7-MS2 RNPs formed in a HeLa cell nuclear extract was inspired from the one adapted by the Lührmann's team to purify spliceosomal complexes.<sup>25</sup> RNP complexes formed upon incubation of an RNA fragment containing several splicing regulatory elements in nuclear extract were expected to be heterogeneous. In view to preserve this overall diversity, we could not use a second purification step. Hence, we adapted the experimental conditions in order: (i) to get efficient and specific binding of the SLS2-A7-MS2 RNA to the beads and, (ii) to limit unspecific retention of the proteins on the beads. As *in vitro* splicing assays are performed in buffer D, we used this buffer to settle the experimental conditions (see Materials and Methods). In parallel to the assay performed on SLS2-A7-MS2 RNA, control experiments were made in the absence of RNA (data not shown) and with a MS2 control RNA that only contained the three MS2 coat protein binding sites (**Fig. 1**). The SLS2-A7-MS2 and MS2 RNPs were eluted in non-denaturing conditions and their protein contents were analyzed in parallel by 1D SDS-PAGE electrophoresis followed by colloidal Coomassie blue staining. Highly reproducible fractionation patterns were obtained in independent experiments and a representative example is shown in **Figure 2B**. The strong difference of pattern obtained for SLS2-A7-MS2 and MS2

RNPs reflected the specificity of the approach. Therefore, the bands of interest in the SLS2-A7-MS2 and MS2 analyses were excised and then in-gel trypsin-digested and further analyzed by nanoLC-MS/MS for protein identification.

With classical MS validation criteria (see Materials and Methods) 44 unique proteins were identified in SLS2-A7-MS2 RNPs (**Table 1**). Among them 33 were identified in SLS2-A7-MS2 RNPs but not in MS2 RNPs. **The proteomic analysis was repeated twice on independent batches of purified RNPs and similar proteins were reproducibly detected** (data not shown).

As will be described below, the specific association of some of the detected proteins with SLS2-A7 RNA was also confirmed by Western blot analysis using specific antibodies (**Fig. 2C**). Finally, in order to verify that protein identification was not biased by the type of affinity chromatography used, we repeated the RNP purification using another kind of RNA tag, the tobramycin high affinity aptamer J6f1<sup>26</sup> which was also used previously for affinity purification of splicing complexes.<sup>23</sup> This aptamer was transcriptionally linked at the 3' end of RNA SLS2-A7 (SLS2-J6f1 RNA, **Fig. 1**) and we used a protocol derived from the one established for splicing complexes for RNP purification (see Materials and Methods). A control experiment was performed using the J6f1 RNA alone. The SDS-PAGE pattern of protein fractionation was rather similar to the one obtained for the SLS2-A7-MS2 RNPs (data not shown). Mass-spectrometry analysis of the excised protein bands revealed the presence of 24 proteins already identified in the SLS2-A7-MS2 RNPs (**Table 1**). The absence of some of the proteins in SLS2-A7-J6f1 RNPs was probably due to the higher pH (9.1) which has to be used in the tobramycin based protocol compared to the MS2-MBP based protocol.

The proteins identified by both affinity methods belong to several classes of cellular proteins (**Table 1**). However, many of them are known to be involved in the biogenesis, metabolism or function of nucleic acids. This global proteomic approach afforded not only to identify

proteins already known to interact with SLS2-A7 RNA such as hnRNP A1 protein, but also unknown candidate partners that might either interact directly with the RNA or be associated through protein-protein interactions. Due to the high number of identified proteins, we next focused the experimental work on some of them and we choose proteins that directly interact with SLS2-A7 RNA. However, we performed a deep investigation on the known biological roles of all the detected proteins (**Table 1**), and the possible functional implications of the association of some of them at site A7 will be discussed.

### **Comparison of proteomic and UV cross-linking data identifies candidate direct partners of SLS2-A7 RNA**

As most of the proteins that bind directly to RNA can be cross-linked to the RNA upon UV irradiation at 254 nm<sup>7, 8</sup>, we used this approach to identify which of the detected SLS2-A7 RNP proteins interact directly with the RNA. UV cross-linking experiments were performed on RNP complexes formed by incubation of a [<sup>32</sup>P]-labeled *in vitro* transcribed SLS2-A7 RNA in HeLa cell nuclear extract (see conditions in the Materials and Methods). Subsequently, the SLS2-A7 RNA was digested with RNase T1. The resulting [<sup>32</sup>P]-labeled proteins were fractionated by SDS-PAGE and their presence and relative labeling levels were revealed by autoradiography. Three very intense and some less intense bands were reproducibly detected in independent experiments. According to the marker proteins, the cross-linked proteins in the intense bands were estimated to have apparent molecular masses of about 100, 65 and 35 kDa, respectively (**Fig. 3A**). HnRNP A1 which is known to interact with SLS2-A7 RNA<sup>8, 12</sup> has a calculated molecular mass of 38.4 kDa. Therefore, we hypothesized that it might correspond to the intense band at about 35 kDa. Then, to try to interpret other cross-linking data, we took into consideration the apparent molecular masses of the proteomic identified proteins as well as their known RNA binding specificities. Nucleolin found in one of the more intense bands in the SDS-PAGE analyses of both SLS2-A7-MS2

RNPs (**Fig. 2B**) and SLS2-A7-J6f1 RNPs (not shown) migrates with an apparent molecular mass of 100-110 kDa on SDS-PAGE.<sup>27</sup> In addition, SLS2-A7 RNA contains three nucleotide sequences (motifs 1, 2 and 3 in **Figure 3B** left) that share strong similarity with known NRE (Nucleolin Recognition Elements, (U/G)CCCG(A/G)) found in human pre-ribosomal RNA.<sup>27, 28</sup> Hence, we hypothesized that the cross-linked 100 kDa protein might be nucleolin. We guessed that hnRNP K with an apparent molecular mass of 65 kDa in the SDS-PAGE analysis of SLS2-A7-MS2 RNPs, might be the cross-linked protein found at about 65 kDa, because it is known to bind to C-rich segments in RNAs<sup>29</sup> and SLS2-A7 RNA contains several stretches of C residues (**Fig. 3B** middle). Finally, as SLS2-A7 RNA also contains a stretch of G residues (**Fig. 3B**) known to have a high affinity for hnRNP H<sup>30, 31</sup> and as hnRNP H was detected with an apparent molecular mass of 50 kDa in the proteomic analysis (**Fig. 2B**), we hypothesized that hnRNP H might also be cross-linked to SLS2-A7 RNA. **Western blot analysis of proteins contained in the SLS2-A7-MS2 and MS2 RNPs performed with specific antibodies directed against nucleolin, hnRNP K/J, hnRNP F/H/H' and hnRNP A1 (Fig. 2C), confirmed the specific association of nucleolin, hnRNP K and hnRNP H with SLS2-A7 RNA as compared to MS2 RNA. Then, we verified that these three proteins interact directly with SLS2-A7 RNA.**

### **Demonstration of the direct interaction of nucleolin, hnRNP K and hnRNP H with SLS2-A7 RNA**

To test for the direct interaction of nucleolin, hnRNP K and hnRNP H with SLS2-A7 RNA, we immunoselected the cross-linked proteins in the UV-irradiated SLS2-A7 RNPs using specific antibodies directed against each of these three proteins (**Fig. 3A**). HnRNP A1 specific antibodies were used as a positive control. As U2AF<sup>65</sup> binds to the 3'ss polypyrimidine tracts (PPTs) and only a small portion of the A7 PPT was present in the SLS2-A7 RNA (**Fig. 2A**), antibodies directed against the U2AF<sup>65</sup> splicing factor were used as a negative control (**Fig.**

**3A**). As expected, hnRNP A1 was identified as one of the cross-linked proteins, whereas no U2AF<sup>65</sup> cross-link was detected. The positive immunoselections obtained with antibodies directed against hnRNP K/J, nucleolin and hnRNP H/H'/F confirmed the direct interaction of these proteins with SLS2-A7 RNA (**Fig. 3A**).

To further characterize these interactions, we performed gel-shift assays (EMSA) using an *in vitro* transcribed [<sup>32</sup>P]-labeled SLS2-A7 RNA and purified full-length His-tagged recombinant proteins (**Suppl. Fig. 1**). Among the three proteins tested, only recombinant hnRNP K protein turned to be soluble in *E. coli* and could be used for EMSA experiments. It was found to bind SLS2-A7 RNA with an apparent K<sub>d</sub> of about 0.7 μM (**Suppl. Fig. 1B**). As the two N-terminal RRM of hnRNP H are known to be the determinant for hnRNP H RNA binding specificity<sup>30</sup>, we produced an His-tagged fragment of hnRNP H containing these two RRM motifs (His<sub>6</sub>-hnRNP H RRM 1,2). We confirmed its binding to SLS2-A7 RNA by EMSA, however its affinity was rather low (apparent K<sub>d</sub> in the range of 1.7 μM) (**Suppl. Fig. 1C**). To evaluate the strength of the nucleolin interaction, the full-length native protein was purified from an HeLa cell extract by two successive chromatography steps as described previously (see Materials and Methods).<sup>32</sup> The apparent K<sub>d</sub> established by EMSA (around 200 nM) (**Suppl. Fig. 1D**) was in the range of the one found for pre-ribosomal RNA.<sup>28</sup>

Having demonstrated the direct binding of nucleolin, hnRNP H and hnRNP K to SLS2-A7 RNA, we then decided to delineate more precisely their respective binding sites.

### **Nucleolin has strong binding sites in the apical part of SLS2-A7 RNA**

Previous studies on nucleolin/pre-rRNA interaction revealed that the two N-terminal RRM of nucleolin bind to a stem-loop structure containing the NRE sequence (5'-U/GCCCCGA-3') in its apical part.<sup>28, 33, 34</sup> Inspection of the SLS2-A7 sequence showed the presence of at least three distinct elements showing a strong degree of similarity with NRE sequences. These three motifs denoted 1, 2 and 3 (5'-CCCCGA-3', 5'-ACCCGA-3', 5'-GCCCCGA-3',

respectively) are located between positions 7985 and 8009 in the apical part of SLS2-A7 (**Fig. 3B**). By some rearrangement of the 2D structure of the apical part of SLS2-A7, motif 2 can adopt a 2D structure similar to that defined for efficient binding of nucleolin RRM 1,2 to pre-rRNA.<sup>28</sup> Interestingly, motif 2 overlaps the ESE2 splicing enhancer (**Fig. 3B**).<sup>12</sup> Furthermore, the 19-nt long sequence containing motifs 2 and 3 (positions 7997-8015 in the BRU HIV-1 RNA) shares 14 common nucleotides with the winner sequence of SELEX experiments performed with nucleolin (5'-CCGACAGGCCCGAAGGAAU-3' bold underlined characters represent identical nucleotides).<sup>33</sup>

To further define nucleolin binding sites in SLS2-A7 RNA, we produced a recombinant His<sub>6</sub>-p30 fragment containing the three N-terminal RRMs, which were previously shown to be sufficient to ensure pre-rRNA binding specificity.<sup>35</sup> First, to test the importance of motifs 1, 2 and 3 for nucleolin binding to SLS2-A7 RNA, we eliminated these three motifs by truncation of the entire apical part of the stem-loop structure (variant SLS2-A7  $\Delta$ NCL, **Suppl. Fig. 1A**). EMSA experiments confirmed that this truncation almost abolished nucleolin binding to the RNA: the estimated K<sub>d</sub> for p30 was shifted from about 3  $\mu$ M for SLS2-A7 WT to more than 12  $\mu$ M for SLS2-A7  $\Delta$ NCL (**Suppl. Fig. 1E**). Then, we tested the effect of individual disruption of each motif on p30 binding to SLS2-A7 RNA. As previous studies showed that C to G substitution at the 5' position in the CCCGA sequence led to a significant decrease of nucleolin affinity<sup>33</sup>, we introduced such C to G substitutions in each of the motifs. EMSA experiments performed with protein p30 and each of these three variant RNAs revealed no marked decrease of the protein affinity upon mutation of a single motif (data not shown). This observation suggests that each of the three sites have equivalent capability to bind nucleolin. Due to steric hindrance only one or two of the sites should be occupied on a given RNA molecule, explaining the absence of effect of their individual mutation.

Then, by fingerprinting experiments, we delineated the p30 protein binding sites on the entire SLS1,2,3-A7 HIV-1 RNA region (positions 7903 to 8092 in the HIV-1 BRU RNA) containing all the regulatory elements acting on site A7.<sup>2, 17</sup> Conditions for p30-SLS1,2,3-A7 complex formation and digestion with T1 and T2 RNases are given in Materials and Methods and in the Legend to **Supplementary Figure 2**. Observation of strong protections in the apical part and ESE3 motif in SLS2-A7 and of lower protections in the terminal loops of SLS1-A7 and SLS3-A7 (**Suppl. Fig. 2**) confirmed that the strongest nucleolin binding sites are located in the apical part of SLS2-A7 and showed that the SLS1-A7 and SLS3-A7 stem loops contain nucleolin binding sites with lower affinity (**Suppl. Fig. 2**).

#### **HnRNP H and hnRNP K both bind to the SLS2-A7 apical part**

HnRNP H was already found to bind several GGG sequences in HIV-1 RNA.<sup>36, 37</sup> As the unique GGG sequence present in the SLS2-A7 RNA is located in the apical part of the SLS2-A7 (positions 7991 to 7994) (**Fig. 3D**), we compared the binding of the recombinant hnRNP H RRM1,2 protein to WT and  $\Delta$ NCL truncated SLS2-A7 RNAs by EMSA (**Suppl. Fig. 1C**). As expected a strong reduction of hnRNP H RRM1,2 protein binding was observed upon deletion of the SLS2-A7 apical part, showing that nucleolin and hnRNP H both bind to this apical part.

As mentioned above, hnRNP K protein binds to C-rich motifs and the consensus sequence established by SELEX experiments for this protein is: 5'-AU(C)<sub>3/4</sub> (U/A)<sub>2</sub>-3'.<sup>29</sup> The 35-nt long sequence located downstream from site A7 (positions 7973 to 8007) in SLS2-A7 is highly enriched in C residues (19 C residues out of 35 nts) (**Fig. 3C**). In addition, it contains a 5'-UCCCAA-3' sequence (positions 7979 to 7984) which fits perfectly to the SELEX consensus sequence. Therefore, we produced variant SLS2- $\Delta$ K RNA with a deletion of the sequence fitting perfectly to the SELEX sequence (**Suppl. Fig. 1A**) and we tested by EMSA whether this truncation and the truncation of the apical part of SLS2-A7 RNA (variant SLS2-

A7  $\Delta$ NCL) decrease hnRNP K affinity. Interestingly, deletion of the apical part had a stronger deleterious effect on hnRNP K binding than deletion of the sequence that fits perfectly to the SELEX sequence (**Suppl. Fig. 1A and B**). Hence, based on gel-shift assays, nucleolin, hnRNP H and hnRNP K all bind preferentially to the SLS2-A7 apical part.

### **HnRNP K like p30 binds ISS, ESE2, ESE3 and ESS3**

Then, we identified the various hnRNP K binding sites present in SLS1,2,3-A7 RNA (**Fig. 4**) by footprinting assays. First, we determined that an hnRNP K concentration of 0.4  $\mu$ M for a [hnRNP K]/[RNA] ratio of 33 allowed the complete engagement of SLS1,2,3-A7 RNA in RNP complexes. Therefore, we performed RNase T1 footprinting assay on complexes formed on 5'-end labeled SLS2-A7 RNA using hnRNP K concentrations of 1.18  $\mu$ M and 2.35  $\mu$ M, respectively. In agreement with EMSA data, fractionation of the digestion products on a sequencing gel (**Fig. 4A**) revealed a strong protection of the SLS2-A7 apical part containing the ESE2 sequence. In addition, protections were detected in ESS3, ESE3 and to a lesser extent in ISS (**Fig. 4A and B**). Hence, like p30, hnRNP K binds to ESE2 and also to sites which are known to be recognized by hnRNP A1.<sup>8, 12</sup>

Furthermore, we observed that addition of protein hnRNP A1 at a 0.1  $\mu$ M concentration significantly reinforced hnRNP K binding to SLS2-A7 RNA (**Fig. 5A**). Indeed, when hnRNP A1 was present alone at this low concentration, no SLS2-A7 RNA gel-shift was observed (data not shown). However, when hnRNP A1 was added together with increasing amounts of hnRNP K, the SLS2-A7 RNA gel-shifts observed were markedly reinforced as compared to shifts obtained at the same hnRNP K concentrations in the absence of hnRNP A1 (**Fig. 5A**). Then we performed UV cross-linking assays to test whether addition of hnRNP K also favors binding of hnRNP A1. Uniformly labeled SLS2-A7 RNA was cross-linked in HeLa cell nuclear extract, in the absence or the presence of increasing amounts of hnRNP K (0.6 and 1.2  $\mu$ M). After digestion of SLS2-A7 RNA with RNases T1 and A the free and cross-linked

hnRNP A1 were immunoselected with a specific antibody bound to Protein G-Sepharose beads. The amount of cross-linked hnRNP A1 was evaluated by SDS-PAGE electrophoresis followed by autoradiography, while the total amount of hnRNP A1 bound to the beads was estimated by Western blot analysis. Three independent experiments were performed and as illustrated in **Figure 5B**, addition of hnRNP K reproducibly increased hnRNP A1 binding to SLS2-A7 RNA.

Therefore, we concluded that hnRNP A1 and hnRNP K do not compete each other for binding to SLS2-A7 RNA. They can even stimulate each other **for association with this RNA**.

### **HnRNP K inhibits *in vitro* splicing at site A7**

HnRNP A1 was described as the major inhibitory protein acting at the A7 splicing site.<sup>8, 12, 13,</sup>  
<sup>15</sup> As we demonstrated that hnRNP K, like hnRNP A1, binds to the ISS, ESE3 and ESS3 splicing silencer elements and as we got evidences for a stimulatory effect of hnRNP K on hnRNP A1 binding to SLS2-A7 RNA, we tested whether hnRNP K may contribute to splicing regulation at site A7. To this end, we used the L47-8072 pre-mRNA containing a mini-D4-A7 intron (**Fig. 1**)<sup>8</sup> to perform *in vitro* splicing assays in the absence or the presence of hnRNP K. Splicing was performed in HeLa cell nuclear extract. We used two other pre-mRNAs as controls: the pSma pre-mRNA derived from  $\beta$ -globin pre-mRNA, which is known not to be affected by hnRNP K addition<sup>38</sup> and the 6A-pyI5-7 pre-mRNA<sup>38</sup> known to be spliced more efficiently in the presence of hnRNP K in the nuclear extract. As expected, addition of hnRNP K at a 1  $\mu$ M concentration had no effect on  $\beta$ -globin pre-mRNA (pSma) splicing, while at the same concentration hnRNP K considerably increased 6A-pyI5-7 splicing efficiency (**Fig. 6**). Interestingly, upon addition of hnRNP K at 0.5 and 1  $\mu$ M final concentrations the splicing efficiency of the L47-8072 pre-mRNA was markedly reduced. Three independent splicing assays were performed and the mean values of M/P decreases were of about 25 and 40%, for addition of hnRNP K at 0.5  $\mu$ M and 1  $\mu$ M final concentrations, respectively (**Fig. 6**).

Therefore, protein hnRNP K, one of the new partners of the HIV-1 *tat/rev* exon 3 that we identified by proteomic analysis, likely has a splicing silencer activity at site A7.

### **Overexpression of hnRNP K alters the *in cellulo* pattern of HIV-1 RNA splicing**

Up to now, whereas the effect of hnRNP A1 and hnRNP H expression on HIV-1 RNA splicing has been thoroughly studied<sup>2, 17, 30</sup>, no study on a possible implication of hnRNP K in the regulation of HIV-1 RNA splicing has been made. Having shown a negative effect of hnRNP K on the *in vitro* utilization of site A7, we decided to test whether hnRNP K may have a more general effect on HIV-1 RNA splicing *in cellulo*. To this end, we used plasmid pΔPSP, a pUC18 derivative containing a truncated version of the HIV-1 pNL4-3 cDNA.<sup>7</sup> The truncation is located within the first intron (D1-A1 intron) and, in HeLa cells, the relative yields of utilization of the splicing sites in transcripts from plasmid pΔPSP are similar to those found for the entire HIV-1 transcripts.<sup>2, 7</sup> Co-transfections of HeLa cells with pXJ41 plasmid derivatives that overexpressed 9G8, ASF/SF2, SC35 or SRp40 were previously found to modify the HIV-1 RNA splicing pattern.<sup>39</sup> Therefore, to test for a possible modulation of the efficiencies of HIV-1 splicing sites by hnRNP K, we co-transfected HeLa cells with plasmid pΔPSP and a pXJ41-hnRNP K derivative which over-expressed hnRNP K. Transfections were performed with two different amounts of pXJ41-hnRNP K plasmid derivative (see Materials and Methods). Western blot analysis of total extracts from the transfected cells revealed increases of the hnRNP K concentration by factors of 1.5 and 2, respectively (**Fig. 7A**). Two control experiments were performed, in one of them plasmid pΔPSP was co-transfected with the empty pXJ41 plasmid, in the second one it was co-transfected with the pXJ41-ASF plasmid derivative which over-expresses protein ASF/SF2 (factor of 4.5, **Fig. 7A**). The maturation products of the pΔPSP transcripts from transfected HeLa cells were analyzed by RT-PCR experiments as previously described, 48 h after the transfection.<sup>39</sup>

Fractionation of the cDNAs obtained on 6% polyacrylamide gel revealed strong variations of the relative levels of HIV-1 mRNAs upon over-expression of hnRNP K (**Fig. 7B**). The fractionated cDNAs were identified by reference to previous experiments performed with plasmid p $\Delta$ PSP and plasmid pXJ41 derivatives. A marked decrease of the *nef2*, and *nef3* mRNAs as well as some decrease of the *rev1,2,3* and *4* mRNAs were detected. In contrast, some increase of the *nef3* and *nef5* mRNAs was detected. The experiment was repeated several times and reproducible results were obtained. In spite of the quite lower over-expression of protein hnRNP K as compared to ASF/SF2 (**Fig. 7A**), the observed modification of the splicing pattern was more important for hnRNP K over-expression than for ASF/SF2 over-expression. Hence, we concluded that hnRNP K should be one important regulator of HIV-1 RNA splicing.

Therefore, we decided to test whether the effect observed at the level of the *Nef* mRNAs reflected variations at the protein level. Indeed, evaluation of the relative expression levels of the Nef protein in the transfected cells by Western blot analysis (**Fig. 7A**) revealed a strong decrease of the amount of this protein upon over-expression of protein hnRNP K.

Therefore, we concluded that hnRNP K is a yet unidentified **regulator of HIV-1 alternative splicing that** strongly decreases Nef protein expression as previously found for overexpression of the SC35 and SRp40 SR proteins.<sup>39, 40</sup>

## Discussion

By combining MS2-MBP and tobramycin based RNP-purification approaches with mass spectrometry analysis of the RNP proteins, we showed that a large number of nuclear proteins can associate with the strategic HIV-1 SLS2-A7 RNA region which is involved in regulation of RNA splicing and RNA export<sup>2, 17, 21</sup> and very likely modulation of the Tat protein translation. As expected, one of the direct protein partners identified was hnRNP A1, which is known to be a strong negative regulator of splicing at site A7.<sup>8, 12, 13, 15</sup> In addition to hnRNP A1, we found that three other hnRNP proteins (hnRNP H, hnRNP K and nucleolin) bind directly to SLS2-A7 RNA. Interestingly the RNA binding sites showing the higher affinity for each of these three proteins are all located in the apical part of the SLS2-A7 structure suggesting that these proteins might compete for binding. Interestingly, the presence of nucleolin is explained by a cluster of three successive canonic NRE sequences in the SLS2-A7 apical part. In addition to our experimental work on nucleolin, hnRNP H and hnRNP K that bind directly to SLS2-A7 RNA, we made a deep analysis of the present knowledge on several of the other identified proteins, especially the ones, common to the RNPs purified by the MS2-MBP and tobramycin based affinity chromatography methods. In most cases their presence can be explained by their possible interactions with nucleolin **or with** hnRNP K. Here, we will give an overview of the possible links between our experimental data and data in the literature. **Detailed information on the various proteins partners of SLS2-A7 RNA identified by proteomic analysis and their possible involvement in HIV-1 biology are given in the Supplementary Tables 1 and 2 as well as in the scheme represented in Figure 8.**

### **Canonical NREs are present in the apical part of SLS2-A7 RNA**

Mass spectrometry analysis unambiguously identified nucleolin in one of the intense band in the SDS-PAGE analysis of SLS2-A7-MS2 RNPs. Nucleolin has a large general affinity for

RNA. Two types of RNA motifs showing a high affinity for nucleolin (NRE for Nucleolin Recognition Element and ECM for Evolutionary Conserved Motif) were previously identified.<sup>34</sup> Nucleolin contains four RRM domains. All four domains are required for binding to the ECM sequence (5'-UAGAUCGAUGU-3'), while the two N-terminal ones are sufficient to ensure the specific recruitment of NRE RNA motifs.<sup>41</sup> Protein p30 contains the three N-terminal RRMs of nucleolin. Two of the three p30 binding sites that we found in the apical part of SLS2 fit perfectly to the NRE consensus sequence (5'-U/GCCCGA-3'). The third one only differs at the 5' position by the presence of an A residue. In the 3D structure which has been established for the nucleolin-NRE complex, the NRE is located in the terminal loop of a stem-loop structure and the RNA loop is sandwiched between the two RRMs and their linker.<sup>42</sup> Limited energy should be required to convert the SLS2 apical structure as represented in **Figure 2A** into a perfect NRE structure.<sup>35</sup> This is as much feasible as nucleolin is known to have an RNA chaperone activity.<sup>28</sup> The strong concentration of NRE elements in the apical part of SLS2-A7 and the fact that only a very limited number of NRE-like motifs are found in the entire HIV-1 RNA molecule (data not shown), are hints for a possible important function of nucleolin binding to the SLS2-A7 segment.

Our observation that p30 also binds to ESE3 and ESS3 and to a lesser extent to ISS may be explained by the presence of UAG sequences in these splicing regulatory elements. In particular, ESS3 includes a UUAG sequence and nucleolin was found to associate with UUAG sequences in cellular mRNAs.<sup>43</sup> Since hnRNP A1 also binds ISS, ESE3 and ESS3<sup>8, 12, 15</sup>, one or the other protein may be bound to the regulatory elements, depending on the relative cellular concentrations of these two proteins. This competition for the same sites may modulate splicing efficiency at site A7. Both hnRNP A1 and nucleolin are highly concentrated in HeLa nuclear extract, this may explain why the addition of nucleolin in this extract had no marked effect on *in vitro* splicing efficiency at site A7 (data not shown).

However, as previously described <sup>18</sup>, we found that L47-8072 RNA with a deletion of the apical part of SLS2-A7 stem-loop structure has a decreased *in vitro* splicing efficiency (data not shown), which is in agreement with a possible negative effect of nucleolin on hnRNP A1 binding at the A7 region.

### **Binding of nucleolin to the SLS2-A7 apical part may serve other functions (Figure 8 and Supplementary Tables 1 and 2)**

Nucleolin was already found to be involved in several aspects of the HIV-1 multiplication cycle (Suppl. Table 1). The new nucleolin binding sites that we discovered in the SLS2-A7 segment may play a role in some of these already established functions or be involved in other yet undiscovered functions. Although predominantly localized in the nucleolus, nucleolin is a shuttling protein. It is also found in the cytoplasm, at the plasma membrane <sup>27, 43</sup> and at the cell surface where it may serve as a facilitator of the HIV-1 particle internalization. <sup>44</sup>

In more direct link with the present data, the presence of HIV-1 RNA in the nucleolus of infected cells was detected a long time ago <sup>45</sup> and the recent use of a nucleolar-localized ribozyme directed against HIV-1 RNA was found to inhibit HIV-1 replication. <sup>46</sup> Transient nucleolar localization of HIV-1 RNAs might also allow their post-transcriptional modification by snoRNAs. The presently identified nucleolin binding sites may be involved in the transient localization of HIV-1 RNA in the nucleolus. Note that Tat and Rev proteins also have transient localization in nucleolus. <sup>46</sup>

Upon HIV-1 infection, part of the nucleolar nucleolin seems to be relocated to the cytoplasm <sup>44</sup>, and may participate together with the viral Gag protein to virion production. <sup>47</sup> Both Gag and nucleolin are supposed to bind to the stem-loop structure containing Psi. <sup>48</sup> and are incorporated together into HIV-1 virions. <sup>47, 48</sup> Consequently, nucleolin was proposed to be

involved in virion assembly and budding.<sup>48</sup> Therefore, we can ask the question, whether the new nucleolin binding sites that we identified may play a role in these processes.

