AVERTISSEMENT

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

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

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

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

LIENS

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

Nancy-Université
Université Henri Poincaré

Faculté des Sciences & Techniques
U.F.R. Sciences et Techniques Biologiques
Ecole Doctorale Biologie, Santé, Environnement

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Biologie Moléculaire

par **Jean-Michel SALIOU**

Etude sur le complexe TAR/Tat/cycline T1 Et Etude des régulations de l'épissage de l'ARN pré-messager du virus HIV-1 : effet global des protéines virales et analyse fine du rôle des protéines SR ASF/SF2 et 9G8 au site accepteur A3

Soutenance publique le 05 Septembre 2008

Membres du Jury :

Rapporteurs :	Mme Joëlle MARIE	Directeur de Recherche INSERM, Gif-sur-Yvette
	Mme AC. DOCK-BREGEON	Directeur de Recherche CNRS, Illkirch
Examinateurs :	Mr James STEVENIN	Directeur de Recherche INSERM, Illkirch
	Mr Alain KROL	Directeur de Recherche CNRS, Strasbourg
Directeur de thèse :	Mme Christiane BRANLANT	Directeur de Recherche CNRS, Nancy

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I,
Faculté des Sciences et Techniques, BP 239, 54506 Vandœuvre lès Nancy Cedex, France.

Nancy-Université
Université Henri Poincaré

Faculté des Sciences & Techniques
U.F.R. Sciences et Techniques Biologiques
Ecole Doctorale Biologie, Santé, Environnement

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Biologie Moléculaire

par **Jean-Michel SALIOU**

Etude sur le complexe TAR/Tat/cycline T1 Et Etude des régulations de l'épissage de l'ARN pré-messager du virus HIV-1 : effet global des protéines virales et analyse fine du rôle des protéines SR ASF/SF2 et 9G8 au site accepteur A3

Soutenance publique le 05 Septembre 2008

Membres du Jury :

Rapporteurs :	Mme Joëlle MARIE	Directeur de Recherche INSERM, Gif-sur-Yvette
	Mme AC. DOCK-BREGEON	Directeur de Recherche CNRS, Illkirch
Examinateurs :	Mr James STEVENIN	Directeur de Recherche INSERM, Illkirch
	Mr Alain KROL	Directeur de Recherche CNRS, Strasbourg
Directeur de thèse :	Mme Christiane BRANLANT	Directeur de Recherche CNRS, Nancy

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I,
Faculté des Sciences et Techniques, BP 239, 54506 Vandœuvre lès Nancy Cedex, France.

REMERCIEMENTS

En premier lieu, je tiens à remercier Mme Christiane Branlant, pour m'avoir accueilli dans son laboratoire, m'avoir permis de réaliser mon travail dans de bonnes conditions matérielles et pour la formation scientifique qu'elle m'a donnée. Ses conseils, sa rigueur, son soutien et sa recherche constante de nouvelles pistes lorsque nous étions face à des difficultés ont été essentiels tout au long du travail expérimental. Je la remercie aussi très fortement pour son effort considérable pour me guider lors de la rédaction de ce manuscrit.

J'adresse aussi tous mes remerciements à Joëlle Marie et Anne-Catherine Dock-Bregeon pour avoir accepté d'être les rapporteurs de ce travail de thèse et de l'avoir lue pendant leurs vacances. Je remercie aussi James Stévenin et Alain Krol pour leur participation au jury de thèse.

J'adresse des remerciements tout particulier à James Stévenin qui a suivi une partie de mes travaux dans le cadre de la collaboration avec notre équipe, qui m'a accueilli en stage dans son équipe, qui nous a fourni les protéines SR qui ont été essentielles à l'étude des régulations de l'épissage. Je le remercie vivement ainsi que Cyril Bourgeois pour leur aide lors de la rédaction de l'article, ainsi que pour toutes les discussions que nous avons eues sur les régulations de l'épissage alternatif. Je ne voudrais pas non plus oublier Renata Gattoni, avec qui j'ai réalisé les premières analyses d'épissage *in cellulo*.

Je voudrais aussi remercier vivement Alain Van Dorsselaer et Sarah Sanglier avec qui nous avons réalisé les analyses protéomiques et chez qui je travaille actuellement. Leur soutien et leur compréhension m'ont permis d'envisager la fin de thèse et la phase de rédaction au plus sereinement.

J'adresse également un large merci à tous les membres du laboratoire, à la fois pour leur contribution de près ou de loin à ce travail, mais aussi pour la bonne ambiance générale à laquelle chacun a contribué. Parmi eux, je remercie spécialement Athanase Visvikis qui m'a formé dans le domaine des cultures cellulaires et transfections, Xavier Manival et Christophe Charron qui m'ont apporté tous leurs conseils pour la partie structurale, Lilia Ayadi pour la partie épissage *in vitro*. Un merci tout particulier aussi à Florence Oillo avec qui j'ai pu travailler étroitement en fin de thèse. Le travail fourni était efficace, motivant et toujours plaisant. Je n'oublie pas non plus les aides des débuts, Delphine et Benoit, avec qui il était bon de discuter et partager scientifiquement ou amicalement.

Ces années de vie au labo 3 auront été aussi l’occasion de construire beaucoup plus qu’une amitié avec mes compagnons de paillasse, Alan et Xavier, alors un grand merci à vous deux, très sincèrement!

Et puis toute ma gratitude et mes sentiments vont vers mes amis et ma famille. Merci pour votre soutien et pour tout ce que vous m’avez apporté depuis déjà longtemps. Cela représente beaucoup pour moi.

SOMMAIRE

ABREVIATIONS 5

PARTIE BIBLIOGRAPHIQUE 7

CHAPITRE 1 : Le rétrovirus HIV-1 9
1. Généralités 9
1.1. La pandémie 9
1.2. Evolution phylogénétique 10
2. Le virion : génome et structure 12
2.1. L'ARN génomique 12
2.2. La structure du virion 14
3. Cycle réplicatif du virus HIV-1 16
3.1. La phase pré-intégrative 16
3.1.1. Entrée du virion dans la cellule infectée 16
3.1.2. Transcription inverse 18
3.1.3. Transport et intégration de l'ADN proviral dans le génome 19
3.2. La phase post-intégrative 21
3.2.1. Expression des gènes viraux 21
3.2.2. Assemblage, bourgeonnement et maturation des nouveaux virions 27
3.2.3. Inhibition de la réponse immunitaire et apoptose 30

CHAPITRE 2 : Description générale de la transcription par l'ARN Pol II : cas particulier de l'ADN proviral de HIV-1 33
1. La transcription : description générale 33
1.1. Généralités 33
1.2. La formation du complexe de pré-initiation 33
1.3. Le rôle central de la phosphorylation du CTD 36
1.4. Le complexe P-TEFb 40
2. La transcription de l'ADNc proviral du virus HIV-1 41
3. La transactivation 43
4. Etat des connaissances structurales du complexe TAR/Tat/cycline T1 44

CHAPITRE 3 : Description générale de l'épissage des ARN 51
1. La réaction d'épissage 52
1.1. La réaction chimique d'épissage 52
1.2. Les séquences définissant les introns 53
1.3. Les composants du spliceosome 54
1.3.1. Biogenèse des UsnRNP chez les mammifères 54
1.3.1.1. Cas des snRNP U1, U2, U4 et U5 54
1.3.1.2. Cas de la snRNP U6 56
1.3.2. Rôles et particularités des UsnRNP 57
1.3.2.1. La snRNP U1 (12S) 57
1.3.2.2. La snRNP U2 (12, 15 et 17S) 58
1.3.2.3. La tri-snRNP U4/U6.U5 58
1.4. L'épissage : assemblage du spliceosome et catalyse de la réaction 61
1.4.1. Le complexe H (pour Heterogeneous) 63
1.4.2. Formation du complexe d'engagement ou complexe E 63
1.4.2.1. Reconnaissance du site 5' d'épissage 64

1.4.2.2. Reconnaissance du site 3' d'épissage 64
1.4.2.3. Rapprochement des sites 5' et 3' d'épissage 65
1.4.3. Formation du complexe A 66
1.4.4. Formation du complexe B 67
1.4.5. Le complexe B* 70
1.4.6. Le complexe C 70
1.4.7. Les complexes spliceosomaux pré-assemblés 72
2. Intégration de l'épissage dans le processus d'expression des gènes 72
2.1. Lien entre transcription et épissage 72
2.1.1. Rôle du CTD de l'ARN Pol II dans le couplage 73
2.1.2. Les protéines en lien avec transcription et épissage 74
2.1.3. Importance du promoteur vis à vis de l'épissage 75
2.1.4. Rôle de la vitesse d'élongation de l'ARN polymérase II 76
2.1.5. Influence de l'épissage sur la transcription 78
2.2. Lien entre export, NMD et traduction : rôle de l'EJC 79
2.2.1. Composants et assemblage du complexe EJC 79
2.2.2. Rôle du complexe EJC dans l'export des ARNm dans le cytoplasme 80
2.2.3. Complexe EJC et système NMD 80
2.2.4. Rôle du complexe EJC dans la traduction des ARNm 81
3. L'épissage alternatif 81
3.1. Les modes d'épissage alternatif 81
3.2. Les sites d'épissage alternatif 84
4. Les régulateurs de l'épissage 85
4.1. Les protéines SR 86
4.1.1. La famille des protéines SR 86
4.1.2. Les séquences reconnues par les protéines SR 89
4.1.2.1. Séquences identifiées par la méthode SELEX 89
4.1.2.2. Les approches bioinformatiques permettant la recherche des ESE 91
4.1.3. Les rôles des protéines SR 92
4.1.3.1. Rôles donnés aux interactions domaine RS-domaine RS 92
4.1.3.2. Rôles donnés aux interactions domaine RS-ARN 97
4.1.3.3. Antagonisme 98
4.1.3.4. Effet inhibiteur 99
4.1.3.5. Rôles des protéines SR en dehors de l'épissage 99
4.2. Les protéines SR-like 101
4.2.1. Les co-activateurs SRm300 et SRm160 101
4.2.2. Les protéines Tra 102
4.3. Les protéines hnRNP 103
4.3.1. Organisation structurale des hnRNP 103
4.3.2. Variants d'épissage et post-traductionnelles des hnRNP 106
4.3.3. Les rôles généraux des protéines hnRNP 107
4.4. Les autres facteurs intervenant dans la régulation de l'épissage 112

OBJECTIFS : 115

RESULTATS 117

CHAPITRE 4 : Travaux en vue d'une étude structurale du complexe TAR/Tat/cyclineT1 119
1. Présentation de la problématique et du partenariat 119

2. Résultats obtenus 121
2.1. Production de l'ARN TAR 121
2.1.1. Le choix de l'ARN produit et de son mode de production 121
2.1.2. Production et purification de l'ARN 123
2.2. Etude de la fixation de la protéine Tat sur l'ARN TAR produit 127
2.3. Production du fragment N terminal de la cycline T1 130
3. Perspectives 137

CHAPITRE 5 : Etude de l'effet de l'expression des protéines virales Tat, Nef et NCp7 sur les taux d'expression des facteurs cellulaires d'épissage et impact sur l'épissage de HIV-1 141
1. Etude de l'effet possible de l'expression de la protéine Tat sur les taux de production de protéines SR et hnRNP 141
1.1. Objectifs 141
1.2. Stratégie utilisée pour la production de la protéine Tat 142
1.3. Etude de l'effet de l'expression de la protéine Tat sur l'expression des protéines SR et hnRNP 144
2. Effet de la surexpression des protéines Tat et Nef sur les taux relatifs d'utilisation des sites d'épissage de l'ARN du virus HIV-1 147
3. Etude de l'effet directe possible de la protéine virale NCp7 sur le taux d'utilisation des sites d'épissage de l'ARN de HIV-1 149
4. Discussion-Perspectives 150

CHAPITRE 6 : Etudes de l'effet de l'élément inhibiteur ESS2 sur la compétition entre les sites d'épissage A1 à A5 155
1. Présentation générale des objectifs de l'article et des travaux complémentaires 155
2. Publication : « Rôle de la structure des ARN et des facteurs protéiques sur l'épissage du virus HIV-1 » 161
3. Données obtenues sur la régulation du site A3 par l'élément ESS2 176
3.1. Etude de l'effet *in cellulo* des mutations 4058 et 4064 176
3.2. Etude de l'effet *in cellulo* de la mutation mutESS2 177
3.3. Etude de l'effet de la structure SLS3 sur l'activité de l'élément ESS2 181

CHAPITRE 7 : Etude de la régulation de l'épissage de l'exon A2-D3 187
1. L'élément ISS situé en aval du site A2 n'influence pas le taux d'utilisation de ce site *in cellulo* contrairement aux données antérieurement obtenues *in vitro* 187
1.1. Etat des connaissances obtenues par L. Damier et S. Jacquenet 187
1.2. Démonstration de l'absence d'effet de l'ISS sur le taux d'ARNm vpr *in cellulo* 190
2. Une faible efficacité d'épissage *in vitro* de l'intron D3-A3 192
3. Comment expliquer le faible taux de production des ARNm vpr par rapport aux autres ARNm viraux et sa forte augmentation en présence d'un excès d'ASF/SF2 ? 195

CHAPITRE 8 : Mise en évidence de régulations du site A3 indépendantes de l'élément ESS2 197
1. Etat des connaissances lorsque j'ai entrepris ce travail 197
2. Résultats obtenus 202
2.1. Etude *in cellulo* 202
2.2. Etude de l'activation par 9G8 *in vitro* 210
2.3. Etude de la fixation de 9G8 et ASF/SF2 par expériences de retard sur gel 214

2.4. Etude de la fixation de 9G8 par la technique d'empreinte en solution 221
2.5. Etude de l'effet de la mutation $C_{5396}U$ sur la fixation des protéines nucléaires sur la tige-boucle SLS 3 223

CONCLUSIONS 237

MATERIELS ET METHODES 243
1. Les constructions génétiques 245
1.1. Les constructions pΔPSP 245
1.2. Les vecteurs d'expression eucaryote 246
1.3. Les constructions utilisées en épissage *in vitro* 247
1.4. La construction pRTR 247
1.5. Les vecteurs d'expression codant la cycline T1 250
1.6. Les constructions MS2 251
2. Analyse du profil d'épissage *in vivo* 251
2.1. Transfection et extraction ARN 251
2.2. Analyse RT-PCR 251
3. Analyse du taux de transcrits des protéines 9G8, SC35, ASF/SF2, hnRNP A1, F, H et H' 252
4. Techniques d'analyse de retard sur gel 252
4.1. Préparation des matrices de transcriptions 252
4.2. Préparation des ARN 253
4.2.1. Synthèse de transcrits radioactifs pour les expériences de retard sur gel. 253
4.2.2. Synthèse de transcrits froids pour les expériences de retard sur gel. 253
4.3. Conditions expérimentales des retards sur gel 254
5. Méthodes spécifiques à l'étude de l'épissage *in vitro* 255
5.1. Préparation de la matrice de transcription 255
5.2. Transcription *in vitro* 255
5.3. Epissage *in vitro* 256
5.4. Détermination de l'efficacité d'épissage 256
6. Production et purification de l'ARN TAR en grande quantité 257
7. Détermination de la structure secondaire de l'ARN TAR 257
7.1. Marquage de l'ARN TAR en 5' 257
7.2. Hydrolyses enzymatiques 258
8. Technique d'empreinte 258
9. Production, purification et cristallogénèse du fragment (1-280) de la cycline T1 259
9.1. Production 259
9.2. Purification 260
9.2.1. Sonication 260
9.2.2. Précipitation des acides nucléiques 260
9.2.3. Chromatographie d'affinité sur colonne His-Trap couplée à un système FPLC 260
9.2.4. Chromatographie d'exclusion de taille sur colonne Superdex 75 couplée à un système FPLC 260
9.3. Cristallogenèse 260
10. Méthode MS2 261

REFERENCES 263

ABREVIATIONS

A	adénine
ADN	acide désoxyribonucléique
ADNc	acide désoxyribonucléique complémentaire
ARN	acide ribonucléique
ARNm	acide ribonucléique messager
ARNt	acide ribonucléique de transfert
ASF/SF2	adénosine triphosphate
BSA	« bovin serum albumin »
C	cytosine
cpm	coups par minute
Da/kDa	dalton/kilodalton
DNase	désoxyribonucléase
dNTP	désoxyribonucléotide triphosphate
DTT	dithiothréitol
dTTP	désoxythymidine triphosphate
EDTA	éthylène diamine tétra-acétate
EN	extrait nucléaire
G	guanine
HeLa	Helen Laug
HEPES	acide (N-[2-hydroxyéthyl] piperazine-N'-[2-éthanesulfonique])
hnRNP	« heterogeneous ribonucleoprotein »
HIV	« human immunodeficiency virus »
kb	kilobase
m2,2,7G	(2, 2, 7) tri-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 »
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
SIV	« simian immunodeficiency virus »
T	thymine
TBE	tampon tris-borate-EDTA
TG	tampon tris-Glycérol
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
v/v	volume/volume
Y	pyrimidine
3’ss	« 3’ splice site » (site 3’ d’épissage)
5’ss	« 5’ splice site » (site 5’ d’épissage)

PARTIE BIBLIOGRAPHIQUE

CHAPITRE 1 : Le rétrovirus HIV-1

CHAPITRE 2 : Description générale de la transcription par l'ARN Pol II : cas particulier de l'ADN proviral de HIV-1

CHAPITRE 3 : Description générale de l'épissage des ARN

CHAPITRE 1 : Le rétrovirus HIV-1

1. Généralités

1.1. La pandémie

Le virus HIV-1[1] ou virus de l'immunodéficience humaine de type 1 (VIH-1), a été isolé en 1983 par Luc Montagnié et Robert Gallo, mais les premières traces d'infection remontent à 1930 (Korber *et al.*, 2000 ; Salemi *et al.*, 2001). Depuis, le virus s'est répandu et est responsable d'une pandémie, le SIDA[2], dont souffrent aujourd'hui 33,2 millions d'individus à travers le monde (ONUSIDA/OMS 2007). Le SIDA conduit à une grave immunodépression exposant le malade à des infections opportunistes et des tumeurs qui lui sont fatales.

Le virus est transmis de la mère à l'enfant et par voies sanguine ou sexuelle. Après contamination d'un individu, le virus est transporté en moins de 48 h dans les ganglions lymphatiques, où il infecte les lymphocytes T auxiliaires et les macrophages. Une fois l'ADN viral intégré dans la cellule infectée, il ne peut être éliminé ou détruit, sauf s'il y a destruction de la cellule elle-même. A la différence des lymphocytes T auxiliaires, l'infection des macrophages par le virus HIV-1 n'est pas cytopathique, ce qui implique que ceux-ci vont servir pour le virus de réservoirs capables de produire de nouveaux virions. C'est également le cas de certaines cellules dendritiques, qui piègent des particules virales sur leur surface et transfèrent ces dernières aux lymphocytes T dans le sang et les tissus lymphoïdes (pour revue, Levy and Fox, 1998).

Lors de la primo-infection, 2 à 6 semaines après la contamination, la multiplication virale est active, le nombre de lymphocytes T auxiliaires diminue progressivement et le virus se dissémine dans l'organisme (pour revue, Sleasman and Goodenow, 2003 ; Fig. 1). L'organisme met 6 à 8 semaines pour organiser une défense immunitaire, qui permet de diminuer la charge virale. La production des anticorps dirigés contre les antigènes du virus marque la séro-conversion. Les symptômes cliniques de la primo-infection sont en général similaires à ceux de la mononucléose.

Le stade suivant de l'infection par le VIH est la phase de latence, qui dure en moyenne une dizaine d'années. Cette phase de séropositivité est asymptomatique, mais la latence n'est qu'apparente car la destruction du système immunitaire progresse. Néanmoins, durant cette

[1] Human Immunodeficiency Virus type 1
[2] Syndrome d'ImmunoDéficience Acquise

période, le système demeure suffisamment intact pour prévenir la plupart des infections. C'est au cours de cette période qu'apparaissent des variants du virus.

Lorsqu'un nombre important de lymphocytes T auxiliaires a été détruit, que la production de nouvelles cellules lymphocytaires ne suffit plus à pallier l'infection, le système immunitaire est débordé, c'est le stade du SIDA clinique. Cette phase symptomatique dure en moyenne 2 à 4 ans et les nombreuses maladies opportunistes qui apparaissent conduisent à la mort du patient.

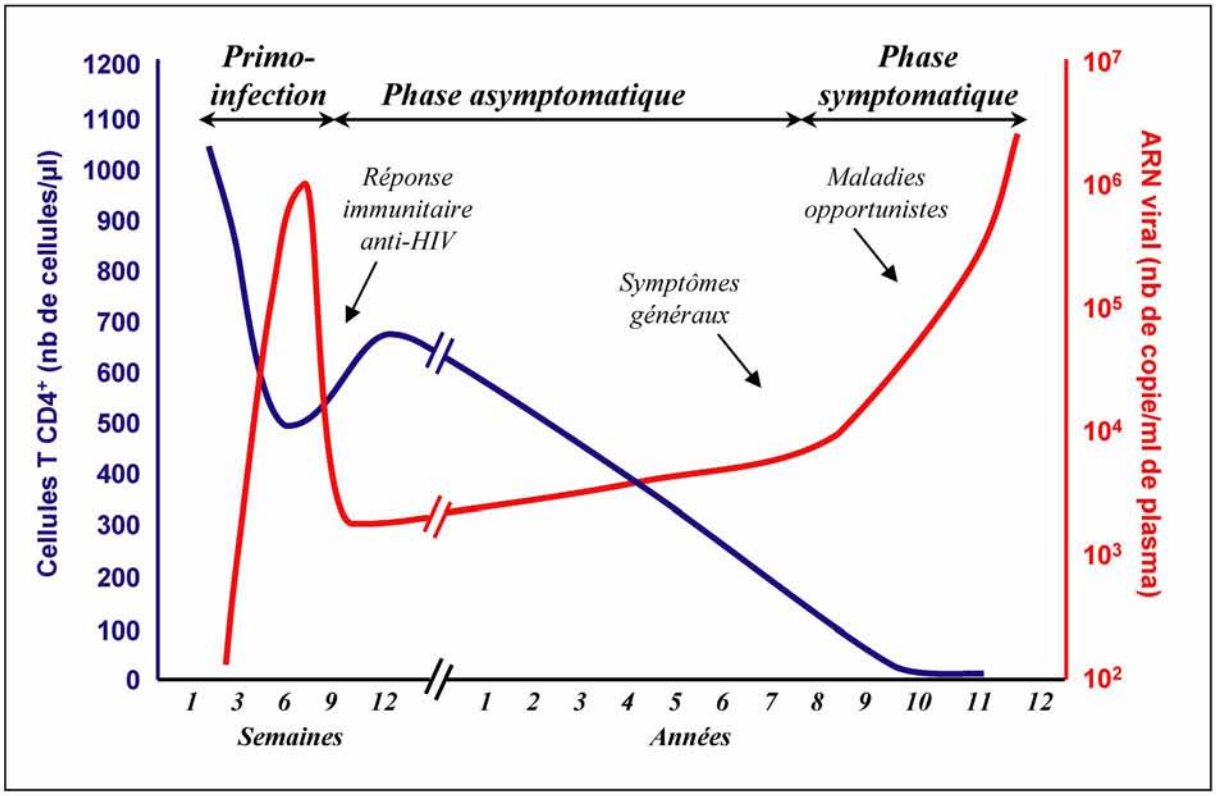


Fig. 1 : Les trois phases de la maladie

Les trois phases de la maladie (primo-infection, phase asymptomatique et phase symptomatique) sont schématiquement délimitées par des flèches. Sur ce graphique, la progression de la perte des cellules T CD4$^+$ est représentée par une courbe verte, tandis que l'évolution virale est représentée par une courbe rouge.

1.2. Evolution phylogénétique

En 1999, il a été proposé que le virus HIV-1 dérive du virus de l'immunodéficience du chimpanzé (SIVcpz), suggérant que l'introduction du virus HIV-1 dans l'espèce humaine résulte d'une transmission inter-espèce. L'Afrique de l'Ouest serait le lieu de transmission du chimpanzé à l'homme, puisque c'est la seule région dans le monde où se côtoient les hommes

infectés par le virus HIV-1 et les chimpanzés *Pan troglodytes troglodytes* infectés par le virus SIVcpz (Gao *et al.*, 1999).

Le virus HIV-2 proviendrait quant à lui d'une transmission à l'homme du virus SIVsm infectant les mangabeys (pour revue, Kuiken, 2001).

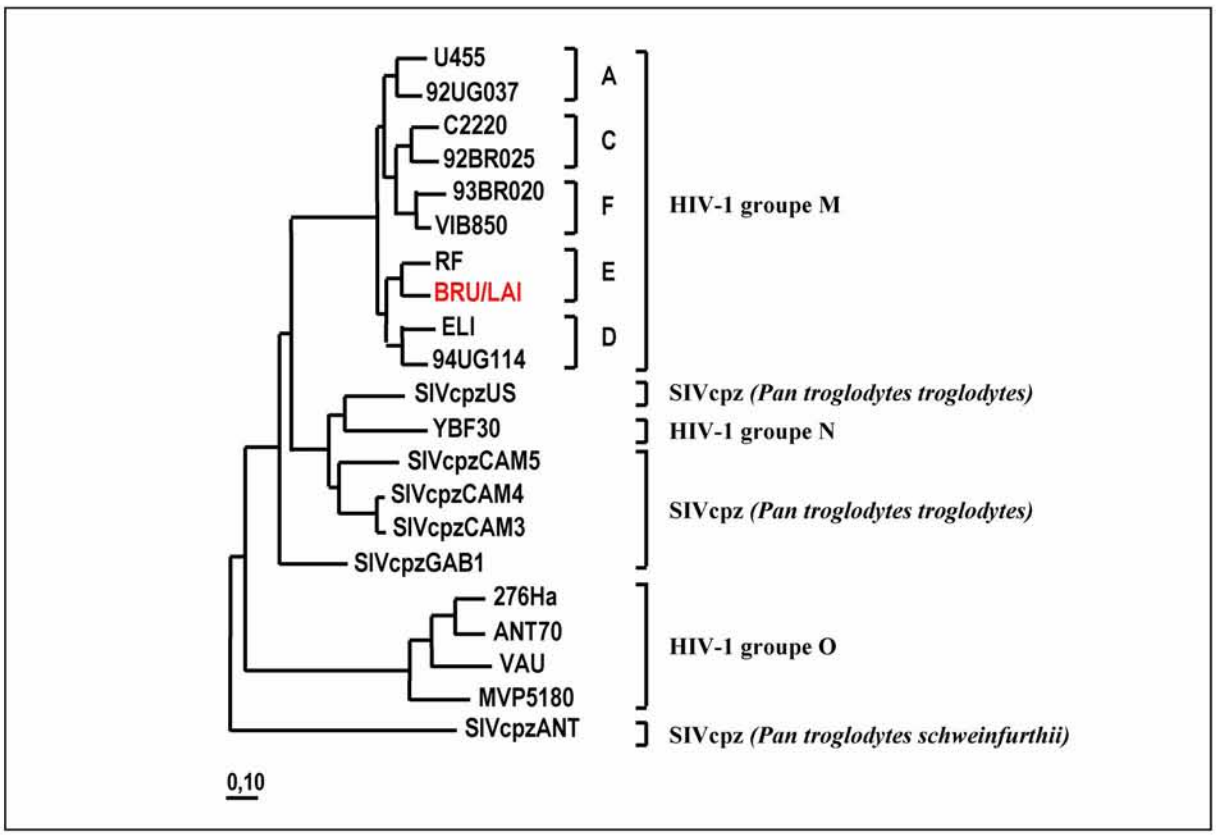


Fig. 2 : Dendrogramme représentant les liens phylogénétiques de souches de virus HIV-1 et SIVcpz

Les 3 groupes du virus VIH-1 (M, N et O) et 5 sous–types du groupe M (A et F) sont représentés sur la droite de la figure. Les espèces de chimpanzés à partir desquelles ont été isolées les virus SIV sont indiquées en italique. L'ARN de la souche Bru, appelée également LAI, qui a été utilisée au cours des travaux de thèse est indiquée en rouge. La barre d'échelle indique un remplacement de 0,1 acide aminé par site. Le dendrogramme est établi par la méthode de maximum de vraisemblance à partir de la séquence de la protéine Env (d'après Hahn et al., 2000)

Les souches caractérisées du virus HIV-1 sont réparties en 3 groupes phylogénétiquement liés, les groupes M[3], N[4] et O[5] (Fig. 2). Elles ont toutes la propriété de conférer le SIDA, mais c'est le groupe M qui est responsable de la pandémie (pour revue, Peeters and Sharp, 2000). 11

[3] Main
[4] Non M et non O
[5] Outlier

sous-types classés de A à K forment ce groupe M. Parmi eux, ce sont les souches de la classe B qui sont les plus étudiées au niveau de l'épissage. Nos études ont porté sur deux des souches de la classe B :

- BRU (ou LAI) qui est retrouvée principalement en Europe occidentale et en Amérique,
- et pNL4-3, qui est une forme recombinante de 2 souches: la partie 5' de son génome est empreint de la souche NY5, tandis que la partie 3' correspond à la souche BRU (Kuiken, 2001).

Les rétrovirus tels que HIV-1 sont capables de subir des recombinaisons intergénome pour créer de nouvelles souches virales. Ainsi, au sein du groupe M, 6 souches recombinantes ont été isolées. Ce phénomène se produit lorsqu'un individu est infecté par plusieurs souches virales. Le génome des souches résultantes est une combinaison de régions de différents sous-types. Par exemple, la souche MAL, largement répandue au Congo (ex-Zaïre), est une forme recombinante des souches A, D et I (Kuiken, 2001). Cet événement est une des raisons pour lesquelles il est très difficile de trouver des traitements**.**

2. Le virion : génome et structure

2.1. L'ARN génomique

Le génome du virus HIV-1 est constitué de 2 molécules d'ARN simple brin identiques, d'environ 9,2 kb, qui sont liées de manière non covalente par leurs extrémités 5'. Ces molécules d'ARN sont issues d'une transcription par l'ARN polymérase II cellulaire et sont donc toutes deux coiffées à leur extrémité 5' et polyadénylées à leur extrémité 3'.

La région 5' non traduite :

L'extrémité 5' de l'ARN génomique renferme les séquences R[6], U5[7] et PBS[8], nécessaires pour l'étape de transcription inverse (Fig. 3). La séquence R contient aussi l'élément TAR[9], nécessaire à l'activation de la transcription virale par la protéine Tat.

La région « leader », située entre le PBS et le codon initiateur AUG du gène *gag*, comporte : (1) une séquence autocomplémentaire appelée DIS[10] essentielle pour la mise en place du

[6] Redondant
[7] Unique 5'
[8] Primer Binding Site
[9] Trans-Activation Response
[10] Dimer Initiation Site

dimère d'ARN, (2) le site donneur d'épissage D1 et (3) l'élément Psi[11] nécessaire à l'encapsidation de l'ARN génomique.

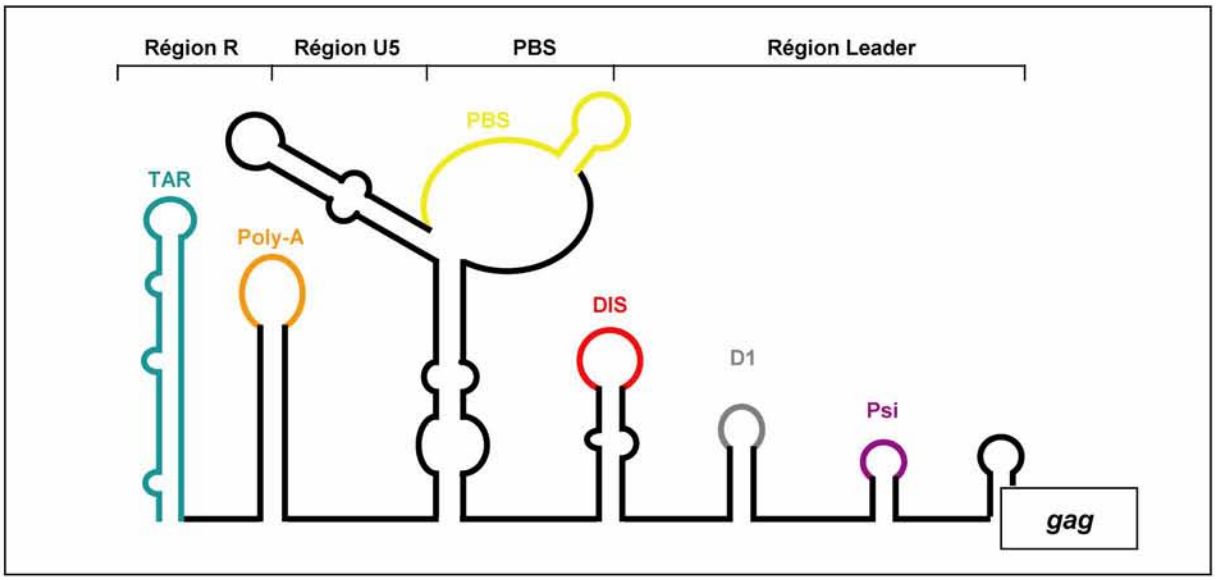


Fig. 3 : Représentation schématique de la structure secondaire de la région 5' non traduite

Les ARN viraux comportent tous l'élément TAR en 5'. Il est suivi par un site de polyadénylation non fonctionnel. La séquence PBS fixe l'ARNtLys3 cellulaire lors de l'initiation de la transcription inverse. La séquence DIS permet la dimérisation de 2 molécules d'ARN génomique. Le site donneur D1 délimite le premier exon de tous les ARNm viraux. Le signal Psi interagit avec la protéine NCp7 lors de l'encapsidation de l'ARN génomique dans les virions

Les sites d'épissage

Chez HIV-1, 5 sites donneurs (D1, D1A, D2, D3 et D4) et 9 sites accepteurs (A1, A1A, A2, A3, A4c, A4a, A4b, A5 et A7) d'épissage sont utilisés en combinaison pour générer les différents ARNm (Fig. 4). 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 HIV-1 contient 9 phases ouvertes de lecture (ORF[12] 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 pour les précurseurs protéiques du même nom (Fig. 4).

[11] Packaging signal
[12] Open Reading Frame

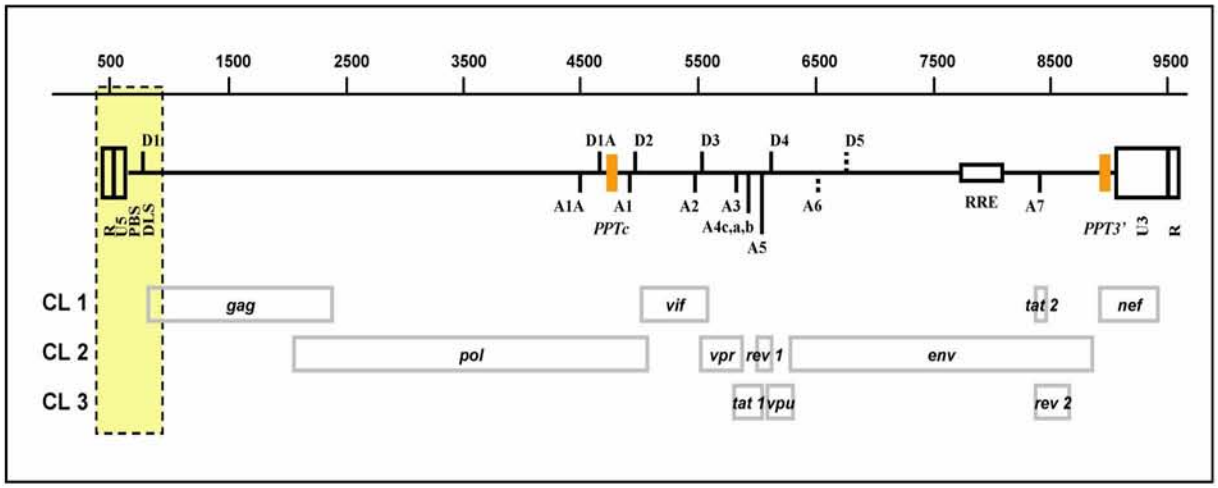


Fig. 4 : Localisation des séquences importantes de l'ARN viral

Localisation des séquences importantes de l'ARN viral : La région 5' non traduite est encadrée en jaune. Les différentes ORF des protéines sont représentées par des rectangles contenant le nom de l'ORF et disposées selon le cadre de lecture (CL) auquel elles appartiennent. Les sites donneurs (notés D) et accepteurs (notés A) d'épissage sont représentés par des barres verticales. Les séquences PPT sont indiquées par des rectangles orange.

Les séquences polypurines PPT' et PPTc[13]

Ces séquences (PPT3' et PPTc) sont importantes pour l'étape de transcription inverse. Elles servent d'amorce pour la synthèse d'ADNc (Charneau *et al.*, 1992).

La région 3' non traduite

L'extrémité 3' renferme la séquence U3[14] et la séquence R, identique à la séquence R présente en 5'. La séquence R contient le signal de polyadénylation (AAUAAA) qui n'est fonctionnel qu'en 3'.

2.2. La structure du virion

Les clichés obtenus par microscopie électronique et cryoélectromicroscopie montrent que les virions matures ont une forme sphérique, d'un diamètre moyen de 120 nm (Fig. 5A). Ces virions sont enveloppés d'une membrane phospholipidique d'origine cellulaire à l'intérieur de laquelle se trouvent la matrice (faible densité aux électrons) et la capside (partie dense aux électrons). Cette dernière comprend l'ARN génomique viral, des ARN d'origine cellulaire, ainsi que des protéines virales et cellulaires (Fig. 5B).

[13] PolyPurine Track
[14] Unique 3'

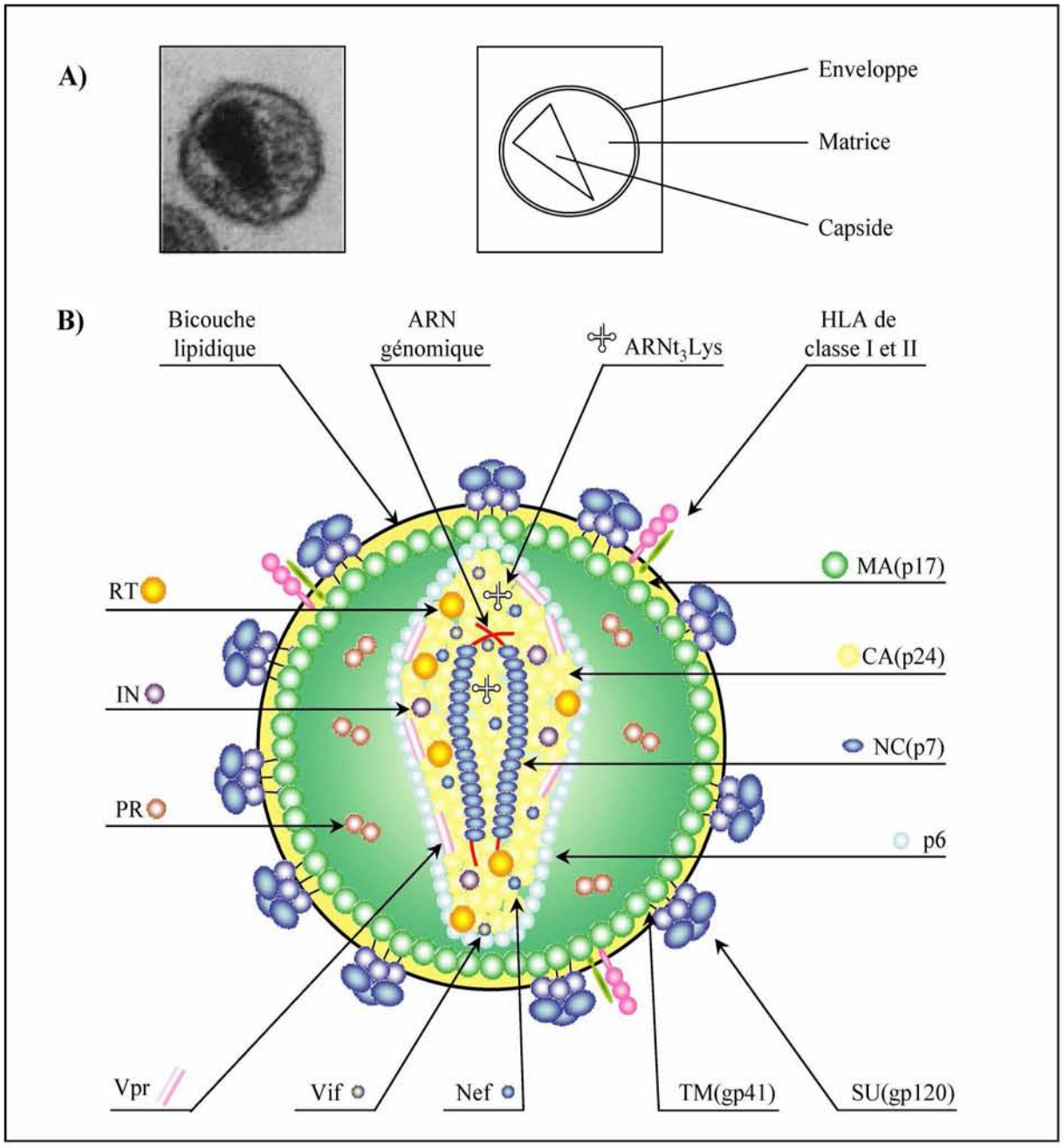


Fig. 5 : Structure du virion mature du virus HIV-1 *(adapté de Coffin et al., 1997)*

A) Particule virale mature vue en cryo microscopie électronique B) Représentation schématique de la particule virale. SU : protéine gp120 de surface ; TM : protéine gp41 transmembranaire ; MA : protéine de matrice ; CA : protéine de capside ; NC : protéine de nucléocapside ; RT : transcriptase inverse ; IN : intégrase ; PR : protéase ; Vpr : protéine virale R ; Vif : protéine virale d'infectivité ; Nef : facteur négatif.

L'enveloppe est constituée d'une bicouche lipidique à laquelle sont associées les protéines virales d'enveloppe gp41 (ou TM) et gp120 (ou SU). Cette bicouche lipidique est issue de la membrane plasmique de la cellule infectée lors du bourgeonnement du virion. Elle comporte

également des protéines dérivées de la membrane cellulaire de l'hôte, telles que les antigènes majeurs d'histocompatibilité (HLA-1 et HLA-2), l'actine et l'ubiquitine.
La matrice est constituée d'environ 2000 copies de la protéine virale p17 (ou MA). L'ensemble tapisse l'enveloppe grâce aux résidus myristilés, situés dans la partie N-terminale de la protéine, et insérés dans la face interne de l'enveloppe du virion (pour revue, Turner and Summers, 1999).

La capside est formée par des multimères de protéines virales p24 (ou CA). Elle renferme les 2 molécules d'ARN génomique entièrement recouverte de protéines virales de nucléocapside p7 (ou NC), mais aussi l'intégrase (IN), la transcriptase inverse (RT), la protéase (PR), la protéine régulatrice Nef et les protéines auxiliaires Vpr et Vif. Parmi les protéines virales, seules les protéines régulatrices Tat, Rev et la protéine auxiliaire Vpu ne sont pas présentes dans les virions. Le virus encapside par ailleurs des molécules provenant de la cellule hôte comme l'ARNt ^{Lys3}, utilisé lors de l'étape de rétro-transcription, et de la cyclophiline A qui interagit avec le précurseur Gag et est essentielle pour la réplication virale (Franke and Luban, 1996).

3. Cycle réplicatif du virus HIV-1

3.1. La phase pré-intégrative

La phase pré-intégrative comprend l'entrée du virion dans une cellule hôte, la transcription inverse de l'ARN génomique, puis le transport et l'intégration de l'ADN proviral dans le génome de la cellule hôte (Fig. 6).

3.1.1. Entrée du virion dans la cellule infectée

Comme mentionné ci-dessus, le virus HIV-1 infecte les cellules à récepteur CD4 comme les lymphocytes T auxiliaires, les cellules de la lignée monocyte-macrophage, les cellules dendritiques et les cellules neuronales. L'infection débute par la fixation spécifique de la glycoprotéine gp120 au récepteur cellulaire CD4 et aux co-récepteurs CCR5 ou CXCR4 qui sont respectivement présents en surface des macrophages ou les cellules T (pour revue, Tang *et al.* 1999). La fusion des membranes ne se réalise qu'après ancrage de gp41 dans la membrane plasmique et rapprochement de la cellule et du virion. Enfin la capside pénètre dans la cellule et subit un désassemblage partiel et progressif.

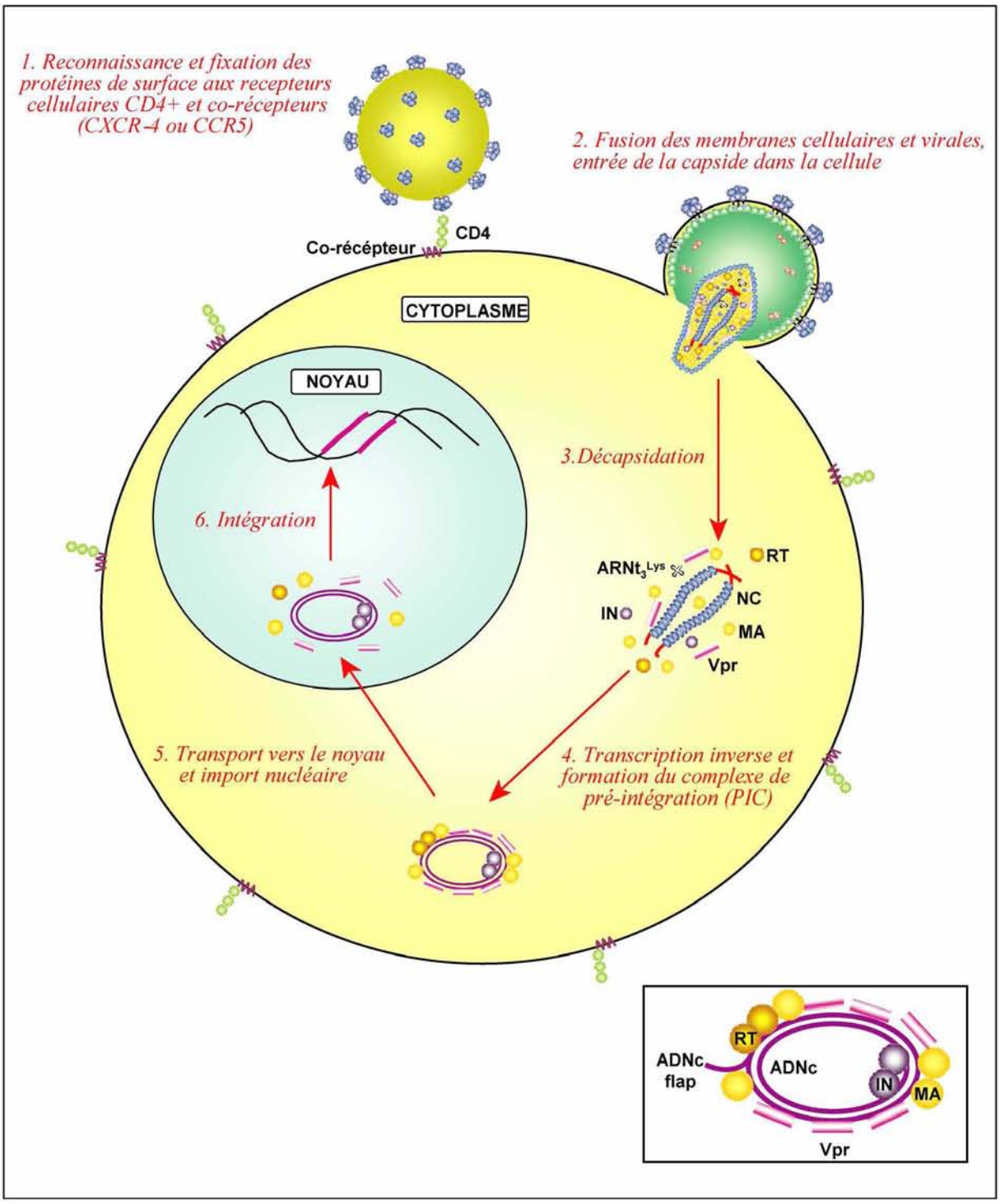


Fig. 6 : La phase pré-intégrative du cycle viral *(adapté de Greene and Peterlin, 2002)*

La phase pré-intégrative est divisée en 6 étapes. Le complexe de pré-intégration PIC est détaillé dans l'encart en bas à droite du schéma. L'ARN génomique et l'ADN proviral sont respectivement représentés en rouge et en violet.

RT : transcriptase inverse ; NC : nucléocapside ; MA : matrice ; IN : intégrase

3.1.2. Transcription inverse

L'ARN génomique viral contenu dans la capside, se retrouve dans le cytoplasme de la cellule infectée, au sein d'un complexe composé de la nucléocapside, de protéines de la matrice, de la protéine Vpr, de la transcriptase inverse et de l'intégrase.

La transcriptase inverse de part ces activités ADN polymérase ARN et ADN dépendante, et RNase H permet la synthèse de l'ADN proviral (Fig. 7).

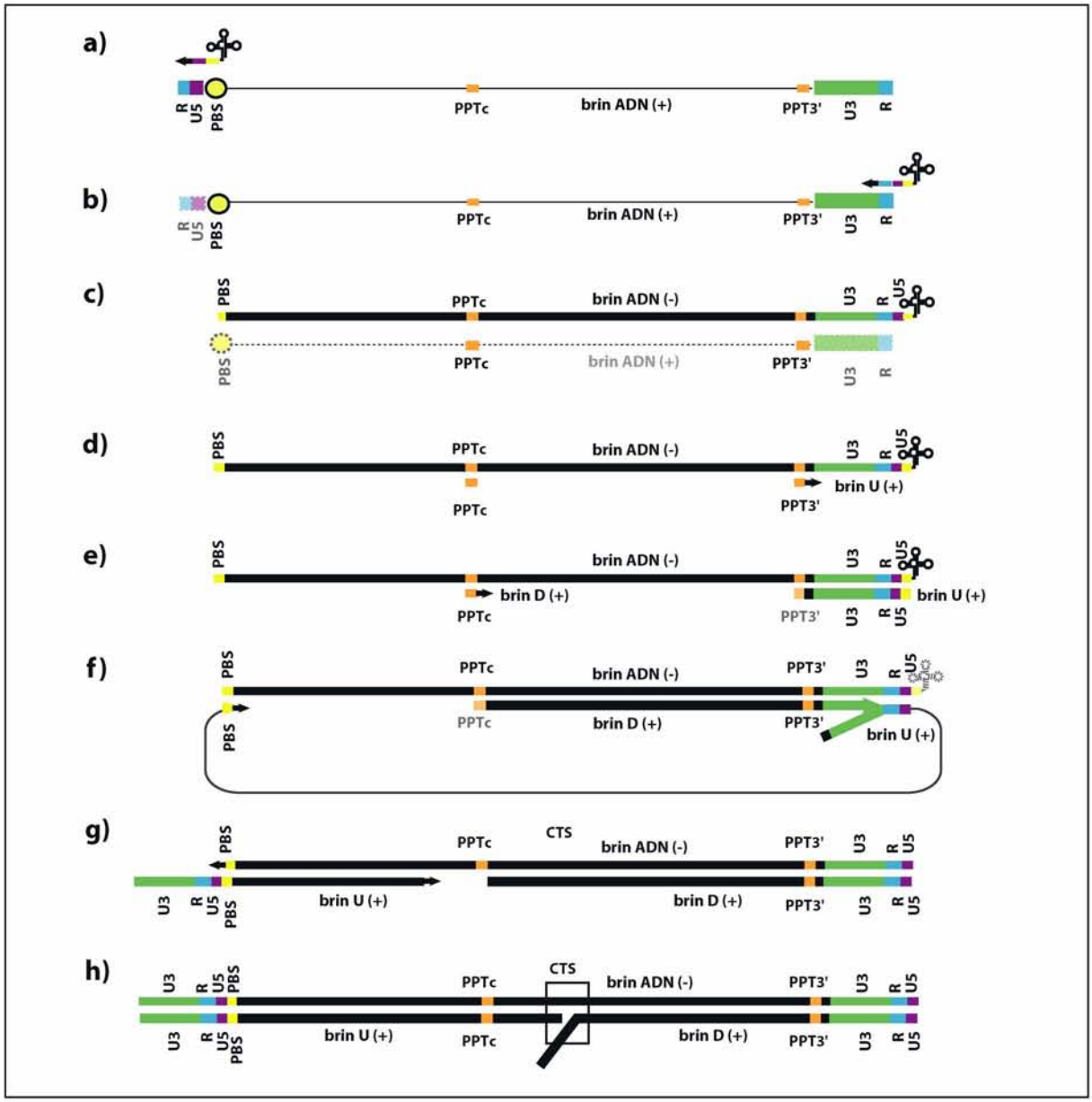


Fig. 7 : La rétro-transcription de l'ARN génomique de HIV-1 *(adapté de Gotte et al., 1999)*

a) Initiation de la synthèse du brin d'ADN (-). b) Dégradation par la RNase H de l'ARN matrice et premier saut de brin. c) Poursuite de la synthèse du brin d'ADN (-) et dégradation par la RNase H de la matrice (en pointillé). d) Initiation de la synthèse du brin d'ADN U (+) au niveau du PPT3'. e) Poursuite de la synthèse de l'ADN (+), dégradation de l'amorce PPT3' et seconde initiation de la

synthèse du brin d'ADN D (+) au niveau du PPTc. f) Dégradation de l'amorce PPTc et de l'amorce ARNt et second saut de brin. g) Terminaison de la synthèse des deux brins d'ADN. h) ADN double brin résultant appelé " DNA flap "car il présente un recouvrement de 100 nt au niveau de la région CTS (séquence de terminaison centrale).

NCp7, de part ses propriétés de protéine chaperon, joue un rôle important dans la transcription inverse, en favorisant les différentes étapes d'hybridation et deshybridation (pour revue, Darlix *et al.*, 2001).

La transcriptase inverse n'a pas la propriété de corriger ses erreurs. Des mutations ponctuelles sont donc introduites dans chaque nouvelle copie du génome viral, à un taux de 1 mutation pour 10 000 nt (pour revue, Sleasman and Goodenow, 2003). Des erreurs peuvent aussi être générées par action de cytidines désaminases cellulaires telles que APOBEC3G (APOB mRNA-Editing enzyme Catalytic polypeptide 3G ou CEM15) qui reconnaît les ADNc étrangers. Pour contrer ce processus d'édition, la protéine Vif inhibe l'encapsidation de l'enzyme APOBEC3G dans les particules virales, et entraîne sa dégradation par ubiquitination. Elle semble aussi limiter l'expression d'APOBEC3G au niveau traductionnel (Stopak *et al.*, 2003 ; Argyris and Pomerantz, 2004). La protéine Vpr joue aussi un rôle dans le maintien de l'information, puisqu'elle facilite le recrutement et l'encapsidation d'une uracile ADN glycosylase qui élimine les résidus uraciles présents dans l'ADNc proviral (pour revue, Le Rouzic and Benichou, 2005).

A l'issue de l'étape de transcription inverse, le génome viral se présente sous la forme d'une copie unique d'ADN double brin, dont les 2 extrémités, nommées LTR[15], sont identiques (séquence U3-R-U5) (pour revue, Gotte *et al.*, 1999).

3.1.3. Transport et intégration de l'ADN proviral dans le génome

Après la transcription inverse, des protéines virales s'assemblent autour de l'ADN proviral pour former le complexe de pré-intégration (ou PIC) (Greene and Peterlin, 2002). Ce complexe, composé de la nucléocapside, de protéines de la matrice, de la protéine Vpr, de la transcriptase inverse et de l'intégrase, migre vraisemblablement le long d'un réseau de microtubules du cytosquelette jusqu'aux pores nucléaires (Bukrinsky and Haffar, 1999 ; Haffar *et al.*, 2000 ; Bouyac-Bertoia *et al.*, 2001). Cet import nucléaire est facilité par la présence de la protéine MA et de Vpr qui présentent des signaux de localisation nucléaire (Gallay *et al.*, 1995 ; Popov *et al.*, 1998).

[15] Long Terminal Repeat

Après import, l'ADNc proviral est intégré dans le génome de la cellule hôte par l'intégrase (Fig. 8). L'enzyme reconnaît les extrémités LTR de l'ADN double brin et s'associe à l'ADN de la cellule hôte. Son rôle est de créer des extrémités 3' protubérantes, puis liguer l'ADNc proviral à l'ADN de la cellule infectée (Hindmarsh and Leis, 1999). Les « gaps » et les séquences excédentaires sont ensuite éliminés par les enzymes cellulaires assurant la réparation de l'ADN.

Les sites d'intégration utilisés ne sont pas définis mais dépendent de la structuration de la chromatine. L'intégration se fait souvent au niveau de sites actifs de transcription (Jung *et al.*, 2002).

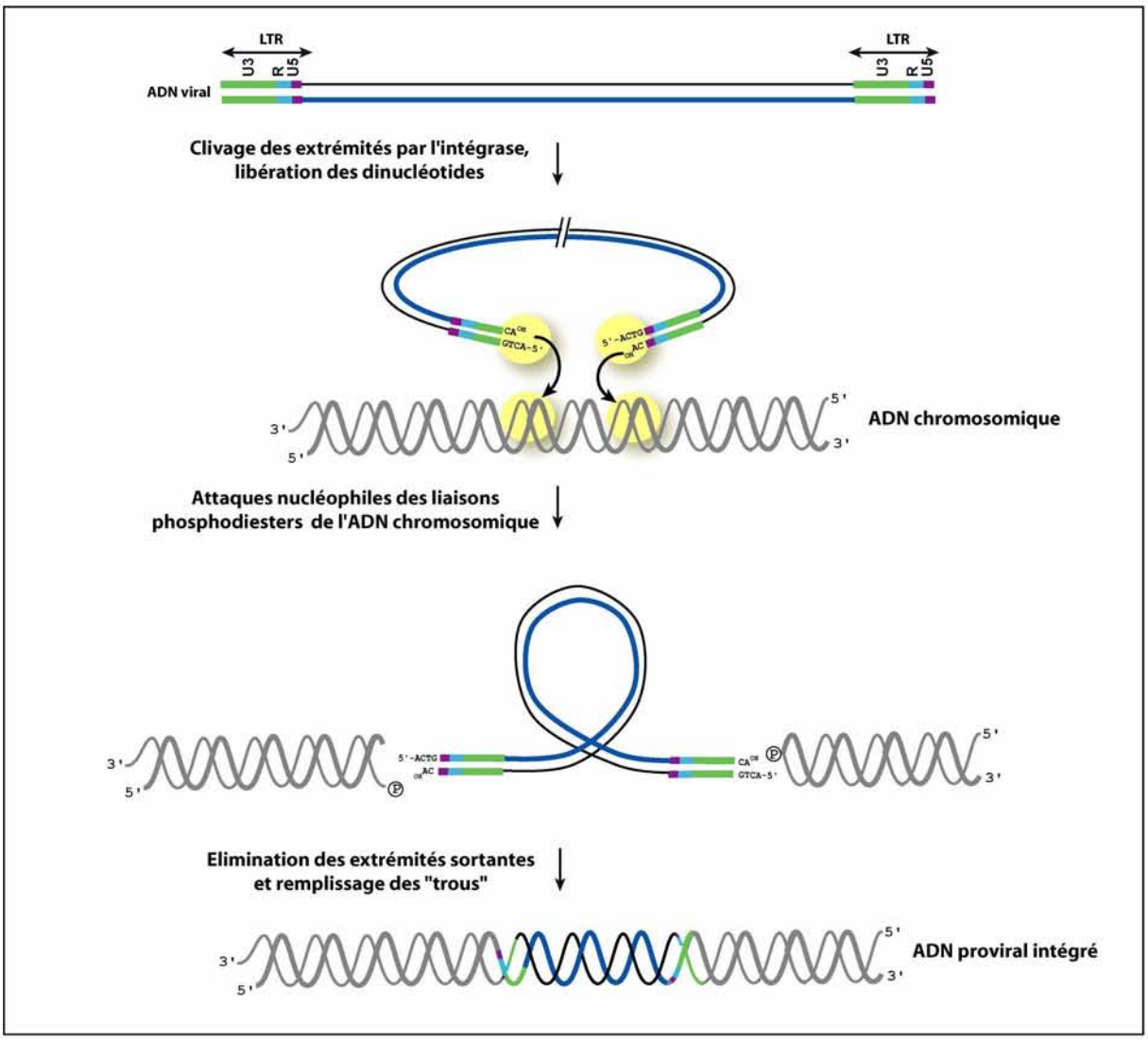


Fig. 8 : Intégration de l'ADN viral dans le génome de la cellule hôte *(d'après Hindmarsh and Leis, 1999)*

3.2. La phase post-intégrative

La phase post-intégrative regroupe les étapes d'expression des gènes viraux, d'assemblage, de bourgeonnement et de maturation des nouveaux virions (Fig. 9).

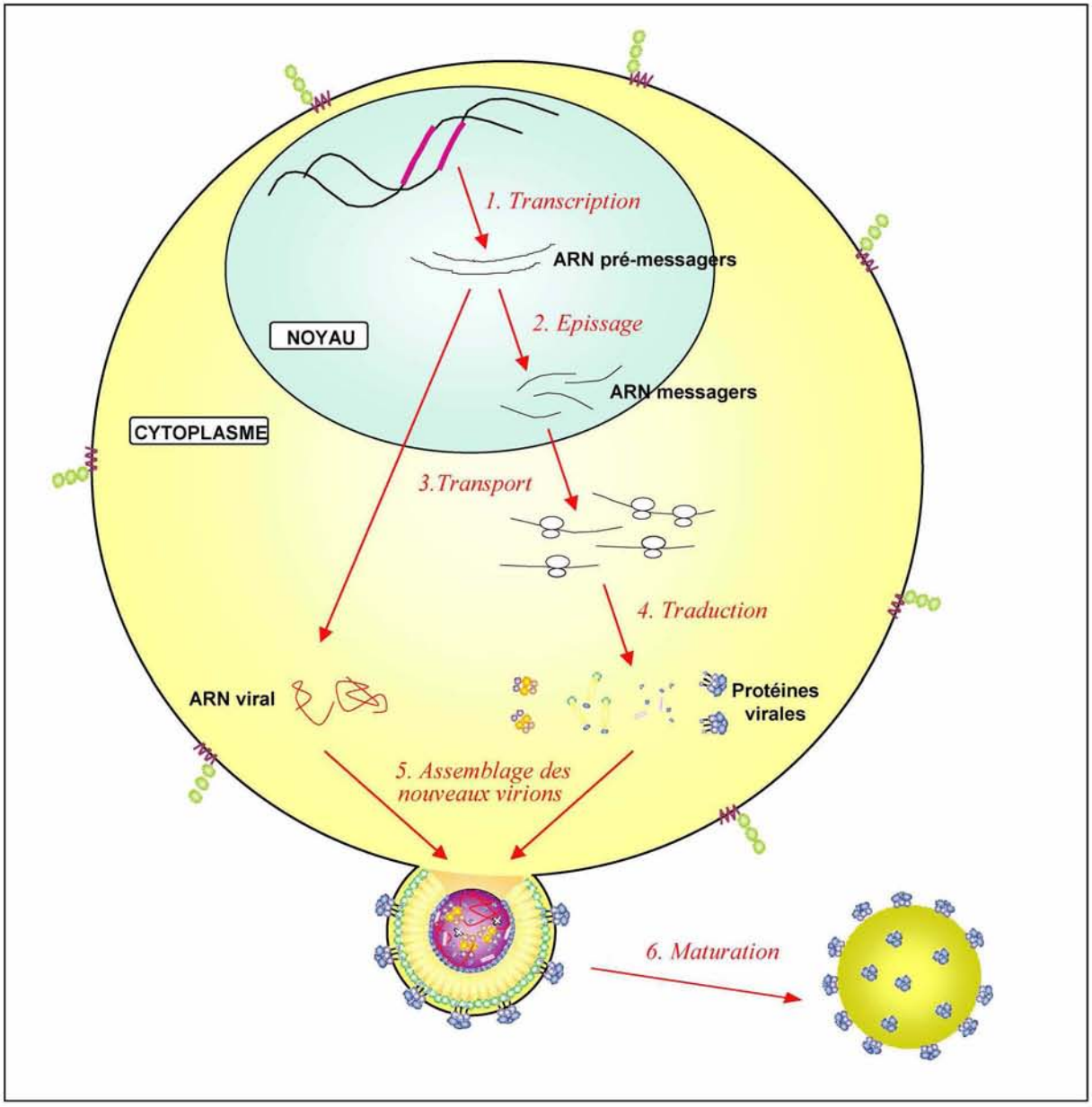


Fig. 9 : La phase post-intégrative du cycle viral *(pour revue, Coffin et al., 1997)*

3.2.1. Expression des gènes viraux

Le processus d'expression des gènes viraux comporte 5 étapes : (1) la transcription de l'ADN proviral et (2) l'épissage du transcrit primaire, que nous décrirons plus largement dans les chapitres 2 et 3 de cette introduction. Puis ont ensuite lieu les étapes (3) d'export nucléaire et (4) de traduction des ARN messagers et enfin (5) les modifications post-traductionnelles.

Nous allons voir que ces diverses étapes sont régulées différemment selon que l'on se trouve dans la phase précoce de l'expression des gènes viraux ou dans sa phase tardive.

- Transcription de l'ADN proviral

La transcription de l'ADN proviral est un point central de l'expression des gènes du virus HIV-1. De l'initiation jusqu'à la terminaison, le virus utilise l'ensemble de la machinerie cellulaire, mais une protéine virale est au cœur de l'efficacité du système : la protéine Tat[16].
Plusieurs séquences régulatrices fixant les facteurs d'activation de la transcription sont présents au niveau du LTR et assure la fixation de l'ARN Polymérase II au niveau du promoteur. En l'état, le complexe de transcription n'est pas efficace et produit des transcrits courts inutilisables. Par contre, en présence de la protéine Tat, le taux de transcription de l'ADN proviral est augmenté de 100 à 300 fois et les transcrits obtenus sont de pleines longueur. Ce phénomène appelé transactivation, ainsi que les divers aspects de la transcription chez HIV-1, feront l'objet d'une description plus détaillée dans le chapitre 2 de cette introduction. Enfin, la terminaison de la transcription s'effectue après la lecture du signal de polyadénylation situé dans la séquence R de la région 3' non traduite.

Autres activités de la protéine Tat

Il faut noter qu'en plus de son rôle sur la transcription, la protéine Tat remplit d'autres fonctions lors de l'infection d'une cellule par le virus HIV-1.
Au début de l'infection, il a été démontré que Tat est nécessaire pour l'initiation de la transcription inverse (pour revue, Harrich and Hooker, 2002). Bien plus tard après l'infection, du fait de la lyse des cellules infectées, la protéine Tat se retrouve dans le milieu extracellulaire. Par interaction avec les protéoglycanes héparanes sulfates présents à la surface des cellules, Tat entre dans les cellules non infectées puis passe dans leur noyau (Frankel and Pabo, 1988 ; Tyagi *et al.*, 2001). Plusieurs changements d'expression des gènes cellulaires importants pour le développement du virus HIV-1 lui sont attribués : 1) la stimulation de l'expression de cytokines immuno-régulatrices, 2) l'induction de l'expression de molécules d'adhésion, 3) la répression des gènes dont les produits limitent la multiplication virale et surtout 4) l'induction de l'apoptose, dont nous parlerons plus loin (pour revue, Huigen *et al.*, 2004).

[16] Transactivator

Implication de la protéine Vpr dans l'activation de la transcription

Dans les cellules nouvellement infectées, la protéine Vpr[17] présente dans le complexe de pré-intégration, semble pouvoir augmenter le taux de transcription en interagissant avec l'ADN proviral et des facteurs d'activation (Sp1 et TFIIB) (pour revue, Le Rouzic and Benichou, 2005). Cette activation aurait lieu avant la synthèse de la protéine Tat. Une autre explication possible de l'effet stimulateur de Vpr sur la transcription est sa capacité à arrêter le cycle cellulaire en phase G2, or c'est pendant cette phase que les LTR sont les plus actifs (Goh *et al.*, 1998).

Implication de la protéine Nef dans l'activation de la transcription

La protéine Nef joue beaucoup de rôles au cours du cycle viral. En particulier, son expression conduit à une augmentation de la production de l'Interleukine 2 qui conduit à une activation du facteur NF-κB, celle de la protéine CDK9 ou encore du facteur SF1, dont on verra plus loin qu'ils participent tous à la stimulation de la transcription (Arora *et al.*, 2002).

- Epissage des ARN pré-messagers

Le transcrit primaire de HIV-1 a plusieurs devenirs. Conservé intact, il sert soit d'ARN messager pour la production du précurseur Gag-Pol, soit d'ARN génomique pour les nouveaux virions produits. Il peut aussi être l'objet d'un épissage alternatif complexe donnant lieu aux ARN messagers permettant l'expression du précurseur Env, des protéines régulatrices et des protéines auxiliaires.

Les neufs ORF présentes sur l'ADN proviral peuvent être exprimées grâce à un épissage alternatif complexe gouverné par les 5 sites donneurs (D1, D1A, D2, D3 et D4) et 9 sites accepteurs (A1, A1A, A2, A3, A4cab, A5 et A7) d'épissage, ainsi que bien d'autres éléments régulateurs que nous décrirons plus précisément tout au long du manuscrit. Les messagers produits sont différents selon la phase, précoce ou tardive, et sont à l'origine des diverses protéines du virus.

Ainsi les protéines régulatrices Tat, Rev et Nef sont issues des ARN messagers de la classe 2 kb, alors que les protéines auxiliaires Vif, Vpu et Vpr, et le précurseur Env sont issus des ARN messagers de la classe 4 kb. Le précurseur Env donne par coupure protéolytique les protéines dites d'enveloppe que sont la protéine transmembranaire gp41 (ou TM) et la

[17] Viral protein

protéine de surface gp120 (ou SU). La dernière classe d'ARN messagers (9 kb) permet d'obtenir les précurseurs Gag et Gag-Pol, qui après traduction et maturation permettront d'obtenir respectivement les protéines structurales et les protéines à activité enzymatique (Fig. 10).

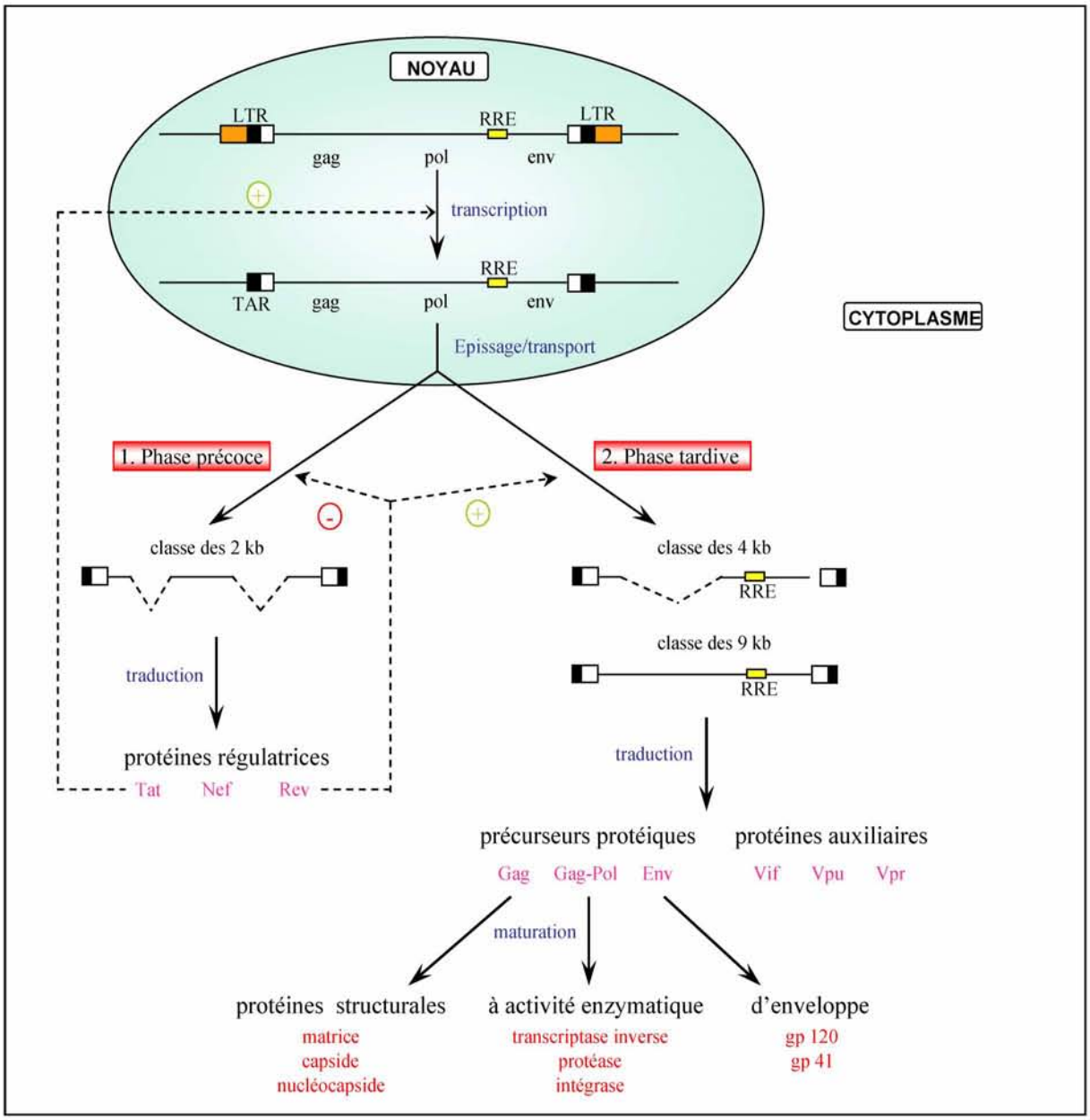


Fig. 10 : Les étapes de l'expression des gènes du virus HIV-1.

En phase précoce, seules les protéines régulatrices sont produites. Lorsqu'une quantité suffisante de protéine Rev a été produite, il y a passage en phase tardive au cours de laquelle les protéines auxiliaires, structurales et enzymatiques sont produites (pour revue ; Stutz and Rosbash, 1998 ; Hope, 1999)

- Transport des ARN messagers et de l'ARN génomique

Le maintien d'un équilibre précis entre les différents ARN (épissés et non épissés) est primordial pour la multiplication du virus. En plus d'une régulation fine de l'épissage, le virus HIV-1 produit la protéine Rev qui facilite le transport des transcrits non épissés.

La protéine régulatrice Rev[18] possède 4 domaines nécessaires à sa fonction: (1) un domaine d'activation qui agit comme signal d'export nucléaire, (2) deux domaines de multimérisation et (3) un domaine de fixation à l'ARN reconnaissant l'élément RRE[19]. Cet élément structuré de l'ARN viral se situe dans l'intron D4-A7 et contient un site primaire de forte fixation de la protéine. En fonction de leur concentration cellulaire, plusieurs autres protéines Rev peuvent ensuite venir se fixer sur RRE (pour revue, Hope, 1999).

En phase précoce, lorsque la protéine est présente en faible quantité dans la cellule infectée, la fixation n'a lieu qu'au niveau du site de haute affinité. L'action de Rev sur le transport est faible et les ARN épissés sont maintenus dans le noyau par interaction avec les facteurs d'épissage, jusqu'à ce qu'ils soient complètement épissés. En phase tardive, de plus grandes quantités de protéine Rev sont synthétisées et s'oligomérisent le long de l'élément RRE favorisant le transport des ARN peu ou pas épissés (Fig. 10).Les interactions créées au sein de l'oligomère ont pour effet de démasquer le signal d'export nucléaire. Le transport dans le cytoplasme se fait ensuite via le système d'export CRM1/Ran GTP (Fig. 11).

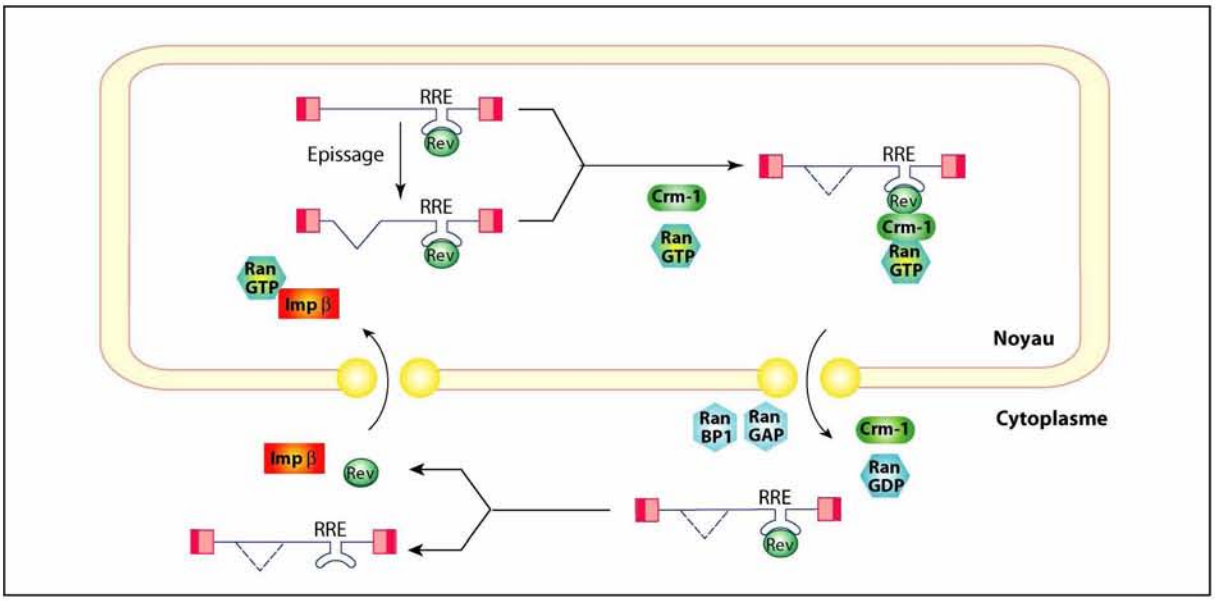


Fig. 11 : Modèle proposé pour le va-et-vient de la protéine Rev entre le noyau et le cytoplasme *(d'après Cullen, 2003 ; Wang et al., 2000)*

[18] Regulator of Virion Expression
[19] Rev Responsive Element

D'autres éléments de séquence semblent importants pour le transport des ARN par Rev. L'élément d'instabilité INS[20] contenu dans l'ORF Gag p17, et des éléments introniques riches en A et U, et répartis sur toute la longueur du génome viral, appelés CRS[21], pourraient intervenir dans la localisation des ARN viraux dans un compartiment spécifique du noyau. Ceci aurait pour but d'éviter la dégradation des ARN non épissés (Huffman and Arrigo, 1997).

Il faut donc noter que la protéine Rev joue un rôle primordial dans l'expression des gènes du virus HIV-1 puisqu'elle permet le passage de la phase précoce d'infection, caractérisée par la production des ARNm multiplement épissés codant les protéines régulatrices, à la phase tardive d'infection, au cours de laquelle les ARNm sont peu ou non épissés et permettent la production des protéines structurales, auxiliaires et à activité enzymatique.

- Traduction des ARN messagers

La synthèse des protéines virales est réalisée classiquement par la machinerie de traduction de la cellule hôte. Les protéines Env et Vpu, et les précurseurs Gag et Gag-Pol, sont obtenus à partir d'ARNm bi-cistroniques.
Dans le cas de Vpu et Env, les deux ORF sont chevauchantes. La traduction de la protéine Env dépend d'un mécanisme appelé « leaky scanning ». La faible efficacité du site d'initiation de la traduction de l'ORF Vpu autorise l'utilisation du codon d'initiation de l'ORF Env situé en aval (Schwartz *et al.*, 1990 ; Schwartz *et al.*, 1992).
La synthèse du précurseur protéique Gag-Pol est obtenue grâce à un décalage du cadre de lecture des ribosomes en fin de traduction du précurseur Gag. La présence d'une séquence « de glissement » associée à une structure tige-boucle permet un décalage du cadre de lecture d'un nucléotide (Jacks *et al.*, 1988 ; Karacostas *et al.*, 1993). Le rapport entre la synthèse de Gag et de Gag-Pol (20 :1) est critique pour la propagation du virus.

- Modifications post-traductionnelles

La protéine Env est glycosylée dans le réticulum endoplasmique, puis clivée par une protéase cellulaire (la furine) dans l'appareil de Golgi. Ceci libère les deux glycoprotéines gp120 et

[20] INstability Sequence
[21] *Cis*-acting Repressor Sequence

gp41 qui restent associés malgré la coupure et qui sont ensuite transportés au niveau de la membrane plasmique du virus pour l'assemblage des virions.

Lors de la traduction, la protéine Gag est myristylée en N-terminal. Cette modification permet son ancrage dans la membrane plasmique de la cellule, au site d'assemblage du nouveau virion (pour revue, Cimarelli and Darlix, 2002).

3.2.2. Assemblage, bourgeonnement et maturation des nouveaux virions

La formation d'une particule virale mature infectieuse requiert, dans un premier temps, le transport du précurseur Gag au niveau de la membrane plasmique et son association avec différents partenaires viraux et cellulaires. Dans un deuxième temps, les virions immatures, non infectieux, sont relargués et subissent une ultime étape de maturation pour devenir infectieux et se propager.

- Recrutement des protéines virales

Capable de s'auto-assembler et de former des pseudo-particules virales (ou VLP[22]), le précurseur Gag est indispensable à l'assemblage et au bourgeonnement de particules virales. Il contient plusieurs domaines fonctionnels :

-La myristilation de son extrémité N-terminale, ainsi qu'une région fortement basique dans le domaine de la matrice (MA), assurent l'ancrage du précurseur Gag dans la membrane plasmique. La partie MA du précurseur est préférentiellement localisée au niveau des radeaux lipidiques, ce qui explique la forte concentration en cholestérol et sphingolipides de la membrane virale (Holm *et al.*, 2003). Durant l'assemblage, la partie MA facilite aussi l'incorporation des protéines de l'enveloppe (pour revue, Tang *et al.*, 1999).

-Le domaine CA joue un rôle structural capital dans l'assemblage et la stabilité du virion (pour revue, Ganser-Pornillos *et al.*, 2008).

-Le domaine p6 est requis pour l'encapsidation de la protéine Vpr. Après la maturation du virion, la protéine Vpr se lie à la NC.

Par ailleurs, la protéine Vif pourrait jouer un rôle non négligeable dans l'assemblage de la particule virale et dans la stabilisation du coeur de la particule virale (Ohagen and Gabuzda, 2000). Son rôle serait lié à sa capacité à interagir à la fois avec l'ARN viral (Khan *et al.*, 2001) et avec les précurseurs protéiques Gag et Gag-Pol (Huvent *et al.*, 1998).

[22] Virus Like Particules

- Recrutement des ARN

Le recrutement de l'ARN viral

Une interaction spécifique entre les motifs de type doigt à zinc de la partie NC du précurseur Gag et une région localisée en 5' de l'ARN viral, est nécessaire à la sélection et l'encapsidation de l'ARN viral. Cette région, appelée domaine Psi ou région « leader » est composée de 4 structures tiges-boucles comprenant le site de dimérisation (DIS), le site donneur d'épissage D1, le site d'encapsidation (Psi) et le site d'initiation de la traduction de l'ORF Gag (pour revue, D'Souza and Summers, 2005).

Plus récemment, il a été démontré que de l'ARN épissé ainsi que certains ARN cellulaires sont aussi encapsidés (Liang *et al.*, 2004 ; Muriaux and Rein, 2003). Une première étude a permis de détecter que l'encapsidation de ces ARN ne dépend pas du signal Psi, mais le mécanisme reste encore incertain (Houzet *et al.*, 2007).

Le recrutement de l'ARNt ^{Lys3}

Durant l'assemblage du virus, l'ARNt ^{Lys3} est sélectivement encapsidé (environ 20 molécules par virion) (Huang *et al.*, 1994). 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 ^{Lys3} dépend de l'incorporation des précurseurs Gag-Pol et plus précisément de l'interaction entre la RT et l'ARNt ^{Lys3} (Khorchid *et al.*, 2000). L'encapsidation de lysyl-ARNt synthétase (LysRS), via son interaction avec le précurseur Gag, est aussi un facteur qui joue sur le taux d'incorporation de l'ARNt ^{Lys3} (Cen *et al.*, 2001 ; Guo *et al.*, 2003).

- Bourgeonnement et maturation du virion

Les particules virales immatures sont relarguées à partir de la membrane plasmique grâce à la machinerie cellulaire de relarguage des corps multi-vésiculaires, dont les composants clés sont les complexes multiprotéiques ESCRT (Martin-Serrano *et al.*, 2003). Le domaine p6, par son interaction avec le facteur Tsg101 qui fait parti du complexe multiprotéique ESCRT-I, joue un rôle essentiel dans cette étape (Gottlinger *et al.*, 1991).

La dernière étape du cycle réplicatif est la maturation des virions, qui a lieu de façon concomitante ou juste après le bourgeonnement. La maturation conduit à des changements morphologiques majeurs de la particule virale. La protéase virale joue un rôle crucial dans

cette étape. Durant l'assemblage, l'interaction de deux molécules de précurseurs Gag-Pol permet la dimérisation des domaines protéase et la formation de l'homodimère correspondant à la forme active de la protéase virale (PR). Cette enzyme initie sa propre maturation par un mécanisme autocatalytique, mais permet aussi d'obtenir les autres protéines à activité enzymatique p66 (RT) et- p32 (IN) (Debouck *et al.*, 1987). Un clivage ordonné et contrôlé des précurseurs protéiques Gag est requis pour permettre la formation d'un virion infectieux (Wiegers *et al.*, 1998). Le premier clivage libère la protéine NC, qui forme un complexe ribonucléoprotéique en se fixant sur le dimère d'ARN génomique. Le clivage suivant libère la protéine CA, permettant la formation de la capside du virion (Fig. 12).

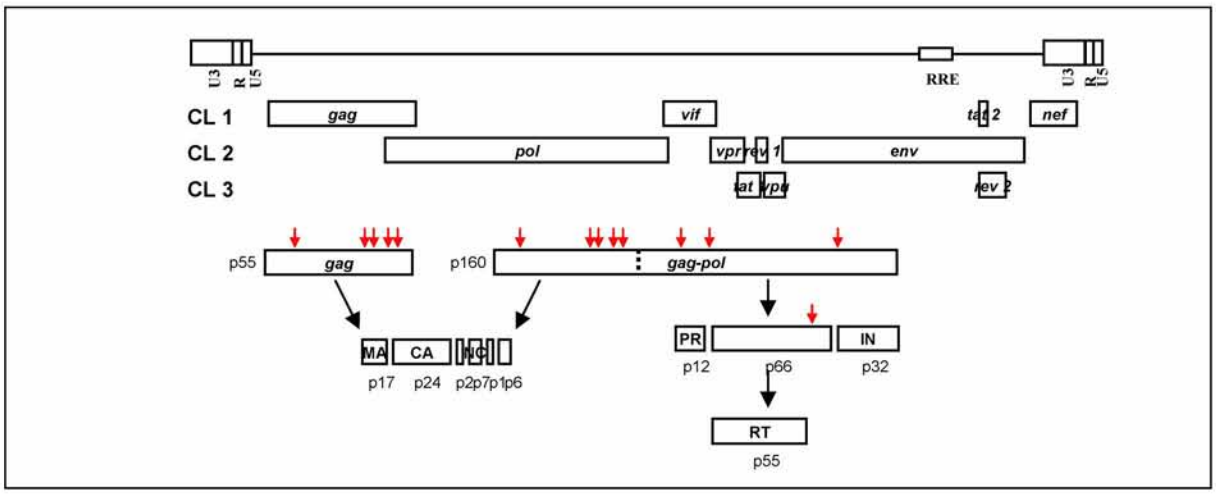


Fig. 12 : Maturation des précurseurs Gag et Gag-pol. *(pour revue, Coffin et al., 1997)*

Le génome du virus HIV-1 est schématiquement représenté. Les précurseurs Gag et Gag-pol sont produits à partir des ARN de la classe 9 kb et maturés par la protéase virale PR. La maturation du précurseur Gag conduit à la production des polypeptides p17 (MA), p24 (CA), p2, p7 (NC), p1 et p6. La maturation du précurseur Gag-pol conduit à la libération des polypeptides p12 (PR), p32 (IN) et p66, ce dernier étant maturé ultérieurement en p55 pour permettre de former l'hétérodimère p55/p66 de la RT.

Récemment, il a été montré que Vif pouvait, à forte concentration, inhiber l'infectivité du virus en bloquant une étape précise de la maturation des précurseurs protéiques (Akari *et al.*, 2004). Vif n'agirait pas comme inhibiteur de protéases mais empêcherait spécifiquement le clivage entre la partie NC et la partie CA.

3.2.3. Inhibition de la réponse immunitaire et apoptose

- Diminution des récepteurs CD4

La synthèse des récepteurs CD4 a lieu au niveau du réticulum endoplasmique. Avant que les molécules de précurseurs Env transitent dans l'appareil de Golgi, des complexes CD4-Env (gp160) peuvent être formés, ce qui conduit à la destruction des CD4. Il faut noter que, dans le cas où des quantités équimolaires de récepteurs CD4 et de protéines Env sont synthétisées, cela conduit non seulement à la perte des récepteurs CD4 à la surface des cellules mais également à celle des protéines Env, conduisant à la formation de virus Env-déficient et donc non infectieux.

Deux protéines virales Vpu et Nef jouent un rôle important dans la dégradation des récepteurs CD4 au niveau du réticulum endoplasmique et à la surface des cellules. La protéine Vpu est une protéine membranaire de 16 kDa qui induit la dégradation rapide des récepteurs CD4 dans le réticulum endoplasmique. Deux résidus sérine conservés de la protéine Vpu doivent être phosphorylés pour induire la dégradation des récepteurs CD4, et interagir avec la protéine β-TrCP[23]. L'interaction simultanée de Vpu avec les CD4 et cette protéine permet le recrutement de Skp1, qui s'associe à la protéine β-TrCP. Ces événements induisent l'ubiquitination et la dégradation protéasomale des récepteurs CD4 avant qu'ils n'atteignent la surface de la cellule infectée (Margottin *et al.*, 1998 ; Bour and Strebel, 2003). La protéine Nef, diminue la quantité de récepteurs CD4 à la surface des cellules infectées par un mécanisme d'endocytose suivi d'une dégradation par la voie endo-lysosomale, impliquant l'interaction directe de Nef avec les CD4 (Salghetti *et al.*, 1995 ; Das and Jameel, 2005). L'action de ces deux protéines contribue à augmenter d'une part, le relarguage des particules virales en évitant la liaison de Env avec les CD4 et d'autre part, l'incorporation de Env dans les virions, ce qui accroît l'infectivité des virions produits.

La diminution des récepteurs CD4 à la surface de la cellule infectée est par ailleurs très importante pour la survie du virus puisque cela évite une surinfection de la cellule par d'autres virus, ce qui conduirait à sa mort rapide.

- La réponse immunitaire est inhibée par les protéines virales

Plusieurs protéines virales sont impliquées dans l'inhibition de la réponse immunitaire, ceci par différents mécanismes. La protéine Nef est capable de réduire l'expression des molécules

[23] β-Transducin-repeat Containing Protein

du complexe majeur d'histocompatibilité de classe I (ou CMH-1) à la surface des cellules infectées, ceci par un mécanisme impliquant l'endocytose des complexes suivis de leur dégradation. Ceci conduit à protéger les cellules infectées de la destruction par les lymphocytes T cytotoxiques (pour revue, Arora *et al.*, 2002). De plus, elle intervient dans un mécanisme limitant le nombre de molécules du complexe majeur d'histocompatibilité de classe II (ou CMH-2) à la surface des cellules infectées en induisant leur dégradation, ce qui limite la reconnaissance des cellules infectées par les cellules T helpers.

La protéine Vpr inhibe aussi la réponse immunitaire en intervenant dans un mécanisme qui permet à la fois la diminution de la production d'anticorps dirigés contre le virus par les lymphocytes T cytotoxiques (CD8) et celle des cytokines pro inflammatoires, telles que IL-12 et TNFα et des chimiokines. La protéine Vpr aurait un effet général sur la progression de la maladie puisque, chez les macaques infectés par un virus SIVmac219 ou le gène *vpr* est délété, une diminution de la charge virale, ainsi qu'une diminution de la progression de la maladie a été observée (Gibbs *et al.*, 1995 ; Desrosiers *et al.*, 1998).

- L'apoptose et les protéines virales

Dans la cellule, il existe différents activateurs cellulaires de l'apoptose dont FasL, TNF-α, Bad et p53. En augmentant l'expression des ligands FasL dans les cellules infectées, la protéine Nef induit l'apoptose des cellules T-cytotoxiques qui présentent des récepteurs Fas en surface et interagissent avec FasL de la cellule infectée. La protéine Nef augmente également l'expression des récepteurs Fas dans les cellules infectées, mais en contre partie elle inhibe toutes les voies d'apoptose de la cellule hôte afin de la protéger. Tout d'abord, Nef inhibe la kinase ASK-1[24], l'intermédiaire clé des voies apoptotiques impliquant Fas et TNF-α. Elle inhibe aussi la fonction apoptotique de la protéine p53 en se liant directement et bloquant son activité. Enfin, la protéine Nef active deux kinases qui phosphorylent Bad et inhibent sa fonction (Das and Jameel, 2005).

Contrairement à la protéine Nef, les protéines Vpu et Tat augmentent l'apoptose des cellules infectées :

- en inhibant le facteur NF-kB, pour la protéine Vpu, ce qui conduirait à une baisse de la synthèse des protéines anti-apoptotiques et à une dérépression des caspases,

[24] Apoptosis aSsociated Kinase-1

- et, pour la protéine Tat, en activant l'expression de la caspase 8 de la voie de transduction du signal apoptotique (Bartz and Emerman, 1999) et de Fas-L (Westendorp *et al.*, 1995 ; pour revue, Karn, 2000).

Il faut enfin noter, que la protéine Tat, lorsqu'elle pénètre dans les cellules non infectées, a la capacité d'y activer la transcription de FasL et ainsi de les sensibiliser à l'induction de l'apoptose par interaction de FasL avec le récepteur Fas. Ceci fait parti du mécanisme qui conduit à l'apoptose massive des cellules lymphoïdes qui caractérise la pathologie SIDA (pour revue, Karn, 2000).

CHAPITRE 2 : Description générale de la transcription par l'ARN Pol II : cas particulier de l'ADN proviral de HIV-1

1. La transcription : description générale

1.1. Généralités

Chez les procaryotes une seule ARN polymérase ADN dépendante effectue la transcription pour tous les types d'ARN. Chez les eucaryotes par contre, ce sont trois ARN polymérases (ARN Pol) qui interviennent :

- l'ARN Pol I pour la transcription des ARN ribosomiques dans le nucléole (28S, 18S et 5,8S),
- l'ARN Pol II pour la transcription des ARN messagers ainsi que celle des snoRNA[1] et des snRNA[2] spliceosomaux à l'exception de l'ARN U6,
- et l'ARN Pol III pour la transcription des ARNt, ARNr 5S et autres petits ARN.

Pour certains virus à ARN, l'ARN est transcrit par une ARN Pol ARN dépendante aussi appelée réplicase et exprimée par le virus. Dans cette introduction, nous nous intéresserons à l'ARN Pol II et plus particulièrement à toute la machinerie qui participe à la transcription des gènes nucléaires codant pour des protéines chez les eucaryotes.

La transcription comporte 4 étapes :

- la formation du complexe de pré-initiation (ou PIC[3]) autour du promoteur,
- l'initiation de la réaction de transcription, avec toutes les régulations qu'elle comporte,
- la phase d'élongation de la transcription qui est concomitante avec la mis en place de la coiffe et l'épissage de l'ARN pré messager,
- et enfin, l'étape de terminaison qui suit celles de clivage et de poly-adénylation de l'ARN.

1.2. La formation du complexe de pré-initiation

Contrairement à l'ARN polymérase des procaryotes, l'ARN Pol II des eucaryotes ne peut initier la transcription seule. Cinq facteurs de transcription généraux interviennent dans

[1] small nucleolar RNA
[2] small nuclear RNA
[3] Pré-Initiation Complex

l'assemblage du complexe de pré-initiation autour du promoteur. Il s'agit des facteurs TFIIB, -D, -E, -F et –H[4] (Fig. 13 ; pour revue Hirose and Ohkuma, 2007).

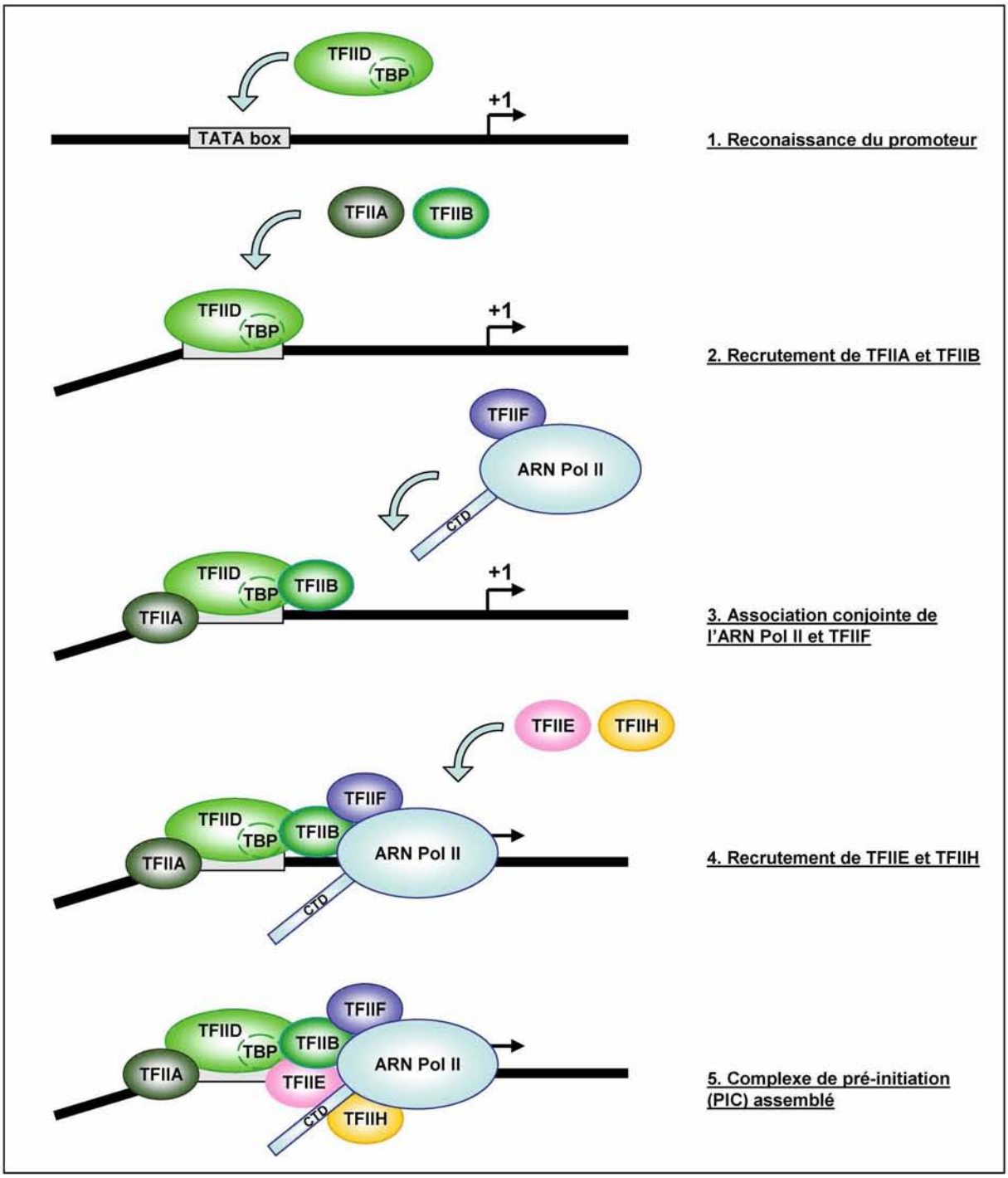


Fig. 13 : Représentation schématique de l'assemblage séquentiel du complexe de pré-initiation de la transcription sur le promoteur basal.

[4] Transcription Factor for RNA polymerase II

La formation du complexe débute par la fixation de la sous unité TBP[5] de TFIID sur la boite TATA qui induit une courbure de l'ADN (Roeder, 1996). Le facteur TFIIA se joint au complexe ADN-TFIID en interagissant directement avec TBP (Geiger *et al.*, 1996) permettant ainsi de maintenir le complexe établi et de poursuivre le processus d'assemblage de la machinerie. TFIIB reconnaît alors TBP et la courbure de l'ADN ainsi qu'une séquence située en aval appelée BRE[6] lorsqu'elle est présente (Lagrange *et al.*, 1998). Par la suite, TFIIB se lie aussi au facteur TFIIF associé à l'ARN Pol II et assiste la polymérase dans l'initiation de la réaction de transcription elle-même (Bushnell *et al.*, 2004). La fixation du facteur TFIIF lié à l'ARN Pol II permet le positionnement de la polymérase au site +1 d'initiation de la transcription. Après cette séquence d'événements, TFIIE stabilise le complexe en établissant plusieurs liens avec tous les facteurs de transcription et l'ARN Pol II (Ohkuma, 1997). Trois rôles supplémentaires lui sont attribués dont certains sont en lien avec TFIIH: (1) la sous unité TFIIEβ se lie à l'ADN à proximité du promoteur (de -9 à +2), assiste TFIIH dans l'ouverture du promoteur grâce à son activité ADN-hélicase, et aide l'ARN Pol II à placer le site d'initiation du brin matrice dans le site actif (Douziech *et al.*, 2000), (2) la sous unité TFIIEα se lie à TFIIH et stimule son activité kinase responsable de la phosphorylation du domaine C terminal de la grande sous unité de l'ARN Pol II (ou CTD[7]) (Ohkuma and Roeder, 1994), et (3) la sous unité TFIIEβ participe enfin au changement conformationnel de l'ARN Pol II qui permet sa conversion de l'état ouvert à l'état fermé (Westover *et al.*, 2004). Comme déjà mentionné ci-dessus, TFIIH, et ses 3 activités CTD-kinase, ADN-hélicase et ATPase, fait aussi parti du complexe de pré-initiation (Svejstrup *et al.*, 1996).

Sur le promoteur, le PIC peut aussi être associé à d'autres complexes appelés Médiateur. Ces derniers sont composés de plus de 20 sous unités et jouent un rôle clé dans la transmission des signaux d'activation de la transcription. Tous favorisent le remodelage de la chromatine grâce à des activités histone-acetylase et activent la transcription grâce à une activité CTD-kinase. De récentes études ont en effet mis en évidence les interactions entre des complexes médiateurs isolés chez l'homme, l'ARN Pol II et la kinase 8 cycline dépendante (ou CDK8[8]). En présence de la cycline C, la kinase phosphoryle le CTD de la polymérase de la même façon que TFIIH (Furumoto *et al.*, 2007 ; Davis *et al.*, 2002).

[5] TATA box Binding Protein
[6] TFIIB Recognition Element
[7] Carboxy Terminal Domain
[8] Cyclin Dependant Kinase 8

1.3. Le rôle central de la phosphorylation du CTD

Une fois le PIC mis en place sur le promoteur, plusieurs changements conformationnels de l'ARN Pol II doivent avoir lieu pour qu'il y ait synthèse du transcrit. Ces modifications de structure s'opèrent de deux manières distinctes :

- la première consiste en plusieurs interactions physiques entre les facteurs TFIIB, -F et -E, ou le Médiateur d'une part, et l'ARN Pol II d'autre part (Hayashi *et al.*, 2005 ; Davis *et al.*, 2002). Ces interactions ont lieu dès la formation du PIC.

- l'autre façon est la phosphorylation du CTD de l'ARN Pol II. Le CTD est constitué de plusieurs heptapeptides répétés (YSPTSPS), qui sont plus ou moins phosphorylés au niveau des serines 2 et 5 au cours de la transcription. Pendant l'initiation, Ser 5 est la principale cible des CTD-kinase de TFIIH et du Médiateur (Komarnitsky *et al.*, 2000 ; Hengartner *et al.*, 1998). Durant l'élongation, c'est Ser 2 qui devient la principale cible d'une 3ème activité CTD kinase, celle portée par le facteur positif d'élongation de la transcription (ou P-TEFb[9]) (Ni *et al.*, 2004 ; Fig. 14).

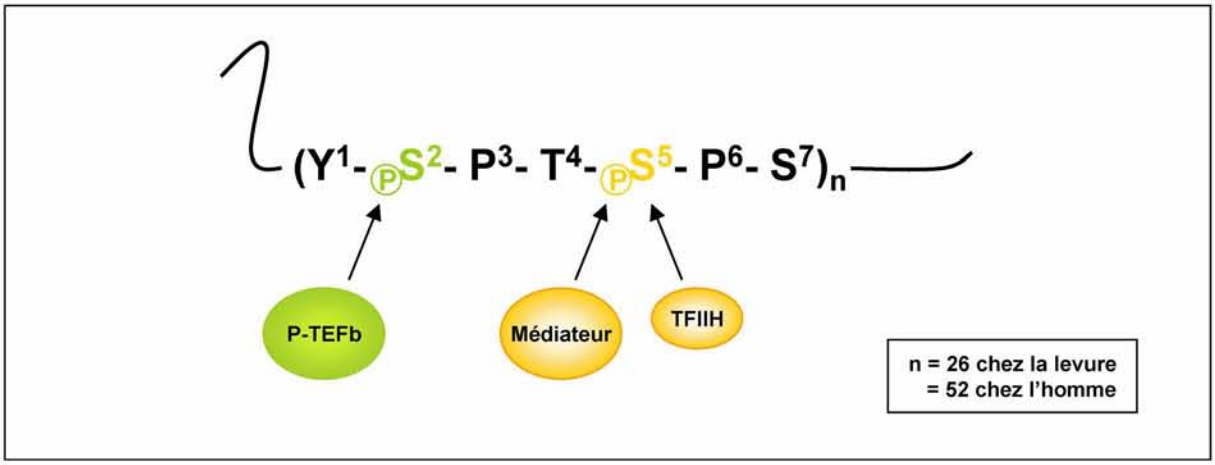


Fig. 14 : Site de phosphorylation du CTD *(D'après Hirose and Ohkuma, 2007)*

Le domaine C-terminal (CTD) de la grande sous unité de l'ARN Pol II est constitué de plusieurs heptapeptides (YSPTSPS) répétés. Le taux de phosphorylation et le type de sérines phosphorylées (2 ou 5) varient au cours de la transcription.

Récemment, il est apparu que la phosphorylation du CTD au niveau des sérines 2 et 5 n'est pas seulement essentielle à la transcription, mais aussi au remodelage de la chromatine du gène en aval du site d'initiation de la transcription, et à tous les processus de maturation de l'ARN associés à la transcription. Des études sur la levure ont mis en évidence que le

[9] Positive Transcription Elongation Factor b

recrutement, après l'initiation de la transcription, du complexe contenant l'histone méthyl-transférase, mais aussi des enzymes de la coiffe et les enzymes de l'épissage, est dépendant de cette phosphorylation. En phase d'élongation, la phosphorylation de Ser 2 permet le recrutement de nouvelles histone méthyl-transférases et des enzymes de la poly-adénylation (Orphanides and Reinberg, 2002 ; Hampsey and Reinberg, 2003).

- l'initiation de la transcription (Fig. 15.1 et 2)

Lors de l'initiation de la transcription, la phosphorylation des Ser 2 du CTD est réalisée par la sous unité CDK7 et la cycline H du facteur TFIIH (Giglia-Mari *et al.*, 2004). Cette phosphorylation permet l'assemblage des premiers nucléotides de l'ARN et s'accompagne du décrochement du complexe d'initiation du promoteur. Le PIC en soi est suffisant pour obtenir une activité transcriptionnelle *in vitro*, mais à très faible taux. Cette activité basale peut être moduler par le Médiateur qui grâce à sa sous unité CDK8 et la cycline C associée, phosphoryle le CTD de manière identique à TFIIH (Davis *et al.*, 2002).

La mis en place de la coiffe en 5' s'effectue aussi pendant l'initiation de la transcription lorsque l'ARN naissant mesure entre 20 et 30 nucléotides (Hirose and Manley, 2000). Ce processus n'intervient que sur les transcrits Pol II. Il dépend non pas de séquence ou structure particulière de l'ARN, mais du niveau de phosphorylation des sérines 5 du CTD de l'ARN Pol II (McCracken *et al.*, 1997), ainsi que d'autres interactions protéine/protéine qui favorisent le recrutement des enzymes de la coiffe auprès de la machinerie transcriptionnelle. Cette étape est gouvernée chez les mammifères, par le facteur transcriptionnel d'élongation DSIF[10] et le facteur négatif d'élongation (NELF[11]), qui induisent un arrêt de la transcription dans la région proximale du promoteur et participent au recrutement des enzymes de la coiffe. Cette interaction s'accompagne d'une activation d'une des enzymes de la coiffe par DSIF et en retour d'une dérépression de NELF (Yamaguchi *et al.*, 1999 ; Mandal *et al.*, 2004).

- l'élongation de la transcription (Fig. 15.3)

La relance de la transcription après la pause orchestrée par le complexe DSIF/NELF correspond au passage de la phase d'initiation à la phase d'élongation (pour revue, Peterlin and Price, 2006). Comme il a été déjà mentionné précédemment, cette étape est largement dépendante de l'activité kinase de P-TEFb. Ces cibles sont nombreuses, il y a :

[10] DRB Sensitivity-Inducing Factor
[11] Negative ELongation Factor

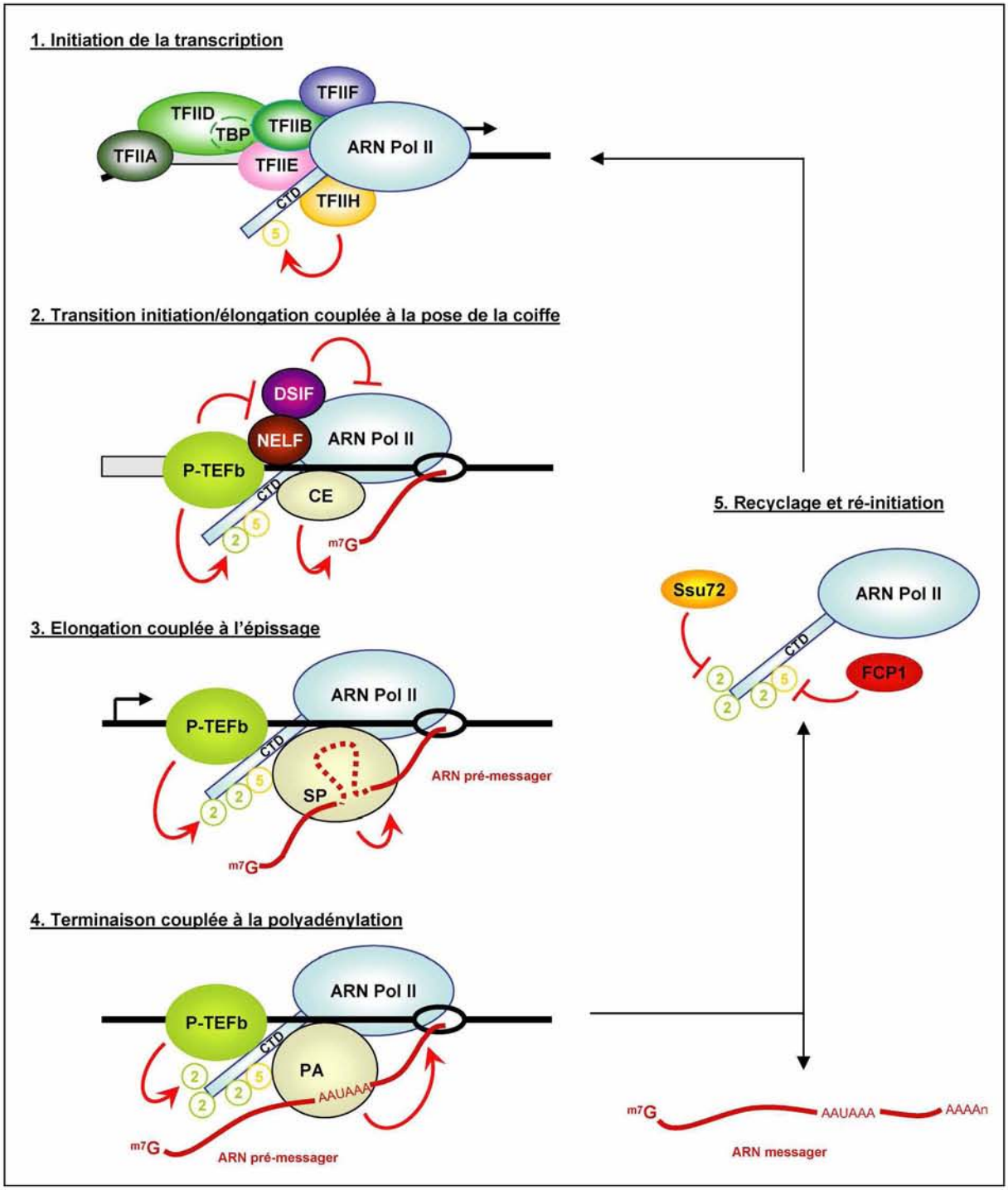


Fig. 15 : Etapes de la transcription par l'ARN polymérase II

5 étapes peuvent être décrites : (1) initiation de la transcription après phosphorylation des sérines 5 du CTD par TFIIH, (2) pause engendrée par la présence des facteurs NELF et DSIF, fixation de la coiffe m7G par les enzymes de la coiffe (CE), puis passage en élongation après phosphorylation des sérines 2 du CTD par P-TEFb, (3) élongation de la transcription couplée à l'activité d'épissage du spliceosome (SP), (4) terminaison de la transcription couplée au clivage de l'extrémité 3' et synthèse de la queue poly-A par la machinerie de polyadénylation (PA), et enfin (5) recyclage de l'ARN Pol II par déphosphorylation des sérines 2 et 5 du CTD par respectivement les facteurs Ssu 72 et FCP1.

- la sous unité Spt5 du facteur DSIF, qui une fois phosphorylée se convertit en un facteur activateur de la transcription (Yamada *et al.*, 2006),

- le domaine RD du facteur NELF qui se décroche de l'ARN une fois phosphorylé, ceci permettant la lecture de la matrice (Fujinaga *et al.*, 2004),

- et aussi le CTD au niveau des Ser 2. Ce phénomène d'hyperphosphorylation entraîne une augmentation de la taille et de la rigidité du CTD et s'accompagne du relargage de plusieurs protéines et facteurs généraux de transcription du PIC. Il apparaît que le CTD hyper-phosphorylé de l'ARN Pol II en phase d'élongation sert de plate-forme aux différents facteurs impliqués dans l'épissage tout d'abord, puis aux enzymes et facteurs de poly-adénylation finalement (pour revue : Hirose and Ohkuma, 2007 ; Peterlin and Price, 2006).

P-TEFb joue un rôle crucial dans la transcription et son activité est fortement régulée, comme nous le verrons plus loin.

- la terminaison (Fig. 15.4 et 5)

L'étape de terminaison suit la reconnaissance du signal de poly-adénylation. Deux réactions ont alors lieu : (1) un clivage endonucléolytique 10 à 30 nt en aval de ce signal (AAUAA), suivi de (2) l'ajout de la queue poly(A) à l'extrémité 3' produite. Ce processus requiert plusieurs facteurs protéiques, dont le facteur spécifique de clivage/poly-adénylation (CPSF), le facteur de stimulation du clivage (CstF), deux facteurs de clivage (CFI et II) et la poly(A) polymérase (PAP) chez les mammifères (Proudfoot *et al.*, 2002).

Plusieurs résultats suggèrent actuellement que le CTD, particulièrement lorsqu'il est phosphorylé à la position Ser 2, orchestre cette étape de terminaison en permettant la formation d'un complexe de poly-adénylation stable et actif (pour revue Hirose and Ohkuma, 2007).

Enfin, il apparaît que la phase de terminaison comporte aussi une étape de recyclage de l'ARN Pol II grâce à un système de boucle dans lequel la région terminale de transcription et en contact avec la région promotrice. Ainsi chez la levure, le facteur Ssu 72 participe à la formation de cette boucle et possède en plus une activité CTD-phosphatase ciblant la Ser 5 phosphorylée et permettant le retour de l'ARN Pol II à un état hypophosphorylé indispensable pour un nouveau cycle de transcription (Ansari and Hampsey, 2005).

1.4. Le complexe P-TEFb

P-TEFb est composé de la kinase cycline dépendante CDK9 et de son partenaire la cycline T1 (ou cT1) (Peng *et al.*, 1998a ; Wei *et al.*, 1998). CDK9 existe sous 2 formes, une de 42 kDa appelée $CDK9_{(42)}$, et l'autre, qui possède une extension de 117 résidus AA, de 55 kDa appelée $CDK9_{(55)}$ (Shore *et al.*, 2003). D'autres cyclines peuvent être associées à CDK9, il s'agit des cyclines T2a et b, et de la cycline K (Peng *et al.*, 1998b). Ici, nous ne nous intéresserons qu'au complexe P-TEFb contenant $CDK9_{(42)}$ et la cycline T1, les autres complexes étant mineurs et se comportant de manière identique dans la plupart des cas.

De récents résultats indiquent que P-TEFb est maintenu dans un équilibre fonctionnel par des régulateurs positifs ou négatifs. Dans les cellules HeLa, la moitié de P-TEFb est séquestré sous une forme catalytiquement inactive dans un complexe contenant l'ARN 7SK et la protéine Hexim1[12] (encore appelée MAQ1[13]) (Yik *et al.*, 2003 ; Michels *et al.*, 2003). Hexim1 se lie à P-TEFb en interagissant directement avec la région C-terminale de la cycline T1, mais l'effet inhibiteur de Hexim1 sur P-TEFb est 7SK dépendant. L'ARN possèderait en effet deux sites de fixation distincts et servirait de lien moléculaire entre Hexim1, qui est capable de se lier seul, et P-TEFb, qui ne se lie qu'en présence d'Hexim1, révélant l'aspect coopératif de l'interaction (Egloff *et al.*, 2006).

En plus du complexe P-TEFb/7SK/Hexim1, appelé 7SK snRNP, P-TEFb se retrouve aussi dans le complexe Brd4/P-TEFb (Jang *et al.*, 2005 ; Yang *et al.*, 2005). Le complexe Brd4/P-TEFb est transcriptionnellement actif. Sa présence dans les complexes transcriptionnels est probablement due à la capacité de Brd4, une protéine à bromo-domaine, à se lier aux histones acétylés.

Dans la cellule, la répartition de P-TEFb sous forme active ou inactive, est dynamique. Plusieurs agents de réponse aux stress sont connus pour dissocier la RNP 7SK et stimuler la formation du complexe Brd4/P-TEFb (Yang *et al.*, 2005). Enfin, des travaux réalisés sur différents types cellulaires suggèrent que le passage d'un état à l'autre des deux fractions de P-TEFb pourrait altérer le développement cellulaire en conduisant soit à une croissance non restreinte, soit à la différentiation terminale (Fig. 16 ; Giacinti *et al.*, 2006 ; De Falco *et al.*, 2005 ; De Falco *et al.*, 2008).

