



HAL
open science

Etude expérimentale du rôle des protéines SR dans la régulation de l'épissage de l'ARN du virus HIV-1, responsable de l'immunodéficience humaine, et modélisation mathématique de ces régulations

Delphine Ropers

► **To cite this version:**

Delphine Ropers. Etude expérimentale du rôle des protéines SR dans la régulation de l'épissage de l'ARN du virus HIV-1, responsable de l'immunodéficience humaine, et modélisation mathématique de ces régulations. Biochimie, Biologie Moléculaire. Université Henri Poincaré - Nancy 1, 2003. Français. NNT : 2003NAN10177 . tel-01748098

HAL Id: tel-01748098

<https://hal.univ-lorraine.fr/tel-01748098>

Submitted on 29 Mar 2018

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

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



AVERTISSEMENT

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

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

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

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

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 1

FACULTÉ DES SCIENCES & TECHNIQUES

U.F.R. Sciences & Techniques Biologiques
Ecole Doctorale BioSE

S.C.D. - U.H.P. NANCY
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique - BP 11
54007 VILLERS-LES-NANCY CEDEX

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Biologie structurale moléculaire et cellulaire

par **Delphine ROPERS**

Etude expérimentale du rôle des protéines SR dans la régulation de l'épissage de l'ARN du virus HIV-1, responsable de l'immunodéficience humaine, et modélisation mathématique de ces régulations

Soutenance publique prévue le 22 octobre 2003

Membres du jury :

Président :	M. James STÉVENIN	Directeur de recherche INSERM, Illkirch
Rapporteurs :	Mme Joëlle MARIE	Directeur de recherche INSERM, Gif-sur-Yvette
	M. Robert VIGNE	Directeur de recherche INSERM, Marseille
Examineurs :	M. Alexander BOCKMAYR	Professeur, UHP, Nancy I
	Mme Christiane BRANLANT	Directeur de recherche CNRS, UHP, Nancy I (Directeur de thèse)

Remerciements

En premier lieu, je tiens à remercier ma directrice de thèse, Mme Christiane Branlant. Ses conseils, sa rigueur et son orientation constante vers des pistes qui nous ont permis d'aboutir à des résultats originaux, m'ont aidée tout au long du travail expérimental et de la rédaction de ce manuscrit. Je la remercie vivement pour la formation scientifique qu'elle m'a apportée.

J'adresse toute ma gratitude à Joëlle Marie et Robert Vigne d'avoir accepté d'être rapporteurs de ce mémoire, ainsi qu'à James Stévenin et Alexander Bockmayr pour m'avoir fait l'honneur de juger ce travail.

Merci à Sentob Saragosti, pour nous avoir donné des fragments d'ADNc des souches HIV-1 YBF30 et SIVcpz ANT, et Martin Stoltzfus, pour nous avoir fourni le plasmide p Δ PSP.

Une partie de cette thèse fait une incursion dans le monde de la modélisation mathématique. Merci à Alexander Bockmayr et Damien Eveillard qui m'ont permis d'acquérir des connaissances de base en modélisation. Merci à Damien pour nos nombreuses discussions scientifiques et pour sa relecture attentive de la partie modélisation du manuscrit.

J'adresse un grand merci à Renata Gattoni et James Stévenin pour l'agréable collaboration que nous avons eue et pour nous avoir fourni de l'extrait cytoplasmique S100 et des protéines SR, sur lesquelles toute notre étude a reposé.

Un grand merci à Lilia Ayadi pour nos nombreuses discussions scientifiques et amicales. Au cours des années, son expérience et sa rigueur, aussi bien dans les études réalisées avec les protéines SR, que dans la rédaction finale de ce manuscrit, m'ont été infiniment précieuses.

Je tiens également à remercier Denise Thomas, que j'ai formée aux techniques de la biologie moléculaire. Elle a rapidement pris son élan et a apporté une qualité remarquable dans la mise en place et la réalisation des expériences. Merci pour sa présence et son aide pendant toute la durée de ma thèse. Je remercie également Annie Mougín, pour m'avoir initiée aux méthodes d'étude de la structure secondaire des ARN, ainsi que Laurence Damier et Sandrine Jacquenet, pour les constructions génétiques qu'elles avaient réalisées et qui ont constitué mon matériel de base.

Je remercie vivement l'équipe des correcteurs qui ont traqué sans relâche les fautes d'orthographe et les non-sens, et qui ont hautement contribué à améliorer le manuscrit. En particulier, j'adresse un grand merci à Virginie Marchand et Houda Hallay, pour leur aide si précieuse dans les dernières corrections.

Merci à tous les membres du laboratoire, pour nos nombreuses discussions scientifiques ou amicales, et sans lesquels chaque jour aurait été semblable à la veille. En particulier, merci à Nathalie Marmier-Gourrier, avec qui j'ai tout partagé !

Je garde pour la fin les personnes de mon cœur, sans qui la vie serait un désert. A ma famille et mes amis, qu'ils soient remerciés pour leur soutien inconditionnel.

Loin de l'image d'Epinal du face à face solitaire entre le doctorant et le sujet de thèse, cette thèse est donc avant tout le fruit d'une aventure collective. Je n'aurais jamais pu réaliser ce travail, sans l'aide et le soutien de toutes ces personnes, dont la générosité, la bonne humeur et l'intérêt manifesté à l'égard de ma recherche, m'ont été fort précieux.

Avant-propos

Depuis sa découverte en 1983, le virus HIV-1 a infecté 42 millions de personnes à travers le monde. En dépit de la découverte et des progrès qu'a permis de faire la trithérapie, la recherche de nouveaux agents thérapeutiques et d'un vaccin est indispensable pour limiter la progression, voire éradiquer, le virus. Dans cet objectif, une compréhension fine de chaque étape du cycle viral est primordiale. C'est dans cette démarche que s'inscrit le travail de l'équipe de Mme Branlant, au sein de laquelle j'ai préparé ma thèse. Nous nous intéressons aux mécanismes moléculaires impliqués dans la régulation de l'épissage alternatif de l'ARN viral.

L'épissage alternatif est, pour HIV-1, comme pour d'autres virus, un moyen de produire une grande variété de protéines et d'en contrôler l'expression, ceci afin d'optimiser la multiplication virale. Cette régulation de l'épissage repose en particulier sur la reconnaissance de séquences de l'ARN viral par des facteurs protéiques cellulaires. Nous nous sommes investis dans l'identification de ces séquences et de leur partenaires protéiques.

Sachant que l'épissage de l'ARN du virus HIV-1 est assuré par la machinerie de la cellule infectée, les mécanismes de base de sa régulation étaient supposés être similaires à ceux des ARN pré-messagers des eucaryotes supérieurs. Par conséquent, je présenterai, dans la revue bibliographique de ce manuscrit, l'épissage de ces ARN pré-messagers eucaryotes et les mécanismes déjà connus comme assurant la régulation de l'étape d'expression des gènes. Je ferai une présentation générale sur le virus HIV-1, en particulier, les étapes du cycle viral et un bilan des connaissances sur l'épissage de l'ARN viral et sa régulation.

Abréviations

A	adénine
ADN	acide désoxyribonucléique
ADNc	acide désoxyribonucléique complémentaire
AMV	« avian myeloblastosis virus »
ARN	acide ribonucléique
ARNm	acide ribonucléique messenger
ARNt	acide ribonucléique de transfert
ASF/SF2	« alternative splicing factor/splicing factor 2 »
ATP	adénosine triphosphate
BSA	« bovin serum albumin »
C	cytosine
Ci/mCi	curie/millicurie
CMCT	1-cyclohexyl-3-(2-morpholinoéthyl) carbodiimide métha-p-toluène
cpm	coups par minute
Da/kDa	dalton/kilodalton
dATP	désoxyadénosine triphosphate
dCTP	désoxycytosine triphosphate
ddNTP	didésoxyribonucléotide
dGTP	désoxyguanosine triphosphate
DMS	diméthylsulfate
DNase	désoxyribonucléase
dNTP	désoxyribonucléotide triphosphate
DTT	dithiothréitol
dTTP	désoxythymidine triphosphate
EDTA	éthylène diamine tétra-acétate
EN	extrait nucléaire
G	guanine
HeLa	Helen Laug
HEPES	acide (N-[2-hydroxyéthyl] piperazine-N'-[2-éthanesulfonique])
hnRNP	« heterogeneous ribonucleoprotein »
HIV	« human immunodeficiency virus »
kb	kilobase
m ^{2,2,7} G	(2, 2, 7) tri-méthyl guanosine
m ⁷ G	7-méthyl guanosine

nt(s)	nucléotide(s)
ORF	« open reading frame » (phase ouverte de lecture)
pb	paire de bases
PCR	« polymerase chain reaction »
PMSF	paraméthylsulfonate
PPT	« polypyrimidine tract »
p/v	poids/volume
qsp	quantité suffisante pour
R	purine
RNase	ribonucléase
rNTP	ribonucléotide triphosphate
rpm	rotation par minute
RT	transcriptase inverse
S	unité de sédimentation Svedberg
RS	arginine-sérine
sb	simple-brin
SDS	sodium dodécylsulfate
UTP	uridine triphosphate
SA	site accepteur
SD	site donneur
SIV	« simian immunodeficiency virus »
T	thymine
TBE	tampon tris-borate-EDTA
TE	tampon tris-EDTA
Tris	tris (hydroxyméthyl) aminométhane
U	uracile
UsnRNA	« uridine rich small nuclear ribonucleic acid »
UsnRNP	« uridine rich small nuclear ribonucleoprotein particle »
UV	ultra-violet
vs	<i>versus</i>
v/v	volume/volume
Y	pyrimidine
3'ss	« 3' splice site » (site 3' d'épissage)
5'ss	« 5' splice site » (site 5' d'épissage)

Sommaire

PARTIE I	REVUE BIBLIOGRAPHIQUE	15
I	Généralités sur l'épissage des ARN pré-messagers nucléaires	17
I.1	Réaction d'épissage des ARN pré-messagers nucléaires	18
I.2	Séquences nécessaires en <i>cis</i> au sein des ARN pré-messagers	19
I.3	Machinerie de base assurant l'épissage des introns majeurs de vertébrés	20
I.3.1	Particules snRNP	21
I.3.1.1	Particule snRNP U1	21
I.3.1.2	Particule snRNP U2	22
I.3.1.3	Particule [U4/U6,U5]	23
I.3.2	Assemblage du spliceosome et facteurs protéiques participant à cet assemblage	26
I.3.2.1	Reconnaissance du site 5' d'épissage	28
I.3.2.2	La reconnaissance du site 3'	35
I.3.2.3	Rapprochement des sites 5' et 3' d'épissage	37
I.3.2.4	Reconnaissance de la séquence de branchement par la snRNP U2	37
I.3.2.5	Incorporation de la tri-snRNP [U4/U6, U5] dans le spliceosome	40
I.3.3	Changements conformationnels conduisant aux complexes actifs	40
I.3.3.1	Transitions structurales	40
I.3.3.2	Rôle des ions divalents	43
I.3.4	Existence d'une structure active stable	44
I.3.5	Intégration de l'étape d'épissage dans le processus d'expression des gènes	44
I.3.5.1	Lien entre les étapes de transcription, épissage et polyadénylation	44
I.3.5.2	Lien entre l'épissage et l'export des ARN messagers	47
II	Régulation de l'épissage des ARN pré-messagers	49
II.1	Généralités sur l'épissage alternatif	49
II.1.1	Modes d'épissage alternatif	49
II.1.2	Les sites d'épissage alternatifs sont des sites d'épissage faibles	51
II.1.2.1	Caractéristiques des sites d'épissage faibles	51
II.1.2.2	L'utilisation des sites d'épissage faibles est régulée par des éléments en <i>cis</i>	51

II.2	Protéines nucléaires jouant un rôle majeur dans la régulation de l'épissage	53
II.2.1	Les protéines hnRNP et leurs sites de fixation	53
II.2.1.1	Protéines hnRNP A/B	55
II.2.1.2	Protéines hnRNP H/F	60
II.2.1.3	Protéine PTB ou hnRNP I	62
II.2.1.4	Protéine hnRNP C	64
II.2.1.5	Protéine TLS/FUS ou hnRNP P2	65
II.2.2	Les protéines SR et leurs sites de fixation	66
II.2.2.1	Identification des sites de fixation des protéines SR	66
II.2.2.2	Activation de l'épissage par les protéines SR fixées à des séquences exoniques	69
II.2.2.3	Activation de l'épissage par les protéines SR fixées à des séquences introniques	74
II.2.2.4	Inhibition de l'épissage par les protéines SR	74
II.2.3	Protéines SRm et Tra à domaine RS	75
II.2.3.1	Les coactivateurs d'épissage SRm160/300	75
II.2.3.2	Protéines Tra	76
II.3	Autres éléments de séquences et facteurs associés intervenant dans la régulation de l'épissage	77
II.3.1	Site donneur aval	77
II.3.2	Séquences riches en A et C	77
II.3.3	Séquences polypyrimidine localisées entre le site 5' et le point de branchement	78
II.3.3.1	Séquences riches en U localisées en aval d'un site 5' faible	78
II.3.3.2	Séquences de type CUCCCC et UUUCCC	79
II.3.4	Séquences CUG/UGC	79
II.3.4.1	Rôle dans l'activation de l'épissage	79
II.3.4.2	Rôle dans l'inhibition de l'épissage	79
II.3.5	Répétitions UG	79
II.3.6	Séquences introniques riches en G	80
II.3.7	Séquences (U)GCAUG	80
II.3.8	Séquence UCAY	81
II.4	Implication de pseudo-sites 5' et 3' dans les régulations de l'épissage	81
II.5	Implication des codons non-sens au sein des phases de lecture dans la régulation de l'épissage	82
II.6	Implication de la structure secondaire des ARN dans la régulation de l'épissage	83
II.6.1	Rôle dans l'activation de l'épissage	83
II.6.2	Rôle dans l'inhibition de l'épissage	84
II.7	L'épissage alternatif : un contrôle combinatoire	85
II.7.1	De multiples régulations contrôlent l'épissage de l'ARN pré-messager <i>c-src</i>	86
II.7.2	La structure du promoteur transcriptionnel régule l'activation de l'épissage de l'ARN pré-messager de la fibronectine par les protéines SR	86

II.7.3	L'épissage de l'ARN pré-messager <i>ADAR2</i> est régulé par édition	87
II.7.4	Régulation de l'expression et de l'activité des facteurs régulateurs d'épissage	88
II.7.4.1	Contrôle de la production des facteurs intervenant dans l'épissage	88
II.7.4.2	Contrôle de l'activité des facteurs d'épissage	89
III	Epissage de l'ARN du virus HIV-1	93
III.1	Description du virus HIV-1	93
III.1.1	Infection par le virus HIV-1	93
III.1.2	Evolution phylogénétique du virus HIV-1	94
III.1.3	Structure du virus HIV-1	95
III.1.4	ARN génomique	96
III.1.4.1	Régions non traduites	96
III.1.4.2	Phases ouvertes de lecture	97
III.1.5	Les protéines du virus HIV-1 et leur fonction dans le cycle de multiplication	97
III.1.5.1	Le cycle viral du virus HIV-1	97
III.1.5.2	Transcriptase inverse	100
III.1.5.3	Protéine Vpr	100
III.1.5.4	Intégrase	100
III.1.5.5	Protéine Tat	100
III.1.5.6	Protéine Rev	102
III.1.5.7	Protéines structurales et auxiliaires intervenant dans l'assemblage et la libération des virions	106
III.1.5.8	Protéine Nef	108
III.2	Régulation de l'épissage de l'ARN du virus HIV-1	108
III.2.1	Production des ARN messagers viraux	108
III.2.2	Impact de l'épissage sur l'activité des protéines virales et l'infectivité du virus	109
III.2.3	Mécanismes d'épissage de l'ARN viral	111
III.2.3.1	Les sites accepteurs d'épissage du virus HIV-1 sont faibles	111
III.2.3.2	Régulation du choix des sites d'épissage	113
	Objectifs et résumé des travaux	121
	 PARTIE II RÉSULTATS	 123
IV	Les protéines SR régulent l'épissage alternatif de l'ARN du virus HIV-1	125
IV.1	Etat des connaissances au départ et stratégie utilisée pour l'étude	125
IV.1.1	Etat des connaissances	125
IV.1.2	Stratégie générale de l'étude	125

IV.2	Etude de l'effet de protéines SR purifiées sur l'efficacité d'épissage des sites A1 à A5	126
IV.2.1	Choix des protéines SR testées	126
IV.2.2	Choix des ARN utilisés pour l'étude <i>in vitro</i>	127
IV.2.3	Choix de la construction utilisée pour l'étude <i>ex vivo</i>	128
IV.2.4	Choix des conditions d'épissage <i>in vitro</i>	129
IV.2.5	Résumé des résultats obtenus à partir de cette analyse <i>in vitro</i>	129
	Publication n° 1 - « 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 revealed by <i>in vitro</i> and <i>ex vivo</i> splicing experiments ».	
	Ropers D., Ayadi L., Gattoni R., Jacquenet S., Damier L., Branlant C. and Stévenin J.	
	En révision pour publication dans J. Biol. Chem.	131
IV.2.6	Expériences complémentaires	174
IV.2.6.1	Résultats d'épissage <i>in vitro</i> obtenus dans le mélange EN/S100 1/7	174
IV.2.6.2	Effet des protéines SR sur la compétition <i>in vitro</i> des sites A1 et A2, d'une part, et A2 et A3, d'autre part	177
IV.3	Conclusions et discussion des résultats	178
IV.3.1	Les protéines SR régulent l'épissage de l'ARN du virus HIV-1, en activant spécifiquement des sites accepteurs d'épissage	178
IV.3.2	Un nouvel exemple de régulation de l'épissage des ARNm viraux par les protéines SR	179
IV.3.3	Impact possible du contrôle de l'épissage alternatif de l'ARN de HIV-1 par les protéines SR, sur la production des protéines virales	180
IV.3.4	Le contrôle de l'expression des protéines SR permettrait au virus HIV-1 de réguler l'expression de ses gènes	182
IV.3.5	Les protéines SR réguleraient l'épissage de l'ARN de HIV-1 en reconnaissant des éléments en <i>cis</i>	183
V	Recherche de nouveaux éléments régulant en <i>cis</i> l'épissage de l'ARN viral	185
V.1	Stratégie de l'étude	185
V.2	Méthodologie expérimentale	185
V.3	Etude du site A2	188
V.3.1	Etat des données lorsque nous avons entrepris ce travail	188
V.3.2	Etude du mécanisme d'activation du site A2 par la protéine ASF/SF2	188
V.3.2.1	Etude de l'effet de délétion ou mutation dans les structures tige-boucles 3 et 4	188
V.3.2.2	Etude de la fixation des protéines ASF/SF2 et hnRNP A1 sur la structure tige-boucle 5 en aval du site A2 par expériences d'empreintes	191

	V.3.2.3	Effet de la mutation de l'élément ESSv sur l'activation du site A2 par ASF/SF2	193
	V.3.3	Conclusions et discussion des résultats	198
V.4		Etude des sites A3 à A5	200
	V.4.1	Etat des connaissances sur le site A3 lorsque j'ai débuté cette étude . . .	200
	V.4.2	Etude de la structure secondaire de la région de l'ARN de HIV-1 BRU renfermant les sites A3 à A5 et étude de l'interaction des composants de l'extrait nucléaire avec cette structure	202
	V.4.2.1	Questions posées	202
	V.4.2.2	Résultats obtenus	203
		Publication n° 2 - « Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its <i>cis</i> -regulatory element : possible involvement in RNA splicing ». Jacquet S., Ropers D., Bilodeau P. S., Damier L., Mougin A., Stoltzfus C.M. and Branlant C (2001) <i>Nucleic Acids Research</i> 29 (2), 464-478.	207
	V.4.2.3	Résultats complémentaires	224
	V.4.3	La séquence riche en purines, située entre les positions 5398 et 5403, est un élément activateur d'épissage	227
	V.4.3.1	L'activation du site A3 par la protéine ASF/SF2 nécessite la présence de la séquence 5398 à 5403 riche en purines	227
	V.4.3.2	L'élément ESEt a des propriétés activatrices d'épissage dans un contexte hétérologue	227
	V.4.4	La protéine ASF/SF2 se fixe à l'élément ESEt	230
	V.4.4.1	L'élément ESEt est important pour la fixation d'ASF/SF2 sur l'ARN	230
	V.4.4.2	Identification du site d'ancrage de la protéine ASF/SF2 sur l'ARN S2.3	230
	V.4.5	La protéine SC35 se fixe aux éléments ESEt et ESS2	234
	V.4.6	Etude des inter-relations entre les éléments régulateurs agissant sur le site A3	237
	V.4.6.1	La présence d'un excès de protéines SR peut compenser l'absence de site D4	237
	V.4.6.2	Rôle des séquences inhibitrices ESS2p et ESS2 dans la régulation de l'utilisation du site A3 par les protéines SR	239
	V.4.6.3	Recherche d'une inter-relation possible entre l'élément ESEt et les éléments inhibiteurs ESS2p et ESS2	240
	V.4.7	Des substitutions de bases mimant des mutations observées dans l'ARN viral de cellules chroniquement infectées affectent l'épissage au site A3	244
	V.4.8	Conclusions et discussion des résultats	246
	V.4.8.1	Le site A3 est soumis à un réseau complexe de régulations par des protéines nucléaires	246

V.4.8.2	L'édition de l'ADNc simple brin et peut être de l'ARN viral : un mécanisme de défense contre l'infection virale	249
VI	Etude du site A3 de l'ARN des virus SIVcpz ANT et HIV-1 YBF30	251
VI.1	Tous les sites d'épissage et de régulation ne sont pas conservés dans les souches HIV-1 YBF30 et SIVcpz ANT	252
VI.2	Comme pour HIV-1 BRU, l'utilisation du site A3 des ARN de HIV-1 YBF30 et de SIVcpz ANT est fortement activée par la protéine SC35	253
VI.3	L'absence d'élément ESS2 dans l'ARN de SIVcpz ANT explique sans doute la forte utilisation du site A3	256
VI.4	Comme pour HIV-1 BRU, le site donneur D4 est essentiel à l'utilisation du site A3 de SIVcpz ANT	257
VI.5	Etude expérimentale de la structure secondaire de la région de l'ARN du virus SIVcpz ANT contenant le site A3	257
VI.6	Conclusions et discussion des résultats	259
VII	Modélisation mathématique de la régulation de l'épissage	263
VII.1	Introduction sur la modélisation mathématique des phénomènes biologiques et objectifs de l'étude	263
VII.2	Modélisation de la régulation de l'épissage au site A3 de l'ARN du virus HIV-1 .	264
VII.2.1	Résumé des connaissances sur le site A3	265
VII.2.2	Hypothèses formulées sur le plan biologique	266
VII.2.2.1	Rôle de l'élément ESEt	266
VII.2.2.2	Rôle de l'élément ESS2	267
VII.2.2.3	Conceptualisation de la régulation du site A3	267
VII.2.3	Formalisation du modèle	268
VII.2.4	Comportement qualitatif du modèle	271
VII.2.5	Etude des propriétés mathématiques du modèle et simulations	272
VII.3	Extension du modèle mathématique à d'autres sites accepteurs de l'ARN du virus HIV-1	273
VII.4	Conclusions	276
VII.5	Annexes	278
	Publication n° 3 - « Multiscale modeling of alternative splicing regulation ». Eveillard D., Ropers D., De Jong H., Branlant C. and Bockmayr A. (2003) <i>Computational Methods in Systems Biology, CSMB'03</i> , C. Priami (ed.), LCNS 2602, Springer-Verlag, Berlin, 75-87.	279
	Publication n° 4 - « A multi-site constraint programming model of alternative spli- cing regulation ». Eveillard D., Ropers D., De Jong H., Branlant C. and Bockmayr A. En révision pour publication dans <i>Theoretical Computer Science</i>	295
	Discussion et perspectives	323

PARTIE III MATÉRIEL ET MÉTHODES	329
VIII Constructions génétiques	331
VIII.1 Constructions renfermant des séquences non contiguës du génome de la souche HIV-1 BRU	331
VIII.2 Constructions hybrides renfermant un fragment du génome HIV-1 BRU et un fragment du génome HIV-1 YBF30 ou SIVcpz ANT	331
VIII.3 Constructions contenant un fragment du génome de l'adénovirus	336
VIII.4 Réalisation de constructions génétiques variantes par mutagenèse dirigée	336
IX Méthodes spécifiques à l'étude de l'épissage <i>in vitro</i>	339
IX.1 Synthèse de transcrits radioactifs <i>in vitro</i>	339
IX.1.1 Préparation du fragment d'ADN à transcrire	339
IX.1.2 Transcription <i>in vitro</i>	339
IX.2 Epissage <i>in vitro</i>	340
IX.2.1 Réaction d'épissage	340
IX.2.2 Fractionnement sur gel et fluorographie	341
IX.2.3 Détermination de l'efficacité d'épissage	341
X Techniques d'analyse de la structure secondaire des ARN	343
X.1 Réaction de transcription <i>in vitro</i>	343
X.2 Séquençage de l'ARN	343
X.2.1 Principe	343
X.2.2 Marquage des oligodésoxyribonucléotides en 5'	344
X.2.3 Réaction de séquençage (Qu <i>et al.</i> 1983)	344
X.2.3.1 Hybridation de l'amorce oligonucléotidique	344
X.2.3.2 Réaction d'élongation	344
X.2.3.3 Fractionnement sur gel	345
X.3 Analyse de la structure secondaire des ARN par des sondes chimiques et enzymatiques	345
X.3.1 Principe	345
X.3.2 Conditions expérimentales	346
X.3.2.1 Modifications chimiques	346
X.3.2.2 Hydrolyses enzymatiques	346
X.3.3 Prédictions de structure secondaire par ordinateur	346
XI Techniques d'analyse des interactions ARN/protéines	347
XI.1 Technique d'empreinte des ARN en extrait nucléaire ou en présence de protéine purifiée	347
XI.1.1 Amplification par PCR des ADN matrices	347
XI.1.2 Synthèse de transcrits froids courts	347
XI.1.3 Principe de la technique d'empreinte des ARN	348

XI.1.4	Conditions expérimentales de l’empreinte d’un ARN en extrait nucléaire ou en présence de protéine purifiée	348
XI.1.4.1	Assemblage des complexes ARN/protéines	348
XI.1.4.2	Hydrolyses enzymatiques et modifications chimiques	349
XI.1.5	Technique de retard sur gel	349
XI.1.5.1	Synthèse de transcrits radioactifs courts	349
XI.1.5.2	Retard sur gel	350

Bibliographie		382
----------------------	--	------------

Première partie

Revue bibliographique

Chapitre I

Généralités sur l'épissage des ARN prémessagers nucléaires

En 1977, des gènes morcelés ont été mis en évidence pour la 1^{re} fois dans le génome de l'adénovirus (Berget *et al.* 1977 ; Berget & Sharp 1977). Depuis lors, la présence de séquences non codantes dans les gènes nucléaires de protéines s'est avérée être générale chez les eucaryotes. Des introns sont également présents dans les gènes d'ARN stables (ARN de transfert, ARN ribosomiques), dans les gènes d'organites cellulaires (chloroplastes, mitochondries), ainsi que dans certains gènes bactériens et d'archaea, mais les mécanismes d'élimination de ces séquences excédentaires sont différents de ceux mis en oeuvre pour les transcrits des gènes nucléaires codant des protéines. En dépit de ces différences de mécanisme, tous les processus d'élimination des séquences excédentaires appelées introns¹ sont rassemblés sous le terme épissage.

Ainsi, au cours de la réaction d'épissage, les séquences introniques des ARN prémessagers sont éliminées et les séquences codantes, dénommées exons², sont jointes les unes aux autres pour former l'ARN messenger. Les ARNm sont ensuite transportés dans le cytoplasme pour y être traduits en protéines.

Bien que chez certains eucaryotes unicellulaires tels que la levure *Saccharomyces cerevisiae*, moins de 5% des gènes nucléaires de protéines contiennent un intron éliminé par la machinerie d'épissage (Bon *et al.* 2003), tous les génomes eucaryotes sont caractérisés par la présence d'introns dans les gènes nucléaires de protéines.

Chez les eucaryotes supérieurs, l'épissage est un phénomène majeur. La plupart des gènes contiennent un intron. De plus, les introns de ces organismes sont en général de grande taille par rapport à ceux d'organismes tels que les levures ou les insectes. Ainsi chez l'homme, la taille moyenne d'un exon est de 75 à 150 nts, alors que les introns font en moyenne 3500 nts (Deutsch & Long 1999) et cette taille peut aller jusque 500 000 nts (Rowen *et al.* 2002).

L'ensemble des introns présents dans tous les types de gènes et chez tous les organismes a été subdivisé en 4 catégories selon les mécanismes d'excision mis en oeuvre :

- les introns des ARN de transfert, dont l'élimination est catalysée par des enzymes (pour revue, Wolin & Matera 1999),

¹ « intragenic regions »

² « expressed regions »

- 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 & Cech 2002),
- les introns autocatalytiques de groupes II et III (pour revues, Bonen & Vogel 2001 ; Doudna & 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),
- les introns nucléaires des ARN pré-messagers, auxquels nous nous sommes intéressés au cours de cette étude, qui sont éliminés par une machinerie nucléaire, le spliceosome.

Si les réactions chimiques mises en oeuvre lors de l'épissage des introns d'ARN pré-messagers sont similaires à celles des introns de groupe II, leur élimination se fait au sein d'un macrocomplexe ribonucléoprotéique, le spliceosome, constitué de cinq UsnRNP³ et de nombreux autres facteurs protéiques (pour revue, Will & Lührmann 2001). Les UsnRNP spliceosomales sont des particules constituées de petits ARN nucléaires, les snRNA U1, U2, U4, U5 et U6, associés à des protéines (pour revue, Will & Lührmann 2001). Ces particules s'assemblent sur les jonctions intron-exon et confèrent à l'ARN pré-messager une configuration spatiale adéquate pour permettre la réaction d'épissage.

L'épissage des ARN pré-messagers nucléaires ne constitue qu'une étape du processus de maturation. En effet, au fur et à mesure de leur synthèse par l'ARN polymérase II, ces ARN sont d'abord coiffés en 5', pour les protéger de l'action d'exonucléases 5'→3', puis épissés et enfin polyadénylés à leur extrémité 3'.

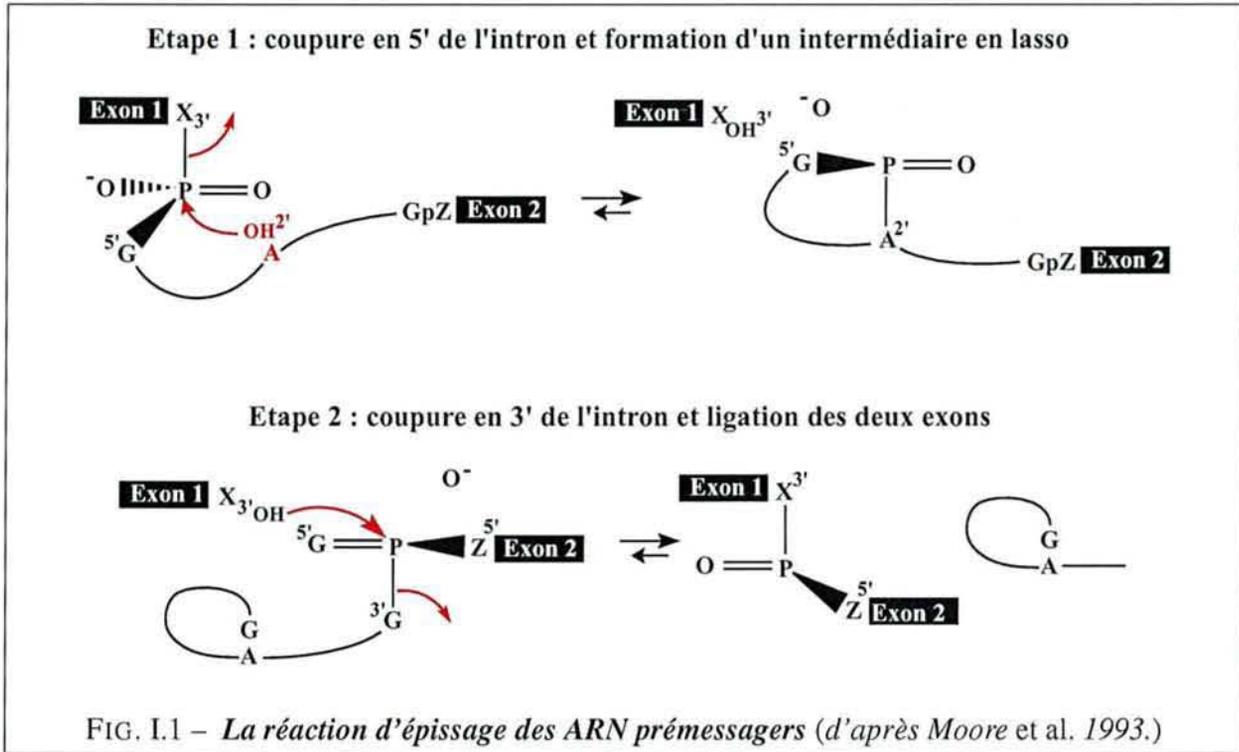
Il y a quelques années, il est apparu qu'il existe en fait 2 types d'introns chez les eucaryotes supérieurs : les introns dits majeurs, éliminés par le spliceosome que j'ai décrit précédemment, et les introns dits mineurs, rencontrés moins fréquemment. Ces introns sont éliminés par un autre type de spliceosome, renfermant les snRNA U11 et U12 à la place des snRNA U1 et U2, les snRNA U4atac et U6atac à la place des snRNA U4 et U6, et le snRNA U5 (pour revue, Tarn & Steitz 1997). Les introns de l'ARN du virus HIV-1 étant des introns majeurs, nous ne présenterons dans cette introduction que l'épissage des introns par le spliceosome majeur.

I.1 Réaction d'épissage des ARN pré-messagers nucléaires

Les introns des ARN pré-messagers sont éliminés par 2 étapes de transestérification (Fig. I.1). 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 l'intron ; ce résidu constitue le point de branchement. Deux intermédiaires réactionnels sont alors produits : l'exon 1, avec une extrémité 3'OH libre, et un intermédiaire contenant l'exon 2 et l'intron structuré en lasso.

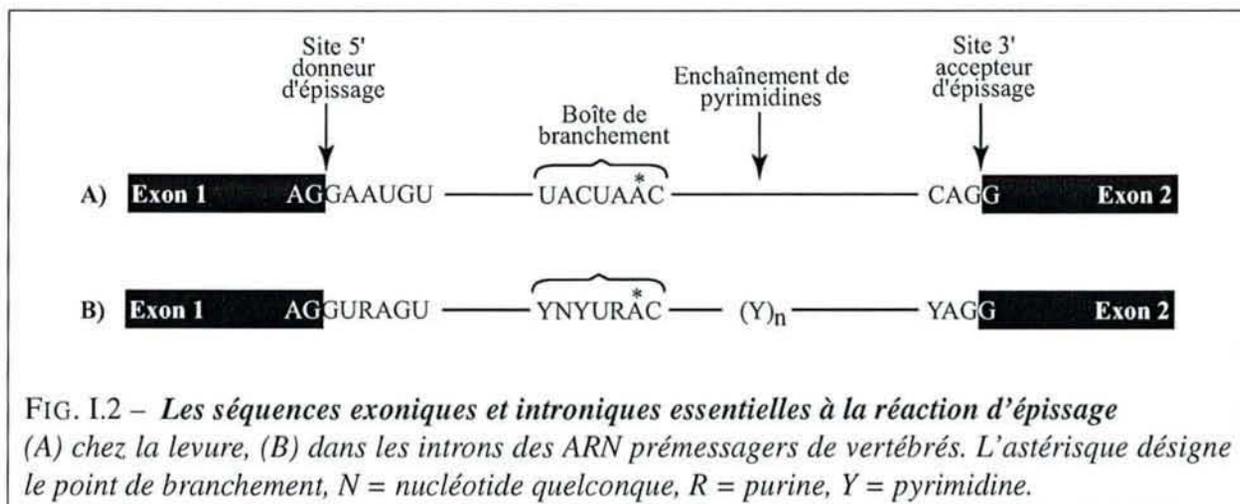
Dans la seconde étape, l'attaque nucléophile de l'extrémité 3'OH de l'exon 1 sur le phosphate en 3' de l'intron conduit à la ligation des deux exons et à la libération de l'intron sous forme de lasso. Ce sont les produits finaux de la réaction d'épissage (pour revue, Moore *et al.* 1993).

³« uridine rich small nuclear ribonucleoparticle »



I.2 Séquences nécessaires en *cis* au sein des ARN pré-messagers

Les jonctions intron-exon et le point de branchement constituent les séquences nécessaires à l'identification d'un intron par la machinerie d'épissage (le spliceosome), qui s'assemble autour de l'ARN pré-messager en reconnaissant ces séquences. Chez la levure, ces signaux de reconnaissance sont fortement conservés dans tous les ARN pré-messagers, alors que ces séquences sont beaucoup plus dégénérées chez les vertébrés (Fig. I.2).



La séquence consensus, établie par comparaison des sites 5' d'épissage des introns majeurs de vertébrés, est AG\GURAGU dans laquelle le dinucléotide GU est le plus conservé (R = purine, \ = jonction exon-intron ; pour revue, Moore *et al.* 1993). Les sites 3' sont eux caractérisés par la séquence YAG\G qui, chez les vertébrés, est précédée d'une séquence de 10 à 20 pyrimidines. Enfin, la 3^e séquence de

reconnaissance renferme le point de branchement, elle est située 18 à 40 nucléotides en amont du site 3' d'épissage. La séquence consensus établie pour les introns majeurs d'eucaryotes est YNYUR^{*}AC (Y = pyrimidine, R = purine, l'astérisque désigne le point de branchement) (Fig. I.2). L'adénosine est majoritairement utilisée comme site de branchement car c'est avec ce résidu que la réaction d'épissage a la meilleure efficacité, en particulier au niveau de la 2^e étape de la réaction (Hornig *et al.* 1986 ; Query *et al.* 1995, 1996). Comme mentionné ci-dessus, les séquences de reconnaissance sont plus conservées chez la levure *S. cerevisiae* avec la séquence AG\GAAUGU en 5' et UACUAC au site de branchement.

I.3 Machinerie de base assurant l'épissage des introns majeurs de vertébrés

L'élimination des introns majeurs de gènes nucléaires de vertébrés met en oeuvre cinq particules ribonucléoprotéiques (snRNP U1, U2, U4, U5, U6), un grand nombre de protéines et de facteurs protéiques spliceosomaux communs à la levure et aux vertébrés, ainsi que des facteurs protéiques spécifiques aux vertébrés, comme les protéines SR.

Les snRNP et les protéines associées au spliceosome reconnaissent les séquences nécessaires à l'épissage décrites dans le paragraphe précédent. Des interactions ARN-ARN et ARN-protéines se forment alors de manière ordonnée et transitoire autour des sites d'épissage, conduisant à la formation du spliceosome au sein duquel ont lieu les 2 réactions de transestérification (pour revues, Hastings & Krainer 2001 ; Will & Lührmann 2001).

La composition en protéines du spliceosome humain a d'abord été étudiée sur des complexes purifiés à partir d'extraits nucléaires de cellules HeLa, par des techniques de gel filtration et de chromatographie d'affinité (Reed 1990 ; Bennett *et al.* 1992). Au cours de ces dernières années, le développement de nouvelles approches de protéomique a permis de faire de grand progrès dans l'identification des composants du spliceosome. Les particules snRNP humaines et de levure ont pu être purifiées à faible concentration saline et l'analyse de leur contenu, par gel suivi d'une analyse par spectroscopie de masse, a permis l'identification de l'ensemble de leurs protéines, y compris celles associées par des interactions faibles (Neubauer *et al.* 1998 ; Gottschalk *et al.* 1999 ; Stevens & Abelson 1999 ; Will *et al.* 2001, 2002 ; Jurica *et al.* 2002 ; Hartmuth *et al.* 2002 ; pour revues, Kramer 1996 ; Kambach *et al.* 1999 ; Will & Lührmann 1997 ; Will & Lührmann 2001). Le même type d'étude a été mené sur des complexes spliceosomaux humains fonctionnels (Zhou *et al.* 2002). L'analyse du contenu en protéines de ces complexes a permis d'identifier environ 145 protéines, dont 58 sont des composants encore non décrits dans les spliceosomes. Environ 30 de ces protéines avaient été caractérisées comme intervenant dans des étapes de l'expression des gènes autres que l'épissage. Ceci s'explique par le fait que les étapes de transcription, formation de la coiffe, épissage et polyadénylation ont lieu de manière ordonnée au sein d'une macrostructure stabilisée par l'interaction d'un grand nombre de composants avec le domaine C-terminal de la grande sous-unité de l'ARN polymérase II.

Nous commencerons par décrire la machinerie spliceosomale de base, nous verrons ensuite les différentes étapes de son assemblage sur les ARN pré-messagers.

I.3.1 Particules snRNP

Les UsnRNA (U1, U2, U4, U5 et U6) de chaque particule ont une petite taille, sont très stables dans la cellule et sont fortement conservés chez les eucaryotes (Branlant *et al.* 1980 ; Krol *et al.* 1981*a, b* ; Branlant *et al.* 1982, 1983 ; Jacob *et al.* 1984 ; Myslinski *et al.* 1984, 1989 ; Myslinski & Branlant 1989 ; pour revue, Will & Lürmann 1997). Ils subissent de nombreuses modifications post-transcriptionnelles (Shimba *et al.* 1995 ; Massenet *et al.* 1999 ; Massenet & Branlant 1999 ; pour revue, Massenet *et al.* 1998), comme des méthylations et des pseudo-uridylation souvent importantes pour leur fonction (Ségault *et al.* 1995 ; Yu *et al.* 1998 ; Newby & Greenbaum 2001).

Certaines protéines des UsnRNP sont retrouvées dans toutes les particules. Ainsi, les snRNP U1, U2, U4 et U5, dont l'assemblage a lieu dans le cytoplasme, possèdent en commun 7 protéines Sm organisées en anneau autour de leur site de fixation sur l'ARN (site SM ; pour revues, Kambach *et al.* 1999 ; Will & Lührmann 2001). La snRNP U6, qui est assemblée dans le noyau, est associée à un ensemble de 7 protéines différentes, les protéines LSm⁴ (pour revue, He & Parker 2000).

Le rôle des UsnRNP dans l'épissage a été très étudié, tant chez la levure que chez les mammifères ; l'emploi de techniques de génétique de la levure a permis en particulier d'obtenir de nombreuses données. Comme les facteurs d'épissage généraux sont très conservés de la levure aux mammifères et jouent un même rôle dans les 2 types d'organisme, nous présenterons, dans cette 1^{re} partie de l'introduction, la fonction des protéines et des ARN impliqués dans l'épissage, en nous appuyant sur des travaux réalisés à la fois chez la levure et chez les mammifères.

I.3.1.1 Particule snRNP U1

La snRNP U1 est une petite particule contenant, outre les protéines Sm, les protéines U1-A, U1-70K et U1-C. Des études par emploi de sondes enzymatiques et chimiques (Branlant *et al.* 1981 ; Mount & Steitz 1981 ; Krol *et al.* 1990 ; Hartmuth *et al.* 1999), puis par microscopie électronique (Kastner & Lührmann 1989 ; Stark *et al.* 2001), ont permis de déterminer l'organisation structurale de la particule, représentée dans la figure I.3-A (pour revues, Kambach *et al.* 1999 ; Will & Lührmann 2001).

Les protéines U1-A, U1-70K et les protéines Sm sont directement ancrées sur l'ARN, alors que la protéine U1-C est essentiellement liée par interaction protéine-protéine.

Au sein de cette particule, la séquence en 5' très conservée du snRNA U1 joue un rôle majeur par son appariement avec l'extrémité 5' de l'intron. Deux protéines jouent également un rôle important dans la fixation de la snRNP U1 à l'ARN pré-messager : la protéine U1-70K qui, par son domaine RS, interagit comme on le verra après, avec le domaine RS des protéines SR, en particulier ASF/SF2. Par ailleurs, la protéine U1-C est capable d'interagir avec l'ARN pré-messager (Heinrichs *et al.* 1990 ; Nelissen *et al.* 1991) et des expériences récentes sur la levure *S. cerevisiae* suggèrent que cette interaction pourrait avoir lieu avant la mise en place de l'appariement de base entre le snRNA U1 et l'extrémité 5' de l'intron (Du & Rosbash 2002).

⁴« Sm-like »

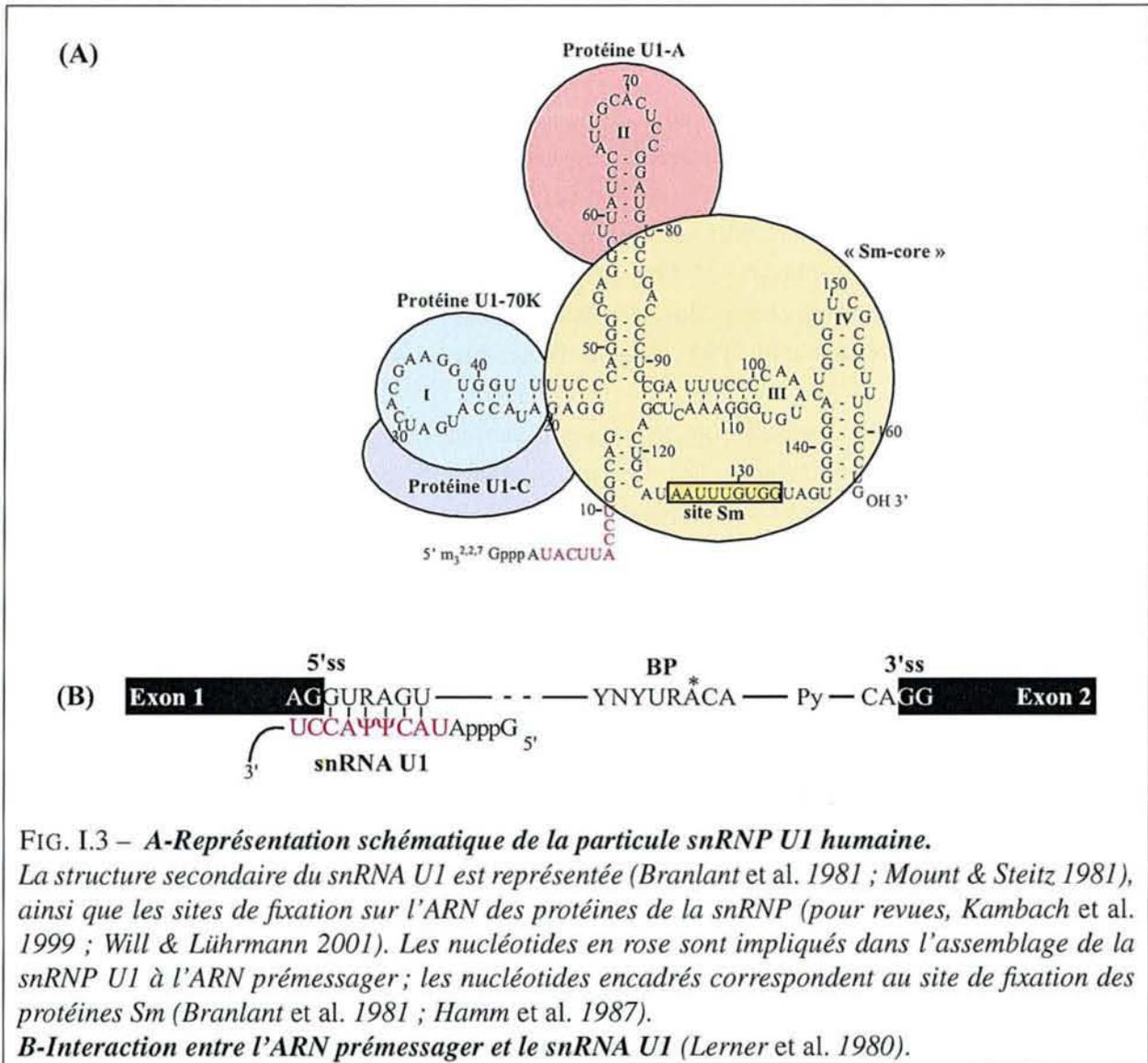


FIG. I.3 – A- Représentation schématique de la particule snRNP U1 humaine.

La structure secondaire du snRNA U1 est représentée (Branlant et al. 1981 ; Mount & Steitz 1981), ainsi que les sites de fixation sur l'ARN des protéines de la snRNP (pour revues, Kambach et al. 1999 ; Will & Lührmann 2001). Les nucléotides en rose sont impliqués dans l'assemblage de la snRNP U1 à l'ARN pré-messager ; les nucléotides encadrés correspondent au site de fixation des protéines Sm (Branlant et al. 1981 ; Hamm et al. 1987).

B-Interaction entre l'ARN pré-messager et le snRNA U1 (Lerner et al. 1980).

En plus de leur rôle dans l'épissage, il faut noter que les protéines U1-70K et U1-A de la snRNP U1 interviennent dans l'assemblage du complexe assurant le clivage et la polyadénylation à l'extrémité 3' des ARN pré-messagers (Gunderson *et al.* 1998 ; pour revue, Hall 2002).

I.3.1.2 Particule snRNP U2

Cette particule est impliquée dans la reconnaissance de la séquence de branchement de l'ARN pré-messager. A faible concentration saline, on purifie une particule de coefficient de sédimentation 17S contenant le snRNA U2, les protéines Sm, les protéines U2A' et U2B'' et l'ensemble des protéines correspondant aux 2 sous-unités SF3a et SF3b du facteur d'épissage SF3 (Figure I.4A ; Brosi *et al.* 1993b ; Kramer *et al.* 1999 ; pour revues, Kambach *et al.* 1999 ; Will & Lührmann 2001). Des données de microscopie électronique sont également disponibles pour la particule snRNP U2 (pour revue, Kambach *et al.* 1999).

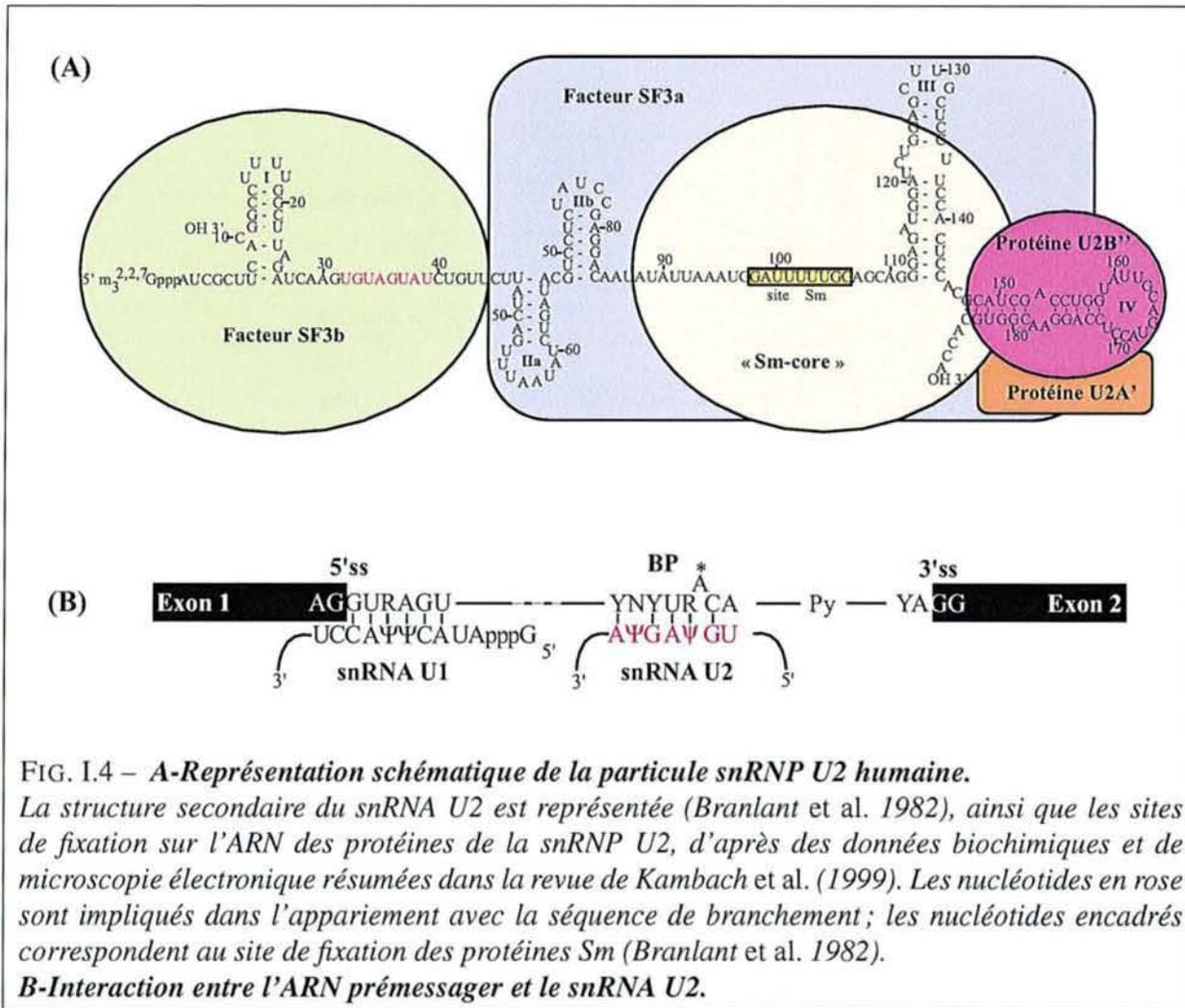


FIG. I.4 – **A- Représentation schématique de la particule snRNP U2 humaine.** La structure secondaire du snRNA U2 est représentée (Branlant et al. 1982), ainsi que les sites de fixation sur l'ARN des protéines de la snRNP U2, d'après des données biochimiques et de microscopie électronique résumées dans la revue de Kambach et al. (1999). Les nucléotides en rose sont impliqués dans l'appariement avec la séquence de branchement ; les nucléotides encadrés correspondent au site de fixation des protéines Sm (Branlant et al. 1982).

B-Interaction entre l'ARN pré-messager et le snRNA U2.

Les facteurs SF3a et SF3b sont constitués de 8 protéines (SAP⁵ 61, 62 et 114 pour SF3a et p14a, SAP 49, 130, 145 et 155 pour SF3b), jouant un rôle important dans l'assemblage du spliceosome (cf § I.3.2.4) (Kramer 1988 ; Kramer & Utans 1991 ; Brosi *et al.* 1993a, b ; Bennett & Reed 1993 ; Das *et al.* 1999 ; Kramer *et al.* 1999). Récemment, 3 nouvelles protéines du facteur SF3b ont été identifiées : p10, p14b et une protéine de la famille des ARN hélicases/RNPases, SF3b125 (Will *et al.* 2002). La protéine SF3b125 est présente dans le facteur SF3b, mais se dissocie du facteur lors de l'intégration de celui-ci dans la particule snRNP U2 17S mature (Will *et al.* 2002).

I.3.1.3 Particule [U4/U6,U5]

Seule la snRNP U4 est présente dans le cytoplasme. A son retour dans le noyau, elle s'associe à la snRNP U6 pour former la di-snRNP U4/U6 (Rinke *et al.* 1985). Cette dernière s'associe ensuite à la snRNP U5 pour former la tri-snRNP [U4/U6,U5] (Hashimoto & Steitz 1984 ; Behrens & Lührmann 1991). Au sein de cette particule, les snRNA U4 et U6 forment une structure en « Y » (Fig. I.5 ; Brow & Guthrie 1988 ; Mougin *et al.* 2002).

⁵ « splicing associated protein »

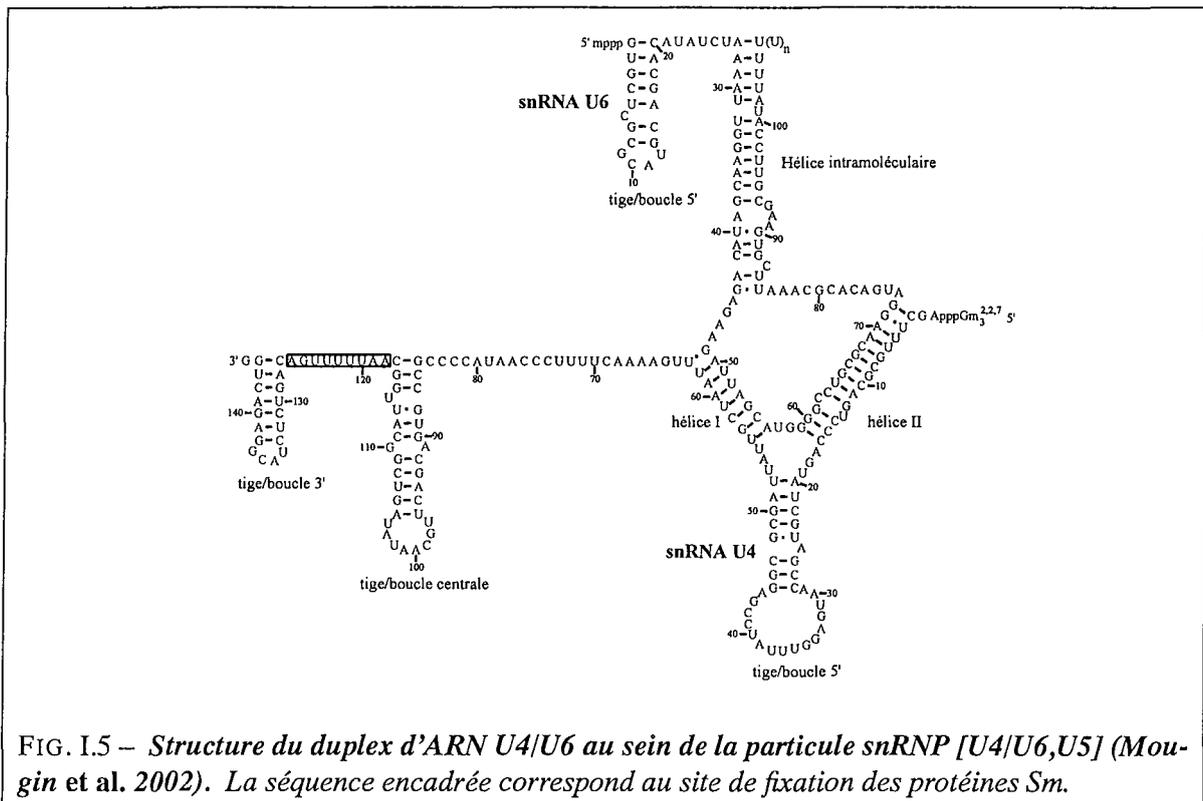


FIG. I.5 – Structure du duplex d'ARN U4/U6 au sein de la particule snRNP [U4/U6,U5] (Mougin et al. 2002). La séquence encadrée correspond au site de fixation des protéines Sm.

C'est sous leur forme tri-snRNP que les particules snRNP U4, U6 et U5 rejoignent le spliceosome. Un grand nombre de réarrangements conformationnels suit cette incorporation, au terme desquels le snRNA U6 se lie au site 5' d'épissage et s'associe avec le snRNA U2 (cf chapitre I.12).

Le snRNA U5 établit quant à lui diverses interactions avec l'ARN pré-messager. Il interagit d'abord avec la bordure 3' du 1^{er} exon, puis à l'issue de la 1^{re} réaction d'épissage, il interagit avec les 2 exons et les aligne dans une configuration adéquate pour permettre la 2^e réaction d'épissage (cf chapitre I.12).

A la différence des snRNP U1 et U2 dont la composition en protéines est assez réduite, la tri-snRNP [U4/U6,U5] contiendrait au moins 29 protéines (Tab. I.1 ; Zhou *et al.* 2002 ; pour revue, Will & Lürmann 1997). Certaines de ces protéines sont également présentes dans la di-snRNP U4/U6 ou la snRNP U5, tandis que d'autres sont spécifiques de la tri-snRNP.

Ainsi, en plus des protéines Sm, 7 protéines de la snRNP U5 ont été identifiées dans la tri-particule : les protéines de 15, 40, 100, 102, 116, 200 et 220 kDa. Parmi elles, la protéine U5-220 (ou hPrp8) interagissant avec le snRNA U5, l'ARN hélicase U5-200, la protéine U5-116 (une GTPase potentielle) et la protéine U5-40 interagissent entre elles pour former un complexe (Achsel *et al.* 1998).

En plus de ces protéines, la tri-snRNP contient des protéines spécifiques de la di-snRNP U4/U6 : la protéine de 61 kDa, essentielle pour la formation de la tri-snRNP (Makarova *et al.* 2002), la protéine de 15,5 kDa, qui interagit avec la tige-boucle 5' du snRNA U4 (Nottrott *et al.* 1999), et 3 protéines qui interagissent entre elles : U4/U6-20K (cyclophiline H), 60K et 90K (Horowitz *et al.* 1997 ; Teigelkamp *et al.* 1998).

Trois protéines spécifiques de la tri-snRNP [U4/U6,U5] sont également présentes : les protéines de 27, 65 et 110 kDa (Fetzer *et al.* 1997 ; Makarova *et al.* 2001). Bien que les protéines de 65 et 110

Protéine	snRNP U5	di-snRNP [U4/U6]	tri-snRNP [U4/U6,U5]
220K (hPrp8)	+	-	+
200K	+	-	+
116K	+	-	+
100K	+	-	+
52K	+	-	-
40K	+	-	+
15K	+	-	+
102K	+	-	+
110K	-	-	+
61K	-	-	+
90K	-	-	+
60K	-	-	+
15,5K	-	-	+
20K (cyclophiline H)	-	-	+
65K	-	-	+
27K	-	-	+
SmB/B'	+	+	+
SmD1	+	+	+
SmD2	+	+	+
SmD3	+	+	+
SmE	+	+	+
SmF	+	+	+
SmG	+	+	+
LSm2	-	-	+
LSm3	-	-	+
LSm4	-	-	+
LSm5	-	-	+
LSm6	-	-	+
LSm7	-	-	+
LSm8	-	-	+

TAB. I.1 – Composition protéique des particules U4, U4/U6 et [U4/U6,U5] humaines.

La présence ou l'absence de protéine est représentée par un signe + ou - (Zhou et al. 2002 ; pour revue, Will & Lürmann 1997).

kDa soient présentes dans la snRNP U5, elles sont systématiquement retrouvées dans la tri-snRNP. Ces protéines favorisent le recrutement de la tri-snRNP dans le spliceosome (Makarova *et al.* 2001). Les protéines LSM de la snRNP U6 sont également présentes dans la tri-particule.

Malgré les études par microscopie électronique des snRNP U4/U6, U5 et de la tri-snRNP [U4/U6,U5] (Kastner *et al.* 1990a, b, 1991), malgré les études par emploi de sondes enzymatiques et chimiques des régions des UsnRNA protégées par les protéines (Mougin *et al.* 2002), et les études des interactions ARN/protéines par pontage covalent (Vidal *et al.* 1999 ; Dix *et al.* 1998 ; Urlaub *et al.* 2001), la connaissance des sites d'interaction des protéines sur les snRNA U4, U5 et U6 est moins précise que dans le cas des snRNA U1 et U2.

Une fois l'assemblage des différentes particules snRNP achevé, celles-ci vont à leur tour s'assembler sur l'ARN prémessager, accompagnées d'un grand nombre de protéines spliceosomales. La structure ainsi formée entre l'ARN prémessager, les UsnRNA et l'ensemble des protéines portera l'activité catalytique.

I.3.2 Assemblage du spliceosome et facteurs protéiques participant à cet assemblage

L'assemblage du spliceosome autour de l'intron à éliminer repose sur des mécanismes de reconnaissance des sites d'épissage. Ce processus d'assemblage est une étape clé, souvent limitante, de l'épissage des ARN prémessagers. En effet, chez les eucaryotes supérieurs, la grande taille des introns (de 3500 nts en moyenne, jusqu'à 10^5 nts) (Hawkins 1988 ; Sterner *et al.* 1996 ; Deutsch & Long 1999), comparée à la petite taille des exons (100 à 150 nts) (Berget 1995) rend difficile le processus de recherche des sites d'épissage par la machinerie spliceosomale, d'où l'hypothèse émise par S. Berget d'une reconnaissance préalable des exons, avant la reconnaissance des introns (Berget 1995). Cette hypothèse a été démontrée par le fait que la mutation d'un site donneur ou accepteur entraîne l'exclusion de l'exon. On appelle « définition des exons » ce processus de reconnaissance des exons au sein de l'ARN prémessager (Berget 1995).

Chez les vertébrés, les 1^{ers} facteurs d'épissage s'associant à l'ARN prémessager ont pour rôle la définition des exons. Une fois les sites 5' et 3' d'épissage définis, les autres facteurs d'épissage s'assemblent à leur tour sur l'ARN prémessager. La définition du premier et du dernier exon de l'ARN prémessager est en revanche différente. Le premier exon est reconnu grâce à l'interaction entre les protéines reconnaissant la coiffe méthylée des ARN prémessagers et les facteurs d'épissage associés au premier site 5' (Ohno *et al.* 1987 ; Izaurralde *et al.* 1994 ; Lewis *et al.* 1996 ; pour revue, Zeng & Berget 2000). Le dernier exon est quant à lui défini par l'interaction entre les facteurs de polyadénylation et les facteurs d'épissage associés au site 3' (Niwa & Berget 1991a, b).

Les études *in vitro* de l'assemblage des complexes spliceosomaux sur les ARN prémessagers ont permis de mettre en évidence la formation successive de 5 complexes possédant une mobilité électrophorétique différente sur gel natif (pour revue, Jurica & Moore 2002) : les complexes H, E, A, B et C. Ces complexes correspondent à l'assemblage ordonné des particules snRNP et de protéines

spliceosomales sur les ARN pré-messagers.

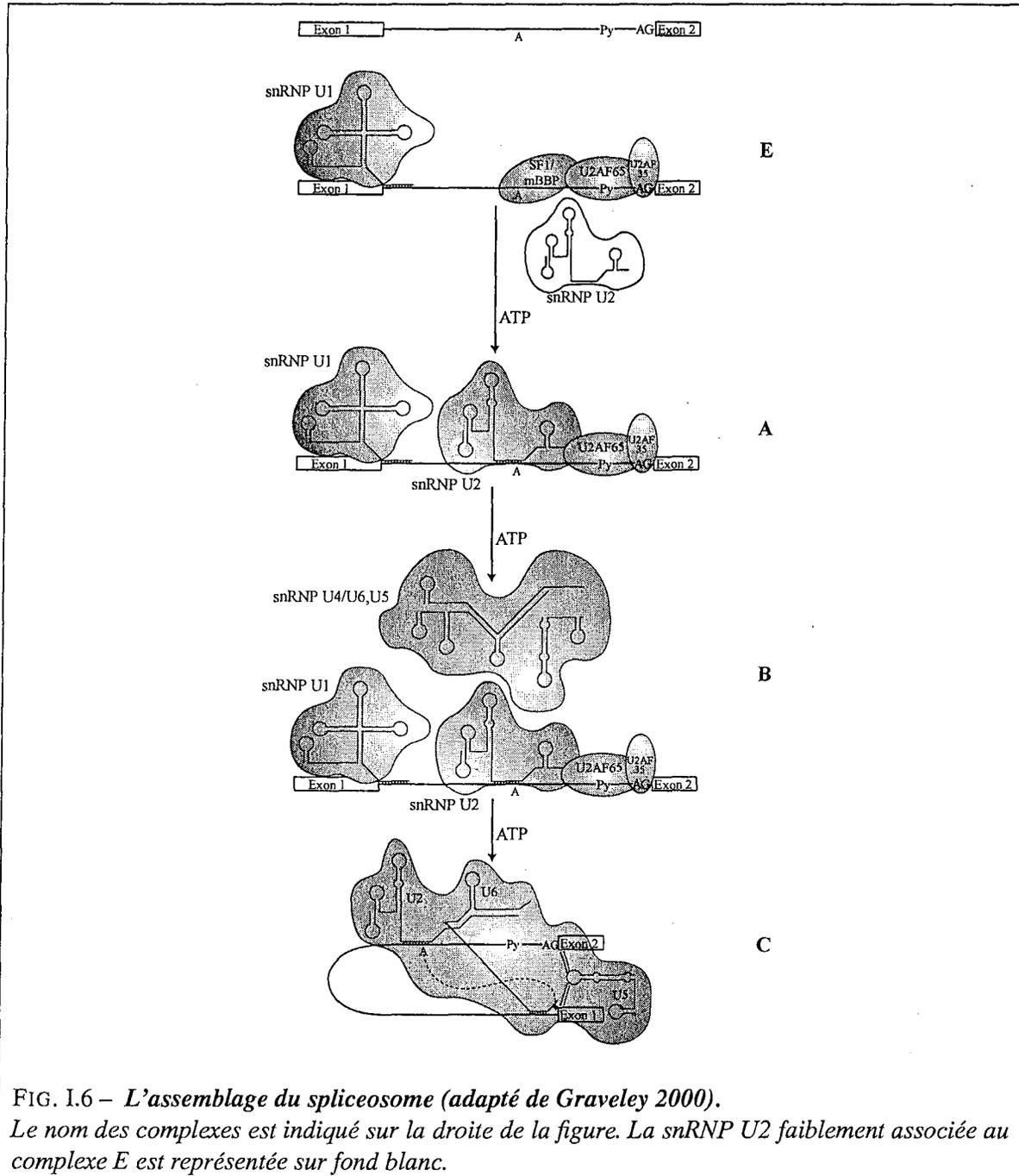


FIG. I.6 – L'assemblage du spliceosome (adapté de Graveley 2000).

Le nom des complexes est indiqué sur la droite de la figure. La snRNP U2 faiblement associée au complexe E est représentée sur fond blanc.

Le complexe H⁶ est obtenu pour tout transcrit de l'ARN polymérase II, même sans intron. Il contient de nombreuses protéines hnRNP⁷ s'assemblant sur les ARN nouvellement synthétisés.

La formation du 1^{er} complexe spécifique de l'épissage, le complexe E⁸, ne nécessite pas d'hydrolyse de l'ATP. Ce complexe est formé par la fixation de la snRNP U1 sur le site 5' d'épissage (Fig. I.6) et la fixation de facteurs d'épissage au niveau du site 3' (SF1/mBBP sur la boîte de branchement, U2AF⁶⁵

⁶ « heterogenous »

⁷ « heterogenous nuclear ribonucleoprotein »

⁸ « early », encore appelé « commitment complex »

sur la séquence polypyrimidine et U2AF³⁵ sur la jonction intron-exon). La particule snRNP U2 est également présente, bien que faiblement associée au complexe E. C'est au sein de ce complexe qu'a lieu la définition des exons.

Le complexe A est ensuite formé grâce à l'hydrolyse d'ATP, ce qui permet l'appariement du snRNA U2 avec la séquence de branchement. Le facteur SF1 lié à cette séquence est relargué. L'arrivée des snRNP U4, U5 et U6, associées sous forme de tri-snRNP [U4/U6, U5], marque la formation du complexe B, au sein duquel de nombreux réarrangements structuraux et protéiques se produisent, conduisant à la formation du complexe C actif pour la 1^{re} étape de la réaction d'épissage (Makarov *et al.* 2002 ; pour revue, Schwer 2001). La dissociation du complexe actif semble aussi passer par différentes étapes intermédiaires (Makarov *et al.* 2002).

De nombreuses protéines spliceosomales n'appartenant pas aux particules UsnRNP assistent l'assemblage des UsnRNP sur l'ARN prémessager. Ces protéines sont soit associées de manière stable au spliceosome, soit de façon transitoire.

Nous allons à présent décrire l'assemblage des UsnRNP conduisant à la formation des différents complexes spliceosomiaux E, A, B et C, tout en présentant les protéines participant à cet assemblage, dans l'ordre de leur intervention. Nous présenterons ainsi la vision admise depuis plusieurs années de l'assemblage des complexes spliceosomiaux, mais des données récentes suggèrent que les processus sont beaucoup plus complexes qu'on ne le pensait auparavant. Nous en discuterons à la fin du chapitre.

I.3.2.1 Reconnaissance du site 5' d'épissage

Protéines SR

Ce sont les 1^{res} protéines intervenant lors de l'assemblage du spliceosome chez les vertébrés. Elles sont ainsi nommées du fait de la richesse de leur extrémité C-terminale en enchaînements Sérine-Arginine (pour revue, Manley & Tacke 1996 ; Graveley 2000). Elles constituent une famille de protéines présentant des caractéristiques communes (Zahler *et al.* 1992) et jouant de multiples rôles dans l'épissage.

Définition des protéines SR : Tout membre de la famille des protéines SR répond aux critères suivants (Zahler *et al.* 1993a) :

- chaque protéine SR est capable d'induire l'épissage quand elle est ajoutée à un extrait cytoplasmique S100 déficient pour cette fonction,
- les protéines SR présentent toutes la même organisation structurale : un domaine N-terminal contenant un ou plusieurs motifs RRM⁹ et un domaine C-terminal riche en enchaînements de dipeptides arginine/sérine ou sérine/arginine appelé domaine RS,
- leurs domaines RS contiennent des sérines phosphorylées reconnues par l'anticorps mAb104 (Roth *et al.* 1991),
- elles sont très conservées en taille et en séquence chez tous les organismes dans lesquels elles existent.

⁹« RNA recognition motif »

Dix protéines répondant à ces critères ont pour l'instant été identifiées chez l'homme (Fig. I.7) : SRp20, ASF/SF2, SC35, SRp30c, SRp40, SRp46, SRp55, SRp75, 9G8 et une protéine plus éloignée en séquence, SRp54 (Zahler *et al.* 1992 ; Ge & Manley 1990 ; Krainer *et al.* 1990b ; Ge *et al.* 1991 ; Krainer *et al.* 1991 ; Fu & Maniatis 1992 ; Screaton *et al.* 1995 ; Zahler *et al.* 1993b ; Cavaloc *et al.* 1994 ; Chaudhary *et al.* 1991 ; Zhang & Wu 1996 ; Soret *et al.* 1998).

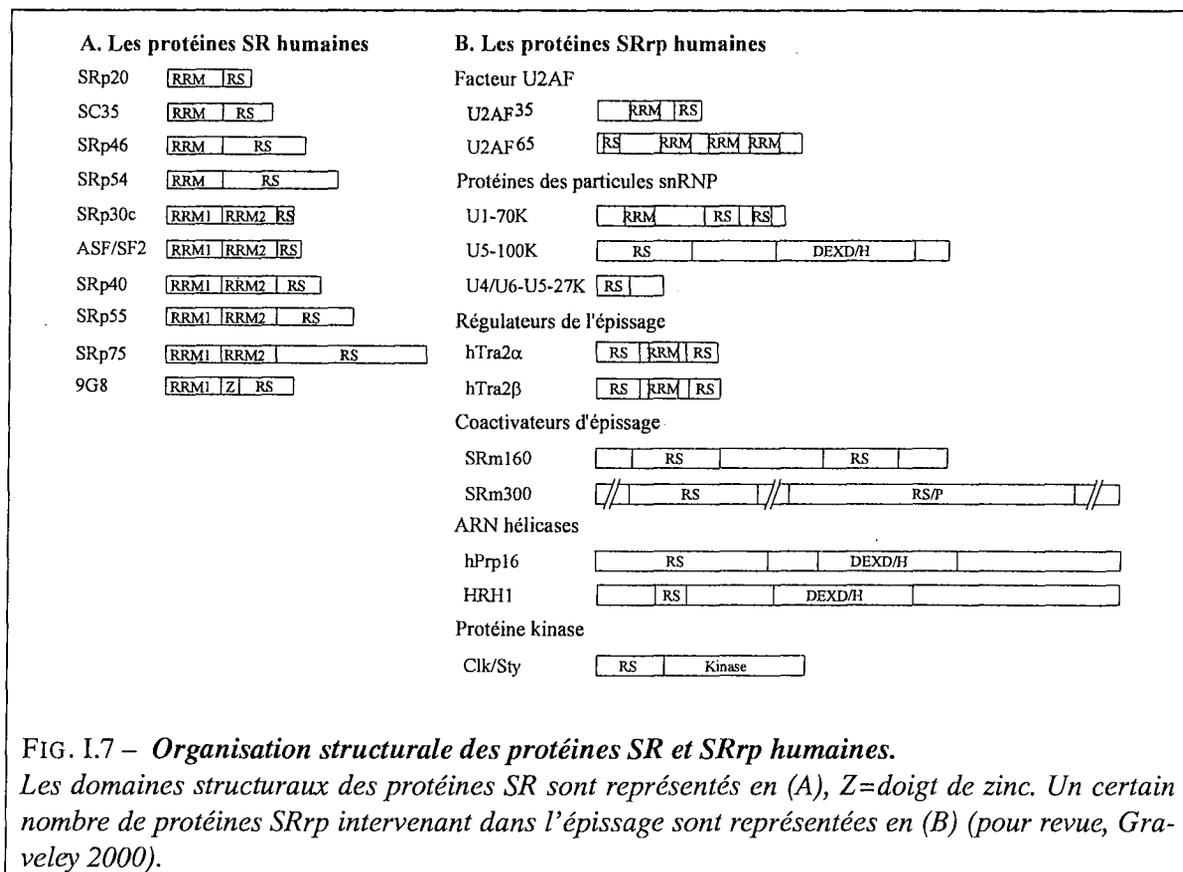


FIG. I.7 – Organisation structurale des protéines SR et SRrp humaines.

Les domaines structuraux des protéines SR sont représentés en (A), Z=doigt de zinc. Un certain nombre de protéines SRrp intervenant dans l'épissage sont représentées en (B) (pour revue, Graveley 2000).

Des protéines SR ont été identifiées chez les métazoaires et les plantes, mais ce n'est pas le cas pour tous les eucaryotes. Par exemple, la levure *S. cerevisiae* n'exprime pas de protéine SR et le nombre de protéines SR est très limité chez la drosophile.

Par rapport aux mammifères et aux plantes, les insectes et la levure présentent la particularité d'avoir des introns très courts. La mise en jeu de nombreuses protéines SR semble donc avoir accompagné l'augmentation en taille des introns.

Par ailleurs, chez la drosophile, il a pu être montré que la protéine B52, homologue de la protéine SRp55 humaine, est essentielle (Ring & Lis 1994), alors que chez *Caenorhabditis elegans*, les gènes de chacune des protéines SR ne se sont pas avérés essentiels. Seule la délétion des 6 gènes codant les protéines SR chez cet organisme conduit à un phénotype létal (pour revue, Graveley 2000).

Organisation structurale : Les protéines SR peuvent être regroupées en deux familles, de type ASF/SF2 ou SC35, suivant que leur domaine de fixation à l'ARN contient un ou deux domaines RRM (Figure I.7).

Les domaines RRM sont rencontrés dans un grand nombre de protéines se liant à l'ARN, comme par

exemple, la protéine hnRNP A1. Nous décrirons plus en détails les caractéristiques structurales des domaines RRM dans la partie consacrée à la protéine hnRNP A1 (cf chapitre II de l'introduction).

L'unique domaine RRM de SC35 et le domaine RRM1 d'ASF/SF2 portent les deux séquences signatures RNP1 (séquence consensus : RGFGFVxF) et RNP2 (séquence consensus : LFVGNL), alors que le domaine RRM2 d'ASF/SF2, encore appelé Ψ -RRM, ne présente pas les 2 séquences consensus. Ce domaine dicte la spécificité de substrat d'ASF/SF2 (Fig. I.8 ; Mayeda *et al.* 1999).

En ce qui concerne la protéine 9G8, son domaine de fixation à l'ARN est constitué d'un domaine RRM et d'un motif à doigt de zinc de type CCHH. Les 4 résidus de ce motif sont localisés sur un brin β et une hélice α arrangés autour de l'ion métallique (pour revue, Mackay & Crossley 1998). Ce motif doigt de zinc contribue à la spécificité de reconnaissance de l'ARN par 9G8 (Fig. I.7 ; Cavaloc *et al.* 1994, 1999).

Le domaine C-terminal des protéines SR porte un nombre variable d'enchaînements de résidus RS ou SR (Fig. I.8). Il permet à ces protéines d'établir des interactions avec les domaines RS d'autres facteurs protéiques. Ces interactions sont de type ionique, impliquant les arginines chargées positivement et les sérines phosphorylées. Lorsqu'elles sont phosphorylées, les protéines SR peuvent donc favoriser l'assemblage des composants spliceosomaux (Roscigno & Garcia-Blanco 1995 ; Xiao & Manley 1997) et la phosphorylation des domaines RS va jouer un rôle important en modulant la force de ce type d'interactions.

Différentes protéines assurent la phosphorylation des domaines RS : les protéines kinases SRPK1, SRPK2 et SRPK3¹⁰ (Gui *et al.* 1994 ; Wang *et al.* 1998 ; Duncan *et al.* 1998), la protéine Clk/Sty (Ben-David *et al.* 1991 ; Howell *et al.* 1991 ; Colwill *et al.* 1996 ; Prasad *et al.* 1999), et de façon surprenante, l'ADN topoisomérase I (Rossi *et al.* 1996 ; pour revue, Tazi *et al.* 1997).

Des phosphatases spécifiques des protéines SR ont également été identifiées : les protéines PP1 (Misteli & Spector 1996) et PP2A (Mermoud *et al.* 1992, 1994). Les protéines SR vont donc subir des cycles de phosphorylation-déphosphorylation de leur domaine RS lors de l'épissage des ARN pré-messagers (Xiao & Manley 1998).

Les ARN pré-messagers des protéines SRp20, 9G8, SC35, ASF/SF2, SRp40 et SRp55 sont épissés de façon alternative, mais les isoformes correspondants ne sont pas tous exprimés (Jumaa *et al.* 1997 ; Popielarz *et al.* 1995 ; Sureau *et al.* 2001 ; Ge *et al.* 1991 ; Du & Taub 1997 ; Sreaton *et al.* 1995). C'est en particulier le cas de l'ARN pré-messager de la protéine SC35, dont l'épissage alternatif conduit à la production d'ARN messagers présentant des stabilités différentes suivant les tissus (Sureau *et al.* 2001).

Les différences de propriétés au niveau de l'action sur l'épissage des différents variants des protéines SR ont été très peu étudiées, à l'exception d'une étude sur l'isoforme ASF3 de la protéine ASF/SF2 (Jumaa & Nielsen 2000). Les différences de structure primaire sont parfois très importantes. On peut ainsi noter des mutations ou des délétions importantes dans des domaines fonctionnels (Fig. I.8). Par exemple, deux variants de la protéine 9G8 portent une délétion presque totale de leur domaine RS, ce qui doit empêcher la protéine d'établir des interactions protéine-protéine avec d'autres protéines portant également des domaines RS. Il en va de même pour le domaine RS de la protéine ASF/SF2

¹⁰« SR-protein specific kinase »

qui, suivant le type de variant, présente soit une délétion, soit une mutation, ce qui doit moduler sa capacité d'interaction.

En plus d'une mutation dans le domaine RS, un variant de la protéine SRp40 porte une délétion du domaine RRM2. Si ce domaine est, comme pour ASF/SF2, responsable de la spécificité de fixation à l'ARN de la protéine (Mayeda *et al.* 1999), on peut s'attendre à des propriétés fonctionnelles très différentes.

Des protéines autres que les protéines SR présentent des domaines RS et parfois des domaines RRM. Elles ont été classées dans la famille SRp¹¹ (Fig. I.7). Certaines de ces protéines sont des facteurs d'épissage comme U2AF³⁵ et U2AF⁶⁵, d'autres sont des protéines des particules UsnRNP (U1-70K, U5-100K et [U4/U6,U5]-27K) (pour revue, Graveley 2000), des protéines régulatrices d'épissage (hTra2 α et hTra2 β), des protéines co-activatrices (SRm160 et SRm300). On trouve également dans cette famille des ARN hélicases (hPrp16, HRH1) et une protéine kinase (Clk/Sty).

Localisation des protéines SR : Certaines protéines SR comme SC35 et SRp40 sont exclusivement localisées dans le noyau, d'autres comme ASF/SF2, 9G8 et SRp20 sont capables de faire la navette entre le noyau et le cytoplasme. Le domaine RS de ces protéines joue un rôle important dans ce trafic entre le cytoplasme et le noyau (Cáceres *et al.* 1998 ; Yeakley *et al.* 1999 ; Cazalla *et al.* 2002). Ainsi, 10 dipeptides RS suffisent pour qu'ASF/SF2 transite entre le noyau et le cytoplasme et pour la localisation de la protéine dans des régions du noyau appelées « speckles » (Cazalla *et al.* 2002). La phosphorylation du domaine RS et en particulier, l'alternance de résidus chargés positivement (arginines) et négativement (sérines phosphorylées) de ce domaine, est importante pour la localisation nucléaire. En effet, le remplacement des séquences RS par des répétitions de di-peptides RE ou RD n'empêche pas la localisation nucléaire.

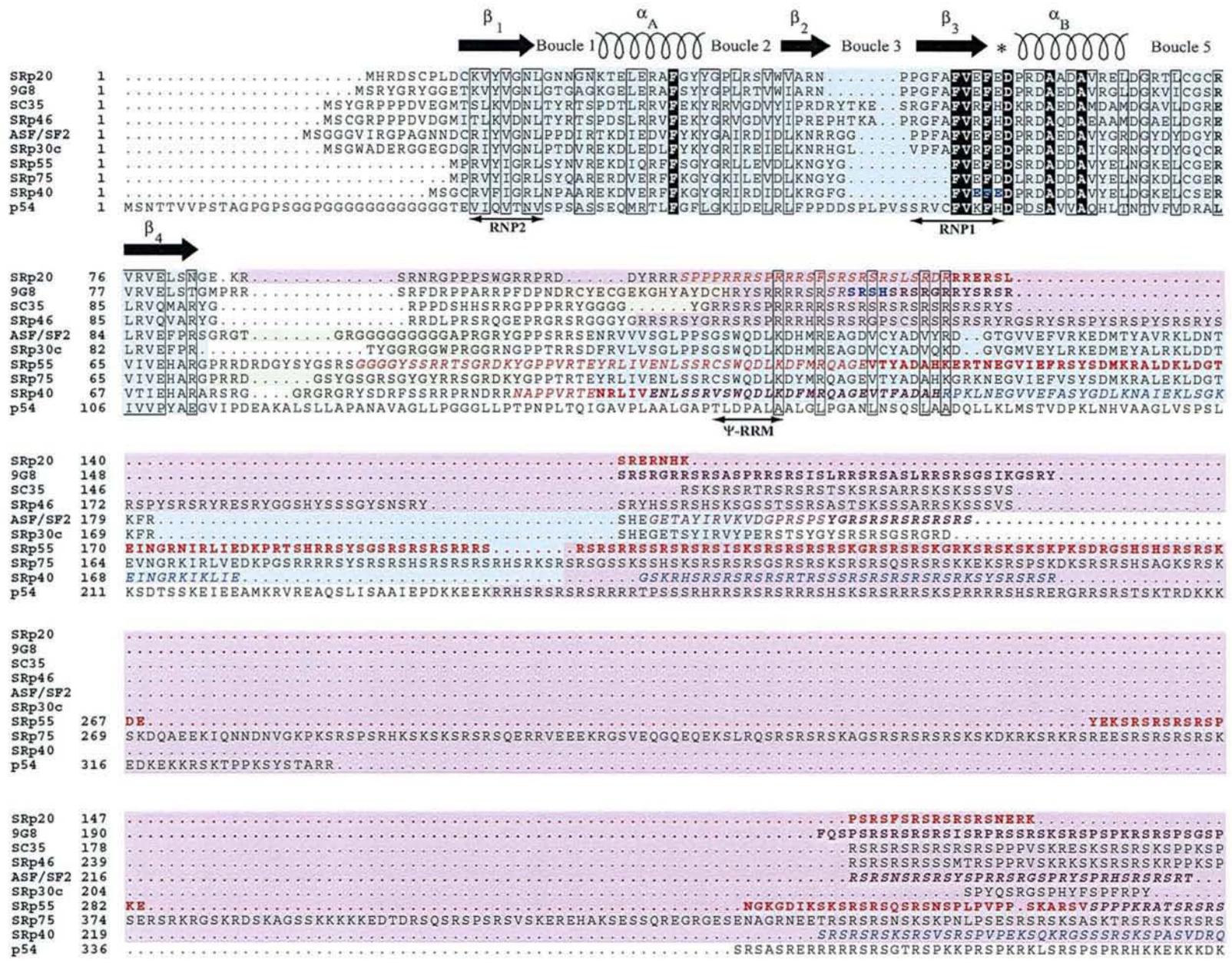
Le signal d'export nucléaire de ces protéines est bipartite. Il est constitué à la fois du domaine RS et du domaine RRM (Cáceres *et al.* 1998 ; Cazalla *et al.* 2002). Bien que la protéine SC35 porte également ce signal, elle reste strictement nucléaire. Cette protéine contient dans son domaine RS un signal de localisation nucléaire, constitué de résidus basiques et de sérines, qui assure la rétention de SC35 dans le noyau (Cazalla *et al.* 2002).

Les protéines impliquées dans l'import nucléaire des protéines SR sont 2 membres de la famille des importines β , les transportines SR1 et SR2 (TRN-SR1 et TRN-SR2) (Kataoka *et al.* 1999 ; Lai *et al.* 2000, 2001). TRN-SR2 permet le transport des protéines SR phosphorylées vers le noyau, en interagissant avec leur domaine RS (Lai *et al.* 2000, 2001).

Rôle des protéines SR dans l'épissage des ARN pré-messagers nucléaires : Les protéines SR interviennent à différentes étapes de l'assemblage et du mécanisme d'action des complexes spliceosomiaux, et elles jouent un rôle majeur (pour revue, Graveley 2000).

Dans l'épissage constitutif, elles contribuent à la définition des exons et elles favorisent l'assemblage du spliceosome, ceci en participant à l'établissement de multiples interactions ARN-protéine et protéine-protéine. Elles constituent une sorte de « ciment » des complexes spliceosomiaux. Pour ce

¹¹ « SR-related protein »



```

SRp20      . . . . .
9G8       229 RRSASPERMD . . . . .
SC35      214 EEEGAVSS . . . . .
SRp46     275 EEEGQMSS . . . . .
ASF/SF2   . . . . .
SRp30c    . . . . .
SRp55     329 RSRSKSRSRSRSSSRD . . . . .
SRp75     479 ASRSPSRSRSRSHRS . . . . .
SRp40     258 RSRSRSRSRVDSGN . . . . .
p54       383 DKERSRDERERSTSKKKKSKDKEKDRERKSESDKDVQVTRDYDEEEQGYDSEKEKKEKPIETGSPKTKECVVEKGTGDSLRESKVNDDHHEEDMMSD

```



FIG. I.8: **Alignement par CLUSTALW des séquences peptidiques des 10 protéines SR humaines**

La position des motifs RNP1 et RNP2, ainsi que celle du motif Ψ -RRM est indiquée sous l'alignement. La structure secondaire du domaine RRM1 est représentée de façon schématique au-dessus de l'alignement : les flèches symbolisent les brins β , les spirales, les hélices α , et l'astérisque, la boucle 4. Le code couleur utilisé pour représenter les domaines structuraux et les acides aminés mutés ou délétés dans les isoformes des protéines est indiqué en-dessous de l'alignement. Les résidus présents dans 100% des cas sont indiqués par un fond noir, les résidus présents dans au moins 80% des cas sont encadrés. Les séquences peptidiques correspondent aux numéros d'accèsion GenBank : P23152 (SRp20), Q16629 (9G8), Q01130 (SC35), NP_115285 (SRp40), Q07955 (ASF/SF2), Q13242 (SRp30c), Q13247 (SRp55), Q08170 (SRp75), Q13243 (SRp40) et Q05519 (p54).

type de fonction générale, les protéines SR n'ont pas de spécificité d'action et c'est pourquoi au départ, elles ont été décrites comme possédant des fonctions redondantes (pour revue, Manley & Tacke 1996).

En effet, comme indiqué précédemment, leur activité a d'abord été testée dans des extraits cytoplasmiques S100. Ce type d'extrait est préparé à partir de cellules HeLa et contient certains des facteurs protéiques et ribonucléoprotéiques nécessaires à l'épissage, mais pas les protéines SR, car elles sont majoritairement localisées dans le noyau et sans doute éliminées lors des étapes de préparation de l'extrait S100. L'ajout de l'une des protéines SR à l'extrait S100 permettant d'obtenir une activité d'épissage *in vitro* vis-à-vis de différents ARN pré-messagers, il a donc été proposé que les protéines n'aient pas de spécificité de substrat (pour revue, Manley & Tacke 1996).

En plus de ce rôle dans le mécanisme de base de l'épissage des ARN pré-messagers nucléaires, les protéines SR jouent un rôle très important dans l'épissage alternatif. Elles influencent le choix de sites d'épissage en compétition les uns avec les autres. Le test couramment utilisé pour déterminer la capacité d'une protéine SR à influencer la sélection des sites d'épissage, consiste à l'ajouter à de l'extrait nucléaire contenant déjà des protéines SR endogènes. C'est par cette technique qu'il est apparu que si les protéines SR ont des fonctions *a priori* redondantes dans l'épissage constitutif, elles possèdent par ailleurs des capacités d'activation spécifiques de l'utilisation de certains sites d'épissage (Manley & Tacke 1996 ; Graveley 2000 ; Smith & Valcárcel 2000). Le ratio de protéines SR variant suivant les tissus (Zahler *et al.* 1993a), le profil d'épissage des ARN pré-messagers soumis à l'épissage alternatif peut donc varier aussi de façon tissu-spécifique.

Cet effet des protéines SR sur la sélection des sites d'épissage est médié par des séquences portées par l'ARN pré-messager au niveau de ces sites d'épissage. Les protéines SR se fixent de manière spécifique sur ces séquences, comme nous le verrons dans la 2^e partie de cette introduction, où nous aborderons l'étude des mécanismes de régulation de l'épissage alternatif par les protéines SR.

Fixation de la snRNP U1

La fixation de la snRNP U1 sur le site 5' d'épissage initie la formation du complexe spliceosomal E. Suivant le degré de complémentarité de bases entre le snRNA U1 et le site 5', les protéines SR, en particulier ASF/SF2 et SC35, vont être plus ou moins importantes pour stabiliser la fixation de la particule au niveau du site d'épissage (Zahler & Roth 1995 ; Staknis & Reed 1994a ; Kohtz *et al.* 1994 ; Wu & Maniatis 1993 ; Jamison *et al.* 1995). Ainsi, en se fixant au voisinage du site 5' d'épissage par son domaine RRM, ASF/SF2 peut, par interaction avec le domaine RS de la protéine spliceosomale U1-70K, stabiliser la fixation de la snRNP U1 (Zuo & Manley 1994 ; Cao *et al.* 1997).

L'importance du rôle des protéines SR dans le processus de reconnaissance du site 5' est reflétée par le fait qu'*in vitro*, un excès de protéine ASF/SF2 ou SC35 permet l'épissage dans un extrait où la snRNP U1 a été éliminée (Crispino *et al.* 1994). De même, lorsque l'interaction entre le site 5' et le snRNA U1 est empêchée par l'ajout d'un oligonucléotide sequestrant l'extrémité 5' du snRNA U1, des complexes spliceosomaux normaux sont détectés sur un gel natif en présence d'un excès de protéines SR (Tarn & Steitz 1994).

I.3.2.2 La reconnaissance du site 3'

Au sein du complexe E, 2 facteurs participent à la reconnaissance du site 3' d'épissage : les facteurs U2AF¹² et SF1/mBBP¹³.

Facteur U2AF

U2AF est un facteur dimérique favorisant la fixation de la protéine SF1/mBBP et de la snRNP U2 au site 3' d'épissage (Fig. I.6 ; Zamore & Green 1989, 1991 ; Zhang *et al.* 1992). Il a d'ailleurs été récemment retrouvé associé à la snRNP U2, bien qu'il soit capable de se fixer à l'ARN pré-messager indépendamment de la particule U2 (Will *et al.* 2002).

La sous-unité U2AF⁶⁵ se lie à l'ARN pré-messager par l'intermédiaire de ses 2 domaines RRM. Elle contient également un domaine RS qui ne semble pas servir à établir des interactions avec d'autres protéines : ce domaine entre en contact avec la séquence de branchement de l'ARN pré-messager, lorsque U2AF⁶⁵ se fixe à la séquence polypyrimidine (Valcárcel *et al.* 1996).

U2AF⁶⁵ se lie préférentiellement aux séquences riches en pyrimidines avec une affinité proportionnelle à la longueur et à la richesse en uridines de la séquence (Zamore *et al.* 1992 ; Wu *et al.* 1999). En fonction des caractéristiques de la séquence polypyrimidine, le facteur U2AF³⁵ sera impliqué ou non dans les étapes précédant la fixation de la snRNP U2. En effet, lorsque la séquence polypyrimidine est sous-optimale, c'est-à-dire que l'enchaînement de pyrimidines est entrecoupé de purines, U2AF³⁵ se fixe à la bordure intron-exon et stabilise la fixation de U2AF⁶⁵ (Fig. I.9 ; Zamore & Green 1991 ; Guth *et al.* 1999 ; Zorio & Blumenthal 1999 ; Wu *et al.* 1999).

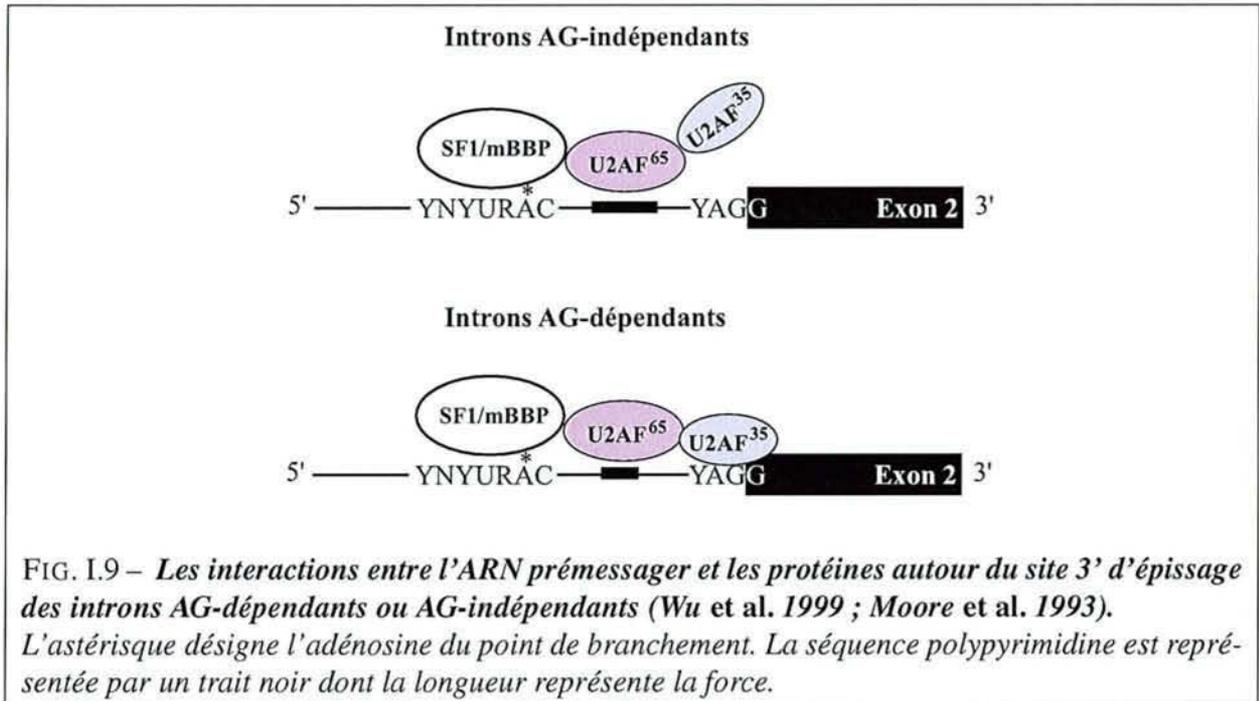
U2AF³⁵ reconnaît des séquences de type YAG (Wu *et al.* 1999). Il peut se fixer au niveau du dinucléotide AG du site 3' d'épissage grâce à son domaine RRM non canonique (Merendino *et al.* 1999 ; Wu *et al.* 1999 ; Zorio & Blumenthal 1999). Ce domaine est également impliqué dans l'interaction d'U2AF³⁵ avec U2AF⁶⁵ (Kielkopf *et al.* 2001).

On distingue ainsi 2 types d'exons, dits AG-dépendants ou AG-indépendants, suivant que la séquence polypyrimidine est sous-optimale ou non (Reed 1989). Chez *S. cerevisiae*, où tous les introns peuvent être considérés comme AG-indépendants, il n'y a pas d'homologue du facteur U2AF³⁵ (Abovich & Rosbash 1997). Par contre, chez *Schyzosaccharomyces pombe*, tous les introns sont AG-dépendants et il existe un orthologue du facteur U2AF³⁵ (Wentz-Hunter & Potashkin 1996 ; Romfo & Wise 1997). Enfin, le facteur U2AF³⁵ possède un domaine RS qui lui permet d'établir des interactions protéine-protéine, avec les protéines SR notamment.

Comme certaines protéines SR, le facteur U2AF fait la navette entre le noyau et le cytoplasme. Il entre et quitte le noyau sous forme de dimère et les domaines RS des deux sous-unités constituent des signaux de localisation redondants : un seul domaine RS suffit en effet pour l'import de U2AF dans le noyau (Gama-Carvalho *et al.* 2001).

¹²« U2 auxiliary factor »

¹³« splicing factor 1/mammalian branch point binding protein »



Récemment, un facteur U2AF²⁶ a été mis en évidence chez les mammifères (Shepard *et al.* 2002). Il possède la même extrémité N-terminale que U2AF³⁵, ce qui lui permet de se fixer à l'ARN au niveau du dinucléotide AG, et également, d'interagir avec U2AF⁶⁵. Il est d'ailleurs capable de stabiliser la fixation de U2AF⁶⁵ sur une séquence polypyrimidine faible.

U2AF²⁶ se différencie par son extrémité C-terminale qui ne possède pas de domaine RS. Or, comme nous le verrons dans le 2^e chapitre de cette introduction, le domaine RS du facteur U2AF³⁵ est primordial pour l'épissage alternatif de certains exons, car il est impliqué dans des interactions avec des protéines SR. La compétition entre les facteurs U2AF³⁵ et U2AF²⁶ pour la fixation au dinucléotide AG peut donc affecter le profil d'épissage de tels exons, d'autant plus que les taux relatifs entre ces deux facteurs diffèrent selon les tissus (Shepard *et al.* 2002).

Facteur SF1/mBBP

SF1/mBBP est le 1^{er} facteur à reconnaître le point de branchement lors de la réaction d'épissage (Berglund *et al.* 1997). C'est une protéine de 75 kDa dont plusieurs isoformes sont produits par épissage alternatif (Kramer *et al.* 1998). Il est à noter que suivant l'isoforme de SF1/mBBP, la protéine FBP11 ou la protéine FBP21 peut interagir avec le domaine C-terminal de la protéine (Bedford *et al.* 1998 ; pour revue, Reed 2000).

SF1 comporte également 2 domaines de fixation à l'ARN dans sa région C-terminale :

- un domaine de type maxi-KH. A la différence des domaines KH, de 70 acides aminés de long et arrangés en tandem, les domaines maxi-KH contiennent environ 100 acides aminés et ne sont pas répétés. Le domaine maxi-KH de SF1/mBBP lui permet d'interagir spécifiquement avec le point de branchement (Berglund *et al.* 1998 ; Liu *et al.* 2001 ; Peled-Zehavi *et al.* 2001),
- un 2^e domaine de fixation à l'ARN, de type doigt de zinc, qui interagit de façon non spécifique avec le squelette phosphodiester de l'ARN. Cette interaction augmenterait l'affinité globale de la protéine pour l'ARN (Berglund *et al.* 1998),

Enfin, SF1 comporte dans sa région N-terminale un segment assurant l'interaction avec U2AF⁶⁵ (Rain *et al.* 1998).

La capacité d'interaction de SF1/mBBP avec la séquence de branchement dépend du degré de similitude de cette séquence avec la séquence consensus. En effet, la protéine se fixe à la séquence du point de branchement d'autant plus fortement que celle-ci se rapproche du consensus (Peled-Zehavi *et al.* 2001). La force de l'interaction de la protéine SF1 avec le point de branchement contribue ainsi à la sélection du site 3' d'épissage.

SF1/mBBP forme ensuite un complexe coopératif avec la protéine U2AF⁶⁵, ce qui permet de renforcer la fixation de chaque protéine sur sa séquence spécifique (Berglund *et al.* 1997, 1998 ; Rain *et al.* 1998 ; Liu *et al.* 2001 ; Peled-Zehavi *et al.* 2001).

La fixation de SF1/mBBP à l'ARN pré-messager est transitoire. En effet, la fixation de la snRNP U2 s'accompagne du départ de SF1 (Berglund *et al.* 1998). Cette protéine semble donc intervenir dans la cinétique d'assemblage des spliceosomes. En effet, l'élimination dans des extraits nucléaires de SF1/mBBP n'abolit pas l'épissage mais affecte la cinétique de formation des complexes spliceosomaux (Guth & Valcárcel 2000). Il a également été proposé que SF1/mBBP et U2AF⁶⁵ participent à l'élimination de facteurs nucléaires fixés de manière non spécifique sur la séquence de branchement avant l'assemblage du spliceosome (Chiara *et al.* 1996 ; Guth & Valcárcel 2000).

I.3.2.3 Rapprochement des sites 5' et 3' d'épissage

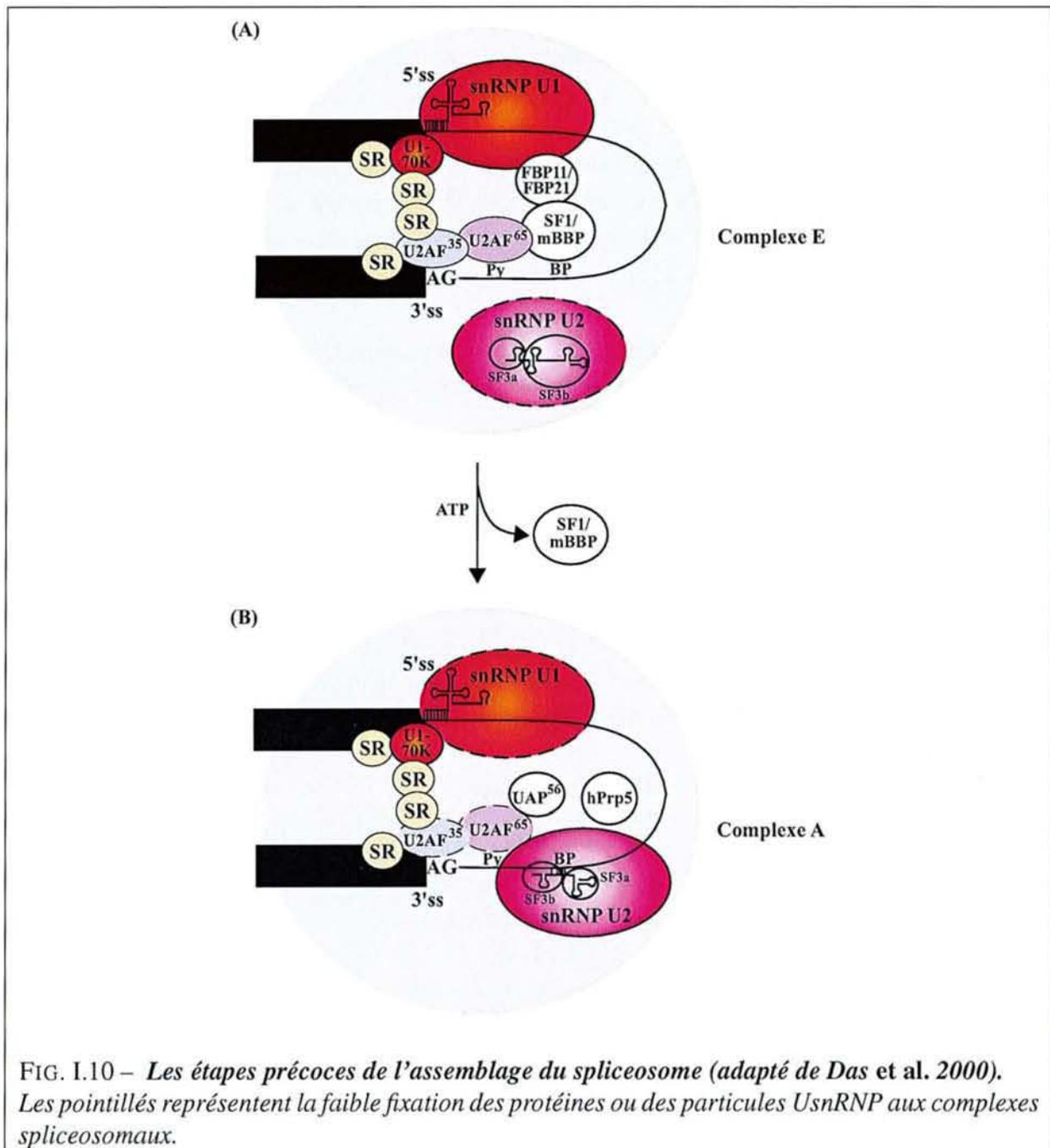
A ce stade précoce de l'assemblage du spliceosome, les sites 5' et 3' ont été reconnus par la machinerie d'épissage. Un important réseau d'interactions s'établit alors entre les protéines liées à l'ARN pré-messager et les protéines des UsnRNP, ce qui permet de rapprocher les extrémités des 2 exons à liguer et de définir les exons et l'intron (Fig. I.10-A).

Les protéines SR sont sans doute fortement impliquées dans ce processus, puisqu'elles pourraient interagir entre elles et ainsi relier le site 5' sur lequel est fixé la snRNP U1 et le site 3', auquel U2AF³⁵ est associé (Robberson *et al.* 1990 ; Fu & Maniatis 1992 ; Wu & Maniatis 1993). La protéine SRp54 participe aussi à ce réseau d'interactions en interagissant avec U2AF⁶⁵ et d'autres protéines SR (Zhang & Wu 1996).

Par ailleurs, SF1/mBBP contribue à établir une communication entre les sites 5' et 3' d'épissage en interagissant avec la protéine FBP11 ou FBP21, qui est capable d'interagir en même temps avec la snRNP U1 (Fig. I.10) (Bedford *et al.* 1998 ; pour revue, Reed 2000).

I.3.2.4 Reconnaissance de la séquence de branchement par la snRNP U2

La formation du complexe spliceosomal A, encore appelé pré-spliceosome, est caractérisée par la fixation stable, sur l'ARN pré-messager, de la snRNP U2 jusqu'alors faiblement associée au complexe E (Das *et al.* 2000). Cette fixation nécessite l'hydrolyse d'ATP et conduit au déplacement du facteur SF1/mBBP (Fig. I.10-B ; Berglund *et al.* 1998).



Interaction de la snRNP U2 avec l'ARN prémessager

Les sous-unités SF3a et SF3b du facteur SF3, présent dans la forme 17S de la snRNP U2, participent à l'ancrage de la particule sur l'ARN prémessager (Fig. I.11). Les protéines des deux sous-unités (à l'exception de SAP130) interagissent en effet de manière séquence-indépendante avec l'ARN, en amont du point de branchement (Bennett *et al.* 1992 ; Champion-Arnaud & Reed 1994 ; Staknis & Reed 1994b ; MacMillan *et al.* 1994 ; Gozani *et al.* 1996 ; Nesic & Kramer 2001). L'ensemble des interactions qui s'établissent au niveau de ce site d'ancrage conduirait à un repliement particulier de l'intron (Fig. I.11) (Gozani *et al.* 1996).

Ainsi, la protéine SAP155 de la sous-unité SF3b se lie à l'ARN de part et d'autre du point de branchement et interagit avec la même région de U2AF⁶⁵ qui, dans l'étape précédente, était liée à

SF1 (Champion-Arnaud & Reed 1994 ; Query *et al.* 1996 ; Gozani *et al.* 1996, 1998). La protéine p14/SF3b14a se fixe, quant à elle, sur le point de branchement (Fig. I.11) (Query *et al.* 1995) et favoriserait l'interaction du snRNA U2 avec l'ARN pré-messager (Will *et al.* 2001 ; pour revue, Will & Lührmann 2001). L'association de la protéine p14 avec le nucléotide de branchement est d'autant plus forte que ce nucléotide est une adénosine (Query *et al.* 1995).

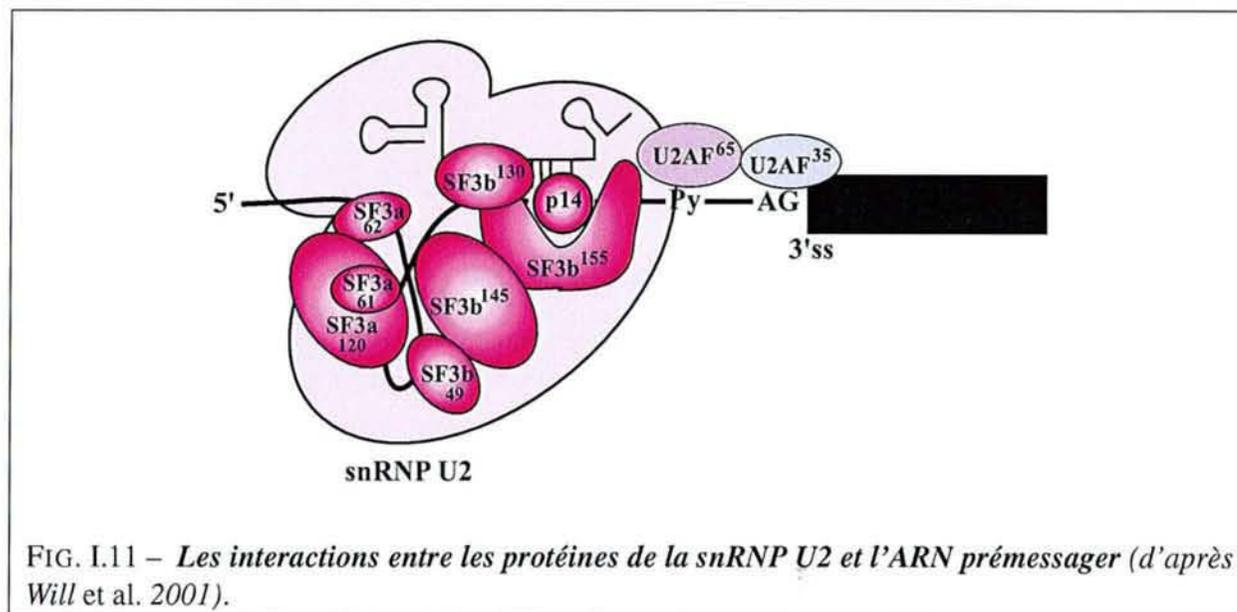


FIG. I.11 – Les interactions entre les protéines de la snRNP U2 et l'ARN pré-messager (d'après Will *et al.* 2001).

Stabilisation de la fixation de la snRNP U2

Lorsqu'il est fixé à la séquence polypyrimidine, le facteur U2AF⁶⁵ stabilise l'association de la snRNP U2 avec l'ARN pré-messager de façon ATP-dépendante (Zamore & Green 1989 ; Gozani *et al.* 1998). Cette dépendance vis-à-vis de l'ATP s'explique par l'implication d'une protéine ARN hélicase à boîte DEAD, UAP56¹⁴, qui interagit avec le facteur U2AF⁶⁵ (Fig. I.10) (Fleckner *et al.* 1997 ; pour revue, Schwer 2001).

La protéine UAP56 participerait aux changements conformationnels qui permettent l'association de la snRNP U2 avec le point de branchement. Elle pourrait favoriser le départ du facteur SF1 (pour revue, Staley & Guthrie 1998). Une autre ARN hélicase associée à la snRNP U2, la protéine hPrp5, a récemment été identifiée. Elle participerait à la réorganisation de la particule, permettant ainsi sa fixation à l'intron (Newnham & Query 2001 ; Zhou *et al.* 2002).

Les protéines SR jouent aussi un rôle dans l'association de la snRNP U2 avec le point de branchement (Fu & Maniatis 1992 ; Staknis & Reed 1994b ; Tarn & Steitz 1995 ; MacMillan *et al.* 1997). En effet, la protéine SC35 peut remplacer fonctionnellement le facteur U2AF⁶⁵ dans un extrait nucléaire où ce facteur a été éliminé (MacMillan *et al.* 1997). De plus, l'ajout de protéines SR à un extrait cytoplasmique S100 inactif en épissage favorise la formation du spliceosome (Krainer *et al.* 1990b) et en particulier, l'association de la particule snRNP U2 avec le point de branchement (Tarn & Steitz 1995).

¹⁴ « U2AF⁶⁵ associated protein, 56 kDa »

Conformation de la séquence de branchement

Lors de la fixation stable de la snRNP U2 sur l'ARN prémessager, l'appariement entre le snRNA U2 et la séquence de branchement permet de faire sortir, hors de l'hélice, l'adénosine dont le 2'OH fera la première attaque nucléophile de la réaction d'épissage (Fig. I.6) (Wu & Manley 1989 ; Wassarman & Steitz 1992 ; Berglund *et al.* 2001 ; Newby & Greenbaum 2001). Un résidu pseudouridine du snRNA U2, qui s'apparie avec le nucléotide voisin de cette adénosine (Massenet & Branlant 1999), stabilise l'interaction entre le snRNA U2 et la séquence de branchement (Fig. I.4-B) (Newby & Greenbaum 2001). Il permet à l'adénosine réalisant l'attaque nucléophile d'adopter une position favorable pour la réalisation de cette attaque.

I.3.2.5 Incorporation de la tri-snRNP [U4/U6, U5] dans le spliceosome

L'assemblage du spliceosome se poursuit avec l'incorporation de la tri-snRNP [U4/U6,U5], ce qui conduit à la formation du complexe spliceosomal B (Bindereif & Green 1987 ; Konarska & Sharp 1987). Ce processus est favorisé par les protéines SR qui participent à la reconnaissance du site 5' par la snRNP U6 (Rosigno & Garcia-Blanco 1995 ; Crispino & Sharp 1995 ; Tarn & Steitz 1995).

D'autres protéines récemment identifiées interviennent également dans le recrutement de la tri-snRNP. Ainsi la protéine SPF30, encore appelée SMNrp¹⁵, est une protéine associée de façon stable avec la snRNP U2, qui interagit également avec la tri-snRNP (Neubauer *et al.* 1998 ; Meister *et al.* 2001 ; Rappsilber *et al.* 2001). En l'absence de SPF30, l'assemblage du spliceosome s'arrête au complexe A (Rappsilber *et al.* 2001).

Les protéines SRrp 65 et 110 de la tri-snRNP [U4/U6,U5] aident aussi à l'entrée de la particule dans le spliceosome, probablement en interagissant avec les protéines SR (Makarova *et al.* 2001). D'autres protéines à domaine RS de la tri-snRNP participeraient également à la stabilisation de la particule avec l'ARN prémessager (Fetzer *et al.* 1997 ; pour revue, Graveley 2000).

Il a récemment été mis en évidence que, chez la levure *S. cerevisiae*, les snRNA U4 et U6 sont capables de s'associer *in vitro* avec le site 5' en absence de point de branchement et de site 3' (Johnson & Abelson 2001). Il est donc probable que la tri-snRNP soit faiblement associée aux complexes spliceosomaux à une étape plus précoce de l'assemblage du spliceosome.

I.3.3 Changements conformationnels conduisant aux complexes actifs

I.3.3.1 Transitions structurales

Avec l'arrivée de la tri-snRNP [U4/U6,U5], un grand nombre de réarrangements conduisent à la formation du complexe C actif en épissage, au cours desquels des interactions entre UsnRNA vont être rompues, au profit de la formation de nouveaux duplex d'ARN.

L'interaction du site 5' d'épissage avec le snRNA U1 est ainsi remplacée par une interaction avec le snRNA U6, suivie d'un appariement des snRNA U2 et U6 (Fig. I.12) (Sawa & Shimura 1992 ; Wassarman & Steitz 1992). Au cours de ces réarrangements conformationnels, le snRNA U2 reste associé au point de branchement.

¹⁵« SMN related protein »

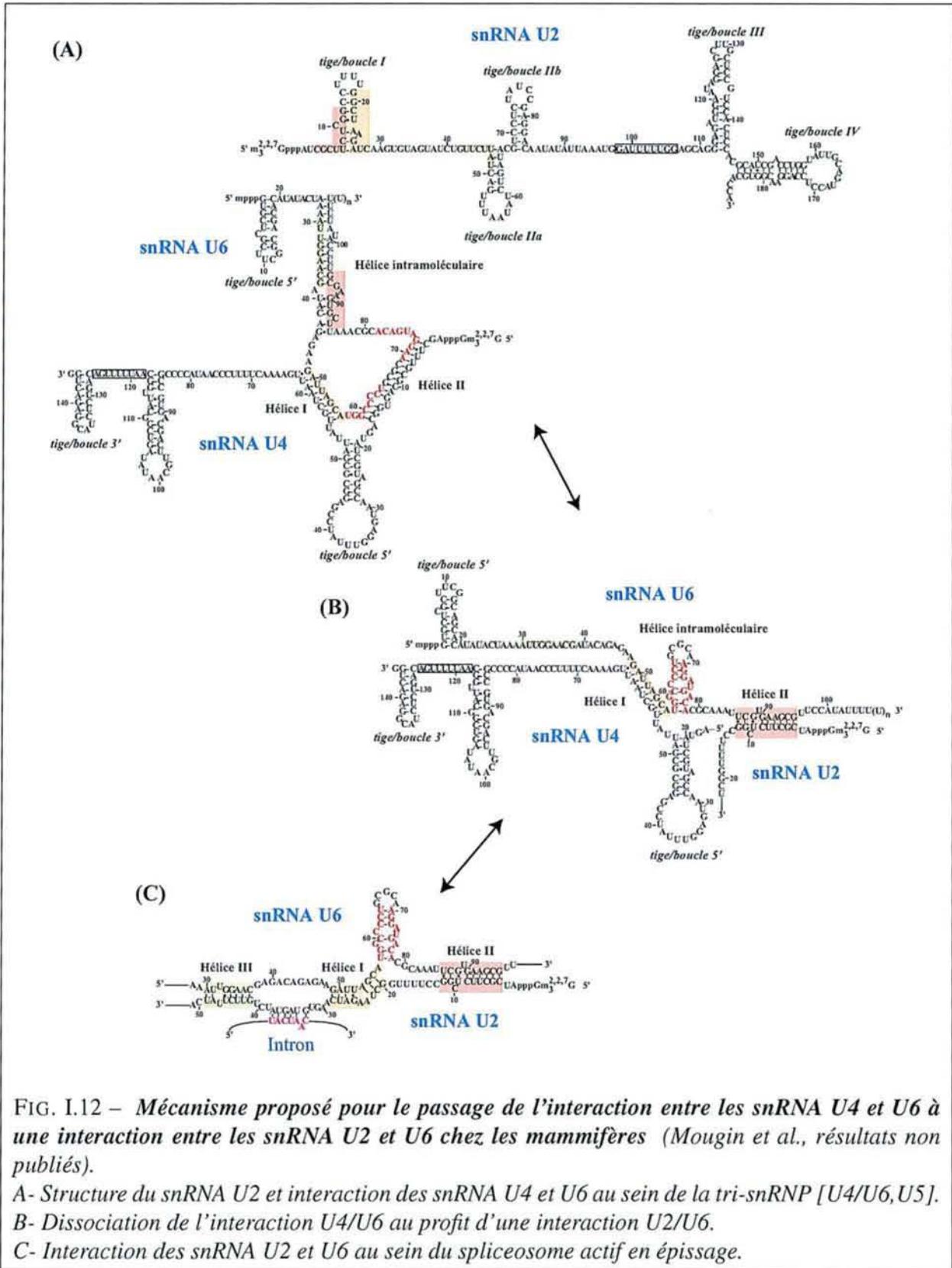


FIG. I.12 – Mécanisme proposé pour le passage de l'interaction entre les snRNA U4 et U6 à une interaction entre les snRNA U2 et U6 chez les mammifères (Mougin et al., résultats non publiés).

A- Structure du snRNA U2 et interaction des snRNA U4 et U6 au sein de la tri-snRNP [U4/U6, U5].

B- Dissociation de l'interaction U4/U6 au profit d'une interaction U2/U6.

C- Interaction des snRNA U2 et U6 au sein du spliceosome actif en épissage.

Certaines des protéines de la tri-snRNP [U4/U6,U5] participeraient à ces transitions structurales par leur activité enzymatique de type ARN hélicase ou proline isomérase (pour revues, Staley & Guthrie 1998 ; Schwer 2001).

Toutes les hélicases n'ont pas été identifiées chez l'homme, contrairement à la levure, où elles sont connues. L'ARN hélicase humaine p68 interviendrait ainsi dans la dissociation du duplex snRNA U1/site 5' d'épissage (Liu 2002). Les protéines U5-100 et U5-200, composantes de la snRNP U5 et de la tri-snRNP [U4/U6,U5], sont des ARN hélicases impliquées dans la disruption de l'appariement entre les snRNA U4 et U6 (Lauber *et al.* 1996 ; Lagerbauer *et al.* 1996, 1998). La protéine U5-116 à activité GTPase pourrait également participer aux transitions structurales (Fabrizio *et al.* 1997). L'hélicase MAH¹⁶ est également associée aux spliceosomes (Molnar *et al.* 1997), mais aucun rôle dans l'épissage ne lui a encore été attribué.

La snRNP U5 participe directement au déplacement de la snRNP U1 du site 5' d'épissage (Fig. I.13-A ; McConnell & Steitz 2001 ; pour revue, Staley & Guthrie 1998). La particule interagit avec la bordure 3' du 1^{er} exon par l'intermédiaire de la boucle I très conservée du snRNA U5. Cette interaction est compatible avec celle du snRNA U6 avec le site donneur. Elle sera maintenue au cours de la 1^{re} et de la 2^e étape de la réaction d'épissage (Wyatt *et al.* 1992 ; Sontheimer & Steitz 1993 ; McConnell & Steitz 2001). Les interactions mises en jeu sont aussi bien canoniques que non canoniques. Elles font intervenir la séquence riche en U — très conservée et fortement modifiée post-transcriptionnellement (pour revue, Massenet *et al.* 1998) — de la boucle I du snRNA U5 et la séquence dégénérée de la bordure exonique.

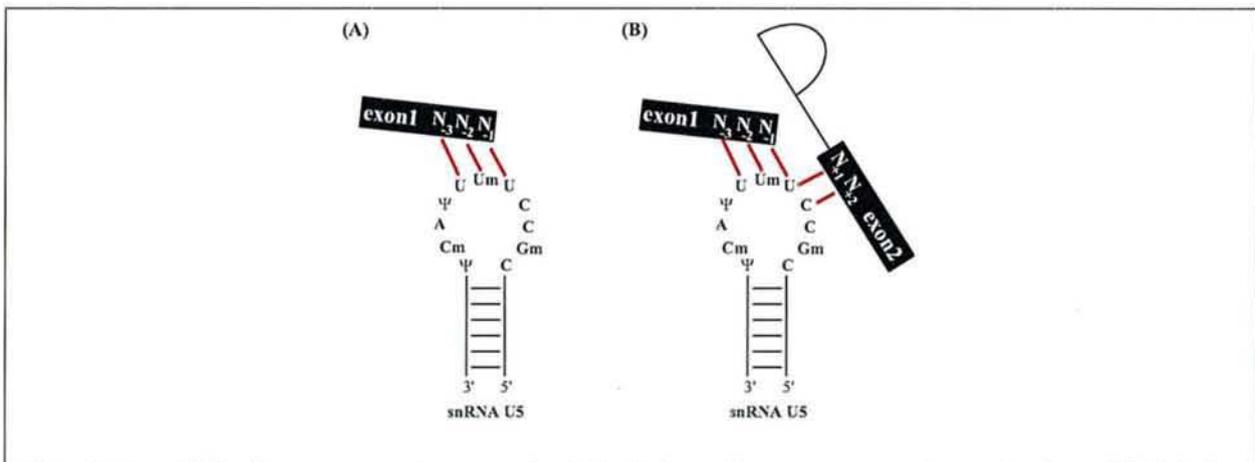


FIG. I.13 – *Détection par pontage covalent des interactions mises en jeu entre le snRNA U5 et les deux exons, au cours de la réaction d'épissage.*

Les panneaux A et B correspondent respectivement à des interactions détectées entre le snRNA U5 et les exons avant la première réaction d'épissage (A) et avant la deuxième réaction d'épissage (B). Les traits rouges entre la boucle I conservée du snRNA U5 et les exons indiquent les résidus pontés (pour revue, Madhani & Guthrie 1994).

Cet appariement entre le snRNA U5 et la bordure exonique est stabilisé par des protéines, en particulier, la protéine U5-220 (Anderson *et al.* 1989). Par irradiation aux UV, des pontages covalents ont

¹⁶ « matrix-associated helicase »

pu être formés en présence de résidus 4-thioU, entre cette protéine et le dinucléotide GU 5' terminal de l'intron, ceci dès la formation du complexe spliceosomal B (Wyatt *et al.* 1992 ; Reyes *et al.* 1996). Par des expériences *in vitro*, V. Ségault a pu montrer au laboratoire, en collaboration avec l'équipe de R. Lührmann, que chez les vertébrés, la protéine U5-220 peut compenser l'absence de la boucle I du snRNA U5, puisqu'un épissage efficace est obtenu en l'absence de cette boucle (Ségault *et al.* 1999). Dans un 2^e temps, la boucle I du snRNA U5, encore fixée au 1^{er} exon avec la protéine U5-220, interagit avec le 2^e exon (Fig. I.13-B). Cette étape a lieu à l'issue de la 1^{re} réaction d'épissage. Les 2 exons sont alors alignés pour la 2^e réaction d'épissage (Sontheimer & Steitz 1993 ; Chiara *et al.* 1996, 1997). Une interaction de type Hoogsteen s'établit également entre les deux résidus G des bordures exoniques (Deirdre *et al.* 1995). La protéine U5-220 stabiliserait également l'interaction du snRNA U5 avec les 2 exons (Teigelkamp *et al.* 1995 ; Beggs *et al.* 1995 ; pour revue, Collins & Guthrie 2000).

La protéine U5-220 pourrait participer à la reconnaissance du site 3' d'épissage. En effet, un pontage covalent aux UV de la protéine est détecté de part et d'autre du dinucléotide AG lors de la 2^e étape d'épissage (Collins & Guthrie 2000 ; Siatecka *et al.* 1999). Une autre protéine, hSlu7, est également impliquée dans la reconnaissance du site 3' d'épissage (Chua & Reed 1999). L'épissage d'un ARN pré-messager dans un extrait nucléaire ne contenant plus de protéine hSlu7 conduit à la sélection d'un mauvais dinucléotide AG lors de la 2^e étape d'épissage (Chua & Reed 1999 ; pour revue, Reed 2000). Des données chez la levure sont en accord avec l'implication de ces 2 facteurs dans la sélection du dinucléotide AG lors de la 2^e étape d'épissage puisque des mutations dans les protéines Slu 7 et Prp8, l'homologue chez la levure de la protéine U5-220 humaine, sont synthétiques létales (Frank *et al.* 1992).

I.3.3.2 Rôle des ions divalents

Les ions, en particulier, le magnésium et le zinc, sont indispensables pour la catalyse par le spliceosome.

Les ions Zn^{2+} sont nécessaires pour la 2^e étape d'épissage. En effet, l'ajout d'un agent chélatant du zinc bloque la réaction d'épissage au niveau de la 1^{re} étape. Mais cette inhibition est réversible, puisque l'épissage est restauré par dialyse en présence d'ions Zn^{2+} (Shomron *et al.* 2002).

Plusieurs protéines spliceosomales (comme U1C, U2AF³⁵, SF1, SF3a60, SF3a66, 9G8, ...) possèdent un motif doigt de zinc leur permettant d'établir des interactions protéine/acide nucléique et protéine/protéine (pour revue, Mackay & Crossley 1998). Or, certaines protéines indispensables à la 2^e étape d'épissage, comme la protéine hSlu7, portent également un motif doigt de zinc (pour revue, Reed 2000). Les ions Zn^{2+} pourraient donc agir sur l'épissage en se fixant à ces protéines, ou alors, ils pourraient jouer un rôle plus général, puisqu'ils sont capables de se fixer à l'ARN de façon non spécifique. Dans ce cas, ils contribueraient à neutraliser les charges négatives du squelette phosphodiester et favoriseraient ainsi des interactions protéines/acides nucléiques (Shomron *et al.* 2002).

Les ions Mg^{2+} sont directement impliqués dans la réaction catalytique (pour revues, Sontheimer 2001 ; Villa *et al.* 2002) en stabilisant les groupes partants des 2 étapes d'épissage (Sontheimer *et al.* 1999 ; Gordon *et al.* 2000 ; Yean *et al.* 2000) et ils activeraient probablement le groupement 2'OH du

point de branchement (Steitz & Steitz 1993).

Un nucléotide du snRNA U6 (U80) serait particulièrement impliqué dans ce processus puisqu'il coordine l'ion Mg^{2+} impliqué dans la stabilisation du groupe partant du site 5' (Yean *et al.* 2000). Ainsi, au sein du spliceosome, l'interaction tertiaire qui s'établit entre les snRNA U2 et U6 rapprocherait le nucléotide U80 du groupe partant lors de la 1^{re} étape d'épissage, ce qui permettrait la stabilisation de ce groupe par les ions Mg^{2+} (Yean *et al.* 2000).

I.3.4 Existence d'une structure active stable

L'assemblage ordonné des UsnRNP, que nous venons de décrire, est le modèle généralement admis. Il a été récemment remis en question par la découverte chez la levure d'une penta-snRNP (U1, U2, U4/U6, U5) fonctionnelle en épissage (Stevens *et al.* 2002 ; Nilsen 2002).

Plusieurs autres études vont dans le sens de l'existence dans les cellules eucaryotes de spliceosomes actifs stables : la présence de la snRNP U2 faiblement associée au complexe E (Das *et al.* 2000), la découverte récente de la présence des protéines ASF/SF2, U2AF³⁵ et U2AF⁶⁵ dans une particule snRNP U2 plus complexe que celle précédemment étudiée (Will *et al.* 2002). De même, une étude menée chez le nématode et dans les cellules HeLa a montré que la tri-snRNP [U4/U6, U5] s'associe fonctionnellement avec l'ARN prémessager dans le complexe E et ceci, indépendamment de la fixation des snRNP U1 et U2 (Maroney *et al.* 2000).

Ces nouvelles données ne sont pas nécessairement incompatibles avec la vision de l'épissage que nous avons jusqu'alors. En effet, ce complexe spliceosomal pré-formé pourrait constituer un moyen de concentrer les facteurs d'épissage à proximité de l'ARN prémessager, afin d'augmenter l'efficacité de la fixation sur l'ARN prémessager et par conséquent, la réaction d'épissage. Dans ce modèle, l'assemblage des snRNP autour de l'ARN prémessager doit plus être vu comme un ensemble d'étapes de stabilisation d'interactions, plutôt que des étapes de recrutement de composants.

I.3.5 Intégration de l'étape d'épissage dans le processus d'expression des gènes

L'expression des gènes eucaryotes débute par l'initiation de la transcription. Les données récentes montrent que les étapes de maturation des ARN ont lieu en même temps que la transcription. Les ARN sont très rapidement coiffés en 5' au cours de leur synthèse, le spliceosome est assemblé sur les ARN en cours de transcription, ainsi que la machinerie de polyadénylation (Fig. I.14) (pour revue, Dreyfuss *et al.* 2002).

I.3.5.1 Lien entre les étapes de transcription, épissage et polyadénylation

La coordination entre la transcription, l'épissage et la polyadénylation est médiée par le domaine CTD¹⁷ de la grande sous-unité de l'ARN polymérase II, ce qui rend tous ces processus interdépendants. Ainsi, le domaine CTD est capable d'activer aussi bien la formation de la coiffe, que l'épissage et la polyadénylation (pour revue, Maniatis & Reed 2002 ; Proudfoot *et al.* 2002). De plus, au niveau

¹⁷« carboxyterminal domain »

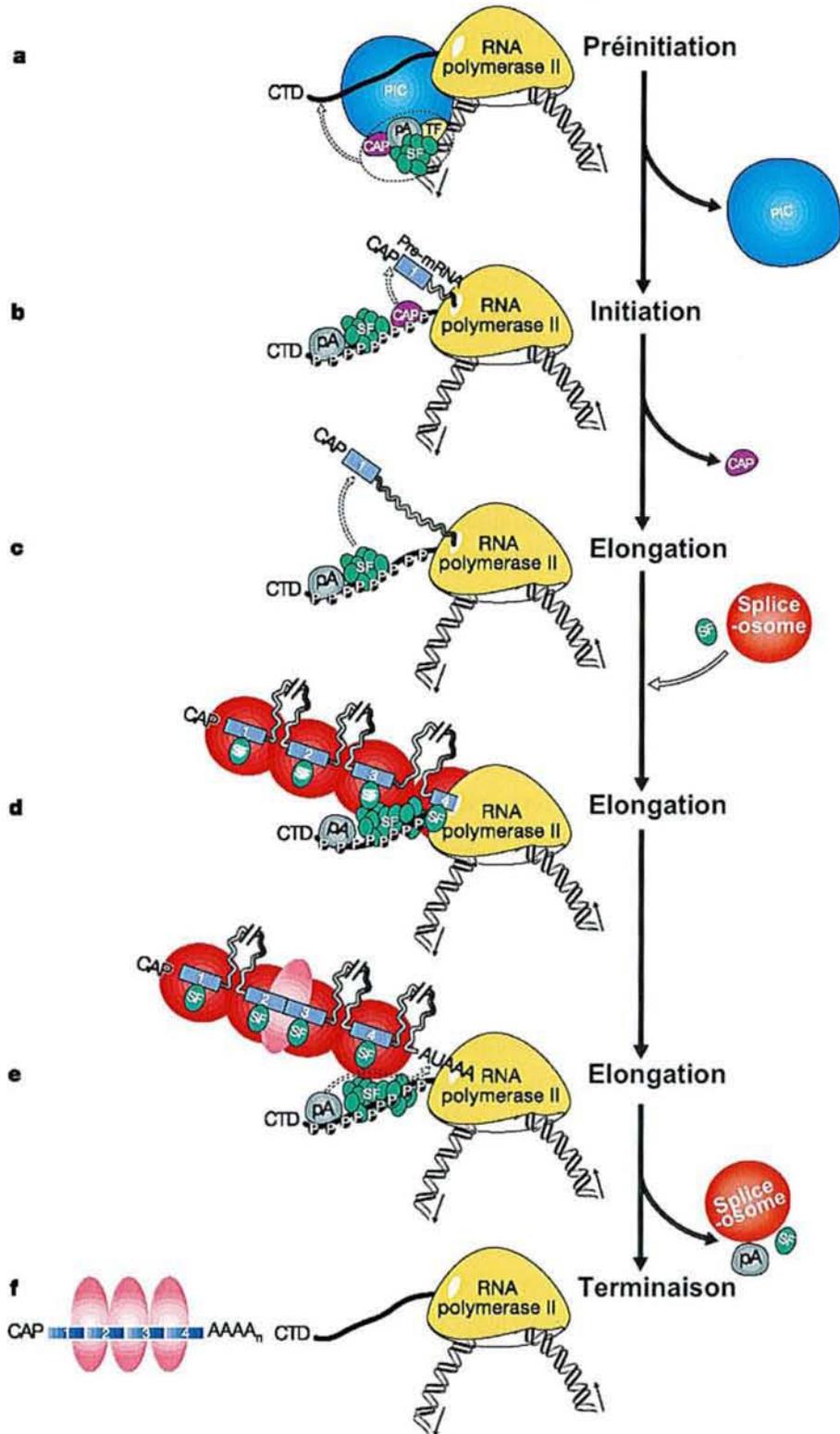


FIG. I.14 – La transcription et la maturation des ARN pré-messagers sont des processus liés (pour revue, Maniatis & Reed 2002). Les machineries de la transcription, la formation de la coiffe, de l'épissage et la polyadénylation sont représentées schématiquement. PIC, complexe de pré-initiation de la transcription ; TF, facteurs de transcription ; CTD, domaine carboxyterminal ; CAP, facteurs de la coiffe ; SF, facteurs d'épissage ; pA, facteurs de polyadénylation ; P, domaine CTD phosphorylé. Les ovaies roses représentent les complexes EJC se retrouvant en amont des jonctions exon-exon après la réaction d'épissage.

de l'étape d'épissage, les protéines fixées à la coiffe après sa synthèse et des complexes protéiques reconnaissant les sites de polyadénylation favorisent respectivement, comme nous l'avons vu précédemment, la reconnaissance du premier et dernier exon de l'ARN prémessager (cf § I.3.2.3). Récemment, il a même été montré que la nature des promoteurs transcriptionnels influençait les profils d'épissage alternatif des ARN prémessagers (Cramer *et al.* 1999).

Le domaine CTD est constitué chez l'homme de 52 répétitions de 7 acides aminés contenant trois résidus sérine. Il agit comme une plate-forme sur laquelle viennent s'assembler les complexes multiprotéiques impliqués dans la formation de la coiffe et de la queue polyadénylée, et des facteurs impliqués dans l'épissage (Fig. I.14) (pour revues, Maniatis & Reed 2002 ; Proudfoot *et al.* 2002).

De nombreuses phosphorylations et déphosphorylations vont ordonner ces différents mécanismes (pour revue, Shatkin & Manley 2000). Ainsi, lors de l'initiation de la transcription, les sérines en cinquième position de chaque répétition du domaine CTD sont progressivement phosphorylées, conduisant à terme, au départ des facteurs d'initiation de la transcription et à l'arrivée des enzymes synthétisant la coiffe. L'ARN en cours de synthèse a alors une longueur de 20 à 40 nucléotides.

Pendant l'élongation de la transcription, les sérines en seconde position des répétitions sont à leur tour phosphorylées, provoquant le départ des protéines de la coiffe et l'arrivée des facteurs d'épissage. Ces résidus sérine sont déphosphorylés avant l'étape de polyadénylation (pour revue, Bentley 2002 ; Proudfoot *et al.* 2002).

Lors de l'épissage, les protéines SR interagissent avec les protéines CASP¹⁸, elles-mêmes associées au domaine CTD de l'ARN polymérase II. Les protéines CASP contiennent en plus de leur domaine de fixation au domaine CTD, des domaines RS qui leur permettent d'établir des interactions avec les protéines SR lorsqu'ils sont phosphorylés ; certaines protéines CASP portent même un domaine de fixation à l'ARN (pour revue, Corden & Patturajan 1997).

Il a été proposé que le couplage entre transcription et épissage assure la jonction des exons dans leur ordre successif dans l'ARN prémessager. Chaque exon nouvellement synthétisé reste associé au domaine CTD jusqu'à la synthèse du prochain exon de l'ARN prémessager, ce qui faciliterait la juxtaposition des exons (Fig. I.14) (pour revues, Zeng & Berget 2000 ; Maniatis & Reed 2002).

Outre son rôle au niveau du domaine CTD, la phosphorylation va jouer également un rôle au niveau de la localisation et de la disponibilité des facteurs de transcription, d'épissage et de polyadénylation. L'épissage des ARN prémessagers eucaryotes a ainsi lieu dans le nucléoplasme des cellules et implique des mouvements de facteurs entre leurs sites de stockage dans le noyau et le lieu de la réaction d'épissage.

Les facteurs d'épissage sont en effet stockés dans des zones précises du noyau, dénommées jusqu'à présent « speckle » et rebaptisées récemment SFC¹⁹ (pour revue, Misteli 2000). Ces compartiments occupent environ 20% du volume total du noyau et ne sont pas délimités par une membrane. Ils possèdent, en plus de concentrations élevées de facteurs d'épissage, des facteurs de transcription, de polyadénylation et des protéines ribosomiques, ce qui représenterait environ 150 protéines au total

¹⁸« CTD-associated SR-like protein »

¹⁹pour « splicing factor compartment »

(Fig. I.15).

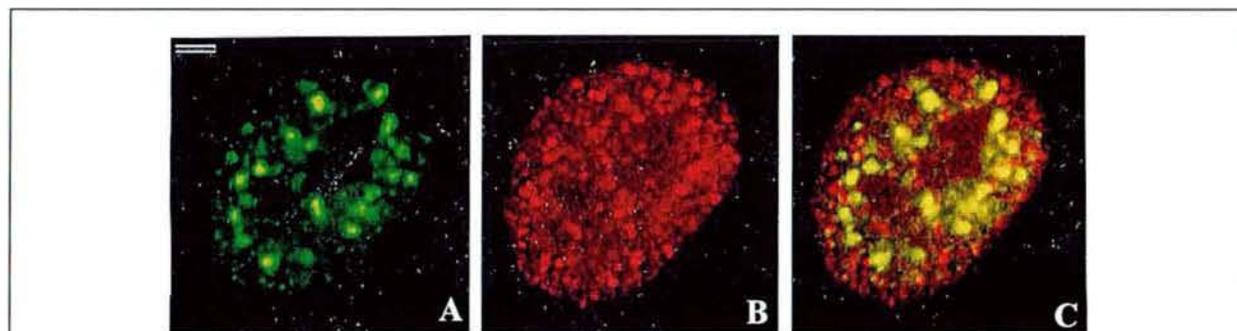


FIG. I.15 – *Les compartiments SFC.* (Zeng et al. 1997 ; pour revue, Misteli 2000).
 (A) Coloration par immunofluorescence de la protéine SC35. Cette protéine est utilisée comme marqueur des compartiments SFC dans les noyaux.
 (B) Coloration par immunofluorescence de l'ARN polymérase II.
 (C) Superposition des deux images.

Sachant que l'épissage de la majorité des ARN pré-messagers se fait co-transcriptionnellement, les facteurs d'épissage doivent donc se déplacer vers les sites de transcription, localisés en général à la périphérie des SFC, ou ailleurs dans le nucléoplasme.

La phosphorylation des facteurs d'épissage va permettre leur migration vers les sites actifs de transcription tandis que leur déphosphorylation les font se réassocier aux SFC (Misteli *et al.* 1998 ; pour revue, Misteli 2000). C'est le cas notamment des protéines SR, dont les protéines kinase les phosphorylant sont également localisées dans les SFC (cf § I.3.2.1). Ces kinases pourraient par ailleurs moduler l'épissage des ARN pré-messagers en jouant sur la disponibilité des protéines SR lors de la réaction d'épissage.

I.3.5.2 Lien entre l'épissage et l'export des ARN messagers

Le couplage entre les étapes d'épissage et d'export implique l'intervention d'une machinerie spécialisée dans l'export des ARN ayant subi un épissage, le complexe EJC²⁰. Les ARNm épissés sont ainsi transportés plus efficacement que les mêmes ARN non épissés (Le Hir *et al.* 2000 ; pour revue, Dreyfuss *et al.* 2002).

Une fois les réactions d'épissage achevées, les introns excisés par la machinerie d'épissage sont dissociés du spliceosome grâce à l'intervention du facteur mDEAH9 (Gee *et al.* 1997) puis, comme les ARN pré-messagers défectueux dans l'assemblage du spliceosome, ils sont pris en charge par des complexes de protéines hnRNP et dégradés (pour revue, Reed & Hurt 2002).

Les ARN matures sont quant à eux libérés du spliceosome grâce à l'ARN hélicase HRH1 (Ohno & Shimura 1996) et rapidement exportés vers le cytoplasme pour y être traduits en protéines (Luo & Reed 1999 ; pour revue, Zenklusen & Stutz 2001). Pour ce faire, le complexe protéique EJC s'est assemblé en amont des jonctions exon-exon de l'ARN messager au cours de la réaction d'épissage

²⁰ « exon-exon junction complex »

(pour revue, Dreyfuss *et al.* 2002). Ce complexe contient les co-activateurs d'épissage SRm160 et RNPS1 (cf chapitre II), la protéine Y14 s'associant spécifiquement aux ARN épissés (Kataoka *et al.* 2000), les facteurs d'export nucléaire TAP (associé au pore nucléaire) (Kataoka *et al.* 2000) et ALY (Zhou *et al.* 2000), ainsi que les protéines Upf3 et magoh (pour revue, Reed & Hurt 2002). La protéine ALY est recrutée au niveau de l'ARN lors de la réaction d'épissage par le facteur UAP⁵⁶ (Luo *et al.* 2001 ; pour revue, Linder & Stutz 2001). Cette protéine permet également le transport des ARN non épissés, mais avec une plus faible efficacité (pour revue, Dreyfuss *et al.* 2002).

