

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ
Institut de Biologie Moléculaire des Plantes du CNRS – UPR2357

THÈSE présentée par :
Clément BOUTON

soutenue le : **23 octobre 2014**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline / Spécialité : Sciences de la Vie et de la Santé / Aspects Moléculaires et
Cellulaires de la Biologie

**Étude de l'épissage alternatif de l'ARN 35S
du virus de la mosaïque du chou-fleur (CaMV)
et de l'export nucléaire des ARN viraux**

THÈSE dirigée par :
M. KELLER Mario

Professeur des Universités, Université de Strasbourg

RAPPORTEURS :
M. FILIPOWICZ Witold
M. JACQUOT Emmanuel

Professeur émérite, FMI, Bâle (Suisse)
Directeur de Recherche, INRA, Montpellier

AUTRES MEMBRES DU JURY :
M. GILMER David
M. HOHN Thomas

Professeur des Universités, Université de Strasbourg
Professeur émérite, Universität Basel, Bâle (Suisse)

Remerciements

Je tiens à remercier le Ministère de l'enseignement supérieur et de la recherche pour le financement de mes travaux de thèse.

Je tiens à remercier monsieur le docteur Emmanuel JACQUOT et messieurs les professeurs Witold FILIPOWICZ, David GILMER et Thomas HOHN, d'avoir accepté de juger ce travail.

Je tiens à remercier tous les membres de l'Institut de Biologie Moléculaire des Plantes, et en particulier les membres de mon laboratoire, pour leurs conseils et/ou leur sympathie.

Je tiens à remercier monsieur le professeur Mario KELLER d'avoir accepté de diriger mes travaux, et pour sa bienveillance à mon égard.

Enfin, je tiens à remercier mes proches collègues, mes amis et ma famille, pour leur gentillesse et leur soutien sans faille.

Avant-propos

Les travaux présentés dans le présent manuscrit sont organisés en deux chapitres et ils s'inscrivent dans une thématique d'étude assez large, consacrée aux étapes nucléaires du cycle infectieux du virus de la mosaïque du chou-fleur (CaMV).

Un aperçu de la biologie du CaMV est présenté en introduction. Nous portons, là, une attention particulière aux sujets où des avancées récentes ont été faites et/ou qui impliquent la protéine virale multifonctionnelle P6/TAV, la clé de voûte du cycle infectieux. Le texte de cette introduction générale est soumis dans le journal *Virologie*.

Le premier chapitre de cette thèse est consacré à l'étude de l'épissage alternatif de l'ARN 35S du CaMV. Il est introduit par une description des différents types introns et des mécanismes impliqués dans leur épissage, suivie d'une partie destinée à apporter une vue d'ensemble sur l'épissage alternatif et ses rôles dans le vivant. Les résultats de notre étude montrent que l'épissage alternatif de l'ARN 35S du CaMV est un phénomène fort complexe et que l'épissage est maintenu de manière permanente malgré l'inactivation de sites d'épissage. Nos résultats soulèvent plusieurs interrogations sur la signification biologique de l'épissage pour le CaMV. En annexe de ce chapitre, est inclus un article en cours de réécriture, dont une première version avait été soumise dans la revue *Journal of Virology*. Cet article reprend seulement une partie des résultats présentés dans ce premier chapitre.

Le deuxième chapitre porte sur l'étude de l'export nucléaire des ARN 35S épissés ou non épissé du CaMV. Alors que la thématique de l'épissage alternatif avait été initiée avant mon arrivée au laboratoire, celle de l'export nucléaire était vierge de toute donnée préliminaire lorsque j'ai débuté mon doctorat. À l'instar du premier chapitre, le sujet est introduit par un aperçu global des mécanismes d'export nucléaire des ARNm chez la levure et les mammifères de celui de certains ARN viraux, car les connaissances sur ce processus chez les plantes sont encore extrêmement sommaires. La partie consacrée aux résultats rapporte principalement nos recherches sur la mise au point de protocoles et de vecteurs viraux destinés à l'étude de la voie d'export des ARN du CaMV et des acteurs en *cis* et en *trans* impliqués dans leur export.

Enfin, une conclusion générale et les perspectives d'étude qui découlent de nos travaux sont discutées dans les dernières pages de ce manuscrit.

Sommaire

Table des matières

Remerciements	I
Avant-propos	III
Sommaire	V
Abréviations	XIII
Introduction générale – Le virus de la mosaïque du chou-fleur (CaMV), encore et toujours	1
Structure de la capside et génome du CaMV	4
Fonctions des protéines du CaMV	5
Le cycle infectieux du CaMV	9
Transcription du génome	9
Épissage de l'ARN 35S	11
Traduction des ARN du CaMV	11
<i>Le saut du ribosome, ou ribosomal shunt</i>	12
<i>Réinitiation de la traduction de l'ARN 35S</i>	12
Réplication du génome du CaMV par rétrotranscription	16
Mouvement de cellule-à-cellule du CaMV	16
Pathogenèse du CaMV	18
Spécificité d'hôtes	18
Symptomatologie	18
Résistance, défense et contre-défense	19
Résistance des plantes au CaMV	20
<i>RNA silencing</i> et la contre-défense du CaMV	20
Défense non spécifique des plantes vis-à-vis du CaMV	21
Conclusion	22
Références	24
Chapitre 1 – Étude de l'épissage alternatif de l'ARN 35S du CaMV	29
INTRODUCTION	31
1) Aperçu général sur l'épissage	31
a) Introduction	31
b) Les différents types d'introns et leur mécanisme d'épissage	31
<i>i. Introns de groupe I</i>	31
<i>ii. Introns de groupe II</i>	32
<i>iii. Introns de pré-ARNm nucléaires ou introns splicéosomaux</i>	33
<i>iv. Introns d'ARNt nucléaires et introns d'archées</i>	34
2) Épissage des introns splicéosomaux de type U2	35
a) Introns splicéosomaux de type U2	35

i.	<i>Caractéristiques générales</i>	35
ii.	<i>Particularités des introns de type U2 de plantes</i>	36
b)	Splicéosome de type U2, dit majeur	36
i.	<i>Composition</i>	36
ii.	<i>Biogénèse des snRNP</i>	36
c)	Assemblage du splicéosome de type U2 et réaction d'épissage	38
i.	<i>Assemblage du splicéosome de type U2</i>	38
ii.	<i>Changements de composition et conformation</i>	40
3)	Épissage des introns splicéosomaux de type U12	40
a)	Caractéristique des introns de type U12.....	40
b)	Réaction d'épissage.....	41
4)	Épissage alternatif	41
a)	Introduction théorique	41
b)	Généralités sur le rôle biologique de l'épissage alternatif.....	41
c)	Régulation de l'épissage alternatif à l'échelle moléculaire.....	42
i.	<i>Facteurs protéiques régulant l'épissage</i>	42
ii.	<i>Régulation transcriptionnelle</i>	43
iii.	<i>Régulation épigénétique</i>	44
d)	Pathologies impliquant l'épissage alternatif	45
e)	Aperçu de rôles de l'épissage alternatif	45
i.	<i>Implication dans le développement chez les animaux</i>	45
ii.	<i>Implication dans l'adaptation à l'environnement chez les plantes</i>	46
iii.	<i>Rôles chez les virus</i>	47
5)	Ouverture sur le sujet d'étude	49
a)	Données disponibles dans la littérature scientifique.....	49
b)	Questions sur le rôle biologique de l'épissage alternatif chez le CaMV	50
	RÉSULTATS	53
1)	Une analyse in silico révèle que des sites d'épissage de l'ARN 35S sont conservés parmi les souches séquencées du CaMV et certains membres du genre des <i>Caulimovirus</i>	53
2)	L'ARN 35S de la souche Cabb B-JI est épissé alternativement	54
3)	Les protéines de fusion P1P2 sont indétectables dans les plantes infectées	56
4)	Un site donneur cryptique est utilisé lorsque les sites D1, D2 et D4 sont mutés	60
5)	L'ARN 35S des souches Cabb-S et Cabb B-JI possède un profil d'épissage complexe	62
6)	Des sites accepteurs cryptiques peuvent être activés pour maintenir l'épissage de l'ARN 35S	65
	DISCUSSION	67
1)	Conservation des sites d'épissage, chez les souches du CaMV et les <i>Caulimovirus</i> ...	67
2)	Découverte d'un profil d'épissage complexe, chez le CaMV	67
3)	Raison d'être de l'épissage alternatif de l'ARN 35S du CaMV	68

4) Autres fonctions possibles de l'épissage, lors du cycle infectieux	71
MATÉRIELS ET MÉTHODES	73
1) Alignement de séquences	73
2) Plantes hôtes, souches virales et inoculation	73
3) Plasmides	73
a) Mutants d'épissage et mutant P1P2stop	73
b) Vecteurs d'expression de versions étiquetées de la protéine GFP et des protéines virales P1, P2, P3, P6/TAV et P1P2(D2)	73
4) Extraction d'ARN totaux et northern-blot	74
5) Synthèse d'ADNc par RT-PCR	74
6) Analyse des ARN 35S épissés par RT-PCR	74
7) Analyse des protéines par western-blot	75
8) Préparation de fractions enrichies en paroi cellulaire	75
9) Expression transitoire par agroinfiltration	75
10) Cinétique d'inhibition de la voie de l'ubiquitine/protéasome	76
11) Expression et marquage au ³² P de protéines virales recombinantes	76
12) Analyse d'interactions protéiques <i>in vitro</i> , par far western-blot	76
ANNEXE – publication « <i>Cauliflower mosaic virus</i> transcriptome reveals a complex alternative splicing pattern necessary for infectivity »	77
Chapitre 2 – Étude de l'export nucléaire des ARN viraux	101
INTRODUCTION	103
1) Les pores nucléaires	103
a) Structure et composition	103
b) Mécanisme général du transport à travers les pores nucléaires	104
c) Assemblage des pores nucléaires	105
i. Assemblage <i>de novo</i> au cours de l'interphase	105
ii. Assemblage <i>post-mitotique</i>	105
2) Export nucléaire des ARN autres que les ARNm, et des protéines	106
a) Export nucléaire des ARNt	106
b) Export nucléaire des miARN	108
c) Export nucléaire des snARN	108
d) Export nucléaire des ARNr	108
e) Export nucléaire des protéines	108
3) Vue générale sur l'export nucléaire des ARNm	109
a) Introduction théorique	109
b) Facteur de transport des ARNm	109
i. Les hétérodimères TAP-p15 et Mex67-Mtr2	109
ii. L'exportine CRM1	109
c) Couplage entre la production des ARNm et leur export nucléaire	110

i.	<i>Le complexe TREX</i>	110
ii.	<i>Épissage et export nucléaire</i>	111
d)	Séquences nucléotidiques impliquées dans l'export nucléaire	111
i.	<i>Régions codantes des séquences signal et des séquences ciblant les mitochondries</i>	111
ii.	<i>Élément constitutif de transport (CTE) des Rétrovirus</i>	111
iii.	<i>Cas des ARNm sans introns</i>	111
e)	Mouvement des mRNP dans le noyau et au sein des pores nucléaires	112
i.	<i>Mouvement dans le noyau</i>	112
ii.	<i>Traversée du pore nucléaire</i>	112
4)	Export nucléaire des ARNm chez les plantes	112
5)	Export nucléaire des ARNm viraux	113
a)	Utilisation de la voie TAP-p15 par les virus	113
i.	<i>Cas de l'herpès simplex virus de type 1 (HSV-1) et de l'herpèsvirus associé au sarcome de Kaposi (KSHV)</i>	113
ii.	<i>Cas des rétrovirus simples</i>	114
b)	Utilisation de la voie CRM1 par les virus	114
i.	<i>Cas du virus du sarcome de Rous (RSV)</i>	114
ii.	<i>Cas des rétrovirus complexes</i>	114
iii.	<i>Cas des foamy virus</i>	115
iv.	<i>Cas du virus de l'influenza A (IAV)</i>	116
6)	Ouverture sur le sujet d'étude	116
	RÉSULTATS	119
1)	Étude de l'interaction entre P6/TAV et la protéine AtXPO1, par GST pull-down	119
2)	Étude de l'implication de P6/TAV dans le processus d'export nucléaire de l'ARN 35S non-épissé, par l'utilisation d'un rapporteur fluorescent	120
a)	Développement des plasmides pEXPORT	120
b)	Transfection et résultats	122
3)	Étude de l'export nucléaire des différentes formes d'ARN 35S par détermination de la quantité d'ARN non-épissé présente au sein de la population globale d'ARN 35S, dans le noyau et le cytoplasme	123
a)	Développement du plasmide pBS-BJI dbA	123
b)	Méthodes de transfection	124
i.	<i>Transfection par biolistique</i>	124
ii.	<i>Transfection de protoplastes au polyéthylène glycol</i>	124
c)	Préparation d'ARN nucléaires et cytoplasmiques	124
d)	Détection des différentes formes de l'ARN 35S	125
i.	<i>Détection par northern-blot</i>	125
ii.	<i>Détection par RT-PCR quantitative (qRT-PCR)</i>	126
	DISCUSSION	129
1)	Les différents ARN viraux produits lors de l'infection	129

2) Voie empruntée pour l'export nucléaire de l'ARN 35S complet.....	129
3) Régulation de l'export nucléaire de l'ARN 35S par une ou plusieurs protéines virales	130
4) Éléments structuraux de l'ARN 35S qui régulent l'export en <i>cis</i>	131
MATÉRIELS ET MÉTHODES	133
1) Plasmides.....	133
a) Vecteurs d'expression d'AtXPO1a et de ses formes tronquées.....	133
b) Vecteurs d'expression de la protéine P6/TAV et de son domaine A.....	133
c) Plasmides du type pEXPORT	133
d) Plasmide pBS-BJI dbA.....	134
2) Analyse d'interactions protéiques <i>in vitro</i> , par GST pull-down	134
a) Transcription et traduction <i>in vitro</i> des protéines AtXPO1a	134
b) Expression des protéines GST-P6/TAV et GST-A.....	134
c) Purification des protéines en « <i>batch</i> ».....	134
d) GST pull-down	135
3) Cellules en culture liquide et plantes.....	135
a) Cellules de tabac BY-2.....	135
b) Cellules d' <i>Arabidopsis</i>	135
c) Plantes	135
4) Transfection par biolistique.....	135
5) Transfection de protoplastes au polyéthylène glycol	136
a) Protoplastes de mésophylles de navet et d' <i>Arabidopsis</i>	136
b) Protoplastes de cellules d' <i>Arabidopsis</i> en culture liquide	136
6) Fractionnement cellulaire	136
7) Analyse des ARN par northern-blot	137
a) ARN de haut poids moléculaires	137
b) ARN de petit poids moléculaires	137
8) Quantification des ARN par qRT-PCR.....	137
Conclusion générale et perspectives	139
1) L'épissage de l'ARN 35S du CaMV dans les faits.....	141
2) Rôles et la régulation de l'épissage chez le CaMV, de nombreuses interrogations subsistent.....	141
3) L'export nucléaire des ARN viraux, un sujet encore aux prémices.....	142
Bibliographie.....	145

Abréviations

ARNg : ARN génomique
 ARS2 : *arsenite resistance 2*
 BCIP : phosphate de 5-Bromo-4-chloro-3-indolyle
 BY-2 : *bright yellow - 2*
 CaMV : *cauliflower mosaic virus* – virus de la mosaïque du chou-fleur
 CAR : *cytoplasmic accumulation regions*
 CERV : *carnation etched ring virus* – virus de la gravure de l'œillet
 CRM1 : *chromosome region maintenance 1*
 CTD : domaine carboxy-terminal
 CTE : *constitutive transport element*
 DMV : *dahlia mosaic virus* – virus de la mosaïque du dahlia
 dpi : *day post inoculation* – jours après infection
 Dscam1 : *Down syndrome cell adhesion molecule 1*
 FMV : *figwort mosaic virus* – virus de la mosaïque de la scrofulaire
 HBSP : *HBV splice-generated protein*
 HBV : *hepatitis B virus* – virus de l'hépatite B
 HIV-1 : *human immunodeficiency virus type 1* – virus de l'immunodéficience humaine de type 1
 HMK : *heart muscle kinase*
 hnRNP : ribonucléoprotéines nucléaires hétérogènes
 HRLV : *horseradish latent virus* – virus latent du raifort
 HSV-1 : *herpes simplex virus type 1* – herpès simplex virus de type 1
 HTLV-1 : *human T-lymphotropic virus* – virus T-lymphotropique humain de type 1
 IAV : *influenza A virus* – virus de l'influenza A
 IEP : *intron encoded protein*
 IPTG : isopropyl- β -D-thiogalactopyranoside
 KSHV : *Kaposi's sarcoma-associated herpesvirus* – l'herpèsvirus associé au sarcome de Kaposi
 LLDAV : *lamium leaf distortion virus* – virus de la déformation des feuilles du lamier
 MES : 2-morpholino éthanesulfonique monohydraté
 miARN : microARN
 MiMV : *mirabilis mosaic virus* – virus de la mosaïque de la belle-de-nuit
 MLV : *murine leukemia virus* – virus de la leucémie murine
 MPMV : *Mason-Pfizer monkey virus* – virus Mason-Pfizer du singe
 MSV : *maize streak virus* – virus de la striure du maïs
 NBT : chlorure de bleu de nitrotétrazolium
 NES : *nuclear export signal*
 NLS : *nuclear localization signal*
 ORF : *open reading frame* – cadre ouvert de lecture
 PFV : *prototype foamy virus*

PHAX : *phosphorylated adapter RNA export*

pré-ARNm : ARN pré-messager

PVDF : polyfluorure de vinylidène

qRT-PCR : *quantitative reverse transcription polymerase chain reaction*

RanGAP : *Ran GTPase activating protein*

RCC1 : *regulator of chromosome condensation 1*

RISC : *RNA-induced silencing complex*

RNase : ribonucléase

RSV : *Rous sarcoma virus* – virus du sarcome de Rous

RTBV : *rice tungro bacilliform virus* - virus bacilliforme du tungro du riz

RT-PCR : *reverse transcription polymerase chain reaction*

SAGA : *Spt-Ada-Gcn5-acetyltransferase*

scaARN : petit ARN spécifique des corps de Cajal

siARN : short interfering RNA – petit ARN interférant

SMN : *survival motor neuron*

snARN : petit ARN nucléaire

snoARN : petit ARN nucléolaire

snRNP : petite particule ribonucléoprotéique nucléaire

SVBV : *strawberry vein banding virus* – virus de la bigarrure des nervures du fraisier

TAP : *Tip-associated protein*

TREX : *transcription export*

U2AF : *U2 auxiliary factor*

vRNP : ribonucléoprotéine virale

Introduction générale

Le virus de la mosaïque du chou-fleur (CaMV), encore et toujours

Texte soumis pour publication dans le journal *Virologie*

Le virus de la mosaïque du chou-fleur (CaMV), encore et toujours

Bouton Clément, Geldreich Angèle, Ryabova Lyuba, Keller Mario et Dimitrova Maria

Institut de Biologie Moléculaire des Plantes du CNRS, Université de Strasbourg, Strasbourg, France

RÉSUMÉ

Son statut de pararétrovirus et les mécanismes traductionnels non-canoniques qu'il emploie pour exprimer ses protéines à partir d'un ARN polycistronique font que le virus de la mosaïque du chou-fleur – ou CaMV pour « *Cauliflower mosaic virus* » – est un modèle d'étude original qui fait l'objet de recherches permanentes. Ces dernières années, des découvertes significatives ont amélioré notre compréhension des mécanismes impliqués dans la transmission du CaMV par les pucerons vecteurs, le mouvement intercellulaire, ainsi que dans l'expression de ses protéines et la contre-défense du virus face aux systèmes de défense des plantes hôtes. Ces deux derniers aspects de la pathogenèse sont sous la dépendance de la protéine virale P6/TAV qui, parce qu'elle est également impliquée dans d'autres étapes de la réplication du virus, est considérée comme la clé de voûte du cycle infectieux du CaMV. Nos connaissances actuelles sur le cycle de multiplication et la pathogenèse du CaMV sont présentées dans cette revue, en insistant sur les étapes où la protéine P6/TAV joue un rôle majeur.

Mots clés : CaMV, P6/TAV, réinitiation traductionnelle, modèle d'étude.

ABSTRACT

As a pararetrovirus and because of its non-canonical translation of a polycistronic pregenomic 35S RNA, *Cauliflower mosaic virus* (CaMV) is an original model system that has been extensively studied. Recent advances have improved our understanding of CaMV aphid transmission, cell-to-cell movement, protein expression and virus counter-defense strategy against host plant defense. Since P6/TAV is involved in many aspects of viral pathogenesis as well as in some replication steps, it is considered as the key player of CaMV infectious cycle. This paper reviews our current knowledge on CaMV multiplication and pathogenesis, with special emphasis on steps in which P6/TAV has a major role.

Keyword: CaMV, P6/TAV, translation reinitiation, model system.

Le virus de la mosaïque du chou-fleur ou CaMV pour « *Cauliflower mosaic virus* », est le membre type du genre des *Caulimovirus* de la famille des *Caulimoviridae*, seule famille de phytovirus ayant un génome à ADN double brin. L'infection de Brassicacées maraîchères (choux, navet, colza) par le CaMV ne constitue pas un préjudice économique suffisamment important qui justifierait en soi, des études approfondies pour élaborer des stratégies de lutte antivirale. Et pourtant, le CaMV est classé en virologie végétale parmi les dix phytovirus les plus importants car il constitue un modèle d'étude original [1]. L'intérêt que suscite le CaMV est en partie lié à des raisons historiques : il fut le premier virus à ADN de plantes identifié et dont le génome a été entièrement séquencé, et il fut le premier phytovirus pour lequel il a été démontré que l'ADN est répliqué par rétrotranscription – on parle de pararétrovirus. Il s'est avéré, suite à cette découverte, que tous les virus à ADN double brin de plantes sont des pararétrovirus, à l'instar des virus de l'hépatite B de l'homme et des animaux. Le CaMV se singularise également des autres phytovirus par l'utilisation d'une part, de mécanismes de traduction très particuliers pour exprimer ses protéines – le saut du ribosome et la réinitiation de la traduction – et d'autre part, de stratégies inédites pour contrecarrer le système de défense antiviral de la plante. Les travaux pionniers sur les stratégies de traduction non-canoniques du CaMV connais-

sent, depuis déjà quelques temps, un regain d'intérêt car elles sont également mises en œuvre par d'autres virus, y compris des virus humains – virus de l'hépatite B, *Spumavirus*, adénovirus, etc –, et même par les cellules, pour exprimer spécifiquement certains ARN messagers.

L'objectif de cette revue est de présenter les connaissances actuelles sur le CaMV et ses interactions avec les plantes hôtes, en portant une attention toute particulière à la protéine virale multifonctionnelle P6/TAV, qui constitue la clé de voûte du cycle infectieux et de la pathogenèse du CaMV.

Structure de la capside et génome du CaMV

Le CaMV est un virus non-enveloppé à capside icosaédrique (T=7), d'environ 50 nm de diamètre. La capside est formée de 420 sous-unités protéiques d'un poids moléculaire de 42, 37, 35 et 32 kDa, les protéines de 42 et 37 kDa étant généralement les plus abondantes [2]. Les protéines de capside résultent de clivages en N- et C-terminal d'un précurseur unique, la protéine P4. La capside comporte aussi une protéine mineure, P3, qui interagit sous forme trimérique avec les protéines structurales, grâce à sa région C-terminale enfoncée dans la capside. L'extrémité N-terminale des molécules de P3 émergent de la capside et interagissent entre elles, *via* des interactions de type *coiled coil*, pour former un réseau autour de la capside [3].

Le génome du CaMV est une molécule d'ADN double-brin circulaire d'environ 8 kpb qui est présente dans la particule virale sous une forme relâchée du fait d'interruptions de séquence sur ses deux brins. Ces interruptions sont générées au cours de la rétrotranscription et ont une structure triple brin, l'extrémité 5' du brin chevauchant commençant par une séquence d'ARN. Le nombre et la position de ces interruptions, appelées Δ , varient selon les souches de CaMV. L'ADN du CaMV possède sept ORF majeurs qui sont tous localisés sur le même brin d'ADN – brin (-) –, et deux régions intergéniques, d'environ 700 pb et 150 pb, qui contiennent les promoteurs ainsi que les séquences régulatrices impliquées dans la transcription du génome en deux ARN majeurs, les ARN 35S et 19S [4]. Les ORF sont soit chevauchants (ORF III, IV et V), soit séparés les uns des autres par quelques nucléotides ; seul l'ORF VI est isolé car il est placé entre les deux régions intergéniques (Figure 1).

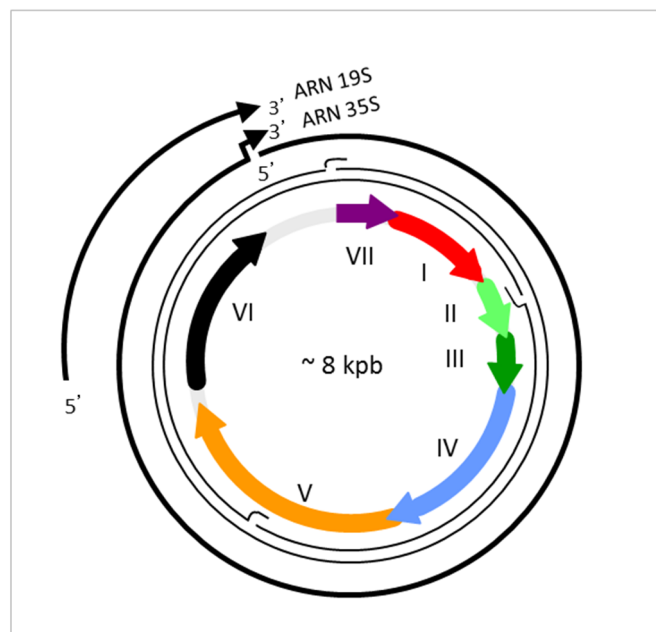


Figure 1 : Organisation génétique de l'ADN du CaMV. Le génome à ADN double brin avec les trois interruptions de séquences de la souche Cabb-B JI est représenté par les deux cercles fins et les ORF (I à VII), par des flèches colorées épaisses. La grande et la petite région intergéniques sont en gris. Les deux transcrits majeurs coiffés et polyadénylés, les ARN 35S et 19S, sont à l'extérieur du cercle.

Fonctions des protéines du CaMV

Une ou plusieurs fonctions sont associées aux six protéines P1 à P6, codées respectivement par les ORF I à VI du CaMV. Bien que l'ORF VII soit pleinement fonctionnel dans des systèmes d'expression hétérologues, son produit d'expression n'a jamais été détecté dans les plantes infectées. Toutes les protéines virales sont exprimées à partir de l'ARN 35S polycistronique ; seule la protéine P6/TAV est également exprimée à partir de son propre ARN, l'ARN 19S monocistronique.

P1 (40 kDa) est la protéine de mouvement (MP) ; elle s'assemble en tubules [5] au niveau des plasmodesmes qui forment un réseau symplastique impliqué dans la communication intercellulaire. Les tubules permettent aux virions de passer de la cellule infectée vers les cellules adjacentes et, par conséquent, d'atteindre les vaisseaux conducteurs du phloème pour se propager dans la plante entière. Le domaine central de P1 est impliqué dans son adressage à la périphérie de la cellule [6] tandis que toute la séquence de P1, sauf l'extrémité C-terminale, est requise pour la formation des tubules [7]. Dans les tubules assemblés, les extrémités N-terminale et C-terminale sont exposées respectivement à la surface et dans le lumen. P1 possède, dans sa région centrale, un domaine de fixation non spécifique à ARN. S'il a été suggéré que le mouvement du virus puisse aussi s'opérer sous la forme de complexes ribonucléoprotéiques constitués de P1 et des ARN viraux, cette hypothèse a été contredite par l'observation que la protéine P1 ne peut assurer le transport d'ARN viraux hétérologues que lorsqu'ils sont encapsidés [8]. P1 interagit avec différentes protéines virales – P3 et P6/TAV – et cellulaires – MIP17, PDLP-1 et PME – ; il a été démontré que certaines de ces interactions sont impliquées dans le mouvement de cellule à cellule du CaMV [2, 9].

P2 (18 kDa) est le facteur assistant de la transmission (FAT) du virus par les pucerons [10]. Elle interagit, grâce à sa région N-terminale, avec un récepteur protéique non-glycosylé inséré dans la chitine de l'extrémité du stylet du puceron – l'acrostyle – et, avec sa partie C-terminale, avec la protéine P3 associée à la capsid. P2 est une protéine abondante qui s'accumule dans les cellules infectées pour former un corps d'inclusion unique, le viroplasma clair – appelé également corps de transmission –, qui renferme, en plus de P2, la protéine P3 et quelques particules de CaMV c'est-à-dire, tous les composants requis pour la transmission du CaMV par le puceron [11]. Après sa synthèse, la protéine P2 est transportée, avec P3, le long des microtubules pour s'agréger en un viroplasma clair [12]. P2 ne semble avoir aucune autre fonction dans le cycle viral et, par conséquent, l'ORF II peut être éliminé du génome viral sans affecter la réplication du CaMV. D'ailleurs, des souches naturelles de CaMV dont l'ORF II est partiellement délété (souche CM1841) ou présente une mutation ponctuelle qui déstabilise P2 (souche Campbell), se répliquent avec une très bonne efficacité lorsqu'elles sont inoculées mécaniquement aux plantes hôtes mais elles nécessitent, évidemment, l'assistance d'un CaMV compétent pour assurer leur transmission de plante à plante.

P3 (15 kDa) est une protéine basique qui est, comme dit précédemment, indispensable à la formation du complexe viral transmissible par le puceron car elle fait le pont entre le FAT et la capsid virale [13]. P3 est également indispensable pour le mouvement intercellulaire des particules de CaMV à travers les tubules [14]. P3 forme, grâce à son extrémité N-terminale structurée en hélices α , des trimères et des tétramères. P3 est associée aux virions sous forme trimérique tandis qu'elle est retrouvée sous forme tétramérique dans le viroplasma clair [3]. Un polypeptide de 11 kDa résultant d'un clivage de P3 à son extrémité C-terminale est fréquemment détecté dans les plantes infectées mais aucune fonction n'a pu être associée à ce clivage. Les régions N- et C-terminales sont indispensables pour le cycle infectieux du CaMV alors qu'une séquence centrale (résidus 60 à 80, sur un total de 129 acides aminés) peut être éliminée sans aucune conséquence pour le virus [2]. En effet, la région N-terminale interagit avec les protéines P1 et P2, alors que la région C-terminale s'associe à la capsid. P3 se lie aussi avec P6/TAV [15] mais son domaine d'interaction n'a pas encore été cartographié à ce jour. Enfin, P3 est capable d'interagir avec l'ADN et l'ARN de manière séquence non-spécifique via une région riche en prolines et en lysines à son extrémité C-terminale [16].

P4 (57 kDa) est le précurseur des protéines de la capsid, l'équivalent de la protéine Gag des rétrovirus. P4 est une protéine très polaire avec des extrémités N- et C-terminales très acides. Après

assemblage de la capside, ces extrémités sont clivées en plusieurs étapes, par la protéinase à acide aspartique codée par l'ORF V et sans doute, par des protéases cellulaires, générant alors des polypeptides de 42, 39, 37 et 35 kDa. Les quantités relatives de P4 et de ses produits de clivage sont variables d'une préparation de virions à l'autre. Ces protéines structurales subissent des modifications post-traductionnelles, des glycosylations et probablement des phosphorylations, car une protéine kinase d'origine encore inconnue est associée aux virions. Elles possèdent toutes, dans leur région C-terminale, un domaine basique et un motif en doigt à zinc (CX₂CX₄HX₄C), caractéristique qui est retrouvée dans la protéine de nucléocapside de tous les rétrovirus, à l'exception des *Spumavirus*. Le motif en doigt à zinc interagit avec une séquence riche en purines, au niveau de la région 5' non traduite – dite région *leader* – de l'ARN 35S pré-génomique, à condition que l'extrémité C-terminale acide du précurseur P4 ait été éliminée au préalable. Le motif en doigt à zinc revêt une importance particulière pour le cycle viral car des mutations qui abolissent son interaction avec l'ARN 35S sont létales pour le CaMV. Le précurseur P4 renferme trois séquences d'instabilité : une à chacune de ses deux extrémités et une séquence de dégradation PEST, qui subsiste à l'extrémité N-terminale de la protéine 42 kDa [17]. P4 et les formes clivées possèdent, à proximité de leur l'extrémité N-terminale, un signal de localisation nucléaire (NLS) qui n'est exposé qu'après maturation des virions. Au début du cycle infectieux, ce signal permet l'acheminement des particules virales matures vers le noyau, vraisemblablement grâce à l'importine α , où elles s'amarrent aux pores nucléaires [18]. La capside n'est pas trouvée à l'intérieur du noyau ce qui suggère que l'ADN viral est libéré au niveau des pores. L'extrémité N-terminale acide du précurseur P4 empêche l'adressage de la protéine vers le noyau car elle masque le NLS. Il s'agit certainement d'un mécanisme de contrôle astucieux pour s'assurer que P4 ne soit pas dirigée directement vers le noyau avant qu'elle ne s'auto-assemble pour former la capside. De cette manière, seuls les virions matures, après les clivages terminaux de P4, seront susceptibles d'atteindre l'enveloppe nucléaire. Il est probable qu'en plus de se diriger vers le noyau d'une nouvelle cellule hôte suite à la transmission, les virions puissent réinfecter la cellule dans laquelle ils ont été formés.

P5 (78 kDa) est une protéine multifonctionnelle essentielle pour la réplication du génome viral ; c'est l'équivalent de la protéine Pol des rétrovirus. En effet, P5 porte, dans sa région C-terminale, une activité ADN polymérase ARN/ADN dépendante, la transcriptase inverse, et une activité RNase H putative, cette dernière ayant été supposée sur la base d'homologies de séquences avec celles des rétrovirus. Cette protéine possède également, dans sa région N-terminale, une activité de type protéinase à acide aspartique (18 kDa) qui est libérée par une auto-maturation générant simultanément la transcriptase inverse [19]. Comme mentionné ci-dessus, cette protéase intervient également dans la maturation de la protéine de capside. L'association de ces trois domaines enzymatiques est une caractéristique structurale et fonctionnelle des pararétrovirus et des rétrovirus. Toutefois, contrairement aux rétrovirus, aucune fonction de type endonucléase/intégrase n'est associée à la protéine P5, l'intégration dans le génome nucléaire de la plante hôte n'étant pas une étape obligatoire pour le déroulement du cycle infectieux des pararétrovirus, à l'inverse du cycle des rétrovirus. S'il y a intégration, celle-ci est purement accidentelle et s'opère par des recombinaisons illégitimes, comme cela a été démontré pour différents pararétrovirus endogènes (EPRV) de plantes [20].

P6/TAV (62 kDa) est exprimée abondamment à partir de son propre messenger, l'ARN 19S, ainsi qu'à partir de l'ARN 35S. C'est une protéine unique aux *Caulimovirus* et au genre voisin des *Soymovirus*, qui ne présente aucune homologie significative avec des protéines virales et cellulaires [21]. Elle est relativement bien conservée chez les différentes souches de CaMV puisque l'identité de séquence va de 92% à 97%, l'extrémité N-terminale étant la région la plus variable. À ce jour, aucune modification post-traductionnelle n'a été décelée au niveau de cette protéine malgré qu'elle renferme des sites consensus pour la phosphorylation et la sumoylation. Bien que principalement localisée dans le cytoplasme, P6/TAV est capable de faire la navette entre le noyau et le cytoplasme puisqu'elle possède un signal d'export nucléaire (NES) riche en leucines et, deux NLS, l'un basique et l'autre non conventionnel qui est similaire à un des NLS de la nucléoprotéine NP du virus influenza de type A. Cette capacité à transiter activement entre le noyau et le cytoplasme est unique vis-à-vis des cinq autres protéines du CaMV, et elle dépend de l'importine α et de l'exportine CRM1 [22, 23].

P6/TAV est une protéine multifonctionnelle qui joue un rôle central dans la pathogenèse du CaMV (Figure 2A) ; nous reviendrons sur toutes ses fonctions de manière plus détaillée, lorsque les différentes étapes du cycle infectieux seront abordées. P6/TAV est le constituant majeur des viroplasmiques denses, des corps d'inclusion dépourvus de membranes et parsemés de zones claires vacuolaires, en fonction de leur stade de développement, qui se distinguent facilement des viroplasmiques clairs déjà évoqués lorsqu'on observe les cellules infectées au microscope électronique. Les viroplasmiques denses sont considérés comme des usines virales car ils sont le site de la réplication de l'ADN viral, de la morphogénèse des virions et de l'accumulation des particules virales nouvellement formées [24]. Leur taille, au stade final de leur formation, est extrêmement variable (de 0,3 µm à 10 µm) en fonction des souches de CaMV et de la plante hôte [25]. Les viroplasmiques denses résultent de l'interaction entre les molécules P6/TAV [22, 26] et ce, indépendamment des autres protéines du CaMV puisque de telles structures sont observées lors d'expérience d'expression transitoire de P6/TAV dans des cellules de tabac, ainsi que dans des plantes transgéniques d'*Arabidopsis thaliana* exprimant constitutivement la protéine [27]. Les viroplasmiques denses se forment par étapes : des petits agrégats sont d'abord produits, puis ils s'assemblent pour générer de grands viroplasmiques denses, généralement dans la région périnucléaire. Leur formation implique l'extrémité N-terminale de la protéine ainsi que deux séquences situées, l'une dans sa partie centrale et l'autre, à son extrémité C-terminale [Geldreich *et al.*, résultats non publiés]. Les petits agrégats de P6/TAV se déplacent le long des microfilaments d'actine mais ils sont incapables de se mouvoir sur les microtubules, contrairement aux agrégats de certains virus à ADN nucléo-cytoplasmiques infectant les animaux, comme les *Asfarviridae* ou les *Poxviridae* [28]. P6/TAV intervient, par ailleurs, dans le déterminisme du spectre d'hôte [29, 30] et dans l'expression des symptômes de la virose [31, 32], son activité de suppresseur du *RNA silencing* [23, 33, 34] permettant au CaMV d'exprimer toute sa virulence chez les plantes hôtes. Toutefois, comme la divinité romaine Janus, la protéine se comporte aussi comme un produit d'avirulence induisant, chez certaines plantes, la résistance à l'infection virale [35]. P6/TAV est indispensable pour l'expression de tous les ORF portés par l'ARN 35S pré-génomique car elle maintient les ribosomes sous une forme compétente afin qu'ils puissent traduire les différents ORF du transcrit, par l'intermédiaire d'un processus appelé réinitiation de la traduction [36]. Cette fonction d'activateur traductionnel et son aptitude à former des viroplasmiques denses sont, d'ailleurs, à l'origine de la dénomination TAV – pour « *transactivator/viroplasm* ». Enfin, l'interaction de P6/TAV avec la protéine de capsid suggère qu'elle pourrait agir comme une chaperonne lors de la morphogénèse des particules virales [37].

Des expériences de mutagenèse dirigée, associées dans la plupart des cas à des tests d'activités *in vitro* et *in vivo*, ont permis de cartographier les domaines impliqués dans les différentes fonctions de P6/TAV (Figure 2B). La région N-terminale renferme les déterminants responsables de la spécificité d'hôte du CaMV et de la symptomatologie et, paradoxalement, les séquences d'avirulence reconnues par les plantes résistantes au CaMV [38]. Elle est également impliquée dans la contre-défense : elle inhibe les voies de l'acide salicylique et du jasmonate, ainsi que le *RNA silencing* [39]. La région N-terminale renferme deux séquences, I1 et I2, totalement conservées chez toutes les protéines P6/TAV, quelle que soit la souche de CaMV. La séquence I1 est un décapeptide dont les trois premiers résidus, EKI, jouent un rôle prépondérant dans la formation des viroplasmiques denses alors que le reste de cette séquence invariante constitue le NES [22]. Une région centrale de P6/TAV comprenant les domaines MAV – pour « *miniTAV* », acides aminés 116 à 242 – et MBD – pour « *multiple binding domain* », acides aminés 242 à 310 – est impliquée dans la réinitiation de la traduction des ARN polycistroniques. Cette région est aussi le siège de multiples interactions avec des composants de la machinerie traductionnelle de la plante hôte – L13, L18, L24, eIF3, RISP et TOR –, [40] ainsi que d'autres facteurs cellulaires comme CHUP1 [41], une protéine impliquée dans le transport des chloroplastes sur les microfilaments d'actine. Le domaine MAV contient une séquence similaire à celle trouvée chez les RNases H, capable de se lier aux duplex ARN-ARN et ARN-ADN [42], et en C-terminal, le NLS non conventionnel [23]. Deux domaines de liaison à l'ARN simple brin, entourant le NLS basique, ont été identifiés dans la deuxième moitié de la protéine, et un doigt à zinc a également été mis en évidence, à son extrémité C-terminale.

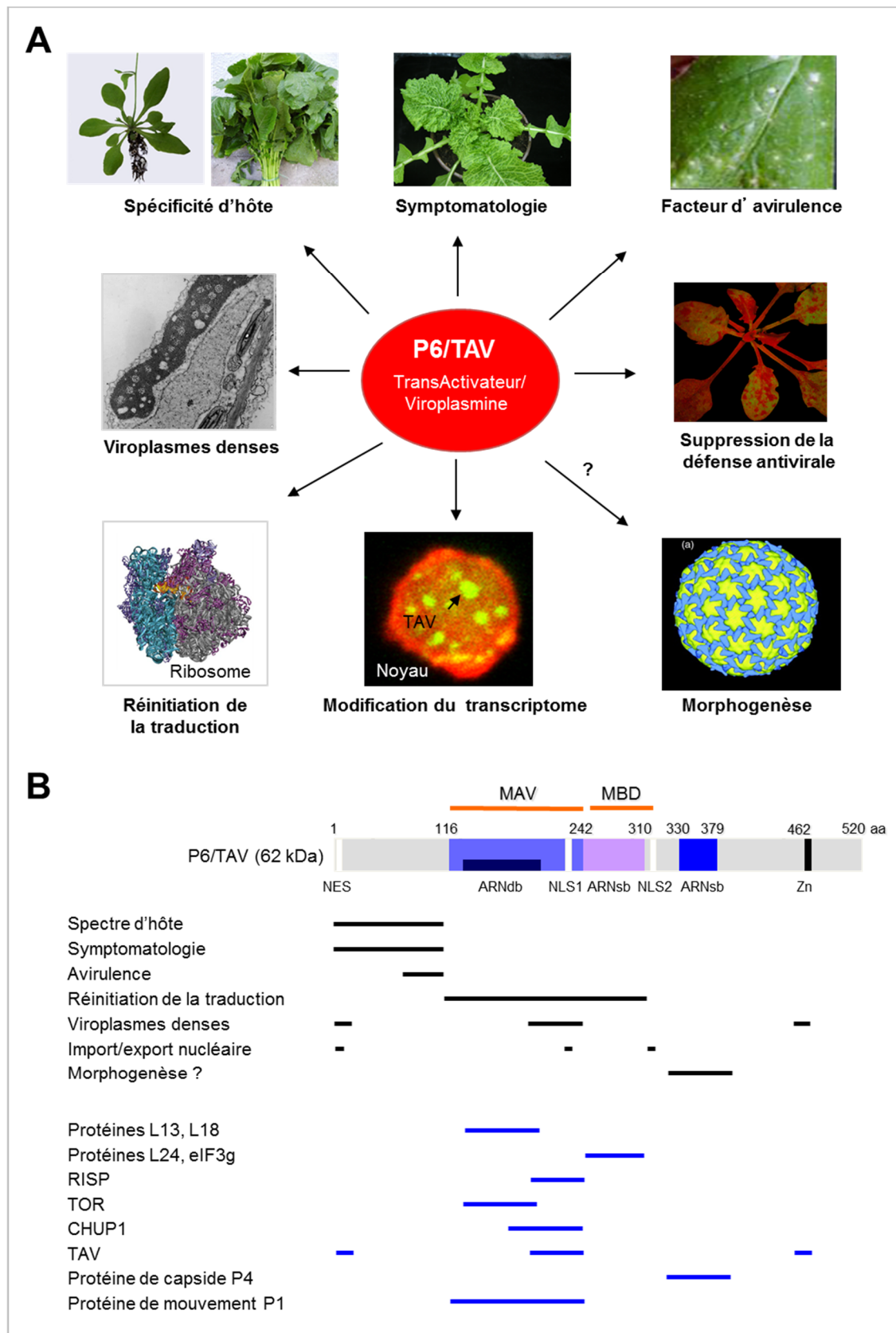


Figure 2 : Les différentes fonctions associées à la protéine P6/TAV du CaMV (A) et les domaines impliqués dans ces fonctions (B). NES : signal d'export nucléaire. NLS : signal de localisation nucléaire. ARN db : domaine de liaison à l'ARN double brin. ARN sb : domaines de liaison à l'ARN simple brin. Zn : motif en doigt à zinc. MAV : *miniTAV*, domaine minimal requis pour la transactivation traductionnelle. MBD : *multiple binding domain*.

Le cycle infectieux du CaMV

Après son relargage dans la cellule par le puceron (Figure 3-1), selon un mécanisme encore méconnu, la particule de CaMV est transportée vers les pores nucléaires grâce à la machinerie d'importation qui reconnaît le NLS des protéines de la capsid. Il est vraisemblable que l'ADN viral soit libéré à l'entrée des pores avant d'être importé (Figure 3-2), car la capsid n'a jamais été détectée dans le noyau. Les interruptions de séquences de l'ADN viral sont ensuite réparées par les enzymes nucléaires en une molécule d'ADN double brin circulaire et fermée de manière covalente. Cette dernière est associée aux histones pour former un minichromosome viral qui comporte 42 ou 43 nucléosomes (Figure 3-3). L'ADN subit des méthylations, une semaine après le début de l'infection des plantes hôtes [43]. Le génome viral est transcrit par l'ARN polymérase II cellulaire en un ARN 35S – dont une fraction de la population subit un épissage alternatif – et un ARN 19S, qui sont tous deux coiffés et polyadénylés (Figure 3-4). La ou les voies d'export nucléaire empruntées par les ARN viraux épissés ou non, n'ont pas encore été identifiées (Figure 3-5). L'expression de l'ARN 19S monocistronique se fait par un mécanisme de traduction coiffe-dépendante classique, alors que celle de l'ARN 35S met en jeu des stratégies non-canoniques, où P6/TAV a un rôle majeur. En plus de s'impliquer dans la traduction de l'ARN 35S, P6/TAV contribue à la formation des viroplasmes denses et elle agit, dans le noyau, comme suppresseur du *RNA silencing* (Figure 3-6). La réplication de l'ADN viral s'effectue dans les viroplasmes denses, par rétrotranscription de l'ARN pré-génomique grâce à la transcriptase inverse (Figure 3-7). Après morphogenèse (Figure 3-8), les nouveaux virions sont transportés à travers les plasmodesmes (Figure 3-9) pour infecter les cellules voisines, tandis que certains réinfectent le noyau.

Nous décrivons les étapes du cycle infectieux qui nous paraissent essentielles et/ou qui viennent d'être éclairées par de nouvelles données, à l'exception de la transmission du CaMV par le puceron qui a déjà fait l'objet de plusieurs excellentes revues [10, 24, 44].

Transcription du génome

Le génome du CaMV est transcrit de manière asymétrique par l'ARN polymérase II cellulaire, en deux ARN majeurs, les ARN 35S et 19S, qui sont coiffés en 5' et polyadénylés en 3'. Ces ARN sont synthétisés à partir de leurs propres promoteurs, les promoteurs 35S et 19S qui sont localisés respectivement dans la grande et la petite région intergénique du génome viral. L'ARN 35S est polycistronique et sert, à la fois, de messenger pour la production de toutes les protéines virales, et d'intermédiaire de réplication – ARN pré-génomique. L'ARN 19S, monocistronique, code pour la protéine P6/TAV et il est colinéaire à l'extrémité 3' de l'ARN 35S ; c'est un ARN subgénomique. Des petits ARN 8S sens et anti-sens, non-polyadénylés, sont également détectés au cours de l'infection. L'ARN 8S sens est probablement produit à partir du promoteur 35S sur un génome viral relâché, puisque la fin de la transcription correspond à l'interruption de séquence $\Delta 1$ sur le brin (-). La forme anti-sens de l'ARN 8S serait produite à partir de l'ARN sens, par un mécanisme encore inconnu. Les ARN 8S semblent jouer un rôle de leurre pour bloquer la machinerie du *RNA silencing* [45]. Enfin, un ARN 22S serait produit à partir d'un promoteur situé dans l'ORF IV [46].

Le promoteur 35S impliqué dans la synthèse de l'ARN 35S pré-génomique, comporte une boîte TATA, une boîte CAAT et des éléments *enhancer* de type GTGGX₃G similaires à ceux trouvés chez d'autres virus à ADN. Il renferme également deux séquences conférant une spécificité tissulaire dans l'expression des gènes : le domaine A (nucléotides - 90 à + 8), qui permet l'expression au niveau des racines, et le domaine B (nucléotides -343 à -90), qui permet l'expression dans les parties aériennes de la plante. Une analyse plus fine du domaine B a permis de caractériser cinq sous-domaines qui sont impliqués dans la transcription spécifique de gènes dans différents tissus ou à différents stades de développement de la plante [47]. Cette expression différentielle est régulée grâce à la fixation de divers facteurs protéiques d'origine nucléaire, dont ASF-2 et TGA1 [48, 49]. Le promoteur 19S dirige la transcription de l'ARN 19S ; il ne possède pas tous les éléments régulateurs décrits chez le promoteur 35S et plus particulièrement, les séquences activatrices de la transcription. Le promoteur

35S est très employé en ingénierie et en biologie fondamentale des plantes car il confère, en l'absence de facteurs viraux, une expression constitutive très élevée des gènes qui lui sont associés et ce, dans la plupart des tissus de la majorité des espèces de plantes.

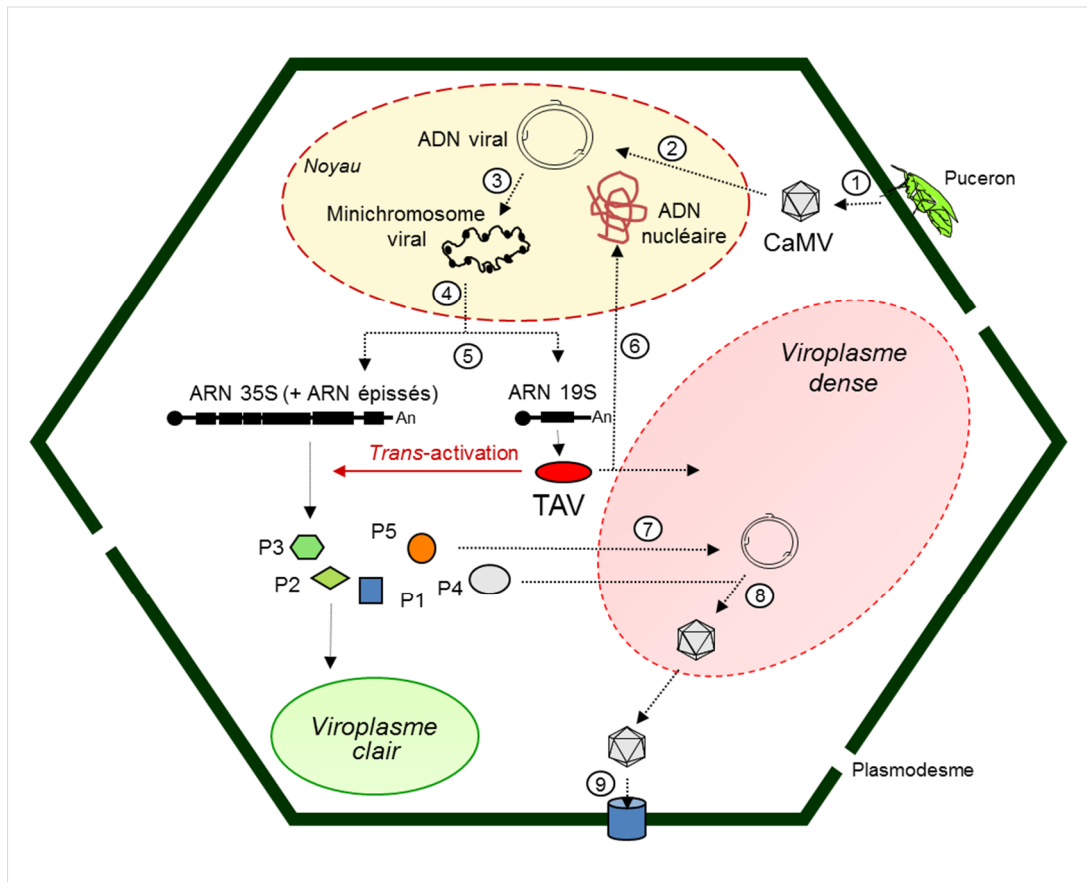


Figure 3 : Cycle infectieux du CaMV. 1 : Transmission du virus par le puceron (*Myzus persicae*). 2 : Entrée de l'ADN viral dans le noyau. 3 : Réparation de l'ADN viral et formation du minichromosome. 4 : Transcription de l'ADN par l'ARN polymérase II cellulaire et épissage alternatif. 5 : Export nucléaire de l'ARN 19S qui est traduit en P6/TAV et, de l'ARN 35S et des ARN épissés qui sont exprimés par réinitiation de la traduction grâce à P6/TAV pour donner naissance à un jeu complet de protéines (P1, P2, P3, P4, P5 et P6/TAV). La protéine P2 s'assemble pour former un viroplasma clair. 6 : Réplication du génome viral par rétrotranscription de l'ARN 35S. 7 : Morphogénèse des virions dans les viroplasmases denses. 8 : Mouvement intercellulaire des virions à travers les tubules formés au niveau des plasmodesmes. 9 : Import nucléaire de P6/TAV, nécessaire à sa fonction de suppresseur du RNA silencing.

La transcription du génome du CaMV serait contrôlée par la phosphorylation du domaine C-terminal de l'ARN polymérase II, par le facteur P-TEFb – pour « *positive transcription elongation factor* ». En effet, des plants d'*A. thaliana* mutés au niveau de gènes nucléaires *CDKC;2* et/ou *CYCT1;5*, dont les produits constituent le P-TEFb, sont résistants à l'infection par le CaMV alors qu'ils restent sensibles à des virus à ARN [50]. Il est connu que la protéine Tat du VIH-1 interagit, au début de la transcription, avec une structure en tige-boucle située à l'extrémité 5' de l'ARN viral, le TAR – pour « *trans-activating responsive element* ». Cette interaction permet le recrutement d'un complexe P-TEFb qui va phosphoryler le domaine C-terminal de l'ARN polymérase II, permettant à cette dernière de synthétiser des transcrits viraux complets, du fait d'une meilleure processivité [51]. Par similitude, il n'est donc pas exclu qu'une protéine du CaMV recrute le complexe P-TEFb au niveau de la structure en tige-boucle de la région *leader* de l'ARN 35S, ce qui augmenterait la processivité de l'ARN polymérase II. P6/TAV pourrait être une excellente candidate pour cette fonction car elle possède des propriétés caryophiles et elle est capable d'interagir avec l'ARN simple- et double-brin.

L'ARN 35S (8,2 kb) est légèrement plus long que le génome parce qu'il comporte des séquences terminales redondantes, qui sont l'équivalent des séquences R des structures LTR (*long terminal repeat*) des rétrovirus. Ces redondances jouent un rôle important dans la rétrotranscription car elles permettent le saut réplicatif. Elles résultent du fait que l'unique signal de terminaison/polyadénylation présent sur le génome, est situé à environ 180 nucléotides en aval du site d'initiation de la transcription. Dans la majorité des cas, le complexe transcriptionnel ignore le signal de polyadénylation lors de son premier passage ; dans le cas contraire, l'arrêt de la transcription génère des petits ARN longs d'environ 180 nucléotides appelés *short-stop*. Le signal de polyadénylation comprend la séquence caractéristique AAUAAA qui détermine le clivage du transcrit primaire à environ 13 nucléotides en aval et, en amont, des séquences répétées UUUUGUA qui agissent en *cis* pour améliorer l'efficacité de la maturation en 3' de l'ARN 35S [52]. Le clivage et la polyadénylation à l'extrémité 3' de l'ARN 19S (1,5 kb) s'effectuent de la même façon que pour l'ARN 35S.

Épissage de l'ARN 35S

Des études réalisées sur la souche de CaMV Cabb-S, montrent qu'environ 70% de la population d'ARN 35S subit un épissage alternatif, tandis que 30% sont préservés, afin de servir de matrice à la transcriptase inverse et de messenger pour la synthèse d'un jeu complet de protéines virales. L'épissage alternatif décrit dans ces études implique 4 sites donneurs, nommés D1 à D4 et localisés dans l'ORF I et dans la région 5' non traduite de l'ARN 35S, et un site accepteur unique A situé dans l'ORF II [53]. Son rôle principal serait de participer à une régulation négative de l'expression de l'ORF II, afin de prévenir une suraccumulation potentiellement toxique de la protéine P2 [54]. Les événements d'épissage conduisent très souvent à la fusion, en phase, d'ORF, notamment entre la partie 5' de l'ORF I et la partie 3' de l'ORF II, créant ainsi des ORF codant pour des protéines de fusion P1P2 dont l'expression reste actuellement hypothétique. La fusion des ORF éviterait la multiplicité de petits ORF qui pourraient inhiber la traduction des ARN 35S épissés. Des molécules d'ADN délétées résultant de la rétrotranscription des ARN épissés sont probablement encapsidées pour être acheminées vers le noyau [55] si l'on en juge par des études de microscopie électronique qui montrent la présence de molécules d'ADN circulaires d'origine virale, plus petites que l'ADN viral génomique, dans les noyaux de cellules infectées [56]. Une étude approfondie, faite récemment sur les souches Cabb-S et Cabb B-JI, montre que le nombre de sites d'épissage est plus élevé que déterminé initialement [Bouton *et al.*, résultats non publiés] ce qui donne lieu à un schéma d'épissage d'une complexité équivalente à celui décrit chez le virus de l'hépatite B [57]. Il a été montré que les ARN épissés du HBV sont souvent détectés chez les patients atteints d'hépatite chronique, et qu'ils s'accumulent préférentiellement lorsque cette hépatite chronique se transforme en un hépatocarcinome [58]. Par ailleurs, comme chez le CaMV, certains ARN épissés du HBV possèdent un nouvel ORF résultant dans ce cas, de la fusion de l'extrémité 5' du gène codant pour la transcriptase inverse avec l'extrémité 3' d'un ORF créé par l'évènement d'épissage. Cette protéine hautement immunogène nommée HBSP – pour « *HBV splice-generated protein* » – n'a pas de rôle bien défini. Des études utilisant de la protéine HBSP produite hors-contexte viral, ont en tout cas démontré qu'elle n'a pas d'influence sur la réplication et la transcription du virus [57]. À l'heure actuelle, on ne dispose d'aucune information sur de possibles autres fonctions et sur la régulation de l'épissage alternatif de l'ARN 35S du CaMV, ni sur les voies d'export empruntées par les ARN 35S épissés et par l'ARN 35S complet.

Traduction des ARN du CaMV

Diverses observations ont conduit à considérer les viroplasmes denses comme le site où s'effectue la traduction des ARNm du CaMV : i) ils sont entourés de nombreux ribosomes, ii) toutes les protéines virales s'accumulent à leur niveau sauf P2 et iii) des protéines hétérologues, exprimées à partir d'un ARN 35S chimère, sont retrouvées au niveau de ces viroplasmes [2].

La protéine P6/TAV du CaMV est synthétisée à partir de l'ARN 19S monocistronique suivant un mécanisme traductionnel coiffe-dépendant canonique. À l'inverse, les autres protéines virales sont

produites à partir de l'ARN 35S polycistronique et de ses formes épissées qui sont également polycistroniques, grâce à deux mécanismes inhabituels, le saut du ribosome et la réinitiation de la traduction. La mise en œuvre de ces deux stratégies traductionnelles s'explique par la structure particulière de l'ARN 35S. En effet, bien que coiffé en 5' et polyadénylé en 3', cet ARN porte les sept ORF du génome viral, ce qui lui confère des caractères à la fois eucaryotes et procaryotes. Par ailleurs, sa région 5' non traduite – région *leader* – est très longue, environ 600 nucléotides, et elle forme une tige-boucle multibranchée de grande taille (478 nucléotides) qui est jalonnée de huit petits ORF (3 à 30 codons). Ces petits ORF (sORF) sont tous localisés dans la structure secondaire de la région *leader*, à l'exception du premier, le sORF A (3 codons), qui se trouve à sept nucléotides en amont de cette structure [59].

Le saut du ribosome, ou ribosomal shunt

Des expériences d'expression transitoire dans des protoplastes ont montré que la région *leader* de l'ARN 35S inhibe fortement (de 70 à 90%) l'expression de gènes rapporteurs suggérant que la traduction de l'ARN 35S s'effectue comme chez d'autres virus – les *Picornavirus*, par exemple – par une initiation interne, la structure secondaire du *leader* faisant office d'IRES. Toutefois, cette hypothèse a été rapidement écartée car l'introduction, en amont cette structure secondaire, d'une structure Kozak – tige-boucle riche en G-C, qui bloque la migration des ribosomes – empêche l'expression des gènes rapporteurs, ce qui démontre sans ambiguïté que l'initiation de la traduction est coiffe-dépendante. Ces gènes sont également exprimés lorsque la structure Kozak est introduite au sein de la tige-boucle de la région *leader*, ce qui signifie que les ribosomes court-circuitent cette région grâce à un saut qui leur permet d'atterrir au niveau du premier ORF majeur, l'ORF VII (Figure 4). Toutefois, il est apparu que certains ribosomes n'effectuent pas le saut mais poursuivent leur migration le long du *leader*, la plupart d'entre eux se décrochant sans doute de l'ARN 35S au fur et à mesure qu'ils traduisent les sORF. Deux éléments structuraux revêtent une importance capitale pour le saut du ribosome : le sORF A et la base – la *stem 1* – de la tige-boucle, qui permet de rapprocher spatialement ce petit ORF et l'ORF VII. Des expériences de mutagenèse dirigée qui déstabilisent les appariements de la *stem 1* inhibent le saut du ribosome. De même, il a été montré clairement que la traduction complète du sORF A, ainsi que le relargage du tripeptide correspondant, est indispensable pour que les ribosomes réalisent le saut et puissent ensuite réinitier la traduction des ORF majeurs. Le saut est très efficace lorsque la distance entre le sORF A et la tige-boucle est de 5 à 10 nucléotides, et si la longueur du sORF A est comprise entre 2 et 15 codons. Il est vraisemblable que la taille réduite du sORF A empêche les ribosomes de se débarrasser des facteurs d'initiation, lors de la courte phase d'élongation de la traduction, ce qui les rend immédiatement compétents pour réinitier la traduction après leur saut [36, 60]. La protéine P6/TAV du CaMV n'est ici pas indispensable, mais elle stimule la réinitiation immédiatement après le saut du ribosome, probablement en favorisant le maintien du facteur d'initiation eIF3 sur le ribosome. Ce facteur, qui est constitué d'une douzaine de sous-unités protéiques, joue un rôle capital dans la formation du complexe de pré-initiation de la traduction – le PIC 43S –, notamment en permettant le recrutement du complexe ternaire eIF2-Met-ARNt^{Met}-GTP.

Le saut du ribosome est mis en œuvre chez tous les virus de la famille des *Caulimoviridae* à l'exception du *Cestrum yellow leaf curling virus*, quel que soit leur genre, car tous possèdent un ARN 35S avec une longue région *leader* structurée [60]. La conservation de cette structure au cours de l'évolution suggère qu'elle joue un rôle important au cours du cycle infectieux et de ce fait, que les virus ont dû adapter leur stratégie de traduction pour contourner cet obstacle. Cette structure secondaire est sans doute impliquée dans l'encapsidation et elle est la source de petits ARN, générés par le *RNA silencing*, dont le rôle serait d'inhiber l'expression de gènes cellulaires [61].

Réinitiation de la traduction de l'ARN 35S

La réinitiation est un mécanisme déjà décrit pour la traduction de petits ORF présents dans la région *leader* de différents ARNm cellulaires. La traduction de sORF est connue depuis longtemps pour jouer un rôle clé dans l'expression de l'ORF majeur d'un messenger, l'exemple le plus connu et

l'un des premiers à avoir été décrit étant celui de l'ARNm codant pour le facteur de transcription GCN4 chez la levure. Par contre, la réinitiation de la traduction de longs ORF dans un système eucaryote était méconnue voire considérée comme impossible, jusqu'à sa découverte chez le CaMV.

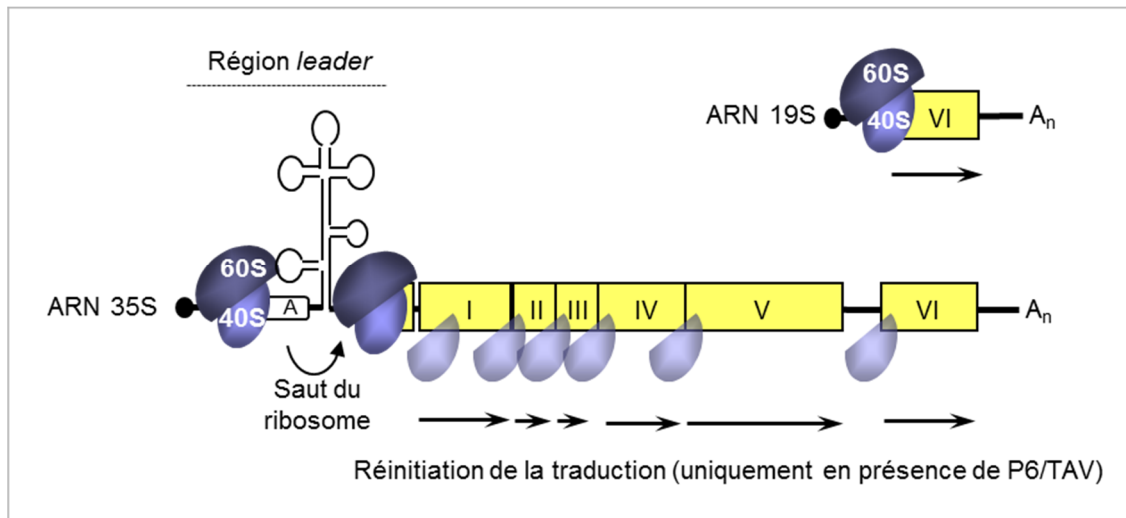


Figure 4 : Représentation schématique de l'organisation structurale des ARN 35S et 19S, et des deux stratégies de traduction non-canoniques mises en œuvre pour exprimer l'ARN 35S, le saut du ribosome et la réinitiation de la traduction. La boîte A représente le sORF A (3 codons), dont la traduction complète est requise pour qu'il y ait saut du ribosome. La structure secondaire de la région *leader* renferme les autres sORF (non représentés). L'ORF VII est caché par le ribosome 80S.

La réinitiation de la traduction des ORF de l'ARN 35S est activée par la protéine virale P6/TAV. Aucune séquence spécifique n'est requise puisque P6/TAV active également la réinitiation de la traduction d'un ARNm bicistronique hétérologue. À l'inverse du système GCN4, la réinitiation est indépendante de la distance entre deux ORF, son efficacité étant identique, que deux ORF consécutifs soient séparés de quelques nucléotides ou de 600 nucléotides. Diverses études, avec des approches expérimentales différentes, ont montré que P6/TAV interagit avec des composants de la machinerie traductionnelle des plantes (Figure 5). Ainsi P6/TAV s'associe directement à la grande sous-unité ribosomale 60S, par l'intermédiaire des protéines ribosomales L13, L18 et L24 – L24 a une localisation différente de L13 et L18 sur le ribosome –, et avec la sous-unité g du facteur d'initiation eIF3. Il a été également montré *in vitro* qu'eIF3 sert de pont entre la petite sous-unité ribosomale 40S et P6/TAV suggérant la formation d'un complexe entre P6/TAV et la sous-unité 40S portant eIF3 [40]. Plus récemment, deux nouveaux partenaires d'interaction de P6/TAV ont pu être identifiés et caractérisés : un facteur protéique spécifique des plantes, jusqu'alors de fonction inconnue, que nous avons appelé RISP – pour « *reinitiation supporting protein* » – et la protéine sérine/thréonine kinase TOR – pour « *target of rapamycin* » [40, 62, 63]. Rappelons que TOR existe chez les animaux et la levure sous la forme de deux complexes, mTORC1 et mTORC2, qui régulent pour le premier, la traduction, la croissance et la prolifération cellulaire et pour le second, la structuration du cytosquelette. Ces diverses actions régulatrices se font en réponse à l'état énergétique de la cellule, au stress, aux facteurs de croissance ou encore, en fonction de l'apport des nutriments. Une fois activée par sa voie de signalisation, TOR phosphoryle la protéine kinase S6K1 et 4E-BP. Contrairement aux animaux et à la levure, la voie de signalisation TOR et ses substrats sont beaucoup moins connus chez les plantes.

Les partenaires de P6/TAV – les protéines L13, L18, L24 de la sous-unité ribosomale 60S, eIF3, RISP et TOR – sont impliqués dans un réseau d'interactions extrêmement complexe (Figure 5) qui est finement régulé, et au sein duquel les domaines MAV et MBD de P6/TAV jouent un rôle central. Ainsi, les protéines L13 et L18 se lient à la partie centrale du domaine MAV alors que L24 s'associe, *via* son extrémité N-terminale, au domaine MBD. TOR interagit, grâce à ses répétitions HEAT présentes en N-terminal, avec la région N-terminale de MAV [64]. Quant à la protéine RISP, qui

est constituée de 4 hélices α – H1 à H4 – elle se lie, *via* H3, à la région C-terminale de MAV et *via* H4, à la région C-terminale de L24. L'interaction de RISP avec les sous-unités a et c d'eIF3 implique l'hélice H2 [62].

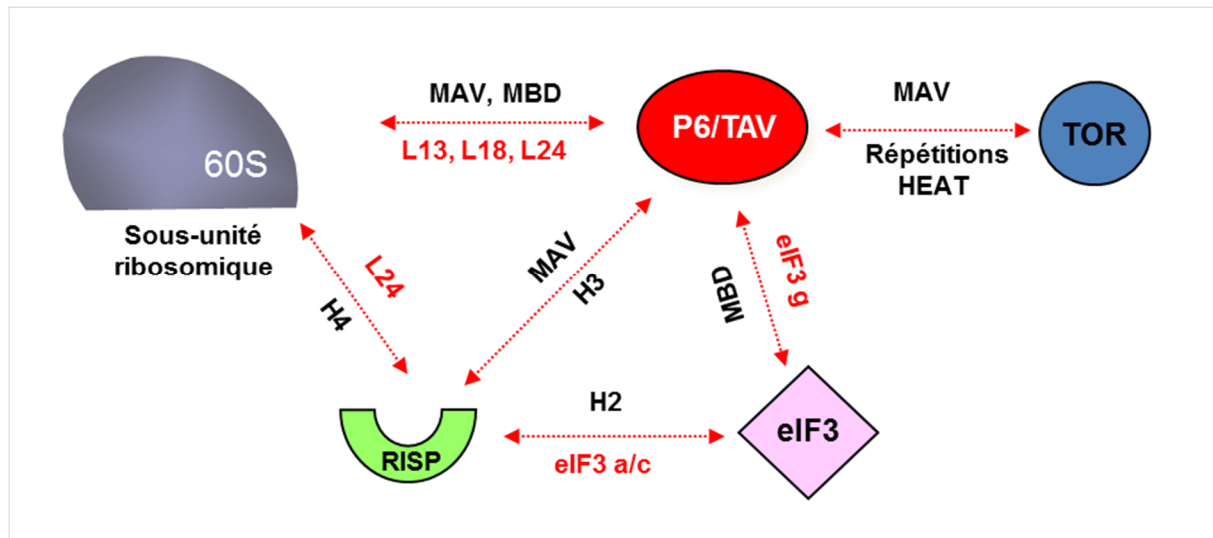


Figure 5 : Réseau d'interactions entre P6/TAV et la machinerie traductionnelle de l'hôte. Les différents partenaires de P6/TAV et de RISP, au sein du ribosome et de eIF3, sont indiqués en rouge et les domaines impliqués dans les interactions sont précisés de part et d'autre des doubles flèches. H : hélice α . MAV : *miniTAV*. MBD : *multiple binding domain*. RISP : *reinitiation supporting protein*. TOR : *target of rapamycin*. D'après Schepe-tilnikov *et al.* [40], modifiée.

