

Université d'Evry Val d'Essonne



Ecole doctorale : « Des génomes aux organismes »

University of Western Australia



School of Biomedical, Biomolecular  
and Chemical Sciences

---

## THESE EN COTUTELLE

présentée pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE D'EVRY VAL D'ESSONNE

et le grade de

DOCTEUR DE L'UNIVERSITE DE WESTERN AUSTRALIA

Spécialité : Biologie cellulaire et moléculaire

### IDENTIFICATION DU FACTEUR CATALYTIQUE DU PROCESSUS D'EDITION DES ARN DES ORGANITES CHEZ LES PLANTES

Présentée par

**Véronique SALONE**

*Soutenance publique prévue le 18 juillet 2008*

---

Composition du jury

**QUETIER Francis**

Directeur de l'école doctorale et représentant de l'Université d'Evry Val d'Essonne

**LURIN Claire**

Directeur de thèse

**SMALL Ian**

Directeur de thèse et représentant de l'Université de Western Australia

**DUJARDIN Geneviève**

Rapporteur

**GUALBERTO José-Manuel**

Rapporteur

**SCHMITZ-LINNEWEBER**

Rapporteur

**« C'était impossible. Ils ne le savaient pas. Ils l'ont fait. »**

**J. Cocteau**

# *Remerciements*

Je suis très reconnaissante envers mes directeurs de thèse Claire Lurin et Ian Small pour m'avoir confié ce projet de doctorat et m'avoir permis de le mener à bien.

Je remercie Michel Caboche pour son accueil à l'URGV ainsi que, Volker Knoop, Charles Bond et Geneviève Dujardin pour leur collaboration.

Je tiens à remercier également l'ensemble des membres du jury d'avoir accepté de juger ce travail, ainsi que les membres de mon comité de thèse pour leurs conseils avisés.

Je remercie les organismes qui ont participé au financement de mon doctorat : le CNRS, l'INRA, la FRM, l'université d'Orsay et l'université de UWA.

Merci à tous mes collègues d'Evry et Perth qui ont œuvré au bon déroulement des expérimentations, en particulier Laure Heurtevin, Jean Colcombet, Etienne Delannoy, Cathie Colas des Francs-Small, Anne-Laure Chateigner-Boutin, Will Stanley, Andéol Falcon de Longevialle et José Gualberto...

....mais aussi un petit clin d'oeil à Jessica Maciel, Sabine Genet, Jenny Gillett, Jude Moyle, Daniel Wipf, Delphine Jublot, Elodie Piednoir, Kristen Feher, Kristina Kühn et Etienne Meyer...

....et une grande mention toute spéciale à Bénédicte Sturbois, Karoline Eifler, et Anne Bersoult ...

Enfin je tiens sincèrement à remercier les personnes qui m'ont soutenu dans cette aventure et toutes les autres.... Je pense tout particulièrement à mes parents, Michel, Yann, Hélène, Timothée et Pauline; et mes amis Théodora, Xavier, Renaud, Arnold, les zap'l'd'air, les petits et grands du CJNA...

Merci aussi à tous ceux que je n'ai pas nommé mais qui se reconnaîtront....

## Sommaire

Principales abréviations utilisées : .....	7
Introduction .....	9
Partie 1 : .....	10
L'ÉDITION DES ARN .....	10
I. Généralités .....	10
I.1 Définition .....	10
I.2 Différentes espèces ; différents processus .....	11
II L'édition par insertion ou délétion de nucléotides : Exemple chez les trypanosomes .	12
II.1 Généralités .....	12
II.2 Le clivage par une endonucléase .....	13
II.3 La délétion d'uridines par une exonucléase .....	13
II.4 L'ajout d'uridines par une TUTase (pour Terminal RNA UridylTransferase) ..	14
II.5 La ligation .....	14
II.6 Rôle d'une Hélicase .....	15
II.7 Autres protéines du complexe 20S éditosome .....	15
II.8 Autres protéines et complexes impliqués dans l'édition .....	16
II.9 Conclusion .....	17
III L'édition des ARN messagers nucléaires par conversion de nucléotides .....	17
III.1 L'édition C-en-U du transcrit de l'apolipoprotéine B (apoB) chez les vertébrés	18
III.1.1 Le gène apoB produit, par un processus d'édition C-en-U de son transcrit,	
deux protéines différentes .....	18
III.1.2 Spécificité de reconnaissance du nucléotide C <sup>666</sup> à désaminer .....	18
III.1.3 Le complexe d'éditosome .....	19
III.1.4 Pour résumé .....	23
III.2 L'édition A-en-I des ARNm chez les animaux .....	23
III.2.1 Généralités .....	23
III.2.2 La famille de protéines ADARs .....	24
III.2.3 Interaction des protéines ADARs avec leur ARN substrat .....	24
III.2.4 L'activité catalytique des protéines ADARs .....	27
III.2.5 Rôle in vivo des protéines ADARs .....	28
III.2.6 Cas particulier des désaminations A-en-I dans les ARNt .....	32
III.2.7 Bilan général sur les désaminations A-en-I .....	34
III.3 Histoire évolutive des enzymes d'édition des ARN .....	34
III.3.1 La superfamille des désaminases zinc-dépendantes .....	34
III.3.2 Corrélations structure-fonction au sein des désaminases .....	35
IV Le processus d'édition chez les plantes .....	36
IV.1 Généralités (Takenaka et al., 2008) (figure n°20) .....	36
IV.2 Quel avantage évolutif à l'édition des ARN des organites chez les plantes ? .....	37
IV.3 Comment une cytidine spécifique est reconnue dans un transcrit pour être	
modifiée ? .....	39
IV.3.1 Une région d'ARN indispensable (dit « cis-élément ») pour la reconnaissance	
du site d'édition .....	39
IV.3.2 Existence de signaux plus éloignés (séquences « enhancer ») et effets longue	
distance .....	40
IV.3.3 Un facteur agissant en trans dans le processus d'édition .....	41
IV.4 Quelle (s) enzyme(s) sont impliquées dans l'édition chez les plantes .....	42
IV.4.1 Est-ce qu'une cytidine-désaminase est impliquée dans l'édition ? .....	42



IV.4.2 Est-ce qu'une transaminase est impliquée dans l'édition ?.....	43
IV.4.3 Est-ce qu'une CTP synthase est impliquée dans l'édition ? .....	44
IV.4.4 Quelle(s) autre(s) enzyme(s) pourraient être impliquée(s) dans l'édition ?..	44
V Bilan sur l'édition des ARNs.....	45
V.1 Avantages et conséquences de ce mécanisme .....	45
V.2 Origine et évolution de ce processus .....	45
Partie 2 : .....	46
LA FAMILLE DE PROTEINES PPR (pour pentatricopeptide repeat).....	46
I Caractéristiques des protéines PPR.....	46
I.1 Mise en évidence d'une grande famille de protéines adressées aux organites chez Arabidopsis.....	46
I.2 Le motif PPR : une séquence de fixation (spécifique) aux ARN .....	47
I.3 Distribution phylogénétique des protéines PPR et caractérisation de différents motifs.....	48
I.4 Quelques données sur l'expression des protéines PPR chez Arabidopsis .....	48
II Rôle des protéines PPR.....	49
II.1 PPR et initiation de la transcription des ARNm .....	49
II.2 PPR et maturation des ARNm .....	49
II.3 PPR et traduction .....	50
II.4 Rôle de protéines PPR dans des mécanismes de régulation de l'expression des gènes propres aux plantes.....	51
II.5 Principales conclusions quant aux rôles des protéines PPR (figure n°26) .....	53
III Présentation du sujet de thèse : « Identification du facteur catalytique impliqué dans le processus d'édition des ARN des organites chez les plantes » .....	53
Présentation des résultats .....	55
Partie 1: .....	56
DES DONNEES EN FAVEUR DE L'IMPLICATION DU DOMAINE DYW DANS LE MECANISME D'EDITION DES ARN CHEZ LES PLANTES .....	56
I Une hypothèse à tester .....	56
II Deux approches parallèles pour tester notre hypothèse.....	57
II.1 Mise en évidence et caractérisation de similarités entre le site actif de cytidines désaminases et le domaine DYW.....	57
II.1.1 Des conservations de séquences primaires entre le domaine DYW et les désaminases .....	57
II.1.2 Le motif catalytique des CDA est comparable à un motif conservé du domaine DYW.....	57
II.1.3 Des conservations de structures secondaires entre le domaine DYW et les cytidines désaminases .....	58
II.1.4 Bilan.....	58
II.2 Une étude de la distribution du domaine DYW dans le règne végétal.....	58
II.2.1 Une analyse bioinformatique.....	58
II.2.2 Mise au point d'une approche par PCR à l'aide d'amorces dégénérées .....	59
II.2.3 PCR à l'aide d'amorces dégénérées sur l'ADN génomique d'espèces d'intérêts.....	59
II.3 "A hypothesis on the identification of the editing enzyme in plant organelles" ...	60
III Etude complémentaire : recherche de la présence éventuelle du domaine DYW chez les dinoflagellées.....	61
III.1 Contexte de l'étude.....	61
III.2 Recherche de la présence du domaine DYW chez les dinoflagellées.....	61
III.3 Validation de la présence d'une PPR de type DYW chez <i>Karenia brevis</i> .....	62

III.4 Bilan .....	62
IV Discussion .....	63
Partie 2: .....	66
LE CANDIDAT <i>AtDYW1</i> .....	66
I Contexte de l'étude .....	66
II Rôle putatif de <i>AtDYW1</i> : une hypothèse de travail.....	66
III Une annotation corrigée pour le locus <i>At1g47580</i> .....	67
IV Localisation subcellulaire de la protéine <i>AtDYW1</i> .....	68
IV.1 Analyses bioinformatiques.....	68
IV.2 Etude de la localisation subcellulaire de protéines de fusion.....	68
IV.3 Expériences de western blot utilisant l'anticorps anti- <i>AtDYW1</i> .....	68
V Interprétations et Bilan .....	69
Partie 3: .....	71
UNE ETUDE GENETIQUE DU GENE <i>AtDYW1</i> .....	71
I L'approche génétique : les grandes lignes .....	71
II Mutants d'insertion.....	71
II.1 Contexte de l'étude.....	71
II.2 Phénotype macroscopique des plantes de génération T4 des lignées SALK-123655 et SALK-012425 .....	72
II.3 Génotypage des plantes de génération T4 des lignées SALK-123655 et SALK-012425 .....	72
II.3.1 Principe du génotypage .....	72
II.3.2 Résultats des réactions de PCR .....	72
II.4 Que signifie l'absence de plantes homozygotes en génération T4, dans la lignée SALK-012425 ? .....	73
II.5 Discussion sur l'absence de mutants d'insertion.....	74
III Lignées « ARNi ».....	74
III.1 Le projet AGRİKOLA .....	74
III.2 Lignées « ARNi » constitutives .....	75
III.2.1 Etablissement des lignées « ARNi » constitutives.....	75
III.2.2 Sélection et génotypage des lignées .....	76
III.2.3 Caractérisation phénotypique des 12 lignées d'intérêt.....	76
III.2.4 Bilan .....	76
III.3 Lignées « ARNi » inductibles de première génération .....	78
III.3.1 Contexte de l'étude.....	78
III.3.2 Etablissement des lignées « ARNi » inductibles de première génération.....	78
III.3.3 Sélection, génotypage et phénotypage des transformants T1 .....	78
III.3.4 Analyse phénotypique des lignées <i>AtDYW1</i> « ARNi » inductible à la génération T2.....	79
III.3.5 Efficacité de l'induction via la dexaméthasone.....	80
III.3.6 Niveau d'expression du gène <i>AtDYW1</i> dans les lignées « ARNi » inductibles .....	82
III.3.7 Recherche de défauts d'édition dans les lignées « ARNi » inductibles 1 et 4 .....	85
III.3.8 Bilan et interprétations des résultats.....	89
III.4 Lignées « ARNi » inductibles de deuxième génération.....	92
III.4.1 Etablissement des lignées « ARNi » inductibles de deuxième génération ...	92
III.4.2 Génotypage et phénotypage de plantes issues des croisements .....	94
III.4.3 Validation de l'efficacité d'induction à la dexaméthasone .....	94
III.4.4 Niveau d'expression du gène <i>AtDYW1</i> après 21 jours d'induction.....	95

III.4.5 Bilan .....	95
III.5 The Arabidopsis AtDYW1 protein: a key component of RNA editing machinery in chloroplasts (manuscript en préparation) .....	96
IV Mutants surexprimant la protéine AtDYW1 .....	97
IV.1 Contexte de l'étude .....	97
IV.2 Mutant de surexpression de première génération .....	97
IV.2.1 Etablissement des lignées de surexpression .....	97
IV.2.2 Sélection des transformants obtenus .....	97
IV.2.3 Phénotypage et génotypage des 29 lignées en génération T2 .....	99
IV.2.4 Analyse des lignées de surexpression en T3 .....	99
IV.2.5 Bilan et interprétations des résultats .....	101
IV.3 Mutant de surexpression de deuxième génération .....	102
IV.3.1 Etablissement des lignées de surexpression de deuxième génération .....	102
IV.3.2 Sélection et génotypage des transformants obtenus .....	103
IV.3.3 Caractérisation phénotypique des transformants obtenus .....	103
IV.3.4 Etude de l'expression du gène AtDYW1 par des expériences de PCR quantitative sur les plantes de génération T1 .....	103
IV.3.5 Recherche de défauts d'édition sur les plantes de génération T1 : Analyse par PPE .....	104
IV.3.6 Etude de l'expression du gène AtDYW1 : expériences de western blot sur des plantes de génération T2 .....	104
IV.3.7 Bilan et interprétations des résultats .....	105
Partie 4: .....	106
UNE ETUDE « FONCTIONNELLE » DE LA PROTEINE AtDYW1 .....	106
I Les objectifs de l'approche .....	106
II Production de la protéine AtDYW1 en système bactérien .....	106
II.1 Construction utilisée .....	106
II.2 Expression de la protéine rDYW1 en système bactérien .....	107
II.3 Production d'un anticorps dirigé contre la protéine AtDYW1 .....	107
II.4 Optimisation de la production de protéine rDYW1 soluble .....	108
II.5 Purification de la protéine rDYW1 .....	109
II.5.1 Chromatographie d'affinité : utilisation du système Ni-NTA (Quiagen) .....	109
II.5.2 Chromatographie par échange d'ions .....	109
II.5.3 Identification des « contaminants » par spectrométrie de masse .....	110
II.6 Tests préliminaires afin d'évaluer, par spectrophotométrie à dichroïsme circulaire, la « renaturation » de la protéine rDYW1 .....	110
III Production de la protéine rDYW1 en système <i>in vitro</i> .....	110
III.1 Généralités sur la technologie RTS (Rapid Translation System) .....	111
III.2 Production à partir d'un extrait d' <i>E. coli</i> (ou <i>E. coli</i> system) ou de germe de blé ( <i>Wheat Germ system</i> ) .....	111
IV Bilan de la production de la protéine AtDYW1 .....	112
V Une étude « fonctionnelle » de la protéine AtDYW1 .....	113
V.1 Mesure d'une activité cytidine-désaminase par spectrophotométrie .....	113
V.2 Mise en évidence d'une activité cytidine-désaminase sur sondes d'ARN radio- marquées .....	114
V.2.1 Description générale du test utilisé .....	114
V.2.2 Mise en place de ce test pour évaluer une éventuelle activité cytidine- désaminase de la protéine AtDYW1 .....	115
V.2.3 Résultats et interprétations .....	116
V.3 Reconstitution d'un complexe « minimal » d'édition .....	117

V.3.1 Objectifs de l'approche .....	117
V.3.2 Reconstitution d'un complexe « minimal » d'édition dans les mitochondries de levure : .....	117
V.3.3 Reconstitution d'un complexe « minimal » d'édition en cytoplasme de cellules de <i>N. benthamiana</i> .....	118
V.4 Bilan sur la recherche d'une activité cytidine-désaminase pour la protéine AtDYW1 .....	121
Discussion générale.....	123
I Le modèle d'étude : un bref rappel .....	124
II Résultats obtenus en faveur du modèle .....	124
II.1 Des arguments phylogéniques en faveur du modèle .....	124
II.2 Des arguments génétiques en faveur du modèle .....	125
III Des interrogations quant au modèle .....	125
III.1 Le domaine DYW et la protéine AtDYW1 sont-ils des facteurs enzymatiques dans la réaction d'édition ?.....	125
III.2 Le modèle est-il valable pour les deux organites ? .....	126
III.3 Combien de protéines PPR sont impliquées dans l'édition ?.....	127
III.4 Les protéines PPR de type PLS sont –elles des facteurs de spécificité dans le processus d'édition? .....	127
III.5 Combien de séquences « cis- élément » une protéine PPR de type PLS peut-elle reconnaître ? .....	129
IV Perspectives.....	129
IV.1 Des données qui pourraient compléter le manuscrit en préparation .....	129
IV.2 Autre perspective : vers une meilleure compréhension de l'interaction PPR/ARN substrat .....	131
V D'autres pistes de travail .....	132
V.1 Vers une compréhension de l'origine et de l'évolution des protéines PPR, du domaine DYW et de la protéine AtDYW1 .....	132
V.2 Edition des ARNt dans les organites des plantes .....	133
V.3 Edition des ARNm nucléaires? .....	134
V.4 Une expression des protéines PPR et de la protéine AtDYW1 extrêmement régulée dans la cellule végétale .....	134
VI Une « critique » de l'approche menée au cours de mon doctorat .....	136
VI.1 Le danger du modèle.....	136
VI.2 Et si c'était par la fin que tout commençait... ..	137
Matériel et méthodes .....	138
I. Matériel biologique.....	139
I.1 Les espèces végétales.....	139
I.2 Les souches des bactéries.....	139
I.3 Conditions de croissance et milieux de culture.....	140
I.3.1 Culture des plantes <i>Arabidopsis thaliana</i> et <i>Nicotiana benthamiana</i> sur terreau .....	140
I.3.2 Culture des plantes <i>Arabidopsis thaliana</i> in vitro .....	140
I.3.3 Culture des souches de bactéries.....	140
I.4 Autre matériel .....	141
II Méthodes de biologie moléculaire.....	141
II.1 Méthodes relatives à l'ADN .....	141
II.1.1 Extraction d'ADN génomique « rapide » d' <i>Arabidopsis</i> .....	141
II.1.2 Echantillons d'ADN génomique provenant d'autres espèces .....	142
II.1.3 Amplification d'ADN par PCR.....	143

II.1.4	Visualisation des produits PCR par électrophorèse.....	144
II.1.5	Purification de produits PCR en vue d'un clonage ou d'un séquençage.....	145
II.1.6	Clonage de produit PCR.....	145
II.1.7	Transformation bactérienne, sélection et test des colonies transformantes..	150
II.1.8	Tranformation d'Agrobactéries, sélection et test des colonies transformantes .....	150
II.1.9	Extraction et purification de plasmide (minipréparation d'ADN).....	151
II.1.10	Séquençage d'ADN.....	151
II-2	Méthodes relatives à l'ARN.....	152
II.2.1	Extraction d'ARN total de tissus végétaux.....	152
II.2.2	Élimination de l'ADN contaminant des extraits d'ARN totaux.....	152
II.2.3	Réaction de transcription « inverse ».....	153
II.3	Mesure du niveau d'édition.....	153
II.3.1	Crible HRM (pour High Resolution Melting).....	153
II.3.2	Analyse par PPE (pour Poisoned Primer Extension).....	154
III	Méthodes relatives aux protéines.....	156
III.1	Extraction de protéines de tissus végétaux.....	156
III.2	Dosage de protéines.....	156
III.3	Migration des protéines sur gel dénaturant.....	156
III.4	Expériences de western blot.....	157
III.5	Production de la protéine rDYW1 en système bactérien.....	158
III.6	Production de la protéine rDYW1 par le système RTS (Rapid Translation System, Roche).....	159
III.7	Purification de la protéine rDYW1.....	159
III.7.1	Par le système Ni-NTA, pour Nickel-Nitrilotriacetic acid (Qiagen).....	159
III.7.2	Chromatographie d'échange d'ions.....	160
III.8	Production d'anticorps polyclonaux.....	160
III.9	Essais de renaturation de la protéine rDYW1 par dichroïsme circulaire.....	161
III.10	Mesure de l'activité cytidine-désaminase par spectrophotométrie.....	161
III.11	Test d'édition sur sondes d'ARN radiomarquées.....	161
IV	Méthodes de biologie cellulaire.....	163
IV.1	Expression transitoire de protéines en cellules de <i>Nicotiana benthamania</i> par agroinfiltration.....	163
IV.2	Obtention de protoplastes et observations microscopiques.....	163
V	Méthodes propres à Arabidopsis.....	164
V.1	Transformation d'Arabidopsis par « floral deeping ».....	164
V.2	Croisement d'Arabidopsis.....	164
V.3	Etude du phénotype.....	164
V.4	Coloration de grains de pollen, méthode d'Alexander (Alexander, 1969).....	164
V.5	Détection de l'activité GUS par coloration directe de plantules d'Arabidopsis..	165
V.6	Détection de l'activité GUS par fluorimétrie.....	165
VI	Outils bioinformatiques.....	166
Annexes	.....	167
Annexe N°1	: Liste des principaux plasmides et vecteurs d'expression utilisés.....	168
Annexe N°2	: Liste des principales constructions réalisées au cours de mon doctorat .....	169
Annexe N°3	: Liste des principaux oligonucléotides utilisés.....	171
Annexe N°4	: Code d'utilisation des bases dégénérées.....	177
Références Bibliographiques	.....	178