L'interaction de TAP avec le pore nucléaire permet l'export de l'ARN messenger et la plupart des facteurs du complexe EJC se dissocient au cours de l'export ou immédiatement après, seules Y14 et magoh restent associées à l'ARNm dans le cytoplasme (pour revue, Dreyfuss *et al.* 2002).

Les protéines hnRNP et SR capables de faire la navette entre le cytoplasme et le noyau participeraient également à l'export des ARN messagers (Mili *et al.* 2001). Les protéines SR agiraient en reconnaissant des séquences ARN précises (Huang & Steitz 2001).

Certains facteurs impliqués dans l'export interviennent encore dans le cytoplasme, dans le mécanisme de reconnaissance des codons STOP prématurés dans les ARN messagers. Ce mécanisme entraîne de tels ARNm vers une voie de dégradation (processus appelé NMD²¹) (pour revues, Reed & Hurt 2002 ; Maniatis & Reed 2002).

La reconnaissance du mauvais positionnement du codon STOP est assurée par les ribosomes en cours de traduction, lors de leur 1^{er} passage sur l'ARN messenger. Dans ce cas, les protéines Y14-magoh du complexe EJC, encore associées à l'ARNm, recruteraient les facteurs impliqués dans le système NMD. Dans le cas d'un bon positionnement du codon STOP, les protéines du complexe EJC se dissocient de l'ARNm à l'issue du 1^{er} cycle de traduction.

²¹ « non sense mediated decay »

Chapitre II

Régulation de l'épissage des ARN pré-messagers

II.1 Généralités sur l'épissage alternatif

Les eucaryotes supérieurs et les virus utilisent l'épissage alternatif pour accroître la capacité codante de leur génome. Chez la drosophile, par exemple, l'épissage de l'ARN pré-messager *Dscam*¹ peut produire potentiellement 38 016 ARN messagers différents (pour revue, Black 2000). Chez l'homme, l'épissage alternatif est un phénomène majeur, car il concerne 35 à 60% des 30 000 gènes (McPherson *et al.* 2001 ; Venter *et al.* 2001 ; Modrek & Lee 2002) et l'altération de cet épissage peut conduire à de graves maladies génétiques dont nous verrons quelques exemples.

L'épissage alternatif est, en outre, un moyen utilisé par les organismes pour contrôler l'expression de leurs gènes. On note, par exemple, des cas d'incorporation de codons STOP dans l'ARN mature, conduisant à la production de protéines tronquées ou d'ARNm instables.

II.1.1 Modes d'épissage alternatif

Selon le mode d'épissage alternatif mis en jeu, on peut définir plusieurs catégories d'exons :

- **Des exons dits exons cassettes** : ces exons sont soit inclus, soit exclus de l'ARN messager mature (Fig. II.1a). Ils sont en général de petite taille, tissu-spécifique, et codent un domaine supplémentaire d'une isoforme de la protéine. Par exemple, l'épissage alternatif de l'ARN pré-messager codant la sous-unité $\gamma 2$ du récepteur à canal-ionique GABA_A permet d'obtenir des récepteurs dont l'ouverture du canal chlore est modulée. L'inclusion d'un exon cassette de 24 nts conduit à la production d'une forme longue de la sous-unité ($\gamma 2L$), ayant une affinité pour certaines substances pharmacologiques, différente de celle de la sous-unité courte $\gamma 2C$ (pour revue, Grabowski & Black 2001). Le proto-oncogène *c-src*, codant une tyrosine kinase, subit également un épissage tissu-spécifique. L'inclusion de l'exon N1 dans les neurones conduit à l'insertion de 6 acides aminés dans le domaine SH3 de la protéine kinase, ce qui altère sa capacité d'interaction avec d'autres protéines (pour revue, Brown & Cooper 1996).
- **Des exons dits exons mutuellement exclusifs** : ces exons sont trouvés en succession dans l'ARN pré-messager et bien qu'ils aient des sites 5' et 3' d'épissage fonctionnels, ils ne sont jamais présents ensemble dans les ARN matures (Fig. II.1b). Ces exons codent généralement une version alternative

¹ « Down syndrome cell adhesion molecule »

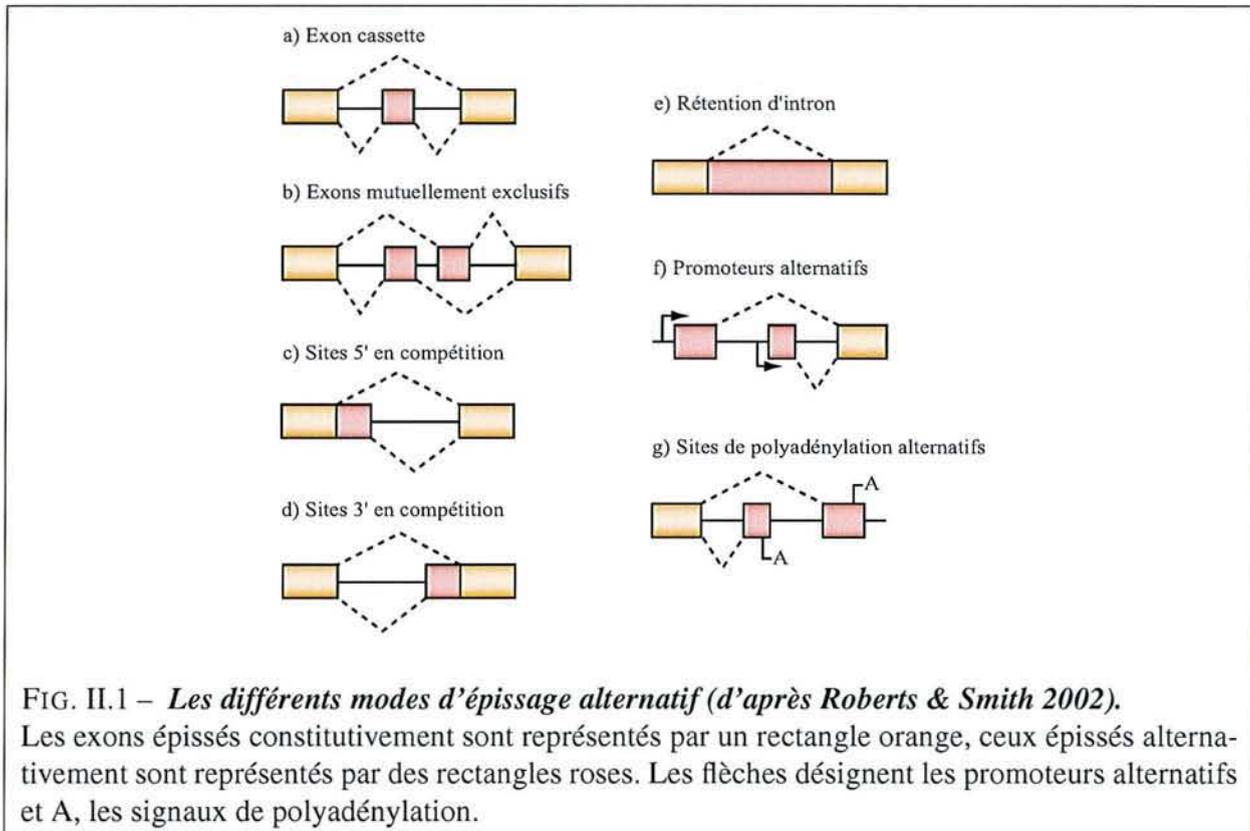


FIG. II.1 – *Les différents modes d'épissage alternatif (d'après Roberts & Smith 2002).*

Les exons épissés constitutivement sont représentés par un rectangle orange, ceux épissés alternativement sont représentés par des rectangles roses. Les flèches désignent les promoteurs alternatifs et A, les signaux de polyadénylation.

du même domaine d'une protéine. Par exemple, l'ARN pré-messager *FGFR-2*² contient 2 exons mutuellement exclusifs, appelés IIIb et IIIc chez la souris, KSAM et BEK chez l'homme. La forme *FGFR-2 IIIb* du récepteur est capable de fixer les facteurs de croissance FGF7 et 10, ce que ne peut pas faire la forme *FGFR-2 IIIc* (Miki *et al.* 1992).

- **Des exons variables du fait de la présence de sites donneurs ou accepteurs alternatifs :** des sites d'épissage alternatifs peuvent être présents dans une séquence exonique et la sélection d'un des sites d'épissage alternatifs entraîne l'exclusion d'une partie de l'exon dans l'ARN messager (Fig. II.1c et d). Ce système est souvent utilisé par les virus, ce qui leur permet de produire de nombreux ARN messagers différents. Par exemple, l'unité E1A de l'adénovirus contient 3 sites 5' en compétition pour un seul site 3'. Leur utilisation permet la production des ARNm 9S, 12S et 13S. L'unité L1 de ce même virus contient, quant à elle, 2 sites 3' en compétition, les sites 55,52K et IIIa (pour revue, Akusjarvi & Stévenin 2003). L'ARN pré-messager du virus HIV-1 contient également des sites d'épissage en compétition, 4 sites donneurs et 8 sites accepteurs, ce qui permet la production d'une quarantaine d'ARN messagers (Purcell & Martin 1993).
- **Des 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 (Fig. II.1e). 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).
- **Des exons alternatifs du fait de la présence de promoteurs ou de sites de polyadénylation alternatifs :** dans ce cas de figure, le choix d'un promoteur ou d'un site de polyadénylation influe

²« fibroblast growth factor receptor 2 »

sur l'identité des exons terminaux (Fig. II.1f et g). Ainsi, l'épissage alternatif du transcrit primaire CT/CGRP³ produit 2 ARN messagers, celui de la calcitonine exprimé dans la thyroïde et celui codant CGRP, exprimé dans les neurones. Dans la thyroïde, l'utilisation du site de polyadénylation de l'exon 4 conduit à l'inclusion de l'extrémité 5' de cet exon dans l'ARN messenger mature (Lou & Gagel 1998).

Les différents modes d'épissage alternatif que nous venons de décrire aboutissent à des ARNm différents, dont la production est régulée selon la nature des sites d'épissage et la présence de séquences régulatrices, selon le type cellulaire, l'état de différenciation de la cellule, l'activation des voies de signalisation ou encore, le génotype sexuel. Tous les ARN pré-messagers dont l'épissage est régulé ont en commun la caractéristique d'avoir des sites d'épissage faibles.

II.1.2 Les sites d'épissage alternatifs sont des sites d'épissage faibles

II.1.2.1 Caractéristiques des sites d'épissage faibles

Alors que les sites d'épissage forts possèdent des signaux correspondant aux séquences consensus, les sites d'épissage faibles sont au contraire divergents. Pour les sites 3' faibles, par exemple, la séquence polypyrimidine est souvent entrecoupée de purines et un point de branchement non canonique, c'est-à-dire autre qu'une adénosine, est utilisé. Ces caractéristiques rendent moins efficace l'interaction des composants spliceosomaux avec les séquences d'épissage.

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

Les ARN pré-messagers présentant des sites d'épissage faibles contiennent des éléments de séquences introniques et/ou exoniques, qui agissent en *cis* pour stimuler ou au contraire, réprimer l'épissage. Ces séquences interagissent en général avec des facteurs protéiques dont nous présenterons le mécanisme d'action dans les paragraphes II.2 et II.3. Nous décrirons, en particulier, le cas des protéines SR et hnRNP, car ce sont les seuls facteurs régulant l'épissage de l'ARN du virus HIV-1 identifiés à ce jour. A titre d'exemple, nous présenterons l'état des connaissances concernant la régulation de l'épissage de quelques ARN pré-messagers, qui constituent des modèles pour la compréhension des mécanismes d'épissage alternatif.

Séquences activatrices

Différentes approches ont été utilisées pour identifier les séquences agissant en *cis*. En particulier, l'approche SELEX⁴ (Tuerk & Gold 1990) a permis d'identifier de nombreuses séquences activatrices d'épissage. La stratégie consistait à utiliser un ARN dérivé du pré-messager *dsx*⁵, qui intervient dans la détermination du sexe chez la drosophile, dont l'épissage nécessite la présence d'une séquence activatrice exonique ESE⁶, et de tester sa capacité à être épissé dans un extrait nucléaire (Schaal & Maniatis 1999b). L'élément ESE était remplacé par une séquence dégénérée et les ARN pré-messagers

³« calcitonin/calcitonin gene related peptide »

⁴« systematic evolution of ligands by exponential enrichment »

⁵« doublesex »

⁶« exonic splicing enhancer »

correspondants étaient épissés dans un extrait nucléaire. Les ARNm produits grâce à la présence de séquences ayant des propriétés activatrices étaient ensuite isolés, et les exons contenant les séquences activatrices étaient amplifiées par RT-PCR et clonés dans l'ARN prémessager test, afin de débiter un nouveau cycle de sélection. Après plusieurs cycles de sélection, des éléments activateurs ont ainsi été mis en évidence et classés en 7 groupes sur la base de leur séquence (Tab. II.1).

Ces séquences activatrices se sont avérées être aussi bien riches en purines qu'en pyrimidines (cf Tab. II.1) et un certain nombre d'entre elles correspond à des sites de fixation des protéines SR, comme nous le verrons dans le § II.2.2.

Type d'activateur	Séquence consensus
Classe I Séquences riches en purines	GGGGA GGAGGA GGAGA AGAGG AGGAG
Classe II Séquences riches en pyrimidines	UCUCC UCUUC UCCUC
Classe III	(U)GGACCNG
Classe IV	(C)CACC(C)
Classe V	YGCCGCC
Classe VI	RGAACYU
Classe VII Activateurs forts sans consensus	CCGACGCCAUGGACGACGAG GGCUGCCAGACGGAAUUGG CCGUGACAGCAUCGGCGG CGUCGGCAGGUGGUCCCG UCUGGAUCCUGCGGAUGG

TAB. II.1 – *Séquences consensus d'éléments ayant des propriétés activatrices de l'épissage lorsqu'ils sont placés en aval d'un site 3'* (Schaal & Maniatis 1999b).

Séquences inhibitrices

Afin d'identifier des séquences exoniques inhibitrices d'épissage présentes dans le génome humain, une étude similaire, utilisant un système dérivé du gène *dhfr*⁷ du hamster, a été réalisée (Fairbrother & Chasin 2000). Lors de l'épissage de l'ARN prémessager correspondant, l'exclusion de l'exon central conduit à un phénotype [*dhfr*⁺] des cellules, tandis que l'inclusion de l'exon provoque un phénotype [*dhfr*⁻]. L'insertion dans l'exon d'une séquence inhibitrice en aval du site 3' d'épissage permet l'exclusion de cet exon et conduit au phénotype [*dhfr*⁺] des cellules.

Des séquences de 50 à 250 paires de bases, produites par hydrolyse de banques d'ADN de placenta humain ou d'*E. coli* (pris comme contrôle), ont été insérées dans l'exon dont l'épissage est régulé.

⁷« dihydrofolate reductase »

Après transfection des cellules CHO⁸ *dhfr*⁻, seules les séquences humaines se sont avérées avoir des propriétés inhibitrices de l'épissage. Parmi elles, on trouve des séquences présentant des similitudes avec les séquences des sites d'épissage (ou pseudo-sites d'épissage), qui fixent des complexes spliceosomaux inactifs en épissage.

Les autres séquences inhibitrices sélectionnées sont riches en pyrimidines, ou correspondent à des motifs de type YAG ou GGG. Ces derniers constituent des sites de fixation des protéines hnRNP, qui sont les principaux facteurs inhibiteurs de l'épissage des ARN pré-messagers chez les vertébrés (Fairbrother & Chasin 2000).

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

II.2.1 Les protéines hnRNP et leurs sites de fixation

Dès leur transcription, les ARN pré-messagers sont pris en charge par un grand nombre de protéines, en particulier, les protéines hnRNP. Cette vingtaine de protéines très conservées, dont la taille varie de 34 à 120 kDa, est présente en quantités importantes dans le noyau (jusqu'à 10 μ M pour la protéine hnRNP A1) (pour revues, Dreyfuss *et al.* 1993, 2002 ; Krecic & Swanson 1999). Elles s'associent aux ARN pré-messagers pour former des particules hétérogènes appelées hnRNP⁹. Dans ces particules, les protéines hnRNP les plus abondantes sont A1, A1^B, A2, B1, C1 et C2 (Beyer *et al.* 1977).

Les protéines hnRNP sont majoritairement localisées dans le nucléoplasme (Dreyfuss *et al.* 1993). Néanmoins, certaines d'entre elles (les protéines hnRNP A/B, D, E, I, et K) sont des protéines « navettes » capables de transiter entre le noyau et le cytoplasme. Dans le noyau, elles sont associées aux ARN pré-messagers et dans le cytoplasme, aux ARNm (Piñol Roma & Dreyfuss 1992, 1993 ; Michael *et al.* 1995a, b).

Le signal permettant l'import de la protéine hnRNP A1 est le motif M9 localisé dans la région N-terminale de la protéine (cf Fig. II.2 ; Siomi & Dreyfuss 1995 ; Weighardt *et al.* 1995). Il fonctionne en interagissant avec un récepteur d'import nucléaire, la transportine (Pollard *et al.* 1996). Le motif M9 constitue également un signal d'export nucléaire pour la protéine hnRNP A1 qui pourrait traverser le pore nucléaire, tout en restant associée à l'ARNm (Piñol Roma & Dreyfuss 1992 ; pour revue, Nakielny & Dreyfuss 1999).

Suivant leur site de fixation sur l'ARN, ces protéines vont interférer avec l'assemblage du spliceosome. Elles ont en effet une large spécificité de reconnaissance, ce qui leur permet de s'associer à tous les ARN pré-messagers (Tab. II.2). Certaines reconnaissent préférentiellement des séquences riches en C (hnRNP E, I et K), d'autres des séquences riches en G (hnRNP A/B, E, H, H', 2H9 et M). Par ailleurs, certaines protéines hnRNP ont en plus la capacité de reconnaître avec une très forte affinité des séquences spécifiques (hnRNP A/B) (pour revue, Dreyfuss *et al.* 2002).

Les protéines hnRNP se fixent à l'ARN par des domaines de type RRM, RGG ou KH. Les autres

⁸ « chinese hamster ovary »

⁹ « heterogenous nuclear ribonucleic particles »

hnRNP	Domaines de fixation à l'ARN	Rôles cellulaires	Sites de fixation sur l'ARN
A1/A1 ^B	2 x RRM, RGG	Epissage, biogenèse des télomères, export des ARNm	uaggg(a/u), UAG
A2/B1	2 x RRM, RGG	Epissage, localisation des ARNm	gcccaaggagcc(rts)
C1/C2	1 x RRM	Epissage, stabilité et rétention nucléaire des ARNm	UAGR
D1/D2	2 x RRM, RGG	Transcription, stabilité des ARNm, recombinaison, biogenèse des télomères	AUUUUA
E1/E2	3 x KH	Traduction, stabilité des ARNm	Poly(G), Poly(C)
F	3 x RRM	Epissage, formation de la coiffe des ARN pré-messagers, transcription	GGGA, Poly(G)
G	1 x RRM, RGG	Epissage, polyadénylation	
H/H'	3 x RRM	Epissage, polyadénylation	GGGA/C, Poly(G)
2H9	2 x RRM, 2 x G/Y	Epissage, réponse au choc thermique	GGGA
I (PTB)	4 x RRM	Epissage, polyadénylation	Poly(U), séquences riches en pyrimidines
K	3 x KH, RGG	Transcription, traduction	Poly(C)
L	4 x RRM	Stabilité et export des ARNm	
M	3 x RRM	Epissage, réponse au choc thermique	Poly(G), Poly(U)
P2 (TLS/FUS)	RRM, RGG	Oncogenèse, épissage	GGUG
Q	RRM, RGG	Epissage	
R	3 x RRM, RGG		
U	RGG	Transcription, rétention nucléaire	

TAB. II.2 – *Les protéines hnRNP majeures* (d'après Swanson & Dreyfuss 1988 ; Dreyfuss et al. 1993 ; Honoré et al. 1999 ; Caputi & Zahler 2001 ; Lerga et al. 2001 ; Kafasla et al. 2002 ; pour revue, Dreyfuss et al. 2002).

Les séquences écrites en minuscules correspondent à des séquences ARN ayant une très forte affinité pour les protéines.

domaines des protéines hnRNP sont composés de différents motifs leur permettant d'établir des interactions protéine-protéine (pour revue, Dreyfuss *et al.* 2002). Certaines protéines hnRNP sont notamment capables d'interagir entre elles, même en absence d'ARN (Kim *et al.* 2000).

Du fait qu'elles sont associées aux ARN pré-messagers dès la synthèse de ces derniers et que certaines d'entre elles peuvent rester associées aux ARN messagers jusque dans le cytoplasme, les protéines hnRNP vont jouer des rôles variés dans les différentes étapes de synthèse, maturation, transport, traduction et dégradation des ARNm. Nous ne décrivons ici que les principales protéines hnRNP jouant un rôle majeur dans l'épissage.

II.2.1.1 Protéines hnRNP A/B

Organisation structurale

Les protéines de la famille hnRNP A/B sont issues de l'épissage alternatif d'ARN pré-messagers produits à partir de 2 gènes distincts. L'ARN pré-messager codant la protéine hnRNP A1 subit un épissage alternatif, conduisant à l'insertion d'un exon supplémentaire produisant la protéine hnRNP A1^B (encore appelée hnRNP B2 ; Buvoli *et al.* 1990). Les protéines hnRNP A2 et B1 sont 2 isoformes produits par épissage alternatif à partir d'un gène différent de celui permettant d'exprimer A1 et A1^B. Les 2 gènes doivent résulter de la duplication d'un gène initial (Biamonti *et al.* 1994). Les protéines hnRNP A2 et B1 diffèrent par l'insertion de 12 acides aminés dans la région N-terminale de la hnRNP B1 (Burd *et al.* 1989 ; Kozu *et al.* 1995). Il existe par ailleurs un isoforme de ces protéines, hnRNP B0, exprimé spécifiquement dans les testicules (Kamma *et al.* 1999).

L'identification de la protéine hnRNP A3 est plus récente (Ma *et al.* 2002). Cette protéine est cytoplasmique et elle pourrait jouer un rôle dans la localisation des ARNm dans le cytoplasme. Plusieurs isoformes de cette protéine sont générés par épissage alternatif (Ma *et al.* 2002).

Les 5 protéines hnRNP A/B portent dans leur région N-terminale 2 domaines RRM très conservés ; les différences de séquences sont plus importantes dans leur domaine C-terminal riche en glycines, contenant une boîte RGG et un signal d'import / export nucléaire M9 (Fig. II.2 ; Burd *et al.* 1989 ; Siomi & Dreyfuss 1995 ; Michael *et al.* 1995a ; Izaurralde *et al.* 1997). Seuls des anticorps monoclonaux spécifiquement dirigés contre des régions variables peuvent donc permettre de distinguer ces protéines. Les anticorps polyclonaux généralement utilisés ne permettent pas une distinction.

La protéine hnRNP A1 est la protéine la plus étudiée de cette famille. Elle se fixe à l'ADN ou à l'ARN avec la même spécificité. Le domaine UP1 de la protéine, constitué des 2 domaines RRM, a été cristallisé puis analysé aux rayons X, seul ou complexé avec un fragment de 12 nucléotides d'ADN télomérique (Shamoo *et al.* 1997 ; Xu *et al.* 1997 ; Ding *et al.* 1999).

Dans le cristal contenant le complexe UP1-ADN, le domaine UP1 est organisé sous forme de dimère et chacun des 2 domaines RRM d'un monomère est associé à un ADN simple-brin (Fig. II.3). Les deux domaines RRM de la protéine hnRNP A1 sont reliés par un enchaînement de 13 acides aminés et sont capables de se replier indépendamment l'un de l'autre. Leur positionnement relatif est défini par la formation de 2 ponts salins entre des acides aminés des 2 domaines (Shamoo *et al.* 1997 ; Xu *et al.* 1997 ; Ding *et al.* 1999).

L'analyse cristallographique par rayons X du domaine UP1 a révélé une organisation similaire à celle

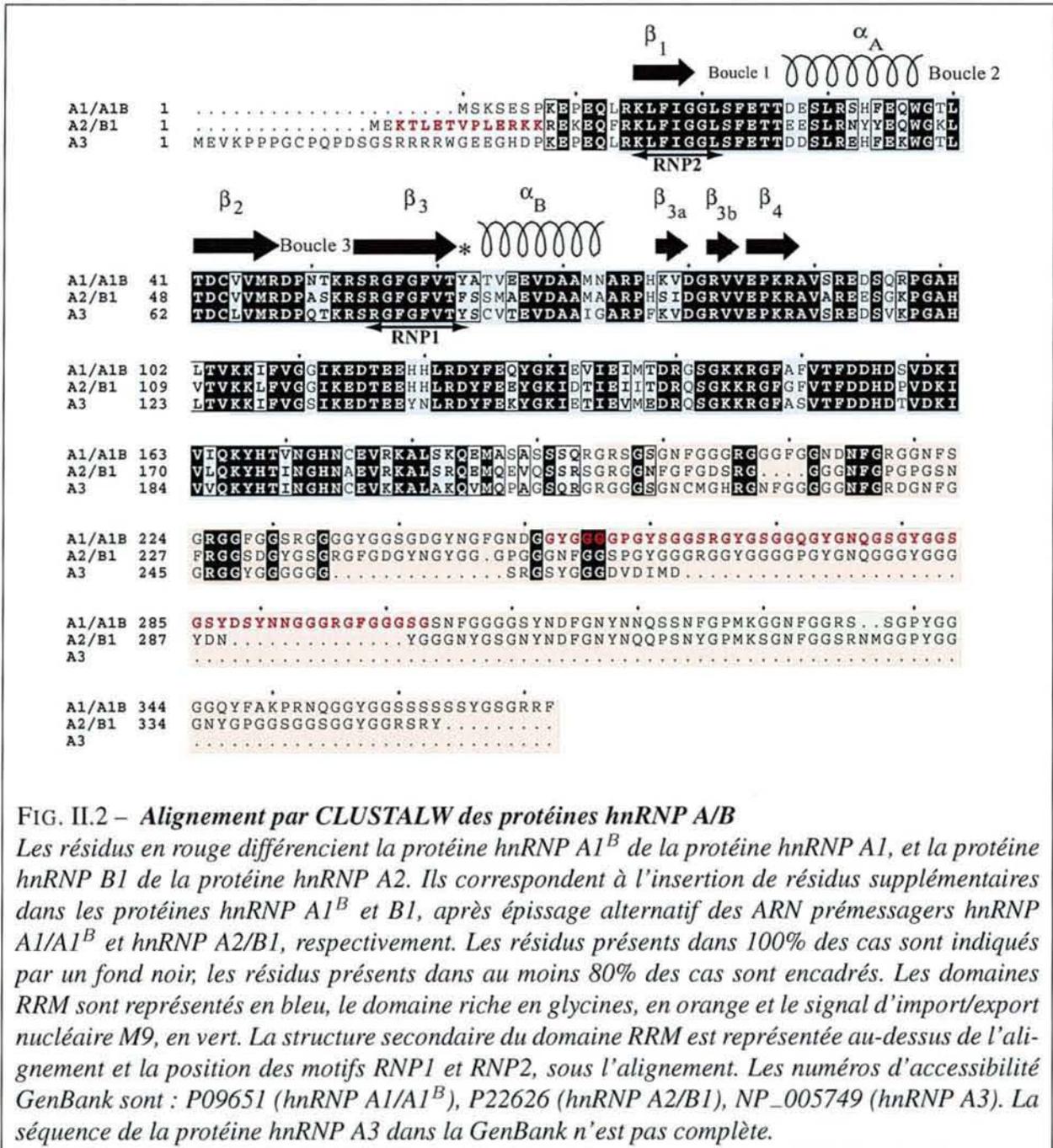
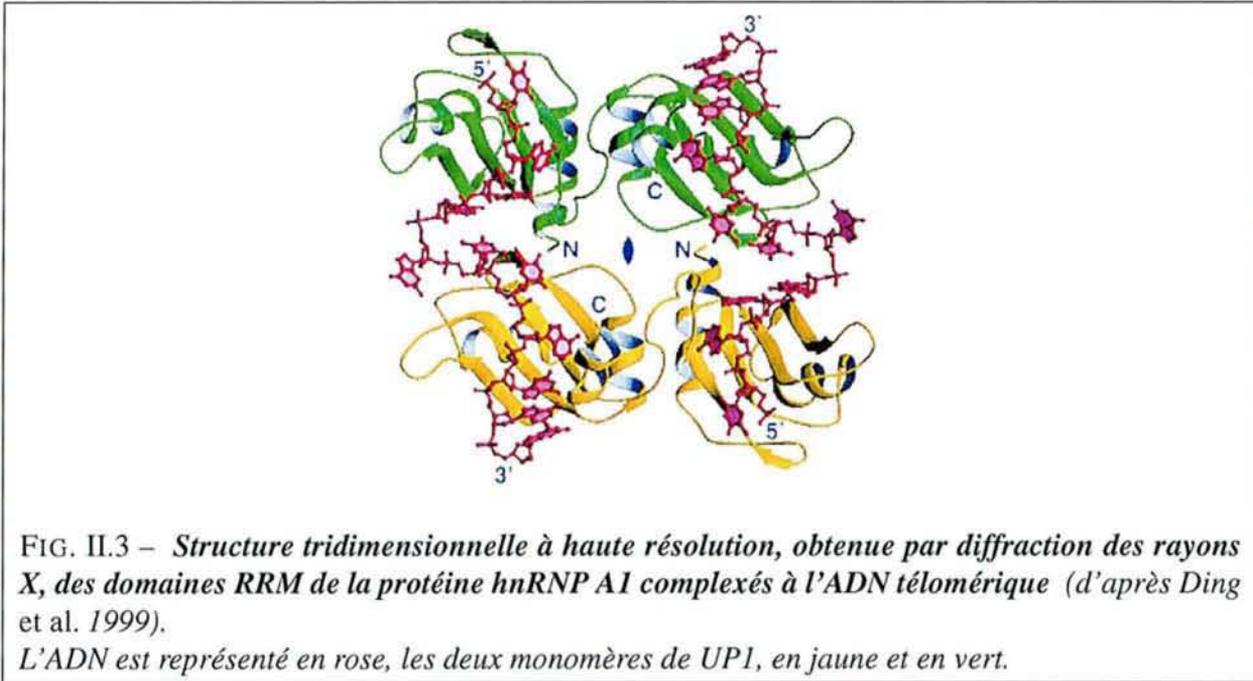


FIG. II.2 – *Alignement par CLUSTALW des protéines hnRNP A/B*

Les résidus en rouge différencie la protéine hnRNP A1^B de la protéine hnRNP A1, et la protéine hnRNP B1 de la protéine hnRNP A2. Ils correspondent à l'insertion de résidus supplémentaires dans les protéines hnRNP A1^B et B1, après épissage alternatif des ARN pré-messagers hnRNP A1/A1^B et hnRNP A2/B1, respectivement. Les résidus présents dans 100% des cas sont indiqués par un fond noir, les résidus présents dans au moins 80% des cas sont encadrés. Les domaines RRM sont représentés en bleu, le domaine riche en glycines, en orange et le signal d'import/export nucléaire M9, en vert. La structure secondaire du domaine RRM est représentée au-dessus de l'alignement et la position des motifs RNP1 et RNP2, sous l'alignement. Les numéros d'accessibilité GenBank sont : P09651 (hnRNP A1/A1^B), P22626 (hnRNP A2/B1), NP_005749 (hnRNP A3). La séquence de la protéine hnRNP A3 dans la GenBank n'est pas complète.

des domaines RRM de la protéine U1-A de la snRNP U1 : chaque motif RRM forme un domaine structural globulaire compact de type (β_1 - α_1 - β_2 - β_3 - α_2 - β_4), dans lequel le feuillet de brins β est anti-parallèle et les 2 hélices α sont positionnées du même côté du feuillet (Nagai *et al.* 1990 ; Shamo *et al.* 1997 ; Xu *et al.* 1997 ; Vitali *et al.* 2002). Au sein de chaque domaine, les acides aminés chargés et aromatiques des séquences RNP1 et RNP2 (les deux séquences signatures des domaines RRM, localisées au milieu des brins β_1 et β_3) sont exposés au solvant et peuvent interagir avec l'ARN par des liaisons hydrogène et par formation d'interactions hydrophobes entre des bases et acides aminés aromatiques (Fig. II.2 et II.3). En effet, ces acides aminés aromatiques, localisés en C-terminal des motifs RNP1 et RNP2, sont orientés de façon à permettre un empilement avec les bases de l'ARN. Ces acides aminés constituent, avec les acides aminés des hélices α , le cœur hydrophobe du domaine.



Les acides aminés conservés des motifs RNP1 et RNP2 sont importants pour la fixation à l'ARN mais ils ne sont pas les éléments déterminants de la spécificité de reconnaissance. Ce sont les acides aminés des régions variables des motifs RNP qui agissent comme déterminants de la spécificité (Kenan *et al.* 1991). Le feuillet β portant les 2 motifs RNP constitue une plate-forme sur laquelle vient se fixer l'ARN (ou l'ADN ; Fig. II.3). L'ARN reste ainsi accessible pour une interaction avec d'autres ARN ou protéines (pour revue, Burd & Dreyfuss 1994b).

Des expériences basées sur l'emploi de la méthode SELEX réalisées par Burd et Dreyfuss ont montré que la protéine hnRNP A1 se lie préférentiellement à des séquences ARN simple-brin de type UAGGG(U/A) présentant des similitudes avec les séquences consensus des sites 5' et 3' d'épissage (cf Tab. II.2 ; Burd & Dreyfuss 1994a). La protéine se fixe également à des séquences régulatrices d'épissage contenant des triplets UAG, c'est le cas notamment dans le mécanisme de production de l'ARNm *tat* du virus HIV-1 (Caputi *et al.* 1999).

La reconnaissance de ces séquences ARN par la protéine hnRNP A1 peut s'expliquer par les informations qu'apporte la structure établie pour le complexe formé entre le domaine UP1 et l'ADN télomérique. En effet, une interaction directe est observée entre le domaine RRM1 et les 4 nucléotides TAGG, et entre le domaine RRM2 et les 5 nucléotides TTAGG. Au sein de ces interactions, les 2 domaines RRM interagissent exactement de la même façon avec chacun des triplets TAG (Ding *et al.* 1999).

Le domaine RGG situé en N-terminal du domaine riche en glycines constitue un autre motif de fixation à l'ARN. Ce motif est fréquemment présent dans des protéines qui possèdent des capacités de reconnaissance large de l'ARN, en particulier les protéines nucléolaires (pour revue, Ginisty *et al.* 1999).

Dans le cas des protéines hnRNP, à l'exception de hnRNP U, où il semble être le seul domaine de fixation à l'ARN, le domaine RGG est présent en plus de domaines forts de fixation de type RRM ou

KH (cf Tab. II.2). En interagissant avec l'ARN par liaisons hydrogène, il faciliterait l'interaction des domaines RRM de la protéine hnRNP A1 avec l'ARN, ce qui augmenterait l'affinité générale de la protéine hnRNP A1 pour l'ARN (Burd & Dreyfuss 1994b).

Une fois fixées à des sites de haute affinité sur l'ARN, les protéines hnRNP A/B sont capables de se lier à des séquences de plus faible affinité et de multimériser le long de l'ARN (Burd & Dreyfuss 1994a, b). Le domaine riche en glycines permettrait la fixation coopérative de ces protéines sur l'ARN en favorisant leur dimérisation (Dreyfuss *et al.* 1993). Le fait que chaque domaine RRM de la protéine hnRNP A1 soit capable d'interagir avec une molécule d'ARN différente (d'après l'analyse cristallographique du complexe UP1-ADN, cf Fig. II.3 ; Ding *et al.* 1999), explique sans doute la capacité qu'ont les protéines hnRNP A/B de favoriser l'appariement d'ARN ayant des séquences complémentaires (pouvoir d'« annealing », Burd & Dreyfuss 1994b ; Blanchette & Chabot 1999). Toutes ces propriétés de fixation à l'ARN de la protéine hnRNP A1 lui permettent de réguler, voire réprimer, l'épissage des ARN pré-messagers nucléaires, en masquant leurs sites d'épissage vis-à-vis de la machinerie spliceosomale. Nous allons à présent aborder ces différents modes de régulation.

Rôle de la protéine hnRNP A1 dans l'épissage

Inhibition des sites 3' d'épissage : ce type d'inhibition par la protéine hnRNP A1 a été observé dans le cas de la régulation de la production de l'ARNm *tat* de HIV-1 (Del Gatto-Konczak *et al.* 1999 ; Caputi *et al.* 1999), ainsi que dans la régulation de l'épissage de l'ARN pré-messager *FGFR2* (Del Gatto-Konczak *et al.* 1999). Au niveau de ces ARN, la protéine hnRNP A1 se fixe sur une séquence inhibitrice exonique contenant des triplets UAG et réprime ainsi l'utilisation du site 3' amont, en empêchant l'assemblage des complexes spliceosomaux autour de ce site d'épissage (Del Gatto-Konczak *et al.* 1999 ; Caputi *et al.* 1999).

Inhibition des sites 5' d'épissage : la protéine hnRNP A1 agirait en empêchant la formation des complexes spliceosomaux sur le site 5' d'épissage, mais les modes d'action semblent plus variés que dans le cas des sites 3'.

Polymérisation de la protéine hnRNP A1 sur l'ARN pré-messager et antagonisme avec la protéine SR ASF/SF2 : comme nous l'avons vu précédemment, la protéine hnRNP A1 possède une large spécificité de reconnaissance de séquences ARN et elle est capable de multimériser. Ainsi, lorsqu'elle est présente en quantités importantes, la protéine hnRNP A1 multimérise le long de l'ARN pré-messager et empêche l'accès des facteurs spliceosomaux, en particulier, de la snRNP U1 au site d'épissage. Sachant que la protéine ASF/SF2 favorise la fixation de la snRNP U1 en se fixant à proximité du site 5' d'épissage, la compétition entre les protéines hnRNP A1 et ASF/SF2 pour leur fixation à l'ARN module le taux d'assemblage de la snRNP U1. Nous décrirons plus en détails le mécanisme de cet antagonisme entre les 2 protéines dans le § II.2.2.

Recrutement de la snRNP U1 sur des pseudo-sites 5' d'épissage : la protéine hnRNP A1 peut participer à la formation de complexes entraînant le recrutement de la snRNP U1 sur des séquences

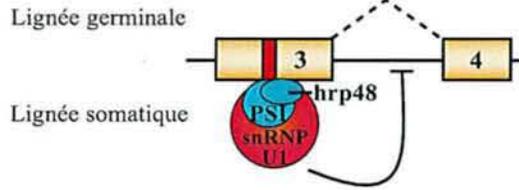


FIG. II.4 – Régulation de l'épissage de l'élément P de la drosophile par un pseudo-site 5' (d'après Smith & Valcárcel 2000).

présentant une similitude avec un site 5' d'épissage. C'est le cas chez *Drosophila melanogaster*, où la protéine homologue à la protéine hnRNP A1 humaine, hrp48, participe au recrutement de la snRNP U1 sur un pseudo-site 5' d'épissage, localisé dans l'exon 3 de la transposase des éléments P. Le 3^e intron de l'ARN, IVS3, n'est épissé que dans la lignée germinale. Il est conservé dans la lignée somatique, ce qui conduit à la production d'une protéine tronquée réprimant la transposition (Fig. II.4). La protéine hrp48 agit au sein d'un complexe protéique constitué de 4 protéines, dont la protéine PSI, exprimée uniquement dans les cellules somatiques, ce qui assure la tissu-spécificité de l'épissage alternatif de l'ARN (Adams *et al.* 1997 ; Siebel *et al.* 1992, 1994, 1995 ; pour revue, Smith & Valcárcel 2000). Cette protéine PSI est capable d'interagir directement avec la snRNP U1 (Labourier *et al.* 2001, 2002).

Séquestration de sites d'épissage par formation de structures secondaires dans l'ARN pré-messager : la protéine hnRNP A1 est capable de réguler l'épissage de son propre ARN pré-messager, contenant

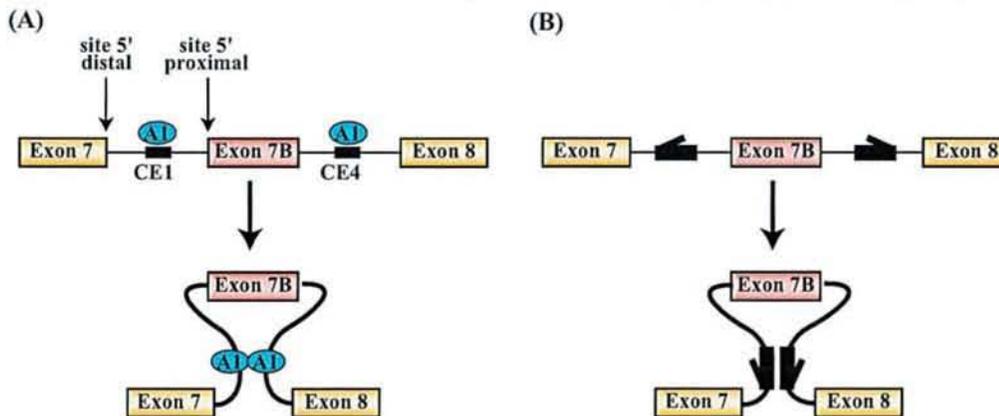


FIG. II.5 – Rôle de la protéine hnRNP A1 dans l'épissage alternatif de son propre ARN pré-messager (d'après Blanchette & Chabot 1999 ; Nasim *et al.* 2002).

(A) L'interaction entre deux molécules de protéine hnRNP A1, fixées sur leur sites de fixation CE1 et CE4 de part et d'autre de l'exon 7B, conduit à l'exclusion de cet exon.

(B) Le remplacement des séquences CE1 et CE4 par deux séquences pouvant s'apparier conduit également à l'exclusion de l'exon 7B.

2 sites 5' d'épissage en compétition. Elle favorise l'utilisation du site 5' distal, ce qui conduit à l'exclusion de l'exon 7B (spécifique de l'isoforme hnRNP A1^B), sans que la fixation de la snRNP U1 au site proximal soit affectée (Fig. II.5-A ; Chabot *et al.* 1997).

Plusieurs études ont permis de comprendre le mécanisme d'action impliqué (Eperon *et al.* 1993, 2000 ; Chabot *et al.* 1997 ; Blanchette & Chabot 1999). Des sites de fixation de haute affinité de la protéine sont localisés de part et d'autre de l'exon 7B. En se fixant à ces sites, la protéine hnRNP A1 va former une boucle dans l'ARN, ce qui, d'une part, séquestre l'exon 7B, et d'autre part, réduit la distance physique entre le site 5' distal et le site 3'. Ceci conduit à l'exclusion de l'exon 7B (Fig. II.5-A).

Le fait que des séquences répétées inversées placées de part et d'autre de l'exon 7B ont exactement le même effet sur l'exclusion de cet exon que les sites de fixation de la protéine hnRNP A1 confirme le modèle (Fig. II.5-A ; Nasim *et al.* 2002).

Rôle des protéines hnRNP A2 et B1 dans le transport

Les protéines hnRNP A2 et B1 interviendraient dans des phénomènes de transport des ARNm dans le cytoplasme des cellules polarisées, comme 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 sur ces ARN (Ainger *et al.* 1997), telles que les séquences A2RE¹⁰ de l'ARN du virus HIV-1 (Mouland *et al.* 2001), les protéines hnRNP A2 et B1 permettent la localisation de certains ARNm dans des zones définies du cytoplasme. L'ARN du HIV-1 serait ainsi dirigé vers les sites d'assemblage des virions. Les protéines hnRNP A2 liées aux séquences A2RE contacteraient des protéines de type moteur moléculaire se déplaçant le long des fibres d'actine ou de tubuline, ce qui assurerait le déplacement des ARNm, guidé par les fibres d'actine ou de tubuline (pour revue, Carson *et al.* 2001).

II.2.1.2 Protéines hnRNP H/F

Organisation structurale

La famille des protéines hnRNP H/F regroupe les protéines hnRNP H, H', F et 2H9, caractérisées par la présence de domaines RRM non canoniques, appelés quasi-RRM (Fig. II.6 ; Honoré *et al.* 1995). Ces domaines diffèrent des domaines RRM classiques par leurs motifs RNP moins bien conservés (Honoré *et al.* 1995). Les protéines hnRNP H/F se fixent sur des séquences d'ARN riches en G (cf Tab. II.2). La séquence GGA est ainsi reconnue par tous les membres de la famille de protéines, alors que la séquence GGC est spécifiquement reconnue par les protéines hnRNP H/H' (Caputi & Zahler 2001).

Régulation de l'épissage par les protéines hnRNP H/F

A la différence des protéines hnRNP A1 et PTB que nous verrons plus loin, les protéines H/F sont impliquées aussi bien dans l'activation que l'inhibition de l'épissage. Ainsi, les protéines H/F font partie, avec la protéine KSRP, d'un complexe protéique activant l'inclusion de l'exon N1 de *c-src*, via une séquence intronique riche en G située en aval de cet exon (Min *et al.* 1995, 1997 ; Chou *et al.* 1999 ; Markovtsov *et al.* 2000).

¹⁰pour « hnRNP A2 Response Element »

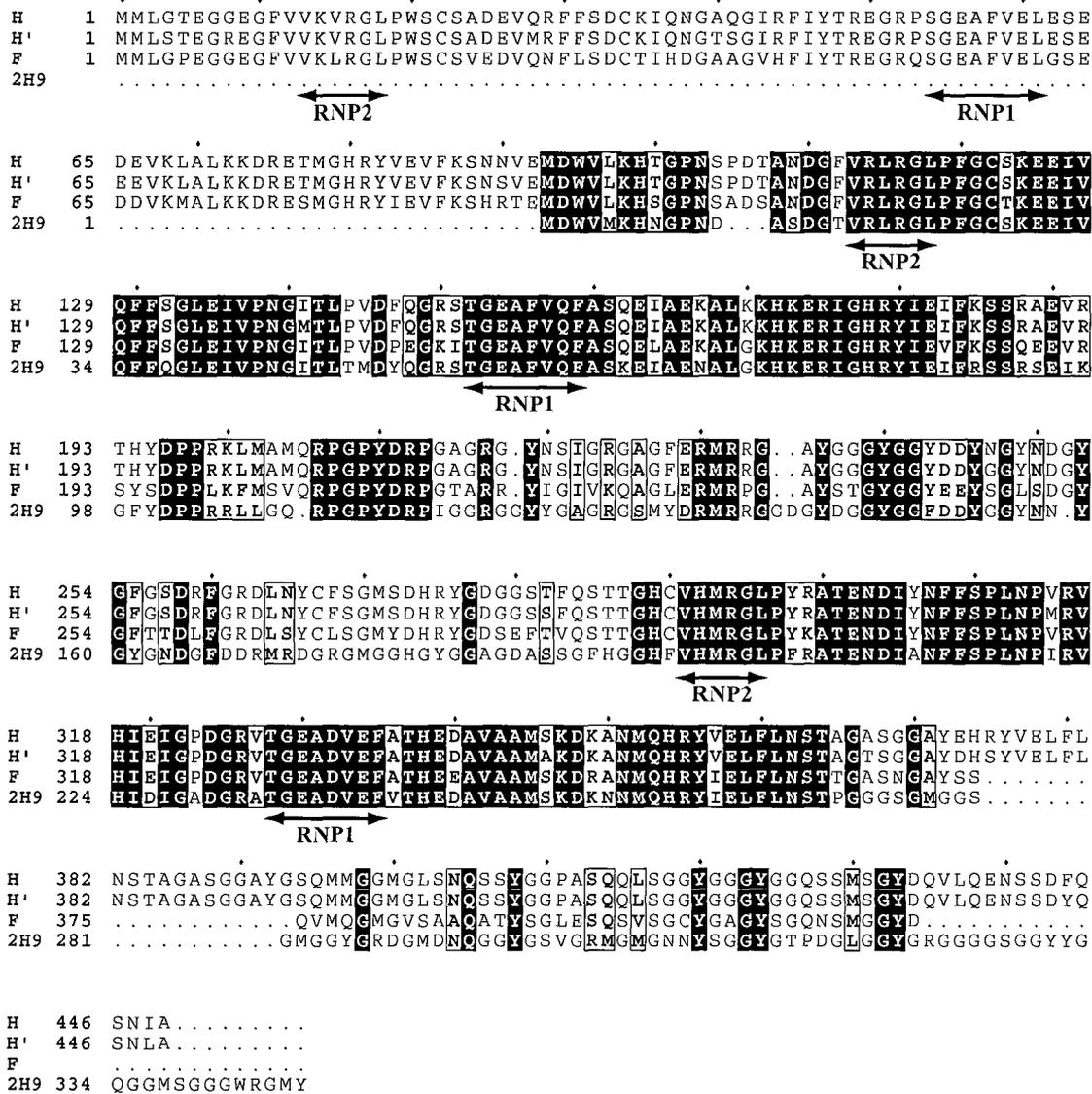


FIG. II.6 – Alignement par CLUSTALW des protéines hnRNP H/F.

Les domaines quasi-RRM sont représentés en bleu. La position des motifs RNP1 et RNP2 est indiquée sous l'alignement. Les résidus présents dans 100% des cas sont indiqués par un fond noir; les résidus présents dans au moins 80% des cas sont encadrés. Les numéros d'accessibilité GenBank sont : P31943 (hnRNP H), P55795 (hnRNP H'), P52957 (hnRNP F) et P31942 (hnRNP 2H9).

Par contre, dans le cas de l'ARN de la β -tropomyosine de rat, la protéine hnRNP H empêche l'inclusion de l'exon 7 dans les cellules non musculaires en se fixant à une séquence ESS¹¹ (Chen *et al.* 1999). De même que dans le cas de l'ARN du virus HIV-1, l'utilisation du site accepteur A3 est inhibé par la fixation de la protéine à une séquence exonique ESS2p (Jacquet *et al.* 2001a). Dans les 2 cas, la protéine pourrait agir en empêchant la fixation de U2AF³⁵ sur la bordure intron-exon (Jacquet *et al.* 2001a ; Chen *et al.* 1999).

¹¹ « exonic splicing silencer »

II.2.1.3 Protéine PTB ou hnRNP I

La protéine hnRNP I ou PTB¹² est impliquée dans de nombreux processus cellulaires comme l'initiation de la traduction coiffe-indépendante, la localisation des ARN dans le cytoplasme, le clivage au site de polyadénylation ou encore, l'épissage alternatif (Moreira *et al.* 1998 ; Côté *et al.* 1999 ; Gosert *et al.* 2000 ; pour revue, Wagner & Garcia-Blanco 2001).

Organisation structurale

L'ARN prémessager de la protéine PTB subit un épissage alternatif conduisant à la production de 4 isoformes. L'isoforme PTB4 est l'isoforme majoritairement produit et c'est celui qui correspond à la protéine hnRNP I. Les 3 autres isoformes PTB1, PTB2 et PTB3 sont également capables de réprimer l'épissage, mais plus faiblement que ne le fait PTB4. Les taux relatifs de ces différentes protéines PTB varient suivant les tissus, ce qui peut donc, suivant le tissu, influencer le profil d'épissage d'un ARN prémessager soumis à l'épissage alternatif (Wollerton *et al.* 2001 ; pour revue, Wagner & Garcia-Blanco 2001).

Du point de vue structural, les isoformes de la protéine PTB portent, en plus d'un signal de localisation nucléaire, 4 domaines RRM non consensus ne possédant pas d'acides aminés hydrophobes aux positions habituellement conservées. Diverses fonctions ont été attribuées à ces différents domaines. Ainsi, le domaine RRM2 est nécessaire pour l'oligomérisation de la protéine (Perez *et al.* 1997 ; Oh *et al.* 1998), le domaine RRM3 permet la fixation à l'ARN et le domaine RRM4 est nécessaire pour la fonction de la protéine, mais pas pour sa liaison à l'ARN (Liu *et al.* 2002 ; pour revue, Wagner & Garcia-Blanco 2001).

Certaines lignées neuronales et les cellules de cerveau sont enrichies en une protéine homologue à la protéine PTB, la protéine nPTB¹³, codée par un gène différent (Markovtsov *et al.* 2000 ; Lillevali *et al.* 2001). On la retrouve également en faible quantité dans un grand nombre de tissus, comme le cœur et le muscle squelettique, mais elle n'est pas détectable dans les cellules HeLa (Markovtsov *et al.* 2000). La protéine nPTB a la même organisation structurale que la protéine PTB, avec 4 domaines RRM et un signal de localisation nucléaire, mais elle réprime plus faiblement l'épissage que la protéine PTB (Markovtsov *et al.* 2000). Les taux relatifs des 2 protéines module donc le profil d'épissage des ARN prémessagers produits dans le cerveau, c'est le cas notamment pour l'ARN *c-src*, que nous présenterons plus loin (§II.7.1 ; Markovtsov *et al.* 2000). Récemment, un isoforme de la protéine nPTB a aussi été mis en évidence (Rahman *et al.* 2002). Il est obtenu par épissage alternatif et n'est exprimé que dans les cellules non-neuronales.

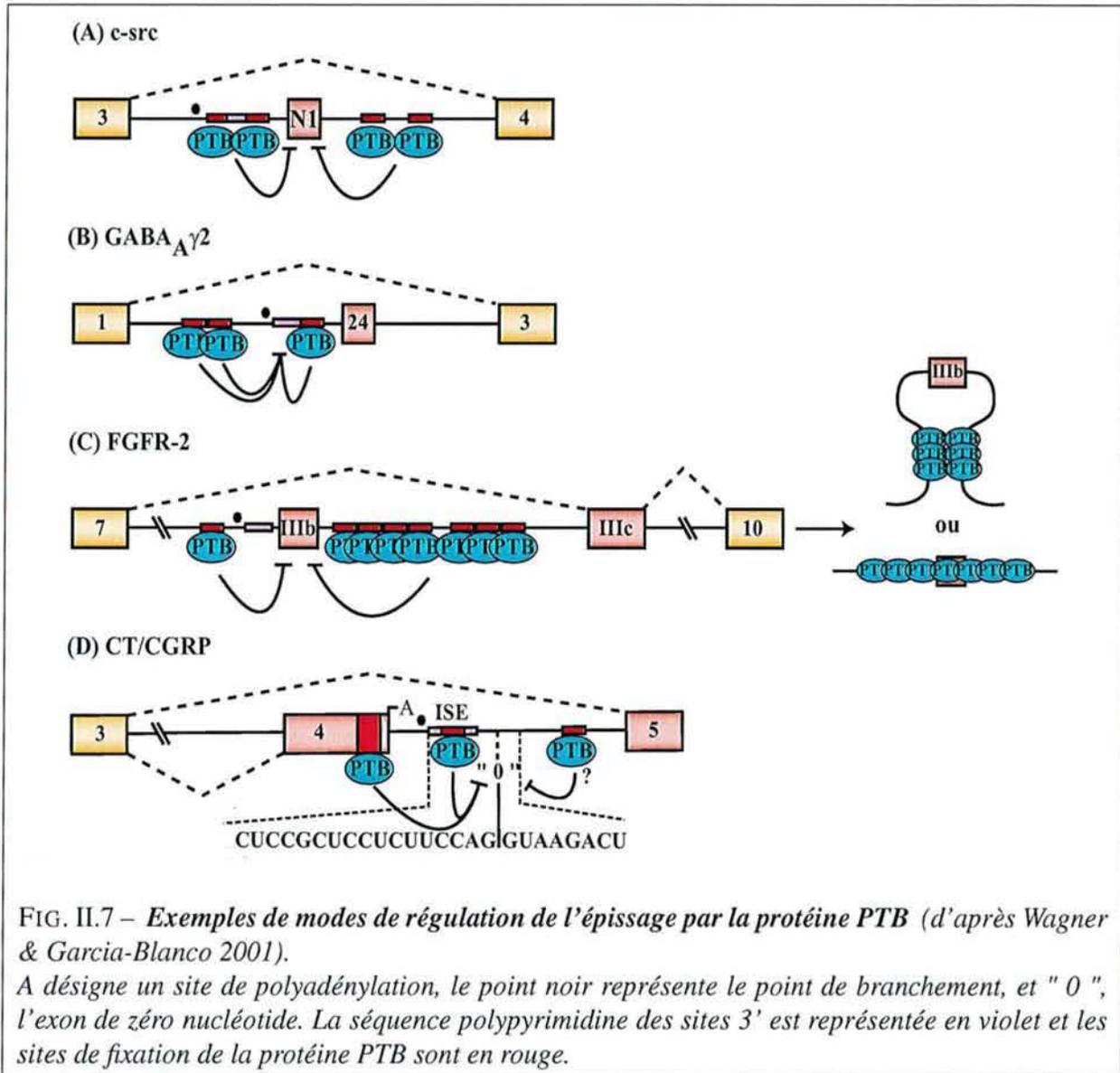
Régulation de l'épissage par la protéine PTB

Les sites de fixation de la protéine PTB ont des positions diverses au sein des ARN prémessagers. Ils peuvent être exoniques ou introniques. Dans le cas de sites exoniques, le mécanisme d'action de la protéine PTB n'est pas connu (Le Guiner *et al.* 2001b).

La protéine PTB a été identifiée initialement comme une protéine se liant aux séquences polypyrimidines des introns riches en résidus U (Garcia-Blanco *et al.* 1989 ; Wang & Pederson 1990). On

¹² « polypyrimidine tract binding protein »

¹³ « neuronal PTB », encore appelée brPTB pour « brain PTB »



la retrouve très souvent associée à la séquence polypyrimidine des sites 3' d'épissage dont elle réprime la reconnaissance, en empêchant l'accès du facteur U2AF et de l'ensemble des composants du spliceosome dépendant de la fixation de ce facteur sur le site 3' (Fig. II.7-A; pour revue, Wagner & Garcia-Blanco 2001). Les motifs ARN reconnus par la protéine PTB, localisés dans les introns flanking l'exon régulé, sont présents dans un contexte riche en pyrimidines et sont de type UCUU(C), UUCUCU ou CUCUCU (Perez *et al.* 1997). Ces séquences introniques peuvent agir à une grande distance de l'exon, par exemple, pour l' α -tropomyosine, elles sont localisées 160 nts en aval de l'exon 3 qu'elles régulent (Gooding *et al.* 1998).

Différents cas de figure observés pour des régulations de l'épissage par la protéine PTB sont représentés schématiquement dans la figure II.7. Dans le cas de l'ARN pré-messager *c-src*, la fixation de la protéine PTB de part et d'autre de l'exon N1 empêche la reconnaissance de cet exon par la machinerie spliceosomale et conduit à son exclusion dans les cellules non neuronales (Fig. II.7-A). Au niveau de l'ARN pré-messager $GABA^A\gamma 2$, trois sites de fixation de la protéine PTB sont situés autour du point

de branchement et de la séquence polypyrimidine. Leur saturation par PTB provoque l'exclusion d'un petit exon de 24 nts dans les cellules non-neuronales (Fig. II.7-B ; Ashiya & Grabowski 1997 ; Zhang *et al.* 1999). En ce sens, la protéine PTB agit de façon similaire à la protéine SXL de *D. melanogaster*, qui empêche la fixation de U2AF⁶⁵ sur l'ARN prémessager *tra* (Valcárcel *et al.* 1993 ; pour revue, Wagner & Garcia-Blanco 2001).

Comme la protéine hnRNP A1, la protéine PTB est capable d'inhiber la définition des exons en multimerisant sur l'ARN prémessager. Les exons régulés sont alors entourés de multiples sites de fixation de la protéine, ce qui les rend inaccessibles au spliceosome en présence de PTB. C'est le cas de l'exon IIIb de l'ARN FGFR-2 (Fig. II.7-C ; pour revue, Wagner & Garcia-Blanco 2001). Deux mécanismes sont possibles : les protéines PTB liées à leurs séquences cibles interagissent entre elles et séquestrent ainsi l'exon ou alors, les protéines PTB s'oligomérisent sur l'exon, ce qui empêche sa reconnaissance. L'élimination de la protéine PTB, par l'approche interférence d'ARN restaure la définition de l'exon IIIb qui est alors inclu (Wagner & Garcia-Blanco 2002).

Là encore, ce mode de régulation est proche de celui de la protéine SXL, lorsqu'elle régule l'épissage de son propre ARNm (Granadino *et al.* 1997 ; pour revue, Wagner & Garcia-Blanco 2001). Elle provoque l'exclusion de l'exon 3 en se fixant à de multiples sites flanquant l'exon. Par contre, il n'est pas connu si elle est capable ou non de s'oligomériser comme la protéine PTB.

La protéine PTB régule également l'épissage de l'ARN prémessager du récepteur FGFR-1, dont l'exon α est inclus dans les cellules gliales (Jin *et al.* 1999b, a). La fixation de la protéine PTB sur des séquences localisées de part et d'autre de l'exon α réprime son inclusion. Dans certains glioblastomes, où une expression anormalement élevée de la protéine PTB est détectée, l'exon α de l'ARNm *FGFR-1* est exclu dans les cellules tumorales (Jin *et al.* 2000). Cela conduit à la production d'une forme du récepteur FGFR-1 plus affine pour le facteur de croissance des fibroblastes, ce qui provoque la prolifération cellulaire (Yamaguchi *et al.* 1994).

Un dernier mode de régulation mis en évidence pour la protéine PTB est lié à sa capacité à réguler également la polyadénylation alternative. L'inclusion et la polyadénylation de l'exon 4 de l'ARN prémessager *CT/CGRP* ont lieu dans la plupart des tissus et dépendent de la reconnaissance d'un élément ISE riche en pyrimidines, renfermant un pseudo-exon de 0 nts, c'est-à-dire qu'un site donneur est accolé à un site accepteur. La fixation du facteur U2AF sur la séquence polypyrimidine favorise la reconnaissance du pseudo-exon et diminue le taux d'inclusion de l'exon 4 et sa polyadénylation. Dans la plupart des tissus, la fixation de la protéine PTB de part et d'autre du pseudo-exon empêche la reconnaissance de cet exon par U2AF et favorise la polyadénylation de l'exon 4 (Fig. II.7-D ; pour revue, Wagner & Garcia-Blanco 2001). La protéine PTB liée à l'élément ISE¹⁴ activerait la polyadénylation soit en interagissant directement avec les facteurs de polyadénylation, soit en stabilisant la fixation de la snRNP U1 sur le site 5' du pseudo-exon (pour revue, Wagner & Garcia-Blanco 2001).

II.2.1.4 Protéine hnRNP C

La protéine hnRNP C est plus un facteur d'activation général de l'épissage qu'un facteur spécifique. Les 2 isoformes de la protéine hnRNP C, les protéines hnRNP C1 et C2, s'associent pour former un tétramère stable. Ce tétramère est capable de se fixer de manière coopérative aux ARN nouvellement

¹⁴ « intronic splicing enhancer »

synthétisés, ceci indépendamment de leur séquence (Gorlach *et al.* 1994 ; McAfee *et al.* 1996). Les 2 protéines se fixent à l'ARN par l'intermédiaire d'un domaine RRM et d'un domaine de type « leucine zipper » (Gorlach *et al.* 1994 ; McAfee *et al.* 1996). La protéine hnRNP C2 diffère de la hnRNP C1 par l'insertion par épissage alternatif de 13 acides aminés entre les 2 domaines de fixation à l'ARN. Les 2 protéines ont été identifiées initialement comme des protéines se fixant à des séquences riches en U, en particulier, aux séquences polypyrimidines des sites 3' d'épissage (Swanson & Dreyfuss 1988) et aux séquences riches en résidus U des UsnRNA (Forne *et al.* 1995 ; Tamsamani & Pederson 1996 ; Shahied-Milam *et al.* 1998). Une séquence riche en U constitue un site de haute affinité pour la protéine hnRNP C1 (Gorlach *et al.* 1994). L'insertion des 13 acides aminés dans la protéine hnRNP C2 conduit à une plus grande sélectivité. Seules les séquences les plus riches en U sont reconnues par cette protéine (Gorlach *et al.* 1994 ; Koloteva-Levine *et al.* 2002).

A ce jour, aucun cas de régulation de l'épissage par les protéines hnRNP C n'est décrit, mais ces protéines sont essentielles pour le processus d'épissage (Choi *et al.* 1986). Elles interviennent dans l'assemblage du spliceosome où elles interagissent avec les UsnRNA. Ces interactions pourraient servir à recruter les UsnRNA au niveau des sites d'épissage (Shahied-Milam *et al.* 1998).

Par ailleurs, il a été proposé que les protéines hnRNP C participent à un mécanisme assurant la rétention des ARN pré-messagers dans le noyau. En effet, elles se fixent préférentiellement aux séquences introniques (Choi & Dreyfuss 1984 ; Piñol Roma & Dreyfuss 1992) et restent confinées dans le noyau. Ainsi, elles peuvent participer à la rétention des ARN pré-messagers dans le nucléoplasme, alors que les protéines hnRNP navettes, qui restent fixées sur les ARN matures, participent à l'export dans le cytoplasme (pour revue, Nakielny & Dreyfuss 1999).

II.2.1.5 Protéine TLS/FUS ou hnRNP P2

La protéine TLS/FUS¹⁵ est impliquée à la fois dans la réparation des dommages de l'ADN et dans l'épissage. Au sein du spliceosome, on la retrouve dans les complexes spliceosomaux H et A (Calvio *et al.* 1995 ; Hartmuth *et al.* 2002). TLS/FUS interagit avec des protéines jouant un rôle central dans l'épissage : le facteur SF1, les protéines SR, en particulier SC35 et SRp75, le co-activateur SRm160, la protéine PTB, la protéine hnRNP A1 et les protéines Sm des particules UsnRNP (Zinszner *et al.* 1994 ; Hackl *et al.* 1994 ; Yang *et al.* 1998 ; Meissner *et al.* 2003). Par ailleurs, TLS/FUS interagit également avec le site 3' lors de la 2^e étape d'épissage (Wu & Green 1997).

La séquence ARN ayant la plus haute affinité pour TLS/FUS est GGUG (Lerga *et al.* 2001), mais la protéine reconnaît également des séquences riches en G, notamment dans l'intron IVSB7 de la β -tropomyosine (Lerga *et al.* 2001).

La surexpression de TLS/FUS dans des cellules HeLa modifie le profil d'épissage du pré-ARNm de l'unité E1A de l'adénovirus, dont elle favorise l'utilisation des sites 5' proximaux 12S et 13S (Hallier *et al.* 1998). Cette régulation peut être modifiée par l'interaction du domaine C-terminal de TLS/FUS avec 2 protéines SR (SC35 et une protéine appelée TASR¹⁶) (Yang *et al.* 1998). Ces 2 protéines ont un effet différent sur l'épissage de l'unité E1A de l'adénovirus : la surexpression de SC35 diminue

¹⁵ « translocated in liposarcoma »

¹⁶ « TLS associated SR protein »

la production des messagers 12S, tandis que la surexpression de TASR conduit à une production accrue des messagers 9, 10 et 11S (Wang & Manley 1995 ; Yang *et al.* 1998). Lorsque TLS/FUS est surexprimée avec l'une ou l'autre de ces protéines SR, elle n'a plus d'effet sur le profil d'épissage du pré-messager E1A, qui ne dépend alors plus que de l'identité de la protéine SR surexprimée (Yang *et al.* 1998).

Il a été observé que dans certaines tumeurs de tissus adipeux humains, une translocation chromosomique a lieu au niveau du gène TLS, conduisant à la production d'une protéine composée de l'extrémité N-terminale de TLS fusionnée à un facteur de transcription. Ainsi, dans un type de leucémie myéloïde, TLS est fusionnée au facteur ERG. Cette protéine de fusion a des capacités de modulation de l'épissage différentes de celles de TLS, ceci du fait qu'elle est incapable d'interagir avec les protéines SR. Ainsi, elle inhibe l'épissage du pré-messager E1A de l'adénovirus et altère le profil d'épissage des ARN messagers CD44 (Yang *et al.* 2000). En interférant avec des facteurs d'épissage importants pour la régulation, ce type de protéine de fusion conduit à de nombreuses anomalies cellulaires.

Les protéines hnRNP que nous venons de décrire jouent un rôle majeur dans l'épissage. Mais d'autres protéines de cette famille ont été montrées comme étant impliquées dans des régulations de l'épissage, comme les protéines hnRNP K, G et M.

La protéine hnRNP K fait ainsi 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 quant à elle l'inclusion de l'exon 7 du pré-messager SMN¹⁷, en interagissant avec hTra2 β 1, une protéine à domaine RS, (Hofmann & Wirth 2002). Des anticorps dirigés contre la protéine hnRNP M inhibent l'épissage *in vitro* (Gattoni *et al.* 1996).

II.2.2 Les protéines SR et leurs sites de fixation

Les membres de la famille des protéines SR sont les principaux activateurs de l'épissage des ARN pré-messagers nucléaires. Ces protéines agissent le plus souvent en se fixant à des séquences exoniques. Quelques cas d'inhibition ont été néanmoins décrits ; ils impliquent la fixation d'une protéine SR sur une séquence intronique.

II.2.2.1 Identification des sites de fixation des protéines SR

Les études qui ont permis de déterminer les séquences fixant les protéines SR avec une haute affinité, ont été menées par emploi de la technique SELEX. Le principe de l'approche consistait à sélectionner, parmi un ensemble d'ARN dont la séquence était dégénérée, ceux fixant avec la plus grande affinité la protéine SR étudiée. Les ARN capables de former des complexes ARN/protéines ont alors été isolés puis soumis à une analyse par RT-PCR et les ADNc correspondants ont été clonés dans un vecteur, sous contrôle d'un promoteur transcriptionnel. Après transcription, les ARN obtenus ont été soumis à nouveau à plusieurs cycles de sélection, avant de séquencer les ADNc correspondants.

La technique SELEX que nous venons de décrire est basée sur un test de fixation. Si les séquences sélectionnées fixent des protéines SR, toutes ne sont pas pour autant capables d'activer l'épissage.

¹⁷ « survival motor neuron »

Protéine SR	Séquence consensus	Test	Auteurs
ASF/SF2	RGAAGAAC AGGACRRAGC SRSASGA GACGAC(R) _n CAGCAG	fixation fixation fonctionnel fonctionnel	Tacke & Manley 1995 Tacke & Manley 1995 Liu <i>et al.</i> 1998 Tian & Kole 2001
SC35	AGSAGAGUA GUUCGAGUA UGUUCSAGWU GWUWCCUGCUA GGGUAUGCUG GAGCAGUAGKS AGGAGAU UGCNGYY GRYYCSYR	fixation fixation fixation fixation fixation fixation fixation fonctionnel fonctionnel	Tacke & Manley 1995 Tacke & Manley 1995 Cavaloc <i>et al.</i> 1999 Cavaloc <i>et al.</i> 1999 Cavaloc <i>et al.</i> 1999 Cavaloc <i>et al.</i> 1999 Cavaloc <i>et al.</i> 1999 Schaal & Maniatis 1999b Liu <i>et al.</i> 2000
9G8	AGACKACGAY ACGAGAGAY GGACGACGA	fixation fixation fonctionnel	Cavaloc <i>et al.</i> 1999 Cavaloc <i>et al.</i> 1999 Schaal & Maniatis 1999b
SRp40	UGGGAGCRGUYRGCUCGY ACDGS	fixation fixation	Tacke <i>et al.</i> 1997 Liu <i>et al.</i> 1998
SRp55	USCGKM	fixation	Liu <i>et al.</i> 1998
SRp20	CUCKUCY	fonctionnel	Schaal & Maniatis 1999b
hTra2	Répétitions GAA	fixation	Tacke <i>et al.</i> 1998

TAB. II.3 – Séquences consensus identifiées comme des sites de liaison pour les protéines SR. Pour chaque séquence, la nature du test (fixation ou fonctionnel) ayant permis son identification est indiquée (R = A ou G ; S = G ou C ; W = A ou U ; Y = C ou U ; M = A ou C ; K = G ou U ; D = A, G ou U ; N = A, C, G ou U).

Pour vérifier leur fonctionnalité en épissage, ces séquences ont ensuite été introduites dans un ARN prémessager contenant un site donneur ou accepteur faible. L'ARN a ensuite été épissé dans un extrait cytoplasmique S100, auquel était ajouté la protéine SR étudiée. Cette méthodologie a été utilisée pour les protéines ASF/SF2 et SC35 (Tab. II.3 ; Tacke & Manley 1995).

D'autres études SELEX ont permis d'identifier directement les séquences fonctionnelles en épissage. La stratégie a consisté à utiliser un ARN prémessager dont l'élément ESE a été remplacé par une séquence dégénérée et les ARN prémessagers correspondants ont été épissés dans un extrait cytoplasmique S100 complété par la protéine SR étudiée. Les ARN messagers produits grâce à la présence de séquences ayant des propriétés activatrices ont ensuite été isolés et les exons contenant ces séquences ont été amplifiés par RT-PCR et sous-clonés dans les ARN prémessagers de départ, de façon à répéter plusieurs fois le cycle de sélection. Cette stratégie a permis d'identifier des séquences activatrices ou ESE répondant aux protéines ASF/SF2, SC35, 9G8, SRp40 et SRp55 (cf Tab. II.3 ; Tacke *et al.* 1997 ; Liu *et al.* 1998, 2000 ; Cavaloc *et al.* 1999).

Ces techniques SELEX ont permis de montrer que les protéines SR sont capables de reconnaître un grand nombre de séquences, en particulier, les protéines ASF/SF2 et SC35. Le cas de la protéine SC35

Séquence essentielle de l'ESE	ARN pré-messager	Position dans l'exon	Protéines fixées	Auteurs
5' UGCUGUU 3'	β -globine	67 nts en aval de l'exon 2	SC35	Schaal & Maniatis 1999a
5' GAGGAGGAG 3'	β -tropomyosine de rat	17 et 68 nts en aval du site 3' de l'exon 6	ASF/SF2	Selvakumar & Helfman 1999
5' GAAGAAGAG 3'		25 nts en aval du site 3' de l'exon 8	ASF/SF2	
5' GGAAG 3'	hormone de croissance bovine	73 à 188 nts en aval du site 3' de l'exon 5	ASF/SF2, hnRNP A1	Sun <i>et al.</i> 1993
5' UCUUCAAUCAACA 3' (élément dsxRE)	<i>dsx</i> , drosophile	6 répétitions des nts 294 à 577 en aval du site 3' de l'exon 4	ASF/SF2, 9G8, Tra et Tra-2	Tian & Maniatis 1994 ; Lynch & Maniatis 1996
5' AAAGGACAAAGGACAAAA 3' (élément PRE)		entre les répétitions 5 et 6 de l'élément dsxRE	ASF/SF2, Tra, Tra-2	
5' GAGGAAGAA 3'	cTNT, poulet	3 à 11 nts en aval du site 3' de l'exon 5	SRp40, SRp55, SRp75, ASF/SF2	Ramchatesingh <i>et al.</i> 1995 ; Stark <i>et al.</i> 1998
5' GAAGAAGA 3'	fibronectine humaine	99 nts en aval du site 3' de l'exon IIIa	ASF/SF2, 9G8	Cramer <i>et al.</i> 1999 ; Lavigueur <i>et al.</i> 1993
5' GGAAGGACAGCA 3'	IgM de souris	20 nts en aval du site 3' de l'exon M2	ASF/SF2	Watakabe <i>et al.</i> 1993 ; Liu <i>et al.</i> 1998
5' UGCUGUG 3'	IgM de souris fibronectine humaine	20 nts et 31 nts en aval du site 3' de l'exon C4	SC35	Chandler <i>et al.</i> 1997 Kuo & Norton 1999 Kuo & Norton 1999
5' AGGAGAAGGGA 3' (PRE1)		150 nts en aval du site 3' de l'exon B	SRp40	
5' GGAGAGAG 3' (PRE2)		30 nts en amont du site 5' de l'exon B	SRp20	
5' UCUUCAAUCAACA 3'	fruitless, drosophile	3 répétitions des nts 38 à 238 en aval du site 5' de l'exon 5	Tra, Tra2	Heinrichs <i>et al.</i> 1998
5' GACGACGAG 3' (Pu1)	unité E1A de l'adénovirus	20 nts en amont du site 12S	9G8	Bourgeois <i>et al.</i> 1999
5' GAUGAAGAG 3' (Pu2)		103 nts en aval du site 3'	SC35, ASF/SF2	

TAB. II.4 – Exemples de séquences ESE fixant les protéines SR.

est l'exemple le plus marquant, puisque ses séquences cibles peuvent être aussi bien riches en purines qu'en pyrimidines.

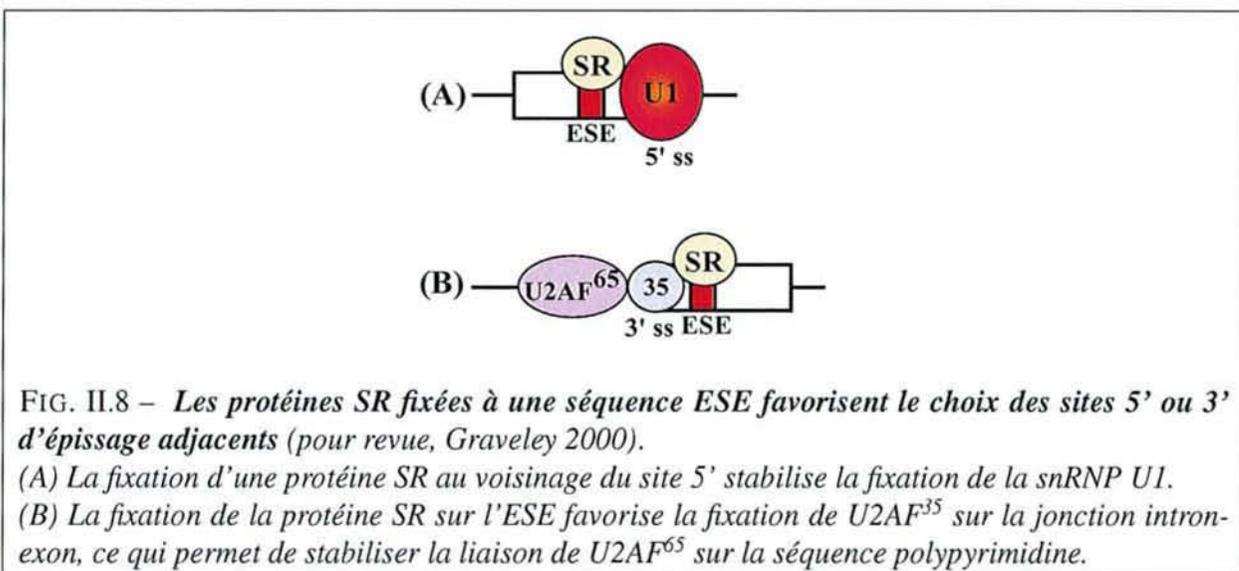
Le spectre de séquences reconnues par les protéines 9G8, SRp40 ou SRp20 est en revanche beaucoup plus restreint. Ainsi, la protéine 9G8 ne reconnaît que des séquences de type GAC. La situation de SRp40 est particulière puisque cette protéine est capable de reconnaître une longue séquence d'ARN (bien qu'il n'ait pas été montré si toute la séquence est requise pour l'activation) et une courte séquence dégénérée (Tab. II.3).

Une grande partie des ESE fixant les protéines SR, identifiés dans les ARN pré-messagers, contiennent les séquences SELEX reconnues par ces protéines. Quelques exemples en sont donnés dans le tableau II.4. Il arrive également que des protéines SR se fixent à des séquences ne correspondant pas aux séquences consensus établies. C'est le cas, par exemple, pour la protéine 9G8, lors de la régulation de l'épissage de son propre ARN pré-messager. Elle se fixe à des séquences riches en purines, alors qu'une fixation d'ASF/SF2 serait attendue sur ces séquences (Lejeune *et al.* 2001).

II.2.2.2 Activation de l'épissage par les protéines SR fixées à des séquences exoniques

A ce jour, les séquences activatrices d'épissage identifiées comme fixant des protéines SR activent la 1^{re} étape de la réaction d'épissage, à l'exception d'un cas que nous décrivons à la fin de ce paragraphe.

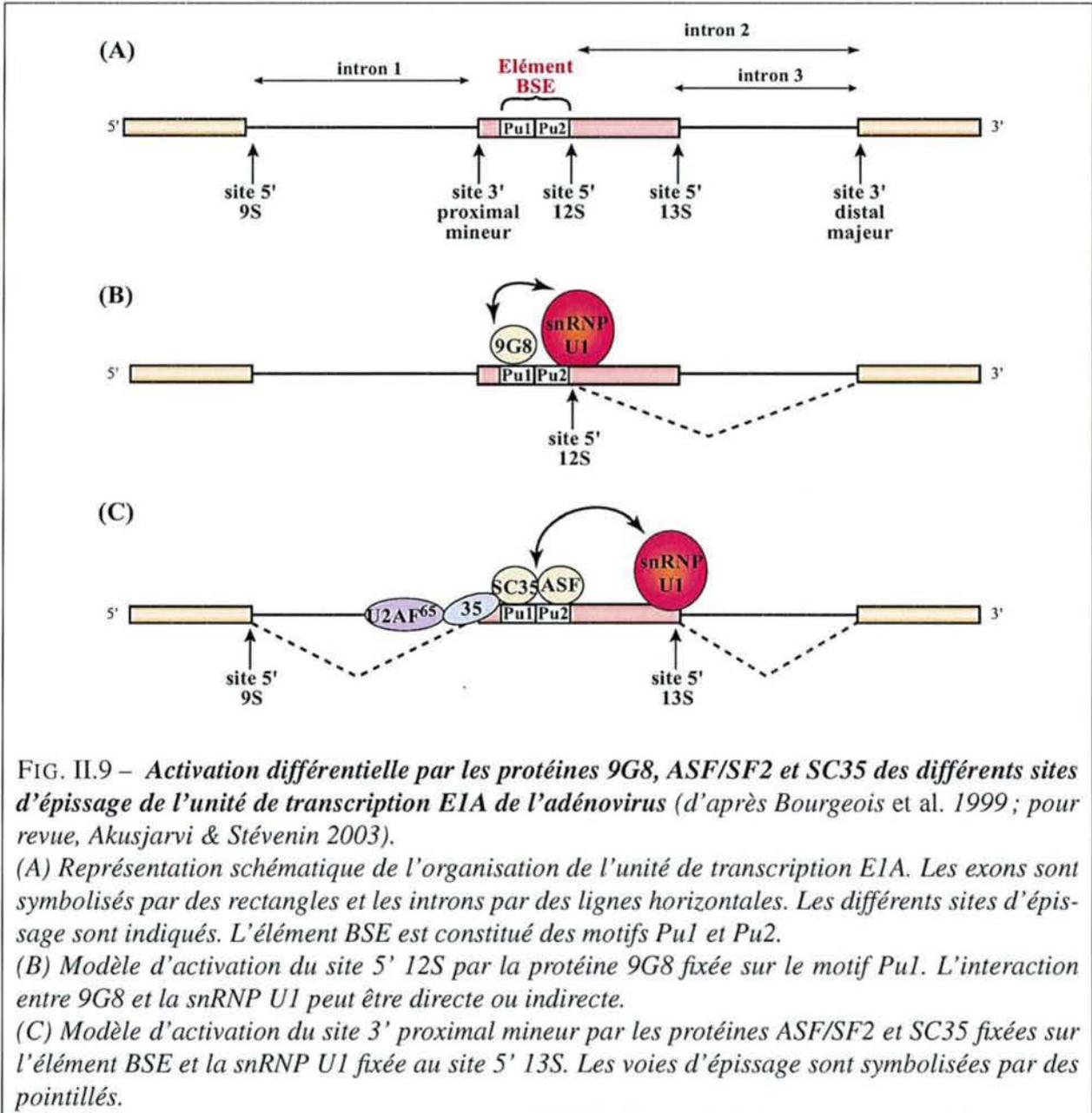
Les protéines SR jouent un rôle majeur dans la sélection des sites d'épissage



Rôle des protéines SR dans la fixation des composants spliceosomaux sur l'ARN pré-messager : Les protéines SR fixées à un élément ESE stabilisent la fixation de la snRNP U1 (Fig. II.8-A) ou du facteur U2AF³⁵ (Fig. II.8-B) sur l'ARN pré-messager par l'établissement d'interactions protéine/protéine (pour revue, Graveley 2000). Ces interactions impliquent le domaine RS des protéines SR et celui de la protéine U1-70K de la snRNP U1 ou du facteur U2AF³⁵. Lorsque U2AF³⁵ est fixé de façon stable sur le site 3' d'épissage, il peut, à son tour, stabiliser la fixation de U2AF⁶⁵ sur la

séquence polypyrimidine. Les protéines SR favorisent donc la définition des exons renfermant un élément ESE.

La localisation des éléments ESE par rapport au site 3' est importante. Ainsi, une séquence ESE forte fonctionne à une plus grande distance de l'intron qu'une séquence ESE faible (Graveley *et al.* 1998). Par ailleurs, la présence de multiples séquences ESE dans l'exon augmente la probabilité d'activer l'épissage (Graveley *et al.* 1998 ; Hertel & Maniatis 1998).



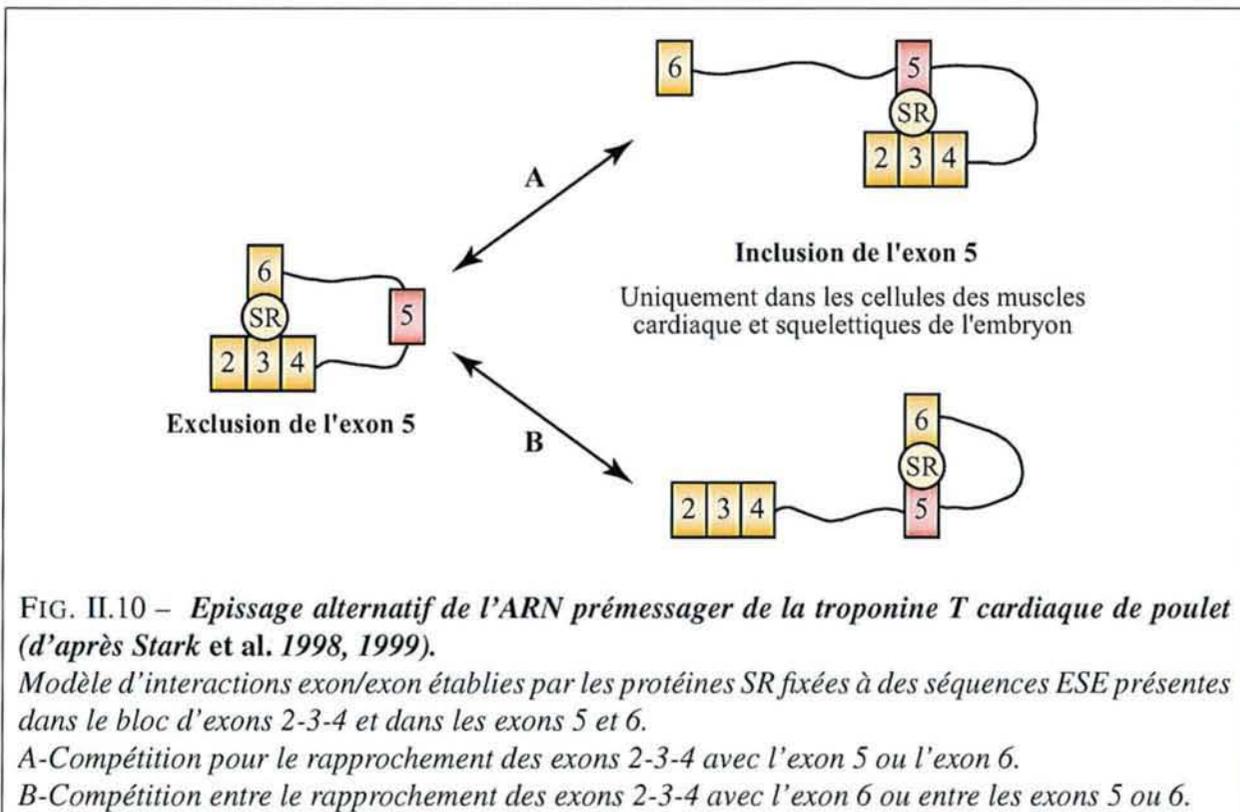
Cas de l'élément bidirectionnel BSE : Certains éléments ESE sont capables de fixer des protéines SR différentes et de favoriser ainsi le choix d'un site par rapport à l'autre. Cette situation a été observée par l'équipe de J. Stévenin pour l'ARN pré-messager de l'unité E1A de l'adénovirus (Fig. II.9 ; Bourgeois *et al.* 1999 ; pour revue, Akusjarvi & Stévenin 2003) : suivant l'identité de la protéine SR

fixée, l'élément BSE¹⁸ active le site 3' amont ou le site 5' aval.

L'élément BSE est localisé immédiatement en amont du site donneur 12S. Il est constitué de 2 séquences riches en purines, Pu1 et Pu2 (Fig. II.9-A). L'élément Pu1 active le site 5' 12S en fixant la protéine 9G8, ce qui permet de stabiliser la fixation de la snRNP U1 sur le site 5' 12S et conduit à l'élimination du 2^e intron (Fig. II.9-B).

En revanche, l'élément Pu1 associé à Pu2 active l'élimination des introns 1 et 3 en fixant les protéines ASF/SF2 et SC35. Celles-ci participeraient à la définition de l'exon délimité par le site 3' proximal et le site 5' 13S, ceci en favorisant d'une part la fixation du facteur U2AF au site 3' proximal et d'autre part, la stabilisation de la snRNP U1 au site 5' 13S (Fig. II.9-C). Les 2 voies d'activation par l'élément BSE sont mutuellement exclusives.

Cas de l'ARN pré-messager de la troponine T cardiaque de poulet : La stabilité des interactions entre les protéines SR fixées sur les différents éléments d'un même ARN pré-messager pourrait définir les choix d'épissage alternatifs. Ce type de régulation a été décrit pour l'ARN pré-messager de la troponine T cardiaque de poulet (cTNT). Les protéines SR se fixent à différents éléments ESE de cet ARN et plutôt que de recruter des facteurs spliceosomaux au niveau des sites d'épissage, elles vont rapprocher les exons entre eux, afin de favoriser leur ligation.



L'exon 5 de l'ARN cTNT est épissé alternativement : son inclusion dans les muscles cardiaque et squelettiques dépend de la présence d'une séquence ESE riche en purines (Ramchatesingh *et al.* 1995 ; Stark *et al.* 1998). La ligation des exons 2, 3 et 4 génèrerait par ailleurs une séquence riche en purines

¹⁸« bidirectional splicing enhancer »

à laquelle pourraient se fixer les protéines SR. Ces protéines interagiraient alors avec les protéines SR liées, soit à l'ESE de l'exon 5, soit à celui de l'exon 6 (Fig. II.10-A ; Stark *et al.* 1998). Cette dernière interaction étant la plus stable, elle conduirait à l'exclusion de l'exon 5 dans la plupart des tissus (Stark *et al.* 1999).

A cette compétition pour le rapprochement des exons 2-3-4 avec l'exon 5 ou l'exon 6, s'ajoute une autre compétition, celle permettant le rapprochement des exons 5 et 6 (Fig. II.10-B). Ce rapprochement, qui a pu être montré directement par microscopie électronique, dépend de la force de l'élément ESE et conduit à l'élimination de l'intron séparant les exons 5 et 6 (Stark *et al.* 1998). On ne connaît pas encore les raisons d'une plus forte stabilisation de l'interaction entre les exons 5 et 6 dans les cellules musculaires.

Ce type de régulation constitue un exemple où l'inclusion d'un exon dépend des événements d'épissage précédents, c'est-à-dire, l'élimination des introns séparant les exons 2, 3 et 4. D'autres cas ont été décrits, comme celui du pré-ARNm de la β -tropomyosine, dont la jonction des exons 6 et 8 crée une séquence activatrice d'épissage, nécessaire à l'épissage de l'intron entre les exons 5 et 6 (Helfman *et al.* 1988 ; Tsukahara *et al.* 1994).

Le rôle des protéines SR peut être contrecarré par l'action des protéines hnRNP A/B

Les 1^{res} études menées sur l'épissage alternatif ont rapidement permis d'identifier les protéines SR et hnRNP A/B comme des facteurs antagonistes (pour revue, Smith & Valcárcel 2000). Leurs sites de fixation sur l'ARN sont en général juxtaposés, ce qui empêche, par encombrement stérique, la fixation simultanée des 2 protéines (Fig. II.11). C'est le cas de l'épissage de l'intron D de l'ARN pré-messager de l'hormone de croissance de bœuf, qui fait intervenir des séquences chevauchantes reconnues par ASF/SF2 et hnRNP A1, ce qui expliquerait l'effet opposé des 2 protéines sur l'épissage de l'intron D de cet ARN (Dirksen *et al.* 2000).

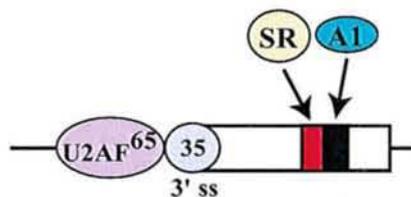
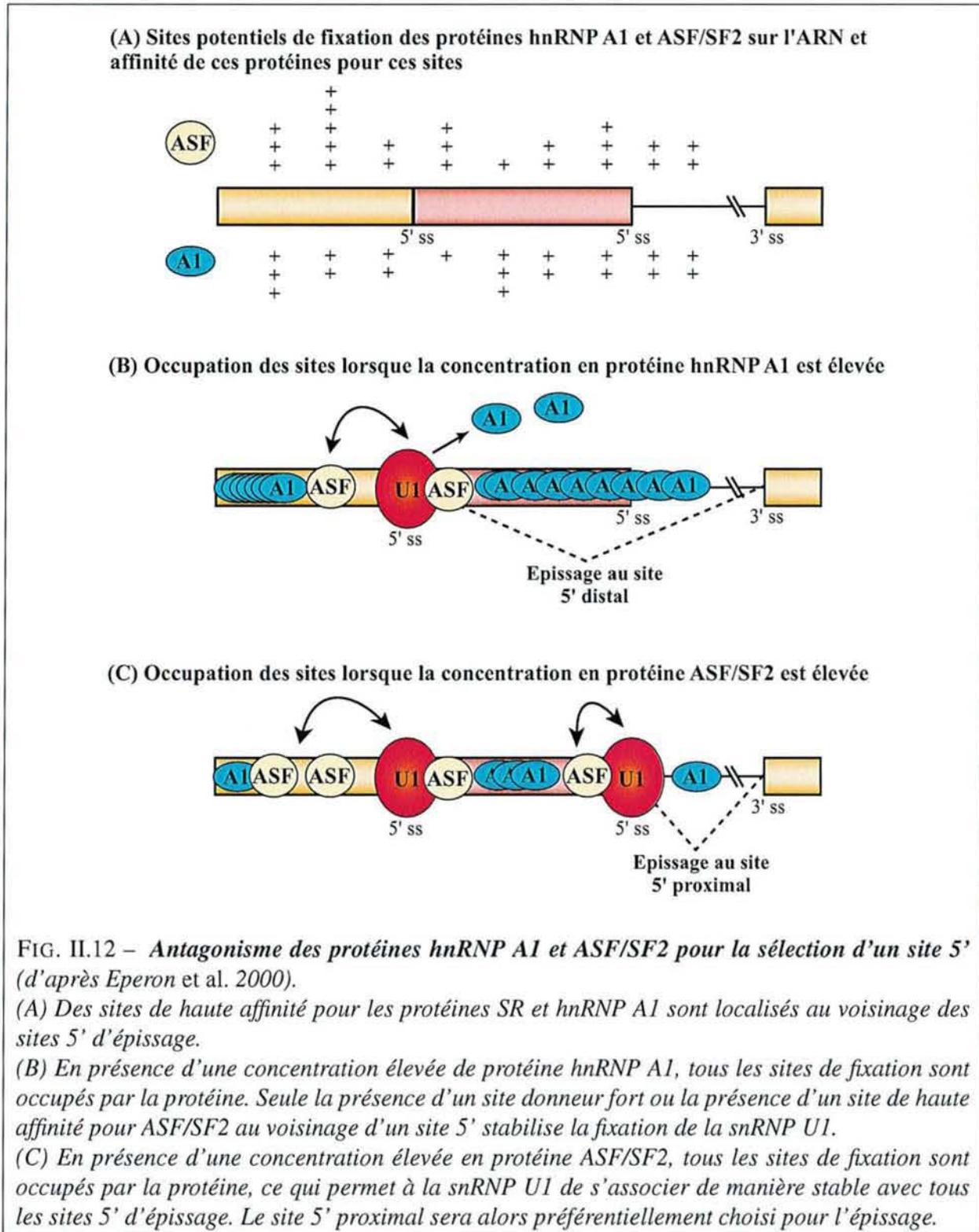


FIG. II.11 – *Les sites de fixation des protéines SR et hnRNP A/B sur l'ARN pré-messager de l'hormone de croissance de bœuf sont chevauchants.* (pour revue, Graveley 2000).

Les protéines SR et hnRNP A/B sont en compétition pour leur fixation à l'ARN. L'élément ESE est représenté en rouge et le site de fixation de la protéine hnRNP A1, en noir.

La fixation d'une protéine SR à un élément ESE peut aussi empêcher la multimérisation de la protéine hnRNP A1 qui bloque normalement le site d'épissage. C'est le cas de la régulation de l'épissage au site accepteur A7 du virus HIV-1 (Tange & Kjems 2001 ; Marchand *et al.* 2002).

Dans le cas de sites 5' d'épissage alternatifs, ce mécanisme a été bien étudié par Eperon *et al.* (2000). Les auteurs ont utilisé un ARN pré-messager modèle dérivé de l'adénovirus, contenant 2 sites donneurs



dupliqués et 1 site accepteur (AdML, Eperon *et al.* 2000). Ils ont alors étudié l'effet des protéines régulatrices ASF/SF2 et hnRNP A1 sur l'association de la snRNP U1 avec les sites 5', ce qui leur a permis de proposer un modèle résumé dans la figure II.12 (Eperon *et al.* 2000). Ce modèle repose sur l'existence de sites de haute affinité pour les protéines SR et hnRNP A1 au voisinage des sites 5' d'épissage (Fig. II.12-A).

Lorsque la concentration en protéine hnRNP A1 est élevée, la protéine multimérise sur l'ARN. Dans

ce cas, la particule snRNP U1 ne peut s'associer qu'à un site 5' fort ou à un site 5' possédant dans son voisinage, un site de haute affinité pour la protéine ASF/SF2 (Fig. II.12-B). Cette dernière est alors capable de déplacer efficacement la protéine hnRNP A1. Dans notre exemple, cela conduit à la fixation stable de la snRNP U1 sur le site distal et donc à son utilisation lors de l'épissage.

Lorsque la concentration en protéine ASF/SF2 est élevée (Fig. II.12-C), la protéine SR stabilise la fixation de la snRNP U1 sur tous les sites 5' d'épissage et le site donneur proximal est préférentiellement choisi.

Le rapport entre les concentrations intra-cellulaires des protéines SR et hnRNP A1 va donc moduler le choix des sites (Fu *et al.* 1992 ; Mayeda & Krainer 1992 ; Eperon *et al.* 2000). De manière générale, ASF/SF2 est présente en excès par rapport à la protéine hnRNP A1, à l'exception des cellules HeLa et des cellules de la rate (Hanamura *et al.* 1998).

L'épissage de l'intron situé entre les exons M1 et M2 de l'immunoglobuline de souris constitue un cas particulier d'activation par les protéines SR. En effet, les protéines SR fixées à un élément ESE favorisent normalement la fixation du facteur U2AF³⁵ sur la jonction intron-exon. La situation est inverse pour l'ARN prémessager IgM : le facteur U2AF³⁵, fixé au site 3', agirait en aval pour stabiliser la fixation de la protéine SR ASF/SF2 sur l'ESE, qui pourrait, à son tour, contrecarrer la fixation d'un facteur antagoniste sur la séquence ESS voisine (Zhu & Krainer 2000).

Dans ce cas particulier, la séquence ESE possède d'autres propriétés : elle permet l'activation des 2 étapes d'épissage, et la protéine ASF/SF2 est indispensable à l'étape de ligation des 2 exons (Chew *et al.* 1999). Cette observation, et le fait que les protéines SR favorisent l'incorporation de la tri-snRNP [U4/U6,U5] et le recrutement de la snRNP U6 au site 5' constituent les seules preuves de l'implication des protéines SR à des étapes ultérieures à la définition des exons, lors de l'épissage des ARN prémessagers.

II.2.2.3 Activation de l'épissage par les protéines SR fixées à des séquences introniques

Les exemples d'activation de ce type sont des cas isolés. L'exemple le plus connu est celui de la séquence S4, localisée en aval de l'exon 6A de la β -tropomyosine de poulet. Cette séquence fixe la protéine ASF/SF2 et active ainsi l'inclusion de cet exon. Au contraire, la fixation de la protéine SC35 sur la séquence S4 n'active pas l'épissage et conduit à l'exclusion de l'exon 6A au profit de l'inclusion de l'exon voisin, l'exon 6B (Gallego *et al.* 1997).

ASF/SF2 agirait au sein d'un complexe protéique, renfermant également les protéines hnRNP K, L et PTB (Expert-Bezancon *et al.* 2002). La présence de la protéine PTB dans ce complexe activateur d'épissage est beaucoup plus surprenante, puisque la protéine est en général un facteur répresseur. Il est probable, qu'au sein du complexe, cette protéine favorise l'inclusion de l'exon 6A en réprimant celle de l'exon 6B (Expert-Bezancon *et al.* 2002).

II.2.2.4 Inhibition de l'épissage par les protéines SR

La position dans l'intron du site de fixation des protéines SR semble être l'élément déterminant de ce type d'inhibition. On peut citer l'exemple de l'élément intronique 3RE, riche en purines, qui réprime l'épissage du site accepteur IIIa du prémessager E1A de l'adénovirus en phase précoce d'infection,

par fixation de la protéine ASF/SF2 (Kanopka *et al.* 1996 ; pour revue, Akusjarvi & Stévenin 2003). Les auteurs avaient avancé l'hypothèse selon laquelle la fixation d'ASF/SF2 à cette séquence empêcherait la fixation de la snRNP U2 sur le point de branchement.

Cette explication a été récemment remise en cause (Dauksaite & Akusjärvi 2002). Il n'y aurait pas de gêne stérique mais le domaine RRM2 d'ASF/SF2 fonctionnerait comme un répresseur d'épissage par l'intermédiaire du motif SWQDLKD qui pourrait interagir avec d'autres facteurs d'épissage impliqués dans la répression. Cette propriété d'inhibition de l'épissage du domaine RRM2 d'ASF/SF2 dépend de la localisation du site de fixation de la protéine : lorsque ce site est localisée dans l'intron de l'ARN pré-messager E1A ou dans celui de la β -globine de lapin, le domaine a une activité répressive tandis qu'il ne fait rien lorsque le site de fixation est exonique (Dauksaite & Akusjärvi 2002).

La séquence 3RE de l'adénovirus joue par ailleurs un double rôle dans la régulation de l'épissage de l'unité L1 de l'adénovirus, car en plus de réprimer le site accepteur IIIa en phase précoce d'infection, elle active en même temps l'épissage du site accepteur 55.52K (Kanopka *et al.* 1996, 1998).

Un autre cas connu d'inhibition de l'épissage par une protéine SR concerne 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 & Chabot 2000, 2002).

II.2.3 Protéines SRm et Tra à domaine RS

II.2.3.1 Les coactivateurs d'épissage SRm160/300

Les 2 protéines SRm¹⁹ 160 et 300 sont associées aux complexes d'épissage. Elles contiennent de multiples répétitions RS, mais à la différence des protéines SR, elles ne possèdent pas de domaine de fixation à l'ARN (Fig. I.7 ; Blencowe *et al.* 1998, 2000).

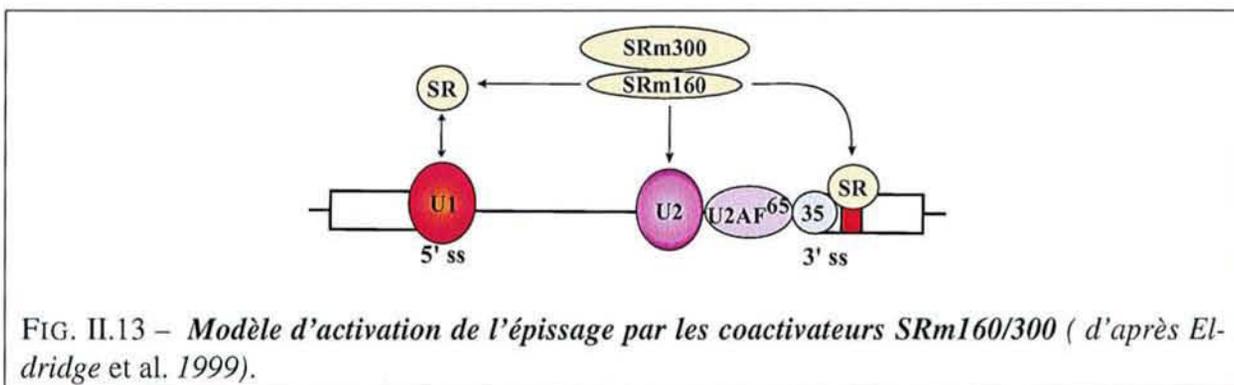


FIG. II.13 – *Modèle d'activation de l'épissage par les coactivateurs SRm160/300 (d'après Eldridge et al. 1999).*

Elles forment un complexe, SRm160/300, activant l'épissage constitutif de certains ARN, mais aussi d'un ARN dérivé du pré-messager *dsx*, contenant une séquence activatrice ESE constituée de 6 répétitions GAA (Eldridge *et al.* 1999). Par ailleurs, alors que ce complexe n'est pas capable de compléter un extrait cytoplasmique S100 pour permettre l'épissage de l'ARN pré-messager de la β -globine, son ajout aux protéines SR dans l'extrait S100 stimule cet épissage (Blencowe *et al.* 1998).

Le complexe SRm160/300 établit des interactions avec des composants du spliceosome et ferait un

¹⁹« SR-related matrix protein »

lien entre les particules snRNP et les protéines SR liées à un ESE (Fig. II.13). En effet, l'association du complexe avec l'ARN prémessager nécessite la fixation préalable de la snRNP U1 au site 5', avec laquelle il pourrait interagir de manière indirecte par l'intermédiaire des protéines SR, et cette association est ensuite stabilisée par l'arrivée de la snRNP U2 (Blencowe *et al.* 1998).

II.2.3.2 Protéines Tra

Chez la drosophile, Tra²⁰ est une protéine régulatrice importante impliquée dans la détermination du sexe. Elle est constituée d'un domaine RRM et de 2 domaines RS, et elle est exprimée aussi bien chez les mâles que les femelles. Elle régule l'épissage alternatif d'ARN prémessagers comme ceux des gènes *fruitless* et *doublesex* (*dsx*).

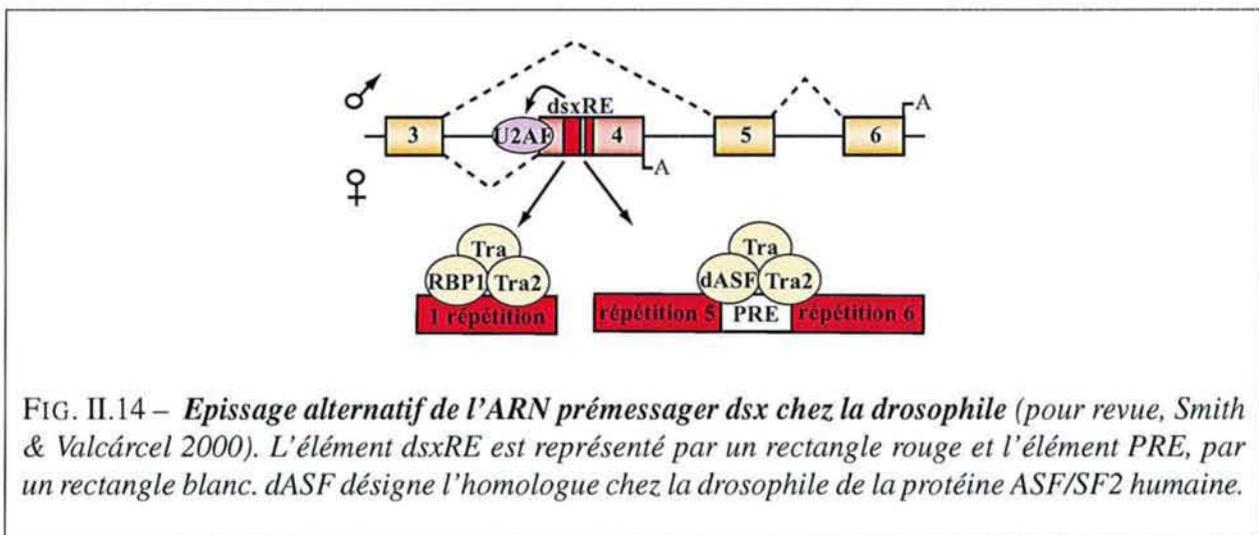


FIG. II.14 – *Epissage alternatif de l'ARN prémessager dsx chez la drosophile* (pour revue, Smith & Valcárcel 2000). L'élément dsxRE est représenté par un rectangle rouge et l'élément PRE, par un rectangle blanc. dASF désigne l'homologue chez la drosophile de la protéine ASF/SF2 humaine.

L'exon 4 de l'ARN *dsx* possède une séquence dsxRE²¹ constituée de 6 répétitions de la séquence UCUUCAUCAACA et d'une séquence riche en purines, PRE²², localisée entre les répétitions 5 et 6 (Fig. II.14; Nagoshi & Baker 1990; Lynch & Maniatis 1995). Chaque répétition de l'élément dsxRE fixe un complexe formé par l'association coopérative des protéines Tra - une protéine exprimée uniquement chez les femelles, Tra2 et d'une protéine SR, RBP1²³, ce qui favoriserait le recrutement du facteur U2AF au site 3' (Fig. II.14; pour revues, Graveley 2000; Smith & Valcárcel 2000) et permet l'inclusion de l'exon 4. La séquence PRE fixe également un complexe protéique activant l'inclusion de l'exon 4, constitué des protéines Tra, Tra2 et dASF/SF2, la protéine homologue à la protéine humaine ASF/SF2 (Fig. II.14; pour revues, Graveley 2000; Smith & Valcárcel 2000). La protéine obtenue, DSX^F, est un facteur de transcription réprimant les gènes de la différenciation mâle (Burtis & Baker 1989).

Chez les mâles, en l'absence de protéine Tra, les interactions entre Tra2, RBP1 et l'ARN prémessager ne sont pas stables, ce qui conduit à l'exclusion de l'exon 4. Le facteur de transcription DSX^M ainsi obtenu est impliqué dans la différenciation cellulaire des mâles (Burtis & Baker 1989).

²⁰ « transformer-2 »

²¹ « dsx repeat element »

²² « purine rich element »

²³ « RNA binding protein 1 »

Le complexe Tra/Tra2/RBP1 régule également l'épissage du pré-messager *fruitless*. Ce sont, dans ce cas, 13 répétitions de la séquence du dsxRE, localisées en amont du site 5' de l'exon 2, qui recrutent le complexe protéique. Le complexe va stabiliser la fixation de la snRNP U1 au site donneur en interagissant avec la protéine U1-70K (pour revues, Graveley 2000 ; Smith & Valcárcel 2000).

Chez les mammifères, 2 protéines homologues à Tra2 ont été identifiées. On retrouve ainsi les protéines hTra2 α et hTra2 β dans des extraits nucléaires de cellules HeLa et des expériences d'épissage en extrait S100 montrent que ces 2 protéines ne peuvent pas remplacer fonctionnellement les protéines SR. D'après une étude SELEX, les 2 protéines hTra2 α et hTra2 β reconnaissent la même séquence, GAA (Tacke *et al.* 1998).

Comme pour les protéines Tra et Tra2 chez la drosophile, les protéines hTra2 α et hTra2 β semblent capables de réguler l'épissage en interagissant avec d'autres protéines SR. C'est le cas de l'exon 7 de l'ARN pré-messager *SMN2*, dont l'inclusion est activée par une séquence exonique riche en purines, à laquelle hTra2 β se fixe et recrute la protéine SRp30c (Young *et al.* 2002).

II.3 Autres éléments de séquences et facteurs associés intervenant dans la régulation de l'épissage

II.3.1 Site donneur aval

Un site donneur localisé en aval du site accepteur utilisé en épissage constitue une séquence activatrice. La fixation de la snRNP U1 sur ce site favorise la définition de l'exon, ce qui augmente l'efficacité d'épissage du site 3' (Robberson *et al.* 1990 ; Kuo *et al.* 1991 ; Berget 1995).

La présence d'un site 5' aval a même un effet activateur plus fort que l'insertion d'un élément ESE fixant des protéines SR. Cet effet dépend néanmoins de la taille de l'exon, puisqu'il devient très faible lorsque le site 5' est placé à plus de 300 nts en aval de l'extrémité 5' de l'exon. L'activation par un élément ESE devient alors nécessaire (Berget 1995 ; Yue & Akusjärvi 1999).

II.3.2 Séquences riches en A et C

Ces séquences riches en A et C constituent une nouvelle classe de séquences activatrices d'épissage, appelées ACE²⁴. Elles activeraient l'épissage en fixant la protéine YB-1²⁵ (Stickeler *et al.* 2001). Cette protéine régule la transcription de certains gènes et elle fait partie de complexes ribonucléo-protéiques se liant aux ARNm qui vont être traduits (pour revue, Wolffe & Meric 1996). Elle est également retrouvée dans le complexe spliceosomal A (Hartmuth *et al.* 2002) et elle est capable de réguler l'épissage de l'ARN E1A de l'adénovirus (Chansky *et al.* 2001). Cette protéine interagit avec la protéine SRp30c (Raffetseder *et al.* 2003) et avec la protéine hnRNP K (Shnyreva *et al.* 2000).

²⁴ « A/C rich sequence enhancer »

²⁵ « Y-Box »

Plusieurs gènes humains, comme *CFTR*²⁶, *CSR*²⁷, *eNOS*²⁸ ou encore, le gène de l'interféron γ possèdent, au niveau de leurs introns, des répétitions polymorphiques CA qui font également partie de la famille des séquences ACE et constituent aussi des séquences activatrices d'épissage. Ainsi, l'efficacité d'épissage de l'intron 13 de l'ARN prémessager *eNOS* dépend du nombre de répétitions CA, qui fixent d'autant plus fortement la protéine hnRNP L, en plus de la protéine YB-1 (Hui *et al.* 2003).

II.3.3 Séquences polypyrimidine localisées entre le site 5' et le point de branchement

Trois types de séquences correspondent à cette classe d'éléments régulateurs.

II.3.3.1 Séquences riches en U localisées en aval d'un site 5' faible

Les séquences de ce type sont systématiquement juxtaposées à un site 5' faible dont elles favorisent l'utilisation. Elles agissent en fixant la protéine TIA-1²⁹. Cette protéine a été au départ considérée comme le précurseur d'une protéine des lymphocytes T cytotoxiques, d'où son nom. En réalité, c'est une protéine nucléaire impliquée dans des phénomènes d'apoptose, majoritairement exprimée dans le cerveau, les testicules et la rate chez l'homme (Beck *et al.* 1996).

La protéine TIA-1, fixée à la séquence riche en U, recrute la snRNP U1 sur le site 5' d'épissage voisin en interagissant directement avec la protéine U1-C de la particule (Del Gatto-Konczak *et al.* 2000 ; Förch *et al.* 2002 ; pour revue, Förch & Valcárcel 2001). Son homologue chez la levure, la protéine Nam8p, est d'ailleurs un composant stable de la snRNP U1 (Tian *et al.* 1991).

La protéine TIA-1 régule ainsi l'épissage alternatif de l'exon 6 de l'ARN prémessager du récepteur Fas, de l'exon K-SAM du récepteur FGFR-2 humain, ainsi que du prémessager *mSl-2*³⁰ de la drosophile (pour revue, Förch & Valcárcel 2001). L'ARN *mSl-2* n'est exprimé que chez les drosophiles mâles. En effet, l'expression de la protéine MSL2 est inhibée chez les femelles par la rétention d'un intron dans la région 5' UTR de l'ARN prémessager. Cette rétention est causée par une protéine présente uniquement chez les femelles, la protéine SXL³¹. Cette protéine se fixe à la séquence riche en U juxtaposée au site 5', ce qui empêche la fixation de la protéine TIA et, par conséquent, la fixation de la snRNP U1 (Förch *et al.* 2001). Par ses propriétés d'« annealing », le facteur U2AF⁶⁵ favoriserait également la fixation de la snRNP U1 sur le site 5' de l'ARN *mSl-2* (Förch *et al.* 2003).

Un facteur humain homologue de TIA-1, la protéine TIAR³², est également capable d'activer l'épissage au niveau de sites 5' faibles en se fixant, comme TIA-1, à une séquence riche en U (Dember *et al.* 1996 ; Le Guiner *et al.* 2001a).

²⁶ « cystic fibrosis transmembrane regulator »

²⁷ « cellular stress response »

²⁸ « endothelial nitric oxyde synthase »

²⁹ « T-cell intracytoplasmic antigen »

³⁰ « male specific lethal 2 »

³¹ « sex lethal »

³² « TIA-1 related »

II.3.3.2 Séquences de type CUCCCC et UUUCCC

Chez *D. melanogaster*, ces séquences polypyrimidines aident à définir l'intron lorsqu'elles sont localisées entre le site 5' et le point de branchement. Les ARN pré-messagers de cet organisme sont en effet caractérisés par la présence d'introns de petite taille et par l'absence de séquence polypyrimidine en aval du point de branchement. Dans l'intron de 59 nts du pré-messager *mle*³³, ce type de séquence est reconnu par l'homologue de la protéine humaine SRp54, qui favorise le recrutement des facteurs spliceosomaux, ce qui conduit à une définition efficace de l'intron (Kennedy *et al.* 1998).

II.3.4 Séquences CUG/UGC

II.3.4.1 Rôle dans l'activation de l'épissage

Ces motifs sont localisés aussi bien dans l'intron amont de l'exon à inclure que dans l'intron aval, et on les retrouve sous forme de séquences répétées de type CUG, CUG/UGC ou encore CUGUCU. Ainsi, l'exon 5 de la protéine cTnT humaine est entouré de séquences CUG et CUGUCU qui permettent son épissage dans les muscles striés embryonnaires (Philips *et al.* 1998). Le motif CUGUCU intervient également dans la régulation de l'épissage de l'ARN pré-messager de la sous-unité $\gamma 2$ du récepteur GABA_A (Zhang *et al.* 1999).

Les éléments de type CUG fixent la protéine CUG-BP³⁴, qui régule aussi bien la traduction que l'épissage chez l'homme (Timchenko *et al.* 1999). Ainsi, dans certains cas de dystrophies musculaires, des expansions de séquences CUG dans la région 3'UTR du gène DMPK, codant une protéine kinase de la myotonine, titrent un isoforme de CUG-BP, la protéine CUG-BP1, ce qui altère l'épissage d'autres ARN pré-messagers et expliquerait l'existence de pathologies associées (Philips *et al.* 1998 ; Singer 1998 ; Charlet *et al.* 2002 ; pour revue, Philips & Cooper 2000).

II.3.4.2 Rôle dans l'inhibition de l'épissage

La fixation d'une protéine autre que CUG-BP peut rendre les séquences CUG/UGC inhibitrices d'épissage. C'est le cas de l'exon 3 de l' α -tropomyosine, dont l'inclusion est réprimée par ce type de séquences dans les cellules du muscle lisse. Toutefois, la surexpression d'une protéine CUG-BP finit par conduire, au contraire, à l'inclusion de l'exon 3 (Gromak & Smith 2002). Les protéines CUG-BP reconnaissent donc probablement la séquence CUG/UGC régulant l'inclusion de l'exon 3, mais leur fixation doit être gênée par la fixation sur la même séquence d'un facteur antagoniste qui n'a pas encore été identifié.

II.3.5 Répétitions UG

L'observation d'une implication de ce type de séquences dans l'épissage est assez récent. On les retrouve juxtaposées à la séquence polypyrimidine de l'intron et elles inhibent l'épissage en fixant le facteur TDP-43³⁵ (Buratti *et al.* 2001). Cette protéine cellulaire est capable d'inhiber *in vitro* la trans-

³³ « maleless »

³⁴ « CUG-binding protein »

³⁵ « HIV-1 TAR DNA binding protein-43 »

cription du virus HIV-1 en se fixant à l'élément TAR³⁶ (Ou *et al.* 1995).

La protéine TDP-43 se fixe à une séquence constituée de 6 répétitions UG et son affinité pour cette séquence augmente avec le nombre de répétitions (Buratti & Baralle 2001).

Les séquences UG et la protéine TDP-43 sont liées à des formes monosymptomatiques de la mucoviscidose. En effet, dans le génome humain, la région située en amont de l'exon 9 du pré-messager CFTR³⁷ présente un polymorphisme de type (TG)_x(T)_y dont la composition fait varier le taux d'exclusion de l'exon 9 (Pagani *et al.* 2000). Chez certains patients, un grand nombre de répétitions UG et/ou un plus petit nombre de U dans l'ARN pré-messager CFTR inhibent l'inclusion de l'exon 9 et conduit à la production d'une protéine inactive. Cette exclusion peut s'expliquer par une compétition entre la fixation de TDP-43 à la séquence (UG)_m et la fixation du facteur U2AF⁶⁵ à la séquence polypyrimidine (U)_n juxtaposée. (Buratti *et al.* 2001).

Dans un cas décrit, celui du gène de l'apolipoprotéine AII, les répétitions UG sont requises pour l'inclusion de l'exon 2, car elles correspondent à une séquence polypyrimidine fonctionnelle (Shelley & Baralle 1987).

II.3.6 Séquences introniques riches en G

Des triplets de G sont fréquemment trouvés dans l'intron au voisinage des sites 5' d'épissage. Ces séquences semblent constituer des identificateurs des exons au sein du génome, au même titre que les sites d'épissage (Nussinov 1988 ; Solovyev *et al.* 1994 ; Lim & Burge 2001). On les retrouve souvent dans les introns de petite taille dont elles permettent un épissage efficace (McCullough & Berget 1997 ; Lim & Burge 2001). Dans l'ARN pré-messager codant l'hormone de croissance humaine I, un motif riche en G situé dans le 3^e intron favorise l'inclusion de l'exon amont, en interagissant directement avec le snRNA U1 (McCarthy & Phillips 1998). Ce ne sont pas les nucléotides 2 à 7 du snRNA U1, généralement associés avec le site donneur d'épissage, mais les nucléotides 8 à 10 qui sont impliqués dans l'interaction. Les auteurs de l'étude n'ont pas pu déterminer si le snRNA U1 pouvait se fixer simultanément sur le site donneur d'épissage et sur ces séquences riches en G (McCullough & Berget 2000).

Dans l'ARN pré-messager *cTNT* de poulet, six répétitions (G)GGGCUG, localisées en aval de l'exon 17, permettent l'inclusion constitutive de cet exon de 6 nucléotides (Carlo *et al.* 1996). Lors des 1^{res} étapes de l'assemblage du spliceosome, la fixation simultanée du facteur SF1/mBBP sur ces séquences et sur le point de branchement favoriserait la reconnaissance de ce petit exon par les facteurs spliceosomaux (Carlo *et al.* 2000).

II.3.7 Séquences (U)GCAUG

Le motif (U)GCAUG active l'épissage lorsqu'il est localisé dans l'intron ou à la fois, dans l'exon et l'intron. On le retrouve, par exemple, chez la souris, en aval de l'exon N1 de *c-src* qui est exprimé

³⁶« transactivation responsive element »

³⁷« cystic fibrosis transmembrane conductance regulator »

uniquement dans les neurones (Modafferi & Black 1997). Ce motif fixe la protéine KSRP³⁸ (Min *et al.* 1997) faisant partie, avec les protéines hnRNP H et F, du complexe protéique activant l'inclusion de l'exon N1 (Min *et al.* 1997 ; Chou *et al.* 1999 ; Markovtsov *et al.* 2000). La protéine KSRP est abondante dans les neurones, mais elle est également exprimée dans les lignées cellulaires non-neuronales, elle ne suffit donc pas pour l'inclusion tissu-spécifique de l'exon N1.

Dans l'ARN pré-messager *CT/CGRP* de rat, cinq répétitions (U)GCAUG (dont trois sont introniques) activent l'inclusion de l'exon 4 dans les cellules exprimant la calcitonine. Néanmoins, l'identité du ou des facteurs se fixant à ces séquences n'est pas connue (Hedjran *et al.* 1997).

II.3.8 Séquence UCAY

Cette séquence activatrice permet la régulation de l'épissage d'ARN pré-messagers du cerveau. Elle fixe la protéine neuronale NOVA-1 qui régule l'épissage alternatif d'ARN pré-messagers comme celui de la sous-unité $\gamma 2$ du récepteur GABA_A et de la sous-unité $\alpha 2$ du récepteur de la glycine (Jensen *et al.* 2000a, b ; pour revues, Perrone-Bizzozero & Bolognani 2002 ; Cáceres & Kornblihtt 2002).

II.4 Implication de pseudo-sites 5' et 3' dans les régulations de l'épissage

Le génome des eucaryotes supérieurs contient statistiquement un grand nombre de séquences présentant une homologie avec des sites d'épissage, sans pour autant constituer des sites d'épissage fonctionnels. Ces pseudo-sites d'épissage constituent le plus souvent des éléments inhibiteurs d'épissage. Les séquences présentant une similitude avec un site 3' d'épissage (UAG) inhibent l'épissage, non pas en recrutant des composants spliceosomaux, mais en fixant la protéine hnRNP A1, comme nous l'avons vu précédemment (Tab. II.2). Quant aux séquences inhibitrices contenant un pseudo-site 5', elles recrutent directement la particule snRNP U1, ce qui empêche la fixation de cette dernière au site 5' naturel. Un pseudo-site 5' régule ainsi l'épissage de l'ARN de l'élément P de la drosophile, dont nous avons vu l'exemple précédemment (Fig. II.4), ainsi que celui du virus du sarcome de Rous (RSV) (Fig. II.15).

L'élément NRS³⁹ du virus du Sarcome de Rous est localisé dans l'intron en amont des 2 sites accepteurs 3' *env* et 3' *src* qu'il réprime. Il contient une séquence riche en purines fixant ASF/SF2, juxtaposée à 2 pseudo-sites 5' chevauchants auxquels peuvent s'associer, soit la particule snRNP U1, soit la particule snRNP U11 spécifique de l'épissage des introns U12-dépendants (Gontarek *et al.* 1993 ; Hibbert *et al.* 1999 ; McNally & McNally 1996, 1998, 1999). La liaison de la snRNP U1 avec le pseudo-site est stabilisée par la protéine ASF/SF2 fixée à sa séquence riche en purines.

L'épissage au site 5' naturel, situé à plus de 300 nts en amont, va dépendre de la compétition entre les snRNP U1 et U11 pour leur fixation sur le NRS (Fig. II.15). Si la snRNP U1 s'y fixe, le site 5' naturel reste vacant et il n'y a pas épissage. Si, au contraire, la snRNP U11 s'y fixe, la snRNP U1 peut

³⁸ « KH-type splicing regulatory protein »

³⁹ « negative regulator of splicing »

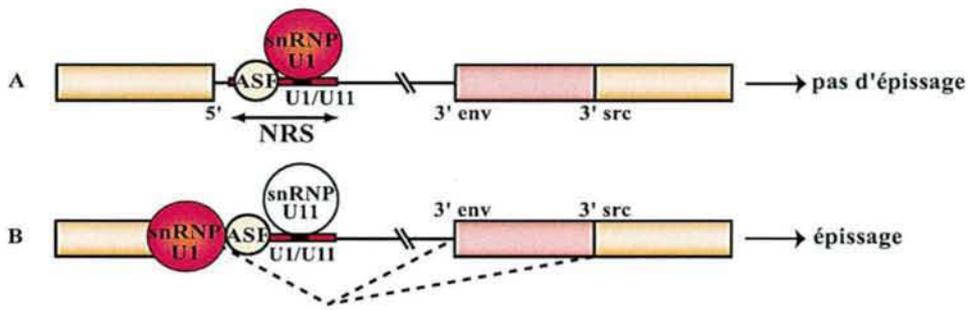


FIG. II.15 – Régulation de l'épissage de l'ARN du virus du sarcome de Rous.

(A) La fixation de la snRNP U1 sur l'élément NRS est stabilisée par ASF/SF2. Elle empêche l'utilisation du site 5' naturel, conduisant à la formation d'un complexe spliceosomal abortif.
 (B) La fixation préalable de la snRNP U11 sur l'élément NRS conduit à la fixation de la snRNP U1 sur le site 5' permettant l'assemblage d'un spliceosome actif.

alors interagir avec le site 5' naturel, ce qui permet l'épissage aux sites 3' *env* et 3' *src* (Hibbert *et al.* 1999).

II.5 Implication des codons non-sens au sein des phases de lecture dans la régulation de l'épissage

Les ARNm portant des codons non-sens au sein des phases de lecture posent un problème pour la cellule, car ils codent des protéines tronquées. Nous avons vu dans le 1^{er} chapitre de l'introduction que ce type d'ARNm était pris en charge par les complexes du NMD, conduisant à la dégradation de ces ARN (§ I.3.5.2).

Une autre réponse cellulaire à la présence de ces codons STOP prématurés a été mise en place (pour revues, Valentine 1998 ; Maquat 2002). Ce phénomène appelé NAS⁴⁰, indépendant du NMD (Wang *et al.* 2002a), contrôle le choix des sites d'épissage afin d'éliminer les codons STOP prématurés et d'assurer le maintien des ORF⁴¹ dans les ARN messagers (pour revues, Mendell & Dietz 2001 ; Maquat 2002). Il implique la reconnaissance, avant l'étape d'épissage, des signaux de traduction sur les ARN pré-messagers. Le mécanisme d'action n'est pour l'instant pas connu, mais des études récentes ont permis d'avancer dans la compréhension de ce phénomène.

Le système NAS dépend de la présence des signaux de traduction, à commencer par le codon initiateur AUG (Wang *et al.* 2002a). Il est déclenché lorsque des mutations de la phase ouverte de lecture sont détectées au cours d'un mécanisme « scannant » l'ARN pré-messager. Par exemple, au cours du développement du lymphocyte T, le gène TCR β ⁴² subit de nombreuses mutations pouvant générer un codon STOP prématuré. Afin d'éviter la production d'une protéine tronquée, 2 sites d'épissage cryptiques sont activés pour produire un ARNm dont l'exon VDJ β a été remplacé par un exon plus court, ne renfermant pas le codon STOP prématuré (Fig. II.16 ; Wang *et al.* 2002b).

⁴⁰ « nonsense-associated altered splicing »

⁴¹ « open reading frame »

⁴² « T cell receptor- β »

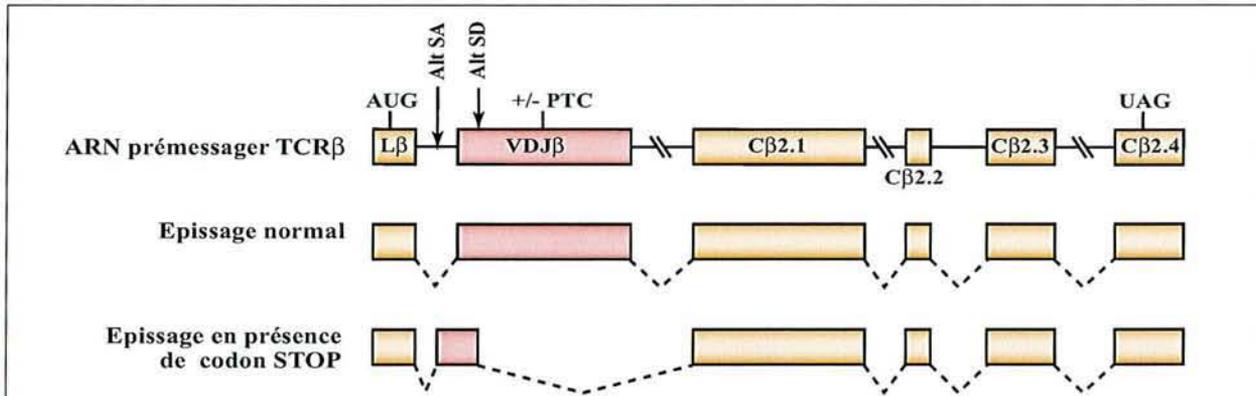


FIG. II.16 – *Epissage alternatif de l'ARN prémessager TCRβ en présence d'un codon STOP prématuré (d'après Wang et al. 2002a).*
 En présence d'un codon STOP prématuré (noté PTC), 2 sites alternatifs notés Alt SA et Alt SD sont utilisés pour générer un ARN messenger ne contenant pas de codon STOP prématuré.

II.6 Implication de la structure secondaire des ARN dans la régulation de l'épissage

II.6.1 Rôle dans l'activation de l'épissage

La structure secondaire joue un rôle important pour l'épissage des ARN prémessagers possédant des introns de grande taille, en favorisant le rapprochement dans l'espace des exons. Ainsi, l'intron du gène RP51B de levure est auto-complémentaire sur 15 nts, ce qui rapproche le site 5' du point de branchement et favorise l'assemblage du spliceosome (Goguel & Rosbash 1993 ; Charpentier & Rosbash 1996).

La structure secondaire régule également l'épissage en rendant des séquences activatrices accessibles. C'est le cas notamment de l'élément ESE de l'exon EDA de la fibronectine. L'étude expérimentale de la structure a montré que l'élément ESE est placé dans une boucle terminale, ce qui le rendrait accessible aux facteurs protéiques qui assurent l'activation, en particulier les protéines SR (Lavigneur *et al.* 1993 ; Muro *et al.* 1999).

Dans le cas de l'ARN prémessager *FGFR-2* renfermant deux exons mutuellement exclusifs, IIIb et IIIc, une structure secondaire engageant 2 séquences régulatrices introniques, ISE-2⁴³ et ISAR⁴⁴, localisées en aval de l'exon IIIb, favorise l'inclusion de l'exon IIIb dans les cellules épithéliales, chez l'homme comme chez le rat (Del Gatto *et al.* 1997 ; Muh *et al.* 2002). Ces deux séquences sont distantes de 736 nts et s'apparient sur 18 nts. Ce ne sont pas les séquences en elles-mêmes, mais plutôt la structure résultant de leur appariement qui est primordiale pour l'inclusion de l'exon IIIb. En effet, toute mutation de la séquence ISE-2 perturbant la structure secondaire affecte la régulation de l'épissage, tandis que des mutations compensatoires de la séquence ISAR restaurent l'inclusion de l'exon (Muh *et al.* 2002). En outre, le remplacement des séquences ISE-2 et ISAR par 2 séquences complémentaires ne perturbe pas la régulation. Des protéines capables de s'associer aux ARN double-brins stabilisent probablement l'appariement des séquences ISE-2 et ISAR, mais elles n'ont pas encore été

⁴³ « intronic splicing enhancer 2 »

⁴⁴ « intronic splicing activator and repressor »

identifiées. Les protéines hnRNP constitueraient de bons candidats (Muh *et al.* 2002). Elles pourraient favoriser l'inclusion de l'exon IIIb en recrutant des facteurs spliceosomaux.

II.6.2 Rôle dans l'inhibition de l'épissage

La structure secondaire d'un ARN pré-messager peut séquestrer les sites d'épissage, qui deviennent alors inaccessibles aux facteurs du spliceosome. C'est le cas de l'ARN de la β -tropomyosine de poulet (Sirand-Pugnet *et al.* 1995) ou encore, de l'ARN pré-génomique du virus de l'hépatite C. Les structures secondaires formées dans ce dernier impliquent 2 régions, A et B, distantes de 1100 nts, complémentaires sur 66 à 68 nts et contenant respectivement les sites 5' et 3' (Loeb *et al.* 2002).

La structure secondaire de l'ARN pré-messager de l'hormone de croissance humaine permet de réguler le choix de deux sites 3' d'épissage en compétition. La formation d'une tige-boucle englobant le site proximal majeur B et son point de branchement conduit à l'utilisation du site distal mineur, ce qui produit un isoforme de l'hormone de croissance (Estes *et al.* 1992).

Dans le système nerveux, la structure secondaire est un facteur important de la régulation des taux de transcrits alternatifs produits : toute mutation perturbant ou renforçant cette structure perturbe le ratio d'ARNm produits (pour revue, Grabowski & Black 2001). C'est le cas de l'ARN messager *tau* codant une protéine impliquée dans la polymérisation et la stabilisation des microtubules et qui contribue ainsi au maintien du cytosquelette neuronal et au transport neuronal (Fig. II.17).

Après épissage alternatif des exons 2, 3 et 10 de l'ARN pré-messager *tau*, différents isoformes ayant des propriétés fonctionnelles distinctes sont obtenus (Fig. II.17-A et B ; Goode *et al.* 2000 ; pour revue, Buée *et al.* 2000) et le maintien du ratio des ARN messagers de ces différents isoformes est primordial pour le neurone. Ainsi, plusieurs maladies neurodégénératives provoquées par des agrégats intra- ou extracellulaires de protéine Tau insoluble sont dues à un déséquilibre entre les quantités d'ARN messagers contenant ou non l'exon 10. C'est le cas de démences fronto-temporales ou de certaines lésions de la maladie d'Alzheimer.

La régulation de l'inclusion de l'exon 10 est complexe, car elle implique de nombreuses séquences régulatrices, ainsi qu'une structure secondaire tige-boucle renfermant le site 5' et limitant ainsi sa reconnaissance par les composants spliceosomaux. Chez les patients souffrant de ce type de maladie neurodégénérative, on retrouve des mutations ponctuelles déstabilisant ou au contraire, renforçant la structure secondaire de l'exon 10, ce qui affecte le taux d'inclusion de l'exon 10 dans l'ARN mature (Fig. II.17-C ; Hutton *et al.* 1998 ; Grover *et al.* 1999 ; Varani *et al.* 1999 ; pour revue, Buée *et al.* 2000).

Compte-tenu du rôle de la structure secondaire dans la régulation de l'épissage des ARN pré-messagers nucléaires, il est possible que des facteurs protéiques soient capables d'ouvrir ou fermer cette structure pour permettre l'accessibilité des signaux d'épissage aux facteurs spliceosomaux. Comme nous l'avons vu précédemment (§ II.2.1.1), la protéine hnRNP A1 constitue l'un de ces facteurs car elle participe à la formation de ces structures. Qu'en est-il des ARN hélicases ?

L'une d'entre elles, l'ARN hélicase p72, a été récemment décrite comme étant impliquée dans l'épis-

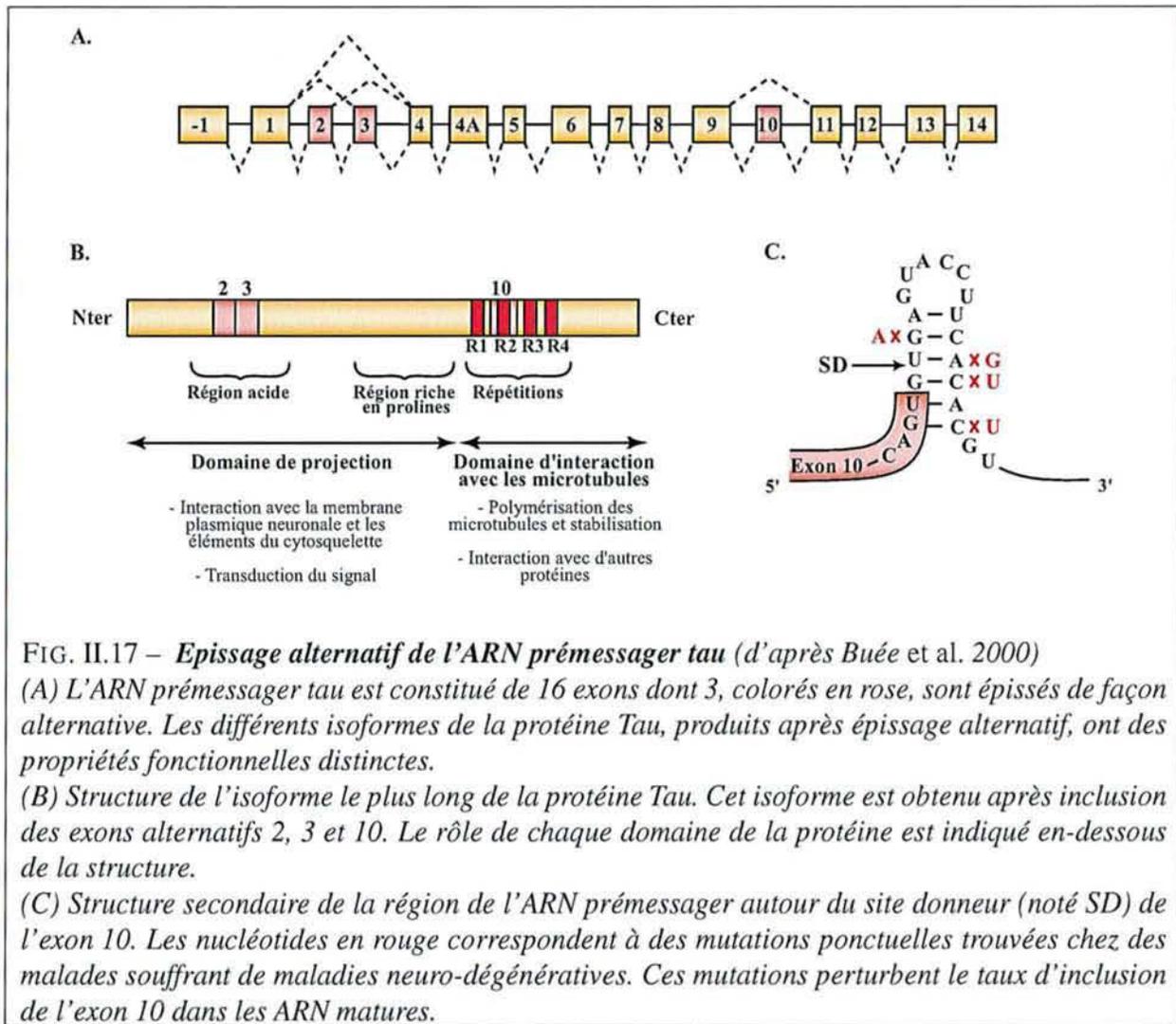


FIG. II.17 – *Epissage alternatif de l'ARN pré-messager tau (d'après Buée et al. 2000)*
 (A) L'ARN pré-messager tau est constitué de 16 exons dont 3, colorés en rose, sont épissés de façon alternative. Les différents isoformes de la protéine Tau, produits après épissage alternatif, ont des propriétés fonctionnelles distinctes.
 (B) Structure de l'isoforme le plus long de la protéine Tau. Cet isoforme est obtenu après inclusion des exons alternatifs 2, 3 et 10. Le rôle de chaque domaine de la protéine est indiqué en-dessous de la structure.
 (C) Structure secondaire de la région de l'ARN pré-messager autour du site donneur (noté SD) de l'exon 10. Les nucléotides en rouge correspondent à des mutations ponctuelles trouvées chez des malades souffrant de maladies neuro-dégénératives. Ces mutations perturbent le taux d'inclusion de l'exon 10 dans les ARN matures.

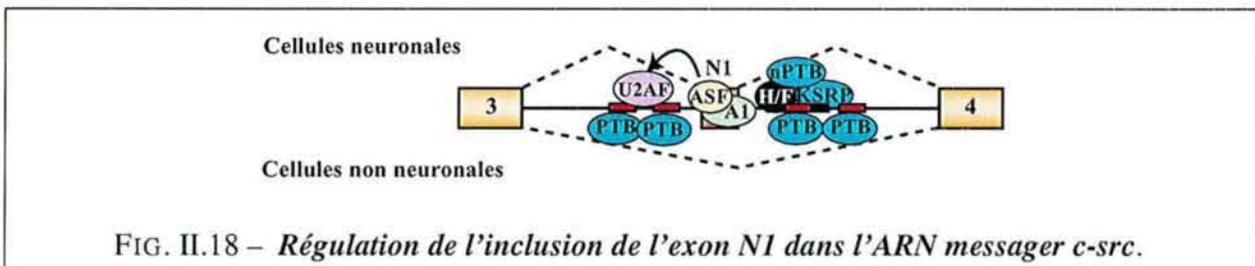
sage alternatif de l'ARN pré-messager CD44 dans un processus ATP-dépendant, au cours duquel elle favorise l'inclusion des exons v4 et v5. La protéine agirait au niveau des interactions ARN/protéine, plutôt que de remodeler les interactions ARN/ARN. En effet, la déstabilisation de la structure tige-boucle renfermant l'exon v4 n'affecte pas l'action de la protéine p72. Par contre, les séquences activatrices de type ACE de l'exon v4 sont nécessaires pour l'activation de l'inclusion de l'exon par l'hélicase p72 (Honig *et al.* 2002).

II.7 L'épissage alternatif : un contrôle combinatoire

Nous avons vu dans les chapitres précédents les mécanismes de base pouvant assurer la régulation de l'épissage alternatif des ARN pré-messagers. Il est apparu que l'épissage peut être soumis au contrôle simultané de plusieurs de ces mécanismes. Dans certains cas, la régulation peut atteindre un degré de complexité supérieur, puisqu'elle fait intervenir d'autres processus biologiques comme la transcription ou encore l'édition. A ces multiples régulations, s'ajoute la régulation de la production, de la localisation ou de l'activité des facteurs d'épissage. Nous donnerons ici un certain nombre d'exemples.

II.7.1 De multiples régulations contrôlent l'épissage de l'ARN pré-messager *c-src*

L'épissage alternatif de l'ARN *c-src*, codant une tyrosine kinase, fait intervenir un exon cassette de 18 nts, l'exon N1. La petite taille de cet exon le rend difficilement reconnaissable par la machinerie spliceosomale. Il va être, par conséquent, la cible de nombreuses régulations, conduisant à son inclusion dans les cellules neuronales et à son exclusion dans les cellules non neuronales (Fig. II.18-A ; pour revue, Smith & Valcárcel 2000). Au total, trois régions de l'ARN *c-src* contribuent à cette régulation.



Dans les cellules non-neuronales, la fixation de la protéine PTB au niveau de 2 éléments introniques situés de part et d'autre de l'exon N1 conduit à l'exclusion de l'exon (Chou *et al.* 2000).

Dans les cellules neuronales, la région intronique située en aval de l'exon N1, qui est normalement reconnue par la protéine PTB, fixe cette fois-ci la protéine nPTB. Comme nous l'avons vu dans le § II.2.1.3, dans les cellules neuronales, cette protéine est majoritaire par rapport à son homologue PTB et elle exerce un effet répressif plus faible que la protéine PTB *in vitro* (Markovtsov *et al.* 2000). La protéine nPTB favorise la formation, sur cette région intronique, d'un complexe constitué en partie des protéines hnRNP H, F et de la protéine neuronale KSRP. Pour le moment, le mécanisme d'action de ce complexe n'est pas compris (Min *et al.* 1995, 1997 ; Chou *et al.* 1999 ; Markovtsov *et al.* 2000). Une 3^e région, riche en purines, située dans l'exon N1 lui-même, intervient dans cette régulation et fixe les protéines ASF/SF2, hnRNP H, F, A1 et A1^B (Rooke *et al.* 2003). L'inclusion de l'exon N1 dépendrait de la fixation à cette région des protéines antagonistes ASF/SF2 et hnRNP A1. La protéine ASF/SF2 pourrait alors stabiliser la fixation du facteur U2AF sur la séquence polypyrimidine dans les cellules neuronales, ce qui empêcherait la fixation de la protéine PTB. La protéine hnRNP A1^B favoriserait également l'inclusion de l'exon N1, en se fixant avec ASF/SF2 sur la séquence riche en purines, plus fortement dans les cellules neuronales que dans les cellules non neuronales. Le rôle des protéines hnRNP H et F n'est pas connu pour le moment (Rooke *et al.* 2003).

II.7.2 La structure du promoteur transcriptionnel régule l'activation de l'épissage de l'ARN pré-messager de la fibronectine par les protéines SR

La structure du promoteur transcriptionnel régulerait l'épissage alternatif. Ainsi, *in vivo*, l'inclusion de l'exon EDA de l'ARN pré-messager de la fibronectine dépend, à la fois, de la présence d'un élément ESE fixant les protéines SR ASF/SF2 et 9G8, et de la capacité de la région promotrice à recruter les protéines SR sur le domaine CTD de l'ARN polymérase II (Fig. II.19) (Cramer *et al.* 1999 ; pour

revue, Cáceres & Kornblihtt 2002).