[12] Hexamethylene bisacetamide-inducible protein 1
[13] Ménage A Quatre 1

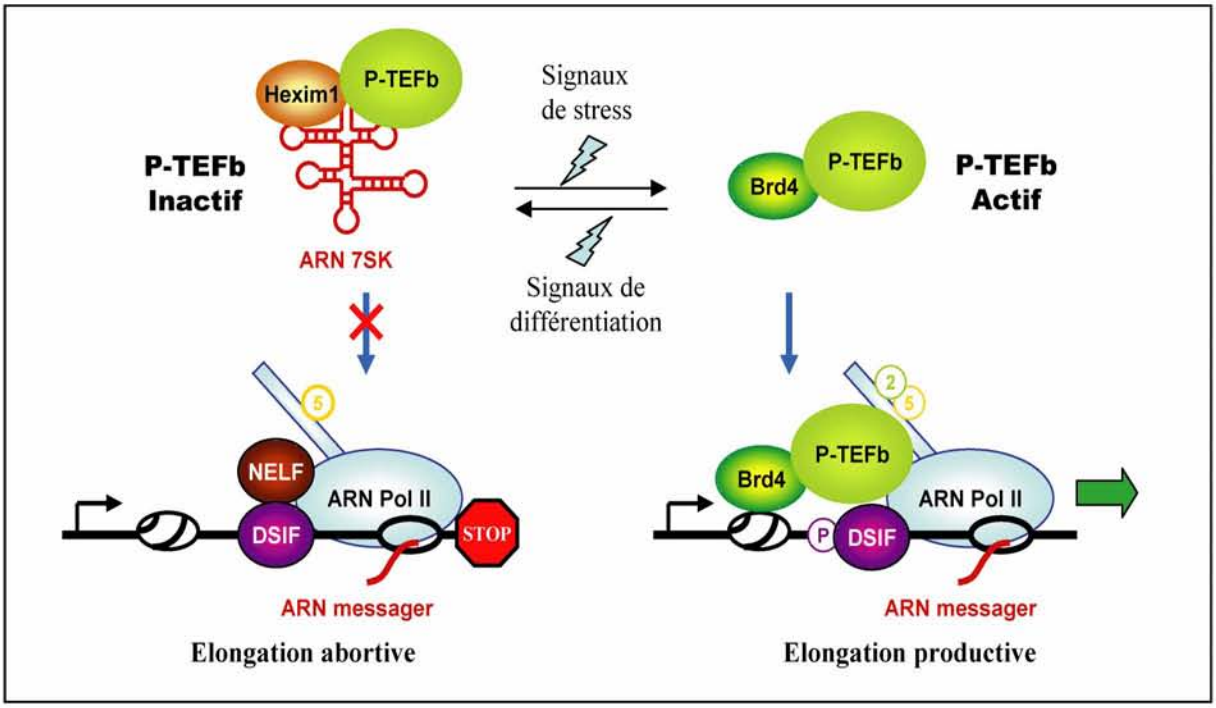


Fig. 16 : Equilibre entre les formes active et inactive de P-TEFb *(D'après Zhou and Yik, 2006).*

2. La transcription de l'ADNc proviral du virus HIV-1

L'initiation de la transcription virale est un processus très efficace. Outre les séquences nécessaires à l'initiation de la transcription (la boîte TATA par exemple), le promoteur du virus HIV-1 est sous le contrôle de plusieurs séquences régulatrices capables de fixer des facteurs activant la transcription. Ainsi le LTR comporte 4 sites de fixation de NF-κB, 4 sites de fixation de AP-1 et 4 sites de fixation de Sp1, qui assurent au virus une efficacité d'initiation de la transcription correcte dans les différents types cellulaires infectés (Fig. 17 ; pour revue, Pereira *et al.*, 2000).

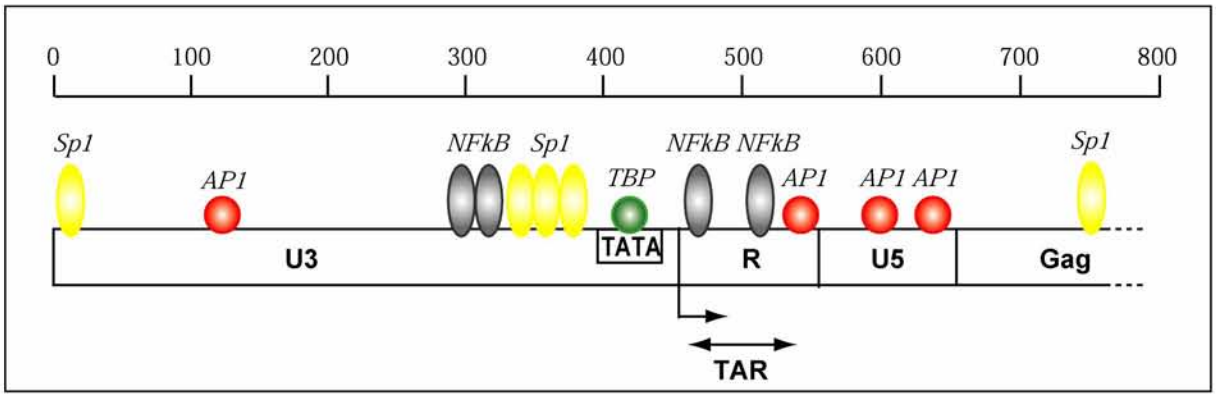


Fig. 17 : Structure du promoteur LTR du virus HIV-1

- en absence de Tat

En absence de Tat, la majorité des complexes de transcription formés ne contiennent pas le facteur p-TEFb. Ceci conduit donc à la synthèse de transcrits courts et non polyadénylés, impropres au développement du virus (Karn, 2000).

L'état de latence qui en découle, ne prend fin que lorsque le facteur P-TEFb est recruté au niveau du complexe de pré-initiation. Cet événement est permis dans un premier temps par les facteurs NF-κB et Sp1 qui assurent le recrutement du facteur P-TEFb au niveau des complexes de pré-initiation (Barboric *et al.*, 2001). Pour se faire, la relocalisation nucléaire du facteur NF-κB, séquestré normalement dans le cytoplasme par le facteur IκBα, a lieu lors de l'infection des lymphocytes T. Bien que de faible niveau dans ce cas, le recrutement du facteur P-TEFb permettrait la synthèse d'un nombre limité de transcrits de pleine longueur et polyadénylés conduisant ainsi à la production des premières protéines virales et de la protéine Tat en particulier (Karn, 1999).

- en présence de Tat

La protéine Tat joue un rôle majeur dans la transcription de l'ARN de HIV-1, et ceci en agissant à plusieurs niveaux.

L'initiation :

En interagissant avec de nombreux facteurs nucléaires possédant des activités acétyl transférase dirigées contre les protéines histones ou des facteurs nucléaires (ex: CREB, CBP, PCAF…), la protéine Tat favorise le remodelage de la chromatine au niveau du LTR. Par ailleurs, Tat semble aussi améliorer la formation du complexe d'initiation de la transcription en interagissant avec des facteurs tels que TBP, TFIIB, TFIIH ou l'ARN Pol II (pour revue, Marcello *et al.*, 2004 ; Huigen *et al.*, 2004).

L'élongation :

Enfin, et c'est son action primordiale, la protéine Tat est à l'origine de l'activation de l'élongation de la transcription, notamment en permettant le recrutement efficace de P-TEFb. Le taux de production de transcrits de pleine longueur est alors fortement augmenté. Ce phénomène est appelé transactivation de la transcription.

3. La transactivation

Lors de la transactivation, la protéine Tat se lie spécifiquement à la cycline T1 et à l'élément TAR localisé à l'extrémité 5' de tous les transcrits viraux naissants (Fig. 18 ; Karn, 1999 ; Taube *et al.*, 1999).

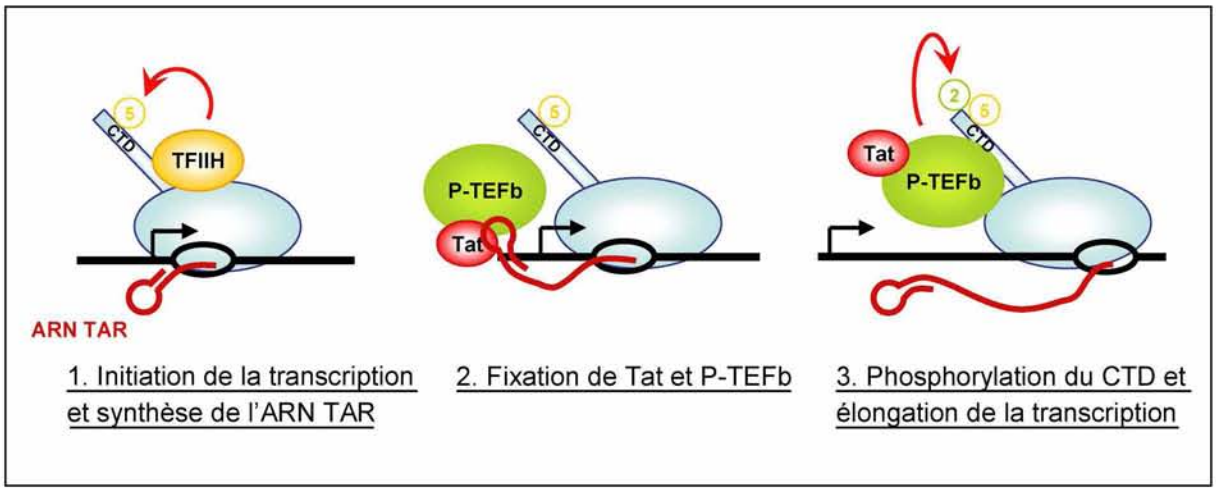


Fig. 18 : Etapes de la transactivation de la transcription par la protéine Tat

La présence de l'ARN TAR sur tous les transcrits HIV-1 produits lors de l'initiation de la transcription (étape 1) permet la fixation coopérative de la protéine Tat associée à P-TEFb (étape 2) et l'hyperphosphorylation du CTD de l'ARN Pol II permettent le passage en phase d'élongation (étape 3).

La forte interaction coopérative entre Tat, TAR et la cycline T1 permet le recrutement P-TEFb à proximité de l'ARN Pol II lors de la pause en phase d'initiation. Comme décrit précédemment, s'en suit alors l'hyperphosphorylation du CTD de la grande sous unité de l'ARN Pol II, mais aussi la phosphorylation des composants de N-TEF qui permet sa conversion en un facteur positif de transcription. Ceci contribue au recrutement des machineries d'épissage et de polyadénylation au niveau du CTD, et donne lieu à un complexe de transcription efficace pour l'élongation. L'interaction de TAR/Tat/cT1 est ensuite rompue par l'acétylation du résidu lysine K_{50} de Tat, ceci permettant la dissociation du complexe et le transfert de Tat sur le complexe d'élongation de l'ARN Pol II (Kaehlcke *et al.*, 2003). Au final, la présence de Tat multiplie 300 à 400 fois le taux de transcription.

Récemment, il a aussi été démontré que Tat dépasse ce rôle de cofacteur en accroissant le taux de P-TEFb libre et actif en transcription dans la cellule. Son interaction avec la cycline T1, d'autant plus forte que TAR est présent, empêche la fixation de Hexim1. Ceci empêche la

formation de la RNP 7SK inactive en transcription, et mieux encore entraîne sa dissociation (Fig. 19 ; Barboric *et al.*, 2007).

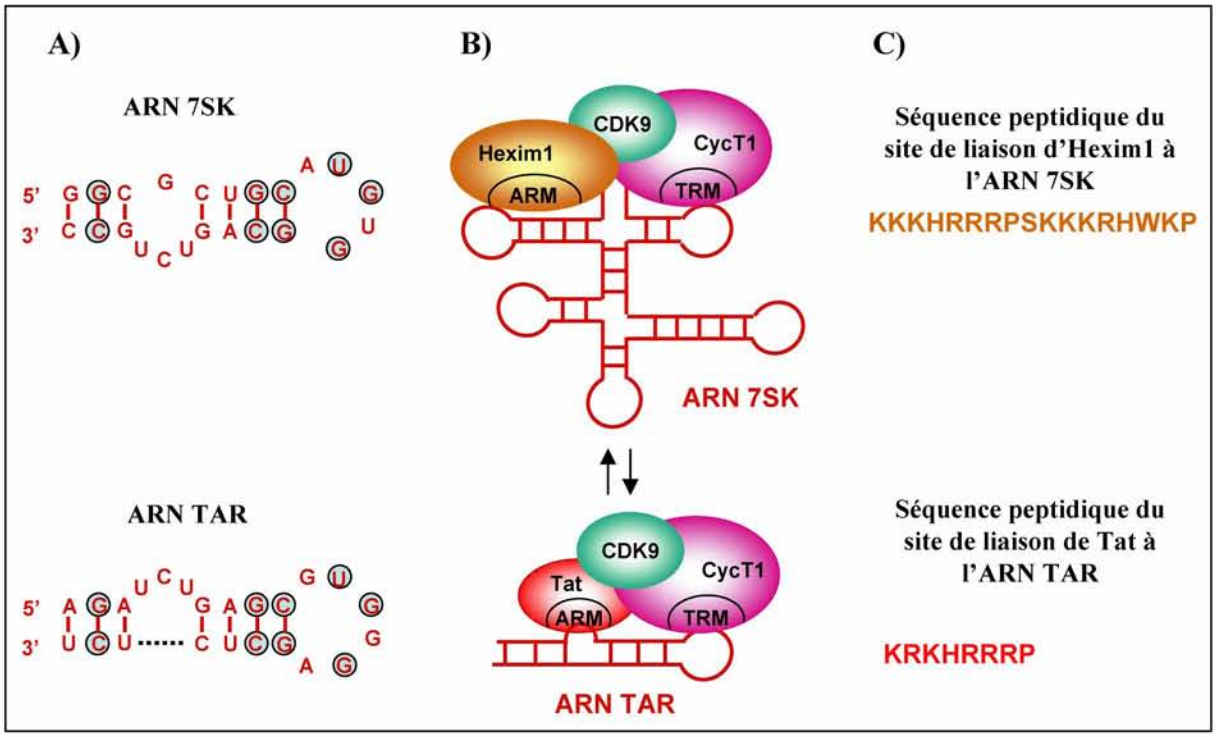


Fig. 19 : Homologie structurale entre les ARN 7SK et TAR, et les motifs les reconnaissant

A) Comparaison des séquences et structures secondaires des tiges-boucles fixant P-TEFb des ARN 7SK et TAR. Les nucléotides conservés sont entourés en bleu. B) Représentation schématique des complexes TAR/Tat/P-TEFb et 7SK/HEXIM1/P-TEFb. Tat et HEXIM1 interagissent directement avec P-TEFb via leur liaison à une région de la cycline T1. Par simplicité, un seul complexe HEXIM1/P-TEFb est représenté alors que deux se fixent sur une molécule ARN 7SK. (TRM: TAR Recognition Motif)C) Les domaines de fixation riches en arginine (ARM) de Tat et HEXIM1 à leurs ARN respectifs sont des motifs fortement homologues et fonctionnellement interchangeables.

4. Etat des connaissances structurales du complexe TAR/Tat/cycline T1

La protéine Tat

Tat est synthétisée dans les deux phases (précoce et tardive) du cycle viral de réplication. Sa présence est essentielle pour la multiplication du virus. Néanmoins, elle n'est pas incorporée dans le virion (Dayton *et al.*, 1986 ; Fisher *et al.*, 1986). Les différents ARNm codant pour Tat conduisent après traduction a deux formes de la protéine : l'une de 72 acides aminés est juste codée à partir du premier exon, alors que l'autre forme possède un domaine C-terminal supplémentaire codé par le second exon (Jones and Peterlin, 1994). Par ailleurs, des

différences de positionnement du codon stop présent dans le second exon, font que la taille de la protéine Tat codées par deux exons varie entre 86 et 101 acides aminés en fonction des isolats et souches. Cette forme codée par deux exons est produite en phase précoce lorsque les transcrits sont épissés entre les sites D4 et A7, alors que Tat 72 n'est synthétisée qu'en phase tardive lorsque les transcrits produits possèdent l'intron D4-A7 (Malim *et al.*, 1988). Les deux formes sont actives en transactivation, mais leurs effets au niveau de l'infection sont différents.

Cinq domaines fonctionnels ont été identifiés dans la protéine Tat 72 (Fig. 20 ; Kuppuswamy *et al.*, 1989 ; Jeang *et al.*, 1999) :

- Le segment N-terminal (1 à 21), est riche en proline et acides aminés acides. Des mutations ponctuelles dans cette région n'altèrent pas la fonction de la protéine.
- Le deuxième segment (22 à 37), est caractérisé par 7 cystéines fortement conservées et impliquées dans la formation de ponts disulfures intra ou intermoléculaires. La mutation individuelle de 6 cystéines sur 7 abolit la fonction de la protéine Tat.
- Le segment « core » (38 à 47), contient le motif RKGLGI très conservé dans toutes les souches de HIV et de SIV.

Ces 3 segments de Tat constituent son domaine d'activation qui est capable d'interagir avec plusieurs protéines cellulaires (Sp1, cycT1, HAT, PKR).

- La région basique (48 à 57), contient le signal de localisation nucléaire RKKRRQRRR permettant à la fois l'import nucléaire de Tat par interaction avec le système des importines, mais aussi l'internalisation de Tat dans des cellules non infectées.
- le segment riche en glutamine (58 à 72), forme avec la région basique le domaine de liaison à l'ARN (TAR en particulier).

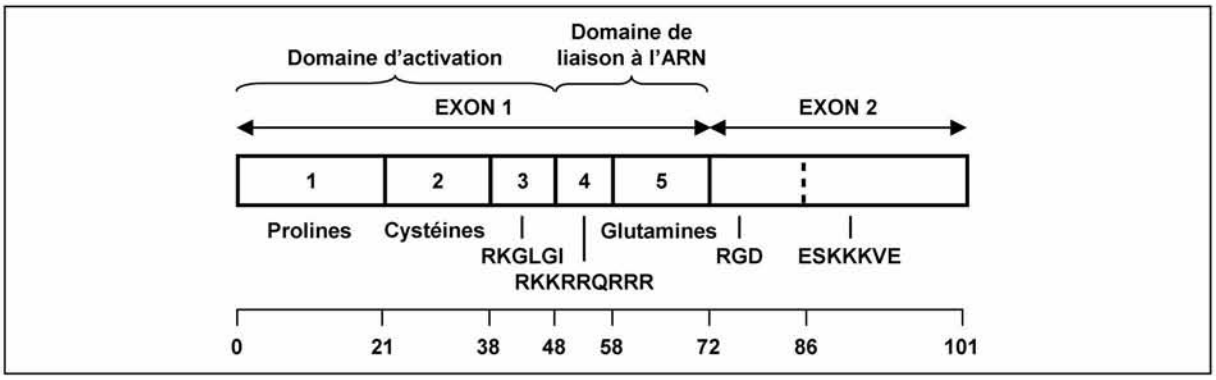


Fig. 20 : Architectures en domaines de la protéine Tat *(d'après Huigen et al., 2004)*

Le second exon de Tat est moins étudié, mais deux motifs ont été identifiés : (1) le motif ESKKKVE hautement conservé, et (2) la séquence RGD, non conservée dans les souches de HIV-2 et de SIV, mais qui pourrait être impliquée dans la liaison de Tat aux intégrines de surface des cellules (Avraham *et al.*, 2004). 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).

L'interaction TAR/Tat

La protéine Tat se lie donc à l'élément TAR, localisé à l'extrémité 5' de tous les transcrits viraux (Muesing *et al.*, 1987). L'élément minimum nécessaire à cette fixation correspond au nucléotide 19 à 44 (Jakobovits *et al.*, 1988). Il comporte une boucle latérale de 3 résidus pyrimidiques (UCU), une boucle terminale de 6 nt (CUGGGA) et 2 hélices standard de forme A (Fig. 21A ; Aboul-ela *et al.*, 1995 ; Al-Hashimi *et al.*, 2002).

L'interaction TAR/Tat est à la fois séquence et structure spécifique. La liaison se fait au niveau de la boucle latérale, qui doit contenir un résidu U à sa base (position 23) et être composée d'au moins 2 nucléotides pour que le U23 soit protubérant. Les deux paires de bases directement adjacentes G26-C39 et A27-U38, ainsi que les groupes phosphate P21 (entre les résidus G21 et A22), P22 et P40 sont aussi fortement impliqués dans l'interaction TAR/Tat (Calnan *et al.*, 1991 ; Weeks and Crothers, 1991 ; Churcher *et al.*, 1993).

L'étude par RMN de TAR seul ou en interaction avec une molécule d'argininamide, puis celle de TAR avec le domaine de liaison à l'ARN de Tat plus tard, ont permis de localiser le site spécifique de liaison de Tat au niveau du sillon majeur à proximité du résidu G26 et de mettre en évidence la formation d'un triplet de base (U23-A27-U38) lors de la liaison (Fig. 21B; Puglisi *et al.*, 1992 ; Long and Crothers, 1999).

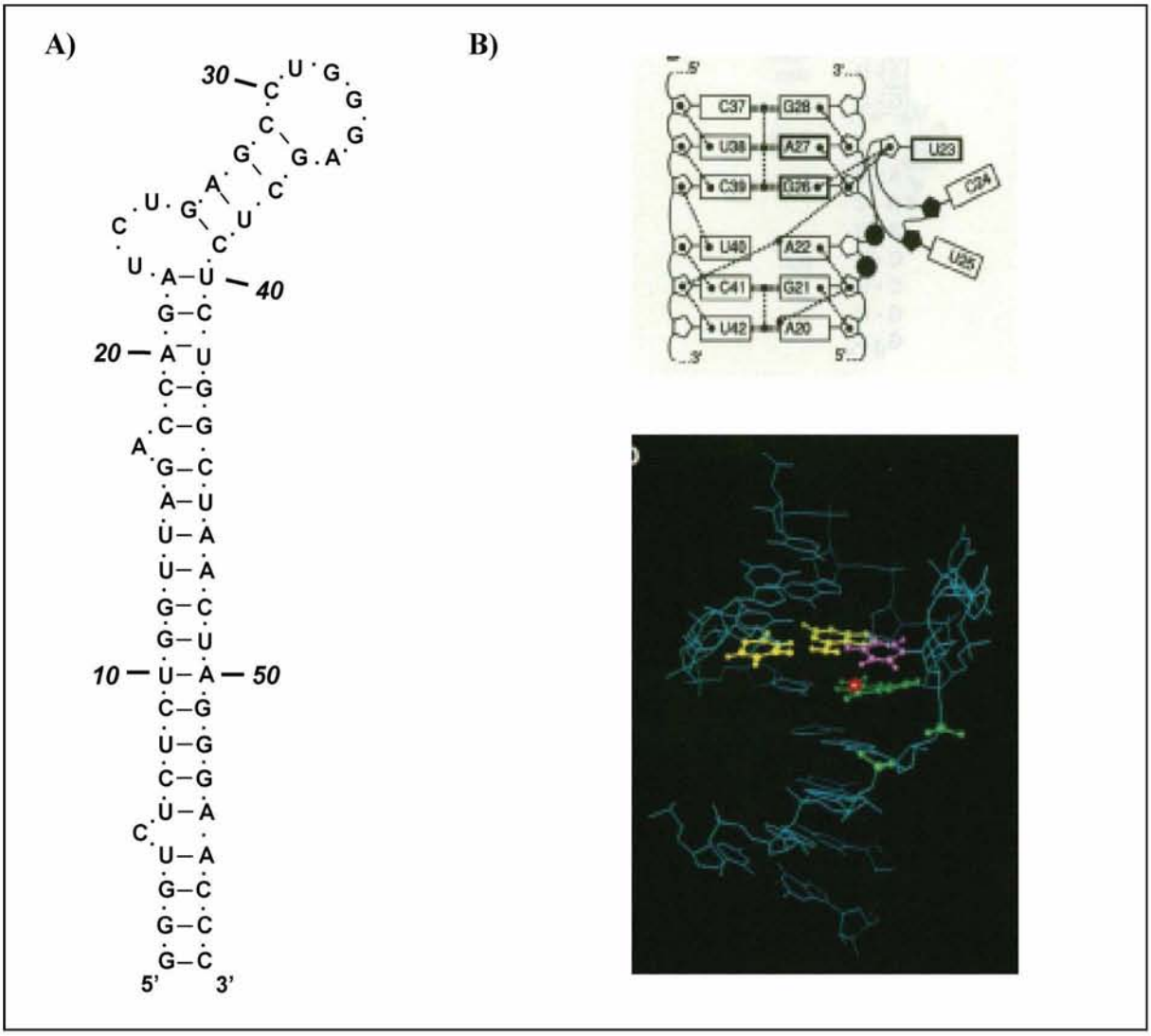


Fig. 21 : L'interaction TAR/Tat

A) Séquence et structure secondaire de l'ARN TAR de la souche NL4-3.

B) Structure RMN de l'ARN TAR en complexe avec l'arginamide (Puglisi et al., 1992).

La cycline T1

La cycline T1 a été initialement identifiée comme ressemblant à la cycline C et interagissant spécifiquement avec Tat (Wei *et al.*, 1998). Dans le même temps, les différentes formes de cycline T (T1, T2a, T2b) ont été mises en évidence, toutes étant des partenaires de la kinase CDK9 et formant un complexe p-TEFb (Peng *et al.*, 1998b). Les cyclines T2a et T2b sont des variants d'épissage. Contrairement à la cycline T1, elles ne sont pas exprimées constitutivement et ne participent pas au phénomène de transactivation (Wimmer *et al.*, 1999 ; Garriga *et al.*, 2003).

La région N-terminale (1 à 272) de la cycline T1 humaine contient 2 domaines conservés dans les cyclines et un motif de reconnaissance de TAR/Tat (TRM[14]) (Fig. 22A ; Peng *et al.*, 1998b). La fixation de CDK9, au niveau du premier domaine cycline conservé, est totalement indépendante des interactions avec les complexes TAR/Tat ou même de 7SK/Hexim qui se fixe dans la même région (Garber *et al.*, 1998 ; Schulte *et al.*, 2005). La région C-terminale des cyclines est beaucoup moins conservée, elle contient un motif de type coiled-coil (370 à 430), une région riche en His (506-530) et une séquence PEST (709-726) (Wei *et al.*, 1998). A part le fait que la région riche en His aide à la reconnaissance du CTD de l'ARN Pol II, il existe peu d'informations sur les propriétés de cette région C-terminale (Lin *et al.*, 2002).

L'interaction TAR/Tat/cT1

La cycline T1 est le troisième partenaire du complexe TAR/Tat. A la différence de Tat qui est capable de se lier seule à l'élément TAR, ni la cycline T1, ni le complexe P-TEFb, ne se lient seuls à l'ARN (Garber *et al.*, 1998). La présence de Tat est donc indispensable, mais il est aussi clairement établi que la formation du complexe ternaire TAR/Tat/cycline T1 est coopérative (Zhang *et al.*, 2000). Seul les 272 premiers AA de la cT1 sont nécessaires pour fixer TAR/Tat *in vitro* et permettre le phénomène de transactivation par Tat in vivo (Wei *et al.*, 1998). L'interaction requiert tout particulièrement la cystéine critique 261 de la cycline T1 humaine pour qu'il y ait formation d'un doigt à Zn intermoléculaire avec les résidus cystéine de Tat (Fig. 22B ; Garber *et al.*, 1998). Cinq autres résidus (R251, L252, R254, I255 et W258) du TRM[15], ainsi que la présence des résidus basiques de la région C-terminale des domaines conservés chez les cyclines, participent à la reconnaissance du complexe TAR/Tat (Garber *et al.*, 1998).

Même si la cycline T1 ne se fixe pas seule à l'élément TAR, elle établit plusieurs contacts avec l'ARN dans le complexe ternaire (Wei *et al.*, 1998). Cette interaction repose sur des critères de séquence (G32 et G34 essentiels) et de structure de l'ARN (importance de la paire de base C30-G34 et du nucléotide en position 35 qui agit comme « spacer ») (Richter *et al.*, 2002).

[14] Tat-TAR Recognition Motif
[15] TAR Recognition Motif

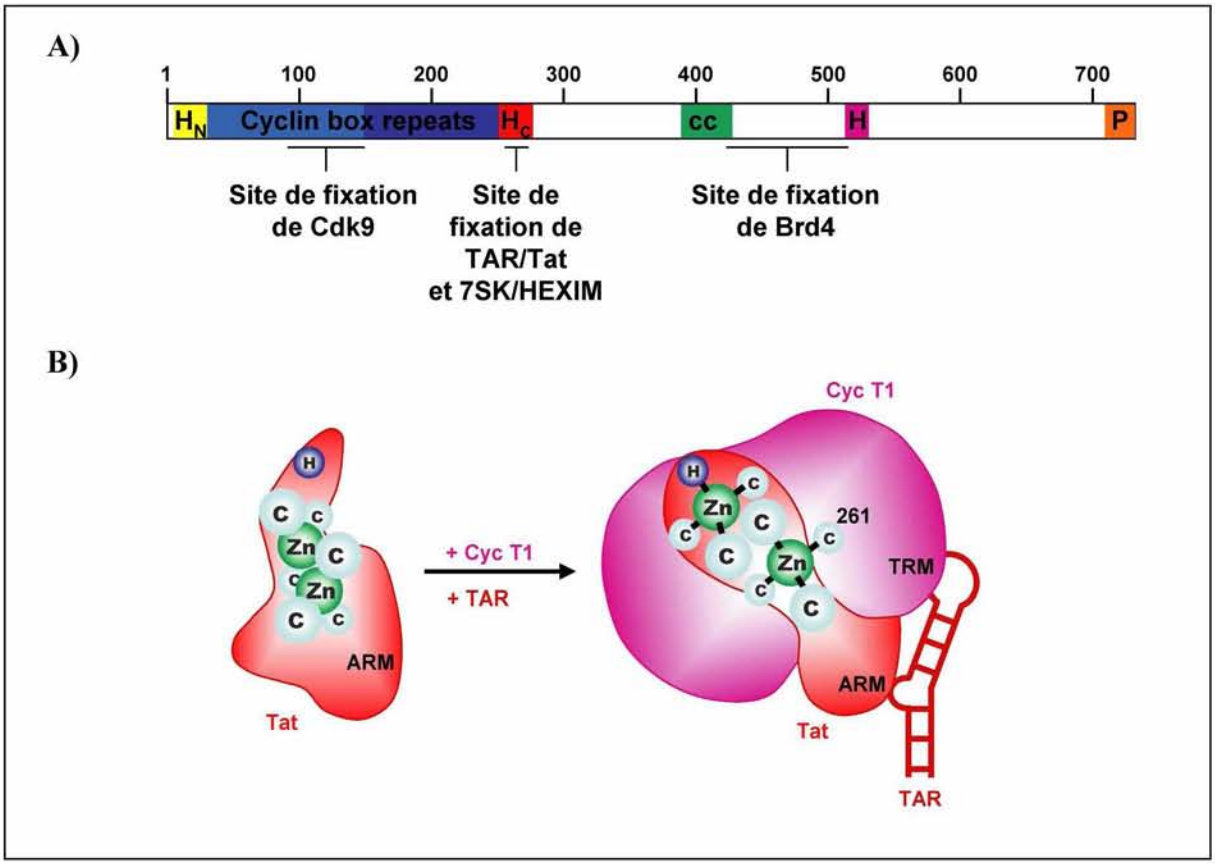


Fig. 22 : L'interaction TAR/Tat/cycline T1 *(D'après Garber et al., 1998)*

A) Domaines structuraux de la cycline T1. Le domaine N-terminal contient les deux domaines cyclines canoniques (cyclin box) et les hélices N et C terminales (HN et HC). La région centrale est beaucoup moins conservée mais contient un segment hélice (cc, coiled coil)) et une région riche en histidine fortement conservée (H). Les 20 derniers résidus en C-terminal sont riches en proline. Les sites de fixation proposés des différents partenaires de la cycline T1 sont présentés en dessous. Le site de reconnaissance de TAR/Tat a été délimité aux résidus 254 à 263.

B) Interaction du complexe TAR/Tat/cycline T1. Deux types d'interactions s'établissent dans la liaison Tat/cycline T1 : (1) une liaison entre les zones hydrophobes des deux protéines et (2) une liaison métallique via l'un des deux ions Zn^{2+}. Cette dernière liaison fait de la cystéine 261 un résidu indispensable à la transactivation.

CHAPITRE 3 : Description générale de l'épissage des ARN

L'idée de gènes morcelés est apparue pour la première fois en 1977, à l'occasion d'une étude du génome de l'adénovirus (Berget *et al.*, 1977). Dès lors, la présence de séquences non codantes insérées dans les gènes s'est avérée être générale pour tous les eucaryotes. Ces séquences non codantes sont nommées introns[1], et il en existe plusieurs types selon le type de gènes et l'organisme. Ils ont été classés en 4 catégories en fonction des mécanismes d'excision mis en œuvre :

- les introns des ARN de transfert, dont l'élimination est catalysée par des enzymes (pour revue, Wolin and Matera, 1999),
- les introns autocatalytiques de groupe I, comprenant notamment les introns des ARN ribosomiques et des introns de gènes de protéines d'organelles (chloroplastes et mitochondries) (pour revue, Doudna and Cech, 2002),
- les introns autocatalytiques de groupes II et III (pour revues, Bonen and Vogel, 2001; Doudna and Cech, 2002) comprenant, pour le type II, des introns de gènes d'organelles et pour le type III, essentiellement des introns de gènes de protéines chloroplastiques de certains protistes (euglènes),
- et enfin les introns des gènes nucléaires codant les protéines qui sont éliminés par un macrocomplexe ribonucléoprotéique appelé spliceosome, et auxquels nous allons nous intéressés exclusivement, puisqu'ils ont fait l'objet de mon travail.

En dépit de ces différences de mécanisme, tous les processus qui conduisent à l'élimination des introns sont rassemblés sous le terme d'épissage.

Chez les eucaryotes, la plupart des gènes nucléaires codant les protéines comportent des introns. Le nombre et la taille de ces introns sont très variables (1 à plus de 50 par gène) et ils peuvent mesurer jusqu'à 500 000 nt (avec une moyenne de 3500 nt). A l'inverse, la taille des séquences codantes, appelées exons[2], est beaucoup plus petite : de 10 à 400 nt (avec une moyenne de 140 nt). L'épissage a lieu dans le noyau et transforme les ARN pré-messagers (pré-ARNm) issu de la transcription, en ARN messagers (ARNm) matures ne contenant que les exons joints les uns aux autres. Les ARNm sont ensuite transportés dans le cytoplasme où ils sont traduits en protéines.

[1] intragenic regions
[2] expressed regions

1. La réaction d'épissage

1.1. La réaction chimique d'épissage

Les introns des ARN pré-messagers sont éliminés par 2 étapes successives de transestérification (Fig. 23). La première étape consiste en l'attaque nucléophile du phosphate en 5' de l'intron par l'hydroxyle en 2' d'un résidu adénosine situé dans une séquence de l'intron, appelée boîte de branchement. Deux intermédiaires réactionnels sont alors produits : l'exon 1, avec une extrémité 3' hydroxyle libre, et un intermédiaire en lasso contenant l'intron, dont la guanosine 5' terminale est covalemment liée au point de branchement par une liaison 2'-5' phosphodiester, et l'exon 2.

Dans la seconde étape, l'attaque nucléophile de l'hydroxyle en 3' de l'exon 1 sur le phosphate en 3' de l'intron conduit à la ligation des deux exons et à la libération de l'intron sous forme de lasso. Ce sont les produits finaux de la réaction d'épissage (Jurica and Moore, 2002).

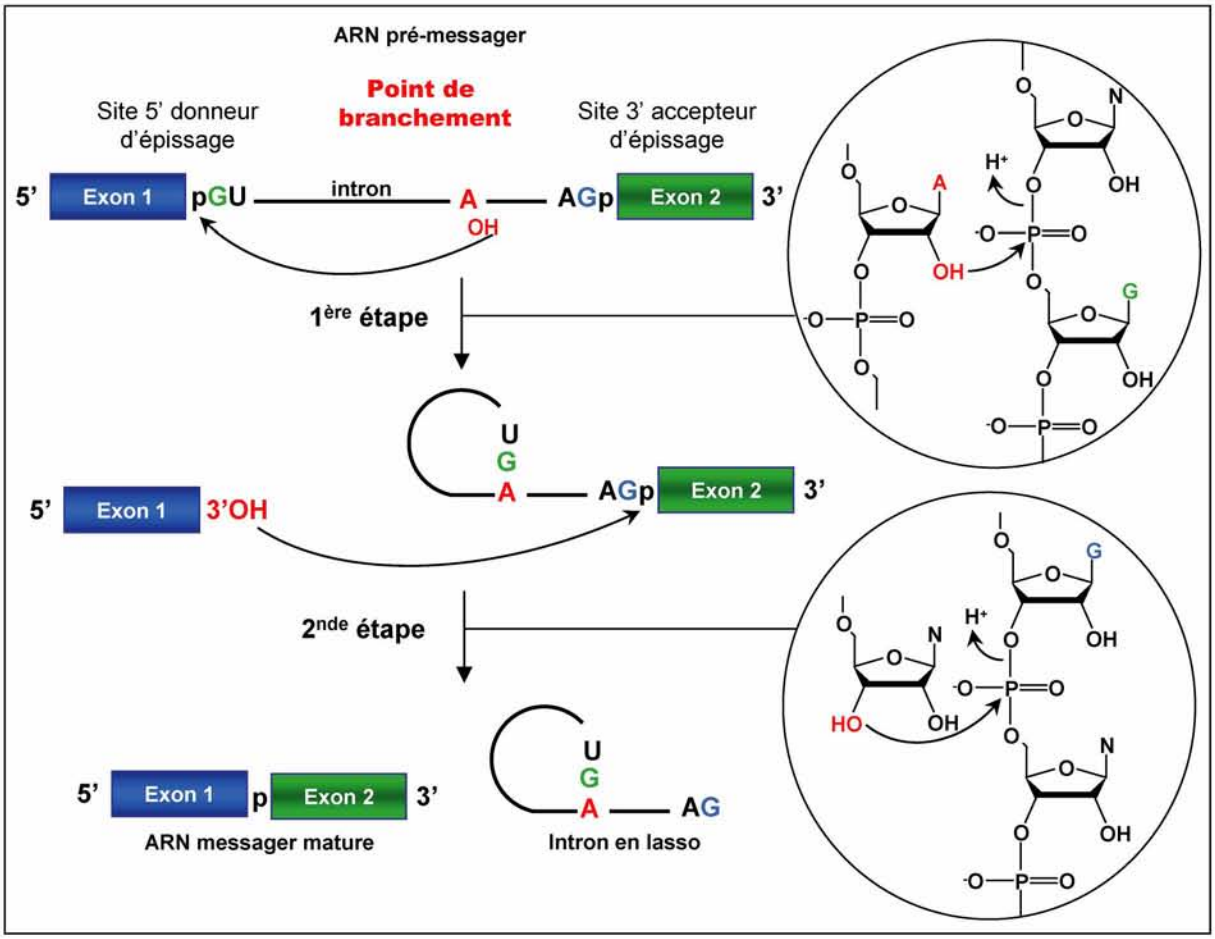


Fig. 23 : Les étapes de la réaction d'épissage des introns des ARN pré-messagers nucléaires

Chaque étape de transestérification est détaillée à droite du schéma dans les cercles. (D'après Jurica and Moore, 2002).

1.2. Les séquences définissant les introns

Les introns renferment un ensemble de séquences conservées permettant leur reconnaissance par la machinerie d'épissage (Fig. 24). 2 types d'introns existent chez les eucaryotes supérieurs : les introns majeurs, éliminés par le spliceosome U2-dépendant ou majeur, et les introns dits mineurs, éliminés par le spliceosome U12-dépendant ou mineur. Les introns mineurs étant beaucoup moins fréquents dans la cellule, mais aussi parce qu'ils sont absents de l'ARN du virus HIV-1, nous ne présenterons dans cette introduction que l'épissage des introns par le spliceosome majeur.

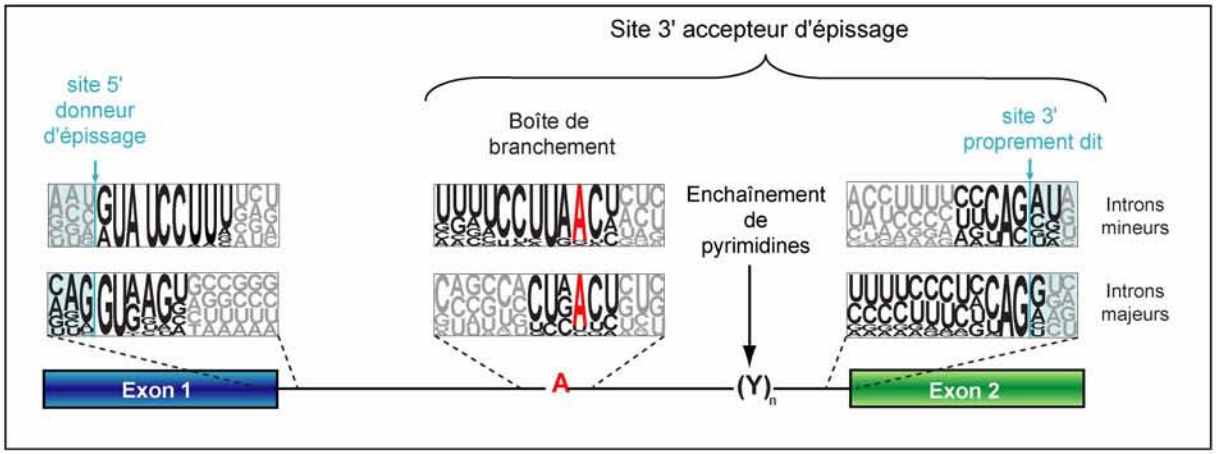


Fig. 24 : Séquences essentielles à la réaction d'épissage dans les introns des ARN pré-messagers de vertébrés.

La hauteur des nucléotides à chaque position représente la fréquence de la présence de ce nucléotide à cette position. Les nucléotides qui semblent impliqués pour la reconnaissance de l'intron sont en noir, les autres en gris. Le nucléotide de branchement est représenté en rouge (d'après Patel and Steitz, 2003).

La borne 5'

La comparaison des bordures exon 5'/intron des introns majeurs chez les vertébrés a permis d'établir la séquence consensus du site 5' donneur d'épissage : AG/**GU**RAGU (R = purine et / = jonction exon/intron), dans laquelle le dinucléotide **GU** en 5' de l'intron est le motif le plus conservé (Moore and Sharp, 1993).

La borne 3'

La définition de la borne 3' est plus complexe. En effet, chez les vertébrés, la borne 3' renferme la boîte de branchement, une séquence polypyrimidine et le site 3' accepteur où a lieu l'épissage.

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 : YNYUR**A**C (Y = pyrimidine, le résidu adénosine conservé correspondant au point de branchement est en gras).
La séquence polypyrimidine correspond à un enchaînement de 10 à 20 résidus pyrimidines, les résidus U étant plus favorables à sa fonction.
Enfin, le site 3' accepteur d'épissage, situé à la bordure intron/exon 3', est caractérisé par une courte séquence conservée YAG/G.

1.3. Les composants du spliceosome

Le spliceosome est un macrocomplexe ribonucléoprotéique architecturé par assemblage de particules ribonucléoprotéiques appelées UsnRNP. Chacune de ces UsnRNP s'architecture elle-même autour d'un ou plusieurs petits ARN nucléaires riches en uridine (ou UsnRNA[3]), et comprend un ensemble de 7 protéines Sm[4] ou Sm-like (Lsm), et plusieurs protéines spécifiques

1.3.1. Biogenèse des UsnRNP chez les mammifères

1.3.1.1. Cas des snRNP U1, U2, U4 et U5

A l'exception du snRNA U6, tous les UsnRNA sont transcrits par l'ARN polymérase II dans le noyau. Ils possèdent une coiffe monométhylée (m7G) en 5' et des nucléotides additionnels à leur extrémité 3'. Une fois exportés dans le cytoplasme, le complexe SMN[5] permet à chaque UsnRNA de s'associer avec 7 protéines Sm pour former une pré-UsnRNP. Cette étape s'accompagne pour chaque UsnRNA, de la conversion de la coiffe monométhylée m7G en une coiffe triméthylée m2,2,7G, et dans le même temps, d'une maturation de l'extrémité 3' (pour revue, Paushkin *et al.*, 2002). A ce niveau, la coiffe et le domaine Sm de chacun des UsnRNA forment alors un signal de localisation nucléaire bipartite spécifiquement requis pour l'interaction avec la snurportine, qui est responsable du passage à travers le pore nucléaire (Plessel *et al.*, 1994). Les pré-UsnRNP s'accumulent ensuite dans les corps de Cajal (ou Cajal bodies), où les snRNA subissent quelques modifications post-transcriptionnelles importantes pour l'assemblage et le fonctionnement du spliceosome (Fig. 25 ; Massenet *et al.*, 1998 ; Jady and Kiss, 2001). La plupart de ces modifications, principalement des

[3] Uridine-rich small nuclear RiboNucleic Acids
[4] Seminoprotein
[5] Survival of Moto Neuron

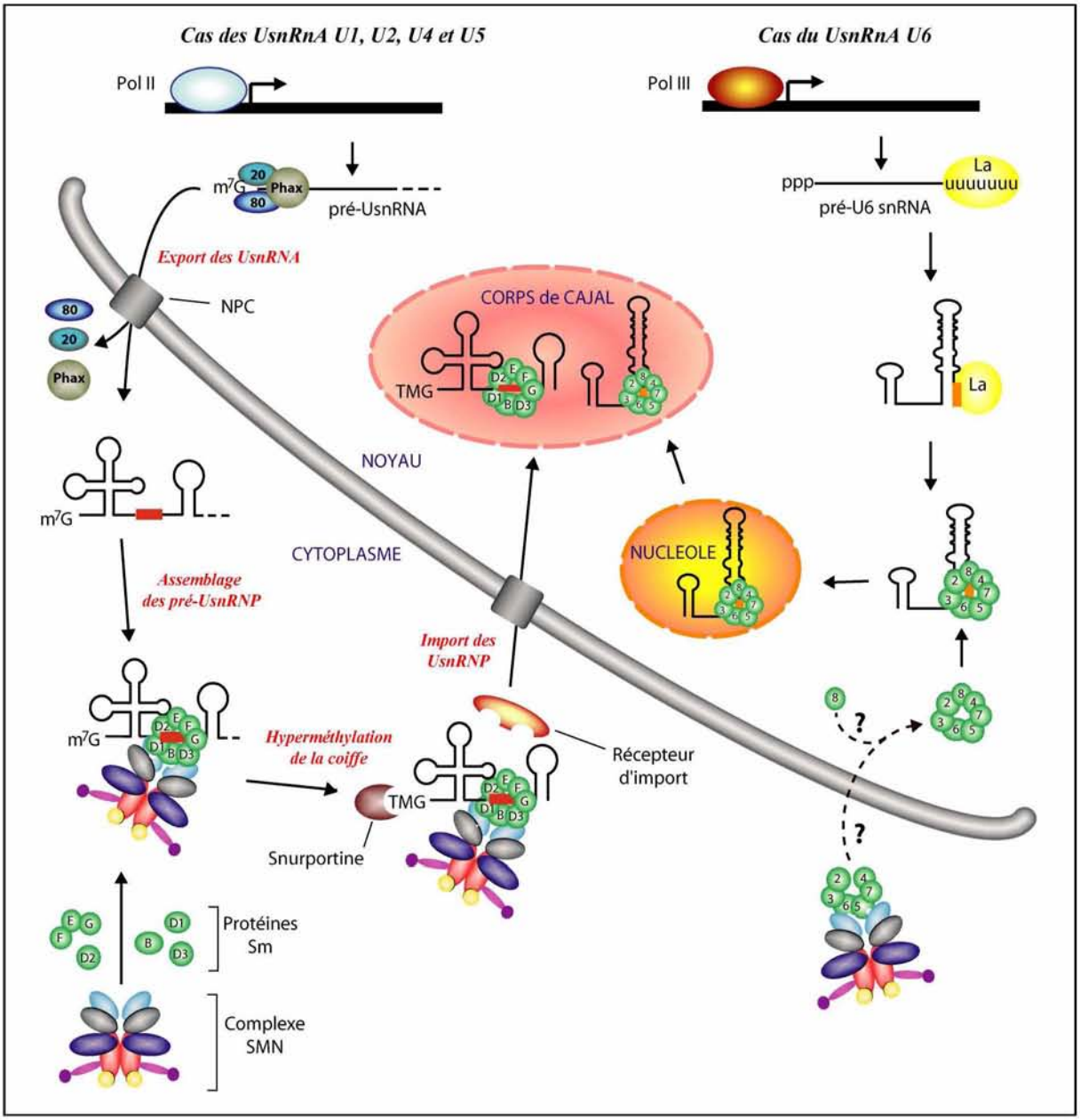


Fig. 25 : Représentation schématique de la biogenèse des UsnRNP chez les eucaryotes supérieurs

Les snRNA U1, U2, U4 et U5 sont transcrits par l'ARN Polymérase II. Ils portent une coiffe incluant un m7G en 5' et des nucléotides supplémentaires en 3' (en pointillé). La coiffe est reconnue par le complexe de liaison à la coiffe composé de CBP 20 et 80 (en bleu). La protéine Phax se lie au complexe et aux snRNA, pour assurer leur passage à travers le pore nucléaire (NPC : "Nuclear pore complex"). Le complexe SMN reconnait ensuite les UsnRNA et les protéines Sm associées, ceci permettant une hyper méthylation de la coiffe et une maturation de l'extrémité 3'. La coiffe et les protéines Sm sont ensuite reconnues par la snurportine, qui interagit avec le récepteur d'import (importine b) pour être importé dans le noyau. Les UsnRNP sont finalement stockées dans les corps de Cajal. Le snRNA U6 est transcrit par l'ARN polymérase III. Il contient à son extrémité 3' une séquence

riche en U, nécessaire à l'interaction avec la protéine La et porte un gamma-méthylphosphate en 5'. La protéine La est ensuite remplacée par les protéines Lsm. Cette association est indispensable au transport dans le nucléole des pré-snRNP U6 nouvellement formées. Elles y subissent des modifications post-transcriptionnelles et termineront leur maturation en transitant dans les corps de Cajal (adapté de Paushkin et al., 2002).

pseudouridylations et des 2'-O méthylations, sont obtenues grâce à des petits ARN guides (ou scaRNA[6]) (pour revue, Kiss, 2004). Les protéines spécifiques de chaque UsnRNP, concentrées dans les corps de Cajal, s'associent enfin pour former des particules snRNP matures qui vont soit être tout de suite assemblées en spliceosome, soit s'accumuler dans des structures de la région interchromatine appelées IGC[7] ou encore « speckles ». Ces structures IGC sont généralement localisées à proximité des sites actifs de transcription et ont plusieurs fonctions dont le stockage/relarguage de différents facteurs impliqués dans les processus de transcription et d'épissage.

1.3.1.2.Cas de la snRNP U6

Contrairement aux autres UsnRNA, le snRNA U6 est transcrit dans le noyau par l'ARN polymérase III (Kunkel *et al.*, 1986). Sa coiffe est γ-méthylée et il possède en 3' une séquence courte et riche en résidus uridines sur laquelle vient se fixer la protéine La[8] (Rinke and Steitz, 1985 ; Dahlberg and Lund, 1988). Ceci permet la stabilisation du UsnRNA et facilite l'assemblage de la UsnRNP (pour revue, Wolin and Cedervall, 2002). Après fixation d'un phosphate 2'-3' cyclique sur le ribose en 3' du snRNA U6 (Lund and Dahlberg, 1992), la protéine La est remplacée par les protéines Sm-like (LSm2 à 8) (Seraphin, 1995 ; Mayes *et al.*, 1999 ; Salgado-Garrido *et al.*, 1999). Contrairement aux autres UsnRNA, les modifications post-transcriptionnelles du snRNA U6 ne seraient pas réalisées dans les corps de Cajal mais dans le nucléole, grâce à des snoRNA[9] guides de modification. Le mode d'association des protéines LSm au snRNA U6, ou encore le rôle du complexe SMN dans la biogenèse de la snRNP U6 n'est pas encore bien connu.

[6] small cajal body RNA
[7] Interchromatin Granules Cluster
[8] Lupus antigen
[9] small nucleolar RNA

1.3.2. Rôles et particularités des UsnRNP

1.3.2.1. La snRNP U1 (12S)

Les composants de la particule snRNP U1 (12S) sont le snRNA U1, les 7 protéines Sm, ainsi que les protéines U1-70K, U1A et U1C (Fig. 26).

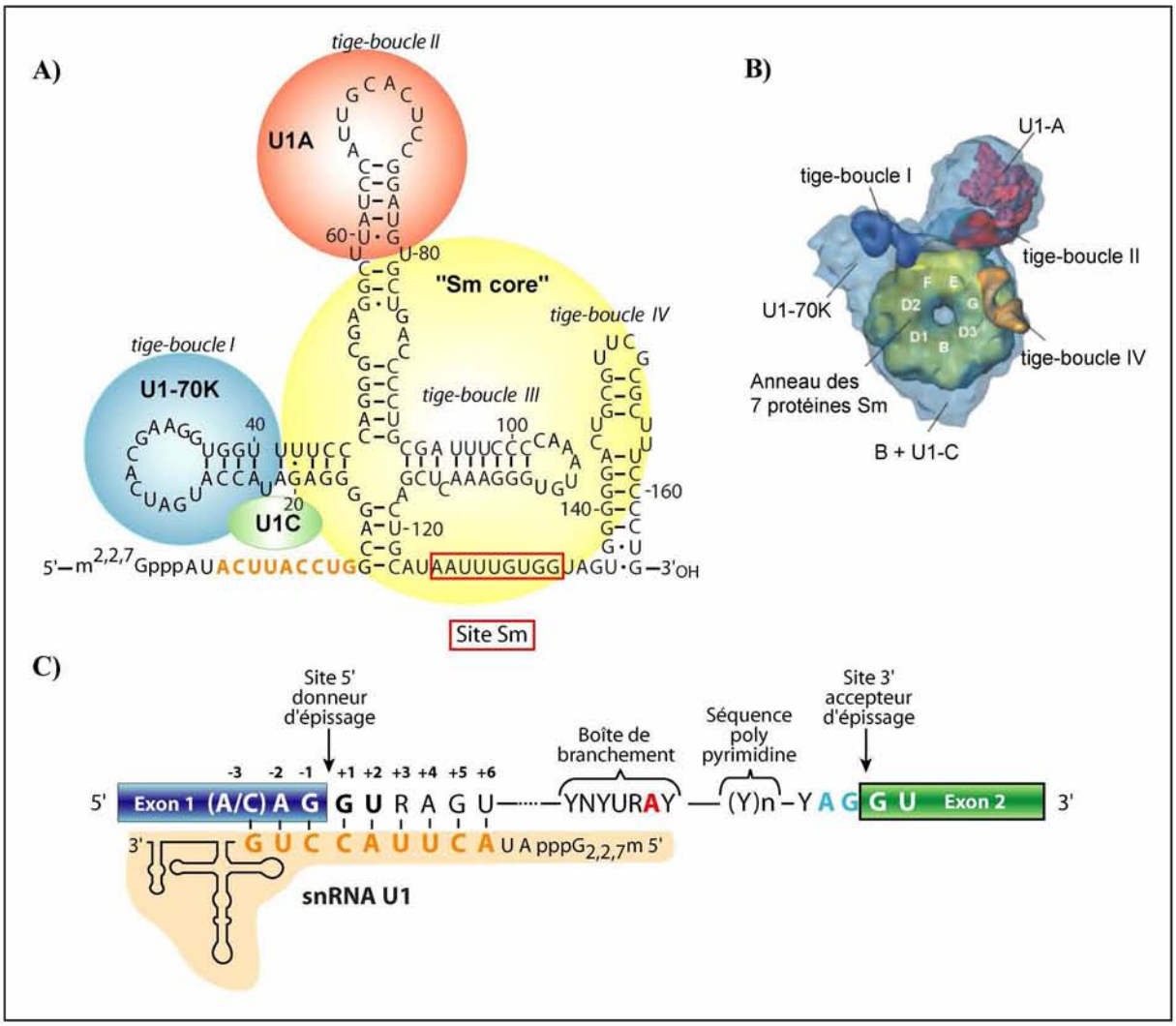


Fig. 26 : La particule snRNP U1

A) Structure secondaire du snRNA U1 et ses protéines associées. La séquence permettant la fixation des protéines Sm est encadrée en rouge et les nucléotides en orange sont nécessaires à l'appariement avec le site 5' d'épissage. B) Modèle de structure 3D proposé pour la snRNP U1 (Stark et al., 2001). C) Appariement entre la snRNP U1 et le site 5' donneur d'épissage d'un ARN pré-messager (Lerner et al., 1980). Les nucléotides N, R et Y désignent un nucléotide quelconque, une purine et une pyrimidine, respectivement (pour revue, Will and Luhrmann, 2001).

Son rôle majeur est la reconnaissance du site 5' donneur d'épissage lors des étapes précoces d'assemblage du spliceosome. Plusieurs de ces constituants y contribuent :

- En premier lieu, l'extrémité 5' du snRNA U1 qui contient une séquence (ACUUACCUG) (en orange dans la Fig. 26) très conservée chez toutes les espèces et complémentaire au site 5' donneur d'épissage des introns épissés par les spliceosomes majeurs (Mount *et al.*, 1983 ; Branlant *et al.*, 1981). Le snRNA U1 peut donc se lier 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 protéine U1-70K contient un domaine RRM[10] reconnaissant spécifiquement une structure tige boucle du snRNA U1, ainsi qu'un domaine RS[11] lui permettant d'interagir avec la famille des protéines SR qui, nous le verrons plus loin, sont largement impliquées dans les mécanismes assurant la reconnaissance et l'utilisation des sites d'épissage (Query *et al.*, 1989a ; Wu and Maniatis, 1993 ; Eperon *et al.*, 1993).

- La protéine U1C participerait elle aussi, par des interactions avec le pré-ARNm, à la reconnaissance entre la snRNP U1 et le site 5' d'épissage (Du and Rosbash, 2002).

1.3.2.2.La snRNP U2 (12, 15 et 17S)

Les composants de la snRNP U2 (17S) sont le snRNA U2, les 7 protéines Sm, les protéines U2A' et U2B'', qui forment ensemble la particule 12S, et les facteurs SF3b[12] puis SF3a qui en s'associant donnent lieu aux particules 15S puis 17S (Fig. 27). La snRNP U2 permet dans un premier temps de reconnaître la boîte de branchement de l'intron, puis dans les étapes plus tardives, elle forme le coeur catalytique du spliceosome avec la particule snRNP U6.
Ainsi, le snRNA U2 contient à la fois une séquence (UGUAGUA) qui s'hybride à la boîte de branchement des pré-ARNm (nucléotides en violet dans la Fig. 27), et une séquence permettant de former des appariements avec le snRNA U6 (nucléotides en vert dans la Fig. 27).

1.3.2.3.La tri-snRNP U4/U6.U5

La tri-snRNP U4/U6.U5 est une tri-particule formée par association de la snRNP U4/U6 à la snRNP U5 (Behrens and Luhrmann, 1991).

[10] RNA Recognition Motif
[11] aRginine Sérine
[12] Splicing Factor 3

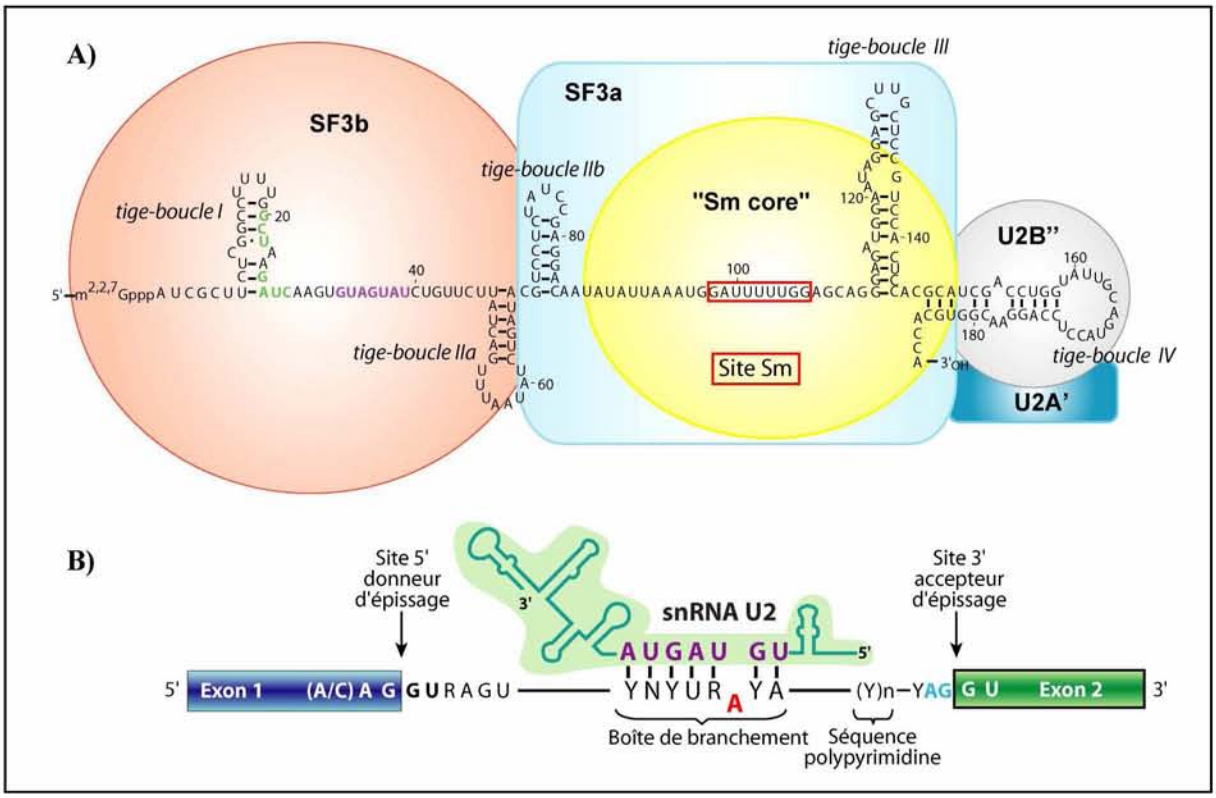


Fig. 27 : Représentation schématique de la particule snRNP U2 et appariements avec le pré-ARNm

A) Les sites de fixation sur le snRNA U2 des protéines de la particule 17S sont représentés. Les nucléotides représentés en violet sont impliqués dans l'appariement avec la séquence de branchement de l'ARN pré-messager. Ceux représenté en vert dans l'interaction avec le snRNA U6. Les nucléotides encadrés correspondent au site Sm (Branlant et al., 1982). B) Appariement entre le snRNA U2 et la boîte de branchement de l'ARN pré-messager. Les nucléotides N, R et Y désignent respectivement, un nucléotide quelconque, une purine et une pyrimidine (D'après Reddy et al., 1981 ; Branlant et al., 1982 ; Ares, 1986).

Les snRNP U4 et U6

Au sein de cette particule, les snRNA U4 et U6 sont associés sous forme de duplex (Fig. 28 ; Hashimoto and Steitz, 1984 ; Mougin *et al.*, 2002).

Ces deux snRNA comportent chacun des séquences leur permettant de s'apparier entre eux (nucléotides en bleu dans la Fig. 28). Le snRNA U6 comporte en plus des séquences lui permettant de s'apparier au snRNA U2 (nucléotides en vert dans la Fig. 27 ; Madhani and Guthrie, 1992), et une séquence très conservée (nucléotides 38-57) nécessaire à la fixation d'un facteur de recyclage SART3/p110 (ou Prp24 chez la levure) (Raghunathan and Guthrie, 1998 ; Bell *et al.*, 2002 ; Rader and Guthrie, 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 and Steitz, 1984 ; Rinke *et al.*, 1985 ; Brow and Guthrie, 1988 ; Mougin *et al.*, 2002).

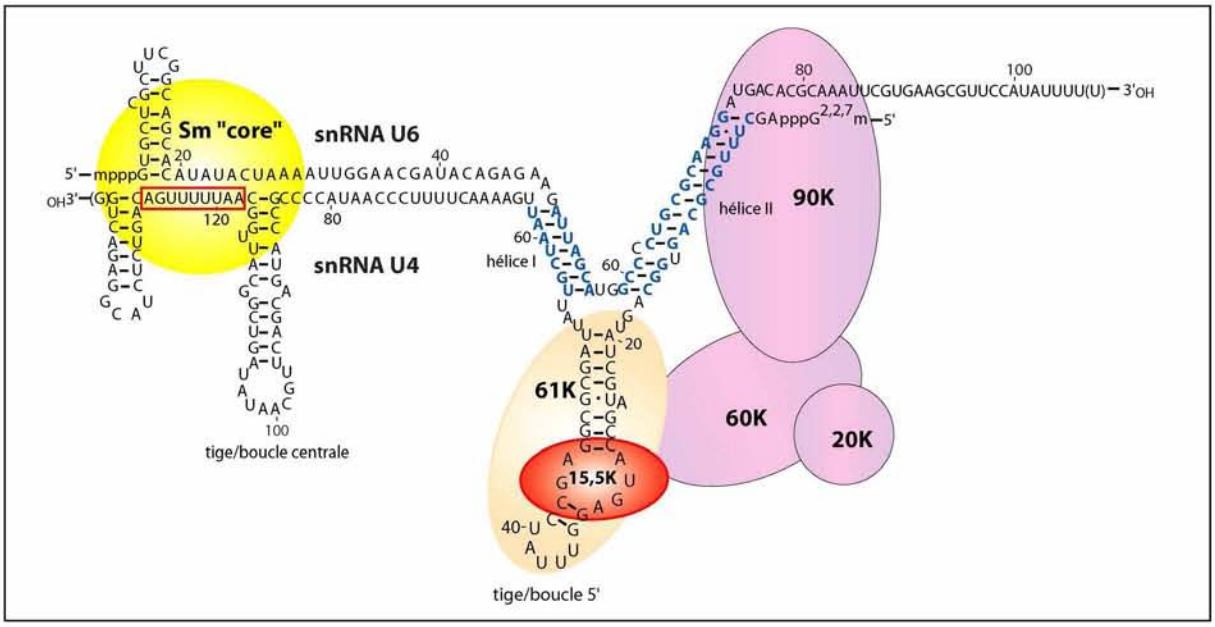


Fig. 28 : Représentation schématique de la particule U4/U6 (13S)

Les nucléotides en bleu sont les séquences nécessaires pour l'appariement des deux UsnRNA sous forme de duplex.

La snRNP U5

Les composants de la particule snRNP U5 (20S) sont le snRNA U5 associé aux 7 protéines Sm, et au moins 9 protéines spécifiques : U5-15K, 40K, 52K, 100K, 102K, 110K, 116K, 200K et 220K) (Fig. 29 ; Behrens and Luhrmann, 1991). Le rôle majeur de la snRNP U5 semble être d'aligner les exons lors de la seconde étape de transestérification.

Les trois protéines U5-200K, 116K et 100K sont impliquées dans le remplacement du snRNA U1 par le snRNA U6 au site 5' d'épissage. La protéine 102K permettrait de former « un pont » entre la snRNP U4/U6 et la snRNP U5 (Legrain and Choulika, 1990 ; Makarov *et al.*, 2000). La protéine U5-220K est impliquée dans de multiples aspects du remodelage du spliceosome. Cette protéine établit des contacts avec de nombreuses régions du snRNA U5 (Mougin *et al.*, 2002), du snRNA U6 (Vidal *et al.*, 1999), ainsi que les sites d'épissage du pré-ARNm.

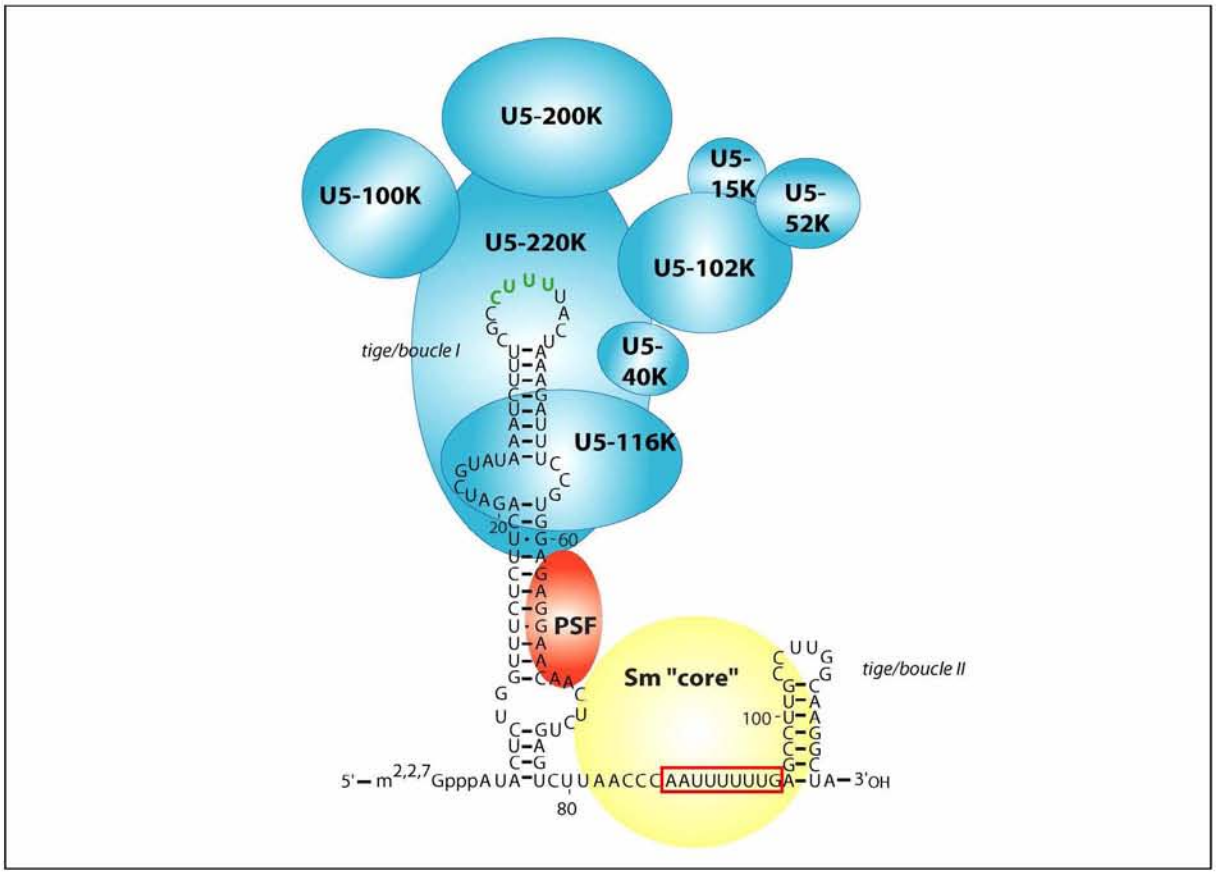


Fig. 29 : Représentation schématique de la snRNP U5

Les nucléotides indiqués en vert sont impliqués dans des appariements avec les deux exons. Seules les protéines U5-220K, U5-116K et PSF ont été montrées s'associer directement au snRNA U5. Les protéines U5-100K, U5-200K, U5-102K, U5-40K et U5-15K ont été représentées en tenant compte des interactions proposées entre les protéines de la snRNP U5.

La tri-snRNP U4/U6.U5

Seules trois protéines (27, 65 et 110K) 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 and Garcia-Blanco, 1995 ; Makarova *et al.*, 2001). Deux des protéines de la tri-snRNP, U5-102K et U4/U6-61K, 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 (Makarova *et al.*, 2002).

1.4. L'épissage : assemblage du spliceosome et catalyse de la réaction

L'assemblage du spliceosome autour de l'intron se fait de manière séquentielle. Ainsi chez les vertébrés, les premières étapes de l'épissage consistent à reconnaître les sites donneur et

accepteur d'épissage. S'assemble ensuite le reste de la machinerie d'épissage qui confère à l'ARN pré-messager une configuration spatiale adéquate pour permettre la réaction d'épissage. Les études *in vitro* de l'assemblage du spliceosome sur l'ARN pré-messager ont permis de mettre en évidence la formation successive de 5 complexes ayant une mobilité électrophorétique différente lorsqu'ils sont fractionnés en gel non dénaturant (Fig. 30 ; Jurica *et al.*, 2002).

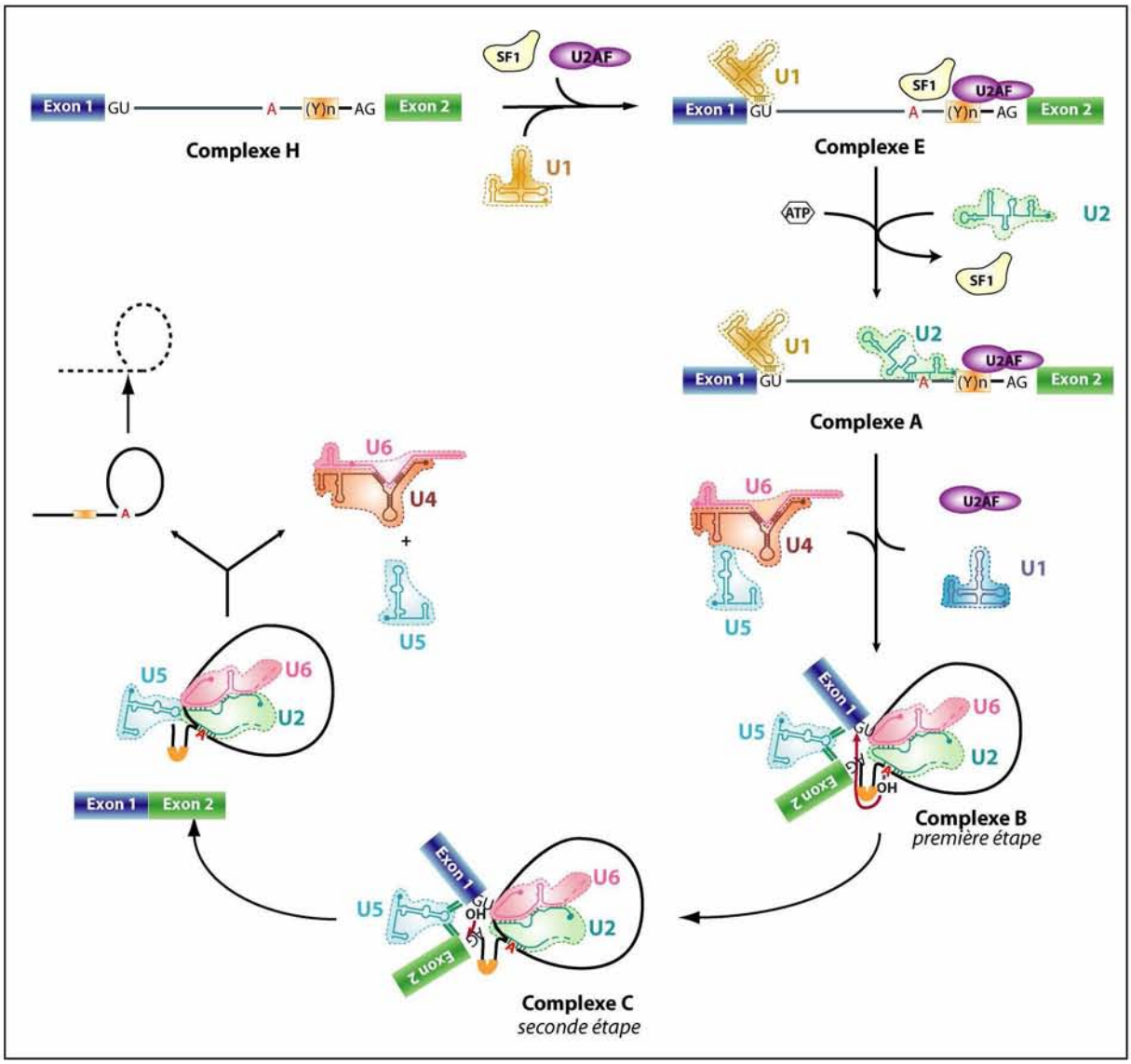


Fig. 30 : L'assemblage du spliceosome autour des introns majeurs

Les particules snRNP U1, U2, U4/U6 et U5 sont schématisées par un trait de couleur entouré d'une aire colorée. U2AF : U2 Auxiliary Factor ; SF1 : Splicing Factor 1. (adapté de Patel and Steitz, 2003)

Ces complexes correspondent à l'assemblage ordonné des différentes UsnRNP sur l'ARN pré-messager. Un grand nombre de protéines spliceosomales s'associent aussi de manière stable ou de façon transitoire.

1.4.1. Le complexe H (pour Heterogeneous)

Dès leur synthèse les transcrits Pol II sont associés à un ensemble de protéines, formant un complexe ribonucléoprotéique appelé complexe H. Sa formation ne dépend pas de la présence d'intron. L'analyse de ce complexe, obtenu *in vitro* simplement en mettant en contact un ARN avec un extrait nucléaire, révèle que la majorité des protéines fixées sont des protéines hnRNP[13] avec en plus quelques protéines plus spécifiques dont l'identité dépend du pré-ARNm (Bennett *et al.*, 1992a; Bennett *et al.*, 1992b). Certaines de ces protéines restent associées aux transcrits pendant sa maturation et peuvent influencer l'efficacité d'épissage. Nous verrons plus loin quelles sont ces protéines et comment elles régulent l'épissage.

1.4.2. Formation du complexe d'engagement ou complexe E

In vitro, après incubation à 30°C et en absence d'ATP, le complexe H contenant un pré-ARNm est rapidement converti en complexe E qui renferme en particulier, la snRNP U1, les facteurs généraux d'épissage U2AF et SF1/mBBP, et des protéines SR. La formation de ce complexe dépend de la reconnaissance des sites donneurs et accepteurs d'épissage, et conduit à leur rapprochement dans l'espace (Fig. 31).

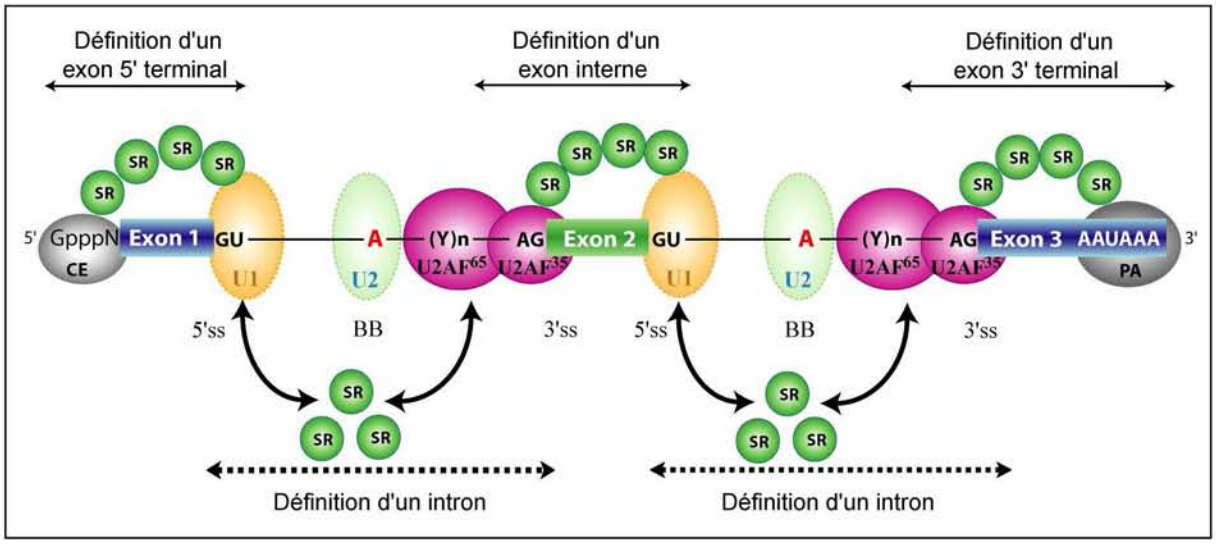


Fig. 31 : La définition des introns et des exons d'un ARN pré-messager au cours de l'assemblage du spliceosome *(adapté de Kohtz et al., 1994)*

5'ss, BB et 3'ss désignent respectivement les sites donneurs 5', les boites de branchement et les sites accepteurs 3' d'épissage. Les ovales gris représentent le complexe de liaison à la coiffe (CE) en 5' et la machinerie de polyadénylation (PA) en 3'. Le nombre de protéines SR représentées est arbitraire.

[13] heterogeneous nuclear RNP

1.4.2.1.Reconnaissance du site 5’ d’épissage

Comme mentionné ci-dessus, l’appariement de bases entre le snRNA U1 et le site donneur d’épissage permet à la snRNP U1 de se positionner sur le site 5’ d’épissage. Cependant, en fonction du degré de complémentarité, la particule peut aussi être stabilisée au niveau du site d’épissage grâce à de multiples interactions avec les protéines SR. L’effet de ces protéines est tel que des études *in vitro* ont démontré qu’à forte concentration, la présence de ces protéines SR peut compenser l’absence du snRNA U1 et induire un épissage (Crispino *et al.*, 1994). Les rôles des différentes protéines SR étant tout à fait multiples et primordiaux, cette famille de protéines fera l’objet d’un paragraphe particulier plus loin dans cette introduction.

1.4.2.2.Reconnaissance du site 3’ d’épissage

Au sein du complexe E, il a été proposé que deux facteurs participent à la reconnaissance du site 3’ d’épissage : U2AF[14] et SF1/mBBP[15] (Fig. 32).

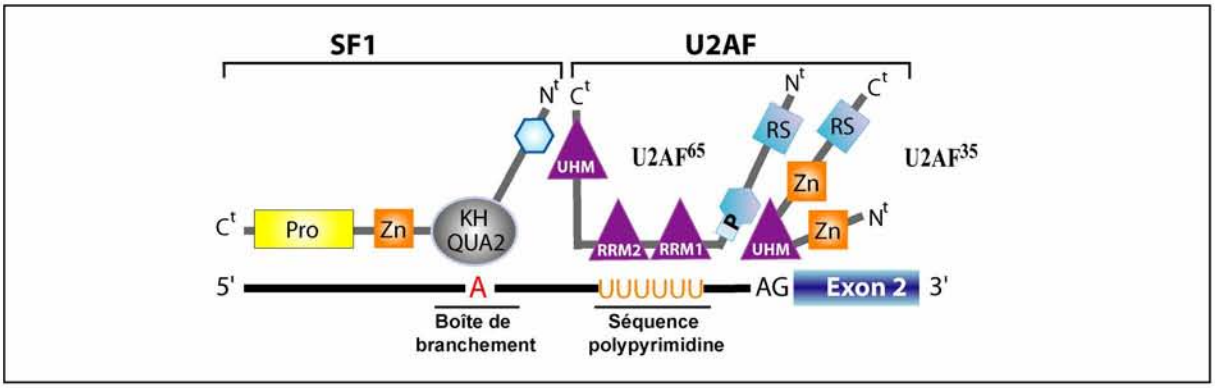


Fig. 32 : Reconnaissance de la borne 3’ par les facteurs U2AF et SF1 (*d'après Selenko et al., 2003)****.***

Le facteur U2AF est constitué de deux sous-unités : U2AF[65] qui se fixe sur la séquence polypyrimidine grâce à deux motifs RRM spécifiques, et U2AF[35] qui se fixe au niveau du dinucléotide AG du site 3’ accepteur d’épissage, grâce à un domaine de liaison à l’ARN non canonique. Lorsque la séquence polypyrimidine est sous optimale, c’est-à-dire que l’enchaînement de pyrimidines est entrecoupé de purines, l’interaction de la petite sous-unité U2AF[35] avec le site 3’ d’épissage est essentielle pour stabiliser la fixation de U2AF[65] à la séquence polypyrimidine. On distingue ainsi 2 types d’exons, dits AG-dépendants ou AG-

[14] U2 snRNP Auxiliary Factor
[15] Splicing Factor 1/mammalian Branch point Binding Protein

indépendants, suivant que la séquence polypyrimidine est sous optimale ou non (Fig. 33 ; Reed, 1989).

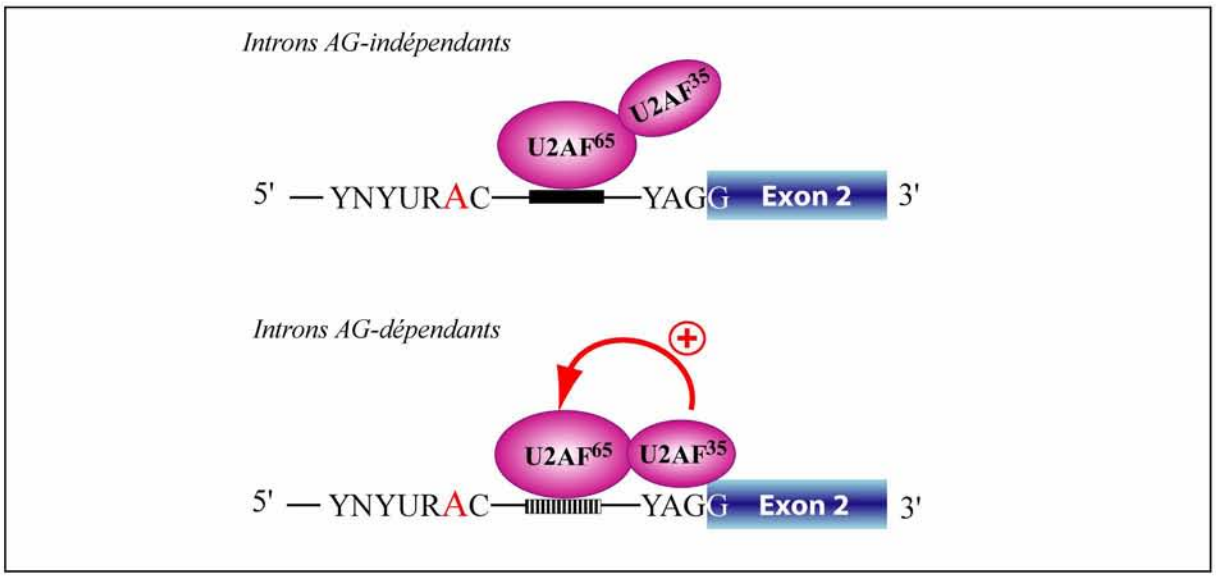


Fig. 33 : Interactions entre l'ARN pré-messager et U2AF autour du site 3' d'épissage des introns AG-dépendants ou AG-indépendants *(pour revue, Wu et al., 1999)*

Pendant des années il a été proposé que le facteur SF1/mBBP se fixe au niveau de la boîte de branchement, et possède une région d'interaction avec U2AF^{65} (Rain *et al.*, 1998 ; Peled-Zehavi *et al.*, 2001). Néanmoins certaines études montrent qu'il n'est pas indispensable à la cellule, tandis que d'autres montrent que son absence entraîne un ralentissement de la cinétique d'assemblage des complexes spliceosomaux. Son rôle exact reste donc encore à démontrer (Guth and Valcarcel, 2000 ; Tanackovic and Kramer, 2005).

1.4.2.3.Rapprochement des sites 5' et 3' d'épissage

A ce stade précoce de l'assemblage des complexes spliceosomaux, les sites 5' et 3' ont été reconnus par la machinerie d'épissage. Un important réseau d'interactions s'établit alors entre les protéines liées au pré-ARNm et les protéines des UsnRNP, ce qui permet de rapprocher les extrémités des 2 exons à liguer. Une étude montre en ce sens que le réseau d'interactions établi dans le complexe E place les sites 5' et 3' d'épissage à une distance inférieur à 20 Å (Kent and MacMillan, 2002).

SF1/mBBP contribuerait, si il joue vraiment un rôle activateur, à ce rapprochement en interagissant avec la protéine FBP 11, elle-même en interaction avec la snRNP U1 (Bedford *et al.*, 1998). De part leurs interactions possibles avec la protéine U1-70K de la snRNP U1 et la

petite sous-unité de U2AF[35] du facteur U2AF, les protéines SR sont sans doute aussi fortement impliquées dans ce processus (Fig. 34A ; Wu and Maniatis, 1993 ; Eperon *et al.*, 1993 ; Kohtz *et al.*, 1994 ; Zhang and Wu, 1996).

Enfin, des études ont montré que la snRNP U2 serait aussi présente dans le complexe E, mais associée par des interactions faibles (Das *et al.*, 2000 ; Huang *et al.*, 2002). Il faut noter par ailleurs qu'un complexe intermédiaire E' peut être purifié en absence d'U2AF (Kent *et al.*, 2005).

1.4.3. Formation du complexe A

En présence d'ATP, le complexe E est converti en complexe A ou pré-spliceosome, caractérisé par l'association stable de la snRNP U2 17S au niveau de la boîte de branchement du pré-ARNm (Fig. 34B ; Ruskin and Green, 1985).

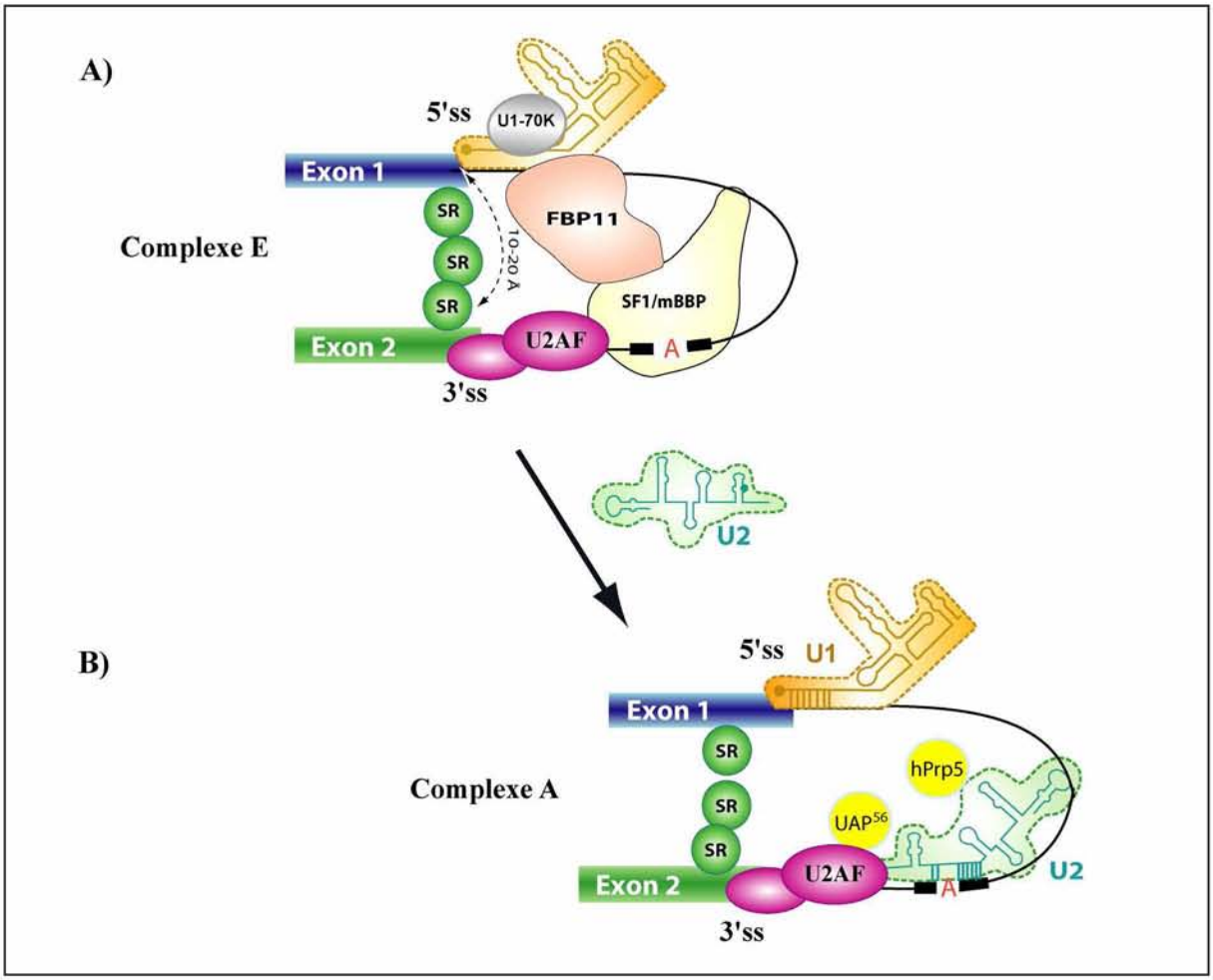


Fig. 34 : Modèle d'assemblage des complexes E et A sur le pré-ARNm *(adapté de Das et al., 2000)*

Pour cela, le snRNA U2 s'apparie à la boite de branchement et toutes les protéines des complexes SF3a et SF3b participent à l'ancrage (Champion-Arnaud and Reed, 1994 ; Gozani *et al.*, 1996 ; Nesic and Kramer, 2001). Un des facteurs essentiels pour cette association est la protéine SF3a155 qui remplace le facteur SF1/mBBP dans ces interactions avec le point de branchement d'une part et la sous-unité U2AF^{65} d'autre part (Gozani *et al.*, 1998). La stabilisation de la particule snRNP U2 est aussi générée par des protéines à activité ARN hélicase, telles que UAP56 et hPrp5 qui participent notamment à la réorganisation de la particule (Perriman *et al.*, 2003 ; Xu *et al.*, 2004). Ces changements conformationnels expliquent pourquoi la formation du complexe A est une étape ATP-dépendante.

En plus de la reconnaissance de la boite de branchement, la snRNP U2 induit aussi la formation, dans l'ARN pré-messager, d'une courte hélice rendant accessible l'adénosine responsable de la première attaque nucléophile de la réaction d'épissage (Massenet *et al.*, 1999 ; Newby and Greenbaum, 2002 ; Zhao and Yu, 2004).

Enfin, il faut noter que la purification du complexe A met encore une fois en évidence la présence de nombreuses protéines SR dont il est probable qu'elles participent à la fois au recrutement de la snRNP U2 mais aussi à la bonne structuration du complexe A (Hartmuth *et al.*, 2002).

1.4.4. Formation du complexe B

Le passage du complexe A au complexe B est aussi une étape ATP-dépendante qui repose tout d'abord sur l'incorporation de la tri-snRNP [U4/U6.U5]. Plusieurs types de facteurs sont impliqués dans cette première étape:

- un ensemble de protéines SR phosphorylées associées à la tri-particule et qui favorisent l'interaction avec le complexe A (Roscigno and Garcia-Blanco, 1995),
- trois protéines SR-like (27K, 65K et 110K), spécifiques de la tri-snRNP [U4/U6.U5],
- et enfin, le facteur SPF30[16], qui semble jouer un rôle lors des réarrangements successifs conduisant au complexe B (Meister *et al.*, 2001 ; Rappsilber *et al.*, 2001)

En effet, cette incorporation s'accompagne de nombreux réarrangements structurels au sein même de la tri-snRNP. En premier lieu l'hétéroduplex entre les snRNA U4 et U6 est rompu au profit d'interactions entre les snRNA U2 et U6 (Fig. 35 ; Hausner *et al.*, 1990 ; Wolff and Bindereif, 1992 ; Mougin *et al.*, 2002).

[16] SPF30

Ces réarrangements conduisent au départ du snRNA U4 une fois le snRNA U6 incorporé dans les complexes spliceosomaux. D'où l'idée que le snRNA U4 est un rôle de masquage des séquences fonctionnelles du snRNA U6 avant son utilisation dans les complexes actifs. Le réseau d'interactions ARN/ARN qui s'établit dans le complexe B, permet ensuite de rapprocher l'extrémité 5' de l'intron et la séquence de branchement pour faciliter la première étape de la réaction d'épissage (Fig. 35).

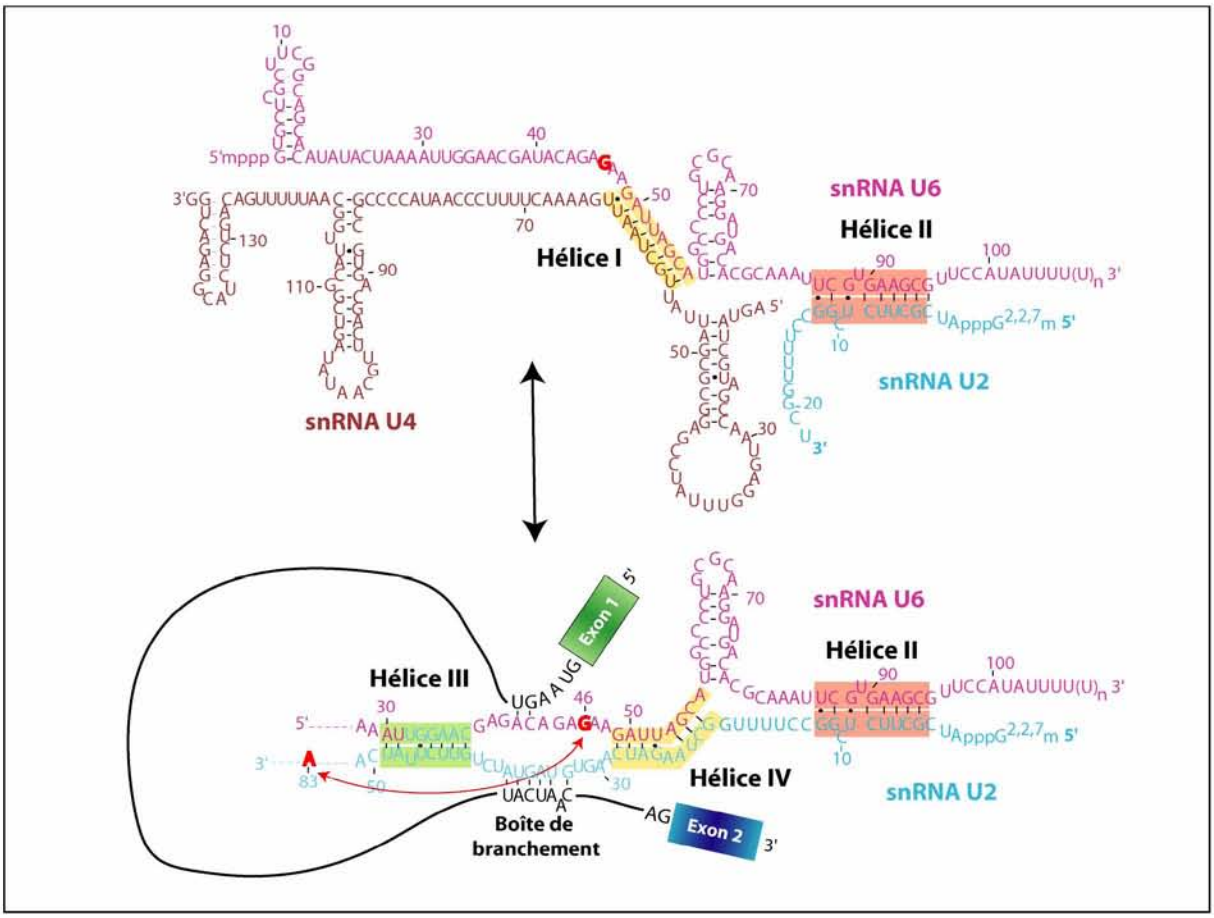


Fig. 35 : Mécanisme proposé pour le passage de l'interaction entre les snRNA U4 et U6 à celle entre les snRNA U2 et U6 chez les mammifères *(Mougin et al., résultats non publiés)*

Une interaction supplémentaire s'établi ensuite entre le snRNA U6 et la séquence en 5' de l'intron au dépend de l'interaction snRNA U1/site 5' (Sawa and Shimura, 1992 ; Wassarman and Steitz, 1993). Ce remplacement est favorisé par la protéine U5-100K, l'hélicase p68 et la protéine SR-like 27K qui remplace la protéine U1-70K dans son interaction avec le site 5' (Fig. 36 ; Ismaili *et al.*, 2001 ; Liu, 2002 ; Fetzer *et al.*, 1997).

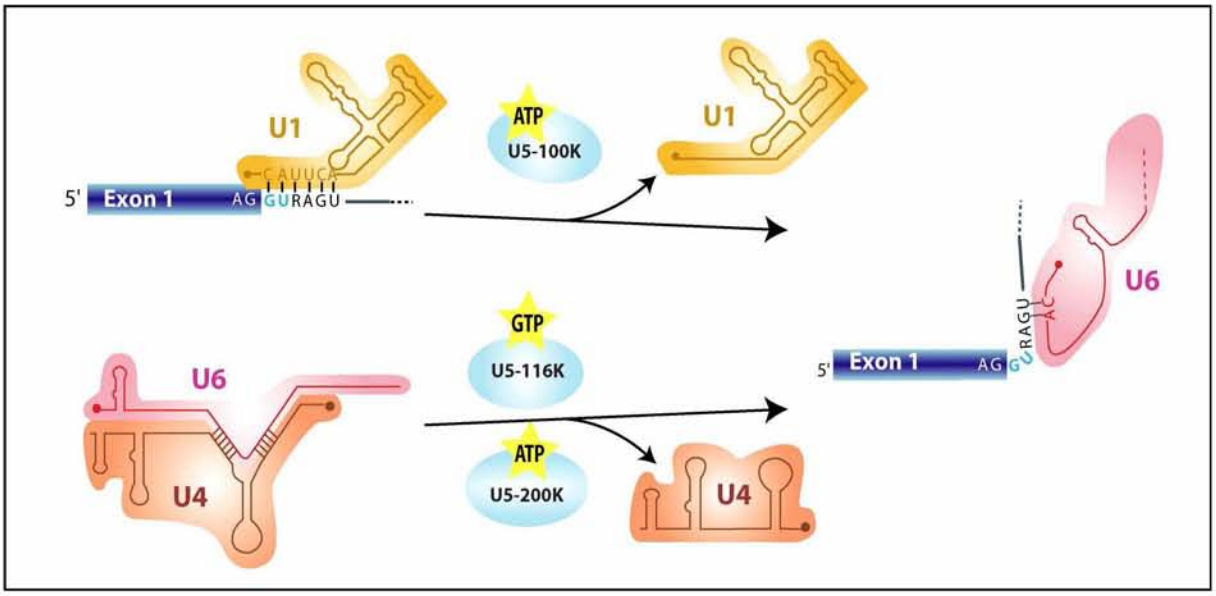


Fig. 36 : Rôles des protéines de la snRNP U5 dans la dissociation de la snRNP U1 avec le site 5' et son remplacement par la snRNP U6 *(adapté de Turner et al., 2004)*

Enfin, le snRNA U5 s'apparie à la bordure 3' de l'exon 1 (Wyatt *et al.*, 1992 ; Sontheimer and Steitz, 1993). La forte dégénérescence de la séquence exonique contraste avec la forte conservation de séquence de la boucle I du snRNA U5 mise en jeu (Ségault *et al.*, 1999), mais la richesse en U de cette boucle ouvre de larges possibilités d'interactions Watson-Crick ou non Watson-Crick (Fig. 37) et l'interaction est stabilisée par des protéines, en particulier, par la protéine U5-220K (Wyatt *et al.*, 1992 ; Reyes *et al.*, 1996).

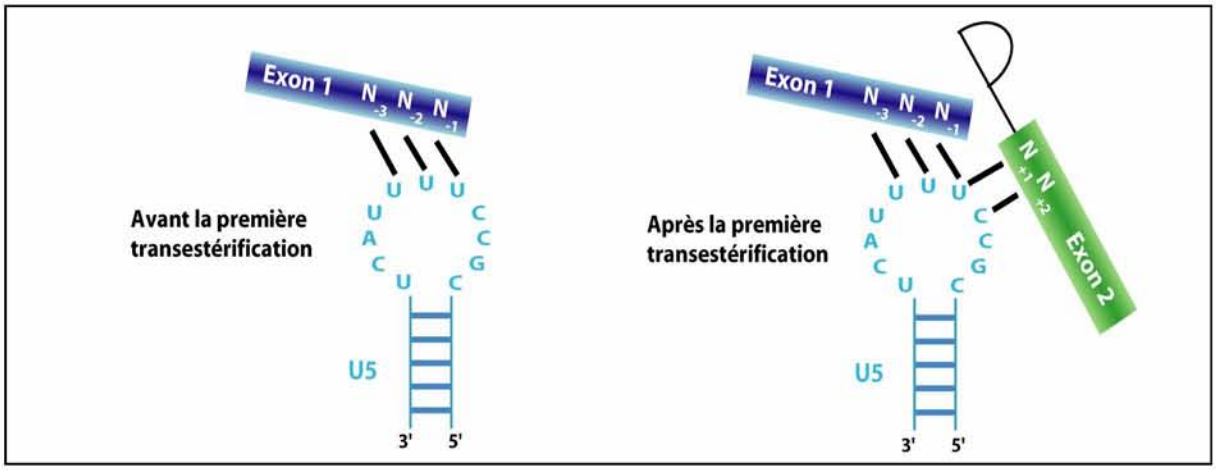


Fig. 37 : Interactions entre le snRNA U5 et les exons avant et après la première réaction d'épissage *(pour revue, Madhani and Guthrie, 1994)*

1.4.5. Le complexe B*

Le complexe B* (45S) correspond au complexe où a lieu la première étape de transestérification de la réaction d'épissage (Makarova *et al.*, 2004). Ce complexe comporte des quantités équimolaires des snRNA U2, U5 et U6. Les snRNA U1 et U4 sont absents après fractionnement. L'analyse par spectrométrie de masse du contenu en protéines du complexe B* (Makarova *et al.*, 2004) a permis d'identifier une centaine de protéines, dont 80 étaient des facteurs d'épissage déjà connus (Tableau 1). Dans ce complexe, tous les composants de la snRNP U2 (17S), ainsi que deux protéines associées à cette snRNP (SR140 et hPrp43) sont présents. Par contre, deux protéines spécifiques à la snRNP U5 (U5-15K et U5-100K), ainsi que toutes les protéines spécifiques à la di-snRNP U4/U6 et deux protéines spécifiques à la tri-snRNP 25S (27K et 61K) sont absentes, indiquant que durant les réarrangements structuraux précédemment décrits, ces facteurs sont dissociés des complexes (Makarova *et al.*, 2004).

1.4.6. Le complexe C

Caractéristiques

Après la première réaction de transestérification, de nouvelles transitions structurales ont lieu et conduisent à la formation du complexe C. La seconde réaction de transestérification a lieu au sein de ce complexe conduisant à la ligation des exons. La composition en protéines du complexe C purifié a été analysée par spectrométrie de masse (Tableau 1 ; Jurica *et al.*, 2002; Jurica and Moore, 2003). Parmi les protéines identifiées, sont incluses des protéines associées aux snRNP U2 et U5, de nombreuses protéines hnRNP et des protéines appartenant aux complexes EJC[17], qui est un complexe de protéines restant sur l'ARN messager après épissage. Ces complexes EJC seront détaillés plus loin dans ce chapitre. Des protéines connues pour être indispensables à la seconde étape d'épissage, telles que hPrp16, hPrp17, hPrp22 et hSlu7, sont aussi associées au complexe C.

[17] Exon Junction Complex

Protéines	N° accès	Caractéristiques	S.c.	BΔU1	B*	C	35S U5
snRNP U2 (17S)							
SF3b 155	AF054284	HEAT	Hsh155	X	X	X	
SF3b 145	NP_006833	SAP	Cus1p	X	X	X	
SF3b 130	NP_036558	CPSF A	Rse1p	X	X	X	
SF3b 49	NP_005841	2 RRM	Hsh49p	X	X		
p14	AAK94041	RRM	Snu17p	X	X		
SF3b 14b	Q9UH06		Rds3p	X	X		
SF3b 10	NP_112577			X	X		
SF3a 120	NP_005868	2 SURP, UBQ	Prp21p	X	X	X	
SF3a 66	NP_009096	C2H2-ZF	Prp11p	X	X	X	
SF3a 60	NP_006793	C2H2-ZF, SAP	Prp9p	X	X	X	
U2 A'	P09661	LRR	Lea1p	X	X	X	
U2 B''	P08579	2 RRM	Msl1p	X	X	X	
protéines associées à la snRNP U2 (17S)							
hPrp43	O43143	DEAH	Prp43p	X	X	X	
SR140	BAA20790	RRM, Surp, SR			X		
SPF45	Q96I25	RRM, G-patch		X			
SPF30	NP_005862	Tudor		X			
snRNP U5							
220K	AAC61776		Prp8p	X	X	X	X
	O75643	DEIH, DxxH	Brr2p	X	X	X	X
116K	NP_004238	GTP-EF2		X	X	X	X
102K	AAF66128	TPR	Prp6p	X	X	X	
100K	NP_004809	DEAD	Prp28p			X	
40K	XP_001618	WD40		X	X	X	X
15K	O14834	thioredoxin-like	Dib1p	X			
tri-snRNP							
110K	AAK49523	SR	Snu66p	X	X	X	
snRNP U4/U6							
90K	NP_004689		Prp3p	X			
60K	O43172	WD40	Prp4p	X		X	
20K	O43447	PPIase		X			
61K	NP_056444	NOP homology	Prp31p	X			
15.5K	NP_004999	ribosomal_L7AE	Snu13p	X			
protéines Sm							
SmB/B'	NP_003082	Sm	SmB1	X	X	X	X
SmD3	NP_004166	Sm	SmD3	X	X	X	X
SmD2	NP_004588	Sm	SmD2	X	X	X	X
SmD1	NP_008869	Sm	SmD1	X	X	X	X
Sm E	NP_003085	Sm	SmE	X	X	X	X
Sm F	NP_003086	Sm	SmF	X	X	X	X
Sm G	NP_003087	Sm	SmG	X	X	X	X
protéines Lsm							
LSm3	Q9Y4Z1	Sm	LSm3	X	X	X	
LSm2	Q9Y333	Sm	LSm2		X	X	
LSm4	NP_036453.1	Sm	LSm4	X			
LSm7	NP_057283.1	Sm	LSm7	X			
LSm8	NP_057284.1	Sm	LSm8	X			
protéines hnRNP							
L	P14866	3 RRM			X		
G	NP_002130	RRM			X	X	
C1/C2	NP_004491	RRM		X	X	X	
A2/B1	NP_112533	2 RRM			X	X	
M	P52272	3 RRM		X		X	
A1	P09651	2 RRM, RGG				X	
C	P07910	RRM				X	
F	P52597	3 RRM				X	
G	P38159	RRM				X	
H	P31943	3 RRM				X	
K	Q07244	3 KH, RGG				X	
R	O43390	3 RRM, RGG				X	
U	Q00839	RGG				X	
Protéines SR							
SRm300	Q9UHA8	SR				X	
hTra-2	Q13595	RRM, SR			X		
9G8	NP_006267	RRM, SR		X	X		
SRp30c	NP_003760	2 RRM, SR			X		
SRp20	P23152	RRM, SR		X	X		
CBC							
CBP 80	A54748	MIF4G	Cbc1	X	X	X	
CBP 20	P52298	RRM	Cbc2p	X	X		
EJC							
ALY	NP_005773	RRM	Yra1p		X	X	
Y14	Q9Y5S9	RRM				X	
Srm160	AAC0932.1	PWI, SR				X	
Magoh	P50606					X	
Nuk-34/eIF4A3	P38919	DEAH			X	X	
Facteurs d'initiation de la traduction							
eIF3, sub6 IP	AAH178871.1	TPR			X		
eIF3, sub6	NP_001559	PCI			X		
eIF3, sub2	NP_003748	5 WD			X		
Protéines ribosomiques							
L14, SA, S2, S3, S3a, S4, S5, S6, S7, S12, S13					X		
S17, S18, S19, S20, S21, S25					X		

Protéines	N° accès	Caractéristiques	S.c.	BΔU1	B*	C	35S U5
Facteurs d'épissage requis pour la 2ème étape							
hPrp22	NP_004932	DEAH, S1	Prp22p		X	X	
Prp16	Q92620	DEAH	Prp16p			X	
Prp17	O60508	7 WD40	Prp17p			X	X
Slu7	NP_006416	Zn-F	Slu7p			X	
protéines associées à la snRNP U5 (35S)							
hPrp19	NP_055317	7 WD40	Prp19p	X	X	X	X
CDC5	NP_001244	MYB-3	Cef1	X	X	X	X
hSyf1	AAH07208	TPR	Syf1p		X	X	X
hSyf3	NP_057736	TPR	Syf3p	X	X	X	X
hIsy1	NP_065752		Isy1p		X	X	X
PRL1	NP_002660	7 WD40	Prp46p		X	X	X
SPF27	NP_005863				X		X
SKIP	NP_036377	SKIP_SNW	Prp45p	X	X	X	X
hECM2	NP_060517	ZF-C3H, RRM	Ecm2p		X	X	X
GCIP p29	NP_056299				X	X	X
AD-002	AAF14858		Cwc15p		X	X	X
KIAA0560	XP_029045	(p-loop)			X	X	X
MGC23918	NP_653317				X		X
G10	P41223	(ZF)	Cwc14p		X		X
Cyp E	NP_006103	RRM, PPIase			X	X	X
PPIase-like 1	NP_057143	PPIase			X	X	X
Protéines spécifiques à BΔU1							
NPW38BP	NP_057396			X			
NPW38	NP_005701	WW		X			
N-CoR1	NP_006302	Myb DNA-binding		X			
hPrp4 kinase	AAH34969	Pkinase		X			
FBP-21	NP_009118	ZnF-U1, WW		X			
TBL1-rp 1	NP_078941	WD40	Sif2p	X			
Smc-1	NP_006297	DA-box, p-loop	Smc1p	X			
p220 (NPAT)	BAA11861			X			
Protéines en relation avec les complexes d'épissage							
MGC13125	NP_116114		Cwc26p	X	X		
p68	NP_004387	DEAH	Dbp2p	X	X	X	
hSnu23	NP_653324	ZnF-U1 (C2H2)	Snu23p	X	X		
	P11940			X	X		
PAPB2	Q15097	4 RRM			X	X	
RED	NP_006074	RD/RE-element		X	X		
hSmu-1	NP_060695	7 WD40		X	X		
MFAP1	NP_005917			X	X		
HsKin17	NP_036443	ZnF-U1, KOW	Rts2p	X	X		
pinin	NP_002678	SR		X	X	X	
UBL5p	NP_077268	UBQ	Hub1p	X	X		
TIP-48	NP_006657	AAA-ATPase		X			
TIP-49	NP_003698	AAA-ATPase	Rvb1p		X		
KIAA1590	BAB13416	kinesin		X	X		
hDbp5	NP_009173	DEAH	Dbp5p		X		
CPSF 160K	XP_066300	CPSF_A			X		
PSF	P23246	2 RRM			X		
CA150	NP_006697	WW, FF			X		
		AQ-repeats					
CDCL2 p110	NP_284923	Sas10/Utp3			X		
HDB/DICE1	NP_036273	VWF, DEAD/H			X		
Cyp-60	S64705	PPIase			X		
KIAA0853	NP_055885	RD/RE			X		
p55PIK	Q92569	SH3			X		
TPX2	CAC10005	(p-loop)			X		
LOC56912	NP_064538				X		
RACK1	P25388	7 WD40	Asc1p		X		
NOSIP	NP_057037				X		
dJ14N1.2	CAC13173	EF-hand			X		
DKFZP586O0120	NP_054796				X		
PPIase-like 3	NP_570981	PPIase			X	X	
TFIP11	NP_036275	G-patch			X	X	
Abstrakt	Q9UJV9	DEAD	(Ded1p)			X	
KIAA0073	BAA07555	PPIase, 3 WD40				X	
KIAA1604	BAB13430	MIF4G				X	
Cactin	AAH04262					X	
DGSI protein	NP_073210					X	
MGC4238	NP_115708	5 WD				X	
SDCCAG10	NP_005860	PPIase				X	
FRG1	Q14331					X	
DKFZp434E2220	NP_060082	PSP				X	
DDX35	NP_068750	DEAH				X	X
FLJ10206	NP_060495	G-patch				X	X
HSP73	NP_006588						X
KIAA0773	NP_055505						X
C25A1.10.p	XP_060969						X
S100 A8	NP_002955	EF-hand					X
S100 A9	NP_002956	EF-hand					X

Tableau 1 : Protéines contenues dans les complexes BDU1, B*, C et U5 (35S) *(Makarova et al., 2004)*

Le nom des protéines associées aux différents complexes est listé avec le nom de certaines protéines homologues de levure. Leur numéro d'accès NCBI est indiqué ainsi que leur caractéristique.

1.4.7. Les complexes spliceosomaux pré-assemblés

L'assemblage ordonné des UsnRNP, que nous venons de décrire, est le modèle généralement admis. Cependant, des études récentes ont remis en question ce modèle, notamment la découverte, chez la levure, d'une penta-snRNP (U1, U2, U4, U5 et U6), non fixée au pré-ARNm (Nilsen, 2002 ; Stevens *et al.*, 2002). En plus, comme nous l'avons vu précédemment, des études par spectrométrie de masse de complexes spliceosomaux ont souvent révélé la présence de protéines non retrouvées dans les complexes fonctionnels. Il reste à savoir si leur présence dans les intermédiaires d'assemblage est due à un problème de purification ou si elles y jouent vraiment un rôle.

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

Pendant longtemps, la transcription, la maturation de l'ARN, c'est-à-dire la pose de la coiffe, l'épissage, le clivage de l'extrémité 3' et la polyadénylation, ainsi que l'export, la traduction et la dégradation des ARNm ont été approchées de manière indépendante. Les études qui ont permis d'aborder les choses de façon plus globale ont révélé plusieurs liens entre ces différentes étapes et parfois même des phénomènes de couplage. Ces interrelations sont dans l'ensemble bien décrites et d'autres fois plus floues.

2.1. Lien entre transcription et épissage

Le plus long gène humain connu (2400 kb), code la dystrophine. Sa transcription dure environ 16 h et il a été démontré que l'ARN pré-messager est épissé co-transcriptionnellement (Tennyson *et al.*, 1995). Le couplage transcription/épissage apparaît ici comme un concept très intuitif. En effet, il est très difficile de concevoir que l'épissage de la douzaine d'introns contenus puisse attendre la fin de la transcription avant de se réaliser. L'observation du transcrit naissant de gènes embryonnaires de drosophile par microscopie électronique a prouvé que l'épissage intervenait co-transcriptionnellement avec une fréquence raisonnable et que la sélection des sites d'épissage avait lieu avant la polyadénylation du transcrit (Beyer and Osheim, 1988). Il apparaît maintenant clairement que transcription et maturation de l'ARN sont des événements couplés (pour revue, Bentley, 2005).

Dans le cas de l'épissage, il est nécessaire de distinguer épissage co-transcriptionnel et couplage épissage/transcription. S'il est évident que certains introns sont éliminés avant la fin de la transcription (de manière donc co-transcriptionnelle), il n'en reste pas moins que certains introns sont épissés après la transcription. Pour exemple, dans le gène BR1 (Balbiani Ring 1),

l'intron 3 situé à 3 kb de l'extrémité 5' du pré-ARNm mesurant 40 kb, est excisé co-transcriptionnellement, alors que l'intron 4 situé 0,6 kb en aval du site de polyadénylation, est à 90% épissé après la transcription (Bauren *et al.*, 1996). C'est pourquoi le terme de couplage est préféré à la notion d'épissage co-transcriptionnel.