### **Principales abréviations utilisées :**

AC : AntiCorps  
ACF : APOBEC1 Complementation Factor  
ADN : Acide DésoxyriboNucléique  
ADNc : ADN complémentaire  
ADNg : ADN génomique  
AGRIKOLA: Arabidopsis Genomic RNAi Knock-out Line Analysis  
AID: Activation-Induced cytidine Deaminase  
apoB : apolipoprotéine B  
APOBEC1 : apoB mRNA editing catalytic component 1  
ARN : Acide RiboNucléique  
ARNg : ARN guide  
ARNm : ARN messenger  
ARNi : ARN interférent  
ARNt : ARN de transfert  
ARNr : ARN ribosomique  
ADAR : Adenosine DeAminases that act on RNA  
ADAT: Adenosine DeAminases that act on tRNA

BLAST : Basic Local Alignment and Search Tool

CATMA: Complete Arabidopsis Transcriptome MicroArray  
CDA : Cytidine DésAminase  
CMS : "Cytoplasmic Male Sterility", stérilité mâle cytoplasmique

dNTP : désoxyribonucléotide  
Ds-RedII ou RFP : Red Fluorescente Protein  
DYW : domaine C-terminal de certaines protéines PPR

GFP : Green Fluorescente Protein  
GST: Gene-specific Sequence Tag  
GUS :  $\beta$ -glucuronidase

hpRNA : hairpin RNA ou ARN en épingle à cheveu  
HRM: High Resolution Melting

IPTG: Isopropyl-Beta-Thio-Galactoside

miARN : microARN  
MS (1/2MS) : milieu Murashige et Skoog

NAD(P) : forme oxydée de la Nicotinamide Adénine Dinucléotide (Phosphate), NAD(P)H pour la forme réduite  
NCBI : National Center for Biotechnology Information  
Ni-NTA : Nickel-Nitrilotriacetic acid  
NLS : séquence de localisation nucléaire

ORF : Open Reading Frame

PCR : "Polymerase chain reaction", réaction en chaîne par polymérase

PEB : Plant Energy Biology

PEP : "Plastidial Encoded Polymerase", polymérase chloroplastique codée par le génome

Chloroplastique

PLS : sous-famille de protéines PPR

PPE : Poisoned Primer Extension

PPR : Pentatricopeptide Repeat

PTGS : Post-Transcriptional Gene Silencing

QTL : Quantitative Locus Trait

RBD : RNA Binding Domain (domaine de liaison à l'ARN)

RMN : Résonance Magnétique Nucléaire

RIP-CHIP : RNA ImmunoPrecipitation and DNA CHIP hybridization

RRM : RNA Recognition Motifs (motif de reconnaissance de l'ARN)

RT-PCR : réaction de transcription inverse suivie d'une PCR

RTS : Rapid Translation System

S: Svedberg, unité de mesure du taux de sédimentation

siRNA : small interfering RNA

TAIR: The Arabidopsis Information Resource

TGS : Transcriptional Gene Silencing

URGV : Unité de Recherche en Génomique Végétale

UTR : UnTranslated Region (région non traduite)

UV : Ultra-violet

# **Introduction**



## **Partie 1 :**

### **L'ÉDITION DES ARN**

#### **I. Généralités**

##### **I.1 Définition**

Il serait opportun de débiter cette introduction en donnant une définition claire du processus d'édition des ARN, mais c'est aussi un exercice périlleux car le terme d'édition des ARN a été utilisé dans la littérature pour décrire une multitude de processus biochimiques différents et la distinction entre les processus d'édition ou de modification est parfois confuse.

Le terme d'édition des ARN a été utilisé pour la première fois en 1986 pour décrire l'insertion de 4 résidus uridines dans le transcrit mitochondrial *coxII* chez le trypanosome (Benne et al., 1986). La communauté scientifique était sceptique et on a alors pensé que ce mécanisme était sans doute spécifique à ce « drôle » de protozoaire.

Puis, rapidement, l'édition d'ARNm a été décrite chez de nombreux organismes eucaryotes, soit pour expliquer des processus d'insertions ou de délétions de nucléotides (qui altèrent le nombre de nucléotides contenus dans la molécule d'ARN) soit pour décrire des conversions ou des remplacements de nucléotides (qui altèrent l'identité des nucléotides contenus dans la molécule d'ARN).

Plus tard, le terme d'édition des ARN a été utilisé pour décrire des désaminations (le plus fréquemment C-en-U, et A-en-I) survenant dans les ARNt et les ARNr d'organismes eucaryotes et procaryotes, mais aussi des modifications mineures des résidus (comme l'ajout de groupement méthyl). De même la polyadénylation de la partie 3' de certains ARNt est aussi communément appelée processus d'édition des ARN. Enfin, un phénomène d'édition cotranscriptionnel des ARN lié au « patinage » de l'ARN polymérase a également été mis en évidence chez certains virus.

Une polémique s'est alors installée entre ceux qui parlent d'édition « vraie » affectant la séquence des ARNm codant pour des protéines, ceux qui définissent l'édition comme un processus modifiant la séquence nucléotidique d'une molécule d'ARN par rapport à sa séquence génomique et ceux pour qui l'édition des ARN regroupe tous les processus moléculaires qui altèrent l'information contenue dans une molécule d'ARN.

En 1998, Price et Gray ont proposé une définition du processus d'édition des ARN qui est aujourd'hui admise par l'ensemble de la communauté scientifique. Selon eux, « l'édition décrit l'ensemble des processus d'altération des ARN conduisant à un transcrit mature dont la séquence diffère de celle codée par le génome au niveau d'un ou plusieurs nucléotides » (Price and Gray, 1998).

## **I.2 Différentes espèces ; différents processus**

L'édition des ARN a été décrite chez de nombreux organismes tels que les mammifères, les trypanosomes, les champignons myxomycètes, certains virus, les bactéries et les plantes. L'édition des ARNm semble restreinte aux eucaryotes, alors que l'édition des ARN non codants, en particulier les ARNt, semble commune aux procaryotes et aux eucaryotes.

L'édition par insertion ou délétion de nucléotides altère le nombre de résidus contenus dans la molécule d'ARN. Comme nous venons de le décrire, ce processus a été mis en évidence chez les trypanosomes (Benne et al., 1986), mais il a été également décrit chez d'autres organismes comme les champignons myxomycètes, certains virus et de nouveaux exemples sont découverts sans cesse.

Dans ce manuscrit, j'ai choisi de détailler uniquement l'édition par insertion et délétion d'uridines chez les trypanosomes car d'une part, il s'agit d'un exemple historique, et d'autre part, c'est pour ce groupe taxonomique que le fonctionnement de la machinerie d'édition est le mieux caractérisé (voir paragraphe II de l'introduction).

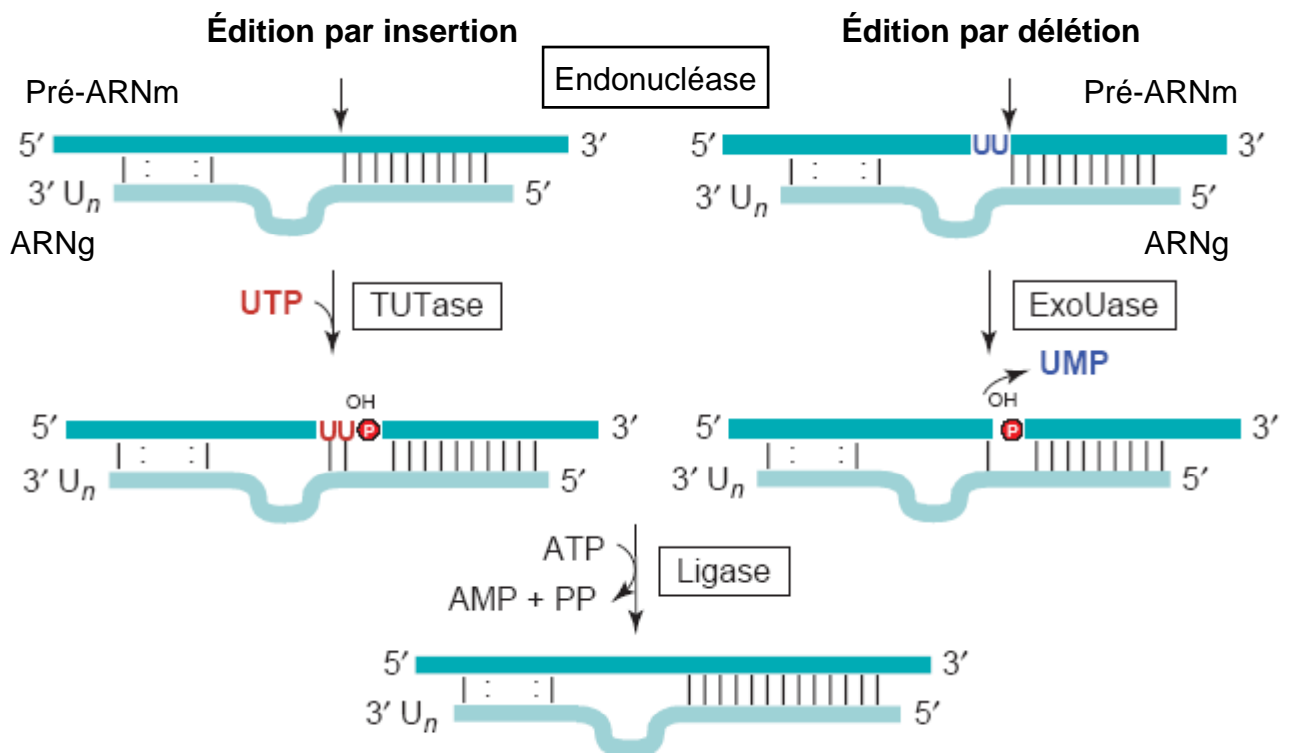
Toutefois, en quelques mots, notons que l'édition chez les champignons myxomycètes, en particulier chez *Physarum polycephalum*, consiste en un mélange de plusieurs processus complexes et est très fréquents dans les ARNm codés par le génome mitochondrial (Horton and Landweber, 2002).

Tout d'abord, l'insertion de tout type de nucléotides (A, C, G et U) sous forme de monomère ou dimère a été décrite. Ce phénomène, bien qu'il reste grandement incompris, pourrait être lié à la machinerie de transcription. Le clonage récent de l'ARN polymérase mitochondriale de *Physarum polycephalum* et sa caractérisation future apporteront sans doute des éléments pour mieux comprendre ce mécanisme (Miller et al., 2006a).

D'autre part, des phénomènes de délétions de nucléotides, de conversions de résidus cytidines en uraciles ou encore le remplacement du nucléotide localisé en 5' de l'ARNm ont été rapportés chez plusieurs espèces de ce groupe. Récemment, ces différentes altérations ont été reproduites dans des tests *in vitro*, ce qui devrait prochainement conduire à une meilleure compréhension de l'édition dans ce groupe (Byrne et al., 2007).

Enfin, certaines espèces combinent les 4 types d'édition. Parfois même, ce qui semble unique chez les eucaryotes, un transcrit est édité par plusieurs mécanismes à la fois. Par exemple, dans le transcrit mitochondrial *coxI*, 66 insertions de nucléotides (dont 59 cytidines, une uridine, et trois dinucléotides différents) ainsi que 4 conversions de résidus cytidines en uraciles, ont été rapportées chez *Physarum polycephalum* (Horton and Landweber, 2000). Chez d'autres espèces de ce groupe un seul type d'édition est observé, comme chez *Clastoderma debaryanum* capable d'édition par insertion d'uridines mais pas d'autres résidus ou de dinucléotides (Horton and Landweber, 2000). Ainsi, l'histoire évolutive de ces 4 formes d'édition au sein de ce groupe taxonomique semble également mystérieuse.

Par ailleurs, l'édition par insertion de nucléotides a été également décrite pour le transcrit de la phosphoprotéine chez les paramyxovirus (aussi nommés NNV pour non segmented negative strand viruses) et le transcrit de la glycoprotéine chez le filovirus de l'ébola (Kolakofsky et al., 2005). Il semble être lié à un « patinage » de l'ARN polymérase, qui d'ordinaire permet l'ajout d'une queue de polyA lors de la réplication de ces virus. L'apparition de ce type d'édition chez ces deux groupes de virus, phylogénétiquement relativement distants, pourrait être survenue indépendamment par un « accident » évolutif.



D'après Stuart *et al.*, 2005

Figure n°1: Mécanisme général d'édition par insertion ou délétion d'uridines chez les trypanosomes

L'édition par conversions ou remplacements de nucléotides altère l'identité des nucléotides contenus dans la molécule d'ARN.

Chez les animaux, l'édition par conversions de nucléotides consiste principalement en la désamination de résidus cytidine en uracile (C-en-U) ou de résidus adénine en inosine (A-en-I). Cependant d'autres modifications ou remplacements de nucléotide, moins largement distribués, existent comme les changements U-en-C, G-en-A ou encore U-en-A. Ces modifications concernent uniquement les ARN nucléaires (ARNm, ARNt et ARNr). Nous détaillerons dans cette introduction, l'édition C-en-U et A-en-I des ARNm, ainsi que de l'édition A-en-I des ARNt (voir paragraphe III de l'introduction).

D'autre part, chez les plantes, l'édition est un processus qui altère uniquement la séquence d'ARNm et ARNt des organites, majoritairement par des conversions C-en-U et moins fréquemment U-en C (voir paragraphe IV de l'introduction).

Enfin, notons que chez de nombreux organismes procaryotes et eucaryotes, l'édition d'ARN non codants, tels que les ARNt et des ARNr, a été décrite. Ce processus consiste en divers types de conversions de nucléotides (le plus fréquemment des désaminations C-en-U et A-en-I). Ces altérations peuvent modifier la structure de ces ARN qui est essentielle à leur fonction biologique. De plus, la modification des bases de l'anticodon permet à un même ARNt de reconnaître plusieurs codons spécifiant un même acide aminé (voir aussi paragraphe III et figure n°15). Pour les ARN non codants, d'autres mécanismes, tels que des modifications mineures des résidus (comme l'ajout de groupement méthyl) ou la polyadénylation de la partie 3' de certains ARNt, ont parfois été appelés « édition » dans la littérature, bien qu'ils s'agissent en réalité de mécanismes de modification des ARN, ceci entraînant une certaine confusion.

**La grande variété d'organismes pour lesquels l'édition des ARN a été décrite, mais aussi la multitude des mécanismes existant, suggèrent que ces processus soient apparus plusieurs fois de façon indépendante au cours de l'évolution. Malgré cette diversité, l'édition des ARN soulève toujours trois questions majeures à savoir, comment les sites d'édition sont-ils spécifiés, comment la réaction d'édition est-elle catalysée et enfin pourquoi ce processus est apparu ?**

## **II L'édition par insertion ou délétion de nucléotides : Exemple chez les trypanosomes**

### **II.1 Généralités**

Chez les trypanosomes, un événement d'édition survient dans le précurseur pré-ARNm de nombreux ARN mitochondriaux. Il s'agit majoritairement d'insertions et plus rarement de délétions d'uridines dans la séquence de ce pré-ARNm. L'édition des transcrits peut avoir des conséquences impressionnantes, comme dans le cas de la sous unité 3 du cytochrome c oxydase (*cox3*) où plus de la moitié du transcrit mature est composée d'uridines insérées de façon post transcriptionnelle.

Le mécanisme général de l'édition par insertions ou délétions de nucléotides chez les trypanosomes est bien connu (figure n°1).

Les ARNm mitochondriaux codent pour des sous-unités des complexes I, III, IV et V de la chaîne respiratoire. Les pré-ARNm sont codés par de larges molécules d'ADN mitochondrial de 2Kb, nommées « maxicircle ». Ces molécules se distinguent des

« minicircles » (~1Kb) qui codent pour les petits ARN guides (ARNg) qui interviennent dans le processus d'édition. Chez *Trypanosoma brucei*, agent responsable de la maladie du sommeil, il existe près de 50 molécules identiques de « maxicircle » et près de 10000 molécules de « minicircle » codant 3 ou 4 ARN guides et constituant un set total supérieur à 1200 ARNg.

L'édition procède en différentes étapes : l'hybridation de l'ARNg et du pré-ARNm, le clivage par une endonucléase, l'ajout d'uridines par une TUTase ou la délétion d'uridines par une exonucléase et enfin une re-ligation (figure n°1).

L'édition est catalysée par un complexe multiprotéique dont la purification a été réalisée chez deux espèces : *Trypanosoma brucei* et *Leishmania tarentolae*. Ce complexe d'une taille de 20 Svedberg a été appelé « 20S éditosome ». L'identification des protéines de ce complexe a été une étape clé dans la compréhension du fonctionnement de la machinerie d'édition chez le trypanosome, comme nous allons le décrire à présent.

## **II.2 Le clivage par une endonucléase**

Le processus d'édition débute par l'hybridation de la partie 5' de l'ARNg avec la partie 3' du pré-ARNm formant ainsi une « ancre » jusqu'à la région qui doit être éditée.

Lors des réactions d'édition *in vitro*, l'endonucléase du complexe d'éditosome reconnaît toujours la première base localisée après l'ancre et effectue une coupure dans le pré-ARNm. Cependant, *in vivo*, la localisation de ce site de coupure est plus variable (Kable et al., 1996; Seiwert et al., 1996; Piller et al., 1997). Il a été proposé qu'en elle-même l'ancre ne soit pas suffisante pour spécifier le site de coupure de l'endonucléase, et que la structure du complexe ARNg/pré-ARNm joue également un rôle (Lawson et al., 2001; Leung and Koslowsky, 2001; Golden and Hajduk, 2006).

Des extraits mitochondriaux purifiés chez *T. brucei* et *L. tarentolae* présentent une activité endonucléase, et d'après les tests d'édition *in vitro*, cette activité est différente dans le cadre d'une édition par délétion ou par insertion de nucléotides (Piller et al., 1997; Alfonzo et al., 1998; Cruz-Reyes et al., 1998; Salavati et al., 2001; Salavati et al., 2002).

Il s'avère que plusieurs protéines, purifiées à partir du complexe d'éditosome de *T. brucei* et *L. tarentolae*, présentent un motif de nucléase. Parmi celles-ci, les protéines KREPB1, KREPB2 et KREPB3, portant toutes trois un motif RNase III et un domaine de fixation à l'ARN double-brin (dsRBD), sont de bonnes candidates pour être les endonucléases de l'éditosome (Worthey et al., 2003).

## **II.3 La délétion d'uridines par une exonucléase**

Lors d'une édition par délétion de nucléotide, une ExoUase enlève les nucléotides uridines contigus et non appariés après le clivage du pré-ARNm par une endonucléase. La protéine KREPC2, purifiée à partir du complexe d'éditosome de *L. tarentolae*, correspond probablement à cette ExoUase (Aphasizhev and Simpson, 2001). En effet, elle contient non seulement en N-terminal un motif 5'-3' exonucléase mais aussi en C-terminal un domaine endonucléase-exonucléase-phosphatase EEP (Worthey et al., 2003). De plus, cette protéine interagit avec les protéines KREL et KREPA2, qui comme nous le verrons par la suite, sont des composants de la machinerie d'édition (Aphasizhev and Simpson, 2001).

#### **II.4 L'ajout d'uridines par une TUTase (pour Terminal RNA UridylylTransferase)**

L'édition par insertion d'uridines s'effectue de l'extrémité 3' vers l'extrémité 5' du pré-ARNm, en utilisant l'ARNg comme matrice, et grâce à une uridylyltransférase (TUTase). Cette activité est favorisée lorsque l'ancre double-brin est riche en purines (ce qui rend également la ligation plus efficace), et inhibée par la présence de pyrimidines dans le pré-ARNm, immédiatement en 5' du site d'édition (Burgess and Stuart, 2000; Igo et al., 2002).

Les TUTases KRET1 et KRET2, purifiées à partir du complexe d'éditosome de *T. brucei* et *L. tarentolae*, présentent un domaine nucléotidyltransférase et un domaine de poly(A) polymérase (PAP). Ces deux protéines appartiennent à la famille des ADN polymérase  $\beta$  (Aphasizhev et al., 2003a). De récentes analyses structurales de ces protéines KRET1 et KRET2 ont permis de montrer qu'elles possèdent effectivement un domaine de fixation de l'UTP et un domaine de fixation à l'ARN (Stagno et al., 2007; Aphasizhev and Aphasizheva, 2008).