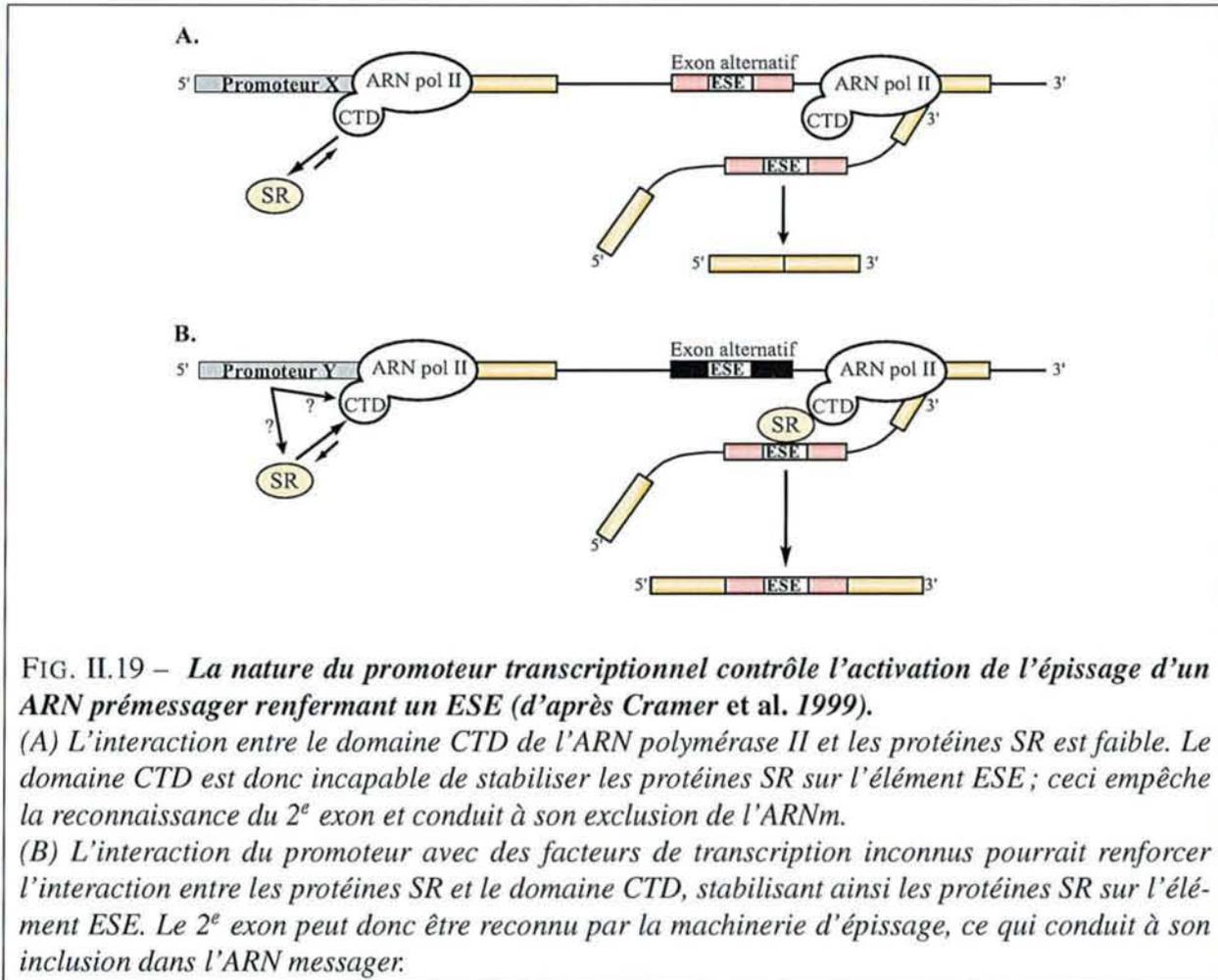


FIG. II.19 – *La nature du promoteur transcriptionnel contrôle l'activation de l'épissage d'un ARN pré-messager renfermant un ESE (d'après Cramer et al. 1999).*

(A) *L'interaction entre le domaine CTD de l'ARN polymérase II et les protéines SR est faible. Le domaine CTD est donc incapable de stabiliser les protéines SR sur l'élément ESE ; ceci empêche la reconnaissance du 2^e exon et conduit à son exclusion de l'ARNm.*

(B) *L'interaction du promoteur avec des facteurs de transcription inconnus pourrait renforcer l'interaction entre les protéines SR et le domaine CTD, stabilisant ainsi les protéines SR sur l'élément ESE. Le 2^e exon peut donc être reconnu par la machinerie d'épissage, ce qui conduit à son inclusion dans l'ARN messager.*

Plus récemment, il a été proposé que la vitesse d'élongation de l'ARN polymérase II joue également un rôle (Kadener *et al.* 2002 ; Nogués *et al.* 2002 ; pour revue, Cáceres & Kornblihtt 2002). Ainsi, une ARN polymérase II peu processive, ou faisant des pauses internes au cours de l'élongation, favoriserait l'inclusion de l'exon alternatif EDA, tandis qu'en présence d'une ARN polymérase II processive, ou ne faisant pas de pauses internes, l'exon EDA serait exclus.

II.7.3 L'épissage de l'ARN pré-messager *ADAR2* est régulé par édition

Quelques enzymes, appartenant à la famille des protéines ADAR⁴⁵, catalysent la conversion d'une adénosine en une inosine dans les ARN pré-messagers. Ce phénomène d'édition change la nature de l'acide aminé codé dans la majorité des cas. Il arrive néanmoins que le changement ait lieu dans une séquence intronique. L'enzyme ADAR2 utilise ce système pour réguler sa propre expression. Elle réalise la conversion A → I au niveau d'une séquence AA d'un intron de son propre ARN pré-messager ; cela crée un dinucléotide AI mimant un site 3' d'épissage. L'utilisation de ce site provoque une insertion de 47 nts dans l'ARN pré-messager, conduisant à un décalage du cadre de lecture et à la

⁴⁵ « adenosine deaminase that acts on RNA »

production d'une protéine tronquée inactive (Rueter *et al.* 1999 ; pour revue, Schaub & Keller 2002). Ce type de régulation constitue un cas isolé.

II.7.4 Régulation de l'expression et de l'activité des facteurs régulateurs d'épissage

Les facteurs d'épissage comme les protéines SR et hnRNP sont les acteurs majeurs des régulations de l'épissage que nous avons abordées, mais leur expression et leur activité sont, elles aussi, régulées.

II.7.4.1 Contrôle de la production des facteurs intervenant dans l'épissage

Au cours de cette introduction, nous avons déjà pu voir que les taux de certains facteurs d'épissage varient suivant les tissus ou encore, le stade du développement, et que l'épissage alternatif de leur propre ARN prémessager permet la production d'isoformes pouvant avoir des propriétés régulatrices différentes (cf § I.3.2.1 et II.2.1.1). C'est le cas de la famille des protéines hnRNP A/B et de la protéine SR ASF/SF2 (Yang *et al.* 1994 ; Mayeda *et al.* 1994 ; Jumaa & Nielsen 2000). Certaines protéines comme SC35 ne possèdent pas d'isoformes, mais contrôlent l'épissage de leur propre ARN prémessager (Sureau *et al.* 2001).

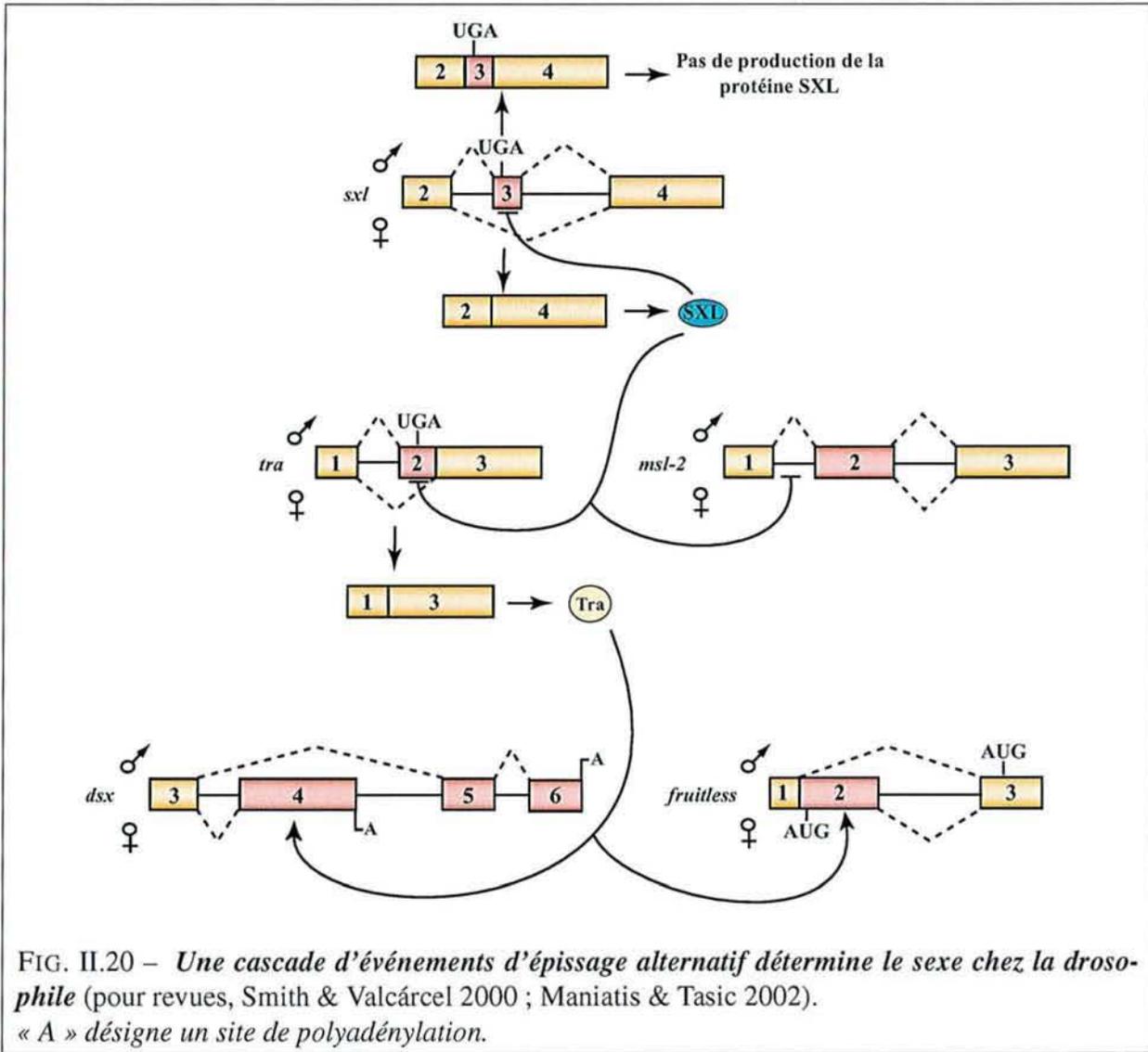
Le niveau d'expression des protéines SR dépend aussi de l'état physiologique de la cellule. Par exemple, lors de l'activation des lymphocytes T, les changements moléculaires à la surface des cellules T sont corrélés à un changement d'expression des protéines SR, qui régulent l'épissage de l'ARN codant la protéine transmembranaire CD45 (Ten Dam *et al.* 2000).

De même, au cours de la progression du cancer du sein, l'inclusion de certains exons dans l'ARN prémessager *CD44*, sous l'effet d'une variation du niveau d'expression des protéines SR, produit une protéine d'adhésion cellulaire impliquée dans la formation des métastases (Stickeler *et al.* 1999 ; pour revue, Smith & Valcárcel 2000).

Chez la drosophile, une cascade d'épissages alternatifs, conduisant à la détermination du sexe, dépend de la production de la protéine SXL (Fig. II.20).

La protéine SXL n'est exprimée que chez les femelles et régule l'expression de son propre ARN prémessager en réprimant l'inclusion de l'exon 3, ce qui conduit à la production d'une protéine fonctionnelle (cf § II.2.1.3 ; pour revues, Smith & Valcárcel 2000 ; Maniatis & Tasic 2002). Chez les mâles, l'exon 3 est systématiquement inclus, ce qui empêche la production de protéine SXL, du fait de la présence d'un codon STOP. La protéine SXL réprime également l'inclusion de l'exon 2 de *tra*, qui contient aussi un codon STOP, ce qui permet la production d'une protéine fonctionnelle chez les drosophiles femelles (cf § II.2.1.3).

La protéine Tra régule à son tour l'épissage des ARN prémessagers *dsx* et *fruitless* (cf § II.2.3). Ainsi, chez les femelles, l'inclusion de l'exon 4 dans l'ARN prémessager *dsx* produit une protéine DSX^F réprimant la transcription des gènes spécifiques des mâles, alors que chez les mâles, l'exclusion de l'exon 4 donne lieu à la protéine DSX^M, réprimant la transcription des gènes spécifiques des femelles. La protéine Tra régule également l'épissage de l'ARN prémessager *fruitless*, provoquant l'inclusion de l'exon 2 chez les femelles.



Enfin, la protéine SXL agit de 2 façons sur la production de la protéine MSL-2. Elle provoque d'une part la rétention du 1^{er} intron chez les drosophiles femelles (cf § II.3.3.1). D'autre part, elle provoque la rétention, dans le noyau, des ARNm contenant cet intron et elle réprime ainsi leur traduction dans le cytoplasme.

II.7.4.2 Contrôle de l'activité des facteurs d'épissage

Contrôle de l'activité des protéines SR par les protéines SRrp

Les protéines SRrp⁴⁶ sont des protéines homologues aux protéines SR. Bien qu'elles ne complètent pas d'extrait cytoplasmique S100, certaines de ces protéines sont capables de réguler directement l'épissage, tandis que d'autres modulent l'activité des protéines SR.

Deux de ces protéines, SRrp35 et SRrp40 (encore appelée SRp38, Shin & Manley 2002) sont homologues à la protéine SC35 et lors de l'épissage, elles agissent comme des facteurs antagonistes des protéines SR, probablement en entrant en compétition pour leur fixation aux ARN pré-messagers (Cowper *et al.* 2001). L'expression des protéines SRrp35 et SRrp40 diffère suivant les tissus, et de ce

⁴⁶« SR repressor protein »

fait, la variation du ratio entre les protéines SRrp et les protéines SR régulerait le profil d'épissage de nombreux ARN pré-messagers (Cowper *et al.* 2001 ; Shin & Manley 2002).

La protéine SRrp86 régule quant à elle l'épissage, mais en contrôlant l'effet des protéines SR : elle va agir soit de façon synergique avec les protéines SR, soit de façon antagoniste. Elle module ainsi la sélection des sites d'épissage en réprimant les protéines ASF/SF2 et SC35, et en activant la protéine SRp20. SRrp86 est capable d'interagir directement avec SC35 et SRp20 (Barnard & Patton 2000 ; Barnard *et al.* 2002). On retrouve également cette propriété d'activation de l'épissage, en synergie avec les protéines SR, pour la protéine RNPS1⁴⁷ (Mayeda *et al.* 1999).

Contrôle de l'activité et de la disponibilité des facteurs d'épissage par leur état de phosphorylation et leur localisation

La phosphorylation est un mécanisme majeur de contrôle de la localisation des facteurs d'épissage et de leur activité.

Les protéines SR y sont particulièrement sensibles puisque, comme nous avons pu le voir dans le 1^{er} chapitre de l'introduction, la phosphorylation joue sur leur capacité à établir des interactions avec l'ARN ou d'autres protéines à domaine RS, et sur leur localisation. Chez l'adénovirus, les protéines SR activent l'épissage des ARNm de la phase précoce d'infection avant d'être la cible de régulations limitant leur activité, ce qui permet le passage à la phase tardive de l'infection (Molin & Akusjärvi 2000) : elles sont ainsi déphosphorylées sous l'action d'une protéase virale (Kanopka *et al.* 1998) et titrées par les transcrits tardifs (Himmelspach *et al.* 1995 ; pour revue, Akusjärvi & Stévenin 2003).

Chez l'homme, des processus cellulaires contrôlent également l'activité des facteurs d'épissage en jouant sur leur degré de phosphorylation. Par exemple, suivant l'état de la cellule, différents profils d'épissage sont attendus. En effet, la réponse de la cellule au stress correspond à l'activation de la voie des MAP⁴⁸ kinases et a pour conséquence la phosphorylation de la protéine hnRNP A1 (van der Houven van Oordt *et al.* 2000). La protéine phosphorylée est alors localisée principalement dans le cytoplasme, ce qui modifie le ratio entre la protéine hnRNP A1 et les protéines SR dans le noyau et perturberait l'épissage des ARN pré-messagers (van der Houven van Oordt *et al.* 2000).

Des hormones comme l'acide rétinolique, la vitamine D ou les hormones thyroïdiennes, moduleraient la phosphorylation du domaine CTD de l'ARN polymérase II, influençant ainsi la capacité de ce domaine à interagir avec les protéines SR. Le degré de phosphorylation de ces dernières serait également régulé par la voie ras et la protéine kinase C, qui peuvent être activées par de l'insuline ou après stimulation par du sérum (pour revue, Akker *et al.* 2001).

Le niveau de phosphorylation régule également l'activité des protéines SRrp. Par exemple, la déphosphorylation de la protéine SRrp40 renforce ses propriétés inhibitrices d'épissage (Shin & Manley 2002). Ainsi, au cours de la phase M du cycle cellulaire, la répression de l'expression des gènes semble corrélée à une inhibition de l'épissage, provoquée par la déphosphorylation de la protéine SRrp40 (Shin & Manley 2002).

⁴⁷ « RNA-binding protein with serine-rich domain »

⁴⁸ « mitogen activated protein »

Un autre paramètre contribue à la régulation de l'activité et de la localisation des facteurs d'épissage. Ainsi, l'interaction entre la protéine SRp30c et la protéine YB-1, qui est en général majoritairement cytoplasmique, entraîne, d'une part, la localisation nucléaire de YB-1 et, d'autre part, perturbe la régulation de l'épissage par SRp30c (Bargou *et al.* 1997 ; Raffetseder *et al.* 2003).

De multiples processus participent donc à la régulation de l'épissage des ARN pré-messagers nucléaires. C'est dans ce cadre que le virus HIV-1 régule son épissage alternatif, puisqu'il va utiliser voire, détourner à son profit, la machinerie cellulaire mise en place, afin d'assurer l'expression de ses protéines et de permettre une multiplication virale efficace.

Chapitre III

Epissage de l'ARN du virus HIV-1

III.1 Description du virus HIV-1

Bien que le virus de l'immunodéficience humaine de type 1 (VIH-1, ou HIV¹-1) n'ait été isolé qu'en 1983 par Luc Montagnié et par Robert Gallo, les 1^{res} traces d'infection de l'homme par le virus remontent aux années 1950. Depuis, le virus s'est répandu et est responsable d'une pandémie, le SIDA², dont souffrent aujourd'hui 42 millions d'individus à travers le monde (ONUSIDA/OMS 2002). Le SIDA conduit à une grave immunodépression exposant le malade à des infections opportunistes et des tumeurs qui lui sont fatales.

III.1.1 Infection par le virus HIV-1

Le virus HIV-1 appartient à la famille des lentivirus, c'est-à-dire des virus dont l'infection est lente, puisque elle peut prendre plusieurs années (pour revue, Sleasman & Goodenow 2003). 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 48h 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 pas 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 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 & Goodenow 2003). 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

¹« human immunodeficiency virus »

²« syndrome de l'immunodéficience acquise »

destruction du système immunitaire progresse. Néanmoins, durant cette 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. Les nombreuses maladies opportunistes qui apparaissent conduisent à la mort du patient.

III.1.2 Evolution phylogénétique du virus HIV-1

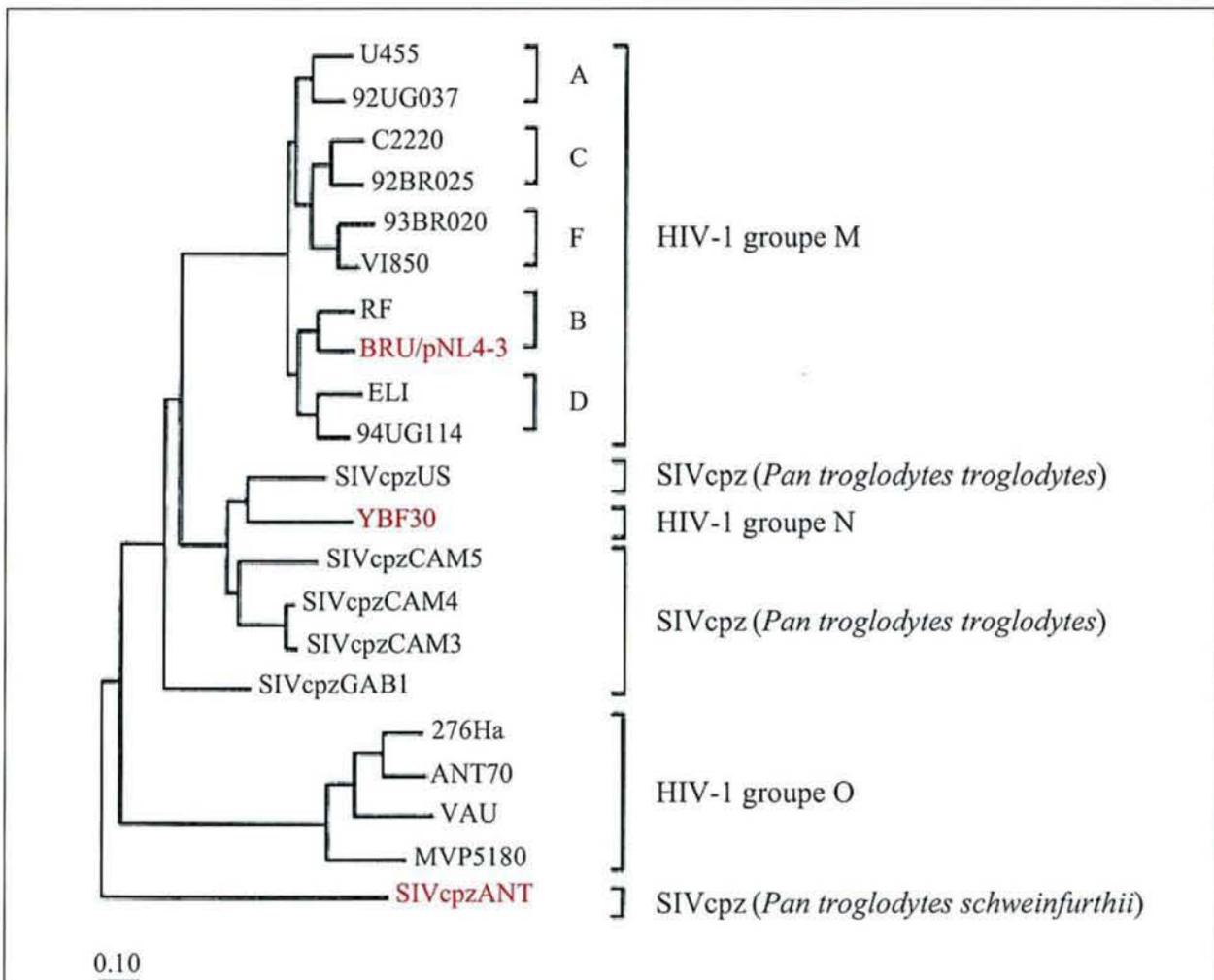


FIG. III.1 – Dendrogramme de souches de virus HIV-1 et SIVcpz établi par la méthode de maximum de vraisemblance à partir de séquences entières de la protéine Env (d'après Hahn et al. 2000).

Les espèces de chimpanzés à partir desquelles ont été isolés les virus sont indiquées en italique. Les souches utilisées au cours de ce travail sont indiquées en rouge. La barre d'échelle indique un remplacement de 0,1 acide aminé par site.

Les souches caractérisées du virus HIV-1 sont réparties en 3 groupes phylogénétiquement liés, les groupes M, N et O. Elles ont toutes la propriété de conférer le SIDA, mais c'est le groupe M³ qui est responsable de la pandémie (Fig. III.1 ; pour revue, Peeters & Sharp 2000). Il est subdivisé en 11 classes (classes A à K). La classe B, à laquelle appartient la souche BRU que nous avons principalement étudiée (encore appelée LAI), est retrouvée principalement en Europe occidentale et en Amérique. Nous avons également utilisé la souche pNL4-3, qui est une forme recombinante de 2 souches de la classe B : la partie 5' de son génome correspond à une souche appelée NY5, tandis que la partie 3' correspond à la souche BRU (Kuiken *et al.* 2001). Les souches de la classe B sont les plus étudiées au niveau de l'épissage.

Les rétrovirus comme HIV-1 sont capables de recombiner entre eux 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 mosaïque de régions de différents sous-types. Par exemple, la souche MAL, largement répandue au Congo (ex-Zaire), est une forme recombinante des souches A, D et I (Kuiken *et al.* 2001).

Le groupe O⁴ est localisé principalement en Afrique de l'Ouest, et le groupe N⁵, découvert récemment, infecte un petit nombre d'individus au Cameroun (Simon *et al.* 1998). Ce groupe, auquel appartient la souche YBF30 que nous avons également étudiée, est éloigné phylogénétiquement du groupe M et proche de certaines souches du virus SIVcpz. Pour l'instant, seules 2 souches de ce groupe ont été isolées.

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 (Gao *et al.* 1999). La date de 1^{re} infection de la population humaine par le virus HIV-1 a été estimée à 1930 (Korber *et al.* 2000 ; Salemi *et al.* 2001). 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 SIVcpz a lui-même évolué suivant l'hôte qu'il a infecté. Ainsi, les virus SIVcpz GAB et US infectent les chimpanzés *Pan troglodytes troglodytes*, tandis que les chimpanzés *Pan troglodytes schweinfurthii* sont infectés par le virus SIVcpz ANT que nous avons étudié (Gao *et al.* 1999). Le virus HIV-2 proviendrait quant à lui d'une transmission à l'homme du virus SIVsm infectant les mangabéys (pour revue, Kuiken *et al.* 2001).

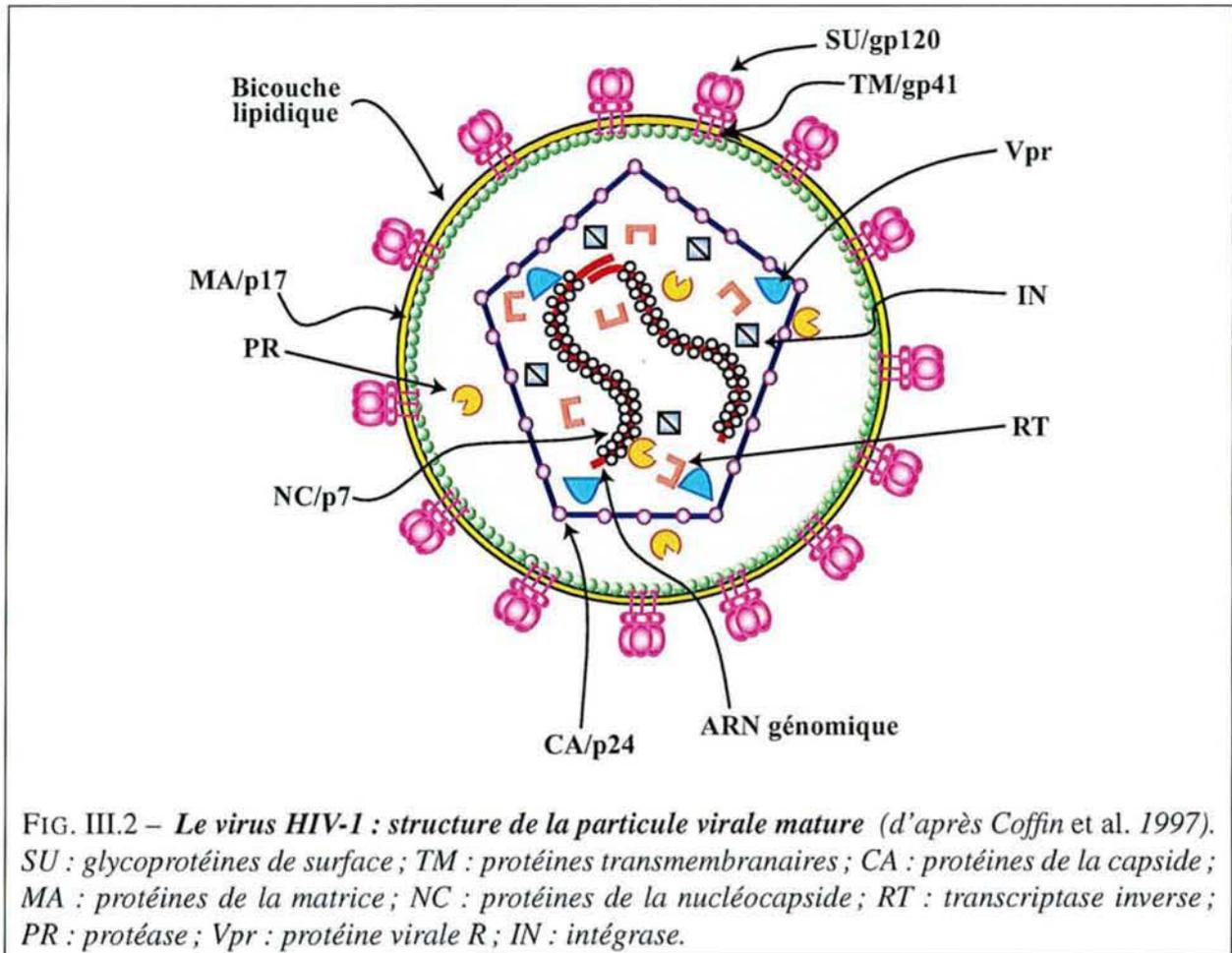
III.1.3 Structure du virus HIV-1

Le virus HIV-1 est une particule d'environ 100 nm de diamètre, délimitée par une enveloppe formée d'une bicouche lipidique dérivant de la membrane de la cellule hôte. Cette enveloppe porte en surface des glycoprotéines (protéines gp120 ou SU), ancrées grâce à leur interaction avec les protéines transmembranaires gp41 ou TM (Fig. III.2). A l'intérieur de la particule, une matrice (protéines p17 ou MA) renferme la capsid (protéines p24 ou CA), qui contient les 2 molécules d'ARN génomique, as-

³« main »

⁴« outlier »

⁵« non M, non O »



sociées avec la protéine de nucléocapside (NCp7), et des protéines à activité enzymatique : l'intégrase (IN), la transcriptase inverse (RT) et la protéase (PR). Le virus encapside également des molécules provenant de la cellule hôte comme des ARN^{Lys3}, de l'ubiquitine et de la cyclophiline A (pour revue, Coffin *et al.* 1997).

III.1.4 ARN génomique

Dans le virion, le génome viral est présent sous la forme de 2 molécules d'ARN identiques d'une longueur de 9 à 10 kb suivant les souches (Fig. III.3), qui sont associées de façon non covalente par leur extrémité 5'. Cette interaction implique en particulier une séquence autocomplémentaire appelée DLS⁶, localisée au niveau du site donneur D1. Comme les ARN pré-messagers cellulaires, ces 2 molécules d'ARN génomique sont coiffées à leur extrémité 5' et polyadénylées à leur extrémité 3'.

III.1.4.1 Régions non traduites

L'extrémité 5' non traduite de l'ARN génomique renferme les séquences R⁷, U5⁸ et PBS⁹, nécessaires pour l'étape de transcription inverse (Fig. III.3). La séquence R contient l'élément TAR, nécessaire à

⁶ « dimer linkage sequence »

⁷ « redondant »

⁸ « unique en 5' »

⁹ « primer binding site »

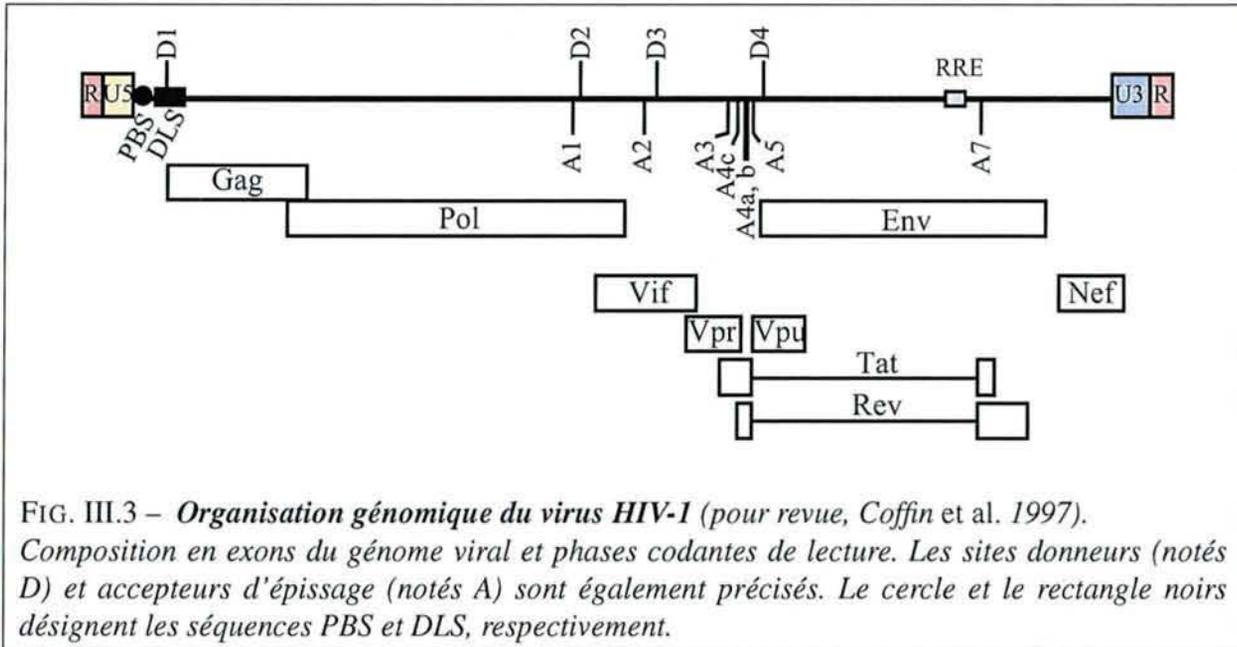


FIG. III.3 – *Organisation génomique du virus HIV-1 (pour revue, Coffin et al. 1997). Composition en exons du génome viral et phases codantes de lecture. Les sites donneurs (notés D) et accepteurs d'épissage (notés A) sont également précisés. Le cercle et le rectangle noirs désignent les séquences PBS et DLS, respectivement.*

l'activation de la transcription virale par la protéine Tat.

L'extrémité 3' renferme quant à elle la séquence U3¹⁰ et la séquence R, copie de la séquence R trouvée en 5', qui contient le signal de polyadénylation.

III.1.4.2 Phases ouvertes de lecture

Le génome viral contient 9 phases ouvertes de lecture ou ORF (Gag, Pol, Env, Tat, Rev, Nef, Vif, Vpr et Vpu), dont 3 sont communes à l'ensemble des rétrovirus : elles codent les précurseurs protéiques Gag, Pol et Env, indispensables pour la multiplication virale. Ces 3 précurseurs sont maturés par coupures protéolytiques, ce qui permet de produire les protéines structurales, les protéines d'enveloppe et les protéines à activité enzymatique (Fig. III.4). Les autres protéines virales (Tat, Rev, Nef, Vif, Vpr et Vpu), ainsi que le précurseur Env, sont produites à partir d'ARN messagers viraux plus ou moins épissés (Fig. III.3 et III.4).

III.1.5 Les protéines du virus HIV-1 et leur fonction dans le cycle de multiplication

III.1.5.1 Le cycle viral du virus HIV-1

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 (Fig. III.5). L'infection d'une cellule par le virus HIV-1 débute par la reconnaissance des glycoprotéines gp120 et gp41 de l'enveloppe virale, par le récepteur cellulaire CD4 et le co-récepteur CCR5 ou CXCR4 (Fig. III.5 ; pour revue, Tang *et al.* 1999).

Cette interaction conduit à la fusion de l'enveloppe du virus avec la membrane plasmique de la cellule hôte et permet à la capsid renfermant l'ARN viral de pénétrer dans la cellule, puis de se disloquer.

¹⁰« unique en 3' »

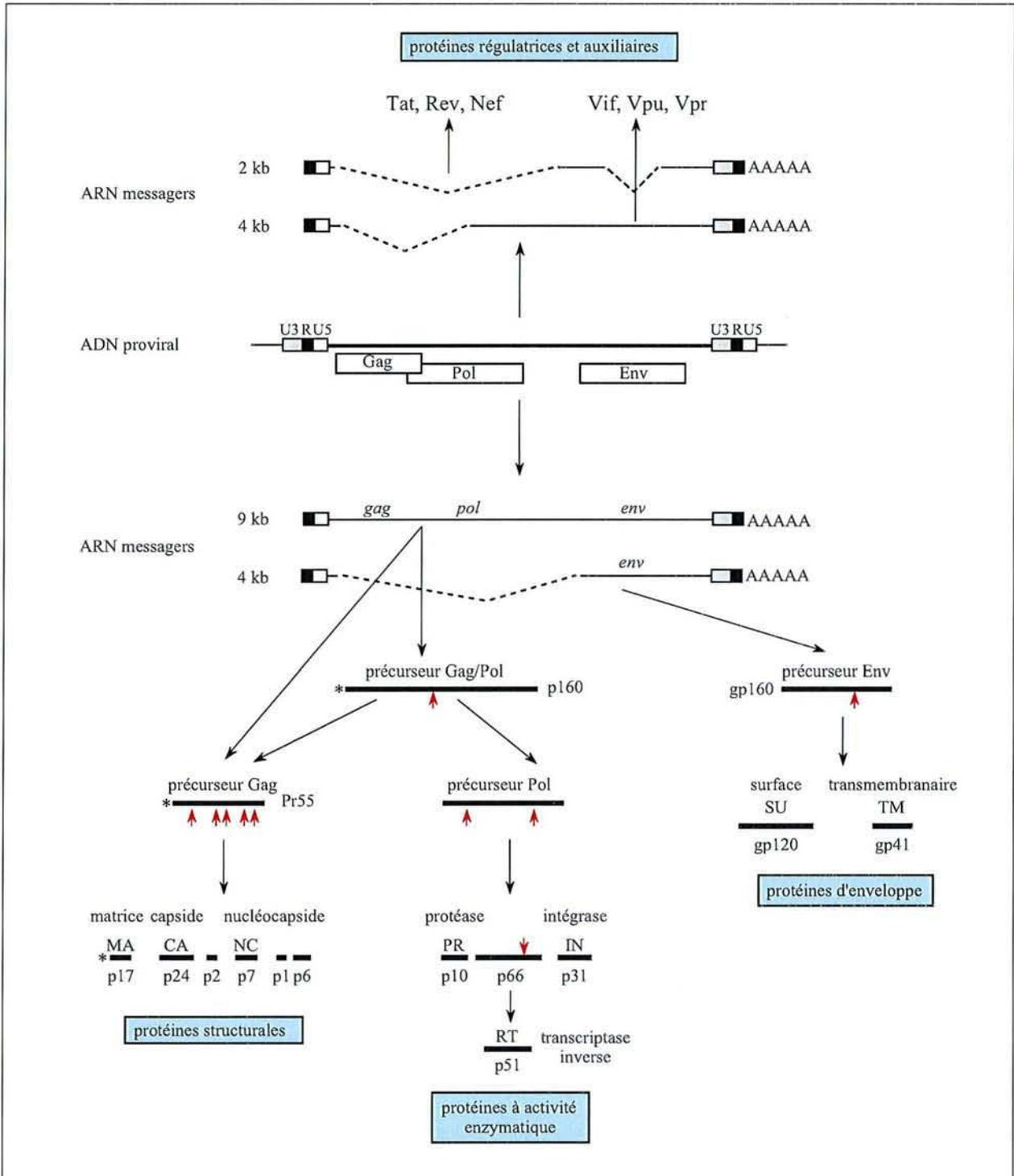
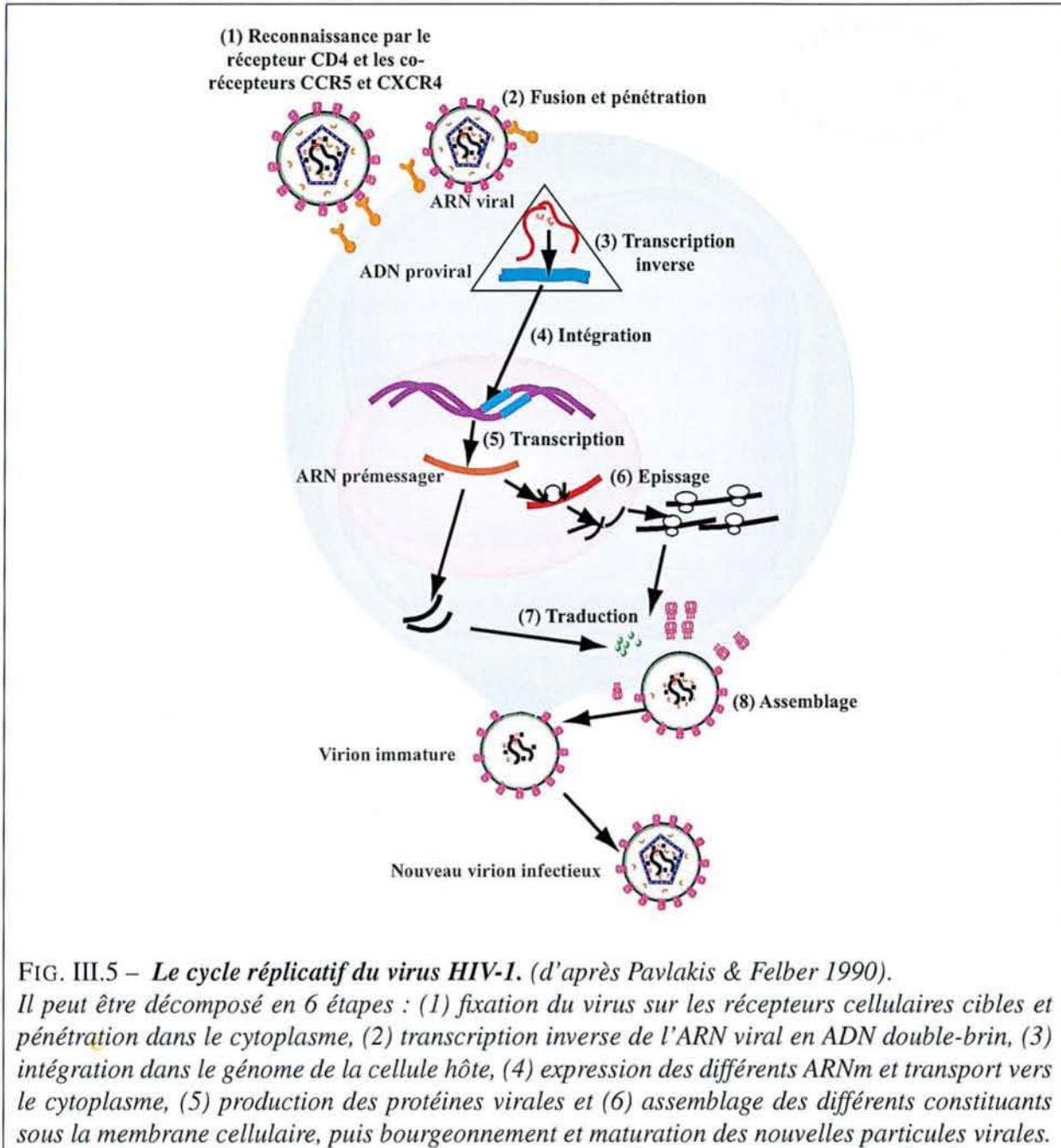


FIG. III.4 – *Synthèse et maturation des protéines du virus HIV-1* (pour revue, Coffin et al. 1997.)
 L'ADN proviral code 9 protéines. Les précurseurs Gag et Pol sont maturés pour donner les protéines structurales et les protéines à activité enzymatique. Le précurseur Env donne les protéines de l'enveloppe. Le symbole * indique la myristylation de la molécule. Les flèches rouges indiquent les sites de protéolyse.



Dans le cytoplasme, l'ARN du virus HIV-1 est converti en ADNc double-brin par la transcriptase inverse. Des protéines virales s'assemblent ensuite autour de cet ADNc pour former le complexe de préintégration, composé de la nucléocapside, de protéines de la matrice, de la protéine Vpr, de la transcriptase inverse et de l'intégrase. Après transport de ce complexe dans le noyau de la cellule hôte, l'ADN viral est intégré dans le génome cellulaire grâce à l'action de l'intégrase, de façon semi-aléatoire, au niveau de sites actifs de transcription (Fig. III.5; pour revue, Tang *et al.* 1999). La transcription de l'ADN proviral conduit à la production d'un seul transcrit, qui a deux rôles : d'une part, sa maturation par épissage alternatif fournit les ARN messagers codant les protéines virales et d'autre part, ce long transcrit primaire peut être conservé intact pour servir d'ARNm pour le précurseur Gag-Pol et d'ARN génomique pour les nouveaux virions (Fig. III.5; pour revue, Tang *et al.* 1999). Les protéines virales nouvellement synthétisées s'assemblent autour de l'ARN géno-

mique pour former un nouveau virion, qui sort de la cellule par bourgeonnement. Après une étape de maturation, les particules virales sont prêtes à infecter de nouvelles cellules. Elles sont soit libérées dans le tissu, soit transmises directement à de nouvelles cellules, ce qui conduit à la fusion de cellules voisines entraînant la formation de syncytia.

III.1.5.2 Transcriptase inverse

La transcriptase inverse du virus HIV-1 est produite par maturation du précurseur Pol (Fig. III.4 ; pour revue, Harrich & Hooker 2002). Elle a une activité ARN dépendante et ADN dépendante. Elle synthétise une copie ADN double-brin du dimère d'ARN génomique présent dans le virion, en utilisant comme amorce l'ARN^{Lys3} cellulaire. L'activité RNase H élimine ensuite l'ARN ayant servi de matrice pour la synthèse du 1^{er} brin d'ADN, ce qui permet alors la synthèse du brin d'ADN complémentaire. Cette polymérase 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 10000 nts (pour revue, Sleasman & Goodenow 2003).

A l'issue de l'étape de transcription inverse, le génome viral se présente sous forme d'une copie unique d'ADN double-brin, dont les 2 extrémités, nommées LTR¹¹, sont identiques (séquence U3-R-U5).

III.1.5.3 Protéine Vpr

La protéine Vpr¹², de 14 kDa, est présente aussi bien dans le virion que dans le cytoplasme des cellules infectées. Elle empêche la multiplication cellulaire en arrêtant la croissance des cellules infectées en phase G2 (pour revue, Bukrinsky & Adzhubei 1999) et participe à l'import nucléaire du complexe de préintégration, grâce au signal d'adressage nucléaire qu'elle contient (pour revue, Sherman & Greene 2002). Cette capacité de pénétration de l'ADNc viral dans le noyau permet au virus HIV-1 d'infecter des cellules ne se divisant pas.

III.1.5.4 Intégrase

L'accessibilité de l'ADN génomique au sein de la chromatine joue un rôle important dans le choix du site d'intégration de l'ADN viral dans le génome de la cellule hôte (pour revue, Tang *et al.* 1999). Cette intégration est catalysée par les activités de l'intégrase (pour revue, Craigie 2001) : par son activité exonucléase, elle élimine 2 résidus sur le brin 3' à chaque extrémité de l'ADN proviral. Par une réaction de transestérification, elle clive de manière asymétrique les 2 brins de l'ADN génomique de la cellule hôte et assure l'insertion de l'ADNc proviral. Un ADN double-brin parfait est ensuite obtenu par action des enzymes cellulaires de réparation de l'ADN.

III.1.5.5 Protéine Tat

La protéine Tat¹³ est essentielle pour la réplication du virus HIV-1, car elle active d'environ 100 fois la transcription des gènes viraux, en agissant au niveau de l'étape d'élongation (pour revue, Karn 2000).

¹¹ « long terminal repeat »

¹² « viral protein R »

¹³ « transactivator »

2000 ; Westendorp *et al.* 1995).

La protéine Tat, libérée par les cellules, a la capacité de pénétrer dans d'autres cellules, ce qui peut conduire à l'apoptose massive des cellules lymphoïdes, caractéristique de la pathologie SIDA (pour revue, Karn 2000). Enfin, Tat régule l'expression de nombreux gènes cellulaires (impliqués dans la transduction du signal, la prolifération, la différenciation, la traduction, la transcription, . . . ; De la Fuente *et al.* 2002 ; Gibellini *et al.* 2002).

Il existe une autre voie d'activation de la transcription de l'ADN proviral, médiée par le facteur NF κ B, qui permet de contrôler la phase de latence du virus HIV-1. En effet, alors que dans les cellules qui se répliquent rapidement, l'ADN proviral est transcrit aussitôt, certains lymphocytes T infectés par le virus entrent en latence après l'intégration de l'ADN proviral dans leur génome. Celui-ci n'est donc pas transcrit, tant que la cellule n'a pas reçu de signal mitogène. Lorsque les lymphocytes T infectés sont activés, les provirus latents qu'elles contenaient sont transférés directement aux cellules voisines, sans que des particules virales neutralisables soient libérées dans les tissus (pour revue, Karn 2000). Cette capacité à infecter des cellules de façon latente permet au virus HIV-1 d'établir une infection persistante, malgré la réponse immunitaire dirigée contre les protéines virales.

Dans les lymphocytes T latents, le facteur NF κ B est séquestré dans le cytoplasme par le facteur I κ B α . Lorsque ces lymphocytes T sont activés par un antigène ou un signal mitogène, le facteur NF κ B est libéré, du fait de la phosphorylation et de l'ubiquitinylation du facteur I κ B α , suivies de sa dégradation. Le facteur NF κ B libre rejoint le noyau de la cellule, où il peut activer la transcription à partir d'un grand nombre de promoteurs, dont celui du virus HIV-1. La transcription virale est activée aussi bien au niveau de l'étape d'initiation (pour revue, Karn 2000), qu'au niveau de l'étape d'élongation (West *et al.* 2001). Enfin, comme la protéine Tat, NF κ B recrute des facteurs à activité acétylase, afin que la structure de la chromatine soit remodelée au niveau du promoteur LTR.

Cette activation de la transcription de l'ADN proviral par NF κ B conduit à la production de faibles quantités de protéine Tat, qui active à son tour la transcription virale de façon très efficace.

III.1.5.6 Protéine Rev

La protéine régulatrice Rev¹⁵ est une protéine de 19 kDa, produite à partir d'ARN messagers fortement épissés. Cette protéine joue un rôle primordial pour l'expression des gènes du virus HIV-1, car elle permet le passage de la phase précoce d'infection, caractérisée par la production des ARNm multiples épissés codant les protéines Tat, Rev, Nef, à la phase tardive d'infection, au cours de laquelle les ARNm sont peu épissés et permettent la production des protéines structurales et auxiliaires (Fig. III.7). Rev agit en reconnaissant un élément structuré de l'ARN viral, l'élément RRE¹⁶, qui contient un site de fixation de haute affinité pour la protéine (Malim *et al.* 1989 ; Cook *et al.* 1991 ; Malim & Cullen 1991 ; Mann *et al.* 1994 ; Charpentier *et al.* 1997 ; pour revue, Hope, 1999).

Les ARN pré-messagers cellulaires contenant des introns encore non épissés sont maintenus dans le noyau par interaction avec les facteurs d'épissage, jusqu'à ce qu'ils soient complètement épissés ou dégradés. Dans les cellules infectées, le caractère sous-optimal des sites accepteurs d'épissage de l'ARN du virus HIV-1 contribue à ralentir l'assemblage des spliceosomes autour des introns. La pro-

¹⁵ « regulator of virion expression »

¹⁶ « Rev response element »

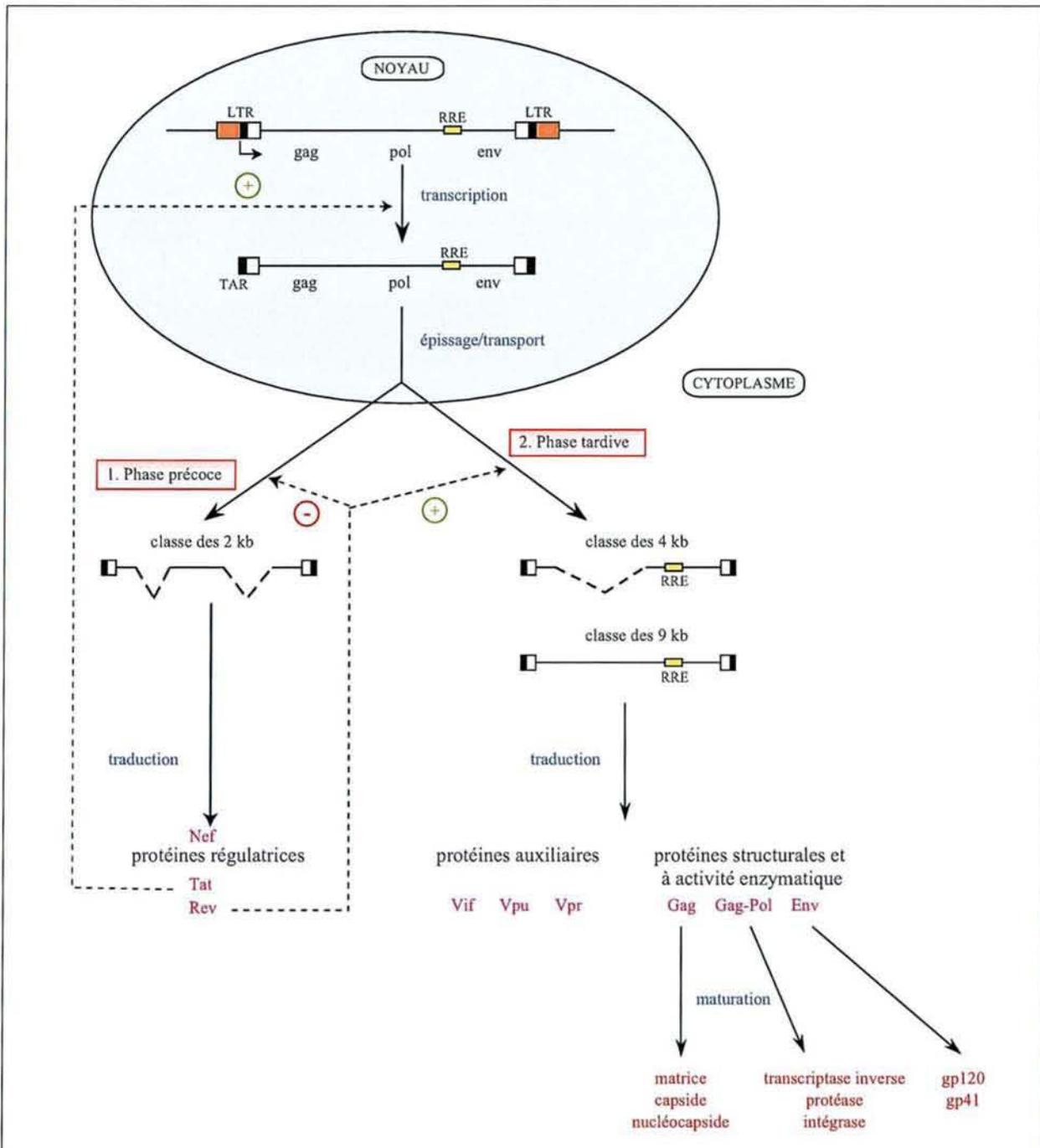
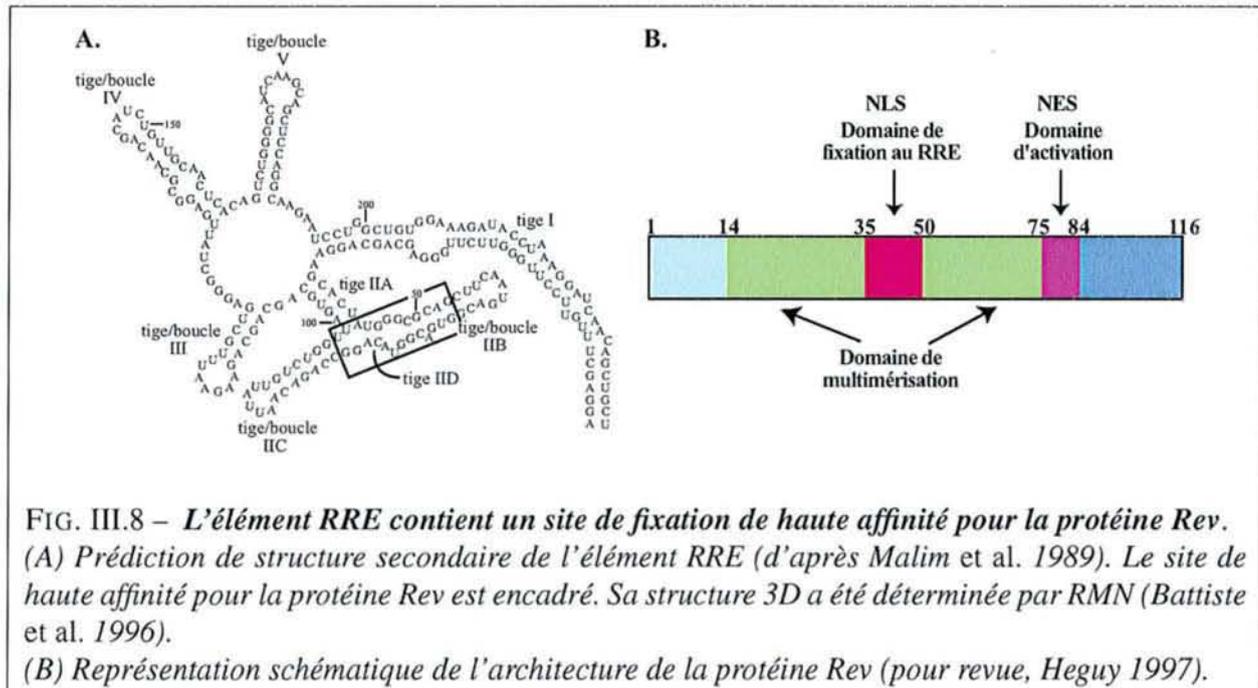


FIG. III.7 – Les étapes de l'expression des gènes du virus HIV-1 (pour revues, Stutz & Rosbash 1998 ; Hope 1999.)

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 cette phase, les protéines structurales, à activité enzymatique et auxiliaires sont produites.

teine Rev peut alors se fixer à l'élément RRE des ARNm non épissés ou épissés une seule fois et leur permettre d'échapper à ce phénomène de rétention, en favorisant leur transport du noyau vers le cytoplasme (Fig. III.7 ; pour revues, Stutz & Rosbash 1998 ; Hope 1999).

Complexe Rev/RRE



L'interaction entre la protéine Rev et l'élément RRE est primordiale pour que le virus HIV-1 se réplique efficacement (pour revue, Stutz & Rosbash 1998). L'élément RRE est très structuré, il renferme plusieurs tige-boucles, dont la IIB et la IIC, qui constituent le site de fixation de haute-affinité de la protéine Rev (Fig. III.8-A ; Malim *et al.* 1989 ; Cook *et al.* 1991 ; Charpentier *et al.* 1997). Lorsque la concentration en protéine Rev augmente dans la cellule infectée, la protéine se fixe à son site de haute affinité, puis multimérise sur l'élément RRE, par fixation coopérative des monomères le long de la structure tige-boucle I du RRE (Malim & Cullen 1991 ; Mann *et al.* 1994 ; pour revue, Stutz & Rosbash 1998).

Des analyses mutationnelles de la protéine Rev ont mis en évidence plusieurs domaines nécessaires à sa fonction : un domaine N-terminal, basique, responsable de la fixation de la protéine à l'élément RRE et de sa localisation nucléaire, flanqué de part et d'autre d'un domaine permettant la multimérisation, et enfin, un domaine C-terminal appelé domaine d'activation, qui agit comme signal d'export nucléaire (Fig. III.8 ; pour revue, Stutz & Rosbash 1998).

Deux fonctions de la protéine Rev découleraient de sa fixation à l'élément RRE : au niveau du transport des ARN et de l'inhibition de l'épissage. S'il a été clairement démontré que la fixation de la protéine Rev à l'élément RRE favorise le transport des ARN peu ou pas épissés dans le cytoplasme, son implication dans l'inhibition de l'épissage est par contre plus controversée.

Rôle possible de la protéine Rev dans l'inhibition de l'épissage

Des données en faveur d'un rôle de la protéine Rev dans l'inhibition de l'épissage ont été obtenues *in vitro*, non pas avec la protéine Rev entière, mais avec le peptide basique se liant au site de forte affinité (Kjems *et al.* 1991 ; Kjems & Sharp 1993). D'autres études ont été réalisées *ex vivo*, dans un contexte hétérologue (intron de la β -globine ; Chang & Sharp 1989). Ces études ont montré qu'un épissage inefficace est nécessaire pour la fonction de la protéine Rev et il a été proposé qu'un peptide, renfermant le domaine de fixation à l'ARN de la protéine Rev, limiterait la fixation de la particule snRNP U2 et bloquerait l'entrée de la tri-snRNP [U4/U6,U5] lors de l'assemblage du spliceosome *in vitro* (Chang & Sharp 1989 ; Kjems *et al.* 1991 ; Kjems & Sharp 1993).

Néanmoins, depuis 10 ans, aucune nouvelle étude n'est venue confirmer ces données, qui ne permettent pas de conclure définitivement sur un rôle de Rev dans l'épissage. En effet, aucune interaction directe entre la protéine Rev et des composants spliceosomaux n'a été mise en évidence. Il a été proposé que Rev entraîne un blocage de l'assemblage des complexes spliceosomaux, par l'interaction de son domaine basique avec la protéine cellulaire p32, une protéine copurifiée avec la protéine SR ASF/SF2 (cf § II.7.4.2 ; Kjems & Sharp 1993 ; Tange *et al.* 1996). Rev pourrait ainsi titrer ASF/SF2 par l'intermédiaire de p32 (Powell *et al.* 1997). Il semble en plus que ASF/SF2, liée à p32, interagisse avec la structure tige-boucle III de l'élément RRE, en présence de la protéine Rev (Powell *et al.* 1997). La formation d'un complexe composé de Rev, RRE, ASF/SF2 et p32 bloquerait l'assemblage des complexes spliceosomaux (Kjems & Sharp 1993 ; Tange *et al.* 1996). Néanmoins, il s'agit essentiellement d'hypothèses qui doivent d'être confirmées.

Rôle de la protéine Rev dans le transport

Au contraire, le rôle de Rev dans le transport des ARN viraux a été clairement établi. La multimérisation de la protéine Rev sur l'élément RRE est indispensable pour l'export des ARN. Lorsque la protéine est présente en faible quantité dans la cellule infectée, elle ne se fixe qu'à son site de haute affinité dans l'élément RRE. Puis, en phase tardive d'infection, lorsque de grandes quantités de la protéine ont été synthétisées, elle multimérise le long de l'élément RRE, ce qui démasque le signal d'export nucléaire NES de la protéine. Il interagit alors avec le système d'export CRM1/Ran GTP (Cmarko *et al.* 2002 ; pour revue, Stutz & Rosbash 1998).

D'autres éléments de séquence semblent importants pour le transport par Rev des ARN non épissés. En effet, l'introduction de l'élément RRE dans l'intron de l'ARN prémessager de la β -globine ne suffit pas à Rev pour stimuler l'export des ARN non épissés (Najera *et al.* 1999). Des séquences additionnelles, comme l'élément d'instabilité INS¹⁷ contenu dans l'ORF Gagp17, restaurent l'export par la protéine Rev d'une grande partie des ARN non épissés. Les éléments introniques INS, encore appelés CRS¹⁸, sont des séquences riches en A et U, réparties sur toute la longueur du génome viral, dans les gènes *env*, *pol*, *gag* et *vif* (Rosen *et al.* 1988 ; Cochrane *et al.* 1991 ; Maldarelli *et al.* 1991 ; Olsen *et al.* 1992 ; Schwartz *et al.* 1992 ; Nasioulas *et al.* 1994 ; Huffman & Arrigo 1997).

Ces éléments pourraient intervenir dans la localisation des ARN viraux dans un compartiment spéci-

¹⁷ « instability sequence »

¹⁸ « cis-acting repressor sequence »

fique du noyau, ce qui éviterait la dégradation des ARN non épissés. Il a été proposé que la protéine hnRNP A1 joue un rôle important dans ce processus (Najera *et al.* 1999).

III.1.5.7 Protéines structurales et auxiliaires intervenant dans l'assemblage et la libération des virions

Protéine Gag

Le produit du gène *gag*¹⁹ est une protéine de 55 kDa, exprimée en très grandes quantités dans les cellules infectées à partir de l'ARNm viral non épissé. Lors de la traduction, cette protéine est myristylée en N-terminal, ce qui permet son ancrage dans la membrane plasmique de la cellule, au site d'assemblage du nouveau virion (pour revue, Cimarelli & Darlix 2002).

Après libération de la particule virale, la protéine Gag est maturée par le produit du gène *pol*, la protéase, en 4 protéines : MA, CA, p6, NCp7 et 2 peptides : p1 et p2 (Fig. III.4).

- MA (p17) correspond à l'extrémité N-terminale de la protéine Gag, elle porte donc également un groupe myristyle. Dans le virion, elle reste associée à la face interne de la bicouche lipidique sous forme d'un trimère, qui recrute les protéines de l'enveloppe (pour revue, Tang *et al.* 1999). Un ensemble de protéines MA est également présent dans le complexe de préintégration qui escorte le génome viral vers le noyau (Gallay *et al.* 1995). En effet, la protéine MA porte un signal reconnu par la machinerie d'import nucléaire des protéines (Bukrinsky *et al.* 1993).
- CA (p24) s'associe en multimère pour constituer le cœur protéique de la particule virale. Au sein du précurseur Gag, elle interagit avec une protéine cellulaire, la cyclophiline A, ce qui explique la présence de cette protéine dans la particule virale (Franke *et al.* 1994 ; Thali *et al.* 1994). Cette interaction est essentielle pour la réplication virale (Franke & Luban 1996).
- La protéine de nucléocapside ou NCp7 s'associe à l'ARN génomique dimérique, qu'elle recouvre entièrement pour former une structure nucléoprotéique semblable à la chromatine, et dont elle favorise l'encapsidation (pour revue, Darlix *et al.* 2001). En plus de son rôle structural, cette protéine joue un rôle important dans la réplication virale : elle a des propriétés de protéine chaperon qui lui permettent d'intervenir lors de la transcription inverse, de l'intégration de l'ADN proviral dans le génome de la cellule hôte et dans la transcription, au cours de laquelle la protéine favorise la production des ARN précoces (Zhang & Crumpacker 2002 ; pour revue, Darlix *et al.* 2001). La protéine reconnaît également le signal d'encapsidation sur l'ARN viral. Ce signal est localisé en aval du site majeur D1 et permet l'incorporation de l'ARN génomique dans le virion (Harrison & Lever 1992). La présence de ce signal dans un intron assure l'intégration de molécules d'ARN non épissées dans la particule virale.
- p6 est une protéine ubiquitinyllée qui intervient dans le bourgeonnement de la particule virale (Strack *et al.* 2000).

Protéine Env

La protéine Env²⁰ est exprimée à partir d'un ARN simplement épissé. A la différence des autres protéines virales, traduites par des ribosomes cytoplasmiques, la traduction de Env est réalisée par des

¹⁹ « group specific antigen »

²⁰ « enveloppe »

ribosomes associés au reticulum endoplasmique. La protéine est ensuite maturée dans le reticulum endoplasmique, au sein duquel elle est glycosylée et clivée par une protéase cellulaire en 2 polypeptides : gp120 et gp41 (Fig. III.4). Ces 2 protéines restent associées et, à travers le Golgi, elles joignent la membrane plasmique pour être incorporées au virion. La partie gp41 correspond au domaine transmembranaire de la protéine Env, tandis que gp120 est localisée à la surface de la particule virale.

Lors de l'infection d'une cellule par le virus HIV-1, l'interaction entre le récepteur cellulaire CD4 et le virus est assurée par des domaines spécifiques de la protéine gp120. Cette protéine possède également des régions dites hyper-variables, dont les séquences varient fortement suivant les isolats. L'une d'entre elles, la boucle V3, n'est pas impliquée dans la fixation au récepteur CD4, mais est responsable du tropisme du virus HIV-1 pour les lignées cellulaires lymphoïdes ou les macrophages. En effet, suivant les souches virales, cette boucle interagit avec le 2^e récepteur cellulaire localisé en surface des lymphocytes et macrophages, CXCR4 ou CCR5 (pour revue, Tang *et al.* 1999). La protéine gp120 interagit par ailleurs avec une protéine cellulaire localisée à la surface des cellules dendritiques, ce qui permet le transport du virion par ces cellules, notamment des muqueuses vers les tissus lymphoïdes.

Protéine Vif

Vif²¹ est une phosphoprotéine de 23 kDa présente en quantités abondantes dans le cytoplasme des cellules infectées, mais également incorporée dans les virions (pour revue, Cimarelli & Darlix 2002). Le rôle de cette protéine n'est pas connu précisément. L'absence de protéine Vif conduit à la production de virions moins infectieux, dont la stabilité, ainsi que l'étape de transcription inverse, est affectée. La protéine Vif pourrait avoir un rôle de chaperon de l'ARN viral, auquel elle s'associe, avec la protéine de nucléocapside. Cette interaction permettrait de protéger l'ARN génomique de l'action d'un inhibiteur endogène, non encore identifié, qui pourrait constituer un mécanisme de défense de la cellule hôte, dirigé contre l'ARN viral (pour revue, Cimarelli & Darlix 2002). La protéine Vif est supposée également empêcher l'activité anti-virale de l'enzyme APOBEC3G²² en empêchant son encapsidation. Lors de la transcription inverse, APOBEC3G catalyse des conversions C→U dans l'ADNc proviral (Mangeat *et al.* 2003 ; Zhang *et al.* 2003 ; Harris *et al.* 2003 ; Lecossier *et al.* 2003), ce qui se traduit par des conversions G→A dans le transcrit primaire après la transcription.

Protéine Vpu

Au sein des lentivirus, la protéine Vpu²³ n'est retrouvée que chez dans les virus HIV-1 et le virus SIV infectant les chimpanzées (SIVcpz). Cette protéine membranaire de 16 kDa participe à la destruction des récepteurs CD4⁺ dans le reticulum endoplasmique et elle favorise le relargage des virions à partir de la membrane plasmique de la cellule infectée, mais elle n'est pas incorporée dans les virions matures (pour revue, Cimarelli & Darlix 2002). Vpu est exprimée à partir du même ARN messager que la protéine Env. Etant donné que le signal d'initiation de la traduction de Vpu est nettement plus

²¹ « viral infectivity factor »

²² « apolipoprotein B m-RNA editing enzyme, catalytic polypeptide-like 3G », encore appelée CEM15

²³ « viral protein U »

faible que celui de Env, la quantité de protéine Vpu est 10 fois plus faible que celle de Env (Schwartz *et al.* 1990).

Protéase virale

En clivant les précurseurs Gag et Gag-Pol, la protéase virale est responsable de profonds changements dans la structure du virion, qui conduisent à la production d'une particule virale mature et infectieuse. Compte-tenu de son rôle majeur dans la maturation du virus, cette enzyme est donc une cible pour les agents antiviraux. La protéase virale agit sous forme de dimère et elle est synthétisée dans la cellule infectée sous forme d'un précurseur Gag-Pol, généré par un décalage de lecture (Fig. III.4). Les ribosomes en cours de traduction rencontrent un motif pseudo-nœud situé en 3' de l'ARN *gag*. Ce motif et l'existence d'une séquence UUUUU conduisent à une fréquence de décalage du cadre de lecture d'environ 5 à 10% permettant l'expression de l'ORF Pol (Parkin *et al.* 1992).

III.1.5.8 Protéine Nef

La protéine Nef²⁴ est très fortement exprimée en phase précoce d'infection, c'est d'ailleurs la 1^{re} protéine à s'accumuler à des niveaux détectables, suite à l'infection d'une cellule par le virus HIV-1 (pour revue, Arora *et al.* 2002). C'est une protéine de 27 kDa, myristylée, codée par un seul exon qui inclut une partie de l'extrémité 3'LTR du génome viral.

La protéine Nef joue un double rôle lors de l'infection par le virus HIV-1 : en activant la production de FasL à la surface des cellules infectées, elle contribue à l'apoptose des lymphocytes T CD4⁺ voisins (Xu *et al.* 1997 ; Zauli *et al.* 1999), tout en protégeant la cellule infectée de l'apoptose, en interagissant avec la protéine kinase ASK1, nécessaire à la transduction du signal apoptotique (Geleziunas *et al.* 2001). Nef protège également la cellule infectée de l'attaque des lymphocytes T cytotoxiques, en réduisant le nombre de récepteurs CD4⁺ et de complexes majeurs d'histocompatibilité (MHC I) sur sa surface (pour revues, Tang *et al.* 1999 ; Arora *et al.* 2002). Le rôle de ces complexes est normalement de présenter des épitopes viraux à la surface des lymphocytes T cytotoxiques. Comme Tat, cette protéine contribue donc à la pathologie SIDA.

Nef possède également des propriétés d'activation de la transcription de gènes cellulaires jouant un rôle important dans la réplication virale, comme la protéine CDK9 ou encore le facteur d'activation de la transcription SF1 (pour revue, Arora *et al.* 2002).

III.2 Régulation de l'épissage de l'ARN du virus HIV-1

III.2.1 Production des ARN messagers viraux

Comme nous l'avons souligné plus tôt, toutes les protéines du virus HIV-1 sont produites à partir d'un même transcrit. Leur taux relatif d'expression et leur ordre chronologique d'expression au cours du cycle viral dépendent donc d'étapes post-transcriptionnelles, comme l'étape d'épissage du transcrit primaire. Cette étape d'épissage est complexe : elle met en jeu 4 sites donneurs (D1, D2, D3, D4) et 8 sites accepteurs (A1, A2, A3, A4c, A4a, A4b, A5 et A7), dont l'utilisation combinée permet au virus

²⁴ « negative factor »

de produire une quarantaine d'ARN messagers, dont les ARN de la classe 2 kb, en phase précoce d'infection, et les ARN de la classe 4 kb, au cours de la phase tardive (Fig. III.9).

Les ARN messagers viraux contiennent des exons codants et des exons non codants. En plus de l'exon terminal délimité par le site D1 présent dans tous les ARNm, les 2 exons non codants, bordés par les couples de sites A1-D2 et A2-D3, peuvent être inclus dans n'importe quelle espèce d'ARN messenger, ce qui augmente de façon considérable le nombre d'ARN messagers pouvant être produits. (Fig. III.9) (Purcell & Martin 1993). Ces 2 exons ont la particularité de moduler le taux d'expression des protéines virales, en jouant sur la stabilité des ARNm. Ainsi, l'incorporation de l'exon 2 stabilise l'ARN messenger, alors que l'exon 3 est plutôt déstabilisateur. Cet exon pourrait contenir un élément d'instabilité INS (Krummheuer *et al.* 2001).

L'épissage de tous les ARNm viraux au site donneur majeur D1 assure l'élimination du signal d'encapsidation du génome viral, contenu dans la séquence intronique localisée en aval du site D1. Ainsi, seuls des ARN non épissés peuvent être incorporés dans les nouveaux virions. Le rôle primordial d'un épissage au site D1 est confirmé par l'observation de l'activation d'un site donneur cryptique localisé 4 nts en aval, lorsque le site D1 est altéré (Purcell & Martin 1993).

En phase précoce d'infection, les ARNm de 2 kb *nef* et *rev* sont les produits majoritaires (56% et 34% respectivement), alors que les ARN messagers *tat* et *vpr* sont présents en faibles quantités (9% et 1% respectivement ; Fig. III.9 ; Purcell & Martin 1993). Quant aux ARNm de la phase tardive d'infection, ils contiennent tous l'intron D4-A7 qui renferme l'élément RRE, permettant leur export par la protéine Rev du noyau vers le cytoplasme. L'ARNm majoritaire de cette classe 4 kb est l'ARNm *env*, il constitue 95% des ARNm, tandis que les ARN messagers *tat*, *vpr* et *vif* ne sont produits qu'à 5% 2% et 1% respectivement.

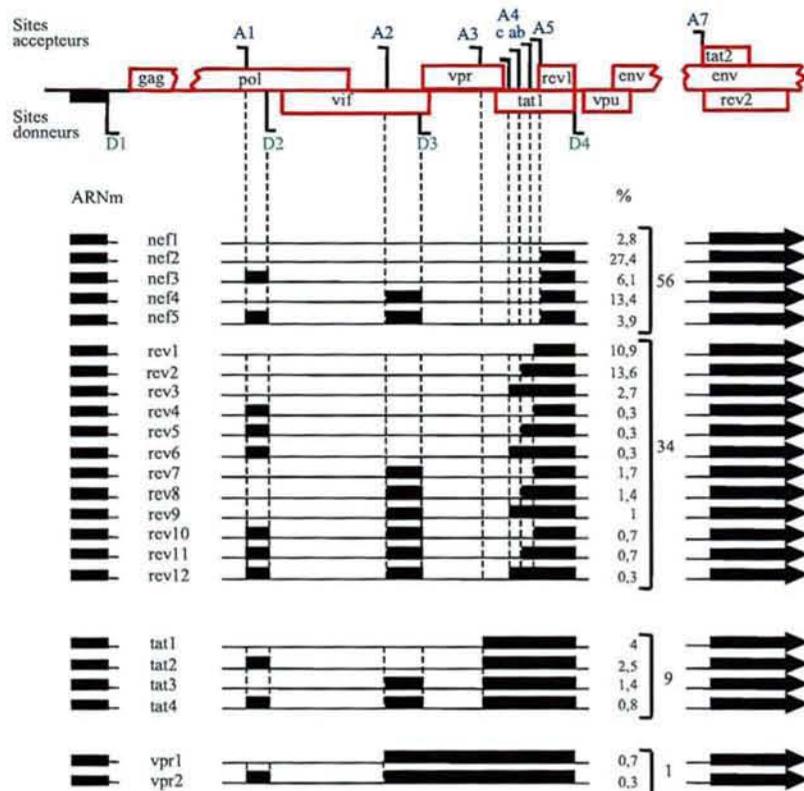
L'expression des protéines virales dépend en partie des quantités d'ARN messagers produits par épissage alternatif. Cette étape du cycle viral permet en outre de moduler l'activité de ces protéines et elle a un lien direct avec l'infectivité du virus HIV-1.

III.2.2 Impact de l'épissage sur l'activité des protéines virales et l'infectivité du virus

Plusieurs études sur la protéine Tat ont mis en évidence le rôle de l'épissage alternatif du transcrite primaire dans la régulation de la fonction des protéines virales.

Au sein du génome viral, les exons *tat1* et *tat2* codent 2 domaines fonctionnellement distincts de la protéine Tat. L'exon *tat1*, bordé par le couple de sites A3-D4, code un domaine de 72 acides aminés impliqué dans la transactivation de la transcription (Viglianti & Mullins 1988 ; Jones & Peterlin 1994 ; Jeang *et al.* 1999). L'exon *tat2* débute au site A7 et code, quant à lui, un domaine protéique de 14 à 32 acides aminés, suivant les souches virales (Neuveut & Jeang 1996). Ce domaine est impliqué en particulier dans la répression de la transcription du complexe majeur d'histocompatibilité de classe I (MHC I) et dans l'apoptose (Bartz & Emerman 1999b ; pour revues, Brown *et al.* 1998 ; Verhoef *et al.* 1998). Il est également important pour permettre à la protéine Tat extracellulaire d'être captée par les

(A) ARN de la classe 2 kb



(B) ARN de la classe 4 kb

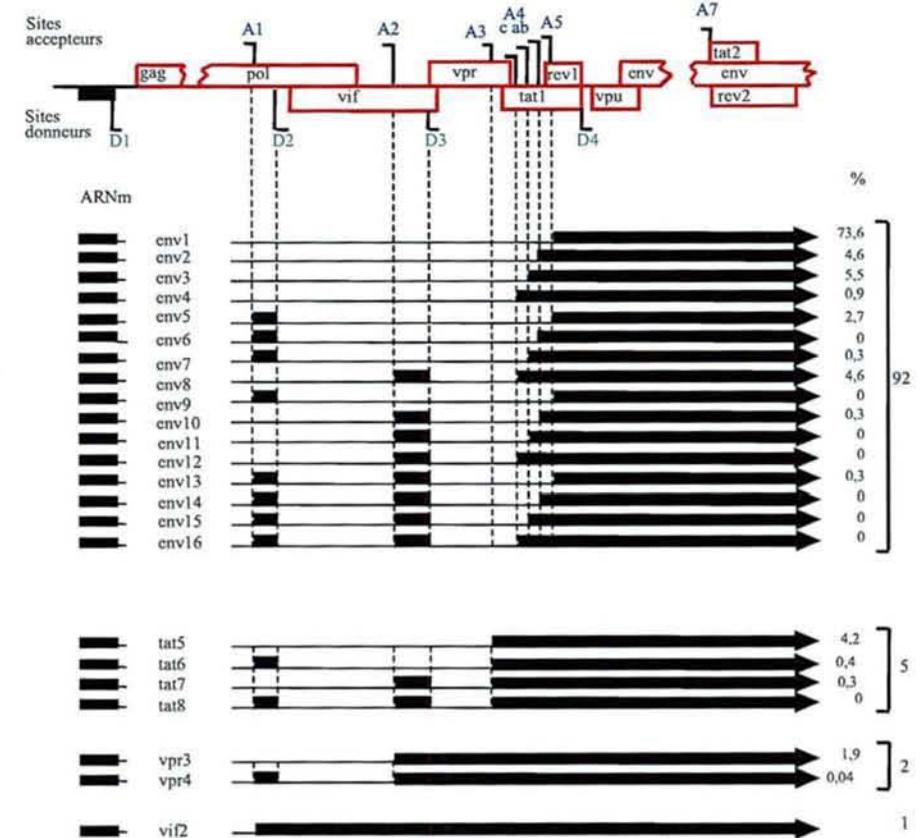


FIG. III.9 – Composition en exons et abondance relative des ARNm du virus HIV-1 (d'après Purcell & Martin 1993).

(A) ARNm de la classe 2 kb provenant d'un épissage multiple.

(B) ARNm de la classe 4 kb obtenus par épissage unique.

Les sites d'épissage ainsi que les protéines codées sont indiquées dans la partie supérieure du schéma représentant l'organisation génomique du virus HIV-1. Les boîtes noires correspondent aux exons constituant l'ARNm. Les proportions de chaque ARNm sont données.

cellules neuronales (Ma & Nath 1997).

Les ARN messagers *tat* produits en phase précoce d'infection contiennent les 2 exons de la protéine Tat (ARNm *tat1* à *tat4*, Fig. III.9). L'utilisation d'un codon STOP localisé en aval du site A7 de ces ARNm permet donc de produire, suivant les souches virales, une isoforme de 86 à 104 acides aminés de la protéine Tat. L'exon *tat2* n'est cependant pas obligatoire pour qu'il y ait production d'une protéine Tat active en transcription. Ainsi, les ARNm *tat* produits en phase tardive d'infection renferment un codon STOP localisé en aval du site donneur D4 et expriment la forme de 72 acides aminés de la protéine Tat, codée par l'exon *tat1* (ARN *tat5* à *tat8*, Fig. III.9). Les protéines Tat exprimées au cours du cycle viral diffèrent donc par la présence ou non du domaine codé par l'exon 2. Par conséquent, on peut s'attendre à des variations d'activité de la protéine Tat lors des différentes phases du cycle viral. Une telle différence d'activité a été mise en évidence au niveau de la régulation de la transcription inverse. En effet, la forme à 2 exons de la protéine Tat réprime l'activité transcriptase inverse en fin de réplication virale, ce dont est incapable l'isoforme de Tat ne contenant qu'un seul exon. Ainsi, des virus HIV-1 exprimant une protéine Tat avec 2 exons présentent une activité de transcriptase inverse plus efficace et une cinétique de réplication plus rapide que les virus n'ayant que l'isoforme Tat-exon 1 (leur exon *tat2* a été délété ; Kameoka *et al.* 2001).

La régulation du taux de production de l'ARNm *tat* par épissage en fonction du type cellulaire joue un rôle important dans les capacités de prolifération du virus dans différents types cellulaires. Ce lien expliquerait notamment la différence de comportement du virus, suivant qu'il infecte des macrophages ou des lymphocytes T. En effet, l'infection des macrophages, qui servent de réservoir pour le virus HIV-1, est lente, chronique et non lytique, alors que l'infection des lymphocytes T est au contraire rapide, lytique et conduit à la production de nombreux virions. Peu de temps après l'infection de macrophages et de lymphocytes T par le virus HIV-1, les profils d'épissage des ARN viraux sont similaires dans les 2 types cellulaires. Par contre, après 14 jours d'infection, une forte diminution de la production d'ARN messagers *tat* et de l'activité de la protéine Tat dans les macrophages se traduit par une forte diminution de la production de virions (Sonza *et al.* 2002). L'ajout de protéine Tat exogène active au contraire la production de virions par les macrophages, confirmant que la différence du taux de production de Tat joue un rôle déterminant dans la définition du type d'infection (Sonza *et al.* 2002). D'une manière générale, l'étude de Sonza *et al.* montre que le contrôle de l'épissage de l'ARN viral, suivant le type cellulaire, conduit le virus à s'adapter aux types de cellules qu'il infecte, ceci en générant une infection de type chronique, ou au contraire, lytique.

III.2.3 Mécanismes d'épissage de l'ARN viral

III.2.3.1 Les sites accepteurs d'épissage du virus HIV-1 sont faibles

Plusieurs études ont en effet montré que les sites d'épissage du virus HIV-1 sont sous-optimaux, et que cette faible efficacité joue un rôle majeur dans la définition des ARNm produits (Purcell & Martin 1993 ; O'Reilly *et al.* 1995 ; Damier *et al.* 1997 ; Damier 1997 ; Dyhr-Mikkelsen & Kjems 1995).

Sachant que les proportions relatives des différents ARNm viraux produits sont identiques dans des lymphocytes T infectés par des particules virales et dans des cellules HeLa transfectées par un provi-

rus (Purcell & Martin 1993), les cellules HeLa ont été choisies pour étudier l'épissage de l'ARN du virus HIV-1, d'autant plus qu'elles sont déjà utilisées pour étudier l'épissage des ARN pré-messagers cellulaires. Deux approches expérimentales, l'une, *ex vivo*, l'autre, *in vitro*, sont généralement employées pour étudier l'efficacité des sites d'épissage des ARN viraux. Ces 2 démarches ont également été utilisées lors de cette étude, en collaboration avec les équipes de J. Stévenin (IGBMC, Strasbourg) et M. Stoltzfus (Université d'Iowa, Etats-Unis).

L'approche *ex vivo* développée par M. Stoltzfus consiste à transfecter des cellules HeLa par un plasmide contenant l'ADNc proviral, portant l'ensemble des sites d'épissage du virus et tronqué dans l'intron D1-A1 au niveau du gène *gag-pol*, ceci afin d'empêcher la production de particules virales (Jacquet *et al.* 2001b). Les ARNm produits sont alors analysés par l'approche RT-PCR semi-quantitative (Purcell & Martin 1993).

Dans l'approche *in vitro*, il est impossible pour des raisons techniques d'utiliser l'ARN viral en entier. Pour cela, des fragments de l'ARN viral, renfermant un ou plusieurs couples de sites d'épissage, avec souvent des introns tronqués, sont utilisés.

En plus de tests réalisés sur des couples de sites donneurs et accepteurs du virus HIV-1, des études visant à mesurer la force relative des sites donneurs ont été réalisées en fusionnant chacun de ces sites avec un site accepteur plus fort, soit de l'adénovirus (expériences d'épissage *in vitro* réalisées au laboratoire par L. Damier 1997), soit de la β -globine (expériences d'épissage *ex vivo* réalisées par l'équipe de K. Beemon, O'Reilly *et al.* 1995). Le même type d'expérience a été réalisé pour mesurer la force relative des sites accepteurs (O'Reilly *et al.* 1995 ; Damier 1997).

Les données obtenues sont en accord avec l'étude globale réalisée par Purcell & Martin (1993). Les sites donneurs D1 et D4 se sont révélés les plus efficaces en épissage, dans toutes les approches utilisées (O'Reilly *et al.* 1995 ; Damier 1997). Cette efficacité est corrélée à la force attendue sur la base de leur complémentarité avec les snRNA U1 et U6. Ceci explique pourquoi ils sont préférentiellement utilisés (en particulier, le site D1 dont l'utilisation est systématique). En ce qui concerne les sites D2 et D3, la limitation d'utilisation pourrait être la conséquence d'un niveau d'appariement plus faible avec le snRNA U6 (O'Reilly *et al.* 1995 ; Damier 1997).

Les différences de force intrinsèque des sites donneurs de l'ARN du virus HIV-1 semblent donc suffisantes pour expliquer leur utilisation lors de la réaction d'épissage, suggérant que ces sites 5' ne sont pas soumis à une régulation complexe, comme c'est le cas des sites accepteurs.

L'analyse des différents ARNm viraux générés dans des lymphocytes T infectés ou dans des cellules HeLa transfectées montre que l'utilisation des différents sites accepteurs est très variable (Fig. III.9). Le site A5 est ainsi utilisé de manière prédominante. Le site A7 est, lui-aussi, fortement utilisé et toujours couplé au site D4, tandis que les autres sites accepteurs sont beaucoup moins utilisés (Purcell & Martin 1993).

L'ensemble des sites accepteurs d'épissage présentent des séquences peu conformes au consensus des sites 3' d'épissage. Néanmoins, le degré de dégénérescence des séquences n'est pas corrélé au taux d'utilisation. En effet, les sites A5 et A7, qui sont les plus utilisés, renferment des séquences polypyrimidines sous-optimales. De plus, les analyses de l'efficacité relative des sites d'épissage *in vitro* montrent que A1 et A2 sont les sites les plus performants.

Les travaux menés par S. Jacquenet au laboratoire ont montré que dans le cas du site A3, par contre, la séquence et la structure secondaire dans laquelle est impliquée la séquence polypyrimidine sont directement responsables de la faible utilisation de ce site. En effet, l'optimisation de cette séquence augmente fortement l'efficacité d'utilisation de ce site (Jacquenot *et al.* 2001b). Le site A7 possède quant à lui 3 points de branchements, dont un correspond à une uridine à la place de l'adénosine habituellement retrouvée à cette position (Dyhr-Mikkelsen & Kjems 1995 ; Damgaard *et al.* 2002). Un nucléotide non canonique de branchement est également utilisé comme point de branchement au site A2 (Damier *et al.* 1997). En ce qui concerne les sites A4a, A4b, A4c et A5, des points de branchement multiples sont utilisés, qui chevauchent bien souvent d'autres sites d'épissage, ce qui réduit leur efficacité d'utilisation par encombrement stérique (Swanson & Stoltzfus 1998).

De même que pour les ARN pré-messagers cellulaires, des éléments régulateurs en *cis* et/ou en *trans* contrôlent le choix de ces sites d'épissage en compétition.

III.2.3.2 Régulation du choix des sites d'épissage

Comme le virus HIV utilise la machinerie spliceosomale de la cellule infectée pour permettre l'épissage de son transcrite primaire, les mécanismes de régulation de l'épissage sont similaires à ceux des ARN pré-messagers cellulaires. Plusieurs séquences régulatrices d'épissage ont été identifiées sur l'ARN viral en aval des sites A2, A3, A6, A7 et en amont du site D4. A7 est le site pour lequel le plus de données sont disponibles.

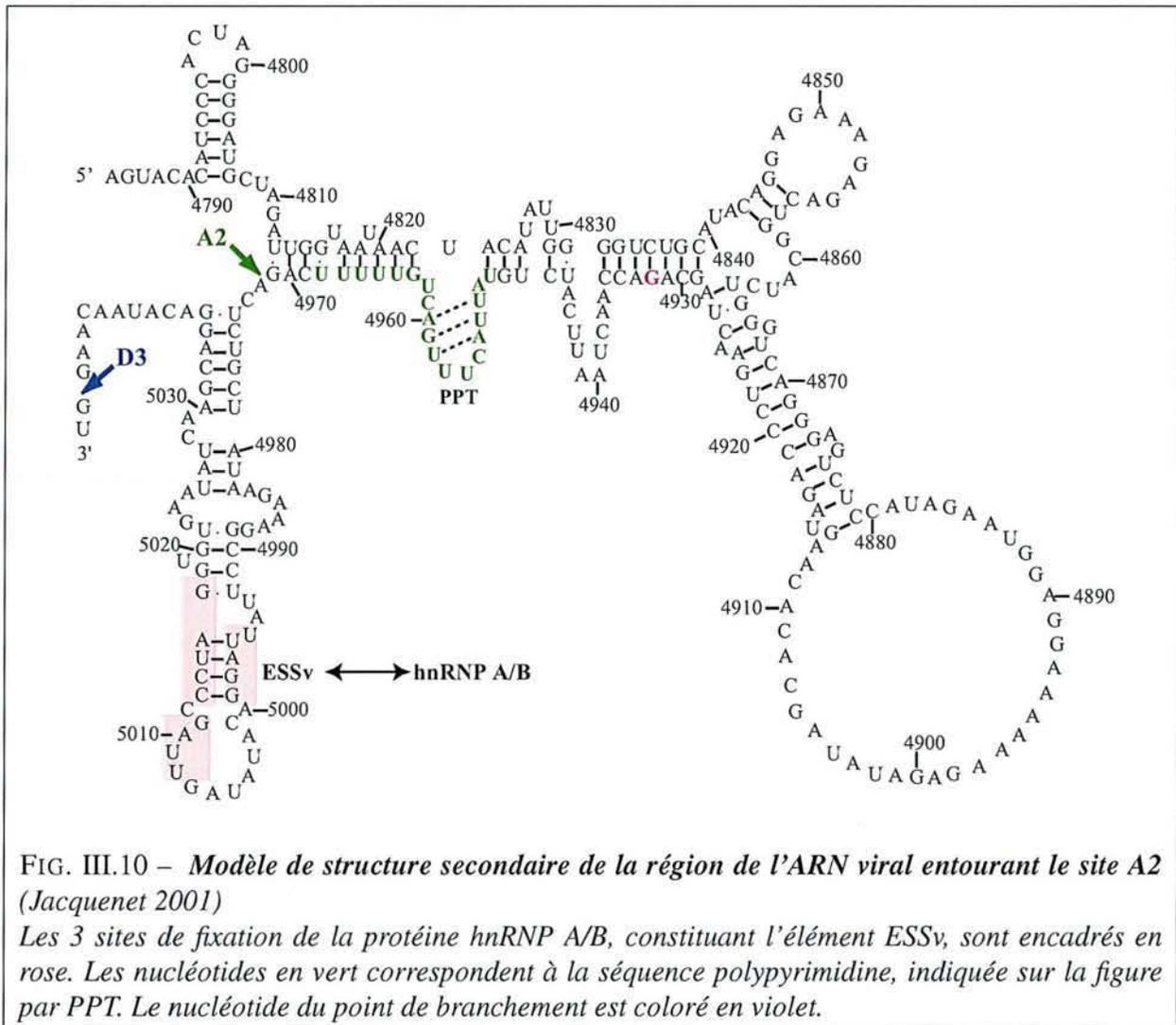
Site A2

Ce site est utilisé pour produire les ARNm *vpr* et il borde également l'exon non-codant délimité par les sites A2-D3. En plus de la présence d'un point de branchement non canonique (une guanosine au lieu d'une adénosine, Damier *et al.* 1997), la localisation du site A2 dans une région structurée de l'ARN peut également expliquer la faible abondance des messagers épissés à ce site, d'après la structure secondaire de cette région (Fig. III.10 ; Jacquenet 2001). Il apparaît que le site d'épissage, comme le point de branchement et la séquence polypyrimidine, est localisé dans une hélice, ce qui peut être limitant pour la fixation des facteurs spliceosomaux.

Un élément inhibiteur appelé ESSv a été récemment mis en évidence en aval du site A2 (Fig. III.10 ; Bilodeau *et al.* 2001). Il est constitué de 3 séquences de type YAG, dont la mutation provoque *ex vivo* l'inclusion de l'exon non-codant A2-D3 dans tous les ARN messagers des classes 2 et 4 kb. Cet élément fixe un membre de la famille des protéines hnRNP A/B. En effet, l'utilisation d'un oligonucléotide biotinylé contenant l'ESSv permet la déplétion d'un extrait nucléaire en protéines hnRNP A/B. L'ajout de protéines de cette famille à l'extrait déplété restitue la répression de l'épissage *in vitro* d'un ARN pré-messager contenant l'élément ESSv (Bilodeau *et al.* 2001).

Sites A3 à A5

Ces sites sont mutuellement exclusifs au sein de l'ARN viral. Le site A3 est exclusivement utilisé pour la production de l'ARNm *tat*, les sites A4c, a et b servent à la production des messagers *rev* et le site A5 permet la production des messagers *env* et *nef*.



Comme pour le site A2, la structure secondaire de l'ARN autour du site A3 module l'efficacité d'épissage à ce site : A3 est localisé à l'extrémité apicale d'une tige-boucle séquestrant la séquence polypyrimidine (Fig. III.11). Au laboratoire, S. Jacquenet a montré que la déstabilisation de cette tige-boucle, qui rend la séquence polypyrimidine plus accessible aux facteurs spliceosomaux, augmente l'efficacité d'épissage au site A3 *in vitro* et *ex vivo* (Jacquenet *et al.* 2001b).

Le site A3 est sous contrôle de 2 éléments inhibiteurs. Le 1^{er}, ESS2, est constitué de 2 séquences CUAGA juxtaposées dans la souche HIV-1 BRU que nous avons principalement étudiée, mais une seule séquence CUAGA est présente dans différentes souches du virus HIV-1 et suffit à réprimer l'épissage (Amendt *et al.* 1994, 1995 ; Si *et al.* 1997 ; Bilodeau *et al.* 1999). L'élément ESS2 est capable de réprimer un site d'épissage hétérologue faible et la force de l'inhibition dépend de la distance entre l'élément ESS2 et le site d'épissage (Amendt *et al.* 1995 ; Si *et al.* 1997).

Le mécanisme d'action de l'élément ESS2 n'est pas connu, par contre, l'élément interfère avec la formation du complexe spliceosomal A et fixe une protéine de la famille hnRNP A/B (Del Gatto-Konczak *et al.* 1999 ; Caputi *et al.* 1999). L'utilisation d'un ARN contenant l'élément ESS2 permet la déplétion d'un extrait nucléaire en protéines hnRNP A/B (Caputi *et al.* 1999). Dans un tel extrait nucléaire, l'inhibition de l'épissage exercée par l'élément ESS2 est levée, mais elle peut être restaurée

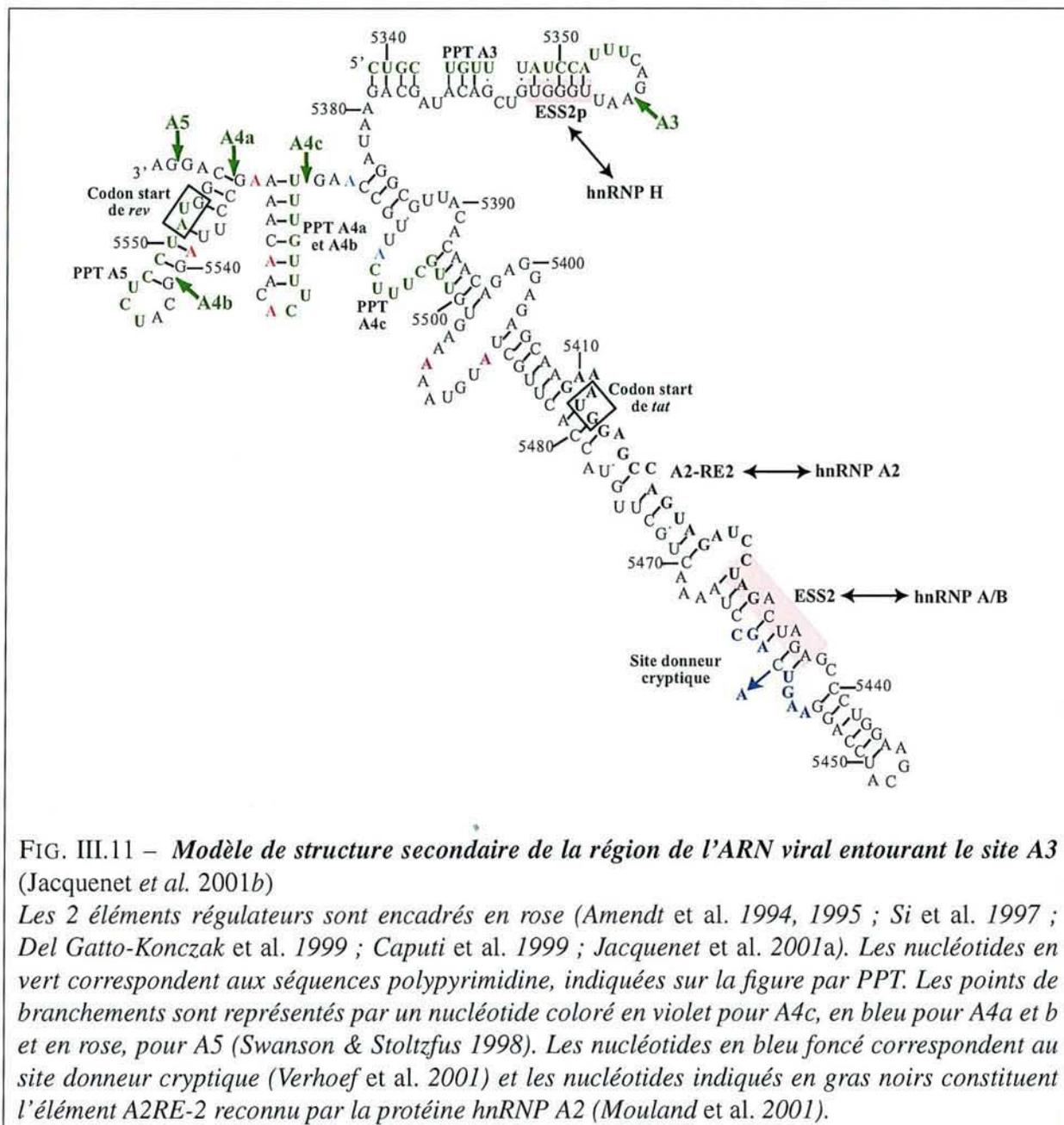


FIG. III.11 – *Modèle de structure secondaire de la région de l'ARN viral entourant le site A3* (Jacquet et al. 2001b)

Les 2 éléments régulateurs sont encadrés en rose (Amendt et al. 1994, 1995 ; Si et al. 1997 ; Del Gatto-Konczak et al. 1999 ; Caputi et al. 1999 ; Jacquenet et al. 2001a). Les nucléotides en vert correspondent aux séquences polypyrimidine, indiquées sur la figure par PPT. Les points de branchements sont représentés par un nucléotide coloré en violet pour A4c, en bleu pour A4a et b et en rose, pour A5 (Swanson & Stoltzfus 1998). Les nucléotides en bleu foncé correspondent au site donneur cryptique (Verhoef et al. 2001) et les nucléotides indiqués en gras noirs constituent l'élément A2RE-2 reconnu par la protéine hnRNP A2 (Mouland et al. 2001).

par ajout de protéines hnRNP A/B (Caputi et al. 1999).

L'élément ESS2 pourrait également réprimer l'utilisation d'un site d'épissage situé en aval. Cet effet inhibiteur a été mis en évidence lors de l'étude d'une souche virale portant une mutation dans le codon AUG de Rev (Verhoef et al. 2001). Après une longue culture de cette souche, l'apparition d'une 2^e mutation restaure la multiplication virale. Cette mutation ponctuelle, localisée dans *tat* active un site 5' d'épissage cryptique fort²⁵, localisé en aval de l'élément ESS2 (Fig. III.11). Ce site cryptique, utilisé en combinaison avec A5, conduit, après utilisation du codon AUG de *tat*, à la production d'une protéine de fusion entre Tat et Rev, appelée T-Rev. La délétion de l'élément ESS2 augmente l'efficacité d'épissage à ce site cryptique (Verhoef et al. 2001). En plus de contrôler l'épissage au site A3, l'élément ESS2 peut donc empêcher l'utilisation de sites d'épissage cryptiques.

²⁵ séquence sauvage : AAGUCAGC, séquence mutée : AA/GUAAGC, séquence consensus d'un site 5' : AG/GUAAGC. Le nucléotide muté est indiqué en gras, la barre oblique représente la jonction intron-exon.

Le second élément inhibiteur identifié, ESS2p²⁶, a été mis en évidence au laboratoire par S. Jacquenet et A. Méreau. Il est juxtaposé au site A3 qu'il contrôle (Fig. III.11), mais son déplacement en aval du site A2 réprime également l'épissage à ce site (Jacquenet *et al.* 2001a). Des expériences de pontages covalents aux UV suivies d'immunoprécipitation ont montré que la protéine hnRNP H se fixe à l'élément ESS2p.

A3 étant un site d'épissage faible, son utilisation nécessite la fixation du facteur U2AF³⁵ sur la jonction intron-exon, comme dans le cas des ARN pré-messagers cellulaires (Wu *et al.* 1999 ; Guth *et al.* 1999 ; Merendino *et al.* 1999) (cf § I.3.2.2). Compte-tenu de la proximité spatiale de l'élément ESS2p et du site A3, l'inhibition de l'épissage serait due à une compétition entre les protéines hnRNP H et U2AF³⁵ pour leur fixation sur l'ARN (Jacquenet *et al.* 2001a). Une simple mutation de l'élément ESS2p, UGGGU→AGGGU, crée un site de fixation de haute affinité pour la protéine hnRNP A1. Cela conduit à une répression encore plus forte de l'épissage, du fait de la plus grande abondance de protéine hnRNP A1 et de sa forte affinité pour le site créé (Jacquenet 2001). Cette donnée conforte l'hypothèse du modèle de répression de l'épissage par encombrement stérique, plutôt que par un effet médié spécifiquement par la protéine hnRNP H.

Un site de fixation de la protéine hnRNP A2 est également présent en aval du site A3. Cette protéine est impliquée dans des mécanismes de transport des ARN et plus particulièrement, dans la localisation cytoplasmique des ARNm du virus HIV-1 dans les oligodendrocytes (Mouland *et al.* 2001 ; pour revue, Carson *et al.* 2001).

Au total, deux sites de fixation de la protéine ont été localisés sur l'ARN du virus HIV-1. La 1^{re}, A2 RE-1, est nécessaire pour le transport de l'ARN *gag*, tandis que A2 RE-2 favorise le transport des messagers *vpr*, mais pas celui des messagers *tat* (Mouland *et al.* 2001). Cette séquence est pourtant localisée dans l'exon *tat1*, en aval du site A3, où elle chevauche partiellement l'élément ESS2 (nts 5409 à 5431). La présence de la séquence A2RE-2 à côté de l'élément ESS2 pourrait permettre un couplage entre l'épissage et la localisation des ARN viraux dans le cytoplasme, ce qui pourrait, par exemple, favoriser la localisation de l'ARN génomique viral ou des ARNm codant les protéines structurales, sur le site d'assemblage du virion (Mouland *et al.* 2001).

Comme nous avons pu le voir précédemment, la région des sites A4c, a, b et A5 est très complexe, puisqu'elle contient 8 points de branchement (Fig. III.11 ; Swanson & Stoltzfus 1998). Le site A4b réprimerait l'utilisation du site A5, puisque sa mutation renforce l'épissage au site A5 (Swanson & Stoltzfus 1998). Aucune autre séquence contrôlant l'utilisation de ces sites n'a pour l'instant été identifiée.

Récemment, une séquence régulatrice riche en purine, de type GAR, a été mise en évidence entre le site A5 et le site D4 (Kammler *et al.* 2001). La reconnaissance de cette séquence (appelée ESE2) par la protéine ASF/SF2 favorise la fixation de la particule snRNP U1 au niveau du site D4. Deux phénomènes en découlent : d'une part, la stabilité de l'ARNm *env* (épissé au site A5) est renforcée (Lu *et al.* 1990 ; Kammler *et al.* 2001), et d'autre part, l'épissage de l'intron D4-A7 est activé (Kammler *et al.* 2001).

²⁶ « ESS2 proximal »

Enfin, la présence du site donneur D4 a été décrite comme nécessaire à l'utilisation des sites A3, A4c,a, b et A5 dans une construction renfermant l'élément ESS2 (Amendt *et al.* 1994).

Site A6

Ce site d'épissage, ainsi que le site donneur D5, n'est présent que dans certaines souches virales comme HXB2, tandis que d'autres souches, comme pNL4-3 et BRU, en sont dépourvues (Kuiken *et al.* 2001). Le site A6 est localisé dans l'exon *env* et constitue, avec SD5, l'exon 6D. Cet exon est toujours exclus des ARNm viraux car son inclusion provoque la production d'ARNm aberrants, conduisant en particulier, à la synthèse d'une protéine chimérique Tat-Rev appelée Tev. Peu de protéine Rev fonctionnelle est alors produite, ce qui explique sans doute la faible efficacité de la réplication virale lorsque l'exon 6D est inclus (Wentz *et al.* 1997).

L'inclusion de l'exon 6D est contrôlée par 2 séquences régulatrices, ESS et ESE (Wentz *et al.* 1997). L'élément ESS renferme 3 motifs AUG consécutifs et une séquence GGGA, reconnue par la protéine hnRNP H, qui favorise la fixation de la snRNP U1, ce qui conduirait à la formation d'un spliceosome inactif (Caputi & Zahler 2002). L'élément ESE, riche en purines, a des propriétés activatrices dans un contexte hétérologue (Wentz *et al.* 1997), mais c'est en réalité une séquence inhibitrice d'épissage, qui agit en fixant des protéines SR (Caputi & Zahler 2002).

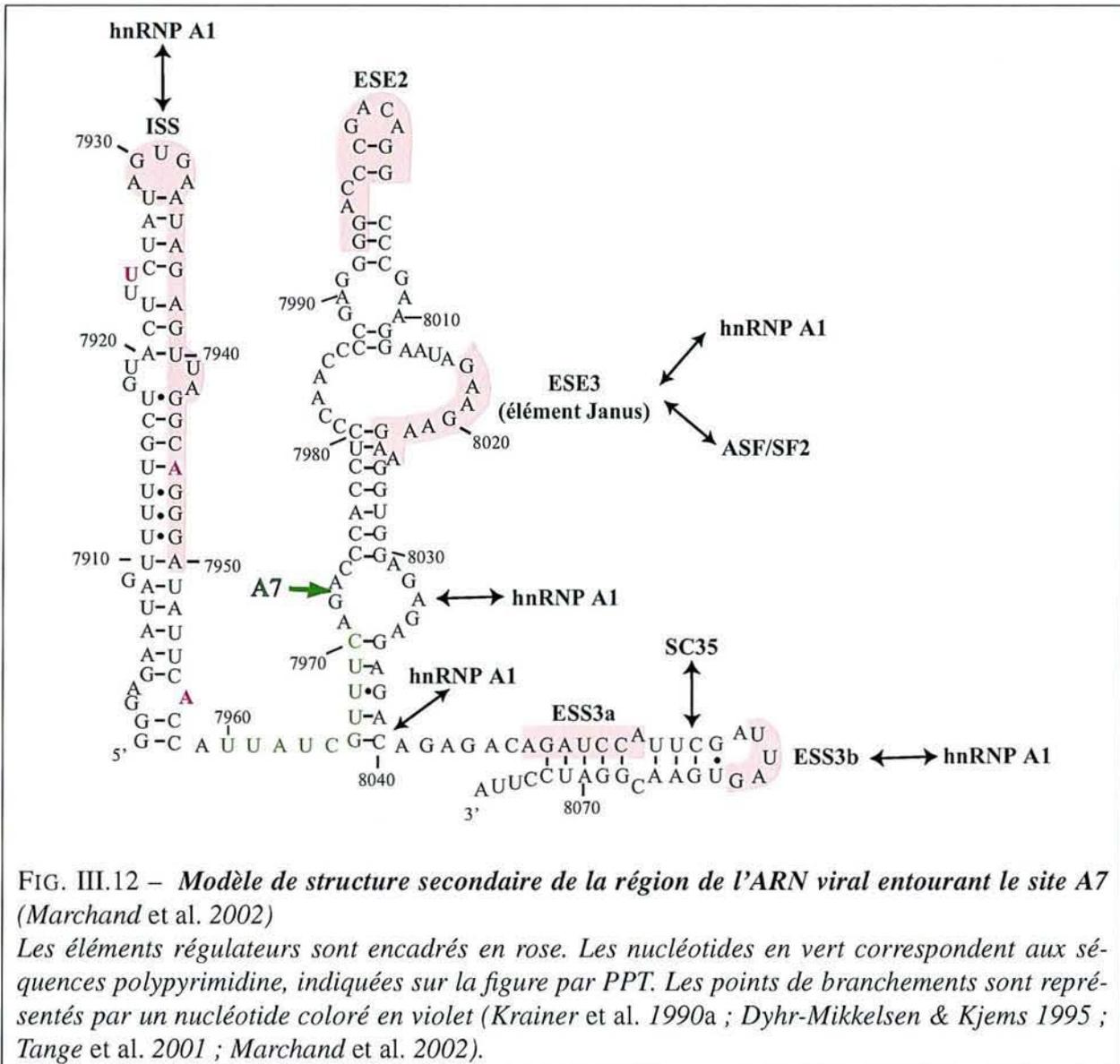
L'étude d'une souche variante du virus HIV-1 HXB2 a montré que l'inclusion de l'exon 6D dans les ARNm *tat*, *rev* et *nef* était due à la présence d'une mutation U→C dans le 3^e motif AUG de l'élément ESS (Wentz *et al.* 1997). Cette mutation transforme l'élément en une séquence activatrice d'épissage fixant la protéine SC35, ce qui conduit à l'inclusion de l'exon 6D dans les ARNm viraux. La protéine SC35 pourrait convertir le complexe formé par la protéine hnRNP H et la snRNP U1 en un complexe actif d'épissage (Caputi & Zahler 2002).

Site A7

La régulation de l'épissage à ce site est à ce jour la plus étudiée et la plus complexe. En effet, de multiples séquences régulatrices contrôlent l'efficacité d'épissage du site A7. Ce site est, rappelons-le, utilisé pour la production des messagers *tat* et *rev*. Comme les sites A2 à A5, il est localisé dans une région structurée de l'ARN viral, ce qui peut donc moduler son efficacité en épissage (Fig. III.12 ; Marchand *et al.* 2002 ; Damgaard *et al.* 2002).

Une multitude d'éléments régulant l'utilisation du site A7 a été mise en évidence (Fu *et al.* 1992 ; Staffa & Cochrane 1995 ; Amendt *et al.* 1995 ; Si *et al.* 1998 ; Mayeda *et al.* 1999 ; Zhu & Krainer 2000 ; Tange *et al.* 2001 ; Tange & Kjems 2001 ; Zhu *et al.* 2001 ; Marchand *et al.* 2002). Ils sont représentés dans la Figure III.12.

Plusieurs éléments inhibiteurs fixant la protéine hnRNP A1 ont ainsi été décrits. L'un d'eux, l'élément intronique ISS, chevauche un des points de branchement du site A7 et contient 3 sites de fixation de la protéine hnRNP A1 (Tange *et al.* 2001 ; Marchand *et al.* 2002 ; Damgaard *et al.* 2002). La fixation de la protéine hnRNP A1 pourrait empêcher la liaison de la protéine SF1/mBBP et de la snRNP U2 sur la séquence de branchement (Damgaard *et al.* 2002). L'autre élément, ESS3, est exonique (Staffa



& Cochrane 1995 ; Si *et al.* 1998 ; Tange *et al.* 2001 ; Zhu *et al.* 2001). La protéine hnRNP A1 se fixe à la séquence ESS3b de cet élément (Fig. III.12 ; Marchand *et al.* 2002 ; Damgaard *et al.* 2002), tandis que la protéine se fixant à la séquence ESS3a n'a pas encore été identifiée.

Des séquences activatrices d'épissage ont également été identifiées en aval du site A7 : l'élément ESE2 (Tange & Kjems 2001), un site de fixation de la protéine SC35 dans la séquence joignant les motifs ESS3a et ESS3b, ce qui explique l'activation de l'épissage de l'intron D4-A7 par cette protéine SR (Mayeda *et al.* 1999), et l'élément Janus ESE3 (Staffa & Cochrane 1995 ; Marchand *et al.* 2002). Cet élément joue un rôle clé dans la régulation du site A7. En effet, la séquence (GAA)₃ de l'élément active l'épissage en fixant la protéine ASF/SF2, tandis que les nucléotides situés en amont de cette séquence forment un élément AUA(GAA)₃ qui réprime l'épissage en fixant la protéine hnRNP A1 (Marchand *et al.* 2002).

Les travaux du laboratoire ont permis de mieux comprendre les inter-relations entre les éléments régulateurs et leur mécanisme d'action (Marchand *et al.* 2002). Une compétition entre les protéines hnRNP A1 et ASF/SF2 pour leur fixation à l'ARN est ainsi à la base de la régulation du site A7

(Fig. III.12).

Si la protéine hnRNP A1 est présente en forte quantité, elle se fixe de manière coopérative sur ses différents sites, ce qui conduit à sa multimérisation le long de l'ARN et empêche ainsi la reconnaissance du site A7 par la machinerie spliceosomale. Si au contraire, le taux d'ASF/SF2 est élevé, la protéine peut se fixer à l'élément ESE3, ce qui empêche la fixation coopérative de la protéine hnRNP A1 et favorise la reconnaissance du site A7 par le spliceosome (Zhu *et al.* 2001 ; Tange *et al.* 2001 ; Marchand *et al.* 2002 ; Damgaard *et al.* 2002).

La région du site A7 est soumise à une autre régulation, celle de l'export des ARNm *env*. En effet, les protéines SR ASF/SF2, SC35 et la protéine hTra2 α seraient capables de réguler l'export par la protéine Rev. L'amplitude de l'effet des protéines SR et de hTra2 α sur la fonction de Rev dépend du type cellulaire et des quantités relatives de ces différentes protéines. Ainsi, la surexpression d'ASF/SF2 et de SC35 inhibe l'export des ARNm *env* par la protéine Rev dans les cellules HeLa et 293, sans pourtant stimuler l'épissage de cet ARNm (Pongoski *et al.* 2002). Le mécanisme d'action des protéines SR n'est pas connu pour le moment. De manière surprenante, ces protéines n'agissent pas en se liant à leur site de fixation en aval du site A7, alors que les éléments ESE3 et ESS3 sont tout de même capables de réguler la fonction de Rev, puisque la délétion de la séquence ESS3 augmente l'efficacité d'export des messagers *env* par Rev, qui est au contraire réduite lorsque la séquence ESE3 est délétée (Pongoski *et al.* 2002).

Objectifs et résumé des travaux

Comme nous avons pu le voir, l'étape d'épissage est primordiale pour la multiplication du virus HIV-1. La compréhension des mécanismes impliqués dans la régulation de l'épissage permettrait d'avoir une maîtrise de cette étape du cycle viral et par là-même, d'essayer de développer des approches thérapeutiques visant à prendre les mécanismes de régulation de l'épissage comme cible.

Lorsque j'ai débuté cette étude, on commençait à savoir que les protéines SR étaient des facteurs importants pour la régulation de l'épissage des ARN pré-messagers cellulaires, mais les effets mis en évidence concernaient surtout leur action sur les sites donneurs. Il en était d'ailleurs de même pour l'effet inhibiteur des protéines hnRNP, qui avait essentiellement été caractérisé au niveau des sites 5'. Ce sont les travaux récents de notre équipe et d'autres équipes, qui ont montré que ces 2 types de protéines agissent aussi sur les sites 3'. Lorsque j'ai débuté ce travail, les données du laboratoire et de la littérature suggéraient que la régulation de l'épissage de l'ARN viral s'exerce essentiellement sur les sites 3' d'épissage. En plus de l'effet inhibiteur de la structure secondaire sur l'utilisation du site A3, quelques éléments régulant *in cis* l'utilisation des sites 3' avaient été identifiés : les 2 éléments ESS, réprimant les sites A3 et A7, et l'élément ESE3, activateur du site A7. Leurs partenaires protéiques n'avaient, par contre, pas été identifiés.

Étant donné que le site A7 fait partie de l'ARN pré-messager *tat23*, utilisé depuis longtemps comme modèle d'étude de l'épissage alternatif, il avait d'abord été montré par Krainer *et al.* (1990a), que la protéine ASF/SF2 active l'utilisation de ce site d'épissage *in vitro*. Par contre, l'effet des protéines SR n'avait pas été testé sur l'utilisation des autres sites d'épissage de l'ARN du virus HIV-1. J'ai donc entrepris, dans un 1^{er} temps, une étude *in vitro* de l'effet des protéines SR sur l'utilisation des sites d'épissage A1 à A5 de l'ARN du virus HIV-1, ceci en collaboration avec L. Ayadi. Afin de confirmer la signification biologique des effets trouvés *in vitro*, nous avons ensuite mené une étude *ex vivo*, en collaboration avec l'équipe de J. Stévenin (IGBMC, Strasbourg).

Les protéines SR se sont révélées avoir des effets différentiels marqués sur le taux d'utilisation des différents sites d'épissage du virus HIV-1. Les sites accepteurs A2 et A3 se sont avérés fortement dépendants de certaines protéines SR, ce qui soulevait la question de la présence, sur l'ARN viral, de séquences régulatrices *in cis* fixant ces protéines. Dans un 2^e temps, j'ai essayé d'identifier ces séquences par différentes approches expérimentales. La caractérisation au laboratoire, par S. Jacquenet et A. Méreau, d'un 2^e élément inhibiteur du site A3, l'élément ESS2p, nous confortait dans l'idée que de multiples régulations s'exercent sur les sites d'épissage et qu'il devait exister des séquences régulatrices capables de contrecarrer l'effet des éléments inhibiteurs d'épissage. En collaboration avec L. Ayadi, j'ai ainsi identifié et caractérisé un 3^e élément régulateur du site A3, l'élément ESEt. Il active

l'utilisation du site A3 en fixant les protéines ASF/SF2 et SC35. Nous avons également pu identifier d'autres sites de fixation des protéines SR sur l'ARN viral, en aval des sites A2 et A3. Les résultats obtenus nous ont permis de proposer un modèle de régulation de l'épissage au site A3 du virus HIV-1.

Compte-tenu de l'effet majeur des protéines SR sur l'épissage de l'ARN du virus HIV-1, nous nous sommes demandés si cette régulation de l'épissage était phylogénétiquement conservée. J'ai donc étudié la régulation s'exerçant sur l'utilisation des sites A3 à A5 des ARN de 2 virus apparentés à HIV-1, mais phylogénétiquement éloignés, HIV-1 YBF30 et SIVcpz ANT.

Dans un dernier volet de mon travail de thèse, j'ai établi une collaboration avec D. Eveillard et A. Bockmayr (projet MODBIO, LORIA, Nancy), afin d'établir un modèle mathématique pouvant rendre compte des variations observées du taux d'utilisation du site A3, en fonction des concentrations respectives en protéines ASF/SF2, SC35 et hnRNP A1 dans les cellules. L'idée est d'établir des modèles mathématiques pouvant refléter les différents modes de régulation possibles d'un site d'épissage, ceci afin de pouvoir, à partir de données expérimentales sur un site donné, essayer de savoir à quel type de régulation il est soumis.

Deuxième partie

Résultats

Chapitre IV

Les protéines SR sont impliquées dans la régulation de l'épissage alternatif de l'ARN du virus HIV-1

IV.1 Etat des connaissances au départ et stratégie utilisée pour l'étude

IV.1.1 Etat des connaissances

Comme je l'ai précisé dans la description de mes objectifs de thèse, lorsque j'ai débuté ma thèse, des études antérieures sur l'épissage de l'ARN du virus HIV-1 avaient montré que les sites accepteurs de l'ARN viral étaient de faible efficacité. Cela supposait l'existence de régulations, comme c'est le cas pour les ARN pré-messagers cellulaires. D'ailleurs, des éléments régulateurs en *cis* avaient déjà été identifiés en aval du site A3 (l'élément inhibiteur ESS2; Amendt *et al.* 1994, 1995; Si *et al.* 1997) et en aval du site A7 (l'élément activateur ESE3 et l'élément inhibiteur ESS3; Staffa & Cochrane 1995; Tange *et al.* 2001; Zhu *et al.* 2001; Si *et al.* 1998) (Fig. IV.1). Néanmoins, les partenaires protéiques de certains de ces éléments n'étaient pas connus à ce moment-là et on pouvait supposer l'existence d'autres éléments régulateurs.

Comme nous l'avons vu dans la revue bibliographique, les protéines SR jouent un rôle majeur dans l'activation des sites d'épissage des ARNm cellulaires. Etant donné que l'ARN viral utilise la machinerie spliceosomale cellulaire pour son épissage, il était vraisemblable que les protéines SR agissent aussi sur les sites d'épissage de l'ARN du virus HIV-1. Une implication de la protéine ASF/SF2 avait d'ailleurs été mise en évidence dans l'activation du site A7 (Krainer *et al.* 1990a). Par contre, aucune étude de l'effet de ces protéines sur les sites accepteurs A1 à A5 n'avait été réalisée. Nous avons donc entrepris une étude systématique de l'effet des protéines SR sur ces 7 sites.

IV.1.2 Stratégie générale de l'étude

Une manière classique d'aborder l'étude de l'effet des protéines SR sur l'efficacité d'utilisation d'un site d'épissage est de réaliser des expériences d'épissage *in vitro* avec des ARN contenant ce site d'épissage et un extrait nucléaire de cellules HeLa auquel on rajoute un mélange de protéines SR purifiées à partir de cellules HeLa ou des protéines SR recombinantes. En général, des études sont

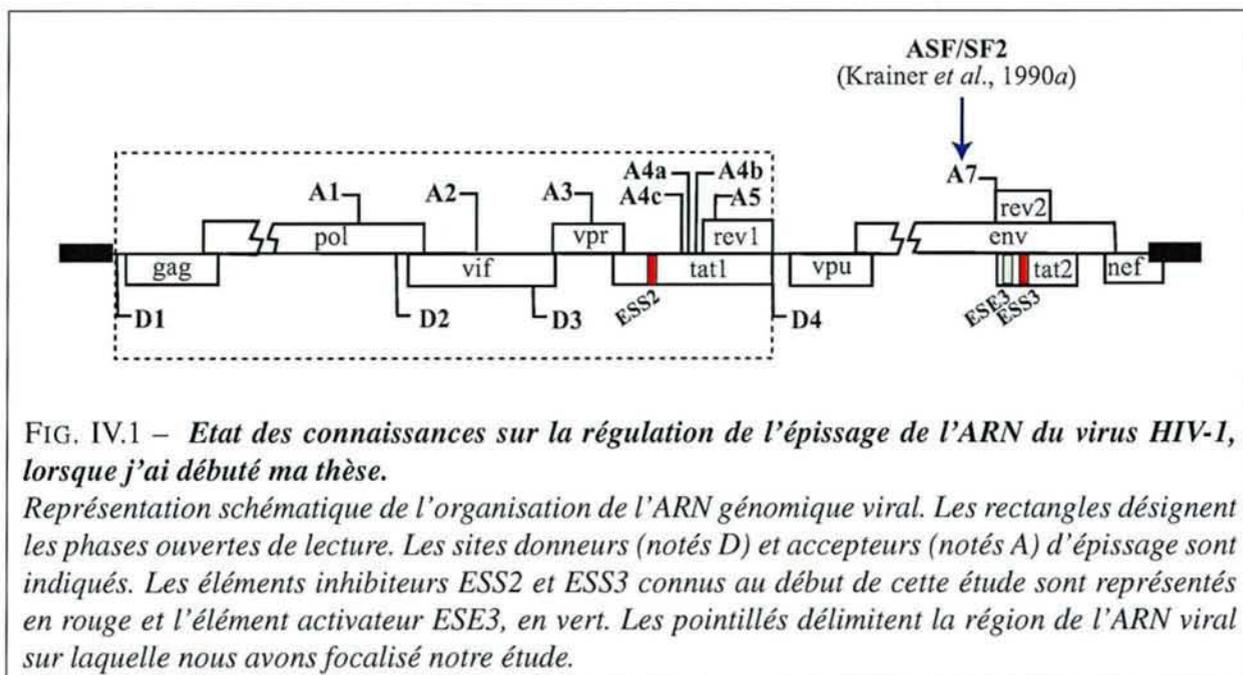


FIG. IV.1 – *Etat des connaissances sur la régulation de l'épissage de l'ARN du virus HIV-1, lorsque j'ai débuté ma thèse.*

Représentation schématique de l'organisation de l'ARN génomique viral. Les rectangles désignent les phases ouvertes de lecture. Les sites donneurs (notés D) et accepteurs (notés A) d'épissage sont indiqués. Les éléments inhibiteurs ESS2 et ESS3 connus au début de cette étude sont représentés en rouge et l'élément activateur ESE3, en vert. Les pointillés délimitent la région de l'ARN viral sur laquelle nous avons focalisé notre étude.

faites en parallèle avec un extrait cytoplasmique S100 de cellules HeLa qui contient beaucoup moins de protéines SR que l'extrait nucléaire. C'est ce que nous avons réalisé.

D'une manière générale, il est important de confirmer des données obtenues *in vitro* par des données *ex vivo*, ce qui fait appel à des expériences basées sur la transfection de cellules HeLa par des plasmides produisant, en expression transitoire, un ARN contenant le site d'épissage étudié et une protéine SR donnée. En collaboration avec l'équipe de J. Stévenin (IGBMC, Strasbourg), cette approche a été employée pour compléter les données que nous avons obtenues *in vitro*.

IV.2 Etude de l'effet de protéines SR purifiées sur l'efficacité d'épissage des sites A1 à A5

IV.2.1 Choix des protéines SR testées

Sur la base de leur spectre de reconnaissance des séquences ARN, les protéines SR peuvent être subdivisées en 2 classes : celles ayant une faible spécificité de reconnaissance, comme ASF/SF2 et SC35, et celles ayant une forte affinité pour un nombre très limité de séquences, comme la protéine 9G8. La protéine SRp40 est particulière, puisqu'elle reconnaît à la fois des séquences de manière peu spécifique et, avec une plus forte affinité, une séquence bien définie (cf § II.3, page 67).

Précisément du fait de leur large spécificité de reconnaissance de l'ARN, les protéines SR SC35 et ASF/SF2 sont très largement impliquées dans l'épissage et sa régulation. Les protéines SR 9G8 et SRp40 sont impliquées dans certaines régulations plus spécifiques de certains ARNm, comme nous l'avons vu dans l'introduction. Il était donc intéressant de tester l'effet de ces différentes protéines sur l'utilisation des sites accepteurs du virus HIV-1, d'autant plus que l'équipe de J. Stévenin, avec qui nous collaborons, pouvait nous fournir ces protéines produites dans un système utilisant le baculovirus. Ce système présente l'avantage de produire des protéines phosphorylées actives en épissage. Nous

avons donc pu tester l'effet de l'ajout de chacune de ces protéines SR recombinantes sur l'utilisation des sites accepteurs A1 à A5.

IV.2.2 Choix des ARN utilisés pour l'étude *in vitro*

Comme nous l'avons indiqué plus haut, nous nous sommes intéressés aux sites A1, A2, A3, A4c, A4b et A5 qui peuvent être utilisés en combinaison avec les 4 sites donneurs D1, D2, D3 et D4. Etant donné que peu d'études avaient été réalisées sur ces sites, à part la démonstration de la présence d'un élément inhibiteur agissant sur le site A3, il était important dans un 1^{er} temps d'étudier l'efficacité relative de ces sites *in vitro*, avec ou sans ajout de protéines SR recombinantes, dans l'extrait nucléaire de cellules HeLa utilisé pour les réactions d'épissage *in vitro*. Pour cela, nous avons utilisé différentes constructions produisant des ARN correspondant à des segments discontinus de l'ARN de HIV-1. Cinq de ces constructions avaient été réalisées par L. Damier avant mon arrivée au laboratoire (C1, C2, C3, C12, L23; Damier 1997), j'en ai réalisé une sixième (L3-U1).

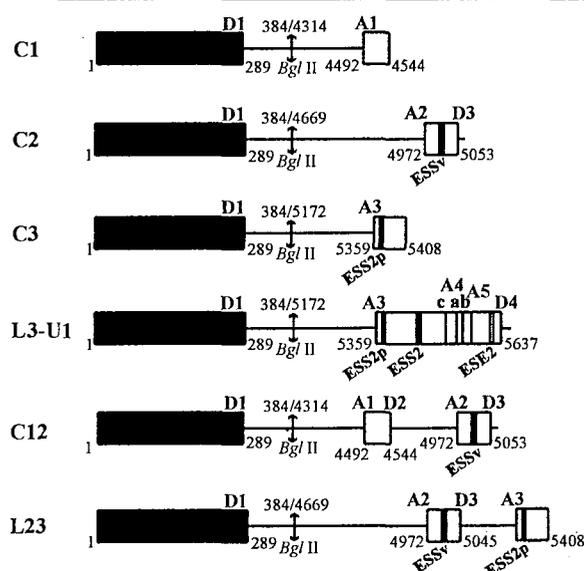


FIG. IV.2 – Constructions génétiques utilisées pour l'étude de l'effet des protéines SR sur l'utilisation des sites A1 à A5 *in vitro*.

Les éléments inhibiteurs sont représentés par un trait noir et l'élément activateur, par un trait gris.

Toutes ces constructions ont été générées à partir de l'ADNc du virus HIV-1 BRU inséré dans le plasmide pBRU3, que nous avait fourni F. Clavel (IMEA/INSERM, Paris). Le choix des ARN produits pour réaliser les tests d'épissage *in vitro* reposait sur le fait que pour pouvoir comparer l'efficacité relative des sites accepteurs de l'ARN de HIV-1 en présence ou en absence de protéines SR recombinantes, il était préférable de coupler ces sites à un même site donneur de HIV-1. En effet, même si la majorité des régulations de l'épissage de l'ARN de HIV-1 s'exercent sur les sites accepteurs, il n'est pas exclu que les différentes protéines SR conduisent à une utilisation différentielle des sites donneurs. Donc, pour une meilleure validité de la comparaison de l'efficacité des sites accepteurs au

sein des ARN testés en épissage, ces sites étaient tous couplés à un même site donneur de HIV-1. Nous avons choisi le site donneur D1, qui est utilisé dans tous les ARNm produits *in vivo*, et qui est le plus efficace des 4 sites donneurs *in vitro* (Damier 1997). De plus, il est vraisemblablement le moins soumis à des régulations.

Ainsi, les constructions C1, C2 et C3 permettaient d'obtenir des ARN renfermant respectivement les couples de sites D1 et A1, D1 et A2, et D1 et A3. La construction L3-U1 conduisait à un ARN renfermant le site donneur D1, ainsi que plusieurs sites accepteurs en compétition A3, A4c, A4a, A4b et A5. De même, les ARN produits à partir de la construction C12 renfermaient les sites accepteurs A1 et A2, et ceux obtenus à partir de la construction L23 renfermaient les sites accepteurs A2 et A3 (Fig. IV.2).

Un autre paramètre dont il était tenu compte dans ces constructions était le fait que l'épissage *in vitro* n'est efficace que pour des ARN ayant une taille inférieure à environ 1000 nts. Ainsi, une grande partie de l'ARN viral situé entre le site D1 et le ou les sites accepteurs testés était déléetée dans les ARN produits. Les ARN contenaient un segment de 384 nts en 5' renfermant le site donneur D1 (positions 1 à 384 de l'ARN HIV-1 BRU), joint à une région de 230 à 740 nts renfermant le ou les sites accepteurs (Fig. IV.2).

IV.2.3 Choix de la construction utilisée pour l'étude *ex vivo*

Nous disposions du plasmide p Δ PSP construit par l'équipe de M. Stoltzfus. Il renferme l'ADNc de la souche HIV-1 pNL4-3, tronqué entre les sites D1 et A1 et contenant l'ensemble des sites donneurs et accepteurs d'épissage de HIV-1 (Fig. IV.3). La souche pNL4-3 est phylogénétiquement très proche de la souche HIV-1 BRU, car elle appartient également à la classe B du groupe M des virus HIV-1 (cf § III.1.2, page 94 de la revue bibliographique).

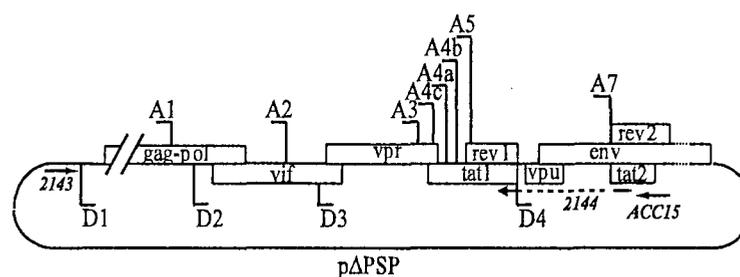


FIG. IV.3 – Représentation schématique du plasmide p Δ PSP utilisé pour étudier l'épissage du transcrit primaire du virus HIV-1 dans des cellules HeLa.

Le plasmide renferme l'ADNc proviral de la souche HIV-1 pNL4-3. Les barres obliques désignent la délétion réalisée dans l'intron D1-A1 de l'ADNc proviral. Les oligonucléotides 2143, 2144 et ACC15, utilisés pour l'analyse des produits d'épissage par la technique RT-PCR sont également représentés.

L'équipe de M. Stoltzfus avait montré qu'après transfection de cellules HeLa avec ce plasmide, l'épissage de l'ARN produit reflète parfaitement l'épissage obtenu pour un ARN viral entier (Jacquet et al. 2001a). Il était donc possible de tester l'effet des protéines SR sur l'utilisation relative des différents sites d'épissage de l'ARN de HIV-1, en cotransfectant des cellules HeLa avec le plasmide

p Δ PSP et un plasmide surproduisant l'une des protéines SR.

Les produits d'épissage obtenus à partir du produit de transcription ont été analysés par la technique RT-PCR en utilisant un oligonucléotide amorce de la transcriptase inverse (ACC15, Fig. IV.3), complémentaire à la séquence de l'ARN de la souche pNL4-3 située, en aval du site A7, entre les positions 8445 à 8463. Pour l'amplification par PCR, le 1^{er} oligonucléotide utilisé avait une séquence complémentaire à la séquence générée après épissage entre les sites D4 et A7. Ainsi, seuls les ARN épissés entre ces 2 sites étaient analysés par cette approche. Le 2^e oligonucléotide utilisé (2143) avait une séquence correspondant à celle de l'ARN de la souche pNL4-3 située en amont du site D1, entre les positions 700 à 727.

IV.2.4 Choix des conditions d'épissage *in vitro*

Comme nous l'avons vu dans l'introduction, un même profil d'épissage de l'ARN de HIV-1 est obtenu dans des cellules HeLa transfectées par l'ADNc proviral et dans des lymphocytes T infectées par les virions. Or, les extraits nucléaires de cellules HeLa sont les plus utilisés pour la réalisation d'expériences d'épissage *in vitro*. D'une manière générale, nous avons donc réalisé tous nos tests de l'efficacité d'utilisation des sites d'épissage de HIV-1 *in vitro*, dans des extraits nucléaires de cellules HeLa.

Au laboratoire, les études d'épissage *in vitro* sont généralement réalisées dans un mélange renfermant 36% d'extrait nucléaire (v/v), soit 8 μ l d'extrait nucléaire dans 22 μ l de mélange réactionnel en tampon D. Ces conditions ont été choisies pour optimiser l'efficacité d'épissage. Afin de mieux mettre en évidence l'effet relatif des protéines SR sur l'épissage des ARN produits à partir des constructions décrites ci-dessus, nous avons réalisé des tests d'épissage en utilisant un extrait cytoplasmique S100, qui contient très peu de protéines SR par rapport à l'extrait nucléaire. Les tests ont été réalisés dans différentes conditions : soit en extrait S100 de cellules HeLa seul, soit dans un mélange comportant 7 μ l d'extrait S100 et 1 μ l d'extrait nucléaire ou encore, 4 μ l d'extrait S100 et 4 μ l d'extrait nucléaire, et ce, toujours avec ou sans ajout de protéines SR recombinantes. En plus, l'essai classique avec 8 μ l d'extrait nucléaire était aussi réalisé, avec et sans ajout de protéines SR.

IV.2.5 Résumé des résultats obtenus à partir de cette analyse *in vitro*

Les données obtenues sur l'effet des protéines SR sur l'épissage de l'ARN du virus HIV-1 sont décrites dans l'article qui suit. Elles ont été évaluées positivement pour publication dans la revue J. Biol. Chem., moyennant l'ajout de quelques expériences complémentaires, incluses dans la version révisée du manuscrit ci-jointe. Un certain nombre de données supplémentaires non incorporées dans l'article sont ajoutées en annexe et un bilan global des résultats, ainsi qu'une discussion sont joints à la suite. En plus des analyses *in vitro* que nous avons réalisées au laboratoire, l'article comporte une analyse complémentaire réalisée *ex vivo* dans des cellules HeLa, en collaboration avec l'équipe de J. Stévenin. R. Gattoni a réalisé ce type d'expérience pour les 4 protéines SR ASF/SF2, SC35, 9G8 et SRp40 et nous avons participé à l'interprétation des données obtenues. En effet, ces données peuvent être très bien interprétées sur la base des données que nous avons obtenues *in vitro*.

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 revealed by *in vitro* and *ex vivo* splicing experiments.

Ropers D., Ayadi L., Gattoni R., Jacquenet S., Damier L., Branlant C. and Stévenin J.

Manuscrit en révision pour publication dans J. Biol. Chem.

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

Delphine Ropers¹, Lilia Ayadi¹, Renata Gattoni², Sandrine Jacquenet^{1*}, Laurence Damier¹,
Christiane Branlant^{1¶} and James Stevenin²

¹*Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR CNRS 7567, Université Henri Poincaré Nancy 1, Boulevard des Aiguillettes, BP239, 54506 Vandoeuvre-lès-Nancy cedex, France*

²*IGBMC, UMR CNRS-INSERM-ULP 7104, 1 rue Laurent Fries, BP 10142, Parc d'innovation, 67404 Illkirch cedex, France*

*Present address: Unité de virologie humaine (U412 INSERM), Ecole Nationale Supérieure de Lyon, 46 allée d'Italie, 69007 Lyon, France

¶To whom correspondence should be addressed:

MAEM, UMR CNRS-UHP 7567, Faculté des Sciences, Boulevard des Aiguillettes, BP239, 54506 Vandoeuvre-lès-Nancy cedex, France

Tel: 03.83.68.43.03o Fax: 03.83.68.43.07, email: Christiane.Branlant@maem.uhp-nancy.fr

Running Title: Effect of SR proteins on the HIV-1 3' splice site utilization

Keywords: virus HIV-1, splicing, SR proteins, tat, vpr, nef

SUMMARY

Splicing is a crucial step for HIV-1 virus multiplication. Eight HIV-1 acceptor sites are used in competition to produce the *vif*, *vpu*, *vpr*, *nef*, *env*, *tat* and *rev* mRNAs. Although SR proteins are known to modulate alternative splicing of pre-messenger RNAs, effects of SR proteins on utilization of the A1 to A5 HIV-1 acceptor sites had not been investigated. To test this effect, we determined the *in vitro* splicing efficiencies of transcripts containing one or several HIV-1 acceptor sites, in S100 extract or mixtures of S100 and nuclear extracts, in the presence or absence of recombinant SR proteins (ASF/SF2, SC35, 9G8 or SRp40). On the other hand, HeLa cells were cotransfected with a construct containing a truncated HIV-1 cDNA with all the splicing sites and a construct overexpressing one of the SR protein mentioned above. The yield of production of the various HIV-1 mRNAs was quantified by RT-PCR experiments. The data demonstrate that two of the A1 to A5 sites, namely A2 and A3, are strongly and specifically regulated by SR proteins. ASF/SF2 activates strongly site A2 and to a lesser extent site A1. SC35 and SRp40 activate strongly site A3. 9G8 has a more modest activation effect on site A3. As a result, upon ASF/SF2 overexpression, the production of *vpr* is specifically increased, whereas upon SC35 or SRp40 overexpression, production of *tat* mRNAs is predominant. Activation by ASF/SF2 and 9G8 was found to be independent of the RS domain. Western-blot analysis with Nef specific antibodies demonstrated the increase of *vpr* or *tat* mRNA

INTRODUCTION

Splicing plays a key role for production of the HIV-1 retroviral proteins. By using the integrated proviral genome as the template, RNA polymerase II of the infected cell produces long primary transcripts which are all identical. Some of these transcripts are transported to the cytoplasm in an intact form to serve as genome for new virions or as messenger RNA for the production of the Gag-Pol protein precursor. The other transcripts undergo alternative splicing to produce mRNAs for the auxiliary and regulatory proteins and the Env precursor protein. Production of mRNAs encoding HIV-1 proteins depends on the alternative utilization of 4 alternative 5'-splice sites D1 to D4 (1) and 8 alternative 3'-splice sites (A1, A2, A3, A4a, b, c, A5 and A7) (1). An additional 3'-splice site (A6) was only found in one HIV-1 strain (2). Donor sites D1, D2 and D3 can be coupled to any of the A1 to A5 acceptor sites, whereas the donor site D4 is exclusively coupled to site A7 and, when this site is present to site A6 (1,2). The combination of these various sites give rise to at least 35 different mRNAs (1). Whereas the relative efficiencies of the HIV-1 donor sites seem to depend mainly upon their complementarity to the U1 snRNA 5'-terminal sequence (3,4), efficiencies of HIV-1 acceptor sites depend upon the presence of *cis*-regulatory elements (5-21). Several studies have shown that HIV-1 acceptor sites are suboptimal: (i) their polypyrimidine tracts are short and interrupted by purines (3,5,9), (ii) their branch-point sequences are not canonical and in some cases a residue other than an adenosine is used as the branch site (22,23), (iii) their accessibility is limited by their sequestering in stable secondary structure (24), and (iv) several *cis*-inhibitory elements have been identified that downregulate the A2, A3 and A7 sites by binding of hnRNP A1 or hnRNP H proteins (11,12,14,16,18,19,21). In spite of these numerous handicaps, the utilization of several of the HIV-1 acceptor sites is

essential for virus multiplication. Sites A3 and A7 are used for production of mRNAs of the Tat transcriptional activator (1; for review, 61). One of the three A4a, b or c sites and site A7 are used to produce mRNAs for the Rev protein (1). This regulatory protein plays an essential role in the transport of intact primary transcripts to the cytoplasm (for review, 25). Site A5 is required for production of mRNAs for the Nef and Vpu proteins. Nef downregulates the cell-surface expression of the HIV-1 receptor glycoprotein, and has many additional roles in the infected cell that favor virus multiplication. It is also responsible for AIDS pathology, because of its role in bystander lymphocyte apoptosis (for reviews, 26,27). Finally, site A2 is needed for production of mRNAs for the Vpr protein. Vpr is implicated in nuclear import of the HIV-1 pre-integration complex and it induces cell cycle arrest in proliferating cells (for review, 28).

The efficiency of the regulated acceptor sites of cellular mRNAs generally depends upon the binding of members of the SR protein family (for reviews, 29,30). To date, 10 distinct members of this family have been identified (for reviews, 29,31). These proteins contain one or two RNA Recognition Motif(s) (RRM), together with a carboxy-terminal domain containing numerous Arginine-Serine aminoacid pairs. SR proteins are involved in constitutive splicing where they seem to have redundant functions in assistance to spliceosomal component assembly. Because of their interactions with the 70K protein of U1 snRNP, the 35 kDa subunit of factor U2AF, and proteins of the U4/U6.U5 tri-snRNP, the SR proteins are essential actors in the complex mechanism of spliceosome assembly (for reviews, 29,32). In addition, SR proteins can regulate alternative splicing by binding to *cis*-regulatory elements. These regulatory elements generally bind specifically one given SR protein, and depending on the identity of this SR protein, different effects are generated. Binding of SR proteins to specific sequences located downstream from the acceptor site generally reinforces the efficiency of U2AF binding to the polypyrimidine

tract, either by direct interaction of the RS domain of the SR protein with the RS domain of U2AF35 or by displacement of the hnRNP A1 protein that blocks the access of spliceosomal components to the 3'-splice site (for reviews, 29,30,32).