As all HIV-1 mRNAs contain the SLS2-A7 segment, through its binding to this segment, nucleolin might regulate HIV-1 mRNA stability and translation, as it is doing for several cellular mRNAs (*ie.* bcl-2 mRNA, IL2 mRNA, beta-globin mRNA).<sup>27</sup> In line with this hypothesis proteins Ybx1 and hnRNP U found in the purified SLS2-A7 RNPs were already detected together with nucleolin in protein complexes which stabilize cellular mRNAs.<sup>49, 50</sup> Finally, the binding of nucleolin to SLS2-A7 RNA may explain the presence of PARP1 and ILF3 (NF90)<sup>51</sup>, as well as the RNA helicase p68 (DDX5), a regulator of transcription and splicing.<sup>52</sup>

### **The synergistic binding of hnRNP A1 and hnRNP K to SLS2-A7 RNA may reinforce splicing inhibition at site A7**

HnRNP K, which is present both in the nucleus and the cytoplasm of human cells, is a prototype protein containing RNA binding motifs of the KH type. It contains 3 KH domains, two in the N-terminal domain and one in the C-terminal domain and a three prong interaction model where each KH domain binds one of the C-rich boxes was proposed.<sup>29</sup> The strong affinity of hnRNP K for the apical part of SLS2-A7 that contains C rich stretches (**Fig. 4**) fits to this model. However, protection of ISS, ESE3 and ESS3 elements by hnRNP K in RNase protection assays also suggests that hnRNP K binds to a large variety of RNA targets.

A two-hybrid screen for possible interactions between hnRNP proteins revealed possible direct interactions of hnRNP K with hnRNP E2, I and L. But, no interaction was detected between hnRNP A1 and hnRNP K by this approach.<sup>53</sup> However, interestingly, here we found that the addition of limited amounts of hnRNP A1 favors hnRNP K binding to SLS2-A7 RNA and *vice versa*. Therefore, not only hnRNP K does not compete hnRNP A1 binding to the

three **splicing silencers** acting at site A7 (ISS, ESE3 and ESS3), but it seems to favor its association. This may explain the negative effect **of hnRNP K on *in vitro* splicing at site A7.** In addition, **when using** plasmid pΔPSP, an increase of the hnRNP K level by a factor of 2 **in HeLa cells** was sufficient to affect profoundly the relative levels **of the *nef* and *rev* mRNAs** and to decrease strongly Nef protein production. Interestingly, a slight over-expression **of hnRNP K** had an effect on Nef expression very similar to that previously observed for over-expression of the SC35 and SRp40 SR proteins.<sup>39, 40</sup> Accordingly, hnRNP K was first described as an activator of splicing which stimulates the utilization of tropomyosin exon 6A.<sup>38</sup> **In contrast, hnRNP K has an inhibitory role at site A7, which is in agreement with several other recent data showing splicing inhibitory properties of protein hnRNP K.**<sup>54, 55</sup> It is therefore not surprising that HIV-1 uses hnRNP K to regulate the complex alternative splicing of its RNA. Protein hnRNP K likely has enhancer or silencer activity depending on the identity of the HIV-1 splicing site. Further, deep studies on the effect of hnRNP K on the various HIV-1 splicing sites will be required to verify this hypothesis.

**Binding of hnRNP K to SLS2-A7 RNA may have additional roles (Figure 8, Suppl. Tables 1 and 2)**

HnRNP K is a multifunctional protein known to be involved in the regulation of transcription, splicing, translation, nuclear transport and signal transduction. It acts as a hub at the intersection of signalling pathways and processes involving nucleic acids.<sup>29</sup> Through its numerous domains it recruits several kinases, as well as transcription, splicing and translation factors. In particular, hnRNP K was shown to interact with Nef and nucleate the Nef-interacting kinases.<sup>56</sup> HnRNP K is also known to interact specifically with Sam68<sup>57</sup> which is a synergistic activator of the Rev function.<sup>58</sup> When over-expressed, Sam68 can even functionally substitute for Rev. HnRNP K does not bind to the RRE motif directly<sup>57</sup>, but can

associate with Sam68 bound to the RRE, and through this interaction limits the Sam68 RRE dependent RNA transport activity (Fig. 8).<sup>57</sup> As the SLS2-A7 stem-loop is expected to be rather close to the RRE element (these two HIV-1 RNA motifs are separated by 312 nts) (Fig. 8), one can imagine that when bound to the SLS1,2,3-A7 RNA region, hnRNP K may interact with Sam68 bound to the RRE and may therefore modulate its positive action on the Rev mediated RNA transport. This may in part explain why mutations in ESE3 were found to influence Rev activity.<sup>59</sup>

Like nucleolin, and sometimes together with nucleolin, hnRNP K increases the stability of cellular mRNAs.<sup>29</sup> It may therefore stabilize HIV-1 mRNAs containing the SLS2-A7 sequence. HnRNP K was also found to mediate translational silencing of cellular and viral mRNAs.<sup>29</sup> Furthermore, the capability of hnRNP K to recruit numerous cellular partners may explain the presence of several of the proteins detected in the purified SLS2-A7 RNPs (Fig. 8), for instance, RNA helicase DDX5 and PSF were already detected in the hnRNP K interactome<sup>60</sup>, DDX1 binds directly to both hnRNP K<sup>61</sup> and its binding to Rev is implicated in the Rev activity.<sup>62</sup>

Interestingly also, the DNA-dependent protein kinase (DNA-PK) found in the purified RNPs was shown to phosphorylate hnRNP proteins (C, K, H and F) as well as the ATP-dependent RNA helicase A (RHA/DHX9) also present in the RNPs,<sup>63</sup> which may explain our observation of the three subunits of DNA-PK (the catalytic subunit PRKDC and its two associated ATP-dependent DNA helicases 1 and 2, also designated as ku70/XRCC6 and Ku86/XRCC5)<sup>63</sup> and perhaps also of RHA in the purified RNPs. Interestingly, the Ku70-Ku86 complex is involved in several steps of HIV-1 biology (transcription and DNA integration) and RHA is required to release RRE containing mRNA from spliceosomes before completion of splicing. RHA only binds weakly to RRE.<sup>64</sup> Noticeably also, activators of the

DNA-PK complex (the Interleukin enhancer-binding factors 2 and 3 also designated as factors NF45 and NF90, respectively)<sup>65</sup> were also present in the purified RNP.

## **Conclusions**

There are several physical or functional links between the proteins that we found associated to SLS2-A7 RNA (**Fig. 8**), which probably explain the large number of protein partners of the small SLS2-A7 RNA region. Nucleolin and hnRNP K, the direct RNA binders, likely play a key role in the recruitment of most of these proteins (**Fig. 8**). Since several of the identified proteins are known to have a functional role in HIV-1 biology, especially in the modulation of the Rev activity, we propose that at least some parts of the network of protein interactions which can be established through nucleolin and hnRNP K association with the SLS2-A7 stem-loop structure (**Fig. 8**) may have a functional importance for HIV-1 biology.

Based on our discovery that hnRNP K modulates the utilization of HIV-1 splicing sites other than site A7, this protein and some of its partners probably also associate with other HIV-1 RNA region.

## Materials and Methods

### Plasmids used and DNA constructions.

*Constructs used for RNA production:* The previously described L47-8490 construct<sup>8</sup> was used for PCR amplifications of HIV-1 BRU genomic DNA fragments. To produce the DNA matrix for SLS2-A7 RNA transcription (**Fig. 1**), a HIV-1 DNA fragment (positions 7965 to 8046 in the HIV-1/BRU genome) was PCR-amplified using an upstream primer containing an SP6 promoter (O-3100 and O-3104 pair, **Table S1**) and it was inserted between the HindIII and XbaI restriction sites of plasmid pUC18 (resulting plasmid pUC18-SP6-SLS2-A7). Then, to produce SLS2-A7-MS2 RNA (**Fig. 1**) for affinity chromatography, the three MS2 coat protein binding sequences from plasmid pAdML-M3<sup>25</sup> were inserted between the BamHI and EcoRI restriction sites of plasmid pUC18-SP6-SLS2-A7 (plasmid pUC18-SP6-SLS2-A7-MS2). The pT7-MS2 construct containing three MS2 coat protein binding sequences from pAdML-M3 was used to produce the control MS2 RNA.<sup>66</sup> Similarly to produce the SLS2-A7-J6f1 RNA for affinity chromatography, the J6f1 tobramycin aptamer DNA sequence<sup>23</sup> was inserted in plasmid pUC18-SP6-SLS2-A7 (addition made at the 3' end of SLS2-A7 RNA, plasmid pUC18-SP6-SLS2-A7-J6f1). J6f1 DNA was also inserted at the BamHI restriction site of plasmid pUC18 (plasmid J6f1). To produce the variant SLS2-A7  $\Delta$ K and SLS2-A7  $\Delta$ NCL RNAs, plasmid pUC18-SP6-SLS2-A7 was subjected to site-directed mutagenesis using the two overlapping oligonucleotide pairs described in **Table S1**. As previously described, the L47-8490 construct was used for PCR amplification of the A7 SLS1,2,3 DNA for *in vitro* transcription and also to transcribe the L47-8072 pre-mRNA (**Fig. 1**) for *in vitro* splicing assays.<sup>8</sup> Plasmids designated for transcription of the pSma pre-mRNA<sup>38</sup> and the 6A-pyI5-7 pre-mRNA.<sup>38</sup> were used for *in vitro* splicing control experiments. The previously described p $\Delta$ PSP plasmid<sup>7</sup> was used for splicing assays in HeLa cells.

*Constructions used for protein productions.* The MS2-MBP plasmid <sup>25</sup> was used to produce the MS2-MBP protein used for affinity purification. The pET15b-p30 construct used for expression of human nucleolin RRM1,2,3 (RNA Recognition Motif) was kindly provided by P. Bouvet.<sup>35</sup> Plasmid pET15b-hnRNP H RRM1,2 expressing the N-terminal part of hnRNP H was built by H Hallay, by insertion of stop codons in the previously described pET15b-hnRNP H plasmid.<sup>67, 68</sup> To produce the recombinant hnRNP K protein, the full-length hnRNP K cDNA from the RZPD collection (<http://www.rzpd.de>) was PCR-amplified (primers O-4695 and O-4696, **Table S1**), subcloned into plasmid pCR2.1 (Invitrogen, USA) and transferred in between the NheI and SalI restriction sites of the pET28b expression vector. To over-express protein hnRNP K in HeLa cells, the full-length hnRNP K cDNA from the RZPD collection was cleaved by EcoRI and XhoI nucleases and inserted between the corresponding sites in plasmid pXJ41.<sup>69</sup> The recombinant pXJ41-ASF plasmid <sup>69</sup> was used for over-expression of protein ASF/SF2 and plasmid pXJ41 as a control.

**Antibodies used and Western blot analysis.** The following antibodies were used in this study: a polyclonal antibody directed against nucleolin (generous gift from S. Muller), a polyclonal anti-hnRNP K/J antibody (ImmuQuest, 3C2), a monoclonal 1G11 antibody directed against hnRNP H/H'/F, a monoclonal 4B10 antibody for hnRNP A1 (generous gift from G. Dreyfuss), a monoclonal anti-U2AF<sup>65</sup> antibody (generous gift from J. Valcarcel), a monoclonal SAP.4G5 anti- $\beta$  tubulin antibody and the MaTG0020 antibody directed against the Nef protein (generous gift from AM Aubertin). Conditions for Western blot analysis were as previously described.<sup>39</sup>

***In vitro* transcription.** For large scale *in vitro* transcription of non-radiolabeled SLS2-A7-MS2 RNA and SLS2-A7-J6f1 RNAs, plasmids pUC18-SP6-SLS2-A7-MS2 and pUC18-SP6-SLS2-J6f1 (10 µg) were linearized by the EcoRI and SmaI nucleases, respectively and then transcribed with the SP6 RNA polymerase. Similarly, for large scale transcription of the MS2 and J6f1 control RNAs, plasmids pUC18-MS2 and pUC18-J6f1 (10 µg) were linearized by the EcoRI and SmaI nucleases, respectively, and then transcribed with the SP6 and T7 RNA polymerases, respectively.

Production of non-radiolabeled and radiolabeled WT and variant SLS2-A7 RNAs, plasmid pUC18-SP6-SLS2-A7 and its variants were digested with BamHI and transcribed with the SP6 RNA polymerase in the absence or the presence of [ $\alpha$ -<sup>32</sup>P]GTP or [ $\alpha$ -<sup>32</sup>P]CTP as previously described.<sup>7, 8</sup> The SLS1,2,3-A7 RNA used for footprinting assays was transcribed by T7 RNA polymerase using a PCR-amplified DNA fragment produced using oligonucleotides O-2065 and O-2382 (**Table S1**) according to Marchand et al.<sup>8</sup> Uniformly [<sup>32</sup>P]-labeled L47-8072 pre-mRNA was produced by *in vitro* transcription of pL47-8072 by T7 RNA polymerase in the presence of [ $\alpha$ -<sup>32</sup>P]UTP.<sup>8</sup> The same procedure was used for *in vitro* transcription of uniformly labeled pSma and 6A-pyI5-7 pre-mRNAs. All transcripts were dissolved in milliQ water and their concentrations were estimated by OD260 nm measurement.

**MS2 RNP affinity chromatography.** The MS2 affinity procedure used for RNP purification was adapted from Deckert et al.<sup>25</sup> and performed according to Maenner et al.<sup>66</sup> To form RNA-MS2-MBP protein complex, about 500 pmol of MS2 or SLS2-A7-MS2 RNAs were denatured by a 10 min incubation at 65 °C and renatured by a slow cooling to room temperature before incubation at 4°C for 15 min with a 5 fold molar excess of purified MS2-MBP fusion protein prepared according to Deckert et al.<sup>25</sup> In parallel, amylose beads (50 µl of 50% slurry, GE

Healthcare, France) were washed three times with splicing Buffer D [20 mM Hepes-KOH, pH 7.9, 100 mM KCl, 3 mM MgCl<sub>2</sub>, 0.2 mM EDTA pH 8.0, 0.5 mM DTT, 0.5 mM PMSF, 20% (vol/vol) glycerol] and were then incubated with the RNA-MS2-MBP complexes for 1 h at 4°C using head-over-tail (HOT) rotation to avoid sedimentation of the matrix. The beads were then washed three times (500 µl each) with splicing Buffer D.

Five hundred µl of HeLa cell nuclear extract (CilBiotech, Mons, Belgium) (about 5 mg of total protein) were dialyzed against buffer D and total yeast tRNA was added at a 5 µM final concentration, before incubation with the amylose beads preloaded with RNA-MS2-MBP complexes. Incubation was performed for 15 min at 4°C with HOT rotation. After three extensive washing steps (30 min) with 500 µl of Buffer D, elution of RNP complexes was performed with 100 µl of Buffer D containing 10 mM maltose for 15 min at room temperature. About 80% of the total amount of eluted fraction was fractionated by 10% SDS - PAGE and stained with a colloidal Coomassie Blue solution. After destaining with distilled water, the visible protein bands were excised from the gel and used for proteomic analysis (see below). The remaining part of the eluate was used for Western blot analysis using specific antibodies (see section Antibodies). Western blots were revealed using suitable HRP-conjugated secondary antibodies and ECL Plus™ from GE Healthcare.

**Tobramycin affinity chromatography.** The procedure used is derived from the method previously described for purification of splicing complexes.<sup>23</sup> Briefly, *N*-hydroxysuccinimide-activated Sepharose 4 Fast Flow (GE Healthcare) was coupled with 5 mM tobramycin.<sup>23</sup> The affinity chromatography was performed at 4°C, as previously described except minor

changes. The BP buffer was used at each of the steps and a 4BP solution (80 mM Tris-HCl, pH 9.1, 4 mM CaCl<sub>2</sub>, 4 mM MgCl<sub>2</sub>, 2 mM DTT) was freshly prepared at the beginning of the experiment. First, 25 µl of tobramycin beads suspension were blocked by addition of 400 µl of blocking buffer (1xBP containing 150 mM KCl, 0.1 mg/ml tRNA, 0.5 mg/ml BSA, 0.01% Nonidet P-40). Incubation was performed overnight with HOT shaking. Then, 100 µl of a mixture containing 500 pmol of tobramycin-tagged RNA and 40 µg of tRNA (1xBP buffer containing 145 mM KCl), were added to 25 µl of matrix. Incubation was performed for 2 h at 4°C. The matrix was then washed three times (750 µl each) with binding buffer (1xBP buffer containing 0.1% Nonidet P-40). Then, 500 µl of HeLa cell nuclear extract dialysed against buffer D adjusted at pH 9.1 (20 mM Hepes-KOH pH 9.1, 100 mM KCl, 3 mM MgCl<sub>2</sub>, 0.2 mM EDTA pH 8.0, 0.5 mM DTT, 0.5 mM PMSF, 20% (vol/vol) glycerol) were added per 25 µl matrix and incubation was for 15 min at 4°C with constant agitation. At the end of the incubation, the matrix was collected by centrifugation and washed three times with 750 µl of washing buffer (1x BP containing 75 mM KCl and 0.1% Nonidet P-40). Finally, RNP complexes were eluted by incubating with 80 µl of elution buffer (1x BP buffer containing 5 mM tobramycin, 75 mM KCl, 2 mM MgCl<sub>2</sub>) with HOT rotation for 10 min at room temperature. Typically, 70% of the bound complexes could be eluted. Proteins were fractionated by 10% SDS-PAGE.

**In-gel digestion.** In-gel digestion was performed with an automated protein digestion system, a MassPrep Station (Waters, Manchester, UK). The gel plugs were washed twice with 50 µl of 25 mM ammonium hydrogen carbonate (NH<sub>4</sub>HCO<sub>3</sub>) and 50 µl of acetonitrile. The cysteine residues were reduced by addition of 50 µl of 10 mM dithiothreitol at 57°C and alkylated by addition of 50 µl of 55 mM iodoacetamide. After dehydration with acetonitrile, the proteins

were cleaved in-gel with a 12.5 ng/μl solution of modified porcine trypsin (Promega, Madison, WI, USA) in 25 mM NH<sub>4</sub>HCO<sub>3</sub> (~ 30 μl). The digestion was performed overnight at room temperature. Tryptic peptides were extracted twice: the first time with 60% acetonitrile in 5% formic acid for 1 hour, and the second time with a 100% acetonitrile solution until the gel pieces were dehydrated. The collected extracts were pooled to a final volume of 60 μl. Excess of acetonitrile was evaporated at 37°C before analysis.

**NanoLC-MS/MS analysis.** Nanocapillary liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) analyses were performed using a CapLC system (Waters, Manchester, UK), coupled to a hybrid Electrospray Quadrupole Time-Of-Flight mass spectrometer (Q-TOF II, Waters, Manchester, UK). A 6.4 μl aliquot of each sample was loaded on a precolumn, before chromatographic separation on a C<sub>18</sub> column (LC packings C<sub>18</sub>, 75 μm i.d., 150 mm length). Separation was performed at 200 nL/min flow rate using a 35 min gradient from 5 to 45% B (eluent A, 0.1% (v/v) formic acid in water and eluent B, 0.1% (v/v) formic acid in acetonitrile). The ESI capillary voltage was set to 3500 V and cone voltage to 35 V. Mass calibration of the TOF was achieved using phosphoric acid (H<sub>3</sub>PO<sub>4</sub>) on the m/z 50-2000 range.<sup>70, 71</sup> The MS survey scan was acquired on the range m/z 300-1500 with a scan time of 1 s and an interscan time of 0.1 s. Automatic switching between MS and MS/MS modes was used. The 3 most intense ions (intensity threshold of 10 counts) were selected on each MS spectrum for further isolation and CID fragmentation. Fragmentation was performed using argon as collision gas and with a collision energy profile optimized for various mass ranges of precursor ions. The scan range for MS/MS acquisition was from m/z 50 to 2000 with a scan time of 1 s and an interscan time of 0.1 s.

Data acquisition was piloted by MassLynx software V4.0 SP1. Mass data collected during a nanoLC-MS/MS analysis were automatically processed and converted into a .pkl using MassLynx software V4.0 SP1. Normal background subtraction type was used for both MS

and MS/MS with 5% threshold, a polynomial correction of order 5 and de-isotoping. Peak list files were submitted to Mascot search engine (Matrix Science, London, UK) installed in its 2.2.04 version on a local server. Searches were performed against a composite target-decoy database including both forward (target) and reversed (decoy) SwissProt protein sequences (21 January 2010, 514212 total entries). Searches were performed with a tolerance on mass measurement of 0.3 Da for both precursor and fragment ions. Carbamidomethylation of cystein residues, oxidation of methionine residues and acetylation of protein N-term residues were searched as variable modifications. Up to one missed cleavage was allowed.

Scaffold V2.02.00.pre36 (Proteome Science, Portland, USA) was used for identification validation and false positive rate estimation for protein identification. For each sample, peptides were filtered out according to the cutoff set for proteins hits with 2 or more peptides (Mascot Ion Score > 20 and Mascot Ion Score - Mascot Identity Score > 7) and no false positive protein was identified. Further data manipulations were performed with Microsoft Excel.

**Production and purification of proteins.** Production and purification of the MS2-MBP protein was according to Deckert et al.<sup>25</sup> Purification of nucleolin from HeLa cell nuclear extract (Cilbiotech, Mons, Belgium) was done according to Belenguer et al.<sup>32</sup> All the recombinant proteins (nucleolin RRM1,2,3 or p30, hnRNP H RRM1,2 and full-length hnRNP K) were His<sub>6</sub>-tagged and expressed in *E. coli* BL21 (DE3) cells grown in LB/Amp (for p30 and hnRNP H RRM1,2) or LB/Kan (for full-length hnRNP K). IPTG induction at 1 mM was made at an A<sub>600</sub> of 0.6. After 5 h of growth at 37°C for hnRNP K, 4 h of growth at 37°C for hnRNP H and 5 h of growth at 30°C for p30, cells were sonicated in lysis buffer [20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM DTT and Protease inhibitor cocktail (Roche)]. Affinity purification was done using Ni-NTA affinity matrix (Ni-NTA agarose, Qiagen). The bound proteins were eluted with 200 mM imidazole. For hnRNP K, hnRNP H and p30, the eluted

fractions were directly applied onto a 5 ml Heparin-Sepharose column and the purified proteins were eluted with a linear KCl gradient 0-1 M. Recombinant proteins were dialyzed against buffer D and stored at -80°C.

**Electrophoretic Mobility Shift Assays (EMSA).** EMSA were performed with 0.5 fmoles of uniformly [<sup>32</sup>P]-labeled transcript and 0.6 pmol of the same non-radiolabeled RNA in the presence of 2 µg of total yeast tRNA. RNP complexes were formed by addition of the one or two recombinant proteins at increasing concentrations, followed by 15 min incubation at 4°C as previously described.<sup>8</sup> Complex formation was analyzed by electrophoresis on 5% non-denaturing polyacrylamide gels as described previously.<sup>8</sup>

**UV cross-linking and immunoselections.** For UV cross-linking assays, complexes were formed by incubation of 50 fmoles of [<sup>32</sup>P]-radiolabeled RNA with 5 µl (50 µg of total protein) of HeLa cell nuclear extract for 15 min at 4°C. Formation of RNP complexes was verified by EMSA. Conditions for UV irradiation at 254 nm were as previously described.<sup>7</sup> The RNA was digested by RNase T1 (50 U) for 1 h at 37°C followed by a 30 min incubation at 50°C. Proteins were fractionated by 10% SDS-PAGE and the UV cross-linked proteins were detected by autoradiography. When required, an additional immunoselection step was performed on cross-linked proteins using specific antibodies bound to Protein G-sepharose beads (see section Antibodies). Immunoselection was performed as previously described.<sup>7</sup> The immunoselected proteins were fractionated by 10% SDS-PAGE and the presence of cross-linked proteins was detected by autoradiography. When necessary, the amount of labeled nucleotides cross-linked with the protein was quantified with a Typhoon (9410) Healthcare scanner.

**Enzymatic probing of RNP-complexes.** Probing of p30-SLS1,2,3-A7 RNP complex was done using 100 ng of non-labeled SLS1,2,3 RNA (1.2 pmol) mixed with 5 µg of total yeast

tRNA in a 15  $\mu$ l total volume containing 4  $\mu$ l of buffer D. The RNA was denatured for 10 min at 65°C and refolded by slow cooling to room temperature. Then, recombinant p30 protein was added at a 8.75  $\mu$ M or 18.25  $\mu$ M concentration and RNP complexes were formed by incubation for 10 min at 30°C. They were then digested with RNase T1 (0.1 U) or RNase T2 (0.6 U) by a further 10 min incubation at the same temperature. Reactions were stopped by addition of 20  $\mu$ g of total yeast tRNA and cooling to 0°C followed by phenol extraction. Enzymatic digestions of naked RNA were performed under the same conditions, except that the protein solution was replaced by 5  $\mu$ l of buffer D. Positions of enzymatic cleavages were identified by extension of the 5'-end [<sup>32</sup>P]-labeled O-2382 primer (see **Table S1**) with avian myeloblastosis virus (AMV) reverse transcriptase (QBiogene) in conditions described previously.<sup>7</sup> After reverse transcription, cDNA products were fractionated on a 7% polyacrylamide/8M urea sequencing gel.

Probing of hnRNP K-SLS1,2,3-A7 RNP complex was done using 0.5 pmol of 5'-end [<sup>32</sup>P]-labeled RNA and 2  $\mu$ g of total yeast tRNA. The RNA was denatured and refolded as described above. Then, the recombinant hnRNP K protein was added at a 1.1  $\mu$ M or 2.15  $\mu$ M concentration. Digestion with RNase T1 (0.5 U) was performed under the same conditions as above and the resulting products were fractionated on a 15% polyacrylamide/8M urea sequencing gel. A reference ladder was obtained by total hydrolysis of naked labeled RNA (0.5 pmol) with 50 mM sodium bicarbonate (pH 9.2) for 2 min at 95°C.

***In vitro* splicing assays.** *In vitro* splicing assays were performed with HeLa cell nuclear extracts using 100,000 cpm of uniformly [<sup>32</sup>P]-labeled pre-mRNA per assay, as previously described.<sup>72</sup> Splicing products were fractionated by electrophoresis on a 5% polyacrylamide/8M urea gels. *In vitro* splicing experiments were repeated three times using different preparations of pre-mRNA transcript. Splicing efficiencies (M/P, amount of mature mRNA *versus* residual pre-mRNA) were established by quantifying the radioactivity in the

bands of gel corresponding to the spliced and unspliced products, respectively. Quantification was achieved using a Typhoon 9410 PhosphorImager and the ImageQuant Software Version 5.2 (1999) Molecular Dynamics and taking into account the number of U residues per molecule.

***In cellulo* splicing assays.** Conditions used were according to Ropers et al.<sup>39</sup> By using the jetPEI™ transfection reagent (Polyplus transfection™) protocol, HeLa cells were co-transfected with 1 µg of ΔPSP plasmid together with 1 or 2 µg of pXJ41-hnRNP K plasmid overexpressing hnRNP K protein or with 250 ng of pXJ41-ASF plasmid overexpressing ASF/SF2. In a control assay, HeLa cells were co-transformed with plasmid pΔPSP and the empty pXJ41 plasmid. Total cellular RNAs were isolated from the transfected HeLa cells 48 h after transfection using the TRIzol® Reagent (Invitrogen™), and were treated by 2 U of Turbo™ DNase (Ambion) as previously described.<sup>39</sup> Three µg of RNA were reverse-transcribed by the Moloney murine leukemia virus-reverse transcriptase according to the supplier's specifications (Invitrogen), using the previously described<sup>7</sup> forward oligonucleotide 2143 and reverse primer 2144 (see **Table S1** in Supplementary data). This pair of oligonucleotides selectively amplified the 1.8-kb HIV size class mRNAs. Amplification was limited to 28 cycles, in order to detect both major and minor bands. DNA products were fractionated by electrophoresis on a 6% polyacrylamide nondenaturing gel and the GeneGenius Gel Imaging System (Syngene) was used to scan the gel after ethidium bromide staining. The fractionated cDNAs were identified by reference to previous data on plasmid ΔPSP.<sup>7,36,39</sup>

The level of protein expression in the transfected cells was estimated by Western blot analysis of total cell extracts using the monoclonal antibodies directed against hnRNP K, ASF/SF2 and Nef proteins (see section antibodies).