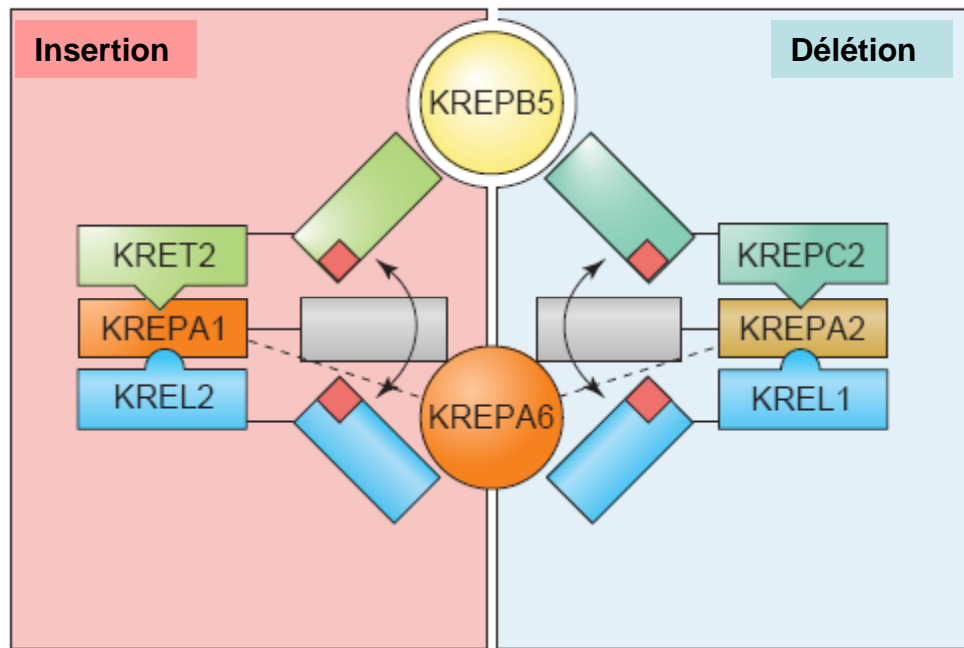
Des lignées « ARNi » affectées dans l'expression des gènes *KRET1* et *KRET2* ont été obtenues chez le trypanosome et leur étude a révélé une diminution de la quantité de transcrits édités (Burgess et al., 1999; Aphasizhev et al., 2003a). Cependant, il apparaît qu'*in vitro*, la présence de KRET2 dans le complexe d'éditosome soit essentielle à l'édition par insertion de nucléotides et que KRET1 puisse ajouter ou retirer des nucléotides (Aphasizhev et al., 2003b; Stagno et al., 2007; Aphasizhev and Aphasizheva, 2008). De plus, KRET2 interagit avec la protéine KREPA1 (composant de la machinerie d'édition) et cette interaction augmente son activité *in vitro* (Ernst et al., 2003; Schnauffer et al., 2003).

Par ailleurs, il a été montré, par des expériences d'« UV-crosslinking », que la protéine KRET1 peut interagir avec des ARNg (Aphasizhev et al., 2002). Ainsi, cette protéine aurait possiblement deux rôles dans l'édition : l'ajout ou la délétion d'uridines et la stabilisation du complexe ARNg/pré-ARNm.

#### **II.5 La ligation**

Deux ARN ligases ont été identifiées dans le complexe d'éditosome 20S : les protéines KREL1 et KREL2. Ces protéines sont très similaires et appartiennent à la super famille des nucléotidyl transférases (Panigrahi et al., 2001; Rusche et al., 2001). De plus, elles présentent des homologies avec l'ARN ligase 2 du bactériophage T4 (Ho et al., 2004). *In vitro*, il a été montré que contrairement aux autres ligases, ces deux enzymes préfèrent les ARN substrats dont l'appariement double-brin n'est pas parfait (Cruz-Reyes and Sollner-Webb, 1996; Kable et al., 1996; Blanc et al., 1999; Igo et al., 2000; Lawson et al., 2001; Cruz-Reyes et al., 2002; Igo et al., 2002).

Des lignées de surexpression d'allèles mutants et des lignées « ARNi » affectant l'expression du gène *KREL1* ont été obtenues et leur étude a révélé que ce gène était essentiel à l'édition (Huang et al., 2001; Schnauffer et al., 2001; Gao and Simpson, 2003). En revanche, l'analyse de lignées « ARNi » du gène *KREL2* n'a pas permis d'identifier un défaut d'édition (Drozd et al., 2002; Gao and Simpson, 2003; O'Hearn et al., 2003). Cependant, il est fort probable que la protéine KREL2 soit tout de même impliquée dans l'édition car il a été montré qu'un sous-complexe protéique contenant la protéine KREL1 est capable d'édition par délétion *in vitro*, et qu'un complexe contenant la protéine KREL2 est capable d'édition par insertion *in*



D'après Stuart *et al.*, 2005

Figure n°2: Interactions entre les protéines du complexe 20S « éditosome »

On distingue deux sous-complexes permettant d'une part l'édition par insertion d'uridines (KREPA1, KREL2, et KRET2); et d'autre part l'édition par délétion d'uridines (KREPA2, KREPC2, et KREL1).

Les protéines KREPA1 et KREPA2 ont un rôle central dans l'organisation de ces sous-complexes.

*vitro* (Schnauffer et al., 2003). Ainsi, dans les lignées « ARNi » affectant l'expression du gène *KREL2* soit l'extinction de ce gène n'est pas suffisante pour affecter l'édition, soit une protéine (par exemple *KREL1*) peut remplacer sa fonction.

La protéine *KREL2* est un bon candidat pour être la ligase intervenant dans l'édition par insertion et *KREL1* dans l'édition par délétion d'uridines. Par ailleurs la protéine *KREL1* pourrait également être impliquée dans les mécanismes de reconnaissance d'ARNg, de réparation des ARN (Koslowsky et al., 1991; Huang et al., 2001). Sa structure a été publiée et présente un site de fixation à l'ATP unique (Deng et al., 2004). Il peut être bloqué par l'action de certaines drogues, ce qui ouvre de nouvelles perspectives thérapeutiques (Amaro et al., 2007).

## **II.6 Rôle d'une Hélicase**

Les ARNg sont utilisés pour éditer les pré-ARNm et doivent être dissociés après la réaction d'édition. Il a été proposé qu'une hélicase appartenant au complexe d'éditosome, la protéine *KREH1*, assure cette fonction (Missel et al., 1997).

L'étude d'un mutant nul pour cette protéine chez *T. brucei* a révélé des individus viables mais dont la croissance cellulaire est ralentie et pour lesquels l'édition est partiellement affectée (Missel et al., 1997). Ainsi cette protéine, bien que non essentielle, pourrait jouer un rôle dans le processus d'édition.

## **II.7 Autres protéines du complexe 20S éditosome**

D'autres protéines, pour lesquelles aucune fonction n'a pu être proposée, ont été identifiées dans le complexe d'éditosome. Parmi celles-ci, six protéines ont été nommées *KREPA1* à *6* (Panigrahi et al., 2001; Worthey et al., 2003). Toutes ces protéines contiennent des domaines en doigts de zinc ou des motifs connus pour être impliqués dans les interactions protéine-ARN ou protéine-protéine.

L'étude d'une lignée « ARNi » affectant l'expression du gène *KREPA1* chez *T. brucei* a révélé des individus à la croissance ralentie et présentant une perte de l'expression du gène *KREL2* et de l'activité d'édition par insertion (Drozd et al., 2002; O'Hearn et al., 2003). De même, l'étude d'une lignée « ARNi » affectant l'expression du gène *KREPA2* chez *T. brucei* a révélé des individus à la croissance ralentie, présentant une perte de l'expression du gène *KREL1* et de l'activité d'édition (Huang et al., 2002). Des lignées « ARNi » pour ces deux gènes a révélé une diminution de l'activité endonucléase de l'édition, et qu'en particulier la *KREPA2* était effectivement essentielle cette l'activité (Drozd et al., 2002; Huang et al., 2002; Law et al., 2008).

Pris ensemble, ces résultats suggèrent que les membres de cette famille sont essentiels à la machinerie d'édition et doivent participer au maintien de la structure et de l'organisation du complexe d'éditosome (figure n°2).



## **II.8 Autres protéines et complexes impliqués dans l'édition**

Le cycle de vie de *T. brucei* consiste en deux étapes : un stade procyclique dans la mouche tsetse, et un stade parasitoïde dans le sang des animaux. Il a été montré que chez *T. brucei*, l'édition varie au cours de ce cycle : ce processus est donc régulé. L'abondance des ARNg ne varie pas au cours du cycle de vie des trypanosomes ; ainsi, cette régulation doit dépendre des mécanismes de reconnaissance et transport des ARNg (Stuart et al., 1997; Cruz-Reyes, 2007). Bien qu'encore largement méconnus, nous allons évoquer à présent différentes protéines et complexes qui pourraient participer à cette régulation.

Les protéines MRP1 et MRP2 forment un complexe MRP qui a été identifié, par des expériences d'« UV-crosslinking », car il présente une haute affinité pour les ARNg (Koller et al., 1997; Blom et al., 2001; Aphasizhev et al., 2003b; Vondruskova et al., 2005).

Il a été montré que les protéines MRP1 et KRET1 (qui comme nous l'avons vu est un composant du complexe d'éditosome) interagissent et cette interaction est sensible à l'action de RNase (Aphasizhev et al., 2003a; Simpson et al., 2004). En outre, l'association de la protéine KRET1 au complexe d'éditosome est également sensible à un traitement à la RNase (Aphasizhev et al., 2002). De plus la protéine MRP1 peut s'associer de façon non-spécifique aux ARN et facilite la formation d'ARN double-brin (Aphasizhev et al., 2003c).

Par ailleurs, l'étude d'un mutant nul pour la protéine MRP1 chez *T. brucei* a révélé des individus viables pour lesquels l'édition n'est que très partiellement affectée (Lambert et al., 1999). En revanche l'analyse d'une lignée « ARNi » affectant l'expression des gènes *MRP1* et *MRP2* chez *T. brucei* a démontré des effets évidents sur l'édition et une réduction dramatique de la quantité de transcrits édités (Vondruskova et al., 2005).

Ainsi, tous ces éléments suggèrent que les protéines KRET1, MRP1 et MRP2 peuvent former un complexe permettant la reconnaissance des ARNg et leur transport dans le complexe d'éditosome.

Trois autres protéines, RBP16, REAP-1 et TbRGG1 peuvent se lier aux ARNg, ARNm et ARNr, mais elles n'appartiennent pas au complexe d'éditosome et ne semblent pas interagir avec des composants de celui-ci (Hayman and Read, 1999; Pelletier et al., 2000).

L'analyse d'une lignée « ARNi » affectant l'expression du gène *RBP16* chez *T. brucei* a révélé une accumulation du transcrit *Cyb* non édité et une diminution du transcrit édité (Pelletier and Read, 2003). Récemment, il a été mis en évidence que cette protéine joue le rôle de protéine chaperon qui stimule, *in vivo* et *in vitro*, la fixation des ARNg sur le pré-ARNm (Miller et al., 2006b; Ammerman et al., 2008).

Le facteur REAP-1 est une protéine qui s'associe préférentiellement aux ARN non édités (Madison-Antenucci et al., 1998; Madison-Antenucci et al., 2002), quant à la protéine TbRGG1, elle présente une affinité pour les ARN riches en uridines (Vanhamme et al., 1998). Le rôle de ces protéines dans l'édition reste à mieux caractériser.

Enfin, la présence d'une queue de polyA en 3' des pré-ARNm les stabilise et ceci paraît indispensable au maintien de l'appariement ARNg/pré-ARNm (Koslowsky et al., 2004; Kao and Read, 2007).

## **II.9 Conclusion**

Tout à fait récemment, l'analyse des séquences non-codantes des gènes nucléaires a permis de révéler l'existence d'un nouveau type d'ARNg capables de reconnaître 29 séquences d'ARNm de gènes non mitochondriaux (Madej et al., 2008). Ainsi l'édition par insertion ou délétion de nucléotide dans ce groupe taxonomique pourrait concerner un plus grand nombre de gènes que ce qui était décrit jusqu'alors.

Nous avons évoqué le fait que l'édition varie au cours du cycle de vie des trypanosomes. En effet, chez *T. brucei*, agent responsable de la maladie du sommeil humaine, l'édition est maximale au stade procyclique dans la mouche tsetse, et réduite au stade parasitoïde dans le sang des animaux, suggérant qu'à ce stade, l'énergie produite par la chaîne respiratoire n'est pas essentielle. Minimiser la taille du génome mitochondrial ne peut pas constituer la « raison d'être » du processus d'édition chez le trypanosome, car la taille totale des « minicircles » produisant les ARNg dépasse celle des « maxicircles » produisant les pré-ARNm. L'édition doit conférer un avantage aux trypanosomes, notamment permettre une régulation de la production d'énergie par la chaîne de respiratoire au cours de son cycle de vie.

L'édition par insertion ou délétion d'uridines dans les ARN mitochondriaux du trypanosome est un type d'édition propre à ce groupe. L'identification de membres du complexe d'éditosome a considérablement aidé à mieux comprendre le fonctionnement de la machinerie d'édition. De nombreuses protéines ainsi caractérisées sont proches des protéines agissant dans les mécanismes de réparation de l'ADN qui impliquent également des étapes de clivage, d'addition ou délétion de nucléotides et de ligation. D'ailleurs toutes ces étapes se déroulent également au sein d'un large complexe. La réparation des ADN est commune à tous les êtres vivants et ces similarités suggèrent l'existence d'un ancêtre commun à la machinerie de ces deux processus.

## **III L'édition des ARN messagers nucléaires par conversion de nucléotides**

Chez les animaux, l'édition par conversion de nucléotides a été décrite au niveau d'un nombre restreint de sites dans les régions codantes de quelques ARNm codés par le génome nucléaire. Actuellement, on découvre que dans les régions non-codantes des ARNm, l'édition pourrait être massive et jouer un rôle fondamental dans des processus de régulation post-transcriptionnelle de l'expression des gènes, mais aussi la défense contre les virus, ou le maintien de l'intégrité du génome.

Deux types majeurs de conversions ont été principalement étudiés : la désamination C-en-U du transcrit codant pour l'apolipoprotéine B chez les vertébrés, et les désaminations A-en-I décrits pour plusieurs transcrits chez de nombreux métazoaires.

### **III.1 L'édition C-en-U du transcrit de l'apolipoprotéine B (apoB) chez les vertébrés**

#### *III.1.1 Le gène apoB produit, par un processus d'édition C-en-U de son transcrit, deux protéines différentes*

Les protéines apoB jouent un rôle central dans le métabolisme des lipides et les mammifères utilisent deux formes protéiques, nommées apoB100 et apoB48, pour transporter le cholestérol et les triglycérides dans le sang. Chez l'Homme, la protéine apoB100 (512kDa) est sécrétée dans le foie, alors que la protéine apoB48 (241kDa) est sécrétée dans l'intestin grêle. Ces deux protéines sont codées par le même gène *apoB* (14kb) (Chen et al., 1987; Powell et al., 1987). En fait, une conversion C-en-U du nucléotide 6666 modifie l'ARNm *apoB* et ceci résulte en le remplacement d'un codon glutamine (CAA) par un codon stop (UAA) prématuré générant ainsi la protéine apo48.

#### *III.1.2 Spécificité de reconnaissance du nucléotide C<sup>6666</sup> à désaminer*

Dans le transcrit *apoB*, la séquence nucléotidique entourant la cytidine à modifier se retrouve extrêmement conservée chez toutes les espèces des marsupiaux à l'Homme ; en revanche, elle ne l'est pas dans d'autres genres où ce transcrit n'est pas édité comme, par exemple, chez les oiseaux (figure n°3B).

Il a été montré qu'une séquence de 26 nucléotides, encadrant la base à modifier, est nécessaire et suffisante à l'édition *in vitro* du transcrit *apoB* (Davies et al., 1989; Driscoll et al., 1989; Shah et al., 1991; Backus and Smith, 1992). Trois éléments essentiels ont été déterminés au sein de ces 26 nucléotides : une séquence de « mooring » ; une séquence « spacer » de 2 à 8 nucléotides séparant la séquence « mooring » de la cytidine à modifier ; et une région riche en AU, en amont de la cytidine à modifier et activant l'édition (région « enhancer » ou « efficiency element ») (Driscoll et al., 1993; Anant et al., 1995; Hersberger et al., 1999; Nakamuta et al., 1999). Par ailleurs, d'autres séquences plus éloignées de la base à modifier peuvent également favoriser le processus d'édition (Hersberger and Innerarity, 1998). L'ensemble de ces régions est représenté sur la figure n°3.

Il apparaît que les séquences conservées autour la cytidine à modifier de l'ARNm *apoB* forment une structure secondaire en « tige-boucle » dans laquelle se trouve cette cytidine C<sup>6666</sup> (Richardson et al., 1998) (figure n°4). Cette structure secondaire semble être un élément déterminant dans le processus d'édition puisqu'elle est reconnue par les facteurs d'édition (Richardson et al., 1998; Anant and Davidson, 2000). Nous verrons, aux paragraphes suivants, qu'en effet il a été montré que non seulement, deux facteurs essentiels au processus d'édition (APOBEC-1 et ACF) reconnaissent cette séquence, mais aussi comment sa structure évolue au cours du processus d'édition.

En résumé, l'édition du transcrit *apoB* au niveau de la cytidine C<sup>6666</sup> se fait avec une très haute spécificité qui dépend non seulement de la structure primaire, mais aussi de la structure secondaire voir tertiaire de cet ARN substrat.

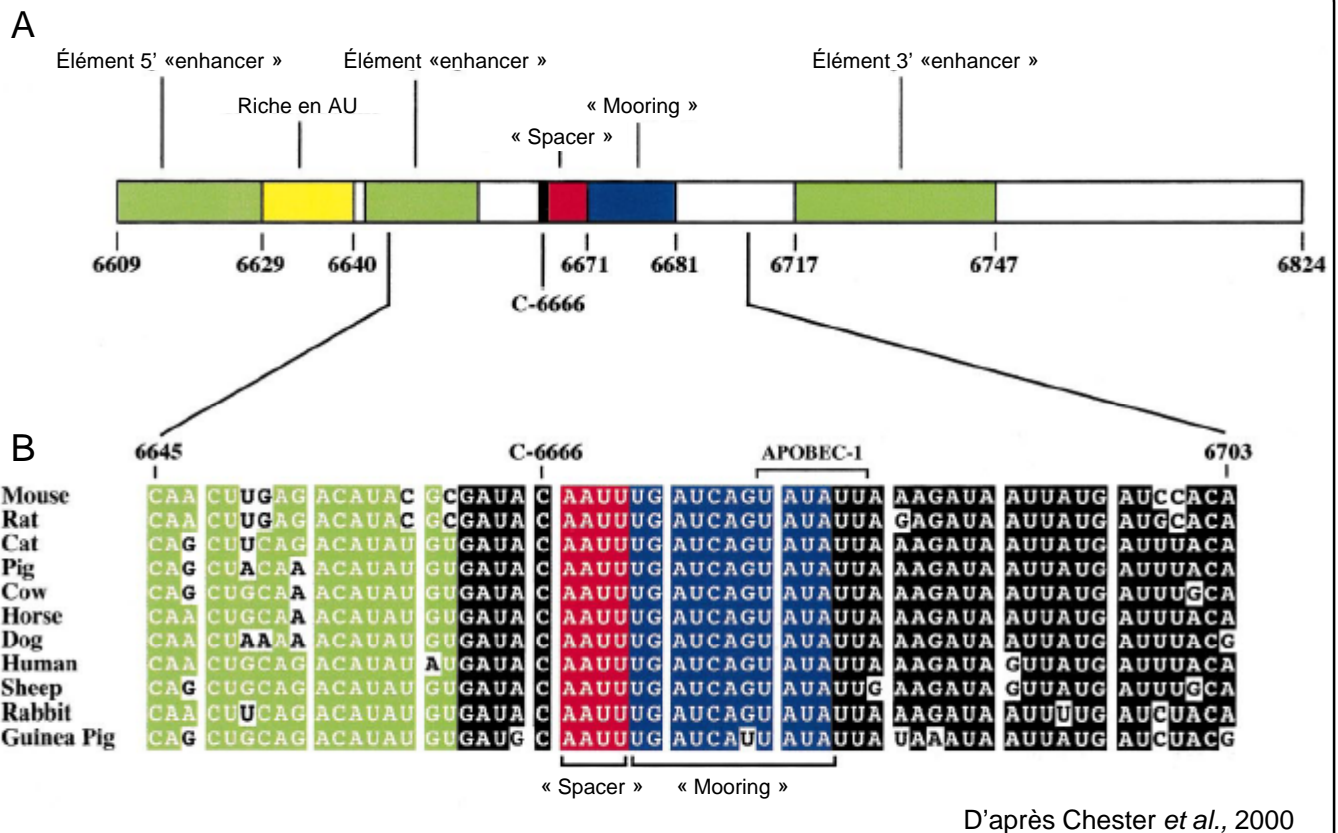


Figure n°3: Organisation des éléments conservés du transcrit *apoB* (A) et alignement multiple des séquences de transcrit *apoB* de différents mammifères (B).