The role of SR proteins on viral multiplication has been demonstrated by previous studies on adenovirus. SR proteins activate the production of the early mRNAs, but a limitation of their activity is required for the late phase of infection (33). This limitation results from titration of the SR proteins by the major late transcripts (34) and/or a virus induced dephosphorylation (35). Activation by SR proteins of the A7 acceptor site was studied and both ASF/SF2 and SC35 activate this site (17,19,20,36,37). However, while variations of the cellular concentration of some of the SR proteins, namely SC35 and 9G8, have been observed after cell infection by virus HIV-1 (38,39), nothing is known about the effects of SR proteins on the HIV-1 A1, A2, A3, A4a, A4b, A4c and A5 sites. These 7 sites compete with each other for coupling to the D1, D2 or D3 sites. To fulfill this absence of knowledge, we performed *in vitro* splicing assays on RNA transcripts containing the D1 donor site and one or several of the A1 to A5 acceptor sites, in the presence or absence of the purified recombinant SR proteins ASF/SF2, SC35, 9G8 or SRp40. We also performed *ex vivo* splicing assays, by cotransfection of HeLa cells with the Δ PSP construct, that produces a truncated HIV-1 RNA with all the splicing sites (13,24), together with a plasmid which overexpresses individual SR protein. In these *ex vivo* experiments, very different HIV-1 RNA splicing patterns were obtained with different overexpressed SR protein. These data could be interpreted by taking into account the *in vitro* data. The possible implication of the observed effects on HIV-1 virus multiplication are discussed.

EXPERIMENTAL PROCEDURES

Plasmids used in this study

pLD-C2, pLD-C3 and pLD-L3-U1 constructs, used for production of the C2, C3 and L3-U1 transcripts, were previously described (11,23,24). To build the plasmid pLD-C1, two DNA fragments were generated by PCR amplifications of plasmid pBRU3 (40) that is containing the HIV-1 BRU/LAI complete cDNA (GenBank accession number: K02013). Amplifications were performed as previously described (23). The DNA fragment encoding the RNA region from position 4313 to 4545 (numbering of 41) was amplified with: a sense primer O-756 (5' TATTCTAGATCTCAGGCTGAACATC 3') containing the HIV-1 BRU RNA sequence from positions 4313 to 4325 (underlined) and generating a *Bgl* II restriction site (*italic*), and an antisense primer O-764 (5' CGGAATTCCCTTCCAGAGGAGCT 3') complementary to the HIV-1 BRU RNA from position 4530 to 4545 (underlined), that generated an *EcoR* I restriction site (*italic*). The *Bgl* II - *EcoR* I fragment of plasmid pLD-C2 (23) was substituted for the *Bgl* II - *EcoR* I fragment of the amplified DNA. The SP1 ex2 inv 9G8-102 plasmid (42) was used for production of an SP6 transcript allowing us to test for protein 9G8 splicing activation properties *in vitro*. Plasmid ΔPSP used for transfection experiments was previously described (11). Plasmids pXJ41-ASF, pXJ41-SRp40, pXJ42-9G8, pXJ42-SC35, or pXJ42-SRp20 were used for overexpression of the corresponding SR proteins. They were constructed by insertion of the cDNA of each SR protein in the multicloning site of pXJ41 or pXJ42 expression vector, as previously described (43).

***In vitro* transcription and *in vitro* splicing assays**

Plasmids pLD-C1 and pLD-C2 linearized by the *EcoR* I nuclease and plasmid pLD-C3 and pLD-L3-U1 linearized with the *Pst* I nuclease were used as templates for the production of capped, uniformly labeled RNAs by the T7 RNA Polymerase in the presence of [α - 32 P] UTP. The same conditions were used for production of uniformly labelled RNA from plasmid SP1 ex2 inv 9G8-102 linearized by *Hind* III, except that the T7 RNA polymerase was replaced by the SP6 RNA polymerase. *In vitro* splicing assays were performed in HeLa cell nuclear extract (from the Computer Cell Culture Center S.A., Belgium), pure cytoplasmic S100 extract prepared as previously described (34), or mixtures of S100 and nuclear extracts (4 μ l:4 μ l and 7 μ l:1 μ l), in 22 μ l assays using classical conditions for *in vitro* splicing reaction (23). Duration of incubation was of 2h30. For both the experiments in pure nuclear extract and in NE/S100 mixtures the results obtained with two different batches of nuclear extracts have been compared. The effects of using different batches of transcripts and SR proteins was also tested. For SR protein assays, 150 or 300 ng of purified 9G8, ASF/SF2, SC35 or SRp40 recombinant protein, prepared as previously described (42), were used. The degree of purity of these proteins was verified by electrophoresis followed by argentic staining. The C1, C2, C3, L3-U1 and Sp1 ex2 inv 9G8-102 splicing products were analyzed by electrophoresis on 5 or 6% polyacrylamide denaturing gels and were visualized by autoradiography. Quantifications were performed with a Molecular Dynamics Phosphorimager using the Image Quant Software (version 5.2). Splicing efficiency was expressed as the ratio of spliced products (M) versus unspliced RNA (P). For each gel electrophoresis of splicing products, M/P values were measured 3 times using the Image Quant Software. The mean value is given with an estimation of the M/P measurement error, shown as bar errors in the histograms representing the M/P values. The factor of splicing

activation by SR proteins was estimated by dividing the M/P ratio obtained in the presence of the recombinant SR protein by that obtained in its absence (fold of activation). The estimated errors on the factors of activation due to the experimental errors on M/P measurements mentioned above was deduced and represented as bar errors in the histograms showing activation factors.

***Ex vivo* splicing assays**

HeLa cells in 60 mm plastic Petri dishes were cotransfected by the calcium phosphate coprecipitation technique as previously described (44), with 1 µg of ΔPSP plasmid and 50 to 400 ng of plasmids overexpressing SR proteins. Carrier BSSK DNA was added to the transfections up to a total amount of DNA of 8 µg. Total cellular RNA was isolated from the transfected HeLa cells 48 hours after transfection and it was treated by DNase. Three µg of RNA were reverse transcribed by the MMLV reverse transcriptase according to the supplier's specifications (Gifco-BRL), using oligonucleotide ACC15 (5' TTCACTAATCGAATGGATC 3', complementary to the HIV-1 pNL4-3 RNA from positions 8445 to 8463) as the primer. About 10% of each of the RT assays were PCR amplified with the forward oligonucleotide 2143 (5' GGCTTGCTGAAGCGCGCACGGCAAGAGG 3', containing the HIV-1 pNL4-3 RNA sequence from positions 700 to 727) and reverse primer 2144 which spans the D4 and A7 splice sites (5' TTGGAGGTGGGTTGCTTTGATAGAG 3', complementary to pNL4-3 RNA from positions 6029 to 6041 and 8369 to 8381). This selectively amplified the 1.8 kb HIV size class mRNAs. Amplification was limited to 30 cycles, in order to detect both major and minor bands. DNA products were analyzed by electrophoresis on a 6% polyacrylamide non-denaturing gel and a Typhoon 8600 imager (Amersham) was used to scan the gel after ethidium bromide

staining. PCR products were quantified using Image Quant of Molecular Dynamics. The fractionated cDNAs were identified by reference to previous data on HeLa cells transfected with plasmid Δ PSP (11,24).

The level of expression of individual SR proteins in the different transfection experiments was estimated by Western-blot analysis using the monoclonal Ab10H3 antibody directed against the RS domain of the SR protein family and specific antibodies directed against ASF/SF2, SC35 or 9G8 (42).

A monoclonal antibody directed against the HIV-1 Nef protein (MaTG0020), a generous gift of A. M. Aubertin, was used to test by Western-blot assays the yield of transitory expression of the Nef protein, in HeLa cell transfected with the Δ PSP plasmid, with or without overexpressing one of the SR protein.

RESULTS

Strategy used for *in vitro* transcription assays

Because among the 8 conserved acceptor sites of HIV-1 RNA, the effects of SR proteins were already tested on site A7 (17,19,21,36,37), we focused our study on the effect of SR proteins on the acceptor sites A1, A2, A3, A4a, b, c and A5. First, we produced *in vitro* transcripts that contained different segments of the HIV-1 BRU RNA and we tested the effects of addition of purified recombinant SR proteins on the *in vitro* splicing efficiencies of these transcripts. They all shared the same donor site (D1) and contained one of the A1, A2 or A3 site, or the block of five A3, A4a, A4b, A4c and A5 sites, respectively (Figs. 1A and 2A). All the transcripts contained the 5' exon of HIV-1 RNA, that is delimited by site D1, and its 95-nt long downstream sequence. The 5' part of the intron was joined to segments of the HIV-1 RNA, containing the studied acceptor site(s). In transcripts C1, C2 and C3, the 178, 303 and 187 nt-long segments, located upstream from sites A1, A2 and A3, were respectively used as the 3' part of the intron. Transcript L3-U1 contained the same 3' intronic sequence as transcript C3. As no information was available on *cis*-regulatory elements at site A1, a 48 nt-long segment (positions 4492 to 4544) was arbitrarily selected as the second exon of transcript C1 (Fig. 1A). In transcript C2, the second exon (positions 4972 to 5053) contained the unique *cis* regulatory element described at site A2 (ESSV, positions 4995 to 5018) (12). In transcript C3, the second exon (positions 5359 to 5408) contained the recently identified *cis* inhibitory element ESS2p, that binds hnRNP H (11). However, the downstream ESS2 inhibitory element (7-9), that binds hnRNP A1 (14,16), was not present in transcript C3 (Fig. 1A). This element was present in the longer L3-U1 transcript, which ends at position 5637 and also contained sites A4a, b, c and A5 (Fig. 2A).

Compared to nuclear extracts, S100 extracts have the advantage to contain very low levels of SR proteins and thus are more suitable than nuclear extracts to test for the splicing activation effects of SR proteins. However, when recombinant SR proteins were added to S100 extract, the *in vitro* splicing efficiencies of the four tested transcripts were low (Figs. 1B, C, D and 2B). Hence, for a quantitative estimation of the activation of SR proteins on splicing at sites A1, A2, A3, A4c, A4a, A4b and A5 in the C1, C2, C3 and L3-U1 RNAs, respectively, splicing reactions were performed in a mixture obtained by addition of equal volumes (4 μ l) of HeLa nuclear and S100 extracts (4:4 NE/S100 mixture). Assays were also performed in a 1 μ l / 7 μ l mixture of nuclear and S100 extracts (1:7 NE/S100 mixture). For simplification, only the results obtained in the 4:4 NE/S100 mixture are shown in details and summary schemes are given for the 1:7 NE/S100 mixture. The splicing activation properties of four recombinant SR proteins namely, 9G8, ASF/SF2, SC35 and SRp40 were tested in these two mixtures and in S100 extract and the observed splicing efficiencies were compared to those in nuclear extract (Fig. 1B, C and D and Fig. 3).

In all these experiments, the splicing efficiencies were estimated as the ratio of the radioactivity contained in the mature RNA (M) *versus* that contained in the precursor RNA (P). The factors of activation, resulting from the addition of individual SR proteins, were calculated by dividing the M/P values obtained upon addition of the SR proteins by those obtained without their addition (Fold activation in Figs. 1 and 2). A comparative analysis of the activation values obtained in four separate experiments performed with different batches of nuclear extract, different preparations of transcripts and SR proteins revealed a high degree of reproducibility of the experiments. Representative examples of splicing products fractionation are shown in figures 1 and 2. Also in support to the validity of our approach, data obtained in the 4:4 NE/S100 mixture

(Figs. 1E3 and 2C3) were coherent with those obtained in the 1:7 NE/S100 mixture (Figs. 1E4 and 2C4).

Comparative analysis of site A1, A2 and A3 efficiencies *in vitro*

As expected, in the absence of SR protein addition, none of the C1, C2 and C3 transcripts was spliced in S100 extract (Fig. 1B, C and D). In the 4:4 NE/S100 mixture without SR protein addition, splicing efficiencies were in the order $C1 \gg C2 > C3$ (Figs. 1B, C, D and E2). The same relative order of splicing efficiencies was found for assays performed in pure nuclear extract (insets in Figs. 1B, C, D and E2). Hence, in the context of the transcripts used in these *in vitro* experiments, site A1 had a higher efficiency than sites A2 and A3. These data confirmed the observation that, even in the absence of the ESS2 inhibitory element, site A3 has a low splicing efficiency (11). In the 4:4 NE/S100 mixture, splicing at site A1 (Fig. 1B) was markedly activated by both ASF/SF2 (lane 6) and SC35 (lane 8) proteins, whereas the activations by proteins 9G8 (lane 4) and SRp40 (lane 10) were weak and undetectable, respectively (see Fig. 1E2 and E3 for quantifications). Splicing at site A2 (Fig. 1C) was also strongly activated by addition of protein ASF/SF2 (lane 6), but to a lower extent by addition of the SC35 (lane 8) and SRp40 (lane 10) proteins. Protein 9G8 (lane 4) had a very weak effect (Fig. 1E2 and E3). Finally, splicing at site A3 (Fig. 1D) was activated by addition of each of the four SR proteins, but again protein 9G8 had the weakest effect. On a quantitative point of view, protein ASF/SF2 had similar activation efficiencies on sites A2 and A3, as well in the 4:4 as in the 1:7 NE/S100 mixtures (factors of about 10 and 12, respectively), and the highest activation efficiencies were found for SRp40 and SC35 proteins at site A3 in the 4:4 NE/S100 (factors of about 15 and 19, respectively, Fig. 1E3).

We concluded that *in vitro*, in transcripts containing individual acceptor sites: (i) proteins ASF/SF2 and SC35 activate site A1 utilization to similar extents, (ii) protein ASF/SF2 has a stronger activation effect than protein SC35 on site A2, (iii) proteins SRp40 and SC35 have a stronger activation capacity than protein ASF/SF2 at site A3, (iv) protein 9G8 has a modest activity on the 3 sites. Data also revealed, that in the presence of very low amounts of SR proteins (S100 extract) (Fig. 1E1), (i) splicing at site A1 is efficient upon addition of protein ASF/SF2 or SC35, (ii) splicing at site A2 is efficient upon addition of protein ASF/SF2 and to a lower extent upon addition of proteins SRp40 and SC35 and, (iii) limited splicing occurs at site A3 upon addition of protein SC35, indicating that individual SR proteins are sufficient to promote the splicing of each of these acceptor sites.

Comparison of site A3, A4a, b and c and A5 efficiencies in the L3-U1 transcript

Transcript L3-U1 (Fig. 2) contained, in downstream position of site D4 position, the U1 snRNA binding region that is required for utilization of site A3 in the presence of ESS2 and also for utilization of sites A4a, b, c and A5 (7). As expected in pure S100 extract, no splicing occurred at any of the 5 sites of the L3-U1 RNA (Fig. 2B and C1). In the 4:4 NE/S100 mixture, site A5 was preferentially used (factor of 2 compared to site A3, Fig. 2C2) and it became predominantly used in pure nuclear extract (factor of 6 compared to site A3, Fig. 3A and C1). Addition of each of the purified SR proteins in S100 extract allowed a low yield of utilization of site A3, whereas the four other sites were not used (Fig. 2B and C1). Based on splicing efficiencies of the L3-U1 transcript in the 4:4 NE/S100 mixture, addition of protein 9G8 had a slight inhibitory effect on the utilization of all the five 3'ss (Fig. 2C3). Protein ASF/SF2 slightly increased utilization of sites A4a and b. However, in contrast with the results obtained for the C3 RNA, it did not

activate site A3 significantly. Protein SRp40 and to a higher extent protein SC35 had strong activation properties on site A3 (factors of about 2 and 5, respectively), but did not modify sites A4a and b utilization and decreased site A5 utilization. When the L3-U1 transcript was spliced in the 1:7 NE/S100 mixture (Fig. 2B and C4), the four SR proteins activated site A3 utilization. However, here again, protein SC35 had the strongest activation effect together with protein SRp40 (factors of about 5.5 and 5, respectively). Also in the 1:7 NE/S100 mixture, activation of sites A4a and b utilization by ASF/SF2 was strong (factor of 4).

When the L3-U1 transcript was spliced in nuclear extract (Fig. 3A, C1 and C2), addition of protein 9G8 again had a very limited effect on the relative efficiencies of the 5 sites, protein ASF/SF2 had a moderate activation effect on site A3 (factor of 2.6) and it slightly decreased site A5 utilization (30% decrease). Protein SRp40 and especially SC35 had a strong activation effect on site A3 (factors of about 5.2 and 15, respectively) and the two proteins had a negative effect on site A5 utilization (60% decrease in both cases) (Fig. 3C2). Interestingly, in the absence of SR protein addition or upon 9G8 addition, a strong accumulation of lariat intermediates was observed in nuclear extract. The yield of this intermediate decreased upon addition of proteins ASF/SF2, SRp40 and especially SC35. Thus, in nuclear extract and in the absence of additional SR proteins, site A3 is weakly used compared to site A5, but when an excess of protein SC35 or SRp40 is added, no more limitation of the second step of the splicing reaction occurs and splicing at site A3 becomes predominant over splicing at the other sites. Utilization of sites A4a and b is moderately sensitive to activation by SR proteins, since only protein ASF/SF2 slightly increased their utilization. Site A5 is not subjected to an activation by SR proteins and may be regulated by default.

To verify that the SR activation effects, that we observed in nuclear extract, were dependent upon the SR protein concentration, we compared the activation effects of proteins SC35 and SRp40 on site A3 utilization in the L3-U1 transcript, using pure nuclear extract and two different amounts of recombinant proteins, namely 150 and 300 ng (Fig. 3B) (this latter amount was the same as that used in the previous assays; Fig. 3A). Estimation of M/P values showed that activation increased for both proteins with increasing protein concentration. However, while increasing the SC35 protein amount by a factor of 2 augmented splicing efficiency at site A3 by a factor of 2.5, the same increase of the amount of protein SRp40 only augmented site A3 utilization by a factor of 1.6 (Fig. 3C3).

The strategy used for *ex vivo* assays

As in the previous *in vitro* assays activation of site A3 by SR proteins was strongly dependent upon the transcript, it was important to develop a strategy to test the effects of individual SR proteins on the splicing pattern of an RNA that contained all the HIV-1 splicing sites. Due to limitation in the length of RNAs that can be spliced *in vitro*, such experiments had to be performed by transfection of HeLa cells. To this end, we used the Δ PSP plasmid developed by Bilodeau and Stoltzfus (11), which contains a HIV-1 cDNA truncated inside the D1-A1 intron. RNA Polymerase II transcripts produced from this plasmid include all the HIV-1 splicing sites and previous experiments in HeLa cells showed that the relative yields of utilization of the splicing sites are identical in the Δ PSP transcript and in the authentic HIV-1 RNA (13,24). HeLa cells were co-transfected with p Δ PSP and 2 amounts (50 and 250 ng) of a recombinant vector, that expressed one SR protein: either the full length SR protein 9G8, ASF/SF2, SC35, SRp40 or SRp20 or a shortened form of protein 9G8 or ASF/SF2, that was missing the RS domain. The

full length SR proteins contained an additional histidine tag which decreased their mobility in SDS-PAGE and allowed us to distinguish them from their endogenous counterpart by means of a general SR detecting antibody comparable to mAb104: 10H3 (42) (see Fig. 4). Our previous experiments (unpublished data) showed that among the SRp30 proteins, proteins 9G8 and ASF/SF2, but not protein SC35, are efficiently recognized by the 10H3 antibody. Hence, using this antibody, two bands corresponding to 32-35 kDa proteins are detected which mainly represent the endogenous 9G8 and ASF/SF2 proteins (Fig. 4A). In addition, a specific SC35 antibody was also used for the analysis of protein SC35 expression (Fig. 4B). The other endogenous SR proteins (SRp75, SRp55, SRp40 and SRp20) are well detected with the 10H3 antibody (Fig. 4A). However, as previously observed by other authors using the mAb104 antibody (45-47), the yield of protein SRp40 detected in HeLa cells may sometimes be lower than those of proteins SRp30 and SRp20. Western-blot analysis of the SR protein expression was performed 3 times on different batches of cells co-transfected with plasmid Δ PSP and the highest amount of SR protein expressing vectors. Reproducible results were obtained and one example is given in Figure 4, showing that all recombinant SR proteins were well expressed. By quantification of the Western-blot signals, and taking as an arbitrary reference the recombinant SRp20 protein amount (1), the relative amounts of recombinant SRp20, ASF/SF2, 9G8 and SRp40 proteins were found to be of 1, 1.1, 1.25 and 1.75, respectively. The relative levels of each recombinant SR protein versus the corresponding endogenous protein was also estimated to 0.75 for SRp20, 1.70 for ASF/SF2, 1.55 for 9G8 and 4.4 for SRp40 (Figs. 4A and C) and 1.75 for SC35 (Figs. 4B and C). As a whole, considering that about 25% of HeLa cells were transfected with an SR expression vector, the recombinant SR proteins, except SRp40, were estimated to be expressed between 3 to 7 times more than the endogenous SR molecules, for the

highest amount of SR expression vectors, and at least twice less with the lowest amount. The values were highest for SRp40, but they are biased due to the lower level of endogenous SRp40 and they should not be considered as abnormal. Taken together, the results of figure 4 suggested that the overexpression of all SR proteins remained enough limited to mimic possible natural variations of SR proteins concentrations.

The effects of the overexpression of SR proteins on HIV-1 RNA splicing were analysed by RT-PCR. For the analysis of the levels of the various HIV-1 mRNAs in the absence of SR protein overexpression, HeLa cells were co-transfected with plasmid Δ PSP and an empty SR protein expression plasmid. In addition, as protein SRp20 was found to have no splicing activation capacity in our conditions, splicing of the Δ PSP transcript in cells overexpressing SRp20 was also used for the estimation of the basal levels of HIV-1 mRNAs (data not shown). Several transfection experiments using various sets of SR protein expression vectors have been performed. They gave similar results and one representative example is given (Fig. 5A). Note that the PCR analyses have been performed with a reverse primer which encompasses the D4-A7 junction, so that all the mRNAs identified were submitted to the D4-A7 splicing reaction. They were identified by comparison with previous data obtained by the same RT-PCR approach (11,24). In addition, the identity of the vpr1-2 mRNAs was verified using oligonucleotides specific for vpr mRNAs (data not shown). The quantification results obtained for the experiment in Figure 4A are shown in Figure 4C.

Overexpression of individual SR proteins strongly modifies the pattern of HIV-1 RNA splicing products

As previously observed by Purcell and Martin (1) and Bilodeau *et al.* (13), splicing of HIV-1 RNA without overexpression of SR proteins led to a higher production of the nef mRNA family, than the tat and rev mRNA families (Fig. 5A, lane 1). This is in agreement with the stronger utilization of site A5 found in nuclear extract *in vitro* (Fig. 3A). In addition, as previously observed (1), the vpr1 and vpr2 mRNAs were present in low yields. As illustrated in Figure 5A and C, the strongest effects of protein ASF/SF2 overexpression were: (i) a strong increase of the vpr mRNAs level, especially vpr1 mRNA and, (ii) a decrease of both the tat1 and the nef2 mRNA levels (Fig. 5A lanes 6-7, and 5C). We noted also a slight increase of the nef3, nef4, nef5 and tat4 mRNA levels when higher amounts of plasmid pXJ41-ASF/SF2 were used for the transfection. On the basis of the *in vitro* data (Fig. 1), the important increase found for vpr1 mRNA and the slight effects on nef4, nef5 and tat4 mRNAs can result from the strong activation of site A2 by protein ASF/SF2. Indeed, production of these 4 mRNAs implies a splicing event at site A2 (Fig. 5B). Whereas site A3 was found to be activated by protein ASF/SF2 *in vitro*, it was probably not activated by protein ASF/SF2 in HeLa cells, since the global amount of the tat mRNAs did not increase significantly upon overexpression of this protein. The global decrease of nef mRNAs upon ASF/SF2 overexpression confirmed the absence of activation of site A5 by this protein that we found *in vitro*.

Overexpression of proteins SRp40 and SC35 led to a huge increase of the tat1 mRNA, which likely reflects a very strong activation of the D1-A3 intron elimination (Fig. 5A lanes 10 to 13, and 5C). The tat2 mRNA level was also increased by protein SC35, but to a lower extent, suggesting that elimination of the D1-A3 intron was more favored than the combined elimination

of the D1-A1 and D2-A3 introns. When protein SRp40 was overexpressed, tat2 mRNA completely disappeared at the benefit of tat3 mRNAs (Fig. 5A lanes 12-13), indicating that protein SRp40 activated markedly site A3 and moderately site A2, which fits with our *in vitro* data (Fig. 1). The strong activation of site A3 by proteins SC35 and SRp40 resulted in a marked reduction of all other mRNA species, including the nef mRNAs (Fig. 4C). Only nef3 mRNA was moderately affected by the SC35 overexpression. However, in agreement with the incapacity of protein SRp40 to activate site A1, nef3 mRNA was strongly reduced by SRp40 overexpression. Taken together in HeLa cells, as found *in vitro* with the L3-U1 transcript (Figs. 2 and 3), site A3 was strongly activated by proteins SRp40 and SC35 at the expense of site A5. Protein 9G8 had the more limited effect on the HIV-1 RNA splicing pattern (Fig. 5A lanes 2-3, and 5B). Like proteins SRp40 and SC35, it increased tat1 mRNA synthesis, but the factor of activation was lower. It also slightly increased the synthesis of the tat2 and nef3 mRNAs, which might result from the activation of site A1. In contrast, nef4 mRNA, which reflects the use of site A2, was slightly decreased upon 9G8 addition.

The activation effects of proteins ASF/SF2 and 9G8 are preserved upon deletion of their activation RS domain

Because it has been shown that protein ASF/SF2, deleted of its RS domain, remains active *in vitro* in stimulating splicing of a HIV-1 tat pre-mRNA containing a shortened D4-A7 intron (17), we tested whether this ASF/SF2 deleted form expressed in transfected cells, preserves its activation property on the HIV-1 pre-mRNA produced from the Δ PSP plasmid. We also tested the activation properties of a Δ RS 9G8 protein. Interestingly, the deleted forms of ASF/SF2 and 9G8 had the same global effect as their wild type counterparts with, however, in case of protein

ASF/SF2, a slight decrease of the capacity to increase the vpr1-2 mRNA synthesis (Fig. 5A). We concluded that the presence of the RS domains of 9G8 and ASF/SF2 are not strictly required for their specific activation properties on site A3 for 9G8 and site A2 for ASF/SF2.

Overexpression of proteins ASF/SF2, SC35 and SRp40 decreases Nef protein expression

Finally, we determined whether the effects of SR proteins, that we observed on the alternative splicing of HIV-1 RNA, could also be detected at the protein level. To this end, after cotransfection assays using the Δ PSP plasmid and the highest amount of SR expressing plasmid, total proteins from transfected cells were extracted and analyzed by Western blot with antibodies directed against several viral proteins. Signals too weak to get quantitative results were obtained with specific antibodies directed against the Tat and Rev proteins. However, a monoclonal antibody directed against Nef, the more expressed protein during the early period of infection, gave with proteins of HIV-1 transfected cells a strong signal for a protein migrating as Nef (Fig 5D). Only a limited reduction of the Nef level was found upon protein 9G8 and SRp20 overexpression. In contrast, in agreement with results observed at the mRNA levels, ASF/SF2, SRp40 and SC35 overexpression resulted in a strong reduction of the Nef level. The effects on Nef protein even appeared to be stronger than that observed on its mRNAs, a result which can be due, at least in part, to the different technical approaches used for their respective detection.

Based on these *ex vivo* experiments, we concluded that the strong activation of site A2 by ASF/SF2 and the strong activation of site A3 utilization by SRp40 and SC35 found *in vitro* occur in cell culture and that the relative levels of SR proteins strongly modify the pattern of splicing of HIV-1 RNA and its protein expression as illustrated by the Nef protein analysis.

DISCUSSION

Splicing of RNAs from lentiviruses is a complex process because of the alternative utilization of several splicing sites including the 8 suboptimal 3' splice sites (1). It is now well established that SR proteins play a key role in the alternative utilization of 3' pre-mRNA splice sites in cellular messenger RNAs (for review, 29). Here, by *in vitro* splicing assays and by transfection experiments in HeLa cells, we demonstrated the specific effects of 4 of the SR proteins ASF/SF2, 9G8, SC35 and SRp40 on the 7 acceptor sites A1, A2, A3, A4a, b, c and A5 of HIV-1 RNA.

Both ASF/SF2 and SC35 activate the A1, A2 and A3 sites when these sites are tested individually

Two of the SR proteins tested, ASF/SF2 and SC35, are very general splicing activators. The two other ones 9G8 and SRp40 are more specific regulators. For instance, 9G8 was found to enhance the splicing of its own messenger (48) and that of the 12S adenovirus E1A transcript (43) and SRp40 was found to be involved in alternative splicing of the cardiac troponin T pre-mRNA (49). Accordingly, ASF/SF2 and SC35 were found to activate the three A1, A2 and A3 sites, when these sites were tested individually (ASF/SF2 and SC35 having the highest activity on sites A2 and A3, respectively) (Fig. 1). In contrast, protein 9G8 had only a limited activation effect, and protein SRp40 had low, medium and strong effects on sites A1, A2 and A3, respectively. As SRp40 did not complement a S100 extract for splicing at site A3 and since its capacity to activate site A3 was augmented when the amount of nuclear extract versus S100 extract was

increased, activation of site A3 by SRp40 probably depends upon one or more nuclear component(s).

***Ex vivo* ASF/SF2 essentially activates site A2, whereas SC35 and SRp40 activate nearly exclusively site A3.**

Whereas site A3 was activated efficiently by ASF/SF2 in the context of the C3 construct, its activation by this SR protein was modest in the L3-U1 RNA and was not detected *ex vivo*. This difference may be due to preferential targeting of ASF/SF2 near site A2 at the expense of site A3 in HIV-1 RNA with all the splicing sites. The competition between sites A2 and A1 for ASF/SF2 should be less strong than that between sites A2 and A3, since the level of mRNAs in HeLa cells, that were spliced at site A1, but not at the A2 site (*tat2*, *nef3*), was slightly increased upon overexpression of ASF/SF2 in HeLa cells. In HeLa cells, SC35 and SRp40 proteins activate nearly exclusively site A3, revealing that the other acceptor sites are subjected to a strong competition with site A3 in these conditions. In this case the *in vitro* data on the L3-U1 RNA and the *ex vivo* data are very similar. Although more modest, 9G8 also has a significant activity on site A3 *ex vivo*. Hence, three of the tested SR proteins activate preferentially site A3 *ex vivo*.

The HIV-1 splicing sites which are the more sensitive to SR activation were previously found to be controlled by ESS elements

Our data show that 2 of the 7 HIV-1 acceptor sites tested (A2 and A3) are highly sensitive to a control by SR protein. The less efficient control by SR proteins at site A1 may be explained by the higher intrinsic efficiency of this site, as evidenced *in vitro* (Fig. 1). Among the three A4a, b and c sites, only A4b is subjected to a moderate regulation by SR proteins. When site A2

utilization was strongly increased by ASF/SF2, or site A3 utilization was strongly activated by SC35 or SRp40, site A5 utilization was markedly decreased. This suggests that site A5 is at least in part regulated by default.

Remarkably, sites A2 and A3, whose activities are strongly dependent upon SR proteins, were previously found to be subjected to negative controls by splicing silencer elements. Three distinct hnRNP A/B binding sites were identified downstream from site A2 and have been found to be responsible for the ESSV inhibitory property (12). As found for site A7 (19,21), ASF/SF2 may activate site A2 by competing the binding of hnRNP A/B to ESSV. Two distinct inhibitory elements binding hnRNP H and hnRNP A1, respectively were found at site A3 (7-9,13,14,16). Interestingly, SC35 and SRp40 have a strong activation capacity on site A3 in the C3 RNA, that contains only one of these inhibitory element (ESS2p), but also on the L3-U1 RNA that contains the two inhibitory elements ESS2 and ESS2p (Figs. 1 and 2). Hence, proteins SC35 and SRp40 are both able to counteract the inhibitory effect of these elements. The question is to know whether they compete hnRNP protein binding to the ESS2p and ESS2 elements. We found that A1 and A4b are also, but to a lesser extent, activated by ASF/SF2, SC35 and 9G8 *ex vivo*. It will also be interesting to test whether these two sites are also controlled by ESS elements.

Protein 9G8 recognizes specific RNA sequences, mainly based on repetitions of GAC triplets (42,50), which may be found in ESE responding to this protein (43,48,51). Such typical sequences cannot be identified in the HIV-1 RNA segment containing sites A1 to A5. This probably explains the modest activation property found for this protein. Protein SRp40 has a define specificity for a long RNA sequence UGGGAGCRGUYRGCUGGY, but it has not been shown whether this entire sequence is required for a full activity (52). Protein SRp40 also has a more broad specificity for RNA (53). Again, no sequence homologous to the specific SRp40

recognition motif was detected in the HIV-1 RNA segment containing sites A1 to A5, which may explain why the action of SRp40 was dependent upon the presence of a nuclear component. Finally, ASF/SF2 and SC35 each recognizes a large variety of sequences (42,53,54) and several putative binding sites for these proteins are present close to the A1, A2 and A3 binding sites, probably explaining why these 3 sites were activated by these two proteins in the C1, C2 and C3 transcripts. Specific analyses will be required to determine with precision the identity of the splicing enhancers which allow the activation of the A1 to A3 splicing sites. In particular, it will be important to clarify whether SC35 and SRp40, which strongly activate the A3 splice site, interact or not with the same *cis*-element of the downstream exon.

Interestingly, as previously found for site A7 (17,19), the activation property of ASF/SF2 on site A2 is not strictly dependent upon the presence of the RS domain. The same is true for 9G8. This reinforced the idea that, at least in part, the action of ASF/SF2 consists in competing the binding of inhibitory proteins on the ESS elements, rather than facilitating directly spliceosome assembly. At site A2, ASF/SF2 may compete hnRNP A1 binding at ESSV, and at site A3, SC35 or SRp40 may compete binding of hnRNP A1 at ESS2 or binding of hnRNP H at ESS2p. Taken together, the present data and previous data on site A7 (17,19,21,36,37,55) show the strong sensitivity of two HIV acceptor sites A2 and A7 to ASF/SF2 and of a third one (A3) to a series of SR proteins (SC35, SRp40, 9G8).

Variations of SR protein composition is expected to alter HIV protein expression in infected cells

Based on the very different effects of ASF/SF2 on the one hand and SC35 and SRp40 on the other hand on the pattern of mRNAs produced from HIV-1 RNA, we can anticipate that

variations of the relative ratios of SR proteins in infected cells will strongly modify the pattern of protein expression. We could indeed demonstrate the strong reduction of Nef protein production upon overexpression of proteins ASF/SF2, SRp40 and especially SC35. Thus, if ASF/SF2 is present in excess in the infected cell, the Vpr protein is expected to be produced in larger amounts, whereas the Nef protein level is expected to decrease. Protein Vpr is required for virus multiplication in non-dividing cells because of its action at an early step of nuclear import of the preintegration complex (56). Protein Vpr also induces cell cycle arrest in proliferating cells, stimulates virus transcription and regulates activation and apoptosis of infected cells (for review, 28). The Nef protein whose level might be decreased is for a large part responsible for the AIDS pathology (for review, 57). Indeed, by activating the production of Fas-L at the surface of the infected cells, Nef participates to the apoptosis of bystander cytotoxic T lymphocytes (58,59). On the contrary, Nef protects the infected cell from apoptosis by interacting with the ASK1 kinase that is needed for transduction of the apoptosis signal (60). Hence, a high amount of ASF/SF2 compared to other SR proteins in infected cells may modify the fate of the infected cell and surrounding lymphocytes. If SC35 or SRp40 are present in excess compared to other SR proteins in the infected cell, the Tat protein is expected to be produced in large amounts. Taking into account the central role played by the Tat protein in transcription of full length HIV-1 RNA (for review, 61), an increase of the SC35 and SRp40 concentration or activity in the infected cell may induce the entry in a viral multiplication cycle. Interestingly, Maldarelli *et al.* recently observed a strong increase of the SC35 concentration after infection of T-cell by HIV-1 virions (38). This increase might favor tat mRNA production and facilitate the initiation of a cycle of virus multiplication. However, this is probably not a general feature since virus HIV-1 is known to be able to remain silent for a long period of time in some infected cell (62). In connection to this

point, it should be noticed that in order to be active in transcription, Tat has to be acetylated at position Lys28 by the HAT acetylase (63).

ACKNOWLEDGMENTS

C.M. Stoltzfus from Iowa University is strongly thanked for his generous gift of plasmid Δ PSP.

F. Clavel is acknowledged for providing plasmid pBru3. We thank the staff of the IGBMC facilities for their assistance.

This work was supported by the french Agence Nationale de Recherche sur le SIDA (ANRS), the Centre National de la Recherche Scientifique, the french Ministère de la Jeunesse, de l'Education Nationale et de la Recherche (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), SIDACTION and local funds from the Région Lorraine.

D. Ropers was a fellow of the french Ministère de la Jeunesse, de l'Education Nationale et de la Recherche, L. Ayadi is a fellow of the french Agence Nationale de Recherche sur le SIDA (ANRS), L. Damier was a fellow of the SIDACTION foundation.

REFERENCES

1. Purcell, D. F., and Martin, M. A. (1993) *J Virol* **67**, 6365-6378.
2. Benko, D. M., Schwartz, S., Pavlakis, G. N., and Felber, B. K. (1990) *J Virol* **64**, 2505-2518.
3. O'Reilly, M. M., McNally, M. T., and Beemon, K. L. (1995) *Virology* **213**, 373-385.
4. Damier, L. (1997) in *Ecole Doctorale Biologie-Santé UFR STB*, Université Henri Poincaré, Nancy, France
5. Staffa, A., and Cochrane, A. (1994) *J Virol* **68**, 3071-3079.
6. Staffa, A., and Cochrane, A. (1995) *Mol Cell Biol* **15**, 4597-4605.
7. Amendt, B. A., Hesslein, D., Chang, L. J., and Stoltzfus, C. M. (1994) *Mol Cell Biol* **14**, 3960-3970.
8. Amendt, B. A., Si, Z. H., and Stoltzfus, C. M. (1995) *Mol Cell Biol* **15**, 4606-4615.
9. Si, Z., Amendt, B. A., and Stoltzfus, C. M. (1997) *Nucleic Acids Res* **25**, 861-867.
10. Si, Z. H., Rauch, D., and Stoltzfus, C. M. (1998) *Mol Cell Biol* **18**, 5404-5413.
11. Jacquenet, S., Mereau, A., Bilodeau, P. S., Damier, L., Stoltzfus, C. M., and Branlant, C. (2001) *J Biol Chem* **276**, 40464-40475.
12. Bilodeau, P. S., Domsic, J. K., Mayeda, A., Krainer, A. R., and Stoltzfus, C. M. (2001) *J Virol* **75**, 8487-8497.
13. Bilodeau, P. S., Domsic, J. K., and Stoltzfus, C. M. (1999) *J Virol* **73**, 9764-9772.
14. Caputi, M., Mayeda, A., Krainer, A. R., and Zahler, A. M. (1999) *EMBO J* **18**, 4060-4067.
15. Caputi, M., and Zahler, A. M. (2002) *EMBO J* **21**, 845-855.

16. Del Gatto-Konczak, F., Olive, M., Gesnel, M. C., and Breathnach, R. (1999) *Mol Cell Biol* **19**, 251-260.
17. Tange, T. O., and Kjems, J. (2001) *J Mol Biol* **312**, 649-662.
18. Tange, T. O., Damgaard, C. K., Guth, S., Valcarcel, J., and Kjems, J. (2001) *EMBO J* **20**, 5748-5758.
19. Zhu, J., Mayeda, A., and Krainer, A. R. (2001) *Mol Cell* **8**, 1351-1361.
20. Pongoski, J., Asai, K., and Cochrane, A. (2002) *J Virol* **76**, 5108-5120.
21. Marchand, V., Mereau, A., Jacquenet, S., Thomas, D., Mougin, A., Gattoni, R., Stevenin, J., and Branlant, C. (2002) *J Mol Biol* **323**, 629-652.
22. Dyhr-Mikkelsen, H., and Kjems, J. (1995) *J Biol Chem* **270**, 24060-24066.
23. Damier, L., Domenjoud, L., and Branlant, C. (1997) *Biochem Biophys Res Commun* **237**, 182-187.
24. Jacquenet, S., Ropers, D., Bilodeau, P. S., Damier, L., Mougin, A., Stoltzfus, C. M., and Branlant, C. (2001) *Nucleic Acids Res* **29**, 464-478.
25. Boris-Lawrie, K., Roberts, T. M., and Hull, S. (2001) *Life Sci* **69**, 2697-2709.
26. Harris, M. (1999) *Curr Biol* **9**, R459-461.
27. Azad, A. A. (2000) *Biochem Biophys Res Commun* **267**, 677-685.
28. Bukrinsky, M., and Adzhubei, A. (1999) *Rev Med Virol* **9**, 39-49.
29. Graveley, B. R. (2000) *RNA* **6**, 1197-1211.
30. Smith, C. W., and Valcarcel, J. (2000) *Trends Biochem Sci* **25**, 381-388.
31. Manley, J. L., and Tacke, R. (1996) *Genes Dev* **10**, 1569-1579.
32. Hastings, M. L., and Krainer, A. R. (2001) *Curr Opin Cell Biol* **13**, 302-309.
33. Molin, M., and Akusjarvi, G. (2000) *J Virol* **74**, 9002-9009.

34. Himmelspach, M., Cavaloc, Y., Chebli, K., Stevenin, J., and Gattoni, R. (1995) *RNA* **1**, 794-806.
35. Kanopka, A., Muhlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C., and Akusjarvi, G. (1998) *Nature* **393**, 185-187.
36. Fu, X. D., Mayeda, A., Maniatis, T., and Krainer, A. R. (1992) *Proc Natl Acad Sci U S A* **89**, 11224-11228.
37. Mayeda, A., Sreaton, G. R., Chandler, S. D., Fu, X. D., and Krainer, A. R. (1999) *Mol Cell Biol* **19**, 1853-1863.
38. Maldarelli, F., Xiang, C., Chamoun, G., and Zeichner, S. L. (1998) *Virus Res* **53**, 39-51.
39. Ryo, A., Suzuki, Y., Arai, M., Kondoh, N., Wakatsuki, T., Hada, A., Shuda, M., Tanaka, K., Sato, C., Yamamoto, M., and Yamamoto, N. (2000) *AIDS Res Hum Retroviruses* **16**, 995-1005.
40. Charneau, P., Alizon, M., and Clavel, F. (1992) *J Virol* **66**, 2814-2820.
41. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., and et al. (1985) *Nature* **313**, 277-284.
42. Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999) *RNA* **5**, 468-483.
43. Bourgeois, C. F., Popielarz, M., Hildwein, G., and Stevenin, J. (1999) *Mol Cell Biol* **19**, 7347-7356.
44. Chen, C., and Okayama, H. (1987) *Mol Cell Biol* **7**, 2745-2752.
45. Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992) *Genes Dev* **6**, 837-847.
46. Gallego, M. E., Gattoni, R., Stevenin, J., Marie, J., and Expert-Bezancon, A. (1997) *EMBO J* **16**, 1772-1784.

47. Tacke, R., Tohyama, M., Ogawa, S., and Manley, J. L. (1998) *Cell* **93**, 139-148.
48. Lejeune, F., Cavaloc, Y., and Stevenin, J. (2001) *J Biol Chem* **276**, 7850-7858.
49. Ramchatesingh, J., Zahler, A. M., Neugebauer, K. M., Roth, M. B., and Cooper, T. A. (1995) *Mol Cell Biol* **15**, 4898-4907.
50. Schaal, T. D., and Maniatis, T. (1999) *Mol Cell Biol* **19**, 1705-1719.
51. Cramer, P., Caceres, J. F., Cazalla, D., Kadener, S., Muro, A. F., Baralle, F. E., and Kornblihtt, A. R. (1999) *Mol Cell* **4**, 251-258.
52. Tacke, R., Chen, Y., and Manley, J. L. (1997) *Proc Natl Acad Sci USA* **94**, 1148-1153.
53. Liu, H. X., Zhang, M., and Krainer, A. R. (1998) *Genes Dev* **12**, 1998-2012.
54. Tacke, R., and Manley, J. L. (1995) *EMBO J* **14**, 3540-3551.
55. Krainer, A. R., Conway, G. C., and Kozak, D. (1990) *Genes Dev* **4**, 1158-1171.
56. 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) *EMBO J* **17**, 909-917.
57. Arora, V. K., Fredericksen, B. L., and Garcia, J. V. (2002) *Microbes Infect* **4**, 189-199.
58. Xu, X. N., Sreaton, G. R., Gotch, F. M., Dong, T., Tan, R., Almond, N., Walker, B., Stebbings, R., Kent, K., Nagata, S., Stott, J. E., and McMichael, A. J. (1997) *J Exp Med* **186**, 7-16.
59. Zauli, G., Gibellini, D., Secchiero, P., Dutartre, H., Olive, D., Capitani, S., and Collette, Y. (1999) *Blood* **93**, 1000-1010.
60. Geleziunas, R., Xu, W., Takeda, K., Ichijo, H., and Greene, W. C. (2001) *Nature* **410**, 834-838.
61. Karn, J. (1999) *J Mol Biol* **293**, 235-254.
62. Tang, H., Kuhen, K. L., and Wong-Staal, F. (1999) *Annu Rev Genet* **33**, 133-170

63. Kiernan, R. E., Vanhulle, C., Schiltz, L., Adam, E., Xiao, H., Maudoux, F., Calomme, C., Burny, A., Nakatani, Y., Jeang, K. T., Benkirane, M., and Van Lint, C. (1999) *EMBO J* **18**, 6106-6118.

FIGURE LEGENDS

Figure 1: Differential effects of individual SR proteins on *in vitro* utilization of HIV-1 sites A1, A2 and A3 in the C1, C2 and C3 RNAs

(A) Schematic representation of the C1, C2 and C3 pre-mRNAs containing the donor site D1 and the acceptor sites A1, A2 and A3. Boxes represent exons and lines represent introns. The positions in the HIV-1 BRU RNA (41) of the extremities of the HIV-1 sequences, which are joined together in the transcripts (*Bgl* II site in the cDNA), are indicated above the constructs. The various splicing sites and splicing regulatory elements that were previously identified (11,12) are also indicated.

(B), (C) and (D) *In vitro* splicing assays. The uniformly labeled C1 (B), C2 (C) or C3 (D) transcripts were incubated 150 min in 8 μ l of buffer D (lanes 1, and lanes 1 in the insets in Panels B, C and D), a mixture of nuclear (NE) and cytoplasmic (S100) extracts (NE/S100 4 μ l:4 μ l) or pure S100 extract, under splicing conditions, in the absence (lanes 2-3) or the presence of 300 ng of the individual SR protein 9G8 (lanes 4-5), ASF/SF2 (lanes 6-7), SC35 (lanes 8-9) or SRp40 (lanes 10-11). The transcripts were also incubated for 150 min in the presence of 8 μ l of HeLa cell nuclear extract (+) (insets in Panel B, C and D). The products of the reaction were separated by electrophoresis in a 5% polyacrylamide sequencing gel. Positions of pre-mRNAs

(P), exon 1 (E), lariat splicing intermediates (L) and splicing products (M) are indicated on the left of the panels. Lanes M correspond to DNA length markers. The length of the DNA fragments in bps are given on the right of the panels.

(E1) and (E2) Histograms showing the splicing efficiencies of the C1, C2 and C3 RNAs in S100 extract (Panel E1) or in the 4:4 NE/S100 mixture (Panel E2), that were estimated by determination of the M/P values, as described in the experimental procedures. The bar errors give an estimation of the M/P measurement errors.

(E3) and (E4) Histograms showing the activation capacity of SR proteins for the C1, C2 and C3 RNAs. The activation factors (fold of activation) were calculated by dividing the M/P values obtained upon addition of recombinant SR protein by the M/P values found without their addition. The activation factors in the 4:4 NE/S100 mixture calculated from data in Panel B, C and D are given in Panel E3. Activation factors in the 1:7 NE/S100 mixture (gel electrophoresis not shown) are given in Panel E4. The bar errors represent the errors on the activation factors.

Figure 2: Differential effects of individual SR proteins on the *in vitro* splicing efficiency of the L3-U1 RNA in S100 extract or in the NE/S100 mixtures

(A) Schematic representation of the L3-U1 pre-mRNA containing the donor site D1 and the acceptor sites A3 to A5. Boxes represent exons and lines represent introns. The position in the HIV-1 BRU RNA of the extremities of the HIV-1 sequences (41), which are joined together in the transcripts, are indicated above the construct (*Bgl* II site used for the junction of the two cDNA fragments). The various splicing sites and splicing regulatory elements that were previously identified are also represented (7-9,11,14,16).

(B) *In vitro* splicing assays with uniformly labeled L3-U1 RNA incubated for 150 min in 8 μ l of buffer D (lane 1), a mixture of nuclear (NE) and cytoplasmic S100 extract (NE/S100 4 μ l: 4 μ l) or pure S100 extract, under splicing conditions, in the absence (lanes 2-3), or the presence of 300 ng of the recombinant SR protein 9G8 (lanes 4-5), ASF/SF2 (lanes 6-7), SC35 (lanes 8-9) or SRp40 (lanes 10-11), respectively. The splicing products were fractionated by electrophoresis in a 5% polyacrylamide sequencing gel. Positions of the precursor (P), first exon (E) and the various maturation products spliced at sites A3 (MA3), A4c (MA4c), A4a, A4b (MA4a, b) or A5 (MA5) are indicated on the left of the panel. Lane M corresponds to a DNA size marker. The size of the fragments in bps are indicated on the left of the panel.

(C1), (C2) Histograms representing the splicing efficiencies at site A3 in S100 extract (Panel C1) and sites A3, A4a, A4b and A5 in the 4:4 NE/S100 mixture, with or without addition of recombinant SR protein. These efficiencies were estimated by dividing the amount of radioactivity found in the MA3, MA4a, b and MA5 splicing products, by that found in the residual precursor (P), respectively.

(C4), (C4) Histograms showing the factor of activation by SR proteins in the 4:4 (Panel C3) or the 1:7 (Panel 4) NE/S100 mixture. The factors of activation in Panel C3 were calculated from data shown in Panel B, the factors of activation in Panel C2 were calculated from data which are not shown.

Figure 3: Differential effects of individual SR proteins on the *in vitro* splicing efficiency of the L3-U1 RNA in nuclear extract

(A) *In vitro* splicing assays with uniformly labeled L3-U1 RNA. Assays were performed in 8 μ l of nuclear extract for 150 min in the absence or the presence of the SR proteins 9G8, ASF/SF2, SC35 or SRp40.

(B) Effect of SR protein concentration on splicing efficiency. Transcript L3-U1 was spliced in a nuclear extract in the presence of 150 or 300 ng of the SR proteins SC35 or SRp40. In both Panels A and B, the positions of the lariat intermediates (L), precursor RNAs (P) and maturation products (MA3, MA4a, b and MA5) are shown on the left of the panel. Lane M is a DNA size marker.

(C1) Histogram showing the M/P values at sites A3 and A5 in the experiment shown in Panel A. Bar errors represent the estimated errors of M/P values, linked to the measurement of the radioactivity in the bands of gel.

(C2) and (C3) Factors of activation by SR proteins observed in the experiments shown in Panels A and B, respectively.

Figure 4: Overexpression of the SR proteins in the HeLa cell transfection experiments

(A) Western-blot analysis of all the SR proteins. Proteins were extracted from transfected cells, 48 hours after the transfection and for each transfection experiment, equal amounts of total proteins were fractionated by SDS-PAGE. Proteins were transferred onto nitrocellulose filters and the presence of SR proteins was tested by Western-blot analysis using the 10H3 antibody that recognizes the endogenous SR proteins except for SC35. The identity of the endogenous SR proteins revealed by the antibody are given on the left side of the panel, while the identities of the SR proteins expressed from the transfected genes are indicated on the right side of the Panel.

(B) Western-blot analysis of SC35 protein. The same experiment was repeated with an antibody specific for protein SC35 (42).

(C) Histogram showing the level of overexpression of the recombinant SR proteins 9G8, ASF/SF2, SC35, SRp40 and SRp20 compared to the endogenous expression of the corresponding proteins. The amounts of radioactivity in the band corresponding to the recombinant protein was divided by that found in the band corresponding to the endogenous protein.

Figure 5: Differential effects of the overexpression of individual SR proteins on HIV-1 RNA alternative splicing in HeLa cells

(A) RT-PCR analysis of HIV-1 spliced products. HeLa cells were cotransfected with plasmid Δ PSP carrying all the HIV-1 splicing sites (24) and a pXJ plasmid derivative overexpressing no (-) or one recombinant WT or mutated SR protein. For each SR protein, two different amounts of plasmid pXJ derivative (100 and 400 ng in A and 50 and 250 ng in B) were used for the cotransfection. For each transfection assay, an equal amount of total RNA was subjected to RT-PCR analysis using the specific HIV-1 primers given in the Experimental Procedures and indicated in Panel B. Lanes marked by M correspond to DNA size markers. The sizes in bps of the DNA fragments are indicated directly on the marker lanes. The identity of the HIV-1 mRNAs corresponding to the fractionated cDNAs are indicated on the left of the 2 Panels. Panel B shows the exon composition of the HIV-1 mRNAs, below a schematic representation of HIV-1 genome. For each mRNA, thick lines represent the exons, their identity are given on the right of the scheme. Positions of the HIV-1 sequence complementary to oligonucleotide ACC15 used for reverse transcription and of oligonucleotides 2143 and 2144 used for PCR amplifications are

indicated below the HIV-1 RNA schematic representation. (C) Variations of nef2, 3 and 4 mRNAs and tat1 mRNA levels in transfected cells depending on the identity and the amount of the SR protein overexpressed. The amounts of nef2, 3 and 4 mRNAs and tat1 mRNA in the experiment shown in Panel B was estimated by measurement of the radioactivity in each band using a Phosphorimager. (D) Western-blot analysis of the amount of Nef protein expressed in the transfected cells. Total proteins were extracted after 48 hours of transfection as described in the Experimental Procedures and for each assay equal amounts of total proteins were fractionated by SDS PAGE. The amount of Nef protein was estimated by Western-blot analysis using a Nef specific antibody (MaTG0020). Quantification of the radioactivity allowed us to determine the percentage of reduction of Nef expression upon SR protein overexpression.

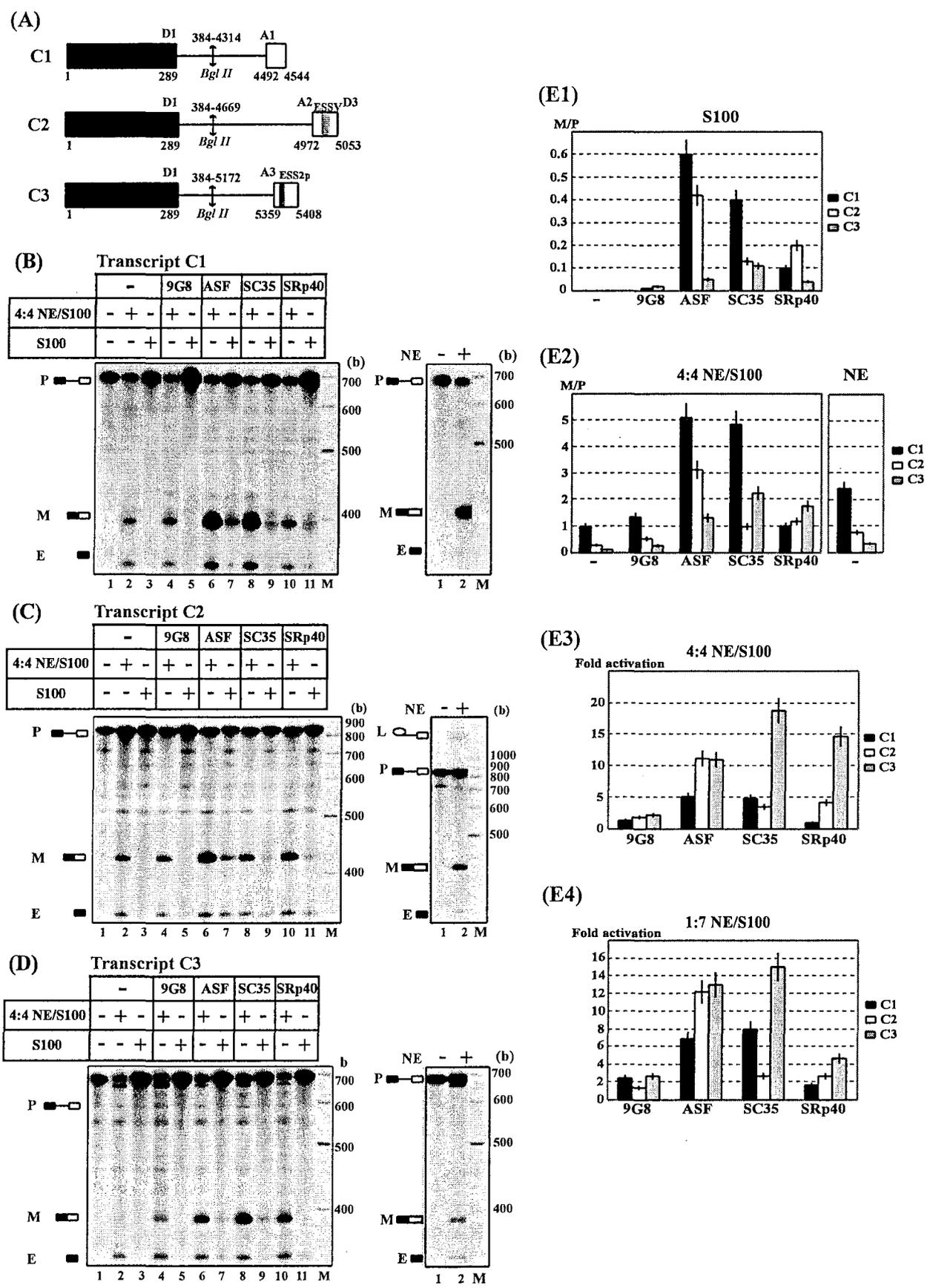


Figure 1

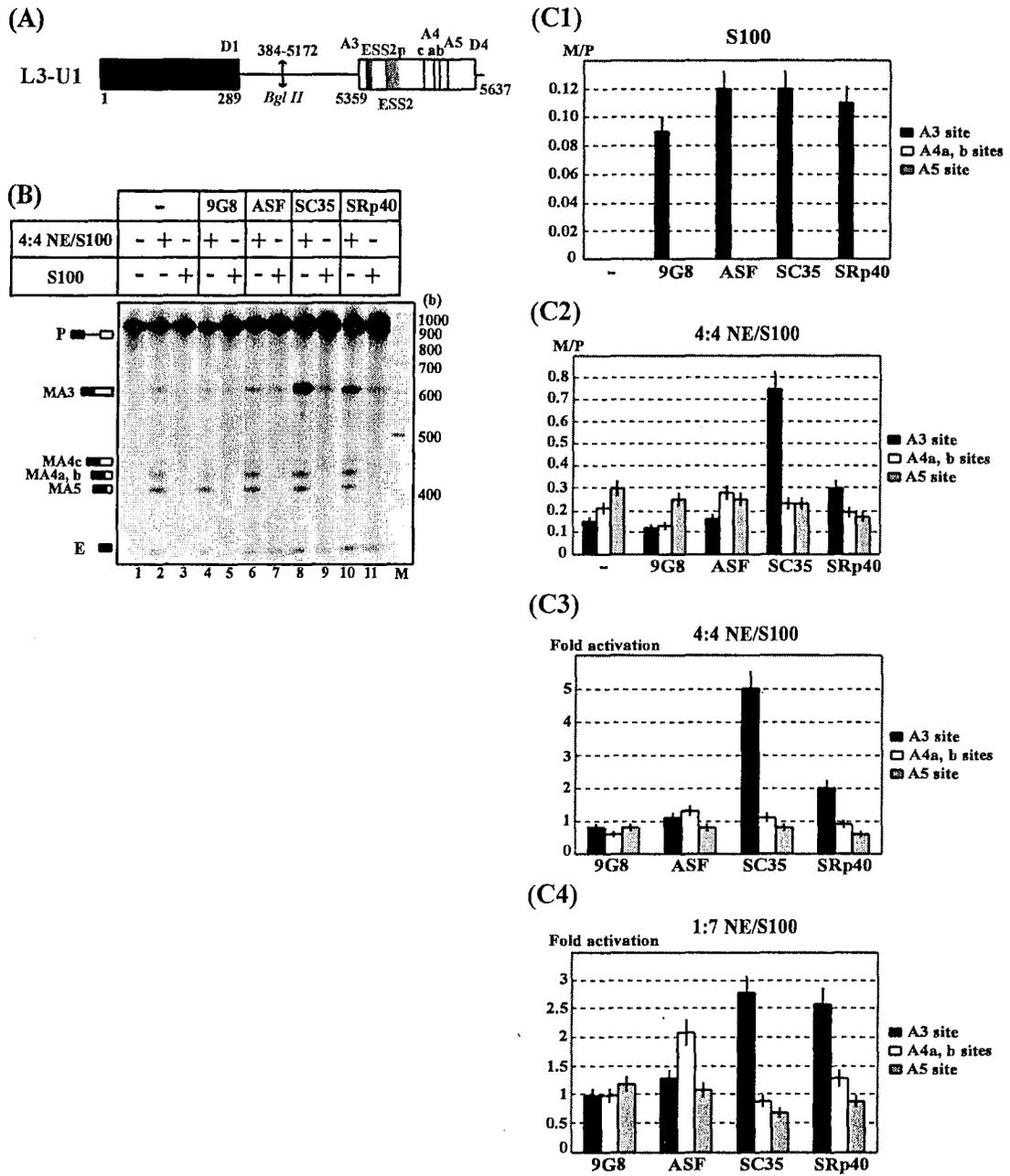


Figure 2

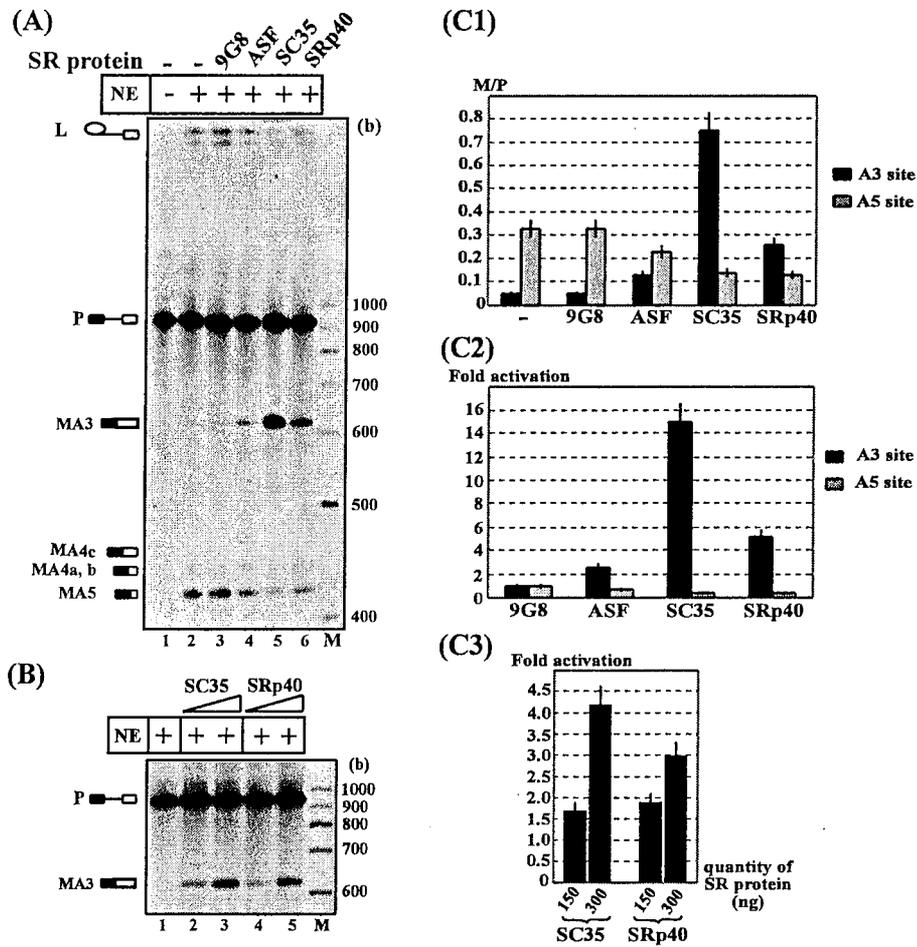


Figure 3

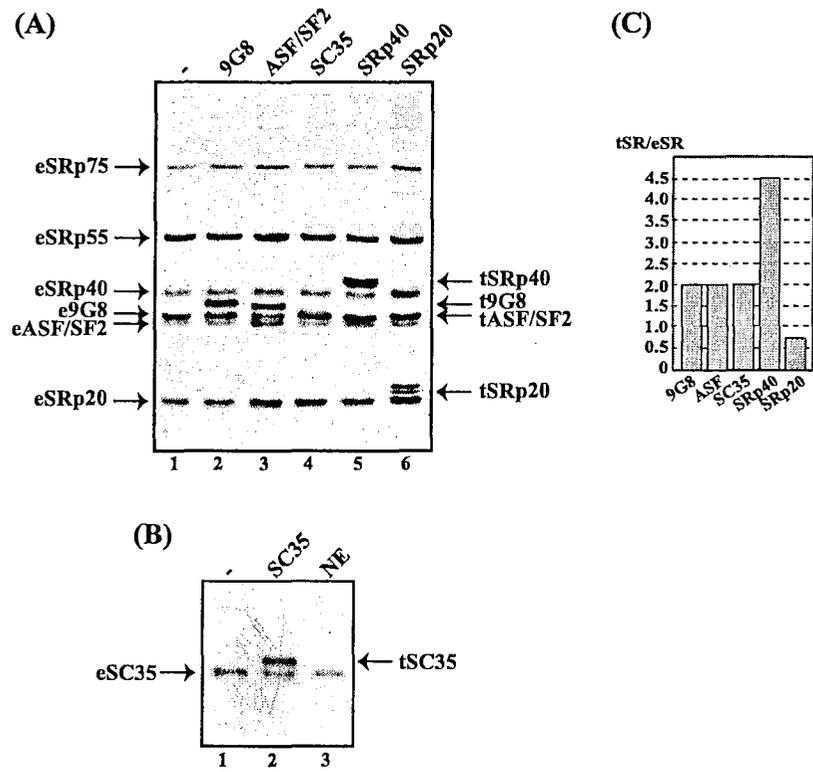


Figure 4

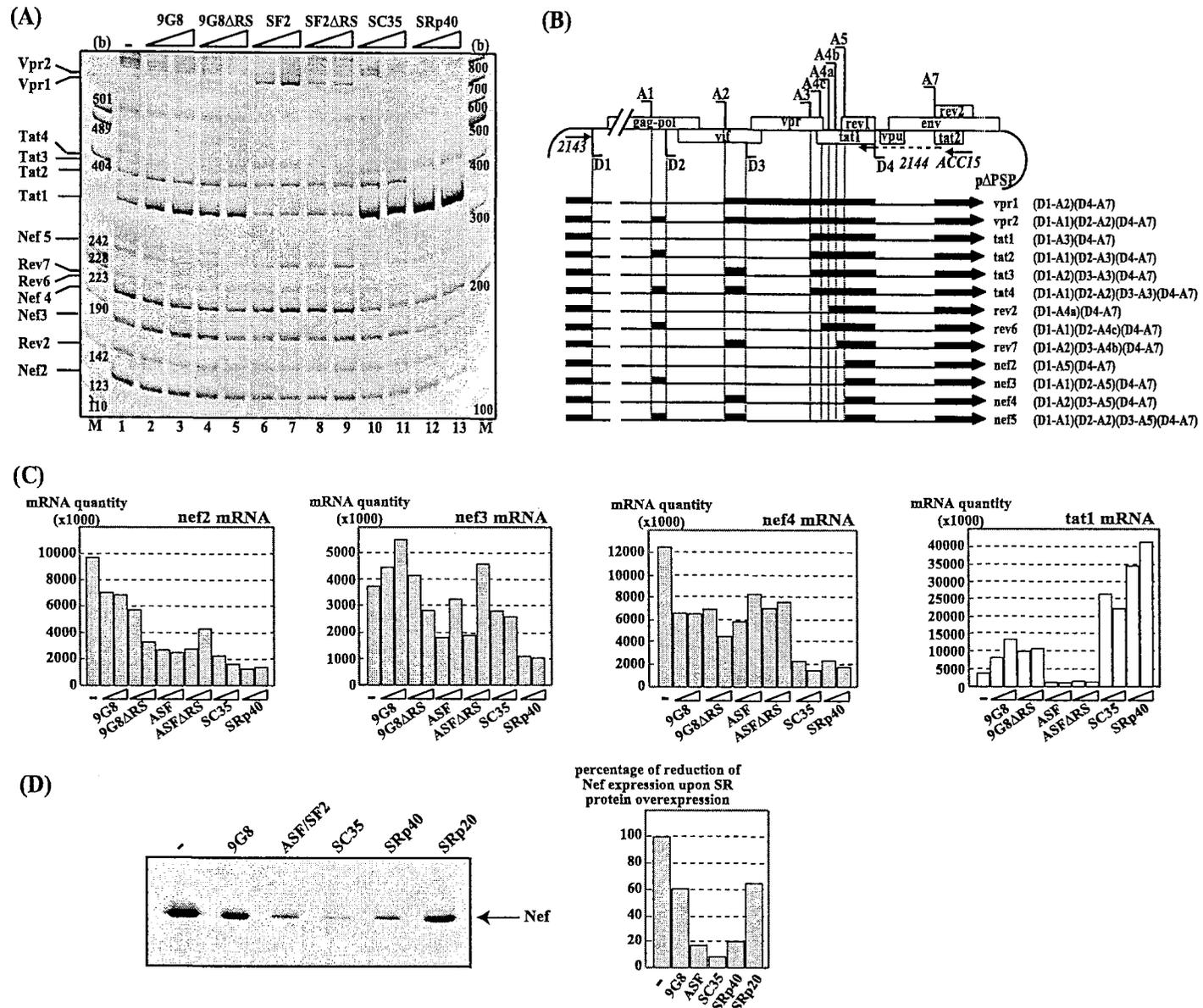


Figure 5

IV.2.6 Expériences complémentaires

IV.2.6.1 Résultats d'épissage *in vitro* obtenus dans le mélange EN/S100 1/7

Comme nous l'avons indiqué dans l'article, nous avons réalisé des expériences d'épissage *in vitro* dans un mélange d'extrait nucléaire et d'extrait S100 en proportions 1/7 (v/v). Pour ne pas surcharger l'article en figures et pour des raisons de simplification, ces données n'avaient pas été illustrées. Nous les présentons donc ici pour les différents ARN C1, C2, C3 et L3-U1 (Fig. IV.4 et IV.5). Pour comparaison, nous présentons également les données d'épissage obtenues dans le mélange EN/S100 4/4.

La figure IV.4 montre les 3 autoradiogrammes obtenus (panneaux A, B et C), ainsi que les valeurs d'efficacité d'épissage M/P présentées sous forme d'histogrammes (panneau D) et les taux d'activation par les différentes protéines SR que nous en avons déduit (panneau E).

Dans le mélange 1/7, les 3 protéines 9G8, ASF/SF2 et SC35 activent l'épissage au site A1 (histogramme supérieur dans le panneau E), plus fortement que dans le mélange 4/4 (histogramme inférieur). Alors que SRp40 n'a pas d'effet activateur dans le mélange 4/4, elle active faiblement l'utilisation de ce site dans le mélange 1/7. Ceci pouvait s'expliquer, soit par la présence dans l'extrait nucléaire d'un composant inhibant l'action des 4 protéines SR testées, soit par la présence d'une quantité déjà suffisamment élevée de protéines SR dans l'extrait, pour permettre l'utilisation du site A1.

En ce qui concerne le site A2, le plus fort effet activateur est observé pour la protéine ASF/SF2 et ce, quelque soit le rapport EN/S100.

De manière intéressante, alors que les taux d'activation observés au site A3 pour 9G8, SC35 et ASF/SF2 sont comparables entre les 2 mélanges, l'effet activateur de SRp40 est fortement stimulé par l'ajout d'extrait nucléaire. Ceci suggère, comme nous l'avons proposé dans l'article, que cette activation nécessite la présence d'autres composants de l'extrait nucléaire.

Les données présentées dans la figure IV.4 suggèrent que les protéines SR présentent non seulement des spécificités de sites d'épissage différentes, mais également une dépendance différente vis-à-vis d'autres composants nucléaires.

La figure IV.5 montre l'autoradiogramme obtenu avec l'ARN L3-U1 (panneau A), ainsi que les valeurs d'efficacité d'épissage M/P (panneau B) et les taux d'activation des sites A3, A4a, b et A5, présentées sous forme d'histogrammes. Dans un contexte où les sites accepteurs sont en compétition, on observe toujours une meilleure activation du site A3 par la protéine SC35, lorsque la quantité d'extrait nucléaire augmente, alors que les taux d'activation ne varient pas pour les autres protéines SR en fonction du mélange (panneau C). Dans l'expérience illustrée, l'ajout d'extrait nucléaire a un effet moins marqué sur l'activation du site A3 par SRp40, que celle que nous observions *ex vivo* (Fig. 5C de l'article). On ne peut cependant pas exclure un problème lié à l'expérimentation.

L'activation du site A3 par ASF/SF2, dans un contexte où ce site d'épissage est en compétition avec les sites en aval, est très faible dans le mélange EN/S100 1/7 et elle diminue même légèrement en fonction de la quantité d'extrait nucléaire ajoutée (panneau C). Cette observation est en accord avec les données obtenues *ex vivo* (cf Fig. 5 de l'article).

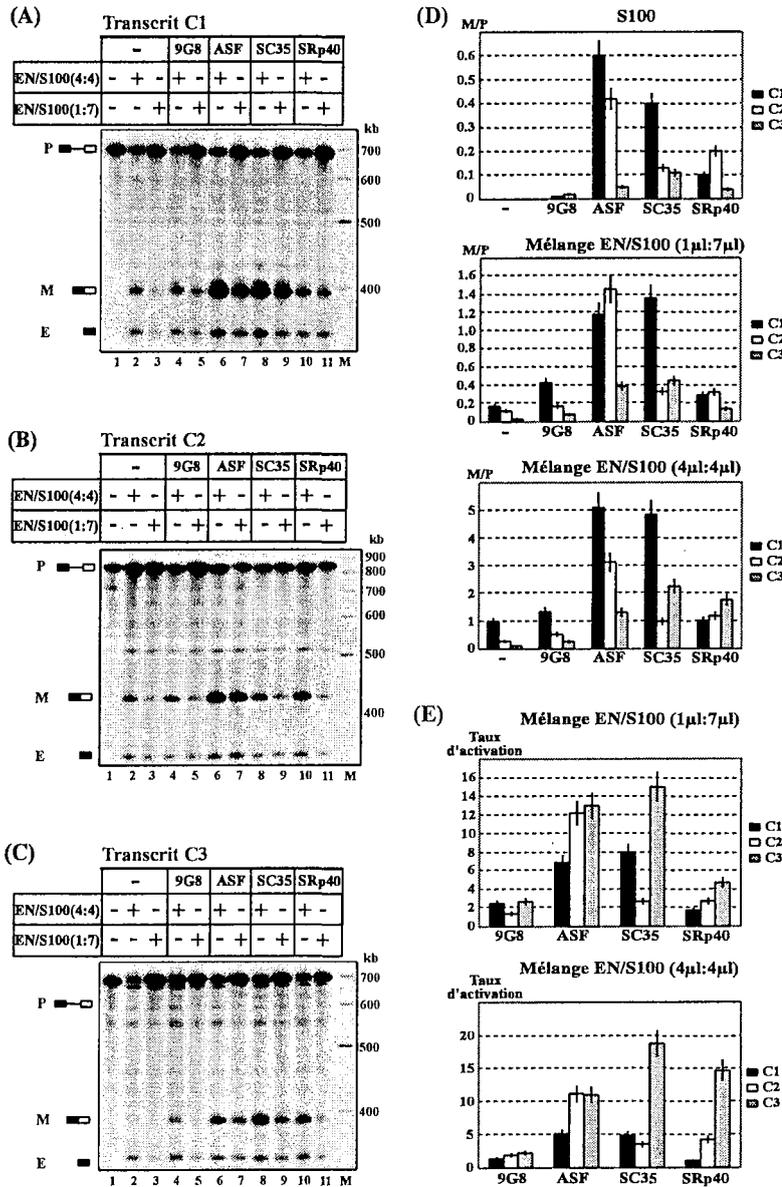


FIG. IV.4 – Effet des protéines SR sur l'épissage des ARN C1, C2 et C3 dans différents mélanges d'extraits nucléaire et cytoplasmique.

(A), (B) et (C) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% produits d'épissage des transcrits C1 (A), C2 (B) et C3 (C) uniformément marqués. Les transcrits ont été incubés 2h30 dans un mélange d'extraits nucléaire et cytoplasmique en proportions 4/4 et 1/7, en absence ou en présence de 300 ng de protéine SR recombinante 9G8, ASF/SF2, SC35 ou SRp40, dans les conditions décrites dans la partie matériel et méthodes. Les contrôles (pistes 1) correspondent aux transcrits non incubés. La position des précurseurs, produits matures et exon 1 est indiquée à gauche des autoradiogrammes.

(D) Histogrammes dans lesquels sont représentées les valeurs des rapports M/P obtenues dans les mélanges EN/S100 1/7 et 4/4, estimées par mesure de la radioactivité au phosphorimager.

(E) Histogrammes dans lesquels sont représentés les valeurs des taux d'activation des sites d'épissage en présence des protéines SR pour les mélanges EN/S100 1/7 et 4/4.

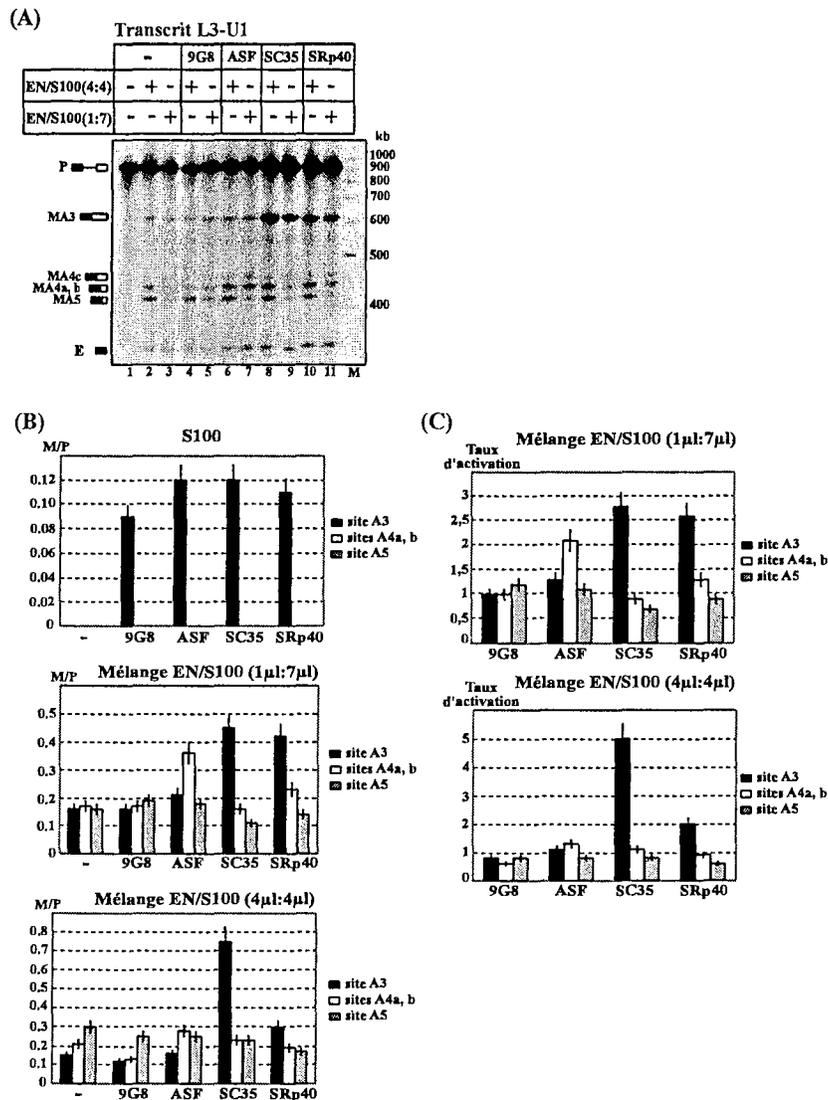


FIG. IV.5 – Effet des protéines SR sur l'épissage de l'ARN L3-U1 dans différents mélanges d'extrait nucléaire et cytoplasmique.

(A) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage du transcrit L3-U1 uniformément marqué. Le transcrit a été incubé 2h30 dans un mélange d'extraits nucléaire et cytoplasmique en proportions 4/4 et 1/7, en absence ou en présence de 300 ng de protéine SR recombinante 9G8, ASF/SF2, SC35 ou SRp40, dans les conditions décrites dans la partie matériel et méthodes. Le contrôle (piste 1) correspond au transcrit non incubé. La position des précurseurs, produits matures et exon 1 est indiquée à gauche de l'autoradiogramme.

(B) Histogrammes dans lesquels sont reportés les valeurs des rapports M/P obtenues dans les mélanges EN/S100 1/7 et 4/4, estimées par mesure de la radioactivité au phosphorimager. (C) Histogrammes dans lesquels sont reportées les valeurs des taux d'activation des sites d'épissage en présence des protéines SR pour les mélanges EN/S100 1/7 et 4/4.

L'utilisation des sites A4a et b est activée par ASF/SF2 à faible concentration en extrait nucléaire, mais la stimulation déjà faible dans le mélange EN/S100 1/7 diminue encore, lorsque la concentration en extrait nucléaire est augmentée.

Nous avons déjà mentionné dans l'article que le site A5 semble régulé par défaut, en présence de SC35 et SRp40, et que 9G8 augmente légèrement l'utilisation de ce site. Cette activation par 9G8 est très faible dans le mélange 1/7 (panneau C, histogramme supérieur) et semble, là-aussi, dépendre d'autres composants nucléaires.

IV.2.6.2 Effet des protéines SR sur la compétition *in vitro* des sites A1 et A2, d'une part, et A2 et A3, d'autre part

Afin d'étudier *in vitro* la compétition entre les sites A1, A2 et A3, nous avons utilisé deux constructions qui avaient été produites par L. Damier (Damier 1997). L'une, C12, produit un ARN contenant les sites donneurs D1 et D2 et les sites accepteurs A1 et A2. L'autre, L23, contient les sites donneurs D1 et D3 et les sites accepteurs A2 et A3 (Fig. IV.6-A).

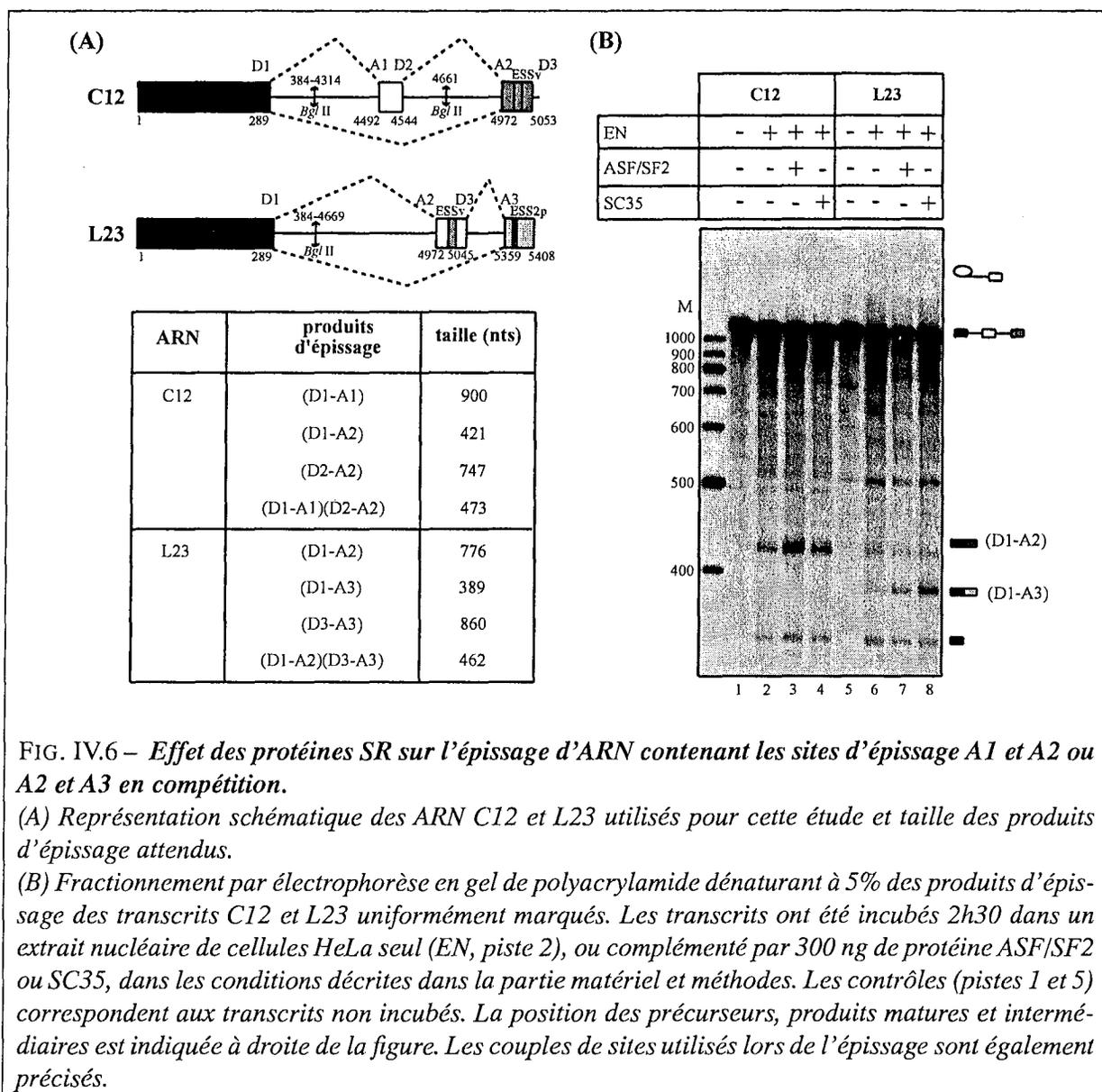
L'ARN C12 est essentiellement épissé entre les sites D1 et A2, aussi bien en extrait nucléaire seul, que lorsque les protéines ASF/SF2 et SC35 sont ajoutées (Fig. IV.6-B). Du fait de la grande taille du produit d'épissage entre les sites D1 et A1 (900 nts), il est difficile de déterminer s'il y a épissage entre ces sites. Par contre, il paraît clair qu'aucun produit d'épissage correspond à un épissage unique entre D2 et A2 ou un double épissage (D1-A1)(D2-A2).

L'utilisation prédominante du site D1 dans l'ARN C12 peut s'expliquer d'une part, parce que ce site est le plus fort et d'autre part, parce qu'il est situé en 5' de l'ARN prémessager, ce qui, comme nous l'avons vu dans la revue bibliographique, favorise sa reconnaissance par la machinerie spliceosomale. En ce qui concerne les sites accepteurs, alors que A1 s'était avéré être plus fort que A2, lorsque les constructions C1 et C2 étaient utilisées, on s'aperçoit ici que A2 est le seul site utilisé. Cela pourrait être dû à l'action d'une séquence qui inhiberait le site A1 et qui serait située en aval de D1.

L'ARN L23 est, quant à lui, essentiellement épissé entre les sites D1 et A3, aussi bien en extrait nucléaire seul, qu'en présence de protéines. D'une manière générale, il semble qu'il y ait une préférence pour l'utilisation du site accepteur distal, dans nos conditions expérimentales.

L'effet des protéines SR sur l'épissage des ARN C12 et L23 est en bon accord avec les données obtenues pour les ARN C2 et C3¹. L'épissage entre les sites D1 et A2 de l'ARN C12 est, comme attendu, activé par ASF/SF2, mais pas par SC35 (Fig. IV.6-B, pistes 3, 4 vs 2). De même, l'épissage de l'ARN L23 entre les sites D1 et A3 est fortement favorisé lorsqu'on ajoute la protéine SC35 (piste 8 vs 6). Pour l'ARN L23, l'observation d'une augmentation de l'épissage entre D1 et A3 par rapport à D1 et A2, lorsque ASF/SF2 est ajoutée, n'est pas en parfait accord avec les données *ex vivo*. En effet, en présence de cette protéine, on ne détecte toujours pas de produit d'épissage résultant de l'utilisation du site A2. Ceci peut être lié à l'absence de la partie 3' de l'exon A3-D4 contenant l'élément ESS2 et les sites A4c, a, b et A5. L'expérience devra être refaite avec une construction contenant cette partie aval de l'exon. C'est pourquoi ces données dans leur état n'ont pas été incluses dans l'article.

¹Il faut en effet comparer les données pour C3 et non L3-U1, car l'exon contenu dans L23 est identique à celui contenu dans l'ARN C3, à savoir que seul l'élément ESS2p est présent (cf Fig. IV.2, page 127).



IV.3 Conclusions et discussion des résultats

IV.3.1 Les protéines SR régulent l'épissage de l'ARN du virus HIV-1, en activant spécifiquement des sites accepteurs d'épissage

Notre étude de l'effet des protéines SR sur l'épissage de l'ARN du virus HIV-1 a montré qu'*ex vivo* ces protéines peuvent modifier fortement le profil d'épissage de l'ARN viral. Les expériences d'épissage *in vitro*, que nous avons menées en parallèle, ont permis d'expliquer les effets observés *ex vivo* : les protéines SR agissent en régulant l'efficacité d'épissage des sites accepteurs A1 à A5 (Fig. IV.3).

Ainsi, après surexpression de la protéine ASF/SF2, l'augmentation de la production de l'ARNm *vpr*, épissé au site A2, est en accord avec la forte activation du site A2, observée *in vitro* après ajout de la protéine ASF/SF2. *In vitro*, la protéine ASF/SF2 a un faible effet activateur sur le site A4b, ce qui est corrélé avec la faible augmentation de la quantité d'ARNm *rev1*.

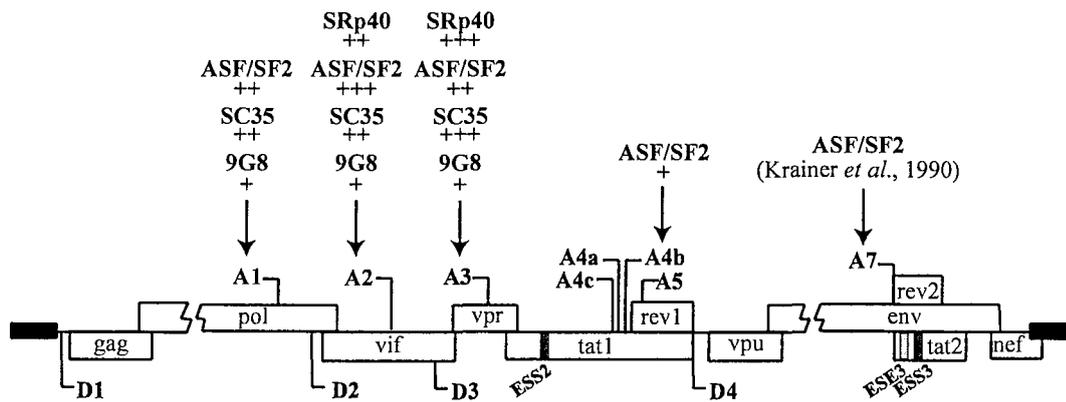


FIG. IV.7 – Effet des protéines SR ASF/SF2, SC35, 9G8 et SRp40 sur l'utilisation des sites accepteurs d'épissage du virus HIV-1.

Les sites d'épissage sont précisés, ainsi que les phases ouvertes de lecture (représentées par des rectangles). Le nombre de signes + est représentatif de la force de l'activation de l'épissage par la protéine SR.

De même, les protéines SC35 et SRp40 activent fortement la production des ARNm *tat1*, ce qui est cohérent avec l'activation du site A3 observée *in vitro* après ajout de ces 2 protéines. La quantité d'ARNm *tat2* est également augmentée par SC35, mais plus faiblement que *tat1*, ce qui suggère que l'élimination de l'intron D1-A3 serait favorisée, par rapport à celle des introns D1-A1 et D2-A3. Par contre, en présence de SRp40, la production d'ARNm *tat3* est augmentée au détriment des ARNm *tat2*, qui disparaissent complètement. Ce résultat est compatible avec l'incapacité de la protéine à activer l'épissage au site A1 *in vitro*.

Enfin, la forte diminution de la quantité d'ARNm *nef* en présence de SC35 ou SRp40, et de *nef2* en présence d'ASF/SF2, se justifie par le fait que ces protéines provoquent, *in vitro*, une diminution de l'utilisation du site A5, au profit du site A3. La compétition entre les sites A3 et A5 que nous observons dans nos études *in vitro* existe donc également *ex vivo*.

La différence entre les données d'épissage *in vitro* et celles obtenues *ex vivo* concerne la protéine 9G8, qui active l'épissage des ARNm *tat1* et *tat2* *ex vivo*, alors qu'*in vitro*, nous avons observé un effet activateur de 9G8 sur le site A3, uniquement dans le contexte de l'ARN C3.

L'activation du site A3 par ASF/SF2 que nous avons observée *in vitro* avec l'ARN court C3 contenant seulement le site A3, n'est que très faiblement retrouvée avec l'ARN long L3-U1 et est inexistante *ex vivo*. Nous discuterons plus loin de ce point, après l'étude détaillée des mécanismes de régulation du site A3.

IV.3.2 Un nouvel exemple de régulation de l'épissage des ARNm viraux par les protéines SR

Notre étude décrit un nouvel exemple de régulation de l'épissage par les protéines SR chez les virus. C'est la 1^{re} étude directe de l'effet des protéines SR, réalisée sur un ARN présentant autant de sites alternatifs d'épissage. Nos données confirment que les protéines sont, pour les virus, comme pour les ARN pré-messagers cellulaires, des acteurs majeurs de l'expression des gènes. Dans le cas du virus

du sarcome de Rous que nous avons décrit dans la revue bibliographique, la protéine ASF/SF2 inhibe l'épissage des ARNm *env* et *src*, en recrutant la snRNP U1 sur un pseudo-site 5' dans l'élément NRS. Le cas de l'adénovirus humain et du papillomavirus bovin est plus proche de celui du virus HIV-1, car les protéines SR participent directement à la sélection des sites d'épissage de ces virus et favorisent l'expression des gènes de la phase précoce, dans le cas de l'adénovirus, ou ceux de la phase tardive, pour le papillomavirus.

Ces différents virus régulent l'activité des protéines SR pour permettre la production des différents ARNm viraux. Ainsi, dans le cas de l'adénovirus, le passage de la phase précoce d'infection à la phase tardive se fait par titration des protéines SR par les transcrits tardifs (pour revue, Akusjarvi & Stévenin 2003). Cette transition dans l'expression des gènes viraux peut être également assurée par déphosphorylation des protéines SR, induite par le virus. C'est le cas pour l'adénovirus, mais aussi pour le virus de la vaccine ou le papillomavirus. Les protéines SR perdent ainsi leur propriétés d'activation de l'épissage (Kanopka *et al.* 1998 ; Huang *et al.* 2002 ; Liu *et al.* 2003).

En ce qui concerne le virus HIV-1, la transition de la phase précoce vers la phase tardive d'infection est assurée par la protéine Rev. Les protéines SR semblent participer à ce processus en modulant l'activité de la protéine Rev (Pongoski *et al.* 2002). Néanmoins, il n'est pas connu pour le moment si une régulation du niveau de phosphorylation des protéines SR participe, comme pour d'autres virus, au passage à la phase tardive du cycle viral. Ce point mérite aussi d'être étudié.

Nos résultats d'épissage suggèrent, qu'en réponse à une variation des quantités intracellulaires de protéines SR, le profil d'épissage des ARNm viraux peut être fortement affecté.

Plusieurs études ont montré que l'expression des ARNm du virus HIV-1 varie, suivant la nature de la cellule infectée. Dans certains cas, cette variation est due à une régulation transcriptionnelle, notamment dans les lymphocytes T entrant en latence après leur infection (Hermankova *et al.* 2003). Dans les macrophages, la régulation pourrait avoir lieu au niveau de l'étape d'épissage. En effet, comme nous l'avons mentionné dans la revue bibliographique, ces cellules servent de réservoir pour le virus HIV-1. Elles subissent une infection lente, chronique et non lytique ; ces caractéristiques sont corrélées avec une diminution de la production des ARN messagers *tat* (Sonza *et al.* 2002) ; les facteurs impliqués dans ce processus n'ont pas encore été identifiés. D'après nos résultats, on pourrait imaginer l'existence de variation de l'activité de certaines protéines SR dans ces cellules, ou la présence de plus fortes concentrations de protéines inhibitrices comme hnRNP A1 ou hnRNP H. Il serait intéressant d'étudier s'il existe une corrélation entre le profil d'épissage des ARNm de HIV-1 et le niveau d'expression des protéines SR dans les différents types de cellules infectées par le virus HIV-1 (lymphocytes T, cellules dendritiques, macrophages ...).

IV.3.3 Impact possible du contrôle de l'épissage alternatif de l'ARN de HIV-1 par les protéines SR, sur la production des protéines virales

Le contrôle de l'épissage exercé par les protéines SR suggère que des variations de la concentration intracellulaire de ces protéines peuvent influencer l'expression des protéines virales. C'est d'ailleurs ce que nous avons observé avec la protéine Nef. Or, des données de Maldarelli *et al.* (1998) montrent que des cellules de la lignée lymphocytaire H9 produisent deux fois plus de protéine SC35, deux jours

après leur infection. Ceci conduirait à une augmentation de la quantité de protéine Tat dans la cellule infectée (Fig. IV.8). Comme nous avons pu le voir précédemment dans l'introduction, la protéine Tat joue un rôle majeur pour le démarrage du cycle viral. En effet, seuls des transcrits courts sont produits en absence de cette protéine. La fixation de la protéine Tat à son élément de réponse TAR conduit, après recrutement de la cycline T1, suivi de celui de la kinase CDK9, à l'hyperphosphorylation du domaine CTD de l'ARN polymérase II. Cette dernière synthétise alors des transcrits de grande taille. Ainsi, en réponse à une plus forte concentration intracellulaire en protéine SC35, la transcription des gènes viraux serait activée et permettrait le démarrage du cycle viral. On peut s'attendre au même résultat, en présence d'une quantité importante de protéine SRp40.

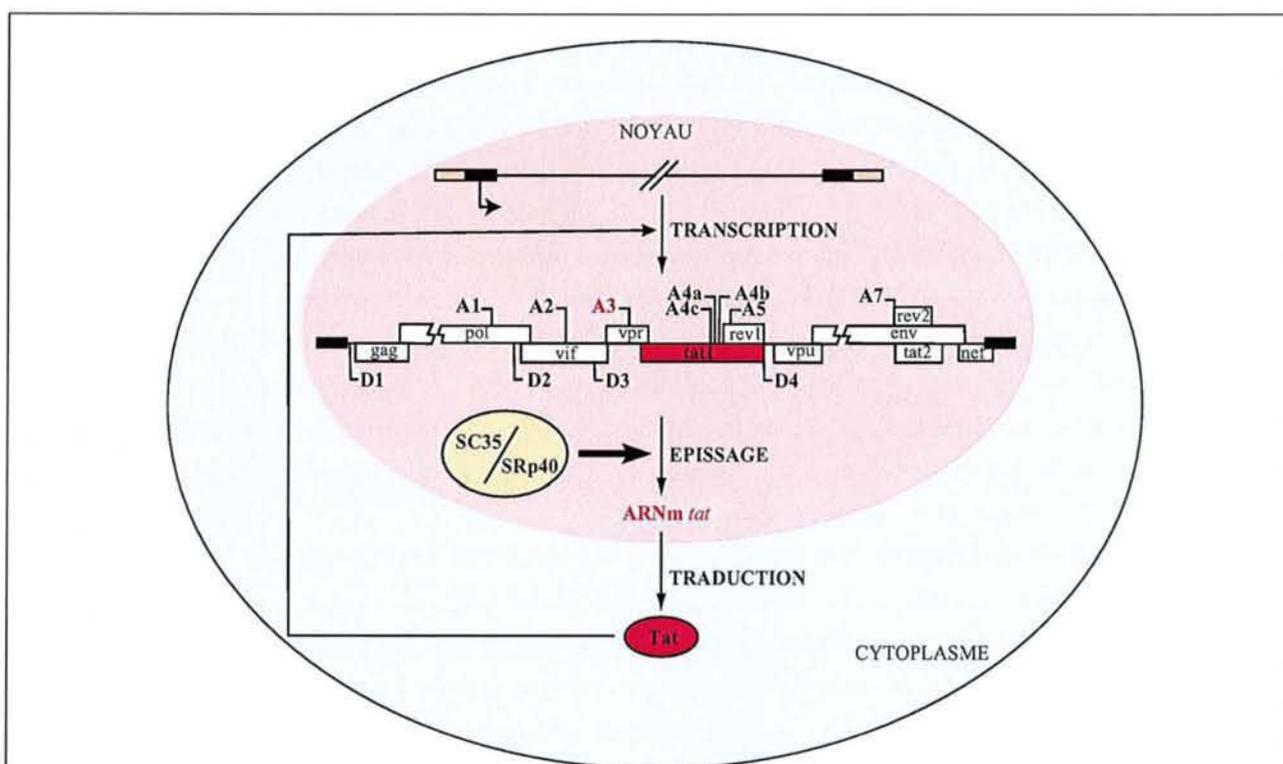


FIG. IV.8 – La surproduction des protéines SC35 ou SRp40 favoriserait le démarrage du cycle viral.

L'augmentation de la concentration intracellulaire en protéine SC35 ou SRp40 activerait l'utilisation du site A3, d'après nos données d'épissage, et conduirait à une production plus importante d'ARNm tat. Une quantité importante de protéine Tat serait produite, ce qui activerait la transcription des gènes viraux, étape nécessaire au démarrage du cycle viral.

Une augmentation de la concentration en protéine ASF/SF2, dans la cellule infectée, conduirait à une production accrue de protéines Vpr. Nous avons vu dans la revue bibliographique que la protéine Vpr bloque les cellules en division en phase G2 et qu'elle favorise la multiplication virale dans les cellules ne se divisant pas, en intervenant dans l'import nucléaire du complexe de préintégration. La protéine Vpr stimule aussi la transcription et régule l'activation et l'apoptose des cellules infectées (pour revue, Bukrinsky & Adzhubei 1999).

La protéine Nef, dont l'expression est au contraire diminuée en présence de concentration élevée

de protéine ASF/SF2, SC35 ou SRp40, intervient également dans des processus multiples comme l'apoptose. Elle a la propriété de protéger la cellule infectée de l'apoptose et en même temps, d'induire l'apoptose de lymphocytes T cytotoxiques voisins (pour revue, Arora *et al.* 2002). Une surexpression de protéine ASF/SF2 dans les cellules infectées est donc susceptible de modifier la relation entre le virus et la cellule infectée, mais aussi entre le virus et les cellules voisines.

IV.3.4 Le contrôle de l'expression des protéines SR permettrait au virus HIV-1 de réguler l'expression de ses gènes

Face au rôle primordial que jouent les protéines virales lors de l'infection par le virus HIV-1, il est fort possible que le virus contrôle l'expression des protéines SR.

En plus de l'étude de Maldarelli *et al.* (1998) que nous avons citée précédemment, quelques autres études ont récemment mis en évidence des modifications de l'expression des facteurs d'épissage dans les cellules infectées par le virus HIV-1. Elles ont ainsi permis de détecter, lors de l'infection, une réduction de l'expression des ARN messagers des protéines 9G8 (Ryo *et al.* 2000), SRp20, ASF/SF2, hTra2 β , hnRNP H, A1, A2/B1 et une légère augmentation de l'expression des ARNm de la protéine SRp40 (van 't Wout *et al.* 2003). D'autres études ont montré que la surexpression de la protéine Tat réprime l'expression de protéines de la voie des MAP kinases (de la Fuente *et al.* 2002 ; Gibellini *et al.* 2002). Comme nous l'avons vu dans l'introduction, cette voie de transduction du signal régule l'activité de la protéine hnRNP A1 en jouant sur son niveau de phosphorylation. Les variations de l'expression des protéines SR par surexpression de la protéine Tat n'ont pas encore été étudiées. C'est un projet que compte développer notre équipe.

Les données de la littérature sont à considérer avec précautions pour le moment, car les conditions expérimentales de chacune de ces études étaient différentes (souches virales et durée d'infection différentes, par exemple) et les résultats obtenus par différentes équipes ne sont pas en parfait accord. Néanmoins, il apparaît que suite à l'infection par le virus HIV-1, l'expression des protéines SR et hnRNP, et vraisemblablement leur activité, varie, ce qui peut avoir des implications importantes sur la production des protéines virales. Les sites A2 et A3 sont réprimés par la protéine hnRNP A1 (Del Gatto-Konczak *et al.* 1999 ; Caputi *et al.* 1999 ; Bilodeau *et al.* 2001). La protéine hnRNP H inhibe également le site A3 (Jacquet *et al.* 2001a). Ainsi, la diminution observée de la quantité des protéines hnRNP A1 et H, ajoutée à une baisse éventuelle du niveau de phosphorylation de la protéine hnRNP A1 et à l'augmentation observée des quantités de protéines SC35 et SRp40 dans la cellule infectée doivent, d'après nos données, conduire à une forte production de protéine Tat.

Quelles pourraient être les conséquences de la baisse observée de la quantité d'ARNm de la protéine ASF/SF2 ? En réponse à une plus faible quantité de protéine ASF/SF2 dans la cellule infectée, l'épissage au site A2 serait, d'une part, moins fortement régulé par ASF/SF2 qui l'active et par la protéine hnRNP A1 qui le réprime. D'autre part, nos expériences *ex vivo* suggèrent que l'activation du site A2 par ASF/SF2 se fait au détriment de celle du site A3. Une baisse de la concentration en protéine ASF/SF2 concourrait donc aussi à une augmentation de l'utilisation du site A3 et donc de la production des ARNm *tat*.

Ces différentes hypothèses ont besoin d'être vérifiées expérimentalement, mais elles soulignent le rôle majeur que les protéines SR pourraient jouer pour la multiplication virale.

IV.3.5 Les protéines SR réguleraient l'épissage de l'ARN de HIV-1 en reconnaissant des éléments en *cis*

Les protéines SR régulent l'épissage des ARN pré-messagers cellulaires en se fixant à des éléments de séquences généralement exoniques. Or, dans le cas du virus HIV-1, le seul site connu de fixation des protéines SR (l'élément ESE3) était localisé en aval du site A7. D'après les données récentes du laboratoire (Marchand *et al.* 2002), la fixation d'ASF/SF2 à la séquence ESE3 empêche la multimérisation de la protéine hnRNP A1 sur cette région de l'ARN viral, permettant ainsi la reconnaissance du site A7 par la machinerie spliceosomale. Un autre site de fixation d'ASF/SF2 a été proposé en aval du site A5, mais aucune démonstration expérimentale n'avait été apportée sur la fixation d'ASF/SF2 et son effet sur l'épissage (Kammler *et al.* 2001). Nous avons vu également dans la revue bibliographique que dans un isolat du virus HIV-1 HXB2, une mutation ponctuelle de l'ARN viral crée un site de fixation de la protéine SC35, qui active l'utilisation des sites d'épissage cryptiques SD5 et SA6. Ceci provoque la production d'ARNm aberrants et conduit à une réplication virale inefficace (Caputi & Zahler 2002). Les régulations que nous avons observées aux sites A2 et A3 doivent vraisemblablement s'expliquer par la fixation des protéines SR à des éléments ESE, qu'il reste encore à identifier. Le fait que les protéines ASF/SF2 et 9G8 aient le même effet *ex vivo* lorsqu'elles sont délétées de leur domaine RS suggère que le rôle activateur des protéines SR aux sites A2 et A3 pourrait être de contrecarrer l'effet négatif des protéines hnRNP.

Les protéines SR que nous avons étudiées ont des spécificités de reconnaissance des séquences ARN variées. Ainsi, comme nous avons pu le voir dans la revue bibliographique, la protéine 9G8 reconnaît un nombre restreint de séquences différentes, comportant des répétitions GAC (Cavaloc *et al.* 1999). Nous n'avons pas détecté ce type de séquences en aval des sites accepteurs étudiés, ce qui pourrait expliquer le faible effet activateur de la protéine sur l'ARN viral. Nos données montrent que la protéine SRp40 a un effet très ciblé, puisqu'elle active uniquement le site A3. Cette protéine a une double spécificité de reconnaissance des séquences ARN, puisqu'elle peut se fixer à la fois sur une longue séquence de 18 nts (UGGGAGCRGUYRGCUCGY) ou sur des séquences courtes et dégénérées. Nous n'avons pas détecté de séquence présentant une forte homologie avec la séquence 18 nts, ce qui pourrait expliquer la dépendance de la protéine SRp40 vis-à-vis d'autres composants de l'extrait nucléaire, lorsqu'elle active l'épissage. En accord avec le fait que les protéines ASF/SF2 et SC35 reconnaissent un grand nombre de séquences ARN, ces protéines sont capables d'activer l'épissage aux sites A1, A2, A3 et même A4b dans le cas d'ASF/SF2. Dans la suite de notre étude, nous avons recherché les sites de fixation des protéines ASF/SF2 et SC35 au voisinage des sites A2 et A3 pouvant expliquer l'effet activateur de ces protéines. Nous avons commencé par focaliser notre étude sur la région du site A2, puis nous avons étudié la région des sites A3 à A5.

Chapitre V

Recherche de nouveaux éléments régulant en *cis* l'épissage de l'ARN viral

V.1 Stratégie de l'étude

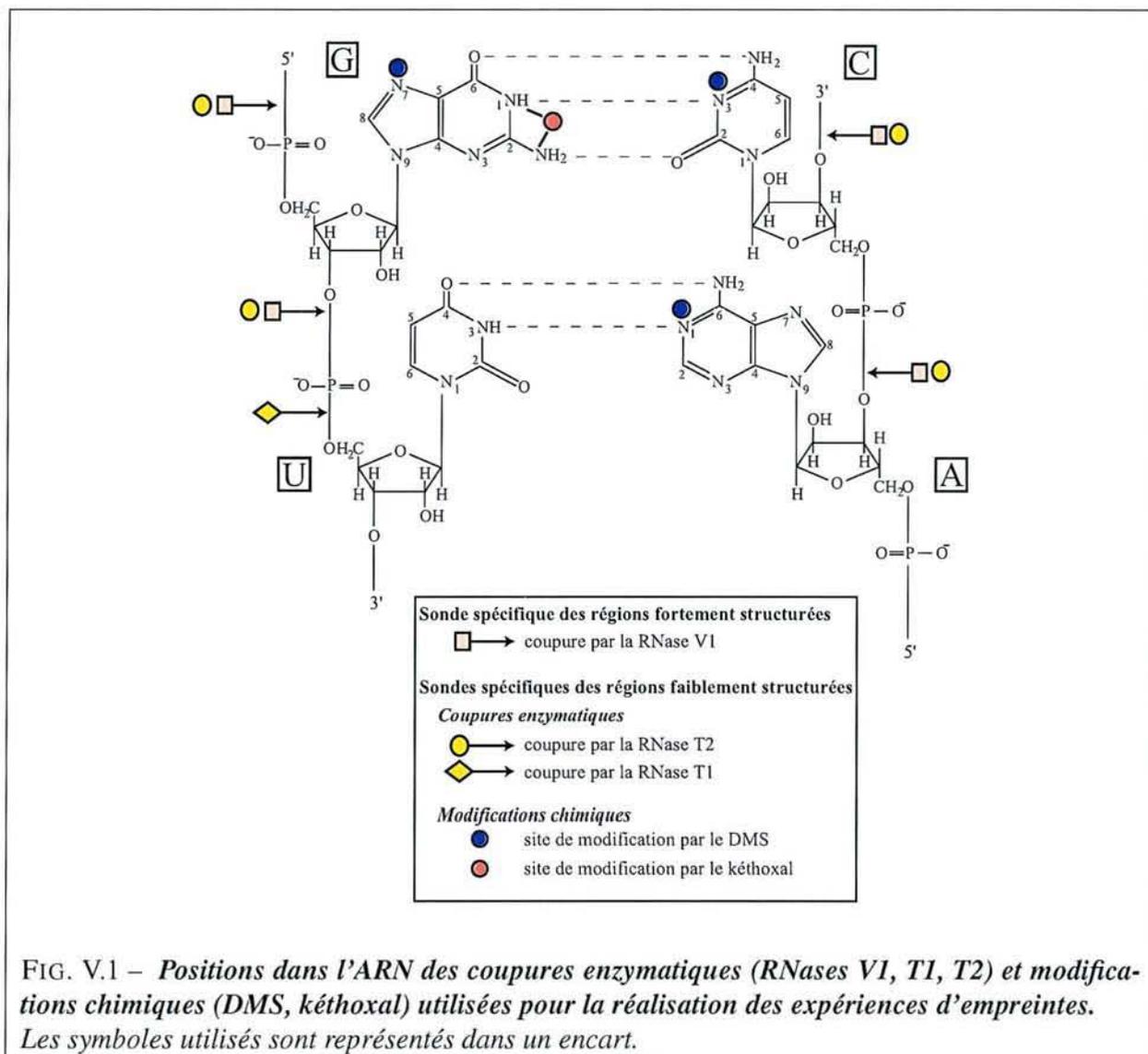
Nous venions de montrer l'existence d'un effet différentiel des protéines SR sur l'activation des sites A1 à A5. Pour comprendre comment ces protéines agissaient, il fallait identifier leur site de fixation sur l'ARN. Sur la base des séquences SELEX reconnues par les protéines ASF/SF2, SC35 et SRp40, on pouvait imaginer réaliser une recherche par informatique des sites potentiels de fixation de ces protéines sur l'ARN du virus HIV-1. Néanmoins, une telle approche était rendue difficile par le grand nombre de séquences que ces protéines pouvaient reconnaître. La petite taille des séquences reconnues et leur degré de dégénérescence sont tels qu'on pouvait s'attendre à ce qu'elles soient statistiquement présentes en grand nombre dans le génome viral. Il aurait ainsi été difficile de distinguer, parmi les nombreuses séquences potentielles de fixation de ces protéines, celles réellement reconnues et jouant un rôle fonctionnel.

Nous avons donc entrepris une étude expérimentale en vue d'identifier les sites de fixation des protéines SR, au voisinage des sites accepteurs, que nous avons montré être régulés par ces protéines et nous avons recherché l'importance fonctionnelle de ces sites. Comme nos travaux *ex vivo* et *in vitro* démontraient une activation du site A2 par ASF/SF2, nous avons recherché la présence de sites de fixation de cette protéine au voisinage de ce site. Par ailleurs, en ce qui concernait le site A3, la situation était plus complexe puisqu'*in vitro*, ce site était activé par ASF/SF2, SC35 et SRp40, alors qu'*ex vivo*, nous n'observions qu'une activation par SC35 et SRp40. Comme nous allons le voir, les études sur ASF/SF2 et SC35 sont bien avancées, puisque nous avons mis en évidence un nouvel élément activateur agissant sur ce site ; la recherche expérimentale n'a pas encore été débutée pour SRp40.

V.2 Méthodologie expérimentale

Différentes approches expérimentales permettent d'identifier des sites de fixation de protéines sur un ARN : des expériences de pontages covalents aux UV, de retards sur gel ou d'empreintes chimiques et enzymatiques des protéines sur l'ARN. J'ai utilisé les expériences de retards sur gel pour définir les conditions de formation de complexes entre des fragments d'ARN de HIV-1 et des protéines SR recombinantes, mais les données obtenues reposent principalement sur l'utilisation des techniques d'empreintes réalisées sur les complexes formées. J'ai employé des techniques d'empreintes à la fois pour étudier des complexes formés *in vitro* entre les séquences d'ARN viral et des protéines SR, mais

aussi pour étudier des complexes formés après incubation de l'ARN avec un extrait nucléaire. Les sondes enzymatiques et chimiques étaient couramment utilisées au laboratoire pour la détermination de la structure secondaire des ARN en solution ou dans des complexes ARN/protéines purifiés (Fig. V.1). Néanmoins, aucune expérience d'empreintes dans un extrait nucléaire n'avait été réalisée et vu la forte concentration en protéines dans l'extrait nucléaire, j'ai dû mettre au point, avec les conseils d'Annie Mougin, des conditions d'analyse des complexes formés en extrait nucléaire. Nous avons utilisé 3 sondes enzymatiques (RNases V1, T1 et T2) et 2 sondes chimiques (DMS et kéthoxal). Le CMCT, couramment utilisé dans ce type d'étude, n'est pas actif dans les conditions utilisées pour les études dans un extrait nucléaire.



En conditions limitantes, les RNases utilisées permettent de distinguer si un résidu est en double-brin (RNase V1) ou en simple-brin (RNases T1 et T2) (Fig. V.1). En effet, la RNase V1 coupe en 5' des liaisons phosphodiester correspondant à des résidus appartenant à des régions structurées de l'ARN. La RNase T1 coupe les liaisons phosphodiester en 3' de guanosines en simple-brin, tandis que la RNase T2 coupe les liaisons phosphodiester en 5' des nucléotides en simple-brin, avec une préfé-

rence pour les adénosines.

En ce qui concerne les 2 sondes chimiques utilisées, le DMS méthyle les résidus A et C en position N1 et N3, respectivement, et les résidus G, en position N7. Les positions N1-A et N3-C étant impliquées dans la formation des appariements de bases canoniques de type Watson-Crick (Fig. V.1), seuls les résidus en simple-brin pouvaient être modifiés à ces positions.

Lors d'une extension d'amorce par la transcriptase inverse, la présence de méthylations en position N1-A ou N3-C génère un arrêt de la transcriptase, d'où la possibilité d'identifier les positions modifiées par comparaison de la position de ces arrêts avec un séquençage réalisé avec la même amorce oligodésoxyribonucléotidique.

L'action du kéthoxal résulte en la formation d'un adduit cyclique entre les positions N1 et N2 des résidus G. Il s'agit, là encore, d'une position impliquée dans la formation des paires de bases Watson-Crick. Donc, seuls les résidus G en simple-brin peuvent être modifiés, et là aussi, la modification bloque l'extension d'une amorce par la transcriptase inverse. De la même manière, les coupures générées par les RNases peuvent être détectées par l'arrêt de l'extension d'amorce qu'elles provoquent.

Lorsque l'on analyse un complexe ARN/protéine par cette approche, les régions de l'ARN où la protéine est directement fixée, ou crée un encombrement stérique, sont protégées. Les segments de l'ARN en interaction ou au proche voisinage de la protéine, peuvent donc être identifiés par comparaison des profils de modifications et de coupures obtenus d'une part, avec l'ARN libre, et d'autre part, avec l'ARN complexé à la protéine. Dans le cas de complexes formés dans un extrait nucléaire, on s'attend à des complexes renfermant plusieurs protéines et donc, à avoir de nombreuses zones de protection. D'une manière générale, pour réaliser ce type d'étude, il faut avoir une bonne connaissance de la structure secondaire de l'ARN étudié. Ces données sont toujours délicates à étudier du fait qu'en plus des arrêts liés à la présence de modifications chimiques ou de coupures enzymatiques, on observe toujours des arrêts dus à des pauses de la transcriptase inverse et l'intensité de ces arrêts peut varier en fonction du tampon dans lequel l'ARN a été incubé. Il est donc nécessaire de réaliser des contrôles d'élongation, dans lesquels aucun agent chimique ou aucune enzyme n'a été ajouté.

Une fois les sites de protection identifiés, l'importance de ces sites pour la fixation de la ou des protéines pour la régulation de l'épissage peut être testée en réalisant une délétion des séquences correspondant à ces sites. Nous avons appliqué cette démarche expérimentale à l'étude des sites A2 et A3.

V.3 Etude du site A2

V.3.1 Etat des données lorsque nous avons entrepris ce travail

Au sein du génome viral, le site accepteur A2 délimite 2 types d'exons : un exon long codant vpr (A2-D4) et un petit exon (A2-D3), de type exon cassette non-codant (Fig. V.2-A). L'inclusion de cet exon dans les ARN messagers viraux joue un rôle déstabilisateur pour ces ARN et permettrait donc de réguler l'expression des protéines virales (Krummheuer *et al.* 2001).

Au laboratoire, L. Damier avait montré, qu'en plus d'une séquence polypyrimidine sous-optimale (Fig. V.2-B), le site A2 utilise un point de branchement non canonique (un résidu G au lieu du résidu A habituellement utilisé ; Damier *et al.* 1997). Par ailleurs, l'équipe de M. Stoltzfus a récemment mis en évidence l'existence d'un élément inhibiteur, ESSv, comportant 3 sites de fixation de la protéine hnRNP A1 (Bilodeau *et al.* 2001).

S. Jacquenet avait mené au laboratoire une étude expérimentale de la structure secondaire de la région de l'ARN de HIV-1 BRU contenant le site A2. D'après cette étude, A2 appartient à une région fortement structurée (Fig. V.2-B), ce qui pourrait être un autre paramètre limitant son utilisation (Jacquenot 2001). Les 3 sites de fixation des protéines hnRNP A1 sont situés dans une structure tige-boucle en 3' du site A2. Nous avons remarqué la présence, dans l'intron en amont du site A2, de deux séquences riches en purines AGGAGAAAGAGA (positions 4844 à 4855) et GGAGGA (positions 4888 à 4893), qui pouvaient potentiellement constituer des sites de fixation de protéines SR. En effet, comme nous l'avons vu dans la revue bibliographique, plusieurs de ces protéines, dont ASF/SF2 et SC35, se lient à des séquences riches en purines. Ces séquences correspondaient respectivement à la boucle terminale de la structure tige-boucle 3 et à une séquence de la boucle terminale de la structure tige-boucle 4.

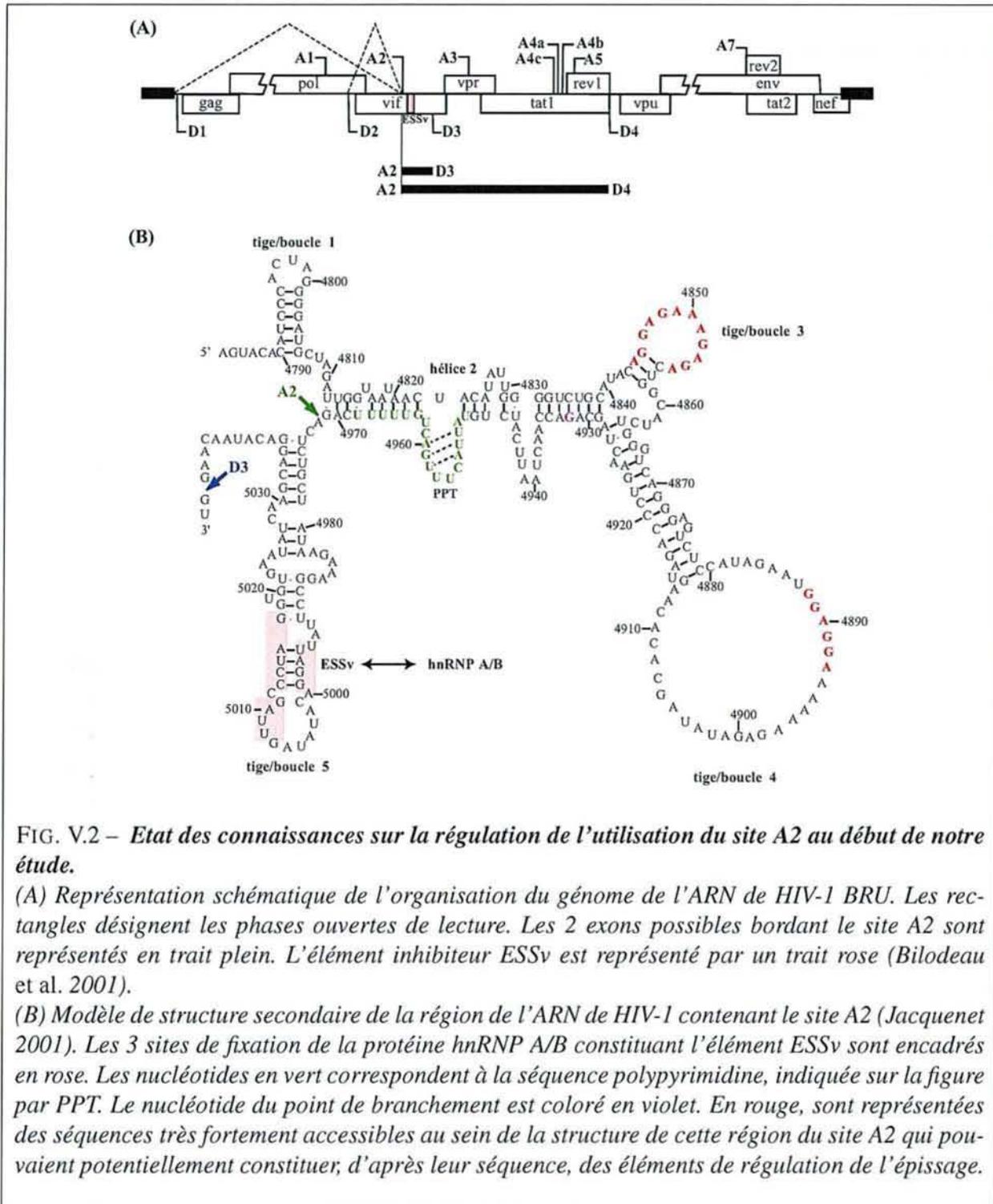
V.3.2 Etude du mécanisme d'activation du site A2 par la protéine ASF/SF2

V.3.2.1 Etude de l'effet de délétion ou mutation dans les structures tige-boucles 3 et 4

Comme nous l'avons indiqué plus haut, les boucles terminales des structures tige-boucles 3 et 4 de l'intron délimité par le site A2 comportent des séquences riches en purines qui pouvaient correspondre à des sites de fixation des protéines SR. Bien que peu de séquences activatrices de l'épissage sont situées dans les introns, nous ne pouvons pas exclure qu'il s'agisse d'éléments de régulation, ceci d'autant plus que la grande boucle de l'élément structural 4 est très conservée dans toutes les souches (Jacquenot 2001).

Comme nous l'avons vu dans l'article inséré dans la partie précédente, l'ARN C2 contenant les sites d'épissage D1 et A2 est épissé efficacement *in vitro* (Damier *et al.* 1997). Nous avons donc généré des mutations dans cette construction par mutagenèse dirigée, de façon à produire l'ARN C2.Δsl3, où la séquence entre les positions 4843 et 4858 de la tige-boucle 3 était délétée, et l'ARN C2.sl4m, dont la séquence entre les positions 4888 et 4893 était mutée (Fig. V.3-A, B).

L'efficacité d'épissage *in vitro* de ces ARN variants a été testée, avec et sans ajout de protéine ASF/SF2 recombinante dans l'extrait nucléaire pur, et comparée à celle de l'ARN C2 sauvage. Au-



cune différence marquée n'a été observée (Fig. V.3-C). Il semblait donc que, comme c'est plus fréquemment le cas, les sites d'activation du site A2 par ASF/SF2 soient situés dans l'exon bordé par ce site. Des travaux sur le site A7, menés en parallèle au laboratoire, montraient que le taux d'utilisation de ce site repose sur la compétition entre les protéines hnRNP A1 et ASF/SF2. Nous nous sommes alors demandé si une telle compétition n'avait pas aussi lieu au site A2. Un autre argument en faveur de cette hypothèse était l'existence d'un site potentiel de fixation d'ASF/SF2, qui chevauchait les sites de fixation de la protéine hnRNP A1 dans la structure tige-boucle 5.

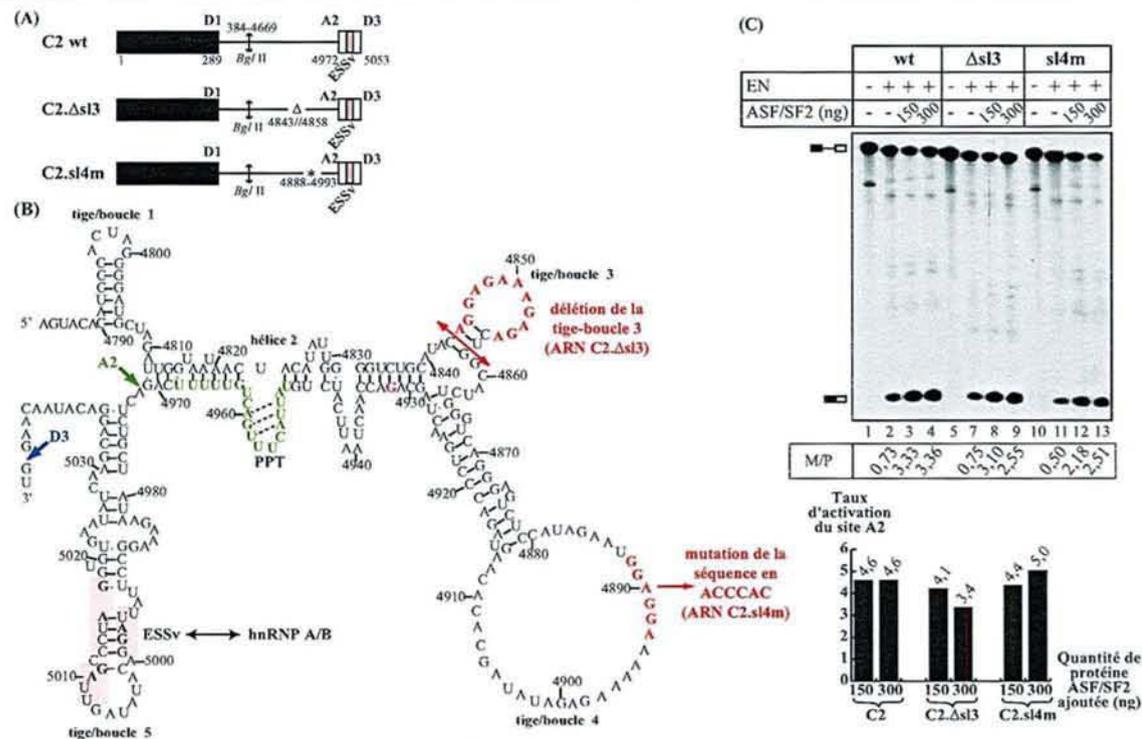


FIG. V.3 – *Etude de l'effet d'une délétion ou d'une mutation dans les tige-boucles 3 ou 4, sur l'activation du site A2 par ASF/SF2.*

(A) Représentation schématique des ARN utilisés pour cette étude. Dans l'ARN C2.Δsl3, la séquence entre les nts 4843 et 4858 a été éliminée (représentée par un symbole Δ sur le schéma). L'ARN C2.sl4m a été obtenu par mutation de la séquence entre les positions 4888 et 4893 (représentée par un astérisque).

(B) Positionnement des sites potentiels de fixation de la protéine ASF/SF2 (représentés en rouge), ainsi que de l'élément ESSv, sur le modèle de structure secondaire de la région de l'ARN de HIV-1 contenant le site A2. Les dinucléotides AG de l'élément ESSv, représentés en gras, ont été mutés en CU dans l'étude du site A2 réalisée par l'équipe de M. Stoltzfus (Bilodeau et al. 2001). La délétion (délimitée par une double-flèche rouge) ou la mutation réalisées sont indiquées sur le schéma.

(C) Effet de l'ajout, à la réaction d'épissage, de 150 et 300 ng de protéine ASF/SF2 sur l'utilisation du site A2. Les ARN C2, C2.Δsl3 et C2.sl4m uniformément marqués ont été incubés 2h30 dans un extrait nucléaire, dans les conditions classiques d'épissage décrites dans la partie matériel et méthodes, et les produits d'épissage ont ensuite été soumis à une électrophorèse en gel dénaturant de polyacrylamide à 5%. La position des précurseurs et des produits matures est indiquée à gauche de l'autoradiogramme. Les valeurs des rapports M/P, estimées par mesure de la radioactivité des bandes au phosphorimager, ainsi qu'une représentation sous forme d'histogramme des taux d'activation du site A2 par les protéines SR sont présentées en-dessous.

V.3.2.2 Etude de la fixation des protéines ASF/SF2 et hnRNP A1 sur la structure tige-boucle 5 en aval du site A2 par expériences d'empreintes

Un moyen de tester la fixation de la protéine SR ASF/SF2 sur la structure tige-boucle 5 était de former un complexe *in vitro* avec la protéine recombinante et de réaliser des empreintes avec des agents chimiques et des RNases sur ce complexe.

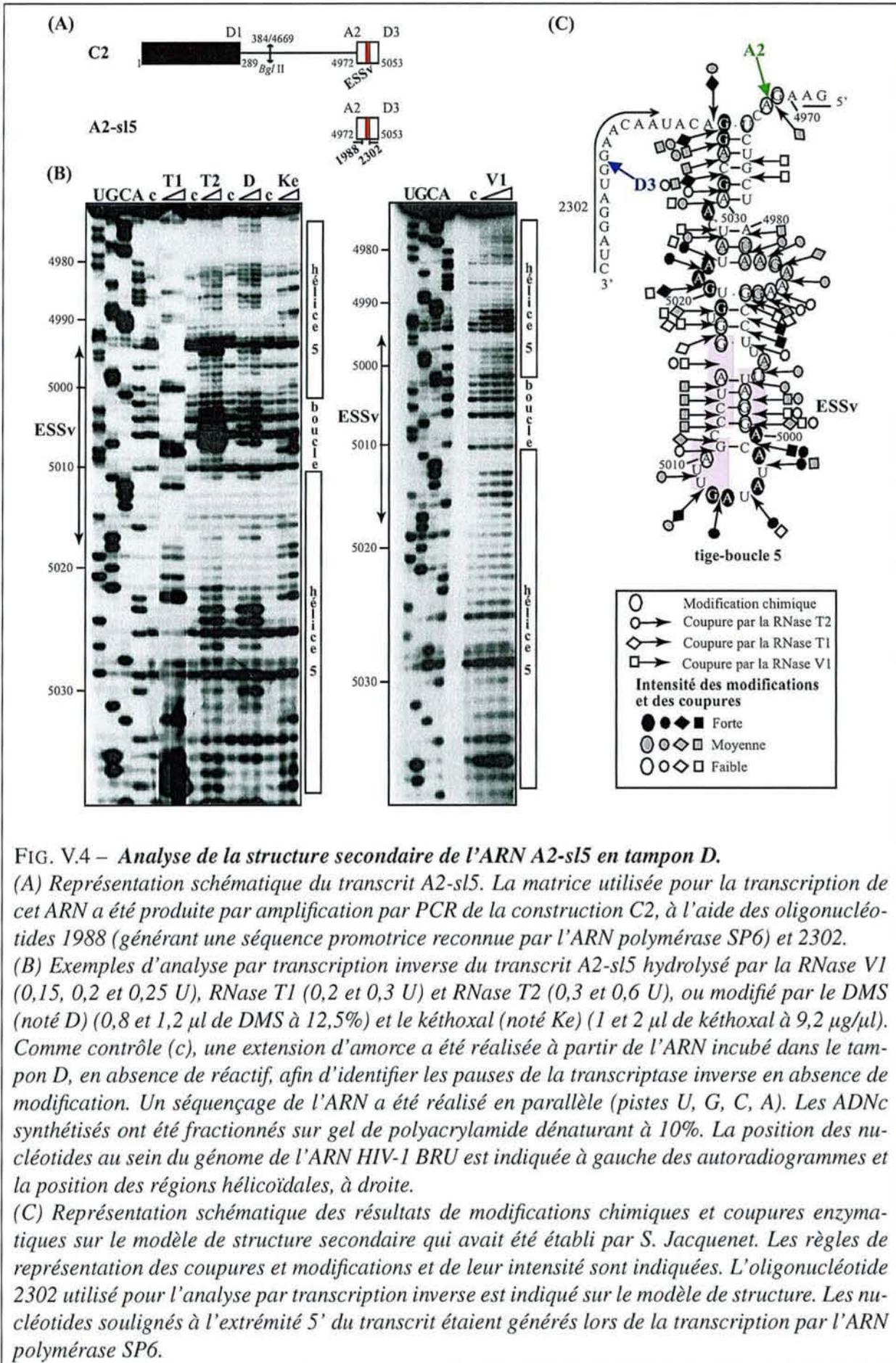
Ce type d'expérience était possible puisque J. Stévenin nous avait fourni de la protéine ASF/SF2 recombinante purifiée. En parallèle, il était intéressant de réaliser le même type d'expérience avec la protéine hnRNP A1. En effet, l'équipe de M. Stoltzfus avait montré que les nucléotides AG des 3 motifs de l'élément ESSv étaient nécessaires pour fixer les protéines hnRNP A/B, mais aucune expérience d'empreintes n'avait été réalisée pour délimiter le site de fixation de ces protéines (Bilodeau *et al.* 2001). Là encore, ce type d'expérience était possible, puisque H. Hallay et A. Méreau avaient produit et purifié une protéine hnRNP A1 recombinante au laboratoire.

Pour simplifier l'interprétation des données, j'ai choisi de réaliser les expériences sur un ARN (A2-s15) correspondant à la structure tige-boucle 5, comportant une séquence simple-brin en 5' qui devait permettre l'appariement de l'oligonucléotide amorce lors des expériences de transcription inverse (Fig. V.4). Nous avons produit par PCR une matrice ADN permettant de produire ce segment de l'ARN de HIV-1, en utilisant, pour l'amplification, le plasmide pLD-C2 et deux oligonucléotides (1988 et 2302), dont l'un générant une séquence promotrice permettant la transcription par l'ARN polymérase du phage SP6 (Fig. V.4-A). Nous avons préféré utiliser cette ARN polymérase car les séquences promotrices reconnues par l'ARN polymérase du phage T7 génèrent, en 5' du transcrit, une séquence reconnue par la protéine hnRNP A1, ce qui perturbe les analyses faites avec cette protéine (Jacquenet *et al.* 2001a).

Etant donné que nous souhaitions tester la compétition possible entre les protéines hnRNP A1 et ASF/SF2 en conditions d'épissage, nous avons réalisé l'étude par empreintes dans le tampon D utilisé pour les réactions d'épissage *in vitro*. Les analyses, faites précédemment par S. Jacquenet, avaient été faites en tampon Tris. Etant donné que la concentration en cations mono- et divalents est différente dans ces 2 tampons, nous avons dû répéter l'analyse de la structure secondaire en employant du tampon D.

Avant l'hydrolyse par les RNases ou les modifications chimiques, l'ARN A2-s15 a été renaturé dans le tampon D, en présence de MgCl₂, à une concentration finale de 2,8 mM. Les conditions d'hydrolyses enzymatiques et de modifications chimiques dérivent de celles précédemment établies au laboratoire par Mougins *et al.* (2002). Les hydrolyses enzymatiques ont été réalisées directement dans le tampon D sans ajout d'autres composés chimiques. Les concentrations d'enzymes et d'ARN, la température, le temps d'hydrolyse, ainsi que les conditions d'arrêt de la réaction sont décrits dans la partie matériel et méthodes. Pour que les modifications chimiques par le DMS et le kéthoxal soient efficaces, du cacodylate de sodium 50 mM pH 7,5 doit être ajouté au tampon D (Mougins *et al.* 2002).

L'analyse que nous avons réalisée sur l'ARN A2-s15 en tampon D est illustrée dans la figure V.4. Elle confirme bien la formation dans le tampon D de la structure tige-boucle 5 proposée par S. Jacquenet. Munis de ces résultats, nous pouvions donc débiter l'étude par empreintes.



L'analyse présentée dans la figure V.4 montre que les sites de fixation de la protéine hnRNP A1, proposés par Bilodeau *et al.* (2001), sont coupés en tampon D par les RNases V1 et T2. Nous avons donc employé ces deux RNases pour la réalisation d'empreintes.

Le complexe entre l'ARN A2-sl5 et la protéine hnRNP A1 recombinante a été formé dans le tampon D avec 1 ; 2,2 et 4,4 μg de protéine et 200 ng d'ARN (soit un excès molaire de protéine par rapport à l'ARN de 2,6 ; 9 et 18). Les résultats obtenus avec 4,4 μg de protéine hnRNP A1 sont présentés dans la figure V.5. Ils révèlent une protection de la partie apicale de la tige-boucle 5, confirmant ainsi les 3 sites de fixation proposés par Bilodeau *et al.* (2001).

Le même type d'expérience a été réalisé avec la protéine ASF/SF2. Les quantités de protéine et d'ARN utilisées pour l'obtention des complexes étaient respectivement de 250 ou 500 ng, et 200 ng (soit un excès molaire de protéine par rapport à l'ARN de 1,2 et 2,4). Etant donné que nous n'avons aucune information sur les sites possibles de fixation, nous avons utilisé une plus large variété d'enzymes (RNases V1, T1 et T2) et d'agents chimiques (kéthoxal et DMS) pour l'analyse de ce complexe. Des données claires ont été obtenues, révélant une protection du segment situé entre les positions 4986 et 5008 (Fig. V.6 et V.7). La protéine ASF/SF2 semble essentiellement fixée sur le brin 5' de l'hélice et sur la boucle terminale. Cette région de l'ARN viral renferme des séquences riches en purines qui pourraient constituer le ou les sites de fixation de la protéine à l'ARN.

V.3.2.3 Effet de la mutation de l'élément ESSv sur l'activation du site A2 par ASF/SF2

Les données d'empreintes montraient bien qu'il y avait un chevauchement des sites de fixation des protéines hnRNP A1 et ASF/SF2. Une autre façon d'étudier la relation entre l'élément inhibiteur ESSv et l'activation du site A2 par ASF/SF2 était de muter l'élément ESSv et de tester par épissage *in vitro* si ASF/SF2 conservait un effet activateur et si oui, à quel niveau.

Dans leur étude du site A2, Bilodeau *et al.* (2001) ont montré que la mutation des dinucléotides AG des 3 motifs de l'élément ESSv abolit la fixation des protéines hnRNP A/B, ce qui lève l'inhibition du site A2. Nous avons introduit ces mutations dans la construction C2 par mutagenèse dirigée. Comme attendu, la mutation de l'élément ESSv augmente l'efficacité d'épissage au site A2 (Fig. V.8-C, piste 5 vs 2). Pour contrôle, nous avons d'abord testé, dans des expériences d'épissage *in vitro*, l'effet de l'ajout de protéine hnRNP A1 recombinante sur l'ARN variant C2-2272, ceci afin de vérifier si on avait bien disparition de l'inhibition du site A2, après élimination de l'élément ESSv. Les résultats présentés dans la figure V.8 montrent que c'est bien le cas, puisque l'inhibition observée pour l'ARN sauvage (facteur d'environ 2) disparaît après mutation de l'élément ESSv (facteur de 1).

En ce qui concerne l'effet d'ASF/SF2 (panneau D), alors que pour l'ARN sauvage le taux d'activation est de 3,9, pour le variant, il diminue à 1,9, soit une perte de presque 50% de l'activation après mutation. Une part importante de l'activation par ASF/SF2 est donc bien liée à la nécessité de déplacer la protéine hnRNP A1 de l'élément ESSv pour qu'il y ait épissage. Néanmoins, même en l'absence d'élément ESSv, ASF/SF2 conserve encore un effet activateur non négligeable. Celui-ci pourrait être

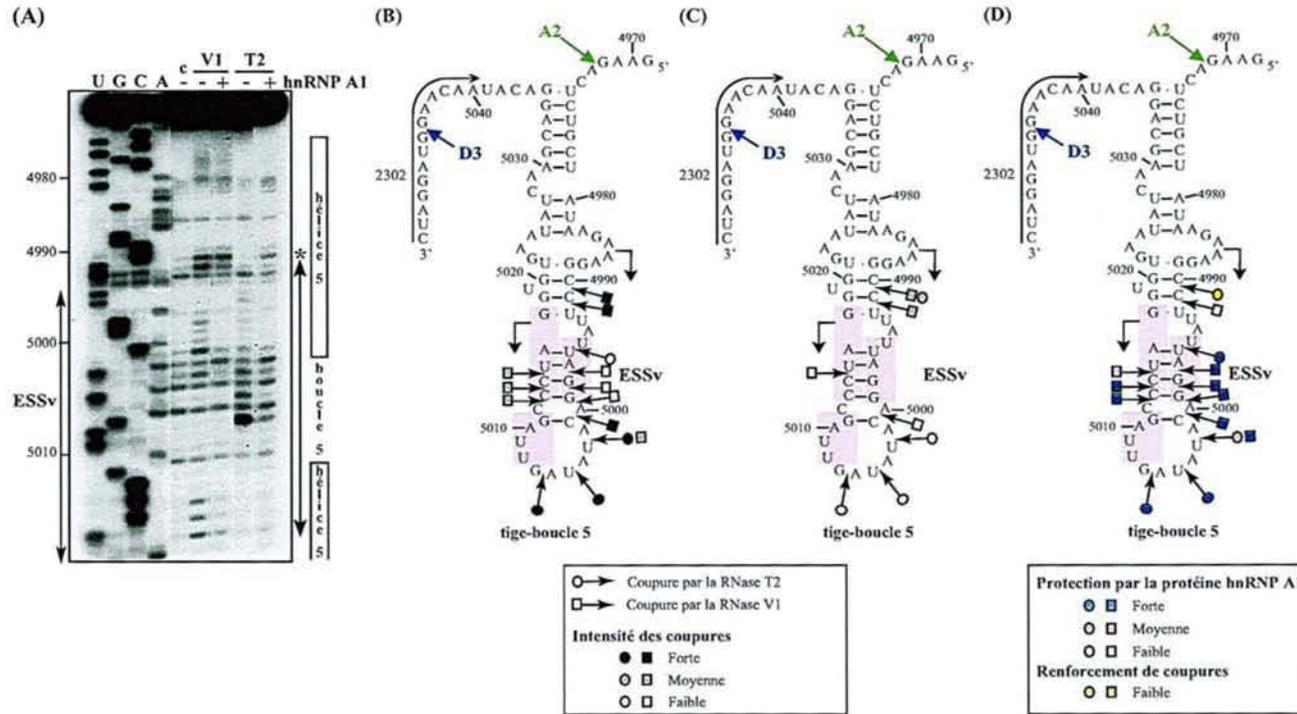


FIG. V.5 – Protection de l'ARN A2-sI5 par la protéine hnRNP A1 dans le complexe ARN A2-sI5/hnRNP A1 reconstitué in vitro.

(A) Exemple d'analyse par transcription inverse du transcrit A2-sI5 hydrolysé par les RNases VI (0,2 U) et T2 (0,3 U), en absence ou en présence de 4,4 µg de protéine hnRNP A1 recombinante. Comme contrôle (c), une extension d'amorce a été réalisée à partir de l'ARN incubé en tampon D, en absence de réactif. Un séquençage de l'ARN a été réalisé en parallèle (pistes U, G, C, A). Les ADNc synthétisés ont été fractionnés par électrophorèse sur un gel de polyacrylamide dénaturant à 10%. La position des nucléotides au sein du génome de l'ARN HIV-1 BRU est indiquée à gauche de l'autoradiogramme. A droite, sont représentées les positions des régions hélicoïdales et de la zone de l'ARN A2-sI5 protégée par la protéine hnRNP A1, qui est délimitée par une flèche. L'astérisque désigne un renforcement de coupure enzymatique en présence de la protéine.

(B) et (C) Représentation schématique des résultats de modifications chimiques et coupures enzymatiques de l'ARN A2-sI5 dans le tampon D (B), ou en présence de la protéine hnRNP A1 (C). La région de l'ARN analysée est délimitée par des flèches noires cassées. La position de la séquence complémentaire à l'oligonucléotide 2302, utilisé pour l'analyse par transcription inverse, est indiquée sur les modèles de structure. Les symboles et code couleur utilisés sont indiqués sous les panneaux dans un encart.

(D) Représentation schématique de la protection par la protéine hnRNP A1 des nucléotides de l'ARN A2-sI5 contre l'action des RNases et des agents chimiques. Le code couleur utilisé pour représenter l'intensité des protections est indiqué dans un encart sous le panneau.

