

Caractérisation de la fonction "La" chez Arabidopsis thaliana et identification d'une structure conservée pour les ARN non-codants de type SINE

Sophie Fleurdépine

▶ **To cite this version:**

Sophie Fleurdépine. Caractérisation de la fonction "La" chez Arabidopsis thaliana et identification d'une structure conservée pour les ARN non-codants de type SINE. Génétique. Université Blaise Pascal - Clermont-Ferrand II; Université d'Auvergne - Clermont-Ferrand I, 2007. Français. NNT : 2007CLF21791 . tel-00718377

HAL Id: tel-00718377

https://theses.hal.science/tel-00718377

Submitted on 16 Jul 2012

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

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

UNIVERSITÉ BLAISE PASCAL **UNIVERSITÉ D'AUVERGNE**
N° d'ordre : D.U. 1791

ÉCOLE DOCTORALE
DES SCIENCES DE LA VIE ET DE LA SANTÉ

Thèse

Présentée à l'Université Blaise Pascal
pour l'obtention du grade de

DOCTEUR D'UNIVERSITÉ
(Spécialité : Physiologie et Génétique Moléculaires)

Soutenue le 23 novembre 2007

Sophie Fleurdépine

Caractérisation de la fonction "La" chez *Arabidopsis thaliana* et identification d'une structure conservée pour les ARN non-codants de type SINE.

Président	:	Pr. Georges Picard, Université Blaise Pascal, Aubière
Rapporteurs	:	Dr. Laurence Maréchal-Drouard, I.B.M.P., Strasbourg
		Pr. Manuel Echeverria, Université de Perpignan
Membres	:	Dr. Cécile Bousquet-Antonelli, Université de Perpignan
		Dr. Martin Crespi, I.S.V., Gif sur Yvette
		Pr. Jean-Marc Deragon, Université de Perpignan, directeur de thèse

U.M.R. C.N.R.S. 6547 GEEM – BIOMOVE
24, avenue des Landais 63177 Aubière Cedex

Il est de coutume au début d'un tel travail de rendre mérite à tous ceux qui ont contribué à le rendre possible, et je suis très heureuse de pouvoir remercier les personnes qui ont pris part à la réalisation de cette thèse par leur aide, leurs conseils et leur soutien.

Merci

à Mme Laurence Drouard, M Manuel Echeverria et M Martin Crespi d'avoir accepté de juger ce travail et pour l'intérêt qu'ils lui ont porté.
à M Georges Picard de m'avoir accueillie au sein de son laboratoire et de me faire l'honneur de présider ce jury.

à Jean-Marc pour ses conseils, son enthousiasme communicatif et la confiance qu'il m'a accordée.
à Cécile pour son aide et sa disponibilité au cours de ces quatre années. Merci de m'avoir transmis une bonne partie de ta ferveur pour la science et de ton organisation de la recherche (même si j'aurai aussi dû adopter ton sens de l'ordre…).

à Pascale Romby de m'avoir accueillie si chaleureusement dans son équipe et de m'avoir enseigné la cartographie en solution des ARN.

aux membres de l'équipe, Alain pour avoir encadrer mes premiers « pas » à la paillasse, Chantal pour sa sympathie, Marie-Noëlle et Thierry pour leurs conseils (ARN et protéines, et plus) et leur bonne humeur, ainsi qu'à tous les membres de Biomove pour leur aide et la gentillesse qu'ils m'ont témoigné au cours de cette thèse.

à Fabienne, Saloua et Isabelle pour m'avoir fait partager leur plaisir d'enseigner et leurs conseils pédagogiques.

à mes camarades du labo, Sophie et Julien, merci pour votre aide et votre compagnie, Jean-Baptiste, Cyril, Carles, Sam, merci pour ta gentillesse ton humour et tes chansons (« salade de fruit » restera ma préférée), merci aussi aux « anciens » Isabelle, Olivier, Guillaume, Sandra et José-Luis.

à Marie-Claude, je te remercie pour toute l'aide que tu m'as apportée, pour ton amitié et ta gaieté.

à Sébastien, mon « alter-ego » du labo (non, je ne suis pas en train de te traiter de bécasse) ces années de thèses n'auraient pas été les mêmes sans toi, et mine de rien, on a peut-être attrapé deux-trois tics de langage, « Marco ! ».

aux informaticiens au sens très large, Pascal et Séverine, Alain, Pierre et Aurélie, Sébastien, Guillaume, François, François, Patrick et les autres pour les repas de midi et les soirées du jeudi, jeudredi ou jardi…

à Sophie et Romain, Claire et Kevin, Anne et Mathieu, Marie-Pauline et Antoine, Lisa, Arnaud pour votre amitié et votre fidélité pendant toutes ces années.

à Myriam et Michael, vous entendre est toujours un joyeux réconfort, vous voir plus souvent, un de mes grands espoirs.

à mes Parents pour leur compréhension, leurs encouragements et leur soutien inconditionnel.

à Jean-Yves pour ta confiance, ta patience et tant d'autre raisons, merci d'être à mes côtés.

Table des matières

Résumé & Abstract.....*1*

Liste des figures......*2*

Tableaux.....*3*

Liste des abréviations......*5*

Synthèse bibliographique.....*6*

I. Les éléments transposables.....*7*

II. Les SINE.....*8*

II.1. Structure des éléments SINE.....8

II.2. Les familles de SINE de *Brassicaceae*.....9

II.3. Expression des éléments SINE.....10

II.3.1) La transcription.....10

II.3.2) Contrôle de la production et de la stabilité des transcrits SINE.....11

II.3.2)a) Contrôle épigénétique.....11

II.3.2)b) Contrôle post-transcriptionnel.....14

II.4. La transposition des SINE.....16

II.4.1) La relation SINE/LINE.....16

II.4.2) Les éléments LINE.....16

II.4.3) La rétroposition des SINE.....18

II.5. Impact des SINE sur l'hôte.....19

II.5.1) Impact génomique.....19

II.5.1)a) Insertions mutagènes.....19

II.5.1)b) Effet sur l'architecture et l'expression des génomes.....19

II.5.2) Rôle riborégulateur des transcrits SINE.....21

II.6. Les particules ribonucléoprotéiques (RNP) des SINE.....23

III. La protéine La.....*25*

III.1. Données de structure.....25

III.1.1) Le domaine La.....26

III.1.2) Le domaine RRM1.....28

III.1.3) Reconnaissance des extrémités UUU-OH des transcrits Pol III.....28

III.1.4) Le domaine carboxy-terminal.....30

III.2. Localisation subcellulaire 30

III.3. Les fonctions de la protéine La 32

III.3.1) Fonctions dans la biosynthèse des ARNt 32

III.3.1)a) Processus de maturation des ARNt 32

III.3.1)b) La protéine La est nécessaire à la maturation normale des précurseurs des ARNt 33

III.3.1)c) Rôle de chaperonne 34

III.3.1)d) Contribution des domaines La et RRM1 de la protéine La à ses activités dans la maturation des ARNt 37

III.3.2) La protéine La impliquée dans la biogenèse de transcrits Pol II 38

III.3.3) Implication de la protéine La dans la biogenèse du petit ARN U6 et inter-connexion de l'action de la protéine La et des protéines Lsm 39

III.3.4) Facteur de transcription de l'ARN polymérase III : un rôle controversé 40

III.3.5) Rôle dans la traduction des ARNm 42

III.3.5)a) La protéine La est impliquée dans la traduction indépendante de la coiffe 5' 42

III.3.5)b) Rôle dans la traduction des composants de la machinerie de traduction 45

III.3.5)c) Interaction de la protéine La avec des ARNm chez S. cerevisiae 46

III.3.5)d) Un nouveau rôle de la protéine La dans la biologie d'ARNm : cas de l'ARNm de la Peptidylglycine α-Amidating Monooygenase (PAM) des mammifères. 47

III.4. La protéine La humaine, des rôles différents selon l'état de phosphorylation 47

Objectifs des travaux de thèse *50*

Résultats *52*

I. Caractérisation de la fonction La chez Arabidopsis thaliana *53*

I.1. Article 1 : A bona fide La protein is required for embryogenesis in *Arabidopsis thaliana* 53

I.2. Données supplémentaires 78

I.2.1) Complémentation des lignées mutantes pour le gène AtLa1 par la fusion GFP-AtLa1 78

I.2.2) AtLa1 est exclue des corps de Cajal 79

I.3. Conclusion 81

II. « Les fonctions La » sont-elles remplies par deux protéines chez Arabidopsis thaliana ? .84

II.1. La protéine At79 présente les caractéristiques d'une authentique protéine La : données structurales et phylogénétiques 84

II.2. Étude fonctionnelle de la protéine At79 86

II.2.1) At79 lie-t-elle les transcrits Pol III ? 86

II.2.2) Localisation subcellulaire de la protéine At79 87

II.2.3) Lignées mutantes pour le gène At79 88

II.3. Étude de la dualité de fonction entre AtLa1 et At79 88
II.3.1) Colocalisation des deux protéines 88
II.3.2) Analyse et comparaison des ARN liés par AtLa1 et At79 89
II.4. Conclusion 92
III. Fonction La sur la biologie des SINE *94*
III.1. Co-immunoprécipitation des transcrits SINE par AtLa1 95
III.2. Conclusion 96
IV. Structure évolutive des ARN SINE dérivés d'ARNt *97*
IV.1. Introduction 97
IV.2. Cartographie en solution des ARN des SINE SB1 de *Brassica napus* et SB2 d'*Arabidopsis thaliana* 98
IV.2.1) SB1 98
IV.2.2) SB2 100
IV.2.3) Mise en parallèle des résultats de structure obtenus pour les SINE SB1 et SB2 100
IV.3. Article 2 : Common evolutionary trends for SINE RNA structures 101
IV.4. Conclusion 118
Discussion *119*
I. Deux authentiques protéines La chez Arabidopsis thaliana *120*
II. Pourquoi deux protéines La chez Arabidopsis thaliana ? *121*
III. Les autres fonctions La chez Arabidopsis thaliana *124*
III.1. Un rôle dans la traduction ? 124
III.2. Un lien entre la fonction La et la réponse au stress 125
IV. La protéine AtLa1 est impliquée dans la biologie des SINE *126*
V. Existe-t-il un lien entre la structure secondaire des ARN SINE et leurs potentiels rôles riborégulateurs ? *127*
Matériel et Méthodes *129*
I. Lignées d'Arabidopsis et suspensions cellulaires *130*
II. Souches de levure et transformation par choc thermique *131*
III. Préparation et transformation transitoire de protoplastes d'Arabidopsis *131*

IV. Préparation des extraits natifs *132*

V. Immunoprécipitations et élutions des ARN et des protéines *132*

VI. RT-PCR des transcrits Pol III et des transcrits SINE *133*

VII. Marquage pCp *133*

VIII. Cartographie en solution des ARN SINE SB1 et SB2 ..... *134*

VIII.1. Préparation de l'ARN 134

VIII.2. Marquage des ARN en 5' 135

VIII.3. Cartographie en solution par digestion enzymatique (méthode de détection directe) 135

VIII.4. Cartographie en solution par modifications chimiques (méthode de détection indirecte) ... 136

IX. Liste des amorces utilisées pour cette étude *138*

Références bibliographiques *142*

Résumé

Les rétroposons SINE sont des éléments d'ADN mobiles présents dans presque tous les génomes eucaryotes. L'étude des SINE a d'abord porté sur le processus de rétroposition qui nécessite la synthèse d'un intermédiaire ARN, l'effet mutagène de leur insertion, et leur impact général sur la structure et l'évolution des génomes. Cependant, il a récemment été proposé que les ARN SINE joueraient en tant que tel, un rôle de riborégulateurs cellulaires. L'étude des éléments *cis* et *trans* intervenant dans le métabolisme des ARN SINE menée chez la plante modèle *Arabidopsis thaliana* a pour but de mieux appréhender la biologie des éléments SINE. Nous avons dans un premier temps déterminé expérimentalement la structure secondaire des ARN des SINE dérivés d'ARNt SB1 de *Brassica napus* et SB2 d'*Arabidopsis*. Ainsi ces deux ARN qui ne présentent pas d'homologie de séquence primaire adoptent des structures secondaires similaires. Suite à cette observation, une étude bio-informatique permettant une analyse évolutive de la structure secondaire des ARN et incluant des séquences d'ARN SINE dérivés d'ARNt provenant de génomes eucaryotes très variés a été conduite par FJ Sun, G Caetano-Anollés et JM Deragon. Cette étude a mis en évidence un schéma évolutif commun au niveau de la structure secondaire des ARN SINE qui pourrait être lié au rôle de ces ARN en tant que riborégulateurs.
Dans le cadre de la recherche de facteurs intervenant *en trans* dans le métabolisme des ARN SINE, nous avons entrepris la caractérisation de la protéine La qui est un facteur ubiquitaire de liaison à l'ARN, impliquée dans le métabolisme de diverses classes d'ARN allant des ARN non-codants aux ARNm cellulaires ou viraux. Contrairement aux autres organismes eucaryotes où il existe une seule protéine La, nous avons identifié chez *Arabidopsis* deux protéines présentant toutes les caractéristiques phylogénétiques et structurales de la protéine La : At32 et At79. Nous avons montré que la protéine At32 (renommée AtLa1) est capable d'endosser les fonctions nucléaires de la protéine La dans la maturation des petits ARN non-codants et notamment des ARN SINE. L'étude de mutants d'insertion d'*Arabidopsis* a montré que la protéine AtLa1 était nécessaire à l'embryogenèse. En revanche, la protéine At79 qui possède la même localisation nucléaire que la protéine AtLa1 n'est pas nécessaire à la viabilité de la plante. Les protéines AtLa1 et At79 ont des niveaux et profils d'expression différents dans la plante. De plus, At79 semble lier des ARN distincts des ARN liés par AtLa1, nous proposons donc qu'il existe chez *Arabidopsis*, deux homologues fonctionnels de la protéine La qui ont acquis un certain niveau de spécialisation dans la fonction La.

Abstract

SINEs are mobile DNA elements found in almost all eukaryotes. Study of SINEs first focused on their retroposition mechanism, mutagenic effect and general impact on the structure and evolution of genomes. However SINE RNAs have recently been proposed to act as cellular riboregulators. Studying *cis* and *trans* elements taking part in the metabolism of SINE RNA in the model plant *Arabidopsis thaliana*, we aim to better understand SINE element biology. First, we experimentally defined the secondary structure of two tRNA-related SINE RNA: SB1 from *Brassica napus* and SB2 from *Arabidopsis*. Although unrelated at the primary sequence level, we found that these RNAs present similar secondary structures. Following this observation, an *in silico* analysis including tRNA-related SINE RNAs from various eukaryotes has been performed by FJ Sun, G Caetano-Anollés and JM Deragon. This study underlines the existence of common evolutionary trends for SINE RNA secondary structure that could be linked with the SINE RNA riboregulators function.
Searching for *trans* factors involved in SINE RNA metabolism, we have chosen to characterise the La protein, an ubiquitous RNA-binding protein involved in the metabolism of various RNAs, from non-coding RNA to cellular or viral mRNA. Unlike other eukaryotes, which have only one La protein, we identified in *Arabidopsis* two proteins with the phylogenetic and structural characteristics of genuine La protein: At32 and At79. We showed that At32 (renamed AtLa1) is able to fulfil La nuclear functions in non-coding RNA maturation, including SINE RNA. We also demonstrate that loss of AtLa1 function leads to embryonic lethality. Although AtLa1 and At79 have the same nuclear localisation, loss of At79 function did not affect viability. AtLa1 and At79 have differing levels and profiles of expression. Furthermore, the AtLa1 and At79 proteins apparently bind distinct sets of RNA. We thus propose that *Arabidopsis* possess two functional homologues of the La protein, which have partially specialised to fulfil different aspects of the La function.

Mots clefs : *Arabidopsis thaliana*, SINE, structure secondaire des ARN, ARN non codants, protéine La

Liste des figures

Figure 1. *Représentation schématique de la structure des différents types d'éléments transposables* *7*

Figure 2. *Exemple de structure d'éléments SINE* *8*

Figure 3. *Modèle simplifié de la méthylation de l'ADN dirigée par l'ARN* *13*

Figure 4. *Mise en évidence des différentes formes de l'ARN du SINE SB1* *15*

Figure 5. *Modèle de rétroposition de l'élément L1* *17*

Figure 6. *Représentation schématique de la structure des protéines La* *25*

Figure 7. *Représentation schématique de la structure du domaine La* *26*

Figure 8. *Représentation schématique de la structure du domaine RRM1* *28*

Figure 9. *Structure tertiaire de la région amino-terminale de la protéine La humaine liée à l'extrémité 3' UUU-OH d'un nonamère d'ARN* *29*

Figure 10. *Représentation schématique de la structure du domaine RRM2 de la protéine La humaine* *30*

Figure 11. *Mouvement des pré-ARNt et des ARNt matures chez S. cerevisiae* *32*

Figure 12. *Modèle de la maturation des pré-ARNt en levures sauvages et en levures dépourvues de protéine La* *33*

Figure 13. *La protéine AtLa1 est exclue des corps de Cajal* *81*

Figure 14. *AtLa1 et At79 sont les seules protéines d'Arabidopis à contenir un domaine La associé à un domaine RRM, dont les caractéristiques correspondent aux autres protéines La* *84*

Figure 15. *Les acides aminés impliqués dans la liaison à l'ARN des motifs La et RRM1 sont conservés chez les protéines AtLa1 et At79* *85*

Figure 16. *Expériences de co-immunoprécipitation de transcrits pol III à partir d'extraits natifs de cellules d'Arabidopsis exprimant la GFP, AtLa1-GFP ou At79-GFP* *86*

Figure 17. *Localisation subcellulaire de la protéine de fusion At79-GFP* *87*

Figure 18. *Caractérisation de deux lignées d'insertion ADN-T pour le gène AT79* *88*

Figure 19. *Colocalisation des protéines de fusion At79-GFP et AtLa1-CherryFP* *89*

Figure 20. *Marquage pCp des ARN co-immunoprécipités par les protéines GFP, AtLa1-GFP et At79-GFP* *90*

Figure 21. *AtLa1-GFP lie les précurseurs de l'ARNr 5S* *91*

Figure 22. *La protéine AtLa1 fixe spécifiquement les transcrits des éléments SINE SB1 et SB4* *95*

Figure 23. *Cartographie en solution de la forme cytoplasmique courte de l'ARN SINE SB1* *98*

Figure 24. *Cartographie en solution de la forme pleine longueur de l'ARN SINE SB1* *99*

Figure 25. *Cartographie en solution de la forme cytoplasmique courte de l'ARN SINE SB2* *100*

Figure 26. *Cartographie en solution de la forme pleine longueur de l'ARN SINE SB2* *101*

Figure 27. *Exemples d'ARN reconnus avec la plus grande affinité par des domaines dsRBD lors des expériences de SELEX menées par Hallegger et al. (2006)* *118*

Figure 28. *Détail des positions d'insertion des ADN-T dans les lignées mutantes pour les gènes ATLA1 et AT79* *130*

Tableaux

Tableau 1. *Présentation des coupures enzymatiques et modifications chimiques* *134*

Tableau 2. *Liste des amorces utilisées pour le clonage des ADNc des protéines GFP, AtLa1, At79 et Lhp1p* *138*

Tableau 3 : *liste des amorces utilisées pour les tests de complémentation en levure S. cerevisiae* *139*

Tableau 4. *Liste des amorces utilisées pour le génotypage des lignées d'insertion ADN-T* *140*

Tableau 5. *Liste des amorces utilisées pour la détection des transcrits Pol III et SINE* *141*

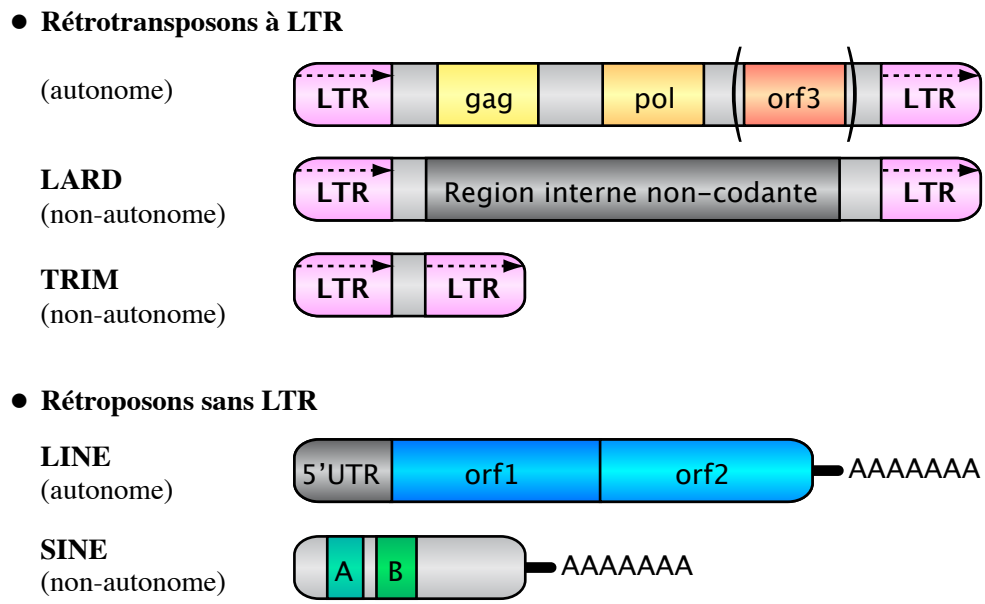
Liste des abréviations

A :	Adénine
aa :	acide aminé
ADN :	Acide DésoxyriboNucléique
ADNc :	ADN complémentaire
ADN-T :	ADN de Transfert
ARN :	Acide RiboNucléique
ARNi :	ARN interférence
ARNm :	ARN méssager
ARNr :	ARN ribosomique
ARNt :	ARN de transfert
ATP :	Adénosine TriPhosphate
dNTP :	Désoxyribo-Nucléotide TriPhosphate
Forme sc :	Forme cytoplasmique courte de l'ARN SINE
Forme fl :	Forme pleine longueur de l'ARN SINE
Forme pA :	Forme de l'ARN SINE se terminant par une queue poly(A)
Kb :	Kilobase
kDa :	KiloDalton
LINE :	Long INterspersed Element
LTR :	Long Terminal Repeat
NLS :	Nuclear Localization Signal
NoLS :	Nucleolar Localization Signal
NRE :	Nuclear Retention Element
NTP :	Nucléotide TriPhosphate
ORF :	Open Reading Frame (Cadre de lecture ouvert)
PAB ou PABP	Poly(A) Binding Protein
piARN :	Piwi-Interacting ARN
Pol I :	ARN polymérase I
Pol II :	ARN polymérase II
Pol III :	ARN polymérase III
Pol IV :	ARN polymérase IV
Pré-ARNt :	Précurseur d'ARNt

RNase :	Ribonucléase
RNP :	RiboNucléoProtéique
RRM :	RNA Recognition Motif
scaARN :	small Cajal bodies specific ARN
siARN :	small interfering ARN
SINE :	Short INterspersed Element
snARN :	small nuclear ARN
snoARN :	small nucleolar ARN
snRNP :	small nuclear RNP
SRP :	Signal Recognition Particule (Particule de reconnaissance du signal)
T :	Thymine
TPRT :	Target-Primed Reverse Transcription
U :	Uracile
UV :	Ultra-Violet

Synthèse bibliographique

Classe I - Rétroéléments



Classe II - Transposons à ADN

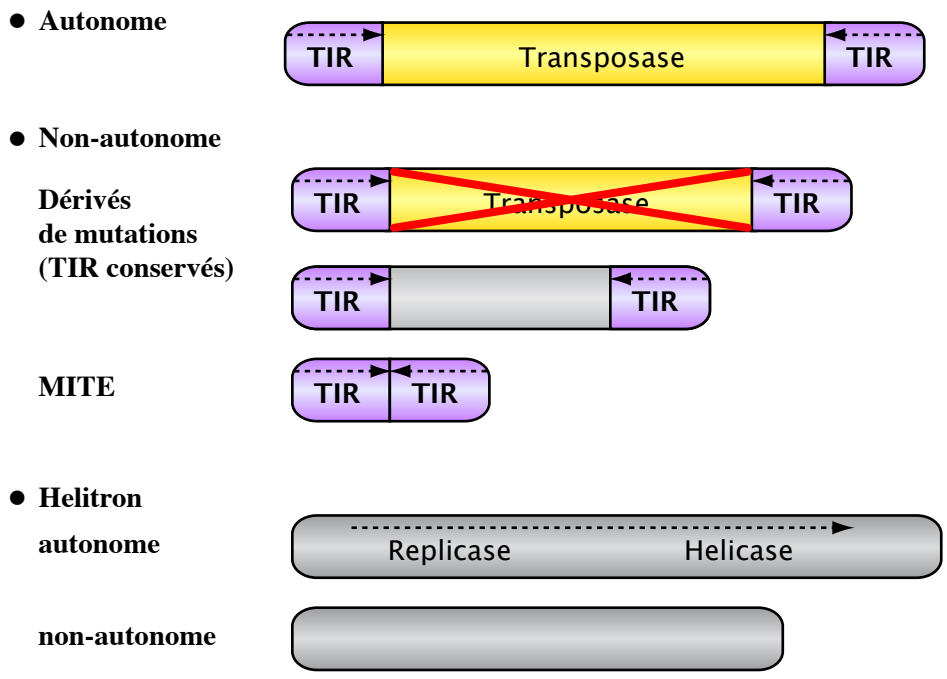


Figure 1. Représentation schématique de la structure des différents types d'éléments transposables

LTR : Long Terminal Repeat, LARD : LArge Retrotransposon Derivative, TRIM :Terminal-repeat Retrotransposon In Miniature, LINE : Long INterspersed Element, SINE : Short INterspersed Element TIR : Terminal Inverted Repeat, MITE : Miniature Inverted-repeat Transposable Element

I. Les éléments transposables

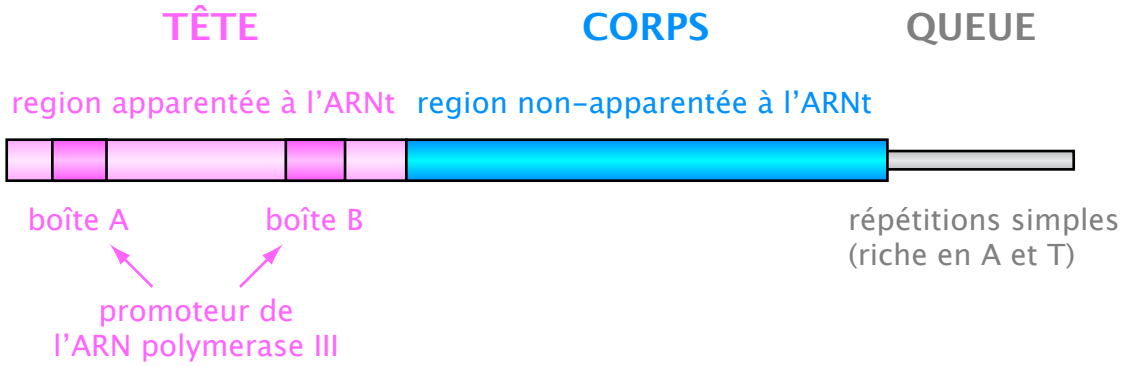
Le séquençage complet du génome humain ainsi que des génomes des organismes modèles comme la souris, la drosophile et la plante *Arabidopsis thaliana* a confirmé que les séquences d'ADN répétées sont des composants majeurs des génomes eucaryotes (Adams *et al.* 2000 ; The Arabidopsis Genome Initiative 2000; Lander *et al.* 2001 ; Waterston *et al.* 2002). Deux sortes de séquences répétées ont été identifiées:

- les séquences répétées en tandem comme les microsatellites, les minisatellites ou les télomères, qui sont le plus souvent limitées à des régions spécifiques des chromosomes ;
- les séquences répétées dispersées, qui sont pour la plupart des copies inactives d'une grande variété d'éléments transposables actuellement ou historiquement actifs dans le génome.

Les éléments transposables ont été détectés dans les génomes de tous les organismes étudiés, des bactéries à l'homme et sont des séquences d'ADN capables de créer de nouvelles copies d'elles-mêmes ou de se déplacer dans d'autres sites du génome. On classe ces éléments selon leur mode de réplication (figure 1) (Prak and Kazazian 2000). Ainsi, les éléments de la classe I ou rétroéléments sont transcrits en ARN et ils créent de nouvelles copies par transcription inverse de cet intermédiaire ARN. Parmi eux se trouvent notamment les rétrotransposons à LTR (Long Terminal Repeat), et les rétroéléments sans LTR : les LINE (Long INterspersed Element) et les SINE (Short INterspersed Element), qui seront présentés plus en détail par la suite. Les rétrotransposons à LTR ont une structure et un mécanisme d'insertion très proche de ceux des rétrovirus. Comme les rétrovirus, les rétrotransposons présentent à leurs extrémités deux longues séquences répétées terminales ou LTR qui sont nécessaires à la régulation de l'expression, à la transcription inverse et à l'intégration de l'élément. Les éléments de la classe II, appelés transposons, créent de nouvelles copies directement par excision et insertion de leur séquence ADN à un nouveau site chromosomique, selon un mode réplicatif ou non-réplicatif.

Dans les deux classes, il existe des éléments autonomes et des éléments non-autonomes. Les éléments autonomes codent les protéines nécessaires à leur propagation. Les éléments non-autonomes ne codent pas les enzymes nécessaires à leur transposition. En revanche, ils partagent une partie des caractéristiques d'éléments autonomes ce qui leur

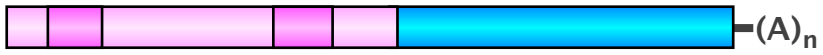
SINE typique



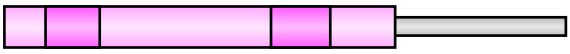
SINE possédant une région 3' apparentée à un LINE



SINE SB1 (*Brassica napus*)



SINE ID (rongeur)



SINE Alu (Humain)

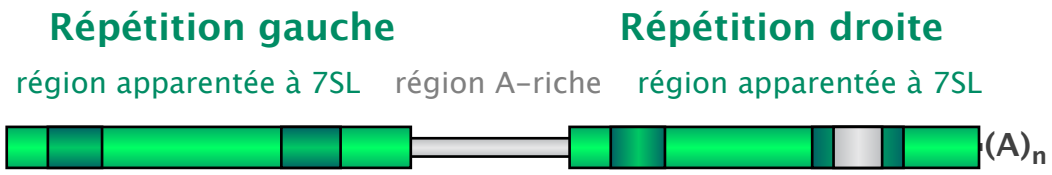


Figure 2. Exemple de structure d'éléments SINE

permet de parasiter la machinerie produite par ces derniers. À l'exception des hélitrons, tous les éléments transposables sont flanqués par des duplications du site cible qui résultent de leur mode d'insertion (Kapitonov and Jurka 2007).

Si tous les génomes contiennent des éléments transposables, la proportion et l'activité des différentes classes d'éléments est très variable entre les génomes.

II. Les SINE

Les éléments SINE ont d'abord été découverts dans les génomes de mammifères où ils sont très représentés (Jelinek *et al.* 1980). Ainsi, dans le génome humain, il existe plus d'un million de copies du SINE Alu, ce qui fait de lui l'élément transposable le plus abondant du génome humain, et aussi l'élément SINE le plus étudié (Lander *et al.* 2001). Chez la souris ce sont quatre familles de SINE qui représentent un total de 1,4 million de copies (Waterston *et al.* 2002).

Depuis, des éléments SINE ont été identifiés dans un grand nombre de génomes de vertébrés, d'invertébrés et de plantes. En revanche ils semblent absents des génomes des levures et des organelles.

Les SINE sont des éléments non-autonomes et dépendent pour leur rétroposition de la machinerie des LINE. On ne connaît effectivement pas de génome qui contiendrait des SINE mais pas de LINE. En revanche les éléments SINE sont absents de certains génomes, comme celui de la drosophile où on trouve des LINE actifs (Kramerov and Vassetzky 2005).

II.1. Structure des éléments SINE

Les SINE sont des séquences courtes, la majorité des SINE ont une longueur de 150 à 200 paires de bases. Ils se composent en général de trois parties (figure 2).

- La première partie en 5' de l'élément est dérivée d'un ARN cellulaire non codant. Cette séquence porte le promoteur interne reconnu par l'ARN polymérase III. La majorité des familles de SINE connues sont dérivées d'ARN de transfert (ARNt). Mais il existe des SINE, notamment le SINE Alu humain et le SINE B1 de souris, dérivés de l'ARN 7SL (Kramerov and Vassetzky 2005). Ce dernier est à la base de l'assemblage de la particule ribonucléoprotéique de reconnaissance du signal (SRP pour Signal Recognition Particule) qui reconnaît le signal d'adressage sur les chaînes

des polypeptides naissants, et dirige les ribosomes vers un récepteur localisé dans la membrane du réticulum endoplasmique. Plus récemment une famille de SINE dérivés de l'ARN ribosomique (ARNr) 5S a été identifiée chez le poisson zèbre (Kapitonov and Jurka 2003).

- Ils ont ensuite une partie centrale dite « corps ». Cette région n'est pas apparentée à l'ARN cellulaire dont le SINE dérive et sa séquence est propre à chaque famille de SINE.
- Ils se terminent par une séquence appelée « queue 3' » composée de nucléotides A (queue poly(A)) ou de répétitions simples.

Il faut noter qu'il existe des éléments ne correspondant pas exactement à ce schéma. Ainsi certains éléments ont des parties corps ou queue 3' très courtes, voire inexistantes, par exemple le SINE ID de rongeur présenté sur la figure 2 n'a pas de partie corps. De même certains SINE, comme l'élément Alu et plusieurs familles de SINE de plantes, ont une forme dimérique.

II.2. Les familles de SINE de *Brassicaceae*

La famille des *Brassicaceae* est une importante famille de plantes dicotylédones, incluant notamment *Brassica oleracea* (le chou), *Brassica napus* (le colza) et la plante modèle *Arabidopsis thaliana*.

La première famille de SINE identifiée chez les *Brassicaceae* a été découverte chez *Brassica napus* et s'appelle SB1 (Deragon *et al.* 1994). Depuis, il a été montré que cette famille est présente chez l'ensemble des *Brassica*, alors qu'elle est absente du génome d'*Arabidopsis thaliana*. À ce jour, 15 familles de SINE ont été identifiées chez les *Brassicaceae* : SB1 à SB15 (Deragon and Zhang 2006). Comme tous les SINE de plantes, les SINE de *Brassicaceae* ont une région 5' dérivée d'ARN de transfert. La taille de leur séquence consensus est très variable, en raison des variations de longueur de leur région corps non-apparentée à l'ARNt. Au contraire d'autres SINE de plantes, ils se terminent tous par une queue poly(A).

La séquence partielle du génome de *Brassica oleracea* compte 4600 copies réparties en 14 familles (SB1 à SB3 et SB5 à SB15), représentant 0,2% du génome (environ 600 millions de paires de bases). Le génome d'*Arabidopis thaliana* (130 millions de paires de bases) compte quant à lui 334 copies réparties en six familles (SB2 à SB7), représentant

0,07% du génome. Chez *Arabidopsis*, les éléments des familles SB2 et SB3 (deux des familles les plus représentées) se trouvent dans des régions géniques, alors qu'ils sont pratiquement absents des régions centromériques et péricentromériques, où se trouvent de nombreux autres éléments transposables (Lenoir *et al.* 2001).

II.3. Expression des éléments SINE

II.3.1) La transcription

Les SINE sont transcrits par l'ARN polymérase III (Pol III). Cette transcription est dirigée par le promoteur interne présent dans la partie 5' héritée de l'ARN cellulaire dont le SINE est dérivé. Une fois initiée par la Pol III, la transcription se poursuit sur la totalité de l'élément jusqu'à un signal de terminaison de transcription correspondant à une séquence d'au moins quatre résidus portant une base thymine (T) sur le brin transcrit (Bogenhagen and Brown 1981 ; Huang and Maraia 2001). Un tel signal n'est pas présent dans la séquence des éléments SINE, ce qui implique que leur transcription se poursuit dans l'ADN génomique flanquant, jusqu'à ce que la Pol III rencontre une séquence faisant office de terminateur. Il en résulte que les transcrits primaires des SINE issus de différents loci sont hétérogènes en taille et en séquence en aval de la queue poly(A).

La séquence promotrice étant située dans la région transcrite des SINE, toute nouvelle copie semble posséder en elle les séquences nécessaires à sa propre transcription. Il a donc été proposé que toutes les copies SINE présentant un promoteur interne intact seraient transcrites et pourraient de ce fait générer de nouveaux éléments (Van Arsdell *et al.* 1981). Or les études de transcription *in vivo* portant sur différents loci de l'élément Alu humain (Aleman *et al.* 2000) le locus BC1 chez la souris (Kim *et al.* 1994), et sur les SINE de *Brassicaceae* (Deragon *et al.* 1996) ont montré que ce n'était pas le cas.

Ainsi, les génomes eucaryotes comptent un grand nombre d'insertions SINE, mais seules quelques copies sont impliquées dans le phénomène d'amplification d'un élément. Ces copies, appelées gènes maîtres ou loci fondateurs réunissent des caractéristiques importantes pour l'activité de l'élément en termes de transcription mais aussi pour le processus de transcription inverse. Ainsi la région en amont de l'insertion doit contenir des séquences amplifiant la transcription de l'élément, lui permettant ainsi d'échapper, même partiellement, à la répression de la transcription et notamment dans les cellules de la lignée germinale (Chesnokov and Schmid 1996 ; Arnaud *et al.* 2001). De même, la région en aval de l'insertion

est impliquée dans la stabilité du transcrit mais aussi dans la capacité de l'ARN à initier la transcription inverse (Aleman *et al.* 2000 ; Pelissier *et al.* 2004). Chaque locus fondateur donnerait ainsi naissance à un nombre important de copies inactives.

Cette hypothèse est soutenue par le fait qu'une étude de la distribution des mutations sur les séquences SINE d'une même famille par rapport au consensus permet de grouper ces éléments en sous-familles. Une famille de SINE correspond à l'ensemble des copies d'un élément donné. Ces copies présentent des divergences nucléotidiques permettant le classement en sous-familles, tous les membres d'une sous-famille ayant hérité leur séquence d'un locus fondateur commun.

La transcription des SINE est fortement réprimée, ainsi les transcrits spécifiques de la Pol III sont très peu représentés (Paulson and Schmid 1986). Cependant, lorsque l'on cible la séquence d'un SINE par hybridation northern blot sur de l'ARN total, cette séquence s'avère être présente dans des ARN de grandes tailles (Kramerov *et al.* 1982; Deragon *et al.* 1996). Cette observation est due au phénomène de co-transcription : une partie des éléments SINE est insérée dans des gènes, le plus souvent dans les introns, donc lors de la transcription de ces gènes par l'ARN polymérase II (Pol II), la séquence du SINE se retrouve dans les transcrits primaires.

II.3.2) Contrôle de la production et de la stabilité des transcrits SINE

II.3.2)a) Contrôle épigénétique

Parmi le grand nombre d'éléments transposables présents dans les génomes eucaryotes, la plupart des éléments sont inactifs en raison de remaniements ou de mutations. Toutefois certains restent intacts, mais silencieux, sous la forme d'éléments cryptiques. En effet, les génomes ont développé des mécanismes épigénétiques de défense contre les éléments transposables, quels qu'ils soient, notamment pour combattre leurs effets potentiellement mutagènes. Ces mécanismes de « silencing » des éléments transposables agissent à la fois au niveau transcriptionnel et au niveau post-transcriptionnel, et mettent en jeu un très grand nombre de facteurs. Ces facteurs ont le plus souvent été identifiés grâce à des lignées mutantes présentant une réactivation des éléments transposables (Slotkin and Martienssen 2007).

Les cellules eucaryotes sont capables d'établir des domaines chromatiniens répressifs ou permissifs pour l'expression des gènes. L'établissement de ces domaines passe par toute

une gamme de modifications covalentes de la chromatine sans altération de la séquence nucléotidique, incluant la méthylation de l'ADN et des modifications des histones.

La méthylation de l'ADN au niveau des résidus cytosines est notamment un moyen de réprimer la transcription des éléments transposables. Les cytosines peuvent être méthylées en contexte symétrique CpG ou CpNpG et asymétrique CpNpN chez les plantes et en contexte symétrique CpG chez les mammifères. La méthylation symétrique de l'ADN est héritable car, lors de la réplication, elle peut être recopiée sur le nouveau brin d'ADN par des méthylases de maintenance, ce qui permet une transmission du silencing pré-établi (Richards and Elgin 2002).

Un autre dispositif de contrôle de l'expression des gènes est la modification des histones. La méthylation de la lysine 9 de l'histone H3 est une caractéristique de la chromatine transcriptionellement répressive, et les nucléosomes associés à des éléments transposables portent le plus souvent cette modification (Richards and Elgin 2002; Henderson and Jacobsen 2007 ; Slotkin and Martienssen 2007). L'état de méthylation de l'histone H3 est lui aussi transmissible, mais les mécanismes qui supportent ce phénomène ne sont pas entièrement élucidés (Henikoff *et al.* 2004).

Des phénomènes d'interconnection entre méthylation de l'ADN et méthylation de la lysine 9 de l'histone H3 ont été mis en évidence. Ainsi, chez *Arabidopsis thaliana*, il semble que la méthylation des histones participe au maintien de la méthylation de l'ADN. De même, la méthylation des cytosines semble capable d'influencer celle de l'histone H3 (Tariq and Paszkowski 2004).

L'établissement du « silencing » chromatinien au niveau de nouveaux éléments fait intervenir des voies dépendantes et indépendantes de l'ARN interférence (ARNi). D'une part, indépendamment de l'ARNi, il a été proposé que l'initiation du silencing sur des éléments nouvellement intégrés pourrait dépendre d'appariements ADN/ADN. Ces appariements entre les copies d'éléments pourraient conduire à la formation de structure d'ADN permettant la transmission des composants impliqués dans la méthylation de l'ADN et la formation d'hétérochromatine, et de cette façon aboutiraient à la transmission du silencing (Ye and Signer 1996).

D'autre part, une voie de méthylation de l'ADN dirigée par l'ARN a été mise en évidence chez les plantes, cette voie faisant intervenir des facteurs impliqués dans les voies d'ARNi permettrait d'établir un contexte chromatinien répressif. Chez *Arabidopsis*, la méthyltransférase DRM2 (Domains Rearranged Methyltransferase 2) est responsable de la

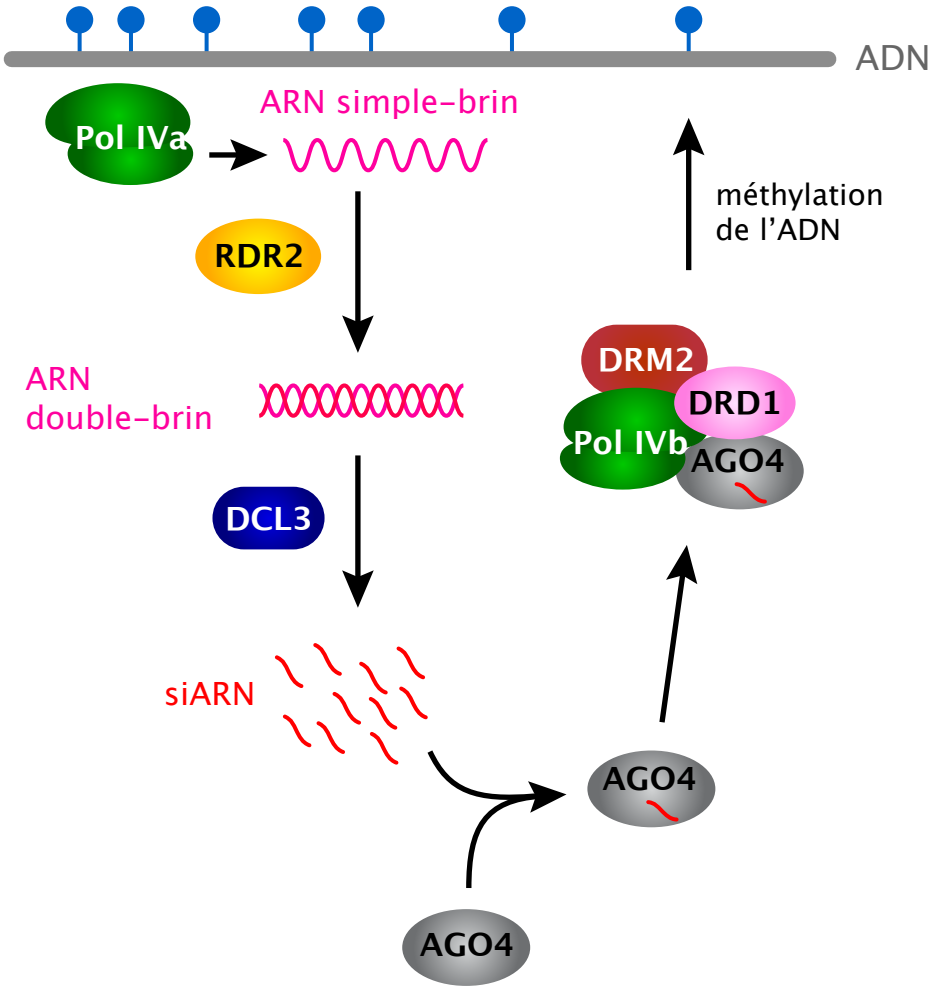


Figure 3. Modèle simplifié de la méthylation de l'ADN dirigée par l'ARN

Dans ce modèle, la Pol IVa interviendrait dans la transcription d'une région ADN partiellement méthylée. L'ARN simple brin produit par la Pol IVa est ensuite copié par RDR2. L'ARN double-brin résultant est clivé par DCL3, qui génère ainsi des siARN. Ces siARN sont pris en charge par la protéine AGO4. AGO4 interagirait alors avec la Pol IVb et la protéine DRD1, formant un complexe capable de diriger la méthylation de l'ADN par DRM2.

Pol IV : ARN polymérase IV, RDR2 : RNA Dependent RNA polymérase2, DCL3 : DICER-like 3, AGO4 : Argaunote 4, DRM2 : Domain Rearranged Methyltransferase 2, DRD1 : Defective in RNA Directed methylation.

Les sites de méthylation de l'ADN sont symbolisés par des boules bleues.

méthylation *de novo* de l'ADN. DRM2 peut être ciblée sur une séquence par des petits ARN interférents (siARN). Ces siARN sont générés par l'enzyme DICER-LIKE 3 (DCL3) et sont ensuite pris en charge par l'une des dix protéines argonaute : AGO4. AGO4 dirigerait les facteurs de remodelage de la chromatine au niveau du locus cible, probablement grâce à l'interaction entre les siRNA qu'elle porte et les transcrits naissants (figure 3) (Henderson and Jacobsen 2007). AGO4 serait aussi impliquée via son activité catalytique dans l'augmentation de la production des siARN en partenariat avec l'ARN polymérase ARN dépendante RDR2 (RNA Dependent RNA Polymerase).

Les deux formes de l'ARN polymérase IV (Pol IV) sont également impliquées dans le processus de méthylation de l'ADN dirigée par l'ARN. La Pol IVa transcrirait les ARN qui sont clivés en siARN par DCL3, mais le substrat sur lequel elle fonctionne reste à déterminer. Plusieurs substrats ont été proposés : l'ADN méthylé, des ARN naissants aux loci cibles produits par les ARN polymérases II ou III, ou des ARN double-brins générés par RDR2 à partir de transcrits des Pol II ou Pol III, ce qui formerait de ce fait une boucle d'amplification. Les deux dernières hypothèses impliquent que la Pol IV possède une activité ARN polymérase dépendante de l'ARN. La Pol IVa mettrait en marche la voie de méthylation de l'ADN dirigée par l'ARN lorsque le niveau de siRNA n'est plus suffisant pour maintenir le silencing du locus (Vaucheret 2005). La Pol IVb intervient elle aussi dans cette voie, et est nécessaire au renforcement de la méthylation (Pontier *et al.* 2005). Si son rôle est encore discuté, il est possible qu'elle ne soit pas impliquée dans la transcription, mais qu'elle coopère avec la protéine DRD1 (Defective in RNA-directed DNA methylation 1), un potentiel facteur de remodelage de la chromatine, pour exposer l'ADN aux méthyltransférases (Henderson and Jacobsen 2007 ; Huettel *et al.* 2007).

Il semble que selon leur position dans le génome et leur taille, le silencing des éléments transposables ne fasse pas intervenir les mêmes facteurs (Zilberman *et al.* 2003 ; Tran *et al.* 2005). Les éléments de petites tailles isolés dans des régions riches en séquences codantes, ce qui est le cas d'une grande partie des copies de SINE, semblent être plus difficile à éteindre que les éléments regroupés en cluster. Dans les grandes régions d'hétérochromatine, le système de méthylation de l'ADN dirigée par l'ARN intervient dans le silencing des éléments transposables. Mais en cas d'inactivation de ce système, la maintenance de la méthylation liée à la réplication apparaît suffisante pour maintenir une structure chromatinienne fermée. En revanche, le système de méthylation de l'ADN

dépendant des siRNA est indispensable pour maintenir le silencing des éléments transposables isolés (Zilberman and Henikoff 2004 ; Henderson and Jacobsen 2007).

Récemment une voie de silencing dépendante de siRNA plus longs, de 26 à 30 nucléotides, a été découverte chez la drosophile (Aravin *et al.* 2003), *Caenorhabditis* (Cox *et al.* 1998) et la souris (Aravin *et al.* 2006 ; Girard *et al.* 2006; Grivna *et al.* 2006). Les protéines caractéristiques de cette voie, identifiées jusqu'ici, appartiennent toutes à une sous-famille des protéines argonaute et sont exprimées spécifiquement dans la lignée germinale. En raison de la levée de la méthylation dans leur lignée germinale, les génomes des animaux semblent avoir développé un mécanisme pour y réguler l'activité des éléments transposables. Les petits ARN impliqués ont été renommés piRNA pour Piwi-Interacting RNA, Piwi étant une des protéines argonaute impliquées dans cette voie. Ces petits ARN ont d'intéressantes caractéristiques. En plus de leur taille, la majorité d'entre eux proviennent d'un petit nombre de loci. Chez la drosophile, des loci capables de contrôler l'expression des éléments transposables avaient précédemment été identifiés (Pelisson *et al.* 1994 ; Castro and Carareto 2004). On sait aujourd'hui que ces loci de contrôle correspondant à des copies défectueuses ou tronquées d'éléments transposables sont à l'origine des piRNA. Les protéines de la sous-famille Piwi n'ont en revanche pas été identifiées chez les plantes, dont la lignée germinale n'est pas différenciée précocement dans l'embryogenèse, au contraire de celle des animaux.

Les SINE sont des éléments non-autonomes, leur transposition est donc impossible si le génome contrôle la transposition des LINE. Le silencing des éléments SINE apparaît comme une sécurité supplémentaire, mais peut aussi être interprété comme un argument en faveur d'un rôle des transcrits SINE en tant que riborégulateurs, dont la production doit donc être très contrôlée.

II.3.2)b) Contrôle post-transcriptionnel

En plus du contrôle transcriptionnel des SINE, il existe des mécanismes de régulation au niveau post-transcriptionnel.

Ainsi, des espèces plus courtes d'ARN SINE ont été identifiées pour l'élément B1 en cellules murines (Adeniyi-Jones and Zasloff 1985). Il a été montré par micro-injection des ARN pleine longueur de B1 dans des oocytes de xénope que la forme courte de l'ARN est issue d'une maturation post-transcriptionnelle du transcrit primaire. Le transcrit court

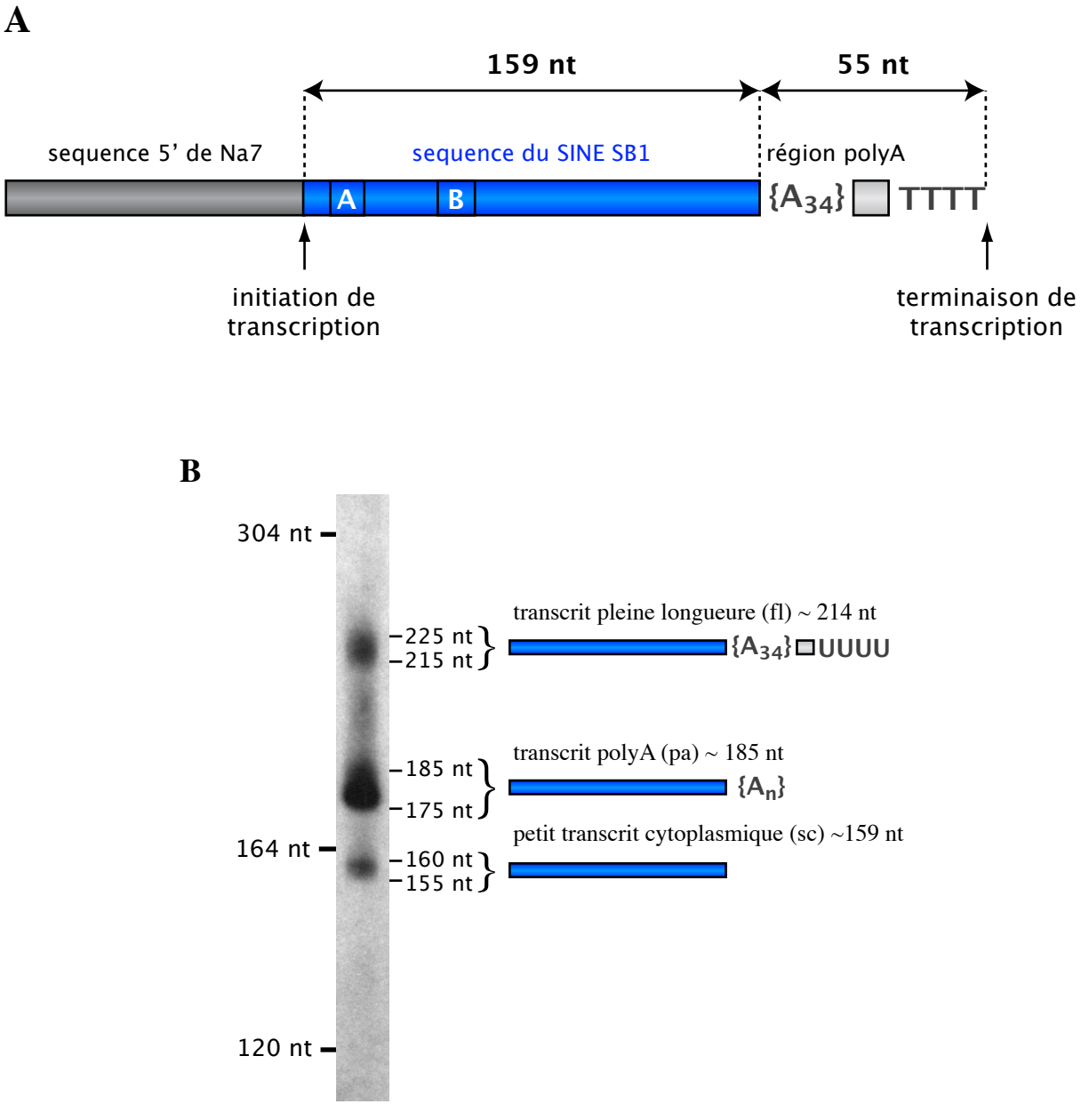


Figure 4. Mise en évidence des différentes formes de l'ARN du SINE SB1

(A) Représentation schématique du locus Na7 du SINE SB1 introduite dans le génome d'*Arabidopsis thaliana*.

(B) Hybridation northern ciblant l'ARN du SINE SB1. La partie droite présente la structure des différentes formes détectées.

(adapté de Pélissier *et al*, 2004)

correspond à la séquence du SINE sans sa queue poly(A) (Adeniyi-Jones and Zasloff 1985). De tels transcrits ont pu être détectés par hybridation northern dans différents tissus de souris (Maraia 1991). Ces transcrits restant localisés dans le cytoplasme, ils ont été nommés ARN cytoplasmiques courts ou « sc » (pour small cytoplasmic). Des résultats similaires ont été obtenus avec le SINE Alu humain (Maraia *et al.* 1993), mais aussi avec le SINE SB1 de *Brassica napus* (Arnaud *et al.* 2001 ; Pelissier *et al.* 2004). L'ARN cytoplasmique court étant dépourvu de queue poly(A), il est considéré comme inapte pour le processus de rétroposition (cf. paragraphe II.4 de la synthèse bibliographique). La maturation de l'ARN pleine longueur en ARN cytoplasmique court pourrait donc être un moyen de limiter la rétroposition des SINE, en limitant la quantité de l'ARN intermédiaire de la rétroposition (Maraia and Sarrowa 1995).

L'étude de la maturation du SINE SB1 dérivé d'ARNt réalisée par T. Pélissier a aussi mis en évidence une nouvelle forme de transcrit SINE (Pelissier *et al.* 2004). Le SINE SB1 étudié est présent dans le génome d'un grand nombre d'espèces de *Brassicaceae*, mais absent de celui d'*Arabidopsis thaliana*. La transcription du SINE SB1 est fortement réprimée dans le génome de *Brassica napus* où il a été découvert (Deragon *et al.* 1996). Un élément SB1 considéré comme fondateur a été introduit dans le génome d'*Arabidopsis* et des lignées de plantes ne possédant qu'un seul exemplaire de l'élément sous contrôle de son propre promoteur ont été sélectionnées. Dans ce contexte génomique, la transcription de l'élément n'est pas réprimée et on peut visualiser les ARN produits par northern blot (figure 4.B). Trois espèces d'ARN sont ainsi clairement identifiables : la forme pleine longueur (fl), la forme cytoplasmique courte (sc) dépourvue de queue poly(A) et de séquence terminatrice et une nouvelle espèce se terminant par une queue poly(A) (forme pA). Ces deux dernières formes sont présentes dans le cytoplasme.

Les modèles proposés jusqu'ici présentent la forme pleine longueur comme l'ARN pris en charge par la machinerie de rétroposition des LINE. Des études menées sur le LINE-1 (L1) humain ont montré que, *in vitro*, sa machinerie est plus efficace pour initier la transcription inverse de l'ARN sur une queue poly(A) terminale que sur une queue poly(A) interne (Cost *et al.* 2002). Un transcrit SINE présentant une queue poly(A) terminale apparaîtrait donc comme un meilleur intermédiaire pour la rétroposition que le transcrit pleine longueur (Pelissier *et al.* 2004).

II.4. La transposition des SINE

Les SINE sont des éléments non-codants, ils ne sont donc pas autonomes pour rétroposer. Pour créer une nouvelle copie, l'élément SINE doit d'abord être transcrit par la Pol III. Son ARN est exporté dans le cytoplasme où il peut être pris en charge par le complexe de rétroposition d'un LINE. Ce complexe va ensuite insérer dans le génome la copie ADN complémentaire du transcrit SINE.

II.4.1) La relation SINE/LINE

Plusieurs découvertes sont à l'origine de l'hypothèse selon laquelle les SINE utiliseraient la machinerie des LINE pour rétroposer. Ainsi, l'analyse *in silico* menée par J. Jurka a révélé que les sites d'insertion dans le génome humain des éléments SINE Alu ont la même séquence concensus d'intégration que les éléments LINE L1, suggérant que le SINE Alu utilise l'endonucléase du LINE L1 (Jurka 1997). Un autre argument en faveur de cette hypothèse a été la découverte de familles de SINE dérivés d'ARNt, possédant une partie de leur séquence corps et leur séquence 3' très proches de séquences 3' d'éléments LINE présents dans le même génome (Okada and Hamada 1997 ; Gilbert and Labuda 1999). Ces familles de SINE seraient donc capables de recruter la machinerie enzymatique des LINE grâce à cette séquence 3' commune (Ohshima *et al.* 1996).

Il existe aujourd'hui des preuves directes de la relation SINE/LINE. En ce qui concerne les partenaires SINE/LINE partageant une séquence 3', Kajikawa et Okada ont montré qu'en ajoutant cette séquence 3' à un ARNm, son taux de rétroposition augmente significativement lors de test en cellules exprimant les protéines de la machinerie de rétroposition du LINE (Kajikawa and Okada 2002). Plus directement, Dewannieux et al. (2003) ont mis en évidence la mobilisation par le LINE L1 d'un élément Alu en culture de cellules. (Dewannieux *et al.* 2003).

II.4.2) Les éléments LINE

Les LINE sont des éléments de 4 à 7 Kb. Ils contiennent un promoteur interne reconnu par l'ARN polymérase II, un ou deux cadres de lecture ouverts (ORF pour Open Reading Frame), un signal de polyadénylation à leur extrémité 3' et une queue poly(A). La majorité des séquences LINE identifiées dans les génomes sont tronquées en 5', ce qui est attribué à un processus de reverse transcription incomplet.

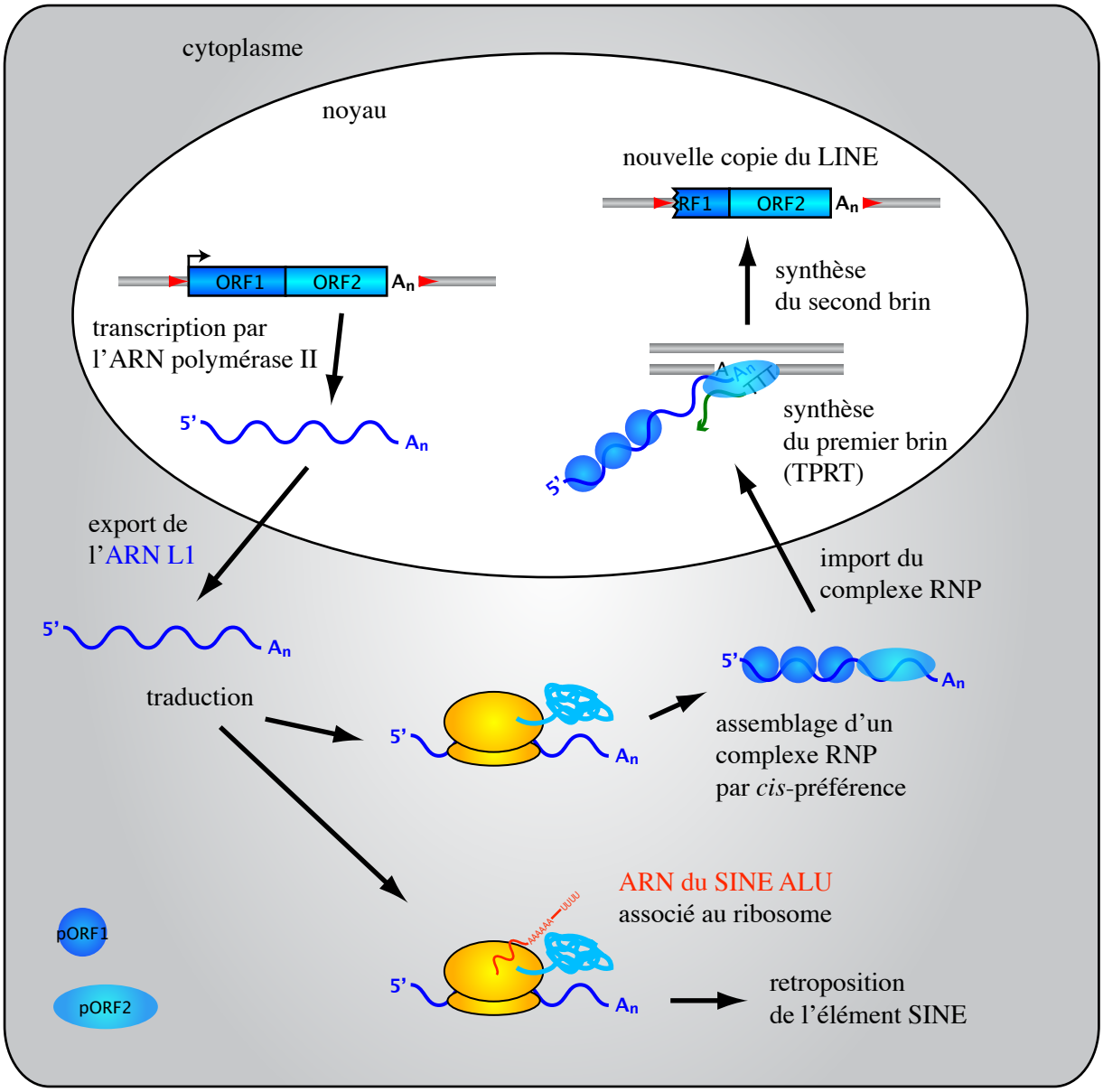


Figure 5. Modèle de rétroposition de l'élément L1

La duplication du site d'insertion est représentée par des flèches rouges.

Le parasitisme de la machinerie du LINE par le SINE Alu est représentée dans la partie inférieure de la figure.

Le LINE L1, qui est l'élément LINE le plus étudié, est présent dans le génome des mammifères. Cependant, des LINE qui lui sont apparentés ont été découverts dans d'autres génomes, par exemple chez les poissons (Duvernell and Turner 1998) ou les plantes (Lavie *et al.* 2002). Dans le génome humain, L1 est le seul LINE actif, il en existe environ 500 000 copies, ce qui représente 17 % de la masse du génome humain (Lander *et al.* 2001). Le premier cadre de lecture (ORF1) de L1 code une protéine de fixation à l'ARN, qui forme un complexe RNP (Ribo-Nucléo-Protéique) avec l'ARN L1 et présente une activité de chaperonne des acides nucléiques. Le deuxième cadre de lecture (ORF2) code une protéine présentant les activités enzymatiques requises pour convertir l'intermédiaire ARN en une nouvelle copie ADN de L1, c'est-à-dire une endonucléase et une transcriptase inverse. Les deux cadres de lectures sont nécessaires pour la rétroposition du LINE (Martin 2006).

La rétroposition de L1 (figure 5) commence par la transcription d'un élément actif, l'ARN est ensuite exporté dans le cytoplasme, puis traduit, donnant naissance à un complexe RNP. Par la suite ce complexe retourne dans le noyau où il permettra l'insertion de l'élément dans l'ADN de l'hôte par le biais du mécanisme de TPRT pour « Target-Primed Reverse Transcription ». L'endonucléase détermine le site d'insertion en créant une cassure simple-brin dans l'ADN cible, le plus souvent au niveau d'un motif 5' TT/AAAA. La transcriptase inverse initie ensuite la synthèse du premier brin d'ADNc au site d'insertion en utilisant comme amorce pour l'élongation l'extrémité 3' hydroxyle générée par l'endonucléase (Cost *et al.* 2002). La fin du processus, c'est-à-dire la coupure du deuxième brin d'ADN et la résolution de l'insertion, n'est pas bien connue, mais elle conduit à une nouvelle copie ADN double brin de l'élément L1 encadrée par une duplication du site d'intégration de taille variable.

Les éléments LINE peuvent être classés en deux deux groupes : stringent ou relâché, en fonction du mécanisme de reconnaissance de la séquence 3' de leur ARN par la transcriptase inverse. La transcriptase-inverse des LINE du type relâché, comme l'élément L1, ne reconnaît pas de séquence autre que la queue poly(A) (Moran *et al.* 1996). En revanche, la transcriptase-inverse des LINE de type stringent reconnaît spécifiquement leur séquence 3', seuls des SINE partageant cette séquence particulière peuvent de ce fait recruter la machinerie de ces LINE.

II.4.3) La rétroposition des SINE

Les LINE rétroposent selon un mécanisme de *cis* préférence, c'est-à-dire que les protéines de leur machinerie lient et rétroposent préférentiellement l'ARN qui les codent. En revanche il a été démontré que les protéines codées par l'ARN L1 peuvent aussi fonctionner en *trans* avec une bien moindre efficacité et rétroposer d'autres ARNm. Ces événements donnent naissance à des pseudogènes procéssés ou rétropseudogènes, c'est-à-dire une copie d'un gène existant sans intron ni promoteur, suivie d'une queue poly(A) et bordée de duplications du site d'insertion (Esnault *et al.* 2000).

Les éléments SINE utilisent cette complémentation en *trans* pour rétroposer. Mais comme le laisse penser le très grand nombre de copies d'Alu dans le génome humain, cet élément présente des caractéristiques qui permettent à son ARN d'interagir avec les protéines de L1 dans une configuration moléculaire équivalant à la relation *cis* de l'ARN L1. Dewannieux et al. (2003) ont montré que la rétroposition du SINE Alu dépendait effectivement de la machinerie du LINE L1 et ils ont aussi identifié certaines caractéristiques nécessaires à un fort taux de rétroposition. Ainsi la queue poly(A) de l'ARN Alu est indispensable à la rétroposition, ce qui renforce l'idée que les protéines du LINE reconnaissent leur substrat grâce à cette séquence, qui est la première à être rétrotranscrite lors du processus de rétroposition des LINE. La reconnaissance de cette séquence pourrait être directe ou se faire par l'intermédiaire des protéines de liaison de la queue poly(A) (PAB pour Poly(A) Binding-protein), qui stabilisent les ARNm et qui lient la queue poly(A) des SINE (Boeke 1997; Muddashetty *et al.* 2002). La forte diminution du taux de rétroposition d'un élément Alu tronqué de la séquence maximale lui permettant de conserver un promoteur actif et une queue poly(A) a montré que la région interne de l'élément est aussi impliquée dans l'efficacité du « parasitisme » de la machinerie des LINE. De plus, la retroposition d'Alu semble presque indépendante de la protéine codée par l'ORF 1 du LINE. Ensemble, ces données vont dans le sens du modèle (Boeke 1997) suivant : l'association du SINE Alu avec l'hétérodimère de protéines de la particule de reconnaissance du signal SRP9/14 (cf. paragraphe II.6 de la synthèse bibliographique) permettrait au SINE de passer outre la *cis* préférence des LINE en ciblant l'ARN Alu aux ribosomes et jouerait un rôle équivalent à celui de la protéine codée par l'ORF 1 dans la formation de la RNP. Mais à ce jour, il n'existe aucune démonstration directe de ce mécanisme moléculaire.

La majorité des SINE, et tous les SINE du génome d'*Arabidopsis thaliana* (Lenoir *et al.* 2001) sont dérivés d'ARN de transfert, certains sont présumés avoir une relation avec un

LINE comparable à la relation Alu/L1, mais il existe très peu d'informations sur les protéines qui pourraient s'associer à leurs ARN pour former un complexe RNP et potentiellement les cibler aux ribosomes. De même en ce qui concerne les couples SINE/LINE partageant une séquence 3', on sait que cette séquence est impliquée dans l'efficacité de rétroposition des SINE, mais le phénomène de *cis* préférence existant aussi pour les LINE de type stringent, il reste à trouver les autres facteurs impliqués dans le parasitisme de la machinerie du LINE.

II.5. Impact des SINE sur l'hôte

II.5.1) Impact génomique

II.5.1)a) Insertions mutagènes

En raison de leur abondance et de leur mobilité, même contrôlée, les éléments SINE ont un impact sur le génome hôte (Deragon and Capy 2000). L'une des conséquences la mieux connue est leur pouvoir mutagène. Bien que beaucoup de copies s'insèrent dans des régions non fonctionnelles, certaines s'insèrent à l'intérieur ou à proximité d'un gène. Une insertion peut inactiver ou perturber l'expression d'un gène de différentes façons, telle que la destruction du promoteur, la destruction ou l'interruption du cadre de lecture. Une insertion dans un intron peut également modifier l'épissage du transcrit. Les mutations qui ont lieu dans les cellules germinales sont génétiquement transmises à la descendance. Les maladies héréditaires liées à des insertions du SINE Alu identifiées chez l'homme correspondent pour la plupart à des insertions dans les exons ou dans les introns à proximité d'un exon, où elles perturbent potentiellement l'épissage (Deininger and Batzer 1999). Il semble qu'on trouve très peu d'insertions dans les régions 5' non-codantes ou codantes, probablement car une insertion dans ces zones est trop perturbante pour la fonction des gènes et de tels événements ne sont donc pas maintenus dans la population (Deininger and Batzer 1999). Des événements de ce type en lignées somatiques sont détectés chez l'homme lorsqu'ils entraînent une dérégulation du cycle cellulaire et une oncogénèse.

II.5.1)b) Effet sur l'architecture et l'expression des génomes

Les SINE sont à l'origine d'autres modifications génétiques par recombinaison homologue inégale. Les éléments SINE sont distribués tout au long du génome, la recombinaison entre des éléments présents à différents loci aboutit de ce fait à des remaniements, incluant des duplications, délétions et même des translocations si les éléments

se trouvent sur des chromosomes différents. Sur de longues périodes évolutives, notamment pour des éléments aussi représentés que l'élément Alu, ce type d'événements est probablement un mécanisme permettant de créer de la diversité génétique au sein du génome humain (Batzer and Deininger 2002).