RISP est un co-facteur de P6/TAV. En effet, sa présence stimule fortement la réinitiation de la traduction des ARN polycistroniques, et cette stimulation requiert une interaction physique entre RISP et P6/TAV. L'importance de cette interaction pour le cycle infectieux du CaMV est confirmée par le retard dans l'infection virale chez des mutants d'*A. thaliana* où un de ses deux gènes codant pour RISP a été inactivé par un transposon – l'inactivation des deux gènes étant létale pour la plante [62]. Cependant, il est apparu assez rapidement que les interactions entre P6/TAV et les différents partenaires de la machinerie traductionnelle, au niveau des polysomes, ne sont pas suffisantes en soi, pour outrepasser les barrières cellulaires à la réinitiation, et que la phosphorylation de certains facteurs est requise pour activer la réinitiation. La découverte de l'interaction de P6/TAV avec la protéine kinase TOR fut naturellement déterminante pour disséquer et comprendre le mécanisme de la réinitiation. Effectivement, nous avons montré, par des approches *in vitro* et *in vivo*, que le CaMV active la voie de signalisation TOR à travers une interaction directe entre P6/TAV et TOR, et que cette activation se manifeste par la phosphorylation de TOR et de son substrat, la protéine kinase S6K, ainsi que celle de RISP, au niveau de la sérine du motif RGRLES impliqué dans la liaison à P6/TAV [63]. Il est probable que la phosphorylation de RISP soit requise pour promouvoir la réinitiation de la traduction par P6/TAV. L'activation de TOR est capitale pour la réplication du CaMV, comme le montre l'incapacité du virus à infecter des plants d'*A. thaliana* dont la synthèse de TOR est abolie par l'expression de siRNA spécifiques – *A. thaliana* ne possède qu'un seul gène TOR – ou lorsque l'ADN viral code une protéine P6/TAV mutée incapable d'interagir avec TOR [63]. Ces résultats constituent le premier exemple démontrant l'implication de TOR dans le cycle infectieux d'un virus de plantes, alors que de nombreux articles concernant TOR et les virus des animaux jalonnent la littérature. En effet, les virus à ADN ou à ARN des animaux, qui traduisent leurs messagers de manière coiffe-dépendante, activent directement le complexe mTORC1, ou indirectement, en agissant sur la voie de signalisation en amont. Ceci aboutit à la phosphorylation de la protéine 4E-BP et ainsi, à la libération du facteur d'initiation eIF-4E qui est indispensable pour la formation du complexe de préinitiation au niveau de la coiffe de l'ARNm. A l'inverse, le CaMV a choisi la stratégie de maintenir, grâce à P6/TAV, la protéine

TOR sous un état de phosphorylation permanent, condition *sine qua non* pour que la réinitiation de la traduction s'effectue tout au long de l'ARN 35S polycistronique.

L'ensemble des données actuelles a conduit à proposer le modèle suivant (Figure 6) : immédiatement après le saut, les ribosomes réinitient la traduction de l'ORF majeur – rappelons que le sORF A vient d'être traduit. Au cours de l'élongation, les facteurs eIF3 et RISP sont transloqués à la surface de la sous-unité 60S des polysomes pour ne pas interférer avec l'élongation et ce, grâce à P6/TAV qui interagit simultanément avec eIF3, RISP et L18/L13. La protéine TOR activée par P6/TAV s'associe également aux polysomes, probablement en même temps qu'eIF3 et RISP, ce qui conduit à la phosphorylation de S6K1 et RISP. Lors de la terminaison de la traduction, le complexe eIF3-RISP-P6/TAV-TOR est relocalisé au niveau de la sous-unité 40S qui scanne l'ARN jusqu'à l'AUG initiateur de l'ORF suivant, tout en recrutant le complexe ternaire eIF2-Met-ARNt^{Met}-GTP et la sous-unité 60S. La phosphorylation de TOR et de RISP associées aux polysomes est essentielle pour la rétention des ribosomes 80S sur l'ARNm polycistronique, et la réutilisation d'eIF3 et de RISP, afin de rendre les ribosomes compétents pour une nouvelle initiation de la traduction.

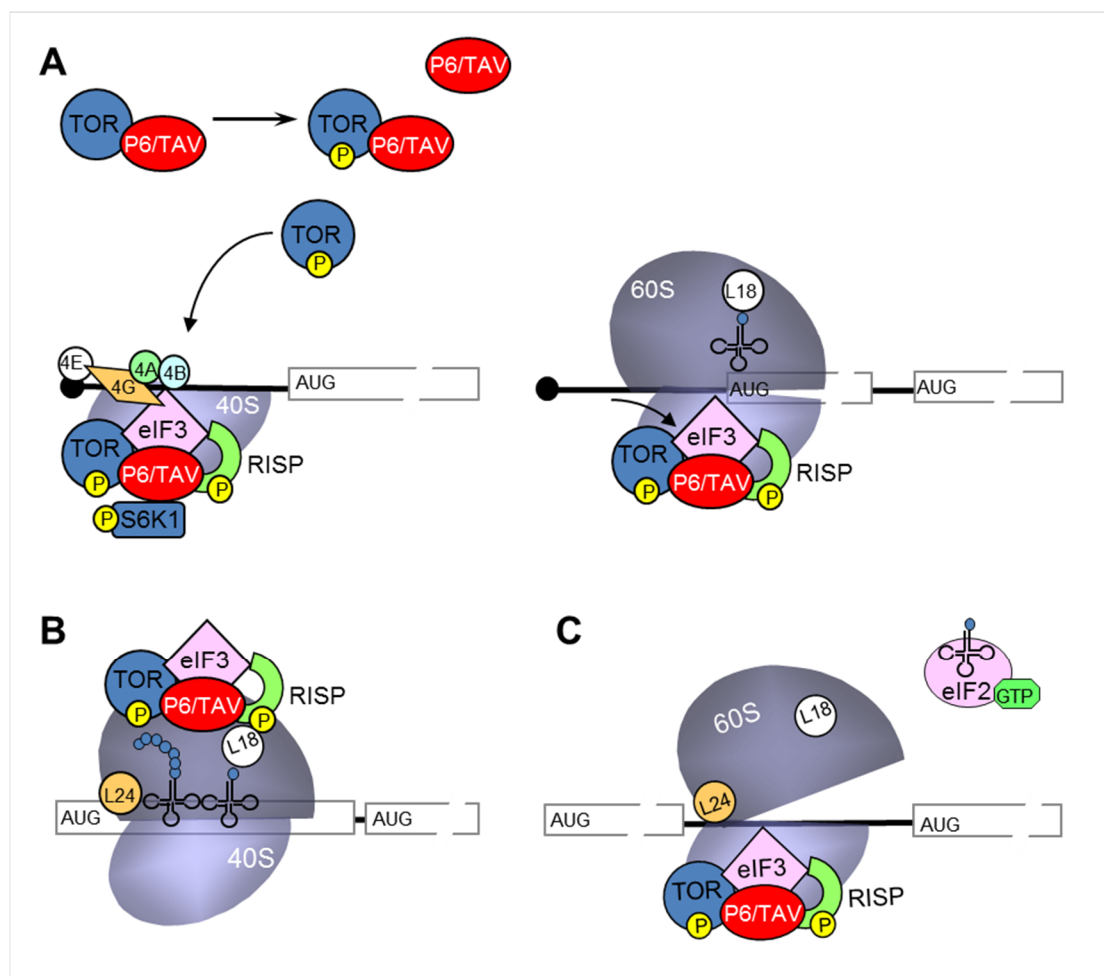


Figure 6 : Modèle hypothétique du mécanisme de la réinitiation de la traduction des ARNm polycistroniques, grâce à la protéine P6/TAV. (A) La protéine kinase TOR activée par P6/TAV se fixe sur le complexe de préinitiation (PIC 43S), vraisemblablement en même temps que eIF3 et RISP et ce, avant la jonction de la sous-unité 60S du ribosome. TOR phosphoryle alors ses substrats RISP et S6K1, qui est relarguée de la sous-unité 40S. Le mécanisme impliqué dans la phosphorylation de TOR est pour l'instant non élucidé. (B) Au cours de l'élongation, le complexe eIF3-P6/TAV-RISP-TOR est transloqué sur la sous-unité 60S puis, (C) avant la terminaison de la traduction de l'ORF, est ramené au niveau de la sous-unité 40S afin de la rendre compétente pour une nouvelle initiation de la traduction.

La réinitiation de la traduction de l'ARN 35S s'effectue chez le genre *Soymovirus* de la famille des *Caulimoviridae*, vraisemblablement selon le même mécanisme de fonctionnement que celui décrit pour le CaMV étant donné que les *Soymovirus* produisent également une viroplasmine (P6/TAV). À l'inverse, les *Cavemovirus* n'ont pas de véritable équivalent à la protéine P6/TAV [82], suggérant que leur ARN 35S polycistronique est traduit selon un mécanisme différent. Les autres genres de *Caulimoviridae* ont fusionné partiellement – *Tungrovirus* – ou totalement leurs ORF – *Petuvirus* –, pour en faire un ORF unique codant pour une polyprotéine [21].

Réplication du génome du CaMV par rétrotranscription

La rétrotranscription de l'ARN 35S pré-génomique du CaMV (Figure 7) s'effectue dans le cytoplasme, au niveau des viroplasmes denses, à la fin du cycle infectieux comme chez tous les pararétrovirus [64]. Le mécanisme de la réplication de l'ADN du CaMV est peu étudié, certainement parce que les premières données obtenues indiquaient de très fortes ressemblances avec celui des rétrovirus. La transcriptase inverse du CaMV nécessite comme amorce l'ARNt^{Met} initiateur, ce qui le distingue du HBV qui se sert de l'enzyme-même pour synthétiser une amorce au niveau de la structure ϵ de l'ARN pré-génomique. L'ARNt^{Met} s'apparie sur l'ARN 35S à environ 600 nucléotides de son extrémité 5', au niveau du site PBS – pour « *primer binding site* ». La synthèse du brin d'ADN complémentaire de polarité négative (-) se fait jusqu'à l'extrémité 5' de la matrice, produisant un ADN dit *strong stop* et recopiant la séquence terminale redondante (R) de l'ARN ; l'ARN hybridé à l'ADN est détruit par l'activité RNase H associée à la transcriptase inverse. Le segment d'ADN s'apparie ensuite à la séquence R située en 3' de l'ARN 35S – c'est le saut réplcatif – et sert d'amorce pour la synthèse du brin complet d'ADN (-). Au cours de cette synthèse, l'ARN matriciel est dégradé par la RNase H, sauf au niveau de séquences, deux à quatre selon les souches de CaMV, particulièrement riches en purines – suites de G. Ces séquences d'ARN sont utilisées comme amorces pour la synthèse discontinue du brin d'ADN (+). L'élongation de l'un des deux brins (+), qui se trouve bloquée lorsque celui-ci arrive à l'extrémité 5' du brin (-), continue après son appariement avec la région 3' du brin (-). Cet appariement est à l'origine de la circularisation de l'ADN viral et de l'apparition de l'interruption de séquence $\Delta 1$; l'ARNt^{Met} initiateur est éliminé partiellement de sorte que quelques ribonucléotides subsistent en 5' du brin chevauchant. La synthèse des brins positifs se poursuit jusqu'à ce qu'elle soit arrêtée par le brin situé en amont. Le déplacement de ce dernier par la transcriptase inverse, sur une distance de quelques nucléotides, conduit à l'apparition des interruptions de séquence Δ sur le brin (+). L'encapsidation de l'ADN du CaMV est une étape non élucidée, mais il est vraisemblable qu'elle soit couplée à la réplication, comme décrit chez les hépadnavirus. Il a d'ailleurs été montré que la protéine structurale P4 interagit avec le sommet de la tige-boucle de la région *leader*, suggérant que la capside s'assemble à ce niveau dès que commence la rétrotranscription.

Mouvement de cellule-à-cellule du CaMV

La protéine de mouvement P1 (MP) du CaMV s'assemble au niveau des plasmodesmes, pour former des tubules qui permettront le passage des virions de la cellule infectée vers les cellules adjacentes. P1 interagit physiquement avec une protéine d'*A. thaliana*, MPI17, qui appartient à une famille de protéines qui régulent le trafic vésiculaire entre les différents compartiments cellulaires y compris le réticulum endoplasmique, l'appareil de Golgi et les endosomes. La protéine MPI17 colocalise avec P1 sous forme de petits foyers au site d'initiation de la formation des tubules [6]. Le transport de P1 de son site de synthèse, en l'occurrence les viroplasmes denses, vers la périphérie de la cellule, implique le système de transport endomembranaire [65]. L'élaboration des tubules est inhibée d'ailleurs en présence de brefeldine A, un inhibiteur spécifique du transport de réticulum endoplasmique vers l'appareil de Golgi.

Récemment, il a été montré que la MP du CaMV interagit, au sein des plasmodesmes, avec une protéine intrinsèque appelée PDL1 – pour « *plasmodesmata-located protein 1* » [66]. Cette protéine cellulaire est associée à la membrane plasmique, l'extrémité N-terminale étant localisée dans

l'apoplasme, et l'extrémité C-terminale dans le cytoplasme. PDLP1 appartient à une petite famille de gènes de protéines d'*A. thaliana* qui modulent le trafic de cellule-à-cellule, et qui sont adressées aux plasmodesmes par la voie sécrétoire grâce à l'hétérodimère COPII qui, rappelons-le, est impliqué dans le transport de vésicules du réticulum endoplasmique vers l'appareil de Golgi. Des données cyto- logiques obtenues à l'aide de sondes fluorescentes suggèrent que P1 interagit avec PDLP1 à la base des tubules. L'interaction entre P1 et PDLP1 est certainement essentielle pour le mouvement interce- lulaire du CaMV, car l'infection généralisée ne s'effectue pas systématiquement chez les plantes d'*A. thaliana* renfermant des gènes PDLP mutés. Le passage des virions à travers les tubules s'effectue grâce à la protéine de capsid mineure P3, P1 étant incapable d'interagir directement avec la capsid. En effet, P3 interagit avec P1 *in vivo* dans des tests de double hybride en levure, et des marquages à l'or colloïdal couplés à la microscopie électronique montrent clairement que P1 et P3 colocalisent avec les virions uniquement dans les plasmodesmes [14]. À ce jour, aucune information n'est disponible en ce qui concerne le transfert des virions des viroplasmes denses, où ils sont stockés, jusqu'aux tubules de P1. Toutefois, comme les viroplasmes denses se déplacent le long des microfilaments d'actine corticaux et vu que P6/TAV interagit avec les protéines de capsid P4 et P3, il n'est pas exclu que les viroplasmes servent de « transporteurs » des particules virales jusqu'à la périphérie de la cellule [67].

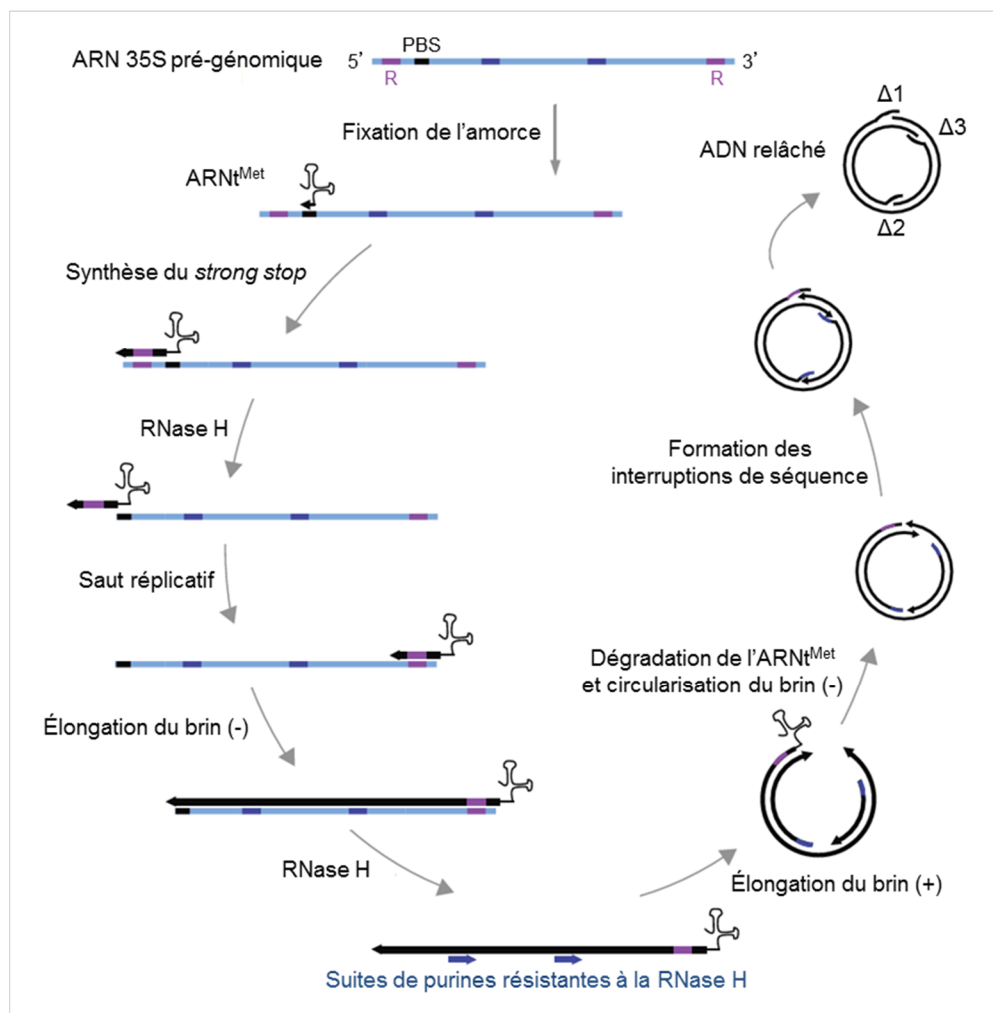


Figure 7 : Rétrotranscription de l'ARN 35S pré-génomique du CaMV. La synthèse de l'ADN complémentaire – brin (-) –, par la transcriptase inverse, est amorcée suite à la fixation d'un ARNt^{Met} initiateur puis complétée, après un saut réplcatif. L'activité RNase H de la transcriptase inverse dégrade l'ARN pré-génomique apparié à l'ADN (-), sauf au niveau de régions riches en purines qui serviront d'amorces pour la synthèse du brin (+). La circularisation des brins se fait suite à l'appariement de la séquence PBS à son complémentaire. Elle génère l'interruption de séquence Δ1 tandis que les interruptions Δ2 et Δ3 sont dues à la synthèse discontinue du brin (+). D'après G. Haas, modifiée.

Pathogénèse du CaMV

Spécificité d'hôtes

Le CaMV a un spectre d'hôte relativement restreint, contrairement à d'autres phytovirus. En effet, la majorité des souches virales, dont Cabb-S, CaMV-S Japon, Bari, D/H et CM1841, n'infectent systématiquement que des plantes de la famille des *Brassicaceae* – anciennement Crucifères – telles que diverses espèces de choux, le navet, le colza et certains écotypes d' *A. thaliana*. Elles se répliquent aussi localement dans les feuilles inoculées de certaines espèces de *Solanaceae* comme *Datura stramonium* – encore connu sous le nom de l'herbe des fous, à cause de ses propriétés psychotropes – et certains membres du genre *Nicotiana* mais elles sont incapables de les infecter systématiquement, à l'exception des souches de CaMV Cabb B-JI, D4, W260 et B29 [29, 68, Geldreich *et al.*, résultats non publiés].

Des expériences réalisées avec des génomes viraux recombinants obtenus à partir de l'ADN de souches ayant des spectres d'hôte différents (Solanacées et/ou Brassicacées) démontrent clairement que la région N-terminale de la protéine P6/TAV (acides aminés 1 à 116) est le déterminant majeur de la spécificité d'hôte du CaMV. Ainsi, la souche de CaMV D4 devient incapable d'infecter systématiquement les solanacées lorsque la région N-terminale de P6/TAV est remplacée par celles de ses homologues chez CM1841 ou Cabb-S. La souche chimère ainsi produite induit seulement, chez les Solanacées, une réponse hypersensible avec l'apparition de lésions locales nécrotiques [68]. Cependant, la situation s'avère plus complexe pour certains pathosystèmes – le système formé par une souche virale et une espèce de plante hôte –, ou même en fonction de certaines conditions comme la luminosité. Des séquences additionnelles de P6/TAV et éventuellement d'autres protéines virales sont alors nécessaires pour que la souche de CaMV puisse infecter son hôte. À titre d'exemple, l'infection généralisée de *Nicotiana bigelovii* par la souche D4 ne requiert que l'extrémité 5' de l'ORF VI alors que celle de *N. edwardsonii* est dépendante de l'ORF VI complet [30]. Concernant le pathosystème composé de *N. bigelovii* et de la souche W260, l'infection implique non seulement l'ORF VI complet, mais également d'autres régions du génome viral comprenant les ORF I, II et IV [69]. L'importance du domaine N-terminal de P6/TAV dans l'infection systémique des plantes hôtes est probablement liée à sa capacité d'interagir avec la protéine de mouvement P1, mais des interactions avec d'autres partenaires viraux et/ou cellulaires contribuent sans doute aussi à déterminer le spectre d'hôte du CaMV.

Symptomatologie

L'infection systémique des plantes hôtes par le CaMV se manifeste par des symptômes qui se différencient par leur nature et leur sévérité selon la souche virale, l'espèce végétale et certaines conditions environnementales comme l'intensité lumineuse et la température. Une large palette de symptômes, allant de chloroses sévères à l'absence de symptômes visibles malgré la présence du virus, est ainsi observée en fonction de la souche virale et de l'écotype d'*A. thaliana*. Les symptômes apparaissent très rapidement – entre 10 et 14 jours après inoculation du virus – pour les souches de CaMV virulentes comme Cabb B-JI, et plus tardivement – trois à quatre semaines – pour certaines souches modérées comme la souche Bari. La chlorose, sous forme de mosaïque, et l'éclaircissement des nervures sont les symptômes foliaires les plus fréquemment observés chez les plantes infectées (Figure 8) ; ils s'accompagnent parfois d'un gaufrage des feuilles. Les plantes infectées sont aussi affectées par un nanisme plus ou moins prononcé, la réduction de taille pouvant atteindre, lors d'une infection par une souche virulente, 75%, en comparaison avec des plantes saines. La photopériode influe également sur la sévérité des symptômes induits par le CaMV, chez *A. thaliana*. Ainsi, les symptômes chez les plantes cultivées en conditions dites de jours longs – soit 16 h de lumière par jour – sont plus modérés que chez celles cultivées en conditions de jours courts – 8 h de lumière par jour –, bien que le taux de répllication du CaMV soit supérieur chez les premières plantes.

Si les études du phénotype induit par des souches de CaMV chimères ont mis en exergue la contribution de plusieurs régions du génome viral dans l'expression de certains types de symptômes,

elles ont surtout démontré que P6/TAV est le déterminant majeur de l'expression et la sévérité des symptômes. En outre, il est frappant de constater la similitude des phénotypes des plantes infectées par le CaMV et d'*A. thaliana* transgéniques exprimant P6/TAV [27,32]. Le type de symptômes et leur sévérité sont contrôlés par la région N-terminale de P6/TAV. Le changement de quelques acides aminés à son extrémité N-terminale est ainsi suffisant pour convertir la souche virulente Cabb-B JI, en une souche modérée chez les Brassicacées [Geldreich *et al.*, résultats non publiés].

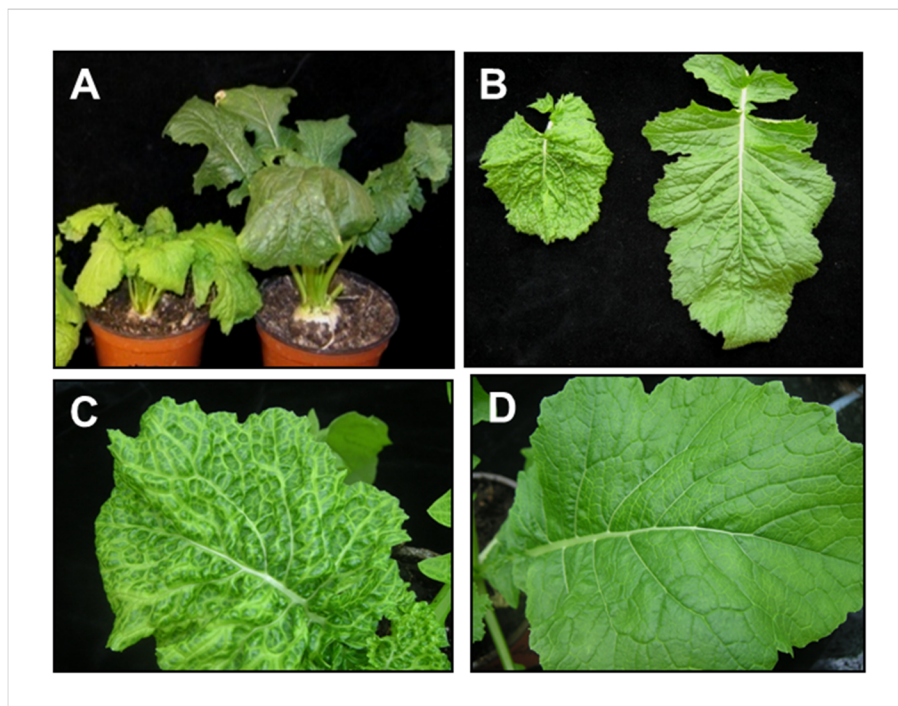


Figure 8 : Symptômes exprimés par des plants de navets infectés par le CaMV. A et B : Nanisme de la plante et des feuilles infectées, en comparaison avec la plante saine. C : Mosaïque, gauchage et éclaircissement des nervures d'une feuille infectée par la souche virulente de CaMV, Cabb B-JI. D : Mosaïque douce chez une feuille infectée par une souche peu virulente de CaMV, Cabb-S.

Le phénotype des plantes infectées est associé à une modification du transcriptome de l'hôte par P6/TAV, la dérégulation de l'expression certains gènes cellulaires présentant, d'ailleurs, des similitudes entre les plants d'*A. thaliana* transgéniques exprimant P6/TAV, et ceux infectés par le CaMV [70]. Il est vraisemblable que P6/TAV agisse indirectement sur l'expression des gènes *via* ses fonctions de supprimeur du *RNA silencing* et d'inhibiteur de la signalisation hormonale, ces deux systèmes étant partiellement voués à la régulation du développement de la plante.

Résistance, défense et contre-défense

Certaines plantes sont résistantes à l'infection virale, car elles possèdent un gène de résistance codant pour un récepteur qui reconnaît une protéine du pathogène, produite par un gène d'it d'avirulence. On parle alors de réaction incompatible, par opposition à la réaction compatible qui permet l'infection virale. Cette reconnaissance, qui obéit au concept du « gène pour gène », déclenche une réaction de défense dite d'hypersensibilité – ou HR, pour « *hypersensitive response* » – qui conduit à la nécrose de la zone infectée pour confiner le pathogène, ainsi qu'à une signalisation systémique – ou SAR, pour « *systemic acquired resistance* » – qui prémunit la plante contre toute attaque ultérieure, par n'importe quel agent pathogène [71, 72]. Lors d'une infection virale, les plantes se défendent grâce au *RNA silencing*, qui conduit à l'extinction séquence-spécifique de l'expression des gènes du pathogène, au niveau transcriptionnel – c'est le TGS, pour « *transcriptional gene silencing* »

–, par méthylation de l'ADN et des histones, et/ou au niveau post-transcriptionnel – c'est le PTGS, pour « *post-transcriptional gene silencing* » –, par clivage de l'ARN cible ou inhibition de sa traduction [73]. Les virus sont à la fois les activateurs, *via* des structures en double-brin des ARN, et la cible de la machinerie du *RNA silencing*. Une fois activé, le *RNA silencing* produit également, à son site d'initiation, un signal composé de petits ARN, qui se propage dans d'autres tissus de la plante pour les prémunir contre l'infection par le virus [74]. Toutefois, les virus de plantes, dont le CaMV, contre-carrent ce système de défense antiviral en codant pour des protéines appelées supprimeurs de *RNA silencing* [75].

Résistance des plantes au CaMV

Certaines souches de CaMV déclenchent, chez des membres du genre *Nicotiana* et certains écotypes d'*A. thaliana*, une réponse hypersensible. Cette réaction est déclenchée par l'interaction de P6/TAV – l'ORF VI est donc le gène d'avirulence – et une protéine de l'hôte codée par le gène de résistance, par exemple le gène semi-dominant *Car-1* chez *A. thaliana*. À titre d'exemple, la souche W260 élicite une HR à cause de P6/TAV, chez *Nicotiana edwardsonii* alors qu'elle infecte *N. bigelovii* et l'écotype Tsu-0 d'*A. thaliana*, lui-même résistant à la souche CM1841 [76]. Par ailleurs, des tabacs transgéniques exprimant la protéine P6/TAV de la souche CaMV-S Japon présentent des taches nécrotiques au niveau des feuilles et des tiges, et ils renferment de nombreuses protéines PR – pour « *pathogenesis-related proteins* » – à activité glucanase, dont l'expression est stimulée en cas de HR. La séquence de P6/TAV responsable de la HR est située entre les acides aminés 41 et 110 de l'extrémité N-terminale et elle se superpose au domaine qui permet au CaMV d'échapper à la résistance des plantes, le domaine RBR-1 pour « *resistance breakage region-1* » [76, 77]. La région N-terminale étant hypervariable, de simples changements de résidus suffisent donc pour dicter l'issue de l'infection des plantes par le CaMV.

RNA silencing et la contre-défense du CaMV

Les plantes infectées par le CaMV renferment des petits ARN dérivant des transcrits viraux, les vsARN – pour « *virus-derived small RNA* ». La structure secondaire de la région *leader* de l'ARN 35S est la source majeure de ces petits ARN chez *A. thaliana*. Elle est la cible de l'action coordonnée et hiérarchique des quatre enzymes DCL1 à DCL4 – pour « *Dicer-like* » – de la machinerie du *RNA silencing*, même si l'on observe une accumulation forte des siARN de 24 et de 21 nucléotides qui sont produits respectivement par DCL3 et DCL4 [45, 61] (Figure 9). Étonnamment, une recherche bioinformatique montre que de nombreux gènes d'*A. thaliana* peuvent être potentiellement ciblés par les vsARN viraux et ainsi, réprimés au cours de l'infection virale.

Le CaMV contrecarre cette défense antivirale grâce à l'activité de suppresseur de *RNA silencing* de la protéine P6/TAV (Figure 9). Cette activité est assujettie à l'importation nucléaire de P6/TAV, étant donné qu'une version délétée des deux NLS est incapable de protéger les transcrits viraux contre la dégradation par la machinerie du *RNA silencing*. P6/TAV interfère avec la voie du *RNA silencing* en inhibant DRB4, un co-facteur qui est requis pour un clivage optimal des ARN double brin par DCL4 [23]. Il a été montré très récemment que le domaine impliqué dans la suppression du *RNA silencing* est localisé dans la région N-terminale de la protéine virale [78]. P6/TAV inhibe également la production des siARN qualifiés de secondaires, dont le rôle est d'amplifier la défense antivirale de la cellule [79]. Les siARN secondaires sont générés par l'enzyme DCL4 qui clive l'ARN cible, converti, au préalable, en ARN double-brin grâce à l'ARN polymérase ARN-dépendante RDR6. L'inhibition partielle de l'activité de DCL4 à travers une interaction entre DRB4 et P6/TAV, au lieu d'une inhibition totale, est sans doute un bon compromis pour le CaMV, dans la mesure où la quantité de siARN de 21 nucléotides est suffisamment réduite pour prévenir la défense antivirale et assez élevée pour inhiber, de façon post-transcriptionnelle, l'expression des gènes cellulaires. L'inhibition de DCL4 explique aussi pourquoi les plantes infectées par le CaMV contiennent une quantité importante de siARN de 24 nucléotides DCL3-dépendants, qui ont le potentiel d'induire la formation d'hétérochromatine. P6/TAV

induirait donc également indirectement, par son activité de suppresseur du *RNA silencing*, l'extinction transcriptionnelle du minichromosome viral, ce qui tempérerait l'accumulation du CaMV et préserverait, ainsi, l'intégrité de la cellule hôte infectée. Cette hypothèse est d'autant plus plausible que l'ADN viral est méthylé quelques jours après le début de l'infection [43]. Des études récentes suggèrent que la machinerie du RNA silencing est détournée de l'ARN 35S, grâce à des ARN 8S produits à partir des molécules d'ADN virales relâchées et qui existent sous les formes sens et anti-sens. Les ARN 8S, vraisemblablement considérés comme aberrants car non-polyadénylés et probablement repliés sous la forme d'une tige-boucle semblable à la région *leader*, sont clivés en vsARN par les quatre DCL d'*A. thaliana*. Ces siARN serviraient ainsi de leurres pour les complexes effecteurs RISC, ce qui préserverait l'ARN 35S de la dégradation [45, 80].

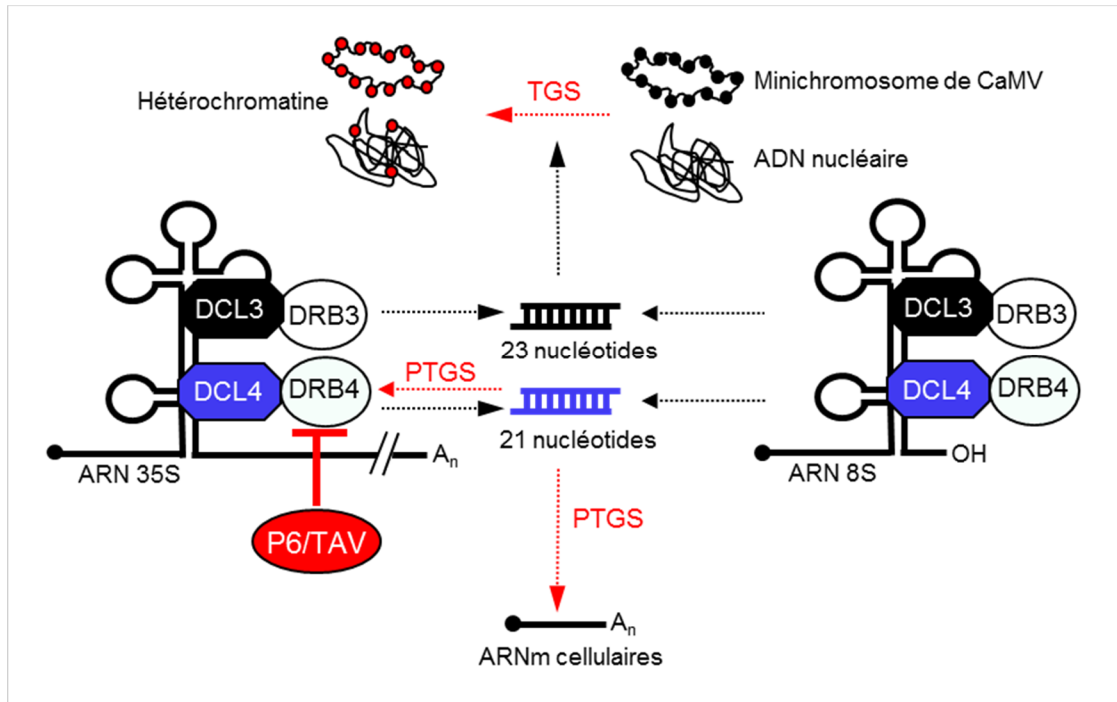


Figure 9 : RNA silencing et contre-défense du CaMV. L'ARN 35S est la cible des quatre DCL d'*A. thaliana* mais est préférentiellement clivé par DCL3 et DCL4. Le CaMV s'y oppose par la protéine P6, son suppresseur de RNA silencing et l'ARN viral 8S, produit en grande quantité, qui joue le rôle de leurre. Seul l'ARN 8S sens est représenté ici. TGS : *transcriptional gene silencing*. PTGS : *post-transcriptional gene silencing*.

La protéine P6/TAV inhibe également la biogenèse des tasiARN d'*A. thaliana* – pour « *trans-acting* siARN) » – qui implique RDR6, DCL4 et son co-facteur DRB4. Cette catégorie de siARN a pour fonction de réguler l'expression des protéines cellulaires, en ciblant les ARNm correspondants. Effectivement, les plants *A. thaliana* transgéniques exprimant P6/TAV, ou ceux infectés par le CaMV, renferment une quantité inhabituellement élevée de précurseurs de tasiARN et, de manière concomitante, un fort taux des ARNm cellulaires qui sont habituellement ciblés par ces tasiARN et clivés par DCL4. Toutefois, la présence de P6/TAV stimule, paradoxalement, l'expression des gènes codant pour DCL4, DRB4 et AGO7, impliqués dans la régulation de l'expression des ARNm par les tasiARN, ce qui suggère que la plante, en cas d'infection ou plus généralement en cas d'expression de P6/TAV, met en place un rétrocontrôle destiné à compenser l'effet inhibiteur de la protéine virale [79].

Défense non spécifique des plantes vis-à-vis du CaMV

De nombreuses plantes compatibles pour un virus déclenchent, en plus du *RNA silencing*, une réaction antivirale contrôlée par les voies de signalisation pilotées par l'acide salicylique, l'acide

jasmonique et l'éthylène. L'activation de ces voies provoque une chute de l'accumulation du virus et/ou affecte le mouvement à longue distance du virus. Ainsi, le CaMV est incapable d'infecter des mutants d'*A. thaliana* dont la voie de l'acide salicylique est activée en permanence, ce qui conduit à une résistance systémique acquise constitutive [81]. Comme le CaMV infecte systématiquement de nombreux écotypes d'*A. thaliana*, il a été supposé qu'une protéine virale devait donc supprimer la voie de signalisation de l'acide salicylique dans les cellules infectées. Les expériences réalisées avec des *A. thaliana* transgéniques montrent que P6/TAV affecte la défense antivirale basale de la plante en empêchant l'expression des gènes contrôlée par la voie de l'acide salicylique, comme la protéine PR-1. L'accumulation de P6/TAV au cours de l'infection d'*A. thaliana* par le CaMV coïncide d'ailleurs avec une baisse de l'abondance des transcrits des gènes dont l'expression est acide salicylique-dépendante. A l'opposé, la présence de P6/TAV stimule l'expression des gènes représentatifs de la voie de l'acide jasmonique, comme *VSP* – pour « *vegetative storage protein* » – et *THI2.1* – pour « *thionin 2.1* » [81]. Ceci s'explique par le fait que cette voie et celle de l'acide salicylique sont régulées d'une manière mutuellement antagoniste, à travers des mécanismes qui reposent, du moins partiellement, sur l'activité de la protéine NRP1, un important régulateur de la communication entre ces deux voies de signalisation. La protéine NRP1 monomérique active la transcription du gène *PR-1* alors que sa forme multimérique est inactive. Curieusement, NRP1 est surexprimée en présence de P6/TAV ce qui laisse supposer que cette dernière favorise la translocation nucléaire d'une forme inactive de NRP1 qui bloquerait ainsi la transcription du gène *PR-1*. En résumé, P6/TAV augmente la susceptibilité d'*A. thaliana* au CaMV en supprimant la réponse de la voie de l'acide salicylique et en augmentant celle de la voie du jasmonate [81]. D'autres protéines virales supprimeurs de *RNA silencing* sont également connues pour leur capacité à modifier l'expression des gènes dépendant de la voie de l'acide jasmonique, comme la protéine HC-Pro des *Potyvirus*, mais à l'inverse de P6/TAV, aucune d'entre elles n'est capable d'interférer avec la voie du salicylate.

L'activité de contre-défense de P6/TAV, qu'il s'agisse de la suppression du *RNA silencing* ou de celle de la défense basale des plantes, n'est aucunement liée à la fonction de transactivateur traductionnel. Cette double activité de supprimeur est portée par la région N-terminale de P6/TAV (aminoacides 1 à 112) [78]. De ce fait, il n'est pas exclu que les voies de signalisation de la plante soient affectées indirectement, par le biais des petits ARN – siARN et/ou miARN – produits par la machinerie du *RNA silencing*.

Conclusion

Le CaMV révèle des stratégies d'expression inédites que lui permettent d'exprimer ses protéines, malgré les obstacles structuraux de l'ARN 35S. Contrairement à d'autres virus, il n'inhibe pas la traduction des ARNm de la cellule hôte pour s'accaparer la machinerie traductionnelle, mais il préfère la modifier subtilement par le biais d'interactions nouvelles et de phosphorylations, afin de l'adapter à des ARNm polycistroniques. Il est probable que quelques ARNm cellulaires particuliers requièrent également une telle adaptation de la machinerie traductionnelle pour leur traduction, puisqu'il s'avère que l'activation de la protéine kinase TOR par P6/TAV mime l'activation de la kinase en réponse à l'auxine, une phytohormone importante pour le développement de la plante [82]. À l'inverse des genres de *Caulimoviridae* qui ont évolué vers la stratégie de synthèse de polyprotéine(s), la stratégie de réinitiation qui est déployée par le CaMV et les genres produisant un ARN 35S polycistronique leur permet, vraisemblablement, de réguler plus finement l'expression de leurs protéines. Étonnement, cette stratégie repose sur une seule protéine virale, la protéine P6/TAV, qui est la clé de voûte du processus infectieux du CaMV, ce qui n'est pas sans risque lorsqu'on sait que l'ADN viral est sujet à des erreurs régulières lors de la réplication par rétrotranscription, ainsi qu'à des événements de recombinaison.

Le CaMV est un modèle d'étude dont l'intérêt dépasse largement le cadre de la virologie végétale car les nombreuses connaissances acquises à son sujet peuvent être transposées au niveau cellulaire et, par conséquent, il fait partie des phytovirus dont l'apport en biologie moléculaire et cellu-

laire n'est plus à démontrer. Il pourrait aussi contribuer à mieux comprendre des mécanismes cellulaires pour lesquels on ne dispose actuellement que d'informations parcellaires, tel que l'export nucléaire des ARNm chez les plantes.

Références

- [1] Scholthof KB, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, Hohn B, Saunders K, Candresse T, Ahlquist P, Hemenway C, Foster GD. Top 10 plant viruses in molecular plant pathology. *Mol Plant Pathol* 2011 ; 12 : 938-954.
- [2] Haas M, Bureau M, Geldreich A, Yot P, Keller M. Cauliflower mosaic virus: still in the news. *Mol Plant Pathol* 2002 ; 3 : 419-429.
- [3] Hoh F, Uzest M, Drucker M, Plisson-Chastang C, Bron P, Blanc S, Dumas C. Structural insights into the molecular mechanisms of cauliflower mosaic virus transmission by its insect vector. *J Virol* 2010 ; 84 : 4706-4713.
- [4] Franck A, Guilley H, Jonard G, Richards K, Hirth, L. Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* 1980 ; 21 : 285-294.
- [5] Perbal MC, Thomas CL, Maule AJ. Cauliflower mosaic virus gene I product (P1) forms tubular structures which extend from the surface of infected protoplasts. *Virology* 1993 ; 195 : 281-285.
- [6] Huang Z, Andrianov V, Han Y, HowellbS. Identification of Arabidopsis proteins that interact with the cauliflower mosaic virus (CaMV) movement protein. *Plant Mol Biol* 2001 ; 47 : 663-675.
- [7] Thomas CL, Maule AJ. Identification of structural domains within the cauliflower mosaic virus movement protein by scanning deletion mutagenesis and epitope tagging. *Plant Cell* 1995 ; 7 : 561-572.
- [8] Sánchez-Navarro J, Fajardo T, Zicca S, Pallàs V, Stavelone L. *Caulimoviridae* Tubule-Guided Transport Is Dictated by Movement Protein Properties. *J Virol* 2010 ; 84 : 4109-4112.
- [9] Schoelz JE, Harries PA, Nelson RS. Intracellular transport of plant viruses: finding the door out of the cell. *Mol Plant* 2011 ; 4 : 813-831.
- [10] Uzest M, Drucker M, Blanc S. La transmission d'un complexe : pas si simple. Cas du virus de la mosaïque du chou-fleur. *Virologie* 2011 ; 15 : 192-204.
- [11] Khelifa M, Journou S, Krishnan K, Gargani D, Espérandieu P, Blanc S, Drucker M. Electron-lucent inclusion bodies are structures specialized for aphid transmission of Cauliflower mosaic virus. *J Gen Virol* 2007 ; 10 : 2872-2880.
- [12] Martinière A, Gargani D, Uzest M, Lautredou N, Blanc S, DruckerM. A role for plant microtubules in the formation of transmission-specific inclusion bodies of cauliflower mosaic virus. *Plant J* 2009 ; 58 : 135-146.
- [13] Leh V, Jacquot E, Geldreich A, Hermann T, Leclerc D, Cerutti M, Yot P, Keller M, Blanc S. Aphid-transmission of cauliflower mosaic virus requires the viral PIII protein. *EMBO J* 1999 ; 18 : 7077-7085.
- [14] Stavelone L, Villani ME, Leclerc D, Hohn T. A coiled-coil interaction mediates Cauliflower mosaic virus cell-to-cell movement. *Proc Natl Acad Sci USA* 2005 ; 102 : 6219-6224.
- [15] Lutz L, Raikhy G, Leisner SM. Cauliflower mosaic virus major inclusion body protein interacts with the aphid transmission factor, the virion-associated protein, and gene VII product. *Virus Res* 2012 ; 170 :150-153.
- [16] Jacquot E, Geldreich A, Keller M, Yot P. Mapping regions of the cauliflower mosaic virus ORF III product required for infectivity. *Virology* 1998 ; 242 : 395-402.
- [17] Karsies A, Hohn T, Leclerc D. Degradation signals within both terminal domains of the cauliflower mosaic virus capsid protein precursor. *Plant J* 2001 ; 27 : 335-343.
- [18] Karsies A, Merkle T, Szurek B, Bonas U, Hohn T, Leclerc D. Regulated nuclear targeting of cauliflower mosaic virus. *J Gen Virol* 2002 ; 83 : 1783-1790.

- [19] Torruella M, Gordon K, Hohn T. Cauliflower mosaic virus produces an aspartic proteinase to cleave its polyproteins. *EMBO J* 1989 ; 8 : 2819-1825.
- [20] Iskra-Caruana ML, Lheureux F, Teycheney PY. Les pararétrovirus endogènes (EPRV), voie nouvelle de transmission des virus de plantes. *Virologie* 2003 ; 7 : 255-265.
- [21] Hohn T, Rothnie H. Plant pararetroviruses : replication and expression. *Curr Opin Virology* 2013 ; 3 : 621-628.
- [22] Haas M, Geldreich A, Bureau M, Dupuis L, Leh V, Vetter G, Kobayashi K, Hohn T, Ryabova L, Yot P, Keller M. The open reading frame VI product of Cauliflower mosaic virus is a nucleocytoplasmic protein: its N terminus mediates its nuclear export and formation of electron-dense viroplasms. *Plant Cell* 2005 ; 17 : 927-943.
- [23] Haas G, Azevedo J, Moissiard G, Geldreich A, Himber C, Bureau M, Fukuhara T, Keller M, Voinnet O. Nuclear import of CaMV P6 is required for infection and suppression of the RNA silencing factor DRB4. *EMBO J* 2008 ; 27 : 2102-2012.
- [24] Bak A, Blanc S, Gargani D, Martinière A, Drucker M. Multiples fonctions des usines virales : l'exemple du virus de la mosaïque du chou-fleur. *Virologie* 2014 ; 4 : 201-210.
- [25] Shalla TA, Shepherd RJ, Petersen LJ. Comparative cytology of nine isolates of Cauliflower mosaic virus. *Virology* 1980 ; 102 : 381-388.
- [26] Li Y, Leisner SM. Multiple domains within the Cauliflower mosaic virus gene VI product interact with the full-length protein. *Mol Plant Microbe Interact*. 2002 ; 10 : 1050-1057.
- [27] Cecchini E, Gong Z, Geri C, Covey SN, Milner JJ. Transgenic Arabidopsis lines expressing gene VI from cauliflower mosaic virus variants exhibit a range of symptom-like phenotypes and accumulate inclusion bodies. *Mol Plant Microbe Interact* 1997 ; 10 : 1094-1101.
- [28] Netherton C, Moffat K, Brooks E, Wileman T. A guide to viral inclusions, membrane rearrangements, factories, and viroplasm produced during virus replication. *Adv Virus Res* 2007 ; 70 : 101-182.
- [29] Schoelz JE, Shepherd RJ, Daubert SD. Gene VI of CaMV encodes a host range determinant. *J Mol Cell Biol* 1986 ; 6 : 2632-2637.
- [30] Wintermantel WM, Anderson EJ, Schoelz JE. Identification of domains within gene VI of cauliflower mosaic virus that influence systemic infection of *Nicotiana bigelovii* in a light dependent manner. *Virology* 1993 ; 196 : 789-798.
- [31] Leisner SM, Howell SH. Symptom variation in different Arabidopsis thaliana ecotypes produced by cauliflower mosaic virus. *Phytopathology* 1992 ; 82 : 1042-1046.
- [32] Yu W, Murfett J, Schoelz JE. Differential Induction of symptoms in *Arabidopsis* by P6 of *Cauliflower mosaic virus*. *Mol Plant Microbe Interact* 2003 ; 16 : 35-42.
- [33] Love AJ, Laird J, Holt J, Hamilton AJ, Sadanandom A, Milner JJ. Cauliflower mosaic virus protein P6 is a suppressor of RNA silencing. *J Gen Virol* 2007 ; 88 : 3439-3444.
- [34] Shivaprasad PV, Rajeswaran R, Blevins T, Schoelz J, Meins F Jr, Hohn T, Pooggin MM. The CaMV transactivator/viroplasm interferes with RDR6-dependent trans-acting a,d secondary siRNA pathways in Arabidopsis. *Nucleic Acids Res* 2008 ; 36 : 5896-5909.
- [35] Kiraly L, Cole AB, Bourque JE, Schoelz JE. Systemic cell death is elicited by the interaction of a single gene in *Nicotiana clevelandii* and gene VI of cauliflower mosaic virus. *Mol Plant Microbe Interact* 1999 ; 12 : 919-925.
- [36] Ryabova LA, Pooggin MM, Hohn T. Viral strategies of translation initiation: ribosomal shunt and reinitiation. *Prog Nucleic Acid Res Mol Biol* 2002 ; 72 : 1-39.

- [37] Himmelbach A, Chapelaine Y, Hohn T. Interaction between cauliflower mosaic virus inclusion body protein and capsid protein: implication for viral assembly. *Virology* 1996 ; 217 : 147-157.
- [38] Palanichelvam K, Schoelz JE. A comparative analysis of the avirulence and translational transactivator functions of gene VI of Cauliflower mosaic virus. *Virology* 2002 ; 293 : 225-233.
- [39] Laird J, McNally C, Carr C, Doddiah S, Yates G, Chrysanthou E, Khattab A, Love AJ, Geri C, Sadanandom A, Smith BO, Kobayashi K, Milner JJ. Identification of the domains of cauliflower mosaic virus protein P6 responsible for suppression of RNA silencing and salicylic acid signalling. *J Gen Virol* 2013 ; 4 : 2777-2789.
- [40] Schepetilnikov M, Ryabova L. Cauliflower mosaic virus (CaMV) upregulates translation reinitiation of its pregenomic polycistronic 35S RNA via interactions with the cell's translation machinery. *Plant virus–Host interactions*. Boston : Academic Press, 2014 : 325-343.
- [41] Angel CA, Lutz L, Yang X, Rodriguez A, Adair A, Zhang Y, Leisner SM, Nelson RS, Schoelz JE. The P6 protein of Cauliflower mosaic virus interacts with CHUP1, a plant protein which moves on actin microfilaments *Virology* 2013 ; 443 :363-374.
- [42] Cerritelli S, Fedoroff O, Reid B, Crouch R. A common 40 amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs. *Nucl Acids Res* 1998 ; 26 : 1834-1840.
- [43] Tang W, Leisner S. Methylation of nonintegrated multiple copy DNA in plants. *Biochem Biophys Res Commun* 1998 ; 245 : 403-406.
- [44] Blanc S, Uzest M, Drucker M. New research horizons in vector-transmission of plant viruses. *Curr Opin Microbiol* 2011 ; 14 : 483-491.
- [45] Blevins T, Rajeswaran R, Aregger M, Borah BK, Schepetilnikov M, Baerlocher L, Farinelli L, Meins F Jr, Hohn T, Pooggin MM. Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. *Nucleic Acids Res* 2011 ; 39 : 5003-5014.
- [46] Plant AL, Covey SN, Grierson D. Detection of a sub-genomic mRNA for gene V, the putative reverse transcriptase gene of cauliflower mosaic virus. *Nucl Acids Res* 1985 ; 13 : 8305-8321.
- [47] Benfey PN, Chua NH. The cauliflower mosaic virus 35S promoter : combinatorial regulation of transcription in plants. *Science* 1990 ; 250 : 959-966.
- [48] Lam, E. Chua, N.H. ASF-2: a factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in Cab promoters. *Plant Cell* 1989 ; 1 : 1147-1156.
- [49] Jupin I, Chua NH. Activation of the CaMV as-1 cis-element by salicylic acid: differential DNA-binding of a factor related to TGA1a. *EMBO J* 1996 ; 15 : 5679-5689.
- [50] Cui X, Fan B, Scholz J, Chen Z. Roles of Arabidopsis cyclin-dependent kinase C complexes in cauliflower mosaic virus infection, plant growth, and development. *Plant Cell* 2007 ; 19 : 1388-1402.
- [51] Battisti PL, Daher A, Bannwarth S, Voortman J, Peden KW, Hiscott J, Moulard AJ, Benarous R, Gagnon A. Additive activity between the trans-activation response RNA-binding protein, TRBP2, and cyclin T1 on HIV type 1 expression and viral production in murine cells. *AIDS Res Hum Retroviruses* 2003 ; 19 : 767-778.
- [52] Sanfaçon H, Brodmann P, Hohn T. A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev* 1991 ; 5 : 141-149.
- [53] Kiss-László Z, Blanc S, Hohn T. Splicing of cauliflower mosaic virus 35S RNA is essential for viral infectivity. *EMBO J* 1995 ; 14 : 3552-3562.
- [54] Froissart R, Uzest M, Ruiz-Ferrer V, Drucker M, Hébrard E, Hohn T, Blanc S. Splicing of Cauliflower mosaic virus 35S RNA serves to downregulate a toxic gene product. *J Gen Virol* 2004 ; 85 : 2719-2726.

- [55] Hirochika H, Takatsuji H, Ubasawa A, Ikeda JE. Site-specific deletion in cauliflower mosaic virus DNA: possible involvement of RNA splicing and reverse transcription. *EMBO J* 1985 ; 4 :1673-1680.
- [56] Olszewski NE, Guilfoyle TJ. Nuclei purified from cauliflower mosaic virus-infected turnip leaves contain subgenomic, covalently closed circular cauliflower mosaic virus DNAs. *Nucleic Acids Res* 1983 ; 11 : 8901-8914.
- [57] Sommer G, Heise T. Posttranscriptional control of HBV gene expression. *Front Biosci* 2008 ; 13 : 5533-5547.
- [58] Bayliss J, Lim L, Thompson AJV, Desmond P, Angus P, Locarmini S, Revill A. Hepatitis B splicing is enhanced prior to development of hepatocellular carcinoma. *J Hepat* 2013 ; 59 : 1022-1028.
- [59] Pooggin MM, Fütterer J, Skryabin KG, Hohn T. A short open reading terminating in front of a stable hairpin is the conserved feature in pregenomic RNA leaders of plant pararetroviruses. *J Gen Virol* 1999 ; 80 : 2217-2228.
- [60] Hohn T, Park HS, Guerra-Peraza O, Stavolone L, Pooggin MM, Kobayashi K, Ryabova LA. Shunting and controlled reinitiation: the encounter of cauliflower mosaic virus with the translational machinery. *Cold Spring Harb Symp Quant Biol* 2002 ; 66 : 269-276.
- [61] Moissiard G, Voinnet O. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. *Proc Natl Acad Sci U S A* 2006 ; 103 : 19593-19598.
- [62] Thiébeauld O, Schepetilnikov M, Park HS, Geldreich A, Kobayashi K, Keller M, Hohn T, Ryabova LA. A new plant protein interacts with eIF3 and 60S to enhance virus-activated translation re-initiation. *EMBO J* 2009 ; 28 : 3171-3184.
- [63] Schepetilnikov M, Kobayashi K, Geldreich A, Caranta C, Robaglia C, Keller M, Ryabova LA. Viral factor TAV recruits TOR/S6K1 signalling to activate reinitiation after long ORF translation. *EMBO J* 2011 ; 30 : 1343-1356.
- [64] Pfeiffer P, Hohn T. Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* 1983 ; 33 : 781-789.
- [65] Carluccio AV, Zicca S, Stavolone L. Hitching a ride on vesicles: cauliflower mosaic virus movement protein trafficking in the endomembrane system. *Plant Physiol* 2014 ; 164 : 1261-1270.
- [66] Amari K, Boutant E, Hofmann C, Schmitt-Keichinger C, Fernandez-Calvino L, Didier P, Lerich A, Mutterer J, Thomas CL, Heinlein M, Mély Y, Maule AJ, Ritzenthaler C. A family of plasmodesmal proteins with receptor-like properties for plant viral movement proteins. *PLoS Pathog* 2010 23 ; 6 : e1001119.
- [67] Harries PA, Palanichelvam K, Yu W, Schoelz JE, Nelson RS. The cauliflower mosaic virus protein P6 forms motile inclusions that traffic along actin microfilaments and stabilize microtubules. *Plant Physiol* 2009 ; 149 : 1005-1016.
- [68] Qiu SG, Schoelz JE. Three regions of cauliflower mosaic virus strain W260 are involved in systemic infection of solanaceous hosts. *Virology* 1992 ; 190 : 773-782.
- [69] Daubert S D, Schoelz J E, Debaio L, Shepherd RJ. Expression of disease symptoms in cauliflower mosaic virus genomic hybrids. *J Mol Appl Genet* 1984 ; 2 : 537-547.
- [70] Geri C, Cecchini E, Giannakou M, Covey S, Milner J. Altered patterns of gene expression in Arabidopsis elicited by cauliflower mosaic virus (CaMV) infection and by a CaMV gene VI transgene. *Mol Plant Microbe Interact* 1999 ; 12 : 377-384.
- [71] Nimchuk Z, Eulgem T, Holt BF, Dangl JL. Recognition and response in the plant immune system. *Annu Rev Genet* 2003 ; 37 : 579-609.

- [72] Zvereva AS, Pooggin M. Silencing and innate immunity in plant defense against viral and non-viral pathogens. *Viruses* 2012 ; 4 : 2578-2597.
- [73] Pumplin N, Voinnet O. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nat Rev Microbiol* 2013 ; 11 : 745-760.
- [74] Brosnan CA, Voinnet O. Cell-to-cell and long-distance siRNA movement in plants: mechanisms and biological implications. *Curr Opin Plant Biol* 2011 ; 14 : 580-587.
- [75] Burgyan J, Havelda Z. Viral suppressors of RNA silencing. *Trends in Plant Sci* 2011 ; 16 : 265-272.
- [76] Agama K, Beach S, Schoelz J, Leisner SM. The 5'-third of cauliflower mosaic virus gene VI conditions resistance breakage in *Arabidopsis* ecotype Tsu-0. *Phytopathology* 2002 ; 92 : 190-196.
- [77] Kobayashi K, Hohn T. The avirulence domain of *Cauliflower mosaic virus* transactivator/viroplasmin is a determinant of viral virulence in susceptible hosts. *Mol Plant Microbe Interact* 2004 ; 17 : 475-483.
- [78] Laird J, McNally C, Carr C, Doddiah S, Yates G, Chrysanthou E, Khattab A, Love AJ, Geri C, Sadanandom A, Smith BO, Kobayashi K, Milner JJ. Identification of the domains of cauliflower mosaic virus protein P6 responsible for suppression of RNA silencing and salicylic acid signalling. *J Gen Virol* 2013 ; 94 : 2777-2789.
- [79] Shivaprasad PV, Rajeswaran R, Blevins T, Schoelz J, Meins F Jr, Hohn T, Pooggin MM. The CaMV transactivator/viroplasmin interferes with RDR6-dependent trans-acting and secondary siRNA pathways in *Arabidopsis*. *Nucleic Acids Res* 2008 ; 36 : 5896-5909.
- [80] Hohn T, Vazquez, F. RNA silencing pathways of plants: silencing and its suppression by plant DNA viruses, *Biochim Biophys Acta* 2012 ; 11 : 588-600.
- [81] Love AJ, Geri C, Laird J, Carr C, Yun BW, Loake GJ, Tada Y, Sadanandom A, Milner JJ. Cauliflower mosaic virus protein P6 inhibits signaling responses to salicylic acid and regulates innate immunity. *PLoS One* 2012 ; 7 : e47535.
- [82] Schepetilnikov M, Dimitrova M, Mancera-Martínez E, Geldreich A, Keller M, Ryabova LA. TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h. *EMBO J* 2013 ; 32 : 1087-1102.

Chapitre 1

Étude de l'épissage alternatif de l'ARN 35S du CaMV

INTRODUCTION

1) Aperçu général sur l'épissage

a) Introduction

L'épissage a été décrit simultanément à la fin des années 1970, dans deux articles portant sur les ARN messagers (ARNm) produits par un adénovirus, tardivement au cours de l'infection (Berget et al., 1977; Chow et al., 1977). Pour ces travaux, Phillip Allen Sharp et Richard John Roberts ont reçu le prix Nobel de Physiologie ou Médecine, en 1993.

La première édition de l'*Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, publiée par Diderot et D'Alembert au cours de la seconde moitié du 18^{ème} siècle, définit l'action d'épisser une corde : « ...c'est l'assembler avec une autre, en entrelassant (*sic*) leurs fils ou cordons l'un avec l'autre, ce qui se fait par le moyen d'une broche de fer appelée *cornet d'épisse* ou *épissoir*. Après un combat, lorsque quelques manœuvres sont coupées ou rompues, on est obligé de les *épisser* quand on n'en a pas de rechange. ».

Par analogie au vocabulaire de marine, l'épissage – ou *splicing* – définit, en biologie, un processus retrouvé dans toutes les branches de l'arbre du vivant, par lequel une séquence intervenante, l'intron, est retirée d'un transcrit primaire pour générer un ARN mature. Les séquences qui sont présentes dans l'ARN mature sont quant à elles nommées exons. Les introns peuvent être classés en quatre types, selon leur mécanisme d'épissage : 1) introns de groupe I, 2) introns de groupe II, 3) introns d'ARN pré-messagers (pré-ARNm) nucléaires ou introns splicéosomiaux, 4) introns d'ARN de transfert (ARNt) nucléaires et introns d'archées (Haugen et al., 2005).

b) Les différents types d'introns et leur mécanisme d'épissage

i. Introns de groupe I

D'une longueur comprise entre 250 et 500 nucléotides (nts), les introns de groupe I sont présents chez les eubactéries et les eucaryotes. Bien qu'ils soient rares, les introns de groupe I des eubactéries ont été décrits dans des gènes codant pour des ARNt, des ARN ribosomiques (ARNr) et des protéines. Retrouvés majoritairement chez les champignons, les plantes, les algues rouges et les algues vertes, les introns de groupe I des eucaryotes semblent restreints aux gènes d'ARNr lorsqu'ils sont nucléaires, tandis qu'ils peuvent être présents dans les gènes plastidiques et mitochondriaux codant des ARNt, des ARNr et des protéines (Haugen et al., 2005). Enfin, des introns de groupe I sont retrouvés dans le génome de bactériophages et certains virus géants (Haugen et al., 2005; Van Etten, 2003).

Les introns de groupe I ont des structures secondaire et tertiaire fortes – parfois stabilisées par des protéines – qui leur permet de s'auto-épisser grâce une activité ribozyme. Certains introns de groupe I possèdent, en plus, un gène codant pour une endonucléase dite de *homing* qui permet leur propagation vers des séquences homologues à celles dans lesquelles ils sont insérés. L'auto-épissage se fait *via* deux réactions de transestérification successives qui sont décrites dans la figure 1. Ce processus nécessite une guanosine exogène qui va se lier au cœur catalytique de l'intron, ainsi que des cations divalents – principalement du Mg^{2+} (Paquin et Shub, 2001).

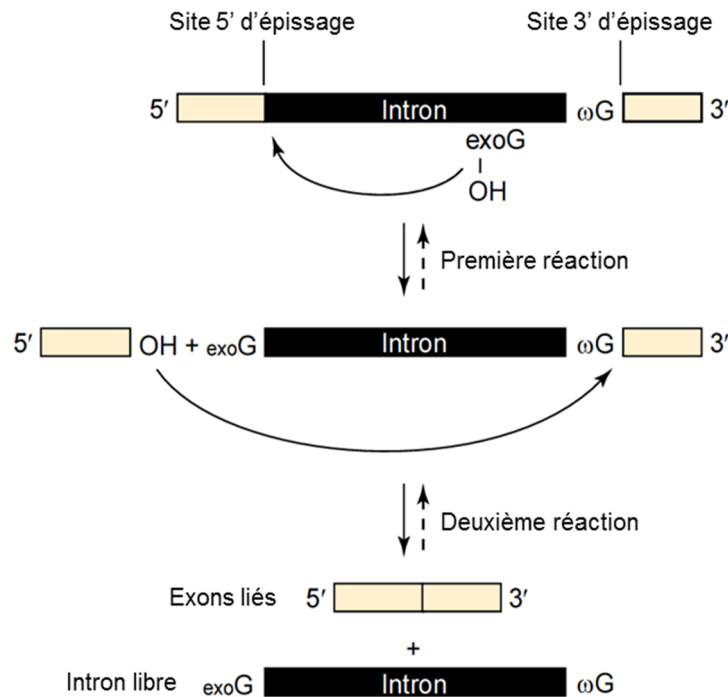


Figure 1 : Auto-épissage d'un intron de groupe I. La première réaction de transestérification correspond à l'attaque nucléophile d'une guanosine exogène (exoG) sur le nucléotide correspondant à l'extrémité 5' de l'intron, nommé site 5' d'épissage. La guanosine se lie de façon covalente à l'extrémité 5' de l'intron, ce qui libère l'exon en situé en amont. Lors de la deuxième réaction, le groupement 3'-hydroxyle (3'-OH) de l'exon en amont réalise une attaque nucléophile sur le site 3' d'épissage, entraînant ainsi la ligation des exons et la libération de l'intron. Le dernier nucléotide de l'intron est toujours une guanosine (ω G). D'après Haugen et al., 2005.

ii. Introns de groupe II

Les introns de groupe II sont, comme les introns de groupe I, des ribozymes. Leur taille est comprise entre 400 et 800 nucléotides et ils ont été caractérisés chez les eubactéries ainsi que dans les génomes mitochondriaux et chloroplastiques de plantes, de champignons et de protistes (Lambowitz et Zimmerly, 2011). La séquence nucléotidique des introns de groupe II est globalement variable, à l'exception des extrémités 5' et 3' dont les séquences conservées sont, respectivement, GUGYG et AY (où U est une uridine et Y est une pyrimidine). À l'inverse, ils possèdent une structure secondaire conservée qui consiste en six domaines, nommés DI à DVI. Ces domaines interagissent dans les trois dimensions pour former une structure fonctionnelle dont le cœur du site actif est situé dans le domaine DV. La plupart des introns de groupe II des eubactéries et la moitié de ceux des eucaryotes ont, au niveau de leur domaine DIV, un cadre ouvert de lecture (ORF) codant pour une protéine multifonctionnelle de type transcriptase inverse, l'IEP – pour « *intron encoded protein* » –, qui est impliquée dans le processus d'épissage – en stabilisant le site catalytique de l'intron par une activité maturase – et dans la propagation de l'intron au sein du génome. Comme les introns de groupe I, les introns de groupe II sont donc des éléments mobiles (Fedorova et Zingler, 2007; Lambowitz et Zimmerly, 2011).

L'épissage des introns de groupe II est effectué grâce à deux réactions de transestérification réversibles nécessitant des cations Mg^{2+} et des protéines stabilisatrices telle l'IEP. Les mécanismes réactionnels sont décrits dans la figure 2.

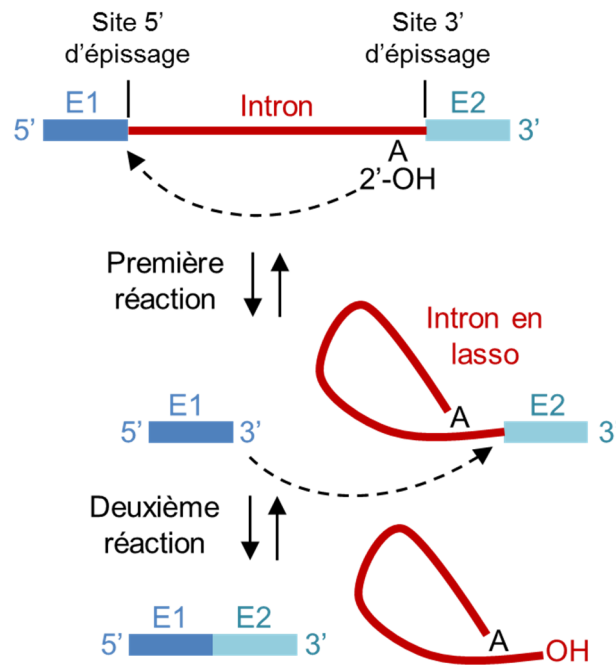


Figure 2 : Auto-épissage d'un intron de groupe II. La première réaction correspond à une attaque nucléophile du groupement 2'-hydroxyle (2'-OH) du nucléotide – généralement une adénosine – nommé point de branchement et situé dans le domaine DVI, sur le site 5' d'épissage. Ceci aboutit à la libération de l'exon en 5' (E1) et à la formation d'un intermédiaire où l'intron, dit en lasso car circularisé par une liaison 2'-5' phosphate, est lié au deuxième exon (E2), en 3'. La deuxième réaction est provoquée par l'attaque nucléophile du groupement 3'-OH de l'exon E1, sur le site 3' d'épissage, qui correspond au dernier nucléotide de l'intron. Les deux exons sont alors liés l'un à l'autre et l'intron en lasso, lié à l'IEP – non représentée sur ce schéma –, est excisé. D'après Lambowitz et Zimmerly, 2011.

La plupart des introns de groupe II eucaryotes ont des défauts, dans leur structure et/ou dans leur ORF, qui entravent leur activité auto-catalytique. L'épissage de ces introns défectueux ne peut alors se faire qu'en présence de protéines – autres que l'IEP – et parfois d'ARN, qui vont jouer le rôle de facteurs d'épissage ou de chaperons. Suite à certains réarrangements génomiques, les introns peuvent être séparés en plusieurs fragments. Ces fragments peuvent alors se réassocier dans l'espace et reformer un ribozyme ce qui va induire l'épissage d'exons situés sur deux transcrits différents. C'est le *trans*-épissage (Lambowitz et Zimmerly, 2011).

Lorsque l'IEP codée par l'intron est pleinement fonctionnelle – ce qui est rarement le cas chez les introns eucaryotes –, elle va assister l'épissage par une activité maturase et permettre la mobilité de l'intron dans le génome, par épissage inverse. Le mécanisme principal d'épissage inverse est appelé « *retrohoming* ». L'ensemble formé par l'intron en lasso et l'IEP va reconnaître une séquence cible d'ADN où l'intron va s'apparier et être rétrotranscrit en ADN grâce à l'activité transcriptase inverse de l'IEP. La séquence cible est reconnue par le biais d'appariements nucléotidiques entre l'intron et l'ADN qui sont identiques à certains appariements formés lors du processus d'épissage. Cela assure ainsi l'insertion de l'intron dans une zone à partir de laquelle il pourra s'auto-épisser ultérieurement (Lambowitz et Zimmerly, 2011).

iii. Introns de pré-ARNm nucléaires ou introns splicéosomaux

Les introns de pré-ARNm nucléaires ne sont retrouvés que chez les eucaryotes et chez certains virus. Leur taille varie de quelques dizaines à plusieurs milliers de nucléotides et leur nombre va d'une dizaine par génome, chez certains protozoaires, à une dizaine par gène, dans le cas de certains métazoaires (Roy et Irimia, 2009). Contrairement aux introns des groupes I et II, les introns splicéosomaux ne sont pas auto-catalytiques. Leur épissage – une suite de deux réactions de transestérifica-

tion – est catalysé par une machinerie ribonucléoprotéique très complexe, le spliceosome. Ce processus sera décrit plus en détail dans la partie 2)c).

iv. Introns d'ARNt nucléaires et introns d'archées

Les introns d'ARNt nucléaires, ainsi que les introns d'archées, ont pour caractéristique commune de n'être épissés que par des protéines. D'une petite taille comprise entre une quinzaine et 100 nucléotides, sans homologie de séquence, ces introns sont présents en nombre chez les archées puisqu'ils sont retrouvés dans environ 70% des gènes d'ARNt ainsi que chez certains gènes d'ARNr et ARNm. Chez les eucaryotes, seuls 6 % des gènes d'ARNt humains et 20% des gènes d'ARNt de levure sont interrompus par des introns (Calvin et Li, 2008; Popow et al., 2012).