2.1.1. Rôle du CTD de l'ARN Pol II dans le couplage

Le chapitre précédent nous a permis d'entrevoir le rôle central du CTD de l'ARN Pol II dans les processus de transcription et de maturation de l'ARN. Nous allons à présent nous intéresser plus en détail aux relations entre le CTD de l'ARN Pol II et les divers constituants de la machinerie d'épissage. Suite à la démonstration que l'ARN Pol II phosphorylée est capable d'activer l'épissage *in vitro* (Hirose *et al.*, 1999), une étude a mis en évidence que cette activation est due au CTD de l'ARN Pol II qu'à la condition que le substrat contienne un exon interne (Zeng and Berget, 2000). L'idée est alors apparue que l'ARN Pol II serait impliquée à la fois dans la reconnaissance des exons, mais aussi dans le rapprochement des exons. Récemment, cette hypothèse a été renforcée par l'observation qu'un clivage au niveau de l'intron naissant n'empêche pas son épissage par la suite (Fig. 38 ; Dye *et al.*, 2006).

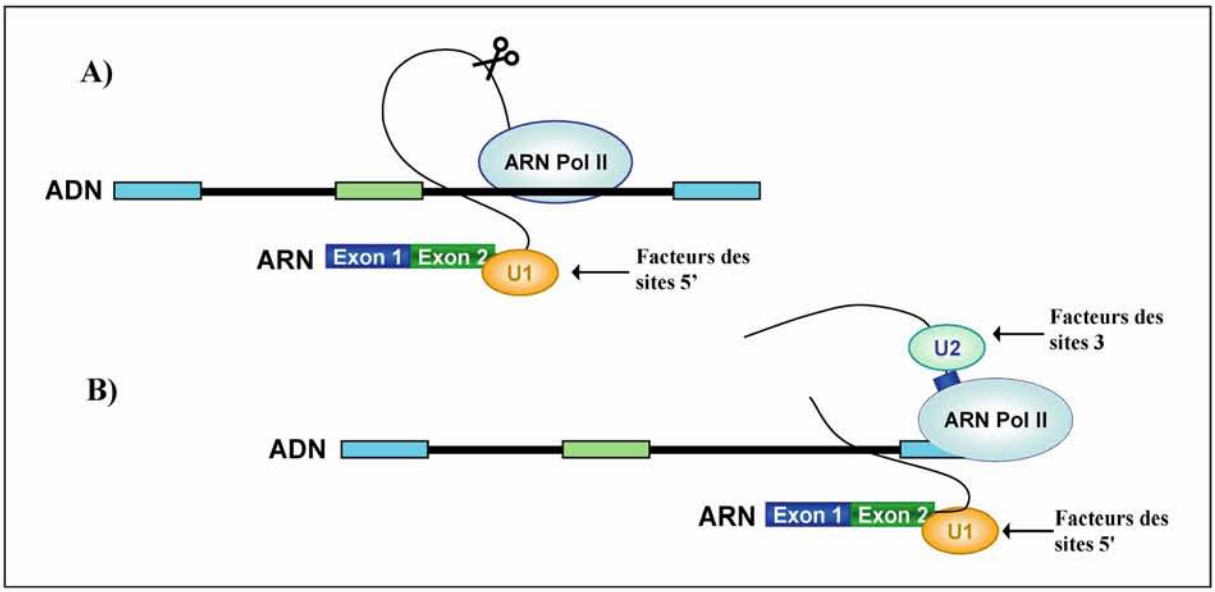


Fig. 38 : Modèle de rapprochement des exons au sein du complexe de transcription

A) Rapprochement des exons de l'ARN pré-messager par l'ARN Pol II. Dans le schéma, la courbe noire représente l'intron naissant et le ciseau représente le clivage de l'intron. B) Epissage co-transcriptionnel. Les facteurs de reconnaissance des sites 5' et 3' d'épissage, y compris les snRNP U1 et U2, sont en interaction avec le complexe transcriptionnel. L'épissage a lieu malgré le clivage de l'intron (D'après Dye et al., 2006).

En parallèle, plusieurs facteurs d'épissage sont maintenant recensés comme étant associés à l'ARN Pol II, en particulier son CTD ou à divers composants du complexe d'élongation :

- Ainsi, des anticorps généraux dirigés contre les protéines SR, les protéines Sm ou encore le facteur U2AF^{65}, ont permis d'immunoprécipiter l'ARN Pol II (Robert *et al.*, 2002 ; Kim *et al.*, 1997)
- Les facteurs PSF (PTB associated Splicing Factor) et p54^{nrb} ont été co-purifiés avec le CTD (Emili *et al.*, 2002).
- Et plus récemment, une étude a mis en évidence les caractéristiques structurales et fonctionnelle de l'interaction entre le snRNA U1 et la cycline H du facteur TFIIH (O'Gorman *et al.*, 2005)

Toutes ces données tendent vers l'idée que le complexe transcriptionnel favorise la reconnaissance des exons, leur rapprochement 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 and Spector, 1999).

2.1.2. Les protéines en lien avec transcription et épissage

L'analyse protéomique du spliceosome humain a permis de mettre en évidence qu'environ 30 des 145 protéines spliceosomales sont communes à d'autre processus de l'expression des gènes (pour revue, Jurica and Moore, 2003). Parmi elles, certaines apportent des exemples concrets du couplage épissage transcription (Tableau 2). Ainsi le cofacteur 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. C'est le cas par exemple de p54^{nrb}et PSF qui s'associent tout deux au site 5' d'épissage d'une part et au CTD de l'ARN Pol II hyperphosphorylé d'autre part (Kameoka *et al.*, 2004). Ces interactions ont lieu au sein de larges complexes fonctionnels incluant l'ARN Pol II et la snRNP U1, voire même la snRNP U2, et des facteurs impliqués dans le contrôle de l'élongation (TAT-SF1, P-TEFb, TFIIF) (Kameoka *et al.*, 2004). Le cas de PGC-1 est aussi un bon exemple puisque cet activateur thermogénique affecte l'épissage alternatif que lorsqu'il est recruté par des complexes associés au promoteur (Monsalve *et al.*, 2000).

Protéines	Lien avec la transcription	Lien avec l'épissage
Chez les mammifères		
SCAFs	Lie le CTD	Protéine SR-like
WT-1	Facteur de transcription	Lie U2AF65
PGC-1	Coactivateur	Régule l'épissage alternatif
CA150	Facteur d'élongation	Lie SF1
		Présent dans le spliceosome
XAB2	Facteur de transcription	Présent dans le spliceosome
SKIP	Facteur de transcription	Présent dans le spliceosome
TAT-SF1	Facteur d'élongation	Lie les UsnRNPs
p54nrb	Facteur de transcription	En complexe avec les UsnRNPs
PSF	Facteur de transcription	En complexe avec les UsnRNPs
TLS	Facteur de transcription	En complexe avec les UsnRNPs
CoAA	Coactivateur	hnRNP-like
		Régule l'épissage alternatif
Chez la levure		
PrP40	Lie le CTD phosphorylé	Facteur d'épissage
Spt5	Facteur d'élongation	Active l'épissage

Tableau 2 : Protéines candidates au lien entre transcription et épissage *(Yuryev et al., 1996 ; Davies et al., 1998 ; Monsalve et al., 2000 ; Goldstrohm et al., 2001 ; Fong and Zhou, 2001 ; Kameoka et al., 2004 ; Auboeuf et al., 2004 ; Morris and Greenleaf, 2000 ; Lindstrom et al., 2003)*

2.1.3. Importance du promoteur vis à vis de l'épissage

Plusieurs études faisant varier le type de promoteur utilisé pour la transcription ont permis de mettre en évidence le lien entre la nature du promoteur et l'efficacité d'épissage. 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). Ainsi, il a été récemment démontré qu'un transcrit Pol II est rapidement associé au spliceosome, alors que l'utilisation du promoteur T7 conduit à la formation d'un complexe H beaucoup moins épissé (Fig. 39 ; Das *et al.*, 2007 ; Das *et al.*, 2006).

Du fait que la plupart des gènes n'ont qu'un promoteur, une façon concevable de comprendre le rôle du promoteur dans le contrôle de l'épissage alternatif *in vivo* consiste à envisager le phénomène d'occupation du promoteur par des facteurs transcriptionnels de nature différentes. Ainsi, il a été démontré que bon nombre d'activateurs transcriptionnels influencent l'efficacité d'épissage (pour revue, Kornblihtt, 2005). Par exemple, la protéine Tat seule n'a qu'un faible effet sur l'épissage de l'exon EDI d'un minigène de laboratoire, alors qu'en présence d'activateurs tels que SW6, Sp1 et CTF, la présence de Tat provoque l'exclusion de cet exon EDI (Nogues *et al.*, 2002).

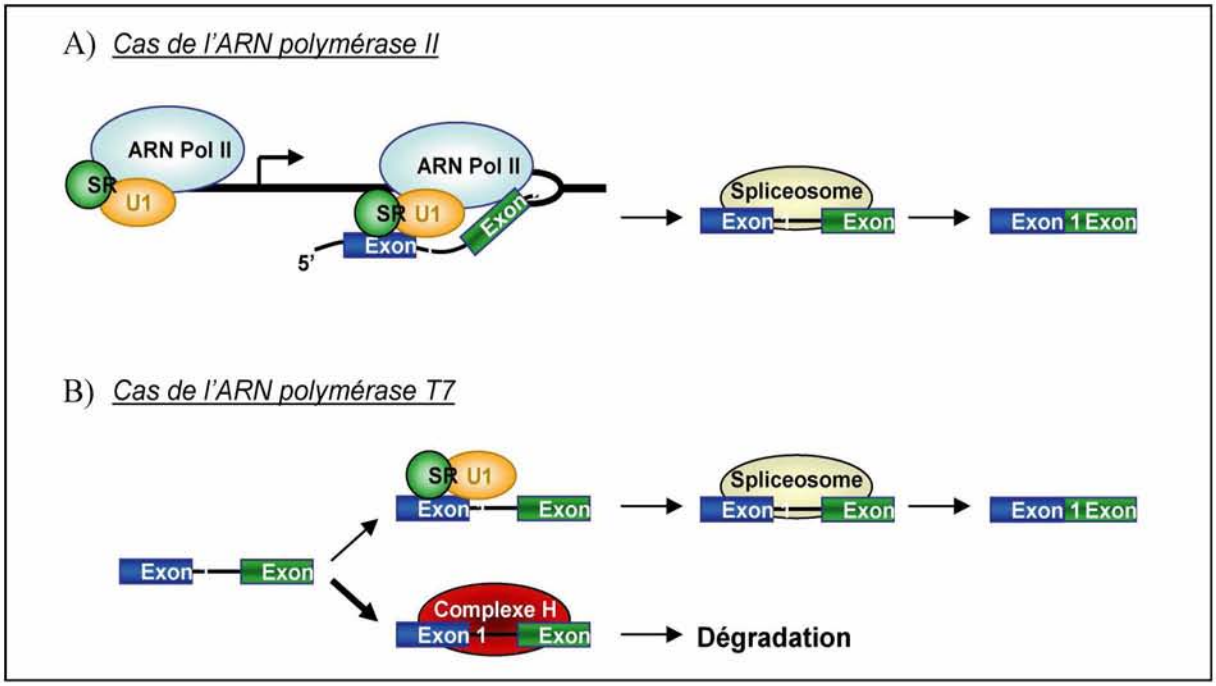


Fig. 39 : Modèle de recrutement co-transcriptionnel des protéines SR et de la snRNP U1 sur le transcrit naissant de l'ARN Polymérase II

A) Cas de l'épissage de transcrit Pol II. Lors de la transcription, les protéines SR et la snRNP U1 sont transférées de l'ARN Pol II au site 5' d'épissage, ce qui favorise l'assemblage du spliceosome et conduit à un épissage efficace. B) Cas de l'épissage de transcrit Pol T7. Les protéines SR et la snRNP U1 sont en compétition avec les protéines hnRNP pour se fixer sur le transcrit T7 nu. Il en résulte soit l'assemblage du spliceosome soit celui du complexe H, et une moindre efficacité d'épissage que le transcrit RNA Pol II.(d'après Das et al., 2007).

2.1.4. Rôle de la vitesse d'élongation de l'ARN polymérase II

Un autre modèle suggère que la transcription pourrait contrôler l'épissage du fait de la variation de processivité de l'ARN Pol II. Dans le cas d'une combinaison d'un site accepteur faible et d'un site accepteur fort en aval, une faible vitesse d'élongation ou des moments de pause pourraient dans certains cas favoriser l'inclusion d'exon par un mécanisme d'exon « skipping », alors qu'une élongation efficace entraînerait plutôt leur exclusion (la vitesse d'élongation de la transcription n'a pas d'influence sur des sites forts constitutifs ; Fig. 40)

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). Prolongeant le concept, la même équipe a pu vérifier que l'acétylation des histones, modification qui facilite le passage de la polymérase, favorise

logiquement l'exclusion de l'exon EDI (Nogues *et al.*, 2002). Puis le modèle a été validé en renforçant le caractère constitutif du site accepteur le plus en amont, les résultats obtenus montrant effectivement que plus l'exon interne est reconnu constitutivement, moins la vitesse d'élongation n'a d'effet sur le degré d'inclusion (Nogues *et al.*, 2003).

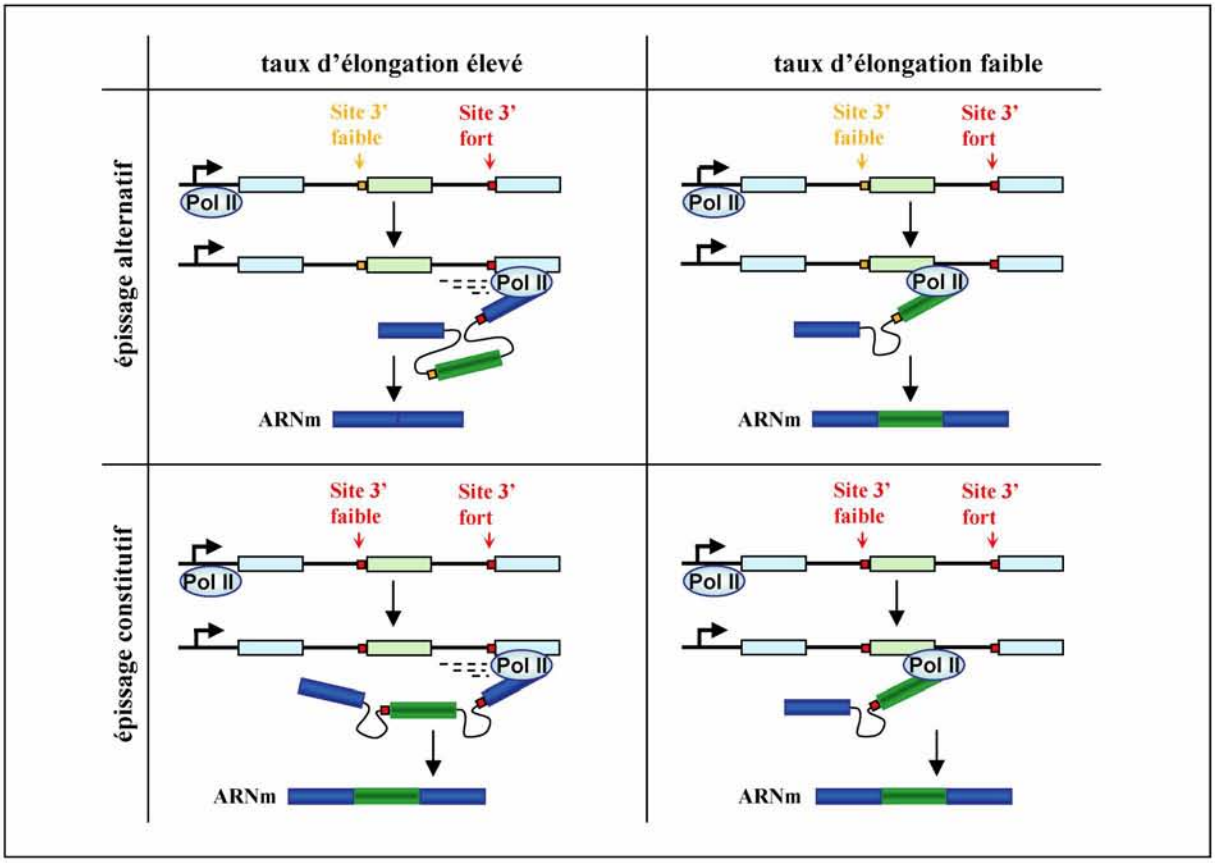


Fig. 40 : Influence de la vitesse d'élongation sur l'épissage alternatif

Dans le cas d'un épissage alternatif (en haut) avec un premier site 3' plus faible que celui de l'intron en aval : un taux d'élongation élevé de l'ARN Pol II favorise l'exclusion de l'exon alternatif (à gauche), alors qu'un taux faible d'élongation favorise l'inclusion (à droite). Dans le cas d'un épissage constitutif (en bas) : l'exon est constitutivement inclus quelque soit le taux d'élongation (d'après Kornblihtt, 2005).

Ce modèle établi dans un contexte hétérologue se vérifie dans le cas intéressant du complexe SWI/SNF (mating-type SWItch/Sucrose NonFermenting) et de la protéine Sam68, une protéine membre de la famille STAR (Signal Transduction Activation RNA). Initialement connue pour son rôle dans le remodelage de la chromatine au niveau des promoteurs, il a été démontré plus récemment que SWI/SNF pouvait être à l'origine d'une baisse du taux de phosphorylation du CTD de l'ARN pol II dans certaines régions codantes des gènes. Le

modèle suggéré propose que le complexe se déplace le long du gène et ce en étant associé à l'ARN Pol II. Lorsque la voie des MAP kinases est activée, la présence de Sam68 sur une séquence régulatrice proche d'un site d'épissage alternatif, pourrait servir de signal au complexe SWI/SNF qui induirait une modification du taux de phosphorylation du CTD de l'ARN pol II, ce qui conduirait à un ralentissement de l'élongation. Comme en plus SWI/SNF est capable d'interagir avec les snRNP U1 et U5, tous ces événements conduiraient à l'inclusion d'exons habituellement exclus (Batsche *et al.*, 2006).

Dans cet exemple, le couplage transcription/épissage se situe à la fois au niveau du recrutement des facteurs d'épissage par la machinerie de transcription et au niveau de l'importance de la vitesse d'élongation de la transcription par rapport à la reconnaissance des sites d'épissage (pour revue, de la Mata and Kornblihtt, 2006).

2.1.5. Influence de l'épissage sur la transcription

Nous venons de voir que l'épissage pouvait être favorisé par la machinerie de transcription, mais quelques données suggèrent qu'en retour la machinerie d'épissage favorise elle aussi la transcription.

- Cette notion a émergé à partir de données montrant que la présence d'intron, ou même simplement d'un site 5' d'épissage, à proximité du promoteur augmente significativement le taux de transcription chez les mammifères comme chez la levure (Furger *et al.*, 2002).

- En même temps, une autre équipe a montré que les UsnRNP interagissent avec le facteur d'élongation de la transcription TAT-SF1 et augmentent considérablement la processivité de l'ARN Pol II, probablement en permettant l'interaction entre P-TEFb et TAT-SF1 (Fong and Zhou, 2001).

- Il a aussi été proposé que le snRNA U1 s'associe au facteur TFIIH et que cette interaction stimule *in vitro* les étapes d'initiation et de ré-initiation de la transcription en plus son rôle sur l'épissage (Kwek *et al.*, 2002).

- Plus récemment enfin, il a été démontre que la présence du facteur $U2AF^{65}$ permet d'éviter le phénomène de pause de l'ARN Pol II engendrée par la synthèse de la séquence polypyrimidine (Ujvari and Luse, 2004).

- Et puis, il a aussi été démontré que le facteur d'épissage SKIP (c-SKi Interacting Protein), associé à P-TEFb est capable de réguler l'élongation de la transcription *in vitro* et *in vivo* (Bres *et al.*, 2005).

Bien que les mécanismes ne soient pas encore clairement établis, l'idée émerge donc que certains facteurs d'épissage associés au transcrit naissant, sont capables de réguler les différentes étapes de la transcription comprenant l'initiation, le passage en phase d'élongation, puis l'élongation (pour revue, Kornblihtt, 2005).

2.2. Lien entre export, NMD et traduction : rôle de l'EJC

Le complexe EJC[18] est un complexe protéique déposé 20 à 24 nt en amont de chaque jonction exon-exon après épissage (Le Hir *et al.*, 2001 ; Kataoka and Dreyfuss, 2004), et qui reste fixé sur l'ARNm jusqu'au passage des ribosomes (Dostie and Dreyfuss, 2002 ; Lejeune *et al.*, 2002). Son rôle est crucial pour l'étape d'export des ARNm, leur surveillance par le NMD[19], leur traduction et leur localisation (pour revue, Lejeune and Maquat, 2005).

2.2.1. Composants et assemblage du complexe EJC

Actuellement 16 protéines ont été identifiées dans les complexes EJC (Fig. 41 ; Ideue *et al.*, 2007). Parmi elles, on trouve :

- les protéines du domaine core : Y14, Magoh, MLN51 et eIF4AIII (Ballut *et al.*, 2005 ; Tange *et al.*, 2005)
- des co-activateurs d'épissage (par exemple RNPS1 (Mayeda *et al.*, 1999a)),
- des facteurs d'export (par exemple UAP56),
- et des facteurs du NMD : Upf3a/b, Upf2 et Upf1.

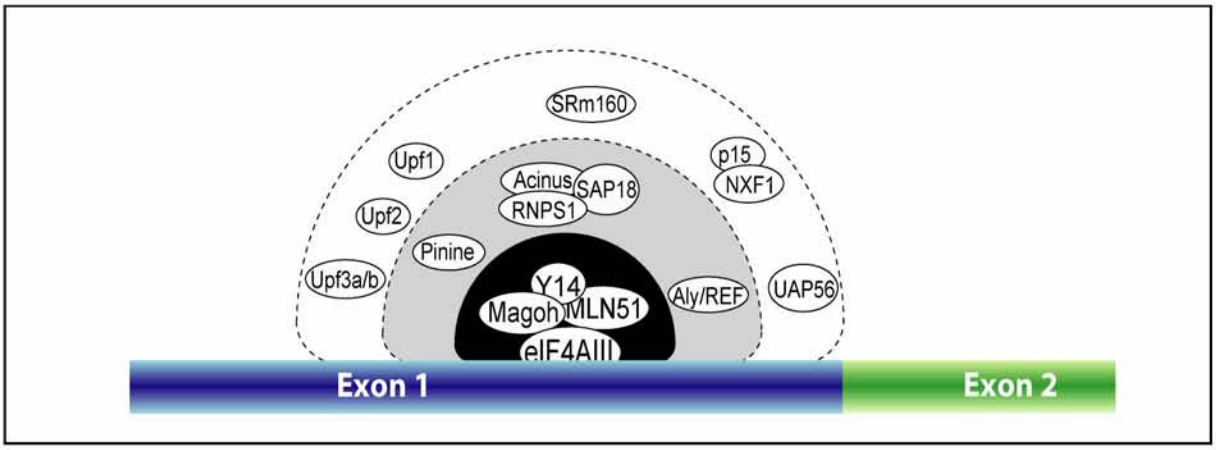


Fig. 41 : Représentation schématique du complexe EJC *(d'après Tange et al., 2005)*

[18] Exon Jonction Complex
[19] Nonsense-mediated mRNA Decay

Le complexe EJC ne s'assemble sur l'ARN messager que lorsqu'il y a épissage, mais son mode d'assemblage n'a pas encore été complètement élucidé. Le composant le mieux caractérisé est l'hétérodimère Y14-Magoh qui est associé aux complexes spliceosomaux lors de la première étape de la réaction d'épissage (Reichert *et al.*, 2002).

L'étude cristallographique récente du domaine « cœur » en interaction avec un ARN poly(U), révèle une pince à ARN dans laquelle les protéines eIF4AIII et MLN51 contactent 6 nt successifs de l'ARN (Bono *et al.*, 2006). En fait, seul le facteur eIF4AIII, qui possède une activité ARN hélicase, fixe directement l'ARN. L'hétérodimère Y14-Magoh permet lui de stabiliser l'interaction du complexe EJC avec l'ARNm en inhibant l'activité ATPase de ce dernier (Ballut *et al.*, 2005). Ce domaine « cœur » qui reste fixé à l'ARNm dans le cytoplasme, sert en fait de plate-forme aux autres facteurs de l'EJC qui s'associent de manière transitoire.

2.2.2. Rôle du complexe EJC dans l'export des ARNm dans le cytoplasme

Des études faites dans les oocytes de Xénope montrent que le complexe EJC augmente l'efficacité d'export des ARNm vers le cytoplasme (Zhou *et al.*, 2000 ; Le Hir *et al.*, 2001). Le facteur UAP56, initialement identifié comme un composant des complexes spliceosomaux, se retrouve aussi dans l'EJC et est essentiel pour l'export des ARNm. En effet, UAP56 et le facteur Aly/REF associé, permettent le recrutement de l'hétérodimère TAP/NXF1:p15 qui assure l'export de la majorité des ARNm par son interaction avec les composants du complexe du pore nucléaire (ou NPC[20]) (Zhou *et al.*, 2000 ; Luo *et al.*, 2001).

2.2.3. Complexe EJC et système NMD

Le NMD est un mécanisme post-transcriptionnel utilisé par les cellules eucaryotes pour dégrader les ARNm présentant des codons stops prématurés. Chez les mammifères, un codon stop est considéré comme prématuré lorsqu'une jonction exon-exon est localisée plus de 50 à 55 nt en aval de ce codon. Cela implique que (1) seuls les ARNm nouvellement synthétisés sont des cibles du NMD, et (2) les transcrits des gènes sans intron (environ 5% des gènes de mammifères) et les transcrits qui comportent des codons stop prématurés à une distance inférieure à 50-55 nt de la jonction exon-exon ne subissent pas le NMD. Il faut noter aussi que certains ARNm cellulaires sont naturellement des cibles du NMD, c'est le cas notamment de

[20] Nuclear Pore Complex

l'ARNm codant l'isoforme c de la sous-unité U2AF[35]. Dans ces cas là, le NMD devient un paramètre de la régulation de l'expression du gène (pour revue, Lejeune and Maquat, 2005).
Ainsi, le complexe EJC est donc un signal permettant la reconnaissance des codons stop prématurés sur l'ARNm (Wagner and Lykke-Andersen, 2002). Mais son rôle ne s'arrête pas là puisqu'il participe aussi à l'assemblage du NMD, grâce à 2 protéines (RNPS1 et Y14) en interactions avec les protéines Upf du NMD (Lykke-Andersen *et al.*, 2001).

2.2.4. Rôle du complexe EJC dans la traduction des ARNm

Le complexe EJC pourrait également jouer un rôle dans l'efficacité de traduction de l'ARNm. En effet, les ARNm ayant subit un épissage conduisent à une plus grande production de protéines par molécule d'ARNm que les mêmes molécules d'ARNm produites sans épissage (Lu and Cullen, 2003 ; Nott *et al.*, 2004). Si la taille du 1er exon est raccourcie de telle sorte qu'elle ne puisse plus recruter le complexe EJC, la stimulation de la traduction est abolie (Wiegand *et al.*, 2003 ; Nott *et al.*, 2004). Les facteurs RNPS1, Y14 et Magoh peuvent augmenter le taux de traduction des ARNm lorsqu'ils sont attachés individuellement à un ARNm rapporteur. L'analyse des polysomes cytoplasmiques a permis de montrer que la présence d'un complexe EJC augmenterait le recrutement des ARNm par les ribosomes par un mécanisme encore inconnu (Nott *et al.*, 2004). Mais, on ne sait pas si ce rôle est indépendant de l'effet positif des protéines SR sur la traduction.

3. L'épissage alternatif

3.1. Les modes d'épissage alternatif

Différents modes d'épissage alternatif existent (Fig. 42):

Les exons dits « cassette »
Les exons « cassettes » sont en général de petite taille, tissu-spécifique, et codent un domaine supplémentaire permettant d'obtenir une isoforme de la protéine.
Par exemple, l'exon N1 de l'ARN pré-messager du proto-oncogène c-src, codant une tyrosine kinase, est un exon cassette. Dans les neurones, son inclusion conduit à l'insertion de 6 acides aminés dans le domaine SH3 de la protéine kinase, ce qui modifie sa capacité d'interaction avec d'autres protéines (pour revue, Brown and Cooper, 1996).

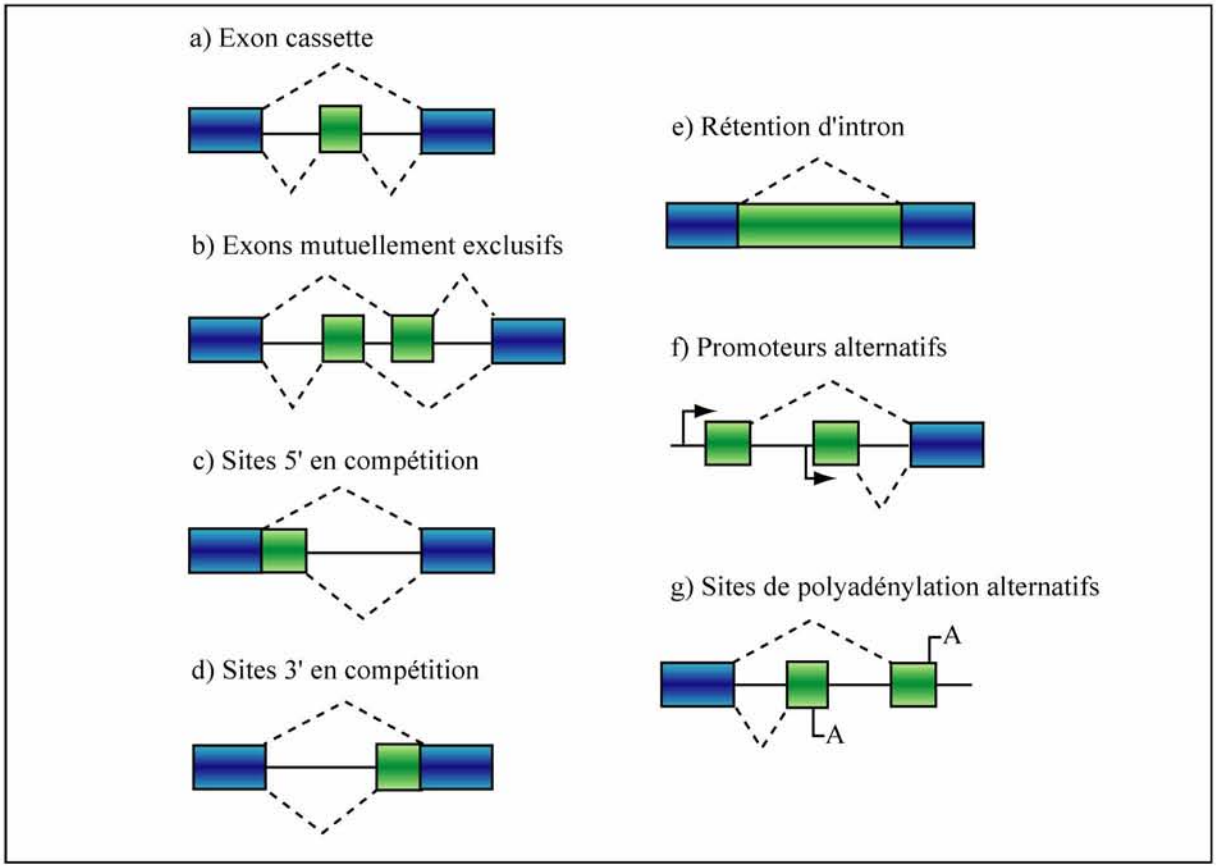


Fig. 42 : Les différents modes d'épissage alternatif

Les exons épissés constitutivement sont représentés par des rectangles bleus, ceux épissés alternativement sont représentés par des rectangles roses. Les flèches désignent les promoteurs alternatifs et les A, les signaux de polyadénylation (adapté de Black, 2003)

Les exons dits mutuellement exclusifs

Bien qu'ils aient des sites 5' et 3' fonctionnels, la caractéristique de ces exons est qu'ils ne sont jamais présents ensemble dans les ARNm matures.

Un cas concret est celui du pré-ARNm *FGFR-2* murin qui contient 2 exons IIIb et IIIc mutuellement exclusifs. Ceci conduit soit à la forme IIIb du récepteur FGFR-2 capable de fixer les facteurs de croissance FGF7 et 10, soit à la forme FGFR-2 IIIc qui ne peut le faire (Miki *et al.*, 1992).

Une forme particulière de ce type d'épissage alternatif est apparue très récemment avec la découverte des sites d'épissage doubles comprenant la partie 3' du consensus d'un site 5' et la partie 5' du consensus d'un site 3', c'est-à-dire une séquence consensus CAG/GU (Fig. 43 ; Zhang *et al.*, 2007). L'étude bio-informatique des événements d'épissage alternatif dans les génomes de l'homme et de la souris, suivie de la validation expérimentale a déjà permis de

détecter une cinquantaine de ces sites à travers les deux génomes. Ce nouveau type de site capable d'être à la fois identifié comme un site donneur d'épissage et un site accepteur soulève des questions importantes au niveau des mécanismes de définition des sites d'épissage, de leur régulation et de leur compétition.

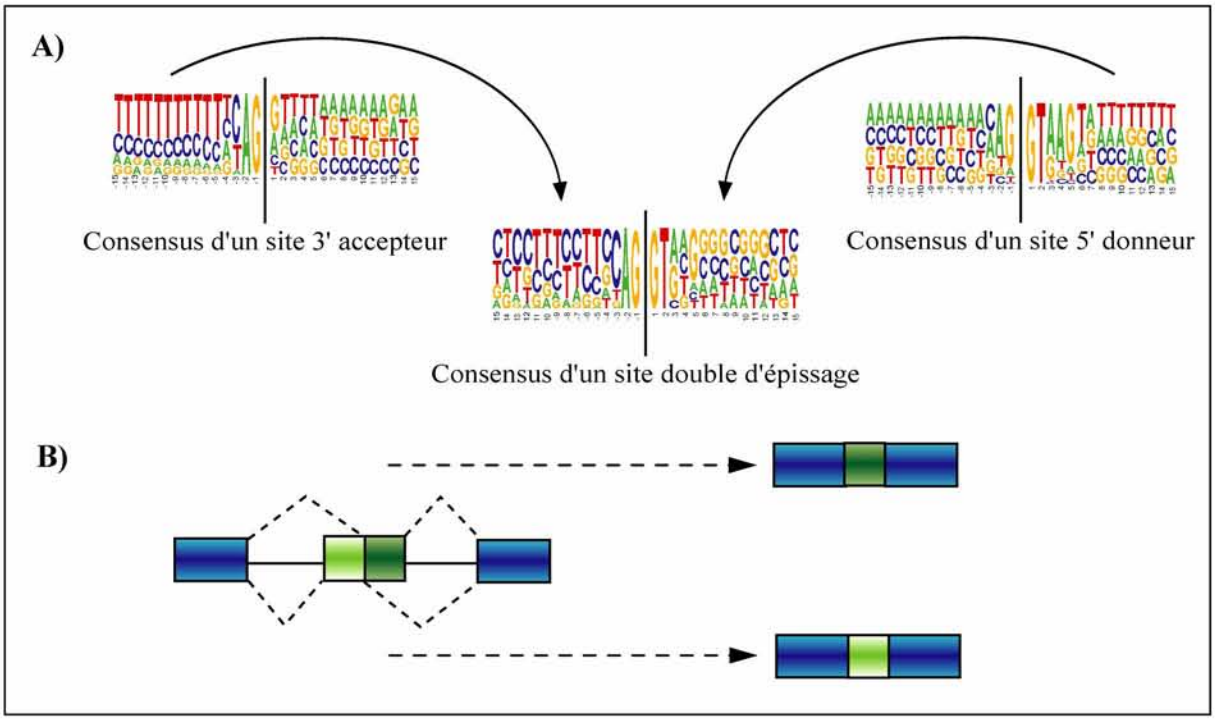


Fig. 43 : Cas du site double d'épissage

A) Séquence consensus d'un site double d'épissage. Le consensus d'un site d'épissage double correspond à la séquence 5' d'un site 3' accepteur associée à la séquence 3' d'un site 5' donneur d'épissage. B) Epissage alternatif en présence d'un site double d'épissage. La composition en exon de l'ARN messager varie selon que le site double d'épissage est utilisé comme site donneur d'épissage (en haut) ou comme site accepteur d'épissage (en bas) (adapté de Zhang et al., 2007).

Les exons présentant plusieurs sites donneur ou accepteur en compétition

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 pour leur permettre 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 donc la production de trois ARNm 9S, 12S et 13S (pour revue, Akusjarvi and Stevenin, 2003).

Les exons résultant d'une rétention d'intron

Suivant les conditions d'épissage, l'intron peut être inclus dans l'ARN messager. Dans ce cas, l'épissage est souvent tissu-spécifique.

Chez la drosophile par exemple, l'épissage du 3^{e} 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).

Les exons terminaux 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, tandis que dans les neurones, le neuropeptide CGRP est produit par utilisation du site de polyadénylation de l'exon 6 (pour revue, Lou and Gagel, 1998).

Plusieurs mécanismes d'épissage alternatif peuvent aussi être combinés dans un même pré-ARNm. La production d'ARNm est alors régulée selon la nature des sites d'épissage et la présence ou non de séquences régulatrices. Le choix des sites d'épissage peut aussi 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.

Il apparaît donc que l'épissage alternatif est, pour la cellule, un excellent moyen de contrôle de l'expression des gènes au cours du développement et de la différenciation cellulaire. Nous allons voir que cette régulation se caractérise par la présence de sites d'épissage sous-optimaux associés à des éléments de séquences agissant en *cis*.

Enfin, il faut noter que, pour les organismes à petit génome, les virus notamment, l'utilisation de l'épissage alternatif est aussi un moyen efficace d'accroître la capacité codante du génome.

3.2. Les sites d'épissage alternatif

Caractéristiques des 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 sont dits sous-optimaux car ils possèdent des séquences plus dégénérées. Un enchaînement de pyrimidines entrecoupé de purines et/ou une séquence de

branchement non canonique, avec parfois même un point de branchement autre que l'adénosine classique, sont aussi des éléments qui affaiblissent les sites 3' d'épissage. Toutes ces caractéristiques rendent moins efficace l'interaction des composants spliceosomaux avec les jonctions exon/intron et intron/exon.

Les éléments de séquences agissant en cis

Lorsque les séquences des sites d'épissage sont très dégénérées, l'assemblage des complexes spliceosomaux est guidé et facilité par des facteurs protéiques qui se fixent à des éléments de séquences agissant en *cis*. Ces séquences activatrices d'épissage peuvent être exoniques, ce sont alors des ESE[21], ou introniques, on parle alors d'ISE[22]. A l'inverse, il existe aussi des sites d'épissage régulés dont les séquences sont peu dégénérées mais dont l'activité est diminuée par la présence de séquences inhibitrices exoniques, appelées ESS[23], ou introniques, appelées ISS[24].

Ces éléments régulateurs agissent par l'intermédiaire de leurs protéines partenaires. De manière générale, les protéines hnRNP sont associées aux éléments répresseurs, tandis que les protéines SR sont associées aux éléments activateurs, mais nous verrons par la suite qu'il existe quelques exemples contraires.

En fait, ces éléments agissant en *cis* ne sont pas spécifiquement associés aux sites d'épissage alternatif et les séquences activatrices et inhibitrices se distribuent différemment dans les exons et les introns. Ainsi, les introns contiennent en majorité des séquences inhibitrices, les exons renferment une plus forte proportion de séquences activatrices, et ceux éliminés de manière alternative possèdent à la fois des séquences activatrices et des séquences inhibitrices (pour revue, Blencowe, 2006).

4. Les régulateurs de l'épissage

Nous décrirons plus particulièrement l'action des protéines SR et des protéines hnRNP, car ce sont les seuls facteurs décrits jusqu'à présent comme régulateurs de l'épissage de l'ARN du virus HIV-1. Puis nous verrons que d'autres paramètres non protéiques peuvent moduler l'épissage des pré-ARNm.

[21] Exonic Splicing Enhancer
[22] Intronic Splicing Enhancer
[23] Exonic Splicing Silencer
[24] Intronic Splicing Silencer

4.1. Les protéines SR

4.1.1. La famille des protéines SR

Les caractéristiques principales des protéines SR sont (1) leur conservation phylogénétique, (2) leur double fonction à la fois comme facteurs généraux de l'épissage et comme régulateurs de l'épissage alternatif et (3) leur ressemblance structurale (pour revue, Graveley, 2000). En effet, toutes sont composées d'un domaine de liaison à l'ARN (ou RBD[25]) en N terminal, suivi d'un domaine riche en dipeptides arginine-serine, appelé domaine RS en C terminal (Fig. 44 ; Birney *et al.*, 1993).

Fig. 44 : Organisation structurales des protéines SR

RRM: motif de reconnaissance de l'ARN, RRM H: domaine RRM homologue, Z: domaine doigt de zinc, RS: domaine riche en dipeptide arginine/sérine (pour revue, Graveley, 2000).

Le domaine RBD

Les différences de composition des RBD des protéines SR ont conduit à distinguer deux sous-groupes (Mayeda *et al.*, 1999b). Le premier regroupe les protéines ayant un motif RRM[26] classique et un second plus atypique et incluant un heptapeptide caractéristique (ASF/SF2 et SRp40 par exemple). L'autre groupe ne contient que les protéines à un seul domaine RRM (SC35 et 9G8 entre autres).

Afin de comprendre les fonctions des protéines SR, plusieurs études ont tenté d'élucider les spécificités de liaison de chaque protéine SR. Le résultat global de ces expériences, que nous détaillerons plus loin, est que 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 donné est également reconnu par d'autres protéines SR (Liu *et al.*, 1998 ; Tacke and Manley, 1995).

[25] RNA Binding Domain
[26] RNA Recognition Motif

Le domaine RS

Le domaine RS est considéré pour la plupart comme le module d'interactions protéine-protéine par lequel les protéines SR s'associent entre elles ou avec les différents composants de la machinerie d'épissage contenant aussi un domaine RS (Wu and Maniatis, 1993). Néanmoins, il est apparu récemment qu'il pourrait aussi contacter l'ARN (Shen *et al.*, 2004). Le domaine RS contient généralement le signal de localisation de la protéine, mais celui–ci peut parfois s'étendre jusque dans le domaine RBD. Enfin, ce domaine est sujet à de nombreux événements de phosphorylation qui permettent de réguler l'activité des protéines, mais aussi parfois leur localisation (Sanford and Bruzik, 2001).

Le domaine RS phosphorylé

Nombre d'études montre que les résidus sérine des domaines RS des protéines SR sont largement phosphorylés. Ce niveau de phosphorylation influence la localisation cellulaire des protéines SR (Gui *et al.*, 1994a ; Colwill *et al.*, 1996), aussi bien que leur fonction dans l'épissage (Wang *et al.*, 1998 ; Xiao and Manley, 1997). Les premières études avaient permis de démontrer que des protéines SR hypo- ou hyper-phosphorylées étaient incapables de stimuler l'épissage *in vitro* (Prasad *et al.*, 1999 ; Sanford and Bruzik, 1999). Plus récemment, deux voies de transduction conduisant à la phosphorylation des protéines ASF/SF2 et 9G8 d'une part, et SRp40 d'autre part, ont été mises en évidence (Blaustein *et al.*, 2005 ; Patel *et al.*, 2005). Dans les deux cas, la phosphorylation de ces protéines SR entraîne l'inclusion des exons qu'elles régulent.

Trois types de protéines kinase ont été identifiée comme pouvant phosphoryler le domaine RS *in vitro*. Il y a la famille des SRPK[27] (SRPK 1 ((Gui *et al.*, 1994b) et SRPK 2 (Wang *et al.*, 1998)), la famille Clk/Sty (Colwill *et al.*, 1996) et les DNA topoisomerase I (Rossi *et al.*, 1996).

Le signal de localisation nucléaire

Les protéines SR possèdent un signal de localisation nucléaire qui leur permet d'interagir avec un récepteur d'import nucléaire spécifique, appelé « SR transportin » (Caceres *et al.*, 1997 ; Kataoka *et al.*, 1999), et d'aller ensuite se localiser dans les « speckles » (pour revue, Lamond and Spector, 2003). Ce signal de localisation nucléaire se situe en général dans le domaine RS de la protéine, mais il a été démontré que ce n'est pas le cas tout le temps. Par exemple, le

[27] SR Protein Kinase

domaine RS d'ASF/SF2 n'est pas suffisant pour permettre une localisation nucléaire, un des deux domaines RRM est nécessaire (Caceres *et al.*, 1997).

Les « speckles » sont à la fois un lieu de stockage des composants de la machinerie d'épissage (Fu RNA 1995), mais aussi l'endroit où est gérer la disponibilité de ces facteurs (Misteli *et al.*, 1997). Ce système fonctionne par l'intermédiaire de kinases (Clk/Sty (Colwill *et al.*, 1996) ; PRP4 (Kojima *et al.*, 2001) ; PSKH1 (Brede *et al.*, 2002)) et de phosphatases (PP1, Trinkle-Mulcahy *et al.*, 2001) directement présentes dans les « speckles ».

Ainsi, il apparaît que les protéines SR subissent le long de leur existence plusieurs évènements de phosphorylation et déphosphorylation qui gouvernent leur localisation cellulaire. Un des exemples maintenant bien connu est celui de la protéine ASF/SF2, dont le domaine RS est tout d'abord phosphorylé par la SRPK1, ce qui lui permet d'aller se localiser au niveau des « speckles » (Aubol *et al.*, 2003). Puis la kinase Clk/Sty phosphoryle à nouveau ASF/SF2 pour lui permettre de quitter les « speckles » et migrer vers les zones d'épissage (Cao *et al.*, 1997 ; Misteli *et al.*, 1998). Nous verrons plus tard que d'autres événements de phosphorylation permettent à ASF/SF2 de participer à l'export cytoplasmique des ARNm.

Les protéines SR-like

Plusieurs autres facteurs, distincts des protéines SR et impliqués dans différents aspects du métabolisme de l'ARN pré-messager, possèdent aussi un domaine RS (Fig. 45). Ces protéines, classées dans la famille des protéines SR-like (ou SRrp[28]), regroupent par exemple les deux sous unités du facteur U2AF (U2AF^{35}, (Zhang *et al.*, 1992) ; U2AF^{35}, (Zamore *et al.*, 1992), des composants des snRNP (U1-70K (Query *et al.*, 1989b), U5-100K (Teigelkamp *et al.*, 1997), U4/U6.U5-27K (Fetzer *et al.*, 1997), des co-activateurs d'épissage SRm 160 et 300 (Blencowe *et al.*, 1998 ; Blencowe *et al.*, 2000), des régulateurs de l'épissage (Tra (Boggs *et al.*, 1987), Tra2 (Amrein *et al.*, 1988)) ou encore des ARN hélicase (hPrp16 (Zhou and Reed, 1998)). Cette liste a été encore rallongée par une étude du génome qui a révélée un large spectre de protéines SR-like qui influencent l'épissage mais aussi le remodelage de la chromatine, la transcription par l'ARN Pol II ou encore la progression du cycle cellulaire (pour revue, Boucher *et al.*, 2001).

[28] SR protein-related protein

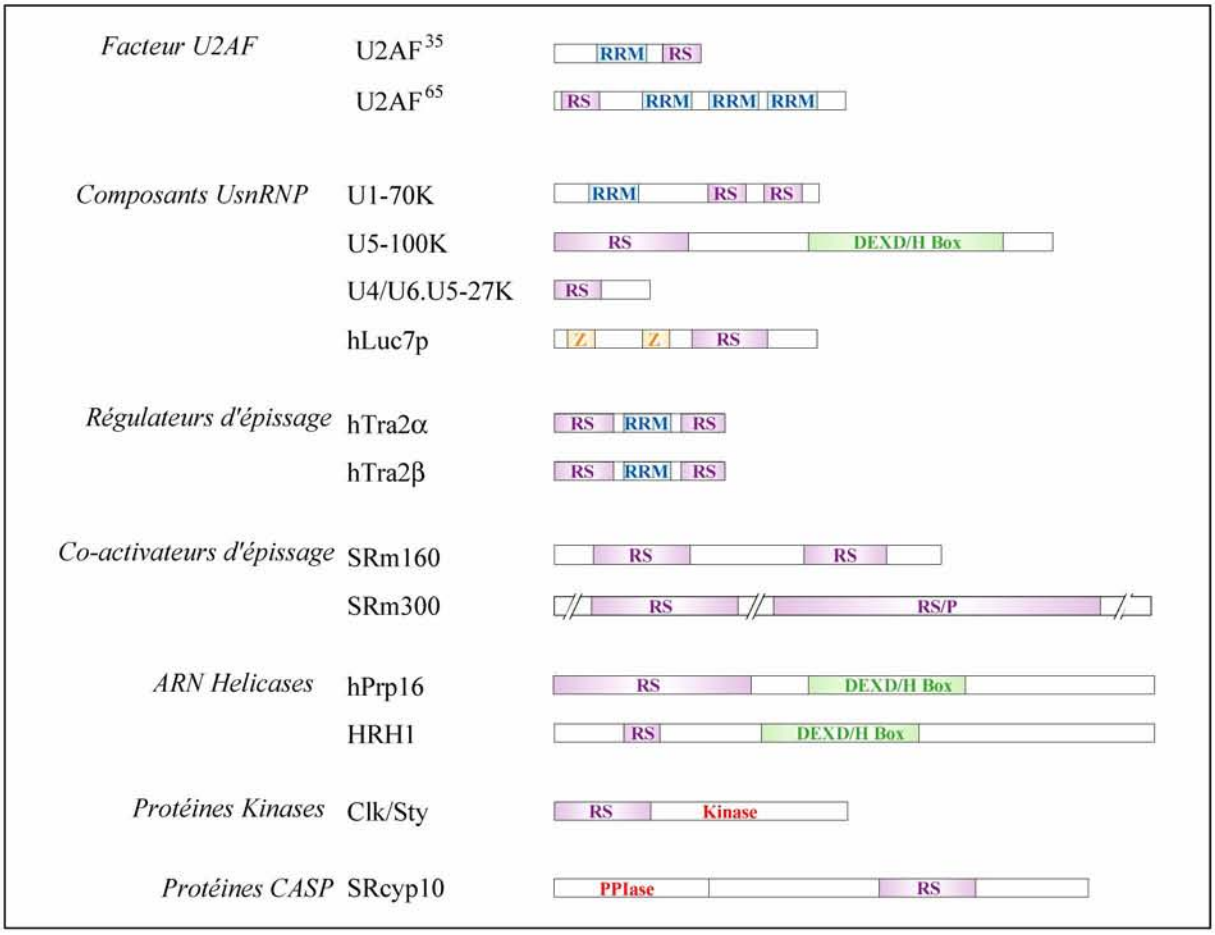


Fig. 45 : Organisation structurale des protéines SR-like

RRM: motif de reconnaissance de l'ARN, Z: domaine doigt de zinc, RS: domaine riche en dipeptide arginine/sérine, DEXD/H Box: motif caractéristique des hélicases à ARN (pour revue, Graveley, 2000).

4.1.2. Les séquences reconnues par les protéines SR

Protéines SR et séquences ESE sont généralement associées car les protéines SR ayant une influence majoritairement positive sur l'épissage et se fixant essentiellement dans les exons, leurs sites de fixation deviennent par définition des ESE. Les premières séquences activatrices d'épissage étaient initialement décrites comme riches en résidus purines, puis l'avènement de la technique SELEX[29] a permis de mieux préciser la séquence reconnue.

4.1.2.1. Séquences identifiées par la méthode SELEX

Plusieurs méthodes de type SELEX ont été utilisées pour déterminer les séquences consensus sur lesquelles se lient les différentes protéines SR.

[29] Systematic Evolution of Ligands through EXponential enrichment

SELEX in vitro basé sur l'affinité de l'ARN

Cette approche consiste à sélectionner parmi un ensemble de séquences aléatoires, celles qui ont la plus haute affinité pour une protéine SR donnée. Après plusieurs cycles de sélection, les ARN capables de former des complexes ARN/protéine sont isolés puis séquencés. Les séquences sélectionnées peuvent ensuite être introduites dans un pré-ARNm contenant au moins un intron afin de tester leur capacité à induire l'épissage *in vitro* en présence d'un extrait cytoplasmique S100 complémenté par la protéine SR. Cette approche SELEX a permis de définir les séquences consensus de plusieurs protéines SR (SRp20, 9G8, ASF/SF2, SC35, SRp40, SRp55) et 2 protéines humaines SR-like (Tra2 ou SRp38) (Tableau 3) (pour revue, Bourgeois *et al.*, 2004).

SELEX fonctionnel par épissage in vitro ou in vivo

Ces approches SELEX visent à identifier directement les séquences ayant des propriétés d'ESE. La stratégie alors utilisée consiste à épisser *in vitro* ou *in vivo*, une mini-construction contenant deux ou trois exons, issue d'un pré-ARNm naturel, et dont l'élément ESE connu est remplacé par la séquence dégénérée. Ces études ont permis de sélectionner soit des séquences activatrices d'épissage (ESE), lorsque les expériences d'épissage *in vitro* étaient réalisées dans un extrait nucléaire, soit des séquences permettant spécifiquement l'activation de l'épissage dans un extrait cytoplasmique S100 en présence d'une protéine SR donnée (Tableau 3) (pour revue, Tacke and Manley, 1999). Une partie de ces données a permis de créer l'outil informatique ESEfinder que nous présenterons plus tard.

Les protéines SR reconnaissent des séquences dégénérées

Les séquences reconnues par les différentes protéines SR sont nombreuses et souvent dégénérées. Par exemple, la protéine SC35 est capable de se lier à des séquences qui peuvent être aussi bien riches en purines qu'en pyrimidines. De plus, les séquences identifiées pour une protéine donnée peuvent également être reconnues par d'autres protéines SR (c'est le cas pour les sites d'ASF/SF2 et de Tra2). Récemment est apparu l'idée que le RBD des protéines SR, tout du moins de certaines, servirait à scanner l'ARN jusqu'au signaux d'épissage où les protéines SR et autres composants du spliceosome se fixeraient de manière coopérative (Shen and Green, 2007).

Protéine SR	SELEX de liaison	SELEX fonctionnel	Protéine SR	SELEX de liaison	SELEX fonctionnel
SRp20	WCAKCUUUA (RBP1) AUCCNNA (RBP1) WCAUCGAY (9G8-Zn^m) YWCUUCAU (9G8-Zn^m) CUWCAAC (9G8-Zn^m)	CCUCGUCC (EN)	SRp40	UGGGAGCRGUYRGCUCGY	ACDGS (S100) YRCRKM (S100)
			SRp55	GRUCAACCNGGCGACMG	USCGKM (S100) YYWCWSG (S100)
9G8	AGACKACGAY ACGAGAGAY	GGACGACGA (EN)	Tra2-α Tra2-β	poly-A AAGAAGAA poly-A	
ASF/SF2	RGAAGAAC AGGACAGAGC	SRSASGA (S100) CRSMSGW (S100)	SRp38	ACAAAGACAAA	
SC35	AGSAGAGUA GUUCGAGUA UGUUCSAGWU GWUWCCUGCUA GGGUAUGCUG GAGCAGUAGKS AGGAGAU	UGCNGYY (EN) GRYYCSYR (S100) GRYYMCYR (S100)	dRSF1	CAACGACGA AAACGCGCG	

Tableau 3 : Séquences identifiées par la méthode SELEX comme étant reconnues par les protéines SR et autres régulateurs d'épissage contenant un domaine RS *(pour revue, Bourgeois et al., 2004)*

Les séquences SELEX sont divisées en deux colonnes selon la méthode utilisée pour les identifier (SELEX de liaison ou fonctionnel) (Heinrichs and Baker, 1995 **;** *Schaal and Maniatis, 1999* **;** *Cavaloc et al., 1999* **;** *Tacke and Manley, 1995* **;** *Tacke and Manley, 1999* **;** *Liu et al., 1998* **;** *Cartegni and Krainer, 2002* **;** *Liu et al., 2000* **;** *Tacke et al., 1997* **;** *Shin and Manley, 2002* **;** *Labourier et al., 1999). Pour les séquences isolées par expériences de SELEX fonctionnel, il est également indiqué si les expériences ont été réalisées en extrait nucléaire (EN) ou avec un extrait cytoplasmique S100 (S100).*

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

4.1.2.2. Les approches bioinformatiques permettant la recherche des ESE

Deux programmes bioinformatiques ont été développés afin de rechercher pour chaque gène, les motifs ESE putatifs.

Le programme « ESEfinder » repose sur l'utilisation d'une collection de motifs ESE, obtenus par SELEX fonctionnel, et permettant l'activation par les protéines ASF/SF2, SC35, SRp40 ou SRp55. Il permet de prédire des ESE putatifs au sein des pré-ARNm avec l'identité des protéines qui pourraient être les activateurs (Cartegni *et al.*, 2003).

L'autre approche informatique appelée « RESCUE[30] » repose sur 10 groupes différents de motifs ESE putatifs. Ce programme recherche des séquences (hexa ou octamer) surreprésentées (1) aux abords des sites d'épissage, (2) dans les exons par rapport aux introns, (3) dans les exons non codant par rapport au pseudo exons non épissés, et (4) dans les régions 5' non traduites des gènes ne possédant pas d'introns (Fairbrother *et al.*, 2002 ; Fairbrother *et al.*, 2004).

Bien que ces deux programmes permettent une prédiction des ESE, les résultats obtenus par ces approches nécessitent obligatoirement des validations expérimentales. En effet, un score de haute affinité pour une protéine n'est pas forcément corrélé à sa liaison au pré-ARNm et encore moins à un effet activateur. Et inversement les protéines SR peuvent se lier à des séquences ayant un score faible.

4.1.3. Les rôles des protéines SR

Comme on l'a vu précédemment, plusieurs protéines SR interviennent dans la formation des différents complexes de la réaction d'épissage, ce qui en fait à la fois des facteurs généraux de l'épissage, mais aussi des régulateurs de l'épissage alternatif. Leurs diverses influences tout au long de l'assemblage du spliceosome ont tout d'abord été expliquées par leur capacité à former des réseaux d'interactions protéine-protéine par l'intermédiaire des domaines RS. Mais récemment est apparu l'idée que les domaines RS sont aussi capables d'interagir avec l'ARN pour définir les différents signaux d'épissage et faciliter leur reconnaissance par les composants du spliceosome.

4.1.3.1. Rôles donnés aux interactions domaine RS-domaine RS

- Définition de l'exon lors de la formation du complexe E

Un des modèles permettant d'expliquer l'effet activateur des protéines SR sur l'épissage est basé sur la capacité des protéines SR à créer un réseau d'interaction protéine-protéine joignant les extrémités de l'exon, et capable d'interagir avec les composants de la machinerie d'épissage responsable de la définition des exons (Fig. 46A). Ainsi, il a pu être démontré qu'ASF et SC35 facilitent les associations de la snRNPU1 au site 5' d'épissage, et de U2AF[35] au site 3' d'épissage (Wu and Maniatis, 1993) et celle de U2AF[65] à la séquence polypyrimidine (Zuo and Maniatis, 1996). Il a ensuite été proposé que la stabilisation de la

[30] Relative Enhancer and Silencer Classification by Unanimous Enrichment

snRNP U1 soit assurée grâce à une interaction entre les domaines RS des protéines ASF/SF2 et U1 70K (Kohtz *et al.*, 1994).

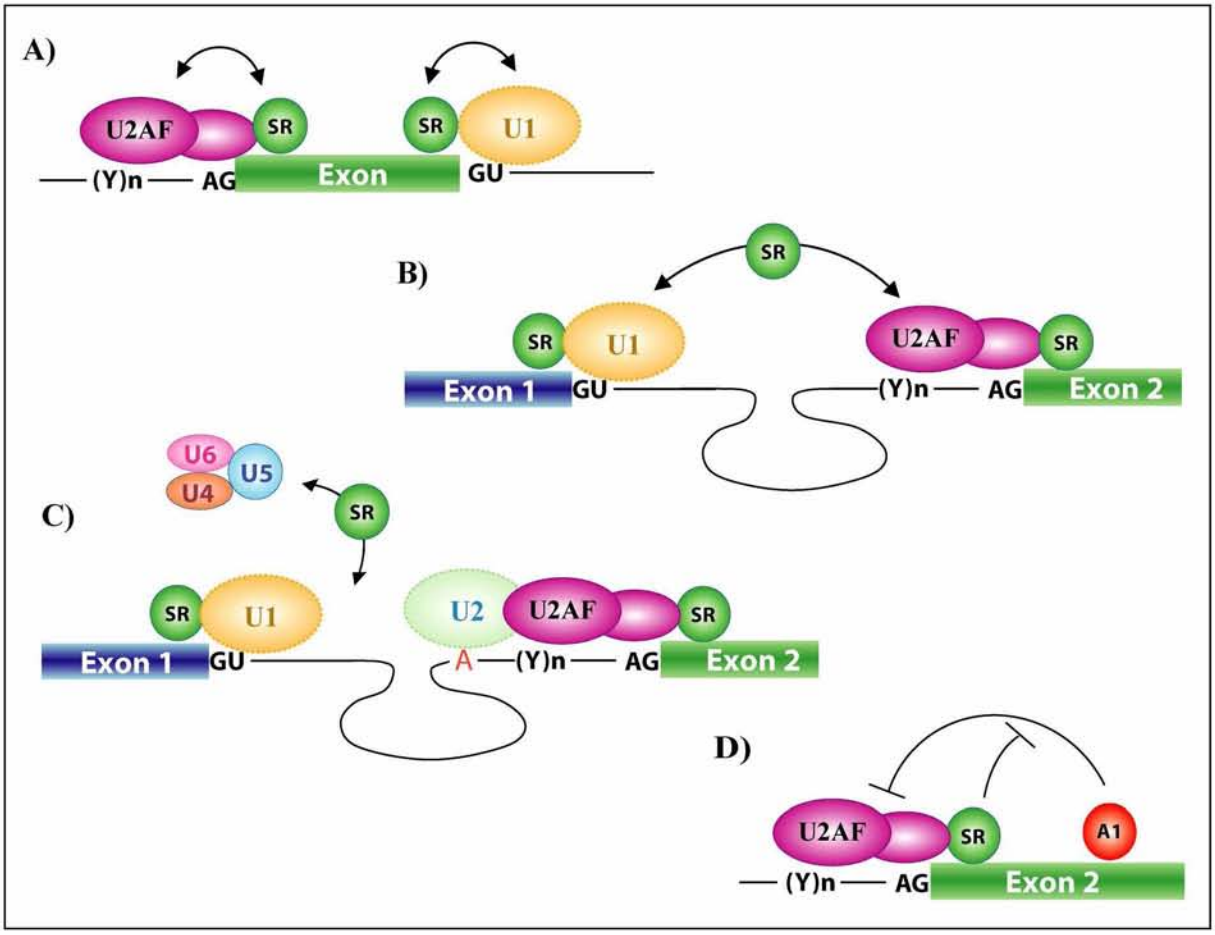


Fig. 46 : Rôle des protéines SR dans l'assemblage du spliceosome *(d'après Sanford et al., 2005)*

A) Définition des exons. Les protéines SR favorisent la reconnaissance des sites 5' et 3' d'épissage par la snRNP U1 et le facteur U2AF respectivement. B) Rapprochement des sites 5' et 3' d'épissage. Les interactions entre domaines RS des protéines SR, de la snRNP U1 et du facteur U2AF favorisent la juxtaposition des sites d'épissage. C) Recrutement de la tri-snRNP. Le même type d'interaction entre domaine RS favorise la formation des complexes B et C. D) Inhibition de l'effet des protéines hnRNP. En plus de favoriser l'utilisation des sites d'épissage, les protéines SR contrecarrent l'effet des protéines hnRNP qui sont majoritairement des répresseurs de l'épissage.

Un des exemples qui illustre le mieux ce rôle de définition des sites accepteur et donneur de l'exon est celui de l'élément bidirectionnel BSE de l'unité E1A de l'adénovirus. Le modèle proposé fait intervenir 3 protéines SR 9G8, SC35 et ASF et 1'élément ESE bipartite composé de 2 séquences riches en purines Pu1 et Pu2. Suivant l'identité de la protéine SR fixée,

l'élément BSE active le site 3' situé en amont et/ou le site 5' en aval. Ainsi la fixation de 9G8 sur l'élément Pu1 favorise la reconnaissance d'un site 5' donneur d'épissage de l'ARN, alors que la fixation d'ASF/SF2 et de SC35 sur la totalité de l'élément BSE favorise la reconnaissance d'un autre site 5' donneur d'épissage, mais aussi celle du site 3' accepteur d'épissage (Fig. 47 ; Bourgeois *et al.*, 1999).

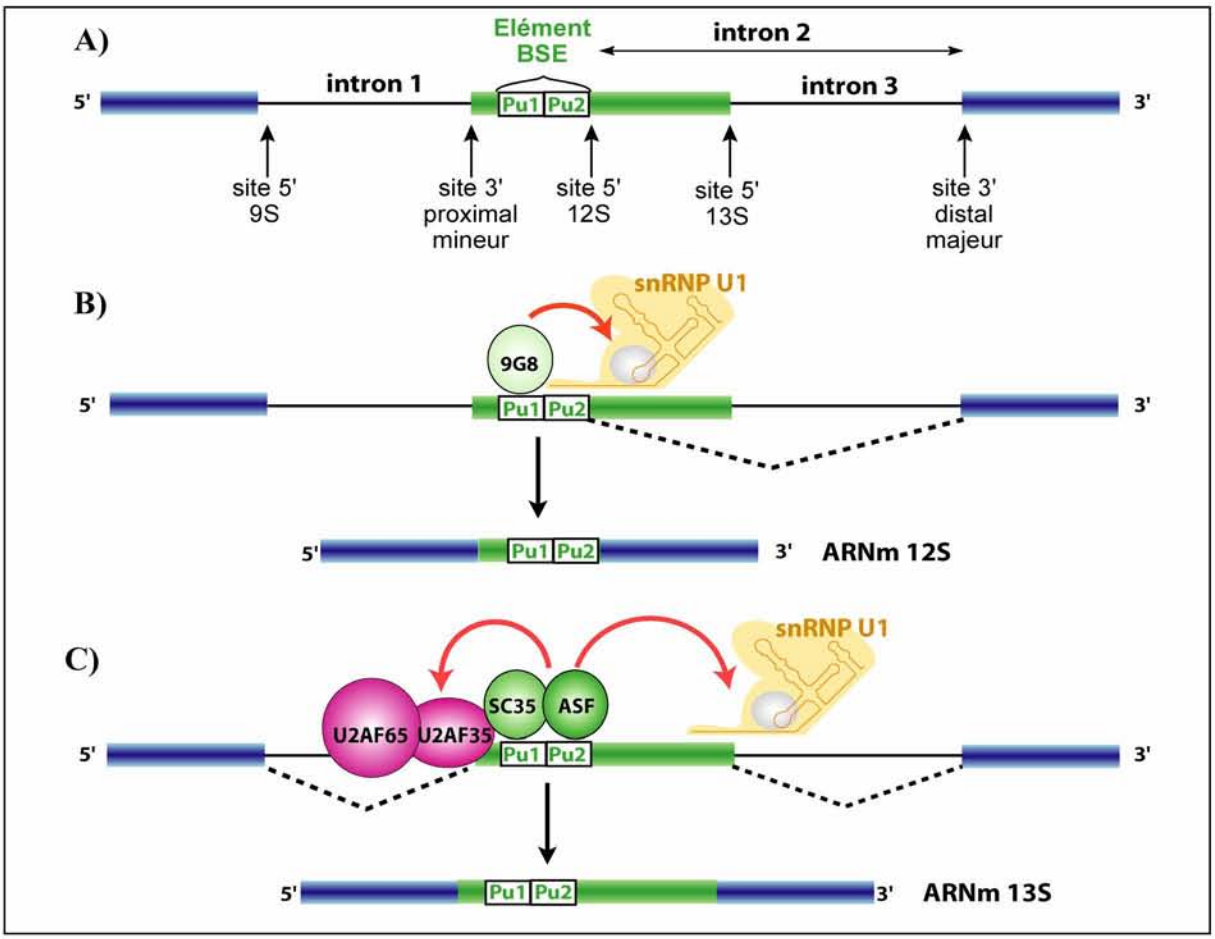


Fig. 47 : Cas de l'ESE bi-directionnelle du pré-ARNm de l'unité de transcription E1A *(d'après Bourgeois et al., 1999; pour revue, Akusjarvi and Stevenin, 2003)*

Un autre modèle proposé pour expliquer l'effet stimulant des protéines SR dans le processus de reconnaissance des exons, est celui qui implique les co-activateurs SRm160/300. Ces protéines possèdent plusieurs répétitions RS, mais pas de domaine RRM. Associé l'un à l'autre, ils activent à la fois l'épissage constitutif et l'épissage ESE-dépendant, probablement grâce à des interactions avec les snRNP U1 et U2 et les protéines SR. Il a été suggéré que les protéines SR favorisent aussi la définition des exons en favorisant le recrutement du complexe SRm160/300 (Fig. 46B ; Blencowe *et al.*, 2000 ; Eldridge *et al.*, 1999).

- *Aide aux rapprochement des sites 5' et 3' d'épissage lors de la formation du complexe E*

A partir des données précédentes, il a aussi été suggéré que le réseau d'interactions des protéines SR puisse favoriser la jonction des deux extrémités dans la dernière étape de formation du complexe E. Le rôle des protéines SR est cette fois-ci d'interagir à travers l'intron entre les sites 5' donneur et 3' accepteur (Fig. 46C).

Un argument en faveur de cette hypothèse était le fait que SC35 soit capable d'initier l'interaction entre les snRNP U1 et U2 (Fu and Maniatis, 1992). Un autre argument a été le fait que les protéines SR peuvent interagir simultanément avec U1 70K et U2AF^{35} (Wu and Maniatis, 1993). Puis, il a été démontré que les protéines SR sont impliquées dans les phénomènes d'épissage en *trans*, réaction dans laquelle les sites 5' et 3' d'épissage se situent sur deux ARN distincts (Bruzik and Maniatis, 1995). Récemment enfin, une étude utilisant des extraits cellulaires de nématode dont les protéines SR sont inactives au stade embryonnaire, puis actives plus tard dans le développement, a permis d'aller plus loin. Il a été démontré que les protéines SR sont essentielles à la fois pour relier les extrémités 5' et 3' d'un intron, mais aussi dans le mécanisme qui conduit la snRNP U1 à positionner la snRNP U2 sur le site de branchement (Boukis *et al.*, 2004). Pour autant, le mécanisme et le rôle exact des protéines SR restent à découvrir.

- *Recrutement de la tri-snRNP lors de la formation du complexe B*

Une autre participation des protéines SR au processus général d'épissage est celle qui favorise le recrutement de la tri-snRNP U4/U6.U5 au niveau du pré-spliceosome lors de la formation du complexe B (Fig. 46D).

Il avait déjà été prouvé qu'ASF, SC35 et SRp55 stimulent la formation du complexe B (Roscigno and Garcia-Blanco, 1995). Une étude plus récente a permis de révéler que les partenaires potentiels des protéines SR lors de cette étape pourraient être deux protéines SR-like de 65 et 110 kDa associées à la tri-snRNP et essentielles à son recrutement (Makarova *et al.*, 2001).

Ces deux derniers mécanismes d'action des protéines SR exon indépendant, ont servi de base de réflexion pour tenter d'expliquer les résultats convergents de quelques publications soulignant l'importance des rôles multiples des protéines SR dans l'épissage des introns (Graveley *et al.*, 1998 ; Tacke *et al.*, 1998 ; Xiao and Manley, 1998 ; Hertel and Maniatis, 1999). En effet, les résultats suggèrent que chaque événement d'épissage requiert une protéine SR pour effectuer une fonction exon dépendante de reconnaissance des bornes de l'exon, et une autre protéine SR pour effectuer les autres fonctions exon indépendantes. Cette hypothèse

va à l'encontre de l'idée qu'une protéine SR est capable d'emplir toutes les fonctions des protéines SR dans l'épissage, et préfère le concept d'une multiplicité de rôles portée par l'ensemble des protéines SR (Fig. 48 ; Graveley, 2000).

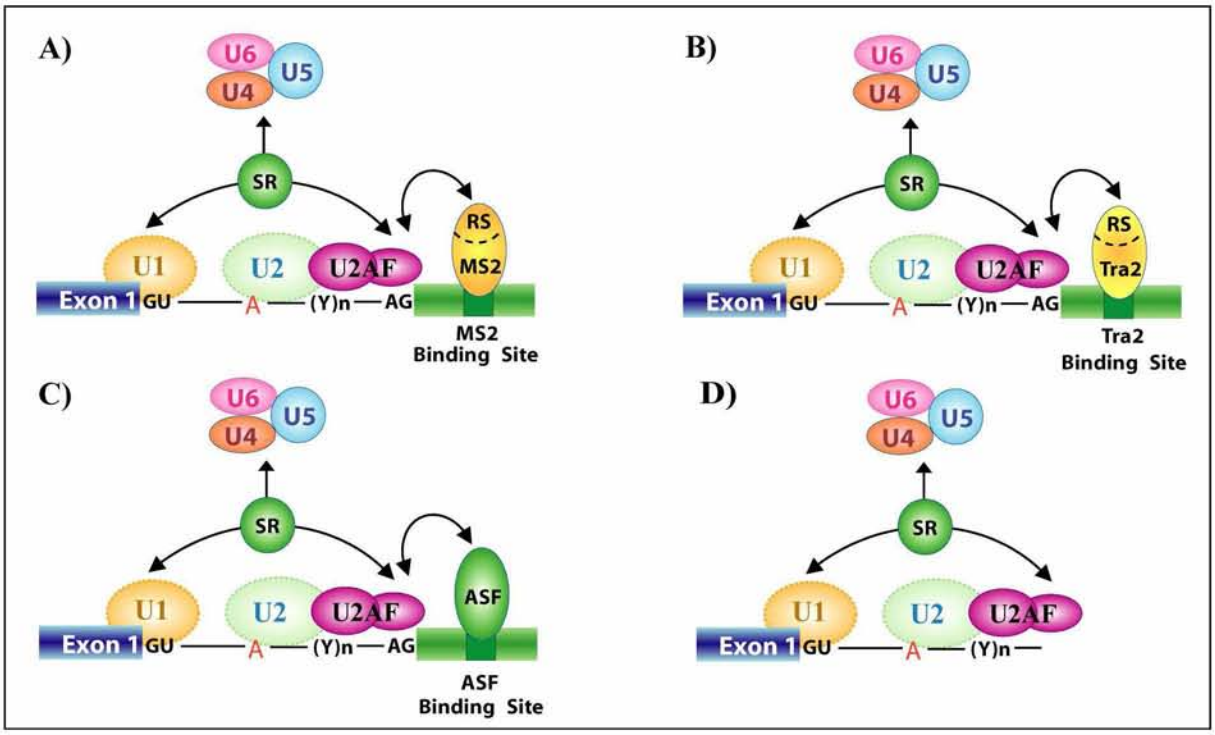


Fig. 48 :Différents rôles des protéines SR dans l'épissage *(d'après Graveley, 2000)*

Liste des résultats mettant en évidence les multiples contributions individuelles des différentes protéines SR dans la réaction d'épissage. Dans les modèles d'études A, B et C, la fusion MS2-RS (en A), le facteur hTra2 (en B) ou la protéine ASF/SF2 thiophophorylée (en C) assurent les fonctions exon (Manley and Tacke, 1996 ; Graveley and Maniatis, 1998 ; Xiao and Manley, 1998). Le dernier modèle étudié (en D) montre que la seule présence de protéines SR assurant les fonctions exons indépendantes permet les premières étapes de formation du spliceosome (Hertel and Maniatis, 1999).

Dans cette optique, l'effet spécifique des protéines SR concernerait leurs fonctions exon dépendantes alors que les fonctions pour lesquelles elles sont redondantes seraient exon indépendantes. Ceci a d'ailleurs en partie été approuvé par une étude montrant que seules les fonctions exon indépendantes sont employées pour l'épissage des introns constitutifs, et que dans ce cas précis toutes les protéines SR peuvent remplir ce rôle (Hertel and Maniatis, 1999).

4.1.3.2. Rôles donnés aux interactions domaine RS-ARN

L'idée que les protéines SR puissent interagir avec la machinerie d'épissage autrement que par des interactions protéine-protéine et des interactions protéine-ARN médiées par le ou les domaines RRM, a été conforté par des études de l'épissage chez *Saccharomyces cerevisiae*. En effet, chez les *Saccharomycetales* en général, famille à laquelle appartient *S. cerevisiae*, il n'existe pas de protéine SR à proprement parlé alors qu'elles sont relativement conservées chez les eucaryotes supérieurs (Kaufer and Potashkin, 2000). Une deuxième particularité est que les séquences de branchement et les sites 5' d'épissage de *S. cerevisiae* dérivent très peu des séquences consensus (pour revue, Lim and Burge, 2001). Une des explications possibles était donc que les protéines SR compensait en partie chez les vertébrés cette faiblesse des signaux d'épissage en favorisant l'interaction des UsnRNA avec leurs séquences cibles dans les introns, et ce en impliquant des interactions protéine-ARN autres que celles impliquant les motifs RRM (Valcarcel *et al.*, 1996 ; Shen and Green, 2004 ; Shen *et al.*, 2004).

Cette hypothèse a été récemment étudiée par des expériences de pontage covalent permettant d'identifier les cibles ARN des domaines RS. En effet, cette approche a permis de démontrer que les domaines RS ont une affinité particulière pour les hybrides ARN-ARN, ce qui leur permet de s'associer sélectivement aux éléments des pré-ARNm transitoirement appariés aux UsnRNA (Shen and Green, 2007).

Par emploi de cette même technique, il est apparu que différents types de domaine RS sont nécessaires au cours de l'épissage, chacun ayant un rôle distinct dans l'assemblage du spliceosome (Shen and Green, 2004 ; Shen *et al.*, 2004). Dans un premier temps, 2 domaines RS, l'un du facteur U2AF^{65} et l'autre d'une protéine SR, sont requis lors de l'assemblage du pré-spliceosome. Ces deux domaines RS contactent le point de branchement séquentiellement pour favoriser l'hybridation du snRNA U2 : c'est tout d'abord, le domaine RS de U2AF^{65} qui se fixe lors de la formation du complexe E, puis il est ensuite remplacé par celui d'une protéine SR dans le pré-spliceosome (ou complexe A). De la même façon, il a été démontré qu'un troisième domaine RS, issu d'une protéine SR, est aussi requis pour stabiliser l'appariement du snRNA U1 au site 5' d'épissage (Fig. 49 ; Shen and Green, 2004 ; Shen *et al.*, 2004). Ces travaux ont par ailleurs montré que les domaines RS des protéines SR devaient être phosphorylés, probablement au niveau des sérines, pour être efficaces.

Si ces résultats modifient considérablement la façon d'envisager le rôle des protéines SR et plus globalement des domaines RS, ils n'excluent pas pour autant les autres modèles qui impliquent des interactions entre les domaines RS des différents facteurs d'épissage incluant

les protéines SR, la protéine U1 70K et U2AF[35]. De manière intéressante d'ailleurs, ces facteurs chez *S. cerevisiae* sont soit absents, c'est le cas des protéines SR et de U2AF[35] (pour revue, Graveley, 2000 ; Sanford *et al.*, 2003), soit sans domaine RS pour la protéine U1 70K (Smith and Barrell, 1991).

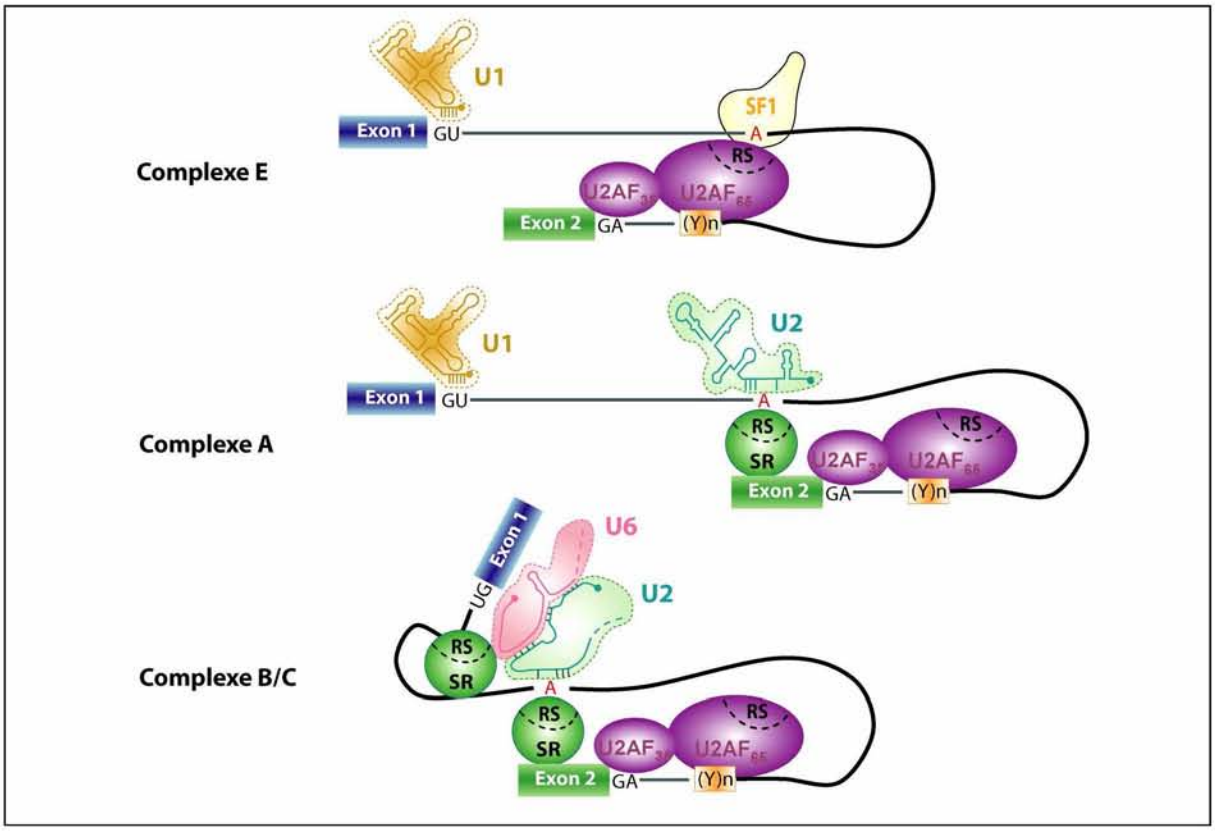


Fig. 49 : Séquence de mise en place des interactions entre les domaines RS et l'ARN pré-messager *(d'après Hertel and Graveley, 2005)*

Le domaine RS du facteur U2AF65 contacte le point de branchement dans le complexe E. Il est ensuite remplacé par le domaine RS d'une protéine SR liée à un ESE dans le complexe A. Finalement, le domaine RS d'une seconde protéine SR non liée à une séquence activatrice d'épissage favorise la liaison du snRNA U6 au site 5' donneur.

4.1.3.3. Antagonisme

Un autre modèle proposé pour comprendre l'effet globalement activateur des protéines SR sur l'épissage est celui du phénomène d'antagonisme par rapport aux protéines hnRNP. Le rôle des protéines SR consiste alors juste à contrecarrer l'effet inhibiteur de protéines hnRNP liées à une séquence inhibitrice d'épissage ESS chevauchante ou juxtaposée à l'élément ESE. Dans ce cas, l'activation de l'épissage est tout à fait indépendante des domaines RS, seuls les

domaines RRM des protéines SR sont nécessaires. Nous reparlerons de ce phénomène dans le paragraphe consacré aux protéines hnRNP.

4.1.3.4. Effet inhibiteur

Bien que les protéines SR aient en forte majorité un effet activateur sur l'épissage, elles peuvent parfois avoir des effets inhibiteurs, et ceci du fait de leur fixation sur des éléments de séquences requis pour l'assemblage des éléments du spliceosome ou leur interaction avec certains de ces composants (pour revue, Pozzoli and Sironi, 2005).

On peut citer l'exemple de l'élément intronique 3RE, riche en purines, qui réprime l'épissage au site accepteur IIIa du pré-messager E1A de l'adénovirus en phase précoce d'infection, par fixation de la protéine ASF/SF2 (Kanopka *et al.*, 1998 ; Akusjarvi and Stevenin, 2003). Les auteurs avaient initialement proposé que la fixation d'ASF/SF2 à cette séquence empêche celle de la snRNP U2 sur le point de branchement. Des données postérieures montrent que l'inhibition provient d'interaction entre le domaine RRM2 d'ASF/SF2 et d'autres facteurs qui eux sont plus directement impliqués dans le mécanisme de répression (Dauksaite and Akusjarvi, 2002). Ce phénomène d'inhibition est lié à la localisation du site de fixation d'ASF dans l'intron car l'inhibition est conservée lorsque cet élément est transporté à une position similaire dans un intron du pré-ARNm de la β-globine de lapin, mais est au contraire perdue s'il est inséré dans un exon (Dauksaite and Akusjarvi, 2002).

Parmi les autres cas connus d'inhibition de l'épissage par une protéine SR, on peut citer :

- l'exclusion de l'exon 4 du pré-ARNm de la protéine SRp20 engendrée par la présence d'ASF/SF2 (Jumaa *et al.*, 1996). Le mécanisme n'a pas encore été défini.
- l'exclusion de l'exon 8 du pré-ARNm de la protéine hnRNP A1 par la protéine SRp30c fixée à la séquence régulatrice CE9 (Simard and Chabot, 2002 ; Simard and Chabot, 2000).
- l'exclusion de l'exon 9 de CFTR par ASF, SRp55 et 75 (Buratti *et al.*, 2001)
- l'exclusion de l'exon 4 de CD45 par ASF et SC35 (Lemaire *et al.*, 1999)

4.1.3.5. Rôles des protéines SR en dehors de l'épissage

Présentes dans le noyau grâce à leur signal de localisation nucléaire, certaines protéines SR sont capables d'effectuer des va-et-vient entre le noyau et le cytoplasme. Ce phénomène appelé « shuttling » a été observé pour ASF/SF2, 9G8 et SRp20 (Caceres *et al.*, 1998). Pour les protéines SR qui ne ressortent pas du noyau, comme SC35, il a été montré qu'elles sont retenues dans le noyau par un signal fort de rétention nucléaire (Cazalla *et al.*, 2002).

Ainsi, de part leur présence sur l'ARN dès sa synthèse et leur propriété de « shuttling », certaines protéines SR jouent un rôle important dans le transport des ARNm, mais aussi dans les évènements cytoplasmiques de localisation et de dégradation des ARNm, et dans la régulation de la traduction.

L'export des ARNm

Le facteur TAP joue un rôle central dans l'export des ARNm. En association avec son cofacteur p15 et d'autres facteurs, il recrute les ARNm, et permet ensuite leur export grâce à son interaction avec les composants du pore nucléaire (Reed and Hurt, 2002 ; Stutz and Izaurralde, 2003). Récemment, il a été démontré que les protéines SR, qui font le va-et-vient entre noyau et cytoplasme (« shuttling »), sont capables d'interagir avec TAP, et donc de servir d'adaptateurs pour le système d'export TAP-dépendant (Huang *et al.*, 2003). Là encore, le taux de phosphorylation des protéines SR semble être directement impliqué dans le mécanisme puisque : (1) la phosphorylation d'ASF/SF2 augmente l'affinité pour U1-70K et décroît l'affinité pour les protéines SR (Xiao and Manley, 1997), et (2) la déphosphorylation de 9G8 et ASF/SF2 augmente fortement leur affinité pour TAP (Huang *et al.*, 2004). Il a donc été postulé que la déphosphorylation des protéines SR servent à désolidariser le complexe ARN/ protéines SR du spliceosome, et dans le même temps à favoriser sa liaison au facteur TAP, pour promouvoir son export (Huang *et al.*, 2004).

Influence sur le NMD

Il a été démontré qu'ASF participe largement à la dégradation de l'ARNm de PKCI-1 dans les fibroblastes de poulet, ceci par son interaction avec une région riche en purine de la région 3'UTR (Lemaire *et al.*, 2002). Plus globalement, il a été montré que la surexpression de protéines SR augmente fortement l'activité du système NMD, cette propriété étant dépendante du domaine RS mais non du phénomène de « shuttling » (Zhang and Krainer, 2004).

Influence sur la traduction

Récemment enfin, il a été démontré pour la première fois que deux des protéines SR transitoirement cytoplasmiques, ASF/SF2 et SRp20, pourraient intervenir dans la traduction. L'étude a montré : (1) que les deux protéines sont associées à la particule ribosomique 80S dans un extrait cytoplasmique et (2) qu'ASF/SF2 stimule la traduction d'une construction reporter dans les cellules HeLa *in vivo* et *in vitro* (Sanford *et al.*, 2004 ; Graveley, 2005).

4.2. Les protéines SR-like

4.2.1. Les co-activateurs SRm300 et SRm160

Deux protéines SRm300 et SRm160 ont été identifiées comme des co-activateurs de l'épissage (Eldridge *et al.*, 1999 ; Blencowe *et al.*, 2000). Elles comportent de multiples répétitions RS mais, à la différence des protéines SR, ne comportent pas de domaine de liaison à l'ARN et ne sont pas capables de complémenter un extrait cytoplasmique S100 pour l'activité d'épissage. Ces co-activateurs d'épissage s'associent en un complexe SRm160/300 et activent l'épissage constitutif ou ESE-dépendant des ARN pré-messagers. L'association du complexe SRm160/300 avec le pré-ARNm « *doublesex* » (*dsx*) de drosophile qui comporte un élément activateur ESE de type GAA requiert la fixation préalable de la snRNP U1 au site 5' d'épissage. Le complexe SRm300/SRm160 pourrait interagir avec la snRNP U1 de manière indirecte par les protéines SR liées à un ESE (Fig. 50) Cette association est stabilisée par la présence de la snRNP U2. Le complexe SRm160/300 établit donc des interactions avec des composants du spliceosome et ferait un lien entre les particules snRNP et les protéines SR liées à un ESE. Le co-activateur SRm160 serait capable d'établir de nombreuses interactions avec les protéines SR et faciliterait ainsi l'épissage des pré-ARNm (Longman *et al.*, 2001).

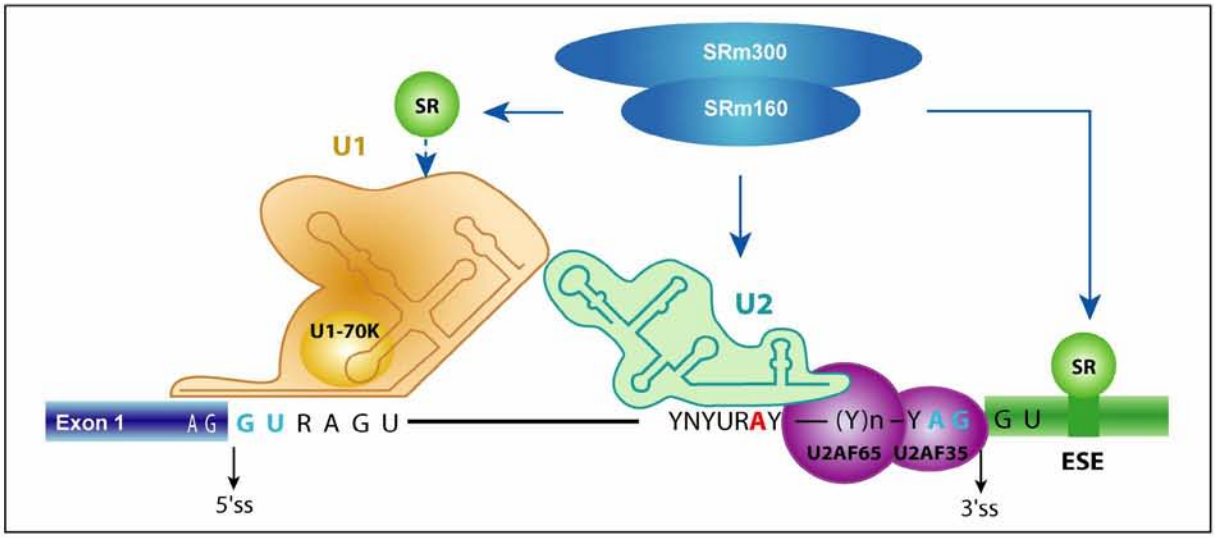


Fig. 50 : Modèle d'activation de l'épissage par les co-activateurs SRm 160/300 *(d'après Eldridge et al., 1999)*

Les co-activateurs SRm160/300 forment un complexe activant l'épissage constitutif mais également l'épissage dépendant des protéines SR. Le complexe SRm160/300 établit des interactions avec des composants du spliceosome (snRNP U1 et U2) et ferait un lien entre les particules snRNP et les protéines SR liées à un ESE.

4.2.2. Les protéines Tra

La protéine Tra a d'abord été découverte chez la drosophile. Elle est en particulier impliquée, au travers de son activité de régulation de l'épissage, dans la détermination du sexe chez la drosophile. Elle est constituée d'un domaine RRM et de deux domaines RS et est uniquement exprimée chez les femelles. Cette protéine régule l'épissage des pré-ARNm *dsx* (« doublesex ») et *fru* (« fruitless ») qui sont épissés de manière alternative chez les mâles et les femelles. Le pré-ARNm *dsx* est constitué de 5 exons constitutifs et d'un exon 4 alternatif. L'exon 4 est soit inclus dans le pré-ARNm, ce qui conduit à un organisme femelle, soit exclu, ce qui conduit à un organisme mâle. L'inclusion de l'exon 4 chez les femelles dépend d'un élément activateur appelé dsxRE, qui est constitué de 6 répétitions de 13 nt, dont la séquence consensus est UC($^{A}/_{U}$)($^{A}/_{U}$)C($^{A}/_{G}$)AUCAACA (Fig. 51). Ces répétitions sont localisées 300 nt en aval du site 3' de l'exon 4 et servent de sites de liaison à un complexe formé par l'association coopérative des protéines Tra, Tra2 et d'une protéine SR, RBP1. Ceci favoriserait le recrutement du facteur U2AF sur le site 3' d'épissage de l'exon 4 et permettrait son inclusion. La protéine obtenue, DSXF, est un facteur de transcription réprimant les gènes de la différenciation mâle. Chez les mâles, en l'absence de protéines Tra, les interactions Tra2, RBP1 et le pré-ARNm ne sont pas stables, ce qui conduit à l'exclusion de l'exon 4. La protéine résultante est DSXM, un facteur de transcription impliqué dans la différenciation cellulaire (Hertel *et al.*, 1996 ; Lynch and Maniatis, 1996).

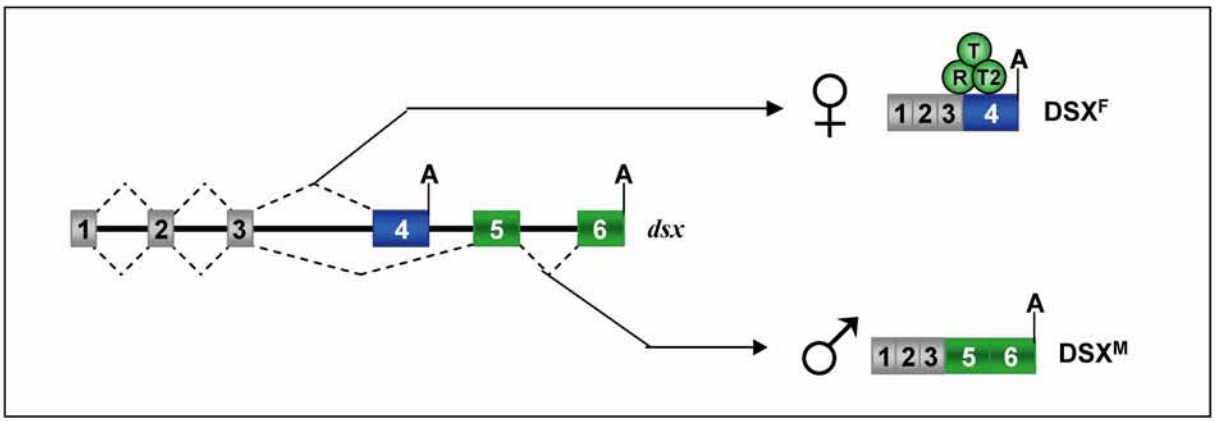


Fig. 51 : Epissage alternatif de l'ARN pré-messager dsx chez la drosophile

L'élément dsxRE présent dans l'exon 4 est le site de fixation du complexe tri-partite formé de Tra (T), Tra2 (T2) et une protéine SR RBP1 (R) qui favoriserait le recrutement du facteur U2AF sur le site 3' de l'exon 4. L'absence de la protéine Tra chez les mâles conduit à l'exclusion de l'exon 4 (pour revue, Smith and Valcarcel, 2000).

Chez l'homme, 2 protéines homologues à la protéine Tra2, hTra2α et hTra2β, ont été identifiées. Comme les protéines Tra et Tra2, elles sont capables de réguler l'épissage en interagissant avec les protéines SR. C'est par exemple le cas pour l'exon 7 du pré-ARNm *SMN2*, dont l'inclusion est activée par une séquence riche en purines de type GAA sur laquelle hTra2β se fixe et recrute la protéine SRp30c (Young *et al.*, 2002).

4.3. Les protéines hnRNP

Dès leur transcription, les ARN produits par l'ARN polymérase II sont associés avec un grand nombre de protéines hnRNP. Des traitements aux RNases de ces complexes hnRNP produisent des monoparticules de 20 nm de diamètre sédimentant à 40S et contenant des segments d'ARN d'environ 700 nt. Dans ces monoparticules, les protéines hnRNP A1, A2, B1, B2, C1, C2 sont les plus représentées. Plus de 25 protéines hnRNP, dont le poids moléculaire varie entre 34 (A1) et 120 kDa (U), ont été mises en évidence (Beyer *et al.*, 1977).

Les protéines hnRNP sont très abondantes dans la cellule, jusqu'à 100 millions de copies par cellule (Dreyfuss *et al.*, 2002).

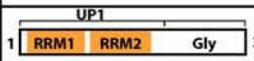
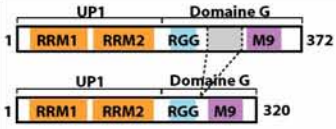
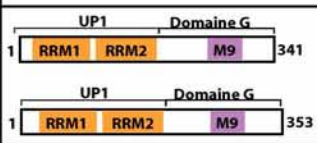
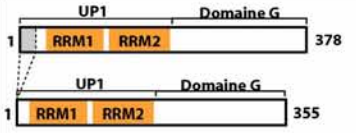
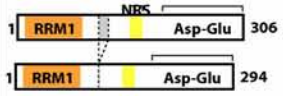
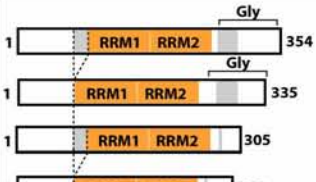
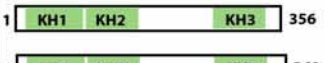
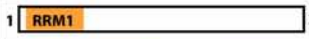
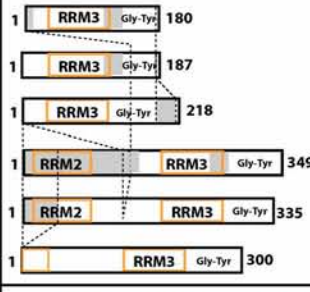
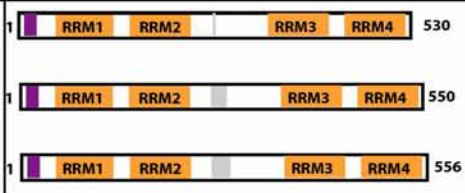
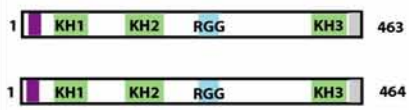
4.3.1. Organisation structurale des hnRNP

Les domaines de liaison à l'ARN

Les protéines hnRNP (Fig. 52) se fixent à l'ARN par leurs domaines de type RRM, RGG ou KH (Dreyfuss *et al.*, 1993). Elles ont toutes une affinité assez large pour l'ARN, avec néanmoins une affinité plus forte pour certains types de séquences : séquences riches en C pour les hnRNP E, I et K, séquences riches en G pour les hnRNP A/B, E, H, H', 2H9, M et U, et séquences riches en U pour les protéines hnRNP C et PTB. Certaines protéines hnRNP ont en plus la capacité de reconnaître avec une forte affinité des séquences spécifiques (hnRNP A/B, M, 2H9, PTB) (pour revue, Dreyfuss *et al.*, 2002). Ainsi, la protéine hnRNP A1 se lie à l'ARN en général avec une affinité de l'ordre du micromolaire et reconnaît par ailleurs la séquence UAGGG(A/U) avec une affinité de 10^{-8} molaire (Burd and Dreyfuss, 1994).

Le domaine d'oligomérisation ou domaine Gly

C'est le domaine C-terminal des protéines de la famille hnRNP A/B, riche en résidus glycine (domaine Gly), qui permet leur oligomérisation. Des résidus phénylalanine et tyrosine dispersés de manière régulière entre les résidus glycine sont requis pour que les protéines de

hnRNP	organisation structurale	PM (kDa)	L	séquence ARN	rôles cellulaires
A0		31,7	N	$G_4(AU_3)_4A$	Contrôle de la traduction
A1B A1		34	N,C	UAGGG(A/U) (Py)UAG	Epissage, export des ARN Maintenance des télomères
A2 B1		36 38	N,C	GCCAAGGAGCC	Epissage, localisation des ARNm, maintenance des télomères
A3(1) A3(2)		41 39	N,C	GCCAAGGAGCC	Localisation des ARNm
C2 C1		41 43	N	riche en U, $AGUA(U)_5GUGGA$	Epissage, contrôle de la traduction, contrôle de la localisation des pré-ARNm, maintenance des télomères
D(1) D(2) D(3) D(4)		45 42 40 37	N	riche en A/U, UUAG/G	Maintenance des télomères, stabilisation ou destabilisation des ARNm, recombinaison
E1 E2		37,5	N,C	Poly(C), Poly(G)	Stabilité des ARNm, contrôle traduction
G		42	N	AAGU	Epissage
F		45,5	N	Poly(G)	Epissage
H H'		56 56	N N	GGGA/C, poly(G)	Epissage, polyadénylation
2H9(6) 2H9(5) 2H9(4) 2H9(1) 2H9(2) 2H9(3)		15,5 16 22 37 35 31,5	N	GGGA	Epissage
PTB(1) PTB(2) PTB(4)		59	N	riche en C/U, UCUU, UCUCU, CUUCUCUCU	Epissage, localisation des ARNm, polyadénylation
K(1) K(2)		62	N,C	poly(C), $AU(C)_{3/4}(U/A)_2$	Epissage, régulation de la transcription, contrôle de la traduction

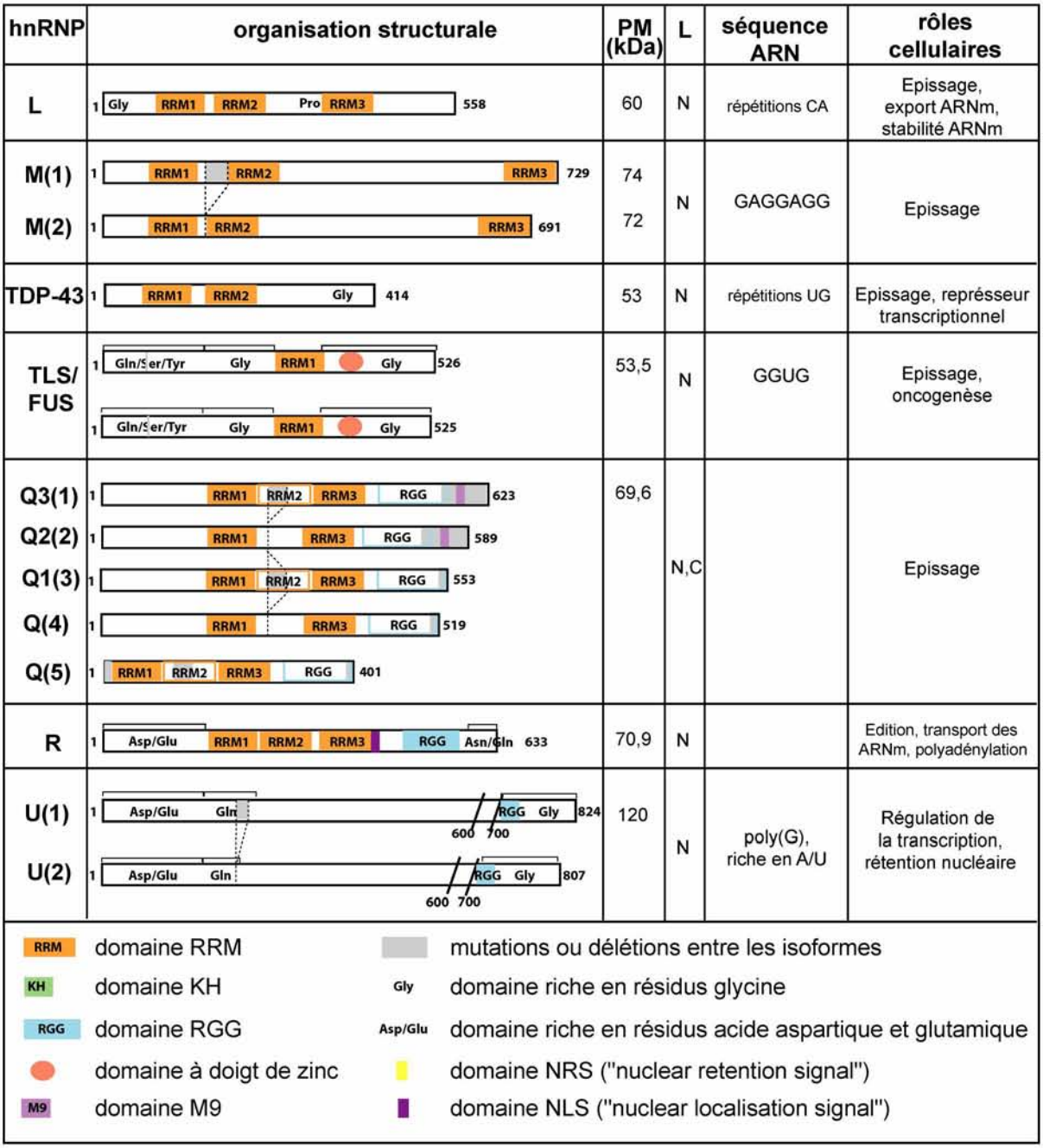


Fig. 52 : Les protéines hnRNP: structures, rôles, et séquences ARN reconnues

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 (d'après Swanson and Dreyfuss, 1988; Dreyfuss et al., 1993; Honore et al., 1999; Caputi and Zahler, 2001; Lerga et al., 2001; pour revues, Krecic and Swanson, 1999; Dreyfuss et al., 2002).

la famille hnRNP A/B interagissent entre elles. Ces résidus interviendraient par formation d'interactions hydrophobes entre les protéines. Un ensemble d'interactions peut être mis en place de manière coopérative, indépendamment de la fixation à l'ADN ou l'ARN, et conduire à une oligomérisation de ces protéines (Cartegni *et al.*, 1996). Par ailleurs, le domaine Gly de ces protéines, en particulier celui de la protéine hnRNP A1, a des propriétés d'«annealing» des acides nucléiques, d'où un rôle possible de chaperonne des acides nucléiques pour la protéine hnRNP A1. Ce domaine est également retrouvé dans beaucoup d'autres protéines hnRNP en particulier hnRNP D, U et H.

Les domaines d'import (M9) et de rétention dans le noyau (NRS)

Bien que les protéines hnRNP soient localisées principalement dans le nucléoplasme des cellules, certaines d'entre elles (les protéines hnRNP A/B, D, E, I, et K), sont capables, comme les protéines SR, de transiter entre le noyau et le cytoplasme. Dans le noyau, elles sont associées aux pré-ARNm ou à l'ADN pour certaines, et dans le cytoplasme, aux ARNm (Pinol-Roma and Dreyfuss, 1992 ; Michael *et al.*, 1995).

Le motif M9 dans la région C-terminale des protéines hnRNP A1/A1B, A2/B1 constitue un signal d'import nucléaire spécifique (pour revue, Nakielny *et al.*, 1999). Ce motif interagit avec un récepteur d'import nucléaire, la transportine (Pollard *et al.*, 1996).

Certaines protéines hnRNP, strictement nucléaires comme la protéine hnRNP C, possèdent un signal de rétention nucléaire NRS[31]. Comme ces protéines se fixent préférentiellement aux séquences introniques des pré-ARNm (pour revue, Nakielny *et al.*, 1999), elles éviteraient aux ARN non épissés d'être exportés dans le cytoplasme (Pinol-Roma and Dreyfuss, 1992). Ces protéines assureraient donc une rétention nucléaire des pré-ARNm.

4.3.2. Variants d'épissage et post-traductionnelles des hnRNP

Les variants d'épissage

Pour la plupart des protéines hnRNP, différentes isoformes ont pu être isolées et sont issues de phénomène d'épissage alternatif. Par exemple, les protéines de la famille hnRNP A/B sont issues de l'épissage alternatif des pré-ARNm produits à partir de 2 gènes distincts résultant probablement de la duplication d'un gène initial (Biamonti *et al.*, 1994). Chaque isoforme possède des caractéristiques définies de liaison à l'ARN et il n'y a pas de lien direct entre efficacité de fixation à l'ARN et efficacité d'action dans l'épissage alternatif (Mayeda *et al.*,

[31] Nuclear Retention Signal

1994). Autre exemple, le pré-ARNm de la protéine PTB subit un épissage alternatif conduisant à la production de 3 isoformes. L'isoforme PTB4 est l'isoforme majoritairement produite. Les 2 autres isoformes PTB1 et PTB2 sont également capables de réprimer l'épissage, mais plus faiblement. Les taux relatifs de ces différentes protéines PTB varient suivant les tissus, ce qui participe aux régulations tissu-spécifique de l'épissage des pré-ARNm (Wollerton *et al.*, 2001 ; pour revue, Wagner and Garcia-Blanco, 2001).

Les modifications post-traductionnelles

Les protéines hnRNP subissent de nombreuses modifications post-traductionnelles qui sont principalement : des phosphorylations sur les résidus sérines ou tyrosines, des méthylations sur les arginines, et des sumoylations sur des résidus. Des différences de taux de modifications post-traductionnelles entraînent des variations dans la capacité des protéines à interagir avec l'ARN. Par exemple, suite à des stress cellulaires (stress oxydatif, réponse à des facteurs de croissance), la protéine hnRNP K est phosphorylée sur des résidus sérines. Ces modifications structurales augmentent l'affinité de la protéine pour d'autres protéines (proto-oncogène, protéines kinases, répresseurs transcriptionnels) et diminuent sa capacité à interagir avec l'ARN *in vitro* (Schullery *et al.*, 1999 ; Ostrowski *et al.*, 2000). De la même manière, la phosphorylation du peptide F, situé en aval du motif M9, gêne l'interaction avec la transportine et diminue l'import nucléaire de hnRNP A1 (Allemand *et al.*, 2005). Autre exemple, les méthylations sur les résidus arginine du domaine RGG de la protéine hnRNP A1, diminuent sa capacité à interagir avec l'ARN et affectent les propriétés d'« annealing » des acides nucléiques de hnRNP A1 (Kim *et al.*, 1998). Enfin, les protéines hnRNP C et M, qui sont exclusivement nucléaires et qui ont la capacité d'interagir avec des ARNm et de les retenir dans le noyau peuvent être sumoylées. Ceci entraîne une diminution de l'affinité des protéines hnRNP pour les ARNm qu'elles retiennent dans le noyau, ce qui faciliterait l'export de ces derniers du noyau vers le cytoplasme (Vassileva and Matunis, 2004).

4.3.3. Les rôles généraux des protéines hnRNP

Du fait qu'elles sont associées aux pré-ARNm dès leur synthèse et que certaines d'entre elles peuvent rester associées aux ARNm jusque dans le cytoplasme, les protéines hnRNP jouent des rôles variés dans les différentes étapes de synthèse, maturation, transport, traduction et dégradation des ARNm (Dreyfuss *et al.*, 1993). Cette affinité pour l'ARN fait des protéines hnRNP un des acteurs majeurs de l'épissage alternatif.

Inhibition de l'utilisation des sites d'épissage :

De part leur capacité à oligomériser le long de l'ARN, le protéines hnRNP jouent la plupart du temps le rôle d'inhibiteur de l'épissage en empêchant la fixation de la machinerie d'épissage. Cette inhibition prend effet sur les sites accepteurs et donneurs sans distinction. Les éléments sur lesquels se fixent les hnRNP peuvent être de nature exonique (ESS) ou intronique (ISS).

Un exemple de ce type de régulation est celui du pré-ARNm de GABAAα2 (sous-unité α2 du récepteur à canal ionique GABAA). Trois séquences inhibitrices fixent la protéine PTB de manière coopérative, ce qui interfère par encombrement stérique avec la liaison de $U2AF^{65}$, et conduit à l'absence de reconnaissance du point de branchement et de la séquence polypyrimidine (Fig. 53 ; Wagner and Garcia-Blanco, 2001).

Un autre exemple de régulation de l'utilisation du site 3' d'épissage est celui de la régulation de l'épissage du pré-ARNm *FGFR -2* codant le récepteur du facteur de croissance des fibroblastes. 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).

Ce mécanisme d'inhibition par encombrement stérique peut aussi intervenir aux sites 5' d'épissage, notamment dans le cas de hnRNP A1 qui est capable d'oligomériser.

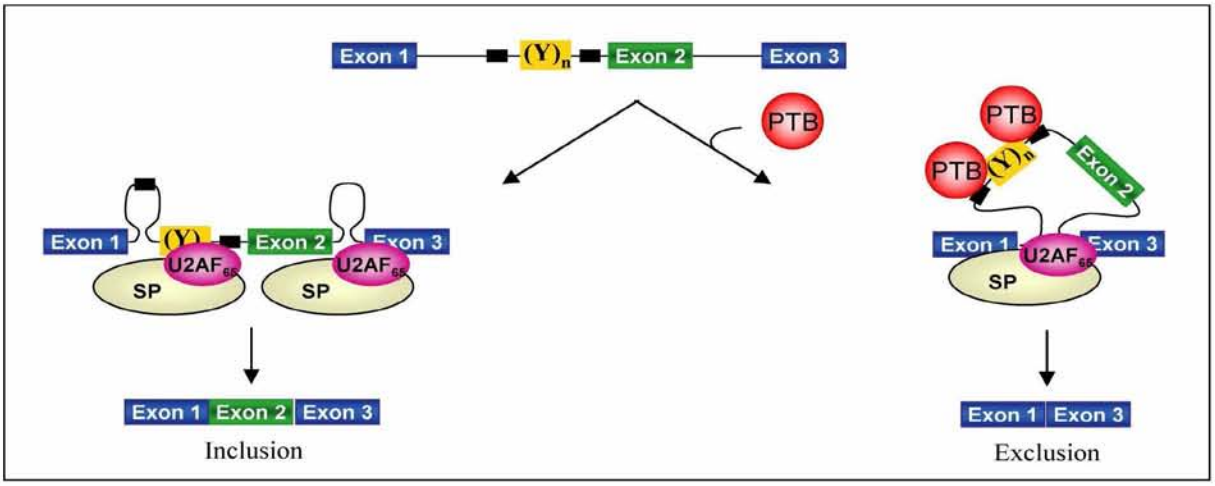


Fig. 53 : Exemple de mécanisme d'inhibition de la reconnaissance de la séquence polypyrimidine dans l'ARN pré-messager GABAAα2 par la protéine PTB/hnRNP I

Inhibition de la reconnaissance par la protéine PTB. La protéine PTB se fixe sur des sites de haute affinité (en noir) situés autour de la séquence polypyrimidine empêchant ainsi la fixation de U2AF65 et favorisant donc l'exclusion de l'exon alternatif (d'après Wagner and Garcia-Blanco, 2001).

Antagonisme entre les protéines hnRNP et les protéines SR

L'inhibition de l'épissage par les protéines hnRNP peut aussi provenir de leur capacité à contrecarrer l'effet activateur des protéines SR liées à une séquence activatrice d'épissage. Cet antagonisme provient du fait que leurs sites de fixation sur l'ARN sont en général juxtaposés voir même chevauchants, ce qui empêche par encombrement stérique la fixation simultanée des 2 types de protéines (pour revue, Smith and Valcarcel, 2000). Ce phénomène a été maintes fois observé et il a conduit à l'hypothèse largement répandue que l'épissage alternatif peut-être régulé de manière tissu-spécifique grâce à des variations des taux relatifs de protéines SR et hnRNP.

L'exemple le plus utilisé est celui de l'antagonisme des protéines ASF/SF2 et hnRNP A1. Une étude de l'effet de ces facteurs sur la reconnaissance par la snRNP U1 de sites 5' en compétition, a permis de mettre en évidence deux paramètres essentiels de cette notion d'antagonisme qui sont la concentration des protéines et leur affinité pour l'ARN. Le modèle proposé indique qu'à forte concentration, hnRNP A1 multimérise sur l'ARN et empêche la particule snRNP U1 de s'associer au site 5' faible qui par ailleurs possède dans son voisinage un site de haute affinité pour la protéine ASF/SF2. Lorsque la concentration en protéine ASF est plus élevée, la compétition est plus forte et le site 5' faible est alors utilisé(Fig. 54 ; Eperon *et al.*, 2000).

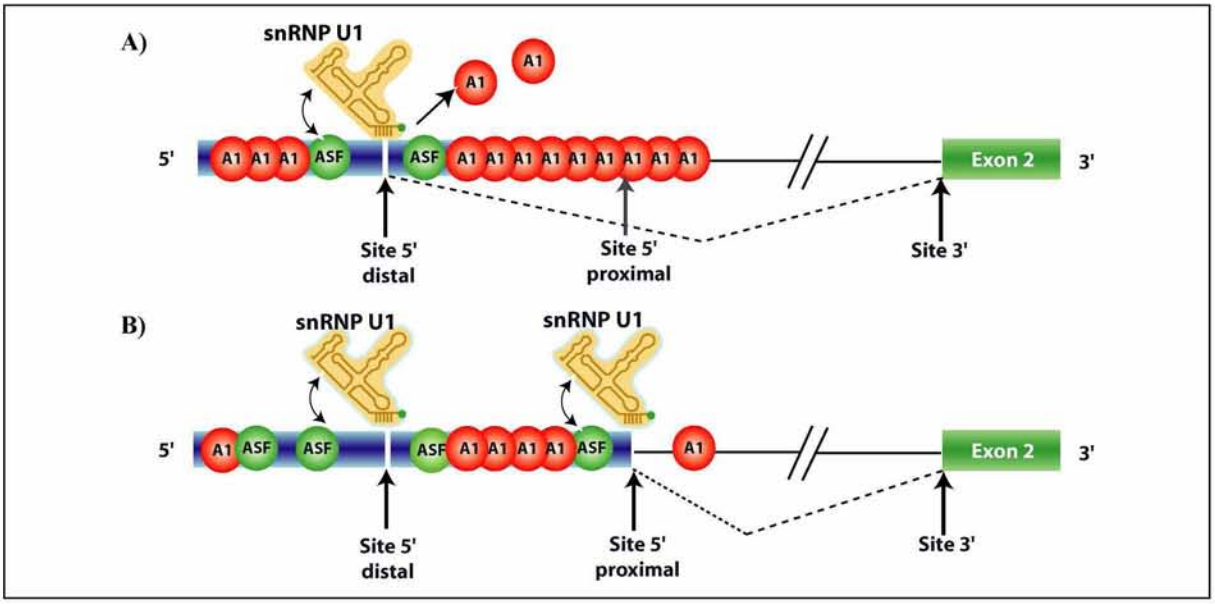


Fig. 54 : Exemple de mécanisme d'antagonisme entre les protéines hnRNP A1 et ASF/SF2 pour la sélection d'un site 5' d'épissage *(d'après Eperon et al., 2000)*

A) Occupation des sites lorsque la concentration en protéine hnRNP A1 est plus élevée que celle en protéine ASF/SF2 dans l'extrait nucléaire. Seuls les sites de haute affinité peuvent être occupés par

ASF/SF2. La liaison de la snRNP U1 aux sites 5' est réduite par compétition avec hnRNP A1. Dans ce cas, le site 5' distal est utilisé du fait de sa proximité avec un site de haute affinité pour ASF/SF2. B) Occupation des sites lorsque la concentration en protéine ASF/SF2 est plus élevée que celle en protéine hnRNP A1 dans l'extrait nucléaire. Tous les sites de faible, moyenne, et haute affinité sont occupés par ASF/SF2. Les deux sites 5' sont occupés par la snRNP U1 et dans ce cas, seul le site 5' proximal est utilisé.