Après l'insertion d'un élément SINE dans une région génique, il est ciblé par la machinerie de « silencing », ceci entraîne une modification du contexte chromatinien et peut avoir un effet sur l'expression des gènes situés à proximité de la nouvelle copie. Chez les plantes, l'élément SINE SB1 de *Brassica* est un élément qui a récemment transposé et certaines copies ne sont pas encore fixées dans la population. Arnaud et al. (2000) ont ainsi pu comparer l'état de méthylation de loci avant et après insertion de SB1. Ils ont montré que les éléments SB1 sont ciblés par la méthylation de l'ADN après leur insertion et que cette méthylation peut être propagée directionnellement aux séquences génomiques flanquantes (Arnaud *et al.* 2000). Les insertions d'éléments SINE peuvent donc avoir un effet délétère sur l'expression d'un gène. Les effets d'une insertion peuvent aussi être positifs. Ainsi, il existe des cas d'insertions ayant permis à des gènes d'acquérir de nouveaux profils d'expression dépendant du stade de développement, ou variant selon les tissus (Batzer and Deininger 2002 ; Kinoshita *et al.* 2007; Lunyak *et al.* 2007). Récemment, il a été montré chez la souris qu'une copie du SINE B2 participe au contrôle temporel et tissulaire de l'expression du gène de l'hormone de croissance (Lunyak *et al.* 2007). Ce phénomène est particulièrement intéressant, puisqu'il ne dépend pas de la seule présence de l'élément B2, mais de sa transcription par les ARN polymérase II et III. En effet, le SINE B2 est transcrit par la Pol III comme les autres SINE, mais certaines copies possèdent également un promoteur Pol II en orientation antisens. Dans le cas de l'hormone de croissance de souris, le gène est normalement réprimé par un contexte chromatinien fermé. La transcription simultanée du SINE B2 par les Pol II et Pol III permet d'établir une région barrière, ce qui contribue à la mise en place d'un contexte chromatinien ouvert.

Chez l'Homme, l'élément SINE Alu est abondamment présent dans les régions intergéniques et les introns. La présence des séquences Alu dans les ARNm influence l'expression des gènes au niveau post-transcriptionnel et la diversité des protéines. Ainsi, le SINE Alu participe de façon importante au phénomène d'exonisation, c'est-à-dire la transformation de séquences introniques en nouvel exon (Sorek 2007). Il contient des séquences ressemblant à des sites d'épissage, plus particulièrement quand il est inséré en

orientation antisens par rapport au gène. Les mutations affectant la séquence Alu peuvent donc faire apparaître de nouveaux sites reconnus par la machinerie d'épissage. Le plus souvent le phénomène d'exonisation des séquences Alu a un effet délétère sur la fonctionnalité de la protéine. Cependant, ces nouveaux sites d'épissage étant plus faibles, ils sont sujets au phénomène d'épissage alternatif, ce qui leur permet d'être tolérés. Ils peuvent alors contribuer à l'acquisition de nouvelles fonctions et/ou de nouveau éléments de régulation.

Le SINE Alu est aussi impliqué dans un processus d'édition des ARNm : la modification co- ou post-transcriptionnelle des adénosines en inosines (édition A>I) (Levanon *et al.* 2005). Cette modification est catalysée par une famille d'enzyme appelées ADAR (pour Adenosine Deaminase Acting on RNA). Un grand nombre des inosines identifiées dans le transcriptome humain résulte de l'édition A>I des éléments Alu présents dans les régions non-codantes des ARNm (UTR et introns). Le rôle exact de l'édition est sujet à spéculation, mais il pourrait affecter l'expression des gènes de plusieurs façons : en altérant la stabilité de l'ARNm, en entraînant la formation d'un nouveau transcrit par altération d'un signal d'épissage ou en étant à l'origine de la substitution d'un acide aminé de la protéine.

II.5.2) Rôle riborégulateur des transcrits SINE

L'impact des SINE sur les génomes, que ce soit en termes de plasticité ou d'expression des gènes, n'est pas lié à une fonction intrinsèque des séquences SINE, mais résulte de leur présence. En revanche, les transcrits SINE semblent avoir un rôle dans la physiologie cellulaire.

Depuis plus d'une dizaine d'années, de nombreux éléments suggèrent que les SINE pourraient avoir un rôle de riborégulateur cellulaire en tant qu'ARN pleine longueur transcrits par la Pol III. Il a ainsi été montré que le nombre de transcrits SINE pleine longueur augmente spécifiquement sous l'effet de différents stress, comme les chocs thermiques, le traitement par la cycloheximide (un inhibiteur de la synthèse des protéines eucaryotes) ou une infection virale. Cette accumulation est due à l'augmentation de la transcription des SINE par la Pol III et elle est spécifique des SINE puisque l'abondance d'autres transcrits Pol III, notamment l'ARNr 5S et l'ARN 7SL ne varie pas. Cette réponse au stress des SINE a été rapportée à propos de SINE dérivés de l'ARN 7SL (Schmid 1998), mais aussi d'ARNt, appartenant à des génomes de mammifères et d'invertébrés (Kimura *et al.* 1999).

Les éléments SINE forment de grandes familles de gènes dispersés tout au long du génome, notamment dans des régions géniques. Leur transcription est réprimée en conditions normales et il existe des mécanismes de régulation post-transcriptionnel pour limiter encore la présence de transcrits primaires ou pleine longueur des SINE dans la cellule. En condition de stress, la dérépression des SINE permettrait d'accumuler une grande quantité d'ARN n'ayant pas de fonction cellulaire en temps normal, ce qui pourrait activer des voies de réponse au stress. Les SINE seraient donc de véritables « gènes de stress ». De plus, les transcrits pleine longueur des SINE ont une durée de vie courte, ce qui présente l'avantage d'avoir une réponse rapidement réversible.

Plusieurs exemples de l'effet des transcrits SINE en réponse au stress ont été décrits chez l'homme. L'ARN Alu est capable de se lier à la Protéine Kinase R (PKR), entraînant une modification de son activité. La PKR sous sa forme active, s'autophosphoryle puis phosphoryle la sous-unité alpha du facteur 2 d'initiation de la traduction eucaryote (eIF2α) ce qui entraîne une inhibition de la synthèse des protéines. Ainsi l'accumulation de transcrits Alu pleine longueur après un stress permettrait à la cellule de « séquestrer » la PKR et donc d'augmenter la synthèse des protéines pour répondre à ce stress (Chu *et al.* 1998; Schmid 1998). De plus, Rubin et al. (2002) ont montré que l'ARN Alu pleine longueur était capable de stimuler la traduction d'ARNm néosynthétisés par une voie inconnue mais indépendante de la PKR. Les auteurs proposent un modèle dans lequel l'ARN Alu stimulerait l'expression des ARNm qui n'étaient pas jusqu'alors engagés dans la machinerie de traduction. Ceci permettrait à la cellule de traduire des ARNm spécifiques de la réponse au stress rapidement et une fois revenue en conditions normales de reprendre la traduction des ARNm engagés dans la machinerie de traduction avant le stress (Rubin *et al.* 2002).

Un effet d'un ARN SINE sur la transcription a récemment été décrit. Ainsi, il a été montré que, chez la souris, l'ARN SINE B2 est capable d'inhiber la transcription par la Pol II en condition de stress thermique (Allen *et al.* 2004). L'ARN B2 peut en effet s'associer spécifiquement à la Pol II, et cette association bloque l'assemblage des complexes de pré-initiation au niveau des promoteurs (Espinoza *et al.* 2004), probablement en empêchant l'association correcte de la polymérase avec l'ADN (Espinoza *et al.* 2007). Lors d'un choc thermique, la transcription Pol III des SINE est spécifiquement augmentée dans les cellules de souris, l'ARN SINE B2 joue alors son rôle d'inhibiteur *in vivo* et bloque la transcription Pol II entraînant la baisse de la transcription observée (Allen *et al.* 2004).

Certains ARN SINE de mammifères peuvent également avoir des fonctions cellulaires en conditions « normales ». BC1 chez les rongeurs et BC200 chez l'homme sont deux loci SINE exprimés spécifiquement dans les neurones (Cao *et al.* 2006). Bien que BC1 résulte de la rétroposition d'un SINE dérivé d'ARNt et BC200 résulte de la rétroposition d'un SINE dérivé de l'ARN 7SL, leurs transcrits adoptent des structures secondaires similaires et sont transportés activement dans les dendrites. Ces ARN semblent impliqués à différents niveaux dans la régulation de la traduction des neurones. Zalfa et al. (2003, 2005) ont montré que BC1 et BC200 interagissaient directement avec la protéine de déficience mentale du X Fragile (FMRP : Fragile X Mental Retardation Protein). La protéine FMRP est une protéine de liaison à l'ARN, qui s'associe à de nombreux ARNm et fonctionne comme répresseur de la traduction (Zhang and Broadie 2005). Les ARN BC1 et BC200 étant capables de former des appariements avec certains ARNm cibles de FMRP, il a été proposé qu'ils seraient impliqués dans la spécificité d'association de FMRP en servant d'adaptateur entre la protéine et ces ARNm cibles (Zalfa *et al.* 2003 ; Zalfa *et al.* 2005). Un rôle plus général d'inhibiteur de la traduction dans les neurones a été aussi été proposé pour BC1 et BC200 (Wang *et al.* 2002 ; Wang *et al.* 2005).

II.6. Les particules ribonucléoprotéiques (RNP) des SINE

Comme présenté précédemment, il existe différentes formes de transcrits SINE. Ces transcrits ont été identifiés dans le noyau mais aussi dans le cytoplasme. Les ARN SINE sont donc amenés à interagir avec différentes protéines qui vont avoir un effet sur leur stabilité, leur trafic entre les compartiments cellulaires et très probablement sur leur rétroposition et/ou leur fonction en tant que riborégulateur. Même si des RNP SINE ont été observées, très peu de protéines interagissant avec les ARN SINE ont été identifiées.

Un premier type d'interaction observé correspond à une reconnaissance de structures spécifiques de l'ARN SINE par une ou plusieurs protéines. C'est par exemple le cas de l'interaction de l'hétérodimère SRP9/SRP14 avec les ARN des SINE Alu et B1 (Chang and Maraia 1993). Ces SINE dérivent tous deux de l'ARN 7SL qui appartient à la particule SRP où il s'associe entre autres aux protéines SRP9 et SRP14. Ces deux éléments SINE ont conservé en partie le repliement de leur ARN ancestral, ce qui leur permet d'interagir avec les protéines SRP9 et SRP14 et contribue à la stabilisation de leurs ARN.

Un autre type d'interaction impliquerait la reconnaissance de séquences communes à un grand nombre de SINE. De nombreux éléments SINE possèdent une queue poly(A) et sont donc probablement associés à la protéine de liaison aux séquences poly(A) (PABP pour Poly(A) Binding Protein). L'existence de cette interaction a été montrée par la co-immunoprécipitation des ARN SINE BC1 et BC200 par des anticorps dirigés contre la PABP (Muddashetty *et al.* 2002 ; West *et al.* 2002).

De même la protéine La, connue entre autres pour lier la séquence poly(U) à l'extrémité 3' hydroxyle des transcrits Pol III, interagit très probablement avec les transcrits primaires des SINE. Chez les mammifères, il a été montré que la protéine La s'associait à certains transcrits SINE (Goodier and Maraia 1998 ; Kremerskothen *et al.* 1998). Même si cette interaction est transitoire et nucléaire, elle serait impliquée dans la maturation post-transcriptionnelle de ces ARN (Maraia *et al.* 1992 ; Goodier and Maraia 1998 ; Schmid 1998). La protéine La pourrait ainsi stabiliser les transcrits primaires des SINE en interagissant avec eux, et de ce fait moduler les mécanismes de maturation.

L'étude des différentes protéines interagissant avec les transcrits SINE apporte de nouvelles informations sur leur biosynthèse mais aussi leurs fonctions en tant que riborégulateurs cellulaires. La recherche de nouvelles protéines interagissant avec les ARN SINE est donc un aspect essentiel de l'étude des éléments SINE.

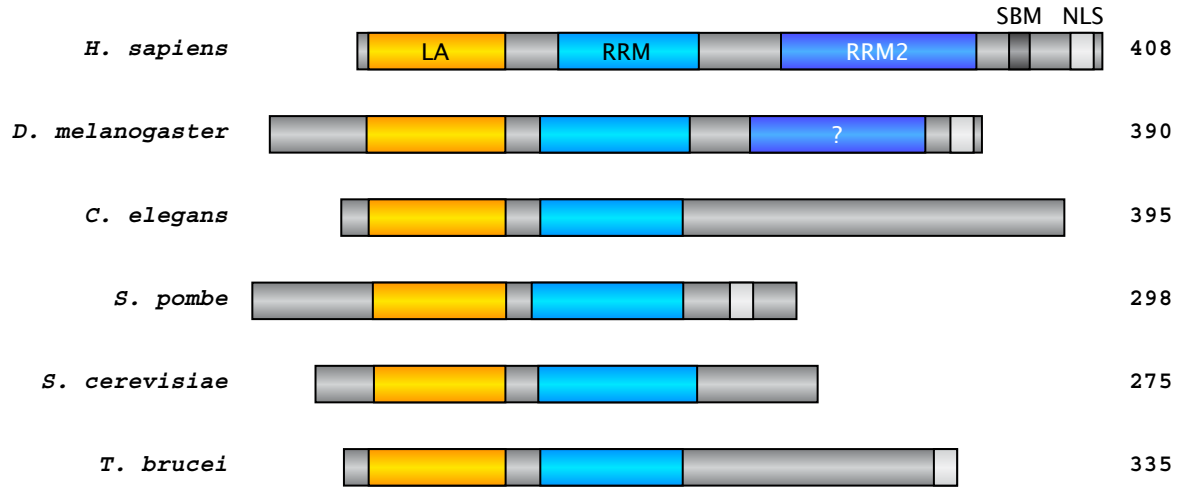


Figure 6. Représentation schématique de la structure des protéines La

Les protéines La de l'homme (*H. sapiens*), de la drosophile (*D. melanogaster*), de caenorhabditis (*C. elegans*), du trypanosome (*T. brucei*) et des deux levures *Saccharomyces cerevisiae* (*S. cerevisiae*) et *Schizosaccharomyces pombe* (*S. pombe*) ont été alignées.

Pour chaque protéine, la longueur en acides aminés est indiquée sur la gauche. Les domaines La et RRM sont respectivement figurés par des boîtes orange et bleues. Dans le cas de la protéine La humaine, un RRM atypique (RRM2) a été identifié dans la partie carboxy-terminale de la protéine. La modélisation de la région carboxy-terminale de la protéine La de la drosophile permet de détecter, mais avec une faible probabilité, ce second RRM.

SBM : motif basique court, NLS : signal de localisation nucléaire.

(adapté de Wolin & Cedervall, 2002)

III. La protéine La

La protéine La est une phosphoprotéine ubiquitaire chez les eucaryotes. Elle a d'abord été identifiée comme autoantigène chez des patients souffrant de maladies auto-immunes : le lupus érythémateux systémique et le syndrome de Sjögren (Mattioli and Reichlin 1974). Son nom « La » vient du nom d'un des patients, dans le sérum duquel ont été identifiés des anticorps dirigés contre elle. La protéine La est aussi connue sous le nom de SS-B pour antigène de type B du Syndrome de Sjögren. Des homologues ont depuis été caractérisés dans un grand nombre d'eucaryotes incluant la drosophile, les levures *Saccharomyces cerevisiae* (Yoo and Wolin 1994) et *Schizosaccharomyces pombe* (Van Horn *et al.* 1997), le trypanosome (Westermann and Weber 2000). La protéine La est une protéine très abondante et majoritairement nucléaire. Chez l'homme, elle a été estimée à 2×10^7 unités par cellule, ce qui en fait une protéine aussi abondante que les protéines ribosomiques (Wolin and Cedervall 2002). C'est une protéine multifonctionnelle dont le rôle le plus ubiquitaire est de fixer, au moins transitoirement, les transcrits néosynthétisés par la Pol III. Elle reconnaît la séquence UUU-OH 3' qui correspond au signal de terminaison de la transcription par la Pol III (Stefano 1984) et protège ces transcrits d'une dégradation exonucléasique. Elle est donc impliquée dans la maturation d'un grand nombre de transcrits Pol III et il semble qu'elle intervienne aussi dans la maturation de certains transcrits Pol II, tels que les petits ARN nucléaires et nucléolaires (Kufel *et al.* 2000 ; Xue *et al.* 2000). Il est aussi proposé qu'elle soit un facteur de terminaison de la transcription Pol III et qu'elle soit impliquée dans la traduction d'ARNm spécifiques endogènes et viraux (Meerovitch *et al.* 1993 ; Pellizzoni *et al.* 1996; Holcik and Korneluk 2000).

III.1. Données de structure

Des comparaisons de séquences de protéines La d'eucaryotes variés ont permis d'identifier les domaines conservés et ont révélé que la région amino-terminale de la protéine (les 200 premiers résidus amino-acides) était plus conservée que la région carboxy-terminale (Maraia and Intine 2001; Wolin and Cedervall 2002).

Ainsi la partie amino-terminale des différentes protéines La caractérisées présente deux domaines protéiques structurés (figure 6). Le premier est appelé domaine La ou motif La, il est long d'environ 60 acides aminés très conservés. La région amino-terminale contient

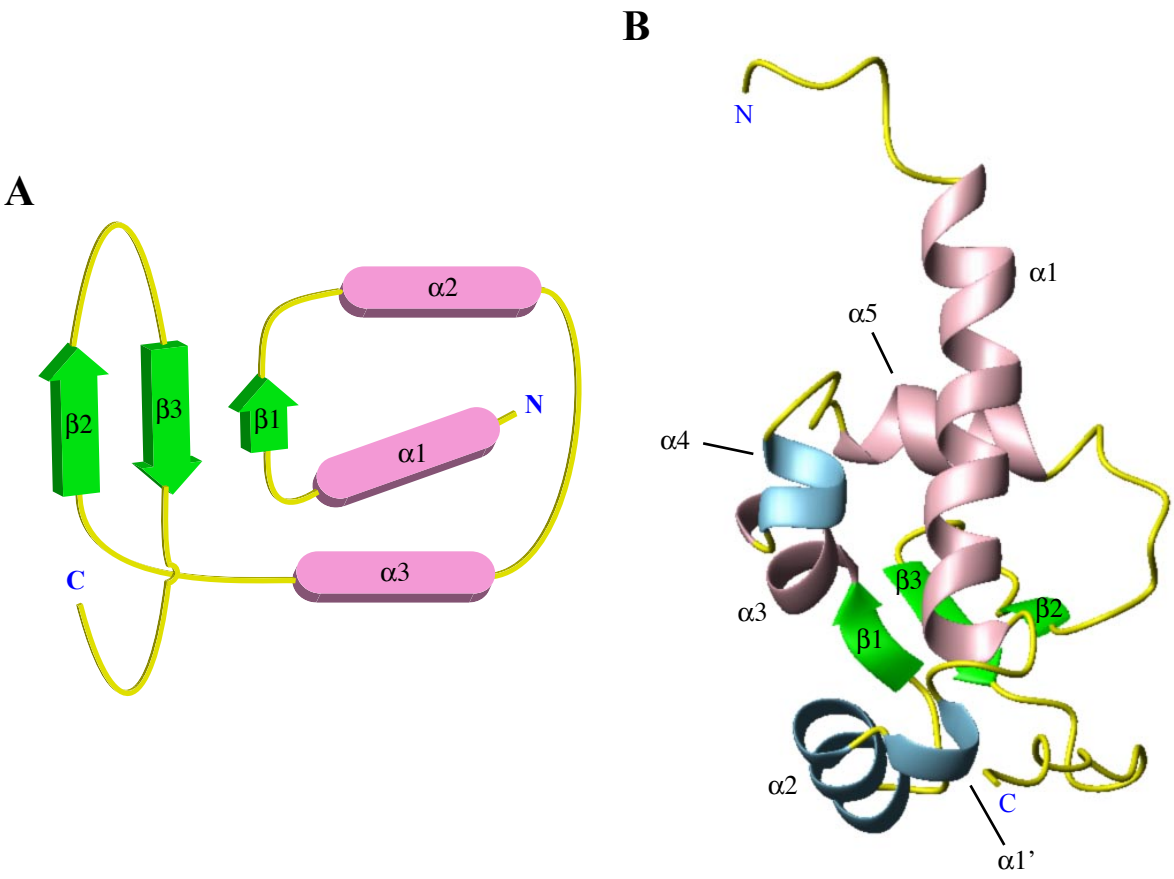


Figure 7. Représentation schématique de la structure du domaine La

(A) Diagramme représentant l'organisation d'un domaine winged helix canonique : α1-β1-α2-α3-β2-β3 (adapté de Gajiwala & Burley, 2000).

(B) Structure tertiaire du domaine La de la protéine humaine (adapaté de Alfano *et al*, 2004).

Les hélices colorées en bleu correspondent à des insertions par rapport au motif winged helix canonique.

ensuite un domaine RRM (pour RNA Recognition Motif) canonique de reconnaissance de l'ARN simple brin. La taille de la partie carboxy-terminale est très variable d'une espèce à l'autre, de 70 acides aminés chez les levures *S. cerevisiae* et *S. pombe* jusqu'à 220 acides aminés chez l'homme et le xénope (Wolin and Cedervall 2002). Cette portion de la protéine apparaît plus longue et plus complexe chez les mammifères. D'autres motifs ont été identifiés dans cette portion de la protéine humaine, notamment un deuxième RRM atypique et un motif basique court composé de 17 acides aminés majoritairement basiques (figure 6). Ce dernier est impliqué dans la localisation nucléolaire de la protéine (Intine *et al.* 2004). Il a été aussi été proposé que la protéine La pourrait former des homodimères grâce à un motif de dimerisation présent dans la région carboxy-terminale (Craig *et al.* 1997). Cependant cette hypothèse est controversée et des données récentes obtenues *in vitro* suggèrent que ce ne serait pas le cas (Jacks *et al.* 2003) (cf. paragraphe III.1.4 de la synthèse bibliographique). Enfin, à l'exception de la protéine de *S. cerevisiae*, le signal de localisation nucléaire des protéines La identifiées se trouve dans la partie carboxy-terminale (Wolin and Cedervall 2002).

On a aujourd'hui une image beaucoup plus précise de la structure de la protéine La grâce aux différentes études présentant la structure des domaines amino-terminal (Alfano *et al.* 2004) et carboxy-terminal (Jacks *et al.* 2003) de la protéine humaine obtenue par spectroscopie de résonance magnétique, la structure du cristal du domaine La de la protéine de *Trypanosoma brucei* (Dong *et al.* 2004), et enfin la structure du cristal du domaine amino-terminal de la protéine humaine associée à un ARN de neuf nucléotides (Teplova *et al.* 2006).

III.1.1) Le domaine La

Le domaine La, qui était parfois présenté comme un domaine RRM atypique, est en fait un domaine winged helix (Alfano *et al.* 2004; Dong *et al.* 2004). Ce type de domaine appartient à la famille des domaines hélice-tour-hélice (HTH), on le retrouve dans des protéines interagissant avec les acides nucléiques, le plus souvent l'ADN, et notamment dans les facteurs de transcriptions (Gajiwala and Burley 2000). Le domaine La se compose de six hélices α et d'un feuillet β à trois brins antiparallèles, organisés comme suit: $\alpha1-\beta1-\alpha3-\alpha5-\beta2-\beta3$ correspondant au module consensus winged helix et les hélices $\alpha1'$, $\alpha2$ et $\alpha4$ considérées comme des insertions dans ce consensus (figure 7). Les hélices

surplombées par la longue hélice α1, se replient au-dessus du feuillet β, qui forme donc une surface exposée sur la face opposée à l'hélice α1. Le seul domaine winged helix connu pour lier de l'ARN appartient au facteur d'élongation procaryotique selB qui est capable d'interagir avec une tige-boucle présente sur un ARN messager (Dong *et al.* 2004 ; Beribisky *et al.* 2007). Dans le cas de selB, les acides aminés impliqués dans la fixation de l'ARN n'ont pas été identifiés. Étant donné les similarités de séquence entre selB et les facteurs de transcription, il est proposé que la surface interagissant avec l'ARN corresponde à celle utilisée par beaucoup de facteurs de transcription pour fixer l'ADN. En comparaison, les résidus correspondant à cette surface sont peu conservés au sein du domaine La, il semble alors peu probable que celui-ci interagisse avec l'ARN de la même manière. En revanche certains résidus amino-acides du domaine La sont conservés dans toutes les protéines La caractérisées, ainsi que dans les autres protéines contenant un domaine La. Ces résidus appartiennent aux hélices α1, α1', α2 et à la boucle entre les hélices α3 et α4 et se retrouvent en surface sur une zone qui semble suffisamment grande pour contacter une ou deux bases (Dong *et al.* 2004). Dans le domaine La du trypanosome, Dong et al. ont identifié un acide aminé, l'aspartate 27, qui semble en grande partie responsable de la reconnaissance du groupement hydroxyle 3'OH.

Des protéines contenant un motif La, distinctes des vraies protéines La, ont été identifiées chez l'homme, la drosophile, *Caenorhabditis elegans* et la levure *S.cerevisiae* (Wolin and Cedervall 2002). À l'exception du domaine La, elles ne sont pas apparentées aux vraies protéines La. Pour la plupart, leur domaine La est en position centrale ou carboxy-terminale au contraire des vraies protéines La dont le domaine La est amino-terminal (Maraia and Intine 2001). Deux protéines contenant un domaine La ont été étudiées chez la levure : Sro9p et Slf1p (Sobel and Wolin 1999). Sro9p a été immunolocalisée dans le cytoplasme et les deux protéines s'associent à des poly-ribosomes, suggérant un rôle dans la traduction. Une autre protéine possédant un domaine La lie l'ARN de la télomérase chez les Euplotes, et semble impliquée dans l'assemblage et/ou la fonction de la télomérase (Aigner *et al.* 2003).

Le domaine La, défini aujourd'hui comme un domaine winged-helix, apparaît donc comme un domaine de liaison aux acides nucléiques représenté dans des protéines qui peuvent avoir des fonctions très diverses.

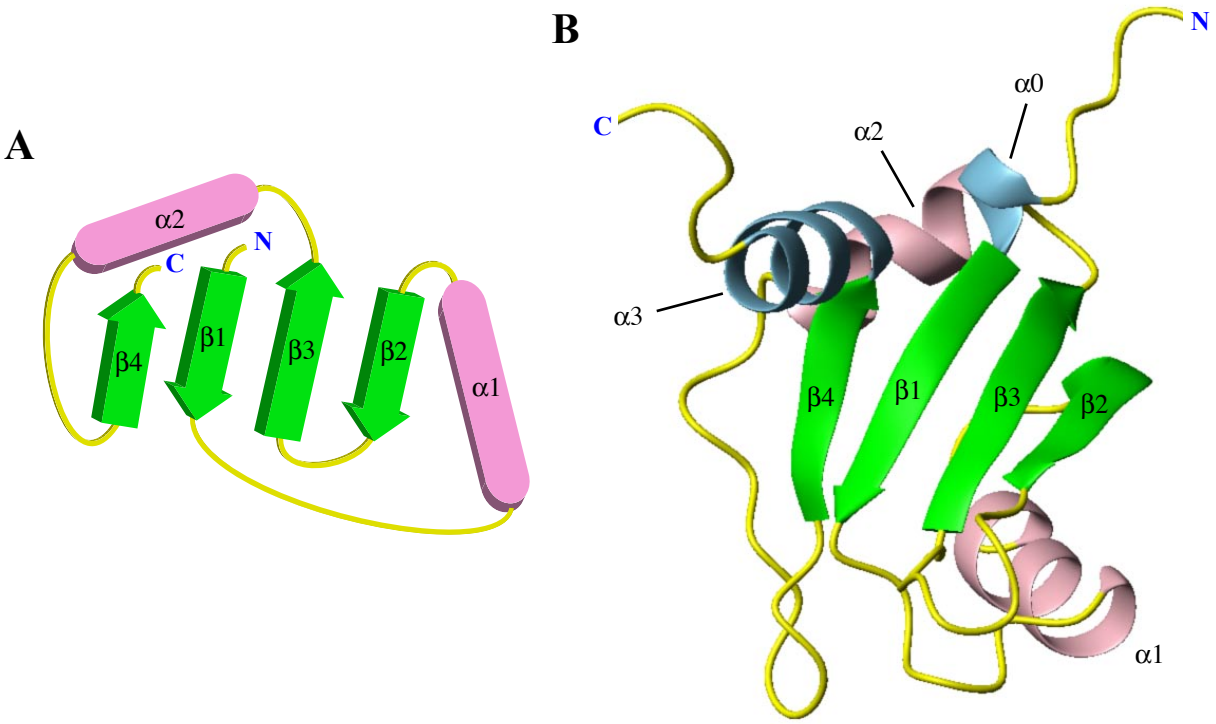


Figure 8. Représentation schématique de la structure du domaine RRM1

(A) Diagramme représentant l'organisation d'un domaine RRM canonique : β1-α1-β2-β3-α2-β4 (adapté de Kielkopf *et al*, 2004).

(B) Structure tertiaire du domaine RRM1 de la protéine humaine (adapaté de Alfano *et al*, 2004).

Les hélices colorées en bleu correspondent à des insertions par rapport au motif RRM canonique.

III.1.2) Le domaine RRM1

Le domaine RRM est un domaine de liaison à l'ARN simple brin, c'est un des motifs les plus fréquemment retrouvé dans les protéines eucaryotes (Maris *et al.* 2005). Les domaines RRM présentent des caractéristiques structurales communes, mais d'un RRM à l'autre les variations de la séquence protéique entraînent des affinité et spécificité de liaison à l'ARN très différentes.

Le domaine RRM1 de la protéine La humaine adopte le repliement d'un RRM classique. Ainsi, la séquence β1–α1–β2–β3–α2–β4–α3 se replie pour former un feuillet β à quatre brin antiparallèles arrangés dans l'espace β4 β1 β3 β2 avec les hélices α1 et α2 disposées en arrière (figure 8). La troisième hélice α3 est principalement hydrophile, elle surplombe en partie la surface de liaison à l'ARN du brin β4. Cette configuration est un peu inhabituelle, mais il semble que la face exposée du feuillet β reste en grande partie disponible pour les interactions avec l'ARN (Alfano *et al.* 2004). Le domaine RRM se caractérise notamment par deux motifs appelés séquences consensus RiboNucléoProtéique RNP1 et RNP2 de six et huit acides aminés pour la plupart chargés positivement ou aromatiques. Ces deux motifs appartiennent aux brins β1 et β3, ils se retrouvent donc sur la partie centrale du feuillet β où ils vont permettre l'interaction avec l'ARN. En effet il a été montré que ces motifs étaient nécessaires et suffisants pour lier une grande variété de molécules d'ARN (Maris *et al.* 2005). Les motifs RNP du domaine RRM1 contiennent deux des trois résidus aromatiques qui sont responsables de l'empilement des bases de l'ARN sur un RRM classique. De plus la boucle entre β2 et β3, qui est souvent impliquée dans la liaison de l'ARN, contient cinq résidus basiques qui peuvent créer des interactions électrostatiques avec le squelette sucre-phosphate de l'ARN. La conformation du domaine RRM1 semble donc compatible avec la liaison de l'ARN (Alfano *et al.* 2004).

III.1.3) Reconnaissance des extrémités UUU-OH des transcrits Pol III

L'extrémité 3' UUU-OH des transcrits Pol III est reconnue par la partie amino-terminale de la protéine La et les deux domaines La et RRM1 sont nécessaires à cette reconnaissance. La partie amino-terminale de la protéine La humaine a été cristallisée avec le nonamère d'ARN 5'-UGCUGUUUU-3' (Teplova *et al.* 2006). Les deux motifs La et RRM1 sont bien impliqués dans la liaison à l'ARN, mais d'une façon inattendue (Figure 9). Ainsi la liaison de l'ARN ne se fait pas à la surface du feuillet β du RRM1, mais c'est un bord du feuillet β en association avec le dos du motif winged helix qui maintiennent l'ARN. Cette

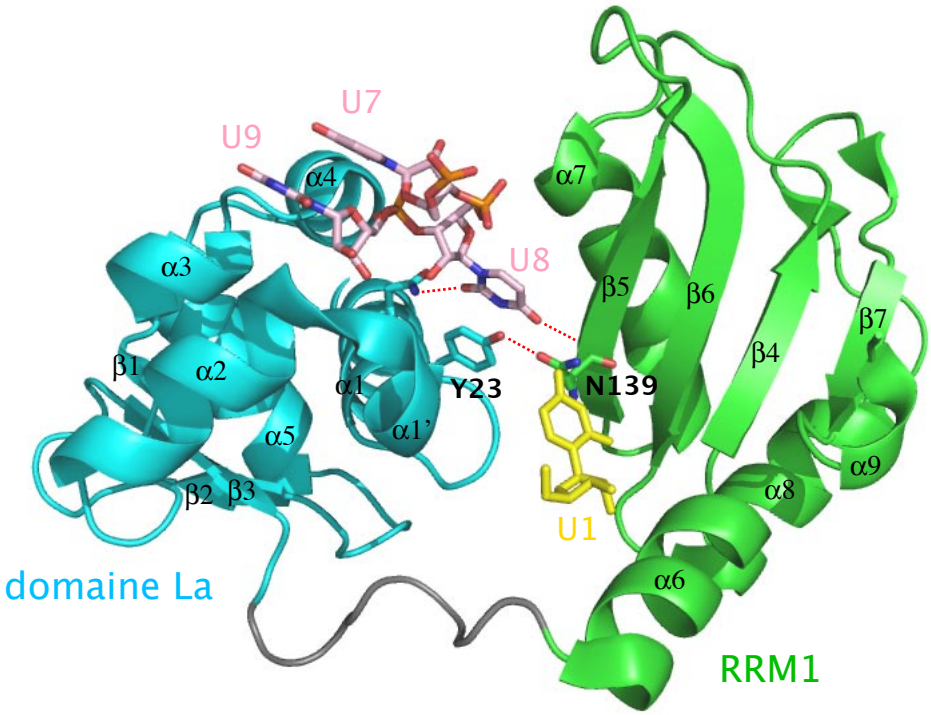


Figure 9. Structure tertiaire de la région amino-terminale de la protéine La humaine liée à l'extrémité 3' UUU-OH d'un nonamère d'ARN

Les domaines La et RRM1 sont respectivement figurés en bleu et vert.

Seule la première base (U1) et les trois bases terminales (U7-U8-U9) du nonamère d'ARN sont représentées.

(adapté de Teplova *et al*, 2006)

structure de la région amino-terminale présente une nouvelle utilisation des domaines La et RRM pour lier l'ARN et de ce fait elle laisse libres leurs surfaces de liaison les plus typiques, ce qui pourrait permettre d'autres interactions.

Dans le cristal, les trois bases terminales de la molécule d'ARN adoptent une conformation dans laquelle U7 s'empile au-dessus de U9 alors que U8 est décalée à l'opposé des deux autres bases. La liaison de l'ARN se fait principalement par U8 et U9. Une des caractéristiques de la protéine La est sa préférence pour les extrémités 3' hydroxyles par rapport aux extrémités 3' phosphates. Cette caractéristique lui permettrait de distinguer les transcrits Pol III naissants des ARN dégradés présentant des extrémités 3' phosphates, ainsi que les séquences UUU présentes aux extrémités de celles internes à une molécule d'ARN (Stefano 1984 ; Long *et al.* 2001). La reconnaissance de l'extrémité hydroxyle 3'OH de l'ARN est effectuée par la chaîne latérale de l'aspartate 33 du domaine La qui contacte par deux liaisons hydrogènes les groupements hydroxyles 2' et 3' OH et qui correspond à l'aspartate 27 identifiée dans le domaine La de la protéine du trypanosome (Dong *et al.* 2004). Le sucre et une partie de la base du nucléotide terminal (U9) interagissent avec deux acides aminés aromatiques du domaine La, mais la base ne semble pas reconnue spécifiquement étant donné qu'elle peut être remplacée par une cytosine, une adénine et même une déoxy-thymidine. En revanche la base du nucléotide U8 (l'avant-dernier) est impliquée dans des interactions avec le domaine La et le domaine RRM. C'est le premier cas de reconnaissance d'une séquence spécifique par un RRM qui n'utilise pas la surface de liaison typique du RRM.

Le nonamère d'ARN utilisé dans ces expériences est vraisemblablement un ligand atypique pour la protéine La. En effet, dans le cristal deux nanomères s'apparient en partie en duplex, ne laissant que les trois dernières bases (U7U8U9-OH) libres. Il ne permet pas de voir les interactions que pourrait établir un quatrième U avec la protéine et de même on ne peut pas exclure qu'un ARN simple brin pourrait interagir plus classiquement avec le feuillet β.

La liaison d'une extrémité UUU-OH 3' semble laisser libres les surfaces de liaison à l'ARN les plus typiques des deux domaines qui sont pourtant conservées. Il est possible que ces surfaces soient impliquées dans une activité plus complexe, notamment l'activité chaperonne qui permet la maturation de certains précurseurs d'ARNt anormaux (Curry and Conte 2006; Maraia and Bayfield 2006 ; Teplova *et al.* 2006) et qui sera décrite plus en détails par la suite (III. 3. 1) c).

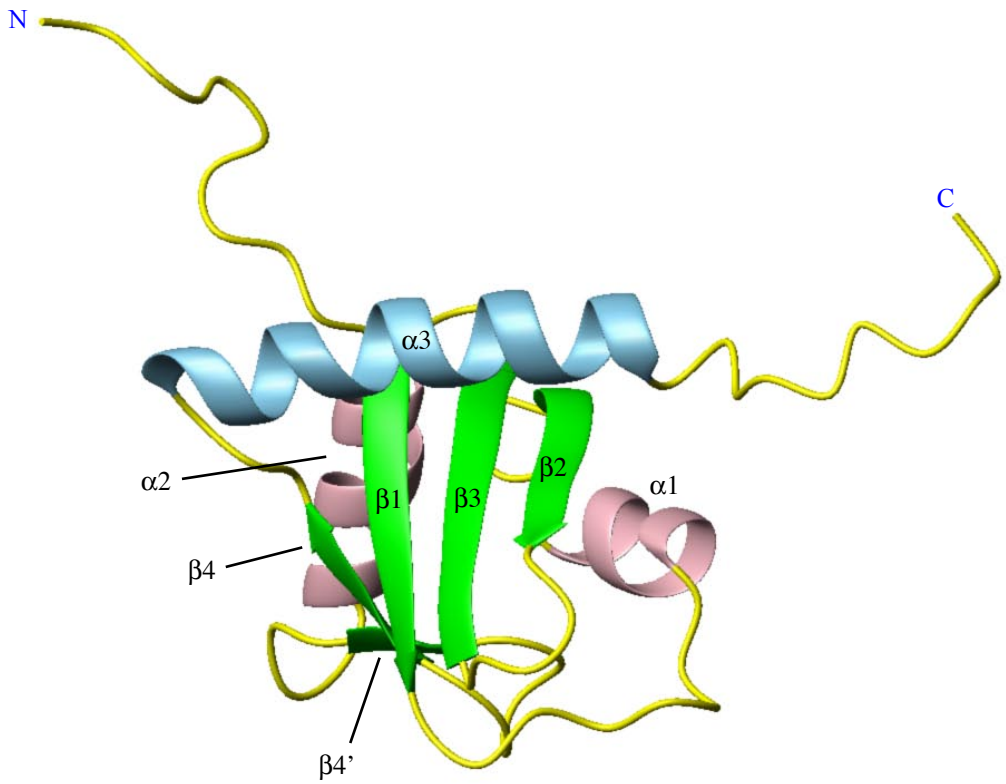


Figure 10. Représentation schématique de la structure du domaine RRM2 de la protéine La humaine

L'hélice colorée en bleu et le feuillet β4' correspondent à des insertions par rapport au motif RRM canonique (voir figure 8.A).

(adapté de Jacks *et al*, 2003)

III.1.4) Le domaine carboxy-terminal

La partie carboxy-terminale de la protéine La présente un faible degré de conservation entre les espèces, cependant des comparaisons de séquences suggèrent que le RRM2 serait présent dans les protéines de vertébrés et même dans les homologues de la drosophile et du trypanosome. L'analyse de la structure de la partie carboxy-terminale de la protéine La humaine indique que le domaine structuré globulaire que constitue le RRM2, est flanqué d'une courte extrémité NH2 flexible et d'une longue extrémité carboxy-terminale non structurée qui contient notamment la sérine 366, cible prédominante de la phosphorylation de la protéine (Jacks *et al.* 2003).

Le deuxième RRM de la protéine humaine est un RRM atypique (figure 10). Ainsi la séquence β1–α1–β2–β3–α2–β4'–β4–α3 se replie en un feuillet β à cinq brins antiparallèles se terminant par une longue hélice α. Au contraire des RRM classiques comme le RRM1, le RRM2 présente des acides aminés acides au niveau de la surface du feuillet β. De tels résidus ne sont pas favorables aux interactions avec le squelette sucre-phosphate de l'ARN et ne sont normalement pas impliqués dans des liaisons hydrogènes avec les bases. La longue hélice qui est reliée par un connecteur très court au brin β4 se replie au-dessus du feuillet β bloquant de ce fait la surface de liaison de l'ARN. Ces éléments de structure pourraient expliquer le fait que le RRM2 ne soit pas capable de lier de l'ARN. Cependant il est possible que ce RRM contribue à la liaison de molécules d'ARN de grande taille et/ou structurées ou qu'il soit impliqué dans des interactions avec d'autres protéines.

Une autre conclusion des expériences menées sur la partie carboxy-terminale est que le motif de dimérisation de la protéine La n'est pas fonctionnel au contraire de ce qui avait été proposé précédemment et que la protéine reste monomérique en solution (Jacks *et al.* 2003).

III.2. Localisation subcellulaire

Les expériences d'immunofluorescence menées dans un grand nombre de tissus et d'espèces ont révélé que la protéine La était majoritairement localisée dans le noyau, à la fois dans le nucléoplasme et le nucléole (Wolin and Cedervall 2002). Mais la protéine La est aussi présente dans le cytoplasme. La part cytoplasmique de protéine La est estimée à 2 à 4 % de la quantité totale dans les ovocytes et les embryons de xénope (Cardinali *et al.* 2003).

Différents signaux de localisation de la protéine La ont depuis été identifiés. La protéine humaine présente un signal de localisation nucléaire (NLS pour Nuclear Localization

Signal) situé à son extrémité carboxy-terminale, des acides aminés 383 à 408, et une séquence nécessaire à la localisation nucléolaire (NoLS pour Nucleolar Localization Signal), des acides aminés 323 à 354 (Horke *et al.* 2004) (cf. paragraphe III.3.1.b de la synthèse bibliographique). Le NLS de la protéine La de *S.cerevisiae* ne se trouve pas dans la portion carboxy-terminale de la protéine comme celui de la protéine humaine ou de la protéine de *S. pombe* et de plus la protéine de *S.cerevisiae* utilise un système d'import nucléaire différent (Rosenblum *et al.* 1997).

La protéine La humaine présente aussi un signal de rétention nucléaire (NRE pour Nuclear Retention Element) au niveau de l'hélice $\alpha 3$ du RRM2 (Intine *et al.* 2002) capable de masquer ou de surpasser une activité d'export nucléaire localisée au niveau du RRM1(Bayfield *et al.* 2007). Un signal NRE a aussi été identifié dans la séquence carboxy-terminale du RRM1 de la protéine Sla1p de *S. pombe*. Des expériences menées dans des cellules de *S. pombe* ont montré que dans les levures exprimant la protéine La humaine ou la protéine Sla1p dépourvues de leurs NRE respectifs, la protéine est majoritairement localisée dans le cytoplasme. Lorsque ces cellules sont traitées avec de la leptomycine B qui est un inhibiteur de la voie exportin 1, la protéine redevient nucléaire. Après traduction, la protéine La est donc correctement importée dans le noyau grâce à son NLS, mais en l'absence de NRE, l'activité d'export nucléaire du RRM1 entraîne sont export vers le cytoplasme par la voie exportin 1 (Bayfield *et al.* 2007). De même, une activité d'export nucléaire du domaine RRM1 de la protéine La de *S. cerevisiae* a été précédemment suspectée, ce qui suggère que cette activité d'export nucléaire du RRM1 serait aussi conservée chez *S. cerevisiae* (Rosenblum *et al.* 1998). En revanche lorsqu'on exprime les protéines La de l'humain ou de *S. pombe* dépourvues de leurs NRE respectifs en cellules de primate, le traitement par la leptomycine B ne restaure que partiellement la localisation nucléaire des protéines. Cette différence de localisation peut être due à une cinétique différente ou à l'utilisation d'une voie d'export différente chez les primates. De même, on ignore aussi si l'activité d'export nucléaire du RRM1 est responsable des déplacements de la protéine La humaine du noyau vers le cytoplasme décrits préalablement (Bayfield *et al.* 2007).

Ainsi, des mouvements de navette de la protéine La entre le noyau et le cytoplasme ont été mis en évidence dans des hétérocaryons résultants de la fusion d'une cellule humaine avec une cellule de souris (Fok *et al.* 2006). Dans ces expériences, la synthèse protéique a été bloquée chimiquement et la protéine La humaine, provenant donc du noyau humain, est observée dans le noyau de souris 4 heures après la fusion des deux cellules. Ces mouvements de navette du noyau vers le cytoplasme ont d'abord été constatés dans des hétérocaryons de

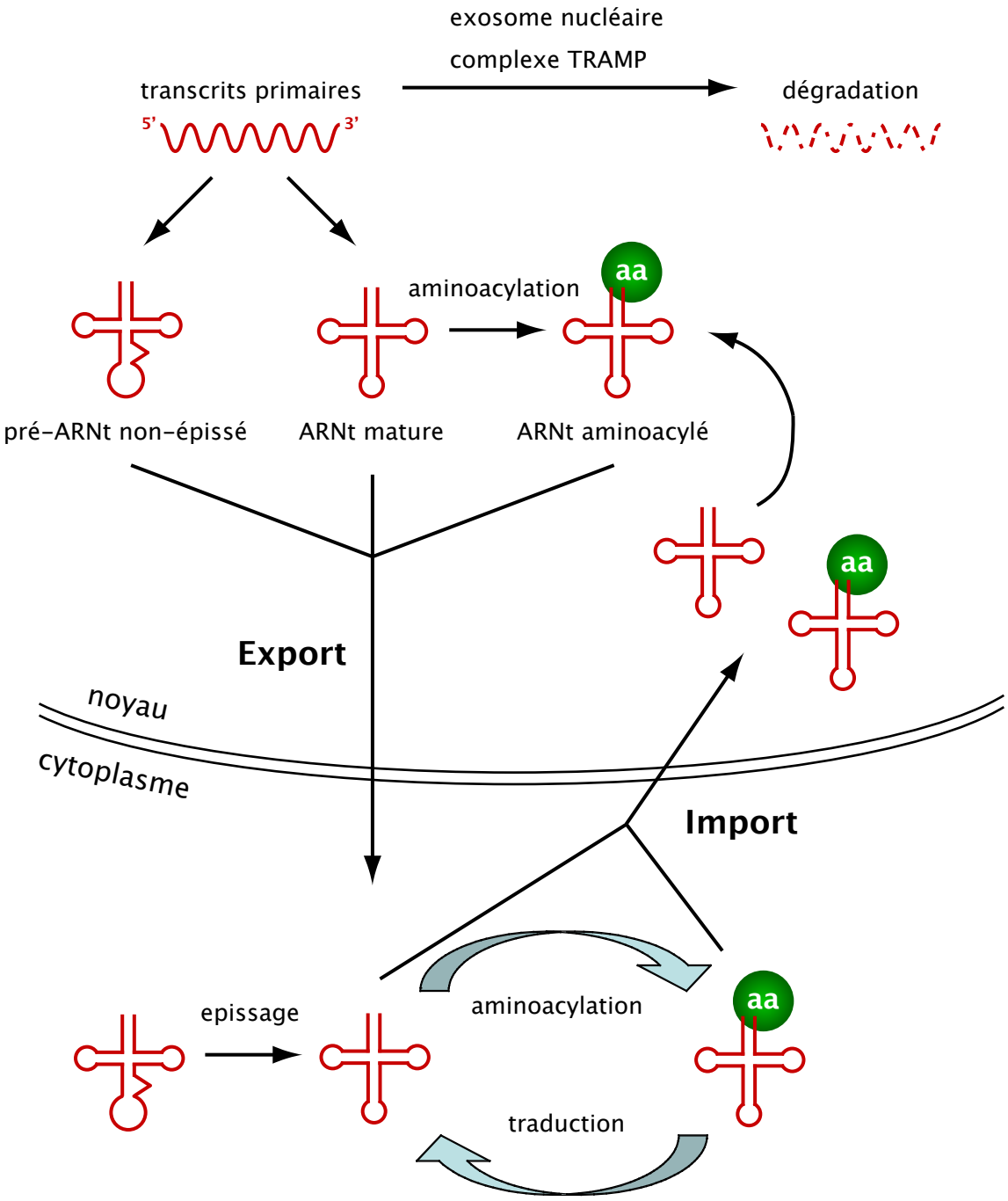


Figure 11. Mouvement des pré-ARNt et des ARNt matures chez *Saccharomyces cerevisiae*

(adaptée de Dahlberg et Lund, 2005)

cellules de souris avec des cellules humaines infectées par le virus d'Epstein-Barr, mais ils ont aussi été observés dans des hétérocaryons de cellules de souris avec des cellules HeLa et des cellules embryonnaires de foie non-infectées, ils ne semblent donc pas spécifiques d'un type cellulaire ni de l'application d'un stress. Les mouvements ne sont pas non plus sensibles à la leptomycine B, l'export de la protéine La observé ici doit donc être indépendant de la voie exportin 1 ou passer par plusieurs voies. La protéine La qui est majoritairement nucléaire dans les cellules de mammifères apparaît donc capable de sortir puis de retourner dans le noyau (Fok *et al*. 2006).

III.3. Les fonctions de la protéine La

III.3.1) Fonctions dans la biosynthèse des ARNt

III.3.1)a) Processus de maturation des ARNt

Tous les ARNt sont transcrits par la Pol III sous forme de précurseurs (pré-ARNt) présentant à leurs extrémités 5' et 3' des séquences qui n'appartiennent pas à l'ARNt mature. De plus, certains de ces précurseurs présentent des introns. Les précurseurs des ARNt subissent donc un processus de maturation complexe et sont soumis à des contrôles qualité tout au long de cette maturation.

Pour la plupart des ARNt eucaryotes, la maturation commence avec le clivage de la séquence 5' par la RNase P, une ribonucléoprotéine (O'Connor and Peebles 1991). La séquence 3' est ensuite éliminée par une activité endonucléolytique ou exonucléolytique. La voie endonucléolytique serait catalysée par une enzyme, la RNase Z, dont le rôle dans la maturation des ARNt a récemment été mis en évidence chez les plantes (Schiffer *et al.* 2002). En effet, des expériences *in vitro* ont montré que les RNase Z de blé et d'*Arabidopsis*, sont capables de cliver l'extrémité 3' d'un pré-ARNt. Des homologues de cette enzyme ont été détectés à la fois chez les procaryotes et les eucaryotes. Chez la drosophile, on retrouve un seul homologue de la RNase Z. Des expériences d'ARNi ciblant le transcrit de la RNase Z ont montré qu'en présence d'une quantité réduite de la protéine, les précurseurs 3' étendus des ARNt nucléaires et mitochondriaux s'accumulaient. Ces observations *in vivo* confirment que cette enzyme a bien un rôle dans l'élimination endonucléolytique de l'extrémité 3' des pré-ARNt (Dubrovsky *et al.* 2004). Après cette étape de maturation, un motif CCA est ajouté à l'extrémité 3' des ARNt par l'ARNt-nucléotidyle-transférase, ce motif étant nécessaire à l'aminoacylation des ARNt (Hopper and Phizicky 2003). Tout au long de leur maturation, les

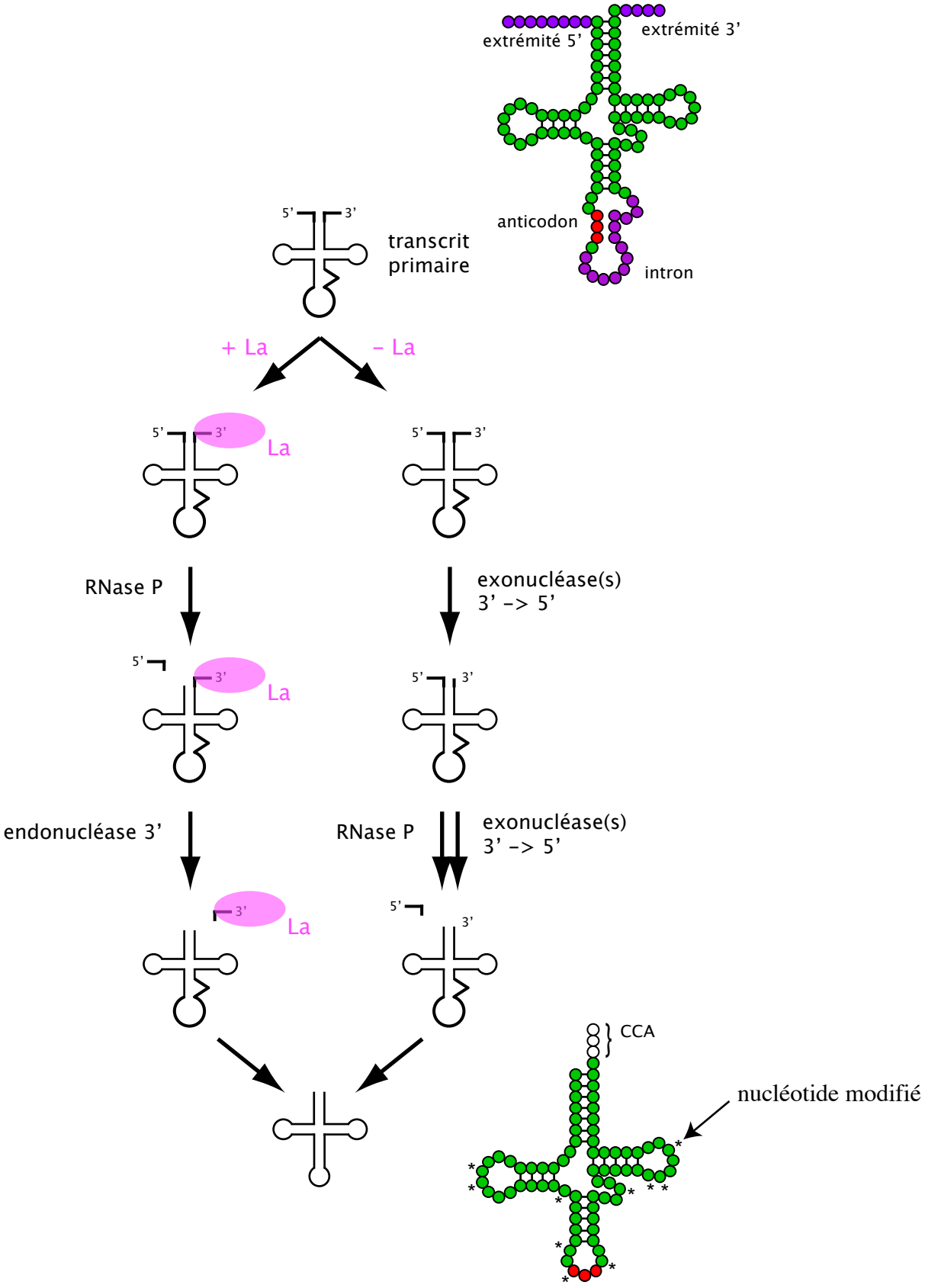


Figure 12. Modèle de la maturation des pré-ARNt en levures sauvages et en levures dépourvues de protéine La

(adaptée de Wolin et Cedervall, 2002)

ARNt subissent un grand nombre de modifications post-transcriptionnelles. Celles-ci sont notamment impliquées dans la stabilisation des ARNt (Wolin and Matera 1999 ; Hopper and Phizicky 2003). Enfin, ils sont exportés du noyau vers le cytoplasme, et il a été démontré que l'amino-acylation nucléaire, même si elle n'est pas indispensable, favorise cet export (Arts *et al.* 1998 ; Lund and Dahlberg 1998).

Des découvertes récentes ont dévoilé des phénomènes de maturation et de circulation des ARNt beaucoup plus complexes que prévu (figure 11). Ainsi contrairement à ce qui est connu chez les vertébrés, l'épissage des introns des ARNt de *S. cerevisiae* n'est pas nucléaire mais cytoplasmique (Yoshihisa *et al.* 2003 ; Yoshihisa *et al.* 2007). Cette localisation de l'épissage permet d'expliquer pourquoi des ARNt non-épissés s'accumulent dans le noyau lorsque la machinerie d'export est mutée, mais pas pourquoi dans certaines conditions, des ARNt épissés se retrouvent dans le noyau. Ces observations ont été le point de départ de la découverte d'un mouvement des ARNt du cytoplasme vers le noyau, un phénomène qui est conservé chez les vertébrés (Shaheen and Hopper 2005; Takano *et al.* 2005), mais dont le rôle reste à élucider.

III.3.1)b) La protéine La est nécessaire à la maturation normale des précurseurs des ARNt

Chez les levures *S. cerevisiae* et *S. pombe*, la protéine La n'est pas nécessaire à la viabilité (Yoo and Wolin 1994 ; Lin-Marq and Clarkson 1995 ; Van Horn *et al.* 1997), ce qui a permis d'étudier l'impact de son absence sur la maturation des ARNt (figure 12). Les homologues de la protéine La sont nommés Lhp1p chez *S. cerevisiae* et Sla1p chez *S. pombe*. Dans les cellules sauvages, la maturation commence par la coupure de l'extrémité 5' par la RNase P. L'extrémité 3' liée par Lhp1p est ensuite enlevée par une activité endonucléolytique. Dans les levures où Lhp1p est absente, il y a un changement dans l'ordre des événements de maturation et dans les mécanismes impliqués. Ainsi dans les cellules mutantes pour le gène *LHP1*, la maturation des ARNt commence par la dégradation de l'extrémité 3' par des exonucléases (Yoo and Wolin 1997). Cette dégradation exonucléasique s'arrêterait dans un premier temps au niveau des appariements entre les bases de la séquence 5' et les bases restantes de la séquence 3'. Une fois la maturation de l'extrémité 5' effectuée par la RNase P, les exonucléases termineraient la digestion de l'extrémité 3' (Wolin and Cedervall 2002). Les levures *S. pombe* mutantes pour le gène *SLA1* présentent les mêmes altérations de la maturation des ARNt que celles décrites chez *S. cerevisiae* (Van Horn *et al.*

1997). D'autre part, la protéine La humaine est capable de complémenter les défauts de maturation des ARNt liés à la mutation du gène *SLA1* de *S. pombe* (Van Horn *et al.* 1997). Ce résultat suggère que les mécanismes de maturation de l'extrémité 3' des ARNt, ainsi que le rôle de la protéine La dans ce processus, sont conservés chez les animaux.Chez *S. cerevisiae*, les étapes précoces de maturation des précurseurs des ARNt, notamment la coupure de l'extrémité 5' par la RNase P, s'effectuent dans le nucléole (Bertrand *et al.* 1998). La localisation nucléolaire de la protéine La, des molécules de pré-ARNt ainsi que de la RNase P indique qu'il en serait de même chez l'Homme (Horke *et al.* 2004).

Si la protéine La n'est pas indispensable pour la formation d'ARNt matures dans les levures sauvages, elle devient indispensable à la viabilité en présence de certaines mutations dans des gènes codant des ARNt essentiels ou des composants de la voie de biosynthèse des ARNt. Ce phénomène, appelé co-létalité, peut se définir comme la conjonction létale de deux mutations qui ne sont pas létales par elles-mêmes, il met le plus souvent en évidence les interactions fonctionnelles voire la redondance de deux facteurs. Ainsi une mutation identifiée dans le gène codant l'ARNt Sérine (CGA) au niveau de la tige de l'anticodon perturbe la structure de l'ARN et rend la présence de la protéine La indispensable à la maturation du précurseur (Yoo and Wolin 1997). L'hypothèse privilégiée aujourd'hui pour expliquer ce phénomène, est que la liaison de la protéine La à la séquence 3' protégerait l'ARNt Ser(CGA) des exonucléases, qui en son absence ne s'arrêteraient pas normalement et entraîneraient la dégradation du précurseur (Yoo and Wolin 1997).

III.3.1)c) Rôle de chaperonne

Un crible génétique a été effectué chez la levure *S. cerevisiae* pour identifier des conditions ou des contextes génétiques dans lesquels la protéine La devient nécessaire à la viabilité (Xue *et al.* 2000). Une mutation identifiée lors de ce crible touche le gène codant l'ARNt Arginine (CGG). Dans une souche portant cet ARNt muté et le gène *LHP1* sous le contrôle d'un promoteur inductible, lorsque l'on éteint le promoteur, on voit qu'après 12h, l'ARNt Arg mature continue d'être produit, mais qu'il n'est plus aminoacylé (Chakshusmathi *et al.* 2003). Étant donné que les extrémités de l'ARNt sont maturées, la protéine La n'est pas ici requise pour protéger l'ARN des dégradations exonucléasiques, c'est donc une autre fonction qui la rend indispensable. La protéine La ne fixe pas l'ARNt mature et elle n'intervient probablement pas directement sur l'aminoacylation. Des analyses de la structure

de l'ARNt muté ont montré que l'ARN adoptait une conformation anormale en hélice *in vitro* et très probablement *in vivo*, et cette structure anormale ne serait pas un substrat pour l'aminoacylation. Dans la souche mutante pour l'ARNt Arg et la protéine Lhp1p, l'introduction d'un gène d'ARNt Arg portant des mutations qui affaiblissent cette structure anormale en hélice de l'ARN rend la souche viable, ce qui suggère que la protéine La aurait un rôle de chaperonne moléculaire précédemment proposé (Yoo and Wolin 1997). Elle permettrait ainsi de stabiliser la structure correcte du précurseur de l'ARNt Arg muté, l'empêchant de se replier anormalement. Une fois la maturation 3' effectuée et donc la protéine La enlevée, la structure correcte de l'ARNt serait maintenue par d'autres protéines, telles que les enzymes de modifications, les ARNt-synthétases (Chakshusmathi *et al.* 2003). Dans beaucoup de transcrits Pol III comme les précurseurs d'ARNt, l'ARN 5S, l'ARN de la particule SRP, les séquences 3' peuvent créer des appariements avec des séquences internes ou 5'. La protéine La pourrait favoriser, par son rôle de chaperonne un repliement correct de ces ARN.

La protéine La est aussi nécessaire à la viabilité de souches de levures présentant des mutations dans les gènes codant les enzymes de modification des ARNt (Copela *et al.* 2006). L'ARNt-1-méthyl-adénosine-58-transférase est l'enzyme responsable du transfert d'un groupement méthyle au niveau de l'azote en position 1 de l'adénosine 58 des ARNt. Elle se compose de deux sous-unités : Trm6 (aussi nommée Gcd10) et Trm61 (ou Gcd14). Le gène *TRM61* est un gène essentiel. Des allèles mutants hypomorphes de ce gène ont été identifiés et l'analyse par hybridation northern blot de certains de leurs ARNt a montré que ces mutants présentaient des défauts de maturation de l'ARNt Méthionine initiateur, Met(i) (Calvo *et al.* 1999). Une souche dont le génome présente à la fois une mutation hypomorphe de *TRM61* et une délétion du gène *LHP1* a un taux de croissance très faible à 28°C, voir nul à 37°C. L'analyse des quantités d'ARNt montre que, dans cette souche, les quantités d'ARN précurseur et mature de l'ARNt Met(i) sont fortement réduites. Ainsi il semble que l'absence de méthylation au niveau de l'adénosine 58 perturbe la structure du précurseur de l'ARNt Mét(i) et requiert la protéine La pour protéger l'extrémité 3' de l'ARNt des dégradations par les exonucléases lors du contrôle qualité (cf. ci-dessous) et obtenir sa maturation correcte (Calvo *et al.* 1999 ; Copela *et al.* 2006).

Les souches mutantes pour la plupart des enzymes de modification des ARNt, comme les souches mutantes pour Lhp1p, n'ont pas de phénotype en conditions normales de croissance. En revanche, des souches mutantes à la fois pour Lhp1p et une enzyme de

modification comme l'ARNt pseudo-uridine synthétase Pus3 ou l'ARNt méthyltransférase Trm1 ne sont pas viables à 37°C (Copela *et al.* 2006). De même, une souche portant une mutation qui réduit l'activité de l'arginyl-ARNt-synthétase et une mutation dans le gène *LHP1* n'est pas viable alors que la première mutation seule n'entraîne qu'une baisse du taux de croissance. Ce phénomène de colétalité observé chez la levure est à l'origine de l'idée selon laquelle la structure correcte des ARNt est assurée par de multiples mécanismes, incluant l'acquisition de nucléotides modifiés et la liaison par la protéine La.

Récemment des processus de contrôle qualité des ARN non-codant ont été découvert (pour une revue récente voir Reinisch and Wolin 2007). L'une de ces voies, décrite chez la levure *S. cerevisiae* fait intervenir un complexe appelé TRAMP (pour Trf4/Air2/Mtr4p Polyadenylation complex) (LaCava *et al.* 2005). En polyadénylant les ARN non-codants anormaux dans le noyau, ce complexe entraînerait leur dégradation par l'exosome. Il a été montré que cette voie était responsable de la dégradation de différents ARN non-codants anormaux et notamment du précurseur de l'ARNt méthionine initiateur hypométhylé présenté précédemment (Kadaba *et al.* 2004 ; Kadaba *et al.* 2006). Cette voie n'est pas la seule à dégrader les ARN non-codants anormaux. Ainsi, certains ARNt matures instables sont dégradés indépendamment du complexe TRAMP ou de l'exosome, suggérant qu'il existe plusieurs voies de dégradation des ARN non-codants anormaux chez *S. cerevisiae* (Alexandrov *et al.* 2006; Copela *et al.* 2006).

On ignore comment les ARN défectueux sont reconnus et ciblés vers les voies de dégradation. Plusieurs hypothèses ont été avancées (pour une revue récente voir Reinisch and Wolin 2007). L'une des possibilités est que des composants du contrôle qualité soient capables de reconnaître des caractéristiques, encore inconnues, communes aux ARN mal repliés. Une alternative serait que les protéines du contrôle qualité ne sont pas capables de distinguer les ARN anormaux, mais que les ARN normaux étant associés en RNP avec des protéines seraient protégés des exonucléases.

La protéine La de levure devient indispensable lorsque des gènes d'ARNt présentent des défauts ou que des gènes codant des enzymes de modification sont mutés. On peut proposer que la protéine La, en liant les précurseurs de ces ARNt anormaux ou anormalement modifiés, joue son rôle de chaperonne ou permet à d'autres protéines de les stabiliser et les protège ainsi de la dégradation par le contrôle qualité nucléaire. Ce rôle pourrait être étendu à d'autres ARN. En effet, une souche de levure exprimant un ARNr 5S muté présente un taux de croissance réduit, en raison de la sous-accumulation de cet ARN. La surexpression de la

protéine Lhp1p complémente en partie le défaut de croissance en permettant une meilleure accumulation de l'ARNr 5S muté (Dechampesme *et al.* 1999).

III.3.1)d) Contribution des domaines La et RRM1 de la protéine La à ses activités dans la maturation des ARNt

La protéine La a donc deux activités distinctes dans la maturation des ARNt : une activité de protection des extrémités 3' et une activité de chaperonne. Une étude récente a évalué la contribution respective des domaines présents dans la région amino-terminale de la protéine dans ces activités (Huang *et al.* 2006). Ces travaux ont été réalisés avec la protéine La humaine exprimée en cellules de levure *S. pombe*. L'activité la plus basique de la protéine La consiste à protéger l'extrémité 3' des ARNt contre l'activité des exonucléases et nécessite une reconnaissance de l'extrémité 3' UUU-OH par le domaine La. La mutation des acides aminés aromatiques du feuillet β du RRM1, qui sont normalement impliqués dans la liaison à l'ARN par ce type de domaine, n'empêche pas cette activité. Ce résultat est en accord avec les données de structure présentées précédemment (III. 1. 3). La maturation de certains précurseurs d'ARNt déficients est dépendante de l'activité chaperonne de la protéine La. Comme l'activité de protection des extrémités 3', l'activité de chaperonne requiert une reconnaissance de l'extrémité 3' UUU-OH par le domaine La, mais elle nécessite aussi un feuillet β du RRM1 intact (Huang *et al.* 2006). Ces données sont donc en accord avec un modèle de maturation des ARNt dans lequel le motif La serait le moyen principal de fixation des ARNt et qu'il donnerait ensuite à l'ARN accès au RRM1 permettant ainsi l'activité chaperonne pour des précurseurs déficients. Les auteurs de cette étude ont aussi observé que pour les deux ARNt normaux qu'ils ont étudiés, les mutations dans le feuillet β du RRM1 entraînaient une baisse de l'abondance des précurseurs et une hausse de l'abondance des ARNt matures. Les mutations du feuillet β du RRM1 semblent donc augmenter l'efficacité de maturation des précurseurs normaux. Même s'il faudrait confirmer avec d'autres ARNt, les auteurs proposent une activité de contrôle-qualité impliquant le domaine RRM1, de ce fait il interagirait avec tous les précurseurs d'ARNt, offrant une activité chaperonne aux précurseurs déficients. Dans cette étude, les précurseurs des ARNt normaux seraient plus efficacement maturés parce qu'ils ne seraient pas soumis au contrôle qualité du RRM1 muté au niveau du feuillet β (Huang *et al.* 2006).

III.3.2) La protéine La impliquée dans la biogenèse de transcrits Pol II

L'épissage des ARN pré-messagers est assuré par un ensemble de complexes ribonucléoprotéiques (RNP) appelés collectivement spliceosome. Les particules RNP contenant les petits ARN nucléaires (snRNP pour small nuclear RNP) sont les composants majeurs du spliceosome. À l'exception de la snRNP U6, chaque snRNP contient un petit ARN nucléaire U1, U2, U4 ou U5 (transcrits par la Pol II) associé à sept protéines Sm communes et à un groupe de protéines spécifique à chaque snRNP. Les protéines Sm B ou B', D1, D2, D3, E et F s'associent pour former un hétéro-heptamère en forme d'anneau et lient les petits ARN nucléaires. Elles participent à leur maturation et appartiennent à leur particule RNP. L'étape d'assemblage des protéines Sm avec les petits ARN nucléaires est cytoplasmique chez les vertébrés, les snRNP retournant ensuite dans le noyau pour exercer leur fonction dans l'épissage des pré-ARNm (Yong *et al.* 2004).

Il existe dans la cellule une autre classe très abondante d'ARN non-codants, majoritairement transcrits par la Pol II, principalement impliqués dans la maturation des précurseurs d'ARN ribosomiques : les petits ARN nucléolaires (snoARN). U3 est un snoARN et il est lui aussi transcrit par le Pol II chez *S. cerevisiae*. Pourtant, la protéine Lhp1p semble impliquée dans la maturation des snARN et du snoARN U3. Ainsi, en l'absence de Lhp1p, certains intermédiaires de maturation de ces ARN ne s'accumulent plus. Bien que leurs transcrits primaires ne se terminent pas en UUU-OH, il semble que la protéine Lhp1 soit capable de lier et de stabiliser certains de leurs intermédiaires de maturation (Kufel *et al.* 2000; Xue *et al.* 2000).

Ces petits ARN nucléaires et nucléolaires subiraient un processus de maturation similaire, très complexe faisant intervenir un grand nombre de protéines. Leur transcrit primaire serait tout d'abord coupé par Rnt1p, l'homologue chez *S. cerevisiae* de la RNase III d'*Escherichia coli*. Ce clivage permettrait l'entrée de l'exosome ou d'autres exonucléases qui dégraderaient la partie 3' de l'ARN jusqu'à une séquence poly(U) à laquelle la protéine La pourrait se fixer (Kufel *et al.* 2000). Elle serait ensuite déplacée par des protéines de la RNP, ce qui permettrait aux exonucléases de terminer la maturation 3' de l'ARN.

En contexte sauvage, la protéine La n'est pas indispensable à la biosynthèse des petits ARN nucléaires ou de U3. En revanche lorsque le gène *SMD1*, codant la protéine Sm D1, est muté, Lhp1p devient indispensable à la croissance à 16°C. Ainsi, dans les cellules mutantes pour *SMD1*, Lhp1 apparaît nécessaire pour l'assemblage de l'ARN U4 en complexe RNP, le snRNP U4/U6. Comme *in vitro*, la protéine Smd1p fixe le précurseur de l'ARN U4 mais pas

l'ARN mature, il est proposé que la protéine La stabiliserait la structure de l'ARN pour permettre la fixation des protéines Sm de la snRNP (Xue *et al.* 2000).