## Figure legends

**Figure 1:** Schematic representation of HIV-1 genome organization and the RNAs used in this study

The HIV-1 genome is shown at the top of the figure. The D1 to D4 5' ss and the A1 to A7 3' ss are indicated. Boxes represent the different open reading frames and the abbreviated names of the corresponding proteins are given. The various RNA transcripts used in this study are represented below. Their names and utilizations are indicated on the left side of the figure. Numbering is that used for HIV-1/BRU RNA sequences.<sup>17</sup> Exon sequences are represented by rectangles, introns by thin lines. The exonic *cis*-regulatory elements acting at site A7 (ESE2, ESE3, ESS3 a and b)<sup>17</sup> are symbolized by dark grey, grey and white vertical rectangles, respectively), the intronic ISS element is represented by a white oval.

**Figure 2:** Structure of the SLS2-A7-MS2 RNA used for affinity chromatography of the RNPs formed in HeLa cell nuclear extract and analysis of their protein content

(A) Secondary structure of the SLS2-A7-MS2 RNA used for affinity chromatography. The 5' part of the RNA folds into the SLS2-A7 stem-loop structure. The three binding sites for the MS2 coat protein at the 3' extremity of the molecule (shaded sequence) can form three small stem-loops structures. The A7 3' ss is indicated by an arrow. ESE2 and ESE3 *cis*-regulatory

elements are delineated by broken arrows. Additional nucleotides at the 5'-end of SLS2-A7-MS2 RNA (grey characters) were included to improve SP6 RNA polymerase transcription. The MS2 RNA used as a control in the proteomic analysis (Panel A) contained the MS2 sequences present in SLS2-A7-MS2 RNA (Panel B) plus an 8 nucleotide sequence at its 5'-end (GGGUCGAC) arising from the restriction sites used for cloning.

(B) SDS PAGE analysis of proteins contained in the affinity purified SLS2-A7-MS2 (Lane SLS2-A7-MS2) and MS2 RNP complexes (Lane MS2). Proteins from the eluted RNP complexes were separated by 10% SDS-PAGE and stained with colloidal Coomassie Blue. Protein markers were fractionated in the two lanes on the left of the gel and their molecular weights in kDa are given. In the MS2 and SLS2-A7-MS2 lanes, positions of the MS2-MBP protein used in the purification process and of the proteins found to interact directly with SLS2-A7 RNA are indicated on the right side of the gel.

(C) Western blot analyses showing the specific interaction of nucleolin, hnRNP K and hnRNP H with SLS2-A7 RNA. The RNP complexes formed by the MS2 and SLS2-A7-MS2 RNAs were eluted under native conditions and subjected to 10% SDS-PAGE followed by Western blotting using specific antibodies directed against nucleolin, hnRNP K/J, H/H'/F, and A1 proteins.

**Figure 3:** Demonstration of the direct interaction of nucleolin, hnRNP A1, H and K with SLS2 A7 RNA by UV cross-linking experiments and prediction of their putative binding sites.

(A) UV cross-linking of proteins from a HeLa cell nuclear extract to SLS2-A7 RNA. The uniformly [<sup>32</sup>P]-labeled RNA was incubated in HeLa nuclear extract and exposed to UV light at 254 nm as described in Materials and Methods. After RNase T1 digestion, one part of the material was directly fractionated by 10% SDS-PAGE (Lane 1), while the other part was divided into aliquots that were immunoselected with different antibodies before fractionation

by 10% SDS-PAGE. The identities of the proteins targeted by the antibodies are indicated above the lanes: hnRNP A1, hnRNP K/J, nucleolin and hnRNP H/H'/F and U2AF<sup>65</sup> which was used as a negative control. Positions of molecular weight markers are shown on the left.

(B) Putative binding sites for nucleolin, hnRNP K and hnRNP H proteins in SLS2-A7 RNA. The SLS2-A7 RNA 2D structure is drawn according to.<sup>8, 12</sup> The putative binding sites for nucleolin, hnRNP K and hnRNP H are indicated by circles nucleotides. These sites are proposed taking into account the known RNA sequence specificities of the proteins.<sup>29</sup> *Cis*-regulatory elements present in SLS2-A7 RNA are delineated by broken arrows and position of the A7 3'ss is shown by a straight arrow.

**Figure 4:** Probing of hnRNP K binding sites on SLS1,2,3-A7 RNA

(A) Footprinting analysis of 5'-end [<sup>32</sup>P]-labeled SLS1,2,3-A7 RNA (0.5 pmol in the presence of 2 µg of tRNA) cleaved by 0.5 U of RNase T1 in 11 µl of buffer D in the absence (-) or in the presence (+) of increasing concentrations of His-hnRNP K protein (1.1 and 2.15 µM, respectively). The two lanes on the left are control assays performed in the absence of RNase T1, without or with addition of 2.15 µM hnRNP K. Numbering of nucleotides and positions of the A7 3'ss and its *cis*-regulatory elements are indicated on the right side of the autoradiogram, positions of stem-loop structures (SLS1, SLS2 and SLS3) are shown on the left side as well as positions of protected nucleotides.

(B) Schematic representation of the results shown in Panel A on the SLS1,2,3-A7 RNA 2D structure. Nucleotides protected by hnRNP K are indicated by white, grey and black circled nucleotides for low, medium and strong protections, respectively. Site A7 and its *cis*-regulatory elements are indicated as in **Figure 2**.

**Figure 5:** Co-stimulation of hnRNP K and hnRNP A1 binding to SLS2-A7 RNA

(A) EMSA were performed using uniformly [<sup>32</sup>P]-labeled RNA incubated with increasing amounts of recombinant His-hnRNP K protein (0.2–2 µM), in the absence (-) or the presence

(+) of 0.1  $\mu$ M of hnRNP A1. The positions of RNP and free RNA are indicated on the left side of the panel.

(B) Effect of the addition of recombinant hnRNP K on the yield of UV cross-linking of hnRNP A1 to SLS2-A7 RNA in HeLa nuclear extract.

Uniformly [ $^{32}$ P]-labeled SLS2-A7 RNA was incubated with 5  $\mu$ l of HeLa cell nuclear extract in the absence or the presence of 0.6 or 1.2  $\mu$ M of hnRNP K and irradiated at 254 nm as described in Material and Methods. After digestion by RNase T1 the cross-linked proteins were immunoselected with the hnRNP A1 monoclonal antibody (4B10) bound to Protein G sepharose beads. A part of the eluted material was fractionated by 10% SDS-PAGE and the radiolabeled protein was detected by autoradiography and quantified (IP anti-hnRNP A1). The remaining part of the cross-linked sample was submitted to Western blot analysis using the same antibody (Western Blot) in order to define the ratio of cross-linked *versus* total immunoselected hnRNP A1 protein. The mean values of the ratios obtained in three independent experiments are given in the graph on the right side of the panel. In each of the three experiments, the value obtained in the absence of hnRNP K were arbitrary defined as the 100% references and values obtained in the presence of hnRNP K were calculated as percentages of these references. Error bars are plotted using the calculated standard deviations.

**Figure 6:** Recombinant hnRNP K inhibits *in vitro* splicing of a D4-A7 mini-intron

Polyacrylamide gel electrophoresis of the *in vitro* splicing products of L47-8072 RNA (left). Uniformly [ $^{32}$ P]-labeled L47-8072 pre-mRNA (100,000 cpm) was incubated in a HeLa cell nuclear extract in conditions described in Materials and Methods in the absence or the presence of increasing amounts of purified His-hnRNP K protein (0.5 or 1.0  $\mu$ M). Unspliced pre-mRNA was fractionated in parallel on the left of the gel. Positions of the L47-8072 pre-mRNA, mRNA and splicing intermediates are schematically represented on the left side of

the autoradiogram. The Middle and right panels corresponds to splicing of the control pre-mRNAs pSma and 6A-PyI5-7. The M/P values calculated as described in Materials and Methods are given in a graph at the bottom of the gel. The values given are mean values of four independent experiments, error bars were established using the calculated standard deviations.

**Figure 7:** Over-expression of hnRNP K in HeLa cells strongly modifies the relative efficiencies of HIV-1 splicing sites

(A) Western blot analyses of the amounts of hnRNP K, ASF/SF2 and Nef proteins in the transfected cells. Total proteins were extracted after 48 h of transfection as described in “Materials and Methods”. For each of the assays equal amounts of total proteins were fractionated by SDS-PAGE. The amount of hnRNP K, ASF/SF2 and Nef protein were estimated by Western blot analysis using specific monoclonal antibodies specific for hnRNP K (3C2 antibody), ASF/SF2 (1D7 antibody) and protein Nef (MaTG0020 antibody). The tubulin amount measured with the anti- $\beta$  tubulin monoclonal SAP.4G5 antibody was used for normalization. Quantification allowed us to determine the percentage of reduction of Nef protein expression upon hnRNP K and ASF/SF2 protein overexpression.

(B) RT-PCR analysis of p $\Delta$ PSP spliced products.

HeLa cells were co-transfected with plasmid p $\Delta$ PSP carrying all the HIV-1 splicing sites<sup>39</sup> and pXJ41 plasmid derivatives over-expressing the hnRNP K (pXJ41-hnRNP K) or ASF/SF2 (pXJ41-ASF) protein. Two different amounts of pXJ41-hnRNP K plasmid (1  $\mu$ g and 2  $\mu$ g) were used for transfection, as indicated above the lanes, 250 ng of pXJ41-ASF plasmid were used for the control experiment. Another control experiment was performed with HeLa cells transfected with 2  $\mu$ g of empty pXJ41 plasmid. For each transfection assay, an equal amount of total RNA was subjected to RT-PCR analysis using the specific HIV-1 primers given in

“Materials and Methods” (**Suppl. Table 3**). The lane marked by M corresponds to DNA size markers, their sizes are indicated in base-pairs. The identity of the HIV-1 mRNAs corresponding to the fractionated cDNAs were deduced from previously published data <sup>7, 39</sup> and indicated on the right side of the panel.

**Figure 8:** Possible physical and functional interactions between the identified protein partners of SLS2-A7 RNA and their possible links with HIV-1 RNA biology.

This figure is drawn based on already published association, direct interactions or functional links between proteins that we identified in the HIV-1 SLS2-A7 RNPs (see **Table 1** and the discussion section for the corresponding references). A special focus is made on protein or RNA-protein associations that were found to have a link with HIV-1 biology. The secondary structures of relevant RNA motifs are outlined as bold solid line (SLS2-A7) and grey lines (RRE, TAR and Psi elements). The position of the A7 3'ss is indicated by an arrow. Already described direct protein-protein interactions are represented by solid lines, proteins found together in RNP or in protein complexes are shown by dashed lines. Red lines with red circles link kinases to their targets. Activation of PRKDC by IFL2 is indicated by a green arrow.

**Table 1** List of proteins identified by LC/MS/MS in the control MS2 RNPs and SLS2 A7 MS2 RNPs. Proteins are classified according to their cellular functions. For clarity, proteins such as trypsin, keratins and MBP were removed from this table.

## References

1. Purcell DF, Martin MA. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol* 1993; 67:6365-78.
2. Stoltzfus MC. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. *Adv Virus Res* 2009; 74:1-40.
3. Nekhai S, Jeang KT. Transcriptional and post-transcriptional regulation of HIV-1 gene expression: role of cellular factors for Tat and Rev. *Future Microbiol* 2006; 1:417-26.
4. Cullen BR. HIV-1 auxiliary proteins: making connections in a dying cell. *Cell* 1998; 93:685-92.
5. Cullen BR. Nuclear mRNA export: insights from virology. *Trends Biochem Sci* 2003; 28:419-24.
6. Li CJ, Friedman DJ, Wang C, Metelev V, Pardee AB. Induction of apoptosis in uninfected lymphocytes by HIV-1 Tat protein. *Science* 1995; 268:429-31.
7. Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, Branlant C. 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* 2001; 276:40464-75.
8. Marchand V, Mereau A, Jacquenet S, Thomas D, Mougin A, Gattoni R, Stevenin J, Branlant C. A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. *J Mol Biol* 2002; 323:629-52.
9. Hallay H, Locker N, Ayadi L, Ropers D, Guittet E, Branlant C. 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* 2006; 281:37159-74.

10. Staffa A, Cochrane A. Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. *Mol Cell Biol* 1995; 15:4597-605.
11. Si ZH, Rauch D, Stoltzfus CM. The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. *Mol Cell Biol* 1998; 18:5404-13.
12. Damgaard CK, Tange TO, Kjems J. 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* 2002; 8:1401-15.
13. Tange TO, Damgaard CK, Guth S, Valcarcel J, Kjems J. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J* 2001; 20:5748-58.
14. Caputi M, Mayeda A, Krainer AR, Zahler AM. hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO J* 1999; 18:4060-7.
15. Zhu J, Mayeda A, Krainer AR. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 2001; 8:1351-61.
16. Jablonski JA, Caputi M. Role of cellular RNA processing factors in human immunodeficiency virus type 1 mRNA metabolism, replication, and infectivity. *J Virol* 2009; 83:981-92.
17. Saliou JM, Bourgeois CF, Ayadi-Ben Mena L, Ropers D, Jacquenet S, Marchand V, Stevenin J, Branlant C. Role of RNA structure and protein factors in the control of HIV-1 splicing. *Front Biosci* 2009; 14:2714-29.
18. Tange TO, Kjems J. SF2/ASF binds to a splicing enhancer in the third HIV-1 tat exon and stimulates U2AF binding independently of the RS domain. *J Mol Biol* 2001; 312:649-62.

19. Mayeda A, Sreaton GR, Chandler SD, Fu XD, Krainer AR. Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol Cell Biol* 1999; 19:1853-63.
20. Okunola HL, Krainer AR. Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol* 2009; 29:5620-31.
21. Woolaway K, Asai K, Emili A, Cochrane A. hnRNP E1 and E2 have distinct roles in modulating HIV-1 gene expression. *Retrovirology* 2007; 4:28.
22. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 2009; 460:711-6.
23. Hartmuth K, Vornlocher HP, Luhrmann R. Tobramycin affinity tag purification of spliceosomes. *Methods Mol Biol* 2004; 257:47-64.
24. Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* 2002; 8:426-39.
25. Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, Stark H, Urlaub H, Luhrmann R. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* 2006; 26:5528-43.
26. Hamasaki K, Killian J, Cho J, Rando RR. Minimal RNA constructs that specifically bind aminoglycoside antibiotics with high affinities. *Biochemistry* 1998; 37:656-63.
27. Mongelard F, Bouvet P. Nucleolin: a multiFACeTed protein. *Trends Cell Biol* 2007; 17:80-6.
28. Allain FH, Bouvet P, Dieckmann T, Feigon J. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J* 2000; 19:6870-81.

29. Bomsztyk K, Denisenko O, Ostrowski J. hnRNP K: one protein multiple processes. *Bioessays* 2004; 26:629-38.
30. Fiset JF, Toutant J, Dugre-Brisson S, Desgroseillers L, Chabot B. hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection. *RNA* 2009; 16:228-38.
31. Schaub MC, Lopez SR, Caputi M. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. *J Biol Chem* 2007; 282:13617-26.
32. Belenguer P, Caizergues-Ferrer M, Labbe JC, Doree M, Amalric F. Mitosis-specific phosphorylation of nucleolin by p34cdc2 protein kinase. *Mol Cell Biol* 1990; 10:3607-18.
33. Bouvet P, Allain FH, Finger LD, Dieckmann T, Feigon J. Recognition of pre-formed and flexible elements of an RNA stem-loop by nucleolin. *J Mol Biol* 2001; 309:763-75.
34. Ginisty H, Amalric F, Bouvet P. Two different combinations of RNA-binding domains determine the RNA binding specificity of nucleolin. *J Biol Chem* 2001; 276:14338-43.
35. Serin G, Joseph G, Ghisolfi L, Bauzan M, Erard M, Amalric F, Bouvet P. Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *J Biol Chem* 1997; 272:13109-16.
36. Jacquenet S, Ropers D, Bilodeau PS, Damier L, Mougou A, Stoltzfus CM, Branlant C. 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. *Nucl Acids Res* 2001; 29:464-78.
37. Caputi M, Zahler AM. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem* 2001; 276:43850-9.
38. Expert-Bezançon A, Le Caer JP, Marie J. Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the

splicing of the alternative exon 6A from chicken beta-tropomyosin pre-mRNA. *J Biol Chem* 2002; 277:16614-23.

39. Ropers D, Ayadi L, Gattoni R, Jacquenet S, Damier L, Branlant C, Stevenin J. 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* 2004; 279:29963-73.

40. Jacquenet S, Decimo D, Muriaux D, Darlix JL. Dual effect of the SR proteins ASF/SF2, SC35 and 9G8 on HIV-1 RNA splicing and virion production. *Retrovirology* 2005; 2:33.

41. Ginisty H, Serin G, Ghisolfi-Nieto L, Roger B, Libante V, Amalric F, Bouvet P. Interaction of nucleolin with an evolutionarily conserved pre-ribosomal RNA sequence is required for the assembly of the primary processing complex. *J Biol Chem* 2000; 275:18845-50.

42. Finger LD, Johansson C, Rinaldi B, Bouvet P, Feigon J. Contributions of the RNA-binding and linker domains and RNA structure to the specificity and affinity of the nucleolin RBD12/NRE interaction. *Biochemistry* 2004; 43:6937-47.

43. Ginisty H, Sicard H, Roger B, Bouvet P. Structure and functions of nucleolin. *J Cell Sci* 1999; 112 ( Pt 6):761-72.

44. Nisole S, Krust B, Hovanessian AG. Anchorage of HIV on permissive cells leads to coaggregation of viral particles with surface nucleolin at membrane raft microdomains. *Exp Cell Res* 2002; 276:155-73.

45. Cochrane AW, Perkins A, Rosen CA. Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function. *J Virol* 1990; 64:881-5.

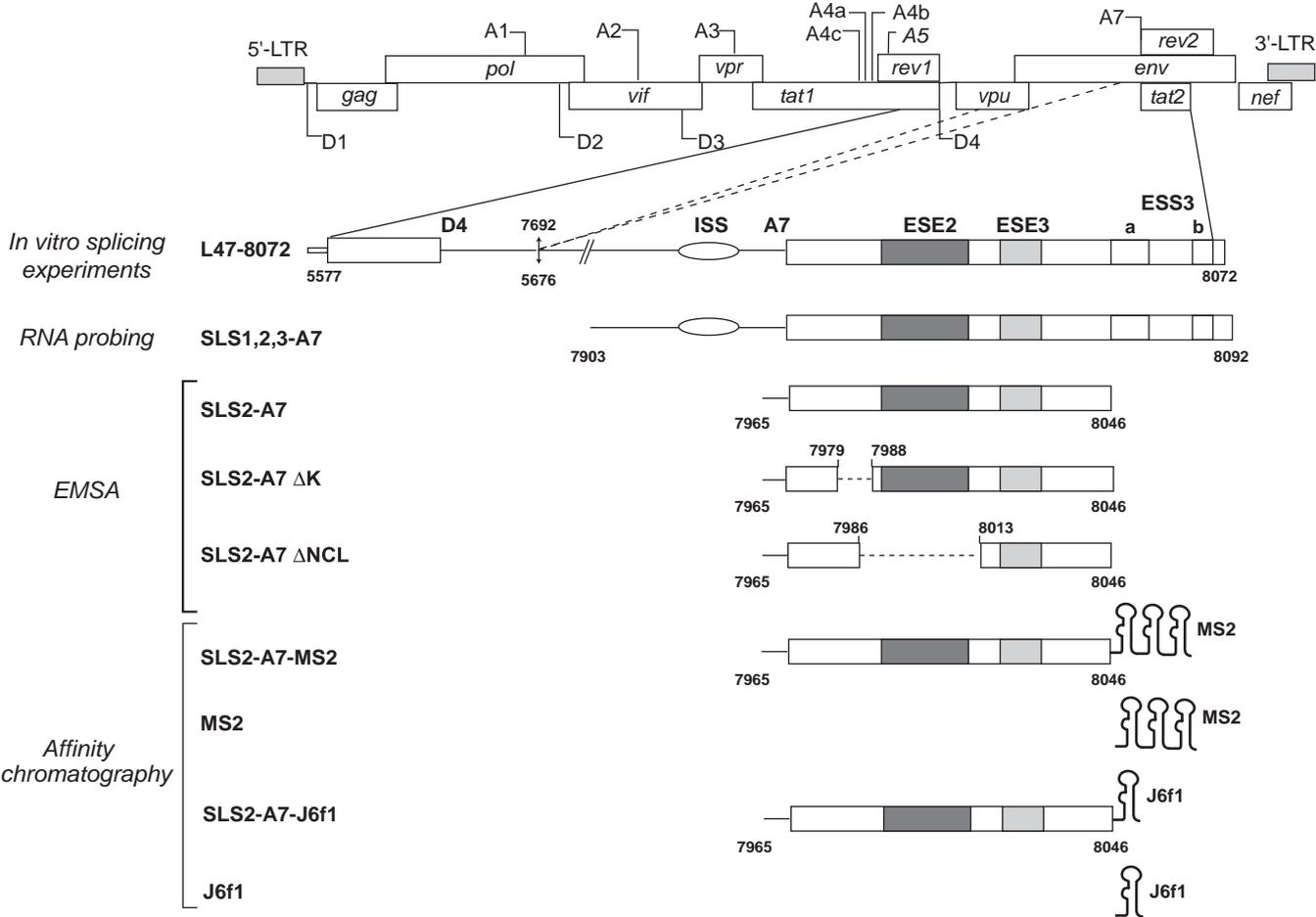
46. Michienzi A, Cagnon L, Bahner I, Rossi JJ. Ribozyme-mediated inhibition of HIV 1 suggests nucleolar trafficking of HIV-1 RNA. *Proc Natl Acad Sci* 2000; 97:8955-60.

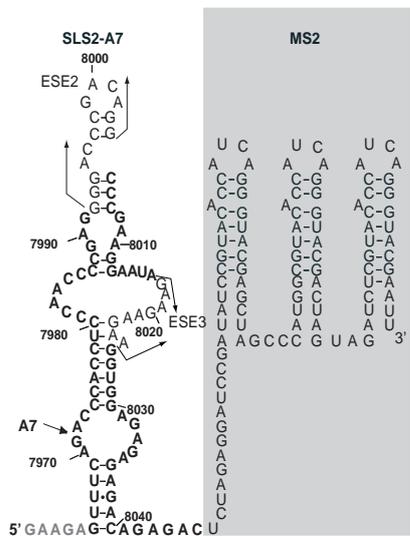
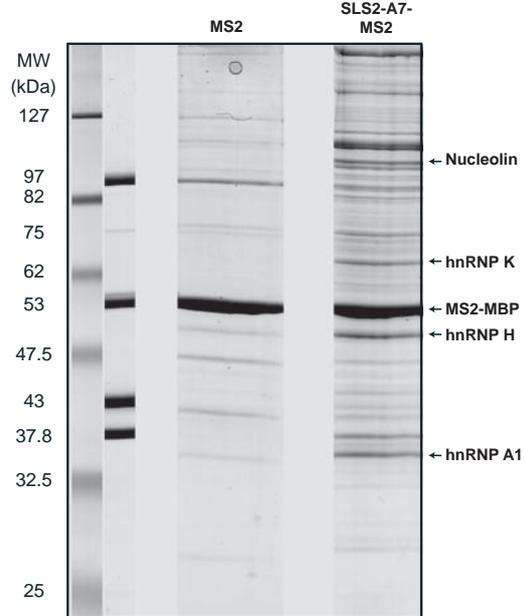
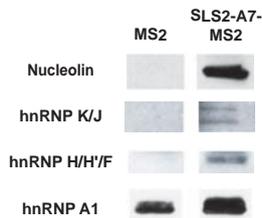
47. Bacharach E, Gonsky J, Alin K, Orlova M, Goff SP. The carboxy-terminal fragment of nucleolin interacts with the nucleocapsid domain of retroviral gag proteins and inhibits virion assembly. *J Virol* 2000; 74:11027-39.
48. Ueno T, Tokunaga K, Sawa H, Maeda M, Chiba J, Kojima A, Hasegawa H, Shoya Y, Sata T, Kurata T, Takahashi H. Nucleolin and the packaging signal, psi, promote the budding of human immunodeficiency virus type-1 (HIV-1). *Microbiol Immunol* 2004; 48:111-8.
49. Chen CY, Gherzi R, Andersen JS, Gaietta G, Jurchott K, Royer HD, Mann M, Karin M. Nucleolin and YB-1 are required for JNK-mediated interleukin-2 mRNA stabilization during T-cell activation. *Genes Dev* 2000; 14:1236-48.
50. Huang Y, Huang K, Boskovic G, Dementieva Y, Denvir J, Primerano DA, Zhu GZ. Proteomic and genomic analysis of PITX2 interacting and regulating networks. *FEBS Lett* 2009; 583:638-42.
51. Borggreffe T, Wabl M, Akhmedov AT, Jessberger R. A B-cell-specific DNA recombination complex. *J Biol Chem* 1998; 273:17025-35.
52. Nicol SM, Fuller-Pace FV. Analysis of the RNA helicase p68 (Ddx5) as a transcriptional regulator. *Methods Mol Biol* 2010; 587:265-79.
53. Kim JH, Hahm B, Kim YK, Choi M, Jang SK. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J Mol Biol* 2000; 298:395-405.
54. Griffith BN, Walsh CM, Szeszel-Fedorowicz W, Timperman AT, Salati LM. Identification of hnRNPs K, L and A2/B1 as candidate proteins involved in the nutritional regulation of mRNA splicing. *Biochim Biophys Acta* 2006; 1759:552-61.
55. Revil T, Pelletier J, Toutant J, Cloutier A, Chabot B. Heterogeneous nuclear ribonucleoprotein K represses the production of pro-apoptotic Bcl-xS splice isoform. *J Biol Chem* 2009; 284:21458-67.

56. Wolf D, Witte V, Clark P, Blume K, Lichtenheld MG, Baur AS. HIV Nef enhances Tat-mediated viral transcription through a hnRNP-K-nucleated signaling complex. *Cell Host Microbe* 2008; 4:398-408.
57. Yang JP, Reddy TR, Truong KT, Suhasini M, Wong-Staal F. Functional interaction of Sam68 and heterogeneous nuclear ribonucleoprotein K. *Oncogene* 2002; 21:7187-94.
58. Reddy TR, Xu W, Mau JK, Goodwin CD, Suhasini M, Tang H, Frimpong K, Rose DW, Wong-Staal F. Inhibition of HIV replication by dominant negative mutants of Sam68, a functional homolog of HIV-1 Rev. *Nat Med* 1999; 5:635-42.
59. McLaren M, Cochrane A. Mapping of determinants involved in the stimulation of HIV-1 expression by Sam68. *Virology* 2009; 385:93-104.
60. Mikula M, Dzwonek A, Karczmariski J, Rubel T, Dadlez M, Wyrwicz LS, Bomszyk K, Ostrowski J. Landscape of the hnRNP K protein-protein interactome. *Proteomics* 2006; 6:2395-406.
61. Chen HC, Lin WC, Tsay YG, Lee SC, Chang CJ. An RNA helicase, DDX1, interacting with poly(A) RNA and heterogeneous nuclear ribonucleoprotein K. *J Biol Chem* 2002; 277:40403-9.
62. Fang J, Kubota S, Yang B, Zhou N, Zhang H, Godbout R, Pomerantz RJ. A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev. *Virology* 2004; 330:471-80.
63. Zhang S, Schlott B, Gorchach M, Grosse F. DNA-dependent protein kinase (DNA-PK) phosphorylates nuclear DNA helicase II/RNA helicase A and hnRNP proteins in an RNA-dependent manner. *Nucl Acids Res* 2004; 32:1-10.
64. Li J, Tang H, Mullen TM, Westberg C, Reddy TR, Rose DW, Wong-Staal F. A role for RNA helicase A in post-transcriptional regulation of HIV type 1. *Proc Natl Acad Sci* 1999; 96:709-14.