Le développement d'un système d'édition *in vitro* du transcrit *apoB* (Driscoll *et al.*, 1989) et la mutagenèse de sites spécifiques de ce transcrit (Davies *et al.*, 1989 ; Shah *et al.*, 1991 ; Backus *et al.*, 1992) ont permis de déterminer avec précision les résidus essentiels pour assurer la reconnaissance de la base à modifier au sein de ce transcrit. Ainsi une séquence de 26 nucléotides est nécessaire et suffisante à l'édition *in vitro* du transcrit *apoB*, et certains éléments extérieurs à cette séquence peuvent influencer sur l'efficacité d'édition. De plus, une séquence de 11 nucléotides, appelée « mooring » et localisée de 4-5 bases après la cytidine C6666 est essentielle à sa modification.

Par la suite, l'introduction de séquences conservées de « mooring » de mammifères dans le transcrit *apoB*, normalement non édité, du poulet ou la mutagenèse des 26 nucléotides entourant la cytidine C6666 de cette séquence la convertissant ainsi en une séquence similaire à celles des mammifères, a permis de confirmer cette observation (Anant *et al.*, 1995 ; Nakamuta *et al.*, 1999). De la même façon il a été montré que les 2 à 8 nucléotides, espaçant la séquence « mooring » de la cytidine à modifier, sont également indispensables au processus l'édition (Driscoll *et al.*, 1993). Enfin, lors d'une comparaison de la séquence du transcrit *apoB* du cochon d'inde, faiblement édité, et de 31 séquences d'autres espèces pour lesquelles le transcrit *apoB* est fortement édité des variations ont été localisées; et des approches similaires à celle menée pour le transcrit *apoB* du poulet ont permis d'identifier une région activatrice de l'édition (région « enhancer » ou « efficiency element ») en amont de la cytidine à modifier (Hersberger *et al.*, 1999). Cette région est riche en résidus AU. Enfin d'autres séquences plus éloignées de la base à modifier, en 5' et 3', peuvent également favoriser le processus d'édition (Hersberger et Innerarity, 1998).

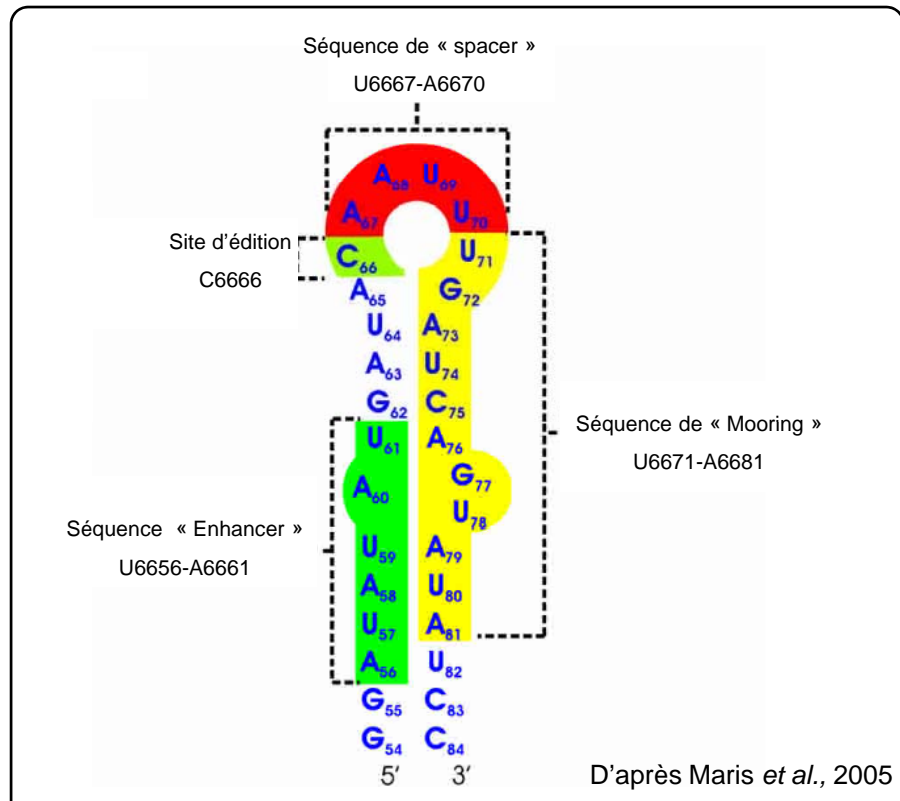
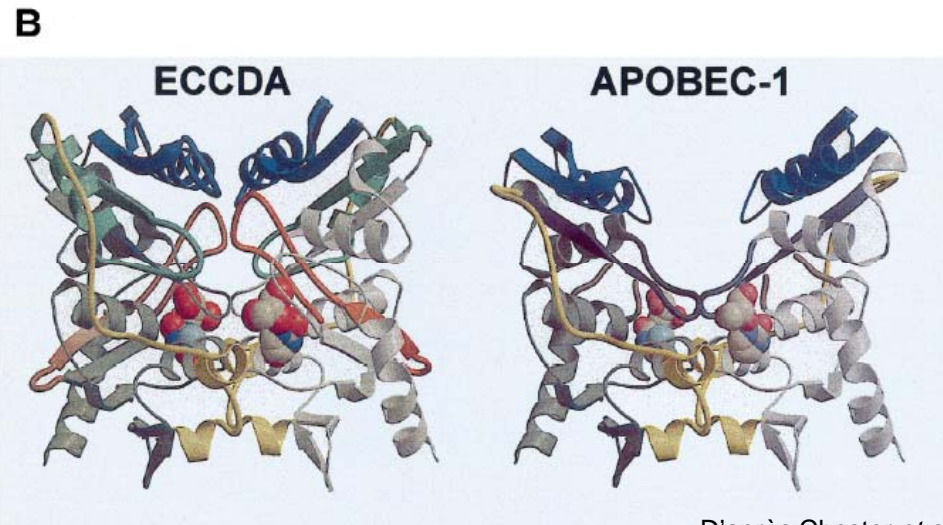
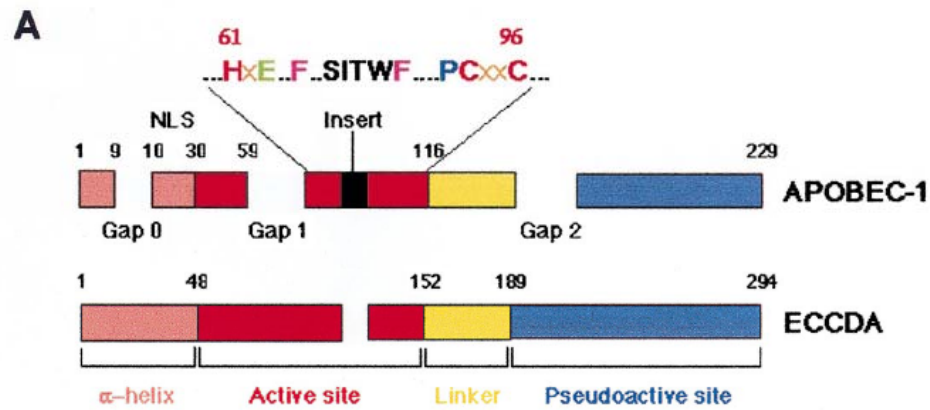


Figure n°4: Représentation de la structure secondaire de la séquence de 31 nucléotides entourant la cytidine C<sup>6666</sup> à modifier dans le transcrit *apoB*.

Cette structure a d'abord été prédite par des analyses bioinformatiques, puis des expériences d'empreinte à la ribonucléase combinées à l'étude biochimique de mutants ont permis de confirmer que les séquences conservées autour la cytidine à modifier de l'ARNm *apoB* pourraient former une structure en « tige-boucle », et montrer que la cytidine C6666 est bien localisée au sein de cette boucle (Richardson *et al.*, 1998).



D'après Chester *et al.*, 2000

Figure n°5: Homologie entre les protéines ECCDA et APOBEC1.

- (A) Il existe une grande conservation de séquences primaires entre la cytidine déaminase d'*E. coli* (ECCDA) agissant sur des mononucléotides, et la protéine APOBEC1, en particulier au niveau du site actif. L'existence d'un motif SITW encadré par deux résidus phénylalanines est une «signature» propre aux protéines APOBEC1. Une mutagenèse extensive du transcrit *apoB* et du site actif de l'enzyme APOBEC-1 associée à des expériences d'UV cross-linking et de compétitions a permis de montrer que les résidus phénylalanines sont impliqués dans la fixation d'APOBEC-1 sur le transcrit *apoB* préférentiellement au niveau de la région riche en AU de la séquence de « mooring » (Navaratnam *et al.*, 1995). De même, il existe 3 espaces ou « gaps » dans le site catalytique d'APOBEC1 absents dans la protéine ECCDA.
- (B) Le cristal de la protéine ECCDA a été obtenu, révélant que chaque monomère de cette enzyme homodimérique fait 31kDa et se compose d'une petite hélice alpha et de deux larges feuillets beta reliés par une séquence dite « flap » de 37 acides aminés (Betts *et al.*, 1994). Une modélisation moléculaire associée à une mutagenèse a permis de proposer une structure tertiaire et quaternaire de la protéine APOBEC1 proche de celle de la protéine ECCDA (Navaratnam *et al.*, 1998). On note simplement que l'arrangement des 3 « gaps » présents dans la protéine APOBEC1 libère un espace de 10kDa qui pourrait correspondre à la masse du transcrit minimal à éditer. Un cristal de la protéine scCDD1, orthologue d'APOBEC1 chez la levure *Saccharomyces cerevisiae* a été obtenu (Xie *et al.*, 2004). Les résultats indiquent que la topologie proposée par Navaratnam *et al.* pour la protéine APOBEC1 est tout à fait plausible, et soulignent que la courte séquence « flap » de cette protéine est impliquée non seulement dans l'oligomérisation d'APOBEC1 et aussi dans la sélection d'un substrat approprié.

### III.1.3 Le complexe d'éditosome

Dès les premières expériences d'édition *in vitro*, il a été montré qu'un large complexe protéique (~1400kDa) s'assemble autour de la base à modifier dans le transcrit *apoB*, suggérant que la machinerie d'édition est composée de nombreux facteurs (Harris et al., 1993). Ce complexe est couramment nommé « éditosome » (Navaratnam et al., 1998). L'identification des composants de ce complexe représente un enjeu fondamental pour comprendre le mécanisme de la réaction d'édition.

#### 1) Le facteur catalytique de la réaction d'édition C-en-U du transcrit *apoB* : la cytidine-désaminase APOBEC1

Une approche de clonage positionnel a permis d'isoler un ADNc codant une protéine de 27 kDa chez le rat : la protéine APOBEC1 (pour *apoB* mRNA editing catalytic component 1) (Teng et al., 1993). Il a été rapidement montré qu'il s'agit de la sous-unité catalytique du complexe d'édition.

Dans un premier temps, il a été décrit que, seule, la protéine APOBEC1 ne permet pas d'éditer *in vitro* le transcrit *apoB*, mais additionnée d'un extrait protéique de poulet ou de cellules HeLa (qui ne contiennent pas d'activité d'édition intrinsèque) l'édition C-en-U est possible (Teng et al., 1993; Davidson et al., 1995; Navaratnam et al., 1995). Cela montre que la protéine APOBEC1 est bien le facteur catalytique de la réaction d'édition, et doit recruter des facteurs auxiliaires non catalytiques pour son activité.

La protéine APOBEC1 est présente et très conservée chez toutes les espèces depuis les marsupiaux jusqu'à l'Homme. De plus, elle présente des homologies avec le site actif de cytidines désaminases fixant du zinc de nombreux organismes agissant sur des substrats monomériques (Scott, 1995) suggérant qu'APOBEC-1 catalyse une désamination de la cytidine C<sup>666</sup> dans le transcrit *apoB*. Des expériences de mutagenèse extensive des protéines APOBEC1 du rat, de l'Homme, de la souris et du lapin ont permis d'identifier les acides aminés essentiels à sa fonction catalytique (Driscoll and Zhang, 1994; Yamanaka et al., 1994; MacGinnitie et al., 1995; Navaratnam et al., 1995). Ainsi le site catalytique de la protéine APOBEC-1 se compose de deux régions conservées, séparées de 30 acides aminés : d'une part, la région (C/HAE) qui contient un résidu (cystéine ou histidine) permettant de lier du zinc et dont le résidu glutamate joue un rôle crucial dans le transfert de protons au cours de la réaction d'édition ; et d'autre part, la région (PCXXC) qui contient une proline conservée et deux cystéines permettant également la fixation de zinc au niveau du site actif. Au centre du site catalytique d'APOBEC1 une différence notable a été révélée, l'existence d'un motif SITW (qui correspond, nous le verrons plus tard, à une « signature » de la famille APOBEC) encadré par deux résidus phénylalanines, impliqués dans la fixation d'APOBEC1 sur le transcrit *apoB* préférentiellement au niveau de la région riche en AU de la séquence de « mooring » (Navaratnam et al., 1995) (figure n°5A).

La cytidine-désaminase d'*E.coli* (ECCDA) agissant sur des mononucléotides fut la première protéine de cette famille dont la structure a été élucidée (Betts et al., 1994). Il existe une conservation de séquences primaires entre ECCDA et APOBEC1 (figure n°5A) qui d'ailleurs agissent toute deux sous forme d'homodimère (Betts et al., 1994; Lau et al., 1994). Par analogie avec la structure d'ECCDA, un modèle structural a été établi pour la protéine APOBEC1 (Navaratnam et al., 1998) révélant l'existence de 3 espaces ou « gaps » dans le site

catalytique d'APOBEC1 (et absents dans la protéine ECCDA) libérant un espace de 10kDa qui pourrait correspondre à la masse du transcrit minimal à éditer. Ainsi, une évolution structurale de la protéine APOBEC1, par rapport aux cytidines désaminases agissant sur des mononucléotides, a été suggérée afin de s'adapter à un nouveau type de substrat : un ARN (figure n°5A).

Dès sa mise en évidence, des investigations expérimentales ont été entreprises afin de caractériser la localisation subcellulaire de la protéine APOBEC1. Il s'avère qu'elle possède un domaine de localisation nucléaire « peu classique » dans sa partie N-terminale, et un signal d'export nucléaire, riche en leucines, dans sa partie C-terminale (Yamanaka et al., 1994; Chester et al., 2003). Dans la cellule, la protéine APOBEC1 est cytosolique, mais grâce à ces deux signaux, elle peut entrer et sortir du noyau (Chester et al., 2003). De plus, bien que ceci reste à démontrer, il est actuellement proposé que des facteurs auxiliaires du complexe d'édition participent à ce mouvement intracellulaire de la protéine APOBEC1.

Par ailleurs, chez l'Homme ou le lapin, l'édition du transcrit *apoB* et l'expression de la protéine APOBEC1 sont spécifiques de l'intestin. En revanche, chez les rongeurs on retrouve l'édition du transcrit *apoB*, tout comme celle de la protéine APOBEC1, dans l'intestin et le foie (Greeve et al., 1993; Teng et al., 1993). Ceci souligne l'existence d'une certaine corrélation entre la distribution tissu-spécifique du processus d'édition du transcrit *apoB* et l'expression de la protéine APOBEC1 chez les vertébrés.

Cependant, chez les rongeurs, la protéine APOBEC1 est exprimée dans d'autres tissus tels le cerveau ou les gonades n'exprimant pas de protéines *apoB*, suggérant que la protéine APOBEC1 puisse avoir d'autres cibles ARN *in vivo* voir un autre rôle notamment dans ces tissus (Teng et al., 1993).

Afin de mieux comprendre la signification biologique de l'édition du transcrit *apoB* et d'autres événements éventuels de régulation par la protéine APOBEC1, l'étude du rôle de cette protéine *in vivo* a été entreprise. Tout d'abord des lignées de souris n'exprimant plus la protéine APOBEC1 (« knock-out ») ont été établies. Ces souris sont viables mais présentent des anomalies dans le métabolisme des lipides. L'édition du transcrit *apoB* est totalement abolie chez ces individus démontrant que cette enzyme est essentielle pour catalyser l'édition de ce transcrit et non redondante (Morrison et al., 1996; Nakamuta et al., 1996).

Parallèlement à ces études, des expérimentations ont été entreprises afin d'évaluer l'utilisation de l'enzyme APOBEC1 dans le cadre d'une thérapie génique chez l'Homme pour lutter contre l'athérosclérose et diminuer le taux de cholestérol LDL dans le sang. La surexpression transitoire de la protéine APOBEC1 en lignée cellulaire a révélé une sur-édition du transcrit *apoB* (Teng et al., 1994; Hughes et al., 1996; Teng et al., 1997). La surexpression de la protéine APOBEC1 dans des souris transgéniques a révélé des individus présentant d'une part une augmentation de l'édition du transcrit *apoB* dans le foie, mais également une dysplasie de cet organe et l'apparition de cancers (Yamanaka et al., 1995). Ainsi, ce résultat compromet le développement d'une thérapie génique chez l'Homme.

Un phénomène d'« hyper-édition » ou d'édition non spécifique pourrait être à l'origine des phénotypes observés dans les souris transgéniques surexprimant la protéine APOBEC1. En effet, plusieurs éléments suggèrent une édition aberrante des transcrits *nat1* et *nfl* (Skuse et al., 1996; Cappione et al., 1997; Yamanaka et al., 1997; Mukopadhyay et al., 2000).

Par ailleurs, il a été mis en évidence que la surexpression de la protéine APOBEC1 augmente la stabilité du transcrit oncogénique *c-myc* car la protéine APOBEC1 reconnaît la séquence riche en AU localisée dans le 3'UTR de cet ARN (Anant and Davidson, 2000). Ceci supporte



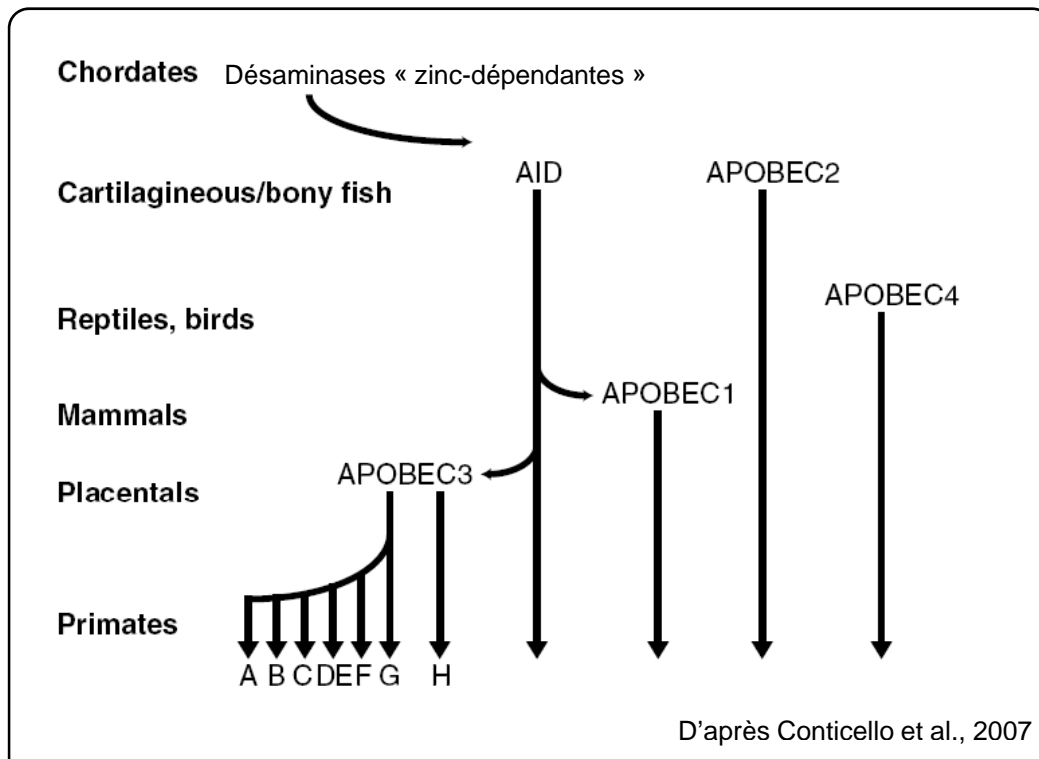


Figure n°6: Évolution de la famille de protéines AID/APOBECs

Il y a près de 500 millions d'années, les deux membres les plus anciens de cette famille, les protéines AID et APOBEC2 (dont la structure génique est remarquablement différente des autres membres de la famille) ont évolué indépendamment à partir de déaminases zinc-dépendantes ancestrales agissant sur des mononucléotides.