Récemment, M. Inada et C. Guthrie ont analysé les ARN associés à la protéine Lhp1p de *S. cerevisiae*. Ils ont tout d'abord isolé les ARN liés par co-immunoprécipitation de la protéine Lhp1p. Ils ont ensuite produit des ADN complémentaires par transcription inverse des ARN et ils les ont hybridés sur une puce portant la quasi-totalité du génome. Dans cette étude, ils ont confirmé la liaison de la protéine La avec les petits ARN nucléaires U1, U2, U4 et U5 ainsi qu'avec le petit ARN nucléolaire U3, mais ils ont aussi constaté que parmi les 66 petits ARN nucléolaires identifiés *S. cerevisiae* plus de la moitié s'associent à la protéine La (Inada and Guthrie 2004).

Le rôle nucléaire de la protéine La n'apparaît donc pas limité à son rôle dans la biogenèse des transcrits Pol III, mais s'étend à un rôle plus général dans la biogenèse des ARN non-codants qui présentent au cours de leur maturation des précurseurs se terminant en 3' UUU-OH. Il existe peu d'éléments permettant de dire si ce rôle plus général de la protéine La est conservé chez les animaux.

III.3.3) Implication de la protéine La dans la biogenèse du petit ARN U6 et inter-connexion de l'action de la protéine La et des protéines Lsm

Comme les ARNt, le petit ARN nucléaire U6 est transcrit par la Pol III, et de ce fait son précurseur est lié par la protéine La, mais cette liaison n'est que transitoire (Pannone *et al.* 1998). Il a en effet été montré chez la levure que Lhp1p était remplacée par l'hétéro-heptamère de protéines Lsm 2-8. Ces protéines restent liées à l'ARN U6 lorsque la particule snRNP U6 s'associe aux particules snRNP U4 et U5. Les protéines Lsm pour Like-Sm ont été nommées ainsi parce qu'elles présentent des similarités de structure avec les protéines Sm, et notamment un domaine particulier des protéines Sm, le motif Sm, qui est impliqué dans la liaison à l'ARN (Beggs 2005). À l'exception des gènes *LSM5*, *LSM6* et *LSM7*, les gènes *LSM* sont des gènes essentiels à la viabilité. Toutefois, selon les souches de laboratoire, la mutation dans le gène *LSM5* a été décrite comme une mutation létale (Mayes *et al.* 1999) ou non (Pannone *et al.* 2001). Les gènes *LSM5*, *LSM6* et *LSM7* ne seraient pas essentiels à la viabilité parce que les protéines qu'ils codent, pourraient être remplacées par d'autres protéines, Lsm ou Sm, dans le complexe Lsm. Une des mutations identifiées lors du crible effectué pour découvrir des conditions et des fonds génétiques (cf. paragraphe III.3.1.c de la synthèse bibliographique) rendant la protéine La nécessaire à la viabilité, touche le motif Sm de la

protéine Lsm8p. En présence de cette protéine hypomorphe, la protéine La devient nécessaire à l'assemblage de la snRNP U6 (Pannone *et al.* 1998). Il a ensuite été montré que la protéine La était nécessaire à la viabilité de souche présentant des mutations dans les gènes non-essentiels *LSM5*, *LSM6* et *LSM7*. Les auteurs proposent donc que dès lors qu'une mutation ralentie ou empêche la liaison du complexe Lsm2-8 normal à l'ARN U6, la protéine La devient indispensable pour stabiliser l'ARN néo-synthétisé et permettre l'assemblage de la particule snRNP (Pannone *et al.* 2001).

Chez *S. cerevisiae*, les protéines Lsm interagissent avec d'autres ARN que l'ARN U6 et notamment des ARN liés par la protéine La, comme le précurseur de l'ARN de la RNase P, les précurseurs des ARNt et les précurseurs des petits ARN nucléolaires (Kufel *et al.* 2002). Plusieurs études suggèrent que pour certains ARN, les interactions des protéines Lsm et de la protéine La sont interdépendantes. Ainsi dans des levures où le gène essentiel *LSM3* est sous le contrôle d'un promoteur conditionnel, 24h après extinction du promoteur, la protéine La ne se lie plus au précurseur de l'ARN de la RNase P, ni au précurseur du gène Scr1 (homologue de l'ARN 7SL de la levure) (Kufel *et al.* 2002), ni aux précurseurs du petit ARN nucléolaire U3 (Kufel *et al.* 2003). Ces résultats suggèrent que la protéine La n'est pas toujours la première protéine à interagir avec les transcrits primaires Pol III. De même en l'absence de la protéine La, la protéine Lsm3 ne s'associe plus aux précurseurs de U3. Le complexe Lsm aurait un rôle de chaperonne en conjonction avec la protéine La pour stabiliser les précurseurs de l'ARN U3 (Kufel *et al.* 2003).

En l'absence des protéines Lsm et de la protéine La, l'abondance de l'ARN U3 mature n'est cependant pas affectée. De même les niveaux des ARNt matures ou des ARN U1, U4 et U5 ne varient pas dans des cellules mutées pour les gènes *LHP1* ou *LSM*. Les voies de maturation des petits ARN doivent donc présenter des redondances qu'il reste à identifier. Cet élément peut aussi être un argument de plus en faveur du modèle qui présente la protéine La comme une chaperonne impliquée dans le contrôle qualité des petits ARN non-codant.

III.3.4) Facteur de transcription de l'ARN polymérase III : un rôle controversé

Un des premiers rôles proposés pour la protéine La humaine était un rôle de facteur de terminaison et de « recyclage » (c'est-à-dire libération de la polymérase pour initier un nouveau cycle) de l'ARN Pol III. Dans les expériences à l'origine de cette hypothèse, la protéine La était retirée d'extraits de cellules HeLa grâce à des anticorps et ces extraits perdaient presque entièrement leur activité de transcription par la Pol III (Gottlieb and Steitz

1989). Les quelques transcrits synthétisés présentant une extrémité 3' raccourcie, un modèle a été proposé dans lequel la protéine La était nécessaire pour une terminaison correcte et le « recyclage » de la Pol III (Gottlieb and Steitz 1989 ; Gottlieb and Steitz 1989).

Au cours des expériences suivantes, l'ajout de protéine La recombinante à des complexes de transcription immobilisés augmentait la libération de la Pol III et était nécessaire à la réinitiation de la transcription. De plus, comme seule la protéine La humaine non-phosphorylée avait cette action stimulatrice de la transcription, il a été proposé que la phosphorylation régulait cette activité (Maraia 1996 ; Goodier and Maraia 1998).

Toutefois, d'autres groupes ont rapporté qu'une Pol III purifiée de vertébré était capable de réaliser la terminaison correcte et la libération des transcrits en l'absence d'autres protéines. Même si les auteurs n'excluent pas la présence de protéine La contaminante dans leurs extraits, leurs expériences ont rendu le rôle de la protéine La dans la transcription controversé (Cozzarelli *et al.* 1983 ; Watson *et al.* 1984 ; Campbell and Setzer 1992). Après ces observations, plusieurs groupes ont testé la transcription Pol III en absence de protéine La chez *S. cerevisiae* (Yoo and Wolin 1997 ; Pannone *et al.* 1998) et le xénope (Lin-Marq and Clarkson 1998). Les extraits de levure dépourvus de Lhp1p sont aussi capables de transcrire les ARNt et l'ARN U6 que les extraits sauvages. De la même façon, les extraits de cellules de xénope dépourvus de protéine La par traitement immunologique se sont avérés aussi capables de transcrire des ARNt que les extraits normaux. Dans ces deux cas, les transcrits obtenus étaient raccourcis en 3', comme les transcrits observés au cours des expériences menées avec la protéine La humaine. Cette caractéristique est attribuée à la fonction de la protéine La dans la protection des transcrits néosynthétisés contre les activités exonucléasiques.

D'autres expériences réalisées avec des extraits de cellules embryonnaires de foie humain ont montré une transcription Pol III complètement active après que la protéine La ait été retirée par traitement immunologique et que l'ajout de protéine La aboutissait à l'augmentation de la quantité de complexes La-ARN, mais ne stimulait pas la transcription (Weser *et al.* 2000). De même la transcription du gène U6 a été obtenue avec des composants partiellement purifiés et la protéine La n'est pas détectable dans le complexe efficace (Chong *et al.* 2001).

Aujourd'hui, le modèle en cours propose que la Pol III utilise un mode de terminaison de la transcription unique par rapport aux autres ARN polymérases dépendantes de l'ADN. Ainsi, elle reconnaît une suite de nucléotides T sur le brin transcrit et cette séquence suffit à déclencher la terminaison de la transcription sans intervention de facteurs extérieurs identifiés. Une séquence de 4 T semble être suffisante pour entraîner la terminaison de la

transcription chez le xénope et l'homme, 4 à 5 T chez *S. pombe* et de 5 à 6 T chez *S. cerevisiae* (Bogenhagen and Brown 1981 ; Huang and Maraia 2001). Mécaniquement, la terminaison de la Pol III implique une longue pause liée à la suite de T. Ainsi même pendant l'élongation, l'ajout de trois résidus U à la suite est particulièrement lent (Matsuzaki *et al.* 1994). La pause de la Pol III au niveau de la suite de T entraîne une hydrolyse de la chaîne ARN par une sous-unité de la Pol III (Shaaban *et al.* 1996).

À travers ces études, la protéine La n'apparaît pas indispensable à la transcription Pol III. De plus, les souches mutantes pour la protéine La étant viables à la fois chez *S. cerevisiae* et *S. pombe*, cette protéine n'est certainement pas un composant essentiel de la machinerie de transcription Pol III chez les levures. En revanche, on ne peut exclure qu'elle interagisse avec cette machinerie pour moduler la transcription *in vivo*.

Récemment des expériences d'immunoprécipitation de la chromatine ont révélé que la protéine La humaine non-phosphorylée était associée avec les gènes transcrits par la Pol III *in vivo*. Cette association est indépendante de la présence de la polymérase et donc des transcrits naissants. Étant donné qu'il existe très peu de séquences ADN conservées entre les différents gènes transcrits par la Pol III, cette association est probablement due à des interactions protéine/protéine, mais le ou les partenaires de la protéine La reste encore à identifier. Cette association de la protéine humaine avec les gènes transcrits par la polymérase III est très intrigante, mais elle reste inexpliquée (Fairley *et al.* 2005).

III.3.5) Rôle dans la traduction des ARNm

De nombreuses études ont impliqué la protéine La dans la traduction d'ARNm spécifiques d'origine virale ou endogènes, notamment des ARNm présentant des séquences 5' particulières. La protéine La serait ainsi impliquée dans des mécanismes particuliers d'initiation de la traduction.

III.3.5)a) La protéine La est impliquée dans la traduction indépendante de la coiffe 5'

Certains ARNm présentent une longue séquence 5' non traduite (5'UTR) capable de diriger l'initiation interne de la traduction au sein de l'ARN indépendamment de sa coiffe 5'. Ces séquences appelées site interne de fixation des ribosomes (IRES pour Internal Ribosome Entry Site) adoptent une structure tertiaire particulière qui est reconnue par les sous-unités ribosomiques elles-mêmes ou par des protéines accessoires qui permettraient le contact avec

ces sous-unités ribosomiques (Vagner *et al.* 2001 ; Spriggs *et al.* 2005). Les protéines cellulaires capables de reconnaître les IRES sont appelées facteurs *trans* d'activation. Il en existerait deux sortes : ceux qui établissent un pont entre l'ARNm et le ribosome et ceux qui agissent comme chaperonne pour modifier la structure de l'ARN. Les protéines ayant des activités de facteur *trans* d'activation ont été identifiées parce qu'elles étaient requises pour l'activité maximale d'une IRES donnée. Cependant, aucune de ces protéines ne semble, jusqu'ici, être un facteur général de la régulation des IRES.

Beaucoup de virus produisent des ARNm qui présentent des IRES. Ils ont ainsi développé des systèmes qui leur permettent de bloquer la traduction normale dépendante de la coiffe de la cellule infectée pour promouvoir la traduction de leurs propres ARNm. De même, les ARNm cellulaires présentant des IRES sont traduits préférentiellement lorsque la traduction dépendante de la coiffe est bloquée. Cet arrêt de la traduction classique se produit lorsque la cellule est soumise à certains stress, comme l'irradiation aux rayons γ, l'hypoxie ou la carence en acides aminés (Vagner *et al.* 2001). La traduction de certains ARNm via leur IRES serait un mécanisme régulé qui permettrait à la cellule de faire face à des stress transitoires, les facteurs *trans* d'activation variant selon les IRES pourraient permettre cette régulation. Des ARNm codant des protéines des voies de l'apoptose sont aussi traduits grâce à des IRES. Ainsi, si le stress ayant entraîné l'arrêt de la traduction classique persiste, ou si les dommages subis ne sont pas réparables, la cellule peut entrer en apoptose (Vagner *et al.* 2001 ; Lewis and Holcik 2005).

La protéine La serait donc un facteur *trans* d'activation. Elle a tout d'abord était identifiée comme la protéine majeure se fixant à la région 5' non-traduite du virus de la poliomyélite lorsque l'ARN de celui-ci est incubé avec un extrait de cellules HeLa et irradié par des rayonnements ultraviolets (Meerovitch *et al.* 1993). Comme les autres picornavirus, le virus de la poliomyélite est traduit par un processus indépendant de la coiffe 5' grâce à une IRES, présente dans la région 5' non-traduite de l'ARNm (Belsham and Sonenberg 2000). L'ARNm du virus de la poliomyélite est peu et mal traduit dans les réticulocytes de lapin, la traduction est en effet initiée de façon aberrante sur l'ARNm. L'ajout d'une forte quantité de protéine La supprime la traduction au niveau de sites aberrants, ce qui suggère qu'elle joue un rôle dans l'initiation correcte de la traduction au niveau de l'IRES (Meerovitch *et al.* 1993). De plus, la protéase 3C du virus de la polio est capable de cliver la protéine La et de libérer une protéine tronquée privée entre autres de son signal de localisation nucléaire, ce qui est corrélé avec une diminution de la quantité de protéine présente dans le noyau et une

augmentation dans le cytoplasme (Shiroki *et al.* 1999). En revanche, comme une grande partie de la protéine La identifiée dans le cytoplasme n'est pas clivée, il semble que la protéine La, normalement nucléaire, est exportée dans le cytoplasme lors de l'infection des cellules par le virus de la poliomyélite (Bachmann *et al.* 1989; Shiroki *et al.* 1999). Le virus induirait la localisation cytoplasmique de la protéine La, où elle pourrait permettre la traduction de l'ARNm viral via son IRES (Maraia and Intine 2001).

Par la suite, différentes équipes ont montré que la protéine La liait et influençait la traduction d'autres ARNm viraux, comme ceux du virus de l'hépatite C, du virus de l'encéphalomyocardite ou le coxsachievirus B3 (Kim and Jang 1999 ; Ali *et al.* 2000 ; Ray and Das 2002).

La protéine La serait également impliquée dans la traduction d'ARNm cellulaires comme ceux codant les protéines Bip et XIAP (Holcik and Korneluk 2000; Kim *et al.* 2001). Bip est la protéine de liaison des chaînes lourdes des immunoglobulines. Elle lie transitoirement une grande variété de protéines de sécrétion et de protéines trans-membranaires et lie, de façon permanente, les protéines mal-repliées qui s'accumulent dans le réticulum endoplasmique. Il est donc proposé qu'elle ait un rôle dans la voie de réponse au stress des protéines mal-repliées (voie UPR pour Unfolded Protein Response). La protéine XIAP (pour X-linked Inhibitor of Apoptosis) est une protéine impliquée dans le processus d'apoptose, elle lie et inhibe certaines caspases, les protéines responsables de la mort cellulaire, jouant ainsi un rôle modulateur dans le déclenchement de l'apoptose. Précocement pendant l'apoptose, la partie carboxy-terminale de la protéine La est clivée. Elle perd alors son signal de localisation nucléaire et, comme lors d'une infection virale, elle passerait dans le cytoplasme, où elle pourrait jouer son rôle dans la traduction de l'ARNm de XIAP (Rutjes *et al.* 1999 ; Ayukawa *et al.* 2000).

Des isoformes alternatives de l'ARNm de la protéine La ont été identifiées dans des lymphocytes (Troster *et al.* 1994 ; Carter and Sarnow 2000). Outre le transcrit majeur qui code la protéine La nucléaire, il existe un transcrit dont la séquence 5' non-codante plus longue contient une IRES, suggérant que la protéine La elle-même continuerait d'être traduite lorsque la traduction dépendante de la coiffe est arrêtée.

Un grand nombre d'éléments se sont donc accumulés en faveur d'un rôle de la protéine La dans l'initiation interne de la traduction d'ARNm endogènes et viraux, mais le mécanisme par lequel elle agit sur la traduction de ces ARNm reste inconnu.

III.3.5)b) Rôle dans la traduction des composants de la machinerie de traduction

Chez les animaux, la protéine La apparaît aussi impliquée dans la traduction des ARNm contenant des séquences oligopyrimidine 5' terminales (5'TOP pour Terminal OligoPyrimidine). Ces ARNm qui codent les protéines ribosomiques et certains autres composants de la machinerie de traduction sont régulés au niveau traductionnel en fonction des conditions de croissance (Meyuhas 2000 ; Hamilton *et al.* 2006). Ainsi en cas de carence en acides aminés ou d'un arrêt du cycle cellulaire, le taux de production des ribosomes est fortement réduit, en partie par l'inhibition spécifique de l'initiation de la traduction des protéines ribosomiques. Une régulation au niveau traductionnel présente l'avantage de pouvoir obtenir un effet très rapide et facilement réversible.

La séquence 5' TOP, correspondant à un résidu C suivi de 4 à 14 résidus pyrimidines consécutifs, confère une expression de l'ARNm dépendante de la croissance. Ce motif 5' TOP serait reconnu par des facteurs *trans* qui réguleraient de cette façon l'initiation de la traduction. En effet, la répression sélective de la traduction d'ARNm 5' TOP peut être levée en ajoutant une grande quantité d'un oligonucléotide d'ARN 5' TOP à l'extrait cellulaire, suggérant qu'il existe un répresseur protéique qu'il est possible de titrer. La protéine La serait donc un de ces facteurs *trans* (Meyuhas 2000).

Ainsi, *in vitro*, la protéine La est capable d'interagir avec des ARNm 5' TOP d'humain et de xénope (Pellizzoni *et al.* 1996). Il a ensuite été montré que la protéine La de xénope sédimentait avec les polysomes d'une façon dépendante de l'ARN et qu'elle liait les ARNm 5' TOP associés aux polysomes (Cardinali *et al.* 2003). Pour connaître l'effet de cette association, des expériences de surexpression de la protéine La ont été menées. Dans des cellules dont la croissance est bloquée, la surexpression de la protéine La entraîne le transfert de l'ARNm 5' TOP RPL4 de la population de polysomes inactifs en traduction vers les polysomes actifs (Crosio *et al.* 2000). Au contraire, l'expression d'une protéine La tronquée de son domaine La dans des cellules en croissance entraîne une plus grande proportion de RPL4 vers la population de polysomes inactifs. Ces observations suggèrent un modèle où la protéine La stimule la traduction des ARNm 5'TOP. Mais d'autres études ont obtenu des résultats en contradiction avec ce modèle. Chez l'homme, la protéine La non-phosphorylée serait préférentiellement associée aux ARNm 5'TOP. La surexpression d'une protéine La mutée sur son site de phosphorylation prédominant en cellule, c'est-à-dire l'augmentation artificielle de la quantité de protéine La non-phosphorylée, entraîne l'augmentation de la quantité d'ARNm 5' TOP RPL37 associée à la protéine et l'abondance de cet ARNm associé

aux polysomes inactifs (Schwartz *et al.* 2004). En accord avec ces derniers résultats, il a été montré que la protéine La humaine réprimerait la traduction d'ARNm 5' TOP *in vitro* (Zhu *et al.* 2001).

La protéine La semble donc bien interagir avec les ARNm 5'TOP et être impliquée dans la régulation de leur traduction, mais son rôle précis reste à déterminer.

III.3.5)c) Interaction de la protéine La avec des ARNm chez S. cerevisiae

L'analyse menée par Inada et Guthrie (2004) des ARN lié par la protéine La de *S. cerevisiae* (cf. paragraphe III.3.2 de la synthèse bibliographique) a permis de montrer qu'elle fixait aussi des ARNm. La protéine Lhp1p fixe entre autres un grand nombre d'ARNm codant des protéines ribosomiques. Au contraire de ceux des animaux, les ARNm codant les protéines ribosomiques de *S. cerevisiae* ne présentent pas de séquences 5' TOP et leur expression n'est pas régulé au niveau traductionnel mais au niveau transcriptionnel, en fonction des conditions de croissance (Powers and Walter 1999). Il n'existe pas d'informations sur l'impact de la liaison Lhp1p à ces ARNm. Mais comme Lhp1p interagit avec un grand nombre d'acteurs de la machinerie de traduction (ARNt, ARN 5S, ARNm codant les protéines ribosomiques) et de la machinerie d'épissage (petits ARN nucléolaires), il est proposé qu'elle soit impliquée dans la coordination de ces événements (Inada and Guthrie 2004).

Un des ARNm liés par Lhp1p identifié au cours de cette étude, est l'ARNm de la protéine Hac1p, facteur de transcription impliqué dans la voie de réponse au stress des protéines mal-repliées. L'implication de la protéine Lhp1p dans la voie UPR a donc été testée. À 30°C (la température optimale de croissance de *S. cerevisiae*), aucune différence n'a été observée entre une souche sauvage et une souche mutante pour le gène *LHP1* lorsqu'elles sont cultivées sur un milieu induisant chimiquement la voie UPR. En revanche à 37°C, la souche mutante pour *LHP1* ne pousse pas sur le milieu induisant la voie UPR. L'observation par northern blot de l'ARNm de Hac1 a montré qu'il n'existe pas de différence de maturation de cet ARN entre les souches sauvage et mutantes pour *LHP1*. Par contre, la quantité de protéine Hac1p est réduite dans la souche mutante pour LHP1. La protéine Lhp1p semble donc impliquée dans la traduction de l'ARNm Hac1, et elle devient nécessaire à la viabilité dans des conditions de croissances difficiles (Inada and Guthrie 2004).

III.3.5)d) Un nouveau rôle de la protéine La dans la biologie d'ARNm : cas de l'ARNm de la Peptidylglycine α-Amidating Monooygenase (PAM) des mammifères.

L'enzyme PAM joue un rôle clef dans les neurones et les cellules endocrines : elle confère, en catalysant leur α-amidation, l'activité biologique aux neuropeptides inactifs destinés à être sécrétés. La protéine La a été identifiée comme la protéine régulatrice fixant la séquence 3' non-traduite de l'ARNm de cette enzyme (Brenet *et al.* 2005). Dans ce cas, la protéine La ne semble pas intervenir directement dans la traduction, mais semble responsable de la rétention nucléaire de l'ARNm. La surexpression de la protéine La entraîne l'augmentation de la quantité d'ARNm de la PAM dans le noyau et diminue en conséquence l'activité de l'enzyme. Il semble donc que la protéine La puisse moduler l'expression de l'enzyme en retenant une partie de ses ARNm dans le noyau.

III.4. La protéine La humaine, des rôles différents selon l'état de phosphorylation

Quatre sites de phosphorylation ont été identifiés sur la protéine La humaine extraite de cellules HeLa, ils correspondent aux acides aminés Thréonine (T) et Sérine (S) : T302, S325, T362 et S366 (Broekhuis *et al.* 2000). La sérine en position 366 (S366) est le site de phosphorylation prédominant de la protéine. C'est aussi le seul site phosphorylé par la caséine kinase 2 (CK2) *in vitro* et l'inhibition ou la diminution de la quantité de cette kinase par ARNi augmente la quantité de protéine La non-phosphorylée, suggérant que la kinase CK2 est bien responsable d'une partie de la phosphorylation de la protéine La *in vivo* (Schwartz *et al.* 2004).

Les sites de phosphorylation de la protéine Lhp1p de levure ont aussi été identifiés. La mutation des sites de phosphorylation n'a pas d'impact ou un impact minime sur les fonctions de la protéine La dans les fonctions les mieux connues de la protéine, comme la maturation des ARNt ou la biogenèse de la snRNP U6. L'état de phosphorylation n'a pas non plus d'impact sur la localisation de la protéine (Long *et al.* 2001).

Au contraire de la protéine de levure, l'état de phosphorylation de la protéine La humaine, et notamment celui de la sérine 366, semble avoir un impact sur les ARN qu'elle lie et par conséquent sur ses fonctions.

Grâce à des anticorps dirigés contre les deux formes de la protéine, phosphorylée ou non au niveau de la sérine 366, Intine et al. (2003) ont observé des localisations et des affinités différentes pour les deux formes de la protéine. La forme phosphorylée de la protéine qui est la plus abondante s'accumule dans le nucléoplasme et le nucléole. En revanche la protéine non-phosphorylée apparaît dans le nucléole et dans le cytoplasme. Aucune différence de localisation n'a été observée entre la protéine La normale et une protéine dont les sites de phosphorylation ont été mutés (Broekhuis *et al.* 2000 ; Intine *et al.* 2003 ; Intine *et al.* 2004), ce qui suggère que l'état de phosphorylation est une conséquence de la localisation et pas le contraire (Bayfield *et al.* 2007).

L'analyse des ARN co-immunoprécipités, à partir d'extraits de cellules HeLa avec les anticorps contre la protéine phosphorylée ou non, a montré que, selon son état de phosphorylation, la protéine La n'était pas associée aux mêmes ARN dans la cellule (Intine *et al.* 2003). Ainsi, les précurseurs des ARNt et de l'ARN 5S sont associés à la protéine La phosphorylée et pas à la protéine non-phosphorylée. La protéine La phosphorylée serait donc responsable du rôle nucléaire de la protéine La dans la biosynthèse des transcrits Pol III. En revanche, la protéine La non-phosphorylée lie les ARNm 5' TOP, ce qui n'est pas le cas de la protéine phosphorylée (Intine *et al.* 2003 ; Schwartz *et al.* 2004).

Il a été préalablement montré que la protéine La humaine non-phosphorylée exprimée chez la levure *S. pombe*, était capable de fixer la séquence UUU-OH des pré-ARNt. Mais il a aussi était montré qu'elle pouvait contacter l'extrémité 5' de ces ARN grâce au motif basique court et que, par cette fixation, elle inhibait la coupure de l'extrémité 5' par la RNase P. La phosphorylation de la sérine 366 située peu après le motif basique empêcherait la fixation de l'extrémité 5' (Intine *et al.* 2000). La protéine La non-phosphorylée serait beaucoup moins concentrée que la protéine phosphorylée dans le nucléoplasme pour ne pas perturber la maturation des ARNt commençant dans ce compartiment (Schwartz *et al.* 2004).

La fonction de la protéine La non-phosphorylée dans le nucléole n'a pas été élucidée. En revanche, la localisation de la protéine dans le nucléole est dépendante de la présence du motif basique court (Horke *et al.* 2004 ; Intine *et al.* 2004). La protéine La interagit via ce motif avec la nucléoline, un facteur impliqué dans la maturation de l'ARNr. Cette interaction serait dépendante de la phosphorylation, suggérant que la protéine La non-phosphorylée s'accumule dans le nucléole parce qu'elle interagit avec la nucléoline (Intine *et al.* 2004). La

protéine La non-phosphorylée se localise plus précisément au niveau du composant fibrillaire dense du nucléole. Cette zone étant le lieu de la maturation des ARNr, Intine et al. proposent que la protéine La non-phosphorylée pourrait être impliquée dans la maturation des transcrits Pol I.

Objectifs des travaux de thèse

L'équipe du professeur Jean-Marc Deragon s'intéresse aux différents aspects de la biologie des éléments SINE. Ainsi, les travaux réalisés au sein de l'équipe s'étendent de l'étude de l'histoire évolutive des SINE, notamment ceux de crucifères, à l'étude fonctionnelle des éléments SINE. Cette étude fonctionnelle des éléments SINE est principalement répartie sur deux axes : d'une part l'étude des mécanismes de rétroposition des SINE (incluant la régulation transcriptionnelle et post-transcriptionnelle des éléments SINE, ainsi que la mobilisation et le fonctionnement de la machinerie de rétroposition) et d'autre part l'étude des fonctions cellulaires des transcrits SINE en tant que riborégulateur.

La protéine La est une protéine multifonctionnelle, dont le rôle le mieux connu est celui qu'elle remplit dans la maturation des transcrits de l'ARN polymerase III. Chez les mammifères, il a été montré que la protéine La s'associait à certains transcrits SINE (Goodier and Maraia 1998 ; Kremerskothen *et al.* 1998). Cette protéine étant conservée chez tous les eucaryotes, il a été proposé qu'elle lierait, au moins transitoirement, tous les transcrits SINE et qu'elle serait impliquée dans la maturation post-transcriptionnelle de ces ARN (Maraia *et al.* 1992 ; Goodier and Maraia 1998 ; Schmid 1998). Afin de tester cette hypothèse, j'ai donc entrepris l'étude de la protéine La dans chez la plante modèle *Arabidopsis thaliana*. La littérature ne faisant état d'aucune étude préalable, mes travaux ont débuté par l'identification de l'homologue de la protéine La d'*Arabidopsis* par une approche bioinformatique. Cette partie de mon travail a révélé une situation originale, *Arabidopsis* ne possédant pas une, mais deux protéines présentant les caractéristiques d'une protéine La. J'ai donc procédé à l'étude de ces deux protéines pour déterminer leur rôle dans la maturation des transcrits pol III. J'ai également débuté les travaux visant à étudier leur impact sur le métabolisme des transcrits SINE.

La plupart des éléments SINE possèdent une partie 5' dérivée d'un ARN de transfert, une famille d'ARN dont la structure secondaire en trèfle est parfaitement connue. Paradoxalement, la structure secondaire des ARN SINE n'a été que très peu étudiée. Parallèlement à mes travaux sur la fonction La chez *Arabidopsis*, j'ai donc participée à une étude visant à étudier la structure secondaire de différentes familles de SINE et à replacer les données générées dans le contexte de l'histoire évolutive des SINE. Ma contribution à cette étude consistait à déterminer la structure en solution des ARN de deux SINE de crucifères : le SINE SB1 de *Brassica napus* et le SINE SB2 d'*Arabidopsis thaliana*.

Résultats

I. Caractérisation de la fonction La chez *Arabidopsis thaliana*

I.1. Article 1 : A bona fide La protein is required for embryogenesis in *Arabidopsis thaliana*

Dans un premier temps, nous avons cherché à identifier l'homologue de la protéine La d'*Arabidopsis thaliana*. À l'aide du programme BLAST, nous avons recherché dans le génome d'*Arabidopsis* des gènes codant des protéines présentant des similarités de séquence avec le consensus du domaine La. Cette étude bioinformatique nous a permis d'identifier 8 protéines d'*Arabidopsis* contenant un domaine La. En plus de ce domaine, la partie amino-terminale des protéines La caractérisées contient un domaine RRM canonique. Parmi les 8 protéines d'*Arabidopsis* identifiées, 5 contiennent un domaine RRM associé en tandem avec le domaine La. Nous avons donc procédé à des analyses phylogénétiques des domaines La et RRM, pour déterminer les protéines d'*Arabidopsis* les plus proches des vrais protéines La. À l'issue de ces analyses, nous avons retenu deux protéines codées par les loci At4g32720 et At1g79880, que nous avons respectivement nommées At32 et At79.

En plus du domaine RRM présent dans la partie amino-terminale, la protéine La humaine contient dans sa partie carboxy-terminale, un second RRM atypique reconnaissable uniquement sur la base de sa structure secondaire (Jacks *et al.* 2003). Nous avons analysé la partie carboxy-terminale de tous les homologues La identifiés à ce jour et à l'exception des protéines des levures *S. cerevisiae* et *S. pombe*, nous avons pu mettre en évidence la présence de ce RRM atypique, nommé RRM2. Des cinq protéines d'*Arabidopsis* présentant un tandem La-RRM dans leur partie amino-terminale, seules les protéines At32 et At79 contiennent ce RRM2, et donc possèdent les caractéristiques structurales d'une protéine La.

Nous avons ensuite cherché à déterminer si ces deux protéines étaient capables de remplir certaines des fonctions connues de la protéine La. Les levures *S. cerevisiae* et *S. pombe* mutantes pour le gène codant la protéine La sont viables et un grand nombre de phénotypes moléculaires liés à l'absence de la protéine ont été mis en évidence, notamment dans les processus de maturation des précurseurs des ARNt. En comparaison, on dispose de relativement peu d'informations sur la maturation des précurseurs des ARNt chez *Arabidopsis*, ce qui rend délicate l'étude directe du rôle des protéines At32 et At79 dans ces processus. La protéine La semble cependant suffisamment conservée pour que la protéine La

humaine complémente les phénotypes observés chez des souches mutantes de *S. pombe*. Nous avons donc testé la capacité des protéines At32 et At79 à complémenter différents phénotypes de souches de *S. cerevisiae* mutantes pour la protéine La (Lhp1). Nous avons complété l'étude fonctionnelle des protéines At32 et At79 en testant leur capacité à lier des transcrits Pol III chez la plante.

La protéine At32 complémente en levure les phénotypes moléculaires choisis et elle fixe en cellules d'*Arabidopsis* les transcrits Pol III que nous avons testés. Ces éléments indiquent qu'elle est au moins responsable des fonctions nucléaires de la protéine La chez *Arabidopsis*, nous l'avons donc renommé AtLa1 pour *Arabidopsis thaliana* La protein 1.

Nous avons finalement complété notre étude avec l'observation de la localisation de la protéine AtLa1 et l'étude de lignées mutantes pour cette protéine, démontrant qu'elle était essentielle à la viabilité de la plante.

Ce travail a donné lieu à la publication d'un article inséré ci-après.

3306–3321 *Nucleic Acids Research, 2007, Vol. 35, No. 10* *Published online 25 April 2007*
doi:10.1093/nar/gkm200

A *bona fide* La protein is required for embryogenesis in *Arabidopsis thaliana*

Sophie Fleurdépine[1], Jean-Marc Deragon[1,2], Martine Devic[2], Jocelyne Guilleminot[2] and Cécile Bousquet-Antonelli[1,*]

[1]CNRS UMR6547 GEEM, Université Blaise Pascal, 63177 Aubière, France and [2]CNRS UMR5096 LGDP, Université de Perpignan Via Domitia, 66860 Perpignan, France

Received March 7, 2007; Revised March 21, 2007; Accepted March 21, 2007

ABSTRACT

Searches in the *Arabidopsis thaliana* genome using the La motif as query revealed the presence of eight La or La-like proteins. Using structural and phylogenetic criteria, we identified two putative genuine La proteins (At32 and At79) and showed that both are expressed throughout plant development but at different levels and under different regulatory conditions. At32, but not At79, restores *Saccharomyces cerevisiae* La nuclear functions in non-coding RNAs biogenesis and is able to bind to plant 3′-UUU-OH RNAs. We conclude that these La nuclear functions are conserved in *Arabidopsis* and supported by At32, which we renamed as AtLa1. Consistently, AtLa1 is predominantly localized to the plant nucleoplasm and was also detected in the nucleolar cavity. The inactivation of AtLa1 in *Arabidopsis* leads to an embryonic-lethal phenotype with deficient embryos arrested at early globular stage of development. In addition, mutant embryonic cells display a nucleolar hypertrophy suggesting that AtLa1 is required for normal ribosome biogenesis. The identification of two distantly related proteins with all structural characteristics of genuine La proteins suggests that these factors evolved to a certain level of specialization in plants. This unprecedented situation provides a unique opportunity to dissect the very different aspects of this crucial cellular activity.

INTRODUCTION

The La protein is a highly abundant phosphoprotein first described in human as an autoantigen in patients suffering from the rheumatic diseases, systemic lupus erythematosus and Sjögren's syndrome (1,2). It is an RNA-binding protein involved in many aspects of RNA metabolism (3–5) and is present in a wide range of eukaryotes including budding and fission yeasts, vertebrates, insects, worm (5) and trypanosome (6). The La protein is one of the first proteins to bind to primary polymerase III (pol III) transcripts due to the specific recognition of the 3′-UUU-OH motif present in these precursors (7). The *Saccharomyces cerevisiae* La protein (named Lhp1p) also binds polymerase II (pol II) transcribed small RNAs that terminate in 3′-UUU-OH such as precursors to the U3 snoRNA (small nucleolar RNA) or U snRNAs (small nuclear RNA) (8–10). From yeast to human, genetic and biochemical studies have shown that La protects these small RNAs from 3′–5′ exonucleases (5,11–13). For example, the binding of La to pre-tRNA precursors prevents exonucleolytic nibbling of their 3′-trailer and promotes its endonucleolytic removal (14–17) and Lhp1p stabilizes U snRNAs and U3 precursors from exonucleolytic degradation (9,10,18). In addition, the La protein fulfils an RNA chaperone activity (19) involved in the assembly of several RNPs (10,18) and in the structural stabilization of pre-tRNAs (20,21). Finally, the La protein most probably takes part in the quality-control mechanism of newly synthesized non-coding RNAs such as pre-tRNAs (17,20,22).

To accomplish its various functions in the biosynthesis of small stable RNAs, most of the La protein logically accumulates in the nucleoplasm as assessed by steady-state subcellular localization experiments (23–25). Nevertheless, the La protein subcellular distribution is highly dynamic as this protein was shown to shuttle through the nucleolus in association with several precursor RNAs (3,24–26) and strongly accumulates in the nucleolus during late G1 and early S phases for yet unknown reasons (24). But not all La protein is present in the nuclear compartment. It has been demonstrated that 2–4% of the *Xenopus* La homologue accumulates in the cytoplasm (27) and that the human La (hLa) protein shuttles between nucleus and cytoplasm (28). Moreover, a major pool of La protein is redistributed to the cytoplasm under various stress conditions such as apoptosis (29) or

*To whom correspondence should be addressed. Tel: +33 (0)4 73407408; Fax: +33 (0)4 73407777; Email: cecile.antonelli@univ-bpclermont.fr

© 2007 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

viral infections (30,31). These observations are in agreement with several reports suggesting that, beside their primary role in pol III and non-coding pol II stable RNAs biosynthesis, La homologues could be implicated in mRNA translation enhancement [for review see (5)]. For example, by binding to their 5′-UTR, the cytoplasmic La protein stimulates the internal ribosome entry site-mediated translation of viral mRNAs (32,33) as well as certain cellular mRNAs (34,35). Also, La is involved in the cap-dependent translation of 5′-terminal oligopyrimidine stretch (TOP) containing mRNAs (27).

La proteins are modular polypeptides whose molecular weight ranges from ~50 kDa in vertebrates to 32 kDa in yeasts. Their NH2-terminal domain (NTD) is extremely well conserved and always contains a 60–80-amino-acid-domain called the La-motif, also found in a number of otherwise unrelated (La-like) proteins (5), closely followed by a canonical RNA recognition motif (RRM1) (36). Diverse structural (37–40) and functional studies (4,5) emphasized the importance for a *bona fide* La protein to display this particular NTD organization at least for efficient and specific binding of 3′-UUU-OH-containing substrates. The COOH-terminal domain (CTD) of La proteins is more variable. The hLa CTD contains an atypical RRM (RRM2) ending with a long helix comprising a nuclear retention element (41,42), followed by a short basic motif (SBM), several phosphorylation sites (23), a nucleolar localization signal (NoLS) (25) and a nuclear localization signal (NLS). The RRM2 motif is found in La proteins from all vertebrates, but is absent from the very short CTD of the yeast proteins and was not detected for La homologues from some metazoans such as fly or worm (5).

In higher plants, a functional homologue of the La protein is yet to be identified. We report here that higher plants are exceptional compared to other eukaryotes by having two distantly related proteins that display every structural feature of genuine La proteins. We show that at least one of these proteins we named AtLa1 (for *Arabidopsis thaliana* La protein 1) is a true functional homologue as it is able to fulfil the nuclear La functions related to RNA pol III transcripts maturation and stability. We also demonstrate that T-DNA insertion in the *AtLa1* gene leads to embryonic lethality showing that AtLa1 function is required for plant viability.

MATERIALS AND METHODS

Molecular phylogenetic analyses

The amino acid sequences of the different La motifs were aligned using ClustalW multiple-alignment program (43). Evolutionary distances were calculated using the probability matrix from blocks (PMB) model (44) of the Protdist program (PHYLIP package version 3.6, available at the following web site: http://evolution.genetics.washington.edu/phylip.html). The coefficient of variation of the gamma distribution (to incorporate rate heterogeneity) was obtained by pre-analysing the data with the Tree-Puzzle program (45), and the significance of the various phylogenetic lineages was assessed by bootstrap analyses (1000 trials) (46). The phylogenetic tree and the consensus tree were inferred using the Neighbour-Joining (NJ) and Consense programs (both from the PHYLIP package), respectively.

Cloning

cDNAs corresponding to At32 and At79 isoforms were PCR amplified from a cDNA library derived from Col0 suspension cell line (47). PCR products were inserted in pBluescript vector and two clones for each cDNA were fully sequenced. Plasmids were named p116 (At32 in pBSK) and p114 (At79 in pBSK), respectively. The centromeric yeast shuttle vector pFL38 (*ARSCEN-URA3*) (48) was used as a platform to assemble the different genes under the control of the *GAL1-10* promoter region (49) with or without the Protein A tag (50). First, a 1169-bp SacI-BamHI and a 264-bp BamHI-HindIII restriction fragments, both from the pGALPATG1L vector (obtained from K. Hellmut and E. Hurt, University of Heidelberg), containing the GAL1-10 promoter region fused with the Protein A tag, and the GAL4 transcription terminator region, respectively, were cloned into pFL38 to produce the pFL38GALProtA vector (p131). The *Arabidopsis* open reading frames (ORFs) were PCR amplified from plasmids p116 and p114, respectively, while the *LHP1* gene was amplified from the pATL vector [generous gift from Sandra Wolin (HHMI, New Haven)]. All PCR amplifications were performed using a 5′-primer with a terminal BamHI site (to produce the tagged version) or with a terminal SphI site (to produce the untagged version) in combination with a 3′-primer ending with an XhoI site. The tagged and untagged final constructs were produced by cloning the different PCR products in pFL38GALProtA after digestions with BamHI-XhoI (tagged version) or with an SphI-XhoI (untagged version). By this way, we obtained the pGAL [numbers p140 (*LHP1*), p139 (At32) and p138 (At79)] and pGALPa [numbers p133 (*LHP1*), p132 (At32) and p135 (At79)] collections. To construct the high-copy vectors, the *ARSCEN* regions of plasmids p131, p133 and p132 were excised by ClaI digestion and replaced with the 2 μL region of pFL44L (48) to produce the p2μGALPa plasmids [p144 (Protein A alone), p150 (*LHP1*), p145 (At32) and p146 (At79)].

Restriction sites (NheI-SmaI at the 5′-end and BamHI-XbaI at the 3′-end) were added by PCR amplification to the eGFP ORF. The PCR amplification product was inserted into pBluescript vector at sites HincII and XbaI to obtain pBSK-eGFP (p315) and verified by sequencing. The At32-coding region was subsequently inserted into the BamHI and XbaI sites of the pBSK-eGFP resulting in NH2 GFP fusion (vector p316), and into the XhoI and NheI sites to give COOH fusion (vector p317). After sequencing, each fusion as well as the eGFP ORF were transferred at the KpnI and XbaI restriction sites of the plant transformation vector pBIN-HYG-TX (51) under the control of the cauliflower mosaic virus 35S promoter (CaMV35S) giving plasmids p320 (GFP), p321 and p322 (GFP-At32 and At32-GFP).

Yeast strains and media

Growth and handling of *S. cerevisiae* were by standard techniques (52). The red/white sectoring medium contained 1% bactopeptone, 0.5% yeast extract and 4% galactose. Fluoroorotate tests were performed on Yeast Nitrogen Base (YNB) medium containing 2% galactose, supplemented as required in amino acids and containing 1 mg/ml of proline and 1 mg/ml of fluoroorotic acid (Melford Lab. Ltd, ref F5001). The BP1 (*Mat*a *lsm8-1 ura3 lys2 ade2 trp1 his3 leu2 lhp1::LEU2*/pATL), CY2 (*Mat*α *ura3 lys2 ade2 trp1 his3 leu2 lhp1::LEU2*) and CY3 (*Mat*α *ura3 lys2 ade2 trp1 his3 leu2*) strains (16,53) used in the present work were kindly provided by Sandra Wolin (HHMI, New Haven). Yeast transformations were performed as described (54) except that 6% DMSO was added prior to heat shock and the final pellet was resuspended in 0.15 M NaCl.

Yeast red/white sectoring and fluoroorotic acid assays

The BP1 strain was transformed with the different sets of yeast plasmids and plated on minimal medium lacking uracil and containing 2% glucose. The sectoring assays were always performed as follow, with two transformants for each plasmid previously selected at least twice on YNB plates lacking uracil and containing 2% galactose. The transformants to be tested were grown overnight at 30°C in liquid minimal medium lacking uracil and containing 2% galactose until cultures reached 1 $OD^{600\ nm}$. They were then diluted in the same medium and monitored for growth until they reached 0.5 $OD^{600\ nm}$. Cells were then plated on sectoring medium at 200–500 cells per plate. After 4 days growth at 30°C, the percentage of red/white sectoring was determined. For further tests, red sectors were streaked on rich galactose medium as many times as necessary to give only solid red colonies. An independent red clone was then tested on appropriate plates for auxotrophy to tryptophan and uracil and for its ability to grow on fluoroorotic-acid-containing plates.

Yeast RNA analysis

The CY2 strain was transformed with the p2μGALPa plasmids and plated on minimal medium lacking uracil and containing 2% glucose. Two transformants for each plasmid were streaked at least twice on YNB plates lacking uracil and containing 2% galactose. Cultures were then conducted as for the sectoring assays and total RNAs were extracted as described (55). To analyse tRNAs and U3 snoRNA, 5 μg of total RNAs were fractionated on polyacrylamide 8.3 M urea gels. RNAs were then blotted on nitrocellulose Hybond N+ membranes (Amersham Biosciences). Hybridizations were performed as described using previously published oligonucleotide probes (9,56).

Yeast western blot analysis

Total protein extracts were prepared as described (57) from 5 $OD^{600\ nm}$ cell cultures prepared for sectoring assays or northern blot analysis. Proteins were fractionated on 10% SDS-PAGE gels and blotted on nitrocellulose membrane with Trans-Blot semi-dry system (BioRad). The blots were reacted with rabbit anti-Nhp2p [generous gift from M. Caizergues-Ferrer (LBME, Toulouse)] (58) at 1/5000 dilution as primary antibody and donkey anti-rabbit IgGs horse radish peroxydase linked (Amersham Biosciences) at 1/10 000 dilution as secondary antibody.

Antisera production and plant western blot analysis

Antibodies were produced using the Eurogentec double X immunization programme followed by affinity-column purifications. Briefly for each protein, rabbits were inoculated with a mix of two synthetic peptides (pep1 and pep2) corresponding to specific regions of the protein. For both proteins, pep1 is located in the central region between RRM1 and RRM2 (At32pep1: H_2N-CQPQKGSANQKNGSDH-$CONH_2$, At79pep1: H_2N-CLGKSESHNEFRRGQI-$CONH_2$,) and pep2 corresponds to the very last 16 or 15 amino acids of the proteins (At32pep2: H_2N-CDSPGGRWNKSQK VEA-COOH, At79pep2: H_2N-CFENVQPTKKARKEP-COOH). Sera from rabbit's final bleeding were divided into two and each sample was affinity purified against pep1 or pep2, respectively. To assess the immunogenicity of each purified serum, we performed western blot analysis on total extracts prepared from yeasts expressing At32 or At79, wild-type *Arabidopsis* cell suspension or 2-weeks-old seedlings. For both At32 and At79, only the sera fraction purified against pep2 gave a satisfying immunogenic response and was used in subsequent western blot analyses.

To prepare total protein extracts, plant materials (plant tissue or cell suspension) were suspended in Laemmli sample buffer supplemented with 0.2 M DTT and 10% Protease Inhibitor Cocktail (P9599 Sigma) in a 1.5-ml microtube and crushed with a small glass pestle and sea sand. Extracts were incubated for 5 min at 95°C and cleared by centrifugation. Gel separation and blotting were performed as described above in the yeast western blot section.

Arabidopsis plant and cell suspension material

The T-DNA insertion lines SAIL 548H11 (59) and GABI 870F12 (60) were obtained from the ABRC Stock Centre and the Max Planck Institute for Plant Breeding Research, respectively.

To prepare transgenic cell suspension lines expressing the different GFP constructs, *Arabidopsis* T87 cells were transformed with the different plant transformation plasmids using *Agrobacterium tumefasciens*, as previously described (61). For each construct, cell lines were screened by western blot analysis with anti-GFP monoclonal antibodies (Clontech) for expression of the transgenic protein. Cell lines were maintained under constant hygromycine selection (25 μg/ml) as liquid suspensions and as calluses growing on solid media, as previously described (61).

Preparation of crude cell extracts and immunoprecipitation experiments

Here, 2 g of *Arabidopsis* cell suspension expressing GFP fusions, obtained by filtration of 3–4-day-old liquid culture is resuspended in 6 ml of ice-cold lysis buffer (150 mM NaCl, 50 mM Tris-HCl pH 8, 2.5 mM $MgCl_2$, 0.1% Triton 100×) with 1 mM DTT, 10 mM vanadyl ribonucleoside complex (Biolabs), 1% protease inhibitor cocktail (P9599 Sigma), 2 mM benzamidine, 1 mM phenyl methyl sulphonyl fluoride and 10 μM decarboxylase inhibitor. The suspension is loaded in the pre-cooled chamber of a 'one-shot' cell disrupter system (Constant Systems Ltd), and cells are lysed under a pressure of 552 bar. The extract is subsequently cleared by centrifugation (30 min, 4200 *g*, 4°C).

For immunoprecipitation, crude extracts were incubated with a 1/300 dilution of anti-GFP full-length polyclonal antibody (Clontech) on a rotary shaker for 3 h at 4°C and then mixed with magnetic nanoparticles conjugated with Protein A (Bio-Adembeads Protein A, Ademtech S.A) and incubated for an additional hour. After three washes with lysis buffer, beads were divided as follows: one-ninth of the beads was resuspended in Laemmli buffer and heated 5 min at 95°C for protein analysis and the remaining beads were eluted with urea 8 M 5 mM EDTA buffer and extracted once with phenol:chloroform:isoamyl alcohol. RNAs were then precipitated by adding 40 μg of glycogen and 2.5 volumes of absolute ethanol and resuspended in RNase-free water. To compare the different transgenic lines, equivalent fractions of proteins or RNAs were used for western blot or RT-PCR analysis. Proteins were analysed by western blot with an anti-GFP monoclonal antibody (Clontech). Chemiluminescent signal was quantified using the VersadocImaging system (BioRad). RNAs were treated with the DNA-free kit from Ambion following the manufacturer's 'Rigorous DNase treatment' protocol. A fraction of the RNA (usually one-fifth) was reverse transcribed with expand reverse transcriptase and hexanucleotides as primer (Roche). cDNAs were then PCR amplified with primer pairs specific to pre-$tRNA^{Met}$, pre-tsnoR43.1 or 5.8S RNA.

Confocal microscopy

Protoplasts were prepared from cell suspensions expressing the GFP fusions as described by Sheen,J. (2002, A transient expression assay using *Arabidopsis* mesophyll protoplasts. http://genetics.mgh.harvard.edu/sheenweb) using an overnight digestion in the dark. Images were obtained with a Zeiss Laser Scanning Microscope LSM 510.

Study of embryonic phenotypes

Wild-type and mutant seeds were collected from hemizygous siliques at different stages of maturity. The seeds were fixed in ethanol:acetic acid (3:1) for 20 min, followed by a slow rehydration in a series of ethanol–water solutions. After rehydration, the seeds were cleared in Hoyer's solution (2.5 g gum arabic, 100 g chloralhydrate, 5 ml glycerol in 30 ml of H_2O) and observed using Nomarski optics with a Zeiss Axioskop2 microscope (Carl Zeiss, Germany).

RESULTS

Identification of La and La-like proteins in *Arabidopsis thaliana* and *Oryza sativa*

We searched (by a protein BLAST at http://www.ncbi.nlm.nih.gov) the *A. thaliana*- and rice (*Oryza sativa*)-expressed genomes for proteins presenting similarities with the eukaryote consensus La-motif (SMART accession number SM00715, http://smart.embl-heidelberg.de). Our search revealed the presence of eight *Arabidopsis* and nine rice putative proteins containing a La motif (Figure 1). We used a phylogenetic approach to test the relationship of *Arabidopsis* and rice La motifs with those present in several eukaryote La and La-like proteins (Figure 1A). Our results show that two *Arabidopsis* (At4g32720 and At1g79880) and two rice (bad19607 and cae03115) proteins form a well-supported cluster (bootstrap of 1000) and that the association of this cluster with the *Schizosaccharomyces pombe* and *S. cerevisiae* La proteins is fairly well supported (bootstrap of 728, Figure 1A). The NTD of La proteins contains a typical RRM (SMART accession number SM00715) (referred to as RRM1) that closely follows the La motif. In addition to At4g32720, At1g79880, bad19607 and cae03115, three *Arabidopsis* and four rice proteins also present a RRM closely following the La motif (Figure 1B), but these proteins group in a cluster intermediate between genuine La and La-like proteins (Figure 1A). The use of the first RRM, instead of the La motif in the phylogenetic analysis also clearly distinguishes these proteins from *bona fide* La homologues (not shown).

The CTD of the hLa protein possesses a novel type of RRM (called RRM2), the structure of which is atypical compared to canonical RRMs (42). This motif is very difficult to detect based on primary sequence analysis alone. We hence searched for the presence of such atypical RRMs in the plant putative La homologues using the hidden-Markov-model (HMM)-based protein structure prediction program SAM-T02 (http://www.cse.ucsc.edu/research/compbio/HMM-apps/T02-query.html). This program was successful in predicting the correct topology for the human atypical RRM2 (see Supplementary Figure 1 online). We found that amongst the plant proteins displaying both the La motif and the canonical RRM, At4g32720 and At1g79880 from *Arabidopsis* and bad19607 and cae03115 from rice are the only ones likely to possess an atypical RRM2 domain in their CTDs (Figure 1B and Supplementary Figure 1 online). We also found that with the exception of the *S. pombe* and *S. cerevisiae* homologues, each of the genuine La proteins we tested is likely to display such motif (Figure 1B, Supplementary Data Figure 1 online). Therefore, the presence of an atypical RRM in the La protein CTDs is apparently not a characteristic restricted to vertebrate proteins, but is likely to be a conserved feature of genuine La proteins with the exception of the yeast homologues,

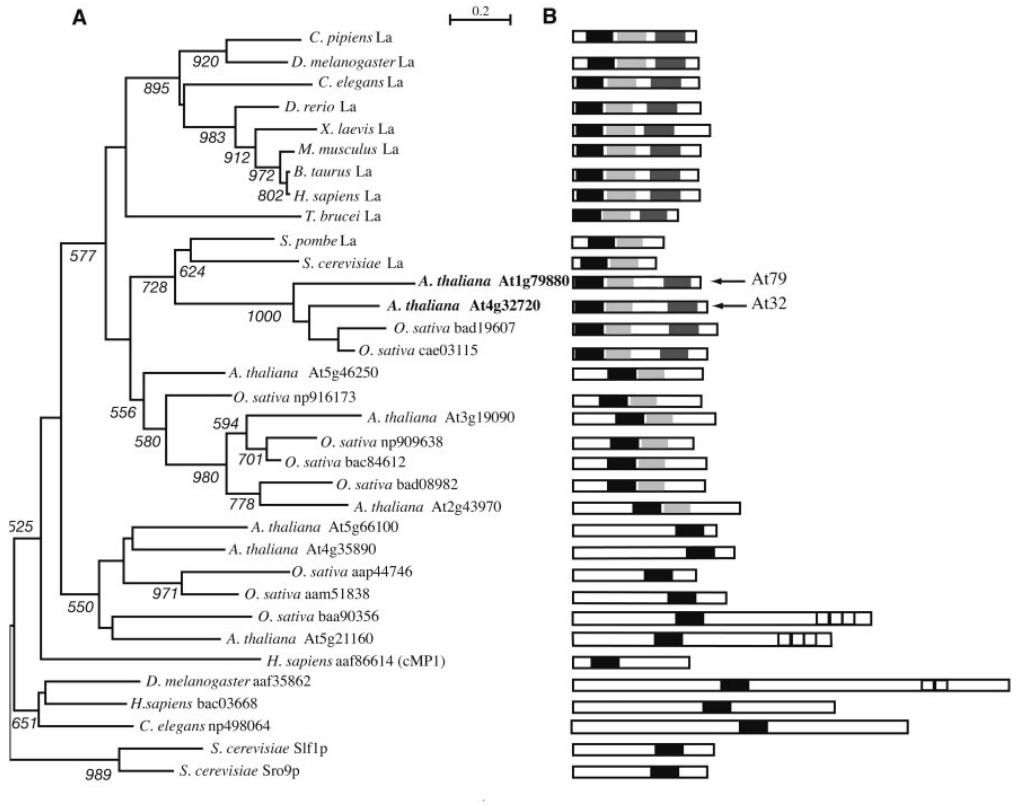


Figure 1. Comparison of La and La-like proteins from various eukaryotes. (**A**) Neighbour-Joining tree representing the phylogenetic relations among La motif-containing proteins. Numbers above each node indicate bootstrap values out of 1000 replicates (only values over 500 are presented). The amino acids divergence scale is indicated. The names of the proteins selected for further analysis are in bold. (**B**) Structural comparison of the different La domain-containing proteins. La domains are represented by black boxes, the first RRM domains by light grey boxes, the atypical second RRM by dark grey boxes and the DM15 motifs (a motif of unknown function) by open boxes. Arrows to the right respectively point to A1g79880 (At79) and At4g32720 (At32).

which possess a much shorter CTD. Altogether these data suggest that *Arabidopsis* At4g32720 and At1g79880 and rice bad19607 and cae03115 are the most likely *bona fide* La homologues of the two species.

The two *Arabidopsis* proteins, At4g32720 and At1g79880, have an overall amino acid identity of only 44%, with the highest identity found in the La motif (54%) and in the first RRM (59%) (see Supplementary Data Figure 1A online), suggesting that they are not produced from recently duplicated genes. In fact our phylogenetic studies show that At4g32720 is more closely related to the two rice proteins than it is to At1g79880. The presence of two distantly related proteins in *Arabidopsis* with characteristics of *bona fide* La proteins is exceptional as the eukaryote La function has always been associated to a single La protein [with the exception of two closely related (>90% amino acids identity) *Xenopus* La proteins (62)].

Expression of At4g32720 and At1g79880 in *Arabidopsis*

Searches of the *Arabidopsis* information resources (TAIR at http://www.arabidopsis.org) and NCBI databases revealed that full-length cDNAs corresponding to both loci have been sequenced. For the At4g32720 locus, all characterized ESTs and full-length cDNAs code for a single protein reported as At4g32720.1. For the At1g79880 locus, three different isoforms are suggested by EST data but only two of those (called At1g79880.1 and At1g79880.2) are supported by full-length cDNAs. The At1g79880.2 protein differs from At1g79880.1 by the deletion of the first 44 amino acids, including more than half of the La motif. Since At1g79880.2 lacks a complete La motif and since the third isoform (At1g79880.3) is only supported by a single EST sequence (corresponding to a small portion of the protein CTD), we decided to conduct further structural and functional studies on the At1g79880.1 protein. From now on, the At4g32720.1 and At1g79880.1 proteins will be referred to as At32 and At79, respectively.

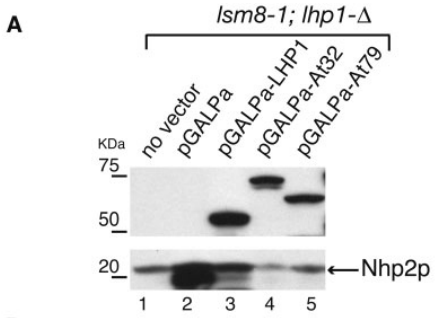
To monitor the developmental expression profiles of At32 and At79, we searched the expression atlas of *Arabidopsis* development microarray data (AtGenExpress) using the Genevestigator web site (https://www.genevestigator.ethz.ch/at/) (63). We observed that mRNAs

corresponding to both genes are present at all developmental stages in every plant tissue tested (see Supplementary Figure 2A online). At32 mRNA levels are highest in tissues composed of actively dividing cells (such as root tips, radicles, seedling, callus or cell suspensions), while At79 mRNAs is in general more abundant in tissues composed of mainly differentiated (non-dividing) cells. In average, At32 mRNAs are four times (4×) more abundant than that of At79 but this ratio is highly variable and is highest (10×) in 'young' tissues and lowest (2×) in 35-day-old senescent and cauline leaves (see Supplementary Figure 2B online). Western blot analysis using antibodies specific for At32 and At79 confirmed that both proteins are present in all developmental stages tested (see Supplementary Figure 2C online).

To further characterize the roles of At32 and At79, we first conducted functional studies in yeast and tested the ability of these plant proteins to complement phenotypes linked to the full depletion of Lhp1p, the *S. cerevisiae* La homologue (16).

At32 restores growth of the *lsm8-1;lhp1-Δ* colethal strain

While Lhp1p is not required for growth in wild-type cells, it becomes essential in specific genetic backgrounds (10,18,21,53). The Sm-like protein 8 (Lsm8p) is an essential member of the Lsm2p–Lsm8p ring-shaped complex (64,65). In the nucleus, Lsm8p associates with U6 snRNA and is important for its stability, for the formation of U6-containing snRNPs and for pre-mRNA splicing (18,64,66,67). The *lsm8-1* allele is not lethal in normal growth conditions but becomes essential when combined with a deletion of the *LHP1* gene. In the *lsm8-1* background, Lhp1p is essential to stabilize newly synthesized U6 RNAs and to facilitate the U6 snRNP assembly (18). The viability of the BP1 (*lsm8-1;lhp1-Δ*, *ade2*) colethal strain is maintained by an extra-chromosomal copy of the *LHP1* gene carried on an *ARSCEN*, *ADE2*, *TRP1* plasmid (pATL). As a red pigment accumulates in *ade2* mutant strains (68), cells that retain pATL will form white colonies, while cells that grow without the plasmid will form solid red colonies (18,53). The ability of plant proteins to restore growth in the *lsm8-1;lhp1-Δ* background can be monitored by the capacity of BP1 to form colonies with red sectors when transformed with a plasmid expressing the coding sequence of interest. DNA fragments corresponding to the *LHP1*, At32 and At79 ORFs were cloned in fusion at their NH2 terminus with the Protein A tag (50) and placed under the control of a galactose-inducible promoter on a low-copy *URA3* vector (48) to create the pGALPa plasmids collection. BP1 was transformed with the different pGALPa plasmids and grown under galactose-inducing conditions. To compare protein production levels in the different transformants, proteins were prepared from an aliquot of each culture and western blot analysis performed using a rabbit anti-Nhp2p (a snoRNP H/ACA protein) antibody (58) that, as for most IgGs, also binds to the Protein A tag (50). As expected, the anti-Nhp2p antibody recognizes in all extracts a single protein migrating above 20 kDa, corresponding to Nhp2p (Figure 2A). Additional bands



B

Expressed prot. / vector	Prot A	Lhp1p	At32	At79
no vector	<0.2			
pGALPa	0.6	22.2	2.9	0.7
pGAL		23.4	1.9	0.5
p2μGALPa	<0.2	45	8	<0.2

Figure 2. Complementation of the *lsm8-1;lhp1-Δ* colethal phenotype by heterologous expression of the At32 and At79 proteins. (**A**) Expression of At32 and At79 proteins in yeast. Western blot analysis of total protein extracts from the untransformed BP1 (*lsm8-1;lhp1-Δ*) strain (lane 1) or the BP1 strain expressing Protein A (lane 2), ProtA-Lhp1p (lane 3), ProtA-At32 (lane 4) or ProtA-At79 (lane 5) from the pGALPa set of vectors using the rabbit anti-Nhp2p antibody (58). The Lhp1p has a predicted molecular weight of 32 kDa but displays an electrophoretic mobility of 38 kDa (69), accordingly the ProtA-Lhp1p migrates above 50 kDa (lane 3). The At32 protein extracted from *Arabidopsis* has an apparent molecular weight of 56 kDa (see Supplementary Figure 2C online) and consistently the fusion protein migrates at ~74 kDa (lane 4). The ProtA-At79 fusion is detected at ~62 kDa (lane 5) as expected from the At79 predicted size of 44 kDa, which was confirmed by western blot analysis of plant extracts with antibodies specific for At79 (see Supplementary Figure 2C). (**B**) Sectoring assays. The BP1 strain transformed with the three sets of plasmids was grown under galactose-inducing conditions and plated on sectoring medium. Two independent experiments were conducted using a different transformant for each plasmid in each case. A total of ~1000 colonies per plasmid were analysed in each test and the percentage of sectoring colonies determined as: [(red colonies plus colonies with at least one red sector)/(number of colonies screened)] × 100. The reported numbers correspond to the average of the percentages obtained in each experiment.

corresponding to the expression of the Protein A alone (Figure 2A, lane 2) and to the different fusion proteins (Figure 2A, lanes 3–5) were detected in extracts from the BP1 strain transformed with the different pGALPa plasmids. Using the Nhp2p signal as a loading control, we can observe that all fusion proteins accumulate to similar levels (Figure 2A).

The galactose-induced cultures of transformed and untransformed BP1 strains were plated on sectoring

medium and the number of colonies with red sectors was counted and expressed as a percentage of the total number of colonies (Figure 2B). As expected, the ProtA-Lhp1p positive control restores red/white sectoring in the BP1 strain giving 22.2% of colonies with red sectors while background, as observed for the untransformed and Protein A-expressing colethal strain, is between <0.2% and 0.6%. Among the *Arabidopsis* proteins, only At32 gives red/white sectoring above background (2.9%). Using fluoroorotic-acid-containing medium (68), we confirmed that the red colonies growth is strictly dependent on the presence of the pGALPa vectors expressing ProtA-Lhp1p or ProtA-At32. The inability of the At79-tagged protein to restore growth in the BP1 strain might be due to the presence of the tag or to the fact that a higher level of this protein is required. To test these possibilities, we placed the untagged plant and *LHP1* ORFs under the control of the *GAL* promoter on the *ARSCEN-URA3* plasmid (pGAL vectors) and the galactose controlled protA-tagged ORFs on a high-copy *2μL-URA3* plasmid (48) (p2μGALPa vectors). Each set of plasmids was transformed into the BP1 strain and tested by western for equivalent protein expression levels using anti-Nhp2p antibodies for p2μGALPa vectors and At32- and At79-specific antibodies for the pGAL vectors (not shown). As with the pGALPa plasmids collection, we tested the ability of the new BP1 transformants to produce red sectors. The untagged proteins gave results similar to the previous ones showing that the presence of the NH2 tag has no clear impact on the expressed proteins (Figure 2B). As expected, the use of high-copy vectors leads to higher levels of red sectors for ProtA-Lhp1p and ProtA-At32 giving 45 and 8% red/white sectoting colonies respectively, whereas the At79 sectoring percentages remain at the background level (Figure 2B).

Altogether these data show that the At32 putative La homologue is able to complement the colethal *lsm8-1;lhp1-Δ* phenotype and suggest that At79 is not able to do so.

At32 is able to restore 3′-end processing of non-coding RNA precursors in an *lhp1-Δ* background

Inactivation of the *LHP1* gene in an otherwise wild-type background leads to several molecular phenotypes including a differential accumulation of U3A snoRNA (referred to as U3 from now on) and tRNA precursor species (9,16). To test the ability of the plant proteins to complement molecular phenotypes observed upon Lhp1p depletion, we expressed At32 and At79 from the p2μGALPa set of vectors in the CY2 (*lhp1::LEU2*) yeast strain which bears a fully inactivated allele of the *LHP1* gene (16). The transformed as well as untransformed CY2 and the isogenic CY3 wild-type strains (69) were grown on galactose-inducing medium. We confirmed by western blot analysis the proper and equivalent expressions of the different fusion proteins in the CY2 background under these growth conditions (data not shown). Total RNAs were extracted from the different transformants and the accumulation of U3 and tRNAs precursors assessed. Three independent experiments were performed and gave similar results.

Lhp1p has been shown to protect two U3-3′ extended forms from exonucleolytic trimming by binding to poly(U) stretches present in the 3′-end region of pre-U3 molecules (Figure 3A). These extended forms, named U3-3′I (U3 + 12) and U3-3′II (U3 + 18) are readily detected by northern blot in wild-type conditions (Figure 3A lane 2 and drawing). Following *LHP1* inactivation, the two extended forms are replaced by a heterogeneous population of molecules whose sizes range from +12 to +8 (9). Consequently, the impact of *LHP1* inactivation on pre-U3 3′-maturation can be assessed by monitoring the level of the U3-3′II extended form. Our analysis shows a strong impact on the accumulation of the U3-3′II form upon inactivation of *LHP1* (Figure 3A lane 1) but by contrast to previous report (9), there is no complete depletion of the band. This might result from strain discrepancies since we are using a different genetic background (16) than the one used by Kufel *et al*. (9). In extracts from the CY2 strain transformed with the p2μGALPa-LHP1 plasmid (Figure 3A lane 4), we observe a strong over-accumulation of U3-3′II (and likely of U3-3′I as well) as compared to the CY3 wild-type strain (Figure 3A compare lanes 2 and 4 and see also Supplementary Figure 3A online for quantification of the signals). Since the expression of the Protein A tag alone under the same strong promoter and from the same high-copy vector, has no influence on U3-3′II accumulation (Figure 3A lane 3), we conclude that the over-expression of Lhp1p is likely to be responsible of this effect (see footnotes on Supplementary Figure 3 online). In the presence of the ProtA-At32 protein, the U3-3′II precursor accumulates ~45% of the wild-type levels, which is around three times the residual levels observed in the untransformed CY2 strain (Figure 3A lane 5 and Supplementary Figure 3A online). Expression of At79 gives only a 5% increase in U3′-II levels as compared to the untransformed or Protein A expressing CY2 strain (Figure 3A lanes 1, 3 and 6 and see Supplementary Figure 3A online).

tRNAs are transcribed with 5′- and 3′-extensions and many contain intervening sequences. Removal of the 3′-extension is catalysed by an endonucleolytic event occurring most of the time after 5′-end excision (56,70,71). The La proteins protect the 3′-trailer from exonucleolytic trimming and, at least in yeasts, stimulate its endonucleolytic maturation (15,16,72,73). Although there is no clear defect in mature tRNA levels in the absence of Lhp1p, the mechanism of 3′-end tRNA maturation and the order by which it occurs are both altered. In *lhp1-Δ* background, the mature 3′-end is produced by exonucleolytic trimming, a process leading to characteristic modifications in the accumulation patterns of most tRNA precursors and intermediates [(15,16) and compare also lanes 1 and 2 Figure 3B]. We monitored these patterns for five intron-containing tRNA families upon expression of the plant proteins in the CY2 (*lhp1-Δ*) strain (Figure 3B and Supplementary Figure 3B online). The identities of tRNA precursors were determined based on their relative electrophoretic mobility, hybridization patterns

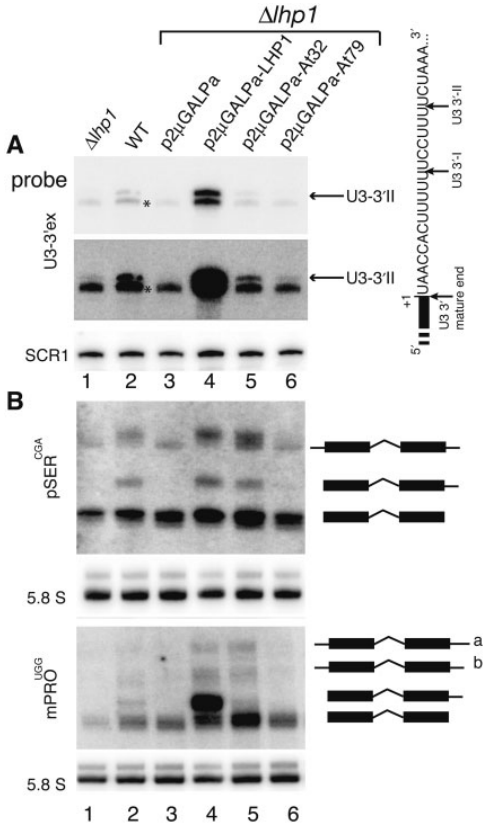


Figure 3. Complementation of *lhp1-Δ* strain pre-U3 and pre-tRNAs processing defects by heterologous expression of At32 and At79. The untransformed *lhp1-Δ* strain (CY2) (lane 1), the isogenic WT strain (CY3) (lane 2), or the CY2 (*lhp1-Δ*) strain transformed with the p2μGALPa set of plasmids allowing the expression of Protein A (lane 3), ProtA-Lhp1p (lane 4), ProtA-At32 (lane 5) or ProtA-At79 (lane 6) were grown in liquid minimal galactose medium and total RNAs extracted for analysis. (**A**) Northern blot analysis of U3 snoRNA precursors pattern. The same blot was probed with an oligonucleotide complementary to the pre-U3 3′-extension (first upper two panels) or to SCR1 (the RNA component of the signal recognition particle) (third panel) as loading control. The second panel corresponds to a longer exposure of the first one. The position of the U3-3′II species is reported on the right and the U3-3′I precursor is marked with an asterisk in the WT lane. The drawing on the right represents a portion of U3 snoRNA 3′-extension with arrows showing the 3′-ends of mature U3 and of U3 3′-I and U3 3′-II precursors, respectively. (**B**) Northern blot analysis of serine^{CGA} and proline^{UGG} tRNAs precursors pattern. Two blots were prepared and respectively probed with oligonucleotides complementary to intronic sequences of ser^{CGA} pre-tRNA (pSER^{CGA}) or to exonic sequences of proline^{UGG} tRNA (mPRO^{UGG}). Each blot was also probed for the 5.8S rRNA as loading control. Schematic representations of tRNA precursors, intermediate and mature species are shown on the right. tRNA^{ProUGG} precursors designated a and b represent two unspliced 5′- and 3′-extensions containing transcripts. Black boxes represent exonic sequences, central broken lines intronic sequences and 5′- and 3′-extensions are depicted by short lines.

and comparison to previous reports (16,56,70). The maturation defects we observed in the *lhp1-Δ* strain for tRNAs serine^{CGA} and proline^{UGG} (Figure 3B lane 1) are as previously reported (16). In these cases, upon Lhp1p depletion, the unspliced 5′- and 3′-extensions containing precursors migrate faster and under accumulate, and the processing intermediates corresponding to unspliced 5′-processed, 3′-unprocessed pre-tRNAs are undetectable (Figure 3B compare lanes 1 and 2).