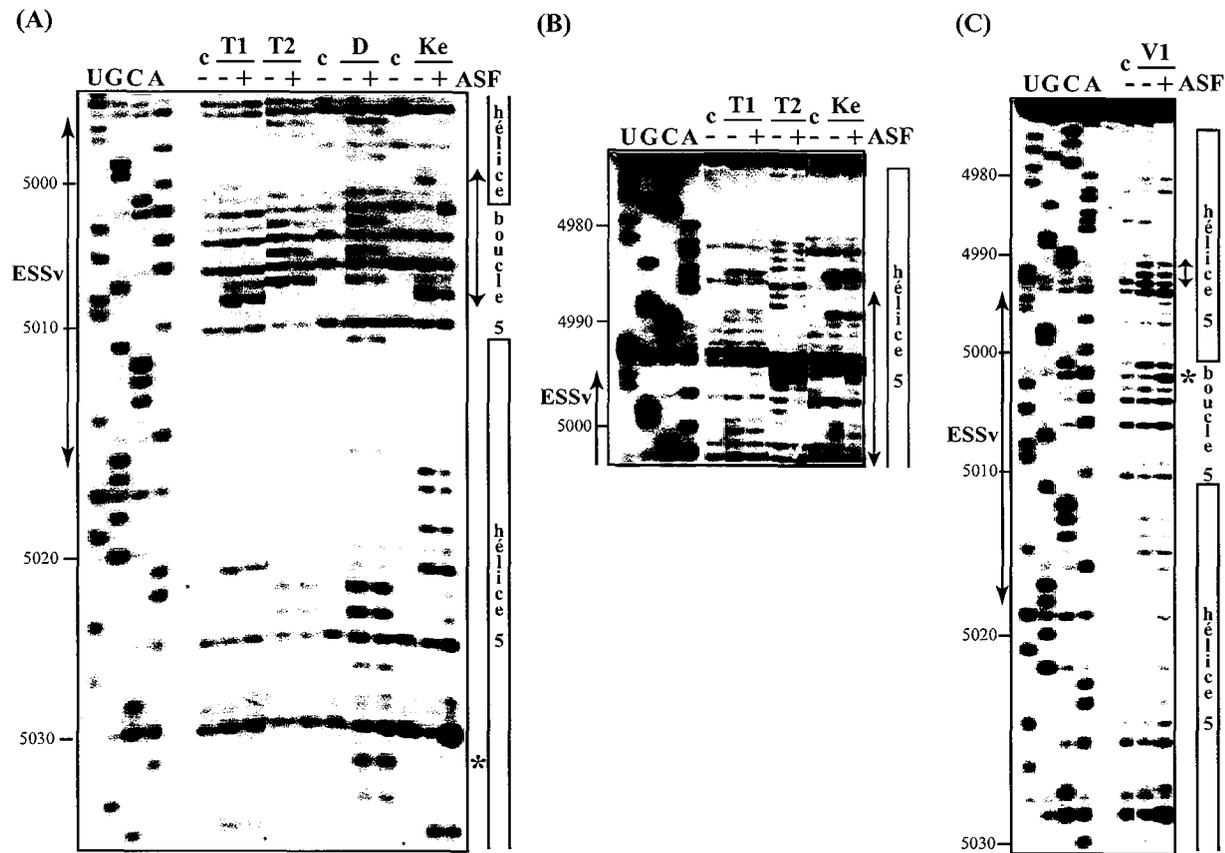


FIG. V.6 – Analyse de l'accessibilité de l'ARN A2-sl5 aux sondes enzymatiques et chimiques en présence de la protéine ASF/SF2.

(A), (B) et (C) Exemples d'analyse par transcription inverse du transcrit A2-sl5 hydrolysé par les RNases VI (0,1 U), T1 (0,025 U) et T2 (0,3 U), ou modifié par 2 μ l de DMS 12,5% ou 1 μ l de kéthoxal à 9,2 μ g/ μ l, en présence de 500 ng de protéine ASF/SF2. L'oligonucléotide 23020 a été utilisé comme amorce. Comme contrôle (c), une extension d'amorce a été réalisée à partir de l'ARN incubé dans le tampon D, en absence de réactif. Un séquençage de l'ARN a été réalisé en parallèle (colonnes U, G, C, A). Les ADNc synthétisés ont été fractionnés par électrophorèse sur gel de polyacrylamide dénaturant à 10%. Le gel présenté dans le panneau (B) correspond à une exposition plus forte du haut du gel présenté dans le panneau (A). La position des nucléotides au sein du génome de l'ARN HIV-1 BRU est indiquée à gauche des autoradiogrammes. A droite, sont représentées les positions des régions hélicoïdales et de la zone de l'ARN A2-sl5 protégée par la protéine ASF/SF2, qui est délimitée par une flèche. L'astérisque désigne un renforcement de l'intensité d'une modification chimique.

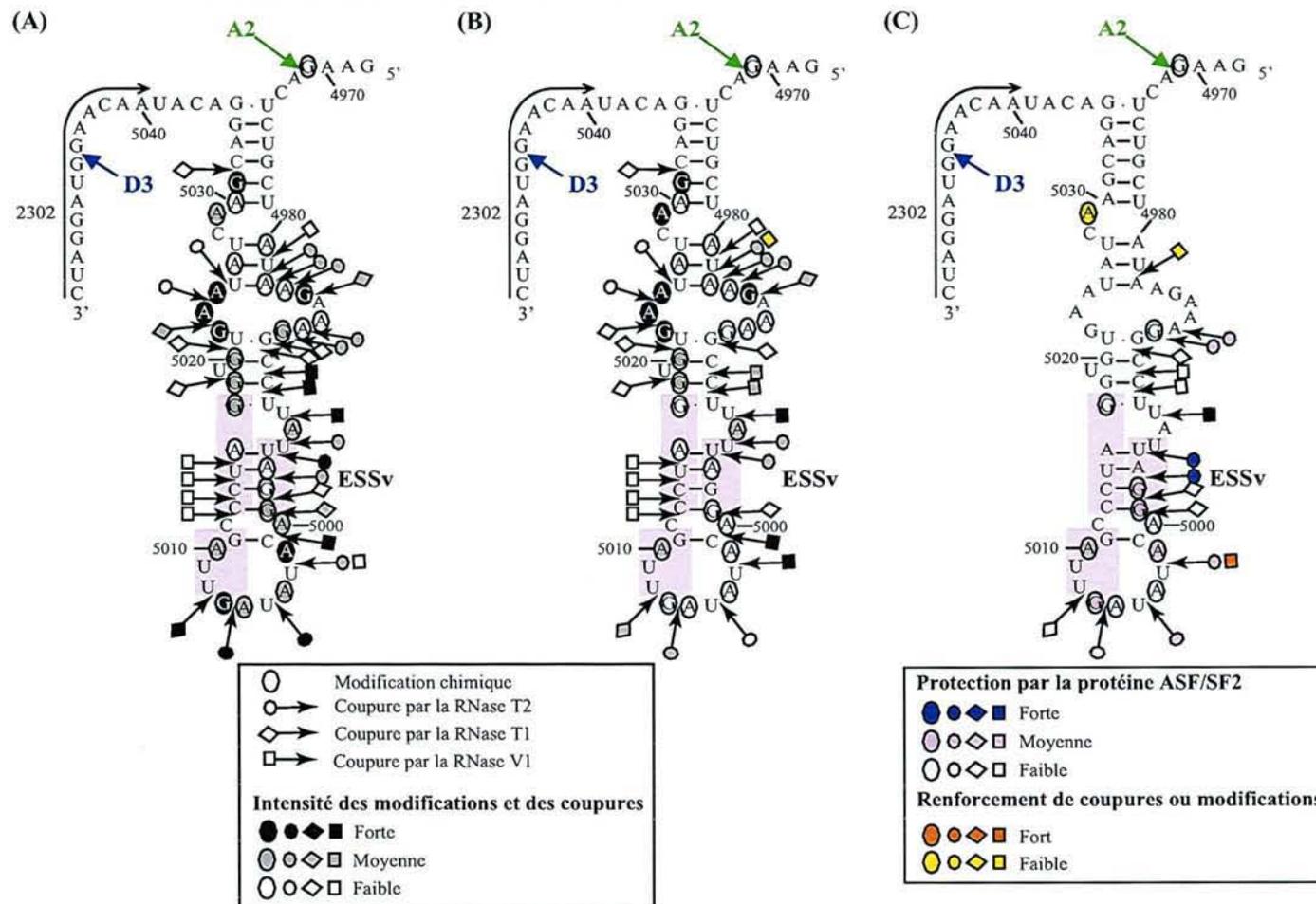


FIG. V.7 – Protection de l'ARN A2-sl5 par la protéine ASF/SF2.

(A) et (B) Représentation schématique des résultats de modifications chimiques et coupures enzymatiques de l'ARN A2-sl5 en absence (A) ou en présence (B) 500 ng d'ASF/SF2. La position de la séquence complémentaire à l'oligonucléotide 2302, utilisé pour l'analyse par transcription inverse, est indiquée sur le modèle de structure. Les symboles et code couleur utilisés sont indiqués sous les panneaux dans un encart.

(D) Représentation schématique de la protection, par la protéine ASF/SF2, des nucléotides de l'ARN A2-sl5, contre l'action des RNases et des agents chimiques. Le code couleur, utilisé pour représenter l'intensité des protections, est indiqué sous le panneau.

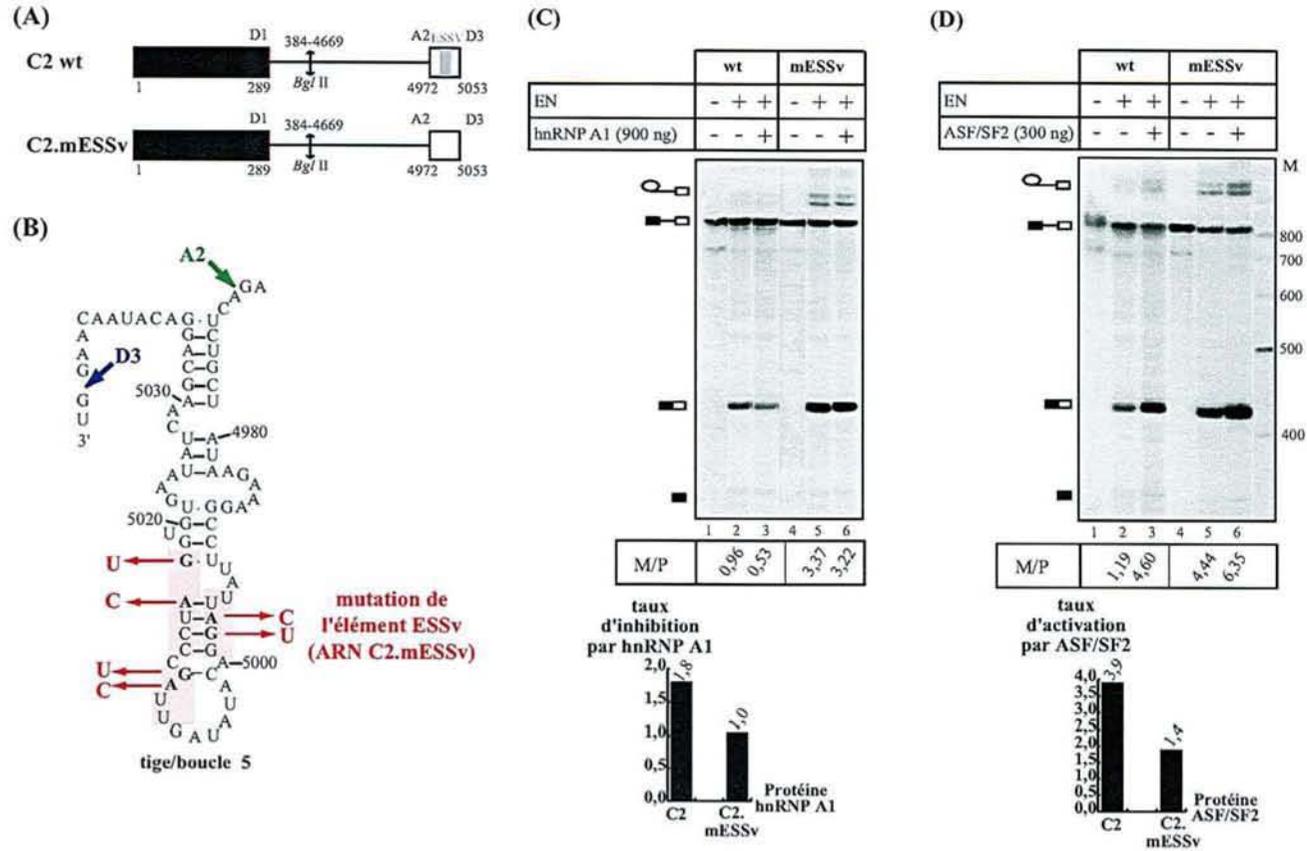


FIG. V.8 – *Effet de la mutation de l'élément ESSv sur la régulation de l'épissage au site A2.*

(A) Représentation schématique des ARN utilisés pour cette étude.

(B) Dans l'ARN C2.mESSv, les dinucléotides de AG des 3 motifs de l'élément ESSv ont été mutés en CU.

(C) Effet de l'ajout de 900 ng de protéine hnRNP A1 sur l'utilisation du site A2 des ARN sauvage et muté. (D) Effet de l'ajout de 300 ng de protéine ASF/SF2 sur l'utilisation du site A2.

Les ARN C2 et C2.mESSv uniformément marqués ont été incubés 2h30 en extrait nucléaire, dans les conditions classiques d'épissage décrites dans la partie matériel et méthodes. Les produits d'épissage ont ensuite été soumis à une électrophorèse en gel de polyacrylamide dénaturant à 5%. La position des précurseurs, intermédiaires et produits matures est indiquée à gauche de l'autoradiogramme. Les valeurs du rapport M/P, estimées par mesure de la radioactivité des bandes au phosphorimager, sont indiquées en-dessous, ainsi qu'une représentation sous forme d'histogramme des taux d'inhibition du site par la protéine hnRNP A1 ou d'activation par la protéine ASF/SF2.

lié à la présence d'un ou plusieurs sites résiduels après élimination de l'élément ESSv. En effet, les protections observées dans la figure V.7 s'étendent en aval de l'élément ESSv.

V.3.3 Conclusions et discussion des résultats

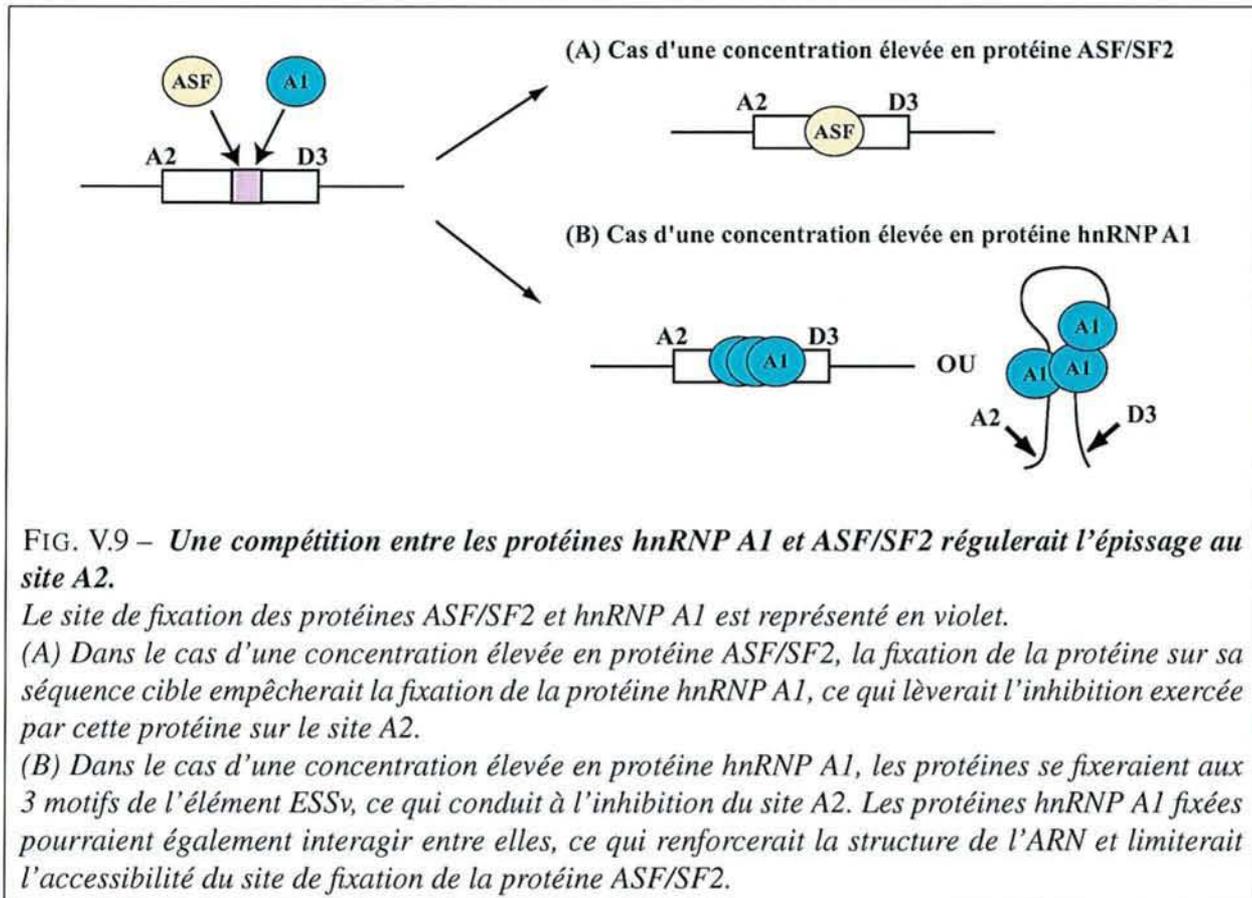
Bien que l'étude des mécanismes de régulation au site A2 ne soit pas encore complète, les premiers résultats que nous avons obtenus permettent de formuler certaines hypothèses.

Les données de nos expériences d'empreinte avec la protéine hnRNP A1 sont en bon accord avec les données de l'équipe de Stoltzfus, mais en plus des trois sites proposés par cette équipe, il semble que la protéine interagisse aussi avec la boucle terminale.

Nos expériences d'empreintes réalisées avec la protéine ASF/SF2 montrent que son ou ses sites de fixation chevauchent ceux de la protéine hnRNP A1. ASF/SF2 protège aussi la boucle terminale. En accord avec la présence de sites de fixation chevauchants pour hnRNP A1 et pour ASF/SF2, la mutation des 3 sites de fixation de hnRNP A1 diminue aussi l'activation par ASF/SF2. L'élément ESSv est donc important à la fois pour l'activation par la protéine ASF/SF2 et pour l'inhibition par hnRNP A1. La boucle terminale de la structure tige-boucle 5 pourrait constituer un site fort de fixation de la protéine ASF/SF2, vu le degré de protection observée en présence de cette protéine. Les sites activateurs d'épissage fixant cette protéine sont en effet souvent situés dans des boucles latérales comme c'est le cas pour le site A7 de HIV-1 (Marchand *et al.* 2002) ou terminale comme c'est le cas pour l'exon EDA de l'ARN pré-messager de la fibronectine humaine (Lavigneur *et al.* 1993 ; Muro *et al.* 1999). La fixation de la protéine ASF/SF2 sur l'élément ESSv et sur la boucle terminale est surprenante. En effet, ces séquences ne ressemblent pas aux séquences décrites comme reconnues par cette protéine (cf Tab. § II.3 de la revue bibliographique). Seule la séquence riche en purines, localisée dans la boucle latérale (nts 4983-4988), pouvait, sur la base de la séquence, être prédite comme un site potentiel de fixation de la protéine ASF/SF2. En fait, seule l'extrémité 3' de cette séquence s'est avéré protégée. Il faut noter qu'une courte séquence riche en purines, AGGA (nts 4997 à 5000) chevauche également le premier motif de l'élément ESSv (Fig. V.7).

D'après les mécanismes connus de régulation de l'épissage des ARN pré-messagers cellulaires, la protéine ASF/SF2 peut activer l'utilisation d'un site 3' de deux façons : soit elle favorise la fixation du facteur U2AF sur le site 3', en établissant des interactions entre son domaine RS et le domaine RS de la protéine U2AF³⁵, ce qui stabilise la fixation de U2AF⁶⁵ sur la séquence polypyrimidine. Cette fixation est elle-même indispensable à la fixation du facteur SF1 sur la boîte de branchement. Soit la protéine SR agit en empêchant la fixation de protéines inhibitrices, comme c'est le cas au site A7. Nos données sur le site A2 suggèrent que l'on a, à ce site, un mode d'activation voisin de celui observé à A7, avec une compétition entre les protéines hnRNP A1 et ASF/SF2 pour leur fixation à un même site. Ceci est en accord avec les données que nous avons obtenues *ex vivo*, qui montraient que la protéine ASF/SF2, dont le domaine RS a été éliminé, est capable d'augmenter, presque autant que la protéine sauvage, la production des ARNm *vpr* épissés au site A2. La protéine ASF/SF2 n'a sans doute pas besoin de son domaine RS pour activer l'épissage au site A2. Des test *in vitro*, avec une

protéine ASF/SF2 dont le domaine RS a été éliminé, devront être réalisés pour le confirmer.



Une compétition entre les protéines ASF/SF2 et hnRNP A1 serait donc à la base de la régulation du site A2 (Fig. V.9). L'efficacité d'utilisation du site A2 dépendrait donc directement des concentrations relatives en protéines hnRNP A1 et ASF/SF2 dans la cellule infectée.

On peut se demander si la structure secondaire de l'ARN joue un rôle dans l'inhibition exercée par la protéine hnRNP A1 ? En effet, deux de ses trois sites de fixation au sein de l'élément ESSv sont appariés. Comme dans le cas du site A7, on pourrait imaginer que la protéine hnRNP A1, fixée à la structure tige-boucle 5, interagit avec d'autres molécules de protéine hnRNP A1, fixées sur un élément intronique. Ceci renforcerait l'effet inhibiteur. Une telle hypothèse n'a pas encore été testée.

Un autre point qui devra être vérifié expérimentalement est la compétition directe entre les protéines ASF/SF2 et hnRNP A1. Pour cela, des expériences de pontages covalents aux UV de l'ARN en présence de protéines ASF/SF2 et hnRNP A1 dans un rapport de concentration variable, suivis d'immunoprécipitation, à l'aide d'anticorps dirigés contre ces protéines, seront réalisés.

V.4 Etude des sites A3 à A5

V.4.1 Etat des connaissances sur le site A3 lorsque j'ai débuté cette étude

Les sites A3, A4c, a, b et A5 sont situés dans une région de 200 nts, localisée dans la partie centrale du génome viral codant les protéines Vpr, Tat et Rev (Fig. V.10). Les séquences situées en aval de A3 et en aval de A4b doivent être très conservées puisqu'elles codent 2 protéines en même temps. En phase précoce d'infection, seul les ARNm de la classe 2 kb sont produits. Le site A3 est utilisé en combinaison avec un des 3 sites donneurs D1, D2, D3 et le couple de sites D4-A7, pour produire des ARNm *tat* contenant les deux exons tat1 et tat2 ; de la même manière, les sites accepteurs A4c, a, b et A7 servent à produire les ARNm *rev*, tandis que l'utilisation des sites A5 et A7 conduit à la production des ARNm *nef*. Au cours de la phase tardive d'infection, il n'y a plus d'épissage de l'intron D4-A7 caractéristique des ARNm de la classe 4 kb. L'utilisation du site A3 conduit à la production d'une protéine Tat, codée uniquement par l'exon tat1, du fait de l'insertion d'un codon STOP localisé en aval du site D4. Les sites A4a, b et c et A5 servent eux à la production des protéines Env et Vpu. La production au cours de cycle viral de l'ensemble des protéines Tat, Rev, Nef et Env nécessite donc une régulation très contrôlée de l'utilisation des sites d'épissage A3 à A5.

Comme l'a montré l'étude que nous avons réalisée sur l'effet des protéines SR sur les sites accepteurs de l'ARN de HIV-1, en absence de protéine SR, le site A5 est nettement plus utilisé que A3, à la fois *in vitro* et *ex vivo*, ce qui est en accord avec les données antérieures de Purcell & Martin (1993). Pourtant, au niveau de l'enchaînement de pyrimidines et de la boîte de branchement, le site A3 n'est pas, parmi les 5 sites, celui prédisposé à être le moins utilisé (Fig. V.11). Sur ce point, certaines données de la littérature apportaient des éléments d'information. Ainsi, le remplacement de la séquence polypyrimidine par une séquence plus optimale augmente le taux d'épissage du site A3 (Si *et al.* 1997). L'équipe de M. Stoltzfus avait aussi décrit la présence d'un élément inhibiteur en *cis*, situé à 70 nts du site A3 et dont l'action était position dépendante (Amendt *et al.* 1994, 1995 ; Si *et al.* 1997). Néanmoins, la présence de cet élément inhibiteur ne permettait pas, à lui seul, d'expliquer le faible taux d'utilisation du site A3 dans l'ARN L3-U1 (cf article n° 1). En effet, dans l'ARN C3 qui ne contient pas cet élément, l'utilisation du site A3 reste faible, ce qui suggérait que d'autres paramètres limitent l'efficacité d'épissage de ce site.

Nous avons vu dans le chapitre précédent qu'en présence de concentrations élevées de protéine SC35 ou SRp40, l'utilisation du site A3 devient prédominante et celle du site A5 diminue, ce qui conduit, comme nous l'avons montré, à une baisse de la production de la protéine Nef.

Afin de décortiquer les mécanismes moléculaires des régulations auxquelles le site A3 est soumis, je me suis intéressée avec S. Jacquenet au rôle de la structure secondaire de la région d'ARN viral renfermant le site A3. Puis, j'ai essayé, avec L. Ayadi, de décortiquer les mécanismes de régulation impliquant les différents éléments régulateurs du site A3. Cette étude nous a conduit à identifier un nouvel élément de régulation qui active l'utilisation du site A3.

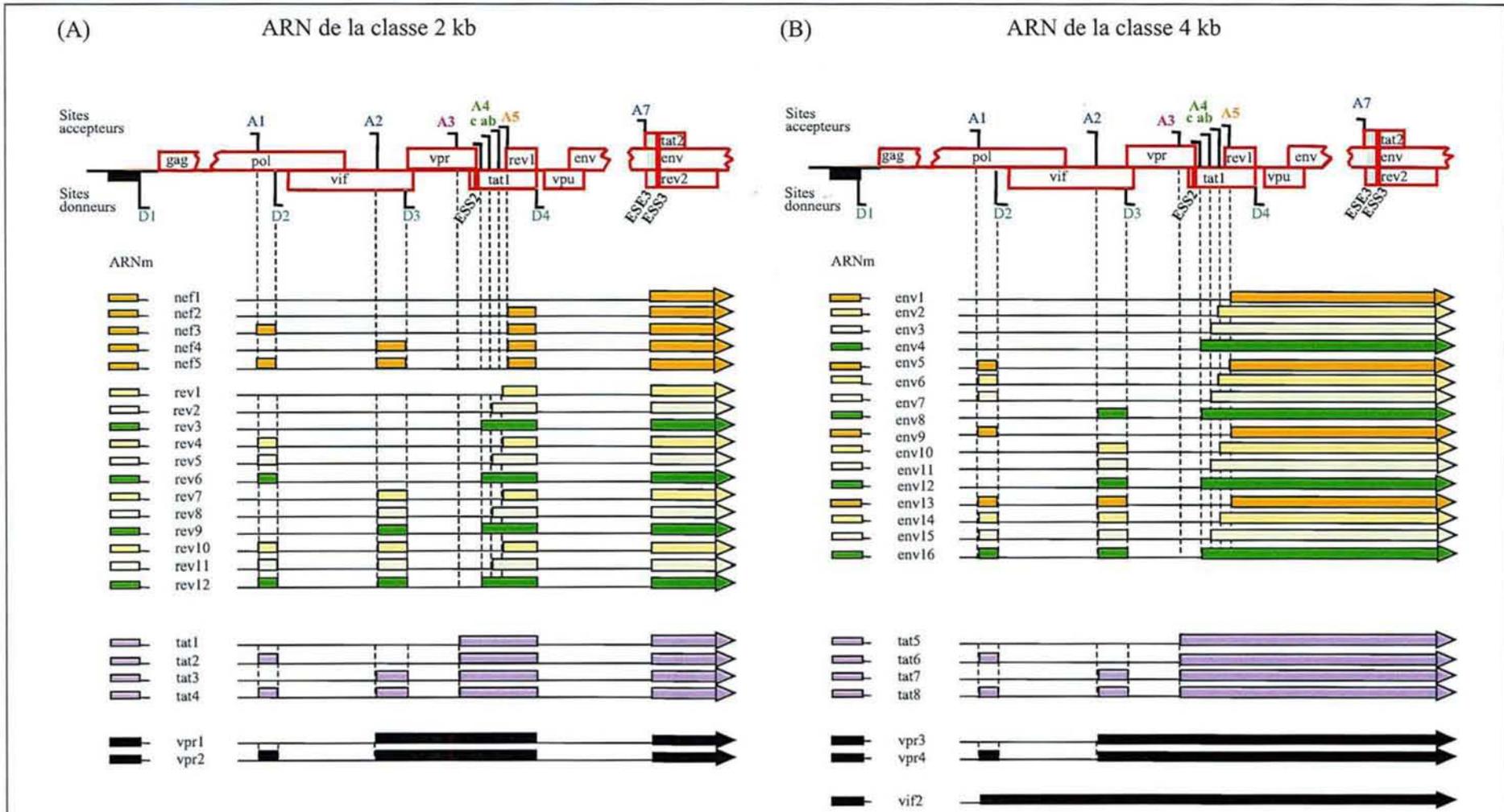


FIG. V.10 – *Etat des connaissances sur la régulation du site A3 au début de notre étude* (d'après Purcell & Martin 1993). Représentation schématique de l'organisation du génome viral. Les rectangles désignent les phases ouvertes de lecture. Les éléments inhibiteurs sont représentés par un trait rouge et les éléments activateurs, en vert. L'identité des facteurs se fixant aux éléments ESS2 et ESS3 n'était pas connue au début de notre étude.

site d'épissage 3'	séquence
consensus	Y (n) N $\begin{matrix} C \\ U \end{matrix}$ AG G
A1	AAUUUUCGGGUUUUUAUACAG G
A2	UAUUACUUUGACUGUUUUUCAG A
A3	ACAACUGCUGUUUAUCCAUUUCAG A
A4a	AGUUUGUUUCACAACAAAAG C
A4b	AGUUUGUUUCACAACAAAAGCCUUAG G
A4c	AAGUGUUGCUUUCUUGCCAAG U
A5	UUAGGCAUCUCCUAUGGCAG G
A7	GGUAUUCACCAUUAUCGUUCAG A

FIG. V.11 – *Comparaison des séquences des sites accepteurs de l'ARN du virus HIV-1 BRU avec la séquence consensus définie pour ces sites (O'Reilly et al. 1995 ; Damier 1997). Les barres verticales représentent les jonctions intron-exon. Y = pyrimidine, N = nucléotide quelconque.*

V.4.2 Etude de la structure secondaire de la région de l'ARN de HIV-1 BRU renfermant les sites A3 à A5 et étude de l'interaction des composants de l'extrait nucléaire avec cette structure

V.4.2.1 Questions posées

La 1^{re} question posée concernait l'effet de la structure secondaire de l'ARN dans cette région sur l'efficacité du site A3. Comme nous l'avons expliqué dans la revue bibliographique, la structure secondaire d'un ARN pré-messager peut moduler son taux d'épissage. Toujours sur le plan de la structure secondaire, on pouvait se demander comment l'élément ESS2, situé à 70 nts en aval du site A3 pouvait générer un phénomène d'inhibition du site A3 ? Est-ce que la structure secondaire de l'ARN favorisait l'action de cet élément ? Existait-il plusieurs conformations, avec pour l'une, un emprisonnement du site A3 dans une structure secondaire impliquant l'élément ESS2 ? Pour apporter des éléments de réponse à cette question, S. Jacquenet avait donc étudié la structure secondaire de cette région de l'ARN viral.

La seconde question était de savoir quels segments de la région de l'ARN viral, contenant les sites A3 à A5, interagissaient avec des composants de l'extrait nucléaire et si les interactions entre composants nucléaires et ARN pouvaient modifier la structure secondaire de cette région de l'ARN de HIV-1. Pour répondre à cette question, il fallait mettre au point des conditions d'analyse de la structure secondaire des ARN et de leur accessibilité dans un extrait nucléaire. Ceci n'avait jamais été fait au

laboratoire. J'ai pris en charge ce travail, sous les conseils d'Annie Mougin, ce qui m'a permis de montrer que la structure secondaire que S. Jacquenet avait établie en solution était conservée dans un extrait nucléaire, et d'identifier les régions protégées de cet ARN. Ces travaux, joints à ceux de S. Jacquenet, ont conduit à un article, où je suis 2^e auteur, publié dans la revue *Nucleic Acids Research* et inséré dans cette partie des résultats.

V.4.2.2 Résultats obtenus

Je commencerai par résumer les résultats de l'analyse de la structure secondaire en solution faite par S. Jacquenet, car ils constituent le point de départ de mon étude. S. Jacquenet avait réalisé une analyse structurale par emploi de sondes enzymatiques dans un tampon Tris (RNase V1) ou AcONa-ZnCl₂ (nucléase S1). Les données obtenues (Fig. 1 de l'article) montraient que le site A3 était localisé dans une boucle terminale d'une structure tige-boucle, très conservée dans les souches appartenant à tous les groupes de virus HIV-1 et même dans le virus SIVcpz GAB (Fig. 2 de l'article). Ceci suggérait une importance fonctionnelle de cette structuration. Par génération de mutants, S. Jacquenet a d'ailleurs pu montrer que cette structure module l'efficacité d'utilisation du site A3 *in vitro* et *ex vivo*. Les sites A4a, b, c et A5 étaient eux situés dans de courtes structures tiges-boucles ou à leur bordure (cf Fig. 1 de l'article).

Enfin, les données montraient que l'élément ESS2 était situé dans une longue structure tige-boucle irrégulière SLS3, près de la boucle terminale, et bien que la séquence de cette région de l'ARN de HIV-1 soit peu conservée dans les souches, la possibilité de former ce type de longue structure irrégulière était conservée (Fig. 1 et 3 de l'article). Il était donc possible que ce type de structuration soit impliqué dans le mécanisme d'action de l'élément ESS2. Pour essayer d'avoir des informations sur ce point, j'ai dans un 1^{er} temps concentré mes efforts sur l'étude du devenir et de la protection de la structure tige-boucle dans un extrait nucléaire en conditions d'épissage (travaux présentés dans l'article).

Les empreintes étant réalisées en tampon D et non dans les tampons utilisés auparavant par S. Jacquenet, j'ai dû étudier la structure de l'élément SLS3 dans le tampon D, en solution et en extrait nucléaire. Pour cela, j'ai réalisé une construction S3 permettant de produire, par transcription *in vitro* en utilisant l'ARN polymérase du phage T7, un ARN correspondant à la tige-boucle SLS3 (Fig. 4A de l'article). J'ai réalisé les analyses à la fois avec des RNases (V1, T1 et T2) et des agents chimiques (DMS et kéthoxal) et j'ai pu montrer que cette structure était conservée en tampon D. La mise au point expérimentale permettant l'analyse dans l'extrait nucléaire en conditions d'épissage a été rendue difficile par la forte concentration en protéines de cet extrait et par la présence de certains réactifs chimiques utilisés pour l'épissage qui gênaient l'activité des enzymes. Nous ne ressortirons ici que les points majeurs de cette mise au point.

Les complexes ARN/protéines ont été formés dans des conditions d'épissage, pendant 5 minutes à 0°C puis 5' à 30°C, en présence de 4 µl d'extrait nucléaire, et 5 µg d'ARNt de levure ont été ajoutés au mélange, afin d'empêcher la formation de complexes aspécifiques. La réactivité des sondes enzymatiques et chimiques dans ces conditions expérimentales a ensuite été étudiée. L'un des composants de la réaction d'épissage, l'alcool polyvinylique (PVA), s'est avéré être un inhibiteur des sondes enzy-

matiques ; nous avons donc cessé d'ajouter ce composé dans nos réactions et le mélange réactionnel sans PVA s'est avéré satisfaisant pour la réalisation des hydrolyses enzymatiques. D'autre part, le pouvoir tampon de la solution D n'était pas suffisamment fort pour compenser la baisse de pH observée lors de l'ajout de sondes chimiques ; il a donc été nécessaire d'ajouter au mélange réactionnel du tampon cacodylate 50 mM (concentration finale) pH 7,5, habituellement utilisé pour l'emploi du DMS et kéthoxal.

Les réactions par les agents chimiques et enzymatiques ont été mises au point pour obtenir des conditions d'hydrolyse ou de modification ménagées. Les réactions ont ainsi été effectuées à 30°C pendant 10 minutes, en présence de 0,4 U de RNase V1, 0,1 U de RNase T1, 2U de RNase T2, 1 μ l de DMS 25% ou 2 μ l de kéthoxal à 37 mg/ml. Afin d'empêcher l'apparition de modifications ou de coupures supplémentaires lors de l'élimination des protéines, l'arrêt des réactions a été effectué avant la déprotéinisation par ajout de 20 μ g d'ARNt de levure à chaque échantillon et incubation dans la glace. Ces conditions étaient suffisantes pour arrêter les RNases T1 et T2. Il fallait, en plus, ajouter 100 μ l d'EDTA 100 mM dans le cas de la RNase V1, 3 μ l de β -mercaptoéthanol 6 M pour le DMS, et 20 μ l de borate de potassium 0,5 M pour le kéthoxal. Enfin, la grande quantité de protéines présentes dans l'extrait nucléaire (13 μ g/ μ l, soit 52 μ g par réaction) nécessitait de prendre beaucoup de précautions lors de l'extraction phénolique, afin de perdre le moins d'ARN possible.

Les données que j'ai obtenues montraient une grande similitude d'accessibilité aux RNases de l'ARN S3 et de la tige-boucle SLS3 au sein de l'ARN L3, ce qui confirmait bien qu'il s'agissait d'un élément autonome de structuration de l'ARN L3. Cela démontrait aussi que la structure secondaire de cet élément était stable à une concentration en magnésium plus faible que celle que S. Jacquenet avait utilisée (2,8 mM au lieu de 10 mM). De plus, grâce à l'utilisation des sondes chimiques, mes données permettaient d'affiner les connaissances sur la structure secondaire de l'élément SLS3. Il en ressortait que le degré de structuration des hélices était variable le long de l'élément SLS3 (Fig. 4C1 de l'article). Alors que les hélices c et g par exemple sont très stables comme en témoigne l'absence de modifications chimiques, les hélices d, e et f - qui contient l'élément ESS2 - sont peu stables. En effet, un nombre élevé de modifications chimiques est observé au niveau de l'élément ESS2.

Autre point important, le niveau de protection des différents éléments qui constituent SLS3 était variable après incubation dans l'extrait nucléaire (Fig. 4C2 de l'article). Ainsi, alors que deux régions étaient fortement protégées, l'une renfermant une séquence simple-brin riche en purines et les nucléotides qui lui font face dans la structure (motif A), l'autre, constituée de la boucle terminale et de l'extrémité 3' de l'élément ESS2 (motif B), la région les séparant, constituée des hélices d, e et f qui contiennent justement l'élément ESS2, était peu protégée. Certains résidus devenaient même plus réactifs aux agents chimiques et coupures enzymatiques dans l'extrait nucléaire.

Ces données soulevaient un problème, car au moment où je réalisais ce travail, Caputi *et al.* (1999) décrivaient la fixation de la protéine hnRNP A1 sur un segment de 60 nts de l'ARN de HIV-1 renfermant l'élément ESS2 et proposaient, d'après d'autres données, que la protéine hnRNP A1 soit responsable de l'activité inhibitrice de l'élément ESS2. La fixation de la protéine hnRNP A1 sur cette même région avait aussi été observée par Del Gatto-Konczak *et al.* (1999). Un site potentiel de fixation pour la protéine hnRNP A1 UAGAGC, positionné de 5434 à 5439, était néanmoins protégé en extrait nucléaire dans nos expériences.

La question se posait donc de savoir si l'absence de protection de l'élément ESS2 dans l'ARN S3 était liée à l'absence de site A3. C'est pourquoi nous avons ensuite repris l'étude sur un ARN plus long renfermant le site A3 et ces données sont décrites en annexe de l'article. Enfin, l'observation d'une protection de la région en simple-brin 5398 à 5403, nous a conduit, comme nous le verrons plus loin à identifier un élément activateur en *cis* du site A3.

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.

Jacquet S., Ropers D., Bilodeau P. S., Damier L., Mouglin A., Stolzfus C.M. and Branlant C (2001) *Nucleic Acids Research* **29** (2), 464-478.

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

Sandrine Jacquenet, Delphine Ropers, Patricia S. Bilodeau¹, Laurence Damier, Annie Mougín, C. Martin Stoltzfus¹ and Christiane Branlant*

Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 UHP-CNRS, Université Henri Poincaré Nancy 1, Boulevard des Aiguillettes, BP239, 54506 Vandoeuvre-lès-Nancy cedex, France and

¹Department of Microbiology, University of Iowa, Iowa City, IA 52242, USA

Received August 26, 2000; Revised October 26, 2000; Accepted November 22, 2000

ABSTRACT

The HIV-1 transcript is alternatively spliced to over 30 different mRNAs. Whether RNA secondary structure can influence HIV-1 RNA alternative splicing has not previously been examined. Here we have determined the secondary structure of the HIV-1/BRU RNA segment, containing the alternative A3, A4a, A4b, A4c and A5 3' splice sites. Site A3, required for *tat* mRNA production, is contained in the terminal loop of a stem–loop structure (SLS2), which is highly conserved in HIV-1 and related SIVcpz strains. The exon splicing silencer (ESS2) acting on site A3 is located in a long irregular stem–loop structure (SLS3). Two SLS3 domains were protected by nuclear components under splicing condition assays. One contains the A4c branch points and a putative SR protein binding site. The other one is adjacent to ESS2. Unexpectedly, only the 3' A residue of ESS2 was protected. The suboptimal A3 polypyrimidine tract (PPT) is base paired. Using site-directed mutagenesis and transfection of a mini-HIV-1 cDNA into HeLa cells, we found that, in a wild-type PPT context, a mutation of the A3 downstream sequence that reinforced SLS2 stability decreased site A3 utilization. This was not the case with an optimized PPT. Hence, sequence and secondary structure of the PPT may cooperate in limiting site A3 utilization.

INTRODUCTION

Retroviruses transcribe their RNA from an integrated proviral genome. The unspliced primary transcript has two functions after transport to the cytoplasm: (i) as messenger RNA for the production of viral proteins, (ii) as genomic RNA which is packaged into newly synthesized virions. HIV-1 is a complex retrovirus of the lentivirus class. Due to the presence of four splicing donor sites (5'ss) and eight splicing acceptor sites

(3'ss) in HIV-1 RNA, the splicing machinery in the infected cell generates at least 40 distinct mRNAs from the unique RNA primary transcript (1). The relative abundance of these mRNAs depends greatly upon the relative efficiencies of splicing at the 3'ss. All of these 3'ss are suboptimal (2–7). Five 3'ss (A3, A4c, A4a, A4b and A5), located in a small central portion of the viral RNA, are in competition with one another. All *tat* mRNAs are spliced at site A3, the *rev* mRNAs are spliced at sites A4a, A4b or A4c and most *nef* and *env* mRNAs are spliced at site A5 (1; for review see 8). These five central 3'ss are all used in the early phase after infection for production of the Tat, Rev and Nef proteins.

The steady-state level of double-spliced *tat* mRNA is considerably lower than the levels of double-spliced *rev* and single-spliced *env/vpu* mRNAs in both lymphoid and non-lymphoid cells (1). It is possible that the *tat* mRNA is less stable than the *rev* and *env* mRNAs. Alternatively, the A3 3'ss may be less efficient than the other downstream 3'ss. Based on studies of the A3 3'ss (2,3,9), the latter explanation appears to be more likely. Indeed, improvement of the efficiency of the A3 3'ss strongly increases the level of the *tat* mRNAs in transfected HeLa cells (10). All HIV-1 splicing sites have a suboptimal polypyrimidine tract (PPT). Surprisingly, in spite of the low splicing efficiency at site A3, its PPT is more uridine-rich than any of the other 3'ss in the central region of the HIV-1 genome. Nevertheless, the A3 PPT sequence is one of the parameters that limits A3 3'ss efficiency, since increasing the number of U residues in the A3 PPT increases the use of site A3 *in vitro* (9). A downstream exon splicing silencer (ESS2) is located 64 nt downstream from site A3 (2,9) and inhibits an early step of spliceosome assembly (3). ESS2 also inhibits splicing when the PPT of the upstream 3'ss is optimized but to a smaller extent (9). This means that the observed low efficiency of the A3 3'ss results from a combined effect of the presence of a suboptimal PPT and ESS2 element. ESS2 shows an inhibitory effect when inserted into a heterologous retroviral context (downstream from Rous sarcoma virus *src* gene 3'ss). In this heterologous context, the effect of ESS2 was found to be distance dependent (2) and a 10 nt sequence of ESS2 was sufficient for the inhibition. However, the magnitude of ESS2

*To whom correspondence should be addressed. Tel: +33 3 83 91 20 91; Fax: +33 3 83 91 20 93; Email: christiane.branlant@maem.uhp-nancy.fr

Table 1. The oligonucleotides used in this study

Primer number	Sequence (5'-3')	Orientation	Restriction site	HIV-1 BRU sequence	HIV-1 MAL sequence	Utilization
O-605	<u>TATTCTGGATCCGGTCTCTCTGGT</u>	i	<i>Bam</i> HI	1-12		PCR
O-606	<u>TATTCTAGATC7CCCATCGATCTA</u>	c	<i>Bg</i> II	373-385		PCR
O-567	<u>AGAAGACCAAGATCTAGAGGGAGCCA</u>	i	<i>Bg</i> II	5155-5165 and 5172-5182	5151-5158 and 5165-5175	PCR
O-608	<u>TATTCTC7GAGTTGCTCTCTCT</u>	c	<i>Pst</i> I	5397-5408		PCR primer extension
O-521	<u>TATTCTC7GCAGTGCITTTGATAGA</u>	c	<i>Pst</i> I	5615-5626		PCR
O-1458	<u>TATTCTC7GCAGTTGCTCTCTCT</u>	c	<i>Pst</i> I		5390-5401	PCR primer extension
O-1459	<u>TATTCTC7GCAGATACTACTTACTGCTC</u>	c			5615-5630	PCR
O-998	<u>TCTAATACGACTCACTATAGGGCGACATAGCAGAATAGGCG</u>	i		5369-5387		PCR
O-954	<u>TATTCTCTGCAGGATCCGGCAATGAAAGC</u>	c	<i>Bam</i> HI	5503-5514		PCR
O-825	<u>AATTCTGCAACAACttcGTTTATtcTTTCAGAAATTGGGTG</u>	i		5326-5340 and 5353-5367		Site-directed mutagenesis
O-826	<u>CCATTTTCAGAAATTGGaTaaACAGCAGAATAGGCGTT</u>	i		5350-5364 and 5375-5389		Site-directed mutagenesis
O-994	<u>ttcTTTCAGAAATgaaaTaaACgaaAGAATAGGCGTT</u>	i		5353-5361 and 5378-5389		Site-directed mutagenesis
O-548	<u>GGCAATGAAAGCAAC</u>	c		5500-5514		Primer extension
O-831	<u>GAGGTCTTCGTCGCTG</u>	c		5571-5585		Primer extension

The number, sequence and utilization of each oligonucleotide are given. Sequences identical (i) or complementary (c) to HIV-1 regions are underlined. Nucleotide positions of these HIV-1 RNA regions are numbered according to the GenBank accession no. K02013 (HIV-1/BRU) and X04415 (HIV-1/MAL). Restriction sites introduced by the oligonucleotides are indicated and the sequences corresponding to those sites are in italic. Lower cases indicate the mutations introduced by site-directed mutagenesis in the HIV-1/BRU RNA.

inhibition in this heterologous context was reduced compared to its effect in its natural context downstream from the A3'ss (9), suggesting that in this natural context, additional sequences or the RNA secondary structure may reinforce ESS2 inhibitory property. The hnRNP A/B proteins were recently proposed to bind ESS2 element and appear to be involved in the observed inhibition (11). However, the precise mechanism of inhibition of spliceosome assembly by ESS2 element is still unknown.

The secondary structure of the HIV-1 RNA region containing site A3 and ESS2 has not been previously studied. It may also contribute to the low efficiency of splicing at site A3. Indeed, several studies have shown an effect of RNA secondary structure on alternative splicing in cellular pre-messenger RNAs and some viral mRNAs. In some cases, RNA secondary structure acts as a splicing enhancer by decreasing physical distances between splice sites (12-15), or by maintaining one or more *cis*-acting regions of pre-mRNA in a favorable configuration for interaction with *trans*-acting factors (16-19). However, in several other cases, sequestration of 5'ss or 3'ss was found to inhibit or to limit splicing site utilization (20-22).

To test for a potential role of stem-loop structures in determining the splicing efficiency at the HIV-1 A3 3'ss, we studied the secondary structure of the HIV-1 RNA region containing site A3 and ESS2, for the BRU and the MAL HIV-1 strains. We also developed conditions to investigate RNA secondary structure in the presence of nuclear extract. Because the A3 3'ss was found to be located in a stem-loop structure

that is highly conserved in HIV-1 and related SIVcpz strains, we tested whether the stability of this stem-loop structure influences splicing at site A3.

MATERIALS AND METHODS

Plasmid construction

Plasmid pBRU3 (23) was used as the source of cDNA sequences from the HIV-1 BRU/LAI strain (GenBank accession no. K02013). PCR amplifications were done according to Nour *et al.* (24). Three distinct HIV-1/BRU cDNA fragments, H1, H2 and H3, corresponding to regions 1-385, 5155-5408 and 5155-5626, respectively, were prepared (numbering according to Ratner *et al.*, 25). PCR amplifications of fragments H1, H2 and H3 were performed using pairs of specific DNA primers (O-605/O-606, O-567/O-608, O-567/O-521, respectively) (Table 1).

All the constructs were made in plasmid pBluescriptKSII⁺ cleaved with the *Bam*HI and *Pst*I nucleases. PCR-amplified fragments were digested using restriction sites generated by the primers (Table 1). In plasmid pLD-C3, fragments H1 and H2 have been ligated together using their *Bg*II restriction sites. Similarly, in plasmid pLD-L3, fragments H1 and H3 have been ligated together. Recombinant plasmids pBluescriptKSII⁺ containing fragment H2 (pA3C) or H3 (pA3L) were prepared.

The HIV-1/MAL region 5148-5401 (fragment H4) and the HIV-1/MAL region 5148-5630 (fragment H5) were PCR amplified with the pairs of primers O-567/O-1458 and O-567/O-1459, respectively (Table 1) using the entire HIV-1/MAL cDNA (26) as

the template. Hybrid recombinant plasmids pBluescriptKSII⁺ (pC3-MAL and pL3-MAL), containing BRU and MAL cDNA sequences, were obtained by insertion of the digestion product of fragment H1 together with the digestion product of fragment H4 or H5, respectively.

The HIV-1/BRU cDNA region encoding the stem-loop 3 RNA fragment containing ESS2 (S3) was PCR amplified from plasmid pLD-L3 with primer O-998 and O-954. Primer O-998 generates a T7 RNA polymerase promoter.

Plasmid pΔPSP, used for the transfection experiments, was constructed in the following way. Infectious HIV-1 plasmid pNL4-3 (GenBank accession no. M19921) was cleaved with *SpeI* and *BalI* to generate an 11.9 kb fragment. This was ligated together with oligonucleotides 5'-CTAGACGCGTTTGG-3' and 5'-CCAAACGCGT-3', which had been previously annealed, to form a double-stranded linker. This created an HIV-1 plasmid deleted between nt 1511 and 4551.

Site-directed mutagenesis

For site-directed mutagenesis of plasmids pLD-C3 and pLD-L3, the inserted *BamHI-PstI* fragment of each plasmid was cloned into phage M13mp9. For both M13mp9 recombinants, site-directed mutagenesis was made according to Kramer *et al.* (27), using the oligonucleotides O-825 and O-826 (Table 1). The mutagenized *BamHI-PstI* DNA fragments were reinserted into plasmid pBluescriptKSII⁺ and the resulting plasmids were designated as pSJ-C3.825, pSJ-C3.826, pSJ-L3.825 and pSJ-L3.826, respectively. Constructs pSJ-C3.994 and pSJ-L3.994 were obtained after a second run of site-directed mutagenesis of the *BamHI-PstI* fragments of plasmids pSJ-C3.825 and pSJ-L3.825. Oligonucleotide O-994 was used to generate the mutation (Table 1). The resulting plasmids were designated as pSJ-C3.994 and pSJ-L3.994, respectively. pΔPSP variants were constructed by replacement of the HIV-1/pNL4-3 *EcoRI-Bsu36I* fragment (nt 5743–5955) with the corresponding HIV-1/BRU *EcoRI-Bsu36I* fragment (nt 5325–5537) from plasmids pLD-L3, pSJ-L3.825, pSJ-L3.826 and pSJ-L3.994. All plasmids with mutated DNA were sequenced to confirm the expected base changes.

In vitro transcription

Prior to transcription with T7 RNA polymerase, all the plasmids were linearized with the *PstI* nuclease. For splicing assays, uniformly labeled capped transcripts were produced as previously described (7). Synthesis of cold transcripts for RNA secondary structure analysis was carried out in a 30 μl mixture, containing 1 μg of linearized plasmid or 1.7 pmol of PCR product, as previously described (15). The RNA was dissolved in 80 μl of sterile water, 1 μl of this solution was used for each chemical reaction or enzymatic digestion.

Chemical and enzymatic probing of RNA in solution

Enzymatic digestions with V1 RNase and S1 nuclease were carried out, as previously described (15), using 1 μl of *in vitro* synthesized RNA in the presence of 1.25 μg of a commercial yeast tRNA mixture (Roche Diagnostics) for V1 RNase digestion and 2.5 μg of yeast tRNA for S1 nuclease digestion. V1 RNase was prepared from *Naja oxiana* venom (28) and S1 nuclease was purchased from Amersham Pharmacia Biotech. Prior to enzymatic digestion, the RNA transcript mixed with the tRNA mixture was pre-incubated for 20 min at 20°C, in 50 μl of a

50 mM KCl, 10 mM MgCl₂, 10 mM Tris-HCl, pH 7.5 solution for V1 RNase digestion and in 50 μl of a 25 mM KCl, 0.5 mM ZnCl₂, 5 mM MgCl₂, 12.5 mM AcONa pH 4.5 solution for S1 nuclease digestion. Incubation was for 10 min at 20°C with 0.1 U of V1 RNase or for 12 min at 20°C with 3 U of S1 nuclease. V1 RNase digestion was stopped by addition of 1 μl of 100 mM EDTA, followed by phenol extraction. S1 nuclease digestion was stopped by adding 20 μg of tRNA, chilling on ice and phenol extraction. After phenol extraction, V1 and S1 digested RNAs were ethanol precipitated, dissolved in 3 μl of distilled water and 1 μl was used for reverse transcription analysis.

Prior to 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide-metho-*p*-toluene sulfonate (CMCT) reaction, 1 μl of *in vitro* synthesized RNA mixed with 10 μg yeast tRNA were pre-incubated for 20 min at 20°C in 50 mM KCl, 10 mM MgCl₂, 100 mM sodium borate pH 8, buffer with 3.2 mg of CMCT per 300 μl assay. Incubation was for 30 min at 20°C. At the end of the reaction, RNA was ethanol precipitated and treated as RNA cleaved by enzymes.

Chemical and enzymatic probing of RNA in nuclear extract

Conditions for enzymatic digestion and chemical modification were modified to allow analysis of RNA secondary structure in the presence of the large amounts of protein contained in the nuclear extract. S1 nuclease could not be used under splicing reaction conditions as its optimum pH of action is of 4.5 and splicing reactions are performed at pH 7.9. Because of this, we used T1 and T2 RNases to probe the single-stranded regions. Polyvinylalcohol (PVA) used for splicing assays has an inhibitory effect on V1, T1 and T2 RNase action and therefore enzymatic digestions and chemical modifications were performed in the splicing buffer in the absence of PVA. Secondary structure analysis was performed with 100 ng of an *in vitro* produced transcript. Prior to enzymatic digestion, the RNA transcript was incubated for 5 min at 0°C, followed by 5 min at 30°C, in 4 μl of HeLa cell nuclear extract in buffer D (0.2 mM EDTA, 0.1 M KCl, 0.5 mM DTT, 0.25 mM PMSF, 20% glycerol, 20 mM HEPES pH 7.9; adapted to Dignam *et al.*, 29) under the conditions used for *in vitro* splicing assays (2.5 mM MgCl₂ and 1 mM ATP), except that PVA was omitted and 5 μg of tRNA were added to the reaction mixture. Incubation was performed for 10 min at 30°C with 0.4 U of V1 RNase, 0.1 U of T1 RNase or 2 U of T2 RNase. V1 RNase digestion was stopped by addition of 2 μl of 100 mM EDTA. T1 and T2 RNase digestions were stopped by cooling the mixture on ice and adding of 20 μg of tRNA. RNA was immediately phenol extracted and ethanol precipitated. It was then dissolved in 4 μl of distilled water and 1 μl was used for primer extension analysis. In parallel, for comparison, V1, T1 and T2 RNase digestions were performed in the same conditions on the same amounts of RNA transcript pre-incubated in the splicing buffer D, instead of the nuclear extract.

We were not able to develop conditions for efficient modification of RNA with CMCT under splicing conditions. To obtain efficient DMS modification, sodium cacodylate had to be added to a 50 mM final concentration. To test for accessible G residues, we used kethoxal. In this case, potassium borate had to be added to a 50 mM final concentration to obtain efficient modification. Incubations with 1 μl of 25% DMS-ethanol solution, or 2 μl of kethoxal (37 mg/ml) were performed for 10 min at 30°C. DMS modifications were stopped by addition

of 3 μ l of 6 M β -mercaptoethanol. Kethoxal modifications were stopped by addition of 20 μ l 0.5 M potassium borate. In both cases, RNA was immediately phenol extracted. RNA modified by kethoxal was dissolved in 4 μ l of 25 mM potassium borate, RNA modified by DMS was dissolved in 4 μ l of sterile water and 1 μ l of each solution was used for reverse transcriptase analysis. Here also, for comparison, DMS and kethoxal modifications were performed in the same conditions on the same amounts of RNA transcript incubated in the splicing buffer D, instead of the nuclear extract.

Identification of modification and cleavage positions and secondary structure prediction

Positions of chemical modifications and enzymatic cleavages were identified by primer extension analysis with the avian myeloblastosis virus reverse transcriptase (Life Science). Four oligonucleotides were used as primers: O-608, O-831, O-548 and O-1458 (Table 1). Oligonucleotide primers were 5'-end labeled with [γ - 32 P]ATP 3000 Ci/mmol (Amersham Pharmacia Biotech). Annealing of primers and primer extension were performed as previously described (15). To prepare sequencing ladders of unmodified RNA, dideoxynucleotide:deoxynucleotide mixtures at a 1:2 ratio were used. The cDNA products were fractionated by electrophoresis on 7% polyacrylamide sequencing gels.

Predictions of secondary structure were made with the MFold program of the GCG software, version 8.1 Unix (1995), which is based on the thermodynamic values proposed by Jaeger *et al.* (30) for an RNA in solution in 1 M NaCl buffer at 37°C. The results of the experimental analysis were introduced in the computer search, in order to define the secondary structure that had the best fit with the experimental data.

The GenBank accession numbers of HIV-1 or related HIV-1 sequences used in the comparative analysis are: K02013 (HIV-1/BRU), M19921 (HIV-1/NL4-3), K03455 (HIV-1/HXB2), M38429 (HIV-1/JRCSF), U39362 (HIV-1/P896), M17449 (HIV-1/MN), D10112 (HIV-1/CAM1), M38431 (HIV-1/NY5CG), K02007 (HIV-1/SF2), X04415 (HIV-1/MAL), K03454 (HIV-1/ELI), L20587 (HIV-1/ANT70C), X52154 (SIVcpzGAB) and U42720 (SIVcpzANT).

In vitro splicing assays

In vitro splicing assays were performed with HeLa cell nuclear extracts from the Computer Cell Culture Center S.A. (Belgium), using 10^5 Cerenkov c.p.m. of RNA transcript per assay, as previously described (7). The reaction mixture was prepared on ice and then incubated at 30°C for 150 min. Spliced products were deproteinized with proteinase K, phenol extracted and analyzed on a 5% polyacrylamide sequencing gel. Splicing efficiency was estimated by scanning the gel with a Molecular Dynamics Phosphorimager using the Image Quant Software, version 3.3. The M:P ratio (moles of mature RNA versus the moles of residual precursor) was determined for each transcript taking into account the estimated radioactivity and the number of uracil residues per molecule.

In vivo splicing assays

HeLa cells were transfected by the modified calcium phosphate coprecipitation technique with 12 μ g plasmid DNA as described above (10). Total cellular RNA was isolated from transfected HeLa cells 48 h post-transfection and 3 μ g of RNA

was reversed transcribed and PCR amplified with forward oligonucleotide primer BSS (5'-GGCTTGCTGAAGCGCG-CACGGCAAGAGG-3'; nt 700-727) and reverse primer SJ4.7A, which spans sites D4 and A7 (5'-TTGGGAGGT-GGGTTGCTTTGATAGAG-3'; nt 8369-8381 and 6032-6044) as described previously (10). After confirmation of amplified spliced product by PAGE, amplification products (100 ng) were radiolabeled by performing a single round of PCR with the addition of 10 μ Ci [α - 32 P]dCTP and the products were analyzed by electrophoresis on a 6% polyacrylamide 7 M urea gel.

RESULTS

Secondary structure analysis of the HIV-1 RNA region containing site A3

We first studied the secondary structure of the region from positions 5303 to 5559 of the HIV-1/BRU RNA in the L3 transcript (Fig. 1A). Transcript L3 contains the 5' extremity of the HIV-1/BRU RNA (positions 1-385) fused to a central portion of the HIV-1/BRU RNA (positions 5172-5626). Transcript L3 carries the D1 5' ss and five 3' ss (A3, A4c, A4a, A4b and A5). In order to detect RNA secondary structure artifacts that may be due to the absence of the authentic partners of the sequences located at the extremities of the studied fragments, analyses of a given HIV-1 RNA region were performed on several RNA fragments of different lengths. To this end, in addition to the L3 transcript, the C3 transcript lacking RNA sequence between nt 5409 and 5626 downstream from site A3, and the A3C and A3L transcripts, which do not contain the 5' part of the HIV-1/BRU RNA, were also subjected to chemical and enzymatic probing experiments (Fig. 1A). The single-stranded specific chemical reagent, CMCT, and two RNases (V1 and S1 with specificities for double- and single-stranded RNA, respectively) were used in the analysis. Positions of chemical modifications and enzymatic cleavages were identified by primer extension analysis, using the oligonucleotide primers given in Table 1. Relevant examples of primer extension analyses of enzymatic cleavages and chemical modifications of transcript L3 are shown in Figure 1B. Identical patterns of chemical modifications and enzymatic cleavages were obtained for the common parts of the four transcripts studied (data not shown). This demonstrated that the presence of ESS2 in the transcript did not alter the secondary structure at the A3 3' ss.

A unique model of RNA secondary structure was derived from all the experimental data and is shown in Figure 1C. In this model, site A3 is located in the terminal loop of stem-loop structure 2 (SLS2). This stem-loop structure is preceded by a smaller stem-loop structure 1 (SLS1) of lower stability and followed by a long irregular stem-loop structure 3 (SLS3), that contains the ESS2 element (Fig. 1C). Although several internal loops and bulges interrupt the helical structure of SLS3, the calculated stability at 37°C of SLS3 is high ($\Delta G = -22.3$ kcal/mol). Two short stem-loop structures 4 and 5 (SLS4 and SLS5) were detected in the segment containing sites A4c, A4b, A4a and A5 (Fig. 1C).

As a test for the biological significance of the proposed structure, we examined the sequences of other HIV-1 and SIV strains for compensatory (replacement of a Watson-Crick base pair by another Watson-Crick base pair) or semi-compensatory (substitution of a Watson-Crick base pair by a G-U or U-G

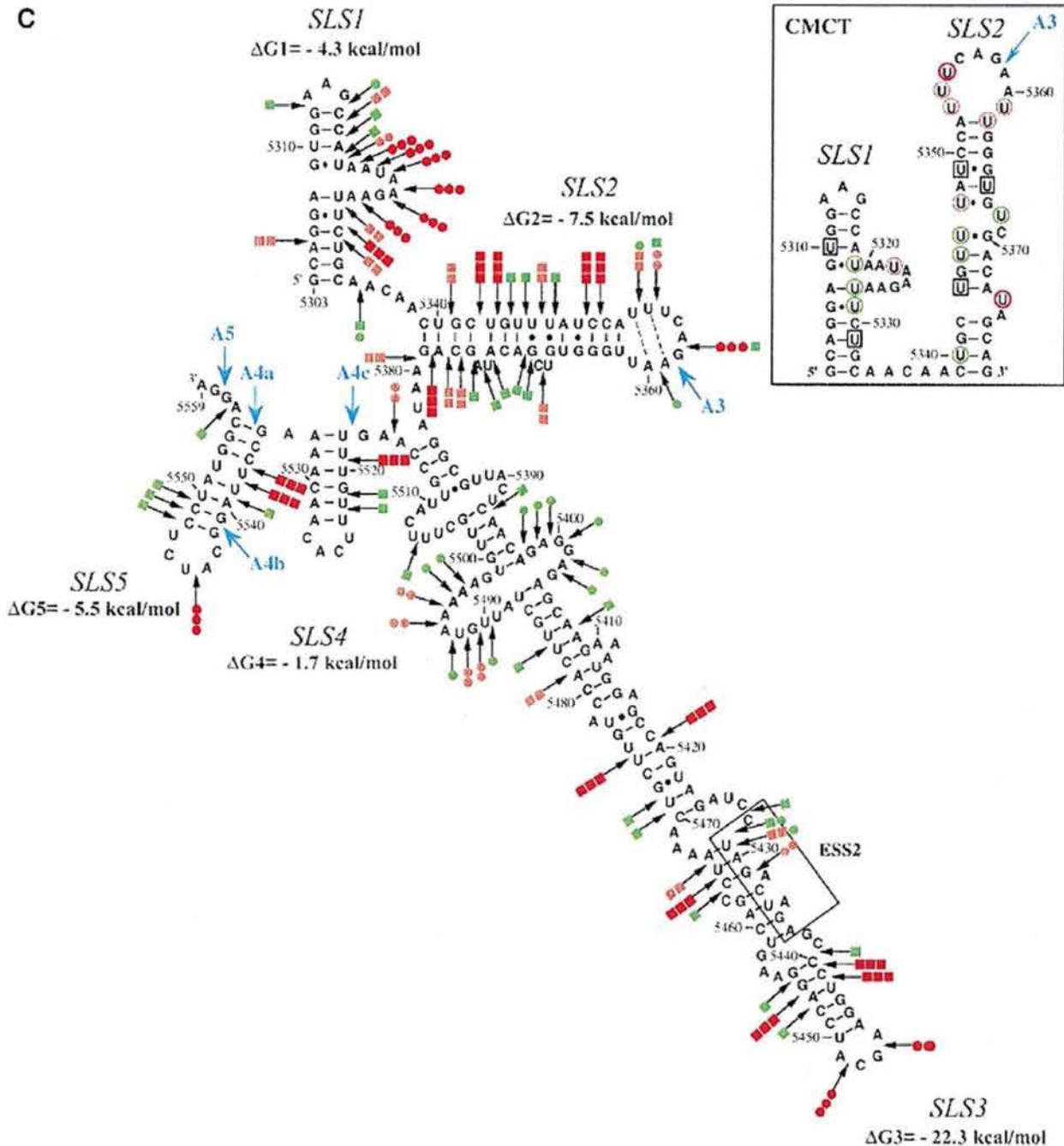


Figure 1. (Opposite and above) Secondary structure analysis of the HIV-1/BRU RNA region from positions 5303 to 5559. (A) Schematic representation of the HIV-1 genome and of the RNAs used for secondary structure analysis. In the HIV-1 genome, the 5'ss (D) and 3'ss (A) are shown; boxes, open reading frames. In the four RNA transcripts (L3 to A3C), numbering of HIV-1/BRU RNA sequences is according to Ratner *et al.* (25), exon sequences are represented by rectangles, introns by thin lines, the thick horizontal lines at the beginning of the transcripts correspond to sequences generated by plasmid pBluescriptKSI⁺. The 5'ss, 3'ss and ESS2 are indicated. In transcripts C3 and L3, the junction between the two HIV-1 RNA regions within the intron is indicated by a vertical broken line. (B) Examples of primer extension analyses of enzymatically digested and chemically modified L3 transcript. Lanes marked by V, S and CMCT correspond to V1 RNase digestion, S1 nuclease digestion and CMCT modification, respectively. Conditions for digestion and modification are given in Materials and Methods. Lanes marked by c correspond to control experiments with the untreated RNA transcript, lanes UGCA to the sequencing ladders. Numbering of nucleotides in the HIV-1/BRU RNA is on the right of the autoradiogram, as well as the positions of the 3'ss (A3, A4c, A4a, A4b and A5). Positions of the various stem-loop structures identified on the basis of this analysis (Fig. 1C) and of ESS2 are shown on the left. The primers used for extension with reverse transcriptase are indicated below each autoradiogram. (C) The results of enzymatic probing are schematically represented on the proposed secondary structure model. Cleavages by enzymes are shown by arrows surmounted with circles for S1 nuclease and squares for V1 RNase. Three red circles or squares indicate a strong cleavage, two orange circles or squares a medium cleavage, one green circle or square a low cleavage. The free energy of the proposed stem-loop structures at 37°C, in 1 M NaCl were calculated with the Mfold software. Positions of nucleotides in the HIV-1/BRU RNA are given. Stem-loop structures are designated as SLS1-5. The various 3'ss are indicated by blue arrows. The ESS2 inhibitory element is squared. Results of CMCT modification of SLS1 and SLS2 are shown in the inset. Red circles indicate a strong, orange a medium and green a low level of modification. Squared nucleotides were not modified.

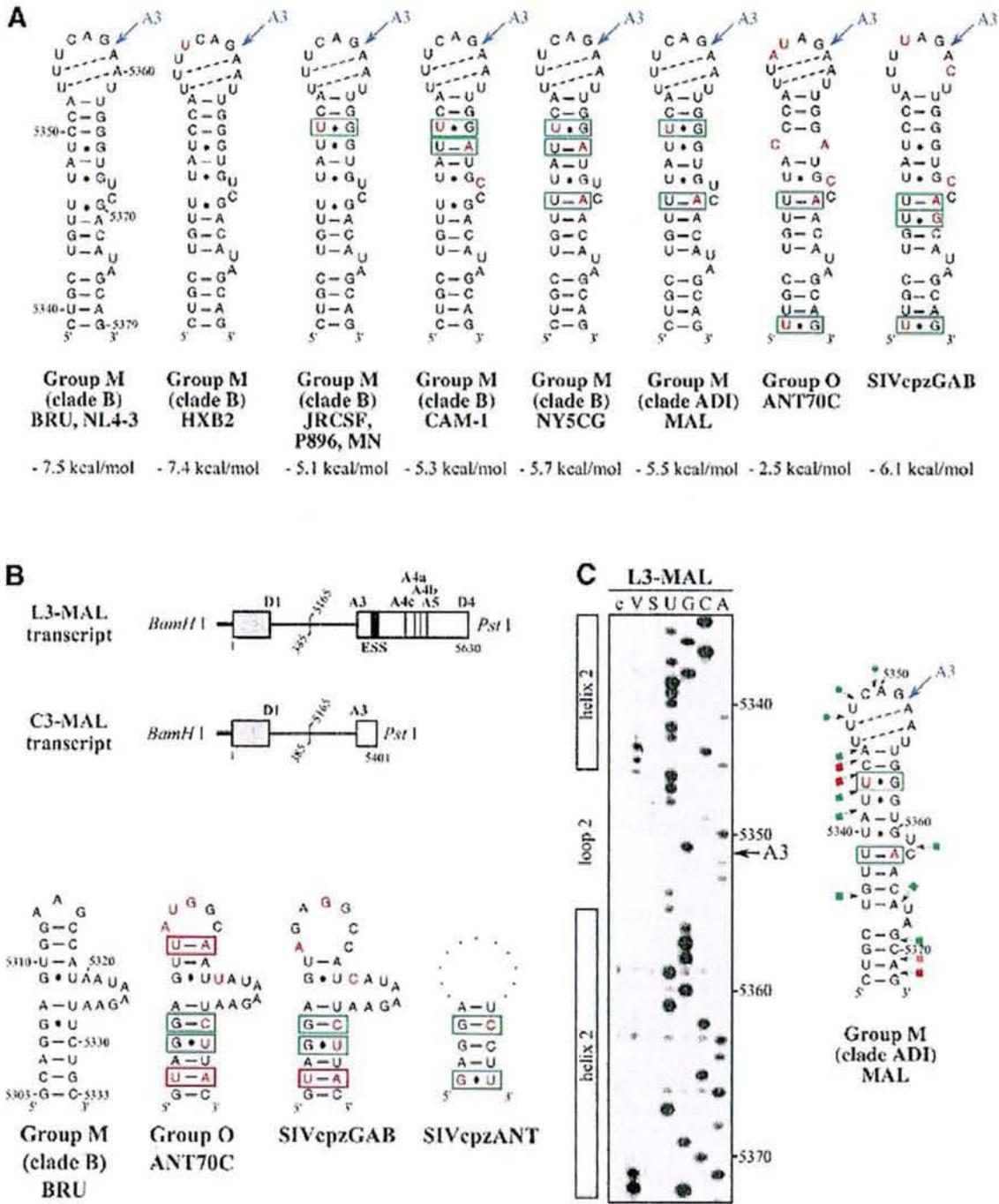


Figure 2. Conservation of SLS2 and SLS1 in the groups M and O of HIV-1 strains and in the SIVcpz strains. Nucleotide sequences in this figure are from the HIV sequence database (<http://hiv-web.lanl.gov/>, Los Alamos National Laboratory, Los Alamos, NM). They were identified by alignment of the HIV-1/BRU RNA sequence with sequences from HIV-1 RNAs of other strains and the SIVcpz strains using the ClustalW program (43). (A) SLS2s are drawn for various strains of the group M of HIV-1 strains, for the ANT70C strain of group O and for the SIVcpzGAB strain. The name of the strains, the clade and the group are indicated below each structure. The free energy of each SLS2, at 37°C, in 1 M NaCl, as calculated by the MFold program is indicated. Sequence variations as compared to BRU are indicated by red nucleotides. Base pairs within green rectangles are conserved by semi-compensatory mutations. (B) Schematic representation of the transcript used for the experimental study of the HIV-1/MAL RNA region containing site A3. The same symbols were used as in Figure 1A. (C) Probing of the secondary structure of the L3-MAL RNA with V1 RNase (lane V) and S1 nuclease (lane S). Primer 1458 was used for reverse transcriptase analysis of cleaved and intact RNA (lane c) and for generating the sequencing ladders (lanes UGCA). Numbering of the HIV-1/MAL RNA sequence on the right of the autoradiogram is according to the Los Alamos HIV sequence database. Beside the autoradiogram, a schematic representation of the cleavages observed is given on the MAL SLS2 model (squares for V1 cleavages; dots for S1 cleavages; red, orange and green for strong, medium and low cleavages, respectively). Helix and loop positions are indicated on the left of the autoradiogram. (D) Based on the experimentally identified SLS1 of the BRU RNA, 'SLS1-like' structures were drawn for the HIV-1/ANT70C, SIVcpzGAB and SIVcpzANT strains. Base pairs in red and green rectangles are conserved by compensatory and semi-compensatory mutations, respectively. The upper part of SLS1, that is not conserved in the SIVcpzANT RNA, is represented by a discontinuous line.

base pair and vice versa) base mutations. The SIVcpzGAB and ANT strains were recently shown to share a common ancestor with virus HIV-1 (31).

As shown in Figure 2A, the possibility to form SLS2 is conserved in clade B and in a recombinant of A, D and I clades (HIV-1/MAL) of group M of HIV-1 strains, as well as in the SIVcpzGAB strain and formation of SLS2 was experimentally verified for the HIV-1/MAL RNA using RNA transcripts L3-MAL and C3-MAL (Fig. 2A, B and C). Although of lower stability, an SLS2 can be formed in the RNA of the ANT70C strain from the group O of HIV-1 strains (Fig. 2A). Such conservation of SLS2 suggested its functional importance in splicing, which was experimentally tested, as described below.

The possibility to form the terminal helix of SLS1 is not as highly conserved in HIV-1 and SIVcpz viruses as SLS2. However, we found that several semi-compensatory and a few compensatory mutations allow formation of the SLS1 bottom helix in the ANT70C strain from group O of HIV-1 strains and in the SIVcpz strains GAB and ANT (Fig. 2D).

SLS3 is highly conserved in strains from clade B of group M of HIV-1 strains (Fig. 3A). Most of the observed base substitutions and the insertion found in HIV-1/SF2, are located in segments which are single-stranded in the HIV-1/BRU SLS3. The few base substitutions in double-stranded regions are semi-conservative: a C-G pair in helix a, found in almost all studied RNAs, is replaced by a U-G pair in the HIV-1/SF2 RNA; a U-G pair in helix a of the HIV-1/BRU RNA is replaced by a U-A pair in all other strains; finally, a U-A pair in helix b, found in almost all studied RNAs, is replaced by a U-G pair in the HIV-1/JRCSF RNA. Only one region of SLS3 showed a variable number of base pairs, depending on the clade B strain studied, and this is the helical part f containing ESS2 (Fig. 3A). Thus, the phylogenetic study of clade B supports the existence of SLS3.

Using enzymatic probes, we studied experimentally the secondary structure of the HIV-1/MAL region corresponding to the HIV-1/BRU SLS3. As shown in Figure 3B and C, a long irregular stem-loop structure was also observed. However, only helix a and the terminal stem-loop structure g are strictly conserved as compared to the HIV-1/BRU SLS3. As mentioned above, the HIV-1/MAL virus is a recombinant of A, D and I clades (26). A structure similar to that established for the MAL RNA can be proposed for the corresponding region of the ELI RNA (Fig. 3C) (The HIV-1/ELI strain belongs to clade D) (26). Interestingly, the terminal part of the BRU SLS3 can be formed in the SIVcpzGAB RNA (Fig. 3D), which favors the functional importance of this terminal part of SLS3. For the remaining part of SLS3, results suggest some variability depending on the clade of HIV-1 strains.

Whereas an active ESS2 was only demonstrated in three HIV-1 strains from clade B of group M, namely NL4-3 (3,9), SF2 (10) and BRU (present data described below), the presence of an ESS2 element has been proposed to be a specific feature of groups M and N of HIV-1 viruses but not of group O viruses (10). In the various HIV-1 strains of clade B that we studied, ESS2 is always located in the irregular helical part f of SLS3, which precedes the conserved terminal region g (Fig. 3A). ESS2 is composed of two repeats of the CUAGA sequence in the HIV-1/BRU RNA. In some HIV-1 strains, only the 3'-terminal C/UUAGA sequence is conserved [this is the case for the SF2 (Fig. 3A), MAL and ELI strains (Fig. 3C)].

Based on the proposed secondary structure, the conserved C/UUAGA motif of ESS2 is at the border of the highly conserved terminal region g.

SLS3 is formed under splicing assay conditions

Because our phylogenetic comparisons were less definitive for SLS3 compared to SLS2 and because SLS3 containing ESS2 was determined using the naked RNA, we tested whether this structure was also formed under conditions for *in vitro* splicing assays. To this end, conditions were developed to probe the RNA secondary structure within a nuclear extract (see Materials and Methods). These conditions were used to analyze the S3 RNA transcript containing all the sequences needed to form SLS3 and part of the SLS2 sequence (Fig. 4A). Three RNases were used in these experiments: (i) V1 RNase to probe accessible double-stranded RNA regions and (ii) T1 and T2 RNases to identify accessible single-stranded regions. Two chemical probes were also used to detect the accessible single-stranded residues, DMS for A and C residues and kethoxal for G residues. To identify the RNA segments protected by the binding of nuclear components, the probing experiments were performed after incubation in a HeLa cell nuclear extract dissolved in buffer D and in buffer D alone.

Examples of primer extension analyses of base modifications and phosphodiester bond cleavages are shown in Figure 4B. Only the probing results showing the larger number of data were selected. Experiments were repeated several times and reproducible results were obtained. For efficient chemical modifications, DMS modifications were performed in the presence of 50 mM sodium cacodylate and kethoxal reactions in the presence of 50 mM potassium borate. However, as shown by the excellent correlation of the results obtained for T1 RNase and kethoxal analysis that both target G residues (for example, see the increased reactivities to both reagents at G residues 5414 and 5415 and the decreased reactivities at G residues 5400 and 5401, Fig. 4B and C), addition of potassium borate in the splicing buffer did not affect the RNA secondary structure and the RNA-protein interactions. Similarly, the very good correlation observed between the T2 RNase and DMS results also revealed that the addition of sodium cacodylate in the splicing buffer did not modify RNA secondary structure and RNA-protein interactions.

Study of the S3 transcript in buffer D, using both enzymatic and chemical probes, confirmed the formation of SLS3 in this buffer. The low level of modification of A and G residues in helical portions d, e and f is the one expected taking into account the expected low thermodynamic stability of these irregular helical portions (Fig. 4C). Using the same probes, SLS3 was also found to be present after incubation of S3 RNA in the nuclear extract (Fig. 4C). In addition, the data suggested that there were two binding sites for nuclear components in SLS3: (i) the structural motif A formed by the segments 5398-5412 and 5480-5499, which includes a large internal loop, the stable helix c and a bulge loop; and (ii) the single-stranded segments of the highly conserved terminal motif B, including the terminal and internal loops (5437-5439 and 5456-5458) (Fig. 4C). A higher degree of protection was observed for motif A as compared to motif B, which suggested a higher affinity of the nuclear components for the former. The accessibility of naked ESS2 to enzymes and chemical reagents is limited in buffer D. However, based on previous studies suggesting the presence of

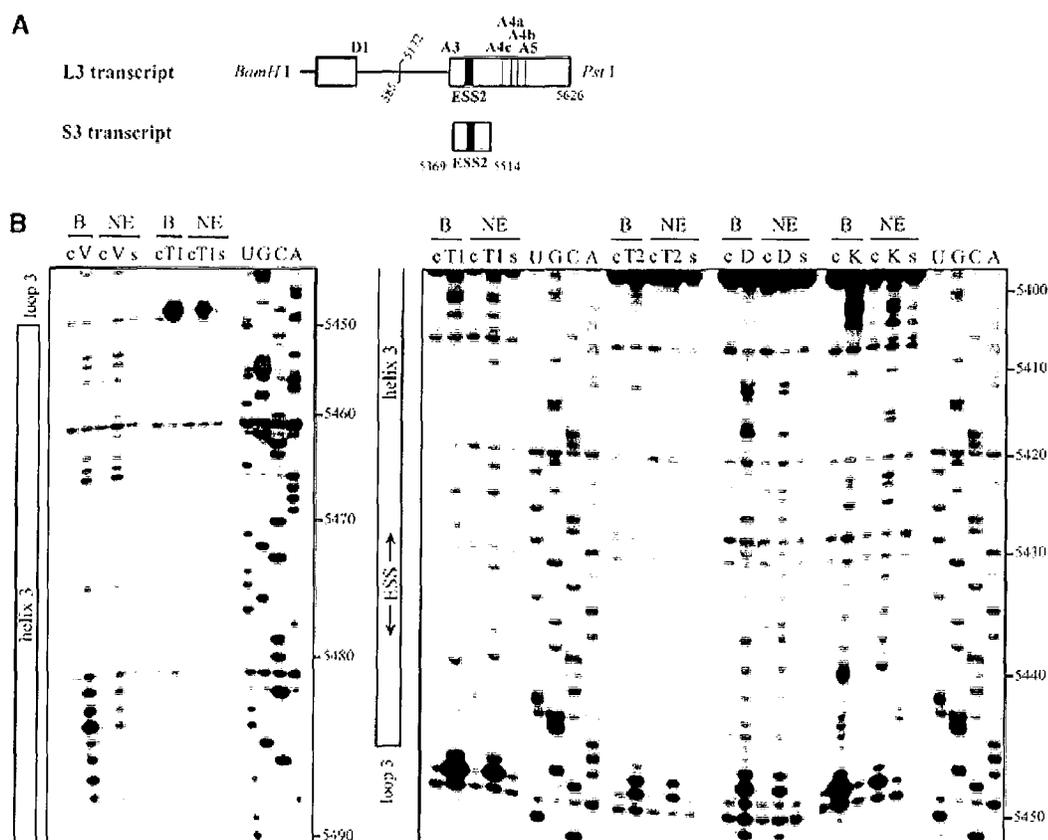


Figure 4. (Above and opposite) The HIV-1/BRU SLS3 is formed in a nuclear extract and two structural motifs A and B are protected by association with nuclear components. (A) The S3 transcript used for the experimental analysis is shown. (B) Examples of primer extension analyses of the S3 transcript cleaved by V1, T1 or T2 nucleases (lanes marked by V, T1, T2, respectively) or modified by kethoxal or DMS (lanes marked by K or D, respectively) are presented. Enzymatic and chemical reactions were either performed in the splicing buffer D (B) or in a nuclear extract (NE). As a control, a primer extension was made with the intact RNA transcript incubated in the absence of reagent, either in buffer D or in the nuclear extract (lanes marked by c). Lanes UGCA correspond to the sequencing ladders. Numbering of the nucleotides in the HIV-1/BRU RNA is indicated on the right. Positions of the loops and helices 3 and of ESS2 are shown on the left. (C) Schematic representation of results of chemical and enzymatic probing of S3, in the splicing buffer D (C1) or in the nuclear extract (C2). In Panel C1, the circled nucleotides were modified by DMS or kethoxal. Cleavages by RNases are shown by arrows, surmounted with a circle for T2 RNase, a star for T1 RNase and a square for V1 RNase. Green, orange and red symbols indicate low, medium and strong modifications or cleavages, respectively. In C2, protection against the action of chemical reagents and nucleases are shown in blue, the intensity of the blue color reflects the level of protection. Increased sensitivity to chemical reagents and nucleases is indicated in yellow (low increase) or orange (medium increase). In both C1 and C2, the oligonucleotide primer 954 used for the reverse transcriptase analysis is indicated. The 3G residues at the 5'-end of the S3 transcript were generated by the T7 RNA polymerase promoter. The portion of the S3 RNA that was analyzed is delimited by the two broken arrows in C1. Due to pause of reverse transcriptase at some of the V1 cleavage sites in the region from positions 5397 to 5438 in the nuclear extract, no estimation of their variation of intensity as compared to naked RNA is given for this part of SLS3. In C3, the HIV-1/BRU functional sequences contained in SLS3 are indicated, namely: the *tat* start codon and *vpr* stop codon (squared in green), the putative SC35 binding site (red letters), the branched sites (circled in blue) and PPT (in blue rectangles) of the A4c 3'ss and the branched sites for A4a and b 3'ss (circled in orange). The limits of the RNA fragments used by Caputi *et al.* (11) and Del Gatto-Konczak *et al.* (32) for hnRNP A/B crosslinking experiments are indicated by black and red arrows, respectively.

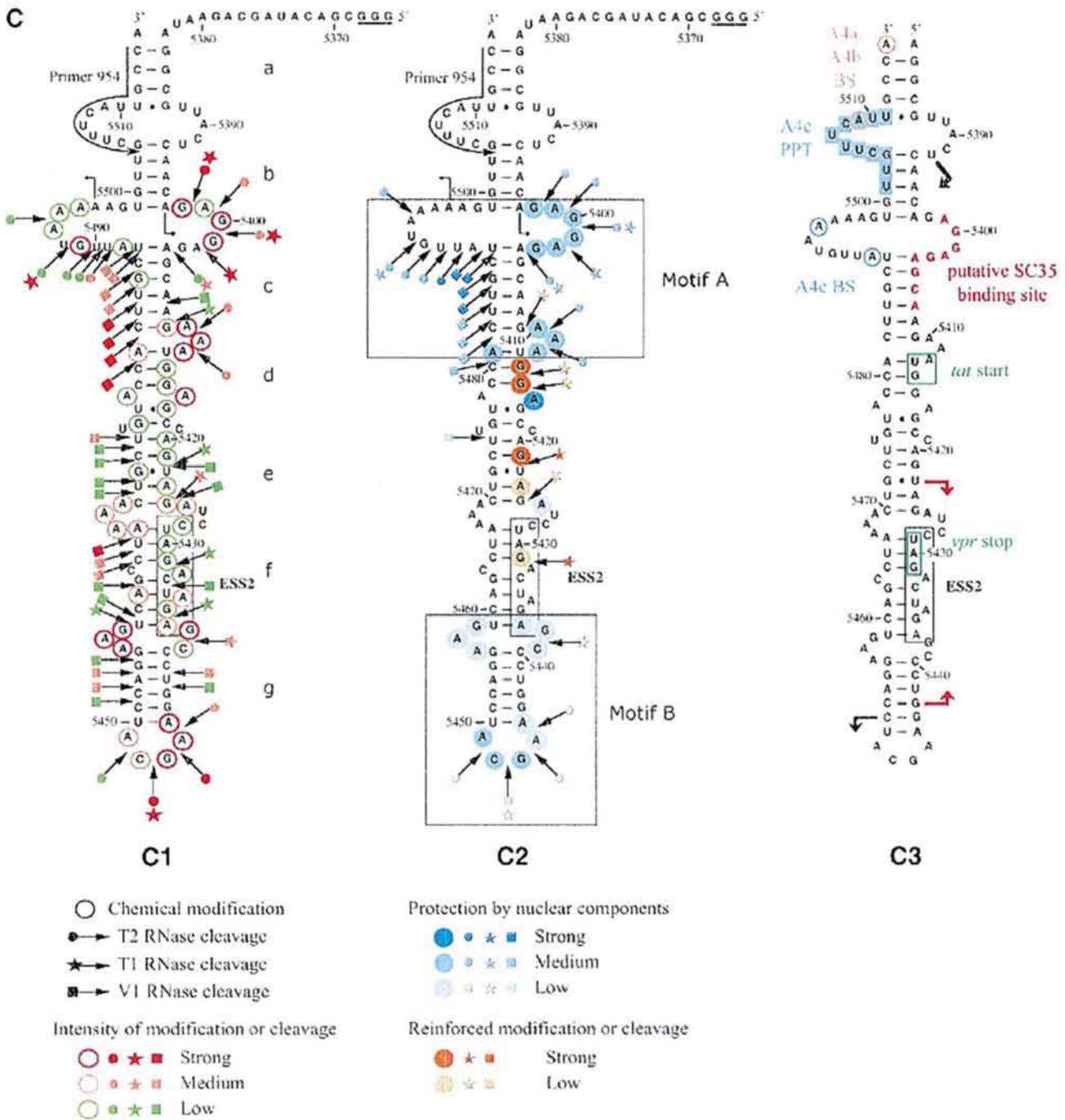
DISCUSSION

There has been little information available previously regarding secondary structure of HIV-1 RNA regions containing splicing sites. Only the secondary structure of the MAL region containing the D1 5'ss has previously been investigated (33). Here, we have described the secondary structure of an HIV-1 RNA region containing the central 3' splice sites: A3, A4a, A4b, A4c and A5 sites. Two main features have emerged from the established structure. Site A3, required for production of the *tat* mRNAs, is located in a terminal loop and its PPT is involved in the helix. ESS2 within *tat* exon 2 and acting on site A3 is located in an irregular helical structure in the vicinity of a highly conserved terminal

motif. These structural features are conserved in HIV-1 strains as well as the related SIVcpz strains.

RNA secondary structure and the PPT sequence likely both contribute to the low A3 site efficiency

Within L3 (900 nt) as well as C3 (684 nt), A3L (510 nt) and A3C (294 nt) transcripts, the HIV-1/BRU RNA region, which is located between positions 5339 and 5379 and contains site A3, always folds into the same stem-loop structure (SLS2). When various strains were compared, we found that all the base substitutions found in SLS2 are either in the single-stranded segments or are such that secondary structure is preserved (semi-compensatory base mutations). Taking into account the strong constraints on evolution exerted by overlapping



and juxtaposed ORFs in HIV-1 RNA, the presence of semi-compensatory mutations in an HIV-1 secondary structure motif reflects an important functional role of this motif. Interestingly, conservation of SLS2 is even observed in the SIVcpzGAB virus from the chimpanzee subspecies *Pan troglodytes troglodytes*, the primary reservoir of HIV-1 (31). One possible reason for the conservation of SLS2 is its contribution to the suboptimal property of the A3 PPT. This may limit the access of the A3 PPT to splicing factors, such as U2AF (34). We found a stronger effect of the RNA secondary

structure at site A3 using *in vitro* splicing assays as compared to transfection assays in HeLa cells. Such a difference between *in vitro* and *ex vivo* assays has already been reported for the regulation of the alternative splicing of the chicken β -tropomyosin pre-mRNA (22).

In vitro, a reinforcement of SLS2 stability had a negative effect on site A3 utilization with a wild-type as well as with an optimized PPT. However, in HeLa cells, using an HIV-1 genomic construct, we showed the dominant effect of the PPT sequence versus the RNA secondary structure. Indeed, reinforcement of

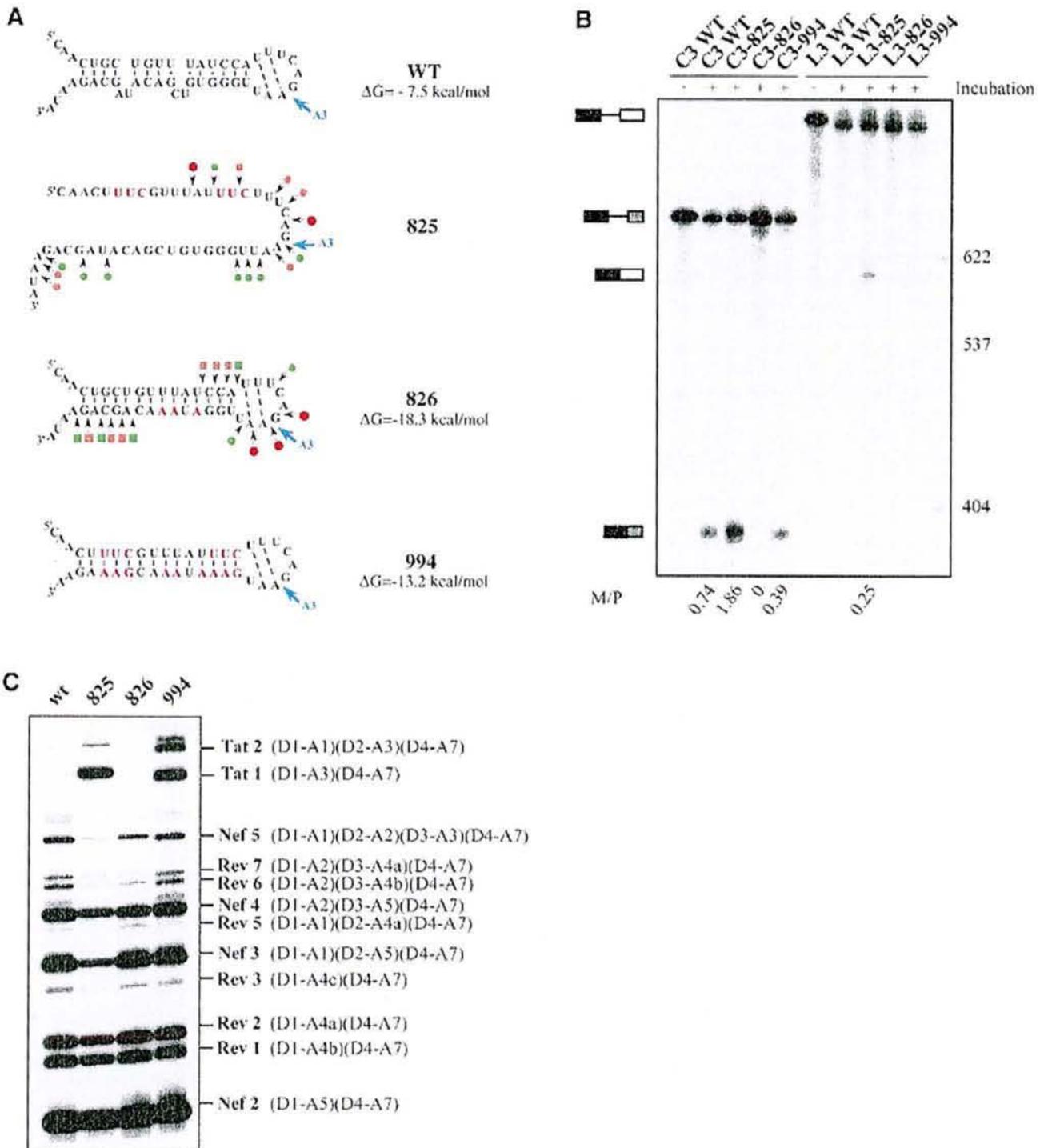


Figure 5. Effects of alteration of SLS2 on site A3 efficiency *in vitro* and *in vivo*. (A) The generated mutations (red nucleotides) and the expected structure of the variant RNAs are shown. For variants C3.825 and C3.826, the secondary structure was probed experimentally, results are schematically represented using the same symbols as in Figure 2C. The free energies of the structures at 37°C in 1 M NaCl, as calculated by the MFold program, are given. (B) PAGE of the *in vitro* splicing products of the wild-type and variant C3 and L3 transcripts. Uniformly labeled RNAs, prepared as described in Materials and Methods, were incubated for 150 min in a HeLa cell nuclear extract in splicing conditions. Untreated C3 and L3 RNAs were fractionated in parallel, as well as size markers (molecular weights on the right of the panel). Positions of the C3 and L3 transcripts and of their spliced products are indicated on the left of the panel. The M:P ratios (given below the lanes) were calculated by estimation with a phosphorimager of the radioactivity in the bands of gel corresponding to the mature RNA and the residual transcript, respectively. (C) PAGE of the RT-PCR products obtained by amplification of total RNA extracted from HeLa cells transfected with the wild-type or the mutated mini-proviral cDNA pΔPSP, with the oligonucleotide primers BSS and SJ4.7A (see Materials and Methods). Nomenclature of the RT-PCR products is according to Purcell and Martin (1).

the stability of the helix (ΔG difference of -5.5 kcal/mol) had no deleterious effect on site A3 utilization when the PPT was optimized (conversion of three Cs and one G into Us and one A and one U into Cs, variant 994). However, site A3 utilization was decreased for variant 826 with a wild-type PPT and a stabilized SLS2 (ΔG difference of -10.8 kcal/mol). In addition to secondary structure reinforcement, disruption of an exonic splicing enhancer (ESE) element in the wild-type sequence may participate in the decreased utilization of site A3 in variant 826. As efficient splicing at site A3 was observed for variant 994, which carried the same three mutations downstream from site A3, as variant 826 together with additional mutations, if an ESE is present downstream from site A3, it is not required in the presence of an optimized PPT.

Our data should also be examined in the light of data from Wu *et al.* (35), Guth *et al.* (36) and Reed (37), who showed that mammalian nuclear introns are AG dependent when their PPT is shorter than 14 nt. The A3 wild-type PPT contains 14 pyrimidine residues but is interrupted by four purine residues. Consequently, splicing at a wild-type A3 site is likely to be AG dependent, while splicing at an A3 site with an optimized PPT is likely to be AG independent. This may, therefore, increase the sensitivity of spliceosome assembly to RNA secondary structure.

Hence, although we cannot discard the possibility of an ESE element downstream from site A3, a reasonable explanation for our *in vivo* and *in vitro* observations is that the secondary structure and the suboptimal sequence of the PPT both participate to the low efficiency of site A3. This may be the reason for the high conservation of SLS2 in HIV-1 and SIVcpz strains.

ESS2 element is located in an irregular flexible structure that binds several nuclear components

Because the secondary structure at site A3 was identical in transcript C3 lacking ESS2 and in transcript L3 containing ESS2, the data clearly demonstrate that ESS2 does not act by a direct RNA-RNA interaction with a functional sequence of site A3. Instead, as previously proposed, it is more likely that the inhibition occurs by binding of inhibitory component(s) acting *in trans* (3,11,32).

In the HIV-1/BRU RNA, ESS2 is located in a long irregular stem-loop structure SLS3, that is formed independently of the length of the studied fragment, the buffer conditions (pH, mono and divalent cation concentrations) and in the presence or the absence of nuclear proteins. Only the terminal loop and helix of SLS3 are strictly conserved in HIV-1 strains. However, a common property of this part of HIV-1 RNAs is the presence of a flexible irregular structure, interrupted by several internal loops and bulges and by a lateral helix in the MAL and ELI RNAs. During the RNA folding process, the conserved stable terminal element may be the nucleation center of this irregular structure.

Based on its location in a region of reduced base-pairing (f), the BRU ESS2 would be expected to be available for RNA-protein interaction. However, its protection in a nuclear extract was limited to the 3' terminal A residue of the downstream CUAGA motif and an increased reactivity of the upstream CUAGA motif to kethoxal and T1 RNase was observed. Interestingly, this upstream CUAGA motif of ESS2 is the less conserved of the two CUAGA motifs. For instance, in the SF2 strain where an active ESS2 was demonstrated, only the downstream CUAGA motif is present (Fig. 3A). ESS2 is adjacent to

the highly conserved terminal motif B, which showed a partial protection in a nuclear extract. It should be noted that the protected A residue of ESS2 and the adjacent protected GC dinucleotide belong to a UAGAGC sequence that is similar to the reported protein hnRNP A1 binding site consensus sequence [UAGGG(A/U)] (38). Therefore, it is possible that binding of hnRNP A1 to HIV-1 RNA segments containing ESS2, detected by cross-linking experiments (11,32) (Fig. 4C), occurs at the UAGAGC sequence overlapping the last CUAGA repeat of ESS2 and the internal loop. In support of this hypothesis, a similar sequence (UAGAGU), which regulates 5' splice site utilization in the hnRNP A1 pre-mRNA, was found to be a binding site for protein hnRNP A1 (39), and binding of a UAGAG sequence to hnRNP A1 RRM is compatible with recent crystallographic data (40). The low protection observed for the core part of ESS2 may be due to the fact that stable binding of the repressor factor(s) to ESS2 needs association with splicing components bound at site A3 (the S3 transcript, that we used for the assays, did not contain the A3 site, Fig. 4A). The presence of numerous internal loops in SLS3 may ensure a greater flexibility allowing free or bound ESS2 to interact with distant components, for instance components bound at the A3 site. Conservation of the irregular architecture of SLS3 may be required for ESS2 mechanism of action.

In contrast to ESS2, the SLS3 motif B was highly protected in a nuclear extract. Interestingly, one of the highly protected sequences in motif A contains the two adenosine residues, used as branch residues for site A4c (Fig. 4C) (41). Protection may be due to the binding of splicing components involved in branch site recognition. Another highly protected single-stranded sequence in motif A (positions 5398-5407) shows homology with one of the consensus sequence established for the binding sites of the SR protein SC35 (Fig. 4C) [HIV-1 sequence 5'-AGGAGAGCA-3', SC35 binding site consensus sequence 5'-AG(C/G)AGAGUA-3'; 42]. It should be noted that, in the proposed secondary structure, the A4c PPT is largely single stranded.

In addition to its importance for the splicing efficiency at site A3, which determines the yield of *tat* mRNA production, the SLS3 architecture may also be important for the regulation of *tat* mRNA translation. Indeed, SLS3 is likely to be present in the spliced *tat* mRNAs. In SLS3 from strains of clade B, the *tat* start codon is located in a short helical portion (d) of low stability (Fig. 4C). However, it is preceded by the stable helix portion (c), which may limit scanning of the *tat* mRNA 5' UTR by the translation initiation complex.

In summary, it is likely that the RNA secondary structure of the HIV-1 RNA region that we determined play important roles in the regulation of RNA splicing and of mRNA translation, as well as increasing HIV-1 RNA resistance to cellular nucleases.

ACKNOWLEDGEMENTS

F. Clavel and J. L. Darlix are thanked for providing plasmid pBRU3 and MAL cDNA, respectively. S.J., D.R. and L.D. were fellows of the French Ministère de la Recherche et de l'Enseignement Supérieur and L.D. was also a fellow of the SIDACTION foundation. This work was supported by the Agence Nationale de Recherche sur le SIDA (ANRS), the Centre National de la Recherche Scientifique (CNRS), the French Ministère de la Recherche et de l'Enseignement

Supérieur and by PHS grant AI36073 to C.M.S. from the National Institute of Allergy and Infectious Disease.

REFERENCES

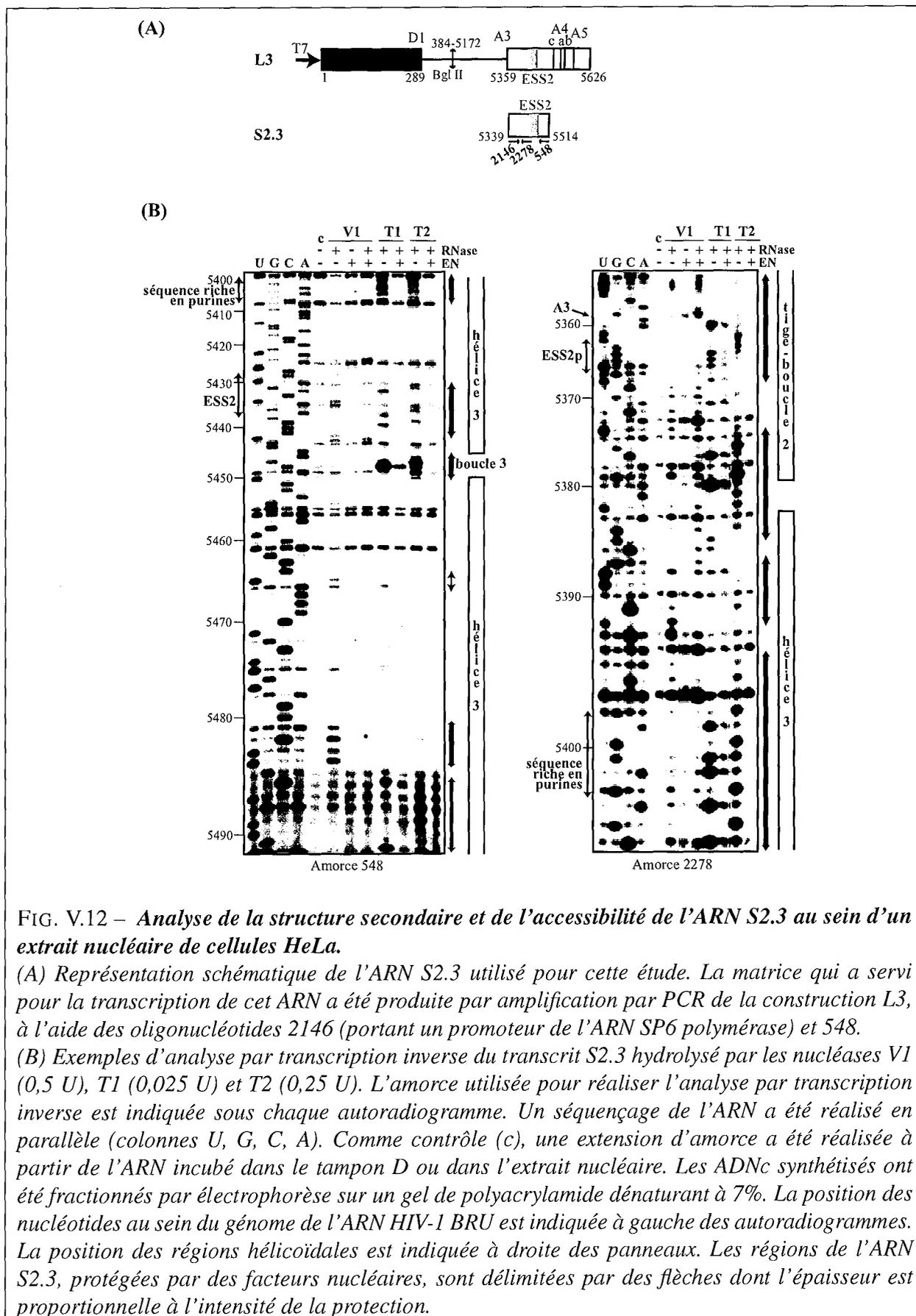
- Purcell, D.F. and Martin, M.A. (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication and infectivity. *J. Virol.*, **67**, 6365–6378.
- Amendt, B.A., Hesselin, D., Chang, L.J. and Stoltzfus, C.M. (1994) Presence of negative and positive *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**, 6480.
- O'Reilly, M.M., McNally, M.T. and Beemon, K.L. (1995) Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology*, **213**, 373–385.
- Staffa, A. and Cochrane, A. (1994) The *tat-rev* intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. *J. Virol.*, **68**, 3071–3079.
- Staffa, A. and Cochrane, A. (1995) Identification of positive and negative splicing regulatory elements within the terminal *tat-rev* exon of human immunodeficiency virus type 1. *Mol. Cell. Biol.*, **15**, 4597–4605.
- Damier, L., Domenjoud, L. and Branlant, C. (1997) The D1-A2 and D2-A2 pairs of splice sites from human immunodeficiency virus type 1 are highly efficient *in vitro*, in spite of an unusual branch site. *Biochem. Biophys. Res. Commun.*, **237**, 182–187.
- Frankel, A.D. and Young, J.A. (1998) HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.*, **67**, 1–25.
- 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.
- Bilodeau, P.S., Domsic, J.K. and Stoltzfus, C.M. (1999) Splicing regulatory elements within *tat* exon 2 of human immunodeficiency virus type 1 (HIV-1) are characteristic of group M but not group O HIV-1 strains. *J. Virol.*, **73**, 9764–9772.
- 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.
- Chebli, K., Gattoni, R., Schmitt, P., Hildwein, G. and Stevenin, J. (1989) The 216-nucleotide intron of the E1A pre-mRNA contains a hairpin structure that permits utilization of unusually distant branch acceptors. *Mol. Cell. Biol.*, **9**, 4852–4861.
- Deshler, J.O. and Rossi, J.J. (1991) Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. *Genes Dev.*, **5**, 1252–1263.
- Charpentier, B. and Rosbash, M. (1996) Intramolecular structure in yeast introns aids the early steps of *in vitro* spliceosome assembly. *RNA*, **2**, 509–522.
- Mougin, A., Grégoire, A., Banroques, J., Ségault, V., Fournier, R., Brulé, F., Chevrier-Müller, M. and Branlant, C. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA*, **2**, 1079–1093.
- Muro, A.F., Caputi, M., Pariyath, R., Pagani, F., Buratti, E. and Baralle, F.E. (1999) Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol. Cell. Biol.*, **19**, 2657–2671.
- 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.
- Shi, H., Hoffman, B.E. and Lis, J.T. (1997) A specific RNA hairpin loop structure binds the RNA recognition motifs of the *Drosophila* SR protein B52. *Mol. Cell. Biol.*, **17**, 2649–2657.
- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999) Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature*, **398**, 579–585.
- Watakabe, A., Inoue, K., Sakamoto, H. and Shimura, Y. (1989) A secondary structure at the 3' splice site affects the *in vitro* splicing reaction of mouse immunoglobulin m chain pre-mRNAs. *Nucleic Acids Res.*, **17**, 8159–8169.
- Clouet d'Orval, B., d'Aubenton-Carafa, Y., Brody, J.M. and Brody, E. (1991) Determination of an RNA structure involved in splicing inhibition of a muscle-specific exon. *J. Mol. Biol.*, **221**, 837–856.
- Libri, D., Piseri, A. and Fizsman, M.Y. (1991) Tissue-specific splicing *in vivo* of the beta-tropomyosin gene: dependence on an RNA secondary structure. *Science*, **252**, 1842–1845.
- Charnau, 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.
- Nour, M., Naimi, A., Beck, G. and Branlant, C. (1995) 16S-23S and 23S-5S intergenic spacer regions of *Streptococcus thermophilus* and *Streptococcus salivarius*, primary and secondary structure. *Curr. Microbiol.*, **31**, 270–278.
- Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., et al. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature*, **313**, 277–283.
- Alizon, M., Wain-Hobson, S., Montagnier, L. and Sonigo, P. (1986) Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell*, **46**, 63–74.
- Kramer, W., Drutsa, V., Jansen, H.W., Kramer, B., Pflugfelder, M. and Fritz, H.J. (1984) The gapped duplex DNA approach to oligonucleotide-directed mutation construction. *Nucleic Acids Res.*, **12**, 9441–9456.
- Vassilenko, S.K. and Bakkina, G.T. (1965) Isolation and properties of the cobra venom nuclease. *Biokhimiya*, **30**, 705–712.
- Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
- Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) Improved predictions of secondary structures for RNA. *Proc. Natl Acad. Sci. USA*, **86**, 7706–7710.
- 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. et al. (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, **397**, 436–441.
- 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.
- Baudin, F., Marquet, R., Isel, C., Darlix, J.L., Ehresmann, B. and Ehresmann, C. (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.*, **229**, 382–397.
- 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.
- 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.
- Guth, S., Martinez, C., Gaur, R.K. and Valcarcel, J. (1999) Evidence for substrate-specific requirement of the splicing factor U2AF35 and for its function after polypyrimidine tract recognition by U2AF65. *Mol. Cell. Biol.*, **19**, 8263–8271.
- Reed, R. (1989) The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.*, **3**, 2113–2123.
- 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.
- Chabot, B., Blanchette, M., Lapierre, I. and La Branche, H. (1997) An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol. Cell. Biol.*, **17**, 1776–1786.
- Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R. and Xu, R.M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.*, **13**, 1102–1115.
- Swanson, A.K. and Stoltzfus, C.M. (1998) Overlapping *cis* sites used for splicing of HIV-1 *env/rev* and *rev* mRNAs. *J. Biol. Chem.*, **273**, 34551–34557.
- 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.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

V.4.2.3 Résultats complémentaires

L'ARN avec lequel nous avons réalisé nos expériences d'empreintes décrites dans l'article contenait uniquement la tige-boucle 2, renfermant l'élément ESS2. Afin de mesurer l'importance du site A3 dans l'interaction des composants de l'extrait avec la tige-boucle 3, nous avons répété ces expériences d'empreintes en extrait nucléaire, avec un ARN plus long (ARN S2.3), contenant, en plus de l'élément ESS2, le site A3 et l'élément ESS2p que S. Jacquenet et A. Méreau venaiënt alors de mettre en évidence (Fig. V.12-A). En effet, en plus des mutants de la tige-boucle 3 décrits dans l'article, S. Jacquenet avait fait d'autres mutants dans le segment situé immédiatement en aval du site A3. Cette fois, les différences d'efficacité d'épissage observées pour ces variants ne pouvaient pas s'expliquer simplement par un effet de la structure secondaire. En recherchant ce qui pouvait expliquer les variations observées à la fois *in vitro* et *ex vivo* dans des cellules HeLa, S. Jacquenet et A. Méreau ont mis en évidence un second élément inhibiteur agissant sur le site A3 (ESS2p) et fixant la protéine hnRNP H (Jacquenet *et al.* 2001a). L'encombrement stérique généré par la fixation de cette protéine doit alors empêcher la fixation de U2AF³⁵ à la bordure intron-exon et bloque ainsi l'assemblage du spliceosome.

Alors que nous avons utilisé les mêmes conditions de formation de complexes que pour l'ARN S3, nous avons obtenu un niveau de protection de l'ARN S2.3 beaucoup plus élevé (Fig. V.12 et V.13). La seule région qui restait légèrement sensible à l'action de la RNase T1 était la région charnière entre les 2 structures tiges-boucles, au niveau du résidu G en position 5379. Ces données suggèrent que la présence de la tige-boucle 2 renfermant le site A3, sa séquence polypyrimidine et l'élément ESS2p, augmente la capacité des composants de l'extrait nucléaire à se lier à la structure tige-boucle 3. Pour affiner l'analyse du comportement de l'ARN S2.3 en extrait nucléaire, il serait intéressant de répéter les expériences en utilisant des quantités croissantes d'extrait nucléaire, ceci en choisissant, pour concentration maximale, celle que j'ai utilisée et pour laquelle la protection est totale. Il serait ainsi possible de définir les segments de l'ARN qui sont saturés en premier par les composants de l'extrait nucléaire.

Nous n'avons pas encore réalisé ces expériences, car une région que nous avons trouvée protégée dans l'ARN S3 et qui l'était également dans l'ARN S2.3, a retenu notre attention et nous a conduit à réaliser un lourd travail expérimental. Il s'agissait du segment en simple-brin riche en purines, GAGGAG, situé entre les positions 5398 et 5403 et constituant la boucle latérale du brin 5' de la structure tige-boucle SLS3. Son niveau très important de protection dans l'ARN S3, sa séquence qui suggérait qu'il pouvait s'agir d'un site de fixation de protéines SR, et sa proximité du site A3, nous ont conduit à nous demander s'il ne s'agissait pas d'un élément activateur exonique agissant sur le site A3.



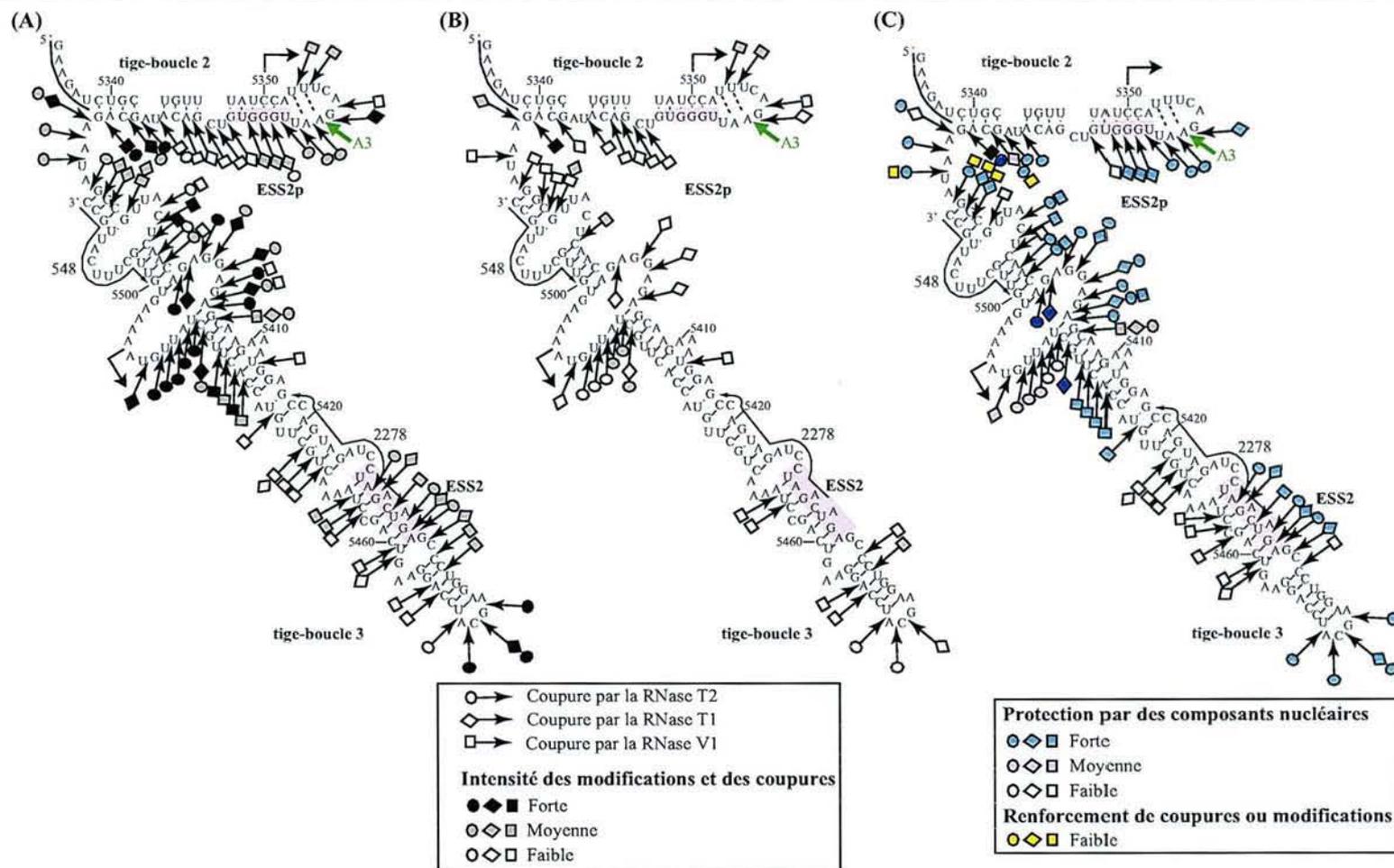


FIG. V.13 – *Protections de l'ARN S2.3 par les composants nucléaires.*

(A) et (B) Représentation schématique des coupures enzymatiques de l'ARN S2.3 dans le tampon D (A) et dans l'extrait nucléaire (B). La région de l'ARN qui a été analysée est délimitée par des flèches noires cassées. La position des séquences complémentaires aux oligonucléotides 548 et 2278 utilisés pour l'analyse par transcription inverse est indiquée sur les modèles de structure. Les nucléotides soulignés à l'extrémité 5' du transcrit sont générés par le promoteur de l'ARN polymérase SP6. Les symboles et code couleur utilisés sont indiqués sous les panneaux dans un encart.

(C) Représentation schématique de la protection, par les composants nucléaires, des nucléotides de l'ARN S2.3 contre l'action des RNases. Le code couleur utilisé pour représenter l'intensité des protections est indiqué dans un encart sous le panneau.

V.4.3 La séquence riche en purines, située entre les positions 5398 et 5403, est un élément activateur d'épissage

V.4.3.1 L'activation du site A3 par la protéine ASF/SF2 nécessite la présence de la séquence 5398 à 5403 riche en purines

Sachant que la séquence riche en purines, comprise entre les nts 5398 et 5403, est protégée en extrait nucléaire et constitue un site potentiel de fixation des protéines ASF/SF2 et SC35, nous avons testé l'effet de sa délétion dans des expériences d'épissage *in vitro*. Dans ce but, nous avons produit par mutagenèse dirigée de la construction C3, l'ARN C3.ΔPu ne contenant plus les nucléotides GAG-GAG, situés entre les positions 5398 et 5403.

Nous avons alors comparé l'efficacité d'épissage *in vitro* des ARN C3 et C3.ΔPu en extrait nucléaire, sans ajout de protéine SR recombinante ou avec 300 ng de protéine SC35 ou ASF/SF2.

Après 2h30 d'incubation, on voit d'après l'autoradiogramme montré dans la figure V.14 que, comme précédemment, l'utilisation du site A3 dans l'ARN C3 est activée par ASF/SF2 et encore plus par SC35. La délétion de la séquence riche en purines a un effet marqué sur l'utilisation du site A3, à tel point que, lorsqu'aucune protéine SR n'est rajoutée, ou lorsqu'ASF/SF2 est rajoutée, aucun produit d'épissage n'est détecté. Nous en avons donc conclu que la séquence 5398 à 5403 riche en purines est vraisemblablement un élément nécessaire à l'activation du site A3 par ASF/SF2. Nous l'avons appelé ESEt¹.

Il était intéressant de noter que SC35, que nous avons trouvée être la protéine SR ayant le plus fort pouvoir activateur sur le site A3, continuait à activer ce site après délétion de l'élément ESEt, mais le taux d'efficacité d'épissage était nettement inférieur à celui obtenu pour l'ARN sauvage (M/P de 2,3 contre 16,5). Elle devait donc agir au moins en partie sur ce site.

V.4.3.2 L'élément ESEt a des propriétés activatrices d'épissage dans un contexte hétérologue

Afin de prouver définitivement que l'élément ESEt est bien un élément activateur d'épissage, nous l'avons déplacé dans un contexte hétérologue. Dans ce but, nous avons inséré sa séquence dans la construction génétique Sp1 ex2-inv, dérivée de l'adénovirus, que nous a fournie J. Stévenin (Fig. V.15-A; Cavaloc *et al.* 1999). Cette construction sert de construction rapporteur lors de la recherche d'une séquence activatrice de l'épissage. Nous en rappellerons brièvement le principe : l'ARN sauvage Sp1 est épissé de façon très efficace dans un extrait nucléaire, comme en témoigne l'autoradiogramme présenté dans la figure V.15-B, piste 2. Lorsque le 2^e exon est inversé, l'épissage de l'ARN devient inefficace (ARN Sp1-ex2-inv; Fig. V.15-B, piste 4). Il est alors nécessaire d'insérer une séquence activatrice dans cet exon pour restaurer cet épissage.

Ainsi, l'insertion de la séquence ASF A10, identifiée par l'équipe de J. Stévenin comme une séquence activatrice d'épissage fixant la protéine ASF/SF2 (Cavaloc *et al.* 1999), restaure l'épissage de cet ARN (piste 10), alors que l'introduction d'une séquence quelconque (appelée gst 80) n'a aucun effet (piste 8).

Lorsque nous avons inséré dans la construction Sp1 ex2-inv la séquence 5398 à 5403 de l'ARN de

¹ « exon splicing enhancer of *tat* mRNA »

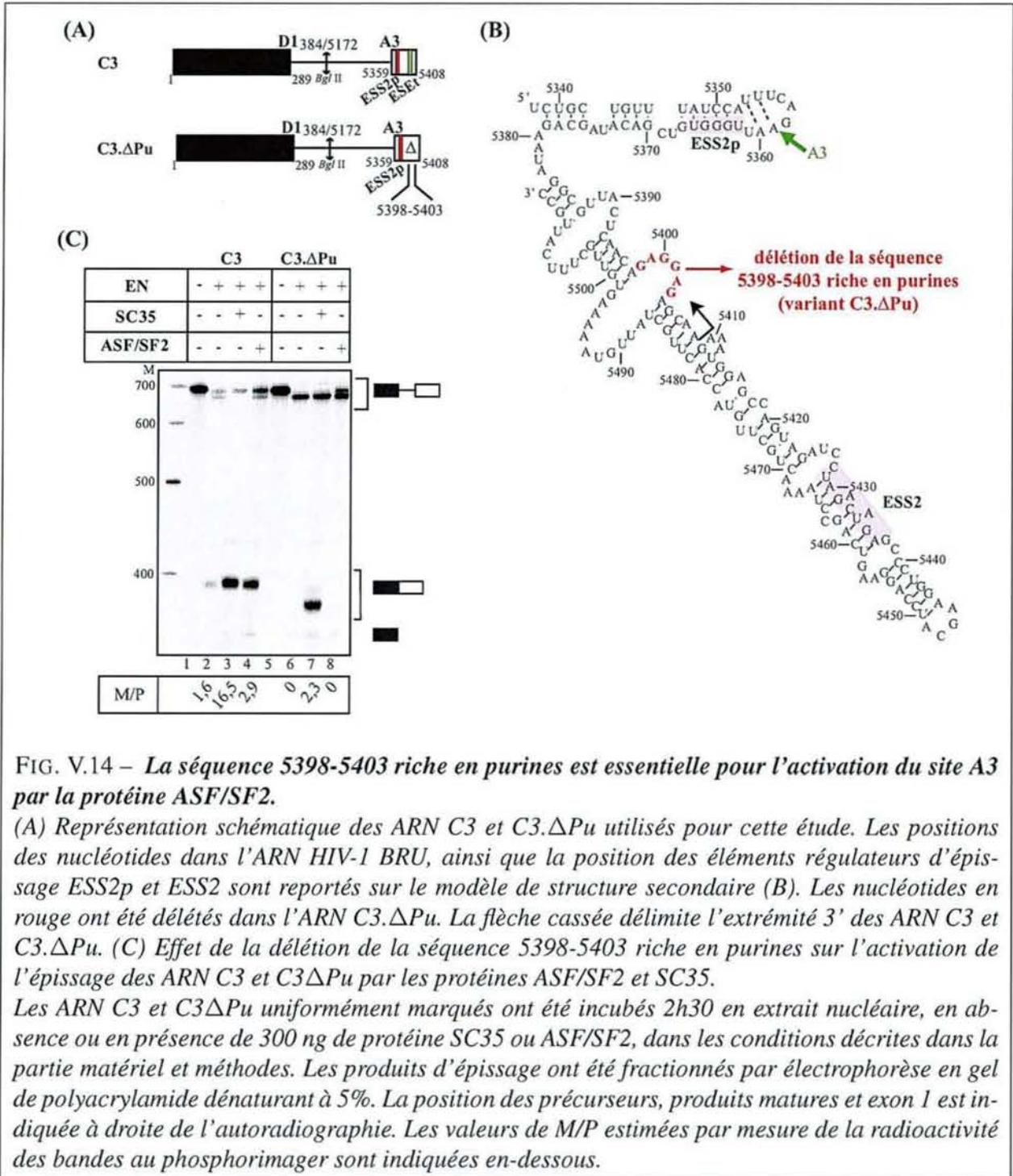


FIG. V.14 – La séquence 5398-5403 riche en purines est essentielle pour l'activation du site A3 par la protéine ASF/SF2.

(A) Représentation schématique des ARN C3 et C3.ΔPu utilisés pour cette étude. Les positions des nucléotides dans l'ARN HIV-1 BRU, ainsi que la position des éléments régulateurs d'épissage ESS2p et ESS2 sont reportés sur le modèle de structure secondaire (B). Les nucléotides en rouge ont été délétés dans l'ARN C3.ΔPu. La flèche cassée délimite l'extrémité 3' des ARN C3 et C3.ΔPu. (C) Effet de la délétion de la séquence 5398-5403 riche en purines sur l'activation de l'épissage des ARN C3 et C3.ΔPu par les protéines ASF/SF2 et SC35.

Les ARN C3 et C3.ΔPu uniformément marqués ont été incubés 2h30 en extrait nucléaire, en absence ou en présence de 300 ng de protéine SC35 ou ASF/SF2, dans les conditions décrites dans la partie matériel et méthodes. Les produits d'épissage ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 5%. La position des précurseurs, produits matures et exon 1 est indiquée à droite de l'autoradiographie. Les valeurs de M/P estimées par mesure de la radioactivité des bandes au phosphorimager sont indiquées en-dessous.

HIV-1 BRU renfermant l'élément ESEt, l'épissage était restauré (piste 6). L'élément ESEt est donc bien un élément activateur d'épissage, capable de fonctionner dans un contexte hétérologue.

Afin de mettre en évidence le rôle de la protéine ASF/SF2 dans cette activation, nous avons reproduit la même expérience, en utilisant, cette fois-ci, de l'extrait cytoplasmique S100 en présence d'une petite quantité d'extrait nucléaire (0,5 µl), complémenté ou non par 150 ng de protéine ASF/SF2 (Fig. V.15-C). Cette protéine n'a aucun effet sur l'efficacité d'épissage de l'ARN Sp1 sauvage (Fig. V.15-C, pistes 1 et 2). Elle ne permet pas l'épissage des ARN Sp1 ex2-inv (pistes 3 et 4) et Sp1 gst80 (pistes 7

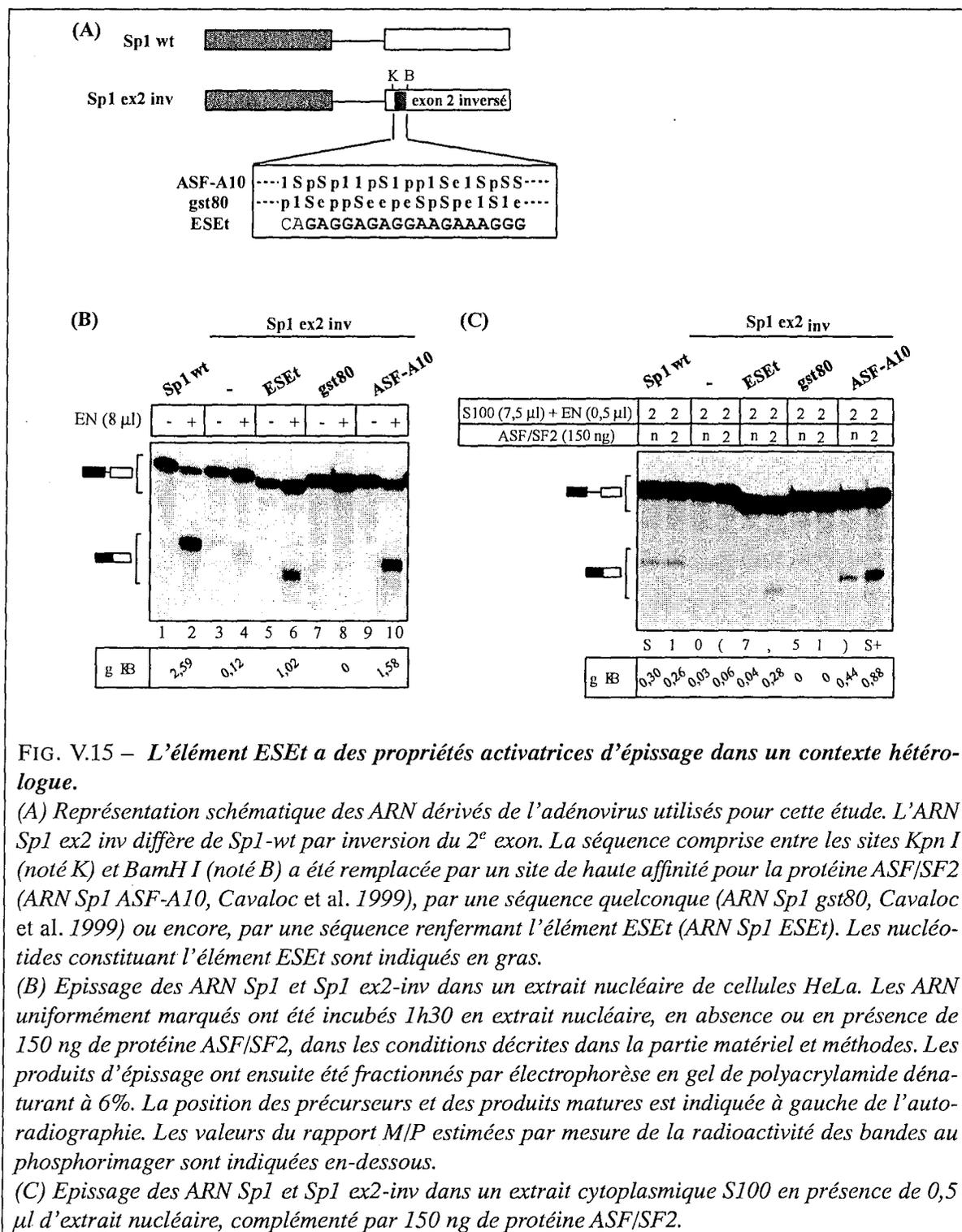


FIG. V.15 – L'élément ESEt a des propriétés activatrices d'épissage dans un contexte hétérologue.

(A) Représentation schématique des ARN dérivés de l'adénovirus utilisés pour cette étude. L'ARN Sp1 ex2 inv diffère de Sp1-wt par inversion du 2^e exon. La séquence comprise entre les sites Kpn I (noté K) et BamHI (noté B) a été remplacée par un site de haute affinité pour la protéine ASF/SF2 (ARN Sp1 ASF-A10, Cavaloc et al. 1999), par une séquence quelconque (ARN Sp1 gst80, Cavaloc et al. 1999) ou encore, par une séquence renfermant l'élément ESEt (ARN Sp1 ESEt). Les nucléotides constituant l'élément ESEt sont indiqués en gras.

(B) Epissage des ARN Sp1 et Sp1 ex2-inv dans un extrait nucléaire de cellules HeLa. Les ARN uniformément marqués ont été incubés 1h30 en extrait nucléaire, en absence ou en présence de 150 ng de protéine ASF/SF2, dans les conditions décrites dans la partie matériel et méthodes. Les produits d'épissage ont ensuite été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 6%. La position des précurseurs et des produits matures est indiquée à gauche de l'autoradiographie. Les valeurs du rapport M/P estimées par mesure de la radioactivité des bandes au phosphorimager sont indiquées en-dessous.

(C) Epissage des ARN Sp1 et Sp1 ex2-inv dans un extrait cytoplasmique S100 en présence de 0,5 μl d'extrait nucléaire, complémenté par 150 ng de protéine ASF/SF2.

et 8), alors qu'elle active fortement l'épissage de l'ARN Sp1 ASF A10 qui peut aussi avoir lieu à un faible taux, sans ajout de protéine (pistes 9 et 10). La protéine ASF/SF2 permet l'épissage de l'ARN Sp1 ESEt qui, lui, n'est pas épissé dans le mélange S100/EN 7,5 μl/0,5 μl (pistes 5 et 6). Ces données concourraient à penser que l'élément ESEt était responsable de l'activation de l'épissage au site A3 par ASF/SF2, puisque dans un contexte hétérologue, il activait bien l'épissage et ceci, de manière ASF/SF2 dépendante.

Lorsque la même expérience a été réalisée en présence de la protéine SC35, nous n'avons pas pu conclure, car la protéine active l'épissage de tous les ARN Sp1 dans l'extrait cytoplasmique S100, y compris les ARN Sp1 ex2-inv et Sp1 gst80, ceci vraisemblablement du fait de la présence d'un site activateur dans les ARN Sp1 qui répond à la protéine SC35.

V.4.4 La protéine ASF/SF2 se fixe à l'élément ESEt

V.4.4.1 L'élément ESEt est important pour la fixation d'ASF/SF2 sur l'ARN

Le fait que l'élément ESEt est nécessaire pour l'activation du site A3 par ASF/SF2 suggérait que cette protéine agissait en se fixant à cette séquence. Dans un 1^{er} temps, j'ai testé, par expérience de retard sur gel, la capacité de la protéine à former un complexe avec un petit fragment de l'exon A3-D4 (positions 5369 à 5408) renfermant l'élément ESEt (Fig. V.16-A). Cet ARN a été produit par transcription d'un produit d'amplification par PCR de la construction L3-U1. En parallèle, un test a été réalisé avec un ARN correspondant à la même région de l'exon A3-D4 dans laquelle l'élément ESEt avait été délété. Cet ARN a été produit à partir d'un produit d'amplification de la construction L3.ΔESEt-U1 (Fig. V.16-A). Les expériences de retard sur gel ont été réalisées avec deux quantités différentes de protéine ASF/SF2 recombinante, 150 et 300 ng, pendant 5 minutes à 0°C puis 15 minutes à 30°C en tampon D. Après fractionnement par électrophorèse en gel non dénaturant, l'ARN libre donne lieu à deux bandes, sans doute du fait de l'existence de deux conformations distinctes ou d'une dimérisation. Après ajout de la protéine, la bande de plus faible mobilité électrophorétique est fortement retardée dans le cas de l'ARN sauvage, mais pas dans le cas de l'ARN muté. Nous en avons déduit qu'ASF/SF2 pouvait bien se lier à l'élément ESEt.

V.4.4.2 Identification du site d'ancrage de la protéine ASF/SF2 sur l'ARN S2.3

Afin d'identifier l'ensemble des sites de fixation de la protéine ASF/SF2 sur le fragment d'ARN formé par les structures tiges-boucles 2 et 3 (ARN S2.3), j'ai alors réalisé des expériences d'empreintes avec des RNases et agents chimiques sur l'ARN libre et l'ARN complexé avec la protéine ASF/SF2. Ceci permettait de vérifier les données d'expérience de retard sur gel et de savoir si la protéine ASF/SF2 se fixait à d'autres séquences sur l'ARN S2.3.

Les complexes ont été formés après renaturation du transcrit en utilisant 200 ng d'ARN et 500 ng de protéine en tampon D pendant 5 minutes à 0°C puis 10 minutes à 30°C. Les hydrolyses enzymatiques (RNases V1, T1 et T2) et les modifications chimiques (DMS et kéthoxal) ont alors été réalisées dans les conditions déjà décrites pour l'ARN A2-sl5 (cf § V.3.2.2, page 191).

Aucune protection n'est observée dans la tige-boucle 2, par contre deux régions de la tige-boucle 3 sont protégées (Fig. V.17 et V.18). L'une l'est très faiblement, il s'agit du segment situé en aval de l'élément ESS2. L'autre, protégée beaucoup plus fortement, correspond à la boucle interne contenant l'élément ESEt et une partie de l'hélice 3 qui jouxte cette boucle. C'est clairement au niveau de l'élément ESEt que la protection est la plus forte (absence totale de coupure par les RNases T1 et T2). Ceci confirme bien que cette séquence polypurine fixe ASF/SF2. La protection du brin opposé de la boucle

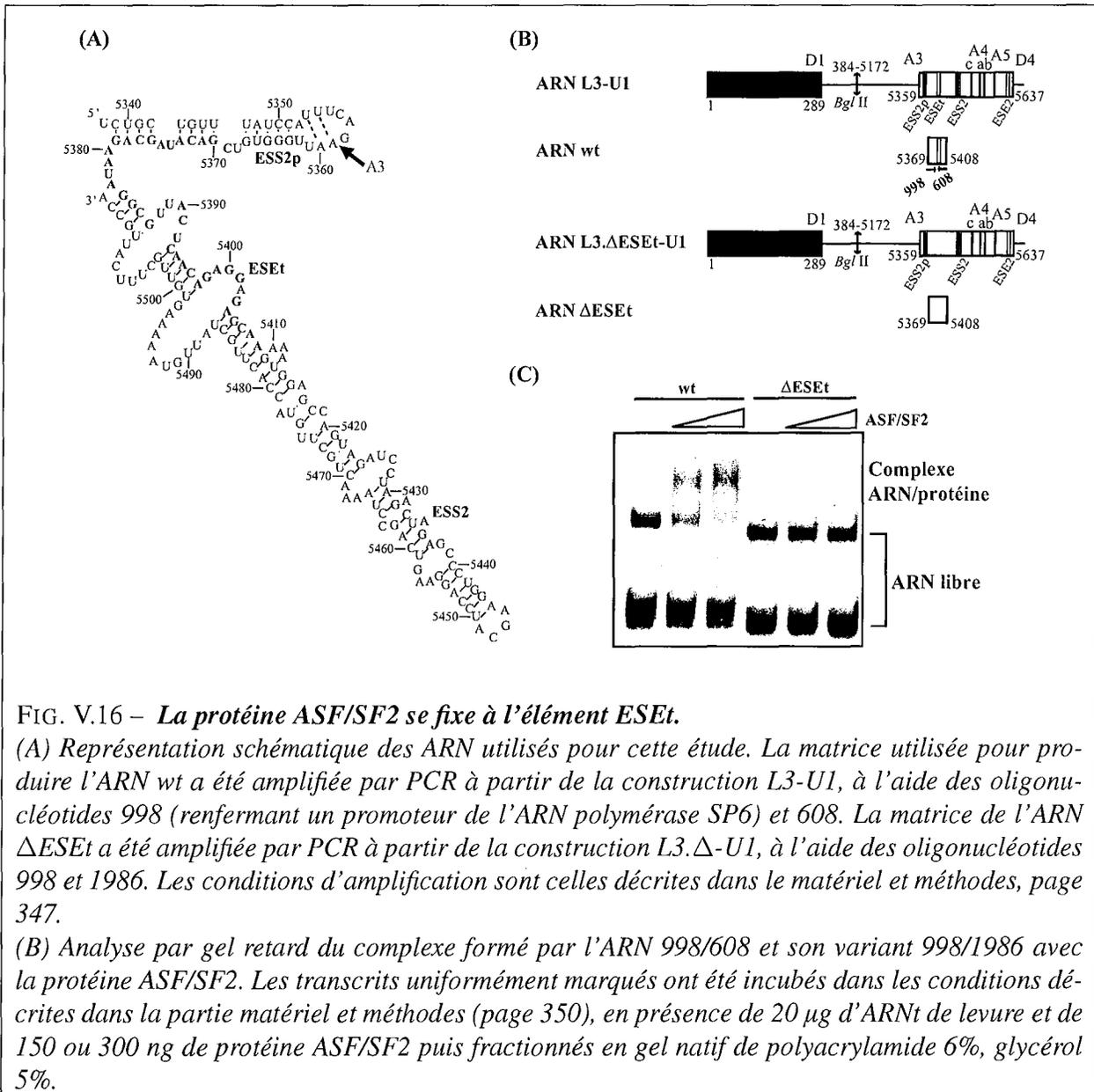


FIG. V.16 – La protéine ASF/SF2 se fixe à l'élément ESEt.

(A) Représentation schématique des ARN utilisés pour cette étude. La matrice utilisée pour produire l'ARN wt a été amplifiée par PCR à partir de la construction L3-U1, à l'aide des oligonucléotides 998 (renfermant un promoteur de l'ARN polymérase SP6) et 608. La matrice de l'ARN ΔESEt a été amplifiée par PCR à partir de la construction L3.Δ-U1, à l'aide des oligonucléotides 998 et 1986. Les conditions d'amplification sont celles décrites dans le matériel et méthodes, page 347.

(B) Analyse par gel retard du complexe formé par l'ARN 998/608 et son variant 998/1986 avec la protéine ASF/SF2. Les transcrits uniformément marqués ont été incubés dans les conditions décrites dans la partie matériel et méthodes (page 350), en présence de 20 μg d'ARNt de levure et de 150 ou 300 ng de protéine ASF/SF2 puis fractionnés en gel natif de polyacrylamide 6%, glycérol 5%.

interne et d'une partie de l'hélice 3 pourrait être liée à un phénomène d'encombrement stérique, car il s'agit uniquement de protections contre l'action des RNases.

Il est intéressant de noter que la protection observée dans le complexe avec la protéine ASF/SF2 recombinante pour cette région de la tige-boucle 3 est identique à celle que nous avons observée au sein de l'extrait nucléaire (Fig. 4 de l'article n° 2 et Fig. V.13 226). Ceci suggère que la protection observée dans l'extrait était essentiellement due à la fixation d'ASF/SF2. Ces données suggèrent, en plus, l'existence d'un site de plus faible affinité, situé en aval de l'élément ESS2. Etant donné que la délétion de l'élément ESEt abolit toute activation par ASF/SF2, ce second site faible ne semble pas intervenir dans l'activation de l'utilisation du site A3.

Les données que j'ai obtenues par les expériences de retard sur gel et d'empreintes ont été complétées par L. Ayadi, par des expériences de compétitions et de pontages aux UV. L'ensemble des données confirme la fixation forte d'ASF/SF2 sur l'élément ESEt.