Puis le gène *APOBEC1* est apparu chez les mammifères, il y a 300-400 millions d'années, suite à une duplication du locus AID. Chez les animaux placentaires, suite à une deuxième duplication du locus AID, le gène *APOBEC3* a émergé, il y a ~170 millions d'années. Chez les primates, une expansion rapide à partir de ce locus *APOBEC3* a donné naissance à la sous-famille de protéines APOBEC3-A à H.

L'origine du gène *APOBEC4* pourrait être antérieure à l'apparition des reptiles et amphibiens, mais demeure mystérieuse à l'heure actuelle.

l'idée qu'*in vivo*, la protéine APOBEC1 pourrait réguler la stabilité de différents ARN dont les protéines correspondantes contrôlent la croissance et la prolifération cellulaire ainsi que la cancérogenèse. Pris ensemble, tous ces éléments suggèrent ainsi que le gène *APOBEC1* pourrait être un proto-oncogène.

La protéine APOBEC1 est la première cytidine-désaminase agissant sur un ARN ayant été identifiée. La présence du motif « signature » SITW au centre de son site catalytique encadré par deux résidus phénylalanines a permis d'identifier d'autres protéines homologues constituant la famille de protéines AID/APOBECs. Cette famille de protéines AID/APOBECs est apparue de façon relativement tardive au cours de l'évolution, il y a environ ~500 millions d'années pour les membres les plus anciens, les protéines AID et APOBEC2 (figure n°6) (Conticello et al., 2007).

A l'heure actuelle, de nombreuses équipes étudient la fonction des différents membres de la cette famille. Ces recherches ont permis de montrer que la protéine AID (pour activation-induced cytidine deaminase) participe, par désamination C-en-U de l'ADN, à la génération de la diversité des anticorps (Conticello et al., 2007). Par ailleurs, la protéine APOBEC3G peut désaminer (conversion C-en-U) l'ADN simple brin *in vitro*, et semble pouvoir, dans certaines conditions, hypermuter l'ADN du virus HIV, *in vivo*, au cours de sa réplication. Ainsi cette protéine est impliquée dans la lutte contre l'infection virale, mais aussi le contrôle de la réplication de rétroéléments endogènes (Conticello et al., 2007).

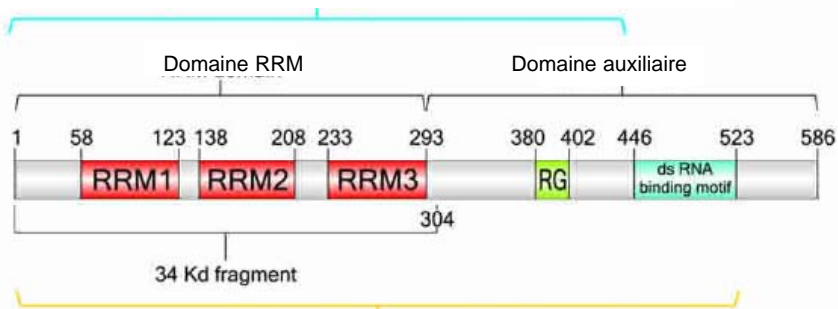
Récemment, les structures des protéines APOBEC2 et APOBEC3G humaines ont été publiées, et leur étude a confirmé des ressemblances avec les désaminases agissant sur des résidus libres, en particulier les cytosines désaminases (Prochnow et al., 2007; Chen et al., 2008). Ainsi, la protéine APOBEC1, facteur catalytique central de l'édition de l'ARNm *apoB*, appartient à une famille de désaminases agissant sur des polynucléotides ARN ou ADN (voir paragraphe III.3 de l'introduction).

## 2) Caractérisation du facteur de spécificité ACF et recherche d'autres facteurs auxiliaires

Comme nous l'avons déjà noté, il a été montré que la machinerie d'édition correspond à un large complexe dit « éditosome » (~1400kDa) (Harris et al., 1993). La protéine APOBEC1, parce qu'elle catalyse la réaction d'édition, a un rôle essentiel au sein de ce complexe, mais elle a besoin de recruter des facteurs auxiliaires pour son activité *in vivo* (Chester et al., 2004). Ces facteurs auxiliaires sont présents y compris dans des cellules ou tissus n'exprimant ni la protéine APOBEC1, ni le transcrit *apoB*. Ils sont sensibles à la dénaturation thermique, saturables en présence d'un excès de substrat, sensibles à l'action de la protéinase K et résistants aux nucléases (Driscoll and Casanova, 1990; Greeve et al., 1991). Ainsi, l'ensemble de ces éléments suggère qu'il s'agisse essentiellement de protéines.

Par des expériences d'« UV cross-linking », une protéine présentant à la fois une affinité pour la protéine APOBEC1 et pour le transcrit *apoB* a été isolée et purifiée. Cette protéine a été nommée ACF (pour APOBEC1 complementation factor) (Mehta and Driscoll, 1998), et son identité a été révélée par spectrométrie de masse (Mehta et al., 2000). La protéine ACF est une protéine de 63,4 kDa qui se compose de 3 motifs RRM (pour RNA recognition motifs) différents, un domaine riche en arginine et un domaine C-terminal de liaison potentiel à l'ARN double-brin (figure n°7). Elle se fixe avec une haute affinité au transcrit *apoB*, dans la séquence de « mooring » riche en AU (Lellek et al., 2000; Mehta et al., 2000). Par des expériences de délétion de domaine protéique, il a été montré que la région N-terminale de la protéine ACF, contenant en particulier les 3 motifs RRM, était indispensable non seulement pour sa liaison au transcrit *apoB* mais aussi pour son interaction avec la protéine APOBEC1

### Région d'interaction avec la protéine APOBEC1



### Région d'interaction avec le transcrit *apoB*

D'après Maris *et al.*, 2005

Figure n°7: Représentation schématique des domaines de la protéine ACF et des régions permettant l'interaction avec la protéine APOBEC1 et le transcrit *apoB*.

En rouge, les 3 motifs RRM (pour RNA recognition motifs)

En vert, le domaine riche en arginine (« RG »)

En turquoise, le domaine potentiel de liaison à l'ARN double-brin (« ds RNA binding motif »)

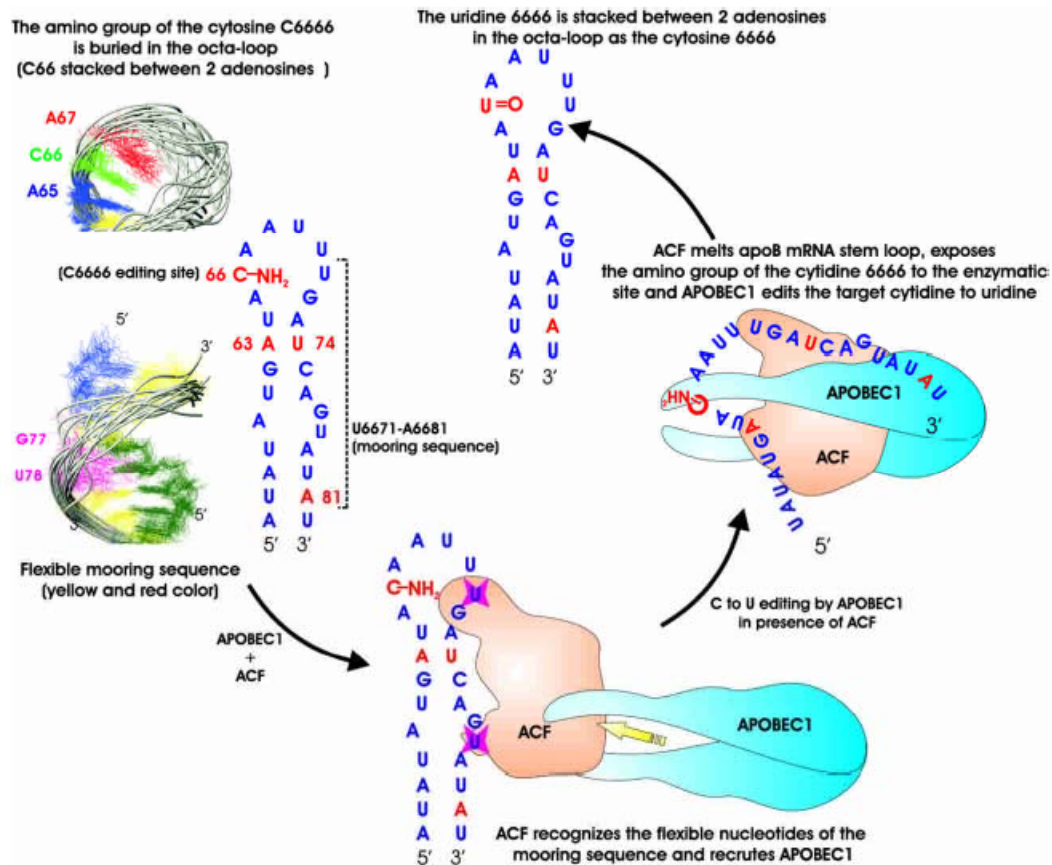
(Blanc et al., 2001a; Mehta and Driscoll, 2002). De façon intéressante, l'existence de plusieurs isoformes de cette protéine, générées par épissage alternatif, a été révélée. Il est actuellement proposé que ces variants protéiques peuvent entrer en compétition pour se lier au transcrit *apoB* ou à la protéine APOBEC1 afin de réguler l'édition (Lellek et al., 2000; Mehta et al., 2000; Sowden et al., 2004). La protéine ACF porte une séquence NLS de localisation nucléaire. Une telle localisation a été validée expérimentalement et par mutagenèse du motif NLS de la protéine ACF, il a été montré que la localisation nucléaire des protéines ACF et APOBEC1 est indépendante (Blanc et al., 2001a).

Par ailleurs, la protéine ACF est exprimée largement dans tous les tissus humains, suggérant qu'elle pourrait participer au métabolisme d'autres ARN (Mehta et al., 2000). A l'heure actuelle, nous disposons de peu d'informations quant au rôle de la protéine ACF *in vivo*. Les souris n'exprimant plus la protéine ACF murine (knock-out) présentent une létalité embryonnaire précoce (Blanc et al., 2005). De même l'extinction de l'expression du gène *hrp-2*, homologue d'*acf* chez *C. elegans*, par interférence à l'ARN, entraîne un arrêt du développement embryonnaire chez cet organisme (Kinnaird et al., 2004). Quel lien existe-t-il entre ces observations phénotypiques et le rôle d'ACF dans l'édition du transcrit *apoB* caractérisé *in vitro* ? Il a été montré que la protéine ACF s'associe avec une haute affinité aux séquences d'ARN riche en AU, et il est actuellement proposé que cette protéine se lie et régule l'expression d'autres ARN et intervient dans le développement précoce.

Dans des tests d'activité *in vitro*, il a été montré que seule la présence des deux protéines APOBEC1 et ACF suffit à l'édition du transcrit *apoB* (Mehta et al., 2000). Ainsi, une vue simplifiée du complexe « minimal » permettant l'édition du transcrit *apoB* a été proposée et se compose de la séquence d'ARN de 26 nucléotides entourant la C<sup>666</sup> à modifier, de l'enzyme homodimérique APOBEC1 et de la protéine ACF. La reconstitution de ce complexe « minimal » d'édition par expression de ces trois partenaires en cellules de levure (où il n'y a normalement pas d'édition des ARN) a confirmé ce modèle (Lellek et al., 2002).

Plus récemment, il a été montré qu'à température physiologique la protéine ACF favorise la réaction d'édition en modifiant la structure secondaire de l'ARN substrat (Chester et al., 2004; Maris et al., 2005). En particulier, par des expériences de résonance magnétique nucléaire, il a été montré que la protéine ACF permet une ouverture de la structure en tige-boucle de l'ARN substrat (formé par les régions conservées) et rend la cytidine C<sup>666</sup> accessible pour l'enzyme APOBEC1. La figure n°8 reprend tous les éléments dont nous disposons quant au mode d'action du complexe « minimal » d'édition ACF /APOBEC1/transcrit *apoB*.

Parallèlement à l'étude du facteur ACF, des recherches ont été menées afin d'identifier d'autres facteurs auxiliaires présents dans le complexe d'« éditosome ». Tout d'abord, la protéine GRY-RBP (pour Glycine-Ariginine-Tyrosine rich RNA Binding Protein), présentant 50% d'homologie avec la protéine ACF, peut inhiber la réaction d'édition dans les tests *in vitro* (Blanc et al., 2001b). Par ailleurs, une protéine nommée ABP-1 (pour Apobec1-Binding-Protein-1) interagit avec la protéine APOBEC1, mais son rôle dans le processus d'édition reste à démontrer à ce jour (Lau et al., 1997). Enfin, une protéine du complexe d'« éditosome », nommée AUX240, a été isolée, et sa présence favorise l'efficacité d'édition dans les tests *in vitro* (Schock et al., 1996). Néanmoins, en dehors de la caractérisation du complexe « minimal » d'édition, nous disposons de peu d'informations sur l'identité et le rôle des autres facteurs du complexe d'« éditosome ». Ces facteurs ne sont pas essentiels à la réaction d'édition en elle-même, et doivent participer à la régulation de l'édition du transcrit *apoB*. De plus, l'efficacité d'édition de ce transcrit varie en fonction d'autres paramètres tels



D'après Marris et al., 2005

Figure n°8: Modèle d'action du complexe « minimal » d'édition du transcrit *apoB* : rôle des protéines APOBEC1 et ACF.

Dans ce modèle, les séquences conservées autour la cytidine à modifier de l'ARNm ApoB forment une structure en « tige-boucle » dans laquelle se trouve la cytidine C6666 (elle est donc peu accessible pour être éditée). La protéine APOBEC1 entre dans le noyau où elle s'associe à la protéine ACF. Cette dernière reconnaît deux résidus uraciles localisés dans la séquence de « mooring » et s'associe spécifiquement à ce transcrit. Cette interaction ACF/*apoB* va modifier la structure secondaire de cet ARN (ouverture de la tige-boucle) de sorte que la cytidine C6666 à modifier devienne accessible pour l'enzyme APOBEC1 qui catalyse ainsi la réaction de déamination C-en-U. Il a été proposé que le complexe APOBEC1/ACF/*apoB* quitte le noyau et se dissocie dans le cytoplasme afin de protéger le transcrit édité d'une dégradation prématurée au cours de son export nucléaire (Chester et al., 2003). Dans le cytoplasme, l'ARN *apoB* peut reprendre sa conformation « tige-boucle » et être traduit en protéine.

que le développement, l'état hormonal et l'alimentation des individus suggérant l'existence d'un autre niveau de régulation dans les individus.

#### *III.1.4 Pour résumé*

Chez les vertébrés, l'édition C-en-U d'un seul nucléotide du transcrit *apoB* résulte en la synthèse de deux protéines différentes à partir d'un seul gène, et augmente ainsi la complexité du protéome. Cependant, face à l'étendue des régions d'ADN non codante du génome des mammifères, coder pour un gène supplémentaire ne devrait pas poser de problème. Ainsi l'édition du transcrit *apoB* est apparue et a été conservée au cours de l'évolution, car elle doit apporter un avantage sélectif aux individus, comme par exemple permettre une réponse plus rapide aux changements métaboliques liés aux nutriments entrant dans l'intestin.

Une étape importante à l'acquisition de l'édition de l'ARN *apoB* chez les mammifères a été l'évolution d'une désaminase agissant sur des substrats monomériques en une famille de protéines, dont la protéine APOBEC1, capable de catalyser la désamination d'une cytidine d'un polynucléotide (ARN ou ADN) (voir aussi paragraphe III.3 de l'introduction).

La protéine APOBEC1 est ainsi le facteur catalytique du processus d'édition, et elle agit dans un large complexe multiprotéique : l'éditosome. La caractérisation d'un autre membre de ce complexe, le facteur spécificité ACF, a permis de mieux comprendre comment la base à désaminer est spécifiée dans le transcrit *apoB*. L'identification et la caractérisation d'autres facteurs intervenant dans ce processus permettront sans doute une compréhension plus fine du fonctionnement et de la régulation de l'activité de la machinerie d'édition dans la cellule.

### **III.2 L'édition A-en-I des ARNm chez les animaux**

#### *III.2.1 Généralités*

L'édition par désamination A-en-I des ARNm chez les animaux est catalysée par les enzymes de la famille ADARs (pour **adenosine deaminases that act on RNA**). Ces protéines agissent sur un substrat ARN entièrement ou largement double-brin.

Historiquement, cette famille de protéines a été tout d'abord découverte chez le Xénope et rapidement des membres de cette famille ont été isolés et clonés chez de nombreux animaux tels que les mammifères, les oiseaux, les grenouilles, les mouches et les vers. De nouveaux membres de cette famille sont constamment ajoutés dans les bases de données, et cette famille de protéines semble restreinte aux métazoaires.

Les protéines ADARs agissent uniquement sur des ARNm nucléaires, et à l'heure actuelle un petit nombre de substrats a été identifié. Les modifications ont d'abord été décrites dans les régions codantes, mais aux cours de ces dernières années, il a été mis en évidence que de nombreuses désaminations A-en-I pouvaient survenir dans les régions non-codantes des ARNm, augmentant considérablement le nombre de cibles potentielles pour ces protéines.

Les protéines ADARs, via leur activité de désamination A-en-I, modifient l'information codée par le génome ou la structure de l'ARN substrat.

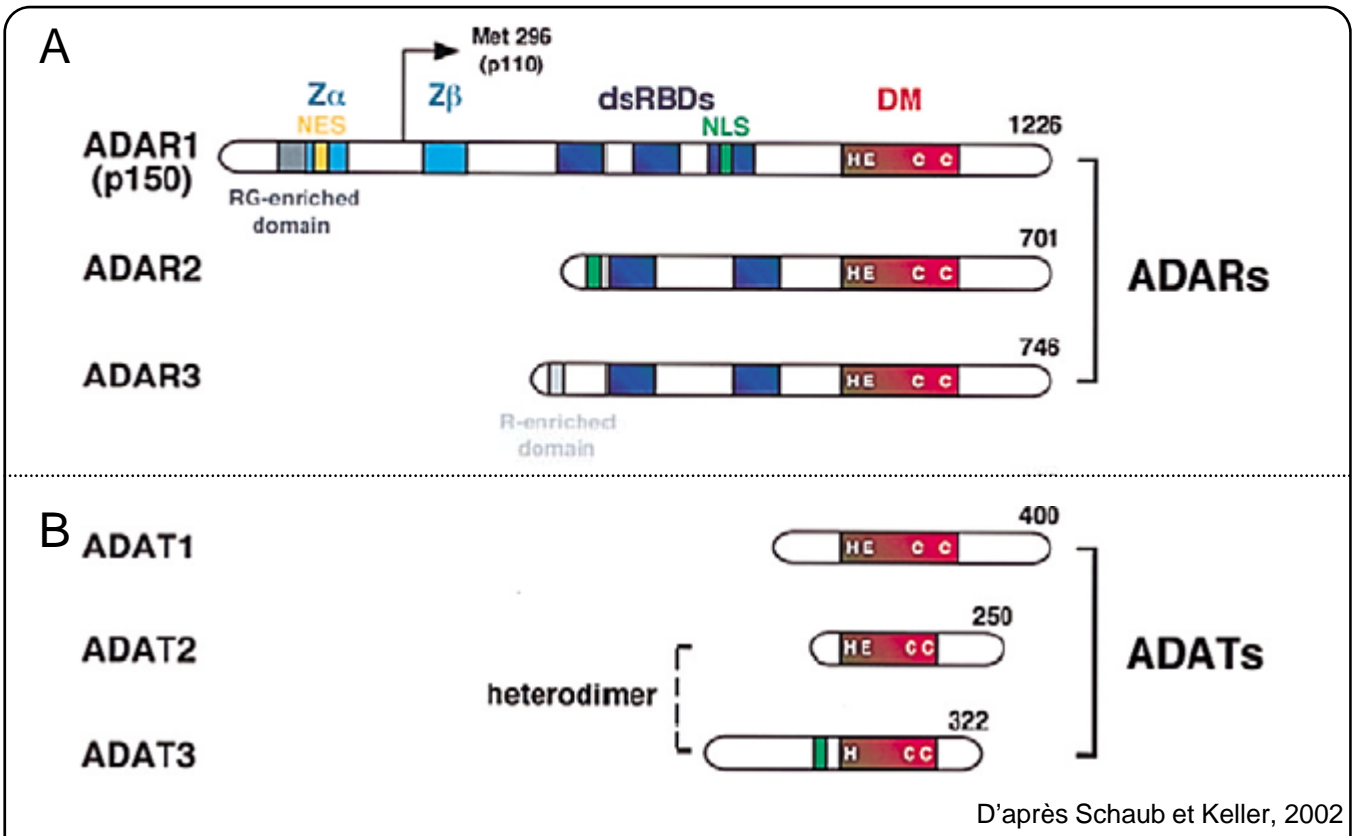


Figure n°9: Organisation des domaines fonctionnels des protéines ADARs humaines et ADATs de levure. Ces protéines portent toutes un site catalytique de désaminase (DM) coloré en rouge, contenant 3 résidus pour fixer le zinc (une histidine et deux cystéines) et un résidu glutamate (E) essentiel à l'activité catalytique.