As for pre-U3 3′-end processing, expression of the Protein A tag alone in CY2 has no impact on the tested pre-tRNA patterns (Figure 3B lane 3 and Supplementary Data Figure 3B online) showing that the effects described below most probably arise from the factor fused to the tag. As expected, expression of the ProtA-Lhp1p in the *lhp1-Δ* background qualitatively restores a wild-type processing pattern for all tested tRNA families (Figure 3B lane 4 and Supplementary Data Figure 3B online). Nevertheless, we can observe a striking over-accumulation of the 3′-extended intermediate of tRNA^{ProUGG} (and of other tested tRNAs, see Supplementary Figure 3B online), a more limited one for tRNA^{SerCGA}, and a slight increase in all unspliced 5′-, 3′-extended transcripts. As for U3, we speculate that all these molecular phenotypes are the consequence of Lhp1p over-expression (see footnotes of Supplementary Figure 3 online).

In the presence of the ProtA-At32 fusion, we clearly observe the restoration of a normal pattern of precursors and intermediates for all tested tRNA families (Figure 3B lane 5 and Supplementary Figure 3B). We also observe, as in the ProtA-Lhp1p-expressing strain but to a more limited extent, the over-accumulation of the 5′-, 3′-extended intron-containing transcripts (Figure 3B compare lanes 2 and 5). Pre-tRNA patterns from the ProtA-At79 expressing CY2 strain are similar to that of the untransformed and Protein A expressing CY2 strains (Figure 3B lane 6 and Supplementary Figure 3B) even after longer exposures (not shown).

Altogether these data show that At32 displays the ability to restore correct precursor patterns of U3 and every tRNA tested. We conclude that At32 is able to restore a wild-type 3′-end maturation process for these precursors and hence to support at least partially Lhp1p function in these processes. On the other hand, expression of At79 in the CY2 background does not allow U3 or tRNA precursors to accumulate. We conclude that expression of At79 is not able to support these Lhp1p functions under such heterologous conditions.

At32 binds to plant RNA polymerase III transcripts *in vivo*

La homologues can associate with highly diverse coding or non-coding RNAs from viral or cellular origins [for review see (5)]. This multi-functionality is in part explained by the capacity of La proteins to specifically bind with high-affinity RNAs presenting a terminal 3′-UUU-OH motif (37,38,40). All primary RNA pol III transcripts end with such a motif but mature forms do not, as it is subsequently removed by a 3′-end processing step. To determine whether At32 is able to bind to 3′-UUU-OH ending RNAs in plant, we asked whether RNA pol III precursors could be co-immunoprecipitated with green fluorescent protein (GFP) tagged versions of the protein. However, since most plant pol III precursors have a very short 3′-extension and are difficult to distinguish

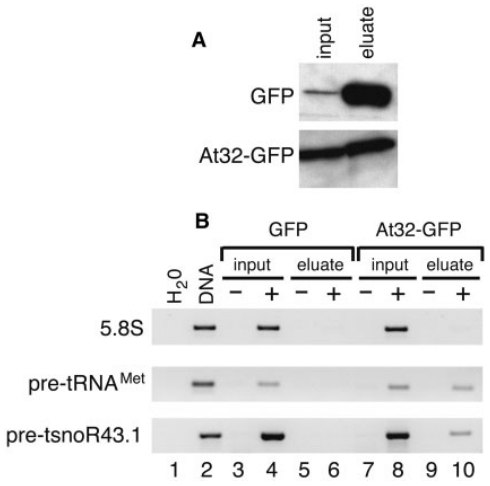


Figure 4. Co-immunoprecipitation experiments with an anti-GFP antibody on crude lysates of GFP or At32-GFP expressing *Arabidopsis* cell lines. (**A**) Western blot analysis of input and immunoprecipitation eluate fractions. Blot was probed with a monoclonal anti-GFP antibody. (**B**) PCR analysis of cDNAs prepared from input and immunoprecipitation eluate fractions. PCR with primer pairs specific to 5.8S cDNA (5.8S), to cDNAs of intron-containing precursors of $tRNA^{Met}$ family (pre-$tRNA^{Met}$) and to the cDNA of the tRNA-snoRNA43.1 precursor (pre-tsnoR43.1) were performed on water or genomic DNA as controls (lanes 1 and 2), on DNase treated but not reverse transcribed input and eluate fractions (minus lanes 3, 5, 7 and 9) and on DNase treated and reverse transcribed input and eluate fractions (plus lanes 4, 6, 8 and 10).

(by hybridization or PCR amplification) from mature forms, we had to select rare situations where this distinction would be possible. One such situation is provided by the presence of an intron in a subgroup of *Arabidopsis* pre-tRNA methionine (pre-$tRNA^{Met}$) (tRNA database: http://lowelab.ucsc.edu/GtRNAdb/) and another by the presence in plant of a dicistronic gene organization, where a tRNA is co-transcribed with a snoRNA in a single precursor molecule (pre-tsnoR43.1) by the RNA pol III complex (74).

Stable transgenic cell suspension lines expressing At32 fused at its NH2 or COOH terminus to the GFP, as well as a cell line expressing the GFP alone as negative control were produced. We prepared native extracts of each of these cell suspension lines and performed immunoprecipitations with anti-GFP antibodies. Two independent experiments that yielded similar results were conducted for each fusion and no significant difference was observed between the NH2 and COOH translational fusions. Western blot analysis performed on identical fractions of each eluate showed that both GFP and At32-GFP are efficiently immunoprecipitated (Figure 4A). We checked for the presence of pre-$tRNA^{Met}$ and pre-tsnoR43.1 in each eluate fraction by RT-PCR analysis (Figure 4B). As negative control, to test the specificity of our immunoprecipitation experiments, we also performed RT-PCR amplifications targeting 5.8S ribosomal RNA since La proteins are not known to associate to this highly abundant transcript. PCR reactions performed on total genomic DNA demonstrate that primers are efficient and give specific signals (Figure 4B lane 2) and since we were not able to amplify from non-reverse transcribed RNAs (Figure 4B lanes 3, 5, 7 and 9), we conclude that the signals obtained with reverse-transcribed samples arise from cDNAs and not from contaminating genomic DNA. Also, in each case, we were able to amplify a specific product using input cDNAs (Figure 4B lanes 4 and 8) showing that each of the three tested RNAs were present in corresponding crude lysates. PCR reactions using cDNAs produced from the At32-GFP immunoprecipitated fraction generated the expected product for pre-$tRNA^{Met}$ and pre-tsnoR43.1, but not for 5.8S rRNA (Figure 4B, lane 10), while PCR reactions using cDNAs produced from the GFP-alone fraction did not produce any PCR product (Figure 4B lane 6). We conclude that pol III transcribed precursors to $tRNA^{Met}$ and tRNA-snoR43.1 specifically co-immunoprecipitate with At32-GFP. Despite the fact that the GFP protein is efficiently immunoprecipitated and is more abundant than At32 in the eluate (Figure 4A), no such RNA is co-immunoprecipitated with GFP alone strongly supporting the conclusion that At32 is able to specifically associate at least to these RNA pol III transcripts *in vivo* and probably to 3′-UUU-OH ending RNAs in general.

To test the capacity of At79 to bind pol III precursor RNAs *in vivo*, we also produced stable transgenic cell suspension lines expressing At79 fused at its NH2 terminus to the GFP protein. Using the same immunoprecipitation procedure as above, we were unable to co-immunoprecipitate pre-$tRNA^{Met}$ or pre-tsnoR43.1 with the At79 fusion protein whereas the protein is efficiently recovered in the eluate fraction (not shown). These negative results were reproduced when performing the PCR reactions on twice the amounts of cDNAs from GFP-At79 eluates.

These results, together with the fact that the At32 is able to complement every tested Lhp1p functions in *S. cerevisiae* strongly suggest that At32 is able to fulfil genuine nuclear La functions in *Arabidopsis* and we hence propose to rename it AtLa1 for *Arabidopsis thaliana* La protein 1. Although displaying characteristic evolutionary and structural features of genuine La proteins, At79 failed to complement yeast phenotypes linked to *LHP1* inactivation and at least in our experimental conditions, does not seem to bind to the tested RNA pol III transcripts in plants.

AtLa1 localizes predominantly to plant nucleoplasm

To get a better understanding of the La function in plant, we further studied AtLa1 in *Arabidopsis*, starting with its subcellular localization. Cell lines expressing GFP (as a control) or AtLa1 fused to the GFP, were prepared as protoplasts and fluorescence was observed by confocal laser scanning microscopy (Figure 5). As expected, the GFP alone is distributed throughout cytoplasm and nucleoplasm with exclusion of the nucleolus (Figure 5A–C). In the vast majority of observed cells, GFP-AtLa1 (NH2 fusion) or AtLa1-GFP (COOH fusion) (Figure 5D–F and not shown) display a diffuse pattern

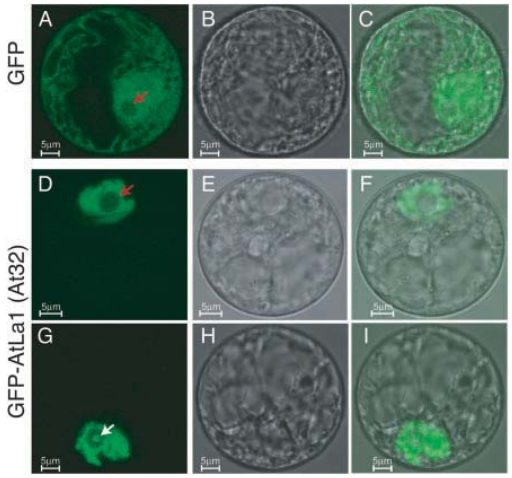


Figure 5. Subcellular distribution of GFP-AtLa1 (At32) fusion protein in *Arabidopsis* cell suspensions. Protoplasts were prepared from *Arabidopsis* cell suspensions stably expressing GFP or GFP-AtLa1 and fluorescent proteins localization assessed by laser confocal microscopy. (**A**–**C**) Protoplast expressing GFP alone. (B) is DIC image of (A), and (C) is merged image of (A) and (B). (**D**–**I**) Protoplasts expressing GFP-AtLa1. (E) is DIC image of (D), and (F) is merged image of (D) and (E). (H) is DIC image of (G), and (I) is merged image of (G) and (H). Red arrows point to the nucleolus and white arrow to the nucleolar cavity.

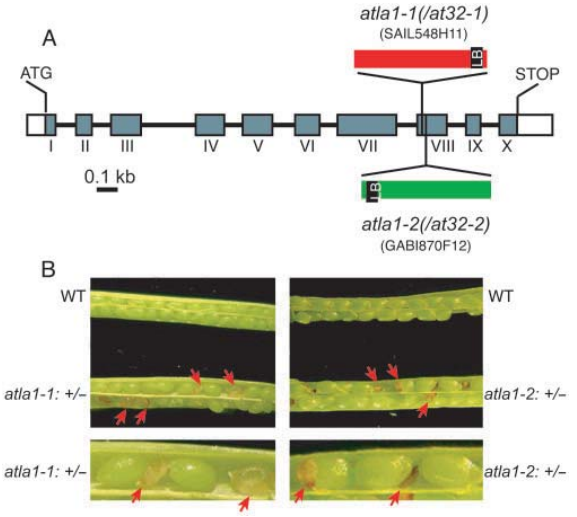


Figure 6. T-DNA insertion in *AtLa1* (/*At32*) gene is lethal. (**A**) Schematic representations of T-DNA insertions at the At4g32720 locus. Comparison of genomic and cDNA (At4g32720.1) sequences revealed that the coding region of the gene contains 10 exons (grey boxes). Open boxes represent 5′- and 3′-untranslated regions. The positions of Sail and Gabi T-DNAs are reported. (**B**) Siliques of hemizygous *AtLa1* insertion lines. Open siliques of self-pollinated *AtLa1*/*atla1-1* (left panels) and *AtLa1*/*atla1-2* (right panels) plants. As a control, siliques of a wild-type plant (*AtLa1*/*AtLa1*) are shown. Red arrows indicate the developmentally arrested siblings.

throughout the nucleoplasm and no obvious general labelling of the nucleolus.

The plant nucleoli have the particularity as compared to other eukaryotes, to display a central entity named nucleolar cavity that appears as a clear space when observed by electron microscopy (75,76). The precise function of the nucleolar cavity is not well defined but amongst other factors small nuclear and nucleolar RNAs were shown to localize in this space (76–78). We observed the accumulation of the AtLa1 NH2 and COOH GFP fusion proteins in the nucleolar cavity for a small fraction of cells (Figure 5G–I and not shown), whereas we never observed the GFP in this nucleolar subcompartment.

This observed subcellular localization is unlikely to be an artefact linked to the production of protoplasts from *Arabidopsis* cells, since the direct observation of a transgenic BY2 tobacco cell line stably expressing GFP-tagged AtLa1 gave the same result (not shown).

AtLa1 is required for viability

To further study AtLa1 functions, we searched several collections of T-DNA insertion mutants for disruption of the At4g32720 locus. We identified two candidates, one from the SAIL (n 548H11) (59) and one from the GABI-KAT (n 870F12) (60) collections. PCR analysis and sequencing allowed us to map each T-DNA insertion sites. Both Sail (allele *atla1-1*) and Gabi (allele *atla1-2*) T-DNAs are inserted at the beginning of the At4g32720 exon VIII, three bases apart (Figure 6A).

Sail and Gabi T-DNAs confer plant resistance to glufosinate ($Glufo^r$) and sulfadiazine (Sul^r) respectively. Plants bearing *atla1-1* and *atla1-2* alleles can hence be selected for growth on medium containing one of these drugs. After selection for several generations, no homozygous plants were recovered for neither allele, suggesting that the homozygous gene At4g32720 inactivation is lethal. The self-progenies of *AtLa1*/*atla1-1* and *AtLa1*/*atla1-2* always yielded a segregation for drug-sensitive to drug-resistant plants of 1:2 (1135 $Glufo^R$:601 $Glufo^S$, $\chi^2 = 1.29$, $P > 0.05$ and 885 Sul^R : 410 Sul^S, $\chi^2 = 1.63$, $P > 0.05$, respectively). These results suggest that the T-DNAs are inserted at a single locus in each case and that the two mutations are nuclear, recessive and sporophytic.

Observation of siliques from hemizygous *atla1* mutant plants under the microscope showed that they contained one-quarter of aborted seeds confirming the lethality of homozygous mutant embryos (Figure 6B and Table 1). Initially aborted seeds appeared smaller and pale brown at a stage where wild-type and hemizygous siblings were green (Figure 6B). Later they were dark brown and completely dried out.

Allelism tests were performed using the two independent T-DNA insertion lines. Immature siliques produced by crosses were opened under the binocular and analysed. The presence of 26.5% of *embryo-defective* seeds (156/589, $\chi^2 = 0.69$, $P > 0.05$) confirmed that the two mutations are allelic and that embryonic lethality directly results from the alteration of AtLa1 function.

Table 1. Comparison of hemizygous siliques from the two mutant lines at the wild-type torpedo/cotyledonary stage

	atla1-1	*atla1-2*
Total number of seeds analysed	533	592
Number of *emb* seeds	137	148
Percentage of *emb* seeds	25.7	25
χ^2(H0 = 1/4) $P > 0.05$	0.14	0
Number of *emb* seeds arrested at early globular stage	83	50
Number of brown and desiccated seeds unusable for phenotype analysis	54	98

atla1 embryos are arrested at early globular stage and display a nucleolar size enlargement

We examined the embryonic phenotype of *atla1* seeds for the two mutant lines (Figure 7 and Table 1). Compared to wild-type embryos at the torpedo/cotyledonary stage, the *atla1* embryos from both lines were much smaller and composed of <50 cells (Figure 7A panels b, c, e and f). Based on their morphology, we conclude that the terminal stage of *atla1* mutant embryo development is early globular (Figure 7A panels c and f).

Comparison of hemizygous siliques from the two mutant lines, taken at the torpedo/cotyledonary stage for wild-type seeds, revealed that the percentage of brown desiccated *atla1-2* seeds is almost twice that of *atla1-1* seeds (Table 1). Moreover embryos proper of *atla1-1* were often marginally larger and composed of more cells than *atla1-2* embryos (Figure 7A compare panels a–c and d–f). These data suggest that although the development of the embryos in both alleles is arrested at the same stage (Figure 7), *atla1-2* mutation is more severe than *atla1-1*.

We also observed that most if not all *atla1* embryo cells display abnormally large nucleoli (Figure 7B). We measured the diameter of nucleoli from *atla1-1* and *atla1-2* embryos arrested at early globular stage (~20 embryos were analysed for each allele with 4–6 nucleoli measured per embryo) and observed that they are 1.8–2 times larger than nucleoli from wild-type embryos at globular stage.

We conclude that the inactivation of AtLa1 function is lethal and leads to an early developmental arrest of the embryo.

DISCUSSION

We report here the identification of two distantly related proteins from *A. thaliana* At32(/AtLa1) and At79 that display every structural feature of *bona fide* La proteins. Microarray expression data from the AtGenexpress project and our western blot analysis clearly indicate that both corresponding loci (At4g32720 and At1g79880) are expressed *in vivo*. Based on its capacity to fulfil in yeast the small nuclear RNA biogenesis function of Lhp1p and to bind *in planta* to RNA pol III transcripts, we conclude that the protein produced at the At4g32720 locus (we named AtLa1 for *Arabidopsis thaliana* La protein 1) is a genuine *A. thaliana* La protein.

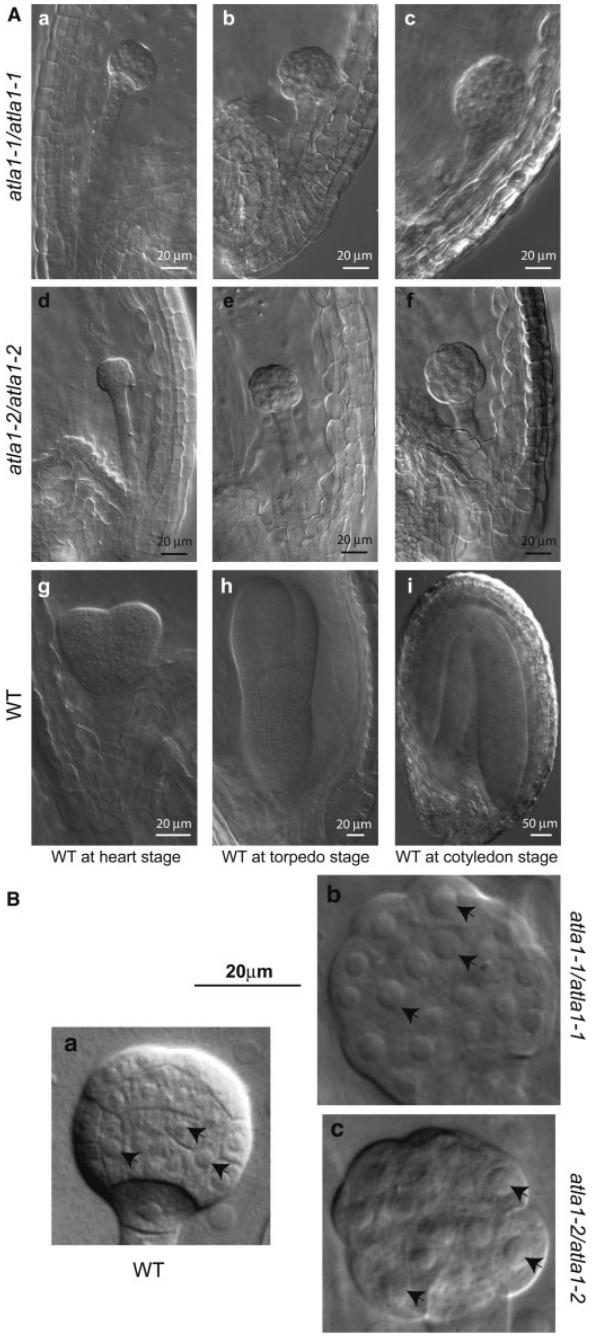


Figure 7. Phenotypes of *atla1* embryos. (**A**) *atla1* mutant embryos are developmentally arrested at early globular stage. Images of cleared seeds containing *atla1-1* embryos (a–c), *atla1-2* embryos (d–f) or wild-type embryos (g–i) are shown. Embryos were analysed at three developmental stages corresponding to wild-type heart-stage (a, d and g), torpedo stage (b, e and h) or cotyledon stage (c, f and i). (**B**) *atla1* mutant cells display a nucleolar size enlargement. Images of cleared seeds containing wild-type (a), *atla1-1* (b) or *atla1-2* (c) embryos at early globular stage. Arrows point to nucleoli. All three images are at the same scale.

The fact that AtLa1 can restore correct U3 and tRNA precursor patterns in *lhp1-Δ* background and growth of the *lsm8-1;lhp1-Δ* strain is a strong indication that it can bind to all RNAs that terminate with an oligouridylate motif (not only to pol III transcripts) and participate in their biogenesis. In yeast, U3 snoRNA is produced by the RNA pol II complex and binding of Lhp1p to precursor molecules is a key step of its maturation process (9). AtLa1 is able to restore correct U3 precursors pattern in an *lhp1-Δ* background showing that it can bind these pol II encoded RNAs and participate in their processing in yeast. This suggests that AtLa1 has the potentiality to bind to *Arabidopsis* pol II-encoded 3′-UUU-OH RNAs such as snoRNAs (79) or snRNAs (80) and to participate in their processing, stability and/or RNP formation.

AtLa1 is also able to restore normal pre-tRNA patterns in an *lhp1-Δ* background and even to stabilize intermediates above wild-type levels. This high level of complementation by a plant protein of the yeast La function in 3′-end pre-tRNAs processing is consistent with previous reports showing that this maturation is evolutionary well conserved in eukaryotes (15,81). This together with the fact that AtLa1 is able to bind *in vivo* to plant tRNA precursors allow us to propose that the mechanism of 3′-processing and protection of pre-tRNAs involves the La function and is conserved in *Arabidopsis*.

Finally, the capacity of AtLa1 to restore growth in the *lsm8-1;lhp1-Δ* strain is a strong indication that it can participate in U6 snRNA biogenesis. In yeast, both the ring-shaped complex Lsm2p-Lsm8p and the Lhp1p proteins are involved in the biogenesis of several non-coding RNAs such as U3 snoRNA, tRNAs and U6 snRNA (18,82,83). However, while *LSM8* or *LHP1* inactivation has no impact on the steady-state accumulation of mature U3 or tRNAs (82,83), U6 snRNA levels are reduced by 50% in the *lsm8-1* background (18,53). A 2–3-fold over-expression of Lhp1p in this background restores U6 snRNA levels to 75% of wild-type and expression of U6 snRNA in extra copies allows viability of the *lsm8-1;lhp1-Δ* strain (18,53). These data suggest that Lhp1p is needed in the *lsm8-1* background to help in U6 snRNP biosynthesis and/or accumulation and the ability of AtLa1 to restore growth of the *lsm8-1;lhp1-Δ* strain is a strong indication that it can at least partially replace Lhp1p in this role.

In summary, our data strongly suggest that in higher plants as in other eukaryotes, oligouridylate ending transcripts, whether encoded by RNA pol II or RNA pol III, require the La function for normal biogenesis and that this function is fulfilled by AtLa1.

A role for AtLa1 in the nucleolar cavity?

Steady-state cellular localization studies using the hLa protein revealed a diffuse nuclear pattern with some labelling of the nucleolus (24,25,84). In fact, hLa was shown to shuttle in and out of the nucleolus (25,26) and its distribution changes throughout the cell cycle with its major pool relocated to the nucleolus during late G1 and early S phases (24). The significance of hLa nucleolar accumulation is not fully understood but it could be involved in early steps of ribosomal RNA biogenesis (84), in the stabilization of some RNA pol III transcripts (such as U6 or pre-tRNAs as part of their nucleolar maturation and/or modification processes) and in binding U3 precursor and forming the corresponding RNP (3,26). As expected for a genuine La protein, we report that AtLa1 mainly localizes to the nucleoplasm in *Arabidopsis* cells. We did not observe a significant level of labelling of the nucleolus using the GFP-tagged AtLa1 proteins under steady-state conditions (except in the nucleolar cavity, see below). In agreement, AtLa1 was not identified in the proteomic analysis of the *Arabidopsis* nucleolus content (85) suggesting that AtLa1 is not a stable component of this nuclear subcompartment. However, it is possible that a small pool of AtLa1 shuttles through nucleolus and/or that the protein accumulates in this compartment at precise steps of the cell cycle. We also observed that, in a small fraction of cells, a significant pool of AtLa1 localizes to the nucleolar cavity. The nucleolar cavity is a highly dynamic entity, devoid of transcription, whose presence and structure is dependent upon cell cycle progression (75,76). Since we did not utilize a synchronous cell population to perform subcellular localization experiments, not every observed protoplast will contain a nucleolar cavity, and since the structure and probably content of this entity varies at least according to the cell cycle, it is possible that AtLa1 does not accumulate in this sub-nucleolar compartment at all times. This explains at least partially why only a sub-fraction of observed cells (whether *Arabidopsis* protoplasts or tobacco cells) displays fluorescent labelling of the nucleolar cavity. Several snoRNAs such as U3, spliceosomal snRNAs or the 7-2/MRP RNA component of the RNase MRP complex were reported to localize at least transiently to the nucleolar cavity (76–78). However, it is neither known if mature or precursor RNAs were detected in these studies, nor if other RNA pol III transcripts such as tRNAs could be present in this structure. The significance and the role of the nucleolar cavity is unknown but it was proposed to be a storage space and/or to host maturation processes (77). We speculate that AtLa1 could fulfil chaperone and/or stabilization functions for RNAs present in this cavity.

AtLa1 function is required early in embryogenesis

The La homologues from budding and fission yeasts are not required for viability in otherwise wild-type conditions (15,69). However, the situation in yeasts, where La proteins have a much shorter C-terminal domain compared to other eukaryotes, is likely to be exceptional. Indeed La homologues from mouse (86), fly (87) and trypanosome (13,88) are required for viability and we show here that this is also the case for AtLa1 in *Arabidopsis*.

The AtLa1 function is apparently not essential during gametogenesis but is required early in embryonic development. *atla1*-deficient embryos cannot reach the globular to heart stage transition and terminate development and desiccate at the early globular stage. This is similar to the situation in mouse where La is required early in development for the production of

healthy blastocysts (86), but contrasts with the *Drosophila melanogaster* situation where La becomes essential only at the late midgut stage (87).

An intriguing characteristic of *atla1* mutant embryos is the presence of unusually large nucleoli. Although a similar nucleolar hypertrophy was previously reported for other *Arabidopsis emb* mutants (*domino1*, *titan* and *plz*) (89–91), it is not a general feature of developmentally arrested embryos. The nucleolus is the site of ribosome biogenesis, and, in wild-type conditions, its size and morphology directly correlates with ribosome production activity (92,93). In genetically altered background, increase in nucleolus size can be a direct consequence of pre-ribosomal RNA processing or ribosome assembly alteration (89,94,95), or can indirectly result in the alteration of other cellular processes interconnected with nucleolar functions (96,97) as this is probably the case in *titan* and *plz* mutants which display, in addition to nucleolar hypertrophy, cell cycle and cytokinesis defects (90,91,98,99). In this case, however, we know that the La protein is likely to be important for ribosome biogenesis. Indeed, the human La protein was reported to interact in the nucleolus with nucleolin (84), a protein involved in ribosomal RNA production, processing, assembly and export of pre-ribosomal particles (100). Also, the yeast Lhp1p protein was shown to participate in the stabilization and folding of 5S rRNA and, when over-expressed, is able to restore 60S ribosomal subunit production in strains expressing *cis*-mutated 5S rRNAs (94). Therefore, we suggest that the nucleolar hypertrophy we observed in *atla1* mutant embryos likely results from a direct rather than indirect effect on ribosome biogenesis and nucleolar functions. However, since AtLa1 is probably a multifunctional protein, the perturbation of nucleolar functions is not necessarily responsible for the lethality phenotype we observed in the mutant.

Two distinct genuine La proteins in *Arabidopsis*?

The presence of two proteins with structural characteristics compatible with the La function is a feature that is likely to be conserved for all higher plants since this situation is present in *Arabidopsis* (a dicotyledon) and rice (a monocotyledon). The two *Arabidopsis* proteins present considerable sequence divergence with a mean amino acid identity of only 44%. This leads us to speculate that this sequence divergence was associated with a certain level of specialization and that the functions of the proteins are at least not fully redundant.

We showed that At79, the second putative *Arabidopsis* La homologue, possesses every structural feature of *bona fide* La proteins. It displays a characteristic La NTD organization and we clearly detected an atypical RRM2 motif in its CTD, which we believe is an additional structural hallmark of genuine La proteins (see Supplementary Figure 1 online). Nevertheless, At79 was not able to complement Lhp1p function in non-coding RNA biosynthesis in yeast, whereas AtLa1 did. This came as a surprise, especially concerning the well-conserved pre-tRNA maturation process (15,81). Several reasons could account why At79 is not functional in yeast among which, improper folding and/or subcellular localization of the protein. But we also failed to co-immunoprecipitate 3′-UUU-OH RNAs with GFP-tagged versions of At79 expressed in *Arabidopsis* cell suspensions. Although we cannot strictly rule out the possibility that the presence of the GFP tag has an inhibitory effect on At79's ability to bind to RNA pol III transcripts, we believe it is unlikely since La proteins from different organisms (hLa, *S. pombe* La, Lhp1p, AtLa1) were previously proven functional as GFP fusions [(25,41,101) and our present results]. We therefore conclude from our immunoprecipitation and yeast complementation results that At79 is probably not able to bind to RNA pol III transcripts and is not involved in their biogenesis.

This situation where a protein structurally similar to a genuine La factor is apparently not competent to fulfil the most 'basic' La function, that is to bind to RNA pol III precursor RNAs, is unprecedented and indeed puzzling. However, it is now clear that there is more to the La function than just binding RNA pol III precursors [for review see (5)] and we propose that At79 is not fully redundant with AtLa1 and that this protein specialized to fulfil other aspects of the La function. In support of this hypothesis, the mRNA patterns of expression of both genes are not superimposed across plant development and their levels are not regulated in the same way in different stress situations such as cold and heat shocks (AtGenExpress, data not shown). Moreover, we observed that AtLa1 is essential for viability, whereas At79 is dispensable for normal plant growth and development (not shown), suggesting that At79 is not able to fulfil entirely the AtLa1 function and that there is at least partial specialization of the two La homologues in *Arabidopsis*. This specialization is unprecedented in eukaryotes and studying the La function in plants should provide unique opportunities to dissect the very different aspects of this crucial cellular activity.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We are grateful to Sandra Wolin (HHMI, New Haven) for generous gift of yeast strains and plasmid pATL and to K. Helmut and E. Hurt (University of Heidelberg) for gift of the plasmid pGALPATG1L. We are thankful to M. Caizergues-Ferrer (LBME, Toulouse) for providing us with anti-Nhp2p antibodies. We would like to thank M.C. Espagnol and C. Goubely for technical help as well as people from UMR384/CBRV (Clermont-Ferrand) for access to the confocal microscope facility. S.F. was the recipient of a French MNERT grant. This work was supported by the CNRS, by the Université Blaise Pascal and by European Community grant (RIBOREG FP6 project). Funding to pay the Open Access publication charges for this article was provided by European community (grant RIBOREG FP6 project).

Conflict of interest statement. None declared.

REFERENCES

1. Alspaugh,M.A. and Tan,E.M. (1975) Antibodies to cellular antigens in Sjögren's syndrome. *J. Clinical Invest*., **55**, 1067–1073.
2. Mattioli,M. and Reichlin,M. (1973) Physical association of two nuclear antigens and mutual occurrence of their antibodies: the relationship of the SM and RNA protein (MO) systems in SLE sera. *J. Immunol*., **110**, 1318–1324.
3. Maraia,R.J. (2001) La protein and the trafficking of nascent RNA polymerase III transcripts. *J. Cell Biol*., **153**, 13–18.
4. Maraia,R.J. and Intine,R.V. (2001) Recognition of nascent RNA by the human La antigen: conserved and diverged features of structure and function. *Mol. Cell. Biol*., **21**, 367–379.
5. Wolin,S.L. and Cedervall,T. (2002) The La protein. *Annu. Rev. Biochem*., **71**, 375–403.
6. Marchetti,M.A., Tschudi,C., Kwon,H., Wolin,S.L. and Ullu,E. (2000) Import of proteins into the trypanosome nucleus and their distribution at karyokinesis. *J. Cell Sci*., **113**, 899–906.
7. Stefano,J.E. (1984) Purified lupus antigen La recognizes an oligouridylate stretch common to the 3′ termini of RNA polymerase III transcripts. *Cell*, **36**, 145–154.
8. Inada,M. and Guthrie,C. (2004) Identification of Lhp1p-associated RNAs by microarray analysis in Saccharomyces cerevisiae reveals association with coding and noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **101**, 434–439.
9. Kufel,J., Allmang,C., Chanfreau,G., Petfalski,E., Lafontaine,D. and Tollervey,D. (2000) Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Mol. Cell. Biol*., **20**, 5415–5424.
10. Xue,D., Rubinson,D.A., Pannone,B.K., Yoo,C.J. and Wolin,S.L. (2000) U snRNP assembly in yeast involves the La protein. *EMBO J*., **19**, 1650–1660.
11. Maraia,R.J. and Intine,R.V. (2002) La protein and its associated small nuclear and nucleolar precursor RNAs. *Gene Expr*., **10**, 41–57.
12. Preiser,P., Vasisht,V., Birk,A. and Levinger,L. (1993) Poly(U)-binding protein inhibits Drosophila pre-5 S RNA 3′-exonuclease digestion. *J. Biol. Chem*., **268**, 11553–11557.
13. Foldynova-Trantirkova,S., Paris,Z., Sturm,N.R., Campbell,D.A. and Lukes,J. (2005) The Trypanosoma brucei La protein is a candidate poly(U) shield that impacts spliced leader RNA maturation and tRNA intron removal. *Int. J. Parasitol*., **35**, 359–366.
14. Intine,R.V., Sakulich,A.L., Koduru,S.B., Huang,Y., Pierstorff,E., Goodier,J.L., Phan,L. and Maraia,R.J. (2000) Control of transfer RNA maturation by phosphorylation of the human La antigen on serine 366. *Mol. Cell*, **6**, 339–348.
15. Van Horn,D.J., Yoo,C.J., Xue,D., Shi,H. and Wolin,S.L. (1997) The La protein in Schizosaccharomyces pombe: a conserved yet dispensable phosphoprotein that functions in tRNA maturation. *RNA*, **3**, 1434–1443.
16. Yoo,C.J. and Wolin,S.L. (1997) The yeast La protein is required for the 3′ endonucleolytic cleavage that matures tRNA precursors. *Cell*, **89**, 393–402.
17. Huang,Y., Bayfield,M.A., Intine,R.V. and Maraia,R.J. (2006) Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. *Nat. Struct. Mol. Biol*., **13**, 611–618.
18. Pannone,B.K., Xue,D. and Wolin,S.L. (1998) A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J*., **17**, 7442–7453.
19. Belisova,A., Semrad,K., Mayer,O., Kocian,G., Waigmann,E., Schroeder,R. and Steiner,G. (2005) RNA chaperone activity of protein components of human Ro RNPs. *RNA*, **11**, 1084–1094.
20. Copela,L.A., Chakshusmathi,G., Sherrer,R.L. and Wolin,S.L. (2006) The La protein functions redundantly with tRNA modification enzymes to ensure tRNA structural stability. *RNA*, **12**, 644–654.
21. Chakshusmathi,G., Kim,S.D., Rubinson,D.A. and Wolin,S.L. (2003) A La protein requirement for efficient pre-tRNA folding. *EMBO J*., **22**, 6562–6572.
22. Kadaba,S., Wang,X. and Anderson,J.T. (2006) Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA*, **12**, 508–521.
23. Broekhuis,C.H., Neubauer,G., van der Heijden,A., Mann,M., Proud,C.G., van Venrooij,W.J. and Pruijn,G.J. (2000) Detailed analysis of the phosphorylation of the human La (SS-B) autoantigen. (De)phosphorylation does not affect its subcellular distribution. *Biochemistry*, **39**, 3023–3033.
24. Deng,J.S., Takasaki,Y. and Tan,E.M. (1981) Nonhistone nuclear antigens reactive with autoantibodies. Immunofluorescence studies of the distribution in synchronized cells. *J. Cell Biol*., **91**, 654–660.
25. Horke,S., Reumann,K., Schweizer,M., Will,H. and Heise,T. (2004) Nuclear trafficking of La protein depends on a newly identified nucleolar localization signal and the ability to bind RNA. *J. Biol. Chem*., **279**, 26563–26570.
26. Horke,S., Reumann,K., Schulze,C., Grosse,F. and Heise,T. (2004) The La motif and the RNA recognition motifs of human La autoantigen contribute individually to RNA recognition and subcellular localization. *J. Biol. Chem*., **279**, 50302–50309.
27. Cardinali,B., Carissimi,C., Gravina,P. and Pierandrei-Amaldi,P. (2003) La protein is associated with terminal oligopyrimidine mRNAs in actively translating polysomes. *J. Biol. Chem*., **278**, 35145–35151.
28. Fok,V., Friend,K. and Steitz,J.A. (2006) Epstein-Barr virus noncoding RNAs are confined to the nucleus, whereas their partner, the human La protein, undergoes nucleocytoplasmic shuttling. *J. Cell Biol*., **173**, 319–325.
29. Ayukawa,K., Taniguchi,S., Masumoto,J., Hashimoto,S., Sarvotham,H., Hara,A., Aoyama,T. and Sagara,J. (2000) La autoantigen is cleaved in the COOH terminus and loses the nuclear localization signal during apoptosis. *J. Biol. Chem*., **275**, 34465–34470.
30. Meerovitch,K., Svitkin,Y.V., Lee,H.S., Lejbkowicz,F., Kenan,D.J., Chan,E.K., Agol,V.I., Keene,J.D. and Sonenberg,N. (1993) La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. *J. Virol*., **67**, 3798–3807.
31. Yocupicio-Monroy,M., Padmanabhan,R., Medina,F. and Del Angel,R.M. (2006) Mosquito La protein binds to the 3′ untranslated region of the positive and negative polarity dengue virus RNAs and relocates to the cytoplasm of infected cells. *Virology*.
32. Costa-Mattioli,M., Svitkin,Y. and Sonenberg,N. (2004) La autoantigen is necessary for optimal function of the poliovirus and hepatitis C virus internal ribosome entry site in vivo and in vitro. *Mol. Cell. Biol*., **24**, 6861–6870.
33. Dasgupta,A., Das,S., Izumi,R., Venkatesan,A. and Barat,B. (2004) Targeting internal ribosome entry site (IRES)-mediated translation to block hepatitis C and other RNA viruses. *FEMS Microbiol. Lett*., **234**, 189–199.
34. Holcik,M. and Korneluk,R.G. (2000) Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: role of La autoantigen in XIAP translation. *Mol. Cell. Biol*., **20**, 4648–4657.
35. Kim,Y.K., Back,S.H., Rho,J., Lee,S.H. and Jang,S.K. (2001) La autoantigen enhances translation of BiP mRNA. *Nucleic Acids Res*., **29**, 5009–5016.
36. Kenan,D.J., Query,C.C. and Keene,J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biol. Sci*., **16**, 214–220.
37. Alfano,C., Sanfelice,D., Babon,J., Kelly,G., Jacks,A., Curry,S. and Conte,M.R. (2004) Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nat. Struct. Mol. Biol*., **11**, 323–329.
38. Dong,G., Chakshusmathi,G., Wolin,S.L. and Reinisch,K.M. (2004) Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch. *EMBO J*., **23**, 1000–1007.
39. Kenan,D.J. and Keene,J.D. (2004) La gets its wings. *Nat. Struct. Mol. Biol*., **11**, 303–305.
40. Maraia,R.J. and Bayfield,M.A. (2006) The La protein-RNA complex surfaces. *Mol. Cell*, **21**, 149–152.

41. Intine,R.V., Dundr,M., Misteli,T. and Maraia,R.J. (2002) Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs. *Mol. Cell*, **9**, 1113–1123.
42. Jacks,A., Babon,J., Kelly,G., Manolaridis,I., Cary,P.D., Curry,S. and Conte,M.R. (2003) Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure*, **11**, 833–843.
43. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
44. Veerassamy,S., Smith,A. and Tillier,E.R. (2003) A transition probability model for amino acid substitutions from blocks. *J. Comput. Biol.*, **10**, 997–1010.
45. Schmidt,H.A., Strimmer,K., Vingron,M. and von Haeseler,A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, **18**, 502–504.
46. Hedge,S.B. (1992) The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Mol. Biol. Evol.*, **9**, 366–369.
47. De Veylder,L., de Almeida Engler,J., Burssens,S., Manevski,A., Lescure,B., Van Montagu,M., Engler,G. and Inze,D. (1999) A new D-type cyclin of Arabidopsis thaliana expressed during lateral root primordia formation. *Planta*, **208**, 453–462.
48. Bonneaud,N., Ozier-Kalogeropoulos,O., Li,G., Labouesse,M., Minvielle-Sebastia,L. and Lacroute,F. (1991) A family of low and high copy replicative, integrative and single-stranded *S. cerevisiae*/*E. coli* shuttle vectors. *Yeast*, **7**, 609–615.
49. Guarente,L., Yocum,R.R. and Gifford,P. (1982) A GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. *Proc. Natl Acad. Sci. USA*, **79**, 7410–7414.
50. Nilsson,B., Moks,T., Jansson,B., Abrahmsen,L., Elmblad,A., Holmgren,E., Henrichson,C., Jones,T.A. and Uhlen,M. (1987) A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng.*, **1**, 107–113.
51. Becker,D. (1990) Binary vectors which allow the exchange of plant selectable markers and reporter genes. *Nucleic Acids Res.*, **18**, 203.
52. Guthrie,C. and Fink,G.F. (1991) *"Guide to Yeast Genetics and Molecular Biology"*, Vol 194, Methods in Enzymlogy.
53. Pannone,B.K., Kim,S.D., Noe,D.A. and Wolin,S.L. (2001) Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein. *Genetics*, **158**, 187–196.
54. Gietz,R.D., Schiestl,R.H., Willems,A.R. and Woods,R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, **11**, 355–360.
55. Beltrame,M. and Tollervey,D. (1992) Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J.*, **11**, 1531–1542.
56. O'Connor,J.P. and Peebles,C.L. (1991) In vivo pre-tRNA processing in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **11**, 425–439.
57. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning, A Laboratory Manual3rd edn*. Cold Spring Harbor Laboratory Press, Cold Spring harbor, New York.
58. Henras,A., Dez,C., Noaillac-Depeyre,J., Henry,Y. and Caizergues-Ferrer,M. (2001) Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic Acids Res.*, **29**, 2733–2746.
59. Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D., Dietrich,B., Ho,P., Bacwaden,J. *et al.* (2002) A high-throughput Arabidopsis reverse genetics system. *Plant Cell*, **14**, 2985–2994.
60. Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B. (2003) An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol.*, **53**, 247–259.
61. Gallego,M.E., Sirand-Pugnet,P. and White,C.I. (1999) Positive-negative selection and T-DNA stability in Arabidopsis transformation. *Plant Mol. Biol.*, **39**, 83–93.
62. Scherly,D., Stutz,F., Lin-Marq,N. and Clarkson,S.G. (1993) La proteins from Xenopus laevis. cDNA cloning and developmental expression. *J. Mol. Biol.*, **231**, 196–204.
63. Zimmermann,P., Hirsch-Hoffmann,M., Hennig,L. and Gruissem,W. (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol.*, **136**, 2621–2632.
64. Achsel,T., Brahms,H., Kastner,B., Bachi,A., Wilm,M. and Luhrmann,R. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J.*, **18**, 5789–5802.
65. Zaric,B., Chami,M., Remigy,H., Engel,A., Ballmer-Hofer,K., Winkler,F.K. and Kambach,C. (2005) Reconstitution of two recombinant LSM protein complexes reveals aspects of their architecture, assembly and function. *J. Biol. Chem.*, **280**, 16066–16075.
66. Mayes,A.E., Verdone,L., Legrain,P. and Beggs,J.D. (1999) Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J.*, **18**, 4321–4331.
67. Salgado-Garrido,J., Bragado-Nilsson,E., Kandels-Lewis,S. and Seraphin,B. (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J.*, **18**, 3451–3462.
68. Hieter,P., Mann,C., Snyder,M. and Davis,R.W. (1985) Mitotic stability of yeast chromosomes: a colony color assay that measures nondisjunction and chromosome loss. *Cell*, **40**, 381–392.
69. Yoo,C.J. and Wolin,S.L. (1994) La proteins from Drosophila melanogaster and Saccharomyces cerevisiae: a yeast homolog of the La autoantigen is dispensable for growth. *Mol. Cell. Biol.*, **14**, 5412–5424.
70. Kufel,J. and Tollervey,D. (2003) 3′-Processing of yeast tRNATrp precedes 5′-processing. *RNA*, **9**, 202–208.
71. Wolin,S.L. and Matera,A.G. (1999) The trials and travels of tRNA. *Genes Dev.*, **13**, 1–10.
72. Fan,H., Goodier,J.L., Chamberlain,J.R., Engelke,D.R. and Maraia,R.J. (1998) 5′ processing of tRNA precursors can be modulated by the human La antigen phosphoprotein. *Mol. Cell. Biol.*, **18**, 3201–3211.
73. Lin-Marq,N. and Clarkson,S.G. (1998) Efficient synthesis, termination and release of RNA polymerase III transcripts in Xenopus extracts depleted of La protein. *EMBO J.*, **17**, 2033–2041.
74. Kruszka,K., Barneche,F., Guyot,R., Ailhas,J., Meneau,I., Schiffer,S., Marchfelder,A. and Echeverria,M. (2003) Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J.*, **22**, 621–632.
75. Gonzalez-Camacho,F. and Medina,F.J. (2006) The nucleolar structure and the activity of NopA100, a nucleolin-like protein, during the cell cycle in proliferating plant cells. *Histochem. Cell Biol.*, **125**, 139–153.
76. Shaw,P.J. and Brown,J.W. (2004) Plant nuclear bodies. *Curr. Opin. Plant Biol.*, **7**, 614–620.
77. Beven,A.F., Lee,R., Razaz,M., Leader,D.J., Brown,J.W. and Shaw,P.J. (1996) The organization of ribosomal RNA processing correlates with the distribution of nucleolar snRNAs. *J. Cell Sci.*, **109**, 1241–1251.
78. Beven,A.F., Simpson,G.G., Brown,J.W. and Shaw,P.J. (1995) The organization of spliceosomal components in the nuclei of higher plants. *J. Cell Sci.*, **108**, 509–518.
79. Brown,J.W., Echeverria,M. and Qu,L.H. (2003) Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci.*, **8**, 42–49.
80. Connelly,S., Marshallsay,C., Leader,D., Brown,J.W. and Filipowicz,W. (1994) Small nuclear RNA genes transcribed by either RNA polymerase II or RNA polymerase III in monocot plants share three promoter elements and use a strategy to regulate gene expression different from that used by their dicot plant counterparts. *Mol. Cell. Biol.*, **14**, 5910–5919.
81. Morl,M. and Marchfelder,A. (2001) The final cut: the importance of tRNA 3′-end processing. *EMBO rep.*, **2**, 17–20.
82. Kufel,J., Allmang,C., Verdone,L., Beggs,J. and Tollervey,D. (2003) A complex pathway for 3′ processing of the yeast U3 snoRNA. *Nucleic Acids Res.*, **31**, 6788–6797.
83. Kufel,J., Allmang,C., Verdone,L., Beggs,J.D. and Tollervey,D. (2002) Lsm proteins are required for normal processing of

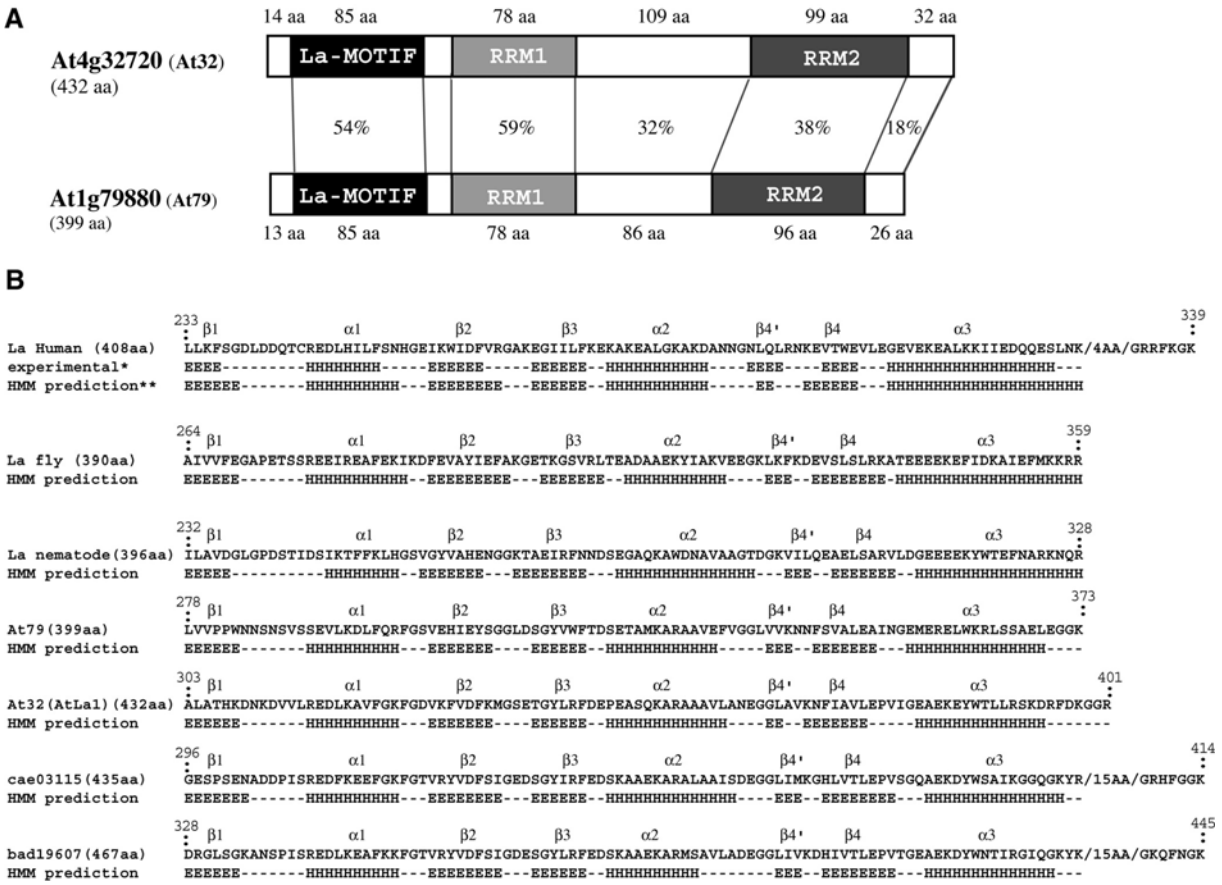
pre-tRNAs and their efficient association with La-homologous protein Lhp1p. *Mol. Cell. Biol.*, **22**, 5248–5256.
84. Intine,R.V., Dundr,M., Vassilev,A., Schwartz,E., Zhao,Y., Zhao,Y., Depamphilis,M.L. and Maraia,R.J. (2004) Nonphosphorylated human La antigen interacts with nucleolin at nucleolar sites involved in rRNA biogenesis. *Mol. Cell. Biol.*, **24**, 10894–10904.
85. Pendle,A.F., Clark,G.P., Boon,R., Lewandowska,D., Lam,Y.W., Andersen,J., Mann,M., Lamond,A.I., Brown,J.W. *et al.* (2005) Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol. Biol. Cell*, **16**, 260–269.
86. Park,J.M., Kohn,M.J., Bruinsma,M.W., Vech,C., Intine,R.V., Fuhrmann,S., Grinberg,A., Mukherjee,I., Love,P.E. *et al.* (2006) The multifunctional RNA-binding protein La is required for mouse development and for the establishment of embryonic stem cells. *Mol. Cell. Biol.*, **26**, 1445–1451.
87. Bai,C. and Tolias,P.P. (2000) Genetic analysis of a La homolog in Drosophila melanogaster. *Nucleic Acids Res.*, **28**, 1078–1084.
88. Arhin,G.K., Shen,S., Perez,I.F., Tschudi,C. and Ullu,E. (2005) Downregulation of the essential Trypanosoma brucei La protein affects accumulation of elongator methionyl-tRNA. *Mol. Biochem. Parasitol.*, **144**, 104–108.
89. Lahmy,S., Guilleminot,J., Cheng,C.M., Bechtold,N., Albert,S., Pelletier,G., Delseny,M. and Devic,M. (2004) DOMINO1, a member of a small plant-specific gene family, encodes a protein essential for nuclear and nucleolar functions. *Plant J.*, **39**, 809–820.
90. Liu,C.M. and Meinke,D.W. (1998) The titan mutants of Arabidopsis are disrupted in mitosis and cell cycle control during seed development. *Plant J.*, **16**, 21–31.
91. Steinborn,K., Maulbetsch,C., Priester,B., Trautmann,S., Pacher,T., Geiges,B., Kuttner,F., Lepiniec,L., Stierhof,Y.D. *et al.* (2002) The Arabidopsis PILZ group genes encode tubulin-folding cofactor orthologs required for cell division but not cell growth. *Genes Dev.*, **16**, 959–971.
92. Hernandez-Verdun,D. (2006) The nucleolus: a model for the organization of nuclear functions. *Histochem. Cell. Biol.*, **126**, 135–148.
93. Hernandez-Verdun,D. (2006) Nucleolus: from structure to dynamics. *Histochem. Cell. Biol.*, **125**, 127–137.
94. Dechampesme,A.M., Koroleva,O., Leger-Silvestre,I., Gas,N. and Camier,S. (1999) Assembly of 5S ribosomal RNA is required at a specific step of the pre-rRNA processing pathway. *J. Cell. Biol.*, **145**, 1369–1380.
95. Sicard,H., Faubladier,M., Noaillac-Depeyre,J., Leger-Silvestre,I., Gas,N. and Caizergues-Ferrer,M. (1998) The role of the Schizosaccharomyces pombe gar2 protein in nucleolar structure and function depends on the concerted action of its highly charged N terminus and its RNA-binding domains. *Mol. Biol. Cell.*, **9**, 2011–2023.
96. Sirri,V., Hernandez-Verdun,D. and Roussel,P. (2002) Cyclin-dependent kinases govern formation and maintenance of the nucleolus. *J. Cell. Biol.*, **156**, 969–981.
97. Dez,C. and Tollervey,D. (2004) Ribosome synthesis meets the cell cycle. *Curr. Opin. Microbiol.*, **7**, 631–637.
98. Liu,C.M., McElver,J., Tzafrir,I., Joosen,R., Wittich,P., Patton,D., Van Lammeren,A.A. and Meinke,D. (2002) Condensin and cohesin knockouts in Arabidopsis exhibit a titan seed phenotype. *Plant J.*, **29**, 405–415.
99. Tzafrir,I., McElver,J.A., Liu,C.M., Yang,L.J., Wu,J.Q., Martinez,A., Patton,D.A. and Meinke,D.W. (2002) Diversity of TITAN functions in Arabidopsis seed development. *Plant Physiol.*, **128**, 38–51.
100. Pontvianne,F., Matia,I., Douet,J., Tourmente,S., Medina,F.J., Echeverria,M. and Saez-Vasquez,J. (2007) Characterization of AtNUC-L1 reveals a central role of nucleolin in nucleolus organization and silencing of AtNUC-L2 Gene in Arabidopsis. *Mol. Biol. Cell.*, **18**, 369–379.
101. Rosenblum,J.S., Pemberton,L.F., Bonifaci,N. and Blobel,G. (1998) Nuclear import and the evolution of a multifunctional RNA-binding protein. *J. Cell Biol.*, **143**, 887–899.

SUPPLEMENTARY DATA

Supplementary Figure 1. Structural comparison of At4g32720 and At1g79880 and identification of atypical RRM2 in genuine La proteins CTD.

(A) Level of amino acids identity between the different domains of At4g32720 (At4g32720.1 protein) and At1g79880 (At1g79880.1 protein).

(B) Topology of the secondary structure elements (named as in (1)) of the atypical second RRM present in La proteins CTD from different species: Human, fly (*D.melanogaster*), worm (*C.elegans*), *Arabidopsis* (At79 (At79880.1) and At32 (At4g32720.1)) and rice (*O. sativa*: cae03115 and bad19607). In each case, the amino acids sequence of the region is given (the number at the start and end of the sequence is in reference to the position of the region in the complete protein). The experimentally determined (*) (1) and/or predicted secondary structure (**) of the region (SAM-T02, http://www.cse.ucsc.edu/research/compbio/HMM-apps/T02-query.html) is presented under the amino acids sequence. H: alpha helix, E: extended strand in beta ladder. The presence of an RRM2 in every genuine La tested but not in La-like factors suggest that it is not restricted to vertebrate La proteins as previously proposed (2), but rather is a conserved feature of *bona fide* La proteins with the exception of the two yeast proteins.



Footnotes on Supplementary Figure 1. Atypical RRM2 motifs were searched in every La and La-related proteins utilised to build the Neighbour-joining tree presented in Figure 1A and was exclusively detected in genuine La homologues. The presence of such RRM2 in every genuine La but not in La-related factors suggests that it is not restricted to vertebrate La homologues as previously proposed (2), but that it is a conserved feature of *bona fide* La proteins in general and that it carries conserved functions.

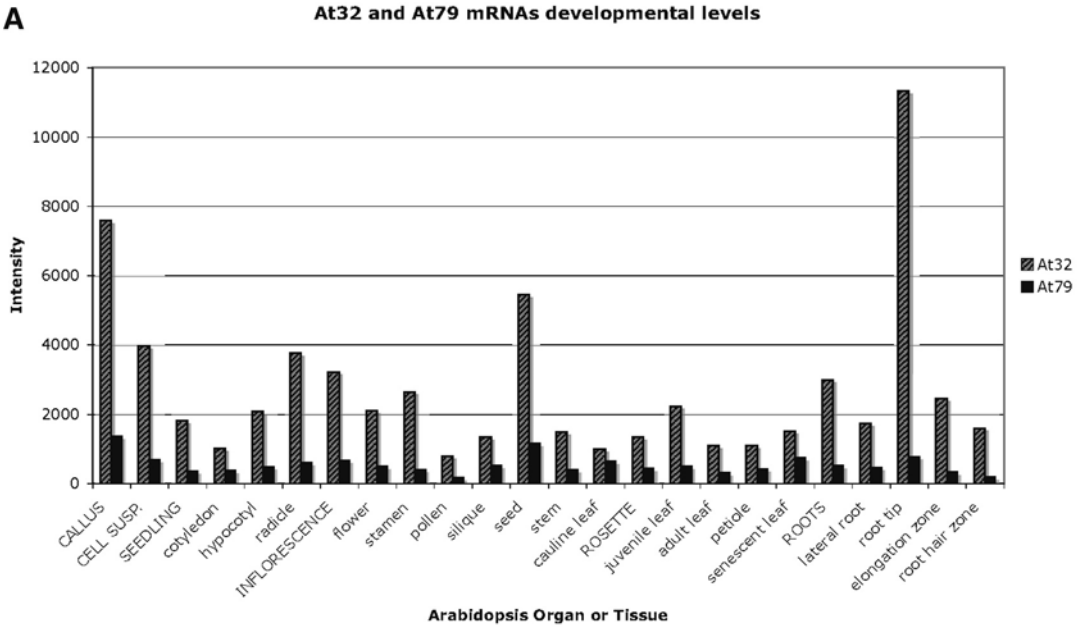
Supplementary Figure 2. Expression profiles of At32 (AtLa1) and At79 in *Arabidopsis.*

(A) Microarray expression profiles of At32 and At79 mRNAs obtained from the Genevestigator web site (https://www.genevestigator.ethz.ch/at/).

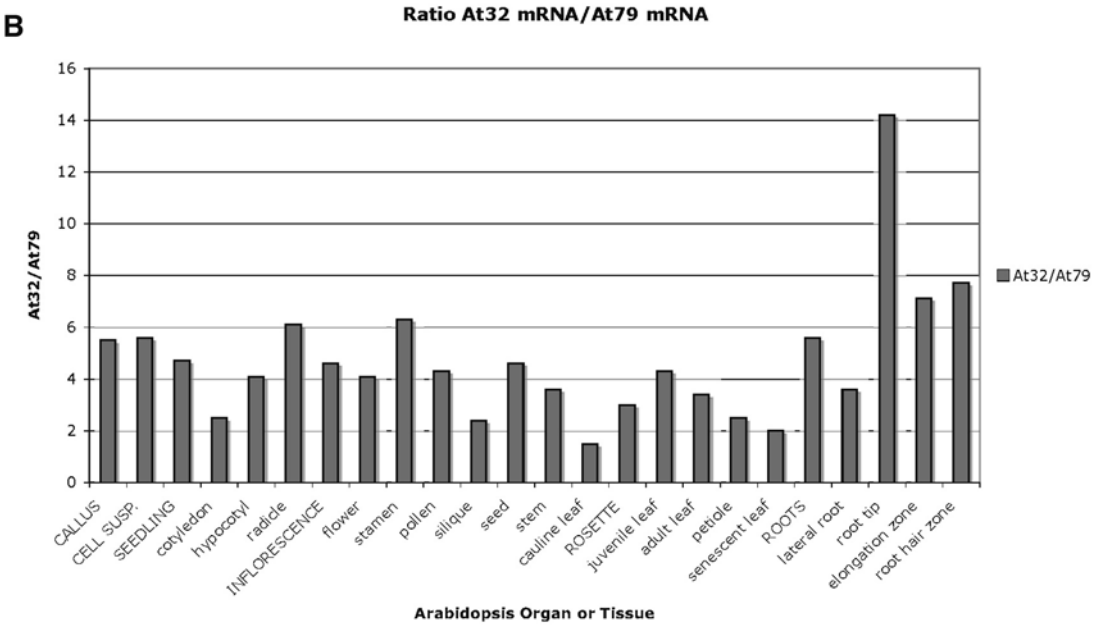
(B) Comparison of At32 and At79 mRNA levels.

(C) Antibodies against specific synthetic peptides were produced in rabbit and affinity purified against the antigen for each protein. These peptides are located outside the conserved regions of At32 (AtLa1) and At79. We performed a western blot analysis on total protein extracts prepared from the CY2 yeast strain expressing At32 (lane 1) or At79 (lane 2), from *Arabidopsis* cell suspensions (lane 3), 7 days old seedling (lane 4), adult leaves (lane 5), flowers (lane 6) and roots (lane 7). Total proteic extracts prepared from *Arabidopsis* material were quantified and comparable amounts loaded. The At32 (AtLa1) antibodies specifically recognize a protein with an apparent molecular weight of 56 kDa in At32 but not in At79 expressing yeast extracts (lanes 1 and 2). This signal is also detected in *Arabidopsis* cellular and plant extracts. We conclude it corresponds to the At32 (AtLa1) protein which migrates slower than its predicted molecular weight of 48 kDa. Such aberrant migration profile was previously observed for both yeast and fly genuine La proteins (3). The At79 antibodies specifically recognize a 44 kDa protein in At79 but not in At32 expressing yeast extracts (lanes 1 and 2). This band, which displays the predicted molecular weight of the At79 protein, is also detected in *Arabidopsis* cellular and plant extracts. We conclude that it corresponds to At79. At79 antibodies also strongly react with a factor of 60 kDa apparent molecular weight of unknown origin. This bands most probably corresponds to an unspecific signal since it is not detected in yeast extracts (lane 2) and is not depleted in T-DNA insertion mutant plants deleted of At79 expression whereas the specific 44 kDa band is undetectable (not shown).

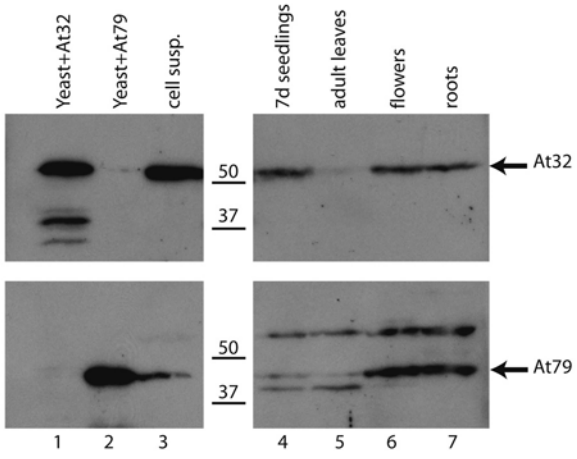
A


At32 and At79 mRNAs developmental levels
Intensity
12000
10000
8000
6000
4000
2000
0
At32
At79
CALLUS
CELL SUSP.
SEEDLING
cotyledon
hypocotyl
radicle
INFLORESCENCE
flower
stamen
pollen
silique
seed
stem
cauline leaf
ROSETTE
juvenile leaf
adult leaf
petiole
senescent leaf
ROOTS
lateral root
root tip
elongation zone
root hair zone
Arabidopsis Organ or Tissue

B

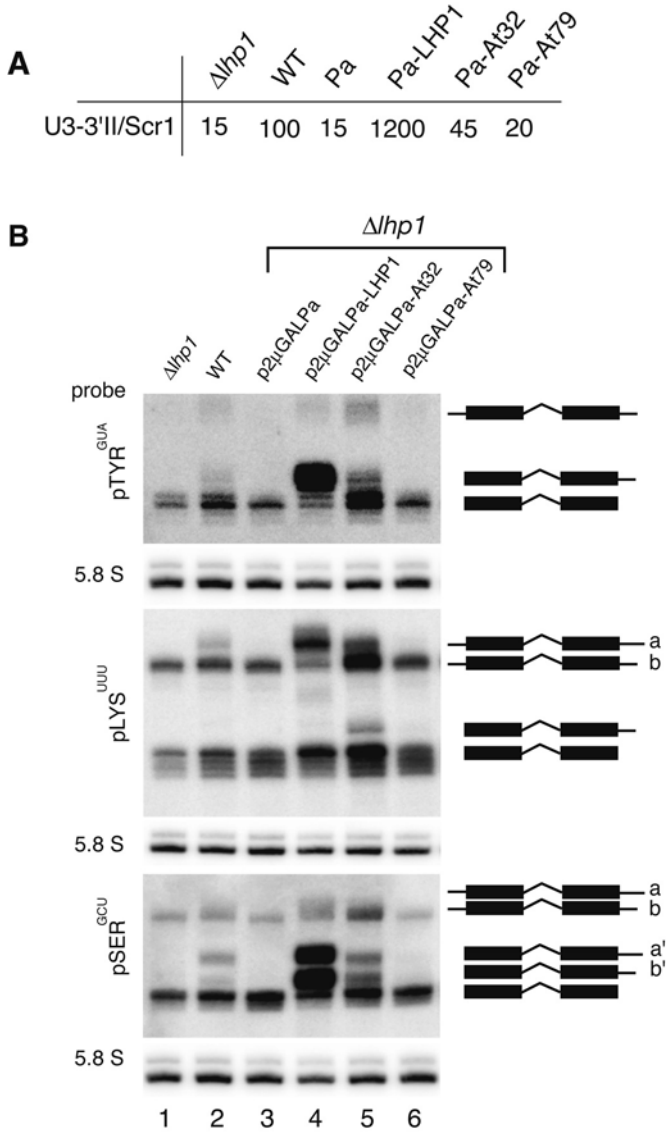

Ratio At32 mRNA/At79 mRNA
At32/At79
16
14
12
10
8
6
4
2
0
At32/At79
CALLUS
CELL SUSP.
SEEDLING
cotyledon
hypocotyl
radicle
INFLORESCENCE
flower
stamen
pollen
silique
seed
stem
cauline leaf
ROSETTE
juvenile leaf
adult leaf
petiole
senescent leaf
ROOTS
lateral root
root tip
elongation zone
root hair zone
Arabidopsis Organ or Tissue

C


Yeast+At32
Yeast+At79
cell susp.
7d seedlings
adult leaves
flowers
roots
50
37
At32
50
37
At79
1
2
3
4
5
6
7

Supplementary Figure 3. Complementation of *lhp1-Δ* molecular phenotypes with At32 or At79 proteins and Lhp1p over-expression phenotypes in yeast.

(A) Phosphorimager quantification of the U3-3'II signal normalized to the SCR1 signal. The values are expressed relatively to the WT which value is arbitrarily set to 100. The reported numbers correspond to the average of values obtained from three independent experiments. **(B)** Northern blot analysis of tyrosine^{GUA}, lysine^{UUU} and serine^{GCU} tRNA precursors patterns. Total RNAs were extracted from the untransformed CY2 (*lhp1-Δ*) strain (lane 1), the isogenic WT CY3 strain (lane 2), or the CY2 strain expressing ProtA (lane 3), ProtA-Lhp1p (lane 4), ProtA-At32 (lane 5) or ProtA-At79 (lane 6) and analysed by northern blot. Blots were probed with ^{32}P-labelled oligonucleotides specific to intronic sequences of pre-tRNA^{TyrGUA} (pTYR^{GUA}), pre-tRNA^{LysUUU} (pLYS^{UUU}) or pre-tRNA^{SerGCU} (pSER^{GCU}). Each blot was also probed for 5.8S rRNA accumulation as loading control. Schematic representations of tRNA precursors, intermediate and mature species are shown on the right. As previously published (4), upon Lhp1p depletion, the unspliced 5' and 3' extended precursors, migrate faster and underaccumulate and the processing intermediates corresponding to unspliced 5'-processed, 3'-extended pre-tRNAs are undetectable. Consistently to literature the 5'-mature, 3'-extended intron-containing pre-tRNA^{LysUUU} is undetectable in WT conditions (lane 2) and the impact of Lhp1p inactivation is monitored by the loss of the longer unspliced 5'-, 3'-extended pre-tRNA^{LysUUU} (labelled a on the schematic representation). For serine^{GCU}, consistently with previous report (5) which identified two unspliced 5'-, 3'-extended transcripts (named a and b on the schematic representation of precursors on the right hand-side) we detect in addition to the already described 3' extended species (labelled a'), an additional band (labelled b') closely migrating above the unspliced, end matured intermediate. Given its size and the fact that together with the a' form, it is completely lost upon inactivation of *LHP1*, we postulate that it is also an unspliced 5'-mature 3'-unprocessed tRNA^{SerGCU} intermediate. In extracts prepared from strain over-expressing Lhp1p (lane 4), we can observe a striking over-accumulation of precursor forms that are depleted upon *LHP1* inactivation. This is also the case for strain expressing At32, but to a lower extent (lane 5). The expression of At79 has no impact on tRNA maturation (lane 6).