L'épissage se fait en deux étapes chez les archées, trois chez les eucaryotes. Lors de la première étape, qui est universelle, l'intron est excisé par une endonucléase d'épissage (Figure 3). Il est à noter que l'extrémité 5' de l'intron se situe, à chaque fois, au niveau du deuxième nucléotide situé en aval de l'anticodon. Les deux exons sont ensuite reliés ensemble par une ARN-ligase, lors de la deuxième étape. Chez les eucaryotes, une troisième étape implique le retrait d'un groupement 2'-phosphate qui était présent à la jonction entre les deux exons, suite à la ligation (Calvin et Li, 2008). L'épissage des introns d'ARNt d'archées est parfois réalisé *in trans*, entre deux voire trois transcrits (Popow et al., 2012).

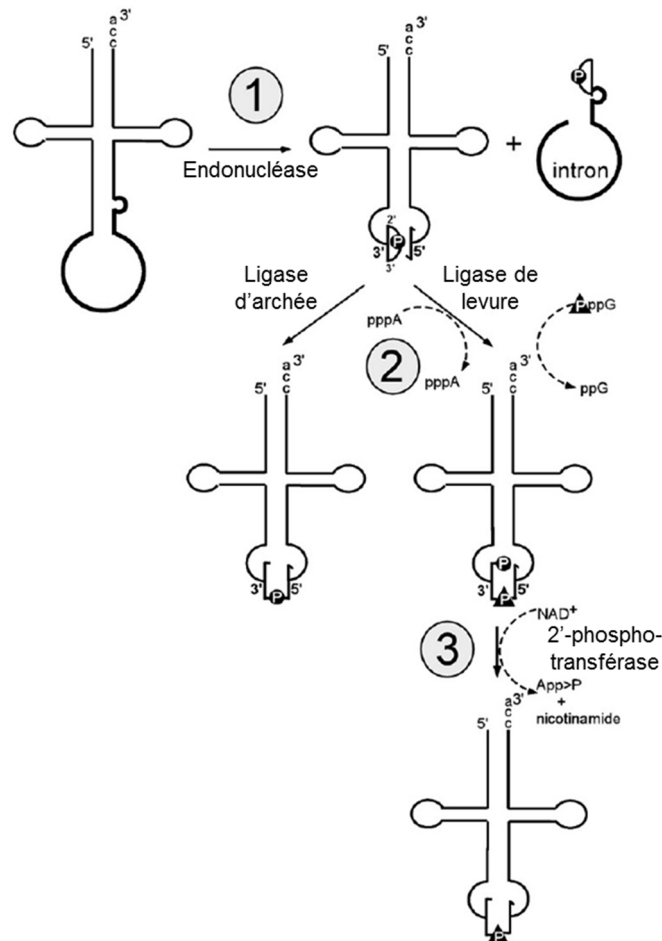


Figure 3 : Épissage enzymatique des introns d'ARNt nucléaires et des introns d'archées. (1) La première étape d'épissage, qui est universelle, correspond à l'excision de l'intron par une endonucléase. (2) Lors de la deuxième étape, les deux exons sont reliés l'un à l'autre par une ARN ligase qui diffère entre les archées et les eucaryotes (ici, la levure). (3) La troisième étape est spécifique aux eucaryotes et elle correspond au retrait d'un groupement 2'-phosphate, présent à la jonction des deux introns, par une 2'-phosphotransférase. D'après Calvin et Li, 2008.

2) Épissage des introns splicéosomaux de type U2

Les introns splicéosomaux sont des séquences intervenantes qui sont propres aux pré-ARNm nucléaires des eucaryotes. De taille et d'abondance variables selon les organismes, ces introns sont épissés, souvent de façon co-transcriptionnelle (Han et al., 2011), par un ensemble de particules ribonucléoprotéiques, le splicéosome. Deux splicéosomes distincts se partagent inégalement l'épissage des introns splicéosomaux : 1) le splicéosome majeur – ou splicéosome U2 – qui épisse 99% des introns, dits de type U2, et 2) le splicéosome mineur – ou splicéosome U12 – qui est responsable de l'épissage des 1% d'introns restant, dits de type U12 et qui ne sont présents que chez certaines espèces (Roy et Irimia, 2009).

a) Introns splicéosomaux de type U2

i. Caractéristiques générales

Le nombre et la taille des introns sont très variables en fonction des organismes (Tableau 1). La levure *Saccharomyces cerevisiae*, par exemple, ne possède que 0,05 intron par gène alors que presque tous les gènes de *Schizosaccharomyces pombe* contiennent 1 intron. Les introns de type U2 sont beaucoup plus nombreux chez les plantes *Arabidopsis thaliana* et *Oryza sativa* (4 introns par gène dans les deux cas) et chez *Homo sapiens* (presque 8 introns par gène). La taille moyenne d'un intron de *Saccharomyces cerevisiae* ou d'*Arabidopsis thaliana* (respectivement 250 et 170 nucléotides) est beaucoup plus courte que celle d'un intron humain (3000 nucléotides).

Organisme	Nombre de gènes de protéines	Intron/gène	Taille moyenne d'un intron (nucléotides)	Références
<i>Saccharomyces cerevisiae</i>	~5800	~0,05	~250	(Kellis et al., 2003; Kupfer et al., 2004)
<i>Schizosaccharomyces pombe</i>	~4900	~0,96	~100	(Kupfer et al., 2004)
<i>Homo sapiens</i>	~20800	~7,8	~3000	(Flicek et al., 2014; Reddy, 2007)
<i>Arabidopsis thaliana</i>	~27400	~4	~170	(Reddy, 2007)
<i>Oryza sativa</i>	~56000	~4	~430	(Ouyang et al., 2007; Reddy, 2007)

Tableau 1 : Ratio intron/gène et taille moyenne des introns de divers organismes modèles

Les introns de type U2 sont définis par la présence de quatre courtes séquences consensus : 1) le site 5' d'épissage – ou site donneur – qui contient un dinucléotide GU conservé et qui délimite la bordure 5' de l'intron, 2) le site 3' d'épissage – ou site accepteur – où l'on retrouve un dinucléotide AG conservé qui marque l'extrémité 3' de l'intron, 3) une courte séquence située entre 15 et 50 nucléotides en amont du site 3' d'épissage et contenant un point de branchement – une adénosine –, 4) une région riche en pyrimidines située entre le point de branchement et le site accepteur d'épissage (Figure 4). Bien que nécessaires pour assurer un épissage précis, les quatre séquences consensus ne sont pas suffisantes pour assurer l'assemblage du splicéosome. De nombreuses et courtes séquences régulatrices – 4 à 18 nucléotides de long – situées dans l'intron et dans les exons adjacents permettent le recrutement d'un ensemble de protéines facteurs d'épissage qui vont réguler posi-

vement ou négativement l'assemblage des composants du spliceosome. Ces séquences sont appelées « *splicing enhancer* » ou « *splicing silencer* », selon qu'elles recrutent des protéines qui assistent ou empêchent l'assemblage du spliceosome et l'épissage subséquent de l'intron (De Conti et al., 2013; Reddy, 2007). Bien qu'impliqué dans l'épissage des introns de type U2 en général (De Conti et al., 2013), leur rôle sera décrit plus en détail dans la partie 3)c)i. consacrée à l'épissage alternatif.

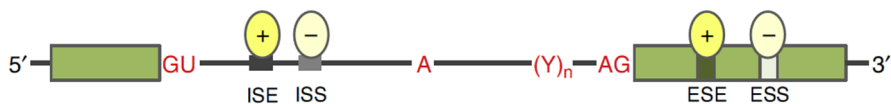


Figure 4 : Représentation schématique d'un intron (ligne noire) entouré de deux exons (rectangles verts). Le dinucléotide GU correspond au site 5' donneur d'épissage et le dinucléotide AG correspond au site 3' accepteur d'épissage. L'adénosine du point de branchement (A) est séparée du site 3' d'épissage par une séquence riche en pyrimidines (Y)_n et longue d'environ 15 à 50 nucléotides. Les ESE et ISE – pour « *exonic et intronic splicing enhancers* » – correspondent courtes séquences présentes, respectivement, dans les exons et les introns et dont la caractéristique est de stimuler l'assemblage du spliceosome. Les ESS et ISS – pour « *exonic et intronic splicing silencers* » – ont une action inverse à celle des ESE et ISE. D'après Syed et al., 2012.

ii. Particularités des introns de type U2 de plantes

Les introns U2 des plantes ont certaines particularités. Premièrement, ils sont relativement courts – environ 170 nucléotides chez *Arabidopsis thaliana* et 430 nucléotides chez *Oryza sativa* – si on les compare aux 3000 nucléotides, en moyenne, des introns humains (Tableau 1). Deuxièmement, alors que les sites 5' et 3' d'épissage ainsi que le point de branchement sont conservés, la région riche en pyrimidines située entre le point de branchement et le site 3' d'épissage est remplacée, chez les plantes, par une région riche en uridine (Reddy et al., 2013). Enfin, les introns des plantes dicotylédones sont riches en adénine et uridine – 67% chez *Arabidopsis thaliana*, contre 52% chez *Homo sapiens* (Reddy, 2007) –, composition qui participe à l'efficacité de leur épissage (Goodall et Filipowicz, 1989).

b) Splicéosome de type U2, dit majeur

i. Composition

Les connaissances générales sur le spliceosome majeur proviennent d'études réalisées sur la levure et l'Homme, mais la plupart des protéines spliceosomales sont conservées chez les plantes (Reddy, 2007). Le spliceosome majeur est composé de 5 petites particules ribonucléoprotéiques nucléaires (snRNP) nommées snRNP U1, U2, U4, U5 et U6, et de toute une myriade de protéines accessoires. Les snRNP sont constituées chacune d'un petit ARN nucléaire (snARN) nommé U1, U2, U4, U5 ou U6 auquel sont associées des protéines dont certaines sont communes à plusieurs snRNP. On estime à environ 45 le nombre de protéines distinctes qui peuvent être associées aux snARN, chez l'Homme (Wahl et al., 2009). Le spliceosome est une machinerie très complexe puisque 300 protéines différentes ont été détectées chez l'Homme et 160 chez la levure – dont toutes ont un homologue chez l'Homme (Reddy et al., 2013). Toutes ces protéines ne sont pas retrouvées simultanément sur le spliceosome. En effet, une des caractéristiques majeures du spliceosome est la dynamique ; de son assemblage à sa dissociation de l'ARN épissé, le spliceosome est constamment remodelé, tout comme l'est le contenu des protéines qui lui sont associées (Wahl et al., 2009).

ii. Biogénèse des snRNP

La biogénèse des snRNP est un processus qui a généralement lieu entre le noyau et le cytoplasme, à l'exception de la snRNP U6 dont l'assemblage est strictement nucléaire. Une snRNP résulte de l'assemblage entre des protéines et un snARN. Les snARN sont des ARN non-codants, abondants et non-polyadénylés. Leur taille varie de 112 nucléotides – pour le snARN U6 – à presque 1200 nucléotides dans le cas du snARN U2 (Matera et Wang, 2014). Ces transcrits sont classés en deux

groupes : 1) snARN de classe « Sm » – snARN U1, U2, U4 et U5 – et 2) snARN de classe « Sm-like » – snARN U6.

1. snRNP U1, U2, U4 et U5

Les snARN de classe Sm sont transcrits par l'ARN polymérase II (Pol II), à partir de promoteurs qui leurs sont propres. Ils sont coiffés d'une 7-méthylguanosine à leur extrémité 5' et mais n'ont pas de queue poly-A à l'extrémité 3'. La terminaison de la transcription des snARN met en jeu à la fois, des protéines chargées sur la Pol II lorsque celle-ci se situe au niveau du promoteur, et un complexe protéique appelé « *integrator* », qui va reconnaître un signal situé en aval du site de clivage – la « 3'-box » – et provoquer le clivage du transcrit. Il a été montré que deux des sous-unités du complexe *integrator*, INTS11 et INTS9, ont des similarités de séquences importantes avec les sous-unités CPSF73 et CPSF100 impliquées dans le clivage et la polyadénylation des pré-ARNm. Il semble que ce soit la sous-unité INTS11 qui clive le transcrit (Matera et al., 2007; Matera et Wang, 2014).

Bien qu'ayant en commun avec les ARNm la coiffe et la transcription par la Pol II, les snARN sont exportés du noyau vers le cytoplasme par une voie d'export différente de la voie TAP-p15, suivie par la plupart des ARNm. C'est la voie d'export CRM1 – pour « *chromosome region maintenance 1* » –, voie d'export des protéines et des sous-unités ribosomiques 40S et 60S, qui est utilisée (Hutten et Kehlenbach, 2007). Les deux sous-unités du complexe lié à la coiffe, CBP20 et CBP80, la protéine ARS2 – pour « *arsenite resistance 2* » – et l'adaptateur d'export PHAX phosphorylé – pour « *phosphorylated adapter RNA export* » – vont recruter

la protéine d'export CRM1 et la GTPase Ran, dans sa forme liée au GTP (Ran-GTP), et ainsi permettre l'export nucléaire du snARN (Figure 5). Une fois dans le cytoplasme, le complexe d'export est désassemblé suite à la déphosphorylation de PHAX, et le snARN est pris en charge par un autre complexe protéique, le complexe SMN – pour « *survival motor neuron* » –, qui est composé de protéines SMN, Sm et GEMINS. Là, la protéine GEMIN5 reconnaît et se fixe sur une séquence du snARN nommée « site Sm ». Par la suite, un hétéroheptamère de protéines Sm – le *Sm core*, formé des protéines Sm B, D1, D2, D3, E, F et G – est assemblé en anneau autour du site Sm et va protéger et stabiliser le snARN. Avant son retour dans le noyau, le snARN va subir des modifications à ses deux extrémités. Le complexe SMN recrute la triméthylguanosine synthase 1, une RNA méthyltransférase qui va former, à l'extrémité 5', une coiffe triméthylée (2,2,7-triméthylguanosine) agissant comme signal d'import. Du côté 3', l'extrémité du snARN est en partie rognée (Figure 5). Ces événements permettent le recrutement subséquent d'un adaptateur d'import nommé Snuportin, au niveau de la coiffe triméthylée, et de l'importine β , aboutissant à l'import nucléaire de la snRNP nouvellement formée (Matera et Wang, 2014). Après dissociation des complexes d'import, les snRNP sont dirigées vers les corps de Cajal, des structures nucléaires, où elles seront associées à d'autres protéines. Là, les snARN subiront des pseudouridylation et des 2'-O-méthylations spécifiques guidées par des scaARN, des petits ARN spécifiques des corps de Cajal (Kiss, 2004).

2. snRNP U6

La biogénèse de la snRNP U6 est strictement nucléaire. Contrairement aux snARN de la classe Sm, le snARN U6 est transcrit par l'ARN polymérase III. Il possède une coiffe γ -monométhylphosphate en 5' et une suite d'uridine à son extrémité 3'. Cette suite d'uridine va permettre le recrutement de la protéine La qui stabilise transitoirement le snARN, avant qu'elle ne soit remplacée par un hétéroheptamère de protéines Lsm (Lsm 2, 3, 4, 5, 6, 7, et 8). Le snARN U6 subit aussi des modifications – 2'-O-méthylations et probablement pseudouridylation – mais celles-ci ont lieu dans le nucléole, sous le contrôle de petits ARN nucléolaires (snoARN) (Kiss, 2004). La snRNP U6 est ensuite transportée vers les corps de Cajal, où elle s'associe à d'autres protéines et à la snRNP U4, par le biais d'appariements de bases entre les snARN U4 et U6. La di-snRNP U4/U6 qui en résulte s'associera elle-même à la snRNP U5, cette fois-ci via des interactions protéine-protéine, pour former une tri-snRNP U4/U6.U5 (Wahl et al., 2009).

neur et le snARN U1, interaction qui est stabilisée par des protéines de la snRNP (Figure 6). À l'autre extrémité de l'intron, la protéine SF1/BPP se lie au point de branchement tandis que le facteur auxiliaire d'U2 – pour « *U2 auxiliary factor* » – se fixe sur la séquence riche en pyrimidines et sur le site accepteur d'épissage, *via*, respectivement, les deux sous-unités U2AF65 et U2AF35 qui le composent. Ces premières interactions forment le complexe splicéosomal E. Lors de la deuxième étape, la snRNP U2 vient déplacer SF1/BPP et interagir, *via* son snARN U2, avec la séquence du point de branchement. Ceci constitue le pré-splicéosome, ou complexe A. La troisième étape voit le recrutement de la tri-snRNP U4/U6.U5 ; on parle alors du complexe B ou splicéosome pré-catalytique. À ce stade, toutes les snRNP sont présentes sur l'intron. Le splicéosome est alors activé par une suite de réarrangements majeurs de conformation et de composition qui aboutissent, notamment, au relargage des snRNP U1 et U4. Activé – c'est le complexe B* –, le splicéosome peut alors catalyser la première des deux réactions de transestérification, qui voit l'attaque nucléophile du groupement 2'-OH de l'adénosine du point de branchement, sur le site donneur d'épissage. Ceci conduit à la formation d'un lasso au niveau de l'intron. L'exon situé en amont est alors libre, tandis que l'exon en aval est lié à l'intron « en lasso ». Le splicéosome est qualifié de catalytique, ou complexe C. De nouveaux réarrangements moléculaires précèdent la deuxième réaction de transestérification, l'attaque nucléophile de l'extrémité 3'-OH de l'exon situé en amont, sur le site accepteur d'épissage. De récents résultats indiquent que le snARN U6 serait responsable de la catalyse des deux réactions de transestérification (Fica et al., 2013, 2014). L'intron « en lasso » est libéré puis généralement dégradé, tandis que les deux exons sont joints l'un à l'autre. Les snRNP U2, U5 et U6 sont, quant à elles, recyclées (Kornbliht et al., 2013; Wahl et al., 2009). Dans certains cas, l'intron excisé est linéarisé et utilisé comme précurseur d'ARN non-codants, certains snoARN et microARN (miARN) par exemple (Brown et al., 2008).

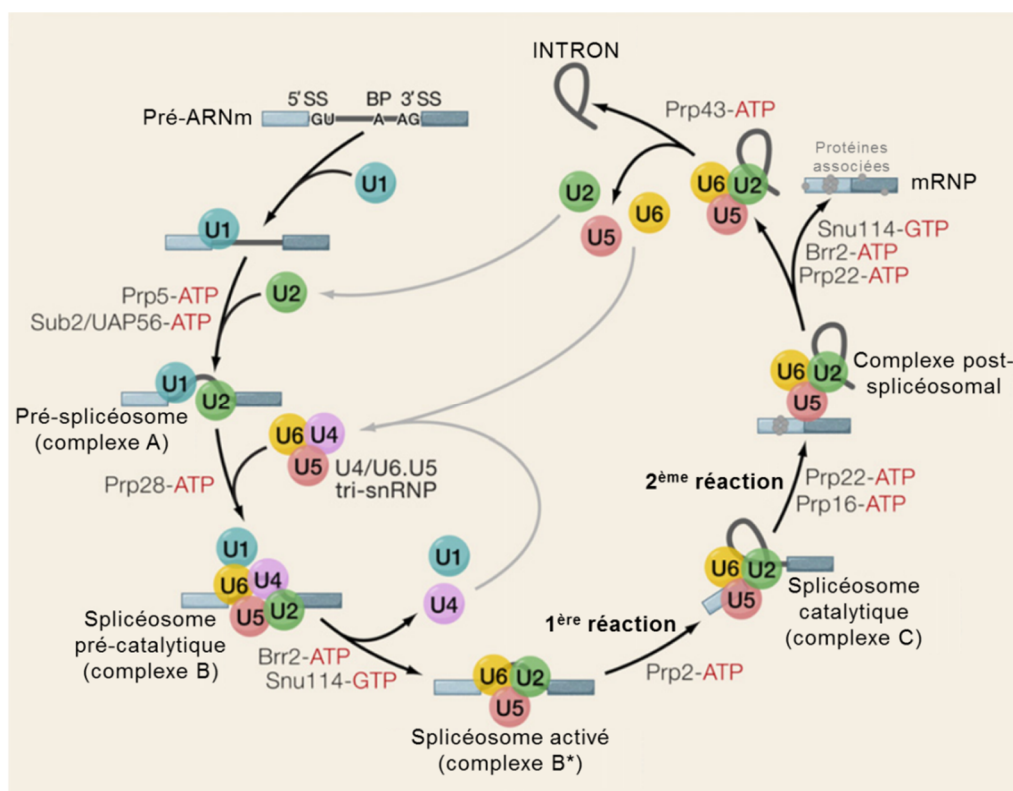


Figure 6 : Assemblage séquentiel du splicéosome de type U2. La liaison des snRNP U1 et U2 au pré-ARNm forme le pré-splicéosome. Le recrutement de la tri-snRNP U4/U6.U5 aboutit à la formation d'un splicéosome pré-catalytique. Des réarrangements destinés à activer le splicéosome conduisent à la libération des snRNP U1 et U4. Le splicéosome activé catalyse la première réaction de transestérification, où le groupement 2'-OH de l'adénosine constituant le point de branchement réalise une attaque nucléophile sur le site donneur d'épissage (5'SS). La deuxième réaction constitue en l'attaque nucléophile de l'extrémité 3'-OH du premier exon sur le site accepteur d'épissage (3'SS), ce qui relie les exons entre eux et libère l'intron sous la forme d'un lasso. Les protéines impliquées dans les transitions entre les différentes étapes sont indiquées. D'après Wahl et al., 2009.

Un complexe protéique nommé EJC – pour « *exon junction complex* » –, et dont beaucoup de composants sont recrutés par le spliceosome lorsque celui adopte les structures « complexes B et C », est déposé entre 20 et 24 nucléotides en amont de la jonction d'épissage. L'EJC est impliqué dans l'export de l'ARNm auquel il est lié, ainsi que dans un mécanisme de contrôle-qualité qui a lieu lors des premiers événements de traduction. En effet, quand un complexe de terminaison de la traduction s'assemble en amont d'un EJC, le codon stop responsable de l'arrêt de la traduction est considéré comme prématuré. L'EJC participe alors au recrutement de protéines qui vont diriger l'ARNm vers une voie de dégradation spécifique, la NMD (pour « *nonsense-mediated decay* »). La présence d'un codon stop prématuré peut être due à un épissage raté. Elle peut aussi résulter d'un événement régulé visant à contrôler l'expression d'un gène en modulant la quantité d'ARNm correspondant (Chang et al., 2007).

ii. Changements de composition et conformation

Les différentes étapes de l'épissage sont marquées par de constantes variations dans la composition du spliceosome. Comme cela a été décrit dans la partie précédente, les snRNP composant le spliceosome ne s'assemblent pas toutes en même temps, et certaines – les snRNP U1 et U4 – sont libérées au cours du processus. À chaque étape, de nombreux facteurs d'épissage et des protéines liées aux snRNP vont être relargués tandis que d'autres seront recrutés, concomitamment. L'ensemble des protéines échangées au cours des différentes étapes du processus d'épissage semble d'ailleurs varier d'un pré-ARNm à l'autre (Wahl et al., 2009).

D'importants changements conformationnels ont lieu tout au long du processus d'épissage. Par exemple, l'arrivée de la tri-snRNP U4/U6.U5, lors de la formation du complexe B, conduit à la mise en place d'interactions entre le snARN U5 et des séquences des deux exons bordant l'intron à épisser. Simultanément, une partie du snARN U6 va s'apparier au snARN U2. Le passage au complexe B* est accompagné d'appariements entre le snARN U6 et le site donneur d'épissage. Il en découle une déstabilisation des interactions entre le snARN U1 et le site donneur d'une part, et entre les snARN U4 et U6 d'autre part. Ceci aboutit au relargage des snRNP U1 et U6. Ces réarrangements sont effectués grâce à l'action de nombreuses protéines dont des hélicases/ATPases ARN-dépendantes du type DExD/H telles que Sub2/UAP56, Prp 28 et Brr2 (Figure 6; Matera et Wang, 2014; Wahl et al., 2009).

3) Épissage des introns spliceosomaux de type U12

a) Caractéristique des introns de type U12

Chez beaucoup d'eucaryotes, un tout petit nombre d'introns spliceosomaux ne partage pas les mêmes caractéristiques qui définissent les introns de type U2. Ces introns, dits de type U12, possèdent des séquences consensus des sites de branchement et 5' d'épissage plus longues et plus conservées que celles des introns de type U2. De plus, ils sont dépourvus d'une séquence riche en pyrimidines et la distance qui sépare le point de branchement du site 3' d'épissage est relativement courte – entre 11 et 13 nucléotides. Les bordures de ces introns sont souvent définies par les dinucléotides GU-AG, et plus rarement par les dinucléotides AU-AC (Turunen et al., 2013).

Des introns de type U12 ont été découverts chez différentes espèces des principaux taxons d'eucaryotes. Leur nombre est estimé à 800 chez l'Homme et 300 chez *Arabidopsis thaliana* – une récente étude suggère qu'il y en aurait environ 2100 chez cette dernière (Marquez et al., 2012) –, mais ils sont absents chez de nombreux protistes et chez les deux organismes modèles *Saccharomyces cerevisiae* et le ver *Caenorhabditis elegans*. Enfin, ces introns ne sont très souvent présents qu'en une seule copie au sein d'un gène donné (Turunen et al., 2013).

b) Réaction d'épissage

Les introns splicéosomiaux de type U12 sont épissés par un splicéosome qui leur est propre, le splicéosome de type U12, dit mineur. Il se compose de 5 snRNP : les snRNP U11, U12, U4atac, U6atac et U5, cette dernière étant partagée avec le splicéosome majeur. La biogénèse des snRNP du splicéosome mineur ne semble guère différer de celle du splicéosome majeur. Les snARN U11, U12 et U4atac appartiennent à la classe Sm, tandis que le snARN U6atac appartient à la classe Sm-like (Matera et Wang, 2014).

Les différentes étapes de l'assemblage du splicéosome mineur sont assez similaires à celles du splicéosome majeur. La différence principale a lieu au début de l'assemblage car les snRNP U11 et U12 sont recrutées concomitamment, sous la forme d'une di-snRNP U11/U12 qui va reconnaître le site donneur d'épissage et le point de branchement. La tri-snRNP U4atac/U6atac.U5 vient ensuite s'assembler sur le splicéosome en formation, formant ainsi le complexe B. Des réarrangements conformationnels – et probablement compositionnels – conduisent à l'activation du splicéosome mineur et à la libération des snRNP U11 et U4atac. Les réactions de transestérification catalysées par les deux splicéosomes sont identiques, mais la cinétique réactionnelle est beaucoup plus lente dans le cas du splicéosome mineur (Turunen et al., 2013).

4) Épissage alternatif

a) Introduction théorique

Les quatre signaux qui définissent les introns de type U2 et les exons ne sont pas suffisants pour assurer l'assemblage du splicéosome. En effet, ces signaux sont peu stringents ; les introns/exons peuvent ainsi avoir des séquences qui diffèrent plus ou moins des signaux consensus. La présence de tels signaux peut aussi conduire à la définition de « pseudoexons » qui ne seront pourtant jamais retrouvés dans les ARNm matures (Gómez Acuña et al., 2013). Enfin, les interactions impliquées dans la mise en place du splicéosome sont relativement faibles (Wahl et al., 2009). Par conséquent, des auxiliaires sont indispensables pour assister la cellule dans la reconnaissance des véritables introns/exons et dans l'assemblage des différents composants du splicéosome. Des facteurs protéiques, agissant en *trans* à la suite de leur interaction avec des séquences ribonucléiques *cis*-régulatrices, vont intervenir favorablement ou défavorablement dans le processus d'épissage. Ainsi, un « dialogue d'interactions » entre facteurs d'épissage et composants du splicéosome se met en place pour chaque intron ou exon. Lorsque ce dialogue aboutit constamment à l'épissage, on parle d'épissage constitutif. Quand ce n'est pas le cas, on qualifie l'épissage d'alternatif. Divers paramètres peuvent influencer, par l'intermédiaire des facteurs d'épissage, l'issue de l'épissage alternatif.

b) Généralités sur le rôle biologique de l'épissage alternatif

L'épissage alternatif permet d'obtenir plusieurs ARNm différents, à partir d'un pré-ARNm unique. Son influence est prépondérante dans l'augmentation de la complexité du transcriptome des organismes qui l'utilise. Les principaux types d'épissage alternatif sont résumés dans la figure 7. Il faut noter que les différents types d'épissage ne sont pas retrouvés avec la même fréquence chez tous les organismes. Ainsi, le saut d'exon, qui est le type d'épissage majeur chez les mammifères n'est que rarement utilisé chez les plantes, les champignons et les protistes. Ceux-ci utilisent majoritairement la rétention d'intron – environ 40% des événements d'épissage chez *Arabidopsis thaliana* (Marquez et al., 2012) – alors que ce type d'épissage est rare chez les mammifères (Keren et al., 2010).

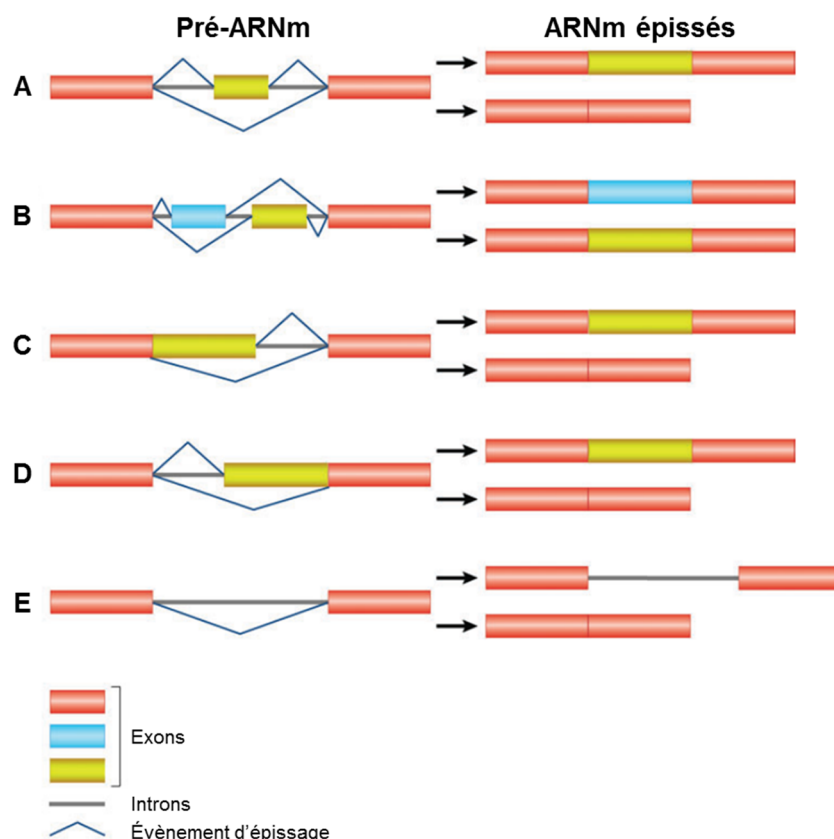


Figure 7 : Principaux types d'épissage alternatif. Les exons épissés alternativement sont représentés en jaune et bleu. (A) Saut d'exon. (B) Exons mutuellement exclusifs. (C) Sélection alternative de sites donneurs d'épissage. (D) Sélection alternative de sites accepteurs d'épissage. (E) Rétention d'intron. D'après Reddy, 2007.

Les conséquences biologiques de l'épissage alternatif sont multiples. Elles s'inscrivent dans le cadre très large de l'expression génétique et sa régulation. L'épissage alternatif augmente la complexité du protéome quand les exons/introns épissés alternativement sont codants (Nilsen et Graveley, 2010). Les protéines synthétisées à partir des transcrits épissés vont potentiellement avoir des fonctions, une stabilité, des localisations ou des modes de régulations différents. La possibilité d'obtenir plusieurs isoformes d'ARNm à partir d'un seul pré-ARNm peut aussi être utilisée pour contrôler l'expression d'un gène par le biais, notamment, de la voie NMD ou de l'ARN interférent. Ainsi, l'inclusion d'un exon dans lequel se trouve un codon stop prématuré ou un site de liaison pour un miARN peut participer à la régulation négative de l'expression d'un gène, en réponse à certains paramètres. La localisation et l'efficacité de traduction d'un transcrit peuvent aussi être modulées par certains événements d'épissage (Reddy et al., 2013; Staiger et Brown, 2013).

c) Régulation de l'épissage alternatif à l'échelle moléculaire

i. Facteurs protéiques régulant l'épissage

Les facteurs régulateurs d'épissage sont les acteurs de base de la régulation de l'épissage alternatif. Ce sont des protéines qui se fixent sur des séquences *cis*-régulatrices et exercent une fonction stimulatrice ou inhibitrice d'épissage. Les séquences *cis*-régulatrices sont nommées d'après leur localisation sur le pré-ARNm et le rôle que le facteur qu'elles recrutent aura sur l'épissage. Ce sont les ESE (*exonic splicing enhancer*), ESS (*exonic splicing silencer*), ISE (*intronic splicing enhancer*) et ISS (*intronic splicing silencer*). Les facteurs régulateurs appartiennent majoritairement à deux groupes de protéines : 1) les protéines dites SR, car riches en sérine (S) et arginine (R), et 2) les ribonucléoprotéines nucléaires hétérogènes (hnRNP). Généralement, les protéines SR ont un rôle stimulateur dans

l'assemblage du spliceosome, tandis que les hnRNP ont plus souvent un rôle inhibiteur. Enfin, il est intéressant de noter que les facteurs régulateurs d'épissage régulent souvent l'expression de leurs propres gènes *via* la voie du NMD. Ils vont influencer sur l'épissage alternatif de leurs transcrits en promouvant ou réprimant – selon leur activité – l'insertion d'exons contenant des codons stop prématurés (Ni et al., 2007).

1. Protéines SR et SR-like

Une nomenclature stricte définit les protéines SR comme des protéines modulaires qui possèdent un ou deux domaines de reconnaissance d'ARN nommé RRM – pour « *RNA recognition motif* » –, à leur extrémité amino-terminale, et un domaine « RS » long d'au moins 50 acides aminés et contenant au moins 40% d'arginine et sérine, à l'extrémité carboxy-terminale (Manley et Krainer, 2010). Les protéines SR reconnaissent et se fixent sur des séquences stimulatrices ESE ou ISE *via* leur RRM, et vont faciliter l'assemblage de certains composants du spliceosome contenant eux aussi un domaine SR, par le biais d'interactions entre domaines SR. U1-70K et U2AF35 sont des exemples de protéines impliquées dans l'assemblage du spliceosome et qui possèdent un domaine RS. 12 protéines SR ont été identifiées chez l'Homme mais c'est chez les plantes que l'on en retrouve le plus grand nombre – environ deux fois plus chez *Arabidopsis thaliana* – (Busch et Hertel, 2012; De Conti et al., 2013; Reddy et al., 2013). D'autres protéines partagent certaines caractéristiques avec les protéines SR mais ne sont pas considérées comme telles. Parmi ces protéines « SR-like », on trouve U1-70K, qui est une des protéines de la snRNP U1 reconnaissant le site 5' d'épissage, et U2AF65, qui est une des sous-unités protéiques définissant le site 3' d'épissage.

2. hnRNP

Les hnRNP sont, comme les protéines SR, des protéines modulaires qui possèdent un ou plusieurs domaines de reconnaissance d'ARN – souvent du type RRM – et des domaines impliqués dans des interactions protéine-protéine. Les hnRNP se lient à leur séquence cible et exercent leur activité inhibitrice en imposant, dans la plupart des cas, des contraintes stériques qui vont jouer en défaveur de l'assemblage du spliceosome. Plus d'une vingtaine d'hnRNP sont connues chez l'Homme et environ deux fois plus chez *Arabidopsis thaliana* (Busch et Hertel, 2012; De Conti et al., 2013; Reddy et al., 2013).

ii. Régulation transcriptionnelle

L'assemblage des facteurs régulateurs d'épissage sur un pré-ARNm a lieu très majoritairement au cours de la transcription et il peut être couplé à cette dernière. Il a en effet été montré que le domaine carboxy-terminal – ou « queue CTD » – de la grande sous-unité de la Pol II peut participer au recrutement de facteurs d'épissage. Par exemple, la queue CTD est nécessaire au recrutement d'une protéine SR nommée SRSF3, qui est un régulateur négatif de l'inclusion de l'exon 33 du pré-ARNm de la fibronectine, une protéine de la matrice extracellulaire (de la Mata et Kornblihtt, 2006). Il a été montré, par des mutations et par l'utilisation de drogues diminuant la vitesse de la Pol II, que la vitesse de la transcription influe aussi sur l'issue de l'épissage alternatif (Kornblihtt et al., 2013). L'irradiation par rayonnement ultra-violet, lorsqu'elle conduit à une inhibition de l'élongation de la transcription *via* une hyper-phosphorylation de la queue CTD de la Pol II, provoque des changements du profil d'épissage de nombreux transcrits, dont certains issus de gènes codant des protéines impliquées dans l'apoptose (Muñoz et al., 2009). En règle générale, il semble qu'une vitesse d'élongation rapide a tendance à promouvoir l'assemblage du spliceosome et/ou de facteurs régulateurs sur des sites « forts », c'est-à-dire des sites dont les séquences sont proches des séquences consensus d'épissage et qui permettent une interaction rapide avec leur(s) partenaire(s) (Figure 8). Inversement, une faible vitesse d'élongation sera moins défavorable à l'utilisation de sites « faibles » (Kornblihtt et al., 2013).

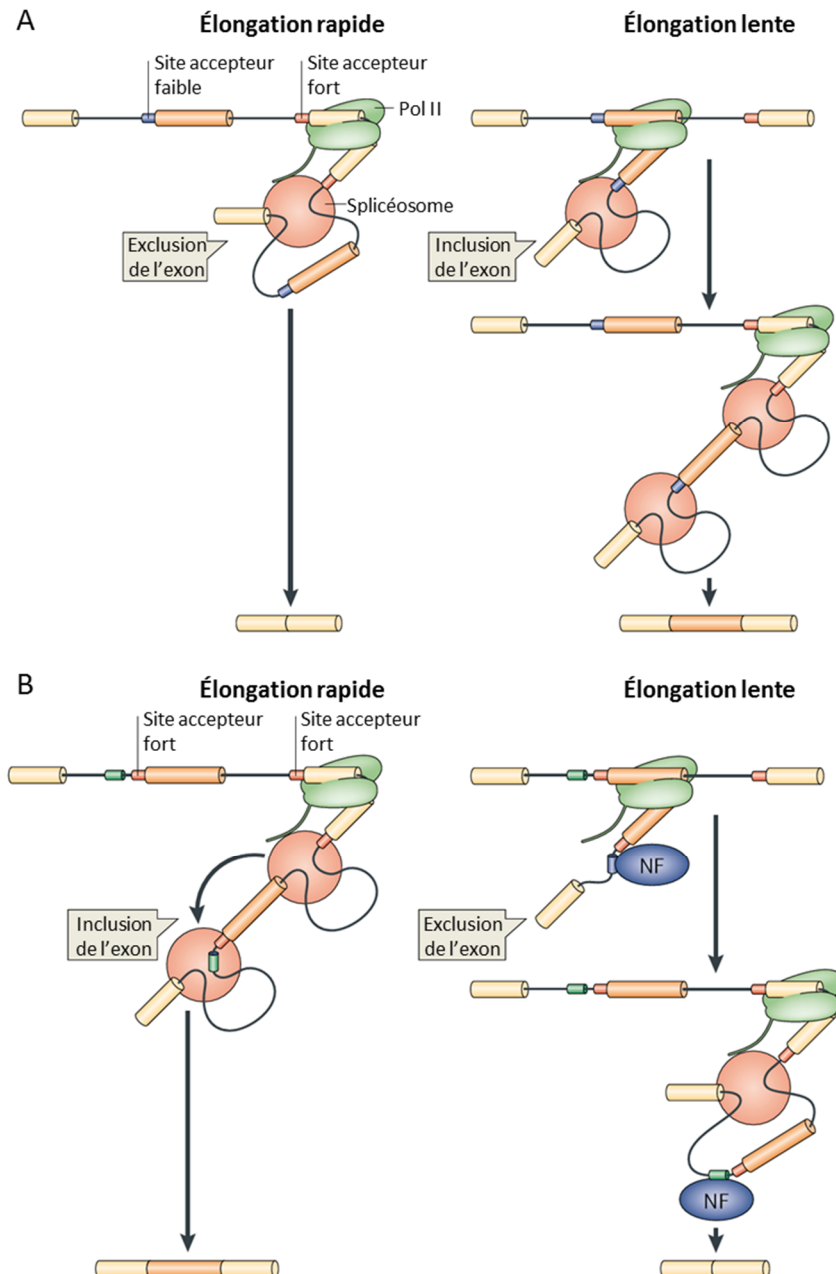


Figure 8 : Influence de la vitesse de transcription sur l'épissage alternatif. La vitesse d'élongation de la Pol II peut avoir des effets opposés sur l'issue de l'épissage alternatif. (A) Une élongation rapide a tendance à favoriser le recrutement des composants du splicéosome sur des sites forts, tandis qu'une élongation lente, du fait qu'elle permet plus temps pour ce recrutement, autorise la sélection de sites faibles. (B) Une élongation lente peut également favoriser le recrutement de facteurs d'épissage dont certains auront une action négative (NF) sur l'assemblage subséquent du splicéosome (B). D'après Kornbliht et al., 2013.

iii. Régulation épigénétique

L'épissage alternatif est influencé par des facteurs épigénétiques. Dans certains génomes dont ceux d'*Homo sapiens* et d'*Arabidopsis thaliana*, il a été montré que le positionnement des nucléosomes ne se fait pas au hasard. En effet, les nucléosomes sont plus abondants au niveau des exons qu'au niveau des introns. De plus, les tailles moyennes d'un exon et d'un fragment d'ADN enroulé autour d'un nucléosome sont très proches – environ 150 nucléotides. Enfin, il a été observé une plus grande abondance de nucléosomes au niveau des exons inclus dans le messager final, par rap-

port aux exons exclus (Luco et al., 2011). Un des modèles proposés est que les nucléosomes, du fait qu'ils induisent un ralentissement de la Pol II, donneraient aux facteurs régulateurs d'épissage et aux composants du spliceosome plus de temps pour s'assembler sur les transcrits en cours de synthèse (Kornblihtt et al., 2013). Par conséquent, ils faciliteraient la reconnaissance des sites « faibles ». Les modifications post-traductionnelles des histones spécifiques à certains types cellulaires ou induites par des stimuli peuvent aussi réguler l'épissage alternatif. En lien avec les exemples cités précédemment, les modifications des histones associées à une activation de la transcription et une élongation rapide – par exemple un groupement méthyl sur la lysine 9 de l'histone H3 – vont avoir tendance à défavoriser les sites « faibles », et inversement (Kornblihtt et al., 2013). Dans d'autres cas, certaines modifications d'histones peuvent recruter des facteurs régulateurs d'épissage par le biais de protéines adaptatrices et ainsi, influencer sur l'inclusion ou l'exclusion de certains exons (Luco et al., 2010).

d) Pathologies impliquant l'épissage alternatif

Des pathologies, dont certains cancers, ont pour cause directe ou indirecte des dérèglements de l'épissage alternatif. Ces dérèglements peuvent être dus à des mutations altérant le profil d'épissage d'un gène en particulier. C'est le cas du syndrome progéria de Hutchinson-Gilford où un site cryptique d'épissage est activé suite à une mutation synonyme dans le gène codant pour la lamine A (Eriksson et al., 2003). D'autres pathologies peuvent résulter de mutations dans des facteurs régulateurs d'épissage ou des composants du spliceosome. L'atrophie musculaire spinale est, par exemple, due à une mutation synonyme qui provoque un saut d'exon et une perte de fonction du gène codant pour la protéine SMN1 qui est impliquée dans la biogénèse des snRNP de classe Sm (Singh et Cooper, 2012).

e) Aperçu de rôles de l'épissage alternatif

L'épissage alternatif est un processus biologique dont le spectre de conséquences – directes et indirectes – sur l'expression génétique est très large. Son rôle est très important dans des mécanismes aussi fondamentaux que le développement des organismes ou leur adaptation à l'environnement.

i. Implication dans le développement chez les animaux

L'épissage alternatif est un phénomène très répandu chez les animaux en général, et les mammifères en particulier. Chez l'Homme, il est estimé que plus de 90% des gènes subissent de l'épissage alternatif (Wang et al., 2008). Les deux exemples décrits ci-dessous illustrent l'importance de ce processus dans le développement des organismes.

L'expression du gène *Dscam1* – pour « *Down syndrome cell adhesion molecule 1* » – de *Drosophila melanogaster* est l'exemple extrême qui montre l'influence que peut avoir l'épissage alternatif sur la complexité du protéome. Le pré-ARNm du gène *Dscam1* contient 95 exons mutuellement exclusifs, répartis en 4 clusters, qui permettent théoriquement la production de plus de 38000 isoformes d'ARNm (Figure 9). *Dscam1* est une protéine membranaire de la super-famille des immunoglobulines. Chez la drosophile, *Dscam1* est impliquée dans la mise en place des réseaux neuronaux : elle maximise la mise en place de connexions inter-neurales en empêchant les auto-connexions entre axones et dendrites d'un même neurone. Pour ce faire, chaque neurone exprime à sa surface une isoforme de *Dscam1* particulière. Un contact qui s'établit entre deux isoformes identiques indique une interaction entre deux branches d'un même neurone et conduit à une répulsion entre ces dernières. La production de nombreuses isoformes de *Dscam1* n'est pas que théorique, il a été montré qu'elle est biologiquement nécessaire (Hattori et al., 2009).

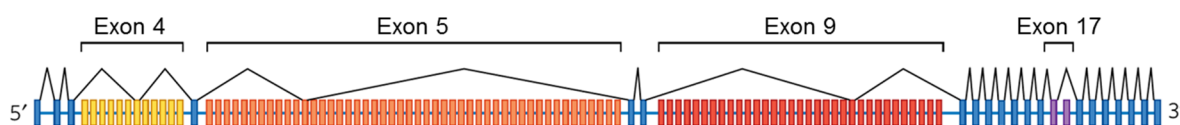


Figure 9 : Pré-ARNm du gène Dscam1 de *Drosophila melanogaster*. Ce pré-ARNm permet, en théorie, la production de plus de 38000 isoformes épissées. Les exons constitutifs sont représentés en bleu, tandis que les 95 exons mutuellement exclusifs sont représentés, selon le cluster auquel ils appartiennent, en jaune, orange, rouge et violet. D'après Nilsen et Graveley, 2010.

L'entrée en différenciation cellulaire est en partie régulée par l'épissage alternatif chez les cellules souches embryonnaires. Chez l'Homme et la souris *Mus musculus*, les cibles génétiques du facteur de transcription FOX1 sont régulées par un événement d'épissage alternatif conservé. Dans les cellules souches embryonnaires, l'inclusion de l'exon 18a aboutit à la production d'une isoforme de FOX1 nommée FOX1-ES, à la fois capable de stimuler l'expression de facteurs de transcription nécessaires au maintien de la totipotence, et de réprimer l'expression de gènes nécessaires à la différenciation cellulaire. Au cours de la différenciation cellulaire, l'inclusion de l'exon 18b au lieu de l'exon 18a provoque la production d'un facteur FOX1 conventionnel dont le domaine de liaison à l'ADN diffère de celui de FOX1-ES. L'action de FOX1 est en faveur de la différenciation et réprime l'expression des gènes requis dans la mise en place et le maintien de la totipotence cellulaire (Gabut et al., 2011).

ii. Implication dans l'adaptation à l'environnement chez les plantes

L'épissage alternatif a longtemps été estimé comme n'étant qu'un phénomène mineur chez les plantes, moins prévalent que chez les animaux. Pourtant, grâce à l'accumulation de données de séquençage à haut-débit, la proportion de gènes connus pour subir de l'épissage alternatif, chez *Arabidopsis thaliana*, est passée de 11% (Iida et al., 2004) à 61% (Marquez et al., 2012). Deux traits particuliers distinguent l'épissage alternatif chez les plantes. Premièrement, l'événement d'épissage le plus commun est la rétention d'intron ; cette caractéristique est partagée avec les protistes et les eucaryotes qualifiés d'inférieurs. Deuxièmement, les plantes possèdent un grand nombre de gènes codant pour des facteurs régulateurs d'épissage ; il y a deux fois plus de gènes codant des protéines SR et des hnRNP chez *Arabidopsis* que chez l'Homme. La multiplicité des gènes de régulateurs d'épissage pourrait avoir son explication dans des événements de duplications et réarrangements de génome auxquels les plantes peuvent être soumises (Reddy et al., 2013). Chez les plantes, organismes statiques, l'épissage alternatif est particulièrement impliqué dans la régulation de l'expression génétique en réponses à des facteurs environnementaux et aux stress (Staiger et Brown, 2013).

L'épissage alternatif participe, par exemple, au maintien de l'horloge circadienne chez les plantes. Ainsi, la mutation de la protéine PRMT5 – pour « *protein arginine methyltransferase5* » –, une protéine impliquée dans la méthylation de certains résidus arginine des protéines Sm, provoque un retard du rythme circadien chez *Arabidopsis*. Ce délai est en partie dû à une altération de l'épissage alternatif d'un des gènes clés de l'horloge circadienne végétale, le gène PRR9 (Sanchez et al., 2010).

Récemment, il a été rapporté que la lumière module l'épissage alternatif de certains gènes d'*Arabidopsis*, dont ceux codant pour la protéine SR-like U2AF65 et les protéines SR SR30 et RS31. Étonnamment, cette modulation dépend d'un signal d'origine chloroplastique – probablement sous le contrôle du statut d'oxydoréduction de la plastoquinone – capable de circuler de façon systémique dans toutes les parties de la plante. L'absence de lumière conduit, par exemple, à la rétention d'introns présents dans le gène RS31. Ce changement de profil d'épissage est aussi observé dans les racines, sauf lorsque celles-ci sont séparées, par dissection, des parties aériennes. La rétention d'introns dans le transcrit RS31 est nécessaire : des plantes où seule l'isoforme totalement épissée du messenger est produite montrent des altérations phénotypiques importantes telles qu'une petite taille et une dégradation rapide des chlorophylles a et b (Petrillo et al., 2014).

iii. Rôles chez les virus

1. Virus des animaux

Les virus des animaux qui produisent leurs transcrits dans le noyau des cellules qu'ils infectent utilisent, comme leurs hôtes, l'épissage alternatif pour enrichir leur protéome et/ou réguler l'expression de leurs gènes. On retrouve, notamment, des transcrits épissés alternativement chez des virus des familles suivantes : *Adenoviridae*, *Herpesviridae*, *Papillomaviridae*, *Polyomaviridae*, *Parvoviridae*, *Bornaviridae*, *Orthomyxoviridae*, *Retroviridae* et *Hepadnaviridae*.

L'exemple typique de l'utilisation de l'épissage alternatif par des virus est celui des adénovirus puisque que c'est chez derniers que l'épissage alternatif a été découvert (Berget et al., 1977; Chow et al., 1977). Le génome des adénovirus est un ADN double-brin linéaire, d'une longueur comprise entre 26 et 45 kpb (Leppard, 2008). Ils possèdent 10 unités de transcription dirigées par leur propre promoteur, et qualifiées de précoces ou tardives selon la phase de l'infection au cours de laquelle elles sont exprimées (Figure 10). Grâce à l'utilisation de sites de polyadénylation différents et de l'épissage alternatif, les adénovirus sont capables de produire une cinquantaine de transcrits différents, au cours de l'infection (Zhao et al., 2014).

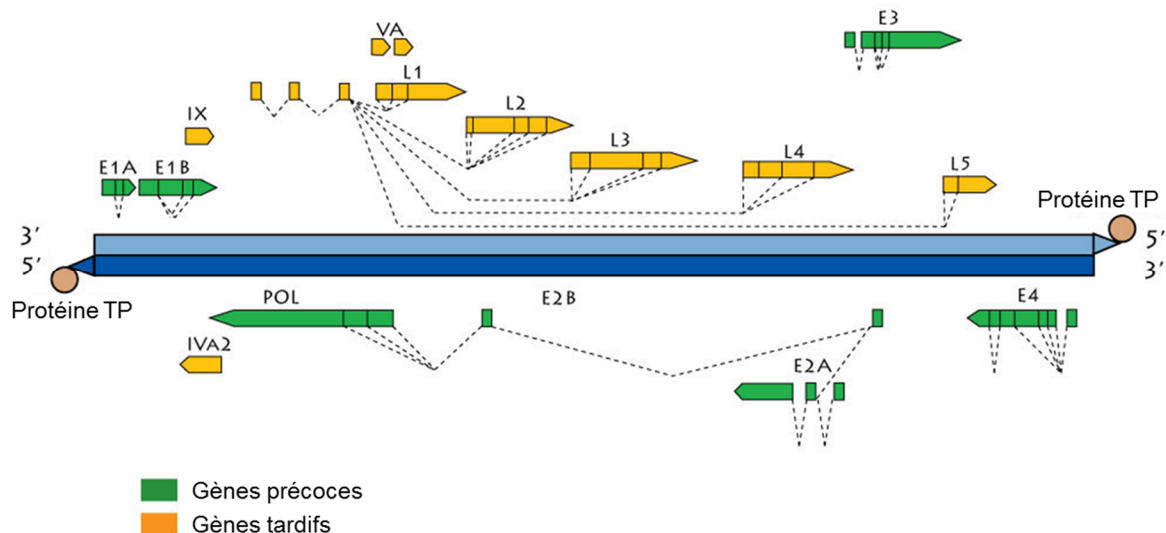


Figure 10 : Carte génomique de l'adénovirus humain de type 5. L'ADN génomique double-brin linéaire a une longueur d'environ 35 kpb. Une séquence inversée d'environ 100 nucléotides est présente aux extrémités de chacun des brins. La protéine terminale TP correspond à une protéine virale liée, de façon covalente, aux deux extrémités 5' du génome viral. D'après Swiss Institute of Bioinformatics.

Chez les adénovirus, l'épissage alternatif permet non seulement d'augmenter la capacité codante du génome mais il est aussi impliqué dans le contrôle temporel de l'expression des gènes viraux. Un exemple bien décrit est celui de l'ARN L1. Cet ARN possède un site donneur et deux sites accepteurs. Bien qu'il appartienne à une unité de transcription tardive, l'ARN L1 est aussi transcrit au cours de la phase précoce de l'infection. A ce stade, c'est le site accepteur proximal qui est utilisé ce qui permet la production des protéines 52/55K. L'utilisation du site accepteur distal, qui permet la production de la protéine IIIa, est réprimée par des protéines SR hyperphosphorylées. Alors que l'infection progresse en phase tardive, la protéine virale E4-ORF4 va interagir, à la fois, avec une phosphatase cellulaire nommée PP2A et des protéines SR hyperphosphorylées. Il en résulte une déphosphorylation des protéines SR qui annihile leur rôle répresseur vis-à-vis de l'utilisation du site accepteur distal (Estmer Nilsson et al., 2001; Kanopka et al., 1998). Il a été montré qu'une autre protéine virale, L4-33K, est un facteur d'épissage qui stimule l'utilisation de sites accepteurs faibles, en

particulier le site qui conduit à la production d'ARNm codant pour la protéine IIIa (Törmänen et al., 2006).

Le papillomavirus humain HPV-16, dont le génome ne possède que deux promoteurs et deux signaux de polyadénylation, fait usage de l'épissage alternatif pour exprimer ses huit gènes. La transition entre les phases précoce et tardive de l'expression des gènes viraux est d'ailleurs accompagnée de modulations dans l'activité de certains sites d'épissage (Johansson et Schwartz, 2013). Le virus de l'influenza A (IAV), un virus à ARN dont le génome est constitué de huit segments d'ARN de polarité négative, utilise l'épissage alternatif pour produire certaines de ses protéines dont la protéine non-structurale 2 (NS2/NEP), impliquée dans l'export nucléaire des ribonucléoprotéines virales (vRNP) (Dubois et al., 2014).

Le virus de l'immunodéficience humaine de type 1 (HIV-1), un rétrovirus complexe, utilise l'épissage alternatif pour générer plus de 40 messagers – dont certains ne sont que partiellement épissés – à partir de son seul pré-ARNm. Certaines protéines virales, comme la protéine Tat, sont produites à partir de multiples ARNm résultant d'événements d'épissage différents. Le profil d'épissage alternatif du HIV-1 est régulé au cours du cycle viral. Les transcrits complètement épissés, qui codent pour des protéines régulatrices, sont les premiers à être produits. Tandis que l'infection progresse, les transcrits partiellement épissés puis non-épissés sont produits et exportés vers le cytoplasme grâce à l'action de la protéine virale régulatrice Rev (Leblanc et al., 2013).

Chez le virus de l'hépatite B (HBV), un pararétrovirus de la famille des *Hepadnaviridae*, cinq sites donneurs et cinq sites accepteurs d'épissage ont été décrits mais leur rôle est encore obscur. Les ARN épissés peuvent être rétrotranscrits et encapsidés, produisant ainsi des virions défectifs. Il a été montré que les ARN épissés sont souvent détectés chez les patients atteints d'hépatite chronique et qu'ils s'accumulent préférentiellement lorsque l'hépatite chronique se transforme en un hépatocarcinome cellulaire (Bayliss et al., 2013). Enfin, certains ARN épissés possèdent un ORF résultant d'une fusion entre l'extrémité 5' du gène codant pour la transcriptase inverse et un ORF créé par l'événement d'épissage. L'ORF de fusion code pour une petite protéine hautement immunogène d'environ 10 kDa, nommée HBSP pour « *HBV splice-generated protein* » et n'ayant pas de rôle bien défini. Des études utilisant de l'HBSP produite hors-contexte viral ont en tout cas montré qu'elle n'a pas d'influence sur la réplication et la transcription du virus (G. H. Lee et al., 2008).

2. Virus des plantes

Les phytovirus sont majoritairement des virus à ARN ; ceci sous-entend qu'ils réalisent leur cycle de réplication dans le cytoplasme, à l'exception des *Nucleorhabdovirus*. L'épissage, qu'il soit constitutif ou alternatif, est rare et seulement cantonné aux virus des familles *Caulimoviridae*, dont le génome est un ADN double-brin circulaire, et des *Geminiviridae*, au génome à ADN simple-brin circulaire. Chez le virus de la striure du maïs (MSV, Figure 11), un virus du genre des *Mastrevirus* (famille des *Geminiviridae*), deux introns alternativement épissés ont été décrits. Les deux brins du MSV sont codants et permettent la production de transcrits dits de sens viraux (V) et complémentaires (C). L'épissage alternatif des introns du MSV est, dans les deux cas, de type rétention d'intron. L'épissage de l'intron situé dans les ORF C1 et C2 conduit à leur fusion en phase. L'ORF de fusion C1:C2 code pour la protéine virale Rep qui est impliquée dans l'initiation de la réplication du génome viral. L'ARN non-épissé permet l'expression de l'ORF C1 qui code pour la protéine RepA qui est impliquée, entre autres, dans la réplication du génome viral. Le rôle de l'épissage de l'intron situé dans l'ORF V1 n'est pas connu (Wright et al., 1997).

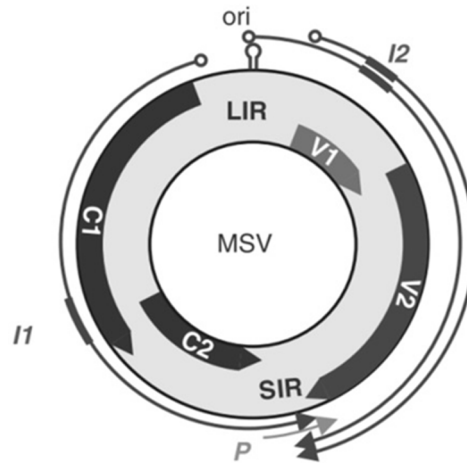


Figure 11 : Carte génomique du virus de la striure du maïs (MSV). Le génome du MSV est un ADN circulaire simple-brin de 2,7 kpb. Des transcrits de sens « viraux » (V) et « complémentaires » (C) sont produits au cours du cycle infectieux. La grande région intergénique (LIR) contient les promoteurs viraux, ainsi qu'une séquence en tige-boucle impliquée dans la réplication du génome viral, par le mécanisme de cercle roulant (*rolling circle*). La petite région intergénique (SIR) contient quant à elle, des signaux de polyadénylation, ainsi qu'un site de fixation pour un fragment d'ADN (P) qui est utilisé pour initier la synthèse du brin complémentaire. Les ORF V1 et V2 codent respectivement pour la protéine de mouvement et la protéine de capsid. L'ORF C1 code pour la protéine RepA, tandis que la fusion entre l'extrémité 5' de l'ORF C1 et l'ORF C2, suite à l'épissage, crée un ORF codant pour la protéine Rep. Les transcrits produits pendant l'infection sont représentés par les flèches noires fines. I1 et I2 sont les deux introns qui ont été décrits. D'après Ruschhaupt et al., 2013.

Chez les *Caulimoviridae*, l'utilisation de l'épissage alternatif n'a été rapporté que pour le virus bacilliforme du tungro du riz (RTBV), le virus de la mosaïque de la scrofulaire (FMV) et le virus de la mosaïque du chou-fleur (CaMV). Dans les cas du RTBV et du FMV, seulement un site donneur et un site accepteur ont été décrits ; un intron unique est épissé sur certaines molécules d'ARN 35S, les ARN non-épissés servant d'ARN pré-génomique. Chez le RTBV, du genre des *Tungrovirus*, l'épissage permet la production d'un ARNm qui servirait à l'expression de l'ORF IV, dont la fonction est encore inconnue (Fütterer et al., 1994). Chez le FMV, qui appartient au genre des *Caulimovirus*, l'épissage conduit à la fusion, en phase, de l'extrémité 5' de l'ORF IV – codant pour la protéine de capsid – avec l'extrémité 3' de l'ORF V – dont le produit est le précurseur de la transcriptase inverse et d'une protéase à acide aspartique (Scholthof et al., 1991) ; l'importance biologique de cette fusion n'est pas connue.

5) Ouverture sur le sujet d'étude

a) Données disponibles dans la littérature scientifique

L'existence de l'épissage alternatif, chez le CaMV, est évoquée pour la première fois dans une publication datée de 1985 où des chercheurs, analysant la population d'ADN viral au cours de l'infection, rapportent la présence de génomes viraux ayant une délétion, chez des choux chinois et des navets infectés par la souche CaMV-S (Hirochika et al., 1985). S'étendant sur 856 nucléotides, la zone délétée est cartographiée entre la fin de l'ORF I et le début de l'ORF II. Il s'avère que la délétion conduit à une fusion, en phase, de la partie 5' de l'ORF I avec la partie 3' de l'ORF II et que ses bordures correspondent aux séquences consensus des sites donneur et accepteur d'épissage. Ainsi, les différents génomes observés témoigneraient d'un événement d'épissage alternatif de type rétention d'intron – se rapporter à la figure 7 de la partie 4)b). Les génomes possédant la délétion seraient produits par rétrotranscription d'un ARN 35S épissé, l'ARN 35S non-épissé étant utilisé pour la production de génomes viraux complets. L'hypothèse de l'implication de l'épissage dans la délétion est renforcée par l'observation, en plantes infectées, d'ARN viraux plus courts d'environ 800 nucléotides que l'ARN

35S « complet ». En outre, des expériences de mutagenèse dirigée sont réalisées. L'inactivation du site donneur d'épissage putatif, par des mutations silencieuses – afin de ne pas modifier la séquence codante de l'ORF I –, conduit à l'utilisation de trois sites donneurs cryptiques et parfois, d'un autre site accepteur situé à proximité immédiate du site accepteur initialement décrit (Hirochika et al., 1985).

Une étude parue en 1995, rapporte l'existence de l'épissage alternatif chez l'ARN 35S de la souche Cabb-S du CaMV (Kiss-László et al., 1995). Quatre sites donneurs, nommés D1 à D4 et localisés dans l'ORF I et dans la région 5' non-traduite de l'ARN 35S, et un site accepteur unique A sont décrits (Figure 12). Selon cette étude, 70% des ARN 35S sont épissés. Le site A est identique au site décrit chez la souche CaMV-S. À l'inverse, aucun des sites donneurs n'est identique au site décrit chez CaMV-S. Seul D3 est identique à un des sites cryptiques ayant émergé suite à la mutation du site donneur initial. L'utilisation des sites donneurs D1 à D3 conduit à la fusion, en phase, de la partie 5' de l'ORF I avec la partie 3' de l'ORF II.

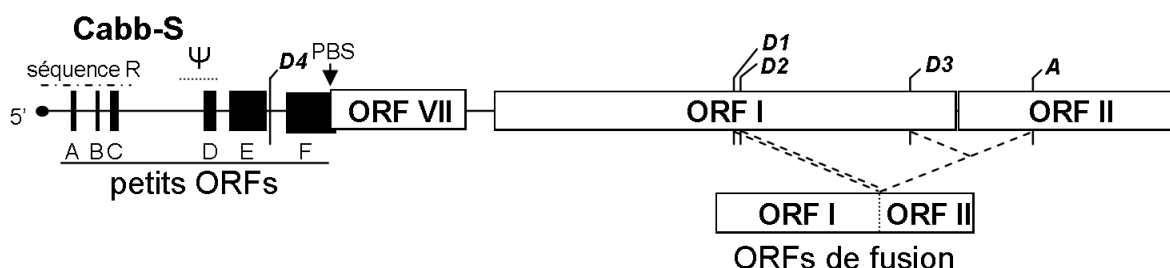


Figure 12 : Carte des 2500 premiers nucléotides de l'ARN 35S de la souche Cabb-S du CaMV. PBS : « *primer binding site* », site d'initiation de la rétrotranscription. Ψ : signal d'encapsidation putatif. La séquence R est une séquence répétée et longue de 180 nucléotides, qui est retrouvée aux deux extrémités de l'ARN 35S. Les événements d'épissage sont représentés par des traits en pointillés. D'après Kiss-László et al., 1995.

Un des résultats majeurs de cet article est la découverte de l'importance qu'à l'épissage pour le cycle viral du CaMV. En effet, des mutations censées rendre le site accepteur unique non-fonctionnel – et par extension empêcher l'épissage de l'ARN 35S – sont létales pour le virus puisque des plantes inoculées avec de tels mutants ne développent pas de symptômes. Il est fait l'hypothèse, à la fin de l'article, que le rôle de l'épissage pourrait être de permettre une meilleure expression des ORF III, IV et V en réduisant le nombre d'ORF situés en amont de l'ORF III. Alors qu'il y a trois ORF présents avant l'ORF III sur l'ARN 35S non-épissé – les ORF VII, I et II –, ce nombre est réduit à deux dans le cas des ARN épissés D1-, D2- et D3-A. L'ORF III est même situé en première position sur l'ARN épissé D4-A (Kiss-László et al., 1995).

Un dernier article, publié en 2004, conclut que l'épissage du CaMV a pour rôle majeur de réguler négativement l'expression de l'ORF II, dont le produit d'expression P2 serait toxique pour le virus en cas de suraccumulation (Froissart et al., 2004). Comme cela a été dit dans l'introduction générale, P2 est une protéine virale impliquée dans la transmission du virus par les pucerons. Elle n'est cependant pas indispensable au déroulement du cycle infectieux dans les plantes hôtes puisqu'il existe des souches défectives pour la transmission – la souche CM4-184 par exemple (Howarth et al., 1981) –, où une bonne partie de l'ORF II est délétée. Le site accepteur A se situe dans l'ORF II, or il a été montré qu'il est indispensable au virus (Kiss-László et al., 1995). Pour aboutir à la conclusion d'un rôle de l'épissage sur l'expression de l'ORF II, les auteurs ont montré que le site accepteur d'épissage n'a d'importance que si le produit d'expression de l'ORF II peut s'accumuler. Ainsi, des mutations censées inactiver le site A, et qui sont létales en contexte normal, n'ont pas d'effet sur le cycle viral si d'autres mutations rendant la protéine P2 instable sont introduites concomitamment (Froissart et al., 2004).

b) Questions sur le rôle biologique de l'épissage alternatif chez le CaMV

Les données présentées dans la littérature scientifique montrent que l'épissage de l'ARN 35S implique, chez la souche Cabb-S, quatre sites donneurs et un site accepteur qui sont tous localisés

dans le premier quart de l'ARN (Kiss-László et al., 1995). Environ 70% des ARN 35S sont épissés, ce qui permettrait une régulation négative de l'expression de l'ORF II dont le produit serait toxique pour le virus (Froissart et al., 2004; Kiss-László et al., 1995). Les 30% d'ARN 35S non-épissé servent à répliquer le génome viral complet ainsi qu'à exprimer la protéine de mouvement (P1) et le facteur assistant de la transmission (P2), deux protéines indispensables à la propagation du CaMV.

Si l'épissage alternatif est nécessaire au CaMV, plusieurs interrogations persistent. La première concerne le nombre de sites donneurs d'épissage. Si le rôle principal de l'épissage alternatif, chez le CaMV, est de réguler négativement l'expression de l'ORF II tout en sauvegardant une fraction d'ARN 35S non-, alors un seul site donneur est nécessaire et ce, quelle que soit sa position. La présence d'un site donneur unique est aussi compatible avec l'hypothèse non-exclusive de l'épissage favorisant l'expression des ORF viraux situés en aval de l'ORF II (Kiss-László et al., 1995). Pourtant, quatre sites donneurs sont utilisés chez Cabb-S, et si l'utilisation d'un seul site donneur a été rapportée chez la souche CaMV-S (Hirochika et al., 1985), cela est probablement dû à des limitations liées aux techniques du moment – ce point sera abordé dans la discussion. On peut ainsi se poser les questions suivantes : pourquoi plusieurs sites donneurs sont utilisés, alors qu'un seul suffirait à effectuer les deux fonctions qui ont été proposées ? L'épissage alternatif est-il impliqué dans d'autres fonctions que celles qui ont été proposées ? Une autre interrogation concerne l'utilisation des trois sites donneurs D1 à D3, qui conduit à la création d'un ORF de fusion entre l'extrémité 5' de l'ORF I et l'extrémité 3' de l'ORF II : des protéines de fusion P1P2 sont donc potentiellement traduites à partir de certains ARN épissés. Ces protéines ont-elles une fonction particulière au cours de l'infection ?

Mon travail de thèse a eu pour but d'acquérir une connaissance plus approfondie de l'épissage alternatif de l'ARN 35S du CaMV, en essayant de répondre aux deux questions énoncées ci-dessus. Le phénomène a été étudié chez deux souches : la souche Cabb-S, utilisée dans deux des trois publications relatives à l'épissage chez le CaMV (Froissart et al., 2004; Kiss-László et al., 1995), et la souche Cabb B-JI, notre souche modèle au laboratoire.

RÉSULTATS

Notre étude sur l'épissage alternatif de l'ARN 35S du CaMV a été abordée de la façon suivante : dans un premier temps, l'existence d'évènements d'épissage putatifs a été déterminée par la recherche *in silico* de sites donneurs et accepteurs d'épissage, dans les séquences génomiques des souches du CaMV. Les sites d'épissage prédits par cette méthode, pour la souche Cabb B-JI, ont ensuite été caractérisés et étudiés *in vivo*.