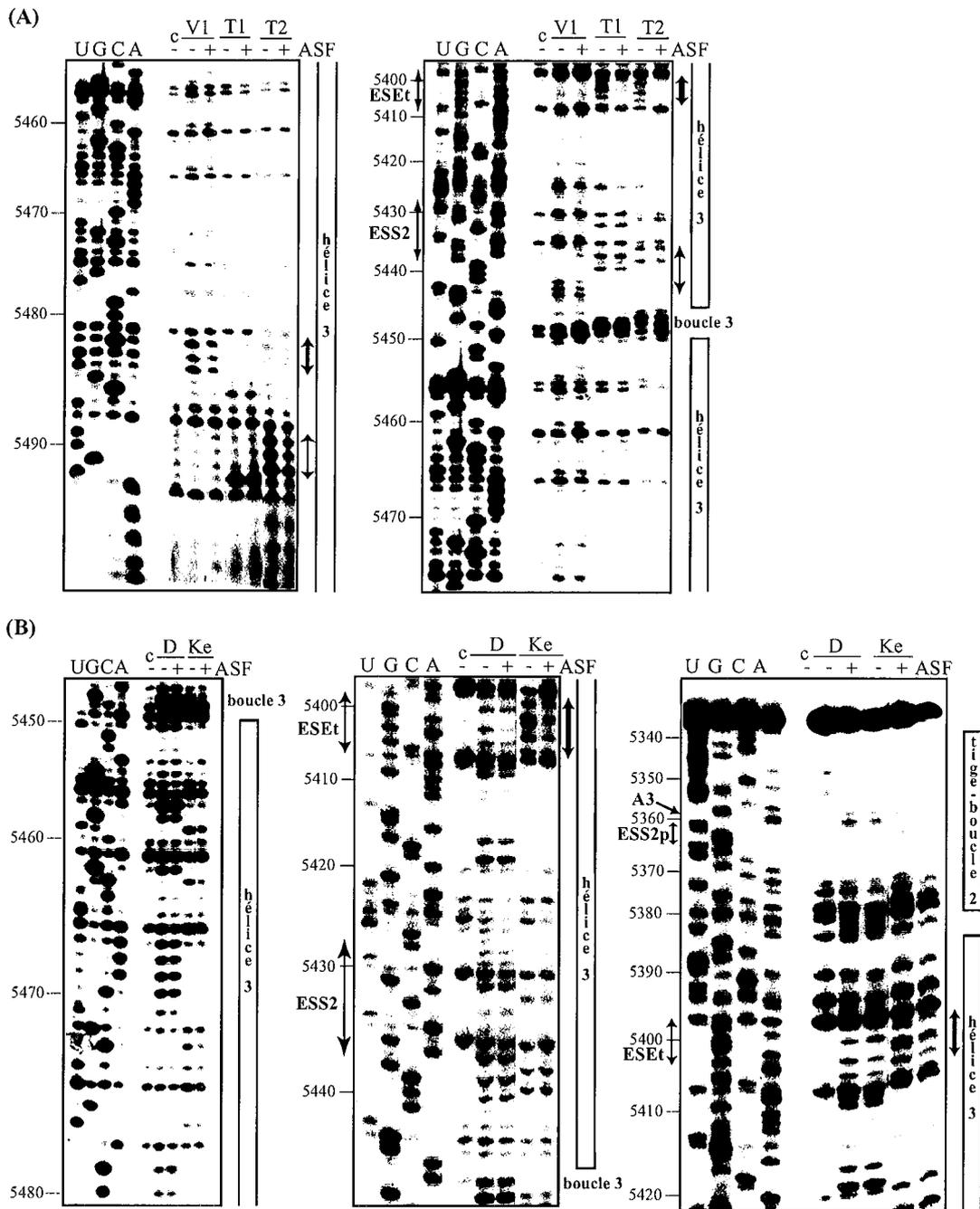


FIG. V.17 – *Etude de l'accessibilité aux RNases et aux agents chimiques de l'ARN S2.3 libre ou complexé à la protéine ASF/SF2.*

(A) et (B) Exemples d'analyse, par extension de l'amorce 548, du transcrit S2.3 hydrolysé par les RNases VI (0,8 U), T1 (0,15 U) et T2 (0,8 U) (A) ou modifié par les sondes chimiques DMS (1 μ l à 12,5%) et kéthoxal (1 μ l à 9,2 μ g/ μ l) (B), en absence ou en présence de 500 ng de protéine ASF/SF2 recombinante. Un séquençage de l'ARN a été réalisé en parallèle (pistes U, G, C, A). Comme contrôle (c), une extension d'amorce a été réalisée à partir de l'ARN incubé en tampon D, en absence de réactif. Les ADNc synthétisés ont été fractionnés par électrophorèse sur gel de polyacrylamide dénaturant à 7% (A) ou à 10% (B). La position des nucléotides au sein du génome de l'ARN HIV-1 BRU est indiquée à gauche des autoradiogrammes. La position des régions hélicoïdales est indiquée à droite des panneaux. Les zones de l'ARN, protégées par la protéine, sont délimitées par des flèches dont l'épaisseur est proportionnelle à la force de la protection.

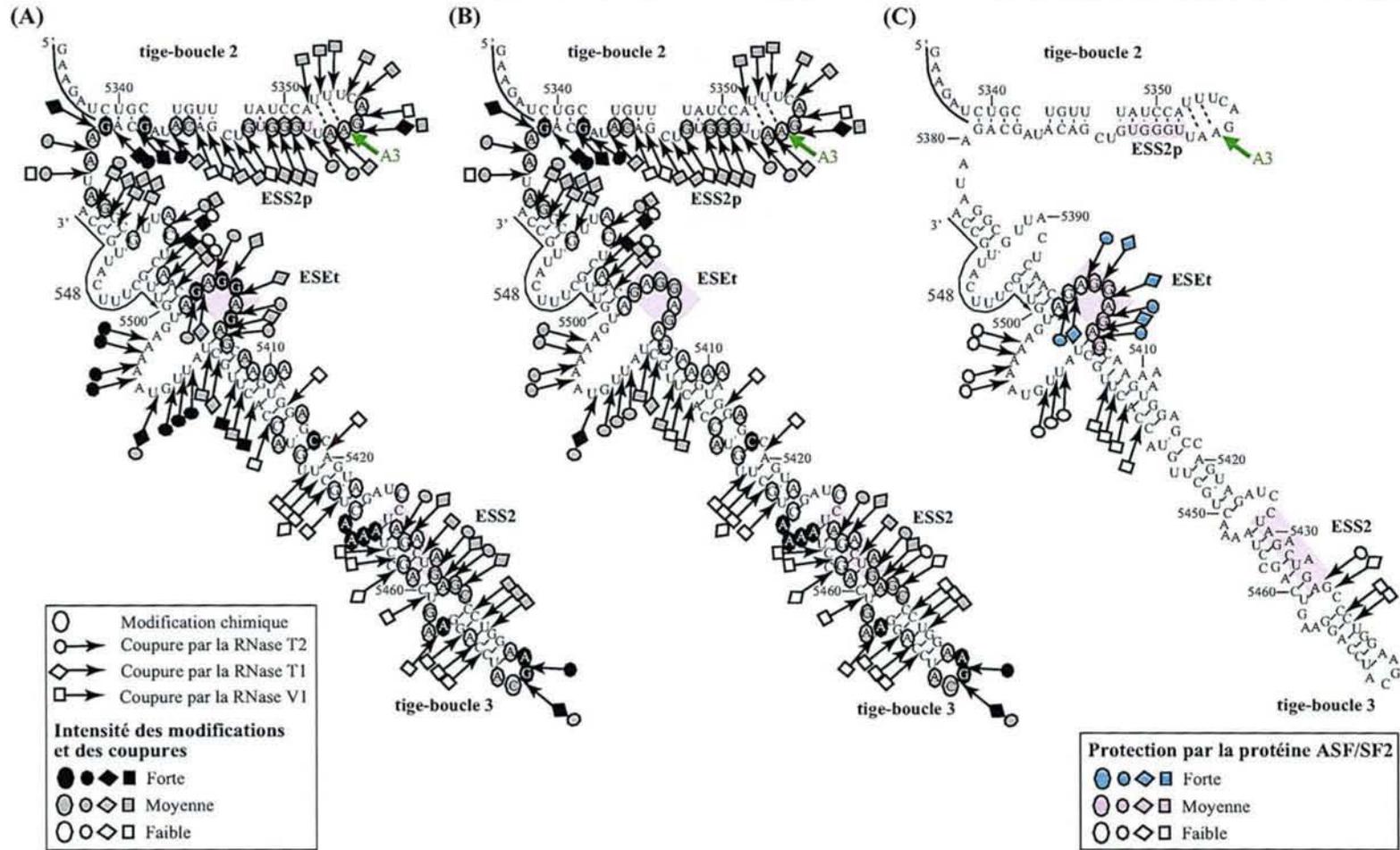


FIG. V.18 – *Protections de l'ARN S2.3 contre l'action des RNases et agents chimiques par la protéine ASF/SF2.*
 (A) et (B) Représentation schématique des coupures enzymatiques et modifications chimiques de l'ARN S2.3 observées en absence (A) ou en présence (B) de 500 ng protéine ASF/SF2 recombinante. La position de la séquence complémentaire à l'oligonucléotide 548, utilisé pour l'analyse par transcription inverse, est indiquée sur le modèle de structure. Les symboles et code couleur utilisés sont indiqués sous les panneaux par un encart.
 (C) Représentation schématique des protections de l'ARN S2.3 par la protéine ASF/SF2, contre l'action des RNases et agents chimiques. Le code couleur utilisé pour représenter l'intensité des protections par la protéine ASF/SF2 est indiqué sous le panneau.

V.4.5 La protéine SC35 se fixe aux éléments ESEt et ESS2

Etant donné que la protéine SC35 active fortement l'utilisation du site A3, mais que la délétion de l'élément ESEt n'abolit pas complètement son effet activateur, j'ai ensuite réalisé de nouvelles expériences d'empreintes en présence de la protéine SC35, de façon à essayer d'identifier son ou ses sites de fixation.

Les complexes ont été formés après renaturation du transcrit en utilisant 200 ng d'ARN et 250 ng de protéine en tampon D pendant 5 minutes à 0°C puis 10 minutes à 30°C. Les hydrolyses enzymatiques (RNases V1, T1 et T2) et les modifications chimiques (kéthoxal) ont alors été réalisées dans les conditions déjà décrites précédemment (cf § V.3.2.2, page 191).

Alors qu'aucune protection n'est observée dans la tige-boucle 2, deux régions de la tige-boucle 3 sont protégées (Fig. V.17 et V.18). La 1^{re} correspond à l'élément ESEt. Dans ce cas, la protection par SC35 semble plus faible que celle obtenue en présence de la protéine ASF/SF2. Comme dans le cas de la protection par ASF/SF2, le brin opposé à l'élément ESEt et une partie de l'hélice 3 sont également protégés, sans doute du fait d'un encombrement stérique. En effet, bien que cette région soit protégée contre l'action des RNases, elle reste accessible aux agents chimiques. La 2^e région, fortement protégée par SC35, correspond à l'élément ESS2 et à la boucle terminale de la tige-boucle 3. Les nucléotides du brin opposé à l'élément ESS2 sont faiblement protégés, sans doute du fait, là-aussi, d'un encombrement stérique.

La protection de cette région au sein du complexe formé avec SC35 est identique à celle que nous avons observée au sein de l'extrait nucléaire (Fig. V.13, page 226). La protection au sein de l'extrait était donc sans doute due à la fixation de la protéine SC35. De même, la protection de l'élément ESEt que nous avons observée dans l'extrait nucléaire doit être due à la fixation des deux protéines ASF/SF2 et SC35. Malgré la protection de l'élément ESEt par SC35, la délétion de cet élément n'abolit pas complètement l'activation du site A3 par cette protéine, dans les expériences d'épissage utilisant la construction courte C3 qui ne contient pas l'élément ESS2 (Fig. V.14). Il existerait peut-être encore d'autres sites de fixation de la protéine SC35 dans la région du site A3, qu'il reste à identifier. La fixation de la protéine SC35 sur les éléments ESS2 et ESEt expliquerait le fort effet activateur de la protéine sur le site A3, en particulier, dans le contexte de l'ARN L3-U1 qui renferme ces 2 sites (cf Fig. 2 et 3 de l'article n° 1).

Les données que j'ai obtenues par les expériences d'empreintes ont été complétées par L. Ayadi, par des expériences de retard sur gel et de compétitions, ainsi que par des expériences de pontages aux UV. L'ensemble des données confirme la fixation de la protéine SC35 sur les éléments ESEt et ESS2.

L'étude par empreintes de la région du site A3 a mis en évidence deux sites de fixation des protéines SR : l'un, chevauchant l'élément ESS2, correspond à un site fort de fixation de la protéine SC35 et à un site faible pour ASF/SF2. L'autre site correspond à l'élément ESEt et fixe fortement ASF/SF2 et SC35. L'étude des propriétés activatrices de cet élément ESEt dans des expériences d'épissage *in vitro* a montré que cet élément est responsable de l'activation du site A3 par ASF/SF2. De plus, la

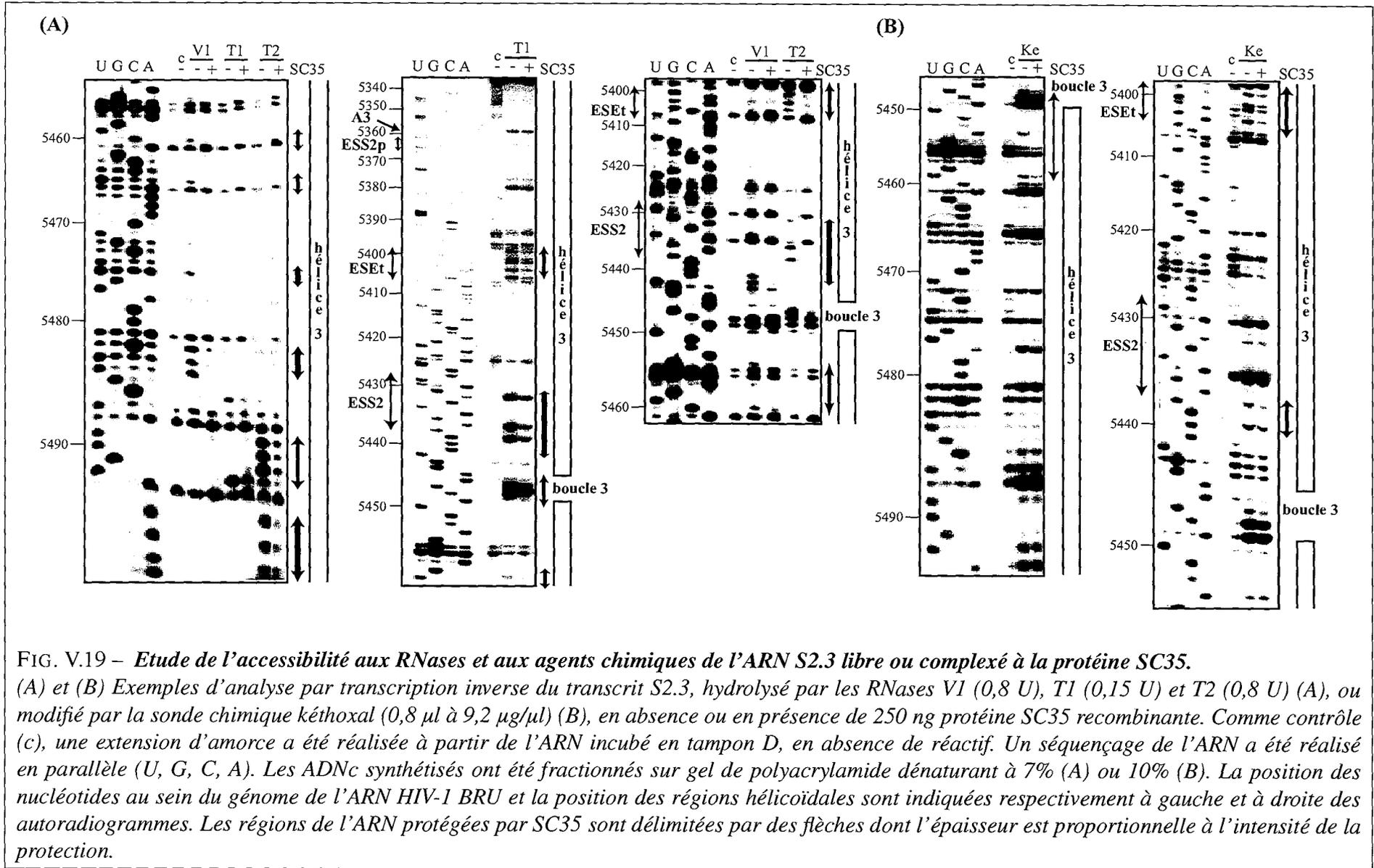


FIG. V.19 – Etude de l'accessibilité aux RNases et aux agents chimiques de l'ARN S2.3 libre ou complexé à la protéine SC35.

(A) et (B) Exemples d'analyse par transcription inverse du transcrit S2.3, hydrolysé par les RNases VI (0,8 U), T1 (0,15 U) et T2 (0,8 U) (A), ou modifié par la sonde chimique kéthoxal (0,8 μ l à 9,2 μ g/ μ l) (B), en absence ou en présence de 250 ng protéine SC35 recombinante. Comme contrôle (c), une extension d'amorce a été réalisée à partir de l'ARN incubé en tampon D, en absence de réactif. Un séquençage de l'ARN a été réalisé en parallèle (U, G, C, A). Les ADNc synthétisés ont été fractionnés sur gel de polyacrylamide dénaturant à 7% (A) ou 10% (B). La position des nucléotides au sein du génome de l'ARN HIV-1 BRU et la position des régions hélicoïdales sont indiquées respectivement à gauche et à droite des autoradiogrammes. Les régions de l'ARN protégées par SC35 sont délimitées par des flèches dont l'épaisseur est proportionnelle à l'intensité de la protection.

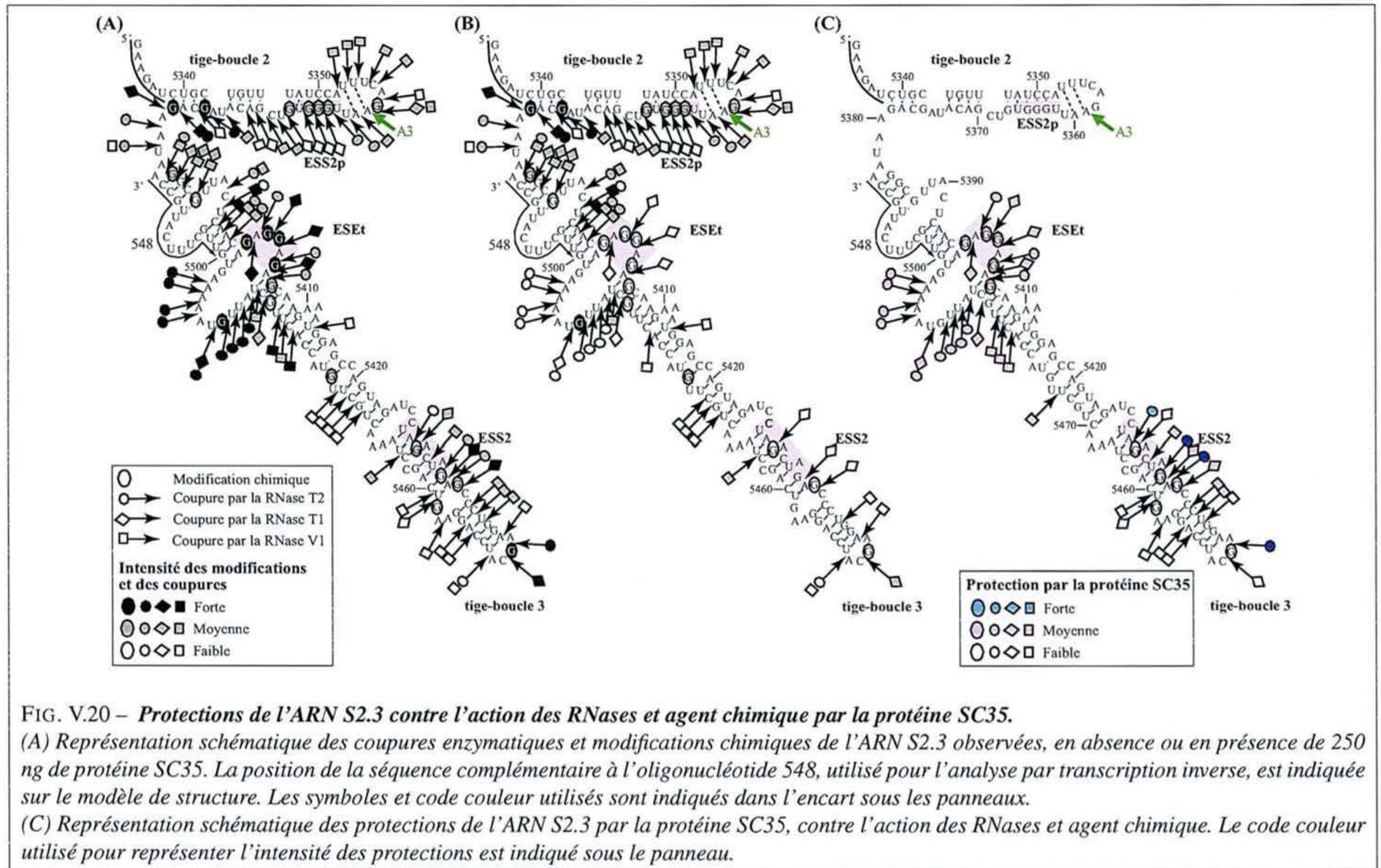


FIG. V.20 – Protections de l'ARN S2.3 contre l'action des RNases et agent chimique par la protéine SC35.

(A) Représentation schématique des coupures enzymatiques et modifications chimiques de l'ARN S2.3 observées, en absence ou en présence de 250 ng de protéine SC35. La position de la séquence complémentaire à l'oligonucléotide 548, utilisé pour l'analyse par transcription inverse, est indiquée sur le modèle de structure. Les symboles et code couleur utilisés sont indiqués dans l'encart sous les panneaux.

(C) Représentation schématique des protections de l'ARN S2.3 par la protéine SC35, contre l'action des RNases et agent chimique. Le code couleur utilisé pour représenter l'intensité des protections est indiqué sous le panneau.

protéine SC35 activerait aussi le site A3 en agissant à la fois sur l'élément ESEt et l'élément ESS2.

L'élément ESEt constitue la 3^e séquence régulatrice du site A3, identifiée à ce jour. Est-ce que la fixation des protéines SR à cette séquence contrecarre l'effet inhibiteur des éléments ESS2p et ESS2 ? De même, quel est le rôle de la fixation de la protéine SC35 sur l'élément ESS2 ? Est-ce d'empêcher la fixation de la protéine hnRNP A1 ? Afin d'apporter des réponses à ces questions, j'ai réalisé, avec L. Ayadi, un ensemble d'expériences d'épissage *in vitro* visant à comprendre quelles sont les inter-relations entre les séquences régulatrices ESS2p, ESS2 et ESEt.

V.4.6 Etude des inter-relations entre les éléments régulateurs agissant sur le site A3

Afin d'étudier les rôles respectifs et les relations entre les séquences régulatrices agissant sur le site A3, nous avons évalué l'efficacité d'épissage *in vitro* de différents variants de l'ARN L3-U1. Les expériences ont été réalisées dans un extrait nucléaire de cellules HeLa, avec ou sans ajout de trois protéines SR recombinantes : 9G8, ASF/SF2 ou SC35. Un mélange de protéines SR totales que nous a fourni J. Stévenin a aussi été utilisé. Il a été préparé à partir d'extraits nucléaires par deux précipitations successives au sulfate d'ammonium, selon le protocole de Zahler *et al.* (1992).

V.4.6.1 La présence d'un excès de protéines SR peut compenser l'absence de site D4

Un 1^{er} point intéressant à tester était de savoir l'importance relative du site D4 et des éléments activateurs et inhibiteurs du site A3. En effet, comme nous l'avons mentionné auparavant, la présence du site de fixation de la snRNP U1 en aval du site D4 a été décrite comme nécessaire à l'utilisation du site A3, ainsi qu'à l'utilisation des sites A4c, a, b et A5, dans une construction renfermant l'élément ESS2 (Amendt *et al.* 1994). Dans ces expériences, les auteurs n'avaient cependant jamais testé l'effet des protéines SR. On pouvait donc se demander si l'ajout de ces protéines pouvait pallier l'absence de site D4 et si tel était le cas, quel allait être le ou les sites les plus utilisés.

Nous rappellerons tout d'abord les résultats précédemment décrits. D'une part, le site A3 peut être utilisé en absence de site donneur aval lorsque l'élément ESS2 est absent (Fig. 1D de l'article n° 1). Dans cet ARN, le site A3 est fortement activé par SC35, mais aussi par ASF/SF2. Dans l'ARN L3-U1 qui renferme les 5 sites accepteurs et le site donneur D4, le site A5 est le plus utilisé dans l'extrait nucléaire, sans ajout de protéines SR. Lorsque les protéines ASF/SF2 et surtout, SRp40 et SC35, sont ajoutées en excès, le site A3 est majoritairement utilisé.

Qu'en est-il en absence de site donneur D4 ? Pour répondre à cette question, nous avons utilisé la construction L3 réalisée par L. Damier, qui diffère de la construction L3-U1 par l'absence du site donneur D4 et du site de fixation de la snRNP U1 (Damier 1997). L'ARN L3 a été incubé en extrait nucléaire, avec ou sans ajout de protéines SR recombinantes 9G8, ASF/SF2 ou SC35, ou d'un mélange naturel de protéines SR dans des conditions permettant l'épissage (Fig. V.21).

Sans ajout de protéines SR, l'ARN L3 n'est pas reconnu efficacement par la machinerie spliceosomale et aucun produit d'épissage n'est détecté (piste 2), ce qui confirme les données publiées par Amendt

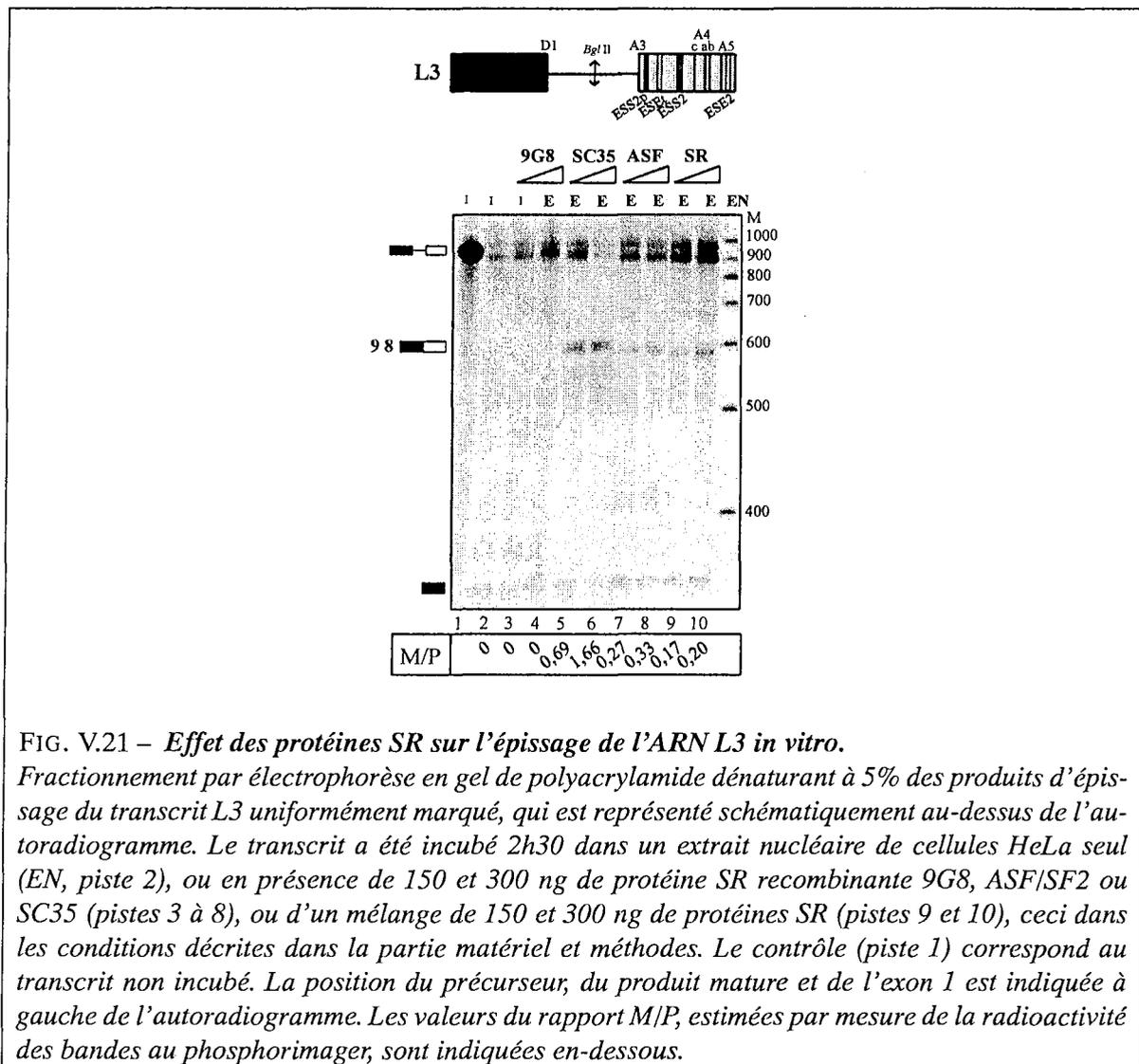


FIG. V.21 – *Effet des protéines SR sur l'épissage de l'ARN L3 in vitro.*

Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage du transcrit L3 uniformément marqué, qui est représenté schématiquement au-dessus de l'autoradiogramme. Le transcrit a été incubé 2h30 dans un extrait nucléaire de cellules HeLa seul (EN, piste 2), ou en présence de 150 et 300 ng de protéine SR recombinante 9G8, ASF/SF2 ou SC35 (pistes 3 à 8), ou d'un mélange de 150 et 300 ng de protéines SR (pistes 9 et 10), ceci dans les conditions décrites dans la partie matériel et méthodes. Le contrôle (piste 1) correspond au transcrit non incubé. La position du précurseur, du produit mature et de l'exon 1 est indiquée à gauche de l'autoradiogramme. Les valeurs du rapport M/P, estimées par mesure de la radioactivité des bandes au phosphorimager, sont indiquées en-dessous.

et al. (1994). On assiste même à une dégradation de cet ARN. Lorsque l'on ajoute un mélange de protéines SR (pistes 9 et 10) ou la protéine SR recombinante ASF/SF2 (pistes 7 et 8) au mélange réactionnel, on observe une faible utilisation du site A3. L'efficacité est plus importante dans le cas de l'ajout de protéine SC35 et une fois de plus, la protéine 9G8 a un effet très faible.

Nous retrouvons donc, pour cette construction dépourvue de site D4, les mêmes effets relatifs des protéines SR sur le site A3, que pour la construction L3-U1. Mais dans ce cas, aucune utilisation des sites A4c, a, b et A5 n'est observée.

Il ressort donc de cette étude que les 4 sites A4c, a, b et A5 sont plus fortement dépendants de la présence du site donneur D4 que le site A3, puisque dans le cas du site A3, l'absence de site D4 peut être compensée par l'ajout de protéines SR, notamment SC35.

V.4.6.2 Rôle des séquences inhibitrices ESS2p et ESS2 dans la régulation de l'utilisation du site A3 par les protéines SR

On pouvait se demander si l'effet des protéines SR au site A3 consistait à compenser l'effet négatif des éléments ESS2p et ESS2 fixant l'un, la protéine hnRNP H, et l'autre, la protéine hnRNP A1 (Caputi *et al.* 1999 ; Del Gatto-Konczak *et al.* 1999 ; Jacquenet *et al.* 2001a). En effet, la protéine hnRNP A1 est connue pour avoir un effet antagoniste de celui des protéines SR. Il était donc intéressant de regarder si les taux d'activation par les protéines SR étaient dépendants de la présence de ces éléments. Dans ce but, nous avons étudié l'effet des protéines SR lorsque l'un ou l'autre, ou encore les 2 éléments inhibiteurs sont éliminés. Nous avons donc réalisé la construction L3.ΔESS2-U1, dans laquelle la séquence renfermant l'élément ESS2, située entre les positions 5429 à 5443, a été délétée (Fig. V.22 ; Jacquenet 2001). Nous avons également utilisé deux constructions qui avaient été réalisées par S. Jacquenet, permettant de produire, la construction L3.ΔESS2p-U1 dans laquelle l'élément ESS2p est délété (position 5360 à 5366) et la construction L3.ΔESS2pΔESS2-U1 dans laquelle les deux éléments ESS2p et ESS2 sont délétés (Fig. V.22).

L'effet de l'ajout des protéines SR recombinantes 9G8, ASF/SF2 et SC35 et d'un mélange de protéines SR sur l'épissage *in vitro* de chacun des ARN produits par ces constructions a été testé, dans les conditions d'épissage décrites précédemment (Fig. V.23).

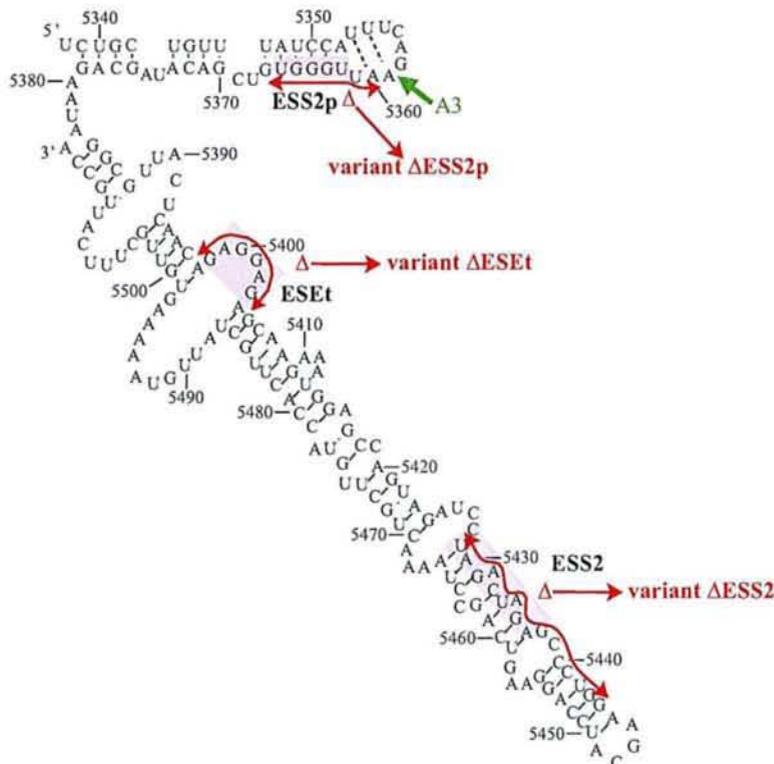


FIG. V.22 – Variants de l'ARN L3-U1 utilisés pour étudier les inter-relations entre les éléments régulateurs agissant sur le site A3.

La séquence délétée dans chaque variant de l'ARN L3-U1 est délimitée par des flèches.

Comme attendu, la délétion de l'élément ESS2p dans l'ARN L3.ΔESS2p-U1 améliore l'efficacité d'épissage au site A3 (d'un facteur 2,3 par rapport à l'ARN L3-U1 ; Fig. V.23-B) (Jacquet *et al.* 2001a). Par contre, aucune diminution n'est observée au niveau du site A5. La protéine 9G8 n'a pas d'effet inhibiteur sur le site A3 comparé à l'ARN L3-U1 (pistes 2 et 3). Les protéines SC35 et ASF/SF2 activent toujours le site A3, mais l'effet activateur de ASF/SF2 est plus marqué que dans le cas de l'ARN L3-U1. Il est de même pour le mélange de protéines SR (pistes 7 à 10). En présence de 300 ng de protéine ASF/SF2 ou du mélange de protéines SR, les taux d'activation sont respectivement de 6,7 et 7,4, au lieu de 3,2 et 3,3 pour l'ARN L3-U1.

Nous en concluons donc que les protéines SR n'ont sans doute pas une action consistant à contrecarrer l'action de l'élément ESS2p, mais qu'au contraire, cet élément limite l'action de la protéine ASF/SF2.

Lorsque l'élément ESS2 est délété (ARN L3.ΔESS2-U1 ; Fig. V.23-C), l'efficacité d'épissage au site A3 est fortement augmentée (d'un facteur 18 par rapport à l'ARN L3-U1). Ceci s'explique par la levée de l'inhibition du site A3 exercée par l'élément ESS2.

La protéine 9G8 a toujours un effet négatif sur A3 à concentration élevée, alors qu'elle stimule l'utilisation du site A5. Les protéines SC35 et ASF/SF2 activent toujours l'épissage au site A3, mais le taux d'activation par la protéine SC35 est fortement diminué (facteur de 6,6 fois contre 80,5 pour l'ARN L3-U1).

Cette observation est en excellent accord avec les données de nos expériences d'empreintes sur l'ARN S2.3, réalisées en présence de SC35 (Fig. V.20), et montrant une forte fixation de SC35 sur l'élément ESS2. On voit que l'activation par ASF/SF2 est peu modifiée en absence de l'élément ESS2, ce qui confirme son absence de fixation sur l'élément ESS2. Les activations par SC35 et ASF/SF2 observées dans le cas de l'ARN L3.ΔESS2-U1 pourraient être générées en grande partie par fixation de ces protéines à l'élément ESEt.

Dans le cas de l'ARN L3.ΔESS2pΔESS2-U1 ne contenant, ni l'élément ESS2p, ni l'élément ESS2 (Fig. V.23-D), le taux d'épissage sans ajout de protéines SR est augmenté d'un facteur 29 par rapport à l'ARN L3-U1.

Les taux d'activation observés pour ASF/SF2 et le mélange de protéines SR sont voisins de ceux observés pour l'ARN L3.ΔESS2-U1. Le taux d'activation par SC35 est en revanche légèrement inférieur, mais il faut signaler que la différence peut être liée à des erreurs de quantification, dans la mesure où ces deux ARN sont très fortement épissés au site A3.

V.4.6.3 Recherche d'une inter-relation possible entre l'élément ESEt et les éléments inhibiteurs ESS2p et ESS2

Afin de savoir si l'élément ESEt contrecarre l'un des deux ou les deux éléments inhibiteurs, nous avons étudié l'effet de délétions simultanées de l'élément ESEt et de l'un des deux éléments inhibiteurs sur l'efficacité d'épissage au site A3, avec ou sans ajout de protéines SR, ceci dans le contexte de la construction L3-U1 (Fig. V.22 et V.24).

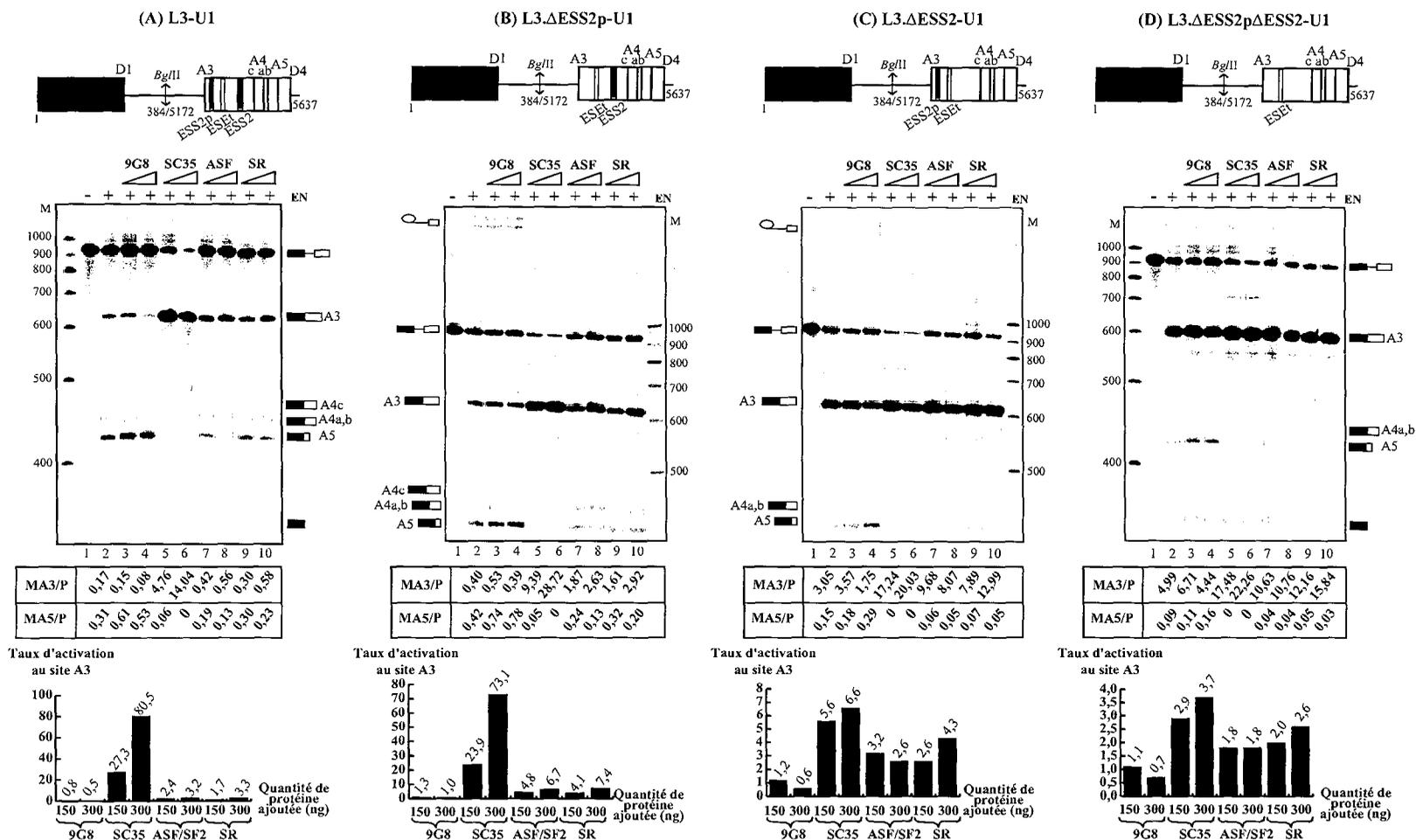
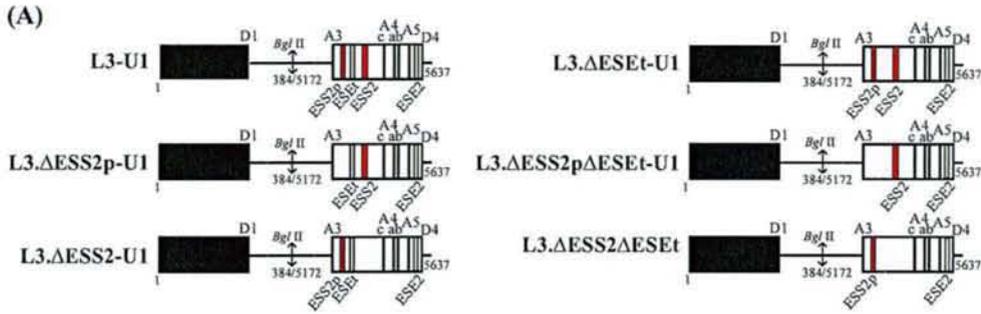


FIG. V.23 – *Etude de l'influence d'un mélange de protéines SR ou de différentes protéines SR purifiées sur l'épissage in vitro de variants de l'ARN L3-U1.*

Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage des transcrits L3-U1 sauvage ou mutés, uniformément marqués. Les ARN utilisés sont représentés de manière schématique au-dessus de chaque autoradiogramme. Les transcrits ont été incubés 2h30 dans un extrait nucléaire de cellules HeLa seul (EN, pistes 2), ou en présence de 150 et 300 ng de protéine SR recombinante 9G8, ASF/SF2 ou SC35 (pistes 3 à 8), ou d'un mélange de 150 et 300 ng de protéines SR (pistes 9 et 10), ceci dans les conditions décrites dans la partie matériel et méthodes. Le contrôle (pistes 1) correspond aux transcrits non incubés. La position des précurseurs, produits matures et intermédiaires est indiquée sur le côté des autoradiogrammes. Les valeurs de M/P, calculées par mesure de la radioactivité pour les sites A3 (MA3/P) et A5 (MA5/P), sont indiquées en-dessous, ainsi qu'une représentation sous forme d'histogramme des taux d'activation du site A3 par les protéines SR.



(B)

	L3-U1	L3.ΔESEt-U1	L3.ΔESS2p-U1	L3.ΔESS2pΔESEt-U1	L3.ΔESS2-U1	L3.ΔESS2ΔESEt-U1
EN	- + + +	- + + +	- + + +	- + + +	- + + +	- + + +
SC35	- - + -	- - + -	- - + -	- - + -	- - + -	- - + -
ASF/SF2	- - - +	- - - +	- - - +	- - - +	- - - +	- - - +

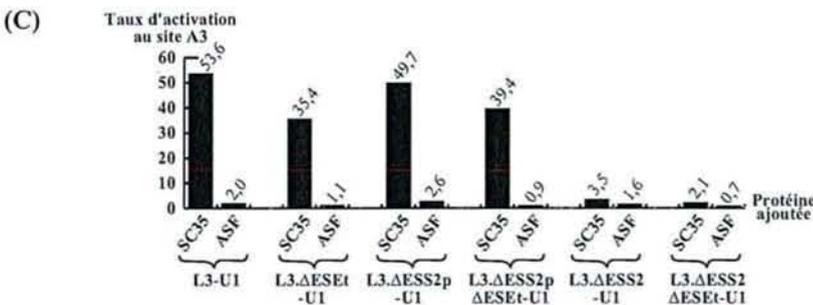
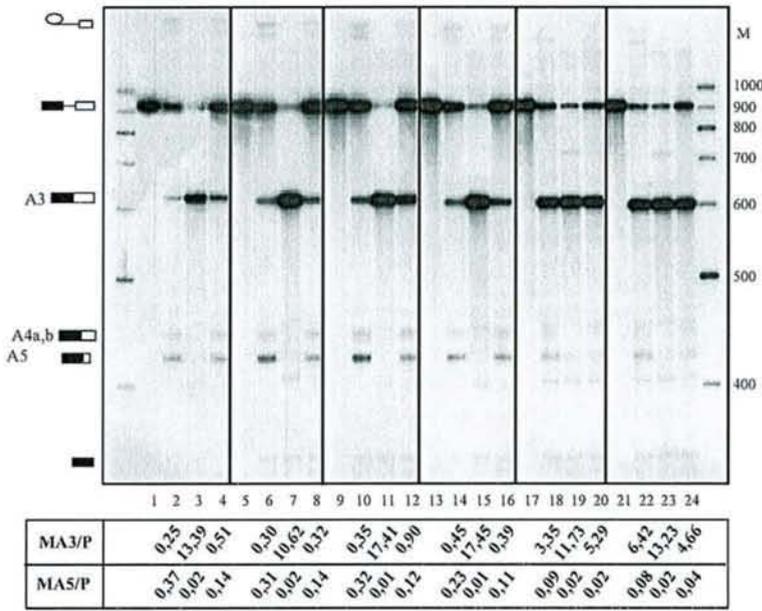


FIG. V.24 – Effet de la délétion des éléments ESEt, ESS2p et ESS2 sur l'activation par les protéines ASF/SF2 et SC35.

(A) Représentation schématique des ARN utilisés pour cette étude.

(B) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage des ARN uniformément marqués. Les transcrits ont été incubés pendant 2h30 dans un extrait nucléaire de cellules HeLa seul (EN), ou complété par 300 ng de protéine purifiée ASF/SF2 ou SC35, dans les conditions décrites dans la partie matériel et méthodes. La position des précurseurs, produits matures et intermédiaires est indiquée à gauche de l'autoradiogramme. Les valeurs du rapport M/P, calculées par mesure de la radioactivité pour les sites A3 (MA3/P) et A5 (MA5/P), sont indiquées en-dessous, ainsi qu'une représentation sous forme d'histogramme des taux d'activation du site A3 par les protéines SR.

Contrairement à ce que nous avons observé pour l'ARN court C3 ne contenant ni l'élément ESS2, ni l'ensemble des sites A4c, A4a, A4b, A5 et D4, la délétion de l'élément ESEt dans la construction L3-U1 (ARN L3.ΔESEt-U1) n'affecte pas l'efficacité d'épissage au site A3 (Fig. V.24, piste 6 vs 2). Aucune activation par ASF/SF2 n'est observée (piste 8 vs 6 ; facteur d'activation de 1,1), ce qui montre que l'élément ESEt est le seul élément responsable de l'activation du site A3 par ASF/SF2, même dans un contexte étendu. Ces données confirment les effets que nous avons précédemment obtenus avec l'ARN C3. L'élément ESE2 décrit comme fixant la protéine ASF/SF2, et situé en amont du site D4 (Kammler *et al.* 2001), n'agit donc pas sur le site A3.

La délétion de l'élément ESEt affecte partiellement l'activation du site A3 par la protéine SC35 (piste 7 vs 6). Le taux d'activation passe en effet d'un facteur 53,6 pour l'ARN sauvage à 35,4 pour l'ARN L3.ΔESEt-U1. Cela confirme que la fixation de SC35 à l'élément ESEt permet l'activation du site A3. L'activation résiduelle de 35,4 serait due à la fixation de la protéine SC35 sur d'autres sites, vraisemblablement sur l'élément ESS2.

Lorsqu'en plus de l'élément ESEt, l'élément ESS2p est délété (ARN L3.ΔESS2pΔESEt-U1 ; pistes 13 à 16), le taux d'épissage est, comme attendu, légèrement augmenté, aussi bien en absence ou en présence de protéine SR ajoutée, mais le taux d'activation par la protéine SC35 est voisin de celui obtenu pour l'ARN L3.ΔESEt-U1 (facteur 39,4). Il en est de même pour l'activation par ASF/SF2. Ces données confirment que la régulation exercée par l'élément ESS2p est indépendante de celle exercée par l'élément ESEt.

Par ailleurs, lorsqu'on délète à la fois les éléments ESS2 et ESEt (ARN L3.ΔESS2ΔESEt-U1 ; pistes 21 à 24), le taux d'épissage est significativement augmenté, mais en contrepartie, les taux d'activation par les protéines SR sont faibles (facteur 2 pour SC35). L'ajout d'ASF/SF2 conduit même, dans ce cas, à une diminution du taux d'épissage. On peut imaginer la fixation d'ASF/SF2 sur un site de plus faible affinité qui, de par sa position, diminuerait l'efficacité d'épissage. La très faible activation observée pour SC35, qui confirme les résultats obtenus pour l'ARN C3 délété de l'élément ESEt, suggère que cette protéine agit bien essentiellement en se fixant aux deux sites, ESEt et ESS2, que nous avons détectés par expériences d'empreintes et de retards sur gel.

De manière générale, les délétions des éléments ESEt et ESS2p n'ont pas d'effet marqué sur l'utilisation du site A5. Par contre, en absence d'élément ESS2, le site A5 est très faiblement utilisé, en particulier après ajout de protéines SR. Dans ce cas, A5 est sans doute régulé par défaut.

Enfin, il est à noter qu'entre les deux expériences d'épissage *in vitro* que nous venons de présenter (Fig. V.23 et V.24), nous avons obtenu des variations du taux d'activation du site A3 par les protéines SR. Ceci s'explique vraisemblablement par un changement de lot d'extrait nucléaire d'une expérience à l'autre.

Il ressort donc de cette étude que le site d'épissage A3 est soumis à une triple régulation :

- celle exercée par l'élément ESS2p, qui semble indépendante des autres régulations et qui limite l'action de la protéine ASF/SF2 ;

- celle médiée par l'élément ESEt. Le fait que la délétion de cet élément conduise à une légère augmentation de l'utilisation du site A3, surtout en absence de l'élément ESS2, suggère qu'il pourrait aussi fixer un composant nucléaire ayant un effet inhibiteur et qui serait déplacé par ASF/SF2 ou SC35 ;
- la dernière régulation s'exerçant sur le site A3 dépend de l'élément ESS2 qui inhibe l'épissage en fixant la protéine hnRNP A1 (Caputi *et al.* 1999 ; Del Gatto-Konczak *et al.* 1999). Nos données montrent qu'il est également capable de fixer la protéine SC35. L'ajout de protéine SC35 déplacerait la protéine hnRNP A1, ce qui lèverait l'inhibition de l'épissage et expliquerait le fort effet activateur de SC35 sur le site A3 que nous avons observé dans les cellules HeLa.

Il est à noter que l'affinité de la protéine SC35 pour les éléments ESEt et ESS2 ne reflète pas l'effet activateur observé à chacun des deux éléments. En effet, la délétion de l'élément ESEt a moins d'effet sur l'activation du site A3 par SC35 que la délétion de l'élément ESS2, alors que d'après des données récentes du laboratoire, obtenues par des expériences de retard sur gel, la protéine a plus d'affinité pour l'élément ESEt. On peut admettre que SC35 n'a pas besoin d'avoir une forte affinité pour l'élément ESS2, pour déplacer la protéine hnRNP A1.

V.4.7 Des substitutions de bases mimant des mutations observées dans l'ARN viral de cellules chroniquement infectées affectent l'épissage au site A3

Sur la base d'une étude réalisée dans des cellules H9 chroniquement infectées par le virus HIV-1, l'équipe de S. Litvak à Bordeaux a proposé l'existence d'un phénomène d'édition de l'ARN pré-messager du virus HIV-1 (Bourara *et al.* 2000). Les sites édités correspondent, l'un, à la conversion d'un C en U et les cinq autres, à des conversions de G en A. Plus récemment, trois études ont montré que les éditions G→A ont en réalité lieu au niveau de l'ADN proviral, lors de la transcription inverse (Mangeat *et al.* 2003 ; Zhang *et al.* 2003 ; Harris *et al.* 2003).

Comme nous l'avons vu dans l'introduction, l'enzyme cellulaire APOBEC3G est responsable de ces éditions de bases, qui auraient un effet négatif sur la multiplication virale, du fait de la destruction de phases ouvertes de lecture ou d'éléments fonctionnels. Ces études ont également révélé que la protéine Vif est capable de contrecarrer l'action de la protéine APOBEC3G ; ce mécanisme de défense a été mis en place par le virus pour conserver l'intégrité de son génome.

Il est vraisemblable que dans la souche virale étudiée par Bourara *et al.* (2000), la conversion C→U correspond bien à une édition de l'ARN, alors que les cinq autres conversions G→A ont lieu lors de la synthèse de l'ADNc simple-brin. La souche sur laquelle Bourara *et al.* ont travaillé exprimait une protéine Vif active. Les substitutions de bases retenues correspondraient à un facteur favorable dans la sélection qui avait été effectuée, à savoir un couple virus/cellule H9 permettant la survie des cellules. De manière intéressante, sur les six sites d'édition détectés par Bourara *et al.* dans l'ARN viral issu de cellules chroniquement infectées, deux sont localisés dans la région du site A3 : l'un correspond à la conversion du nucléotide G5363 en A dans l'élément ESS2p (UGGGU→5UAGGU) et l'autre correspond à l'édition du nucléotide C5396 en U en amont de l'élément ESEt (CAGAGGAG→UAGAGGAG). Ces 2 substitutions de base apparaissent le plus souvent de manière concomitante.

Compte-tenu de la proximité des sites d'édition avec les séquences régulatrices du site A3, il était

intéressant d'étudier si ces substitutions de bases affectaient le taux d'épissage à ce site.

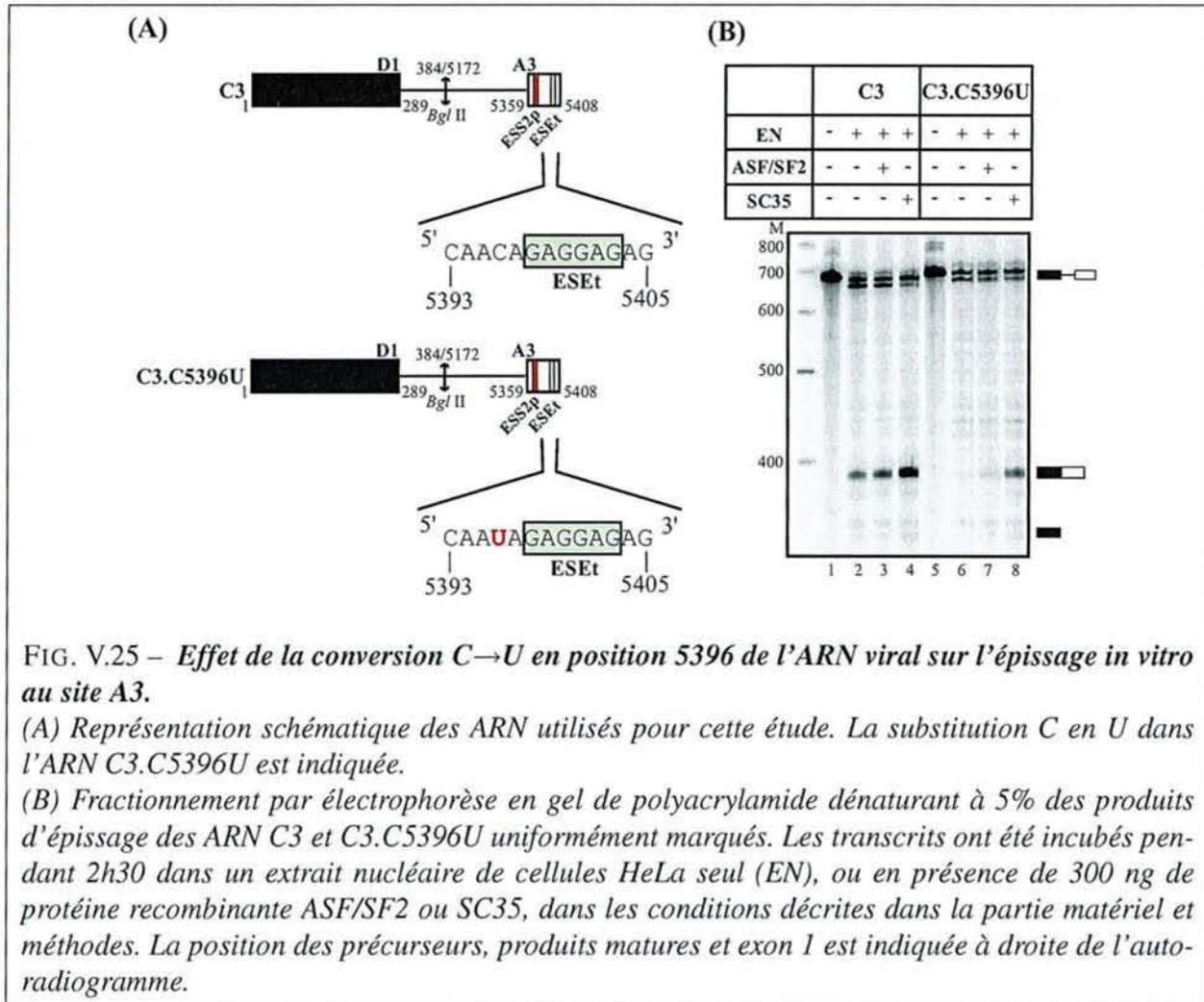


FIG. V.25 – Effet de la conversion C→U en position 5396 de l'ARN viral sur l'épissage in vitro au site A3.

(A) Représentation schématique des ARN utilisés pour cette étude. La substitution C en U dans l'ARN C3.C5396U est indiquée.

(B) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage des ARN C3 et C3.C5396U uniformément marqués. Les transcrits ont été incubés pendant 2h30 dans un extrait nucléaire de cellules HeLa seul (EN), ou en présence de 300 ng de protéine recombinante ASF/SF2 ou SC35, dans les conditions décrites dans la partie matériel et méthodes. La position des précurseurs, produits matures et exon 1 est indiquée à droite de l'autoradiogramme.

Nous avons étudié l'effet de la substitution de base C5396U située en amont de l'élément ESEt. L'effet de cette mutation sur l'efficacité d'épissage au site A3 dans l'ARN C3 est présenté dans la figure V.25. Il apparaît que la substitution C5396U abolit l'utilisation du site A3 dans l'extrait nucléaire (Fig. V.25-B, piste 6) et affecte fortement l'activation du site A3 par ASF/SF2 et SC35 (pistes 7 et 8).

En collaboration avec l'équipe de M. Stoltzfus, nous avons ensuite testé, par transfection de cellules HeLa avec le plasmide pΔPSP, l'effet des 2 éditions G5363A et C5396U sur la production des ARNm *tat*, qui sont épissés au site A3 (Fig. V.26). Les résultats obtenus montrent que la substitution G→A affecte légèrement la production des ARNm *tat*, que la substitution C→U a un effet plus marqué et que la double substitution abolit totalement la synthèse d'ARNm *tat1*.

Les ARN viraux isolés par Bourara *et al.* (2000) portaient donc deux substitutions de bases, sans doute générées l'une au niveau de l'ADNc proviral, l'autre au niveau de l'ARN, dont l'effet, en plus d'interrompre l'ORF Vpr, était de bloquer l'utilisation du site A3.

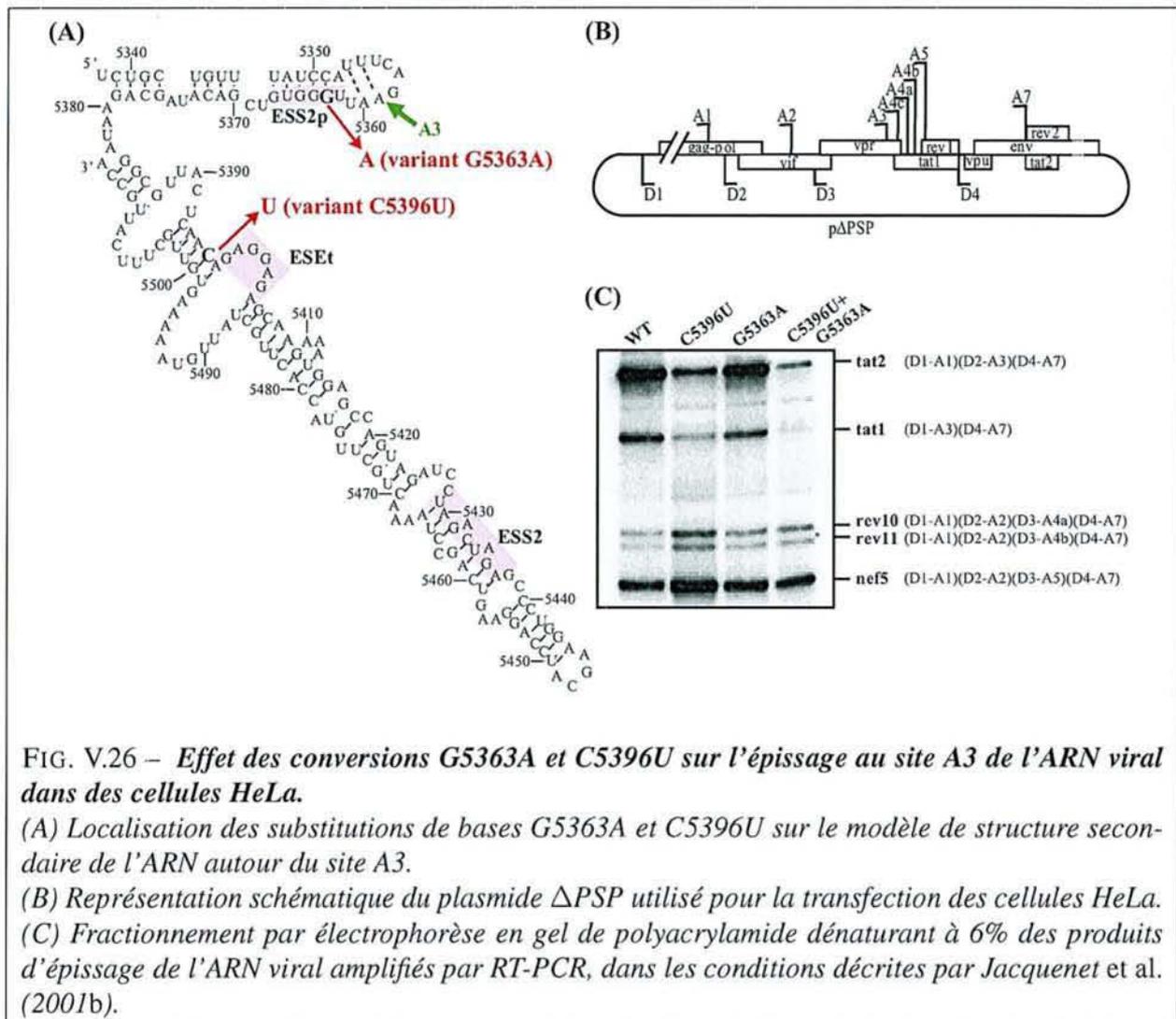


FIG. V.26 – *Effet des conversions G5363A et C5396U sur l'épissage au site A3 de l'ARN viral dans des cellules HeLa.*

(A) *Localisation des substitutions de bases G5363A et C5396U sur le modèle de structure secondaire de l'ARN autour du site A3.*

(B) *Représentation schématique du plasmide ΔPSP utilisé pour la transfection des cellules HeLa.*

(C) *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).*

V.4.8 Conclusions et discussion des résultats

V.4.8.1 Le site A3 est soumis à un réseau complexe de régulations par des protéines nucléaires

Un nouvel élément de régulation, ESEt, fixe les protéines ASF/SF2 et SC35

Par des expériences d'empreintes biochimiques et d'épissage *in vitro*, nous avons mis en évidence un 3^e élément régulant l'utilisation du site A3, ESEt (Fig. V.27). Il agit en fixant les protéines SR ASF/SF2 et SC35 et conserve ses propriétés activatrices dans un contexte hétérologue.

L'effet positif de l'élimination de l'élément ESEt sur l'épissage au site A3 dans la construction longue, suggère que cet élément fixe également un composant nucléaire à activité négative qui reste à identifier. La fixation des protéines ASF/SF2 et/ou SC35 sur l'élément ESEt chasse sans doute ce composant. D'après des données de H. Hallay, obtenues par expériences d'empreintes sur l'ARN S2.3, ce composant pourrait être la protéine hnRNP A1.

La séquence de l'élément ESEt (GAGGAG) correspond aux séquences généralement reconnues par les protéines ASF/SF2 et SC35, d'après les données des études SELEX qui ont été publiées (cf Tab. II.3, page 67 de la revue bibliographique). Cette séquence a également été identifiée dans des sé-

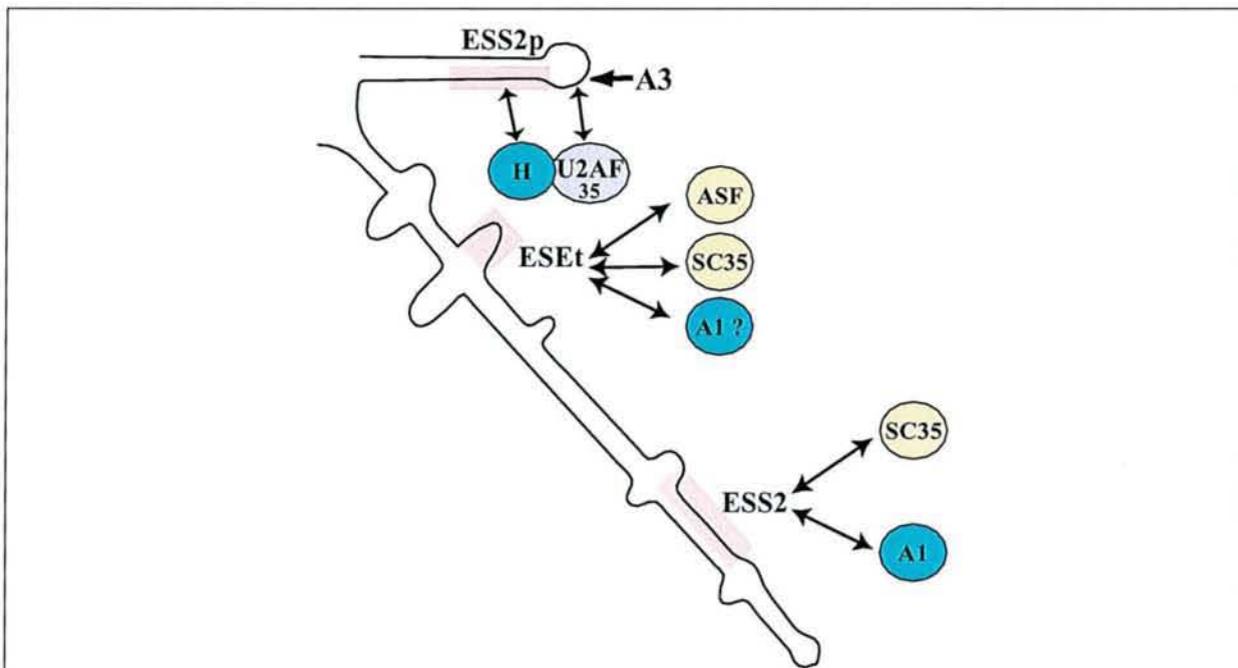


FIG. V.27 – Un réseau complexe d'interactions avec des protéines nucléaires contrôle le taux d'utilisation du site A3.

Au moins trois éléments régulateurs contrôlent l'efficacité d'épissage du site A3. La fixation de la protéine hnRNP H à l'élément ESS2p réprime le site A3 en empêchant la fixation du facteur U2AF³⁵ (Jacquenot et al. 2001a). L'élément ESEt active l'utilisation du site A3 en fixant les protéines ASF/SF2 et SC35, qui chassent sans doute un composant nucléaire inhibiteur fixé sur ce site. Enfin, une compétition s'exerce entre les protéines hnRNP A1 et SC35 pour leur fixation à l'élément ESS2 : la protéine hnRNP A1 fixée à l'élément ESS2 inhiberait l'épissage du site A3. Cette inhibition serait levée par l'association de la protéine SC35 à l'élément ESS2.

quences activatrices d'épissage naturelles. Par exemple, dans l'ARN pré-messager de la β -tropomyosine de rat, deux séquences constituées de trois répétitions GAG fixent ASF/SF2 et activent l'inclusion de l'exon 6 (cf Tab. II.4 page 68 de la revue bibliographique ; Selvakumar & Helfman 1999).

Il est à noter que, dans l'exon IIIA de l'ARN pré-messager de la fibronectine, la protéine SRp40 reconnaît une séquence riche en purines (AGGAGAAGGGA ; Kuo & Norton 1999), qui diffère nettement des séquences de haute affinité déterminées pour cette protéine (UGGGAGCRGUYRGCUCGY et ACDGS ; cf Tab. II.3, page 67). Compte-tenu du fort effet activateur de la protéine SRp40 sur le site A3, il sera intéressant de tester si cette protéine agit en se fixant sur l'élément ESEt.

La fixation des protéines SR ASF/SF2 et SC35 sur l'élément ESEt expliquerait une observation faite par Caputi *et al.* (1999). Au cours d'expériences de déplétion de l'extrait nucléaire en protéine hnRNP A1, à l'aide d'un ARN contenant l'élément ESS2, ces auteurs ont également observé une perte de protéines SR. Ceci les a obligé à compléter l'extrait nucléaire déplété, en protéines SR, pour les expériences d'épissage *in vitro*. En réalité, l'ARN utilisé était assez grand et contenait l'élément ESEt, ce qui devait conduire à une élimination des protéines SR fixées à cet élément.

Lorsque les études sont réalisées *ex vivo* sur l'ARN produit à partir de la construction p Δ PSP, aucune augmentation de l'utilisation du site A3 n'est observée lorsqu'on surexprime ASF/SF2, l'effet activateur d'ASF/SF2 étant plus fort au site A2 (Fig. 5 de l'article n° 1). A l'issue de cette étude, nous avons conclu qu'au niveau des sites A2 et A3 en compétition, l'activation du site A2 par ASF/SF2 se faisait aux dépens de celle du site A3. Mais on peut envisager qu'en absence d'élément ESEt, on n'ait plus du tout d'utilisation du site A3 lorsqu'ASF/SF2 est surexprimée. Dans ces conditions, l'élément ESEt pourrait permettre le maintien d'une utilisation minimale du site A3.

On pourrait ainsi imaginer qu'il existe dans des ARN multiplement épissés tels que celui de HIV-1, deux types d'éléments activateurs :

- des activateurs forts, qui induisent l'utilisation préférentielle d'un site d'épissage donné, en présence d'un excès d'une protéine SR donnée ou d'un facteur régulateur les reconnaissant,
- des éléments de régulation plus faibles, permettant de préserver un niveau de base minimum des sites autres que celui fortement activé. Cela pourrait être un rôle possible de l'élément ESEt.

SC35 active le site A3 en se liant à la fois aux éléments ESEt et ESS2

En plus de l'élément ESEt, l'ensemble de nos résultats a montré une fixation de la protéine SC35 sur l'élément ESS2. La forte activation du site A3 par SC35 pourrait être liée au fait que cette protéine SR entre en compétition avec la protéine hnRNP A1 au niveau de l'élément régulateur ESS2. Cette hypothèse est actuellement testée au laboratoire.

Une telle compétition entre les protéines SC35 et hnRNP A1 a également été proposée au site A7. Dans ce cas, la fixation de la protéine SC35 sur la séquence joignant les éléments ESS3a et ESS3b, empêcherait la fixation de la protéine hnRNP A1 à l'élément inhibiteur ESS3b (Mayeda *et al.* 1999 ; Zhu *et al.* 2001).

Il est intéressant de rappeler que les éléments ESS3 et ESS2 présentent de fortes homologies de séquences. Nos données suggèrent qu'ils fonctionnent de manière similaire, en fixant à la fois hnRNP A1 qui a un fort potentiel inhibiteur, et SC35, qui est un activateur.

La séquence UAGAGCCCUG protégée par SC35 (Fig. V.20), et chevauchant l'extrémité 3' de l'élément ESS2, des positions 5434 à 5444, pourrait constituer le site de fixation de SC35 sur l'élément ESS2. En effet, cette séquence est similaire à celles identifiées par l'approche SELEX comme fixant la protéine SC35 avec une haute affinité (Schaal & Maniatis 1999b ; Liu *et al.* 2000) (cf tableau II.3, page 67). Nos expériences d'épissage sur l'ARN variant L3. Δ ESS2-U1 montrent que l'élimination de cette séquence affecte l'activation du site A3 par SC35 (cf Fig. V.22 et V.23-C). Ainsi, en plus d'empêcher la fixation de la protéine hnRNP A1 à l'élément ESS2, la protéine SC35 pourrait activer directement l'épissage au site A3 en se fixant à cette séquence. Il paraissait donc important d'étudier le rôle potentiel de cette séquence dans l'activation du site A3 par SC35, par des expériences de mutagenèse dirigée, suivies d'expériences de retard sur gel et d'épissage *in vitro*. Ces expériences sont en cours au laboratoire.

Etant donné qu'*ex vivo* nous avons trouvé un fort effet activateur de la protéine SRp40 sur le site A3, il sera intéressant de tester si cette protéine agit aussi en contrecarrant la fixation de la protéine hnRNP A1. Nous n'avons, pour l'instant, pas trouvé de site potentiel de fixation de cette protéine au niveau

de l'élément ESS2.

En ce qui concerne l'élément ESS2p, il agit de manière indépendante, mais limite les performances de la protéine ASF/SF2 au site A3. La protéine hnRNP H empêcherait donc la fixation du facteur U2AF³⁵, sans que les protéines SR entrent en compétition avec sa fixation sur l'élément ESS2p.

Comment les protéines SR régulent-elles le taux de production des ARNm viraux ?

Nos expériences d'épissage *in vitro* montrent clairement que les 2 éléments ESS2p et ESS2 sont requis pour qu'on puisse avoir utilisation du site A5 et par conséquent, production de protéine Nef. En effet, le site A5 est le plus utilisé en l'absence de protéine SR ajoutée. Si on ajoute ASF/SF2 et SC35, le taux d'utilisation du site A5 diminue. Cette diminution en présence des 2 protéines SR est plus marquée si l'élément ESS2p est délété et encore plus, en absence de l'élément ESS2. Le site A5 est donc sans doute régulé par défaut. Ainsi, dans une situation où la concentration en protéine SC35 est augmentée dans les cellules, comme cela a été décrit tôt après l'infection, le site A3 peut être plus fortement utilisé que le site A5, ce qui peut conduire à une production élevée de protéine Tat.

Une activation du site A5 par la protéine 9G8 est néanmoins possible puisque, d'après nos données d'épissage *in vitro*, cette protéine diminue légèrement l'utilisation du site A3 *in vitro*, mais pas *ex vivo*. Des données récentes de Schaal *et al.* (2003) présentées au congrès « Eukaryotic mRNA Processing 2003 », montrent que l'élément ESE2 situé en amont du site D4 fixe la protéine 9G8, ce qui expliquerait l'activation du site A5 que nous observons.

Les sites A4 sont toujours très faiblement utilisés dans nos conditions expérimentales d'épissage *in vitro* et les protéines SR ont peu d'effet sur leur taux d'utilisation, à l'exception de la protéine ASF/SF2 sur les sites A4a et b. Néanmoins, une étude plus fine de ces sites nécessite la réalisation de constructions dans lesquelles le 2^e exon ne renferme que les 3 sites A4c, A4a et A4b. Il sera ainsi possible de tester l'effet des protéines SR sur l'épissage des ARN produits à partir de ces nouvelles constructions.

V.4.8.2 L'édition de l'ADNc simple brin et peut être de l'ARN viral : un mécanisme de défense contre l'infection virale

L'étude de l'effet de la substitution C5396U en amont de l'élément ESEt a permis de mettre en évidence un effet très marqué de cette substitution de base sur l'épissage au site A3 et sur son activation par les protéines ASF/SF2 et SC35. La substitution G5363A réprime, elle aussi, l'utilisation du site A3 mais à un plus faible niveau. La combinaison des 2 mutations conduit à une très faible production des ARNm *tat1*.

Comment expliquer un tel effet sur l'épissage au site A3 *in vitro* et *ex vivo* ? Ces substitutions de bases créent des éléments UAGG dans l'élément ESS2p et UAGA au niveau de l'élément ESEt, pouvant être reconnus par la protéine hnRNP A1. Ceci aurait pour conséquence de changer l'identité de la protéine hnRNP fixée à l'élément ESS2p (Fig. V.27). La fixation de la protéine hnRNP A1 à l'élément ESS2p muté a été confirmée par expérience de retard sur gel par H. Hallay et son effet sur l'épissage *in vitro*, montré par S. Jacquenet (2001). Des expériences récentes de H. Hallay et L. Ayadi montrent aussi que la substitution C5396U conduit à une meilleure fixation de la protéine hnRNP A1 à l'élément ESEt,

aux dépens des protéines ASF/SF2 et SC35.

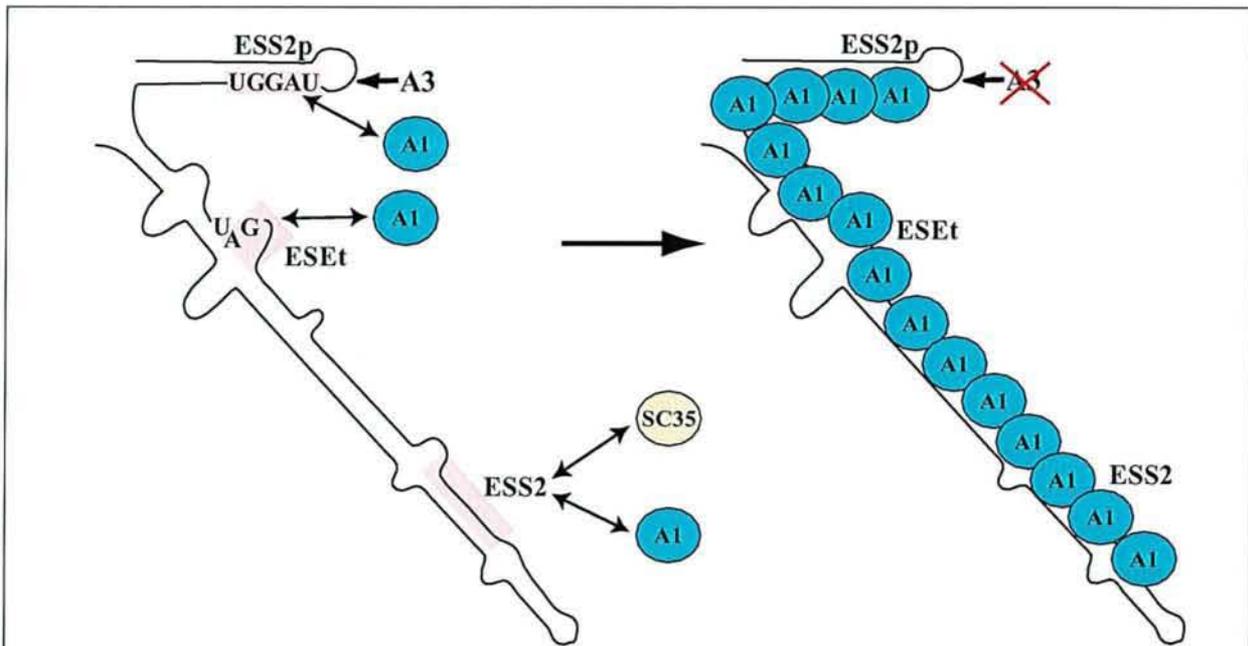


FIG. V.28 – *L'édition de l'ARN viral change l'identité des protéines fixées aux séquences régulatrices du site A3.*

La mutation G→A dans l'élément ESS2p permet la fixation de la protéine hnRNP A1 à cet élément (Jacquet et al. 2001a). La mutation C→U en amont de l'élément ESEt pourrait convertir une partie de cet élément en site de fixation de la protéine hnRNP A1. L'existence de plusieurs autres séquences de type UAG dans cette région de l'ARN viral pourrait conduire à une fixation coopérative de la protéine hnRNP A1 le long de l'ARN, qui empêcherait la reconnaissance du site A3 par les facteurs spliceosomaux. Ceci pourrait expliquer le fort effet inhibiteur des 2 sites d'édition que nous avons étudiés sur l'épissage au site A3 *in vitro* et *ex vivo*.

Après édition de l'ADN, en position 5396, et de l'ARN, en position 5363, on aurait donc génération de 2 sites forts de fixation de la protéine hnRNP A1, ce qui conduirait sans doute à une fixation coopérative de la protéine hnRNP A1 le long de l'ARN (Fig. V.28), empêchant la reconnaissance du site A3 par les facteurs spliceosomaux, comme c'est le cas au site A7 (Marchand *et al.* 2002). Cette limitation de la production de l'ARNm *tat* et donc de la protéine Tat conduirait à une faible production de transcrits viraux plus longs et limiterait l'apoptose des cellules infectées. Ceci pourrait expliquer que ce phénomène d'édition de l'ARN viral ait été observé dans des cellules chroniquement infectées, c'est-à-dire ayant résisté à plusieurs cycles successifs d'infection. Il a été maintenant montré que l'édition au niveau de l'ADNc simple-brin est, en effet, un mécanisme de défense de la cellule hôte contre l'infection (cf § III.1.5.7 de la revue bibliographique). Lorsque les mutations générées par édition conduisent à une forme silencieuse du virus, la cellule infectée ne subit que des dommages limités et très peu de virions sont produits.

Chapitre VI

Etude du site A3 de l'ARN des virus SIVcpz ANT et HIV-1 YBF30

Etant donné que nous avons observé un très fort taux de régulation du site A3, susceptible de jouer un rôle important dans le cycle de multiplication virale, il était intéressant de tester si cette régulation, mise en évidence chez HIV-1 BRU, était conservée dans des virus qui, tout en étant apparentés à HIV-1 BRU, étaient relativement distants phylogénétiquement. En effet, une conservation de la régulation du site A3 devait être un argument très fort en faveur d'une haute importance fonctionnelle de cette régulation.

Comme nous l'avons vu dans la revue bibliographique, l'ancêtre commun des souches actuelles des virus SIVcpz et HIV-1 aurait eu les chimpanzés *Pan troglodytes* comme hôte. Une fois transféré à l'homme, les souches virales auraient fortement divergé indépendamment l'une de l'autre, ceci dans le cadre de l'adaptation de la souche à ce nouvel hôte. Parmi les différentes lignées de souches de virus SIVcpz actuellement étudiées, deux sont phylogénétiquement reliées au virus HIV-1 : l'une, correspondant à la souche SIVcpz ANT, infecte le chimpanzé *Pan troglodyte schweinfurtii* (cf Fig. III.1, page 94 ; Vanden Haesevelde *et al.* 1996 ; Gao *et al.* 1999), l'autre, regroupe des souches infectant le chimpanzé *Pan troglodyte troglodyte*. Des transmissions inter-espèces répétées de cette lignée de virus à l'homme auraient conduit à l'apparition des groupes de souches M, N et O du virus HIV-1 (cf Fig. III.1, page 94 ; Gao *et al.* 1999 ; pour revue, Hahn *et al.* 2000). La souche YBF30 appartient quant à elle au groupe N (Simon *et al.* 1998). Il s'agit d'une souche mosaïque, issue de la recombinaison d'une souche HIV-1 du groupe M (pour la partie 5' du génome) et de la souche SIVcpz US (pour la partie 3' du génome) (Simon *et al.* 1998). Cette souche est par conséquent éloignée phylogénétiquement des autres souches du virus HIV-1 (cf Fig. III.1, page 94 de l'introduction).

La majorité des données sur l'épissage de l'ARN du virus HIV-1 provient de l'étude de souches appartenant à la classe B du groupe M, qui sont responsables de la pandémie. Les souches HIV-1 pNL4-3 et HIV-1 BRU, que nous avons utilisées dans la partie des travaux décrite précédemment, appartiennent à ce groupe. Seule une étude menée par l'équipe de M. Stoltzfus a concerné l'épissage de l'ARN de différentes souches des groupes M et O (Bilodeau *et al.* 1999). En dépit de la proximité phylogénétique de ces différentes souches, Bilodeau *et al.* (1999) ont observé des différences pour certains éléments de régulation de l'épissage. Ainsi, la présence d'un élément ESS2 comportant deux répétitions CUAGA n'est pas une caractéristique générale. Certaines souches ne comportent qu'un seul des deux éléments de la répétition. C'est le cas pour les souches HIV-1 SF2 et ELI (cf Fig. 3 de l'article n° 2). Par ailleurs, l'équipe de M. Stoltzfus avait mis en évidence l'inhibition de l'utilisation du site A5 par le site A4b (Swanson & Stoltzfus 1998). Les travaux de Bilodeau *et al.* (1999) mon-

traient que cette propriété était une caractéristique des souches du groupe M et non du groupe O.

Dans un 1^{er} temps, j'ai mené une étude sur la souche HIV-1 MAL, issue d'une recombinaison de souches des classes A, D et I du groupe M. Les efficacités des sites accepteurs de l'ARN de HIV-1 MAL ont été comparées à celles de l'ARN de HIV-1 BRU. Pour cela, j'ai réalisé des constructions permettant de produire des ARN comportant la même partie 5' que les ARN C2, C3 et L3-U1, à savoir, l'exon 1 de HIV-1 BRU, avec une région de l'ARN viral de HIV-1 MAL contenant soit le site A2, soit A3 ou encore, l'ensemble des sites A3 à A5. L'étude de l'épissage des ARNm produits à partir de telles constructions a révélé une efficacité d'épissage des sites accepteurs HIV-1 MAL, similaire à celle des sites accepteurs HIV-1 BRU correspondants. En dépit d'une différence de séquence de l'ordre de 13 % entre le génome des souches HIV-1 BRU et MAL, les régulations s'appliquant sur les sites d'épissage semblent donc conservées, ce qui peut s'expliquer par la conservation, dans la souche HIV-1 MAL, de séquences de régulation connues.

Cette absence de différences entre BRU et MAL renforçait l'intérêt d'étudier, dans un 2^e temps, des souches beaucoup plus divergentes, telles que HIV-1 YBF30 et SIVcpz ANT dont les ARN génomiques respectifs présentent seulement 74% et 66% d'identité avec celui de HIV-1 BRU.

VI.1 Tous les sites d'épissage et de régulation ne sont pas conservés dans les souches HIV-1 YBF30 et SIVcpz ANT

D'après l'alignement de séquences présenté dans la figure VI.1, un point commun aux souches HIV-1 BRU, HIV-1 YBF30 et SIVcpz ANT, est la présence de séquences polypyrimidines sous-optimales en amont de chacun des sites A3. En revanche, les ARN des souches HIV-1 YBF30 et SIVcpz ANT ne possèdent pas tous les sites d'épissage présents dans l'ARN de la souche HIV-1 BRU. Ainsi, parmi les 3 sites A4, seul le site A4a est retrouvé dans les trois souches. A4b, qui a un effet négatif sur l'utilisation du site A5 (Swanson & Stoltzfus 1998), n'est présent que dans l'ARN de HIV-1 BRU et YBF30, tandis que seule la souche HIV-1 BRU contient le site A4c.

En ce qui concerne les éléments régulateurs en *cis*, l'élément ESS2 n'est retrouvé que dans les souches HIV-1 BRU et YBF30, sous forme de 2 répétitions CUAGA dans la souche BRU et d'un seul motif dans la souche YBF30. Enfin, alors que les répétitions de type GAR (R = purine) de l'élément ESE2 décrit par Kammler *et al.* (2001) sont conservées dans les 3 souches, les séquences des éléments ESS2p et ESEt présentent chacune une substitution de base. Ainsi dans la souche YBF30, la séquence GAGGAG de l'élément ESEt (BRU) devient AAGGAG, suivie de purines qui peuvent renforcer la fixation des protéines SR. Par contre, la substitution d'un G en C dans la série de G de l'élément ESS2p dans l'ARN de YBF30 altère sans doute la fixation de la protéine hnRNP H. L'élément ESS2p ne serait donc pas actif dans cette souche. De plus, dans la souche SIVcpz ANT, il est absent.

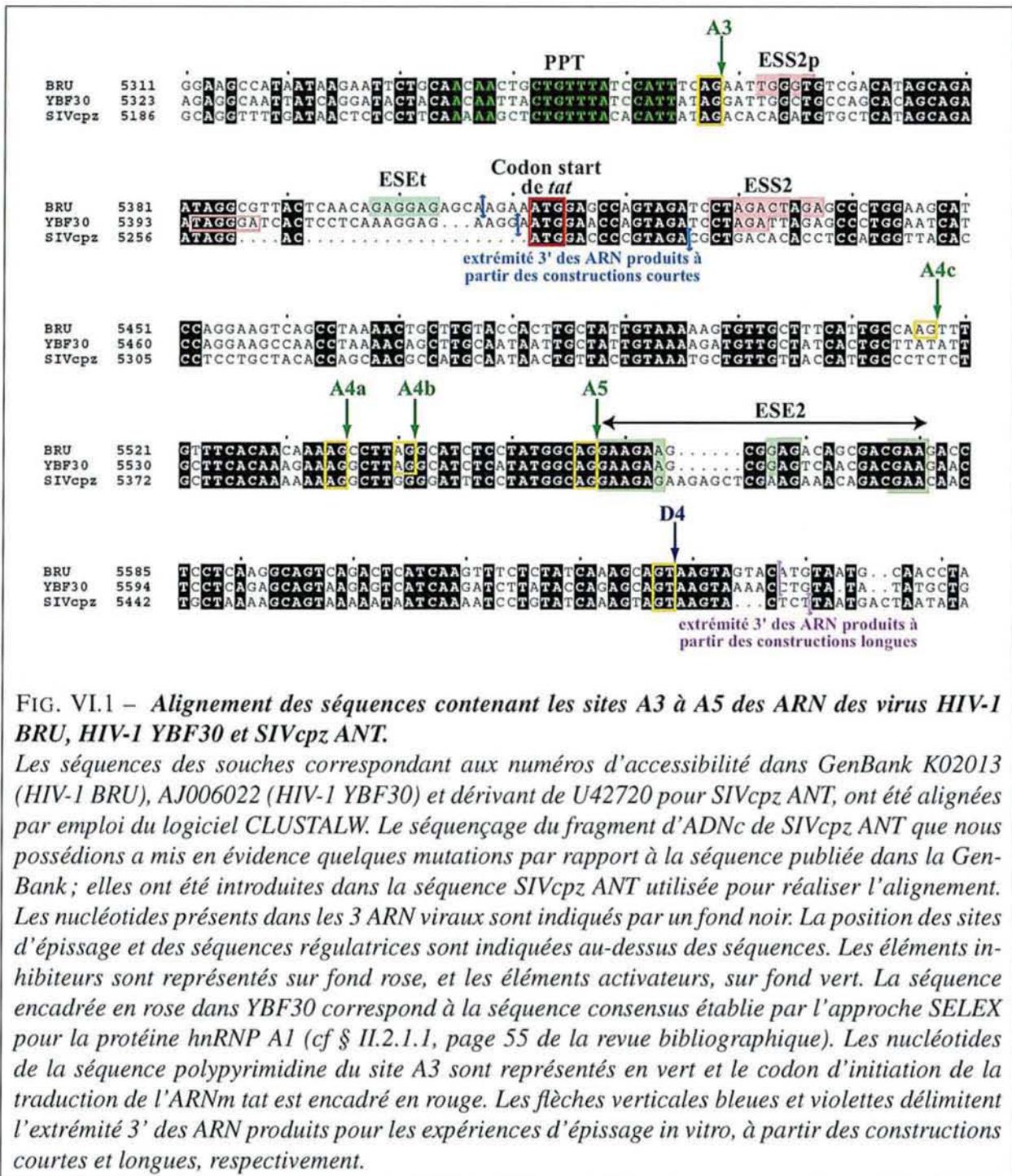
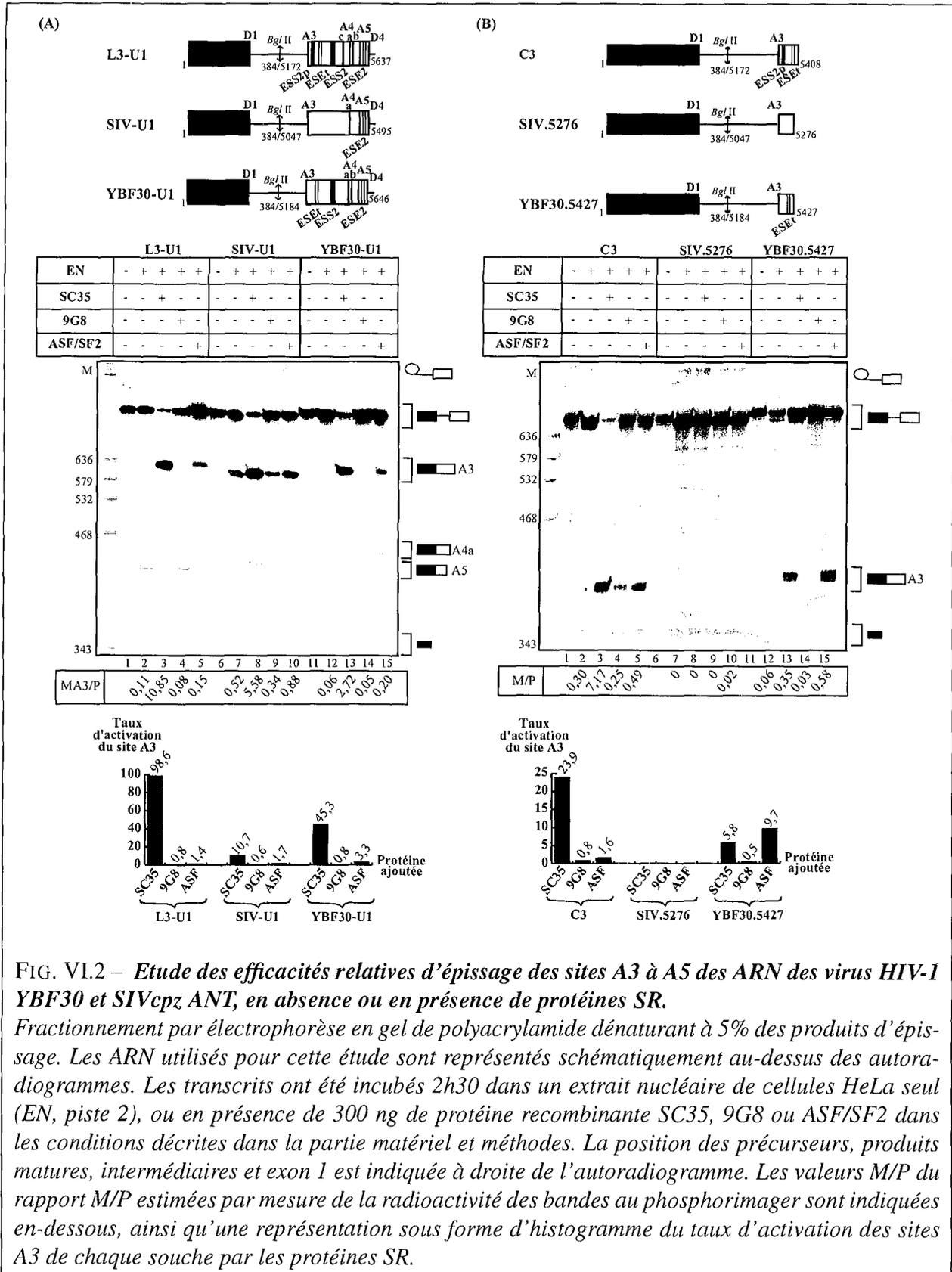


FIG. VI.1 – Alignement des séquences contenant les sites A3 à A5 des ARN des virus HIV-1 BRU, HIV-1 YBF30 et SIVcpz ANT.

Les séquences des souches correspondant aux numéros d'accessibilité dans GenBank K02013 (HIV-1 BRU), AJ006022 (HIV-1 YBF30) et dérivant de U42720 pour SIVcpz ANT, ont été alignées par emploi du logiciel CLUSTALW. Le séquençage du fragment d'ADNc de SIVcpz ANT que nous possédions a mis en évidence quelques mutations par rapport à la séquence publiée dans la GenBank ; elles ont été introduites dans la séquence SIVcpz ANT utilisée pour réaliser l'alignement. Les nucléotides présents dans les 3 ARN viraux sont indiqués par un fond noir. La position des sites d'épissage et des séquences régulatrices sont indiquées au-dessus des séquences. Les éléments inhibiteurs sont représentés sur fond rose, et les éléments activateurs, sur fond vert. La séquence encadrée en rose dans YBF30 correspond à la séquence consensus établie par l'approche SELEX pour la protéine hnRNP A1 (cf § II.2.1.1, page 55 de la revue bibliographique). Les nucléotides de la séquence polypyrimidine du site A3 sont représentés en vert et le codon d'initiation de la traduction de l'ARNm *tat* est encadré en rouge. Les flèches verticales bleues et violettes délimitent l'extrémité 3' des ARN produits pour les expériences d'épissage *in vitro*, à partir des constructions courtes et longues, respectivement.

VI.2 Comme pour HIV-1 BRU, l'utilisation du site A3 des ARN de HIV-1 YBF30 et de SIVcpz ANT est fortement activée par la protéine SC35

Nous avons focalisé notre étude de l'épissage des ARN des virus HIV-1 YBF30 et SIVcpz ANT sur les sites A3 à A5. Nous avons adopté la même démarche expérimentale que pour les souches HIV-1 BRU et MAL : des constructions génétiques ont été réalisées, en couplant soit le site A3 seul, soit l'ensemble des sites A3 à A5 des ARN HIV-1 YBF30 et SIVcpz ANT au site donneur D1 de l'ARN



HIV-1 BRU (Fig. VI.2).

Les constructions contenaient donc la même partie 5' de la construction L3-U1 de HIV-1 BRU. Les constructions longues SIV-U1 et YBF30-U1 renfermaient l'ensemble de l'exon A3-D4 et le site de fixation de la snRNP U1 en aval du site D4. Comme par rapport à l'ARN HIV-1 BRU, les séquences des ARN de HIV-1 YBF30 et SIVcpz ANT présentent des délétions en aval du site A3, les constructions courtes SIV.5276 et YBF30.5427 ont été conçues de façon à conserver la taille de l'exon 2.

Les séquences des constructions SIV-U1, SIV.5276, YBF30-U1 et YBF30.5427 ont été amplifiées par PCR à partir de fragments d'ADNc qui nous avaient été fournis par Sentob Saragosti (INSERM U 552, Paris) (cf chapitre VIII de la partie matériel et méthodes). Des ARN uniformément marqués ont été produits à partir des 6 constructions représentées dans les panneaux A et B de la figure VI.2, puis soumis à un épissage en extrait nucléaire, avec ou sans ajout de l'une des 3 protéines SR recombinantes SC35, 9G8 et ASF/SF2.

L'épissage de l'ARN YBF30-U1 est aussi faible en extrait nucléaire que l'ARN L3-U1, contrairement à ce qu'on observe pour l'ARN SIV-U1 (Fig. VI.2-A). Ainsi, pour cet ARN, l'ordre d'efficacité d'épissage est : A3>>A4a≈A5 (piste 7), alors que pour L3-U1, l'ordre est : A5>A3>>A4a, b, c (piste 2). Ceci pourrait refléter une plus faible, voire une absence totale d'efficacité des éléments ESS2p et ESS2 agissant sur le site A3 de l'ARN SIV-U1, comme le suggère l'alignement de séquences (cf Fig. VI.1).

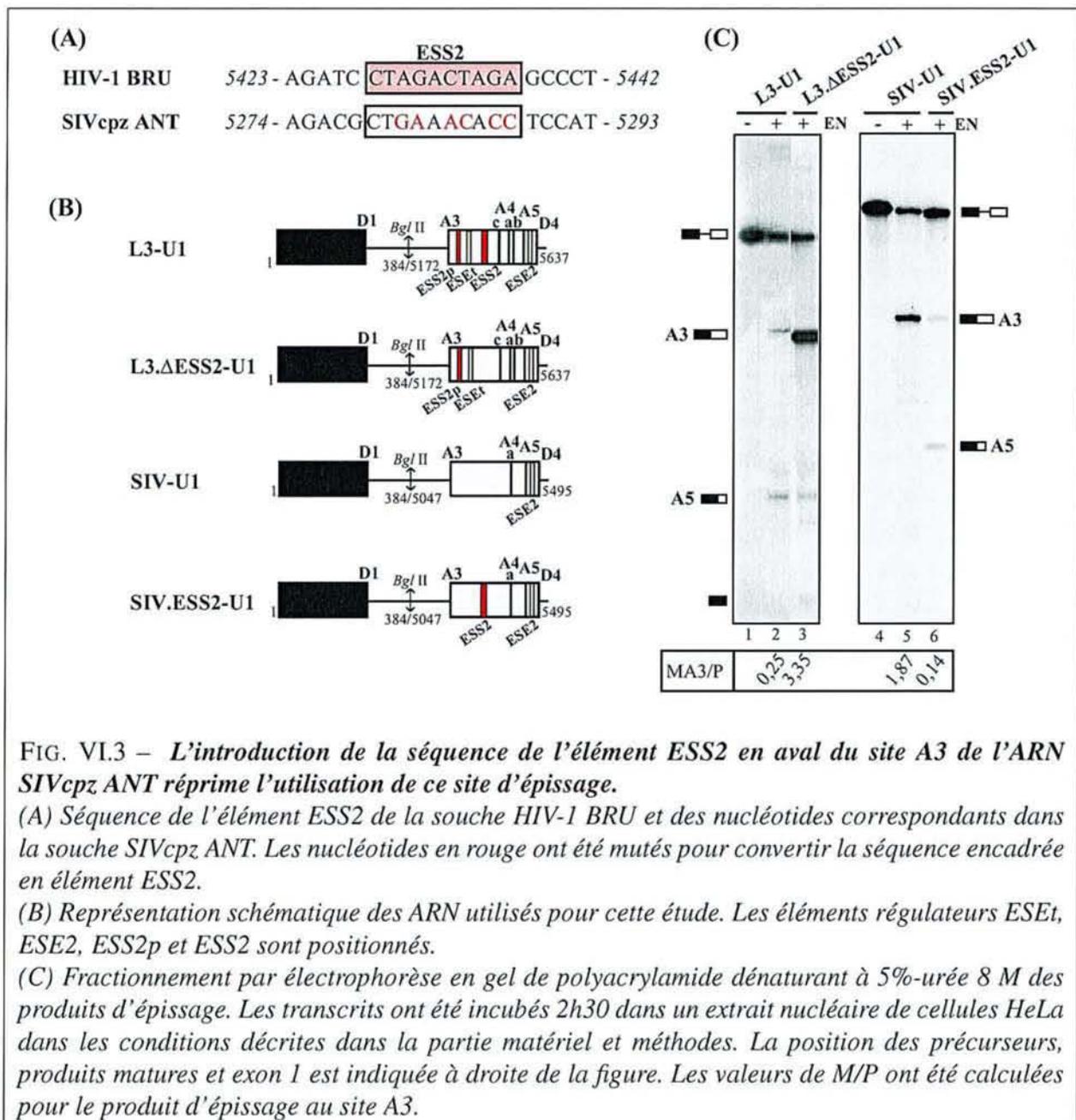
Comme pour la souche HIV-1 BRU, l'ajout des protéines ASF/SF2 et SC35 a des effets marqués sur l'efficacité du site A3, ceci aussi bien pour HIV-1 YBF30 que pour SIVcpz ANT (Fig. VI.2-A) : la protéine SC35 active fortement le site A3 (pistes 8, 13 vs 3) et la protéine ASF/SF2 a un effet activateur sur le site A3, plus fort dans les ARN SIV-U1 et YBF30-U1 que dans l'ARN L3-U1 (pistes 10, 15 vs 5). Quant à elle, la protéine 9G8 a un faible effet (pistes 9, 14 vs 4). Ainsi, en dépit des différences de séquences entre les souches, notamment en aval du site A3, celui-ci est toujours la cible de régulations par les protéines SR, SC35 étant un meilleur activateur qu'ASF/SF2.

Dans le contexte des constructions courtes, aucun épissage au site A3 de l'ARN SIV.5276 n'est détecté (Fig. VI.2-B, piste 7) et l'ajout des protéines SR ne permet pas de restaurer l'épissage à ce site (pistes 8, 9 et 10). La réaction semble bloquée après la 1^{re} étape d'épissage, puisqu'on détecte un produit en lasso et l'exon 1. Dans le cas de YBF30, l'épissage a bien lieu mais de manière surprenante, la protéine ASF/SF2 a un effet activateur plus fort que SC35 sur le site A3 au sein de l'ARN court (pistes 13 et 15), contrairement à ce qui est observé pour l'ARN C3 (pistes 3 et 5).

Il est à noter que 2 séquences AAGGA sont présentes en succession au niveau de l'élément ESEt de YBF30, ce qui renforce peut être l'efficacité de fixation d'ASF/SF2 par rapport à celle de SC35 sur l'élément ESEt (cf Fig. VI.1). Enfin, l'absence d'épissage dans le cas de SIVcpz pourrait être due à l'absence d'élément ESEt.

VI.3 L'absence d'élément ESS2 dans l'ARN de SIVcpz ANT explique sans doute la forte utilisation du site A3

Sur la base de l'alignement de séquence présenté dans la figure VI.1, nous pouvions expliquer la forte efficacité du site A3 dans l'ARN de la souche SIVcpz ANT par de nombreux changements au niveau de l'élément ESS2 (Fig. VI.3-A). D'autre part, cette efficacité était tout à fait comparable à celle observée dans l'ARN HIV-1 BRU, lorsque l'élément ESS2 était délété (mutant L3. Δ ESS2-U1 ; Fig. VI.3-B et C, piste 5 vs 3).



Afin de tester si la forte utilisation du site A3 dans l'ARN SIV-U1 était bien liée à l'absence d'élément ESS2, nous avons recréé par mutation une séquence similaire à celle de HIV-1 BRU dans l'ARN de la

souche SIVcpz ANT (Fig. VI.3-A). Les tests d'épissage montrent que l'efficacité du site A3 est alors fortement réduite et on observe une utilisation du site A5 (Fig. VI.3-C, piste 6). L'absence d'élément ESS2 dans l'ARN SIVcpz ANT explique donc vraisemblablement la forte efficacité du site d'épissage A3 dans l'ARN de cette souche. Dans cet ARN comme pour celui de BRU, on assiste à un équilibre entre l'utilisation des sites A3 et A5 : si le taux de A3 baisse, celui de A5 augmente.

VI.4 Comme pour HIV-1 BRU, le site donneur D4 est essentiel à l'utilisation du site A3 de SIVcpz ANT

Dans un 2^e temps, nous avons essayé de comprendre pourquoi le site A3 de l'ARN SIVcpz ANT n'est pas utilisé lorsqu'il est dans le contexte de la construction courte, alors qu'il l'est au sein de la construction longue. Une explication possible était que l'ARN de la construction longue renferme une séquence activatrice d'épissage, absente dans la construction courte, qui pourrait être l'élément ESE2 identifié par Kammler *et al.* (2001). Au moment où nous avons réalisé cette étude, l'élément ESE2 n'avait pas encore été mis en évidence. Nous avons donc réalisé une construction intermédiaire (SIV.5465), produisant un ARN renfermant 189 nucléotides de plus que l'ARN SIV.5276 et 30 nucléotides de moins que l'ARN SIV-U1 (Fig. VI.4-A et B).

Une autre possibilité était que le site D4 soit absolument indispensable pour qu'il y ait épissage efficace au site A3. Pour tester cette hypothèse, nous avons réalisé la construction SIV.ΔU1, permettant de produire un ARN où le site de fixation de la snRNP U1 en aval de D4 était éliminé (Fig. VI.4-B).

L'épissage *in vitro* de ces 2 ARN montre que l'élimination de la séquence 5466-5495 dans l'ARN SIV.5465 (Fig. VI.4-C, piste 6), comme l'élimination du site donneur D4 dans l'ARN SIV.ΔU1 (Fig. VI.4-C, piste 4) suffit à réduire très fortement l'utilisation du site A3.

Ainsi, il ressort des expériences d'épissage *in vitro* réalisées avec des ARN SIVcpz ANT de différentes tailles (SIV-U1, SIV.ΔU1, SIV.5465 et SIV.5276), que seule la présence du site donneur D4 permet l'utilisation du site d'épissage A3. Ce site joue donc, dans l'ARN de SIVcpz ANT, un rôle plus marqué pour l'activation du site A3 que dans l'ARN de HIV-1 BRU, sans doute en augmentant la définition de l'exon A3-D4.

VI.5 Etude expérimentale de la structure secondaire de la région de l'ARN du virus SIVcpz ANT contenant le site A3

S. Jacquenet avait montré au laboratoire que la structure secondaire de l'ARN du virus HIV-1 BRU modulait l'efficacité d'épissage (Jacquenet *et al.* 2001b ; cf article n° 2). Par ailleurs, il semble que la structure tige-boucle 3 renfermant l'élément ESS2 soit conservée dans les souches virales comportant cet élément, ceci sans doute en lien avec le mécanisme d'action de cet élément. Par conséquent, dans la mesure où l'élément ESS2 n'est pas présent dans l'ARN de SIVcpz ANT, on peut s'attendre à ce que la pression de sélection pour conserver la structure tige-boucle 3 n'existe pas dans le cas de SIVcpz ANT.

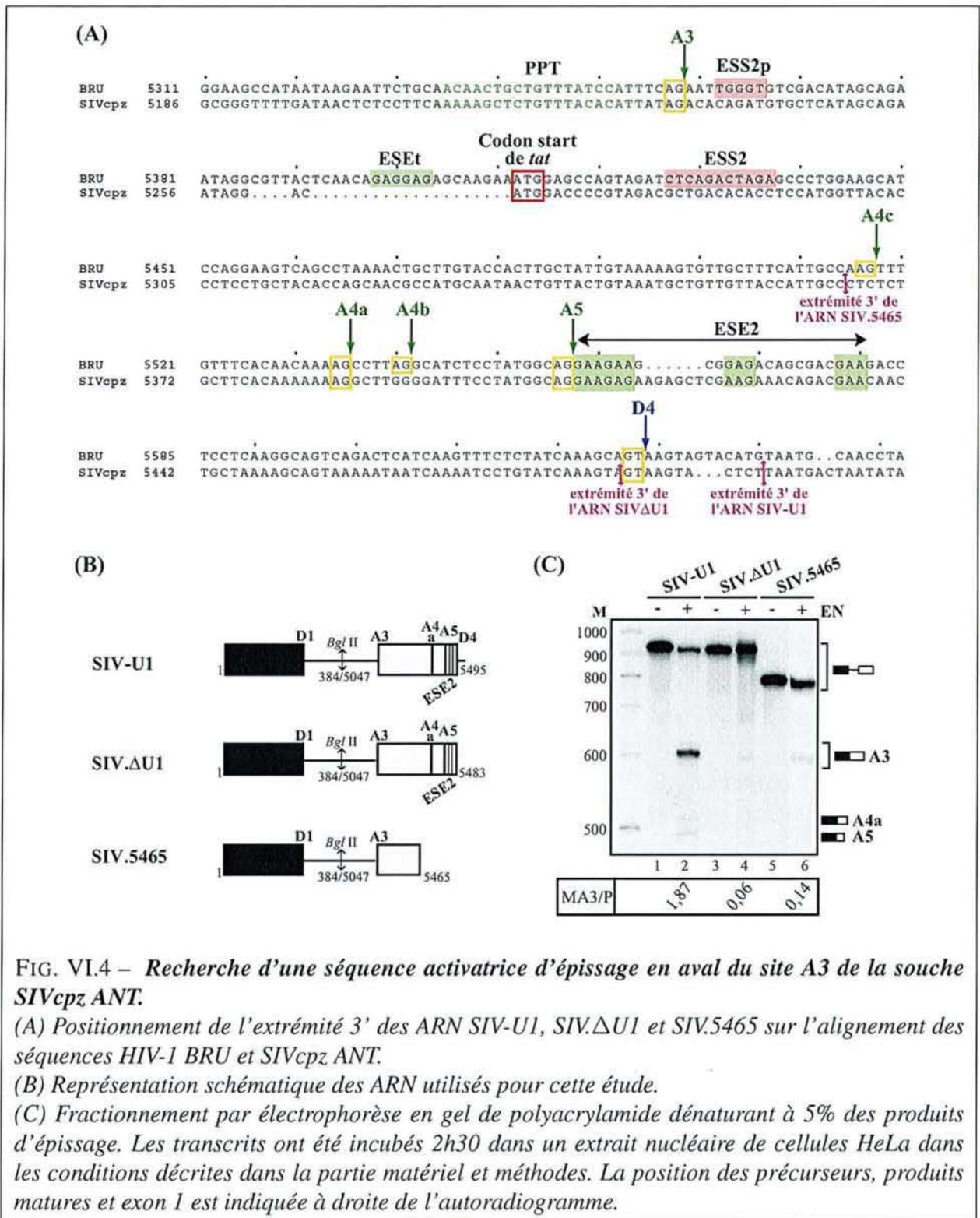


FIG. VI.4 – Recherche d'une séquence activatrice d'épissage en aval du site A3 de la souche SIVcpz ANT.

(A) Positionnement de l'extrémité 3' des ARN SIV-U1, SIV.ΔU1 et SIV.5465 sur l'alignement des séquences HIV-1 BRU et SIVcpz ANT.

(B) Représentation schématique des ARN utilisés pour cette étude.

(C) Fractionnement par électrophorèse en gel de polyacrylamide dénaturant à 5% des produits d'épissage. Les transcrits ont été incubés 2h30 dans un extrait nucléaire de cellules HeLa dans les conditions décrites dans la partie matériel et méthodes. La position des précurseurs, produits matures et exon 1 est indiquée à droite de l'autoradiogramme.

Par ailleurs, étant donné la divergence de séquence entre HIV-1 BRU et SIVcpz ANT, il n'était pas possible de prédire la structure secondaire de la région de cet ARN renfermant les sites A3 à A5. Il nous a donc semblé important, pour comparer les virus HIV-1 et SIVcpz ANT, d'établir expérimentalement la structure secondaire de cette région de l'ARN du virus SIVcpz. C'est ce que j'ai réalisé pour le segment compris entre les nucléotides 5164 et 5446 de l'ARN SIV-U1 (Fig. VI.5-A). Pour cela, j'ai utilisé les RNases V1, T1 et T2, et les sondes chimiques DMS et kéthoxal. Les 1^{ers} essais

ont été réalisés dans les conditions expérimentales utilisées pour l'étude de l'ARN de HIV-1 BRU, à savoir en présence de MgCl₂ 10 mM et de KCl 50 mM (Jacquet *et al.* 2001b). Dans ces conditions, l'ARN semblait adopter un ensemble hétérogène de conformations. Pour obtenir des résultats clairs, j'ai dû augmenter la concentration en ions monovalents. J'ai ainsi utilisé une concentration en KCl de 350 mM, qui avait déjà été utilisée dans des études antérieures au laboratoire (Mougin *et al.* 1996).

L'analyse pour l'établissement d'un modèle de structure secondaire a été possible pour le segment allant des nucléotides 5165 à 5260 (Fig. VI.5-B et C). On retrouve dans ce modèle les structures tige-boucles 1 et 2 établies dans le cas de l'ARN de HIV-1, avec le site A3 en boucle terminale et la séquence polypyrimidine correspondante dans l'hélice. Par contre, la structure de la tige-boucle 3 qui renferme l'élément ESS2 dans l'ARN HIV-1 n'est pas conservée, mais nous n'avons pas encore trouvé de modèle de repliement satisfaisant.

Des expériences d'empreintes réalisées sur la séquence de l'ARN SIVcpz ANT correspondant aux tige-boucles 2 et 3 de HIV-1 BRU (ARN SIV.sl24) ont mis en évidence une très forte protection de la séquence allant des positions 5275 à 5288 contre les coupures par la RNase T1 (Fig. VI.6). Cette séquence, qui est localisée dans une région simple-brin, constituerait donc un site de fixation de protéines nucléaires qui pourraient être importantes pour moduler le taux d'utilisation du site A3 de l'ARN SIVcpz ANT. Il sera intéressant de tester cette hypothèse par mutagenèse dirigée suivie de tests d'épissage *in vitro*.

VI.6 Conclusions et discussion des résultats

Notre étude comparative sur les 3 souches HIV-1 BRU, HIV-1 YBF30 et SIVcpz ANT montre que 3 des 5 sites A3, A4c, A4a, A4b et A5 sont conservés et utilisés dans les 3 virus. La présence des sites A4c et A4b pourrait être une acquisition récente dans cette famille de virus. Non seulement les 3 sites A3, A4a et A5 sont conservés, mais l'effet relatif des protéines SR sur ces 3 sites est conservé, à savoir, que le site A3 est la cible majeure des régulations, avec une activation par la protéine ASF/SF2 et surtout, par la protéine SC35. Le site A5 semble régulé par défaut.

Il est intéressant de constater qu'il y a conservation des effets des protéines SR, sans que l'on ait pour autant une conservation des mécanismes de régulation, en particulier, dans le cas de SIVcpz ANT. Ceci suggère fortement une grande importance pour la multiplication virale des effets des protéines SR sur le site A3. Lorsque la séquence virale est mutée, de nouveaux éléments de régulation sont sélectionnés afin de préserver l'effet des protéines SR. Bien que l'élément ESS2 de la souche YBF30 présente une substitution C→U en 5' du second motif de cet élément (CUAGAUUAGA au lieu de CUAGACUAGA), l'effet de l'élément ESS2 de la souche YBF30 semble similaire à celui de la souche BRU (cf Fig. VI.1). D'après nos données, la fixation de la protéine SC35 à cet élément doit aussi lever l'inhibition au site A3. Il sera donc intéressant de tester les capacités de fixation de SC35 à cette séquence ESS2 comportant une substitution C→U.

D'après la séquence de l'ARN de YBF30, un autre site potentiel de fixation de la protéine hnRNP A1 semble situé en amont de l'élément ESEt (positions 5394 à 5399, cf Fig. VI.1). Cette séquence correspond à la séquence consensus établie par l'approche SELEX pour la protéine hnRNP A1 (Burd

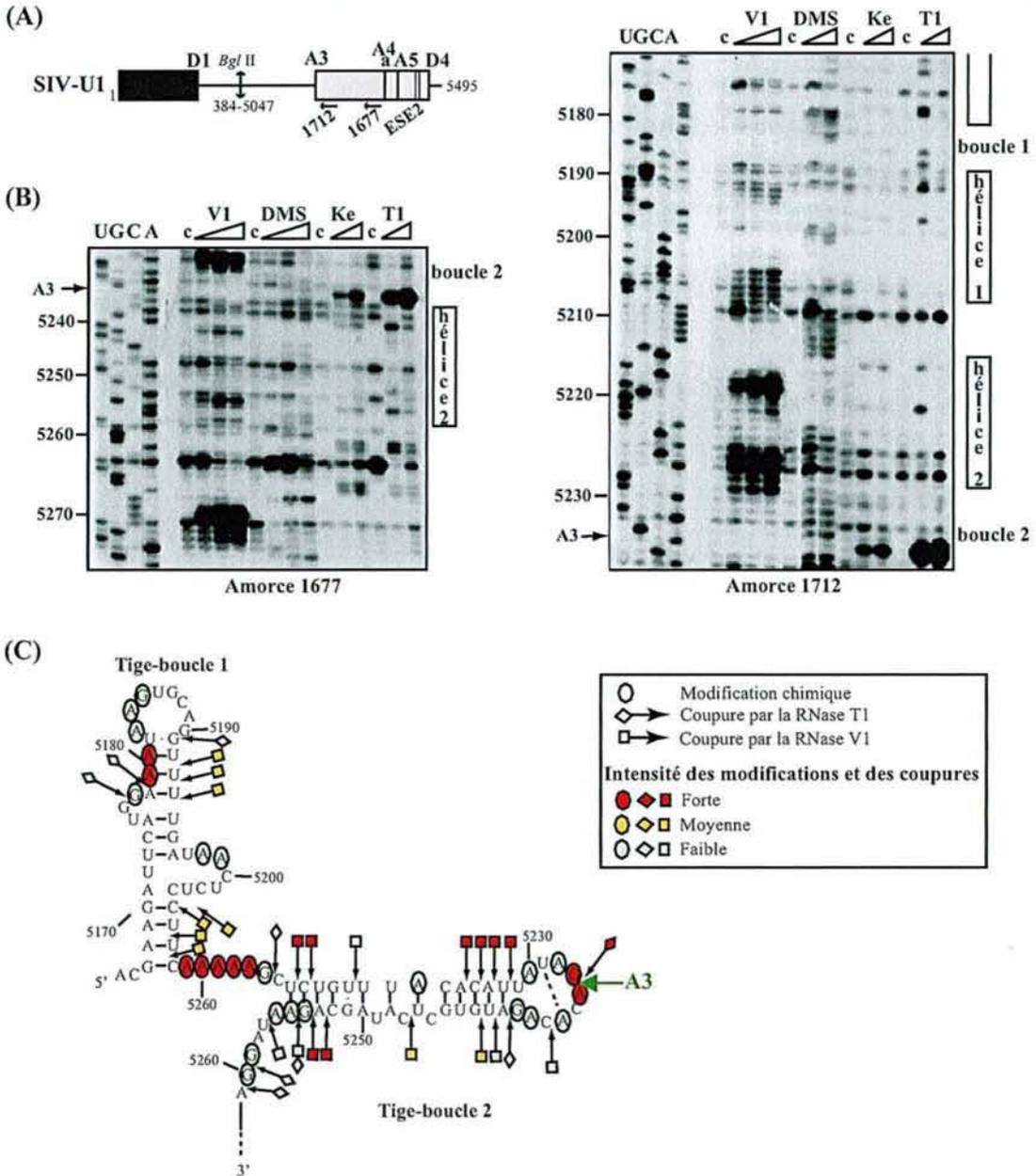


FIG. VI.5 – *Etude expérimentale de la structure secondaire de la région de l'ARN de SIVcpz ANT contenant le site A3.*

(A) Représentation schématique du transcrit SIV-U1 utilisé pour cette étude. L'élément ESE2 est symbolisé par un trait vert. La position des oligonucléotides 1677 et 1712, utilisés comme amorces pour les analyse par la transcriptase inverse, est indiquée.

(B) Exemples d'analyses de structure secondaire réalisées sur le transcrit SIV-U1 dans les conditions décrites dans la partie matériel et méthodes. Le transcrit a été soumis à des hydrolyses ménagées par des quantités croissantes de RNase VI (0,025, 0,05 et 0,1 U) et RNase T1 (0,05 et 0,1 U), et à des modifications chimiques par des quantités croissantes de DMS (0,5 et 1 µl de DMS 25%) et kéthoxal (1 et 2 µl de kéthoxal à 9,25 µg/µl). Les colonnes U, G, C, A correspondent au séquençage de l'ARN. La position des nucléotides au sein du génome de l'ARN SIVcpz ANT est indiquée à gauche des autoradiogrammes. Les ADNc ont été fractionnés sur gel de polyacrylamide dénaturant à 7%. L'identité de l'oligonucléotide amorce utilisé pour l'analyse est indiquée sous chaque autoradiogramme. La position des hélices identifiées est indiquée à droite des panneaux.

(C) Modèle de structure secondaire de la région de l'ARN du virus SIVcpz ANT contenant le site A3. La légende des couleurs est précisée dans un encart.

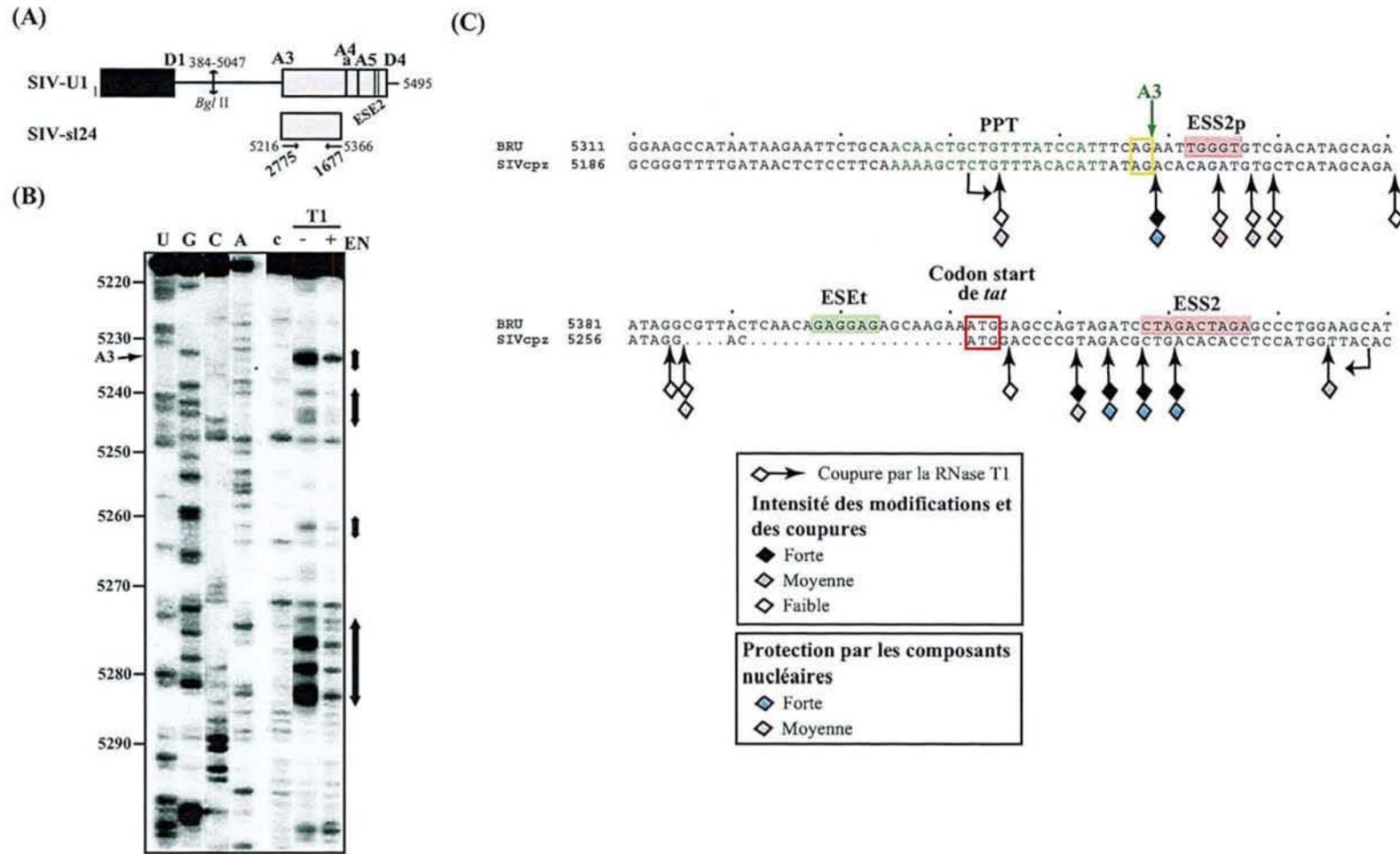


FIG. VI.6 – *Protection de l'ARN SIV-sI24 par les composants de l'extrait nucléaire de cellules HeLa contre l'action de la RNase T1.*

(A) *Représentation schématique du transcrit SIV-sI24. La matrice utilisée pour la transcription de cet ARN a été produite par amplification par PCR de la construction SIV-U1, à l'aide des oligonucléotides 2775 (portant un promoteur pour l'ARN SP6 polymérase) et 1677.*

(B) *Exemple d'analyse par transcription inverse du transcrit SIV-sI24 hydrolysé par 0,5 U de RNase T1, en absence (-) ou en présence (+) d'extrait nucléaire de cellules HeLa (noté EN). Les colonnes U, G, C, A correspondent au séquençage de l'ARN. La position des nucléotides au sein du génome de l'ARN SIVcpz ANT est indiquée à gauche des autoradiogrammes et les flèches, dont l'épaisseur est représentative de l'intensité de la protection, sont placées à droite.*

(C) *Représentation schématique des résultats de coupure à la RNase T1 de l'ARN SIV-sI24 dans le tampon D ou dans l'extrait nucléaire. La légende des couleurs est indiquée sous la figure. Les flèches cassées délimitent la région d'ARN analysée.*

& Dreyfuss 1994a). Il est possible que les protéines SR liées à l'élément ESEt limitent la fixation de la protéine hnRNP A1 sur ce site potentiel et vice versa. La modification en séquence de l'élément ESEt dans la souche YBF30 renforcerait peut-être la fixation d'ASF/SF2, ce qui assurerait une meilleure compétition avec la protéine hnRNP A1. Ceci pourrait expliquer la plus forte activation du site A3 par ASF/SF2 observée pour la construction courte YBF30.5427 (cf Fig. VI.2). On pourrait imaginer que l'efficacité d'inhibition par l'élément ESS2 de la souche YBF30 soit un peu plus faible, du fait qu'il ne possède qu'un seul motif CUAGA, mais que ceci soit compensé par la présence d'un autre site fort pour la protéine hnRNP A1, situé entre les positions 5394 à 5399. Il sera aussi intéressant de tester cette hypothèse par mutagenèse dirigée suivie de tests d'épissage *in vitro*.

Notre étude de structure secondaire montre la très forte conservation, dans un virus très éloigné de HIV-1 BRU (SIVcpz ANT), de la structure secondaire de la région renfermant les séquences fonctionnelles du site A3 : le site A3 en lui-même, la séquence polypyrimidine et la séquence refermant le point de branchement. Par contre, la structure secondaire des régions renfermant les éléments régulateurs de l'épissage semble moins conservée.

Il est à noter que la forte utilisation du site A3 due à l'absence d'élément ESS2 dans l'ARN SIVcpz ANT a aussi été observée pour les souches du groupe O de virus HIV-1 (Bilodeau *et al.* 1999). Pour ces virus, on peut s'attendre à une plus forte production de protéine Tat. Or la synthèse de cette protéine doit être régulée. Bilodeau *et al.* (1999) avaient proposé :

- soit que la protéine Tat, produite en plus grande quantité dans le groupe O (et dans SIVcpz ANT), soit moins active que celle du groupe M,
- soit que d'autres séquences dans le génome des souches du groupe O (et de SIVcpz ANT), distantes de l'exon A3-D4 et de ses séquences flanquantes, compensent l'absence d'élément ESS2.

En conclusion, sachant que les virus HIV-1 et SIVcpz ont un ancêtre commun, la régulation du site A3 par les protéines SR semble donc avoir été conservée au cours de l'évolution de ces virus. Les protéines SR constituent donc vraisemblablement des facteurs indispensables à la multiplication virale.

Chapitre VII

Modélisation mathématique de la régulation de l'utilisation des sites accepteurs d'épissage de l'ARN du virus HIV-1

L'étude expérimentale du site A3 de l'ARN du virus HIV-1 que nous avons présentée a révélé une régulation complexe de l'utilisation de ce site, dont l'efficacité d'épissage peut varier fortement, en fonction des concentrations relatives en protéines SR et protéine hnRNP A1 dans les cellules infectées. Sur la base des résultats expérimentaux de cette étude, nous avons établi une collaboration avec D. Eveillard et A. Bockmayr pour modéliser ce type de régulation (projet MODBIO, LORIA, Nancy). Notre objectif était de développer des outils mathématiques visant à modéliser les mécanismes de régulation de l'épissage, ceci en prenant comme référence, les régulations que nous avons identifiées pour l'ARN du virus HIV-1.

De manière générale, la démarche de modélisation nécessite une simplification du système. Tout en ayant soin d'être au maximum en accord avec la réalité biologique, nous avons commencé par établir un modèle simple, simulant la variation du taux d'épissage au site A3, en fonction des concentrations relatives en protéines ASF/SF2, SC35 et hnRNP A1. Par la suite, nous avons complexifié ce modèle, de façon à prendre progressivement en compte d'autres sites d'épissage. Toute cette étude en est encore à ses débuts et nécessite d'être validée expérimentalement, mais les premiers résultats obtenus sont encourageants.

Dans ce chapitre, nous ne décrivons pas en détails les modèles mathématiques qui ont été établis par l'équipe de modélisateurs, mais nous expliciterons l'ensemble des étapes qui ont abouti à la formalisation des connaissances biologiques. Nous expliciterons aussi le choix de nos hypothèses, partie du travail dans laquelle je me suis plus particulièrement investie.

VII.1 Introduction sur la modélisation mathématique des phénomènes biologiques et objectifs de l'étude

Selon sa construction et l'objectif qui lui est assigné, un modèle donne une vision simplifiée de la réalité. La simplification se limite alors généralement à une réduction d'échelle. Par exemple, une maquette (c'est-à-dire, un modèle réduit) de navire est construit afin de procéder à des essais de flottaison en bassin. En biologie, l'utilisation de modèles animaux ou cellulaires emprunte la même démarche. Ainsi, un modèle animal particulier peut être choisi pour la brièveté de son cycle de reproduction,

ce qui permet des expérimentations nécessitant un renouvellement rapide des générations. Le modèle peut être également une représentation simplifiée d'un objet ou d'un processus. Par exemple, en biochimie, le modèle de Michaëlis-Menten décrit simplement des cinétiques enzymatiques complexes. Une extension de ce modèle appliquée aux micro-organismes est le modèle de Monod, décrivant la croissance bactérienne, tout aussi complexe.

Les modèles sont utilisés à des fins diverses. Par exemple, certains modèles permettent de synthétiser les connaissances d'un système biologique, pour mettre en évidence d'éventuelles incohérences, tandis que des modèles prédictifs vont permettre le test d'hypothèses biologiques. Ces deux types de modèles sont complémentaires. Ils sont ensuite étudiés avec des méthodes de sensibilité permettant de tester les paramètres du système, ce qui permet ensuite au modélisateur de réaliser des simulations après validation, en adéquation avec les données à disposition.

Dans un premier temps, nous avons souhaité étudier l'évolution de l'efficacité d'épissage au site A3 en fonction des quantités de protéines ASF/SF2, SC35 et hnRNP A1. Pour cela, nous avons eu recours à une modélisation par contraintes, qui permettait l'étude de systèmes complexes. La stratégie était d'utiliser une modélisation mathématique pour décrire le processus étudié (la régulation de l'épissage). Ce type de démarche permet d'analyser des situations complexes, impliquant de multiples variables, difficiles à analyser sans une mise en équations du problème.

En se focalisant sur le site A3, notre objectif premier a été d'établir un modèle fournissant un cadre théorique, rendant compte des observations expérimentales et en accord avec les conclusions tirées de ces données.

VII.2 Modélisation de la régulation de l'épissage au site A3 de l'ARN du virus HIV-1

La modélisation d'un processus biologique implique de fortes interactions entre les mathématiques, l'informatique et la biologie. Le protocole de modélisation débute par l'énoncé des hypothèses sur le processus biologique. Celles-ci sont ensuite traduites en langage mathématique, dans lequel les variables ont une signification sur le plan biologique. On étudie ensuite le modèle du point de vue mathématique, puis on retraduit les résultats mathématiques en connaissances biologiques. La biologie intervient donc en entrée et en sortie de ce modèle. Par ailleurs, l'élaboration d'un modèle peut être vue comme un processus itératif, puisqu'à chaque étape on affine les hypothèses faites sur le processus biologique (pour revue, Pavé 1994).

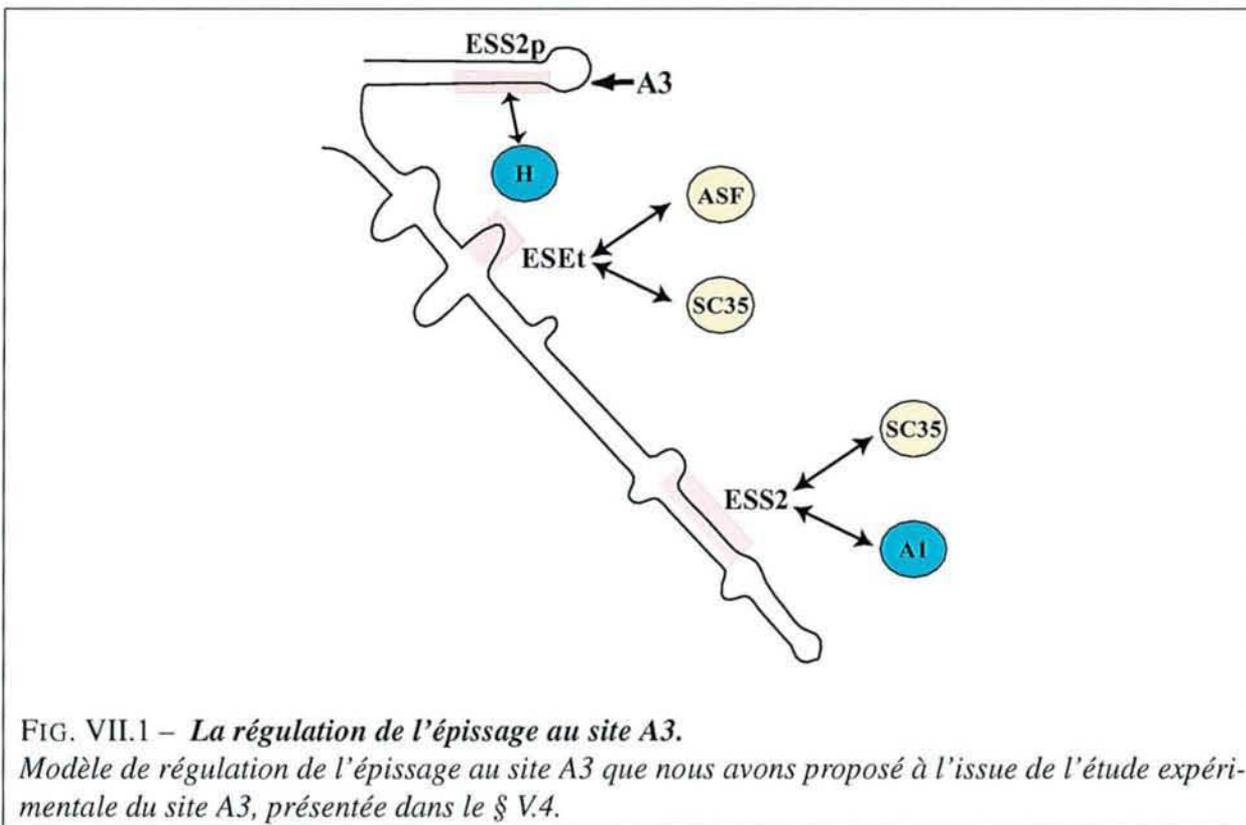
Dans un premier essai de modélisation, nous avons voulu établir un modèle mathématique rendant compte de la variation d'efficacité d'épissage au site A3 de l'ARN du virus HIV-1, en fonction des concentrations relatives des protéines SR ASF/SF2 et SC35 et de la protéine hnRNP A1. L'établissement de ce modèle repose sur les données obtenues lors de notre étude expérimentale du site A3, que nous avons présentée dans le § V.4, page 200.

VII.2.1 Résumé des connaissances sur le site A3

La figure VII.1 rappelle les connaissances sur la régulation du site A3, obtenues à l'issue de l'étude expérimentale. D'après le modèle de régulation proposé, l'utilisation de ce site apparaît comme dépendante de plusieurs paramètres :

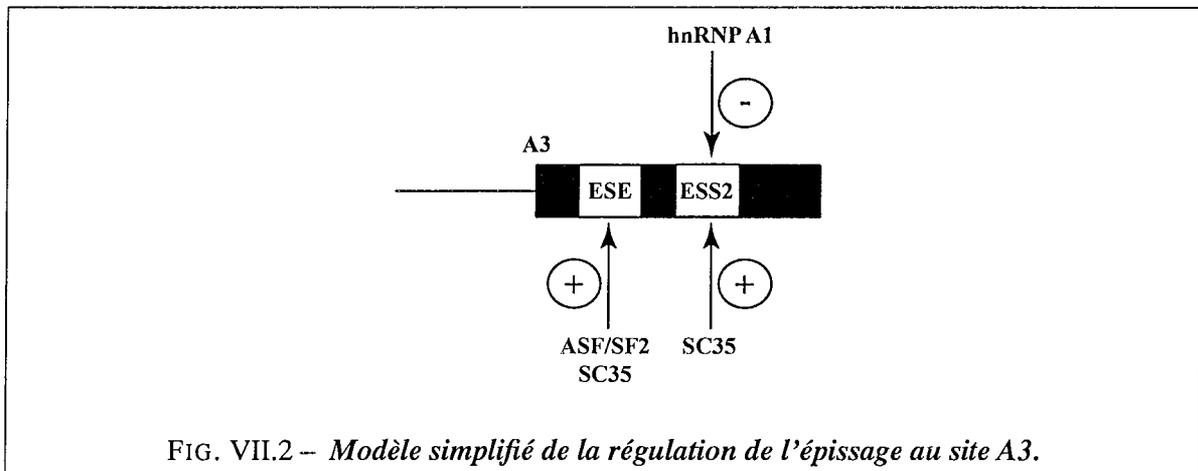
- la structure secondaire : le site A3 est en effet localisé dans une structure tige-boucle qui limite son accessibilité aux facteurs d'épissage et donc réduit son efficacité ;
- la présence de séquences inhibitrices d'épissage : deux éléments, ESS2p et ESS2, répriment le site A3 en fixant respectivement les protéines hnRNP H et hnRNP A1. La régulation exercée par l'élément ESS2p ne dépend pas des protéines SR ;
- la présence d'une séquence activatrice d'épissage : l'élément ESE1 active l'utilisation du site A3 en fixant les protéines ASF/SF2 et SC35 ;
- une compétition pour la fixation des protéines SC35 et hnRNP A1 à l'élément ESS2 : en effet, la protéine SC35 est également capable de se fixer à l'élément ESS2. Nous avons donc fait l'hypothèse que la fixation de la protéine SR à l'élément ESS2 empêche la fixation de la protéine hnRNP A1 et vice versa.

Il faut noter que la protéine SRp40 active également l'utilisation du site A3, mais nous ne connaissons pas son mode d'action. De plus, 10 protéines SR ont été identifiées chez l'homme, mais nous n'avons testé que l'effet de quatre d'entre elles.



Comme nous l'avons mentionné précédemment, un modèle est une représentation simplifiée de la réalité. Il s'agit d'une volonté délibérée, de la part du modélisateur, de faire abstraction de certains détails (par exemple, une carte d'état-major est un modèle réduit de la réalité environnante, dont les

détails tels que chaque arbre, caillou ou brin d'herbe ont été omis, afin d'avoir une vue synthétique des lieux représentés). Notre objectif était donc de représenter au mieux le processus biologique étudié, avec le minimum d'hypothèses. Pour cela, il fallait restreindre le référentiel en nous limitant à un modèle rendant compte des variations d'efficacité d'épissage au site A3, en fonction des quantités de protéines SR ASF/SF2 et SC35 et de la protéine hnRNP A1. Il était donc nécessaire de faire abstraction des paramètres de la régulation indépendants de ces deux types de protéines tels que, par exemple, la concentration en protéine hnRNP H se fixant sur l'élément ESS2p ou la concentration en facteurs spliceosomaux. Nous avons également omis l'effet activateur de la protéine SRp40, dont nous ne connaissons pas le mode d'action. La régulation de l'épissage au site A3 était alors décrite de façon simplifiée, selon le schéma de la figure VII.2.



VII.2.2 Hypothèses formulées sur le plan biologique

Nous voulons que le modèle mathématique rende compte de la dynamique de la régulation du site A3. Pour cela, il fallait s'intéresser à l'aspect mécanistique de notre système biologique. Dans ce but, nous avons décrit le rôle que nous attribuons à chaque élément régulateur du système, ceci sur la base de nos connaissances expérimentales et bibliographiques. Cela nous a amené à formuler des hypothèses au niveau biologique, sur lesquelles a reposé l'élaboration du modèle.

VII.2.2.1 Rôle de l'élément ESEt

Les protéines ASF/SF2 et SC35 activent l'épissage au site A3 en se fixant à l'élément ESEt. L'état de nos connaissances ne nous permettait pas de savoir si ces deux protéines se fixent de manière coopérative ou, au contraire, entrent en compétition pour se fixer à l'élément ESEt. Nous avons considéré qu'il n'y a pas de coopérativité lors de la fixation et que les 2 protéines sont capables de se compenser mutuellement, c'est-à-dire, que leur mode d'action est le même. En ce qui concerne ce mode d'action, il était vraisemblable qu'en se fixant à l'élément ESEt, ASF/SF2 et SC35 activent l'assemblage des complexes spliceosomaux, d'après les mécanismes de régulation décrits dans la revue bibliographique.

D'après notre étude expérimentale du site A3, l'élément ESEt semble agir de façon indépendante de l'élément ESS2 car, d'une part, cet élément est seul responsable de l'activation du site A3 par

ASF/SF2, d'autre part, la délétion de l'élément ESS2 n'empêche pas l'activation du site A3 par la protéine fixée à l'élément ESEt et vice versa.

VII.2.2.2 Rôle de l'élément ESS2

Le mode d'inhibition du site A3 par la protéine hnRNP A1 fixée à l'élément ESS2 n'est pas connu, mais d'après les mécanismes d'action de cette protéine, décrits dans la revue bibliographique, elle empêche vraisemblablement la reconnaissance du site d'épissage par les facteurs spliceosomaux, ce qui bloque l'assemblage du spliceosome, et conduit donc à une inhibition à la fois de la 1^{re} et de la 2^e étape d'épissage.

Enfin, d'après notre étude expérimentale du site A3, la protéine SC35 empêcherait la fixation de la protéine hnRNP A1 à l'ARN, ce qui lèverait l'inhibition exercée par la protéine et conduirait à l'activation indirecte du site A3.

VII.2.2.3 Conceptualisation de la régulation du site A3

Sur la base des hypothèses formulées, nous avons pu représenter la dynamique de régulation du site A3 selon le schéma de la figure VII.3. Nous avons pris en compte les 2 étapes de la réaction d'épissage, au niveau desquelles l'assemblage du spliceosome est la cible des régulations par les protéines ASF/SF2, SC35 et hnRNP A1. Le fait que chaque étape d'épissage soit décrite comme une réaction réversible est une astuce que nous avons utilisée pour la modélisation, mais qui ne représente pas une réalité biologique, notamment au niveau de la 2^e étape d'épissage.

Le taux de produits de la 1^{re} et de la 2^e réaction sont affectés par la régulation, mais le rapport des quantités d'ARN messager et intermédiaire est fixe et indépendante de la régulation. Notre modèle prenait donc en compte 3 espèces d'ARN : l'ARN prémessager, les intermédiaires réactionnels de la 1^{re} étape et l'ARN messager mature, dont les concentrations varient au cours de la réaction. Il s'agissait d'un modèle par compartimentation.

Dans notre démarche de modélisation, nous nous sommes placés dans le même contexte que dans nos expériences d'épissage *in vitro*, dans lesquelles les études étaient réalisées sur un seul site accepteur ou quelques sites accepteurs en compétition. Nous avons en plus considéré que notre système était isolé, sans échanges extérieurs.

Enfin, dans nos expériences d'épissage *in vitro*, nous avons ajouté un excès de l'une des protéines SR à la réaction, sachant que ces protéines sont déjà présentes naturellement dans l'extrait nucléaire. Cet extrait renferme par ailleurs une forte quantité de protéines hnRNP A1. Nous avons donc considéré que du fait de ce large excès de protéines ASF/SF2, SC35 et hnRNP A1, et malgré leur fixation à l'ARN au cours de la réaction d'épissage, la quantité totale de protéines régulatrices était constante.

Lors de nos expériences d'épissage *in vitro*, nous avons traduit l'efficacité d'épissage d'un site accepteur par le rapport M/P (quantité d'ARNm par rapport à la quantité d'ARN prémessager restant). Nous avons donc utilisé des valeurs M/P que nous avons obtenues lors d'expériences d'épissage *in vitro* de l'ARN L3-U1, réalisées dans 8 μ l d'extrait nucléaire de cellules HeLa (M/P en l'absence de protéine SR ajoutée = 0,20).