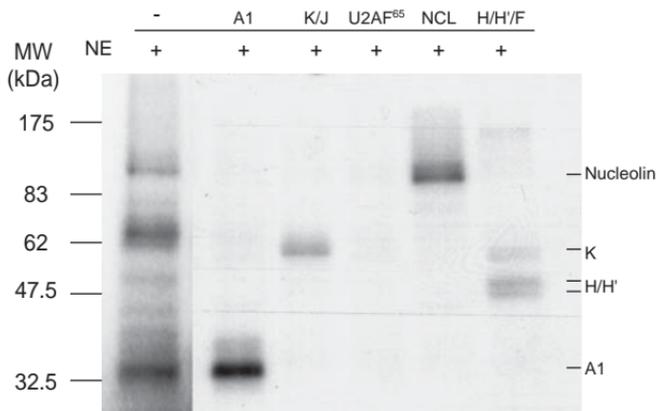
65. Ting NS, Kao PN, Chan DW, Lintott LG, Lees-Miller SP. DNA-dependent protein kinase interacts with antigen receptor response element binding proteins NF90 and NF45. *J Biol Chem* 1998; 273:2136-45.
66. Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, Dubois A, Sanglier-Cianferani S, Van Dorsselaer A, Clerc P, Avner P, Visvikis A, Branlant C. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol* 2010; 8:e1000276.
67. Chou MY, Rooke N, Turck CW, Black DL. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol Cell Biol* 1999; 19:69-77.
68. Hallay H. Study on the mechanisms of alternative splicing of the RNA from the human immunodeficiency virus 1: Binding and mechanism of action of hnRNP and SR proteins at the A3 acceptor site. Vandoeuvre les Nancy: H. Poincaré University, 2006.
69. Bourgeois CF, Popielarz M, Hildwein G, Stevenin J. Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol Cell Biol* 1999; 19:7347-56.
70. Miguet L, Bechade G, Fornecker L, Zink E, Felden C, Gervais C, Herbrecht R, Van Dorsselaer A, Mauvieux L, Sanglier-Cianferani S. Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res* 2009; 8:3346-54.
71. De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2007; 2:778-91.

72. Damier L, Domenjoud L, Branlant C. 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* 1997; 237:182-7.

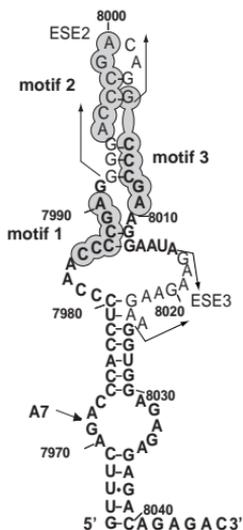


**A****B****C**

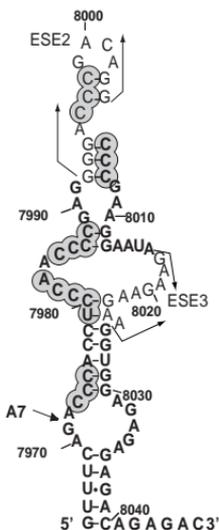
## SLS2-A7 RNA



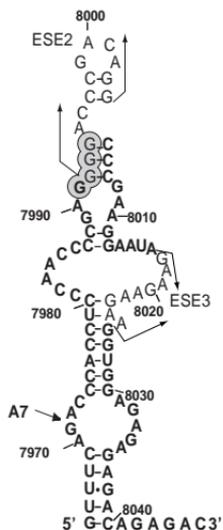
**B** Putative nucleolin binding sites

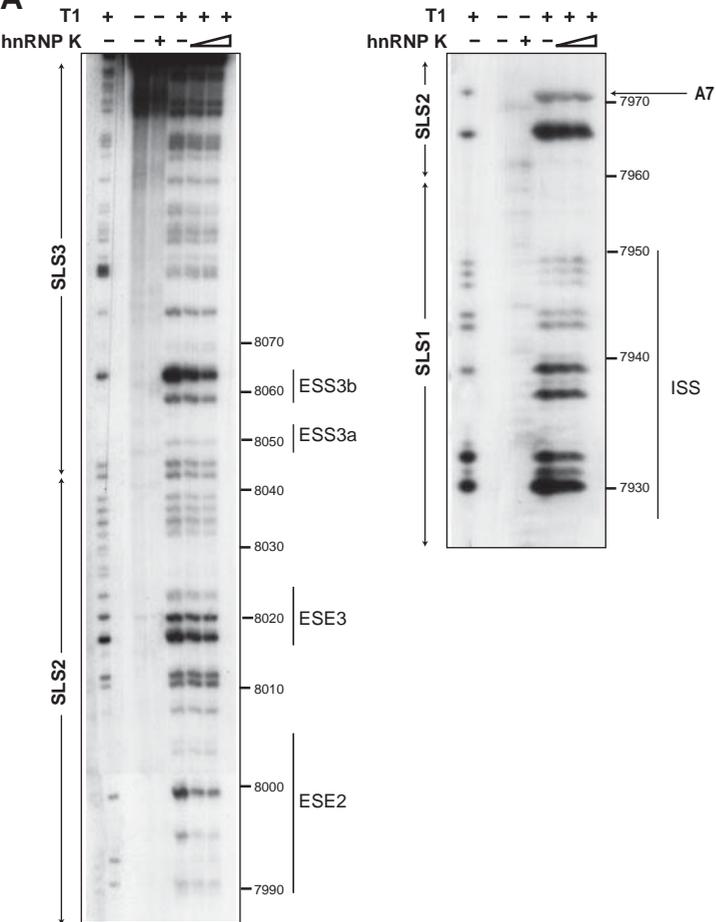
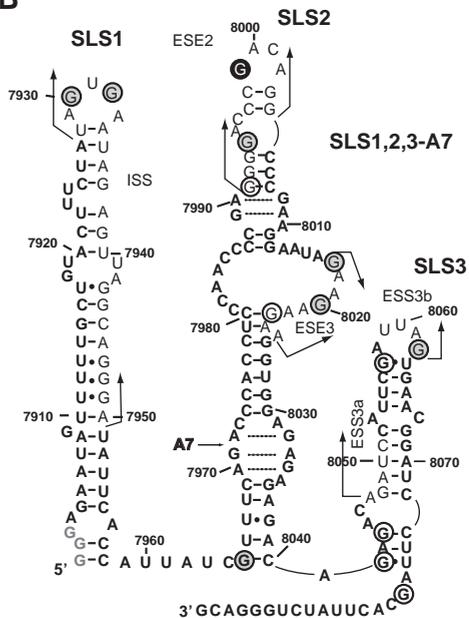


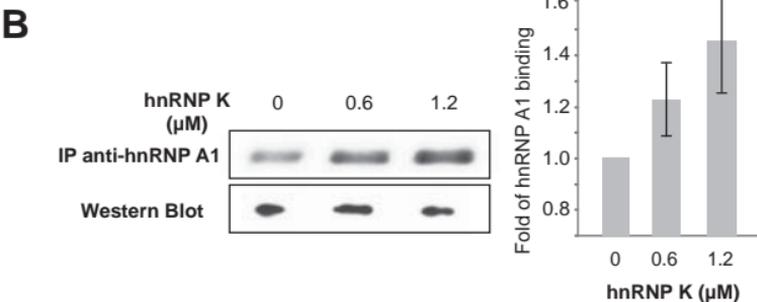
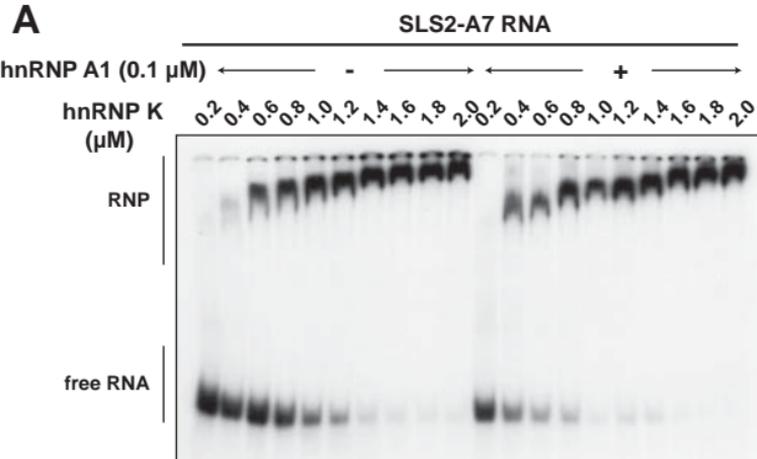
**C** Putative hnRNP K binding sites

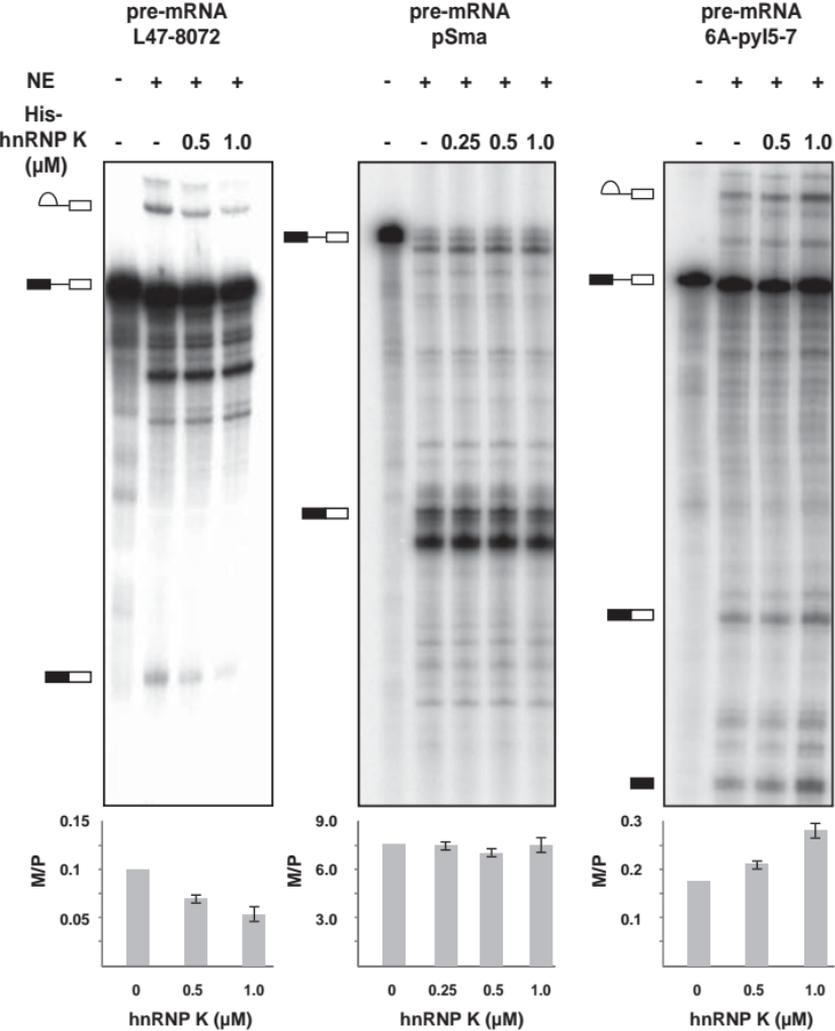


**D** Putative hnRNP H binding sites

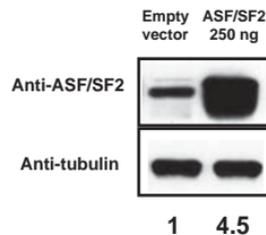
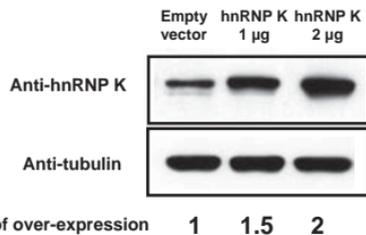
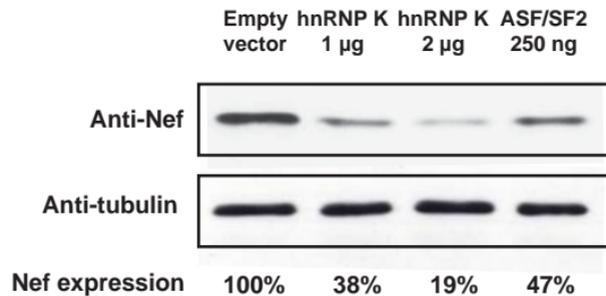


**A****B**

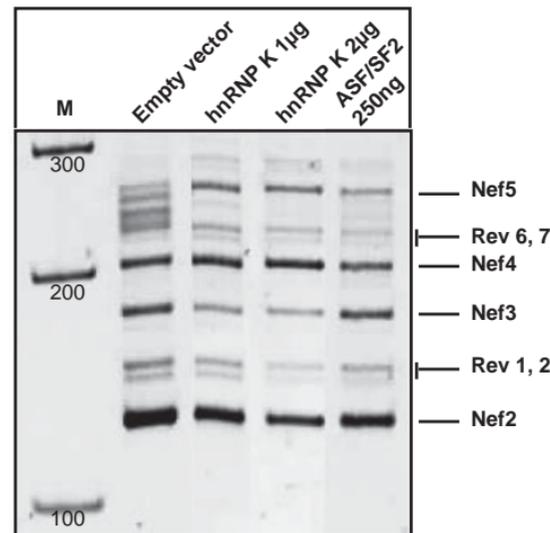


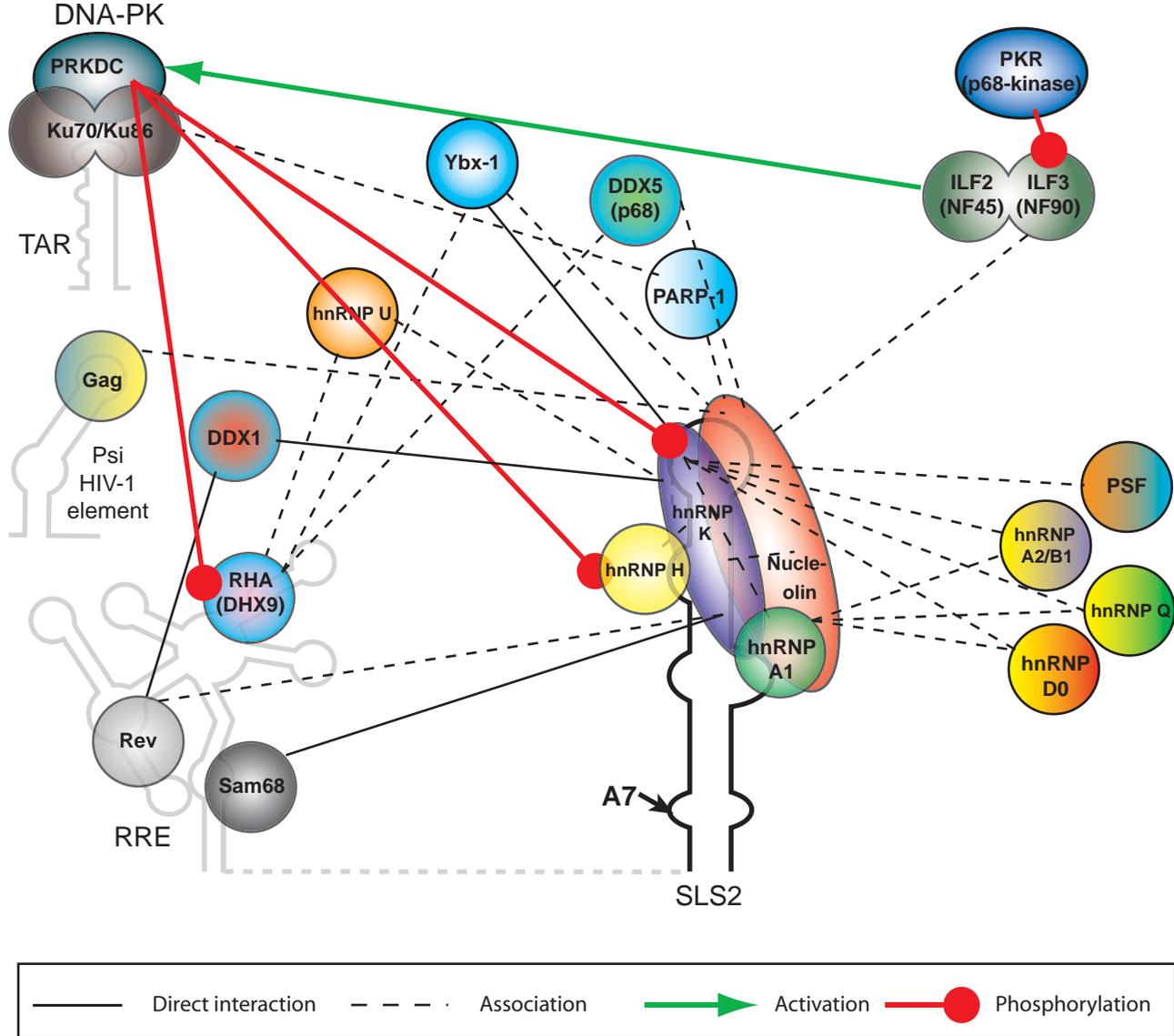


A



B

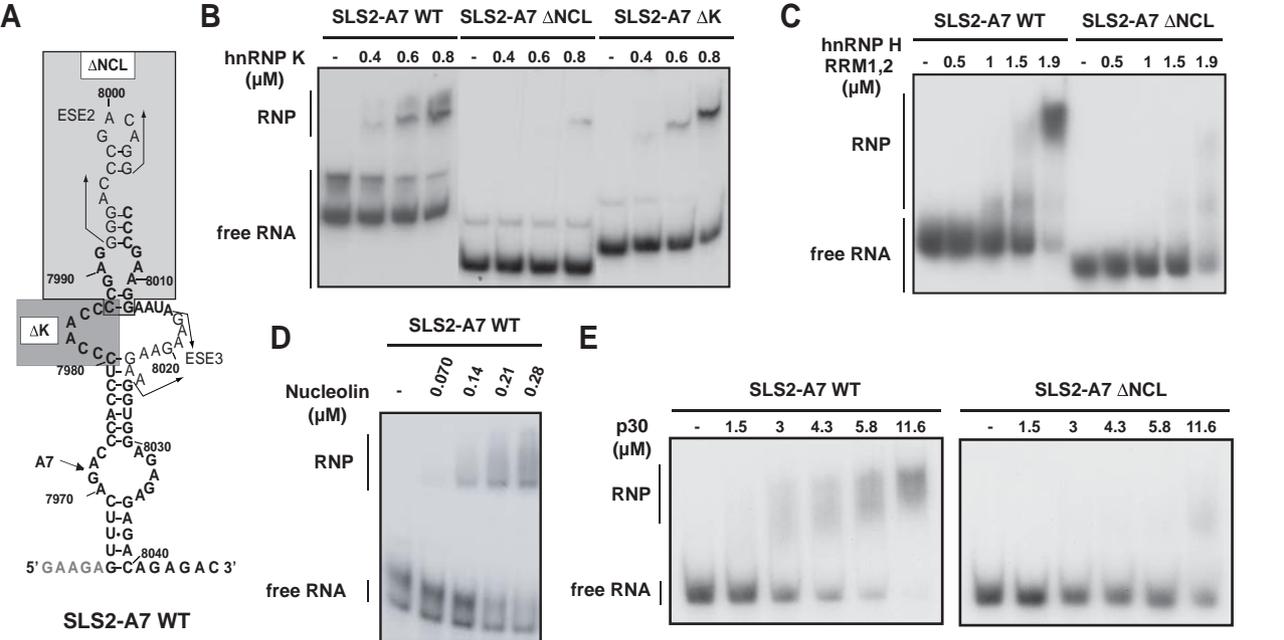




**Figure 8: Possible physical and functional interactions between the identified protein partners of SLS2-A7 RNA and their possible links with HIV-1 RNA biology.**

This figure is drawn based on already published association, direct interactions or functional links between proteins that we identified in the HIV-1 SLS2-A7 RNPs (for details on the interactions, see Supplementary Tables 1 and 2). A special focus is made on protein or RNA-protein associations that were found to have a link with HIV-1 biology. The secondary structures of relevant RNA motifs are outlined as bold solid line (SLS2-A7) and grey lines (RRE, TAR and Psi elements). The position of the A7 3' ss is indicated by an arrow. Already described direct protein-protein interactions are represented by solid lines, proteins found together in RNP or in protein complexes are shown by dashed lines. Red lines with red circles link kinases to their targets. Activation of PRKDC by ILF2 is indicated by a green arrow.

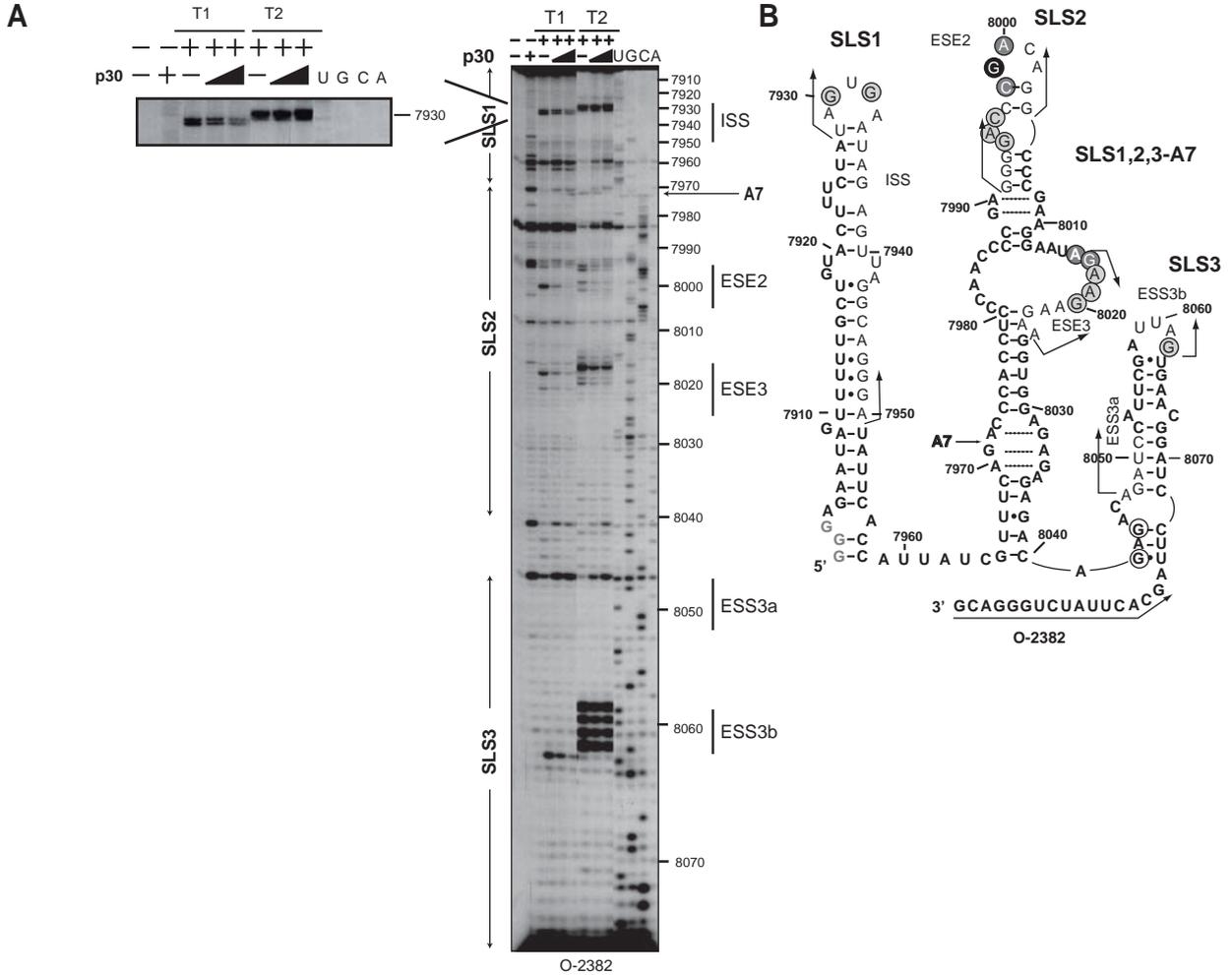
	Protein name	Alternative Name	Accession numbers	Molecular weight (Da)	Unique peptide			
					MS2	SLS2 -MS2	J6f1	SLS2 -J6f1
<b>hnRNP proteins</b>	Heterogeneous nuclear ribonucleoprotein A1		P09651	34 194	2	8		6
	Heterogeneous nuclear ribonucleoprotein A3		P51991	39 634		3		5
	Heterogeneous nuclear ribonucleoprotein D0		Q14103	38 337		4		4
	Heterogeneous nuclear ribonucleoprotein H		P31943	49 182		3		3
	Heterogeneous nuclear ribonucleoprotein K		P61978	50 961		8		11
	Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	O60506	69 616		2		7
	Heterogeneous nuclear ribonucleoprotein R		O43390	70 926		5		8
Heterogeneous nuclear ribonucleoprotein U		Q00839	90 567		7	2	5	
Heterogeneous nuclear ribonucleoproteins A2/B1		P22626	37 460		7	3	8	
<b>Nucleolin partners</b>	Nuclease-sensitive element-binding protein 1	Ybx1	P67809	35 712		2		5
	Heat shock 70 kDa protein 1	Hsp70	P08107	70 169		3		
	Nucleolin		P19338	76 842		9	2	12
	Poly [ADP-ribose] polymerase 1	PARP1	P09874	113 070	3	7		16
<b>Splicing factors</b>	Splicing factor U2AF 35 kDa subunit		Q01081	27 797		2		2
	Splicing factor, proline- and glutamine-rich		P23246	75 424		2		
<b>RNA helicases</b>	ATP-dependent RNA helicase A	DHX9, RHA	Q08211	140 903	6	7		11
	ATP-dependent RNA helicase DDX1		Q92499	82 415		6		
	Probable ATP-dependent RNA helicase DDX5	p68	P17844	69 304		6		5
<b>Transcription factors</b>	Bromodomain adjacent to zinc finger domain protein 1A		Q9NRL2	178 689		3		
	General transcription factor 3C polypeptide 3		Q9Y5Q9	101 258		2		
	Probable global transcription activator SNF2L4		P51532	184 632		4		4
	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5		O60264	121 893		6		
	Transcription intermediary factor 1-beta		Q13263	88 531		2		
<b>Translation factors</b>	Elongation factor 1-alpha 1		P68104	50 167	5	5		3
	Selenocysteine-specific elongation factor		P57772	65 288		2		2
<b>DNA binding proteins</b>	ATP-dependent DNA helicase 2 subunit 1	XRCC6, Ku70	P12956	69 828		3		7
	ATP-dependent DNA helicase 2 subunit 2	XRCC5, Ku86	P13010	82 689		5		
	DNA topoisomerase 2-alpha		P11388	174 373		2		
	DNA-dependent protein kinase catalytic subunit	PRKDC	P78527	469 078		17		10
<b>Cell division proteins</b>	Anaphase-promoting complex subunit 7		Q9UJX3	63 117		2		
	Sister chromatid cohesion protein PDS5 homolog A		Q29RF7	150 818		2		
<b>Interleukins</b>	Interleukin enhancer-binding factor 2	ILF2, NF45	Q12905	43 045	11	9		8
	Interleukin enhancer-binding factor 3	ILF3, NF90	Q12906	95 321	7	7		10
<b>Other proteins</b>	E3 ubiquitin/ISG15 ligase TRIM25		Q14258	70 971		2		
	E3 ubiquitin-protein ligase UHRF1		Q96T88	89 796		2		
	Glycogen phosphorylase, liver form		P06737	97 134		3		
	GTP-binding nuclear protein Ran		P62826	24 405		4		
	Importin-5		O00410	123 614	3			
	Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	P19525	62 079	4	3		2
	Nucleophosmin		P06748	32 542	3			
	Ras GTPase-activating-like protein IQGAP1		P46940	189 241		3		13
UPF0568 protein C14orf166 homolog		Q9Y224	28 051		6			



**Supplementary figure 1: Deletion of the apical stem-loop ( $\Delta$ NCL variant) affects the binding of recombinant proteins p30, hnRNP K and hnRNP H to SLS2-A7 RNA.**

**(A)** Representation of the two SLS2-A7 RNA deletion mutants used in EMSA experiments. The deleted sequences in variant SLS2-A7  $\Delta$ K (nts 7980-7987) and SLS2-A7  $\Delta$ NCL (nts 7987-8012) RNAs are indicated on the SLS2-A7 RNA 2D structure by grey and dark grey boxes, respectively.