- (A) Les protéines ADARs contiennent, en plus du site catalytique, des domaines de fixation à l'ARN double brins (dsRBDs), colorés en bleu. Si toutes les protéines ADARs ont une organisation structurale commune, il existe cependant des différences remarquables spécifiques aux différents membres de la famille. Tout d'abord, les protéines ADAR3 et ADAR1 (isoforme p150) possèdent un domaine riche en arginine et en lysine (R-domain) dans sa partie N-terminale (coloré en gris). Par ailleurs, la protéine ADAR1 (isoforme p150) possède, à son extrémité N-terminale, un domaine de fixation à l'ADN (Z-DNA), coloré en turquoise. Ce domaine pourrait permettre à la protéine ADAT1 de se fixer à proximité d'un site d'activation de la transcription dans le génome, afin qu'elle puisse exercer son activité de désamination sur les transcrits nouvellement synthétisés (Herbert et al., 1997). De plus, ce domaine est décrit comme essentiel à l'édition des petits ARN double-brin de 15pb (Herbert et al., 2001). La protéine ADAR1 (isoforme p150) contient également un signal d'exportation nucléaire (NES) coloré en jaune et un signal de localisation nucléaire (commun aux deux isoformes) indiqué en vert. La méthionine m296 de l'isoforme ADAR1p110 est également indiquée.
- (B) les protéines ADATs possèdent un domaine catalytique C-terminal similaire à celui des protéines ADARs, mais aucun motif de fixation à l'ARN double brin. Il est couramment admis que les protéines de la famille ADARs ont évolué à partir des protéines des protéines ADATs par l'acquisition des domaines dsRBD.

En effet, lorsqu'un résidu inosine est introduit dans une séquence codante d'un ARNm, il va être « interprété » comme une guanosine (G) par la machine de traduction, ce qui peut altérer le codon spécifié et donc la séquence protéique. En outre, cela peut affecter également l'épissage de ce transcrit, sa stabilité, ou son efficacité de traduction. Par ailleurs, la désamination des résidus adenosine (A) résulte en le remplacement d'une paire de base A : U par un mésappariement I : U. Comme les inosines s'hybrident préférentiellement avec des cytidines, cela peut introduire une déstabilisation de la structure double-brin de l'ARN substrat.

Initialement, il a été montré que ces changements peuvent conduire à une diversification des protéines produites par un même gène ou à une altération de la fonction de la protéine, mais la mise en évidence récente de désaminations dans les régions non codantes suggèrent, comme nous le verrons, d'autres rôles pour les protéines ADARs.

### *III.2.2 La famille de protéines ADARs*

Chez tous les organismes, malgré une faible conservation de la séquence primaire, les protéines ADARs présentent une organisation structurale similaire comprenant un domaine catalytique dans la partie C-terminale et 2 à 3 répétitions de motifs de liaison à l'ARN double-brins (dsRBDs) (figure n°9) (Kim et al., 1994; O'Connell et al., 1995).

Chez les vertébrés, 3 gènes ADARs distincts (ADAR1, 2 et 3) sont très conservés des poissons jusqu'à l'Homme (Slavov et al., 2000a; Slavov et al., 2000b). Chez les invertébrés, un seul gène ADAR est exprimé chez la drosophile (dADAR), alors que chez le vers *C. elegans*, il existe deux gènes ADAR moins conservés (ceADAR1 et ceADAR2) (Kim et al., 1994; Hough et al., 1999; Palladino et al., 2000a).

Les protéines ADARs sont exprimées dans le noyau, à l'exception de l'isoforme p150 de la protéine ADAR1 (Patterson and Samuel, 1995; George and Samuel, 1999a; Kawakubo and Samuel, 2000). En effet, chez les vertébrés, la présence de deux promoteurs permet d'obtenir deux isoformes de la protéine ADAR1 : une forme courte (p110) et une forme longue (p150) (figure n°9) (Patterson and Samuel, 1995; George and Samuel, 1999a; Kawakubo and Samuel, 2000). L'isoforme p110 est constitutivement exprimée dans le noyau, tandis que l'isoforme p150 est produite à partir d'un promoteur inductible par un interféron (INF) et a la particularité de pouvoir entrer et sortir du noyau (Eckmann et al., 2001; Poulsen et al., 2001).

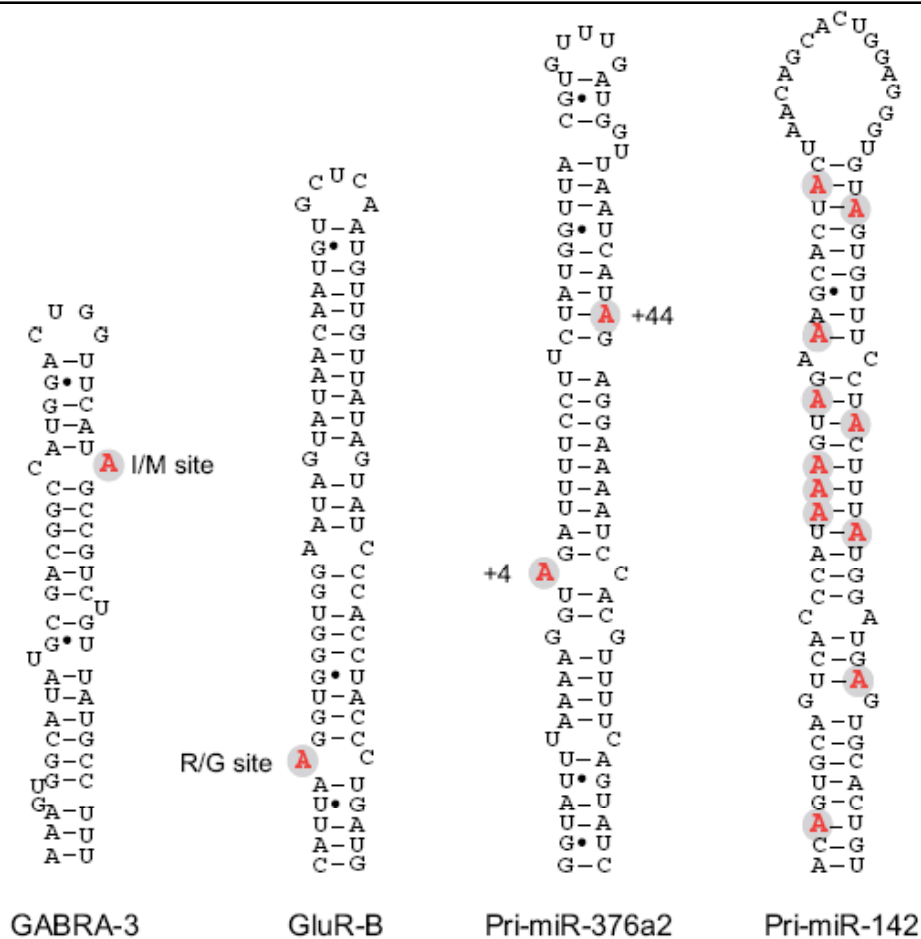
### *III.2.3 Interaction des protéines ADARs avec leur ARN substrat*

#### Les protéines ADARs se fixent à un ARN double-brin

Les domaines dsRBDs des protéines ADARs sont comparables à ceux identifiés chez d'autres protéines et il a été montré, par mutagenèse, que le motif KKxxK de ces domaines est essentiel à la fixation sur l'ARN (Lai et al., 1995; Liu and Samuel, 1996; Gallo et al., 2003; Sansam et al., 2003). Ainsi les protéines ADARs appartiennent à la superfamille des protéines de liaison à l'ARN double-brin (Carlson et al., 2003; Saunders and Barber, 2003).

Les protéines ADARs contiennent plusieurs répétitions de domaine dsRBDs, il n'est donc pas surprenant qu'elles fixent l'ARN double-brin avec une haute affinité. En revanche, les





D'après Ohman, 2007

Figure n°10: Structure secondaire d'ARN substrat des protéines ADARs.

Les ARNm substrats des protéines ADARs forment généralement une longue molécule d'ARN entièrement ou largement double brin. Les appariements font généralement plusieurs centaines de paires de base à quelques exceptions près comme celle de l'ARN murin GABRA-3 codant pour la sous-unité  $\alpha 3$  du récepteur GABA qui constitue le plus petit substrat naturel actuellement connu pour les protéines ADARs (~20 paires de bases) (présenté à gauche). Les ARN substrats sont rarement totalement double brin, et il apparaît que la présence de boucles et de mésappariements confère une spécificité non seulement pour la reconnaissance de l'enzyme avec son substrat mais aussi pour spécifier les résidus à éditer.

GABRA-3 et GluR-B sont les structures prédites pour les transcrits de souris codants pour la sous-unité  $\alpha 3$  du récepteur GABA et la sous-unité b du récepteur au glutamate, respectivement. Ces deux ARN substrats sont édités sélectivement (en rouge est indiquée la base désaminée).

Pri-miR-376a2 (homme) et Pri-miR-142 (souris) correspondent à des microARN non-codants et se retrouvent hérités.

données actuelles indiquent que les protéines ADARs reconnaissent leur substrat ARN avec une faible sélectivité toutefois le nombre de répétitions de motifs dsRBDs qu'elles contiennent, ainsi que la distance les séparant, pourraient conférer une certaine spécificité de reconnaissance du substrat, *in vivo* (Bass, 2002). De façon surprenante, il a été montré que la protéine ADAR3 reconnaît les ARN double-brin avec une haute affinité bien qu'elle ne semble pas capable de désaminer ces substrats *in vitro*, suggérant que dans le cerveau elle pourrait recruter d'autres facteurs pour assurer sa fonction catalytique (Melcher et al., 1996; Chen et al., 2000).

### Structure de l'ARN substrat des protéines ADARs

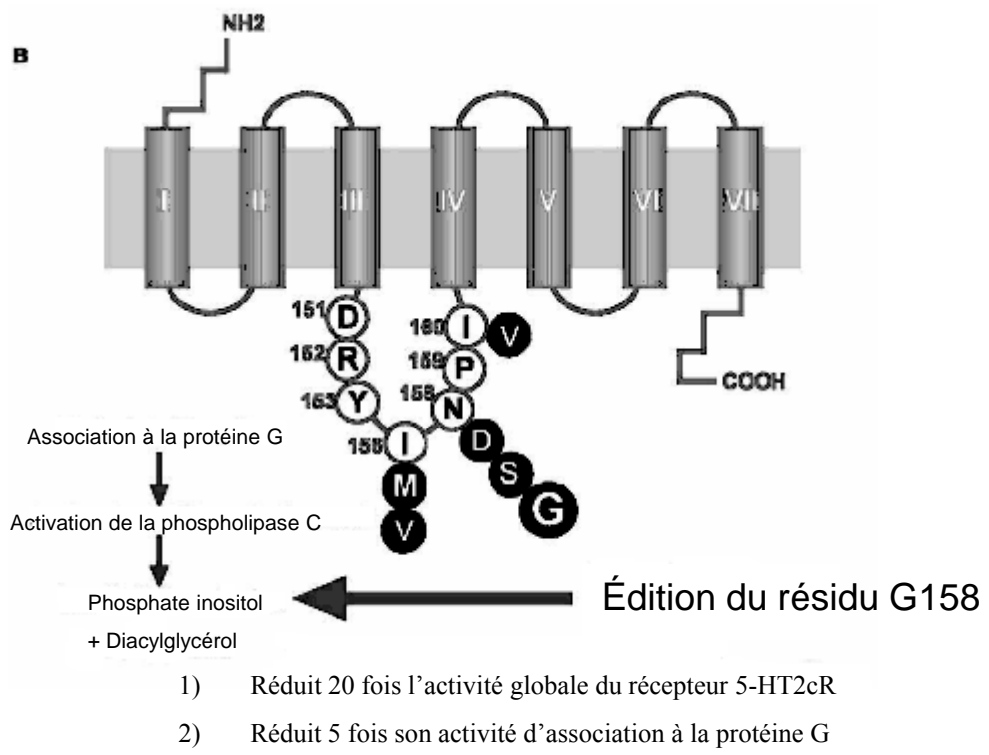
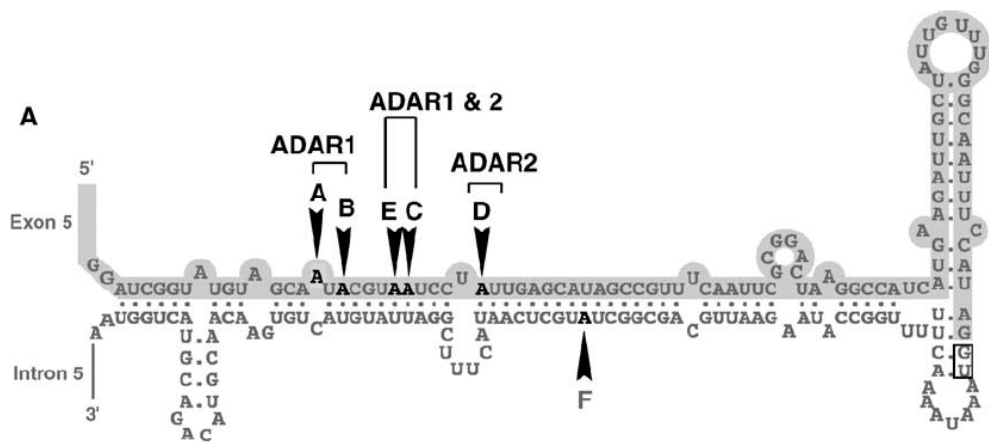
Les ARNm substrats des protéines ADARs forment généralement une longue molécule d'ARN entièrement ou largement double-brin dans laquelle il apparaît que la présence de boucles et de mésappariements confère une spécificité non seulement pour la reconnaissance de l'enzyme avec son substrat mais aussi pour spécifier les résidus à éditer (Lehmann and Bass, 1999; Dawson et al., 2004) (figure n°10).

De nombreux ARN substrats des protéines ADAR forment des structures double-brin par un repliement « intramoléculaire ». Dans les régions codantes, les pré-ARNm forment des structures secondaires par association de la séquence exonique avec une séquence intronique complémentaire nommée ECS (pour Editing Complementary Sequence) (Higuchi et al., 1993; Lomeli et al., 1994; Burns et al., 1997). Le fait que la mise en place de ces structures secondaires fasse intervenir des séquences introniques signifie que l'édition a lieu avant ou de façon simultanée avec l'épissage (Raitskin et al., 2001; Bratt and Ohman, 2003). Cependant, il existe quelques rares exemples de formation d'ARN substrat double-brin par association naturelle avec un ARN antisens (Saccomanno and Bass, 1999).

Bien souvent, les structures secondaires des ARN substrats des protéines ADARs ont été prédites à l'aide d'outils bioinformatiques, biochimiques, phylogéniques, et biophysiques, et n'ont pas été rigoureusement démontrées. Le substrat hairpin R/G est une exception puisque sa structure a été vérifiée par mutagenèse compensatoire (Lomeli et al., 1994; Aruscavage and Bass, 2000). Il est intéressant de noter qu'à la différence de ce qui est observé pour les ARNt et les ARNr, les résidus les plus conservés de ce substrat sont localisés dans les hélices d'ARN double-brin, plutôt que dans les boucles et autres mésappariements (Aruscavage and Bass, 2000). Une observation similaire a été faite pour les transcrits du récepteur de la sérotonine (Niswender et al., 1998). Ceci va de paire avec l'idée que les boucles et autres mésappariements sont essentiels pour maintenir la spécificité de reconnaissance enzyme/substrat, ainsi que pour spécifier la base à modifier.

### Spécificité des protéines ADARs

*In vitro*, il a été montré que dans un substrat totalement apparié et d'une longueur de 50pb, ~50% des résidus adénosines vont être désaminés avant que la réaction ne s'arrête. Si les molécules d'ARN sont plus courtes ou portent des mésappariements, elles ne seront désaminées qu'au niveau de quelques sites uniquement (Nishikura et al., 1991; Polson and Bass, 1994). De plus, *in vivo*, on constate que les régions codantes sont éditées plus spécifiquement que les régions non codantes qui se retrouvent souvent hyperéditées (Ohman, 2007) (figure n°10). Ainsi les protéines ADARs ont une très grande spécificité. Cette spécificité est intrinsèque aux protéines ADARs et à leur ARN substrat ; elle ne fait pas intervenir d'autre co-facteur (Polson et al., 1996).



D'après Valente et Nishikura, 2005

Figure n°11: (A) Fragment du transcrite codant pour le récepteur humain de la sérotonine 5-HT<sub>2C</sub> contenant 6 sites d'édition. Les sites A-E sont localisés dans l'exon 5 qui correspond à une boucle intracellulaire II à laquelle s'associe la protéine G indispensable à l'activité de ce récepteur. Le site F est localisé dans l'intron et son édition affecte l'épissage du transcrite 5-HT<sub>2C</sub>R. L'exon 5 et l'intron adjacent forment un ARN double brin. *In vitro* les protéines ADAR1 et ADAR2 éditent ce transcrite selon différents « patterns » pouvant générer jusqu'à 24 isoformes de ce récepteur.

(B) Le récepteur 5-HT<sub>2C</sub> contient 7 domaines transmembranaires. Les 7 acides aminés DRYINPI (indiqués dans un cercle blanc) de la boucle intracellulaire II sont indispensables à la liaison à la protéine G. Encerclés en noir sont indiqués les acides aminés variants produits par édition. Selon les sites édités, le couplage à la protéine G est plus ou moins efficace. Par exemple, l'apparition d'un acide aminé glycine (G) en position 158 réduit de 5 fois l'efficacité de couplage à la protéine G. De plus la conformation de la boucle varie considérablement en fonction du « pattern » d'édition.

Néanmoins, les tests *in vitro* ont montré que les protéines ADARs peuvent se fixer sur n'importe quel ARN double-brin et le désaminer à certains sites. Ainsi, l'utilisation du terme spécificité pour caractériser ces enzymes est assez ambiguë; de ce fait les termes de préférence et de sélectivité ont été définis, comme décrit ci-après (Bass, 2002).

#### Préférences des protéines ADARs

Au sein d'un transcrit, les protéines ADARs vont désaminer certains résidus adénosines préférentiellement. Ainsi, dans les tests *in vitro*, les protéines ADARs désaminent leur substrat endogène de façon identique à ce qu'elles font *in vivo* (Polson and Bass, 1994; Lehmann and Bass, 2000).

Il a été montré que les protéines ADAR1 et ADAR2 de certains vertébrés désaminent préférentiellement les adénines qui ont en 5' un résidu A ou U, et bien qu'il manque d'évidences expérimentales, les protéines ADARs des invertébrés pourraient préférer un résidu G en 5' de l'adénine à modifier (Polson and Bass, 1994; Yang et al., 1995; Patton et al., 1997; Morse and Bass, 1999; Lehmann and Bass, 2000; Reenan et al., 2000). En outre, dans l'ARN double-brin, la base qui se trouve en face de l'adénine à modifier semble également jouer un rôle, en particulier les mésappariements A: C sont plus efficacement édités (Wong et al., 2001).

De plus, il a été montré, par mutagenèse, que les bases entourant l'adénine à modifier, sans doute parce qu'elles sont en contact avec le site catalytique, pourraient également influencer l'efficacité d'induction (Yi-Brunozzi et al., 2001; Kallman et al., 2003). Par ailleurs, il apparaît que les protéines ADAR1 et ADAR2 ne désaminent pas les mêmes résidus au sein d'un transcrit, et des expériences d'échange de leurs domaines catalytiques ont confirmé qu'ils sont importants dans cette sélectivité (figure n°11A) (Valente and Nishikura, 2005).

#### Sélectivité des protéines ADARs (modèle établi par Bass, 2002)

Clairement ces préférences pour les nucléotides voisins ne suffisent pas à expliquer l'incroyable spécificité de reconnaissance des adénines à éditer par les protéines ADARs. Par exemple, le transcrit de l'antigénome HDV contient plus de 300 adénines et une seule est précisément désaminée (Polson et al., 1996; Polson et al., 1998). Il apparaît que la structure et la stabilité de l'ARN substrat ont également un rôle essentiel dans le choix des bases à modifier.

Nous l'avons dit, les protéines ADARs présentent uniquement une activité sur des ARN double-brin. Dans une cellule, il existe un équilibre entre une forme double-brin et une forme simple brin pour une molécule d'ARN, et tout événement qui modifie cet équilibre en favorisant la forme simple brin de l'ARN substrat va diminuer l'affinité des protéines ADARs pour ce substrat.