1) Une analyse *in silico* révèle que des sites d'épissage de l'ARN 35S sont conservés parmi les souches séquencées du CaMV et certains membres du genre des *Caulimovirus*

Dans la littérature, l'épissage alternatif a été suggéré chez la souche CaMV-S (Hirochika et al., 1985), puis décrit chez la souche Cabb-S (Froissart et al., 2004; Kiss-László et al., 1995). Nous avons évalué l'existence potentielle de ce phénomène chez d'autres souches virales, par une analyse *in silico*. Les séquences nucléotidiques de l'ARN pré-génomique 35S des neuf souches dont la séquence complète est publiée sur Genbank ont été alignées sur la séquence de la souche Cabb-S, à l'aide du programme Clustal Omega (Figure 13).

	D4	D1	D2	D3	A
Cabb-S	AGAG GU ACGUU 	AAAG GU AUCUAGCAUACG GU AAGUUU K G N L A Y G K F	AUUGGUAUAACCUUCAAAUAG GU AAC I G N N L Q I G N		AUUGU AG GAAGAAGC I V G R S
U1 snARN 3'-GUCCAUUCA-5'			ORF I		ORF II
CMV-1	AGAG GU AC <u>A</u> UU	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUAG <u>U</u> CUUCAAAUAGGAAAC		AUUGU AG GAAGAAGC
NY8153	AGAG GU ACGUU	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUAG <u>U</u> CUUCAAAUAGGAAAC		AUUGU AG GAAGAAGC
CM1841	AGAG GU ACGUU	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUAG <u>U</u> CUUCAAAUAGGAAAC		AUUGU AG GAAGAAGC
BBC	AGAG GU ACG <u>C</u> U	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUA <u>U</u> CUUCAAAUAGGAAAC		AUUGU AG GAAGAAGC
B29	AGAG GU ACGUU	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUAG <u>U</u> CUUCAAAUAGGAAAC		AUUGU AG GAAGAAGC
W260	AGAG GU ACGUU	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUAG <u>U</u> CUUCAAAUAGGAAAC		AUUGU AG GAAGAAGC
D/H	AGAG GU ACG <u>C</u> U	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUAG <u>U</u> CUUCAAAUAGGAA <u>A</u> U		AUUGU AG GAAGAAGC
Xinjiang	AGAG GU ACG <u>C</u> U	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUAG <u>U</u> CUUCAAAUAGGAAAC		AUUGU AG GAAGAAGC
Cabb B-JI	AGAG GU ACG <u>C</u> U	AAAG GU AUCUAGCAUACG GU AAGUUU	AUUGGUAUAUAG <u>U</u> CUUCAAAUAGGAAAC		AUUGU AG GAAGAAGC

Figure 13 : Alignement de séquences de souches du CaMV. Seules les séquences homologues à celles situées à proximité des sites d'épissage décrits chez Cabb-S sont montrées. Les dinucléotides GU et AG, qui correspondent respectivement aux trois sites donneurs et à l'unique site accepteur, sont en gras. Les nucléotides qui diffèrent de la séquence de la souche Cabb-S sont soulignés. Les acides aminés sont indiqués sous les séquences nucléotidiques situées dans les ORF I et II de Cabb-S. Les interactions entre le snARN U1 et la jonction exon-intron située au niveau du site donneur D4 sont indiquées.

L'alignement de séquences révèle que trois des quatre sites donneurs décrits chez la souche Cabb-S – D4, qui est localisé dans la région 5' non-traduite, D1 et D2, qui sont tous deux localisés dans l'ORF I – ainsi que le site accepteur A sont présents dans l'ARN 35S des différentes souches du CaMV analysées. Les jonctions exon-intron – AG**GU** pour les sites D4, D1 et D3, CG**GU** pour le site D2 – et intron-exon – **AGGA** – sont conservés parmi les différentes souches virales (les dinucléotides GU et AG, qui délimitent les introns splicéosomaux, sont en gras). Enfin, il est intéressant de noter que les sites donneurs D4, D1 et D2 ainsi que le site accepteur A sont également conservés chez les 67 souches dont les séquences génomiques ont été récemment publiées (Yasaka et al., 2014). À l'inverse des autres sites, le site donneur D3 n'est présent que chez la souche Cabb-S (Kiss-László et al., 1995) ; chez les autres souches, le dinucléotide GU est remplacé par un dinucléotide GA (Figure 13). Une séquence UGGU située immédiatement en amont du site D3, et qui est conservée parmi toutes les souches analysées, pourrait indiquer la présence d'un site donneur putatif. Son utilisation réelle par le splicéosome est cependant discutable, car l'appariement avec le snARN U1 de plantes

hôte du CaMV (Filipowicz et al., 1990) est faible – 5 nucléotides sur 9 possibles – si l'on compare avec les sites démontrés expérimentalement – 7 nucléotides sur 9 pour le site donneur D4, par exemple (Figure 13). Quelques sites donneurs, dont certains semblent avoir la capacité de s'apparier avec le snARN U1, et accepteurs putatifs sont retrouvés dans d'autres régions de l'ARN 35S. Par exemple, la séquence GAGGUAUGU, localisée dans l'ORF V de Cabb-S à la position 5308, s'apparie correctement avec le snARN U1 – 7 nucléotides sur 9 – et est conservée parmi toutes les souches testées dans l'alignement de la figure 13.

L'analyse de l'ARN 35S d'autres membres du genre des *Caulimovirus*, avec qui le CaMV partage la même organisation génomique, montre que les sites D1 et/ou D2 ainsi que le site accepteur A sont parfois conservés (Figure 14). Les deux sites donneurs ainsi que le site accepteur A sont présents chez le virus de la mosaïque de la scrofulaire (FMV), le virus de la déformation des feuilles du lamier (LLDV) et le virus de la mosaïque du dahlia (DMV). Ceci laisse à penser que l'épissage décrit chez l'ARN 35S de la souche Cabb-S pourrait avoir lieu chez d'autres *Caulimovirus*. Chez le virus de la gravure de l'œillet (CERV), seul un des sites donneurs est conservé. Le site A est inexistant chez ce virus, le motif AGGA de la souche Cabb-S étant remplacé par la séquence CGGU, mais un dinucléotide AG situé à proximité immédiate est conservé avec le CaMV (Figure 14). Seul le site donneur D1 est conservé entre le virus de la bigarrure des nervures du fraisier (SVBV) et le CaMV, tandis qu'aucun des sites d'épissage présents sur l'ARN 35S du CaMV n'existe chez le virus de la mosaïque de la belle-de-nuit (MiMV) et le virus latent du raifort (HRLV). Enfin, les deux sites d'épissage caractérisés chez le FMV (Scholthof et al., 1991) ne sont pas présents sur l'ARN 35S du CaMV.

	D1	D2	A
Cabb-S	AAAG <u>GU</u> AAUCUAGCAUACG <u>GU</u> AAGUUU		AUUGU <u>AGGA</u> AGA
CERV	CACGGAAACCUAGUCUACG <u>GU</u> AAAUUU		GUUUGCGGUAGA
DMV	AAAG <u>GU</u> AAUUUAUGUUACG <u>GU</u> AAAUUU		GUAGUUGC U AAA
FMV	AGAG <u>GU</u> AAUCUCGCAUAUG <u>GU</u> AAGUUU		AUAAC <u>AG</u> CACGA
LLDAV	AAAG <u>GU</u> AAUUUAGCUUAUG <u>GU</u> AAAUUU		AUAUU <u>AGGC</u> AGA
SVBV	AAAG <u>GU</u> AACCUUCCUACGGGAAACUU		-UUGUAUGUGGA
Cabb B-JI	AAAG <u>GU</u> AAUCUAGCAUACG <u>GU</u> AAGUUU		AUUGU <u>AGGA</u> AGA

Figure 14 : Alignement de séquences de plusieurs membres du genre des *Caulimovirus*. CERV : virus de la gravure de l'œillet ; DMV : virus de la mosaïque du dahlia ; FMV : virus de la mosaïque de la scrofulaire ; LLDV : virus de la déformation des feuilles du lamier ; SVBV : virus de la bigarrure des nervures du fraisier.

En conclusion, les analyses *in silico* que nous avons réalisées suggèrent fortement que l'épissage alternatif de l'ARN pré-génomique 35S est un phénomène conservé parmi les souches du CaMV, qui pourrait aussi avoir lieu chez d'autres espèces appartenant au genre des *Caulimovirus*.

2) L'ARN 35S de la souche Cabb B-JI est épissé alternativement

L'analyse *in silico* réalisée sur les différentes souches du CaMV indique que les sites donneurs D4, D1 et D2, ainsi que le site accepteur A, sont présents chez la souche Cabb B-JI, notre modèle au laboratoire. Contrairement à la souche Cabb-S, Cabb B-JI est capable d'infecter de façon systémique de nombreux écotypes d'*Arabidopsis thaliana* (Cecchini et al., 1998) ainsi que certaines plantes de la famille des *Solanaceae* tel que *Nicotiana benthamiana* (Mével et Kerlan, 1990). Pour savoir si l'ARN 35S de la souche Cabb B-JI est épissé alternativement au cours de l'infection virale, les ARN totaux ont été extraits à partir de plants de navets, à 21 jours après inoculation (dpi, pour « days post inoculation »), séparés par électrophorèse sur gel d'agarose-urée, puis analysés par northern-blot (Figure 15).

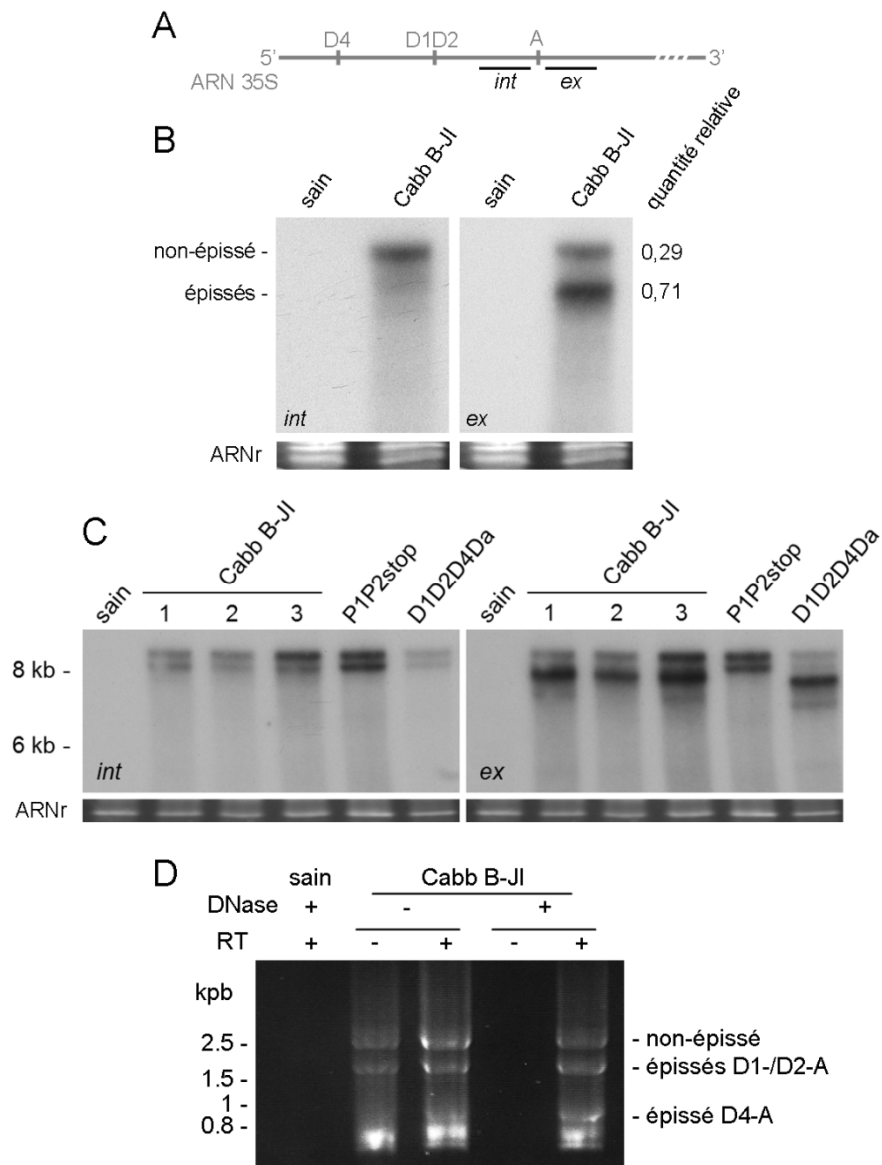


Figure 15 : Analyse par northern-blot, des différentes formes d'ARN 35S. (A) Des sondes oligonucléotidiques ciblant l'intron (*int*) ou l'exon (*ex*) situés autour du site accepteur A putatif ont été utilisées pour détecter l'ARN 35S complet et ses isoformes épissées. (B) et (C) Northern-blot sur les ARN totaux extraits à 21 dpi, provenant de trois réplicats indépendants de navets infectés par Cabb B-JI, ainsi que de navets infectés par le mutant de protéines de fusion P1P2stop et le mutant d'épissage D1D2D4Da. Les ARN ont été séparés par électrophorèse sur gel d'agarose (0,8% pour le B, 0,65% pour le C) - urée (7 M pour le B, 6 M pour le C), transférés sur une membrane de nylon puis hybridés avec les sondes *int* et *ex*, radiomarquées au ^{32}P . (C) RT-PCR sur les ARN totaux extraits à partir de navets, à 21 dpi. La région 5' de l'ARN 35S a été amplifiée, et les produits séparés par électrophorèse sur gel d'agarose.

L'ARN 35S complet – d'une taille d'environ 8200 nucléotides – a été détecté avec une sonde oligonucléotidique (sonde *int* – Figure 15A) ciblant une séquence située immédiatement en amont du site accepteur putatif (Figure 15B). Pour détecter les ARN épissés, une deuxième sonde ciblant la région située immédiatement en aval du site accepteur putatif a été utilisée (sonde *ex* – Figure 15A). L'ARN 35S complet ainsi que des isoformes plus courtes – de taille inférieure à 7600 nucléotides – ont été détectés avec une telle sonde. Aucun signal n'a été révélé avec les ARN totaux de plantes saines. Une quantification par densitométrie des signaux obtenus avec la sonde *ex* a indiqué que l'isoforme non-épissée représente environ 30% de la population globale d'ARN 35S (Figure 15B). Dans certains cas, deux bandes ont été détectées en proportions équivalentes avec la sonde *int* (Figure 15C). Leur signification n'est pas connue mais il pourrait s'agir d'une ou d'autres isoformes épissées impliquant

un autre site accepteur que le site A. Alternativement, ce doublet pourrait résulter de la migration différentielle de deux formes distinctes de l'ARN 35S, bien que la présence d'urée à une concentration quasi dénaturante tende à infirmer cette hypothèse. Des formes distinctes de l'ARN 35S pourraient être expliquées, par exemple, par les diverses conformations que peut adopter la région 5' *leader* (Hemmings-Mieszczyk et al., 1997). Les profils des pistes P1P2stop et D1D2D4Da seront commentés dans les parties 3 et 4.

La région 5' de l'ARN 35S englobant les sites d'épissage putatifs a été amplifiée par RT-PCR – pour « *reverse transcription - polymerase chain reaction* ». L'analyse des produits d'amplification, par électrophorèse sur gel d'agarose, a montré que les amplicons majeurs avaient une taille conforme à la taille théorique des ADN complémentaires (ADNc) correspondant aux isoformes non-épissées (2,5 kilo paires de bases – kpb), épissés D1- et D2-A (1,8 kpb) et épissés D4-A (0,8 kpb) de l'ARN 35S (Figure 15C). Le séquençage de ces amplicons a révélé qu'il s'agissait effectivement des quatre isoformes prédites et que, par conséquent, l'ARN 35S de la souche Cabb B-JI possède un schéma d'épissage alternatif similaire à celui de la souche Cabb-S, exception faite du site donneur D3 qui n'est pas conservé entre les deux souches. En effet, comme le laissait supposer l'analyse *in silico*, aucune isoforme de type D3-A n'a été détectée chez des navets infectés par la souche Cabb B-JI, ce qui confirme que le site donneur d'épissage D3 n'existe pas chez Cabb B-JI. Enfin, nous n'avons pas découvert d'isoformes épissées dans d'autres régions de l'ARN 35S, en particulier dans l'ORF V et ce, malgré la présence du site donneur putatif décrit dans la section précédente. Des produits d'amplification ont été obtenus lorsque les ARN totaux n'ont été ni traités à la désoxyribonucléase (DNase), ni incubés avec la RT – piste « - DNase/-RT » –, indiquant la présence d'ADN viral contaminant. Leur profil de migration était similaire à ceux obtenus dans les pistes « +RT ». De plus, le séquençage a révélé que ces ADN viraux possédaient des délétions identiques à celles engendrées par des événements d'épissage de type D1- et D2-A, ce qui a confirmé la rétrotranscription d'isoformes épissées D1-A et D2-A de l'ARN 35S, au cours de l'infection. De telles délétions ont également été retrouvées dans de l'ADN viral extrait de virion (A. Geldreich, données non-publiées). Il est important de noter que nous avons observé, chez certains clones séquencés, des fragments assez courts – de 600 à 800 nucléotides – contenant des délétions sans lien direct avec l'épissage. Certaines d'entre-elles possédaient des bordures situées autour des deux discontinuités de séquences $\Delta 1$ et $\Delta 3$, des reliquats de la rétrotranscription qui sont situés respectivement au début de l'ORF VII et à la fin de l'ORF II (Richards et al., 1981). Ceci sous-entend que des recombinaisons peuvent avoir lieu entre ces deux discontinuités.

3) Les protéines de fusion P1P2 sont indétectables dans les plantes infectées

La sélection par la machinerie d'épissage, des sites donneurs D1 et D2 – ainsi que D3 dans le cas de la souche Cabb-S – et du site accepteur A, crée des ORF de fusion entre la partie 5' de l'ORF I et la partie 3' de l'ORF II. Alors que la taille de la portion issue de l'ORF I varie en fonction du site donneur sélectionné, celle provenant de l'ORF II est identique à tous les ORF de fusion. Les protéines P1P2 (D1 et D2) résultant de l'expression de ces ORF ont une masse moléculaire théorique d'environ 31 kDa. Malgré de nombreuses tentatives, les protéines de fusion P1P2 n'ont jamais pu être immunodétectées à l'aide d'anticorps polyclonaux anti-P1 et anti-P2, alors que la protéine P1 (46 kDa), son produit de clivage (38 kDa) (Thomas et Maule, 1995) et la protéine P2 (18 kDa) ont pu être révélées avec ces mêmes anticorps (Figure 16A). Les protéines de fusion ont également été recherchées dans des fractions cellulaires enrichies en paroi pecto-cellulosique car la protéine P1 colocalise avec les plasmodesmes, des canaux traversant la paroi pour connecter les cellules végétales les unes aux autres (Stavolone et al., 2005). Les fractions cellulaires testées n'ont pas révélé la présence de protéines P1P2 (Figure 16B). Ces résultats suggèrent que les protéines de fusion P1P2 sont faiblement exprimées et/ou qu'elles sont dégradées rapidement après leur biosynthèse.

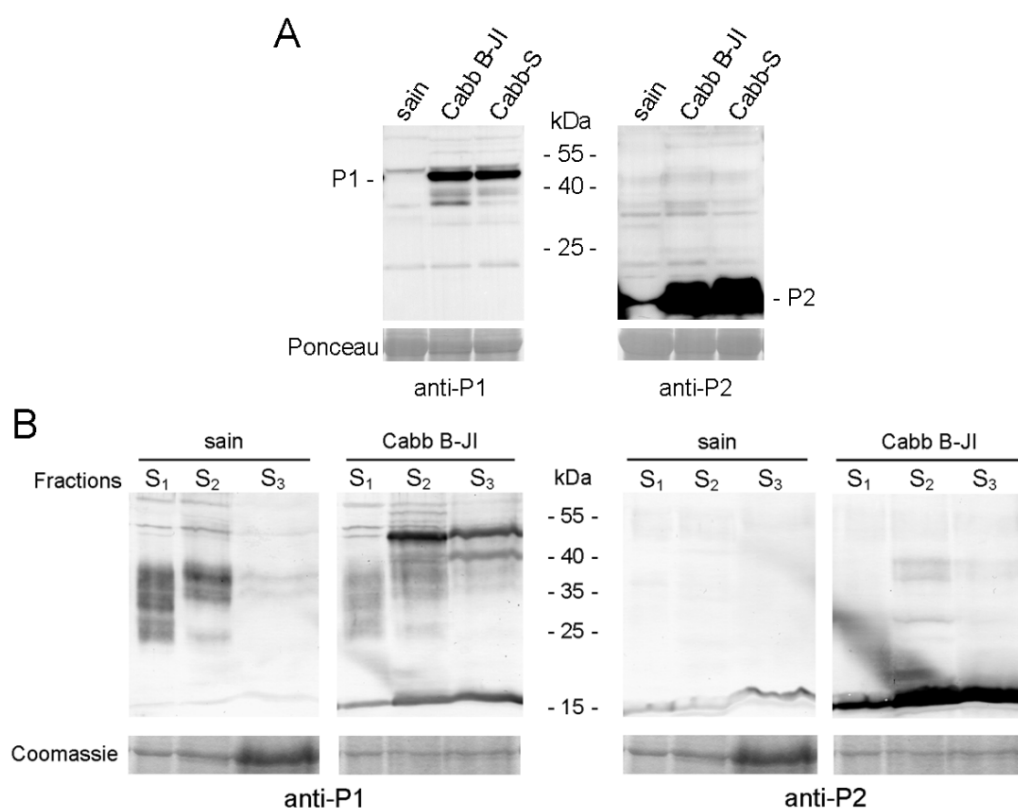


Figure 16 : Immunodétection des protéines de fusion P1P2 chez des navets infectés par les souches Cabb B-JI ou Cabb-S. Les protéines totales (A) ou issues de fractions enrichies en paroi pecto-cellulosique (B) ont été analysées par électrophorèse sur gel de polyacrylamide, en présence de SDS (SDS-PAGE) ; un western-blot a ensuite été réalisé avec des anticorps polyclonaux anti-P1 ou anti-P2. Les fractions S₁, S₂ et S₃ correspondent aux différents stades d'enrichissement en paroi. Les membranes de la partie (A) ont volontairement été surexposées afin de maximiser les chances de détection des protéines P1P2. La coloration des membranes et des gels, respectivement au rouge Ponceau et au bleu de Coomassie a permis d'apprécier les quantités de protéines analysées.

La détection de la protéine P1P2(D2) a été également infructueuse lorsque des cellules de tabac *Nicotiana tabacum* BY-2 – pour « *bright yellow - 2* » – ont été transfectées avec un plasmide recombinant codant pour une protéine de fusion entre P1P2(D2) et la protéine fluorescente verte eGFP – pour « *enhanced green fluorescent protein* » –, alors que cette dernière, exprimée seule, a été parfaitement observée (Angèle Geldreich, données non-publiées). Par ailleurs, l'expression de la P1P2(D2) fusionnée aux étiquettes HA- ou myc- a été testée par agroinfiltration, sur feuilles de *Nicotiana benthamiana*. Deux constructions codant pour la GFP étiquetée avec HA- ou myc- ont été utilisées en parallèle, comme témoins positifs. Les deux protéines HA- et myc-GFP ont été détectées avec des anticorps dirigés contre leurs étiquettes, dans toutes nos expériences d'expression transitoire. À l'inverse, des protéines ayant une masse moléculaire similaire à celle des P1P2(D2) étiquetées (environ 39 kDa) n'ont été mises en évidence que de rares fois et ce, en très faibles quantités (Figures 17 et 19). La plupart du temps, les P1P2(D2) étiquetées n'ont pas été immunodétectées dans des extraits protéiques totaux préparés à partir de feuilles agroinfiltrées (Figure 17A) et ce, alors que les ARNm correspondants ont pu être détectés par RT-PCR (Figure 17B).

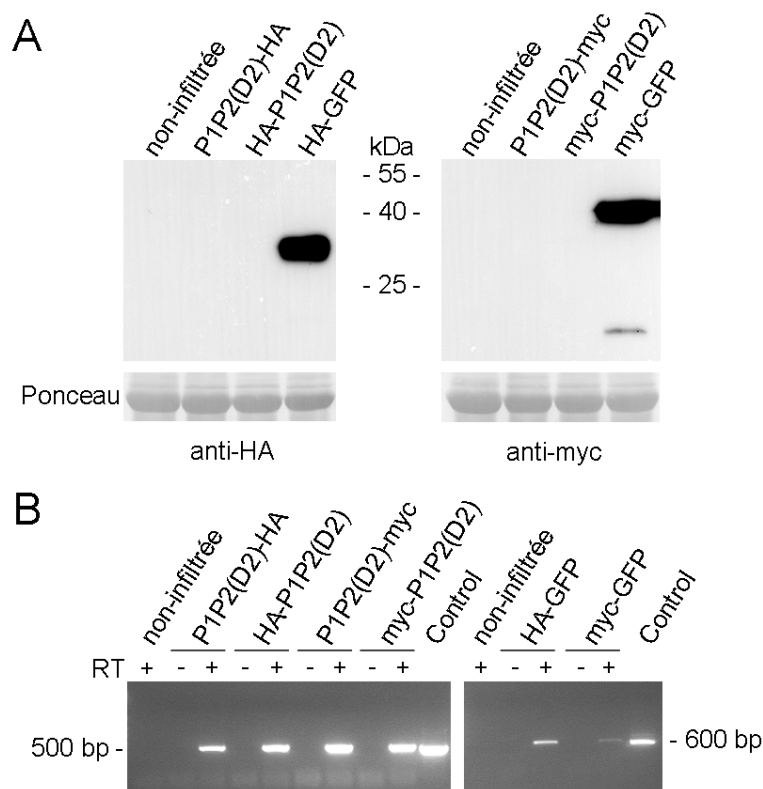


Figure 17 : Immunodétection des protéines de fusion P1P2(D2) étiquetées HA- et myc- et exprimées de façon transitoire, par agroinfiltration de feuilles de *Nicotiana benthamiana*. (A) Les protéines totales ont été extraites à partir des feuilles agroinfiltrées 2,5 jours après infiltration et les protéines de fusion P1P2(D2) ont été recherchées par western-blot, avec des anticorps anti-HA ou anti-myc. Des feuilles agroinfiltrées avec des constructions codant pour HA-GFP ou myc-GFP ont été utilisées comme contrôles positifs. Les images résultent d'une surexposition volontaire des membranes, afin d'augmenter les chances de détection des protéines P1P2. (B) La présence de transcrits correspondant aux différentes constructions agroinfiltrées a été vérifiée par RT-PCR, à partir d'ARN totaux extraits à partir de feuilles agroinfiltrées.

Il a été montré que P2 interagit avec elle-même (Hébrard et al., 2001), ainsi qu'avec P3 (Leh et al., 1999) et P6/TAV (Lutz et al., 2012). En plus d'oligomériser pour former les tubules permettant le mouvement viral de cellule à cellule (Perbal et al., 1993), P1 peut interagir avec P3 (Stavolone et al., 2005), via son extrémité carboxy-terminale, et P6/TAV (Hapiak et al., 2008). Nous avons fait l'hypothèse que les protéines de fusion pourraient interagir avec les partenaires viraux des protéines P1 et/ou P2 complètes. Pour évaluer les possibilités d'interactions entre les protéines P1P2 et d'autres protéines virales, des expériences de far western-blot ont été réalisées. La protéine P1P2(D2) fusionnée à la GST – pour « glutathion S-transférase » – et exprimée chez *Escherichia coli* a ainsi été incubée avec des protéines virales P1, P3 et P6/TAV recombinantes et radiomarquées au ^{32}P , ainsi qu'avec des particules virales purifiées. *In vitro*, GST-P1P2(D2) a interagit avec les particules virales purifiées, les protéines P2 et P3 mais ni avec P1, ni avec P6/TAV (Figure 18). Aucune interaction n'a été détectée entre la GST seule et les protéines/particules virales. Des expériences de GST pull-down et de double-hybride, visant à confirmer les interactions observées par far western-blot, sont en cours de réalisation.

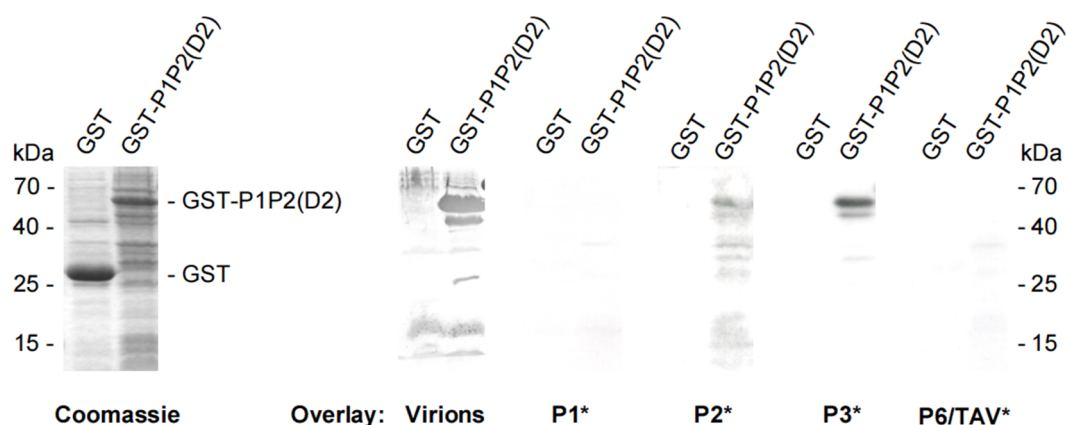


Figure 18 : Tests d'interaction par far western-blot, entre la GST-P1P2(D2) exprimée en système bactérien, et des protéines virales recombinantes et des particules purifiées. La GST-P1P2(D2) et la GST seule ont été fractionnées par SDS-PAGE, transférées sur membrane de polyfluorure de vinylidène (PVDF), puis cette dernière a été incubée avec des virions purifiés ou des protéines virales P1, P3 ou P6/TAV recombinantes et radiomarquées au ^{32}P (overlay).

Le fait que les protéines de fusion P1P2 ne soient pas détectées au cours de l'infection mais puissent interagir *in vivo* avec certains partenaires viraux de P1 et P2, suggère que ces protéines sont rapidement dégradées après leur expression, possiblement afin de prévenir la mise en place de telles interactions qui interféreraient avec les fonctions de P1 et/ou P2. Pour savoir si la voie de l'ubiquitine/protéasome est impliquée dans la dégradation des protéines de fusion, nous avons réalisé des cinétiques d'inhibition du protéasome. Deux constructions codant pour deux versions étiquetées de la protéine P1P2(D2) ont été agroinfiltrées dans des feuilles de *Nicotiana benthamiana*. 1,5 jour après infiltration, les feuilles ont été traitées avec du MG132, un inhibiteur de la voie de l'ubiquitine/protéasome, ainsi qu'avec de la cycloheximide, un inhibiteur de la traduction eucaryote. La protéine DELLA RGA – pour « repressor of the *ga1-3 mutant* » –, liée à une étiquette HA-, a été utilisée comme témoin positif. Cette protéine, dont le rôle est de réprimer la voie de signalisation des gibbérellines, est connue pour être une cible de la voie de l'ubiquitine/protéasome (Feng et al., 2008).

Alors que la protéine RGA-HA a été stabilisée par le traitement au MG132 (Figure 19), il n'a pas été observé de tel effet pour la protéine de fusion P1P2(D2) exprimée chez *N. benthamiana*. Même si la forte difficulté à exprimer de façon reproductible la protéine de fusion P1P2(D2) ne nous permet pas d'éliminer complètement l'hypothèse d'une dégradation par la voie de l'ubiquitine/protéasome, il semble que les protéines P1P2 soient intrinsèquement instables.

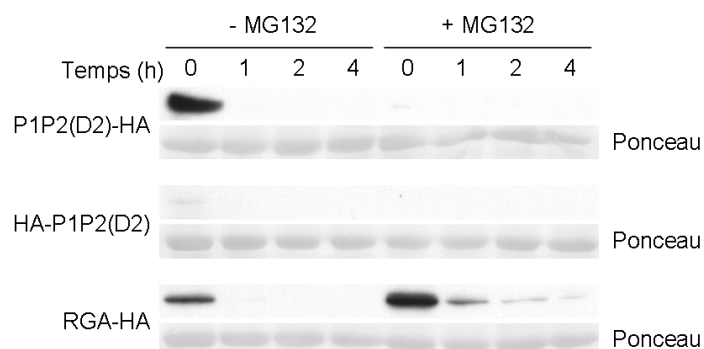


Figure 19 : Cinétique d'inhibition de la voie de l'ubiquitine/protéasome, par traitement au MG132. Des feuilles de *Nicotiana benthamiana* ont été agroinfiltrées avec deux constructions codant pour la P1P2(D2) étiquetée HA-. 1,5 jour après infiltration, les feuilles ont été incubées avec du MG132 puis, 1,5 h après le début d'incubation, de la cycloheximide a été ajoutée au mélange. Le moment de l'ajout de la cycloheximide correspond au temps 0. La présence des protéines étiquetées HA- a été recherchée par western-blot, à l'aide d'anticorps anti-HA.

Pour évaluer l'importance des protéines de fusion P1P2 au cours du cycle infectieux, nous avons inséré un codon d'arrêt de la traduction immédiatement en aval du site accepteur A – mutant P1P2stop, où la séquence **AGGAAGAAGCUUACUCGGA** est remplacée par **AGGAAGAAGCUUACUCUGA**. L'inoculation de plants de navets avec ce mutant a conduit à l'apparition de symptômes systémiques identiques à ceux provoqués par la souche Cabb B-JI sauvage ; aucun délai dans l'apparition de ces symptômes n'a été observé. Par ailleurs, les protéines de capsid résultant de clivages du précurseur P4 ont été retrouvées en même quantité chez le mutant et le virus sauvage (Figure 20). Ceci indique que, dans nos conditions expérimentales, les protéines de fusion, ou du moins la portion issue de l'ORF II, n'ont pas d'importance pour le cycle infectieux. L'analyse des formes d'ARN 35S présentes chez les plantes infectées par le mutant P1P2stop a révélé une absence relative de formes épissées (Figure 15C, échantillon P1P2stop avec la sonde ex). Ce résultat surprenant indique que les protéines de fusion P1P2 et/ou P2 – puisque le codon stop inséré conduit à l'expression d'une P2 tronquée au deux tiers – pourraient réguler positivement l'épissage de l'ARN 35S.

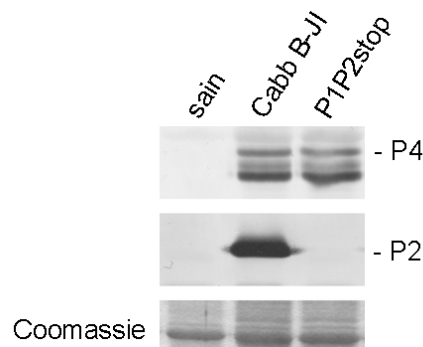


Figure 20 : Immunodétection des protéines P4 et P2 chez des navets infectés par le mutant P1P2stop et le virus Cabb B-JI sauvage. Les protéines totales ont été extraites à 25 dpi et analysées par western-blot, comme décrit précédemment.

4) Un site donneur cryptique est utilisé lorsque les sites D1, D2 et D4 sont mutés

Des études précédentes ont suggéré que l'épissage de l'ARN 35S est essentiel au CaMV puisqu'il permet de réguler négativement l'expression des protéines P1 et P2, dont la suraccumulation – en particulier dans le cas de P2 – serait délétère pour le cycle infectieux (Froissart et al., 2004; Kiss-László et al., 1995). Cependant, étant donné qu'un site donneur serait suffisant pour assurer cette régulation négative d'expression, et puisque les sites donneurs

D1, D2 et D4 sont conservés parmi les différentes souches du CaMV, nous avons émis l'hypothèse que l'épissage alternatif pourrait avoir d'autres fonctions régulatrices pour le virus. Pour évaluer l'importance biologique des différents sites donneurs, nous les avons inactivés un à un ou en combinaison, par mutagenèse dirigée sur le vecteur viral pMD324 contenant le génome complet de la souche Cabb B-JI. Les mutations introduites au niveau des jonctions exon-intron devaient abolir l'interaction avec le snARN U1 tout en maintenant, dans le cas des sites D1 et D2, la séquence codante de l'ORF I, afin de ne pas altérer la fonction de la protéine de mouvement P1 (Figure 21A). En effet, ces deux sites font partie de codons codant pour deux glycines qui sont conservées chez d'autres *Caulimovirus* comme le FMV et le CERV. De plus, il a été observé que la mutation de résidus situés à leur proximité immédiate conduisait à une incapacité à développer une infection systémique (Thomas et Maule, 1995).

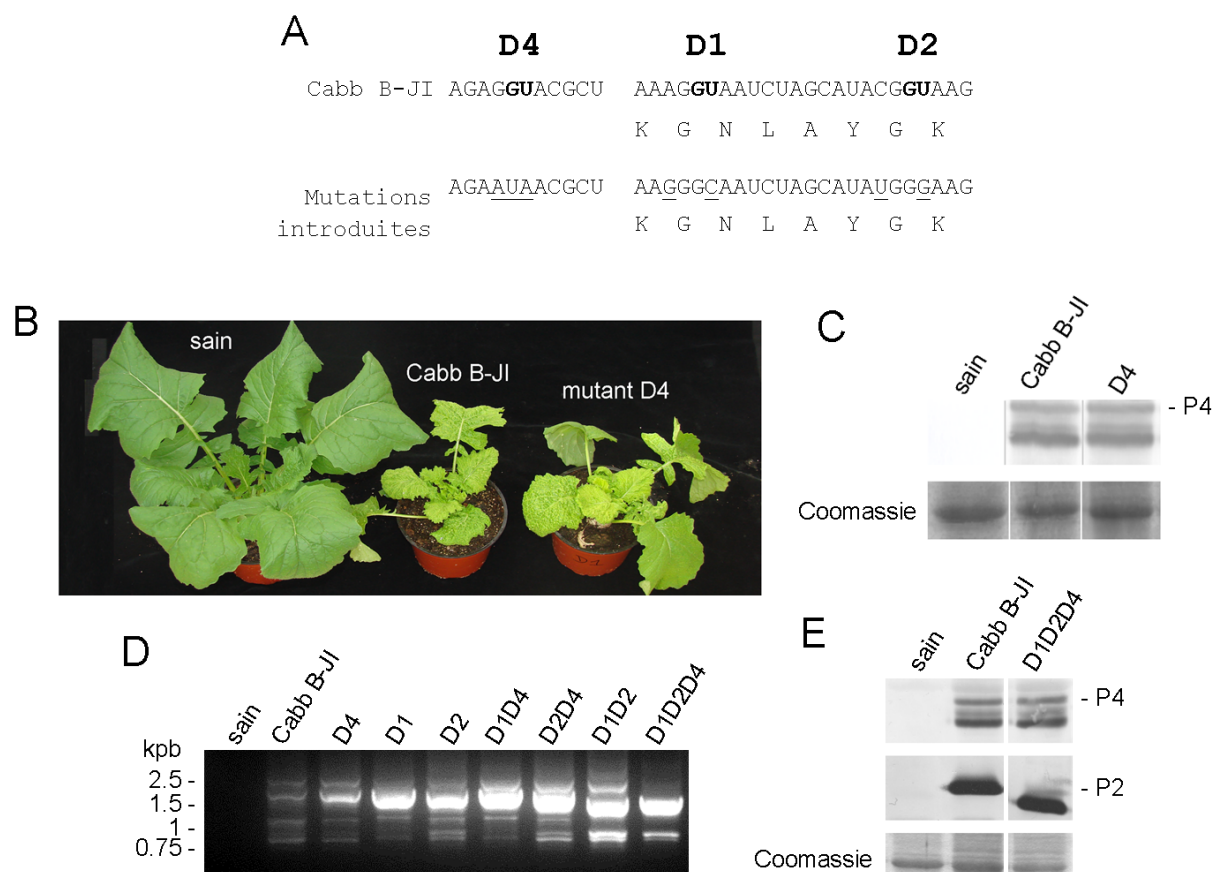


Figure 21: Les mutations des sites donneurs de la souche Cabb B-JI sont contrebalancées par l'utilisation d'un site donneur cryptique. (A) Les jonctions exon-intron des sites donneurs ont été mutées indépendamment ou en combinaison. Les sites donneurs GU sont en gras, et les nucléotides mutés sont soulignés. Les acides aminés correspondant à la séquence codante de l'ORF I sont indiqués sous la séquence nucléotidique des sites D1 et D2. (B) Symptômes présentés par des navets infectés par le mutant du site donneur D4 et le virus sauvage, à 25 dpi, et (C) immunodétection de la protéine de capsid P4. (D) Analyse, par RT-PCR, du profil d'épissage des mutants des sites donneurs. Les ARN totaux de navets infectés ont été extraits à 21 dpi et la région 5' de l'ARN 35S – soit les 2500 premiers nucléotides – a été amplifiée. Les amplicons ont été séparés par électrophorèse sur gel d'agarose. (E) Immunodétection de la protéine de capsid P4 et de la protéine P2, chez des navets infectés par le triple mutant D1D2D4 et le virus sauvage, à 25 dpi. Le décalage observé pour la protéine P2 chez le mutant D1D2D4 est un artefact de migration.

Des mutants simples, doubles et un mutant triple ont été générés. Les plasmides résultants ont été inoculés mécaniquement à des plants de navets au stade trois-feuilles. Dix plantes ont été inoculées avec chaque mutant d'épissage, lors de quatre expériences indépendantes. Tous les mutants testés se sont révélés être infectieux chez le navet, y compris le triple mutant D1D2D4 pour lequel on s'attendait à ce que les mutations introduites soient létales pour le virus, du fait qu'elles abolissent théoriquement l'épissage (Froissart et al., 2004; Kiss-László et al., 1995). Deux semaines après inoculation, tous les plants de navets ont présenté des symptômes systémiques – mosaïques foliaires, jaunissement des nervures et rabougrissement – identiques à ceux des plantes inoculées avec le génome sauvage (Figure 21B). En outre, aucun délai dans l'apparition des symptômes n'a été observé entre les différents mutants et le virus sauvage. Enfin, aucune différence dans l'accumulation du précurseur P4 des protéines de capsid n'a mise en évidence (Figures 21C et 21E). Le séquençage de l'ADN viral extrait à partir des feuilles systémiques n'a pas révélé de réversion de mutation. Les ARN totaux des plantes infectées par les différents mutants ou le virus sauvage ont été extraits, et la région 5' de l'ARN 35S englobant les sites d'épissage a été amplifiée par RT-PCR. Les produits d'amplification obtenus ont ensuite été analysés par électrophorèse sur gel d'agarose. Le profil global d'épissage présenté par les différents mutants est assez similaire, avec un amplicon ma-

jeur d'environ 1,8 kpb, et plusieurs amplicons mineurs dont un d'environ 2,4 kpb correspondant à l'ADNc de l'ARN 35S non-épissé. Il est intéressant de noter que l'amplicon principal obtenu avec le mutant D1D2 est plus court que celui observé chez les autres mutants – 1,6 kpb contre 1,8 kpb. Alors que le triple mutant D1D2D4 ne devrait conduire qu'à un amplicon de 2,4 kpb, l'amplicon de 1,6 kpb est également observé. L'amplicon de 1,6 kpb a été extrait du gel d'agarose, cloné dans un vecteur pGEM-T Easy (Promega) et séquencé. La séquence obtenue a révélé qu'il s'agissait d'une nouvelle isoforme épissée de l'ARN 35S impliquant un site donneur que nous avons nommé Da, localisé dans l'ORF I à la position 653 (Tableau 2 et Figure 22A), et le site accepteur A. Cet événement d'épissage conduit, comme lorsque les sites D2 et D3 sont utilisés, à une fusion en phase des ORF I et II. La détection d'une quantité similaire de P2, chez des navets infectés par le virus sauvage et le triple mutant (Figure 21E), laisse d'ailleurs penser à une régulation de l'expression de l'ORF II équivalente. Ce site Da a été muté dans le contexte du triple mutant D1D2D4 – la séquence AGGU a été remplacée par GGGC, sans que cela modifie la séquence codante de P1 – et le quadruple mutant D1D2D4Da s'est également révélé être infectieux, suggérant, ici aussi, l'utilisation d'un ou plusieurs sites cryptiques. L'analyse des ARN viraux par northern-blot est allée dans ce sens puisque l'utilisation de la sonde *ex*, ciblant toutes les formes de l'ARN 35S, a révélée des bandes de taille inférieure à l'ARN 35S non-épissé, identifié quant à lui spécifiquement par la sonde *int* (Figure 15C, échantillon D1D2D4Da). La différence dans le profil des bandes de taille inférieure à l'ARN 35S non-épissé, entre l'échantillon D1D2D4Da et les trois échantillons Cabb B-JI (Figure 15C, sonde *ex*) tend, d'ailleurs, à indiquer la présence d'isoformes épissées propres au quadruple mutant.

En conclusion, nos résultats montrent que les sites donneurs D1, D2 et D4 de l'ARN 35S ne sont pas importants, en soi, pour l'infectiosité et la virulence du CaMV puisque leur inactivation ne provoque pas de délais dans l'apparition des symptômes, qui sont d'ailleurs identiques à ceux provoqués par le CaMV sauvage, ni de variations d'accumulation de la protéine de capsid P4 dans les feuilles systémiques. À l'inverse, l'épissage, en tant que tel, est important puisque le virus utilise des sites donneurs cryptiques pour pallier leur inactivation.

5) L'ARN 35S des souches Cabb-S et Cabb B-JI possède un profil d'épissage complexe

Comme dit précédemment, un quadruple mutant D1D2D4Da de la souche Cabb B-JI s'est également révélé infectieux, lorsqu'il a été inoculé à des plants de navets, ce qui a suggéré l'existence d'autres sites d'épissage sur l'ARN 35S. Ceci nous a amené à étudier en détail l'épissage alternatif au sein de l'ARN 35S de la souche Cabb B-JI et ce, en nous focalisant sur sa partie 5' puisque l'amplification par RT-PCR des régions localisées en aval du site accepteur n'a pas révélé l'existence d'autres formes épissées. La région 5' de l'ARN 35S – de l'extrémité 5' à la fin de l'ORF II – a été amplifiée par RT-PCR, à partir d'ARN totaux extraits de navets infectés et traités à la DNase. Après clonage des amplicons dans le vecteur pGEM-T Easy, la taille des fragments insérés a été évaluée par électrophorèse sur gel d'agarose. Cette étape de triage a permis d'éliminer des clones contenant des inserts identiques et surreprésentés. Les clones sélectionnés, à la suite de ce tri, ont été séquencés ; les séquences obtenues ont été alignées avec celle de l'ARN 35 non-épissé. En plus des isoformes qui ont été déjà mentionnées – non-épissée, D4-A, Da-A, D1-A et D2-A –, nous avons détectés 11 nouvelles isoformes épissées provenant de l'utilisation de deux autres sites donneurs – Db et Dc – et de six sites accepteurs – Aa à Af – (Tableau 2, Figures 22A et 23). Tous ces nouveaux sites d'épissage possèdent les séquences consensus GU – sites donneurs – ou AG – sites accepteurs – ; ils ont été nommés selon l'ordre de leur découverte.

Le site donneur Db est localisé dans le petit ORF E, à proximité du site D4, au sein de la région 5' non-traduite. Le site donneur Dc, ainsi que les six nouveaux sites accepteurs, sont tous localisés au sein de l'ORF I (Figure 22A). Les différentes isoformes épissées que nous avons analysées sont le résultat de cinq événements de simple, neuf de double et un de triple épissage. Ils contiennent tous au moins un ORF de fusion impliquant le petit ORF E, les ORF I et II, ainsi que des séquences

d'autres cadres de lecture localisés dans la région des ORF I et II (Figure 23). Il est intéressant de noter que l'utilisation du site donneur Db, qui se situe dans le petit ORF E de la région 5' non-traduite, conduit aussi à la formation d'ORF de fusion comme c'est le cas chez les isoformes Db-A et Db-Ab-D2-A (Figure 23).

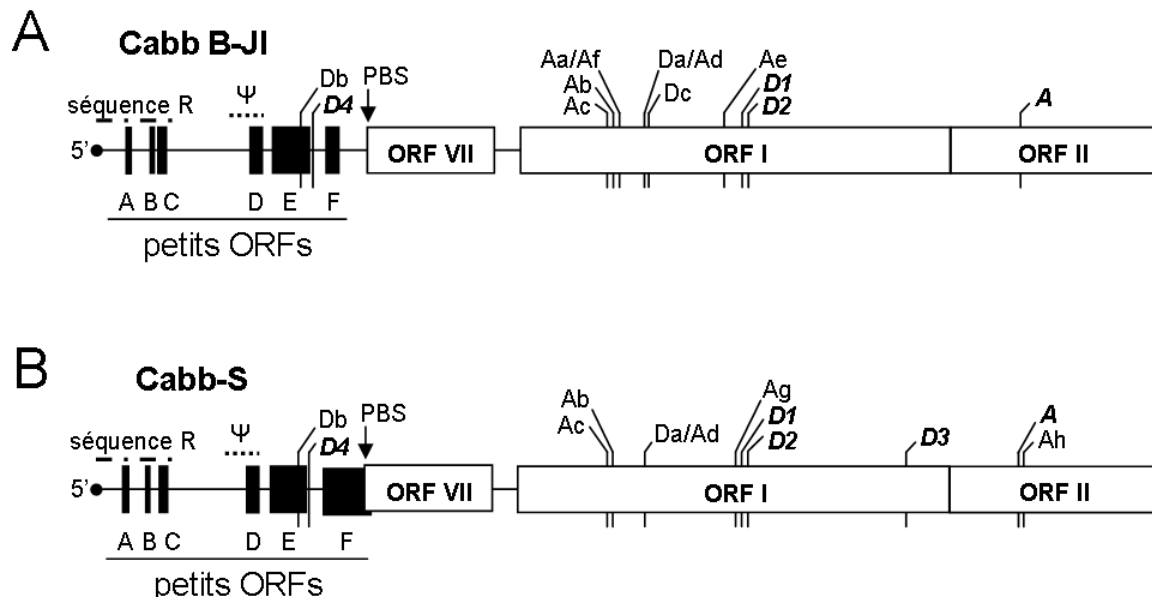


Figure 22 : Sites donneurs et accepteurs d'épissage retrouvés chez les souches Cabb B-JI et Cabb-S. Les 2500 premiers nucléotides de l'ARN 35S des souches Cabb B-JI (A) et Cabb-S (B) sont schématisés. Les sites d'épissage découverts précédemment sont en gras italique. Le signal d'encapsidation putatif (Ψ) et le site d'initiation de la rétrotranscription (PBS) sont représentés. La séquence R correspond à une séquence répétée longue d'environ 180 nucléotides, qui est retrouvée aux deux extrémités de l'ARN 35S.

	Cabb B-JI	Cabb-S
Sites donneurs d'épissage (position)	D1 (872); D2 (887); D4 (7915); Da (653); Db (7890); Dc (658)	D1 (872); D2 (887); D3 (1247); D4 (7909); Da (653); Db (7884)
Sites accepteur d'épissage (position)	A (1509); Aa (584); Ab (577); Ac (565); Ad (652); Ae (830); Af (587)	A (1508); Ab (577); Ac (565); Ad (652); Ag (864); Ah (1515)

85 et 87 clones, issus de deux expériences indépendantes, ont respectivement été séquencés pour les souches Cabb B-JI et Cabb-S. Les sites Da à Dc et Aa à Af ont été nommés selon l'ordre de leur découverte chez la souche Cabb B-JI.

Tableau 2 : Sites donneurs et accepteurs d'épissage retrouvés chez les souches Cabb B-JI et Cabb-S.

Lorsque nous avons analysé les ARN viraux extraits de plants de navets infectés par la souche Cabb-S, nous avons également observé un profil d'épissage complexe dans la partie 5' de l'ARN 35S puisque nous avons pu caractériser 12 isoformes épissées différentes (Figure 23) ; seule l'isoforme D3-A n'a pas été détectée, malgré sa description par Kiss-László et al.. La plupart des sites d'épissage ont une position identique chez les souches Cabb-S et Cabb B-JI (sites Da, Db, Ab, Ac, Ad ; Figures 22A, 22B et Tableau 2). Dans de très rares cas, nous avons constaté l'utilisation d'un site accepteur (Ah, à la position 1515) situé en aval du site accepteur A – en position 1508 – considéré précédemment comme unique ; nous n'avons jamais mis en évidence l'utilisation de ce site chez la souche Cabb B-JI, sauf lorsque nous avons inactivé le site accepteur A – ceci sera décrit la partie

suivante. L'étude, par alignement de séquences, de la conservation potentielle des nouveaux sites détectés indique que tous sont présents parmi les différentes souches virales.

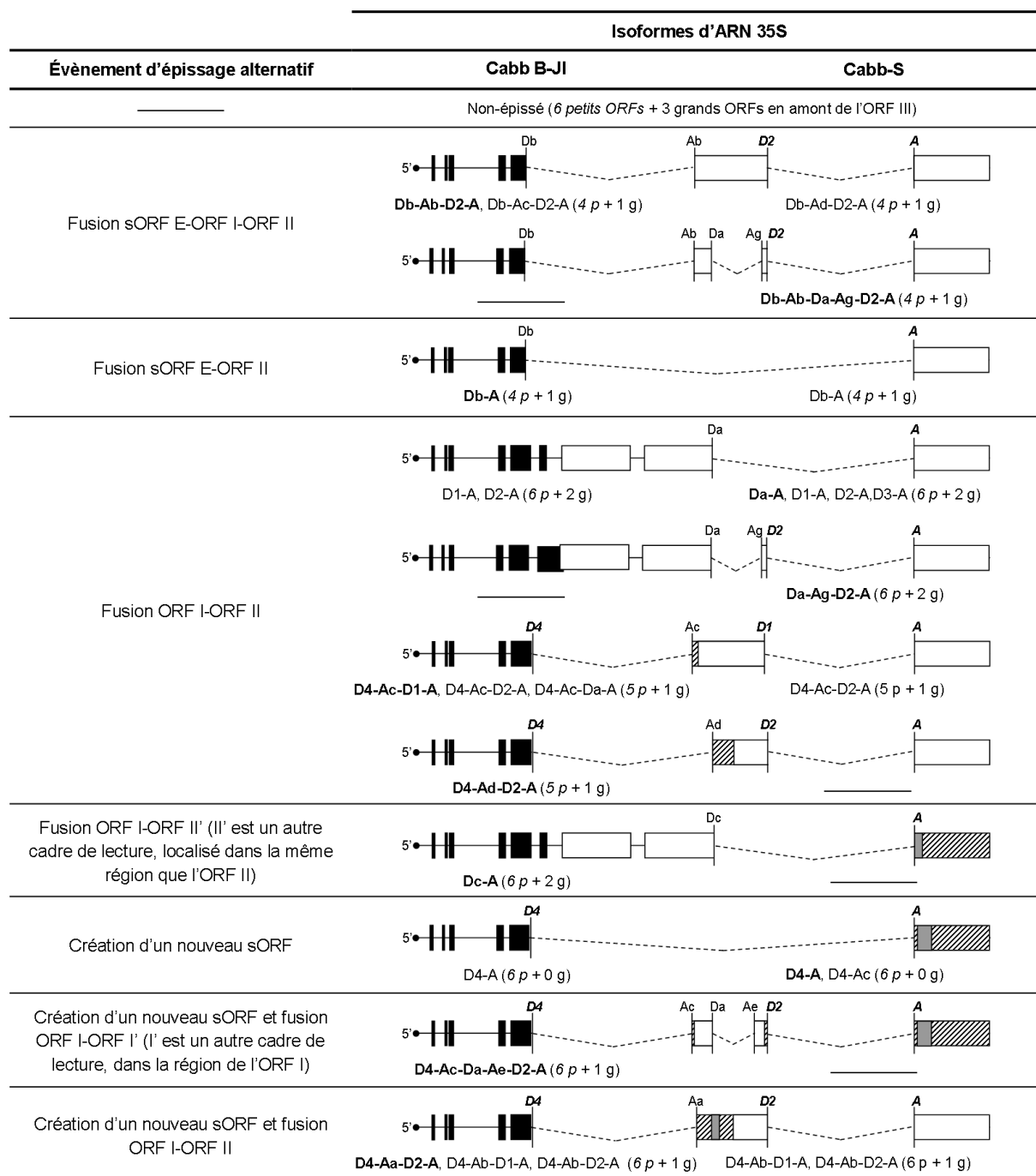


Figure 23 : Évènements d'épissage et isoformes correspondantes, chez les souches Cabb B-JI et Cabb-S. L'épissage est représenté par des lignes pointillées. Pour chaque type d'évènement d'épissage, les isoformes sont indiquées et une ou plusieurs isoformes représentatives – indiquées en gras – sont schématisées. Le nombre de petits (p) et grands (g) ORF situés en amont de l'ORF III est donné entre parenthèses. Les régions des ORF I et II qui, chez certaines isoformes, ne sont plus codantes du fait d'un décalage du cadre de lecture, sont hachurées. Les séquences codantes qui sont localisées dans les régions des ORF I et II, mais qui possèdent un cadre de lecture différent, sont grisées.

Notre analyse des différentes isoformes de l'ARN 35S montre clairement que ce dernier est l'objet, chez les souches Cabb B-JI et Cabb-S, d'un d'épissage alternatif complexe qui implique de

multiples sites donneurs et accepteurs situés majoritairement dans l'ORF I. La conservation de tous ces sites, chez les différentes souches virales, indique que cet épissage complexe semble être une propriété commune à l'ensemble des souches du CaMV.

6) Des sites accepteurs cryptiques peuvent être activés pour maintenir l'épissage de l'ARN 35S

Nos travaux montrent que l'inactivation de sites donneurs d'épissage peut être contrecarrée par l'utilisation d'une multitude d'autres sites. La découverte, parmi les nouveaux sites que nous avons caractérisés, d'un site accepteur situé à proximité du site A (Figure 22B, site accepteur Ah) suggère que ce site Ah – ou d'autres sites accepteurs – peut suppléer le site accepteur A et par conséquent, que ce dernier n'est pas aussi crucial pour l'infectiosité que ce qui était décrit auparavant (Froissart et al., 2004; Kiss-László et al., 1995). Pour vérifier cette hypothèse, nous avons introduit une mutation considérée par Kiss-László et al. (1995) et Froissart et al. (2004) comme suffisante pour empêcher l'utilisation du site A : ainsi, la jonction intron/exon **AGG** a été mutée en **AGA** et le mutant correspondant a été nommé B-JI ma3 (Figure 24A). Dans nos conditions expérimentales – c'est-à-dire une inoculation mécanique de nos plantes avec 8 µg d'ADN viral linéarisé et une culture en serre, en conditions

de jours longs –, les plants de navets inoculés avec le mutant B-JI ma3 ont développé des symptômes systémiques identiques à ceux provoqués par la souche Cabb B-JI sauvage (Figure 24B). L'accumulation des protéines virales dans les feuilles systémiques s'est révélée similaire entre la souche sauvage et le mutant B-JI ma3, à l'exception notable de la protéine P2, laquelle est quasiment indétectable chez les plantes infectées par le mutant (Figure 24D). Pour déterminer si le comportement observé était dû à la souche virale, nous avons évalué l'infectiosité d'un mutant de la souche Cabb-S contenant la même mutation (mutant S ma3 ; Figure 24A). En contradiction avec les études publiées précédemment (Froissart et al., 2004; Kiss-László et al., 1995), ce mutant s'est avéré être aussi infectieux que le virus sauvage (Figure 24B). Comme pour le mutant B-JI ma3, la protéine P2 s'est trouvée être difficilement détectable chez les plantes infectées avec le mutant S ma3, alors qu'aucune différence n'a été observée pour les autres protéines virales testées, par rapport aux plantes infectées avec la souche Cabb-S (Figure 24D). Pour les deux souches Cabb B-JI et Cabb-S, le séquençage de la partie 5' de l'ARN 35S de la descendance virale n'a montré aucune réversion de mutation. De même, aucune autre modification qui aurait pu expliquer l'infectiosité de ces mutants, telle qu'une délétion ou l'insertion d'un codon d'arrêt de la traduction dans l'ORF II, n'a été observée. Fait surprenant, l'amplification par RT-PCR et le séquençage de la partie 5' de l'ARN 35S ont révélé que le site accepteur A – qui est en position 1509 chez Cabb B-JI, et 1508 chez Cabb-S – était toujours actif et ce, en dépit de la mutation introduite (Figures 24A et 24C, clones 14 et 43). Par ailleurs, chez certaines isoformes épissées que nous avons caractérisées, le dinucléotide AG situé immédiatement en aval du site A – c'est-à-dire le site accepteur Ah – a été utilisé en tant que site accepteur d'épissage (Figures 24A et 24C, clone 17).

La mutation ma3 n'affecte pas directement l'utilisation du site accepteur A. Pour inactiver totalement ce site, chez la souche Cabb B-JI, nous avons introduit une mutation au sein même du dinucléotide AG : le site accepteur **AGG** a été muté en UGC (mutant B-JI mutA ; Fig. 21A). Comme avec les mutants testés précédemment, les plants de navets inoculés avec le mutant B-JI mutA ont présenté des symptômes systémiques de nature et de sévérité identiques à ceux produits par la souche sauvage (Figure 24B). Aucun délai dans l'apparition de ces symptômes n'a été noté entre le mutant et le virus sauvage. En outre, les plantes inoculées avec le mutant B-JI mutA ont accumulé des quantités de protéines virales similaires, y compris de P2, à celles des plantes inoculées avec le virus sauvage (Figure 24D). Enfin, le séquençage des ARN viraux a confirmé l'absence de réversion de mutation du site A, et a révélé que deux dinucléotides AG situés en aval du site muté, dont le site Ah, sont occasionnellement utilisés comme sites accepteurs d'épissage (Figure 24A).

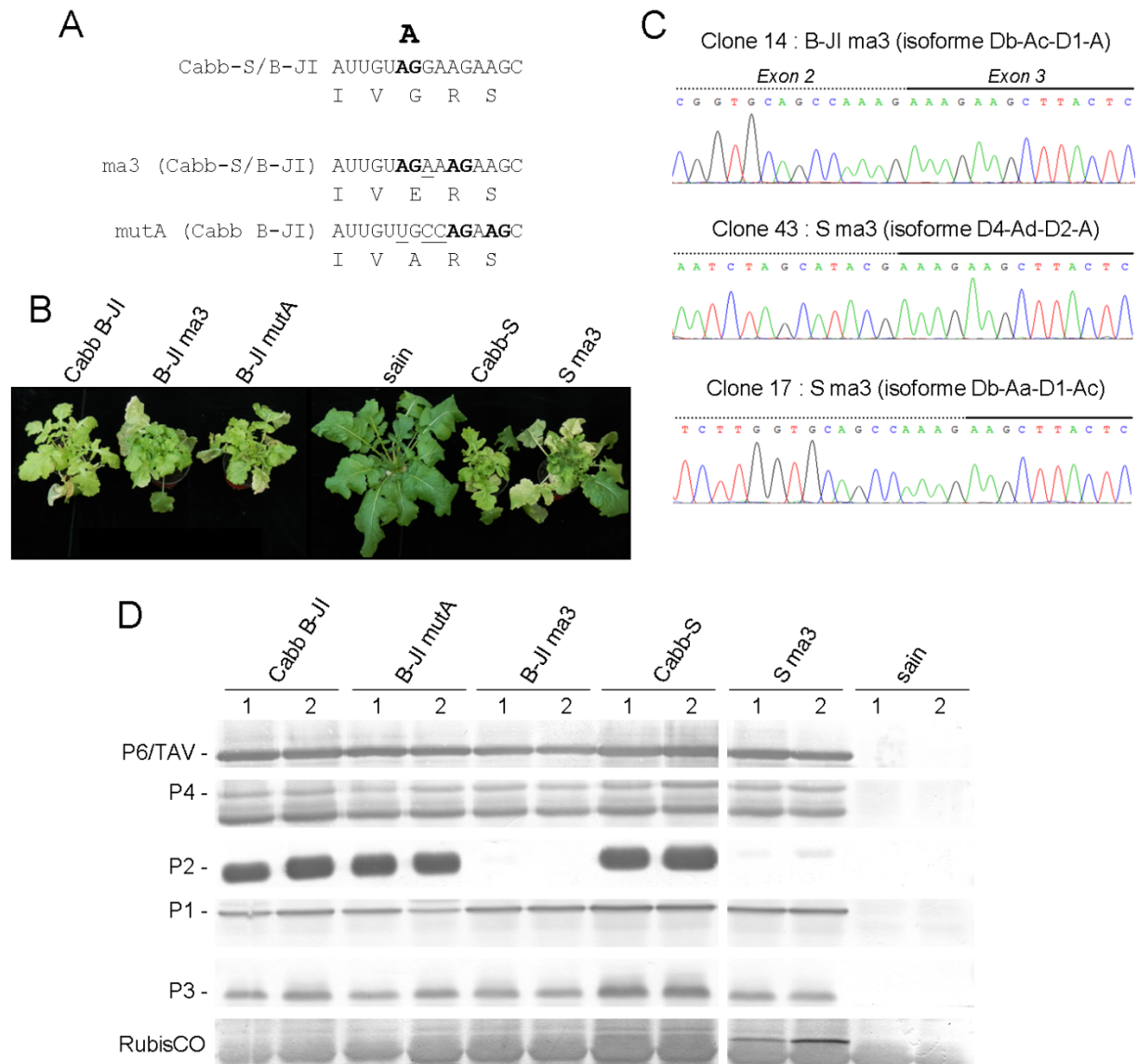


Figure 24 : Des mutations du site accepteur A conduisent à l'utilisation de sites accepteurs cryptiques. (A) Le site accepteur d'épissage A a été muté chez les souches Cabb B-JI et Cabb-S. Les dinucléotides AG utilisés comme sites accepteurs chez les virus sauvages et mutants sont en gras ; les mutations insérées sont soulignées. La séquence codante de l'ORF II est indiquée sous les séquences nucléotidiques du virus sauvage et des deux mutants. (B) Symptômes présentés par des navets infectés par les virus sauvages et les différents mutants, à 45 dpi. (C) Séquences nucléotidiques de la jonction d'épissage située autour du site accepteur A, chez trois isoformes épissées, clonées à partir de navets infectés avec les mutants B-JI ma3 et S ma3. L'exon 2 est représenté par une ligne pointillée et l'exon 3 par une ligne solide. (D) Immunodétection des protéines virales P6/TAV, P4, P2, P1 et P3 chez des navets infectés par les mutants B-JI mutA, B-JI ma3, S ma3 ou les souches sauvages correspondantes. Deux réplicats sont présentés pour chaque condition.

En conclusion, nos résultats montrent que lorsque le site accepteur A est rendu inutilisable, l'épissage de l'ARN 35S est maintenu grâce à la sélection, par le splicéosome, de sites accepteurs cryptiques présents à proximité du site A. Ceci explique pourquoi l'inactivation du site A n'est pas délétère pour l'infectiosité du virus.

DISCUSSION

1) Conservation des sites d'épissage, chez les souches du CaMV et les *Caulimovirus*

Nos données démontrent l'existence de l'épissage alternatif chez une troisième souche de CaMV, la souche Cabb B-JI, après les travaux réalisés sur les souches CaMV-S (Hirochika et al., 1985) et Cabb-S (Kiss-László et al., 1995). Comme le montre notre analyse *in silico* (Figure 13), il est probable que l'épissage alternatif de l'ARN 35S soit un phénomène très largement répandu parmi les souches du CaMV, puisque trois des quatre sites donneurs et le site accepteur décrits par Kiss-László et al. sont conservés chez toutes les souches virales séquencées. Les sites D1 et D2 font partie de codons Gly qui se situent dans l'ORF I, au niveau de régions conservées parmi les différentes souches du CaMV et certains *Caulimovirus* (Thomas et Maule, 1995). Pourtant, la conservation de ces deux sites donneurs n'est pas à mettre au crédit de l'importance des deux glycines pour lesquelles ils codent, pour le mouvement du virus de cellule à cellule, puisque les codons de la glycine sont de type GGX (où X est n'importe quel nucléotide), et non seulement GGU. Enfin, tous les nouveaux sites d'épissage que nous avons pu caractériser chez Cabb B-JI et/ou Cabb-S sont conservés chez les autres souches du CaMV.

Outre le CaMV, le seul *Caulimovirus* pour lequel l'existence de l'épissage alternatif a été prouvée est le FMV, dont l'organisation génomique est très similaire (Richins et al., 1987; Scholthof et al., 1991). L'évènement d'épissage qui a été décrit implique un site donneur situé dans l'ORF IV et un site accepteur situé dans l'ORF V. Ces deux sites ne sont pas conservés chez le CaMV. Inversement, les sites D1, D2 et A du CaMV sont conservés chez le FMV et le LLDV. Il est donc également possible que l'épissage alternatif dans la région 5' de l'ARN 35S soit un phénomène existant chez d'autres *Caulimovirus* que le CaMV.

2) Découverte d'un profil d'épissage complexe, chez le CaMV

L'épissage alternatif de l'ARN 35S du CaMV est un phénomène bien plus complexe que ce que les premiers articles publiés sur le sujet ont laissé transparaître (Hirochika et al., 1985; Kiss-László et al., 1995). En effet, nos résultats montrent que l'ARN 35S de la souche Cabb B-JI possède au moins six sites donneurs et sept sites accepteurs. Leur utilisation par le splicéosome conduit, en théorie, à la production de plusieurs dizaines d'isoformes épissées. Nous avons pu caractériser 15 de ces isoformes, parmi lesquelles certaines résultent de l'excision de deux voire trois introns. Une telle complexité est également observée chez la souche Cabb-S puisque six sites donneurs et six sites accepteurs sont utilisés au cours de l'infection ; nous avons caractérisé 13 isoformes épissées différentes. De même que chez Cabb B-JI, des isoformes proviennent de l'excision de deux ou trois introns. Enfin, tous les nouveaux sites que nous avons caractérisés sont conservés chez l'ensemble des souches du CaMV. Les évènements d'épissage que nous avons décrits ne concernent que la partie 5' de l'ARN 35S. Malgré la présence de plusieurs sites putatifs dans d'autres régions de l'ARN 35S, nous n'avons pas détecté d'épissage en aval de l'ORF II, ce qui suggère que l'épissage de la partie 5' pourrait être impliqué, entre autres, dans la régulation du transport *in planta* du CaMV et/ou de sa transmission, par les pucerons vecteurs.

La découverte de l'existence d'une telle complexité, dans l'épissage alternatif de l'ARN 35S du CaMV, soulève inéluctablement la question de sa signification biologique. L'approche méthodologique que nous avons utilisée peut expliquer, en partie, ce résultat, ainsi que la différence entre nos données et celles obtenues précédemment (Hirochika et al., 1985; Kiss-László et al., 1995). Dans l'article d'Hirochika et al. (1985), l'existence de l'épissage alternatif chez la souche CaMV-S a été fortement suggérée après que différentes formes d'ADN génomique viral aient été clonées puis séquencées. La

difficulté que représentait une telle tâche, à cette époque, a limité le nombre de molécules d'ADN viral séquencées à quatre. La méthode employée par Kiss-László et al. (1995) se rapproche de la nôtre puisqu'elle repose sur le séquençage d'une cinquantaine de clones d'ADNc produit à partir d'ARN de protoplastes d'*Arabidopsis* transfectés avec des constructions artificielles contenant les premiers ORF du CaMV. Les nouveaux sites d'épissage que nous présentons ici ont été caractérisés à partir d'ADNc synthétisé à partir de navets infectés. De nombreux clones – environ 90 par souche – ont été séquencés et ce, après une étape de tri où les clones contenant l'ADNc d'isoformes épissées surreprésentées ont été éliminés. Notre méthodologie, qui privilégie la diversité des fragments de RT-PCR clonés, a ainsi permis la caractérisation d'au moins 12 isoformes épissées par souche. Il est certain que notre façon de procéder a induit un biais en privilégiant la caractérisation d'isoformes dont certaines ne sont peut-être que rarement produites. Toutefois, l'étendue même de ce biais est difficile à évaluer car il faudrait, pour cela, pouvoir mesurer la proportion relative de chacune des isoformes de l'ARN 35S.