**(B) to (E)** EMSA experiments. RNP complexes were formed by incubation of uniformly labeled WT and variant RNAs with increasing concentrations of the recombinant proteins: in panel B the recombinant hnRNP K used at 0.4, 0.6, 0.8  $\mu$ M concentrations, in panel C the recombinant His-hnRNP H RRM1,2 was used at 0.5, 1, 1.5 and 1.9  $\mu$ M concentrations, in panel d nucleolin purified from HeLa cell nuclear extract according to Belenguer <sup>1</sup> was used at 0.07, 0.14, 0.21 and 0.28  $\mu$ M concentrations and in panel E the recombinant nucleolin RRM1,2,3 (p30 protein) was used at 1.5, 3, 4.3, 5.8 and 11.6  $\mu$ M concentrations. In each of the EMSA autoradiograms, the positions of RNPs and free RNAs are indicated.



### Supplementary figure 2: Probing of p30 protein binding sites on SLS1,2,3-A7 RNA

Prior to this experiment the p30 protein concentration needed to get a complete gel-shift of an *in vitro* transcribed SLS1,2,3-A7 RNA was defined (about 14  $\mu\text{M}$ ). Thus we selected two p30 protein concentrations (8.75  $\mu\text{M}$  and 18.23  $\mu\text{M}$  corresponding to [p30]/[RNA] ratios of 104 and 260, respectively) to form the RNP complexes that were subjected to RNase T1 and T2 digestions.

**(A)** Primer extension analysis of SLS1,2,3-A7 RNA (1.12 pmol in the presence of 5  $\mu\text{g}$  of tRNA) cleaved by 0.1 U of RNase T1 (lanes T1) or 0.6 U of RNase T2 (lanes T2) in 9  $\mu\text{l}$  of buffer D, in the absence (-) or in the presence (+) of increasing concentrations of p30 protein (8.75 and 18.25  $\mu\text{M}$ , respectively). The lane on the left of the gel is a control extension performed on the same RNA incubated in the same conditions in the absence of RNase and protein p30. The second lane corresponds to a control extension performed on the RNA incubated with p30 protein. Lanes U, G, C and A correspond to the sequencing ladder. Numbering of nucleotides in the HIV-1/BRU RNA is shown on the right, as well as positions of the A7 3'ss and its cis-regulatory elements. Stem-loop structures (SLS1, SLS2 and SLS3) are indicated on the left side as well as positions of protected nucleotides. A part of the gel corresponding to the ISS sequence is enlarged and shown in the insert on the left side of the panel.

**(B)** Schematic representation of the digestion results shown in Panel A on the SLS1,2,3-A7 RNA 2D structure. The nucleotides protected by protein p30 are indicated by white, grey and black circles for low, medium and strong protections, respectively. The hybridization site of the primer O-2382 used for reverse transcription is shown. Site A7 and its cis-regulatory elements are indicated as in Figure 2. The three additional G residues at the 5'-end of the transcript that were introduced to improve T7 RNA transcription are in grey.

At the lowest protein concentration, a marked protection is detected in both the GACCCGA sequence (positions 7994 to 8000 in ESE2, motif 2) and the UAGAAG sequence (positions 8015 to 8020, ambivalent ESE3 element). Residue G8062 in ESS3b is partially protected at both concentrations. In addition, some protection of the two G residues in the terminal loop of SLS1 is detected at the highest p30 concentration. The absence of enzymatic cleavages in motifs 1 and 3 limited the analysis of these motifs. However, these footprinting data confirmed p30 protein binding to motif 2 in ESE2. The detected protection in ESE3 might result either from the direct association of p30 with ESE3 or, from p30 binding to the neighboring motifs 1 and 3, that could not be probed by RNase digestion. Therefore, SLS2-A7 contains strong nucleolin binding sites and SLS1-A7 and SLS3-A7 likely contain sites with lower affinities.

## Functional links of nucleolin and hnRNP K with proteins detected in SLS2-A7 RNPs.

Protein identity	Proposed function	Reference
Nucleolin	<p>- Nucleolin is found in association with Ybx1, hnRNP U and hnRNP K in protein complexes which stabilize cellular mRNAs.</p> <p>- Nucleolin was found in association with Parp1 and ILF3 (NF90).</p> <p>-Nucleolin was found in protein complex associated with hepatitis C which contain the RNA helicase p68 (DDX5).</p>	<p>Chen CY, Gherzi R, Andersen JS, Gaietta G, Jurchott K, Royer HD, Mann M, Karin M. Nucleolin and YB-1 are required for JNK-mediated interleukin-2 mRNA stabilization during T-cell activation. <i>Genes Dev</i> 2000; 14:1236-48</p> <p>Huang Y, Huang K, Boskovic G, Dementieva Y, Denvir J, Primerano DA, Zhu GZ. Proteomic and genomic analysis of PITX2 interacting and regulating networks. <i>FEBS Lett</i> 2009; 583:638-42.</p> <p>Skalweit A, Doller A, Huth A, Kahne T, Persson PB, Thiele BJ. Posttranscriptional control of renin synthesis: identification of proteins interacting with renin mRNA 3'-untranslated region. <i>Circ Res</i> 2003; 92:419-27.</p> <p>Lee PT, Liao PC, Chang WC, Tseng JT. Epidermal growth factor increases the interaction between nucleolin and heterogeneous nuclear ribonucleoprotein K/poly(C) binding protein 1 complex to regulate the gastrin mRNA turnover. <i>Mol Biol Cell</i> 2007; 18:5004-13.</p> <p>Borggreffe T, Wabl M, Akhmedov AT, Jessberger R. A B-cell-specific DNA recombination complex. <i>J Biol Chem</i> 1998; 273:17025-35.</p> <p>Lee JH, Nam IY, Myung H. Nonstructural protein 5B of hepatitis C virus. <i>Mol Cells</i> 2006; 21:330-6.</p> <p>Nicol SM, Fuller-Pace FV. Analysis of the RNA helicase p68 (Ddx5) as a transcriptional regulator. <i>Methods Mol Biol</i> 2010; 587:265-79.</p>
hnRNP K	<p>- Proteins found in hnRNP K interactome, Ybx1, DEAD box RNA helicase DDX1, RNA helicase DDX5, hnRNP U, hnRNP H, hnRNP A1, hnRNP A2/B1, hnRNP Q, hnRNP D0, PSF (direct binding for Ybx1 and DDX1).</p> <p>- hnRNP K, together with hnRNP H and RHA, is phosphorylated by the DNA-independant protein kinase DNA-PK containing the catalytic subunit PRKDC and its two associated ATP-dependant DNA helicases 1 and 2, also designated as Ku70/XRCC6 and Ku86/XRCC5).</p> <p>- PKR phosphorylates the ILF2-ILF3 complex which activate DNA-PK</p> <p>- HnRNP K was found to mediate translational silencing of cellular and viral mRNAs and this inhibitory property is increased by its phosphorylation by the cell cycle dependent ERK kinases, which drives hnRNP K into the cytoplasm.</p>	<p>Mikula M, Dzwonek A, Karczmariski J, Rubel T, Dadlez M, Wyrwicz LS, Bomsztyk K, Ostrowski J. Landscape of the hnRNP K protein-protein interactome. <i>Proteomics</i> 2006; 6:2395-406.</p> <p>Chen HC, Lin WC, Tsay YG, Lee SC, Chang CJ. An RNA helicase, DDX1, interacting with poly(A) RNA and heterogeneous nuclear ribonucleoprotein K. <i>J Biol Chem</i> 2002; 277:40403-9.</p> <p>Shnyreva M, Schullery DS, Suzuki H, Higaki Y, Bomsztyk K. Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein. <i>J Biol Chem</i> 2000; 275:15498-503.</p> <p>Zhang S, Schlott B, Gorlach M, Grosse F. DNA-dependent protein kinase (DNA-PK) phosphorylates nuclear DNA helicase II/RNA helicase A and hnRNP proteins in an RNA-dependent manner. <i>NAR</i> 2004; 32:1-10.</p> <p>Ting NS, Kao PN, Chan DW, Lintott LG, Lees-Miller SP. DNA-dependent protein kinase interacts with antigen receptor response element binding proteins NF90 and NF45. <i>J Biol Chem</i> 1998; 273:2136-45.</p> <p>Langland JO, Kao PN, Jacobs BL. Nuclear factor-90 of activated T-cells: A double-stranded RNA-binding protein and substrate for the double-stranded RNA-dependent protein kinase, PKR. <i>Biochemistry (Mosc)</i> 1999; 38:6361-8.</p> <p>Bomsztyk K, Denisenko O, Ostrowski J. hnRNP K: one protein multiple processes. <i>Bioessays</i> 2004; 26:629-38.</p>

## Implication in HIV-1 biology of proteins detected in SLS2-A7 RNPs.

Protein identity	Proposed function	Reference
Nucleolin	<p>- Cell surface nucleolin can mediate intracellular import of specific ligands and serve as a facilitator of HIV-1 particles internalization.</p> <p>- In HIV-1 infected cells, cytoplasmic nucleolin binds the stem-loop structure containing Psi together with Gag.</p> <p>- Nucleolin is proposed to be involved in virion assembly and budding</p>	<p>-Hovanessian AG, Puvion-Dutilleul F, Nisole S, Svab J, Perret E, Deng JS, Krust B. The cell-surface-expressed nucleolin is associated with the actin cytoskeleton. <i>Exp Cell Res</i> 2000; 261:312-28. Nisole S, Krust B, Hovanessian AG. Anchorage of HIV on permissive cells leads to coaggregation of viral particles with surface nucleolin at membrane raft microdomains. <i>Exp Cell Res</i> 2002; 276:155-73.</p> <p>- Bacharach E, Gonsky J, Alin K, Orlova M, Goff SP. The carboxy-terminal fragment of nucleolin interacts with the nucleocapsid domain of retroviral gag proteins and inhibits virion assembly. <i>J Virol</i> 2000; 74:11027-39.</p> <p>- Ueno T, Tokunaga K, Sawa H, Maeda M, Chiba J, Kojima A, Hasegawa H, Shoya Y, Sata T, Kurata T, Takahashi H. Nucleolin and the packaging signal, psi, promote the budding of human immunodeficiency virus type-1 (HIV-1). <i>Microbiol Immunol</i> 2004; 48:111-8.</p>
hnRNP K	<p>- hnRNP K interacts with Nef and nucleate the Nef-interacting kinases, including Lck, PKC<math>\delta</math>, and PI-3 kinase, leading to Lck and Erk1/2 activation.</p> <p>- hnRNP K interacts specifically with Sam68 which is a synergistic activator of the Rev function, Sam68 can even functionally substitute for Rev. Sam68 increases cleavage and subsequent polyadenylation of HIV-1 RNA.</p> <p>- hnRNPs K, A1, Q, R and U associate with the Rev protein.</p>	<p>- Wolf D, Witte V, Clark P, Blume K, Lichtenheld MG, Baur AS. HIV Nef enhances Tat-mediated viral transcription through a hnRNP-K-nucleated signaling complex. <i>Cell Host Microbe</i> 2008; 4:398-408.</p> <p>- Yang JP, Reddy TR, Truong KT, Suhasini M, Wong-Staal F. Functional interaction of Sam68 and heterogeneous nuclear ribonucleoprotein K. <i>Oncogene</i> 2002; 21:7187-94. Reddy TR, Xu W, Mau JK, Goodwin CD, Suhasini M, Tang H, Frimpong K, Rose DW, Wong-Staal F. Inhibition of HIV replication by dominant negative mutants of Sam68, a functional homolog of HIV-1 Rev. <i>Nat Med</i> 1999; 5:635-42. McLaren M, Cochrane A. Mapping of determinants involved in the stimulation of HIV-1 expression by Sam68. <i>Virology</i> 2009; 385:93-104. McLaren M, Asai K, Cochrane A. A novel function for Sam68: enhancement of HIV-1 RNA 3' end processing. <i>RNA</i> 2004; 10:1119-29.</p> <p>-Hadian K, Vincendeau M, Mausbacher N, Nagel D, Hauck SM, Ueffing M, Loyter A, Werner T, Wolff H, Brack-Werner R. Identification of a heterogeneous nuclear ribonucleoprotein-recognition region in the HIV Rev protein. <i>J Biol Chem</i> 2009; 284:33384-91.</p>
DDX1	<p>- DDX1 binds to Rev and modifies its subcellular localization. The absence of DDX1 expression in human astrocytes explains the low ratio of unspliced versus spliced HIV-1 RNA in these cells and the restriction of HIV-1 virus in astrocytes.</p>	<p>- Fang J, Acheampong E, Dave R, Wang F, Mukhtar M, Pomerantz RJ. The RNA helicase DDX1 is involved in restricted HIV-1 Rev function in human astrocytes. <i>Virology</i> 2005; 336:299-307. Fang J, Kubota S, Yang B, Zhou N, Zhang H, Godbout R, Pomerantz RJ. A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev. <i>Virology</i> 2004; 330:471-80.</p>
Ku70-Ku86 complex	<p>- The Ku70/XRCC6-Ku86/XRCC5 complex binds strongly to TAR HIV-1 RNA region. The Ku70-Ku86 complex is involved in HIV-1 DNA integration steps.</p>	<p>- Kaczmarek W, Khan SA. Lupus autoantigen Ku protein binds HIV-1 TAR RNA in vitro. <i>Biochem Biophys Res Commun</i> 1993; 196:935-42. Jeanson L, Subra F, Vaganay S, Hervy M, Marangoni E, Bourhis J, Mouscadet JF. Effect of Ku80 depletion on the preintegrative steps of HIV-1 replication in human cells. <i>Virology</i> 2002; 300:100-8.</p>
RHA	<p>- RHA is required to release RRE containing mRNA from spliceosomes before completion of splicing.</p>	<p>- Li J, Tang H, Mullen TM, Westberg C, Reddy TR, Rose DW, Wong-Staal F. A role for RNA helicase A in post-transcriptional regulation of HIV type 1. <i>PNAS</i> 1999; 96:709-14.</p>

Primer name	Sequence (5'-3')	Positions	Utilization
<b>Amplification for RNA preparation</b>			
O-3858	TTCAGACCCACCTCGAGGGGACCCG	Bru 7967-7999	Mutagenesis pUC18-SLS2 ΔK
O-3859	CGGGTCCCCTCGAGGTGGGTCTGAA	Bru 7967-7999	Mutagenesis pUC18-SLS2 ΔK
O-4705	GGTTGGGAGGTGGGTCTG	Bru 7969-7986	Mutagenesis pUC18-SLS2 ΔNCL
O-4706	AATAGAAGAAGAAGGTGG	Bru 8013-8030	Mutagenesis pUC18-SLS2 ΔNCL
O-4707	AGCTTAATACGACTCACTATAGGG		pUC18-T7-MS2
O-4708	TCGACCCTATAGTGAGTCGTATTA		pUC18-T7-MS2
O-4695	<b>GCTAGC</b> ATGGAACTGAACAGCC		hnRNP K cDNA amplification
O-4696	<b>GTCGAC</b> TTAGAATCCTTCAACAT		hnRNP K cDNA amplification
O-3100	CCCAAGCTTGGCATTAGGTGACACTATAGAAGAGTTTCAGACCCACCTC		SP6 promoter pUC18 cloning
O-3104	GCTCTAGAGTCTCTGTCTCTCTC		SLS2-A7 RNA
O-2143	GGCTTGCTGAAGCGCGCACGGCAAGAGG	pNL4-3 nts 700-727	<i>In cellulo</i> splicing assays
O-2144	TTGGGAGGTGGGTTGCTTTGATAGAG	pNL4-3 nts 8381-8369 and 6041-6029	<i>In cellulo</i> splicing assays
<b>2D structure of SLS1, 2, 3 stem-loop</b>			
O-2065	CGGGATCCTCTAATACGACTCACTATAGGGAGAATAGTTTTGCTGTAC	Bru 7900-7921	SLS1,2,3-A7 RNA
O-2382	CGTCCCAGATAAGTGC	Bru 8077-8092	SLS1,2,3-A7 RNA

Table S1: Oligonucleotides used in this study with indication of their target sites in HIV-1 Bru or pNL4.3 RNA and their utilization.

# **BIBLIOGRAPHIE**



- 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): 3090-8.
- Abbink, T. E., Ooms, M., Haasnoot, P. C. and Berkhout, B.** (2005). The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry* **44**(25): 9058-66.
- Abovich, N. and Rosbash, M.** (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**(3): 403-12.
- 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): 5789-802.
- Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K. and Dreyfuss, G.** (1986). mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol Cell Biol* **6**(8): 2932-43.
- Ainger, K., Avossa, D., Diana, A. S., Barry, C., Barbarese, E. and Carson, J. H.** (1997). Transport and localization elements in myelin basic protein mRNA. *J Cell Biol* **138**(5): 1077-87.
- Albini, A., Benelli, R., Presta, M., Rusnati, M., Ziche, M., Rubartelli, A., Paglialunga, G., Bussolino, F. and Noonan, D.** (1996). HIV-tat protein is a heparin-binding angiogenic growth factor. *Oncogene* **12**(2): 289-97.
- Allain, F. H., Bouvet, P., Dieckmann, T. and Feigon, J.** (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *Embo J* **19**(24): 6870-81.
- Amarasinghe, G. K., Zhou, J., Miskimon, M., Chancellor, K. J., McDonald, J. A., Matthews, A. G., Miller, R. R., Rouse, M. D. and Summers, M. F.** (2001). Stem-loop SL4 of the HIV-1 psi RNA packaging signal exhibits weak affinity for the nucleocapsid protein. structural studies and implications for genome recognition. *J Mol Biol* **314**(5): 961-70.
- Amendt, B. A., Si, Z. H. and Stoltzfus, C. M.** (1995). 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., Hesslein, D., Chang, L. J. and Stoltzfus, C. M.** (1994). Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. *Mol Cell Biol* **14**(6): 3960-70.
- 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): 1025-35.
- Arora, V. K., Fredericksen, B. L. and Garcia, J. V.** (2002). Nef: agent of cell subversion. *Microbes Infect* **4**(2): 189-99.
- Asai, K., Platt, C. and Cochrane, A.** (2003). Control of HIV-1 env RNA splicing and transport: investigating the role of hnRNP A1 in exon splicing silencer (ESS3a) function. *Virology* **314**(1): 229-42.
- Ashley, C. T., Jr., Wilkinson, K. D., Reines, D. and Warren, S. T.** (1993). FMR1 protein: conserved RNP family domains and selective RNA binding. *Science* **262**(5133): 563-6.
- Ayouba, A., Souquieres, S., Njinku, B., Martin, P. M., Muller-Trutwin, M. C., Roques, P., Barre-Sinoussi, F., Mauciere, P., Simon, F. and Nerrienet, E.** (2000). HIV-1 group N among HIV-1-seropositive individuals in Cameroon. *AIDS* **14**(16): 2623-5.

- Bach, M., Winkelmann, G. and Luhrmann, R.** (1989). 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc Natl Acad Sci U S A* **86**(16): 6038-42.
- Bach, M., Bringmann, P. and Luhrmann, R.** (1990). Purification of small nuclear ribonucleoprotein particles with antibodies against modified nucleosides of small nuclear RNAs. *Methods Enzymol* **181**: 232-57.
- Bacharach, E., Gonsky, J., Alin, K., Orlova, M. and Goff, S. P.** (2000). The carboxy-terminal fragment of nucleolin interacts with the nucleocapsid domain of retroviral gag proteins and inhibits virion assembly. *J Virol* **74**(23): 11027-39.
- Bagga, P. S., Arhin, G. K. and Wilusz, J.** (1998). DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. *Nucleic Acids Res* **26**(23): 5343-50.
- Bakkour, N., Lin, Y. L., Maire, S., Ayadi, L., Mahuteau-Betzer, F., Nguyen, C. H., Mettling, C., Portales, P., Grierson, D., Chabot, B., Jeanteur, P., Branlant, C., Corbeau, P. and Tazi, J.** (2007). Small-molecule inhibition of HIV pre-mRNA splicing as a novel antiretroviral therapy to overcome drug resistance. *PLoS Pathog* **3**(10): 1530-9.
- Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B. and Le Hir, H.** (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol* **12**(10): 861-9.
- Balvay, L., Libri, D., Gallego, M. and Fisman, M. Y.** (1992). Intronic sequence with both negative and positive effects on the regulation of alternative transcripts of the chicken beta tropomyosin transcripts. *Nucleic Acids Res* **20**(15): 3987-92.
- Bandziulis, R. J., Swanson, M. S. and Dreyfuss, G.** (1989). RNA-binding proteins as developmental regulators. *Genes Dev* **3**(4): 431-7.
- Barillari, G., Gendelman, R., Gallo, R. C. and Ensoli, B.** (1993). The Tat protein of human immunodeficiency virus type 1, a growth factor for AIDS Kaposi sarcoma and cytokine-activated vascular cells, induces adhesion of the same cell types by using integrin receptors recognizing the RGD amino acid sequence. *Proc Natl Acad Sci U S A* **90**(17): 7941-5.
- Barnard, D. C., Li, J., Peng, R. and Patton, J. G.** (2002). Regulation of alternative splicing by SRp86 through coactivation and repression of specific SR proteins. *Rna* **8**(4): 526-33.
- Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L.** (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**(4599): 868-71.
- Bartolome, C., Maside, X. and Charlesworth, B.** (2002). On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol* **19**(6): 926-37.
- Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. and Williamson, J. R.** (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**(5281): 1547-51.
- Baudin, F., Marquet, R., Isel, C., Darlix, J. L., Ehresmann, B. and Ehresmann, C.** (1993). Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J Mol Biol* **229**(2): 382-97.
- Bauren, G. and Wieslander, L.** (1994). Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76**(1): 183-92.

- Behrens, S. E. and Luhrmann, R.** (1991). Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev* **5**(8): 1439-52.
- Beil, B., Screatton, G. and Stamm, S.** (1997). Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing. *DNA Cell Biol* **16**(6): 679-90.
- Belenguer, P., Caizergues-Ferrer, M., Labbe, J. C., Doree, M. and Amalric, F.** (1990). Mitosis-specific phosphorylation of nucleolin by p34cdc2 protein kinase. *Mol Cell Biol* **10**(7): 3607-18.
- Bell, M., Schreiner, S., Damianov, A., Reddy, R. and Bindereif, A.** (2002). p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *Embo J* **21**(11): 2724-35.
- Benko, D. M., Schwartz, S., Pavlakis, G. N. and Felber, B. K.** (1990). A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. *J Virol* **64**(6): 2505-18.
- 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): 3165-75.
- Bentley, D.** (2002). The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol* **14**(3): 336-42.
- Berget, S. M.** (1995). Exon recognition in vertebrate splicing. *J Biol Chem* **270**(6): 2411-4.
- 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): 858-67.
- Bergman, C. M. and Kreitman, M.** (2001). Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res* **11**(8): 1335-45.
- Berkhout, B.** (1996). Structure and function of the human immunodeficiency virus leader RNA. *Prog Nucleic Acid Res Mol Biol* **54**: 1-34.
- Berthet-Colominas, C., Monaco, S., Novelli, A., Sibai, G., Mallet, F. and Cusack, S.** (1999). Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. *Embo J* **18**(5): 1124-36.
- Beyer, A. L. and Osheim, Y. N.** (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev* **2**(6): 754-65.
- Beyer, A. L., Christensen, M. E., Walker, B. W. and LeSturgeon, W. M.** (1977). Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell* **11**(1): 127-38.
- 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): 8487-97.
- Birney, E., Kumar, S. and Krainer, A. R.** (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* **21**(25): 5803-16.
- Biswas, D. K., Salas, T. R., Wang, F., Ahlers, C. M., Dezube, B. J. and Pardee, A. B.** (1995). A Tat-induced auto-up-regulatory loop for superactivation of the human immunodeficiency virus type 1 promoter. *J Virol* **69**(12): 7437-44.
- Black, D. L.** (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**: 291-336.

- Blanchette, A. R., Fuentes Medel, Y. F. and Gardner, P. D.** (2006). Cell-type-specific and developmental regulation of heterogeneous nuclear ribonucleoprotein K mRNA in the rat nervous system. *Gene Expr Patterns* **6**(6): 596-606.
- 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): 1939-52.
- Blencowe, B. J.** (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* **25**(3): 106-10.
- Blencowe, B. J.** (2006). Alternative splicing: new insights from global analyses. *Cell* **126**(1): 37-47.
- Blencowe, B. J., Issner, R., Nickerson, J. A. and Sharp, P. A.** (1998). A coactivator of pre-mRNA splicing. *Genes Dev* **12**(7): 996-1009.
- Blencowe, B. J., Bowman, J. A., McCracken, S. and Rosonina, E.** (1999). SR-related proteins and the processing of messenger RNA precursors. *Biochem Cell Biol* **77**(4): 277-91.
- Bogerd, H. P., Echarri, A., Ross, T. M. and Cullen, B. R.** (1998). Inhibition of human immunodeficiency virus Rev and human T-cell leukemia virus Rex function, but not Mason-Pfizer monkey virus constitutive transport element activity, by a mutant human nucleoporin targeted to Crm1. *J Virol* **72**(11): 8627-35.
- 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): 739-47.
- Bohne, J., Wodrich, H. and Krausslich, H. G.** (2005). Splicing of human immunodeficiency virus RNA is position-dependent suggesting sequential removal of introns from the 5' end. *Nucleic Acids Res* **33**(3): 825-37.
- Bomsztyk, K., Denisenko, O. and Ostrowski, J.** (2004). hnRNP K: one protein multiple processes. *Bioessays* **26**(6): 629-38.
- Bomsztyk, K., Van Seuning, I., Suzuki, H., Denisenko, O. and Ostrowski, J.** (1997). Diverse molecular interactions of the hnRNP K protein. *FEBS Lett* **403**(2): 113-5.
- Bonifaci, N., Moroianu, J., Radu, A. and Blobel, G.** (1997). Karyopherin beta2 mediates nuclear import of a mRNA binding protein. *Proc Natl Acad Sci U S A* **94**(10): 5055-60.
- Bono, F., Ebert, J., Lorentzen, E. and Conti, E.** (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**(4): 713-25.
- Bono, F., Ebert, J., Unterholzner, L., Guttler, T., Izaurralde, E. and Conti, E.** (2004). Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. *EMBO Rep* **5**(3): 304-10.
- Bouvet, P., Allain, F. H., Finger, L. D., Dieckmann, T. and Feigon, J.** (2001). Recognition of pre-formed and flexible elements of an RNA stem-loop by nucleolin. *J Mol Biol* **309**(3): 763-75.
- Branlant, C., Krol, A., Ebel, J. P., Gallinaro, H., Lazar, E. and Jacob, M.** (1981). The conformation of chicken, rat and human U1A RNAs in solution. *Nucleic Acids Res* **9**(4): 841-58.
- Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Haendler, B. and Jacob, M.** (1982). U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *Embo J* **1**(10): 1259-65.
- Branlant, C., Krol, A., Lazar, E., Haendler, B., Jacob, M., Galego-Dias, L. and Pousada, C.** (1983). High evolutionary conservation of the secondary structure and of certain nucleotide sequences of U5 RNA. *Nucleic Acids Res* **11**(23): 8359-67.

- Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J. and Jeanteur, P.** (1980). Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. *Nucleic Acids Res* **8**(18): 4143-54.
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P.** (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc Natl Acad Sci U S A* **75**(10): 4853-7.
- 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): 1211-26.
- Briggs, J. A., Simon, M. N., Gross, I., Krausslich, H. G., Fuller, S. D., Vogt, V. M. and Johnson, M. C.** (2004). The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol* **11**(7): 672-5.
- Brosi, R., Hauri, H. P. and Kramer, A.** (1993). Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem* **268**(23): 17640-6.
- Brow, D. A. and Guthrie, C.** (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* **334**(6179): 213-8.
- Buonaguro, L., Buonaguro, F. M., Giraldo, G. and Ensoli, B.** (1994). The human immunodeficiency virus type 1 Tat protein transactivates tumor necrosis factor beta gene expression through a TAR-like structure. *J Virol* **68**(4): 2677-82.
- Buratti, E., Muro, A. F., Giombi, M., Gherbassi, D., Iaconcig, A. and Baralle, F. E.** (2004). RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol Cell Biol* **24**(3): 1387-400.
- Burd, C. G. and Dreyfuss, G.** (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**(5172): 615-21.
- Burd, C. G., Swanson, M. S., Gorlach, M. and Dreyfuss, G.** (1989). Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts. *Proc Natl Acad Sci U S A* **86**(24): 9788-92.
- Burnham, A. J., Gong, L. and Hardy, R. W.** (2007). Heterogeneous nuclear ribonuclear protein K interacts with Sindbis virus nonstructural proteins and viral subgenomic mRNA. *Virology* **367**(1): 212-21.
- 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): 1706-9.
- Calero, G., Wilson, K. F., Ly, T., Rios-Steiner, J. L., Clardy, J. C. and Cerione, R. A.** (2002). Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. *Nat Struct Biol* **9**(12): 912-7.
- Campbell, G. R., Watkins, J. D., Esquieu, D., Pasquier, E., Loret, E. P. and Spector, S. A.** (2005). The C terminus of HIV-1 Tat modulates the extent of CD178-mediated apoptosis of T cells. *J Biol Chem* **280**(46): 38376-82.
- Campillos, M., Lamas, J. R., Garcia, M. A., Bullido, M. J., Valdivieso, F. and Vazquez, J.** (2003). Specific interaction of heterogeneous nuclear ribonucleoprotein A1 with the -219T allelic form modulates APOE promoter activity. *Nucleic Acids Res* **31**(12): 3063-70.
- 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): 43850-9.

- Caputi, M. and Zahler, A. M.** (2002). SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *Embo J* **21**(4): 845-55.
- 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): 4060-7.
- Caputi, M., Freund, M., Kammler, S., Asang, C. and Schaal, H.** (2004). A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *J Virol* **78**(12): 6517-26.
- 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): 468-83.
- 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): 1776-86.
- 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.
- Chai, Q., Zheng, L., Zhou, M., Turchi, J. J. and Shen, B.** (2003). Interaction and stimulation of human FEN-1 nuclease activities by heterogeneous nuclear ribonucleoprotein A1 in alpha-segment processing during Okazaki fragment maturation. *Biochemistry* **42**(51): 15045-52.
- Chan, C. C., Dostie, J., Diem, M. D., Feng, W., Mann, M., Rappsilber, J. and Dreyfuss, G.** (2004). eIF4A3 is a novel component of the exon junction complex. *Rna* **10**(2): 200-9.
- Chan, D. C., Chutkowski, C. T. and Kim, P. S.** (1998). Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc Natl Acad Sci U S A* **95**(26): 15613-7.
- Chang, H. K., Gallo, R. C. and Ensoli, B.** (1995). Regulation of Cellular Gene Expression and Function by the Human Immunodeficiency Virus Type 1 Tat Protein. *J Biomed Sci* **2**(3): 189-202.
- Chang, J. W., Koike, T. and Iwashima, M.** (2009). hnRNP-K is a nuclear target of TCR-activated ERK and required for T-cell late activation. *Int Immunol* **21**(12): 1351-61.
- Chang, Y. F., Imam, J. S. and Wilkinson, M. F.** (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* **76**: 51-74.
- Charnay, N., Ivanyi-Nagy, R., Soto-Rifo, R., Ohlmann, T., Lopez-Lastra, M. and Darlix, J. L.** (2009). Mechanism of HIV-1 Tat RNA translation and its activation by the Tat protein. *Retrovirology* **6**: 74.
- Charneau, P., Alizon, M. and Clavel, F.** (1992). A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *J Virol* **66**(5): 2814-20.
- 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. H., Chang, J. G., Lu, R. M., Peng, T. Y. and Tarn, W. Y.** (2008). The RNA binding protein hnRNP Q modulates the utilization of exon 7 in the survival motor neuron 2 (SMN2) gene. *Mol Cell Biol* **28**(22): 6929-38.
- 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): 1389-400.

- Chiara, M. D., Gozani, O., Bennett, M., Champion-Arnaud, P., Palandjian, L. and Reed, R.** (1996). Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Mol Cell Biol* **16**(7): 3317-26.
- Chou, M. Y., Rooke, N., Turck, C. W. and Black, D. L.** (1999). hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol Cell Biol* **19**(1): 69-77.
- Chou, T. B., Zachar, Z. and Bingham, P. M.** (1987). Developmental expression of a regulatory gene is programmed at the level of splicing. *Embo J* **6**(13): 4095-104.
- Clery, A., Blatter, M. and Allain, F. H.** (2008). RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **18**(3): 290-8.
- Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V. and Wilson, S. H.** (1988). Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. Large-scale overproduction in Escherichia coli and cooperative binding to single-stranded nucleic acids. *J Biol Chem* **263**(2): 1063-71.
- Cohen, J. B., Snow, J. E., Spencer, S. D. and Levinson, A. D.** (1994). Suppression of mammalian 5' splice-site defects by U1 small nuclear RNAs from a distance. *Proc Natl Acad Sci U S A* **91**(22): 10470-4.
- 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): 8196-205.
- Conte, M. R., Grune, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S. and Curry, S.** (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *Embo J* **19**(12): 3132-41.
- Cowper, A. E., Caceres, J. F., Mayeda, A. and Screaton, G. R.** (2001). Serine-arginine (SR) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing. *J Biol Chem* **276**(52): 48908-14.
- Cramer, P., Caceres, J. F., Cazalla, D., Kadener, S., Muro, A. F., Baralle, F. E. and Kornblihtt, A. R.** (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell* **4**(2): 251-8.
- Crowder, S. M., Kanaar, R., Rio, D. C. and Alber, T.** (1999). Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. *Proc Natl Acad Sci U S A* **96**(9): 4892-7.
- Cui, Y., Hagan, K. W., Zhang, S. and Peltz, S. W.** (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes Dev* **9**(4): 423-36.
- Cullen, B. R.** (1986). Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell* **46**(7): 973-82.
- Cullen, B. R.** (1991). Regulation of human immunodeficiency virus replication. *Annu Rev Microbiol* **45**: 219-50.
- Cullen, B. R.** (1998). HIV-1 auxiliary proteins: making connections in a dying cell. *Cell* **93**(5): 685-92.
- Currie, J. R. and Brown, W. T.** (1999). KH domain-containing proteins of yeast: absence of a fragile X gene homologue. *Am J Med Genet* **84**(3): 272-6.
- Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F. and Weiss, R. A.** (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**(5996): 763-7.

- 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): 1401-15.
- Damgaard, C. K., Andersen, E. S., Knudsen, B., Gorodkin, J. and Kjems, J.** (2004). RNA interactions in the 5' region of the HIV-1 genome. *J Mol Biol* **336**(2): 369-79.
- Damier, L.** (1997). Etude de la régulation de l'épissage du transcrit primaire du virus de l'immunodéficience humaine de type 1, HIV-1. UFR des Sciences et Techniques Biologiques. Vandoeuvre les Nancy, Université Henri Poincaré.
- 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): 182-7.
- Darnell, R. B.** (2006). Developing global insight into RNA regulation. *Cold Spring Harb Symp Quant Biol* **71**: 321-7.
- Das, R., Zhou, Z. and Reed, R.** (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell* **5**(5): 779-87.
- 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): 1100-9.
- 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): 867-81.
- Dauksaite, V. and Akusjarvi, G.** (2002). Human splicing factor ASF/SF2 encodes for a repressor domain required for its inhibitory activity on pre-mRNA splicing. *J Biol Chem* **277**(15): 12579-86.
- Dauksaite, V. and Akusjarvi, G.** (2004). The second RNA-binding domain of the human splicing factor ASF/SF2 is the critical domain controlling adenovirus E1A alternative 5'-splice site selection. *Biochem J* **381**(Pt 2): 343-50.
- de Alava, E. and Gerald, W. L.** (2000). Molecular biology of the Ewing's sarcoma/primitive neuroectodermal tumor family. *J Clin Oncol* **18**(1): 204-13.
- 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): 5528-43.
- Degot, S., Le Hir, H., Alpy, F., Kedinger, V., Stoll, I., Wendling, C., Seraphin, B., Rio, M. C. and Tomasetto, C.** (2004). Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module. *J Biol Chem* **279**(32): 33702-15.
- Del Gatto-Konczak, F., Olive, M., Gesnel, M. C. and Breathnach, R.** (1999). hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol Cell Biol* **19**(1): 251-60.
- 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): 6287-99.
- Del Gatto, F. and Breathnach, R.** (1995). Exon and intron sequences, respectively, repress and activate splicing of a fibroblast growth factor receptor 2 alternative exon. *Mol Cell Biol* **15**(9): 4825-34.
- Delelis, O., Carayon, K., Guiot, E., Leh, H., Tauc, P., Brochon, J. C., Mouscadet, J. F. and Deprez, E.** (2008). Insight into the integrase-DNA recognition mechanism. A

- specific DNA-binding mode revealed by an enzymatically labeled integrase. *J Biol Chem* **283**(41): 27838-49.
- Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R. and Landau, N. R.** (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**(6584): 661-6.
- Denisenko, O. N. and Bomsztyk, K.** (1997). The product of the murine homolog of the *Drosophila* extra sex combs gene displays transcriptional repressor activity. *Mol Cell Biol* **17**(8): 4707-17.
- Denisenko, O. N., O'Neill, B., Ostrowski, J., Van Seuning, I. and Bomsztyk, K.** (1996). Zik1, a transcriptional repressor that interacts with the heterogeneous nuclear ribonucleoprotein particle K protein. *J Biol Chem* **271**(44): 27701-6.
- di Marzo Veronese, F., Copeland, T. D., DeVico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C. and Sarngadharan, M. G.** (1986). Characterization of highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. *Science* **231**(4743): 1289-91.
- Diem, M. D., Chan, C. C., Younis, I. and Dreyfuss, G.** (2007). PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat Struct Mol Biol* **14**(12): 1173-9.
- Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R. and Xu, R. M.** (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* **13**(9): 1102-15.
- 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): 8762-72.
- Dredge, B. K. and Darnell, R. B.** (2003). Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. *Mol Cell Biol* **23**(13): 4687-700.
- Dreyfuss, G., Swanson, M. S. and Pinol-Roma, S.** (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci* **13**(3): 86-91.
- Dreyfuss, G., Kim, V. N. and Kataoka, N.** (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* **3**(3): 195-205.
- Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. and Burd, C. G.** (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**: 289-321.
- 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.
- Du, Q., Melnikova, I. N. and Gardner, P. D.** (1998). Differential effects of heterogeneous nuclear ribonucleoprotein K on Sp1- and Sp3-mediated transcriptional activation of a neuronal nicotinic acetylcholine receptor promoter. *J Biol Chem* **273**(31): 19877-83.
- Dumler, I., Stepanova, V., Jerke, U., Mayboroda, O. A., Vogel, F., Bouvet, P., Tkachuk, V., Haller, H. and Gulba, D. C.** (1999). Urokinase-induced mitogenesis is mediated by casein kinase 2 and nucleolin. *Curr Biol* **9**(24): 1468-76.
- Durova, O. M., Vorobiev, II, Smirnov, I. V., Reshetnyak, A. V., Telegin, G. B., Shamborant, O. G., Orlova, N. A., Genkin, D. D., Bacon, A., Ponomarenko, N. A., Friboulet, A. and Gabibov, A. G.** (2009). Strategies for induction of catalytic antibodies toward HIV-1 glycoprotein gp120 in autoimmune prone mice. *Mol Immunol*.

- Dyhr-Mikkelsen, H. and Kjems, J.** (1995). Inefficient spliceosome assembly and abnormal branch site selection in splicing of an HIV-1 transcript in vitro. *J Biol Chem* **270**(41): 24060-6.
- Edery, I., Petryshyn, R. and Sonenberg, N.** (1989). Activation of double-stranded RNA-dependent kinase (dsl) by the TAR region of HIV-1 mRNA: a novel translational control mechanism. *Cell* **56**(2): 303-12.
- 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): 6125-30.
- Ensoli, B., Barillari, G., Salahuddin, S. Z., Gallo, R. C. and Wong-Staal, F.** (1990). Tat protein of HIV-1 stimulates growth of cells derived from Kaposi's sarcoma lesions of AIDS patients. *Nature* **345**(6270): 84-6.
- Ensoli, B., Buonaguro, L., Barillari, G., Fiorelli, V., Gendelman, R., Morgan, R. A., Wingfield, P. and Gallo, R. C.** (1993). Release, uptake, and effects of extracellular human immunodeficiency virus type 1 Tat protein on cell growth and viral transactivation. *J Virol* **67**(1): 277-87.
- 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): 3607-17.
- Evans, J. R., Mitchell, S. A., Spriggs, K. A., Ostrowski, J., Bomsztyk, K., Ostarek, D. and Willis, A. E.** (2003). Members of the poly (rC) binding protein family stimulate the activity of the c-myc internal ribosome entry segment in vitro and in vivo. *Oncogene* **22**(39): 8012-20.
- 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): 3921-31.
- Expert-Bezancon, A., Le Caer, J. P. and Marie, J.** (2002). Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken beta-tropomyosin pre-mRNA. *J Biol Chem* **277**(19): 16614-23.
- Feng, Y., Broder, C. C., Kennedy, P. E. and Berger, E. A.** (1996). HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**(5263): 872-7.
- Ferraiuolo, M. A., Lee, C. S., Ler, L. W., Hsu, J. L., Costa-Mattioli, M., Luo, M. J., Reed, R. and Sonenberg, N.** (2004). A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. *Proc Natl Acad Sci U S A* **101**(12): 4118-23.
- 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): 344-55.
- Finger, L. D., Johansson, C., Rinaldi, B., Bouvet, P. and Feigon, J.** (2004). Contributions of the RNA-binding and linker domains and RNA structure to the specificity and affinity of the nucleolin RBD12/NRE interaction. *Biochemistry* **43**(22): 6937-47.
- Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W. and Luhrmann, R.** (1995). The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**(3): 475-83.
- Fisette, J. F., Toutant, J., Dugre-Brisson, S., Desgroseillers, L. and Chabot, B.** (2009). hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection. *Rna*.

- 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): 1864-72.
- Fong, Y. W. and Zhou, Q.** (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**(6866): 929-33.
- Fornerod, M., Ohno, M., Yoshida, M. and Mattaj, I. W.** (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**(6): 1051-60.
- 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): 16176-81.
- Frankel, A. D. and Young, J. A.** (1998). HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **67**: 1-25.
- Fribourg, S., Gatfield, D., Izaurralde, E. and Conti, E.** (2003). A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat Struct Biol* **10**(6): 433-9.
- Fridell, R. A., Truant, R., Thorne, L., Benson, R. E. and Cullen, B. R.** (1997). Nuclear import of hnRNP A1 is mediated by a novel cellular cofactor related to karyopherin-beta. *J Cell Sci* **110** ( Pt 11): 1325-31.
- Fu, X. D.** (1995). The superfamily of arginine/serine-rich splicing factors. *RNA* **1**(7): 663-80.
- Fu, X. D. and Maniatis, T.** (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**(6257): 437-41.
- Furger, A., O'Sullivan, J. M., Binnie, A., Lee, B. A. and Proudfoot, N. J.** (2002). Promoter proximal splice sites enhance transcription. *Genes Dev* **16**(21): 2792-9.
- Galati, D., Paiardini, M., Cervasi, B., Albrecht, H., Bocchino, M., Costantini, A., Montroni, M., Magnani, M., Piedimonte, G. and Silvestri, G.** (2003). Specific changes in the posttranslational regulation of nucleolin in lymphocytes from patients infected with human immunodeficiency virus. *J Infect Dis* **188**(10): 1483-91.
- Gallego, M. E., Balvay, L. and Brody, E.** (1992). cis-acting sequences involved in exon selection in the chicken beta-tropomyosin gene. *Mol Cell Biol* **12**(12): 5415-25.
- Gallego, M. E., Gattoni, R., Stevenin, J., Marie, J. and Expert-Bezancon, A.** (1997). The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A. *Embo J* **16**(7): 1772-84.
- Gallo, R. C. and Montagnier, L.** (2002). Historical essay. Prospects for the future. *Science* **298**(5599): 1730-1.
- Ganot, P., Jady, B. E., Bortolin, M. L., Darzacq, X. and Kiss, T.** (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol* **19**(10): 6906-17.
- Garber, M. E., Wei, P. and Jones, K. A.** (1998). HIV-1 Tat interacts with cyclin T1 to direct the P-TEFb CTD kinase complex to TAR RNA. *Cold Spring Harb Symp Quant Biol* **63**: 371-80.
- Garber, M. E., Mayall, T. P., Suess, E. M., Meisenhelder, J., Thompson, N. E. and Jones, K. A.** (2000). CDK9 autophosphorylation regulates high-affinity binding of the human immunodeficiency virus type 1 tat-P-TEFb complex to TAR RNA. *Mol Cell Biol* **20**(18): 6958-69.
- Garneau, D., Revil, T., Fiset, J. F. and Chabot, B.** (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem* **280**(24): 22641-50.
- Gatignol, A.** (2007). Transcription of HIV: Tat and cellular chromatin. *Adv Pharmacol* **55**: 137-59.

- Gaynor, R.** (1992). Cellular transcription factors involved in the regulation of HIV-1 gene expression. *Aids* **6**(4): 347-63.
- Ge, H., Zuo, P. and Manley, J. L.** (1991). Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell* **66**(2): 373-82.
- Ghetti, A., Pinol-Roma, S., Michael, W. M., Morandi, C. and Dreyfuss, G.** (1992). hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res* **20**(14): 3671-8.
- Gheysen, D., Jacobs, E., de Foresta, F., Thiriart, C., Francotte, M., Thines, D. and De Wilde, M.** (1989). Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell* **59**(1): 103-12.
- Ghisolfi, L., Kharrat, A., Joseph, G., Amalric, F. and Erard, M.** (1992). Concerted activities of the RNA recognition and the glycine-rich C-terminal domains of nucleolin are required for efficient complex formation with pre-ribosomal RNA. *Eur J Biochem* **209**(2): 541-8.
- Ginisty, H., Amalric, F. and Bouvet, P.** (2001). Two different combinations of RNA-binding domains determine the RNA binding specificity of nucleolin. *J Biol Chem* **276**(17): 14338-43.
- Ginisty, H., Sicard, H., Roger, B. and Bouvet, P.** (1999). Structure and functions of nucleolin. *J Cell Sci* **112** ( Pt 6): 761-72.
- Ginisty, H., Serin, G., Ghisolfi-Nieto, L., Roger, B., Libante, V., Amalric, F. and Bouvet, P.** (2000). Interaction of nucleolin with an evolutionarily conserved pre-ribosomal RNA sequence is required for the assembly of the primary processing complex. *J Biol Chem* **275**(25): 18845-50.
- Giorgi, C. and Moore, M. J.** (2007). The nuclear nurture and cytoplasmic nature of localized mRNPs. *Semin Cell Dev Biol* **18**(2): 186-93.
- Goff, S. P.** (1990). Retroviral reverse transcriptase: synthesis, structure, and function. *J Acquir Immune Defic Syndr* **3**(8): 817-31.
- Good, P. J., Chen, Q., Warner, S. J. and Herring, D. C.** (2000). A family of human RNA-binding proteins related to the Drosophila Bruno translational regulator. *J Biol Chem* **275**(37): 28583-92.
- Goux-Pelletan, M., Libri, D., d'Aubenton-Carafa, Y., Fiszman, M., Brody, E. and Marie, J.** (1990). In vitro splicing of mutually exclusive exons from the chicken beta-tropomyosin gene: role of the branch point location and very long pyrimidine stretch. *Embo J* **9**(1): 241-9.
- 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): 4752-60.
- Graveley, B. R.** (2000). Sorting out the complexity of SR protein functions. *RNA* **6**(9): 1197-211.
- Graveley, B. R., Hertel, K. J. and Maniatis, T.** (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA* **7**(6): 806-18.
- Greene, W. C. and Peterlin, B. M.** (2002). Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nat Med* **8**(7): 673-80.
- Griffith, B. N., Walsh, C. M., Szeszel-Fedorowicz, W., Timperman, A. T. and Salati, L. M.** (2006). Identification of hnRNPs K, L and A2/B1 as candidate proteins involved in the nutritional regulation of mRNA splicing. *Biochim Biophys Acta* **1759**(11-12): 552-61.
- Grishin, N. V.** (2001). KH domain: one motif, two folds. *Nucleic Acids Res* **29**(3): 638-43.

- Guatelli, J. C., Gingeras, T. R. and Richman, D. D.** (1990). Alternative splice acceptor utilization during human immunodeficiency virus type 1 infection of cultured cells. *J Virol* **64**(9): 4093-8.
- Hadian, K., Vincendeau, M., Mausbacher, N., Nagel, D., Hauck, S. M., Ueffing, M., Loyter, A., Werner, T., Wolff, H. and Brack-Werner, R.** (2009). Identification of a heterogeneous nuclear ribonucleoprotein-recognition region in the HIV Rev protein. *J Biol Chem* **284**(48): 33384-91.
- 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): 37159-74.
- Hamilton, B. J., Burns, C. M., Nichols, R. C. and Rigby, W. F.** (1997). Modulation of AUUUA response element binding by heterogeneous nuclear ribonucleoprotein A1 in human T lymphocytes. The roles of cytoplasmic location, transcription, and phosphorylation. *J Biol Chem* **272**(45): 28732-41.
- Hamilton, B. J., Nagy, E., Malter, J. S., Arrick, B. A. and Rigby, W. F.** (1993). Association of heterogeneous nuclear ribonucleoprotein A1 and C proteins with reiterated AUUUA sequences. *J Biol Chem* **268**(12): 8881-7.
- Hanahan, D.** (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**(4): 557-80.
- 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): 430-44.
- 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): 5126-37.
- 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): 16719-24.
- Heguy, A.** (1997). Inhibition of the HIV Rev transactivator : a new target for therapeutic intervention. *Front Biosci* **2**: d283-97.
- Henics, T., Sanfridson, A., Hamilton, B. J., Nagy, E. and Rigby, W. F.** (1994). Enhanced stability of interleukin-2 mRNA in MLA 144 cells. Possible role of cytoplasmic AU-rich sequence-binding proteins. *J Biol Chem* **269**(7): 5377-83.
- Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H. and Luhrmann, R.** (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *Embo J* **14**(9): 2076-88.
- Hirose, Y., Tacke, R. and Manley, J. L.** (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev* **13**(10): 1234-9.
- Hodnett, J. L. and Busch, H.** (1968). Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei. *J Biol Chem* **243**(24): 6334-42.
- Honore, B., Rasmussen, H. H., Vorum, H., Dejgaard, K., Liu, X., Gromov, P., Madsen, P., Gesser, B., Tommerup, N. and Celis, J. E.** (1995). Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. *J Biol Chem* **270**(48): 28780-9.
- Hope, T. J.** (1999). The ins and outs of HIV Rev. *Arch Biochem Biophys* **365**(2): 186-91.

- Houzet, L., Paillart, J. C., Smagulova, F., Maurel, S., Morichaud, Z., Marquet, R. and Mougel, M. (2007). HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Res* **35**(8): 2695-704.
- Hovhannisyann, R. H. and Carstens, R. P. (2007). Heterogeneous ribonucleoprotein m is a splicing regulatory protein that can enhance or silence splicing of alternatively spliced exons. *J Biol Chem* **282**(50): 36265-74.
- Huang, P. R., Tsai, S. T., Hsieh, K. H. and Wang, T. C. (2008). Heterogeneous nuclear ribonucleoprotein A3 binds single-stranded telomeric DNA and inhibits telomerase extension in vitro. *Biochim Biophys Acta* **1783**(2): 193-202.
- 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): 9666-70.
- Huang, Y., Mak, J., Cao, Q., Li, Z., Wainberg, M. A. and Kleiman, L. (1994). Incorporation of excess wild-type and mutant tRNA(3Lys) into human immunodeficiency virus type 1. *J Virol* **68**(12): 7676-83.
- Hui, J., Hung, L. H., Heiner, M., Schreiner, S., Neumuller, N., Reither, G., Haas, S. A. and Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *Embo J* **24**(11): 1988-98.
- Hung, L. H., Heiner, M., Hui, J., Schreiner, S., Benes, V. and Bindereif, A. (2008). Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *Rna* **14**(2): 284-96.
- 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): 29745-52.
- Ibrahim, E. C., Schaal, T. D., Hertel, K. J., Reed, R. and Maniatis, T. (2005). Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc Natl Acad Sci U S A* **102**(14): 5002-7.
- Irimura, S., Kitamura, K., Kato, N., Saiki, K., Takeuchi, A., Gunadi, Matsuo, M., Nishio, H. and Lee, M. J. (2009). HnRNP C1/C2 may regulate exon 7 splicing in the spinal muscular atrophy gene SMN1. *Kobe J Med Sci* **54**(5): E227-36.
- Ishikawa, F., Matunis, M. J., Dreyfuss, G. and Cech, T. R. (1993). Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. *Mol Cell Biol* **13**(7): 4301-10.
- 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): 314-27.
- Iwanaga, K., Sueoka, N., Sato, A., Hayashi, S. and Sueoka, E. (2005). Heterogeneous nuclear ribonucleoprotein B1 protein impairs DNA repair mediated through the inhibition of DNA-dependent protein kinase activity. *Biochem Biophys Res Commun* **333**(3): 888-95.
- Jablonski, J. A. and Caputi, M. (2009). Role of cellular RNA processing factors in human immunodeficiency virus type 1 mRNA metabolism, replication, and infectivity. *J Virol* **83**(2): 981-92.
- Jacquet, S., Decimo, D., Muriaux, D. and Darlix, J. L. (2005). Dual effect of the SR proteins ASF/SF2, SC35 and 9G8 on HIV-1 RNA splicing and virion production. *Retrovirology* **2**: 33.
- Jacquet, S., Mereau, A., Bilodeau, P. S., Damier, L., Stoltzfus, C. M. and Branlant, C. (2001). A second exon splicing silencer within human immunodeficiency virus type 1

- tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J Biol Chem* **276**(44): 40464-75.
- Jacquet, 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): 464-78.
- Jumaa, H. and Nielsen, P. J.** (1997). The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *Embo J* **16**(16): 5077-85.
- Jurica, M. S. and Moore, M. J.** (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12**(1): 5-14.
- Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. and Moore, M. J.** (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *Rna* **8**(4): 426-39.
- Kadlec, J., Izaurralde, E. and Cusack, S.** (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol* **11**(4): 330-7.
- Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Luhrmann, R., Li, J. and Nagai, K.** (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**(3): 375-87.
- 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): 1782-91.
- Kamma, H., Portman, D. S. and Dreyfuss, G.** (1995). Cell type-specific expression of hnRNP proteins. *Exp Cell Res* **221**(1): 187-96.
- 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.
- Kammler, S., Leurs, C., Freund, M., Krummheuer, J., Seidel, K., Tange, T. O., Lund, M. K., Kjems, J., Scheid, A. and Schaal, H.** (2001). The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. *Rna* **7**(3): 421-34.
- Kanopka, A., Muhlemann, O. and Akusjarvi, G.** (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**(6582): 535-8.
- Karacostas, V., Wolffe, E. J., Nagashima, K., Gonda, M. A. and Moss, B.** (1993). Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* **193**(2): 661-71.
- Karn, J.** (1999). Tackling Tat. *J Mol Biol* **293**(2): 235-54.
- Kataoka, N., Diem, M. D., Kim, V. N., Yong, J. and Dreyfuss, G.** (2001). Magoh, a human homolog of *Drosophila mago nashi* protein, is a component of the splicing-dependent exon-exon junction complex. *Embo J* **20**(22): 6424-33.
- Kataoka, N., Yong, J., Kim, V. N., Velazquez, F., Perkinson, R. A., Wang, F. and Dreyfuss, G.** (2000). Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* **6**(3): 673-82.
- Kenan, D. J., Query, C. C. and Keene, J. D.** (1991). RNA recognition: towards identifying determinants of specificity. *Trends Biochem Sci* **16**(6): 214-20.
- Khorchid, A., Halwani, R., Wainberg, M. A. and Kleiman, L.** (2002). Role of RNA in facilitating Gag/Gag-Pol interaction. *J Virol* **76**(8): 4131-7.