Comme nous l'avons déjà mentionné, ce phénomène est assez fréquent et concerne toutes les protéines. Par exemple, il a été démontré récemment que 9G8 d'une part, et hnRNP H et F d'autre part possèdent un seul et même site de fixation pour réguler positivement (9G8), et négativement (hnRNP H), l'épissage de l'exon 2 de l'ARN de l'α-Tropomyosin (Crawford and Patton, 2006).

Modulation de l'épissage par rapprochement d'intron :

Quelques études ont mis en évidence le fait que les protéines hnRNP A/B puissent relier des séquences introniques très distantes l'une de l'autre. Ce phénomène peut avoir 2 impacts sur l'épissage selon la position des deux sites de fixation.

Le premier impact, si les sites se trouvent dans un même intron, est une stimulation de l'épissage par rapprochement des sites 5' donneur et 3' accepteur. Ce phénomène a déjà été observé dans un contexte hétérologue avec un ARN contenant un large intron, faiblement épissé *in vitro*. L'insertion de séquences hnRNP A1 près des jonctions augmente significativement son épissage (Fig. 55A ; Garneau *et al.*, 2005).

Le second effet possible est un effet inhibiteur indirect car lorsque les sites de fixations de la protéine encadrent un exon, celui-ci peut se trouver alors exclu de l'épissage. C'est le cas du pré-ARNm de hnRNP A1 qui régule ainsi l'épissage de son propre pré-ARNm. La présence de sites de fixation de haute affinité localisés de part et d'autre de l'exon 7B, induit la formation d'une boucle dans l'ARN, qui conduit à l'exclusion de l'exon 7B sans même empêcher la fixation de la snRNP U1 (Fig. 55B ; Chabot *et al.*, 2003).

Activation de l'épissage:

Les mécanismes sont encore assez mal connus, mais il semblerait que certaines protéines hnRNP, dans certains contextes, puissent effectivement jouer un rôle positif sur l'épissage. En voici quelques exemples :

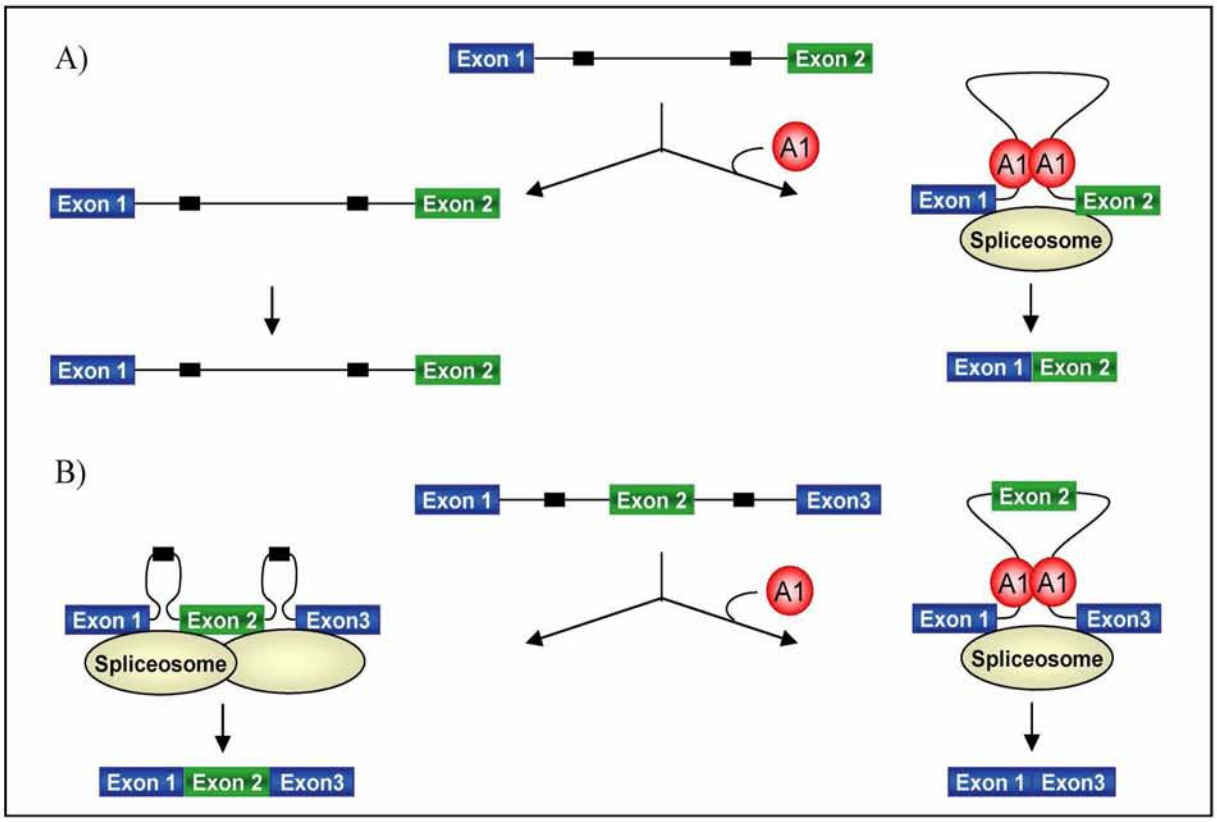


Fig. 55 : Exemple de régulation de l'épissage par rapprochement de séquences ARN par oligomérisation de la protéine hnRNP A1

A) Lorsque les sites de fixations sont à l'intérieur d'un même intron. Ce cas étudié in vitro a montré que la présence de la protéine hnRNP A1 favorise l'épissage lorsque la taille de l'intron est importante (Garneau et al., 2005). B) Lorsque les sites de fixation sont sur deux introns. La capacité d'oligomérisation de certaine hnRNP entraîne l'exclusion de l'exon central. C'est le cas de hnRNP A1 sur son propre ARN pré-messager (Chabot, 2003), mais aussi celui de PTB sur l'exon 3b de l'ARN pré-messager de FGFR-2 (Wagner and Garcia-Blanco, 2001).

- La protéine hnRNP K fait partie d'un complexe activant l'épissage de l'exon 6A de la β-tropomyosine de poulet (Expert-Bezancon *et al.*, 2002).

- La protéine hnRNP G active l'inclusion de l'exon 7 du pré-ARNm *smn17*, en interagissant avec hTra2_1, une protéine à domaine RS (Hofmann and Wirth, 2002).

- Et enfin, la protéine hnRNP M semble avoir un rôle positif puisque des anticorps dirigés contre elle inhibent l'épissage *in vitro* (Gattoni *et al.*, 1996).

Rôle des protéines hnRNP dans la localisation des ARN

Certaines protéines hnRNP sont impliquées dans la localisation des ARN : rétention nucléaire ou export dans le cytoplasme, la fonction dépend du signal de localisation de la protéine.

Ainsi, les protéines hnRNP A2 et B1 interviennent dans des phénomènes de transport des ARNm dans le cytoplasme des cellules polarisées, telles que les oligodendrocytes. Ce transport au sein du cytoplasme est souvent un mécanisme important pour permettre l'expression de protéines spécifiques dans une région particulière de ces cellules. Par leur fixation à des séquences cibles de l'ARN les protéines hnRNP A2 et B1 permettent la localisation de certains ARNm dans des zones définies du cytoplasme (Ainger *et al.*, 1997).
A contrario, une étude récente montrent que la protéine hnRNP H se fixe aux longues répétitions CUG correspondant à des amplifications CTG dans le gène DMPK chez les patients atteints de dystrophie musculaire de type 1 et participerait à la rétention de ces transcrits dans le noyau (Kim *et al.*, 2005).

4.4. Les autres facteurs intervenant dans la régulation de l'épissage

De nombreux autres facteurs de régulation de l'épissage alternatif existent en plus des protéines hnRNP et SR.

La protéine YB-125

La protéine YB-125 se fixe sur une classe particulière de séquences activatrices d'épissage, riches en dinucléotides CA, appelées ACE (Stickeler *et al.*, 2001). Cette protéine est connue pour réguler la transcription de certains gènes et faisant partie de complexes ribonucléoprotéiques se liant aux ARNm avant traduction (pour revue, Wolffe and Meric, 1996). Mais elle est également retrouvée dans le complexe spliceosomal A (Hartmuth *et al.*, 2002) et interagit avec la protéine SRp30c (Raffetseder *et al.*, 2003) ou les protéines hnRNP K et L (Shnyreva *et al.*, 2000 ; Hui *et al.*, 2005). Il a par exemple été montré qu'elle est capable de réguler l'épissage de l'ARN E1A de l'adénovirus (Chansky *et al.*, 2001).

Les protéines CELF

Les protéines CELF[32], également connues sous le nom de BRUNO-like, régulent l'épissage d'un certain nombre de gènes comme, par exemple, le gène cTNT (exon 5) (Ladd *et al.*, 2001) ou le gène de la PTB (exon 11) (Wollerton *et al.*, 2001). Ces protéines sont exprimées de manière tissu-spécifique et leur expression dépend aussi du stade de développement. Elles se fixent sur des séquences riches en U/G situées dans les introns à proximité des exons.

[32] CUG-BP and ETR-3-like factor

La famille de protéines Fox

Cette famille est constituée des protéines Fox-1 et Fox-2 et de leurs isoformes respectifs. Ce sont des facteurs qui régulent l'épissage de manière tissu-spécifique et qui possèdent un domaine RRM. Les expériences SELEX réalisées avec la protéine Fox-1 humaine montrent qu'elle reconnaît spécifiquement la séquence UGCAUG (Ponthier *et al.*, 2006). Une étude récente montre que ces séquences UGCAUG sont conservées phylogéniquement et retrouvées le plus fréquemment en aval d'exons alternatifs qui sont surtout inclus dans des ARNm produits dans le cerveau ou dans les muscles (Minovitsky *et al.*, 2005). Ces séquences sont retrouvées dans plusieurs gènes comme celui de la fibronectine, c-src, calcitonine/CGRP, NMHC-B, 4.1R, FGFR-2 (Minovitsky *et al.*, 2005).

La famille de protéines Nova

Dernier exemple, les protéines Nova-1 et Nova-2 sont des facteurs de régulation de l'épissage exprimés exclusivement dans les neurones du système nerveux central (Buckanovich *et al.*, 1993). Comme la protéine hnRNP K, chaque protéine Nova contient 3 domaines KH. Des expériences de SELEX montrent qu'elles se fixent sur des séquences répétées UCAY. Des données récentes montrent que 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 GABAA et de la sous-unité α2 du récepteur de la glycine (Ule *et al.*, 2003).

OBJECTIFS :

Dans cette introduction nous avons présenté en lien avec les travaux que nous avons réalisé en thèse : (1) le virus HIV-1 et son mode de multiplication, (2) la transcription chez les eucaryotes et comment le virus détourne ce mécanisme à son profit et enfin (3) l'épissage alternatif chez les vertébrés, ainsi que tous les éléments impliqués pour sa régulation. En effet, au cours de ce travail de thèse, j'ai abordé différentes thématiques : 1) l'étude du complexe TAR/Tat/cycline T1 impliqué dans la transactivation de la transcription de l'ARN du virus HIV-1, et 2) différentes facettes de la régulation de l'épissage de l'ARN du virus HIV-1. Ces deux axes s'inscrivaient dans le cadre de deux programmes de recherche financés par l'ANRS, le premier sur la structure de la protéine Tat était coordonné par R. Haser, le second sur l'épissage de l'ARN viral était coordonné par C. Branlant.

Comme nous l'avons vu dans l'introduction, l'action de la protéine virale Tat est primordiale pour activer la transcription et permettre l'obtention d'ARN viraux de pleine longueur. Le mécanisme permettant cela fait intervenir l'élément TAR présent à l'extrémité 5' des ARN du virus HIV-1 d'une part, et la cycline T1, composant du complexe p-TEFb responsable de l'hyperphosphorylation de l'ARN polymérase II d'autre part. L'objectif principal du projet coordonné par R. Haser était l'étude structurale de Tat au sein de ce complexe TAR/Tat/cycline T1. C'est donc dans cette optique, qu'il nous a été proposé à Mme Branlant et moi-même de produire l'ARN TAR et la cycline T1. La protéine Tat nous étant fournie par un des industriels participant au projet (Pierre Fabre), il nous fallait tester la formation de complexes entre les trois partenaires (TAR, Tat et cycline T1) afin de déterminer les meilleures conditions permettant l'étude du complexe par radiocristallographique. Pour la réalisation de ce travail, en plus des conseils de C. Branlant, j'ai bénéficié de l'expertise de plusieurs chercheurs et ingénieurs permanents de l'équipe, X Manival (CR1 CNRS) pour les productions de macromolécules biologiques et C Charron (IR CNRS) pour la cristallogénèse.

Le second volet de ma thèse était consacré à l'étude de régulations de l'épissage alternatif de l'ARN du virus HIV-1, épissage alternatif, qui comme on l'a vu constitue une étape majeure du cycle de multiplication du virus HIV-1. Cette thématique était déjà largement étudiée au laboratoire lors de mon arrivée en thèse et il m'a été proposé à la fois de prolonger certaines études déjà entamées par D. Ropers, L. Ayadi et H. Hallay sur les régulations de l'utilisation du site A3, mais aussi d'ouvrir un autre axe consistant en l'étude de l'impact de certaines

protéines du virus sur l'épissage alternatif de son transcrit. En ce qui concerne le site A3, l'objectif était de continuer la caractérisation fine du mécanisme de régulation faisant intervenir des éléments ESE et ESS associés à ce site et les protéines SR et hnRNP se fixant sur ces éléments. Pour le nouvel axe de recherche, l'objectif était en particulier de tester l'effet des protéines Tat, Nef et NCp7. Un point commun à ces différents thèmes de recherche, qui constituait une originalité des travaux, était de compléter toutes les études *in vitro*, par des études *in cellulo* utilisant un ARN dérivé de l'ARN du virus HIV-1, comportant les 6 sites donneurs et 9 sites accepteurs d'épissage. Ce travail impliquait donc la mise au point de la technique de mutagénèse dirigée du plasmide pΔPSP (12 kb) au laboratoire, ce qui n'avait pas été fait auparavant au laboratoire. Pour ce travail sur les régulations de l'épissage du virus HIV-1, en plus des conseils de C Branlant, j'ai aussi bénéficié de ceux de A. Visvikis (Professeur UHP) et L. Ayadi (IE CNRS) au laboratoire, et ceux de l'équipe de J. Stévenin (IGBMC, Strasbourg) qui en plus nous a fourni toutes les protéines SR recombinantes que nous avons utilisées pour cette étude. Enfin, pour les tous derniers travaux qui ont consisté à purifier les complexes ARN/protéines formés autour de la région contenant le site A3 et ses éléments de régulation, j'ai pu bénéficier de la collaboration avec L. Fouillen, S. Sanglier et A. Van Dorsselaer (LSMBO, Strasbourg) pour l'identification par spectrométrie de masse des protéines liées.

RESULTATS

CHAPITRE 4 : **Travaux en vue d'une étude structurale du complexe TAR/Tat/cyclineT1**

CHAPITRE 5 : **Etude de l'effet de l'expression des protéines virales Tat, Nef et NCp7 sur les taux d'expression des facteurs cellulaires d'épissage et impact sur l'épissage de HIV-1**

CHAPITRE 6 : **Etudes de l'effet de l'élément inhibiteur ESS2 sur la compétition entre les sites d'épissage A1 à A5**

CHAPITRE 7 : **Etude de la régulation de l'épissage de l'exon A2-D3**

CHAPITRE 8 : **Mise en évidence de régulations du site A3 indépendantes de l'élément ESS2**

CHAPITRE 4 : Travaux en vue d'une étude structurale du complexe TAR/Tat/cyclineT1

1. Présentation de la problématique et du partenariat

Nous avons vu que la protéine Tat joue un rôle majeur pour la multiplication du virus HIV-1. Aussi de nombreuses études avaient déjà été réalisées dans l'objectif d'apporter des informations structurales sur cette protéine. L'idée étant bien sûr à partir de cette structure, de pouvoir définir des molécules chimiques capables de bloquer son action. Malheureusement, malgré les différents travaux menés sur des protéines Tat de différents isolats ou homologues d'autres lentivirus, peu de données convergent finalement (Fig. 56).

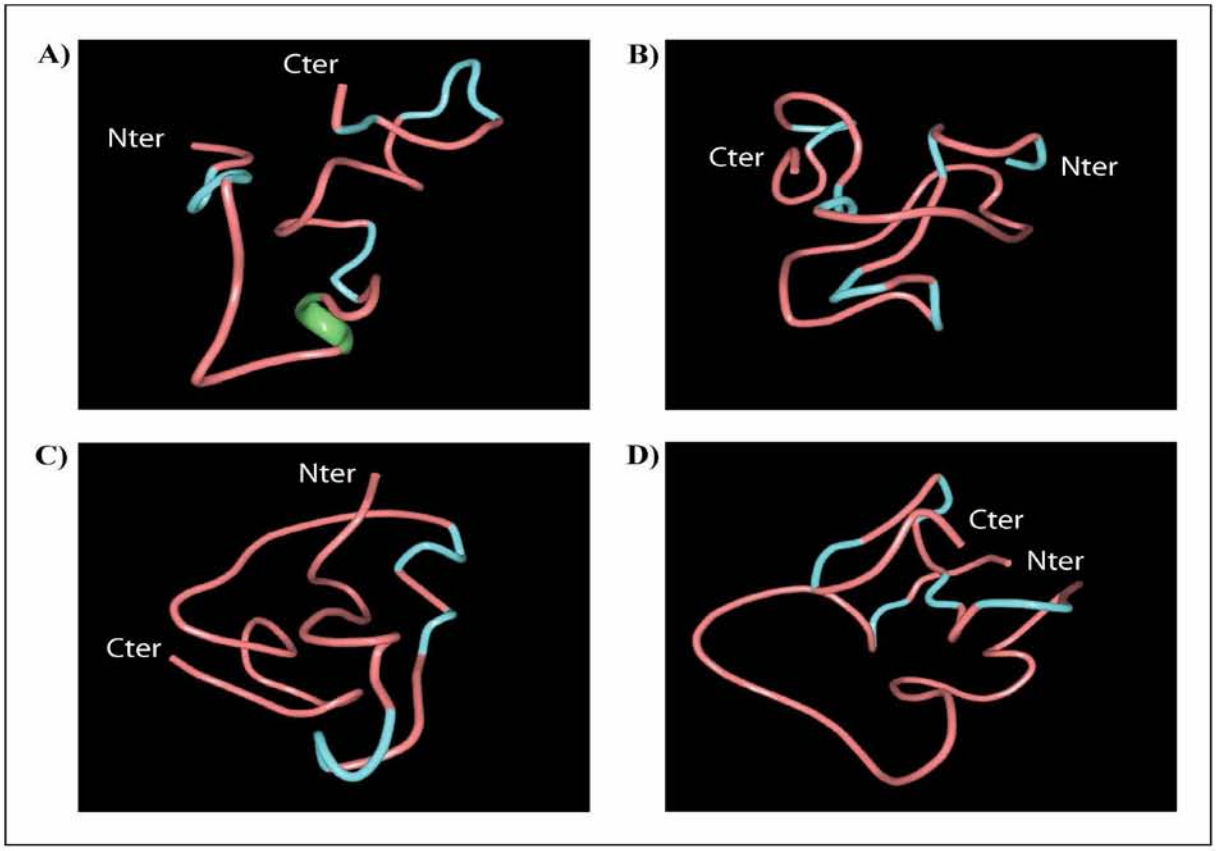


Fig. 56 : Structure 3D des protéines Tat obtenus :

Présentation des structures 3D obtenues pour les protéines (A) EIAV Tat (1-72) (1TVT, Willbold et al., 1994), (B) HIV-1 Tat (1-87) (1K5K, Gregoire et al., 2001), (C) HIV-1 Tat (1-86) (1JFW, Peloponese et al., 2000) et (D) HIV-1 Tat (1-86) (1TIV, Bayer et al., 1995). Les segments en bleu correspondent à des boucles et ceux en verts sont prédits comme formant une structure hélice. Les valeurs RMS sur le squelette par rapport à la structure 1TVT sont de 8,16 Å pour 1K5K, 9,30 Å pour 1JFW et 7,4 6 Å pour 1TIV.

Une première étude RMN qui était couplée à des simulations moléculaires par informatique a été réalisée sur la protéine Tat86. Les données suggèrent qu'elle comporte un domaine globulaire structuré correspondant aux régions appelées « core » et région riche en Gln, les régions basique et riche en Cys étant elles hautement flexibles à pH 6,3 en conditions réductrices (Fig. 56D ; Bayer *et al.*, 1995). L'étude de la protéine EIAV Tat (1-72) a aussi révélé la présence d'une région « core » globulaire. Contrairement à HIV-1 Tat, certaines séquences apparaissent comme formant des structures en hélices α (Fig. 56A ; Willbold *et al.*, 1994). Il a ensuite été montré que EIAV Tat a une structuration proche de celles des homéodomaines responsables de la fixation de facteur de transcription à l'ADN (Willbold *et al.*, 1998). La présence d'une partie globulaire avec quelques fragments flexibles (particulièrement la région basique) a aussi été proposée à partir d'une étude RMN à faible résolution sur une protéine Tat synthétique de 86 résidus. Cette étude étant réalisée à un pH non physiologique (pH 4,5) et en absence d'agent réducteur (Fig. 56C ; Peloponese *et al.*, 2000). Un modèle correspondant à un degré plus élevé de structuration a été obtenu avec une protéine Tat de 87 résidus fusionnée à la protéine Mal E, l'étude ayant là aussi été faite à pH 4,5 en conditions oxydantes. Dans le modèle ainsi établi, trois résidus de la région « core » le Trp11, Phe 38 et Tyr47 forment une face hydrophobique. La région basique adopte une conformation étendue, le domaine riche en Cys contient un feuillet beta et le domaine riche en Gln une hélice α (Fig. 56B ; Gregoire *et al.*, 2001).

Les études RMN réalisées sur des fragments de Tat en interaction avec TAR ont surtout donné des informations sur la structure de TAR (Puglisi *et al.*, 1992 ; Aboul-ela *et al.*, 1995, Long and Crothers, 1999). L'une de ces études a néanmoins permis de mettre en évidence un changement conformationnel de Tat (32-72) lors de son association avec TAR (Metzger *et al.*, 1997). Dans le cadre de l'étude des interactions formées entre Tat et le complexe nucléaire à activité histone acétyl transférase CBP[72]/p300 et son facteur associé PCAF[73], deux analyses structurales par RMN de peptides de Tat ont aussi été réalisées. La première a concerné le bromodomaine de PCAF (719-832) en interaction avec un peptide de Tat (SYGR-AcK-KRRQR, 47-54). Elle a permis de mettre en évidence des interactions très spécifiques impliquant l'acétylation de K_{50} (et pas celle de K_{28} ou K_{51}) et les résidus Y_{47}, R_{53} et Q_{54}. L'interaction entraîne des variations locales de conformation du bromodomaine (Mujtaba *et al.*, 2002). Cette étude a aussi révélé que l'interaction PCAF/Ac-Tat inhibe la fixation de Tat sur TAR. Néanmoins, elle n'a pas donné de réelles informations sur la structure de Tat.

[72] CREB Binding Protein
[73] p300/CBP associated factors

La seconde étude, par RMN a permis d'établir un modèle 3D de l'interaction entre le domaine KIX de CBP et le fragment N-terminal de Tat (1-24) dans lequel ce fragment forme une hélice α (Vendel and Lumb, 2004).

Au final, les différences importantes observées entre ces structures établies révèlent le caractère très flexible de la protéine Tat. Ceci peut s'expliquer par la multiplicité des rôles de Tat et donc de ces partenaires, mais complique aussi largement l'approche « drug design » qui ne peut viser l'inhibition de toutes les interactions. L'idée générale du projet initié par l'ANRS, porté par R. Haser (cristallographe à l'IBCP, Lyon) et pour lequel le laboratoire a été sollicité, était d'étudier la structure de Tat non plus seule ou sous forme de peptide associé à TAR, mais au sein du complexe TAR/Tat/cycline T1. De cette manière, il semblait possible à la fois d'éviter les problèmes de flexibilité de la protéine Tat étudiée seule, mais aussi d'obtenir la structure de Tat dans une de ces conformations actives les plus importantes. Cette hypothèse était d'autant plus raisonnable, qu'au sein de ce complexe, des résidus Cys appartenant à Tat et à la cycline T1 participent à la formation d'une structure doigt à Zn intermoléculaire (Garber *et al.*, 1998) et que ces deux protéines se fixent à TAR de manière coopérative (Richter *et al.*, 2002). Bien que la transactivation de la transcription virale soit dépendante de l'association de l'élément TAR à la protéine Tat et au facteur p-TEFb, qui contient la cycline T1 mais aussi la protéine kinase CDK9, cette dernière n'avait pas été prévue dans le complexe produit d'une part pour simplifier l'étude, et d'autre part car aucune interaction n'avait été mise en évidence entre Tat et CDK9.

Des industriels étaient associés à ce projet (Pierre Fabre et Sanofi). La protéine Tat nécessaire à cette étude devait être synthétisée chimiquement par l'équipe de A. Beck (Institut de Recherches Pierre Fabre, St Julien en Genevois). Notre laboratoire était en charge de la production de l'ARN TAR et de la cycline T1 et l'étude structurale devait être effectuée par l'équipe de R. Haser (IBCP, Lyon). Nous avions aussi en charge de tester la formation et la qualité des complexes.

2. Résultats obtenus

2.1. Production de l'ARN TAR

2.1.1. Le choix de l'ARN produit et de son mode de production

L'étude cristallographique de complexes ARN/protéines impose un certain nombre de contraintes en ce qui concerne la production de l'ARN. Celui-ci doit être totalement pur, sans traces de nucléases, mais aussi parfaitement homogène en terme de taille et de repliement.

Les techniques traditionnelles de transcription *in vitro* sont basées sur l'utilisation de promoteurs transcriptionnels de phages qui assurent le recrutement des ARN polymérases de ces phages, enzymes qui sont commercialisées. Les systèmes les plus fréquemment utilisés sont ceux issus des bactériophages T7, SP6 et T3 (Fig. 57).

Fig. 57 : Séquences consensus des promoteurs transcriptionnels des bactériophages T7, SP6 et T3.

La base en gras correspon à l'origine de transcription est donc au nucléotide +1 dans l'ARN issu de la transcription. La séquence soulignée indique la séquence minimum requise pour une transcription efficace, le reste des bases est nécessaire pour une efficacité optimale.

La terminaison de la transcription, qui est un mécanisme relativement complexe dans la cellule, est résolue *in vitro*, le plus souvent, par arrêt de la transcription à l'extrémité de la matrice. Ainsi, la transcription *in vitro* se fait en général à partir d'un produit PCR contenant le promoteur en 5', ou d'un plasmide linéarisé par coupure par une enzyme de restriction en aval de la séquence à transcrire.

Ce mode de production de l'ARN par transcription *in vitro* a deux inconvénients majeurs pour l'étude des ARN par radio-cristallographie. Le premier provient du fait que la spécificité très étroite de reconnaissance des séquences promotrices par les ARN polymérases de phages impose pour que la transcription soit efficace, la présence de quelques résidus en 5' de l'ARN (Fig. 57), ce qui peut entraîner des modifications de la structure du transcrit par rapport à celle de l'ARN produit *in vivo*. Par ailleurs, l'arrêt de la transcription réalisée *in vitro* n'est pas un mécanisme précis. Des nucléotides peuvent être ajoutés ou non transcrits, conduisant à des produits de transcription de taille différentes +1, +2 ou -1, -2 nucléotides.

La technique que nous avons employée pour contourner ces obstacles, a été de produire *in vitro* un transcrit primaire dans lequel la séquence TAR était flanquée de deux ribozymes à tête de marteaux (Fig. 58A). Ces structures ribozymes incorporées catalysent leur propre clivage de manière très précise, ce qui permettait de libérer l'ARN TAR avec des extrémités

5' et 3' bien définies (Price *et al.*, 1995). Il était par ailleurs possible dans ces conditions d'utiliser un promoteur dont la séquence permettait une transcription d'efficacité optimale, puisque la séquence en 5' de l'ARN était ensuite éliminée par le clivage du ribozyme. Ces structures ribozymes, qui sont trouvées naturellement dans certaines régions des ARN satellites de virus des plantes (Haseloff and Gerlach, 1989), n'imposent que la présence de résidus uridines situés en 3' de TAR pour que le ribozyme en 3' fonctionne (Fig. 58B).

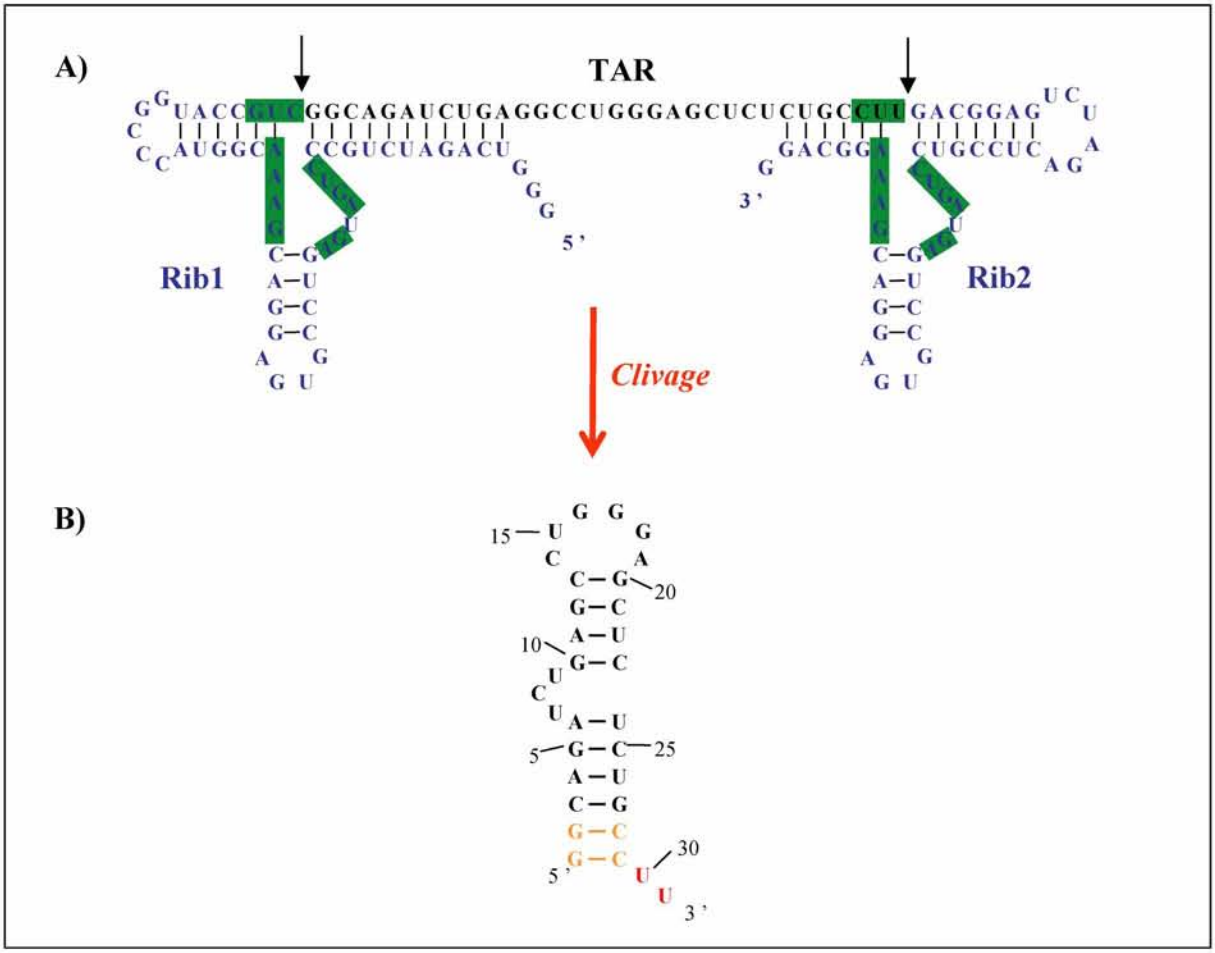


Fig. 58 : Stratégie de production de l'ARN TAR.

A) Séquence et structure secondaire du transcrit produit par l'ARN polymérase du phage T7. Les séquences encadrées sont les régions conservées des ribozymes qui sont nécessaires au clivage. Les flèches indiquent les liaisons phosphodiester clivées. B) Structure secondaire de l'ARN TAR libéré par clivage.

2.1.2. Production et purification de l'ARN

-Production de l'ADN matrice

Six oligonucléotides s'hybridant les uns aux autres ont permis par un jeu de PCR successives d'obtenir l'ADN codant l'ARN hybride sous contrôle du promoteur de l'ARN polymérase du

phage T7 (Fig. 59). Les 3 oligonucléotides 5'-3' ont été conçus de façon à ce qu'ils puissent être réutilisés pour la génération d'autres constructions de matrice visant la production d'autres ARN hybrides, autrement dit seuls les 3 oligonucléotides 3'-5' étaient spécifiques de la séquence TAR à produire. Le fragment d'ADN matrice ainsi obtenu a ensuite été cloné dans le plasmide pUC19 (Fermentas), ceci afin de produire de grandes quantités de la matrice ADN et éventuellement de modifier aisément la séquence de l'ARN TAR codé par mutagénèse dirigée.

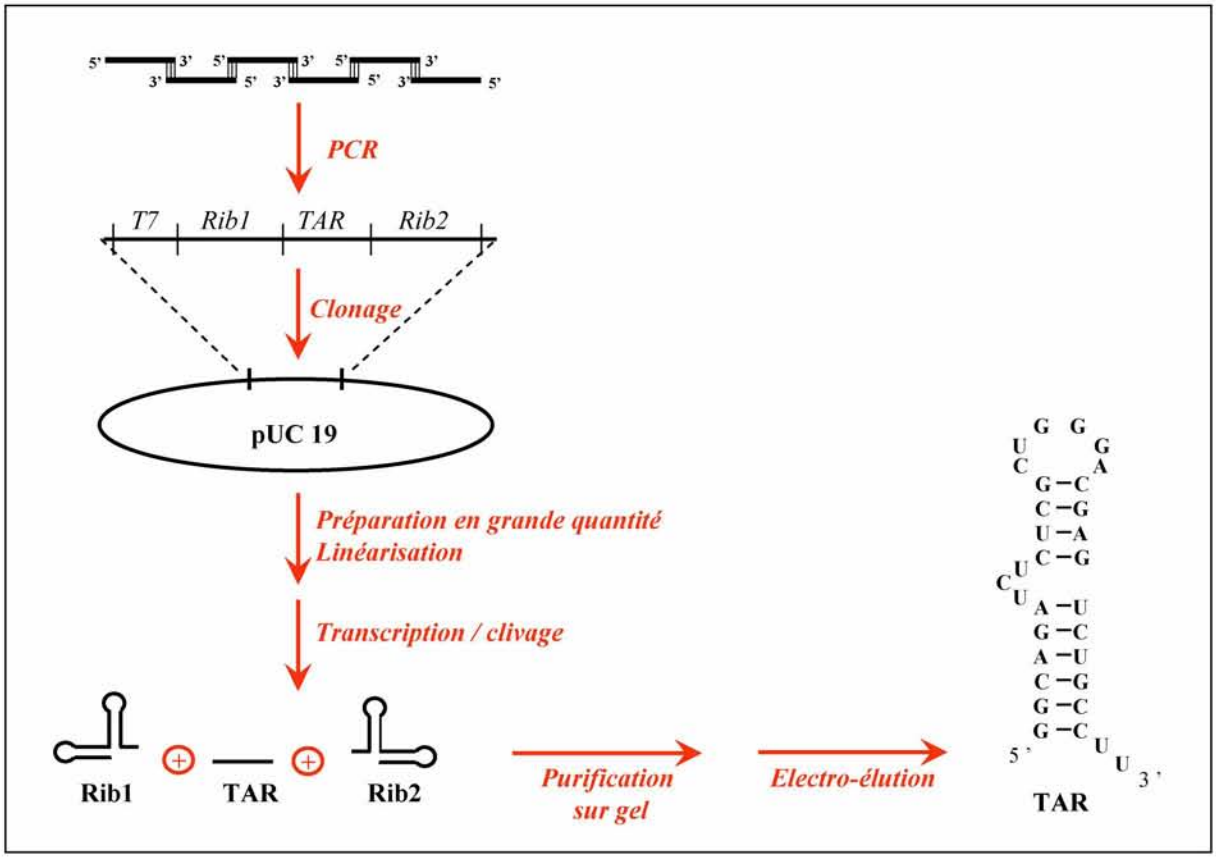


Fig. 59 : Protocole de production de l'ARN TAR.

La première étape consiste en la production du gène T7-Rib1-TAR-Rib2 par un jeu de PCR successives et son clonage dans le vecteur pUC19 (Fermentas). Le plasmide est ensuite produit en grande quantité et linéarisé, puis soumis à l'étape qui couple transcription et clivage. L'ARN TAR est ensuite purifié par une électrophorèse en gel dénaturant et enfin électro-élué.

-Production de l'ARN

Pour pouvoir produire de grandes quantités d'ARN hybride, il était nécessaire d'utiliser de grandes quantités d'ARN polymérase T7, nous avons donc produit cette enzyme sous forme recombinante en grande quantité à partir d'un clone disponible au laboratoire. Un protocole de transcription *in vitro* à grande échelle a ensuite dû être mis au point.

L'activité autocatalytique des ribozymes est Mg^{2+} dépendante et a lieu de manière concomitante à l'étape de transcription. Nous avons donc testé quelle était la concentration optimale en Mg^{2+} pour obtenir à la fois une transcription efficace et un bon rendement de clivage des ribozymes. Cette étude nous a permis de mettre en évidence : (1) que le clivage du ribozyme 1 (en 5') est quasiment totale même sans Mg^{2+}, et (2) qu'à l'opposé le clivage du ribozyme 2 (en 3') est beaucoup moins efficace mais que son efficacité augmente avec la concentration en Mg^{2+}, jusqu'à une valeur maximum de 25 mM, condition que nous avons donc sélectionnée pour toutes les productions (Fig. 60).

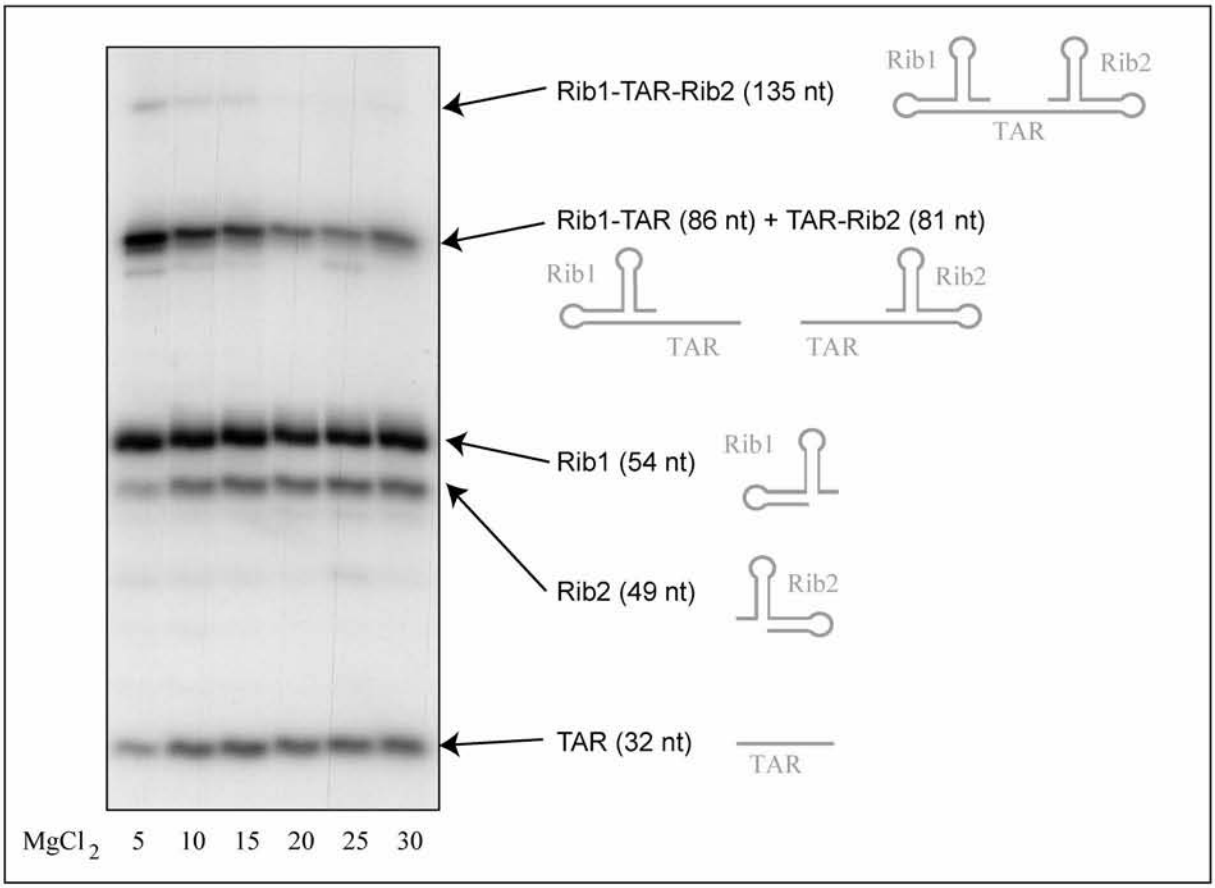


Fig. 60 : Expérience visant à définir la concentration en $MgCl_2$ optimale pour l'étape de transcription/clivage.

La transcription a été effectuée en présence d'UTP [a-32P] pendant 3 h dans un tampon contenant 5 à 30 mM de $MgCl_2$. Les produits de transcription / clivage ont été déposés sur un gel polyacrylamide 14 %, Urée 8 M.

Une hypothèse pouvait être que la formation de la structure secondaire du ribozyme 2 puisse être en compétition avec celle de TAR, contrairement au ribozyme 1 qui lui est synthétisé avant TAR, ce qui doit lui permettre de se structurer correctement. Afin d'atténuer la perte de

matériel due à la non coupure du ribozyme 2, nous avons tenté de dénaturer et renaturer les produits de la transcription afin d'augmenter l'efficacité de clivage du ribozyme en aval de TAR. Néanmoins, ce traitement n'a apporté aucune amélioration du rendement de production de TAR.

Les autres paramètres de réaction qui ont été testés sont les concentrations de matrice linéarisée, d'ARN Polymérase T7 et de nucléotides ; 3 constituants qu'il nous fallait minimiser pour des questions de coût et de temps de travail, tout en conservant un bon rendement de transcription. Ainsi nous avons pu diminuer les quantités de T7 utilisées car elle n'était pas limitante dans les conditions de base, mais nous n'avons pu modifier la quantité de matrice (100 µg/ml final) car le rendement en était directement dépendant. L'étude de la concentration de nucléotides nous a permis de constater un meilleur rendement lorsque les quantités de chacun étaient pondérées en fonction de la composition de l'ARN voulu. Au final, grâce à cette modification, nous avons pu diminuer la concentration moyenne à 2 mM final par nucléotide. La purification de l'ARN TAR libéré par auto-clivage des ribozymes a été réalisée par électrophorèse en gel dénaturant, suivie d'une extraction de l'ARN par électro-élution et une étape de précipitation comme décrit dans le chapitre Matériel et méthodes.

Après de nombreuses expériences visant à optimiser à la fois la transcription, le clivage et la purification de l'ARN, cette méthode nous a permis d'obtenir 0,5 à 1 mg d'ARN TAR pur par expérience de transcription, ce qui était largement suffisant en vue de l'étude cristallographique envisagée.

Il était alors important de tester si l'ARN adoptait la structure secondaire escomptée. Pour cela, j'ai employé les RNases V1, T1 et T2 en tant que sondes de la structure 2D des ARN. La nucléase V1 est spécifique des régions appariées ou empilées, les RNases T1 et T2 sont utilisées dans des conditions où elles coupent préférentiellement les régions en simple brin. Au final, l'ARN obtenu adoptait bien la structure secondaire attendue (Fig. 61). Par ailleurs, aucune dégradation n'a été constatée au cours du temps lorsque l'ARN était conservé à -20°C, ce qui nous permettait d'envisager des tests de cristallisation.

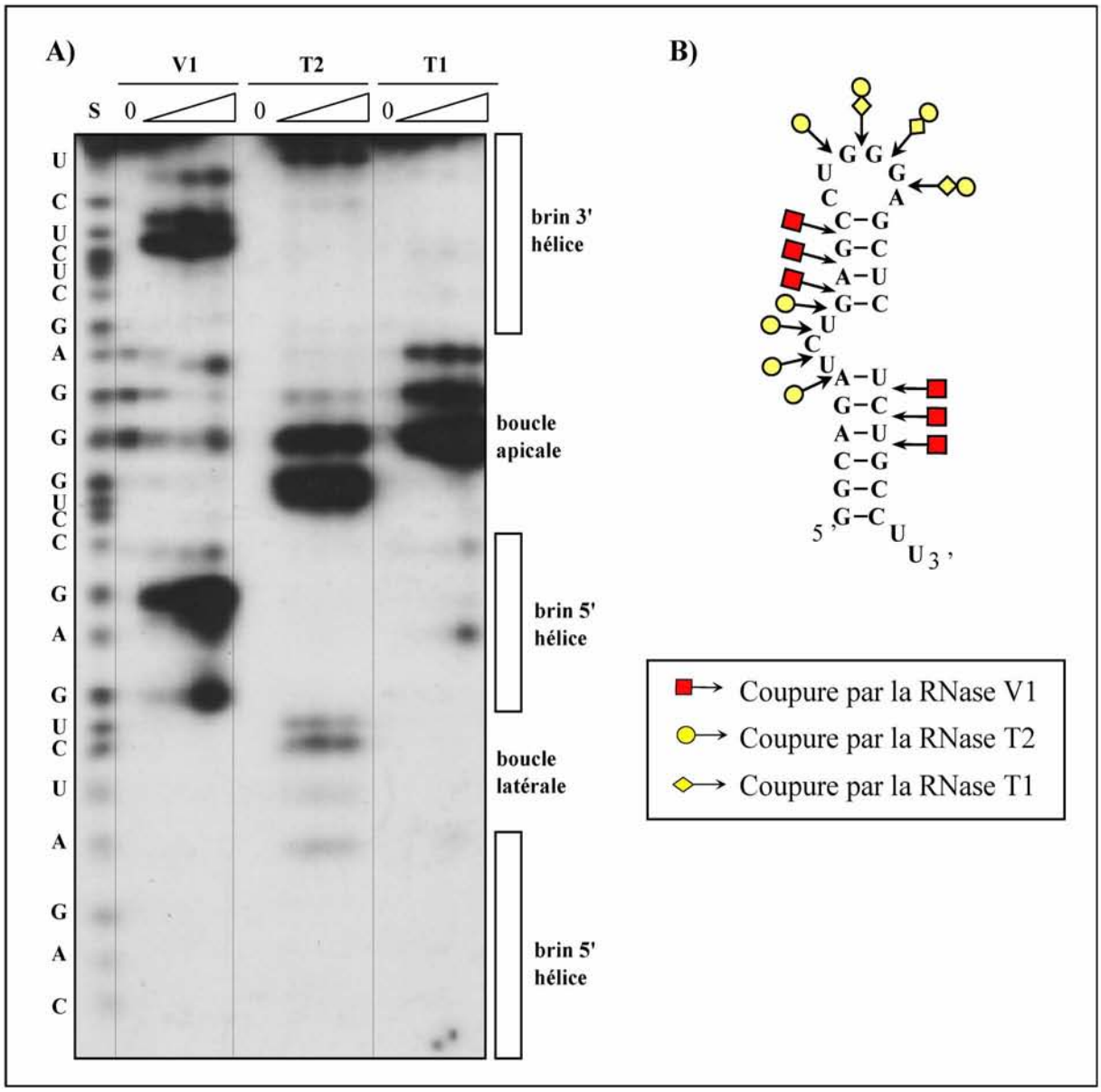


Fig. 61 : Analyse de la structure secondaire en solution adoptée par l'ARN TAR produit.

A) L'ARN TAR issu du protocole de transcription/clivage décrit précédemment a été marqué en 5' puis soumis soit à une hydrolyse au bicarbonate de sodium (piste S), soit à des hydrolyses ménagées par les RNases V1 (0, 4, 8, 16.10^{-5}U), T2 (0, 0.5, 1, 1.5 U) et T1 (0, 0.5, 1, 1.5 U) comme décrit dans la partie Matériel et Méthodes. B) Les différentes coupures représentées selon la symbolique indiquée dans l'encart ont été positionnées sur la structure secondaire théorique de l'ARN.

2.2. Etude de la fixation de la protéine Tat sur l'ARN TAR produit

Deux « formes » de la protéine Tat avaient été synthétisées chimiquement par l'équipe de A. Beck :

- la protéine Tat 2.1 de 101 acides aminés, provenant de la souche ACH320.2A.2.1 du VIH-1,
- et la protéine Tat 133 de 101 acides aminés, issue d'une souche virale primaire isolée du patient 133 de la cohorte de patients de l'Hôpital de la Croix-Rousse (Lyon).

Ces protéines ont été incubées à température ambiante en présence d'ARN TAR marqué en 5' avec de l'ATP [γ-32P], des ARNt compétiteurs (0,1 µg/µl final) dans le tampon D (HEPES pH 7.9 20 mM, KCl 100 mM, Glycérol 20%, EDTA 0.2 mM, DTT 0.5 mM, PMSF 0.5 mM), puis fractionnées sur gel acrylamide 4% natif comme décrit dans le chapitre Matériel et Méthodes. De manière reproductible, nous pouvions observé une bande correspondant à la formation du complexe TAR/Tat pour les concentrations en protéine Tat 2.1 de 3 et 8 µM (Fig. 62). Néanmoins, à de plus fortes concentrations, l'ARN était réparti le long de la piste suggérant un phénomène d'agrégation.

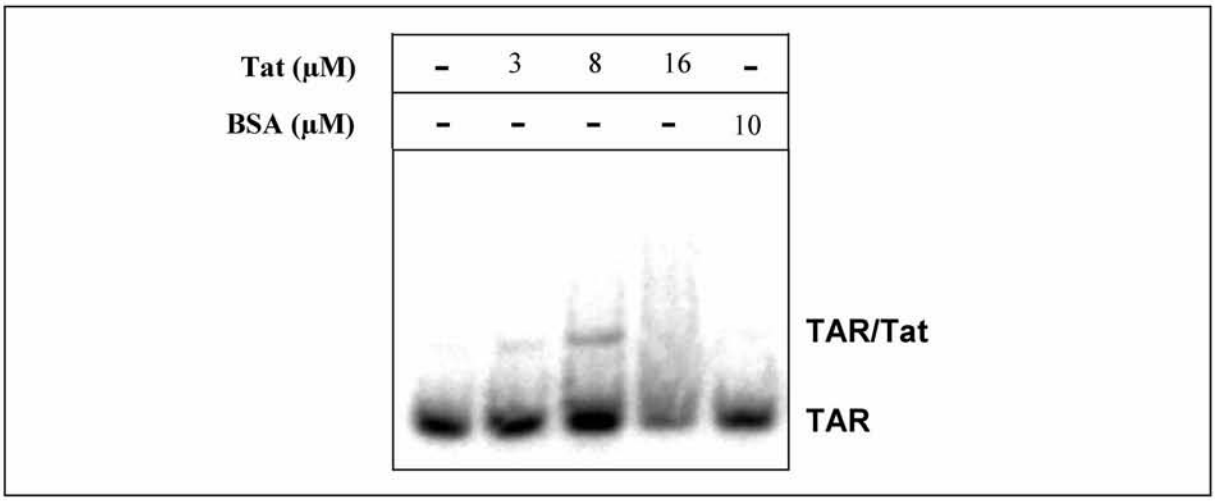


Fig. 62 : Etude de la fixation de la protéine Tat 2.1 sur l'ARN TAR par expérience de retard sur gel.

Analyse par expériences de retard sur gel des interactions entre l'ARN TAR et la protéine Tat 2.1 et la BSA (en témoin négatif). L'ARN marqué en 5' (de l'ordre de 10 nM) est incubé pendant 10 min à 25°C en présence de concentrations croissantes de protéine Tat (0 à 16 µM), puis les complexes sont fractionnés sur gel non dénaturant à 4%. L'ARN libre et les complexes formés sont indiqués sur les cotés.

Le taux de complexe ainsi obtenu étant trop faible pour pouvoir envisager une analyse structurale, nous nous sommes demandés si le faible taux de complexe n'était pas lié à une mauvaise qualité de la protéine synthétisée chimiquement . Nous avons alors testé plusieurs autres protéines Tat :

- la protéine synthétique Tat 86 de 86 acides aminés, provenant de la souche Lai du VIH-1 qui nous a été fournie par S. Muller (IBMC, Strasbourg) (Chaloin *et al.*, 2005),
- la protéine recombinante GST-Tat 101 de 101 acides aminés, provenant de la souche Bru du VIH-1 qui nous a été fournie par M. Benkirane (IGH, Montpellier),
- et la protéine recombinante GST-Tat 101 recombinante de 101 acides aminés, provenant de la souche Bru du VIH-1 dont le clone avait été produit au laboratoire par S. Massenet.

En plus de ces tests sur la nature des protéines, nous avons fait varier plusieurs des paramètres de la réaction de retard sur gel :

- En premier lieu, l'ARN a subit plusieurs types de traitements de dénaturation suivis d'une renaturation (chauffage 5 min à 95°C puis refroidissement 5 min dans la glace, ou chauffage 10 min à 65°C puis refroidissement progressif jusqu'à température ambiante ou aucun traitement).

- Les effets de l'utilisation de différentes concentrations de protéine et d'ARN, ainsi que de différents ratios entre protéine Tat et d'ARN TAR ont été étudiés.

- Plusieurs solutions ont été utilisées pour la formation du complexe, nous avons fait varier la concentration saline globale (de 100 à 300 mM final), celle en $MgCl_2$ (de 0 à 5 mM final), l'effet de la présence ou non de détergents (ex. NP-40), de DTT ou de glycérol ont été testés.

- L'influence de la concentration en ARNt compétiteurs a aussi été évaluée (0 à 100 ng/µl).

- Et enfin, de nombreuses conditions d'électrophorèse ont été testées, ceci en faisant varier la température de migration (temp. ambiante ou à 4°C, tampon de migration TBE ou TG, type de gel agarose ou acrylamide).

Malgré tous ces essais, le rendement de fixation de Tat sur TAR n'a pas été amélioré. Par contre, nous avons toujours observé une meilleure affinité pour l'ARN TAR des protéines Tat recombinantes par rapport aux protéines Tat synthétiques.

Bien que n'étant pas favorables, ces résultats étaient néanmoins en accord avec les données obtenues par les différents groupes travaillant sur les caractéristiques physicochimiques de l'interaction entre la séquence TAR et la protéine Tat complète (Garber *et al.*, 1998 ; Marozzi *et al.*, 1998, Bres *et al.*, 2002 ; Fujinaga *et al.*, 2002 ; Endo-Munoz *et al.*, 2005 ; Chaloin *et al.*, 2005, Marquez *et al.*, 2006), et les quelques expériences permettant d'avoir de meilleurs rendements utilisaient en fait des peptides de Tat (Delling *et al.*, 1992 ; Smith *et al.*, 1998).

Les données de la littérature indiquaient par contre que l'interaction TAR/Tat est fortement renforcée en présence de cycline T1 (Garber *et al.*, 1998 ; Fujinaga *et al.*, 2002). C'est pourquoi nous avons essayé de produire le fragment minimum de la cycline T1 décrit pour renforcer l'interaction TAR/Tat.

2.3. Production du fragment N terminal de la cycline T1

Comme nous l'avons décrit dans l'introduction de ce manuscrit, le segment correspondant aux 272 acides aminés N-terminaux de la cycline T1 est suffisant pour établir une interaction avec l'ARN TAR et la protéine Tat. Initialement, M. Benkirane nous avais fourni un clone permettant de produire un fragment 1-300 de la cycline T1, fusionné à la GST, mais avec cette construction, nous nous sommes heurtés à de nombreux problèmes d'insolubilité de la protéine produite. Elle se retrouvait essentiellement dans des corps d'inclusion après sonication, et le rendement de production était faible, ceci quelle que soit la souche d'*E. coli* (BL21 DE3, BL21 DE3 codon + et BL21 DE3 PLysS) et les températures de culture utilisées (16, 25 ou 37°C) (Fig. 63).

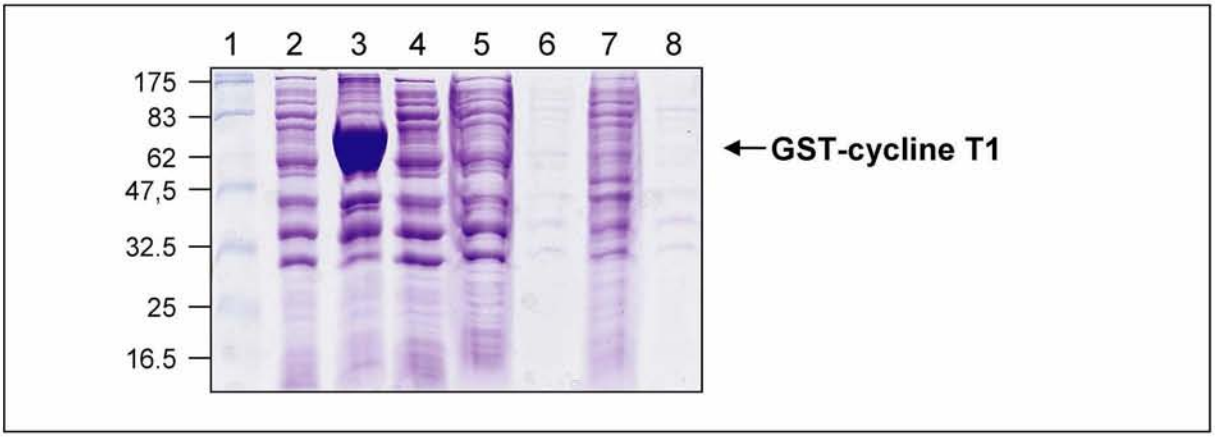


Fig. 63 : Analyse de la purification de la fusion GST-cycline T1 (1-300).

L'analyse a été réalisée par fractionnement des protéines en gel de polyacrylamide dénaturant à 12,5%, suivie d'une coloration au bleu de Coomassie. Les pistes 2 et 3 contiennent respectivement les fractions soluble et insoluble obtenues après sonication et centrifugation de la culture après induction. La fraction soluble a ensuite été fractionnée sur billes de glutathione sepharose. Les pistes suivantes représentent la fraction non retenue sur les billes (piste 4), les fractions de lavage (pistes 5 et 6) et celles d'élution (pistes 7 et 8). Les tailles (en kDa) des marqueurs (piste 1) sont indiquées sur la gauche du gel.

Nous avons alors décidé de produire le fragment de la cycline T1 (1-280) qui avait été décrit comme contenant tous les éléments de reconnaissance du complexe TAR/Tat et comme étant soluble (Fujinaga *et al.*, 2002). La région d'ADN codant ce fragment de la cycline T1 a été clonée dans les plasmides pGex 2TK et pET 28b. Une fusion de ce fragment d'ADN avec l'ADN codant la protéine Tat 101 a aussi été produite et clonée dans le vecteur pGex 4T1. Une telle construction avait déjà été utilisée par l'équipe de BM. Peterlin et leurs données indiquaient que la fusion Tat-cycline T1 obtenue était active et soluble du moins aux concentrations utilisée pour la réalisation des expériences de biologie moléculaire (Fujinaga *et al.*, 2002).

Ces différents vecteurs plasmidiques ont été choisis en tenant compte de leurs caractéristiques particulières. Les vecteurs pGex (Amersham Biosciences) permettent de produire des protéines fusionnées à la GST (Glutathion S Transferase), qui peut améliorer la solubilité des protéines à laquelle elle est fusionnée. Le vecteur pET 28b (Novagen) permet d'exprimer des protéines fusionnées à une étiquette comportant 6 résidus His. C'est le système qui est actuellement le plus utilisé pour la synthèse de protéines destinées à la radiocristallographie.

Après de multiples essais dans différentes souches d'*E. coli* (BL21 DE3; BL21 DE3 codon +; BL21 DE3 pLysS, Stratagene) et le test de plusieurs conditions de culture, il s'est avéré qu'aucune des constructions dérivées des vecteurs pGex ne permettait de produire des protéines suffisamment solubles pour réussir une production à grande échelle. C'est donc la construction réalisée dans le vecteur pET, produisant le fragment 1-280 de la cycline T1 fusionné à 6 résidus histidine (pET-T1(1-280)) que nous avons sélectionnée pour la suite des travaux. Nous avons sélectionné la souche BL 21 DE3 codon + pour la production, la croissance cellulaire étant réalisée à 37°C en milieu LB-kanamycine jusqu'à obtention d'une DO de 0,7. Ensuite, une induction de l'expression de la protéine est alors réalisée pendant 3 heures à 30°C avec 0,2 mM IPTG, 100 μM ZnSO4 et 200 μM MgCl2. La présence de l'étiquette 6His nous permettait enfin de réaliser une purification sur colonne Ni^{2+}-NTA agarose (Quiagen).

La Fig. 64 présente le suivi par électrophorèse (SDS PAGE) et par analyse « Western blot » de la production et de la purification de la protéine 6His-cyclineT1. L'examen des gels révèle que: (1) 70% de la protéine produite est présente dans la fraction soluble après lyse par sonication, et que (2) le degré de pureté finale est d'environ 95%.

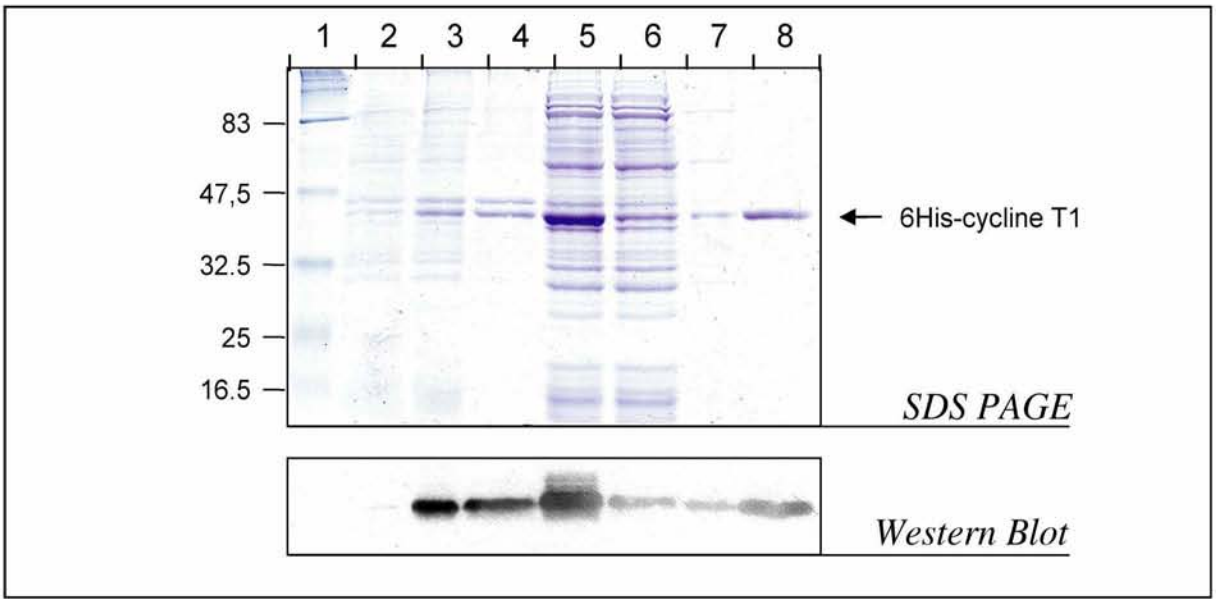


Fig. 64 : Analyse de la purification de la fusion 6His-cycline T

Les protéines ont été fractionnées en gel de polyacrylamide dénaturant à 12,5%, puis mises en évidence par coloration au bleu de Coomassie ou révélée par la technique de Western blot avec un anticorps spécifique de l'étiquette 6His. Les cellules BL21 DE3 codon+ transformées par le plasmide pET-T1(1-280) ont été cultivées à 37°C dans du milieu LB. Des aliquotes de la culture ont été prélevés avant induction (piste 2) et après induction (piste 3),et déposés sur gel après chauffage à 96°C pendant 5 min. Les pistes 4 et 5 correspondent respectivement aux fractions insoluble et soluble obtenues après sonication et centrifugation de la culture à la fin de l'induction. La fraction soluble a ensuite été fractionnée sur billes de Ni^{2+}-NTA. Les pistes 6, 7 et 8 correspondent respectivement à la fraction non retenue sur les billes, la fraction de lavage et celle d'élution. Les tailles (en kDa) des marqueurs (piste 1) sont indiquées sur la gauche du gel.

Quelques mises au point ont ensuite été réalisées au niveau des étapes de purification : notamment l'utilisation de PEI pour éliminer les acides nucléiques, puis le passage sur une colonne HisTrap 5 ml (Amersham Biosciences) avec l'optimisation des conditions de lavage (NaCl à 2 M, puis à 300 mM) et enfin le dessalage par gel filtration sur colonne Superdex 75 (GE Healthcare) et la concentration sur Centriprep (Millipore). Ces diverses modifications ont permis d'adapter le protocole à la production de la cycline T1 à grande échelle, ceci tout en conservant le rendement et la qualité de la préparation obtenus à petite échelle. Ainsi, le protocole final qui est explicité dans le chapitre Matériel et méthodes, permet de produire 2 mg de protéine 6His-cyclineT1 par litre de culture, ceci à une concentration de 10 mg/ml dans un tampon Hepes 50 mM pH 8,5 ; NaCl 300 mM, ce qui représentaient une quantité et une concentration suffisantes pour envisager des essais de cristallisation. Etant donné qu'à

l'époque aucune structure 3D n'était disponible pour la cycline T1, nous avons essayé de cristalliser le fragment ainsi préparé avec C. Charron (IR cristallographe au laboratoire). Une telle étude pouvait aussi nous renseigner sur les potentialités de cristallisation du complexe TAR/Tat/cycline T1.
Trois séries de tests de cristallisations représentant 1000 conditions de cristallogénèse ont été effectuées.

La première série de tests a été non concluante car aucun cristal, pas même un précipité, n'a été détecté, ce qui laissait penser que la concentration de la protéine (9 mg/ml) était trop faible. La deuxième série de tests, effectuée avec une concentration de protéine plus élevée (16 mg/ml), a effectivement permis de visualiser des précipités dans certaines des conditions de cristallisation testées, mais aucun cristal n'a été obtenu. Ceci pouvait être dû au fait que le fragment ainsi concentré présentait des signes d'insolubilité lors de la formation des gouttes lors des essais de cristallisation. D'autre part la concentration saline de l'échantillon de départ était élevée (NaCl 300 mM) par rapport à celles des solutions utilisées pour réaliser les tests de cristallisation. Pour remédier à ce problème, nous avons alors testé la possibilité de solubiliser la protéine dans d'autres tampons en vue de préserver sa stabilité tout en diminuant la concentration saline. Afin de ne pas perturber l'ensemble du processus de purification déjà établi, nous avons décidé d'effectuer le changement de tampon par dialyse après l'étape de purification sur la colonne d'affinité HisTrap.
Différentes dialyses ont donc été effectuées en faisant varier le pH du tampon, la concentration saline et la nature des ions et contre-ions. Chacune de ces expériences a été réalisées sur des aliquotes d'un même lot de protéine purifiée, de sorte qu'après dialyse et centrifugation des dialysats, les différentes fractions solubles ont pu être comparées par dépôt sur gel SDS PAGE et par calcul des concentrations finales de fragment (1-280) de cycline T1.
L'analyse des effets des variations de pH a révélé une perte de solubilité de la protéine aux pH extrêmes (5.6 et 10.5) et une augmentation du phénomène de multimérisation dans la gamme basique (8.5, 9.5 et 10.5) d'autre part. De manière assez surprenante, compte tenu du pHi de la protéine (pHi=6.83), c'est à des pH de 6.5 et 7.5 que nous avons obtenu la meilleure solubilité et c'est finalement la valeur de pH 7.5 qui a été retenue (Fig. 65A). En ce qui concerne l'identité des ions et contre-ions, 6 contre-ions (citrate, formate, acétate, cacodylate, HPO_4^{2-} et SO_4^{2-}) et 5 ions (NH_4^+, K^+, Li^+, Mg^{2+} et Ca^{2+}) ont été testés, couplés respectivement avec l'ion Na^+ et le contre-ion Cl^-. Bien que l'utilisation de Na_2HPO_4 ou de Na_2SO_4 ne

semblait pas perturber la solubilité du fragment de cycline T1 lors de la dialyse, c'est le citrate de sodium qui a été sélectionné pour la suite des tests (Fig. 65B).

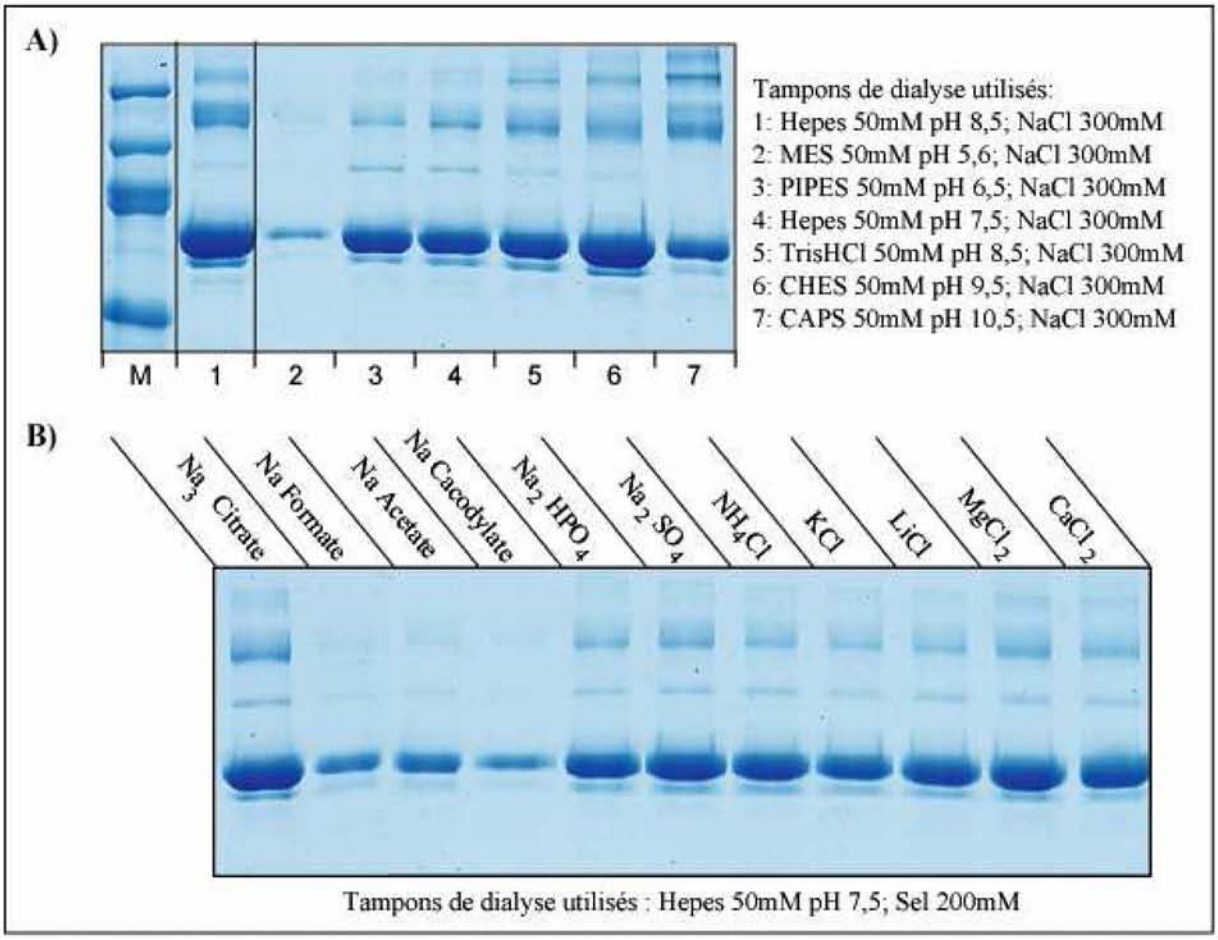


Fig. 65 : Test de l'effet du pH et de la nature des ions et contre-ions sur la solubilité de la protéine 6His-cycline T1

A) Analyse des effets induits par le changement de pH lors de la dialyse. Les aliquotes d'un même lot de protéine purifiée ont été dialysés contre le tampon de lyse (piste 1) ou contre 6 autres tampons dont le pH variait de 6,5 à 10,5 (piste 2 à 7). Les protéines ont ensuite été fractionnées en gel de polyacrylamide dénaturant à 12,5%. (La piste M correspond au marqueur de taille).

B) Analyse des effets induits par le changement de la nature du sel lors de la dialyse. Les aliquotes d'un même lot de protéine purifiée ont été dialysés contre divers tampons dont la composition en sel variait comme indiqué en haut de chaque piste, puis les protéines ont ensuite été fractionnées en gel de polyacrylamide dénaturant à 12,5%.

Au cours d'une dernière étape d'optimisation nous avons encore essayé de trouver le meilleur compromis entre la baisse nécessaire de la concentration saline du tampon de dialyse et la nécessité de conserver un degré suffisant de solubilité de la fusion 6His-cycline T1. Des

dialyses ont à nouveau été effectuées sur des fractions aliquotes d'un même lot de protéine purifiée, une étape de concentration a été ajoutée au protocole afin de vérifier la stabilité de la protéine à haute concentration (Fig. 66). Comme attendu, la baisse de la concentration saline affecte fortement la solubilité de la protéine mais une concentration de 50 mM de citrate de sodium, nous a parue satisfaisante, car même après concentration à 13 mg/ml la protéine restait soluble.

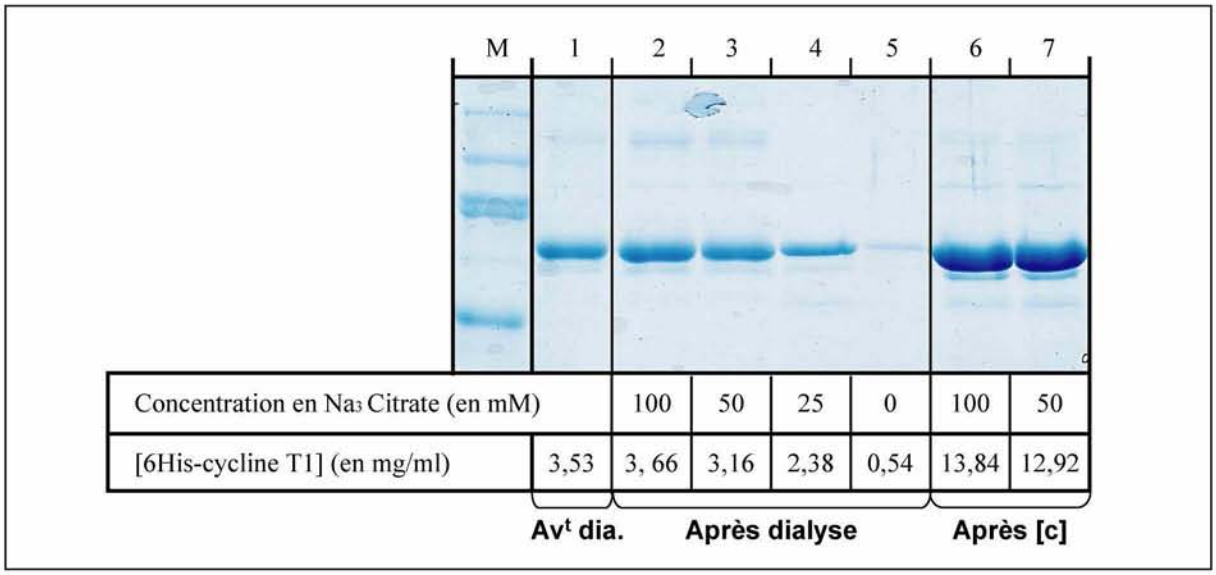


Fig. 66 : Test de l'effet de la concentration saline sur la solubilité de la protéine 6His-cycline T1.

Fractionnement en gel de polyacrylamide dénaturant (12,5%) de 5 aliquotes d'un même lot de protéine purifiée ayant subi différents traitements: le premier aliquote a été conservé dans le tampon d'élution (Hepes 50 mM pH 8,5, NaCl 300 mM, Imidazole 500 mM) (piste 1), les 4 autres ont été dialysés contre 4 tampons Hepes 20 mM pH 7,5 dont la concentration en citrate de sodium variait de 0 à 100 mM (pistes 2 à 5). Pour 2 de ces conditions (100 et 50 mM), les aliquotes ont ensuite été concentrés 8× (pistes 6 et 7). Les concentrations finales des protéines obtenues après les différents traitements sont indiquées.

Nous aboutissions ainsi à des conditions qui paraissaient satisfaisante pour la cristallisation (bonne stabilité de la protéine à une concentration permettant de réaliser des tests de cristallogénèse et concentration saline compatible avec les solutions de cristallisation). Néanmoins, aucun cristal n'a pu être obtenu lors de tests de cristallogénèse.

Cette tentative de cristallisation a été la dernière puisque la structure cristallographique du fragment (1-281) de la cycline T1 humaine fusionnée à un segment de la protéine Tat a été publiée quelques semaines après (Anand *et al.*, 2007). En confirmation de ce que nous avons

obtenu, les auteurs indiquent qu'aucun des essais de cristallisation effectués avec différents fragments de cycline T1 n'avaient donné des cristaux et ceci, selon eux, à cause de problème d'agrégation non spécifique. Tout comme nous, ils ont eu l'idée de fusionner Tat et la cycline T1. Leur choix par contre s'est porté sur le fragment (22-78) de EIAV Tat qui est supposé ne contenir qu'un doigt de Zn formé par 1 résidus His et 2 résidus Cys, plutôt que la protéine Tat de HIV-1 qui possède 7 résidus Cys compliquant potentiellement son expression et son repliement (aspect que nous avions effectivement pu noter en tentant de produire la fusion Cyc T1 (1-280)-Tat 101). Au final, l'ajout de l'élément Tat de EIAV a eu l'effet stabilisateur escompté puisqu'il a permis d'obtenir la structure du fragment 1-281 de la cycline T1 avec une résolution de 2,67 Å (Fig. 67).

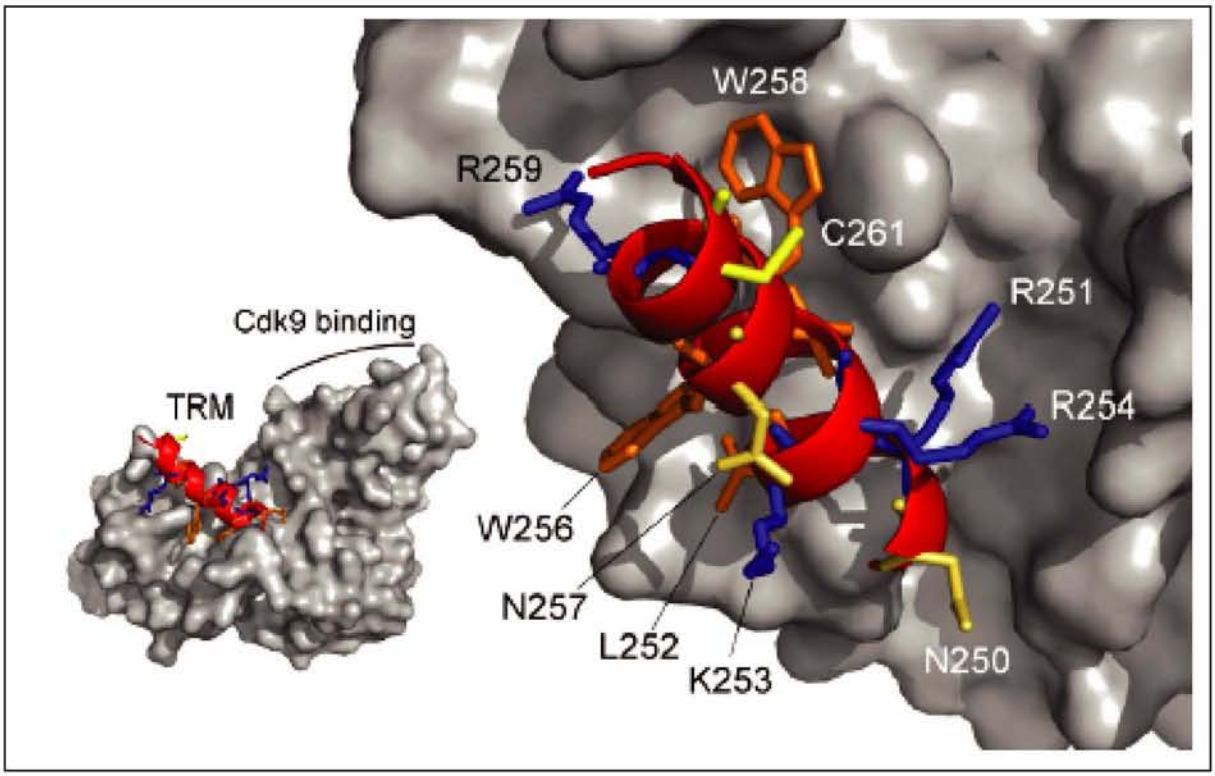


Fig. 67 : Conformation du motif de reconnaissance de TAR/Tat dans la structure 3D de fragment 1-281 la cycline T1

Conformation des résidus 247 à 263 de la région C terminale de la cycline T1. Les résidus chargés positivement K253 et R254, les résidus asparagine polaires N250 et N257, le résidu aromatique W256 et plus particulièrement la cystéine261 sont largement accessibles, contrairement aux résidus R251, L252, I255 et W258 qui font contact avec les répétitions du motif cycline (Anand et al., 2007).

L'analyse de cette structure montre que le domaine N-terminal de la cycline T1 qui, comme pour toutes les cyclines interagit avec les kinases cycline-dépendantes dans toutes les familles

de cyclines, présente une structure 3D homologue à celle trouvée dans les autres cyclines. Le domaine situé en aval est plus divergent par rapport à ceux trouvés dans les autres cyclines et apparaît moins structuré, ce qui pourrait expliquer la spécificité de substrat cycline T1/CDK9. La structure 3D établie montre aussi que les résidus du motif TRM qui interagissent avec le complexe TAR/Tat sont partiellement exposés et accessibles pour se fixer à ce complexe. La cystéine 261, et notamment son atome de soufre, sont tout particulièrement bien orientés pour former la structure intermoléculaire en doigt de zinc avec les résidus cystéines de Tat (Fig. 67). Par ailleurs, l'extrémité C-terminale de l'hélice contenant le motif TRM et l'hélice H1 du premier domaine cycline semblent former un sillon chargé positivement. Les auteurs proposent que ce sillon interagisse avec une hélice α qui soit le site de fixation à la fois de Tat et de la protéine Hexim1. Ceci expliquerait pourquoi les fixations de Hexim1 et de Tat soient mutuellement exclusives (Anand *et al.*, 2007).

Si ces travaux ont permis d'obtenir la structure du fragment de cycline T1, ils confirment aussi les difficultés qu'il y a à réaliser une analyse structurale de la protéine Tat. En effet, bien que la protéine de fusion étudiée soit capable de fixer l'ARN TAR, la structure de la partie de Tat présente n'a pu être résolue. Le fragment de EIAV Tat étant trop flexible sa densité électronique n'était pas définie , ceci en accord avec des résultats récents de RMN qui proposent que la protéine Tat de HIV-1 appartienne à la famille des protéines naturellement non repliée (Shojania and O'Neil, 2006). Sa structure 3D serait induite comme nous le pensions par ses interactions avec ses partenaires.

3. Perspectives

Malgré cette publication de la structure de la partie la plus intéressante de la cycline T1, les perspectives de ce travail restent larges. En effet, l'objectif initial qui était de co-cristalliser les différents partenaires du complexe TAR/Tat/cycline T1 n'a encore pas été réalisé par nos concurrents et les nouvelles données confirment bien que c'est sans doute la meilleure voie pour obtenir une structure 3D fonctionnelle de la protéine Tat.

Un important travail de mise au point de la formation du complexe et de sa cristallisation reste à faire, mais un ensemble d'outils et de données sont maintenant disponibles pour y parvenir.

Un point important dont il faudra tenir compte est la séquence de l'ARN TAR utilisé et plus précisément la séquence de la boucle latérale. En effet, deux séquences sont communément utilisées : la séquence UCU présente dans l'ARN de la souche NY5 (pNL4.3), séquence que

nous avons utilisée, et la séquence UUU, qui est celle de la souche LAI (Bru), fréquemment utilisée dans des travaux de la littérature. Bien qu'un grand nombre d'études mutationnelles ait été réalisé sur l'élément TAR, à notre connaissance, aucune d'entre elles n'a comparé la capacité de fixation de protéines Tat issues de différentes souches à ces deux séquences TAR. Deux autres observations renforcent l'idée que l'identité de cette boucle latérale pourrait moduler l'affinité : la boucle latérale de l'élément TAR de HIV-2 ne comporte pas 3 résidus mais un dinucléotide UU, par ailleurs un aptamère issu d'une expérience SELEX réalisées avec le peptide (32-72) de Tat comporte deux boucles latérales l'une correspondant au dinucléotide UU, l'autre au dinucléotide UC (Fig. 68 ; Yamamoto *et al.*, 2000 ; Matsugami *et al.*, 2004). Enfin, il est intéressant de noter qu'en utilisant l'élément TAR de la souche LAI (UUU) et une protéine Tat 86 synthétisée chimiquement, l'équipe de S. Muller a décrit l'obtention d'une fixation très forte de cette protéine sur l'ARN TAR (résultats non publiés). Il sera donc important de tester l'efficacité de formation des complexes avec des ARN présentant les différents types de boucle latérale possibles.

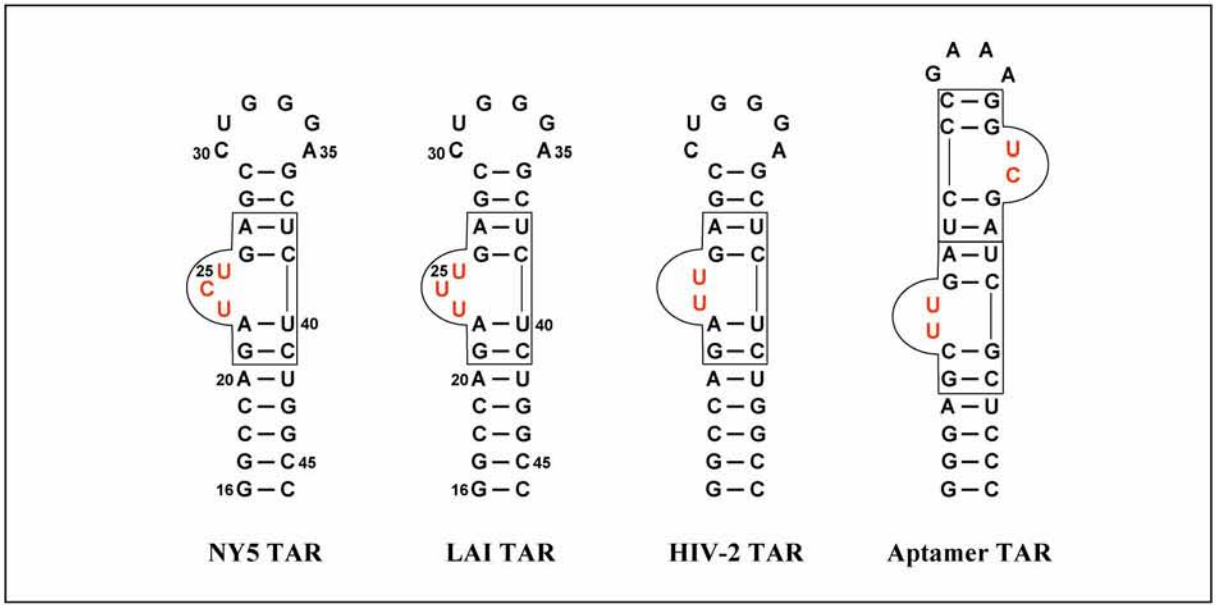


Fig. 68 : Séquences des éléments TAR de différents virus HIV.

Les éléments TAR des souches HIV-1 NY5 et LAI, ainsi que HIV-2, ne diffèrent que par la séquence de leur boucle latérale. Le meilleur candidat issu d'une étude par SELEX des capacités de fixation à l'ARN de Tat, comporte deux sites de fixation. Les sites postulés de liaison de Tat sont encadrés. Les paires de bases supposées impliquées sont indiquées par un trait.

Pour obtenir de grandes quantités de complexe utilisable en radio-cristallographie, il faudra encore essayer d'augmenter la capacité de chacun des partenaires à former le complexe

TAR/Tat/cycline T1. Ainsi, d'autres protéines Tat mutées ou modifiées post-traductionnellement (acétylation, méthylation, ubiquitinylation), pourraient être testées et comparées dans le but de trouver le meilleur candidat possible. Des expériences décrites dans la littérature semblent indiquer que la fixation de fragments de la protéine Tat est meilleure que celle de la protéine complète (ex : peptide CQ (37-72) ; Yamamoto *et al.*, 2000). Une définition de fragments de Tat formant plus favorablement le complexe TAR/Tat/cycline T1 pourrait être réalisée.

CHAPITRE 5 : Etude de l'effet de l'expression des protéines virales Tat, Nef et NCp7 sur les taux d'expression des facteurs cellulaires d'épissage et impact sur l'épissage de HIV-1

1. Etude de l'effet possible de l'expression de la protéine Tat sur les taux de production de protéines SR et hnRNP

1.1. Objectifs

Nous avons vu en introduction, que l'unique ARN pré-messager issu de la transcription du génome du virus HIV-1, contient 9 ORF, dont l'expression est dépendante de l'utilisation régulée de 5 sites donneurs et 9 sites accepteurs d'épissage. Le taux et le moment d'expression de ces ORF dépendent de deux paramètres, d'une part, l'efficacité relative des sites d'épissage et leur régulation et, d'autre part, l'action de la protéine Rev. En effet, comme nous l'avons déjà décrit, les protéines régulatrices sont exprimées pendant la phase précoce, alors que les précurseurs des protéines enzymatiques et structurales et les protéines auxiliaires sont produits en phase tardive. L'existence de ces deux phases dans l'expression des gènes du VIH est largement gouvernée par la protéine Rev qui permet le transport des ARN peu ou non épissés. Dans la phase précoce qui précède l'action de Rev, la régulation de l'utilisation des sites dépend d'un jeu complexe d'interactions entre les régions régulatrices de l'ARN viral entourant les sites d'épissage (ESE, ESS et ISS dans la plupart des cas connus) et les protéines SR et hnRNP. Les effets antagonistes de ces facteurs cellulaires ont pu être mis en évidence à plusieurs de ces sites *in vitro* et *in cellulo*.

Les données obtenues *in cellulo* en surexprimant des protéines SR (Ropers *et al.*, 2004) montraient que des modifications des taux relatifs de ces protéines peuvent altérer les taux relatifs d'utilisation des différents sites d'épissage de l'ARN de HIV-1. La question se posait donc de savoir si l'expression de certaines protéines du virus peut changer le taux d'expression des différentes protéines SR et hnRNP cellulaires, ceci pouvant moduler en retour les taux relatifs d'utilisation des sites alternatifs d'épissage des ARN pré-messagers cellulaires et du virus. Des travaux déjà décrits dans la littérature montraient ainsi que l'infection par le virus HIV-1 modifie le taux d'expression des protéines SR 9G8 et SC35 (Maldarelli *et al.*, 1998 ; Ryo *et al.*, 2000).

D'autres données de la littérature montraient par ailleurs, que la protéine Tat module l'expression d'un certain nombre de gènes cellulaires (voir Introduction). L'idée était donc de

tester la capacité de Tat à modifier le taux d'expression de certaines protéines SR et hnRNP. Pour cela, nous avons à la fois (1) mesuré les variations du taux des ARN messagers correspondant par PCR semi-quantitative, et (2) suivi les variations du taux de protéines dans les noyaux par Western blot.

1.2. Stratégie utilisée pour la production de la protéine Tat

Sous les conseils et avec l'aide d'A. Visvikis qui avait déjà utilisé le système Tet-Off et son vecteur d'expression pTRE (BD Biosciences Clontech), j'ai réalisé une construction pTRE-Tat-Flag permettant d'exprimer la protéine Tat-Flag de manière transitoire dans les cellules HeLa en absence de virus. Ce système Tet-Off est basé sur l'utilisation d'éléments régulateurs essentiels de l'opéron assurant la résistance à la tétracycline chez *E.coli.* Ils ont été utilisés pour générer un système permettant de réguler et d'optimiser l'expression de protéines dans les cellules mammifères:

- les cellules HeLa Tet-Off sont une lignée stable contenant le vecteur pTet-Off-Advanced qui permet d'exprimer la protéine de fusion transactivatrice tTA composée du domaine de liaison à l'ADN TetR capable de se fixer sur l'élément TRE (Tetracycline Resistance Element) et de trois domaines d'activation de la transcription issus de protéines du virus HSV VP16. Cet activateur ne se fixe sur l'élément TRE qu'en absence de doxycycline, un dérivé de la tétracycline.

- le vecteur pTRE contenant l'élément TRE situé en amont d'une version minimale du promoteur précoce du cytomégalovirus (PminCMV) assure l'expression du gène d'intérêt en présence de l'activateur tTA et devient progressivement silencieux en fonction de la concentration en doxycycline qui inhibe l'activité de tTA (Fig. 69A).

L'intérêt de ce système d'expression est qu'il est inductible, sans fuite d'expression et que cette expression est modulable en fonction de la quantité de doxycycline.

En parallèle, nous voulions aussi nous assurer que la protéine Tat exprimée était bien fonctionnelle. Afin de tester le pouvoir de transactivation de la protéine Tat, j'ai construit le plasmide pLTR-Luc contenant le gène de la luciférase placé sous contrôle du promoteur LTR (Fig. 69B).

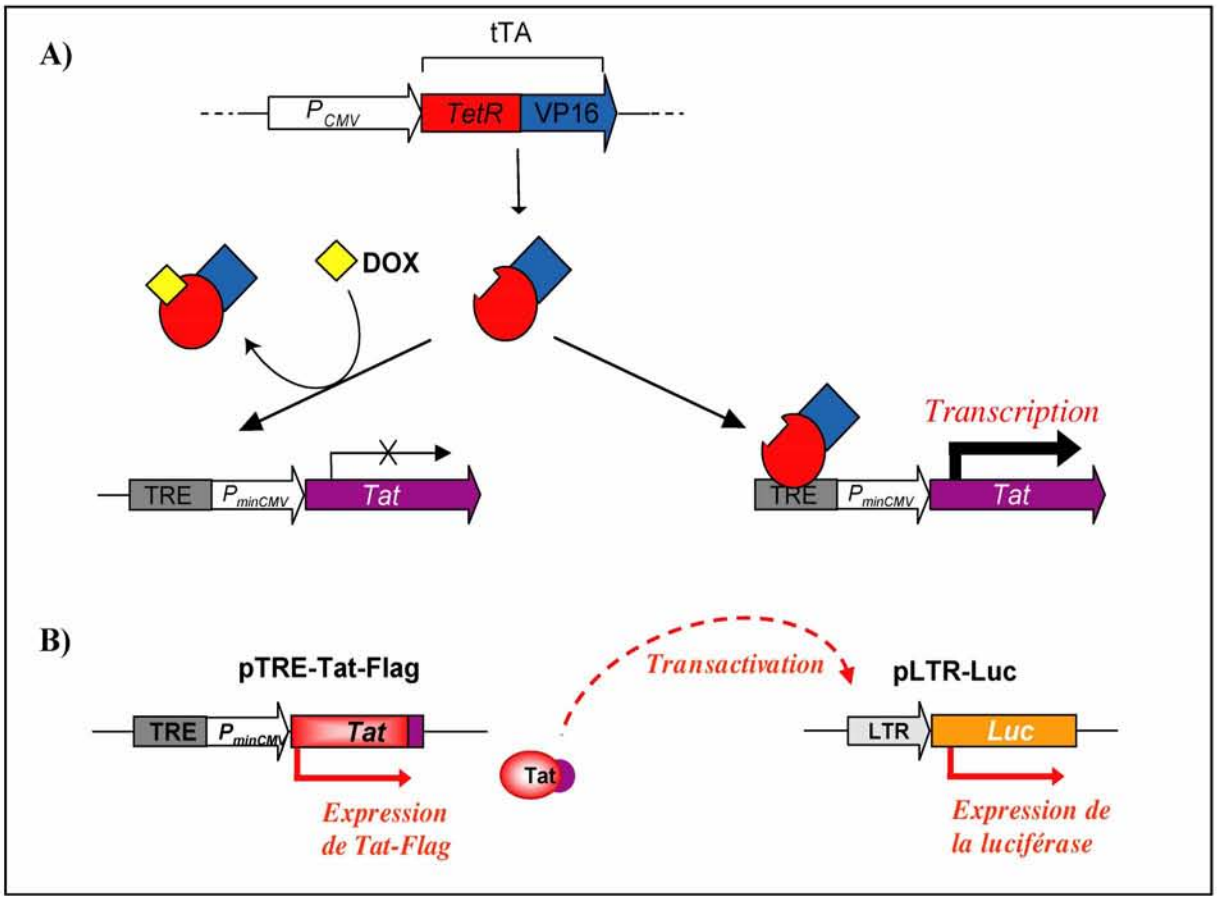


Fig. 69 : Schéma de régulation des gènes du système Tet-Off et principe du test de fonctionnalité de la protéine Tat-Flag exprimée

A) Le gène tTA est sous contrôle du promoteur du cytomégalovirus P_{CMV}. Il est exprimé constitutivement dans les cellules HeLa Tet-Off. La présence de doxycycline permet à tTA de se fixer à l'élément TRE et d'activer le promoteur précoce du cytomégalovirus (P_{minCMV}). B) Principe du test de fonctionnalité de la protéine Tat-Flag exprimée.

Les cellules HeLa Tet-Off ont été co-transfectées par les deux plasmides pTRE-Tat-Flag et pLTR-Luc, et la même expérience a été effectuée en parallèle avec le vecteur pTRE-GGT à la place du plasmide pTRE-Tat-Flag afin d'obtenir un témoin négatif. Le suivi de l'expression transitoire de la protéine Tat (souche HIV-1 Bru) a été réalisé par Western blot, grâce au peptide Flag en C-terminal pour lequel nous disposions d'un anticorps spécifique. Les résultats montraient que la protéine était bien exprimée (Fig. 70A). La mesure de l'activité luciférase obtenue grâce au plasmide pLTR-Luc a permis de confirmer l'effet transactivateur de la protéine Tat produite (Fig. 70B). Enfin, nous avons pu aussi vérifier en utilisant plusieurs concentrations de doxycycline dans le milieu de culture, que nous disposions bien d'un système de production de la protéine Tat dépendant de la concentration en doxycycline

(Fig. 70C). Nous pouvions donc utiliser ce système pour tester l'effet de l'expression de Tat sur les gènes cellulaires des protéines SR et hnRNP.

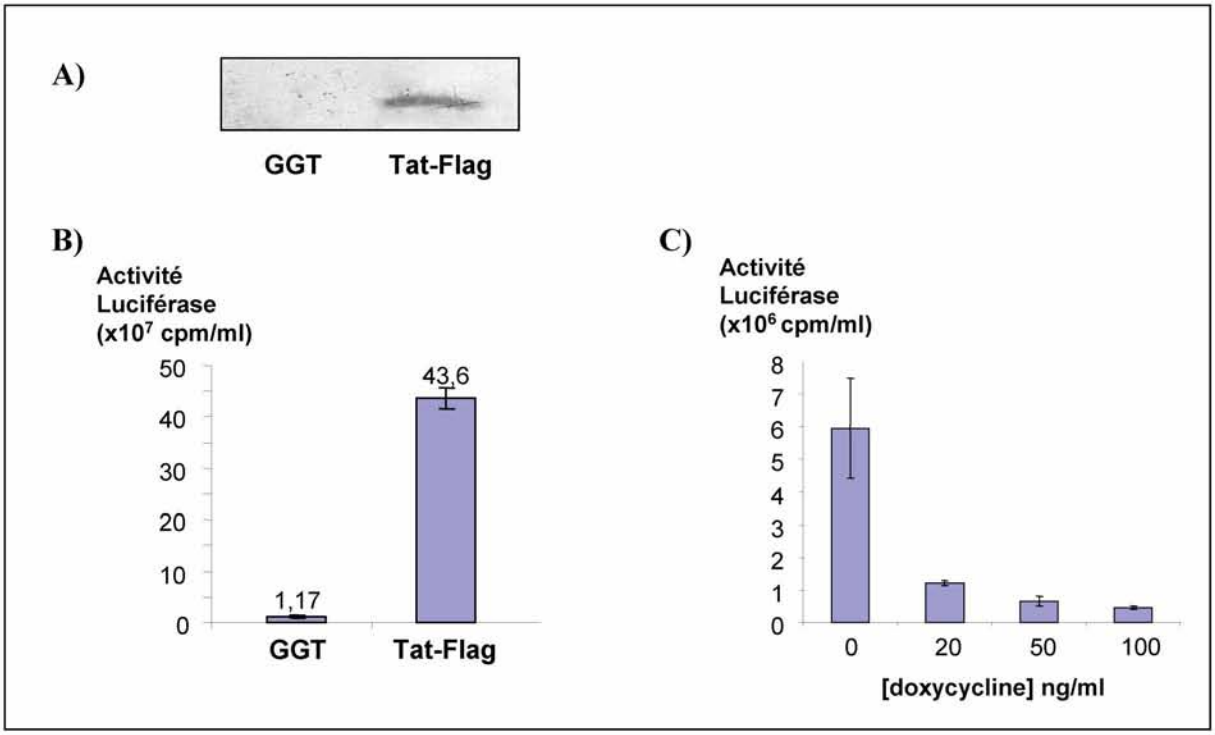


Fig. 70 : Analyse de la production et de l'activité de la protéine Tat par le système Tet-Off

A) Analyse de la production de la protéine Tat-Flag par Western blot avec l'anticorps anti-Flag. B) Analyse de l'activité de la protéine Tat-Flag par suivi de l'activité luciférase obtenue grâce à l'activation de la transcription du plasmide reporter pLTR-Luc. C) Contrôle de l'inhibition de la production de la protéine Tat-Flag par ajout de doxycyline.

1.3. Etude de l'effet de l'expression de la protéine Tat sur l'expression des protéines SR et hnRNP

Test au niveau du taux de transcrits :

J'ai étudié l'effet de l'expression de la protéine Tat sur le taux des transcrits des protéines SR et hnRNP par RT-PCR semi-quantitatives. Les études ont porté sur 3 des protéines SR (9G8, SC35, ASF/SF2) et les 4 protéines hnRNP A1, F, H1 et H2. Pour cela, un couple d'oligonucléotides spécifiques de chaque ORF a été défini (cf. Mat et met), ainsi qu'un couple d'oligonucléotides permettant de mesurer le taux d'ARNm de la β_2-microglobuline que nous avons utilisé comme témoin interne. Les résultats obtenus sont présentés sur la Fig. 71. Aucune variation marquée du taux des ARNm n'était observée dans les conditions utilisées.

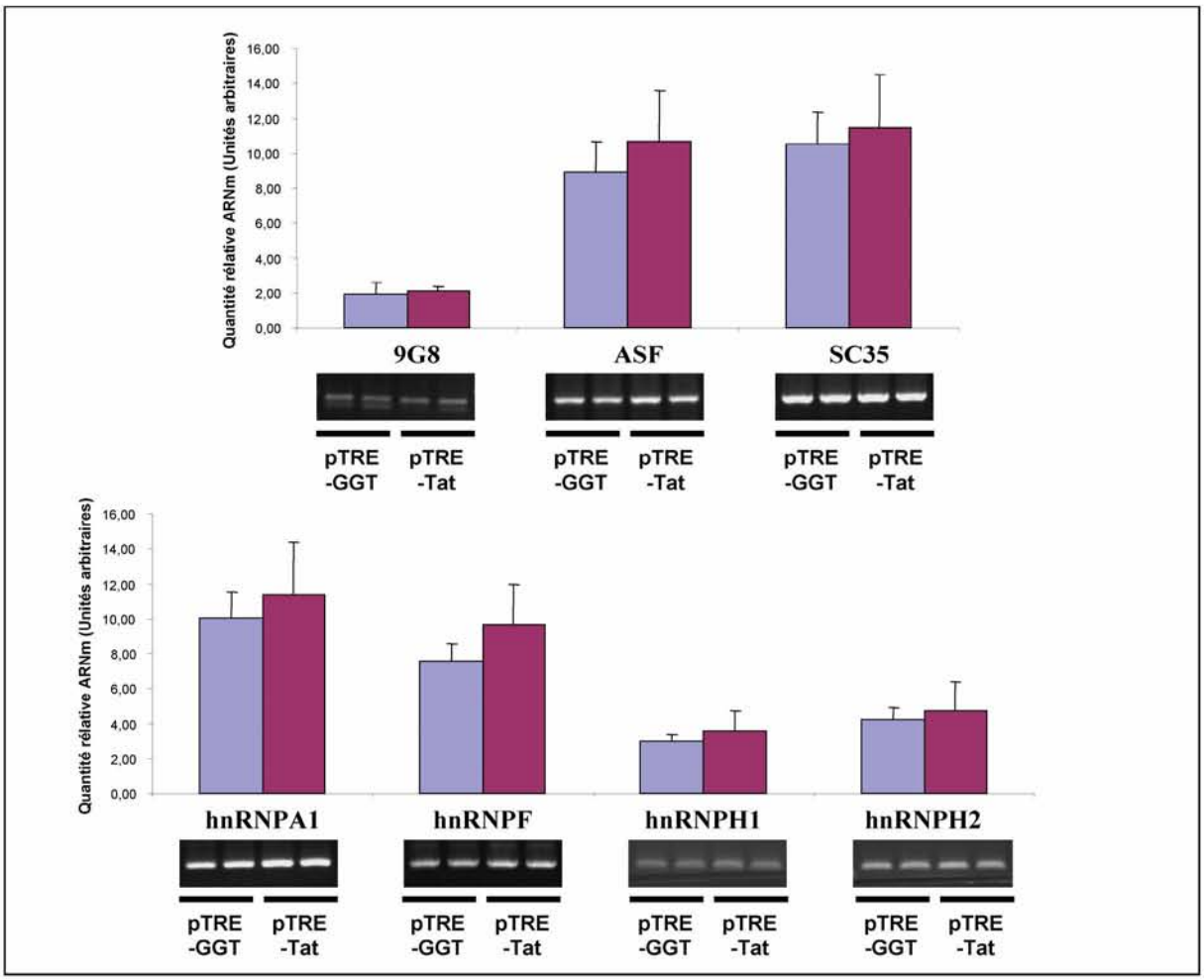


Fig. 71 : Effet de l'expression de la protéine Tat sur le taux de transcrits des protéines 9G8, SC35, ASF/SF2, hnRNP A1, F, H1 et H2

Analyse par RT-PCR des ARN messagers des protéines 9G8, ASF, SC35, hnRNP A1, F, H1 et H2 extraits de cellules HeLa Tet-Off obtenues lors de deux expériences distinctes de co-transfection par les vecteurs pTRE-GGT (témoin) ou pTRE-Tat-Flag. Chaque produit PCR a été quantifié comme décrit dans la partie Matériel et méthodes et les résultats sont indiqués au dessus des inserts correspondant au gel de fractionnement des produits de RT-PCR.

Test au niveau du taux de protéines :

N'ayant pas observé d'effet marqué au niveau du taux des ARNm, nous avons testé s'il pouvait néanmoins y en avoir un au niveau des protéines. Pour cela, nous avons utilisé des anticorps dirigés contre les protéines ASF/SF2, SC35 et 9G8 que nous a fourni J. Stévenin (IGBMC, Strasbourg) et un anticorps dirigé contre la protéine hnRNP A1 fourni par G. Dreyfuss (Philadelphie, USA). Bien que certaines de ces protéines soient à la fois nucléaire et cytoplasmique comme c'est le cas pour ASF/SF2 et hnRNP A1, leur action de régulation de l'épissage s'effectue au niveau du noyau. Il était donc important d'étudier leur variation de

concentration dans le noyau en fonction de l'expression de la protéine Tat. Pour cela, nous avons préparé des extraits nucléaires à partir des cellules HeLa transfectées par les plasmides pTRE-GGT (expérience témoin) ou pTRE-Tat-Flag. Le taux de protéines SR et hnRNP dans ces extraits a été analysé par Western blot. En présence de la protéine Tat, nous observons une augmentation de la quantité de la protéine SC35 dans le noyau, une légère augmentation de la quantité d'ASF/SF2 mais, aucune variation significative du taux des protéines hnRNP A1, 9G8 et ASF/SF2 (Fig. 72).

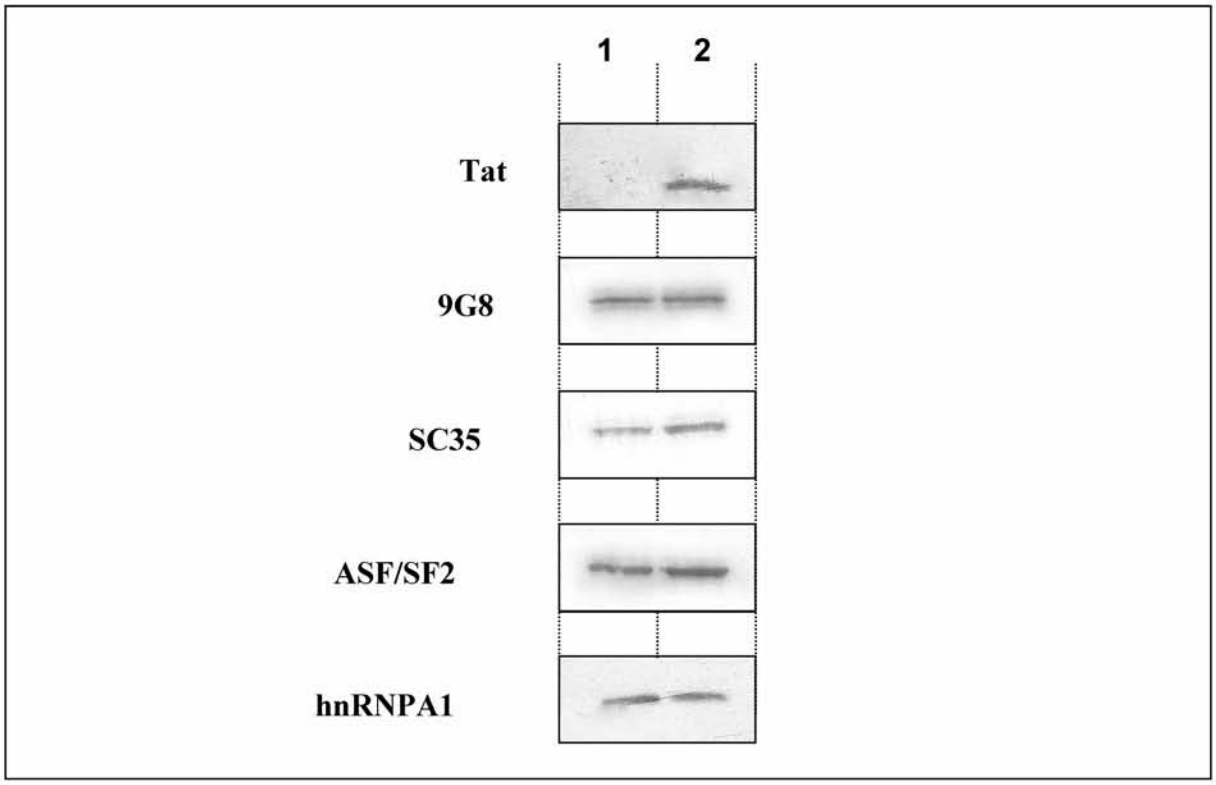


Fig. 72 : Etude de l'effet de l'expression de la protéine Tat sur le taux de protéines 9G8, SC35, ASF/SF2 et hnRNP A1 dans les noyaux

Des extraits de cellules HeLa transfectées par le plasmide témoin pTRE-GGT (piste 1) ou par le plasmide pTRE-Tat-Flag (piste 2) ont été préparés. Des quantités égales de protéines totales ont été fractionnées par SDS-PAGE. Les protéines ont ensuite été transférées sur une membrane de nitrocellulose et les variations de taux des protéines 9G8, SC35, ASF/SF2 et hnRNP A1 ont été analysées par Western blot, en utilisant des anticorps spécifiques de chacune des protéines. Nous observons une augmentation du taux des protéines SC35 et ASF/SF2.

2. Effet de la surexpression des protéines Tat et Nef sur les taux relatifs d'utilisation des sites d'épissage de l'ARN du virus HIV-1

Dans ce nouveau volet de notre étude, il nous a paru intéressant d'étudier aussi le rôle de la protéine Nef qui fait partie des protéines virales comme Tat dans la phase précoce du cycle viral, et dont nous avons vu dans l'introduction qu'elle pouvait potentialiser l'activité de Tat ainsi que celle de différentes activités cellulaires dont des activités kinases. Un moyen indirect de tester si la surexpression des protéines Tat et Nef peut avoir une influence sur le taux nucléaire des facteurs cellulaires pouvant réguler l'épissage, était d'étudier l'effet de la surexpression de chacune de ces protéines virales sur le profil d'épissage de l'ARN du viral. Bien sûr, les résultats attendus pouvaient aussi refléter un effet direct des protéines virales sur la régulation de l'épissage.

Pour réaliser cette étude, nous disposions du plasmide pΔPSP construit par l'équipe de M. Stoltzfus (Jacquenet *et al.*, 2001b) et qui renferme l'ADNc de la souche HIV-1 pNL4.3, tronqué au sein de l'intron délimité par les sites D1 et A1 (1550-4550), ce qui abolit son pouvoir pathogène (Fig. 73). L'équipe de M. Stoltzfus avait montré qu'après transfection de cellules HeLa avec ce plasmide, les différents ARNm viraux pouvaient être analysés et que le profil d'épissage reflétait parfaitement celui obtenu pour un ARN viral entier dans les lymphocytes T infectés (Jacquenet *et al.*, 2001b).

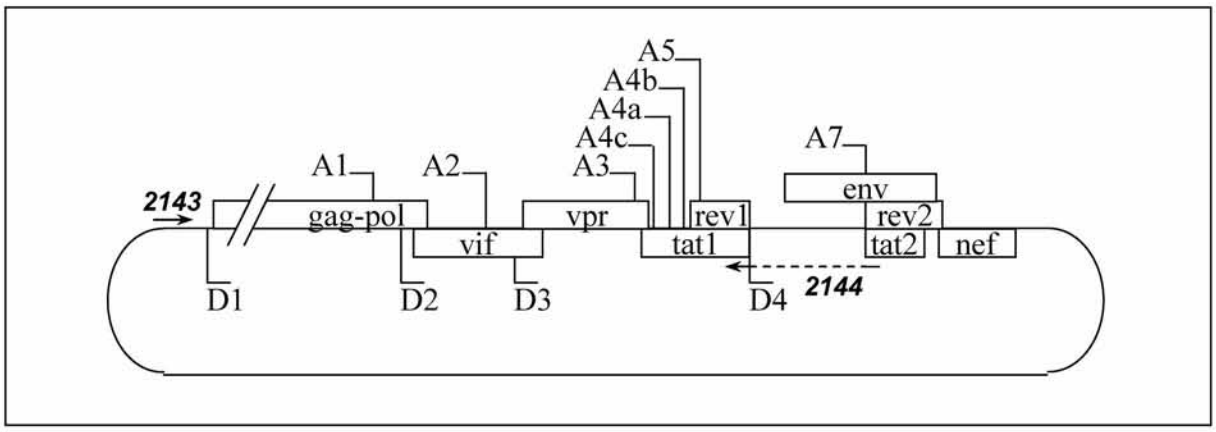


Fig. 73 : Schéma représentant le plasmide pΔPSP

Le plasmide pΔPSP est dérivé du plasmide pBR322 dans lequel a été inséré l'ADNc proviral de la souche HIV-1 pNL4.3 délété de la région 1501-4551 de l'ORF gag. Cette délétion est indiquée par les deux barres parallèles inclinées. Les positionnements de l'amorce 2144 utilisée pour la transcription inverse et l'amplification par PCR, ainsi que l'amorce 2143 uniquement utilisée pour l'amplification par PCR sont indiqués par des flèches.

Ainsi des cellules HeLa ont été co-transfectées par une des constructions pTRE-Tat-Flag ou pTRE-Nef avec le plasmide pΔPSP, puis cultivées pendant 48 h avant extraction des ARN totaux. Les produits d'épissage du transcrit du plasmide pΔPSP ont été analysés par RT-PCR en utilisant l'oligonucléotide 2144 pour la transcription inverse et le couple 2143/2144 pour l'amplification par PCR (Fig. 73). L'oligonucléotide 2143 s'hybride sur les positions 700 à 727 (souche pNL4-3) situées en amont du site D1, et l'oligonucléotide 2144 à une séquence complémentaire à la séquence générée après épissage entre les sites D4 et A7. Ainsi, seuls les ARN épissés entre ces 2 sites sont analysés par cette approche.