Footnotes on Supplementary Figure 3. Whether on U3 or tRNA precursors, Lhp1p over-expression has a strong previously unpublished impact on their steady-state levels (see also Figure 3 in the manuscript). U3 3'-extension is trimmed to mature 3'-end by the exosome (6) which resumes processing only after the removal of Lhp1p from the U3 3'-trailer. Lhp1p displacement was suggested to be consecutive to the binding of the C/D snoRNP core proteins to U3 and proposed to be steric (7). We postulate that the strong accumulation of U3-3'II (and likely of U3-3'I as well) reflects a delay in its final trimming, resulting from the difficult displacement of Lhp1p in over-expressing strains. For tRNAs, La binds pre-tRNAs 3'-trailer, protects it from exonucleolytic nibbling and allows its endonucleolytic removal (8). Whether Lhp1p needs to be displaced from the 3'-trailer before this endonuclease can cut is unknown. As for U3, we propose that the striking over-accumulation of pre-tRNA 5'-processed 3'-unprocessed intermediates in presence of high levels of Lhp1p is due to a delay or even a

partial block of the 3'-extension excision process. This implies that the RNase might not be able to proceed while Lhp1p is bound to the 3' trailer and that Lhp1p needs to be displaced enzymatically or sterically for proper tRNA 3'-processing. In addition to these possible impacts on U3 and tRNAs maturation processes, a general stabilization of otherwise degraded primary transcripts and/or 3'-extended intermediates is expected at high Lhp1p concentration.

REFERENCES on SUPPLEMENTARY DATA

1. Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P.D., Curry, S. and Conte, M.R. (2003) Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure*, **11**, 833-843.
2. Topfer, F., Gordon, T. and McCluskey, J. (1993) Characterization of the mouse autoantigen La (SS-B). Identification of conserved RNA-binding motifs, a putative ATP binding site and reactivity of recombinant protein with poly(U) and human autoantibodies. *J. Immunol.*, **150**.
3. Yoo, C.J. and Wolin, S.L. (1994) La proteins from Drosophila melanogaster and Saccharomyces cerevisiae: a yeast homolog of the La autoantigen is dispensable for growth. *Mol. Cell. Biol.*, **14**, 5412-5424.
4. Yoo, C.J. and Wolin, S.L. (1997) The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell*, **89**, 393-402.
5. O'Connor, J.P. and Peebles, C.L. (1991) In vivo pre-tRNA processing in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **11**, 425-439.
6. Raijmakers, R., Schilders, G. and Pruijn, G.J. (2004) The exosome, a molecular machine for controlled RNA degradation in both nucleus and cytoplasm. *Eur. J. Cell Biol.*, **83**, 175-183.
7. Kufel, J., Allmang, C., Chanfreau, G., Petfalski, E., Lafontaine, D. and Tollervey, D. (2000) Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Mol. Cell. Biol.*, **20**, 5415-5424.
8. Wolin, S.L. and Matera, A.G. (1999) The trials and travels of tRNA. *Genes Dev.*, **13**, 1-10.

I.2. Données supplémentaires

I.2.1) Complémentation des lignées mutantes pour le gène AtLa1 *par la fusion GFP-AtLa1*

Nous avons montré précédemment que la protéine de fusion GFP-AtLa1 était capable de lier des précurseurs de transcrits Pol III en culture de cellules (figure 4 de l'article 1). Afin de construire un outil permettant l'étude des ARN et protéines interagissant avec la protéine AtLa1, sans la concurrence de la protéine endogène, nous avons entrepris de complémenter les lignées mutantes pour AtLa1 avec cette fusion.

Des plantes hétérozygotes des deux lignées mutantes pour le gène *AtLa1* (SAIL 548H11 = *atla1-1* et GABI 870F12 = *atla1-2*) ont été transformées, via *Agrobacterium tumefaciens*, avec un plasmide permettant l'expression d'une protéine de fusion d'AtLa1 avec la GFP (Green Fluorescent Protein), le transgène *GFP-AtLa1* étant placé sous le contrôle du promoteur 35S du Virus de la Mosaïque du Chou-fleur (CaMV 35S). Pour chacune des lignées d'insertions *atla1-1* et *atla1-2*, nous avons sélectionné quarante transformants primaires présentant à la fois le transgène *GFP-AtLa1* et l'ADN-T responsable de la mutation du gène *AtLa1*. Nous les avons ensuite génotypés pour connaître l'état hétérozygote ou homozygote de l'ADN-T, ce dernier étant normalement absent de la population puisque létal. Si le transgène *GFP-AtLa1* complémente efficacement les mutations *atla1-1* et/ou *atla1-2*, nous devrions être en mesure d'identifier des plantes homozygotes pour l'ADN-T responsable de la mutation. Nous n'avons cependant pas obtenu de plante homozygote pour les mutations *atla1-1* ou *atla1-2*, que ce soit parmi les transformants primaires ou leur descendance (pour chacune des lignées mutantes, un total de 80 plantes issues de 2 transformants primaires hétérozygotes ont été analysées). Bien que le transgène *GFP-AtLa1* soit correctement exprimés dans les tissus somatiques des transformants primaires et de leur descendance (analyses par western blot, données non présentées), il ne semble pourtant pas capable de complémenter le phénotype de létalité embryonnaire (figure 7 de l'article 1) observé chez les plantes mutantes *atla1-1* et *atla1-2*.

Nous ne pouvons pas exclure la possibilité que la fusion de la GFP en amont de la protéine AtLa1 affecte partiellement la fonctionnalité de cette dernière. Nous avons donc également tenté de complémenter le phénotype de létalité embryonnaire des lignées mutantes *atla1-1* et *atla1-2* en exprimant l'ADNc de la protéine AtLa1 sans tag sous le contrôle du promoteur CaMV 35S. De la même façon que pour le transgène *GFP-AtLa1*, nous avons d'abord sélectionné des transformants primaires présentant à la fois l'ADNc de la protéine

AtLa1 et l'ADN-T responsable de la mutation du gène *AtLa1*. Nous avons ensuite déterminé le génotype des transformants primaires et de leur descendance pour identifier des plantes homozygotes pour les mutations *atla1-1* et *atla1-2*. Mais là encore, nous n'avons pu obtenir de plantes homozygotes *atla1-1/atla1-1* ou *atla1-2 /atla1-2*. L'expression d'AtLa1 sous contrôle du promoteur CaMV35S ne rétablit donc pas la viabilité des plantes inactivées pour la fonction AtLa1.

Dans les deux constructions utilisées pour complémenter les mutations *atla1-1* et *atla1-2*, les séquences codantes des transgènes *AtLa1* et *GFP-AtLa1* ont été placées sous contrôle du promoteur CaMV 35S. Nous avons recherché dans la littérature d'autres exemples de mutations entraînant un phénotype de létalité embryonnaire, que les auteurs auraient tenté de complémenter en exprimant le cDNA du gène affecté sous contrôle du promoteur CaMV 35S. C'est le cas de mutations affectant les gènes *AtCUL1* et *EDD1* (Uwer *et al.* 1998; Shen *et al.* 2002). Comme dans notre cas, les auteurs n'ont pu complémenter le phénotype de létalité embryonnaire en utilisant le promoteur CaMV 35S, mais ils ont pu le complémenter en utilisant un autre promoteur : son propre promoteur dans le cas du gène *AtCUL1* et le promoteur USP (pour Unknown Seed Protein) dans le cas du gène *EDD1*. Le promoteur CaMV 35S est un promoteur fort et il est généralement considéré comme ubiquitaire. Il existe cependant des données indiquant que le promoteur CaMV 35S n'est pas actif pendant l'embryogenèse précoce (Custers *et al.* 1999), ce qui pourrait expliquer l'échec de la complémentation, étant donné le stade d'arrêt du développement des graines mutantes pour le gène *AtLa1* (figure 7 de l'article 1).

Afin de compléter cette partie de l'étude et d'obtenir un outil pour étudier les ARN et les protéines interagissant avec AtLa1, nous projetons de construire des lignées de plante exprimant la protéine de fusion GFP-AtLa1 sous le contrôle du promoteur endogène du gène *AtLa1*.

I.2.2) AtLa1 est exclue des corps de Cajal

Les corps de Cajal sont des structures nucléaires que l'on retrouve chez les animaux et les plantes, et des structures correspondantes ont été identifiées chez la levure (Cioce and Lamond 2005). Les corps de Cajal sont plus abondants dans les cellules où la transcription est active et notamment dans les cellules en prolifération. Ils sont plus petits que le nucléole, partagent beaucoup de ses composants et sont souvent observés à proximité de celui-ci. Ils

sont impliqués dans la biogenèse et le trafic des particules ribonucléoprotéiques des petits ARN nucléaires (snRNP) et des particules ribonucléoprotéiques des petits ARN nucléolaires (snoRNP) mais leurs rôles n'ont pas encore été entièrement élucidés (Stanek and Neugebauer 2006). Une étape importante, mais pas entièrement comprise, de la biogenèse des petits ARN nucléaires U1, U2, U4, U5, U6 est la modification post-transcriptionnelle de leurs nucléotides. Les snARN U1, U2, U4, U5 sont transcrits par la Pol II et sont exportés dans le cytoplasme, où ils vont s'associer aux protéines Sm et subir la maturation de leur coiffe 5'. Ils vont ensuite revenir dans le noyau, puis dans les corps de Cajal où ils vont subir des modifications post-transcriptionnelles (Matera and Shpargel 2006). Il a récemment été montré que les corps de Cajal contiennent des petits ARN spécifiques, qui ressemblent aux petits ARN nucléolaires responsables des modifications post-transcriptionnelles des ARNr. Les petits ARN des corps de Cajal (scaARN) remplissent une fonction similaire, puisqu'ils guident les modifications des nucléotides des petits ARN nucléaires (Darzacq *et al.* 2002 ; Jady *et al.* 2003).

Au contraire des autres petits ARN, le snARN U6 (transcrit par la Pol III) ne semble pas sortir du noyau et subit des modifications post-transcriptionnelles lors de son passage dans le nucléole (Stanek and Neugebauer 2006). U6 repasse ensuite dans le nucléoplasme, où il s'associe aux protéines Lsm. Finalement, il va lui aussi dans les corps de Cajal, où il s'assemble aux snARN modifiés U4 et U5, pour former le complexe tri-snRNP : U4/U5-U6 impliqué dans l'épissage (Stanek and Neugebauer 2006). Certains petits ARN nucléolaires et notamment le snoARN U3, passent aussi à travers les corps de Cajal. En effet, il est proposé que chez les vertébrés, l'assemblage de la particule snoRNP, la maturation 3' et l'hyperméthylation de la coiffe d'U3 aient lieu dans les corps de Cajal (Verheggen *et al.* 2002). La localisation subcellulaire de ces événements semble variable d'un organisme à l'autre, puisque chez la levure *S. cerevisiae*, ils surviennent dans le nucléole (Verheggen *et al.* 2002). Chez les plantes, des snARN et des précurseurs de snoARN ont été détectés dans les corps de Cajal, suggérant qu'une partie de la maturation de ces petits ARN s'y déroulerait (Shaw *et al.* 1998 ; Docquier *et al.* 2004 ; Shaw and Brown 2004).

Récemment, il a été proposé que les siARN de la voie de méthylation de l'ADN dépendante de l'ARN (cf. paragraphe II.3.1 de la synthèse bibliographique) seraient produits et pris en charge par les acteurs de la voie dans des corps nucléaires situés dans ou en périphérie du nucléole (Li *et al.* 2006; Pontes *et al.* 2006). En effet, des siARN de la voie de méthylation de l'ADN dépendante de l'ARN s'accumulent dans ces corps nucléaires et colocalisent avec les protéines RDR2 et DCL3 impliquées dans leur maturation, ainsi qu'avec

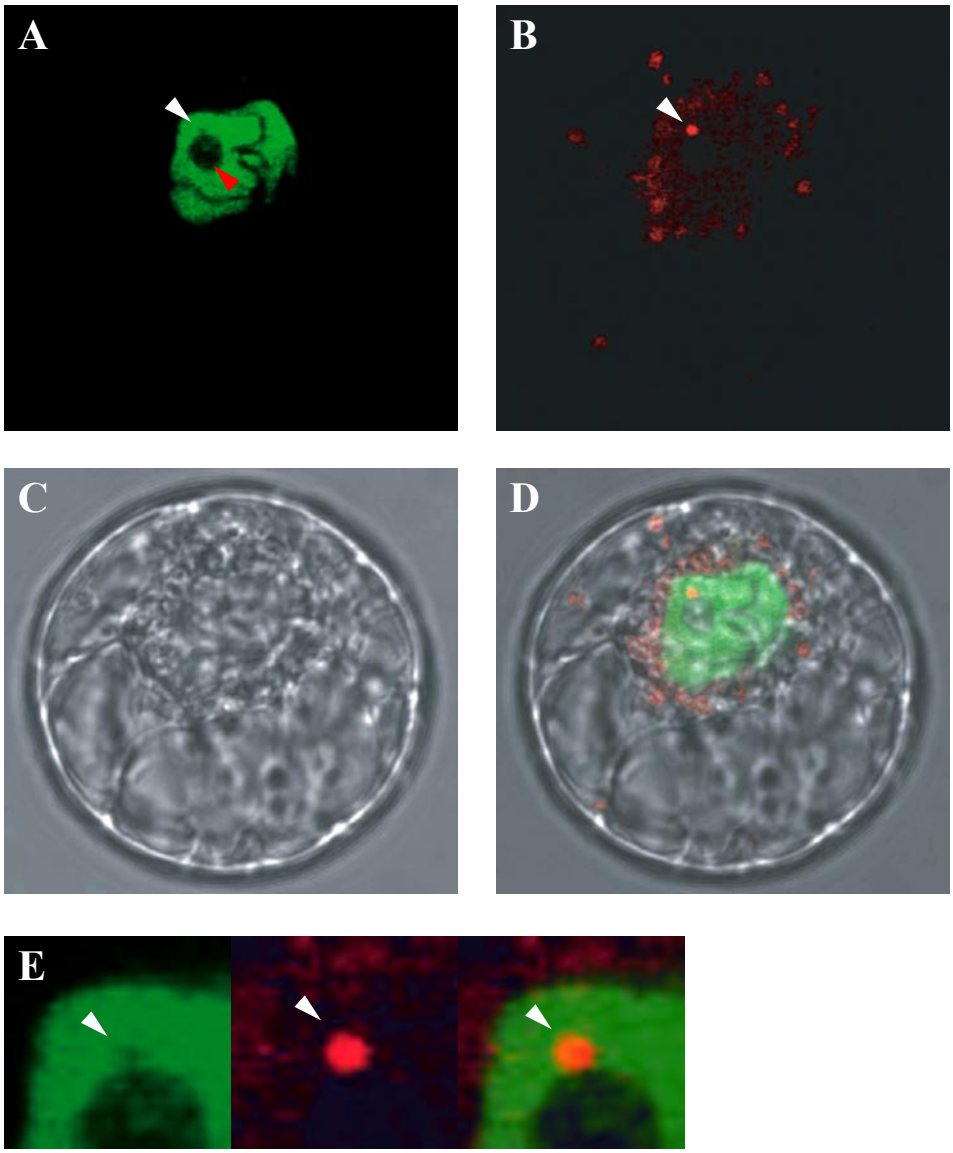


Figure 13. La protéine AtLa1 est exclue des corps de Cajal

Des protoplastes de cellules d'*Arabidopsis* ont été co-transformés avec des constructions permettant l'expression transitoire des protéines de fusion AtLa1-GFP et d'U2B''-RFP. Ces protoplastes ont ensuite été observés à l'aide d'un microscope confocal.

Les photos A et B présentent respectivement les localisations subcellulaires des protéines AtLa1-GFP et U2B''-RFP. La photo C présente le même protoplaste observé en contraste interférentiel différentiel (DIC) et le panneau D, la superposition des trois photos précédentes. Le panneau E présente un agrandissement du corps de Cajal. Les flèches blanches et rouges indiquent respectivement un corps de Cajal et le nucléole. En B et D, les signaux rouges qui entourent le noyau correspondent à l'autofluorescence de plastes.

les protéines AGO4 et Pol IVb qui s'associent à ces siARN. Les données de Li et al. (2006) suggèrent que ces corps nucléaires seraient en fait les corps de Cajal.

Afin d'étudier plus finement la localisation de la protéine AtLa1 nous avons co-transformé de façon transitoire des protoplastes de cellules d'*Arabidopsis* avec des plasmides codant respectivement la protéine de fusion AtLa1-GFP et la protéine U2B" fusionnée avec la protéine fluorescente rouge (RFP pour Red Fluorescent Protein) (Lorkovic *et al.* 2004). La protéine U2B" appartient à la particule ribonucléoprotéique du petit ARN nucléaire U2, qui en s'assemblant avec les particules ribonucléoprotéiques U1, U4, U5 et U6 forment le spliceosome (Valadkhan 2005). La protéine U2B" est localisée dans le nucléoplasme et dans la cavité nucléolaire (présentée en conclusion de cette partie, paragraphe I.3 des résultats), elle semble en revanche absente du reste du nucléole. Dans le nucléoplasme, elle apparaît sous forme de paquets, ce qui est en accord avec la localisation des autres facteurs d'épissage. De plus, la protéine U2B" apparaît fortement concentrée dans les corps de Cajal ; elle est donc fréquemment utilisée comme un marqueur de ces corpuscules nucléaires (Lorkovic *et al.* 2004).

Sur les images de microscopie confocale présentées en figure 13, on voit que les protéines AtLa1-GFP et U2B"-RFP sont toutes les deux présentes dans le nucléoplasme. Par comparaison avec U2B", la localisation de la protéine AtLa1 apparaît plus uniforme et elle semble également plus abondante. En revanche la protéine AtLa1 semble exclue des corps de Cajal, où l'on voit une forte accumulation de la protéine U2B". Au cours de cette expérience, nous n'avons observé la présence de la protéine AtLa1 dans aucun des corps de Cajal marqués par U2B".

I.3. Conclusion

Nos travaux sur la protéine La d'*Arabidopsis* ont permis l'identification de deux homologues potentiels : les protéines At32 et At79, respectivement codées aux loci At4g32720 et At1g79880.

Des expériences de complémentation de levures *S. cerevisiae* déficientes pour la protéine La, ainsi que des études de l'association de ces protéines aux transcrits Pol III en système cellulaire d'*Arabidopsis*, ont démontré que la fonction nucléaire de la protéine La dans la maturation des transcrits Pol III existe chez *Arabidopsis* et vraisemblablement chez les

plantes supérieures. Cette fonction est remplie par la protéine At32, que nous avons donc renommée AtLa1.

Les connaissances concernant la maturation des ARNt chez les plantes viennent de systèmes d'étude *in vitro*. En effet, il semble que le très grand nombre de gènes codant des ARNt (630 chez *Arabidopsis*, contre 273 chez *S. cerevisiae*), ainsi que la rapidité des événements de maturation ne permettent pas de suivre la maturation des ARNt *in vivo* par la technique d'hybridation northern, comme cela est possible chez *S. cerevisiae*.

La maturation des précurseurs des ARNt, observée *in vitro* dans des extraits cellulaires, semble suivre les mêmes étapes que chez *S. cerevisiae*, commençant par le clivage de l'extrémité 5' par une activité endonucléase de type RNase P (Franklin *et al.* 1995 ; Yukawa *et al.* 2001). Cette enzyme qui possède un composant ARN chez les levures et les animaux, semble être différente chez les plantes. En effet, jusqu'ici aucun ARN homologue à l'ARN de la RNase P n'a été identifié dans les génomes de plante. Si l'enzyme responsable de la maturation 5' des ARNt chez les plantes possède un composant ARN, celui-ci ne doit pas être conservé en séquence par rapport aux autres organismes (Franklin *et al.* 1995 ; Piccinelli *et al.* 2005). Suite à la maturation de l'extrémité 5', l'extrémité 3' du précurseur est ensuite clivée par une activité endonucléase (Yukawa *et al.* 2001), récemment attribuée à la RNase Z (cf. paragraphe III.3.1.a de la synthèse bibliographique) (Mayer *et al.* 2000). Lorsque le précurseur présente un intron, les étapes de maturation de ses extrémités précèdent l'étape d'épissage. Des tests *in vitro* de maturation de précurseurs d'ARNt ont montré que les ARNt de plantes pouvaient subir une maturation correcte en extraits de cellules HeLa (van Tol *et al.* 1987). De même les précurseurs d'ARNt du xénope, de l'homme et de la levure *S. cerevisiae* subissent une maturation correcte de leurs extrémités 3' et 5' en extraits de germe de blé (Stange *et al.* 1988).

Nous avons montré que la protéine AtLa1 fixait les précurseurs des ARNt chez *Arabidopsis* (figure 4 de l'article 1) et restaurait les défauts de maturation des ARNt dans une souche de levure *S. cerevisiae* mutante pour le gène *LHP1* (figure 3.B de l'article 1). Ces résultats s'ajoutent aux éléments présentés précédemment et pointent vers une conservation des processus de maturation des extrémités 3' et 5' des précurseurs des ARNt chez les plantes, ainsi que du rôle rempli par la protéine La au sein de ces processus.

À l'aide de protéines de fusion avec la GFP, nous avons montré que la protéine AtLa1 est localisée majoritairement dans le nucléoplasme, mais elle ne semble pas s'accumuler dans

le nucléole (figure 5 de l'article 1). Ce résultat est assez surprenant, puisque des études similaires ont montré que chez la levure *S. cerevisiae* et l'homme, la protéine La a une localisation nucléolaire (Graus *et al.* 1985 ; Long *et al.* 2001). Cette localisation est associée aux étapes précoces de maturation des précurseurs des ARNt, notamment la coupure de l'extrémité 5' par la RNase P qui s'effectuent dans le nucléole (Bertrand *et al.* 1998 ; Horke *et al.* 2004). Sur nos images, il est difficile de conclure sur la présence d'une faible quantité ou l'exclusion complète de la protéine AtLa1 du nucléole. Ces données laissent penser que les étapes précoces de maturation des pré-ARNt ne s'y effectueraient pas. Il est également envisageable que le passage de la protéine La dans le nucléole soit trop transitoire pour entraîner une accumulation.

Nos observations de la localisation subcellulaire de la protéine AtLa1 ont également mis en évidence une présence, au moins transitoire, dans la cavité nucléolaire (figure 5 de l'article 1). La cavité nucléolaire est un compartiment nucléolaire très intrigant, uniquement retrouvé chez les plantes (Shaw and Brown 2004). Ses fonctions biologiques restent à élucider, mais il a été proposé qu'elle servirait d'espace de stockage ou de recyclage de certains composants cellulaires. Il a également été observé que la cavité nucléolaire pouvait déverser son contenu dans le nucléoplasme. La formation de la cavité nucléolaire apparaît liée au cycle cellulaire, ainsi il semble qu'il existe deux types de cavité nucléolaire : l'une visible en G1 et l'autre en G2 (Gonzalez-Camacho and Medina 2006). La cavité visible en phase G2 n'est pas systématiquement présente dans le nucléole. De plus, leur observation par microscopie électronique a mis en évidence des différences morphologiques, en particulier une taille plus importante et une forme plus régulière de la cavité visible en G2. La morphologie et la taille des cavités nucléolaires dans lesquelles nous avons observé la protéine AtLa1 nous laissent penser qu'il s'agit de cavités visibles en G2. Ceci pourrait expliquer que nous ne les ayons observées que dans un faible nombre de cellule.

Nous avons également montré que l'absence de la protéine AtLa1 était létale à l'état embryonnaire (figures 6 et 7 de l'article 1). Ce résultat est concordant avec ceux obtenus chez d'autres organismes pluricellulaires : la souris et la drosophile, où l'inactivation du gène codant la protéine La entraîne un arrêt du développement au stade embryonnaire. Au contraire, les levures *S.cerevisiae* et *S. pombe* mutantes pour le gène codant la protéine La, sont viables en conditions normales de croissance. Comparées aux protéines des levures, les protéines La des eucaryotes supérieurs présentent une région carboxy-terminale étendue, contenant notamment un second RRM atypique. Cette partie de la protéine pourrait être

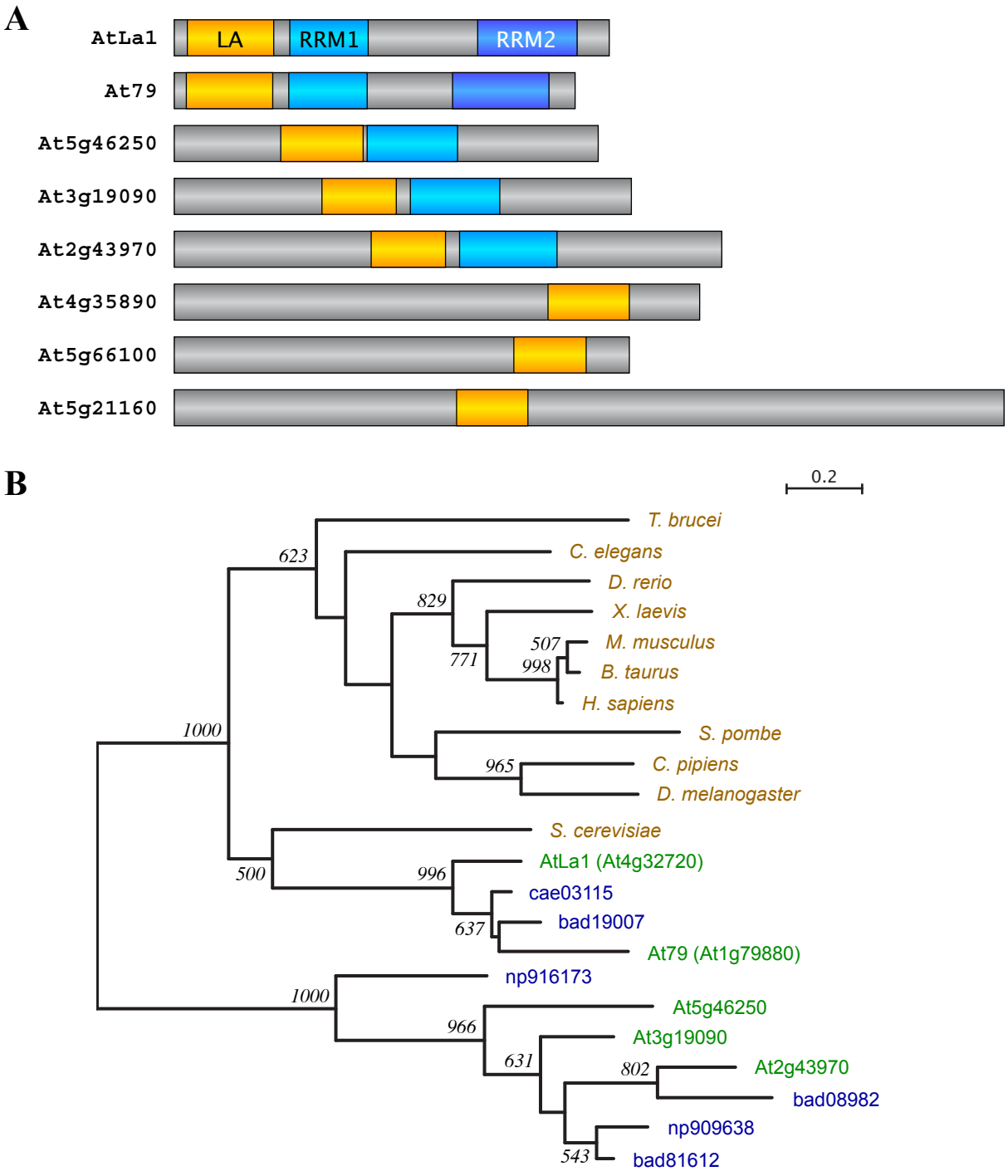


Figure 14. AtLa1 et At79 sont les seules protéines d'*Arabidopis* à contenir un domaine La associé à un domaine RRM, dont les caractéristiques correspondent aux autres protéines La

(A) Représentation schématique de la structure des huit protéines d'*Arabidopsis* dont la séquence contient un domaine La. Les domaines La sont représentés par une boîte orange et les domaines RRM par des boîtes bleues.

(B) Arbre phylogénétique représentant les relations existant entre les domaines RRM1 de type canonique des protéines La et des protéines d'*Arabidopsis* et de riz contenant à la fois un domaine La et un domaine RRM. Les nombres à côté des embranchements indiquent les valeurs bootstrap obtenues pour 1000 répétition (seules les valeurs supérieures à 500 sont présentées). Les protéines La sont indiquées en marron, les protéines d'*Arabidopsis* en vert et celles de riz en bleu.

associée à certaines fonctions absentes chez la levure, et responsable du phénotype de létalité. Ces fonctions restent cependant à déterminer. De plus, le phénotype de létalité embryonnaire a empêché de déterminer si la protéine La était nécessaire au développement somatique des eucaryotes supérieurs. Afin de répondre à cette question chez les plantes, nous proposons d'établir un système d'ARNi inductible, qui permettrait d'éteindre le gène *ATLA1* après la formation de la graine.

II. « Les fonctions La » sont-elles remplies par deux protéines chez *Arabidopsis thaliana* ?

Les données présentées dans l'article 1 démontrent que la protéine AtLa1(/At 32) est une authentique protéine La, au moins capable d'endosser chez la plante les fonctions nucléaires des protéines La dans la maturation des transcrits Pol III. En revanche la situation concernant la protéine At79 est beaucoup plus intrigante. Nous présenterons dans ce chapitre des données qui soutiennent l'hypothèse que At79 est bien une authentique protéine La qui pourrait être spécialisée dans certaines fonctions La.

II.1. La protéine At79 présente les caractéristiques d'une authentique protéine La : données structurales et phylogénétiques

Comme nous l'avons vu dans l'article 1, le génome d'*Arabidospsis* contient huit gènes codant des protéines présentant un domaine La. Les protéines La sont entre autres caractérisées par la présence d'un tandem La-RRM. Parmi ces huit protéines, cinq présentent un domaine RRM associé au domaine La (figure 14.A). Sur les arbres phylogénétiques construits à partir des séquences protéiques du domaine La (figure 1 de l'article 1) et du domaine RRM de type canonique (RRM1) (figure 14.B), on constate que les protéines AtLa1 et At79 appartiennent à la même branche que les séquences de deux protéines du riz cae03115 et bad19607 (soutenue par une valeur bootstrap de 996/1000), ainsi qu'à celle de la protéine La de *S. cerevisiae*. En revanche, les trois autres protéines d'*Arabidopsis* présentant un domaine La et un domaine RRM (At5g46250, At3g19090 et At2g43970) forment une branche distincte avec d'autres protéines du riz. Si les domaines La de ces trois protéines sont relativement proches de ceux des vraies protéines La (figure 1 de l'article 1), leurs domaines RRM forment un groupe clairement distinct (valeur bootstrap de 1000) de celui des vraies

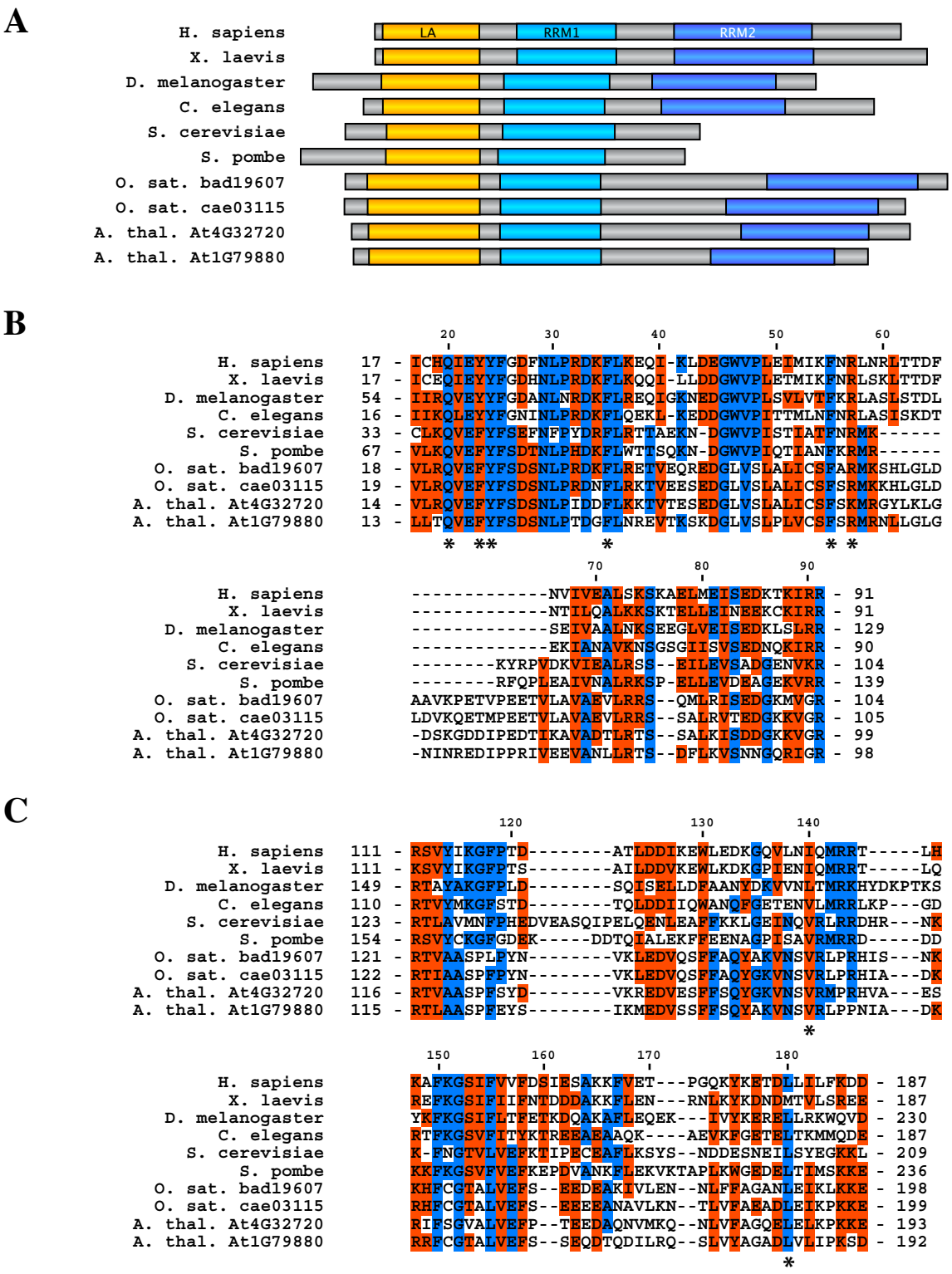


Figure 15. Les acides aminés impliqués dans la liaison à l'ARN des motifs La et RRM1 sont conservés chez les protéines AtLa1 et At79

(A) Représentation schématique de la structure des protéines La de de l'homme, du xénope, de la drosophile, de caenorhabditis, des levures S. cerevisiae et S. pombe, ainsi que des homologues potentiels du riz (O. sat.) et des protéines AtLa1 et At79 d'*Arabidopsis*. Les motifs La et RRM sont respectivement figurés par des boîtes orange et bleues.

(B-C) Alignement des séquences protéiques correspondant aux domaines La (B) et RRM1 (C). Les identités de séquences et les substitutions sont respectivement indiquées en bleu et en rouge. Les acides aminés impliqués dans la liaison à l'ARN sont indiqués par une étoile (*****).

protéines La (figure 14.B). De plus, dans tous les homologues de la protéine La identifiés à ce jour, les domaines La et RRM1 sont organisés en tandem avec le domaine La très proche de l'extrémité amino-terminale de la protéine (Maraia and Intine 2001). Les domaines La et RRM1 d'AtLa1 et At79 correspondent à cette organisation, au contraire de ceux des protéines At5g46250, At3g19090 et At2g43970 qui sont en position centrale (figure 14.A). De plus, à l'exception des protéines La des levures *S. cerevisiae* et *S. pombe*, les protéines La identifiées présentent toutes un second motif RRM, nommé RRM2 (article 1 et figure 15.A). Le motif RRM2 a initialement été mis en évidence dans la protéine La humaine. En utilisant le logiciel de prédiction de la structure locale des protéines SAM-T02, basé sur la méthode des modèles de chaînes de Markov cachées, nous avons recherché la présence de ce RRM2 dans l'ensemble des protéines contenant un domaine La. Cette analyse nous a permis de montrer que seules les authentiques protéines La présentent un tel domaine (figure supplémentaire 1 de l'article 1). Nous proposons donc que les protéines La des eucaryotes supérieurs sont caractérisées par la présence d'un tandem La-RRM1 et d'un RRM2. Chez *Arabidopsis*, les protéines AtLa1 et At79 sont les seules à contenir un RRM2 en plus du domaine La associé en tandem à un domaine RRM (figure 14.A).

La figure 15 (B et C) présente un alignement des domaines La et RRM1 des protéines La de l'homme, du xénope, de la drosophile, de *C. elegans*, des levures *S. pombe* et *S. cerevisiae*, ainsi que les potentiels homologues du riz bad19607 et cae03115 et d'*Arabidopsis* AtLa1 et At79. Nous avons reporté sur cette figure les acides aminés impliqués dans les liaisons à l'ARN, ils sont distingués par des étoiles. La majorité de ces acides aminés sont conservés chez l'ensemble de ces organismes. Lorsque la conservation n'est pas parfaite, l'acide aminé remplaçant présente les mêmes propriétés physico-chimiques et cette substitution est le plus souvent observée chez plusieurs organismes.

À la lumière de ces données phylogénétiques et structurales, les protéines AtLa1 et At79 ont toutes les caractéristiques des protéines La. Les résultats de notre étude de la protéine AtLa1, présentés dans le chapitre précédent, montrent que la protéine AtLa1 assure les fonctions de la protéine La dans les processus de maturation des transcrits présentant une extrémité 3' UUU-OH. De plus, nous avons montré qu'une mutation du gène *AtLa1* est létale à l'état homozygote, la protéine At79 n'est donc pas capable de compenser l'absence de la protéine AtLa1. Ce résultat indique que les protéines AtLa1 et At79 ne sont pas complètement redondantes, ce qui nous a amené à proposer que la protéine At79 se serait spécialisée pour

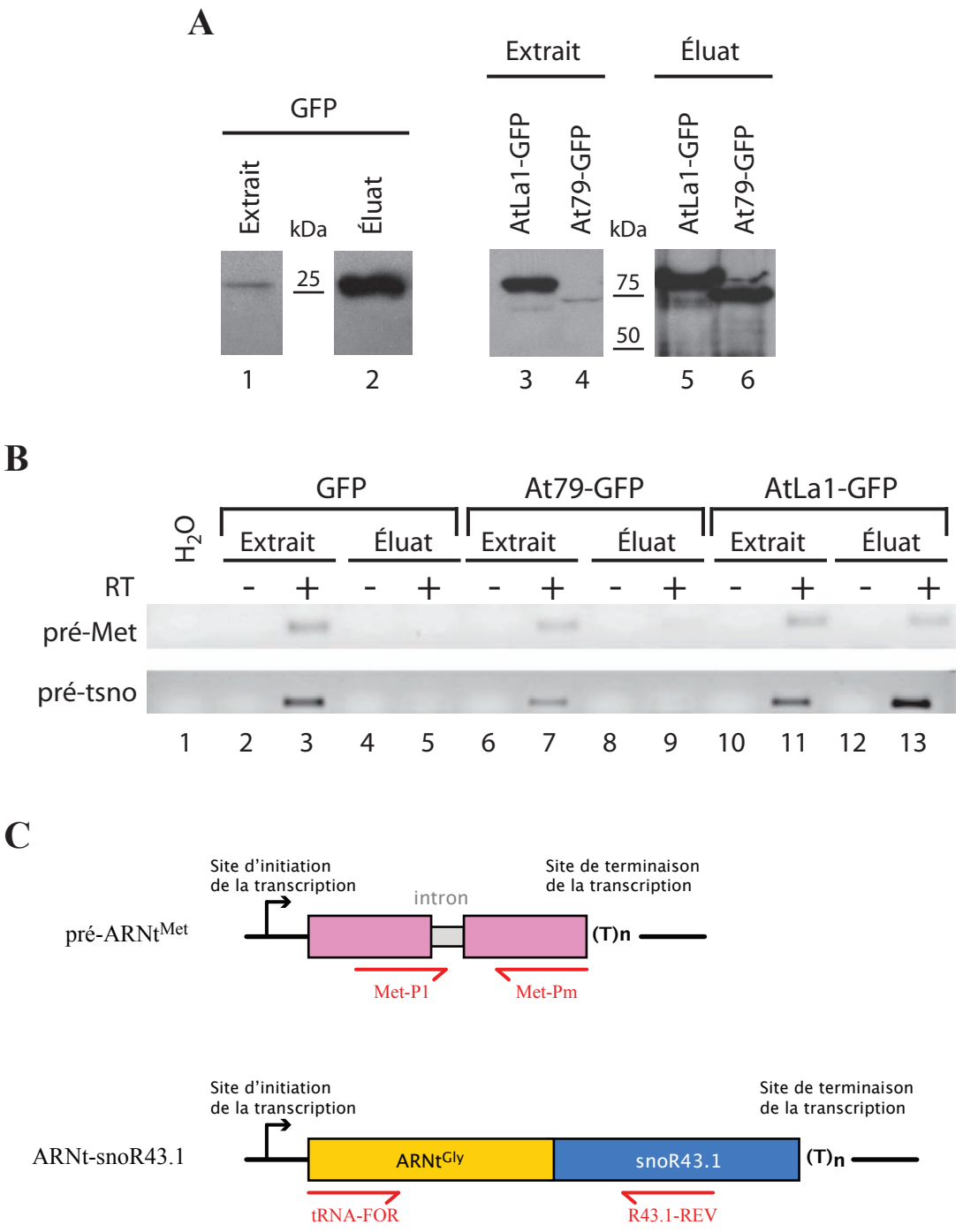


Figure 16. Expériences de co-immunoprécipitation de transcrits pol III à partir d'extraits natifs de cellules d'*Arabidopsis* exprimant la GFP, AtLa1-GFP ou At79-GFP

(A) Analyse par western blot des protéines présentes dans les extraits natifs et dans l'éluat des immunoprécipitations réalisées avec un anticorps polyclonal de lapin dirigé contre la GFP. Les western blot ont été hybridés avec un anticorps monoclonal de souris dirigé contre la GFP.

(B) Analyse par PCR des ADNc préparés à partir des extraits natifs et des éluats des immunoprécipitations. Les amplifications PCR ont été réalisées avec des paires d'oligonucléotides spécifiques de l'ADNc du précurseur de l'ARNt méthionine (pré-Met) et de l'ADNc du précurseur de l'ARNt-snoR43.1 (pré-tsno). Tous les échantillons ont été préalablement traités à la DNase, puis rétrotranscrit (+) ou non (-).

(C) Représentation schématique des ARN ciblés. Les amorces utilisées en (B) sont figurées par des flèches rouges.

assurer certaines des fonctions La. Afin de compléter notre étude des fonctions La chez *Arabidopsis* nous nous sommes donc intéressés aux rôles que pouvait remplir la protéine At79.

II.2. Étude fonctionnelle de la protéine At79

II.2.1) At79 lie-t-elle les transcrits Pol III ?

Au contraire de la protéine AtLa1, la protéine At79 s'est avérée incapable de complémenter les phénotypes des souches de *S. cerevisiae* mutantes pour la protéine Lhp1p (article 1). La complémentation des phénotypes choisis nécessite la fixation des extrémités 3' UUU-OH des ARN. L'échec de la complémentation par At79 peut s'expliquer entre autres par l'expression en système hétérologue, il est en effet possible que la protéine ne soit pas correctement produite (repliement anormal, absence de certaines modifications post-traductionnelles ?) ou localisée dans les levures. Nous avons donc testé la capacité de la protéine At79 à fixer les extrémités 3' UUU-OH de transcrits Pol III en cellules d'*Arabidopsis*.

Nous avons procédé de la même façon que lors de l'étude de la protéine AtLa1 (article1). Nous avons produit des lignées de cellules d'*Arabidopsis* exprimant la protéine At79 fusionnée à un tag GFP en position amino-terminale (GFP-At79) ou carboxy-terminale (At79-GFP). Nous avons ensuite préparé des extraits natifs à partir de ces lignées cellulaires et réalisé des immunoprécipitations avec un anticorps dirigé contre la GFP. Parallèlement, les mêmes manipulations ont été réalisées avec la lignée cellulaire n'exprimant que la GFP, utilisée comme témoin négatif de l'association aux transcrits Pol III, et avec les lignées exprimant les protéines de fusion AtLa1-GFP et GFP-AtLa1 utilisées comme témoin positif.

Les mêmes résultats ayant été obtenus avec les lignées exprimant les fusions amino-terminales (GFP-AtLa1 et GFP-At79) et celles exprimant les fusions carboxy-terminales (AtLa1-GFP et At79-GFP), seuls les résultats obtenus avec les fusions carboxy-terminales sont présentés figure 16.

Dans un premier temps, l'efficacité de l'immunoprécipitation a été vérifiée par western blot (figure 16.A). Pour chacune des trois lignées utilisées (GFP, AtLa1-GFP et At79-GFP) la même masse de cellules a été utilisée pour préparer un extrait total natif, et les mêmes volumes d'extraits ou d'éluats ont été utilisés pour l'analyse des protéines. Il apparaît que les protéines GFP (figure 16.A, piste 1) et At79-GFP (figure 16.A, piste 4) sont moins fortement exprimées dans nos lignées que ne l'est la protéine AtLa1-GFP (figure 16.A, piste 3).

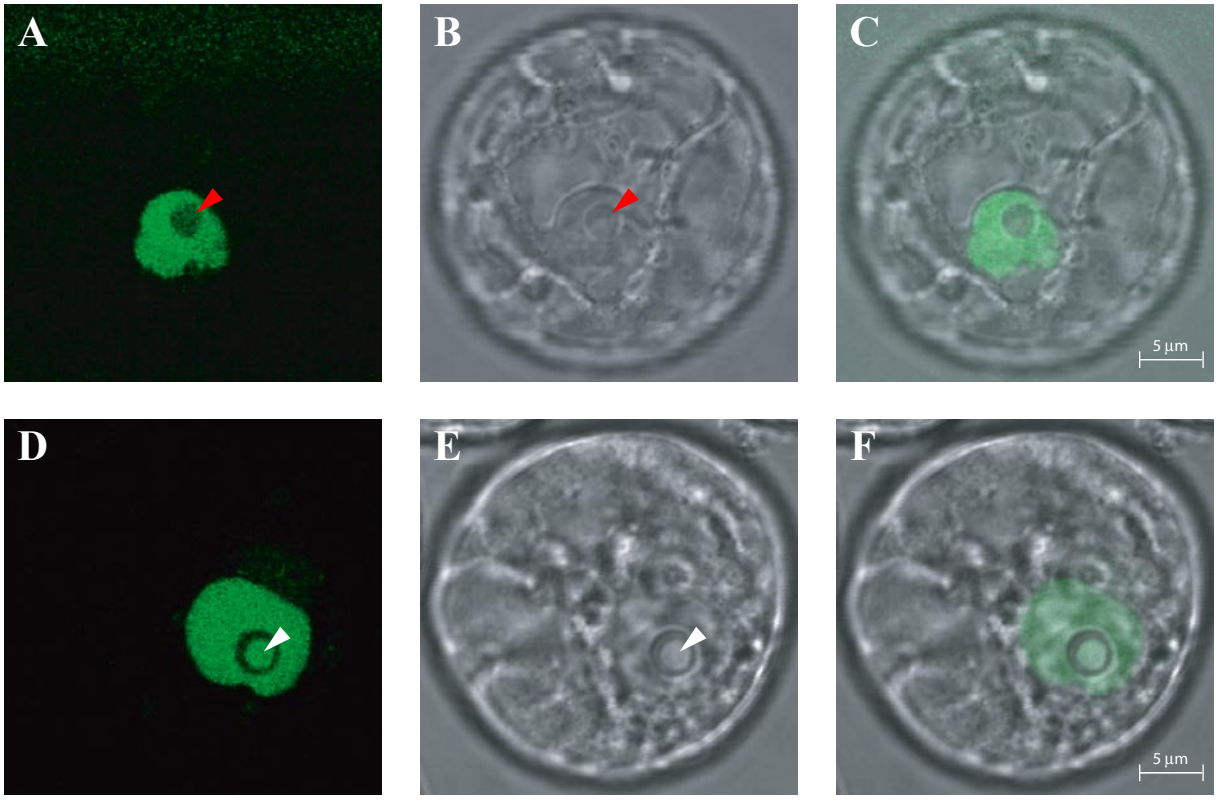


Figure 17. Localisation subcellulaire de la protéine de fusion At79-GFP

Des protoplastes ont été préparés à partir de cellules d'*Arabidopsis* exprimant de façon stable la protéine de fusion At79-GFP. Ces protoplastes ont ensuite été directement observés à l'aide d'un microscope confocal.

(A-C) Le plus souvent, la protéine At79-GFP apparaît diffuse dans le noyau, à l'exception du nucléole.

(D-F) Dans un nombre plus rare de cas, la protéine At79-GFP apparaît diffuse dans le noyau et fortement concentrée dans la cavité nucléolaire, mais toujours absente du reste du nucléole.

(A et D) Signal correspondant à la protéine de fusion At79-GFP. (B et E) Images DIC des protoplastes. (C et F) Superposition des photos A-B et D-E.

Les flèches rouges et blanches indiquent respectivement le nucléole et la cavité nucléolaire.

Toutefois, la quantité de protéines immunoprécipitées est comparable pour ces trois lignées (figures 16.A, pistes 2, 5 et 6). Nous avons ensuite testé par RT-PCR la présence du précurseur de l'ARNt méthionine et du précurseur dicistronique appelé ARNt-snoR43.1 (figure 16.B). Le précurseur ARNt-snoR43.1 est un représentant d'un nouveau mode d'expression des snoARN spécifique aux plantes. Ce type de gène code un précurseur dicistronique contenant un ARNt et un snoARN, qui est transcrit par la Pol III grâce au promoteur interne de l'ARNt. La maturation de ce précurseur génère un ARNt et un snoARN (Kruszka *et al.* 2003). Quand les ARN n'ont pas été rétrotranscrits (figure 16.B, pistes 2, 4, 6, 8, 10 et 12), nous n'avons obtenu aucune amplification, les échantillons sont donc exempts de toute contamination par de l'ADN. De plus, les deux précurseurs sont correctement amplifiés à partir des ADNc de nos contrôles positifs : les ARN présents dans les extraits totaux des trois lignées testées (figure 16.B, pistes 3, 7 et 11), ainsi que les ARN co-immunoprécipités avec la protéine de fusion AtLa1-GFP (figure 16.B, piste 13). En revanche, nous n'avons pas obtenu d'amplification à partir des ADNc des ARN présents dans l'éluat de la lignée At79-GFP (figure 16.B, piste 9). Les protéines de fusion At79-GFP et GFP-At79 ne semblent donc pas capables de lier les extrémités 3' UUU-OH des transcrits Pol III que nous avons testés. Il est bien sûr possible que la protéine At79 fixe d'autres transcrits Pol III ou que les protéines de fusion soient inactives en raison de la présence du tag. Cependant, d'autres expériences réalisées et présentées ultérieurement (cf. paragraphe II.3.2 des résultats) nous laissent penser que la protéine At79-GFP soit en mesure de lier des ARN.

II.2.2) Localisation subcellulaire de la protéine At79

Dans le dessein d'identifier des fonctions cellulaires de la protéine At79, nous avons cherché de nouvelles pistes en déterminant la localisation subcellulaire de la protéine. Nous avons donc tiré parti des lignées des cellules exprimant les protéines de fusion At79-GFP et GFP-At79 que nous avons produites pour étudier le rôle de cette protéine dans la biogenèse des transcrits Pol III. Afin de pouvoir observer des cellules isolées, nous avons incubé les suspensions cellulaires exprimant les protéines de fusion avec des enzymes capables de digérer les parois cellulaires, créant ainsi des protoplastes. Ces protoplastes ont ensuite été directement observés en microscopie confocale (figure 17).

La protéine At79 apparaît diffuse dans le nucléoplasme, et ne semble pas s'accumuler dans le nucléole (figure 17.A à 17.C), excepté dans certaines cellules où elle apparaît concentrée dans la cavité nucléolaire (figure 17.D à 17.F). Cette localisation correspond à

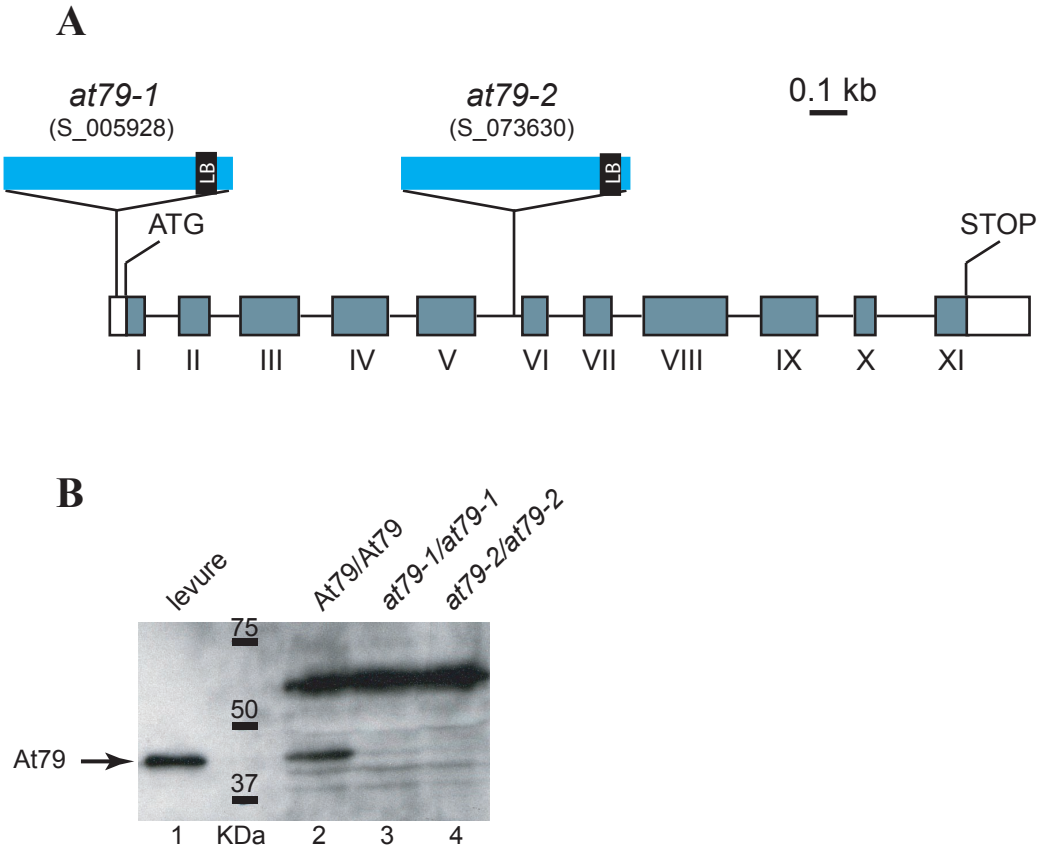


Figure 18. Caractérisation de deux lignées d'insertion ADN-T pour le gène *AT79*

(A) Représentation schématique des insertions ADN-T au locus At1g79880. La comparaison de la séquence génomique de ce locus et des séquences ADNc identifiés (At1g79880.1) a révélé que le gène *AT79* comporte 11 exons (figurés par des boîtes grises), séparés par 10 introns. Les boîtes blanches représentent les régions 5' et 3' non traduites. Les positions des insertions ADN-T ont été reportées sur la représentation schématique du locus At1g79880 et sont représentées par des boîtes bleues.

(B) Analyse par western blot d'extraits protéiques de levure exprimant la protéine At79 (piste 1), de plantes sauvages (piste 2) et de plantes homozygotes pour l'allèle mutant *at79-1* ou *at79-2* (pistes 3 et 4). L'anticorps utilisé est un anticorps de lapin dirigé contre un peptide de la partie carboxy-terminale de la protéine At79 (cet anticorps a été décrit dans l'article 1).

celle observée pour la protéine AtLa1 fusionnée à un tag GFP (figure 5.D à 5.I de l'article 1). Ce résultat montre que comme AtLa1, At79 est une protéine majoritairement nucléaire et que sous certaines conditions, qui restent à déterminer, elle s'accumule dans la cavité nucléolaire. En revanche, comme nous n'avons pas fait d'expérience de colocalisation avec U2B", nous ne pouvons pas apporter d'information sur sa présence éventuelle dans les corps de Cajal.

II.2.3) Lignées mutantes pour le gène At79

Avec pour même objectif de définir les fonctions cellulaires de la protéine At79, nous avons recherché dans les collections de lignées d'insertion T-DNA, des lignées porteuses d'une insertion au locus *At1g79880* (*AT79*). Nous avons choisi deux lignées avec des sites d'insertions différents dans la collection SALK : S_005928 et S_073630, que nous avons respectivement nommées *at79-1* et *at79-2*.

Dans un premier temps, nous avons vérifié par séquençage la position de ces insertions dans le gène *AT79* (figure 18.A), l'ADN-T inséré dans la lignée *at79-1* se situe dans la région 5' non-traduite du gène, alors que celui inséré dans la lignée *at79-2* se situe dans le cinquième intron.

Nous avons ensuite sélectionné des plantes homozygotes pour chacune des deux lignées d'insertion T-DNA et vérifié par western blot qu'elles ne produisaient plus la protéine At79 (figure 18.B). Malgré l'absence de la protéine At79, les plantes homozygotes pour les mutations *at79-1* ou *at79-2* ne présentent pas de phénotypes évidents affectant leur croissance ou leur morphologie (données non-présentées). Ce résultat montre que la protéine At79 n'est pas indispensable à la viabilité d'*Arabidopsis* dans des conditions de croissance standard.

La protéine At79 présente les caractéristiques structurales d'une authentique protéine La et une localisation nucléaire cohérente avec les fonctions La. Cependant, elle ne lie visiblement pas les transcrits Pol III. Nous avons donc cherché à établir une comparaison entre les fonctionnalités de AtLa1 et At79.

II.3. Étude de la dualité de fonction entre AtLa1 et At79

II.3.1) Colocalisation des deux protéines

Au cours de nos expériences visant à déterminer la localisation subcellulaire des protéines AtLa1 et At79 (cf. article 1 et paragraphe II.2.2 des résultats), nous avons montré

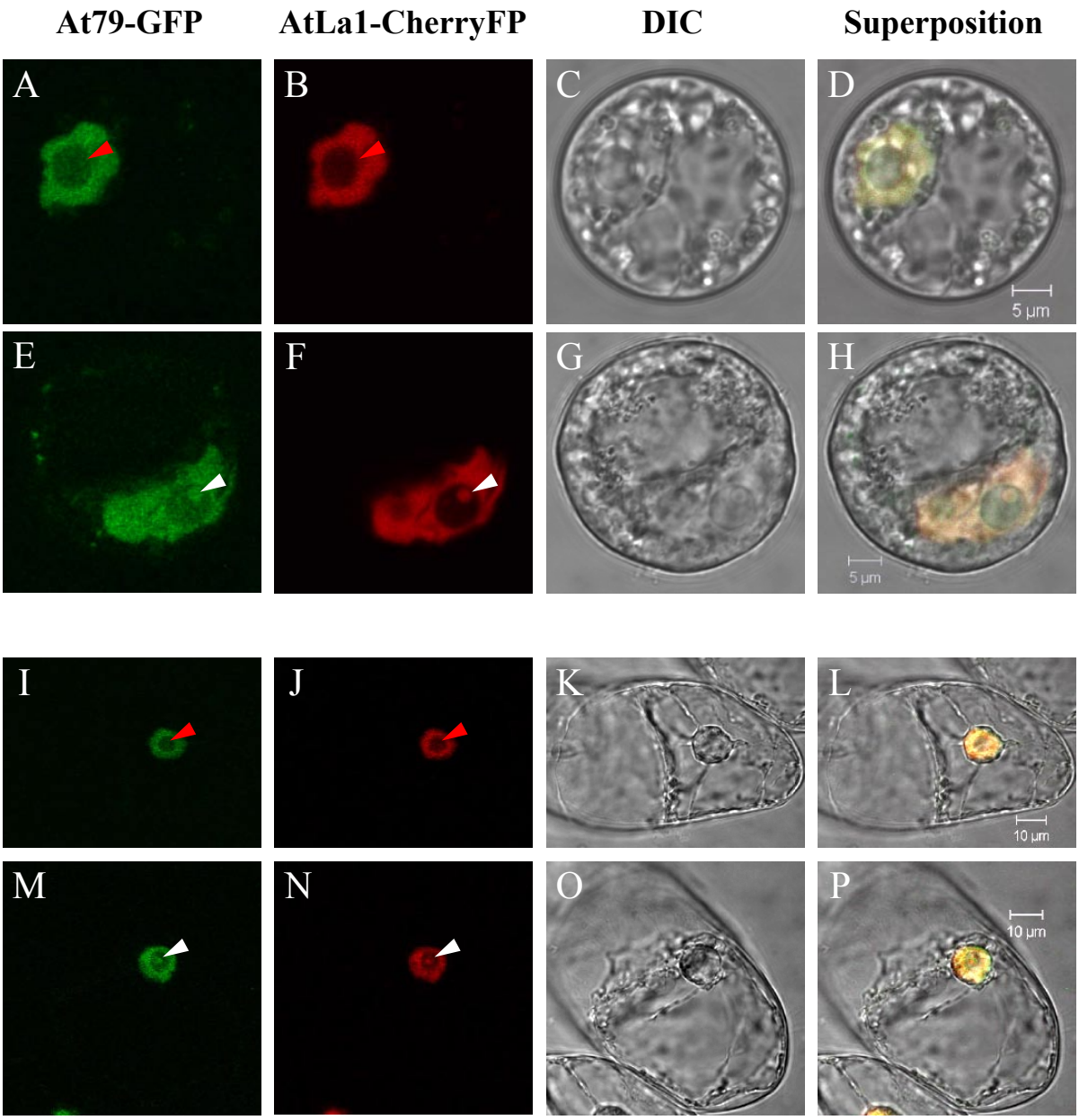


Figure 19. Colocalisation des protéines de fusion At79-GFP et AtLa1-CherryFP

Des protoplastes ont été préparés à partir de cellules d'*Arabidopsis* exprimant de façon stable les protéines de fusion At79-GFP et AtLa1-CherryFP. Ces protoplastes (A-H), ainsi que des cellules de tabac BY2 exprimant les mêmes protéines de fusions (I-P), ont été directement observés à l'aide d'un microscope confocal.

(A-D et I-L) Le plus souvent, les protéines At79-GFP et AtLa1-CherryFP apparaissent diffuses dans le noyau, à l'exception du nucléole.

(E-H et M-P) Dans un nombre plus rare de cas, les deux protéines de fusion apparaissent diffuses dans le noyau et fortement concentrée dans la cavité nucléolaire, mais toujours absente du reste du nucléole.

Lors de ces observations, nous avons toujours constaté une colocalisation des protéines At32-GFP et AtLa1-CherryFP.

Les flèches rouges et blanches indiquent respectivement le nucléole et la cavité nucléolaire.

que ces protéines étaient majoritairement nucléaires et peu abondantes ou absentes du nucléole. Sous certaines conditions qui restent à déterminer, ces deux protéines peuvent également être visualisées dans la cavité nucléolaire. Même si les protéines AtLa1 et At79 semblent se situer dans les mêmes compartiments cellulaires, il est cependant possible qu'elles s'y trouvent séparément, par exemple à différents stades du cycle cellulaire (dans les cavités nucléolaires visibles en G1 ou en G2).

Pour répondre à cette question, nous avons construit des lignées de cellules d'*Arabidopsis* et de tabac (BY2) exprimant les protéines de fusion At79-GFP et AtLa1-CherryFP. La protéine Cherry Fluorescent Protein (CherryFP) a les mêmes longueurs d'onde d'excitation et d'émission que la RFP et comme elle, elle apparaît rouge. Au contraire des cellules d'*Arabidopsis,* les cellules BY2 de tabac s'agrègent peu et poussent donc en suspensions très fines. Cette caractéristique permet de les observer directement sans les digérer pour obtenir des protoplastes individualisés. Comme le montre la figure 19, les deux protéines colocalisent parfaitement que ce soit en protoplastes de cellules d'*Arabidopsis* (figure 19.A à 19.H) ou en cellules de tabac (figure 19.I à 19.P). En ce qui concerne le cas particulier de la cavité nucléolaire, nos observations ont révélé que lorsqu'une des deux protéines est présente, l'autre l'est aussi. Étant donné que les localisations observées sont les mêmes dans les protoplastes de cellules d'*Arabidopsis* et dans les cellules de tabac, la présence des protéines AtLa1 et At79 dans la cavité nucléolaire (figure 19.E à 19.H et 19.M à 19.P) n'est pas un artefact dû à la production des protoplastes. Nous pouvons donc conclure que, dans les conditions testées, les protéines AtLa1 et At79 sont présentes, ensemble, dans les mêmes compartiments cellulaires.

II.3.2) Analyse et comparaison des ARN liés par AtLa1 et At79

Lors de la caractérisation de la fonction La, nous avons testé la capacité d'AtLa1 et At79 à lier les transcrits Pol III (figure 4 de l'article 1 et paragraphe II.2.1 des résultats). Nous avons construit des lignées de cellules exprimant des fusions des deux protéines avec la GFP. Grâce à des anticorps dirigés contre la GFP nous avons co-immunoprécipité les ARN liés aux deux protéines. Nous avons choisi une approche par RT-PCR visant des précurseurs de transcrits Pol III différenciables des transcrits matures. Seule AtLa1 s'est révélée capable de lier nos candidats.

Figure 20. Marquage pCp des ARN co-immunoprécipités par les protéines GFP, AtLa1-GFP et At79-GFP

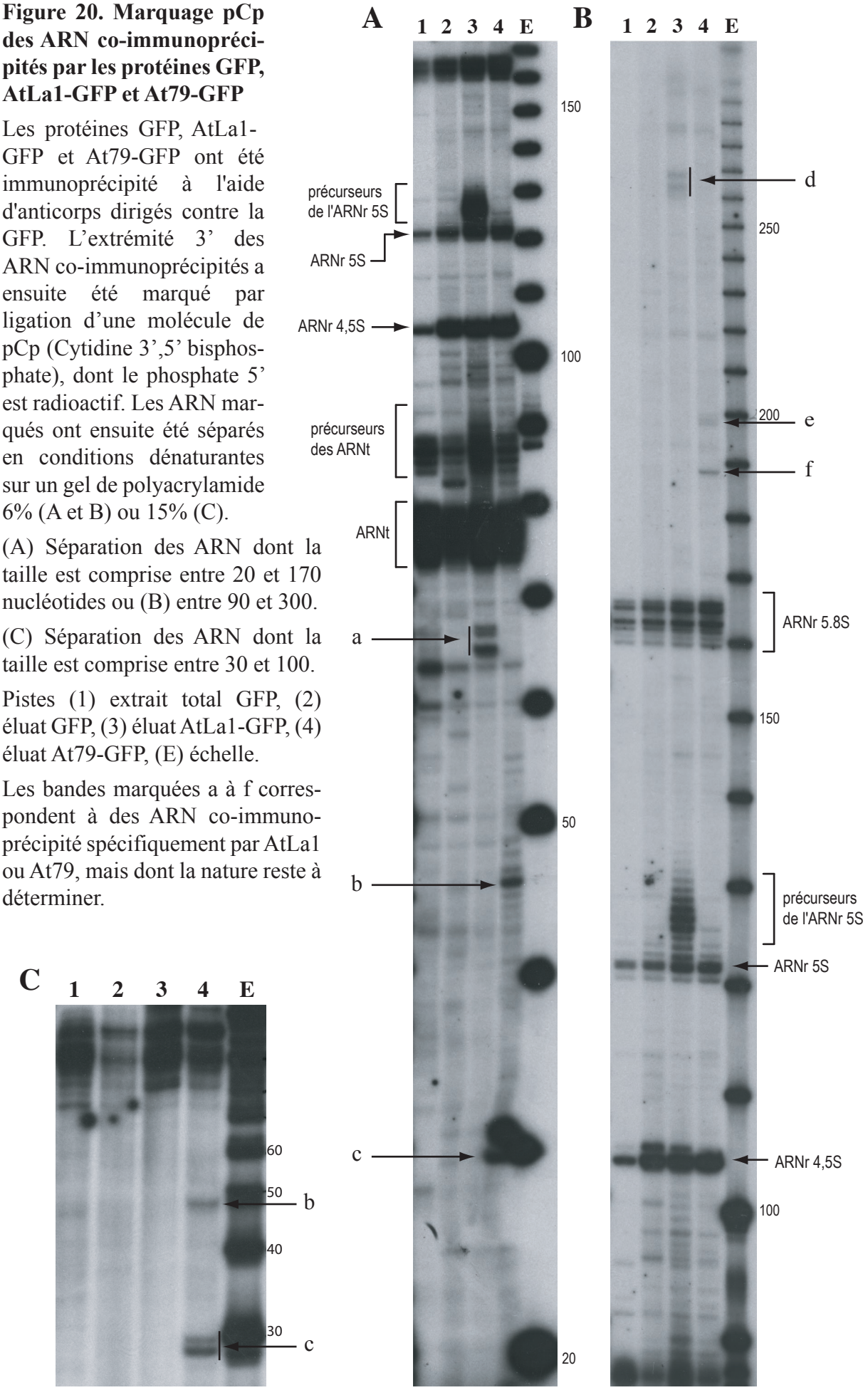
Les protéines GFP, AtLa1-GFP et At79-GFP ont été immunoprécipité à l'aide d'anticorps dirigés contre la GFP. L'extrémité 3' des ARN co-immunoprécipités a ensuite été marqué par ligation d'une molécule de pCp (Cytidine 3',5' bisphosphate), dont le phosphate 5' est radioactif. Les ARN marqués ont ensuite été séparés en conditions dénaturantes sur un gel de polyacrylamide 6% (A et B) ou 15% (C).

(A) Séparation des ARN dont la taille est comprise entre 20 et 170 nucléotides ou (B) entre 90 et 300.

(C) Séparation des ARN dont la taille est comprise entre 30 et 100.

Pistes (1) extrait total GFP, (2) éluat GFP, (3) éluat AtLa1-GFP, (4) éluat At79-GFP, (E) échelle.

Les bandes marquées a à f correspondent à des ARN co-immunoprécipité spécifiquement par AtLa1 ou At79, mais dont la nature reste à déterminer.



Afin de déterminer si la protéine At79 fixe des ARN, et si en dehors des candidats testés jusqu'ici, il existe une redondance dans la fixation des ARN entre les deux protéines, nous avons mis en place une approche différente. Nous avons procédé au marquage de l'extrémité 3' des ARN co-immunoprécipités, par ligation d'une molécule de Cytidine 3',5' bisphosphate (pCp) portant un phosphate radioactif en 5'. Ainsi, tous les ARN présentant une extrémité 3'OH accessible sont marqués. Les ARN marqués ont ensuite été séparés par électrophorèse sur un gel de polyacrylamide dénaturant. Les résultats obtenus sont présentés en figure 20.

Parallèlement aux ARN co-immunoprécipités avec les fusions AtLa1-GFP et At79-GFP, nous avons marqué les ARN d'un extrait natif ainsi que les ARN co-immunoprécipités par la GFP seule afin de pouvoir identifier les bandes correspondant à des transcrits piégés aspécifiquement lors de l'immunoprécipitation, notamment des transcrits très représentés. Sur le gel 6% utilisé, on sépare les ARN dont les tailles sont comprises entre 20 et 300 nucléotides (figure 20.A et 20.B). Grâce au marqueur de taille, nous pouvons proposer une identité à certaines bandes correspondant à des ARN attendus de taille et d'abondance concordantes. Mais pour pouvoir affirmer que les ARN proposés correspondent effectivement aux bandes observées, il faudrait vérifier leurs identités par hybridation northern blot. Ainsi, parmi les bandes du bruit de fond, le groupe de bandes entre 70 et 80 nucléotides (nt) peut correspondre aux ARNt, la bande au-dessus de 100 nt à l'ARNr chloroplastique 4,5S, la bande au-dessus de 120 nt à l'ARNr 5S, et dans le groupe de bandes d'environ 150-160 nt se trouvent probablement les deux bandes correspondant à l'ARNr 5,8S.