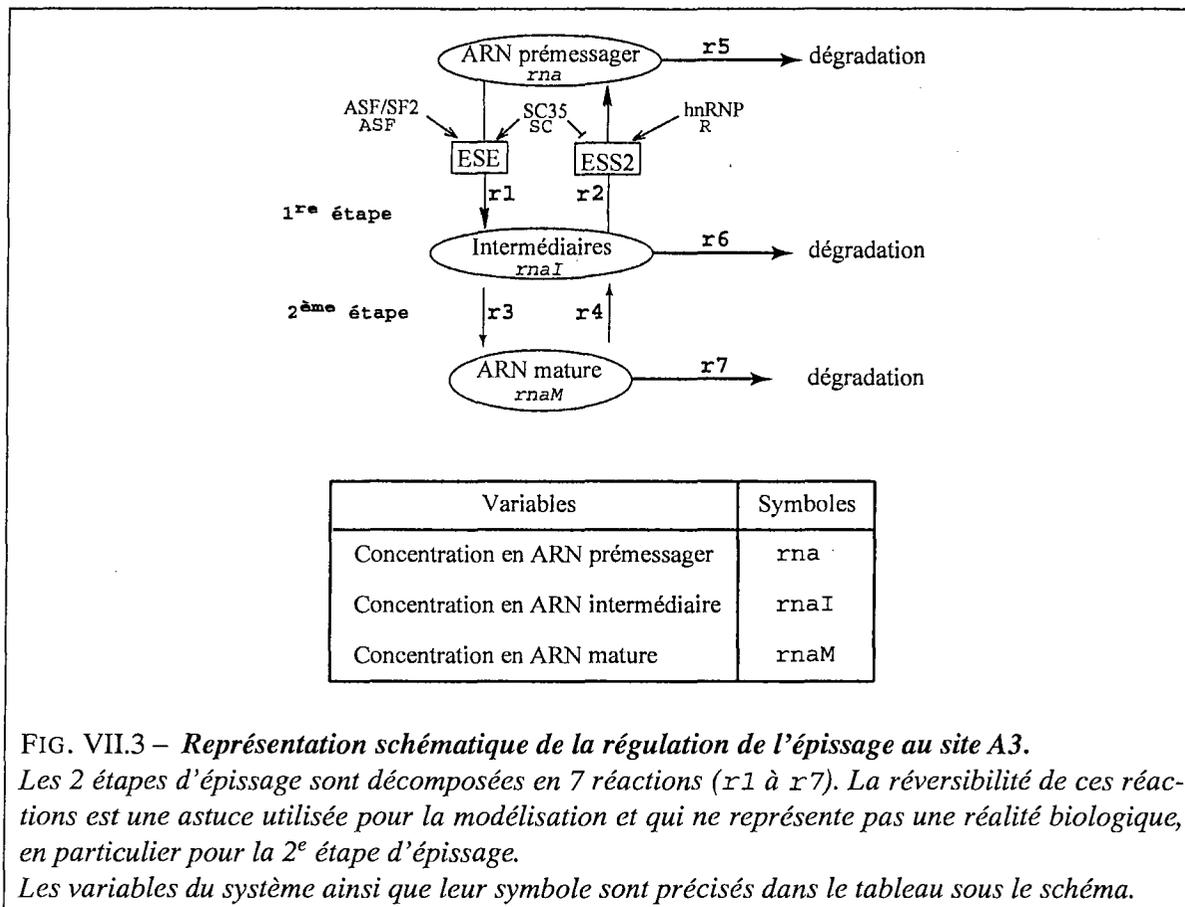


FIG. VII.3 – Représentation schématique de la régulation de l'épissage au site A3.

Les 2 étapes d'épissage sont décomposées en 7 réactions ($r1$ à $r7$). La réversibilité de ces réactions est une astuce utilisée pour la modélisation et qui ne représente pas une réalité biologique, en particulier pour la 2^e étape d'épissage.

Les variables du système ainsi que leur symbole sont précisés dans le tableau sous le schéma.

La quantité d'ARN pré-messager au départ de la réaction a été estimée à $0,006 \mu\text{M}$ d'ARN pré-messager L3-U1, dans nos expériences d'épissage *in vitro*. Nous avons eu par contre des difficultés pour connaître la quantité totale des protéines hnRNP A1, ASF/SF2 et SC35 présentes dans nos réactions. Il fallait, en effet, pouvoir estimer la quantité de ces protéines régulatrices présentes au départ dans l'extrait nucléaire, avant l'ajout d'un excès de protéine ASF/SF2 ou SC35. Nous avons essayé d'attribuer des valeurs pour les concentrations de ces différentes protéines, sur la base de données de la littérature. Ainsi, d'après Michael *et al.* (1995a), la protéine hnRNP A1 est présente à $10 \mu\text{M}$ dans les extraits nucléaires de cellules HeLa, ce qui représentait une concentration finale de $3,5 \mu\text{M}$, dans nos conditions expérimentales. Une étude réalisée par Mayeda & Krainer (1992) avait permis d'évaluer à 2 le rapport des quantités de protéines hnRNP A1 et ASF/SF2 dans les cellules HeLa, ce qui portait à $1,75 \mu\text{M}$ la concentration en protéine ASF/SF2 dans l'extrait nucléaire. Enfin, nous n'avions pas de données disponibles pour la protéine SC35, nous avons donc fixé arbitrairement sa concentration à $2 \mu\text{M}$.

VII.2.3 Formalisation du modèle

La modélisation d'un processus biologique est en général obtenue en se basant sur une analogie existant avec un système présentant des caractères de similitude, et dont les lois sont mieux connues. Par exemple, la comparaison du cœur avec une pompe a permis de prouver que la circulation sanguine relevait des lois de l'hydraulique. Dans notre cas, aucun modèle mathématique sur l'épissage n'avait

encore été établi, nous avons donc du chercher un système qui pourrait rendre compte de la régulation étudiée.

Une représentation formelle de l'intervention des protéines ASF/SF2, SC35 et hnRNP A1 était une cinétique de transfert, inspirée des taux de croissance limites (pour revue, Pavé 1994). Dans ce cas, les concentrations des protéines régulatrices ASF/SF2, SC35 et hnRNP A1 étaient les facteurs limitant du transfert. La cinétique de chaque réaction dépendait donc des quantités d'ARN, des quantités de protéines régulatrices et de leur affinité pour les éléments ESEt et ESS2. Ce mode de représentation formelle avait l'avantage de tenir compte d'un possible effet de saturation des éléments ESEt et ESS2 par les protéines régulatrices. Le tableau VII.1 résume l'ensemble des variables et des paramètres du système. Pour une meilleure lisibilité des équations, les variables de concentration ont été écrites sans crochets.

Symboles	Variables et paramètres	détermination de la valeur	unités
rna	Concentration en ARN pré-messager	expérimentale	μM
$rnaI$	Concentration en ARN intermédiaire		μM
$rnaM$	Concentration en ARN messenger		μM
ASF	Concentration en protéine ASF/SF2	bibliographique	μM
SC	Concentration en protéine SC35	arbitraire	μM
R	Concentration en protéine hnRNP A1	bibliographique	μM
φ_{ESE}	Vitesse maximale de fixation des protéines ASF/SF2 et SC35 pour l'élément ESEt	méthode de sensibilité	s^{-1}
φ_R	Vitesse maximale de fixation de la protéine hnRNP A1	méthode de sensibilité	s^{-1}
k_{ESE}	Coefficient de demi-saturation de l'élément ESEt	méthode de sensibilité	μM
k_{SC}	Coefficient de demi-saturation de l'élément ESS2 par la protéine SC35	méthode de sensibilité	μM
k_R	Coefficient de demi-saturation de l'élément ESS2 par la protéine hnRNP A1	méthode de sensibilité	μM
κ	Vitesse de réaction	déterministe	s^{-1}
κ'	Vitesse de réaction	déterministe	s^{-1}
λ	Coefficient de dégradation	arbitraire	s^{-1}

TAB. VII.1 – *Variables biologiques et paramètres du système.*

Le mode de détermination des variables et paramètres est précisé dans chaque cas. Les valeurs sont soit obtenues par des méthodes de sensibilité, soit issues de l'expérience, de la bibliographie ou encore, fixées de manière arbitraire.

Dans la figure VII.3, nous avons représenté la régulation du site A3 par 7 réactions de type chimique, au cours desquelles les quantités d'ARN pré-messager, intermédiaire et messenger varient.

La 1^{re} réaction (r_1) représente la transformation de l'ARN pré-messager en ARN intermédiaire. Dans notre modèle, la cinétique de cette réaction dépend de la quantité d'ARN pré-messager et de la fixation des protéines ASF/SF2 et SC35 à l'élément ESEt. De plus, nous considérons que les protéines ASF/SF2 et SC35 se compensent mutuellement. Nous pouvons donc écrire :

$$r_1 = \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna$$

D'après notre modèle conceptuel de régulation, la réaction r_2 représente la transformation de l'ARN intermédiaire en ARN prémessager. Dans cette étape, la protéine SC35 contrecarre l'action de la protéine hnRNP A1. La cinétique de la réaction dépend donc des concentrations relatives en protéines hnRNP A1 et SC35, ainsi que de la quantité d'ARN intermédiaire ($rnaI$) :

$$r_2 = \frac{\varphi_R \times R}{k_R \left(1 + \frac{SC}{k_{SC+R}}\right)} rnaI$$

La réaction r_3 représente la transformation de l'ARN intermédiaire en ARN mature. Comme nous considérons la cinétique de réaction du 1^{er} ordre, la réaction ne va dépendre que de la quantité d'ARN intermédiaire et d'une constante κ . Le principe est le même pour la réaction inverse, au cours de laquelle l'ARN mature est transformé en ARN intermédiaire. Nous pouvons donc écrire :

$$\begin{aligned} r_3 &= \kappa \cdot rnaI \\ r_4 &= \kappa' \cdot rnaM \end{aligned}$$

Les réactions r_5 , r_6 et r_7 correspondent à la dégradation naturelle des ARN prémessager, intermédiaire et mature, respectivement. Comme nous n'avons aucune donnée sur l'importance du phénomène de dégradation de chaque espèce d'ARN au cours de la réaction d'épissage, nous avons considéré, pour les 1^{ers} essais de modélisation, que la cinétique de dégradation est dépendante des quantités d'ARN et d'un facteur de dégradation λ . Sur ce point, nous ne sommes pas en parfait accord avec la réalité biologique. Nous avons donc :

$$\begin{aligned} r_5 &= \lambda \cdot rna \\ r_6 &= \lambda \cdot rnaI \\ r_7 &= \lambda \cdot rnaM \end{aligned}$$

Sur la base du modèle conceptuel présenté dans la figure VII.3, nous pouvons ensuite décrire les variations des quantités d'ARN au cours de la réaction d'épissage. Ainsi, la quantité d'ARN prémessager diminue par transformation de cet ARN en ARN intermédiaire (réaction r_1) et par dégradation (réaction r_5); dans ce modèle, elle augmente par transformation de l'ARN intermédiaire en ARN prémessager (réaction r_2), ce qui peut s'écrire de la manière suivante :

$$\frac{d(rna)}{dt} = r_2 - r_1 - r_5$$

De la même façon, selon le schéma conceptuel, la quantité d'ARN intermédiaire augmente par la transformation de l'ARN prémessager en ARN intermédiaire (r_1) et de l'ARN messenger en ARN intermédiaire (r_4). Elle diminue par transformation de l'ARN intermédiaire en ARN prémessager (r_2) et en ARN mature (r_3), et par dégradation (r_6). Autrement dit :

$$\frac{d(rnaI)}{dt} = r_1 + r_4 - r_2 - r_3 - r_6$$

Nous avons décrit sur le même principe la variation de concentration de la quantité d'ARN messenger :

$$\frac{d(rnaM)}{dt} = r_3 - r_4 - r_7$$

Au final, nous aboutissons donc à un système de 3 équations différentielles ordinaires (ODE), décrivant les variations des quantités d'ARN au cours de la réaction d'épissage, en fonction des quantités de protéines régulatrices hnRNP A1, ASF/SF2 et SC35 :

$$\begin{aligned} \frac{d(rna)}{dt} &= \frac{\varphi_R \times R}{k_R(1 + \frac{SC}{k_{SC+R}})} rnaI - \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna - \lambda \cdot rna \\ \frac{d(rnaI)}{dt} &= \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna - \frac{\varphi_R \cdot R}{k_R(1 + \frac{SC}{k_{SC+R}})} rnaI - \kappa \cdot rnaI \\ &\quad + \kappa' \cdot rnaM - \lambda \cdot rnaI \\ \frac{d(rnaM)}{dt} &= \kappa \cdot rnaI - \kappa' \cdot rnaM - \lambda \cdot rnaM \end{aligned}$$

VII.2.4 Comportement qualitatif du modèle

Une fois le modèle établi, il nous fallait vérifier que son comportement, autrement dit, ses propriétés qualitatives, étaient en accord avec la connaissance biologique que nous avons sur la régulation du site A3.

Comme dans nos expériences d'épissage *in vitro*, nous avons traduit l'efficacité d'épissage d'un site accepteur par le rapport M/P (quantité d'ARN messenger par rapport à la quantité d'ARN prémessenger restant), nous avons décrit de la même façon l'efficacité d'épissage au site A3 (*eff*) au cours de la réaction d'épissage :

$$eff(t) = \frac{rnaM(t)}{rna(t)}$$

Dans nos expériences, il était vraisemblable qu'après une phase transitoire, l'efficacité d'épissage du site A3 atteigne un équilibre, ce qui nous permettait de déterminer sa valeur. A l'équilibre, nous avons donc :

$$\begin{aligned} r_1 &= r_2 \\ r_3 &= r_4 \end{aligned}$$

et nous pouvons alors décrire l'efficacité d'épissage au site A3 à l'équilibre de la manière suivante :

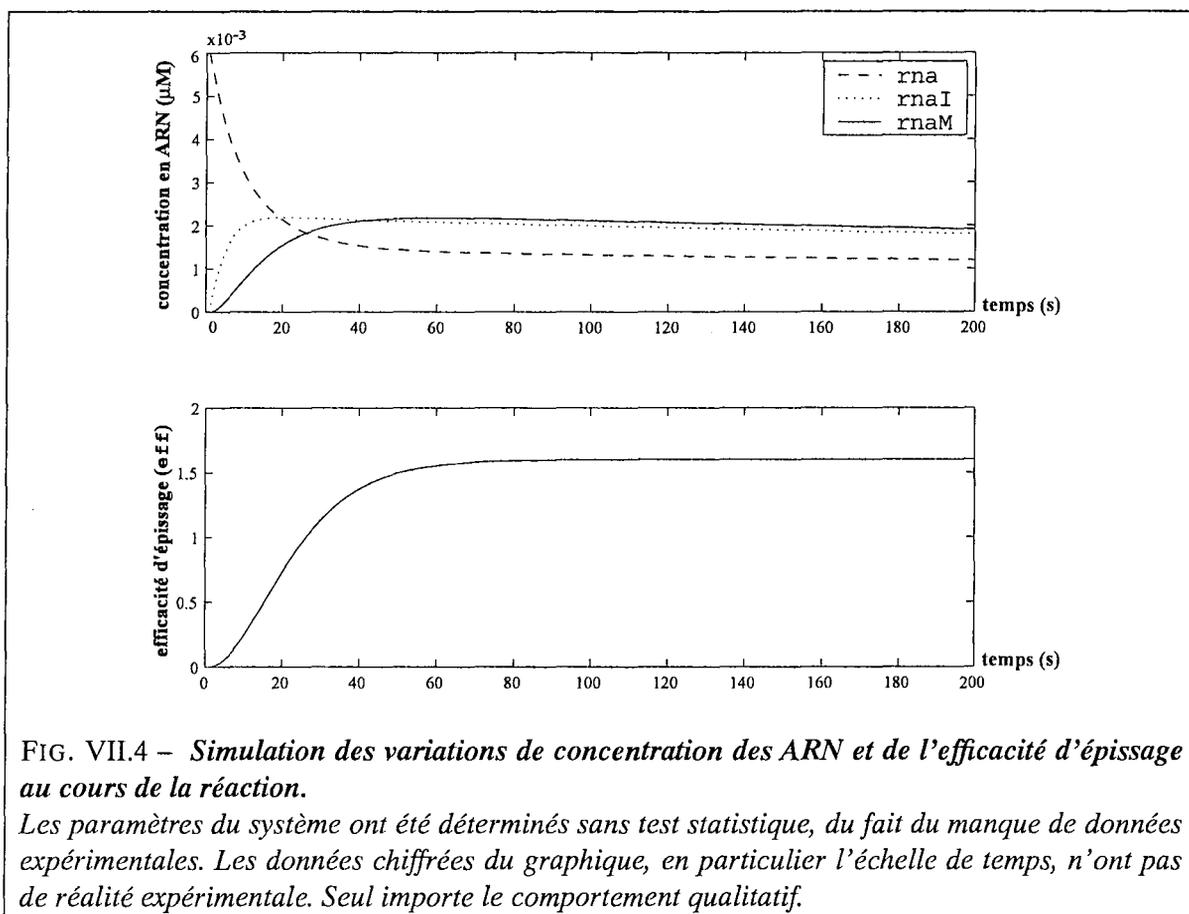
$$eff_{eq} = \frac{\kappa \cdot \varphi_{ESE}(ASF + SC)(k_R \cdot k_{SC} + k_R \cdot SC + R \cdot k_{SC})}{\kappa'(k_{ESE} + ASF + SC) \cdot \varphi_R \cdot R \cdot k_{SC}}$$

L'analyse de la formule indique que l'efficacité d'épissage à l'équilibre est une fonction croissante des concentrations en protéines ASF/SF2 et SC35 et une fonction décroissante de la concentration en protéine hnRNP A1, ce qui est en accord avec les données expérimentales que nous avons obtenues sur le site A3. Le comportement qualitatif à l'équilibre du modèle était donc cohérent avec les observations expérimentales.

VII.2.5 Etude des propriétés mathématiques du modèle et simulations

Il fallait ensuite étudier les propriétés mathématiques du modèle et pouvoir réaliser des simulations. Il était donc nécessaire de transcrire le modèle mathématique dans un langage qui fait appel à la programmation par contraintes (Bockmayr & Courtois 2002). Un des avantages du modèle est que l'efficacité d'épissage peut être assimilée à une variable observateur, qui rend compte de l'état du système.

Avant de réaliser des simulations, il était nécessaire de déterminer les paramètres cinétiques du système par des méthodes de sensibilité et d'étudier les propriétés qualitatives du modèle. Ces études ont été réalisées par D. Eveillard, qui a ensuite pu réaliser une simulation de la variation des quantités d'ARN pré-messager, intermédiaire et mature, et de l'efficacité du site A3, au cours de la réaction d'épissage. Comme nous manquons de données expérimentales pour réaliser cette simulation, les valeurs des paramètres du système devaient être fixées arbitrairement. Les résultats obtenus sont présentés dans la figure VII.4. Les données biologiques étant trop restreintes, nous ne pouvons qu'analyser sur le plan qualitatif les variations des quantités d'ARN et de l'efficacité d'épissage.



Les variations de concentration et d'efficacité observées au cours de la réaction d'épissage correspondent à ce qu'on observe *in vitro*. Ainsi, d'après les données présentées dans la figure VII.4, les différentes quantités d'ARN et l'efficacité d'épissage passent par une phase de transition avant d'atteindre un équilibre. La quantité d'ARN pré-messager diminue fortement dans un 1^{er} temps, puis

diminue plus progressivement. L'ARN intermédiaire est la 1^{re} espèce d'ARN synthétisée, puis apparaît l'ARN messager. Cette variation des quantités d'ARN correspond à ce que l'on observe dans nos réactions d'épissage *in vitro*.

Sur la base des hypothèses que nous avons formulées dans le § VII.2.2, nous avons pu réaliser un 1^{er} modèle mathématique continu, dont la simulation concorde avec nos observations expérimentales. Le fait que, du point de vue mathématique, nous puissions simuler l'efficacité d'épissage au site A3, simplement sur la base de la régulation exercée par les protéines hnRNP A1, ASF/SF2 et SC35 sur ce site, souligne la cohérence des hypothèses que nous avons formulées sur le mécanisme d'action de ces protéines, et renforce l'intérêt, sur le plan expérimental, d'approfondir notre étude sur les mécanismes de régulation du site A3. En particulier, compte-tenu du fort effet activateur de la protéine SRp40, à la fois dans nos expériences d'épissage *in vitro* et *ex vivo*, il sera important d'inclure la protéine SRp40 dans le modèle, dès que son mécanisme d'action sera connu.

Il est certain que cette modélisation n'en est encore qu'à ses débuts et qu'elle est perfectible. Elle a encore besoin d'être affinée et le modèle doit être validé quantitativement. Mais pour cela, des données expérimentales supplémentaires seront nécessaires. Il faut en effet pouvoir déterminer avec plus de précision les paramètres du système et à l'heure actuelle, nous manquons de données. Dans cet objectif, il serait important de réaliser de nouvelles expériences d'épissage *in vitro*, dans lesquelles nous ferons en particulier varier deux paramètres :

- la quantité de protéine SR ajoutée à l'extrait nucléaire. C'est ce que nous avons déjà réalisé dans notre étude expérimentale du site A3, mais la gamme de concentrations testée doit être élargie ;
- la cinétique. En effet, dans toutes les expériences, nous avons déterminé les efficacités d'épissage après 2h30 d'incubation. Il serait important de pouvoir suivre la variation de la valeur M/P au cours de la réaction d'épissage.

La cohérence des données de simulation, obtenues avec le modèle sur la base des hypothèses biologiques formulées, renforçait l'intérêt de prendre en compte d'autres sites d'épissage du virus. Nous avons donc, dans un 2^e temps, essayé d'élaborer un modèle plus global, qui prenne en compte les sites d'épissage A3 et A5.

VII.3 Extension du modèle mathématique à d'autres sites accepteurs de l'ARN du virus HIV-1

D'un point de vue biologique, il était intéressant d'élargir le modèle de régulation du site A3 en tenant compte de la présence d'autres sites en compétition, comme c'est le cas au sein du génome du virus HIV-1. Cette compétition joue un rôle important, comme nous l'avons montré expérimentalement. Nous avons donc commencé par élargir le modèle mathématique en incluant le site A5, puisque nous avons observé dans nos expériences un équilibre entre l'utilisation des sites A3 et A5.

Alors que nous avions une bonne connaissance de la régulation au site A3 par les protéines hnRNP A1, ASF/SF2 et SC35, nos données étaient moins nombreuses pour le site A5. Il nous semblait que

A5 était utilisé par défaut et que l'activation du site A3 par les protéines SR, en particulier, ASF/SF2 et SC35, permettait l'utilisation de A3 aux dépens de A5.

Sur la base de l'équilibre observé entre l'utilisation des sites A3 et A5, nous avons construit un modèle global, prenant en compte deux modèles locaux (Fig. VII.5) :

- le modèle de régulation du site A3, que nous avons présenté précédemment. Nous mesurons en sortie de ce modèle une efficacité d'épissage au site A3, dont la valeur dépend des concentrations relatives de protéines ASF/SF2, SC35 et hnRNP A1. Il s'agit d'un observateur, qui décrit l'état du système à l'échelle du site A3.
- le modèle de régulation du site A5. De façon similaire au site A3, la variable observateur en sortie de ce modèle traduit l'efficacité d'épissage au site A5. Sa valeur ne dépend pas des concentrations en protéines hnRNP A1, ASF/SF2 et SC35.

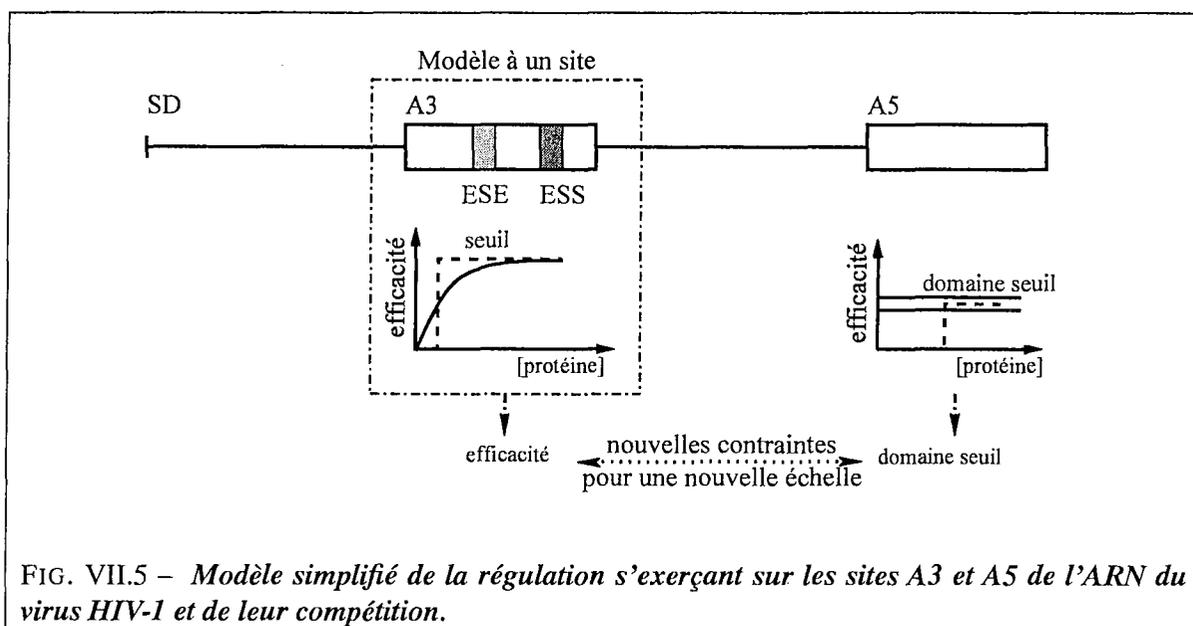


FIG. VII.5 – *Modèle simplifié de la régulation s'exerçant sur les sites A3 et A5 de l'ARN du virus HIV-1 et de leur compétition.*

Le principe de la modélisation a donc été le suivant : chaque modèle local, continu en temps, fonctionne indépendamment l'un de l'autre, et les observateurs correspondants (les efficacités d'épissage des sites A3 et A5) sont ensuite injectés dans le modèle plus global.

Dans ce modèle global, nous avons changé d'échelle, puisque nous considérons un niveau supérieur de complexité de la régulation, au niveau duquel nous décrivons l'utilisation par défaut du site A5 et sa compétition avec le site A3. Nous aboutissons dans ce cas à un modèle hybride. Dans ce modèle, le site A5 est préférentiellement utilisé lors de la réaction d'épissage, jusqu'à ce que la régulation du site A3 par les protéines ASF/SF2 et SC35 permette une utilisation plus importante de A3. Dans ce cas, le système bascule d'une utilisation préférentielle du site A5 vers celle de A3.

Afin de pouvoir gérer la notion de comportement par défaut du site A5, le modèle a été écrit, comme précédemment, dans un langage qui fait appel à la programmation par contraintes hybride (Bockmayr & Courtois 2002). Pour le modèle que nous voulons établir, l'avantage de ce type de programmation est qu'elle permet de gérer le manque d'informations en introduisant des contraintes, aussi bien au niveau du modèle local, que dans le modèle plus global. En particulier, l'utilisation des contraintes

permettait d'introduire la notion d'équilibre dans l'utilisation des sites A3 à A5.

Comme dans le cas du modèle établi pour le site A3, nous avons utilisé des données expérimentales d'efficacité d'épissage du site A5 de l'ARN L3-U1 (M/P en absence de protéine SR ajoutée=0,35). Elles proviennent de la même expérience d'épissage que les données utilisées précédemment pour le modèle au site A3. Les données obtenues pour A5 sont résumées dans le tableau ??.

Après étude des contraintes du système et détermination des valeurs des paramètres cinétiques, D. Eveillard a pu réaliser des simulations à partir du modèle global, qui sont présentées dans la figure VII.6 :

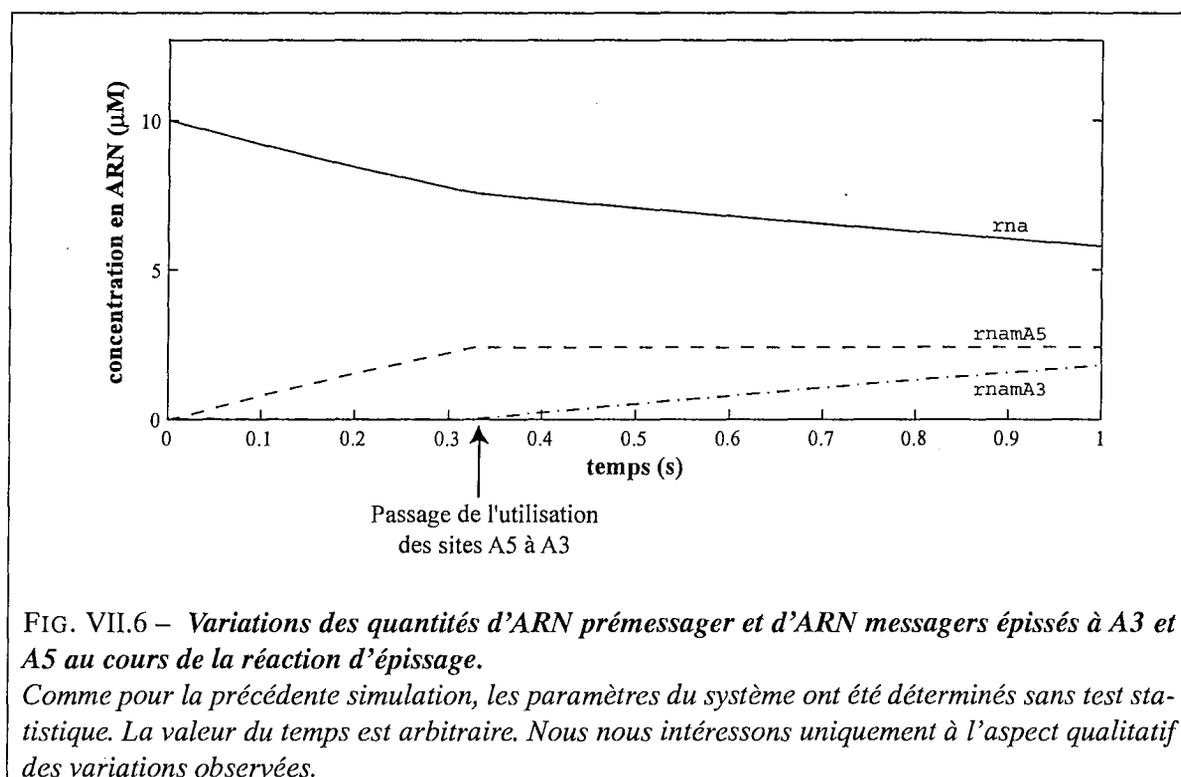


FIG. VII.6 – Variations des quantités d'ARN pré-messager et d'ARN messagers épissés à A3 et A5 au cours de la réaction d'épissage.

Comme pour la précédente simulation, les paramètres du système ont été déterminés sans test statistique. La valeur du temps est arbitraire. Nous nous intéressons uniquement à l'aspect qualitatif des variations observées.

Comme pour le modèle local du site A3 que nous avons présenté précédemment, les paramètres du système ont été déterminés de manière déterministe, par des méthodes de sensibilité, et par conséquent, nous ne pouvons qu'étudier le comportement qualitatif du modèle et essayer de juger s'il satisfait aux hypothèses.

D'après l'état actuel du modèle, la simulation obtenue des quantités d'ARN épissés aux sites A3 et A5 fait apparaître une transition au cours de la réaction, qui correspond au passage de l'utilisation du site A5 à A3. La quantité d'ARN pré-messager diminue ainsi en 2 temps. La 1^{re} espèce d'ARNm apparaissant est celui épissé au site A5 et sa quantité atteint rapidement un plateau. Le 2^e ARNm correspond à celui épissé au site A3. Son quantité croît pour se rapprocher de celle du site A5. Les simulations réalisées avec le modèle global sont donc cohérentes avec les observations que nous avons faites expérimentalement. Elles traduisent bien une utilisation plus forte du site A5, par rapport à celle de A3, en l'absence de protéine SR ajoutée.

VII.4 Conclusions

Dans notre démarche de modélisation, nous avons essayé de faire un modèle mathématique qui puisse expliquer au mieux nos observations expérimentales. Il est très simplifié pour l'instant. Des données expérimentales devront être obtenues pour progresser dans la qualité de la modélisation.

Dans un 1^{er} temps, nous avons essayé de modéliser la régulation qui nous paraissait la plus simple, à savoir, la compétition que nous avons observée au site A3 *in vitro*, entre les protéines SR (ASF/SF2, SC35) et la protéine hnRNP A1, dans l'ARN C3, comme dans l'ARN L3-U1 renfermant les sites A3 à A5. Nous observons pour cet ARN un équilibre entre l'utilisation des sites A3 et A5. Dans un 2^e temps, nous avons donc essayé de construire un modèle reflétant cet équilibre, tout en tenant compte de la régulation du site A3 par la protéine hnRNP A1 et les protéines SR.

Dans les deux cas, les progrès du modèle dépendront de la réalisation de gammes d'expériences visant à tester l'effet de concentrations croissantes des protéines régulatrices sur les efficacités d'épissage aux deux sites, et à définir les cinétiques d'épissage correspondantes. Les données ainsi obtenues devraient permettre d'affiner les paramètres du modèle et d'obtenir un outil de prédiction qui pourra être utilisé pour l'étude de la régulation d'autres sites d'épissage, pour lesquels on a aussi observé une compétition entre protéines SR et hnRNP A1 et compétition entre plusieurs sites d'épissage. Il faudra dans ce cas aussi disposer des résultats expérimentaux assez fournis pour définir avec précision les paramètres correspondant au système étudié. Mais comme l'outil mathématique est déjà construit, le travail sera fortement facilité.

Par ailleurs, lorsque les paramètres majeurs de la régulation de l'épissage de l'ARN du virus HIV-1 auront été complètement définis, il sera possible d'essayer de construire un modèle encore plus complet, incorporant tous ces paramètres. En particulier, les régulations complexes au site A7, déjà largement décryptées au laboratoire, et la régulation au site A2, dont nous avons débuté l'étude, devront être incluses dans le modèle, en plus des régulations aux sites A3 et A5.

Il est clair que vue la complexité des processus biologiques que nous étudions, les modèles ainsi établis ne peuvent pas traduire l'ensemble de la complexité. Par contre, lorsque les modèles auront atteint un degré supérieur d'optimisation, ils vont permettre de formuler et de tester des hypothèses sur l'influence relative de certains paramètres de cette complexité.

L'obtention de modèles affinés passe par l'étape initiale de recherche des meilleurs modes de modélisation. C'est cette étape à laquelle nous avons participé. Nous sommes conscients que le travail de modélisation est loin d'avoir atteint son terme. Il sera poursuivi entre le LORIA et le MAEM. Un point important à souligner est que la participation à une telle initiation d'une modélisation, représentée, pour un biologiste, une réflexion sur ses résultats expérimentaux permettant d'en extraire les éléments essentiels. Il est aussi nécessaire d'essayer, avec l'aide des mathématiques, de les traduire en termes mathématiques, ce qui amène à poser de nouvelles questions biologiques et à apprendre à extraire les données les plus essentielles dans des expériences de biologie. Cela correspondait, pour moi, à une formation complémentaire à celle, classique, de la biologie moléculaire.

La construction des deux modèles à laquelle nous avons participé a fait l'objet d'un article publié dans le cadre de la conférence « Computational Methods in Systems Biology'03 », et d'une publication

qui est actuellement en révision pour la revue *Theoretical Computer Science*. Les manuscrits de ces deux articles sont joints en annexe.

VII.5 Annexes

Multiscale modeling of alternative splicing regulation

Eveillard D., Ropers D., De Jong H., Branlant C. and Bockmayr A. (2003)

Computational Methods in Systems Biology, CSMB'03, C. Priami (ed.), LCNS 2602,
Springer-Verlag, Berlin, 75-87.

Multiscale Modeling of Alternative Splicing Regulation

Damien Eveillard¹, Delphine Ropers², Hidde de Jong³, Christiane Branlant²,
and Alexander Bockmayr^{1*}

¹ LORIA, Université Henri Poincaré
BP 239, 54506 Vandœuvre-lès-Nancy, France
Damien.Eveillard@loria.fr

² Laboratoire de Maturation des ARN et Enzymologie Moléculaire
UMR 7567 CNRS-UHP, BP 239, 54506 Vandœuvre-lès-Nancy, France

³ INRIA Rhône-Alpes, Helix Project
655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier, France

Abstract. Alternative splicing is a key process in post-transcriptional regulation, by which several kinds of mature RNA can be obtained from the same premessenger RNA. Using a constraint programming approach, we model the alternative splicing regulation at different scales (single site vs. multiple sites), thus exploiting different types of available experimental data.

1 Introduction

Alternative splicing is a biological process occurring in post-transcriptional regulation of RNA. Through the elimination of selected introns, alternative splicing allows generating several kinds of mRNA from the same premessenger RNA. The combinatorial effect of splicing contributes to biological diversity, especially in the case of the human immunodeficiency virus (HIV-1). Recent biological studies show the impact that SR proteins have on the dynamics of post-transcriptional regulation via the control of the splicing process [8]. SR proteins can be divided into two functional classes: they may either activate or inhibit splicing. Due to the complexity of alternative splicing regulation, the knowledge that can be gained from experiments is limited. Each experiment focus on one splicing site. In a first approach, we model SR regulation in this restricted context. Using differential equations, we develop a qualitative model for the A3 splicing site in HIV-1. The qualitative behavior depends on the values of the reaction kinetic parameters. Experimental results available to us validate this first approach in the equilibrium phase. Our second approach aims at validation on a higher scale. The ultimate goal is to obtain a model that can be validated qualitatively both on the scale of a single splicing site and on the scale of the whole HIV-1, in order to represent the global effect of alternative splicing in the HIV-1 cycle.

* Part of this work was done within the ARC INRIA "Process Calculi and Biology of Molecular Networks", <http://contraintes.inria.fr/cpbio>

Our models are developed in a constraint programming framework [2, 3]. Constraint programming seems well-suited for modeling biological systems because it allows one to handle partial or incomplete information on the system state. Each constraint gives one piece of information on the system. The overall knowledge is accumulated in the constraint store. The constraint engine available in constraint programming systems operates on the constraint store. It may add new information to the store or check whether some property is entailed by the information already available.

While a constraint model may be refined whenever additional biological knowledge becomes available, it allows one to make useful inferences even from partial and incomplete information. Therefore, constraint programming seems to be a natural computational approach to face the current situation in systems biology as it is described by B. Palsson [16]: “Because biological information is incomplete, it is necessary to take into account the fact that cells are subject to certain constraints that limit their possible behaviors. By imposing these constraints in a model, one can then determine what is possible and what is not, and determine how a cell is likely to behave, but never predict its behavior precisely.”

The organisation of the paper is as follows: we start in Sect. 2 with a description of the biological process of alternative splicing regulation. Based on a number of biological hypotheses, we develop a continuous model of the regulation at one splicing site. This model includes competition and compensation of different proteins on two binding sites, ESE and ESS2. The single-site model is validated in a qualitative way by extracting from the model a splice efficiency function, which can be measured in experiments. In Sect. 3, we first simulate the single-site continuous model in the hybrid concurrent constraint programming language *Hybrid cc* [9, 10]. Then we derive a more global model involving two generic splicing sites, which may be generalized to multiple sites. This means that we model at two different scales, using the splice efficiency as an abstraction of the local model of one site in the more global context of different sites. The two-site model uses the constraint solving and default reasoning facilities of *Hybrid cc*. This allows us to make predictions on the global behavior even in absence of detailed local information on some of the splicing sites.

2 The Biological Problem of Alternative Splicing Regulation

2.1 Biological Process

Eukaryotic and virus gene expression is based on production of RNA containing intronic sequences. The process of splicing allows for intron elimination and junction of exonic sequences [14], see Fig. 1. Splicing is a major biological process in the HIV-1 life cycle: the viral RNA either remains unchanged to serve as genomic RNA for new virions, or it is alternatively spliced to allow for the production of virion proteins [26]. The viral genome, initially RNA, is integrated

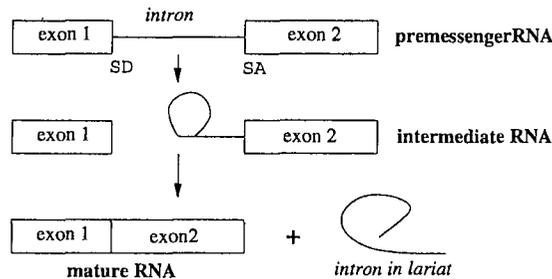


Fig. 1. Representation of the splicing process

into the host genome. In the HIV-1 case, splicing regulation is a complex phenomenon involving 4 donors sites (SD) and 8 acceptors sites (SA), which may yield 40 proteins [17].

This combinatorial complexity is achieved by regulating the selection of the acceptor site [15, 17]. Protein factors control the regulation via binding sites. We focus in our study on the acceptor site A3, where splicing can be repressed by hnRNP A/B via the ESS2 binding site [4, 6], see Fig. 2. Recent experimental studies carried out in our group [18] show that an ESE sequence can activate splicing at the A3 site when SC35 and ASF/SF2 proteins bind to it. More specifically, the ratio of hnRNP A/B and SR proteins determines the splice efficiency at the A3 site.

2.2 Biological Hypotheses

We model the regulation by SR proteins in the restricted context of the A3 splicing site (see Fig. 2) under the following hypotheses:

- We study only one splicing site. Thus, we consider regulation at the scale corresponding to our experimental results. These yield the splice efficiency as the ratio of the mature RNA over pre-messenger RNA.

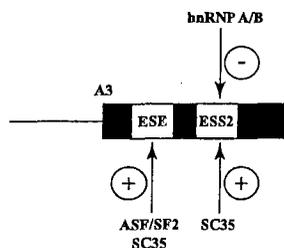


Fig. 2. Regulatory elements of the A3 splicing site

78 Damien Eveillard et al.

- We suppose that the splicing process involves two steps, relating three functional classes of RNA: immature, intermediate, and mature RNA, see Fig. 1. Intermediate RNA corresponds to immature RNA activated by proteins. Mature RNA corresponds to mature RNA and introns in lariat.
- The protein concentration in experiments is saturated. Therefore, we assume that it is constant, despite the binding of proteins to the RNA during regulation.
- SR proteins have two functions. They regulate the splicing process, and they initialize the splicing machinery.
- Regulation is controlled by the ESE and ESS2 binding sites, which are independent. Thus, only indirect interaction is possible between ESE and ESS2.
- The SR proteins ASF/SF2 and SC35 may activate the first splicing reaction by binding to the site ESE. We assume that these two proteins compensate each other.
- The hnRNP proteins may inhibit the first splicing reaction by binding to the site ESS2. On the other hand, if the SC35 proteins bind to ESS2, this activates the first splicing reaction. Therefore we have a competition between hnRNP and SC35.
- Due to a lack of experimental results, we assume Michaelis-Menten kinetics for all proteins. Our goal is a qualitative model validated at equilibrium. Thus, we do not consider transient states.

Our biological hypotheses are summarized in Fig. 3

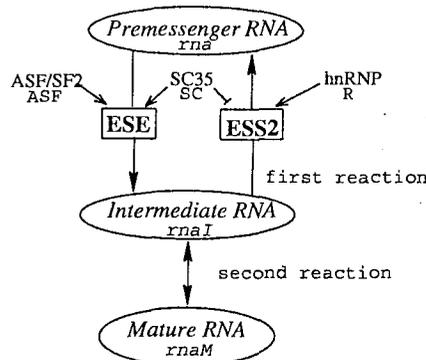


Fig. 3. Schematic representation of the splicing site regulation

2.3 Formal Model

The biological hypotheses can be translated into a mathematical formalism, which leads to a system of ordinary differential equations based on the Michaelis-Menten relation [22]. The single-site model will later be integrated into a larger

multi-site model, see Sect. 3. We describe the splicing process by three ordinary differential equations (ODEs) corresponding to the three functional classes of RNA. Two terms represent the first splicing reaction. The first term represents the cooperation between ASF/SF2 and SC35 in the regulation of ESE. Since we assume compensation, only the sum of activator proteins is important. We represent this interaction by a Michaelis-Menten function depending on the quantity of immature RNA and controlled by the sum of the proteins ASF/SF2 and SC35 [13]:

$$\frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna$$

The symbols used are given in Tab. 1. The second term captures the antagonistic function of hnRNP and SC35 proteins on the site ESS2. In this case, the Michaelis-Menten function represents the inhibitive competition between two proteins : hnRNP and SC35 [27]. It depends on the quantity of intermediary RNA:

$$\frac{\varphi_R \times R}{k_R(1 + \frac{SC}{k_{SC+R}})} rnaI$$

The second splicing reaction is modeled by a simple first order kinetic with constant parameters κ and κ' . Different RNAs decrease proportional to the same degradation factor λ . We formalize the biological process by the system of differential equations:

$$\begin{aligned} \frac{d(rna)}{dt} &= \frac{\varphi_R \times R}{k_R(1 + \frac{SC}{k_{SC+R}})} rnaI - \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna - \lambda \cdot rna \\ \frac{d(rnaI)}{dt} &= \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna - \frac{\varphi_R \cdot R}{k_R(1 + \frac{SC}{k_{SC+R}})} rnaI - \kappa \cdot rnaI \\ &\quad + \kappa' \cdot rnaM - \lambda \cdot rnaI \\ \frac{d(rnaM)}{dt} &= \kappa \cdot rnaI - \kappa' \cdot rnaM - \lambda \cdot rnaM \end{aligned}$$

2.4 Validation of the Regulatory System

The formal model of regulation at a single-site can be directly simulated in the constraint programming language Hybrid cc, see Sect. 3.1. However, before doing this, it should first be validated with respect to existing biological knowledge.

A mathematical analysis of the ODE system in Sect. 2.3 shows that the partial derivatives in the Jacobian matrix have two characteristic properties:

- the partial derivatives on the diagonal elements are negative.
- the partial derivatives on the extra diagonal elements are positive.

Table 1. Symbols and units for the biological variables and parameters

Symbol	Variables and Parameters	unit
rna	Immature RNA	μM
$rnaI$	Intermediary RNA	μM
$rnaM$	Mature RNA	μM
ASF	Protein ASF/SF2	μM
SC	Protein SC35	μM
R	Protein hnRNP	μM
φ_{ESE}	Maximal affinity for the enhancer	s^{-1}
φ_R	Maximal affinity of hnRNP	s^{-1}
k_{ESE}	Half saturation coefficient for the enhancer	μM
k_{SC}	Half saturation coefficient for SC35	μM
k_R	Half saturation coefficient for hnRNP	μM
κ	Reaction rate	s^{-1}
κ'	Reaction rate	s^{-1}
λ	Degradation coefficient	s^{-1}

In our model, the RNA concentrations do not reach an equilibrium, but continue to decrease until total degradation of RNA. However, the above two properties imply that the *splice efficiency* defined by

$$efficiency(t) = \frac{rnaM(t)}{rna(t)}$$

reaches an equilibrium [1]. From the condition $d\,efficiency/dt = 0$, we may derive the following formula for the splice efficiency in the equilibrium phase:

$$efficiency = \frac{\kappa \cdot \varphi_{ESE}(ASF + SC)(k_R \cdot k_{SC} + k_R \cdot SC + R \cdot k_{SC})}{\kappa'(k_{ESE} + ASF + SC) \cdot \varphi_R \cdot R \cdot k_{SC}}$$

It is reasonable to assume that this equilibrium is reached after a short transient phase, so that it can be measured by experiments. According to our formula, the splice efficiency is

- an increasing function of the activators SC and ASF .
- a decreasing function of the inhibitor R .

Experimental results show that

- $rnaM/rna$ increases with an increase of activator proteins.
- $rnaM/rna$ decreases with an increase of inhibitor proteins.

These results of our model correlate with available experimental data. In summary, the model may be considered as qualitatively validated under the hypotheses described in Sect. 2.2. We next consider simulation in the concurrent constraint language Hybrid cc.

3 Multiscale Modeling and Simulation with Hybrid cc

3.1 Hybrid Concurrent Constraint Programming

In constraint programming, the user specifies constraints on the behavior of the system that is being studied. Each constraint expresses some partial information on the system state. The constraint solver may check constraints for consistency or infer new constraints from the given ones. In concurrent constraint programming (cc), different computation processes may run concurrently. Interaction is possible via the *constraint store*. The store contains all the constraints currently known about the system. A process may *tell* the store a new constraint, or *ask* the store whether some constraint is entailed by the information currently available, in which case further action is taken [19]. One major difficulty in the original cc framework is that cc programs can detect only the presence of information, not its absence. To overcome this problem, [20] proposed to add to the cc paradigm a sequence of phases of execution. At each phase, a cc program is executed. At the end, absence of information is detected, and used in the next phase. This results in a synchronous reactive programming language, Timed cc. But, the question remains how to detect negative information instantaneously. Default cc extends cc by a negative ask combinator $\text{if } c \text{ else } A$, which imposes the constraints of A unless the rest of the system imposes the constraint c . Logically, this can be seen as a default. Introducing phases as in Timed cc leads to Timed Default cc [21]. Only one additional construct is needed: hence A , which starts a copy of A in each phase after the current one.

Table 2. Combinators of Hybrid cc

Agents	Propositions
c	c holds now
$\text{if } c \text{ then } A$	if c holds now, then A holds now
$\text{if } c \text{ else } A$	if c will not hold now, then A holds now
$\text{new } X \text{ in } A$	there is an instance $A[T/X]$ that holds now
A, B	both A and B hold now
$\text{hence } A$	A holds at every instant after now
$\text{always } A$	same as $(A, \text{hence } A)$
$\text{when}(c) A$	same as $(\text{if } c \text{ then } A, \text{hence } (\text{if } c \text{ then } A))$
$\text{unless}(c) A \text{ else } B$	same as $(\text{if } c \text{ then } B, \text{if } c \text{ else } A)$

Hybrid cc [9, 10] is an extension of Default cc over continuous time. First continuous constraint systems are allowed, i.e., constraints may involve differential equations that express initial value problems. Second, the hence operator is interpreted over continuous time. It imposes the constraints of A at every real time instant after the current one. The evolution of a system in Hybrid cc is piecewise continuous, with a sequence of alternating point and interval phases.

82 Damien Eveillard et al.

All discrete changes take place in a point phase, where a simple Default cc program is executed. In a continuous phase, computation proceeds only through the evolution of time. The interval phase, whose duration is determined in the previous point phase, is exited as soon as the status of a conditional changes [10]. Tab. 2 summarizes the basic combinators of Hybrid cc.

Hybrid cc is well-suited for modeling dynamic biological systems, as argued in [2, 3].

3.2 Single-Site Model: Local Modeling

The single-site model from Sect. 2.3 with experimental values can be expressed directly in Hybrid cc.

```
interval t, rna, rnaI, rnaM;
t=0; rna = 0.06; rnaI = 0; rnaM = 0;
always{ t' = 200;
  rna' = (Pr*R*rnaI)/(kr*(1+(SC/ksc))+R)
        -(Pese*(ASF+SC)*rna)/(kese+ASF+SC)-delta*rna;
  rnaI' = (Pese*(ASF+SC)*rna)/(kese+ASF+SC)
          -(Pr*R*rnaI)/(kr*(1+(SC/ksc))+R)-k*rnaI+kk*rnaM-delta*rnaI;
  rnaM' = k*rnaI-kk*rnaM-delta*rnaM;
}
```

During the simulation, we obtain the predicted equilibrium for the splice efficiency, see Fig. 4. Under our hypotheses, which include protein competition and compensation, the model correctly simulates the alternative splicing activity. This supports the hypotheses made in the model such as the role of the ESE and ESS2 binding sites.

3.3 Two Site Model: Global Modeling

A realistic model of alternative splicing has to reflect the combinatorial complexity discussed in Sect. 2.1. Assuming that regulation is modular [12], the single-site model may be seen as one module inside a larger framework. The qualitative validation given in Sect. 2.4 justifies the introduction of the single-site model into a larger scale model involving several splicing sites. To illustrate this, we consider the generic example of two splice acceptor sites (SA) associated with one donor site (SD), see Fig. 5.

The behavior at one splicing site can be captured by a single function, the splice efficiency, which depends on the protein concentrations. This function is used in a larger-scale global model that describes the choice between two acceptor sites A3 and A4. In the HIV-1 case, the A3 site is the default splicing site. Only if the splice efficiency *eff1* gets smaller than *eff2*, regulation switches to the other state.

Recent work [5] shows the linearity of the splicing kinetics. Thus, on the larger scale, we may consider splicing as a linear process described by two ordinary

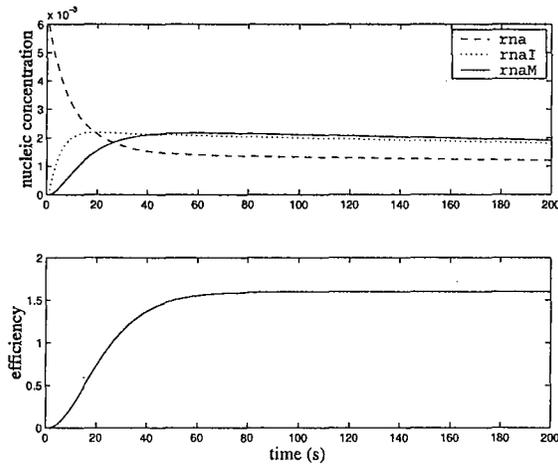


Fig. 4. Variation of the pool of RNA and the splice efficiency in the splicing reaction

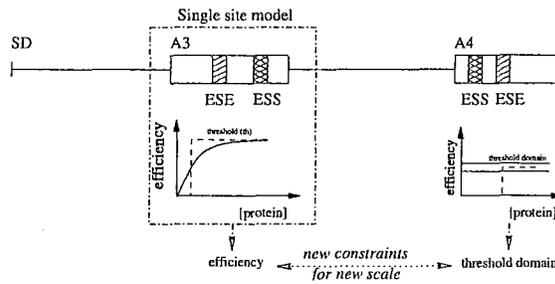


Fig. 5. Single-site model inside a general splicing regulation model

differential equations involving some kinetic constant. The first system represents the default behavior characterized by the constant $Ka4$. Following local changes on a single-site, the model may exhibit a different behavior, characterized by the kinetic constant $Ka3$. Thus, in this example, only the kinetic constants change, while the overall structure of the ODE system remains the same.

This default behavior can be naturally expressed in Hybrid cc using the combinator `unless(c) A`.

84 Damien Eveillard et al.

```

always {
  if (eff1 <= eff2) {rna' = -Ka3*rna;
                    rnam1' = 0;
                    rnam2' = Ka3*rna;};
  unless ((eff1 <= eff2)) {rna' = - Ka4*rna;
                           rnam1' = Ka4*rna;
                           rnam2' = 0;};
}

```

Note that `unless(c) A` is not equivalent to `if ¬c then A`. The second alternative will be chosen if the solver cannot verify that `(eff1 <= eff2)`. This may have *two* reasons:

- `(eff1 <= eff2)` is false, i.e. `(eff1 > eff2)`, or
- `(eff1 <= eff2)` is unknown (default behavior).

Simulation in Hybrid cc yields the behavior illustrated in Fig.6. We observe the switch from the first system of ODEs to the second when `eff2` passes the upper threshold for `eff1`.

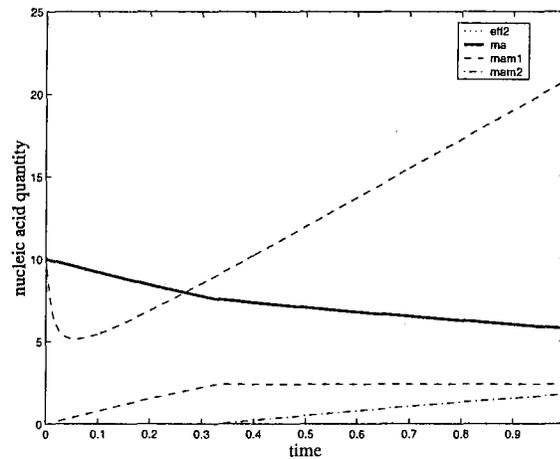


Fig. 6. Variation of RNA with a variation of splice efficiency

4 Conclusion and Further Research

Our approach combines mathematical and computational methods. Mathematical analysis allows us to validate the single-site model in a qualitative way. This is possible using the experimental data obtained in our group. The validation

shows the consistency of our biological hypotheses. Based on this, we can extract the splice efficiency as a suitable abstraction of the local behavior at one site inside a more global model involving different sites. For the experimental biologist, the single-site model may serve as a computational tool to evaluate his knowledge on a fine-grained biological process.

On the computational side, the constraint solving and default reasoning capabilities of Hybrid cc allow us to exploit as much as possible the incomplete knowledge that is typical for ongoing biological research. Indeed, default behavior may compensate the lack of experimental data. Thus, using constraint programming, we can delimit with our model the possible splicing behavior. Similar to mathematical analysis, constraint reasoning therefore may provide a powerful qualitative validation.

The combination of mathematical analysis and computational methods is also the key to the multiscale modeling developed in this paper. It leads to the qualitative validation represented by the extraction of the splice efficiency function. The splice efficiency characterizes the modularity of the regulation. Thus, the smaller-scale behavior is represented in the larger-scale model, based on the single-site splice efficiency. The extraction of a suitable criterion on the smaller scale is crucial to understanding an experimental process from a systems biology perspective. Furthermore, constraints can be used to handle the problem of missing data in a multiscale model. Different scales usually correspond to biological experiments yielding different types of results. Despite the variety of possible experiments, these must be integrated into a global model in order to better understand the biological process.

Multiscale modeling requires a close interaction between biological and computational approaches, as illustrated by Fig.7.

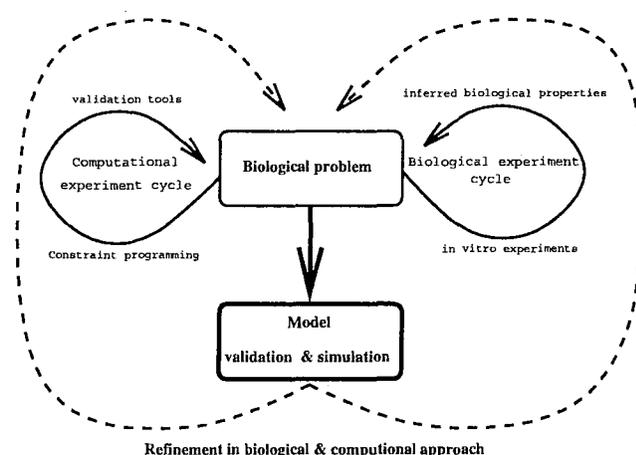


Fig. 7. Representation of ideal interactions for modeling biology

In the context of alternative splicing regulation, we are currently working on new experimental data for the quantitative validation of our models. On the computational side, we have integrated our model into a general HIV-1 model [11]. Preliminary results show that the modification of a splice constant may induce different behaviors in the HIV-1 life cycle model. Using the extended model, we may validate several biological hypotheses on the global effect of alternative splicing in the full HIV-1 life cycle.

References

- [1] Bernard, O., and Gouzé J.-L.: Transient behavior of biological models as a tool of qualitative validation - Application to the Droop model and to a N-P-Z model. *J. Biol. Syst.*, 4(3) (1996) 303–314 80
- [2] Bockmayr, A., Courtois, A.: Modeling biological systems in hybrid concurrent constraint programming (Abstract). In 2nd Int. Conf. Systems Biology, ICSB'01, Pasadena, CA (2001) 76, 82
- [3] Bockmayr, A., Courtois, A.: Using hybrid concurrent constraint programming to model dynamic biological systems. In 18th International Conference on Logic Programming, ICLP'02, Copenhagen Springer, LNCS 2401 (2002) 76, 82
- [4] Caputi, M. A., Mayeda, M. A., Krainer, A. R., and Zahler, A. M.: hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO* 18(14) (1999) 4060–7 77
- [5] Dautry, F., Weill, D.: Kinetic analysis of mRNA metabolism. In Interdisciplinary School on imaging, modelling and manipulating transcriptional regulatory networks, Ambleteuse (2002) 20/92 82
- [6] Del Gatto-Konczak, F., Olive, M., Gesnel, M. C., and Breathnach, R.: hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol. Cell. Biol.* 19 (1) (1999) 251–60 77
- [7] Graveley, B. R., Hertel, K. J., Maniatis, T.: A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO* 17(22) (1998) 6747–6756
- [8] Graveley, B. R.: Sorting out the complexity of SR protein functions. *RNA* 6 (2000) 1197–1211 75
- [9] Gupta, V., Jagadeesan, R., Saraswat V.: Computing with continuous change. *Science of computer programming* 30(1-2) (1998) 3–49 76, 81
- [10] Gupta, V., Jagadeesan, R., Saraswat V., Bobrow, D. G.: Programming in hybrid constraint languages. In *Hybrid Systems II*, 226–251. Springer, LNCS 999 (1995) 76, 81, 82
- [11] Hammond, B. J.: Quantitative Study of the Control of HIV-1 Gene Expression. *J. Theor. Biol.* 163 (1993) 199–221 86
- [12] Hartwell, L. H., Hopfield, J. J., Leibler, S., Murray, A. W.: From molecular to modular cell biology. *Nature* 402 (1999) C47–C52 82
- [13] Heinrich, R., Schuster, S.: *The regulation of cellular systems*. Int'l Thomson Publishing, New York (1996) 79
- [14] Moore, M. J., Query, C. C., Sharp, P. A.: Splicing of precursors to mRNA by the spliceosome. In *The RNA World*, Cold Spring Harbor Laboratory Press (1993) 303–357 76

- [15] O'Reilly, M., McNally, M. T., Beemon, K. L.: Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology* 213(2) (1995) 373–85 77
- [16] Palsson, B.: The challenges of in silico biology. *Nature Biotechnology* 18 (2000) 1147 – 1150 76
- [17] Purcell, D. F., Martin, M. A.: Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* 67(11)(1993) 6365–6378 77
- [18] Ropers, D., Ayadi, L., Jacquenet, S., Méreau, A., Thomas, D., Mougin, A., Bilodeau, P., Stoltzfus, C. M., Gattoni, R., Stévenin, J., Branlant, C.: A complex regulation of the central A3 to A5 acceptor sites in HIV-1 RNA. In *Eukaryotic mRNA Processing* (2001) 77
- [19] Saraswat, V. A.: *Concurrent constraint programming*. ACM Doctoral Dissertation Awards. MIT Press (1993) 81
- [20] Saraswat, V. A., Jagadeesan, R., Gupta, V.: Foundations of timed concurrent constraint programming. In 9th Symp. Logic in Computer Science, LICS'94, Paris IEEE (1994) 71 – 80 81
- [21] Saraswat, V. A., Jagadeesan, R., Gupta, V.: Timed default concurrent constraint programming. *Journal of Symbolic Computation* 22(5/6) (1996) 475–520 81
- [22] Segel, L. A.: *Modelling dynamic phenomena in molecular and cellular biology*. Cambridge University Press (1984) 78
- [23] Si, Z. H., Amendt, B. A., Stoltzfus, C. M.: 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. *N. A. R.* 25(4) (1997) 861–867
- [24] Si, Z. H., Rauch, D., Stoltzfus, C. M.: The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. *Mol. Cell. Biol* 18(9) (1998) 5404–13
- [25] Staffa, A., Cochrane, A.: The tat/rev intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. *J. Virol.* 68(5) (1994) 3071–9
- [26] Tang, H., Kuhen, K. L., Wong-Staal, F.: Lentivirus replication and regulation. *Annu. Rev. Genet* 33 (1999) 133–170 76
- [27] Voit, E.: *Computational Analysis of Biochemical Systems*. Cambridge University Press (2000) 79

A multi-site constraint programming model of alternative splicing regulation

Eveillard D., Ropers D., De Jong H., Branlant C. and Bockmayr A.

En révision pour publication dans Theoretical Computer Science.

A Multi-Site Constraint Programming Model of Alternative Splicing Regulation

Damien Eveillard^{a,b,1} Delphine Ropers^b Hidde de Jong^c
Christiane Branlant^b Alexander Bockmayr^{a,2}

^aLORIA, Université Henri Poincaré, BP 239, 54506 Vandœuvre-lès-Nancy,
France

^bLaboratoire de Maturation des ARN et Enzymologie Moléculaire,
UMR 7567 CNRS-UHP, BP 239, 54506 Vandœuvre-lès-Nancy, France

^cINRIA Rhône-Alpes, Helix Project, 655 avenue de l'Europe, Montbonnot, 38334
Saint Ismier, France

Abstract

Alternative splicing is a key process in post-transcriptional regulation, by which several kinds of mature RNA can be obtained from the same pre-messenger RNA. The resulting combinatorial complexity contributes to biological diversity, especially in the case of the human immunodeficiency virus HIV-1. Using a constraint programming approach, we develop a model of the alternative splicing regulation in HIV-1. Our model integrates different scales (single site vs. multiple sites), and thus allows us to exploit several types of experimental data available to us.

Key words: computational biology, constraint programming, modeling, hybrid system, alternative splicing, HIV-1

1 Introduction

Alternative splicing is a biological process occurring in post-transcriptional regulation of RNA. Through the elimination of selected introns, alternative splicing allows the generation of several kinds of mature RNA from the same pre-messenger RNA. The resulting combinatorial complexity contributes to biological diversity, especially in the case of the human immunodeficiency virus

¹ Corresponding author: damien.eveillard@loria.fr

² Part of this work was done within the ARC INRIA "Process Calculi and Biology of Molecular Networks", <http://contraintes.inria.fr/cpbio>

(HIV-1). Recent biological studies show the impact that SR proteins have on the dynamics of post-transcriptional regulation via the control of the splicing process (Graveley, 2000). SR proteins can be divided into two functional classes: they may either activate or inhibit splicing. Due to the complexity of alternative splicing regulation, the knowledge that can be gained from experiments is limited. Each experiment focuses on one splicing site. In a first approach, we model SR regulation in this restricted context. Using differential equations, we develop a model for the regulation of the A3 splicing site in HIV-1. The qualitative behavior depends on the values of the reaction kinetic parameters. Experimental results available to us validate this first approach in the equilibrium phase. Our second approach aims at validation on a higher scale. The ultimate goal is to obtain a model that can be validated qualitatively both on the scale of a single splicing site and on the scale of the whole HIV-1, in order to represent the global effect of alternative splicing in the HIV-1 cycle.

Our models are developed in a constraint programming framework (Bockmayr and Courtois, 2001, 2002). Constraint programming seems well-suited for modeling biological systems because it allows one to handle partial or incomplete information. Each constraint gives one piece of information on the system that is studied. The overall knowledge is accumulated in the constraint store. The constraint engine available in constraint programming systems operates on the constraint store. It may add new information to the store or check whether some property is entailed by the information present in the store. While a constraint model may be refined whenever additional biological knowledge becomes available, it allows one to make useful inferences even from partial and incomplete information. Therefore, constraint programming seems to be a natural computational approach to face the current situation in systems biology as it is described by Palsson (2000): “Because biological information is incomplete, it is necessary to take into account the fact that cells are subject to certain constraints that limit their possible behaviors. By imposing these constraints in a model, one can then determine what is possible and what is not, and determine how a cell is likely to behave, but never predict its behavior precisely.”

The organisation of the paper is as follows: we start in Section 2 with a description of the biological process of alternative splicing regulation. Based on a number of biological hypotheses, we develop in Section 3 a continuous model of the regulation at one splicing site. This model includes competition and compensation of different proteins on two binding sites, ESE and ESS2. The single-site model is validated in a qualitative way by extracting from the model a splice efficiency function, which can be measured in experiments. In Section 4, we briefly present the hybrid concurrent constraint programming language Hybrid cc (Gupta et al., 1998, 1995), and explain how it can be used for modeling dynamic biological systems. In Section 5, we first simu-

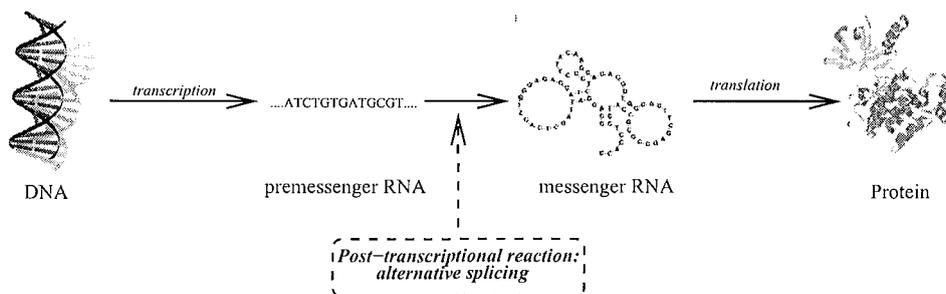


Fig. 1. Information flow in molecular biology

late the single-site continuous model in this language. Then we derive a more global model involving three generic splicing sites, which may be generalized to multiple sites. This means that we model at two different scales, using the splice efficiency function as a time-scale abstraction of the local model of one site in the more global context of different sites. The three-site model uses the constraint solving and default reasoning facilities of Hybrid cc. This allows us to make predictions on the global behavior even in absence of detailed local information on some of the splicing sites.

2 Alternative splicing: A biological problem for formals methods

2.1 The biological problem of alternative splicing regulation

Molecular biology studies the information flow in the *transcription* of DNA into RNA, and the *translation* of RNA in proteins, see Fig. 1. In eukaryotes and viruses, the transcriptional process, where information contained in the DNA molecule is extracted and transcribed into an RNA molecule, is followed by another process, which is *alternative splicing* (Moore et al., 1993). In fact, the DNA molecule first yields a *pre-messenger* RNA molecule. The pre-messenger RNA can be decomposed in exons and introns. During splicing, introns are removed. Only the exons are kept for the *messenger* or *mature* RNA (mRNA).

The regulation of the splicing process depends on several sites. The first one is the *donor site* *SD*, located at the end of one exon, see Fig. 2. Its main characteristic is the GU nucleic acid sequence motif, which contains the phosphate domain. The other site is the *acceptor site* *SA* located at the beginning of the next exon. It is characterized by an AG motif containing another phosphate domain. Together, they define the intron to be excised from the pre-messenger RNA. They permit the binding of a huge ribonucleoproteic complex: the spliceosome. This complex is partially activated by another motif, the *branching point* *BP*. This is another binding site contained inside the intron. The three sites permit the regulation of the splicing process by activation of the

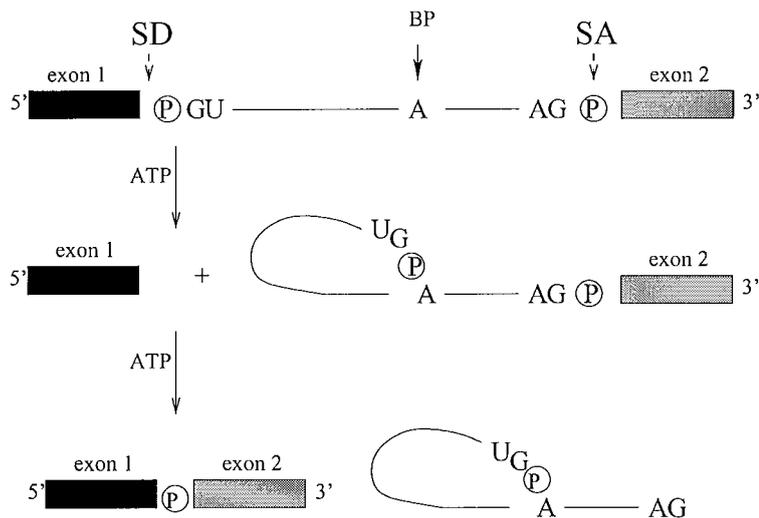


Fig. 2. The splicing process operates in the intronic region of a pre-messenger RNA that lies between two exons. The exons are delimited by the SD and SA binding sites. A first reaction cuts the RNA at the SD binding site. A second reaction cuts in the phosphate domain of the SA binding site. Each reaction requires ATP energy.

spliceosome complex. The key to the regulation is the choice of one acceptor and one donor site.

2.2 Combinatorial complexity

The splicing process is a key source of complexity in the cell. As illustrated by Fig. 3, various messenger RNA can be obtained from a unique pre-messenger RNA through the elimination of different introns, and the junction of the remaining exonic sequences. This process is regulated through the choice of the donor and acceptor sites. The complexity of the regulation increases with the capacity of an exon to become an intron in a given cellular condition.

In Fig. 3, seven types of intron elimination are shown. In each case, the dotted lines above and below the pre-messenger RNA give two concrete examples of intron elimination. The first case describes classical exons. The second case can be seen as a competition between different exons. In the third and the fourth case, regulation depends on the position of the acceptor and the donor site respectively. In the fifth case, an intron is partitioned between two exons. In the last two cases, the choice between different promoters or different polyadenylation sites increases the complexity of the alternative splicing.

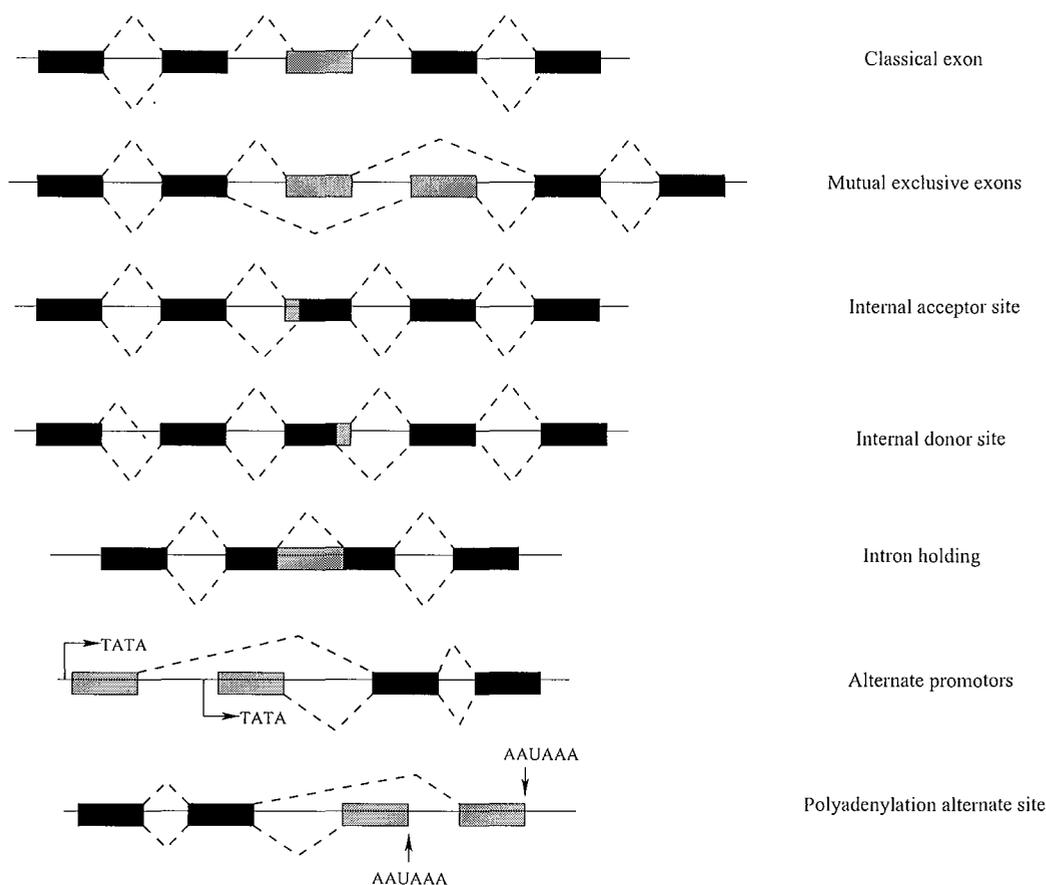


Fig. 3. Complexity induced by alternative splicing: seven types of intron elimination

2.3 Alternative splicing in the context of HIV-1

Eukaryotic and viral gene expression involves the production of RNA containing intronic sequences. The process of splicing allows for intron elimination and junction of exonic sequences. Splicing is a major biological process in the HIV-1 life cycle: the viral RNA either remains unchanged to serve as genomic RNA for new virions, or it is alternatively spliced to allow for the production of virion proteins (Tang et al., 1999). The viral genome, initially RNA, is integrated into the host genome. In the HIV-1 case, splicing regulation is a complex phenomenon involving 4 donor sites (SD) and 8 acceptor sites (SA), which may yield 40 mature messenger RNAs (Purcell and Martin, 1993). This combinatorial complexity is achieved by regulating the selection of the acceptor site (O'Reilly et al., 1995; Purcell and Martin, 1993). Protein factors control the regulation via binding sites. We focus in our study on the acceptor site A3, see Fig. 4, where splicing can be repressed by hnRNP A/B proteins via the ESS2 binding site (Caputi et al., 1999; Del Gatto-Konczak et al., 1999). Recent experimental studies carried out in our group (Ropers et al., 2003) show that an ESE sequence can activate splicing at the A3 site when SC35 and ASF/SF2 proteins bind to it. SC35 can also bind to the ESS2 site. More

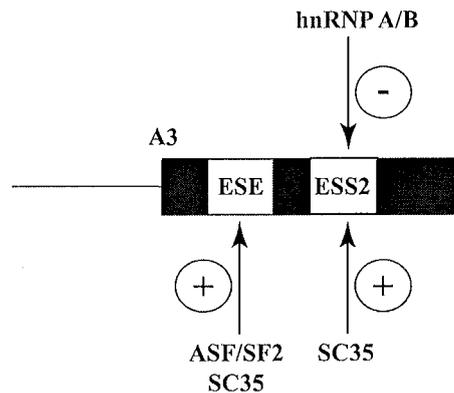


Fig. 4. Regulatory elements of the A3 splicing site. The exon delimited by the A3 acceptor site contains the ESE and ESS2 binding sites, which bind ASF/SF2, SC35 and hnRNP A/B proteins. These regulatory elements activate or repress the splicing reaction on the A3 site.

specifically, the ratio of hnRNP A/B and SR proteins determines the splice efficiency at the A3 site.

3 Modeling one splicing site

3.1 Biological hypotheses

We model the regulation by SR proteins in the restricted context of the A3 splicing site under the following hypotheses, see Fig. 4:

- We study only one splicing site. Thus, we consider regulation at the scale corresponding to our experimental results, which are measurements of the splice efficiency given as the ratio of mature RNA over pre-messenger RNA.
- We suppose that the splicing process involves two reactions, relating three functional classes of RNA, see Fig. 2: immature RNA (*rna*), intermediate RNA (*irna*), and mature RNA (*mrna*). Intermediate RNA corresponds to immature RNA activated by proteins. Mature RNA corresponds to mature RNA and introns in lariat.
- The protein concentration in experiments is saturated. Therefore, we assume that it is constant, despite the binding of proteins to the RNA during regulation.
- SR proteins regulate the splicing process by initialization of the splicing machinery.
- Regulation is controlled by the ESE and ESS2 binding sites, which are independent.
- The SR proteins ASF/SF2 and SC35 may activate the first splicing reaction

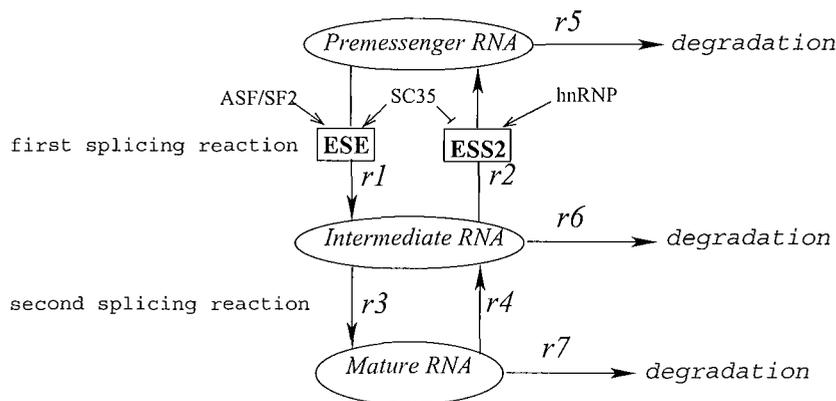


Fig. 5. Schematic representation of the splicing site regulation. The two splicing reactions are composed by 7 kinetic reactions.

by binding to the site ESE. We assume that these two proteins compensate each other.

- The hnRNP A/B proteins may inhibit the first splicing reaction by binding to the site ESS2. On the other hand, if the SC35 proteins bind to ESS2, this activates the first splicing reaction. Therefore we have a competition between hnRNP A/B and SC35.
- Due to a lack of experimental results, we assume Michaelis-Menten kinetics for the reactions involved in the splicing process.

The biological hypotheses are summarized in Fig. 5

3.2 Mathematical model

Our biological hypotheses can be represented by a system of ordinary differential equations based on the Michaelis-Menten relation (Segel, 1984). The single-site model that we obtain will later be integrated into a larger multi-site model, see Section 5. We will describe the splicing process by seven kinetic reactions. The symbols used are given in Tab. 1.

The reaction r_1 represents the transformation of premessenger RNA (rna) to intermediate RNA ($irna$). It requires cooperation between ASF/SF2 and SC35 proteins for the regulation of ESE. Since we assume compensation, only the sum of the activator proteins is important. We represent this interaction by a Michaelis-Menten function depending on the quantity of immature RNA, and controlled by the sum of the proteins ASF/SF2 and SC35 (Heinrich and Schuster, 1996)

$$r_1 = \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna$$

The reaction r_2 represents the transformation of intermediate RNA to pre-messenger RNA. It captures the antagonistic function of hnRNP A/B and SC35 proteins on the site ESS2. In this case, the Michaelis-Menten function represents the inhibitive competition between two proteins: hnRNP A/B and SC35 (Voit, 2000). It depends on the quantity of intermediate RNA:

$$r_2 = \frac{\varphi_R \cdot R}{k_R(1 + \frac{SC}{k_{SC+R}})} irna$$

The reaction r_3 represents the transformation of intermediate RNA to mature RNA ($mrna$). We assume for this reaction a simple first-order kinetics with a constant parameter κ . Similarly, r_4 represents the reaction which transforms mature RNA to intermediate RNA:

$$r_3 = \kappa \cdot irna, \quad r_4 = \kappa' \cdot mrna.$$

r_5 , r_6 and r_7 respectively represent the degradation reaction of immature RNA, intermediate RNA and mature RNA. Different RNAs decrease proportional to the same degradation factor λ :

$$r_5 = \lambda \cdot rna, \quad r_6 = \lambda \cdot irna, \quad r_7 = \lambda \cdot mrna.$$

We formalize the splicing process at site A3 by the system of differential equations:

$$\begin{aligned} \frac{d(rna)}{dt} &= r_2 - r_1 - r_5, \\ \frac{d(irna)}{dt} &= r_1 + r_4 - r_2 - r_3 - r_6, \\ \frac{d(mrna)}{dt} &= r_3 - r_4 - r_7, \end{aligned}$$

which corresponds to

$$\begin{aligned} \frac{d(rna)}{dt} &= \frac{\varphi_R \cdot R}{k_R(1 + \frac{SC}{k_{SC+R}})} irna - \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna - \lambda \cdot rna, \\ \frac{d(irna)}{dt} &= \frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna - \frac{\varphi_R \cdot R}{k_R(1 + \frac{SC}{k_{SC+R}})} irna \end{aligned}$$

Table 1
 Symbols and units for the biological variables and parameters

Symbol	Variables and Parameters	unit
rna	Immature RNA	μM
$irna$	Intermediate RNA	μM
$mrna$	Mature RNA	μM
ASF	Protein ASF/SF2	μM
SC	Protein SC35	μM
R	Protein hnRNP A/B	μM
φ_{ESE}	Maximal affinity for the enhancer	s^{-1}
φ_R	Maximal affinity of hnRNP A/B	s^{-1}
k_{ESE}	Half saturation coefficient for the enhancer	μM
k_{SC}	Half saturation coefficient for SC35	μM
k_R	Half saturation coefficient for hnRNP A/B	μM
κ	Reaction rate	s^{-1}
κ'	Reaction rate	s^{-1}
λ	Degradation coefficient	s^{-1}

$$-\kappa \cdot irna + \kappa' \cdot mrna - \lambda \cdot irna,$$

$$\frac{d(mrna)}{dt} = \kappa \cdot irna - \kappa' \cdot mrna - \lambda \cdot mrna.$$

3.3 Validation of the regulatory system

The mathematical model of regulation at a single site can be directly simulated in the constraint programming language Hybrid cc, as will be shown in Section 4. However, it should first be validated with respect to existing biological knowledge. In our model, the RNA concentrations do not reach an equilibrium, but continue to decrease until total degradation of RNA. However, we may assume that the splicing reactions quickly reach an equilibrium.

In the equilibrium phase, we have $r_1 = r_2, r_3 = r_4$, which is equivalent to

$$\frac{\varphi_{ESE}(ASF + SC)}{k_{ESE} + ASF + SC} rna = \frac{\varphi_R \cdot R}{k_R(1 + \frac{SC}{k_{SC+R}})} irna, \quad \kappa \cdot irna = \kappa' \cdot mrna.$$

If we define the *splice efficiency* by

$$efficiency(t) = \frac{mrna(t)}{rna(t)},$$

we obtain the following formula for the splice efficiency in the equilibrium phase:

$$efficiency_{eq} = \frac{\kappa \cdot \varphi_{ESE}(ASF + SC)(k_R \cdot k_{SC} + k_R \cdot SC + R \cdot k_{SC})}{\kappa'(k_{ESE} + ASF + SC) \cdot \varphi_R \cdot R \cdot k_{SC}}$$

According to our formula, the splice efficiency is

- an increasing function of the activators SC and ASF .
- a decreasing function of the inhibitor R .

Experimental results show that

- $(mrna/rna)_{eq}$ increases with an increase of activator proteins.
- $(mrna/rna)_{eq}$ decreases with an increase of inhibitor proteins.

Thus, the results of our model correlate with available experimental data. Therefore, we may consider the model to be qualitatively validated under the hypotheses described in Section 3.1. We next consider simulation in the concurrent constraint language Hybrid cc.

4 Hybrid concurrent constraint programming

To model alternative splicing regulation, we will use hybrid concurrent constraint programming. The general idea of *constraint programming* is that the user specifies constraints on the behavior of the system that is being studied. Each constraint expresses some partial information on the system state. The constraint solver may check constraints for consistency or infer new constraints from the given ones. In *concurrent constraint programming* (cc), different computational processes may run concurrently. Interaction is possible via the *constraint store*. The store contains all the constraints currently known about the system. A process may *tell* the store a new constraint, or *ask* the store whether some constraint is entailed by the information currently available, in which case further action is taken (Saraswat, 1993). One major difficulty in

Table 2
Combinators of Hybrid cc

Agents	Propositions
c	c holds now
if c then A	if c holds now, then A holds now
if c else A	if c will not hold now, then A holds now
new X in A	there is an instance $A[T/X]$ that holds now
(A, B)	both A and B hold now
hence A	A holds at every instant after now
always A	same as $(A, \text{hence } A)$
unless(c) A else B	same as (if c then B , if c else A)

the original cc framework is that cc programs can detect only the presence of information, not its absence. To overcome this problem, Saraswat et al. (1994) proposed to add to the cc paradigm a sequence of phases of execution. At each phase, a cc program is executed. At the end, absence of information is detected, and used in the next phase. This results in a synchronous reactive programming language, Timed cc. But, the question remains how to detect negative information instantaneously. Default cc extends cc by a negative ask combinator if c else A , which imposes the constraints of A unless the rest of the system imposes the constraint c . Logically, this can be seen as a default. Introducing phases as in Timed cc leads to Timed Default cc (Saraswat et al., 1996). Only one additional construct is needed: hence A , which starts a copy of A in each phase after the current one.

Hybrid cc (Gupta et al., 1998, 1995) is an extension of Default cc over continuous time. First continuous constraint systems are allowed, i.e., constraints may involve differential equations that express initial value problems. Second, the hence operator is interpreted over continuous time. It imposes the constraints of A at every real time instant after the current one. The evolution of a system in Hybrid cc is piecewise continuous, with a sequence of alternating point and interval phases. All discrete changes take place in a point phase, where a simple Default cc program is executed. In a continuous phase, computation proceeds only through the evolution of time. The interval phase, whose duration is determined in the previous point phase, is exited as soon as the status of a conditional changes (Gupta et al., 1995). Tab. 2 summarizes the basic combinators of Hybrid cc.

It has been argued in Bockmayr and Courtois (2001, 2002) that Hybrid cc is well-suited for modeling dynamic biological systems. In addition to the general

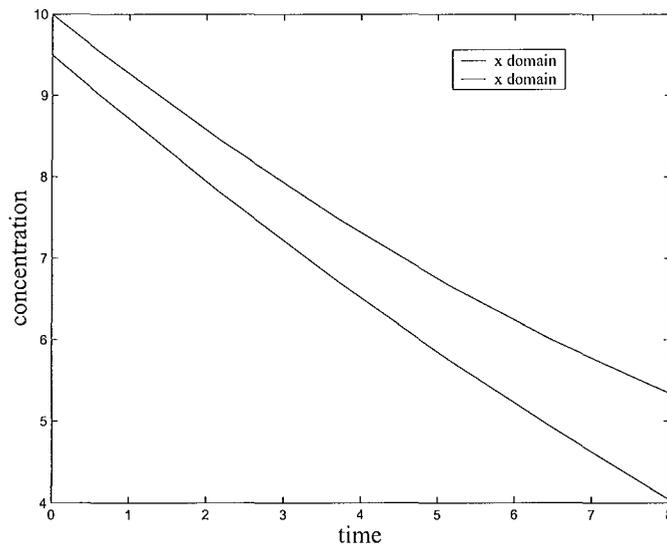


Fig. 6. Reachability region for a molecular species with Michaelis-Menten kinetics

discussion in Bockmayr and Courtois (2002), we illustrate here by a number of small examples, how the basic combinators of *Hybric cc* can be applied naturally to the study of biological systems.

4.1 Interval constraints and continuous dynamics

The *Hybric cc* language that we are using is based on interval constraints (Carlson and Gupta, 1998). This means that variables are defined over an interval of real numbers, and computations are done in interval arithmetic. This is very useful in biology, where typically parameters and values are not exactly known.

We illustrate this by a very simple example in *Hybrid cc* involving a single constraint on an interval variable x , see Fig. 6. Since we are reasoning about dynamical systems, we use the *always A* combinator, expressing that A holds at every time instant.

```
interval x;
x = [9.5,10];
always { x' = -(2*x)/(15+x);
        }
sample(x);
```

4.2 Parallel composition

Hybrid cc allows for parallel composition of constraints. (A, B) imposes the constraints of both A and B . Operationally, the program (A, B) behaves like the simultaneous execution of both A and B . A and B may share common variables, and thus communicate via the constraint store.

We illustrate parallel composition by a small Hybrid cc program specifying a Michaelis-Menten kinetics. Consider two molecular species X and Y with concentrations x and y , and suppose X is transformed into Y . The initial concentration of X lies in the interval $[14, 14.5]$. The production rate of Y depends on the concentration of X according to the formula $y' = (A_{max} * x) / (k_s + x)$, for some constants A_{max} and k_s . The concentration of X is reduced at the same rate. We add constraints $x, y \geq 0$ to say that concentrations are non-negative, and constraints $s = x + y, s' = 0$ to express conservation of matter. The constraint solver computes reachability regions for x and y , see Fig. 7. In particular, we can observe that at the end of the experiment, the concentration of y will be greater than the concentration of x . Interval constraints are particularly useful in sensibility studies, where we can easily test the importance of one variable versus the others. Another advantage of interval constraints is the capability to detect the possible reachability area for a biological system.

```
#define ks 1.5
#define Amax 2
interval x,y,s;
x = [14,14.5];          /* Initialization */
y = 0;
always {
  x' = -(Amax*x)/(ks+x); /* Michaelis-Menten kinetics */
  y' = (Amax*x)/(ks+x);
  x >= 0;                /* Non-negative concentrations */
  y >= 0;
  s = x + y;            /* Conservation of matter */
  s' = 0;
}
sample(x, y);
```

4.3 Conditionals and discrete change

In general, the dynamics of a system will depend on conditions. In Hybrid cc, we may use the combinator `if c then A` expressing that if c holds now,

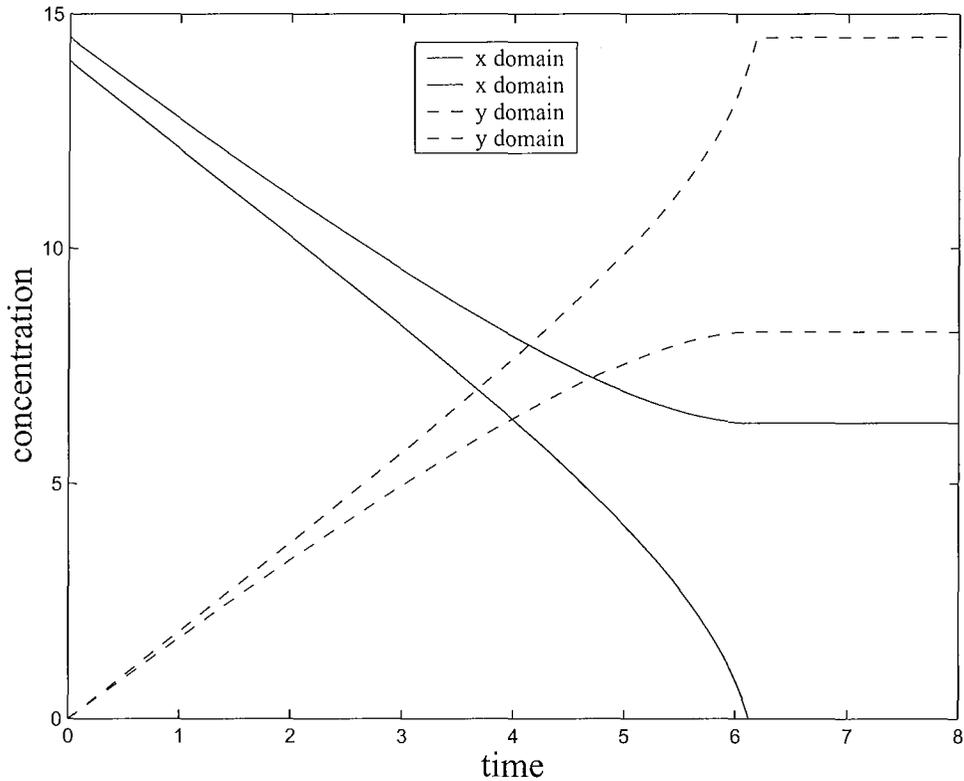


Fig. 7. Reachability regions for two molecular species with Michaelis-Menten kinetics

then A holds now. This allows one to make discrete changes to switch from one dynamics to another. The next program models the situation that the transformation of X to Y gets activated if a certain protein P reaches a threshold.

```

interval x, y, p;
x=[14,14.5];
y=0;
p=0.75;
always {
  p' = 1.5;
  if (p >= 3)
    { x' = -(Amax*x)/(ks+x);
      y' = (Amax*x)/(ks+x);
    }
  if (p < 3)
    { x' = 0;
      y' = 0;
    }
}
sample(p, x ,y);

```

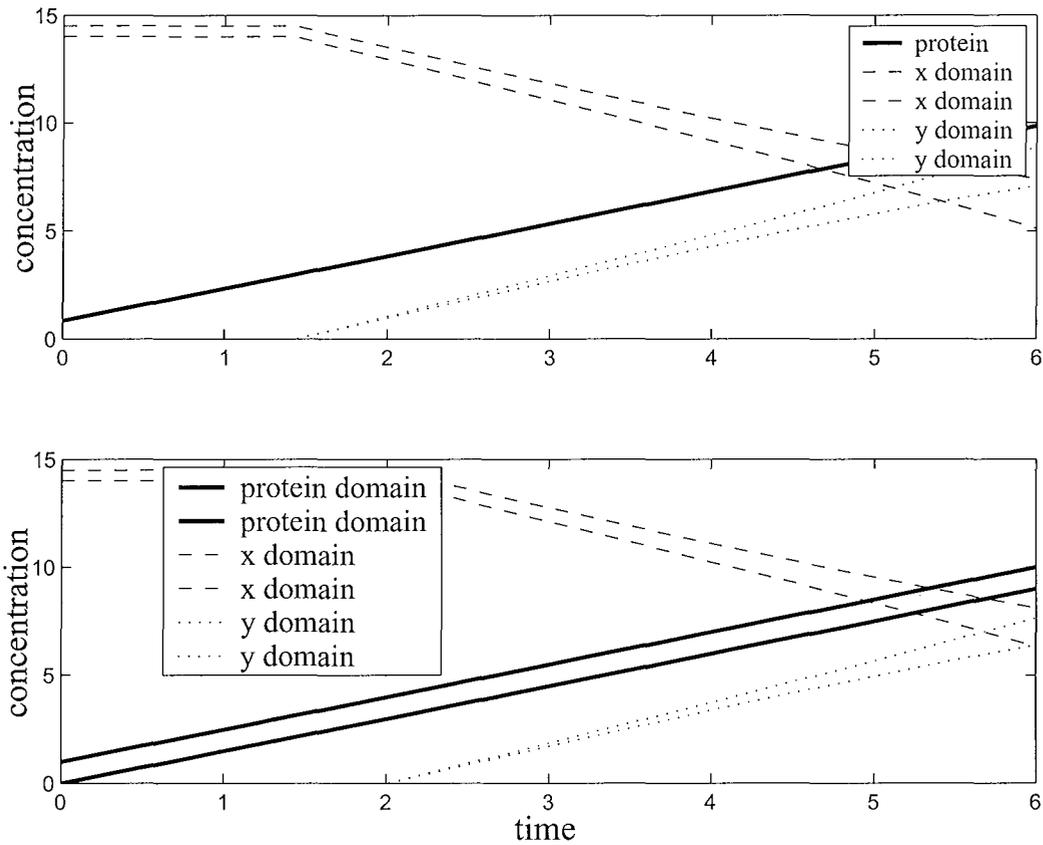


Fig. 8. Switching behaviour for conditional (top) and default combinator (bottom).

Here, we have assumed that there is no uncertainty on the initial value of p . Without this hypothesis, the constraint solver cannot decide between the two alternatives $p \geq 1$ and $p < 1$. In order to handle conditions in presence of uncertainty, we use default reasoning that we describe next.

4.4 Default behavior

The default combinator `if c else A` (or `unless(c) A`) expresses that A holds now, if c will not hold now. Operationally, this means that the current store on quiescence does *not* entail c . Note that `unless(c) A` is not equivalent to `if $\neg c$ then A` . If A is executed, this may have two reasons:

- The current store entails $\neg c$ (in this case `unless(c) A` behaves like `if $\neg c$ then A`), or
- the current store neither entails c nor $\neg c$, i.e., it is not known whether or not c holds. In this case, A is executed *by default*.

A is not executed, if the current store entails c .

We use the same example as before. The only difference is that the variable p representing the protein concentration is initialized with the interval $[0, 1]$. As we can see in Fig. 8, the reaction gets activated when the *lower* bound for p reaches the threshold.

```

interval x, y, p;
x=[14,14.5];
y=0;
p=[0,1];
always {
  p' = 1.5;
  if ( p >= 3 )
    { x' = -(Amax*x)/(ks+x);
      y' = (Amax*x)/(ks+x);
    }
  unless (p >= 3)
    { x' = 0;
      y' = 0;
    }
}
sample(p, x ,y);

```

The default combinator is a convenient way of handling incomplete knowledge in biology. In particular, we will use it in our multi-site model of alternative splicing regulation in Section 5.2.

5 Modeling the alternative splicing regulation with Hybrid cc

5.1 Single-site model: local modeling

The single-site model from Section 3.2 with experimental values can be expressed directly in Hybrid cc.

```

# define Pese 0.01           # define kr 0.01
# define Psc 0.2             # define k 0.19
# define Pr 0.4              # define kk 0.01
# define kese 0.35          # define SC 2
# define ksc 2               # define ASF 1.75
# define R 0.35

interval t, rna, irna, mrna;
t=0; rna = 0.06; irna = 0; mrna = 0;

```

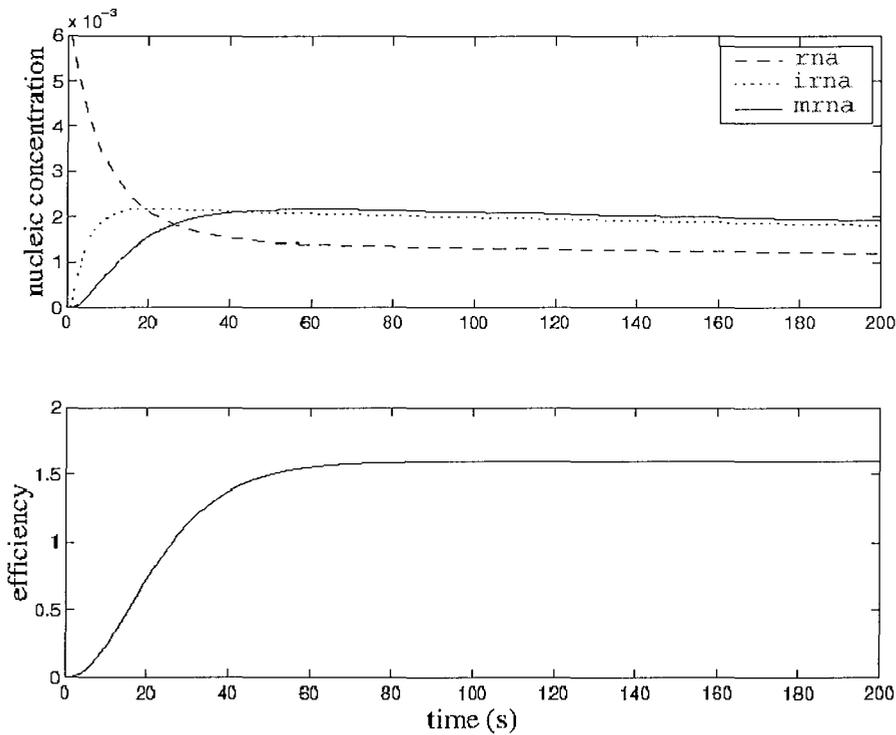


Fig. 9. Variation of RNA pool and splice efficiency in the splicing reaction

```

always{
  rna' = (Pr*R*irna)/(kr*(1+(SC/ksc))+R)
        -(Pese*(ASF+SC)*rna)/(kese+ASF+SC)
        -delta*rna;
  irna' = (Pese*(ASF+SC)*rna)/(kese+ASF+SC)
        -(Pr*R*irna)/(kr*(1+(SC/ksc))+R)
        -k*irna+kk*mrna-delta*irna;
  mrna' = k*irna-kk*mrna-delta*mrna;
}
sample(rna, irna, mrna);
    
```

During the simulation, we obtain the predicted equilibrium for the splice efficiency, see Fig. 9. Under our hypotheses, which include protein competition and compensation, the model correctly simulates the alternative splicing activity at site A3. This supports the hypotheses made in the model such as the role of the ESE and ESS2 binding sites.

5.2 Three-site model: global modeling

A realistic model of alternative splicing has to reflect the combinatorial complexity discussed in Section 2.3. Assuming that regulation is modular (Hartwell et al., 1999), the single-site model may be seen as one module inside a larger

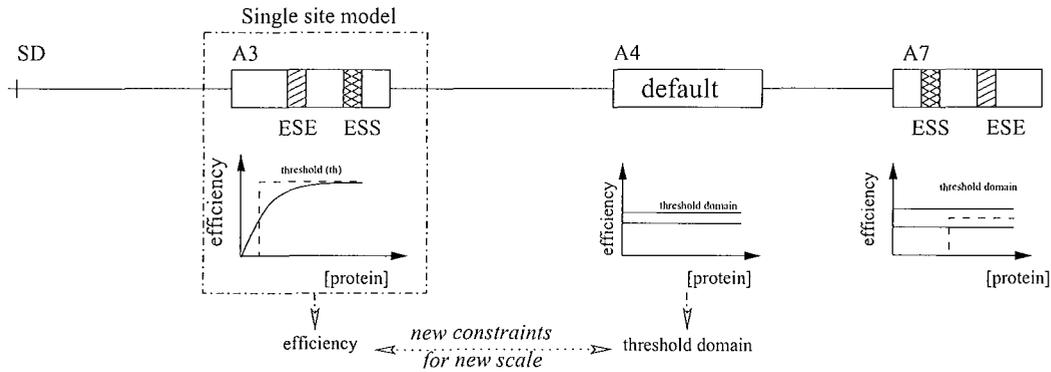


Fig. 10. Single-site model inside a more general multi-site regulation model

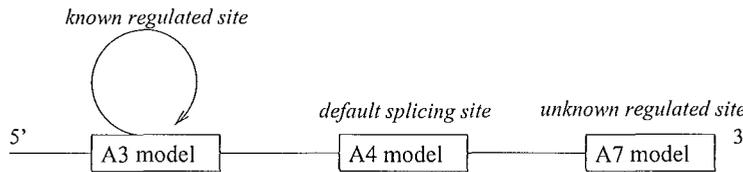


Fig. 11. Biological information on three acceptor sites A3, A4, A7

framework. The qualitative validation given in Section 3.3 justifies the introduction of the single-site model into a larger-scale model involving several splicing sites. To illustrate this, we consider the generic example of three acceptor sites (A3, A4 and A7) associated with one donor site (SD), see Fig. 10.

The behavior at one splicing site can be captured by a single function, the splice efficiency, which depends on the protein concentrations. This function is used in a larger-scale global model that describes the choice between three acceptor sites A3, A4 and A7. In the HIV-1 case, the A4 site is the default splicing site. Only if the efficiency of A3 (effA3) or A7 (effA7) gets larger than the efficiency of A4 (effA4), regulation switches to the other state. The sites A3, A4, and A7 exhibit three generic behaviors, see also Fig. 11:

- A3 is a regulated site with known behavior.
- A7 is a regulated site with unknown behavior.
- A4 is an unregulated site, i.e., the behavior does not depend on protein concentrations.

Current biological experiments give information on the local behavior at one site. However, modeling the local behavior is not enough. In order to understand the global splicing process, we must integrate several types of knowledge. On the one hand, we have information on the local behavior at individual acceptor sites. On the other hand, we have some information on the global behavior, like the default role of A4 or the competition between different acceptor sites. Constraint programming allows us to integrate this information, and to produce a global model.

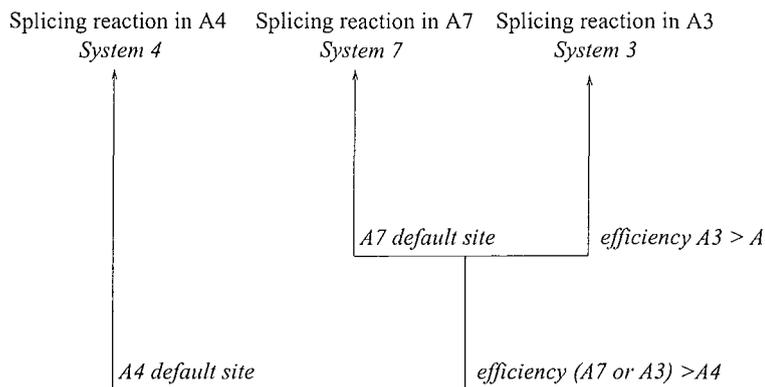


Fig. 12. Choice of the acceptor site A3, A4 or A7 depending on the splice efficiency

Recent work (Dautry and Weill, 2002) shows the linearity of the splicing kinetics. Thus, on the larger scale, we may consider splicing as a linear process described by three systems of ordinary differential equations. For each acceptor site A_i , $i \in \{3, 4, 7\}$, we introduce one system with four differential equations:

- r_{i1} represents the consumption of immature RNA if A_i is dominating.
- r_{i3} represents the production of mature RNA at A3.
- r_{i4} represents the production of mature RNA at A4.
- r_{i7} represents the production of mature RNA at A7.

k_{ij} is the kinetic constant for reaction r_{ij} .

A4 is the default splicing site. It is dominating unless the splice efficiency of A3 or A7 gets larger than the splice efficiency of A4. If this happens, A7 becomes the default splicing site unless the efficiency of A3 gets larger than the efficiency of A7, see Fig. 12. The local behavior at A3 has been described by the single-site model given in Section 5.1. This model predicts the splice efficiency of A3 depending on the protein concentrations.

In the Hybrid cc program given below, the concentration of SC35 is increased linearly. Depending on the corresponding variation of the splice efficiency at A3, the three-site model exhibits different behaviors, characterized by the choice of one of the three differential equation systems. The default behavior discussed before can be expressed naturally in Hybrid cc using the combinator `unless(c) A`.

```
#define c1 2
#define c2 0.5
#define c3 0.8
#define c4 1
#define c5 0.9
```

```

#define c6 0.1
#define R 3.5

interval t, prot, effA3, effA4, effA7, rna, mrnaA3,
    mrnaA4, mrnaA7;
t = 0;
rna = 10;
mrnaA3 = 0; /*known regulated acceptor site */
mrnaA4 = 0; /*unregulated acceptor site*/
mrnaA7 = 0; /*unknown regulated acceptor site*/

always { t' = 10;
prot = 0.1*t;          /* protein variation */

/* A3 efficiency depends on the protein concentration
   A3 efficiency represents the local behavior of A3
   (observer of A3) */
effA3 = c1*(prot+c2)*(c3*prot+c4)/(c5*(prot+c6));

6 <= effA4; effA4 <= 8; /* effA4 : efficiency domain of A4*/
7 <= effA7; effA7 <= 9; /* effA7 : efficiency domain of A7*/
}

/* The behavior depends on the efficiency of
   the 3 acceptor sites*/

always {
/* if A3 or A7 dominant */
if (effA3 >= effA4 || effA7 >= effA4) {
    if (effA7 <= effA3) { /* splicing on A3 */
        rna' = -0.51 * rna - 0.01*rna;
        mrnaA3' = 0.4 * rna - 0.1*mrnaA3; /* A3 kinetics */
        mrnaA4' = 0.01 * rna - 0.1*mrnaA4; /* A4 kinetics */
        mrnaA7' = 0.1 * rna - 0.1*mrnaA7; /* A7 kinetics */
    };
    unless ((effA7 <= effA3)) { /*default splicing on A7*/
        rna' = -0.51 * rna - 0.01*rna;
        mrnaA3' = 0.1 * rna - 0.1*mrnaA3; /* A3 kinetics */
        mrnaA4' = 0.01 * rna - 0.1*mrnaA4; /* A4 kinetics */
        mrnaA7' = 0.4 * rna - 0.1*mrnaA7; /* A7 kinetics */
    };
}
/* default splicing on A4 */
unless (effA3 >= effA4 || effA7 >= effA4) {
    rna' = -0.32 * rna - 0.01*rna;
}
}

```

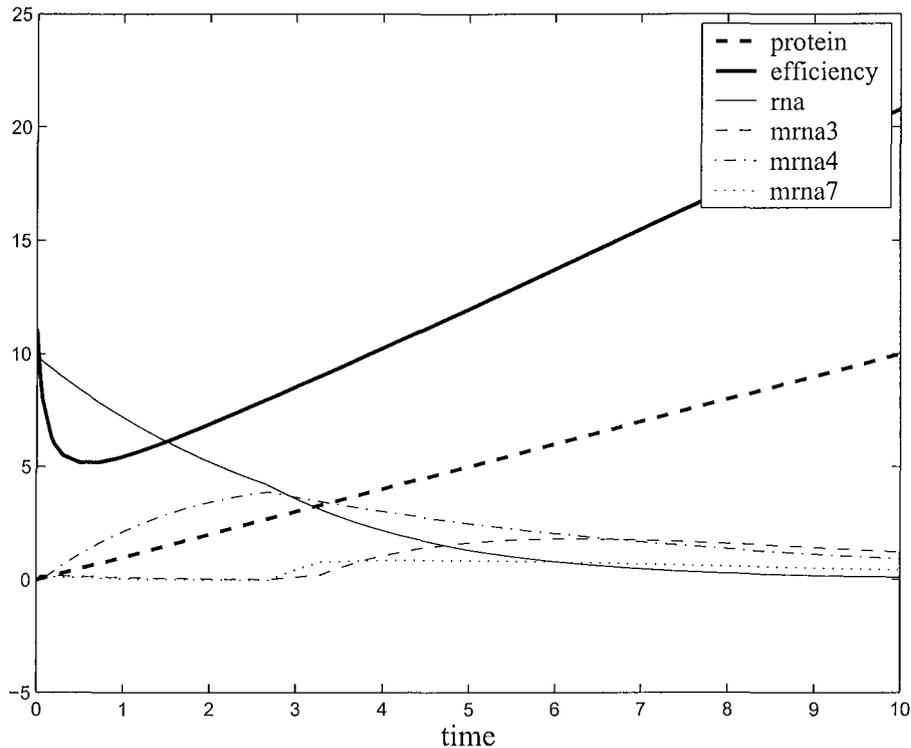


Fig. 13. Variation of mRNA production depending on variation of SR proteins

```

mrnaA3' = -0.01 * rna - 0.1*mrnaA3; /* A3 kinetics */
mrnaA4' = 0.3 * rna - 0.1*mrnaA4; /* A4 kinetics */
mrnaA7' = -0.01 * rna - 0.1*mrnaA7; /* A7 kinetics */
}
}
sample(prot, effA3, rna, mrnaA3, mrnaA4, mrnaA7)
    
```

According to the semantics of the default combinator, the A4 site will be chosen if the solver cannot deduce that $(\text{effA3} \geq \text{effA4})$ or $(\text{effA7} \geq \text{effA4})$. This may have *two* reasons:

- $(\text{effA3} \geq \text{effA4})$ or $(\text{effA7} \geq \text{effA4})$ is false, i.e., $(\text{effA3} < \text{effA4})$ and $(\text{effA7} < \text{effA4})$, or
- it is not known whether $(\text{effA3} \geq \text{effA4})$ or $(\text{effA7} \geq \text{effA4})$ holds (default behavior).

Thus the A4 site is the default site if the splice efficiency of A3 and A7 is not sufficiently high. If A3 or A7 dominate A4, then A7 is the default splicing site, unless A3 dominates A7.

Simulation in Hybrid cc yields the behavior shown in Fig. 13. First mrnaA4 is produced, i.e., the default site A4 is active. When effA3 passes the upper threshold for effA4 , site A7 gets activated, and mrnaA7 is produced. Finally,

when `effA3` further increases and passes the upper threshold for `effA7`, site A3 gets activated and we observe production of `mrnA3`, while the concentrations of `mrnA4` and `mrnA7` become stationary.

Basically, the model gives to the biologist three qualitative states : first a splicing at the A4 site, second a splicing at the A7 site, and finally a splicing at the A3 site. The constraint programming system can extract reachability regions for the three biological states, despite the variation in the concentration of SR proteins. The reachability region is an important qualitative information to extend the single-site to a multi-site model. `Hybrid cc` permits a qualitative validation of the model, although the currently available information on the alternative splicing regulation in HIV-1 is incomplete.

6 Conclusion and further research

Our approach combines mathematical and computational methods. Mathematical analysis allows us to validate the single-site model in a qualitative way, based on the experimental data obtained in our group. The validation shows the consistency of our biological hypotheses. In a second step, we can extract the splice efficiency as a time-scale abstraction of the local behavior at one site inside a more global model involving different sites. For the experimental biologist, the single-site model may serve as a computational tool to evaluate his knowledge on a fine-grained biological process.

On the computational side, the constraint solving and default reasoning capabilities of `Hybrid cc` allow us to exploit as much as possible the incomplete knowledge of our system. Default behavior may compensate the lack of experimental data. Using constraint programming, we can delimit with our model the possible splicing behavior. This provides a powerful tool for qualitative validation.

Combining mathematical analysis and computational methods is the key to extending the single-site model to a multi-site model as described in this paper. It leads to the qualitative validation represented by the extraction of the splice efficiency function. The splice efficiency characterizes the modularity of the regulation. Thus, the one-site behavior is represented in the three-site model, based on the single-site splice efficiency. The extraction of a suitable criterion on the smaller scale is crucial to understanding an experimental process from a systems biology perspective. Furthermore, constraints can be used to handle the problem of missing data in time-scale abstraction of a single-site model in a more global multi-site model. Different scales usually correspond to biological experiments yielding different types of results. Despite the variety of possible experiments, these must be integrated into a global model in order to better

understand the biological process.

Modeling alternative splicing requires a close interaction between biological and computational approaches. In the context of alternative splicing regulation, we are currently working on new experimental data for the quantitative validation of our models. On the computational side, we have integrated our model into a general HIV-1 model (Hammond, 1993). Preliminary results show that the modification of a splice constant may induce different behaviors in the HIV-1 life cycle model. Using the extended model, we may validate several biological hypotheses on the global effect of alternative splicing in the full HIV-1 life cycle.

References

- Bernard, O., Gouzé, J.-L., 1999. Nonlinear qualitative signal processing for biological systems: application to the algal growth in bioreactors. *Mathematical Biosciences* 157, 357–372.
- Bockmayr, A., Courtois, A., 2001. Modeling biological systems in hybrid concurrent constraint programming (Abstract). In: 2nd Int. Conf. Systems Biology, ICSB'01, Pasadena, CA. p. 106.
- Bockmayr, A., Courtois, A., 2002. Using hybrid concurrent constraint programming to model dynamic biological systems. In: 18th International Conference on Logic Programming, ICLP'02, Copenhagen. Springer, LNCS 2401, pp. 85–99.
- Caputi, M., Mayeda, M., Krainer, A., Zahler, A., 1999. hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO* 18 (14), 4060–4067.
- Carlson, B., Gupta, V., 1998. Hybrid cc and interval constraints. In: Hybrid Systems: Computation and Control, HSCC'98. Springer, LNCS 1386, pp. 80 – 95.
- Dautry, F., Weill, D., 2002. Kinetic analysis of mRNA metabolism. In: Interdisciplinary School on Imaging, Modelling and Manipulating Transcriptional Regulatory Networks. Ambleteuse, p. 20.
- Del Gatto-Konczak, F., Olive, M., Gesnel, M., Breathnach, R., 1999. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol. Cell. Biol.* 19 (1), 251–260.
- Graveley, B., 2000. Sorting out the complexity of SR protein functions. *RNA* 6, 1197–1211.
- Gupta, V., Jagadeesan, R., Saraswat, V., 1998. Computing with continuous change. *Science of computer programming* 30 (1-2), 3–49.
- Gupta, V., Jagadeesan, R., Saraswat, V., Bobrow, D. G., 1995. Programming in hybrid constraint languages. In: Hybrid Systems II. Springer, LNCS 999, pp. 226–251.

- Hammond, B., 1993. Quantitative study of the control of HIV-1 gene expression. *J. Theor. Biol.* 163, 199–221.
- Hartwell, L., Hopfield, J., Leibler, S., Murray, A., 1999. From molecular to modular cell biology. *Nature* 402, C47–C52.
- Heinrich, R., Schuster, S., 1996. *The regulation of cellular systems*. Thomson Publishing, New York.
- Moore, M., Query, C., Sharp, P., 1993. Splicing of precursors to mRNA by the spliceosome. In: *The RNA World*. Cold Spring Harbor Laboratory Press, pp. 303–357.
- O'Reilly, M., McNally, M., Beemon, K., 1995. Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology* 213 (2), 373–385.
- Palsson, B., 2000. The challenges of in silico biology. *Nature Biotechnology* 18, 1147 – 1150.
- Purcell, D., Martin, M., 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* 67 (11), 6365–6378.
- Ropers, D., Ayadi, L., Gattoni, R., Jacquenet, S., Damier, L., Branlant, C., Stévenin, J., 2003. 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 revealed by in vitro and ex vivo splicing experiments. Submitted to *J. Biol. Chemistry* .
- Saraswat, V. A., 1993. *Concurrent constraint programming*. MIT Press.
- Saraswat, V. A., Jagadeesan, R., Gupta, V., 1994. Foundations of timed concurrent constraint programming. In: *9th Symp. Logic in Computer Science, LICS'94*, Paris. IEEE, pp. 71 – 80.
- Saraswat, V. A., Jagadeesan, R., Gupta, V., 1996. Timed default concurrent constraint programming. *Journal of Symbolic Computation* 22 (5/6), 475–520.
- Segel, L., 1984. *Modelling dynamic phenomena in molecular and cellular biology*. Cambridge University Press.
- Tang, H., Kuhlen, K., Wong-Staal, F., 1999. Lentivirus replication and regulation. *Annu. Rev. Genet* 33, 133–170.
- Voit, E. O., 2000. *Computational analysis of biochemical systems*. Cambridge Univ. Press.

Discussion et perspectives

L'épissage étant une étape clé du cycle viral, la compréhension des mécanismes impliqués dans sa régulation devrait participer à l'obtention d'une meilleure maîtrise du cycle viral. Elle pourrait aussi permettre d'envisager le développement d'approches thérapeutiques, visant à utiliser, comme cibles, les mécanismes spécifiques de régulation de l'épissage de l'ARN du virus HIV-1. Il est donc important d'étudier, de manière détaillée, les processus de régulation de l'épissage.

Les études poussées de l'efficacité des sites d'épissage, menées au laboratoire, par L. Damier (Damier 1997) et S. Jacquenet (Jacquenet 2001), ainsi que par l'équipe de K. Beemon (O'Reilly *et al.* 1995), avaient permis de montrer que les régulations majeures s'exercent sur les sites 3' d'épissage. En effet, ces sites sont faibles et localisés dans des régions structurées, ce qui module leur efficacité d'utilisation. Ils sont contrôlés par des éléments activateurs et inhibiteurs d'épissage.

Jusqu'à-là, à l'exception du site A7, on avait surtout une information sur les éléments inhibiteurs. Notre travail apporte des données sur les éléments activateurs *in cis* et les protéines les reconnaissant. Nous avons en effet montré une action forte et différentielle des protéines SR sur certains des sites d'épissage.

En plus du rôle général que jouent les protéines SR dans toute réaction d'épissage, qu'il s'agisse d'un épissage constitutif ou régulé, nos données montrent que les protéines SR ont un effet différentiel sur l'utilisation des différents sites d'épissage de l'ARN du virus HIV-1.

Au total, trois sites sont la cible de fortes régulations (nos données) et A7 (les données de Marchand *et al.* au laboratoire et d'autres équipes). Il est intéressant de noter que deux de ces sites, A3 et A7, sont utilisés pour la production de l'ARNm *tat*, dont le produit de traduction joue un rôle majeur dans l'initiation d'un cycle de multiplication virale. Nos données mettent en évidence un rôle prédominant de la protéine SC35 au site A3. Le site de fixation de la protéine ASF/SF2, que nous avons mis en évidence et qui fixe aussi SC35, semble plus servir à contrecarrer la forte activation du site A2 par la protéine ASF/SF2, qu'à générer une activation de A3 par cette protéine. Pourtant au niveau du site A3, la fixation d'ASF/SF2 à l'élément activateur ESEt, que nous avons identifié, ne semble pas se heurter à une forte compétition par une protéine hnRNP A1, alors qu'au site A2, nos données suggèrent que le mécanisme d'activation par ASF/SF2 repose sur sa capacité à déplacer la protéine hnRNP A1 fixée sur le site inhibiteur ESSv (Bilodeau *et al.* 2001). Ceci, ajouté au fait que l'effet des protéines SR que nous avons observé *ex vivo*, est très similaire, lorsque ces protéines possèdent leur domaine RS ou lorsque ce dernier a été tronqué, suggère très fortement que l'action majeure des protéines SR, dans les régulations que nous avons observées, repose sur leur capacité à entrer en compétition avec la fixation des protéines hnRNP, en particulier hnRNP A1.

Deux mécanismes alternatifs de régulation de l'utilisation des sites 3' par les protéines SR avaient été proposés :

- simplement favoriser la fixation des facteurs d'épissage tels que U2AF, possédant aussi un domaine RS, par interaction de ce domaine avec celui de la protéine SR fixée sur l'élément activateur ;
- contrecarrer l'effet négatif des protéines hnRNP.

Dans le cas de HIV-1, aussi bien les données que nous présentons, que celles du laboratoire sur le site A7, suggèrent que le second mécanisme joue un rôle majeur et définit quels sont les sites potentiels fortement stimulés aux dépens des autres, en présence d'une protéine SR donnée. Les autres

sites activateurs, tels que l'élément ESEt, que nous avons mis en évidence, pourraient préserver un niveau minimal d'utilisation des sites en compétition, en conditions de forte activation d'un autre site, dans le cas d'ARN soumis à un épissage alternatif complet, comme c'est le cas pour HIV-1. C'est une notion nouvelle pas encore énoncée, car peu d'études de régulation d'épissage aussi complexes que celles de HIV-1 n'avaient été menées jusque-là. Les études réalisées par transfection *ex vivo* portaient, jusque-là, sur un nombre beaucoup plus limité de sites d'épissage en compétition les uns avec les autres. Par ailleurs, le fait qu'un grand nombre d'équipes avaient travaillé sur la régulation de l'épissage de l'ARN du virus HIV-1 permet d'avoir une connaissance approfondie de l'ensemble du système. Ainsi, l'épissage de l'ARN de HIV-1 nous semble être un modèle de choix pour une analyse globale des paramètres intervenant dans le processus d'épissage alternatif.

Un autre point qui ressort de notre étude, corollaire au précédent, est que les protéines hnRNP, et en particulier la protéine hnRNP A1, jouent un rôle majeur dans la régulation de l'épissage de l'ARN du virus HIV-1, et vraisemblablement ailleurs. D'autres équipes l'ont déjà proposé, pour l'épissage alternatif des ARN pré-messagers cellulaires. Dans le cas du virus HIV-1, la protéine hnRNP A1 intervient au niveau des 3 sites fortement régulés A2, A3 et A7. Les régulations observées reposent sur sa capacité à limiter l'utilisation de ces sites en se fixant à des éléments inhibiteurs.

Une grande question reste posée à ce niveau : comment en se fixant à ces éléments inhibiteurs, la protéine hnRNP A1 limite-t-elle les capacités d'assemblage des complexes spliceosomaux ? Les travaux de V. Marchand au laboratoire, en appui sur des travaux publiés auparavant par d'autres équipes, apportent une réponse pour le site A7. A savoir que, dans ce cas, la protéine se lie à la fois à des éléments introniques et exoniques et que via des interactions protéine/protéine entre le domaine Gly de différentes protéines hnRNP A1, la fixation aux différents sites se fait de manière coopérative et conduit à une complète occlusion du site A7 au sein d'un complexe de protéines hnRNP A1.

Dans le cas des sites A2 et A3, que nous avons étudiés, les études n'ont, pour le moment, pas encore mis en évidence de sites de fixation de la protéine hnRNP A1 au niveau intronique. Il nous paraît important que les études futures visent à déterminer s'il existe aussi de tels sites introniques, conduisant à une situation similaire à celle observée au site A7, ou si les mécanismes très forts d'inhibition de l'utilisation des sites A2 et A3 par la protéine hnRNP A1 relèvent d'autres mécanismes. On pourrait imaginer une interaction avec d'autres protéines hnRNP fixées sur l'ARN, avec toujours un phénomène de stabilisation de l'ensemble par effet coopératif, un recrutement de protéines à domaine Gly ayant un effet inhibiteur, ou une action directe négative sur la fixation de certains facteurs d'épissage. Cette hypothèse est plus difficile à envisager pour un élément inhibiteur tel que l'élément ESS2, situé loin du site A3. Néanmoins, comme nous l'avons déjà mentionné, l'élément ESS2 est situé dans une longue structure tige-boucle, très flexible du fait de nombreuses boucles internes. Ceci permet d'envisager que la ou les protéines fixées sur l'élément ESS2 puissent interagir avec des protéines liées au site A3.

Donc au site A3, nos données suggèrent fortement que SC35 entre en compétition avec hnRNP A1. L. Ayadi apporte actuellement les preuves expérimentales définitives par expériences de pontage, mais en plus, les expériences futures devront viser à comprendre comment la protéine hnRNP A1 a un effet inhibiteur.

Une technique nouvelle va être utilisée en vue d'apporter des éléments d'information sur ce point,

ceci, en collaboration avec l'équipe de R. Lührmann (Max Planck Institut, Göttingen). Cette technique consiste à insérer dans un ARN, pour lequel on recherche les partenaires protéiques, une séquence qui permettra de retenir le complexe ARN/protéines sur colonne. Nous utiliserons une séquence aptamérique qui a été identifiée par l'approche SELEX comme ayant une forte affinité pour un antibiotique de type aminoglycoside, la tobramycine (Bachler *et al.* 1999). Cette séquence sera fixée à l'extrémité 5' ou 3' de l'ARN. L'ARN étudié (dans notre cas, soit la structure tige-boucle 3 renfermant l'élément ESS2, sauvage ou muté, soit l'ensemble du fragment allant du site A3 au site D4, ou de plus grands fragments de l'ARN de HIV-1) seront fusionnés à la séquence aptamérique. Après incubation des ARN fusionnés dans un extrait nucléaire de cellules HeLa, les complexes formés seront retenus sur une colonne où la tobramycine aura été fixée, et leur contenu en protéines sera étudié par électrophorèse, couplée à la spectrométrie de masse. L'équipe de R. Lührmann a déjà utilisé cette approche pour analyser le contenu en protéines de complexes spliceosomaux, formés sur un intron épissé de manière constitutive (Hartmuth *et al.* 2002). Le modèle HIV-1 apportera des informations sur ce qui se passe dans un système régulé.

L'emploi de cette approche nous permettra de connaître quelles sont les autres protéines nucléaires qui se fixent à la région de l'ARN de HIV-1 contenant le site A3 et ses éléments de régulation. Cette information est indispensable pour pouvoir répondre à la question : comment la protéine hnRNP A1, en se fixant à l'élément ESS2, peut inhiber l'utilisation du site A3, situé 70 nts en amont ?

Toujours dans le cadre de l'étude des sites A2 et A3, une autre question générale que soulève nos données, est la possible fixation coopérative des protéines SR. Alors que les capacités de fixation de la protéine hnRNP A1 sont étayées par de nombreuses données expérimentales, le problème a été peu étudié dans le cas des protéines SR. Nos données suggèrent l'existence de plusieurs sites de fixation de la protéine ASF/SF2 au niveau de l'élément ESSv régulant le site A2. C'est aussi le cas pour le site A7 (Marchand *et al.* 2002). Nos données démontrent la présence de deux sites de fixation de la protéine SC35 en aval du site A3. Il paraît très important de tester si la fixation de ces protéines SR à plusieurs sites voisins se fait de manière coopérative, ce qui pourrait renforcer leur capacité à contrecarrer l'effet de la protéine hnRNP A1.

Un élément négatif vis-à-vis de cette hypothèse est, tout de même, notre observation que les protéines SR, privées de leur domaine RS, que l'on pouvait supposer intervenir dans de tels phénomènes de coopérativité, ont des effets très similaires à ceux des protéines entières *ex vivo*. En contrepartie, il faut tout de même noter que, bien que les effets que nous avons observés vont dans le même sens pour les protéines avec ou sans domaine RS, ils sont tout de même atténués en leur absence.

Enfin, toujours dans le cadre de ces phénomènes de coopérativité et de l'action négative de la protéine hnRNP A1, nos observations sur l'effet de mutations sélectionnées dans des ARN de HIV-1, isolés de cellules chroniquement infectées sont très intéressantes. En effet, il apparaît que deux substitutions de bases (l'une dans l'élément ESS2p, l'autre, en amont de l'élément ESEt), sont susceptibles de générer de nouveaux sites de fixation de la protéine hnRNP A1 et donc, de favoriser sa polymérisation le long de l'ARN, entraînant une inhibition presque complète de l'utilisation du site A3.

Ces données confortent la notion d'une action de la protéine hnRNP A1, basée sur un mécanisme de multimérisation. En plus, elles montrent qu'un nombre très limité de mutations exoniques peuvent ainsi modifier très fortement les régulations de l'épissage alternatif. Ceci est en accord avec l'obser-

vation, au niveau de maladies génétiques, de nombreuses mutations silencieuses, au niveau traductionnel, dans les exons des ARN pré-messagers cellulaires (Cáceres & Kornblihtt 2002). Ces données soulèvent aussi le problème d'un rôle vraisemblable des phénomènes d'édition, à la fois de l'ADNc simple-brin et de l'ARN, dans les mécanismes de résistance des cellules.

Nos travaux sur des virus très éloignés, SIVcpz ANT et HIV-1 YBF30, apportent aussi un certain nombre de notions générales intéressantes.

La 1^{re} est que la régulation de l'épissage alternatif est un paramètre important dans le cycle viral. Cette importance se traduit par le fait qu'en dépit des nombreuses mutations que subissent les ARN rétroviraux, ces régulations peuvent être conservées, comme nous l'avons montré pour l'activation du site A3 par la protéine SC35. La 2^e notion est que la conservation de l'activation d'un site d'épissage donné, par une protéine SR donnée, n'implique pas la stricte conservation du mécanisme de régulation. Du fait des mutations fréquentes, un site inhibiteur va pouvoir avoir un effet diminué, mais un autre aura alors un effet renforcé ou sera généré. Il en sera de même pour les éléments activateurs, mais à la fin, l'effet global, négatif, de la protéine hnRNP A1 va être conservé, et celui, positif, de la protéine SR, également conservé, de manière à préserver la régulation. Les rétrovirus se prêtent tout particulièrement à ce type d'analyse, vue leur vitesse de mutation.

Cette conservation de l'activation du site A3 par la protéine SC35 doit être mise en liaison avec une observation de l'activation du site A7 par cette même protéine (Mayeda *et al.* 1999). Les deux sites sont requis pour produire l'ARNm *tat* et SC35 serait surexprimée tôt après infection de la cellule (Maldarelli *et al.* 1998). Non seulement, il y a régulation des sites d'épissage de l'ARN du virus HIV-1, mais cette régulation doit avoir une implication au niveau de la régulation temporelle des protéines virales. D'ailleurs, ceci explique sans doute d'ailleurs pourquoi ces régulations sont si conservées. L'effet de la protéine SC35 au site A7 n'a pas été étudié de manière très détaillée. Il paraîtrait important de reprendre cette étude, à la lumière des données que nous avons obtenues au site A3 et ceci, aussi, en vue de faire progresser notre travail de modélisation mathématique.

En effet, comme nous venons de le dire, du fait d'un grand nombre de données et de la possibilité de faire une étude globale comme nous l'avons faite, l'épissage alternatif de l'ARN de HIV-1 se prête bien à la construction d'un modèle global, essayant de prendre en compte l'ensemble des paramètres majeurs régissant la régulation. Dans ce cadre, il sera important de mieux définir la situation au site A1. Existe-t-il des éléments de régulation *in cis* à ce site, ou est-ce que l'effet positif des protéines SR, assez large *in vitro*, mais d'amplitude modérée *ex vivo*, que nous avons observé, repose simplement sur l'activation de l'assemblage des complexes spliceosomaux que les protéines SR exercent sur tout couple de sites d'épissage ?

Les sites A4a, b et c, dont l'un au moins, doit être utilisé pour produire la protéine Rev, semblent seulement beaucoup moins activés par la protéine ASF/SF2. La faible utilisation de ces sites *in vitro* soulève des questions, dans la mesure où Rev est une protéine essentielle, qui doit être exprimée en quantités suffisantes pour jouer son rôle dans le transport des ARN non épissés. En effet, elle agit aussi en multimérisant sur l'élément RRE et son action dépend de sa concentration (Hope 1999). Là-aussi, il reste encore du travail à faire, pour mieux comprendre ce qui se passe à ces sites. Comme

nous l'avons déjà dit, pour réaliser une telle étude *in vitro*, il est nécessaire de détruire les sites A3 et A5, qui entrent trop fortement en compétition avec les trois sites A4.

Comme nous l'avons aussi dit, l'optimisation des modèles mathématiques déjà établis nécessitent aussi d'autres types d'expériences, visant à mieux définir les caractéristiques cinétiques des sites, ceci en fonction des concentrations relatives en protéines SR.

Enfin, nous avons beaucoup raisonné sur les concentrations de protéines SR, mais ce qui est aussi très important pour leur activité, de même que pour les protéines hnRNP, est leur niveau de phosphorylation. On peut très bien avoir des variations fortes d'activité des protéines SR, sans pour autant avoir de variations de leur concentration. C'est aussi un champ d'étude qu'il faudra aborder, en tenant compte du fait que certaines des protéines virales, comme Tat, peuvent avoir un effet possible, à la fois sur la transcription des gènes des protéines SR, mais aussi, sur ceux des kinases qui les phosphorylent.

Nous avons donc apporté une contribution dans l'étude de la régulation de l'épissage de l'ARN du virus HIV-1 qui, nous l'espérons, sera une base solide pour la suite des études qui devront être nombreuses pour répondre à toutes les questions posées.

Troisième partie

Matériel et méthodes

Chapitre VIII

Constructions génétiques

L'ADNc de plusieurs souches virales a été utilisé au cours de ce travail. Nous avons ainsi disposé du plasmide pBRU3, qui nous a été fourni par F. Clavel et qui comporte la séquence entière du génome de la souche BRU/LAI du virus HIV-1 (numéro d'accessibilité dans la GenBank : K02013) des nucléotides +1 à +9229, +1 correspondant au premier nucléotide de la séquence R située à l'extrémité 5' du génome (Charneau *et al.* 1992). Des fragments d'ADNc contenant la région des sites d'épissage A3 à A5 des souches HIV-1 YBF30 (AJ006022; nts 5194 à 5647) et SIVcpz ANT (U42720; nts 5047-5495) nous ont été fournis par S. Saragosti (INSERM U 552, Paris).

VIII.1 Constructions renfermant des séquences non contiguës du génome de la souche HIV-1 BRU

Par la technique d'amplification par PCR sur le plasmide pBRU3, L. Damier avait construit, au cours de sa thèse au laboratoire, des plasmides pBluescript KSII+ recombinants renfermant des séquences non contiguës du génome du virus HIV-1 BRU. Ces constructions permettent de générer, par transcription *in vitro*, des ARN contenant des couples de sites d'épissage et dont la taille est compatible avec les expériences d'épissage *in vitro*. Les constructions utilisées dans ce travail sont schématisées dans la figure VIII.1.

Selon le même principe, nous avons généré de nouveaux plasmides pBluescript KSII+ recombinants renfermant des séquences non contiguës du génome du virus HIV-1 BRU. Ces nouvelles constructions sont représentées dans la figure VIII.2-A. Les séquences des oligonucléotides utilisés pour réaliser ces constructions sont indiquées dans le tableau VIII.1.

VIII.2 Constructions hybrides renfermant un fragment du génome HIV-1 BRU et un fragment du génome HIV-1 YBF30 ou SIVcpz ANT

Afin d'étudier de nouvelles souches des virus HIV-1 et SIV, nous avons choisi de créer des constructions génétiques hybrides HIV-1 BRU-HIV-1 YBF30 et HIV-1 BRU-SIVcpz ANT. Pour ce faire, nous sommes partis d'un des plasmides pBluescript recombinants construits par L. Damier, le plasmide pLD-L3. Ce plasmide contient, entre les sites *BamH* I et *Bgl* II, un 1^{er} fragment HIV-1 BRU que nous voulions conserver, constitué de l'exon 1 bordé par le site donneur D1, suivi de 95 paires de bases

Nom	Séquence 5' → 3'	Utilisation	Hybridation	Site de restriction ou promoteur
548	GGCAATGAAAGCAAC	RT	HIV-1 BRU, nts 5500-5514	
567	AGAAGACCAAGATCTAGAGGGAGCCA	Construction	HIV-1 BRU, nts 5155-5182	<i>Bgl</i> II
605	TATTCTGGATCCGGTCTCTCTGGT	Construction	HIV-1 BRU, nts 1-12	<i>Bam</i> H I
608	TATTCTCTGCAGTTGCTCTCCTCT	Construction, RT	HIV1 BRU nts 5408-5397	<i>Pst</i> I
954	TATTCTCTGCAGGATCCGGCAATGAAAGC	Construction, RT	HIV-1 BRU, nts 5503-5514	<i>Pst</i> I
998	TCTAATACGACTCACTATAGGGCGACATAGCAGAA-TAGGCG	PCR	HIV-1 BRU, nts 5369-5387	Promoteur T7
1114	GCTTAAGGCTGCAGGTACTACTACTGCTTTGATAGAG-AAACTTG	Construction	HIV-1 BRU, nts 5626-5637	<i>Pst</i> I
1677	TATTCTCTGCAGGGCAATGGTAACAACAG	PCR, RT	SIVcpz ANT, nts 5349-5366	<i>Pst</i> I
1712	TATTCTCTGCAGTCTACGGGGTCCATG	Construction, RT	SIVcpz ANT, nts 5262-5276	<i>Pst</i> I
1714	GCAGTTGTTCTGCTGTT	RT	SIVcpz ANT, nts 5428-5444	
1713	TATTCTCTGCAGTTCCAATCCITCTCCTTT	Construction	HIV-1 YBF30, nts 5426-5409	<i>Pst</i> I
1791	CCGGAATTCAAGAGTACTTACTACTTTGATAC	Construction	SIVcpz ANT, nts 5473 à 5495	<i>Eco</i> R I
1792	TATGGAAGATCTAGAGAGAACCAATGAATGAGTG	Construction	SIVcpz ANT, nts 5047 à 5068	<i>Bgl</i> II
1794	TATGGAAGATCTAGAGGGAACCCTATAATGAATG	Construction	HIV-1 YBF30, nts 5184-5205	<i>Bgl</i> II
1795	CCGGAATTCTGGTTTACTTACTGCTCTGGTAT	Construction	HIV-1 YBF30 nts, 5625-5647	<i>Eco</i> R I
1895	CATATTGGGGTCTGCATACGGCATCTGGGTCAGGG-AGTCTG	Mutagenèse dirigée	HIV-1 BRU, nts 4825-4873	
1896	GAGACTCCCTGACCCAGATGCCGTATGCAGACCCC-AATATG	Mutagenèse dirigée	Complémentaire de l'oligonucléotide 1895	
1897	GGAGTCTCCATAGAATACCCACAAAAGAGATATAG-CACACA	Mutagenèse dirigée	HIV-1 BRU, nts 4872-4912	
1898	TGTGTGCTATATCTCTTTTGTGGGTATTCTATGGA-GACTCC	Mutagenèse dirigée	Complémentaire de l'oligonucléotide 1897	
1901	GGCGTACTCAACAAGCAAGAAATGGAGCC	Mutagenèse dirigée	HIV-1 BRU, nts 5384-5419, délétion des nts 5398-5403	
1902	GGCTCCATTTCTTGCTTGTTGAGTAACGCC	Mutagenèse dirigée	Complémentaire de l'oligonucléotide 1901	
1986	TATTCTCTGCAGTTGCTTGTTGAGTAACGCC	Construction	HIV-1 BRU, nts 5384-5408	<i>Pst</i> I
1988	CGCGATTTAGGTGACACTATAGAAGACTCTGCT-ATAAGAAAGGC	PCR	HIV-1 BRU, nts 4970-4990	Promoteur SP6

Nom	Séquence 5' → 3'	Utilisation	Hybridation	Site de restriction ou promoteur
2028	TATCCGGAATTCTACTTTGATACAGGATTTTG	Construction	SIVcpz ANT, nts 5464-5483	<i>EcoR I</i>
2029	TATCCGGAATTCTCGCAATGGTAACAACAG	Construction	SIVcpz ANT, nts 5350-5365	<i>EcoR I</i>
2030	CCCGTAGACGCTAGACTAGATCCATGGTTACACCC	Mutagenèse dirigée	SIVcpz ANT, nts 5269-5303	
2031	GGGTGTAACCATGGATCTAGTCTAGCGTCTACGGG	Mutagenèse dirigée	Complémentaire de l'oligonucléotide 2030	
2143	GGCTTGCTGAAGCGCGCACGGCAAGAGG	PCR	HIV-1 pNL4-3, nts 700-727	
2144	TTGGGAGGTGGGTTGCTTTGATAGAG	RT-PCR	HIV-1 pNL4-3, nts 8381-8369 et 6041-6029	
2146	CGCGATTTAGGTGACACTATAGAAGATCTGCTGTTT-ATCCATTC	PCR	HIV-1 BRU, nts 5339-5356	Promoteur SP6
2157	CTACTCAACAGAGGAGAGCAAG	Construction	HIV 1 BRU, nts 5388-5408	1/2 site <i>Kpn I</i> en 5', 1/2 site <i>BamH I</i> en 3'
2158	GATCCTTGCTCTCCTCTGTTGAGTAGGTAC	Construction	Complémentaire de l'oligonucléotide 2157	1/2 site <i>BamH I</i> en 5', 1/2 site <i>Kpn I</i> en 3'
2159	GTGTGTTCTAGACACAG	Séquençage d'ADN	pSP65-SP1-ex2 inv, nts 248-264	
2772	GAAAGGCCTTATTCTGACATATAGTTCTCCCTCTGTGTGAATATCAAGC	Mutagenèse dirigée	HIV-1 BRU, nts 4984 - 5032	
2773	GCTTGATATTCACACAGAGGGAGAACTATATGTCAGATAAAGGCCTTTC	Mutagenèse dirigée	Complémentaire de l'oligonucléotide 2772	
2278	GTCTAGGATCTACTGG	RT	HIV-1 BRU, nts 5418-5433	
2302	GATCCTACCTTGTT	RT	HIV-1 BRU, nts 5040-5053	
2775	CGCGATTTAGGTGACACTATAGATCTGTTTACACATT-ATAGAC	PCR	SIVcpz ANT, nts 5216-5235	Promoteur SP6

TAB. VIII.1 – Oligonucléotides utilisés au cours de ce travail.

La séquence des oligonucléotides est précisée, ainsi que leur utilisation et la séquence à laquelle ils s'hybrident. Le site de restriction (indiqué en italiques sur les séquences) ou le promoteur (indiqué en gras) apporté par l'oligonucléotide est également donné.

Nom	Schéma de la construction	Linéarisation
C1		<i>EcoR</i> I
C2		<i>EcoR</i> I
C3		<i>Pst</i> I
C12		<i>EcoR</i> I
L23		<i>Pst</i> I
C3.2001		<i>Pst</i> I
L3		<i>Pst</i> I
L3.957		<i>Pst</i> I
L3.1728-U1		<i>Pst</i> I
L3.957.1728-U1		<i>Pst</i> I

FIG. VIII.1 – Constructions génétiques utilisées au cours de ce travail.

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 *Bgl* II utilisé pour réaliser la jonction des fragments d'ADNc. La flèche horizontale désigne le promoteur transcriptionnel. Dans la construction C3.2001, l'astérisque désigne une mutation ponctuelle C→T en position 5401 de la construction C3. Les sites uniques de restriction utilisés pour produire les transcrits sont indiqués, pour chaque construction, dans la colonne « linéarisation ».

d'intron. Le 2^e fragment HIV-1 BRU a été cloné entre les sites *Bgl* II et *Pst* I, juxtaposé à un site *EcoR* I. C'est ce fragment *Bgl* II-*Pst* I et dans certains cas, *Bgl* II-*EcoR* I, que nous avons remplacé par un fragment d'une des 2 nouvelles souches étudiées, HIV-1 YBF30 ou SIVcpz ANT. Les constructions ainsi réalisées sont représentées dans la figure VIII.2-B. Les séquences des oligonucléotides utilisés pour réaliser ces constructions sont indiquées dans le tableau VIII.1.

(A)	Nom	Vecteur de clonage	ADN matrice	Oligonucléotides utilisés pour l'amplification par PCR	Enzymes de restriction utilisées pour le clonage	Schéma de la construction	Linéarisation
	L3-U1	pLD-L3	L3	567-1114	<i>Bgl</i> II - <i>Pst</i> I		<i>Pst</i> I
	L3.ΔESS2-U1	pLD-L3	L3.957	567-1114	<i>Bgl</i> II - <i>Pst</i> I		<i>Pst</i> I
	C3.ΔESE	pLD-L3	C3	605-1986	<i>Bam</i> HI - <i>Pst</i> I		<i>Pst</i> I
(B)	Nom	Vecteur de clonage	ADN matrice	Oligonucléotides utilisés pour l'amplification par PCR	Enzymes de restriction utilisées pour le clonage	Schéma de la construction	Linéarisation
	SIV-U1	pLD-L3	fragment d'ADNc	1791-1792	<i>Bgl</i> II - <i>Eco</i> R I		<i>Hind</i> III
	SIV.5276	pLD-L3	SIV-U1	605-1712	<i>Bam</i> HI - <i>Pst</i> I		<i>Hind</i> III
	SIV.5365	pLD-L3	SIV-U1	1792-2029	<i>Bgl</i> II - <i>Eco</i> R I		<i>Hind</i> III
	SIV.ΔU1	pLD-L3	SIV-U1	1792-2028	<i>Bgl</i> II - <i>Eco</i> R I		<i>Hind</i> III
	YBF30.U1	pLD-L3	fragment d'ADNc	1794-1795	<i>Bgl</i> II - <i>Eco</i> R I		<i>Hind</i> III
	YBF30.5427	pLD-L3	YBF30-U1	605-1713	<i>Bam</i> HI - <i>Pst</i> I		<i>Hind</i> III

FIG. VIII.2 – Constructions génétiques réalisées au cours de ce travail.