Nous avons observé des variations marquées des taux d'utilisation relatifs des sites A2, A3, A4c,a,b et A5 lorsque nous avons surexprimé soit Tat, soit Nef, soit les deux protéines (Fig. 74). De manière intéressante, lorsque Tat était surexprimée seule, une disparition presque totale des ARNm tat1, tat2 et tat3 était observée, ce qui pourrait correspondre à une régulation négative de la protéine Tat sur son propre épissage ou une action négative de cette protéine sur la stabilité de ses ARNm. Lorsque Tat est surexprimée, nous observons aussi une diminution du taux des ARNm nef 3, 4 et 5 (Fig. 74). Ces résultats sont inattendus car sur la base de l'augmentation observée du taux de SC35 dans le noyau on pouvait s'attendre au contraire à une augmentation de l'utilisation du site A3 qui est activé par cette protéine SR. Néanmoins, comme il sera discuté plus loin, plusieurs mécanismes peuvent être envisagés pour expliquer ce résultat.

La surexpression de Nef s'est traduite par une augmentation du taux d'ARNm vpr1 et vpr2 similaire à celle observée pour une augmentation du taux d'ASF/SF2 (Ropers *et al.*, 2004) (Fig. 74). L'observation de variations différentes du profil des ARNm suivant que Tat ou Nef étaient surexprimées, est un argument en faveur du fait que ces variations dépendent bien de la surexpression des protéines virales. Dans le cas d'une co-surexpression de Tat et Nef, nous retrouvions bien une baisse de la quantité des ARNm tat, mais le taux d'autres ARNm qui étaient augmentés en présence de Nef seule, était diminué dans ce cas. Ces résultats obtenus en fin de thèse paraissent très intéressants. Il sera important de répéter les expériences une seconde fois afin de s'assurer de la reproductibilité des résultats. S'ils sont reproductibles, il sera important de les compléter, comme nous allons le décrire plus loin dans la partie Discussion-Perspectives.

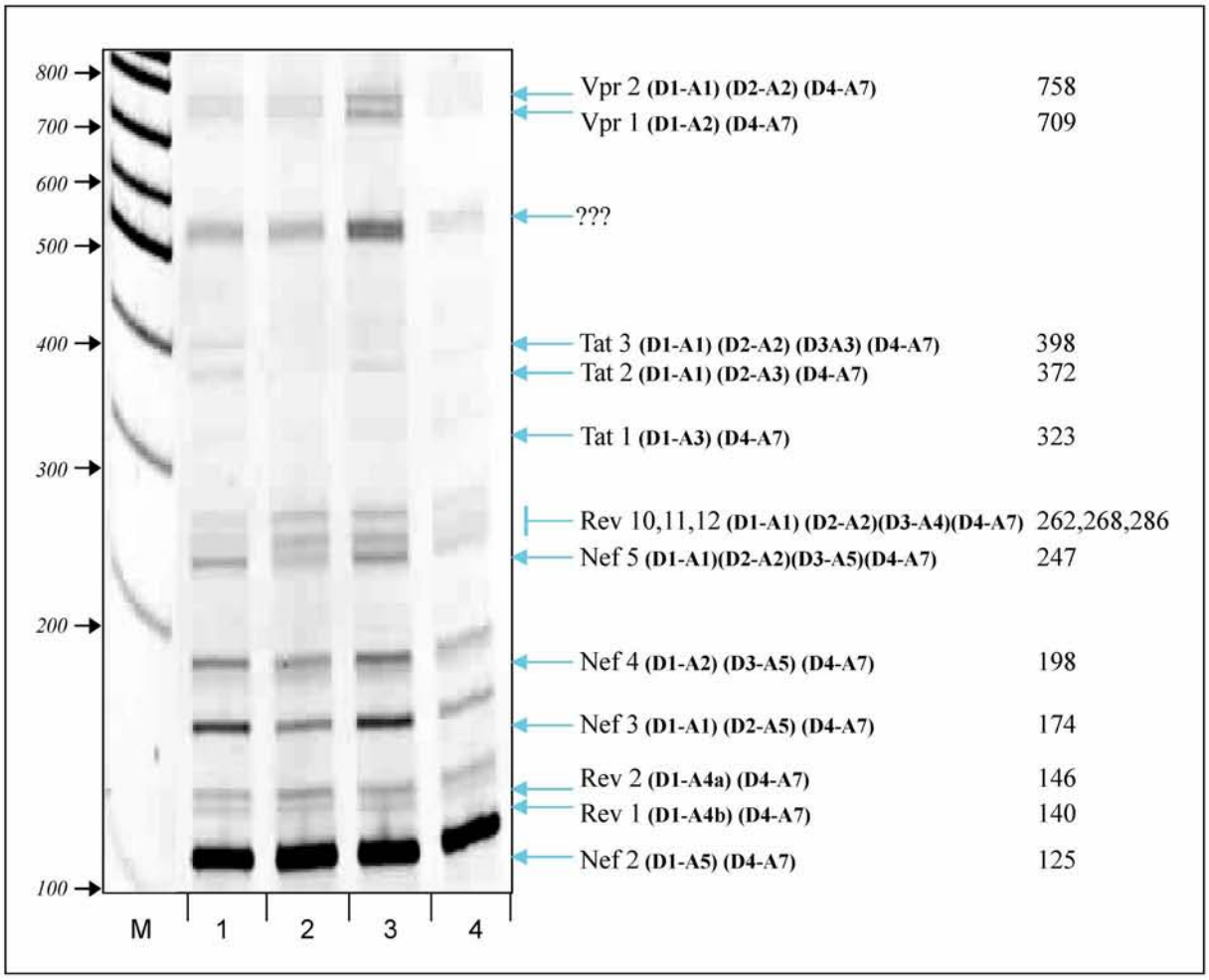


Fig. 74 : Influence de la surexpression des protéines Tat et Nef sur l'épissage alternatif de l'ARN du virus HIV-1

Analyse RT-PCR des produits de l'épissage de l'ARN produit par le plasmide pΔPSP. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP seul (piste 1) ou avec les plasmides pTRE-Tat-Flag et pTRE-Nef permettant de sur exprimer Tat-Flag (piste 2), Nef (piste 3) ou les 2 protéines ensemble (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 HIV-1, comme décrit dans le chapitre Matériels et méthodes. L'identité des ADNc fractionnés correspondant aux différents ARNm de HIV-1 est indiquée à gauche de l'autoradiogramme selon la nomenclature de Purcell & Martin (Purcell and Martin, 1993).

3. Etude de l'effet directe possible de la protéine virale NCp7 sur le taux d'utilisation des sites d'épissage de l'ARN de HIV-1

Une autre protéine du virus HIV-1, la nucléocapside NCp7, pourrait modifier le profil d'épissage de la cellule infectée. Cette protéine s'associe à l'ARN génomique qu'elle recouvre entièrement pour former une structure nucléoprotéique, et favorise la dimérisation et

l'encapsidation de l'ARN viral. Elle possède aussi des propriétés de protéine chaperon qui lui permettent d'intervenir lors des étapes de transcription inverse et d'intégration (pour revue, Darlix *et al.*, 2001). Un autre groupe a pu aussi démontré que NCp7 migre du cytoplasme vers le noyau des cellules infectées 10 heures après l'infection (Zhang and Crumpacker, 2002). De part sa grande affinité pour l'ARN viral, il est possible d'envisager un effet de la protéine NCp7 sur le processus d'épissage. Nous avons donc abordé l'étude de l'effet possible de NCP7 sur l'efficacité de deux des sites d'épissage de l'ARN de HIV-1.

Les protéines NCp7 sauvage (1-72aa) et mutante (12-53aa) synthétisées chimiquement par l'équipe de B. Roques nous ont été fournies par JL. Darlix. La protéine tronquée (12-53) nous a servi de témoin, puisqu'elle ne se fixe pas sur l'ARN de HIV-1 (Gabus *et al.*, 2001). Nous disposions au laboratoire de tout un ensemble de constructions permettant de produire par transcription *in vitro* des ARN renfermant le site donneur D1 de la souche HIV-1 Bru et un ou plusieurs sites accepteurs de l'ARN de cette même souche. L'efficacité d'épissage *in vitro* des ARN produits à partir de ces constructions avait déjà été bien étudiée *in vitro* au laboratoire avec ou sans ajout de protéines SR (Jacquenet *et al.*, 2001a ; Ropers *et al.*, 2004). Nous avons donc testé l'effet de la présence de NCp7 sur l'efficacité de cet épissage *in vitro.*

Après plusieurs essais, nous avons observé des effets plus marqués lorsque la protéine était incubée avec l'ARN avant que l'extrait nucléaire soit ajouté par rapport aux effets observés lorsqu'elle était ajoutée en même temps que l'extrait nucléaire. Nous avons observé dans ces conditions une augmentation de l'efficacité d'épissage au site A1 couplé au site D1 et au contraire une inhibition du site A2 couplé à ce même site D1. On note que l'effet observé au site A2 est très similaire à celui observé lors de l'ajout de la protéine hnRNP A1 (Fig. 75).

4. Discussion-Perspectives

L'ensemble des données que nous venons de décrire suggèrent fortement que certaines des protéines du virus HIV-1 peuvent modifier les taux relatifs d'utilisation des sites d'épissage du virus HIV-1, ceci soit directement, soit par modulation de l'expression ou de l'activité des protéines SR.

Nous avons observé une augmentation du taux de protéines SR ASF/SF2 et surtout SC35 localisées dans le noyau lorsque la protéine Tat était surexprimée par transfection transitoire, ceci alors que nous n'avons pas détecté de variation significative du taux des ARNm par RT-PCR semi-quantitative. Il sera important de refaire plusieurs fois les expériences de quantifications du taux des protéines SR dans les noyaux, sachant que l'étape de préparation

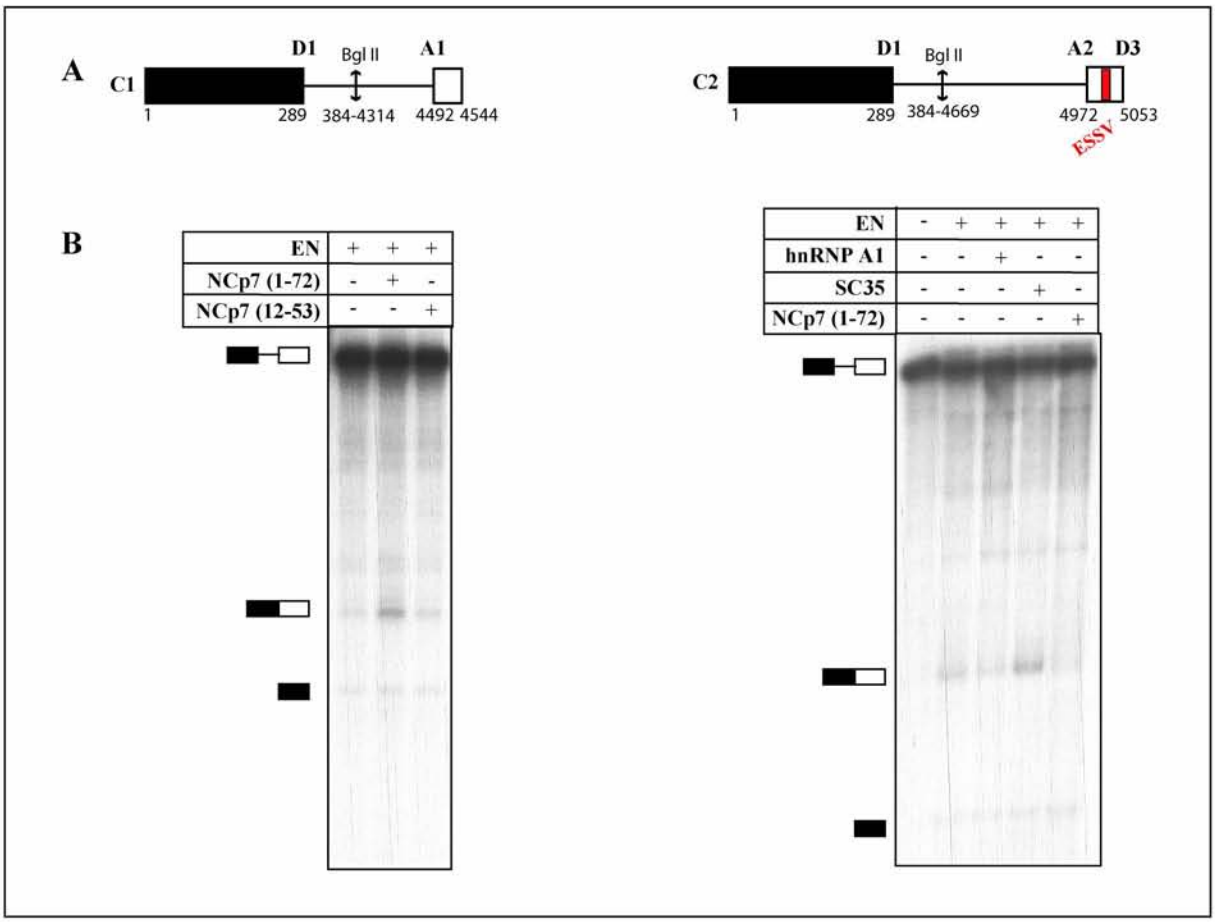


Fig. 75 : Etude de l'effet de la protéine NCp7 sur l'utilisation in vitro des sites accepteurs A1 et A2 de l'ARN du virus HIV-1

A) Représentation schématique des constructions C1 et C2. L'ARN C1 est composé des séquences 1-384 et 5172-5408 de l'ARN de la souche Bru, séquences qui contiennent respectivement les sites d'épissage D1 et A3. L'ARN C2 contient les séquences 1-384 et 4669-5053 de l'ARN de la souche Bru séquences qui contiennent respectivement les sites d'épissage D1, A2 et D3, ainsi que l'élément ESSV (en rouge). B) Effet des fragments (1-72) et (12-53) de la protéine NCp7 sur l'épissage des ARN C1 et C2. Les ARN ont été incubés en présence d'extrait nucléaire de cellules HeLa (EN) seul ou avec 1µg de protéine NCp7 ou 400 ng de protéines SC35 ou hnRNP A1 dans les conditions décrites dans le chapitre Matériel et méthodes. Les produits d'épissage ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 5%. La position des précurseurs est indiquée à droite de l'autoradiographie.

des extraits nucléaires est plus difficile à contrôler que celle de préparation des ARN totaux. Par ailleurs, pour une quantification plus précise des taux d'ARNm, nous pouvons envisager d'utiliser des expériences de PCR quantitative, la technique ayant maintenant été introduite au laboratoire par C. Aigueperse.

Un autre point sur lequel il sera important de se pencher, sera de savoir si la durée d'expression des gènes Tat et Nef transfectés est suffisante pour que les protéines virales aient le temps d'avoir une action sur les gènes cellulaires. L'effet observé sur la concentration nucléaire pourrait simplement refléter un effet sur le niveau de phosphorylation des protéines SR, qui module leur trafic entre le cytoplasme et le noyau. Pour obtenir une expression prolongée, nous pourrons envisager de construire et utiliser des lignées exprimant Tat ou Nef de manière stable, ceci sous contrôle du système Tet-Off.

Dans l'interprétation de ces données, il faut aussi tenir compte des données récentes sur la régulation de l'expression des protéines SR et hnRNP. Il a été proposé que l'épissage soit régulé de façon tissu-spécifique ou selon les stades du développement grâce à une régulation des taux de protéines SR et hnRNP (pour revue, Caceres and Kornblihtt, 2002 ; Sanford *et al.*, 2005).

En parallèle, trois études plus spécifiques de l'expression des protéines SR ont démontré l'implication directe des protéines SRp20, 9G8 et SC35 dans l'épissage de leur propres ARNm, avec un phénomène de rétrocontrôle pour SC35 (Jumaa *et al.*, 1996 ; Sureau *et al.*, 2001 ; Lejeune *et al.*, 2001). Très récemment, ce modèle a été généralisé à l'ensemble des protéines SR sur la base d'études des liens existants entre des éléments ultra-conservés des gènes de ces protéines et la régulation de leur expression, chez l'homme et la souris (Lareau *et al.*, 2007). Une étude similaire réalisées sur les gènes de protéines hnRNP suggèrent qu'elles sont aussi soumises au même type de régulation (Ni *et al.*, 2007). En effet, dans les deux cas, des éléments très conservés sont présents soit dans des exons cassettes dits « poisons », car contenant un codon stop prématuré, soit dans des introns alternatifs de la région 3' non traduite. L'inclusion de ces éléments crée des ARNm non productifs car ciblés puis détruits par le système NMD. Sur cette base, les auteurs de ces deux articles proposent une autorégulation voire une régulation croisée assurant une homéostasie globale des facteurs d'épissage (Fig. 76).

En plus de ces régulations du taux de production des protéines SR, nous avons pu voir en introduction que leur niveau de phosphorylation et leur disponibilité dans le noyau sont aussi d'autres moyens de régulation. Nos données peuvent donc s'expliquer par une action des protéines Tat et Nef à ces différents niveaux, et des études devront être réalisées en ce sens.

L'effet négatif sur l'utilisation du site A3 observé pour la protéine Tat était inattendu. La protéine Tat ayant des capacités de fixation directe à l'ARN, une explication possible des données pourrait être que Tat se fixe sur le site A3 et ainsi qu'elle l'inhibe par encombrement stérique. Il sera donc important dans un premier temps de savoir si Tat a la capacité de se lier

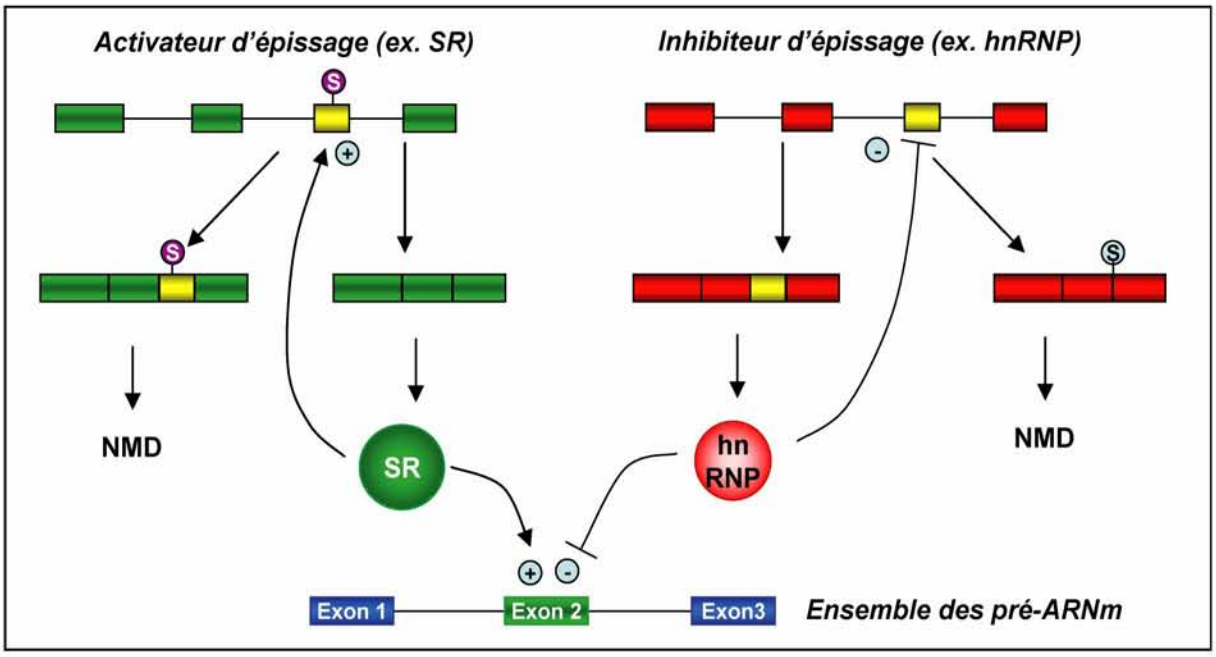


Fig. 76 : Modèle d'autocontrôle de la production des facteurs d'épissage

directement à la structure tige boucle contenant le site A3, où encore au segment contenant sa boite de branchement. Ces tests peuvent être réalisés par expériences de retard sur gel avec la protéine recombinante produite comme décrit ci-dessus, ou encore par des expériences de pontage aux UV entre des ARN radioactifs et les protéines d'extraits nucléaires de cellules exprimant la protéine Tat-Flag, suivies d'une hydrolyse de l'ARN et de la sélection par l'anticorps anti-Flag. Des expériences d'épissage *in vitro* en présence ou en absence de la protéine GST-Tat pourront aussi être réalisées pour confirmer un effet direct possible.

L'effet de Tat peut aussi être indirect soit par activation des sites en compétition avec le site A3, soit par recrutement de facteurs.

D'une manière générale, il pourrait être intéressant de tester si la protéine Tat possède d'autres sites de fixation que la région TAR dans l'ARN du virus HIV-1. Ceci pourrait se faire en développant une puce à ADN contenant des fragments de l'ADNc pro-viral jalonnant l'ARN du virus HIV-1, ainsi que des oligonucléotides permettant de mettre en évidence les choix d'épissage alternatif. Le pontage aux UV de Tat-Flag à l'ARN suivi d'une hydrolyse ménagée et de la sélection par un peptide anti-Flag, puis d'une analyse par RT-PCR avec les puces à ADN devrait permettre de définir les sites éventuels de fixation. Comme nous le verrons plus loin, nous avons abordé une étude par protéomique des facteurs se liant sur les tige-boucles SLS2 contenant le site A3, et SLS3 contenant ses éléments régulateurs, il serait intéressant de tester l'effet de la présence de la protéine Tat sur les protéines fixées.

Jusque là, la protéine Nef n'a pas été décrite comme ayant des propriétés de fixation à l'ARN. Il est plus vraisemblable que les effets observés sur l'épissage soient liés à une action indirecte de cette protéine au niveau des kinases qui modulent l'activité des protéines SR. L'augmentation observée de l'ARNm vpr suggère une activation forte d'ASF/SF2, puisque le même niveau d'activation de cet ARNm est été observé en surexprimant ASF/SF2 (Ropers *et al.*, 2004). Nous ne pouvons néanmoins pas exclure qu'une autre protéine SR encore non testée jusque là, active aussi ce site. Aussi bien pour Tat que pour Nef, il serait intéressant de produire des lignées de cellules HeLa exprimant ces protéines de manière stable et d'analyser les modifications portées par les protéines SR, ainsi que leur localisation dans ces cellules.

Bien qu'étant préliminaires, les résultats obtenus avec la protéine virale NCp7 sont intéressants et nouveaux, puisque aucune donnée d'un rôle potentiel de NCp7 sur l'épissage de l'ARN de HIV-1 n'est disponible dans la littérature. Afin de poursuivre le travail, nous pourrons tenter de compléter l'étude en testant le rôle de NCp7 sur les autres sites d'épissage A3, A4c, a, b et A5, puis éventuellement identifier les sites de fixation de NCp7 au voisinage des sites d'épissage par empreinte avec des nucléases et/ou par expérience de retard sur gel.

CHAPITRE 6 : Etudes de l'effet de l'élément inhibiteur ESS2 sur la compétition entre les sites d'épissage A1 à A5

Une partie des travaux que j'ai réalisé sur le rôle de l'élément inhibiteur d'épissage ESS2 qui agit au site A3 a été incluse dans un article général qui fait le point sur l'état des connaissances sur les régulations s'exerçant à différents niveaux sur l'épissage alternatif de l'ARN du virus HIV-1. Nous commencerons cette dernière partie des résultats comprenant les chapitres 6, 7 et 8 en présentant cet article qui donne une vue générale sur les régulations de l'épissage alternatif de l'ARN du virus HIV-1, et contient des données précises que j'ai obtenues sur le site A3.

1. Présentation générale des objectifs de l'article et des travaux complémentaires

La régulation de l'épissage alternatif de l'ARN du virus HIV-1 dépend fortement du caractère sous-optimal et de la régulation des sites accepteurs d'épissage de cet ARN. L'équipe avait réalisé depuis plusieurs années une étude sur la régulation d'un grand nombre de sites accepteurs d'épissage de l'ARN du virus HIV-1. Il nous a été demandé d'écrire un article pour le journal « Frontiers in Bioscience » qui présente les connaissances publiées et certaines de nos données non publiées sur la régulation des sites accepteurs d'épissage de l'ARN du virus.

Ainsi dans cet article à la rédaction duquel j'ai largement participé, nous faisons le bilan des connaissances sur les régulations de l'épissage alternatif de l'ARN du virus HIV-1, et décrivons en détails les effets antagonistes des protéines SR et hnRNP aux sites accepteurs A2, A3 et A7, et leur impact sur la production des ARNm vpr, tat et rev notamment.

Les données obtenues *in vitro* par notre équipe sur la fixation des protéines hnRNP A1 et ASF/SF2 au site A2 n'avaient pas encore étaient publiées. Elles démontrent l'existence d'une compétition entre hnRNP A1 et ASF/SF2 sur une très courte région de l'intron A2-D3.

D'autres résultats non publiés concernant le site A3 et l'ESS2 sont aussi décrits. Ils ont été obtenus par analyse de l'effet sur l'épissage *in cellulo*, de mutations introduites sur l'ARN transcrit à partir du plasmide pΔPSP déjà décrit ci-dessus. Ces mutations avaient déjà été largement étudiées auparavant par d'autres méthodes et différents groupes, c'est pourquoi nous allons les présenter en préambule à l'article.

L'élément ESS2 a été décrit pour la première fois par l'équipe de MT. Stoltzfus qui montrait que cette séquence en aval du site A3 a un effet inhibiteur sur l'utilisation de celui-ci au moment de l'épissage (Amendt *et al.*, 1994, Amendt *et al.*, 1995). Afin d'obtenir une identification plus précise, les mêmes auteurs ont ensuite procédé à la substitution de chaque dinucléotides de la séquence par les dinucléotides complémentaires, et ont pu définir l'élément ESS2 comme étant la séquence conservée 5'-CUAGACUAGA-3' (située entre les positions 5846 et 5855 de l'ARN de la souche pNL4-3 ou les positions 5428 et 5437 de l'ARN de la souche Bru) (Fig. 77 ; Si *et al.*, 1997). Puis l'idée d'immobiliser cette séquence sur des billes d'agarose et de l'employer pour dépléter l'extrait nucléaire utilisé pour les réactions d'épissages *in vitro*, a permis à M. Caputi de démontrer que la protéine hnRNP A1 se fixe sur cette région de l'ARN et est responsable de son activité inhibitrice (Caputi *et al.*, 1999). Le mécanisme de cette inhibition par un élément situé à longue distance du site A3 (75 nt) restait alors à définir.

L'analyse de la structure 2D de l'ARN viral autour du site A3, réalisée par S. Jacquenet et D. Ropers au laboratoire, a tout d'abord permis de montrer que l'élément ESS2 se situe dans la partie apicale d'une grande structure tige-boucle irrégulière (SLS 3-A3) (Fig. 77 ; Jacquenet *et al.*, 2001b). H. Hallay et L. Ayadi ont alors réalisé au laboratoire un travail de mutagénèse dirigée contre une région de l'ARN viral plus large que celle qui avait été mutée par MT. Stoltzfus, mais toujours avec le même principe de substitution de dinucléotides par les dinucléotides complémentaires. L'effet des mutations a alors été étudié à la fois sur les capacités de fixation de la protéine hnRNP A1 sur les structures SLS 2-A3 et 3-A3, et sur l'efficacité d'épissage du site A3 *in vitro* (Hallay *et al.*, 2006). Les résultats obtenus ont permis de montrer que l'élément ESS2 joue un rôle central dans la fixation de la protéine hnRNP A1 sur l'ensemble de la structure SLS 3-A3 et le segment en simple brin reliant les structures tige-boucle SLS2 et SLS3. Par ailleurs, *in vitro*, un lien a pu être mis en évidence entre cette capacité de l'élément ESS2 à permettre la fixation et la multimérisation de hnRNPA1 sur cette région de l'ARN viral et l'inhibition de l'utilisation du site A3 par l'élément ESS2. En utilisant le plasmide pΔPSP décrit plus haut et en surexprimant une des protéines SR ASF/SF2, 9G8, SRp40 et SC35 par co-transfection dans des cellules HeLa, D. Ropers avait montré en collaboration avec J. Stévenin que SC35 et SRp40, et à un moindre degré 9G8, activent fortement l'utilisation du site A3 *in vitro* et *in cellulo* au détriment de l'utilisation des autres sites en compétition (Ropers *et al.*, 2004). D. Ropers, H. Hallay et L. Ayadi ont ensuite montré au cours de leur travail *in vitro* que les protéines SC35 et SRp40

se fixent à des sites de l'ARN viral qui chevauchent ceux où se lie hnRNPA1, et entrent donc en compétition d'où leur pouvoir activateur (Hallay *et al.*, 2006).

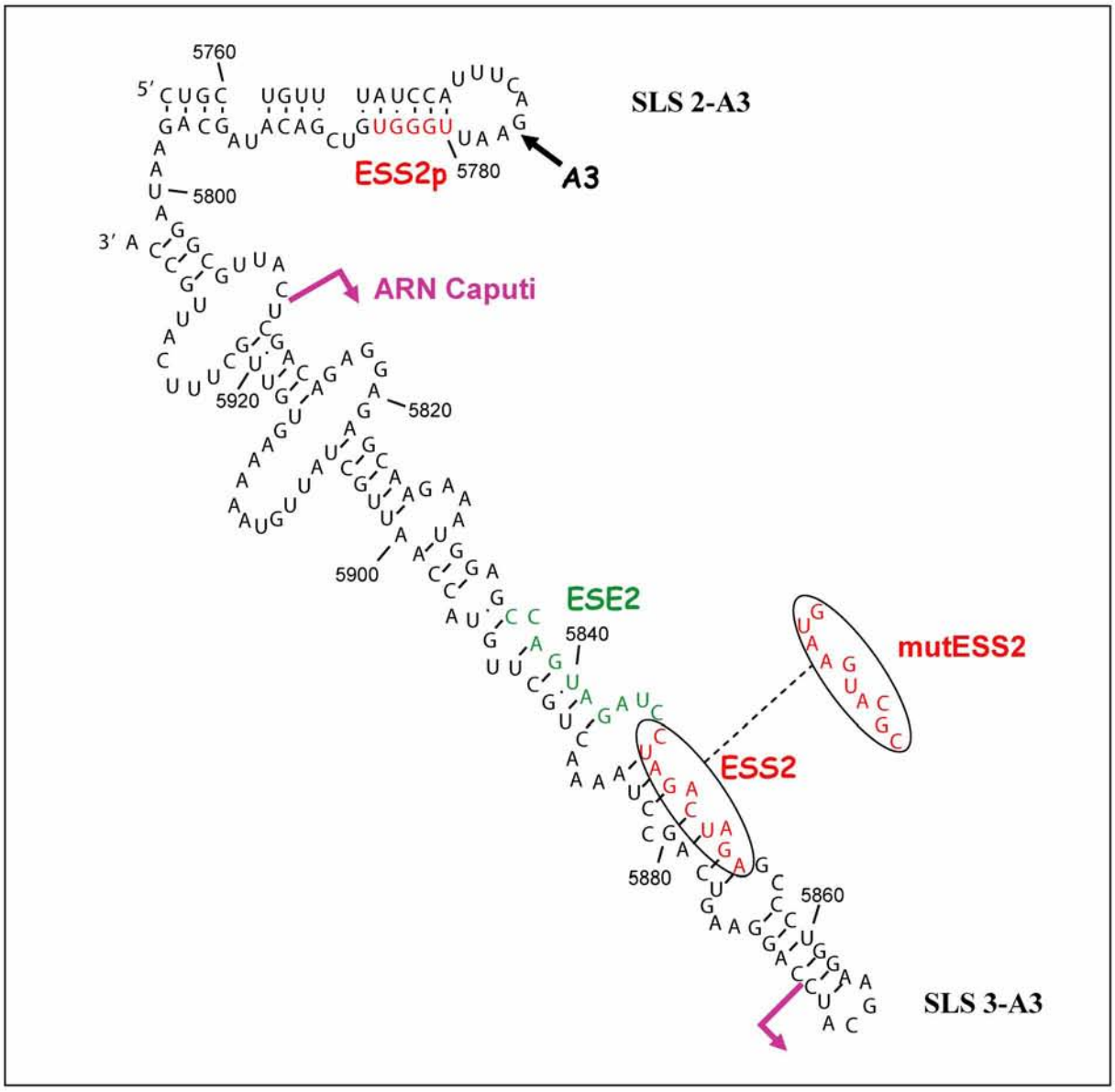


Fig. 77 : Structure secondaire de la région de l'ARN du virus HIV-1 contenant le site A3 et ses éléments régulateurs et mutant utilisé pour étudier ces éléments régulateurs

Structure secondaire établie pour la région de l'ARN du VIH-1 contenant le site A3 et ses éléments régulateurs (Jacquenet et al., 2001b). Les structures tiges-boucles SLS 2 et SLS 3 sont indiqués. Les éléments inhibiteurs ESS2p et ESS2 sont indiqués par les lettres rouges et l'élément activateur ESE2 en vert. Les flèches roses délimitent l'ARN utilisé par l'équipe de Caputi pour mettre en évidence la fixation de hnRNP A1 sur l'élément ESS2. La substitution mutESS2 dans l'élément ESS2 est également indiquée (Caputi et al., 1999). La numérotation des nucléotides est celle de l'ARN de la souche pNL4-3.

Puisqu'il était important de vérifier ces données obtenues *in vitro*, certaines des mutations pour lesquelles une diminution de l'activité de l'élément ESS2 était observée *in vitro* ont été introduite dans le plasmide pΔPSP, ceci afin de tester si le même effet était bien retrouvé *in cellulo*.

Ces mutations intervenaient dans trois motifs UAG, dont deux se situent dans l'élément ESS2 et le 3^{ième} en amont de cet élément, dans ce qui avait été décrit comme un élément activateur nommé ESE2 par l'équipe de M. Caputi (Zahler *et al.*, 2004) (Fig. 78).

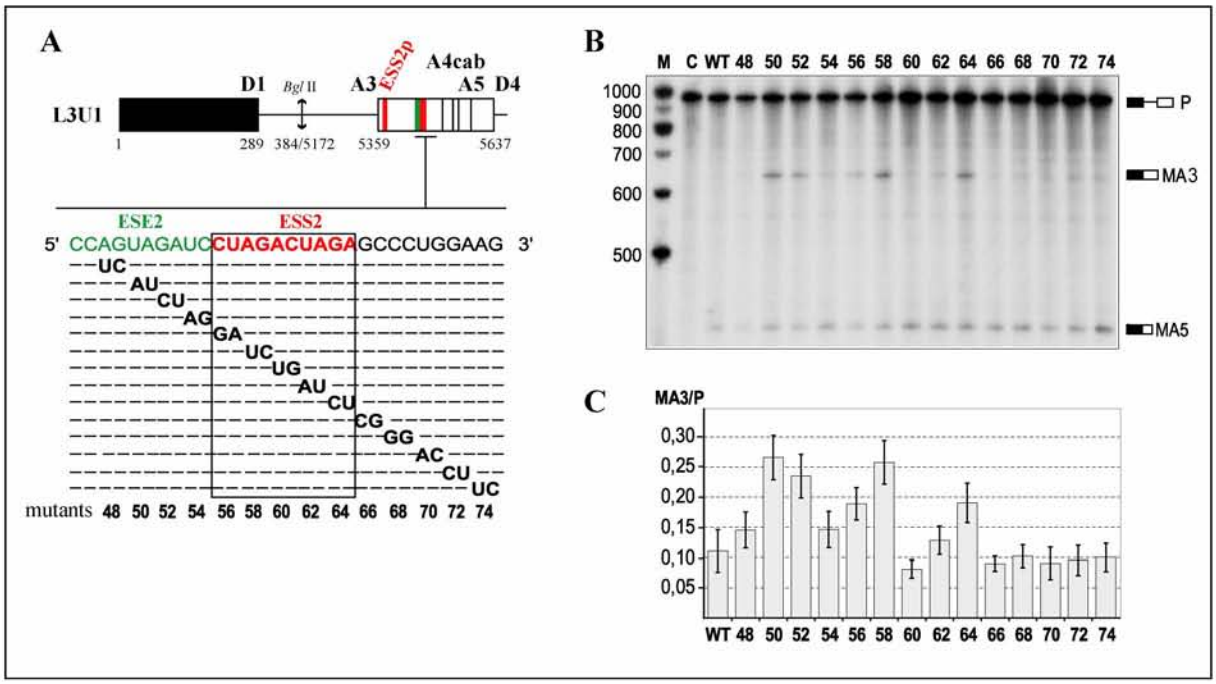


Fig. 78 : Effets des substitutions des dinucléotides de la région ARN contenant ESE2 et ESS2 sur l'utilisation de A3 en épissage in vitro

A) Schéma de l'ARN L3U1 et des mutants utilisés pour l'épissage in vitro. Les numéros donnés correspondent à la position du dinucléotide substitué. B) Effet des substitutions sur l'épissage de l'ARN L3U1. Les ARN ont été incubés en présence d'extrait nucléaire de cellules HeLa dans les conditions décrites dans le chapitre Matériel et méthodes. Les produits d'épissage ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 5%. La position des précurseurs est indiquée à droite de l'autoradiographie. Un ARN non traité a été fractionné en parallèle dans la piste C. C) Ratio MA3/P obtenu pour chaque ARN.

J'ai donc réalisé une mutation dans chacun des trois triplets UAG au sein du plasmide pΔPSP (mutant 4050 dans l'élément ESE2 et mutant 4058 et 4064 dans l'élément ESS2) comme décrit dans la partie Matériel et méthodes. Nous avons aussi choisi comme témoin une mutation qui n'affectait pas l'épissage *in vitro* (mutant 4054).

Les données obtenues pour les mutants de l'élément ESS2 sont décrites dans l'article inséré ci-dessous. Elles confirment les effets observés *in vitro* puisqu'elles montrent que deux substitutions de bases (AG en UC variant 4058 et GA en CU variant 4064) dans cet inhibiteur suffisent à limiter fortement son effet négatif et que le taux d'utilisation du site A3 est dans ce cas très augmenté par rapport à celui des autres sites accepteurs sans que l'on ait à ajouter un excès de protéines SR.

En plus de ces remplacements de dinucléotides, MT. Stoltzfus avait généré et étudié l'effet d'une mutation plus large (appelée ici mutESS2) remplaçant la séquence CUAGACUAGA de l'élément ESS2 par la séquence GUAAGUACGC (Amendt *et al.*, 1995). Cette mutation s'était avérée lever l'inhibition générée par l'élément ESS2 dans une construction hétérologue. Plus récemment, M. Caputi s'était à nouveau servi de cette mutation pour valider l'existence de l'ESE2 et la fixation de la protéine SC35 sur cette séquence juxtaposée à l'ESS2 (Zahler *et al.*, 2004).

Néanmoins, lorsque H. Hallay et L. Ayadi ont réalisé des substitutions de dinucléotides par leurs séquences complémentaires dans cette séquence ESE2, au lieu d'observer une diminution de l'efficacité du site A3 comme attendu pour la mutation d'une séquence activatrice, une augmentation du taux d'épissage à A3 a été observée comme mentionné ci-dessus (Fig. 78). Ces données étaient concordantes avec la présence d'un motif UAG pouvant correspondre à un site de fixation de hnRNP A1 et avec les données d'empreintes. Par ailleurs, des expériences de retard sur gel réalisées avec des fragments de l'ARN de la souche Bru contenant les éléments ESE2 et ESS2, montraient que la délétion de l'un ou l'autre des éléments affectait de la même façon la fixation de SC35 (Hallay *et al.*, 2006). Ces données n'étaient donc pas en bon accord avec le fait que l'élément appelé ESE2 soit vraiment un élément activateur du site A3.

Pour essayer de mieux comprendre les différences observées entre les résultats de notre équipe et celle de M. Caputi, j'ai donc introduit la longue substitution mutESS2 dans l'élément ESS2 dans le plasmide pΔPSP. De manière inattendue, comme nous allons le montrer dans la publication qui suit et l'expliquer en détails dans les résultats complémentaires, la mutation mutESS2 génère en fait un nouveau site accepteur dans l'ARN viral, ce qui a dû fausser les interprétations faites par l'équipe de Caputi.

Comme le montre la synthèse des données sur la régulation au site A3 qui est présentée dans l'article qui suit, un paramètre important dans cette régulation semble être la structure adoptée par l'ARN. La structure tige-boucle apicale de SLS 3-A3, appelé motif B, est très conservée

en séquence dans toutes les souches de virus HIV-1 du groupe M (Jacquenet *et al.*, 2001b). Des données obtenues en collaboration entre le laboratoire et l'équipe d'E. Guittet (LCBS, Gif sur Yvette) (thèse H. Hallay) montraient que cette structure tige-boucle joue un rôle important dans la structuration de l'élément ESS2. Néanmoins l'influence de cette structuration sur l'activité inhibitrice de l'épissage n'avait jamais été testée *in cellulo.*

Deux mutations supplémentaires ont donc été introduites sur le plasmide pΔPSP de façon à produire des ARN dans lesquels la structuration de l'élément ESS2 était modifiée, et pouvoir tester l'effet de ces modifications sur l'efficacité relative d'utilisation du site A3 par rapport aux autres sites accepteurs en compétition. Ces travaux sont présentés à la suite de l'article dans ce chapitre.

2. Publication : « Rôle de la structure des ARN et des facteurs protéiques sur l'épissage du virus HIV-1 »

Role of RNA structure and protein factors in the control of HIV-1 splicing

Jean-Michel Saliou [1], Cyril F. Bourgeois [2,3,4,5], Lilia Ayadi-Ben Mena [1], Delphine Ropers[1], Sandrine Jacquenet[1], Virginie Marchand[1], James Stevenin[2,3,4,5], Christiane Branlant [1]

[1] MAEM, UMR 7567 CNRS-UHP, Nancy Universite, Faculte des sciences et techniques, BP239, Vandoeuvre Les Nancy CEDEX, France, [2] IGBMC, Department of Functional Genomics, Illkirch, F-67400 France, [3] INSERM, U596, Illkirch, F-67400 France, [4] CNRS, UMR7104, Illkirch, F-67400 France, [5] Universite Louis Pasteur, Strasbourg, F-67000 France

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Overview of HIV-1 alternative splicing in the context of viral infection
4. A small regulatory element that folds into a stem-loop structure controls site A2 utilization
5. A complex extended exon region regulates site A3 utilization
6. Site A7 is regulated by downstream and upstream elements
7. Regulatory elements located in other HIV-1 RNA regions
8. HnRNP A1 a general host down-regulator of HIV-1 RNA splicing
9. Viral proteins may limit HIV-1 RNA splicing
10. Conclusions: lessons from HIV-1 splicing for the design of future therapies
11. Acknowledgments
12. References

1. ABSTRACT

Alternative splicing plays a key role in the production of numerous proteins by complex lentiviruses such as HIV-1. The study of HIV-1 RNA splicing has provided useful information not only about the physiology of the virus, but also about the general mechanisms that regulate mammalian pre-mRNA alternative splicing. Like all retroviruses, a fraction of HIV-1 transcripts remains intact to serve as genomic RNA and to code for Gag and Gag-Pol protein precursors. In addition, splicing is important for controlling the production of some viral proteins, which could otherwise have a negative effect on the infected cell. Here, we summarize how the utilization of HIV-1 splicing sites is limited by the binding of nuclear factors to cis-acting silencer elements, taking into account the role of RNA secondary structure in these mechanisms. We also describe how the poorly efficient HIV-1 acceptor sites are nevertheless activated by serine/arginine-rich proteins. Finally, we discuss how nuclear factors that interact with both the transcription and splicing machineries also participate in the control of HIV-1 RNA splicing.

2. INTRODUCTION

The choice between alternative splice sites is prevalent in higher eukaryotic genomes, since about 75% of the genes are regulated by alternative splicing of their pre-messenger RNAs (1). On the basis of data accumulated over the last 15-20 years, a number of general features have started to emerge to explain how alternative splicing is regulated in cells, leading the scientists from this field to believe that the elucidation of a "splicing code" can now be envisaged (2). One of those general rules is the existence of antagonistic mechanisms that allow either the activation or the repression of alternative splice sites, driven by a variety of cis-acting splicing enhancers and silencers. Two main families of proteins are considered as mediators between the cis regulatory elements and the spliceosomal machinery: serine/arginine-rich (SR) proteins and hnRNP proteins that recognize splicing enhancers and splicing silencers, respectively, although there are exceptions to this general view (3). In addition, other parameters are known to influence alternative splicing, including the dynamic coupling between transcription and other RNA processing events (4), or the presence of RNA secondary structure (5).

RNA structure and protein factors in HIV-1 splicing

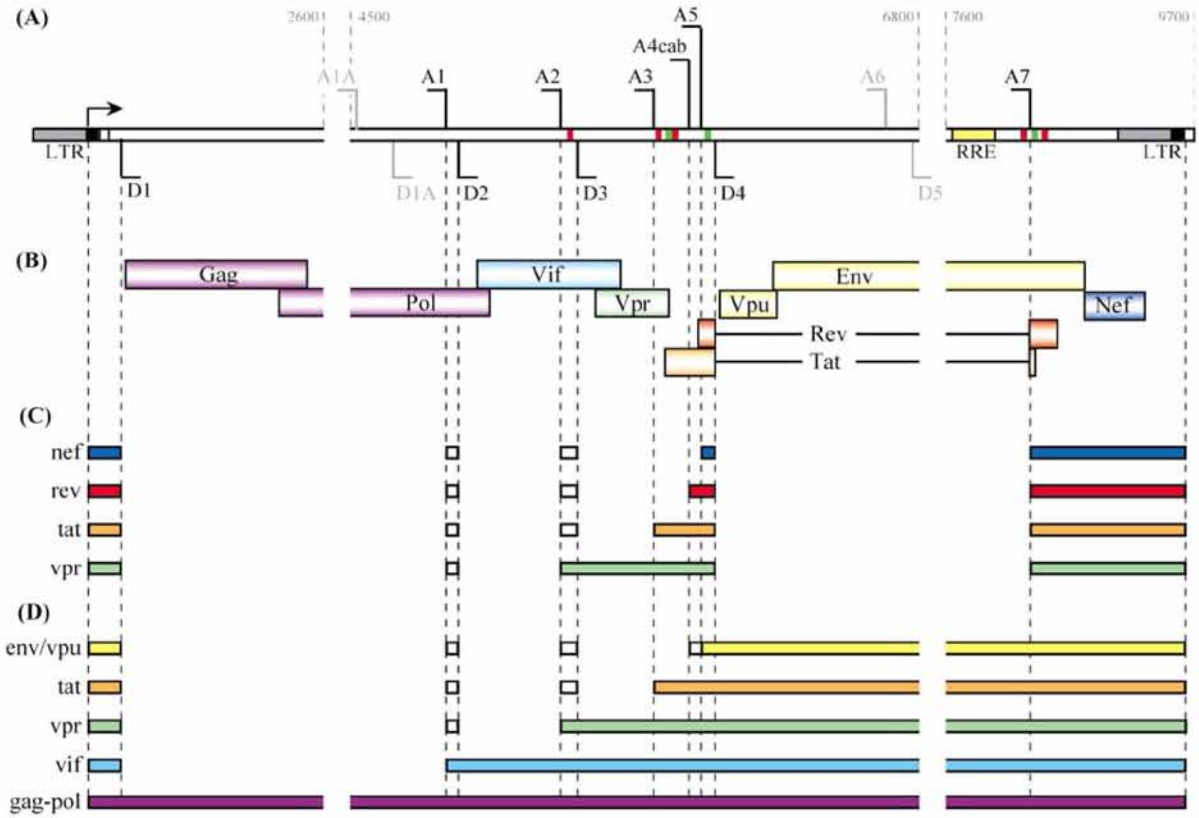


Figure 1. Schematic organisation of the HIV-1 proviral genome. (A) HIV-1 genome contains 5 donor splice sites (D1, D1A, D2, D3 and D4) and 9 acceptor splice sites (A1, A1A, A2, A3, A4c,a,b, A5 and A7) (8,11). Sites A1A and D1A have been reported to be preferentially involved in pre-mRNA stabilization (11). In addition, strains from the IIIB family of HIV viruses contain the additional A6 and D5 pair of sites (95). The LTRs which flank HIV-1 DNA in the host genome are composed of U5 (grey box), R (black box) and U3 (white box). The Rev Response Element (RRE) is represented by a yellow box. Little red boxes represent the ESSV, ESS2p, ESS2, ISS and ESS3 splicing silencer elements. Little green boxes represent ESE2, GAR ESE and ESE3 splicing enhancer elements (ESE3 is both an enhancer and a silencer) (38). Regions 2600 to 4500 (downstream from site A1A) and 6800 to 7600 (upstream of site A7) were not represented to simplify the figure. The nucleotide numbering corresponds to the NL4-3 HIV-1 strain. (B) Organization of the open reading frames of the HIV-1 genome. The nine ORFs code for protein precursors Gag, Gag-Pol and Env, regulatory proteins Tat, Rev and Nef, and auxiliary proteins Vif , Vpr and Vpu. (C) and (D) Messenger RNAs produced during early (C) and late (D) phases of HIV-1 infection. White boxes represent exons which are not systematically included in messenger RNAs.

3. OVERVIEW OF HIV-1 ALTERNATIVE SPLICING IN THE CONTEX T OF VIRAL INFECTION

Alternative splicing plays a key role in the complex process of HIV-1 multiplication. After integration of the proviral cDNA into the host genome, RNA polymerase II (RNA Pol II) recognizes the LTR promoter. In the absence of the viral protein Tat, the vast majority of the transcripts that are produced are very short (TAR RNA), because of transcriptional pausing of RNA Pol II (6,7). In contrast, when Tat is present, its association with the TAR element, located at the 5' extremity of the transcripts, allows the recruitment of cyclin T1 and CDK9, components of p-TEFb. Consequently CDK9 phosphorylates the CTD of the large subunit of RNA Pol II and other transcription factors, resulting in the synthesis of full length viral RNA (9229 nucleotides in the case of the HIV-1 Bru strain). This RNA has numerous functions in virus multiplication. It is encapsidated as genomic RNA to produce new viruses and also used as messenger RNA for translation of the Gag and Gag-Pol precursor proteins. On the other hand, it has to be spliced to produce mRNAs encoding the Env precursor protein as well as numerous regulatory and auxiliary proteins (Figure 1B). One of the regulatory proteins, Rev, plays a key role in the transport of a fraction of the unspliced transcripts into the cytoplasm.

HIV-1 RNA splicing is very complex, as HIV-1 RNA contains at least 5 splicing donor sites (5' ss) and 8 to 9 splicing acceptor sites (3' ss) (8-11) (Figure 1A). The combined utilization of these splice sites leads to the production of at least 40 distinct mRNAs in infected cells (8,12). In early stages of the multiplication cycle, transcripts are multiply spliced. They predominantly

Donor splicing site		Acceptor splicing site	Branch point	Pyrimidine tract	Splicing site
Consensus	A/C AG | GU A AGU	Consensus	YNYURAY	(Y)n	YAG | G
D1 (95)	CUG | GUGAGU	A1 (26)	nd	UUUUCGGGUUUAUUACAG	G
D1A (nd)	CAG | GUAAGA	A1A (nd)	nd	UACUUCCUCUUAAAAUUAG	C
D2 (38)	AAG | GUGAAG	A2 (60)	UAGCAGA	UAUUUUGAUUGUUUUUCAG	A
D3 (45)	AAG | GUAGGA	A3 (55)	nd	CUGCUGUUUAUCCAUUUCAG	A
		A4c (48)	UUGCUAUUGUAAA	UGUUGCUUUCAUUGCCAAG	U
		A4a (48)	CUUUCAUUGCCAAG	UUUGUUUCAUGACAAAAG	C
		A4b (48)	CUUUCAUUGCCAAG	UUUGUUUCAUGACAAAAGCUUAG	G
		A5 (48)	GUUUCAUGACAAAAGCCUUAG	UCUCCUAUGGCAG	G
D4 (83)	GCA | GUAAGU	A6 (nd)	nd	UGUGUUAGUUUAAAG	U
D5 (nd)	AAG | GUGCAG	A7 (41)	UACUUUC ---- UAGGCAGGGAUAU	UUAUCGUUUCAG	A

Figure 2. Sequence of the HIV-1 5' (donor) and 3' (acceptor) splice sites. Nucleotides of 5' ss which do not base-pair with U1 snRNA are indicated in red, and those which are expected to base-pair with U6 snRNA are marked by an asterisk. Exon-intron and intron-exon junctions are represented by verticals bars. Numbers in brackets indicate the relative efficiency of utilization (in %) of splice sites in vivo (17). Underlined nucleotides represent identified or putative branch points (18,34,35). N=A, C, G or U; Y=C or U; R=A or G.

encode the regulatory Tat and Rev proteins, which are absolutely required for virus multiplication, as well as Nef, whose absence results in a lower efficiency of multiplication (13). During this early stage, alternative splicing of HIV-1 RNA takes place on the basis of the relative strengths of the various 5' and 3' splice sites (Figure 1C).

In a second step of cell infection by virus HIV-1, the Rev protein plays a key role in the fate of viral transcripts. Indeed, when enough Rev has been produced, it recognizes a strong binding site within the Rev Response Element (RRE) and multimerizes on this element, allowing the CRM1-dependent export of unspliced or singly spliced HIV-1 RNAs into the cytoplasm (14). At this late stage of the multiplication cycle, a large fraction of the transcripts is transported in an intact form into the cytoplasm. The singly spliced mRNAs encode the Env precursor and the Vif, Vpr and Vpu auxiliary proteins (Figure 1). Vpr, which has an early function in the transport of the pre-integration complex into the nucleus, plays several important functions for virus multiplication (15). Vif is involved in virus assembly (16). Therefore, at this stage, both the Rev action and the relative strengths of the splice sites D1, D2 and D3 and A1 to A5, define the splicing pattern of HIV-1 RNA (Figure 1D).

Although the HIV-1 5' ss are not optimal, most of them have a rather good complementary with U1 snRNA, especially D1 and D4 (Figure 2). The relative strengths of the donor sites were evaluated both in vitro and in vivo (17,18), and they correlate well with the level of complementarity with U1 snRNA. This suggests that regulation at the 5' ss is limited and up to now, only a few splicing enhancers were found to act at 5' ss: for example the GAR element located upstream of site D4, binds the SR proteins ASF/SF2 and SRp40, and activates site D4 utilization (19,20).

Site D1 is the most efficient of all the HIV-1 5' ss. All the downstream splicing events depend on its presence since no splicing occurs when it is mutated (21). It is located in an RNA region that contains several important functional elements and has a highly conserved 2D RNA structure in all HIV-1 strains (22,23). A dependency of site D1 efficiency on this peculiar 2D structure was recently shown (24). Recent studies have shown that the low strength of the D2 and D3 5' ss is important to limit splicing at the upstream 3' ss A1 and A2, respectively (25,26) and therefore maintains the required level of unspliced RNA. Interestingly, unusual functions, not related to splicing, have been attributed to some of the HIV-1 5' ss. For instance, site D1 is responsible for preventing premature cleavage and polyadenylation at a 3' processing site located 70 nts downstream from the cap structure (27,28). Another example is the role of site D4 in stabilization of the env mRNA, which relies on the base-pairing of U1 snRNA with the 5' ss and on the stabilization

of U1 snRNP by ASF/SF2 bound to the upstream GAR element (19). A recently identified additional D1A site is also proposed to participate to the stability of unspliced RNA (11). In addition, U1 snRNA complementary sequences in HIV-1 RNA were also recently found to facilitate the recruitment of some components of the transcription machinery (29), reinforcing the idea of cross-talk between transcription and splicing complexes (see also Section 9).

All the HIV-1 3' ss are suboptimal (Figure 2). They include suboptimal polypyrimidine tracts (PPT) (30-32) and non-canonical branch-site sequences (33-35). For some 3' ss, two or more branch points can be used, including nucleotides other than adenosine residues and they exhibit moderate or low intrinsic efficiency (31,33,36,37). In addition, 3' ss are frequently included within highly stable 2D RNA structures, as shown for sites A3, A4b,a,c, A5 and A7 (32,38,39). Their utilization is down regulated by the presence of numerous cis splicing-silencer elements (ESS) that bind hnRNP proteins, most frequently hnRNP A1/B (38,40-45). By counteracting all these negative characteristic features of HIV-1 3' ss, SR proteins (ASF/SF2, SC35, SRp40 or 9G8) allow viral RNA splicing (35,44,46-48). However, this is a slow process as compared to some efficiently spliced cellular pre-mRNAs, so that RNA folding is likely to take place prior to splicing. In addition, HIV-1 RNA splicing is an ordered process, as introns are eliminated according to their order of appearance, starting from the 5' extremity of the RNA molecule (21). Furthermore, the equilibrium between spliced and unspliced RNAs depends also on the presence of the suboptimal D2 and D3 5' ss (25,26).

Studying the complex regulation of HIV-1 RNA alternative splicing is very important to fully understand the biology of this virus and to define new therapeutic strategies to limit its multiplication. It also represents a good model to study the mechanisms of alternative splicing in vertebrates. Three of the eight 3' ss (A2, A3 and A7), which are conserved in all HIV strains, are subjected to very tight regulation. They have been intensively studied, by us and others: sites A2 and A3 are used for vpr and tat mRNA production, respectively. Site A7 is used for both tat and rev mRNA production (Figure 1). A competition between hnRNP A1/B and SR proteins was found to occur at these three sites. However, the mechanisms involved at each of these sites display specific characteristics, due to a different organization of the protein binding sites and to the influence of the RNA 2D structure. In the next part of the review, we will focus on these 3 splice sites and will highlight the role of the HIV-1 RNA secondary structure that is known to be crucial for the synthesis of full-length RNA (TAR element) and for the export of unspliced RNA (RRE). When appropriate, we will also complete our review with some additional unpublished data. For complementary information see the review "Modulating HIV-1 RNA processing and utilization" by McLaren et al. in this series and a previously published review by Stoltzfus and Madsen (49).

4. A SMALL REGULATORY ELEMENT THAT FOLDS INTO A STEM-LOOP STRUCTURE CONTROLS SITE A2 UTILIZ ATION

MT Stoltzfus and colleagues first identified a 24-nt sequence downstream from site A2 in the HIV-1 RNA. This sequence, containing three PyUAG motifs (numbered 1, 2 and 3 in Figure 3) was called ESSV (V for Vpr) because its binding to hnRNP A1 was found to have a strong negative effect on site A2 utilization (41) (Figure 3). They also showed that this negative effect on site A2 results from the limited association of U2AF65 to the PPT (50). In parallel, we found that the SR protein ASF/SF2 strongly activates site A2 utilization, both in vitro and in cellulo (48). Among the four SR proteins that we tested in cellulo (SC35, SRp40, 9G8 and ASF/SF2), only ASF/SF2 strongly favoured site A2 utilization at the expense of sites A1, A3, A4b,a,c and A5 (48) and this was confirmed by the Darlix team, using an HIV-1 pNL4-3 clone (51). In agreement with these data, the IDC16 indole derivative that specifically suppresses splicing activation by ASF/SF2 is a strong inhibitor of site A2 utilization in vivo and in vitro, and the addition of a large excess of recombinant ASF/SF2 in vitro can compensate partially the inhibitory effect of IDC16 (52). More recently, Madsen and Stoltzfus refined the ESSV limits by generation of dinucleotide substitutions in the 24-nt sequence of ESSV within plasmid p PSP (25). This plasmid contains an HIV-1 pNL4-3 cDNA truncated within the D1-A1 intron, but it retains all the HIV-1 splice sites which have the same relative efficiency as in the entire HIV-1 RNA (32,41). ESSV was redefined as a 16-nt element containing PyUAG motifs 2 and 3, and an important AUAG element (motif 4 in Figure 3). Mutations in the (Py/A)UAG elements 2-4 have a strong positive effect on splicing at site A2. In contrast, mutations in the sequence between elements 1 and 4 have a negative effect on splicing, suggesting that they contain an enhancer element (25). Importantly, when some of the mutations in the restricted ESSV were generated in the entire pNL4-3 HIV-1 virus, replication was reduced by 95% and exon 3 inclusion was increased as expected from the increased A2 activity. In addition, the levels of unspliced RNA and p55-Gag were strongly reduced (25).

We recently obtained additional structural information which explains the competitive action of hnRNP A1 and ASF/SF2 at site A2 (submitted data). First, the 2D structure of the HIV-1 RNA region, which contains exon A2-D3 and both the functional and regulatory elements of site A2, was established in vitro, using chemical and enzymatic probes in conditions of in vitro splicing assays (Figure 3). As shown in figure 3, exon A2-D3 folds into a stem-loop structure (SLS 4-A2). The possibility to form this 2D structure is highly conserved in HIV-1 strains. In the established structure, the identified A2 branch-point sequence (BP) is involved in a stable base-pair interaction. In contrast, the polypyrimidine tract (PPT) is only partially base-paired and site A2 is located at the junction between two helices. The 4 (Py/A)UAG elements of the ESSV element are located in the apical part of SLS 4-A2 (Figure 3).

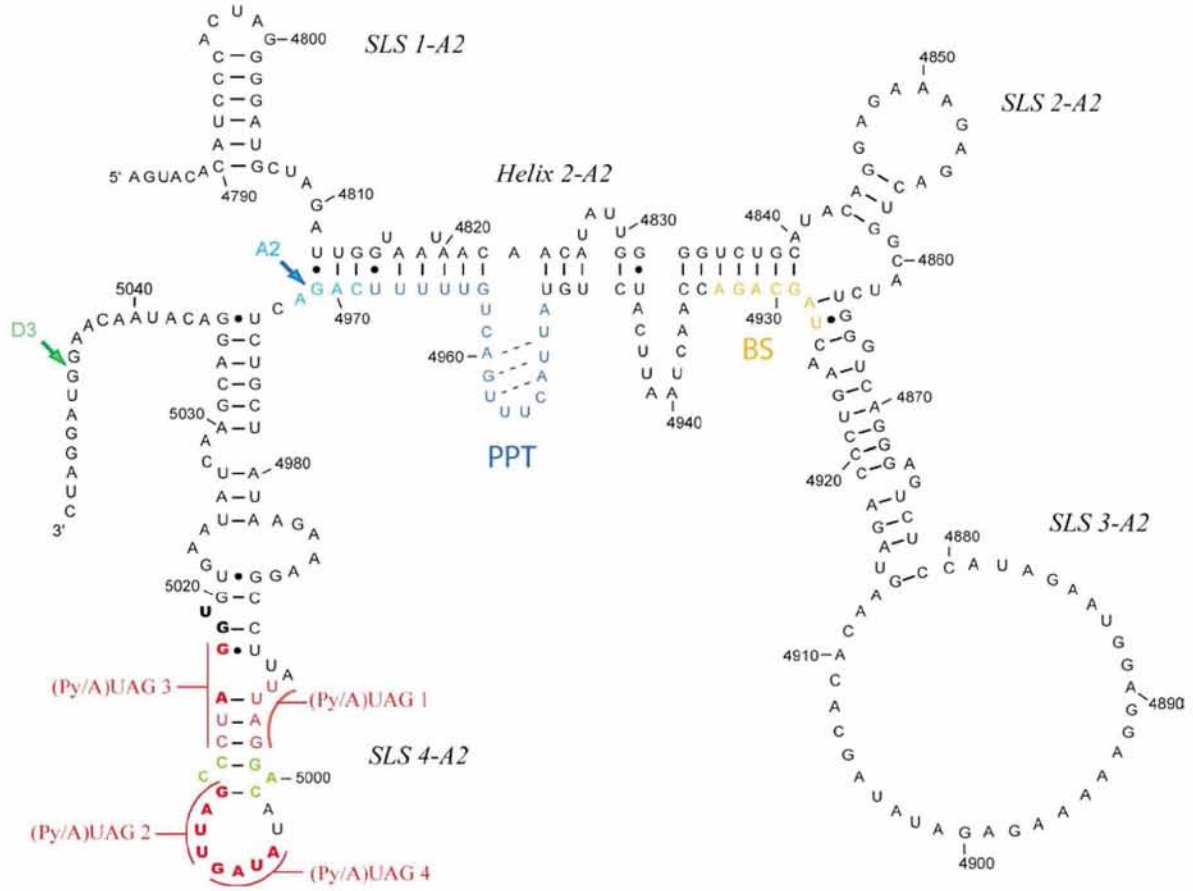


Figure 3. Proposed secondary structure for the HIV-1 Bru region containing the A2 3' ss and its functional and regulatory element. The structure of stem-loop sequences SLS 1,2,3,4-A2 and helix 2-A2 was defined by enzymatic and chemical probing (submitted data). The branch-site sequence (BS) is in orange. The polypyrimidine tract (PPT) and the A2 3' ss are in blue, site D3 is in green and the intron-exon and exon-intron junctions are shown by arrows. Sequences in red correspond to the three PyUAG elements numbered 1, 2 and 3 initially identified as putative hnRNP A1 binding sites in ESSV (41) and the AUAG motif ((Py/A)UAG 4) that is important for splicing inhibition in cellulo together with elements 2 and 3 (25). Residues required for ESSV activity according to Madsen and Stoltzfus are in bold (black or red), and residues which may correspond to an enhancer element are in green.

A precise delineation of the hnRNP A1 binding sites on SLS 4-A2 was obtained by footprinting analysis of complexes formed between a HIV-1 Bru RNA fragment (4807 to 5039) and recombinant hnRNP A1 (data submitted). As expected, protections were detected in the apical part of SLS 4-A2 that contains ESSV and the strong protection of the G residue 5007 in the AUAG motif confirmed hnRNP A1 binding to this motif. In addition, the 5' strand of the internal loop I in SLS 4-A2 (4983 to 4989), as well as site A2, were also protected (Figure 4A). In contrast, the entire PPT sequence became highly accessible in the presence of hnRNP A1, suggesting that the previously demonstrated inhibition of U2AF65 binding to PTT (50) does not result from a simple masking of the PTT by hnRNP A1. A complete protection of the SLS 4-A2 region from position 4970 to 5018, including the intron-exon junction, was also observed when RNA probing was performed using HeLa cell nuclear extract rather than pure hnRNP A1 (data submitted). Once again, only a very limited protection of the PTT sequence was obtained (Figure 4A) in agreement with Stoltzfus's data showing a weak binding of U2AF65 (50). These data demonstrate that competition between hnRNP A1 and splicing factors takes place at the intron-exon junction and not at the PPT.

Several polypurine-rich sequences known to be recognized by ASF/SF2 (53,54) are present in the intron sequence located upstream of site A2. However, their mutation or deletion had no effect on splicing efficiency at site A2, or on its activation by ASF/SF2 (submitted data). In contrast, AG to CU substitutions in the (Py/A)UAG 2, 3 and 4 of ESSV increased the splicing activity of the resulting substrate in vitro and in cellulo (25). For a more precise delineation of the ASF/SF2 binding sites, the complexes formed between an HIV-1 Bru RNA fragment (4986 to 5010) and recombinant ASF/SF2 were analysed

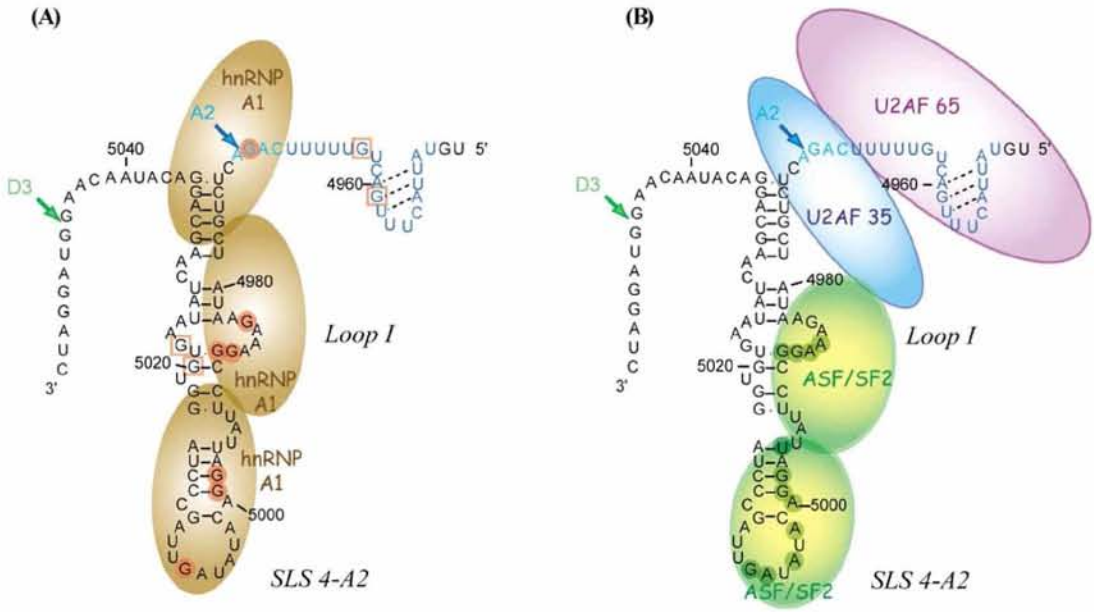


Figure 4. Proposed model for hnRNP A1 and ASF/SF2 binding to the HIV-1 RNA region containing the 3' ss A2. (A) Schematic model of binding and silencing activity of hnRNP A1 based on footprinting data on complexes formed with recombinant hnRNP A1 (submitted data). (B) Schematic model of binding and enhancing activity of ASF/SF2 based on footprinting data on complexes formed with recombinant ASF/SF2 (submitted data). Nucleotides in full circles correspond to nucleotides which are protected by hnRNP A1 (T1 RNase, panel A) or ASF/SF2 (chemical modifications and RNAse digestions, panel B). In panel A the nucleotides in squares show an increased reactivity to T1 RNase in the presence of hnRNP A1. The polypyrimidine tract (PPT) and 3' ss A2 are in blue.

by enzymatic footprinting assays (data submitted). Strong protections were found to extend from position 4986 to 5008 including the AUAG motif 4 of ESSV, suggesting the binding of at least 2 molecules of ASF/SF2 (Figure 4B). In agreement with Madsen and Stoltzfus's data, ASF/SF2 interacts with the segment linking motifs 1 and 4 that was shown to decrease splicing efficiency upon mutation in plasmid p PSP (25). Within the ASF/SF2 protected sequence, only the AAAGGC sequence (4985 to 4990) shows a good homology with the high affinity ASF/SF2 binding sites defined by Tacke and Manley, but we did not identify any sequence matching to the C/G-rich functional ASF/SF2 binding sites as defined by Smith et al. (55,56).

In conclusion, altogether the data show that the 50-nt long segment located downstream from site A2 contains the overlapping silencer and enhancer acting at site A2. HnRNP A1 is able to cover the entire segment, but it does not mask the PPT sequence. As this sequence is suboptimal, its stable recognition by U2AF65 is expected to require the binding of U2AF35 to the intron-exon junction. However, binding of hnRNPA1 to this junction might preclude U2AF35 association and therefore impair the stable binding of U2AF65 to the PTT. In contrast, when ASF/SF2 binds downstream of site A2, it may on the one hand compete out hnRNP A1 binding (RRM-dependent effect). On the other hand, through an interaction of its RS domain with those of both U2AF subunits, it may favour the binding of U2AF35 and U2AF65 to the intron-exon junction and to the PPT, respectively, and activate splicing at site A2. An involvement of both the RRM and the RS domain of ASF/SF2 is in agreement with our previous in cellulo data showing that over-expression of an ASF/SF2 protein truncated of its RS domain, in the presence of plasmid p PSP, leads to an increase of vpr mRNA production in HeLa cells, but at a level lower than that obtained when the full-length protein is over-expressed (48).

5. A COMPLEX EX TENDED EX ON REGION REGULATES SITE A3 UTILIZ ATION

Two splicing silencer elements have been found to limit site A3 utilization: the ESS2p element, located very close to site A3, that binds protein hnRNP H (32), and the ESS2 element, located farther downstream (70 nts downstream of site A3), that binds hnRNP A1 (40,57). Our previous structural analysis demonstrated that site A3 is located in the terminal loop of a stem-loop structure called SLS 2-A3 (32) (Figure 5). Both the PPT and ESS2p element are involved in the stem (5' and 3' strands, respectively). The PPT is intrinsically suboptimal and its base-pair interaction with the exon sequence likely limits its accessibility (32) (Figure 5). Hence, as seen for site A2, the efficient binding of U2AF65 to the PPT of site A3 should depend on the association of U2AF35 with the intron-exon junction. We previously proposed that the binding of hnRNP H to this element might limit U2AF35 association with site A3 by steric hindrance, explaining how ESS2p decreases site A3 activity (32).

RNA structure and protein factors in HIV-1 splicing

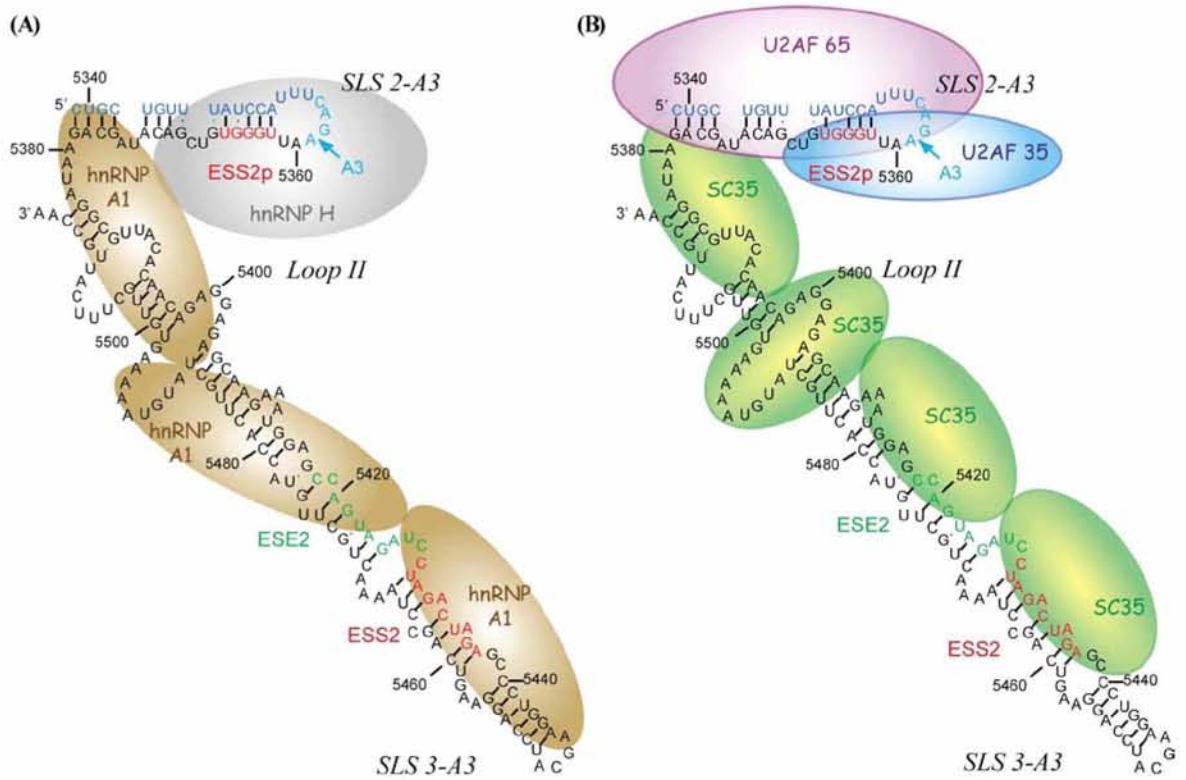


Figure 5. Two dimensional structure of the HIV-1 RNA region containing the 3' ss A3 and model for site A3 silencing by hnRNP proteins and enhancing by protein SC35 (46,47). Red and green sequences represent the splicing silencer (ESS2p and ESS2) and the splicing enhancer (ESE2), respectively. The polypyrimidine tract (PPT) and splice site A3 are in blue. (A) Model of site A3 inhibition by hnRNP A1 and H. (B) Model of site A3 activation by SC35.

Using several molecular approaches as well as NMR analyses, we recently succeeded in explaining how hnRNP A1 binding to ESS2 strongly limits site A3 activity, despite its far downstream location (47). According to the 2D structure that we previously established (32), ESS2 is located in the apical part of a long irregular stem-loop structure (SLS 3-A3) (Figure 5). By site-directed mutagenesis of SLS 3-A3, footprinting assays of complexes formed between hnRNP A1 and wild type or mutated RNAs, as well as gel-shift analyses, we showed that ESS2 plays a key role in the association of hnRNP A1 with the entire SLS 3-A3 region and with the junction between SLS 2-A3 and SLS 3-A3 (47). We also provided detailed information on how hnRNP A1 associates with this HIV-1 RNA region.

HnRNP A1 contains two RNA Recognition Motifs (RRMs) and a Gly-rich domain that allows its multimerization (58,59). The latter domain was also proposed to stabilize the interaction of hnRNP A1 with nucleic acids (60). By NMR analysis, we demonstrated the direct binding of the two RRMs of hnRNP A1 (UP1 protein) to ESS2 and to the terminal stem-loop of SLS 3-A3 but surprisingly, this interaction did not disturb the RNA 2D structure (47). By enzymatic footprinting experiments using complexes formed between an RNA containing SLS 2-A3 and SLS 3-A3 (SLS 2,3-A3 RNA) and increasing amounts of either hnRNP A1 or UP1, we showed that in both cases, two molecules of protein bind simultaneously to the SLS 3-A3 structure: one interacts with the two CUAG motifs of ESS2 and the terminal stem-loop, while the second one binds to an upstream UAG motif contained in the recently identified ESE2 enhancer (46) and to the 3' strand of the internal loop II in SLS 3-A3 (Figure 5A). We propose that this simultaneous binding takes place through an interaction of the RRM domains of UP1 or full-length hnRNP A1 (47). However, the stability of the assembled complex is higher for hnRNP A1 than for UP1. This means that, as previously suggested (58,59), the Gly-rich domain of hnRNP A1 can stabilize its interaction with RNA, either by electrostatic interaction with RNA or by establishing interactions with another nearby Gly-rich domain.

In addition, we observed that mutating one of the two UAG motifs in ESS2 is sufficient to strongly reduce hnRNP A1 binding to the entire SLS 2,3-A3 RNA (47). Taken together, our data support the following model for hnRNP A1 interaction at site A3: hnRNP A1 uses an entry site with highly characteristic features (ESS2) that allows, together with the surrounding sequences, the establishment of a stable RNA-protein interaction involving two hnRNP A1 molecules. Then, other molecules of hnRNP A1 can interact with sites of lower affinity, the interaction being

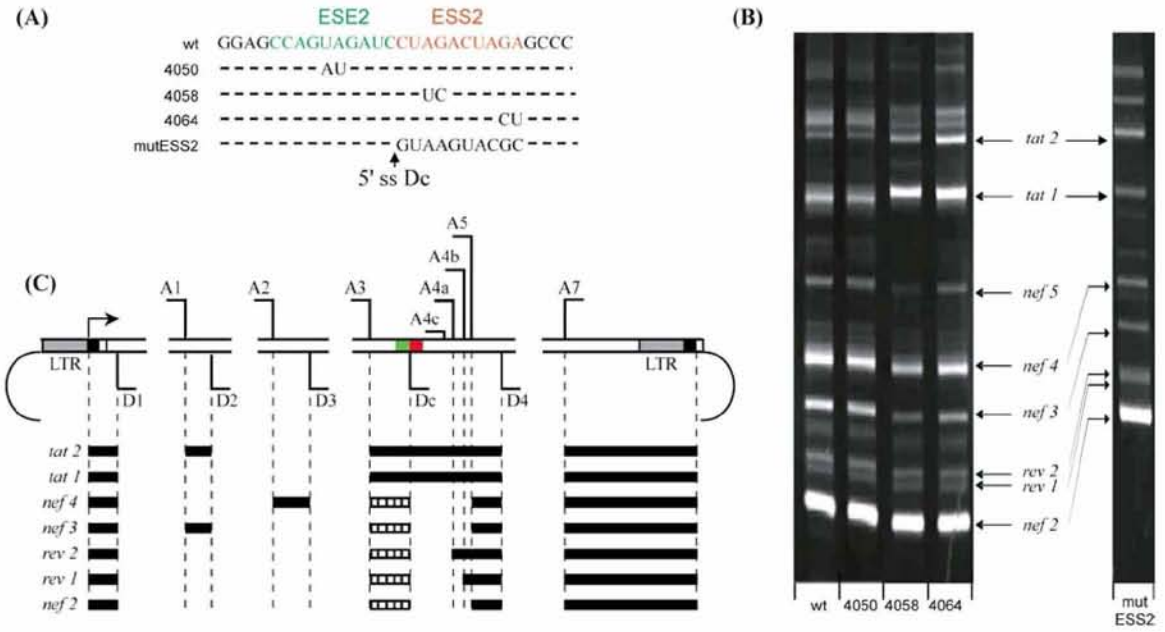


Figure 6. Effect of mutations in ESE2 and ESS2 on the splicing pattern of HIV-1 RNA. (A) Mutations introduced in the p PSP plasmid used in this study (42). Sequences of ESE2 and ESS2 that are located downstream from site A3 are in green and red, respectively. The mutant with substitution of most residues in the ESS2 element (mutESS2) was previously analyzed (40,46,57). (B) RT-PCR analysis of HIV-1 spliced products. HeLa cells were transfected with the p PSP plasmid. For each transfection assay, equal amounts of total RNA were subjected to RT-PCR analysis as previously described (48). The identity of the different mRNAs was confirmed by sequencing when necessary. (C) Schematic representation of HIV-1 genome and exon composition of the different HIV-1 mRNAs. For each mRNA, exons are represented by black boxes. Hashed boxes represent the ectopic exon which is included in the nef and rev mRNAs when the donor splice site Dc is introduced in the mutESS2 construct.

stabilized by protein-protein interactions through the Gly domains. This results in the wrapping of SLS 3-A3 and the SLS 2-A3/SLS 3-A3 junction with hnRNP A1, probably occluding directly the access to the PPT (Figure 5A).

In spite of the strong negative effect of hnRNP A1, site A3 is used in infected cells (8). We previously showed that its activity is highly increased when the SR proteins SC35 or SRp40 are over-expressed in HeLa cells or when recombinant SC35 or SRp40 protein is added in in vitro splicing assays (48). This was in agreement with the description of an SC35-dependent splicing enhancer element (ESE2) adjacent to ESS2 (46) and with results obtained in infected cells (51). By exhaustive footprinting analysis of complexes formed between the SLS 2,3-A3 RNA region and recombinant SC35 or SRp40 protein, we showed that the splicing activation can be explained by binding of these two proteins on several sites which overlap the hnRNP A1 binding sites (47). Interestingly, we found that the strongest SC35 binding site corresponds to the 5' strand of internal loop II in SLS3-A3 (Figure 5B). However, its deletion only reduced the splicing efficiency by a factor of 2.5 (47). Although the affinity of SC35 for ESE2 and ESS2 is lower than for loop II, the strong activation observed upon SC35 and SRp40 binding to ESE2 and ESS2 is due to a direct competition with hnRNP A1 binding to its key entry site in the SLS 2,3-A3 region. These data illustrate the fact that strong SR protein binding sites located downstream from splice sites may have limited functions in splicing regulation. However, they may have other functions, since SR proteins are known to be involved in exon junction complex (EJC) formation, in mRNA transport and in activation of translation (61,62). Therefore, when analysing splicing regulation, it is important to keep in mind that strong SR protein binding sites in exons may not be sufficient per se to define splicing enhancers.

In conclusion, the strong inhibitory activity of ESS2 is due to the capacity of this sequence to generate a complex array of hnRNP A1 interactions on the entire RNA region that extends up to the PPT. Accordingly, mutation in each of the UAGA motifs within ESS2 resulted in the disruption of hnRNP A1 binding to the entire region and to an improved splicing efficiency at site A3 (47). The data previously obtained in vitro using mini HIV-1 RNA substrates (47) or in cellulo using chimeric RNAs (63) were completed here by introducing the mutations in the p PSP plasmid (42) to evaluate their effects in an RNA containing all HIV-1 splice sites. Whereas the UA to AU dinucleotide substitution in ESE2 (variant 4050) had no detectable effect on the relative levels of tat, rev and nef mRNAs, dinucleotide mutations in the two UAG motifs of ESS2 (variants 4058 and 4064) resulted in a strong increase of the steady state levels of tat1 and tat2 mRNAs, at the expense of nef3 to nef5 mRNAs (Figure 6B). These data highlight the central role of the two UAG motifs of ESS2 in site A3 regulation. Accordingly, over-expression of SC35 or

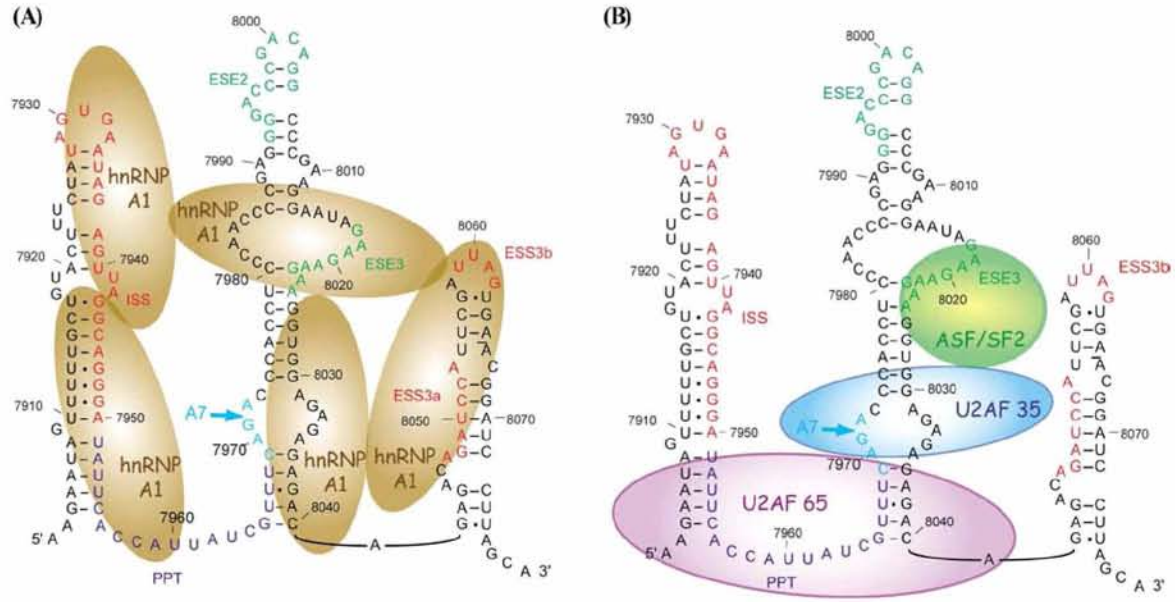


Figure 7. Two dimensional structure established for the HIV-1 region containing the 3' ss A7 and a model of silencing by hnRNP A1 and enhancing by ASF/SF2. Splicing silencers ISS, ESS3a and ESSb are represented in red (31,43,44). Splicing enhancers ESE2 and ESE3 are in green (35,39,64,96). ESE3 is both an enhancer and a silencer element according to footprinting and splicing data (38). The polypyrimidine tract (PPT) and splice site A7 are in blue. (A) Model of site A7 inhibition by hnRNP A1. (B) Model of site A7 activation by ASF/SF2.

SRp40 in HeLa cells transfected with the mutated constructs 4058 and 4064 did not result in a further increase of tat mRNAs (unpublished results). This finding reinforced the idea that the main function of SC35 is to antagonize the action of hnRNP A1 initiated by its binding to ESS2. In previous studies, the function of ESS2 was also tested on chimeric RNAs in which this entire element had been mutated (40,46,57). Hence, we also transferred this ESS2 mutation in the p PSP plasmid and looked at the splicing pattern in HeLa cells (Figure 6). Surprisingly, the splicing pattern was found to be profoundly modified (Figure 6B). By cDNA sequencing, we found that mutation of the original CUAGACUAGA ESS2 sequence to the variant GUAAGUACGC sequence created a new strong 5' ss in ESS2 that we called Dc. As Dc is located between the acceptor sites A3 and A4c, it is used in all the transcripts spliced at sites A4c, A4a, A4b and A5 (Figure 6C). Therefore, the results obtained for site A3 using this mutant might be biased by the generation of this new 5' ss and should be reanalysed carefully.

6. SITE A7 IS REGULATED BY DOWSTREAM AND UPSTREAM ELEMENTS

The numerous studies performed on site A7 revealed the presence of an intronic silencer sequence (ISS) that overlaps the branch site sequences (43), and a bipartite exonic silencer (ESS3), which binds hnRNP A1 (30,31,44,45,64). One exonic enhancer (ESE3) was proposed to bind ASF/SF2 (30,64). Our group and the Kjems group established the secondary structure of the A7 region of HIV-1 RNA (38,39). In this structure, ISS, ESE3 and ESS3 are located in three distinct stem-loop structures, SLS 1-A7, SLS 2-A7 and SLS 3-A7, respectively (Figure 7). In addition, our two groups showed that ESE3 constitutes a nucleation site for binding of hnRNP A1 to the entire HIV-1 region including the 3 SLSs (38,39). In agreement with previous data (44), footprinting assays revealed a multimerization of hnRNP A1 on the segment joining ESE3 to ESS3 (38,39) (Figure 7A). In this case, hnRNP A1 binds cooperatively to both exonic and intronic sequences and allows the propagation of its binding to the entire A7 region. Therefore, site A7 and the branch-site sequences are poorly accessible and splicing is highly reduced. In contrast, ASF/SF2 association with ESE3 is expected to destabilize the entire hnRNP A1 association to this region, allowing the assembly of splicing complexes. We consequently described ESE3 as a Janus element since it has at the same time a negative and a positive activity (38). The apical part of SLS 2-A7 was proposed to have an ASF/SF2-dependent enhancer activity on site A7 (35). However, our recent purification of complexes assembled onto SLS 2-A7 in a nuclear extract suggests the binding of other nuclear proteins at this site (unpublished data).

Finally, in addition to be involved in the modulation of site A7 activity, the ESS3a motif of the bipartite ESS3 silencer element was proposed to limit the Rev-mediated RNA transport via the binding of a cellular factor, although this factor was not identified (45,65). Recent analyses identified hnRNP E1 as a potential ESS3a-binding factor. However, hnRNP E1 was not found to modulate site A7 utilization, but rather to block translation of viral RNAs (66).

7. REGULATORY ELEMENTS LOCATED IN OTHER HIV-1 REGIONS

In addition to the RRE, whose secondary structure has been solved, splicing regulatory elements were detected in two other regions of HIV-1 RNA whose 2D structure has not yet been studied. First, one bidirectional element that binds both ASF/SF2 and SRp40 was found between the A5 3' ss and the D4 5' ss (19,20). SR protein binding to this enhancer is required for efficient utilization of site A5 and for association of the U1 snRNP to site D4. As mentioned above, the ASF/SF2-mediated binding of U1 snRNP to site D4 is required to stabilize the env mRNA (20). However, it should be noted that upon ASF/SF2 over-expression in HeLa cells transfected with the p PSP plasmid or in infected cells, no increase of the utilization of site A5 was detected. Only site A2 utilization was increased with the appearance of a larger amount of vpr mRNAs at the expense of nef mRNAs and, accordingly, production of Nef protein was reduced (48,51). Similarly, upon SRp40 over-expression, site A3 was predominantly activated at the expense of the other sites in the A3 to A5 cluster of 3' ss. Here again, this resulted in the reduced production of Nef (48,51). Therefore, among the A1 to A5 3' ss, sites A2 and A3 are the most strongly activated by ASF/SF2 and SRp40, respectively. Either the ESE element located downstream from site A5 is just used to maintain a basal utilization of site A5 upon ASF/SF2 increase, or the simultaneous binding of the ASF/SF2 and SRp40 proteins to this site is required for activation of site A5. To address this question, it will be interesting to cotransfect HeLa cells with plasmids over-expressing both ASF/SF2 and SRp40, in the presence of plasmid p PSP. Altogether, these data illustrate how results obtained using a pre-mRNA containing a few splice sites have to be compared with data obtained in the full genomic context. The possible concerted action of some SR proteins at given sites should also be considered.

As mentioned above, the low efficiency of the downstream 5' ss D2 and D3 is also involved in the low utilization of the A1 and A2 3' ss, respectively (25,26). The Stoltzfus team recently showed that both a GGGG silencer located immediately downstream from site D2 and an enhancer element found downstream from site A1, which binds the SR protein SRp75, modulate site D2 utilization and consequently site A1 utilization (67). The partner of the GGGG silencer has not yet been identified.

8. HNRNP A1 IS A GENERAL CELLULAR DOWN-REGULATOR OF HIV-1 RNA SPLICING

Taking into account its negative action at most of the HIV-1 3' ss and at some of the 5' ss, hnRNP A1 protein can be considered as one of the most essential cellular factors for limiting HIV-1 RNA splicing, and thus facilitating the Rev-dependent RNA transport of unspliced or partly spliced RNAs. As shown previously for regulation of cellular 5' ss (68), hnRNP A1 regulates HIV-1 3' ss by interacting with an entry site which contains at least one UAG motif. This is followed by a multimerization of hnRNP A1 along the RNA region. Interestingly, the UAG motifs in the entry site may be partially double stranded, like in ESSV and ESS2, but such folding does not prevent hnRNP A1 binding, as shown by our NMR analysis (see Section 5) (47). According to HIV-1 probing data, the presence of polypurine sequences may favour the propagation of hnRNP A1 along the RNA. Such polypurine sequences, especially when they are single-stranded, can also be recognized by SR proteins ASF/SF2, SC35 and SRp40, limiting hnRNP A1 propagation. Noticeably, whereas several of the HIV-1 splicing sites are activated by ASF/SF2, activation by SC35 is mostly restricted to tat mRNA production through the activation of the A3 3'ss (48,51). In line with this observation, the SC35 concentration in macrophages is increased during the first two weeks after infection, whereas the hnRNP A/B and H concentrations are decreased (69). Around the peak of virus production the opposite situation is observed. Concentration of hnRNP A/B proteins increases, while that of SC35 decreases. Therefore, soon after infection, the virus could alter the expression of host genes so that the level of splicing inhibitory proteins is reduced and that conditions are optimized to activate the synthesis of the tat mRNAs, allowing virus replication. Note that an increased expression of SC35 was also demonstrated two days after HIV-1 infection of the H9 lymphocyte cells (70), whereas expression of 9G8 was downregulated after infection of T cells (71). Interestingly, a severe dephosphorylation of some SR proteins, which is known to be detrimental for their function was observed in the course of HIV-1 infection (72). A similar phenomenon was also observed after infection by adenovirus or herpes virus (73,74, Akusjarvi in this series).

In contrast to hnRNP A1, hnRNP H was only found to play a negative role at the A3 3' ss (42). Moreover, intronic hnRNP H binding sites that have a positive role on splicing were recently identified upstream of site D3 (75), and it was proposed that proteins of the hnRNP F/H family may loop out the RNA and therefore favour some splicing events, by a mechanism similar to that found for proteins of the hnRNP A1/B family (76). Similarly, it should also be noted that in the strains from the IIIB family, activation of the additional A6D site depends on an enhancer that is bound by both hnRNP H and SC35 (77). Finally, another cellular protein, p32, also plays an essential role in HIV-1 splicing inhibition, through its interaction with Tat (see below).

9. VIRAL PROTEINS MAY LIMIT HIV-1 RNA SPLICING

By enabling the nuclear export of unspliced or partly spliced HIV-1 RNAs, Rev can be considered as a factor which allows the viral pre-mRNA to escape from the splicing process. However, how Rev interferes with spliceosome assembly remains unclear (49). Besides Rev, other viral proteins have the potential to influence splicing. Indeed, it was recently shown that Vpr, through its interaction with the SAP145 splicing factor, may limit cellular pre-mRNA splicing by interfering with the formation of the SAP145-SAP49 complex (78,79). Therefore, Vpr might also modulate viral RNA splicing.

RNA structure and protein factors in HIV-1 splicing

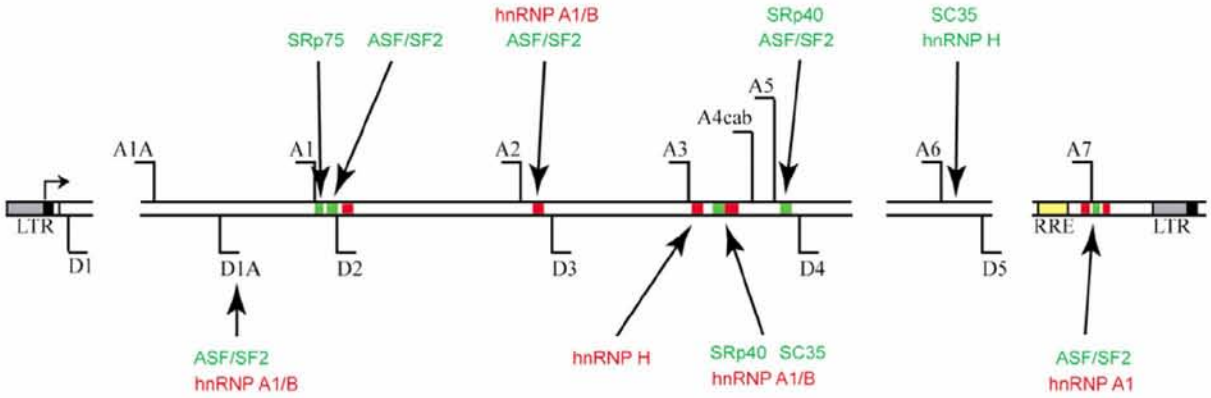


Figure 8. General overview of the regulation of HIV-1 RNA splicing by negative and positive nuclear factors. Cis- and trans-acting splicing activators and repressors are in green and red, respectively. Little red boxes represent (from left to right) GGGG (downstream from D2), ESSV, ESS2p, ESS2, ISS and the ESS3 silencers. Little green boxes represent (from left to right) the SRp75-binding ESEVif and the ESE-M1/M2 bound by ASF/SF2 (upstream of site D2), ESE2, GAR ESE and ESE3.

The Tat protein can modulate HIV-1 RNA splicing through various mechanisms that largely depend on the coupling between transcription and splicing. For instance, a recent study showed that replacing the HIV U3 region of the 5' LTR by a CMV promoter, in a simplified **env/nef** reporter (extending from D4 5' ss to the 3' LTR, making Tat synthesis impossible) led to more efficient splicing, and this effect was reversed upon co-expression of Tat (80). In line with this, another report showed that Tat can induce exon skipping in a non-HIV reporter, a result that was attributed to an increased transcription efficiency (81). This effect of Tat on splicing could also be indirect and mediated by proteins such as Tat-SF1 or SKIP, two components of the splicing machinery that can associate with pTEFb and stimulate Tat-dependent transcription elongation (82,83).

Another likely effector of Tat-mediated HIV-1 splicing inhibition is the cellular p32 protein. The first hint of its role came from the comparison of HIV-1 infection in human or murine cells. It is known for many years that murine cells do not support efficient HIV-1 multiplication, which has impeded scientists working on HIV from using the mouse as an animal model (84). This is due at least to specific amino acid changes between mouse and human in 2 proteins: cyclin T1 and p32, which respectively explain a lack of pTEFb/Tat/TAR-mediated activation of transcription and overly efficient HIV-1 RNA splicing in mouse cells (85,86). Therefore the p32 protein appears to be critical to explain the natural inefficiency of HIV-1 splicing in human cells, which is necessary to ensure the production of unspliced and partly spliced mRNA. Interestingly, while described a few years ago as an inhibitor of the SR protein ASF/SF2 (87), p32 is also known to have functions during transcription, especially in virus-infected cells. For example, p32 induces an hyperphosphorylation of the RNA Pol II CTD during adenovirus infection and blocks the expression of the adenovirus major late unit (88). Recently, Berro and colleagues showed that the acetylated form of HIV-1 Tat interacts efficiently and specifically with p32 protein (89). Interestingly, treatment of HIV-1 infected cells with the deacetylase inhibitor trichostatin A induced the recruitment of p32 to the LTR promoter and a shift of splicing from multiply to singly spliced isoforms. Tat acetylation is essential for HIV-1 multiplication since its effects on the Tat/pTEFb complex lead to the establishment and maintenance of an active transcription elongation complex (90,91). The strong processivity of this complex should not by itself favour splicing, especially at the suboptimal HIV-1 splicing sites (92). We can therefore anticipate that the recruitment of p32 by acetylated Tat helps to further inhibit splicing during the late phase of viral infection through a molecular mechanism which has yet to be determined, but which could involve a sequestration of ASF/SF2, one of the major activators of HIV-1 pre-mRNA splicing.

10. CONCLUSIONS: LESSONS FROM HIV-1 SPLICING FOR THE DESIGN OF FUTURE THERAPIES

Overall modulation of viral splicing during HIV-1 infection may be divided into 2 major phases. Early, during infection, the cellular splicing machinery is fully operational and the activity of SR proteins secures the production of **tat** and **rev** mRNAs, in some kind of "default" pathway. When a threshold concentration of Tat protein is reached in a later phase, the tremendous increase of RNA Pol II processivity might by itself alter the splicing pattern by preventing the utilization of the weakest splice sites. Altogether, the concerted action of Tat and cellular factors such as hnRNP A1 or p32, the action of Rev and a possible modulation of the activity of SR proteins, lead to a strong inhibition of HIV-1 splicing (especially the D4-A7 splicing) and to the synthesis of the proteins required for the assembly of new viruses. On the basis of this dual phase process, novel therapeutic strategies targeting splicing can be envisaged in two directions. The first approach would aim at decreasing splicing efficiency in the early phase to prevent Tat and Rev production. An alternative approach would be to mimic the situation that prevails in mouse cells, and improve the splicing of viral pre-mRNA during

the late phase of infection, preventing the production of the structural proteins.

So far, most efforts have been concentrated on targeting SR proteins, building on the work that has already been done in this field by several groups. For example, Fukuhara and colleagues have analysed the effect of a specific inhibitor of the SR protein kinases SRPK1 and SRPK2 (72). Unfortunately, even though it downregulated SRp75, that chemical compound had no significant effect on HIV-1 replication during a cell line infection assay. Another approach using an indole derivative that selectively inhibits the ESE-dependent splicing activity of ASF/SF2 has been developed recently (93). The results obtained with this kind of compound is more promising, since the drug IDC16 strongly decreased the production of multiply spliced HIV-1 transcripts, including tat, rev and nef mRNAs (52,93). Furthermore, the production of new viruses (measured by the quantification of the viral capsid protein p24) was reduced by at least 90% in macrophages or in activated peripheral blood mononuclear cells. This makes this drug a very promising candidate for anti-viral therapy, since it has a weaker effect than AZT on the expression of a panel of cellular genes (52). Finally, another strategy consists in targeting directly the HIV-1 RNA using U7 snRNA-coupled antisense sequences (94). These molecules, which are designed to mask splice sites and/or ESEs selectively, induce the skipping of specific exons and a strong decrease of the amounts of the targeted mRNA (in this case the tat/rev mRNAs). The consequence of this treatment was an inhibition, albeit incomplete, of HIV-1 production in human T lymphocytes. We can foresee that these different strategies will soon be developed further and that they will allow the emergence of new therapies to complement or improve the current medical treatment of AIDS.

11. ACKNOWLEDGEMENTS

JMS and CFB contributed equally to this review. We thank the « Centre National de la Recherche Scientifique » (CNRS), the « Institut National de la Sante et de la Recherche Medicale » (INSERM), the french « Ministere de l'Enseignement Superieur et de la Recherche », the « Agence Nationale de la Recherche sur le Sida et les hepatites virales » (ANRS 05195/06194), the « Agence Nationale pour la Recherche » (ANR-05-BLAN-0261-01) and the European Alternative Splicing Network (EURASNET, 6th Framework Program) for financial support. JMS was supported by a PhD studentship from ANRS. We are grateful to Chris Browning for his proofreading of the manuscript and to A. Cochrane for sending his review prior to publication in the same series.

12. REFERENCES

1. Johnson, J. M., J. Castle, P. Garrett-Engele, Z. Kan, P. M. Loerch, C. D. Armour, R. Santos, E. E. Schadt, R. Stoughton and D. D. Shoemaker: Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science, 302, 2141-4 (2003)

2. Wang, Z. and C. B. Burge: Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. RNA (2008)

3. Matlin, A. J., F. Clark and C. W. Smith: Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol , 6, 386-98 (2005)

4. Maniatis, T. and R. Reed: An extensive network of coupling among gene expression machines. Nature, 416, 499-506 (2002)

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

6. Karn, J.: Tackling Tat. J Mol Biol , 293, 235-54. (1999)

7.Peterlin, B. M. and D. H. Price: Controlling the elongation phase of transcription with P-TEFb. Mol Cell, 23, 297-305 (2006)

8. Purcell, D. F. and M. A. Martin: Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J. Virol., 67, 6365-78 (1993)

9. Benko, D. M., S. Schwartz, G. N. Pavlakis and B. K. Felber: A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. J Vir ol, 64, 2505-18 (1990)

10. Salemi, M., K. Strimmer, W. W. Hall, M. Duffy, E. Delaporte, S. Mboup, M. Peeters and A. M. Vandamme: 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, 276-8 (2001)

11. Lutzelberger, M., L. S. Reinert, A. T. Das, B. Berkhout and J. Kjems: A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. J Biol Chem, 281, 18644-51 (2006)

12. Schwartz, S., B. K. Felber, D. M. Benko, E. M. Fenyo and G. N. Pavlakis: Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J Vir ol, 64, 2519-29 (1990)

13. Kestler, H. W., 3rd, D. J. Ringler, K. Mori, D. L. Panicali, P. K. Sehgal, M. D. Daniel and R. C. Desrosiers: Importance of the nef gene for maintenance of high virus loads and for development of AIDS. Cell, 65, 651-62 (1991)

14. Hope, T. J.: The ins and outs of HIV Rev. Arch Biochem Biophys, 365, 186-91 (1999)

15. Popov, S., M. Rexach, G. Zybarth, N. Reiling, M. A. Lee, L. Ratner, C. M. Lane, M. S. Moore, G. Blobel and M. Bukrinsky: Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. EMBO J , 17, 909-17 (1998)

16. Ohagen, A. and D. Gabuzda: Role of Vif in stability of the human immunodeficiency virus type 1 core. J Vir ol, 74, 11055-66 (2000)

17. O'Reilly, M. M., M. T. McNally and K. L. Beemon: Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. Virology, 213, 373-85 (1995)

18. Damier, L.: Etude de la régulation de l'épissage du transcrit primaire du virus de l'immunodéficience humaine de type 1. Université de Nancy 1 (1997).
19. Kammler, S., C. Leurs, M. Freund, J. Krummheuer, K. Seidel, T. O. Tange, M. K. Lund, J. Kjems, A. Scheid and H. Schaal: 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, 421-34 (2001)
20. Caputi, M., M. Freund, S. Kammler, C. Asang and H. Schaal: A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. J Vir ol, 78, 6517-26 (2004)
21. Bohne, J., H. Wodrich and H. G. Krausslich: Splicing of human immunodeficiency virus RNA is position-dependent suggesting sequential removal of introns from the 5' end. Nucleic Acids Res, 33, 825-37 (2005)
22. Baudin, F., R. Marquet, C. Isel, J. L. Darlix, B. Ehresmann and C. Ehresmann: Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. J Mol Biol , 229, 382-97 (1993)
23. Abbink, T. E., M. Ooms, P. C. Haasnoot and B. Berkhout: The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. Biochemistry, 44, 9058-66 (2005)
24. Abbink, T. E. and B. Berkhout: RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. J Vir ol, 82, 3090-8 (2008)
25. Madsen, J. M. and C. M. Stoltzfus: An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. J Vir ol, 79, 10478-86 (2005)
26. Madsen, J. M. and C. M. Stoltzfus: 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 (2006)
27. Ashe, M. P., L. H. Pearson and N. J. Proudfoot: The HIV-1 5' LTR poly (A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. EMBO J , 16, 5752-63 (1997)
28. Ashe, M. P., A. Furger and N. J. Proudfoot: Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. RNA, 6, 170-7 (2000)
29. Damgaard, C. K., S. Kahns, S. Lykke-Andersen, A. L. Nielsen, T. H. Jensen and J. Kjems: A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. Mol Cell , 29, 271-8 (2008)
30. Staffa, A. and A. Cochrane: The tat/rev intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. J Vir ol, 68, 3071-9 (1994)
31. Si, Z., B. A. Amendt and C. M. Stoltzfus: 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, 861-7 (1997)
32. Jacquenet, S., D. Ropers, P. S. Bilodeau, L. Damier, A. Mougin, C. M. Stoltzfus and C. Branlant: Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing. Nucleic Acids Res, 29, 464-78 (2001)
33. Damier, L., L. Domenjoud and C. Branlant: The D1-A2 and D2-A2 pairs of splice sites from human immunodeficiency virus type 1 are highly efficient in vitro, in spite of an unusual branch site. Biochem Biophys Res Commun 237, 182-7 (1997)
34. Swanson, A. K. and C. M. Stoltzfus: Overlapping cis sites used for splicing of HIV-1 env/nef and rev mRNAs. J Biol Chem, 273, 34551-7 (1998)
35. Tange, T. O. and J. Kjems: 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, 649-62 (2001)
36. Dyhr-Mikkelsen, H. and J. Kjems: Inefficient spliceosome assembly and abnormal branch site selection in splicing of an HIV-1 transcript in vitro. J Biol Chem, 270, 24060-6 (1995)
37. Kammler, S., M. Otte, I. Hauber, J. Kjems, J. Hauber and H. Schaal: The strength of the HIV-1 3' splice sites affects Rev function. Retrovirology, 3, 89 (2006)
38. Marchand, V., A. Mereau, S. Jacquenet, D. Thomas, A. Mougin, R. Gattoni, J. Stevenin and C. Branlant: A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. J Mol Biol , 323, 629-52 (2002)
39. Damgaard, C. K., T. O. Tange and J. Kjems: 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, 1401-15 (2002)
40. Caputi, M., A. Mayeda, A. R. Krainer and A. M. Zahler: hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. EMBO J , 18, 4060-7 (1999)
41. Bilodeau, P. S., J. K. Domsic, A. Mayeda, A. R. Krainer and C. M. Stoltzfus: 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 Vir ol, 75, 8487-97 (2001)
42. Jacquenet, S., A. Mereau, P. S. Bilodeau, L. Damier, C. M. Stoltzfus and C. Branlant: 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, 40464-75 (2001)
43. Tange, T. O., C. K. Damgaard, S. Guth, J. Valcarcel and J. Kjems: The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. EMBO J , 20, 5748-58 (2001)
44. Zhu, J., A. Mayeda and A. R. Krainer: Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell , 8, 1351-61 (2001)
45. Asai, K., C. Platt and A. Cochrane: Control of HIV-1 env RNA splicing and transport: investigating the role of hnRNP A1 in exon splicing silencer (ESS3a) function. Virology, 314, 229-42 (2003)
46. Zahler, A. M., C. K. Damgaard, J. Kjems and M. Caputi: 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, 10077-84 (2004)

47. Hallay, H., N. Locker, L. Ayadi, D. Ropers, E. Guittet and C. Branlant: Biochemical and NMR study on the competition between proteins SC35, SRp40 and hnRNP A1 at the HIV-1 Tat exon 2 splicing site. J Biol Chem (2006)
48. Ropers, D., L. Ayadi, R. Gattoni, S. Jacquenet, L. Damier, C. Branlant and J. Stevenin: Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA. J Biol Chem, 279, 29963-73 (2004)
49. Stoltzfus, C. M. and J. M. Madsen: Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. Curr HIV Res, 4, 43-55 (2006)
50. Domsic, J. K., Y. Wang, A. Mayeda, A. R. Krainer and C. M. Stoltzfus: 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, 8762-72 (2003)
51. Jacquenet, S., D. Decimo, D. Muriaux and J. L. Darlix: Dual effect of the SR proteins ASF/SF2, SC35 and 9G8 on HIV-1 RNA splicing and virion production. Retrovirology, 2, 33 (2005)
52. Bakkour, N., Y. L. Lin, S. Maire, L. Ayadi, F. Mahuteau-Betzer, C. H. Nguyen, C. Mettling, P. Portales, D. Grierson, B. Chabot, P. Jeanteur, C. Branlant, P. Corbeau and J. Tazi: Small-molecule inhibition of HIV pre-mRNA splicing as a novel antiretroviral therapy to overcome drug resistance. PLoS Pathog, 3, 1530-9 (2007)
53. Bourgeois, C. F., F. Lejeune and J. Stevenin: Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. Prog Nucleic Acid Res Mol Biol , 78, 37-88 (2004)
54. Zheng, Z. M.: Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. J Biomed Sci, 11, 278-94 (2004)
55. Tacke, R. and J. L. Manley: The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. EMBO J. , 14, 3540-51 (1995)
56. Smith, P. J., C. Zhang, J. Wang, S. L. Chew, M. Q. Zhang and A. R. Krainer: An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet, 15, 2490-508 (2006)
57. Amendt, B. A., Z. H. Si and C. M. Stoltzfus: 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, 4606-15 (1995)
58. Dreyfuss, G., M. J. Matunis, S. Pinol-Roma and C. G. Burd: hnRNP proteins and the biogenesis of mRNA. Annu. Rev. Biochem., 62, 289-321 (1993)
59. Cartegni, L., M. Maconi, E. Morandi, F. Cobianchi, S. Riva and G. Biamonti: hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. J Mol Biol , 259, 337-48 (1996)
60. Casas-Finet, J. R., J. D. Smith, Jr., A. Kumar, J. G. Kim, S. H. Wilson and R. L. Karpel: Mammalian heterogeneous ribonucleoprotein A1 and its constituent domains. Nucleic acid interaction, structural stability and self-association. J Mol Biol , 229, 873-89 (1993)
61. Sanford, J. R., J. Ellis and J. F. Caceres: Multiple roles of arginine/serine-rich splicing factors in RNA processing. Biochem Soc Trans, 33, 443-6 (2005)
62. Huang, Y. and J. A. Steitz: SRprises along a messenger's journey. Mol Cell , 17, 613-5 (2005)
63. Amendt, B. A., D. Hesslein, L. J. Chang and C. M. Stoltzfus: 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, 3960-70 (1994)
64. Mayeda, A., G. R. Screaton, S. D. Chandler, X. D. Fu and A. R. Krainer: 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, 1853-63 (1999)
65. Pongoski, J., K. Asai and A. Cochrane: Positive and negative modulation of human immunodeficiency virus type 1 Rev function by cis and trans regulators of viral RNA splicing. J Vir ol, 76, 5108-20 (2002)
66. Woolaway, K., K. Asai, A. Emili and A. Cochrane: hnRNP E1 and E2 have distinct roles in modulating HIV-1 gene expression. Retrovirology, 4, 28 (2007)
67. Exline, C. M., Z. Feng and C. M. Stoltzfus: Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Vir ol, 82, 3921-31 (2008)
68. Eperon, I. C., O. V. Makarova, A. Mayeda, S. H. Munroe, J. F. Caceres, D. G. Hayward and A. R. Krainer: Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. Mol Cell Biol , 20, 8303-18 (2000)
69. Dowling, D., S. Nasr-Esfahani, C. H. Tan, K. O'Brien, J. L. Howard, D. A. Jans, D. F. Purcell, C. M. Stoltzfus and S. Sonza: HIV-1 infection induces changes in expression of cellular splicing factors that regulate alternative viral splicing and virus production in macrophages. Retrovirology, 5, 18 (2008)
70. Maldarelli, F., C. Xiang, G. Chamoun and S. L. Zeichner: The expression of the essential nuclear splicing factor SC35 is altered by human immunodeficiency virus infection. Virus Re s, 53, 39-51. (1998)
71. Ryo, A., Y. Suzuki, M. Arai, N. Kondoh, T. Wakatsuki, A. Hada, M. Shuda, K. Tanaka, C. Sato, M. Yamamoto and N. Yamamoto: Identification and characterization of differentially expressed mRNAs in HIV type 1-infected human T cells. AIDS Res Hum Retroviruses, 16, 995-1005 (2000)
72. Fukuhara, T., T. Hosoya, S. Shimizu, K. Sumi, T. Oshiro, Y. Yoshinaka, M. Suzuki, N. Yamamoto, L. A. Herzenberg and M. Hagiwara: Utilization of host SR protein kinases and RNA-splicing machinery during viral replication. Proc Natl Acad Sci U S A , 103, 11329-33 (2006)
73. Kanopka, A., O. Muhlemann, S. Petersen-Mahrt, C. Estmer, C. Ohrmalm and G. Akusjarvi: Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. Nature, 393, 185-7 (1998)
74. Sciabica, K. S., Q. J. Dai and R. M. Sandri-Goldin: ICP27 interacts with SRPK1 to mediate HSV splicing inhibition by altering SR protein phosphorylation. EMBO J , 22, 1608-19 (2003)

75. Schaub, M. C., S. R. Lopez and M. Caputi: Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem, 282, 13617-26 (2007)
76. Martinez-Contreras, R., J. F. Fisette, F. U. Nasim, R. Madden, M. Cordeau and B. Chabot: Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. PLoS Biol , 4, e21 (2006)
77. Caputi, M. and A. M. Zahler: SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. EMBO J , 21, 845-55 (2002)
78. Hashizume, C., M. Kuramitsu, X. Zhang, T. Kurosawa, M. Kamata and Y. Aida: Human immunodeficiency virus type 1 Vpr interacts with spliceosomal protein SAP145 to mediate cellular pre-mRNA splicing inhibition. Microbes Infect, 9, 490-7 (2007)
79. Kuramitsu, M., C. Hashizume, N. Yamamoto, A. Azuma, M. Kamata, Y. Tanaka and Y. Aida: A novel role for Vpr of human immunodeficiency virus type 1 as a regulator of the splicing of cellular pre-mRNA. Microbes Infect, 7, 1150-60 (2005)
80. Bohne, J., A. Schambach and D. Zychlinski: New way of regulating alternative splicing in retroviruses: the promoter makes a difference. J Vir ol, 81, 3652-6 (2007)
81. Nogues, G., S. Kadener, P. Cramer, D. Bentley and A. R. Kornblihtt: Transcriptional activators differ in their abilities to control alternative splicing. J Biol Chem, 277, 43110-4 (2002)
82. Fong, Y. W. and Q. Zhou: Stimulatory effect of splicing factors on transcriptional elongation. Nature, 414, 929-33 (2001)
83. Bres, V., N. Gomes, L. Pickle and K. A. Jones: A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. Genes Dev, 19, 1211-26 (2005)
84. Bieniasz, P. D. and B. R. Cullen: Multiple blocks to human immunodeficiency virus type 1 replication in rodent cells. J Vir ol, 74, 9868-77 (2000)
85. Garber, M. E., P. Wei, V. N. KewalRamani, T. P. Mayall, C. H. Herrmann, A. P. Rice, D. R. Littman and K. A. Jones: The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. Genes Dev, 12, 3512-27 (1998)
86. Zheng, Y. H., H. F. Yu and B. M. Peterlin: Human p32 protein relieves a post-transcriptional block to HIV replication in murine cells. Nat Cell Biol , 5, 611-8 (2003)
87. Petersen-Mahrt, S. K., C. Estmer, C. Ohrmalm, D. A. Matthews, W. C. Russell and G. Akusjarvi: The splicing factor-associated protein, p32, regulates RNA splicing by inhibiting ASF/SF2 RNA binding and phosphorylation. EMBO J , 18, 1014-24 (1999)
88. Ohrmalm, C. and G. Akusjarvi: Cellular splicing and transcription regulatory protein p32 represses adenovirus major late transcription and causes hyperphosphorylation of RNA polymerase II. J Vir ol, 80, 5010-20 (2006)
89. Berro, R., K. Kehn, C. de la Fuente, A. Pumfery, R. Adair, J. Wade, A. M. Colberg-Poley, J. Hiscott and F. Kashanchi: Acetylated Tat regulates human immunodeficiency virus type 1 splicing through its interaction with the splicing regulator p32. J Vir ol, 80, 3189-204 (2006)
90. Kaehlcke, K., A. Dorr, C. Hetzer-Egger, V. Kiermer, P. Henklein, M. Schnoelzer, E. Loret, P. A. Cole, E. Verdin and M. Ott: Acetylation of Tat defines a cyclinT1-independent step in HIV transactivation. Mol Cell , 12, 167-76 (2003)
91. Ott, M., A. Dorr, C. Hetzer-Egger, K. Kaehlcke, M. Schnolzer, P. Henklein, P. Cole, M. M. Zhou and E. Verdin: Tat acetylation: a regulatory switch between early and late phases in HIV transcription elongation. Novartis Found Symp 259, 182-93; discussion 193-6, 223-5 (2004)
92. Nogues, G., S. Kadener, P. Cramer, M. de la Mata, J. P. Fededa, M. Blaustein, A. Srebrow and A. R. Kornblihtt: Control of alternative pre-mRNA splicing by RNA Pol II elongation: faster is not always better. IUBMB L ife, 55, 235-41 (2003)
93. Soret, J., N. Bakkour, S. Maire, S. Durand, L. Zekri, M. Gabut, W. Fic, G. Divita, C. Rivalle, D. Dauzonne, C. H. Nguyen, P. Jeanteur and J. Tazi: Selective modification of alternative splicing by indole derivatives that target serine-arginine-rich protein splicing factors. Proc Natl Acad Sci U S A, 102, 8764-9 (2005)
94. Asparuhova, M. B., G. Marti, S. Liu, F. Serhan, D. Trono and D. Schumperli: Inhibition of HIV-1 multiplication by a modified U7 snRNA inducing Tat and Rev exon skipping. J Gene Med, 9, 323-34 (2007)
95. Gottlinger, H. G., T. Dorfman, E. A. Cohen and W. A. Haseltine: The role of the tnv protein and tnv RNA splicing signals in replication of HIV-1 IIIB isolates. Virology, 189, 618-28 (1992)
96. Staffa, A. and A. Cochrane: Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. Mol Cell Biol , 15, 4597-605 (1995)

Abbreviations: HIV-1: human immunodeficiency virus type 1; SR protein: serine/arginine protein; hnRNP: heterogeneous ribonucleoprotein; LTR: long terminal repeat; TAR: trans-activation response; pTEF-b: positive transcription elongation factor b; CDK9: cyclin dependant kinase 9; CTD: C-terminal domain; RRE; Rev responsive element; ESE: exonic splicing enhancer; ESS: exonic splicing silencer; ISS: intronic splicing silencer; PPT: polypyrimidine tract; SLS: stem-loop sequence; NMR: nuclear magnetic resonance; RRM: RNA recognition motif; RNA Pol II: RNA polymerase II.