En comparant les pistes correspondant aux ARN co-immunoprécipités par les protéines AtLa1-GFP et At79-GFP, on voit qu'en dehors du bruit de fond, elles ne fixent pas les mêmes ARN. Dans la piste des ARN co-immunoprécipités par AtLa1-GFP (figure 20.A à 20.C, piste 3), on peut identifier certaines bandes correspondant probablement à des ARN transcrits par l'ARN polymérase III. Ainsi, on voit un enrichissement du signal au-dessus du bruit de fond dans la piste AtLa1-GFP au niveau 85-90 nt (figure 20.A, piste 3), qui est la zone de migration attendue des précurseurs d'ARNt. Il semble donc que contrairement à la GFP et la protéine de fusion At79-GFP, la protéine AtLa1-GFP lie les précurseurs d'ARNt, ce qui est cohérent avec les résultats que nous avons obtenus par RT-PCR dans l'article 1.

Au contraire des autres ARNr transcrits par la Pol I, l'ARNr 5S est transcrit par la Pol III et il a été montré que la protéine La s'associait à ces précurseurs chez les mammifères (Rinke and Steitz 1982). Chez *Arabidopsis*, le nombre de gènes d'ADNr 5S est estimé à

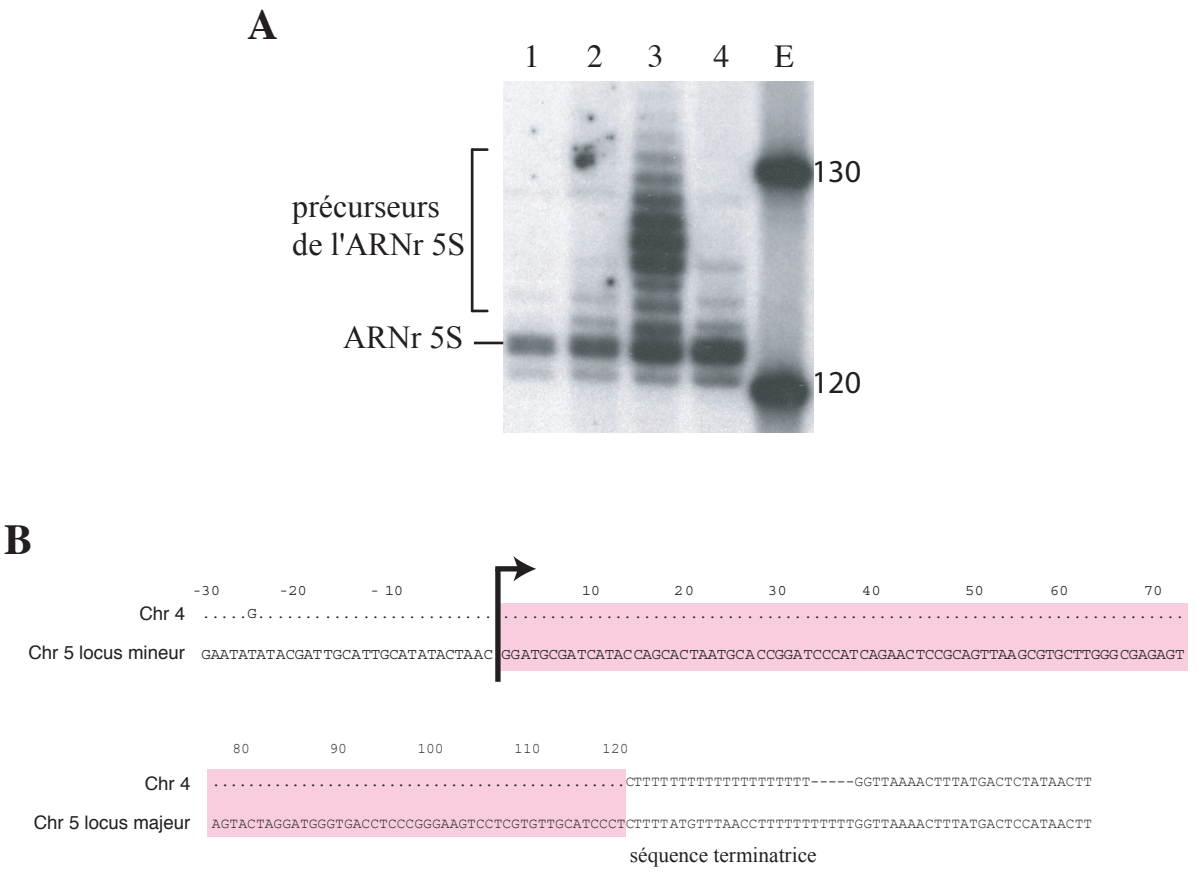


Figure 21. AtLa1-GFP lie les précurseurs de l'ARNr 5S

Les protéines GFP, AtLa1-GFP et At79-GFP ont été immunoprécipités à l'aide d'anticorps dirigés contre la GFP. L'extrémité 3' des ARN co-immunoprécipités a ensuite été marqué par ligation d'une molécule de pCp, dont le phosphate 5' est radioactif. Les ARN marqués ont ensuite été séparés en conditions dénaturantes sur un gel de polyacrylamide 6% (A).

(A) Détail de la figure 8.

Pistes (1) extrait total GFP, (2) éluat GFP, (3) éluat AtLa1-GFP, (4) éluat At79-GFP, (E) échelle.

(B) Séquences consensus des loci exprimés d'ARNr 5S d'*Arabidopsis thaliana*. La séquence de l'ARNr 5S mature est surlignée en rose.

environ 1000 copies par génome haploïde. Ces gènes sont répétés en tandem, ils sont situés dans l'hétérochromatine péricentromérique et sont organisés en cinq loci, avec trois loci sur le chromosome 3, un locus sur le chromosome 4 et 2 loci sur le chromosome 5. Seul le locus du chromosome 4 et le locus majeur du chromosome 5 sont exprimés (Cloix *et al.* 2002). Dans la piste 3 de la figure 21.A on voit une échelle de 9 bandes au-dessus de l'ARNr 5S mature. Ces bandes ont des tailles cohérentes avec les séquences transcrites de l'ARN 5S présentées en figure 21.B. Les bandes très intenses, correspondant en taille à la séquence de l'ARN 5S mature plus 4 ou 5 nt, sont très probablement des précurseurs issus de la transcription des deux loci. Ces transcrits se terminent donc par un C, qui n'appartient pas à la séquence de l'ARN 5S mature, et 3 ou 4 U correspondant au signal de terminaison de la Pol III. Les bandes plus hautes coïncident vraisemblablement avec des transcrits issus du locus du chromosome 4 se terminant par 5 à 8 U.

AtLa1-GFP fixe d'autres ARN que nous n'avons pas identifiés, plusieurs bandes migrant aux environs de 65 nt (figure 20.A bande a), et d'autres aux environs de 265 nt (figure 20.B bande d). Peu d'ARN ayant une taille d'environ 65 nt ont été décrits chez *Arabidopsis.* Des petits ARN nucléolaires (Marker *et al.* 2002) et certains précurseurs de microARN (Reinhart *et al.* 2002) ont des tailles correspondantes. Étant donné que les précurseurs des petits ARN nucléolaires et des microARN sont faiblement abondants, il paraît peu probable qu'ils forment des bandes aussi intenses. Il serait donc intéressant de cloner les bandes correspondantes et de les séquencer. En ce qui concerne les bandes autour de 265 nt, il pourrait s'agir de précurseurs de l'ARN 7-2/MRP, un autre transcrit Pol III appartenant à une particule ribonucléoprotique appelée RNase MRP, et qui serait impliquée dans la maturation des ARN ribosomiques (Kiss *et al.* 1992).

Si l'on s'intéresse aux ARN co-immunoprécipités par la protéine At79-GFP, le résultat le plus évident est l'absence des bandes correspondant aux précurseurs des transcrits Pol III. En effet, dans les pistes correspondant aux ARN co-immunoprécipités par la protéine At79-GFP (figure 20.A et 20.B, piste 4), il n'y a pas d'enrichissement des bandes supposées être les précurseurs des ARNt ou les précurseurs de l'ARNr 5S. Au contraire d'AtLa1, At79 ne semble donc pas capable de lier les transcrits Pol III, ce qui confirme les résultats obtenus après RT-PCR (figure 16). En revanche, on note la présence de quatre groupes de bandes clairement et spécifiquement enrichies dans l'éluat de la protéine At79-GFP. Les plus grandes ont des tailles de 190 nt et 200 nt (figure 20.B, bandes e et f) et grâce à l'utilisation d'un marqueur de taille ARN, nous avons déterminé que les plus petites ont des tailles de 28-29 nt

et 44 nt (figure 20.B et 20.C, bandes b et c). Il est pour l'instant difficile, uniquement sur la base de leur taille, de déterminer la nature de ces ARN associés directement ou indirectement avec At79.

L'ensemble des résultats présentés dans ce chapitre ont été reproduits à l'identique dans une seconde expérience complètement indépendante. Cependant, alors que nous projetions d'identifier par clonage et séquençage la nature des ARN associés spécifiquement à At79, nous avons rencontré plusieurs difficultés techniques (probablement liées à une dégradation des ARN) qui nous ont empêchés de reproduire les profils d'immunoprécipitation présentés ici.

II.4. Conclusion

Arabidopsis thaliana présente une situation originale, puisqu'il n'existe pas une, mais deux protéines La : AtLa1 et At79. Nous avons établi que la protéine AtLa1 assurait au moins les fonctions La dans la maturation des transcrits Pol III. En ce qui concerne At79, la situation reste plus floue.

Le gène *AT79* n'est pas un pseudogène puisque la protéine At79 endogène est détectable dans des extraits protéiques de cellules et de plantes sauvages, et absente dans les extraits protéiques des plantes des deux lignées d'insertion ADN-T (*at79-1* et *at79-2*) pour le gène *AT79* que nous avons caractérisées (figure 18).

Contrairement à la protéine AtLa1, la protéine At79 ne semble pas en mesure de lier les transcrits Pol III. Lors des expériences de RT-PCR, nous n'avons pas détecté le précurseur de l'ARNt méthionine ou le précurseur dicistronique ARNt-snoR43.1 (figure 16.B), parmi les ARN co-immunoprécipités avec les protéines de fusion GFP-At79 et At79-GFP. De plus, après un marquage global des ARN (cf. paragraphe II.3.2 des résultats), on ne voit pas, dans les immunoprécipitations At79-GFP, les précurseurs de transcrits Pol III clairement identifiés parmi les ARN co-immunoprécipités par la protéine AtLa1-GFP. Étant donné la présence de bandes spécifiques dans les pistes correspondant aux immunoprécipitations At79-GFP, la protéine de fusion At79-GFP est au moins partiellement fonctionnelle. La présence de ces bandes spécifiques est aussi un argument en faveur d'une spécialisation de fonction des deux protéines.

On ne peut pas exclure que la protéine At79 ait une affinité pour les transcrits Pol III plus faible qu'AtLa1. Dans les cellules transgéniques, les protéines de fusion sont en compétition avec les protéines endogènes et dans nos lignées cellulaires, la protéine de fusion At79-GFP est plus faiblement exprimée que la protéine de fusion AtLa1-GFP. Il est donc possible que la protéine At79-GFP soit en condition défavorable par rapport à la protéine AtLa1 endogène pour la fixation des transcrits Pol III. Cependant, en défaveur de cette hypothèse, la protéine At79 sans tag, en l'absence de compétiteur, n'a pas complémenté les défauts de maturation des transcrits Pol III observés dans des souches mutantes pour le gène *LHP1* (article 1).

L'observation par microscopie confocale de protoplastes d'*Arabidopsis* et de cellules de tabac co-exprimant les protéines de fusion At79-GFP et AtLa1-CherryFP a montré qu'elles avaient la même localisation dans le nucléoplasme. Ce résultat confirme que la protéine At79 ait une localisation subcellulaire cohérente avec les fonctions La.

Nous avons également observé la présence des deux protéines de fusion dans la cavité nucléolaire. Il semble exister deux types de cavité nucléolaire, l'une visible en G1, l'autre visible en G2 (Gonzalez-Camacho and Medina 2006). Il serait donc intéressant d'observer la localisation des protéines AtLa1 et At79 dans des cultures synchronisées de cellules de tabac, afin de déterminer à quel moment du cycle cellulaire elles se trouvent dans la cavité nucléolaire.

Un autre argument en faveur de la spécialisation de fonction nous a été fourni par l'étude des lignées d'insertion ADN-T dans les gènes *ATLA1* et *AT79*. Alors que la mutation du gène *ATLA1* est létale à l'état homozygote (figure 6 de l'article 1), les lignées mutantes pour le gène *AT79* sont viables et ne présentent pas de phénotypes (données non-présentées). Il semble donc que même si les deux protéines AtLa1 et At79 sont proches, elles ne soient pas ou pas entièrement redondantes puisque la protéine At79 n'est pas capable de compenser l'absence de la protéine AtLa1 dans les graines portant la mutation du gène *ATLA1* à l'état homozygote (figure 6.B de l'article 1).

Selon nous, l'ensemble de ces résultats indique que la protéine At79 serait une authentique protéine La et qu'elle serait spécialisée dans des fonctions autres que la maturation des transcrits Pol III.

III. Fonction La sur la biologie des SINE

Comme nous l'avons vu précédemment, les éléments SINE sont transcrits par l'ARN polymérase III. Leurs transcrits primaires sont donc très probablement amenés à interagir avec la protéine La. Nous avons donc voulu vérifier si l'une des deux protéines d'*Arabidopsis* AtLa1 ou At79 était capable de lier les transcrits SINE *in vivo*.

Dans le génome d'*Arabidopsis*, six familles différentes de SINE (SB2 à SB7) ont été identifiées (Deragon and Zhang 2006) (cf. paragraphe II.2 de la synthèse bibliographique). Pour cette étude, nous avons choisi de nous restreindre aux transcrits de deux de ces familles. La première que nous avons choisie est la famille du SINE SB2 (également appelée RathE1), qui est la famille la plus représentée au sein du génome d'*Arabidopsis*. Elle compte 144 copies, soit près de la moitié des éléments SINE identifiés chez *Arabidopsis*, qui sont dispersées dans tout le génome et notamment dans les régions géniques. Certains éléments de la famille SB2 sont même insérés dans des introns et sont donc co-transcrits dans des ARNm (Lenoir *et al.* 2001). Nous avons également choisi la famille du SINE SB4 (ou RathE3) qui est moins représentée dans le génome (64 copies), mais dont les copies sont plus rares dans les régions géniques et pour laquelle il existe donc très peu de cas de co-transcription.

De plus, nous avons réalisé cette étude en utilisant comme fond génétique une lignée cellulaire d'*Arabidopsis* appelée Na7 (Pelissier *et al.* 2004). Cette lignée a été créée au laboratoire par T. Pélissier et a la particularité de présenter dans son génome une copie du SINE SB1 de *Brassica napus*. Cet élément SINE appartient à une famille qui est normalement absente du génome d'*Arabidopsis* et le locus introduit, appelé locus Na7, présente toutes les caractéristiques d'un locus fondateur. Le locus Na7 est la seule copie du SINE SB1 à être conservée dans toutes les espèces de *Brassica* testées (Tatout *et al.* 1999). Chez *Brassica napus*, la séquence du SINE SB1 présente au locus Na7 correspond au transcrit majeur détecté *in vivo* (Deragon *et al.* 1996 ; Arnaud *et al.* 2001). En effet, ce locus contient, en amont du SINE, des séquences amplificatrices de la transcription, qui lui permettent d'échapper partiellement à la répression transcriptionnelle (Arnaud *et al.* 2001).

Dans la lignée cellulaire d'*Arabidopsis*, le locus Na7 est transcriptionnellement actif et les transcrits pleine longueur, poly(A) terminal et cytoplasmique court du SINE SB1 peuvent être visualisés par hybridation northern (Pelissier *et al.* 2004) (cf. paragraphe II.3.2.b de la synthèse bibliographique). Le SINE SB1 étant normalement absent du génome d'*Arabidopsis*,

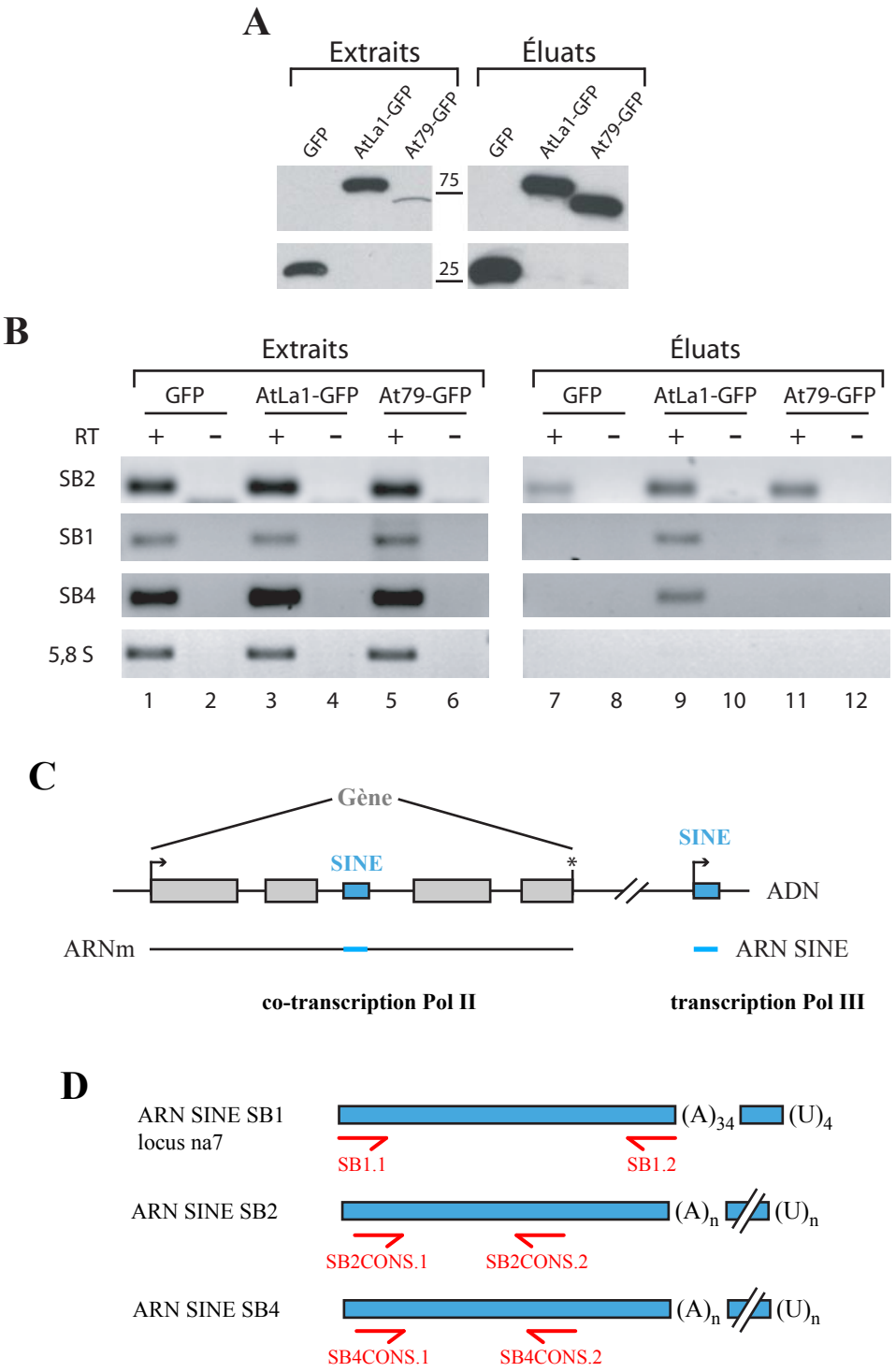


Figure 22. La protéine AtLa1 fixe spécifiquement les transcrits des éléments SINE SB1 et SB4

(A) Analyse par western blot des protéines présentes dans les extraits natifs et dans l'éluat des immunoprécipitations réalisées avec un anticorps polyclonal de lapin dirigé contre la GFP. Les western blot ont été hybridés avec un anticorps monoclonal de souris dirigé contre la GFP.

(B) Analyse par PCR des ADNc préparés à partir des extraits natifs et des éluats des immunoprécipitations. Les amplifications PCR ont été réalisées avec des paires d'oligonucléotides spécifiques des ADNc des transcrits des SINE SB1, SB2 et SB4, ainsi que de l'ADNc de l'ARNr 5,8S. Tous les échantillons ont été préalablement traités à la DNase, puis rétrotranscrit (+) ou non (-).

(C) Représentation schématique du phénomène de co-transcription. Les exons sont figurés par des boîtes grises et les éléments SINE par des boîtes bleues.

(D) Positions des amorces utilisées détecter les ARN SINE en (B).

l'ensemble des transcrits détectés proviennent du transgène Na7. L'utilisation de cette lignée cellulaire, nous permet donc de nous affranchir du problème de co-transcription dans le cas particulier de l'élément SINE SB1.

Nous avons donc transformé la lignée Na7 de façon stable pour qu'elle exprime la GFP ou les protéines de fusion AtLa1-GFP et At79-GFP. Nous avons ensuite testé l'association des protéines avec les transcrits des SINE endogènes SB2 et SB4 et les transcrits du SINE SB1 issus du locus Na7 introduit artificiellement dans le génome des cellules.

III.1. Co-immunoprécipitation des transcrits SINE par AtLa1

Comme précédemment, nous avons immunoprécipité les protéines GFP, AtLa1-GFP et At79-GFP à l'aide d'anticorps dirigés contre la GFP. Après avoir vérifier l'efficacité des immunoprécipitations (figure 22.A), nous avons recherché par RT-PCR la présence de transcrits SINE issus des éléments SB1, SB2, et SB4 parmi les ARN co-immunoprécipités avec les anticorps contre la GFP dans les extraits de nos différentes lignées.

Dans le cas où les ARN n'ont pas été rétrotranscrits (figure 22.B, pistes 2, 4, 6, 8, 10 et 12), nous n'avons obtenu aucune amplification, les échantillons sont donc exempts de toute contamination ADN. Pour éprouver la spécificité des immunoprécipitations, nous avons utilisé comme contrôle négatif une paire d'amorces ciblant l'ARN ribosomal 5,8S, un ARN particulièrement abondant, mais que la protéine La n'est pas censée lier (figure 22). Comme attendu l'ARNr 5,8S peut être détecté dans les extraits natifs (figure 22.B, pistes 1, 3 et 5), mais pas dans les éluats (figure 22.B, pistes 7, 9 et 11).

Après rétrotranscription, les transcrits correspondant aux trois familles d'éléments choisis sont également détectés dans les extraits natifs utilisés pour les immunoprécipitations (figure 22.B, pistes 1, 3 et 5). Avec les amorces ciblant les SINE endogènes SB2 et SB4, les produits d'amplification obtenus peuvent provenir de vrais transcrits SINE, mais aussi d'ARNm contenant dans leur séquence une copie SINE (phénomène de co-transcription). En revanche, les produits d'amplification obtenus avec les amorces ciblant le SINE SB1 sont obligatoirement dus aux transcrits issus du locus Na7.

Avec les amorces ciblant la séquence du SINE SB2, nous avons obtenu des produits d'amplification à partir des ARN co-immunoprécipités par les trois protéines GFP, AtLa1-GFP, At79-GFP (Figure 22.B, pistes 7, 9 et 11). Ces produits sont donc sûrement dus

au phénomène de co-transcription et correspondent à du bruit de fond (figure 22.C). Ces résultats ne nous permettent donc pas de conclure sur la liaison de nos protéines sur les transcrits SINE de la famille SB2. En revanche, les amorces ciblant les transcrits des SINE SB1 et SB4 permettent d'obtenir des produits d'amplification à partir des ARN co-immunoprécipités par la protéine de fusion AtLa1-GFP (figure 22.B, piste 9), mais pas à partir de ceux co-immunoprécipités par la GFP ou la protéine At79-GFP (figure 22.B, pistes 7 et 11). Pour SB4 qui est un élément SINE endogène du génome d'*Arabidopsis*, cela peut donc signifier qu'AtLa1 fixe des transcrits de ce SINE, ou, même si cela paraît peu probable, qu'elle fixe des transcrits dans lesquels la séquence de SB4 est présente. Étant donné que le SINE SB1 n'existe pas naturellement dans le génome d'*Arabidopsis*, les produits d'amplifications obtenus avec les amorces ciblant cet élément ne peuvent pas être dus au phénomène de co-transcription, il semble donc que la protéine AtLa1 lie les transcrits SINE Pol III.

III.2. Conclusion

Comme nous pouvions nous y attendre, les résultats de co-immunoprécipitation présentés ici suggèrent que seule la protéine AtLa1 lie les transcrits SINE. Les amorces utilisées pour la RT-PCR se trouvent dans la séquence du SINE avant la queue poly(A) (figure 22.D), on ne peut donc pas savoir à quelle forme du transcrit SINE : pleine longueur, poly(A) terminale, ou cytoplasmique courte, la protéine AtLa1 se lie. Il paraît plus probable que la protéine AtLa1 lie la forme pleine longueur qui se termine par une séquence 3' UUU-OH, et appartient donc, ne serait-ce que transitoirement à la particule ribonucléoprotéique des SINE. Étant donné que la protéine La est capable de fixer des ARN ne se terminant pas en UUU-OH (ARNm TOP, IRES), il serait intéressant de vérifier par cartographie à la RNase sur les ARN co-immunoprécipités, que la protéine AtLa1 ne fixe que cette forme des transcrits SINE.

De même, il serait intéressant d'étudier quel peut-être l'impact de la protéine AtLa1 sur les processus de maturation des transcrits SINE. Ainsi, dans des lignées Na7 surexprimant la protéine AtLa1, le suivi des différentes formes du SINE SB1 : pleine longueur, poly(A) terminale et cytoplasmique courte par hybridation northern devrait permettre de déterminer si l'accumulation des différentes formes de transcrits est affectée, par rapport à celle observée normalement dans les cellules Na7.

Dans les expériences d'immunoprécipitation de la protéine At79, nous n'avons pas détecté la présence des transcrits SINE testés. Ce résultat est en accord avec les données que nous avons obtenus sur les interactions entre At79 et d'autres transcrits Pol III. Contrairement à la protéine AtLa1, la protéine At79 ne semble en effet pas capable de fixer l'extrémité 3' UUU-OH des transcrits Pol III, il n'est donc pas surprenant que nous ne détections pas d'association de la protéine At79 aux SINE.

IV. Structure évolutive des ARN SINE dérivés d'ARNt

La plupart des éléments SINE possèdent une partie 5' dérivée d'un ARN de transfert, dont la structure secondaire en trèfle est parfaitement connue. Paradoxalement, on dispose de peu de données expérimentales sur la structure secondaire qu'adoptent les ARN SINE. Dans le cadre d'une étude portant sur la structure secondaire de familles de SINE provenant de différents eucaryotes, j'ai déterminé la structure en solution des ARN de deux SINE de crucifères : le SINE SB1 de *Brassica napus* et le SINE SB2 d'*Arabidopsis thaliana.*

IV.1. Introduction

Lors de la transcription, l'ARN néosynthétisé se replie aussi vite qu'il est produit. On distingue trois niveaux dans la structure d'un ARN. La structure primaire correspond à la séquence en nucléotides, la structure secondaire est le résultat des interactions physico-chimiques entre les bases de l'ARN, tandis que la structure tertiaire consiste en la représentation exacte de la forme tridimensionnelle de la molécule. Les fonctions biologiques d'un ARN au sein de la cellule sont liées à la conformation spatiale qu'il adopte. On peut avoir accès à la structure tridimensionnelle de l'ARN par des méthodes biologiques, comme la cristallographie et la RMN, qui sont des méthodes sûres mais difficiles à mettre en place et coûteuses. La structure secondaire des ARN, plus accessible, permet donc de poser de bonnes bases pour la connaissance de la structure tridimensionnelle qui est le reflet de leur fonction. Il est possible de prédire la structure secondaire des ARN par des méthodes informatiques selon plusieurs approches. La plus utilisée pour étudier la structure secondaire des SINE est l'approche thermodynamique. Elle suppose que la structure secondaire d'un ARN est la plus stable parmi toutes les structures possibles, c'est-à-dire celle d'énergie minimale. De même il

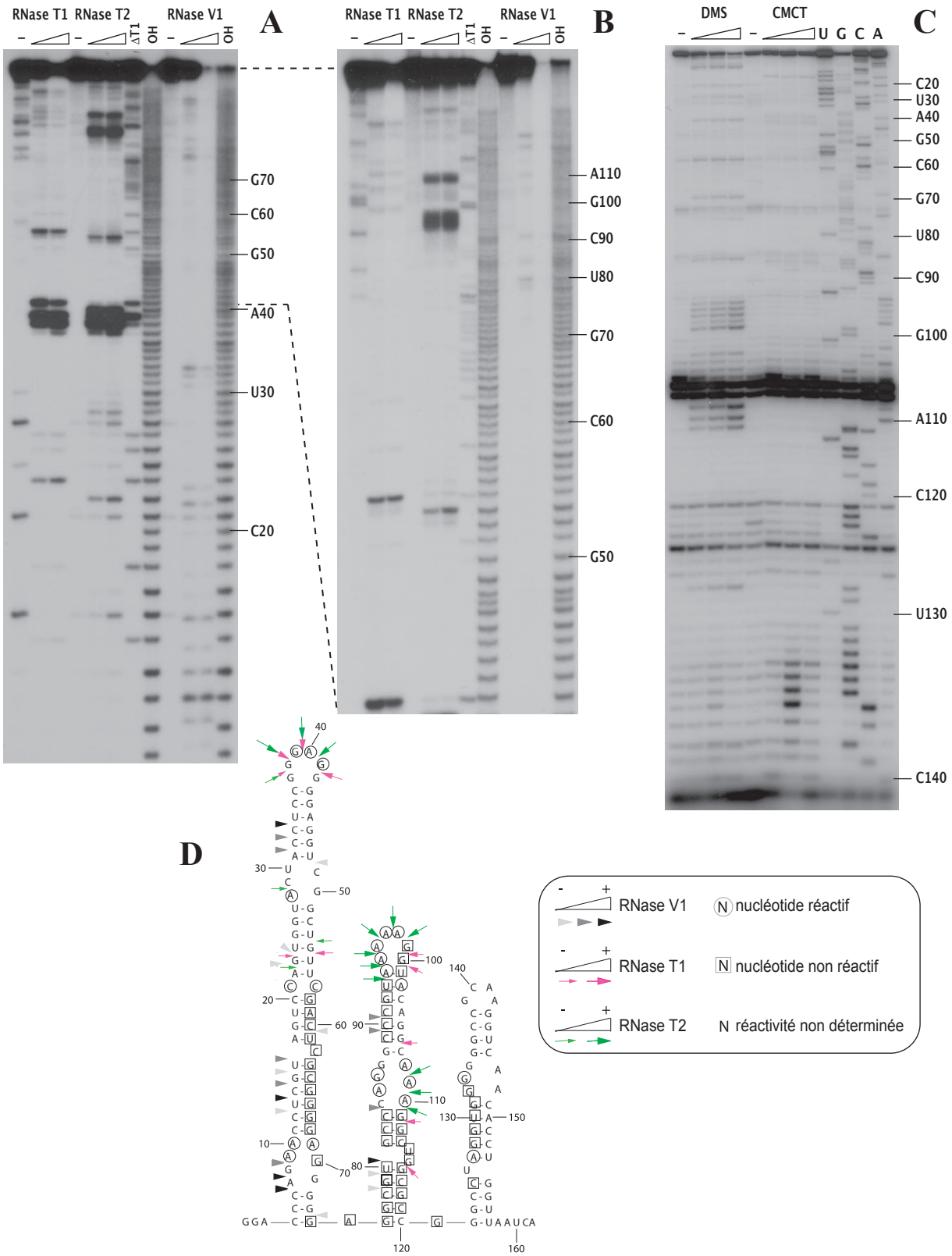


Figure 23. Cartographie en solution de la forme cytoplasmique courte de l'ARN SINE SB1

(A et B) Digestions enzymatiques de l'ARNsc marqué en 5'. ΔT1 : hydrolyse de l'ARN par la RNase T1 en conditions dénaturantes, statistiquement tous les G seront coupés. OH : échelle obtenue par hydrolyse alcaline de l'ARN marqué.

(C) Modification chimique par le DMS et le CMCT suivi d'une extension d'amorce. UGCA : pistes correspondant au séquençage. L'arrêt de la transcriptase inverse se fait une base avant la base modifiée.

(D) Coupures et modifications reportées sur la structure secondaire prédite par Mfold.

(-) : incubation de l'ARN dans les mêmes conditions mais sans traitement.

est possible d'avoir accès expérimentalement à la structure secondaire des ARN par des méthodes enzymatiques et chimiques (Ehresmann *et al.* 1987).

IV.2. Cartographie en solution des ARN des SINE SB1 de *Brassica napus* et SB2 d'*Arabidopsis thaliana*

IV.2.1) SB1

Nous avons analysé la structure en solution des formes cytoplasmique courte (sc, figure 23) et pleine longueur (fl, figure 24) de l'ARN du SINE SB1 par traitements enzymatiques et chimiques (Ehresmann *et al.* 1987). La séquence utilisée correspond à celle du SINE SB1 au locus fondateur Na7 dans le génome de *Brassica napus*, également utilisé dans notre étude de l'association des protéines AtLa1 et At79 aux transcrits SINE (cf. paragraphe III des résultats).

Pour le traitement enzymatique, nous avons utilisé la RNase V1 qui coupe les régions double brin ou empilées, la RNase T1 qui coupe les résidus G simple brin, ainsi que la RNase T2 qui coupe les résidus A simple brin et dans une moindre mesure les résidus G, C et U. Les traitements enzymatiques sont détectés directement. Les transcrits sont marqués radioactivement en 5', ils sont ensuite soumis à une digestion ménagée par les différentes enzymes et les fragments obtenus sont séparés par électrophorèse sur un gel de polyacrylamide dénaturant. En raison de leurs tailles importantes les enzymes ne permettent pas une analyse fine de la structure, mais elles renseignent sur l'accessibilité des différentes régions de la molécule. Pour plus de précision, nous avons donc complété les traitements enzymatiques appliqués à l'ARNsc de SB1 par des traitements chimiques qui permettent de tester l'appariement, canonique ou non, des différentes bases. Le diméthylsulfate (DMS) et le dérivé de la carbodiimide (CMCT) modifient les bases sur une de leur position d'appariement Watson-Crick, si celles-ci ne sont pas impliquées dans une liaison hydrogène. Ainsi le DMS modifie les positions N1 des adénines et N3 des cytosines, et le CMCT modifie les positions N3 des uraciles et N1 des guanines. Les nucléotides modifiés sont révélés indirectement par la technique d'élongation d'amorce à la transcriptase inverse. L'ARN traité est hybridé à un oligonucléotide marqué radioactivement en 5', qui sert d'amorce à l'élongation par la transcriptase inverse. Celle-ci synthétise des chaînes d'ADNc jusqu'à ce que sa progression soit arrêtée par une modification. Les fragments d'ADNc obtenus sont séparés par électrophorèse sur gel de polyacrylamide dénaturant. Un séquençage de l'ARN non modifié

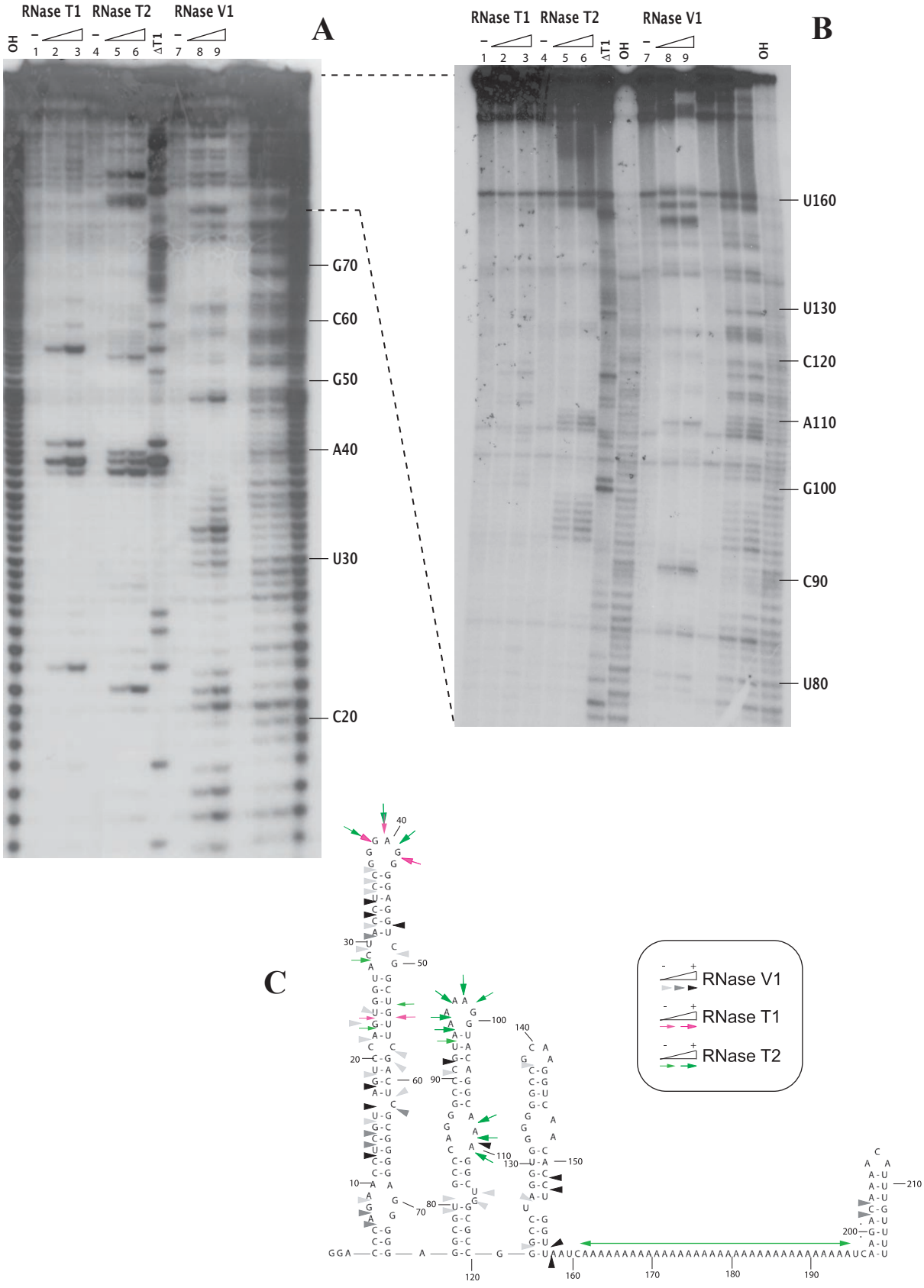


Figure 24. Cartographie en solution de la forme pleine longueur de l'ARN SINE SB1

(A et B) Digestions enzymatiques de l'ARNfl marqué en 5'. ΔT1 : hydrolyse de l'ARN par la RNase T1 en conditions dénaturantes. OH : échelle obtenue par hydrolyse alcaline de l'ARN marqué. (-) : incubation de l'ARN dans les mêmes conditions mais sans traitement.

(C) Coupures reportées sur la structure secondaire prédite par Mfold.

est effectué parallèlement de la même façon avec le même oligonucléotide pour replacer les arrêts de la transcriptase inverse et ainsi localiser les nucléotides modifiés. Préalablement à tous les traitements, les molécules d'ARN, marquées ou non, sont dénaturées à 100°, puis renaturées en présence de magnésium à 23°C pour leur permettre de se replier dans des conditions physiologiques.

Les figures 23 et 24 présentent respectivement les résultats des traitements enzymatiques et chimiques effectués sur l'ARN de la forme cytoplasmique courte de SB1 et les résultats des traitements enzymatiques effectués sur l'ARN pleine longueur de SB1. Nous avons positionné les sites de clivage et les modifications sur les modèles de structure secondaire prédits par le programme Mfold (Mathews *et al.* 1999 ; Zuker 2003).

Le modèle de structure pour l'ARNsc du SINE présente trois tiges-boucles (figure 23.D). La première qui comprend les nucléotides 4 à 74 est la plus longue (comprend les boîtes A nt 16 à 26 et B nt 53 à 63 reconnues par la Pol III). Elle se termine par une boucle, formée des nucléotides 37 à 41, fortement coupée par les RNase T1 et T2, qui se trouve au bout d'une hélice coupée par la RNase V1 et présentant des nucléotides non-réactifs. Cette hélice est interrompue par des zones non appariées : boucles internes ou renflements. Les nucléotides 9, 10 et 69 réactifs au DMS suggèrent qu'il existe un premier renflement dans le début de l'hélice. De même la réactivité des nucléotides 21 et 57, ainsi que les coupures concomitantes des RNases V1 et T2 indiquent dans cette zone une structure dynamique de l'hélice, et la présence possible d'une boucle interne. La deuxième tige-boucle est plus courte, elle se compose d'une hélice présentant des nucléotides non réactifs et des coupures par la RNase V1 et d'une boucle terminale, formée des nucléotides 93 à 100, coupée par les RNases T1 et T2. Elle présente aussi une boucle interne révélée par des nucléotides réactifs (86, 87 et 108, 109, 110) et des coupures par les RNases T1 (G111) et T2 (A108, 109, 110). Nous avons peu d'informations sur la partie 3' de l'ARN, mais elles sont compatibles avec le repliement en tige-boucle prédit par Mfold.

Les résultats expérimentaux présentés en figure 24 révèlent une structure de l'ARN pleine longueur du SINE SB1 au locus Na7 cohérente avec les prédictions de Mfold et confirment les résultats obtenus sur la forme courte. Ainsi sur les 120 premiers nucléotides on obtient presque exactement les mêmes coupures enzymatiques sur les ARN sc et fl. Les coupures qu'on ne retrouve pas sont des coupures faibles, donc des coupures moins

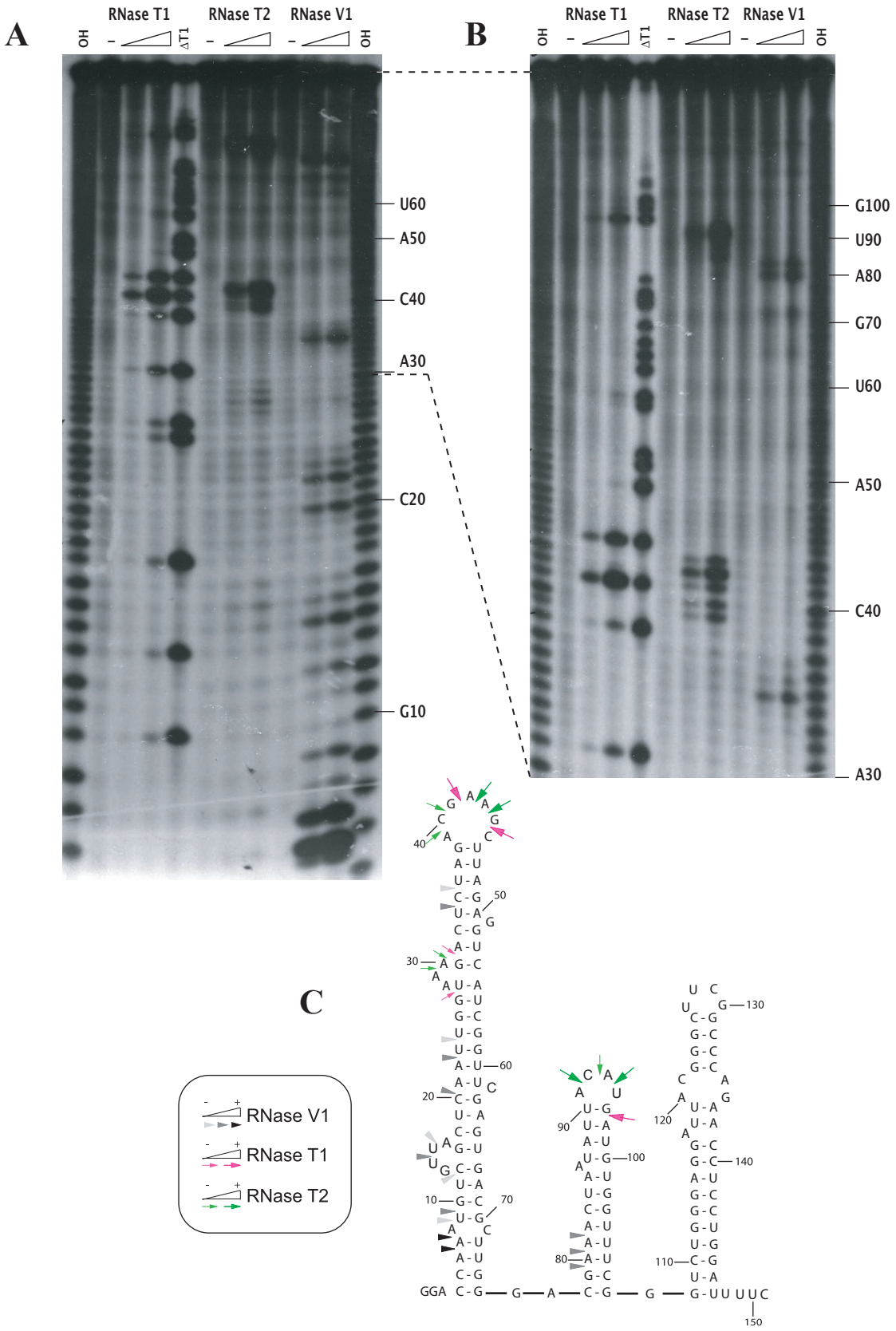


Figure 25. Cartographie en solution de la forme cytoplasmique courte de l'ARN SINE SB2

(A et B) Digestions enzymatiques de l'ARNsc marqué en 5'. ΔT1 : hydrolyse de l'ARN par la RNase T1 en conditions dénaturantes. OH : échelle obtenue par hydrolyse alcaline de l'ARN marqué. (-) : incubation de l'ARN dans les mêmes conditions mais sans traitement.

(C) Coupures reportées sur la structure secondaire prédite par Mfold.

reproductibles. La forme pleine longueur comporte la séquence du SINE avec la queue poly(A), mais aussi le terminateur apporté par la séquence où est inséré le SINE (figure 24.B). Dans le haut du gel présenté en figure 24.B on voit un “smear” dans les pistes 5 et 6 correspondant à un grand nombre de coupures par la RNase T2, ce qui signifie que la queue poly(A) reste simple brin. La queue poly(A) et le terminateur ne perturbent donc pas la structure de la séquence correspondant à la forme sc du SINE.

IV.2.2) SB2

En ce qui concerne le SINE B2 nous avons analysé par traitement enzymatique la structure en solution de ses formes cytoplasmique courte (sc, figure 25) et pleine longueur (fl, figure 26). Dans les deux cas, pour les 151 premiers nucléotides, on obtient un profil de coupures quasi identique cohérent avec le modèle de structure prédit par Mfold. Les nucléotides 4 à 75 se replient en tige-boucle. Les nucléotides 38 à 44 formant la boucle terminale sont coupés par les RNases T1 et T2. Des coupures par la RNase V1 sont réparties tout le long de l’hélice double brin. On note la présence d’une boucle interne révélée par des coupures par les RNase T1 et T2 au niveau des nucléotides 26, 29, 30 et 31. En ce qui concerne la deuxième tige-boucle, on note les coupures par les RNases T1 et T2 après les nucléotides 91 à 93 et 95 indiquant la boucle terminale et des coupures par la RNase V1 dans la base de l’hélice. On peut voir des coupures par la RNase T2 plus haut dans le gel, qui pourraient correspondre à la troisième tige-boucle, mais on ne peut les positionner exactement sur la séquence. Pour résoudre plus précisément la structure de la deux et de la troisième tige-boucle nous avons en projet de faire des digestions enzymatiques sur l’ARN marqué en 3’. Sur le haut du gel présentant le traitement enzymatique de l’ARN pleine longueur en figure 26.D on voit un “smear” correspondant à des coupures par la RNase T2, ce qui indique que la queue poly(A) reste simple brin.

IV.2.3) Mise en parallèle des résultats de structure obtenus pour les SINE SB1 et SB2

Nous avons analysé expérimentalement les structures secondaires de quatre ARN correspondant aux formes cytoplasmiques courtes et pleines longueurs des SINE SB1 et SB2 (figures 23 à 26). Les résultats obtenus sont pour tous cohérents avec les modèles prédits informatiquement avec le logiciel Mfold. Ce logiciel basé sur une approche thermodynamique semble donc être un bon outil pour étudier la structure secondaire des ARN SINE.

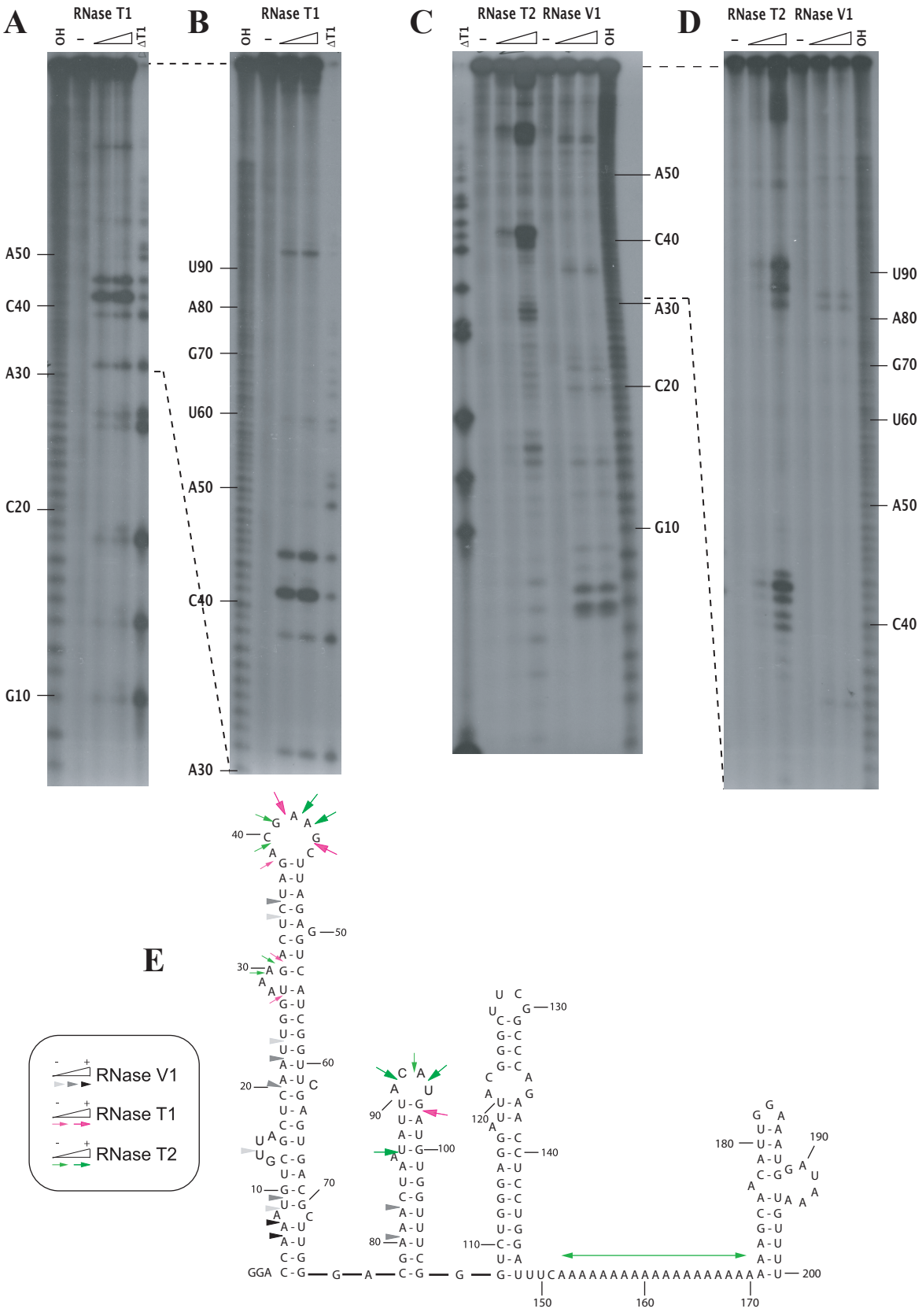


Figure 26. Cartographie en solution de la forme pleine longueur de l'ARN SINE SB2

Digestions enzymatiques de l'ARNfl marqué en 5' par les RNases T1 (A et B), T2 et V1 (C et D). ΔT1 : hydrolyse de l'ARN par la RNase T1 en conditions dénaturantes. OH : échelle obtenue par hydrolyse alcaline de l'ARN marqué. (-) : incubation de l'ARN dans les mêmes conditions mais sans traitement.

(E) Coupures reportées sur la structure secondaire prédite par Mfold.

Les expériences de cartographie en solution présentées ici ont été réalisées sur des ARN issus de séquences d'ADN choisies parce qu'elles correspondent à des loci fondateurs des éléments SINE étudiés. Un locus est considéré comme fondateur s'il est exprimé et peut permettre ou a permis de créer de nouvelles copies du SINE. Lors de la rétroposition, seule la séquence du SINE avec la queue poly(A) s'insère dans le génome. Or lorsqu'un élément SINE est exprimé, les transcrits primaires comportent la séquence du SINE avec la queue poly(A) mais aussi la séquence suivant au locus d'insertion jusqu'à un signal de terminaison de l'ARN polymérase III (Pol III). La séquence terminatrice apportée par la région 3' flanquante est un facteur intervenant dans la stabilité des transcrits. Pour les loci des SINE SB1 et SB2 utilisés dans cette étude, cette séquence est assez courte, elle ne déstabilise pas la structure de l'élément et laisse une séquence poly(A) simple brin, ce qui est considéré comme une caractéristique nécessaire pour qu'un locus soit fondateur.

Les SINE SB1 et SB2 sont tous deux dérivés d'ARN de transfert (ARNt). Les résultats obtenus montrent que les ARN de ces SINE ne se replient plus selon la structure secondaire caractéristique des ARNt. Bien que ces deux éléments n'aient pas d'homologies de séquences, ils produisent des ARN qui adoptent des structures secondaires très similaires.

Ce résultat expérimental associé à une étude bioinformatique de la structure secondaire de SINE dérivés d'ARNt provenant de divers eucaryotes a permis d'émettre l'hypothèse que la structure secondaire de ces ARN SINE aurait une histoire évolutive commune. Ces travaux sont l'objet de l'article 2.

IV.3. Article 2 : Common evolutionary trends for SINE RNA structures

Une étude bio-informatique permettant une analyse évolutive de la structure secondaire des ARN a été réalisée par Jean-Marc Deragon sur un grand nombre de séquences d'ARN SINE dérivés d'ARNt provenant de génomes eucaryotes aussi variés que les plantes, les poissons ou les mammifères.

Dans un premier temps, les structures secondaires des ARN SINE obtenues avec Mfold ont été comparées. Les différents éléments étudiés semblent adopter des profils de repliement similaires en une seule, deux ou trois tiges-boucles, comportant plusieurs

renflements et boucles internes, et même des embranchements secondaires s'ils sont de grande taille.

Une analyse cladistique a ensuite été réalisée par Feng Jie Sun et Gustavo Caetano-Anollés (Université de l'Illinois) pour établir un arbre reflétant les relations de parenté entre les différentes structures d'ARN observées. Vingt SINE représentatifs d'eucaryotes ont été choisis pour mener cette étude. L'analyse cladistique est basée sur la notion d'homologie. Elle établit des relations de parenté sur la base du partage de caractères. Le caractère désigne tout attribut utilisé pour reconnaître, décrire ou différencier les taxa. Dans notre cas, les caractères sont les caractéristiques structurales telles que les tiges-boucles, les renflements, les boucles internes. L'arbre retenu est le plus parcimonieux, c'est-à-dire celui qui requiert le plus petit nombre d'étape pour résoudre les relations entre les taxa. Cet arbre montre que les SINE évoluent en se complexifiant. Ainsi, exception faite des SINE du groupe I qui sont des cas particuliers, les SINE les moins évolués ont une structure simple, avec un segment apparenté à un ARNt, et ils évoluent par addition de nouvelles tiges-boucles et de structures interne des hélices. Cette étude a permis de mettre en évidence une histoire évolutive de la structure prenant en compte l'apparition des tiges-boucles, mais aussi l'ordre d'apparition de nouvelles structures et même l'évolution des appariements dans les différentes boucles. Bien que les SINE ne soient pas conservés au niveau de leur séquence primaire, notre étude suggère qu'ils sont sous pression de sélection pour adopter un schéma évolutif commun au niveau de leur structure secondaire ARN.

Opinion TRENDS in Genetics Vol.23 No.1

Full text provided by www.sciencedirect.com ScienceDirect

Common evolutionary trends for SINE RNA structures

Feng-Jie Sun[1], Sophie Fleurdépine[2], Cécile Bousquet-Antonelli[2], Gustavo Caetano-Anollés[1] and Jean-Marc Deragon[2,3]

[1] Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[2] CNRS UMR 6547 BIOMOVE and GDR 2157, Université Blaise Pascal Clermont-Ferrand II, 63177 Aubière Cedex, France
[3] CNRS UMR 5096 Université de Perpignan, 66860, Perpignan, France

Short interspersed elements (SINEs) and long interspersed elements (LINEs) are transposable elements in eukaryotic genomes that mobilize through an RNA intermediate. Understanding their evolution is important because of their impact on the host genome. Most eukaryotic SINEs are ancestrally related to tRNA genes, although the typical tRNA cloverleaf structure is not apparent for most SINE consensus RNAs. Using a cladistic method where RNA structural components were coded as polarized and ordered multistate characters, we showed that related structural motifs are present in most SINE RNAs from mammals, fishes and plants, suggesting common selective constraints imposed at the SINE RNA structural level. Based on these results, we propose a general multistep model for the evolution of tRNA-related SINEs in eukaryotes.

Introduction

Short interspersed elements (SINEs) and long interspersed elements (LINEs) are mobile retroelements that are widely distributed in eukaryotic genomes [1]. These dispersed repetitive elements have been the subject of intense study since their discovery in human and rodent genomes more than two decades ago. Until recently, research interest was centered on the retroposition mechanism, the mutagenic effects of the retroelements and their general impact on the structure and evolution of genomes. However, the recent discovery that human LINEs can modulate human gene transcription [2–4] and that mammalian SINE RNAs can in stress situations fulfill basic regulatory functions linked to transcription or translation (Box 1) has promoted the idea that SINEs and LINEs can be important contributors to the genetic control architecture of higher eukaryotes.

SINEs are 80–500 bp long and their genomic copy number usually ranges from a few hundred to more than a million in the case of the human Alu family [1,5,6]. SINEs are nonautonomous and must use the enzymatic machinery of an autonomous LINE partner for retroposition [7,8] (see Figure 1 for a general model of SINE and LINE retroposition). This SINE–LINE partnership is based on common 3′ regions or poly(A) tails. SINEs are ancestrally derived from tRNAs, or more exceptionally from 7SL or 5S RNAs [5,9]. They usually have a composite structure consisting of a 5′ tRNA-related segment followed by a tRNA-unrelated segment of variable size [6]; an exception is the recently described t-SINE superfamily composed exclusively of tRNA-related elements [10]. In all cases, SINEs harbor in their tRNA-related segment the internal promoter (composed of A and B boxes) recognized by the RNA polymerase III (Pol III) machinery [11]. Therefore, SINE-specific transcription is completely dependent on the Pol III transcription system.

Although most SINE families have a limited (species- or genus-specific) distribution, the central region of several vertebrate and invertebrate SINE families was found to be surprisingly similar. Based on the nature of the conserved region, these SINEs were grouped into three superfamilies called CORE-SINE [12], V-SINE [13] and AmnSINE [14]. CORE-SINE probably originated from a single SINE family present 550 million years ago (Mya) in the common ancestor of vertebrates and invertebrates that later spread and diversified in different taxa [12]. All CORE-SINE consensus sequences have similar 5′ and CORE domains but completely different 3′ regions, suggesting that CORE-SINE families are evolving by the shuffling of their 3′ regions, possibly to keep pace with variations in populations of active LINE partners [12]. V-SINEs are also widely spread in vertebrates and all contain a strongly conserved central motif [13]. Like CORE-SINEs, V-SINEs might also be around 550 million years old and are evolving essentially by the shuffling of 3′ regions. Recently, a third SINE superfamily with an Amniota-wide occurrence (AmnSINEs) was described [14]. A novel shared central domain and an unusual 5′ chimeric domain composed of 5S rRNA and tRNA-derived sequences characterize AmnSINEs. The strong conservation of the central domain of CORE-SINEs, V-SINEs and AmnSINEs over a long evolutionary period suggests that this region might have been subject to some form of positive selection [12–14] (Box 1).

SINE RNA secondary structure

Little is known about SINE RNA structure. SINEs derived from 7SL RNA (including Alu and B1) conserved part of the RNA folding pattern of the ancestral molecule [15,16]. However, the situation is less clear for tRNA-related SINEs. Only two tRNA-related SINE structures have been experimentally studied. One is the BC1 RNA, a founder

Corresponding author: Deragon, J-M. (j-marc.deragon@univ-bpclermont.fr).
Available online 28 November 2006.

Opinion TRENDS in Genetics Vol.23 No.1 27

Box 1. The impact of SINEs on gene expression

SINEs probably had a significant impact on the evolution of eukaryotic genomes [1]. Although genomic integration of SINEs can be deleterious, it can also stimulate genome reorganization (e.g. by genomic duplication resulting from SINE–SINE unequal crossover), promote the emergence of new transcriptional regulation pathways (by associating genes with new positive or negative *cis*-elements, by transcriptional interference, or by modifying the methylation status of a promoter), and generate variability at the posttranscriptional level (by introducing new splicing, polyadenylation and editing signals in pre-mRNAs). In addition, several experimental results suggest that SINE Pol III-specific transcripts could, in stress situations, have a role as noncoding riboregulators [43–46]. In mammals and insects, the induction of different cellular stresses (such as heat shock or viral infection) leads to a specific increase in SINE RNA levels [1]. In mammals, SINE RNAs have been shown to regulate translation by interacting with protein kinase R (PKR) (an enzyme that phosphorylates eukaryotic initiation factor (eIF)2α and downregulates translation [46,47]), or by a PKR-independent mechanism [45,48]; or to regulate transcription by binding to the RNA polymerase II complex [43,44]. In this context, mammalian SINEs can be considered 'stress responsive-genes' capable of modulating the level of mRNAs and proteins in different stress situations. SINEs in mammals can also regulate cellular function in 'normal' (unstressed) situations. BC1 (in rodent) and BC200 (in human) are two SINE-related RNAs specifically produced in brain tissues. These RNAs can associate with the fragile X mental retardation protein (FRMP) and could work as an adaptor between this protein and its target mRNAs, leading to a selective translational repression [40,41]. Also BC1 and BC200 could have a more general effect as suppressors of translation in neurons and other cell types [38,39].

molecule responsible for the amplification of identifier (ID) elements in rodents [17]. Although, the tRNA-related segment of BC1 RNA is nearly 80% identical to its ancestral sequence ($tRNA^{Ala}$), it does not fold into the tRNA-like cloverleaf structure, but instead assumes a single extended stem-loop structure [17]. By contrast, salmon SmaI SINE RNA folds into three separate domains: a 5′ cloverleaf structure, a linker and a 3′ extended stem-loop [18]. Moreover, the 5′ domain of SmaI can be modified by HeLa pseudouridylate synthetase(s) at positions identical to those found in the ancestral sequence ($tRNA_1^{Lys}$) [19] and is proposed to form an L-shaped tertiary structure similar to tRNAs [18], suggesting a good conservation of the RNA folding properties of the ancestral molecule. Recently, we probed the secondary structure of two tRNA-related plant SINE RNAs from *Brassica oleracea* and *Arabidopsis thaliana* (SB1 and SB2 [20]) and found that these molecules did not conserve the folding patterns of ancestral tRNAs but,

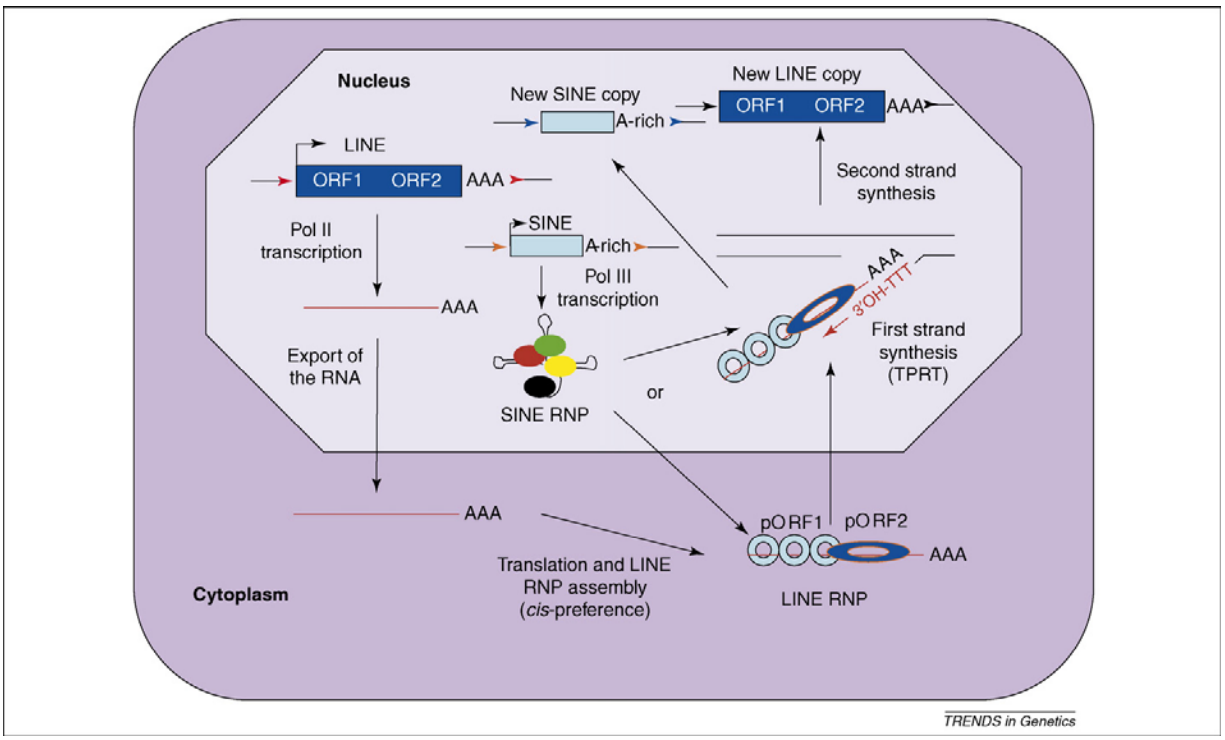


Figure 1. A general model of SINE and LINE retroposition (for a general review, see Ref. [33]). LINE elements are transcribed from an internal RNA polymerase II (Pol II) promoter to generate full-length sense-strand LINE RNAs that subsequently are exported to the cytoplasm. Following translation, pORF1 and pORF2 co-assemble with their encoding RNA (by a process named *cis*-preference) to form the LINE RNP. This complex is imported into the nucleus (or enters during mitosis) and engages in the TPRT process that leads to the first-strand cDNA synthesis. Subsequent steps such as second-strand cleavage, second-strand cDNA synthesis and ligation of the resultant cDNA to genomic DNA have yet to be explained. SINE elements are transcribed from an internal Pol III promoter to generate a full-length SINE RNA that is probably guided through several post-transcriptional modifications [49] before or during RNP assembly. We suggest that, to engage in TPRT and to retropose, the SINE and LINE RNPs must first interact, either in the cytoplasm or in the nucleus ('or'). In addition, a common 3′ region or poly(A) tail between SINE and LINE RNAs is needed for the SINE RNA to capture the LINE machinery [7,50]. As for LINEs, the subsequent steps leading to SINE integration are unknown.

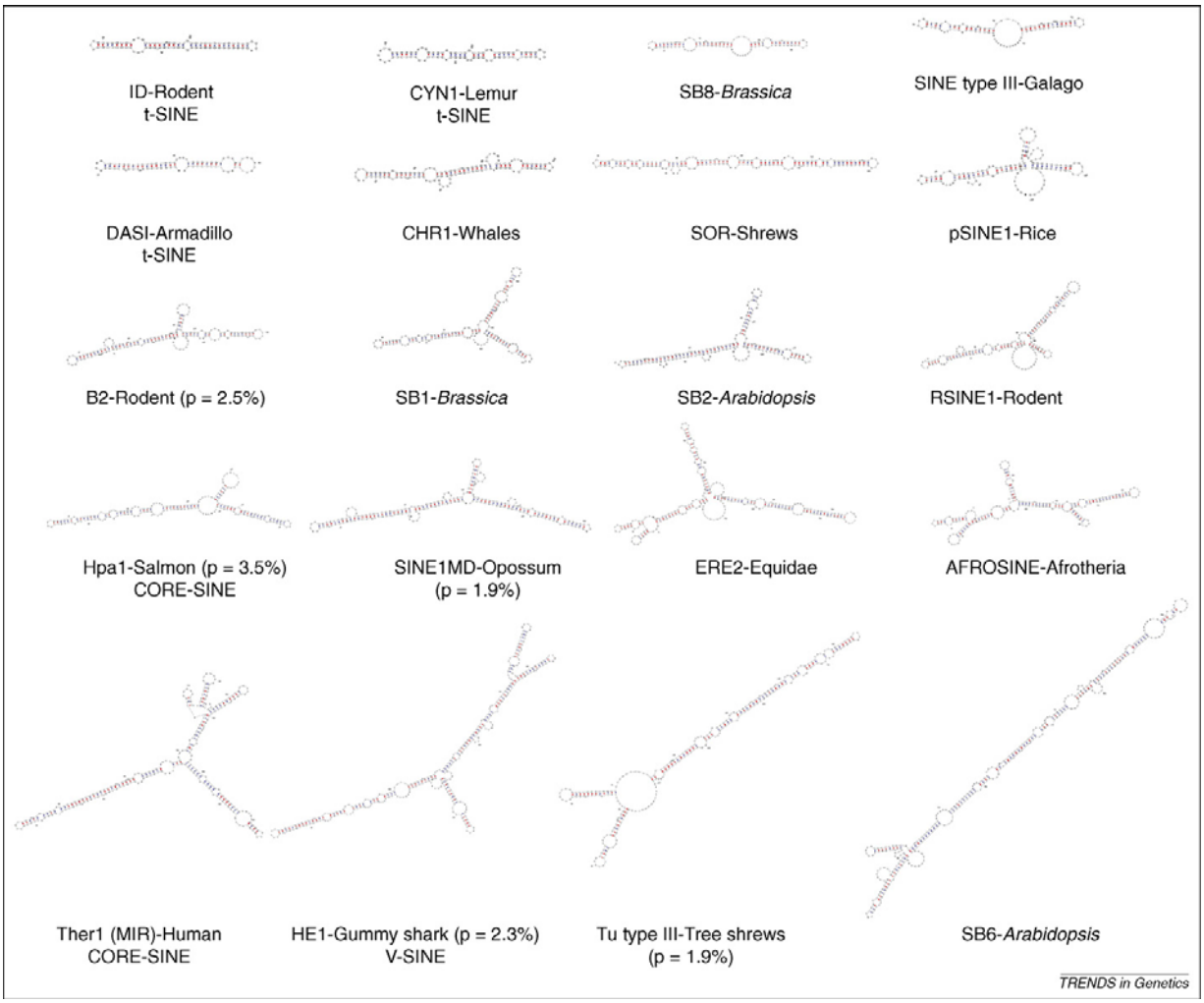


Figure 2. Predicted RNA secondary structures of 20 representative SINE consensus sequences. The structures were obtained using Mfold version 3.2 (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/old.cgi) at 37 °C for all mammalian sequences, and version 2.3 at 25 °C for all plant and fish sequences. In most cases, the structure given is the most optimal (i.e. with the lowest free energy). When this is not the case and a suboptimal structure was selected, the percent suboptimality, p, is given and represents the increase in free energy compared with the most optimal structure [51,52]. The origin and association of a SINE to a described superfamily (t-SINE, CORE-SINE or V-SINE) is indicated.

although unrelated at the primary sequence level, they still presented similar RNA structures (see Figure 2 for a representation of the SB1 and SB2 RNA structures). This prompted us to predict, using a bioinformatic approach, the RNA secondary structure of the 15 different SINEs corresponding to the known SINE families in *Brassica* and *Arabidopsis*. Although generally the best accuracy for RNA structure predictions can be achieved by using comparative methods [21,22], this is often not the case with swiftly evolving RNA sequences (such as SINEs) because imperfect alignment can destroy any covariation signal present in this type of data [21]. We therefore decided to use the Mfold RNA free-energy minimization algorithm on the 15 different SINE consensus sequences, because these reconstituted molecules best approximate the biologically active (i.e. founder) copy of each SINE family [23]. For SB1 and SB2, the minimal free-energy structures were fully compatible with our experimental data (not shown), indicating that, at least in these two cases, Mfold efficiently predicted the correct RNA structure. Although the 15 consensus sequences tested have no significant primary sequence homology (apart from the short A and B motifs of the internal Pol III promoter) and are of different lengths (from 99 to 332 nt [20]), similar folding patterns were observed (Figure S1 in the supplementary material online).