Ainsi, le concept de sélectivité propose que l'introduction, par édition, d'un mésappariement ou plusieurs mésappariements I: U peut déstabiliser l'ARN substrat qui devient simple brin, arrêtant ainsi l'activité des protéines ADARs. Cette idée simple expliquerait en outre pourquoi les protéines ADARs désaminent moins efficacement les substrats de petites longueurs et ceux portant des boucles et mésappariements (figure n°12).

De plus, d'après ce modèle, la présence de boucles internes dans une longue molécule d'ARN doit diminuer le nombre de désamination possible et augmenter la sélectivité (figure n°12). Nous avons évoqué précédemment que les résidus conservés dans les ARN substrats semblent

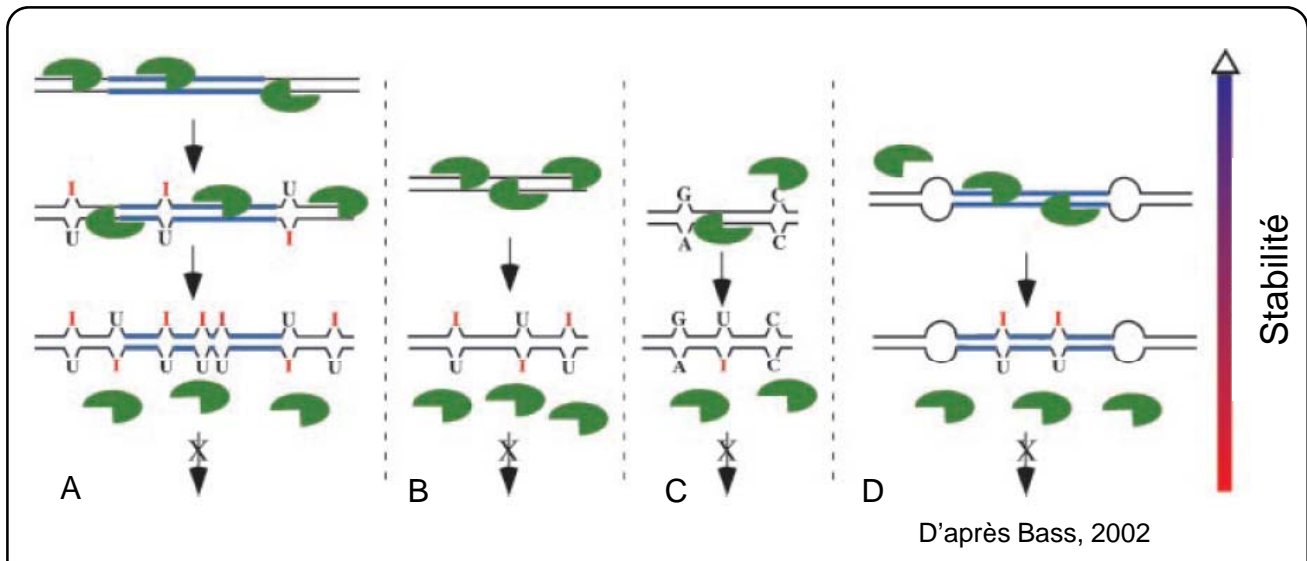


Figure n°12: La spécificité des protéines ADARs dépend de la structure de l'ARN substrat

Les protéines ADARs (en vert) agissent sur 4 ARN de stabilité différente. Leur action introduit des mésappariements I:U, de sorte que les substrats deviennent de moins en moins stables et retrouvent finalement une structure simple. Le modèle propose que ceci arrête la réaction d'édition par les protéines ADARs. Ainsi, les substrats plus courts (B) ou qui contiennent des mésappariements (C) sont désaminés plus sélectivement car peu d'événements de désamination suffisent à modifier l'« équilibre » entre l'état double brin et simple brin. Pour les séquences spécifiques présentées en (A) et (D), on note que la sélectivité augmente en présence de boucles internes. En effet un long substrat contenant plusieurs boucles internes est reconnu comme une série de substrats plus courts entre chaque boucle. Ainsi les protéines ADARs ne désaminent que quelques bases au sein d'un tel transcrit.

localisés dans les régions d'hélices double-brin afin de maintenir la structure secondaire de ces substrats et donc la sélectivité qui opère.

Enfin, il apparaît que la structure secondaire de certains ARN substrats doit permettre aux protéines ADARs de se fixer au niveau d'un site unique et spécifier ainsi la base à modifier.

**En résumé, bien que les protéines ADARs puissent avoir une légère préférence pour des adénosines dans un contexte nucléotidique particulier, c'est essentiellement la structure de l'ARN substrat qui détermine cette spécificité.**

### *III.2.4 L'activité catalytique des protéines ADARs*

Quand les protéines ADARs désaminent un résidu adénosine (A) en présence de molécule d'eau portant un atome d'oxygène marqué, cet oxygène lourd se retrouve incorporé au niveau du carbone 6 du résidu inosine ainsi produit (Polson et al., 1991). Ainsi les protéines ADARs catalysent une désamination hydrolytique, sans doute par un mécanisme similaire à celui utilisé par les désaminases agissant sur les nucléotides ou les nucléosides telle que les adénosines désaminases (ADAs) et les cytidines désaminases (CDAs) (Wilson et al., 1991; Betts et al., 1994).

Les protéines ADAs et CDAs sont toutes deux des métalloenzymes qui utilisent le zinc pour catalyser les réactions de désamination hydrolytique. Ces deux types d'enzymes ont un résidu glutamate, essentiel à leur activité, près du site de fixation du zinc. Cependant les ADAs utilisent 3 résidus histidine et un résidu aspartate pour lier le zinc, alors que les CDAs utilisent un résidu histidines et deux résidus cystéines. Des études cristallographiques des membres de ces deux familles ont montré que ces différences dans la séquence primaire avaient des répercussions structurales importantes dans la façon dont le zinc est fixé (Carter, 1995; Carter, 1998).

De façon surprenante, le site catalytique des protéines ADARs présente des homologies avec celui de cytidines désaminases, plutôt que celui des adénosines désaminases. Ainsi, par des expériences de mutagenèse, il a été démontré qu'au sein du triplet conservé HAE, le résidu glutamate est essentiel à l'activité de ces protéines (Lai et al., 1995). Par ailleurs, par analogie avec les cytidines désaminases, il a été proposé que les tripeptides HAE et PCG pourraient fixer du zinc (Lai et al., 1995).

Comme pour certaines cytidines désaminases, il a été montré que les protéines ADARs agissent sous forme de dimères chez les mammifères et les mouches. Les protéines ADAR1 et ADAR2 ne peuvent pas former d'hétérodimères, et doivent former un homodimère pour être actives (Cho et al., 2003). De façon intéressante, la protéine ADAR3 ne se dimérise pas, ce qui pourrait expliquer son absence d'activité sur différents ARN bien qu'elle puisse s'y lier avec une haute affinité (Cho et al., 2003). Selon les espèces, des résultats indiquent que la dimérisation des protéines ADARs peut s'effectuer avant leur fixation à l'ARN substrat, ou au contraire par l'intermédiaire de ce substrat (Cho et al., 2003; Gallo et al., 2003).

Enfin, il a été proposé que les protéines ADARs utilisent un mécanisme de « base-flipping » pour faire sortir l'adénosine à modifier de la molécule d'ARN double-brin et la rendre ainsi accessible au site actif de l'enzyme (Polson and Bass, 1994; Stephens et al., 2000; Yi-Brunozzi et al., 2001).

### III.2.5 Rôle *in vivo* des protéines ADARs

#### 1) Phénotype *in vivo*

Les conséquences physiologiques des désaminations A-en-I catalysées par les protéines ADAR ont été étudiées chez plusieurs espèces.

Chez *C. elegans* et chez la drosophile, les individus d'une lignée n'exprimant plus de protéine(s) ADAR sont viables et présentent des anomalies de locomotion et comportementales (Palladino et al., 2000b; Tonkin et al., 2002). Ainsi, chez les invertébrés, les gènes *ADARs* ne semblent pas essentiels mais entraînent des désordres neurologiques.

A l'opposé, chez les mammifères, les souris n'exprimant plus le gène *ADAR1* ou *ADAR2* meurent au cours du développement embryonnaire ou quelques jours après leur naissance suite à des crises épileptiques répétées (Higuchi et al., 2000; Wang et al., 2000a; Wang et al., 2004). Chez l'Homme, des mutations dans le locus *ADAR1* résultent en l'apparition d'un symptôme neurologique dit « dyschromosis symmetrica hereditaria » (Miyamura et al., 2003; Zhang et al., 2004).

#### 2) Désaminations qui modifient la séquence protéique

Les protéines ADARs jouent un rôle clé dans l'édition des transcrits codant pour des protéines du système nerveux central, comme de nombreux canaux ioniques et des récepteurs de neurotransmetteurs (Seeburg and Hartner, 2003).

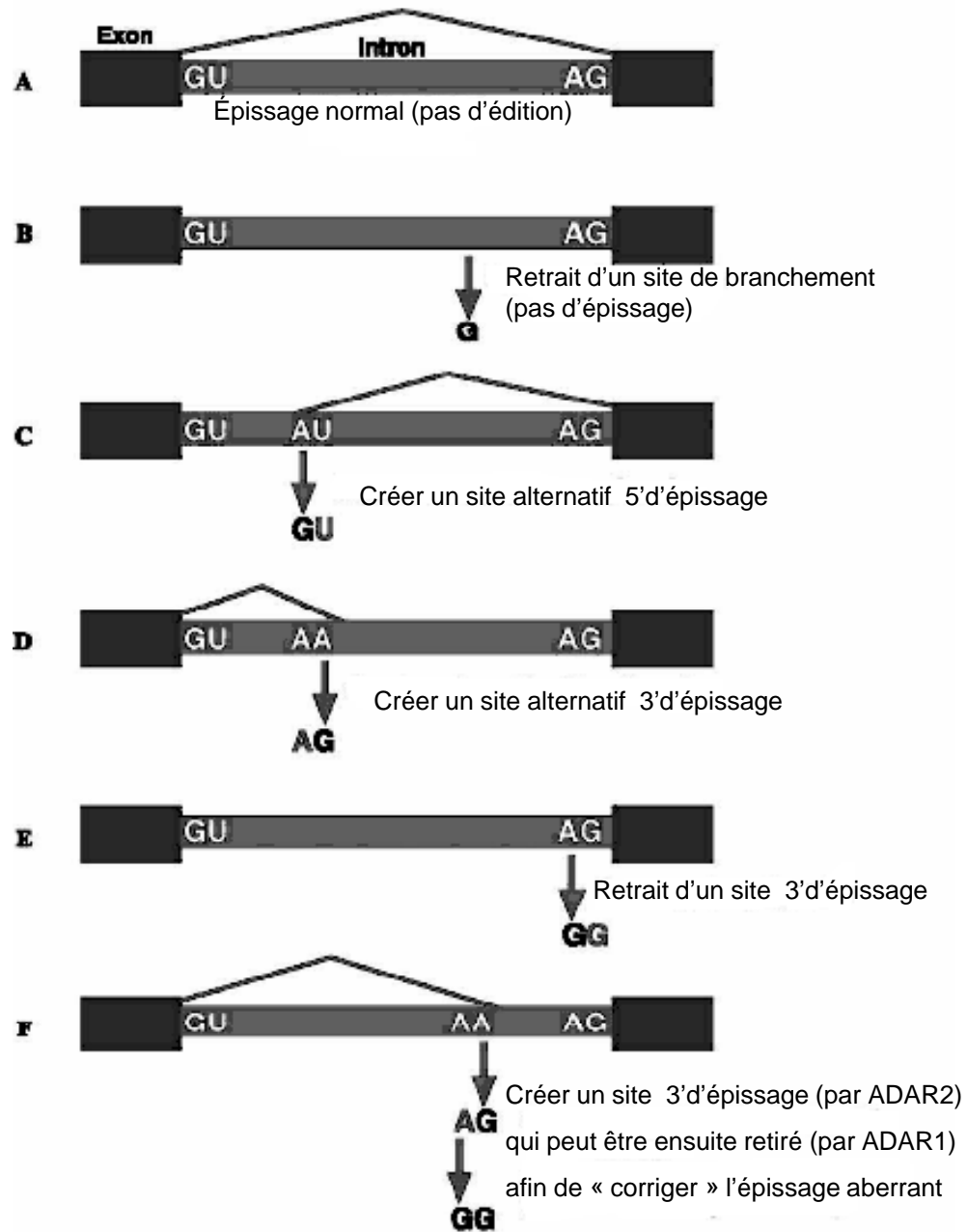
En effet, un résidu inosine dans l'ARNm va être « interprété » comme une guanosine (G) par la machine de traduction et ce processus d'édition va permettre de produire différentes isoformes, et en créant une diversité protéique, de réguler finement le fonctionnement de ces différents composants cérébraux. Ainsi l'édition doit conférer un avantage, c'est pourquoi quand ce processus est apparu il a non seulement été conservé mais aussi le nombre de sites d'édition dans les transcrits cibles n'a cessé d'augmenter (Bass, 2002; Maas et al., 2003).

#### La protéine ADAR2 chez les mammifères

La protéine *ADAR2*, *in vivo*, a une cible majeure : le site Q/R du transcrit codant pour la sous-unité B du récepteur *GluR* qui est édité avec 100% d'efficacité (Higuchi et al., 2000; Seeburg et al., 2001). En effet, la létalité des souris *ADAR2*<sup>-/-</sup> peut être complétée par l'expression du transcrit *GluR-B* édité (Higuchi et al., 2000). La protéine *ADAR2* édite cependant de nombreux autres sites de ce transcrit, mais l'édition de ces derniers n'apparaît pas essentielle à la survie des individus (Higuchi et al., 2000).

Il a été montré que l'édition du site Q/R est essentiel à un assemblage correct des 4 sous-unités de ce récepteur ainsi qu'à sa perméabilité au calcium (Sommer et al., 1991; Egebjerg and Heinemann, 1993; Seeburg et al., 2001; Greger et al., 2002; Greger et al., 2003). Enfin l'édition de ce site a des conséquences sur l'épissage correct du transcrit *GluR-B* (Higuchi et al., 2000).

Notons que dans le cerveau humain, les canaux récepteurs de glutamate (*GluR*) sont édités et un défaut d'édition dans ces transcrits cibles pourrait être associé à de nombreux troubles tels



D'après Valente et Nishikura, 2005

Figure n°13: L'édition par les protéines ADARs peut altérer l'épissage d'un transcrit

(A) Épissage correct d'un transcrit (B) l'édition affecte le site de branchement et abolit l'épissage du transcrit (C) l'édition crée un site alternatif d'épissage en 5' (D) ou en 3' (E) l'édition affecte le site 3' accepteur et abolit l'épissage du transcrit. (F) La protéine ADAR2 édite son propre transcrit créant un site accepteur 3' d'épissage. La protéine produite par cet épissage alternatif est non fonctionnelle. De façon intéressante, ce site d'épissage alternatif peut être supprimé sous l'action de la protéine ADAR1.



que l'épilepsie, la maladie d'Alzheimer, la maladie d'Huntington, la schizophrénie (Akbarian et al., 1995; Kawahara et al., 2003; Kawahara et al., 2004; Vollmar et al., 2004).

### La protéine ADAR1 chez les mammifères

La protéine ADAR1, *in vivo*, édite 6 sites (A-F) du transcrite codant pour le récepteur de la sérotonine de type 2C, appelé 5-HT<sub>2C</sub>R (figure n°11B). Selon le « *pattern* » d'édition, différentes isoformes sont synthétisées, et une étude pharmacologique a permis de montrer que la protéine G s'associait plus ou moins à ces différentes isoformes du récepteur 5-HT<sub>2C</sub>R (Burns et al., 1997; Fitzgerald et al., 1999; Herrick-Davis et al., 1999; Niswender et al., 1999; Wang et al., 2000b). Ainsi la protéine ADAR1, par édition différentielle du transcrite 5-HT<sub>2C</sub>R peut diminuer la réponse à la sérotonine d'un organisme ce qui conduit à un dysfonctionnement neuropsychiatrique (Seeburg and Hartner, 2003). En ce sens, chez l'Homme, il apparaît que l'édition de ce neurotransmetteur est très affectée chez des patients schizophrènes (Sodhi et al., 2001) ou victimes de suicide (Niswender et al., 2001; Gurevich et al., 2002).

Par une étude phylogénétique, une nouvelle cible de la protéine ADAR1 a été identifiée dans le gène *Kv1.1* codant pour un canal potassique (Hoopengardner et al., 2003). Chez l'Homme, le transcrite correspondant *hKv1.1* est édité spécifiquement de sorte que l'isoleucine à la position 400 soit remplacée par une valine. Cet acide aminé est localisé dans le pore de ce canal et l'édition à ce site entraîne une inactivation rapide de son fonctionnement (Bhalla et al., 2004). Il apparaît que des mutations dans cette région du pore, chez l'Homme, sont associées à une maladie neurologique (EA-1 pour episodic ataxia type1) (Kuryatov et al., 1997; Maylie et al., 2002).

### Les protéines ADARs chez les invertébrés

Moins d'études ont été menées chez les invertébrés, cependant il apparaît que dans ce groupe, les protéines ADARs jouent également un rôle clé dans l'édition des transcrits codant pour des protéines du système nerveux central, en particulier de nombreux canaux ioniques.

Ainsi, chez la drosophile, la protéine dADAR est impliquée dans l'édition d'un canal sodium codé par le gène *PARA* (Reenan et al., 2000). De plus, un orthologue du gène *Kv1.1* (canal potassique), a été identifié comme une cible probable de cette protéine dADAR (Hoopengardner et al., 2003). De même, chez le *C. elegans*, des canaux potassiques semblent être édités par les protéines ADARs (Patton et al., 1997; Rosenthal and Bezanilla, 2002).

### *3) Désaminations qui affectent l'épissage*

Comme nous l'avons déjà évoqué, la formation d'un ARN double-brin se fait généralement par un repliement « intramoléculaire » associant une séquence exonique avec une séquence intronique complémentaire nommée ECS. L'édition peut donc survenir dans un intron et modifier l'épissage de ce dernier, et donc étendre considérablement le protéome d'une espèce (voir figure n°13).

Supportant cette idée, il a été montré que dans la lignée de souris ADAR2<sup>-/-</sup>, la cinétique d'épissage du transcrite GluR-B diminue d'un facteur 10 (Higuchi et al., 2000). De plus, l'édition dans le transcrite de la tyrosine phosphatase PTPN6 fait disparaître un site 3' accepteur d'épissage, et le variant protéique ainsi produit semble impliqué dans la formation

de leucémies (Beghini et al., 2000). Chez l'Homme, l'épissage du transcrite du récepteur de la sérotonine (5-HT<sub>2C</sub>R) est également régulé par l'action des protéines ADAR qui peuvent désaminer différenciellement 5 sites exoniques et 1 site intronique (Flomen et al., 2004). Enfin, l'exemple le mieux caractérisé démontrant que l'édition A-en-I affecte l'épissage chez les mammifères est celui de « l'auto-édition » de l'ARN pré-messager *ADAR2*. En effet, la protéine *ADAR2* elle-même édite la position -1 de son transcrite, introduisant un site d'épissage alternatif 3' accepteur. La protéine ainsi produite est totalement non fonctionnelle suggérant que cette « auto-édition » permet de contrôler le niveau de protéine *ADAR2* dans la cellule (Rueter et al., 1999; Maas et al., 2001). De façon intéressante, il a été souligné que la protéine *ADAR1* pourrait éditer la position -2 de ce même transcrite en abolissant ce site alternatif d'épissage (Dawson et al., 2004) (figure n°13).

#### 4) Désaminations dans les séquences non codantes

Depuis la découverte de l'édition A-en-I, seul un nombre restreint de sites d'édition dans les ARNm ont été identifiés. A l'ère postgénomique, les outils bioinformatiques ont permis de montrer qu'en réalité, chez les primates, la vaste majorité des sites d'édition A-en-I pourraient se retrouver dans les éléments répétés et abondants comme les transposons Alu et LINE1 (Kikuno et al., 2002; Athanasiadis et al., 2004; Blow et al., 2004; Kim et al., 2004; Levanon et al., 2004). Par exemple, des sites potentiels d'édition ont été rapportés pour plus de 2600 ARNm (Kim et al., 2004; Levanon et al., 2004). Il est connu que ces éléments, grâce à leur séquences inversées-répétées, peuvent former de longues structures d'ARN épingle à cheveu (ou hairpin) (Batzer and Deininger, 2002; Jasinska and Krzyzosiak, 2004). Ces sites sont localisés en grande majorité (>80%) dans les transposons Alu localisés en 5'UTR, 3'UTR et surtout dans les introns des transcrits (Athanasiadis et al., 2004; Blow et al., 2004; Kim et al., 2004; Levanon et al., 2004).