Key w ords: HIV-1, alternative splicing, RNA structure, SR proteins, hnRNP proteins, splicing enhancer, splicing silencer

Send corres pondence to: Christiane Branlant, MAEM UMR 7567, Universite Henri Poincare - Nancy I, BP 239, 54506 Vandoeuvre-les-Nancy, France, Tel: 33-383684303, Fax: 33-383684307, E-mail: Christiane.Branlant@maem.uhp-nancy.fr

3. Données obtenues sur la régulation du site A3 par l'élément ESS2

3.1. Etude de l'effet *in cellulo* des mutations 4058 et 4064

Comme mentionné ci-dessus, seules certaines substitutions de dinucléotides dans et autour de l'élément ESS2 augmentaient l'efficacité d'épissage *in vitro* au site A3 (Fig. 78). Nous avons vu dans l'article que les deux mutations 4058 et 4064 réalisées dans chacun des deux motifs UAG de l'élément ESS2 augmentent fortement le taux des ARNm tat1 et tat2 au dépend des messagers nef lorsque des cellules HeLa sont transfectées avec les plasmides pΔPSP mutés. Ceci peut s'expliquer par une forte augmentation de l'utilisation du site A3 au dépend du site A5 qui est en compétition avec A3 (Fig. 79).

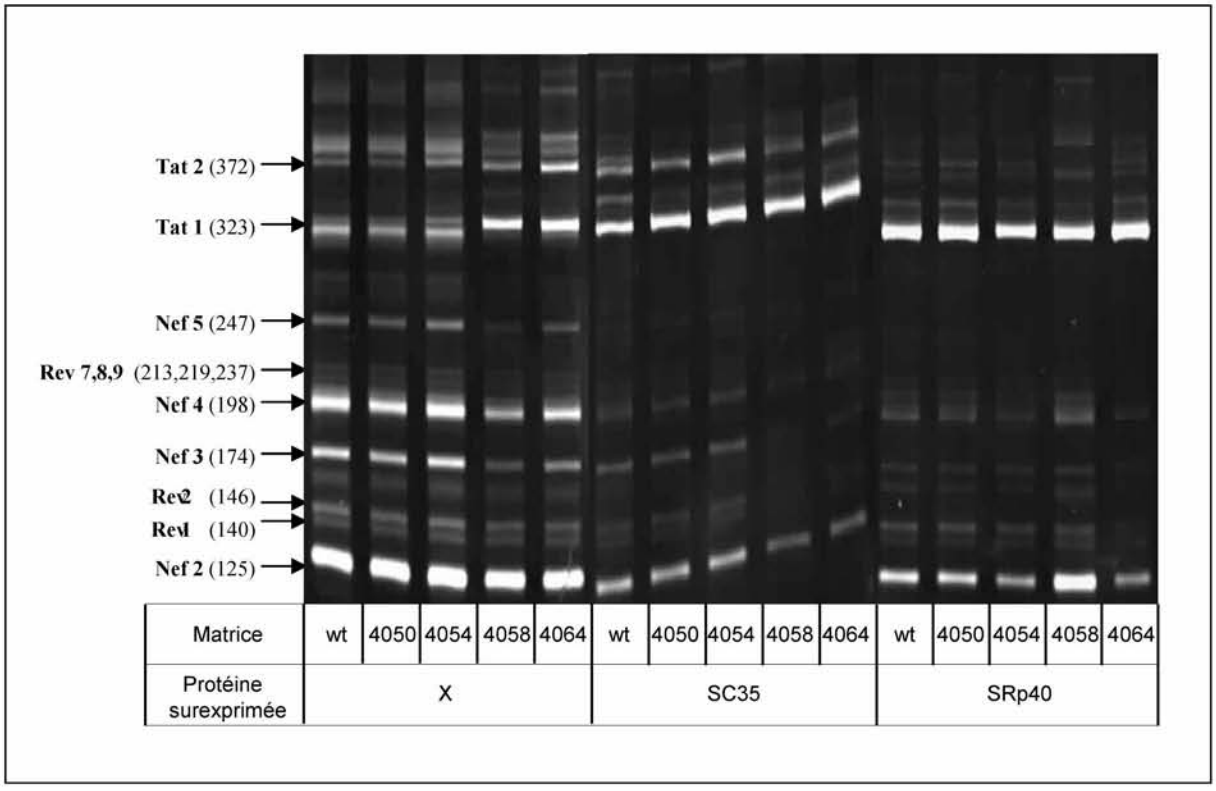


Fig. 79 : Effet des mutations 4050, 4054, 4058 et 4064 sur l'épissage du transcrit du plasmide pΔPSP dans des cellules HeLa

Analyse RT-PCR des produits d'épissage de l'ARN du plasmide pΔPSP sauvage ou portant les mutations 4050, 4054, 4058 et 4064. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP sauvage ou mutés, et le plasmide pXJ vide (X) ou contenant un gène permettant la surexpression des protéines SR SC35 ou SRp40. Chaque analyse par RT-PCR a été réalisée avec la même quantité d'ARN total extrait et avec les oligonucléotides 2143 et 2144 comme décrit dans le chapitre Matériels et méthodes. L'identité des ADNc fractionnés est indiquée sur la gauche de la figure.

Comme on pouvait s'y attendre par ailleurs, nous avons pu constater que le taux des ARNm tat1 et tat2 reste identique quelque soit le plasmide pΔPSP (wt, 4058 ou 4064) co-transfecté avec un plasmide surexprimant les protéines SR SC35 ou SRp40 (Fig. 79). Ceci démontrant à nouveau que le rôle de la fixation de ces protéines SR sur cette séquence et de contrecarrer la fixation, la multimérisation et l'effet inhibiteur de hnRNP A1.

Comme attendu, cet effet des protéines SR restait par contre tout aussi important pour le mutant 4054 qui avait été choisi comme contrôle du fait qu'aucune modification du taux d'épissage n'avait été observée *in vitro* en présence de cette mutation. De même, dans le cas du mutant 4050 qui affecte le motif UAG situé dans l'élément ESE2, une augmentation très forte du taux des ARNm tat1 et tat2 lors de la surexpression des protéines SC35 et SRp40 était observée, ceci du fait que cette mutation n'entraînait aucune augmentation significative des ARNm tat1 et tat2 en absence de surexpression (Fig. 79).

Nos données confirmaient donc l'importance cruciale des deux motifs UAG de l'élément ESS2 dans l'inhibition du site A3, et la capacité des protéines SR SC35 et SRp40 à lever cette inhibition. Les résidus G de ces deux motifs sont en contact direct avec le domaine UP1 de hnRNP A1 comme l'avait montré l'étude RMN réalisée par N. Locker (Hallay *et al.*, 2006). Nous montrions au contraire que l'élément appelé ESE2 a un effet très faible sur l'activité du site A3, ce qui suggère que le site où la compétition entre hnRNP A1 et SC35 est la plus forte est plus l'élément ESS2 que l'élément ESE2 (Hallay *et al.*, 2006).

3.2. Etude de l'effet *in cellulo* de la mutation mutESS2

Nous avons décrit dans l'article le fait que la substitution de la séquence CUAGACUAGA de l'élément ESS2 par la séquence GUAAGUACGC conduit à la génération d'un nouveau site 5' d'épissage dans l'ARN viral (Fig. 6 de l'article ou Fig. 80).

En effet, dans l'analyse RT-PCR obtenue avec le plasmide $p\Delta PSP_{mutESS2}$, le profil de fractionnement était différent de celui obtenu pour le plasmide pΔPSP sauvage. La bande correspondant à l'ARNm tat1 semblait présente mais celles correspondant aux ARNm rev et nef semblaient déplacées vers le haut du gel, suggérant la présence d'une séquence supplémentaire dans ces ARNm. Les ADNc contenus dans les bandes nef2' et tat1 ont donc été extraits du gel, clonés et séquencés, afin de déterminer quels événements d'épissage conduisaient à ces ARNm.

L'analyse de la séquence de l'ADNc de la bande supposée correspondre à tat1 a confirmé son identité. Comme attendu du fait de la mutation de l'élément ESS2, l'ARNm tat1 est bien

produit à un taux plus élevé à partir des transcrits de pΔPSP$_{mutESS2}$, qu'à partir de pΔPSP sauvage (Fig. 6 de l'article ou Fig. 80).

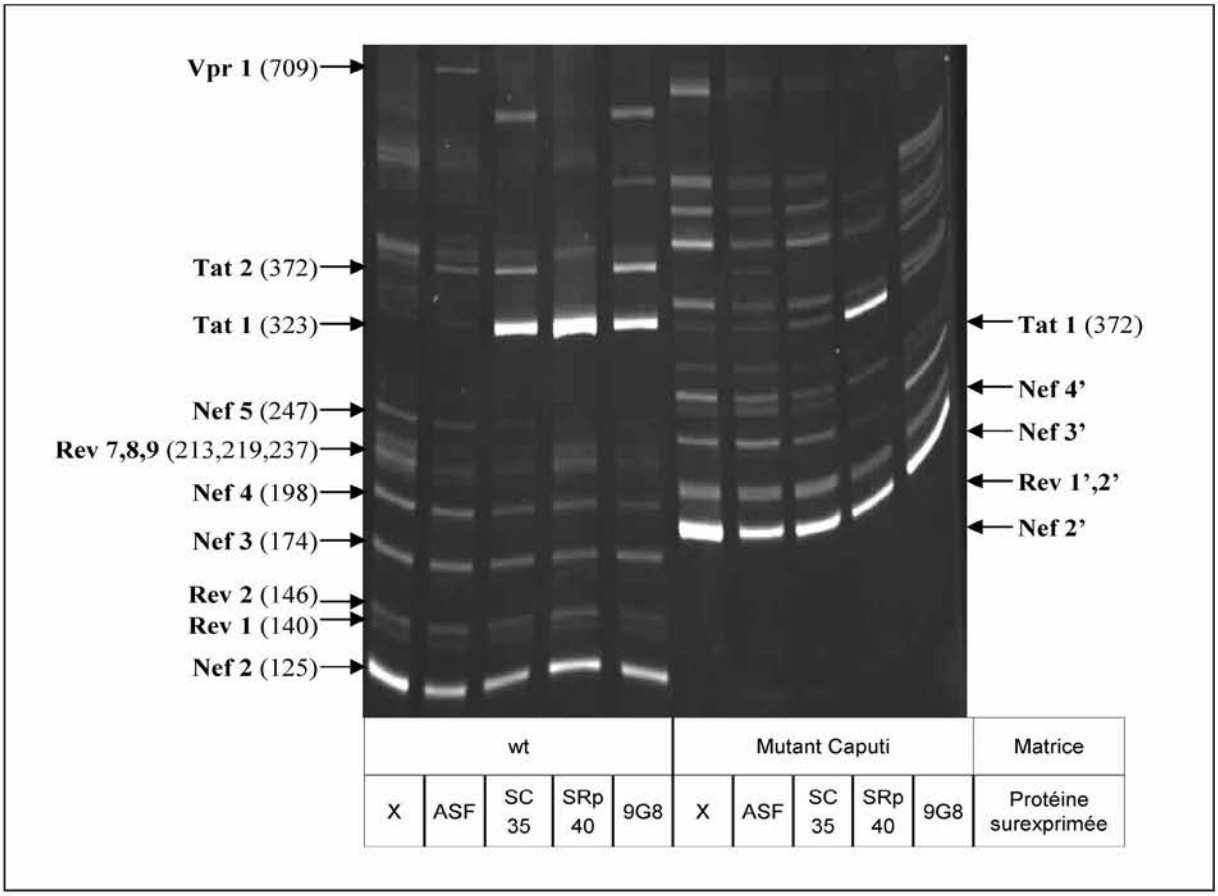


Fig. 80 : Effet de la mutation mutESS2 sur l'épissage du transcrit du plasmide pΔPSP dans des cellules HeLa

Analyse RT-PCR des produits d'épissage de l'ARN du plasmide pΔPSP sauvage ou portant la mutation mutESS2. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP sauvage ou muté, et le plasmide pXJ vide (X) ou contenant un gène permettant la surexpression d'une protéine SR. Chaque analyse par RT-PCR a été réalisée avec la même quantité d'ARN total extrait et avec les oligonucléotides 2143 et 2144 comme décrit dans le chapitre Matériels et méthodes. L'identité des ADNc fractionnés est indiquée sur la gauche de la figure. Dans le cas du plasmide pΔPSP sauvage elle a été définie par référence aux profils déjà caractérisés. Dans le cas du mutant mutESS2, les bandes correspondant aux ARNm nef2' et tat1 ont été séquencées. L'identité des autres bandes a été déduite à partir de cette analyse.

L'analyse de la séquence obtenue pour la bande dénommée nef2' montrait qu'un nouvel exon délimité par le site A3 et un site donneur créé par la mutation était présent en plus des exons

normaux de l'ARNm nef2. Ce site donneur, que nous avons appelé Dc, comporte une séquence parfaitement complémentaire au snRNA U1. Seule la séquence à la bordure 3' de l'exon n'est pas conforme au consensus (Fig. 81A et B).

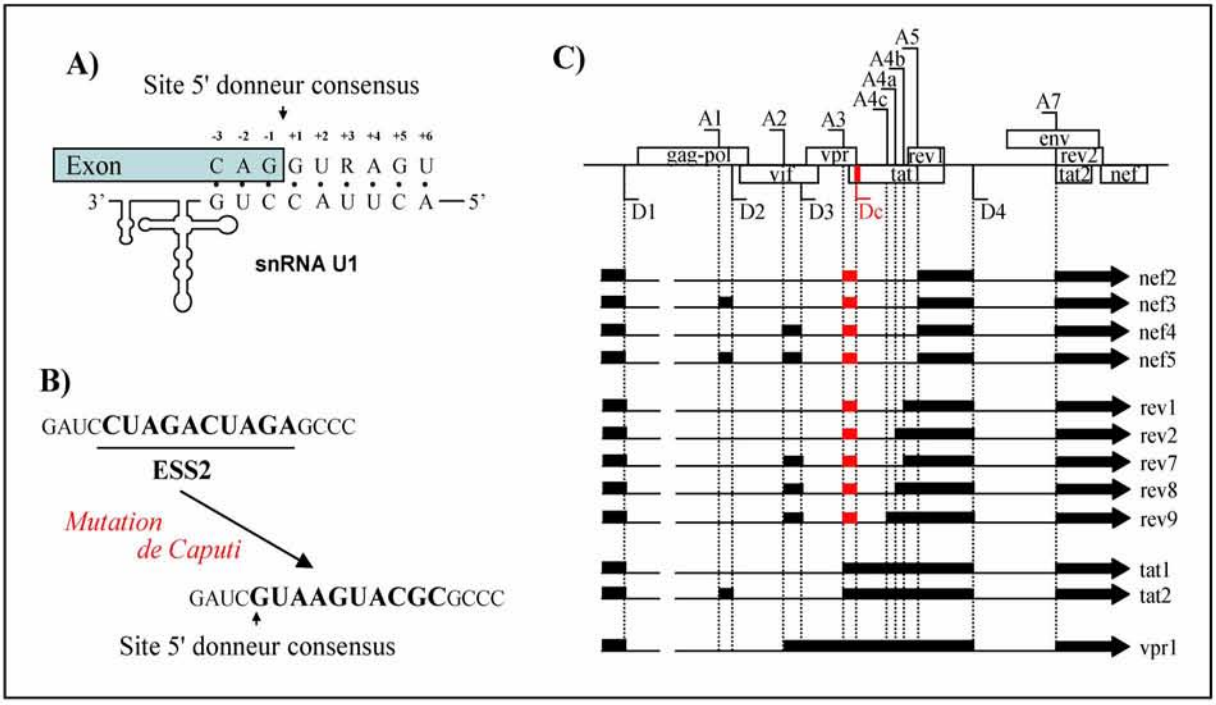


Fig. 81 : ARNm supposés produits par le plasmide pΔPSP portant la mutation mutESS2

A) Rappel de l'appariement formé entre le snRNA U1 et un site 5' d'épissage dont la séquence est celle du consensus des sites 5' d'épissage. B) La mutation mutESS2 génère un site 5' d'épissage. C) Effet supposé de la mutation mutESS2 sur la composition en exons des ARNm produits à partir du transcrit du plasmide pΔPSP et détectables par la technique RT-PCR que nous avons utilisée. Le site donneur Dc et l'exon additionnel créés par la mutation sont indiqués en rouge.

Nous en avons donc conclu que la diminution de mobilité électrophorétique observée pour tous les ARNm qui ne sont pas normalement épissés au site A3 (tous les ARNm nef et rev, Fig. 81C) correspondait à l'insertion quasiment constitutive de l'exon A3-Dc dans ces ARNm. Comme la séquence exonique A3-Dc est contenue dans l'exon A2-D4 et dans l'exon A3-D4, qui sont respectivement présents dans les ARNm tat et vpr, aucun rajout de séquence n'est observé dans ces ARNm, ce qui explique que ces messagers conservent leur mobilité électrophorétique. Néanmoins, des bandes supplémentaires sont observées au dessus de celle correspondant à l'ARNm tat2. Il est possible que la génération d'un site donneur fort au sein de l'ARN du virus HIV-1 rompe l'utilisation ordonnée des sites d'épissage et que des ARN

épissés entre D1 et A1 puissent par exemple être épissés entre Dc et un des sites A4c, a, b ou A5, sans pour autant être épissés entre A2 et D2, et entre A3 et D3.

Lorsque la protéine SR SC35 a été surexprimée en présence du plasmide pΔPSP$_{mutESS2}$, il est intéressant de noter que le taux d'ARNm tat1 n'augmentait pas de manière marquée (Fig. 80). Ceci confirme une fois de plus que l'effet positif de SC35 sur l'activité du site A3 consiste majoritairement à contrecarrer l'action de l'élément ESS2 et plus précisément la fixation de hnRNP A1 sur cet élément. D'ailleurs, quelque soit la protéine SR surexprimée, nous n'avons observé aucun changement marqué des taux relatifs des ARNm tat, rev et nef. Il est intéressant de noter, que le taux de l'ARNm tat1 augmentait lorsque SRp40 était surexprimée. Ceci suggère fortement qu'en plus de son action au niveau de l'élément ESS2, cette protéine a une autre action activatrice de la production de l'ARNm tat1, ce qui est en bon accord avec les résultats antérieurs montrant (1) qu'elle augmente encore plus que SC35 le taux d'ARNm tat1, et (2) que l'ARNm tat2 est peu abondant lors de sa surexpression (Ropers *et al.*, 2004). L'augmentation du taux de tat1 lorsque le plasmide pΔPSP$_{mutESS2}$ est co-exprimé avec SRp40 s'accompagne d'une diminution du taux des ARNm rev et nef, comme c'est le cas pour pΔPSP sauvage. Nous mettons donc en évidence au travers de ces expériences, le fait que SRp40 agit à deux niveaux distincts pour augmenter le taux d'ARNm tat1. Il restera à définir en quoi consiste ce deuxième niveau d'intervention de SRp40.

Un autre point intéressant à noter est l'augmentation du taux d'ARNm nef2' dans le cas du plasmide pΔPSP$_{mutESS2}$, par rapport à ce qui est observé pour le plasmide pΔPSP sauvage. Ceci suggère un couplage très efficace d'une part entre le site donneur D1 et le site accepteur A3, et d'autre part entre le site donneur Dc et le site accepteur A5, puisque ces deux introns sont éliminés dans les ARNm nef2'. D'ailleurs, l'intégration de l'exon A3-Dc dans la quasi-totalité des ARNm nef 2, 3, 4 et rev1, 2 (ARNm nef 2', 3', 4' et rev1', 2') démontre une très forte efficacité du site Dc qui doit participer à une définition optimale de l'exon A3-Dc. Comme les taux relatifs des ARNm rev et nef incluant l'exon A3-Dc étaient peu différents de ceux obtenus pour le plasmide pΔPSP sauvage, à l'exception de nef2, on peut supposer que l'élimination des introns Dc-A4 c, a, b reste une étape limitante dans le processus.

Les données obtenues sur l'effet de la surexpression d'ASF/SF2 sur l'augmentation du taux d'ARNm vpr ne sont pas très claires car des bandes sont présentes à ce niveau même en absence de surexpression de protéine SR. Nous ne sommes pas dans l'état actuel des analyses en mesure d'interpréter les données de manière fiable.

Il faut noter, qu'en plus des données que nous venons de décrire, nos résultats remettent en cause l'interprétation des données faites par l'équipe de M. Caputi lorsqu'elle a utilisé la mutation mutESS2 (Caputi *et al.*, 1999 ; Zahler *et al.*, 2004). Différentes études publiées par cette équipe était en effet basées sur l'analyse de l'épissage *in vitro* de la mini-construction pHS1-X contenant le site donneur D1 couplé aux sites accepteurs A3, A4c,a,b et A5, construction qui avait été réalisée par l'équipe de CM. Stoltzfus (Amendt *et al.*, 1994). Les résultats obtenus montraient que la mutation de l'élément ESS2 dans cette construction, conduisait à une forte activation de l'épissage au niveau du site A3. Néanmoins, compte tenu de nos résultats, il est vraisemblable que l'activation observée était non pas le résultat de la destruction de l'élément inhibiteur, mais plutôt celui de la création d'un site donneur capable d'activer l'utilisation de A3. Ce phénomène avait déjà pu être constaté par l'équipe de CM. Stoltzfus, qui démontrait notamment que la présence de D4 à l'extrémité 3' de la construction favorisait l'épissage aux sites accepteurs A3, A4c,a,b et A5 (Amendt *et al.*, 1994). Par ailleurs, vu la complexité de la construction et de la mutation utilisée pour démontrer l'existence de l'activateur ESE2 et vu le peu d'effet que nous avons observé après mutation de sa séquence, on peut remettre en question l'existence de cet élément activateur.

3.3. Etude de l'effet de la structure SLS3 sur l'activité de l'élément ESS2

Importance de degré d'appariement de l'élément ESS2

Selon les données obtenues au laboratoire, l'élément ESS2 participe à la formation d'une hélice mais deux de ses résidus A restent non appariés (Jacquenet *et al.*, 2001b ; Fig. 82). H. Hallay et N. Locker avaient montré par des techniques d'empreinte en solution avec des sondes chimiques et enzymatiques, mais aussi par étude de l'empreinte de la protéine UP1 sur l'ARN par RMN, que l'interaction de SLS 3-A3 avec les deux RRM N–terminaux de hnRNP A1 se fait sans altération de la structure 2D de l'ARN (Hallay *et al.*, 2006). Néanmoins, aucune expérience n'avait été réalisée en vue de savoir si la structure 2D de l'élément ESS2 était importante pour son activité d'inhibition du site A3.

Pour tenter de répondre à cette question, j'ai construit deux plasmides pΔPSP variants (ESS2open et ESS2closed), produisant des ARN dans lesquels la structuration de l'élément ESS2 était modifiée, sans que sa séquence soit altérée. Dans l'ARN variant ESS2open la séquence du brin 3' de SLS 3-A3 hybridé à ESS2 a été muté pour abolir les capacités d'appariement avec l'élément ESS2. Dans l'ARN variant ESS2closed, cette même séquence a

au contraire été mutée de façon à générer une hélice parfaite et stable avec l'élément ESS2 (Fig. 82).

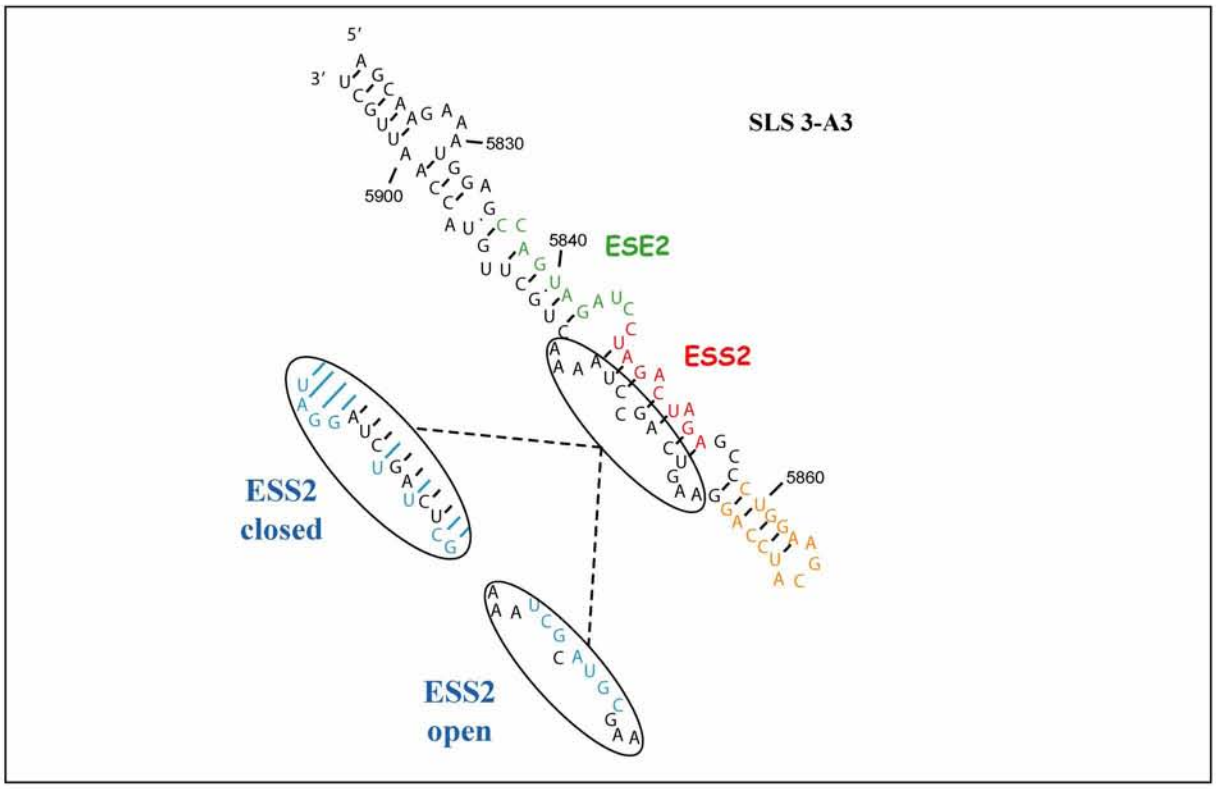


Fig. 82 : Mutations réalisées dans la structure tige-boucle SLS 3-A3 pour étudier l'importance fonctionnelle de la structuration de l'élément ESS2

Les nucléotides mutés sont indiqués en bleu. La délétion de la séquence du motif B (en orange) permet de recréer une boucle apicale sans modification théorique de la structure tige-boucle restante. La numérotation est celle de la souche pNL4-3.

L'effet des mutations a été testé par transfection de cellules HeLa. Trois expériences distinctes de transfections transitoires ont été réalisées et les ARN cellulaires extraits ont été analysés par RT-PCR. Un grand degré de reproductibilité a été obtenu et les résultats présentés dans la Fig. 83 sont représentatifs des 3 séries d'expériences. Les profils d'épissage obtenus en absence de surexpression des protéines SR sont très similaires pour les deux mutants. De manière surprenante en présence d'un excès d'ASF/SF2 l'augmentation du taux de l'ARNm vpr n'est plus observée. Alors qu'un renforcement des taux d'ARNm tat1 et tat2 est détecté surtout pour le mutant ESS2open. Le taux de l'ARNm nef4 est aussi augmenté dans ces conditions. Un des effets les plus visibles des mutations ESS2open et surtout ESS2closed est la baisse du taux d'augmentation de l'ARNm tat1 lorsque les cellules HeLa surexpriment les activateurs connus du site accepteur A3, soit SC35, SRp40 et 9G8 (Fig. 83).

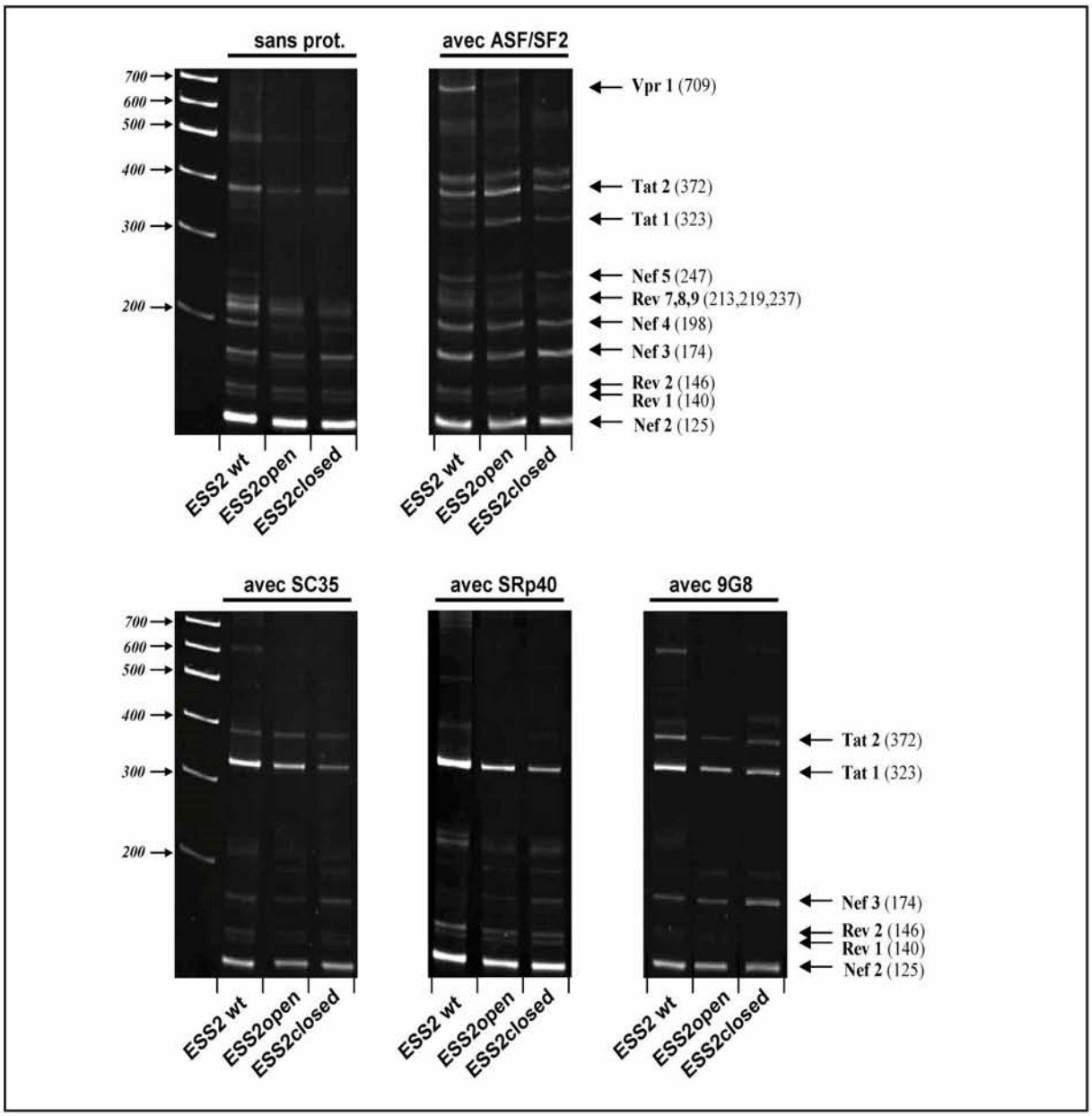


Fig. 83 : Effet des mutations ESS2 open et closed sur l'épissage de l'ARN transcrit à partir du plasmide pΔPSP dans des cellules HeLa ne surexprimant pas ou surexprimant transitoirement certaines protéines SR

Analyse par RT-PCR des produits d'épissage de l'ARN du plasmide pΔPSP sauvage ou muté. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP sauvage ou mutés, et les plasmides pXJ surexprimant une protéine SR recombinante ou pour contrôle le plasmide pXJ vide (X). Chaque analyse par RT-PCR a été réalisée avec la même quantité d'ARN total extrait des cellules HeLa et avec les oligonucléotides 2143 et 2144 comme décrit dans le chapitre Matériels et méthodes. L'identité des ADNc fractionnés correspondant aux différents ARNm de HIV-1 a été déduite des marqueurs de taille et des profils déjà caractérisés. Elle est indiquée sur la gauche des gels.

Comme les variations quantitatives de l'ARNm vpr sont difficiles à étudier par RT-PCR en utilisant le couple d'oligonucléotides 2143 et 2144 du fait de la faiblesse des quantités de messagers vpr. Nous avons complété l'étude par une autre analyse RT-PCR basée sur l'utilisation de l'oligonucléotide 2027 spécifique des ARNm vpr (Fig. 84A). Cette analyse ne permettait pas de distinguer vpr1 de vpr2, mais vpr1 est toujours largement majoritaire (Ropers *et al.*, 2004, observations personnelles). Le résultat de cette analyse est présenté dans la Fig. 84B, il confirme le fait que les mutations ESS2open et ESS2closed engendrent une diminution de la production de ARNm vpr.

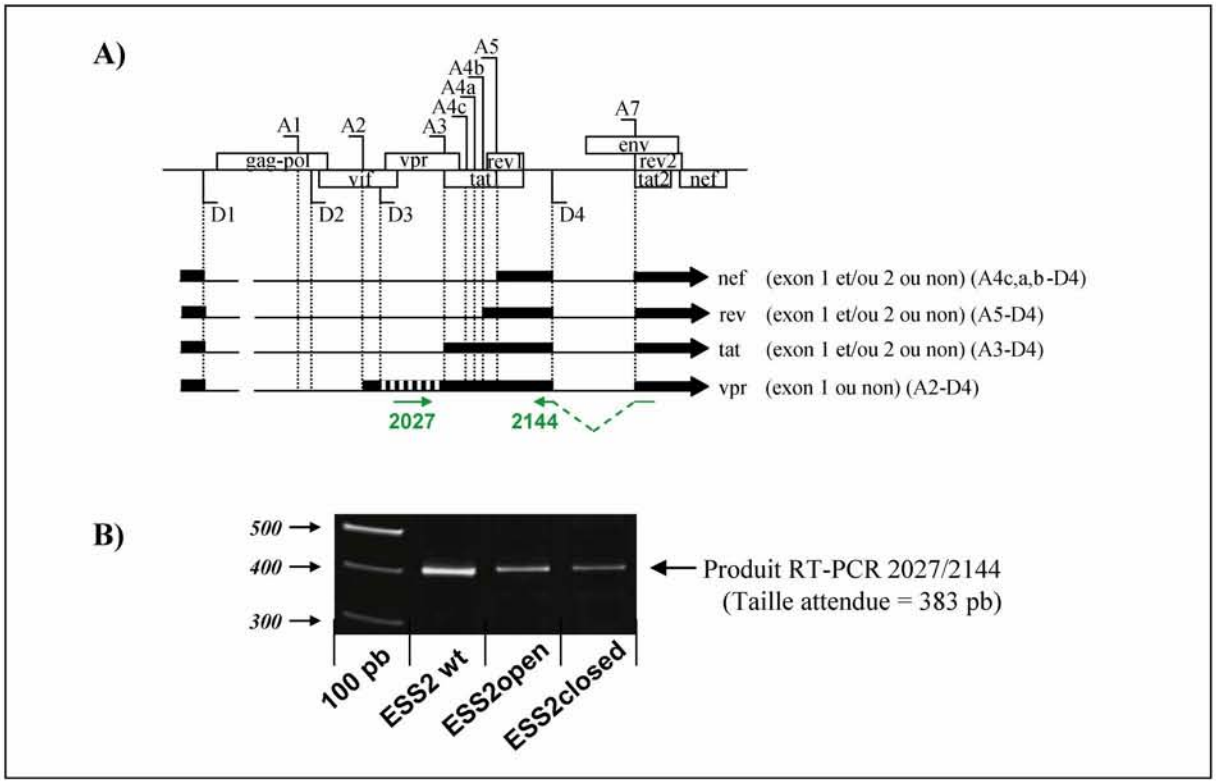


Fig. 84 : Etude de l'effet des mutations ESS2 open et closed sur la production des ARNm vpr par épissage de l'ARN transcrits à partir des plasmides pΔPSP sauvage ou mutés dans des cellules HeLa surexprimant transitoirement la protéine ASF/SF2.

A) Représentation schématique des sites d'hybridation des oligonucléotides 2027 et 2144 utilisés pour l'analyse par RT-PCR (représentés en vert). La séquence hachurée représente la seule séquence intronique qui différencie les transcrits vpr des autres ARNm. B) Analyse par RT-PCR des quantités d'ARNm vpr issus de l'épissage des ARN produits par les plasmides pΔPSP. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP sauvage ou mutés, et le plasmide pXJ-ASF surexprimant la protéine ASF/SF2. Chaque analyse par RT-PCR a été réalisée avec la même quantité d'ARN total extrait des cellules et avec les oligonucléotides spécifiques 2027 et 2144.

En conclusions, nous notons un nombre important d'effets sur le profil d'épissage de l'ARN produit par le plasmide pΔPSP lors d'un changement de la séquence située entre les positions 5874 et 5887 (souche pNL4.3).

Plusieurs paramètres doivent être considérés pour commenter ces données. Bien sûr, les mutations réalisées modifiaient la structuration de l'élément ESS2 et il n'est pas impossible que des changements de la force relative des sites d'épissage, liée à un changement de l'activité de l'élément ESS2 entraîne des variations du taux d'utilisation du site A3 et aussi de celle du site A2. Néanmoins, il n'est pas exclu que les variations observées ne soient pas uniquement dues aux changements de structure secondaire de ESS2. En effet, la séquence mutée pourrait aussi intervenir dans la régulation de l'épissage au site A3 en fixant des facteurs de régulation, en particulier les protéines SR SC35 et SRp40 comme le suggérait les résultats d'empreintes enzymatiques réalisées par D. Ropers (Hallay *et al.*, 2006) La prise en compte de ces données et de celles obtenues plus haut pour les mutants 4050 et 4054 affectant l'élément ESE2, suggèrent que les sites de fixation de SC35 et SRp40 ayant la potentialité de chasser hnRNP A1 de son site fort ESS2 sont plutôt la séquence 5874 à 5887 qui s'apparie à ESS2, que la séquence amont.

Dans l'état actuel des données, il n'est donc pas facile de donner des explications précises aux résultats obtenus. Néanmoins, il apparaît clairement à partir de ces données que la substitution d'une séquence au sein d'une région régulatrice d'un des sites d'épissage a des répercussions marquées sur plusieurs des sites et une fois de plus nos données montrent l'importance d'utiliser la méthode globale d'étude des sites d'épissage. Bien sûr elle devra être compléter par des études *in vitro* de fixation des facteurs nucléaires sur les ARN mutés pour progresser dans l'interprétation des données.

Importance du motif B

Comme nous l'avons déjà mentionné, les travaux réalisés en collaboration entre notre équipe et celle d'E. Guittet par RMN, montraient que le motif B correspondant à la structure tige-boucle apicale de la tige-boucle SLS 3-A3 est important pour préserver la structuration de l'élément ESS2. Nous avons donc décidé de tester l'effet d'une délétion de ce motif B (Fig. 82) toujours en utilisant le plasmide pΔPSP et en surexprimant ou pas les protéines SR.

Il est intéressant de noter que les effets observés sont assez similaires à ceux observés pour les mutant ESS2closed et ESS2open, à savoir une diminution de l'augmentation du taux de l'ARNm tat1 par les protéines SR SC35 et SRp40 (elle est légèrement moins forte que pour les deux mutants précédents) et une absence d'apparition de l'ARNm vpr en surexprimant

ASF/SF2 (Fig. 85). Encore une fois, ces effets peuvent être liés au changement de structure de l'élément ESS2 modifiant la force relative des sites d'épissage, mais aussi à la modification des sites de fixation des protéines SR SC35 et 9G8 ou encore à l'absence d'expression de la protéine Tat.

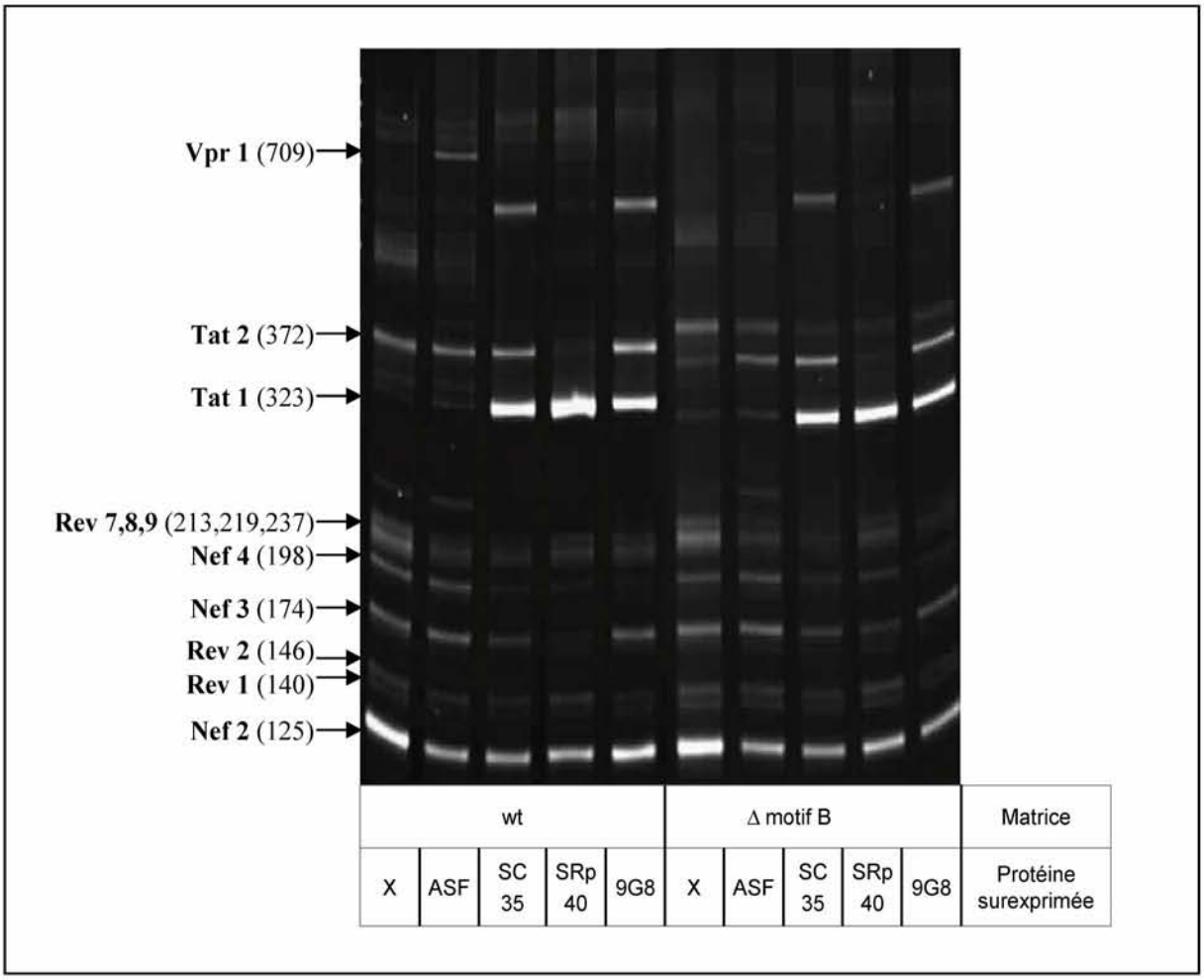


Fig. 85 : Effet de la délétion du motif B sur l'épissage de l'ARN transcrit à partir du plasmide pΔPSP dans des cellules HeLa ne surexprimant pas ou surexprimant transitoirement certaines protéines SR

Analyse par RT-PCR des produits d'épissage de l'ARN du plasmide pΔPSP sauvage ou muté. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP sauvage ou mutés, et les plasmides pXJ surexprimant une protéine SR recombinante ou pour contrôle le plasmide pXJ vide (X). Chaque analyse par RT-PCR a été réalisée avec la même quantité d'ARN total extrait des cellules HeLa et avec les oligonucléotides 2143 et 2144 comme décrit dans le chapitre Matériels et méthodes. L'identité des ADNc fractionnés correspondant aux différents ARNm de HIV-1 a été déduite des marqueurs de taille et des profils déjà caractérisés. Elle est indiquée sur la gauche des gels.

CHAPITRE 7 : Etude de la régulation de l'épissage de l'exon A2-D3

Un point marquant des données présentées dans l'article inséré dans le chapitre précédent était l'observation que l'activation déjà décrite du site A2 par ASF/SF2 (Ropers *et al.*, 2004 ; Jacquenet *et al.*, 2005) est dû au fait que cette protéine empêche la fixation de la protéine hnRNP A1 à l'élément inhibiteur ESSv. Les données obtenues suggèrent qu'ASF/SF2 peut à la fois déplacer hnRNP A1 qui empêche la fixation du facteur U2AF35 à la jonction intron/exon et en plus augmenter la fixation de ce facteur par interaction de son domaine SR avec celui de U2AF35.

La forte augmentation du taux de transcrits vpr en présence d'ASF/SF2 est d'autant plus marquante, que ces transcrits sont produits à un taux très faible sans surexpression d'ASF/SF2. C'est à la fois le cas dans les lymphocytes T transfectés par un ADNc entier (56% des transcrits correspond à des messagers nef, 34% à des messagers rev, 9% à des messagers tat, et seulement 1% % à des messagers vpr ; Purcell and Martin, 1993) et lors de transfection de cellules HeLa par le plasmide pΔPSP, puisqu'ils ne sont pas détectables par RT-PCR sans surexpression d'ASF/SF2 (Ropers *et al.*, 2004). Une explication qui avait été antérieurement émise par l'équipe était la présence d'un inhibiteur de l'utilisation du site A2 dans le segment D3-A3. Cet inhibiteur pouvait être l'élément ISS qui avait été étudié *in vitro* par S. Jacquenet (thèse Jacquenet), d'où nos travaux décrits dans le chapitre suivant.

1. L'élément ISS situé en aval du site A2 n'influence pas le taux d'utilisation de ce site *in cellulo* contrairement aux données antérieurement obtenues *in vitro*

1.1. Etat des connaissances obtenues par L. Damier et S. Jacquenet

En testant la capacité d'épissage de pré-ARNm produits à partir de constructions génétiques contenant les sites d'épissage D1 et A2 de l'ARN du virus HIV-1 et des séquences de tailles variables en aval du site donneur D3, L. Damier avait mis en évidence au laboratoire une séquence ARN (nommée ISS), localisée en aval de ce site donneur D3, qui bloque l'épissage de l'intron D2-A2 *in vitro* (Fig. 86). Le caractère inhibiteur de la séquence ISS avait aussi été vérifié en utilisant un autre site donneur d'épissage issu de l'adénovirus.

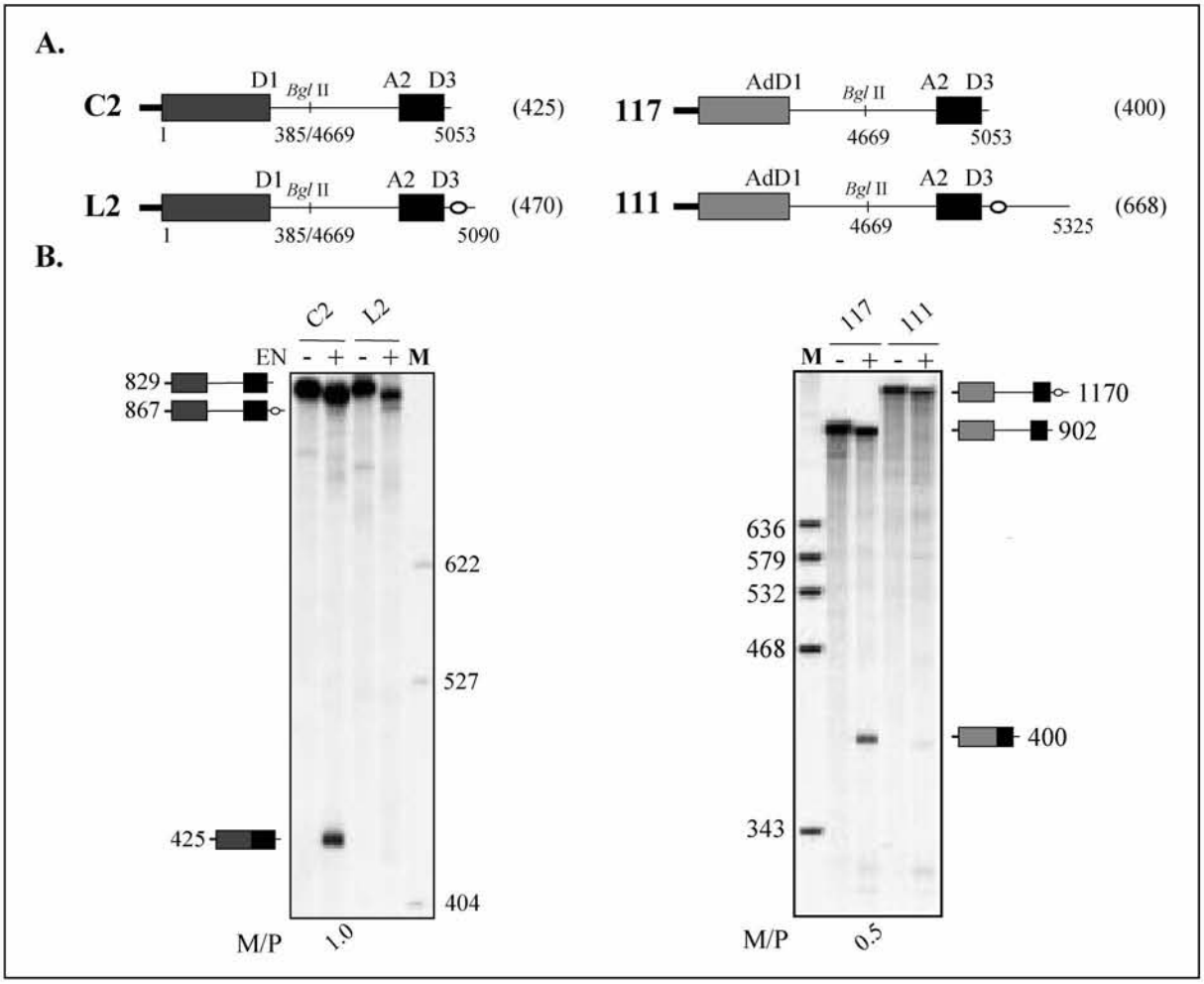


Fig. 86 : Mise en évidence de l'existence d'un élément inhibiteur ISS en aval du site D3 *(Thèse de S. Jacquenet, 2001)*

A) Représentation schématique des transcrits utilisés pour démontrer l'existence d'un élément inhibiteur en aval du site D3. Les transcrits C2 et L2 ne contiennent que des séquences provenant de l'ARN du virus HIV-1, les positions des extrémités des fragments au sein de l'ARN de la souche Bru sont indiquées sur le schéma. Les transcrits 117 et 111 contiennent une partie 5' issue du transcrit tardif de l'adénovirus et une partie 3' issue du virus HIV-1. La taille du produit d'épissage attendu pour chaque transcrit est indiquée entre parenthèse. B) Electrophorèse en gel de polyacrylamide des produits d'épissage des transcrits C2, L2 et 117, 111 obtenus en présence d'extrait nucléaire (EN). Un marqueur de taille (M) et les ARN non épissés sont fractionnés en parallèle.

D'après les données de L. Damier, cet élément semblait agir de manière particulière, en bloquant l'entrée de la tri-snRNP [U4/U6.U5] lors de l'assemblage du spliceosome. Elle avait aussi délimité les résidus de cet élément en définissant un élément nécessaire à l'apparition de l'inhibition (mini ISS) et une séquence adjacente renforçant les propriétés d'inhibition (Fig. 87).

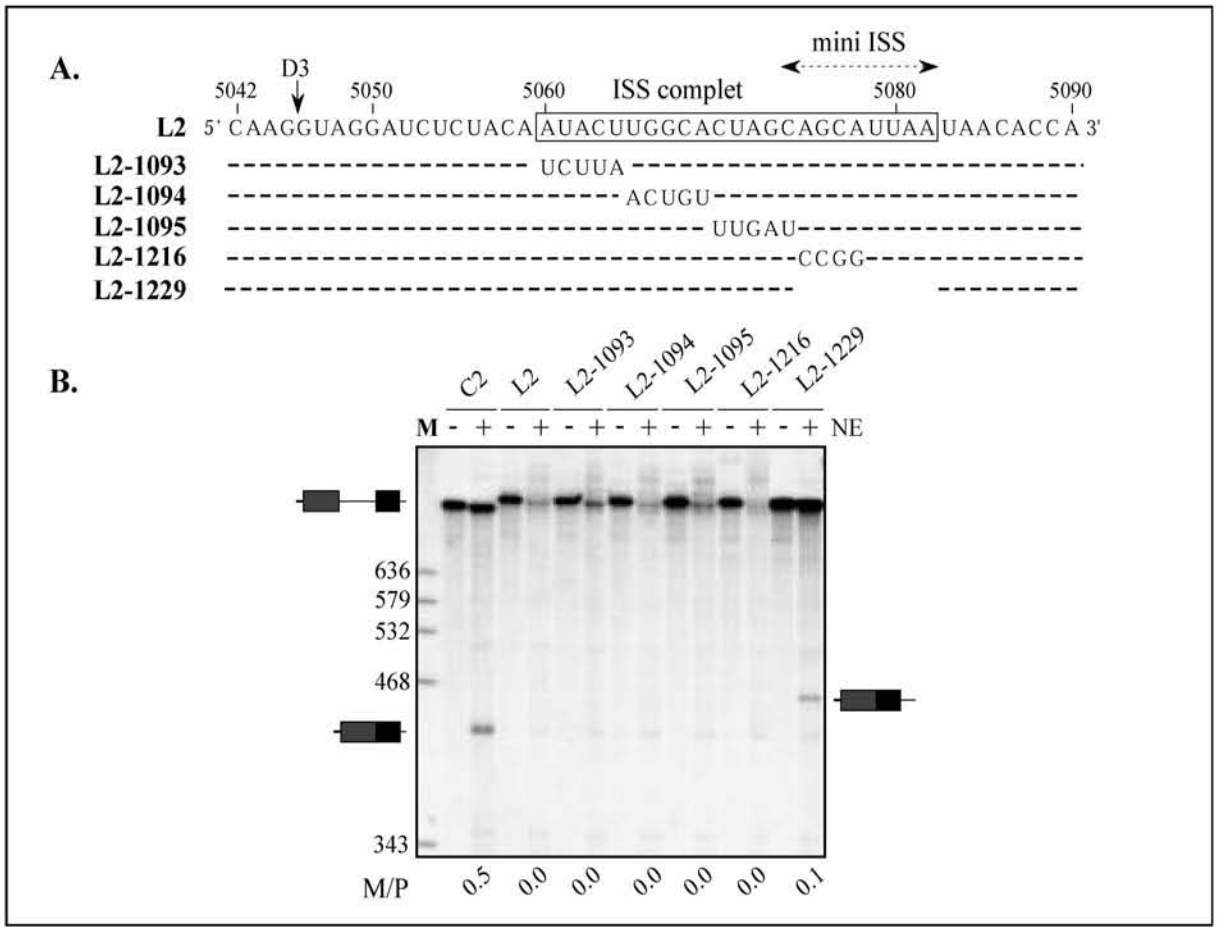


Fig. 87 : Délimitation de l'élément ISS minimal et de la séquence adjacente renforçant son activité *(Thèse de S Jacquenet, 2001)*

A) Séquence située en aval du site D3 dans le pré-ARNm L2 et ses différents variants L2-1093, L2-1094, L2-1095, L2-1216 et L2-1229. B) Electrophorèse en gel polyacrylamide des produits d'épissage in vitro des transcrits C2, L2 et L2 mutés obtenus en présence d'extrait nucléaire (NE). Un marqueur de taille (M) et les ARN non épissés sont fractionnés en parallèle. Les produits d'épissage obtenus pour la construction C2 et le variant L2-1229 sont indiqués de part et d'autre de l'autoradiogramme. La séquence miniISS située entre les positions 5074 et 5082 de l'ARN de la souche Bru est essentielle à l'effet inhibiteur, la séquence 5060 à 5073 renforce l'effet inhibiteur.

Par contre, aucune étude de l'effet de l'élément inhibiteur n'avait été réalisée *in cellulo*, car à l'époque où S. Jacquenet avait réalisé ce travail, elle ne maîtrisait pas la mutagénèse dirigée du plasmide pΔPSP, ce que j'ai implanté au laboratoire. Je disposais ainsi des outils nécessaires pour essayer de compléter ce travail par une analyse *in cellulo*, qui avait été demandée par les « reviewers » lorsque le travail de S. Jacquenet avait été soumis pour publication.

1.2. Démonstration de l'absence d'effet de l'ISS sur le taux d'ARNm vpr *in cellulo*

Les délétions de l'ISS complet et de l'élément ISS minimum au sein du plasmide pΔPSP ont tout d'abord été réalisées sur le plasmide intermédiaire pG-D1D3 contenant le fragment *Bss*HII/*Eco*RI de pΔPSP (positions 711 à 5743 de l'ARN de la souche pNL4-3), puis le fragment muté a été réinséré dans le plasmide pΔPSP dans lequel la séquence correspondante avait été éliminée. Les différentes étapes de la production de ces plasmides mutants $p\Delta PSP_{5288}$ (délétion complète de l'ISS) et $p\Delta PSP_{5290}$ (délétion de l'ISS minimum) sont décrites dans le chapitre Matériel et méthodes.

Ces plasmides ont été utilisés pour transfecter des cellules HeLa et les profils d'épissage des transcrits produits ont été analysés par RT-PCR comme décrit au chapitre 5. L'ISS étant supposé avoir un effet inhibiteur sur l'utilisation du site accepteur A2, le résultat attendu des délétions était une augmentation des transcrits épissés au niveau de A2, et plus particulièrement de l'ARNm vpr, ceci au dépend des autres transcrits. Contrairement à cette attente, aucune variation significative du taux d'ARNm vpr n'a été observée, quelle que soit la délétion.

Nous avons alors testé si un effet pouvait avoir lieu en surexprimant des protéines SR, ceci en co-transfectant le plasmide pΔPSP avec un des plasmides permettant la surexpression d'ASF/SF2, SC35, SRp40 ou de 9G8. Là encore, aucune augmentation significative du taux d'ARNm vpr n'a été observée en dehors des variations classiquement observées lors de la surexpression des protéines SR ASF/SF2, SC35, SRp40 ou 9G8 et ceci dans les trois séries distinctes d'expériences que nous avons réalisées (Fig. 88).

Nous n'avons donc pas observé de différence marquante entre l'épissage *in cellulo* des ARN contenant ou ne contenant pas l'élément ISS et ceci avec ou sans protéine SR. Seule une bande supplémentaire était détectée au dessus de la bande correspondant à rev2 lorsque l'ISS était délété et qu'ASF/SF2 était surexprimée. Néanmoins, cette différence n'a pas été observée de manière reproductible. Nous en avons donc conclu que l'élément ISS n'est pas responsable du faible taux de production des ARNm vpr, puisque ce taux ne varie pas après sa délétion.

Des différences entre résultats obtenus par épissage *in vitro*, tels que ceux obtenus par S. Jacquenet et des résultats obtenus *in cellulo* comme ceux que j'ai obtenus, peuvent s'expliquer par le fait qu'*in vitro* dans un extrait nucléaire, certains facteurs peuvent être limitants et que l'épissage est découplé de la transcription contrairement à ce qui se passe *in* .

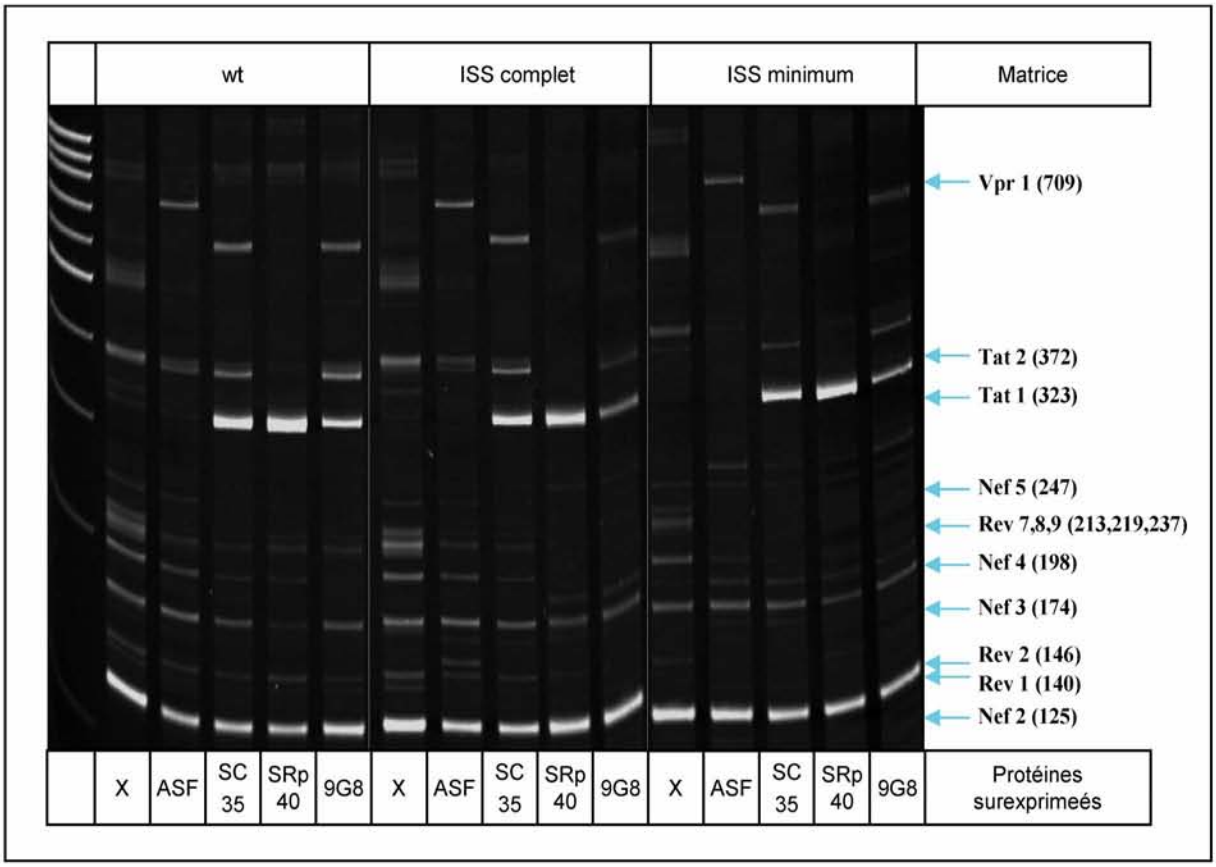


Fig. 88 : Test de l'effet de la délétion partielle ou totale de l'élément ISS sur l'épissage de l'ARN transcrit à partir du plasmide pΔPSP dans des cellules HeLa avec ou sans surexpression transitoire de protéines SR

Analyse par RT-PCR des produits d'épissage des ARN produits à partir des plasmides $p\Delta PSP_{wt}$ ou $p\Delta PSP_{5288}$ et $p\Delta PSP_{5290}$ portant la délétion totale ou partielle de l'élément ISS. Les cellules HeLa ont été co-transfectées avec les différents plasmides pΔPSP et l'un des plasmides pXJ-ASF, pXJ-SC35, pXJ-SRp40 ou pXJ-9G8 permettant respectivement la surexpression des protéines SR ASF/SF2, SC35, SRp40 ou 9G8, ou du plasmide PXJ témoin n'exprimant aucune protéine (X). Chaque analyse RT-PCR a été réalisée avec la même quantité d'ARN total extrait des cellules comme décrit dans le chapitre Matériels et méthodes. L'identité des ADNc fractionnés correspondant aux différents ARNm connus de HIV-1 est indiquée à la droite de la figure.

vivo. En tous cas, les données obtenues démontraient clairement que le faible taux de production des ARNm vpr n'est pas lié à la présence de l'élément ISS. Nous en avons conclu, que s'il existait bien un élément inhibiteur de l'utilisation du site A2, cet élément inhibiteur devait correspondre à une autre séquence que la séquence ISS.

2. Une faible efficacité d'épissage *in vitro* de l'intron D3-A3

Une autre hypothèse pour expliquer le faible taux de production des ARNm vpr pouvait être la forte efficacité du site D3 et donc une élimination fréquente des introns délimités par le site D3 et les sites accepteurs A3 à A5 (Fig. 89).

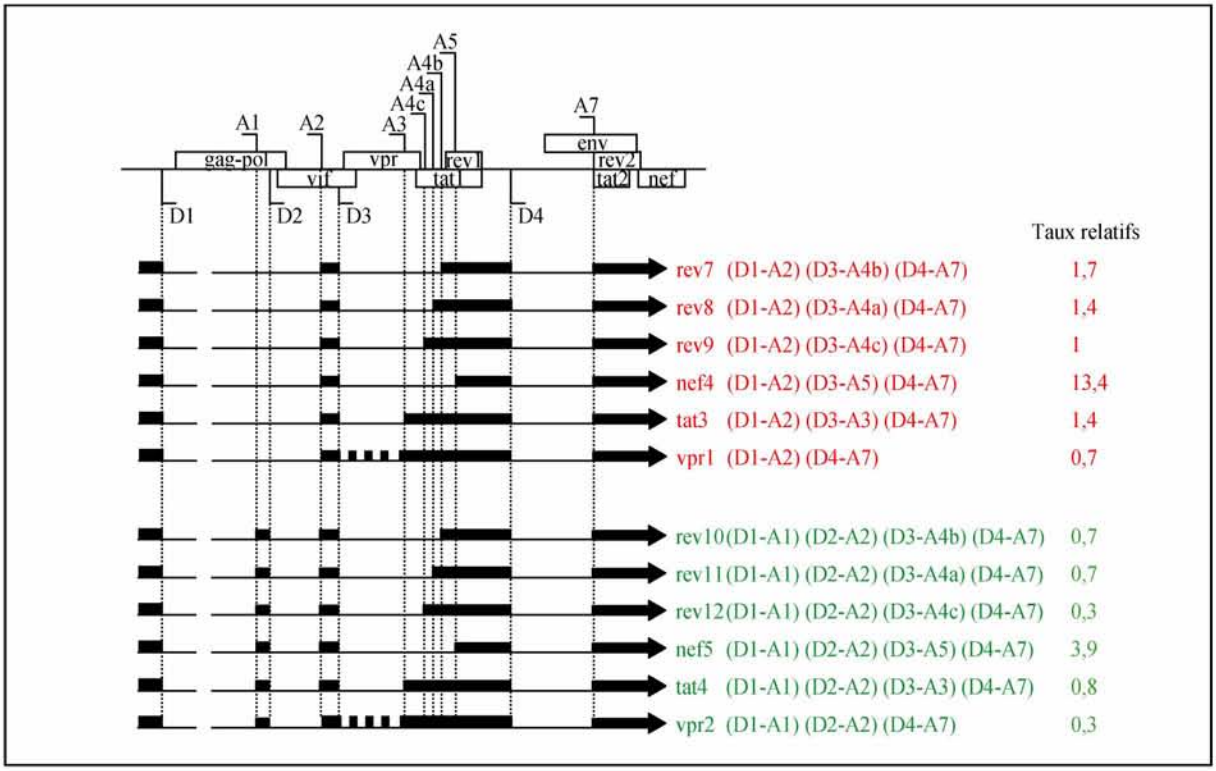


Fig. 89 : Liste des différents ARNm épissés au site accepteur A2

Les ARNm sont classés selon qu'ils contiennent (en vert) ou ne contiennent pas (en rouge) l'exon A1-D2. La séquence additionnelle qui différencie les ARNm vpr des autres ARNm épissés au site A2 est représentée par un trait discontinu. Les taux relatifs de chaque ARNm dans les lymphocytes T transfectés par un cDNA entier du virus HIV-1 sont indiqués (Purcell and Martin, 1993)

Deux constructions produites par L. Damier au laboratoire, permettaient de tester les capacités d'épissage de l'intron D3-A3. Il s'agit des plasmides 102 et 106 contenant l'exon A2-D3 couplé aux sites accepteurs A3, A4c, a, b et A5 pour 102, et A3 seul pour 106. Une autre construction, numérotée 112 était aussi intéressante puisque le site D3 est couplé à un site accepteur de l'adénovirus (Fig. 90A).

L. Damier avait pu constaté que les ARN produits à partir de ces 3 constructions n'étaient pas épissés *in vitro* (L. Damier, thèse). Néanmoins, nous venons de voir que même s'il ne modifiait pas le taux d'épissage *in cellulo*, il avait été démontré que l'élément ISS avait un

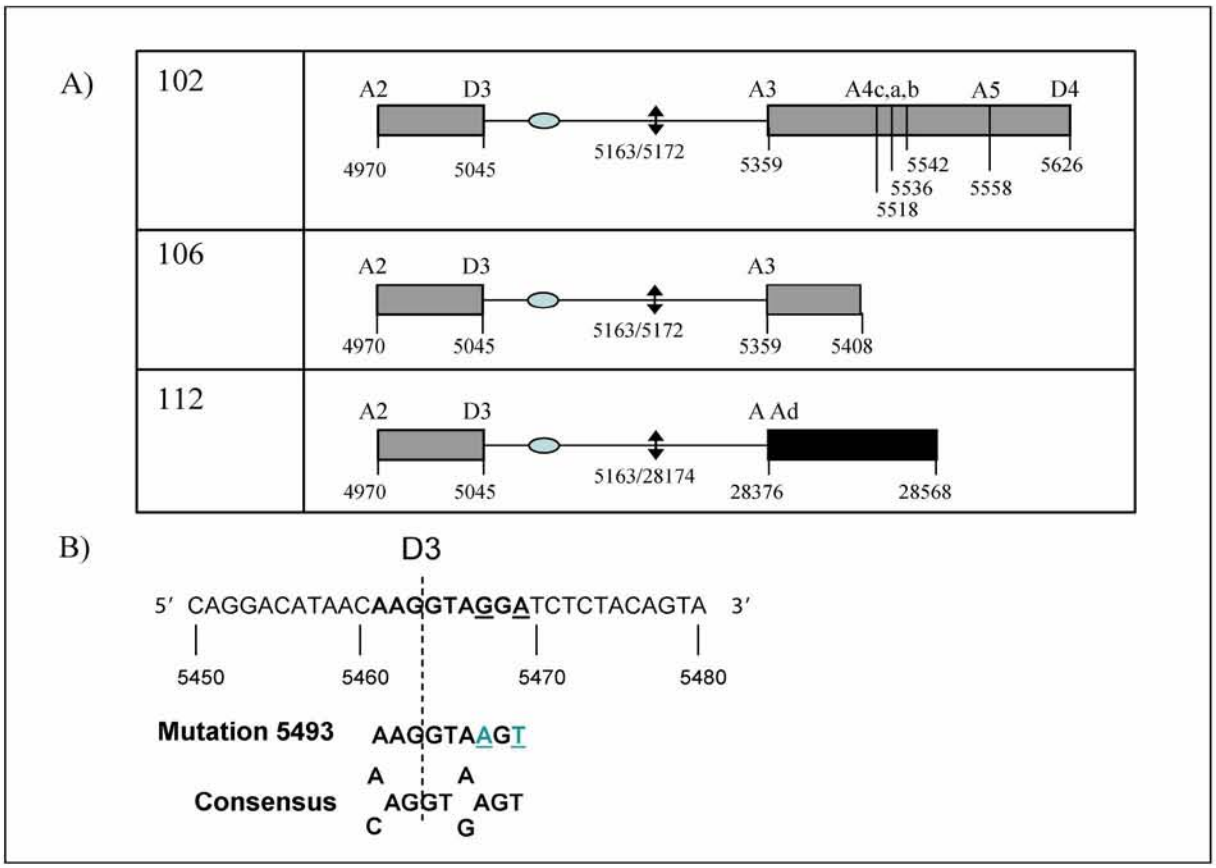


Fig. 90 : Schéma des constructions utilisées pour étudier l'épissage in vitro entre les sites D3 et A3

A) Représentation schématique des ARN produits à partir des constructions produites antérieurement par L. Damier. Les exons sont représentés par des rectangles, l'intron par un trait fin. Les chiffres indiqués correspondent aux positions au sein de l'ARN de la souche Bru. La double flèche verticale représente les délétions effectuées dans l'intron. L'élément ISS est schématisé par l'ovale bleu. B) Mutation 5493 effectuée sur les constructions 102, 106 et 112. Les nucléotides substitués dans les mutants 5493 sont soulignés. Les mutations de G en A et A en T permettent d'obtenir un site donneur consensus.

effet inhibiteur de l'épissage *in vitro* dans le cas d'ARNm contenant les sites D1 et A2 (Fig. 86). Afin de s'assurer que cet effet de l'ISS ne se retrouve pas sur les constructions 102 et 106, l'élément ISS a été délété par mutagenèse dirigée. Avec ces délétions, nous n'avons pu détecter aucune amélioration du taux d'épissage *in vitro* des ARN 102 et 106 et au final, toutes ces données étaient en faveur d'une faible efficacité du couple de site D3-A3 *in vitro*.

La faible efficacité d'épissage des ARN 102, 106 et 112 pouvait aussi être liée au fait que la séquence consensus du site D3 diffère de deux nucléotides par rapport à celle de la séquence consensus des sites 5' d'épissage. Pour tester cette hypothèse, j'ai réalisé une mutagénèse

dirigée du site D3 permettant de ramener sa séquence de fixation du snRNA U1, à la séquence consensus établie pour les sites 5' d'épissage, ceci dans les 3 constructions (Fig. 90B). L'épissage *in vitro* des ARN produit a montré que seul l'ARN 112 est épissé *in vitro*, même lorsque le site D3 est optimisé. Une possibilité était que la fixation d'ASF/SF2 sur l'exon A2-D3 que nous avons décrite ci-dessus, puisse à la fois activer le site A2 et par encombrement stérique limité l'utilisation du site D3, ce qui aurait pu expliquer pourquoi la production de l'ARNm vpr1 est très fortement activée lorsque ASF/SF2 est surexprimée. Dans ce cas, on pouvait espérer qu'en réalisant les épissages *in vitro* dans un extrait cytoplasmique qui contient peu de protéines SR et donc de la protéine ASF/SF2 ou dans un extrait cytoplasmique auquel une faible quantité d'extrait nucléaire était ajoutée (ratio de 1/7, respectivement), il serait possible d'observer un épissage des ARN 102, 103 et 112. En fait seul l'ARN 112 a pu être épissé *in vitro* et ceci en présence d'extrait nucléaire (Fig. 91).

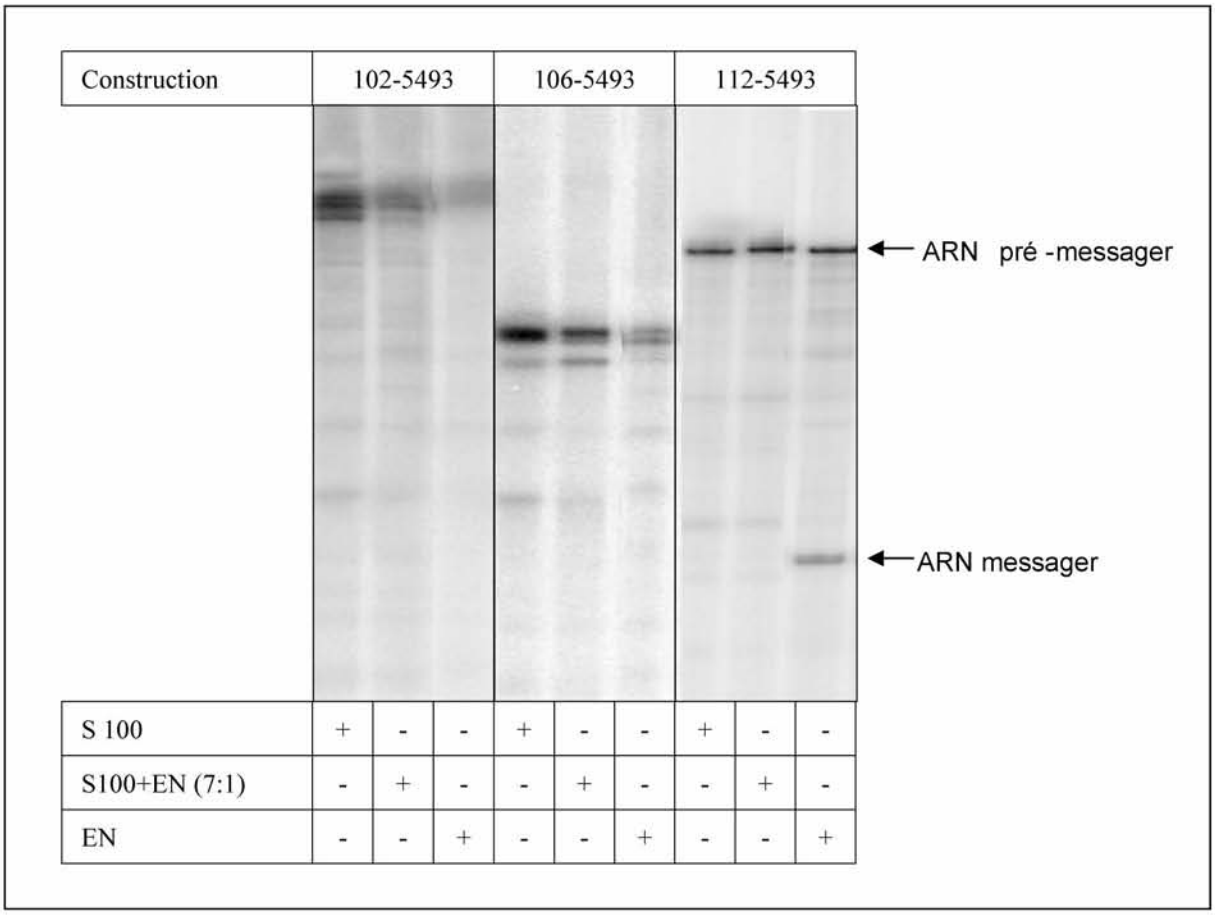


Fig. 91 : Effet de la mutation 5493 sur l'épissage in vitro des ARN 102,106 et 112.

Electrophorèse en gel polyacrylamide des produits d'épissage in vitro des transcrits 102-5493, 106-5493 et 112-5493 obtenus en présence d'extrait cytoplasmique S100, d'extrait nucléaire (EN), ou d'un mélange des deux (S100+EN 7:1).

Ainsi, contrairement à la majorité des couples de sites d'épissage de l'ARN du VIH-1, le couple de sites D3-A3 n'est pas fonctionnel en épissage *in vitro,* ceci même après élimination de l'élément ISS et optimisation de la séquence du site D3. Ces données renforcent l'idée d'un comportement très particulier de la séquence de l'intron D3-A3. Son élément ISS bloque l'épissage *in vitro* de l'intron D1-A2 situé en amont et le reste de la séquence ne permet pas l'obtention d'un épissage entre les sites D3 et A3. L'obtention d'un épissage pour l'ARN 112 indique soit que le site A2 de l'adénovirus est plus efficace que le site A3 de HIV-1, soit qu'un inhibiteur de l'élimination de l'intron D3-A3 est contenu dans la partie de cet intron éliminé dans l'ARN 112. Il est intéressant de noter, que d'une manière générale, le taux des ARNm viraux épissés entre le site D3 et l'un des sites A3 à A5 est faible par rapport à ceux directement épissés entre le site D1 et ces mêmes sites accepteurs, ceci à l'exception des ARNm nef4 et nef 5 qui sont issus d'un épissage entre D3 et A5. L'hypothèse d'un inhibiteur de l'élimination de D3-A3 ou D3-A4a, b ou c situé dans la séquence éliminée dans l'ARN 112 est donc envisageable.

3. Comment expliquer le faible taux de production des ARNm vpr par rapport aux autres ARNm viraux et sa forte augmentation en présence d'un excès d'ASF/SF2 ?

Les données décrites ci-dessus démontrent clairement que le faible taux de production des ARNm vpr *in cellulo,* n'est pas lié comme cela avait été proposé précédemment à la présence de l'élément ISS. La forte augmentation du taux d'ARNm vpr1 en présence d'ASF/SF2 peut être due au fort taux d'activation du site A2 par ASF/SF2. En effet comme nous l'avons vu, l'élément inhibiteur ESSV situé au proche voisinage de ce site dans l'exon A2-D3 fixe plusieurs molécules d'hnRNP A1, mais ASF/SF2 est à la fois capable de déplacer hnRNP A1 et d'aider U2AF35 à se fixer, comme SC35 et SRp40 le font au site A3. Néanmoins, il faut noter que le taux d'augmentation de l'ARNm vpr1 en présence d'un excès d'ASF/SF2 est plus élevé que celui des autres ARNm épissés au site A2, tels que : vpr2, tat3, tat4, rev7, nef4 et nef5 (Fig. 89).

Une explication possible pour cette différence peut être que ces ARNm sont produits par de multiples étapes d'épissage, contrairement à vpr1 qui n'est épissé qu'entre D1 et A2 et entre D4 et A7 (Fig. 89). On peut aussi imaginer des actions possibles d'ASF/SF2 autres que l'activation du site A2. Une des nombreuses explications possibles peut être la stabilisation des ARNm vpr par fixation d'ASF/SF2. Ainsi, comme nous l'avons mentionné dans l'article ci-dessus, les ARNm env sont fortement stabilisés en présence d'ASF/SF2. Ceci est du au fait que

la fixation de cette protéine sur une séquence en amont du site D4 stabilise la fixation de la snRNP U1 au site D4 (Caputi *et al.*, 2004). Comme indiqué dans l'article, l'ARNm env n'est pas épissé au site D4, mais l'invalidation de ce site entraîne une instabilité des ARNm env (Kammler *et al.*, 2006). On pourrait donc imaginer une situation semblable au site A2, où la fixation d'ASF/SF2 sur l'exon A2-D3, stabiliserait la fixation de la snRNP U1 au site D3, sans pour autant qu'il y ait assemblage de spliceosomes, ce qui pourrait stabiliser les ARNm vpr.

Comme nous l'avons déjà proposé ci-dessus on pourrait envisager qu'au contraire la fixation d'ASF/SF2 en amont du site D3 puisse par encombrement stérique bloquer la fixation de la snRNP U1 au site D3. Dans les expériences de protection par les RNases dont nous avons représenté les résultats dans l'article ci-dessus, le site D3 n'était pas masqué par ASF/SF2. Par ailleurs, l'utilisation d'extrait S100 contenant moins de protéine ASF/SF2 n'a pas favorisé l'utilisation du site D3 *in vitro*.

Il faut aussi noter, qu'ASF/SF2 a sans doute d'autres rôles *in cellulo* dans le jeu complexe de compétition qui a lieu entre les sites d'épissage A1 à A5. Ainsi, on s'attend à ce qu'elle bloque l'utilisation du site A4 par occlusion de sa boîte de branchement. En effet, comme nous le verrons plus loin, ASF/SF2 se lie à une boucle de la structure tige-boucle SLS3-A3 contenant cette boîte de branchement et elle masque complètement cette séquence. Ce type de blocage peut aussi exister à d'autres sites.

L'application de l'approche protéomique, que nous avons utilisé pour étudier les éléments régulateurs du site A3 et que nous décrirons plus loin pourra peut être apporter des informations précieuses sur les comportements particuliers des sites A2 et D3 et de l'intron D3-A3.

.

CHAPITRE 8 : Mise en évidence de régulations du site A3 indépendantes de l'élément ESS2

1. Etat des connaissances lorsque j'ai entrepris ce travail

En plus de l'action des éléments ESS2p et ESS2 qui régulent négativement l'utilisation du site A3 et du fait que les protéines SR SC35 et SRp40 contrecarrent l'effet de l'élément ESS2, D. Ropers et L. Ayadi (Ropers *et al.*, 2004) avaient montré par des expériences d'épissage *in vitro* que la protéine ASF/SF2 peut activer l'utilisation du site A3 dans l'ARN C3 (Fig. 92). Dans cet ARN, les sites D1 et A3 sont directement couplés et les sites A4c, a, b et A5 ainsi que l'élément ESS2 sont absents. Il est donc plus facile d'étudier les régulations n'impliquant pas l'élément ESS2 avec cette construction.

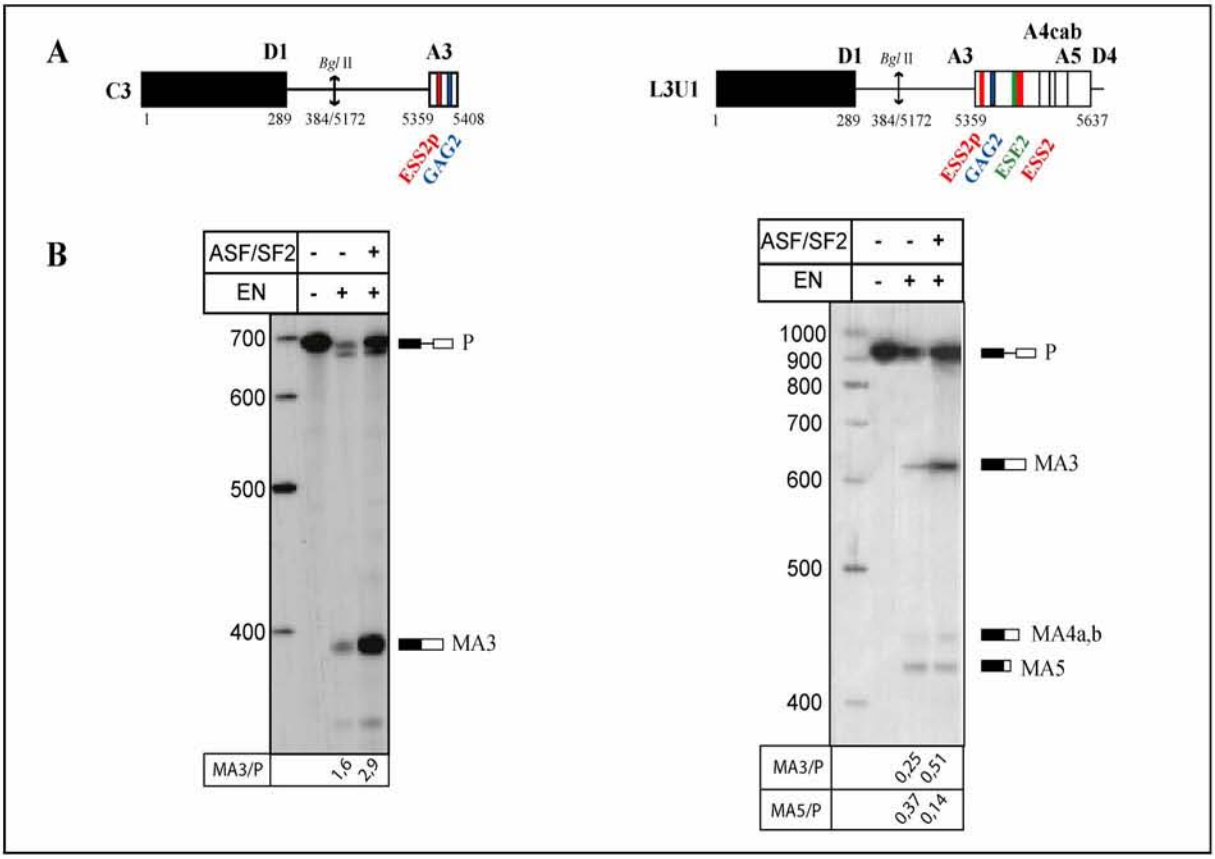


Fig. 92 : Effet de la protéine SR ASF/SF2 sur l'efficacité d'épissage in vitro au site A3 dans les ARN C3 et L3U1

A) Représentation schématique des constructions C3 et L3U1. Les éléments régulateurs sont représentés en couleur et les sites d'épissages A4c, a, b et A5 par des barres noires. B) Effet de la protéine ASF/SF2 sur l'épissage in vitro des ARN C3 et L3U1. Les ARN ont été incubés en

présence d'extrait nucléaire de cellules HeLa (EN) seul ou avec 300ng de protéines ASF/SF2 dans les conditions décrites dans le chapitre Matériel et méthodes. Les produits d'épissage ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 5%. La position des pre-ARNm est indiquée à droite de l'autoradiographie. Les valeurs de M/P (rapport de la quantité d'ARNm M sur celle de pré-messager P) estimées par mesure de la radioactivité des bandes au Phosphorimager sont indiquées en dessous des pistes.

Dans le cas de la construction L3U1, qui contient l'ESS2 et les sites en compétition avec A3, l'activation de l'utilisation du site A3 par ASF/SF2 est plus faible mais non nulle (Fig. 92 ; Ropers *et al.*, 2004). Par contre, comme nous l'avons vu déjà plusieurs fois dans la description de nos résultats, cet effet ne se répercute pas sur l'augmentation du taux d'ARNm tat1, vraisemblablement du fait de la forte compétition avec le site A2 qui est lui très fortement activé en présence d'un excès d'ASF/SF2 (Ropers *et al.*, 2004 et résultats ci-dessus).

D. Ropers avait réalisé des expériences d'empreinte de la protéine ASF/SF2 recombinante sur l'ARN SLS 2/3 contenant les deux structures tige-boucle, ceci en soumettant les complexes formés à des hydrolyses enzymatiques et des modifications chimiques. Les données obtenues montraient qu'ASF/SF2 protège fortement à la fois le brin 5' et le brin 3' de la boucle interne II (Fig. 93B). Par contre, ASF/SF2 ne protégeait pas l'élément ESS2, ce qui explique que cette protéine ne soit pas, contrairement à SC35 et SRp40, capable de contrecarrer l'élément inhibiteur ESS2.

Des expériences de retard sur gel réalisées par L. Ayadi et H. Hallay avec l'ARN SLS 2/3 sauvage ou portant une délétion du brin 5' de la boucle interne II, ont ensuite permis de montrer que l'élimination du brin 5' de la boucle interne II diminue très fortement la fixation d'ASF/SF2 à l'ARN SLS 2/3 (Fig. 93C). La fixation d'ASF/SF2 à la tige boucle SLS 3 dépend donc fortement de la présence de la séquence GAGGAG du brin 5' de la boucle interne II. D'ailleurs lorsqu'un complexe est formé avec l'ARN sauvage, ce dernier peut dissocier le complexe formé lorsqu'on l'ajoute en concentration élevée, ce qui n'est pas le cas pour l'ARN où la séquence GAGGAG a été éliminée (Fig. 93D). Cette séquence est d'ailleurs en bon accord avec les données obtenues par l'approche SELEX indiquant que les séquences ARN ayant une forte affinité pour ASF/SF2 sont riches en purine (Tacke and Manley, 1995 ; Liu *et al.*, 1998).

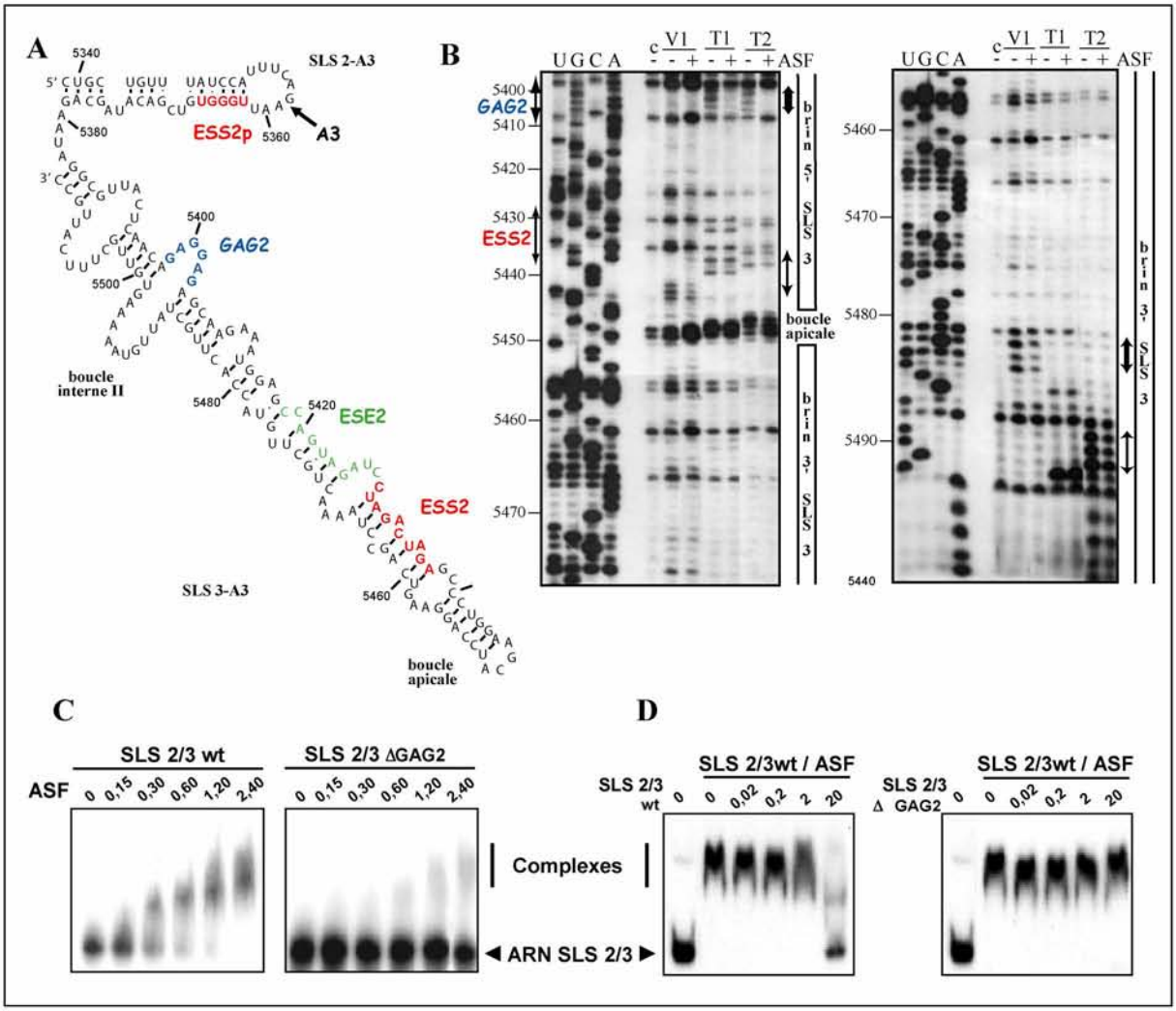


Fig. 93 : Etudes de la fixation d'ASF/SF2 sur l'ARN SLS 2/3

A) Représentation de l'ARN SLS 2/3 utilisé pour les expériences d'empreintes et de retard sur gel présentées dans les panneaux B, C et D. B) Effet de la présence de la protéine ASF/SF2 sur les coupures de l'ARN SLS2/3 par les RNases V1, T1 et T2. Après incubation en présence de 500 ng de la protéine ASF, le transcrit SLS 2/3 a été soumis à des hydrolyses ménagées par 0,8 U de RNase V1 ou 0,15 U de RNase T2 ou 0,8 U de RNase T1, puis les produits d'hydrolyse ont été analysés par extension d'amorce par la transcriptase inverse avec l'oligonucléotide 548 comme décrit dans le chapitre Matériel et méthodes. Les coupures ont été identifiées par comparaison avec le séquençage de l'ARN non traité dont les produits de réaction ont été déposés dans les 4 premières pistes. C) Analyse par expériences de retard sur gel des interactions qui peuvent être établies entre les ARN SLS 2/3$_{wt}$ et SLS 2/3$_{\Delta GAG2}$ et la protéine ASF/SF2. Les ARN ont été marqué en 5' puis incubés en présence de concentrations croissantes en protéine ASF (0 à 2,4 µM). Les complexes formés ont été fractionnés par électrophorèse en gel de polyacrylamide. L'ARN libre et les complexes formés sont indiqués sur les côtés des autoradiogrammes. D) Analyse par expériences de retard sur gel de la déstabilisation du complexe formé entre l'ARN SLS 3$_{wt}$ radiomarqué et la protéine ASF, par ajout de quantités croissantes d'ARN froid SLS 3$_{wt}$ et SLS 3$_{\Delta GAG2}$ (0 à 20pmol).

Afin de tester le rôle fonctionnel de cette interaction, D. Ropers et L. Ayadi avaient alors réalisé des expériences d'épissage *in vitro* en extrait nucléaire de cellules HeLa en utilisant l'ARN de la construction C3 sauvage ou délétée de la séquence GAGGAG (Fig. 94A). La délétion GAGGAG entraîne une perte de l'activation de l'épissage de l'intron D1-A3 par ASF/SF2 (Fig. 94B). Afin de valider la présence d'un élément activateur d'épissage, D. Ropers et L. Ayadi ont alors placé la séquence située entre les positions 5396 à 5415 de l'ARN de la souche Bru (nommé ESEt1) dans un contexte hétérologue, à savoir le plasmide Sp1-ex2inv (Fig. 94C). Cette construction est dérivée du plasmide Sp1 qui permet de produire un mini ARN pré-messager E1A de l'adénovirus fortement épissé en présence d'extrait nucléaire (Schmitt *et al.*, 1987). Lorsque l'exon 2 est inversé (cas de Sp1-ex2inv), le transcrit est très faiblement épissé et pour augmenter l'efficacité d'épissage, il est nécessaire d'inclure une séquence activatrice dans l'exon 2 inversé (Cavaloc *et al.*, 1999). Ce test des propriétés activatrices de l'épissage d'une séquence avait en particulier été réalisé par insertion de la séquence consensus (ASF-A10) de fixation d'ASF/SF2 obtenue après SELEX (Cavaloc *et al.*, 1999). Ainsi il a pu être démontré que l'insertion de la séquence ESEt1 contenant l'hexamère GAGGAG permet d'activer fortement l'épissage, alors que l'insertion de la séquence aléatoire contrôle GST-80 n'a aucun effet (Fig. 94D). L'ensemble de ces données suggérait donc que le brin 5' de la boucle interne II, correspondait à un activateur de l'épissage au site A3, alors appelé ESEt pour Exon Splicing Enhancer de l'ARNm tat.

En parallèle, à cette découverte d'un nouvel activateur du site A3 réalisés au laboratoire, l'équipe de S. Litvak (REGER, Bordeaux) avait mis en évidence deux substitutions de bases dans la région régulant l'utilisation du site A3, ceci dans des ARN viraux extraits de lymphocytes T infectés par le virus HIV-1 (souche Bru) mais résistantes à cette infection après plusieurs passage du virus. L'une de ces deux substitutions de bases (G en A) se situait en position 5363 de l'ARN viral dans l'élément ESS2p. L'autre (C en U) se trouvait en position 5396 juxtaposée à la séquence GAGGAG (Bourara *et al.*, 2000). Au total les auteurs n'avaient trouvé que 5 substitutions de bases dans l'ensemble des régions de l'ARN viral qu'ils avaient séquencés, dont ces deux dans la région régulant le site A3.
Ces auteurs proposaient que ces substitutions de bases soient générées directement par édition de l'ARN. Depuis, d'autres articles ont décrit une activité de déamination convertissant les résidus C de l'ADN viral simple brin en résidus U au cours de la rétro-transcription. Cette activité désaminase est portée par la protéine cellulaire APOBEC3G (Harris *et al.*, 2003 ;

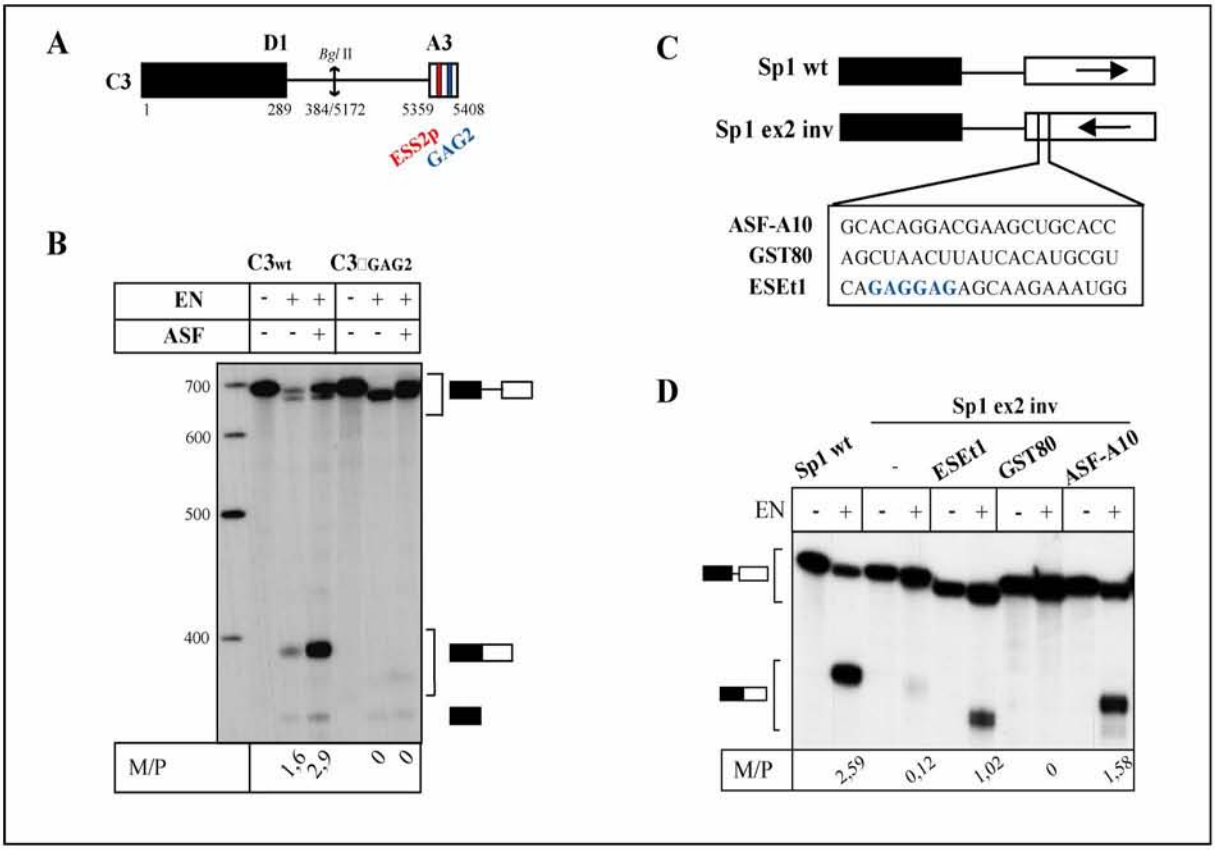


Fig. 94 : Effet de la délétion de la séquence GAGGAG sur l'efficacité d'épissage in vitro de l'ARN C3 et test des propriétés activatrices de cette séquence dans un contexte hétérologue

A) Représentation schématique de la construction C3. L'ARN C3 est composé des séquences 1-384 et 5172-5408 de la souche Bru de HIV-1, il contient les sites d'épissage D1 et A3, ainsi que l'élément ESS2p (en rouge) et la séquence GAGGAG (en bleu). B) Effet de la délétion de la séquence GAGGAG sur l'activation de l'épissage des ARN C3 par la protéine ASF/SF2. Les ARN ont été incubés en présence d'extrait nucléaire de cellules HeLa (EN) seul ou avec 300ng de protéines ASF/SF2 dans les conditions décrites dans le chapitre Matériel et méthodes. Les produits d'épissage ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 5%. La position des différents ARN est indiquée à droite de l'autoradiographie. Les valeurs de M/P (rapport de la quantité d'ARNm M sur celle de pré-messager P) estimées par mesure de la radioactivité des bandes au Phosphorimager sont indiquées en dessous des pistes. C) Représentation schématique des constructions Sp1wt et Sp1-ex2inv. Les séquences ASF-A10, GST80 et ESEt1 contenant la séquence GAGGAG (en bleu), ont été insérées en aval du site accepteur pour tester leur effet activateur. D) Effet de la séquence ESEt1 sur l'épissage de la construction Sp1-ex2inv. Les ARN issus de la transcription des constructions ont été incubés en absence ou en présence d'extrait nucléaire (EN) dans les conditions décrites dans le chapitre Matériel et méthodes. La séquence GST80 sert de témoin négatif et la séquence ASF-A10 sert de témoin positif.

Mangeat *et al.*, 2003). Il est vraisemblable que les substitutions G en A, détectées par l'équipe de S. Litvak étaient liées à cette activité désaminase vis-à-vis de l'ADNc viral et pas à une activité d'édition directe de l'ARN. La question d'une édition de l'ARN reste ouverte dans le cas de la substitution $C_{5396}U$. Cette substitution peut être due à une substitution G en A lors de la synthèse du second brin d'ADN, ou réellement à une édition de l'ARN. Les ARN viraux qui contenaient ces substitutions de bases ne produisaient que de très faibles quantités de virions. Une possibilité était le blocage de la multiplication du virus à l'étape d'épissage. Dans un premier temps, S. Jacquenet, H. Hallay et A. Méreau avaient montré que la substitution $G_{5363}A$ dans l'élément ESS2p augmente le pouvoir inhibiteur de cet élément en convertissant le site hnRNP H qu'il contient en un site hnRNP A1.

L. Ayadi, D. Ropers, H. Halay ont ensuite montré que la mutation $C_{5396}U$ a un effet inhibiteur encore plus fort sur le site d'épissage A3, ceci par des expériences d'épissage *in vitro*, mais aussi *in cellulo* en générant soit l'un, soit l'autre des deux substitutions de bases ou les deux à la fois. Les mutagenèses dirigées du plasmide pΔPSP avaient été réalisées en collaboration avec l'équipe de MT. Stoltzfus. Les résultats obtenus *in vitro* avec la construction L3-U1 montraient que le taux d'épissage au site A3 diminue fortement dans le cas de la mutation $C_{5396}U$ et de la double mutation, même en présence d'un excès de protéines SR ASF et SC35 (Fig. 95).

Par contre, aucune étude de l'effet des mutations n'avait été réalisée *in cellulo* en surexprimant les protéines SR et rien n'était connu sur les bases moléculaires de l'inhibition générée par la mutation $C_{5396}U$. Le mécanisme d'activation de l'utilisation du site A3 par l'élément ESEt jouxtant la substitution $C_{5396}U$ restait aussi à décrypter. J'ai essayé d'apporter des informations sur ces deux questions en utilisant différentes approches que nous allons décrire.

2. Résultats obtenus

2.1. Etude *in cellulo*

Dans un premier temps, j'ai comparé les effets de la délétion de la séquence GAGGAG (ΔGAG_2), de la mutation $C_{5396}U$ et de la double mutation (ΔGAG_2 + $C_{5396}U$ nommée Dble) sur l'épissage *in cellulo* de l'ARN produit par le plasmide pΔPSP. Pour cela, les 3 plasmides pΔPSP mutants (ΔGAG_2, $C_{5396}U$, Dble) ont été produits comme décrit dans le chapitre Matériels et méthodes. Chacune de ces constructions a ensuite été co-transfectée dans les

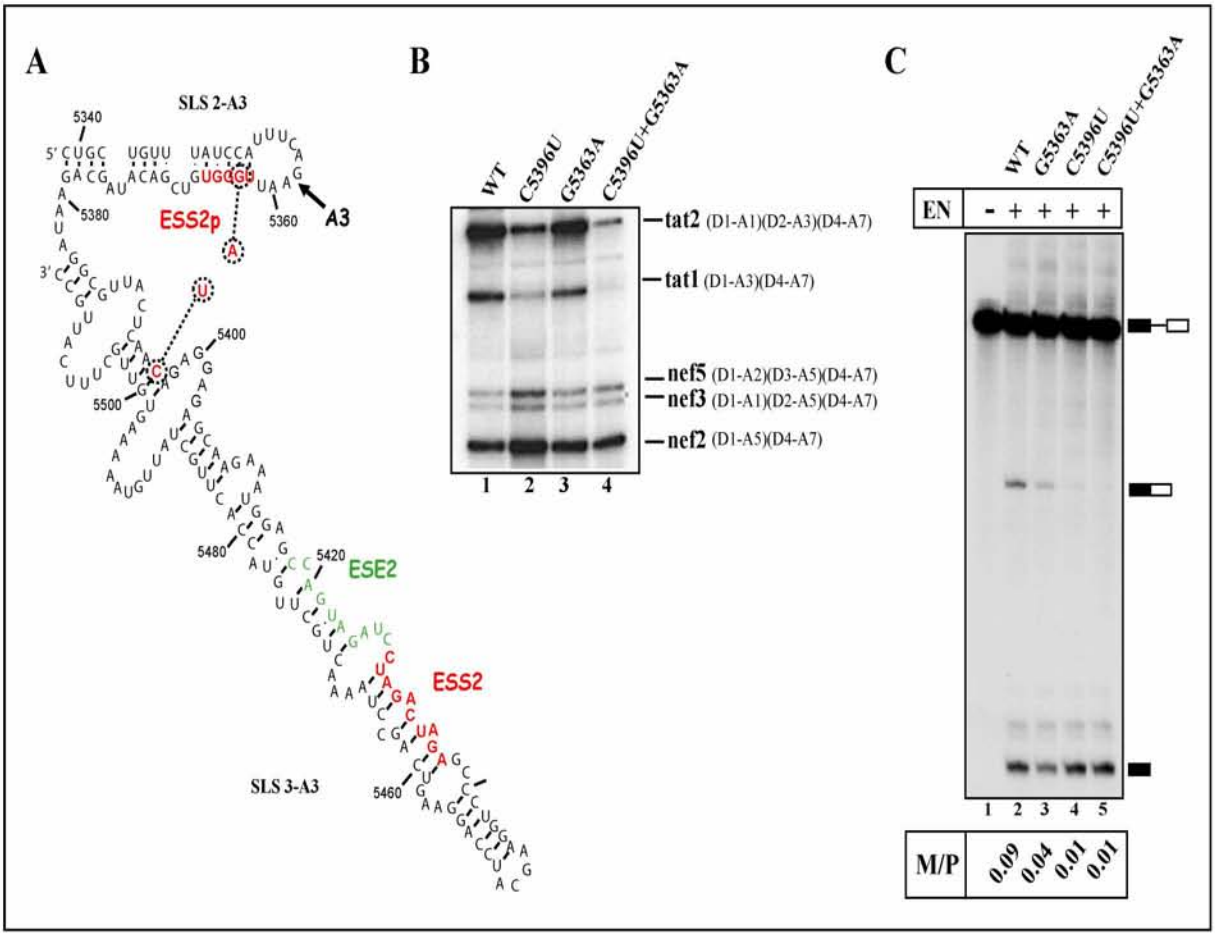


Fig. 95 : Effet des conversions G5363A et C5396U sur l'épissage au site A3 de l'ARN viral in vitro et dans des cellules HeLa.

A) Localisation des substitutions de bases G5363A et C5396U sur le modèle de structure secondaire de l'ARN autour du site A3. B) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 6% des produits d'épissage de l'ARN viral amplifiés par RT-PCR dans les conditions décrites par Jacquenet et al. (2001b). C) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage des ARN uniformément marqués. Les transcrits L3U1 wt et mutés ont été incubés pendant 2h30 dans un extrait nucléaire de cellules Hela comme décrit dans la partie Matériels et méthodes. La position des précurseurs et des produits épissés est indiquée à gauche de l'autoradiogramme.

cellules HeLa avec un vecteur pXJ vide ou exprimant l'une des protéines SR : ASF/SF2, SC35, SRp40 ou 9G8. Les produits d'épissage contenus dans les extraits cellulaires ont été caractérisés par RT-PCR comme dans les expériences décrites dans les chapitres précédents. Chaque expérience a été répétée 3 fois, ceci en utilisant des lots différents de cellules co-

transfectées. En absence de mutation, des profils de fractionnement des ADNc identiques à ceux observés dans les expériences décrites dans les chapitres précédents ont été obtenus.

Sans surexpression d'une des protéines SR, la mutation qui s'est avérée avoir le plus d'effet sur le profil d'épissage est la mutation $C_{5396}U$. En effet, comme observé précédemment en collaboration avec MT. Stoltzfus, nous notons que les ARNm tat1 et tat2 ne sont pas détectables pour les deux mutants portants cette mutation, ce qui reflète une baisse d'utilisation du site A3 (Fig. 96). Sans que l'on puisse l'expliquer en l'état actuel des connaissances, une légère augmentation du taux des ARNm tat est observée pour la délétion de la séquence GAGGAG.

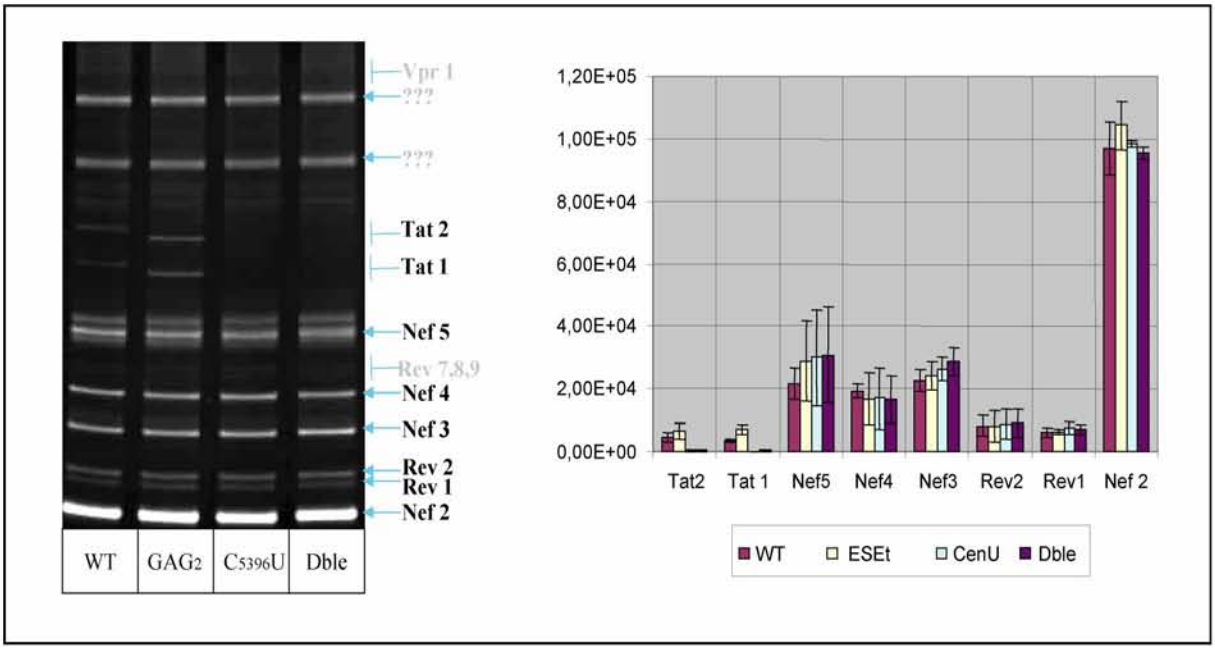


Fig. 96 : Effets des mutations ΔGAG$_2$, C$_{5396}$U et Dble sur l'épissage de l'ARN du plasmide pΔPSP dans les cellules HeLa

Analyse par RT-PCR des produits d'épissage de l'ARN des plasmides pΔPSP sauvage et mutés. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP$_{wt}$, pΔPSP$_{ΔGAG2}$, pΔPSP$_{C5396U}$ ou pΔPSP$_{Dble}$ et le plasmide pXJ-X ne surexprimant aucune protéine. Chaque analyse par RT-PCR a été réalisée avec la même quantité d'ARN total extrait des cellules HeLa et avec les oligonucléotides 2143 et 2144 comme décrit dans le chapitre Matériels et méthodes. La correspondance entre les ADNc fractionnés et les différents ARNm produits est indiquée à gauche de l'autoradiogramme. La quantité de chaque ARNm a été estimée à partir de l'intensité des bandes mesurée comme décrit dans le chapitre Matériel et méthodes. Les valeurs obtenues pour 3 expériences indépendantes, ainsi que les écart-types sont indiqués dans le panneau de droite.

Lorsque de la protéine ASF est surexprimée, les effets observés des mutations sont quasi identiques à ceux observés sans surexpression de protéine SR, à savoir une disparition des ARNm tat1 et tat2 lorsque la mutation $C_{5396}U$ est présente (Fig. 97). Comme attendu, la surexpression d'ASF augmente le taux d'ARNm vpr1, mais celui-ci reste inchangé quelque soit la mutation (Fig. 97). On note que le baisse du taux de tat1 et tat2 a surtout comme répercussion une légère augmentation de nef3 indiquant comme souvent observé précédemment une utilisation du site A5 au dépend du site A3.