Les vecteurs utilisés pour le clonage, ainsi que les constructions utilisées comme matrice pour l'amplification de fragments par PCR sont indiqués, avec les couples d'oligonucléotides et les enzymes de restriction. Les schémas représentent les fragments d'ADNc de la souche HIV-1 BRU (A) ou des souches HIV-1 YBF30 et SIVcpz ANT (B) insérés dans le vecteur pBluescript KSII+.

VIII.3 Constructions contenant un fragment du génome de l'adénovirus

Au cours de ce travail, nous avons utilisé un certain nombre de constructions réalisées par l'équipe de J. Stévenin, représentées dans la figure VIII.3-A. Ces constructions contiennent un fragment d'ADN de l'adénovirus placé sous contrôle du promoteur transcriptionnel de l'ARN polymérase du phage SP6. Nous sommes partis de l'une de ces constructions (SP1-ex2 inv) pour réaliser la construction SP1-ESEt, dans laquelle le fragment d'ADN compris entre les sites de restriction *Kpn* I et *Bam*H I a été remplacé par un fragment d'ADNc du virus HIV-1 (figure VIII.3-B).

VIII.4 Réalisation de constructions génétiques variantes par mutagenèse dirigée

Un certain nombre de mutants ont été produits à partir des constructions des figures VIII.1 et VIII.2 par la technique de mutagenèse par PCR. Le principe de cette technique repose sur le fait que l'ADN plasmidique sauvage, servant de matrice pour la PCR, a été produit chez *E. coli* XL1-blue (*recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lacF'*, *proAB*, *lacI^qZΔM15*, Tn10 (*tet^r* ; Bullock *et al.* 1987) et porte, par conséquent, des méthylations. Ce plasmide est amplifié par l'ADN polymérase thermostable de l'Archae *Pyrococcus furiosus* (*Pfu*) à partir de deux oligonucléotides complémentaires l'un de l'autre et contenant tous deux la mutation à introduire. Après amplification par PCR, la matrice est dégradée par l'enzyme de restriction *Dpn* I qui reconnaît des sites méthylés (5' GA_mTC 3'), par conséquent, des sites qui sont présents sur le plasmide produit chez *E. coli* mais pas sur le produit PCR. Les produits PCR sont ensuite utilisés pour transformer des bactéries qui produiront de nouveaux plasmides.

Le mélange réactionnel est le suivant :

Plasmide à muter	10 ng
Tampon de PCR 10X (Tris-HCl 200 mM pH=8,0 ; MgCl ₂ 20 mM ; KCl 100 mM ; (NH ₄) ₂ SO ₄ 60 mM ; Triton X100 1% ; BSA 100 mg/ml)	5 μl
Amorce 5'	125 ng
Amorce 3'	125 ng
Mélange des 4 dNTP (1,25 mM chacun)	8 μl
<i>Pfu</i> native (2,5 U/μl) (Stratagène)	1 μl
H ₂ O	qsp 50 μl

Après un cycle de dénaturation de 5 minutes à 95°C, l'amplification se déroule en 18 cycles de 1 minute à 55°C, 8 minutes à 68°C, 30 secondes à 95°C. Les produits PCR sont visualisés sur gel d'agarose à 0,8 %. La matrice est ensuite digérée par 10 U d'enzyme *Dpn* I pendant 1 heure à 37°C. Enfin, les bactéries *E. coli* XL1-blue, rendues compétentes par traitement au CaCl₂ (Sambrook *et al.* 1989), sont transformées avec 1 ou 4 μl de produit PCR traité par la *Dpn* I.

Les oligonucléotides utilisés pour introduire les mutations sont récapitulés dans le tableau VIII.1 et les constructions contenant les mutations sont présentées dans la figure VIII.4. La séquence du plasmide recombinant obtenu a été vérifiée par séquençage réalisé sur l'ADN double brin (Tabor & Richardson 1987). Les techniques générales de biologie moléculaire utilisées pour réaliser les différents clonages sont décrites dans le manuel de laboratoire de Sambrook *et al.* (1989).

Nom	Mutation réalisée à partir de la construction	Oligonucléotides utilisés	Séquence mutée	Linéarisation
C2			ctgcatacaggagaaagagactggcatctg 4836 4865	EcoR I
C2.Δsl3	C2	1895/1896	ctgcatac-----ggcatctg 4836 4865	EcoR I
C2			tctccatagaatggaggaaaaaagagatat 4876 4904	EcoR I
C2.sl4m	C2	1897/1898	tctccatagaatacccacaaaaaagagatat 4876 4904	EcoR I
C2			ttatttaggacatatagtttagccctagggtg 4992 5021	EcoR I
C2.mESSv	C2	2272/2273	ttattctgacatatagttctccctagggtg 4992 5021	EcoR I
SIV-U1			cccgtagacgctgacacacctccatgggta 5269 5298	Hind III
SIV.ESS2-U1	SIV-U1	2030/2031	cccgtagacgctagactagatccatgggta 5269 5298	Hind III
L3-U1			gcggttactcaacagaggagagcaagaaatg 5385 5414	Pst I
L3.ΔESE-U1	L3-U1	1901/1902	gcggttactcaaca-----agcaagaaatg 5385 5414	Pst I

FIG. VIII.4 – Variants réalisés par mutagenèse dirigée.

Les plasmides utilisés comme matrice et les oligonucléotides utilisés pour réaliser les mutagenèses sont indiqués. Les séquences des constructions sauvages et mutées sont données. Les nucléotides mutés sont représentés en rouge et les nucléotides délétés sont remplacés par des tirets rouges. Les éléments régulateurs sont encadrés sur fond rose.

Chapitre IX

Méthodes spécifiques à l'étude de l'épissage *in vitro*

IX.1 Synthèse de transcrits radioactifs *in vitro*

IX.1.1 Préparation du fragment d'ADN à transcrire

Afin de limiter la transcription à la région nous intéressant, l'ADN servant de matrice est préalablement hydrolysé par une enzyme de restriction, au niveau d'un site localisé juste en aval de la séquence clonée. Le site de linéarisation de chacune des constructions utilisées est indiqué dans les figures VIII.1 à VIII.4.

Après hydrolyse par une enzyme de restriction, l'ADN matrice est soumis à une extraction au phénol-chloroforme (1/1) et à une extraction par un mélange chloroforme-alcool isoamylique (24/1). L'ADN est ensuite précipité à l'éthanol en présence d'acétate de sodium 0,3 M (concentration finale). L'ADN linéarisé est repris dans de l'eau et ajusté à la concentration de 500 ng/ μ l.

IX.1.2 Transcription *in vitro*

La synthèse de transcrits uniformément marqués au ^{32}P est réalisée par réaction de polymérisation par l'ARN polymérase du phage T7 ou SP6 en présence d'un ribonucléotide marqué au ^{32}P . L'ADN linéarisé (250 ng) est transcrit dans le milieu réactionnel suivant :

Tampon de transcription US 5X (MgCl ₂ 30 mM ; NaCl 50 mM ; Tris-HCl 0,2 M, DTT 200 mM)	0,5 μ l
Inhibiteur de RNases (RNasin, 31 U/ μ l) (Amersham Biosciences)	0,5 μ l
Mélange de rNTP : ATP (5 mM), CTP (5 mM), GTP (2 mM)	1 μ l
UTP 0,5 mM	1 μ l
[α - ^{32}P] UTP (800 Ci/mmol ; 10 mCi/ml) (ICN)	0,9 μ l
m ⁷ GpppG 10 mM (NEB)	1,25 μ l
ARN polymérase T7 (61 U/ μ l) (USB) ou SP6 (20 U/ μ l) (MBI Fermentas)	0,5 μ l ou 1 μ l
BSA (2 mg/ml)	0,5 μ l
H ₂ O	qsp 10 μ l

On diminue la concentration en GTP par rapport aux autres rNTPs, afin de favoriser l'incorporation de la coiffe m⁷GpppG à l'extrémité 5' du transcrit, ce qui protège l'ARN de la dégradation. Le mélange

est incubé 1 heure à 37°C. La matrice ADN est ensuite dégradée par incubation en présence de 10 U de DNase I, exempte de RNases (Amersham), pendant 30 minutes à 37°C. Après ajout de 20 µg d'ARNt de levure servant d'entraîneur de précipitation et de 190 µl d'eau stérile, les transcrits sont extraits avec un volume de phénol-chloroforme (1/1), puis avec un volume de chloroforme-alcool isoamylique (24/1). Les ARN sont ensuite soumis à deux précipitations successives à 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 à l'éthanol à 80%, puis séché et repris dans 10 µl d'eau stérile. La quantité de transcrits radioactifs obtenue est déterminée par mesure de la radioactivité contenue dans 1 µl d'ARN au compteur à scintillation (Beckman). L'ARN est ensuite dilué dans de l'eau stérile à la concentration de 100 000 cpm (mesurés en Cerenkov)/µl. 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).

IX.2 Epissage *in vitro*

Pour réaliser les expériences d'épissage *in vitro*, nous avons utilisé des extraits nucléaires de cellules HeLa provenant de la société Computer Cell Culture Center (Belgique). L'extrait nucléaire peut être décongelé et recongelé 1 à 2 fois sans perte notable d'activité. Les extraits cytoplasmiques S100 nous ont été fournis par J. Stévenin (IGBMC, Strasbourg) (Himmelspach *et al.* 1995). Les protéines SR, également fournies par J. Stévenin, sont des protéines recombinantes exprimées dans le système baculovirus (Cavaloc *et al.* 1999).

IX.2.1 Réaction d'épissage

Chaque réaction d'épissage *in vitro* est réalisée avec 1 µl de transcrit radioactif, soit environ 100 000 cpm. Le transcrit est ajouté au mélange réactionnel suivant :

MgCl ₂ 62,5 mM	1 µl
ATP 25 mM	1 µl
Créatine phosphate 625 mM	1 µl
RNasin (31 U/µl)	0,2 µl
Tampon RNasin (KCl 50 mM, DTT 5 mM, HEPES-KOH pH 7,6 20 mM, glycérol 50%)	0,2 µl
Alcool polyvinylique 13% (p/v)	6 µl
Tampon D (EDTA 0,2 mM; KCl 0,1 M; DTT 0,5 mM; PMSF 0,25 mM; HEPES 20 mM pH 7,9; glycérol 20%)	3,6 µl

La réaction est initiée par ajout à ce mélange de 8 µl d'extrait nucléaire, préalablement traité 10 minutes à 30°C afin d'éliminer l'ATP endogène. Le mélange est placé 5 minutes sur glace, puis incubé 2h30 à 30°C. Dans le cas des ARN dérivés de l'adénovirus (Fig. VIII.3), l'incubation est réalisée pendant 1h30.

Lors des tests de l'effet des protéines SR sur l'efficacité d'épissage *in vitro*, 150 à 300 ng de protéines

SR recombinantes purifiées sont ajoutés au milieu réactionnel, en même temps que l'extrait nucléaire. Les échantillons sont traités de la même façon qu'au cours des expériences réalisées en absence de protéines SR.

La réaction d'épissage est arrêtée par ajout de 20 μg de protéinase K (Roche) dans un tampon contenant : EDTA 12,5 mM ; NaCl 150 mM ; SDS 1% ; Tris-HCl 100 mM, pH 7,4. Après une incubation de 30 minutes à 30°C, les protéines sont extraites au phénol-chloroforme, puis au chloroforme-alcool isoamylique (24/1) et les ARN sont précipités en présence d'éthanol absolu sans ajout de sels. Le précipité est lavé puis séché et repris dans 5 μl d'une solution de bleu de formaldéhyde (EDTA 20 mM, bleu de bromophénol 0,05%, bleu de xylène cyanol 0,05%, dans de la formaldéhyde désionisée).

IX.2.2 Fractionnement sur gel et fluorographie

Les produits d'épissage sont dénaturés pendant 10 minutes à 65°C puis plongés dans la glace, avant d'être analysés par électrophorèse sur gel de polyacrylamide dénaturant dont la composition est la suivante : acrylamide 5% (p/v) et N, N'-méthylène bisacrylamide dans un rapport 20/1 (p/p), tampon TBE 1x (acide borique 89 mM, EDTA 2 mM, Tris borate 89 mM, pH 8,3), urée 8 M, TEMED 0,08% (v/v) (catalyseur de la polymérisation) et persulfate d'ammonium 0,1% (p/v) (initiateur de la polymérisation). Après une pré-électrophorèse à 45 V/cm, destinée à préchauffer le gel à 48°C, 1500 cpm Cerenkov sont déposés sur le gel. L'électrophorèse est poursuivie à 45 V/cm pendant 4 heures à une température d'environ 48°C, le bleu de xylène cyanol étant utilisé comme référence de migration. Dans le cas des ARN dérivés de l'adénovirus (Fig. VIII.3), l'analyse des produits d'épissage est réalisée sur un gel d'acrylamide 6%, pendant 1h15.

Les ARN sont ensuite fixés par immersion du gel pendant 5 min dans une solution d'acide acétique 10 % (v/v). Le gel est transféré sur du papier Whatmann 3MM, recouvert d'un film plastique, puis séché par chauffage sous vide à 80°C. Le gel est alors placé en fluorographie à -80°C au contact d'un film Fuji X-Ray, en présence d'écrans amplificateurs.

IX.2.3 Détermination de l'efficacité d'épissage

Le gel est scanné par un appareil PhosphorImager (Molecular Dynamics) et la radioactivité contenue dans les bandes d'intérêt est quantifiée à l'aide du logiciel Image Quant, version 5.2. Les quantités de molécules de précurseurs et de produits d'épissage sont calculées sur la base du contenu en résidus uridine de chaque type d'ARN. L'efficacité d'épissage est exprimée par le rapport de la radioactivité contenue dans la bande d'ARN messenger (M) et celle contenue dans la bande de précurseur restant à la fin de la réaction (P). Les expériences ont été répétées au moins deux fois afin de vérifier leur reproductibilité.

Chapitre X

Techniques d'analyse de la structure secondaire des ARN

X.1 Réaction de transcription *in vitro*

Les conditions de réaction sont les suivantes :

ADN linéarisé	2 μg
Tampon de transcription TSB 5x (HEPES-KOH pH 7,6 400 mM ; DTT 150 mM ; spermidine 10 mM ; MgCl ₂ 60 mM)	10 μl
RNasin (31 U/ μl)	0,5 μl
Mélange de rNTP (10 mM chacun)	10 μl
ARN polymérase T7 (61 U/ μl)	1 μl
H ₂ O	qsp 50 μl

Le mélange est incubé 2h à 37°C, puis la matrice d'ADN est dégradée par incubation pendant 30 minutes à 37°C, en présence de 10 U de DNase I exempte de RNases. Les ARN sont purifiés par extraction avec un volume de phénol/chloroforme (1/1), puis avec un volume de chloroforme-alcool isoamylique (24/1), et ils sont précipités à l'éthanol en présence d'acétate d'ammonium 2,5 M (concentration finale). Le précipité est repris dans de l'eau stérile à raison de 50 ng/ μl et les ARN dissous sont conservés à -80°C.

X.2 Séquençage de l'ARN

X.2.1 Principe

Un oligonucléotide, marqué en 5' au ³²P, est hybridé à une région donnée de l'ARN à séquencer. Il sert d'amorce à la transcriptase inverse qui synthétise, en présence des 4 désoxynucléotides dNTP, un brin d'ADN complémentaire à l'ARN. Des didésoxy-nucléotides ddNTP sont ajoutés au milieu réactionnel en concentration telle que la synthèse du brin d'ADN est arrêtée statistiquement. Un témoin permettant d'identifier les pauses naturelles de la transcriptase inverse est effectué en absence de ddNTP.

X.2.2 Marquage des oligodésoxyribonucléotides en 5'

Les oligonucléotides sont marqués en 5' en présence d'ATP [γ - 32]P par action de la polynucléotide kinase du bactériophage T4. Le mélange réactionnel utilisé pour la réaction de phosphorylation est le suivant :

[γ - 32]P ATP (10 mCi/ml, 3000 Ci/mmol, Amersham)	1 μ l
Oligodésoxyribonucléotide	100 ng
Tampon kinase 10x (MgCl ₂ 0,1 M ; DTT 0,05 M ; Tris-HCl 0,5 M pH 7,6)	1 μ l
Polynucléotide kinase (10 U/ μ l) (MBI Fermentas)	1 μ l
H ₂ O	qsp 10 μ l

Après incubation 45 minutes à 37°C, le mélange est chauffé pendant 10 minutes à 65°C pour inactiver la kinase, puis les oligodésoxyribonucléotides sont dilués à 1 ng/ μ l dans de l'eau stérile.

Certains oligonucléotides doivent être purifiés sur gel de polyacrylamide 10% dénaturant. Dans ce cas, à l'issue de la réaction de marquage, 5 μ l de bleu de formaldéhyde sont ajoutés et le mélange est déposé sur le gel. L'électrophorèse est réalisée à 17 V/cm dans du tampon TBE 1x. Le gel est ensuite autoradiographié et la bande contenant l'oligonucléotide est excisée et incubée dans 400 μ l de tampon d'élution (NaCl 0,25 M ; EDTA 0,01 M ; Tris-HCl pH 7,5 0,02 M) à 37°C sur la nuit. La phase aqueuse est ensuite prélevée et centrifugée 1 min à 12000 rpm pour éliminer la présence éventuelle de fins débris de gel. L'oligodésoxynucléotide contenu dans le surnageant est alors précipité en présence d'éthanol absolu et de 1 μ l de glycogène (10 μ g/ μ l), sans ajout de sels. Le culot est ensuite lavé, séché puis repris dans de l'eau à raison de 100000 cps/ μ l.

X.2.3 Réaction de séquençage (Qu *et al.* 1983)

X.2.3.1 Hybridation de l'amorce oligonucléotidique

Les conditions d'hybridation sont les suivantes :

Oligodésoxyribonucléotide marqué en 5' au 32 P	1 μ l
ARN transcrit <i>in vitro</i>	1 μ l
Tampon AMV RT 10X (MgCl ₂ 60 mM ; KCl 0,4 M ; Tris-HCl 0,5 M, pH 8,3)	0,25 μ l
H ₂ O	qsp 2,5 μ l

Le mélange est placé durant 40 secondes à 96°C, puis est immédiatement plongé dans la glace.

X.2.3.2 Réaction d'élongation

Aux 2,5 μ l du mélange d'hybridation sont ajoutés :

Mélange de dNTP (5 mM chacun)	0,1 μ l
Tampon AMV RT 10X	0,25 μ l
Transcriptase inverse AMV (2 U/ μ l) (Q-Biogen)	0,25 μ l
H ₂ O	qsp 5 μ l

A cette étape, 0,5 μl de l'un des ddNTP 500 μM est ajouté dans chacun des 4 tubes A, C, G et T du séquençage. Le mélange réactionnel est incubé pendant 30 minutes à 45°C. Les réactions sont stoppées par ajout de 4 μl de bleu de dépôt de formaldéhyde.

X.2.3.3 Fractionnement sur gel

Les ADNc sont dénaturés 2 minutes à 96°C, puis placés dans la glace. L'analyse des produits de séquençage et d'élongation se fait par séparation sur gel de polyacrylamide dénaturant à 7%. Les échantillons sont déposés sur le gel (2 μl), l'électrophorèse est réalisée à 45 V/cm en tampon TBE 1x, les bleus de bromophénol et de xylène cyanol étant utilisés comme référence de migration. Le gel est ensuite traité comme au paragraphe IX.2.2.

Les produits de séquençage et d'élongation ont parfois été séparés sur gel de polyacrylamide dénaturant à 10%. Dans ce cas, suite à l'électrophorèse, les ARN sont fixés par immersion du gel pendant 5 minutes dans une solution d'éthanol 30% et acide acétique 10% (v/v). Le gel est transféré sur un film plastique puis recouvert de papier Whatmann 3MM avant d'être séché sous vide et mis en fluorographie.

X.3 Analyse de la structure secondaire des ARN par des sondes chimiques et enzymatiques

X.3.1 Principe

Pour identifier les régions simple-brin ou double-brin de l'ARN, nous avons réalisé des modifications chimiques (par le diméthylsulfate ou DMS et le kéthoxal) et des digestions enzymatiques (par les RNases V1, T1 ou T2) en conditions statistiques (Qu *et al.* 1983 ; Ehresmann *et al.* 1987). Les cibles de ces agents sont présentées dans la partie résultats de ce manuscrit.

Les hydrolyses enzymatiques et les modifications chimiques sont réalisées dans des conditions ménagées en présence d'ARNt, de manière à avoir en moyenne une coupure ou une modification par molécule. Les liaisons phosphodiester coupées par les enzymes et les bases modifiées par les sondes chimiques sont mises en évidence par extension d'une amorce oligodésoxyribonucléotide par la transcriptase inverse. L'oligodésoxyribonucléotide amorce est marqué en 5' au ^{32}P . Une réaction d'élongation est effectuée sur l'ARN intact, ce qui permet de différencier les arrêts d'élongation dus aux modifications chimiques et coupures enzymatiques, des pauses de la transcriptase inverse en cours d'élongation. Les produits d'élongation sont séparés par électrophorèse en gel de polyacrylamide dénaturant, en parallèle avec les produits de la réaction de séquençage réalisée avec la transcriptase inverse et le même oligodésoxyribonucléotide amorce. La position des bases modifiées et des coupures enzymatiques est déduite par comparaison avec le séquençage.

X.3.2 Conditions expérimentales

Les modifications chimiques et enzymatiques sont effectuées sur 50 ng de transcrit synthétisé *in vitro*. La qualité du transcrit est préalablement vérifiée par un séquençage à la transcriptase inverse. Les expériences ont été répétées au moins deux fois afin de vérifier l'obtention de résultats reproductibles.

X.3.2.1 Modifications chimiques

Avant de réaliser les modifications chimiques au DMS, 50 ng d'ARN sont dilués dans 50 μ l du tampon (KCl 300 mM ; MgCl₂ 10 mM ; cacodylate de sodium 50 mM, pH 7,5) dans lequel la réaction sera réalisée, et sont préincubés 5 min à 65°C puis 10 min à 20°C. Les modifications au DMS (Aldrich) sont réalisées de façon ménagée par ajout de 0,5 ou 1 μ l d'une solution de DMS 100% diluée au 1/4 dans de l'éthanol absolu (concentration finale 0,03 M) par essai. Après une incubation de 10 minutes à 20°C, les réactions sont stoppées par l'ajout de 10 μ l de solution G-Stop (AcONa 1,5 M pH 5,3 ; β -mercaptoéthanol 1 M ; Tris-acétate 1 M pH 7,5 ; EDTA 0,1 mM) suivi d'une précipitation à l'éthanol absolu. Le précipité d'ARN est ensuite lavé à l'éthanol à 70%, séché puis dissous dans 4 μ l d'eau stérile. Un μ l de cette solution d'ARN est utilisé pour une réaction d'élongation à la transcriptase inverse, dans les conditions décrites au paragraphe X.2.3.

Les ARN traités par le kéthoxal sont incubés dans le même mélange que les ARN traités par le DMS. On les préincube également 5 min à 65°C puis 10 min à 20°C. Les modifications au kéthoxal sont réalisées par ajout de 1 ou 2 μ l de kéthoxal à 37 mg/ml, dilué au 1/4 dans de l'eau. Après 10 minutes d'incubation à 20°C, les réactions sont stoppées par ajout de 10 μ l de borate de potassium 0,5 M, pH 7,0. Les précipités d'ARN sont lavés et repris dans 4 μ l de borate de potassium 25 mM, pH 7,0. Un μ l est utilisé par réaction d'élongation à la transcriptase inverse.

X.3.2.2 Hydrolyses enzymatiques

L'ARN (50 ng de transcrit et 1,25 μ g d'ARNt pour la RNase V1 ou 2 μ g d'ARNt pour les RNases T1 et T2) est préincubé 20 minutes à 20°C dans 50 μ l de tampon (KCl 350 mM ; MgCl₂ 10 mM ; Tris-HCl 10 mM, pH 7,5) avant ajout d'une RNase. Les hydrolyses enzymatiques sont réalisées de façon ménagée par incubation, 10 minutes à 20°C en présence de 0,025, 0,05 ou 0,1 U de RNase V1 (Vassilenko *et al.* 1981), de 0,05 ou 0,1 U de RNase T1 (Roche) ou 0,2 ou 0,5 U de RNase T2 (Gibco GRL). Les réactions sont stoppées par ajout de 1 μ l d'EDTA 100 mM pour les digestions à la RNase V1, et par ajout de 20 μ g d'ARNt de levure pour les digestions à la RNase T1 ou T2, immédiatement suivi dans tous les cas, d'une extraction phénolique. Après précipitation à l'éthanol en présence d'acétate de sodium 0,3 M final et lavage à l'éthanol à 70%, les précipités d'ARN sont repris dans 4 μ l d'eau stérile. Des analyses par élongation d'amorce par la transcriptase inverse sont réalisées avec 1 μ l de chacune des préparations d'ARN obtenues après hydrolyse enzymatique.

X.3.3 Prédiction de structure secondaire par ordinateur

Les prédictions de structure secondaire ont été réalisées par utilisation du programme MFold (version 3.1) (<http://www.bioinfo.rpi.edu/~zukerm/>; Zuker 2003).

Chapitre XI

Techniques d'analyse des interactions ARN/protéines

XI.1 Technique d'empreinte des ARN en extrait nucléaire ou en présence de protéine purifiée

XI.1.1 Amplification par PCR des ADN matrices

Un produit PCR, servant de matrice pour la réaction de transcription *in vitro*, est synthétisé par utilisation d'une amorce introduisant un promoteur pour l'ARN polymérase du phage T7 ou SP6, dans les conditions suivantes :

Plasmide	20 ng
Tampon de PCR 10X (Tris-HCl 200 mM pH=8,0; 20 mM MgCl ₂ ; 100 mM KCl; 60 mM (NH ₄) ₂ SO ₄ ; 1% Triton X100; 100 ml/mg BSA)	5 μ l
Amorce 5' portant le promoteur transcriptionnel	125 ng
Amorce 3'	125 ng
Mélange des 4 dNTP (1,25 mM chacun)	8 μ l
Dynazyme TM II ADN polymérase (2 U/ μ l) (Finnzyme)	0,5 μ l
H ₂ O	qsp 50 μ l

Après une dénaturation de 2 minutes à 94°C, l'ADN est amplifié pendant 30 cycles de 30 sec à 94°C, 30 sec à 45°C, 1 min 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.

XI.1.2 Synthèse de transcrits froids courts

Les conditions de réaction sont les suivantes :

Matrice PCR	500 ng à 1 μ g
Tampon de transcription TSB 5x	6 μ l
MgCl ₂ 0,25 M	1,95 μ l
RNasin (31 U/ μ l)	0,5 μ l
Mélange de rNTP (12,5 mM chacun)	9,6 μ l
ARN polymérase T7 (69 U/ μ l) (ou SP6 (20 U/ μ l))	1 (ou 2) μ l
H ₂ O	qsp 30 μ l

Le mélange est incubé 2h à 37°C, puis la matrice d'ADN est dégradée par incubation pendant 30 minutes à 37°C, en présence de 10 U de DNase I, exempte de RNases. Après ajout de 170 μ l d'eau, les ARN sont purifiés par extraction avec un volume de phénol/chloroforme, puis avec un volume de chloroforme-alcool isoamylique, et ils sont précipités à l'éthanol absolu une première fois en présence d'acétate d'ammonium 0,25 M et une seconde fois en présence de NaCl 0,2 M. Le précipité est repris dans de l'eau stérile à raison de 200 ng/ μ l et les ARN dissous sont conservés à -80°C.

Ces petits transcrits sont séquencés ou élongués par la transcriptase inverse dans les conditions décrites au paragraphe X.2.3.

XI.1.3 Principe de la technique d'empreinte des ARN

La fixation d'une protéine sur l'ARN protège cet ARN de l'action des sondes enzymatiques ou chimiques habituellement utilisées dans les analyses de structure secondaire des ARN.

L'ARN est incubé en tampon D dans des conditions permettant l'épissage, seul, ou en présence d'extrait nucléaire ou de protéine purifiée. Les hydrolyses enzymatiques et les modifications chimiques sont ensuite réalisées. L'analyse par transcription inverse des ARN met en évidence, au niveau de la zone d'interaction ARN/protéine, un certain nombre de coupures ou modifications présentes au niveau de l'ARN seul, mais absentes lorsque l'ARN est incubé en présence de protéine.

XI.1.4 Conditions expérimentales de l'empreinte d'un ARN en extrait nucléaire ou en présence de protéine purifiée

XI.1.4.1 Assemblage des complexes ARN/protéines

L'ARN dont on veut identifier les sites de fixation de facteurs protéiques est incubé dans le mélange suivant :

Transcrit froid	200 ng
ARNt de levure	qsp 5 μ g d'ARN total
Tampon D	3,6 μ l
MgCl ₂ 62,5 mM	1 μ l
RNasin (31 U/ μ l)	0,5 μ l
(BSA)	(qsp 2 μ g de protéine)
(Cacodylate de sodium 1 M, pH 7,5)	(1 μ l)
H ₂ O	qsp 18 μ l

L'ajout de RNasin peut être omis si la préparation d'extrait nucléaire ou de protéine n'induit pas de coupure de l'ARN. Cet ajout ne gêne pas l'action des nucléases. Le cacodylate de sodium est ajouté au mélange réactionnel lorsque les sondes chimiques DMS et kéthoxal sont utilisées. Le tampon D n'a en effet pas de pouvoir tampon suffisant pour compenser la baisse de pH due à l'ajout des sondes chimiques. Lorsque les empreintes sont réalisées en présence de protéine purifiée, de la BSA est ajoutée afin de porter à 2 μg la quantité totale de protéines dans chaque échantillon.

Le mélange est ensuite placé 5 min à 65°C, puis 5 min à température ambiante et enfin, 5 min dans la glace afin d'obtenir une population d'ARN avec une structuration secondaire homogène. Quatre μl d'extrait nucléaire ou de protéine diluée dans le tampon D¹ sont ensuite ajoutés et le mélange est incubé 5 min dans la glace, puis 5 min à 30°C.

XI.1.4.2 Hydrolyses enzymatiques et modifications chimiques

On fait ensuite agir les sondes enzymatiques et chimiques pendant 10 min à 30°C. Les quantités d'agent chimique ou enzymatique généralement ajoutées et les conditions d'arrêt des réactions sont les suivantes :

Agent	Quantité ajoutée	Arrêt de la réaction
RNase V1	0,1 à 0,25 U	2 μl EDTA 100 mM, 20 μg ARNt de levure, incubation dans la glace
RNase T1	0,2 U à 0,3 U	20 μg ARNt, glace
RNase T2	0,3 U ou 0,6 U	20 μg ARNt, glace
DMS	dilué à 12,5% dans de l'éthanol absolu, 0,8 à 1,2 μl	3 μl de β -mercaptoéthanol 6 M, 20 μg ARNt, glace
Kéthoxal	dilué à 12,5% dans l'eau, 1 à 2 μl	20 μl de borate de potassium 0,5 M, 20 μg ARNt, glace

Les échantillons sont aussitôt extraits au phénol/chloroforme (1/1) puis au chloroforme-alcool isoamylique (24/1) et précipités à l'éthanol en présence d'acétate de sodium 0,3 M. Les culots sont ensuite lavés, séchés et repris dans 4 μl d'eau pour les échantillons traités par les sondes enzymatiques et le DMS, et dans 4 μl de borate de potassium 25 mM pH 7,0 pour les ARN traités par le kéthoxal.

Les ARN sont ensuite analysés par transcription inverse dans les conditions décrites au § X.2.3.

XI.1.5 Technique de retard sur gel

XI.1.5.1 Synthèse de transcrits radioactifs courts

Des transcrits uniformément marqués ont été produits à partir de matrices PCR produites dans les conditions décrites dans le § XI.1.1. Le mélange réactionnel de transcription est le suivant :

¹les quantités de protéine utilisées sont précisées dans la partie résultats

Tampon de transcription US 5X	6 μ l
DTT 200 mM	1,5 μ l
MgCl ₂ 0,25 M	0,95 μ l
RNasin (31 U/ μ l)	0,5 μ l
Mélange de rNTP : ATP (5 mM) CTP (5 mM) GTP (5 mM)	1 μ l
UTP 0,1 mM	1 μ l
α - ³² P UTP (800 Ci/mmol ; 10 mCi/ml) (ICN)	2 μ l
ARN polymérase T7 (61 U/ μ l) ou SP6 (20 U/ μ l)	1 μ l
BSA (2 mg/ml)	1 μ l
H ₂ O	qsp 30 μ l

Le mélange est incubé 2 heures à 37°C. La matrice ADN est ensuite dégradée par incubation en présence de 10 U de DNase I exempte de RNases, pendant 30 minutes à 37°C.

A l'issue de la réaction de transcription, 15 μ l de bleu de formamide sont ajoutés et le mélange est déposé sur le gel de polyacrylamide dénaturant à 6%. L'électrophorèse est réalisée à 17 V/cm dans du tampon TBE 0,5x. Le gel est ensuite autoradiographié et la bande contenant l'ARN est excisée et incubée dans 180 μ l de tampon d'éluion (NaCl 0,25 M ; EDTA 0,01 M ; Tris-HCl 0,02 M pH 7,5) et 20 μ l de phénol à température ambiante sur la nuit. La phase aqueuse est ensuite prélevée et centrifugée 1 min à 12000 rpm pour éliminer la présence éventuelle de fins débris de gel. L'ARN est extrait au phénol-chloroforme puis chloroforme-alcool isoamylique et précipité en présence d'éthanol absolu et de 1 μ l de glycogène (10 μ g/ μ l), sans ajout de sels. Le précipité d'ARN est lavé deux fois à l'éthanol à 80%, puis séché et repris dans de l'eau stérile.

XI.1.5.2 Retard sur gel

Cette technique aussi appelée EMSA² 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.

10 000 cpm d'ARN marqué sont incubés dans le mélange réactionnel utilisé pour assembler les complexes ARN/protéines (décrit dans le § XI.1.4.1), en présence de 20 μ g d'ARNt de levure. Après 5 min d'incubation dans la glace suivies de 15 min d'incubation à 30°C, 6 μ l de bleu de dépôt non dénaturant (HEPES 200 mM pH 7,9 ; glycérol 40%, bleu de bromophénol 0,03% ; xylène cyanol 0,03%) 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 , glycérol 0,5%. Le gel est ensuite séché et exposé en fluorographie pendant une nuit.

² pour « Electrophoretic Mobility Shift Assay »

Bibliographie

- Abovich, N. & Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*, **89** (3), 403–12.
- Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S. & Lührmann, R. (1998) The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol Cell Biol*, **18** (11), 6756–66.
- Adams, M. D., Tarnag, R. S. & Rio, D. C. (1997) The alternative splicing factor PSI regulates P-element third intron splicing *in vivo*. *Genes Dev*, **11** (1), 129–38.
- Ainger, K., Avossa, D., Diana, A. S., Barry, C., Barbarese, E. & Carson, J. H. (1997) Transport and localization elements in myelin basic protein mRNA. *J Cell Biol*, **138** (5), 1077–87.
- Akker, S. A., Smith, P. J. & Chew, S. L. (2001) Nuclear post-transcriptional control of gene expression. *J Mol Endocrinol*, **27** (2), 123–31.
- Akusjarvi, G. & Stévenin, J. (2003) Remodelling of the host cell RNA splicing machinery during an adenovirus infection. *Curr Top Microbiol Immunol*, **272**, 253–86.
- Amendt, B. A., Hesslein, D., Chang, L. J. & Stoltzfus, C. M. (1994) Presence of negative and positive *cis*-acting RNA splicing elements within and flanking the first *tat* coding exon of human immunodeficiency virus type 1. *Mol Cell Biol*, **14** (6), 3960–70.
- Amendt, B. A., Si, Z. H. & 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** (8), 4606–15.
- Anderson, G. J., Bach, M., Lührmann, R. & Beggs, J. D. (1989) Conservation between yeast and man of a protein associated with U5 small nuclear ribonucleoprotein. *Nature*, **342** (6251), 819–21.
- Arora, V., Fredericksen, B. & Garcia, J. V. (2002) Nef: agent of cell subversion. *Microbes Infect*, **4** (2), 189–99.
- Ashiya, M. & Grabowski, P. J. (1997) A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA*, **3** (9), 996–1015.
- Bachler, M., Schroeder, R. & Ahsen, U. (1999) Streptotag: a novel method for the isolation of RNA-binding proteins. *RNA*, **5** (11), 1509–16.
- Bargou, R. C., Jurchott, K., Wagener, C., Bergmann, S., Metzner, S., Bommert, K., Mapara, M. Y., Winzer, K. J., Dietel, M., Dorken, B. & D., R. H. (1997) Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic MDR1 gene expression. *Nat Med*, **3** (4), 447–50.
- Barnard, D. C., Li, J., Peng, R. & Patton, J. G. (2002) Regulation of alternative splicing by SRp86 through coactivation and repression of specific SR proteins. *RNA*, **8** (4), 526–33.
- Barnard, D. C. & Patton, J. G. (2000) Identification and characterization of a novel serine-arginine-rich splicing regulatory protein. *Mol Cell Biol*, **20** (9), 3049–57.
- Bartz, S. R. & Emerman, M. (1999a) Human immunodeficiency virus type 1 *Tat* induces apoptosis and increases sensitivity to apoptotic signals by up-regulating FLICE/caspase-8. *J Virol*, **73** (3), 1956–63.

- Bartz, S. R. & Emerman, M. (1999b) Human immunodeficiency virus type 1 Tat induces apoptosis and increases sensitivity to apoptotic signals by up-regulating FLICE/caspase-8. *J Virol*, **73** (3), 1956–63.
- Battiste, J., Mao, H., Rao, N., Tan, R., Muhandiram, D., Kay, L., Frankel, A. & Williamson, J. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, **273** (5281), 1547–51.
- Beck, A. R., Medley, Q. G., O'Brien, S., Anderson, P. & Streuli, M. (1996) Structure, tissue distribution and genomic organization of the murine rrm-type RNA binding proteins TIA-1 and TIAR. *Nucleic Acids Res*, **24** (19), 3829–35.
- Bedford, M. T., Reed, R. & Leder, P. (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. *Proc Natl Acad Sci U S A*, **95** (18), 10602–7.
- Beggs, J. D., Teigelkamp, S. & Newman, A. J. (1995) The role of PRP8 protein in nuclear pre-mRNA splicing in yeast. *J Cell Sci Suppl*, **19**, 101–5.
- Behrens, S. E. & Lührmann, R. (1991) Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev*, **5** (8), 1439–52.
- Ben-David, Y., Letwin, K., Tannock, L., Bernstein, A. & Pawson, T. (1991) A mammalian protein kinase with potential for serine/threonine and tyrosine phosphorylation is related to cell cycle regulators. *EMBO J*, **10** (2), 317–25.
- Benkirane, M., Chun, R. F., Xiao, H., Ogryzko, V. V., Howard, B. H., Nakatani, Y. & Jeang, K. T. (1998) Activation of integrated provirus requires histone acetyltransferase. p300 and P/CAF are coactivators for HIV-1 Tat. *J Biol Chem*, **273** (38), 24898–905.
- Bennett, M., Michaud, S., Kingston, J. & Reed, R. (1992) Protein components specifically associated with pre-spliceosome and spliceosome complexes. *Genes Dev*, **6** (10), 1986–2000.
- Bennett, M. & Reed, R. (1993) Correspondence between a mammalian spliceosome component and an essential yeast splicing factor. *Science*, **262** (5130), 105–8.
- Bentley, D. (2002) The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol*, **14** (3), 336–42.
- Berget, S. M. (1995) Exon recognition in vertebrate splicing. *J Biol Chem*, **270** (6), 2411–4.
- Berget, S. M., Moore, C. & Sharp, P. A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A*, **74** (8), 3171–5.
- Berget, S. M. & Sharp, P. A. (1977) A spliced sequence at the 5'-terminus of adenovirus late mRNA. *Brookhaven Symp Biol*, **29**, 332–44.
- Berglund, J. A., Abovich, N. & Rosbash, M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev*, **12** (6), 858–67.
- Berglund, J. A., Chua, K., Abovich, N., Reed, R. & Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, **89** (5), 781–7.
- Berglund, J. A., Fleming, M. L. & Rosbash, M. (1998) The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA*, **4** (8), 998–1006.
- Berglund, J. A., Rosbash, M. & Schultz, S. C. (2001) Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA*, **7** (5), 682–91.
- Beyer, A. L., Christensen, M. E., Walker, B. W. & LeStourgeon, W. M. (1977) Identification and characterization of the packaging proteins of core 40s hnRNP particles. *Cell*, **11** (1), 127–38.
- Biamonti, G., Ruggiu, M., Saccone, S., Della Valle, G. & Riva, S. (1994) Two homologous genes, originated by duplication, encode the human hnRNP proteins A2 and A1. *Nucleic Acids Res*, **22** (11), 1996–2002.
- Bilodeau, P. S., Domsic, J. K., Mayeda, A., Krainer, A. R. & Stoltzfus, C. M. (2001) RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element. *J Virol*, **75** (18), 8487–97.

- Bilodeau, P. S., Domsic, J. K. & Stoltzfus, C. M. (1999) Splicing regulatory elements within tat exon 2 of human immunodeficiency virus type 1 (HIV-1) are characteristic of group M but not group O HIV-1 strains. *J Virol*, **73** (12), 9764–72.
- Bindereif, A. & Green, M. R. (1987) An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J*, **6** (8), 2415–24.
- Black, D. L. (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell*, **103** (3), 367–70.
- Blanchette, M. & Chabot, B. (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J*, **18** (7), 1939–52.
- Blencowe, B. J., Bauren, G., Eldridge, A. G., Issner, R., Nickerson, J. A., Rosonina, E. & Sharp, P. A. (2000) The SRm160/300 splicing coactivator subunits. *RNA*, **6** (1), 111–20.
- Blencowe, B. J., Issner, R., Nickerson, J. A. & Sharp, P. A. (1998) A coactivator of pre-mRNA splicing. *Genes Dev*, **12** (7), 996–1009.
- Bockmayr, A. & Courtois, A. (2002) Using hybrid concurrent constraint programming to model dynamic biological systems. In *18th International Conference on Logic Programming, ICLP'02, Copenhagen*, (Springer, ed.), vol. 2401, pp. 85–99 LNCS.
- Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuvéglise, C., Munsterkotter, M., Guldener, U., Mewes, H. W., Van Helden, J., Dujon, B. & Gaillardin, C. (2003) Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res*, **31** (4), 1121–35.
- Bonen, L. & Vogel, J. (2001) The ins and outs of group II introns. *Trends Genet*, **17** (6), 322–31.
- Bourara, K., Litvak, S. & Araya, A. (2000) Generation of G-to-A and C-to-U changes in HIV-1 transcripts by RNA editing. *Science*, **289** (5484), 1564–6.
- Bourgeois, C. F., Popielarz, M., Hildwein, G. & Stevenin, J. (1999) Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol Cell Biol*, **19** (11), 7347–56.
- Branlant, C., Krol, A., Ebel, J. P., Gallinaro, H., Lazar, E. & Jacob, M. (1981) The conformation of chicken, rat and human U1A RNAs in solution. *Nucleic Acids Res*, **9** (4), 841–58.
- Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J. & Jeanteur, P. (1980) Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. *Nucleic Acids Res*, **8** (18), 4143–54.
- Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Haendler, B. & Jacob, M. (1982) U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *EMBO J*, **1** (10), 1259–65.
- Branlant, C., Krol, A., Lazar, E., Haendler, B., Jacob, M., Galego-Dias, L. & Pousada, C. (1983) High evolutionary conservation of the secondary structure and of certain nucleotide sequences of U5 RNA. *Nucleic Acids Res*, **11** (23), 8359–67.
- Brosi, R., Groning, K., Behrens, S. E., Lührmann, R. & Kramer, A. (1993a) Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science*, **262** (5130), 102–5.
- Brosi, R., Hauri, H. P. & Kramer, A. (1993b) Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem*, **268** (23), 17640–6.
- Brow, D. A. & Guthrie, C. (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, **334** (6179), 213–8.
- Brown, J. A., Howcroft, T. K. & Singer, D. S. (1998) HIV Tat protein requirements for transactivation and repression of transcription are separable. *J Acquir Immune Defic Syndr Hum Retrovirol*, **17** (1), 9–16.
- Brown, M. T. & Cooper, J. A. (1996) Regulation, substrates and functions of src. *Biochim Biophys Acta*, **1287** (2-3), 121–49.

- Buée, L., Bussièrre, T., Buee-Scherrer, V., Delacourte, A. & Hof, P. R. (2000) Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res Brain Res Rev*, **33** (1), 95–130.
- Bukrinsky, M. & Adzhubei, A. (1999) Viral protein R of HIV-1. *Rev Med Virol*, **9** (1), 39–49.
- Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M. & Stevenson, M. (1993) A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature*, **365** (6447), 666–9.
- Bullock, W. O., Fernandez, J. M. & Short, J. M. (1987) XL1-blue: A high efficiency plasmid transforming *recA* *Escherichia coli* strain with beta-galactosidase selection. *BioTechniques*, **5**, 376–378.
- Buratti, E. & Baralle, F. E. (2001) Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J Biol Chem*, **276** (39), 36337–43.
- Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M. & Baralle, F. E. (2001) Nuclear factor TDP-43 and SR proteins promote *in vitro* and *in vivo* CFTR exon 9 skipping. *EMBO J*, **20** (7), 1774–84.
- Burd, C. G. & Dreyfuss, G. (1994a) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J*, **13** (5), 1197–204.
- Burd, C. G. & Dreyfuss, G. (1994b) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265** (5172), 615–21.
- Burd, C. G., Swanson, M. S., Grolach, M. & Dreyfuss, G. (1989) Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts. *Proc Natl Acad Sci U S A*, **86** (24), 9788–92.
- Burtis, K. C. & Baker, B. S. (1989) Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell*, **56** (6), 997–1010.
- Buvoli, M., Cobianchi, F., Bestagno, M. G., Mangiarotti, A., Bassi, M. T., Biamonti, G. & Riva, S. (1990) Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein. *EMBO J*, **9** (4), 1229–35.
- Cáceres, J. F. & Kornblihtt, A. R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet*, **18** (4), 186–93.
- Cáceres, J. F., Sreaton, G. R. & Krainer, A. R. (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev*, **12** (1), 55–66.
- Calvio, C., Neubauer, G., Mann, M. & Lamond, A. I. (1995) Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. *RNA*, **1** (7), 724–33.
- Cao, G., Jamison, S. F. & Garcia-Blanco, M. A. (1997) Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing *in vitro*. *RNA*, **3** (12), 1456–67.
- Caputi, M., Mayeda, A., Krainer, A. R. & Zahler, A. M. (1999) hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO J*, **18** (14), 4060–7.
- Caputi, M. & Zahler, A. M. (2001) Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/f/2H9 family. *J Biol Chem*, **276** (47), 43850–9.
- Caputi, M. & Zahler, A. M. (2002) SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *EMBO J*, **21** (4), 845–55.
- Carlo, T., Sierra, R. & Berget, S. M. (2000) A 5' splice site-proximal enhancer binds SF1 and activates exon bridging of a microexon. *Mol Cell Biol*, **20** (11), 3988–95.
- Carlo, T., Sterner, D. A. & Berget, S. M. (1996) An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. *RNA*, **2** (4), 342–53.
- Carson, J. H., Cui, H. & Barbarese, E. (2001) The balance of power in RNA trafficking. *Curr Opin Neurobiol*, **11** (5), 558–63.

- Cavaloc, Y., Bourgeois, C. F., Kister, L. & Stévenin, J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA*, **5** (3), 468–83.
- Cavaloc, Y., Popielarz, M., Fuchs, J. P., Gattoni, R. & Stévenin, J. (1994) Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. *EMBO J*, **13** (11), 2639–49.
- Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A. R. & Cáceres, J. F. (2002) Nuclear export and retention signals in the RS domain of SR proteins. *Mol Cell Biol*, **22** (19), 6871–82.
- Chabot, B., Blanchette, M., Lapierre, I. & La Branche, H. (1997) An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol Cell Biol*, **17** (4), 1776–86.
- Champion-Arnaud, P. & 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** (16), 1974–83.
- Chandler, S. D., Mayeda, A., Yeakley, J. M., Krainer, A. R. & Fu, X. D. (1997) RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc Natl Acad Sci U S A*, **94** (8), 3596–601.
- Chang, D. D. & Sharp, P. A. (1989) Regulation by HIV Rev depends upon recognition of splice sites. *Cell*, **59** (5), 789–95.
- Chansky, H. A., Hu, M., Hickstein, D. D. & Yang, L. (2001) Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer Res*, **61** (9), 3586–90.
- Charlet, B. N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A. & Cooper, T. A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol Cell*, **10** (1), 45–53.
- Charneau, P., Alizon, M. & Clavel, F. (1992) A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *J Virol*, **66** (5), 2814–20.
- Charpentier, B. & Rosbash, M. (1996) Intramolecular structure in yeast introns aids the early steps of *in vitro* spliceosome assembly. *RNA*, **2** (6), 509–22.
- Charpentier, B., Stutz, F. & Rosbash, M. (1997) A dynamic *in vivo* view of the HIV-I Rev-RRE interaction. *J Mol Biol*, **266** (5), 950–62.
- Chaudhary, N., McMahon, C. & Blobel, G. (1991) Primary structure of a human arginine-rich nuclear protein that colocalizes with spliceosome components. *Proc Natl Acad Sci U S A*, **88** (18), 8189–93.
- Chen, C. D., Kobayashi, R. & Helfman, D. M. (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev*, **13** (5), 593–606.
- Chew, S. L., Liu, H. X., Mayeda, A. & Krainer, A. R. (1999) Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing. *Proc Natl Acad Sci U S A*, **96** (19), 10655–60.
- Chiara, M. D., Gozani, O., Bennett, M., Champion-Arnaud, P., Palandjian, L. & Reed, R. (1996) Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Mol Cell Biol*, **16** (7), 3317–26.
- Chiara, M. D., Palandjian, L., Feld Kramer, R. & Reed, R. (1997) Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. *EMBO J*, **16** (15), 4746–59.
- Choi, Y. D. & Dreyfuss, G. (1984) Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): a unique supramolecular assembly. *Proc Natl Acad Sci U S A*, **81** (23), 7471–5.
- Choi, Y. D., Grabowski, P. J., Sharp, P. A. & Dreyfuss, G. (1986) Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science*, **231** (4745), 1534–9.
- Chou, M. Y., Rooke, N., Turck, C. W. & Black, D. L. (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol Cell Biol*, **19** (1), 69–77.
- Chou, M. Y., Underwood, J. G., Nikolic, J., Luu, M. H. & Black, D. L. (2000) Multisite RNA binding and release

- of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing.
Mol Cell, **5** (6), 949–57.
- Chua, K. & Reed, R. (1999) The RNA splicing factor hSlu7 is required for correct 3' splice-site choice.
Nature, **402** (6758), 207–10.
- Cimarelli, A. & Darlix, J. L. (2002) Assembling the human immunodeficiency virus type 1.
Cell Mol Life Sci, **59** (7), 1166–84.
- Cmarko, D., Boe, S. O., Scassellati, C., Szilvay, A. M., Davanger, S., Fu, X. D., Haukenes, G., Kalland, K. H. & Fakan, S. (2002) Rev inhibition strongly affects intracellular distribution of human immunodeficiency virus type 1 RNAs.
J Virol, **76** (20), 10473–84.
- Cochrane, A. W., Jones, K. S., Beidas, S., Dillon, P. J., Skalka, A. M. & Rosen, C. A. (1991) Identification and characterization of intragenic sequences which repress human immunodeficiency virus structural gene expression.
J Virol, **65** (10), 5305–13.
- Coffin, J. M., Hughes, S. H. & Varmus, H. E. (1997) *Retroviruses*.
C. S. H. L. Press, New-York.
- Collins, C. A. & Guthrie, C. (2000) The question remains: is the spliceosome a ribozyme?
Nat Struct Biol, **7** (10), 850–4.
- Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C. & Duncan, P. I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution.
EMBO J, **15** (2), 265–75.
- Cook, K. S., Fisk, G. J., Hauber, J., Usman, N., Daly, T. J. & Rusche, J. R. (1991) Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate.
Nucleic Acids Res, **19** (7), 1577–83.
- Corden, J. L. & Patturajan, M. (1997) A CTD function linking transcription to splicing.
Trends Biochem Sci, **22** (11), 413–6.
- Côté, C. A., Gautreau, D., Denegre, J. M., Kress, T. L., Terry, N. A. & Mowry, K. L. (1999) A xenopus protein related to hnRNP I has a role in cytoplasmic RNA localization.
Mol Cell, **4** (3), 431–7.
- Cowper, A. E., Cáceres, J. F., Mayeda, A. & Sreaton, G. R. (2001) Serine-arginine (sr) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing.
J Biol Chem, **276** (52), 48908–14.
- Craigie, R. (2001) Hiv integrase, a brief overview from chemistry to therapeutics.
J Biol Chem, **276** (26), 23213–6.
- Cramer, P., Cáceres, J. F., Cazalla, D., Kadener, S., Muro, A. F., Baralle, F. E. & Kornblihtt, A. R. (1999) Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer.
Mol Cell, **4** (2), 251–8.
- Crispino, J. D., Blencowe, B. J. & Sharp, P. A. (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP.
Science, **265** (5180), 1866–9.
- Crispino, J. D. & Sharp, P. A. (1995) A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing.
Genes Dev, **9** (18), 2314–23.
- Damgaard, C. K., Tange, T. O. & Kjems, J. (2002) hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure.
RNA, **8** (11), 1401–15.
- Damier, L.
Etude de la régulation de l'épissage du transcrit primaire du virus de l'immunodéficience humaine de type 1, HIV-1.
Thèse Université Henri Poincaré, Nancy I.
- Damier, L., Domenjoud, L. & Branlant, C. (1997) The D1-A2 and D2-A2 pairs of splice sites from human immunodeficiency virus type 1 are highly efficient in vitro, in spite of an unusual branch site.
Biochem Biophys Res Commun, **237** (1), 182–7.
- Darlix, J. L., Cristofari, G. & Roques, B. (2001) La protéine de nucléocapside du VIH1: un peptide à doigt au cœur du virus et de sa réplication.
Virologie, **5** (5), 329–38.
- Das, B. K., Xia, L., Palandjian, L., Gozani, O., Chyung, Y. & Reed, R. (1999) Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155.

- Mol Cell Biol*, **19** (10), 6796–802.
- Das, R., Zhou, Z. & Reed, R. (2000) Functional association of U2 snRNP with the ATP-independent spliceosomal complex E.
Mol Cell, **5** (5), 779–87.
- Dauksaite, V. & Akusjärvi, G. (2002) Human splicing factor ASF/SF2 encodes for a repressor domain required for its inhibitory activity on pre-mRNA splicing.
J Biol Chem, **277** (15), 12579–86.
- de la Fuente, C., Santiago, F., Deng, L., Eadie, C., Zilberman, I., Kehn, K., Maddukuri, A., Baylor, S., Wu, K., Lee, C., Pumfery, A. & Kashanchi, F. (2002) Gene expression profile of HIV-1 Tat expressing cells: a close interplay between proliferative and differentiation signals.
BMC Biochem, **3** (1), 14.
- De la Fuente, C., Santiago, F., Deng, L., Eadie, C., Zilberman, I., Kehn, K., Maddukuri, A., Baylor, S., Wu, K., Lee, C. G., Pumfery, A. & Kashanchi, F. (2002) Gene expression profile of HIV-1 Tat expressing cells: a close interplay between proliferative and differentiation signals.
BMC Biochem, **3** (1), 14.
- Deirdre, A., Scadden, J. & Smith, C. W. (1995) Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs.
EMBO J, **14** (13), 3236–46.
- Del Gatto, F., Plet, A., Gesnel, M. C., Fort, C. & Breathnach, R. (1997) Multiple interdependent sequence elements control splicing of a fibroblast growth factor receptor 2 alternative exon.
Mol Cell Biol, **17** (9), 5106–16.
- Del Gatto-Konczak, F., Bourgeois, C. F., Le Guiner, C., Kister, L., Gesnel, M. C., Stévenin, J. & Breathnach, R. (2000) The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site.
Mol Cell Biol, **20** (17), 6287–99.
- Del Gatto-Konczak, F., Olive, M., Gesnel, M. C. & Breathnach, R. (1999) hnRNP A1 recruited to an exon *in vivo* can function as an exon splicing silencer.
Mol Cell Biol, **19** (1), 251–60.
- Dember, L. M., Kim, N. D., Liu, K. Q. & Anderson, P. (1996) Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities.
J Biol Chem, **271** (5), 2783–8.
- Deutsch, M. & Long, M. (1999) Intron-exon structures of eukaryotic model organisms.
Nucleic Acids Res, **27** (15), 3219–28.
- Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R. & Xu, R. M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA.
Genes Dev, **13** (9), 1102–15.
- Dirksen, W. P., Li, X., Mayeda, A., Krainer, A. R. & Rottman, F. M. (2000) Mapping the SF2/ASF binding sites in the bovine growth hormone exonic splicing enhancer.
J Biol Chem, **275** (37), 29170–7.
- Dix, I., Russell, C. S., O'Keefe, R. T., Newman, A. J. & Beggs, J. D. (1998) Protein-RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*.
RNA, **4** (12), 1675–86.
- Doudna, J. A. & Cech, T. R. (2002) The chemical repertoire of natural ribozymes.
Nature, **418** (6894), 222–8.
- Dreyfuss, G., Kim, V. N. & Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry.
Nat Rev Mol Cell Biol, **3** (3), 195–205.
- Dreyfuss, G., Matunis, M. J., Piñol Roma, S. & Burd, C. G. (1993) hnRNP proteins and the biogenesis of mRNA.
Annu Rev Biochem, **62**, 289–321.
- Du, H. & Rosbash, M. (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing.
Nature, **419** (6902), 86–90.
- Du, K. & Taub, R. (1997) Alternative splicing and structure of the human and mouse SFRS5/HRS/SRp40 genes.
Gene, **204** (1-2), 243–9.
- Duncan, P. I., Stojdl, D. F., Marius, R. M., Scheit, K. H. & Bell, J. C. (1998) The Clk2 and Clk3 dual-specificity protein kinases regulate the intranuclear distribution of SR proteins and influence pre-mRNA splicing.
Exp Cell Res, **241** (2), 300–8.
- Dyhr-Mikkelsen, H. & Kjems, J. (1995) Inefficient spliceosome assembly and abnormal branch site selection in splicing of an HIV-1 transcript *in vitro*.
J Biol Chem, **270** (41), 24060–6.

- Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P. & Ehresmann, B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res*, **15** (22), 9109–28.
- El Kharroubi, A., Piras, G., Zensen, R. & Martin, M. A. (1998) Transcriptional activation of the integrated chromatin-associated human immunodeficiency virus type 1 promoter. *Mol Cell Biol*, **18** (5), 2535–44.
- Eldridge, A. G., Li, Y., Sharp, P. A. & Blencowe, B. J. (1999) The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc Natl Acad Sci U S A*, **96** (11), 6125–30.
- Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A. & Krainer, A. R. (1993) Pathways for selection of 5' splice sites by u1 snRNPs and SF2/ASF. *EMBO J*, **12** (9), 3607–17.
- Eperon, I. C., Makarova, O. V., Mayeda, A., Munroe, S. H., Cáceres, J. F., Hayward, D. G. & Krainer, A. R. (2000) Selection of alternative 5' splice sites: role of u1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol*, **20** (22), 8303–18.
- Estes, P. A., Cooke, N. E. & Liebhaber, S. A. (1992) A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms. *J Biol Chem*, **267** (21), 14902–8.
- Expert-Bezancon, A., Le Caer, J. P. & Marie, J. (2002) Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6a from chicken beta-tropomyosin pre-mRNA. *J Biol Chem*, **277** (19), 16614–23.
- Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W. S. & Lührmann, R. (1997) An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J*, **16** (13), 4092–106.
- Fairbrother, W. G. & Chasin, L. A. (2000) Human genomic sequences that inhibit splicing. *Mol Cell Biol*, **20** (18), 6816–25.
- Fetzer, S., Lauber, J., Will, C. L. & Lührmann, R. (1997) The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase. *RNA*, **3** (4), 344–55.
- Fleckner, J., Zhang, M., Valcárcel, J. & Green, M. R. (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev*, **11** (14), 1864–72.
- Förch, P., Martínez, C., L., M. & Valcárcel, J. (2001) Modulation of msl-2 5' splice site recognition by Sex-lethal. *RNA*, **7** (9), 1185–91.
- Förch, P., Merendino, L., Martínez, C. & Valcárcel, J. (2003) The U2 snRNP Auxiliary Factor u2AF 65 can promote U1 snRNP recruitment to 5' splice sites. *Biochem J*, **31**.
- Förch, P., Puig, O., Martínez, C., Séraphin, B. & Valcárcel, J. (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *EMBO J*, **21** (24), 6882–92.
- Förch, P. & Valcárcel, J. (2001) Molecular mechanisms of gene expression regulation by the apoptosis-promoting protein TIA-1. *Apoptosis*, **6** (6), 463–8.
- Forne, T., Rossi, F., Labourier, E., Antoine, E., Cathala, G., Brunel, C. & Tazi, J. (1995) Disruption of base-paired U4.U6 small nuclear RNAs induced by mammalian heterogeneous nuclear ribonucleoprotein C protein. *J Biol Chem*, **270** (27), 16476–81.
- Frank, D., Patterson, B. & Guthrie, C. (1992) Synthetic lethal mutations suggest interactions between U5 small nuclear RNA and four proteins required for the second step of splicing. *Mol Cell Biol*, **12** (11), 5197–205.
- Franke, E. K. & 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** (1), 279–82.
- Franke, E. K., Yuan, H. E. & Luban, J. (1994) Specific incorporation of cyclophilin A into HIV-1 virions. *Nature*, **372** (6504), 359–62.
- Fu, X. D. & Maniatis, T. (1992) Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science*, **256** (5056), 535–8.

- Fu, X. D., Mayeda, A., Maniatis, T. & Krainer, A. R. (1992) General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5' and 3' splice site selection. *Proc Natl Acad Sci U S A*, **89** (23), 11224–8.
- Gallay, P., Swingler, S., Song, J., Bushman, F. & Trono, D. (1995) HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell*, **83** (4), 569–76.
- Gallego, M. E., Gattoni, R., Stévenin, J., Marie, J. & Expert-Bezancon, A. (1997) The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6a. *EMBO J*, **16** (7), 1772–84.
- Gama-Carvalho, M., Carvalho, M. P., Kehlenbach, A., Valcárcel, J. & Carmo-Fonseca, M. (2001) Nucleocytoplasmic shuttling of heterodimeric splicing factor U2AF. *J Biol Chem*, **276** (16), 13104–12.
- 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. & Hahn, B. H. (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, **397** (6718), 436–41.
- Garcia-Blanco, M. A., Jamison, S. F. & Sharp, P. A. (1989) Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev*, **3** (12A), 1874–86.
- Gattoni, R., Mahe, D., Mahl, P., Fischer, N., Mattei, M. G., Stévenin, J. & 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** (13), 2535–42.
- Ge, H. & Manley, J. L. (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell*, **62** (1), 25–34.
- Ge, H., Zuo, P. & Manley, J. L. (1991) Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell*, **66** (2), 373–82.
- Gee, S., Krauss, S. W., Miller, E., Aoyagi, K., Arenas, J. & Conboy, J. G. (1997) Cloning of mDEAH9, a putative RNA helicase and mammalian homologue of *Saccharomyces cerevisiae* splicing factor prp43. *Proc Natl Acad Sci U S A*, **94** (22), 11803–7.
- Geleziunas, R., Xu, W., Takeda, K., Ichijo, H. & Greene, W. C. (2001) HIV-1 Nef inhibits ASK1-dependent death signalling providing a potential mechanism for protecting the infected host cell. *Nature*, **410** (6830), 834–8.
- Gibellini, D., Re, M. C., La Placa, M. & Zauli, G. (2002) Differentially expressed genes in HIV-1 tat-expressing CD4(+) T-cell line. *Virus Res*, **90** (1-2), 337–45.
- Ginisty, H., Sicard, H., Roger, B. & Bouvet, P. (1999) Structure and functions of nucleolin. *J Cell Sci*, **112** (Pt 6), 761–72.
- Goguel, V. & Rosbash, M. (1993) Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell*, **72** (6), 893–901.
- Gontarek, R. R., McNally, M. T. & Beemon, K. (1993) Mutation of an RSV intronic element abolishes both U11/U12 snRNP binding and negative regulation of splicing. *Genes Dev*, **7** (10), 1926–36.
- Goode, B. L., Chau, M., Denis, P. E. & Feinstein, S. C. (2000) Structural and functional differences between 3-repeat and 4-repeat tau isoforms. implications for normal tau function and the onset of neurodegenerative disease. *J Biol Chem*, **275** (49), 38182–9.
- Gooding, C., Roberts, G. C. & Smith, C. W. (1998) Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. *RNA*, **4** (1), 85–100.
- Gordon, P. M., Sontheimer, E. J. & Piccirilli, J. A. (2000) Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA*, **6** (2), 199–205.
- Gorlach, M., Burd, C. G. & Dreyfuss, G. (1994) The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein c proteins. *J Biol Chem*, **269** (37), 23074–8.

- Gosert, R., Chang, K. H., Rijnbrand, R., Yi, M., Sangar, D. V. & Lemon, S. M. (2000) Transient expression of cellular polypyrimidine-tract binding protein stimulates cap-independent translation directed by both picornaviral and flaviviral internal ribosome entry sites in vivo. *Mol Cell Biol*, **20** (5), 1583–95.
- Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Luhrmann, R. & Fabrizio, P. (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J*, **18** (16), 4535–48.
- Gozani, O., Feld, R. & Reed, R. (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev*, **10** (2), 233–43.
- Gozani, O., Potashkin, J. & Reed, R. (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol*, **18** (8), 4752–60.
- Grabowski, P. J. & Black, D. L. (2001) Alternative RNA splicing in the nervous system. *Prog Neurobiol*, **65** (3), 289–308.
- Granadino, B., Penalva, L. O., Green, M. R., Valcárcel, J. & Sanchez, L. (1997) Distinct mechanisms of splicing regulation *in vivo* by the *Drosophila* protein Sex-lethal. *Proc Natl Acad Sci U S A*, **94** (14), 7343–8.
- Graveley, B. R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6** (9), 1197–211.
- Graveley, B. R., Hertel, K. J. & Maniatis, T. (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J*, **17** (22), 6747–56.
- Gromak, N. & Smith, C. W. (2002) A splicing silencer that regulates smooth muscle specific alternative splicing is active in multiple cell types. *Nucleic Acids Res*, **30** (16), 3548–57.
- Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K. & Hutton, M. (1999) 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J Biol Chem*, **274** (21), 15134–43.
- Gui, J. F., Tronchère, H., Chandler, S. D. & Fu, X. D. (1994) Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc Natl Acad Sci U S A*, **91** (23), 10824–8.
- Gunderson, S. I., Polycarpou-Schwarz, M. & Mattaj, I. W. (1998) U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell*, **1** (2), 255–64.
- Guth, S., Martínez, C., Gaur, R. K. & Valcárcel, J. (1999) Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2af(65). *Mol Cell Biol*, **19** (12), 8263–71.
- Guth, S. & Valcárcel, J. (2000) Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2af. *J Biol Chem*, **275** (48), 38059–66.
- Hackl, W., Fischer, U. & Lührmann, R. (1994) A 69-kD protein that associates reversibly with the sm core domain of several spliceosomal snRNP species. *J Cell Biol*, **124** (3), 261–72.
- Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. (2000) AIDS as a zoonosis: scientific and public health implications. *Science*, **287** (5453), 607–14.
- Hall, T. M. (2002) Poly(A) tail synthesis and regulation: recent structural insights. *Curr Opin Struct Biol*, **12** (1), 82–8.
- Hallier, M., Lerga, A., Barnache, S., Tavitian, A. & Moreau-Gachelin, F. (1998) The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. *J Biol Chem*, **273** (9), 4838–42.
- Hamm, J., Kazmaier, M. & Mattaj, I. W. (1987) *In vitro* assembly of U1 snRNPs. *EMBO J*, **6** (11), 3479–85.
- Hanamura, A., Cáceres, J. F., Mayeda, A., Franza, B. R., J. & Krainer, A. R. (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA*, **4** (4), 430–44.
- Harrich, D. & Hooker, B. (2002) Mechanistic aspects of HIV-1 reverse transcription initiation. *Rev Med Virol*, **12** (1), 31–45.

- Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., P.-M. S. K., Watt, I. N., Neuberger, M. S. & Malim, M. H. (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell*, **113** (6), 803–9.
- Harrison, G. P. & Lever, A. M. (1992) The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J Virol*, **66** (7), 4144–53.
- Hartmuth, K., Raker, V. A., Huber, J., Branlant, C. & Lührmann, R. (1999) An unusual chemical reactivity of sm site adenosines strongly correlates with proper assembly of core U snRNP particles. *J Mol Biol*, **285** (1), 133–47.
- Hartmuth, K., Urlaub, H., Vornlocher, H. P., Will, C. L., Gentzel, M., Wilm, M. & Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A*, **99** (26), 16719–24.
- Hashimoto, C. & Steitz, J. A. (1984) U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res*, **12** (7), 3283–93.
- Hastings, M. L. & Krainer, A. R. (2001) Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol*, **13** (3), 302–9.
- Hawkins, J. D. (1988) A survey on intron and exon lengths. *Nucleic Acids Res*, **16** (21), 9893–908.
- He, W. & Parker, R. (2000) Functions of Lsm proteins in mRNA degradation and splicing. *Curr Opin Cell Biol*, **12** (3), 346–50.
- Hedjran, F., Yeakley, J. M., Huh, G. S., Hynes, R. O. & Rosenfeld, M. G. (1997) Control of alternative pre-mRNA splicing by distributed pentameric repeats. *Proc Natl Acad Sci U S A*, **94** (23), 12343–7.
- Heguy, A. (1997) Inhibition of the HIV Rev transactivator: a new target for therapeutic intervention. *Front Biosci*, **2**, d283–97.
- Heinrichs, V., Bach, M., Winkelmann, G. & Lührmann, R. (1990) U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science*, **247** (4938), 69–72.
- Heinrichs, V., Ryner, L. C. & Baker, B. S. (1998) Regulation of sex-specific selection of fruitless 5' splice sites by transformer and transformer-2. *Mol Cell Biol*, **18** (1), 450–8.
- Helfman, D. M., Ricci, W. M. & Finn, L. A. (1988) Alternative splicing of tropomyosin pre-mRNAs *in vitro* and *in vivo*. *Genes Dev*, **2** (12A), 1627–38.
- Hermankova, M., Siliciano, J., Zhou, Y., Monie, D., Chadwick, K., Margolick, J., Quinn, T. & Siliciano, R. (2003) Analysis of human immunodeficiency virus type 1 gene expression in latently infected resting CD4+ T lymphocytes *in vivo*. *J Virol*, **77** (13), 7383–92.
- Hertel, K. J. & Maniatis, T. (1998) The function of multisite splicing enhancers. *Mol Cell*, **1** (3), 449–55.
- Hibbert, C. S., Gontarek, R. R. & Beemon, K. L. (1999) The role of overlapping U1 and U11 5' splice site sequences in a negative regulator of splicing. *RNA*, **5** (3), 333–43.
- Himmelspach, M., Cavaloc, Y., Chebli, K., Stévenin, J. & Gattoni, R. (1995) Titration of serine/arginine (SR) splicing factors during adenoviral infection modulates E1A pre-mRNA alternative splicing. *RNA*, **1** (8), 794–806.
- Hofmann, Y. & Wirth, B. (2002) hnRNP-G promotes exon 7 inclusion of survival motor neuron (smn) via direct interaction with Htra2-beta1. *Hum Mol Genet*, **11** (17), 2037–49.
- Honig, A., Auboeuf, D., Parker, M. M., O'Malley, B. W. & Berget, S. M. (2002) Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. *Mol Cell Biol*, **22** (16), 5698–707.
- Honoré, B., Rasmussen, H. H., Vorum, H., Dejgaard, K., Liu, X., Gromov, P., Madsen, P., Gesser, B., Tommerup, N. & Celis, J. E. (1995) Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. *J Biol Chem*, **270** (48), 28780–9.
- Honoré, B., Vorum, H. & Baandrup, U. (1999) hnRNPs H, H' and F behave differently with respect to posttranslational cleavage and subcellular localization. *FEBS Lett*, **456** (2), 274–80.

- Hope, T. J. (1999) The ins and outs of HIV Rev.
Arch Biochem Biophys, **365** (2), 186–91.
- Hornig, H., Aebi, M. & Weissmann, C. (1986) Effect of mutations at the lariat branch acceptor site on beta-globin pre-mRNA splicing *in vitro*.
Nature, **324** (6097), 589–591.
- Horowitz, D. S., Kobayashi, R. & Krainer, A. R. (1997) A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs.
RNA, **3** (12), 1374–87.
- Howell, B. W., Afar, D. E., Lew, J., Douville, E. M., Icely, P. L., Gray, D. A. & Bell, J. C. (1991) STY, a tyrosine-phosphorylating enzyme with sequence homology to serine/threonine kinases.
Mol Cell Biol, **11** (1), 568–72.
- Huang, T., Nilsson, C., Punga, T. & Akusjärvi, G. (2002) Functional inactivation of the SR family of splicing factors during a vaccinia virus infection.
EMBO Rep, **3** (11), 1088–93.
- Huang, Y. & Steitz, J. A. (2001) Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA.
Mol Cell, **7** (4), 899–905.
- Huffman, K. M. & Arrigo, S. J. (1997) Identification of *cis*-acting repressor activity within human immunodeficiency virus type 1 protease sequences.
Virology, **234** (2), 253–60.
- Hui, J., Stangl, K., Lane, W. S. & Bindereif, A. (2003) HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats.
Nat Struct Biol, **10** (1), 33–7.
- Hutton, M., Lendon, C. L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., Hackett, J., Adamson, J., Lincoln, S., Dickson, D., Davies, P., Petersen, R. C., Stevens, M., de Graaff, E., Wauters, E., van Baren, J., Hillebrand, M., Joosse, M., Kwon, J. M., Nowotny, P., Heutink, P. & et al. (1998) Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.
Nature, **393** (6686), 702–5.
- Izaurrealde, E., Jarmolowski, A., Beisel, C., Mattaj, I. W., Dreyfuss, G. & Fischer, U. (1997) A role for the m9 transport signal of hnRNP A1 in mRNA nuclear export.
J Cell Biol, **137** (1), 27–35.
- Izaurrealde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. & Mattaj, I. W. (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing.
Cell, **78** (4), 657–68.
- Jacob, M., Lazar, E., Haendler, B., Gallinaro, H., Krol, A. & Branlant, C. (1984) A family of small nucleoplasmic RNAs with common structural features.
Biol Cell, **51** (1), 1–9.
- Jacquet, S.
Etude de la régulation de l'épissage du transcrit primaire du virus de l'immunodéficience humaine de type 1.
Thèse Université Henri Poincaré, Nancy I.
- Jacquet, S., Méreau, A., Bilodeau, P. S., Damier, L., Stoltzfus, C. M. & 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** (44), 40464–75.
- Jacquet, S., Ropers, D., Bilodeau, P. S., Damier, L., Mougin, A., Stoltzfus, C. M. & 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** (2), 464–78.
- Jamison, S. F., Pasmán, Z., Wang, J., Will, C., Lührmann, R., Manley, J. L. & Garcia-Blanco, M. A. (1995) U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements.
Nucleic Acids Res, **23** (16), 3260–7.
- Jeang, K. T., Xiao, H. & Rich, E. A. (1999) Multifaceted activities of the HIV-1 transactivator of transcription, Tat.
J Biol Chem, **274** (41), 28837–40.
- Jensen, K. B., Dredge, B. K., Stefani, G., Zhong, R., Buckanovich, R. J., Okano, H. J., Yang, Y. Y. & Darnell, R. B. (2000a) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability.
Neuron, **25** (2), 359–71.
- Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K. & Darnell, R. B. (2000b) The tetranucleotide UCAY directs the specific recognition of RNA by the Nova-K-homology 3 domain.
Proc Natl Acad Sci U S A, **97** (11), 5740–5.

- Jin, W., Bi, W., Huang, E. S. & Côté, G. J. (1999a) Glioblastoma cell-specific expression of fibroblast growth factor receptor-1beta requires an intronic repressor of RNA splicing. *Cancer Res*, **59** (2), 316–9.
- Jin, W., Huang, E. S., Bi, W. & Côté, G. J. (1999b) Redundant intronic repressors function to inhibit fibroblast growth factor receptor-1 alpha-exon recognition in glioblastoma cells. *J Biol Chem*, **274** (39), 28035–41.
- Jin, W., McCutcheon, I. E., Fuller, G. N., Huang, E. S. & Côté, G. J. (2000) Fibroblast growth factor receptor-1 alpha-exon exclusion and polypyrimidine tract-binding protein in glioblastoma multiforme tumors. *Cancer Res*, **60** (5), 1221–4.
- Johnson, T. L. & Abelson, J. (2001) Characterization of U4 and U6 interactions with the 5' splice site using a *S. cerevisiae* *in vitro* trans-splicing system. *Genes Dev*, **15** (15), 1957–70.
- Jones, K. A. & Peterlin, B. M. (1994) Control of RNA initiation and elongation at the HIV-1 promoter. *Annu Rev Biochem*, **63**, 717–43.
- Jumaa, H., Guenet, J. L. & Nielsen, P. J. (1997) Regulated expression and RNA processing of transcripts from the SRp20 splicing factor gene during the cell cycle. *Mol Cell Biol*, **17** (6), 3116–24.
- Jumaa, H. & Nielsen, P. J. (2000) Regulation of SRp20 exon 4 splicing. *Biochim Biophys Acta*, **1494** (1-2), 137–43.
- Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N. & Moore, M. J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, **8** (4), 426–39.
- Jurica, M. S. & Moore, M. J. (2002) Capturing splicing complexes to study structure and mechanism. *Methods*, **28** (3), 336–45.
- Kadener, S., Fededa, J. P., Rosbash, M. & Kornblihtt, A. R. (2002) Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc Natl Acad Sci U S A*, **99** (12), 8185–90.
- Kafasla, P., Patrino-Georgoula, M., Lewis, J. D. & Gualis, A. (2002) Association of the 72/74-kDa proteins, members of the heterogeneous nuclear ribonucleoprotein M group, with the pre-mRNA at early stages of spliceosome assembly. *Biochem J*, **363** (Pt 3), 793–9.
- Kambach, C., Walke, S. & Nagai, K. (1999) Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol*, **9** (2), 222–30.
- Kameoka, M., Rong, L., Gotte, M., Liang, C., Russell, R. S. & Wainberg, M. A. (2001) Role for human immunodeficiency virus type 1 Tat protein in suppression of viral reverse transcriptase activity during late stages of viral replication. *J Virol*, **75** (6), 2675–83.
- Kamma, H., Horiguchi, H., Wan, L., Matsui, M., Fujiwara, M., Fujimoto, M., Yazawa, T. & Dreyfuss, G. (1999) Molecular characterization of the hnRNP A2/B1 proteins: tissue-specific expression and novel isoforms. *Exp Cell Res*, **246** (2), 399–411.
- Kammler, S., Leurs, C., Freund, M., Krummheuer, J., Seidel, K., Tange, T. O., Lund, M. K., Kjems, J., Scheid, A. & Schaal, H. (2001) The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. *RNA*, **7** (3), 421–34.
- Kanopka, A., Mühlemann, O. & Akusjärvi, G. (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature*, **381** (6582), 535–8.
- Kanopka, A., Mühlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C. & Akusjärvi, G. (1998) Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature*, **393** (6681), 185–7.
- Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. (1987) Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*, **330** (6147), 489–93.
- Karn, J. (1999) Tackling Tat. *J Mol Biol*, **293** (2), 235–54.
- Karn, J. (2000) Tat, a novel regulator of hiv transcription and latency. In *HIV Sequence Compendium 2000*, (Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J.,

- Mullins, J., Wolinsky, S. & Korber, B., (eds), Theoretical Biology and Biophysics Group, Los Alamos National Laboratory Los Alamos pp. 2–19.
- Kastner, B., Bach, M. & Lührmann, R. (1990a) Electron microscopy of snRNPs U2, U4/6 and U5: evidence for a common structure-determining principle in the major UsnRNP family. *Mol Biol Rep*, **14** (2-3), 171.
- Kastner, B., Bach, M. & Lührmann, R. (1990b) Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. *Proc Natl Acad Sci U S A*, **87** (5), 1710–4.
- Kastner, B., Bach, M. & Lührmann, R. (1991) Electron microscopy of U4/U6 snRNP reveals a Y-shaped U4 and U6 RNA containing domain protruding from the U4 core RNP. *J Cell Biol*, **112** (6), 1065–72.
- Kastner, B. & Lührmann, R. (1989) Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5' RNA terminus. *EMBO J*, **8** (1), 277–86.
- Kataoka, N., Bachorik, J. L. & Dreyfuss, G. (1999) Transportin-SR, a nuclear import receptor for SRproteins. *J Cell Biol*, **145** (6).
- Kataoka, N., Yong, J., Kim, V. N., Velazquez, F., Perkinson, R. A., Wang, F. & Dreyfuss, G. (2000) Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell*, **6** (3), 673–82.
- Kenan, D. J., Query, C. C. & Keene, J. D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem Sci*, **16** (6), 214–20.
- Kennedy, C. F., Kramer, A. & Berget, S. M. (1998) A role for SRp54 during intron bridging of small introns with pyrimidine tracts upstream of the branch point. *Mol Cell Biol*, **18** (9), 5425–34.
- Kessler, M. & Mathews, M. B. (1992) Premature termination and processing of human immunodeficiency virus type 1-promoted transcripts. *J Virol*, **66** (7), 4488–96.
- Kielkopf, C. L., Rodionova, N. A., Green, M. R. & Burley, S. K. (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell*, **106** (5), 595–605.
- Kim, J. H., Hahn, B., Kim, Y. K., Choi, M. & Jang, S. K. (2000) Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J Mol Biol*, **298** (3), 395–405.
- Kjems, J., Frankel, A. D. & Sharp, P. A. (1991) Specific regulation of mRNA splicing *in vitro* by a peptide from HIV-1 Rev. *Cell*, **67** (1), 169–78.
- Kjems, J. & Sharp, P. A. (1993) The basic domain of Rev from human immunodeficiency virus type 1 specifically blocks the entry of U4/U6.U5 small nuclear ribonucleoprotein in spliceosome assembly. *J Virol*, **67** (8), 4769–76.
- Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Lührmann, R., Garcia-Blanco, M. A. & Manley, J. L. (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, **368** (6467), 119–24.
- Koloteva-Levine, N., Amichay, M. & Elroy-Stein, O. (2002) Interaction of hnRNP-C1/C2 proteins with RNA: analysis using the yeast three-hybrid system. *FEBS Lett*, **523** (1-3), 73–8.
- Konarska, M. M. & Sharp, P. A. (1987) Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell*, **49** (6), 763–74.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. & Bhattacharya, T. (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science*, **288** (5472), 1789–96.
- Kozu, T., Henrich, B. & Schafer, K. P. (1995) Structure and expression of the gene (HNRPA2B1) encoding the human hnRNP protein A2/B1. *Genomics*, **25** (2), 365–71.
- Krainer, A. R., Conway, G. C. & Kozak, D. (1990a) The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell*, **62** (1), 35–42.

- Krainer, A. R., Conway, G. C. & Kozak, D. (1990b) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev*, **4** (7), 1158–71.
- Krainer, A. R., Mayeda, A., Kozak, D. & Binns, G. (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell*, **66** (2), 383–94.
- Kramer, A. (1988) Presplicing complex formation requires two proteins and U2 snRNP. *Genes Dev*, **2** (9), 1155–67.
- Kramer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem*, **65**, 367–409.
- Kramer, A., Gruter, P., Groning, K. & Kastner, B. (1999) Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. *J Cell Biol*, **145** (7), 1355–68.
- Kramer, A., Quentin, M. & Mulhauser, F. (1998) Diverse modes of alternative splicing of human splicing factor SF1 deduced from the exon-intron structure of the gene. *Gene*, **211** (1), 29–37.
- Kramer, A. & Utans, U. (1991) Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *EMBO J*, **10** (6), 1503–9.
- Krecic, A. M. & Swanson, M. S. (1999) hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol*, **11** (3), 363–71.
- Krol, A., Branlant, C., Lazar, E., Gallinaro, H. & Jacob, M. (1981a) Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. homologies with U1 and U5 RNAs. *Nucleic Acids Res*, **9** (12), 2699–716.
- Krol, A., Gallinaro, H., Lazar, E., Jacob, M. & Branlant, C. (1981b) The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes. *Nucleic Acids Res*, **9** (4), 769–87.
- Krol, A., Westhof, E., Bach, M., Lührmann, R., Ebel, J. P. & Carbon, P. (1990) Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res*, **18** (13), 3803–11.
- Krummheuer, J., Lenz, C., Kammler, S., Scheid, A. & Schaal, H. (2001) Influence of the small leader exons 2 and 3 on human immunodeficiency virus type 1 gene expression. *Virology*, **286** (2), 276–89.
- Kuiken, C. L., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Wolinsky, S. & Korber, B. (2001) *HIV Sequence Compendium 2001*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.
- Kuo, B. A. & Norton, P. A. (1999) Accurate selection of a 5' splice site requires sequences within fibronectin alternative exon B. *Nucleic Acids Res*, **27** (19), 3945–52.
- Kuo, H. C., Nasim, F. H. & Grabowski, P. J. (1991) Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science*, **251** (4997), 1045–50.
- Labourier, E., Adams, M. D. & Rio, D. C. (2001) Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol Cell*, **8** (2), 363–73.
- Labourier, E., Blanchette, M., Feiger, J. W., Adams, M. D. & Rio, D. C. (2002) The KH-type RNA-binding protein PSI is required for drosophila viability, male fertility, and cellular mRNA processing. *Genes Dev*, **16** (1), 72–84.
- Laggerbauer, B., Achsel, T. & Lührmann, R. (1998) The human U5-200kD DEXH-box protein unwinds U4/u6 RNA duplexes *in vitro*. *Proc Natl Acad Sci U S A*, **95** (8), 4188–92.
- Laggerbauer, B., Lauber, J. & Lührmann, R. (1996) Identification of an RNA-dependent ATPase activity in mammalian U5 snRNPs. *Nucleic Acids Res*, **24** (5), 868–75.
- Lai, M. C., Lin, R. I., Huang, S. Y., Tsai, C. W. & Tarn, W. Y. (2000) A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J Biol Chem*, **275** (11), 7950–7.
- Lai, M. C., Lin, R. I. & Tarn, W. Y. (2001) Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc Natl Acad Sci U S A*, **98** (18), 10154–9.

- Lauber, J., Fabrizio, P., Teigelkamp, S., Lane, W. S., Hartmann, E. & Lührmann, R. (1996) The HeLa 200 kDa U5 snRNP-specific protein and its homologue in *Saccharomyces cerevisiae* are members of the DEXH-box protein family of putative RNA helicases. *EMBO J*, **15** (15), 4001–15.
- Lavigueur, A., La Branche, H., Kornblihtt, A. R. & Chabot, B. (1993) A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes Dev*, **7** (12A), 2405–17.
- Le Guiner, C., Lejeune, F., Galiana, D., Kister, L., Breathnach, R., Stévenin, J. & Del Gatto-Konczak, F. (2001a) TIA-1 and TIAR activate splicing of alternative exons with weak 5' splice sites followed by a U-rich stretch on their own pre-mRNAs. *J Biol Chem*, **276** (44), 40638–46.
- Le Guiner, C., Plet, A., Galiana, D., Gesnel, M. C., Del Gatto-Konczak, F. & Breathnach, R. (2001b) Polypyrimidine tract-binding protein represses splicing of a fibroblast growth factor receptor-2 gene alternative exon through exon sequences. *J Biol Chem*, **276** (47), 43677–87.
- Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. (2000) The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J*, **19** (24), 6860–9.
- Lecossier, D., Bouchonnet, F., Clavel, F. & Hance, A. (2003) Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science*, **300** (5622), 1112.
- Lejeune, F., Cavaloc, Y. & 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** (11), 7850–8.
- Lerga, A., Hallier, M., Delva, L., Orvain, C., Gallais, I., Marie, J. & Moreau-Gachelin, F. (2001) Identification of an RNA binding specificity for the potential splicing factor TLS. *J Biol Chem*, **276** (9), 6807–16.
- Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980) Are snRNPs involved in splicing? *Nature*, **283** (5743), 220–4.
- Levy, J. A. (1998) *HIV and the pathogenesis of AIDS*. 2nd edition,, American Society for Microbiology.
- Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, I. W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev*, **10** (13), 1683–98.
- Lillevali, K., Kulla, A. & Ord, T. (2001) Comparative expression analysis of the genes encoding polypyrimidine tract binding protein (PTB) and its neural homologue (brPTB) in prenatal and postnatal mouse brain. *Mech Dev*, **101** (1-2), 217–20.
- Lim, L. P. & Burge, C. B. (2001) A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A*, **98** (20), 11193–8.
- Linder, P. & Stutz, F. (2001) mRNA export: travelling with DEAD box proteins. *Curr Biol*, **11** (23), R961–3.
- Liu, H., Zhang, W., Reed, R. B., Liu, W. & Grabowski, P. J. (2002) Mutations in RRM4 uncouple the splicing repression and RNA-binding activities of polypyrimidine tract binding protein. *RNA*, **8** (2), 137–49.
- Liu, H. X., Chew, S. L., Cartegni, L., Zhang, M. Q. & Krainer, A. R. (2000) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol*, **20** (3), 1063–71.
- Liu, H. X., Zhang, M. & Krainer, A. R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev*, **12** (13), 1998–2012.
- Liu, X., Mayeda, A., Tao, M. & Zheng, Z. (2003) Exonic splicing enhancer-dependent selection of the bovine papillomavirus type 1 nucleotide 3225 3' splice site can be rescued in a cell lacking splicing factor ASF/SF2 through activation of the phosphatidylinositol 3-kinase/Akt pathway. *J Virol*, **77** (3), 2105–15.
- Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngrinou-Molango, S., Sprangers, R., Zanier, K., Kramer, A. & Sattler, M. (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*, **294** (5544), 1098–102.

- Liu, Z. R. (2002) p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex.
Mol Cell Biol, **22** (15), 5443–50.
- Loeb, D. D., Mack, A. A. & Tian, R. (2002) A secondary structure that contains the 5' and 3' splice sites suppresses splicing of duck hepatitis B virus pregenomic RNA.
J Virol, **76** (20), 10195–202.
- Lou, H. & Gagel, R. F. (1998) Alternative RNA processing—its role in regulating expression of calcitonin/calcitonin gene-related peptide.
J Endocrinol, **156** (3), 401–5.
- Lu, X. B., Heimer, J., Rekosh, D. & Hammarskjold, M. L. (1990) U1 small nuclear RNA plays a direct role in the formation of a rev-regulated human immunodeficiency virus env mRNA that remains unspliced.
Proc Natl Acad Sci U S A, **87** (19), 7598–602.
- Luo, M. J. & Reed, R. (1999) Splicing is required for rapid and efficient mRNA export in metazoans.
Proc Natl Acad Sci U S A, **96** (26), 14937–42.
- Luo, M. L., Zhou, Z., Magni, K., Christoforides, C., Rappaport, J., Mann, M. & Reed, R. (2001) Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly.
Nature, **413** (6856), 644–7.
- Lynch, K. W. & Maniatis, T. (1995) Synergistic interactions between two distinct elements of a regulated splicing enhancer.
Genes Dev, **9** (3), 284–93.
- Lynch, K. W. & Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer.
Genes Dev, **10** (16), 2089–101.
- Ma, A. S., Moran-Jones, K., Shan, J., Munro, T. P., Snee, M. J., Hoek, K. S. & Smith, R. (2002) Heterogeneous nuclear ribonucleoprotein A3, a novel RNA trafficking response element-binding protein.
J Biol Chem, **277** (20), 18010–20.
- Ma, M. & Nath, A. (1997) Molecular determinants for cellular uptake of Tat protein of human immunodeficiency virus type 1 in brain cells.
J Virol, **71** (3), 2495–9.
- Mackay, J. P. & Crossley, M. (1998) Zinc fingers are sticking together.
Trends Biochem Sci, **23** (1), 1–4.
- MacMillan, A. M., McCaw, P. S., Crispino, J. D. & Sharp, P. A. (1997) SC35-mediated reconstitution of splicing in U2AF-depleted nuclear extract.
Proc Natl Acad Sci U S A, **94** (1), 133–6.
- MacMillan, A. M., Query, C. C., Allerson, C. R., Chen, S., Verdine, G. L. & Sharp, P. A. (1994) Dynamic association of proteins with the pre-mRNA branch region.
Genes Dev, **8** (24), 3008–20.
- Madhani, H. D. & Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome.
Annu Rev Genet, **28**, 1–26.
- Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M. & Lührmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome.
Science, **298** (5601), 2205–8.
- Makarova, O. V., Makarov, E. M., Liu, S., Vormlocher, H. P. & Lührmann, R. (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing.
EMBO J, **21** (5), 1148–57.
- Makarova, O. V., Makarov, E. M. & Lührmann, R. (2001) The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes.
EMBO J, **20** (10), 2553–63.
- Maldarelli, F., Martin, M. A. & Strebel, K. (1991) Identification of posttranscriptionally active inhibitory sequences in human immunodeficiency virus type 1 RNA: novel level of gene regulation.
J Virol, **65** (11), 5732–43.
- Maldarelli, F., Xiang, C., Chamoun, G. & Zeichner, S. L. (1998) The expression of the essential nuclear splicing factor SC35 is altered by human immunodeficiency virus infection.
Virus Res, **53** (1), 39–51.
- Malim, M. H. & Cullen, B. R. (1991) HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency.
Cell, **65** (2), 241–8.

- Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V. & Cullen, B. R. (1989) The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA.
Nature, **338** (6212), 254–7.
- Mangeat, B., Turelli, P., Caron, G., Friedly, M., Perrin, L. & Trono, D. (28 May 2003).
Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.
Nature advance online publication.
- Maniatis, T. & Reed, R. (2002) An extensive network of coupling among gene expression machines.
Nature, **416** (6880), 499–506.
- Maniatis, T. & Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans.
Nature, **418** (6894), 236–43.
- Manley, J. L. & Tacke, R. (1996) SR proteins and splicing control.
Genes Dev, **10** (13), 1569–79.
- Mann, D. A., Mikaelian, I., Zimmel, R. W., Green, S. M., Lowe, A. D., Kimura, T., Singh, M., Butler, P. J., Gait, M. J. & Karn, J. (1994) A molecular rheostat. Cooperative rev binding to stem I of the rev- response element modulates human immunodeficiency virus type-1 late gene expression.
J Mol Biol, **241** (2), 193–207.
- Maquat, L. E. (2002) NASTy effects on fibrillin pre-mRNA splicing: another case of ESE does it, but proposals for translation-dependent splice site choice live on.
Genes Dev, **16** (14), 1743–53.
- Marchand, V., Méreau, A., Jacquenet, S., Thomas, D., Mougin, A., Gattoni, R., Stévenin, J. & Branlant, C. (2002) A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding.
J Mol Biol, **323** (4), 629–52.
- Markovtsov, V., Nikolic, J. M., Goldman, J. A., Turck, C. W., Chou, M. Y. & Black, D. L. (2000) Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein.
Mol Cell Biol, **20** (20), 7463–79.
- Maroney, P. A., Romfo, C. M. & Nilsen, T. W. (2000) Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly.
Mol Cell, **6** (2), 317–28.
- Massenet, S. & Branlant, C. (1999) A limited number of pseudouridine residues in the human atac spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs.
RNA, **5** (11), 1495–503.
- Massenet, S., Motorin, Y., Lafontaine, D. L., Hurt, E. C., Grosjean, H. & 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** (3), 2142–54.
- Massenet, S., Mougin, A. & Branlant, C. (1998) Posttranscriptional modifications in the U small nuclear RNAs. In *Modification and Editing of RNA*, (Grosjean, H. & Benne, R., eds), ASM Press Washington, DC.
- Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M. Q., Gardiner, E. M. & R., K. A. (1999) Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing.
EMBO J, **18** (16), 4560–70.
- Mayeda, A. & Krainer, A. R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2.
Cell, **68** (2), 365–75.
- Mayeda, A., Munroe, S. H., Cáceres, J. F. & Krainer, A. R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins.
EMBO J, **13** (22), 5483–95.
- Mayeda, A., Sreaton, G. R., Chandler, S. D., Fu, X. D. & Krainer, A. R. (1999) Substrate specificities of SR-proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mrna exonic elements.
Mol Cell Biol, **19** (3), 1853–63.
- McAfee, J. G., Shahied-Milam, L., Soltaninassab, S. R. & LeSturgeon, W. M. (1996) A major determinant of hnRNP C protein binding to RNA is a novel bZIP-like RNA binding domain.
RNA, **2** (11), 1139–52.

- McCarthy, E. M. & Phillips, J. A., r. (1998) Characterization of an intron splice enhancer that regulates alternative splicing of human GH pre-mRNA. *Hum Mol Genet*, **7** (9), 1491–6.
- McConnell, T. S. & Steitz, J. A. (2001) Proximity of the invariant loop of U5 snRNA to the second intron residue during pre-mRNA splicing. *EMBO J*, **20** (13), 3577–86.
- McCullough, A. J. & Berget, S. M. (1997) G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol Cell Biol*, **17** (8), 4562–71.
- McCullough, A. J. & Berget, S. M. (2000) An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5' splice sites. *Mol Cell Biol*, **20** (24), 9225–35.
- McNally, L. M. & McNally, M. T. (1996) SR protein splicing factors interact with the rous sarcoma virus negative regulator of splicing element. *J Virol*, **70** (2), 1163–72.
- McNally, L. M. & McNally, M. T. (1998) An RNA splicing enhancer-like sequence is a component of a splicing inhibitor element from Rous sarcoma virus. *Mol Cell Biol*, **18** (6), 3103–11.
- McNally, L. M. & McNally, M. T. (1999) U1 small nuclear ribonucleoprotein and splicing inhibition by the rous sarcoma virus negative regulator of splicing element. *J Virol*, **73** (3), 2385–93.
- McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., Fulton, R., Kucaba, T. A., Wagner-McPherson, C., Barbazuk, W. B., Gregory, S. G., Humphray, S. J., French, L., Evans, R. S., Bethel, G., Whittaker, A., Holden, J. L., McCann, O. T., Dunham, A., Soderlund, C., Scott, C. E., Bentley, D. R., Schuler, G., Chen, H. C., Jang, W., Green, E. D., Idol, J. R., Maduro, V. V., Montgomery, K. T., Lee, E., Miller, A., Emerling, S., Kucherlapati, Gibbs, R., Scherer, S., Gorell, J. H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garcia, D., de Jong, P. J., Catanese, J. J., Nowak, N., Osoegawa, K., Qin, S., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V. G., Kirsch, I. R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J. F., Hawkins, T., Myers, R. M., Shang, J., Ramirez, L., Schmutz, J., Velasquez, O., Dixon, K., Stone, N. E., Cox, D. R., Haussler, D., Kent, W. J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenberg, J., Chen, X. N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H. S., Sakaki, Y., Shimizu, N., Asakawa, S. *et al.* (2001) A physical map of the human genome. *Nature*, **409** (6822), 934–41.
- Meissner, M., Lopato, S., Gotzmann, J., Saueremann, G. & Barta, A. (2003) Proto-oncoprotein TLS/FUS is associated to the nuclear matrix and complexed with splicing factors PTB, SRm160, and SR proteins. *Exp Cell Res*, **283** (2), 184–95.
- Meister, G., Hannus, S., Plottner, O., Baars, T., Hartmann, E., Fakan, S., Lagerbauer, B. & Fischer, U. (2001) SM-Nrp is an essential pre-mRNA splicing factor required for the formation of the mature spliceosome. *EMBO J*, **20** (9), 2304–14.
- Mendell, J. T. & Dietz, H. C. (2001) When the message goes away: disease-producing mutations that influence mRNA content and performance. *Cell*, **107** (4), 411–4.
- Merendino, L., Guth, S., Bilbao, D., Martínez, C. & Valcárcel, J. (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature*, **402** (6763), 838–41.
- Mermoud, J. E., Cohen, P. & Lamond, A. I. (1992) Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res*, **20** (20), 5263–9.
- Mermoud, J. E., Cohen, P. T. & Lamond, A. I. (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J*, **13** (23), 5679–88.
- Michael, W. M., Choi, M. & Dreyfuss, G. (1995a) A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell*, **83** (3), 415–22.
- Michael, W. M., Siomi, H., Choi, M., Piñol Roma, S., Nakielny, S., Liu, Q. & Dreyfuss, G. (1995b) Signal sequences that target nuclear import and nuclear export of pre-mRNA-binding proteins. *Cold Spring Harb Symp Quant Biol*, **60**, 663–8.

- Miki, T., Bottaro, D. P., Fleming, T. P., Smith, C. L., Burgess, W. H., Chan, A. M. & 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** (1), 246–50.
- Mili, S., Shu, H. J., Zhao, Y. & Piñol Roma, S. (2001) Distinct RNP complexes of shuttling hnRNP proteins with pre-mRNA and mRNA: candidate intermediates in formation and export of mRNA.
Mol Cell Biol, **21** (21), 7307–19.
- Min, H., Chan, R. C. & Black, D. L. (1995) The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event.
Genes Dev, **9** (21), 2659–71.
- Min, H., Turck, C. W., Nikolic, J. M. & Black, D. L. (1997) A new regulatory protein, ksrp, mediates exon inclusion through an intronic splicing enhancer.
Genes Dev, **11** (8), 1023–36.
- Misteli, T. (2000) Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function.
J Cell Sci, **113** (Pt 11), 1841–9.
- Misteli, T., Cáceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F. & Spector, D. L. (1998) Serine phosphorylation of SRproteins is required for their recruitment to sites of transcription *in vivo*.
J Cell Biol, **143** (2), 297–307.
- Misteli, T. & Spector, D. L. (1996) Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors.
Mol Biol Cell, **7** (10), 1559–72.
- Modafferi, E. F. & Black, D. L. (1997) A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon.
Mol Cell Biol, **17** (11), 6537–45.
- Modrek, B. & Lee, C. (2002) A genomic view of alternative splicing.
Nat Genet, **30** (1), 13–9.
- Molin, M. & Akusjärvi, G. (2000) Overexpression of essential splicing factor ASF/SF2 blocks the temporal shift in adenovirus pre-mRNA splicing and reduces virus progeny formation.
J Virol, **74** (19), 9002–9.
- Molnar, G. M., Crozat, A., Kraeft, S. K., Dou, Q. P., Chen, L. B. & Pardee, A. B. (1997) Association of the mammalian helicase MAH with the pre-mRNA splicing complex.
Proc Natl Acad Sci U S A, **94** (15), 7831–6.
- Moore, M., Query, C. & Sharp, P. A. (1993) Splicing of precursors to messenger RNAs by the spliceosome. In *The RNA World*. Cold Spring Harbor Laboratory Press.
- Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J. L. & Proudfoot, N. J. (1998) The upstream sequence element of the C2 complement poly(a) signal activates mRNA 3' end formation by two distinct mechanisms.
Genes Dev, **12** (16), 2522–34.
- Mougin, A., Gottschalk, A., Fabrizio, P., Lührmann, R. & Branlant, C. (2002) Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles.
J Mol Biol, **317** (5), 631–49.
- Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M. & Branlant, C. (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency.
RNA, **2** (11), 1079–93.
- 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. & Carson, J. H. (2001) RNA trafficking signals in human immunodeficiency virus type 1.
Mol Cell Biol, **21** (6), 2133–43.
- Mount, S. M. & Steitz, J. A. (1981) Sequence of U1 RNA from *Drosophila melanogaster*: implications for U1 secondary structure and possible involvement in splicing.
Nucleic Acids Res, **9** (23), 6351–68.
- Muh, S. J., Hovhannisyanyan, R. H. & Carstens, R. P. (2002) A non-sequence-specific double-stranded RNA structural element regulates splicing of two mutually exclusive exons of fibroblast growth factor receptor 2 (fgfr2).
J Biol Chem, **277** (51), 50143–54.
- Muro, A. F., Caputi, M., Pariyath, R., Pagani, F., Barranti, E. & Baralle, F. E. (1999) Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display.
Mol Cell Biol, **19** (4), 2657–71.

- Myslinski, E. & Branlant, C. (1989) A new snRNA with a trimethylated cap structure in the slime mold *Physarum polycephalum*. *Nucleic Acids Res*, **17** (22), 9470.
- Myslinski, E., Branlant, C., Wieben, E. D. & Pederson, T. (1984) The small nuclear RNAs of *Drosophila*. *J Mol Biol*, **180** (4), 927–45.
- Myslinski, E., Wilhelm, F. X. & Branlant, C. (1989) A structural analysis of *P. polycephalum* U1 RNA at the RNA and gene levels. Are there differentially expressed U1 RNA genes in *P. polycephalum*? U1RNA evolution. *Nucleic Acids Res*, **17** (3), 1019–34.
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348** (6301), 515–20.
- Nagoshi, R. N. & Baker, B. S. (1990) Regulation of sex-specific RNA splicing at the *drosophila* doublesex gene: *cis*-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes Dev*, **4** (1), 89–97.
- Najera, I., Krieg, M. & Karn, J. (1999) Synergistic stimulation of HIV-1 rev-dependent export of unspliced mRNA to the cytoplasm by hnRNP A1. *J Mol Biol*, **285** (5), 1951–64.
- Nakielnny, S. & Dreyfuss, G. (1999) Transport of proteins and RNAs in and out of the nucleus. *Cell*, **99** (7), 677–90.
- Nasim, F. U., Hutchison, S., Cordeau, M. & Chabot, B. (2002) High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA*, **8** (8), 1078–89.
- Nasioulas, G., Zolotukhin, A. S., Taberner, C., Solomin, L., Cunningham, C. P., Pavlakis, G. N. & Felber, B. K. (1994) Elements distinct from human immunodeficiency virus type 1 splice sites are responsible for the Rev dependence of env mRNA. *J Virol*, **68** (5), 2986–93.
- Nelissen, R. L., Heinrichs, V., Habets, W. J., Simons, F., Lührmann, R. & van Venrooij, W. J. (1991) Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic Acids Res*, **19** (3), 449–54.
- Nesic, D. & Kramer, A. (2001) Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17s U2 snRNP, and pre-spliceosome formation. *Mol Cell Biol*, **21** (19), 6406–17.
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A. & Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet*, **20** (1), 46–50.
- Neuveut, C. & Jeang, K. T. (1996) Recombinant human immunodeficiency virus type 1 genomes with *tat* unconstrained by overlapping reading frames reveal residues in *tat* important for replication in tissue culture. *J Virol*, **70** (8), 5572–81.
- Newby, M. I. & Greenbaum, N. L. (2001) A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA*, **7** (6), 833–45.
- Newnham, C. M. & Query, C. C. (2001) The ATP requirement for U2 snRNP addition is linked to the pre-mRNA region 5' to the branch site. *RNA*, **7** (9), 1298–309.
- Nilsen, T. W. (2002) The spliceosome: no assembly required? *Mol Cell*, **9** (1), 8–9.
- Niwa, M. & Berget, S. M. (1991a) Polyadenylation precedes splicing *in vitro*. *Gene Expr*, **1** (1), 5–14.
- Niwa, M. & Berget, S. M. (1991b) Mutation of the AAUAAA polyadenylation signal depresses *in vitro* splicing of proximal but not distal introns. *Genes Dev*, **5** (11), 2086–95.
- Nogués, G., Kadener, S., Cramer, P., Bentley, D. & Kornblihtt, A. R. (2002) Transcriptional activators differ in their abilities to control alternative splicing. *J Biol Chem*, **277** (45), 43110–14.
- Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R. & Lührmann, R. (1999) Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J*, **18** (21), 6119–33.

- Nussinov, R. (1988) Conserved quartets near 5' intron junctions in primate nuclear pre-mRNA.
J Theor Biol, **133** (1), 73–84.
- Oh, Y. L., Hahm, B., Kim, Y. K., Lee, H. K., Lee, J. W., Song, O., Tsukiyama-Kohara, K., Kohara, M., Nomoto, A. & Jang, S. K. (1998) Determination of functional domains in polypyrimidine-tract-binding protein.
Biochem J, **331** (Pt 1), 169–75.
- Ohno, M., Sakamoto, H. & Shimura, Y. (1987) Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure.
Proc Natl Acad Sci U S A, **84** (15), 5187–91.
- Ohno, M. & Shimura, Y. (1996) A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome.
Genes Dev, **10** (8), 997–1007.
- Olsen, H. S., Cochrane, A. W. & Rosen, C. (1992) Interaction of cellular factors with intragenic *cis*-acting repressive sequences within the HIV genome.
Virology, **191** (2), 709–15.
- ONUSIDA/OMS (2002) *Le point sur l'épidémie de SIDA*. Programme commun des Nations-Unies sur le VIH/SIDA (ONUSIDA) et l'Organisation Mondiale de la Santé (OMS) Genève.
- O'Reilly, M. M., McNally, M. T. & Beemon, K. L. (1995) Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA.
Virology, **213** (2), 373–85.
- Ou, S. H., Wu, F., Harrich, D., Garcia-Martinez, L. F. & Gaynor, R. B. (1995) Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 target DNA sequence motifs.
J Virol, **69** (6), 3584–96.
- Pagani, F., Buratti, E., Stuaní, C., Romano, M., Zuccato, E., Niksic, M., Giglio, L., Faraguna, D. & Baralle, F. E. (2000) Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element.
J Biol Chem, **275** (28), 21041–7.
- Parkin, N. T., Chamorro, M. & Varmus, H. E. (1992) Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo.
J Virol, **66** (8), 5147–51.
- Pavé, A. (1994) *Modélisation en biologie et en écologie*. Aléas.
- Pavlakakis, G. N. & Felber, B. K. (1990) Regulation of expression of human immunodeficiency virus.
New Biol, **2** (1), 20–31.
- Peeters, M. & Sharp, P. M. (2000) Genetic diversity of HIV-1: the moving target.
AIDS, **14** (Suppl 3), S129–40.
- Peled-Zehavi, H., Berglund, J. A., Rosbash, M. & 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** (15), 5232–41.
- Perez, I., McAfee, J. G. & Patton, J. G. (1997) Multiple RRM domains contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein.
Biochemistry, **36** (39), 11881–90.
- Perrone-Bizzozero, N. & Bolognani, F. (2002) Role of HuD and other RNA-binding proteins in neural development and plasticity.
J Neurosci Res, **68** (2), 121–6.
- Philips, A. V. & Cooper, T. A. (2000) RNA processing and human disease.
Cell Mol Life Sci, **57** (2), 235–49.
- Philips, A. V., Timchenko, L. T. & Cooper, T. A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy.
Science, **280** (5364), 737–41.
- Piñol Roma, S. & Dreyfuss, G. (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm.
Nature, **355** (6362), 730–2.
- Piñol Roma, S. & Dreyfuss, G. (1993) Cell cycle-regulated phosphorylation of the pre-mRNA-binding (heterogeneous nuclear ribonucleoprotein) C proteins.
Mol Cell Biol, **13** (9), 5762–70.
- Pollard, V. W., Michael, W. M., Nakielny, S., Siomi, M. C., Wang, F. & Dreyfuss, G. (1996) A novel receptor-mediated nuclear protein import pathway.
Cell, **86** (6), 985–94.
- Pongoski, J., Asai, K. & Cochrane, A. (2002) Positive and negative modulation of human immunodeficiency virus

- type 1 Rev function by *cis* and *trans* regulators of viral RNA splicing.
J Virol, **76** (10), 5108–20.
- Popielarz, M., Cavaloc, Y., Mattei, M. G., Gattoni, R. & Stévenin, J. (1995) The gene encoding human splicing factor 9G8. structure, chromosomal localization, and expression of alternatively processed transcripts.
J Biol Chem, **270** (30), 17830–5.
- Powell, D. M., Amaral, M. C., Wu, J. Y., Maniatis, T. & Greene, W. C. (1997) HIV Rev-dependent binding of SF2/ASF to the Rev response element: possible role in Rev-mediated inhibition of HIV RNA splicing.
Proc Natl Acad Sci U S A, **94** (3), 973–8.
- Prasad, J., Colwill, K., Pawson, T. & Manley, J. L. (1999) The protein kinase Clk/Sty directly modulates SRprotein activity: both hyper- and hypophosphorylation inhibit splicing.
Mol Cell Biol, **19** (10), 6991–7000.
- Proudfoot, N. J., Furger, A. & Dye, M. J. (2002) Integrating mRNA processing with transcription.
Cell, **108** (4), 501–12.
- Purcell, D. F. & Martin, M. A. (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity.
J Virol, **67** (11), 6365–78.
- Qu, H. L., Michot, B. & Bachellerie, J. P. (1983) Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5' terminal domain of eukaryotic 28S rRNA.
Nucleic Acids Res, **11** (17), 5903–20.
- Query, C. C., Strobel, S. A. & Sharp, P. A. (1995) The branch site adenosine is recognized differently for the two steps of pre-mRNA splicing.
Nucleic Acids Symp Ser, **33**, 224–5.
- Query, C. C., Strobel, S. A. & Sharp, P. A. (1996) Three recognition events at the branch-site adenine.
EMBO J, **15** (6), 1392–402.
- Raffetseder, U., Frye, B. C., Rauen, T., Jurchott, K., Royer, H. D., Lynen-Jansen, P. & 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, **25**, 25.
- Rahman, L., Bliskovski, V., Reinhold, W. & Zajac-Kaye, M. (2002) Alternative splicing of brain-specific PTB defines a tissue-specific isoform pattern that predicts distinct functional roles.
Genomics, **80** (3), 245–9.
- Rain, J. C., Rafi, Z., Rhani, Z., Legrain, P. & 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** (5), 551–65.
- Ramchatesingh, J., Zahler, A. M., Neugebauer, K. M., Roth, M. B. & Cooper, T. A. (1995) A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer.
Mol Cell Biol, **15** (9), 4898–907.
- Rappsilber, J., Ajuh, P., Lamond, A. I. & Mann, M. (2001) SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome.
J Biol Chem, **276** (33), 31142–50.
- Reed, R. (1989) The organization of 3' splice-site sequences in mammalian introns.
Genes Dev, **3** (12B), 2113–23.
- Reed, R. (1990) Protein composition of mammalian spliceosomes assembled *in vitro*.
Proc Natl Acad Sci U S A, **87** (20), 8031–5.
- Reed, R. (2000) Mechanisms of fidelity in pre-mRNA splicing.
Curr Opin Cell Biol, **12** (3), 340–5.
- Reed, R. & Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing.
Cell, **108** (4), 523–31.
- Reyes, J. L., Kois, P., Konforti, B. B. & Konarska, M. M. (1996) The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome.
RNA, **2** (3), 213–25.
- Ring, H. Z. & Lis, J. T. (1994) The SRprotein B52/SRp55 is essential for *Drosophila* development.
Mol Cell Biol, **14** (11), 7499–506.
- Rinke, J., Appel, B., Digweed, M. & Lührmann, R. (1985) Localization of a base-paired interaction between small

- nuclear RNAs U4 and U6 in intact U4/u6 ribonucleoprotein particles by psoralen cross-linking.
J Mol Biol, **185** (4), 721–31.
- Rio, D. C., Laski, F. A. & Rubin, G. M. (1986) Identification and immunochemical analysis of biologically active *Drosophila* P element transposase.
Cell, **44** (1), 21–32.
- Robberson, B. L., Côté, G. J. & Berget, S. M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons.
Mol Cell Biol, **10** (1), 84–94.
- Roberts, G. C. & Smith, C. W. (2002) Alternative splicing: combinatorial output from the genome.
Curr Opin Chem Biol, **6** (3), 375–83.
- Romfo, C. M. & Wise, J. A. (1997) Both the polypyrimidine tract and the 3' splice site function prior to the first step of splicing in fission yeast.
Nucleic Acids Res, **25** (22), 4658–65.
- Rooke, N., Markovtsov, V., Cagavi, E. & Black, D. L. (2003) Roles for SR Proteins and hnRNP A1 in the Regulation of c-src Exon N1.
Mol Cell Biol, **23** (6), 1874–84.
- Roscigno, R. F. & Garcia-Blanco, M. A. (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome.
RNA, **1** (7), 692–706.
- Rosen, C. A., Terwilliger, E., Dayton, A., Sodroski, J. G. & Haseltine, W. A. (1988) Intragenic *cis*-acting art gene-responsive sequences of the human immunodeficiency virus.
Proc Natl Acad Sci U S A, **85** (7), 2071–5.
- Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C. & Tazi, J. (1996) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I.
Nature, **381** (6577), 80–2.
- Roth, M. B., Zahler, A. M. & Stolk, J. A. (1991) A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription.
J Cell Biol, **115** (3), 587–96.
- Rowen, L., Young, J., Birditt, B., Kaur, A., Madan, A., Philipps, D. L., Qin, S., Minx, P., Wilson, R. K., Hood, L. & Graveley, B. R. (2002) Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity.
Genomics, **79** (4), 587–97.
- Rueter, S. M., Dawson, T. R. & Emeson, R. B. (1999) Regulation of alternative splicing by RNA editing.
Nature, **399** (6731), 75–80.
- Ryo, A., Suzuki, Y., Arai, M., Kondoh, N., Wakatsuki, T., Hada, A., Shuda, M., Tanaka, K., Sato, C., Yamamoto, M. & Yamamoto, N. (2000) Identification and characterization of differentially expressed mRNAs in HIV type 1-infected human T cells.
AIDS Res Hum Retroviruses, **16** (10), 995–1005.
- Salemi, M., Strimmer, K., Hall, W. W., Duffy, M., Delaporte, E., Mboup, S., Peeters, M. & Vandamme, A. M. (2001) Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution.
Faseb J, **15** (2), 276–8.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular cloning. A laboratory manual*. 2nd edition., C. S. H. L. Press, New-York.
- Sawa, H. & Shimura, Y. (1992) Association of U6 snRNA with the 5'-splice site region of pre-mRNA in the spliceosome.
Genes Dev, **6** (2), 244–54.
- Schaal, H., Freund, M., Kammler, S., Asang, C. & Caputi, M. (2003) A bidirectional SR protein-dependent exonic splicing enhancer regulates *rev*, *env*, *vpu* and *nef* gene expression.
In Eukaryotic mRNA Processing.
- Schaal, T. D. & Maniatis, T. (1999a) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA.
Mol Cell Biol, **19** (1), 261–73.
- Schaal, T. D. & Maniatis, T. (1999b) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences.
Mol Cell Biol, **19** (3), 1705–19.
- Schaub, M. & Keller, W. (2002) RNA editing by adenosine deaminases generates RNA and protein diversity.
Biochimie, **84** (8), 791–803.
- Schwartz, S., Felber, B. K., Fenyo, E. M. & Pavlakis, G. N. (1990) Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs.
J Virol, **64** (11), 5448–56.

- Schwartz, S., Felber, B. K. & Pavlakis, G. N. (1992) Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of Rev protein. *J Virol*, **66** (1), 150–9.
- Schwer, B. (2001) A new twist on RNA helicases: DExH/D box proteins as RNPs. *Nat Struct Biol*, **8** (2), 113–6.
- Screaton, G. R., Cáceres, J. F., Mayeda, A., Bell, M. V., Plebanski, M., Jackson, D. G., Bell, J. I. & Krainer, A. R. (1995) Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. *EMBO J*, **14** (17), 4336–49.
- Ségault, V., Will, C. L., Polycarpou-Schwarz, M., Mattaj, I. W., Branlant, C. & 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** (4), 2782–90.
- Ségault, V., Will, C. L., Sproat, B. S. & Lührmann, R. (1995) *In vitro* reconstitution of mammalian U2 and U5 snRNPs active in splicing: Sm proteins are functionally interchangeable and are essential for the formation of functional U2 and U5 snRNPs. *EMBO J*, **14** (16), 4010–21.
- Selby, M. J., Bain, E. S., Luciw, P. A. & Peterlin, B. M. (1989) Structure, sequence, and position of the stem-loop in TAR determine transcriptional elongation by tat through the HIV-1 long terminal repeat. *Genes Dev*, **3** (4), 547–58.
- Selvakumar, M. & Helfman, D. M. (1999) Exonic splicing enhancers contribute to the use of both 3' and 5' splice site usage of rat beta-tropomyosin pre-mRNA. *RNA*, **5** (3), 378–94.
- Shahied-Milam, L., Soltaninassab, S. R., Iyer, G. V. & LeStourgeon, W. M. (1998) The heterogeneous nuclear ribonucleoprotein C protein tetramer binds U1, U2, and U6 snRNAs through its high affinity RNA binding domain (the bZIP-like motif). *J Biol Chem*, **273** (33), 21359–67.
- Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. & Steitz, T. A. (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol*, **4** (3), 215–22.
- Shatkin, A. J. & Manley, J. L. (2000) The ends of the affair: capping and polyadenylation. *Nat Struct Biol*, **7** (10), 838–42.
- Shelley, C. S. & Baralle, F. E. (1987) Deletion analysis of a unique 3' splice site indicates that alternating guanine and thymine residues represent an efficient splicing signal. *Nucleic Acids Res*, **15** (9), 3787–99.
- Shepard, J., Reick, M., Olson, S. & Graveley, B. R. (2002) Characterization of U2AF(26), a splicing factor related to U2AF(35). *Mol Cell Biol*, **22** (1), 221–30.
- Sherman, M. P. & Greene, W. C. (2002) Slipping through the door: HIV entry into the nucleus. *Microbes Infect*, **4** (1), 67–73.
- Shimba, S., Bokar, J. A., Rottman, F. & Reddy, R. (1995) Accurate and efficient N⁶-adenosine methylation in spliceosomal U6 small nuclear RNA by HeLa cell extract *in vitro*. *Nucleic Acids Res*, **23** (13), 2421–6.
- Shin, C. & Manley, J. L. (2002) The SR protein SRp38 represses splicing in M phase cells. *Cell*, **111** (3), 407–17.
- Shnyreva, M., Schullery, D. S., Suzuki, H., Higaki, Y. & Bomsztyk, K. (2000) Interaction of two multifunctional proteins. heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein. *J Biol Chem*, **275** (20), 15498–503.
- Shomron, N., Malca, H., Vig, I. & Ast, G. (2002) Reversible inhibition of the second step of splicing suggests a possible role of zinc in the second step of splicing. *Nucleic Acids Res*, **30** (19), 4127–37.
- Si, Z., Amendt, B. A. & Stoltzfus, C. M. (1997) Splicing efficiency of human immunodeficiency virus type 1 tat RNA is determined by both a suboptimal 3' splice site and a 10 nucleotide exon splicing silencer element located within tat exon 2. *Nucleic Acids Res*, **25** (4), 861–7.
- Si, Z. H., Rauch, D. & Stoltzfus, C. M. (1998) The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. *Mol Cell Biol*, **18** (9), 5404–13.

- Siatecka, M., Reyes, J. L. & Konarska, M. M. (1999) Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev*, **13** (15), 1983–93.
- Siebel, C. W., Admon, A. & Rio, D. C. (1995) Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev*, **9** (3), 269–83.
- Siebel, C. W., Fresco, L. D. & Rio, D. C. (1992) The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control u1 snRNP binding. *Genes Dev*, **6** (8), 1386–401.
- Siebel, C. W., Kanaar, R. & Rio, D. C. (1994) Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes Dev*, **8** (14), 1713–25.
- Simard, M. J. & Chabot, B. (2000) Control of hnRNP A1 alternative splicing: an intron element represses use of the common 3' splice site. *Mol Cell Biol*, **20** (19), 7353–62.
- Simard, M. J. & Chabot, B. (2002) SRp30c is a repressor of 3' splice site utilization. *Mol Cell Biol*, **22** (12), 4001–10.
- Simon, F., Mauciere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F. & Brun-Vezinet, F. (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med*, **4** (9), 1032–7.
- Singer, R. H. (1998) Triplet-repeat transcripts: a role for DNA in disease. *Science*, **280** (5364), 696–7.
- Siomi, H. & Dreyfuss, G. (1995) A nuclear localization domain in the hnRNP A1 protein. *J Cell Biol*, **129** (3), 551–60.
- Sirand-Pugnet, P., Durosay, P., Clouet d'Orval, B. C., Brody, E. & Marie, J. (1995) beta-tropomyosin pre-mRNA folding around a muscle-specific exon interferes with several steps of spliceosome assembly. *J Mol Biol*, **251** (5), 591–602.
- Sleasman, J. W. & Goodenow, M. M. (2003) HIV-1 infection. *J Allergy Clin Immunol*, **111** (2 Suppl), S582–92.
- Smith, C. W. & Valcárcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, **25** (8), 381–8.
- Solovyev, V. V., Salamov, A. A. & Lawrence, C. B. (1994) Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res*, **22** (24), 5156–63.
- Sontheimer, E. J. (2001) The spliceosome shows its metal. *Nat Struct Biol*, **8** (1), 11–3.
- Sontheimer, E. J., Gordon, P. M. & Piccirilli, J. A. (1999) Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev*, **13** (13), 1729–41.
- Sontheimer, E. J. & Steitz, J. A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262** (5142), 1989–96.
- Sonza, S., Mutimer, H. P., O'Brien, K., Ellery, P., Howard, J. L., Axelrod, J. H., Deacon, N. J., Crowe, S. M. & Purcell, D. F. (2002) Selectively reduced tat mRNA heralds the decline in productive human immunodeficiency virus type 1 infection in monocyte-derived macrophages. *J Virol*, **76** (24), 12611–21.
- Soret, J., Gattoni, R., Guyon, C., Sureau, A., Popielarz, M., Le Rouzic, E., Dumon, S., Apiou, F., Dutrillaux, B., Voss, H., Ansorge, W., Stévenin, J. & Perbal, B. (1998) Characterization of SRp46, a novel human SRsplicing factor encoded by a PR264/SC35 retroseudogene. *Mol Cell Biol*, **18** (8), 4924–34.
- Staffa, A. & Cochrane, A. (1995) Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. *Mol Cell Biol*, **15** (8), 4597–605.
- Staknis, D. & Reed, R. (1994a) SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol Cell Biol*, **14** (11), 7670–82.
- Staknis, D. & Reed, R. (1994b) Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol Cell Biol*, **14** (5), 2994–3005.

- Staley, J. P. & Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, **92** (3), 315–26.
- Stark, H., Dube, P., Lührmann, R. & Kastner, B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, **409** (6819), 539–42.
- Stark, J. M., Bazett-Jones, D. P., Herfort, M. & Roth, M. B. (1998) SR proteins are sufficient for exon bridging across an intron. *Proc Natl Acad Sci U S A*, **95** (5), 2163–8.
- Stark, J. M., Cooper, T. A. & Roth, M. B. (1999) The relative strengths of SR protein-mediated associations of alternative and constitutive exons can influence alternative splicing. *J Biol Chem*, **274** (42), 29838–42.
- Steitz, T. A. & Steitz, J. A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A*, **90** (14), 6498–502.
- Sterner, D. A., Carlo, T. & Berget, S. M. (1996) Architectural limits on split genes. *Proc Natl Acad Sci U S A*, **93** (26), 15081–5.
- Stevens, S. W. & Abelson, J. (1999) Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc Natl Acad Sci U S A*, **96** (13), 7226–31.
- Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D. & Abelson, J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell*, **9** (1), 31–44.
- Stickeler, E., Fraser, S. D., Honig, A., Chen, A. L., Berget, S. M. & 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** (14), 3821–30.
- Stickeler, E., Kittrell, F., Medina, D. & Berget, S. M. (1999) Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. *Oncogene*, **18** (24), 3574–82.
- Strack, B., Calistri, A., Accola, M. A., Palu, G. & Gottlinger, H. G. (2000) A role for ubiquitin ligase recruitment in retrovirus release. *Proc Natl Acad Sci U S A*, **97** (24), 13063–8.
- Stutz, F. & Rosbash, M. (1998) Nuclear RNA export. *Genes Dev*, **12** (21), 3303–19.
- Sun, Q., Mayeda, A., Hampson, R. K., Krainer, A. R. & Rottman, F. M. (1993) General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes Dev*, **7** (12B), 2598–608.
- Sureau, A., Gattoni, R., Dooghe, Y., Stévenin, J. & Soret, J. (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J*, **20** (7), 1785–96.
- Swanson, A. K. & Stoltzfus, C. M. (1998) Overlapping *cis* sites used for splicing of HIV-1 *env/nef* and *rev* mRNAs. *J Biol Chem*, **273** (51), 34551–7.
- Swanson, M. S. & Dreyfuss, G. (1988) RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J*, **7** (11), 3519–29.
- Tabor, S. & Richardson, C. C. (1987) DNA sequence analysis with a modified T7 DNA polymerase. *Proc Natl Acad Sci U S A*, **84** (14), 4767–71.
- Tacke, R., Chen, Y. & Manley, J. L. (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc Natl Acad Sci U S A*, **94** (4), 1148–53.
- Tacke, R. & Manley, J. L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J*, **14** (14), 3540–51.
- Tacke, R., Tohyama, M., Ogawa, S. & Manley, J. L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93** (1), 139–48.
- Tang, H., Kuhen, K. L. & Wong-Staal, F. (1999) Lentivirus replication and regulation. *Annu Rev Genet*, **33**, 133–70.
- Tange, T. O., Damgaard, C. K., Guth, S., Valcarcel, J. & Kjems, J. (2001) The hnRNP A1 protein regulates HIV-1 *tat* splicing via a novel intron silencer element. *EMBO J*, **20** (20), 5748–58.
- Tange, T. O., Jensen, T. H. & Kjems, J. (1996) *In vitro* interaction between human immunodeficiency virus type 1

- Rev protein and splicing factor ASF/SF2-associated protein, p32.
J Biol Chem, **271** (17), 10066–72.
- Tange, T. O. & Kjems, J. (2001) SF2/ASF binds to a splicing enhancer in the third HIV-1 tat exon and stimulates U2AF binding independently of the RS domain.
J Mol Biol, **312** (4), 649–62.
- Tarn, W. Y. & Steitz, J. A. (1994) SR proteins can compensate for the loss of U1 snRNP functions *in vitro*.
Genes Dev, **8** (22), 2704–17.
- Tarn, W. Y. & Steitz, J. A. (1995) Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps.
Proc Natl Acad Sci U S A, **92** (7), 2504–8.
- Tarn, W. Y. & Steitz, J. A. (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge.
Trends Biochem Sci, **22** (4), 132–7.
- Tazi, J., Rossi, F., Labourier, E., Gallouzi, I., Brunel, C. & Antoine, E. (1997) DNA topoisomerase I: customs officer at the border between DNA and RNA worlds?
J Mol Med, **75** (11–12), 786–800.
- Teigelkamp, S., Achsel, T., Mundt, C., Gothel, S. F., Cronshagen, U., Lane, W. S., Marahiel, M. & Lührmann, R. (1998) The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins.
RNA, **4** (2), 127–41.
- Teigelkamp, S., Newman, A. J. & Beggs, J. D. (1995) Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA.
EMBO J, **14** (11), 2602–12.
- Temsamani, J. & Pederson, T. (1996) The C-group heterogeneous nuclear ribonucleoprotein proteins bind to the 5' stem-loop of the U2 small nuclear ribonucleoprotein particle.
J Biol Chem, **271** (40), 24922–6.
- Ten Dam, G. B., Zilch, C. F., Wallace, D., Wieringa, B., Beverley, P. C., Poels, L. G. & Screaton, G. R. (2000) Regulation of alternative splicing of CD45 by antagonistic effects of SR protein splicing factors.
J Immunol, **164** (10), 5287–95.
- Thali, M., Bukovsky, A., Kondo, E., Rosenwirth, B., Walsh, C. T., Sodroski, J. & Gottlinger, H. G. (1994) Functional association of cyclophilin A with HIV-1 virions.
Nature, **372** (6504), 363–5.
- Tian, H. & Kole, R. (2001) Strong RNA splicing enhancers identified by a modified method of cycled selection interact with SR protein.
J Biol Chem, **276** (36), 33833–9.
- Tian, M. & Maniatis, T. (1994) A splicing enhancer exhibits both constitutive and regulated activities.
Genes Dev, **8** (14), 1703–12.
- Tian, Q., Streuli, M., Saito, H., Schlossman, S. F. & Anderson, P. (1991) A polyadenylate binding protein localized to the granules of cytolytic lymphocytes induces DNA fragmentation in target cells.
Cell, **67** (3), 629–39.
- Timchenko, N. A., Welm, A. L., Lu, X. & Timchenko, L. T. (1999) CUG repeat binding protein (cugbp1) interacts with the 5' region of C/EBPbeta mRNA and regulates translation of c/ebpbeta isoforms.
Nucleic Acids Res, **27** (22), 4517–25.
- Tsukahara, T., Casciato, C. & Helfman, D. M. (1994) Alternative splicing of beta-tropomyosin pre-mRNA: multiple cis-elements can contribute to the use of the 5'- and 3'-splice sites of the nonmuscle/smooth muscle exon 6.
Nucleic Acids Res, **22** (12), 2318–25.
- Tuerk, C. & Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.
Science, **249** (4968), 505–10.
- Urlaub, H., Raker, V. A., Kostka, S. & Lührmann, R. (2001) Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure.
EMBO J, **20** (1–2), 187–96.
- Valcárcel, J., Gaur, R. K., Singh, R. & Green, M. R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected].
Science, **273** (5282), 1706–9.
- Valcárcel, J., Singh, R., Zamore, P. D. & Green, M. R. (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA.
Nature, **362** (6416), 171–5.
- Valentine, C. R. (1998) The association of nonsense codons with exon skipping.
Mutat Res, **411** (2), 87–117.

- van der Houven van Oordt, W., Diaz-Meco, M. T., Lozano, J., Krainer, A. R., Moscat, J. & Cáceres, J. F. (2000) The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J Cell Biol*, **149** (2), 307–16.
- Van Lint, C., Emiliani, S., Ott, M. & Verdin, E. (1996) Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO J*, **15** (5), 1112–20.
- van 't Wout, A., Lehrman, G. K., Mikheeva, S. A., O'Keeffe, G. C., Katze, M. G., Bumgarner, R. E., Geiss, G. K. & I., M. J. (2003) Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4⁺-T-cell lines. *AIDS Res Hum Retroviruses*, **77** (2), 1392–1402.
- Vanden Haesevelde, M. M., Peeters, M., Jannes, G., Janssens, W., van der Groen, G., Sharp, P. M. & Saman, E. (1996) Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee. *Virology*, **221** (2), 346–50.
- Varani, L., Hasegawa, M., Spillantini, M. G., Smith, M. J., Murrell, J. R., Ghetti, B., Klug, A., Goedert, M. & Varani, G. (1999) Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc Natl Acad Sci U S A*, **96** (14), 8229–34.
- Vassilenko, S. K., Carbon, P., Ebel, J. P. & Ehresmann, C. (1981) Topography of 16 S RNA in 30 S subunits and 70 S ribosomes accessibility to cobra venom ribonuclease. *J Mol Biol*, **152** (4), 699–721.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C. *et al.* (2001) The sequence of the human genome. *Science*, **291** (5507), 1304–51.
- Verhoef, K., Bauer, M., Meyerhans, A. & Berkhout, B. (1998) On the role of the second coding exon of the HIV-1 Tat protein in virus replication and MHC class I down-regulation. *AIDS Res Hum Retroviruses*, **14** (17), 1553–9.
- Verhoef, K., Bilodeau, P. S., van Wamel, J. L., Kijms, J., Stoltzfus, C. M. & Berkhout, B. (2001) Repair of a Revminus human immunodeficiency virus type 1 mutant by activation of a cryptic splice site. *J Virol*, **75** (7), 3495–500.
- Vidal, V. P., Verdone, L., Mayes, A. E. & Beggs, J. D. (1999) Characterization of U6 snRNA-protein interactions. *RNA*, **5** (11), 1470–81.
- Viglianti, G. A. & Mullins, J. I. (1988) Functional comparison of transactivation by simian immunodeficiency virus from rhesus macaques and human immunodeficiency virus type 1. *J Virol*, **62** (12), 4523–32.
- Villa, T., Pleiss, J. A. & Guthrie, C. (2002) Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell*, **109** (2), 149–52.
- Vitali, J., Ding, J., Jiang, J., Zhang, Y., Krainer, A. R. & Xu, R. M. (2002) Correlated alternative side chain conformations in the RNA-recognition motif of heterogeneous nuclear ribonucleoprotein A1. *Nucleic Acids Res*, **30** (7), 1531–8.
- Wagner, E. J. & Garcia-Blanco, M. A. (2001) Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol*, **21** (10), 3281–8.
- Wagner, E. J. & Garcia-Blanco, M. A. (2002) RNAi-mediated PTB depletion leads to enhanced exon definition. *Mol Cell*, **10** (4), 943–9.

- Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C. & Fu, X. D. (1998) SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol*, **140** (4), 737–50.
- Wang, J., Chang, Y. F., Hamilton, J. I. & Wilkinson, M. F. (2002a) Nonsense-associated altered splicing: a frame-dependent response distinct from nonsense-mediated decay. *Mol Cell*, **10** (4), 951–7.
- Wang, J., Hamilton, J. I., Carter, M. S., Li, S. & Wilkinson, M. F. (2002b) Alternatively spliced TCR mRNA induced by disruption of reading frame. *Science*, **297** (5578), 108–10.
- Wang, J. & Manley, J. L. (1995) Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing *in vivo* in diverse ways. *RNA*, **1** (3), 335–46.
- Wang, J. & Pederson, T. (1990) A 62,000 molecular weight spliceosome protein crosslinks to the intron polypyrimidine tract. *Nucleic Acids Res*, **18** (20), 5995–6001.
- Wassarman, D. A. & Steitz, J. A. (1992) Interactions of small nuclear RNA's with precursor messenger RNA during *in vitro* splicing. *Science*, **257** (5078), 1918–25.
- Watakabe, A., Tanaka, K. & Shimura, Y. (1993) The role of exon sequences in splice site selection. *Genes Dev*, **7** (3), 407–18.
- Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H. & Jones, K. A. (1998) A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*, **92** (4), 451–62.
- Weighardt, F., Biamonti, G. & Riva, S. (1995) Nucleocytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. *J Cell Sci*, **108** (Pt 2), 545–55.
- Wentz, M. P., Moore, B. E., Cloyd, M. W., Berget, S. M. & Donehower, L. A. (1997) A naturally arising mutation of a potential silencer of exon splicing in human immunodeficiency virus type 1 induces dominant aberrant splicing and arrests virus production. *J Virol*, **71** (11), 8542–51.
- Wentz-Hunter, K. & Potashkin, J. (1996) The small subunit of the splicing factor U2AF is conserved in fission yeast. *Nucleic Acids Res*, **24** (10), 1849–54.
- West, M. J., Lowe, A. D. & Karn, J. (2001) Activation of human immunodeficiency virus transcription in T cells revisited: NF-kappaB p65 stimulates transcriptional elongation. *J Virol*, **75** (18), 8524–37.
- Westendorp, M. O., Frank, R., Ochsenbauer, C., Stricker, K., Dhein, J., Walczak, H., Debatin, K. M. & Krammer, P. H. (1995) Sensitization of T cells to CD95-mediated apoptosis by HIV-1 Tat and gp120. *Nature*, **375** (6531), 497–500.
- Will, C. L. & Lührmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol*, **13** (3), 290–301.
- Will, C. L. & Lürmann, R. (1997) snRNP structure and function. In *Eukaryotic mRNA Processing*, (Kraimer, A. R., ed.), IRL Press Cold Spring Habor pp. 131–173.
- Will, C. L., Schneider, C., MacMillan, A. M., Katopodis, N. F., Neubauer, G., Wilm, M., Lührmann, R. & Query, C. C. (2001) A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J*, **20** (16), 4536–46.
- Will, C. L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M. & Lührmann, R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J*, **21** (18), 4978–88.
- Wolffe, A. P. & Meric, F. (1996) Coupling transcription to translation: a novel site for the regulation of eukaryotic gene expression. *Int J Biochem Cell Biol*, **28** (3), 247–57.
- Wolin, S. L. & Matera, A. G. (1999) The trials and travels of tRNA. *Genes Dev*, **13** (1), 1–10.
- Wollerton, M. C., Gooding, C., Robinson, F., Brown, E. C., Jackson, R. J. & Smith, C. W. (2001) Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB). *RNA*, **7** (6), 819–32.

- Wu, J. & Manley, J. L. (1989) Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev*, **3** (10), 1553–61.
- Wu, J. Y. & Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75** (6), 1061–70.
- Wu, S. & Green, M. R. (1997) Identification of a human protein that recognizes the 3' splice site during the second step of pre-mRNA splicing. *EMBO J*, **16** (14), 4421–32.
- Wu, S., Romfo, C. M., Nilsen, T. W. & Green, M. R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, **402** (6763), 832–5.
- Wyatt, J. R., Sontheimer, E. J. & Steitz, J. A. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev*, **6** (12B), 2542–53.
- Xiao, S. H. & Manley, J. L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev*, **11** (3), 334–44.
- Xiao, S. H. & Manley, J. L. (1998) Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J*, **17** (21), 6359–67.
- Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A. & Krainer, A. R. (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure*, **5** (4), 559–70.
- Yamaguchi, F., Saya, H., Bruner, J. M. & Morrison, R. S. (1994) Differential expression of two fibroblast growth factor-receptor genes is associated with malignant progression in human astrocytomas. *Proc Natl Acad Sci U S A*, **91** (2), 484–8.
- Yang, L., Embree, L. J. & Hickstein, D. D. (2000) TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins. *Mol Cell Biol*, **20** (10), 3345–54.
- Yang, L., Embree, L. J., Tsai, S. & Hickstein, D. D. (1998) Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. *J Biol Chem*, **273** (43), 27761–4.
- Yang, X., Bani, M. R., Lu, S. J., Rowan, S., Ben-David, Y. & Chabot, B. (1994) The A1 and A1B proteins of heterogeneous nuclear ribonucleoproteins modulate 5' splice site selection *in vivo*. *Proc Natl Acad Sci U S A*, **91** (15), 6924–8.
- Yeakley, J. M., Tronchère, H., Olesen, J., Dyck, J. A., Wang, H. Y. & Fu, X. D. (1999) Phosphorylation regulates *in vivo* interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *J Cell Biol*, **145** (3), 447–55.
- Yean, S. L., Wuenschell, G., Termini, J. & Lin, R. J. (2000) Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature*, **408** (6814), 881–4.
- Young, P. J., DiDonato, C. J., Hu, D., Kothary, R., Androphy, E. J. & Lorson, C. L. (2002) SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. *Hum Mol Genet*, **11** (5), 577–87.
- Yu, Y. T., Shu, M. D. & Steitz, J. A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J*, **17** (19), 5783–95.
- Yue, B. G. & Akusjärvi, G. (1999) A downstream splicing enhancer is essential for *in vitro* pre-mRNA splicing. *FEBS Lett*, **451** (1), 10–4.
- Zahler, A. M., Lane, W. S., Stolk, J. A. & Roth, M. B. (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev*, **6** (5), 837–47.
- Zahler, A. M., Neugebauer, K. M., Lane, W. S. & Roth, M. B. (1993a) Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science*, **260** (5105), 219–22.
- Zahler, A. M., Neugebauer, K. M., Stolk, J. A. & Roth, M. B. (1993b) Human SR proteins and isolation of a cDNA encoding SRp75. *Mol Cell Biol*, **13** (7), 4023–8.
- Zahler, A. M. & Roth, M. B. (1995) Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. *Proc Natl Acad Sci U S A*, **92** (7), 2642–6.

- Zamore, P. D. & Green, M. R. (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A*, **86** (23), 9243–7.
- Zamore, P. D. & Green, M. R. (1991) Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J*, **10** (1), 207–14.
- Zamore, P. D., Patton, J. G. & Green, M. R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, **355** (6361), 609–14.
- Zauli, G., Gibellini, D., Secchiero, P., Dutartre, H., Olive, D., Capitani, S. & Collette, Y. (1999) Human immunodeficiency virus type 1 Nef protein sensitizes CD4(+) T lymphoid cells to apoptosis via functional upregulation of the CD95/CD95 ligand pathway. *Blood*, **93** (3), 1000–10.
- Zeng, C. & Berget, S. M. (2000) Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. *Mol Cell Biol*, **20** (21), 8290–301.
- Zeng, C., Kim, E., Warren, S. L. & Berget, S. M. (1997) Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *EMBO J*, **16** (6), 1401–12.
- Zenklusen, D. & Stutz, F. (2001) Nuclear export of mRNA. *FEBS Lett*, **498** (2-3), 150–6.
- Zhang, H., Yang, B., Pomerantz, R., Zhang, C., Arunachalam, S. & Gao, L. (28 May 2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature advance online publication.
- Zhang, J. & Crumpacker, C. S. (2002) Human immunodeficiency virus type 1 nucleocapsid protein nuclear localization mediates early viral mRNA expression. *J Virol*, **76** (20), 10444–54.
- Zhang, L., Liu, W. & Grabowski, P. J. (1999) Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. *RNA*, **5** (1), 117–30.
- Zhang, M., Zamore, P. D., Carmo-Fonseca, M., Lamond, A. I. & Green, M. R. (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc Natl Acad Sci U S A*, **89** (18), 8769–73.
- Zhang, W. J. & Wu, J. Y. (1996) Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol Cell Biol*, **16** (10), 5400–8.
- Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419** (6903), 182–5.
- Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E. & Reed, R. (2000) The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature*, **407** (6802), 401–5.
- Zhu, J. & Krainer, A. R. (2000) Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev*, **14** (24), 3166–78.
- Zhu, J., Mayeda, A. & Krainer, A. R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell*, **8** (6), 1351–61.
- Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B. & Price, D. H. (1997) Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation *in vitro*. *Genes Dev*, **11** (20), 2622–32.
- Zinszner, H., Albalat, R. & Ron, D. (1994) A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by CHOP. *Genes Dev*, **8** (21), 2513–26.
- Zorio, D. A. & Blumenthal, T. (1999) Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature*, **402** (6763), 835–8.
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31** (13), 3406–15.
- Zuo, P. & Manley, J. L. (1994) The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. *Proc Natl Acad Sci U S A*, **91** (8), 3363–7.

RÉSUMÉ

L'épissage est une étape majeure du cycle de multiplication du virus HIV-1. L'utilisation combinée des 4 sites 5' et 8 sites 3' d'épissage de l'ARN de HIV-1 permet la production d'une quarantaine d'ARNm. Nous avons montré que les sites 3' A1 à A5 sont régulés de façon différentielle par les protéines SR ASF/SF2, SC35, 9G8 et SRp40 (régulateurs cellulaires de l'épissage). Notre étude fine du site A3 met en évidence un élément activateur en *cis*, ESEt, qui fixe les protéines ASF/SF2 et SC35. Elle montre aussi que la protéine SC35 se lie à l'élément inhibiteur ESS2, et entre en compétition avec la protéine hnRNP A1. Nous avons montré la conservation de l'activation du site A3 par les protéines SC35 et ASF/SF2 chez les virus SIVcpz et HIV-1. Enfin, sur la base de nos résultats expérimentaux, des modèles mathématiques reflétant l'un, les régulations qui s'exercent au site A3 et l'autre, la compétition entre les sites A3 à A7 ont été construits par l'équipe d'A. Bockmayr (projet MODBIO, LORIA, Nancy).

TITLE

Experimental study of the SR protein role in the splicing regulation of the HIV-1 virus RNA, responsible of the human immunodeficiency, and mathematical modelisation of these regulations

SUMMARY

Splicing is a key step for virus HIV-1 multiplication. Four donor and eight acceptor splicing sites are used in combinaison to produce about 40 mRNAs. We showed that the acceptor sites A1 to A5 are differentially regulated by the SR proteins ASF/SF2, SC35, 9G8 and SRp40 (regulator of cellular splicing). Our deep study of the regulation at site A3 shows the presence of a splicing activatory element, ESEt, that binds both the SC35 and ASF/SF2 proteins. We also showed that protein SC35 binds the ESS2 inhibitory element, and compete protein hnRNP A1 for binding to this site. We showed that regulation of site A3 by the SR proteins is conserved in SIVcpz virus and HIV-1. Based on our experimental results, mathematical models simulating splicing regulations at site A3 and competition of the A3-A7 sites, respectively, were established by the team of A. Bockmayr (project MODBIO, LORIA, Nancy).

DISCIPLINE

Biologie structurale moléculaire et cellulaire

MOTS-CLÉS

Virus HIV-1, SIVcpz, épissage, protéines SR, éléments activateur et inhibiteur en *cis*, modèle mathématique

ADRESSE DU LABORATOIRE

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