- Khorchid, A., Javanbakht, H., Wise, S., Halwani, R., Parniak, M. A., Wainberg, M. A. and Kleiman, L.** (2000). Sequences within Pr160gag-pol affecting the selective packaging of primer tRNA(Lys3) into HIV-1. *J Mol Biol* **299**(1): 17-26.
- Kielkopf, C. L., Lucke, S. and Green, M. R.** (2004). U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* **18**(13): 1513-26.
- 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.
- 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): 2655-64.
- King, J. E., Eugenin, E. A., Buckner, C. M. and Berman, J. W.** (2006). HIV tat and neurotoxicity. *Microbes Infect* **8**(5): 1347-57.
- Kingsman, S. M. and Kingsman, A. J.** (1996). The regulation of human immunodeficiency virus type-1 gene expression. *Eur J Biochem* **240**(3): 491-507.
- Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C. and Montagnier, L.** (1984). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**(5996): 767-8.
- Kleiman, L. and Cen, S.** (2004). The tRNALys packaging complex in HIV-1. *Int J Biochem Cell Biol* **36**(9): 1776-86.
- Klimek-Tomczak, K., Wyrwicz, L. S., Jain, S., Bomsztyk, K. and Ostrowski, J.** (2004). Characterization of hnRNP K protein-RNA interactions. *J Mol Biol* **342**(4): 1131-41.
- Kofuji, P., Lederer, W. J. and Schulze, D. H.** (1994). Mutually exclusive and cassette exons underlie alternatively spliced isoforms of the Na/Ca exchanger. *J Biol Chem* **269**(7): 5145-9.
- 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): 119-24.
- Koradi, R., Billeter, M. and Wuthrich, K.** (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**(1): 51-5, 29-32.
- Korber, B. (1998). Numbering positions in HIV relative to HXB2CG. Human retroviruses and AIDS. L. A. Los Alamos National Laboratory, NM: p. 102-11.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. and Bhattacharya, T.** (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**(5472): 1789-96.
- Kornblihtt, A. R.** (2005). Promoter usage and alternative splicing. *Curr Opin Cell Biol* **17**(3): 262-8.
- Krainer, A. R., Mayeda, A., Kozak, D. and Binns, G.** (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell* **66**(2): 383-94.
- Kramer, A., Gruter, P., Groning, K. and Kastner, B.** (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. *J Cell Biol* **145**(7): 1355-68.
- Krecic, A. M. and Swanson, M. S.** (1999). hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11**(3): 363-71.
- Krol, A., Branlant, C., Lazar, E., Gallinaro, H. and Jacob, M.** (1981). Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. *Nucleic Acids Res* **9**(12): 2699-716.

- Kumar, A., Williams, K. R. and Szer, W.** (1986). Purification and domain structure of core hnRNP proteins A1 and A2 and their relationship to single-stranded DNA-binding proteins. *J Biol Chem* **261**(24): 11266-73.
- Kuppuswamy, M., Subramanian, T., Srinivasan, A. and Chinnadurai, G.** (1989). Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis. *Nucleic Acids Res* **17**(9): 3551-61.
- Kuroyanagi, H.** (2009). Fox-1 family of RNA-binding proteins. *Cell Mol Life Sci.*
- Kwek, K. Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N. J. and Akoulitchev, A.** (2002). U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat Struct Biol* **9**(11): 800-5.
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. and Hendrickson, W. A.** (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**(6686): 648-59.
- Ladd, A. N. and Cooper, T. A.** (2004). Multiple domains control the subcellular localization and activity of ETR-3, a regulator of nuclear and cytoplasmic RNA processing events. *J Cell Sci* **117**(Pt 16): 3519-29.
- Ladd, A. N., Charlet, N. and Cooper, T. A.** (2001). The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol Cell Biol* **21**(4): 1285-96.
- Lamond, A. I. and Spector, D. L.** (2003). Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4**(8): 605-12.
- Larrucea, S., Cambroner, R., Gonzalez-Rubio, C., Fraile, B., Gamallo, C., Fontan, G. and Lopez-Trascasa, M.** (1999). Internalization of factor J and cellular signalization after factor J-cell interaction. *Biochem Biophys Res Commun* **266**(1): 51-7.
- Lau, C. K., Diem, M. D., Dreyfuss, G. and Van Duyne, G. D.** (2003). Structure of the Y14-Magoh core of the exon junction complex. *Curr Biol* **13**(11): 933-41.
- Laughrea, M. and Jette, L.** (1994). A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry* **33**(45): 13464-74.
- Laughrea, M. and Jette, L.** (1996). Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248-271 are dispensable for dimer formation. *Biochemistry* **35**(5): 1589-98.
- Le Hir, H. and Andersen, G. R.** (2008). Structural insights into the exon junction complex. *Curr Opin Struct Biol* **18**(1): 112-9.
- Le Hir, H., Izaurralde, E., Maquat, L. E. and Moore, M. J.** (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *Embo J* **19**(24): 6860-9.
- Le Hir, H., Gatfield, D., Braun, I. C., Forler, D. and Izaurralde, E.** (2001). The protein Mago provides a link between splicing and mRNA localization. *EMBO Rep* **2**(12): 1119-24.
- Lee, B. S. and Culbertson, M. R.** (1995). Identification of an additional gene required for eukaryotic nonsense mRNA turnover. *Proc Natl Acad Sci U S A* **92**(22): 10354-8.
- Lee, D. Y., Teyssier, C., Strahl, B. D. and Stallcup, M. R.** (2005). Role of protein methylation in regulation of transcription. *Endocr Rev* **26**(2): 147-70.
- Lee, H. C., Choe, J., Chi, S. G. and Kim, Y. K.** (2009). Exon junction complex enhances translation of spliced mRNAs at multiple steps. *Biochem Biophys Res Commun* **384**(3): 334-40.

- Lee, P. T., Liao, P. C., Chang, W. C. and Tseng, J. T. (2007). Epidermal growth factor increases the interaction between nucleolin and heterogeneous nuclear ribonucleoprotein K/poly(C) binding protein 1 complex to regulate the gastrin mRNA turnover. *Mol Biol Cell* **18**(12): 5004-13.
- Leeds, P., Peltz, S. W., Jacobson, A. and Culbertson, M. R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev* **5**(12A): 2303-14.
- Lejeune, F. and Maquat, L. E. (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol* **17**(3): 309-15.
- Lemey, P., Pybus, O. G., Rambaut, A., Drummond, A. J., Robertson, D. L., Roques, P., Worobey, M. and Vandamme, A. M. (2004). The molecular population genetics of HIV-1 group O. *Genetics* **167**(3): 1059-68.
- Levy, J. B., Dorai, T., Wang, L. H. and Brugge, J. S. (1987). The structurally distinct form of pp60c-src detected in neuronal cells is encoded by a unique c-src mRNA. *Mol Cell Biol* **7**(11): 4142-5.
- Li, C., Lin, R. I., Lai, M. C., Ouyang, P. and Tarn, W. Y. (2003). Nuclear Pnn/DRS protein binds to spliced mRNPs and participates in mRNA processing and export via interaction with RNPS1. *Mol Cell Biol* **23**(20): 7363-76.
- Libri, D., Piseri, A. and Fiszman, M. Y. (1991). Tissue-specific splicing in vivo of the beta-tropomyosin gene: dependence on an RNA secondary structure. *Science* **252**(5014): 1842-5.
- Lightfoote, M. M., Coligan, J. E., Folks, T. M., Fauci, A. S., Martin, M. A. and Venkatesan, S. (1986). Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. *J Virol* **60**(2): 771-5.
- Lin, S. and Fu, X. D. (2007). SR proteins and related factors in alternative splicing. *Adv Exp Med Biol* **623**: 107-22.
- 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.
- 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.
- Lotz, M., Clark-Lewis, I. and Ganu, V. (1994). HIV-1 transactivator protein Tat induces proliferation and TGF beta expression in human articular chondrocytes. *J Cell Biol* **124**(3): 365-71.
- Lu, X., Timchenko, N. A. and Timchenko, L. T. (1999). Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. *Hum Mol Genet* **8**(1): 53-60.
- Lu, X. J. and Olson, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31**(17): 5108-21.
- Luhrmann, R. and Stark, H. (2009). Structural mapping of spliceosomes by electron microscopy. *Curr Opin Struct Biol* **19**(1): 96-102.
- Lutz, C. S., Cooke, C., O'Connor, J. P., Kobayashi, R. and Alwine, J. C. (1998). The snRNP-free U1A (SF-A) complex(es): identification of the largest subunit as PSF, the polypyrimidine-tract binding protein-associated splicing factor. *Rna* **4**(12): 1493-9.

- 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): 18010-20.
- Ma, X. M., Yoon, S. O., Richardson, C. J., Julich, K. and Blenis, J.** (2008). SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* **133**(2): 303-13.
- Maddon, P. J., Dagleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A. and Axel, R.** (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**(3): 333-48.
- Madsen, J. M. and Stoltzfus, C. M.** (2005). An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. *J Virol* **79**(16): 10478-86.
- Madsen, J. M. and Stoltzfus, C. M.** (2006). A suboptimal 5' splice site downstream of HIV-1 splice site A1 is required for unspliced viral mRNA accumulation and efficient virus replication. *Retrovirology* **3**: 10.
- 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): 2205-8.
- 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): 2553-63.
- 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): 2381-91.
- Makeyev, A. V. and Liebhaber, S. A.** (2002). The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *Rna* **8**(3): 265-78.
- Makeyev, A. V., Kim, C. B., Ruddle, F. H., Enkhmandakh, B., Erdenechimeg, L. and Bayarsaihan, D.** (2005). HnRNP A3 genes and pseudogenes in the vertebrate genomes. *J Exp Zool A Comp Exp Biol* **303**(4): 259-71.
- Makinson, A. and Reynes, J.** (2009). The fusion inhibitor enfuvirtide in recent antiretroviral strategies. *Curr Opin HIV AIDS* **4**(2): 150-8.
- Malim, M. H. and Cullen, B. R.** (1991). HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* **65**(2): 241-8.
- Malim, M. H., Tiley, L. S., McCarn, D. F., Rusche, J. R., Hauber, J. and Cullen, B. R.** (1990). HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* **60**(4): 675-83.
- Maniatis, T. and Reed, R.** (2002). An extensive network of coupling among gene expression machines. *Nature* **416**(6880): 499-506.
- Manley, J. L. and Krainer, A. R.** A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev* **24**(11): 1073-4.
- Mann, D. A., Mikaelian, I., Zimmel, R. W., Green, S. M., Lowe, A. D., Kimura, T., Singh, M., Butler, P. J., Gait, M. J. and Karn, J.** (1994). A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J Mol Biol* **241**(2): 193-207.
- Mansky, L. M. and Temin, H. M.** (1995). Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69**(8): 5087-94.

- Mansky, L. M., Preveral, S., Selig, L., Benarous, R. and Benichou, S.** (2000). The interaction of vpr with uracil DNA glycosylase modulates the human immunodeficiency virus type 1 In vivo mutation rate. *J Virol* **74**(15): 7039-47.
- Maquat, L. E.** (1995). When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *Rna* **1**(5): 453-65.
- Maquat, L. E.** (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* **5**(2): 89-99.
- Marais, G., Nouvellet, P., Keightley, P. D. and Charlesworth, B.** (2005). Intron size and exon evolution in Drosophila. *Genetics* **170**(1): 481-5.
- Marasco, W. A., Szilvay, A. M., Kalland, K. H., Helland, D. G., Reyes, H. M. and Walter, R. J.** (1994). Spatial association of HIV-1 tat protein and the nucleolar transport protein B23 in stably transfected Jurkat T-cells. *Arch Virol* **139**(1-2): 133-54.
- Marchand, V., Mereau, A., Jacquenet, S., Thomas, D., Mouglin, 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): 629-52.
- Maris, C., Dominguez, C. and Allain, F. H.** (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *Febs J* **272**(9): 2118-31.
- Markovtsov, V., Nikolic, J. M., Goldman, J. A., Turck, C. W., Chou, M. Y. and Black, D. L.** (2000). Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol Cell Biol* **20**(20): 7463-79.
- Marshall, H. M., Ronen, K., Berry, C., Llano, M., Sutherland, H., Saenz, D., Bickmore, W., Poeschla, E. and Bushman, F. D.** (2007). Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One* **2**(12): e1340.
- Martin Stoltzfus, C.** (2009). Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. *Adv Virus Res* **74**: 1-40.
- Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fiset, J. F., Revil, T. and Chabot, B.** (2007). hnRNP proteins and splicing control. *Adv Exp Med Biol* **623**: 123-47.
- Mattaj, I. W.** (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46**(6): 905-11.
- Mattick, J. S.** (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* **2**(11): 986-91.
- Mayeda, A. and Krainer, A. R.** (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**(2): 365-75.
- Mayeda, A., Helfman, D. M. and Krainer, A. R.** (1993). Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol Cell Biol* **13**(5): 2993-3001.
- Mayeda, A., Sreaton, G. R., Chandler, S. D., Fu, X. D. and Krainer, A. R.** (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol Cell Biol* **19**(3): 1853-63.
- Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M. Q., Gardiner, E. M. and Krainer, A. R.** (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *Embo J* **18**(16): 4560-70.
- Mayrand, S. H. and Pederson, T.** (1990). Crosslinking of hnRNP proteins to pre-mRNA requires U1 and U2 snRNPs. *Nucleic Acids Res* **18**(11): 3307-18.

- Mazza, C., Segref, A., Mattaj, I. W. and Cusack, S.** (2002). Co-crystallization of the human nuclear cap-binding complex with a m7GpppG cap analogue using protein engineering. *Acta Crystallogr D Biol Crystallogr* **58**(Pt 12): 2194-7.
- Mazza, C., Ohno, M., Segref, A., Mattaj, I. W. and Cusack, S.** (2001). Crystal structure of the human nuclear cap binding complex. *Mol Cell* **8**(2): 383-96.
- McBride, A. E. and Silver, P. A.** (2001). State of the arg: protein methylation at arginine comes of age. *Cell* **106**(1): 5-8.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M. and Bentley, D. L.** (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**(6614): 357-61.
- McDonald, B. and Martin-Serrano, J.** (2009). No strings attached: the ESCRT machinery in viral budding and cytokinesis. *J Cell Sci* **122**(Pt 13): 2167-77.
- McKnight, G. L., Reasoner, J., Gilbert, T., Sundquist, K. O., Hokland, B., McKernan, P. A., Champagne, J., Johnson, C. J., Bailey, M. C., Holly, R. and et al.** (1992). Cloning and expression of a cellular high density lipoprotein-binding protein that is up-regulated by cholesterol loading of cells. *J Biol Chem* **267**(17): 12131-41.
- Michael, W. M., Eder, P. S. and Dreyfuss, G.** (1997). The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. *Embo J* **16**(12): 3587-98.
- Michelotti, G. A., Michelotti, E. F., Pullner, A., Duncan, R. C., Eick, D. and Levens, D.** (1996). Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Mol Cell Biol* **16**(6): 2656-69.
- Michienzi, A., Cagnon, L., Bahner, I. and Rossi, J. J.** (2000). Ribozyme-mediated inhibition of HIV 1 suggests nucleolar trafficking of HIV-1 RNA. *Proc Natl Acad Sci USA* **97**(16): 8955-60.
- Mikula, M., Dzwonek, A., Karczmarski, J., Rubel, T., Dadlez, M., Wyrwicz, L. S., Bomsztyk, K. and Ostrowski, J.** (2006). Landscape of the hnRNP K protein-protein interactome. *Proteomics* **6**(8): 2395-406.
- Milani, D., Mazzoni, M., Borgatti, P., Zauli, G., Cantley, L. and Capitani, S.** (1996). Extracellular human immunodeficiency virus type-1 Tat protein activates phosphatidylinositol 3-kinase in PC12 neuronal cells. *J Biol Chem* **271**(38): 22961-4.
- Mili, S., Shu, H. J., Zhao, Y. and Pinol-Roma, S.** (2001). Distinct RNP complexes of shuttling hnRNP proteins with pre-mRNA and mRNA: candidate intermediates in formation and export of mRNA. *Mol Cell Biol* **21**(21): 7307-19.
- 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): 2659-71.
- Min, H., Turck, C. W., Nikolic, J. M. and Black, D. L.** (1997). A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev* **11**(8): 1023-36.
- Mishra, S. H., Shelley, C. M., Barrow, D. J., Jr., Darby, M. K. and Germann, M. W.** (2006). Solution structures and characterization of human immunodeficiency virus Rev responsive element IIB RNA targeting zinc finger proteins. *Biopolymers* **83**(4): 352-64.
- Misteli, T. and Spector, D. L.** (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell* **3**(6): 697-705.
- Moore, M. J. and Sharp, P. A.** (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**(6444): 364-8.

- 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): 631-49.
- Mouland, A. J., Xu, H., Cui, H., Krueger, W., Munro, T. P., Prasol, M., Mercier, J., Rekosh, D., Smith, R., Barbarese, E., Cohen, E. A. and Carson, J. H.** (2001). RNA trafficking signals in human immunodeficiency virus type 1. *Mol Cell Biol* **21**(6): 2133-43.
- Mount, S. M. and Steitz, J. A.** (1981). Sequence of U1 RNA from *Drosophila melanogaster*: implications for U1 secondary structure and possible involvement in splicing. *Nucleic Acids Res* **9**(23): 6351-68.
- 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): 509-18.
- Mueller, M. C. and Bogner, J. R.** (2007). Treatment with CCR5 antagonists: which patient may have a benefit? *Eur J Med Res* **12**(9): 441-52.
- Munro, T. P., Magee, R. J., Kidd, G. J., Carson, J. H., Barbarese, E., Smith, L. M. and Smith, R.** (1999). Mutational analysis of a heterogeneous nuclear ribonucleoprotein A2 response element for RNA trafficking. *J Biol Chem* **274**(48): 34389-95.
- Mura, C., Cascio, D., Sawaya, M. R. and Eisenberg, D. S.** (2001). The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. *Proc Natl Acad Sci U S A* **98**(10): 5532-7.
- Muriaux, D., Fosse, P. and Paoletti, J.** (1996). A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry* **35**(15): 5075-82.
- Muriaux, D., Girard, P. M., Bonnet-Mathoniere, B. and Paoletti, J.** (1995). Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J Biol Chem* **270**(14): 8209-16.
- Nabell, L. M., Raja, R. H., Sayeski, P. P., Paterson, A. J. and Kudlow, J. E.** (1994). Human immunodeficiency virus 1 Tat stimulates transcription of the transforming growth factor alpha gene in an epidermal growth factor-dependent manner. *Cell Growth Differ* **5**(1): 87-93.
- Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F., Wilson, S. H. and Williams, K. R.** (1991). Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. *Biochemistry* **30**(11): 2968-76.
- Nagy, E. and Maquat, L. E.** (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* **23**(6): 198-9.
- Nakielny, S. and Dreyfuss, G.** (1997). Nuclear export of proteins and RNAs. *Curr Opin Cell Biol* **9**(3): 420-9.
- 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): 1078-89.
- Nasim, M. T., Chernova, T. K., Chowdhury, H. M., Yue, B. G. and Eperon, I. C.** (2003). HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism in RNA binding. *Hum Mol Genet* **12**(11): 1337-48.

- Neugebauer, K. M. (2002). On the importance of being co-transcriptional. *J Cell Sci* **115**(Pt 20): 3865-71.
- Neuveut, C., Scoggins, R. M., Camerini, D., Markham, R. B. and Jeang, K. T. (2003). Requirement for the second coding exon of Tat in the optimal replication of macrophage-tropic HIV-1. *J Biomed Sci* **10**(6 Pt 1): 651-60.
- Ng, L. F., Chan, M., Chan, S. H., Cheng, P. C., Leung, E. H., Chen, W. N. and Ren, E. C. (2005). Host heterogeneous ribonucleoprotein K (hnRNP K) as a potential target to suppress hepatitis B virus replication. *PLoS Med* **2**(7): e163.
- Nichols, R. C., Wang, X. W., Tang, J., Hamilton, B. J., High, F. A., Herschman, H. R. and Rigby, W. F. (2000). The RGG domain in hnRNP A2 affects subcellular localization. *Exp Cell Res* **256**(2): 522-32.
- Nisole, S., Said, E. A., Mische, C., Prevost, M. C., Krust, B., Bouvet, P., Bianco, A., Briand, J. P. and Hovanessian, A. G. (2002). The anti-HIV pentameric pseudopeptide HB-19 binds the C-terminal end of nucleolin and prevents anchorage of virus particles in the plasma membrane of target cells. *J Biol Chem* **277**(23): 20877-86.
- Nogues, G., Kadener, S., Cramer, P., Bentley, D. and Kornblihtt, A. R. (2002). Transcriptional activators differ in their abilities to control alternative splicing. *J Biol Chem* **277**(45): 43110-4.
- 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): 5527-38.
- 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): 6119-33.
- O'Donnell, W. T. and Warren, S. T. (2002). A decade of molecular studies of fragile X syndrome. *Annu Rev Neurosci* **25**: 315-38.
- 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): 373-85.
- Oberstrass, F. C., Lee, A., Stefl, R., Janis, M., Chanfreau, G. and Allain, F. H. (2006). Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. *Nat Struct Mol Biol* **13**(2): 160-7.
- Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. and Allain, F. H. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**(5743): 2054-7.
- Ohagen, A. and Gabuzda, D. (2000). Role of Vif in stability of the human immunodeficiency virus type 1 core. *J Virol* **74**(23): 11055-66.
- Ohno, M. and Shimura, Y. (1996). A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. *Genes Dev* **10**(8): 997-1007.
- Okunola, H. L. and Krainer, A. R. (2009). Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol*.
- Olson, S., Blanchette, M., Park, J., Savva, Y., Yeo, G. W., Yeakley, J. M., Rio, D. C. and Graveley, B. R. (2007). A regulator of Dscam mutually exclusive splicing fidelity. *Nat Struct Mol Biol*.

- Ostareck-Lederer, A. and Ostareck, D. H.** (2004). Control of mRNA translation and stability in haematopoietic cells: the function of hnRNPs K and E1/E2. *Biol Cell* **96**(6): 407-11.
- Ostareck-Lederer, A., Ostareck, D. H., Cans, C., Neubauer, G., Bomsztyk, K., Superti-Furga, G. and Hentze, M. W.** (2002). c-Src-mediated phosphorylation of hnRNP K drives translational activation of specifically silenced mRNAs. *Mol Cell Biol* **22**(13): 4535-43.
- Ostareck, D. H., Ostareck-Lederer, A., Shatsky, I. N. and Hentze, M. W.** (2001). Lipoxygenase mRNA silencing in erythroid differentiation: The 3'UTR regulatory complex controls 60S ribosomal subunit joining. *Cell* **104**(2): 281-90.
- Ostrowski, J., Wyrwicz, L., Rychlewski, L. and Bomsztyk, K.** (2002). Heterogeneous nuclear ribonucleoprotein K protein associates with multiple mitochondrial transcripts within the organelle. *J Biol Chem* **277**(8): 6303-10.
- Ostrowski, J., Van Seuning, I., Seger, R., Rauch, C. T., Sleath, P. R., McMullen, B. A. and Bomsztyk, K.** (1994). Purification, cloning, and expression of a murine phosphoprotein that binds the kappa B motif in vitro identifies it as the homolog of the human heterogeneous nuclear ribonucleoprotein K protein. Description of a novel DNA-dependent phosphorylation process. *J Biol Chem* **269**(26): 17626-34.
- Ott, D. E., Coren, L. V., Johnson, D. G., Kane, B. P., Sowder, R. C., 2nd, Kim, Y. D., Fisher, R. J., Zhou, X. Z., Lu, K. P. and Henderson, L. E.** (2000). Actin-binding cellular proteins inside human immunodeficiency virus type 1. *Virology* **266**(1): 42-51.
- Paiardini, M., Cervasi, B., Dunham, R., Sumpter, B., Radziewicz, H. and Silvestri, G.** (2004). Cell-cycle dysregulation in the immunopathogenesis of AIDS. *Immunol Res* **29**(1-3): 253-68.
- Paillart, J. C., Shehu-Xhilaga, M., Marquet, R. and Mak, J.** (2004). Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* **2**(6): 461-72.
- Paillart, J. C., Marquet, R., Skripkin, E., Ehresmann, B. and Ehresmann, C.** (1994). Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J Biol Chem* **269**(44): 27486-93.
- Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C. and Marquet, R.** (1996). A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A* **93**(11): 5572-7.
- Pandolfo, M., Valentini, O., Biamonti, G., Morandi, C. and Riva, S.** (1985). Single stranded DNA binding proteins derive from hnRNP proteins by proteolysis in mammalian cells. *Nucleic Acids Res* **13**(18): 6577-90.
- Parkin, N. T., Cohen, E. A., Darveau, A., Rosen, C., Haseltine, W. and Sonenberg, N.** (1988). Mutational analysis of the 5' non-coding region of human immunodeficiency virus type 1: effects of secondary structure on translation. *Embo J* **7**(9): 2831-7.
- Passiatore, G., Rom, S., Eletto, D. and Peruzzi, F.** (2009). HIV-1 Tat C-terminus is cleaved by calpain 1: implication for Tat-mediated neurotoxicity. *Biochim Biophys Acta* **1793**(2): 378-87.
- Patel, A. A. and Steitz, J. A.** (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* **4**(12): 960-70.
- Patel, S. B. and Bellini, M.** (2008). The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **36**(20): 6482-93.
- Patterson, B. and Guthrie, C.** (1987). An essential yeast snRNA with a U5-like domain is required for splicing in vivo. *Cell* **49**(5): 613-24.

- Patton, J. G., Porro, E. B., Galceran, J., Tempst, P. and Nadal-Ginard, B. (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev* 7(3): 393-406.
- Paziewska, A., Wyrwicz, L. S., Bujnicki, J. M., Bomsztyk, K. and Ostrowski, J. (2004). Cooperative binding of the hnRNP K three KH domains to mRNA targets. *FEBS Lett* 577(1-2): 134-40.
- Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., Dibanga, G., Koumare, B., Saidou, M., Esu-Williams, E., Lombart, J. P., Badombena, W., Luo, N., Vanden Haesevelde, M. and Delaporte, E. (1997). Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* 11(4): 493-8.
- Peltz, S. W., Brown, A. H. and Jacobson, A. (1993). mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes Dev* 7(9): 1737-54.
- Peng, R., Dye, B. T., Perez, I., Barnard, D. C., Thompson, A. B. and Patton, J. G. (2002). PSF and p54nrb bind a conserved stem in U5 snRNA. *Rna* 8(10): 1334-47.
- Pereira, L. A., Bentley, K., Peeters, A., Churchill, M. J. and Deacon, N. J. (2000). A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res* 28(3): 663-8.
- Perez Canadillas, J. M. and Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *Embo J* 22(11): 2821-30.
- Pettit, S. C., Sheng, N., Tritch, R., Erickson-Viitanen, S. and Swanstrom, R. (1998). The regulation of sequential processing of HIV-1 Gag by the viral protease. *Adv Exp Med Biol* 436: 15-25.
- Philips, A. V. and Cooper, T. A. (2000). RNA processing and human disease. *Cell Mol Life Sci* 57(2): 235-49.
- Piguet, V., Schwartz, O., Le Gall, S. and Trono, D. (1999). The downregulation of CD4 and MHC-I by primate lentiviruses: a paradigm for the modulation of cell surface receptors. *Immunol Rev* 168: 51-63.
- Plantier, J. C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemee, V., Damond, F., Robertson, D. L. and Simon, F. (2009). A new human immunodeficiency virus derived from gorillas. *Nat Med* 15(8): 871-2.
- Pollard, V. W., Michael, W. M., Nakielnny, S., Siomi, M. C., Wang, F. and Dreyfuss, G. (1996). A novel receptor-mediated nuclear protein import pathway. *Cell* 86(6): 985-94.
- Popov, S., Rexach, M., Zybarth, G., Reiling, N., Lee, M. A., Ratner, L., Lane, C. M., Moore, M. S., Blobel, G. and Bukrinsky, M. (1998). Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. *Embo J* 17(4): 909-17.
- Price, S. R., Evans, P. R. and Nagai, K. (1998). Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* 394(6694): 645-50.
- Pumfery, A., Deng, L., Maddukuri, A., de la Fuente, C., Li, H., Wade, J. D., Lambert, P., Kumar, A. and Kashanchi, F. (2003). Chromatin remodeling and modification during HIV-1 Tat-activated transcription. *Curr HIV Res* 1(3): 343-62.
- 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): 6365-78.