Chez Cabb-S, la proportion globale d'ARN 35S épissés a été estimée à 70%, et les quatre isoformes alors décrites ont été quantifiées par la technique de protection aux ribonucléases (RNases) A et T1 (Kiss-László et al., 1995). Selon ces données, l'isoforme D4-A représente 10 % de la population totale d'ARN 35S, tandis que les trois autres isoformes D1-A, D2-A et D3-A représentent chacune 20% de cette population. Une estimation de nos résultats obtenus par northern-blot et analysés en densitométrie donne, à première vue, une valeur similaire pour la souche Cabb B-JI puisqu'on estime à 71% la proportion d'ARN épissés, au sein de la population totale d'ARN 35S. (Figure 15B). Pourtant, nos résultats remettent doublement en cause ces estimations. Premièrement, la quantification des isoformes s'est faite en ciblant les différentes jonctions d'épissage. Or ces jonctions peuvent être partagées par plusieurs isoformes lorsque ces dernières procèdent de l'excision de plusieurs introns : les isoformes Db-Ab-D2-A, D2-A et D4-Ad-D2-A partagent, par exemple, la jonction D2-A. Deuxièmement, pour ce qui est de l'estimation de la proportion globale d'ARN 35S épissés, nous montrons que certains ARN épissés peuvent être rétrotranscrits, comme cela a été observé chez la souche CaMV-S (Hirochika et al., 1985). L'ARN 35S possède, à ses extrémités 5' et 3', une séquence répétée d'environ 180 nucléotides qui est impliquée dans des sauts réplicatif intra- ou intermoléculaires ayant lieu au cours de la rétrotranscription. En outre, il possède, aux alentours de l'extrémité 5' de l'ORF VII, un site de liaison pour l'ARNt-Met initiateur utilisé pour initier la rétrotranscription, ainsi qu'une à trois séquences riches en purines, qui permettent d'initier la synthèse du brin complémentaire de l'ADNc (Guilley et al., 1983; Pfeiffer et Hohn, 1983). En théorie, tout ARN 35S possédant ces signaux – et c'est le cas de chacune des isoformes épissées, à l'exception de celles où les sites donneurs Db ou D4 ont été sélectionnés – peut être rétrotranscrit. Nous avons détecté, expérimentalement, plusieurs formes d'ADN viral délété, issues de la rétrotranscription d'ARN épissés. Il se peut que ces différentes formes d'ADN délété correspondent aux molécules d'ADN circulaires de tailles variées qui ont été observées dans le noyau de plantes infectées par le CaMV (Olszewski et Guilfoyle, 1983). Ces génomes délétés pourraient probablement participer à la production d'ARN analogues aux isoformes épissées – dits *spliced-like* –, dont la présence fausserait les tentatives de quantification. Plus globalement, la complexité des profils d'épissage que nous avons observés, ainsi que la relative homogénéité de taille parmi les différents formes épissées – l'excision de l'intron le plus long, Db-A, réduit la taille de l'ARN 35S de seulement 20% –, rendent la quantification précise des isoformes décrites très ardue.

3) Raison d'être de l'épissage alternatif de l'ARN 35S du CaMV

Un des points, dans notre questionnement de départ, était de comprendre pourquoi plusieurs sites donneurs d'épissage sont utilisés. Cette interrogation s'inscrivait alors dans un schéma d'épissage relativement simple, où trois sites donneurs sur quatre pouvaient conduire à l'expression de protéines de fusion P1P2. La génération de protéines de fusion, au cours de l'épissage, est assez commune. L'épissage alternatif, chez l'HBV, génère deux protéines de fusion différentes nommées HBSP et HBDSP (Chen et al., 2010; Soussan et al., 2000). Chez le MSV, c'est un ORF de fusion procédant de l'épissage qui code pour la protéine Rep impliquée dans la réplication du génome viral

(Wright et al., 1997). Plus généralement, la création d'ORF est un phénomène retrouvé dans de nombreux cas d'épissage alternatif. À l'échelle des virus, nombreuses sont les protéines produites à partir d'ARN épissés alternativement ; c'est le cas, par exemple, des antigènes t et T du SV40 – un membre des *Polyomaviridae* – qui sont synthétisés à partir de deux isoformes issues d'un même pré-ARNm. On citera également, en référence aux exemples énoncés dans l'introduction, les protéines du HIV-1 dites accessoires, Rev et Tat, ainsi que la protéine NS2/NEP de l'IAV.

Même si nos données n'apportent pas une réponse claire à la question, il semble que ce ne soit pas dans le but d'exprimer des protéines de fusion que de multiples sites donneurs sont utilisés. Il était admis, précédemment, que 60% des ARN 35S de la souche Cabb-S codaient pour de telles protéines (Kiss-László et al., 1995). On pouvait alors supposer un rôle pour ces protéines. Nous avons étudié, *in vitro*, leurs interactions potentielles avec des protéines virales. Les protéines P1P2 sont capables d'interagir *in vitro* avec les protéines virales P2, P3 ainsi qu'avec les virions. Ces interactions se font probablement par le biais de la partie P2 des protéines de fusion (Hébrard et al., 2001; Leh et al., 1999). Dans le cas de l'interaction avec les virions, il est probable que cela se fasse par l'intermédiaire de P3, puisque cette dernière est associée aux particules virales (Dautel et al., 1994). Des interactions *in planta* entre protéines de fusion et P2 et P3 pourraient donc moduler la transmission (Armour et al., 1983; Leh et al., 1999, 2001) ou le mouvement de cellule à cellule du CaMV (Stavolone et al., 2005). Cependant, comme cela a été exposé précédemment, le pourcentage des ARN codant les protéines P1P2 est probablement inférieur à 60%, du fait de l'existence de nombreuses autres isoformes épissées – se référer à la partie précédente pour l'explication sur la quantification des isoformes épissées. De plus, les mutants où les sites donneurs D1 et D2 ont été inactivés se sont révélés être parfaitement infectieux, dans nos conditions d'inoculation et de culture des plantes, tout comme l'a été le mutant conduisant à la production de protéines de fusion tronquées de leur fragment issu de P2, suite à l'introduction d'un codon stop immédiatement en aval du site A (mutant P1P2stop). Enfin, nous n'avons jamais détecté les protéines de fusion en contexte infectieux. Il semble que ces protéines soient très faiblement exprimées et/ou instables puisque nous ne les avons détectées que très rarement et en faibles quantités, lors d'expériences d'expression transitoire, *in planta*. Cette possible instabilité ne semble pas dépendre de la voie de l'ubiquitine/protéasome, puisque nous n'avons pas observé d'accumulation des protéines de fusion suite à l'inhibition de cette voie de dégradation, par le MG132. Il se peut que les protéines P1P2 soient dégradées par des protéases cellulaires – propres aux cellules végétales puisque nous avons pu exprimer les P1P2 en système bactérien –, ou bien qu'elles possèdent une ou plusieurs séquences d'instabilité, comme cela a été montré pour la protéine de capsid (Karsies et al., 2001). Une dernière explication serait que les ARN viraux épissés ne sont pas traduits, bien que cela semble très improbable étant donné que ces ARN ont une organisation très proche – à l'exception des isoformes épissées au niveau des sites D1 et Db – de l'ARN 35S non-épissé.

Parmi tous les isoformes épissés que nous avons caractérisés, la très grande majorité codent pour des ORF de fusion qui impliquent non seulement les ORF I et II, mais aussi le petit ORF E ainsi que des séquences d'autres cadres de lecture localisés dans la région des ORF I et II. Il est possible qu'à défaut de coder pour des protéines ayant une fonction biologique, ces ORF aient un rôle en *cis*. Ainsi, en conduisant à la formation d'un long ORF, ils permettraient de réguler négativement l'expression de l'ORF II, tout en prévenant l'apparition de nombreux ORF dont la présence serait de nature à défavoriser la réinitiation traductionnelle P6/TAV-dépendante des ORF situés en aval, tels que les ORF III et IV. Cette hypothèse a été émise, en des termes assez similaires, par Kiss-László et al. (1995).

S'il a été proposé que l'épissage alternatif de l'ARN 35S du CaMV permette une meilleure expression des ORF III, IV et V, en réduisant le nombre d'ORF situés en amont de l'ORF III (Kiss-László et al., 1995), il a été montré que le rôle principal de l'épissage est de réguler négativement l'expression de l'ORF II et ce, afin d'empêcher une trop forte accumulation de son produit d'expression, P2, qui serait toxique pour le bon déroulement du cycle viral (Froissart et al., 2004). Il nous a été impossible de reproduire les résultats des articles de Kiss-László et al. (1995) et Froissart

et al. (2004), en ce qui concerne l'inactivation du site accepteur d'épissage A. Dans ces deux articles, la jonction intron-exon située autour du site A a été mutée, ce qui a provoqué une inactivation du site A et de l'épissage de l'ARN 35S. Nous avons introduit cette même mutation (mutation ma3) sur nos deux souches d'études – dont l'une, Cabb-S, a été utilisée dans ces deux articles – sans pour autant observer une inactivation de l'épissage. Au contraire, nous avons montré que cette mutation, dans nos conditions expérimentales, n'est pas capable de désactiver le site A. Le seul effet notable que nous ayons observé est une absence de P2 dans les plantes infectées par ces mutants. Dans une seconde tentative, l'inactivation complète du site A – chez le mutant mutA – a conduit à l'utilisation de dinucléotides AG situés à proximité immédiate de ce site. Notre incapacité à inhiber véritablement l'épissage de l'ARN 35S a été la limite principale de nos investigations et elle a annihilé, jusqu'à ce jour, nos efforts visant à comprendre la ou les fonctions de l'épissage alternatif, chez le CaMV. La détection de quantités similaires de P1 et de P2, chez le mutant B-JI mutA et la souche sauvage (Figure 24D), semble indiquer que les sites accepteurs cryptiques qui pallient la mutation du site A sont reconnus efficacement par le spliceosome. En d'autres termes, la quantité d'ARN 35S non-épissé, dans les cellules infectées par ce mutant, devrait être la même que celle présente dans les cellules infectées par le virus sauvage, étant donné que les protéines P1 et P2 ne sont exprimées qu'à partir de la forme complète de l'ARN 35S. Dans le cas contraire, on détecterait plus de protéines P1 et P2, chez les plantes infectées par ce mutant, par rapport à celles infectées par la souche sauvage. Comme évoqué précédemment, il a été envisagé que les ARN épissés favoriseraient l'expression des ORF III et IV, en réduisant le nombre d'ORF situés en amont de l'ORF III. Il s'avère que les plants de navets infectés par le mutant B-JI mutA et la souche sauvage, renferment des quantités de protéines virales P3 et P4 quasi-identiques. Le fait que la protéine P2 ne soit pas détectée dans les feuilles systémiques des plantes infectées par les mutants de type ma3, alors qu'elle s'accumule normalement dans les plantes infectées par le mutant B-JI mutA, pourrait s'expliquer par une différence dans l'acide aminé introduit par la mutation : la glycine de la P2 sauvage est remplacée par une alanine, chez le mutant B-JI mutA – un acide aminé chimiquement très semblable à la glycine –, ou par un acide aspartique, chez les deux mutants ma3. Il se peut que la présence d'un acide aspartique conduise à des modifications dans les propriétés physico-chimiques de la P2, ce qui aurait un effet déstabilisateur comme cela a déjà été rapporté dans le cas de la mutation d'autres de ses résidus (Hébrard et al., 2001).

Enfin, il est envisageable que l'utilisation de nombreux sites d'épissage, et donc la présence de nombreuses isoformes épissées de l'ARN 35S, soient le résultat d'une modulation de l'épissage causée par l'infection. Certaines protéines virales peuvent influencer sur l'épissage de certains transcrits – viraux ou cellulaires – en agissant de façon ciblée, ou parfois de manière plus globale. La protéine NS1 du virus de l'influenza A, par exemple, peut déréguler l'épissage, grâce à sa capacité à interagir avec plusieurs composants du spliceosome dont les snARN U2 et U6 (Dubois et al., 2014). Par analogie, on peut imaginer qu'une ou plusieurs protéines du CaMV puissent perturber l'épissage dans les cellules infectées. La protéine P6/TAV, de par sa localisation partiellement nucléaire (M. Haas et al., 2005) et sa capacité à interagir avec l'ARN simple- (De Tapia et al., 1993) et double-brin (Cerritelli et al., 1998), pourrait être un candidat potentiel à cette fonction mais d'autres protéines, en particulier P2 et P3 – dont les petites tailles autorisent une diffusion nucléaire – pourraient également être impliquées dans ce phénomène. L'incapacité à détecter par northern-blot, des formes épissées chez des plantes infectées par un mutant produisant des protéines P1P2 et P2 tronquées – le mutant P1P2stop – semble aller, au moins partiellement, dans le sens de cette dernière hypothèse. Finalement, il n'est pas impossible que les nombreuses isoformes de l'ARN 35S soient, en partie, le résultat d'un processus d'épissage aléatoire. Cette hypothèse, selon laquelle les sites d'épissage sont souvent sélectionnés de façon stochastique, a été proposée récemment (Melamud et Moul, 2009a, 2009b) et elle expliquerait une partie importante des événements d'épissage chez l'Homme. Ainsi, de nombreuses isoformes sans fonction particulière seraient produites et soumises à une pression évolutive faible, tant qu'elles n'influenceraient pas négativement le cycle infectieux du virus.

4) Autres fonctions possibles de l'épissage, lors du cycle infectieux

Tous les nouvelles isoformes que nous avons caractérisées ne codent pas pour les ORF I et II, et aucune d'entre elles ne semble être défavorable à l'expression des ORF III, IV et V, puisqu'elles ont moins d'ORF en amont de l'ORF III que n'en a l'ARN 35S non-épissé. Même si les conclusions de Froissart et al. (2004) ne vont pas particulièrement dans cette direction, il est possible que l'épissage ait également pour rôle de réguler négativement l'expression de l'ORF I, peut-être de façon temporelle comme chez certains virus. La protéine de mouvement du virus de la mosaïque du tabac est, par exemple, exprimée au début de l'infection (Lehto et al., 1990; Watanabe et al., 1984).

Il n'est pas à exclure que l'épissage alternatif puisse avoir d'autres fonctions sans lien direct avec l'expression des gènes viraux. Les introns contiennent parfois des séquences de précurseurs de miARN (Brown et al., 2008). Ce n'est apparemment pas le cas chez le CaMV, puisque la région allant de la fin de la séquence *leader* – le début de l'ORF VII – jusqu'au site A, c'est-à-dire la majorité de l'intron Db-A, n'est pas la source d'une forte production de petits ARN viraux par la machinerie cellulaire, contrairement à la région *leader* et au début de l'ARN 19S (Blevins et al., 2011; Moissiard et Voinnet, 2006). Récemment, il a été montré, chez *Arabidopsis*, que les ARN possédant des introns sont moins sensibles au phénomène de *RNA silencing* que les transcrits exempts d'introns. En particulier, les ARN avec introns sont moins sujets à la production de petits ARN interférents (siARN) médiée par l'ARN polymérase ARN-dépendant RDR6 (Christie et al., 2011). Au cours de l'infection virale, l'ARN 35S est sujet à la dégradation en siARN, grâce à l'action des 4 types de protéines DCL – pour « *dicer-like* » – effectrices du *RNA silencing* (Blevins et al., 2006; Moissiard et Voinnet, 2006). Ce processus, et plus particulièrement la production de siARN dépendante de RDR6, est en partie contrecarré par la protéine P6/TAV (Shivaprasad et al., 2008). Il est possible que la présence d'introns sur l'ARN 35S renforce l'action inhibitrice de P6/TAV envers RDR6. Enfin, une dernière hypothèse est que les sites donneurs localisés dans la région 5' *leader* – Db et D4 – pourraient être impliqués dans la suppression d'événements de polyadénylation prématurés, lors de la transcription de l'ARN 35S, de façon analogue à ce qui a été décrit chez les *Retroviridae* comme le HIV-1 (Ashe et al., 1995; Sajic et al., 2007) ou les *foamy virus* (Schrom et al., 2013). En effet, la snRNP U1, en plus de son rôle dans la réaction d'épissage, interfère avec les événements de formation de l'extrémité 3' des pré-ARNm et protège certains pré-ARNm d'une polyadénylation prématurée (Kaida et al., 2010). Ce mécanisme participe au maintien de l'intégrité du transcriptome, en assurant une longueur convenable aux ARNm (Langemeier et al., 2013). Chez le HIV-1 et les *foamy virus*, le signal de polyadénylation est retrouvé à chacune des deux extrémités du génome proviral, au niveau de séquences répétées nommées 5' et 3'LTR – pour « *long terminal repeat* » – qui sont générées au cours de la rétrotranscription. Il a été démontré que le signal de polyadénylation localisé dans la 5'LTR est ignoré grâce à un mécanisme qui implique la reconnaissance, par la snRNP U1, d'un site donneur d'épissage situé à quelques centaines de nucléotides (Ashe et al., 1995; Schrom et al., 2013). La reconnaissance de ce site donneur, à proximité du signal de polyadénylation situé à l'extrémité 5' du génome, empêche un arrêt prématuré de la transcription et donc, participe à la production d'ARN viraux complets. Sur le génome du CaMV, le signal de polyadénylation est situé à environ 180 nucléotides en aval du site d'initiation de la transcription. Il va de soi que la polyadénylation, chez le CaMV, doit être régulée afin de permettre une transcription complète du génome et d'empêcher la production de petits ARN viraux dits *short-stop*, résultant d'un arrêt de la transcription suite à la première rencontre avec le signal de polyadénylation (Sanfaçon et Wicczorek, 1992). Les sites donneurs D4 et Db sont localisés à environ 300 nucléotides en aval du signal de polyadénylation et lui font face, au sein de la tige-boucle de la région 5' *leader*. L'étude d'une possible implication de ces sites dans la régulation de la polyadénylation pourrait permettre de savoir si le CaMV utilise une stratégie similaire à celle qui est employée par certains membres de la famille des *Retroviridae*.

MATÉRIELS ET MÉTHODES

1) Alignement de séquences

Les séquences complètes des souches de CaMV et de membres du genre *Caulimovirus* proviennent de Genbank. Les numéros d'accèsion des souches de CaMV sont les suivants : Cabb-S (NC_001497.1), CMV-1 (M90543.1), NY8153 (M90541.1), CM1841 (V00140.1), BBC (M90542.1), B29 (X79465.1), W260 (JF809616.1), D/H (M10376.1), Xinjiang (AF140604.1), Cabb B-JI (KJ716236). Les numéros d'accèsion des *Caulimovirus* analysés sont les suivants : *Carnation etched ring virus* (CERV; NC_003498.1), *Dahlia mosaic virus* (DMV; NC_018616.1), *Figwort mosaic virus* (FMV; NC_003554.1), *Lamium leaf distortion associated virus* (LLDAV; NC_010737.1), *Strawberry vein banding virus* (SVBV; NC_001725.1). L'alignement de séquences a été réalisé à partir du programme informatique Clustal Omega (Sievers et al., 2011), avec des paramètres de base.

2) Plantes hôtes, souches virales et inoculation

Des plants de navets (*Brassica rapa* cv. Tokyo) au stade « trois feuilles » sont inoculés mécaniquement avec 8 µg de vecteur viral linéarisé, contenant le génome des souches sauvages Cabb-S (vecteur pCa37 linéarisé par *Sall* ; Franck et al., 1980) et Cabb B-JI (vecteurs pMD324 et pBS-BJI, linéarisés respectivement par *Sall* et *SpeI* ; Delseny et Hull, 1983), ou de mutants de celles-ci. Après inoculation, les plants sont cultivés en conditions de jour long (16 h de lumière à 22°C, 8 h d'obscurité à 18°C).

3) Plasmides

a) Mutants d'épissage et mutant P1P2stop

Tous les mutants de sites donneurs et accepteurs ont été générés par la technique standard de mutagenèse dirigée, à partir des plasmides pMD324 (mutants des sites donneurs et mutant P1P2stop, chez Cabb B-JI), pBS-BJI (mutants des sites accepteurs, chez Cabb B-JI) et pCa37 (mutant du site accepteur, chez Cabb-S).

b) Vecteurs d'expression de versions étiquetées de la protéine GFP et des protéines virales P1, P2, P3, P6/TAV et P1P2(D2)

L'intron D2-A a été délété par PCR, sur le plasmide pMD324, pour créer la séquence codante de la protéine P1P2(D2). Les séquences codantes des gènes de la protéine fluorescente verte (GFP) et de la protéine P1P2(D2) ont été fusionnées aux séquences codantes des étiquettes hémaglutinine (HA, provenant de l'hémaglutinine du virus de l'influenza humaine) et myc (provenant du facteur de transcription c-Myc), par le système de clonage Gateway (Invitrogen), selon les instructions du fabricant et à partir des vecteurs binaires pGWB14, 15, 17 et 18 (Nakagawa et al., 2007).

Pour fusionner la séquence codante de la glutathion S-transférase à celle de la protéine de fusion P1P2(D2), la séquence de cette dernière a été amplifiée par PCR (oligonucléotides : 5'-CAAGG-TACCGGATCCATGGATTTGTATCCAGAAGAAAATA-3' et 5'-CAAGAGCTCGTCTGACTTAGCCAAT-AATATTCTTTAATCCTT-3'), puis clonée par *Bam*HI et *Sall*, dans le vecteur pGEX-6P-1 (GE Healthcare).

Les séquences codantes des protéines virales P1, P2, P3 et P6/TAV ont été amplifiées par PCR (oligonucléotides 5'- CAAGGTACCGGATCCATGGATTTGTATCCAGAAGAAAATA-3' et 5'-CAA-GAGCTCGTCTGACTTATTCTCCACAGATTTCTTTTAATTT-3' pour P1, 5'-CAAGGTACCGGATCCAT-

GAGCATTACGGGTCAACCG-3' et 5'- CAAGAGCTCGTCGACTTAGCCAATAATATTCTTTAATCCTT-3' pour P2, 5'-CAAGGTACCGGATCCATGGCTAATCTTAATCAAATCCAAAA-3' et 5'-CAAGAGCTCGTCGACCTAAAATTGATTCGGCCATCCTG-3' pour P3, 5'-CAAGGTACCGGATCCATGGAGAACA-TAGAAAACT-3' et 5'-CAAGAGCTCGTCGACTCAATCCACTTGCTTTGAAG-3' for P6/TAV), puis clonées par *KpnI* et *SacI* dans le vecteur pETK (Leh et al., 2000), un vecteur qui conduit à la production de protéines possédant, à leur extrémité amino-terminale, la séquence RRASV, qui est ciblée par la kinase bovine de muscle cardiaque – HMK, pour « *heart muscle kinase* ». Toutes les constructions décrites dans ce paragraphe ont été vérifiées par séquençage.

4) Extraction d'ARN totaux et northern-blot

Trois disques foliaires (1cm de diamètre) issus de feuilles systémiques de navets, sont récoltés à 25 dpi, congelés dans l'azote liquide puis broyés dans un homogénéiseur Precellys 24 (Bertin Technologies). Les ARN totaux sont extraits avec du TRI-Reagent (Sigma-Aldrich), selon les instructions du fabricant, puis solubilisés dans de l'eau stérile. 15 µg d'ARN totaux sont séparés par électrophorèse sur gel d'agarose (0,65% ou 0,8%) – urée (6M ou 7M), dans du tampon Tris-acétate-EDTA (40 mM de Tris-acétate, 1 mM d'EDTA). Les ARN sont ensuite électro-transférés sur une membrane de nylon (Amersham Hybond-N+; GE Healthcare), pendant 75 min à 350 mA, dans un tampon Tris-borate-EDTA 0,5x (45 mM de Tris-borate, 1mM d'EDTA) à 4°C, puis fixés à la membrane par une irradiation aux UVs (1200 mJ dans un Stratalinker 2400, Stratagene). La membrane est pré-hybridée dans du tampon PerfectHyb Plus (Sigma-Aldrich) pendant 15 min à 42°C, puis incubée, pendant une nuit à cette même température, avec des sondes oligonucléotidiques radiomarquées au ³²P (sonde *int* : 5'-CTACAATCTCATTGAGGTTGTTAAGATGA-3', sonde *ex* : 5'- CTTCCATATTCC-GAGTAAGCTTCTTC-3'). Les sondes liées à la membrane de façon non-spécifique sont éliminées par un premier lavage de 30 min à 50°C, dans du tampon *saline sodium citrate* 5x (SSC ; 0,75 M de NaCl et 75 mM de citrate de sodium, pH 7) - SDS 0,1%, puis par un second lavage de 5 min à 50°C, dans du tampon SSC 2x - SDS 0,1%. Une étape de lavage supplémentaire, identique à la précédente, est parfois effectuée.

5) Synthèse d'ADNc par RT-PCR

2,5 µg d'ARN totaux sont traités avec 2,5 U de DNase I (Promega), pendant 30 min à 37°C, puis extraits avec du TRI-Reagent. Les ADNc sont synthétisés par l'enzyme SuperScript III (Invitrogen), en présence d'amorce de type oligo(dT)₁₈, puis ils servent de matrice à l'ADN polymérase GoTaq Flexi (Promega). Pour la PCR, des oligonucléotides ciblant la région 5' de l'ARN 35S (5'-CAAGCGCGCTCTAGACTGAAATCACCAGTCTCTCTC-3' et 5'-CAAGAGCTCCTGCAGTTAGCCAA-TAATATTCTTTAATCCTT-3'), le gène de la GFP (5'-CTTAGCAAGGGCGAGGAGCTGT-3' et 5'-CTTGTACAGCTCGTCCATGCC-3') et le gène de la protéine de fusion P1P2(D2) (5'-CGAGATTCACTA-CTGCGTC-3' et 5'-CAACTGCAGTCTAGACTTTTAATTCTAGTATTTT-3') sont utilisés.

6) Analyse des ARN 35S épissés par RT-PCR

La partie 5' de l'ARN 35S est amplifiée par RT-PCR, selon le protocole décrit ci-dessus. Les amplicons obtenus sont ensuite séparés sur un gel d'agarose à 1%, puis extraits à l'aide du kit NucleoSpin Gel and PCR Clean-up (Macherey-Nagel), et enfin clonés dans le vecteur pGEM-T Easy (Promega). Dans le cas des mutants du site accepteur A, la PCR faite à la suite de la rétrotranscription est réalisée avec l'ADN polymérase Phusion (Thermo Scientific), une polymérase beaucoup plus fidèle que la polymérase GoTaq. Après obtention des clones, la taille des inserts est vérifiée par une électrophorèse sur gel d'agarose. Les clones contenant des inserts identiques et surreprésentés sont éliminés, tandis que les clones sélectionnés sont envoyés à séquencer. Les séquences d'inserts obtenues sont alignées sur la séquence de l'ARN 35S, grâce au logiciel Serial Cloner 2.6.

7) Analyse des protéines par western-blot

Trois disques foliaires (1cm de diamètre) issus de feuilles systémiques ou agroinfiltrées, sont broyés dans 80 µL de tampon de broyage (250 mM de Tris-HCl pH 7,75, 1 mM DTT), et mélangés à 40 µL de tampon d'extraction (75 mM de Tris-HCl pH 6,8, 9 M d'urée, 4,3% de SDS et 7,5 % de β-mercaptoéthanol). Les extraits protéiques obtenus sont dénaturés pendant 5 min à 95°C, puis séparés par SDS-PAGE. Après électrophorèse, les protéines sont transférées sur une membrane de PVDF (Immobilon-P, Merck Millipore). Les protéines virales P1, P2, P3, P4 et P6/TAV sont détectées à l'aide d'anticorps polyclonaux de lapin, dilués 10000 fois. L'étiquette HA- est immunodétectée par un anticorps monoclonal de souris (Thermo Scientific Pierce) dilué 10000 fois ; l'étiquette myc- est, quant à elle, reconnue par des anticorps polyclonaux de lapin (Santa Cruz Biotechnology) dilués 2500 fois. Après incubation avec les anticorps primaires, la membrane est lavée, puis incubée avec des anticorps secondaires couplés à la peroxydase de raifort et dilués 20000 fois. La fixation de ces anticorps est révélée par chimioluminescence, à l'aide du kit Light Plus ECL (Roche), puis elle est visualisée par un système d'acquisition Fusion FX (Vilber Lourmat). Alternativement, l'incubation peut être réalisée avec des anticorps secondaires couplés à la phosphatase alcaline et dilués 5000 fois. La fixation des anticorps est alors révélée avec une solution composée de NBT et de BCIP (Promega).

8) Préparation de fractions enrichies en paroi cellulaire

Des fractions enrichies en paroi cellulaire ont été préparées d'après le protocole décrit par Albrecht et al., (1988). 8 g de feuilles systémiques sont récoltés à 28 dpi, et les feuilles sont dénervurées, finement hachées, puis broyées dans 7 mL de tampon de broyage (400 mM de saccharose, 100 mM de Tris-HCl pH 8,1, 10 mM de KCl, 10 mM de β-mercaptoéthanol, 5 mM de MgCl₂ et 10 % de glycérol). Le mélange est filtré à travers une toile à bluter dont le vide de maille est de 10 µm, puis les débris retenus sur la toile sont repris dans 3,5 mL de tampon de broyage ; le mélange est de nouveau filtré. Le filtrat recueilli constitue la fraction F. Les débris retenus sur la toile sont repris dans 3,5 mL de tampon de broyage additionné de 2% de Triton X-100, et le mélange obtenu est incubé 15 min à température ambiante puis filtré. Le deuxième filtrat recueilli constitue la fraction S₁. Les débris retenus sont repris dans 3,5 mL de tampon d'extraction (9 M d'urée, 75 mM de Tris-HCl pH 6,8, 1 M de β-mercaptoéthanol et 4,5% de SDS), et le mélange est incubé 15 min à 100°C. Le filtrat recueilli après une nouvelle filtration constitue la fraction S₂. Les résidus obtenus sont traités à l'identique, ce qui conduit à l'obtention d'une dernière fraction, nommée S₃. En parallèle, un extrait protéique total est préparé en broyant 100 mg de feuilles systémiques dans 100 µL d'urée à 8 M. Les différentes fractions sont analysées par western-blot, selon le protocole décrit précédemment.

9) Expression transitoire par agroinfiltration

Les vecteurs codants pour les versions étiquetées HA- et/ou myc- des protéines GFP, P1P2(D2) et RGA sont introduits dans *Agrobacterium tumefaciens* GV3101, par électroporation. Les agrobactéries transformées sont cultivées pendant une nuit à 28°C, dans du milieu Luria-Bertani complété avec 10 mM de MES, 40 µM d'acétosyringone et des antibiotiques appropriés. Le matin suivant, les cultures sont sédimentées par centrifugation (10 min à 3500 X g), puis elles sont suspendues dans un tampon composé de 10 mM de MES, 10 mM de MgCl₂ et 150 µM d'acétosyringone, et incubées pendant 5 h, à température ambiante et sous agitation constante. Après cette phase d'incubation, la DO des cultures est mesurée, à une longueur d'onde de 600 nm. Les cultures sont diluées à une DO de 0,5, et mélangées avec un volume équivalent d'une culture d'agrobactéries transformées avec le plasmide pBIN61-P19 et diluée à la même DO. Le plasmide pBIN61-P19 code pour P19, un inhibiteur du *RNA silencing* (Voinnet et al., 2003) qui est destinée à optimiser l'expression transitoire des protéines. Des plants de *Nicotiana benthamiana* au stade « six à huit feuilles » sont infiltrés avec une seringue sans aiguille, puis cultivés en conditions de jour long (16 h de lumière à 22°C, 8 h d'obscurité à 18°C). Les échantillons sont récoltés 2,5 jours après infiltration –

ou 1,5 jour, dans le cas des échantillons destinés aux cinétiques d'inhibition de la voie de l'ubiquitine/protéasome.

10) Cinétique d'inhibition de la voie de l'ubiquitine/protéasome

1,5 jour après agroinfiltration, les feuilles de *N. benthamiana* sont infiltrées avec une solution de MG132 à 100 μ M. Des disques foliaires (5 mm de diamètre) sont immédiatement prélevés et incubés dans le milieu d'infiltration. 1,5 h après le début de l'incubation, de la cycloheximide (Sigma-Aldrich) à 100 μ M est additionnée au milieu, et les disques sont incubés dans une cloche à vide pendant 10 min, afin de faire pénétrer dans les tissus, cet inhibiteur de la traduction eucaryote. Des disques sont prélevés aux temps 0 h, 1 h, 2 h et 4 h après l'ajout de la cycloheximide. Les protéines sont extraites et analysées par western-blot, selon le protocole décrit précédemment.

11) Expression et marquage au 32 P de protéines virales recombinantes

Les protéines virales P1, P2, P3, P6/TAV et GST-P1P2(D2) sont exprimées dans la souche BL21(DE3)pLysS d'*Escherichia coli*. L'expression des protéines est induite par l'ajout de 0,2 mM d'IPTG, lorsque la culture bactérienne est en phase exponentielle de croissance (DO à 600 nm environ égale à 0,6). Les bactéries sont ensuite cultivées pendant 4 h à 28°C, puis sédimentées par centrifugation et suspendues dans du tampon HMK (20 mM de Tris-HCl pH 7,5, 100 mM de NaCl et 12 mM de $MgCl_2$). Après trois étapes de sonication de 10 s chacune, le lysat bactérien est centrifugé pendant 10 min à 11000 X *g*, puis les corps d'inclusions sont suspendus dans 0,5 mL de tampon HMK. 30 μ L de cette suspension sont phosphorylés en présence de 25 μ Ci de [γ - 32 P]ATP et 25 U de kinase bovine de muscle cardiaque (Sigma-Aldrich), dans 150 μ L de tampon HMK, pendant 90 min à température ambiante. Après marquage, l'ATP non-incorporé est éliminé dans une colonne illustra MicroSpin (GE Healthcare).

12) Analyse d'interactions protéiques *in vitro*, par far western-blot

Un lysat de bactéries exprimant la protéine GST-P1P2(D2) est séparé par SDS-PAGE et transféré sur membrane de PVDF. La membrane est ensuite saturée dans du tampon phosphate salin 1x (137 mM de NaCl, 2,7 mM de KCl, 10 mM de Na_2HPO_4 et 2 mM de KH_2PO_4) supplémenté de 5% de lait et 1% de Tween 20, pendant 12 h à température ambiante. Après saturation, la membrane est incubée pendant 12 h à 4°C, soit avec des protéines virales recombinantes et radiomarquées au 32 P, soit avec des particules virales purifiées, puis elle est lavée pendant 10 min dans du tampon phosphate salin 1x supplémenté de 1% de Tween 20. Les interactions entre la protéine de fusion GST-P1P2(D2) et les protéines recombinantes et radiomarquées sont détectées par autoradiographie, tandis que celles impliquant les particules virales sont détectées par western-blot, avec des anticorps polyclonaux dirigés contre la protéine P4.

ANNEXE

Cauliflower mosaic virus transcriptome reveals a complex alternative splicing pattern necessary for infectivity

Article en réécriture après une première soumission à *Journal of Virology*

***Cauliflower mosaic virus* transcriptome reveals a complex alternative splicing pattern necessary for infectivity**

Clément Bouton, Angèle Geldreich, Laëticia Ramel, Lyubov A Ryabova, Maria Dimitrova # and Mario Keller #

Institut de Biologie Moléculaire des Plantes du CNRS, Université de Strasbourg, Strasbourg, France

Address correspondence to Maria Dimitrova, maria.dimitrova@ibmp-cnrs.unistra.fr or Mario Keller, mario.keller@ibmp-cnrs.unistra.fr

Running title: Complex alternative splicing pattern of CaMV

Number of words 249 + 145 (abstract), 7748 (text).

ABSTRACT

The pararetrovirus *Cauliflower mosaic virus* (CaMV) uses alternative splicing to generate several isoforms from its polycistronic pregenomic 35S RNA. This process is essential for infectivity. Previous works have identified four splice donor sites and a single splice acceptor site in the 35S RNA 5' region and suggested that the main role of CaMV splicing is to downregulate expression of open reading frames (ORFs) I and II. In this study, we show that alternative splicing is a conserved process among CaMV strains. In Cabb B-JI and Cabb-S strains, splicing can lead to fusion between ORFs I and II. Corresponding fusion proteins are not detected during infection and are almost never transiently expressed *in planta*. When expressed in *E. coli* they interact with viral proteins P2 and P3 *in vitro*, which suggests rapid degradation after synthesis, thus avoiding functional interference with the respective full-length proteins. Through RT-PCR-based analysis we demonstrate that CaMV 35S RNA exhibits a complex splicing pattern, as we identify new splice donor and acceptor sites whose selection leads to more than twelve 35S RNA isoforms in infected plants. Interestingly, inactivating splice donor or acceptor sites is not lethal for the virus, since disrupted sites are systematically rescued by the activation of cryptic and/or seldom used splice sites. Taken together, our data depict a conserved and robust process, involving multiple sites, that ensures splicing of 35S RNA. Our results also reinforce the hypothesis that CaMV uses alternative splicing as a way to downregulate expression of ORFs I and II.

IMPORTANCE

The pararetrovirus *Cauliflower mosaic virus* (CaMV) is one of the few plant viruses for which splicing has been reported. Alternative splicing of the CaMV polycistronic pregenomic 35S RNA has previously been shown to be essential for infectivity, probably by downregulating the expression of viral ORFs I and II. Here we demonstrate that alternative splicing is a process conserved among CaMV strains. The CaMV splicing pattern was thought to be simple, but our data reveal a more complex picture in which the new splice sites we characterize lead to the production of more than twelve 35S RNA isoforms in CaMV-infected plants. We also show that cryptic and/or seldom used sites can substitute inactivated sites. The observed flexibility in the use of splice sites, together with their multiplicity, stresses the robustness of CaMV 35S RNA alternative splicing and highlights the importance of this process during viral multiplication.

INTRODUCTION

Alternative RNA splicing is intensively performed in eukaryotes to increase the proteome through the formation of numerous mRNA species from a primary transcript and to regulate the expression of proteins in different organs and cell types (1, 2). Splicing can also enhance the expression level of protein genes because the spliceosome is tightly connected to the transcriptional machinery (3) and promotes mRNA export in mammalian cells (4). Many animal DNA viruses, complex retroviruses (5) and type A *influenza virus* use this strategy to express their proteins during the infectious

cycle (6). In the animal pararetroviruses *Hepatitis B virus* (HBV) and *Duck hepatitis virus* (DHBV), alternative splicing produces several mRNA species encoding fusion proteins whose roles remain elusive. However, it has been noticed that splicing is enhanced during chronic infection, notably prior to development of hepatocarcinoma, suggesting that spliced RNAs may contribute to the pathogenesis of *Hepadnaviruses* (7, 8). Unlike animal viruses, splicing rarely occurs with plant viruses, essentially because the vast majority has an RNA genome. Consequently, except for the *Nucleorhabdovirus* genus members, they realize their replication cycle in the cytoplasm. However, several plant DNA viruses from the *Geminiviridae* and *Caulimoviridae* families perform splicing to express some of their proteins (9, 10). The introns found in their genomes perfectly follow the GU...AG rule defined for nuclear mRNA introns, and they possess A+U rich sequences as described for plant introns (11). In *Mastreviruses* (family *Geminiviridae*), splicing of the C transcript generates an mRNA that codes for the fusion protein Rep which is required to initiate DNA replication by rolling circle (9). In the *Caulimoviridae* family, splicing has only been described in two genera - *Tungrovirus* and *Caulimovirus* - and it has been proposed to enhance the expression of distal ORFs in the spliced RNAs. In *Rice tungro bacilliform virus* (RTBV), the polycistronic 35S RNA undergoes a single splicing event that leads to removal of a large intron (6.3 kb), resulting in a monocistronic mRNA specific for protein P4 (12). In *Figwort mosaic virus* (FMV), splicing of the polycistronic 35S RNA creates an in-frame fusion between the 5' end of ORF IV and the 3' region of ORF V (13). However, the biological relevance of splicing in RTBV and FMV has never been investigated.

Cauliflower mosaic virus (CaMV) is a plant pararetrovirus and the type member of the genus *Caulimovirus*. Its double-stranded DNA genome is replicated by reverse transcription of the full-length pregenomic 35S RNA (14). This polycistronic RNA also serves as mRNA for the production of a complete set of proteins thanks to translational reinitiation mediated by the multifunctional protein P6/transactivator-viroplasm (P6/TAV), which is expressed from the subgenomic 19S RNA (15). Splicing occurs in the 35S RNA of CaMV-S strain, as evidenced by the characterization of a viral DNA carrying a deletion that perfectly corresponds to an intron in infected turnip plants (16). The 35S RNA of CaMV Cabb-S (Strasbourg) strain undergoes alternative splicing that generates four mRNA species in the course of its infectious cycle (17). Spliced RNAs represent approximately 70% of total viral RNA in CaMV-infected plants (17), thus indicating that splicing is regulated, as in retroviruses, to preserve a fraction of full-length 35S RNA. Alternative splicing in Cabb-S involves four splice donor sites, one located within the 5' untranslated region (UTR) and three in ORF I, and a single splice acceptor site within ORF II (Fig. 1A). Three spliced 35S RNAs contain a fusion ORF between the 5' part of ORF I which encodes the movement protein (P1) involved in the cell-to-cell spread of CaMV (18) and the 3' region of ORF II which encodes a factor (P2) involved in the transmission of CaMV by aphids (19). Up to now, no function has been assigned to the fusion proteins P1P2. Splicing is vital for CaMV Cabb-S since mutations that inactivate the unique splice acceptor site totally abolish infectivity in turnip plants (17, 20). It has been hypothesized that alternative splicing of the 35S RNA provides appropriate mRNAs to enhance the expression of the virion-associated protein (VAP) and the capsid protein (CP) encoded by ORFs III and IV, respectively, and modulates the expression of ORF II (20) and ORF I. Indeed, it has been shown that splicing of the 35S RNA is dispensable in CaMV mutants harboring a complete deletion of ORF II. In addition, the acceptor site has been found to be no longer required if ORF II is replaced by a foreign ORF (i.e. ORF coding for thioredoxin) or if mutations introduced in ORF II lead to a protein P2 which does not accumulate in infected cells. Therefore, a function of 35S RNA splicing would be to downregulate ORF II, whose excessive expression would be toxic for CaMV (20). However, a role has yet to be assigned to CaMV alternative splicing, since a single splicing event is sufficient to downregulate the expression of ORFs I and II and simultaneously to permit an efficient translation of downstream ORFs III and IV.

To gain a better understanding of 35S RNA alternative splicing and its functional relevance in CaMV infectivity, we focused our investigations on the different splice sites and on the putative fusion proteins that are encoded by some of the spliced isoforms. In this study, we show that CaMV alternative splicing is a conserved phenomenon which is more complex than previously described. We identi-

fy new splice donor and acceptor sites that allow the production of at least twelve 35S RNA isoforms during infection. The fusion proteins P1P2 encoded by some spliced RNAs do not seem to be important for the virus since they accumulate neither upon infection nor in plant transient expression assays. We show that inactivation of splice sites by mutagenesis is not lethal for CaMV since alternative splicing is kept functional by activation of cryptic splice donor and acceptor sites. These observations suggest that this process plays a key role in CaMV biology.

MATERIALS AND METHODS

Sequence alignments.

Complete sequences of CaMV strains and *Caulimovirus* species were retrieved from Genbank. Accession numbers of the CaMV strain genomes are the following: Cabb-S (NC_001497.1), CMV-1 (M90543.1), NY8153 (M90541.1), CM1841 (V00140.1), BBC (M90542.1), B29 (X79465.1), W260 (JF809616.1), D/H (M10376.1), Xinjiang (AF140604.1), Cabb B-JI (KJ716236). Accession numbers of the aligned *Caulimovirus* genomes are the following: *Carnation etched ring virus* (CERV; NC_003498.1), *Dahlia mosaic virus* (DMV; NC_018616.1), *Figwort mosaic virus* (FMV; NC_003554.1), *Lamium leaf distortion associated virus* (LLDAV; NC_010737.1), *Strawberry vein banding virus* (SVBV; NC_001725.1). Multiple sequence alignment was performed using Clustal Omega program (21) with default settings.

Host plant and virus inoculation.

Three-leaf stage turnips (*Brassica rapa* cv. Tokyo) were mechanically inoculated with 8 µg of linearized recombinant plasmids containing mutated or wild-type genome of two CaMV strains, Cabb-S (22) (*Sall*-linearized pCa37) and Cabb B-JI (23) (*Sall*-linearized pMD324 or *SpeI*-linearized pBS-BJI). Inoculated plants were grown in a greenhouse under long-day conditions (16 h/22°C day-8 h/18°C night).

Plasmids.

All donor and acceptor site mutants were generated by standard site-directed mutagenesis from pMD324 (donor site mutants of the Cabb B-JI strain), pBS-BJI (acceptor site mutants of the Cabb B-JI strain) or pCa37 (acceptor site mutants of the Cabb-S strain). The D2-A intron from pMD324 was deleted by PCR to create the P1P2(D2) coding sequence. Human influenza hemagglutinin (HA) and myc-tagged green fluorescent protein (GFP) and P1P2(D2) coding sequences were generated with the Gateway cloning system (Invitrogen) from binary vectors pGWB14, 15, 17 and 18 (24). To fuse the glutathione S-transferase (GST) to P1P2(D2), the P1P2(D2) cloning sequence was amplified by PCR with the following oligonucleotides (5'-CAAGGTACCGGATCCAT GGATTTGTATCCAGAAGAAAATA-3' and 5'-CAAGAGCTCGTCGACTTAGCCAATAATATTCTTTAATCCTT-3') and cloned via *Bam*HI and *Sall* into the pGEX-6P-1 vector (GE Healthcare). Coding sequences of the viral proteins P1, P2, P3 and P6/TAV were amplified with the following oligonucleotides (5'-CAAGGTACCGGATCCATGGATT-TGTATCCAGAAGAAAATA-3' and 5'-CAAGAGCTCGTCGACTTATTCTCCACAGATTTCTTTTAATTT-3' for P1, 5'-CAAGGTACCGGATCCATGAGCATTACGGGTCAACCG-3' and 5'-CAAGAGCTCGTCG-ACTTAGCCAATAATATTCTTTAATCCTT-3' for P2, 5'-CAAGGTACCGGATCCATGGCTAATCTTAA-TCAAATCCAAAAA-3' and 5'-CAAGAGCTCGTCGACCTAAATTGATTTCGGCCATCCTG-3' for P3, 5'-CAAGGTACCGGATCCATGGAGAACATAGAAAACT-3' and 5'-CAAGAGCTCGTCGACTCAATCC-ACTTGCTTTGAAG-3' for P6/TAV) and cloned into the *Kpn*I and *Sac*I restriction sites of the pETK vector (25) which allows N-terminal fusion to the coding sequence RRASV targeted by the bovine heart muscle kinase (HMK). All constructs were verified by sequencing.

Total RNA extraction and northern-blotting.

Three discs (diameter: 1 cm) from turnip systemic leaves harvested at 25 dpi were ground in a Precellys 24 homogenizer (Bertin Technologies). Total RNA was extracted using TRI-Reagent (Sigma-Aldrich) according to the manufacturer's instructions and solubilized in sterile water. 15 µg of RNA were separated by electrophoresis on an agarose (0.8%)-urea (7 M) gel in Tris-acetate-EDTA buffer (40 mM Tris-acetate, 1 mM EDTA). The RNA was electro-transferred onto a nylon membrane (Amersham Hybond-N+; GE Healthcare) in 0.5x Tris-borate-EDTA buffer (45 mM Tris-borate, 1 mM EDTA) at 350 mA for 75 min at 4°C and UV-crosslinked at 1,200 mJ in a Stratalinker 2400 (Stratagene). The membranes were pre-hybridized in PerfectHyb Plus buffer (Sigma-Aldrich) at 42°C for 15 min and incubated overnight at 42°C with ³²P-radiolabeled DNA oligonucleotide probes ("int": 5'-CTACAATCTCATTGAGGTTGTTAAGATGA-3'; "ex": 5'-CTTCCATATTCCGAGTAAGCTTCTTC-3'). Non-specifically bound probes were removed by washing the membranes in 5x saline-sodium citrate buffer (SSC; 0.75 M NaCl, 75 mM saline-sodium, pH 7), 0.1% sodium-dodecyl-sulfate (SDS) and in 2x SSC, 0.1% SDS at 50°C for 30 min and 5 min, respectively.

RT-PCR.

2.5 µg of total RNA were treated with 2.5 U of DNase I (Promega) at 37°C for 30 min and extracted with TRI-Reagent. cDNAs were synthesized with the SuperScript III reverse transcriptase (Invitrogen) in the presence of oligo(dT)₁₈ primer and used as template for the GoTaq Flexi DNA polymerase (Promega). For PCR, primers targeting the 5' region of the 35S RNA were 5'-CAAGCGC-GCTCTAGACTGAAATCACCAGTCTCTCTC-3' and 5' CAAGAGCTCCTGCAGTTAGCCAATAATAT-TCTTTAATCCTT-3'. Primers to amplify the GFP coding sequence were 5'-CTTAGCAAGGGCGA GGAGCTGT-3' and 5'-CTTGTACAGCTCGTCCATGCC-3' and the ones to amplify the ORF encoding P1P2(D2) were 5'-CGAGATTCACTACTGCGTC-3' and 5'-CAACTGCAGTCTAGACTTTTAATTCTAG-TATTTT-3'.

Analysis of spliced 35S RNA.

The 5' part of the 35S RNA was amplified by RT-PCR using the aforementioned conditions. Amplicons were fractionated on a 1% agarose gel, extracted with the NucleoSpin Gel and PCR Cleanup kits (Macherey-Nagel) and cloned into pGEM-T Easy vector (Promega). In the case of acceptor mutants, the PCR following the RT reaction was performed with the Phusion High-Fidelity DNA polymerase (Thermo Scientific) and the amplicons were cloned into pJET1.2 vector (Thermo Scientific). Inserts were sequenced and DNA sequences were aligned with the 35S RNA sequence using the Serial Cloner 2.6 software.

Protein analysis.

Three discs (1 cm diameter) of systemic or agroinfiltrated leaves were ground in 80 µl 250 mM Tris-HCl, pH 7.75 1 mM DL-Dithiothreitol and mixed with 40 µl of extraction buffer (75 mM Tris-HCl, pH 6.8, 9 M urea, 4.3% SDS, and 7.5% β-mercaptoethanol). Extracts were denatured at 95°C for 5 min. Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and electro-transferred to a polyvinylidene fluoride membrane (PVDF; Immobilon-P, Merck Millipore). Anti-P1, anti-P2 and anti-P4 polyclonal antisera were used at a 1:10,000 dilution. Monoclonal anti-HA (Thermo Scientific Pierce) and polyclonal anti-myc (Santa Cruz Biotechnology) antibodies were used at 1:10,000 and 1:2,500 dilutions, respectively. Horseradish peroxidase-labeled secondary antibodies were used at a 1:20,000 dilution and revealed with Lumi-Light Plus ECL (Roche). Chemiluminescence was visualized with a Fusion FX (Vilber Lourmat) acquisition system.

Transient expression assays by agroinfiltration.

HA- and myc-tagged P1P2(D2) or GFP coding vectors were transferred into *Agrobacterium tumefaciens* GV3101. Transformed *Agrobacteria* were grown overnight at 28°C in Luria-Bertani broth

supplemented with 10 mM 2-(N-Morpholino)ethanesulfonic acid (MES) and 40 μ M acetosyringone in the presence of appropriate antibiotics. Overnight cultures were pelleted at 3,500 X *g* for 10 min, suspended in 10 mM MES, 10 mM MgCl₂ and 150 μ M acetosyringone and incubated at room temperature for 5 h under constant shaking. Cultures were diluted at 0.5 OD and mixed with the same volume of a 0.5 OD *Agrobacterium* culture carrying the pBIN61-P19 vector prior to infiltration. Six to eight week-old *Nicotiana benthamiana* plants were infiltrated using a needleless syringe. All samples were harvested at 2.5 days post infiltration.

Recombinant expression and labeling of viral proteins.

Recombinant viral proteins P1, P2, P3, P6/TAV and GST-P1P2(D2) were expressed in *Escherichia coli* BL21(DE3)pLysS. Their expression was induced in bacteria culture in exponential phase by adding 1mM Isopropyl β -D-1-thiogalactopyranoside. Bacteria were then grown at 28°C for 4 h, pelleted and suspended in heart muscle kinase (HMK) buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 12 mM MgCl₂). After being sonicated 3 times for 10 s, lysates were centrifuged at 11,000 X *g*, for 10 min at 4°C and the inclusions bodies were suspended in 0.5 mL HMK buffer. 30 μ L of the inclusion bodies suspension were phosphorylated in the presence of 25 μ Ci [γ -³²P]ATP and 25 U of HMK (Sigma-Aldrich) in 150 μ L HMK buffer, at room temperature for 1.5 h. ATP in excess was eliminated using an illustra MicroSpin G-25 Column (GE Healthcare).

Far Western assays.

The GST-P1P2(D2)-containing bacterial fraction was separated by SDS-PAGE, electrotransferred to a PVDF membrane (Immobilon-P; Merck Millipore) and saturated in phosphate-buffered saline 1x (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄), Tween 20 1%, milk 5% at room temperature for 12 h. The membrane was incubated either with ³²P-radiolabelled viral proteins or with purified virus particles at 4°C for 12 h and washed in phosphate-buffered saline 1x-Tween 20 1% for 10 min. The interaction between the fusion protein and virus particles was detected with anti-P4 polyclonal antibodies (1:10,000 dilution) and alkaline phosphatase-labeled secondary antibodies (1:5,000 dilution).

RESULTS

In silico analysis reveals that splice sites of 35S RNA are conserved among CaMV strains and other Caulimoviruses.

The comparison of the complete DNA sequence of ten CaMV strains revealed that the 5' part of their pregenomic 35S RNA possesses three splice donor sites described for CaMV Cabb-S strain (D4 in the 5' UTR and D1 and D2 in ORF I; nomenclature from (17)), and the splice acceptor site (A in ORF II). The exon-intron boundaries are conserved in all the sequenced CaMV strains (Fig. 1B): AG-**GU** sequence for D1 and D4, CG**GU** for D2 and **AGGA** for A (the 5' and 3' invariant dinucleotides characteristic of the eukaryotic intron extremities are in bold). Moreover, these three sites are conserved in all the 67 genome sequences that have recently been released (26; C. Bouton, unpublished data). By contrast, the donor site D3 used for splicing in Cabb-S (17) is degenerated in the other strains; notably, the dinucleotide GU is mutated into GA. The 35S RNA contains a putative donor site upstream of D3 bearing a sequence that is conserved among all CaMV strains except strain BBC. However, compared to the other donor sites, its use by the spliceosome is questionable since the corresponding exon-intron junction weakly anneals with the U1 small nuclear RNA (U1 snRNA; Fig. 1B) of CaMV host plants such as *Arabidopsis thaliana* and turnip (27). Several other cryptic donor and acceptor sites are present along the 5' region of CaMV pregenomic RNA. Analysis of the 5' part of the 35S RNA sequence of some other members of the *Caulimovirus* genus like *Figwort mosaic virus*, *Lamium leaf distortion associated virus* and *Dahlia mosaic virus* showed that D1 and/or D2 are con-

served in these virus species, as well as the acceptor site (Fig. 1C). This strongly suggests that splicing might also be important for the infectivity of *Caulimoviruses* in general.

Taken together, *in silico* analyses strongly suggest that alternative splicing of the pregenomic 35S RNA is a conserved phenomenon among CaMV strains and that it may also occur in other *Caulimovirus* species.

35S RNA of CaMV Cabb B-JI strain undergoes alternative splicing.

To determine if 35S RNA of CaMV Cabb B-JI, our model strain, is spliced during the infectious cycle, total RNA was extracted from CaMV-infected turnip plants at 21 dpi, separated by agarose-urea gel electrophoresis and analyzed by northern blot. The full-length 35S RNA (8,200 nucleotides) was detected with an oligonucleotide probe (*int*) targeting a sequence immediately upstream of the splice acceptor site (Fig. 2A and 2B). Full-length and smaller 35S RNA-derived versions (< 7,600 nucleotides) were detected with a probe (*ex*) that is complementary to a sequence immediately downstream of the splice acceptor site (Fig. 2A and 2B). No band was present when RNA from mock-inoculated plants was tested with both probes (Fig. 2B). RT-PCR on total RNA was performed to amplify the 5' region of the 35S RNA encompassing the putative splicing sites (Fig. 2C). Analysis of the amplification products by electrophoresis showed that the major amplicons were consistent with the predicted sizes of cDNAs corresponding to the non-spliced 35S RNA (2,400 bp) and its D1-, D2- (1,800 bp) and D4- (800 bp) spliced isoforms. Sequencing of these amplicons unambiguously showed that they correspond to the three predicted 35S RNA spliced species and, consequently, that CaMV Cabb B-JI follows the same alternative splicing pattern as Cabb-S strain (except for the D3 site). As expected from the *in silico* analysis, a cDNA corresponding to D3-A spliced RNA described for Cabb-S could never be amplified from RNAs extracted from Cabb B-JI-infected turnips, confirming that this splice donor site does not exist in this strain. Amplification products were obtained in the minus-DNase/minus-RT lane but not in the plus-DNase/minus-RT lane (Fig. 2C), indicating the presence of contaminating viral DNA. The profile observed after gel electrophoresis was very similar to the ones obtained after RT, and sequencing revealed deletions identical to D1- and D2-A splicing, confirming the reverse-transcription of these spliced isoforms upon infection.

Fusion proteins P1P2 are undetectable in infected plants.

The selection of splice donor sites D1 and D2 (and D3 in Cabb-S strain) and acceptor site A creates fusion ORFs between the 5' part of ORF I (whose size varies depending on the donor site) and the 3' part of ORF II, which remains unchanged in the fusion ORFs. The resulting P1P2 (D1 and D2) fusion proteins have a theoretical molecular mass of about 31 kDa. In spite of numerous attempts, fusion P1P2 proteins were not detected when we used polyclonal anti-P1 and anti-P2 antibodies, whereas the P1 protein (46 kDa), its cleavage product (38 kDa) (28) and P2 (18 kDa) were revealed with these antibodies (Fig. 3A), suggesting that the fusion proteins are weakly expressed and/or degraded shortly after their synthesis. Detection of the P1P2 fusion proteins was also unsuccessful in cultured BY2 tobacco cells transfected with a recombinant plasmid coding for this viral protein fused at its N-terminus with eGFP, whereas the latter expressed alone diffused in the cytoplasm (A. Geldreich, unpublished data). Moreover, the expression of HA- and myc-tagged P1P2(D2) was investigated by agroinfiltration of *Nicotiana benthamiana* leaves. Agrobacteria carrying plasmids encoding HA-eGFP and myc-eGFP were used to infiltrate control leaves. Both control proteins were always revealed with antibodies raised against their tag. Only in one experiment did we detect proteins having the expected molecular mass of tagged P1P2(D2) (about 39 kDa). Mostly, the tagged P1P2(D2) proteins were not immunodetected in crude protein extracts (Fig. 3B) although their mRNAs were detected by RT-PCR in the infiltrated leaves (Fig. 3C). It has been shown that P2 self-interacts (29), interacts with P3 (30)

and P6/TAV (31). P1, in addition to oligomerizing to form tubules (18) allowing cell-to-cell movement, also interacts with P3 (32) and P6/TAV (33). We therefore hypothesized that fusion proteins could interact with viral partners of full-length P1 and/or P2 proteins. To assess the putative interactions between P1P2 fusion proteins and viral proteins, we performed far western experiments on GST-P1P2(D2) fusion protein expressed in *E. coli*, using recombinant and radiolabelled viral proteins P1, P2, P3 and P6 or purified CaMV particles as overlay. GST-P1P2(D2) interacted with purified viral particles, P2 and P3, but neither with P1 nor with P6/TAV (Fig. 3D), whereas GST alone did not interact with any of these proteins.

The fact that fusion proteins are not detected in infected plants but can interact *in vivo* with some viral partners of P1 and P2 suggests that these proteins are rapidly degraded after their expression, possibly in order to prevent potential interference with the functions of P1 and/or P2.

A cryptic donor site rescues splicing upon mutation of D1, D2 and D4 sites.

Previous studies have suggested that splicing of the 35S RNA is essential for CaMV since it downregulates the expression of P1 and P2, whose early accumulation could be deleterious for the infectious cycle (20). However, as one donor site should be sufficient for this purpose, and because all CaMV strains have conserved the D1, D2 and D4 donor sites, we hypothesized that alternative splicing must have other regulatory functions. To assess the biological relevance of the splice donor sites, we inactivated them by site-directed mutagenesis in the viral vector pMD324 containing the complete Cabb B-JI genome. The mutations introduced into the exon-intron junctions abolished the interaction with U1 snRNA but, in the case of D1 and D2, maintained the coding sequence of ORF I in order to safeguard the function of movement protein P1 (Fig. 4A). Single, double and triple mutants were generated, and the resulting plasmids were mechanically inoculated in three-leaf stage turnip plants. Ten plants were inoculated with each splicing mutant in four independent experiments. All CaMV mutants resulted to be infectious on turnip plants including, surprisingly, the triple mutant D1D2D4 that was expected to be lethal for the virus due to predicted abolition of splicing (17, 20). Two weeks after inoculation, all turnip plants expressed the same systemic symptoms (mosaic, vein-clearing and stunting) as plants inoculated with the wild-type (wt) CaMV genome did, and no delay in symptom appearance was observed. Sequencing of viral DNA extracted from systemic leaves did not reveal any reversion of mutations. Total RNA was extracted from plants infected with wt CaMV or splicing mutants, and the 35S RNA 5' region encompassing the donor and acceptor sites was amplified by RT-PCR. Amplification products obtained from wt CaMV, all single mutants and D1D4 and D2D4 double mutants were analyzed by gel electrophoresis and revealed a similar pattern, displaying a major band (1.8 kb) and several minor bands, among which the unspliced 35S RNA (2.4 kb) (Fig. 4B). Interestingly, the major band obtained with D1D2 and D1D2D4 mutants was smaller (1.6 kb) than those observed with the other splicing mutants. This amplicon was extracted from the agarose gel, cloned and sequenced. The sequence corresponded to a new spliced isoform of 35S RNA involving a donor site that is located at position 653 in ORF I (donor site Da, Table 1 and Fig. 5A) and the splice acceptor site A. This splicing event, as for D1 and D2, created an in-frame fusion between ORF I and ORF II.

In conclusion, our results show that the donor splice sites of the 35S RNA are not relevant *per se* for CaMV infectivity. On the other hand, splicing itself appears to be, since the virus used a cryptic splice donor site upon their inactivation.

CaMV Cabb B-JI and Cabb-S 35S RNA exhibit a complex splicing pattern.

CaMV Cabb B-JI was still infectious when the quadruple mutant (D1, D2, D4 and Da) was inoculated in turnip plants, strongly suggesting the presence of other splice sites in the 35S RNA. This prompted us to investigate in detail the splicing taking place within Cabb B-JI 35S RNA by focusing on

its 5' part, since amplification by RT-PCR of the sequence located downstream of the acceptor site did not reveal any spliced RNA (C. Bouton, unpublished data). Amplicons were obtained by RT-PCR performed on DNase-treated total RNA from turnip plants infected with wt Cabb B-JI, using a pair of primers flanking the 5' region, from the 35S RNA 5' terminus to the end of ORF II. After cloning in pGEM-T Easy vector and sequencing, sequences of the amplicons were aligned with 35S RNA. In addition to the aforementioned spliced RNAs, we found eleven new spliced RNA isoforms that arose from the alternative use of two other donor sites (Db and Dc) and six acceptor sites (Aa to Af; Table 1 and Fig. 5A and 5C). All these new splice sites conform to splice donor and acceptor site consensus sequences GU...AG and were named according to order of discovery. Donor site Db is located in sORF E of the 5' UTR close to the D4 site, whereas donor site Dc and the six new acceptor sites are in ORF I (Fig. 5A). Spliced isoforms resulted from five single, nine double and one triple splicing events, and contained various fusion ORFs involving sORF E, ORF I, ORF II and sequences in other reading frames localized in ORF I and ORF II regions (Fig. 5C). Remarkably, the use of donor site Db, which is localized in the 5'-UTR, also leads to the formation of fusion ORFs. In fact, several isoforms possessing fusion ORFs involving sORF E were detected (for example isoforms Db-A and Db-Ab-D2-A; Fig. 5C). When we analyzed the viral RNAs extracted from turnip plants infected with CaMV Cabb-S strain, we also discovered a complex splicing pattern of the 35S RNA 5' region characterized by twelve spliced isoforms (Fig. 5C). Several splice sites are located as in Cabb B-JI (sites Da, Db, Ab, Ac and Ad; Fig. 5A and 5B, Table 1). The Cabb-S strain in a few instances also used a splice acceptor site (Ah, position 1515) downstream of the splice acceptor site that was previously considered unique (A, position 1508). We did not detect any spliced RNA involving this acceptor site in turnip plants infected with Cabb B-JI strain, except when we inactivated the major acceptor site (see below). In some sequenced clones and for both strains, it is noteworthy that we found deletions unrelated to splicing. Some of them involved borders around the two sequence discontinuities $\Delta 1$ and $\Delta 3$ that are remnants of the reverse transcription process and located at the beginning of ORF VII and the end of ORF II, respectively (34). This implies that recombination occurs between these discontinuities.

Taken together, the CaMV transcriptome analysis clearly demonstrates that the 35S RNAs of CaMV Cabb B-JI and Cabb-S strains both exhibit a complex alternative splicing pattern involving multiple splice donor and acceptor sites, most of them being located within ORF I.

Cryptic splice acceptor sites are activated to maintain splicing of CaMV 35S RNA.

The fact that disruption of splice donor sites can be rescued by the use of other sites suggests that acceptor site A is not as crucial for infectivity as it was previously thought to be (17, 20). We therefore introduced a mutation previously considered sufficient to prevent the use of the acceptor site A in Cabb B-JI genome (**AGG** to AGA, mutant JI-ma3; Fig. 6A). In our experimental conditions (mechanical inoculation of 8 μ g of linearized viral DNA per plant), turnip plants inoculated with the mutant JI-ma3 developed the same systemic symptoms (mosaic, vein clearing and stunting) as those inoculated with the wild-type strain (Fig. 6B). To confirm this result, we tested the infectivity of a Cabb-S genome containing the same mutation (mutant S ma3; Fig. 6A), finding this mutant to be infectious too (Fig. 6B). Sequencing of the 5' part of 35S RNA in the viral progeny of both strains showed neither a reversion of the introduced mutations nor other modifications such as deletion and/or introduction of a stop codon in ORF II, which could have explained infectivity of these mutants. In fact, RT-PCR amplification and sequencing of the 5' region of the CaMV 35S RNA isoforms revealed that splice acceptor site A (position 1509 in Cabb B-JI and 1508 in Cabb-S) was still active despite the introduced mutation (Fig. 6A and 6C, clones 14 and 43). For some isoforms, the AG dinucleotide immediately downstream (acceptor site Ah, Fig. 5B) was used as acceptor site (Fig. 6A and 6C, clone 17). To fully disrupt site A, we introduced a mutation into the dinucleotide AG (**AGG** to UGC, mutant B-JI mutA; Fig. 6A) and inoculated this mutant in turnip plants. Plants inoculated with B-JI mutA expressed typical CaMV systemic symptoms with no delay compared to Cabb B-JI-infected plants (Fig. 6B) and they accumulated similar

quantities of viral proteins P2 and CP (Fig. 6D). Sequence analysis of the spliced viral RNAs confirmed mutation at site A and revealed that two other downstream AG dinucleotides were used as acceptor sites for splicing (Fig. 6A).

In conclusion, these results show that CaMV maintains splicing of its 35S RNA thanks to the use of cryptic splice acceptor sites in ORF II, thus explaining why inactivation of acceptor site A was not lethal for CaMV. They also reinforce the hypothesis that splicing of 35S RNA is essential for CaMV replication.

DISCUSSION

Alternative splicing of 35S RNA potentially occurs in many CaMV strains since they possess three of the four splice donor sites and the splice acceptor site described for Cabb-S strain (17). Some of these splice sites have also been found in the 35S RNA of other members of the *Caulimovirus* genus, suggesting that splicing might be a key common process for these viruses. Analysis of amplified cDNAs obtained by RT-PCR of viral RNAs extracted from Cabb B-JI-infected turnips revealed that the splicing pattern of 35S RNA is more complex than previously described for Cabb-S strain: in fact, we identified 15 isoforms resulting from the use of 6 splice donor sites and 7 acceptor sites. Such complexity was also observed in Cabb-S strain where 12 isoforms of 35S RNA were found. We never detected spliced RNAs arising from splicing events downstream of ORF II, which is in contrast with the behavior observed in *Figwort mosaic virus* infection, where splicing occurs in the 3' part of the 35S RNA, leading to an in-frame fusion between the 5' region of ORF IV and the 3' region of ORF V that code for the capsid protein and reverse transcriptase, respectively (13). The spliced RNAs possess at their 5' and 3' ends a repeated region (R region) involved in intra- or inter-molecular replication jumps, the polypurine tracts for the synthesis of the second DNA strand and, except for the isoforms where splicing involved the donor sites located upstream of ORF VII, the primer binding site (PBS) where the cellular Met-tRNA initiator used as primer by the reverse transcriptase binds the 35S RNA (35, 36). As expected, some spliced RNAs are reverse transcribed in CaMV-infected cells into double-stranded circular DNA molecules, excluding the aforementioned RNA species deprived of PBS by the splicing event. Naturally occurring site-specific deletion mutants resulting from reverse transcription of a spliced RNA have been previously observed in plants infected with CaMV-S Japan isolate (16) and in plants infected with FMV (13). Small circular CaMV DNA molecules of various sizes have been detected through electron microscopy in the nuclei isolated from CaMV-infected turnip plants (37). Some of these molecules may correspond to reverse transcription products of spliced RNAs. Viral particles containing genomes with deletions due to reverse transcription of spliced RNAs have been isolated from HBV-infected patients and from FMV- and CaMV-infected plants (8, 13, 16, 38, A. Geldreich, unpublished data) suggesting that they are transported into neighboring cells. Consequently, these shorter genomes could participate in the production of spliced-like RNAs upon transcription and artificially increase the pool of spliced RNAs among the global population of viral RNAs.

Multiple spliced RNA isoforms arising from single-, double- or triple-splicing events were characterized in CaMV-infected plants. Almost all spliced RNAs contained a new ORF coming from a fusion involving the sORF E, the ORFs I and II or even other reading frames located within ORFs I and II. A complex population of spliced viral transcripts comprising at least 15 spliced isoforms was also characterized in patients chronically infected by hepatitis B virus (HBV) (39, 40). Singly spliced RNAs code for a fusion protein which is involved in the HBV pathogenesis and may participate in the hemostatic abnormality observed in HBV-related liver disease (40, 41), whereas the doubly-spliced RNAs encode a novel protein that is able to activate transcription (42). In CaMV, the fused ORF I-ORF II found in some 35S RNA isoforms code for chimeric proteins (P1P2) formed by the N-terminus of P1 and the C-terminus of P2. The function, if any, of these P1P2 proteins remains unknown. It has been hypothesized that they could play a role in the development of a systemic infection in plants growing under specific conditions, especially at high light intensity (17, 43). The present data speak in favor of

a structural rather than functional meaning of the multiple fusion ORFs generated upon alternative splicing. In-frame fusions in the 5' part of the 35S RNA allow both to downregulate expression of ORFs I and II to prevent the appearance of multiple ORFs, which could be deleterious for the P6/TAV-mediated translation of the downstream ORFs. Indeed, none of the isoforms we detected contained more ORFs than the unspliced 35S RNA. Furthermore, these putative proteins were not detected by western blots performed on crude protein extracts and cell wall-enriched fractions from CaMV-infected plants, using anti-P1 and anti-P2 antibodies in parallel (antibodies directed specifically against the peptide forming the P1P2 junctions were not available). This suggested that the fusion proteins are seldom expressed from spliced RNAs during CaMV infection or, alternatively, that they are rapidly degraded upon their synthesis *in planta*. This last hypothesis is reinforced by the fact that myc-/HA-tagged or GFP-fused P1P2 proteins were almost never detected in agroinfiltrated *Nicotiana benthamiana* leaves, whereas their mRNAs were detected by RT-PCR. These proteins could be degraded by the proteasome since analysis of their amino acid sequence did not reveal any instability sequence (PEST) similar to those found in the CaMV P4 protein (44). The rapid degradation of P1P2 would prevent their interference with the activities of full-length P1 and P2 proteins since the fusion proteins are able to interact *in vitro*, in farwestern experiments, with P2, P3 and purified virions. In the latter case, the interaction is probably mediated by P3, since this protein is associated to viral particles (45). These interactions, which should operate through the P2 C-terminus (29, 30) of the fusion proteins, could hinder the systemic spread and the viral transmission by aphids through competition with interactions involving the full-length P1, P3 and viral particles required for intracellular and cell-to-cell movements of CaMV (32, 46), and through prevention of the formation of the viral transmission complex (47).