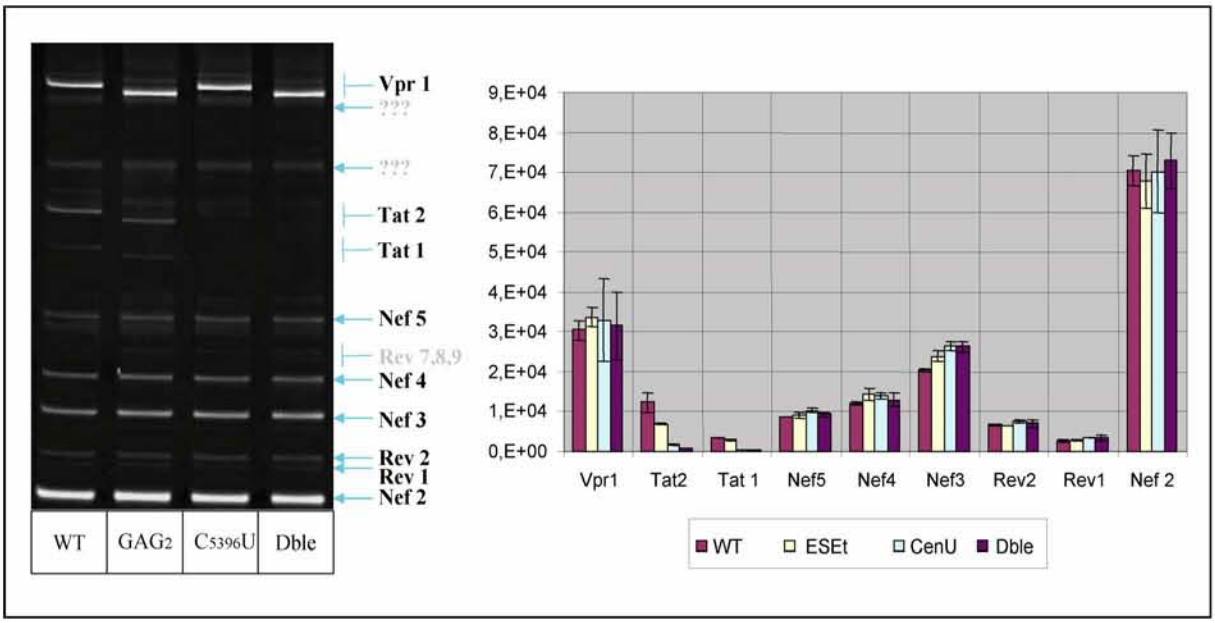


Fig. 97 : Effet des mutations ΔGAG$_2$, C$_{5396}$U et Dble sur l'épissage de l'ARN du plasmide pΔPSP dans les cellules HeLa surexprimant ASF/SF2

Analyse par RT-PCR des produits d'épissage de l'ARN du plasmide pΔPSP sauvage ou muté. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP$_{wt}$, pΔPSP$_{ΔGAG2}$, pΔPSP$_{C5396U}$ ou pΔPSP$_{Dble}$ et le plasmide pXJ-ASF. Chaque analyse par RT-PCR est réalisée avec la même quantité d'ARN total extrait des cellules et avec les oligonucléotides 2143 et 2144 comme décrit dans le chapitre Matériels et méthodes. Les ARNm viraux correspondant à chaque bande sont indiqués à gauche de l'autoradiogramme. La quantité de chaque ARNm a été estimée à partir de l'intensité des bandes mesurée comme décrit dans le chapitre Matériel et méthodes. Les résultats obtenus lors de trois expériences indépendantes sont représentés dans le panneau de droite.

Comme précédemment décrit, dans le cas de l'ARN HIV-1 sauvage, la présence d'un excès des protéines SC35 ou 9G8 entraîne une hausse de l'épissage au niveau du site A3 et donc une augmentation du taux de messagers tat1 et tat2. Lorsque la mutation $C_{5396}U$ est introduite, à nouveau on note une inhibition de l'utilisation de A3. La différence dans ces conditions vient

par contre du fait que la délétion de la séquence GAGGAG a le même effet inhibiteur que la mutation $C_{5396}U$, et que ces 2 effets sont cumulatifs (Fig. 98). Pour le double mutant en effet, le taux de messagers tat1 diminue des facteurs 2,5 et 5 respectivement en présence de SC35 et de 9G8. De façon plus étonnante, les fortes variations de tat1 n'entraînent pas de variations de nef2 (qui est pourtant en plus faible quantité du fait même que les messagers tat sont plus nombreux), alors que les messagers nef3 augmentent en conséquence (Fig. 98). De plus il apparaît dans ces 2 conditions (+SC35 ou +9G8) que le taux de messagers tat2 stagne voire augmente légèrement en présence des mutations.

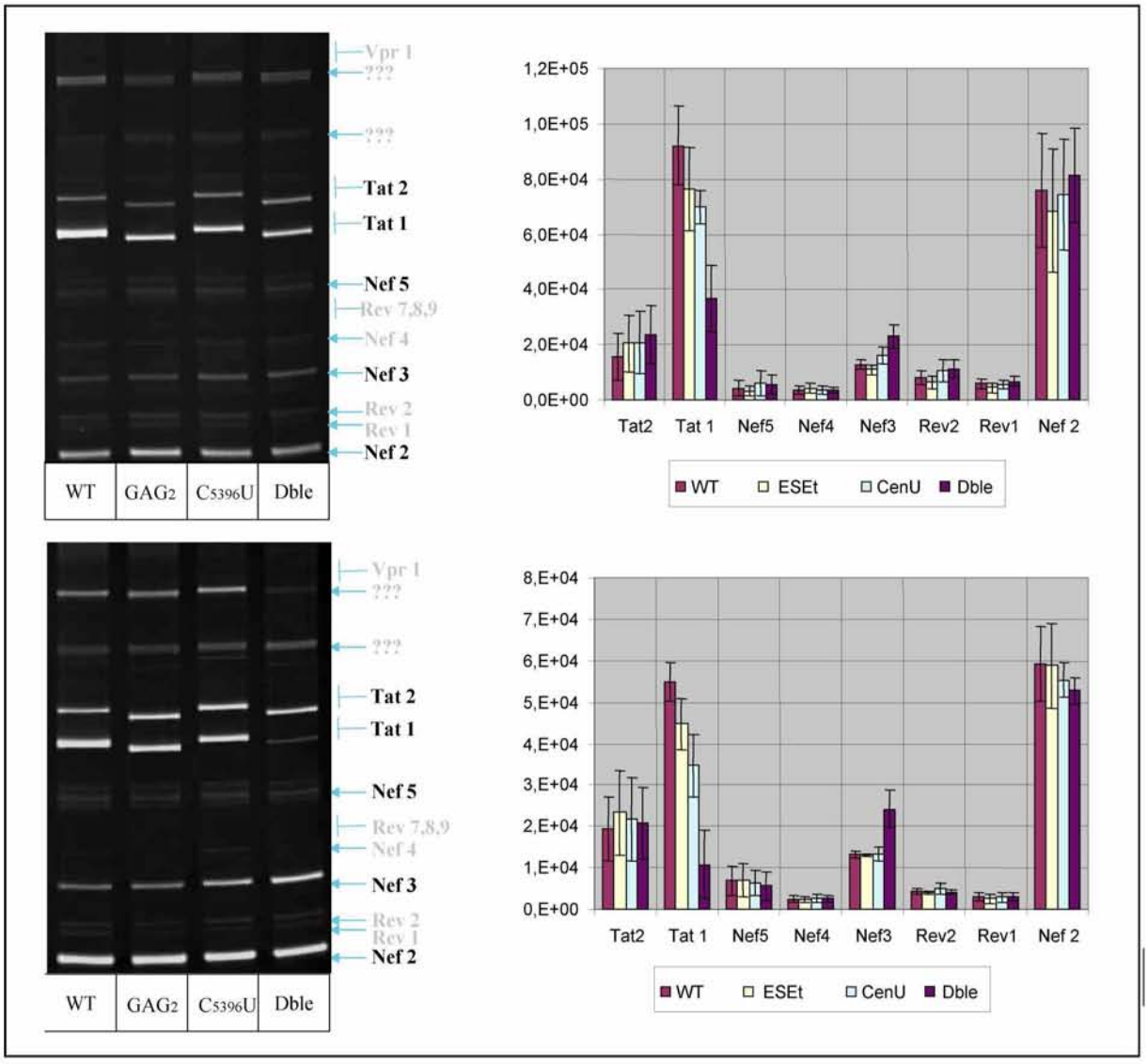


Fig. 98 : Effet des mutations ΔGAG$_2$, C$_{5396}$U et Dble sur l'épissage de l'ARN du plasmide pΔPSP dans des cellules HeLa surexprimant SC35 ou 9G8

Analyse par RT-PCR des produits d'épissage de l'ARN du plasmide pΔPSP sauvage ou muté. Les cellules HeLa ont été co-transfectées avec les plasmides pΔPSP$_{wt}$, pΔPSP$_{ΔGAG2}$, pΔPSP$_{C5396U}$ ou pΔPSP$_{Dble}$ et les plasmides pXJ-SC35 (panneau haut) ou pXJ-9G8 (panneau bas). Chaque analyse par

RT-PCR a été réalisée avec la même quantité d'ARN total extrait des cellules et avec les oligonucléotides 2143 et 2144 comme décrit dans le chapitre Matériels et méthodes. La correspondance entre les ADNc fractionnés et les ARNm de HIV-1 est indiquée à gauche de l'autoradiogramme. La quantité de chaque ARNm a été estimée à partir de l'intensité des bandes mesurée comme décrit dans le chapitre Matériels et méthodes. Les résultats obtenus pour trois expériences indépendantes sont représentés dans le panneau de gauche.

Enfin, lorsque SRp40 est surexprimée dans les cellules transfectées par les plasmides pΔPSP mutés, aucun effet détectable n'est observé sur le profil d'épissage. Pour chacun des ARN, l'épissage de l'intron D1-A3 est largement majoritaire et conduit à la production de l'ARNm tat1 (Fig. 99). Ceci confirme la constatation déjà faite dans les chapitres précédents que SRp40 a sans doute des effets plus complexes que SC35. L'ARNm tat2 est quasiment toujours absent lorsqu'elle est surexprimée et c'est elle qui a le plus fort effet activateur de l'élimination de l'intron D1-A3. En plus son effet n'est affecté par aucune des 3 mutations que nous avons réalisées dans ces expériences.

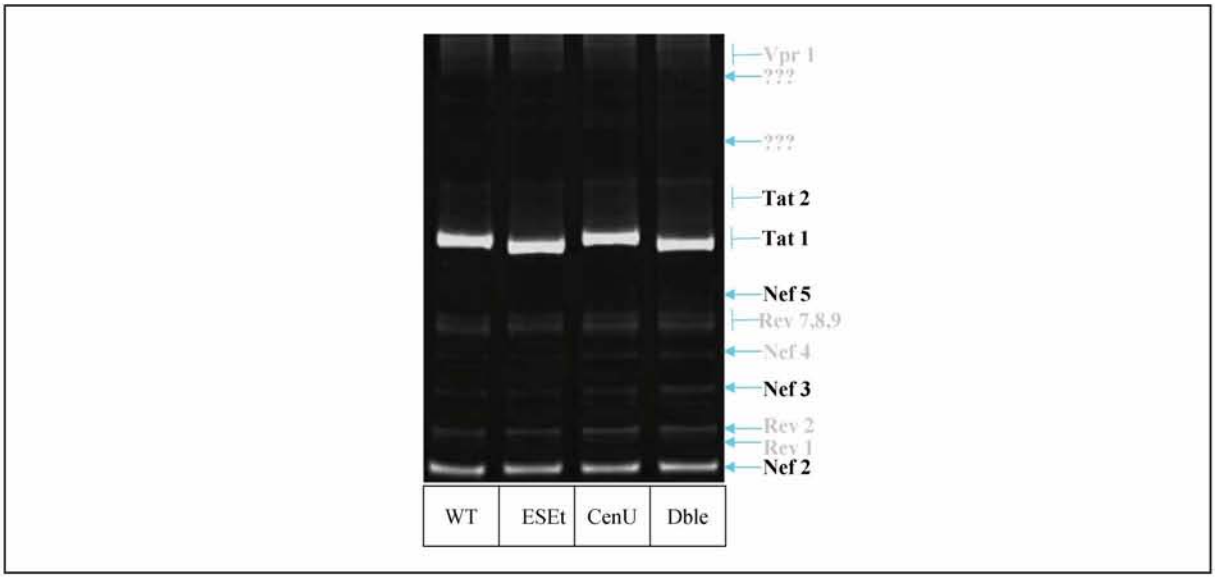


Fig. 99 : Effet des mutationsΔGAG_2, $C_{5396}U$ et Dble sur l'épissage de l'ARN du plasmide pΔPSP sauvage ou muté dans les cellules HeLa surexprimant SRp40

Analyse RT-PCR des produits de l'épissage de l'ARN de HIV-1. Les cellules HeLa ont été co-transfectées avec les plasmides $p\Delta PSP_{wt}$, $p\Delta PSP_{\Delta GAG2}$, $p\Delta PSP_{C5396U}$ et $p\Delta PSP_{Dble}$ et le plasmide pXJ-SRp40. Chaque analyse RT-PCR est réalisée avec la même quantité d'ARN total extrait et avec les oligonucléotides spécifiques de l'ARN de HIV-1 comme décrit dans le chapitre Matériels et méthodes. L'identité des ADNc fractionnés correspondant aux différents ARNm de HIV-1 est indiquée sur la gauche de la figure.

Pris dans leur globalité, une des conclusions qui peut être tirée des résultats, est que cette région de l'ARN contenant la séquence GAGGAG et le nucléotide C_{5396}, a effectivement un impact sur l'utilisation du site d'épissage A3, même si cet effet ne touche pas la production de messagers tat1 et tat2 de façon identique. En effet, nous avons vu qu'en absence de protéine SR ou en présence d'ASF, les faibles taux de messagers tat1 et tat2 diminuent tout les deux tandis qu'en présence des protéines SC35 ou 9G8, seul le taux de messagers tat1 diminue pendant que celui de tat2 reste stable. Une explication plausible de ce phénomène est que l'inhibition de A3 conduise à l'insertion de l'exon A1-D2 et qu'à nouveau le choix se porte entre un épissage de l'exon D2-A3 (conduisant au messager tat2) et un épissage entre D2 et les sites accepteurs A4c,a,b ou A5. Dans le cas d'une faible utilisation de A3 initiale (celui sans protéine SR et avec ASF), il est probable que de faibles quantités de messager tat2 soient à nouveau reproduites et donc que tat 1 et tat2 disparaissent en même temps. Dans le cas d'une forte utilisation de A3 par contre (celui avec les protéines SC35 et 9G8), il est plus probable que du messager tat2 soit reproduit et par conséquent que la diminution de tat1 compense celle de tat2.

Permettant d'expliquer les variations des messagers tat1 et tat2, cette hypothèse convient aussi au fait que les seules augmentations de messagers ne concernent que nef3. En effet, si l'on considère que l'inhibition de l'épissage entre les sites D1 et A3 conduit essentiellement non pas à l'épissage entre les sites D1 d'une part et A4c, a, b ou A5 d'autre part, mais à celui entre les couples de sites D1/A1, puis D2/A4c, a, b ou D2/A5, il apparaît comme plus évident que seuls les messagers rev4, 5, 6 ou nef3 augmentent à l'instar de rev1, 2, 3 ou nef2. Dans tous les cas, l'utilisation des sites A4c, a, b étant tellement faible, on comprend pourquoi seul le messager nef3 augmente de manière significative dans nos analyses (Fig. 100).

Finalement, sur la base des variations observées du taux de production des ARNm tat1 et tat2, la substitution $C_{5396}U$ a un effet négatif marqué sur l'utilisation du site A3 sans surexpression de protéines SR et elle diminue très fortement son activation par SC35 et 9G8. De plus, dans le cas de 9G8 l'effet est fortement renforcé lorsque la séquence GAGGAG est en plus éliminée. Comme rien n'était connu sur le mode d'activation du site A3 par 9G8, nous nous sommes demandés si la séquence GAGGAG et le résidu C_{5396} ne faisaient pas partis d'un même élément en *cis* au travers duquel 9G8 active le site A3. Nous avons entrepris une série d'études *in vitro* pour tester cette hypothèse que nous allons décrire ci-dessous.

Un autre point intéressant apporté par ces résultats était le fait que l'activation du site A3 par SRp40 se distingue de celle par SC35 car elle n'est affectée ni par la délétion de la séquence GAGGAG, ni par la substitution $C_{5396}U$.

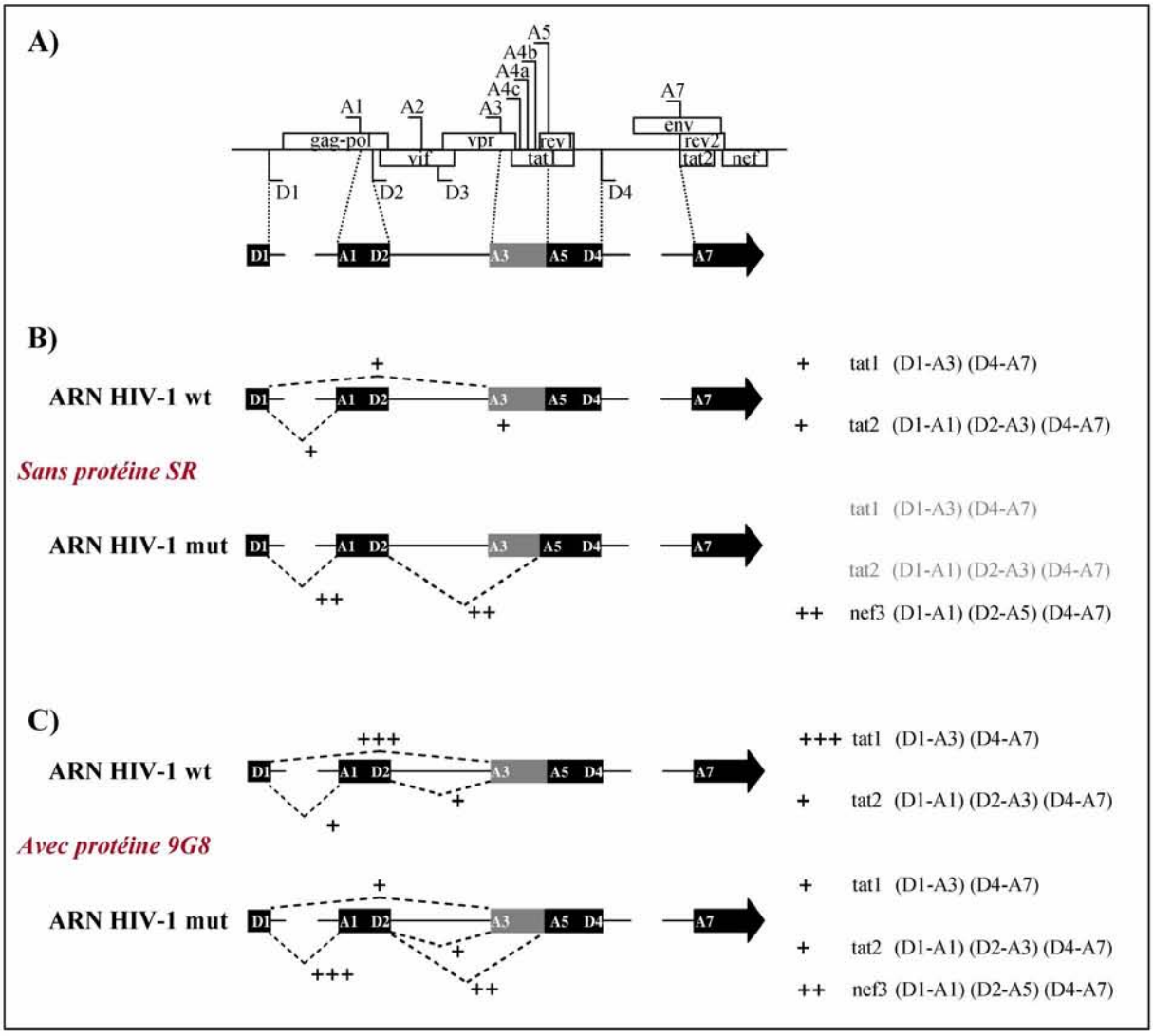


Fig. 100 : Modèle proposé permettant d'expliquer les variations d'expression des ARN messagers du virus HIV-1 en fonction des mutations et du type de protéine SR surexprimée

Autre observation importante, des mutations situées en aval du site A3 favorisent donc l'inclusion de l'exon A1-D2 situé très en amont. On peut imaginer que soit une structure à longue distance générée par des interactions ARN/ARN ou entre protéines liées à des régions distantes de l'ARN soit rompue, soit que l'équilibre des forces des sites d'épissage soit altéré, ce qui pourrait modifier l'efficacité d'inclusion de certains exons et les cinétiques relatives d'utilisation des sites d'épissage.

Il faut aussi noter que ces mutations altèrent la région codant la protéine Vpr. Or il a été montré récemment que la protéine Vpr peut agir sur l'épissage de certains messagers cellulaires. Elle bloque l'épissage d'un pré-mRNA β-globine pris comme modèle, ceci aussi bien *in vivo* qu'*in vitro* (Kuramitsu *et al.*, 2005). On sait quelle se lie aux facteurs d'épissage

SAP145 et empêche son interaction avec le facteur SAP49 (Hashizume *et al.*, 2007). Ces facteurs sont deux sous-unités du facteur d'épissage SF3b requis pour l'arrimage de la snRNP U2 à la boîte de branchement. Il n'est donc pas exclu que Vpr intervienne dans la régulation de l'épissage alternatif de l'ARN viral. Certains des effets observés des mutations pourraient donc être liés aux mutations générées dans Vpr.

2.2. Etude de l'activation par 9G8 *in vitro*

L'impact de 9G8 sur l'utilisation de A3 dans l'ARN de HIV-1 n'avait pour l'instant pas été étudié ni dans la littérature, ni au laboratoire. De façon surprenante aucun des mécanismes découverts comme régulant les sites d'épissage de HIV-1, ne montrait l'implication de 9G8, alors que plusieurs autres récentes études révélaient son importance seule et/ou en présence d'autres facteurs. Ainsi la protéine 9G8 jouait un rôle dans :

- la régulation de l'épissage de l'exon v9 de CD44 en interaction avec ASF et SRp20 (Galiana-Arnoux *et al.*, 2003)

- la régulation de l'épissage du mRNA de la GnRH avec l'aide de Tra2 et SRp30c (Park *et al.*, 2006)

- la régulation de l'exon 2 de l'αTM en compétition avec les hnRNP H et F (Crawford and Patton, 2006)

- la régulation de l'épissage de l'exon 10 de tau par 9G8 seule (Gao *et al.*, 2007).

La diminution de l'efficacité du site A3 par la mutation $C_{5396}U$ pouvait s'expliquer soit par le fait que le nucléotide C_{5396} est un nucléotide important pour l'activité ESE de cette région de l'ARN, soit par le fait que la mutation crée un site de fixation d'un inhibiteur fort de l'épissage au site A3. La séquence GAGGAG nous l'avons vu, est capable de fixer ASF et son élimination supprime l'activation du site A3 par ASF. Par ailleurs, le fait de cumuler la délétion de cette séquence et la substitution $C_{5396}U$ abolissait l'activation par 9G8. Etant donné qu'une étude sur l'épissage de l'intron D de l'ARNm de la β–globine avait permis de montrer qu'ASF et 9G8 sont capables d'agir en synergie (Li *et al.*, 2000), nous nous sommes demandés si ces 2 protéines SR ne pouvaient pas aussi agir en synergie au niveau du site A3.

-Etude de l'effet activateur 9G8 dépendant de la séquence contenant la séquence GAGGAG en contexte hétérologue

Un premier point à tester était de savoir si la région contenant la séquence GAGGAG et le résidu C_{5396} pouvait être une séquence ARN assurant l'activation de l'épissage par 9G8. Pour

cela, D. Ropers et L. Ayadi ont donc utilisé la construction Sp1-ESEt1, ainsi que la construction Sp1-ESEt2 pour laquelle la séquence insérée a été recentrée autour du nucléotide C_{5396} (ESEt2) (positions 5379 à 5406 dans l'ARN Bru), et une version mutée ($ESEt2_{C5396U}$) (Fig. 101A). Des épissages *in vitro* ont été réalisés sur les ARN produits par ces 2 plasmides et la construction initiale contenant le segment ESEt1, ceci en présence ou en absence d'ajout de protéine ASF/SF2 recombinante. Pour une meilleure observation de l'effet d'ASF/SF2, un mélange 7/1 d'extrait cytoplasmique S100 et d'extrait nucléaire (EN) qui contient peu de protéines SR a été utilisé. On note aussi une diminution de l'effet activateur d'ASF/SF2 en présence de la substitution $C_{5396}U$, elle n'est néanmoins pas très marquée (Fig. 101B).

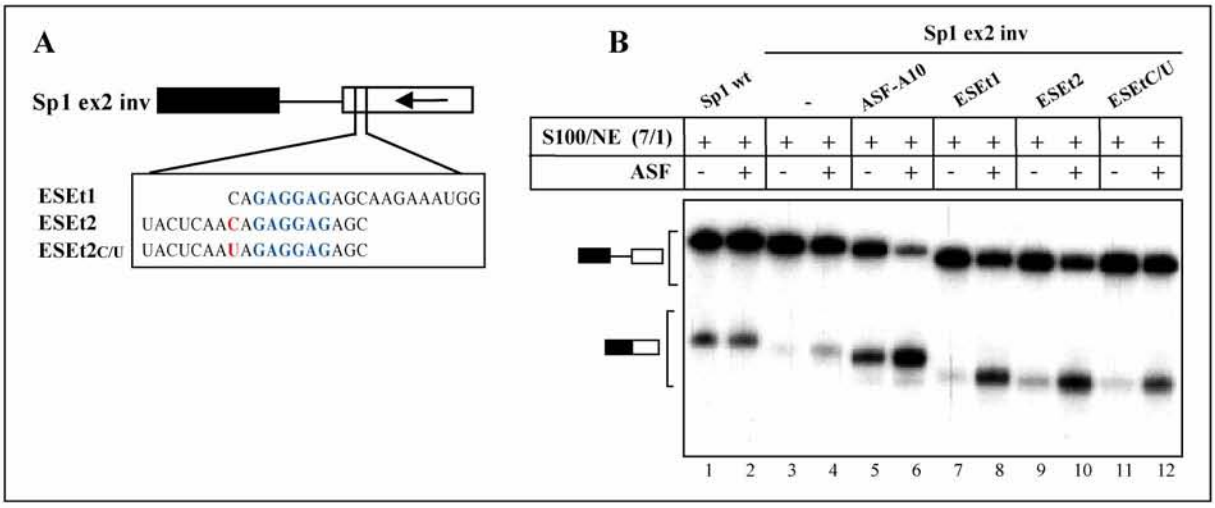


Fig. 101 : Effet de la mutation $C_{5396}U$ sur les propriétés d'activateur ASF/SF2 dépendant de l'élément ESEt

A) Représentation schématique des constructions Sp1-ex2inv utilisées dans le panneau B. B) Effet de la mutation $C_{5396}U$ sur l'épissage des constructions Sp1. Les ARN issus de la transcription des constructions ont été incubés en présence d'un mélange 7/1 d'extrait cytoplasmique S100 et d'extrait nucléaire (EN), seul ou en présence de 0,2 µM de protéine ASF, dans les conditions décrites dans le chapitre Matériels et méthodes. La séquence GST80 servait de témoin négatif et la séquence ASF-A10 de témoin positif.

Avec F. Oillo, nous avons ensuite réalisé des expériences similaires avec la protéine 9G8. Les 3 constructions Sp1-ex2inv contenant les séquences ESEt1, ESEt2 et $ESEt2_{C5396U}$ ont été utilisées pour la production des ARN, ainsi qu'une construction où la séquence codant un site de fixation connu de 9G8 (9G8-102) avait été inséré, construction Sp1-9G8-102 qui servait de témoin positif. La construction Sp1-GST-80 servait toujours de témoin négatif (Cavaloc *et al.*,

1999). Les expériences d'épissage *in vitro* ont été réalisées à la fois en extrait nucléaire et en extrait S100, sans ou avec ajout de protéine 9G8 recombinante.

En extrait nucléaire, 9G8 n'avait pas d'effet activateur sur l'ARN de la construction Sp1 où l'exon 2 n'était pas inversé. Il en était de même lorsque la séquence de l'exon 2 était inversée et dans ce cas le taux d'épissage de base était faible comme attendu. L'activation par 9G8 de l'ARN produit à partir de la construction Sp1-9G8-102 était plus faible qu'attendu. Aucune activation n'était détectée lorsque le fragment ESEt1 était inséré. Au contraire, une activation du taux d'épissage d'un facteur 2 était observée lorsque le segment ESEt2 sauvage était inséré, et l'activation diminuait en présence de la substitution $C_{5396}U$ (Fig. 102B).

Dans le cas où l'expérience était réalisée en extrait S100 contenant peu de protéine 9G8, le taux d'épissage de l'ARN Sp1-9G8-102 était fortement augmenté comme attendu. Dans ce cas, à la fois l'insertion du fragment ESEt1 et celle du fragment ESEt2 conduisaient à une légère augmentation du taux d'épissage. Néanmoins, cette activation était faible par rapport à celle observée pour le contrôle positif Sp1-9G8-102 (Fig. 102C).

Nous observions donc bien un effet activateur de 9G8 lorsque la séquence en amont de la séquence GAG_2 était présente et non mutée et ceci en extrait nucléaire. L'effet était nettement moins marqué lorsque l'épissage était réalisé en S100. Comme en plus ASF s'était avérée avoir des propriétés activatrices en présence de GAG_2 et sa séquence aval, l'ensemble de ces données était en bon accord avec notre hypothèse de départ : à savoir une action conjuguée d'ASF et de 9G8 au niveau de cet élément activateur, avec sans doute (1) la séquence GAG_2 comme cible d'ASF puisque sa délétion élimine l'activation et la fixation d'ASF, et (2) la séquence en amont du motif GAG_2 comme cible de 9G8 puisque la séquence ESEt2 est plus efficace que la séquence ESEt1 en présence de 9G8. Par ailleurs, en accord avec les résultats obtenus *in cellulo*, la mutation $C_{5396}U$ semble bien diminuer l'activation en général.

-Etude le l'effet activateur de la protéine 9G8 sur les transcrits L3-U1 et C3

Ayant mis en évidence des propriétés d'activateur d'épissage 9G8 et ASF/SF2 dépendantes de ces séquences ESEt de l'ARN Bru, il était important de tester l'effet de ces propriétés activatrices *in vitro* dans le contexte HIV-1 authentique. Ainsi, avec F. Oillo et L. Ayadi, nous avons utilisé les constructions L3-U1 et C3 déjà décrites pour tester l'impact des diverses mutations (ΔGAG_2, $C_{5396}U$, Dble) sur l'activation de l'épissage par ASF/SF2 et 9G8. Comme la mutation $C_{5396}U$ diminue la stabilité de la structure tige boucle SLS3 en générant une paire de bases non appariée, nous avons testé en plus dans le cas de l'ARN L3-U1, l'effet de la

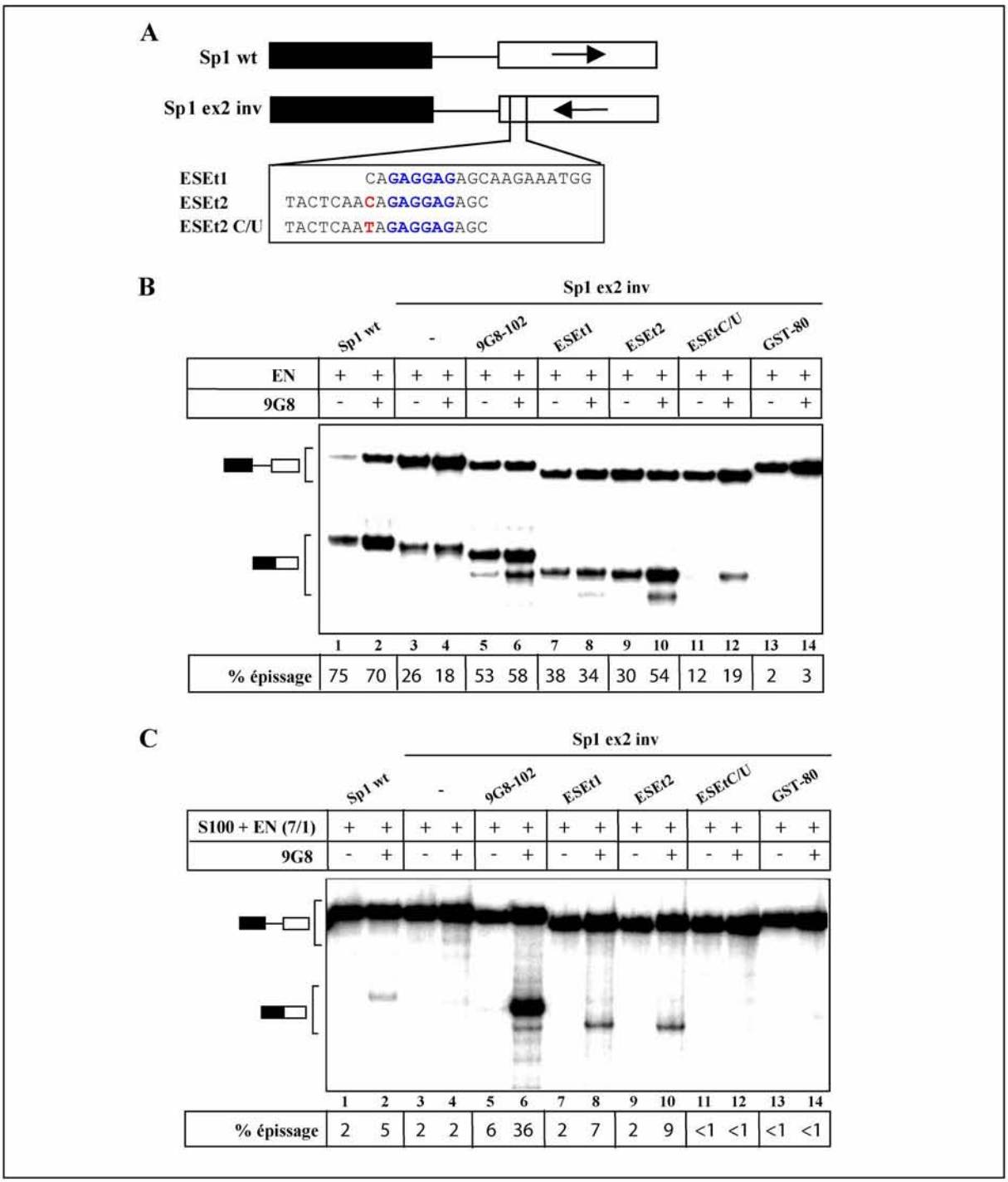


Fig. 102 : Etude des propriétés activatrices 9G8 dépendantes de la région de ESEt

A) Représentation schématique des constructions Sp1wt et des diverses constructions Sp1-ex2inv utilisées. B, C) Effet des séquences ESEt1, ESEt2$_{wt}$ et ESEt2$_{C5396U}$ sur l'épissage de la construction Sp1-ex2inv. Les ARN issus de la transcription des constructions ont été incubés en présence d'extrait nucléaire (panneau B) ou d'un mélange 7/1 d'extrait cytoplasmique S100 et d'extrait nucléaire (panneau C), seul ou avec 400 ng de protéine 9G8, dans les conditions décrites dans le chapitre Matériels et méthodes. La séquence GST80 servait de témoin négatif et la séquence 9G8-102 de témoin positif.

génération d'une mutation $G_{5500}A$ permettant de restaurer la paire de bases. Des ARN L3-U1 portant cette mutation $G_{5500}A$ seule et la double mutation $G_{5500}A$ et $C_{5396}U$ ont été produits. Les épissages ont été réalisés dans un mélange contenant 50% d'extrait nucléaire, 50% de S100. Alors que chacune des mutations ΔGAG_2 et $C_{5396}U$ se sont avérées avoir un effet très négatif sur l'épissage de l'ARN C3, la mutation ΔGAG_2 n'avait pas un effet très marqué sur l'épissage au site A3 dans l'ARN L3-U1 (Fig. 103), ce qui était en bon accord avec les résultats antérieurs (Hallay *et al.*, 2006). En bon accord aussi avec les données obtenues *in cellulo*, la mutation $C_{5396}U$ bloquait complètement l'épissage au site A3 dans la construction L3-U1 et aucune activation n'était détectable ni par ASF/SF2, ni par 9G8. La mutation $G_{5500}A$ seule n'avait pratiquement aucun effet sur le taux de base d'épissage et sur l'activation par ASF/SF2 et 9G8. Par contre, de manière inattendue, l'ajout de la mutation compensatoire augmentait fortement les capacités d'activation du site A3 par ASF/SF2 et surtout par 9G8. Ce qui suggère que le changement de structuration de l'ARN entraîné par la substitution $C_{5396}U$ participe à l'effet négatif que cette mutation a sur l'épissage (Fig. 103).
Ces données confirmaient bien le rôle de cette région de l'ARN contenant le motif GAGGAG pour l'activation de l'utilisation du site A3 par 9G8 et montré que la structuration de l'ARN pourrait jouer un rôle dans cette activation.

2.3. Etude de la fixation de 9G8 et ASF/SF2 par expériences de retard sur gel

Pour continuer à approfondir l'étude du mécanisme d'activation d'ASF/SF2 et de 9G8 au site A3, il était alors important d'essayer de localiser avec précision leurs positionnements relatifs au niveau du motif GAGGAG et des séquences bordantes. Les expériences de retard sur gel et d'empreintes qui avaient été réalisées par D. Ropers, H. Hallay et L. Ayadi montraient que la séquence GAGGAG est un site fort de fixation d'ASF/SF2. Les données fonctionnelles sur l'épissage décrites ci-dessus suggéraient que 9G8 agit en se liant à la séquence amont. Pour conforter cette hypothèse, nous avons réalisé des expériences de retard sur gel.

Dans une première série d'expériences, nous avons réalisé les expériences de retard sur gel sur l'ARN SLS 2/3 uniformément marqué par de l'UTP [α-^{32}P] qui a été incubé en présence des protéines ASF ou 9G8 dans des conditions déjà utilisées par l'équipe de J. Stévenin pour l'étude SELEX de 9G8 (cf. chapitre Matériels et méthodes ; Cavaloc *et al.*, 1999). Comme attendu, la protéine ASF se fixe sur les ARN SLS $2/3_{wt}$ et SLS $2/3_{C5396U}$, et le retard disparaît

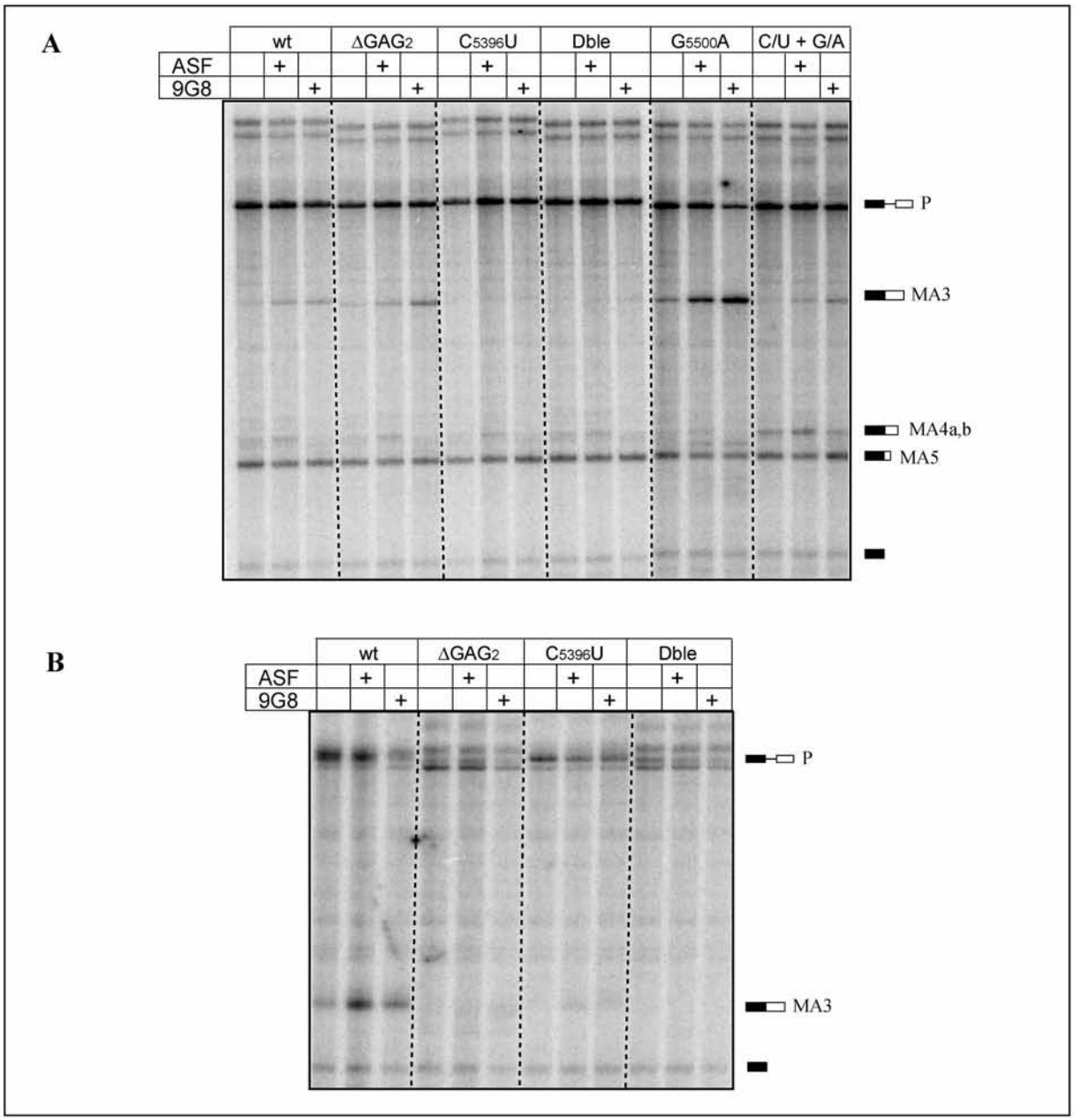


Fig. 103 : Effet des diverses mutations sur l'épissage in vitro au site A3 de l'ARN viral et son activation par les protéines SR ASF/SF2 et 9G8

Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage des ARN uniformément marqués L3U1 (A) ou C3 (B) wt ou mutés. Les transcrits ont été incubés pendant 2h30 dans un mélange 50% extrait nucléaire 50% extrait cytoplasmique S100 sel ou en présence de 400 ng des protéines SR ASF/SF2 ou 9G8 comme décrit dans la partie Matériels et méthodes. La position des précurseurs et des produits épissés est indiquée à gauche de l'autoradiogramme.

lorsque la séquence GAGGAG est délétée (mutant ΔGAG_2 et Dble). Bien que les gels obtenus pour 9G8 soient un peu flous, ils démontraient pour la première fois, la fixation de 9G8 sur l'ARN SLS 2/3 (Fig. 104). Nous notions néanmoins une baisse du taux de fixation en présence de la mutation $C_{5396}U$.

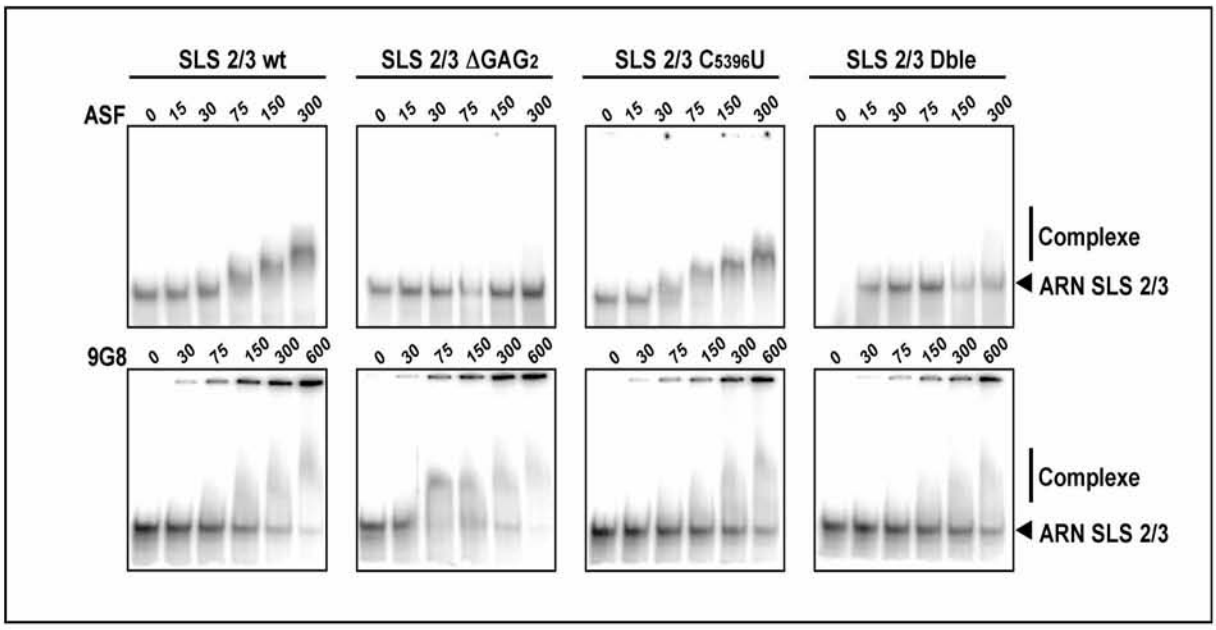


Fig. 104 : Effet des mutations ΔGAG$_2$, C$_{5396}$U et Dble sur la fixation des protéines ASF et 9G98 sur l'ARN SLS 2/3

Analyse par expériences de retard sur gel des interactions entre les ARN SLS 2/3$_{wt}$, SLS 2/3$_{\Delta GAG2}$, SLS 2/3$_{C5396U}$ et SLS 2/3$_{Dble}$, et les protéines ASF ou 9G8. Les ARN uniformément marqués ont été incubés en présence de concentrations croissantes des protéines SR (0 à 600nM), puis les complexes ont été fractionnés en gel de polyacrylamide non dénaturant à 10%. La position de l'ARN libre et des complexes formés est indiquée à droite des autoradiogrammes.

D'après les données de Li et *al.* (Li *et al.*, 2000), qui avaient mis en évidence une synergie d'action d'ASF/SF2 et 9G8, on pouvait se demander si la présence d'ASF pouvait augmenter la fixation de 9G8. Au contraire, pour des problèmes d'encombrement stérique, il était possible que la fixation des deux protéines soit exclusive. Pour tester ces hypothèses, des expériences de retard sur gel ont donc été effectuées en incubant l'ARN avec 9G8 seule ou en présence d'ASF/SF2. Comme il était possible que les protéines puissent interagir par leurs domaines SR, l'expérience a été réalisée avec une protéine 9G8 entière et une protéine délétée de son domaine RS 9G8ΔRS.

Les résultats obtenus ne suggèrent pas l'existence d'une synergie de fixation. Par contre, on constate que la fixation de 9G8 n'empêche pas celle d'ASF/SF2 et *vice versa, c*e qui suggérait l'existence de sites non chevauchants (Fig. 105).

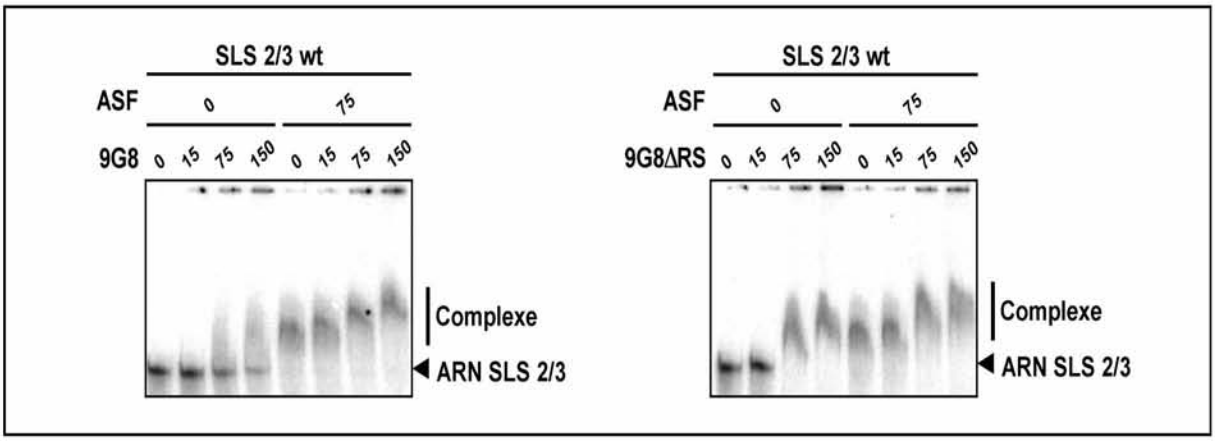


Fig. 105 : Test d'une synergie possible de fixation des protéines ASFSF2 et 9G8

Analyse par expériences de retard sur gel des interactions entre l'ARN SLS 2/3$_{wt}$ et les protéines ASF/SF2 et 9G8 ou 9G8ΔRS. L'ARN a été incubé en présence des protéines 9G8 ou 9G8ΔRS à des concentrations allant de 15 à 150 nM et pour tester l'effet de la présence d'ASF/SF2, les deux mêmes expériences ont été réalisées en présence de 75nM de protéine ASF. Les complexes ont été fractionnés en gel de polyacrylamide non dénaturant. La position de l'ARN libre et des complexes formés est indiquée à droite des autoradiogrammes.

La suite du travail s'est ensuite orientée vers la recherche du site de fixation de 9G8 et donc l'utilisation de différents fragments de l'ARN SLS 2/3. L'un des pré requis de cette étude étant de pouvoir comparer la capacité de fixation de 9G8 pour ces différents ARN, il nous a donc fallu travailler avec les mêmes quantités d'ARN et de radioactivité quelque soit la taille et la composition de l'ARN, d'où l'utilisation d'une faible quantité d'ARN uniformément marqué (permettant d'avoir la même quantité de radioactivité dans chaque expérience) et d'un excès bien quantifié d'ARN froid (permettant d'avoir la même quantité d'ARN dans chaque expérience) (*cf.* Matériels et méthodes).

Le premier transcrit étudié a été le fragment 5366-5409 (souche HIV-1 Bru), appelé SLS E. Il contient une partie du brin 3' de la tige boucle SLS 2-A3 et le début du brin 5' de la tige-boucle SLS 3-A3 contenant la séquence GAGGAG. Trois ARN contrôles avaient été utilisés dans ces expériences : un ARN contenant la séquence aléatoire ne fixant ni ASF/SF2, ni 9G8 (GST-80), qui servait de contrôle négatif, les 2 autres ARN contenaient les séquences de forte affinité, l'une pour ASF/SF2 (ASF-A10), l'autre pour 9G8 (9G8-102) déjà utilisées plus haut.

Les constructions permettant de les produire nous avaient été fournie par l'équipe de J. Stévenin, qui les avait déjà utilisés comme contrôle (Cavaloc *et al.*, 1999). Les résultats obtenus montrent que SLS E, comme SLS 2/3, fixe à la fois ASF et 9G8 (Fig. 106).

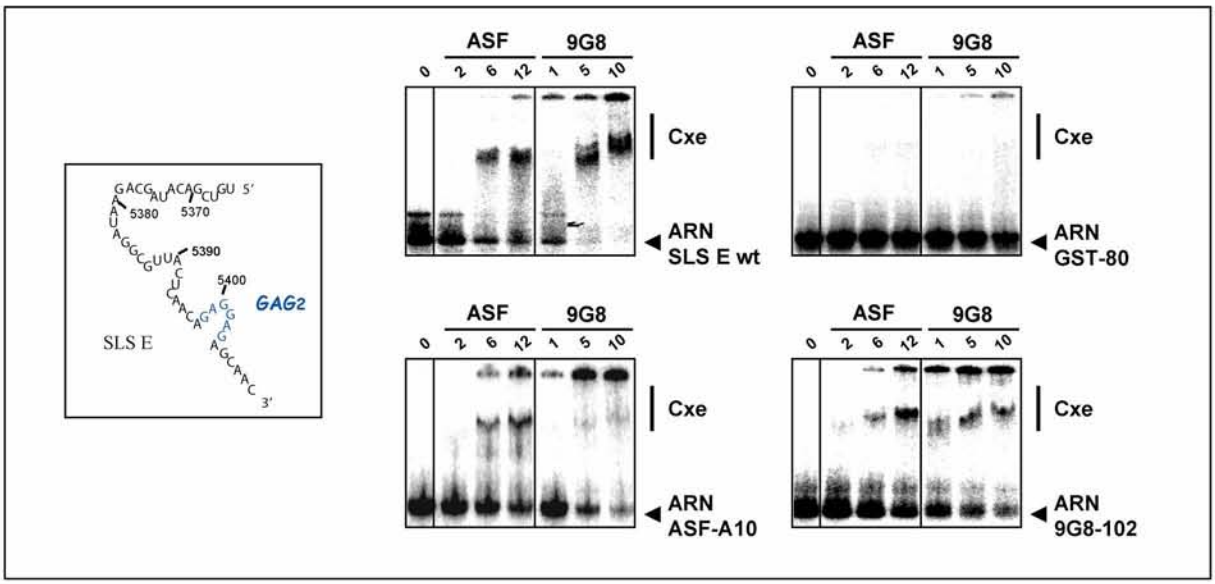


Fig. 106 : Etude de la fixation des protéines ASF et 9G8 sur l'ARN SLS E_{wt} et des ARN contrôle

Analyses par expérience de retard sur gel des interactions s'établissant entre les ARN SLS E_{wt}, GST-80, ASF-A10 et 9G8-102, et les protéines ASF ou 9G8. Les ARN ont été incubés en présence de concentrations croissantes des protéines SR (0 à 12µM), puis les complexes ont été fractionnés sur gel de polyacrylamide non dénaturant. La position de l'ARN libre et des complexes formés est indiquée à gauche des autoradiogrammes.

Nous observions bien une interaction forte à la fois d'ASF/SF2 et de 9G8 avec ce fragment de l'ARN du virus HIV-1 et aucune interaction avec l'ARN servant de contrôle négatif. ASF/SF2 formait bien un complexe avec l'ARN ASF-A10 contenant un site fort de fixation pour cette protéine. On notait la formation d'un agrégat restant dans les poches lorsque 9G8 était incubée avec cet ARN, ceci avec une quantité faible de complexe non agrégé. Dans le cas de l'ARN contrôle ayant une forte affinité pour 9G8, à la fois ASF/SF2 et 9G8 étaient capables de se lier, l'affinité pour ASF/SF2 étant tout de même plus faible comme déjà décrit par Cavaloc et *al.* (Cavaloc *et al.*, 1999). Ces données démontraient bien l'interaction d'ASF/SF2 et 9G8 avec la région de l'ARN viral contenant la séquence GAGGAG.

Par la même approche et avec le même transcrit SLE, nous avons ensuite testé l'effet des mutations ΔGAG_2, $C_{5396}U$ et Dble sur la fixation d'ASF/SF2 et de 9G8 (Fig. 107A).

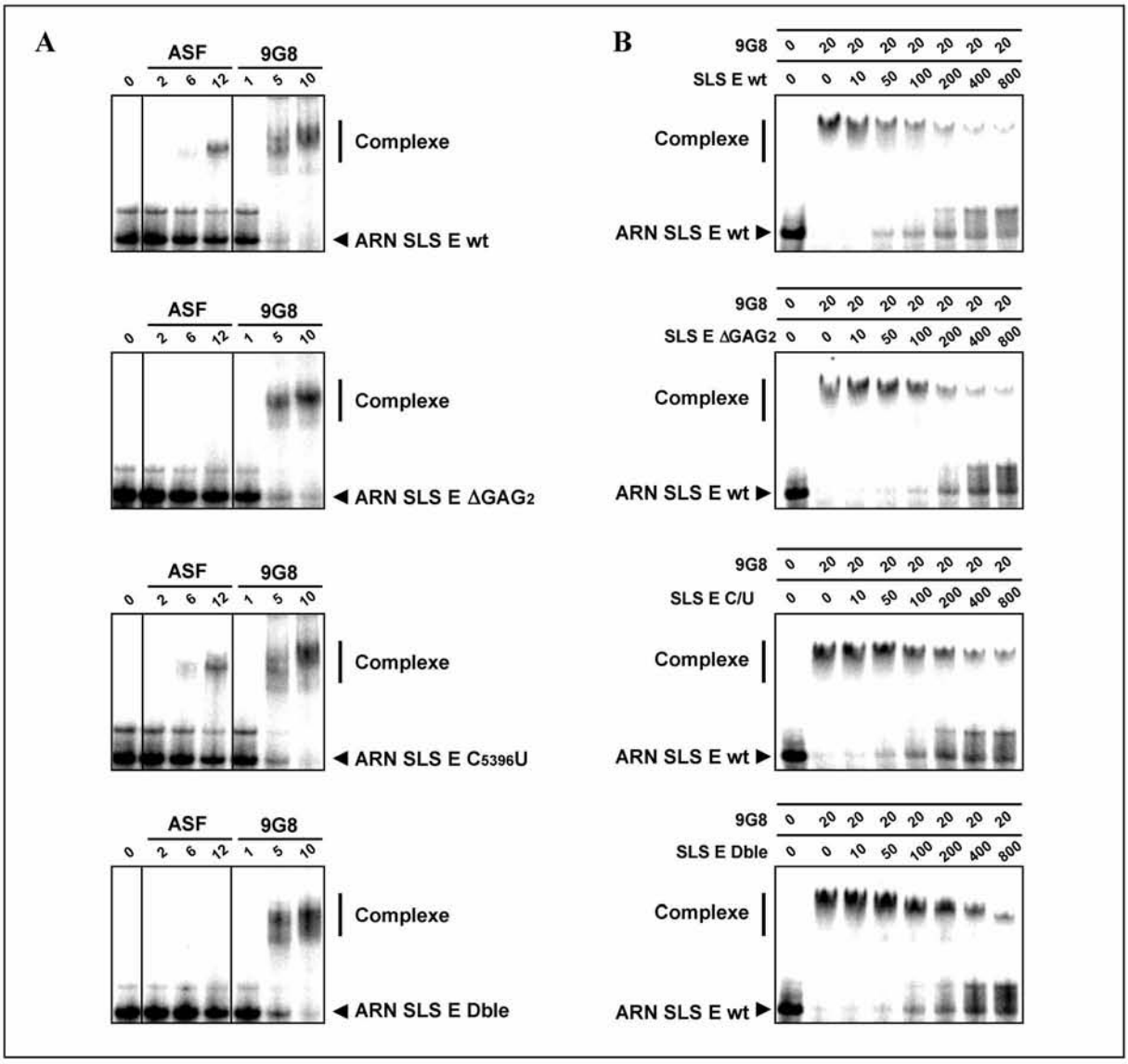


Fig. 107 : Etude des effets des mutations ΔGAG_2 et $C_{5396}U$ et Dble sur la fixation des protéines ASF et 9G8 sur l'ARN SLSE

A) Analyses par expérience de retard sur gel des interactions mises en place entre les ARN SLS E_{wt}, SLS $E_{\Delta GAG2}$, SLS E_{C5396U} et SLS $E_{Dble,}$ et les protéines ASF ou 9G8. Les ARN ont été incubés en présence de concentrations croissantes de protéine ASF ou 9G8 (0 à 12 µM), puis les complexes ont été fractionnés en gel de polyacrylamide non dénaturant. Les positions de l'ARN libre et des complexes formés sont indiquées à droite des autoradiogrammes. B) Analyses par expérience de retard sur gel de la déstabilisation du complexe formé entre l'ARN SLS E_{wt} radiomarqué et la protéine 9G8 à une concentration de 20 µM, par ajout de quantités croissantes d'ARN froid SLS E_{wt}, SLS $E_{\Delta GAG2}$, SLS E_{C5396U} ou SLS E_{Dble} (excès de 0 à 800× d'ARN froid compétiteur).

Comme attendu, la délétion de la séquence GAGGAG entraîne le perte de fixation d'ASF/SF2, mais aucune des mutations ne s'est avérée avoir d'effet sur la fixation de 9G8. Pour compléter ces données, nous avons réalisé des expériences de compétition, les complexes étaient formés entre la protéine 9G8 et l'ARN SLS E_{wt}, et des quantités croissantes d'ARN froids sauvage ou mutés étaient rajoutées. La dissociation du complexe SLS E/9G8 était très similaire quelque soit l'ARN compétiteur utilisé (wt, ΔGAG_2, $C_{5396}U$ ou Dble) (Fig. 107B). Ceci était par ailleurs très surprenant compte tenu des effets cumulatifs observés au niveau épissage. La question se posait donc de savoir si cet effet n'était pas lié à la création par la mutation d'un site de fixation pour une protéine venant contrecarrer la fixation de 9G8.

Pour essayer de mieux cerner les sites de fixation de 9G8 sur la région en aval du site A3, nous avons répété avec F. Oillo, les expériences de retard sur gel avec des ARN chevauchants jalonnant cette région. Quatre ARN, A, AB, BC et C (Fig. 108A) ont alors été utilisés.

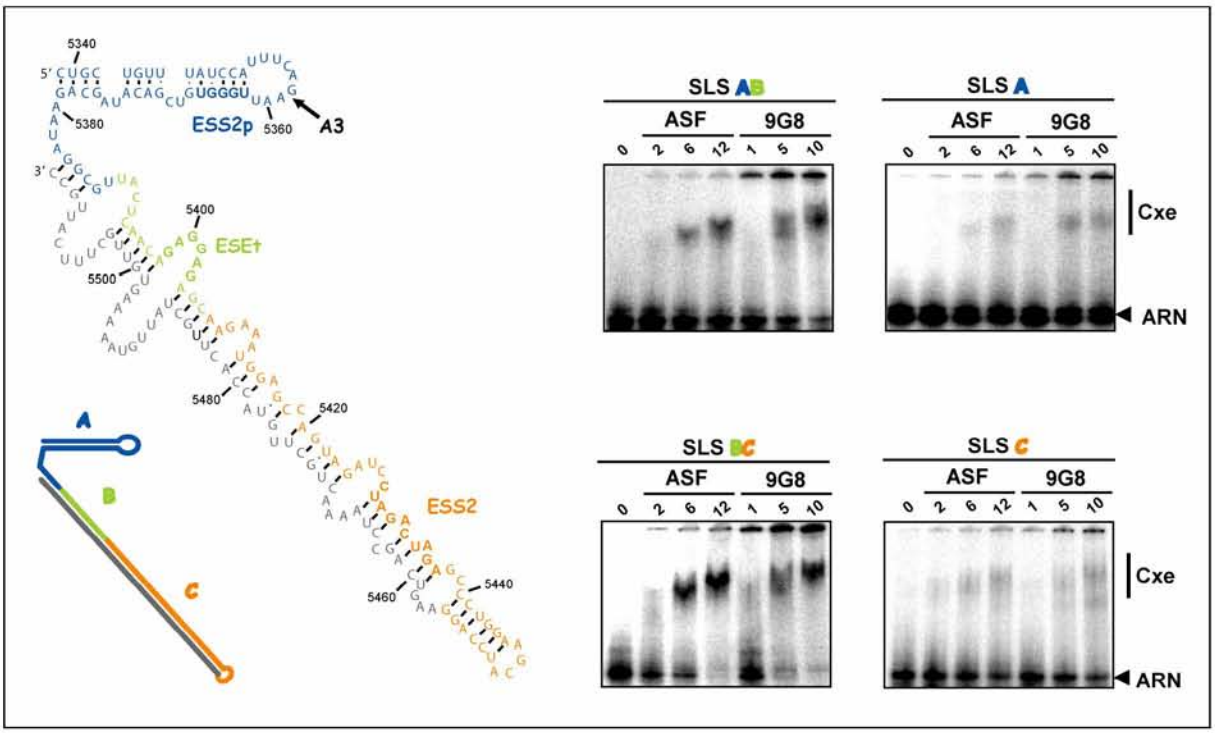


Fig. 108 : Analyse de la fixation des protéines ASF et 9G8 sur différents fragments de l'ARN SLS 2/3

Analyses par expérience de retard sur gel des interactions mises en place entre les ARN composés des séquences A, B et/ou C (schématisées sur la gauche), et les protéines ASF ou 9G8. Les ARN ont été incubés en présence de concentrations croissantes des protéines SR (0 à 12 µM), puis les complexes ont été fractionnés en gel de polyacrylamide non dénaturant. Les positions de l'ARN libre et des complexes formés sont indiquées à gauche des autoradiogrammes.

Les résultats montrent que seuls les ARN AB et BC fixent fortement à la fois ASF/SF2 et 9G8, indiquant donc que le fragment B correspondant à la séquence ESEt2 insérée dans la construction Sp1-ESEt2, fixe bien à la fois ASF/SF2 au niveau du motif GAGGAG et 9G8 en amont.

2.4. Etude de la fixation de 9G8 par la technique d'empreinte en solution

Un autre moyen de définir les sites de fixations de 9G8 était de réaliser comme cela avait été fait pour ASF/SF2, des expériences testant la protection contre l'action des RNases générée sur l'ARN SLS 2/3 par la fixation de 9G8. Ainsi, sur la base du ratio ARN/protéine déjà utilisé pour les expériences de retard sur gel, F. Oillo a formé le complexe avec SLS 2/3 et 9G8, puis l'a traité par la RNase T1 dans des conditions où les coupures ont préférentiellement lieu sur les liaisons phosphodiester en 3' des guanosines en simple brin. Les expériences ont été réalisées avec de l'ARN froid et les positions des clivages ont été analysées par extension d'amorce (comme décrit dans la partie Matériels et méthodes) et comparaison de la position des arrêts à un séquençage réalisé avec la même amorce (amorce 548 s'hybridant à l'extrémité 3' de l'ARN SLS 2/3 des positions 5514 à 5500). Les données obtenues montrent que la protéine 9G8 protége essentiellement la séquence GAGGAG (Fig. 109).

Il est possible que la fixation aie lieu sur la séquence amont comme le suggèrent les résultats que nous avons décrit ci-dessus et que la protection de la séquence GAGGAG aie lieu par encombrement stérique. Il est aussi possible que la fixation de 9G8 sur la séquence en amont de la séquence GAGGAG favorise la fixation d'une seconde molécule sur cette dernière. Les expériences de retard sur gel montraient en effet une affinité de 9G8 pour les séquences riches en purines reconnues par ASF/SF2. Les données montrent que l'élément ESS2 et le motif B peuvent aussi être protégés, ce qui peut participer au mécanisme d'activation du site A3 par 9G8. Néanmoins les empreintes réalisées avec SC35, SRp40 et 9G8 montrent clairement que pour les deux premières protéines, le site primaire de fixation est l'ESS2 et le motif B (Hallay *et al.*, 2006), alors que pour 9G8 c'est la séquence GAGGAG. Dans le cas d'ASF/SF2 qui n'active pas fortement le site A3 seule, aucune fixation n'est observée sur l'élément ESS2. Au travers de l'ensemble des données obtenues, nous arrivions à comprendre pourquoi les protéines SC35, SRp40, 9G8 et ASF/SF2 ont un comportement différent vis-à-vis du site A3.

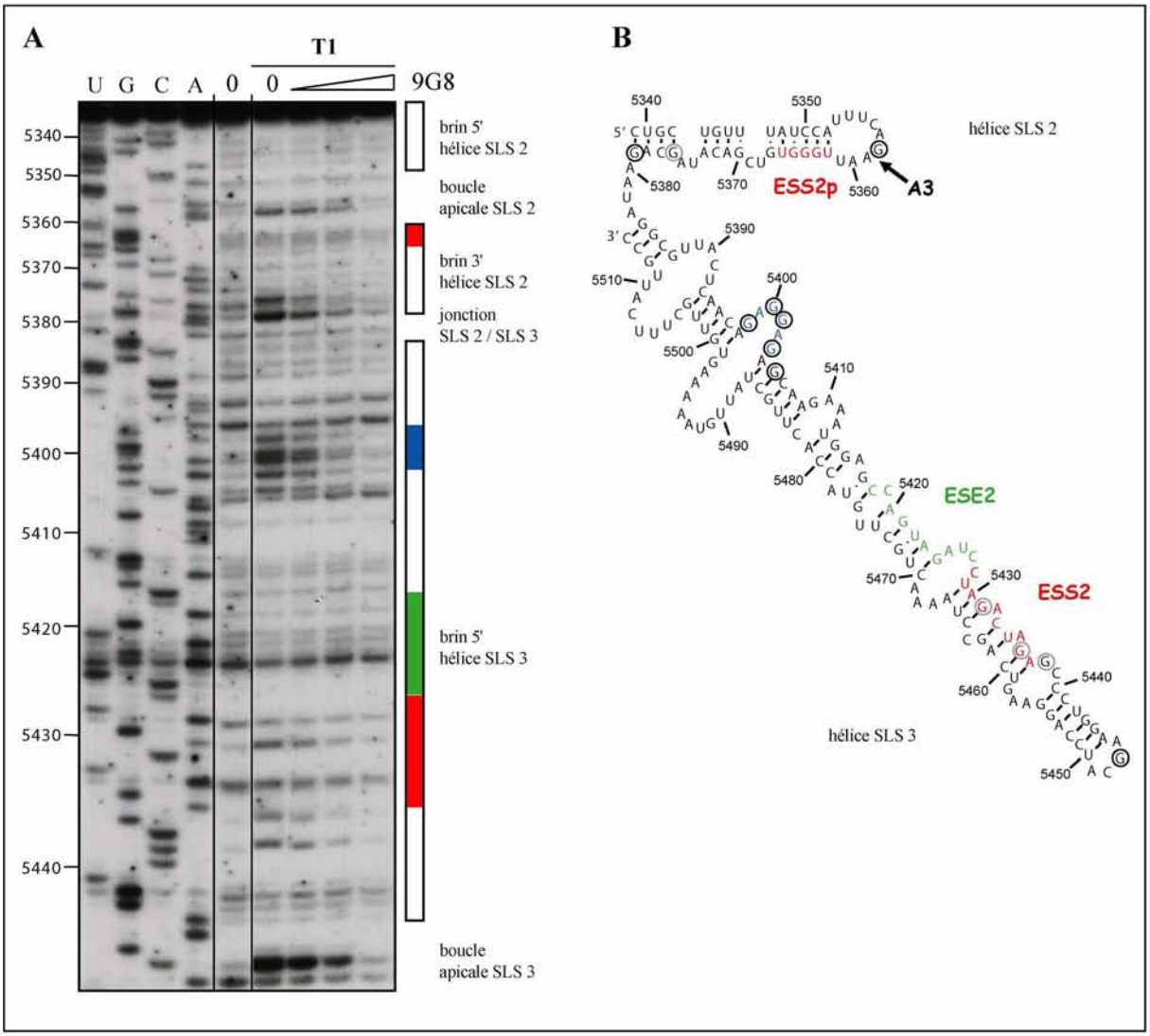


Fig. 109 : Analyse par la RNase T1 de la protection générée par la fixation de la protéine 9G8 sur l'ARN SLS 2/3

Effet de la présence de la protéine 9G8 à des concentrations croissantes sur les coupures de l'ARN SLS 2/3 par la RNase T1. Après incubation en présence de 0, 4, 8 et 12 µM de la protéine 9G8, le transcrit SLS 2/3 a été soumis à des hydrolyses ménagées par 0 ou 0,5 U de RNase T1 puis les produits d'hydrolyse ont été analysés par extension d'amorce par la transcriptase inverse avec l'oligonucléotide 548 comme décrit dans le chapitre Matériels et méthodes. Les coupures ont été identifiées par comparaison avec le séquençage de l'ARN non traité dont les produits de réaction ont été déposés dans les 4 premières pistes. B) Représentation schématique de la structure secondaire de l'ARN SLS 2/3, ainsi que les positions des coupures par la RNase T1. Les nucléotides protégés par la protéine 9G8 à concentration faible sont entourés par un rond noir celle protégées à concentration élevée par un rond gris.

Néanmoins un problème n'était toujours pas résolu : pourquoi la mutation $C_{5396}U$ avait elle un tel effet négatif sur l'épissage au site A3? Nos données suggéraient que l'effet observé ne pouvait pas correspondre seulement à une perte de l'activation par 9G8. Il restait donc la possibilité que la mutation permette la fixation d'une molécule inhibitrice. Pour le tester, nous avons utilisé l'approche protéomique afin d'identifier les protéines nucléaires capables de se fixer sur l'ARN sauvage et sur l'ARN muté.

2.5. Etude de l'effet de la mutation $C_{5396}U$ sur la fixation des protéines nucléaires sur la tige-boucle SLS 3

L'ensemble des données collectées convergeant vers la possibilité que la mutation $C_{5396}U$ génère un site de fixation pour un facteur inhibant le site A3, nous avons appliqué pour la recherche de cet inhibiteur, une approche protéomique qui avait été implantée au laboratoire par V. Marchand, Y. Motorine et C. Aigueperse (thèse V. Marchand). Cette approche est basée sur l'utilisation d'une étape de chromatographie d'affinité permettant de purifier les complexes formés entre l'ARN étudié et les protéines d'un extrait cellulaire. Cette étape de chromatographie est suivie d'un fractionnement des protéines associées à l'ARN par électrophorèse SDS-PAGE. Les protéines contenues dans les bandes de gel sont ensuite identifiées par spectrométrie de masse nanoLC-MS/MS. Le principe est donc de rajouter transcriptionnellement une séquence à une des extrémités de l'ARN étudié. Deux types de séquences ont été utilisées au laboratoire pour la réalisation de l'étape de chromatographie d'affinité, (1) l'ajout d'une séquence isolée par SELEX comme ayant une forte affinité pour la tobramycine (dans ce cas des billes où la tobramycine est fixée sont utilisées), et (2) une séquence de l'ARN du phage MS2, correspondant au site de fixation de la protéine de l'enveloppe de ce phage. Ce dernier système s'est avéré plus performant car les billes utilisées dans ce cas là sont des billes d'amylose qui ont un faible pouvoir de rétention des protéines en comparaison avec des billes de sépharose.

Les billes d'amylose permettent de fixer de manière stable une protéine de fusion MS2-MBP contenant en C-terminal la Maltose Binding Protein qui a une forte affinité pour l'amylose et en N-terminal, le domaine de fixation à l'ARN de la protéine de l'enveloppe du phage MS2. La partie N-terminale de cette protéine de fusion est donc capable d'interagir avec le motif ARN du phage MS2 se liant à la protéine de l'enveloppe. Ce motif forme une structure tige-boucle et 3 structures tige-boucle MS2 sont ajoutées en général à l'extrémité 3' de l'ARN étudié. Ainsi, un ARN d'intérêt contenant ces 3 motifs (M3) peut être immobilisé sur les

billes après fixation de la protéine de fusion au domaine M3 (Fig. 110A). Il est ensuite possible d'incuber les billes associées à l'ARN avec un extrait cellulaire. Les complexes ribonucléoprotéiques formés restent ainsi fixés sur les billes. Les protéines associées avec une faible affinité ou retenues de manière aspécifique sont ensuite éliminées par des étapes de lavage. Ses lavages ne peuvent néanmoins pas être réalisés à des concentrations salines très élevées, sous peine de rompre l'interaction entre la protéine de fusion et l'ARN. Cette interaction ne résiste en effet pas à une concentration en cations monovalents supérieure à 300 mM. Le complexe MS2-MBP/M3-ARN/protéines de l'extrait peut ensuite être élué des billes par ajout de maltose (Fig. 110B). Après fractionnement des protéines de ce complexe en gel de polyacrylamide dénaturant, soit les bandes majeures sont découpées, soit le gel est découpé de manière uniforme. Les protéines contenues dans les bandes peuvent ensuite être analysées sur un spectromètre de masse ESI-MS/MS[1] couplé à un système chromatographique nanoLC[2] (nanoLC-MS/MS).

Pour tester si la mutation $C_{5396}U$ modifie les partenaires de la région SLS3 de l'ARN du virus HIV-1, nous avons utilisé l'ARN SLS AB wt ou muté ($C_{5396}U$) que nous avons décrit ci-dessus dans le cadre des études de la fixation de 9G8 par expériences de gel retard. Nous avons ajouté le domaine M3 en 3' des deux ARN (soit $SLSAB_{wt}$-M3 et $SLSAB_{C5396U}$-M3). Les expériences antérieures du laboratoire montraient que plus l'ARN fusionné au motif M3 est long, plus un grand nombre de protéines liées de manière systématique à l'ARN soit double brin, soit simple brin, sont retenues dans ces expériences. Vu sa faible taille (68 nt), un faible nombre de liaisons aspécifiques était attendue pour l'ARN SLS AB. De plus il contient le site A3 et sa séquence polypyrimidine, et d'après ce que nous avons montré ci-dessus, des sites de fixation d'ASF/SF2 et 9G8. Les 2 ARN ainsi que l'ARN M3 non fusionné servant de témoin, ont donc été incubés en présence de la protéine MS2-MBP fixés sur les billes, puis incubés avec un extrait nucléaire de cellules HeLa commercial. Après lavage avec un tampon composé d' Hepes KOH pH 7,9 20mM ; KCl 100 mM ; Glycérol 20% ; EDTA 0,2 mM ; DTT 0,5 mM ; MgCl2 3 mM, les complexes ont été élués à l'aide de maltose 10 mM, puis déposés sur gel SDS-PAGE (Fig. 111).

[1] Electro Spray Ionisation MS/MS
[2] nano Liquid Chromatography

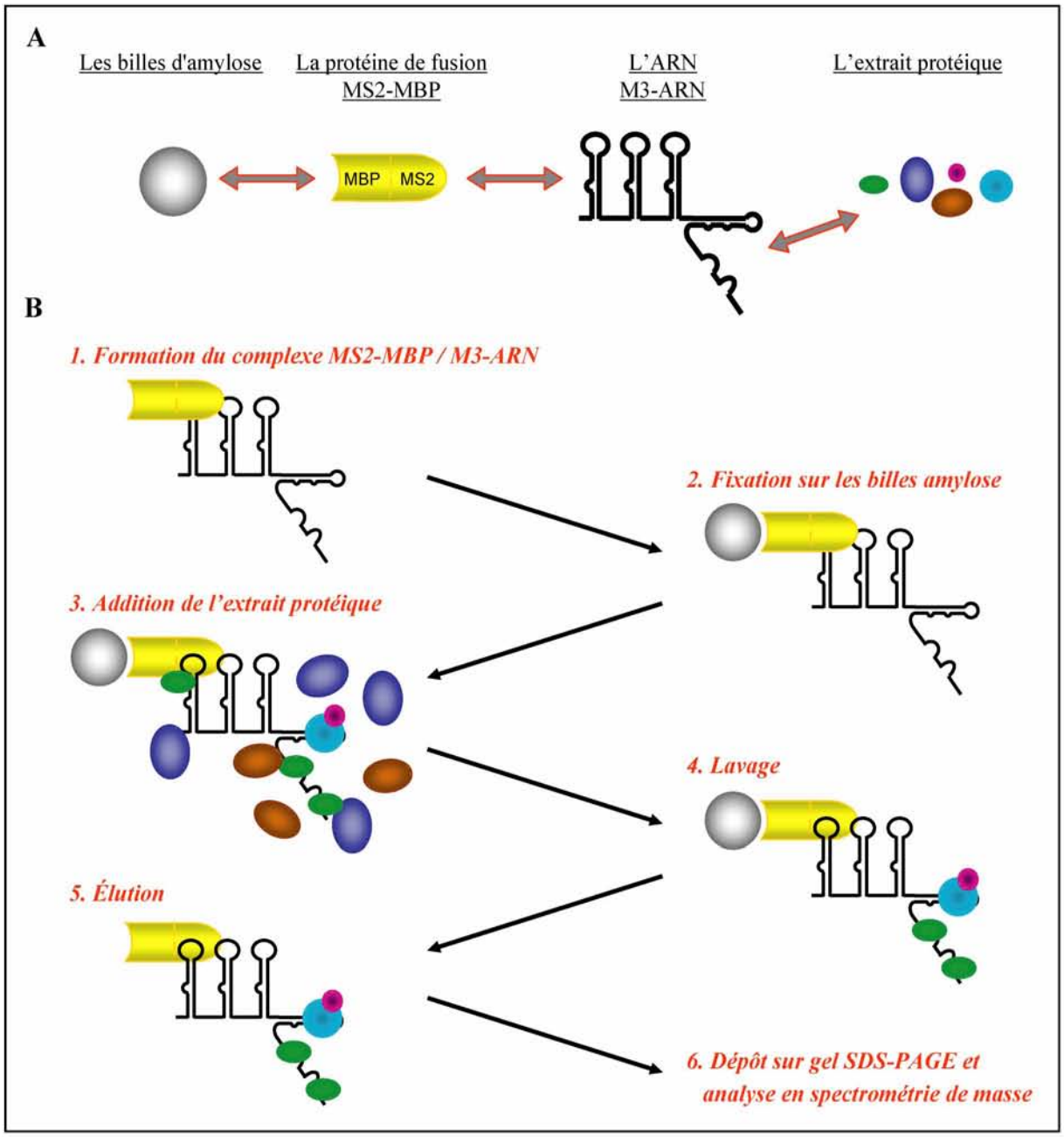


Fig. 110 : Principe de la purification de complexe ribonucléoprotéiques par chromatographie d'affinité reposant sur l'emploi de la protéine de l'enveloppe du phage MS2 et de la protéine MBP

A) Les différents partenaires utilisés dans cette approche. B) Etapes de la purification des complexes RNP. Le transcrit composé de l'ARN d'intérêt et du motif M3 est incubé en présence de protéine MS2-MBP (1), les billes d'amylose sont ensuite ajoutées (2) et le tout est incubé en présence de l'extrait protéique (3). Les complexes formés sont ensuite lavés (4), puis élués par ajout de maltose et les protéines qu'ils contiennent sont directement fractionnées par électrophorèse en gel SDS-PAGE (6). Les protéines sont ensuite colorées au bleu colloïdal et les bandes du gel sont découpées afin de réaliser une analyse par spectrométrie de masse.

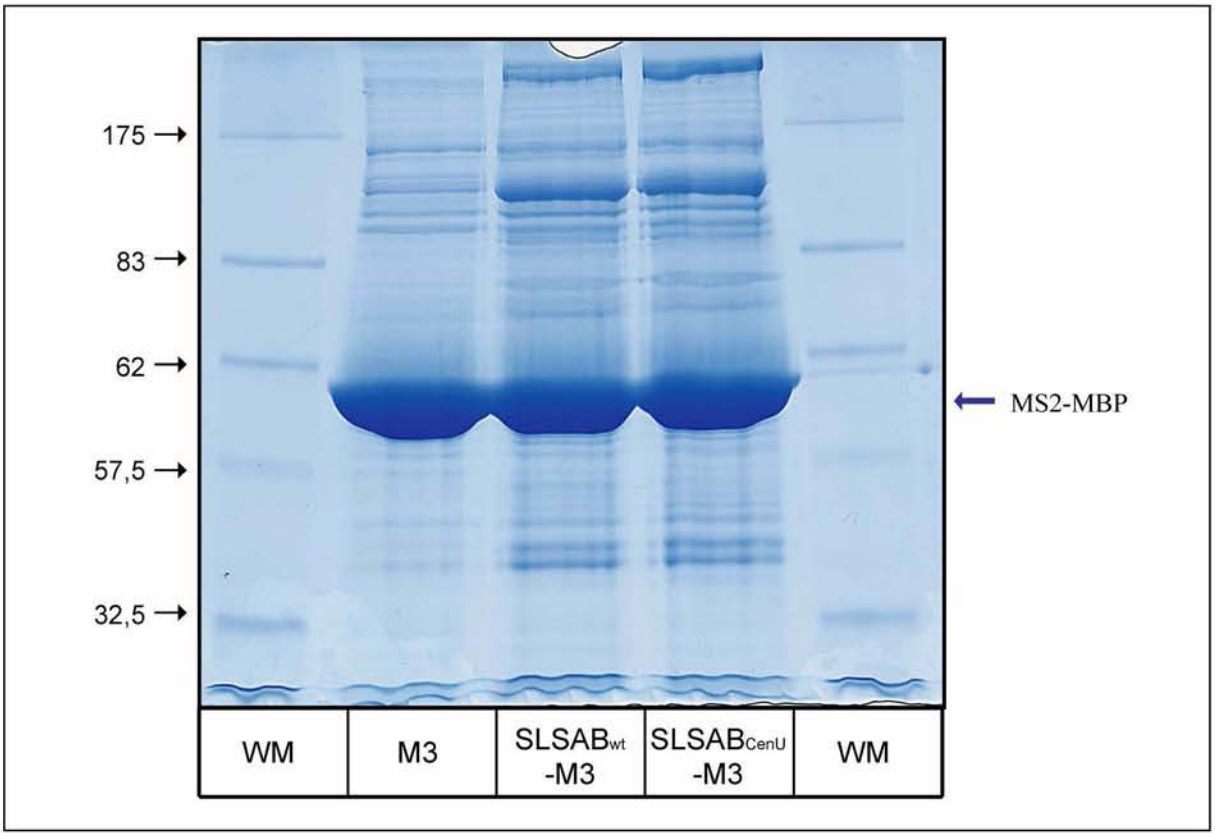


Fig. 111 : Analyse par électrophorèse SDS-PAGE des protéines contenues dans les complexes ribonucléoprotéiques purifiés par chromatographie d'affinité reposant sur l'emploi de la protéine de fusion MS2-MBP

Environ 400 pmoles d'ARN M3, SLS AB_{wt}-M3 ou SLS AB_{CenU}-M3 ont été incubés en présence de 2 nmol de protéine MS2-MBP, puis 50 µl de billes et enfin 500 µl d'extrait nucléaire de cellules HeLa. Après lavage, les complexes RNP ont été élués par ajout de maltose 10 mM et déposés sur gel SDS-PAGE 10%. Le gel a ensuite été coloré par une solution de bleu de Coomassie colloïdal. Un marqueur de poids moléculaire a été déposé à gauche du gel (WM).

Aucune différence marquante ne ressortait entre le profil de fractionnement des protéines obtenu pour l'ARN sauvage et celui obtenu pour l'ARN muté. Aussi nous avons décidé de réaliser une découpe systématique des 3 pistes d'intérêt du gel (les deux ARN de fusion et le contrôle MS2), ceci tout les 1,5 mm. Une analyse nanoLC-MS/MS a ensuite été réalisée sur le contenu de chacune des bandes. Ainsi, 44 protéines ont été retrouvées dans la piste témoin et chacune des pistes correspondant aux essais. Nous n'avons donc pas tenu compte de ces protéines qui doivent se lier sans spécificité aux ARN ou qui se lient spécifiquement aux structures tige boucle de l'ARN MS2. Par ailleurs, 71 protéines n'ont été identifiées que dans les pistes correspondant aux ARN SLS sauvage ou muté (Voir annexe en fin de chapitre). La majorité de ces protéines étaient communes aux 2 pistes. Dans cette catégorie, nous avons pu

trouver notamment les facteurs U2AF 35 et 65, des sous unités du facteur Sf3b (Sf3b1 et Sf3b4) et des protéines hnRNP (A/B, A0, D0 et M). Ceci était attendu du fait de la présence de la séquence polypyrimidine du site A3 et du fait que les protéines hnRNP sont en général toujours retrouvées dans ce type d'expérience. De part la grande quantité de résidus lysine présents dans le domaine RS, les protéines SR donnent très peu de peptides ionisables et leur détection est plus difficile. Malgré cela, nous avons pu retrouver les protéines ASF et SC35 dans les 2 cas, ceci en accord avec nos données et celle de H. Hallay et *al.*. La protéine SRp20 était aussi présente. Point dont il faudra tenir compte dans de futures études. Une analyse plus détaillée des spectres a aussi permis de détecter la présence des peptides de 9G8 à la fois pour $SLSAB_{wt}$-M3 et $SLSAB_{C5396U}$-M3, ce qui à nouveau confirme nos données.

L'objectif initial étant de trouver soit des facteurs activateurs disparaissants, soit des facteurs inhibiteurs apparaissants avec la mutation, les protéines étant identifiées seulement dans l'une ou l'autre des pistes des ARN SLS sauvage ou muté, nous intéressait tout particulièrement. Ainsi en accord avec l'effet inhibiteur de la mutation $C_{5396}U$ sur l'épissage nous avons pu constater la disparition de 3 facteurs associés à la snRNP U2 (SR140, p14 et SF3b125) et un facteur associé à la snRNP U5 (PRP8) dans les complexes formés avec l'ARN muté, ceci en dépit de l'association des facteurs U2AF 35 et 65, des sous unités du facteur Sf3b1 et Sf3b4 du facteur Sf3B avec cet ARN muté comme mentionné ci-dessus. Il est aussi intéressant de noter l'apparition du facteur PUF60 sur l'ARN muté. Ce facteur a été initialement découvert comme étant impliqué dans des réactions d'épissage de faible efficacité (Jurica and Moore, 2003 ; Page-McCaw *et al.*, 1999), et ce n'est que récemment qu'il a pu être démontré qu'il possède les mêmes capacités que le facteur U2AF65 (Hastings *et al.*, 2007). A noter aussi que les protéines hnRNP A3 et L, qui sont des partenaires du facteur PUF60 (Ewing *et al.*, 2007), étaient identifiées dans un plus grand nombre de bande et étaient donc présentes en plus grande quantité sur l'ARN $SLSAB_{C5396U}$ par rapport à l'ARN sauvage.

Parmi les protéines déjà décrites comme inhibitrices de l'épissage et que l'on observait plus spécifiquement sur l'ARN muté, on peut noter les protéines hnRNP F et G, dont on sait qu'elles peuvent se lier à l'élément inhibiteur ESS2p contenu dans l'ARN sauvage. Il reste donc à savoir si leur fixation est renforcée sur l'ARN muté.

Le facteur DAZAP1 est aussi apparu comme un facteur potentiellement très intéressant, car il n'était présent que sur l'ARN $SLSAB_{C5396U}$ et que pour le moment, il n'a quasiment jamais été détecté dans d'autres complexes nucléoprotéique analysés au laboratoire. Cette protéine contient 2 domaines RRM, ce qui suggère qu'elle interagit avec l'ARN (Tsui *et al.*, 2000). Autre point très intéressant, DAZAP1 a été très récemment décrit comme un inhibiteur

d'épissage, ceci dans deux articles publiés par l'équipe de F. Baralle (ICGEB, Trieste, Italie). Dans ces deux articles, la protéine DAZAP1 est décrite comme ayant un caractère inhibiteur de l'épissage en association avec les protéines hnRNP A1 et A2 (Skoko *et al.*, 2008 ; Goina *et al.*, 2008). La mise en évidence de la protéine DAZAP1 par ces deux équipes a aussi été obtenue par approche protéomique. Les deux groupes d'auteurs ont chacun démontré que la liaison de cette protéine est due à une mutation ponctuelle détectée chez des patients Dans un cas, la mutation est localisée dans l'exon 18 du gène BRCA1 et dans l'autre cas, elle est située dans l'exon 37 du gène NF1. Ces deux mutations génèrent un motif UAG, conduisant à la fixation du facteur DAZAP1 et des protéines hnRNP A1 et A2, ce qui entraîne l'exclusion de l'exon contenant la mutation (Skoko *et al.*, 2008 ; Goina *et al.*, 2008). En comparant les 2 séquences mutées dans les gènes BRCA1 et NF1 et la séquence contenant la mutation $C_{5396}U$ dans l'ARN du virus HIV-1, on peut constater qu'elles conduisent toutes à la génération d'un motif UAG dont on connaît l'importance pour la fixation de hnRNP A1 (Fig. 112).

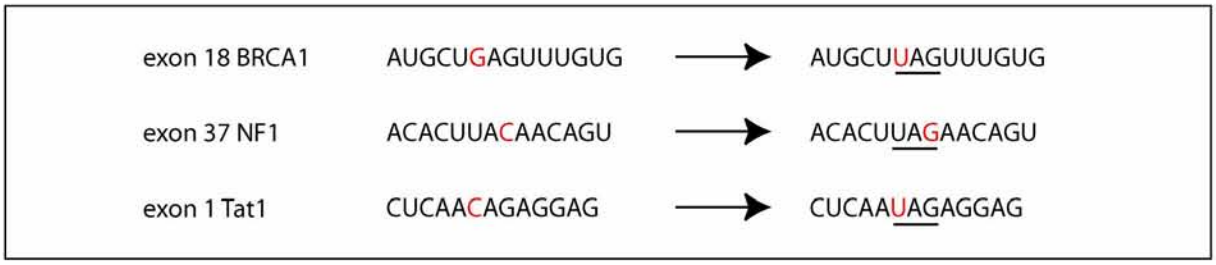


Fig. 112 : Comparaison de la séquence de l'exon 2 sauvage et muté avec celles des ARN décrits comme fixant les protéines DAZAP1, hnRNP A1 et A2

En rouge sont indiquées les mutations conduisant à la formation du motif AUG (souligné) et à la fixation les protéines DAZAP1, hnRNP A1 et A2 dans les exons 18 de BRCA1 et 37 de NF1.

A part ce motif UAG, les 3 séquences ont peu de points communs. Comme les motifs UAG sont connus pour être importants pour la fixation de hnRNPA1, on peut se demander si c'est hnRNP A1 qui est fixée sur les motifs UAG générés. On peut imaginer que l'interaction entre hnRNP A1 et DAZAP1 puisse renforcer la stabilité du complexe formé entre hnRNP A1 et l'ARN. Une autre possibilité est que DAZAP1 interagisse elle-même directement avec des séquences ARN contenant un motif UAG. Ceci, en interagissant en plus avec hnRNP A1 et hnRNP A1, elles mêmes fixées à des sites voisins. On sait que hnRNP A1 est capable de se lier à la fois au brin 5' (ESEt) et au brin 3' de la boucle interne II de la tige boucle SLS3, qui est située juste en aval de la séquence mutée (Hallay *et al.*, 2006). Par ailleurs, hnRNP A2 se

lie aussi sur une séquence de la structure tige boucle SLS3 située au voisinage de la boucle interne II (Mouland *et al.*, 2001).

Pour définir laquelle de ces hypothèses est la plus vraisemblable, il nous faudra dans un premier temps produire la protéine DAZAP1 recombinante, sachant que les protéines hnRNP A1 et A2, sont déjà disponibles au laboratoire. Nous pourrons alors tester les interrelations existantes lors de leur fixation à l'ARN SLS 3, ceci sur l'ARN sauvage et sur l'ARN muté. Les techniques de gel retard, d'empreintes chimiques et enzymatiques et de pontages aux UV pourront être utilisées, ainsi que la purification et l'analyse supramoléculaire par spectrométrie de masse de complexes formés entre SLS3 et les 3 protéines. Il sera bien sûr aussi intéressant d'étudier l'effet de la déplétion de la protéine DAZAP1 sur l'épissage au site A3 dans les constructions C3 ou L3-U1 portant la mutation $C_{5396}U$. Si DAZAP1 est bien responsable de l'inhibition, on devrait pouvoir restaurer un épissage actif des ARN mutés par cette déplétion. Nous pouvons bien sûr aussi envisager de limiter le taux de protéine DAZAP1 *in cellulo* par utilisation de siRNA et de tester l'effet produit sur l'épissage de l'ARN des plasmides pΔPSP sauvage et mutés. Ces différentes expériences devraient nous permettre de valider ou d'invalider l'hypothèse d'un rôle essentiel de DAZAP1 dans l'effet très négatif de la mutation $C_{5396}U$.

Annexe : Résultats de protéomique

AC	Protéine	Taxonomie	Masse	pI	MS2	SLS2/E Wt	SLS2/E CU
hnRNP							
O43390 HNRPR_HUMAN	Heterogeneous nuclear ribonucleoprotein R	Homo sapiens	70899	8.23	X	x	x
O60506 HNRPQ_HUMAN	Heterogeneous nuclear ribonucleoprotein Q	Homo sapiens	69560	8.68	X	x	x
P09651 ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	Homo sapiens	38822	9.26	X	x	x
P14866 HNRPL_HUMAN	Heterogeneous nuclear ribonucleoprotein L	Homo sapiens	60149	6.65			x
P22626 ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1	Homo sapiens	37407	8.97	X	x	x
P31942 HNRH3_HUMAN	Heterogeneous nuclear ribonucleoprotein H3	Homo sapiens	36903	6.37		x	
P38159 HNRPG_HUMAN	Heterogeneous nuclear ribonucleoprotein G	Homo sapiens	42306	10.06			x
P51991 ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	Homo sapiens	39571	9.10	X	x	x
P52272 HNRPM_HUMAN	Heterogeneous nuclear ribonucleoprotein M	Homo sapiens	77464	8.84		x	x
P52597 HNRPF_HUMAN	Heterogeneous nuclear ribonucleoprotein F	Homo sapiens	45643	5.38			x
P61978 HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	Homo sapiens	50944	5.39	X	x	x
Q00839 HNRPU_HUMAN	Heterogeneous nuclear ribonucleoprotein U	Homo sapiens	90457	5.76	X	x	x
Q13151 ROA0_HUMAN	Heterogeneous nuclear ribonucleoprotein A0	Homo sapiens	30822	9.34		x	x
Q14103 HNRPD_HUMAN	Heterogeneous nuclear ribonucleoprotein D0	Homo sapiens	38410	7.62		x	x
Q99729 ROAA_HUMAN	Heterogeneous nuclear ribonucleoprotein A/B	Homo sapiens	36202	8.22		x	x
Q9BUJ2 HNRL1_HUMAN	Heterogeneous nuclear ribonucleoprotein U-like protein 1	Homo sapiens	95679	6.49	X	x	x
Helicase							
O00571 DDX3X_HUMAN	ATP-dependent RNA helicase DDX3X	Homo sapiens	73198	6.73	X	x	x
O43143 DHX15_HUMAN	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	Homo sapiens	90875	7.12	X	x	x
P12956 KU70_HUMAN	ATP-dependent DNA helicase 2 subunit 1	Homo sapiens	69799	6.23	X	x	x
P13010 KU86_HUMAN	ATP-dependent DNA helicase 2 subunit 2	Homo sapiens	82652	5.55	X	x	x
P17844 DDX5_HUMAN	Probable ATP-dependent RNA helicase DDX5	Homo sapiens	69105	9.06	X	x	x
Q08211 DHX9_HUMAN	ATP-dependent RNA helicase A	Homo sapiens	140869	6.41		x	x
Q86XP3 DDX42_HUMAN	ATP-dependent RNA helicase DDX42	Homo sapiens	102912	6.54		x	
Q92499 DDX1_HUMAN	ATP-dependent RNA helicase DDX1	Homo sapiens	82380	6.81	X	x	x

Q92841 DDX17_HUMAN	Probable ATP-dependent RNA helicase DDX17	Homo sapiens	72326	8.82		x	
Q9H2U1 DHX36_HUMAN	Probable ATP-dependent RNA helicase DHX36	Homo sapiens	114704	7.57	X	x	x
Q9NR30 DDX21_HUMAN	Nucleolar RNA helicase 2	Homo sapiens	87290	9.32	X	x	x
Q9UHX1 PUF60_HUMAN	Poly(U)-binding-splicing factor PUF60	Homo sapiens	59838	5.19			x

Splicing factor

O75533 SF3B1_HUMAN	Splicing factor 3B subunit 1	Homo sapiens	145723	6.58	X	x	x
P23246 SFPQ_HUMAN	Splicing factor, proline- and glutamine-rich	Homo sapiens	76102	9.45	X	x	x
P26368 U2AF2_HUMAN	Splicing factor U2AF 65 kDa subunit	Homo sapiens	53467	9.19			x
P84103 SFRS3_HUMAN	Splicing factor, arginine/serine-rich 3	Homo sapiens	19318	11.64		x	x
Q01081 U2AF1_HUMAN	Splicing factor U2AF 35 kDa subunit	Homo sapiens	27854	9.09		x	
Q01130 SFRS2_HUMAN	Splicing factor, arginine/serine-rich 2	Homo sapiens	25461	11.86	X	x	
Q07955 SFRS1_HUMAN	Splicing factor, arginine/serine-rich 1	Homo sapiens	27728	10.37	X	x	x
Q12874 SF3A3_HUMAN	Splicing factor 3A subunit 3	Homo sapiens	58812	5.27		x	
Q15393 SF3B3_HUMAN	Splicing factor 3B subunit 3	Homo sapiens	135492	5.13	X	x	x
Q15427 SF3B4_HUMAN	Splicing factor 3B subunit 4	Homo sapiens	44357	8.54		x	x
Q6P2Q9 PRP8_HUMAN	Pre-mRNA-processing-splicing factor 8	Homo sapiens	273427	8.95		x	

Small nuclear ribonucleoprotein

O75643 U520_HUMAN	U5 small nuclear ribonucleoprotein 200 kDa helicase	Homo sapiens	244353	5.73	X	x	x
P09661 RU2A_HUMAN	U2 small nuclear ribonucleoprotein A'	Homo sapiens	28398	8.72	X	x	x
P14678 RSMB_HUMAN ou P63162	Small nuclear ribonucleoprotein-associated proteins B and B' ou N	Homo sapiens	24594	11.20		x	
P62306 RUXF_HUMAN	Small nuclear ribonucleoprotein F	Homo sapiens	9719	4.70		x	
P62314 SMD1_HUMAN	Small nuclear ribonucleoprotein Sm D1	Homo sapiens	13273	11.56		x	x
P62316 SMD2_HUMAN	Small nuclear ribonucleoprotein Sm D2	Homo sapiens	13518	9.92		x	x
P62318 SMD3_HUMAN	Small nuclear ribonucleoprotein Sm D3	Homo sapiens	13907	10.33			x

Ribosomal protein

O76021 RL1D1_HUMAN	Ribosomal L1 domain-containing protein 1	Homo sapiens	54939	10.13		x	x
P05388 RLA0_HUMAN	60S acidic ribosomal protein P0	Homo sapiens	34252	5.71	X		

P08865 RSSA_HUMAN	40S ribosomal protein SA	Homo sapiens	32833	4.79	X	x	x
P15880 RS2_HUMAN	40S ribosomal protein S2	Homo sapiens	31305	10.25			x
P18124 RL7_HUMAN	60S ribosomal protein L7	Homo sapiens	29207	10.66	X	x	x
P23396 RS3_HUMAN	40S ribosomal protein S3	Homo sapiens	26671	9.68	X	x	x
P35268 RL22_HUMAN	60S ribosomal protein L22	Homo sapiens	14778	9.21		x	x
P46782 RS5_HUMAN	40S ribosomal protein S5	Homo sapiens	22862	9.73		x	x
P61247 RS3A_HUMAN	40S ribosomal protein S3a	Homo sapiens	29926	9.75		x	x
P62081 RS7_HUMAN	40S ribosomal protein S7	Homo sapiens	22113	10.09	X		
P62249 RS16_HUMAN	40S ribosomal protein S16	Homo sapiens	16435	10.21			x
P62269 RS18_HUMAN	40S ribosomal protein S18	Homo sapiens	17708	10.99			x
P62424 RL7A_HUMAN	60S ribosomal protein L7a	Homo sapiens	29977	10.61			x
P62701 RS4X_HUMAN	40S ribosomal protein S4, X isoform	Homo sapiens	29579	10.16			x
P62753 RS6_HUMAN	40S ribosomal protein S6	Homo sapiens	28663	10.85		x	
Q02878 RL6_HUMAN	60S ribosomal protein L6	Homo sapiens	32708	10.59	X	x	x

Autres

O15042 SR140_HUMAN	U2-associated protein SR140	Homo sapiens	118219	8.59		x	
O43707 ACTN4_HUMAN	Alpha-actinin-4	Homo sapiens	104788	5.27	X	x	x
O75369 FLNB_HUMAN	Filamin-B	Homo sapiens	278021	5.49	X		
O95602 RPA1_HUMAN	DNA-directed RNA polymerase I subunit RPA1	Homo sapiens	194068	6.80	X	x	x
P02545 LMNA_HUMAN	Lamin-A/C	Homo sapiens	74095	6.57	X	x	
P04406 G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase	Homo sapiens	36030	8.57			x
P05455 LA_HUMAN	Lupus La protein	Homo sapiens	46808	6.68	X	x	x
P06748 NPM_HUMAN	Nucleophosmin	Homo sapiens	32555	4.64	X	x	x
P07355 ANXA2_HUMAN	Annexin A2	Homo sapiens	38580	7.57	X	x	x
P07900 HS90A_HUMAN	Heat shock protein HSP 90-alpha	Homo sapiens	84607	4.94			x
P08107 HSP71_HUMAN	Heat shock 70 kDa protein 1	Homo sapiens	70009	5.48	X	x	
P08238 HS90B_HUMAN	Heat shock protein HSP 90-beta	Homo sapiens	83212	4.97	X		
P09874 PARP1_HUMAN	Parp 1	Homo sapiens	113012	8.99	X	x	x
P0AEX9 MALE_ECOLI	Maltose-binding periplasmic protein precursor	Escherichia coli	43360	5.53	X	x	x

P11021 GRP78_HUMAN	78 kDa glucose-regulated protein precursor	Homo sapiens	72288	5.07	X	x	
P11142 HSP7C_HUMAN	Heat shock cognate 71 kDa protein	Homo sapiens	70854	5.37	X	x	x
P11387 TOP1_HUMAN	DNA topoisomerase 1	Homo sapiens	90669	9.33		x	x
P11388 TOP2A_HUMAN	DNA topoisomerase 2-alpha	Homo sapiens	174276	8.82	X		x
P13639 EF2_HUMAN	Elongation factor 2	Homo sapiens	95277	6.41	X	x	x
P14625 ENPL_HUMAN	Endoplasmin precursor	Homo sapiens	92411	4.76	X		
P15252 REF_HEVBR	Rubber elongation factor protein	Hevea brasiliensis	14713	5.04	X		
P19338 NUCL_HUMAN	Nucleolin	Homo sapiens	76568	4.60	X	x	x
P19525 E2AK2_HUMAN	Interferon-induced, double-stranded RNA-activated protein kinase	Homo sapiens	62056	8.58	X	x	x
P20029 GRP78_MOUSE	78 kDa glucose-regulated protein precursor	Mus musculus	72377	5.07			x
P21333 FLNA_HUMAN	Filamin-A	Homo sapiens	280564	5.70	X		
P23284 PPIB_HUMAN	Peptidyl-prolyl cis-trans isomerase B precursor	Homo sapiens	22728	9.33		x	
P23528 COF1_HUMAN	Cofilin-1	Homo sapiens	18491	8.22		x	x
P35250 RFC2_HUMAN	Replication factor C subunit 2	Homo sapiens	39132	6.04		x	x
P35251 RFC1_HUMAN	Replication factor C subunit 1	Homo sapiens	128175	9.38			x
P35606 COPB2_HUMAN	Coatomer subunit beta'	Homo sapiens	102422	5.15		x	
P35637 FUS_HUMAN	RNA-binding protein FUS	Homo sapiens	53394	9.40	X	x	x
P37108 SRP14_HUMAN	Signal recognition particle 14 kDa protein	Homo sapiens	14561	10.05		x	x
P40108 ALDH_CLAHE	Aldehyde dehydrogenase	Davidiella tassiana	53520	6.23	X		
P42166 LAP2A_HUMAN	Lamina-associated polypeptide 2 isoform alpha	Homo sapiens	75446	7.56	X	x	x
P42285 SK2L2_HUMAN	Superkiller viralicidic activity 2-like 2	Homo sapiens	117729	6.12	X	x	x
P42680 TEC_HUMAN	Tyrosine-protein kinase Tec	Homo sapiens	73534	8.69		x	
P46940 IQGA1_HUMAN	Ras GTPase-activating-like protein IQGAP1	Homo sapiens	189134	6.08	X	x	
P49458 SRP09_HUMAN	Signal recognition particle 9 kDa protein	Homo sapiens	10105	7.77		x	
P49916 DNL3_HUMAN	DNA ligase 3	Homo sapiens	102625	9.03			x
P51532 SMCA4_HUMAN	Probable global transcription activator SNF2L4	Homo sapiens	184530	7.83		x	x
P53621 COPA_HUMAN	Coatomer subunit alpha	Homo sapiens	138244	7.70	X		x
P53999 TCP4_HUMAN	Activated RNA polymerase II transcriptional coactivator p15	Homo sapiens	14386	9.60		x	x
P60709 ACTB_HUMAN	Actin, cytoplasmic 1	Homo sapiens	41710	5.29	X	x	x
P60842 IF4A1_HUMAN	Eukaryotic initiation factor 4A-I	Homo sapiens	46125	5.32		x	

P61981 1433G_HUMAN	14-3-3 protein gamma	Homo sapiens	28285	4.80			x
P62258 1433E_HUMAN	14-3-3 protein epsilon	Homo sapiens	29155	4.63		x	x
P63104 1433Z_HUMAN	14-3-3 protein zeta/delta	Homo sapiens	27728	4.73	X	x	x
P63261 ACTG_HUMAN	Actin, cytoplasmic 2	Homo sapiens	41766	5.31	X	x	x
P67809 YBOX1_HUMAN	Nuclease sensitive element-binding protein 1	Homo sapiens	35903	9.87	X	x	x
P68400 CSK21_HUMAN	Casein kinase II subunit alpha	Homo sapiens	45115	7.29		x	x
P69171 COAT_BPZR	Coat protein	Enterobacteria phage ZR	13719	8.67	X	x	x
P78527 PRKDC_HUMAN	DNA-dependent protein kinase catalytic subunit	Homo sapiens	468788	6.75	X	x	x
P82979 HCC1_HUMAN	Nuclear protein Hcc-1	Homo sapiens	23656	6.10			x
Q00610 CLH1_HUMAN	Clathrin heavy chain 1	Homo sapiens	191493	5.48	X	x	x
Q01082 SPTB2_HUMAN	Spectrin beta chain, brain 1	Homo sapiens	274439	5.39	X	x	
Q08211 DHX9_HUMAN	ATP-dependent RNA helicase A	Homo sapiens	140869	6.41	X		
Q08945 SSRP1_HUMAN	FACT complex subunit SSRP1	Homo sapiens	81024	6.45			x
Q12905 ILF2_HUMAN	Interleukin enhancer-binding factor 2	Homo sapiens	43035	5.19	X	x	x
Q12906 ILF3_HUMAN	Interleukin enhancer-binding factor 3	Homo sapiens	95279	8.86	X	x	x
Q13148 TADBP_HUMAN	TAR DNA-binding protein 43	Homo sapiens	44711	5.85		x	x
Q13283 G3BP1_HUMAN	Ras GTPase-activating protein-binding protein 1	Homo sapiens	52132	5.36	X	x	x
Q13813 SPTA2_HUMAN	Spectrin alpha chain, brain	Homo sapiens	284364	5.22	X		x
Q14152 IF3A_HUMAN	Eukaryotic translation initiation factor 3 subunit 10	Homo sapiens	166468	6.38	X	x	x
Q14498 RBM39_HUMAN	RNA-binding protein 39	Homo sapiens	59343	10.10		x	
Q15019 SEPT2_HUMAN	Septin-2	Homo sapiens	41461	6.15			x
Q15084 PDIA6_HUMAN	Protein disulfide-isomerase A6 precursor	Homo sapiens	48091	4.95			x
Q15233 NONO_HUMAN	Non-POU domain-containing octamer-binding protein	Homo sapiens	54197	9.01	X	x	x
Q52LJ0 FA98B_HUMAN	Protein FAM98B	Homo sapiens	37167	5.96		x	x
Q7Z2W4 ZCC2_HUMAN	Zinc finger CCCH type antiviral protein 1	Homo sapiens	101380	8.68	X		
Q86V81 THOC4_HUMAN	THO complex subunit 4	Homo sapiens	26872	11.15		x	x
Q8N5C6 SRBD1_HUMAN	S1 RNA-binding domain-containing protein 1	Homo sapiens	111705	8.94		x	x
Q92522 H1X_HUMAN	Histone H1x	Homo sapiens	22474	10.76		x	
Q96EP5 DAZP1_HUMAN	DAZ-associated protein 1	Homo sapiens	43356	8.73			x
Q9BQG0 MBB1A_HUMAN	Myb-binding protein 1A	Homo sapiens	148762	9.34	X	x	x

Q9BVC5 ASHWN_HUMAN	Ashwin	Homo sapiens	25843	9.76		x	
Q9H0A0 NAT10_HUMAN	N-acetyltransferase 10	Homo sapiens	115631	8.50		x	x
Q9H0D6 XRN2_HUMAN	5'-3' exoribonuclease 2	Homo sapiens	108513	7.26			x
Q9H9Y6 RPA2_HUMAN	DNA-directed RNA polymerase I subunit RPA2	Homo sapiens	128146	8.07	X		
Q9P2I0 CPSF2_HUMAN	Cleavage and polyadenylation specificity factor subunit 2	Homo sapiens	88431	4.98			x
Q9Y224 CN166_HUMAN	Protein C14orf166	Homo sapiens	28051	6.19	X	x	x
Q9Y262 IF36I_HUMAN	Eukaryotic translation initiation factor 3 subunit 6-interacting protein	Homo sapiens	66684	5.93	X		
Q9Y266 NUDC_HUMAN	Nuclear migration protein nudC	Homo sapiens	38219	5.27	X		
Q9Y3B4 PM14_HUMAN	Pre-mRNA branch site protein p14	Homo sapiens	14576	9.41		x	
Q9Y3I0 CV028_HUMAN	UPF0027 protein C22orf28	Homo sapiens	55175	6.77		x	x
Q9Y678 COPG_HUMAN	Coatomer subunit gamma	Homo sapiens	97655	5.32			x

CONCLUSIONS

ET

PERSPECTIVES

Au travers de la partie bibliographique de ce manuscrit nous avons abordé deux étapes de l'expression des gènes eucaryotes tout à fait primordiales pour le cycle de multiplication du virus HIV-1 :

- la transcription d'une part, où dans le cas du virus HIV-1, la protéine Tat et la séquence TAR de l'ARN de l'ARN viral, détournent le facteur P-TEFb, et avec lui la machinerie transcriptionnelle de la cellule hôte, de leurs cibles cellulaires, ceci afin de transcrire l'ensemble du génome viral en un grand nombre de copies.

- et l'épissage, qui dans le cas de ARN du virus HIV-1, est une étape très complexe par le grand nombre de combinaisons possibles des 6 sites donneurs et 9 sites accepteurs d'épissage, et surtout par la diversité des régulations qui gouverne leur utilisation.

En ce qui concerne le phénomène de transactivation par le complexe TAR/Tat/cycline T1, l'établissement de la structure tridimensionnelle de la protéine Tat et de ce complexe s'est avéré être très difficile mais représente toujours un très important défi dans la lutte contre le SIDA. Notre travail a permis de mettre au point des conditions optimales pour la production de l'ARN TAR et un fragment de la cycline T1. La structure tridimensionnelle de ce fragment de cycline T1 en fusion avec un fragment de la protéine Tat du virus équin a pu être établie par nos concurrents, mais ces données n'apportent toujours pas d'information sur la structure de la protéine Tat au sein de ce complexe. D'autres approches devront donc être utilisées pour essayer d'apporter des informations structurales sur le complexe TAR/Tat/cycline T1, soit la coproduction de tous les partenaires chez *E. coli,* soit l'utilisation d'approches autres que la radiocristallographie, comme par exemple la microscopie électronique.

La seconde partie de mon travail de thèse concernait l'épissage de l'ARN du virus HIV-1. Par deux approches différentes nous aboutissons à la conclusion que certaines des protéines du virus HIV-1 pourraient de manière directe ou indirecte avoir un effet sur les taux relatifs d'utilisation des sites d'épissage de l'ARN du virus HIV-1. Nous avons en effet observé un effet direct ou indirect, de la surexpression des protéines Tat et Nef, sur l'épissage d'un ARN contenant tous les sites d'épissage de l'ARN viral. La protéine virale NCp7 recombinante s'est avérée influencer directement le taux d'épissage in vitro aux sites A1 et A2, et ceci de manière opposée. Enfin, des mutations dans les régions codant les protéines Tat ou Vpr se sont avérées avoir des effets à longue distance sur les sites d'épissage qui pourraient être dus à une action directe ou indirecte de ces protéines au niveau de ces sites d'épissage. Cette notion d'un effet de l'expression des protéines virales sur les taux relatifs d'utilisation des différents

sites d'épissage de l'ARN du virus HIV-1, n'a pas encore été abordée par d'autres équipes. Elle doit être confirmée et son étude ouvre des perspectives importantes pour la suite des travaux sur la régulation de l'épissage de l'ARN du virus HIV-1.

Le travail approfondi que nous avons réalisé sur le site A3 a permis de définir quelles protéines SR interviennent au niveau de l'élément ESEt et aussi de comprendre le rôle très négatif sur l'utilisation du site A3 de la substitution $C_{5396}U$ en amont de l'élément ESEt.
Nos données prouvent en effet la fixation des protéines SR ASF/SF2 et 9G8 à un segment de l'ARN viral incluant la séquence GAGGAG (ESEt) et leur effet positif sur l'utilisation du site A3. Il est possible que la fixation de ces deux protéines empêche la fixation d'une protéine inhibitrice. Cette fixation pourrait au contraire être renforcée par la mutation $C_{5396}U$. En effet, l'identification des partenaires protéiques de l'ARN portant la mutation $C_{5396}U$ par l'approche protéomique, nous a permis de mettre en évidence la présence de la protéine DAZAP1 et des protéines hnRNP A1 et A2. Il restera à déterminer si la présence de ces protéines, en particulier de DAZAP1 décrite récemment comme un inhibiteur d'épissage, est responsable de l'effet très délétère de la mutation $C_{5396}U$ sur le taux d'utilisation du site A3. Notre observation va permettre d'établir une collaboration avec l'équipe de F. Baralle, faisant aussi parti du réseau Eurasnet, qui a découvert les propriétés inhibitrices de l'épissage de DAZAP1. Les démonstrations pourront être faites par emploi de la protéine recombinante et de la panoplie d'essais *in vitro* et *in cellulo* que nous avons mis au point, ainsi que par l'approche siRNA qui a été implantée au laboratoire par C. Aigueperse.
Ces compléments devraient nous permettre d'amener ces données à un très bon niveau de publication.

Alors que dans le cas du site A3 la mutation $C_{5396}U$ est nécessaire pour qu'en extrait nucléaire la protéine DAZAP1 soit fixée à l'ARN, en traitant des données de spectrométrie de masse obtenues sur des complexes formés avec un segment d'ARN contenant le site d'épissage A7 (thèse de V. Marchand en collaboration avec S. Sanglier), nous avons pu observer que cette protéine se fixe dans ce cas sans nécessité de mutation. L'analyse des complexes formés par différents fragments d'ARN permet d'identifier le site de fixation avec précision. Il s'agit de la structure tige-boucle SLS 3-A7 contenant l'élément inhibiteur ESS3. L'effet potentiellement inhibiteur de DAZAP1 au site A7 pourra donc aussi être étudié. Il faut aussi noter que V. Marchand et C. Aigueperse ont montré qu'à la fois ASF/SF2 et 9G8 activent

l'utilisation du site A7. Les sites A3 et A7 tous deux requis pour la production de l'ARNm de la protéine Tat semblent donc régulés par les mêmes facteurs.
Comme nous l'avons vu, la protéine hnRNP A1 est un inhibiteur fort du site A2 et son action négative est contrecarrée par la protéine ASF/SF2. Il est aussi possible que d'autres facteurs nucléaires interviennent dans la régulation de l'utilisation de ce site. L'emploi de l'approche de purification des complexes par ajout de la séquence étiquette MS2 dans l'ARN, suivie de l'analyse des protéines de ces complexes par spectrométrie de masse devraient permettre de répondre à cette question.