- Purvis, S. F., Georges, D. L., Williams, T. M. and Lederman, M. M.** (1992). Suppression of interleukin-2 and interleukin-2 receptor expression in Jurkat cells stably expressing the human immunodeficiency virus Tat protein. *Cell Immunol* **144**(1): 32-42.
- Query, C. C., Bentley, R. C. and Keene, J. D.** (1989). A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**(1): 89-101.
- Quivy, V. and Van Lint, C.** (2002). Diversity of acetylation targets and roles in transcriptional regulation: the human immunodeficiency virus type 1 promoter as a model system. *Biochem Pharmacol* **64**(5-6): 925-34.
- 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): 1378-92.
- Raghunathan, P. L. and Guthrie, C.** (1998). A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science* **279**(5352): 857-60.
- Ramos, A., Hollingworth, D. and Pastore, A.** (2003). G-quartet-dependent recognition between the FMRP RGG box and RNA. *Rna* **9**(10): 1198-207.
- Rappsilber, J., Ryder, U., Lamond, A. I. and Mann, M.** (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**(8): 1231-45.
- Ratner, A. V., Okada, R. D., Newell, J. B. and Pohost, G. M.** (1985). The relationship between proton nuclear magnetic resonance relaxation parameters and myocardial perfusion with acute coronary arterial occlusion and reperfusion. *Circulation* **71**(4): 823-8.
- Rautonen, J. and Rautonen, N.** (1992). Tat and Kawasaki disease. *Immunol Today* **13**(5): 190-1.
- 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): 590-9.
- Reed, R.** (1989). The organization of 3' splice-site sequences in mammalian introns. *Genes Dev* **3**(12B): 2113-23.
- Reed, R.** (1990). Protein composition of mammalian spliceosomes assembled in vitro. *Proc Natl Acad Sci U S A* **87**(20): 8031-5.
- Reed, R.** (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr Opin Cell Biol* **12**(3): 340-5.
- Reed, R.** (2003). Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol* **15**(3): 326-31.
- Revil, T., Pelletier, J., Toutant, J., Cloutier, A. and Chabot, B.** (2009). Heterogeneous nuclear ribonucleoprotein K represses the production of pro-apoptotic Bcl-xS splice isoform. *J Biol Chem* **284**(32): 21458-67.
- Richter-Cook, N. J., Howard, K. J., Cirino, N. M., Wohrl, B. M. and Le Grice, S. F.** (1992). Interaction of tRNA(Lys-3) with multiple forms of human immunodeficiency virus reverse transcriptase. *J Biol Chem* **267**(22): 15952-7.
- Riggs, N. L., Little, S. J., Richman, D. D. and Guatelli, J. C.** (1994). Biological importance and cooperativity of HIV-1 regulatory gene splice acceptors. *Virology* **202**(1): 264-71.
- Rinke, J., Appel, B., Digweed, M. and Luhrmann, R.** (1985). Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. *J Mol Biol* **185**(4): 721-31.
- Rizzuto, C. D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P. D., Hendrickson, W. A. and Sodroski, J.** (1998). A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* **280**(5371): 1949-53.
- Robberson, B. L., Cote, G. J. and Berget, S. M.** (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**(1): 84-94.

- Robert-Guroff, M., Popovic, M., Gartner, S., Markham, P., Gallo, R. C. and Reitz, M. S.** (1990). Structure and expression of tat-, rev-, and nef-specific transcripts of human immunodeficiency virus type 1 in infected lymphocytes and macrophages. *J Virol* **64**(7): 3391-8.
- 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): 1874-84.
- 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): 29963-73.
- 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.
- Rosen, C. A., Sodroski, J. G., Goh, W. C., Dayton, A. I., Lippke, J. and Haseltine, W. A.** (1986). Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III. *Nature* **319**(6054): 555-9.
- Roth, M. B., Murphy, C. and Gall, J. G.** (1990). A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J Cell Biol* **111**(6 Pt 1): 2217-23.
- Roth, M. B., Zahler, A. M. and Stolk, J. A.** (1991). A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J Cell Biol* **115**(3): 587-96.
- Ruskin, B., Zamore, P. D. and Green, M. R.** (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**(2): 207-19.
- Russell, R. S., Hu, J., Beriault, V., Mouland, A. J., Laughrea, M., Kleiman, L., Wainberg, M. A. and Liang, C.** (2003). Sequences downstream of the 5' splice donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA. *J Virol* **77**(1): 84-96.
- Salemi, M., Strimmer, K., Hall, W. W., Duffy, M., Delaporte, E., Mboup, S., Peeters, M. and Vandamme, A. M.** (2001). Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *Faseb J* **15**(2): 276-8.
- 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**: 2714-29.
- Sauliere, J., Sureau, A., Expert-Bezancon, A. and Marie, J.** (2006). The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol Cell Biol* **26**(23): 8755-69.
- Sawicka, K., Bushell, M., Spriggs, K. A. and Willis, A. E.** (2008). Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem Soc Trans* **36**(Pt 4): 641-7.
- Scala, G., Ruocco, M. R., Ambrosino, C., Mallardo, M., Giordano, V., Baldassarre, F., Dragonetti, E., Quinto, I. and Venuta, S.** (1994). The expression of the interleukin 6 gene is induced by the human immunodeficiency virus 1 TAT protein. *J Exp Med* **179**(3): 961-71.
- Scherer, W. F., Syverton, J. T. and Gey, G. O.** (1953). Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med* **97**(5): 695-710.

- Schwartz, S., Felber, B. K. and Pavlakis, G. N.** (1992). Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. *Mol Cell Biol* **12**(1): 207-19.
- Schwartz, S., Felber, B. K., Benko, D. M., Fenyo, E. M. and Pavlakis, G. N.** (1990). Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol* **64**(6): 2519-29.
- 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): 965-76.
- SenGupta, D. N. and Silverman, R. H.** (1989). Activation of interferon-regulated, dsRNA-dependent enzymes by human immunodeficiency virus-1 leader RNA. *Nucleic Acids Res* **17**(3): 969-78.
- SenGupta, D. N., Berkhout, B., Gatignol, A., Zhou, A. M. and Silverman, R. H.** (1990). Direct evidence for translational regulation by leader RNA and Tat protein of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* **87**(19): 7492-6.
- Seraphin, B.** (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *Embo J* **14**(9): 2089-98.
- Serin, G., Joseph, G., Ghisolfi, L., Bauzan, M., Erard, M., Amalric, F. and Bouvet, P.** (1997). Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *J Biol Chem* **272**(20): 13109-16.
- Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. and Steitz, T. A.** (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol* **4**(3): 215-22.
- Sharp, P. M. and Hahn, B. H.** (2008). AIDS: prehistory of HIV-1. *Nature* **455**(7213): 605-6.
- 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): 363-73.
- 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): 367-76.
- Shibuya, T., Tange, T. O., Sonenberg, N. and Moore, M. J.** (2004). eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* **11**(4): 346-51.
- Shnyreva, M., Schullery, D. S., Suzuki, H., Higaki, Y. and Bomsztyk, K.** (2000). Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein. *J Biol Chem* **275**(20): 15498-503.
- Shuster, E. O. and Guthrie, C.** (1990). Human U2 snRNA can function in pre-mRNA splicing in yeast. *Nature* **345**(6272): 270-3.
- 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): 861-7.
- Si, Z. H., Rauch, D. and Stoltzfus, C. M.** (1998). The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. *Mol Cell Biol* **18**(9): 5404-13.
- Sickmier, E. A., Frato, K. E. and Kielkopf, C. L.** (2006). Crystallization and preliminary X-ray analysis of a U2AF65 variant in complex with a polypyrimidine-tract analogue by use of protein engineering. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**(Pt 5): 457-9.

- Simard, M. J. and Chabot, B.** (2000). Control of hnRNP A1 alternative splicing: an intron element represses use of the common 3' splice site. *Mol Cell Biol* **20**(19): 7353-62.
- Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F. and Brun-Vezinet, F.** (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* **4**(9): 1032-7.
- Simon, J. H., Carpenter, E. A., Fouchier, R. A. and Malim, M. H.** (1999). Vif and the p55(Gag) polyprotein of human immunodeficiency virus type 1 are present in colocalizing membrane-free cytoplasmic complexes. *J Virol* **73**(4): 2667-74.
- Simpson, P. J., Monie, T. P., Szendroi, A., Davydova, N., Tyzack, J. K., Conte, M. R., Read, C. M., Cary, P. D., Svergun, D. I., Konarev, P. V., Curry, S. and Matthews, S.** (2004). Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure* **12**(9): 1631-43.
- Sinck, L., Richer, D., Howard, J., Alexander, M., Purcell, D. F., Marquet, R. and Paillart, J. C.** (2007). In vitro dimerization of human immunodeficiency virus type 1 (HIV-1) spliced RNAs. *RNA* **13**(12): 2141-50.
- Singh, G., Charlet, B. N., Han, J. and Cooper, T. A.** (2004). ETR-3 and CELF4 protein domains required for RNA binding and splicing activity in vivo. *Nucleic Acids Res* **32**(3): 1232-41.
- Siomi, H., Siomi, M. C., Nussbaum, R. L. and Dreyfuss, G.** (1993). The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell* **74**(2): 291-8.
- Siomi, H., Matunis, M. J., Michael, W. M. and Dreyfuss, G.** (1993). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res* **21**(5): 1193-8.
- Skalka, A. M. and Katz, R. A.** (2005). Retroviral DNA integration and the DNA damage response. *Cell Death Differ* **12 Suppl 1**: 971-8.
- Slijepcevic, P. and Al-Wahiby, S.** (2005). Telomere biology: integrating chromosomal end protection with DNA damage response. *Chromosoma* **114**(4): 275-85.
- Smith, C. W. and Valcarcel, J.** (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* **25**(8): 381-8.
- Smith, D. J., Query, C. C. and Konarska, M. M.** (2008). "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol Cell* **30**(6): 657-66.
- Smith, G. C. and Jackson, S. P.** (1999). The DNA-dependent protein kinase. *Genes Dev* **13**(8): 916-34.
- Smith, S. M., Pentlicky, S., Klase, Z., Singh, M., Neuveut, C., Lu, C. Y., Reitz, M. S., Jr., Yarchoan, R., Marx, P. A. and Jeang, K. T.** (2003). An in vivo replication-important function in the second coding exon of Tat is constrained against mutation despite cytotoxic T lymphocyte selection. *J Biol Chem* **278**(45): 44816-25.
- Solis, A. S., Peng, R., Crawford, J. B., Phillips, J. A., 3rd and Patton, J. G.** (2008). Growth hormone deficiency and splicing fidelity: two serine/arginine-rich proteins, ASF/SF2 and SC35, act antagonistically. *J Biol Chem* **283**(35): 23619-26.
- Spellman, R. and Smith, C. W.** (2006). Novel modes of splicing repression by PTB. *Trends Biochem Sci* **31**(2): 73-6.
- Srivastava, M. and Pollard, H. B.** (1999). Molecular dissection of nucleolin's role in growth and cell proliferation: new insights. *FASEB J* **13**(14): 1911-22.

- Staffa, A. and Cochrane, A.** (1994). The tat/rev intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. *J Virol* **68**(5): 3071-9.
- 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): 4597-605.
- Stark, H., Dube, P., Luhrmann, R. and Kastner, B.** (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**(6819): 539-42.
- Stoltzfus, M. C.** (2009). Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. *Adv. Virus Res.* **74**: 1-40.
- 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): 113-30.
- Stutz, F. and Rosbash, M.** (1998). Nuclear RNA export. *Genes Dev* **12**(21): 3303-19.
- Swanson, A. K. and Stoltzfus, C. M.** (1998). Overlapping cis sites used for splicing of HIV-1 env/nef and rev mRNAs. *J Biol Chem* **273**(51): 34551-7.
- Swanson, M. S. and Dreyfuss, G.** (1988). RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *Embo J* **7**(11): 3519-29.
- Swanson, M. S., Nakagawa, T. Y., LeVan, K. and Dreyfuss, G.** (1987). Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol Cell Biol* **7**(5): 1731-9.
- 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): 3540-51.
- 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): 1148-53.
- Takimoto, M., Tomonaga, T., Matunis, M., Avigan, M., Krutzsch, H., Dreyfuss, G. and Levens, D.** (1993). Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro. *J Biol Chem* **268**(24): 18249-58.
- Tanaka, E., Fukuda, H., Nakashima, K., Tsuchiya, N., Seimiya, H. and Nakagama, H.** (2007). HnRNP A3 binds to and protects mammalian telomeric repeats in vitro. *Biochem Biophys Res Commun* **358**(2): 608-14.
- Tange, T. O. and Kjems, J.** (2001). SF2/ASF binds to a splicing enhancer in the third HIV-1 tat exon and stimulates U2AF binding independently of the RS domain. *J Mol Biol* **312**(4): 649-62.
- Tange, T. O., Nott, A. and Moore, M. J.** (2004). The ever-increasing complexities of the exon junction complex. *Curr Opin Cell Biol* **16**(3): 279-84.
- Tange, T. O., Shibuya, T., Jurica, M. S. and Moore, M. J.** (2005). Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *Rna* **11**(12): 1869-83.
- 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): 5748-58.
- Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E. and Hammer, S. M.** (2008). The challenge of HIV-1 subtype diversity. *N Engl J Med* **358**(15): 1590-602.
- Tennyson, C. N., Klamut, H. J. and Worton, R. G.** (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat Genet* **9**(2): 184-90.

- Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeyer, C., Hentze, M. W. and Kulozik, A. E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *Embo J* **17**(12): 3484-94.
- Thisted, T., Lyakhov, D. L. and Liebhaber, S. A. (2001). Optimized RNA targets of two closely related triple KH domain proteins, heterogeneous nuclear ribonucleoprotein K and alphaCP-2KL, suggest Distinct modes of RNA recognition. *J Biol Chem* **276**(20): 17484-96.
- Tintaru, A. M., Hautbergue, G. M., Hounslow, A. M., Hung, M. L., Lian, L. Y., Craven, C. J. and Wilson, S. A. (2007). Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep* **8**(8): 756-62.
- Tomonaga, T. and Levens, D. (1995). Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J Biol Chem* **270**(9): 4875-81.
- Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**(4968): 505-10.
- Turelli, P., Doucas, V., Craig, E., Mangeat, B., Klages, N., Evans, R., Kalpana, G. and Trono, D. (2001). Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol Cell* **7**(6): 1245-54.
- Tycowski, K. T., You, Z. H., Graham, P. J. and Steitz, J. A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol Cell* **2**(5): 629-38.
- Ueno, T., Tokunaga, K., Sawa, H., Maeda, M., Chiba, J., Kojima, A., Hasegawa, H., Shoya, Y., Sata, T., Kurata, T. and Takahashi, H. (2004). Nucleolin and the packaging signal, psi, promote the budding of human immunodeficiency virus type-1 (HIV-1). *Microbiol Immunol* **48**(2): 111-8.
- Ujvari, A. and Luse, D. S. (2004). Newly Initiated RNA encounters a factor involved in splicing immediately upon emerging from within RNA polymerase II. *J Biol Chem* **279**(48): 49773-9.
- Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A. and Darnell, R. B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**(5648): 1212-5.
- Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J. and Darnell, R. B. (2006). An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**(7119): 580-6.
- Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J. S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., Zeeberg, B. R., Kane, D., Weinstein, J. N., Blume, J. and Darnell, R. B. (2005). Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37**(8): 844-52.
- 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): 1706-9.
- Van Lint, C. (2000). Role of chromatin in HIV-1 transcriptional regulation. *Adv Pharmacol* **48**: 121-60.
- Varani, G. and Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct* **27**: 407-45.
- 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): 8229-34.

- Venables, J. P., Koh, C. S., Froehlich, U., Lapointe, E., Couture, S., Inkel, L., Bramard, A., Paquet, E. R., Watier, V., Durand, M., Lucier, J. F., Gervais-Bird, J., Tremblay, K., Prinos, P., Klinck, R., Elela, S. A. and Chabot, B. (2008). Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol Cell Biol* **28**(19): 6033-43.
- Vera, J., Jaumot, M., Estanyol, J. M., Brun, S., Agell, N. and Bachs, O. (2006). Heterogeneous nuclear ribonucleoprotein A2 is a SET-binding protein and a PP2A inhibitor. *Oncogene* **25**(2): 260-70.
- Verhoef, K., Bilodeau, P. S., van Wamel, J. L., Kjems, J., Stoltzfus, C. M. and Berkhout, B. (2001). Repair of a Rev-minus human immunodeficiency virus type 1 mutant by activation of a cryptic splice site. *J Virol* **75**(7): 3495-500.
- Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. and Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell* **6**(6): 1331-42.
- Waggoner, S. and Sarnow, P. (1998). Viral ribonucleoprotein complex formation and nucleolar-cytoplasmic relocalization of nucleolin in poliovirus-infected cells. *J Virol* **72**(8): 6699-709.
- Wahl, M. C., Will, C. L. and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**(4): 701-18.
- Wang, Z. and Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna* **14**(5): 802-13.
- Watkins, N. J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C. and Luhrmann, R. (2000). A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**(3): 457-66.
- Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L. and Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**(7256): 711-6.
- Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H. and Jones, K. A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**(4): 451-62.
- Weinberg, R. A. and Penman, S. (1968). Small molecular weight monodisperse nuclear RNA. *J Mol Biol* **38**(3): 289-304.
- Wentz, M. P., Moore, B. E., Cloyd, M. W., Berget, S. M. and Donehower, L. A. (1997). A naturally arising mutation of a potential silencer of exon splicing in human immunodeficiency virus type 1 induces dominant aberrant splicing and arrests virus production. *J Virol* **71**(11): 8542-51.
- Widlak, P. and Garrard, W. T. (1998). Nucleosomes and regulation of gene expression. Structure of the HIV-1 5'LTR. *Acta Biochim Pol* **45**(1): 209-19.
- Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., Giddings, M. C. and Weeks, K. M. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**(4): e96.
- Will, C. L. and Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* **13**(3): 290-301.
- 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): 4978-88.

- Williams, K. R., Stone, K. L., LoPresti, M. B., Merrill, B. M. and Planck, S. R.** (1985). Amino acid sequence of the UP1 calf thymus helix-destabilizing protein and its homology to an analogous protein from mouse myeloma. *Proc Natl Acad Sci U S A* **82**(17): 5666-70.
- Winkermann, G., Bach, M. and Luhrmann, R.** (1989). Evidence from complementation assays in vitro that U5 snRNP is required for both steps of mRNA splicing. *Embo J* **8**(10): 3105-12.
- Wolf, D., Witte, V., Clark, P., Blume, K., Lichtenheld, M. G. and Baur, A. S.** (2008). HIV Nef enhances Tat-mediated viral transcription through a hnRNP-K-nucleated signaling complex. *Cell Host Microbe* **4**(4): 398-408.
- Wolffe, A. P. and Meric, F.** (1996). Coupling transcription to translation: a novel site for the regulation of eukaryotic gene expression. *Int J Biochem Cell Biol* **28**(3): 247-57.
- Wu, J. Y. and Maniatis, T.** (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**(6): 1061-70.
- 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): 832-5.
- Wyatt, R., Kwong, P. D., Desjardins, E., Sweet, R. W., Robinson, J., Hendrickson, W. A. and Sodroski, J. G.** (1998). The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**(6686): 705-11.
- Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A. and Krainer, A. R.** (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure* **5**(4): 559-70.
- Xue, D., Rubinson, D. A., Pannone, B. K., Yoo, C. J. and Wolin, S. L.** (2000). U snRNP assembly in yeast involves the La protein. *Embo J* **19**(7): 1650-60.
- Yamamoto, H., Tsukahara, K., Kanaoka, Y., Jinno, S. and Okayama, H.** (1999). Isolation of a mammalian homologue of a fission yeast differentiation regulator. *Mol Cell Biol* **19**(5): 3829-41.
- Yang, J. P., Reddy, T. R., Truong, K. T., Suhasini, M. and Wong-Staal, F.** (2002). Functional interaction of Sam68 and heterogeneous nuclear ribonucleoprotein K. *Oncogene* **21**(47): 7187-94.
- Yang, X., Bani, M. R., Lu, S. J., Rowan, S., Ben-David, Y. and Chabot, B.** (1994). The A1 and A1B proteins of heterogeneous nuclear ribonucleoproteins modulate 5' splice site selection in vivo. *Proc Natl Acad Sci U S A* **91**(15): 6924-8.
- Young, P. J., DiDonato, C. J., Hu, D., Kothary, R., Androphy, E. J. and Lorson, C. L.** (2002). SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. *Hum Mol Genet* **11**(5): 577-87.
- 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): 10077-84.
- 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): 9243-7.
- Zaric, B., Chami, M., Remigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F. K. and Kambach, C.** (2005). Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J Biol Chem* **280**(16): 16066-75.

- Zekeng, L., Gurtler, L., Afane Ze, E., Sam-Abbenyi, A., Mbouni-Essomba, G., Mpoudi-Ngolle, E., Monny-Lobe, M., Tapka, J. B. and Kaptue, L. (1994). Prevalence of HIV-1 subtype O infection in Cameroon: preliminary results. *AIDS* **8**(11): 1626-8.
- Zeng, C. and Berget, S. M. (2000). Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. *Mol Cell Biol* **20**(21): 8290-301.
- Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L. and Charneau, P. (2000). HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101**(2): 173-85.
- Zhang, H., Pomerantz, R. J., Dornadula, G. and Sun, Y. (2000). Human immunodeficiency virus type 1 Vif protein is an integral component of an mRNP complex of viral RNA and could be involved in the viral RNA folding and packaging process. *J Virol* **74**(18): 8252-61.
- Zhang, J., Sun, X., Qian, Y., LaDuca, J. P. and Maquat, L. E. (1998). At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. *Mol Cell Biol* **18**(9): 5272-83.
- Zhang, W. J. and Wu, J. Y. (1996). Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol Cell Biol* **16**(10): 5400-8.
- Zhang, W. W., Zhang, L. X., Busch, R. K., Farres, J. and Busch, H. (1993). Purification and characterization of a DNA-binding heterodimer of 52 and 100 kDa from HeLa cells. *Biochem J* **290** ( Pt 1): 267-72.
- Zhou, Z., Licklider, L. J., Gygi, S. P. and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**(6903): 182-5.
- Zhou, Z., Sim, J., Griffith, J. and Reed, R. (2002). Purification and electron microscopic visualization of functional human spliceosomes. *Proc Natl Acad Sci U S A* **99**(19): 12203-7.
- Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E. and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**(6802): 401-5.
- 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): 1351-61.



## **Résumé :**

L'épissage est une étape clef de la multiplication du VIH-1. Par utilisation de 4 sites donneurs et 8 sites accepteurs d'épissage, plus de 40 ARNm différents sont produits. Une approche protéomique nous a permis d'identifier de nouvelles protéines interagissant avec la région de l'ARN viral contenant le site A7. Nous avons démontré l'interaction directe avec l'ARN viral de 5 des protéines identifiées (nucléoline, hnRNP A1/B, hnRNP H et hnRNP K). Nous avons montré que hnRNP K a plusieurs sites de fixation dans la région du site A7 et que hnRNP A1 et hnRNP K se lient de façon coopérative. Nous avons montré un effet inhibiteur de hnRNP K sur l'épissage au site A7.

Comme la protéine hnRNP A1 est un régulateur négatif de plusieurs sites accepteurs d'épissage (A1, A2, A3, A7), nous avons testé si la protéine hnRNP K pouvait renforcer l'inhibition à ces sites. En fait, hnRNP K active l'épissage *in vitro* des introns entre le site donneur D1 et les sites accepteurs A1, A2 et A3. Nous avons montré que la protéine hnRNP K renforce fortement l'activité de ASF/SF2 au site A2, ce qui indique que selon le contexte, la protéine hnRNP K peut être activatrice ou inhibitrice de l'épissage du VIH-1.

J'ai observé de plus que la surexpression de la protéine hnRNP K dans des cellules HeLa, transfectées avec le plasmide p $\Delta$ PSP contenant le virus VIH-1 dépourvu de ses capacités d'encapsidation, produit un changement très marqué de l'épissage alternatif de l'ARN  $\Delta$ PSP, ce qui confirme la forte influence de hnRNP K sur l'épissage alternatif du VIH-1. L'augmentation de la concentration cellulaire de hnRNP K dans les cellules HeLa conduit aussi à une diminution de la protéine virale Nef.

La protéine hnRNP K intervient donc non seulement dans la régulation du site A7, mais aussi dans celle de la majorité des sites d'épissage régulés de l'ARN du VIH. L'action de cette protéine sur plusieurs des sites d'épissage montre que la protéine hnRNP K est probablement un régulateur général de l'épissage de VIH-1.

**Mots-clés :** épissage alternatif, virus VIH-1, hnRNP K, protéine nef, protéines hnRNP.

## **Abstract:**

HIV-1 pre-mRNA splicing depends upon 4 donor and 8 acceptor sites, which are used in combination to produce more than 40 different mRNAs. To further characterize nuclear factors involved in these processes, we purified RNP complexes formed by incubation of SLS2-A7 transcripts in HeLa cell nuclear extracts by affinity chromatography to identify new associated proteins. We showed that, in addition to the well known hnRNP A1 inhibitor of site A7, nucleolin, hnRNP H and hnRNP K interact directly with SLS2-A7 RNA. We demonstrated that hnRNP K has multiple binding sites in the vicinity of site A7 and that binds cooperatively to hnRNP A1 to the A7 RNA region and limits the A7 utilization *in vitro*.

As hnRNP A1 is a negative regulator of several HIV-1 splicing sites (A1, A2, A3), we tested whether hnRNP K may also reinforce hnRNP A1 inhibition at these sites. Surprisingly, hnRNP K activated *in vitro* splicing of the D1-A1, D1-A2 and D1-A3 introns. Interestingly, hnRNP K was found to reinforce strongly the ASF/SF2 activity at site A2, which indicates that depending on the splicing site hnRNP K can be a splicing activator or inhibitor.

To test how hnRNP K influences the relative utilization of HIV-1 splicing sites *in cellulo*, we used plasmid p $\Delta$ PSP containing all the HIV-1 splicing sites and tested the effect of over-expression in HeLa cells on alternative splicing of the  $\Delta$ PSP RNA. Doubling the amount of hnRNP K in HeLa cells led to a drastic change of the  $\Delta$ PSP RNA alternative splicing, which confirms the strong influence of hnRNP K on alternative splicing. Moreover, increase of cellular concentration of hnRNP K strongly decrease the viral Nef protein production.

hnRNP K protein affects A7 splicing regulation but also regulates the majority of regulated splicing sites of HIV. By extension of the study of hnRNP K effect to other HIV-1 splicing sites, we discovered that hnRNP K is a general regulator of HIV-1 splicing.

**Key words:** alternative splicing, HIV-1 virus, hnRNP K, nef protein, hnRNP proteins.