Another aim of our study was to understand the functional relevance of 35S RNA alternative splicing in CaMV pathogenesis. We noticed that the splice donor sites D1 and D2 in ORF I are situated within a GGT triplet coding for glycine, even if this amino acid can be specified by four GGN codons. This could hint toward a selective pressure to conserve both the splice sites and the glycine codons (Fig. 1B). Glycine residues at these positions are found in the movement protein of all CaMV strains and two other Caulimoviruses, *Figwort mosaic virus* and *Carnation Etched Ring virus* (28). They evidently play an important role in cell-to-cell movement, since mutations within the donor site leading to their substitution by another amino acid abolish CaMV infectivity (C. Bouton, unpublished data). We expected that modifications of the alternative splicing scheme would alter the virulence of CaMV on host plants. However, individual or collective mutation of the first described donor sites (17) did not alter the pathogenicity of Cabb B-JI strain in turnips. Similarly, mutation of CaMV Cabb-S donor sites did not modify the phenotype of infected turnip plants. Apparently, mutation of splice sites stimulates the selection of seldom-used sites and/or activates cryptic splice sites (Fig. 4B) to maintain a high level of spliced RNAs in infected tissues. A similar activation of cryptic splice sites following the inactivation of a donor site has been observed in the CaMV-S strain (16) and in *Mastrevirus* Maize Streak Virus (9) infections. The multiplicity of splice sites in CaMV stimulated us to determine if alternative splicing of the 35S RNA has roles outside the downregulation of the expression of aphid transmission factor P2 and movement protein P1 (17, 20). The presence of splice sites also explains why mutations disrupting splice sites of CaMV Cabb B-JI and Cabb-S strains are not lethal for the virus in our experimental conditions. These observations contradict previous experiments, in which Cabb-S strain was not infectious in turnip plants when the use of splice acceptor site A was prevented (17), except for two plants which finally displayed symptoms with a three-week delay. Sequencing of the viral genome revealed that the mutated acceptor site was not restored, but a small deletion (58 bp) introducing a premature termination codon in ORF II was found (17). We observed neither reversion of mutations nor any modification in the ORF II of our mutants upon infection. Our results show that nearby AG dinucleotides can be used as rescue acceptor sites upon disruption of acceptor site A. This suggests that alternative splicing in CaMV achieves strong versatility through multiplicity of splice sites within the 5' part of 35 S RNA and flexibility in their use, thus constituting a robust means to downregulate expression of ORFs I and II.

Alternative splicing could have other roles besides downregulating the expression of ORFs I and II. Recently, it was shown in *Arabidopsis* that spliced RNAs are less effective substrates for RNA silencing mediated by the RNA-dependent RNA polymerase RDR6, compared to intron-less transcripts (48, 49). During CaMV infection, the 35S RNA is subject to degradation by Dicer-like proteins (50–52), but processing of endogenous RDR6-derived double-stranded RNA is impaired, suggesting that introns partially preserve viral RNAs from excessive degradation by the silencing machinery. Furthermore, the splice donor sites located in the leader region (Db and D4) might be involved in the suppression of premature polyadenylation at the 5' end of the 35S RNA, as described for *Retroviridae* HIV-1 (53, 54) and foamy viruses (55). The U1 small nuclear ribonucleoprotein (U1 snRNP), in addition to its role in the splicing process, interferes with the pre-mRNA 3' end processing and protects some pre-mRNAs from premature polyadenylation (56). This regulatory mechanism ensures transcriptome integrity by regulating mRNA length (57). In HIV-1 and foamy viruses, the polyadenylation signal (poly(A) signal) is found at both extremities of the viral progenome, in the redundant 5' and 3' long terminal repeats (LTRs). It has been shown that the poly(A) signal at the 5' LTR is skipped through a mechanism involving the U1 snRNP-mediated recognition of a splice donor site located a few hundred nucleotides upstream (HIV-1) (53) or downstream (foamy viruses) (55), thus avoiding a premature termination of transcription. The CaMV poly(A) signal is located about 180 nucleotides downstream of the transcription start site, so polyadenylation needs to be regulated to allow transcription of the complete genome. The Db and/or D4 donor sites are about 300 nucleotides downstream of the poly(A) signal. The putative role of D4 and Db in the regulation of CaMV premature polyadenylation should therefore be assessed in the future.

The splice donor and acceptor sites described in the present study allow CaMV to produce multiple spliced 35S RNAs, which do not encode for P1 and P2 proteins while still being appropriate mRNAs for P6/TAV-mediated expression of viral proteins via translation reinitiation. Any inactivation of a splice site would be ineffective in impairing CaMV infectivity. The synthesis of intron-deleted DNA by reverse transcription could be a strategy to increase the population of spliced RNAs in a spliceosome-independent fashion. Whether CaMV alternative splicing is regulated by *cis*-elements (splicing enhancers and silencers) is another point which must be elucidated in the future, to gain insight on the temporal regulation and the biological relevance of the various 35S RNA spliced isoforms.

ACKNOWLEDGMENTS

We thank Martin Drucker for providing us with the pBS-BJI viral vector, Andy Maule for the anti-P1 antiserum and Marco Incabone for critical reading of the manuscript. C.B. was supported by a PhD fellowship from the French Ministry of Research.

REFERENCES

1. **Maniatis T, Tasic B.** 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**:236–243.
2. **Reddy ASN.** 2007. Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol.* **58**:267–294.
3. **Damgaard CK, Kahns S, Lykke-Andersen S, Nielsen AL, Jensen TH, Kjems J.** 2008. A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol. Cell* **29**:271–278.
4. **Valencia P, Dias AP, Reed R.** 2008. Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **105**:3386–3391.
5. **Zheng Z-M.** 2010. Viral oncogenes, noncoding RNAs, and RNA splicing in human tumor viruses. *Int. J. Biol. Sci.* **6**:730–755.

6. **Inglis SC, Brown CM.** 1981. Spliced and unspliced RNAs encoded by virion RNA segment 7 of influenza virus. *Nucleic Acids Res.* **9**:2727–2740.
7. **Bayliss J, Lim L, Thompson AJV, Desmond P, Angus P, Locarnini S, Revill PA.** 2013. Hepatitis B virus splicing is enhanced prior to development of hepatocellular carcinoma. *J. Hepatol.* **59**:1022–1028.
8. **Lee GH, Wasser S, Lim SG.** 2008. Hepatitis B pregenomic RNA splicing--the products, the regulatory mechanisms and its biological significance. *Virus Res.* **136**:1–7.
9. **Wright EA, Heckel T, Groenendijk J, Davies JW, Boulton MI.** 1997. Splicing features in maize streak virus virion- and complementary-sense gene expression. *Plant J. Cell Mol. Biol.* **12**:1285–1297.
10. **Kiss-László Z, Hohn T.** 1996. Pararetro- and retrovirus RNA: splicing and the control of nuclear export. *Trends Microbiol.* **4**:480–485.
11. **Goodall GJ, Filipowicz W.** 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* **58**:473–483.
12. **Fütterer J, Potrykus I, Valles Brau MP, Dasgupta I, Hull R, Hohn T.** 1994. Splicing in a plant pararetrovirus. *Virology* **198**:663–670.
13. **Scholthof HB, Wu FC, Richins RD, Shepherd RJ.** 1991. A naturally occurring deletion mutant of figwort mosaic virus (caulimovirus) is generated by RNA splicing. *Virology* **184**:290–298.
14. **Haas M, Bureau M, Geldreich A, Yot P, Keller M.** 2002. Cauliflower mosaic virus: still in the news. *Mol. Plant Pathol.* **3**:419–429.
15. **Ryabova LA, Pooggin MM, Hohn T.** 2006. Translation reinitiation and leaky scanning in plant viruses. *Virus Res.* **119**:52–62.
16. **Hirochika H, Takatsuji H, Ubasawa A, Ikeda JE.** 1985. Site-specific deletion in cauliflower mosaic virus DNA: possible involvement of RNA splicing and reverse transcription. *EMBO J.* **4**:1673–1680.
17. **Kiss-László Z, Blanc S, Hohn T.** 1995. Splicing of cauliflower mosaic virus 35S RNA is essential for viral infectivity. *EMBO J.* **14**:3552–3562.
18. **Perbal MC, Thomas CL, Maule AJ.** 1993. Cauliflower mosaic virus gene I product (P1) forms tubular structures which extend from the surface of infected protoplasts. *Virology* **195**:281–285.
19. **Armour SL, Melcher U, Pirone TP, Lyttle DJ, Essenberg RC.** 1983. Helper component for aphid transmission encoded by region II of cauliflower mosaic virus DNA. *Virology* **129**:25–30.
20. **Froissart R, Uzest M, Ruiz-Ferrer V, Drucker M, Hébrard E, Hohn T, Blanc S.** 2004. Splicing of Cauliflower mosaic virus 35S RNA serves to downregulate a toxic gene product. *J. Gen. Virol.* **85**:2719–2726.
21. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**:539.
22. **Franck A, Guilley H, Jonard G, Richards K, Hirth L.** 1980. Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* **21**:285–294.
23. **Delsený M, Hull R.** 1983. Isolation and characterization of faithful and altered clones of the genomes of cauliflower mosaic virus isolates Cabb B-JI, CM4-184, and Bari I. *Plasmid* **9**:31–41.
24. **Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T.** 2007. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104**:34–41.
25. **Leh V, Yot P, Keller M.** 2000. The cauliflower mosaic virus translational transactivator interacts with the 60S ribosomal subunit protein L18 of *Arabidopsis thaliana*. *Virology* **266**:1–7.
26. **Yasaka R, Nguyen HD, Ho SYW, Duchêne S, Korkmaz S, Katis N, Takahashi H, Gibbs AJ, Ohshima K.** 2014. The Temporal Evolution and Global Spread of Cauliflower mosaic virus, a Plant Pararetrovirus. *PLoS One* **9**:e85641.
27. **Filipowicz W, Kiss T, Marshallsay C, Waibel F.** 1990. U-snRNA genes, U-snRNAs and U-snRNPs of higher plants. *Mol. Biol. Rep.* **14**:125–129.

28. **Thomas CL, Maule AJ.** 1995. Identification of structural domains within the cauliflower mosaic virus movement protein by scanning deletion mutagenesis and epitope tagging. *Plant Cell* **7**:561–572.
29. **Hébrard E, Drucker M, Leclerc D, Hohn T, Uzest M, Froissart R, Strub JM, Sanglier S, van Dorsselaer A, Padilla A, Labesse G, Blanc S.** 2001. Biochemical characterization of the helper component of Cauliflower mosaic virus. *J. Virol.* **75**:8538–8546.
30. **Leh V, Jacquot E, Geldreich A, Hermann T, Leclerc D, Cerutti M, Yot P, Keller M, Blanc S.** 1999. Aphid transmission of cauliflower mosaic virus requires the viral PIII protein. *EMBO J.* **18**:7077–7085.
31. **Lutz L, Raikhy G, Leisner SM.** 2012. Cauliflower mosaic virus major inclusion body protein interacts with the aphid transmission factor, the virion-associated protein, and gene VII product. *Virus Res.* **170**:150–153.
32. **Stavolone L, Villani ME, Leclerc D, Hohn T.** 2005. A coiled-coil interaction mediates cauliflower mosaic virus cell-to-cell movement. *Proc. Natl. Acad. Sci. U. S. A.* **102**:6219–6224.
33. **Hapiak M, Li Y, Agama K, Swade S, Okenka G, Falk J, Khandekar S, Raikhy G, Anderson A, Pollock J, Zellner W, Schoelz J, Leisner SM.** 2008. Cauliflower mosaic virus gene VI product N-terminus contains regions involved in resistance-breakage, self-association and interactions with movement protein. *Virus Res.* **138**:119–129.
34. **Richards KE, Guilley H, Jonard G.** 1981. Further characterization of the discontinuities in cauliflower mosaic virus DNA. *FEBS Lett.* **134**:67–70.
35. **Pfeiffer P, Hohn T.** 1983. Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* **33**:781–789.
36. **Guilley H, Richards KE, Jonard G.** 1983. Observations concerning the discontinuous DNAs of cauliflower mosaic virus. *EMBO J.* **2**:277–282.
37. **Olszewski NE, Guilfoyle TJ.** 1983. Nuclei purified from cauliflower mosaic virus-infected turnip leaves contain subgenomic, covalently closed circular cauliflower mosaic virus DNAs. *Nucleic Acids Res.* **11**:8901–8914.
38. **Roberts K, Love AJ, Laval V, Laird J, Tomos AD, Hooks MA, Milner JJ.** 2007. Long-distance movement of Cauliflower mosaic virus and host defence responses in Arabidopsis follow a predictable pattern that is determined by the leaf orthostichy. *New Phytol.* **175**:707–717.
39. **Sommer G, Heise T.** 2008. Posttranscriptional control of HBV gene expression. *Front. Biosci. J. Virtual Libr.* **13**:5533–5547.
40. **Soussan P, Tuveri R, Nalpas B, Garreau F, Zavala F, Masson A, Pol S, Brechot C, Kremsdorf D.** 2003. The expression of hepatitis B spliced protein (HBSP) encoded by a spliced hepatitis B virus RNA is associated with viral replication and liver fibrosis. *J. Hepatol.* **38**:343–348.
41. **Park G-S, Kim H-Y, Shin H-S, Park S, Shin H-J, Kim K.** 2008. Modulation of hepatitis B virus replication by expression of polymerase-surface fusion protein through splicing: implications for viral persistence. *Virus Res.* **136**:166–174.
42. **Chen W-N, Chen J-Y, Lin W-S, Lin J-Y, Lin X.** 2010. Hepatitis B doubly spliced protein, generated by a 2.2 kb doubly spliced hepatitis B virus RNA, is a pleiotropic activator protein mediating its effects via activator protein-1- and CCAAT/enhancer-binding protein-binding sites. *J. Gen. Virol.* **91**:2592–2600.
43. **Qiu SG, Wintermantel WM, Sha Y, Schoelz JE.** 1997. Light-dependent systemic infection of solanaceous species by cauliflower mosaic virus can be conditioned by a viral gene encoding an aphid transmission factor. *Virology* **227**:180–188.
44. **Karsies A, Hohn T, Leclerc D.** 2001. Degradation signals within both terminal domains of the cauliflower mosaic virus capsid protein precursor. *Plant J. Cell Mol. Biol.* **27**:335–343.
45. **Dautel S, Guidasci T, Pique M, Mougeot JL, Lebourier G, Yot P, Mesnard JM.** 1994. The full-length product of cauliflower mosaic virus open reading frame III is associated with the viral particle. *Virology* **202**:1043–1045.
46. **Sánchez-Navarro J, Fajardo T, Zicca S, Pallás V, Stavolone L.** 2010. Caulimoviridae tubule-guided transport is dictated by movement protein properties. *J. Virol.* **84**:4109–4112.

47. **Leh V, Jacquot E, Geldreich A, Haas M, Blanc S, Keller M, Yot P.** 2001. Interaction between the open reading frame III product and the coat protein is required for transmission of cauliflower mosaic virus by aphids. *J. Virol.* **75**:100–106.
48. **Christie M, Croft LJ, Carroll BJ.** 2011. Intron splicing suppresses RNA silencing in Arabidopsis. *Plant J. Cell Mol. Biol.* **68**:159–167.
49. **Christie M, Carroll BJ.** 2011. SERRATE is required for intron suppression of RNA silencing in Arabidopsis. *Plant Signal. Behav.* **6**:2035–2037.
50. **Moissiard G, Voinnet O.** 2006. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. *Proc. Natl. Acad. Sci. U. S. A.* **103**:19593–19598.
51. **Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, Park H-S, Vazquez F, Robertson D, Meins F Jr, Hohn T, Pooggin MM.** 2006. Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.* **34**:6233–6246.
52. **Shivaprasad PV, Rajeswaran R, Blevins T, Schoelz J, Meins F Jr, Hohn T, Pooggin MM.** 2008. The CaMV transactivator/viroplasm interferes with RDR6-dependent trans-acting and secondary siRNA pathways in Arabidopsis. *Nucleic Acids Res.* **36**:5896–5909.
53. **Ashe MP, Griffin P, James W, Proudfoot NJ.** 1995. Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev.* **9**:3008–3025.
54. **Sajic R, Lee K, Asai K, Sakac D, Branch DR, Upton C, Cochrane A.** 2007. Use of modified U1 snRNAs to inhibit HIV-1 replication. *Nucleic Acids Res.* **35**:247–255.
55. **Schrom E-M, Moschall R, Hartl MJ, Weitner H, Fecher D, Langemeier J, Böhne J, Wöhrl BM, Bodem J.** 2013. U1snRNP-mediated suppression of polyadenylation in conjunction with the RNA structure controls poly (A) site selection in foamy viruses. *Retrovirology* **10**:55.
56. **Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G.** 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**:664–668.
57. **Langemeier J, Radtke M, Böhne J.** 2013. U1 snRNP-mediated poly(A) site suppression: beneficial and deleterious for mRNA fate. *RNA Biol.* **10**:180–184.

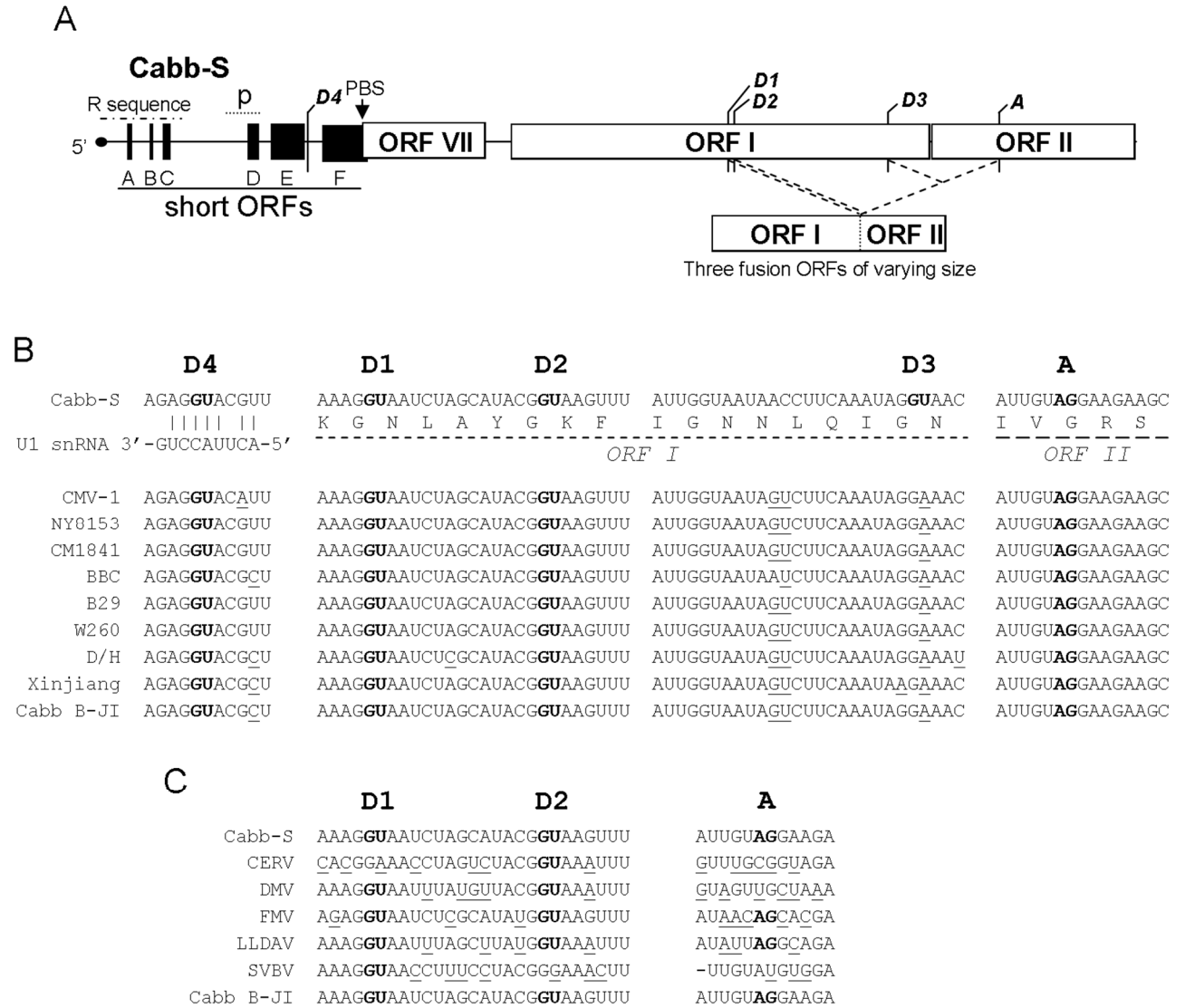


FIG 1: Conservation of 35S RNA splice sites among CaMV strains and other *Caulimoviruses*. (A) Map of the 5' region (about 2,500 nucleotides) of 35S RNA of Cabb-S strain. Empty boxes represent major ORFs, while black boxes represent short ORFs (sORFs) in the leader region. D1 to D4 are the splice donor sites and A, the acceptor splice site described by Kiss-László *et al.* (1995). Dotted lines mark the splicing events leading to in-frame fusions between the 5' and 3' extremities of ORF I and ORF II, respectively. The putative encapsidation signal (p) and the primer binding site (PBS) where reverse transcription is initiated are shown. The repeated sequence R corresponds to an approximately 180 nucleotide-long sequence found in both termini of the 35S RNA. (B) Alignment of 35S RNA sequences from various CaMV strains. The sequences surrounding Cabb-S splice sites are shown. The GU and AG nucleotides of donor and acceptor sites, respectively, are in bold. Nucleotides differing from the Cabb-S sequence are underlined. Amino acids encoded by the regions containing the spliced sites are provided below the nucleotide sequences. The interaction between U1 snRNA and exon-intron junctions is shown below D4. (C) Alignment of the 35S RNA sequence from *Caulimovirus* species.

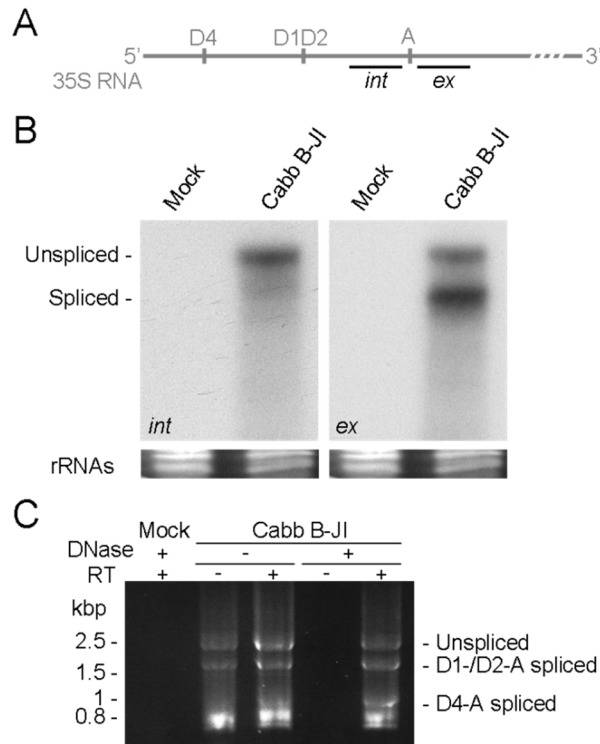


FIG 2: 35S RNA of CaMV Cabb B-JI strain is alternatively spliced. (A) Oligonucleotide probes targeting the intronic (*int*) or the exonic (*ex*) region surrounding the acceptor site were used to detect 35S RNA and its putative spliced isoforms. (B) Northern-blotting analysis of turnip total RNA extracted at 21 dpi. RNA was separated by electrophoresis on an agarose (0.8%)-urea (7 M) gel, transferred onto a nylon membrane and probed with ³²P-labeled *int* or *ex* probes. (C) RT-PCR on turnip total RNA extracted at 21 dpi. The 5' region of 35S RNA was amplified and the products separated on agarose gel.

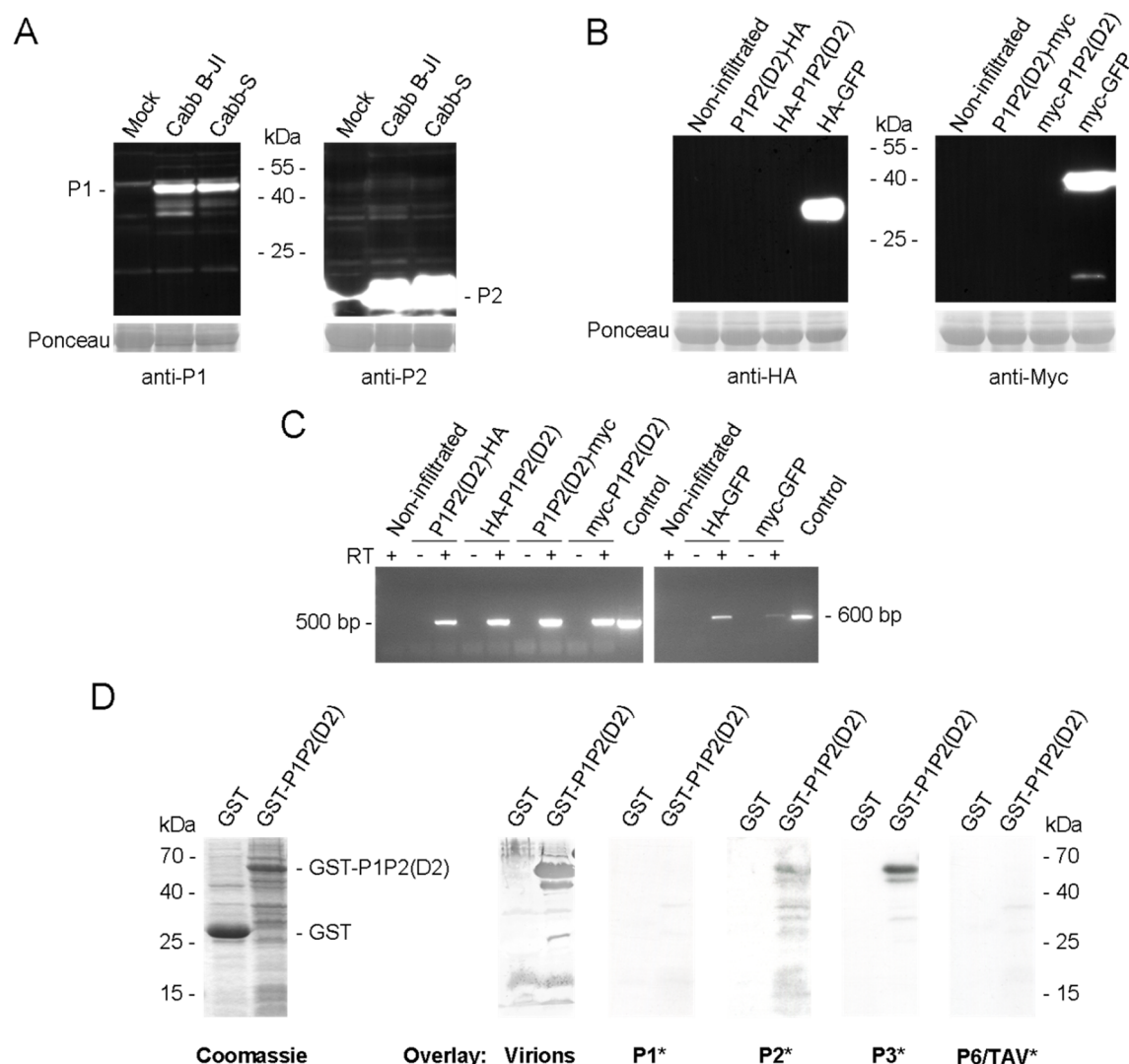


FIG 3: P1P2 fusion proteins are unstable in plants both in a viral context and in transient expression assays. (A) Immunodetection of P1P2 fusion proteins in CaMV Cabb B-JI- or Cabb-S-infected turnip plants. Total proteins were analyzed by SDS-PAGE and western blotting was performed with anti-P1 and anti-P2 polyclonal antibodies. (B) Immunodetection of HA- or myc-tagged P1P2(D2) protein upon *Agrobacterium tumefaciens*-mediated transient expression assay in *Nicotiana benthamiana* leaves. Total proteins were extracted from agroinfiltrated leaves and the presence of P1P2(D2) fusion protein was assessed by western assays with anti-HA or anti-myc antibodies. Leaves agroinfiltrated with HA- or myc-GFP expression vectors were used as positive controls. (C) Tagged-P1P2(D2) and GFP transcripts were assessed by RT-PCR on total RNA from infiltrated leaves. (D) Far-western assays on GST or GST-P1P2(D2) expressed in *E. coli*. Recombinant proteins were submitted to SDS-PAGE and transferred to PVDF membranes. 32 P-labeled recombinant P1, P2, P3 and TAV proteins or purified virions were used as overlays.

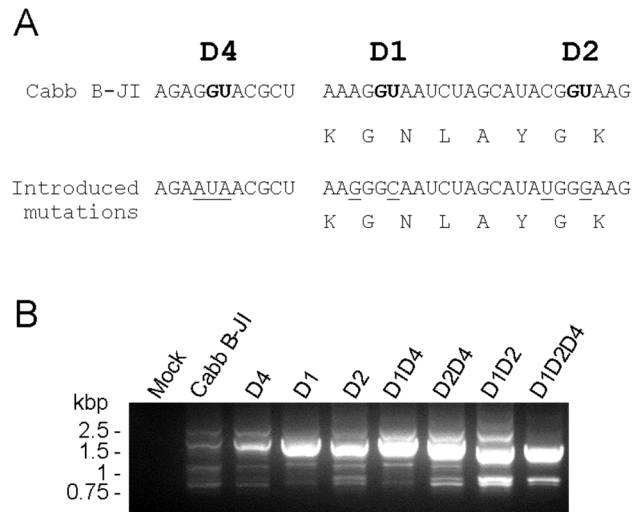


FIG 4: Mutations of splice donor sites are rescued by the use of a cryptic donor site. (A) Exon-intron junctions of splice donor sites of CaMV Cabb B-JI were mutated independently or in combination. The GU donor sites are in bold, and mutated nucleotides are underlined. Corresponding amino acids are provided below the nucleotide sequences surrounding D1 and D2 sites. (B) Analysis by RT-PCR of the splicing pattern of the donor site mutants. Total RNA from infected turnips was extracted at 21 dpi and the 5' region (about 2,500 nts) of 35S RNA was amplified. Amplicons were separated by agarose gel electrophoresis.

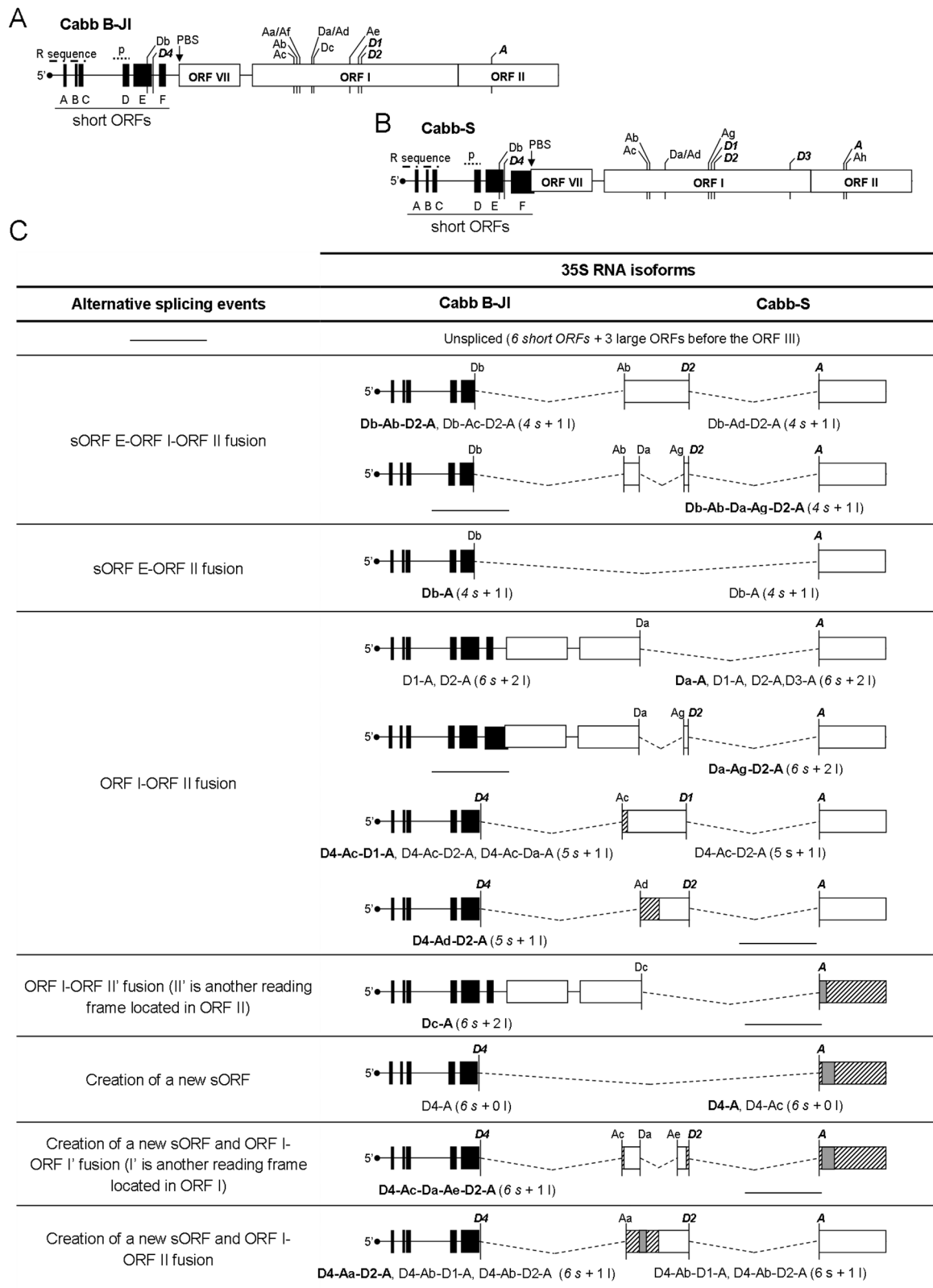


FIG 5: The 5' part of Cabb B-JI and Cabb-S 35S RNA contains numerous splice sites. The 5' region of Cabb B-JI (A) and Cabb-S (B) 35S RNA (about 2,500 nts) is schematized. Previously discovered sites are in bold italic. The putative encapsidation signal (p) and the primer binding site (PBS) where reverse transcription is initiated are shown. The repeated sequence R corresponds to an approximately 180 nucleotide-long sequence found in both termini of the 35S RNA. (C) Splicing events and corresponding isoforms in Cabb B-JI and Cabb-S strains. Splicing is represented by dotted lines. For each

event, isoforms are indicated and one or several representative isoforms (in bold) are schematized. Isoforms that are very similar to the ones schematized are only indicated. The number of short (s) and large (l) ORFs that are upstream of ORF III is given in brackets. Regions of ORFs I or II that are not coding anymore in some spliced isoforms are hatched. Coding sequences located within ORFs I or II but in a different reading frame are in grey.

	Cabb B-JI	Cabb-S
Splice donor sites (position)	D1 (872); D2 (887); D4 (7915); Da (653); Db (7890); Dc (658)	D1 (872); D2 (887); D3 (1247); D4 (7909); Da (653); Db (7884)
Splice acceptor sites (position)	A (1509); Aa (584); Ab (577); Ac (565); Ad (652); Ae (830); Af (587)	A (1508); Ab (577); Ac (565); Ad (652); Ag (864); Ah (1515)

85 and 87 clones from two independent experiments were sequenced for Cabb B-JI and Cabb-S strains, respectively. Sites Da to Dc and Aa to Af were named according to their order of discovery in Cabb B-JI strain.

TABLE 1 Splice donor and acceptor sites found in Cabb B-JI and Cabb-S strains by RT-PCR.

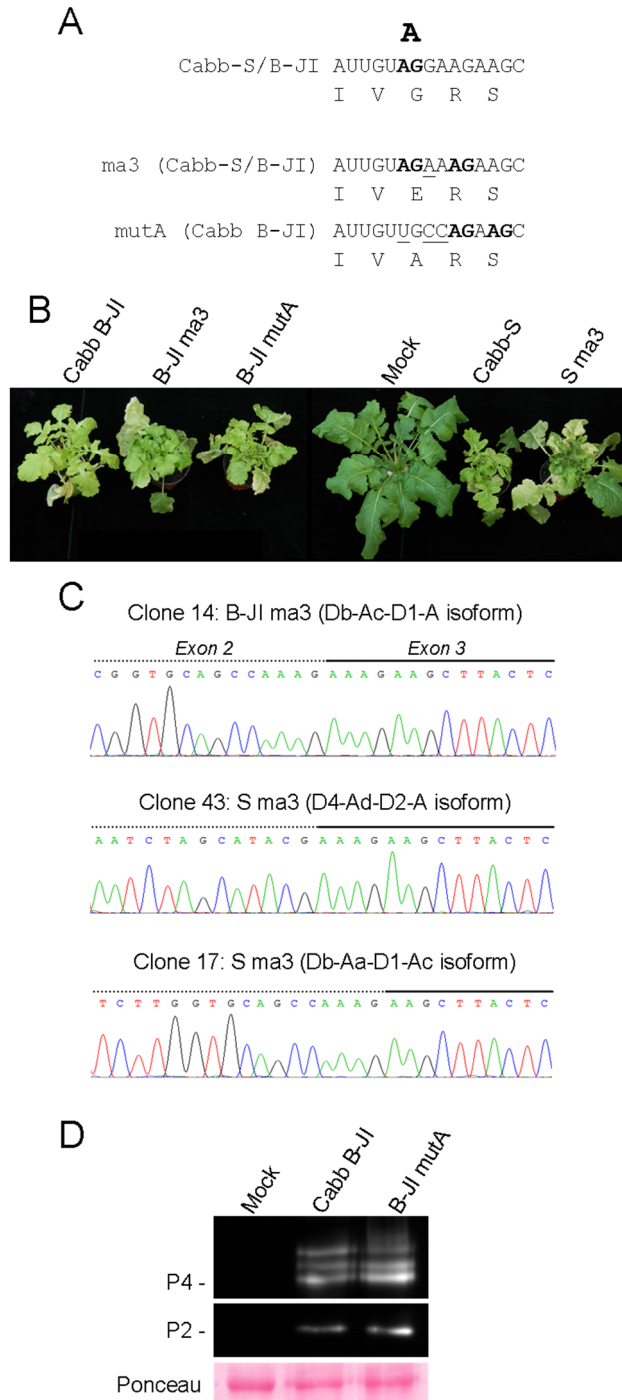


FIG 6: Mutations of splice acceptor site A lead to the use of downstream AG dinucleotides as acceptor sites. (A) Splice acceptor site A was mutated both in Cabb-S and Cabb B-JI strains. The AG dinucleotides used as acceptor sites in the wild type and mutant viruses are in bold; mutated nucleotides are underlined. Corresponding amino acids are provided below the nucleotide sequences. (B) Symptoms on turnip plants at 45 dpi. (C) Sequences at the second splice junction of three isoforms cloned from splice acceptor mutant-infected turnips (B-JI ma3 and S ma3) after RT-PCR amplification on total RNA. Exons 2 and 3 are represented by a dotted line and a solid line, respectively. (D) Immunodetection of P2 and capsid protein (CP) in CaMV Cabb B-JI- or Cabb B-JI mutA-infected turnips plants. Total proteins were analyzed by SDS-PAGE, and western blotting was performed with anti-P2 and anti-CP polyclonal antibodies.

Chapitre 2

Étude de l'export nucléaire des ARN viraux

INTRODUCTION

Dans le monde des procaryotes, la transcription et la traduction sont directement couplées. À l'inverse, chez les eucaryotes, ces deux processus sont séparés physiquement par l'enveloppe nucléaire : les ARNm sont transcrits et épissés dans le noyau, puis exportés vers le cytoplasme où ils seront traduits. Les données et modèles présentés dans cette introduction sont essentiellement issus de travaux réalisés sur la levure et l'Homme, les deux modèles qui sont les plus communément utilisés dans le champ de recherche que représente l'étude du transport nucléo-cytoplasmique.

1) Les pores nucléaires

La traversée des pores nucléaires est un passage obligé dans le transport des ARN, du noyau vers le cytoplasme. Ces complexes protéiques sont insérés dans la double-membrane qui constitue l'enveloppe nucléaire et ils permettent un passage sélectif et bidirectionnel des molécules dont la masse moléculaire est supérieure à 60 kDa.

a) Structure et composition

Le pore nucléaire est le plus grand des complexes protéiques eucaryotes : sa masse moléculaire est d'environ 60 MDa chez la levure et 90-120 MDa chez l'Homme (Raices et D'Angelo, 2012). Il peut être décrit comme la répétition de 8 unités formant chacune le rayon d'une roue de chariot. Chaque unité a une face nucléaire et une face cytoplasmique d'où émergent deux structures, une structure en forme de panier du côté nucléaire et une autre en filaments du côté cytoplasmique (Figure 1). Ces deux extrémités du pore nucléaire interviennent dans le transport nucléo-cytoplasmique, en interagissant avec des molécules en transit. L'intérieur du pore consiste en un canal central d'un diamètre d'environ 35 nm (Oeffinger et Zenklusen, 2012). La densité de pores nucléaires est assez variable d'un organisme à l'autre, et d'un type cellulaire à l'autre. Généralement, on trouve de 10 à 20 pores par μm^2 . Ce chiffre peut s'élever à 50 pour le noyau d'une cellule de *Nicotiana tabacum* BY-2 et 60 pour le noyau d'un ovocyte de xénope *Xenopus laevis* (Fiserova et al., 2009; Grossman et al., 2012). En dépit de sa taille gigantesque, un pore nucléaire n'est composé que d'une trentaine de protéines différentes, les nucléoporines – ou Nups –, qui sont présentes en 8, 16, 32 ou 48 exemplaires – donc un multiple de 8 (Grossman et al., 2012). Ces protéines sont relativement bien conservées et elles sont classées en différentes catégories, selon leurs caractéristiques biochimiques et/ou leur localisation au sein du pore : ce sont les Nups membranaires, les Nups d'échafaudage, et les Nups-FG (Figure 1).

Les Nups membranaires permettent l'ancrage du pore dans l'enveloppe nucléaire (Figure 1). Les Nups d'échafaudage sont localisées tout le long du pore, où elles constituent l'un des trois anneaux nucléaire, interne ou cytoplasmique. Certaines de ces Nups ont la particularité de partager des similitudes structurales avec les protéines du complexe « coatomer » formant certaines vésicules impliquées dans le trafic cellulaire. Ces similarités expliqueraient l'incurvation de la membrane nucléaire, autour du pore (Grossman et al., 2012). Les Nups-FG sont des protéines non-structurales qui tapissent l'intérieur du pore nucléaire et forment la barrière sélective qui permet de contrôler la nature de molécules en transit dans le canal central. Ces Nups-FG ont la particularité de posséder de 4 à 48 répétitions de différents motifs FG à 4 acides aminés où la phénylalanine (F) et la glycine (G) sont majoritaires. La répétition des motifs FG confère aux FG-Nups une structure intrinsèquement désordonnée (Grossman et al., 2012).

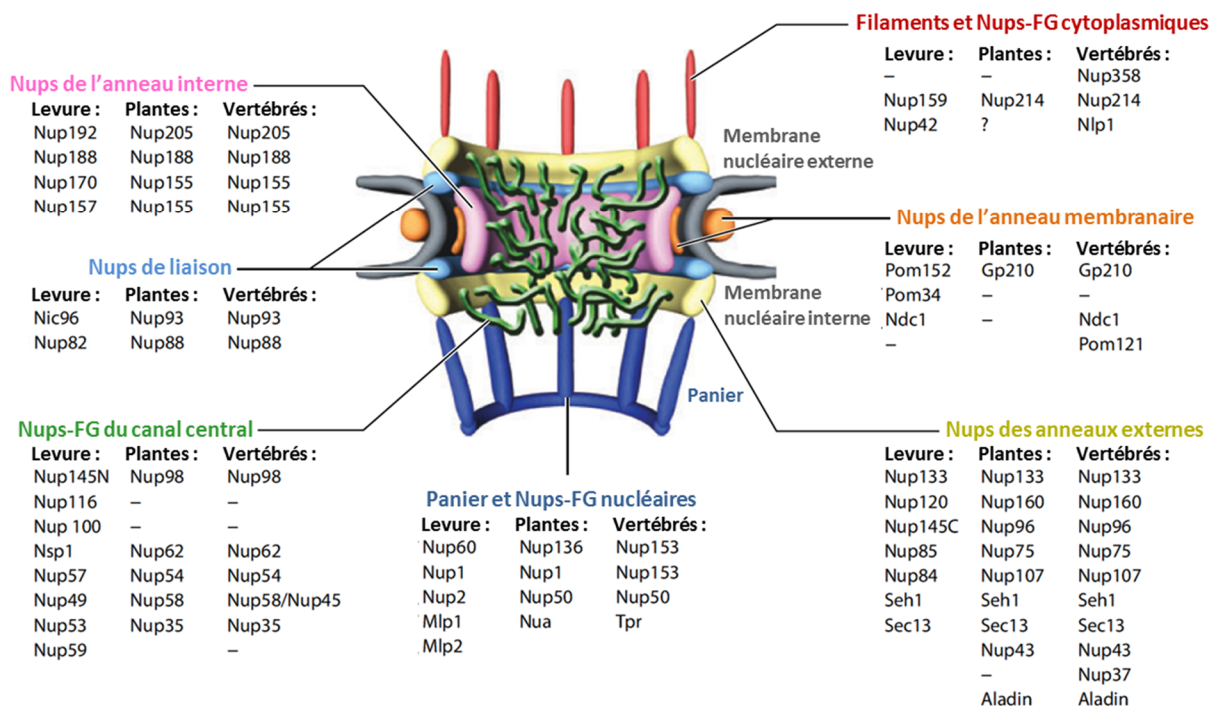


Figure 1 : Représentation schématisée d'un pore nucléaire, vu en coupe longitudinale. Les noms des principales nucléoporines (Nups) et de leurs homologues sont donnés pour la levure, les plantes et les vertébrés. D'après Grossman et al., 2012.

b) Mécanisme général du transport à travers les pores nucléaires

Les pores nucléaires n'entravent pas le passage passif des ions et des petites molécules dont la masse moléculaire est inférieure à 60 kDa. Les grosses molécules, à l'inverse, ne peuvent traverser les pores nucléaires que de façon active, grâce à l'assistance d'un ou plusieurs facteurs de transport. Le seuil des 60 kDa n'est, d'ailleurs, qu'une indication car certaines molécules dont la masse est légèrement inférieure se font, elles aussi, assister par des facteurs de transport (Grünwald et al., 2011). Ces facteurs sont des protéines dont le type varie selon l'espèce de molécule prise en charge ; ils seront décrits ultérieurement. Cependant, quel que soit le facteur de transport utilisé, le mécanisme général de transport à travers le pore nucléaire est relativement similaire et il implique les FG-Nups. Ces dernières forment un réseau dense dans le canal central du pore, ce qui empêche le transport des grosses molécules par diffusion. Seuls les facteurs de transport sont capables d'interagir avec les répétitions de motifs FG qui composent les FG-Nups (Figure 2). Ces interactions, bien que de faible affinité, sont très spécifiques. Ainsi, les complexes formés par les facteurs de transport et leurs charges – ou cargos – traversent la barrière sélective que forment les FG-Nups, suite à de nombreuses interactions avec ces dernières, tandis que les molécules n'étant pas prises en charge par un facteur de transport – et donc n'ayant pas à être transportées – ne pourront pas passer cet obstacle (Grossman et al., 2012; Strambio-De-Castillia et al., 2010). La directionnalité dans le transport des cargos – n'est, cependant, pas imprimée par les FG-Nups. Elle dépend de la nature des molécules transportées et elle sera abordée dans la suite de ce chapitre.

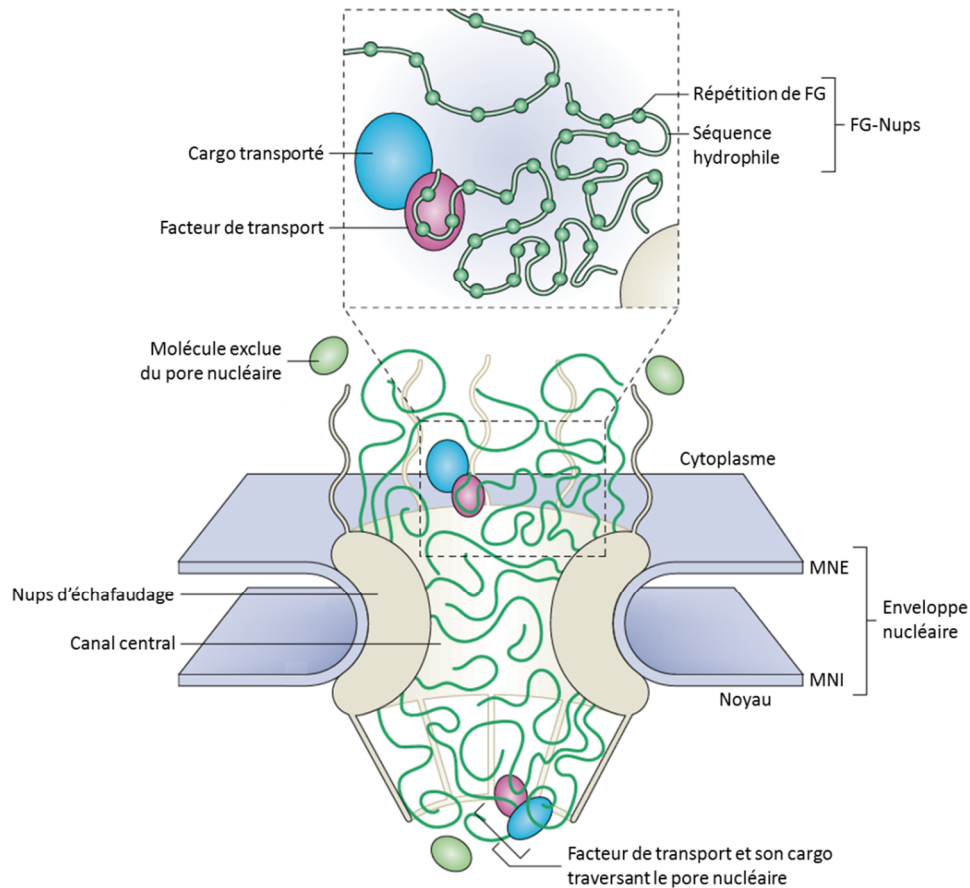


Figure 2 : Passage des facteurs de transport nucléaires associés à leurs cargos, à travers le pore nucléaire. Le passage des grosses macromolécules, à travers le pore nucléaire, est rendu possible grâce à leur association avec un facteur de transport. Ce dernier est capable de réaliser toute une suite d'interactions avec les répétitions FG des FG-Nups, les Nups qui forment une barrière sélective et tapissent le canal central du pore. MNE : membrane nucléaire externe. MNI : membrane nucléaire interne. D'après Strambio-De-Castillia et al., 2010.

c) Assemblage des pores nucléaires

Les pores nucléaires peuvent s'assembler à deux moments du cycle cellulaire : lors de l'interphase et après une mitose ouverte – c'est-à-dire si la mitose s'accompagne d'une désagrégation de l'enveloppe nucléaire –, lors du réassemblage de l'enveloppe nucléaire (Fernandez-Martinez et Rout, 2009).

i. Assemblage de novo au cours de l'interphase

Au cours de l'interphase, le nombre de pores nucléaires présents dans l'enveloppe nucléaire augmente, en particulier chez les cellules en prolifération. De nouveaux pores sont assemblés *de novo*, à travers l'enveloppe nucléaire. Il semble que les Nups membranaires soient les premières à s'assembler. Elles sont suivies par des Nups d'échafaudage qui sont délivrées de part et d'autre de l'enveloppe nucléaire. Ce processus d'assemblage des pores nucléaires ne ferait intervenir que des Nups nouvellement synthétisées (D'Angelo et al., 2006; Grossman et al., 2012).

ii. Assemblage post-mitotique

La mitose, selon qu'elle s'accompagne d'une désagrégation ou d'un maintien de l'enveloppe nucléaire, est qualifiée d'ouverte ou fermée. On retrouve souvent la mitose ouverte chez les vertébrés

et les plantes, tandis que la mitose fermée est plutôt l'apanage des eucaryotes qualifiés d'inférieurs. L'entrée en mitose ouverte provoque un désassemblage des pores nucléaires, notamment suite à la phosphorylation de certains Nups (Fernandez-Martinez et Rout, 2009). À la fin de la mitose, alors que les chromosomes se décondensent, un complexe nommé pré-pore et composé des Nups d'échafaudage Nup107 et 160, est recruté à proximité de la chromatine grâce à la protéine Elys/Mel28 qui interagit, à la fois, avec la surface de la chromatine et le pré-pore. Ce recrutement s'effectue grâce à la GTPase Ran, dans sa forme liée au GTP (Ran-GTP), qui libère le pré-pore alors complexé à l'importine β en interagissant avec cette dernière. Par la suite, d'autres Nups, dont les Nups membranaires, s'associent au pré-pore pour s'assembler en un pore nucléaire alors que l'enveloppe nucléaire se reforme progressivement. À l'inverse de l'assemblage à l'interphase, l'assemblage post-mitotique des pores nucléaires se fait à partir de Nups qui composaient les pores déconstruits lors de l'entrée en mitose. Cette voie d'assemblage est, par ailleurs, plus rapide que l'assemblage *de novo* (Grossman et al., 2012).

2) Export nucléaire des ARN autres que les ARNm, et des protéines

Les mécanismes d'export nucléaire des ARN autres que les ARNm, ainsi que celui des protéines, ont en commun de reposer sur l'utilisation de facteurs de transport nommés karyophérines, membres de la super-famille des importines β . Ces facteurs sont aussi nommés exportines, par opposition aux importines *sensu stricto* qui sont impliquées dans l'import de cargos vers le noyau. L'association entre les karyophérines et les molécules qu'elles prennent en charge est modulée par la GTPase Ran que l'on peut retrouver sous deux formes, associée au GTP (Ran-GTP) ou au GDP (Ran-GDP). La présence d'un facteur d'échange de nucléotide dépendant de Ran, nommé RCC1 – pour « *regulator of chromosome condensation 1* » – et associé à la chromatine, contribue à maintenir un taux élevé de Ran-GTP dans le noyau (Figure 3). Dans le cytoplasme, et particulièrement au niveau des filaments cytoplasmiques des pores nucléaires, se trouve une protéine activant l'activité GTPase de Ran, RanGAP – pour « *Ran GTPase activating protein* ». Ainsi, la forme de Ran qui est majoritaire dans le cytoplasme est Ran-GDP. Par conséquent, un gradient de concentration en Ran-GTP est maintenu entre le noyau et le cytoplasme, et c'est ce gradient qui confère la directionnalité du transport nucléo-cytoplasmique dépendant des karyophérines (Grossman et al., 2012; Hutten et Kehlenbach, 2007). Ainsi, une exportine ne peut se lier à la molécule qu'elle transporte que si elle est associée à Ran-GTP. C'est l'opposé du mécanisme d'import puisque la liaison de Ran-GTP au complexe d'import – composé de la molécule à importer, d'un adaptateur et d'une importine – provoque la dissociation de ce dernier (Figure 3). Le complexe de transport, composé de la molécule à exporter, de la karyophérine et de Ran-GTP peut alors traverser le pore nucléaire, par un mécanisme encore incomplètement élucidé mais dépendant d'interactions entre l'exportine et les FG-Nups (Grossman et al., 2012). Au sortir du pore nucléaire, une fois dans le cytoplasme, l'hydrolyse du GTP par Ran conduit à la dissociation du complexe de transport (Köhler et Hurt, 2007).

a) Export nucléaire des ARNt

L'export des ARNt est principalement médié par l'exportine-t. Suite à leur transcription par l'ARN polymérase III (Pol III), les pré-ARNt subissent plusieurs étapes de maturation. Leurs deux extrémités sont rognées : la RNase P coupe l'extrémité 5' et plusieurs nucléases, dont la RNase Z, coupent l'extrémité 3'. Une séquence CCA, nécessaire à l'aminocacylation des ARNt, est ensuite ajoutée à l'extrémité 3' par une nucléotidyl transférase. Enfin, les ARNt sont épissés – voir la brève description de cet épissage enzymatique dans le premier chapitre –, lorsqu'ils possèdent un intron, et subissent une grande variété de modifications – méthylations, thiolations, etc... – sur certaines de leurs bases (El Yacoubi et al., 2012; Hopper, 2013). Une fois ces étapes de modifications post-transcriptionnelles effectuées, l'extrémité 3', qui est sortante à la base de l'ARNt, est reconnue par l'exportine-t liée à Ran-GTP (S. J. Lee et al., 2011). Le complexe exportine-t/Ran-GTP/ARNt est ensuite transporté à

travers le pore nucléaire, vers le cytoplasme. Il est important de noter que l'exportine-t n'est pas capable de discriminer entre les ARNt épissés et non-épissés. Ainsi, il arrive que certains ARNt soient exportés vers le cytoplasme, en dépit de la présence d'un intron. L'export nucléaire des ARNt n'est pas uniquement dirigé par l'exportine-t. Certaines protéines, en particulier l'exportine-5 qui est le facteur de transport des miARN, sont impliquées dans ce processus. Le rôle de l'exportine-5 dans l'export nucléaire des ARNt est surtout notable chez la levure *Saccharomyces cerevisiae* (Hopper, 2013).

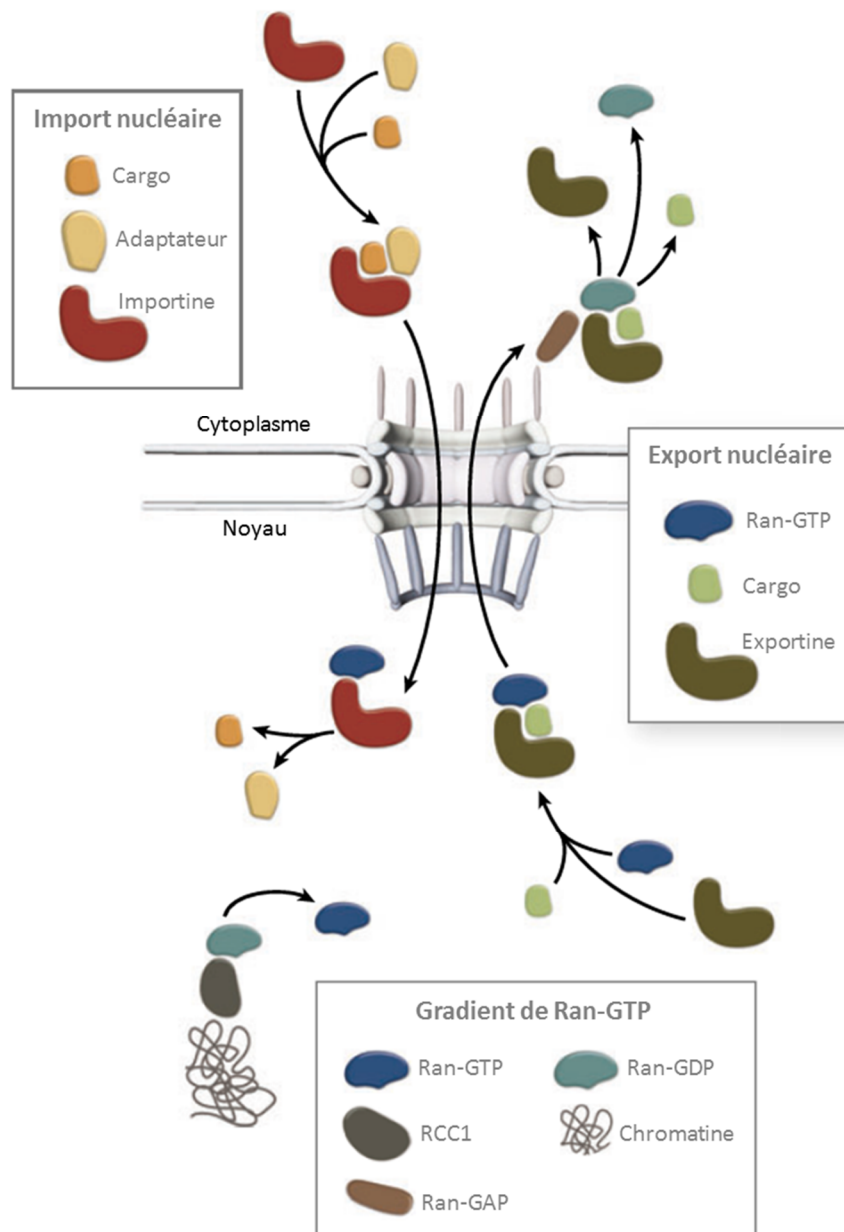


Figure 3 : Transport nucléocytoplasmique médié par les karyophérines. L'import nucléaire d'un cargo se fait par interaction de celui-ci avec une karyophérine de type importine, souvent par le biais d'un adaptateur. Dans le noyau, le complexe est dissocié lorsqu'une protéine Ran liée au GTP s'associe à l'importine. Inversement, Ran-GTP promeut la formation d'un complexe d'export entre une karyophérine de type exportine et son cargo. Dans le cytoplasme, l'hydrolyse du GTP par Ran aboutit à la dissociation du complexe d'export. La directionnalité du transport nucléocytoplasmique médié par les karyophérine est dépendante d'un gradient de Ran-GTP à travers l'enveloppe nucléaire. Dans le noyau, la présence de la protéine RCC1 associée à la chromatine transforme la Ran-GDP en Ran-GTP. Dans le cytoplasme, la Ran-GAP stimule l'activité GTPase de Ran, et donc favorise la présence de Ran-GDP. D'après Grossman et al., 2012.

b) Export nucléaire des miARN

Les miARN peuvent être produits soit à partir d'introns – on parle alors de mirtrons –, soit à partir de transcrits propres nommés pri-miARN et synthétisées par les Pols II et III. Les pri-miARN sont constituées d'une structure en tige-boucle entourée de séquences adjacentes. La première étape de maturation d'un pri-miARN consiste à éliminer ces séquences adjacentes, par une RNase nommée Drosha, pour générer une tige-boucle d'environ 65 pb, nommée pré-miARN. Cet intermédiaire, produit à la fin de la première étape de maturation, est exporté vers le cytoplasme grâce à l'exportine-5 liée à Ran-GTP, qui reconnaît la base du pré-miARN, et en particulier son extrémité 3' débordante, généralement longue de deux nucléotides (Köhler et Hurt, 2007). Après le transport vers le cytoplasme et la libération du complexe pré-miARN/exportine-5/Ran-GDP, le pré-miARN est clivé, par une autre RNase nommée Dicer, en un fragment d'environ 22 pb à l'appariement imparfait. La dernière étape de maturation consiste en la dégradation de l'un des deux brins, puis en l'insertion du brin restant au sein d'un complexe multiprotéique nommé RISC – pour « *RNA-induced silencing complex* » –, qui va diriger la reconnaissance des cibles du miARN.

c) Export nucléaire des snARN

L'export nucléaire des snARN a été partiellement décrit dans le premier chapitre de cette thèse, lorsque la biosynthèse des snRNP a été abordée. Les snARN, hormis le snARN U6, sont transcrits par la Pol II. Ils possèdent une coiffe 7-méthylguanosine en 5', mais n'ont pas de queue poly-A en 3'. Leur export nucléaire est dépendant de l'exportine-1, plus communément appelée CRM1. Alors que les exportines-t et -5 reconnaissent des motifs structuraux sur les ARNt et les miARN, CRM1 reconnaît une séquence d'acides aminés qui est nommée NES – pour « *nuclear export signal* » –, et qui est riche en leucine et autres acides aminés à chaîne latérale hydrophobe (Köhler et Hurt, 2007). L'organisation générale d'un NES est la suivante : $\Phi x_{2-3} \Phi x_{2-3} \Phi x \Phi$, où Φ est généralement une leucine mais peut aussi être une isoleucine, une valine, une phénylalanine ou une méthionine, et où x est n'importe quel acide aminé (Hutten et Kehlenbach, 2007). La coiffe des snARN permet le recrutement des deux sous-unités protéiques CBP20 et CBP80, de la protéine ARS2 et de l'adaptateur d'export PHAX sous sa forme phosphorylée. C'est au niveau de PHAX que l'on retrouve un NES qui va permettre la formation d'un complexe d'export avec CRM1 et Ran-GTP. Dans le cytoplasme, ce complexe est dissocié suite à l'hydrolyse du GTP de Ran et à la déphosphorylation de PHAX. Le snARN est libéré puis assemblé à des protéines pour former une snRNP (Köhler et Hurt, 2007).

d) Export nucléaire des ARNr

Les quatre espèces d'ARNr qui composent les ribosomes eucaryotes sont les ARN 5,8S, 18S et 28S, transcrits par l'ARN polymérase I (Pol I) en un même pré-ARNr 45S, et l'ARNr 5S, transcrit par la Pol III. Ces quatre ARNr ne sont pas exportés tels quels vers le cytoplasme, mais une fois associés à de nombreuses protéines ribosomales, en pré-sous-unités ribosomales 40S et 60S. Ces pré-sous-unités font partie des plus grosses entités à traverser les pores nucléaires. Le facteur de transport principal des pré-sous-unités ribosomales est l'exportine CRM1. Son mécanisme d'association à la pré-sous-unité 40S est peu clair. En ce qui concerne la pré-sous-unité 60S, CRM1 est recrutée par une protéine adaptatrice nommée Nmd3 et possédant un NES. D'autres facteurs de transport peuvent être utilisés accessoirement par les deux pré-sous-unités pour quitter le noyau, en particulier le couple Mex67-Mtr2 – équivalent chez *Saccharomyces cerevisiae* du couple TAP-p15 –, impliqué principalement dans l'export nucléaire des ARNm (Köhler et Hurt, 2007; Thomson et al., 2013).

e) Export nucléaire des protéines

L'exportine CRM1 est le facteur de transport majeur des protéines. La liaison entre CRM1 et une protéine avec un NES est réalisée coopérativement avec Ran-GTP, mais elle semble aussi être

stimulée par l'action d'une protéine nommée RanBP3 – pour « *Ran-binding protein 3* » – qui a pour fonctions de favoriser le maintien d'une forte concentration de Ran-GTP à proximité de CRM1, et d'augmenter l'affinité de cette dernière pour les protéines à NES (Hutten et Kehlenbach, 2007). La dissociation du complexe d'export, à la suite de la traversée du pore nucléaire, se fait par hydrolyse du GTP lié à Ran. Cette hydrolyse est stimulée par la protéine RanGAP ainsi que par la protéine RanBP1, dont le rôle est de rendre Ran sensible à l'action de RanGAP. Il est à noter qu'une fraction de RanGAP est associée aux filaments cytoplasmiques des pores nucléaires *via* une interaction indirecte avec la Nup358, ce qui favorise une dissociation du complexe d'export dès sa sortie du canal central du pore nucléaire puisque Nup358 – également nommée RanBP2 – possède également des domaines de liaison à Ran (Hutten et Kehlenbach, 2007).

3) Vue générale sur l'export nucléaire des ARNm

a) Introduction théorique

C'est sous la forme de ribonucléoprotéines messagères – ou mRNP – que les ARNm sont exportés vers le cytoplasme, où ils seront traduits (Köhler et Hurt, 2007). Tout au long de la biogénèse des ARNm, de nombreuses protéines – facteurs de transcription, d'épissage, etc... – s'associent aux transcrits, de façon plus ou moins transitoire, par recrutement direct ou assisté par la Pol II (Kelly et Corbett, 2009; Oeffinger et Zenklusen, 2012). Parmi elles, sont retrouvés des facteurs impliqués directement ou indirectement dans l'export nucléaire. L'export nucléaire des ARNm est donc intimement lié aux différentes étapes de biogénèse et de maturation des transcrits, en ce sens que toutes ces étapes peuvent participer, en des proportions qui varient selon le type d'organisme, au recrutement du facteur de transport des ARNm.

Le mode d'export nucléaire des ARNm est bien conservé entre la levure et l'Homme. La différence majeure entre l'export nucléaire des autres ARN et des protéines et celui des ARNm est que ce dernier n'est dirigé ni par des karyophérines, ni par le gradient de Ran-GTP présent autour de l'enveloppe nucléaire (Köhler et Hurt, 2007).

b) Facteur de transport des ARNm

i. Les hétérodimères TAP-p15 et Mex67-Mtr2

L'export nucléaire des ARNm est principalement gouverné par un hétérodimère nommé TAP-p15 ou NXF1-NXT1 chez les métazoaires, et Mex67-Mtr2 chez la levure. TAP – pour « *Tip-associated protein* » – et son homologue Mex67 possèdent un motif RRM de reconnaissance de l'ARN à leur extrémité amino-terminale, et un motif de type NTF2-like qui permet l'interaction avec la protéine p15 ou son analogue structural Mtr2 (Valkov et al., 2012). Si l'interaction directe entre TAP-p15 et l'ARNm à transporter est possible, le recrutement de l'hétérodimère se fait généralement avec le concours de facteurs d'export présents sur le transcrit. Enfin, tout comme les différentes exportines qui ont été abordées précédemment, TAP-p15 peut interagir avec les répétitions de motifs FG des FG-Nups ce qui permet aux mRNP sur lesquelles elles sont fixées, de traverser les pores nucléaires (Köhler et Hurt, 2007).

ii. L'exportine CRM1

Quelques ARNm ont été décrits comme étant dépendants de l'exportine CRM1 pour leur export nucléaire. Les transcrits des gènes codant pour l'interféron IFN- α 1 et pour la cycline D1 sont exportés par cette voie (Kimura et al., 2004). CRM1 se fixe aux ARNm à transporter grâce à des protéines adaptatrices, telles HuR et NXF3 (Culjkovic-Kraljacic et Borden, 2013; Hutten et Kehlenbach, 2007). Enfin, certains virus utilisent la voie CRM1 pour exporter certains de leurs transcrits. Ils sont décrits par la suite.

c) Couplage entre la production des ARNm et leur export nucléaire

Production des ARNm et export nucléaire sont deux processus couplés. En effet, l'altération d'une ou plusieurs étapes de la biogénèse des transcrits aboutit souvent à leur rétention dans le noyau (Oeffinger et Zenklusen, 2012). Les étapes où le couplage entre la production des ARNm et leur export est le plus intense, varient selon les organismes. Ainsi, chez la levure, le recrutement de Mex67-Mtr2 semble fait principalement au cours de l'élongation et de la terminaison de la transcription tandis que chez l'humain, TAP-p15 est recruté préférentiellement suite à l'ajout de la coiffe, ainsi que lors de l'épissage (Kelly et Corbett, 2009).

i. Le complexe TREX

La relation qui existe entre transcription et export nucléaire est matérialisée par le complexe multiprotéique TREX – pour « *transcription export* » – de la levure. TREX participe au recrutement et à la liaison de facteurs d'export, sur les ARNm en cours de formation. Chez la levure, TREX se compose d'un sous-complexe nommé THO – lui-même composé des sous-unités Hpr1, Mft1, Thp2 et Tho2 –, ainsi que des facteurs d'export Sub2 et Yra1 (Kelly et Corbett, 2009). THO est associé à la Pol II en cours d'élongation, via des interactions avec la queue CTD. À la fin de l'élongation, et de concert avec la sous unité Pcf11 du facteur de clivage et de polyadénylation CF1A, THO participe au recrutement d'Yra1 (Katahira, 2012; Meinel et al., 2013). De plus, ses sous-unités Sub2 et Hpr1 peuvent interagir avec l'hétérodimère Mex67-Mtr2. Yra1 est un adaptateur d'export à proprement parler puisque son rôle est de faire le lien entre l'ARNm et Mex67-Mtr2 ; Yra1 est transféré de Pcf11 vers Sub2, une ARN hélicase, puis est déposé sur l'ARNm grâce à cette dernière (Johnson et al., 2009). Yra1 stimulerait l'affinité de Mex67 pour l'ARN et permettrait donc la fixation de Mex67-Mtr2 au transcrit, avant d'être dissocié du transcrit peu de temps avant son passage dans le pore nucléaire, probablement à la suite d'une ubiquitinylation par l'E3 ubiquitine-ligase Tom1 (Katahira, 2012; Valkov et al., 2012).