On voit au travers de notre travail combien l'étude des mécanismes d'épissage de l'ARN de HIV-1 est importante pour la compréhension de son cycle de multiplication. De telles études peuvent ouvrir des perspectives dans le domaine thérapeutique, comme en atteste les récents travaux du laboratoire en collaboration avec l'équipe de J. Tazi, qui ont permis de montrer qu'une molécule inhibant l'effet activateur d'ASF/SF2 limite la multiplication du virus (Bakkour *et al.*, 2007).

MATERIELS ET METHODES

MATERIELS & METHODES :

1. Les constructions génétiques

1.1. Les constructions pΔPSP

La construction pΔPSP, fournie par CM. Stoltzfus et utilisée en transfection, a été réalisée à partir du plasmide infectieux pNL4-3 (GenBankTM accession number M19921). La délétion réalisée entre les positions 1511 et 4551 et qui rend pΔPSP non-infectieux a été obtenu par digestion de pNL4-3 par les nucléases *Spe*I et *Bal*I, puis religation avec les oligonucleotides 5'-CTAGACGCGTTTGG-3' et 5'-CCAAACGCGT-3'.

La longueur du plasmide rendant compliqué toute tentative de mutagénèse, deux plasmides intermédiaires de plus petite taille et contenant les séquences à muter, ont été crée. Il s'agit de :

- la construction pCR-D4A7 réalisée à partir du vecteur pCR 2.1 (Invitrogen) dans lequel le fragment *Eco*RI/*Bam*HI de pΔPSP (5743-8465) a été cloné aux sites *Eco*RI et *Bam*HI,
- et la construction pG-D1D3 réalisée à partir du vecteur pGex 2TK (GeLifeSciences) dans lequel le fragment *Bss*HII/*Eco*RI de pΔPSP (711-5743) a été cloné aux sites *Bss*HII et *Eco*RI.

Les différents mutants envisagés ont été effectués sur l'une de ces deux constructions puis les fragments mutés ont été re-clonés dans la construction pΔPSP digérée par les enzymes appropriées. Chacun de ces mutants a été séquencé sur l'ensemble du fragment re-cloné.

Les mutations réalisées et les séquences des oligonucléotides utilisés pour réaliser ces constructions sont indiquées dans le tableau suivant.

Construction	Matrice	Oligo 1 Oligo 2
pΔPSP ΔESEt	pCR-D4A7	4462 5′-GGCGTTACTCGACAAGCAAGAAATGGAGCC-3′ 4463 5′-GGCTCCATTTCTTGCTTGTCGAGTAACGCC-3′
pΔPSP Caputi	pCR-D4A7	3186 5′-GCCAGTAGATCGTAAGTACGCGCCCTGGAAGCA-3′ 3187 5′-TGCTTCCAGGGCGCGTACTTACGATCTACTGGC-3′
pΔPSP 4050	pCR-D4A7	4050 5′-AAATGGAGCCAGATGATCCTAGACTA-3′ 4051 5′-TAGTCTAGGATCATCTGGCTCCATTT-3′
pΔPSP 4054	pCR-D4A7	4054 5′-GGAGCCAGTAGAAGCTAGACTAGAGC-3′ 4055 5′-GCTCTAGTCTAGCTTCTACTGGCTCC-3′
pΔPSP 4058	pCR-D4A7	4058 5′-CCAGTAGATCCTTCACTAGAGCCCTG-3′ 4059 5′-CAGGGCTCTAGTGAAGGATCTACTGG-3′

pΔPSP 4064	pCR-D4A7	4064	5′-GATCCTAGACTACTGCCCTGGAAGCA-3′
		4065	5′-TGCTTCCAGGGCAGTAGTCTAGGATC-3′
pΔPSP ΔESEt	pCR-D4A7	5207	5′-GGCGTTACTCGATAGAGGAGAGCAA-3′
		5208	5′-TTGCTCTCCTCTATCGAGTAACGCC-3′
pΔPSP C→U	pCR-D4A7	5209	5′-GGCGTTACTCGATAAGCAAGAAATGGAGCCAG-3′
		5210	5′-GGCTCCATTTCTTGCTTATCGAGTAACGCCTATTC-3′
pΔPSP Dble	pCR-D4A7	5211	5′-GCATCCAGGAAGCGTACGCTAAACTGCTTGTACC-3′
		5212	5′-GGTACAAGCAGTTTAGCGTACGCTTCCTGGATGC-3′
pΔPSP ΔBcl ap	pCR-D4A7	3199	5′-CAGTAGATCCTAGACTAGAGCCGAAGTCAGCCTAAAACTGCTTGT-3′
		3200	5′-ACAAGCAGTTTTAGGCTGACTTCGGCTCTAGTCTAGGATCTACTGC-3′
pΔPSP ΔLoopII	pCR-D4A7	4587	5′-GCTTGTACCACTTGCTTGTTGCTTTCATTG-3′
		4588	5′-CAATGAAAGCAACAAGCAAGTGGTACAAGC-3′
pΔPSP 5288	pG-D1D3	5288	5′-TGTAGAGATCCTACCTTG-3′
		5289	5′-TAAAACCAAAACAGATAAAG-3′
pΔPSP 5290	pG-D1D3	5290	5′-AGTACTTGGCACTAGTAAAACCAAAACAGATAAAG-3′
		5291	5′-CTGTTTTGGTTTTACTAGTGCCAAGTACTGTAG-3′

1.2. Les vecteurs d'expression eucaryote

Les constructions pTRE-Tat-Flag, -Nef, -Flag-hnRNP H' et -Flag-hnRNP F, utilisés en transfection, sont toutes issues du vecteur d'expression pTRE2 (Clontech). Les origines et mode de clonage des ADNc des protéines exprimées sont divers:

- l'ADNc $Tat_{(101)}$-Flag a été amplifié à partir du plasmide pEV280-Tat-Flag fourni par M. Benkirane, puis cloné dans pTRE2 aux sites *Bam*HI/*Xba*I,
- l'ADNc de Nef a été amplifié à partir de la construction pΔPSP, puis cloné dans pTRE2 aux sites *Bam*HI/*Xba*I,
- l'ADNc des hnRNP H' et F ont été amplifié à partir de cDNA totaux de cellule HeLa, puis cloné dans pcDNA-Flag fourni par S. Massenet, et enfin la fusion Flag-hnRNP a été inséré aux sites *Hind*III /*Xba*I.

Les oligonucléotides utilisés pour réaliser ces constructions sont indiqués dans le tableau qui suit.

Les autres plasmides d'expression utilisés en transfection, pTRE-GGT d'une part et pXJ-ASF, -9G8, -SC35 et -SRp40 d'autre part, ont été respectivement fourni par J. Stevenin et A. Visvikis.

Construction	Matrice	Oligo 1 Oligo 2
pTRE-Tat-Flag	pEV-Tat-	3442 5′-CGGGATCCATGGAGCCAGTAGATCC-3′
	Flag	3443 5′-GCTCTAGACTACTTGTCATCGTCC-3′
pTRE-Nef	pΔPSP	3574 5′-CGGGATCCATGGGTGGCAAGTGGTC-3′
		3575 5′-GCTCTAGATCAGCAGTTCTTGAAGTAC-3′
pTRE-Flag-H’	$cDNA_{tot}$	3887 5′-GGATCCCTGAGCACGGAAGGCAGG-3′
	HeLa	3888 5′-TCTAGACTAAGCAAGGTTTGACTGATAGTC-3′
pTRE-Flag-F	$cDNA_{tot}$	3883 5′-GGATCCCTGGGCCCTGAGGGAGGT-3′
	HeLa	3884 5′-TCTAGACTAGTCATAGCCACCCATGCT-3′

1.3. Les constructions utilisées en épissage *in vitro*

Les constructions dérivent du plasmide pBRU3, qui nous a été fourni par F. Clavel (Institut Pasteur). Il contient la séquence entière du génome du virus HIV-1 de la souche HIV-1/BRU (GenBankTM accession number K02013), 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*. Les différentes constructions utilisées au cours de ce travail, ainsi que les oligonucléotides permettant d’obtenir les mutants, sont présentés dans la Fig. 113.

1.4. La construction pRTR

Le gène T7-Rib1-TAR-Rib2 a été obtenu par plusieurs PCR successives effectuées avec les 6 oligonucléotides décrits dans la Fig. 114. Le premier jeu de PCR se fait entre les couples $^{2895}/_{2898}$, $^{2896}/_{2899}$ et $^{2897}/_{2900}$, il donne respectivement les fragments A, B et C. La deuxième PCR est réalisée entre les fragments A et B en présence des oligonucléotides 2895 et 2899, et donne le fragment AB. Enfin la dernière PCR est effectuée entre les fragments AB et C en présence des oligonucléotides 2895 et 2900 pour obtenir le gène T7-Rib1-TAR-Rib2. Le premier jeu de PCR se fait dans les conditions suivantes : 1 µg de chaque oligonucléotides ; tampon Pfu 10X (Promega) ; 16 µl dNTP (1,25mM chacun) ; 1 µl Pfu ADN polymérase (2U/µl) (Promega) ; H2O qsp 50 µl. Après une dénaturation 2 min à 94°C, l’ADN est amplifié pendant 30 cycles de 30 sec à 94°C, 30 sec à 42°C, 30 sec à 74°C. Les deux autres PCR se

font dans des conditions différentes : 20 µl de chaque produit PCR ; 1 µg de chaque oligonucléotides ; tampon Pfu 10X (Promega) ; 1 µl dNTP (1,25mM chacun) ; 0,5 µl Pfu ADN polymérase (2U/µl) (Promega) ; H2O qsp 50 µl. Après une dénaturation 5 min à 94°C, l'ADN est amplifié pendant 30 cycles de 30 sec à 94°C, 30 sec à 55°C, 30 sec à 74°C. Le gène et le vecteur pUC 19 (Fermentas) sont finalement digéré par *Pst*I et *Bam*HI, puis ligué pour obtenir la construction pRTR.

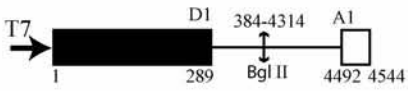
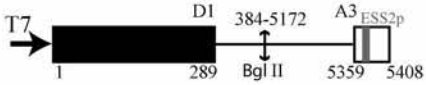
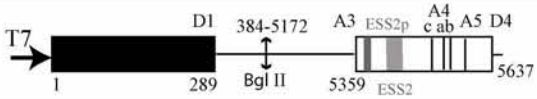
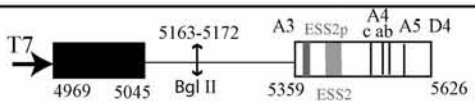
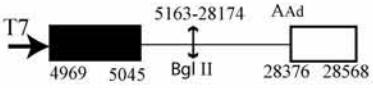
Nom	Schéma de la construction	Linéarisation
C1		EcoR I
C2		EcoR I
C3		Pst I
L3U1		Pst I
102		Pst I
106		Pst I
112		Pst I

Fig. 113 : Constructions génétiques utilisées pour l'épissage in vitro

Les schémas représentent les fragments d'ADNc insérés dans le vecteur de clonage pBluescript KSII+ : les exons sont représentés par des rectangles et les introns, par des traits. Les positions des fragments d'ADNc au sein de l'ADNc proviral entier sont indiquées. La double flèche verticale représente le site de restriction BglII utilisé pour réaliser la jonction des fragments d'ADNc. Les sites uniques de restriction utilisés pour produire les transcrits sont indiqués, pour chaque construction, dans la colonne "linéarisation".

A)

Nom	Séquence 5' - 3'	Descriptif
2895	AA*CTGCAG***TCTAATACGACTCACTATA**GGG	Site *Pst*I + séquence T7 (en gras)
2896	CTGATGAGTCCGTGAGGACGAAACGGTACCCGGTACCGT	Séquence Rib 1 conservée
2897	GACGGAGTCTAGACTCCGTCCTGATGAGTCCGTGAGGAC	Séquence Rib 2 conservée
2898	CTCATCAGGGCAGATCTGACCCTATAG	Séquence Rib1dépendante de TAR
2899	ACTCCGTCAAGGCAGAGAGCTCCCAGGCTCAGATCTGCCGACGGTACC	Séquence TAR (en rouge)
2900	CGGGATCCCTGCCTTTCGTCCTCAC	Séquence Rib2 dépendante de TAR + site *Bam*HI

B)

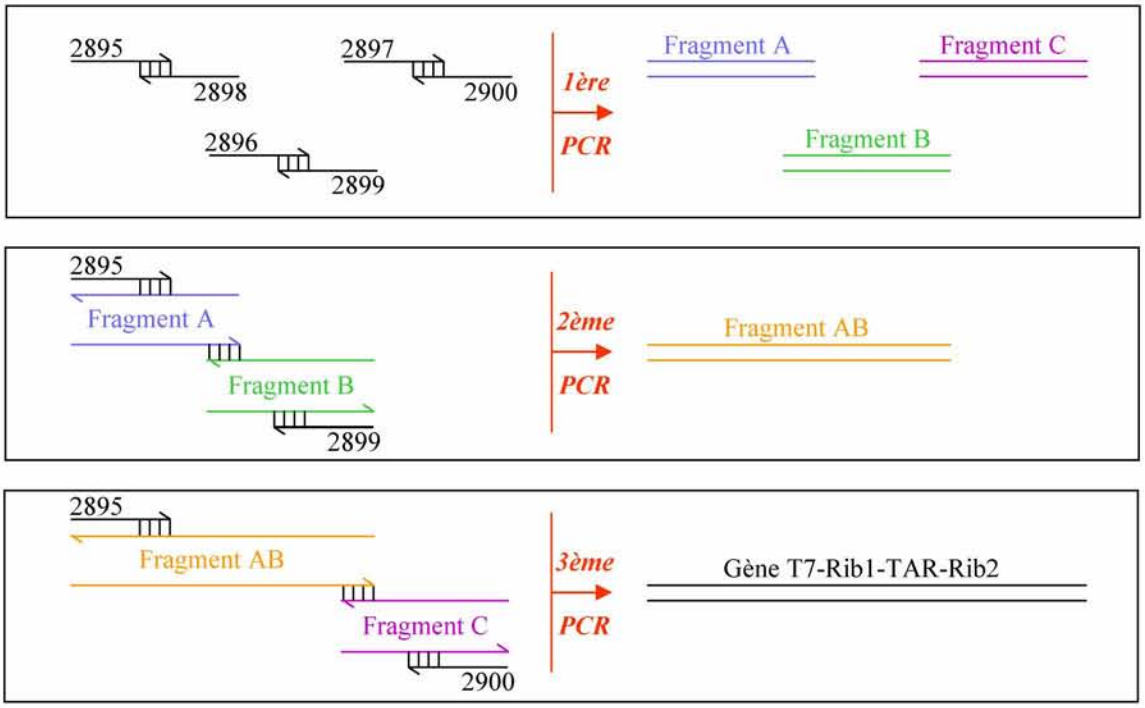


Fig. 114 : Construction du gène T7-Rib1-TAR-Rib2.

A) Tableau des oligonucléotides utilisés. Les sites de restriction sont indiqués en italiques. Les séquences soulignées correspondent aux séquences chevauchantes. B) Stratégie utilisée pour l'obtention du gène T7-Rib1-TAR-Rib2. Les oligonucléotides 2895 et 2899 ajoutés pour la PCR 2 permettent de reformer des brins 5'→3' du fragment A et des brins 3'→5' du fragment B, et ainsi favoriser la formation du fragment AB (technique rééditée pour la 3ème PCR).

1.5. Les vecteurs d'expression codant la cycline T1

Le plasmide pG-T1(1-300) nous a été fourni par M. Benkirane. Il a servi de plasmide d'expression du fragment 1-300 de la cycline T1, mais aussi de matrice pour réaliser les autres constructions. Ainsi l'utilisation des oligonucléotides $^{3399}/_{3400}$ à permis d'amplifier l'ADNc du fragment 1-280 de la cycline T1 et de l'insérer dans le vecteur pGex 2TK (GeLifeSciences) aux sites *Bam*HI et *Eco*RI pour obtenir la construction pG-T1(1-280). La fusion pG-T1-Tat a été produite par clonages successifs du fragment *Bam*HI/*Eco*RI de pG-T1(1-280), et du fragment *Eco*RI/*Xho*I de la protéine Tat 101 (amplifié par les oligonucléotides $^{3401}/_{3628}$ à partir du plasmide pEV280-Tat-Flag), dans le vecteur pGex 4T1 (GeLifeSciences). Enfin la construction pET-T1(1-280) a été obtenu par amplification de l'ADNc du fragment 1-280 de la cycline T1 par le oligonucléotides $^{3140}/_{3400}$, puis clonage de ce dernier dans le vecteur pET 28b (Novagen) aux sites *Nde*I/*Eco*RI. Les séquences des oligonucléotides utilisés pour réaliser ces constructions ainsi que les schémas de ces constructions sont présentés dans la Fig. 115.

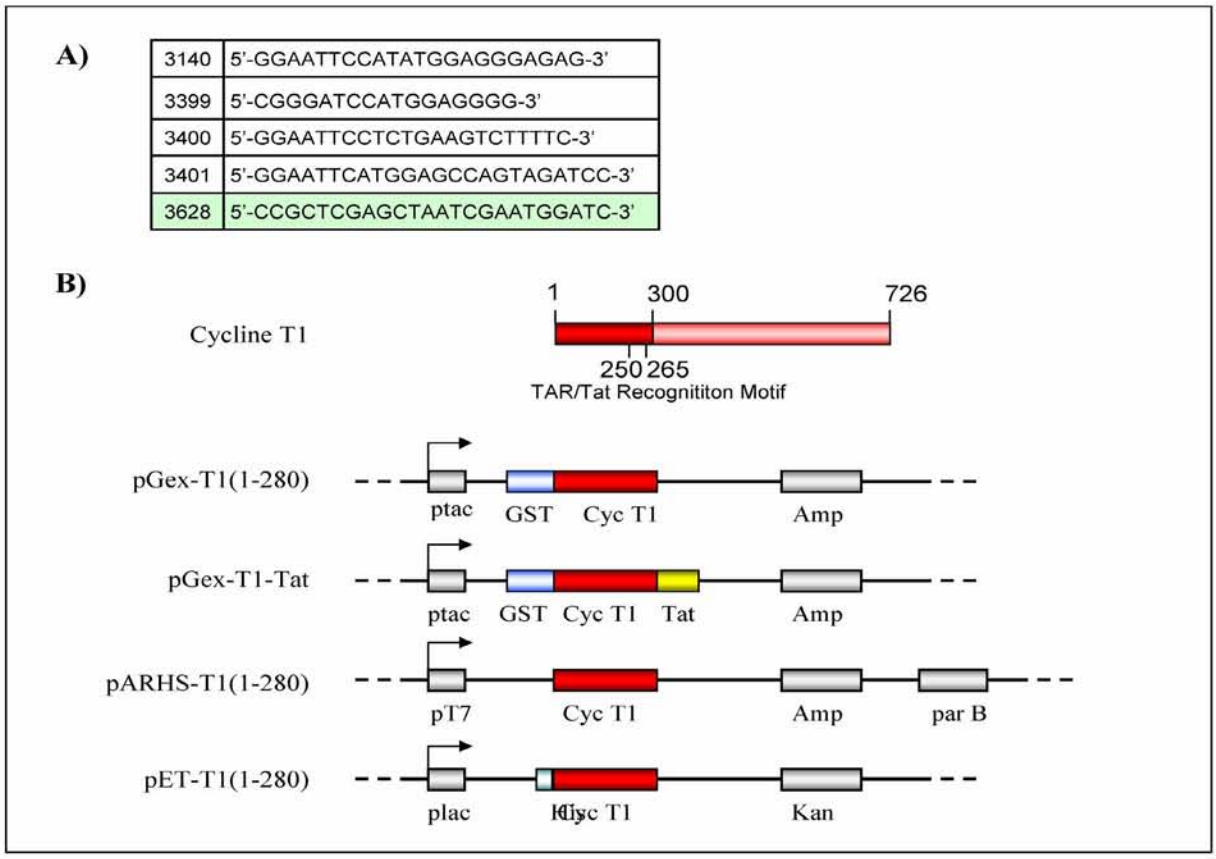


Fig. 115 : Les vecteurs d'expression utilisés pour exprimer la cycline T1.

A) Tableau des oligonucléotides utilisés pour les clonages. B) Représentation schématique des vecteurs d'expression utilisés pour la production de la cycline T1.

1.6. Les constructions MS2

La séquence MS2, issue du plasmide pAdML-M3, contient les 3 tiges-boucles de fixation de la protéine MS2. La construction pUC-MS2 a été obtenue par clonage de cette séquence entre les sites *Eco*RI et *Bam*HI du vecteur pUC18 (Fermentas). La construction pUC-2/3-MS2 a été obtenue par L Ayadi en insérant contenant la séquence (5339-5514 souche Bru) correspondant à l'ARN SLS 2/3 au site *Xba*I. Les constructions pUC-2/Ea-MS2 et pUC-2/Ec-MS2 ont été obtenues par mutagénèse de pUC-2/3-MS2 avec les couples d'oligonucléotides $^{5658}/_{5662}$ et $^{5660}/_{5662}$ respectivement (5658 : 5'-GTTGCTCTCCTCTGTTG-3', 5660 : 5'-GTTGCTCTCCTCTATTG-3' et 5662 : 5'-TCTAGAGGATCCGATATC-3').

2. Analyse du profil d'épissage *in vivo*

2.1. Transfection et extraction ARN

Les cellules HeLa ($2X10^5$ cellules par puit) sont transfectées dans des plaques 12 puits avec 1,33 µg de plasmide pΔPSP et 0,66 µg de plasmide pXJ surexprimant les protéines SR, 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_2PO_4 1,8mM ; NaCl 137mM ; Na_2HPO_4 10,1mM). Les ARN totaux cellulaires sont finalement extraits des cellules HeLa au Trizol selon les spécifications du fournisseur (Invitrogen).

2.2. 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_2O, puis dénaturé 5min à 65°C et renaturé dans la glace ; 0,5µl d'oligonucléotide (100µM) ; Tampon MuMLV 5X (Invitrogen) ; 0,25µl DTT (1M) ; 1,25µl dNTP (10mM chacun) ; 0,5µl MuMLV-RT (200U/µl) (Invitrogen) ; H2O 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 $MgCl_2$ (25mM) ; 0,5µl dNTP (10 mM chacun) ; 0,125µl GoTaq Flexi ADN polymérase (5U/µl) (Promega) ; H2O 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 au données précédemment obtenues (Ropers *et al.*, 2004).

3. Analyse du taux de transcrits des protéines 9G8, SC35, ASF/SF2, hnRNP A1, F, H et H'

Pour cette analyse, les ARN totaux de cellules HeLa sont extraits et rétro transcrits de la même façon que précédemment (cf § 2.1 Transfection et extraction ARN et 2.2 Analyse par RT-PCR). Les oligonucléotides permettant de cibler les messagers des protéines 9G8, SC35, ASF/SF2, hnRNP A1, F, H et H' sont décrit dans le tableau qui suit :

Nom	Séquence	Utilisation
3867	GGCTACTAGGTGGCATAGTG	Oligo 5' pour analyse ARNm de 9G8
3868	TGTTGTGGTTGGGAACTGCA	Oligo 3' pour analyse ARNm de 9G8
3869	TGTAAACGGAGGACTTGGGG	Oligo 5' pour analyse ARNm de SC35
3870	CCTTAAAAGTCAGGAGGCCAC	Oligo 3' pour analyse ARNm de SC35
3871	GGTGGGTAGCAATGCCAGTA	Oligo 5' pour analyse ARNm de ASF/SF2
3872	ATCCAGTGAGCCCTCTCCAA	Oligo 3' pour analyse ARNm de ASF/SF2
3873	GGAGGTGGTGGAAGCTACAA	Oligo 5' pour analyse ARNm de hnRNP A1
3874	CCTGTAGCTTCCCTGTCACT	Oligo 3' pour analyse ARNm de hnRNP A1
3875	GCCTGGAGAGCCAGTCAGTG	Oligo 5' pour analyse ARNm de hnRNP F
3876	GCACGTCTCCCCAGTGTGTT	Oligo 3' pour analyse ARNm de hnRNP F
3877	GGGAGGCATGGGCTTGTCAA	Oligo 5' pour analyse ARNm de hnRNP H1
3878	CCACTACTGTAGTAGCTGCT	Oligo 3' pour analyse ARNm de hnRNP H1
3879	GGCTTATCCAACCAGTCTAG	Oligo 5' pour analyse ARNm de hnRNP H2
3880	ACTAAACATCCCTCCCATTC	Oligo 3' pour analyse ARNm de hnRNP H2

4. Techniques d'analyse de retard sur gel

4.1. Préparation des matrices de transcriptions

Un produit PCR, servant de matrice pour la réaction de transcription *in vitro*, est synthétisé en utilisant une amorce introduisant un promoteur pour l'ARN polymérase du phage SP6, dans

les conditions suivantes : Plasmide 20 ng Tampon 10X (; 20 mM MgCl2 ;100 mM KCl ; 60 mM (NH4)2SO4 ; 1% Triton X100 ; 100 ml/mg BSA, Tris-HCl 200 mM pH=8,0) 5 µl Amorce 5' portant le promoteur SP6 100 ng Amorce 3' 100 ng Mélange des 4 dNTP (1,25 mM chacun) 8 µl DynazymeTM II ADN polymérase (2 U/µl) (Finnzyme) 0,5 µl H2O qsp 50 µl Après une dénaturation de 2 minutes à 94°C, l'ADN est amplifié pendant 30 cycles de 30 secondes à 94°C, 30 secondes à 45°C, 1 minute à 74°C. 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.

4.2. Préparation des ARN

4.2.1. Synthèse de transcrits radioactifs pour les expériences de retard sur gel.

La transcription est réalisée pendant 2 heures à 37°C dans le mélange réactionnel suivant : 1 µl de matrice (500 ng/µl), 2 µl tampon de transcription US 5X (Hepes 0,25 M pH 7,5 ; $MgCl_2$ 0,1 M ; DTT 33 mM), 0,5 µl DTT (0,2M), 0,25 µl RNasin 31 U/µl (Amersham biosciences), 0,5 µl BSA (2mg/ml), 1 µl mélange de rNTP (ATP, CTP, GTP (5/5/2 mM) (Fermentas)), 1 µl UTP (50µM) (Fermentas), 0,9 µl α-32P UTP (800 Ci/mmol ; 10 mCi/ml) (ICN), 0,5 µl ARN polymérase SP6 (20 U/µl) (MBI Fermentas), H2O qsp 10 µl. Après incubation, la matrice est digérée avec 1 µl de DNase (10 U/µl) pendant 30 min à 37°C. L'ARN est ensuite extrait au phénol/chloroforme et au chloroforme/alcool isoamylique (24/1), puis précipité 2 fois en présence de 2,5 M d'acétate d'ammonium et en présence de Tris NaCl. L'ARN est finalement lavé, séché, repris dans de l'eau stérile et la radioactivité est quantifiée grâce au compteur à scintillation (Beckman).

4.2.2. Synthèse de transcrits froids pour les expériences de retard sur gel.

La transcription est réalisée pendant 3 heures à 37°C dans le mélange réactionnel suivant : tampon de transcription TB 5X (Hepes KOH pH 7,5 0,4 M, MgCl2 120 mM; Spermidine 10 mM; DTT 0,2M), 0,25 µl RNasin 31 U/µl (Amersham biosciences), avec 4 µl de matrice, 3 µl d'un mélange de rNTP (25 mM) (Fermentas), 1 µl d'ARN polymérase SP6 (20 U/µl) (MBI Fermentas), H2O qsp 10 µl. Après incubation, la matrice est digérée avec 1 µl de DNase (10 U/µl) pendant 30 min à 37°C. L'ARN est ensuite extrait au phénol/chloroforme et au chloroforme/alcool isoamylique (24/1), puis précipité 2 fois en présence de 2,5 M d'acétate d'ammonium. L'ARN est ensuite fractionné sur gel acryl-urée 10%. L'ARN est ensuite élué

de la bande d'intérêt en présence de 250 µl de tampon d'élution (NaCl 250 mM ; EDTA 10 mM ; Tris-HCl 20 mM pH 7,5) pendant une nuit à 4°C. L'ARN est enfin précipité à l'éthanol en présence d'acétate de sodium (0,3 M final).

4.3. 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.

-en présence d'ARN chaud

Après dénaturation 5 min à 65°C et renaturation 10 min à température ambiante, 10 fmol d'ARN uniformément marqué dans un volume de 3 µl sont incubés en présence de 3 µl de tampon B 3X (Hepes pH 7,9 30mM ; KCl 600mM ; NaCl 60mM ; NP 40 (Igepal) 0,3% ; Glycérol 30% ; DTT 3mM), 1 µl d'ARNt de levure (1 µg/µl), 0,2 µl de RNasin 31 U/µl (Amersham biosciences), H2O qsp 6 µl. Puis les gammes de protéines (de l'ordre de 0 à 5 pmol dans 2 µl) sont ajoutées. Après 20 min d'incubation à température ambiante, 5 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 (38 : 2) à 6% 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.

-en présence d'ARN froid et d'ARN chaud

25 fmol d'ARN froid et 100 cps d'ARN uniformément marqué (représentant moins d'une fmol) sont mélangés dans 0,5 µl d'H_2O. Après dénaturation 5 min à 65°C et renaturation 10 min à température ambiante, les ARN sont incubés en présence de 0,6 µl de tampon B 3X (Hepes pH 7,9 30mM ; KCl 600mM ; NaCl 60mM ; NP 40 (Igepal) 0,3% ; Glycérol 30% ; DTT 3mM), 0,1 µl d'ARNt de levure (2,5 µg/µl), 0,05 µl de RNasin 31 U/µl (Amersham biosciences), H2O qsp 1,25 µl. Puis les gammes de protéines (de l'ordre de 0 à 15 pmol dans un volume de 0,5 µl) sont ajoutées. Après 20 min d'incubation à température ambiante, 1 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 (38 : 2) à 6% 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.

5. Méthodes spécifiques à l'étude de l'épissage *in vitro*

5.1. Préparation de la matrice de transcription

Afin de limiter la transcription à la région étudiée, le plasmide matrice est préalablement hydrolysé par une enzyme de restriction adéquate (Fig. 113). L'ADN est ensuite soumis à deux extractions, l'une au phénol/chloroforme (1/1) et la seconde par un mélange chloroforme/alcool isoamylique (24/1). Puis l'ADN est précipité à l'éthanol en présence d'acétate de sodium 0,3 M et finalement repris dans l'eau.

5.2. Transcription *in vitro*

La transcription est réalisée par l'ARN polymérase du phage T7 en présence d'un ribonucléotide marqué au 32P. L'ADN linéarisé (250 ng) est transcrit dans le milieu réactionnel suivant : Tampon US 5X (MgCl2 30 mM; Spermidine 5mM; Tris-HCl 0,2 M) ; 2 µl DTT (0,2M) ; 0,5 µl RNasin (31U/µL) (Amersham Biosciences) ; 0,25 µl rNTP (ATP 5 mM, CTP 5 mM, GTP 2 mM) ; 1 µl UTP (0,5 mM) ; 1 µl [α-32]P UTP (800 Ci/mmol ; 10 mCi/ml) ; 0,9 µl Coiffe m7GpppG (10mM) (NEB) ; 1,25 µl ARN polymérase T7 (61 U/µl) (USB) ; 0,5 µl BSA (2mg/ml) (Sigma) ; H2O qsp 10 µl. Les différences de concentration en GTP et UTP par rapport aux autres rNTPs, favorisent respectivement l'incorporation de la coiffe m7GpppG à l'extrémité 5' du transcrit, et l'incorporation de [α-32]P UTP uniformément dans le transcrit. Le mélange est incubé 1 h à 37°C, puis la matrice est 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). 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).

5.3. Epissage *in vitro*

Les expériences d'épissage *in vitro* sont réalisées avec des extraits nucléaires de cellules HeLa provenant de la société CilBiotech ou des extrait cytoplasmique S100 fourni par J. Stévenin. Ces extraits peuvent être complémentés par les protéines SR recombinantes issues du système bacculovirus, également fournies par J. Stévenin (Cavaloc *et al.*, 1999).

Chaque réaction d'épissage *in vitro* est réalisée dans les conditions suivantes :

1 µl de transcrit radioactif (15fmol/µl) ; 1 µl $MgCl_2$ (62,5mM) ; 1 µl ATP (25mM) ; 1 µl Créatine phosphate (625mM) ; 0,4 µl RNasin (15U/µl) (Amersham Biosciences) ; 6 µl Alcool polyvinylique 13% ; 3,6 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%) complémenté ou non en protéines (400 ng de protéines SR recombinantes par exemple) et 8 µl d'extrait nucléaire, préalablement traité 10 minutes à 30°C afin d'éliminer l'ATP endogène. Le mélange est placé 5 minutes dans la glace, puis incubé 2h30 à 30°C.

La réaction d'épissage est ensuite arrêtée par ajout de 20 µg de protéinase K (Roche) fonctionnant dans le tampon suivant : 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 (1/1), 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 formaldéhyde (EDTA 20 mM, bleu de bromophénol 0,05%, bleu de xylène cyanol 0,05%, dans de la formaldéhyde désionisée). Les produits d'épissage sont finalement séparés par électrophorèse sur gel polyacrylamide dénaturant (acryl-bisacrylamide (20:1) 5% ; TBE 1x ; urée 8 M). Puis le gel est séché et analysé par le système d'images Typhoon (GeLifeSciences).

5.4. Détermination de l'efficacité d'épissage

La radioactivité contenue dans les bandes d'intérêt est quantifiée à l'aide du logiciel Image Quant (version 5.2), puis les quantités correspondantes sont calculées en fonction du nombre de résidus uridine contenus dans chaque type d'ARN. Pour chaque épissage sont déterminées, la quantité de précurseur (P), celle du (ou des) ARN messager(s) (M_1, M_2, …, M_n) et celle du bruit de fond (T). Le pourcentage d'épissage est exprimée par la formule $(M-T)/[P+M_{total}-$

(n+1)T]x100. Les expériences ont été répétées au moins deux fois afin de vérifier leur reproductibilité.

6. Production et purification de l'ARN TAR en grande quantité

La construction pRTR linéarisée par *Pst*I (250 µg) est transcrit dans le tampon θ 10X (Tris-HCl pH 8 0,4M ; $MgCl_2$ 0,3M ; Spermidine 23mM), en présence de 50 µl DTT (1M), 1 ml d'un mélange rNTP respectant les ratio A/C/G/U de l'ARN Rib1-TAR-Rib2 (6,6mg ATP; 7,9mg CTP; 10,5mg GTP; 6,5mg UTP soit environ 20mM chacun), 10 µl Triton X100 (10%), 140 µl ARN polymérase T7 produite au laboratoire, dans un volume de 10 ml. Le mélange est incubé une nuit à 37°C, puis centrifugé à 5000 rpm pendant 5 min à 4°C. Les ARN sont ensuite précipités dans de l'éthanol absolu en présence de 500 µl NaCl (4M). Puis le précipité est lavé à l'éthanol 70%, séché et repris dans 1 ml de bleu de formaldéhyde (EDTA 20 mM, bleu de bromophénol 0,05%, bleu de xylène cyanol 0,05%, dans de la formaldéhyde désionisée). La séparation des différents produits de la réaction transcription/clivage s'effectue par électrophorèse sur gel de polyacrylamide dénaturant (polyacrylamide 16%, urée 8 M, TBE 0,5x). L'ARN TAR est enfin électro-élué pendant 4h dans un volume final de 500 µl.

7. Détermination de la structure secondaire de l'ARN TAR

7.1. Marquage de l'ARN TAR en 5'

L'ARN TAR délivré par clivage des ribozymes contient une extrémité 5'-OH qui permet son marquage en 5' en présence d'ATP [□-32P] par action de la polynucléotide kinase du bactériophage T4. Le mélange réactionnel utilisé pour la réaction de phosphorylation est le suivant : 400 ng d'ARN TAR, 1 µl de tampon kinase 10x (MgCl2 0,1 M; DTT 0,05 M; Tris-HCl 0,7 M pH 7,6), 1 µl de Polynucléotide kinase (10 U/µl) (Biolabs), 2 µl de [γ-32P] ATP (10 mCi/ml, 3000 Ci/mmol, Amersham), 0,5 µl de RNasine (20 U/µl) (Pharmacia) et H2O qsp 10 µl. Après une incubation de 1 h à 37°C, le mélange est chauffé pendant 2 min à 95°C et déposé sur gel polyacrylamide dénaturant (polyacrylamide 10%, urée 8M, TBE 1x). L'ARN est ensuite identifié par autoradiographie, puis la bande d'intérêt et découpée et éluée en présence de 250 µl de tampon d'élution (NaCl 250 mM ; EDTA 10 mM ; Tris-HCl 20 mM pH 7,5) pendant une nuit à 4°C. L'ARN est enfin précipité à l'éthanol en présence d'acétate de sodium (0,3 M final). Après centrifugation, le précipité est lavé à l'éthanol à 70 %, séché.

La radioactivité est quantifiée grâce au compteur à scintillation (Beckman) afin de dissoudre le précipité dans de l'eau stérile à raison de 20 000 cps/µl d'ARN.

7.2. Hydrolyses enzymatiques

L'ARN marqué en 5' (20 000 cps) est préincubé en présence de 2 µg d'ARNt, 2 µl de tampon tris 10X (Tris-HCl 0,5 M pH 7,6 ; MgCl2 0,1 M ; KCl 1,5 M) et H2O qsp 20 µl pendant 5 min à 65°C puis 20 min à température ambiante.Les hydrolyses enzymatiques sont ensuit réalisées par ajout des RNases V1, T1 ou T2 (gamme de quantité + tampon tris 1X qsp 4 µl) et incubation 12 min à 30°C. La réaction est stoppée par ajout de 175 µl H2O et 1 µl d'ARNt (20µg/µl), suivis d'une extraction phénolique. Après précipitation à l'éthanol en présence d'acétate de sodium 0,3M et lavage à l'éthanol à 70%, les précipités d'ARN sont repris dans 5 µl de bleu de dépôt dénaturant (EDTA 20 mM ; bleu de bromophénol 0,05% ; bleu de xylène cyanol) et déposé sur gel polyacrylamide dénaturant (polyacrylamide 10%, urée 8M, TBE 1x).

8. Technique d'empreinte

1,5 µg d'ARN froid sont incubés en présence de 3,6 µl de tampon D (EDTA 0,2 mM; KCl 0,1 M; DTT 0,5 mM; PMSF 0,25 mM; HEPES 20 mM pH 7,9 ; glycérol 20%), 0,25 µl d'ARNt de levure (20 µg/µl), 0,2 µl de RNasin 31 U/µl (Amersham biosciences), H2O qsp 13 µl. Après dénaturation 5 min à 65°C et renaturation 10 min à température ambiante, 1 µl e $MgCl_2$ (62,5 mM) et 8 µl de protéine (17 à 232 pmol de protéine 9G8) sont ajoutés, puis le tout est incubé 5 min dans la glace et 5 min 30°C.

La RNase T1 est ensuite ajoutée (1 µl à 0,5 U/µl) pendant 10 min à 30°C, puis la réaction est stoppée par ajout de 1 µl dARNt et H_2O qsp 200 µl. Les protéines sont ensuite extraites au phénol/chloroforme (1/1), puis au chloroforme/alcool isoamylique (24/1) et les ARN sont précipités en présence d'éthanol absolu sans ajout de sels. Le précipité est repris dans 4 µl d'H_2O.

Les coupures sot ensuite analysées par élongation d'amorce comme décrit ci-dessous. Un oligonucléotide marqué en 5' est hybridé à l'ARN modifié dans le cas de l'analyse d'empreinte ou sur le transcrit de départ dans le cas d'un séquençage. L'extrémité 3' de cet oligonucléotide sert d'amorce à la transcriptase inverse (AMV reverse transcriptase, Q-BIOgene) qui synthétise un brin d'ADNc complémentaire à l'ARN en présence d'un mélange des 4 dNTP. Lorsque l'ARN a été traité par les sondes chimiques ou enzymatiques, ce sont les

modifications chimiques ou les coupures enzymatiques qui sont responsables des arrêts d'élongation. Pour le séquençage, c'est l'incorporation de ddNTP en condition statistique qui permet de stopper la synthèse. Les détails de la réaction est indiqué ci-dessous :

A. Hybridation de l'oligonucléotide

Oligonucléotide marqué à l'extrémité 5'	100 cpm
Transcrit modifié (ou non dans le cas du séquençage)	1 µl
Tampon AMV RT 10X (Tris-HCl 0,5 M pH 8,3 ; $MgCl_2$ 60 mM ; KCl 0,4 M)	0,25 µl
H_2O	qsp 2,5 µl

Le mélange est incubé 40 sec à 96°C, puis plongé dans la glace.

B. Réaction d'élongation

dNTP (5 mM chacun)	0,1 µl
Tampon AMV RT 10X	0,25 µl
Transcriptase inverse (2 U/µl)	0,25 µl
H_2O	1,9 µl

Pour obtenir une extension de l'amorce, 2,5 µl de ce mélange sont ajoutés aux 2,5 µl du mélange d'hybridation. Le tout est incubé 30 min à 45°C. La réaction est arrêtée par 3 µl de bleu dénaturant. Dans le cas d'un séquençage, 0,5 µl du didéoxynucléotide voulu à une concentration de 0,5 mM sont rajoutés au mélange d'élongation.

C. Fractionnement et fluorographie

Avant électrophorèse en gel de polyacrylamide dénaturant à 7%, les échantillons sont placés 2 min à 96°C et plongés dans la glace.

Après migration et séchage, le gel est placé en fluorographie une nuit à -80°C en présence d'écrans amplificateurs (Amersham, NEN).

9. Production, purification et cristallogénèse du fragment (1-280) de la cycline T1

9.1. Production

La cycline T1 est produite par la souche *E. coli* BL21 DE3 codon + (Stratagene) transformée par le plasmide pET-T1(1-280). La culture réalisée en milieu LB supplémenté en kanamycine (100 µg/ml) est ensemencée à 0,1 unité de DO_{600} par ml de culture grâce à une pré-culture d'une nuit dans le même milieu et ensemencée par des cellules souches conservées à -80°C dans du LB 10% glycérol. La culture est agitée à 37°C jusqu'à obtention d'une absorbance de 0,6 unité de DO_{600}/ml, puis la production est induite par l'ajout d'IPTG (0,2mM), $MgCl_2$ (0,2mM) et $ZnSO_4$ (0,1mM). L'agitation se poursuit à 30°C pendant 3h. Les cellules sont finalement récoltées par centrifugation à 4000 g pendant 30min et stockées à - 20°C.

9.2. Purification

9.2.1. Sonication

Les culots cellulaires sont repris dans 25 ml de tampon Hepes (20mM Hepes NaOH pH 8,5 ; 1M NaCl) par litre de culture. Les cellules sont ensuite traitées aux ultrasons sous une puissance de 40 W pendant 2x2min. Après sonication, le lysat bactérien est centrifugé pendant 30min à 4°C à 15 000 rpm. On travaille ensuite avec la partie soluble de la protéine, située dans le surnageant.

9.2.2. Précipitation des acides nucléiques

Le surnageant de sonication est ensuite traité au PEI (0,06%) pendant 15min à 4°C sous agitation et les acides nucléiques sont précipités par centrifugation pendant 30min à 4°C à 15 000 rpm.

9.2.3. Chromatographie d'affinité sur colonne His-Trap couplée à un système FPLC

Le surnageant de PEI est chargé sur la colonne His-Trap 5ml (GeLifeSciences) préalablement équilibré avec du tampon Hepes. Les étapes de lavage et d'élution sont effectuées avec le même tampon (20mM Hepes NaOH pH 8,5) et différentes concentrations de sels : le premier lavage est réalisé avec 2M NaCl, le second avec 1M NaCl et 10% Imidazol, et l'élution avec 1M NaCl et 50% Imidazol.

9.2.4. Chromatographie d'exclusion de taille sur colonne Superdex 75 couplée à un système FPLC

Les pics d'élution issus de la précédente chromatographie sont déposés sur la colonne Superdex 75 (GeLifeSciences) préalablement équilibré avec du tampon Hepes. Les fractions sous le pic sont ensuite cumulées et concentrées sur centriprep (Fabricant) jusqu'à 10 mg/ml. Les concentrations sont déterminées par mesure du coefficient d'extinction molaire et/ou dosage Bradford.

9.3. Cristallogenèse

Les tests de cristallisation ont été réalisés selon la méthode de diffusion de vapeur en gouttes assises dans des boîtes de cristallisation de 96 puits en utilisant le robot de cristallisation Mosquito (TTP Labtech). Pour chaque condition testée, une goutte constituée de 100 nl de

protéine et 100 nl de solution de cristallisation était mise en présence de 100 µl de solution de cristallisation dans le réservoir. Plus de 1000 conditions de cristallisation différentes ont ainsi été utilisées, comprenant les kits Crystal Screen I & II, Index, PEG/Ion, SaltRX (Hampton Research), Structure Screen I & II, Clear Strategy Screen I & II (Molecular Dimensions), and JBScreen 1-10 (Jena Bioscience). Les boîtes de cristallisation ont ensuite été placées à 20°C.

10. Méthode MS2

L'ARN SLS2/E-M3 contient trois sites de fixation de la protéine MS2 du phage R17 (phage R17-MS2 coat protein) et la protéine de fusion MS2-MBP est composée de la protéine MS2 et de MBP (Maltose Binding Protein)

L'ARN SLS2/E-M3 (0,4 nmol) est repris dans 50µl de tampon MS2 (Hepes KOH pH7,9 20mM ; KCl 100mM ; Glycérol 20% ; EDTA 0,2mM ; DTT 0,5mM ; $MgCl_2$ 3mM), dénaturé pendant 10min à 65°C, puis renaturé 10min à température ambiante et placé dans la glace. La protéine MS2-MBP (2 nmol) est ensuite ajoutée et le volume final est ajusté à 300 µl avant d'être incubé 20' dans la glace. Le complexe formé SLS2/E-M3 / MS2-MBP est ensuite fixé sur 50 µl de billes d'amylose pendant 2h à 4°C en agitation. Les billes sont isolées par centrifugation 1min à 11000 rpm, puis lavées avec 500 µl de Tp MS2, avant d'être incubée 20min à 4°C en agitation, en présence de 500 µl d'extrait nucléaire. A nouveau les billes sont isolées par centrifugation 1' à 11000 rpm, puis lavées 3 fois avec 500 µl de Tp MS2 et finalement éluée avec 100 µl de tampon MS2 complémenté en maltose 10 mM.

20 µl de bleu SBL (Tris pH6,8 250mM ; SDS 8% ; β-mercaptoéthanol 20% ; bleu de bromophénol 0,04%) sont ajoutés à 60 µl d'éluat, puis les protéines sont fractionnées sur gel SDS PAGE 10% Acryl-bisacryl (29/1). Après migration, le gel est coloré au bleu colloïdale (sulfate ammonium 10% ; bleu de Coomassie G250 0,1% ; ac. Orthophosphorique 1,6% ; EtOH 20%).

REFERENCES

Aboul-ela, F., Karn, J. and Varani, G. (1995) The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J Mol Biol*, **253**, 313-332.

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**, 1077-1087.

Akari, H., Fujita, M., Kao, S., Khan, M.A., Shehu-Xhilaga, M., Adachi, A. and Strebel, K. (2004) High level expression of human immunodeficiency virus type-1 Vif inhibits viral infectivity by modulating proteolytic processing of the Gag precursor at the p2/nucleocapsid processing site. *J Biol Chem*, **279**, 12355-12362.

Akusjarvi, G. and Stevenin, J. (2003) Remodelling of the host cell RNA splicing machinery during an adenovirus infection. *Curr Top Microbiol Immunol*, **272**, 253-286.

Al-Hashimi, H.M., Gosser, Y., Gorin, A., Hu, W., Majumdar, A. and Patel, D.J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J Mol Biol*, **315**, 95-102.

Allemand, E., Guil, S., Myers, M., Moscat, J., Caceres, J.F. and Krainer, A.R. (2005) Regulation of heterogenous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. *Proc Natl Acad Sci U S A*, **102**, 3605-3610.

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**, 3960-3970.

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**, 4606-4615.

Amrein, H., Gorman, M. and Nothiger, R. (1988) The sex-determining gene tra-2 of Drosophila encodes a putative RNA binding protein. *Cell*, **55**, 1025-1035.

Anand, K., Schulte, A., Fujinaga, K., Scheffzek, K. and Geyer, M. (2007) Cyclin box structure of the P-TEFb subunit cyclin T1 derived from a fusion complex with EIAV tat. *J Mol Biol*, **370**, 826-836.

Ansari, A. and Hampsey, M. (2005) A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev*, **19**, 2969-2978.

Ares, M., Jr. (1986) U2 RNA from yeast is unexpectedly large and contains homology to vertebrate U4, U5, and U6 small nuclear RNAs. *Cell*, **47**, 49-59.

Argyris, E.G. and Pomerantz, R.J. (2004) HIV-1 Vif versus APOBEC3G: newly appreciated warriors in the ancient battle between virus and host. *Trends Microbiol*, **12**, 145-148.

Arora, V.K., Fredericksen, B.L. and Garcia, J.V. (2002) Nef: agent of cell subversion. *Microbes Infect*, **4**, 189-199.

Auboeuf, D., Dowhan, D.H., Li, X., Larkin, K., Ko, L., Berget, S.M. and O'Malley, B.W. (2004) CoAA, a nuclear receptor coactivator protein at the interface of transcriptional coactivation and RNA splicing. *Mol Cell Biol*, **24**, 442-453.

Aubol, B.E., Chakrabarti, S., Ngo, J., Shaffer, J., Nolen, B., Fu, X.D., Ghosh, G. and Adams, J.A. (2003) Processive phosphorylation of alternative splicing factor/splicing factor 2. *Proc Natl Acad Sci U S A*, **100**, 12601-12606.

Avraham, H.K., Jiang, S., Lee, T.H., Prakash, O. and Avraham, S. (2004) HIV-1 Tat-mediated effects on focal adhesion assembly and permeability in brain microvascular endothelial cells. *J Immunol*, **173**, 6228-6233.

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**, 1530-1539.

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**, 861-869.

Barboric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N. and Peterlin, B.M. (2001) NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol Cell*, **8**, 327-337.

Barboric, M., Yik, J.H., Czudnochowski, N., Yang, Z., Chen, R., Contreras, X., Geyer, M., Matija Peterlin, B. and Zhou, Q. (2007) Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. *Nucleic Acids Res*, **35**, 2003-2012.

Bartz, S.R. and Emerman, M. (1999) Human immunodeficiency virus type 1 Tat induces apoptosis and increases sensitivity to apoptotic signals by up-regulating FLICE/caspase-8. *J Virol*, **73**, 1956-1963.

Batsche, E., Yaniv, M. and Muchardt, C. (2006) The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol*, **13**, 22-29.

Bauren, G., Jiang, W.Q., Bernholm, K., Gu, F. and Wieslander, L. (1996) Demonstration of a dynamic, transcription-dependent organization of pre-mRNA splicing factors in polytene nuclei. *J Cell Biol*, **133**, 929-941.

Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R. and Rosch, P. (1995) Structural studies of HIV-1 Tat protein. *J Mol Biol*, **247**, 529-535.

Bedford, M.T., Reed, R. and Leder, P. (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. *Proc Natl Acad Sci U S A*, **95**, 10602-10607.

Behrens, S.E. and Luhrmann, R. (1991) Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev*, **5**, 1439-1452.

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**, 2724-2735.

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

Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G. and Reed, R. (1992b) Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol Cell Biol*, **12**, 3165-3175.

Bentley, D.L. (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol*, **17**, 251-256.

Berget, S.M., Moore, C. and Sharp, P.A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A*, **74**, 3171-3175.

Beyer, A.L., Christensen, M.E., Walker, B.W. and LeStourgeon, W.M. (1977) Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell*, **11**, 127-138.

Beyer, A.L. and Osheim, Y.N. (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev*, **2**, 754-765.

Biamonti, G., Ruggiu, M., Saccone, S., Della Valle, G. and Riva, S. (1994) Two homologous genes, originated by duplication, encode the human hnRNP proteins A2 and A1. *Nucleic Acids Res.*, **22**, 1996-2002.

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**, 5803-5816.

Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, **72**, 291-336.

Blaustein, M., Pelisch, F., Tanos, T., Munoz, M.J., Wengier, D., Quadrana, L., Sanford, J.R., Muschietti, J.P., Kornblihtt, A.R., Caceres, J.F., Coso, O.A. and Srebrow, A. (2005) Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol*, **12**, 1037-1044.

Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell*, **126**, 37-47.

Blencowe, B.J., Bauren, G., Eldridge, A.G., Issner, R., Nickerson, J.A., Rosonina, E. and Sharp, P.A. (2000) The SRm160/300 splicing coactivator subunits. *Rna*, **6**, 111-120.

Blencowe, B.J., Issner, R., Nickerson, J.A. and Sharp, P.A. (1998) A coactivator of pre-mRNA splicing. *Genes Dev*, **12**, 996-1009.

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**, 739-747.

Bonen, L. and Vogel, J. (2001) The ins and outs of group II introns. *Trends Genet*, **17**, 322-331.

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**, 713-725.

Boucher, A., Kharfi, A., Al-Akoum, M., Bossu, P. and Akoum, A. (2001) Cycle-dependent expression of interleukin-1 receptor type II in the human endometrium. *Biol Reprod*, **65**, 890-898.

Boukis, L.A., Liu, N., Furuyama, S. and Bruzik, J.P. (2004) Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. *J Biol Chem*, **279**, 29647-29653.

Bour, S. and Strebel, K. (2003) The HIV-1 Vpu protein: a multifunctional enhancer of viral particle release. *Microbes Infect*, **5**, 1029-1039.

Bourara, K., Litvak, S. and Araya, A. (2000) Generation of G-to-A and C-to-U changes in HIV-1 transcripts by RNA editing. *Science*, **289**, 1564-1566.

Bourgeois, C.F., Lejeune, F. and Stevenin, J. (2004) Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog Nucleic Acid Res Mol Biol*, **78**, 37-88.

Bourgeois, C.F., Popielarz, M., Hildwein, G. and Stevenin, J. (1999) Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol Cell Biol*, **19**, 7347-7356.

Bouyac-Bertoia, M., Dvorin, J.D., Fouchier, R.A., Jenkins, Y., Meyer, B.E., Wu, L.I., Emerman, M. and Malim, M.H. (2001) HIV-1 infection requires a functional integrase NLS. *Mol Cell*, **7**, 1025-1035.

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**, 841-858.

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**, 1259-1265.

Brede, G., Solheim, J. and Prydz, H. (2002) PSKH1, a novel splice factor compartment-associated serine kinase. *Nucleic Acids Res*, **30**, 5301-5309.

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**, 1211-1226.

Bres, V., Kiernan, R., Emiliani, S. and Benkirane, M. (2002) Tat acetyl-acceptor lysines are important for human immunodeficiency virus type-1 replication. *J Biol Chem*, **277**, 22215-22221.

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H. and Luhrmann, R. (1984) Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *Embo J*, **3**, 1357-1363.

Brow, D.A. and Guthrie, C. (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, **334**, 213-218.

Brown, M.T. and Cooper, J.A. (1996) Regulation, substrates and functions of src. *Biochim Biophys Acta*, **1287**, 121-149.

Bruzik, J.P. and Maniatis, T. (1995) Enhancer-dependent interaction between 5' and 3' splice sites in trans. *Proc Natl Acad Sci U S A*, **92**, 7056-7059.

Buckanovich, R.J., Posner, J.B. and Darnell, R.B. (1993) Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron*, **11**, 657-672.

Bukrinsky, M.I. and Haffar, O.K. (1999) HIV-1 nuclear import: in search of a leader. *Front Biosci*, **4**, D772-781.

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M. and Baralle, F.E. (2001) Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *Embo J*, **20**, 1774-1784.

Burd, C.G. and Dreyfuss, G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *Embo J*, **13**, 1197-1204.

Bushnell, D.A., Westover, K.D., Davis, R.E. and Kornberg, R.D. (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science*, **303**, 983-988.

Caceres, J.F. and Kornblihtt, A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet*, **18**, 186-193.

Caceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L. and Krainer, A.R. (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol*, **138**, 225-238.

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

Calnan, B., Tidor, B., Biancalana, S., Hudson, D. and Frankel, A. (1991) Arginine-mediated RNA recognition: the arginine fork. *Science*, **252**, 1167-1171.

Cao, W., Jamison, S.F. and Garcia-Blanco, M.A. (1997) Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *Rna*, **3**, 1456-1467.

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**, 6517-6526.

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**, 4060-4067.

Caputi, M. and Zahler, A.M. (2001) Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem*, **276**, 43850-43859.

Cartegni, L. and Krainer, A.R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet*, **30**, 377-384.

Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S. and Biamonti, G. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol*, **259**, 337-348.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. and Krainer, A.R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*, **31**, 3568-3571.

Cavaloc, Y., Bourgeois, C.F., Kister, L. and Stevenin, J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA*, **5**, 468-483.

Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A.R. and Caceres, J.F. (2002) Nuclear export and retention signals in the RS domain of SR proteins. *Mol Cell Biol*, **22**, 6871-6882.

Cen, S., Khorchid, A., Javanbakht, H., Gabor, J., Stello, T., Shiba, K., Musier-Forsyth, K. and Kleiman, L. (2001) Incorporation of lysyl-tRNA synthetase into human immunodeficiency virus type 1. *J Virol*, **75**, 5043-5048.

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.

Chaloin, O., Peter, J.C., Briand, J.P., Masquida, B., Desgranges, C., Muller, S. and Hoebeke, J. (2005) The N-terminus of HIV-1 Tat protein is essential for Tat-TAR RNA interaction. *Cell Mol Life Sci*, **62**, 355-361.

Champion-Arnaud, P. and Reed, R. (1994) The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev*, **8**, 1974-1983.

Chansky, H.A., Hu, M., Hickstein, D.D. and Yang, L. (2001) Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer Res*, **61**, 3586-3590.

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**, 2814-2820.

Churcher, M.J., Lamont, C., Hamy, F., Dingwall, C., Green, S.M., Lowe, A.D., Butler, J.G., Gait, M.J. and Karn, J. (1993) High affinity binding of TAR RNA by the human immunodeficiency virus type-1 tat protein requires base-pairs in the RNA stem and amino acid residues flanking the basic region. *J Mol Biol*, **230**, 90-110.

Cimarelli, A. and Darlix, J.L. (2002) Assembling the human immunodeficiency virus type 1. *Cell Mol Life Sci*, **59**, 1166-1184.

Coffin, J.M., Hughes, S.H. and Varmus, H.E. (1997) *Retroviruses*.

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**, 10470-10474.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C. and Duncan, P.I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *Embo J*, **15**, 265-275.

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**, 251-258.

Crawford, J.B. and Patton, J.G. (2006) Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol Cell Biol*, **26**, 8791-8802.

Crispino, J.D., Blencowe, B.J. and Sharp, P.A. (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science*, **265**, 1866-1869.

Cullen, B.R. (2003) Nuclear mRNA export: insights from virology. *Trends Biochem Sci*, **28**, 419-424.

Dahlberg, J.E. and Lund, E. (1988) *The genes and transcription of the major small nuclear RNAs*. Springer-Verlag, New-York, NY.

Darlix, J.L., Cristofari, G. and Roques, B.P. (2001) La protéine de nucléocapside du VIH1: un peptide à doigt au coeur du virus et de sa réplication. *Virologie*, **5**, 329-338.

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**, 1100-1109.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P. and Reed, R. (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell*, **26**, 867-881.

Das, R., Zhou, Z. and Reed, R. (2000) Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell*, **5**, 779-787.

Das, S.R. and Jameel, S. (2005) Biology of the HIV Nef protein. *Indian J Med Res*, **121**, 315-332.

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**, 12579-12586.

Davies, R.C., Calvio, C., Bratt, E., Larsson, S.H., Lamond, A.I. and Hastie, N.D. (1998) WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes Dev*, **12**, 3217-3225.

Davis, J.A., Takagi, Y., Kornberg, R.D. and Asturias, F.A. (2002) Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell*, **10**, 409-415.

Dayton, A.I., Sodroski, J.G., Rosen, C.A., Goh, W.C. and Haseltine, W.A. (1986) The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell*, **44**, 941-947.

De Falco, G., Bellan, C., D'Amuri, A., Angeloni, G., Leucci, E., Giordano, A. and Leoncini, L. (2005) Cdk9 regulates neural differentiation and its expression correlates with the differentiation grade of neuroblastoma and PNET tumors. *Cancer Biol Ther*, **4**, 277-281.

De Falco, G., Leucci, E., Onnis, A., Bellan, C., Tigli, C., Wirths, S., Cerino, G., Cocco, M., Crupi, D., De Luca, A., Lanzavecchia, A., Tosi, P., Leoncini, L. and Giordano, A. (2008) Cdk9/Cyclin T1 complex: a key player during the activation/differentiation process of normal lymphoid B cells. *J Cell Physiol*, **215**, 276-282.

de la Mata, M. and Kornblihtt, A.R. (2006) RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol*, **13**, 973-980.

Debouck, C., Gorniak, J.G., Strickler, J.E., Meek, T.D., Metcalf, B.W. and Rosenberg, M. (1987) Human immunodeficiency virus protease expressed in Escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. *Proc Natl Acad Sci U S A*, **84**, 8903-8906.

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**, 251-260.

Delling, U., Reid, L.S., Barnett, R.W., Ma, M.Y., Climie, S., Sumner-Smith, M. and Sonenberg, N. (1992) Conserved nucleotides in the TAR RNA stem of human

immunodeficiency virus type 1 are critical for Tat binding and trans activation: model for TAR RNA tertiary structure. *J Virol*, **66**, 3018-3025.
Desrosiers, R.C., Lifson, J.D., Gibbs, J.S., Czajak, S.C., Howe, A.Y., Arthur, L.O. and Johnson, R.P. (1998) Identification of highly attenuated mutants of simian immunodeficiency virus. *J Virol*, **72**, 1431-1437.
Dostie, J. and Dreyfuss, G. (2002) Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr Biol*, **12**, 1060-1067.
Doudna, J.A. and Cech, T.R. (2002) The chemical repertoire of natural ribozymes. *Nature*, **418**, 222-228.
Douziech, M., Coin, F., Chipoulet, J.M., Arai, Y., Ohkuma, Y., Egly, J.M. and Coulombe, B. (2000) Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking. *Mol Cell Biol*, **20**, 8168-8177.
Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol*, **3**, 195-205.
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.
D'Souza, V. and Summers, M.F. (2005) How retroviruses select their genomes. *Nat Rev Microbiol*, **3**, 643-655.
Du, H. and Rosbash, M. (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature*, **419**, 86-90.
Dye, M.J., Gromak, N. and Proudfoot, N.J. (2006) Exon tethering in transcription by RNA polymerase II. *Mol Cell*, **21**, 849-859.
Egloff, S., Van Herreweghe, E. and Kiss, T. (2006) Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. *Mol Cell Biol*, **26**, 630-642.
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**, 6125-6130.
Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B.J. and Ingles, C.J. (2002) Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *Rna*, **8**, 1102-1111.
Endo-Munoz, L., Warby, T., Harrich, D. and McMillan, N.A. (2005) Phosphorylation of HIV Tat by PKR increases interaction with TAR RNA and enhances transcription. *Virol J*, **2**, 17.
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**, 3607-3617.
Eperon, I.C., Makarova, O.V., Mayeda, A., Munroe, S.H., Caceres, J.F., Hayward, D.G. and Krainer, A.R. (2000) Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol*, **20**, 8303-8318.
Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y.V., Ethier, M., Sheng, Y., Vasilescu, J., Abu-Farha, M., Lambert, J.P., Duewel, H.S., Stewart, II, Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S.L., Moran, M.F., Morin, G.B., Topaloglou, T. and Figeys, D. (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol*, **3**, 89.
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**, 16614-16623.
Fairbrother, W.G., Yeh, R.F., Sharp, P.A. and Burge, C.B. (2002) Predictive identification of exonic splicing enhancers in human genes. *Science*, **297**, 1007-1013.
Fairbrother, W.G., Yeo, G.W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P.A. and Burge, C.B. (2004) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res*, **32**, W187-190.
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**, 344-355.
Fisher, A.G., Feinberg, M.B., Josephs, S.F., Harper, M.E., Marselle, L.M., Reyes, G., Gonda, M.A., Aldovini, A., Debouk, C., Gallo, R.C. and et al. (1986) The trans-activator gene of HTLV-III is essential for virus replication. *Nature*, **320**, 367-371.
Fong, Y.W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, **414**, 929-933.
Franke, E.K. and Luban, J. (1996) Inhibition of HIV-1 replication by cyclosporine A or related compounds correlates with the ability to disrupt the Gag-cyclophilin A interaction. *Virology*, **222**, 279-282.
Frankel, A.D. and Pabo, C.O. (1988) Cellular uptake of the tat protein from human immunodeficiency virus. *Cell*, **55**, 1189-1193.
Fu, X.D. and Maniatis, T. (1992) Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science*, **256**, 535-538.
Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T. and Peterlin, B.M. (2004) Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol*, **24**, 787-795.
Fujinaga, K., Irwin, D., Taube, R., Zhang, F., Geyer, M. and Peterlin, B.M. (2002) A minimal chimera of human cyclin T1 and tat binds TAR and activates human immunodeficiency virus transcription in murine cells. *J Virol*, **76**, 12934-12939.
Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A. and Proudfoot, N.J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev*, **16**, 2792-2799.
Furumoto, T., Tanaka, A., Ito, M., Malik, S., Hirose, Y., Hanaoka, F. and Ohkuma, Y. (2007) A kinase subunit of the human mediator complex, CDK8, positively regulates transcriptional activation. *Genes Cells*, **12**, 119-132.
Gabus, C., Derrington, E., Leblanc, P., Chnaiderman, J., Dormont, D., Swietnicki, W., Morillas, M., Surewicz, W.K., Marc, D., Nandi, P. and Darlix, J.L. (2001) The prion protein has RNA binding and chaperoning properties characteristic of nucleocapsid protein NCP7 of HIV-1. *J Biol Chem*, **276**, 19301-19309.
Galiana-Arnoux, D., Lejeune, F., Gesnel, M.C., Stevenin, J., Breathnach, R. and Del Gatto-Konczak, F. (2003) The CD44 alternative v9 exon contains a splicing enhancer responsive to the SR proteins 9G8, ASF/SF2, and SRp20. *J Biol Chem*, **278**, 32943-32953.
Gallay, P., Swingler, S., Song, J., Bushman, F. and Trono, D. (1995) HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell*, **83**, 569-576.
Ganser-Pornillos, B.K., Yeager, M. and Sundquist, W.I. (2008) The structural biology of HIV assembly. *Curr Opin Struct Biol*, **18**, 203-217.
Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M. and Hahn, B.H. (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*, **397**, 436-441.

Gao, L., Wang, J., Wang, Y. and Andreadis, A. (2007) SR protein 9G8 modulates splicing of tau exon 10 via its proximal downstream intron, a clustering region for frontotemporal dementia mutations. *Mol Cell Neurosci*, **34**, 48-58.

Garber, M.E., Wei, P., KewalRamani, V.N., Mayall, T.P., Herrmann, C.H., Rice, A.P., Littman, D.R. and Jones, K.A. (1998) The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev*, **12**, 3512-3527.

Garneau, D., Revil, T., Fisette, 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**, 22641-22650.

Garriga, J., Bhattacharya, S., Calbo, J., Marshall, R.M., Truongcao, M., Haines, D.S. and Grana, X. (2003) CDK9 is constitutively expressed throughout the cell cycle, and its steady-state expression is independent of SKP2. *Mol Cell Biol*, **23**, 5165-5173.

Gattoni, R., Mahe, D., Mahl, P., Fischer, N., Mattei, M.G., Stevenin, J. and Fuchs, J.P. (1996) The human hnRNP-M proteins: structure and relation with early heat shock-induced splicing arrest and chromosome mapping. *Nucleic Acids Res*, **24**, 2535-2542.

Geiger, J.H., Hahn, S., Lee, S. and Sigler, P.B. (1996) Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science*, **272**, 830-836.

Giacinti, C., Bagella, L., Puri, P.L., Giordano, A. and Simone, C. (2006) MyoD recruits the cdk9/cyclin T2 complex on myogenic-genes regulatory regions. *J Cell Physiol*, **206**, 807-813.

Gibbs, J.S., Lackner, A.A., Lang, S.M., Simon, M.A., Sehgal, P.K., Daniel, M.D. and Desrosiers, R.C. (1995) Progression to AIDS in the absence of a gene for vpr or vpx. *J Virol*, **69**, 2378-2383.

Giglia-Mari, G., Coin, F., Ranish, J.A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N.G., Raams, A., Argentini, M., van der Spek, P.J., Botta, E., Stefanini, M., Egly, J.M., Aebersold, R., Hoeijmakers, J.H. and Vermeulen, W. (2004) A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat Genet*, **36**, 714-719.

Goh, W.C., Rogel, M.E., Kinsey, C.M., Michael, S.F., Fultz, P.N., Nowak, M.A., Hahn, B.H. and Emerman, M. (1998) HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo. *Nat Med*, **4**, 65-71.

Goina, E., Skoko, N. and Pagani, F. (2008) Binding of DAZAP1 and hnRNPA1/A2 to an exonic splicing silencer in a natural BRCA1 exon 18 mutant. *Mol Cell Biol*, **28**, 3850-3860.

Goldstrohm, A.C., Albrecht, T.R., Sune, C., Bedford, M.T. and Garcia-Blanco, M.A. (2001) The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol Cell Biol*, **21**, 7617-7628.

Gotte, M., Li, X. and Wainberg, M.A. (1999) HIV-1 reverse transcription: a brief overview focused on structure- function relationships among molecules involved in initiation of the reaction. *Arch Biochem Biophys*, **365**, 199-210.

Gottlinger, H.G., Dorfman, T., Sodroski, J.G. and Haseltine, W.A. (1991) Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A*, **88**, 3195-3199.

Gozani, O., Feld, R. and Reed, R. (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev*, **10**, 233-243.

Gozani, O., Potashkin, J. and Reed, R. (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell. Biol.*, **18**, 4752-4760.

Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197-1211.

Graveley, B.R. (2005) Coordinated control of splicing and translation. *Nat Struct Mol Biol*, **12**, 1022-1023.

Graveley, B.R., Hertel, K.J. and Maniatis, T. (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *Embo J*, **17**, 6747-6756.

Graveley, B.R. and Maniatis, T. (1998) Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell*, **1**, 765-771.

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**, 673-680.

Gregoire, C., Peloponese, J.M., Jr., Esquieu, D., Opi, S., Campbell, G., Solomiac, M., Lebrun, E., Lebreton, J. and Loret, E.P. (2001) Homonuclear (1)H-NMR assignment and structural characterization of human immunodeficiency virus type 1 Tat Mal protein. *Biopolymers*, **62**, 324-335.

Gui, J.F., Lane, W.S. and Fu, X.D. (1994a) A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature*, **369**, 678-682.

Gui, J.F., Tronchere, H., Chandler, S.D. and Fu, X.D. (1994b) Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc. Natl. Acad. Sci. USA*, **91**, 10824-10828.

Guo, F., Cen, S., Niu, M., Javanbakht, H. and Kleiman, L. (2003) Specific inhibition of the synthesis of human lysyl-tRNA synthetase results in decreases in tRNA(Lys) incorporation, tRNA(3)(Lys) annealing to viral RNA, and viral infectivity in human immunodeficiency virus type 1. *J Virol*, **77**, 9817-9822.

Guth, S. and Valcarcel, J. (2000) Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF. *J Biol Chem*, **275**, 38059-38066.

Haffar, O.K., Popov, S., Dubrovsky, L., Agostini, I., Tang, H., Pushkarsky, T., Nadler, S.G. and Bukrinsky, M. (2000) Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. *J Mol Biol*, **299**, 359-368.

Hahn, B.H., Shaw, G.M., De Cock, K.M. and Sharp, P.M. (2000) AIDS as a zoonosis: scientific and public health implications. *Science*, **287**, 607-614.

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 hnRNP A1 at the HIV-1 Tat exon 2 splicing site. *J Biol Chem*.

Hampsey, M. and Reinberg, D. (2003) Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell*, **113**, 429-432.

Harrich, D. and Hooker, B. (2002) Mechanistic aspects of HIV-1 reverse transcription initiation. *Rev Med Virol*, **12**, 31-45.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S. and Malim, M.H. (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell*, **113**, 803-809.

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**, 16719-16724.

Haseloff, J. and Gerlach, W.L. (1989) Sequences required for self-catalysed cleavage of the satellite RNA of tobacco ringspot virus. *Gene*, **82**, 43-52.

Hashimoto, C. and Steitz, J.A. (1984) U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res*, **12**, 3283-3293.

Hashizume, C., Kuramitsu, M., Zhang, X., Kurosawa, T., Kamata, M. and Aida, Y. (2007) Human immunodeficiency virus type 1 Vpr interacts with spliceosomal protein SAP145 to mediate cellular pre-mRNA splicing inhibition. *Microbes Infect*, **9**, 490-497.

Hastings, M.L., Allemand, E., Duelli, D.M., Myers, M.P. and Krainer, A.R. (2007) Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF65. *PLoS ONE*, **2**, e538.

Hausner, T.P., Giglio, L.M. and Weiner, A.M. (1990) Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev*, **4**, 2146-2156.

Hayashi, K., Watanabe, T., Tanaka, A., Furumoto, T., Sato-Tsuchiya, C., Kimura, M., Yokoi, M., Ishihama, A., Hanaoka, F. and Ohkuma, Y. (2005) Studies of Schizosaccharomyces pombe TFIIE indicate conformational and functional changes in RNA polymerase II at transcription initiation. *Genes Cells*, **10**, 207-224.

Heinrichs, V. and Baker, B.S. (1995) The Drosophila SR protein RBP1 contributes to the regulation of doublesex alternative splicing by recognizing RBP1 RNA target sequences. *Embo J*, **14**, 3987-4000.

Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S. and Young, R.A. (1998) Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol Cell*, **2**, 43-53.

Hertel, K.J. and Graveley, B.R. (2005) RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem Sci*, **30**, 115-118.

Hertel, K.J., Lynch, K.W., Hsiao, E.C., Liu, E.H. and Maniatis, T. (1996) Structural and functional conservation of the Drosophila doublesex splicing enhancer repeat elements. *Rna*, **2**, 969-981.

Hertel, K.J. and Maniatis, T. (1999) Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. *Proc Natl Acad Sci U S A*, **96**, 2651-2655.

Hindmarsh, P. and Leis, J. (1999) Retroviral DNA integration. *Microbiol Mol Biol Rev*, **63**, 836-843, table of contents.

Hirose, Y. and Manley, J.L. (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev*, **14**, 1415-1429.

Hirose, Y. and Ohkuma, Y. (2007) Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J Biochem (Tokyo)*, **141**, 601-608.

Hirose, Y., Tacke, R. and Manley, J.L. (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev*, **13**, 1234-1239.

Hofmann, Y. and Wirth, B. (2002) hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Hum Mol Genet*, **11**, 2037-2049.

Holm, K., Weclewicz, K., Hewson, R. and Suomalainen, M. (2003) Human immunodeficiency virus type 1 assembly and lipid rafts: Pr55(gag) associates with membrane domains that are largely resistant to Brij98 but sensitive to Triton X-100. *J Virol*, **77**, 4805-4817.

Honore, B., Vorum, H. and Baandrup, U. (1999) hnRNPs H, H' and F behave differently with respect to posttranslational cleavage and subcellular localization. *FEBS Lett*, **456**, 274-280.

Hope, T.J. (1999) The ins and outs of HIV Rev. *Arch Biochem Biophys*, **365**, 186-191.

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**, 2695-2704.

Huang, T., Vilardell, J. and Query, C.C. (2002) Pre-spliceosome formation in S.pombe requires a stable complex of SF1-U2AF(59)-U2AF(23). *Embo J*, **21**, 5516-5526.

Huang, Y., Gattoni, R., Stevenin, J. and Steitz, J.A. (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell*, **11**, 837-843.

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**, 7676-7683.

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**, 9666-9670.

Huffman, K.M. and Arrigo, S.J. (1997) Identification of cis-acting repressor activity within human immunodeficiency virus type 1 protease sequences. *Virology*, **234**, 253-260.

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**, 1988-1998.

Huigen, M.C., Kamp, W. and Nottet, H.S. (2004) Multiple effects of HIV-1 trans-activator protein on the pathogenesis of HIV-1 infection. *Eur J Clin Invest*, **34**, 57-66.

Huvent, I., Hong, S.S., Fournier, C., Gay, B., Tournier, J., Carriere, C., Courcoul, M., Vigne, R., Spire, B. and Boulanger, P. (1998) Interaction and co-encapsidation of human immunodeficiency virus type 1 Gag and Vif recombinant proteins. *J Gen Virol*, **79 (Pt 5)**, 1069-1081.

Ideue, T., Sasaki, Y.T., Hagiwara, M. and Hirose, T. (2007) Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev*, **21**, 1993-1998.

Ismaili, N., Sha, M., Gustafson, E.H. and Konarska, M.M. (2001) The 100-kda U5 snRNP protein (hPrp28p) contacts the 5' splice site through its ATPase site. *Rna*, **7**, 182-193.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J. and Varmus, H.E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, **331**, 280-283.

Jacquenet, 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.

Jacquenet, S., Mereau, A., Bilodeau, P.S., Damier, L., Stoltzfus, C.M. and Branlant, C. (2001a) 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**, 40464-40475.

Jacquenet, S., Ropers, D., Bilodeau, P.S., Damier, L., Mougin, A., Stoltzfus, C.M. and Branlant, C. (2001b) Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing. *Nucleic Acids Res*, **29**, 464-478.

Jady, B.E. and Kiss, T. (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *Embo J*, **20**, 541-551.

Jakobovits, A., Smith, D.H., Jakobovits, E.B. and Capon, D.J. (1988) A discrete element 3' of human immunodeficiency virus 1 (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV trans activator. *Mol Cell Biol*, **8**, 2555-2561.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N. and Ozato, K. (2005) The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell*, **19**, 523-534.

Jeang, K.T., Xiao, H. and Rich, E.A. (1999) Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *J Biol Chem*, **274**, 28837-28840.

Jones, K.A. and Peterlin, B.M. (1994) Control of RNA initiation and elongation at the HIV-1 promoter. *Annu Rev Biochem*, **63**, 717-743.

Jumaa, P., Wren, B. and Tabaqchali, S. (1996) Epidemiology and typing of Clostridium difficile. *Eur J Gastroenterol Hepatol*, **8**, 1035-1040.

Jung, A., Maier, R., Vartanian, J.P., Bocharov, G., Jung, V., Fischer, U., Meese, E., Wain-Hobson, S. and Meyerhans, A. (2002) Multiply infected spleen cells in HIV patients. *Nature*, **418**, 144.

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**, 426-439.

Jurica, M.S. and Moore, M.J. (2002) Capturing splicing complexes to study structure and mechanism. *Methods*, **28**, 336-345.

Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, **12**, 5-14.

Kaehlcke, K., Dorr, A., Hetzer-Egger, C., Kiermer, V., Henklein, P., Schnoelzer, M., Loret, E., Cole, P.A., Verdin, E. and Ott, M. (2003) Acetylation of Tat defines a cyclinT1-independent step in HIV transactivation. *Mol Cell*, **12**, 167-176.

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**, 1782-1791.

Kammler, S., Otte, M., Hauber, I., Kjems, J., Hauber, J. and Schaal, H. (2006) The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology*, **3**, 89.

Kanopka, A., Muhlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C. and Akusjarvi, G. (1998) Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature*, **393**, 185-187.

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**, 661-671.

Karn, J. (1999) Tackling Tat. *J Mol Biol*, **293**, 235-254.

Karn, J. (2000) Tat, a novel regulator of HIV transcription and latency. In Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Mullins, J., Sodroski, J., Wolinsky, S. and Korber, B. (eds.), *HIV Sequence Compendium 2000*. Los Alamos National Laboratory, Los Alamos, pp. 2-19.

Kataoka, N., Bachorik, J.L. and Dreyfuss, G. (1999) Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol*, **145**, 1145-1152.

Kataoka, N. and Dreyfuss, G. (2004) A simple whole cell lysate system for in vitro splicing reveals a stepwise assembly of the exon-exon junction complex. *J Biol Chem*, **279**, 7009-7013.

Kaufer, N.F. and Potashkin, J. (2000) Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. *Nucleic Acids Res*, **28**, 3003-3010.

Kent, O.A. and MacMillan, A.M. (2002) Early organization of pre-mRNA during spliceosome assembly. *Nat Struct Biol*, **9**, 576-581.

Kent, O.A., Ritchie, D.B. and Macmillan, A.M. (2005) Characterization of a U2AF-independent commitment complex (E') in the mammalian spliceosome assembly pathway. *Mol Cell Biol*, **25**, 233-240.

Khan, M.A., Aberham, C., Kao, S., Akari, H., Gorelick, R., Bour, S. and Strebel, K. (2001) Human immunodeficiency virus type 1 Vif protein is packaged into the nucleoprotein complex through an interaction with viral genomic RNA. *J Virol*, **75**, 7252-7265.

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**, 17-26.

Kim, D.H., Langlois, M.A., Lee, K.B., Riggs, A.D., Puymirat, J. and Rossi, J.J. (2005) HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res*, **33**, 3866-3874.

Kim, E., Du, L., Bregman, D.B. and Warren, S.L. (1997) Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. *J Cell Biol*, **136**, 19-28.

Kim, S., Park, G.H. and Paik, W.K. (1998) Recent advances in protein methylation: enzymatic methylation of nucleic acid binding proteins. *Amino Acids*, **15**, 291-306.

Kiss, T. (2004) Biogenesis of small nuclear RNPs. *J Cell Sci*, **117**, 5949-5951.

Kleiman, L., Halwani, R. and Javanbakht, H. (2004) The selective packaging and annealing of primer tRNALys3 in HIV-1. *Curr HIV Res*, **2**, 163-175.

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**, 119-124.

Kojima, T., Zama, T., Wada, K., Onogi, H. and Hagiwara, M. (2001) Cloning of human PRP4 reveals interaction with Clk1. *J Biol Chem*, **276**, 32247-32256.

Komarnitsky, P., Cho, E.J. and Buratowski, S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*, **14**, 2452-2460.

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**, 1789-1796.

Kornblihtt, A.R. (2005) Promoter usage and alternative splicing. *Curr Opin Cell Biol*, **17**, 262-268.

Krecic, A.M. and Swanson, M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol*, **11**, 363-371.

Kuiken, C.L., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Wolinsky, S. and Korber, B. (2001) *HIV Sequence Compendium 2001*.

Kunkel, G.R., Maser, R.L., Calvet, J.P. and Pederson, T. (1986) U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc Natl Acad Sci U S A*, **83**, 8575-8579.

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**, 3551-3561.

Kuramitsu, M., Hashizume, C., Yamamoto, N., Azuma, A., Kamata, M., Tanaka, Y. and Aida, Y. (2005) A novel role for Vpr of human immunodeficiency virus type 1 as a regulator of the splicing of cellular pre-mRNA. *Microbes Infect*, **7**, 1150-1160.

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 TFIIH and regulates transcriptional initiation. *Nat Struct Biol*, **9**, 800-805.

Labourier, E., Allemand, E., Brand, S., Fostier, M., Tazi, J. and Bourbon, H.M. (1999) Recognition of exonic splicing enhancer sequences by the Drosophila splicing repressor RSF1. *Nucleic Acids Res*, **27**, 2377-2386.

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**, 1285-1296.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D. and Ebright, R.H. (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev*, **12**, 34-44.

Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol*, **4**, 605-612.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C. and Brenner, S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, **446**, 926-929.

Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M.J. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *Embo J*, **20**, 4987-4997.

Le Rouzic, E. and Benichou, S. (2005) The Vpr protein from HIV-1: distinct roles along the viral life cycle. *Retrovirology*, **2**, 11.

Legrain, P. and Choulika, A. (1990) The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cysteine/histidine motif common to several splicing factors. *Embo J*, **9**, 2775-2781.

Lejeune, F., Cavaloc, Y. and Stevenin, J. (2001) Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene. Identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. *J Biol Chem*, **276**, 7850-7858.

Lejeune, F., Ishigaki, Y., Li, X. and Maquat, L.E. (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *Embo J*, **21**, 3536-3545.

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**, 309-315.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.L. and Lafyatis, R. (2002) Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev*, **16**, 594-607.

Lemaire, R., Winne, A., Sarkissian, M. and Lafyatis, R. (1999) SF2 and SRp55 regulation of CD45 exon 4 skipping during T cell activation. *Eur J Immunol*, **29**, 823-837.

Lerga, A., Hallier, M., Delva, L., Orvain, C., Gallais, I., Marie, J. and Moreau-Gachelin, F. (2001) Identification of an RNA binding specificity for the potential splicing factor TLS. *J Biol Chem*, **276**, 6807-6816.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) Are snRNPs involved in splicing? *Nature*, **283**, 220-224.

Levy, J.A. and Fox, S.E. (1998) The outreach-assisted model of partner notification with IDUs. *Public Health Rep*, **113 Suppl 1**, 160-169.

Li, X., Shambaugh, M.E., Rottman, F.M. and Bokar, J.A. (2000) SR proteins Asf/SF2 and 9G8 interact to activate enhancer-dependent intron D splicing of bovine growth hormone pre-mRNA in vitro. *Rna*, **6**, 1847-1858.

Liang, C., Hu, J., Russell, R.S., Kameoka, M. and Wainberg, M.A. (2004) Spliced human immunodeficiency virus type 1 RNA is reverse transcribed into cDNA within infected cells. *AIDS Res Hum Retroviruses*, **20**, 203-211.

Lim, L.P. and Burge, C.B. (2001) A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A*, **98**, 11193-11198.

Lin, X., Taube, R., Fujinaga, K. and Peterlin, B.M. (2002) P-TEFb containing cyclin K and Cdk9 can activate transcription via RNA. *J Biol Chem*, **277**, 16873-16878.

Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., 3rd and Hartzog, G.A. (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol Cell Biol*, **23**, 1368-1378.

Liu, H.X., Chew, S.L., Cartegni, L., Zhang, M.Q. and Krainer, A.R. (2000) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol*, **20**, 1063-1071.

Liu, H.X., Zhang, M. and Krainer, A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev*, **12**, 1998-2012.

Liu, Z.R. (2002) p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Mol Cell Biol*, **22**, 5443-5450.

Long, K.S. and Crothers, D.M. (1999) Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. *Biochemistry*, **38**, 10059-10069.

Longman, D., McGarvey, T., McCracken, S., Johnstone, I.L., Blencowe, B.J. and Caceres, J.F. (2001) Multiple interactions between SRm160 and SR family proteins in enhancer-dependent splicing and development of C. elegans. *Curr Biol*, **11**, 1923-1933.

Lou, H. and Gagel, R.F. (1998) Alternative RNA processing--its role in regulating expression of calcitonin/calcitonin gene-related peptide. *J Endocrinol*, **156**, 401-405.

Lu, S. and Cullen, B.R. (2003) Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *Rna*, **9**, 618-630.

Lund, E. and Dahlberg, J.E. (1992) Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science*, **255**, 327-330.

Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M. and Reed, R. (2001) Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature*, **413**, 644-647.

Lykke-Andersen, J., Shu, M.D. and Steitz, J.A. (2001) Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science*, **293**, 1836-1839.

Lynch, K.W. and Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes Dev*, **10**, 2089-2101.

Madhani, H.D. and Guthrie, C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71**, 803-817.

Madhani, H.D. and Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet*, **28**, 1-26.

Makarov, E.M., Makarova, O.V., Achsel, T. and Luhrmann, R. (2000) The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions. *J Mol Biol*, **298**, 567-575.

Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P. and Luhrmann, R. (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *Embo J*, **21**, 1148-1157.

Makarova, O.V., Makarov, E.M. and Luhrmann, R. (2001) The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *Embo J*, **20**, 2553-2563.

Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M. and Luhrmann, R. (2004) A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *Embo J*, **23**, 2381-2391.

Maldarelli, F., Xiang, C., Chamoun, G. and Zeichner, S.L. (1998) The expression of the essential nuclear splicing factor SC35 is altered by human immunodeficiency virus infection. *Virus Res*, **53**, 39-51.

Malim, M.H., Hauber, J., Fenrick, R. and Cullen, B.R. (1988) Immunodeficiency virus rev trans-activator modulates the expression of the viral regulatory genes. *Nature*, **335**, 181-183.

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J. and Reinberg, D. (2004) Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc Natl Acad Sci U S A*, **101**, 7572-7577.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L. and Trono, D. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature*, **424**, 99-103.

Manley, J.L. and Tacke, R. (1996) SR proteins and splicing control. *Genes Dev*, **10**, 1569-1579.

Marcello, A., Lusic, M., Pegoraro, G., Pellegrini, V., Beltram, F. and Giacca, M. (2004) Nuclear organization and the control of HIV-1 transcription. *Gene*, **326**, 1-11.

Margottin, F., Bour, S.P., Durand, H., Selig, L., Benichou, S., Richard, V., Thomas, D., Strebel, K. and Benarous, R. (1998) A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol Cell*, **1**, 565-574.

Marozzi, A., Meneveri, R., Giacca, M., Gutierrez, M.I., Siccardi, A.G. and Ginelli, E. (1998) In vitro selection of HIV-1 TAR variants by the Tat protein. *J Biotechnol*, **61**, 117-128.

Marquez, N., Sancho, R., Macho, A., Moure, A., Masip, I., Messeguer, A. and Munoz, E. (2006) Anti-Tat and anti-HIV activities of trimers of n-alkylglycines. *Biochem Pharmacol*, **71**, 596-604.

Martin-Serrano, J., Zang, T. and Bieniasz, P.D. (2003) Role of ESCRT-I in retroviral budding. *J Virol*, **77**, 4794-4804.

Massenet, S., Motorin, Y., Lafontaine, D.L., Hurt, E.C., Grosjean, H. and Branlant, C. (1999) Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol Cell Biol*, **19**, 2142-2154.

Massenet, S., Mougin, A. and Branlant, C. (1998) *Posttranscriptionnal modifications in the U small nuclear RNAs.*

Matsugami, A., Tamura, Y., Kudo, M., Uesugi, S., Yamamoto, R., Kumar, P. and Katahira, M. (2004) Structure of RNA aptamer for HIV Tat complexed with Tat-derived peptide. *Nucleic Acids Symp Ser (Oxf)*, 111-112.

Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M.Q., Gardiner, E.M. and Krainer, A.R. (1999a) Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *Embo J*, **18**, 4560-4570.

Mayeda, A., Munroe, S.H., Caceres, J.F. and Krainer, A.R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.*, **13**, 5483-5495.

Mayeda, A., Screaton, G.R., Chandler, S.D., Fu, X.D. and Krainer, A.R. (1999b) 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**, 1853-1863.

Mayes, A.E., Verdone, L., Legrain, P. and Beggs, J.D. (1999) Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *Embo J*, **18**, 4321-4331.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S. and Bentley, D.L. (1997) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev*, **11**, 3306-3318.

Meister, G., Hannus, S., Plottner, O., Baars, T., Hartmann, E., Fakan, S., Laggerbauer, B. and Fischer, U. (2001) SMNrp is an essential pre-mRNA splicing factor required for the formation of the mature spliceosome. *Embo J*, **20**, 2304-2314.

Metzger, A.U., Bayer, P., Willbold, D., Hoffmann, S., Frank, R.W., Goody, R.S. and Rosch, P. (1997) The interaction of HIV-1 Tat(32-72) with its target RNA: a fluorescence and nuclear magnetic resonance study. *Biochem Biophys Res Commun*, **241**, 31-36.

Michael, W.M., Choi, M. and Dreyfuss, G. (1995) A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell*, **83**, 415-422.

Michels, A.A., Nguyen, V.T., Fraldi, A., Labas, V., Edwards, M., Bonnet, F., Lania, L. and Bensaude, O. (2003) MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol Cell Biol*, **23**, 4859-4869.

Miki, T., Bottaro, D.P., Fleming, T.P., Smith, C.L., Burgess, W.H., Chan, A.M. and Aaronson, S.A. (1992) Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene. *Proc Natl Acad Sci U S A*, **89**, 246-250.

Minovitsky, S., Gee, S.L., Schokrpur, S., Dubchak, I. and Conboy, J.G. (2005) The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Res*, **33**, 714-724.

Misteli, T., Caceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F. and Spector, D.L. (1998) Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J Cell Biol*, **143**, 297-307.

Misteli, T., Caceres, J.F. and Spector, D.L. (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature*, **387**, 523-527.

Misteli, T. and Spector, D.L. (1999) RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell*, **3**, 697-705.

Monsalve, M., Wu, Z., Adelmant, G., Puigserver, P., Fan, M. and Spiegelman, B.M. (2000) Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol Cell*, **6**, 307-316.

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**, 364-368.

Morris, D.P. and Greenleaf, A.L. (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, **275**, 39935-39943.

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**, 631-649.

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**, 2133-2143.

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**, 509-518.

Muesing, M.A., Smith, D.H. and Capon, D.J. (1987) Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. *Cell*, **48**, 691-701.

Mujtaba, S., He, Y., Zeng, L., Farooq, A., Carlson, J.E., Ott, M., Verdin, E. and Zhou, M.M. (2002) Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. *Mol Cell*, **9**, 575-586.

Muriaux, D. and Rein, A. (2003) Encapsidation and transduction of cellular genes by retroviruses. *Front Biosci*, **8**, d135-142.

Nakielny, S., Shaikh, S., Burke, B. and Dreyfuss, G. (1999) Nup153 is an M9-containing mobile nucleoporin with a novel Ran-binding domain. *Embo J*, **18**, 1982-1995.

Nesic, D. and Kramer, A. (2001) Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17S U2 snRNP, and prespliceosome formation. *Mol Cell Biol*, **21**, 6406-6417.

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**, 651-660.

Newby, M.I. and Greenbaum, N.L. (2002) Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat Struct Biol*, **9**, 958-965.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E. and Ares, M., Jr. (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev*, **21**, 708-718.

Ni, Z., Schwartz, B.E., Werner, J., Suarez, J.R. and Lis, J.T. (2004) Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol Cell*, **13**, 55-65.

Nilsen, T.W. (2002) The spliceosome: no assembly required? *Mol Cell*, **9**, 8-9.

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**, 43110-43114.

Nogues, G., Kadener, S., Cramer, P., de la Mata, M., Fededa, J.P., Blaustein, M., Srebrow, A. and Kornblihtt, A.R. (2003) Control of alternative pre-mRNA splicing by RNA Pol II elongation: faster is not always better. *IUBMB Life*, **55**, 235-241.

Nott, A., Le Hir, H. and Moore, M.J. (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev*, **18**, 210-222.

O'Gorman, W., Thomas, B., Kwek, K.Y., Furger, A. and Akoulitchev, A. (2005) Analysis of U1 small nuclear RNA interaction with cyclin H. *J Biol Chem*, **280**, 36920-36925.

Ohagen, A. and Gabuzda, D. (2000) Role of Vif in stability of the human immunodeficiency virus type 1 core. *J Virol*, **74**, 11055-11066.

Ohkuma, Y. (1997) Multiple functions of general transcription factors TFIIE and TFIIH in transcription: possible points of regulation by trans-acting factors. *J Biochem (Tokyo)*, **122**, 481-489.

Ohkuma, Y. and Roeder, R.G. (1994) Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature*, **368**, 160-163.

Orphanides, G. and Reinberg, D. (2002) A unified theory of gene expression. *Cell*, **108**, 439-451.

Ostrowski, J., Schullery, D.S., Denisenko, O.N., Higaki, Y., Watts, J., Aebersold, R., Stempka, L., Gschwendt, M. and Bomsztyk, K. (2000) Role of tyrosine phosphorylation in the regulation of the interaction of heterogenous nuclear ribonucleoprotein K protein with its protein and RNA partners. *J Biol Chem*, **275**, 3619-3628.

Page-McCaw, P.S., Amonlirdviman, K. and Sharp, P.A. (1999) PUF60: a novel U2AF65-related splicing activity. *Rna*, **5**, 1548-1560.

Park, E., Han, J., Son, G.H., Lee, M.S., Chung, S., Park, S.H., Park, K., Lee, K.H., Choi, S., Seong, J.Y. and Kim, K. (2006) Cooperative actions of Tra2alpha with 9G8 and SRp30c in the RNA splicing of the gonadotropin-releasing hormone gene transcript. *J Biol Chem*, **281**, 401-409.

Patel, A.A. and Steitz, J.A. (2003) Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*, **4**, 960-970.

Patel, N.A., Kaneko, S., Apostolatos, H.S., Bae, S.S., Watson, J.E., Davidowitz, K., Chappell, D.S., Birnbaum, M.J., Cheng, J.Q. and Cooper, D.R. (2005) Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase CbetaII alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40. *J Biol Chem*, **280**, 14302-14309.

Paushkin, S., Gubitz, A.K., Massenet, S. and Dreyfuss, G. (2002) The SMN complex, an assemblyosome of ribonucleoproteins. *Curr Opin Cell Biol*, **14**, 305-312.

Peeters, M. and Sharp, P.M. (2000) Genetic diversity of HIV-1: the moving target. *Aids*, **14 Suppl 3**, S129-140.

Peled-Zehavi, H., Berglund, J.A., Rosbash, M. and Frankel, A.D. (2001) Recognition of RNA branch point sequences by the KH domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. *Mol Cell Biol*, **21**, 5232-5241.

Peloponese, J.M., Jr., Gregoire, C., Opi, S., Esquieu, D., Sturgis, J., Lebrun, E., Meurs, E., Collette, Y., Olive, D., Aubertin, A.M., Witvrow, M., Pannecouque, C., De Clercq, E., Bailly, C., Lebreton, J. and Loret, E.P. (2000) 1H-13C nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. *C R Acad Sci III*, **323**, 883-894.

Peng, J., Marshall, N.F. and Price, D.H. (1998a) Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *J Biol Chem*, **273**, 13855-13860.

Peng, J., Zhu, Y., Milton, J.T. and Price, D.H. (1998b) Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev*, **12**, 755-762.

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**, 663-668.

Perriman, R., Barta, I., Voeltz, G.K., Abelson, J. and Ares, M., Jr. (2003) ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. *Proc Natl Acad Sci U S A*, **100**, 13857-13862.

Peterlin, B.M. and Price, D.H. (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol Cell*, **23**, 297-305.

Pinol-Roma, S. and Dreyfuss, G. (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature*, **355**, 730-732.

Plessel, G., Fischer, U. and Luhrmann, R. (1994) m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. *Mol Cell Biol*, **14**, 4160-4172.

Pollard, V.W., Michael, W.M., Nakielny, S., Siomi, M.C., Wang, F. and Dreyfuss, G. (1996) A novel receptor-mediated nuclear protein import pathway. *Cell*, **86**, 985-994.

Ponthier, J.L., Schluepen, C., Chen, W., Lersch, R.A., Gee, S.L., Hou, V.C., Lo, A.J., Short, S.A., Chasis, J.A., Winkelmann, J.C. and Conboy, J.G. (2006) Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. *J Biol Chem*, **281**, 12468-12474.

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**, 909-917.

Pozzoli, U. and Sironi, M. (2005) Silencers regulate both constitutive and alternative splicing events in mammals. *Cell Mol Life Sci*, **62**, 1579-1604.

Prasad, J., Colwill, K., Pawson, T. and Manley, J.L. (1999) The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol Cell Biol*, **19**, 6991-7000.

Price, S.R., Ito, N., Oubridge, C., Avis, J.M. and Nagai, K. (1995) Crystallization of RNA-protein complexes. I. Methods for the large- scale preparation of RNA suitable for crystallographic studies. *J Mol Biol*, **249**, 398-408.

Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501-512.

Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. and Williamson, J.R. (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*, **257**, 76-80.

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**, 6365-6378.

Query, C.C., Bentley, R.C. and Keene, J.D. (1989a) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, **57**, 89-101.

Query, C.C., Bentley, R.C. and Keene, J.D. (1989b) A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol Cell Biol*, **9**, 4872-4881.

Rader, S.D. and Guthrie, C. (2002) A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *Rna*, **8**, 1378-1392.

Raffetseder, U., Frye, B., Rauen, T., Jurchott, K., Royer, H.D., Jansen, P.L. and Mertens, P.R. (2003) Splicing factor SRp30c interaction with Y-box protein-1 confers nuclear YB-1 shuttling and alternative splice site selection. *J Biol Chem*, **278**, 18241-18248.

Raghunathan, P.L. and Guthrie, C. (1998) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol*, **8**, 847-855.

Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P. and Kramer, A. (1998) Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. *Rna*, **4**, 551-565.

Rappsilber, J., Ajuh, P., Lamond, A.I. and Mann, M. (2001) SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome. *J Biol Chem*, **276**, 31142-31150.

Reddy, R., Henning, D., Epstein, P. and Busch, H. (1981) Primary and secondary structure of U2 snRNA. *Nucleic Acids Res*, **9**, 5645-5658.

Reed, R. (1989) The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.*, **3**, 2113-2123.

Reed, R. and Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell*, **108**, 523-531.

Reichert, V.L., Le Hir, H., Jurica, M.S. and Moore, M.J. (2002) 5' exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. *Genes Dev*, **16**, 2778-2791.

Reyes, J.L., Kois, P., Konforti, B.B. and Konarska, M.M. (1996) The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA*, **2**, 213-225.

Richter, S., Cao, H. and Rana, T.M. (2002) Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. *Biochemistry*, **41**, 6391-6397.

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**, 721-731.

Rinke, J. and Steitz, J.A. (1985) Association of the lupus antigen La with a subset of U6 snRNA molecules. *Nucleic Acids Res*, **13**, 2617-2629.

Rio, D.C., Laski, F.A. and Rubin, G.M. (1986) Identification and immunochemical analysis of biologically active Drosophila P element transposase. *Cell*, **44**, 21-32.

Robert, F., Blanchette, M., Maes, O., Chabot, B. and Coulombe, B. (2002) A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. *J Biol Chem*, **277**, 9302-9306.

Roeder, R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*, **21**, 327-335.

Ropers, D., Ayadi, L., Gattoni, R., Jacquenet, S., Damier, L., Branlant, C. and Stevenin, J. (2004) Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA. *J Biol Chem*, **279**, 29963-29973.

Roscigno, R.F. and Garcia-Blanco, M.A. (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *Rna*, **1**, 692-706.

Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J.F., Antoine, E., Cathala, G., Brunel, C. and Tazi, J. (1996) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature*, **381**, 80-82.

Ruskin, B. and Green, M.R. (1985) An RNA processing activity that debranches RNA lariats. *Science*, **229**, 135-140.

Ryo, A., Suzuki, Y., Arai, M., Kondoh, N., Wakatsuki, T., Hada, A., Shuda, M., Tanaka, K., Sato, C., Yamamoto, M. and Yamamoto, N. (2000) Identification and characterization of differentially expressed mRNAs in HIV type 1-infected human T cells. *AIDS Res Hum Retroviruses*, **16**, 995-1005.

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**, 276-278.

Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. and Seraphin, B. (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *Embo J*, **18**, 3451-3462.

Salghetti, S., Mariani, R. and Skowronski, J. (1995) Human immunodeficiency virus type 1 Nef and p56lck protein-tyrosine kinase interact with a common element in CD4 cytoplasmic tail. *Proc Natl Acad Sci U S A*, **92**, 349-353.

Sanford, J.R. and Bruzik, J.P. (1999) Developmental regulation of SR protein phosphorylation and activity. *Genes Dev*, **13**, 1513-1518.

Sanford, J.R. and Bruzik, J.P. (2001) Regulation of SR protein localization during development. *Proc Natl Acad Sci U S A*, **98**, 10184-10189.

Sanford, J.R., Ellis, J. and Caceres, J.F. (2005) Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochem Soc Trans*, **33**, 443-446.

Sanford, J.R., Gray, N.K., Beckmann, K. and Caceres, J.F. (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes Dev*, **18**, 755-768.

Sanford, J.R., Longman, D. and Caceres, J.F. (2003) Multiple roles of the SR protein family in splicing regulation. *Prog Mol Subcell Biol*, **31**, 33-58.

Sawa, H. and Shimura, Y. (1992) Association of U6 snRNA with the 5'-splice site region of pre-mRNA in the spliceosome. *Genes Dev.*, **6**, 244-254.

Schaal, T.D. and Maniatis, T. (1999) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.*, **19**, 1705-1719.

Schmitt, P., Gattoni, R., Keohavong, P. and Stevenin, J. (1987) Alternative splicing of E1A transcripts of adenovirus requires appropriate ionic conditions in vitro. *Cell*, **50**, 31-39.

Schullery, D.S., Ostrowski, J., Denisenko, O.N., Stempka, L., Shnyreva, M., Suzuki, H., Gschwendt, M. and Bomsztyk, K. (1999) Regulated interaction of protein kinase Cdelta with the heterogeneous nuclear ribonucleoprotein K protein. *J Biol Chem*, **274**, 15101-15109.

Schulte, A., Czudnochowski, N., Barboric, M., Schonichen, A., Blazek, D., Peterlin, B.M. and Geyer, M. (2005) Identification of a cyclin T-binding domain in Hexim1 and biochemical analysis of its binding competition with HIV-1 Tat. *J Biol Chem*, **280**, 24968-24977.

Schwartz, S., Felber, B.K., Fenyo, E.M. and Pavlakis, G.N. (1990) Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J Virol*, **64**, 5448-5456.

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**, 207-219.

Ségault, V., Will, C.L., Polycarpou-Schwarz, M., Mattaj, I.W., Branlant, C. and Lührmann, R. (1999) Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic steps of pre-mRNA splicing in HeLa nuclear extracts. *Mol. Cell. Biol.*, **19**, 2782-2790.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A. and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell*, **11**, 965-976.

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**, 2089-2098.

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**, 363-373.

Shen, H. and Green, M.R. (2007) RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nat Struct Mol Biol*, **14**, 597-603.

Shen, H., Kan, J.L. and Green, M.R. (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell*, **13**, 367-376.

Shin, C. and Manley, J.L. (2002) The SR protein SRp38 represses splicing in M phase cells. *Cell*, **111**, 407-417.

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**, 15498-15503.

Shojania, S. and O'Neil, J.D. (2006) HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1-72) by NMR spectroscopy. *J Biol Chem*, **281**, 8347-8356.

Shore, S.M., Byers, S.A., Maury, W. and Price, D.H. (2003) Identification of a novel isoform of Cdk9. *Gene*, **307**, 175-182.

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**, 861-867.

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**, 7353-7362.
Simard, M.J. and Chabot, B. (2002) SRp30c is a repressor of 3' splice site utilization. *Mol Cell Biol*, **22**, 4001-4010.
Skoko, N., Baralle, M., Buratti, E. and Baralle, F.E. (2008) The pathological splicing mutation c.6792C>G in NF1 exon 37 causes a change of tenancy between antagonistic splicing factors. *FEBS Lett*, **582**, 2231-2236.
Sleasman, J.W. and Goodenow, M.M. (2003) 13. HIV-1 infection. *J Allergy Clin Immunol*, **111**, S582-592.
Smith, C.A., Crotty, S., Harada, Y. and Frankel, A.D. (1998) Altering the context of an RNA bulge switches the binding specificities of two viral Tat proteins. *Biochemistry*, **37**, 10808-10814.
Smith, C.W. and Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, **25**, 381-388.
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**, 44816-44825.
Smith, V. and Barrell, B.G. (1991) Cloning of a yeast U1 snRNP 70K protein homologue: functional conservation of an RNA-binding domain between humans and yeast. *Embo J*, **10**, 2627-2634.
Sontheimer, E.J. and Steitz, J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989-1996.
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**, 539-542.
Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D. and Abelson, J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell*, **9**, 31-44.
Stickeler, E., Fraser, S.D., Honig, A., Chen, A.L., Berget, S.M. and Cooper, T.A. (2001) The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *Embo J*, **20**, 3821-3830.
Stopak, K., de Noronha, C., Yonemoto, W. and Greene, W.C. (2003) HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell*, **12**, 591-601.
Stutz, F. and Izaurralde, E. (2003) The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell Biol*, **13**, 319-327.
Stutz, F. and Rosbash, M. (1998) Nuclear RNA export. *Genes Dev*, **12**, 3303-3319.
Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J. and Soret, J. (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *Embo J*, **20**, 1785-1796.
Svejstrup, J.Q., Vichi, P. and Egly, J.M. (1996) The multiple roles of transcription/repair factor TFIIH. *Trends Biochem Sci*, **21**, 346-350.
Swanson, M.S. and Dreyfuss, G. (1988) Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell. Biol.*, **8**, 2237-2241.
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. USA*, **94**, 1148-1153.
Tacke, R. and Manley, J.L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.*, **14**, 3540-3551.

Tacke, R. and Manley, J.L. (1999) Determinants of SR protein specificity. *Curr Opin Cell Biol*, **11**, 358-362.
Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93**, 139-148.
Tanackovic, G. and Kramer, A. (2005) Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo. *Mol Biol Cell*, **16**, 1366-1377.
Tang, H., Kuhen, K.L. and Wong-Staal, F. (1999) Lentivirus replication and regulation. *Annu Rev Genet*, **33**, 133-170.
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**, 1869-1883.
Taube, R., Fujinaga, K., Wimmer, J., Barboric, M. and Peterlin, B.M. (1999) Tat transactivation: a model for the regulation of eukaryotic transcriptional elongation. *Virology*, **264**, 245-253.
Teigelkamp, S., Mundt, C., Achsel, T., Will, C.L. and Luhrmann, R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *Rna*, **3**, 1313-1326.
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**, 184-190.
Trinkle-Mulcahy, L., Sleeman, J.E. and Lamond, A.I. (2001) Dynamic targeting of protein phosphatase 1 within the nuclei of living mammalian cells. *J Cell Sci*, **114**, 4219-4228.
Tsui, S., Dai, T., Roettger, S., Schempp, W., Salido, E.C. and Yen, P.H. (2000) Identification of two novel proteins that interact with germ-cell-specific RNA-binding proteins DAZ and DAZL1. *Genomics*, **65**, 266-273.
Turner, B.G. and Summers, M.F. (1999) Structural biology of HIV. *J Mol Biol*, **285**, 1-32.
Turner, I.A., Norman, C.M., Churcher, M.J. and Newman, A.J. (2004) Roles of the U5 snRNP in spliceosome dynamics and catalysis. *Biochem Soc Trans*, **32**, 928-931.
Tyagi, M., Rusnati, M., Presta, M. and Giacca, M. (2001) Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans. *J Biol Chem*, **276**, 3254-3261.
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**, 49773-49779.
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**, 1212-1215.
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**, 1706-1709.
Vassileva, M.T. and Matunis, M.J. (2004) SUMO modification of heterogeneous nuclear ribonucleoproteins. *Mol Cell Biol*, **24**, 3623-3632.
Vendel, A.C. and Lumb, K.J. (2004) NMR mapping of the HIV-1 Tat interaction surface of the KIX domain of the human coactivator CBP. *Biochemistry*, **43**, 904-908.
Vidal, V.P., Verdone, L., Mayes, A.E. and Beggs, J.D. (1999) Characterization of U6 snRNA-protein interactions. *Rna*, **5**, 1470-1481.
Wagner, E. and Lykke-Andersen, J. (2002) mRNA surveillance: the perfect persist. *J Cell Sci*, **115**, 3033-3038.
Wagner, E.J. and Garcia-Blanco, M.A. (2001) Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol*, **21**, 3281-3288.
Wang, H.Y., Lin, W., Dyck, J.A., Yeakley, J.M., Songyang, Z., Cantley, L.C. and Fu, X.D. (1998) SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol*, **140**, 737-750.

Wang, W.K., Chen, M.Y., Chuang, C.Y., Jeang, K.T. and Huang, L.M. (2000) Molecular biology of human immunodeficiency virus type 1. *J Microbiol Immunol Infect*, **33**, 131-140.

Wassarman, D.A. and Steitz, J.A. (1993) A base-pairing interaction between U2 and U6 small nuclear RNAs occurs in > 150S complexes in HeLa cell extracts: implications for the spliceosome assembly pathway. *Proc Natl Acad Sci U S A*, **90**, 7139-7143.

Weeks, K.M. and Crothers, D.M. (1991) RNA recognition by Tat-derived peptides: interaction in the major groove? *Cell*, **66**, 577-588.

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**, 451-462.

Westendorp, M.O., Frank, R., Ochsenbauer, C., Stricker, K., Dhein, J., Walczak, H., Debatin, K.M. and Krammer, P.H. (1995) Sensitization of T cells to CD95-mediated apoptosis by HIV-1 Tat and gp120. *Nature*, **375**, 497-500.

Westover, K.D., Bushnell, D.A. and Kornberg, R.D. (2004) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*, **119**, 481-489.

Wiegand, H.L., Lu, S. and Cullen, B.R. (2003) Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. *Proc Natl Acad Sci U S A*, **100**, 11327-11332.

Wiegers, K., Rutter, G., Kottler, H., Tessmer, U., Hohenberg, H. and Krausslich, H.G. (1998) Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol*, **72**, 2846-2854.

Will, C.L. and Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol*, **13**, 290-301.

Willbold, D., Metzger, A.U., Sticht, H., Gallert, K.C., Voit, R., Dank, N., Bayer, P., Krauss, G., Goody, R.S. and Rosch, P. (1998) Equine infectious anemia virus transactivator is a homeodomain-type protein. *J Mol Biol*, **277**, 749-755.

Willbold, D., Rosin-Arbesfeld, R., Sticht, H., Frank, R. and Rosch, P. (1994) Structure of the equine infectious anemia virus Tat protein. *Science*, **264**, 1584-1587.

Wimmer, J., Fujinaga, K., Taube, R., Cujec, T.P., Zhu, Y., Peng, J., Price, D.H. and Peterlin, B.M. (1999) Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b. *Virology*, **255**, 182-189.

Wolff, T. and Bindereif, A. (1992) Reconstituted mammalian U4/U6 snRNP complements splicing: a mutational analysis. *EMBO J.*, **11**, 345-359.

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**, 247-257.

Wolin, S.L. and Cedervall, T. (2002) The La protein. *Annu Rev Biochem*, **71**, 375-403.

Wolin, S.L. and Matera, A.G. (1999) The trials and travels of tRNA. *Genes Dev*, **13**, 1-10.

Wollerton, M.C., Gooding, C., Robinson, F., Brown, E.C., Jackson, R.J. and Smith, C.W. (2001) Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB). *Rna*, **7**, 819-832.

Wu, J.Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**, 1061-1070.

Wu, S., Romfo, C.M., Nilsen, T.W. and Green, M.R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, **402**, 832-835.

Wyatt, J.R., Sontheimer, E.J. and Steitz, J.A. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev.*, **6**, 2542-2553.

Xiao, S.H. and Manley, J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev*, **11**, 334-344.

Xiao, S.H. and Manley, J.L. (1998) Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *Embo J*, **17**, 6359-6367.

Xu, Y.Z., Newnham, C.M., Kameoka, S., Huang, T., Konarska, M.M. and Query, C.C. (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *Embo J*, **23**, 376-385.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T. and Handa, H. (2006) P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell*, **21**, 227-237.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J. and Handa, H. (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, **97**, 41-51.

Yamamoto, R., Katahira, M., Nishikawa, S., Baba, T., Taira, K. and Kumar, P.K. (2000) A novel RNA motif that binds efficiently and specifically to the Ttat protein of HIV and inhibits the trans-activation by Tat of transcription in vitro and in vivo. *Genes Cells*, **5**, 371-388.

Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K. and Zhou, Q. (2005) Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell*, **19**, 535-545.

Yik, J.H., Chen, R., Nishimura, R., Jennings, J.L., Link, A.J. and Zhou, Q. (2003) Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol Cell*, **12**, 971-982.

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**, 577-587.

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M. and Corden, J.L. (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A*, **93**, 6975-6980.

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**, 10077-10084.

Zamore, P.D., Patton, J.G. and Green, M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, **355**, 609-614.

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**, 8290-8301.

Zhang, C., Hastings, M.L., Krainer, A.R. and Zhang, M.Q. (2007) Dual-specificity splice sites function alternatively as 5' and 3' splice sites. *Proc Natl Acad Sci U S A*, **104**, 15028-15033.

Zhang, J. and Crumpacker, C.S. (2002) Human immunodeficiency virus type 1 nucleocapsid protein nuclear localization mediates early viral mRNA expression. *J Virol*, **76**, 10444-10454.

Zhang, J., Tamilarasu, N., Hwang, S., Garber, M.E., Huq, I., Jones, K.A. and Rana, T.M. (2000) HIV-1 TAR RNA enhances the interaction between Tat and cyclin T1. *J Biol Chem*, **275**, 34314-34319.

Zhang, M., Zamore, P.D., Carmo-Fonseca, M., Lamond, A.I. and Green, M.R. (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc. Natl. Acad. Sci. USA*, **89**, 8769-8773.

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**, 5400-5408.

Zhang, Z. and Krainer, A.R. (2004) Involvement of SR proteins in mRNA surveillance. *Mol Cell*, **16**, 597-607.

Zhao, X. and Yu, Y.T. (2004) Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in Xenopus oocytes. *Rna*, **10**, 681-690.

Zhou, Q. and Yik, J.H. (2006) The Yin and Yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. *Microbiol Mol Biol Rev*, **70**, 646-659.

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**, 401-405.

Zhou, Z. and Reed, R. (1998) Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *Embo J*, **17**, 2095-2106.

Zuo, P. and Maniatis, T. (1996) The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev*, **10**, 1356-1368.

Résumé :

Ce travail de thèse a comporté deux parties distinctes : l'une porte sur l'étude du complexe TAR/Tat/cycline T1 impliqué dans la transactivation de la transcription de l'ARN du virus HIV-1, l'autre concerne différentes facettes de la régulation de l'épissage de l'ARN du virus HIV-1.

L'interaction de la protéine virale Tat avec l'élément TAR présent à l'extrémité 5' des ARN du virus HIV-1 d'une part, et la cycline T1, composant du complexe p-TEFb responsable de l'hyperphosphorylation de l'ARN polymérase II d'autre part, est primordial pour obtenir des ARN viraux de pleine longueur. Dans l'objectif de réaliser une étude structurale du complexe TAR/Tat/cycline T1, l'ARN TAR et un fragment de la cycline T1 ont été produits en grandes quantités. De nombreuses tentatives de complexation des trois partenaires (TAR, Tat et cycline T1) ont été effectuées, mais la qualité des cristaux n'était pas suffisante pour une étude radiocristallographique du complexe.

L'épissage est une étape majeure du cycle de multiplication du virus HIV-1. Son ARN comporte 5 sites donneurs et 8 sites accepteurs dont l'utilisation combinée permet la production des 9 ORF virales. Les variations de l'épissage alternatif de l'ARN du virus HIV-1 en fonction de l'expression de protéines virales Tat et Nef ont été étudiées. Nous avons par ailleurs étudié l'effet des protéines SR sur l'utilisation des sites accepteurs A2 et A3. L'étude fine de l'élément régulateur ESEt du site A3 a révélé l'implication de la protéine SR 9G8 dans le schéma complexe de régulation de ce site.

Mots-clés:

Virus HIV-1, complexe TAR/Tat/cycline T1, épissage alternatif, protéines SR, élément ESE