Common evolutionary trends for eukaryote SINE RNA structures

We next tested if similar RNA folding patterns could be observed for tRNA-related SINE families outside the Brassicaceae. More than 50 tRNA-derived SINE families have been described in mammals, fishes and plants [1]. Consensus sequences from many of these SINE families are not robust (they were generated from the alignment of a small number of copies) but were nevertheless used to

predict the corresponding RNA structures. Surprisingly, the diversity of structures observed for most of these SINE families is similar to that observed for SINEs of Brassicaceae and shows a clear conservation of patterns. Consensus RNAs from 20 representative families are presented in Figure 2. RNAs from the shorter SINE elements composed only of a tRNA-related domain (such as ID-Rodent, CYN1-Lemur, and DASI-Armadillo, all members of the t-SINE superfamily [10,24,25]), form either a single (ID and CYN1) or two (DASI) short extended stem-loop structures containing several bulges and internal loops (Figure 2). The single other t-SINE, Vic1 from camel [26], is apparently an exception to this rule because it exhibits a tRNA-like folding pattern. RNAs from SINEs with short tRNA-unrelated domains fused to tRNA-related domains (SB8-*Brassica*, SINE type III-Galago and CHR1-Whales) also form two short stem-loops. For SINEs with longer tRNA-unrelated segments, their RNAs either fold as two extended stem-loops (SOR-Shrews) or adopt a structure comprising three stem-loops, which is the most common structure for tRNA-related SINEs (Figure 2). The length of the three stem-loops can vary (in correlation with the increase in size of the tRNA-unrelated segment), and in some cases multifurcation (branching) in one or two stem-loops can be observed (ERE2-Equidae, AFROSINE-Afrotheria, Ther1-Human, HE1-Gummy shark). The common structures could be observed for many CORE-SINEs and V-SINEs (Figure S2 and Figure S3 in the supplementary material online). The observation that consensus RNAs for some of the V-SINE families could not be folded in the characteristic triple stem-loop structure could be the result of the extreme weakness of some consensus sequences of this superfamily, which are usually derived from the alignment of only two to six independent copies [13]. Alternatively, true exceptions might exist inside the V-SINE superfamily.

Use of a cladistic method to compare the different SINE RNA structures

To test more rigorously the relatedness of the different SINE RNA structures, we used a cladistic approach that recovers phylogenetic information directly from the structure of nucleic acid molecules and can be applied to the study of molecules of highly divergent lineage [27,28]. Maximum parsimony analyses of 37 substructural components derived from 20 representative eukaryotic SINE RNAs resulted in two minimal-length trees (see the supplementary material online, including Table S1 and Table S2, for methods and data), the strict consensus of which is presented in Figure 3. Although bootstrap support (BS) estimates were generally low (with eight out of 17 values exceeding 50%), decay indices were generally high. Among the 17 decay indices, 11 are from three to five steps, and six

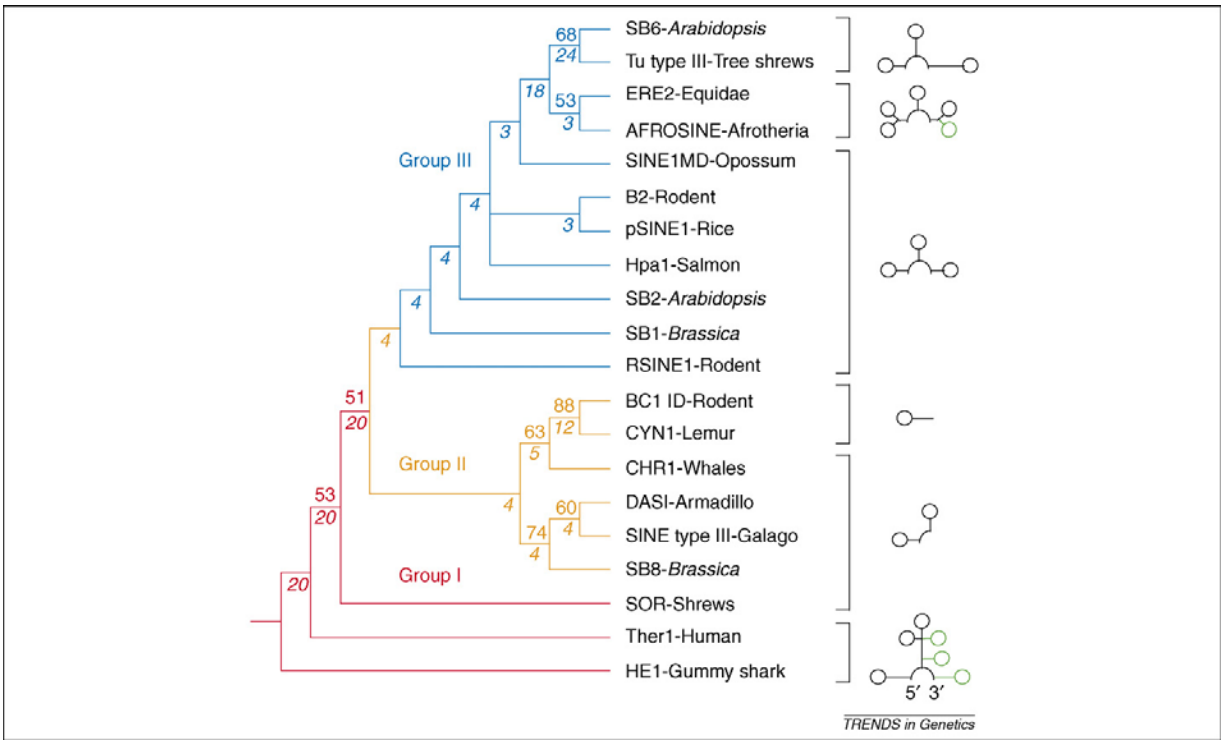


Figure 3. Reconstruction of a eukaryotic phylogenetic tree of SINE structures. The tree represents a strict consensus of two optimal trees of 1512 steps each derived from equally weighted maximum parsimony analyses of 37 substructures from 20 eukaryotic SINE RNAs. Consistency indices (CI) of 0.443 and 0.439, with and without uninformative characters, respectively, and a retention index (RI) of 0.569, showed measures of homology and character fit. There was also significant phylogenetic signal in the data ($g_1 = -0.577$, where g_1 is skewness). Numbers above branches indicate bootstrap estimates for 1000 replicate analyses; values <50% are not indicated. Numbers in italics below branches are decay indices. Overall structures are described with lines (stems) and circles (hairpin loops), with varying substructures colored in green. SINE groups I, II and III are indicated with different colors on the tree.

are from 12 to 24 steps. A few branches having low BS values are supported by high decay indices, whereas some other branches having >50% BS values are supported by relatively low decay indices. The tree is well resolved and shows three clear groups. Group I is basal and contains three SINE structures exhibiting either stem multifurcation (Ther1-Human and HE1-Gummy shark) or longer tRNA-unrelated segments (SOR-Shrews). These basal groups are followed by group II, a clade composed of six SINEs each containing a tRNA-related domain (ID-Rodent, CYN1-Lemur and DASI-Armadillo), sometimes associated with short tRNA-unrelated segments (SINE type III-Galago, SB8-*Brassica*, and CHR1-Whales). Two of these SINEs (ID-Rodent and CYN1-Lemur) form a single stem-loop structure, and the other four form double stem-loop molecules. Finally, group III contains the remaining SINEs, all with structures characterized by three stem-loops.

Results suggest a common trend of structural evolution in which molecules with either one or two stem-loops evolve by addition of new stem-loops through multiloop formation or terminal branching. Consequently, we can envision that SINE RNAs started as simple structures with a single tRNA-related segment and became more complex in some lineages or stayed simple in others. It is noteworthy that SINE length associates with increases in structural complexity, suggesting SINE sequences increase in length during evolution.

Two exceptions to these common trends are Ther1-Human (a representative of the CORE-SINEs) and HE1-Gummy shark (a representative of the V-SINEs); these have structures with, respectively, two and three stem-loops, and multifurcation, and both appear at the base of the tree. This basal placement could imply the third stem-loop structure is ancestral, was lost in SOR-Shrews and group II, but was maintained in group III SINEs. However, both HE1-Shark and Ther1-Human are long (315 and 258 nt, respectively) and extensively multifurcated. It is therefore possible that the conserved central domain of most CORE-SINEs and V-SINEs facilitates acquisition and exchange of tRNA-unrelated segments (as suggested in refs [12,13]), speeding the generation of complex structures in these molecular lineages.

We also reconstructed trees using a distance-based method (Figure S4 and methods in the supplementary material online). The topology of the neighbor-joining tree is largely congruent to the topologies of the maximum parsimony trees. The positions of only three SINE RNAs in group III (pSINE1-Rice, SB1-*Brassica* and RSINE1 Rodent) were incongruent; pSINE1-Rice is sister to group II RNAs, and SB1-*Brassica* and RSINE1 Rodent are basal and follow group I SINEs. These positions matched regions in the maximum parsimony trees with low BS values. The common trend of structural evolution inferred on the maximum parsimony trees was maintained on the neighbor-joining tree.

A general model for tRNA-related SINE evolution

To define a general model of structural evolution for tRNA-related SINEs, we analyzed the 37 substructures derived from 20 representative eukaryotic SINE RNAs (Figure 3) using a novel phylogenetic approach that generates trees of substructural components that stabilize or destabilize molecules (F-J. Sun and G. Caetano-Anollés, unpublished). This involves defining new kinds of taxa (substructures) and characters (molecules), and uses a criterion of primary homology pertaining to substructures based on molecular lineages (see the supplementary material online). Phylogenies describing the evolution of substructures established clearly that the SINE RNA molecule has an origin in a GC-rich tRNA-related segment (Figure 4). A timeline for the evolution of SINE molecules can be inferred directly from the phylogenetic trees of

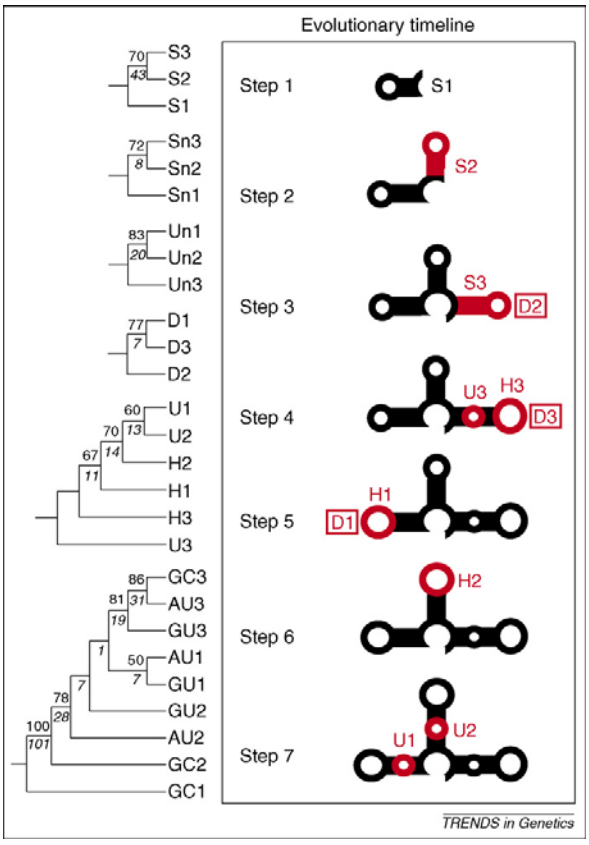


Figure 4. Phylogenetic trees of tRNA-related SINE RNA substructures and inference of a model of structural evolution. Single optimal trees describing the evolution of stems (S1–S3; 1647 steps; CI = 0.902, RI = 0.509; $g_1 = 0.385$), stem components (Sn1–Sn3; 1608 steps; CI = 0.984, RI = 0.500; $g_1 = 0.135$), destabilizing hairpins (H1–H3) and unpaired segments (U1–U3) (537 steps; CI = 0.756, RI = 0.463; $g_1 = 0.246$), destabilizing unpaired components (Un1–Un3; 210 steps; CI = 0.881, RI = 0.638; $g_1 = 0.092$), degree of multiloop (D1–D3; 1603 steps; CI = 0.998, RI = 0.7508; $g_1 = 0.377$), and stabilizing base pair components (GU1–GU3, GC1–GC3, and AU1–AU3; 1849 steps; CI = 0.860, RI = 0.640; $g_1 = -1.199$) were derived from equally weighted maximum parsimony analyses of 20 eukaryotic SINE RNA structures using branch-and-bound searches. Numbers above branches indicate bootstrap estimates for 1000 replicate analyses; values <50% are not indicated. Numbers in italics below branches are decay indices. In step 1 of the inferred evolutionary model, SINEs originate in a tRNA-related domain (S1) that increases in size. During step 2, molecules add an additional stem-loop structure (S2), representing either tRNA-related or tRNA-unrelated segments, or both. In step 3, a tRNA-unrelated portion (S3) is incorporated to form the structure with three stem-loops that is characteristic of most eukaryotic SINEs. In steps 4–7, hairpin loops (H) increase in size and bulges and internal loops (U) are added to stems in defined order. Stems also multifurcate (D) by adding new hairpin loops. This leads to the formation of complex and high copy number SINEs. Abbreviations: CI, consistency index; RI, retention index.

Opinion TRENDS in Genetics Vol.23 No.1 31

Box 2. CYN and DAS families as examples of SINE evolution

The oldest CYN and DAS SINEs, found in the genomes of lemur and armadillo, respectively, consist exclusively of monomeric elements containing only the tRNA-related domain [10,24,25]. These old elements are present at low copy number and harbor one or two simple stem-loops. CYN sequences became more complex following dimerization (CYN-ALPHA and CYN-BETA) and trimerization (CYN-GAMMA) of the tRNA-related portion (Figure I) [10,25]. This process also led to more complex RNA structures, changing from one extended stem-loop to two and then three (with some multifurcations). The trimers constitute the youngest family with the highest copy number [10,25], suggesting that the presence of a more complex RNA structure was associated with an increase in amplification success. Elements from the DAS family [24] also increased in complexity, in part by dimerization of the tRNA-related domain, RNA structures changing from two extended stem-loops (DASI) to three (DASIIa) (Figure I). However, in this case different tRNA-unrelated portions were fused to the tRNA-related portion of DASI to generate DASIII2b and DASIII3b families. This led to the transition of RNA structures from two to three extended stem-loops with some multifurcations, apparently associated with higher amplification success [24]. Therefore, at least in these two cases, the increase in size of the element and the acquisition of a characteristic RNA structure containing three stem-loops was associated with a better retroposition efficiency. The addition or shuffling of a new (complete or partial) tRNA-unrelated segment to an existing tRNA-related domain (as seen for the DASI to DASIII2b and DASIII3b transition) is probably a general mechanism by which SINEs change and become more complex and was suggested as the major mechanism in CORE-SINE [12], V-SINE [13], AmnSINE [14] and Brassicaceae SINE [20] evolution.

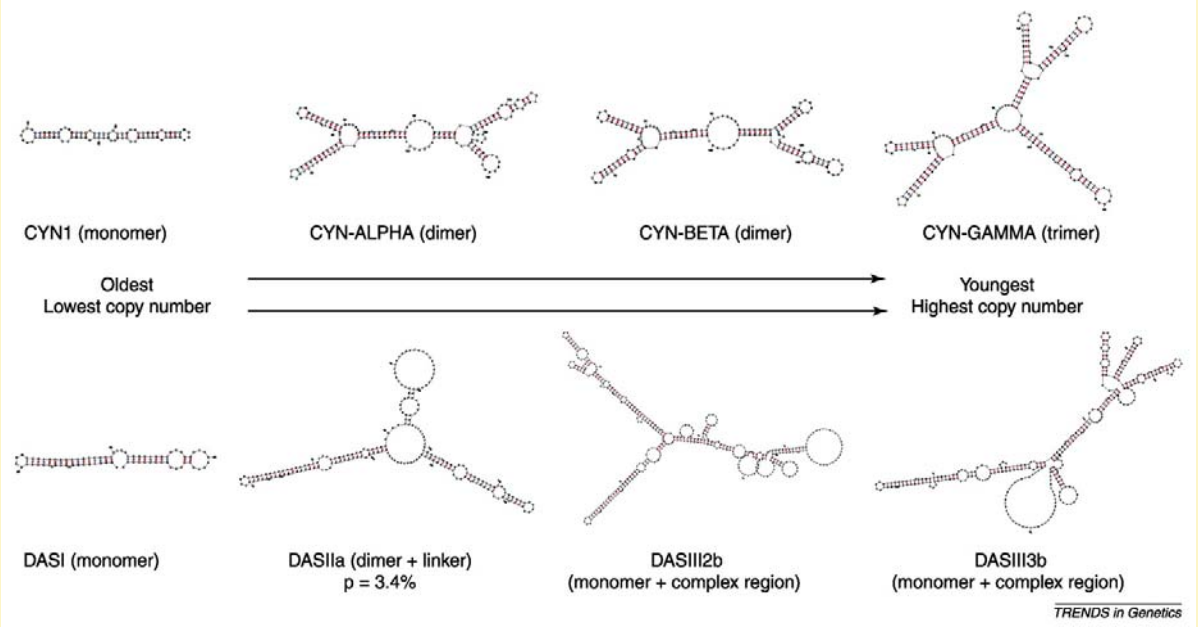


Figure I. Proposed evolution of SINEs belonging to the CYN and DAS families. Abbreviation: p, percent suboptimality.

substructures. This timeline represents a global model of evolution of SINE RNA structure that suggests modern molecules resulted from gradual addition of structural components to a tRNA-related founder. Overall, the proposed model shows that the SINEs could have evolved and amplified from a single stem-loop to two and three, as inferred from cladistic analyses. To accomplish this process, SINEs either amplify the tRNA-related segment or add short or long tRNA-unrelated segments to their molecules. A good example supporting the general model of tRNA-related SINE evolution is illustrated by the diversification process of CYN and DAS SINE families (Box 2).

Concluding remarks

The 7SL-related SINEs probably benefit from the binding of some of the signal recognition particle (SRP) proteins (SRP9 and SRP14 are core members of 7SL-related SINE ribonucleoproteins (RNPs) [29]), and are therefore under selective pressure to keep at least part of the folding pattern of the ancestral molecule. By contrast, the majority of SINEs derived from tRNAs do not fold as their ancestral molecule. This might be essential to avoid recognition by tRNA-processing enzymes, in particular the 3′ endonuclease (RNase Z) [30], which would cleave the 3′ end of the SINE RNA. We show here that most tRNA-related SINEs have the same general evolutionary history resulting in similar RNA structures. This suggests that tRNA-related SINE RNAs must interact with yet unknown conserved proteins producing similar constraints at the RNA level in most eukaryotes. The only proteins known to interact with tRNA-related SINEs are poly(A)-binding proteins and the La autoantigen protein [31,32]. Because these proteins bind to the unstructured 3′ A-rich or the terminal T-rich regions of SINE RNAs, it is unlikely that they are directly responsible for the selection pressure imposed on the RNA structure. Characterization of SINE RNPs is needed to reveal the nature of the conserved core proteins binding to tRNA-related SINE RNAs.

We suggest that the selection pressure imposed on the SINE RNA structure has more to do with its putative function as riboregulator (see Box 1) than with its capacity to parasitize the LINE retroposition machinery. In the SINE–LINE partnership, it is likely that SINE and LINE RNPs must interact at some point to enable SINE retroposition (see the model in Figure 1). Because LINEs are evolving fast with little conservation (across species) of key RNP proteins (such as the open reading frame (*ORF*)*1* product) [33], we should expect a concerted evolution of SINE RNA-binding proteins and consequently of SINE RNA structures, resulting in greatly divergent SINE RNA structures across species. Instead, we propose that the conserved trend we observed can best be explained by a selection pressure to keep basic regulatory functions of the SINE RNAs and that their capacity to interact with the LINE machinery and to retropose is only a side effect that might or might not happen. If true, we should be able to find in genomes Pol III-transcribed riboregulators with characteristics of SINEs but present at a single (or low) copy. The adenovirus VA1 and Epstein–Barr EBER-1 and EBER-2 viral loci could represent examples of 'single copy SINE-like elements' because the corresponding RNAs are produced by the Pol III machinery, adopt a structure similar to SINEs and are riboregulators of key cellular functions [34,35]. In conclusion, we propose that SINEs are members of a large group of Pol III-transcribed riboregulators (which include the cellular 7SL [36], 7SK [37], BC1 and BC200 [38–41], vault [42], and the viral VA1 [34], EBER-1 and EBER-2 [35] RNAs) and that their amplification in a large family of retroposons (i.e. retropseudogenes) is linked to a predisposed but fortuitous capacity to parasitize the LINE retroposition machinery.

Acknowledgements

This research was supported by CNRS, the European Commission (RIBOREG STREP program) (to J.M. Deragon) and NSF Project MCB-0343126 (to G. Caetano-Anollés). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2006.11.005.

References

1 Kramerov, D.A. and Vassetzky, N.S. (2005) Short retroposons in eukaryotic genomes. *Int. Rev. Cytol.* 247, 165–221
2 Perepelitsa-Belancio, V. and Deininger, P. (2003) RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* 35, 363–366
3 Han, J.S. *et al.* (2004) Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* 429, 268–274
4 Muotri, A.R. *et al.* (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910
5 Jurka, J. (1995) Origin and evolution of Alu repetitive elements. In *The Impact of Short Interspersed Elements (SINEs) on the Host Genome* (Maraia, R., ed.), pp. 25–41, R.G. Landes, Springer
6 Okada, N. and Ohshima, K. (1995) Evolution of tRNA-derived SINEs.. In *The Impact of Short Interspersed Elements (SINEs) on the Host Genome* (Maraia, R., ed.), pp. 61–80, R.G. Landes, Springer
7 Dewannieux, M. *et al.* (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35, 41–48
8 Kajikawa, M. and Okada, N. (2002) LINEs mobilize SINEs in the eel through a shared 3′ sequence. *Cell* 111, 433–444
9 Kapitonov, V.V. and Jurka, J. (2003) A novel class of SINE elements derived from 5S rRNA. *Mol. Biol. Evol.* 20, 694–702
10 Piskurek, O. *et al.* (2003) Unique mammalian tRNA-derived repetitive elements in dermopterans: the t-SINE family and its retrotransposition through multiple sources. *Mol. Biol. Evol.* 20, 1659–1668
11 Arnaud, P. *et al.* (2001) Analysis of the SINE S1 pol III promoter from *Brassica*; impact of methylation and influence of external sequences. *Plant J.* 26, 295–305
12 Gilbert, N. and Labuda, D. (2000) Evolutionary inventions and continuity of CORE-SINEs in mammals. *J. Mol. Biol.* 298, 365–377
13 Ogiwara, I. *et al.* (2002) V-SINEs: A new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome Res.* 12, 316–324
14 Nishihara, H. *et al.* (2006) Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* 16, 864–874
15 Labuda, D. and Zietkiewicz, E. (1994) Evolution of secondary structure in the family of 7SL-like RNAs. *J. Mol. Evol.* 39, 506–518
16 Sinnett, D. *et al.* (1991) Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. *J. Biol. Chem.* 266, 8675–8678
17 Rozhdestvensky, T.S. *et al.* (2001) Neuronal BC1 RNA structure: evolutionary conversion of a tRNA(Ala) domain into an extended stem-loop structure. *RNA* 7, 722–730
18 Kawagoe-Takaki, H. *et al.* (2006) Probing the secondary structure of salmon SmaI SINE RNA. *Gene* 365, 67–73
19 Matsumoto, K. *et al.* (1986) Gene for lysine tRNA1 may be a progenitor of the highly repetitive and transcribable sequences present in the salmon genome. *Proc. Natl. Acad. Sci. U. S. A.* 83, 3156–3160
20 Deragon, J.M. and Zhang, X. Short interspersed elements (sines) in plants: origin, classification and use as phylogenetic markers. *Syst. Biol.* (in press)
21 Gardner, P.P. and Giegerich, R. (2004) A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*,5, 140, DOI: 10.1186/1471-2105-5-140 (http://www.biomedcentral.com)
22 Reeder, J. *et al.* (2006) Beyond Mfold: recent advances in RNA bioinformatics. *J. Biotechnol.* 124, 41–55
23 Jurka, J. (1998) Repeats in genomic DNA: mining and meaning. *Curr. Opin. Struct. Biol.* 8, 333–337
24 Churakov, G. *et al.* (2005) A novel abundant family of retroposed elements (DAS-SINEs) in the nine-banded armadillo (*Dasypus novemcinctus*). *Mol. Biol. Evol.* 22, 886–893
25 Schmitz, J. and Zischler, H. (2003) A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates. *Mol. Phylogenet. Evol.* 28, 341–349
26 Lin, Z. *et al.* (2001) Characterization of a SINE species from vicuna and its distribution in animal species including the family Camelidae. *Mamm. Genome* 12, 305–308
27 Caetano-Anollés, G. (2002) Evolved RNA secondary structure and the rooting of the universal tree of life. *J. Mol. Evol.* 54, 333–345
28 Caetano-Anollés, G. (2002) Tracing the evolution of RNA structure in ribosomes. *Nucleic Acids Res.* 30, 2575–2587
29 Maraia, R. and Sarrowa, J. (1995) Alu-family SINE RNA: interacting proteins and pathways of expression. In *The Impact of Short Interspersed Elements (SINEs) on the Host Genome* (Maraia, R., ed.), pp. 163–196, R.G. Landes, Springer
30 Morl, M. and Marchfelder, A. (2001) The final cut: the importance of tRNA 3′-end processing. *EMBO Rep.* 2, 17–20
31 Muddashetty, R.S. *et al.* (2002) Poly(A)-binding protein is associated with neural BC1 and BC200 ribonucleoprotein particles. *J. Mol. Biol.* 321, 433–445
32 West, N. *et al.* (2002) Shared protein components of SINE RNPs. *J. Mol. Biol.* 321, 423–432
33 Moran, J. and Gilbert, N. (2002) Mammalian LINE-1 retrotransposons and related elements. In *Mobile DNA II* (Craig, N. *et al.*, eds), pp. 836–869, ASM Press
34 Maran, A. and Mathews, M.B. (1988) Characterization of the double-stranded RNA implicated in the inhibition of protein synthesis in cells

infected with a mutant adenovirus defective for VA RNA. *Virology* 164, 106–113
35 Kitagawa, N. *et al.* (2000) Epstein–Barr virus-encoded poly(A)$^-$ RNA supports Burkitt's lymphoma growth through interleukin-10 induction. *EMBO J.* 19, 6742–6750
36 Doudna, J.A. and Batey, R.T. (2004) Structural insights into the signal recognition particle. *Annu. Rev. Biochem.* 73, 539–557
37 Michels, A.A. *et al.* (2004) Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J.* 23, 2608–2619
38 Kondrashov, A.V. *et al.* (2005) Inhibitory effect of naked neural BC1 RNA or BC200 RNA on eukaryotic *in vitro* translation systems is reversed by poly(A)-binding protein (PABP). *J. Mol. Biol.* 353, 88–103
39 Wang, H. *et al.* (2005) Dendritic BC1 RNA in translational control mechanisms. *J. Cell Biol.* 171, 811–821
40 Zalfa, F. *et al.* (2005) Fragile X mental retardation protein (FMRP) binds specifically to the brain cytoplasmic RNAs BC1/BC200 via a novel RNA-binding motif. *J. Biol. Chem.* 280, 33403–33410
41 Zalfa, F. *et al.* (2003) The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* 112, 317–327
42 van Zon, A. *et al.* (2001) Multiple human vault RNAs. Expression and association with the vault complex. *J. Biol. Chem.* 276, 37715–37721
43 Allen, T.A. *et al.* (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol. Biol.* 11, 816–821
44 Espinoza, C. *et al.* (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat. Struct. Mol. Biol.* 11, 822–829
45 Rubin, C.M. *et al.* (2002) Selective stimulation of translational expression by Alu RNA. *Nucleic Acids Res.* 30, 3253–3261
46 Schmid, C.W. (1998) Does SINE evolution preclude Alu function? *Nucleic Acids Res.* 26, 4541–4550
47 Williams, B.R. (1999) PKR; a sentinel kinase for cellular stress. *Oncogene* 18, 6112–6120
48 Hasler, J. and Strub, K. (2006) Alu RNP and Alu RNA regulate translation initiation *in vitro*. *Nucleic Acids Res.* 34, 2374–2385
49 Pelissier, T. *et al.* (2004) Synthesis and processing of tRNA-related SINE transcripts in *Arabidopsis thaliana*. *Nucleic Acids Res.* 32, 3957–3966
50 Ohshima, K. and Okada, N. (2005) SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet. Genome Res.* 110, 475–490
51 Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science* 244, 48–52
52 Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415

Common evolutionary trends for SINE RNA structures

Feng-Jie Sun[1], Sophie Fleurdépine[2], Cécile Bousquet-Antonelli[2], Gustavo Caetano-Anollés[1] and Jean-Marc Deragon[2,3]

[1]Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[2]CNRS UMR 6547 BIOMOVE and GDR 2157, Université Blaise Pascal Clermont-Ferrand II, 63177 Aubière Cedex, France
[3]CNRS UMR 5096 Université de Perpignan, 66860, Perpignan, France
Corresponding author: Deragon, J-M. (j-marc.deragon@univ-bpclermont.fr).

Here, we reconstruct evolutionary history directly from the structure of RNA using cladistic principles and considerations in statistical mechanics (Caetano-Anollés, 2001; 2002a,b; 2005). Structural features in tRNA-related SINE RNA molecules are treated as ordered multi-state phylogenetic characters and the transformation of one character state to another polarized by invoking an evolutionary tendency towards molecular order. This award winning approach 'embeds structure and function directly into phylogenetic analysis' (Pollock, 2003). We also reconstruct diversification histories of molecular repertoires using a novel extension of our cladistic method (Sun and Caetano-Anollés, unpublished). The method generates phylogenetic trees of substructural components, and these phylogenies are used to define a structural origin and a general model of SINE RNA evolution.

Character coding

The folded secondary structure of each of 20 eukaryotic SINE RNAs (see Figure 2 in the main text) was decomposed into three possible stem-loop structural components. Attributes describing structural features in these stem-loop structures were then coded as phylogenetic characters. These attributes define the 'shape' (geometry) of molecules analyzed. A total of 37 substructural features were identified and scored from the sequences and secondary structures following the direction from 5′ to 3′ ends (Table S1). Coded character states were generally based on the length (the total number of bases or base pairs) and number of these substructures, and are represented in alphanumeric format using numbers 0 to 9, letters A to Z and a to z, and symbols / and \. Missing structures were given the minimum state (0). Each stem (helical region) was scored as two complementary sequences (5′ and 3′ sides) for unpaired areas. A data matrix (SINEs × substructures) was generated (Table S2) and encoded in the NEXUS format. Phylogenetic relationships between SINE RNA structures were inferred directly on the most parsimonious trees (see Figure 3 in the main text). Partitioned data matrices (substructures × SINEs) were also generated using characters collected either from the substructures that help stabilize the RNA molecules (e.g., stems and base pairings) or from the substructures that help destabilize the RNA molecules (e.g., bulges, internal loops, and hairpin loops). Phylogenetic relationships among substructures of a same kind were inferred directly on the most parsimonious trees (see Figure 4 in the main text).

Character argumentation

Structural features were treated as linearly ordered and polarized characters by applying an evolutionary tendency towards molecular order. The validity of assumptions related to character argumentation has been discussed in detail elsewhere (Caetano-Anollés, 2001; 2002a,b; 2005). The polarization was determined by fixing the direction of character state change using a transformation sequence that distinguishes ancestral states as those that are more thermodynamically stable. The maximum character states were used as the ancestral states for substructures stabilizing the RNA molecules, while the minimum states (0) are given as the ancestral states for substructures destabilizing the RNA molecules. Hypothetical ancestors were included in the searches for the most parsimonious trees, which were rooted at the point where the ancestors connect to the trees.

Table S1. Substructural features used in the phylogenetic analyses of 20 SINE RNA molecules. Stems are separated by internal loops. Unpaired regions may include bulges and internal loops. Features 2-12, 14-24, and 26-36 are scored from the first, second, and third stem-loop structures, respectively, from 5′ to 3′ ends.

1. Length of unpaired region at the 5′ end
2. Number of stems
3. Length of stems
4. Number of unpaired regions
5. Length of unpaired regions
6. Length of hairpin loops
7. Number of GC pairs
8. Number of AU pairs
9. Number of GU pairs
10. Presence or absence of multiloop
11. Degree of multiloop
12. Length of multiloop
13. Length of unpaired region between the first and second stem-loops
14. Number of stems
15. Length of stems
16. Number of unpaired regions
17. Length of unpaired regions
18. Length of hairpin loops
19. Number of GC pairs
20. Number of AU pairs
21. Number of GU pairs
22. Presence or absence of multiloop
23. Degree of multiloop
24. Length of multiloop
25. Length of unpaired region between the second and third stem-loops
26. Number of stems
27. Length of stems
28. Number of unpaired regions
29. Length of unpaired regions
30. Length of hairpin loops
31. Number of GC pairs
32. Number of AU pairs
33. Number of GU pairs
34. Presence or absence of multiloop
35. Degree of multiloop
36. Length of multiloop
37. Length of unpaired region at the 3′ end

Phylogenetic analyses

All data matrices were analyzed using equally weighted maximum parsimony (MP). MP tree reconstructions were sought using either branch-and-bound or heuristic search strategies in PAUP* (Swofford, 2003). When the heuristic search strategy was used, 10,000 heuristic searches were initiated using random addition starting taxa, tree bisection reconnection (TBR) branch swapping, and selecting the multrees option. One shortest tree was saved from each search. Bootstrap values (Felsenstein, 1985) were calculated from 1,000 replicate analyses using a heuristic search strategy, 500 random replicates of addition sequence of taxa, and TBR branch swapping. Only those values compatible with the 50% majority rule consensus tree were recorded. Decay analyses (Bremer, 1988; Donoghue *et al*., 1992) were performed using reverse constraints in PAUP*. Each reverse constraint analysis was conducted with TBR branch swapping and multrees on 1,000 random addition sequence starting trees. The g_1 statistic of skewed tree length distribution was calculated from 10,000 random parsimonious trees, and was used to assess the amount of nonrandom structure in the data (Hillis and Huelsenbeck, 1992). Distance trees of the SINE RNAs were reconstructed using the neighbor-joining method (Saitou and Nei, 1987) implemented in PAUP*.

Trees of substructures and models of RNA evolution

Phylogenetic trees of substructures depict the evolutionary relationships of molecular components that make up the RNA molecules analyzed. These trees reveal evolutionary patterns of structural (chemical) diversification, showing how RNA structure changes in the course of evolution. Patterns suggest, by definition, a structural origin and a series of steps describing how individual substructures are incorporated into the evolving RNA molecules.

In these trees, terminal leaves (taxa) are RNA substructures present in extant molecules and internal nodes are substructures of ancient molecular repertoires. Substructures can be for example structural domains, stems, hairpin loops, internal loops, bulges, unpaired segments in multiloops or external segments, base pairs, or modified bases. Internal nodes delimit the birth of substructures in the context of other substructures of the molecular repertoire. Consequently, tree topologies define evolutionary timelines of structural diversification.

Phylogenetic characters represent features of structures, describing for example the 'shape' of homologous substructures of the same kind, that are characteristic of individual molecules of the molecular repertoire examined. Therefore, characters depict molecular lineages generally associated with a family of molecules that is evolving within an organismal lineage. In Table S2, characters (rows of the data matrix) are SINE RNA representatives of SINE families. While characters can be considered relatively free of covariation patterns and consequently of tendencies of overconfidence in phylogenies, the validity of the reconstruction exercise depends on an adequate sampling of the molecular repertoire. With SINE RNAs, each structure represents a consensus of numerous structures within SINE families, guaranteeing an extensive sampling of molecular variants within each family. However, we do not know how many eukaryotic SINE families are present in nature. Consequently, poor sampling could result in missing substructural components and models of molecular evolution that are deficient.

Table S2. Data matrix used in the cladistic analyses of tRNA-related SINE RNA molecules. See Table S1 for a detailed description of substructural features.

SINEs	Substructural features 10 20 30 • • •
B2-Rodent	02S7E6GA2000215008230000004G7C68620009
BC1 ID-Rodent	03U6A3G6800000000000000000000000000003
SB1-*Brassica*	15O9G6E5500023E5B6B2100012D3648230004
SB2-*Arabidopsis*	11U493F9600022A24535200012F464A230004
SB6-*Arabidopsis*	03I4D3A71000718116512000 1B/S\9SeC0005
ERE2-Equidae	06O6CAL1213725I9B4C51000A4Y7G8GD5000G
RSINE1-Rodent	15P9L5H3500052J679A81000416115600000J
pSINE1-Rice	73J5C3C43000214008220000919006261000 9
SINE1MD-Opossum	02b7K5LA60001181545300001 3S583DD20001
Hpa1-Salmon	56ZAO4EAB00061600C33000003J5646B20002
AFROSINE-Afrotheria	05O4CBE8213612A23653200005a57BGC81351
Tu type III-Shrews	02D39693100041C228831000A9rKe4VD9000C
HE1-Gummy shark	26dDU5P9500037tBM9NEI13552E2949320004
CYN1-Lemur	06QAG7AA60000000000000000000000000003
SB8-*Brassica*	02I5C4C5100042C36482200000000000000005
Char1-Whales	04Q8N6H450004160034110000000000000001
DAS1-Armadillo	01G23495200042A26862200000000000000001
Type III-Galago	03E574A1300051D234A2100000000000000D
SOR-Shrews	06QCL3J5200045RBH5I630000000000000001
Ther1-Human	14aBE4JA70004AuAHKMR716D0000000000001

Phylogenetic trees of substructures provide by definition a model of structural evolution. When using shape characters, this model describes an evolutionary process in which RNA molecules evolve from an originating substructure by incorporation of base pairs to stems, nucleotides to unpaired substructures, and stems and unpaired substructures to other components in stem-loops. This occurs in defined order. Consequently, the topologies of trees of substructures can be used to build models of RNA evolution such as the one in Figure 4 in the main text. Generally, the phylogenetic relationship of stems determines the evolution of the overall shape of the molecule that is typical of the molecular repertoire studied, and should be considered first. Trees of unpaired structural components provide patterns of 'decoration' of the evolving stems. Features describing base pairs in stems generate trees that depict preference for usage of nucleotides in base pairing interactions. Numbers of paired (Sn) and unpaired (Un) regions describe interruptions in coaxial stacking of helical segments and relative frequency of bulges and internal loops in these segments and generate trees that describe evolution of these geometrical features in RNA molecules.

References

Bremer, K. (1988) The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42, 795-803

Caetano-Anollés, G. (2001) Novel strategies to study the role of mutation and nucleic acid structure in evolution. *Plant Cell Tissue Org. Cult.* 67, 115-132

Caetano-Anollés, G. (2002a) Evolved RNA secondary structure and the rooting of the universal tree. *J. Mol. Evol.* 54, 333-345

Caetano-Anollés, G. (2002b) Tracing the evolution of RNA structure in ribosomes. *Nucleic Acids Res.* 30, 2575-2587

Caetano-Anollés, G. (2005) Grass evolution inferred from chromosomal rearrangements and geometrical and statistical features in RNA structure. *J. Mol. Evol.* 60, 635-652

Donoghue, M.J. *et. al.* (1992) Phylogenetic relationships of Dipsacales based on *rbc*L sequences. *Ann. Missouri Bot. Gard.* 79, 333-345

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791

Hillis, D.M. and Huelsenbeck, J.P. (1992) Signal, noise, and reliability in molecular phylogenetic analyses. *J. Hered.* 83, 189-195

Pollock, D.D. (2003) The Zuckerkandl Prize: structure and evolution. *J. Mol. Evol.* 56, 375-376

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing evolutionary trees. *Mol. Biol. Evol.* 4, 406-425

Swofford, D. L. (2003) PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

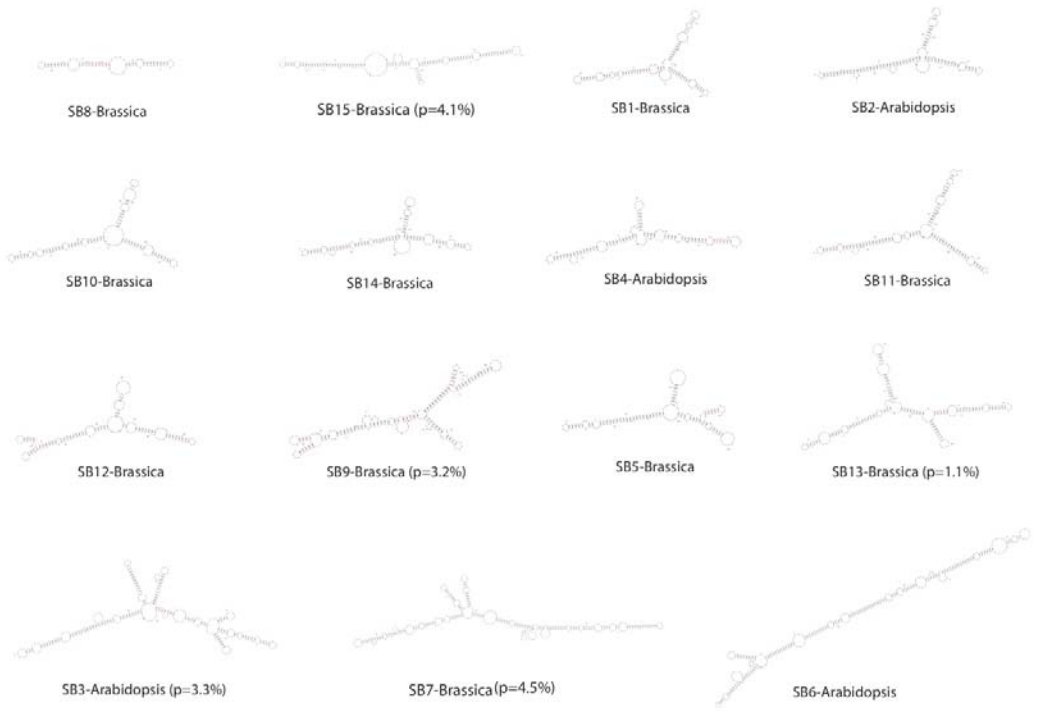


Figure S1. SINE RNA structures of the 15 different consensus sequences representing the known SINE families in *Arabidopsis* and *Brassica*. The structures were obtained using the Mfold program as described in Figure 2

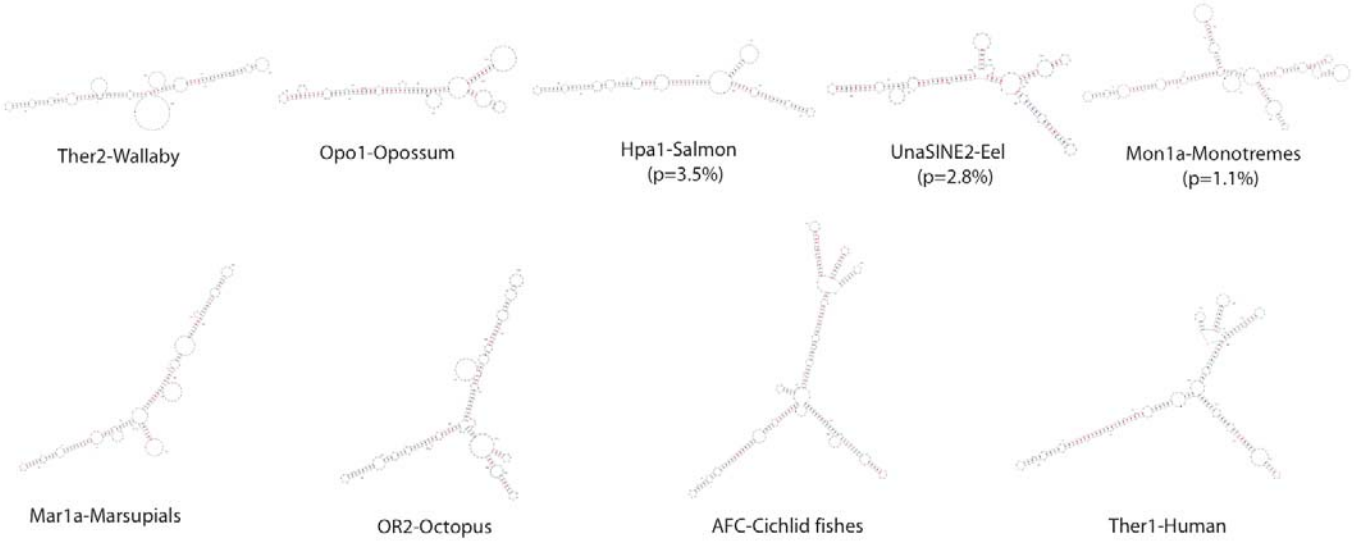


Figure S2. CORE-SINE RNA structures. The structures were obtained using the Mfold program as described in Figure 2

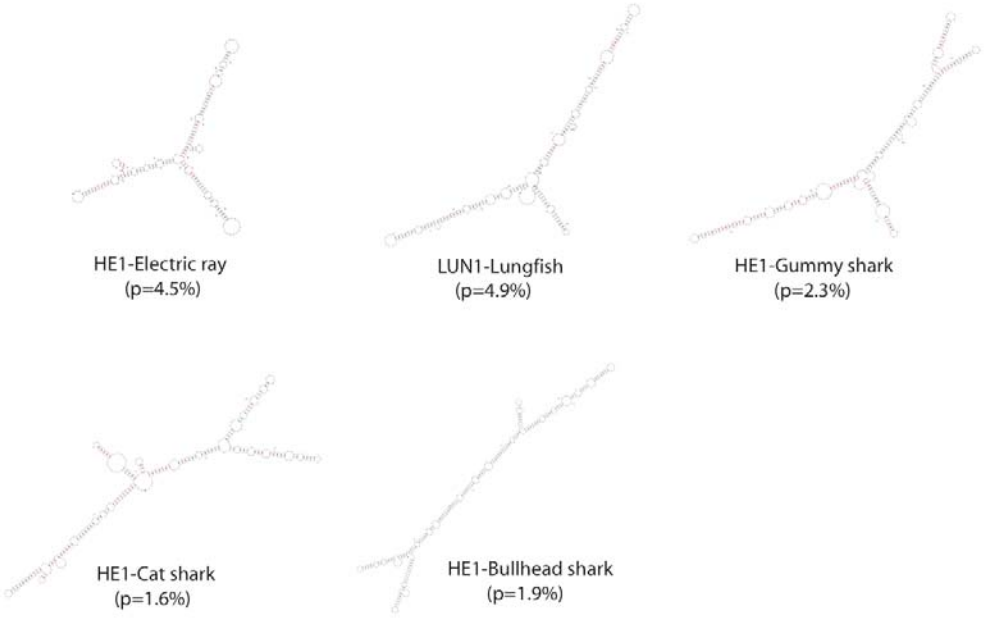


Figure S3. V-SINE RNA structures. The structures were obtained using the Mfold program as described in Figure 2

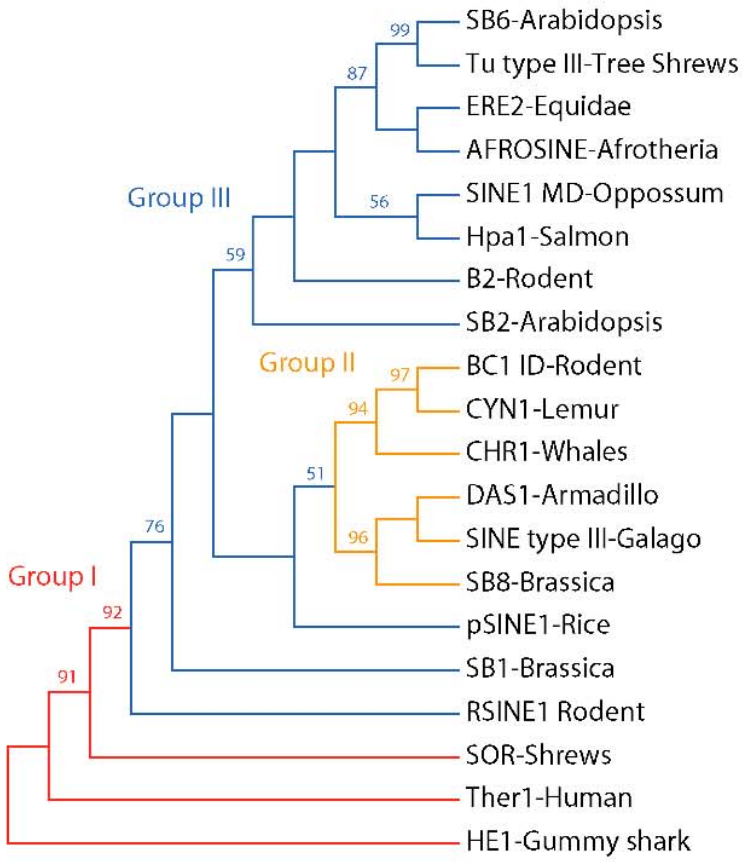


Figure S4. Neighbor-joining tree of 20 tRNA-related SINE RNAs inferred from 37 structural characters. Based on the results of the maximum parsimony analyses (Figure 3), HE1-Gummy shark is chosen as the outgroup. Numbers above branches indicate bootstrap estimates for 1,000 replicate analyses; values <50% are not indicated. Groups I, II, and III of SINE RNAs circumscribed in Figure 3, based on the results of the maximum parsimony analyses of the structural characters, are indicated.

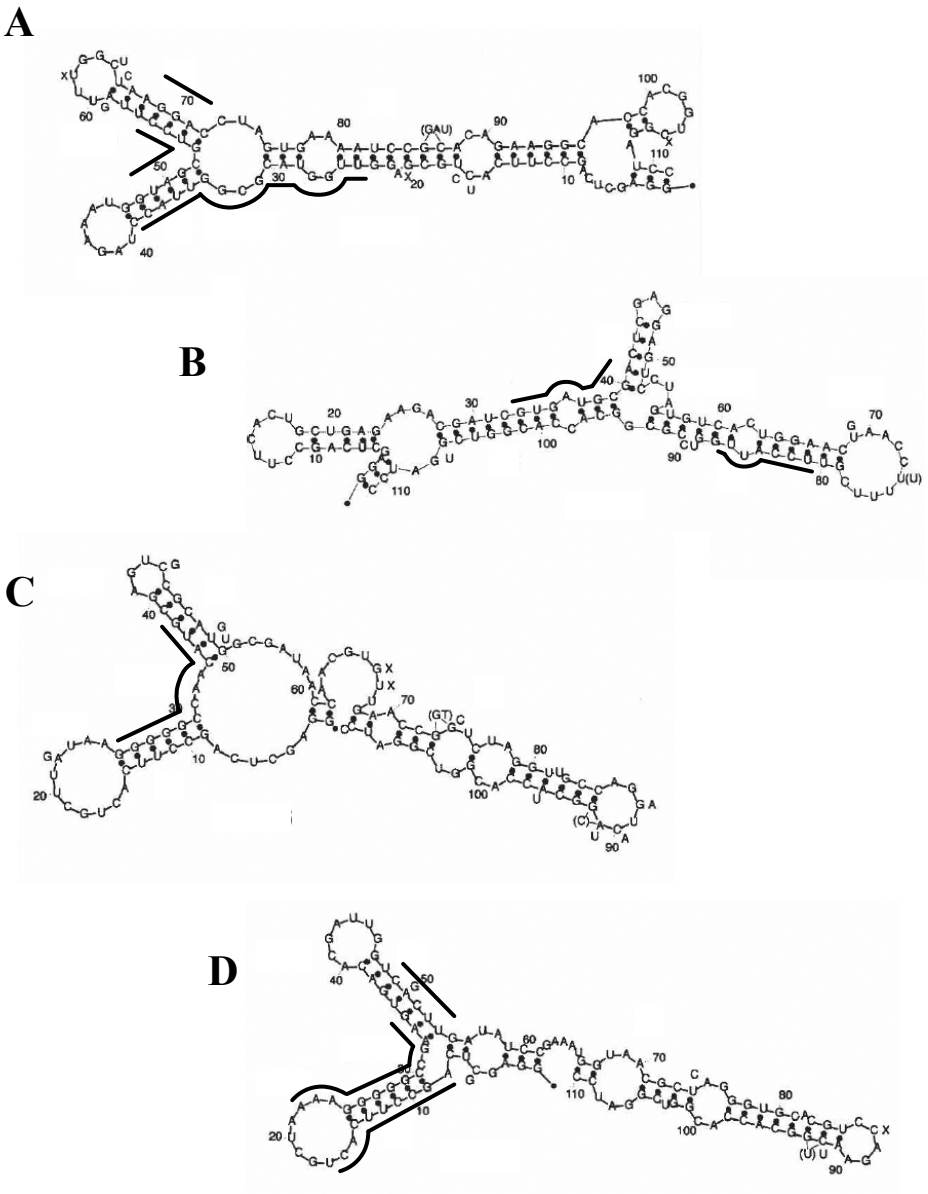


Figure 27. Exemples d'ARN reconnus avec la plus grande affinité par des domaines dsRBD lors des expériences de SELEX menées par Hallegger et al. (2006)

Les ARN représentés en (A) et (B) ont été identifiés avec le deuxième domaine dsRBD de la protéine ADAR1 du xénope.

Les ARN représentés en (C) et (D) ont été identifiés avec le deuxième domaine dsRBD de la protéine Xlrbpa du xénope.

Les lignes noires représentent les zones protégées par les domaines dsRBD lors d'expérience de protection à la RNase V1.

(adapté de Hallegger et al., 2006)

IV.4. Conclusion

Comme nous l'avons vu précédemment les SINE utilisent la machinerie des LINE pour rétroposer. Les LINE évoluent vite et avec peu de conservation entre les espèces. S'il existait une pression de sélection en faveur de la rétroposition des SINE dérivés d'ARNt, on devrait observer au sein de chaque génome une évolution conjointe des SINE avec leur partenaire LINE (i.e. une évolution concertée), ce qui n'est pas le cas. L'évolution observée peut en revanche s'expliquer par une pression de sélection similaire sur la structure des transcrits SINE pour la plupart des eucaryotes. Il existe un grand nombre de protéines capables de lier l'ARN, et différents domaines permettant cette interaction ont été identifiés. La reconnaissance de l'ARN peut être dépendante d'une séquence particulière, mais certains motifs comme les motifs de liaison à l'ARN double brin interagissent avec le squelette carboné de l'ARN et reconnaissent donc des structures. Les analyses de complexes ARN-Protéine contenant des domaines de liaison à l'ARN double brin ont montré que ces protéines peuvent être spécifiques de structures aussi variées que les tiges-boucles, les boucles internes, les renflements et les hélices présentant des mésappariements (revu dans(Fierro-Monti and Mathews 2000; Stefl *et al.* 2005)). Récemment Hallegger et al. (Hallegger *et al.* 2006) ont abordé le problème dans l'autre sens en cherchant quelles structures étaient reconnues par des domaines de liaison à l'ARN double brin. *In vitro* par des expériences de SELEX, ils ont utilisé les domaines de deux protéines pour cribler les transcrits de 74 nt issus d'une banque d'ADN de séquences aléatoires. Les ARN pour lesquels les domaines ont le plus d'affinités se replient en trois tiges-boucles et forment des structures ressemblant à celles observées pour les SINE dérivés d'ARNt (figure 27). Notre hypothèse est qu'il existe une pression s'exerçant pour que les transcrits SINE adoptent une structure particulière leur permettant de lier des protéines conservées, notamment des protéines à domaine de liaison à l'ARN double brin, et puissent ainsi exercer des fonctions riborégulatrices dans la cellule.

Discussion

La protéine La est une protéine impliquée dans de nombreux aspects de la biologie des ARN. Des homologues ont été identifiés et caractérisés chez un grand nombre d'organismes eucaryotes, allant de la levure à l'homme. En revanche, aucun homologue de plante n'avait jusqu'alors été identifié. Mes travaux de thèse constituent la première étude de la fonction La chez les plantes.

I. Deux authentiques protéines La chez *Arabidopsis thaliana*

Les protéines La précédemment identifiées présentent toutes une région amino-terminale caractéristique portant un domaine La et un domaine RRM (RRM1) associés en tandem. À l'exception des protéines des levures *S. cerevisiae* et *S. pombe*, les autres protéines La ont une région carboxy-terminale longue. Cette région de la protéine La humaine contient un second domaine RRM de structure atypique (RRM2), dont la présence était aussi suspecté dans la protéine de drosophile.

Sur la base de ces données de structure, nous avons recherché l'homologue d'*Arabidopsis* de la protéine La. La situation que nous avons mise à jour est exceptionnelle comparée aux autres eucaryotes puisqu'il existe deux protéines présentant toutes les caractéristiques structurales d'une authentique protéine La (cf. paragraphe II.1 des résultats). Nos analyses phylogénétiques montrent que les domaines La et RRM1 des deux protéines d'*Arabidopsis* sont proches de ceux de la protéine La de *S. cerevisiae*. En revanche, comme les autres protéines La des eucaryotes supérieurs, elles présentent des régions carboxy-terminales longues contenant notamment un domaine RRM2. Par des méthodes bioinformatiques, nous avons détecté ce RRM2 dans l'ensemble des authentiques protéines La d'eucaryotes supérieurs (figure supplémentaire 1 de l'article 1). Il est par contre absent de toutes les autres protéines contenant un domaine La, que nous avions incluses dans l'analyse. Ces nouvelles données indiquent que, chez les eucaryotes supérieurs, les authentiques protéines La contiennent à la fois un tandem La-RRM1 et un RRM2 atypique.

L'absence du RRM2 dans les protéines de levures ainsi que l'étude de versions tronquées de la protéine La humaine suggèrent que le RRM2 n'est pas impliqué dans la liaison aux extrémités 3' UUU-OH (Wolin and Cedervall 2002; Horke *et al.* 2004). La présence de ce domaine atypique dans les protéines des eucaryotes supérieurs est très probablement lié à de nouvelles fonctions La qui restent à identifier.

II. Pourquoi deux protéines La chez *Arabidopsis thaliana* ?

Les données structurelles et phylogénétiques soutiennent très fortement que les protéines codées aux loci At34g32720 et At1g79880 sont d'authentiques facteurs La. Les études fonctionnelles ont montré de façon claire que la protéine AtLa1 endosse très vraisemblablement la fonction La nucléaire dans la maturation et la protection de l'extrémité 3' des transcrits pol III et qu'elle participe aussi au processus de maturation de transcrits pol II non-codants tels que des snoARN ou des snARN et/ou au processus d'assemblage de certaines de leurs RNP (comme c'est le cas lors de l'assemblage de la RNP U6 chez la levure, cf. paragraphe III.3.3 de la synthèse bibliographique et article 1).
Le profil d'expression de l'ARNm AtLa1 est cohérent avec ce rôle nucléaire (figure supplémentaire 2 de l'article 1), ainsi l'ARNm AtLa1 est présent dans tous les tissus en division active où il est nécessaire de produire des ARNt, des ribosomes, des protéines et moins en phase de sénescence.

Pour la protéine At79, la situation est moins claire. Si elle s'associe bien à des ARN, dont la nature reste à déterminer (cf. paragraphe II.3.2 des résultats), elle s'est avérée incapable de complémenter les levures mutantes pour le gène codant la protéine La de levure (Lhp1p), ou de lier des transcrits pol III dans nos expériences de co-immunoprécipitation. Ces résultats nous ont paru surprenants. D'après les comparaisons de séquence que nous avons effectuées, les domaines La et RRM1 des protéines AtLa1 et At79 ont toutes les caractéristiques de domaines fonctionnels (figures 15.B et 15.C). En effet, les acides aminés impliqués dans la liaison à l'ARN, lors de l'étude de Teplova et al (2004), sont conservés chez AtLa1 et At79. Sur la seule base des séquences des régions amino-terminales des protéines, rien n'indique que la protéine At79 n'ait pas la capacité de lier l'extrémité 3' UUU-OH des transcrits pol III.

Si les régions amino-terminales des protéines AtLa1 et At79 présentent un bon niveau de conservation, la comparaison de leurs séquences carboxy-terminales révèle un plus fort degré de divergence (figure supplémentaire 1 de l'article 1). Cette constatation nous a conduit à proposer l'hypothèse selon laquelle la région carboxy-terminale de la protéine At79 pourrait interférer avec l'association aux extrémités 3' UUU-OH.

Plusieurs expériences sont envisageables pour vérifier cette hypothèse. Nous proposons de construire des vecteurs exprimant des protéines recombinantes dont les parties

carboxy-terminales ont été inversées : NH_2(AtLa1)-COOH(At79) et NH_2(At79)-COOH(AtLa1), ainsi que des protéines recombinantes dont les parties carboxy-terminales ont été remplacées par celle de *S. pombe* (plus courte, sans RRM2, mais qui conserve un signal de localisation nucléaire NLS pour adresser la protéine vers le noyau) ou un NLS de plante. Ces constructions pourront ensuite être utilisées dans des expériences de complémentation des phénotypes mutants de levures *S. cerevisiae* déficientes pour la protéine La. Des constructions exprimant une version taguée de ces protéines recombinantes pourront également être utilisées *in planta* pour tester leur capacité à lier des précurseurs de transcrits pol III dans des expériences de co-immunoprécipitation. Ces dernières expériences devraient aussi apporter des indications sur l'importance des régions carboxy-terminales des protéines de plantes dans la reconnaissance des extrémités 3' UUU-OH. Pour mieux comprendre les différences d'affinité et de spécificité existant entre les protéines AtLa1 et At79, il serait intéressant de produire des protéines recombinantes (portant des délétions ou des mutations) et de procéder à des tests *in vitro* de liaison à différents substrats ARN.

La fonction La présente donc chez *Arabidopsis thaliana* une situation originale, puisqu'il n'existe pas une, mais deux protéines présentant toutes les caractéristiques structurales d'une protéine La. Un cas de figure similaire a récemment été observé chez *Arabidopsis,* pour la nucléoline, une protéine très abondante, localisée majoritairement dans le nucléole et impliquée dans la biogenèse des ribosomes (Ginisty *et al.* 1999).

Lorsque F. Pontvianne *et al.* ont cherché à identifier l'homologue de la nucléoline chez *Arabidopsis,* ils ont trouvé deux gènes, qu'ils ont nommé *AtNUC-L1* et *AtNUC-L2,* codant des protéines présentant 49 % d'homologie (Pontvianne *et al.* 2007). La situation chez *Arabidopsis* est donc différente de celle observée chez les levures et les animaux, où il n'existe qu'un seul gène codant la nucléoline. Toutefois, en conditions normales de croissance, seule la protéine AtNUC-L1 est exprimée. Les plantes mutantes pour le gène *AtNUC-L1* sont viables, mais présentent des phénotypes de croissance et de développement très sévères. Dans ces plantes mutantes, le gène *AtNUC-L2* est activé, suggérant que la protéine AtNUC-L2 est capable de compenser au moins partiellement l'absence de la protéine AtNUC-L1.

Si la situation de départ paraît comparable, une fonction qui est normalement assurée par une seule protéine semble l'être par deux chez *Arabidopsis,* nos résultats ont montré que ce n'était pas le cas. En effet, alors que seul le gène codant la protéine AtNUC-L1 est actif dans des conditions normales de croissance, nous avons montré que dans les mêmes

conditions les gènes codant les protéines AtLa1 et At79 sont tous les deux actifs (figure supplémentaire 2 de l'article 1). De plus, la protéine AtNUC-L2 est capable de partiellement compenser l'absence de la protéine AtNUC-L1, ce qui n'est pas le cas pour les protéines AtLa1 et At79, puisque la protéine At79 n'est pas capable de compenser l'absence de la protéine AtLa1 au moins au cours de l'embryogenèse (figures 6 et 7 de l'article 1). Dans le cas des protéines AtLa1 et At79, nous pensons donc être en présence de protéines qui se sont spécialisées pour remplir des aspects différents de la fonction La.

Notre étude de la fonction La chez *Arabidopsis* nous a permis d'accumuler des arguments en faveur de cette hypothèse. (1) L'analyse des données du transcriptome d'*Arabidopsis thaliana* (projet AtGenExpress, http://www.uni-tuebingen.de/plantphys/AFGN/) a montré que les deux gènes *ATLA1* et *AT79* étaient exprimées de façon constitutive, mais que le ratio entre leurs niveaux d'expression était variable selon le tissu et le stade du développement. Ce qui suggère que les protéines AtLa1 et At79 sont régulées indépendamment et selon nous soutient l'idée qu'elles assurent différents aspects de la fonction La. (2) L'étude de plantes mutantes pour les gènes *ATLA1* et *AT79* a montré que les protéines n'étaient pas complètement redondantes au moins pendant l'embryogenèse. (3) L'analyse globale des ARN associés aux deux protéines a montré qu'elles liaient des ARN différents.

Pour continuer à explorer les relations existant entre les protéines AtLa1 et At79, la caractérisation de leurs promoteurs respectifs a été entreprise au laboratoire par Cécile Bousquet-Antonelli. Afin d'étudier et comparer leurs profils d'expression, des plantes transgéniques exprimant un gène rapporteur (type Glucuronidase) sous contrôle de ces promoteurs sont en construction. Nous prévoyons également d'effectuer de nouvelles expériences de complémentation *in planta*, en exprimant la protéine At79 sous contrôle du promoteur du gène *ATLA1*. Cette approche devrait nous permettre de tester l'hypothèse selon laquelle la protéine At79 ne compense pas l'absence d'AtLa1 au cours de l'embryogenèse parce que le niveau d'expression du gène *AT79* est trop faible à ce stade du développement. Si cette complémentation est possible, nous disposerons d'un outil pour analyser les ARN fixés par la protéine At79 en l'absence de son potentiel compétiteur AtLa1, notamment les transcrits Pol III dans l'hypothèse qu'At79 aurait une plus faible affinité qu'AtLa1 pour ces transcrits. Dans le cas où la complémentation s'avérerait impossible, il serait intéressant d'utiliser une approche d'ARNi inductible, ciblant l'ARNm d'AtLa1. Nous pourrons alors

déterminer l'effet de l'absence ou d'une sous-accumulation de la protéine AtLa1 au niveau somatique et tester dans ces conditions l'association de la protéine At79 aux transcrits Pol III.

Nos données indiquent que la protéine AtLa1 remplit au moins la fonction La nucléaire dans la maturation et la protection de l'extrémité 3' des transcrits Pol III et que la protéine At79 c'est vraisemblablement spécialisée pour assurer d'autres aspects de la fonction La, qu'il nous reste à déterminer.

III. Les autres fonctions La chez *Arabidopsis thaliana*

III.1. Un rôle dans la traduction ?

De nombreuses études, menées principalement chez les mammifères, ont mis en évidence l'implication de la protéine La dans les mécanismes de traduction. Ces études s'intéressent essentiellement à deux aspects : d'une part un potentiel rôle régulateur dans la production de la machinerie de traduction et d'autre part un rôle dans l'initiation de la traduction à partir de certains IRES. Pour montrer l'existence de telles fonctions La chez les plantes et définir les rôles respectifs des protéines AtLa1 et At79, nous souhaitons dans un premier temps tester l'association de ces protéines à des ARNm codant des protéines ribosomiques et à des ARN traduit à partir d'IRES.

Chez les eucaryotes, la protéine La semble jouer un rôle dans la production de la machinerie de traduction. Ainsi, chez la levure *S. cerevisiae*, il a été montré que la protéine La s'associait à des ARNm codant des protéines ribosomiques (Inada and Guthrie 2004). De même, chez les animaux, la protéine La s'associe à des ARNm codant les protéines ribosomiques, ainsi que certains autres composants de la machinerie de traduction (Pellizzoni *et al.* 1996 ; Meyuhas 2000 ; Cardinali *et al.* 2003). Chez les animaux, ces ARNm présentent une séquence 5' non-traduite particulière : le 5'TOP, pour 5' Terminal OligoPyrimidine. Des expériences *in vitro* indiquent que la protéine La se fixerait au niveau de cette séquence 5'TOP (Pellizzoni *et al.* 1996). Bien que sa fonction exacte soit controversée, il semble que la protéine La soit impliquée dans la traduction de ces ARNm (Crosio *et al.* 2000 ; Zhu *et al.* 2001; Cardinali *et al.* 2003 ; Schwartz *et al.* 2004).

Chez les plantes, des analyses de séquences des gènes codant les protéines ribosomiques suggèrent que pour certaines d'entre-elles, leurs ARNm contiendraient des séquences 5'TOP (Shama and Meyuhas 1996; Barakat *et al.* 2001 ; Casati and Walbot 2004). De plus, des expériences *in vitro* ont montré que des extraits de germes de blé étaient capables de discriminer des ARNm présentant un 5'TOP des autres ARNm. Ces éléments suggèrent que des voies de régulation de la traduction utilisant ce type de séquence existent chez les plantes.

Nous proposons donc de rechercher par RT-PCR des transcrits codant des protéines ribosomiques, présentant ou non un 5'TOP, parmi les ARN co-immunoprécipités par les protéines AtLa1 et At79.

Les IRES (Internal Ribosome Entry Site) sont des séquences qui permettent l'initiation de la traduction d'un ARN indépendamment de la coiffe (Komar et Hatzoglou, 2005). Pour certains IRES, l'assemblage du complexe d'initiation de la traduction nécessite le recrutement préalable de facteurs *trans* d'activation.

Des expériences *in vitro* et *in vivo* ont montré que la protéine La humaine est capable de reconnaître et fixer des IRES d'ARN viraux (Meerovitch *et al.* 1993; Ali *et al.* 2000) et d'ARNm endogènes (Holcik and Korneluk 2000; Kim *et al.* 2001), ce qui a conduit à l'hypothèse que la protéine La serait un facteur *trans* d'activation pour ces IRES.

Des IRES ont également été identifiés dans des ARN de virus de plantes (Basso *et al.* 1994 ; Dorokhov *et al.* 2006), ainsi que dans l'ARNm de la protéine HSP101 du maïs (Dinkova *et al.* 2005). En raison du faible nombre d'IRES connus chez les plantes, aucun facteur *trans* d'activation n'a encore été identifié. Dans ce contexte, les protéines AtLa1 et At79 sont de bons candidats et il serait intéressant d'étudier leur capacité à lier des ARN contenant des IRES connus, par exemple dans des expériences *in vitro* en utilisant des protéines recombinantes produites en bactérie. Si l'une des protéines se révélait capable de lier certains IRES, l'identification d'ARNm associés en conditions favorisant la traduction indépendante de la coiffe (infection virale) pourrait alors contribuer à la caractérisation d'IRES chez les plantes.

III.2. Un lien entre la fonction La et la réponse au stress

Différentes données de la littérature suggèrent qu'il existe un lien entre la protéine La et les voies de réponse aux stress. Ainsi, des expériences de co-immunoprécipitation ont mis

en évidence une association de la protéine Lhp1p de *S. cervisiae* avec l'ARNm de la protéine Hac1p, un facteur de transcription nécessaire à la réponse au stress des protéines mal repliées (UPR) (Inada and Guthrie 2004). Ces mêmes auteurs ont montré que la délétion du gène *LHP1* entraîne une sensibilité accrue au stress UPR, qui découle d'une plus faible accumulation de la protéine Hac1p. De même, il a été montré *in vitro* que la protéine La humaine s'associe à l'IRES de l'ARNm de la protéine Bip, un acteur de la voie de réponse au stress UPR (Kim *et al.* 2001). Chez l'homme, il a également été observé que la protéine La était relocalisée vers le cytoplasme dans certaines conditions de stress, notamment lors d'une infection par le virus de la poliomyélite (Shiroki *et al.* 1999). Le même phénomène a par ailleurs été observé en cellules HeLa dans des conditions induisant l'apoptose (Ayukawa *et al.* 2000).

Nous disposons au laboratoire de deux lignées mutantes pour le gène *AT79*. Ces lignées sont viables et ne présentent pas de phénotype évident. Dans la mesure où la protéine At79 ne semble pas impliquée dans la biogenèse des ARN se terminant par une extrémité 3' UUU-OH, nous avons émis l'hypothèse qu'elle assurerait d'autres aspect de la fonction La. Il nous semble donc très intéressant d'étudier le comportement de ses lignées en condition de stress (comme un traitement chimique induisant un stress UPR, une infection virale ou un choc thermique). Ces travaux sont en cours au sein de l'équipe et ont été initiés par Cécile Bousquet-Antonelli. Cette étude prévoit également d'utiliser les plantes transgéniques exprimant les protéines de fusion GFP-AtLa1 et GFP-At79, pour mettre en évidence des conditions induisant une relocalisation de l'une ou l'autre de ces protéines du noyau vers le cytoplasme.