En plus de son rôle dans l'export nucléaire, le complexe THO/TREX de la levure stimule l'élongation de la transcription et empêche la formation d'hybrides entre le transcrit naissant et le brin d'ADN non-codant – surnommés les *R-loops* –, qui peuvent être une source d'hyper-recombinaisons (Katahira, 2012).

Chez les métazoaires, TREX diffère du complexe de levure, par son mode de recrutement et sa composition. Le complexe THO des métazoaires n'est pas tout à fait identique à celui de la levure puisqu'il se compose de THOC1 et THOC2, les homologues respectifs de Hpr1 et Tho2, ainsi que de THOC5, THOC6 et THOC7, dont on ne retrouve pas d'homologues chez la levure. Les homologues de Sub2 sont UAP56 et URH49, celui d'Yra1 est ALY/REF (Palazzo et Akef, 2012). Le mode de recrutement des composants du complexe TREX, chez les métazoaires, diffère de celui décrit chez la levure puisqu'il est lié, principalement, à la présence du complexe de liaison à la coiffe – en particulier sa grande sous-unité CBP80 –, et à la présence d'événements d'épissage (Katahira, 2012; Masuda et al., 2005). En effet, il a été montré que la sous-unité ALY est capable d'interagir directement avec la CBP80. De plus, ALY peut être également recrutée par UAP56 qui est, par ailleurs, un facteur d'épissage impliqué dans les premières étapes de l'assemblage du spliceosome (Köhler et Hurt, 2007; Luo et al., 2001). UAP56 est ensuite relarguée, ce qui permet à ALY de recruter TAP-p15 sur l'ARNm, puis d'être libérée à son tour (Palazzo et Akef, 2012).

Un second complexe protéique, nommé TREX-2, participe au couplage entre transcription et export nucléaire. TREX-2 se compose, chez la levure, des sous-unités Sac3, Thp1, Cdc31 et Sus1 ; il est bien conservé chez l'humain et *Arabidopsis thaliana* (García-Oliver et al., 2012). TREX-2 se localise au niveau du panier nucléaire des pores, où il interagit avec la Nup1. Chez la levure, il semble aussi que TREX-2 interagisse physiquement avec un complexe activateur de la transcription nommé SAGA – pour « *Spt-Ada-Gcn5-acetyltransferase* ». Composé de 21 protéines, SAGA est impliqué dans le remodelage de la chromatine, en particulier grâce à une activité histone acétyltransférase. Ainsi, l'interaction TREX-2/SAGA permet de recruter directement les gènes activés par SAGA, au

niveau du pore nucléaire. De plus, la sous-unité Sac3 de TREX-2 peut lier Mex67-Mtr2, ce qui contribue à coupler physiquement la production des ARNm et leur export nucléaire (García-Oliver et al., 2012; Valkov et al., 2012). Cette connexion physique entre certains gènes transcrits et les pores nucléaires confirme, en partie, l'hypothèse dite du *gene gating* selon laquelle tout gène est lié physiquement à un pore nucléaire (Blobel, 1985). Loin d'être la règle, cette hypothèse est au moins partiellement valide pour les gènes dont l'expression est contrôlée par le complexe SAGA.

ii. Épissage et export nucléaire

Le recrutement de TREX et des facteurs de transport, chez les métazoaires, semble dépendre beaucoup plus de l'épissage qu'il ne l'est chez la levure. Par exemple, UAP56 sert, à la fois, de facteur d'épissage et de facteur d'export (Palazzo et Akef, 2012). De même UAP56 et ALY sont toutes deux associées aux EJC, ces complexes qui sont déposés sur les ARNm, suite à l'épissage (Chang et al., 2007). Par ailleurs, il a été montré que certaines protéines SRs – dont les rôles majeurs dans l'épissage ont été décrits dans le premier chapitre – peuvent participer directement au recrutement de TAP-p15 (Huang et al., 2003). Enfin, il existe un système de surveillance qui est basé sur la rétention d'ARNm partiellement ou encore non-épissés, aux abords du pore nucléaire, puis leur dégradation par l'exosome nucléaire. Même si son mécanisme n'est pas connu, il est clair que les Nups Mlp1 et Mlp2 – Tpr chez l'humain –, localisées au niveau du panier nucléaire, sont impliquées (Coyle et al., 2011; Palazzo et Akef, 2012).

d) Séquences nucléotidiques impliquées dans l'export nucléaire

Certaines séquences nucléotidiques peuvent être impliquées, plus ou moins directement, dans l'export nucléaire des transcrits sur lesquels elles se trouvent, en participant au recrutement de facteurs d'export.

i. Régions codantes des séquences signal et des séquences ciblant les mitochondries

Les régions codantes des séquences signal et des séquences ciblant les mitochondries sont capables de stimuler l'export nucléaire, notamment du fait de leur composition nucléotidique. Il a été montré que ces deux types de séquences ont fortement tendance à être localisés au niveau du premier exon, et qu'elles codent souvent pour des acides aminés hydrophobes. Par conséquent, elles sont particulièrement riches en uridines car la plupart des codons spécifiant les acides aminés hydrophobes possèdent une uridine en position 2. À l'inverse, ces séquences sont particulièrement pauvres en adénines. S'il a été montré que la composition nucléotidique de ces deux types de séquences stimule l'export nucléaire des ARNm *via* l'hétérodimère TAP-p15, le mécanisme exact qui sous-tend cette stimulation n'est pas connu (Palazzo et al., 2013).

ii. Élément constitutif de transport (CTE) des Rétrovirus

Certains rétrovirus comme le virus Mason-Pfizer du singe (MPMV) utilisent un élément constitutif de transport – ou CTE, pour « *constitutive transport element* » – pour exporter leurs ARN non-épissés. Le CTE adopte une structure en tige-boucle qui permet une interaction directe avec TAP (Shida, 2012). C'est d'ailleurs l'étude du MPMV qui a conduit à la première description de TAP comme étant un facteur d'export nucléaire d'ARNm (Grüter et al., 1998).

iii. Cas des ARNm sans introns

Chez les mammifères, environ 5% des pré-ARNm ne possèdent pas d'introns et donc ne bénéficient pas du recrutement de facteurs de transport assisté par l'épissage. (Shabalina et al., 2010). Ces transcrits, comme ceux codant pour l'histone H2A, c-Jun ainsi que les interférons IFN- α 1 et IFN- β 1, sont pourtant exportés vers le cytoplasme. Dans le cas du messager de l'histone H2A, un élément de 22 nucléotides permet la fixation de deux protéines SR qui, à leur tour, recrutent le facteur d'export

TAP (Huang et al., 2003; Huang et Steitz, 2001). Le transcrit de l'interféron $\alpha 1$ possède, dans la partie centrale de sa séquence codante, une région structurée sur laquelle se fixe une protéine adaptatrice – qui pourrait être le facteur d'épissage RPS1 – capable de faire le lien avec l'exportine CRM1 (Kimura et al., 2004, 2010).

e) Mouvement des mRNP dans le noyau et au sein des pores nucléaires

i. Mouvement dans le noyau

Les mRNP, une fois libérées du site de transcription, se déplacent par diffusion et de manière aléatoire dans le noyau. Leurs mouvements peuvent être discontinus, selon l'encombrement moléculaire des zones dans lesquelles elles se déplacent. La taille du noyau, variable selon les organismes, influe sur la durée entre la libération d'une mRNP et le moment de sa rencontre avec un pore nucléaire – généralement de l'ordre de quelques minutes. La position du gène au sein du noyau peut moduler la durée de la phase de diffusion nucléaire : les transcrits produits à la périphérie nucléaire sont susceptibles d'atteindre rapidement un pore nucléaire (Oeffinger et Zenklusen, 2012).

ii. Traversée du pore nucléaire

La traversée du pore nucléaire par une mRNP est un évènement assez rapide puisqu'estimé à quelques centaines voire quelques dizaines de millisecondes (ms). Au sein du pore nucléaire, l'hétérodimère TAP-p15 lié à la mRNP interagit avec des FG-Nups et se déplace de façon aléatoire. Au sortir du pore, la mRNP est remodelée : le dimère TAP-p15 ainsi que d'autres protéines fixées à l'ARNm sont détachés. C'est ce remodelage, dirigé par les protéines Dbp5 et Gle1, qui imprime la directionnalité du transport des mRNP au sein du pore nucléaire (Figure 4) (Oeffinger et Zenklusen, 2012).

Dbp5 est une hélicase à ATP ARN-dépendante qui est à la fois capable de se fixer sur l'ARNm, lors de la transcription, et d'interagir avec des Nups – dont la Nup159 – situées au niveau des filaments cytoplasmiques. Là, elle interagit également avec Gle1. Cette interaction, stabilisée par l'inositol hexakiphosphate, stimule très fortement l'activité ATPase de Dbp5. Ceci entraîne un remodelage dans la composition de la mRNP qui aboutit au détachement de TAP-p15. Ainsi, l'ARNm libéré de ses facteurs de transport ne pourra pas repénétrer dans le pore nucléaire (Oeffinger et Zenklusen, 2012; Stewart, 2007; Valkov et al., 2012).

4) Export nucléaire des ARNm chez les plantes

Les connaissances sur l'export nucléaire des ARNm chez les plantes sont très sommaires. À ce jour, il n'a pas été découvert d'homologue pour TAP et p15, chez les plantes. Cependant, des homologues de certaines sous-unités de TREX ainsi que d'UAP56 et ALY ont été décrits (Gaouar et Germain, 2013; Kammel et al., 2013; Meier, 2012). Il en est de même pour Dbp5 et certaines sous-unités de TREX-2 (Meier, 2012). MOS11, un homologue de la protéine humaine CIP29 qui capable d'interagir avec ALY et UAP56, est impliqué dans l'export des ARNm, chez *Arabidopsis* (Germain et al., 2010). L'homologie de fonction n'a pas été prouvée pour toutes les protéines listées ci-dessus, même si des défauts dans l'export nucléaire des ARNm ont été constatés chez des mutants de plusieurs Nups – telle que NUA, l'homologue de Mlp1 –, de LOS4 – l'orthologue de Dbp5 –, ainsi que des complexes TREX et TREX-2 (Meier, 2012).

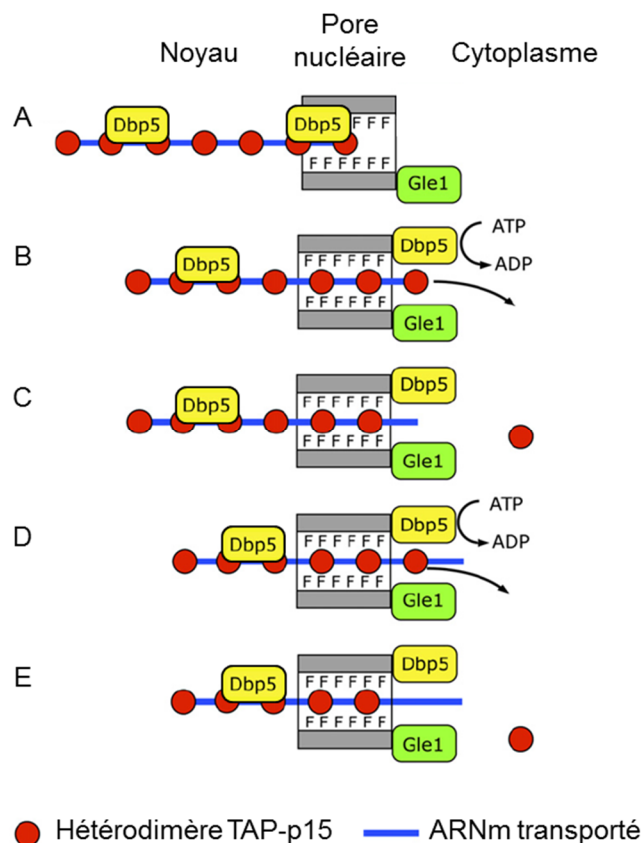


Figure 4 : Remodelage d'une mRNP, à la sortie d'un pore nucléaire. Alors que la directionnalité du transport nucléo-cytoplasmique est imprimée par un gradient de Ran-GTP à travers la membrane plasmique, celle de l'export nucléaire des mRNP dépend de la dissociation progressive des facteurs d'export par l'hélicase ATP-dépendante Dbp5, dans le cytoplasme. L'activité de Dbp5 est fortement stimulée par son interaction avec la protéine Gle1. Modifié d'après (Stewart, 2007).

5) Export nucléaire des ARNm viraux

Comme c'est le cas pour de nombreux autres processus cellulaires, les virus ont développé diverses stratégies pour exploiter les machineries nucléaires afin d'exporter leurs ARNm vers le cytoplasme. Alors que certains, comme les herpèsvirus ou les rétrovirus simples, utilisent la voie « usuelle » médiée par TAP-p15 pour exporter tout ou partie de leurs transcrits, d'autres utilisent à cette fin une voie rarement réservée à l'export nucléaire d'ARNm, la voie dépendante de CRM1.

a) Utilisation de la voie TAP-p15 par les virus

i. Cas de l'herpès simplex virus de type 1 (HSV-1) et de l'herpèsvirus associé au sarcome de Kaposi (KSHV)

L'HSV-1 est un membre de la famille des *Herpesviridae*. Il possède donc un génome linéaire à ADN double-brin. La protéine virale ICP27 a un rôle majeur de modulateur du transport nucléo-cytoplasmique, en plus de sa capacité à perturber fortement l'épissage chez l'hôte. ICP27 est capable, d'une part, de bloquer l'import de protéines vers le noyau en interagissant avec la FG-Nup62 et, d'autre part, d'assurer l'export nucléaire des ARNm viraux – qui sont majoritairement exempts d'introns – grâce à sa capacité à se fixer sur l'ARN et à interagir directement avec TAP-p15 et ALY (Yarbrough et al., 2014).

La protéine ORF57 du KSHV est également capable de stimuler l'export nucléaire des transcrits viraux en fixant, à la fois, ARN et la protéine adaptatrice ALY. Ceci permet de recruter le reste du complexe TREX, ainsi que le dimère TAP-p15 (Schumann et al., 2013).

ii. Cas des rétrovirus simples

Les rétrovirus sont qualifiés de simples lorsqu'ils ne possèdent que les trois gènes *gag*, *pol* et *env*. Chez le MPMV, et comme chez tous les rétrovirus, le gène *env* codant pour la glycoprotéine d'enveloppe est exprimé à partir d'un ARN épissé tandis que *gag* et *pol*, codant respectivement pour la protéine de capsid et la transcriptase inverse, sont tous deux exprimés à partir d'un ARN non-épissé, qui sert également d'ARN génomique (ARNg ; Figure 5).



Figure 5 : Génome et transcrits du MPMV. Les différents ARN produits, et les gènes qu'ils portent, sont indiqués. Les LTR – pour « *long terminal repeat* » – sont des répétitions générées lors de la rétrotranscription, et qui contiennent le promoteur et le terminateur de transcription. CTE : *constitutive transport element*. D'après Shida, 2012.

L'expression des gènes *gag* et *pol* est permise par la relative faiblesse des deux sites d'épissage, ainsi que par la présence, à l'extrémité 3' de l'ARN viral, d'une séquence nommée CTE qui permet de recruter directement TAP. Ainsi, une fraction suffisante d'ARN viraux non épissés est exportée vers le cytoplasme (Shida, 2012).

Le virus du sarcome de Rous utilise une stratégie analogue pour exporter ses ARN non-épissés, à la différence que TAP est recrutée par deux séquences répétées et nommées *direct repeats* (Shida, 2012).

Récemment, il a été décrit que le virus de la leucémie murine (MLV) utilise également la voie TAP-p15 pour exporter ses ARN épissés et non-épissés. Le recrutement de TAP semble dépendre d'une séquence localisée à la fin du gène *pol* et nommée CAE – pour « *cytoplasmic accumulation element* » (Pessel-Vivares et al., 2014; Sakuma et al., 2014).

b) Utilisation de la voie CRM1 par les virus

i. Cas du virus du sarcome de Rous (RSV)

Certaines données suggèrent qu'en plus d'utiliser la voie TAP-p15 pour exporter ses ARN non-épissés, le RSV pourrait utiliser un second mécanisme, qui dépendrait de la protéine virale Gag. Gag a la capacité de se fixer à l'ARN viral, en particulier au niveau d'un élément d'assemblage, et de s'associer à CRM1. En interagissant avec CRM1 et une fraction des ARN viraux non-épissés, Gag pourrait exporter séparément l'ARNg du reste des transcrits non-épissés, destinés à l'expression de *gag* et *pol* (Shida, 2012).

ii. Cas des rétrovirus complexes

À l'inverse des rétrovirus simples, les rétrovirus complexes possèdent, en plus des gènes *gag*, *pol* et *env* communs à tous les rétrovirus, un ou plusieurs gènes codant des protéines dites accessoires, mais qui en réalité jouent un rôle prépondérant dans la pathogénèse de ces virus. Leurs ARN peuvent subir plusieurs événements d'épissage. Ces virus ont dû développer des stratégies pour favoriser l'export de leurs ARN partiellement épissés et non-épissés, afin de permettre l'expression de

certaines protéines et de maintenir l'intégrité génomique, alors que la machinerie cellulaire aurait tendance à seulement exporter leurs ARN complètement épissés.

Le génome du virus T-lymphotropique humain de type 1 (HTLV-1) code pour un gène accessoire nommé *rex*, qui est exprimé à partir d'un ARN doublement épissé (Figure 6A). La protéine Rex possède un signal de localisation nucléaire (NLS) et un NES. Elle est donc capable de faire la navette entre le noyau et le cytoplasme. Dans le noyau, Rex interagit spécifiquement avec les ARN viraux, en se liant, sous une forme multimérique, à une séquence fortement structurée et localisée en 3', la séquence RXRE – pour « *Rex response element* ». Par son NES, Rex recrute CRM1, ce qui conduit à l'export nucléaire des ARN viraux, en particulier ceux qui partiellement et non-épissés (Bogerd et al., 1996; Hidaka et al., 1988; Shida, 2012).

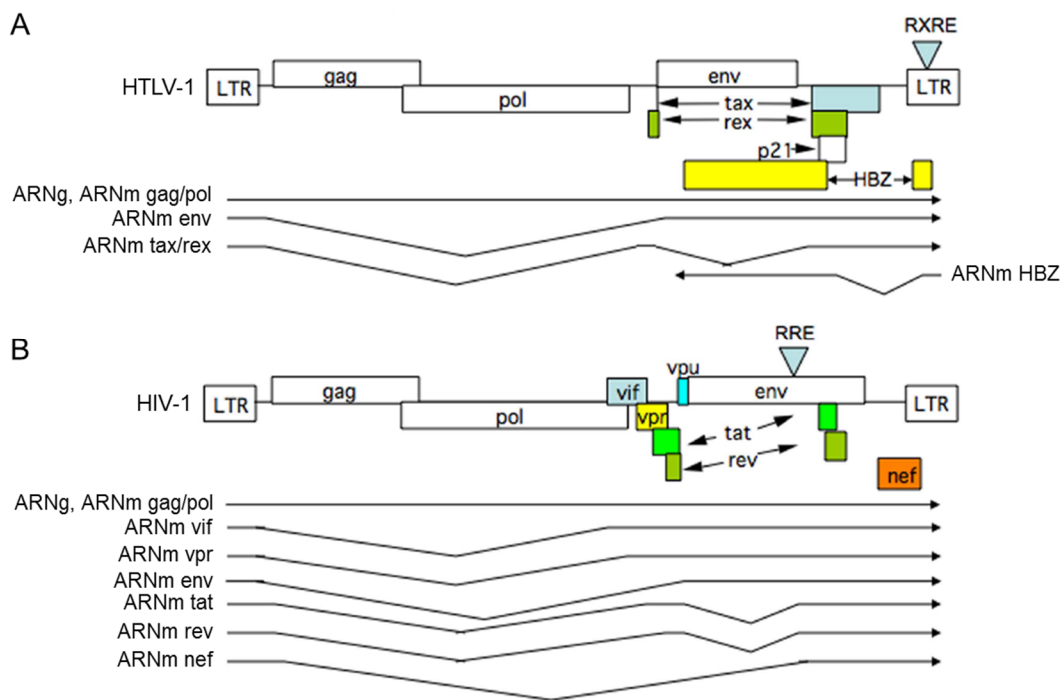


Figure 6 : Génomes et transcrits du HTLV-1 (A) et du HIV-1 (B). Les différents ARN produits, et les gènes qu'ils portent, sont indiqués. LTR : long terminal repeat. RXRE : Rex response element. RRE : Rev response element. D'après Shida, 2012.

Le virus de l'immunodéficience humaine de type 1 (HIV-1) utilise une stratégie analogue pour exporter certains de ses ARN, grâce à sa protéine Rev. La séquence de fixation de Rev, RRE – pour « *Rev response element* » –, est localisée dans le gène *env* et n'est présente que sur les ARN partiellement et non-épissés (Figure 6B). Comme Rex, Rev est capable de multimériser sur RRE et de recruter CRM1 pour l'export des ARN viraux non-épissés et partiellement épissés (Fischer et al., 1994; Malim et al., 1990; Shida, 2012).

iii. Cas des foamy virus

Le *prototype foamy virus* (PFV) est un rétrovirus et membre type du genre *Spumavirus*. En plus du promoteur habituel des rétrovirus qui se situe dans la région 5' LTR, le PFV possède un promoteur interne localisé dans la région du gène *env*. L'export des ARN non-épissés et partiellement épissés du PFV repose sur la voie CRM1 mais, à l'inverse des autres rétrovirus, aucune protéine virale n'est impliquée. C'est la protéine cellulaire HuR, agissant également dans l'export nucléaire de quelques ARNm exportés par la voie CRM1, qui est recruté par le PFV, pour ce processus. HuR se

fixe probablement à une séquence virale qui n'a pas encore été caractérisée puis, avec l'assistance des protéines ANP32A et ANP32B, recrute l'exportine CRM1 pour exporter les ARN viraux vers le cytoplasme (Bodem et al., 2011).

iv. Cas du virus de l'influenza A (IAV)

L'IAV est un virus à ARN simple-brin multipartite et de polarité négative, membre de la famille des *Orthomyxoviridae*, et dont la transcription et la réplication, effectuées par une ARN polymérase ARN-dépendante virale, ont lieu dans le noyau des cellules infectées. Les segments du génome viral sont produits sous forme de vRNP : la polymérase virale – un hétérotrimère – réplique le génome viral en utilisant les ARN antigénomiques comme matrice, et de nombreuses protéines NP se fixent à chaque ARN nouvellement synthétisé. Les vRNP sont exportés par la voie CRM1 grâce aux protéines virales M1 et NEP tandis que les cRNP, qui sont constituées des RNP contenant l'ARN antigénomique, restent dans le noyau. La protéine M1 peut à la fois interagir avec les protéines NP et NEP. Grâce aux différents NES localisés sur M1 et NEP, les différentes vRNP peuvent être exportées vers le cytoplasme puis s'accumuler au niveau de la membrane plasmique, où le processus d'assemblage des virions pourra commencer (Hutchinson et Fodor, 2013). L'IAV est capable d'inhiber, en parallèle, l'export nucléaire des ARNm de l'hôte *via* des interactions entre la protéine virale NS1 et certains composants clés de la machinerie d'export, tels que TAP et p15 (Satterly et al., 2007).

6) Ouverture sur le sujet d'étude

À ce jour, aucune donnée concernant l'export nucléaire des ARN du CaMV n'a été publiée. De plus, les connaissances sur l'export nucléaire des ARNm, chez les plantes, sont minces voire inexistantes en ce qui concerne l'export des ARN viraux des quelques virus de plantes dont le génome est à ADN (*Caulimoviridae*, *Geminiviridae* et *Nanoviridae*). Ainsi, la difficulté majeure du problème réside dans le fait qu'il ne peut être abordé qu'à la lumière de ce qui existe dans des systèmes relativement éloignés de notre sujet, chez des virus animaux par exemple.

Une fraction d'ARN 35S du CaMV doit nécessairement être exportée sous la forme non-épissée afin de garantir l'intégrité génomique du virus ainsi que l'expression des protéines P1 et P2, impliquées respectivement dans le mouvement du CaMV de cellule à cellule et dans sa transmission de plante à plante. Chez les animaux, où les pré-ARNm, lorsqu'ils ne sont pas encore épissés, sont retenus dans le noyau, les rétrovirus utilisent deux stratégies pour exporter des ARN viraux partiellement et non épissés. D'une part, leur sites d'épissage sont souvent faibles, par nature ou grâce à des séquences régulant négativement leur utilisation par le spliceosome (Leblanc et al., 2013). D'autre part, ils possèdent des séquences, et parfois codent également pour des protéines, qui promeuvent l'export de leurs propres ARN *via* la voie d'export utilisée par les ARNm cellulaires et/ou celle utilisée par les protéines.

Une différence notable existe chez les plantes : l'évènement d'épissage alternatif qui est le plus fréquent est la rétention d'intron, alors que ce phénomène représente moins de 5% des évènements d'épissage chez l'humain (Reddy et al., 2013). Ainsi, nombre d'ARNm de plantes sont exportés du noyau alors même qu'ils contiennent un ou plusieurs introns. Il n'est donc pas impossible que les mécanismes régulant l'export d'ARNm avec intron(s) diffèrent entre les animaux et les plantes. Alors que la voie majeure d'export des ARNm de plantes n'a pas encore été décrite, l'export nucléaire de leurs protéines est mieux caractérisé. Deux homologues de CRM1, codés par les gènes *XPO1A* et *XPO1B*, ont été mis en évidence, et le processus d'export est très similaire à celui des animaux (Blanchvillain et al., 2008; Merkle, 2011). Ce processus peut d'ailleurs être inhibé par l'inhibiteur classique qu'est la leptomycine B. Ainsi, l'étude d'une possible implication de la voie CRM1 dans l'export des ARN du CaMV semble envisageable.

Les trois questions principales qui ont sous-tendu notre approche expérimentale ont donc été les suivantes : quel(s) mécanisme(s) utilise le CaMV pour préserver en partie de l'épissage l'ARN

35S, et exporter l'ARN 35S non-épissé vers le cytoplasme ? Quelles sont les protéines qui interviennent en *trans* dans ce processus ? Quelles sont les séquences qui gouvernent, en *cis*, l'export de l'ARN 35S ? Pour tenter d'y répondre, la mise en place de protocoles expérimentaux a été nécessaire et beaucoup de temps y a été consacré.

RÉSULTATS

La thématique de l'export nucléaire des ARN de virus de plantes était, à notre connaissance, encore inexplorée lorsque j'ai commencé mon doctorat. Par conséquent, il était impératif de mettre au point des protocoles expérimentaux qui nous permettent d'apporter des éléments de réponse sur l'export de l'ARN 35S du CaMV. J'ai donc consacré le plus clair de mon temps à la conception de vecteurs viraux idoines, à l'obtention et à la transfection de protoplastes et à l'expérimentation d'approches différentes pour évaluer l'export nucléaire des ARN viraux. Deux systèmes ont été développés pour pouvoir étudier l'export de l'ARN 35S et les facteurs impliqués en *cis* et *trans* dans ce processus. Le premier système présenté repose sur l'expression d'un rapporteur fluorescent comme témoin de l'export de l'ARN 35S non-épissé ; le second est basé sur une détection ciblée des ARN 35S, dans le noyau et le cytoplasme. Dans les deux cas, la souche Cabb B-JI a été utilisée comme modèle d'étude.

1) Étude de l'interaction entre P6/TAV et la protéine AtXPO1, par GST pull-down

Comme cela a été exposé dans l'introduction de ce chapitre, il est commun que les virus utilisent en *trans* une protéine codée par leur génome pour exporter, par la voie CRM1, leurs propres ARN vers le cytoplasme et en particulier, ceux qui ne sont pas ou partiellement épissés. La protéine P6/TAV nous a semblé être le meilleur candidat à une telle fonction, pour le CaMV, puisqu'elle a un profil de localisation nucléo-cytoplasmique, grâce aux deux NLS – pour « *nuclear localization signal* » – et au NES qu'elle possède (G. Haas et al., 2008; M. Haas et al., 2005), des motifs qui sont d'ailleurs indispensables à l'infectiosité du virus (G. Haas et al., 2008). La présence d'un NES – P6/TAV est d'ailleurs la seule protéine du CaMV pour laquelle l'existence d'un NES a été démontrée (M. Haas et al., 2005) – sous-entend fortement une interaction directe avec AtXPO1a et/ou AtXPO1b, les homologues fonctionnellement redondants de CRM1, chez *Arabidopsis* (Haasen et al., 1999). Pour vérifier cette hypothèse, l'interaction entre P6/TAV et AtXPO1a a été étudiée par la technique GST pull-down.

Les ADNc de la protéine AtXPO1a (1075 acides aminés) et de ses parties amino-terminale (Nter, acides-aminés 1 à 667) et carboxy-terminale (Cter, acides-aminés 668 à 1075) ont synthétisés puis clonés dans un vecteur d'expression. Les protéines correspondantes ont été exprimées *in vitro*, dans un système de réticulocytes de lapin en présence de [³⁵S]-méthionine, puis elles ont été testées pour leur capacité à interagir avec la protéine P6/TAV fusionnée à la GST et exprimée en système *E. coli*. Comme le montre la figure 7, ni la protéine AtXPO1a complète ni ses deux versions tronquées n'ont été capables de se lier à la GST seule. Un très faible signal a été observé pour les essais d'interaction de GST-P6/TAV avec la luciférase et la partie Cter d'AtXPO1a, laissant penser à des interactions aspécifiques. À l'inverse, un signal de forte intensité a été détecté avec la version complète et la partie Nter d'AtXPO1a. Dans le cas de GST-A, qui correspond à une fusion de la GST avec le domaine A de P6/TAV – ce domaine comprend les 112 premiers acides-aminés de la protéine et renferme le NES (M. Haas et al., 2005) –, un très faible signal a été détecté avec AtXPO1a et sa partie Nter (Figure 7D). Au vu de la faible quantité de GST-A purifiée, par rapport à GST-P6/TAV (Figure 7C), cette observation n'exclut pas que le domaine A de P6/TAV interagisse avec AtXPO1a, *via* l'extrémité amino-terminale de cette dernière.

En résumé, ces résultats indiquent que P6/TAV interagit spécifiquement avec AtXPO1, très probablement *via* un ou des domaines situés dans la partie amino-terminale de l'exportine. Il a été montré que la portion centrale de la protéine CRM1 humaine est impliquée dans la liaison d'un NES, notamment une cystéine en position 531 au niveau de laquelle se lie de façon covalente, la leptomyicine B, un des inhibiteurs de la voie d'export CRM1 (Dong et al., 2009). Cette cystéine est également présente chez AtXPO1a (Haasen et al., 1999), ainsi que sur la construction AtXPO1a Nter que nous

avons testée. Ceci suggère, d'une part, que P6/TAV et AtXPO1a interagissent de manière directe, et d'autre part, que cette interaction se fait entre la zone de fixation des NES située sur AtXPO1a, et le NES de P6/TAV localisé dans le domaine A (M. Haas et al., 2005). Cette dernière hypothèse devra être vérifiée par une autre approche expérimentale et en testant des mutants ponctuels et/ou de délétion de P6/TAV, notamment le mutant P6/TAVm1 où les 3 leucines du NES ont été mutées (M. Haas et al., 2005).

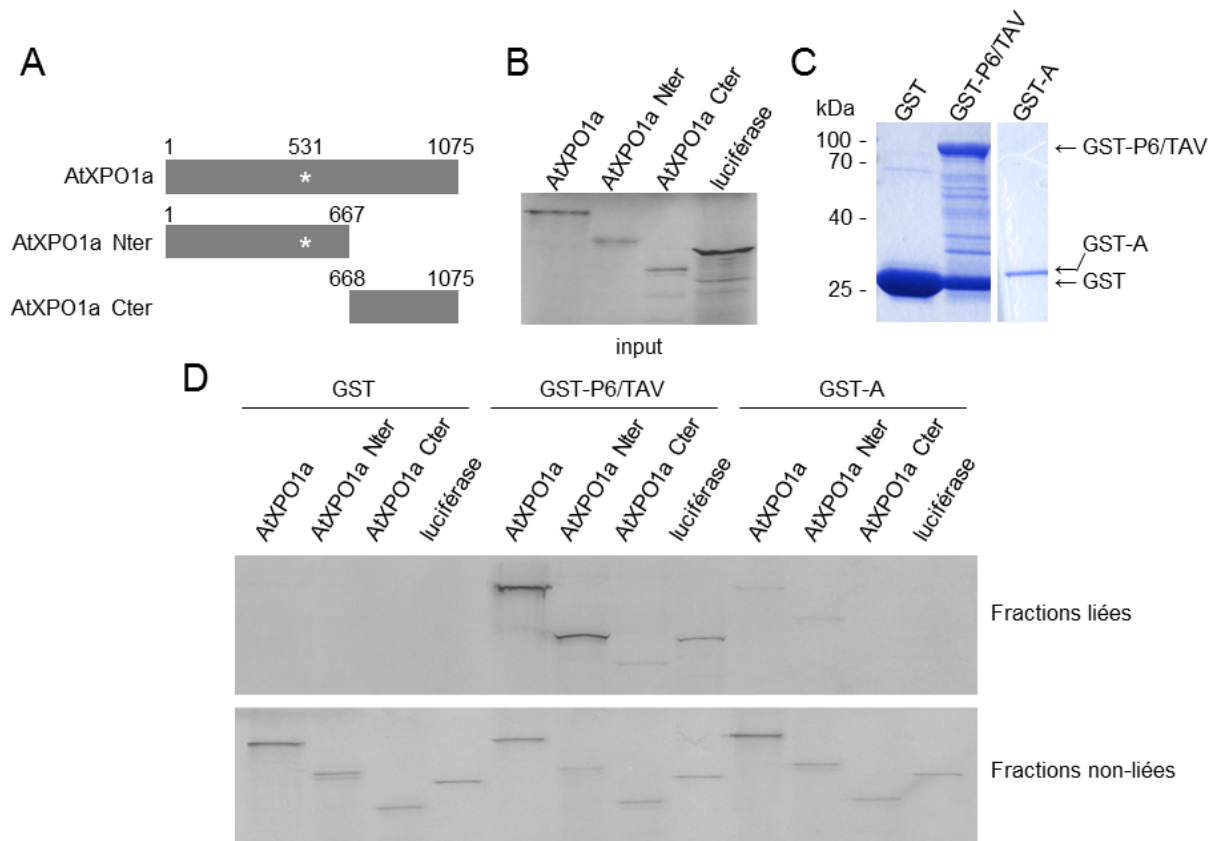


Figure 7 : Étude de l'interaction entre P6/TAV et AtXPO1a, par GST pull-down. La protéine AtXPO1a complète ainsi que deux fragments (A) ont été synthétisées *in vitro* et marquées au ^{35}S (B). La capacité d'interaction de ces protéines a été testée contre de la GST – pour « glutathion S-transférase » – seule ou fusionnée soit à la protéine P6/TAV complète, soit au domaine A de P6/TAV, exprimées chez *E. coli* (C) et liées à des billes de Sepharose. Pour chaque essai, les fractions liées et non-liées ont été séparées par SDS-PAGE (D). La luciférase de luciole *Photinus pyralis* a été utilisée comme témoin négatif d'interaction avec P6/TAV. Le site de fixation putatif de la leptomycine B dans AtXPO1a et AtXPO1a Nter est indiqué par une étoile blanche (A).

2) Étude de l'implication de P6/TAV dans le processus d'export nucléaire de l'ARN 35S non-épissé, par l'utilisation d'un rapporteur fluorescent

Pour étudier une possible implication de P6/TAV dans l'export de la forme non-épissée de l'ARN 35S, nous avons développé un système transitoire dans lequel l'export de l'ARN 35S non-épissé conduit à l'expression d'un gène rapporteur codant pour une protéine fluorescente, en présence ou en absence de P6/TAV.

a) Développement des plasmides pEXPORT

La zone comprise entre le site donneur D2 et le site accepteur A, bien que propre à la forme non-épissée de l'ARN 35S, n'est pas la plus appropriée pour insérer un gène destiné à « rapporter »

l'export nucléaire. En effet, l'expression d'un tel gène à partir d'un ARN 35S polycistronique recombinant serait inéluctablement dépendante de l'activité de P6/TAV dans la réinitiation de la traduction. Afin de s'affranchir des fonctions traductionnelles de P6/TAV, nous avons développé un plasmide nommé pEXPORT (Figure 8), qui permet la production d'un ARN 35S complet possédant le gène rapporteur codant pour la protéine fluorescente verte eGFP – pour « *enhanced green fluorescent protein* » – fusionné en phase, au petit ORF A – sORF A, constitué de 3 codons. Ce petit ORF est le tout premier ORF présent sur l'ARN 35S. Il est systématiquement traduit et cet évènement est d'ailleurs essentiel à l'expression des ORF situés en aval de la région *leader*, via le mécanisme du shunt du ribosome (Pooggin et al., 2000). La production d'ARN 35S non-épissé a été forcée en supprimant le site accepteur A, par introduction de la mutation « mutA » – se référer au premier chapitre pour plus de détails sur cette mutation. Deux versions du vecteur pEXPORT ont été produites : l'une possédant un ORF VI sauvage (pEXPORTmutA), et l'autre possédant un ORF VI inexprimable suite à la suppression du codon initiateur et l'ajout de deux codons de stop en aval (pEXPORTmutAmutP6).

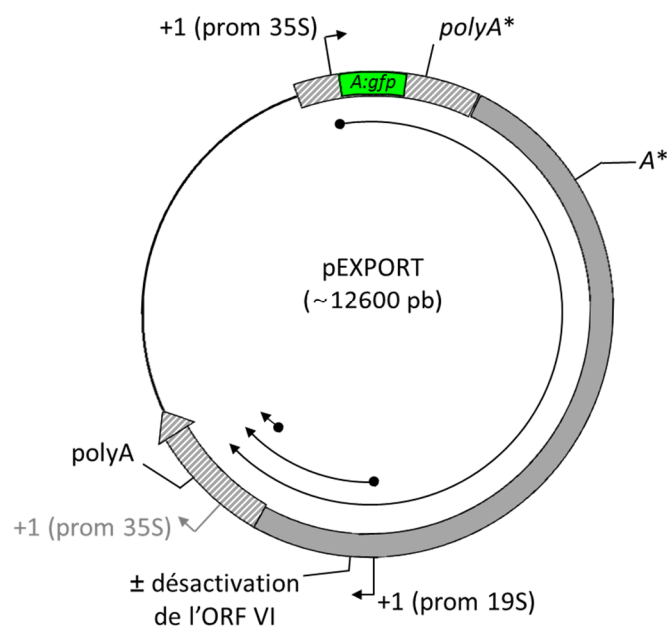


Figure 8 : Carte du plasmide pEXPORT. Les séquences qui sont dupliquées aux deux extrémités sont hachurées. Les sites d'initiation de la transcription sont représentés par des flèches coudées. Les mutations sont signalées par le signe *. Les différentes espèces d'ARN susceptibles d'être produites à partir du plasmide sont indiquées par des arcs de cercle fléchés.

Les vecteurs viraux conventionnels utilisés lors des expériences d'inoculations mécaniques ne sont pas adaptés à la production d'ARN 35S, en système transitoire. Seuls des vecteurs qui possèdent une duplication de séquence permettant la transcription des répétitions présentes aux deux extrémités de l'ARN 35S peuvent être utilisés (Khelifa et al., 2010; Kobayashi et al., 2002). La stratégie de clonage suivante, basée sur les travaux de Khelifa et al. (2010), a été utilisée : une séquence d'environ 1100 nucléotides contenant le promoteur et le terminateur 35S a été sous-clonée à partir du plasmide pCaMV-GFP (Khelifa et al., 2010), dans le vecteur LITMUS 28i (New England BioLabs). Cette séquence virale a été modifiée : le gène de l'eGFP a été fusionné au sORF A, et le signal de polyadénylation a été inactivé. La séquence modifiée a ensuite été insérée directement en amont du génome viral présent dans un vecteur viral nommé pBS-BJI. Le vecteur pEXPORT obtenu contient une duplication de séquence (portions hachurées sur la figure 8) qui permet la transcription des répétitions présentes aux deux extrémités de l'ARN 35S. Ce dernier peut donc être transcrit directement, sans qu'il y ait besoin de linéariser préalablement le génome viral à partir de du vecteur, comme lors

des expériences d'inoculation mécanique. Enfin, la suppression du signal de polyadénylation sur la première séquence virale dupliquée – le signal AATAAA est muté en AAGACA sur la répétition localisée en 5' – permet d'éviter la production d'un ARN de type *short-stop* qui ne contiendrait que le sORF A fusionné au gène de l'eGFP, ainsi que les deux sORF – B et C en l'occurrence – présents en amont du signal.

b) Transfection et résultats

Les deux versions du vecteur pEXPORT, pEXPORTmutA et pEXPORTmutAmutP6, ont été introduites par biolistique, dans des cellules de tabac *Nicotiana tabacum* BY-2 et d'*Arabidopsis thaliana* (écotype Landsberg erecta) issues de cultures liquides. Pour cela, des microbilles de tungstène recouvertes de vecteur pEXPORT ont été bombardées sur les cellules, à l'aide d'un canon à particules. Il faut noter que parmi les deux types cellulaires utilisés, seule *Arabidopsis* est un hôte véritable puisque la souche Cabb B-JI ne provoque que des lésions locales sur *Nicotiana tabacum* (Mével et Kerlan, 1990). La transfection de ces deux types cellulaires avec les constructions pEXPORTmutA et pEXPORTmutAmutP6 a conduit, à chaque fois, à la production de la protéine A-eGFP (Figure 9). Ainsi, dans ce modèle, l'expression de cette protéine montre que l'export nucléaire de l'ARN 35S non-épissé n'est pas assujettie à la présence de P6/TAV.

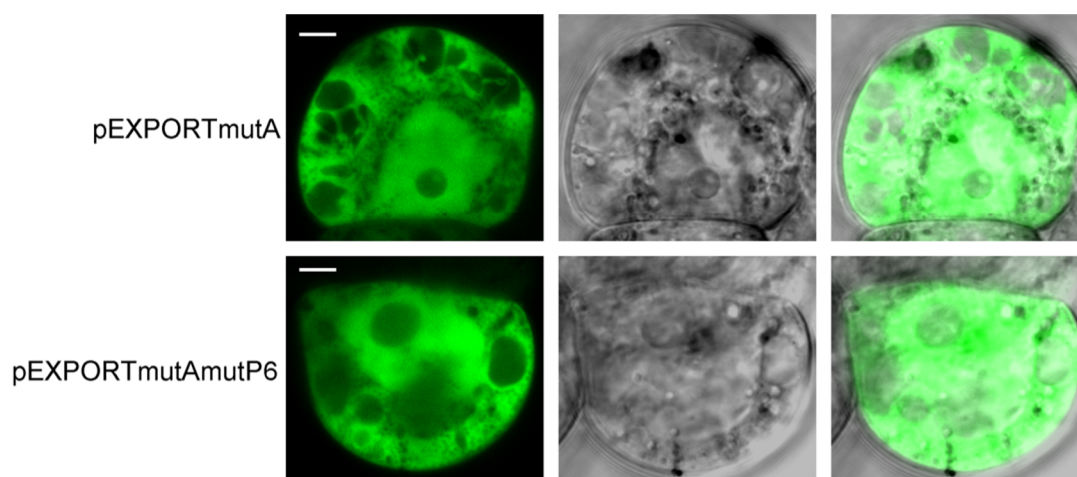


Figure 9 : Observation en microscopie confocale, de cellules d'*Arabidopsis* transfectées, par biolistique, avec les vecteurs pEXPORTmutA et pEXPORT mutAmutP6. La barre blanche représente 5 µm. À gauche, images de fluorescence collectées par CLSM ; au centre, images en contraste d'interférence ; à droite, superpositions des deux types d'images.

Dans un premier temps, ces données ont été interprétées comme démontrant que la protéine P6/TAV n'est nullement nécessaire à l'export de la forme non-épissée de l'ARN 35S bien qu'en l'absence de quantification des ARN exportés, nous ne pouvions pas exclure qu'elle puisse simplement stimuler ce processus. Cependant, nos travaux sur l'épissage, réalisés en parallèle, ont indiqué que la mutation mutA du site accepteur n'inhibe pas l'épissage puisqu'elle conduit à l'utilisation de deux sites accepteurs cryptiques, situés à proximité du site A. Les transcrits conduisant à la production de la protéine A-eGFP pouvaient donc avoir été préalablement épissés. Ainsi, ce système expérimental ne nous a pas permis de conclure sur le rôle de P6/TAV dans l'export nucléaire de l'ARN 35S puisqu'il n'autorise pas la distinction entre formes épissées et non-épissées. À cela s'ajoute la possibilité que la mutation destinée à supprimer le premier signal de polyadénylation (Figure 8) ne soit pas efficace et donc que la protéine A-eGFP résulte de la traduction d'un ARN *short-stop*. Nous n'avons malheureusement pas pu vérifier ces deux hypothèses, car les taux de transfection des deux types

cellulaires – tabac BY-2 et *Arabidopsis* – n'ont jamais été suffisamment élevés – généralement inférieurs à 5% – pour détecter, par northern-blot, les transcrits produits à partir des constructions. Enfin, comme cela a été évoqué précédemment, notre système, tel qu'il a été développé, ne permettait pas non plus de statuer entre une non-implication et un rôle régulateur de P6/TAV.

3) Étude de l'export nucléaire des différentes formes d'ARN 35S par détermination de la quantité d'ARN non-épissé présente au sein de la population globale d'ARN 35S, dans le noyau et le cytoplasme

Pour pallier aux limites énoncées ci-dessus, nous avons mis au point un système d'étude de l'export nucléaire de l'ARN 35S qui permet de s'affranchir de l'expression d'un rapporteur et qui repose sur la mesure directe d'un ratio témoignant de la quantité d'ARN 35S non-épissé au sein de la population d'ARN 35S. Ce ratio est calculé pour les ARN nucléaires et cytoplasmiques, et les valeurs obtenues dans différentes conditions sont comparées. Par exemple, l'accumulation d'ARN 35S non-épissé dans le noyau, suite à un traitement avec un inhibiteur de l'export nucléaire, peut être détectée. En théorie, ce système est plus adaptable que celui que nous avons développé précédemment car il permet l'étude de l'influence que peuvent avoir les éléments en *cis* – séquences virales – et en *trans* – protéines –, ainsi que certains inhibiteurs comme la leptomycine B, sur l'export de l'ARN 35S.

a) Développement du plasmide pBS-BJI dbA

Le plasmide de base pour ce système, nommé pBS-BJI dbA, est représenté dans la figure 10. Il permet la production de l'ARN 19S et, comme les plasmides pEXPORT, celle de l'ARN 35S complet avec ses répétitions de séquences aux extrémités 5' et 3'. La stratégie de clonage utilisée est similaire à celle mise en œuvre pour les plasmides pEXPORT : la séquence virale d'environ 1100 nucléotides, issue du plasmide pCaMV-GFP (Khelifa et al., 2010) contenant le promoteur et le terminateur 35S (région hachurée en 5' sur la figure 10) sous-clonée dans le vecteur LITMUS 28i a été insérée en amont du génome viral complet présent dans le vecteur pBS-BJI.

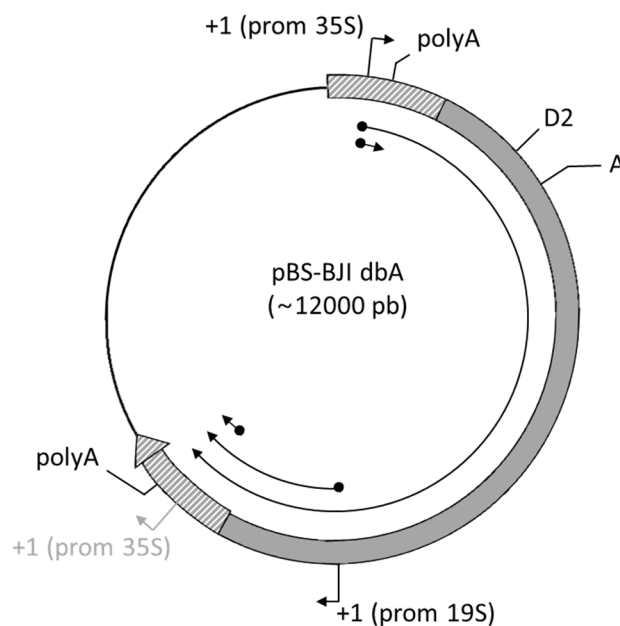


Figure 10 : Carte du plasmide pBS-BJI dbA. Les séquences qui sont dupliquées aux deux extrémités sont hachurées. Les sites d'initiation de la transcription sont représentés par des flèches coudées. Les différentes espèces d'ARN susceptibles d'être produites à partir du plasmide sont indiquées par des arcs de cercle fléchés.

Inoculée mécaniquement à des plants de navets sans digestion préalable par des enzymes de restriction, le plasmide pBS-BJI dbA, qui renferme donc environ 1,1 copie du génome viral, conduit au développement de symptômes systémiques identiques à ceux de plantes inoculées avec un extrait brut et ce, sans délai dans leur apparition. Une observation similaire a déjà été faite avec un vecteur ayant des caractéristiques très proches du notre (Kobayashi et al., 2002). Ce résultat reflète certainement la capacité qu'a le plasmide à conduire directement à la production de l'ARN 35S pré-génomique. En outre, la présence de deux séquences génomiques répétées de 1100 paires de bases offre également la possibilité d'excision et de circularisation de l'ADN viral, par recombinaison homologue.

b) Méthodes de transfection

Plusieurs méthodes de transfection ont été testées, le but étant d'obtenir un taux de transfection suffisamment élevé pour permettre l'analyse des ARN transcrits à partir du vecteur viral, par northern-blot et/ou qRT-PCR. Ce taux de transfection a été évalué par co-transfection des cellules ou protoplastes avec le vecteur pCK-eGFP, qui permet l'expression du gène codant l'eGFP sous le contrôle du promoteur 35S.

i. Transfection par biolistique

La première méthode de transfection que nous avons testée est la biolistique, à partir de cellules de tabac BY-2 et d'*Arabidopsis* provenant de cultures liquides. Dans les deux cas, le système s'est révélé être insatisfaisant, surtout du fait du très faible taux de transfection obtenu – en général inférieur à 5% – et ce, quels que soient les paramètres sélectionnés – quantité d'ADN sur les billes, distance de tir, etc...

ii. Transfection de protoplastes au polyéthylène glycol

Des tests de transfection ont été réalisés sur trois types de protoplastes : des protoplastes de mésophylle de navet et d'*Arabidopsis*, ainsi que des protoplastes de cellules d'*Arabidopsis* en culture liquide. Les meilleurs taux de transfection – jusqu'à 70% – ont été obtenus avec des protoplastes de mésophylle de navet. Malheureusement, nous avons observé d'énormes variations puisque ce taux est souvent descendu à 10%, ce qui était incompatible avec une utilisation en routine. Des taux de transfection assez moyens – généralement situés aux alentours de 20% – et également fortement variables d'une expérience à l'autre, ont été obtenus à partir des protoplastes de mésophylle d'*Arabidopsis*. Seuls les protoplastes préparés à partir de cellules d'*Arabidopsis* en culture liquide ont donné des taux de transfection relativement satisfaisants – allant de 30% à 40% – et assez reproductibles.

c) Préparation d'ARN nucléaires et cytoplasmiques

L'étude de l'export nucléaire de l'ARN 35S, à l'aide du plasmide pBS-BJI dbA, repose sur la mesure de la quantité de l'ARN 35S non-épissé au sein de la population d'ARN 35S, dans le noyau et le cytoplasme. Les ARN nucléaires et cytoplasmiques ont été isolés après fractionnement subcellulaire des protoplastes transfectés. Les protoplastes ont été incubés dans un tampon hypotonique puis vortexés. Le lysat ainsi obtenu a été soumis à plusieurs étapes de centrifugation qui ont conduit à l'obtention d'une fraction nucléaire et d'une fraction cytoplasmique, d'où les ARN ont été extraits. La qualité du fractionnement a été estimée par détection de snoARN, qui sont des ARN exclusivement nucléaires, par northern-blot. Comme le montre la figure 11, le protocole de fractionnement cellulaire a permis d'obtenir des ARN cytoplasmiques relativement purs puisque les snoARN U25a et 37 n'ont été détectés que dans les fractions nucléaires ou à partir des ARN totaux.

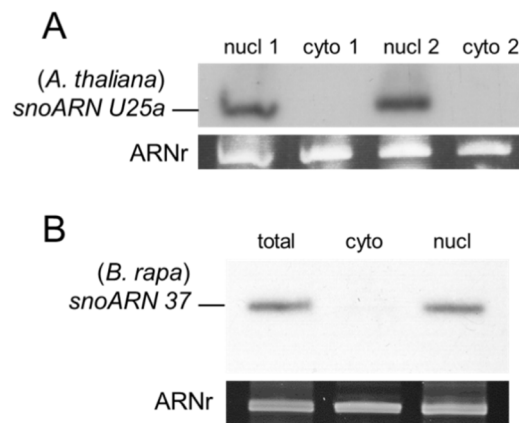


Figure 11 : Détection de snoARN dans des fractions nucléaires et cytoplasmiques obtenues à partir de protoplastes d'*Arabidopsis* et de navet *B. rapa*. (A) Détection du snoARN U25a dans des fractions nucléaires (nucl) et cytoplasmiques (cyto) de protoplastes préparés à partir de cellules d'*Arabidopsis* en culture liquide. Deux expériences indépendantes sont présentées. (B) Détection du snoARN 37 dans les ARN totaux et dans des fractions nucléaire et cytoplasmique de protoplastes de mésophylle de navet. Les ARN ont été séparés par électrophorèse sur gel de polyacrylamide (8%) - urée (8M), transférés sur membrane de nylon puis hybridés avec des sondes oligonucléotidiques radiomarquées au ^{32}P .

d) Détection des différentes formes de l'ARN 35S

i. Détection par northern-blot

Plusieurs tentatives de détection des différentes formes de l'ARN 35S ont été faites par northern-blot, en utilisant une sonde radioactive dirigée contre une région située immédiatement en aval du site accepteur. Des essais ont d'abord été réalisés avec des ARN totaux de protoplastes de mésophylle de navet et de cellules d'*Arabidopsis* en culture liquide, extraits 24h après leur transfection avec le plasmide pBS-BJI dbA. Dans nos conditions expérimentales, la technique du northern-blot s'est avérée insuffisamment sensible pour pouvoir détecter les différentes formes de l'ARN 35S dans les protoplastes de navet transfectés. Il est intéressant de noter que nos difficultés à détecter l'ARN 35S ont été encore plus accentuées lorsque les protoplastes d'*Arabidopsis* ont été utilisés. Il semble d'ailleurs que ce différentiel dans l'accumulation d'ARN viraux soit généralisable aux plantes infectées étant donné que la quantité d'ARN 35S de la souche Cabb B-JI est plus faible chez *Arabidopsis* que chez le navet (Figure 12 ; comparer le signal à 21 dpi chez le navet à celui à 19 dpi chez *Arabidopsis*), alors que le moment d'apparition des premiers symptômes systémiques – 10 à 15 dpi – et le temps de réplication sont similaires chez les deux espèces (Cecchini et al., 1998; Khelifa et al., 2010). La faible accumulation des ARN viraux dans les protoplastes transfectés avec le plasmide pBS-BJI dbA et ce, quelle que soit l'espèce utilisée, n'a pas permis l'analyse, par northern-blot, des différentes formes de l'ARN 35S présentes dans les fractions nucléaire et cytoplasmique.

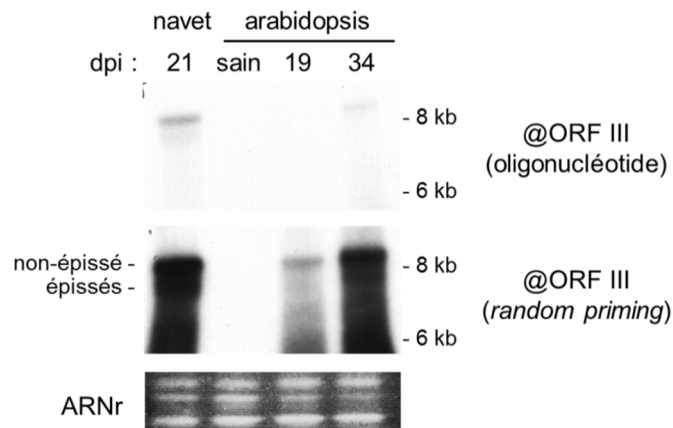


Figure 12 : Détection de l'ARN 35S par northern-blot, chez des plants de navets et d'*Arabidopsis* infectés. Les ARN totaux ont été séparés par électrophorèse sur gel d'agarose (0,8%) - urée (7M), transférés sur membrane de nylon puis hybridés soit avec une sonde oligonucléotidique dirigée contre l'ORF III, soit avec des sondes générées aléatoirement par *random priming*, et dirigées contre ce même ORF.

ii. Détection par RT-PCR quantitative (qRT-PCR)

La détection de l'ARN 35S par northern-blot étant limitée par le taux de transfection des protoplastes, une méthode de détection plus sensible, la qRT-PCR, a été choisie. Les protoplastes de mésophylle de navet et ceux de cellules d'*Arabidopsis* en culture liquide ont tous les deux été utilisés comme matériel expérimental. Deux régions de l'ARN 35S ont été ciblées : l'une située entre les sites D2 et A pour ne détecter que la forme non-épissée, et l'autre située en aval du site A pour détecter l'ensemble des formes de l'ARN 35S. Ces deux régions ont été quantifiées de façon absolue afin de pouvoir être comparées l'une à l'autre, en utilisant un composé marquant l'ADN de façon non-spécifique, le SYBR Green I (Life Technologies). La quantification absolue permet ici le calcul du ratio déterminant la part représentée par la forme non-épissée, au sein de la population globale d'ARN 35S. L'optimisation des paramètres pour cette approche – oligonucléotides utilisés, paramètres de rétrotranscription, etc... – a été particulièrement longue et ardue. Elle était encore en cours à la fin de mon doctorat, étant donné que les premiers résultats que nous avons obtenus se sont révélés être aberrants : une première expérience complète a été réalisée avec des protoplastes transfectés et traités à la leptomycine B, un inhibiteur de la voie d'export CRM1, pour tenter de déterminer si cette

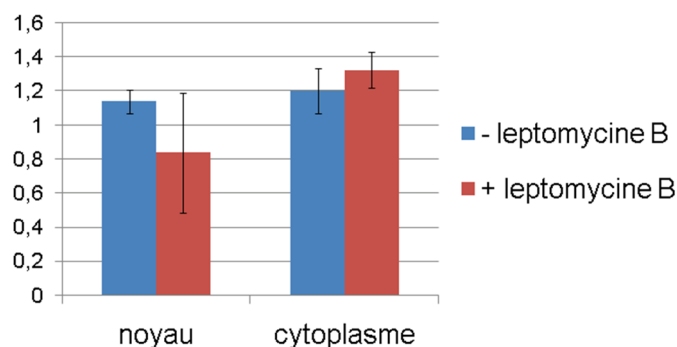


Figure 13 : Évolution du ratio (ARN 35S non-épissé)/(ARN 35S), en fonction de la présence de leptomycine B. Des protoplastes de mésophylle de navet ont été transfectés avec le plasmide pBS-BJI dbA, incubés dans 50 μ M de leptomycine B et récoltés 16 h après transfection. Les quantités de la forme non-épissée et de la population globale d'ARN 35S ont été déterminées dans les fractions nucléaire et cytoplasmique par qRT-PCR, à partir de deux réplicats biologiques indépendants.

dernière était impliquée dans l'export nucléaire de l'ARN 35S. Les données obtenues par qRT-PCR n'ont pas permis de répondre à cette question puisqu'elles ont abouti à des ratios (ARN 35S non-épissé)/(ARN 35S total) supérieurs à 1 (Figure 13). Ce résultat n'est pas dû à la fixation non-spécifique des oligonucléotides utilisés puisque aucune amplification n'a été détectée lorsque des ARN issus de protoplastes transfectés avec un vecteur vide, le vecteur pBS, ont été utilisés. Le système d'étude de l'export nucléaire de l'ARN 35S par détermination de la quantité d'ARN non-épissé présente au sein des populations nucléaire et cytoplasmique d'ARN 35S est actuellement en cours d'optimisation.

DISCUSSION

1) Les différents ARN viraux produits lors de l'infection

La présence des ARN du CaMV dans le compartiment cytoplasmique est consécutive à plusieurs processus nucléaires – transcription, épissage et transport nucléaire notamment – dont certains auront sans doute été détournés à profit par le virus. Le CaMV est un modèle d'étude particulièrement intéressant pour ce qui a trait à l'export nucléaire des ARNm car les différents types d'ARN viraux exportés du noyau, pendant l'infection, ont des caractéristiques bien distinctes. L'ARN monocistronique 19S, codant pour P6/TAV, est un ARNm qui n'a pas été épissé, étant donné que son pré-ARNm est dépourvu d'introns. Cette caractéristique est partagée avec un nombre limité de transcrits produits par la plante hôte, vu que plus de 80% des pré-ARNm de gènes nucléaires possèdent au moins un intron chez les plantes (Reddy, 2007). À ce point de vue, les ARN 35S dits *spliced-liked* se rapprochent de l'ARN 19S car ils n'ont pas d'introns au stade pré-ARNm, et si leur séquence est identique à celle d'un ARN épissé, c'est qu'ils sont transcrits à partir de génomes issus de la rétrotranscription d'ARN épissés. Les pré-ARN 35S transcrits à partir du génome viral complet contiennent des introns qui seront épissés ou non, avant l'export nucléaire. Ainsi, au cours du cycle infectieux, environ 30% des molécules d'ARN 35S échappent au processus d'épissage et sont directement exportées vers le cytoplasme. Au sein de ce compartiment, elles auront la double fonction d'ARN pré-génomique et d'ARNm pour permettre, à la fois, la réplication du génome viral et l'expression des ORF I et II, ce qui assurera la pérennité du virus. Considérant les caractéristiques variées qui viennent d'être énoncées, il est possible que les différents transcrits produits par le CaMV mettent en jeu des mécanismes variés pour recruter leurs facteurs d'export, voire même qu'ils utilisent des voies d'export différentes, comme c'est le cas, dans le monde animal, des rétrovirus complexes. Pour finir, le cas des ARN 8S sens et anti-sens, qui sont probablement transcrits à partir de la forme relâchée de l'ADN viral, est plus énigmatique. Ces ARN, qui sont non-polyadénylés, ne sont pas utilisés à des fins d'expression mais plutôt clivés par les quatre protéines *Dicer-like* (Blevins et al., 2011), qui semblent avoir une localisation majoritairement nucléaire (Hiraguri et al., 2005; Xie et al., 2004). Par conséquent, rien ne prouve que ces ARN soient exportés vers le cytoplasme.

On peut donc supposer que des voies d'exportation nucléaire et/ou des machineries d'export différentes sont utilisées pour l'export des différents ARN viraux. À ce titre, l'élucidation des mécanismes d'export des ARN du CaMV contribuerait, sans aucun doute, à une meilleure compréhension de l'export des ARNm cellulaires chez les plantes, d'autant plus que les informations à ce sujet sont clairessemées. Notre projet sur l'export des ARN du CaMV comprenait trois questions principales : quelle est la voie empruntée pour l'export nucléaire de l'ARN 35S complet ? L'export est-il contrôlé par une ou plusieurs protéines virales ou est-il entièrement sous la dépendance de la machinerie cellulaire ? Quels sont les éléments structuraux de l'ARN 35S qui régulent cet export en *cis* ?

2) Voie empruntée pour l'export nucléaire de l'ARN 35S complet

L'étude de l'export de l'ARN 35S non-épissé du CaMV s'est heurtée d'emblée à deux obstacles. D'une part, il a été difficile d'obtenir des protoplastes de plantes hôtes, ou plus généralement un système d'étude transitoire, avec un fort taux de transfection et une bonne reproductibilité. D'autre part, il s'est avéré impossible de générer un ARN 35S muté échappant complètement à l'épissage puisque l'inactivation du site accepteur A conduit à l'utilisation de sites accepteurs situés à proximité – voir le chapitre 1. Le remplacement des ORF I et II est envisageable mais il nous exposerait invariablement au risque de retirer des séquences contrôlant l'export en *cis*. Ceci explique pourquoi la première approche reposant sur l'expression d'un gène rapporteur placé en début d'un ARN 35S muté au niveau du site accepteur A, a été remplacée par une seconde approche basée sur la détermination de la quantité d'ARN non-épissé présente au sein de la population globale d'ARN 35S, dans le noyau et

le cytoplasme. Les rétrovirus complexes et les *Spumavirus* exportent leurs ARN pré-génomiques par la voie CRM1. Nous avons donc testé l'effet de la leptomycine B, qui inhibe la voie CRM1, sur l'export de l'ARN 35S dans des protoplastes transfectés avec le vecteur pBS-BJI dbA. Ces tests ont été peu concluants, car la détermination par qRT-PCR, de la quantité d'ARN 35S non-épissé parmi les populations nucléaire et cytoplasmique d'ARN 35S, a conduit à des valeurs aberrantes. L'optimisation de notre système est donc encore nécessaire. Pour vérifier la précision de la quantification par qRT-PCR, des mélanges de quantités connues de différents vecteurs contenant les ADNc d'isoformes non-épissée et épissées seront utilisés comme matrice d'amplification. Ainsi, les mesures obtenues par qPCR, à partir de ces différents mélanges, pourront être confrontées aux valeurs théoriques des ratios.

3) Régulation de l'export nucléaire de l'ARN 35S par une ou plusieurs protéines virales

La première question à laquelle nous avons essayé de répondre consistait à savoir si la protéine P6/TAV est impliquée dans l'export de l'ARN 35S non-épissé, à l'instar de certaines protéines virales comme Rev, chez le HIV-1 (Fischer et al., 1994; Malim et al., 1990), ou Rex, chez le HTLV-1 (Bogerd et al., 1996; Hidaka et al., 1988). En effet, P6/TAV nous a paru être un bon candidat puisqu'elle est capable, à la fois, de se lier aux ARN simple- et double-brin (Cerritelli et al., 1998; De Tapia et al., 1993), et de faire la navette entre le noyau et le cytoplasme – via deux NLS et un NES (G. Haas et al., 2008; M. Haas et al., 2005). Un rôle dans l'export des ARN semblait d'autant plus plausible que l'abolition de l'export nucléaire de P6/TAV, par mutagenèse du NES, est létale pour le CaMV et ce, bien que la protéine virale mutée conserve sa capacité à activer la réinitiation de la traduction de l'ARN 35S (G. Haas et al., 2008) et à supprimer le *RNA silencing* (Laird et al., 2013). Les résultats de nos expériences de GST pull-down indiquent que P6/TAV interagit, probablement de façon directe, avec AtXPO1a, un homologue de CRM1 chez *Arabidopsis*. Cependant, rien ne prouve que l'interaction P6/TAV-AtXPO1a ait une fonction dans l'export nucléaire de l'ARN 35S. Il est possible que cette interaction empêche une accumulation de P6/TAV dans le noyau qui pourrait être délétère pour le cycle infectieux du CaMV. En effet, la présence nucléaire de trop fortes quantités de P6/TAV pourrait conduire à des corps d'inclusion perturbant les fonctions et les structures nucléaires. Elle pourrait aussi, à cause de l'activité de supprimeur du *RNA silencing* de P6/TAV, donner lieu à une inhibition exacerbée de la machinerie du *RNA silencing* qui serait néfaste pour l'hôte et, par conséquent, pour le CaMV. Pour apporter des éléments de réponse sur l'implication de la protéine P6/TAV dans l'export de l'ARN 35S, nous avons développé des vecteurs viraux capables d'exprimer ou non la protéine P6/TAV et ayant, à la fois, un gène rapporteur – le gène de l'eGFP, en 5' du génome viral – et le site accepteur d'épissage A inactivé. Pour les raisons évoquées précédemment, aucune conclusion ne peut être formulée à partir de nos résultats puisque nous n'avons pas pu étudier une population homogène constituée exclusivement d'ARN 35S non-épissé.

P6/TAV est, théoriquement, le candidat viral le plus sérieux à un rôle dans l'export des ARN viraux. Cependant, il se peut que d'autres protéines virales soient impliquées dans ce processus, comme par exemple la protéine P3, bien que celle-ci ait déjà des fonctions dans la transmission du CaMV (Leh et al., 1999) et son mouvement de cellule à cellule (Stavolone et al., 2005). Si P3 est impliquée dans l'export des ARN du CaMV, son rôle n'est peut-être pas prépondérant car le virus se réplique dans des protoplastes transfectés avec un vecteur viral n'exprimant pas P3 à cause d'un décalage de cadre de lecture dans l'ORF III (Kobayashi et al., 2002). Cependant, ces travaux n'excluent pas que l'export de l'ARN 35S non-épissé ait été perturbé en l'absence de P3. En effet, ces auteurs ont suivi la réplication virale par immunodétection de la protéine de capsid P4 et par amplification d'une région d'ADN viral n'incluant pas la zone sujette à délétion lors de l'épissage. Des ARN épissés traduits et rétrotranscrits pourraient expliquer, à eux seuls, ces résultats.

Comme P6/TAV, la protéine P3 est capable de lier l'ARN (Jacquot et al., 1996, 1998). Sa petite taille – 15 kDa – lui permet de gagner et quitter le noyau seulement par diffusion puisqu'à ce jour, il

n'a été caractérisé ni NLS – généralement riches en lysines et arginines – ni NES – souvent une séquence de type $\Phi x_{2-3} \Phi x_{2-3} \Phi x \Phi$, où Φ est généralement une leucine et parfois une isoleucine, une valine, une phénylalanine ou une méthionine, et où x est n'importe quel acide aminé. Cependant, des analyses *in silico* font apparaître, dans la région N-terminale de P3, la séquence MKADIKAILLELL qui pourrait s'apparenter à un NES, mais sa fonctionnalité n'a pas encore été étudiée. Quoi qu'il en soit, P3 pourrait également coopérer, voire agir en synergie, avec P6/TAV étant donné que ces deux protéines virales peuvent interagir entre elles *in vitro* (Lutz et al., 2012). Un rôle putatif de P3 et/ou P6/TAV pourra être étudié en insérant des mutations *ad hoc* dans le vecteur pBS-BJI dbA. Toutefois, il convient de rappeler que si la voie CRM1 est utilisée par les rétrovirus complexes, les protéines qui interviennent en *trans* ne sont pas nécessairement virales. Le *prototype foamy virus*, un rétrovirus qui ressemble par bien des égards au CaMV, utilise des protéines nucléaires adaptatrices pour exporter son ARN non-épissé par la voie CRM1.

4) Éléments structuraux de l'ARN 35S qui régulent l'export en *cis*

Le vecteur pBS-BJI dbA permet également la recherche de séquences engagées en *cis*, dans l'export nucléaire des ARN viraux. En effet, tous les rétrovirus utilisent des séquences en *cis* pour exporter leurs ARN et ce, indépendamment de la voie d'export utilisée et de l'implication d'une protéine virale dans le phénomène. Les séquences d'export des rétrovirus forment des structures secondaires auxquelles se fixent des protéines adaptatrices cellulaires ou virales. Dans le cas du CaMV, il est légitime de penser que la région 5' *leader*, connue pour être fortement structurée, pourrait jouer un rôle en recrutant des protéines cellulaires et/ou virales nécessaires l'export de l'ARN 35S (Dominguez et al., 1996). L'implication de séquences localisées dans d'autres régions de l'ARN 35S ne doit cependant pas être exclue. Une séquence localisée à l'extrémité 3' de l'ARN 35S – comme c'est le cas du CTE chez le MLV et du RexRE chez le HTLV-1 – pourrait concourir, à la fois, à l'export de l'ARN 19S et des différentes formes de l'ARN 35S. En parallèle de l'utilisation du vecteur pBS-BJI dbA, une des approches possibles pour caractériser la séquence d'export serait d'effectuer des expériences de gel retard en utilisant des extraits nucléaires ou des protéines virales – notamment P6/TAV et P3 – et des transcrits de petite taille – 500 à 750 nucléotides – couvrant la totalité de la séquence de l'ARN 35S. Ce type d'étude permettrait, par la même occasion, d'apporter des éléments de réponse quant à la nature virale ou cellulaire des protéines adaptatrices impliquées dans l'export de l'ARN 35S.

MATÉRIELS ET MÉTHODES

1) Plasmides

a) Vecteurs d'expression d'AtXPO1a et de ses formes tronquées

2,5 µg d'ARN totaux d'*Arabidopsis* sont traités avec 2,5 U de Dnase I (Promega) pendant 30 min à 37°C. Les ARN sont ensuite extraits au phénol/chloroforme puis repris dans de l'eau stérile. La rétrotranscription en ADNc est faite avec l'enzyme Maxima Reverse Transcriptase (Thermo Scientific), en présence d'oligo(dT)₁₈. L'ADNc correspondant à l'ARNm d'AtXPO1a a été amplifié par PCR avec l'ADN polymérase Phusion (Thermo Scientific) (oligonucléotides : 5'-GGGGACAAGTTTGTACAAA-AAAGCAGGCTACATGGCGGCTGAGAAGTTAAGGG-3' et 5'-GGGGACCACTTTGTACAAGAAAGC-TGGGTATTATGAGTCCACCATCTCGTCT-3') puis cloné dans le vecteur pDONR/Zeo par Gateway (Invitrogen), selon les instructions du fabricant ; le vecteur ainsi obtenu a été nommé pENTR-AtXPO1a. La séquence codante d'AtXPO1a et ses versions tronquées ont ensuite été sous-clonées dans un vecteur d'expression pGBKT7 (Clontech).

b) Vecteurs d'expression de la protéine P6/TAV et de son domaine A

L'ORF VI, ainsi qu'une version tronquée codant pour le domaine A – qui correspond au 112 premiers acides-aminés de P6/TAV –, ont été introduits dans le vecteur d'expression pGEX-6P1 (GE Healthcare) qui permet une fusion de l'extrémité 5' de la séquence codante d'intérêt, avec la séquence codante de la GST.

c) Plasmides du type pEXPORT

Un fragment du plasmide pCaMV-GFP (Khelifa et al., 2010), d'environ 1100 nucléotides et contenant le promoteur et le terminateur 35S, a été digéré par *KpnI* et *BsWI* et sous cloné dans un vecteur Litmus 28i (New England BioLabs). Ce fragment a été nommé « séquence A ». Le gène codant pour l'eGFP a été fusionné au sORF A. Pour cela, un site *SalI* a été inséré à la fin de l'ORF A, par PCR (oligonucléotides 5'-AAATAATGTGTGAGTCGACTCCCGATAAGGG-3' et 5'-ATGGAGAAA-ATAGAGAGAGATAGATTTGTAG-3'). L'ORF codant pour l'eGFP a été amplifié par PCR, à partir du plasmide pCKeGFP et avec des oligonucléotides possédant chacun un site *SalI* (séquences : 5'-ATTGTCGACATGGTGAGCAAGGGCGAGG-3' et 5'-ATTGTCGACTTACTTGTACAGCTCGTCCA-3'). Ensuite, un nucléotide situé entre le sORF A et le gène de l'eGFP a été supprimé afin de fusionner ces deux gènes (oligonucléotides : 5'-AATGTGTGAGTCGAGCATGGT-3' et 5'-ATTTATGGAGAAAA-TAGAGAGAGAT-3'). Pour finir, le signal de polyadénylation situé dans le fragment inséré a été inactivé (oligonucléotides : 5'-ACTTCTATCAAGACAATTTCTAATTCCT-3' et 5'-ATTTTACAAA-TACAAATACATACTAAGG-3').

En parallèle, le vecteur viral pBS-BJI (don du Dr. Martin Drucker, INRA Montpellier) qui contient le génome de la souche Cabb B-JI cloné par *SpeI* a été modifié par mutagenèse dirigée. Le site accepteur A a été inactivé (mutation mutA ; oligonucléotides 5'-ATGAGATTGTTGCCAGAAGCTTATCTC-3' et 5'-TGAGGTTGTTAAGATGATTAATTA-3'). L'ORF VI possède, dans ces 50 premiers nucléotides, trois codons ATG dans la même phase de lecture. Pour empêcher la production de P6/TAV, trois mutations ont été introduites (mutP6). Le codon ATG initiateur a été inactivé (oligonucléotides 5'-CTCCAAGCATAGAGAACATAG-3' et 5'-GTCTGATTTTCTCAGTCTCCA-3'), puis deux codons stop ont été insérés, l'un entre le deuxième et le troisième ATG et l'autre, après le troisième ATG (oligonucléotides 5'-ACTAATGCTATAACTCGATCTAGT-3' et 5'-ATTTACTCTTGCATGAGGAG-3'). La séquence A contenant l'ORF de fusion A-eGFP et le signal de polyadénylation désactivé a été amplifiée par PCR (séquences des oligonucléotides 5'-CTATAGGGCAGATCTTCGAAT-3' et 5'-

TATTCCGCGGTAATACGACTCACTATAGGCC-3') et clonée dans les vecteurs pBS-BJlmutA et pBS-BJlmutAmutP6 par *Sac*II et *Bs*WI pour donner, respectivement les vecteurs pEXPORTmutA et pEXPORTmutAmutP6.

d) Plasmide pBS-BJl dbA

La séquence A non-mutée clonée dans le plasmide Litmus28i a été amplifiée par PCR avec les oligonucléotides 5'-CTATAGGGCAGATCTTCAAT-3' et 5'-TATTCCGCGGTAATACGACTCATATAGGCC-3' puis clonée dans le vecteur pBS B-Jl par *Sac*II et *Bs*WI.

2) Analyse d'interactions protéiques *in vitro*, par GST pull-down

a) Transcription et traduction *in vitro* des protéines AtXPO1a

Les protéines AtXPO1a ont été synthétisées *in vitro* grâce au système de transcription couplée à la traduction dans des lysats de réticulocytes de lapin (TNT T7 Coupled Reticulocyte Lysate System, Promega), selon les instructions du fabricant. En bref, les ARNm codant les protéines d'intérêt sont synthétisés par la T7 ARN polymérase, à partir d'un promoteur T7 présent dans les plasmides pGBKT7-AtXPO1a, et ils sont ensuite traduits en présence de 20 µCi de [³⁵S]-méthionine, dans un mélange réactionnel de 50 µL.

b) Expression des protéines GST-P6/TAV et GST-A

Les vecteurs d'expression de GST-P6/TAV et GST-A sont exprimés dans la souche BL21(DE3) d'*Escherichia coli*. Les bactéries transformées sont mises en préculture dans 20 mL de milieu LB additionné d'ampicilline (100 µg/mL). La préculture est incubée pendant une nuit, à 37°C et sous agitation. 500 mL de milieu LB additionné du même antibiotique sontensemencés par 10 mL de la préculture, puis mis sous agitation à 37°C pendant 2 h, jusqu'à ce que la culture bactérienne atteigne sa phase exponentielle de croissance (DO à 600 nm de 0,6 environ). L'induction de l'expression des protéines recombinantes se fait par l'addition d'IPTG à une concentration finale de 0,2 mM, et les bactéries sont mises sous agitation à 20°C pendant 4 h. Les bactéries sont ensuite sédimentées par centrifugation à 7 000 X *g* pendant 15 min, puis congelées à -20°C.

c) Purification des protéines en « *batch* »

La présence de l'étiquette GST en position N-terminale de la protéine de fusion permet à la protéine d'intérêt d'interagir fortement avec le glutathion couplé à des billes de Sepharose 4B (GE Healthcare). La technique du « *batch* » dans des microtubes est utilisée. Toutes les étapes d'extraction et de purification des protéines sont réalisées à 4°C.

Les culots bactériens provenant de 500 mL de culture sont repris avec 30 mL de tampon NET (100 mM de NaCl, 0,1 mM d'EDTA, 50 mM de Tris-HCl pH 7,5, 0,5% Nonidet P-40) additionné de 2 mM de DTT et d'un cocktail complet d'inhibiteurs de protéases (Roche). Les bactéries sont lysées par une sonication douce en *pulse*, durant 3 cycles de 4 min, puis elles sont centrifugées à 25 000 X *g* pendant 30 min, afin de sédimenter les protéines insolubles. Le surnageant contenant les protéines solubles est filtré à travers un filtre de 0,45 µm de diamètre de pore, pour enlever les éventuels agrégats. Le surnageant est ensuite mis à incubé en présence de 500 µL de billes de glutathion Sepharose 4B, pendant 12 h à 4°C et sous agitation continue. Les billes sont sédimentées par une brève centrifugation et lavées 3 fois avec 500 µL de tampon NET.

d) GST pull-down

100 µL de billes de glutathion Sepharose 4B ayant fixé les protéines P6 sont mis en présence de 10 µL du mélange réactionnel de transcription et traduction *in vitro*, dans lequel ont été synthétisées les protéines AtXPO1a, pendant 2 h à température ambiante et sous agitation. Les billes sont lavées comme décrit précédemment, et remises en suspension dans 50 µL de tampon de Laemmli (125 mM de Tris-HCl pH 6,8, 20% de glycérol, 10% de β-mercaptoéthanol, 4% de SDS et 0,05% de bleu de bromophénol). Après une dénaturation de 5 min à 95°C et une brève centrifugation, les protéines du surnageant sont séparées par SDS-PAGE et analysées par autoradiographie.

3) Cellules en culture liquide et plantes

a) Cellules de tabac BY-2

Les cellules de tabac (*Nicotiana tabacum* cv. Bright Yellow 2) sont cultivées sous agitation (130 rotations par minute – rpm), à l'obscurité et à une température de 23°C, dans un milieu basal de Murashige-Skoog (pH 5,8 ; référence M0221, Duchefa Biochemie) complété avec du saccharose à 30 g/L, du myo-inositol à 100 mg/L, de la thiamine à 1 mg/L, du KH₂PO₄ à 200 mg/L et de l'acide 2,4-dichlorophénoxyacétique à 0,2 mg/L. Les cellules sont repiquées chaque semaine, par ensemencement de 2 mL de culture dans 100 mL de milieu frais.

b) Cellules d'*Arabidopsis*

Les cellules d'*Arabidopsis thaliana* (écotype Landsberg erecta) sont cultivées sous agitation (130 rpm), à lumière constante et à une température de 23°C, dans un milieu de Murashige-Skoog avec vitamines de Nitch (référence M0256, Duchefa Biochemie) complété avec du saccharose à 30 g/L, de la kinétine à 230 nM et de l'acide α-naphtylèneacétique à 5,4 µM. Les cellules sont repiquées chaque semaine, par ensemencement de 2 mL de culture dans 100 mL de milieu frais.

c) Plantes

Les plants de navets (*Brassica rapa* cv. Tokyo) et d'*Arabidopsis thaliana* sont cultivés sous serres, dans des conditions de jour long (16 h de lumière à 22°C, 8 h d'obscurité à 18°C).

4) Transfection par biolistique

Les cellules de tabac BY-2 et d'*Arabidopsis* destinées à être transfectées sont récoltées en phase exponentielle, trois jours après repiquage. 8 mL de culture sont déposés sur une couche de papier filtre stérile et le milieu en excès est éliminé par filtration sous vide, à l'aide d'un entonnoir Büchner. Le filtre couvert de cellules est ensuite déposé sur du milieu basal de Murashige-Skoog solide contenant 0,8% d'agarose. Les cellules sont bombardées avec des microbilles de tungstène (1,1 µm de diamètre) recouvertes du ou des plasmides d'intérêt. Deux mg billes sont désagrégés et stérilisés en les vortexant pendant 3 min, dans 1 mL d'éthanol absolu. Les billes sont sédimentées par centrifugation (1 min à 52 X g) puis l'éthanol en excès est retiré et les billes sont mises à sécher. 33 µL de glycérol à 50%, 10 µg d'ADN plasmidique, 33 µL de CaCl₂ à 2 M et 15 µL de spermidine à 0,5 M sont successivement ajoutés aux billes, sous agitation constante. Les billes couvertes d'ADN sont incubées avec le mélange pendant 30 min à température ambiante, sédimentées par centrifugation (1 min à 52 X g) et 150 µL d'éthanol à 70% sont ajoutés après retrait du surnageant. Les billes sont de nouveau sédimentées par centrifugation (1 min à 52 X g) et le surnageant est remplacé par 150 µL d'éthanol absolu. Après une dernière étape de centrifugation, les billes sont séchées sous vide puis suspendues dans 30 µL d'éthanol absolu et conservées dans de la glace, jusqu'au moment de leur utilisation. Les cellules sont bombardées par deux fois avec 8 µL de billes en suspension, dans un canon à parti-

cules. La distance de tir est de 110 mm, sous une pression de 7 bars et pendant 25 ms. Après bombardement, les cellules sont incubées à l'obscurité pendant 24 h à 23°C puis observées au microscope confocal à balayage laser (LSM 700, ZEISS), sous un spectre d'excitation de longueur d'onde égale à 488 nm.

5) Transfection de protoplastes au polyéthylène glycol

a) Protoplastes de mésophylles de navet et d'*Arabidopsis*

Des feuilles de plants de navets et d'*Arabidopsis* au stade 6-8 feuilles sont récoltées, dilacées finement puis mises à digérer dans 20 mL d'un milieu préparé de la façon suivante : une solution composée de 20 mM d'acide 2-morpholino éthanesulfonique monohydraté (MES, pH 5,7), 0,4 M de mannitol et 20 mM de KCl est préchauffée à 70 °C, pendant de 3 à 5 min. 1,5% de cellulase R10 et 0,4% de macérozyme R10 (Yakult Pharmaceutical) sont dissous pendant environ 30 min sous agitation légère. Le milieu est ensuite chauffé à 55°C pendant 10 min, refroidi à température ambiante, puis complétement avec 10 mM de CaCl₂ et 0,1% de sérum d'albumine bovine, et filtré à travers une membrane de 0,45 µm de diamètre de pore. La digestion se fait sans agitation pendant 1 h puis sous agitation (45 rpm) pendant 2 h, à 25°C. Les débris non-digérés sont séparés par filtration sur une double couche de Miracloth (Millipore) et les résidus d'enzymes sont éliminés par trois étapes de lavage successives lors desquelles les protoplastes sont sédimentés par centrifugation (3 min à 100 X g, dans une centrifugeuse à rotor libre et sans frein) et suspendus dans 20 mL de mannitol à 0,5 M. Les protoplastes sont lavés une dernière fois dans du mannitol à 0,5 M supplémenté de 1mM de CaCl₂, comptés à l'aide d'une cellule de numération Fuchs-Rosenthal puis finalement suspendus dans le même milieu, à une concentration de $2,5 \times 10^6$ cellules par mL. Dans un tube de 13 mL en polypropylène (SARSTEDT), 400 µL de protoplastes sont mélangés à 50 µL de solution T (0,5 M de mannitol, 40 mM de CaCl₂ et 10 mM de MES, pH 5,8) contenant 47,5 µg du plasmide d'intérêt et 2,5 µg de pCKeGFP, un plasmide rapporteur codant pour l'eGFP. 450 µL d'une solution de polyéthylène glycol 4000 (Fluka) à 30% (protoplastes d'*Arabidopsis*) ou 40 % (protoplastes de navet) – préparée 1h avant transfection dans la solution T puis filtrée à travers une membrane de 0,45 µm de diamètre de pore – sont ajoutés goutte à goutte aux protoplastes et mélangés par une légère agitation. Le tube contenant le mélange de transfection est placé à l'horizontal sur de la glace et incubé pendant 30 min, sous agitation (75 rpm). La transfection est stoppée par ajout de 8 mL de solution T. Les protoplastes et la solution T sont mélangés précautionneusement, par un ou deux retournements, puis le tube est incubé à la verticale dans de la glace, pendant 30 min et sans agitation. Les protoplastes sont sédimentés par centrifugation (3 min à 100 X g, sans frein), lavés une fois dans 3 mL de solution Hi (0,5 M de mannitol, 50 mM de CaCl₂ et 50 mM de glycine, pH 8), suspendus dans 1 mL de milieu basal de Murashige-Skoog (Sigma-Aldrich) complétement avec du saccharose à 30 g/L, et incubés dans un puits d'une plaque de culture (à 6 ou 12 puits), pendant 16 h à 23°C et à l'obscurité. Dans le cas d'un traitement à la leptomycine B (Sigma-Aldrich), le composé est ajouté à une concentration de 50 ng/mL, 1 h après transfection.

b) Protoplastes de cellules d'*Arabidopsis* en culture liquide

Environ 10 à 15 mL d'une culture cellulaire d'*Arabidopsis*, à 10 jours post-repiquage, sont sédimentés (3 min à 100 X g, sans frein) et lavés deux fois dans du MES à 0,5 M. La digestion et la transfection des protoplastes sont réalisées à l'identique de celles des protoplastes de mésophylle de navet et d'*Arabidopsis*.

6) Fractionnement cellulaire

Les protoplastes sont récoltés dans un tube Falcon de 50 mL, sédimentés (3 min à 100 X g, sans frein) et le surnageant est retiré par pipetage. Environ 1,5 mL d'un tampon hypotonique (10 mM

de Tris-HCl pH 8, et 2 mM de MgCl₂) est ajouté et les protoplastes sont vortexés pendant 3 s. Le lysat est réparti dans des microtubes de 2 mL et soumis à une première centrifugation (10 min à 1500 X g, à 4°C). Le surnageant est prélevé et centrifugé une seconde fois (15 min à 15000 X g, à 4°C) tandis que le culot, qui représente la fraction nucléaire, est immédiatement suspendu dans 1 mL de TRI-Reagent (Sigma-Aldrich). À la fin de la deuxième centrifugation, le surnageant constituant la fraction cytoplasmique est prélevé et mélangé à 3 volumes de TRI-Reagent (Sigma-Aldrich). Les ARN sont extraits en suivant le protocole décrit dans la partie « matériels et méthodes » du premier chapitre.

7) Analyse des ARN par northern-blot

a) ARN de haut poids moléculaires

La détection des ARN de haut poids moléculaires se fait suite à leur séparation par électrophorèse sur gel d'agarose (0,8%) - urée (7M) et leur transfert sur membrane de nylon (Amersham Hybond-N+; GE Healthcare), en suivant le protocole décrit dans la partie « matériels et méthodes » du premier chapitre. La détection des différentes formes de l'ARN 35S se fait soit à l'aide d'un oligonucléotide dirigé contre l'ORF III et radiomarké au ³²P (voir le premier chapitre pour le protocole de radiomarquage des oligonucléotides ; séquence : 5'- GCTGCAACAGCTTCTAAGCTTTCT -3'), soit à l'aide de sondes générées aléatoirement par *random priming*, et dirigées contre l'ORF III. Dans ce dernier cas, l'ORF III a été amplifié avec les oligonucléotides 5'-ACGTAAGTGCCACTAGTATGGCTA-ATCTTAA-3' et 5'-AGGGCTGTGGGGCCCCTAAAATTGATTCGGC-3', puis l'amplicon obtenu a été utilisé comme matrice pour la synthèse de sondes marquées au ³²P, à l'aide du kit de marquage Prime-a-Gene (Promega) et selon les instructions du fabricant. La membrane de nylon est pré-hybridée dans du tampon PerfectHyb Plus (Sigma-Aldrich) pendant 15 min à 42°C puis avec les sondes, quel que soit leur type, pendant une nuit à 42°C. Les sondes non-fixées à la membrane sont éliminées par un premier lavage de 30 min à 50°C, dans du tampon SSC 5x - SDS 0,1%, puis par un second lavage de 5 min à 50°C, dans du tampon SSC 2x - SDS 0,1%. Une étape de lavage supplémentaire identique à la précédente est parfois effectuée.

b) ARN de petit poids moléculaires

20 µg d'ARN sont séparés par électrophorèse sur un gel de polyacrylamide (8% - ratio acrylamide/bis-acrylamide à 29/1) - urée (8M), dans du tampon TBE 1x (90 mM de Tris-borate, 2 mM d'EDTA) et sous une tension de 80V. Les ARN sont ensuite électro-transférés sur une membrane de nylon (Amersham Hybond-N+; GE Healthcare) pendant 75 min à 350 mA, dans du TBE 0,5x à 4°C, puis fixés à la membrane par une irradiation aux UVs (1200 mJ dans un Stratalinker 2400, Stratagene). Les oligonucléotides utilisés pour la détection des snoARN sont radiomarkés au ³²P. La séquence de l'oligonucléotide dirigé contre le snoARN U25a d'*Arabidopsis* est : 5'-ACTCAGTCCCTT-AGATGTTTCATAATAATGC-3' ; celle dirigée contre le snoARN 37 de navet est 5'-CAGCCAAGGGTT-GTATCTTCAACTA-3'. L'hybridation et les lavages sont effectués comme décrits précédemment.

8) Quantification des ARN par qRT-PCR

2,5 µg d'ARN de protoplastes sont traités avec 2,5 U de Dnase I (Promega) et 10 U d'inhibiteur de RNase (RiboLock, Thermo Scientific) pendant 50 min à 37°C. Les ARN sont extraits au phénol/chloroforme puis repris dans de l'eau stérile. La rétrotranscription en ADNc est faite avec l'enzyme SuperScript III (Invitrogen), en présence de 5 µM d'hexamères nucléotidiques de séquences aléatoires. Le mélange réactionnel est incubé 10 min à 25°C, puis 50 min à 55°C et 15 min à 70°C. La PCR quantitative est effectuée dans un thermocycleur en temps réel (LightCycler 480, Roche), avec un mélange enzymatique contenant l'ADN polymérase et le SYBR Green I (LightCycler 480 SYBR Green I Master, Roche), et à partir d'ADNc dilué 30 fois. Les oligonucléotides ciblant l'ARN 35S non-épissée de l'ARN 35S sont utilisés à une concentration de 350 nM et ont pour séquences 5'-GTGCTT-

CATCCTCTAATACTGAAA-3' et 5'-GTTGACCCGTAATGCTCATAAT-3'. Les oligonucléotides ciblant l'ensemble des formes de l'ARN 35S sont utilisés à une concentration de 450 nM et ont pour séquences 5'-CTCAACATCGACTGCCAAACCA-3' et 5'-AGTCCGATTTCTGCTGCCC-3'. La température d'hybridation utilisée est de 57°C. Les formes de l'ARN 35S – non-épissé et population totale – sont quantifiées de manière absolue. Une courbe étalon, nécessaire à ce type de quantification, est réalisée à partir de six dilutions du vecteur viral pMD324 décrit dans le chapitre 1.

Conclusion générale et perspectives

1) L'épissage de l'ARN 35S du CaMV dans les faits

L'épissage alternatif de l'ARN 35S et l'export nucléaire des ARN viraux sont deux étapes du cycle infectieux du CaMV qui sont liées, à la fois, dans l'espace – le noyau – et le temps – pendant et après la transcription. Des deux types principaux d'ARN chez le CaMV, l'ARN 35S est le seul à être épissé. La première description de l'épissage alternatif de l'ARN 35S a fait état, chez la souche Cabb-S, de l'existence de quatre sites donneurs et d'un site accepteur A, tous localisés entre l'extrémité 5' de l'ARN 35S et la fin de l'ORF II (Kiss-László et al., 1995). L'analyse détaillée que nous avons menée sur différentes isoformes de l'ARN 35S a révélé l'existence d'autres sites donneurs et accepteurs au sein de cette même région, chez la souche Cabb-S, ainsi que chez notre modèle au laboratoire, la souche Cabb B-JI. En tout, ce sont au minimum 14 et 16 isoformes qui sont respectivement produites par les souches Cabb-S et Cabb B-JI. La plupart des nouveaux sites décrits sont communs aux deux souches, et une analyse *in silico* montre que tous sont conservés parmi les autres souches de CaMV séquencées. Ceci suggère que l'épissage de la partie 5' de l'ARN 35S est un phénomène assez commun chez le CaMV, et qui semble s'étendre à d'autres espèces de *Caulimovirus* comme l'atteste la conservation de certains sites, dont le site accepteur A, chez le FMV ou le LLDV. Beaucoup d'isoformes épissées possèdent toutes les informations nécessaires pour servir de matrice à la transcriptase inverse et effectivement, des génomes viraux contenant des délétions dues à l'épissage sont détectés chez les plantes infectées. Ces génomes participent probablement à la production d'ARN dits *spliced-like*, qui sont identiques aux véritables ARN épissés et peuvent être encapsidés comme l'ADN viral complet.

Chez les virus comme chez leurs hôtes, l'épissage alternatif est une stratégie de régulation de l'expression génétique qui contribue, entre autres, à augmenter la complexité du protéome. Chez le CaMV, la principale fonction de l'épissage serait de réguler négativement l'expression de l'ORF II codant pour la protéine P2 (Froissart et al., 2004). Cette protéine, qui est uniquement impliquée dans la transmission de plante à plante par le puceron vecteur, n'est pas nécessaire pour la viabilité du virus *in planta* puisque certaines souches naturelles dépourvues d'une partie de l'ORF II, comme la souche CM4-184 (Howarth et al., 1981), sont parfaitement infectieuses. La nécessité de réguler négativement l'expression de P2 s'expliquerait par le fait qu'un excès de la protéine pourrait bloquer le cycle viral. La toxicité potentielle de la protéine P2 pourrait notamment être liée à son affinité pour le réseau microtubulaire (Froissart et al., 2004). Naturellement, cette dérégulation est associée à celle de l'expression de la protéine de mouvement P1, codée par l'ORF I. De manière concomitante, l'épissage stimulerait l'expression des ORF III et IV en réduisant le nombre d'ORF présent en amont de l'ORF III (Kiss-László et al., 1995). Les isoformes épissées possèdent des ORF de fusion issus, notamment, de la fusion, en phase, de l'extrémité 5' de l'ORF I avec l'extrémité 3' de l'ORF II. Les protéines de fusion P1P2 correspondantes interagissent *in vitro* avec les protéines virales P2 et P3. Pourtant, elles ne sont pas retrouvées dans les plantes infectées par le CaMV et elles ne sont que très rarement détectées lors d'expériences d'expression transitoire *in planta*, ce qui suggère qu'elles sont peu exprimées et/ou instables. Par ailleurs, l'inactivation des sites d'épissage, de manière à abolir la synthèse des protéines P1P2 chez Cabb B-JI, ne provoque ni perte d'infectiosité, ni délai dans l'apparition des symptômes. L'ensemble de ces observations plaide en défaveur d'un rôle des protéines P1P2 dans le cycle infectieux du CaMV. La formation quasi systématique d'ORF de fusion, chez les ARN 35S épissés, n'a donc apparemment pas pour fonction d'enrichir le protéome du CaMV, mais plutôt d'empêcher une prolifération d'ORF suite à l'épissage, dont la présence pourrait créer un contexte défavorable à la réinitiation traductionnelle dirigée par P6/TAV.

2) Rôles et la régulation de l'épissage chez le CaMV, de nombreuses interrogations subsistent

La complexité du schéma d'épissage chez le CaMV rappelle celui décrit pour le virus de l'hépatite B, pour lequel aucune explication tangible n'a pu être fournie quant à son rôle dans la pa-

thogénèse, si ce n'est que les ARN épissés s'accumulent plus fréquemment lorsque les infections chroniques aboutissent à des hépatocarcinomes cellulaires (Bayliss et al., 2013; G. H. Lee et al., 2008). Chez le CaMV, la sélection des différents sites donneurs semble être un processus très polyvalent puisque l'inactivation d'un ou plusieurs sites conduit le splicéosome à se reporter sur des sites cryptiques et/ou moins utilisés. De même, l'inactivation partielle ou totale du site accepteur A conduit à l'utilisation de sites accepteurs situés à proximité. Cette observation est en complète contradiction avec les conclusions de travaux antérieurs à ce doctorat où l'inactivation du site A était décrite comme létale pour le virus (Froissart et al., 2004; Kiss-László et al., 1995), mais paradoxalement, elle ne contredit pas les conclusions qui ont été faites sur l'importance de l'épissage dans le cycle infectieux et son rôle dans l'expression de l'ORF II, étant donné qu'aucune des mutations que nous avons étudiées n'a provoqué d'arrêt total de l'épissage de l'ARN 35S ni de suraccumulation de P2 dans les cellules infectées. Ce dernier point semble d'ailleurs indiquer le maintien d'un taux d'épissage constant dans la population globale d'ARN 35S, malgré les mutations présentes. Il apparaît plus que jamais nécessaire d'introduire dans l'ARN 35S des mutations réellement capables d'abolir l'épissage, pour trancher définitivement ces différents points. De plus, d'autres questions relatives aux fonctions de l'épissage au cours du cycle infectieux du CaMV pourraient aussi être étudiées : l'épissage est-il bénéfique pour l'efficacité de réplication du virus ? Module-t-il la virulence du virus ? La première question pourrait être facilement abordée en système de protoplastes, en remplaçant les ORF I et II qui ne sont pas nécessaires à la réplication virale, par un ORF unique dépourvu d'intron. *In planta*, cette stratégie paraît plus délicate, vu les fonctions de l'ORF I dans le mouvement du virus. Dans ce cas, les points de branchement putatifs et des séquences régulatrices prédites *in silico* pourraient être des cibles de choix pour l'insertion de mutations.

Il convient d'admettre que nous n'avons pas d'explication fonctionnelle quant au rôle des multiples sites d'épissage dans le cycle infectieux du CaMV puisqu'une seule isoforme épissée suffit, en théorie, à réguler l'expression des ORF II, III et IV. Il se peut que l'épissage ait d'autres fonctions que celles énoncées précédemment, telles qu'une stimulation de la transcription de l'ARN 35S (Damgaard et al., 2008) ou sa protection vis-à-vis de la dégradation par le *RNA silencing* (Christie et al., 2011). La présence de sites donneurs d'épissage dans la région *leader* de l'ARN 35S, à proximité du signal de polyadénylation, pourrait aussi être impliquée dans la suppression d'événements de polyadénylation prématurés, comme cela a été démontré chez différents virus et pour certains ARNm cellulaires (Kaida et al., 2010; Schrom et al., 2013; West, 2012) ; cette hypothèse est actuellement en cours d'étude au laboratoire. Une autre question restant sans réponse concerne les mécanismes de régulation de l'épissage chez le CaMV. En effet, il est clair que l'épissage alternatif de l'ARN 35S doit être régulé au cours de temps, si on se réfère aux travaux de Kobayashi et al. (1998) sur la cinétique d'apparition des protéines du CaMV, en protoplastes. La protéine P1 s'accumule précocement – comme P6/TAV et P5 –, suivie des protéines P3 et P4 alors que P2 apparaît tardivement au cours du cycle infectieux, ce qui signifie que l'intégrité de l'ARN 35S est maintenue en début et en fin du cycle viral. L'hypothèse d'une implication de protéines virales telles que P6/TAV, P3 et/ou P2 dans cette régulation est un axe de recherche qu'il serait intéressant de développer d'autant plus que certains résultats – la non-détection, par northern-blot, de formes épissées si l'ORF II est tronqué en 5' par un codon stop – suggèrent que la région C-terminale de P2 pourrait être impliquée dans ce processus.

3) L'export nucléaire des ARN viraux, un sujet encore aux prémices

Si l'épissage apparaît nécessaire à la régulation de l'expression génétique chez le CaMV, les deux types d'ARN 35S – épissés et non-épissé – doivent être pris en charge par la machinerie d'export afin de rejoindre le cytoplasme où ils seront traduits et/ou rétrotranscrits. La forme non-épissée de l'ARN 35S est d'ailleurs indispensable au virus puisqu'elle est la seule qui permette l'expression des ORF I et II, ainsi que la rétrotranscription en ADN viral complet. L'export nucléaire des ARNm est un processus qui est relativement bien décrit chez la levure et les mammifères, mais très peu connu chez les plantes. Une raison à cette disparité est que les facteurs de transport principaux des ARNm sont conservés et bien caractérisés chez les mammifères et la levure alors qu'on ne

leur connaît pas d'homologues chez les végétaux. Chez les mammifères, l'export nucléaire est la suite logique de l'épissage, alors que l'export d'ARN contenant encore un ou plusieurs introns est généralement réprimé, étant donné que la présence de ces derniers est souvent considérée comme résultant d'un défaut d'épissage. Les rétrovirus sont amenés, de par leur stratégie d'expression, à détourner les mécanismes de surveillance entourant le processus d'export nucléaire afin d'exporter, vers le cytoplasme, des transcrits partiellement et non-épissés. Si les moyens de détournement mis en place varient d'un virus à l'autre, tous reposent sur l'utilisation de séquences virales en *cis* et de facteurs protéiques, parfois viraux, en *trans*. Les questions afférentes à l'export des ARN du CaMV sont encore sans réponses mais un système transitoire destiné à l'étude de ce processus est au stade terminal de développement. Ce système est basé sur la quantification de la proportion représentée par la forme non-épissée, au sein des populations nucléaire et cytoplasmique d'ARN 35S ; cette valeur est comparée entre différentes conditions. Notre système, une fois au point, permettra d'identifier la ou les séquences virales impliquées en *cis*, en comparant les effets de différentes délétions dans l'ARN 35S. L'implication éventuelle d'une ou de plusieurs protéines virales dans l'export de l'ARN 35S sera également évaluée. Le candidat le plus crédible est la protéine P6/TAV car celle-ci peut se lier aux ARN et probablement interagir avec AtXPO1a, le facteur d'export des protéines chez les plantes et homologue de CRM1.

L'élucidation du mécanisme d'export nucléaire de l'ARN 35S constitue, au même titre que la régulation de l'épissage alternatif, un défi incontournable pour compléter nos connaissances sur le cycle infectieux du CaMV. Étonnamment, les interactions entre le CaMV et le noyau des plantes hôtes ont suscité jusqu'à ce jour très peu d'intérêt, alors qu'elles conditionnent le bon déroulement du cycle de réplication et, plus généralement, la pathogenèse du CaMV. Il est vraisemblable que l'étude de ces processus permettra, au-delà du CaMV, d'apporter des informations intéressantes sur l'export nucléaire et l'épissage des ARN chez les plantes.

Bibliographie

- Albrecht, H., A. Geldreich, J. M. de Murcia, D. Kirchherr, J. M. Mesnard, et G. Lebeurier. (1988). Cauliflower mosaic virus gene I product detected in a cell-wall-enriched fraction. *Virology* 163, 503-508.
- Armour, S. L., U. Melcher, T. P. Pirone, D. J. Lyttle, et R. C. Essenberg. (1983). Helper component for aphid transmission encoded by region II of cauliflower mosaic virus DNA. *Virology* 129, 25-30.
- Ashe, M. P., P. Griffin, W. James, et N. J. Proudfoot. (1995). Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes & Development* 9, 3008-3025.
- Bayliss, J., L. Lim, A. J. V. Thompson, P. Desmond, P. Angus, S. Locarnini, et P. A. Revill. (2013). Hepatitis B virus splicing is enhanced prior to development of hepatocellular carcinoma. *Journal of Hepatology* 59, 1022-1028.
- Berget, S. M., C. Moore, et P. A. Sharp. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 74, 3171-3175.
- Blanvillain, R., L. C. Boavida, S. McCormick, et D. W. Ow. (2008). Exportin1 genes are essential for development and function of the gametophytes in *Arabidopsis thaliana*. *Genetics* 180, 1493-1500.
- Blevins, T., R. Rajeswaran, M. Aregger, B. K. Borah, M. Schepetilnikov, L. Baerlocher, et al. (2011). Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. *Nucleic Acids Research* 39, 5003-5014.
- Blevins, T., R. Rajeswaran, P. V. Shivaprasad, D. Beknazariants, A. Si-Ammour, H.-S. Park, et al. (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Research* 34, 6233-6246.
- Blobel, G. (1985). Gene gating: a hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 82, 8527-8529.
- Bodem, J., T. Schied, R. Gabriel, M. Rammling, et A. Rethwilm. (2011). Foamy virus nuclear RNA export is distinct from that of other retroviruses. *Journal of Virology* 85, 2333-2341.
- Bogerd, H. P., R. A. Fridell, R. E. Benson, J. Hua, et B. R. Cullen. (1996). Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization-selection assay. *Molecular and Cellular Biology* 16, 4207-4214.
- Brown, J. W. S., D. F. Marshall, et M. Echeverria. (2008). Intronic noncoding RNAs and splicing. *Trends in Plant Science* 13, 335-342.
- Busch, A. et K. J. Hertel. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdisciplinary Reviews. RNA* 3, 1-12.
- Calvin, K. et H. Li. (2008). RNA-splicing endonuclease structure and function. *Cellular and Molecular Life Sciences: CMLS* 65, 1176-1185.
- Cecchini, E., N. S. Al-Kaff, A. Bannister, M. E. Giannakou, D. G. McCallum, A. J. Maule, et al. (1998). Pathogenic interactions between variants of cauliflower mosaic virus and *Arabidopsis thaliana*. *Journal of Experimental Botany* 49, 731-737.
- Cerritelli, S. M., O. Y. Fedoroff, B. R. Reid, et R. J. Crouch. (1998). A common 40 amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs. *Nucleic Acids Research* 26, 1834-1840.

- Chang, Y.-F., J. S. Imam, et M. F. Wilkinson. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annual Review of Biochemistry* 76, 51-74.
- Chen, W.-N., J.-Y. Chen, W.-S. Lin, J.-Y. Lin, et X. Lin. (2010). Hepatitis B doubly spliced protein, generated by a 2.2 kb doubly spliced hepatitis B virus RNA, is a pleiotropic activator protein mediating its effects via activator protein-1- and CCAAT/enhancer-binding protein-binding sites. *The Journal of General Virology* 91, 2592-2600.
- Chow, L. T., R. E. Gelinas, T. R. Broker, et R. J. Roberts. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1-8.
- Christie, M., L. J. Croft, et B. J. Carroll. (2011). Intron splicing suppresses RNA silencing in Arabidopsis. *The Plant Journal: For Cell and Molecular Biology* 68, 159-167.
- Coyle, J. H., Y.-C. Bor, D. Rekosh, et M.-L. Hammariskjold. (2011). The Tpr protein regulates export of mRNAs with retained introns that traffic through the Nxf1 pathway. *RNA (New York, N.Y.)* 17, 1344-1356.
- Culjkovic-Kraljacic, B. et K. L. B. Borden. (2013). Aiding and abetting cancer: mRNA export and the nuclear pore. *Trends in Cell Biology* 23, 328-335.
- Damgaard, C. K., S. Kahns, S. Lykke-Andersen, A. L. Nielsen, T. H. Jensen, et J. Kjems. (2008). A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Molecular Cell* 29, 271-278.
- D'Angelo, M. A., D. J. Anderson, E. Richard, et M. W. Hetzer. (2006). Nuclear pores form de novo from both sides of the nuclear envelope. *Science (New York, N.Y.)* 312, 440-443.
- Dautel, S., T. Guidasci, M. Pique, J. L. Mougeot, G. Lebeurier, P. Yot, et J. M. Mesnard. (1994). The full-length product of cauliflower mosaic virus open reading frame III is associated with the viral particle. *Virology* 202, 1043-1045.
- De Conti, L., M. Baralle, et E. Buratti. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews. RNA* 4, 49-60.
- De la Mata, M. et A. R. Kornblihtt. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nature Structural & Molecular Biology* 13, 973-980.
- Delseny, M. et R. Hull. (1983). Isolation and characterization of faithful and altered clones of the genomes of cauliflower mosaic virus isolates Cabb B-JI, CM4-184, and Bari I. *Plasmid* 9, 31-41.
- De Tapia, M., A. Himmelbach, et T. Hohn. (1993). Molecular dissection of the cauliflower mosaic virus translation transactivator. *The EMBO Journal* 12, 3305-3314.
- Dominguez, D. I., T. Hohn, et W. Schmidt-Puchta. (1996). Cellular proteins bind to multiple sites of the leader region of cauliflower mosaic virus 35S RNA. *Virology* 226, 374-383.
- Dong, X., A. Biswas, K. E. Süel, L. K. Jackson, R. Martinez, H. Gu, et Y. M. Chook. (2009). Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* 458, 1136-1141.
- Dubois, J., O. Terrier, et M. Rosa-Calatrava. (2014). Influenza viruses and mRNA splicing: doing more with less. *mBio* 5, e00070-00014.
- El Yacoubi, B., M. Bailly, et V. de Crécy-Lagard. (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annual Review of Genetics* 46, 69-95.
- Eriksson, M., W. T. Brown, L. B. Gordon, M. W. Glynn, J. Singer, L. Scott, et al. (2003). Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* 423, 293-298.

- Estmer Nilsson, C., S. Petersen-Mahrt, C. Durot, R. Shtrichman, A. R. Krainer, T. Kleinberger, et G. Akusjärvi. (2001). The adenovirus E4-ORF4 splicing enhancer protein interacts with a subset of phosphorylated SR proteins. *The EMBO Journal* 20, 864-871.
- Fedorova, O. et N. Zingler. (2007). Group II introns: structure, folding and splicing mechanism. *Biological Chemistry* 388, 665-678.
- Feng, S., C. Martinez, G. Gusmaroli, Y. Wang, J. Zhou, F. Wang, et al. (2008). Coordinated regulation of Arabidopsis thaliana development by light and gibberellins. *Nature* 451, 475-479.
- Fernandez-Martinez, J. et M. P. Rout. (2009). Nuclear pore complex biogenesis. *Current Opinion in Cell Biology* 21, 603-612.
- Fica, S. M., M. A. Mefford, J. A. Piccirilli, et J. P. Staley. (2014). Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nature Structural & Molecular Biology* 21, 464-471.
- Fica, S. M., N. Tuttle, T. Novak, N.-S. Li, J. Lu, P. Koodathingal, et al. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature* 503, 229-234.
- Filipowicz, W., T. Kiss, C. Marshallsay, et F. Waibel. (1990). U-snRNA genes, U-snRNAs and U-snRNPs of higher plants. *Molecular Biology Reports* 14, 125-129.
- Fischer, U., S. Meyer, M. Teufel, C. Heckel, R. Lührmann, et G. Rautmann. (1994). Evidence that HIV-1 Rev directly promotes the nuclear export of unspliced RNA. *The EMBO Journal* 13, 4105-4112.
- Fiserova, J., E. Kiseleva, et M. W. Goldberg. (2009). Nuclear envelope and nuclear pore complex structure and organization in tobacco BY-2 cells. *The Plant Journal: For Cell and Molecular Biology* 59, 243-255.
- Flicek, P., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, et al. (2014). Ensembl 2014. *Nucleic Acids Research* 42, D749-755.
- Franck, A., H. Guilley, G. Jonard, K. Richards, et L. Hirth. (1980). Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* 21, 285-294.
- Froissart, R., M. Uzest, V. Ruiz-Ferrer, M. Drucker, E. Hébrard, T. Hohn, et S. Blanc. (2004). Splicing of Cauliflower mosaic virus 35S RNA serves to downregulate a toxic gene product. *The Journal of General Virology* 85, 2719-2726.
- Fütterer, J., I. Potrykus, M. P. Valles Brau, I. Dasgupta, R. Hull, et T. Hohn. (1994). Splicing in a plant pararetrovirus. *Virology* 198, 663-670.
- Gabut, M., P. Samavarchi-Tehrani, X. Wang, V. Slobodeniuc, D. O'Hanlon, H.-K. Sung, et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* 147, 132-146.
- Gaouar, O. et H. Germain. (2013). mRNA export: threading the needle. *Frontiers in Plant Science* 4, 59.
- García-Oliver, E., V. García-Molinero, et S. Rodríguez-Navarro. (2012). mRNA export and gene expression: the SAGA-TREX-2 connection. *Biochimica Et Biophysica Acta* 1819, 555-565.
- Germain, H., N. Qu, Y. T. Cheng, E. Lee, Y. Huang, O. X. Dong, et al. (2010). MOS11: a new component in the mRNA export pathway. *PLoS Genetics* 6, e1001250.
- Gómez Acuña, L. I., A. Fiszbein, M. Alló, I. E. Schor, et A. R. Kornblihtt. (2013). Connections between chromatin signatures and splicing. *Wiley Interdisciplinary Reviews. RNA* 4, 77-91.

- Goodall, G. J. et W. Filipowicz. (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* 58, 473-483.
- Grossman, E., O. Medalia, et M. Zwerger. (2012). Functional architecture of the nuclear pore complex. *Annual Review of Biophysics* 41, 557-584.
- Grünwald, D., R. H. Singer, et M. Rout. (2011). Nuclear export dynamics of RNA-protein complexes. *Nature* 475, 333-341.
- Grüter, P., C. Tabernero, C. von Kobbe, C. Schmitt, C. Saavedra, A. Bachi, et al. (1998). TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Molecular Cell* 1, 649-659.
- Guilley, H., K. E. Richards, et G. Jonard. (1983). Observations concerning the discontinuous DNAs of cauliflower mosaic virus. *The EMBO Journal* 2, 277-282.
- Haasen, D., C. Köhler, G. Neuhaus, et T. Merkle. (1999). Nuclear export of proteins in plants: AtXPO1 is the export receptor for leucine-rich nuclear export signals in *Arabidopsis thaliana*. *The Plant Journal: For Cell and Molecular Biology* 20, 695-705.
- Haas, G., J. Azevedo, G. Moissiard, A. Geldreich, C. Himber, M. Bureau, et al. (2008). Nuclear import of CaMV P6 is required for infection and suppression of the RNA silencing factor DRB4. *The EMBO Journal* 27, 2102-2112.
- Haas, M., A. Geldreich, M. Bureau, L. Dupuis, V. Leh, G. Vetter, et al. (2005). The open reading frame VI product of Cauliflower mosaic virus is a nucleocytoplasmic protein: its N terminus mediates its nuclear export and formation of electron-dense viroplasms. *The Plant Cell* 17, 927-943.
- Han, J., J. Xiong, D. Wang, et X.-D. Fu. (2011). Pre-mRNA splicing: where and when in the nucleus. *Trends in Cell Biology* 21, 336-343.
- Hapiak, M., Y. Li, K. Agama, S. Swade, G. Okenka, J. Falk, et al. (2008). Cauliflower mosaic virus gene VI product N-terminus contains regions involved in resistance-breakage, self-association and interactions with movement protein. *Virus Research* 138, 119-129.
- Hattori, D., Y. Chen, B. J. Matthews, L. Salwinski, C. Sabatti, W. B. Grueber, et S. L. Zipursky. (2009). Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature* 461, 644-648.
- Haugen, P., D. M. Simon, et D. Bhattacharya. (2005). The natural history of group I introns. *Trends in Genetics: TIG* 21, 111-119.
- Hébrard, E., M. Drucker, D. Leclerc, T. Hohn, M. Uzest, R. Froissart, et al. (2001). Biochemical characterization of the helper component of Cauliflower mosaic virus. *Journal of Virology* 75, 8538-8546.
- Hemmings-Mieszczak, M., G. Steger, et T. Hohn. (1997). Alternative structures of the cauliflower mosaic virus 35 S RNA leader: implications for viral expression and replication. *Journal of Molecular Biology* 267, 1075-1088.
- Hidaka, M., J. Inoue, M. Yoshida, et M. Seiki. (1988). Post-transcriptional regulator (rex) of HTLV-1 initiates expression of viral structural proteins but suppresses expression of regulatory proteins. *The EMBO Journal* 7, 519-523.
- Hiraguri, A., R. Itoh, N. Kondo, Y. Nomura, D. Aizawa, Y. Murai, et al. (2005). Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant Molecular Biology* 57, 173-188.

- Hirochika, H., H. Takatsuji, A. Ubasawa, et J. E. Ikeda. (1985). Site-specific deletion in cauliflower mosaic virus DNA: possible involvement of RNA splicing and reverse transcription. *The EMBO Journal* 4, 1673-1680.
- Hopper, A. K. (2013). Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast *Saccharomyces cerevisiae*. *Genetics* 194, 43-67.
- Howarth, A. J., R. C. Gardner, J. Messing, et R. J. Shepherd. (1981). Nucleotide sequence of naturally occurring deletion mutants of cauliflower mosaic virus. *Virology* 112, 678-685.
- Huang, Y., R. Gattoni, J. Stévenin, et J. A. Steitz. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Molecular Cell* 11, 837-843.
- Huang, Y. et J. A. Steitz. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Molecular Cell* 7, 899-905.
- Hutchinson, E. C. et E. Fodor. (2013). Transport of the influenza virus genome from nucleus to nucleus. *Viruses* 5, 2424-2446.
- Hutten, S. et R. H. Kehlenbach. (2007). CRM1-mediated nuclear export: to the pore and beyond. *Trends in Cell Biology* 17, 193-201.
- Iida, K., M. Seki, T. Sakurai, M. Satou, K. Akiyama, T. Toyoda, et al. (2004). Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. *Nucleic Acids Research* 32, 5096-5103.
- Jacquot, E., A. Geldreich, M. Keller, et P. Yot. (1998). Mapping regions of the cauliflower mosaic virus ORF III product required for infectivity. *Virology* 242, 395-402.
- Jacquot, E., L. S. Hagen, M. Jacquemond, et P. Yot. (1996). The open reading frame 2 product of cacao swollen shoot badnavirus is a nucleic acid-binding protein. *Virology* 225, 191-195.
- Johansson, C. et S. Schwartz. (2013). Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nature Reviews. Microbiology* 11, 239-251.
- Johnson, S. A., G. Cubberley, et D. L. Bentley. (2009). Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. *Molecular Cell* 33, 215-226.
- Kaida, D., M. G. Berg, I. Younis, M. Kasim, L. N. Singh, L. Wan, et G. Dreyfuss. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664-668.
- Kammel, C., M. Thomaier, B. B. Sørensen, T. Schubert, G. Längst, M. Grasser, et K. D. Grasser. (2013). *Arabidopsis* DEAD-box RNA helicase UAP56 interacts with both RNA and DNA as well as with mRNA export factors. *PloS One* 8, e60644.
- Kanopka, A., O. Mühlemann, S. Petersen-Mahrt, C. Estmer, C. Ohrmalm, et G. Akusjärvi. (1998). Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* 393, 185-187.
- Karsies, A., T. Hohn, et D. Leclerc. (2001). Degradation signals within both terminal domains of the cauliflower mosaic virus capsid protein precursor. *The Plant Journal: For Cell and Molecular Biology* 27, 335-343.
- Katahira, J. (2012). mRNA export and the TREX complex. *Biochimica Et Biophysica Acta* 1819, 507-513.
- Kellis, M., N. Patterson, M. Endrizzi, B. Birren, et E. S. Lander. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241-254.

- Kelly, S. M. et A. H. Corbett. (2009). Messenger RNA export from the nucleus: a series of molecular wardrobe changes. *Traffic (Copenhagen, Denmark)* 10, 1199-1208.
- Keren, H., G. Lev-Maor, et G. Ast. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews. Genetics* 11, 345-355.
- Khelifa, M., D. Massé, S. Blanc, et M. Drucker. (2010). Evaluation of the minimal replication time of Cauliflower mosaic virus in different hosts. *Virology* 396, 238-245.
- Kimura, T., I. Hashimoto, T. Nagase, et J.-I. Fujisawa. (2004). CRM1-dependent, but not ARE-mediated, nuclear export of IFN- α 1 mRNA. *Journal of Cell Science* 117, 2259-2270.
- Kimura, T., I. Hashimoto, M. Nishizawa, S. Ito, et H. Yamada. (2010). Novel cis-active structures in the coding region mediate CRM1-dependent nuclear export of IFN- α 1 mRNA. *Medical Molecular Morphology* 43, 145-157.
- Kiss-László, Z., S. Blanc, et T. Hohn. (1995). Splicing of cauliflower mosaic virus 35S RNA is essential for viral infectivity. *The EMBO Journal* 14, 3552-3562.
- Kiss, T. (2004). Biogenesis of small nuclear RNPs. *Journal of Cell Science* 117, 5949-5951.
- Kobayashi, K., S. Tsuge, H. Nakayashiki, K. Mise, et I. Furusawa. (1998). Requirement of cauliflower mosaic virus open reading frame VI product for viral gene expression and multiplication in turnip protoplasts. *Microbiology and Immunology* 42, 377-386.
- Kobayashi, K., S. Tsuge, L. Stavelone, et T. Hohn. (2002). The cauliflower mosaic virus virion-associated protein is dispensable for viral replication in single cells. *Journal of Virology* 76, 9457-9464.
- Köhler, A. et E. Hurt. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nature Reviews. Molecular Cell Biology* 8, 761-773.
- Kornblihtt, A. R., I. E. Schor, M. Alló, G. Dujardin, E. Petrillo, et M. J. Muñoz. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature Reviews. Molecular Cell Biology* 14, 153-165.
- Kupfer, D. M., S. D. Drabenstot, K. L. Buchanan, H. Lai, H. Zhu, D. W. Dyer, et al. (2004). Introns and splicing elements of five diverse fungi. *Eukaryotic Cell* 3, 1088-1100.
- Laird, J., C. McNally, C. Carr, S. Doddiah, G. Yates, E. Chrysanthou, et al. (2013). Identification of the domains of cauliflower mosaic virus protein P6 responsible for suppression of RNA silencing and salicylic acid signalling. *The Journal of General Virology* 94, 2777-2789.
- Lambowitz, A. M. et S. Zimmerly. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harbor Perspectives in Biology* 3, a003616.
- Langemeier, J., M. Radtke, et J. Böhne. (2013). U1 snRNP-mediated poly(A) site suppression: beneficial and deleterious for mRNA fate. *RNA Biology* 10, 180-184.
- Leblanc, J., J. Weil, et K. Beemon. (2013). Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. *Wiley Interdisciplinary Reviews. RNA* 4, 567-580.
- Leclerc, D., L. Stavelone, E. Meier, O. Guerra-Peraza, E. Herzog, et T. Hohn. (2001). The product of ORF III in cauliflower mosaic virus interacts with the viral coat protein through its C-terminal proline rich domain. *Virus Genes* 22, 159-165.
- Lee, G. H., S. Wasser, et S. G. Lim. (2008). Hepatitis B pregenomic RNA splicing--the products, the regulatory mechanisms and its biological significance. *Virus Research* 136, 1-7.

- Lee, S. J., C. Jiko, E. Yamashita, et T. Tsukihara. (2011). Selective nuclear export mechanism of small RNAs. *Current Opinion in Structural Biology* 21, 101-108.
- Lehto, K., P. Bubrick, et W. O. Dawson. (1990). Time course of TMV 30K protein accumulation in intact leaves. *Virology* 174, 290-293.
- Leh, V., E. Jacquot, A. Geldreich, M. Haas, S. Blanc, M. Keller, et P. Yot. (2001). Interaction between the open reading frame III product and the coat protein is required for transmission of cauliflower mosaic virus by aphids. *Journal of Virology* 75, 100-106.
- Leh, V., E. Jacquot, A. Geldreich, T. Hermann, D. Leclerc, M. Cerutti, et al. (1999). Aphid transmission of cauliflower mosaic virus requires the viral PIII protein. *The EMBO Journal* 18, 7077-7085.
- Leh, V., P. Yot, et M. Keller. (2000). The cauliflower mosaic virus translational transactivator interacts with the 60S ribosomal subunit protein L18 of *Arabidopsis thaliana*. *Virology* 266, 1-7.
- Leppard, K. N. (2008). Adenoviruses: Molecular Biology. In B. W. J. Mahy et M. H. V. V. Regenmortel (Eds.), *Encyclopedia of Virology (Third Edition)*, p. 17-23. Oxford: Academic Press.
- Luco, R. F., M. Allo, I. E. Schor, A. R. Kornblihtt, et T. Misteli. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16-26.
- Luco, R. F., Q. Pan, K. Tominaga, B. J. Blencowe, O. M. Pereira-Smith, et T. Misteli. (2010). Regulation of alternative splicing by histone modifications. *Science* 327, 996-1000.
- Luo, M. L., Z. Zhou, K. Magni, C. Christoforides, J. Rappsilber, M. Mann, et R. Reed. (2001). Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* 413, 644-647.
- Lutz, L., G. Raikhy, et S. M. Leisner. (2012). Cauliflower mosaic virus major inclusion body protein interacts with the aphid transmission factor, the virion-associated protein, and gene VII product. *Virus Research* 170, 150-153.
- Malim, M. H., L. S. Tiley, D. F. McCarn, J. R. Rusche, J. Hauber, et B. R. Cullen. (1990). HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* 60, 675-683.
- Manley, J. L. et A. R. Krainer. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes & Development* 24, 1073-1074.
- Marquez, Y., J. W. S. Brown, C. Simpson, A. Barta, et M. Kalyna. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Research* 22, 1184-1195.
- Masuda, S., R. Das, H. Cheng, E. Hurt, N. Dorman, et R. Reed. (2005). Recruitment of the human TREX complex to mRNA during splicing. *Genes & Development* 19, 1512-1517.
- Matera, A. G., R. M. Terns, et M. P. Terns. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews. Molecular Cell Biology* 8, 209-220.
- Matera, A. G. et Z. Wang. (2014). A day in the life of the spliceosome. *Nature Reviews. Molecular Cell Biology* 15, 108-121.
- Meier, I. (2012). mRNA export and sumoylation-Lessons from plants. *Biochimica Et Biophysica Acta* 1819, 531-537.
- Meinel, D. M., C. Burkert-Kautzsch, A. Kieser, E. O'Duibhir, M. Siebert, A. Mayer, et al. (2013). Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. *PLoS Genetics* 9, e1003914.

- Melamud, E. et J. Moul. (2009a). Stochastic noise in splicing machinery. *Nucleic Acids Research* 37, 4873-4886.
- Melamud, E. et J. Moul. (2009b). Structural implication of splicing stochastics. *Nucleic Acids Research* 37, 4862-4872.
- Merkle, T. (2011). Nucleo-cytoplasmic transport of proteins and RNA in plants. *Plant Cell Reports* 30, 153-176.
- Mevel, S. et C. Kerlan. (1990). Caractérisation biologique de différents isolats du virus de la mosaïque du chou fleur (CaMV). 1. Plantes hôtes : Solanées. *Agronomie* 10, 749-758.
- Moissiard, G. et O. Voinnet. (2006). RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. *Proceedings of the National Academy of Sciences of the United States of America* 103, 19593-19598.
- Muñoz, M. J., M. S. Pérez Santangelo, M. P. Paronetto, M. de la Mata, F. Pelisch, S. Boireau, et al. (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* 137, 708-720.
- Nakagawa, T., T. Kurose, T. Hino, K. Tanaka, M. Kawamukai, Y. Niwa, et al. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering* 104, 34-41.
- Ni, J. Z., L. Grate, J. P. Donohue, C. Preston, N. Nobida, G. O'Brien, et al. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Development* 21, 708-718.
- Nilsen, T. W. et B. R. Graveley. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457-463.
- Oeffinger, M. et D. Zenklusen. (2012). To the pore and through the pore: a story of mRNA export kinetics. *Biochimica Et Biophysica Acta* 1819, 494-506.
- Olszewski, N. E. et T. J. Guilfoyle. (1983). Nuclei purified from cauliflower mosaic virus-infected turnip leaves contain subgenomic, covalently closed circular cauliflower mosaic virus DNAs. *Nucleic Acids Research* 11, 8901-8914.
- Ouyang, S., W. Zhu, J. Hamilton, H. Lin, M. Campbell, K. Childs, et al. (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Research* 35, D883-887.
- Palazzo, A. F. et A. Akef. (2012). Nuclear export as a key arbiter of « mRNA identity » in eukaryotes. *Biochimica Et Biophysica Acta* 1819, 566-577.
- Palazzo, A. F., K. Mahadevan, et S. P. Tarnawsky. (2013). ALREX-elements and introns: two identity elements that promote mRNA nuclear export. *Wiley Interdisciplinary Reviews. RNA* 4, 523-533.
- Paquin, B. et D. A. Shub. (2001). Introns: Group I Structure and Function. In eLS. John Wiley & Sons, Ltd.
- Perbal, M. C., C. L. Thomas, et A. J. Maule. (1993). Cauliflower mosaic virus gene I product (P1) forms tubular structures which extend from the surface of infected protoplasts. *Virology* 195, 281-285.
- Pessel-Vivares, L., M. Ferrer, S. Lainé, et M. Mougél. (2014). MLV requires Tap/NXF1-dependent pathway to export its unspliced RNA to the cytoplasm and to express both spliced and unspliced RNAs. *Retrovirology* 11, 21.

- Petrillo, E., M. A. G. Herz, A. Fuchs, D. Reifer, J. Fuller, M. J. Yanovsky, et al. (2014). A chloroplast retrograde signal regulates nuclear alternative splicing. *Science (New York, N.Y.)* 344, 427-430.
- Pfeiffer, P. et T. Hohn. (1983). Involvement of reverse transcription in the replication of cauliflower mosaic virus: a detailed model and test of some aspects. *Cell* 33, 781-789.
- Pooggin, M. M., T. Hohn, et J. Fütterer. (2000). Role of a short open reading frame in ribosome shunt on the cauliflower mosaic virus RNA leader. *The Journal of Biological Chemistry* 275, 17288-17296.
- Popow, J., A. Schleiffer, et J. Martinez. (2012). Diversity and roles of (t)RNA ligases. *Cellular and Molecular Life Sciences: CMLS* 69, 2657-2670.
- Raices, M. et M. A. D'Angelo. (2012). Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. *Nature Reviews. Molecular Cell Biology* 13, 687-699.
- Reddy, A. S. N. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual Review of Plant Biology* 58, 267-294.
- Reddy, A. S. N., Y. Marquez, M. Kalyna, et A. Barta. (2013). Complexity of the alternative splicing landscape in plants. *The Plant Cell* 25, 3657-3683.
- Richards, K. E., H. Guilley, et G. Jonard. (1981). Further characterization of the discontinuities in cauliflower mosaic virus DNA. *FEBS Letters* 134, 67-70.
- Richins, R. D., H. B. Scholthof, et R. J. Shepherd. (1987). Sequence of figwort mosaic virus DNA (caulimovirus group). *Nucleic Acids Research* 15, 8451-8466.
- Roy, S. W. et M. Irimia. (2009). Splicing in the eukaryotic ancestor: form, function and dysfunction. *Trends in Ecology & Evolution* 24, 447-455.
- Sajic, R., K. Lee, K. Asai, D. Sakac, D. R. Branch, C. Upton, et A. Cochrane. (2007). Use of modified U1 snRNAs to inhibit HIV-1 replication. *Nucleic Acids Research* 35, 247-255.
- Sakuma, T., J. I. Davila, J. A. Malcolm, J.-P. A. Kocher, J. M. Tonne, et Y. Ikeda. (2014). Murine leukemia virus uses NXF1 for nuclear export of spliced and unspliced viral transcripts. *Journal of Virology* 88, 4069-4082.
- Sanchez, S. E., E. Petrillo, E. J. Beckwith, X. Zhang, M. L. Rugnone, C. E. Hernando, et al. (2010). A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature* 468, 112-116.
- Sanfaçon, H. et A. Wieczorek. (1992). Analysis of cauliflower mosaic virus RNAs in Brassica species showing a range of susceptibility to infection. *Virology* 190, 30-39.
- Satterly, N., P.-L. Tsai, J. van Deursen, D. R. Nussenzveig, Y. Wang, P. A. Faria, et al. (2007). Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proceedings of the National Academy of Sciences of the United States of America* 104, 1853-1858.
- Scholthof, H. B., F. C. Wu, R. D. Richins, et R. J. Shepherd. (1991). A naturally occurring deletion mutant of figwort mosaic virus (caulimovirus) is generated by RNA splicing. *Virology* 184, 290-298.
- Schrom, E.-M., R. Moschall, M. J. Hartl, H. Weitner, D. Fecher, J. Langemeier, et al. (2013). U1snRNP-mediated suppression of polyadenylation in conjunction with the RNA structure controls poly (A) site selection in foamy viruses. *Retrovirology* 10, 55.

- Schumann, S., B. R. Jackson, B. Baquero-Perez, et A. Whitehouse. (2013). Kaposi's sarcoma-associated herpesvirus ORF57 protein: exploiting all stages of viral mRNA processing. *Viruses* 5, 1901-1923.
- Shabalina, S. A., A. Y. Ogurtsov, A. N. Spiridonov, P. S. Novichkov, N. A. Spiridonov, et E. V. Koonin. (2010). Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. *Molecular Biology and Evolution* 27, 1745-1749.
- Shida, H. (2012). Role of Nucleocytoplasmic RNA Transport during the Life Cycle of Retroviruses. *Frontiers in Microbiology* 3, 179.
- Shivaprasad, P. V., R. Rajeswaran, T. Blevins, J. Schoelz, F. Meins Jr, T. Hohn, et M. M. Pooggin. (2008). The CaMV transactivator/viroplasm interferes with RDR6-dependent trans-acting and secondary siRNA pathways in Arabidopsis. *Nucleic Acids Research* 36, 5896-5909.
- Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7, 539.
- Singh, R. K. et T. A. Cooper. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends in Molecular Medicine* 18, 472-482.
- Soussan, P., F. Garreau, H. Zylberberg, C. Ferray, C. Brechot, et D. Kremsdorf. (2000). In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA. *The Journal of Clinical Investigation* 105, 55-60.
- Staiger, D. et J. W. S. Brown. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *The Plant Cell* 25, 3640-3656.
- Stavolone, L., M. E. Villani, D. Leclerc, et T. Hohn. (2005). A coiled-coil interaction mediates cauliflower mosaic virus cell-to-cell movement. *Proceedings of the National Academy of Sciences of the United States of America* 102, 6219-6224.
- Stewart, M. (2007). Ratcheting mRNA out of the nucleus. *Molecular Cell* 25, 327-330.
- Strambio-De-Castillia, C., M. Niepel, et M. P. Rout. (2010). The nuclear pore complex: bridging nuclear transport and gene regulation. *Nature Reviews. Molecular Cell Biology* 11, 490-501.
- Thomas, C. L. et A. J. Maule. (1995). Identification of structural domains within the cauliflower mosaic virus movement protein by scanning deletion mutagenesis and epitope tagging. *The Plant Cell* 7, 561-572.
- Thomson, E., S. Ferreira-Cerca, et E. Hurt. (2013). Eukaryotic ribosome biogenesis at a glance. *Journal of Cell Science* 126, 4815-4821.
- Törmänen, H., E. Backström, A. Carlsson, et G. Akusjärvi. (2006). L4-33K, an adenovirus-encoded alternative RNA splicing factor. *The Journal of Biological Chemistry* 281, 36510-36517.
- Turunen, J. J., E. H. Niemelä, B. Verma, et M. J. Frilander. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews. RNA* 4, 61-76.
- Valkov, E., J. C. Dean, D. Jani, S. I. Kuhlmann, et M. Stewart. (2012). Structural basis for the assembly and disassembly of mRNA nuclear export complexes. *Biochimica Et Biophysica Acta* 1819, 578-592.
- Van Etten, J. L. (2003). Unusual life style of giant chlorella viruses. *Annual Review of Genetics* 37, 153-195.

- Voinnet, O., S. Rivas, P. Mestre, et D. Baulcombe. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *The Plant Journal: For Cell and Molecular Biology* 33, 949-956.
- Wahl, M. C., C. L. Will, et R. Lührmann. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701-718.
- Wang, E. T., R. Sandberg, S. Luo, I. Khrebtkova, L. Zhang, C. Mayr, et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470-476.
- Watanabe, Y., Y. Emori, I. Ooshika, T. Meshi, T. Ohno, et Y. Okada. (1984). Synthesis of TMV-specific RNAs and proteins at the early stage of infection in tobacco protoplasts: transient expression of the 30K protein and its mRNA. *Virology* 133, 18-24.
- West, S. (2012). The increasing functional repertoire of U1 snRNA. *Biochemical Society Transactions* 40, 846-849.
- Wright, E. A., T. Heckel, J. Groenendijk, J. W. Davies, et M. I. Boulton. (1997). Splicing features in maize streak virus virion- and complementary-sense gene expression. *The Plant Journal: For Cell and Molecular Biology* 12, 1285-1297.
- Xie, Z., L. K. Johansen, A. M. Gustafson, K. D. Kasschau, A. D. Lellis, D. Zilberman, et al. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biology* 2, E104.
- Yarborough, M. L., M. A. Mata, R. Sakthivel, et B. M. A. Fontoura. (2014). Viral subversion of nucleocytoplasmic trafficking. *Traffic (Copenhagen, Denmark)* 15, 127-140.
- Yasaka, R., H. D. Nguyen, S. Y. W. Ho, S. Duchêne, S. Korkmaz, N. Katis, et al. (2014). The Temporal Evolution and Global Spread of Cauliflower mosaic virus, a Plant Pararetrovirus. *PloS One* 9, e85641.
- Zhao, H., M. Chen, et U. Pettersson. (2014). A new look at adenovirus splicing. *Virology* 456-457, 329-341.

THÈSE

présentée par : **Clément BOUTON**

Étude de l'épissage alternatif de l'ARN 35S du virus de la mosaïque du chou-fleur (CaMV) et de l'export nucléaire des ARN viraux

RÉSUMÉ : L'épissage alternatif de l'ARN pré-génomique et polycistronique 35S et son export nucléaire sont deux étapes du cycle infectieux du virus de la mosaïque du chou-fleur (CaMV) qui ont fait l'objet de peu d'études. Quatre sites donneurs et un site accepteur localisés dans la région 5' de l'ARN 35S, ont déjà été décrits. Nos résultats montrent que cette région contient en réalité plus d'une douzaine de sites donneurs et accepteurs impliqués dans la génération d'une quinzaine d'isoformes d'ARN 35S par des épissages simples, doubles ou triples. L'épissage alternatif ne semble pas avoir pour fonction d'augmenter la complexité du protéome viral, bien que des séquences codantes résultant de fusions entre ORF soient présentes sur la plupart des isoformes épissées. Si une fonction dans la régulation négative de l'expression de l'ORF II a précédemment été décrite, le rôle de l'épissage alternatif apparaît encore énigmatique et difficile à évaluer puisque l'inactivation de sites donneurs ou accepteurs est systématiquement contrecarrée par l'utilisation de sites cryptiques, par la machinerie d'épissage. Pour maintenir l'intégrité du génome viral, certaines molécules d'ARN 35S doivent être préservées de l'épissage et exportées vers le cytoplasme sous une forme complète, où elles seront utilisées comme messager pour l'expression des gènes viraux, et comme ARN pré-génomique pour la rétrotranscription en ADN viral complet. L'essentiel de nos travaux sur ce sujet a porté sur le développement d'approches expérimentales destinées à l'étude de la voie d'export nucléaire de l'ARN 35S, ainsi que des séquences et protéines virales potentiellement impliquées dans ce processus. Les possibles acteurs agissant en *cis* et en *trans* sont discutés.

Mots clés : épissage alternatif, export nucléaire, CaMV, virus

A study of Cauliflower mosaic virus 35S RNA alternative splicing and viral RNAs nuclear export

ABSTRACT: Alternative splicing and nuclear export of the pregenomic polycistronic 35S RNA are two steps of the infectious cycle of *Cauliflower mosaic virus* that are poorly understood. Four splice donor sites and one splice acceptor site, all located within the 35S RNA 5' region, have been described in previous reports. Our results show that this region actually possesses more than a dozen of splice donor and acceptor sites which are involved in the generation of at least fifteen 35S isoforms via single, double and triple splicing. Alternative splicing does not appear to be used to expand CaMV proteome, although most of spliced isoforms contain new coding sequences arising from fusions between ORFs. While a function in downregulation of ORF II expression has been previously described, the role of alternative splicing for CaMV is still confusing and difficult to assess. Indeed, inactivation of splice donor or acceptor sites is constantly rescued by the use of cryptic sites. In order to maintain CaMV genomic integrity, some unspliced 35S RNA must be exported to the cytoplasm where they will serve both as replicative intermediate and as mRNA for the synthesis of viral proteins. The important part of our work has been focused on developing experimental tools dedicated to study the 35S RNA nuclear export pathway as well as viral sequences and proteins potentially involved in this process. Putative *cis*- and *trans*-acting factors are discussed.

Keywords: alternative splicing, nuclear export, CaMV, virus