Nous envisageons finalement d'utiliser ces lignées transgéniques pour co-immunoprécipiter les ARN associés aux protéines AtLa1 et At79 et, à l'aide de puces à ADN, d'identifier des cibles potentielles lors de la réponse au stress.

IV. La protéine AtLa1 est impliquée dans la biologie des SINE

Nos expériences de co-immunoprécipitation nous ont permis de mettre en évidence une association de la protéine AtLa1 avec des transcrits SINE dérivés d'ARN de transfert. Comme la protéine AtLa1 fixe les transcrits néosynthétisés par la pol III, il paraît très probable que son association aux ARN SINE s'effectue via l'extrémité 3' UUU-OH de la

forme pleine longueur (fl). Pour confirmer cette hypothèse et étudier le rôle de la protéine La dans la maturation des SINE, nous projetons de surexprimer les protéines AtLa1 et At79 dans les lignées de plantes et les lignées cellulaires dont le génome contient le locus Na7 du SINE SB1 de *Brassica napus* (Pelissier *et al.* 2004). En suivant les différentes espèces de transcrits SB1, nous pourrons voir si la surexpression d'AtLa1 permet de stabiliser la forme fl et change le profil de maturation. De plus, les lignées d'*Arabidopsis* crées au laboratoire, qui expriment le SINE SB1 présentent des phénotypes sévères de développement (Thierry Pélissier), un argument en faveur d'un rôle de riborégulateur de l'ARN SB1 chez la plante. Il serait intéressant de voir l'impact sur ces phénotypes de la surexpression de la protéine AtLa1 et de l'ARNi ciblant l'ARNm la codant.

V. Existe-t-il un lien entre la structure secondaire des ARN SINE et leurs potentiels rôles riborégulateurs ?

Les résultats de l'article 2 supportés par nos données expérimentales de structures des ARN SINE SB1 et SB2 ont permis de proposé une histoire évolutive commune de la structure des ARN SINE dérivés d'ARNt chez les eucaryotes. En effet, cette étude a montré que les ARN SINE dérivés d'ARNt et provenant d'eucaryotes variés adoptent des structures secondaires similaires. Dans le modèle proposé, ces ARN SINE auraient pour la plupart évolués d'une seule tige-boucle vers une structure à deux, puis trois tiges-boucles, présentant des structures internes comme des renflements ou des boucles internes. Il semble donc qu'une pression de sélection s'exerce pour que les ARN SINE acquièrent ces structures de plus en plus complexes.

Le motif de liaison à l'ARN double-brin dsRBD (pour double-stranded RNA Binding Domain) est présent dans des protéines de fonctions très variées. On le trouve notamment dans des enzymes qui vont modifier la structure de l'ARN (hélicases, enzymes d'édition ou nucléase des voies de l'ARNi) et des protéines dont l'activité est régulée par l'ARN double-brin (kinases ou phosphatases). Il est le plus souvent présent en plusieurs exemplaires dans les protéines. L'analyse de domaines dsRBD complexé à se l'ARN double brin indique une absence de contact entre les chaînes latérales des acides aminés et les groupements caractéristique des bases, sugérant qu'il n'y pas de reconnaissance d'une séquence spécifique (Ryter and Schultz 1998 ; Ramos *et al.* 2000 ; Wu *et al.* 2004). En revanche, des données de

structure et des expériences de biochimie ont montré que les protéines contenant des dsRBD avaient des spécificités de liaison pour des structures particulières, certaines de ces protéines semblent ainsi reconnaître des tiges-boucles, d'autres des boucles internes, des renflements ou encore des hélices présentant des mésappariements (Fierro-Monti and Mathews 2000; Stefl *et al.* 2005).

Les ARN SINE dérivés d'ARNt adoptent des structures secondaires qui présentent des zones structurées ressemblant aux structures reconnues par les protéines contenant des dsRBD. Nous proposons que la capacité d'interagir avec des protéines à domaines dsRBD serait à la base de la pression de sélection sur les ARN SINE et qu'ils joueraient donc un rôle modulateur de la fonction de ces protéines. Selon les éléments SINE et le génome dans lequel ils se trouvent, les ARN SINE serait amenés à s'associer à des protéines différentes, expliquant les variations de structures et les différents rôles riborégulateurs observés.

Matériel et Méthodes

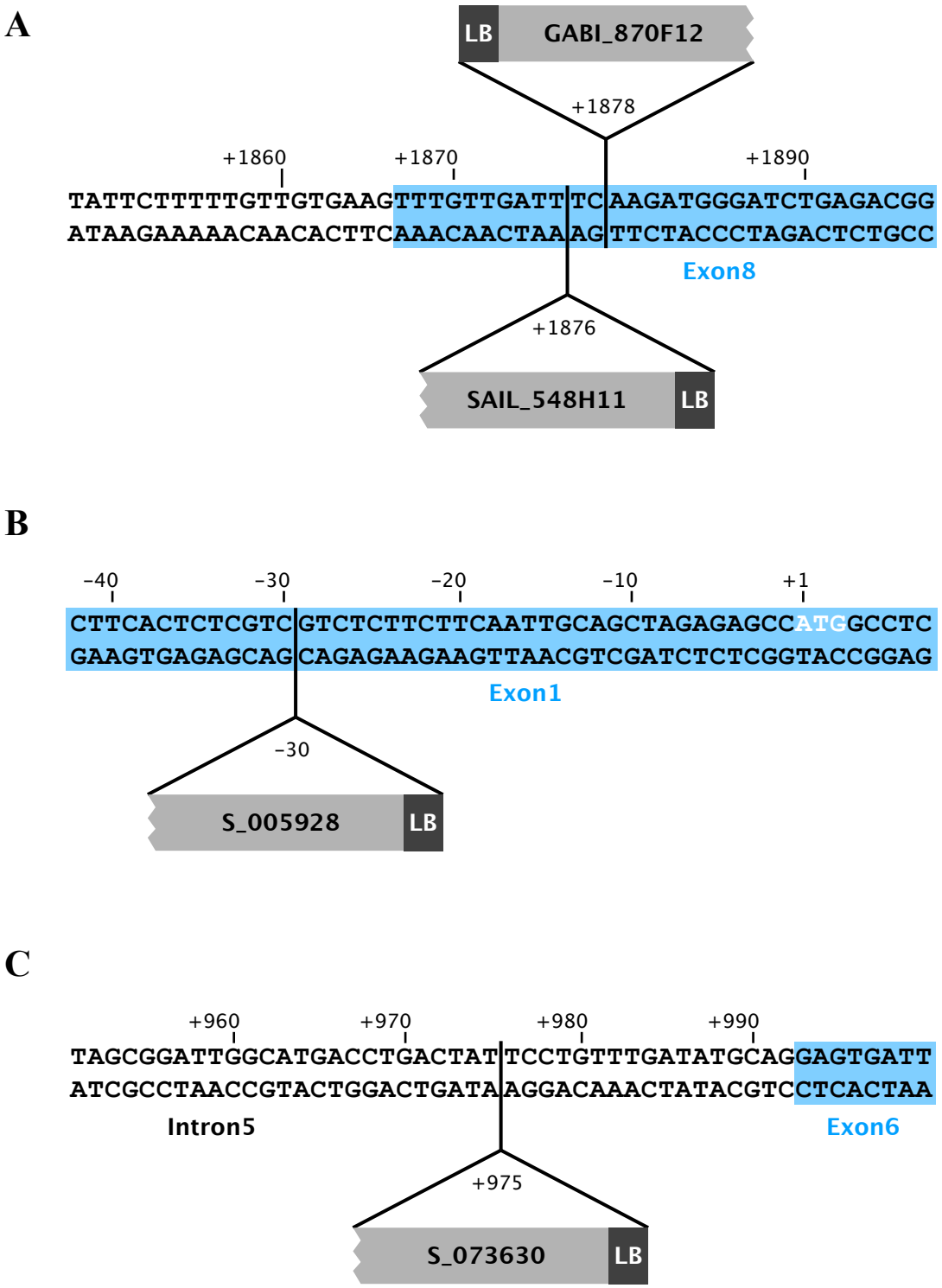


Figure 28. Détail des positions d'insertion des ADN-T dans les lignées mutantes pour les gènes *ATLA1* et *AT79*

(A) Positions des insertions ADN-T SAIL_548H11 et GABI_870F12 dans le gène *ATLA1*.

(B) Position de l'insertion ADN-T SALK S_005928 dans le gène *AT79*.

(C) Position de l'insertion ADN-T SALK S_073630 dans le gène *AT79*.

La position +1 correspond au codon initiateur de la traduction. Les boîtes bleues indiquent des exons. Les ADN-T sont figurés par des boîtes grises.

I. Lignées d'*Arabidopsis* et suspensions cellulaires

Les lignées d'insertion pour le gène *ATLA1* : SAIL_548H11 (Sessions *et al.* 2002) et GABI_870F12 (Rosso *et al.* 2003) ont respectivement été obtenues auprès de l'*Arabidopsis* Biological Resource Center et du Max Planck Institute for Plant Breeding Research. Les lignées d'insertion pour le gène *AT79* : ADN-T S_005928 et S_073630 ont été identifiées dans la banque publique de lignées établie par le Salk Institute Genomic Analysis Laboratory (Alonso *et al.* 2003) et les graines obtenues auprès du Nottingham *Arabidopsis* Stock Centre (Scholl *et al.* 2000). Les positions des insertions sont indiquées figure 28.

Les plantes sont cultivées en serre suivant les conditions standard (16 heures jour, 8 heures nuit).

Les lignées de cellules d'*Arabidopsis* exprimant les protéines taguées GFP sont dérivées de la lignée T87. Ces lignées de cellules ont été produites en utilisant la transformation par *Agrobacterium tumefaciens* (Gallego *et al.* 1999). Le plasmide portant la construction d'intérêt est introduit par électroporation dans la souche d'*Agrabocterium* LBA4404. 100 µl d'une culture d'*Agrobacterium* de 24-48h sont culottés, rincés avec 1 mL de GB5 (milieu de culture des cellules d'*Arabidopsis* : préparation Gamborg's B5, saccharose 30 g/L, pH 5,8), et repris dans 1 mL de GB5. 100 µL de cette suspension sont mélangés à 10 mL d'une culture de cellules d'*Arabidopsis* âgée de 2-3 jours et co-cultivés à 23°C pendant 48h. Après avoir éliminé le surnageant du mélange cellules-*Agrobacterium*, les cellules sont rincées deux fois dans du GB5, puis étalées sur milieu solide GB5 dans lequel on a ajouté l'hormone Acide Naphtalène Acétique (ANA 25 µg/mL), l'antibiotique Céfotaxine (200 µg/mL) et l'antibiotique permettant la sélection du plasmide d'intérêt. Après deux semaines de culture à 23°C, les cellules portant la construction d'intérêt forment des cals. Une douzaine de cals isolés sont repiqués, mis en culture indépendamment puis testés par hybridation western pour l'expression de la construction (avec des anticorps monoclonaux dirigés contre la GFP, Clontech JL8). Les lignées cellulaires sélectionnées sont ensuite cultivées sous pression de sélection constante.

II. Souches de levure et transformation par choc thermique

Les souches de *Saccharomyces cerevisiae* BP1 (*MATa lsm8-1 ura 3 lys2 ade2 trp1 his3 leu2 lhp1::LEU2*/pATL), CY2 (*MATα ura 3 lys2 ade2 trp1 his3 leu2 lhp1::LEU2*) et CY3 (*MATα ura 3 lys2 ade2 trp1 his3 leu2*) ont été données par Sandra Wolin (HHMI, New Haven)

Les souches de levures exprimant les protéines d'*Arabidopsis* ou la protéine Lhp1p seules ou taggées par la protéine A (*Staphylococcus aureus)* ou la protéine A seule ont été produites par transformation des levures par choc thermique.

Un volume (V) d'une culture (poussée jusqu'à une densité optique de 0,5) de la souche à transformer est centrifugé à 4000 g pendant 5 minutes. Le culot est repris dans ¼ V de TE-LiAc (TrisHCl 10mM pH8, EDTA 1mM, Acétate de Lithium 0,1M), centrifugé à 4000 g pendant 5 min, puis resuspendu dans dans 1/50 V de TE-LiAc. Le volume requis pour effectuer la ou les transformation(s) (40 µL par transformation) est prélevé. 50 µg d'ADN entraîneur sont ajoutés par transformation, et le mélange est réparti dans les microtubes contenant l'ADN transformant (100 ng). 4 volumes de PEG 50%-TE-LiAc sont ajoutés et les transformations sont incubées 45 min à 30°C. Du DMSO est ensuite ajouté pour arriver à une concentration finale de 6%, et les tubes sont incubés 15 min à 42°C. Après ajout de 2 volumes de NaCl à 9°/oo, les tubes sont centrifugés 15 sec, le surnageant est éliminé et le culot repris dans 100 mL de NaCl 9°/oo est étalé sur des boîtes de milieu minimum permettant la sélection des levures portant le plasmide d'intérêt, c'est-à-dire un milieu dépourvu de l'acide aminé dont la voie de biosynthèse est rétablie grâce au gène porté par le plasmide. Pour chaque construction, deux colonies isolées sont sélectionnées et amplifiées pour les manipulations suivantes.

III. Préparation et transformation transitoire de protoplastes d'*Arabidopsis*

Des cellules cultivées en suspension sont collectées, puis reprises dans une solution enzymatique (préparer un mélange : 1% cellulase, 1,25 ‰ macérozyme, 400 mM mannitol, 20 mM KCl, 20 mM MES-KOH pH 5,8 ; chauffer 10 minutes à 55ºC puis ajouter 10 mM $CaCl_2$, 0,1% BSA ; stériliser par filtration) et incubées sur la nuit à température ambiante. La suspension de protoplastes est passée sur un filtre de nylon (35-75 microns) pour éliminer les cals qui n'ont pas été digérés. Les protoplastes sont culottés en centrifugeant 1 min à 100 g

(accélération et freinage minimums), puis lavés une fois dans 30 mL de solution W5 (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 5mM Glucose, 0,03% MES, ajuster le pH à 5.8 avec une solution de KOH 1 M). Les protoplastes sont finalement repris dans du W5 à une concentration d'environ 1 x 10^5 protoplastes par mL (approximativement 5 mL de W5 pour 25 mL de cellules en suspension).

Pour une transformation transitoire, les protoplastes sont incubés 30 minutes sur de la glace, puis repris dans une solution de MaMg (15mM $MgCl_2$, 400 mM Mannitol et 5 mM Mes-KOH pH 5,8), à une concentration d'environ 1 x 10^5 protoplastes par mL. 10 µg d'ADN plasmidique (1 µg/µL) sont mélangés avec 100 µL de la suspension de protoplastes, puis à 1 volume (110 µL) de solution 40% PEG (pour 10 mL : 4g de PEG4000 (#81240, Fluka), 2,5 mL de Mannitol 800 mM, 1 mL de $CaCl_2$ 1M et 3 mL d'H_2O). La suspension est incubée 30 minutes à température ambiante, puis diluée dans 4 volumes de solution W5 (mélanger très doucement). Les protoplastes sont ensuite culottés par centrifugation, 5 minutes à 80 g, et le surnageant écarté. Les protoplastes sont finalement repris dans 1 mL de W5 et incubés sur la nuit à 23ºC, dans l'obscurité.

IV. Préparation des extraits natifs

Les extraits natifs sont préparés à partir de cellules cultivées en suspension. Les cellules d'une culture de 3-4 jours sont collectées, puis reprises dans 3 volumes (3 ml pour 1 g) de tampon de lyse (150 mM NaCl, 50 mM Tris-HCl pH 8, 2,5 mM $MgCl_2$, 0,1% triton X-100) contenant : 1 mM de dithiothreitol (DTT), 10 mM de vanadyl ribonucleoside complex (New England Biolabs), 1% d'un cocktail inhibiteur de protéases (P9599, Sigma), 2 mM de benzamidine, 1 mM de de phenyl méthyl sulfoxy fluoride (PMSF) et 10 µM d'inhibiteur de decarboxylase (DCI). Les cellules sont ensuite lysées sous une pression de 552 bars, à l'aide d'un désintégrateur de cellules "One Shot" (Constant Systems Ltd). Les extraits sont finalement clarifiés par centrifugation, 30-45 minutes à 16000 g. L'ensemble des opérations est réalisées à 4ºC.

V. Immunoprécipitations et élutions des ARN et des protéines

Pour immunoprécipiter les protéines taguées par la GFP, les extraits natifs sont incubés avec une dilution 1:300 d'un anticorps polyclonal de lapin dirigé contre la GFP

(Clontech), 3 heures à 4ºC sur une roue. Pendant ce temps, des billes magnétiques Bio-Adembeads couplées à la protéine A (Ademtech S.A.) sont équilibrées dans 2 volumes de tampon de lyse sans inhibiteurs. Les échantillons sont ensuite mélangés avec les billes magnétiques suivant les indications du fabricant (10 µL de billes par µg d'anticorps) et incubés une heure de plus à 4ºC. Après trois rinçages avec du tampon de lyse, 1/9ème des billes est prélevé, puis chauffé 5 minutes à 95ºC dans 30 µL de tampon Laemmli pour procéder à l'analyse des protéines par hybridation western. Le reste des billes est utilisé pour l'élution des ARN co-immunoprécipités. Les billes sont reprises dans une solution d'urée à 8 M et 5 mM d'EDTA, puis incubées 10 minutes à température ambiante. Après une extraction phénol:chloroforme:alcool isoamylique, les ARN sont précipités par addition de 40 mg de glycogène et 2,5 volumes d'éthanol absolu. Les ARN sont finalement repris dans de l'eau ultrapure Milli-Q®.

VI. RT-PCR des transcrits Pol III et des transcrits SINE

Les ARN précipités à partir d'extraits natifs ou après immunoprécipitation des protéines taguées GFP sont d'abord traités avec le kit "DNA-free" (Ambion), suivant le protocole "Rigorous DNase treatment" du fabricant. Une partie des ARN (généralement 1/5ème) est rétrotranscrite à l'aide de l'"expand reverse transcriptase" (Roche) et d'amorces aléatoires (hexanucléotides). Les ADNc ainsi obtenus sont ensuite amplifiés par PCR à l'aide de couples d'amorces spécifiques des précurseurs de l'ARNt méthionine et de l'ARNt-ARNsnoR43.1, des éléments SINE SB1, SB2 et SB4 et de l'ARNr 5.8S.

VII. Marquage pCp

L'ARN ligase T4 (#AM2140, Ambion) est utilisé pour lier de la cytidine 3',5' bisphosphate [5'^{32}P] (pCp) à l'extrémité 3'-OH des ARN, suivant les recommandations du fabricant. L'ARN ligase T4 est éliminée par une extraction phénolique, puis les ARN sont précipités à l'éthanol en présence de glycogène (20 µg). Le culot d'ARN marqués est ensuite repris dans 40 µl de tampon de charge (7 M urée, 0,05% bleu de bromophénol et 0,05% xylène cyanol). Les ARN marqués (5 µl) sont finalement séparés sur un gel de polyacrylamide 6% ou 15%, en conditions dénaturantes.

VIII. Cartographie en solution des ARN SINE SB1 et SB2

Tableau 1 : Présentation des coupures enzymatiques et modifications chimiques

Enzymes	**Coupure**	**Détection**
Rnase T1	G non apparié	détection directe
Rnase T2	A non apparié	
Rnase V1	N apparié ou empilé	

Agent chimiques	**Modification**	**Détection**
DMS	A ----> A(N1-CH_3)	extension d'amorce
	C ----> C(N3-CH_3)	
CMCT	G ----> G(N1-C=N-R) | NH-R’	
	U ----> U(N3-C=N-R) | NH-R’	

VIII.1. Préparation de l’ARN

La séquence des ARN SINE SB1 ou SB2 est insérée directement en aval d’un promoteur T7 dans le vecteur pUT7 (Serganov *et al.* 1997). Le vecteur est d’abord linéarisé en aval de la séquence d’intérêt. La transcription *in vitro* est réalisée avec de l’ARN polymérase T7 (Roche), à partir de 10 µg d’ADN (tampon ARN polymérase T7 1x, 4 mM de chaque nucléotide triphosphate, 5 mM de guanosine monophosphate, 16 mM de $MgCl_2$, 40 unités de Rnasine et 80 unités d’ARN polymérase T7, dans un volume final de 100 µL). Les ARN transcrits *in vitro* sont purifiés sur un gel de polyacrylamide 8% - urée 8 M et élués sur la nuit à 4ºC dans une solution d’acétate d’ammonium 0,5 M, EDTA 1 mM. Ils sont ensuite précipités par 2,5 volumes d’éthanol absolu, lavés deux fois à l’éthanol 80 % et repris dans de l’eau ultrapure.

VIII.2. Marquage des ARN en 5'

Les ARN sont déphosphorylés par de la phosphatase alcaline de bactérie (BAP), qui est ensuite éliminée par une extraction phénolique. Les ARN sont précipités comme précédemment. Le marquage en 5' est effectué par la polynucléotide kinase T4 (T4 PNK) qui transfert le phosphate γ du [γ^{32}P]-ATP sur le groupe hydroxyle terminal. Les ARN marqués sont ensuite purifiés sur un gel polyacrylamide, comme décrit précédemment.

Avant d'être utilisés, les différents ARN produits sont dénaturés par une incubation de 2 minutes à 90ºC, puis refroidis sur la glace 1 minute. Ils sont ensuite renaturés 15 minutes à 20ºC dans le tampon utilisé pour la digestion enzymatique ou les modifications chimiques.

VIII.3. Cartographie en solution par digestion enzymatique (méthode de détection directe)

La digestion enzymatique est réalisée sur une quantité d'ARN marqués en 5' correspondant à 50 000 coups par minute (cpm). Les ARN sont incubés 5 minutes à 20ºC dans un volume total de 10 µL (tampon N : 25 mM Hepes-KOH pH 7,5, 5 mM acétate de magnésium et 50 mM acétate de potassium), en présence de 2 µg d'ARNt total (SIGMA) et de différentes quantités de l'une des enzymes Ambion suivantes : RNase T1 (0,2 u et 0,4 u), RNase T2 (0,05 u et 0,1 u) ou RNase V1 (0,05 u et 0,1 u).

Après digestion, la réaction est diluée avec 40 µL d'acétate de sodium 0,3 M et l'enzyme est éliminée par une extraction phénolique (vérifier la présence des ARN marqués à l'aide d'un compteur de radioactivité). Les ARN sont précipités par 3 volumes d'éthanol absolu (vérifier la présence des ARN dans le culot à l'aide d'un compteur radioactivité), puis rincer une fois avec 150 µL d'éthanol 80%. Le culot est séché et repris dans 10 µL de tampon de charge (8 M urée, 0,05 % bleu de bromophénol et 0,05% xylène cyanol). Pour chaque échantillon, 3 µL sont directement chargés sur un gel de polyacrylamide 12%-urée 8M.

Afin de pouvoir déterminer les positions de coupure, deux types de marqueurs de taille sont préparés. L'ARN marqué est hydrolysé en conditions alcalines pour obtenir une échelle base à base. 1 µl (50 000 cpm) d'ARN marqué est incubé 3 min à 90ºC en présence d'1 µg d'ARNt total, dans un volume total de 10 µL de carbonate de sodium (Na_2CO_3/$NaHCO_3$) 0,1 M pH 9. Pour le second marqueur, l'ARN marqué est hydrolysé par la RNase T1 en conditions dénaturantes (l'enzyme coupe après tous les "G"). 1 µl (50 000 cpm) d'ARN

marqué est incubé 5 min à 50ºC en présence d'1 µg d'ARNt total, dans 5 µL de tampon ΔT1 (20 mM citrate de sodium pH 4,5, 1 mM EDTA, 7 M urée 0,05 % bleu de bromophénol et 0,05% xylène cyanol). Une unité de RNase T1 est ajoutée et le mélange est incubé 10 min de plus à 50ºC. Pour les deux marqueurs, un volume de tampon de charge est ajouté à chaque échantillon. Les marqueurs peuvent être conservés à -20ºC.

VIII.4. Cartographie en solution par modifications chimiques (méthode de détection indirecte)

Le DMS (DiMéthylSulfate) doit être manipulé sous la hotte, les cônes utilisés et les surnageants doivent être placés dans une solution de NaOH 1 N (bouteille placée sous la hotte) qui détruit instantanément toute trace de DMS. Le DMS est dilué dans l'éthanol.

Le CMCT (1-Cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate) doit être manipulé sous la hotte, les cônes utilisés et les surnageants doivent être placés dans une solution d'acide acétique 10%. Le CMCT est dilué extemporanément dans de l'eau ultrapure.

Les modifications chimiques ont été effectuées sur 0,2 µg d'ARN non marqués.

Pour la modification par le DMS, les ARN sont incubés 5 minutes à 20ºC dans 20 µL d'une dilution 1/200ème, 1/100ème ou 1/40ème dans du tampon N (25 mM Hepes-KOH pH 7,5, 5 mM acétate de magnésium et 50 mM acétate de potassium), en présence de 2 µg d'ARNt total (SIGMA).

Pour la modification par le CMCT, les ARN sont incubés 10 minutes à 20ºC dans 20 µL d'une solution 6 mg/mL, 8 mg/mL ou 12 mg/mL dans du tampon C (50 mM Borate-NaOH pH 8, 5 mM acétate de magnésium et 50 mM acétate de potassium), en présence de 2 µg d'ARNt total (SIGMA).

Les ARN modifiés par le DMS ou le CMCT sont ensuite précipités à l'éthanol absolu, après ajout d'acétate de sodium, et sont finalement repris dans 4 µL d'eau ultrapure.

Les sites de modifications sont déterminés par extension d'amorce. Les ARN modifiés sont mélangés avec 1 µL d'amorce marquée en 5' (100.000 cpm) et purifiée sur gel comme décrit pour les ARN. Le mélange est incubé 1 minute à 90ºC, puis 1 minute sur la glace. L'hybridation de l'amorce est ensuite réalisée dans le tampon de la transcriptase inverse AMV (RT-AMV, Roche Applied Science), 15 minutes à 20ºC. Pour l'extension d'amorce, la

réaction est incubée 30 minutes à 37ºC dans un volume total de 15 µL (tampon TR-AMV 1X, 0,3 mM de chaque dNTP et 2 u de RT-AMV).

Parallèlement, un séquençage de l'ARN étudié est réalisé pour identifier les bases modifiées. Quatre extensions d'amorce sont réalisées dans les conditions décrites précédemment, excepté que dans chacune d'elle un ddNTP différent est ajouté à une concentration finale de 3 µM, le dNTP correspondant à 30 µM et les autres dNTP à 100 µM.

Toutes les réactions sont arrêtées par précipitation éthanolique. Les ARN sont finalement repris dans du tampon de charge et la moitié de l'échantillon est séparée sur un gel de polyacrylamide 12%-urée 8M.

IX. Liste des amorces utilisées pour cette étude

Tableau 2 : liste des amorces utilisées pour le clonage des ADNc des protéines GFP, AtLa1, At79 et Lhp1p

Nom	Séquence	Utilisation
GFP-F(NheI-SmaI)	GACGGAGCTAGCCCCGGGATGGTGAGCAAGGGCGAG	Clonage de la GFP
GFP-R(Bam-Xba)	GCTCTAGATTACGGATCCTTGTACAGCTCGTCCA	Clonage de la GFP
at1g79F	GGGGGAATTCCACTCTCGTCGTCTCTTCTTC	Clonage *AT79*
at1g79R	GGGGGGATCCACGGCAATACGTCAAGTCATTA	Clonage *AT79*
at4g32F	GGGGGAATTCAGTACCCTTTTTCTCTCTTATTC	Clonage *ATLA1*
at4g32R	GGGGGGATCCGTGTGCATAAGAAGGCACATAA	Clonage *ATLA1*
Bam32ATG	CGCGGATCCGAATGTCGATTCCTTGTCTAACCG	Clonage *ATLA1*
Sph32ATG	ACTTGCATGCTCATGTCGATTCCTTGTCTAACCG	Clonage *ATLA1*
Xho32STOP	CCGCTCGAGTCATGCTTCCACCTTCTGAGAT	Clonage *ATLA1*
Bam79ATG	CGCGGATCCGAATGGCCTCCTCCTTCAACGA	Clonage *AT79*
Sph79ATG	ACTTGCATGCTCATGGCCTCCTCCTTCAACGAAG	Clonage *AT79*
Xho79STOP	CCGCTCGAGTTAGGGCTCCTTTCGGGCC	Clonage *AT79*
Bamlhp1ATG	CGCGGATCCGAATGTCTGAAAAACCACAACAAGA	Clonage *LHP1*
Sphlhp1ATG	ACTTGCATGCTCATGTCTGAAAAACCACAACAAGAG	Clonage *LHP1*
Xholhp1STOP	CCGCTCGAGTCACTCCTTGTGCTCCTCATC	Clonage *LHP1*

Tableau 3 : liste des amorces utilisées pour les tests de complémentation en levure *S. cerevisiae*

Nom	Séquence	Origine	Utilisation
ySER-intron	AGCCGAACTTTTTATTCCATTC	O'Connor & Peebles, 1991	Hybridation northern
yPRO-mature	ACCCAGGGCCTCTC	Yoo & Wolin, 1997	Hybridation northern
yLEU-intron	AATTTAGAGGTTAAATCCA	O'Connor & Peebles, 1991	Hybridation northern
yTYR-intron	AAGATTTCGTAGTGATAAATT	O'Connor & Peebles, 1991	Hybridation northern
ySER(GCU)intron	AATTGCTTTTCTGAGGAAA	O'Connor & Peebles, 1991	Hybridation northern
yLYS-intron	ATCCTTGCTTAAGCAAATGCGC	O'Connor & Peebles, 1991	Hybridation northern
yU3-1	TCCTATGAAGTACGTCGAC	Kufel *et al*.,2003	Hybridation northern
yU3-2(mature)	GTACCCACCCATAGAGCCCT	Kufel *et al*.,2003	Hybridation northern
yU3-3'ex	GTGGTTAACTTGTCA	Kufel *et al*.,2003	Hybridation northern
yU3-intron	CAAAAGCTGCTGCAATGG	Kufel *et al*.,2003	Hybridation northern
yeast5,8S	CATCGATGCGAGAACCAAGAG		Hybridation northern

Tableau 4 : Liste des amorces utilisées pour le génotypage des lignées d'insertion ADN-T

Nom	Séquence	Origine	Utilisation
LBa1	TGGTTCACGTAGTGGGCCATCG	Alonso *et al*, 2003	Génotypage des lignées SALK (LB)
S79-1(RP)	GCAACTTCCTCCACAATCCTC		Génotypage de la lignée SALK S_005928
S79-1(LP)	TCCAGTCCCTAAGAGAGAGAGAAAG		Génotypage de la lignée SALK S_005928
S79-2(RP)	TTGATCAGTGTTGTCCTTGCAC		Génotypage de la lignée SALK S_073630
S79-2(LP)	CCTCCCTTTGGTTTGTTTGTC		Génotypage de la lignée SALK S_073630
T-DNAGabi	CCCATTTGGACGTGAATGTAGACAC	Rosso *et al*, 2003	Génotypage des lignées Gabi (LB)
Gabi32(RP)	GAAAGTAAACAAAGCCAGCAG		Génotypage de la lignée Gabi 870F12
Gabi32(LP)	AACATCACGAGTCACCAGTCG		Génotypage de la lignée Gabi 870F12
Sail LB-2	GCTTCCTATTATATCTTCCCAAATTACCAATACA	Sessions *et al*, 2002	Génotypage des lignées SAIL (LB)
Sail32-2(5')	ACCAACCTCAGAAGGGATCTG		Génotypage de la lignée SAIL 548H11
Sail32-2(3')	CCGATCTTTGCTTCGTAGGAG		Génotypage de la lignée SAIL 548H11

Tableau 5 : Liste des amorces utilisées pour la détection des transcrits Pol III et SINE

Nom	Séquence	Utilisation
5.8S-DT	AAAACGACTCTCGGCAAC	RT-PCR
5.8S-Rev	TTGTGACACCCAGGCAC	RT-PCR
Met-P1	GCGCGTAGGTCTCATAGCTA	RT-PCR
Met-Pm	TGGGGTGAGAGAGGCTCG	RT-PCR
tRNA-FOR	GCACCAGTGGTCTAGTGGTAGA	RT-PCR
R43.1-Rev	GAGAATGCATTGGACCCAACCAATAC	RT-PCR
R43.12Rev	TCAGATTATAGCGAGGTTAATATC	RT-PCR
SB1.1	ACCCAGAACCTCGTAGTCCAG	RT-PCR
SB1.2	TTGATTAACCAGGTGTTGACCT	RT-PCR
SB2CONS.1	TGTCGTTAGCTCAATTGGTAAA	RT-PCR
SB2CONS.2	TTTCGTCCCAAGCGTCACTCGAA	RT-PCR
SB4CONS.1	GYTRTTGGYCYAGTGGTAAAT	RT-PCR
SB4CONS.2	ACACCAGRGTTAAATTCAT	RT-PCR

Pour les amorces dégénérées SB4CONS.1 et SB4CONS.2, les nucléotides notées R correspondent à une purine (A ou G) et les nucléotides notées Y à une pyrimidine (C ou T).

Références bibliographiques

Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of Drosophila melanogaster." Science **287**(5461): 2185-95.

Adeniyi-Jones, S. and M. Zasloff (1985). "Transcription, processing and nuclear transport of a B1 Alu RNA species complementary to an intron of the murine alpha-fetoprotein gene." Nature **317**(6032): 81-4.

Aigner, S., J. Postberg, et al. (2003). "The Euplotes La motif protein p43 has properties of a telomerase-specific subunit." Biochemistry **42**(19): 5736-47.

Aleman, C., A. M. Roy-Engel, et al. (2000). "Cis-acting influences on Alu RNA levels." Nucleic Acids Res **28**(23): 4755-61.

Alexandrov, A., I. Chernyakov, et al. (2006). "Rapid tRNA decay can result from lack of nonessential modifications." Mol Cell **21**(1): 87-96.

Alfano, C., D. Sanfelice, et al. (2004). "Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein." Nat Struct Mol Biol **11**(4): 323-9.

Ali, N., G. J. Pruijn, et al. (2000). "Human La antigen is required for the hepatitis C virus internal ribosome entry site-mediated translation." J Biol Chem **275**(36): 27531-40.

Allen, T. A., S. Von Kaenel, et al. (2004). "The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock." Nat Struct Mol Biol **11**(9): 816-21.

Alonso, J. M., A. N. Stepanova, et al. (2003). "Genome-wide insertional mutagenesis of Arabidopsis thaliana." Science **301**(5633): 653-7.

Aravin, A., D. Gaidatzis, et al. (2006). "A novel class of small RNAs bind to MILI protein in mouse testes." Nature **442**(7099): 203-7.

Aravin, A. A., M. Lagos-Quintana, et al. (2003). "The small RNA profile during Drosophila melanogaster development." Dev Cell **5**(2): 337-50.

Arnaud, P., C. Goubely, et al. (2000). "SINE retroposons can be used in vivo as nucleation centers for de novo methylation." Mol Cell Biol **20**(10): 3434-41.

Arnaud, P., Y. Yukawa, et al. (2001). "Analysis of the SINE S1 Pol III promoter from Brassica; impact of methylation and influence of external sequences." Plant J **26**(3): 295-305.

Arts, G. J., S. Kuersten, et al. (1998). "The role of exportin-t in selective nuclear export of mature tRNAs." Embo J **17**(24): 7430-41.

Ayukawa, K., S. Taniguchi, et al. (2000). "La autoantigen is cleaved in the COOH terminus and loses the nuclear localization signal during apoptosis." J Biol Chem **275**(44): 34465-70.

Bachmann, M., D. Falke, et al. (1989). "Intracellular distribution of the La antigen in CV-1 cells after herpes simplex virus type 1 infection compared with the localization of U small nuclear ribonucleoprotein particles." J Gen Virol **70 (Pt 4)**: 881-91.

Barakat, A., K. Szick-Miranda, et al. (2001). "The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome." Plant Physiol **127**(2): 398-415.

Basso, J., P. Dallaire, et al. (1994). "Evidence for an internal ribosome entry site within the 5' non-translated region of turnip mosaic potyvirus RNA." J Gen Virol **75 (Pt 11)**: 3157-65.

Batzer, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." Nat Rev Genet **3**(5): 370-9.

Bayfield, M. A., T. E. Kaiser, et al. (2007). "Conservation of a masked nuclear export activity of La proteins and its effects on tRNA maturation." Mol Cell Biol **27**(9): 3303-12.

Beggs, J. D. (2005). "Lsm proteins and RNA processing." Biochem Soc Trans **33**(Pt 3): 433-8.

Belsham, G. J. and N. Sonenberg (2000). "Picornavirus RNA translation: roles for cellular proteins." Trends Microbiol **8**(7): 330-5.

Beribisky, A. V., T. J. Tavares, et al. (2007). "The three-dimensional structure of the Moorella thermoacetica selenocysteine insertion sequence RNA hairpin and its interaction with the elongation factor SelB." Rna.

Bertrand, E., F. Houser-Scott, et al. (1998). "Nucleolar localization of early tRNA processing." Genes Dev **12**(16): 2463-8.

Boeke, J. D. (1997). "LINEs and Alus--the polyA connection." Nat Genet **16**(1): 6-7.

Bogenhagen, D. F. and D. D. Brown (1981). "Nucleotide sequences in Xenopus 5S DNA required for transcription termination." Cell **24**(1): 261-70.

Brenet, F., N. Dussault, et al. (2005). "Mammalian peptidylglycine alpha-amidating monooxygenase mRNA expression can be modulated by the La autoantigen." Mol Cell Biol **25**(17): 7505-21.

Broekhuis, C. H., G. Neubauer, et al. (2000). "Detailed analysis of the phosphorylation of the human La (SS-B) autoantigen. (De)phosphorylation does not affect its subcellular distribution." Biochemistry **39**(11): 3023-33.

Calvo, O., R. Cuesta, et al. (1999). "GCD14p, a repressor of GCN4 translation, cooperates with Gcd10p and Lhp1p in the maturation of initiator methionyl-tRNA in Saccharomyces cerevisiae." Mol Cell Biol **19**(6): 4167-81.

Campbell, F. E., Jr. and D. R. Setzer (1992). "Transcription termination by RNA polymerase III: uncoupling of polymerase release from termination signal recognition." Mol Cell Biol **12**(5): 2260-72.

Cao, X., G. Yeo, et al. (2006). "Noncoding RNAs in the mammalian central nervous system." Annu Rev Neurosci **29**: 77-103.

Cardinali, B., C. Carissimi, et al. (2003). "La protein is associated with terminal oligopyrimidine mRNAs in actively translating polysomes." J Biol Chem **278**(37): 35145-51.

Carter, M. S. and P. Sarnow (2000). "Distinct mRNAs that encode La autoantigen are differentially expressed and contain internal ribosome entry sites." J Biol Chem **275**(36): 28301-7.

Casati, P. and V. Walbot (2004). "Crosslinking of ribosomal proteins to RNA in maize ribosomes by UV-B and its effects on translation." Plant Physiol **136**(2): 3319-32.

Castro, J. P. and C. M. Carareto (2004). "Drosophila melanogaster P transposable elements: mechanisms of transposition and regulation." Genetica **121**(2): 107-18.

Chakshusmathi, G., S. D. Kim, et al. (2003). "A La protein requirement for efficient pre-tRNA folding." Embo J **22**(24): 6562-72.

Chang, D. Y. and R. J. Maraia (1993). "A cellular protein binds B1 and Alu small cytoplasmic RNAs in vitro." J Biol Chem **268**(9): 6423-8.

Chesnokov, I. and C. W. Schmid (1996). "Flanking sequences of an Alu source stimulate transcription in vitro by interacting with sequence-specific transcription factors." J Mol Evol **42**(1): 30-6.

Chong, S. S., P. Hu, et al. (2001). "Reconstitution of transcription from the human U6 small nuclear RNA promoter with eight recombinant polypeptides and a partially purified RNA polymerase III complex." J Biol Chem **276**(23): 20727-34.

Chu, W. M., R. Ballard, et al. (1998). "Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR." Mol Cell Biol **18**(1): 58-68.

Cioce, M. and A. I. Lamond (2005). "Cajal bodies: a long history of discovery." Annu Rev Cell Dev Biol **21**: 105-31.

Cloix, C., S. Tutois, et al. (2002). "Analysis of the 5S RNA pool in Arabidopsis thaliana: RNAs are heterogeneous and only two of the genomic 5S loci produce mature 5S RNA." Genome Res **12**(1): 132-44.

Copela, L. A., G. Chakshusmathi, et al. (2006). "The La protein functions redundantly with tRNA modification enzymes to ensure tRNA structural stability." Rna **12**(4): 644-54.

Cost, G. J., Q. Feng, et al. (2002). "Human L1 element target-primed reverse transcription in vitro." Embo J **21**(21): 5899-910.

Cox, D. N., A. Chao, et al. (1998). "A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal." Genes Dev **12**(23): 3715-27.

Cozzarelli, N. R., S. P. Gerrard, et al. (1983). "Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes." Cell **34**(3): 829-35.

Craig, A. W., Y. V. Svitkin, et al. (1997). "The La autoantigen contains a dimerization domain that is essential for enhancing translation." Mol Cell Biol **17**(1): 163-9.

Crosio, C., P. P. Boyl, et al. (2000). "La protein has a positive effect on the translation of TOP mRNAs in vivo." Nucleic Acids Res **28**(15): 2927-34.

Curry, S. and M. R. Conte (2006). "A terminal affair: 3'-end recognition by the human La protein." Trends Biochem Sci **31**(6): 303-5.

Custers, J., S. C. Snepvangers, et al. (1999). "The 35S-CaMV promoter is silent during early embryogenesis but activated during nonembryogenic sporophytic development in microspore culture." Protoplasma **208**: 257-264.

Darzacq, X., B. E. Jady, et al. (2002). "Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs." Embo J **21**(11): 2746-56.

Dechampesme, A. M., O. Koroleva, et al. (1999). "Assembly of 5S ribosomal RNA is required at a specific step of the pre-rRNA processing pathway." J Cell Biol **145**(7): 1369-80.

Deininger, P. L. and M. A. Batzer (1999). "Alu repeats and human disease." Mol Genet Metab **67**(3): 183-93.

Deragon, J. M. and P. Capy (2000). "Impact of transposable elements on the human genome." Ann Med **32**(4): 264-73.

Deragon, J. M., N. Gilbert, et al. (1996). "A transcriptional analysis of the S1Bn (Brassica napus) family of SINE retroposons." Plant Mol Biol **32**(5): 869-78.

Deragon, J. M., B. S. Landry, et al. (1994). "An analysis of retroposition in plants based on a family of SINEs from Brassica napus." J Mol Evol **39**(4): 378-86.

Deragon, J. M. and X. Zhang (2006). "Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers." Syst Biol **55**(6): 949-56.

Dewannieux, M., C. Esnault, et al. (2003). "LINE-mediated retrotransposition of marked Alu sequences." Nat Genet **35**(1): 41-8.

Dinkova, T. D., H. Zepeda, et al. (2005). "Cap-independent translation of maize Hsp101." Plant J **41**(5): 722-31.

Docquier, S., V. Tillemans, et al. (2004). "Nuclear bodies and compartmentalization of pre-mRNA splicing factors in higher plants." Chromosoma **112**(5): 255-66.

Dong, G., G. Chakshusmathi, et al. (2004). "Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch." Embo J **23**(5): 1000-7.

Dorokhov, Y. L., P. A. Ivanov, et al. (2006). "An internal ribosome entry site located upstream of the crucifer-infecting tobamovirus coat protein (CP) gene can be used for CP synthesis in vivo." J Gen Virol **87**(Pt 9): 2693-7.

Dubrovsky, E. B., V. A. Dubrovskaya, et al. (2004). "Drosophila RNase Z processes mitochondrial and nuclear pre-tRNA 3' ends in vivo." Nucleic Acids Res **32**(1): 255-62.

Duvernell, D. D. and B. J. Turner (1998). "Swimmer 1, a new low-copy-number LINE family in teleost genomes with sequence similarity to mammalian L1." Mol Biol Evol **15**(12): 1791-3.

Ehresmann, C., F. Baudin, et al. (1987). "Probing the structure of RNAs in solution." Nucleic Acids Res **15**(22): 9109-28.

Esnault, C., J. Maestre, et al. (2000). "Human LINE retrotransposons generate processed pseudogenes." Nat Genet **24**(4): 363-7.

Espinoza, C. A., T. A. Allen, et al. (2004). "B2 RNA binds directly to RNA polymerase II to repress transcript synthesis." Nat Struct Mol Biol **11**(9): 822-9.

Espinoza, C. A., J. A. Goodrich, et al. (2007). "Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription." Rna **13**(4): 583-96.

Fairley, J. A., T. Kantidakis, et al. (2005). "Human La is found at RNA polymerase III-transcribed genes in vivo." Proc Natl Acad Sci U S A **102**(51): 18350-5.

Fierro-Monti, I. and M. B. Mathews (2000). "Proteins binding to duplexed RNA: one motif, multiple functions." Trends Biochem Sci **25**(5): 241-6.

Fok, V., K. Friend, et al. (2006). "Epstein-Barr virus noncoding RNAs are confined to the nucleus, whereas their partner, the human La protein, undergoes nucleocytoplasmic shuttling." J Cell Biol **173**(3): 319-25.

Franklin, S. E., M. G. Zwick, et al. (1995). "Characterization and partial purification of two pre-tRNA 5'-processing activities from Daucus carrota (carrot) suspension cells." Plant J **7**(4): 553-63.

Gajiwala, K. S. and S. K. Burley (2000). "Winged helix proteins." Curr Opin Struct Biol **10**(1): 110-6.

Gallego, M. E., P. Sirand-Pugnet, et al. (1999). "Positive-negative selection and T-DNA stability in Arabidopsis transformation." Plant Mol Biol **39**(1): 83-93.

Gilbert, N. and D. Labuda (1999). "CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs." Proc Natl Acad Sci U S A **96**(6): 2869-74.

Ginisty, H., H. Sicard, et al. (1999). "Structure and functions of nucleolin." J Cell Sci **112 (Pt 6)**: 761-72.

Girard, A., R. Sachidanandam, et al. (2006). "A germline-specific class of small RNAs binds mammalian Piwi proteins." Nature **442**(7099): 199-202.

Gonzalez-Camacho, F. and F. J. Medina (2006). "The nucleolar structure and the activity of NopA100, a nucleolin-like protein, during the cell cycle in proliferating plant cells." Histochem Cell Biol **125**(1-2): 139-53.

Goodier, J. L. and R. J. Maraia (1998). "Terminator-specific recycling of a B1-Alu transcription complex by RNA polymerase III is mediated by the RNA terminus-binding protein La." J Biol Chem **273**(40): 26110-6.

Gottlieb, E. and J. A. Steitz (1989). "Function of the mammalian La protein: evidence for its action in transcription termination by RNA polymerase III." Embo J **8**(3): 851-61.

Gottlieb, E. and J. A. Steitz (1989). "The RNA binding protein La influences both the accuracy and the efficiency of RNA polymerase III transcription in vitro." Embo J **8**(3): 841-50.

Graus, F., C. Cordon-Cardo, et al. (1985). "Immunohistochemical localization of La nuclear antigen in brain. Selective concentration of the La protein in neuronal nucleoli." J Neuroimmunol **9**(5): 307-19.

Grivna, S. T., E. Beyret, et al. (2006). "A novel class of small RNAs in mouse spermatogenic cells." Genes Dev **20**(13): 1709-14.

Hallegger, M., A. Taschner, et al. (2006). "RNA aptamers binding the double-stranded RNA-binding domain." Rna **12**(11): 1993-2004.

Hamilton, T. L., M. Stoneley, et al. (2006). "TOPs and their regulation." Biochem Soc Trans **34**(Pt 1): 12-6.

Henderson, I. R. and S. E. Jacobsen (2007). "Epigenetic inheritance in plants." Nature **447**(7143): 418-24.

Henikoff, S., T. Furuyama, et al. (2004). "Histone variants, nucleosome assembly and epigenetic inheritance." Trends Genet **20**(7): 320-6.

Holcik, M. and R. G. Korneluk (2000). "Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: role of La autoantigen in XIAP translation." Mol Cell Biol **20**(13): 4648-57.

Hopper, A. K. and E. M. Phizicky (2003). "tRNA transfers to the limelight." Genes Dev **17**(2): 162-80.

Horke, S., K. Reumann, et al. (2004). "Nuclear trafficking of La protein depends on a newly identified nucleolar localization signal and the ability to bind RNA." J Biol Chem **279**(25): 26563-70.

Huang, Y., M. A. Bayfield, et al. (2006). "Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation." Nat Struct Mol Biol **13**(7): 611-8.

Huang, Y. and R. J. Maraia (2001). "Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human." Nucleic Acids Res **29**(13): 2675-90.

Huettel, B., T. Kanno, et al. (2007). "RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants." Biochim Biophys Acta **1769**(5-6): 358-74.

Inada, M. and C. Guthrie (2004). "Identification of Lhp1p-associated RNAs by microarray analysis in Saccharomyces cerevisiae reveals association with coding and noncoding RNAs." Proc Natl Acad Sci U S A **101**(2): 434-9.

Intine, R. V., M. Dundr, et al. (2002). "Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs." Mol Cell **9**(5): 1113-23.

Intine, R. V., M. Dundr, et al. (2004). "Nonphosphorylated human La antigen interacts with nucleolin at nucleolar sites involved in rRNA biogenesis." Mol Cell Biol **24**(24): 10894-904.

Intine, R. V., A. L. Sakulich, et al. (2000). "Control of transfer RNA maturation by phosphorylation of the human La antigen on serine 366." Mol Cell **6**(2): 339-48.

Intine, R. V., S. A. Tenenbaum, et al. (2003). "Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs." Mol Cell **12**(5): 1301-7.

Jacks, A., J. Babon, et al. (2003). "Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element." Structure **11**(7): 833-43.

Jady, B. E., X. Darzacq, et al. (2003). "Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm." Embo J **22**(8): 1878-88.

Jelinek, W. R., T. P. Toomey, et al. (1980). "Ubiquitous, interspersed repeated sequences in mammalian genomes." Proc Natl Acad Sci U S A **77**(3): 1398-402.

Jurka, J. (1997). "Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons." Proc Natl Acad Sci U S A **94**(5): 1872-7.

Kadaba, S., A. Krueger, et al. (2004). "Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae." Genes Dev **18**(11): 1227-40.

Kadaba, S., X. Wang, et al. (2006). "Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA." Rna **12**(3): 508-21.

Kajikawa, M. and N. Okada (2002). "LINEs mobilize SINEs in the eel through a shared 3' sequence." Cell **111**(3): 433-44.

Kapitonov, V. V. and J. Jurka (2003). "A novel class of SINE elements derived from 5S rRNA." Mol Biol Evol **20**(5): 694-702.

Kapitonov, V. V. and J. Jurka (2007). "Helitrons on a roll: eukaryotic rolling-circle transposons." Trends Genet **23**(10): 521-9.

Kim, J., J. A. Martignetti, et al. (1994). "Rodent BC1 RNA gene as a master gene for ID element amplification." Proc Natl Acad Sci U S A **91**(9): 3607-11.

Kim, Y. K., S. H. Back, et al. (2001). "La autoantigen enhances translation of BiP mRNA." Nucleic Acids Res **29**(24): 5009-16.

Kim, Y. K. and S. K. Jang (1999). "La protein is required for efficient translation driven by encephalomyocarditis virus internal ribosomal entry site." J Gen Virol **80 (Pt 12)**: 3159-66.

Kimura, R. H., P. V. Choudary, et al. (1999). "Silk worm Bm1 SINE RNA increases following cellular insults." Nucleic Acids Res **27**(16): 3380-7.

Kinoshita, Y., H. Saze, et al. (2007). "Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats." Plant J **49**(1): 38-45.

Kiss, T., C. Marshallsay, et al. (1992). "7-2/MRP RNAs in plant and mammalian cells: association with higher order structures in the nucleolus." Embo J **11**(10): 3737-46.

Kramerov, D. A., I. V. Lekakh, et al. (1982). "The sequences homologous to major interspersed repeats B1 and B2 of mouse genome are present in mRNA and small cytoplasmic poly(A) + RNA." Nucleic Acids Res **10**(23): 7477-91.

Kramerov, D. A. and N. S. Vassetzky (2005). "Short retroposons in eukaryotic genomes." Int Rev Cytol **247**: 165-221.

Kremerskothen, J., M. Nettermann, et al. (1998). "Identification of human autoantigen La/SS-B as BC1/BC200 RNA-binding protein." DNA Cell Biol **17**(9): 751-9.

Kruszka, K., F. Barneche, et al. (2003). "Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z." Embo J **22**(3): 621-32.

Kufel, J., C. Allmang, et al. (2000). "Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding." Mol Cell Biol **20**(15): 5415-24.

Kufel, J., C. Allmang, et al. (2003). "A complex pathway for 3' processing of the yeast U3 snoRNA." Nucleic Acids Res **31**(23): 6788-97.

Kufel, J., C. Allmang, et al. (2002). "Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p." Mol Cell Biol **22**(14): 5248-56.

LaCava, J., J. Houseley, et al. (2005). "RNA degradation by the exosome is promoted by a nuclear polyadenylation complex." Cell **121**(5): 713-24.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lavie, L., J. L. Prieto, et al. (2002). "Bali1A, a Possible LINE Partner for the SINE S1 in Brassica napus." Genome Letters **1**(3): 123-130.

Lenoir, A., L. Lavie, et al. (2001). "The evolutionary origin and genomic organization of SINEs in Arabidopsis thaliana." Mol Biol Evol **18**(12): 2315-22.

Levanon, K., E. Eisenberg, et al. (2005). "Letter from the editor: Adenosine-to-inosine RNA editing in Alu repeats in the human genome." EMBO Rep **6**(9): 831-5.

Lewis, S. M. and M. Holcik (2005). "IRES in distress: translational regulation of the inhibitor of apoptosis proteins XIAP and HIAP2 during cell stress." Cell Death Differ **12**(6): 547-53.

Li, C. F., O. Pontes, et al. (2006). "An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana." Cell **126**(1): 93-106.

Lin-Marq, N. and S. G. Clarkson (1995). "A yeast RNA binding protein that resembles the human autoantigen La." J Mol Biol **245**(2): 81-5.

Lin-Marq, N. and S. G. Clarkson (1998). "Efficient synthesis, termination and release of RNA polymerase III transcripts in Xenopus extracts depleted of La protein." Embo J **17**(7): 2033-41.

Long, K. S., T. Cedervall, et al. (2001). "Phosphorylation of the Saccharomyces cerevisiae La protein does not appear to be required for its functions in tRNA maturation and nascent RNA stabilization." Rna **7**(11): 1589-602.

Lorkovic, Z. J., J. Hilscher, et al. (2004). "Use of fluorescent protein tags to study nuclear organization of the spliceosomal machinery in transiently transformed living plant cells." Mol Biol Cell **15**(7): 3233-43.

Lund, E. and J. E. Dahlberg (1998). "Proofreading and aminoacylation of tRNAs before export from the nucleus." Science **282**(5396): 2082-5.

Lunyak, V. V., G. G. Prefontaine, et al. (2007). "Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis." Science **317**(5835): 248-51.

Maraia, R. and J. Sarrowa (1995). " Alu-Family SINE RNA : interacting proteins and pathways of expression." The Impact of Short Interspersed Elements (SINEs) on the Host Genome: 163-196.

Maraia, R. J. (1991). "The subset of mouse B1 (Alu-equivalent) sequences expressed as small processed cytoplasmic transcripts." Nucleic Acids Res **19**(20): 5695-702.

Maraia, R. J. (1996). "Transcription termination factor La is also an initiation factor for RNA polymerase III." Proc Natl Acad Sci U S A **93**(8): 3383-7.

Maraia, R. J. and M. A. Bayfield (2006). "The La protein-RNA complex surfaces." Mol Cell **21**(2): 149-52.

Maraia, R. J., D. Y. Chang, et al. (1992). "The RNA polymerase III terminator used by a B1-Alu element can modulate 3' processing of the intermediate RNA product." Mol Cell Biol **12**(4): 1500-6.

Maraia, R. J., C. T. Driscoll, et al. (1993). "Multiple dispersed loci produce small cytoplasmic Alu RNA." Mol Cell Biol **13**(7): 4233-41.

Maraia, R. J. and R. V. Intine (2001). "Recognition of nascent RNA by the human La antigen: conserved and divergent features of structure and function." Mol Cell Biol **21**(2): 367-79.

Maris, C., C. Dominguez, et al. (2005). "The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression." Febs J **272**(9): 2118-31.

Marker, C., A. Zemann, et al. (2002). "Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana." Curr Biol **12**(23): 2002-13.

Martin, S. L. (2006). "The ORF1 Protein Encoded by LINE-1: Structure and Function During L1 Retrotransposition." J Biomed Biotechnol **2006**(1): 45621.

Matera, A. G. and K. B. Shpargel (2006). "Pumping RNA: nuclear bodybuilding along the RNP pipeline." Curr Opin Cell Biol **18**(3): 317-24.

Mathews, D. H., J. Sabina, et al. (1999). "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure." J Mol Biol **288**(5): 911-40.

Matsuzaki, H., G. A. Kassavetis, et al. (1994). "Analysis of RNA chain elongation and termination by Saccharomyces cerevisiae RNA polymerase III." J Mol Biol **235**(4): 1173-92.

Mattioli, M. and M. Reichlin (1974). "Heterogeneity of RNA protein antigens reactive with sera of patients with systemic lupus erythematosus. Description of a cytoplasmic nonribosomal antigen." Arthritis Rheum **17**(4): 421-9.

Mayer, M., S. Schiffer, et al. (2000). "tRNA 3' processing in plants: nuclear and mitochondrial activities differ." Biochemistry **39**(8): 2096-105.

Mayes, A. E., L. Verdone, et al. (1999). "Characterization of Sm-like proteins in yeast and their association with U6 snRNA." Embo J **18**(15): 4321-31.

Meerovitch, K., Y. V. Svitkin, et al. (1993). "La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate." J Virol **67**(7): 3798-807.

Meyuhas, O. (2000). "Synthesis of the translational apparatus is regulated at the translational level." Eur J Biochem **267**(21): 6321-30.

Moran, J. V., S. E. Holmes, et al. (1996). "High frequency retrotransposition in cultured mammalian cells." Cell **87**(5): 917-27.

Muddashetty, R., T. Khanam, et al. (2002). "Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles." J Mol Biol **321**(3): 433-45.

O'Connor, J. P. and C. L. Peebles (1991). "In vivo pre-tRNA processing in Saccharomyces cerevisiae." Mol Cell Biol **11**(1): 425-39.

Ohshima, K., M. Hamada, et al. (1996). "The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements." Mol Cell Biol **16**(7): 3756-64.

Okada, N. and M. Hamada (1997). "The 3' ends of tRNA-derived SINEs originated from the 3' ends of LINEs: a new example from the bovine genome." J Mol Evol **44 Suppl 1**: S52-6.

Pannone, B. K., S. D. Kim, et al. (2001). "Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein." Genetics **158**(1): 187-96.

Pannone, B. K., D. Xue, et al. (1998). "A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts." Embo J **17**(24): 7442-53.

Paulson, K. E. and C. W. Schmid (1986). "Transcriptional inactivity of Alu repeats in HeLa cells." Nucleic Acids Res **14**(15): 6145-58.

Pelissier, T., C. Bousquet-Antonelli, et al. (2004). "Synthesis and processing of tRNA-related SINE transcripts in Arabidopsis thaliana." Nucleic Acids Res **32**(13): 3957-66.

Pelisson, A., S. U. Song, et al. (1994). "Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene." Embo J **13**(18): 4401-11.

Pellizzoni, L., B. Cardinali, et al. (1996). "A Xenopus laevis homologue of the La autoantigen binds the pyrimidine tract of the 5' UTR of ribosomal protein mRNAs in vitro: implication of a protein factor in complex formation." J Mol Biol **259**(5): 904-15.

Piccinelli, P., M. A. Rosenblad, et al. (2005). "Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes." Nucleic Acids Res **33**(14): 4485-95.

Pontes, O., C. F. Li, et al. (2006). "The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center." Cell **126**(1): 79-92.

Pontier, D., G. Yahubyan, et al. (2005). "Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis." Genes Dev **19**(17): 2030-40.

Pontvianne, F., I. Matia, et al. (2007). "Characterization of AtNUC-L1 reveals a central role of nucleolin in nucleolus organization and silencing of AtNUC-L2 gene in Arabidopsis." Mol Biol Cell **18**(2): 369-79.

Powers, T. and P. Walter (1999). "Regulation of ribosome biogenesis by the rapamycin-sensitive TOR-signaling pathway in Saccharomyces cerevisiae." Mol Biol Cell **10**(4): 987-1000.

Prak, E. T. and H. H. Kazazian, Jr. (2000). "Mobile elements and the human genome." Nat Rev Genet **1**(2): 134-44.

Ramos, A., S. Grunert, et al. (2000). "RNA recognition by a Staufen double-stranded RNA-binding domain." Embo J **19**(5): 997-1009.

Ray, P. S. and S. Das (2002). "La autoantigen is required for the internal ribosome entry site-mediated translation of Coxsackievirus B3 RNA." Nucleic Acids Res **30**(20): 4500-8.

Reinhart, B. J., E. G. Weinstein, et al. (2002). "MicroRNAs in plants." Genes Dev **16**(13): 1616-26.

Reinisch, K. M. and S. L. Wolin (2007). "Emerging themes in non-coding RNA quality control." Curr Opin Struct Biol **17**(2): 209-14.

Richards, E. J. and S. C. Elgin (2002). "Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects." Cell **108**(4): 489-500.

Rinke, J. and J. A. Steitz (1982). "Precursor molecules of both human 5S ribosomal RNA and transfer RNAs are bound by a cellular protein reactive with anti-La lupus antibodies." Cell **29**(1): 149-59.

Rosenblum, J. S., L. F. Pemberton, et al. (1997). "A nuclear import pathway for a protein involved in tRNA maturation." J Cell Biol **139**(7): 1655-61.

Rosenblum, J. S., L. F. Pemberton, et al. (1998). "Nuclear import and the evolution of a multifunctional RNA-binding protein." J Cell Biol **143**(4): 887-99.

Rosso, M. G., Y. Li, et al. (2003). "An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics." Plant Mol Biol **53**(1-2): 247-59.

Rubin, C. M., R. H. Kimura, et al. (2002). "Selective stimulation of translational expression by Alu RNA." Nucleic Acids Res **30**(14): 3253-61.

Rutjes, S. A., P. J. Utz, et al. (1999). "The La (SS-B) autoantigen, a key protein in RNA biogenesis, is dephosphorylated and cleaved early during apoptosis." Cell Death Differ **6**(10): 976-86.

Ryter, J. M. and S. C. Schultz (1998). "Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA." Embo J **17**(24): 7505-13.

Schiffer, S., S. Rosch, et al. (2002). "Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes." Embo J **21**(11): 2769-77.

Schmid, C. W. (1998). "Does SINE evolution preclude Alu function?" Nucleic Acids Res **26**(20): 4541-50.

Scholl, R. L., S. T. May, et al. (2000). "Seed and molecular resources for Arabidopsis." Plant Physiol **124**(4): 1477-80.

Schwartz, E. I., R. V. Intine, et al. (2004). "CK2 is responsible for phosphorylation of human La protein serine-366 and can modulate rpL37 5'-terminal oligopyrimidine mRNA metabolism." Mol Cell Biol **24**(21): 9580-91.

Serganov, A., A. Rak, et al. (1997). "Ribosomal protein S15 from Thermus thermophilus--cloning, sequencing, overexpression of the gene and RNA-binding properties of the protein." Eur J Biochem **246**(2): 291-300.

Sessions, A., E. Burke, et al. (2002). "A high-throughput Arabidopsis reverse genetics system." Plant Cell **14**(12): 2985-94.

Shaaban, S. A., E. V. Bobkova, et al. (1996). "In vitro analysis of elongation and termination by mutant RNA polymerases with altered termination behavior." Mol Cell Biol **16**(11): 6468-76.

Shaheen, H. H. and A. K. Hopper (2005). "Retrograde movement of tRNAs from the cytoplasm to the nucleus in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **102**(32): 11290-5.

Shama, S. and O. Meyuhas (1996). "The translational cis-regulatory element of mammalian ribosomal protein mRNAs is recognized by the plant translational apparatus." Eur J Biochem **236**(2): 383-8.

Shaw, P. J., A. F. Beven, et al. (1998). "Localization and processing from a polycistronic precursor of novel snoRNAs in maize." J Cell Sci **111 (Pt 15)**: 2121-8.

Shaw, P. J. and J. W. Brown (2004). "Plant nuclear bodies." Curr Opin Plant Biol **7**(6): 614-20.

Shen, W. H., Y. Parmentier, et al. (2002). "Null mutation of AtCUL1 causes arrest in early embryogenesis in Arabidopsis." Mol Biol Cell **13**(6): 1916-28.

Shiroki, K., T. Isoyama, et al. (1999). "Intracellular redistribution of truncated La protein produced by poliovirus 3Cpro-mediated cleavage." J Virol **73**(3): 2193-200.

Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-85.

Sobel, S. G. and S. L. Wolin (1999). "Two yeast La motif-containing proteins are RNA-binding proteins that associate with polyribosomes." Mol Biol Cell **10**(11): 3849-62.

Sorek, R. (2007). "The birth of new exons: Mechanisms and evolutionary consequences." Rna **13**(10): 1603-8.

Spriggs, K. A., M. Bushell, et al. (2005). "Internal ribosome entry segment-mediated translation during apoptosis: the role of IRES-trans-acting factors." Cell Death Differ **12**(6): 585-91.

Stanek, D. and K. M. Neugebauer (2006). "The Cajal body: a meeting place for spliceosomal snRNPs in the nuclear maze." Chromosoma **115**(5): 343-54.

Stange, N., H. J. Gross, et al. (1988). "Wheat germ splicing endonuclease is highly specific for plant pre-tRNAs." Embo J **7**(12): 3823-8.

Stefano, J. E. (1984). "Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts." Cell **36**(1): 145-54.

Stefl, R., L. Skrisovska, et al. (2005). "RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle." EMBO Rep **6**(1): 33-8.

Takano, A., T. Endo, et al. (2005). "tRNA actively shuttles between the nucleus and cytosol in yeast." Science **309**(5731): 140-2.

Tariq, M. and J. Paszkowski (2004). "DNA and histone methylation in plants." Trends Genet **20**(6): 244-51.

Tatout, C., S. Warwick, et al. (1999). "SINE Insertions as Clade Markers for Wild Crucifer Species " Mol Biol Evol **16**(11): 1614-21.

Teplova, M., Y. R. Yuan, et al. (2006). "Structural basis for recognition and sequestration of UUU(OH) 3' temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen." Mol Cell **21**(1): 75-85.

The Arabidopsis Genome Initiative (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana." Nature **408**(6814): 796-815.

Tran, R. K., D. Zilberman, et al. (2005). "Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in Arabidopsis." Genome Biol **6**(11): R90.

Troster, H., T. E. Metzger, et al. (1994). "One gene, two transcripts: isolation of an alternative transcript encoding for the autoantigen La/SS-B from a cDNA library of a patient with primary Sjogrens' syndrome." J Exp Med **180**(6): 2059-67.

Uwer, U., L. Willmitzer, et al. (1998). "Inactivation of a glycyl-tRNA synthetase leads to an arrest in plant embryo development." Plant Cell **10**(8): 1277-94.

Vagner, S., B. Galy, et al. (2001). "Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites." EMBO Rep **2**(10): 893-8.

Valadkhan, S. (2005). "snRNAs as the catalysts of pre-mRNA splicing." Curr Opin Chem Biol **9**(6): 603-8.

Van Arsdell, S. W., R. A. Denison, et al. (1981). "Direct repeats flank three small nuclear RNA pseudogenes in the human genome." Cell **26**(1 Pt 1): 11-7.

Van Horn, D. J., C. J. Yoo, et al. (1997). "The La protein in Schizosaccharomyces pombe: a conserved yet dispensable phosphoprotein that functions in tRNA maturation." Rna **3**(12): 1434-43.

van Tol, H., N. Stange, et al. (1987). "A human and a plant intron-containing tRNATyr gene are both transcribed in a HeLa cell extract but spliced along different pathways." Embo J **6**(1): 35-41.

Vaucheret, H. (2005). "RNA polymerase IV and transcriptional silencing." Nat Genet **37**(7): 659-60.

Verheggen, C., D. L. Lafontaine, et al. (2002). "Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments." Embo J **21**(11): 2736-45.

Wang, H., A. Iacoangeli, et al. (2005). "Dendritic BC1 RNA in translational control mechanisms." J Cell Biol **171**(5): 811-21.

Wang, H., A. Iacoangeli, et al. (2002). "Dendritic BC1 RNA: functional role in regulation of translation initiation." J Neurosci **22**(23): 10232-41.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Watson, J. B., D. W. Chandler, et al. (1984). "Specific termination of in vitro transcription by calf thymus RNA polymerase III." Nucleic Acids Res **12**(13): 5369-84.

Weser, S., M. Bachmann, et al. (2000). "Transcription efficiency of human polymerase III genes in vitro does not depend on the RNP-forming autoantigen La." Nucleic Acids Res **28**(20): 3935-42.

West, N., A. M. Roy-Engel, et al. (2002). "Shared protein components of SINE RNPs." J Mol Biol **321**(3): 423-32.

Westermann, S. and K. Weber (2000). "Cloning and recombinant expression of the La RNA-binding protein from Trypanosoma brucei." Biochim Biophys Acta **1492**(2-3): 483-7.

Wolin, S. L. and T. Cedervall (2002). "The La protein." Annu Rev Biochem **71**: 375-403.

Wolin, S. L. and A. G. Matera (1999). "The trials and travels of tRNA." Genes Dev **13**(1): 1-10.

Wu, H., A. Henras, et al. (2004). "Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III." Proc Natl Acad Sci U S A **101**(22): 8307-12.

Xue, D., D. A. Rubinson, et al. (2000). "U snRNP assembly in yeast involves the La protein." Embo J **19**(7): 1650-60.

Ye, F. and E. R. Signer (1996). "RIGS (repeat-induced gene silencing) in Arabidopsis is transcriptional and alters chromatin configuration." Proc Natl Acad Sci U S A **93**(20): 10881-6.

Yong, J., L. Wan, et al. (2004). "Why do cells need an assembly machine for RNA-protein complexes?" Trends Cell Biol **14**(5): 226-32.

Yoo, C. J. and S. L. Wolin (1994). "La proteins from Drosophila melanogaster and Saccharomyces cerevisiae: a yeast homolog of the La autoantigen is dispensable for growth." Mol Cell Biol **14**(8): 5412-24.

Yoo, C. J. and S. L. Wolin (1997). "The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors." Cell **89**(3): 393-402.

Yoshihisa, T., C. Ohshima, et al. (2007). "Cytoplasmic splicing of tRNA in Saccharomyces cerevisiae." Genes Cells **12**(3): 285-97.

Yoshihisa, T., K. Yunoki-Esaki, et al. (2003). "Possibility of cytoplasmic pre-tRNA splicing: the yeast tRNA splicing endonuclease mainly localizes on the mitochondria." Mol Biol Cell **14**(8): 3266-79.

Yukawa, Y., H. Fan, et al. (2001). "A tobacco nuclear extract supporting transcription, processing, splicing and modification of plant intron-containing tRNA precursors." Plant J **28**(5): 583-94.

Zalfa, F., S. Adinolfi, et al. (2005). "Fragile X mental retardation protein (FMRP) binds specifically to the brain cytoplasmic RNAs BC1/BC200 via a novel RNA-binding motif." J Biol Chem **280**(39): 33403-10.

Zalfa, F., M. Giorgi, et al. (2003). "The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses." Cell **112**(3): 317-27.

Zhang, Y. Q. and K. Broadie (2005). "Fathoming fragile X in fruit flies." Trends Genet **21**(1): 37-45.

Zhu, J., A. Hayakawa, et al. (2001). "Binding of the La autoantigen to the 5' untranslated region of a chimeric human translation elongation factor 1A reporter mRNA inhibits translation in vitro." Biochim Biophys Acta **1521**(1-3): 19-29.

Zilberman, D., X. Cao, et al. (2003). "ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation." Science **299**(5607): 716-9.

Zilberman, D. and S. Henikoff (2004). "Silencing of transposons in plant genomes: kick them when they're down." Genome Biol **5**(12): 249.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-15.