Pris ensemble, ceci augmente considérablement le nombre de cibles potentielles des protéines ADAR et l'enjeu actuel est de comprendre quelles sont les conséquences fonctionnelles de l'édition de ces séquences. Des études bioinformatiques ont permis de proposer que l'édition pourrait stabiliser ou déstabiliser les éléments Alu (Athanasiadis et al., 2004; Blow et al., 2004; Levanon et al., 2004). De plus, des éléments suggèrent que l'hyper-édition de séquences LINE1 et Alu pourrait être impliquée dans la défense antivirale (Zhang and Carmichael, 2001) ou interférer avec le mécanisme de transposition de ces éléments (Schramke and Allshire, 2003, 2004). Enfin, l'édition de ces éléments pourrait affecter la maturation, la stabilité, le transport, la localisation l'épissage ou encore la traduction d'un ARNm.

#### 5) Rôle dans la défense contre les virus

Plusieurs études indiquent que les protéines ADARs, via leur activité de désamination A-en-I, participent à la défense antivirale (Bass, 2002). Les ARN viraux peuvent former des structures d'ARN double-brin de type hairpin (O'Hara et al., 1984; Hajjar and Linial, 1995) ou par association avec un ARN antisens (Kumar and Carmichael, 1997). De ce fait, les protéines ADAR peuvent introduire des mutations A-en-I sur le brin sens ou U-en-C sur le brin complémentaire des ARNs viraux.

Il a été clairement démontré que les protéines ADARs peuvent hypermuter de façon non spécifique l'ARN viral de la rubéole (Billeter et al., 1994; Cattaneo, 1994) et du polyoma

virus (Kumar and Carmichael, 1997). Ces hypermutations semblent permettre une rétention des ARN viraux dans le noyau (Zhang and Carmichael, 2001).

Par ailleurs, il a été montré que l'antigénome du HDV (Hepatitis Delta Virus) est édité par la protéine ADAR1 au niveau d'un site spécifique convertissant un codon stop en un codon tryptophane (Kos et al., 1986; Polson et al., 1996; Taylor, 1999). Rappelons que l'expression de la protéine ADAR1p150 est cytoplasmique et inductible par un interféron (George and Samuel, 1999b). De plus dans le cytoplasme, l'existence d'une activité ribonucléase (I-RNase) spécifique des paires I : U et U : I a été révélée (Scadden and Smith, 1997).

Pris ensemble, tous ces éléments ouvrent des pistes quant au rôle des protéines ADARs, en particulier la protéine ADAR1, dans la défense contre les virus, mais il faudra attendre des études supplémentaires pour mieux caractériser leur action.

### 6) Rôle dans l'interférence à l'ARN

Plusieurs éléments suggèrent l'existence d'un lien entre la voie d'édition A-en-I et la voie de l'interférence à l'ARN.

En effet, chez *C. elegans*, les individus d'une lignée ayant une double délétion des gènes *ceADAR1* et *ceADAR2* présentent des anomalies de locomotion et ce phénotype peut être complété par croisement avec une lignée déficiente pour l'« ARNi » (Tonkin and Bass, 2003). Pour expliquer une telle observation, il a été proposé que des transcrits normalement édités chez un individu sauvage « échappent » aux voies d'« ARNi ». Par ailleurs, chez les souris *ADAR1*<sup>-/-</sup>, l'efficacité de « silencing » de certains siRNA est significativement augmentée (Yang et al., 2005). Il apparaît donc que la protéine ADAR1 pourrait limiter le silencing par la voie des siRNA.

En outre, l'édition A-en-I de plusieurs microARN a été rapportée (Chen et al., 2004; Luciano et al., 2004; Pfeffer et al., 2005). Les conséquences fonctionnelles de l'édition A-en-I sont doubles.

D'une part, l'édition A-en-I des microARN par les protéines ADARs pourrait réguler l'expression des microARN. Par exemple, il a été montré pour le microARN 142 que son édition, par les protéines ADAR1 et ADAR2, altère sa structure en tige-boucle si bien qu'il n'est plus reconnu par les protéines de la voie d'« ARNi », telle que la protéine Drosha (Chen et al., 2004). De plus la présence de mésappariements I : U et U : I favorise sa dégradation (Caudy et al., 2003; Scadden, 2005). Enfin, dans les souris *ADAR1*<sup>-/-</sup> et *ADAR2*<sup>-/-</sup>, il a été montré que le taux de miARN142 mature est augmenté par rapport à un individu sauvage (Yang et al., 2006).

D'autre part, une étude récente a montré que l'édition d'un pré-miARN (miARN376) par les protéines ADARs peut ne pas affecter sa maturation, mais qu'un seul changement A-en-I dans sa séquence est suffisant pour qu'il reconnaisse une nouvelle cible (Kawahara et al., 2007). Dans ce sens, on sait que les protéines ADARs « hyper-éditent » les régions d'ARN non traduites, en particulier le 3'UTR des transcrits contenant des séquences Alu chez les primates. Or, les cibles préférentielles des microARN sont justement ces régions 3'UTR. Ainsi, l'hyperédition des régions 3'UTR par les protéines ADARs pourrait faire apparaître ou disparaître des cibles pour des miARN. Une récente étude bioinformatique et statistique a permis de prédire que l'hyperédition des régions 3'UTR par les protéines ADARs pourrait effectivement modifier 10% des cibles des miARN (Liang and Landweber, 2007).

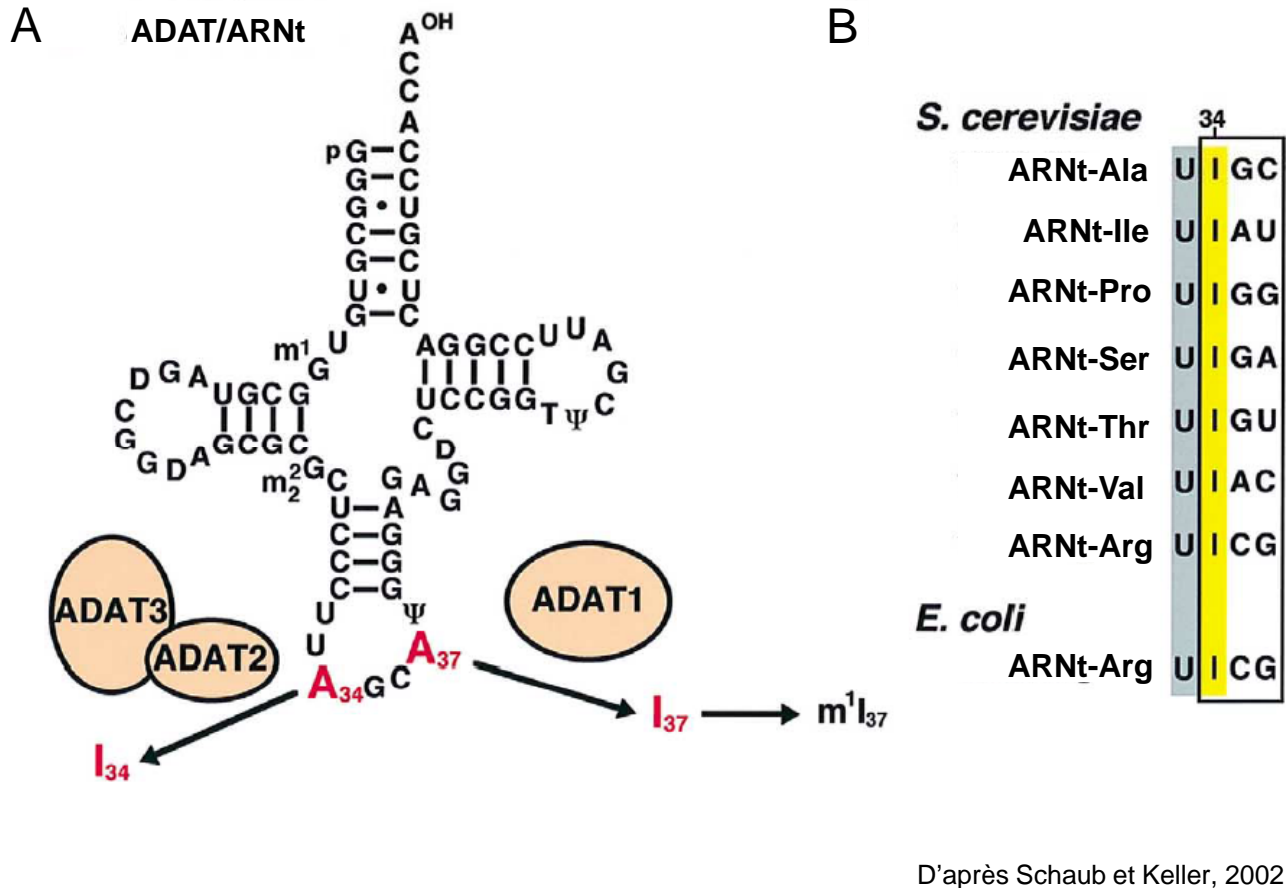


Figure n°14: Les protéines ADATs désaminent des résidus A-en-I dans les ARNt

- (A) Chez la levure, l'ARNt<sup>Ala</sup> présente une inosine en position 34 qui correspond à la première position de l'anticodon et une N1-méthylinosine (m1I) à la position 37 qui correspond à la première base adjacente en 3' de l'anticodon. La protéine ADAT1 est impliquée dans la désamination de l'adénosine A37; les protéines ADAT2 et ADAT3 qui forment un hétérodimère pour désaminer l'adénosine A34.
- (B) Les ARNt de *S. cerevisiae* et *E. coli* contenant des inosines

Ainsi, la désamination A-en-I par les protéines ADARs pourrait réguler à différents niveaux la voie de « ARNi » chez les métazoaires.

### III.2.6 Cas particulier des désaminations A-en-I dans les ARNt

#### 1) La famille ADATs

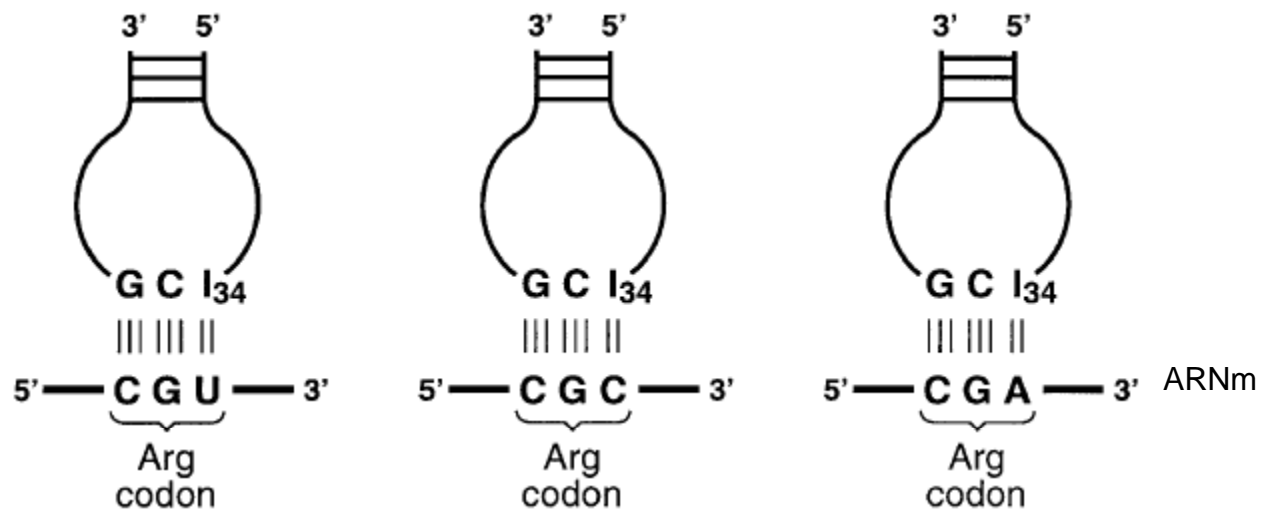
L'édition par désamination A-en-I des ARNt est décrite chez un grand nombre d'organismes des bactéries jusqu'à l'Homme. En effet pour une majorité d'eucaryotes, l'ARNt<sup>Ala</sup> présente une inosine en position 34 (qui correspond à la première position de l'anticodon) et une N<sup>1</sup>-methylinosine (m<sup>1</sup>I) à la position 37 (qui correspond à la première base adjacente en 3' de l'anticodon) (Holley et al., 1965; Grosjean et al., 1996; Sprinzl et al., 1998; Bjork et al., 2001) (figure n°14A). De plus, chez certains eucaryotes supérieurs, 8 autres ARNt (et seulement 7 autres chez la levure) présentent une inosine uniquement à la position 34 (Sprinzl et al., 1998) (figure n°14B). Chez les bactéries et dans chloroplastes, seul l'ARNt<sup>Arg</sup> présente une inosine en position 34, et aucune inosine n'a été identifiée dans les ARNt des mitochondries et des archéobactéries (Grosjean et al., 1995; Sprinzl et al., 1998).

La réaction de désamination des ARNt est catalysée par les enzymes de la famille ADATs (pour **adenosine deaminases that act on tRNA**). Les protéines de la sous-famille ADATs ont été identifiées par leur homologie de séquence avec les protéines ADARs. En effet, les protéines ADATs possèdent un domaine catalytique C-terminal similaire à celui des protéines ADARs, mais aucun motif de fixation à l'ARN double-brin (figure n°9) (Gerber and Keller, 1999). Il est couramment admis que les protéines de la famille ADARs ont évolué à partir des protéines des protéines ADATs par l'acquisition des domaines dsRBD.

Contrairement aux protéines ADARs qui semblent spécifiques des métazoaires, les protéines ADATs sont présentes chez de nombreux, si ce n'est pas tout les organismes. La famille de protéines ADATs consiste en trois membres : les protéines ADAT1, ADAT2 et ADAT3 qui sont conservées de la levure jusqu'à l'Homme et confèrent une spécificité de désamination (Gerber et al., 1998; Gerber and Keller, 1999; Maas et al., 1999; Maas et al., 2000).

En effet, chez la levure, la protéine ADAT1 agit sous forme d'homodimère pour désaminer l'adénosine A37 dans l'anticodon, et n'est pas essentielle à la survie cellulaire (Gerber et al., 1998). En revanche, les protéines ADAT2 et ADAT3 qui forment un hétérodimère pour désaminer l'adénosine A34, sont essentielles à la survie cellulaire (Gerber and Keller, 1999).

Plus récemment, les protéines ecTadA et saTadA, orthologues vraisemblables de la protéine ADAT2, ont été identifiées chez *E. coli* et chez *Staphylococcus aureus* respectivement, où elles agissent sous la forme d'un homodimère et désaminent exclusivement le résidu A34 de l'ARNt<sup>Arg2</sup> (Wolf et al., 2002; Losey et al., 2006). De même, les protéines TbADAT2 et TbADAT3, orthologues des protéines ADAT2 et ADAT3, ont été identifiées chez le trypanosome (Rubio et al., 2007) où elles forment un hétérodimère pour désaminer non seulement le résidu A34 en inosine de l'ARNt<sup>th</sup>, mais aussi, de façon plus surprenante, le résidu C32 en uracile.



D'après Schaub et Keller, 2002

Figure n°15: Modèle de fonctionnement de l'ARNt<sup>Arg</sup> contenant une inosine en position 34.

Le résidu inosine pourrait s'apparier avec une base U, C ou A à la troisième position du codon et insérer un acide aminé Arginine dans la protéine correspondante.

## 2) *Activité catalytique et reconnaissance du substrat*

A l'heure actuelle, nous disposons seulement d'un petit nombre d'informations quant au mode de fonctionnement des enzymes ADATs. Cependant, les sites catalytiques des protéines ADARs et ADATs sont très similaires. De ce fait, les protéines ADATs doivent catalyser une désamination hydrolytique en fixant du zinc. De façon intéressante, l'absence du résidu glutamate dans le site catalytique de la protéine ADAT3 suggère que dans l'hétérodimère ADAT2/ADAT3 la désamination A-en-I soit catalysée par la protéine ADAT2 (Gerber and Keller, 1999).

Les protéines de la famille ADATs, comme les ADARs, pourraient avoir évolué à partir de cytidines désaminases agissant sur des mononucléotides (CDAs). De même, le fait que, chez le trypanosome, le complexe TbADAT2/TbADAT3 présente à la fois une activité adénosine désaminase et cytidine-désaminase, renforce d'autant plus cette hypothèse (Rubio et al., 2007). Récemment, un cristal des protéines ecTadA et saTadA a été obtenu (Kim et al., 2006; Losey et al., 2006). Leur étude a permis de mettre en évidence que ces protéines semblent plus proches de cytosines désaminases que des cytidines désaminases (voir paragraphe III.3).

L'absence de domaine de liaison dsRBD chez les protéines ADATs implique qu'elles reconnaissent leur substrat via un nouveau mode d'interaction protéine/ARN. D'ailleurs, en dépit de leur ressemblance, les protéines ADATs et ADARs ne peuvent pas agir sur les mêmes ARN substrats (Schaub and Keller, 2002).

Chez les eucaryotes, des expériences de mutagenèse dirigée dans l'ARNt<sup>Ala</sup> suggèrent que sa structure tridimensionnelle est indispensable pour sa reconnaissance par la protéine ADAT1 (Gerber et al., 1998). De plus, la présence d'une purine en position 3' de l'adénine à modifier semble augmenter l'efficacité d'édition par les protéines ADATs (Auxilien et al., 1996; Grosjean et al., 1996). En revanche, chez les procaryotes, la désamination du résidu A34 semble être entièrement déterminée par l'identité des nucléotides de l'anticodon (Grosjean et al., 1996; Maas et al., 1999).

## 3) *Rôle de l'édition A-en-I des ARNt*

Il existe 61 codons qui spécifient 20 acides aminés différents, conservés universellement chez tous les organismes vivants, et la reconnaissance de ces codons est assurée par un petit nombre d'ARN de transfert. La présence d'inosine dans l'anticodon d'un ARNt joue un rôle essentiel dans la traduction car cela permet d'augmenter la capacité de reconnaissance de codons par les ARNt en autorisant de nouveaux appariements de bases (figure n°15).

Comme nous l'avons dit en introduction, plusieurs autres types de conversion de nucléotides ont été décrits dans les ARNt de nombreux organismes (les désaminations C-en-U, et A-en-I étant les plus fréquentes). Par ailleurs, il existe d'autres types de modifications dans les ARNt, comme la modification de certains nucléotides ou la polyadénylation de la partie 3'. Ces modifications résultent également en des changements significatifs dans l'appariement des bases. Bien souvent la distinction entre ces modifications et le processus d'édition est confuse. Par exemple, la désamination de l'adénine en position 37 de l'ARNt<sup>Ala</sup> est suivie d'une étape de méthylation résultant en la présence d'une N<sup>1</sup>-methylinosine (m<sup>1</sup>I) à la première base adjacente en 3' de l'anticodon (Gerber and Keller, 1999).

### *III.2.7 Bilan général sur les désaminations A-en-I*

Chez les métazoaires, des désaminations A-en-I par les protéines ADARs ont été mises en évidence pour un petit nombre de substrats, à l'heure actuelle.

Tout d'abord, il a été montré qu'elles affectent les régions codantes des transcrits des récepteurs de neurotransmetteurs et des canaux ioniques. Ces modifications conduisent à une diversification des protéines produites par un même gène et permettent une régulation fine du fonctionnement de ces différents composants cérébraux. Il apparaît donc que l'édition confère un avantage. Ainsi, lorsqu'il est apparu, ce processus a été non seulement conservé mais aussi étendu à un plus grand nombre de sites au sein des transcrits.

Dans ce sens, une analyse bioinformatique d'une banque d'ADNc de rat, suggère qu'il existe un très grand nombre de résidus inosine dans les ARNm, en particulier pour les transcrits spécifiques du cerveau où il a été trouvé en moyenne un résidu inosine tous les 17000 nucléotides (Paul and Bass, 1998). Ainsi, de nouvelles cibles des protéines ADARs pourraient venir à être identifiées.

Par ailleurs, aux cours de ces dernières années, il a été mis en évidence que nombreuses désaminations A-en-I pouvaient survenir dans les régions non-codantes des ARNm, augmentant considérablement le nombre de cibles de ces protéines. Ainsi d'autres rôles pourraient être attribués à l'édition A-en-I chez les métazoaires tels que : la régulation post-transcriptionnelle de l'expression des gènes, mais aussi la défense contre les virus, ou le maintien de l'intégrité du génome.

L'acquisition de l'édition A-en-I chez les métazoaires s'est faite par l'évolution d'une cytidine ou cytosine désaminase, en la famille de protéines ADATs (capable d'éditer les ARNt simple brin) puis ADARs (capable d'éditer les ARNm double-brin). Nous notons, dès à présent, un évident parallélisme entre l'acquisition de l'édition par conversions C-en-U chez les vertébrés, et l'acquisition de l'édition A-en-I chez les métazoaires. De plus, l'activité catalytique (désamination hydrolytique) et les rôles des protéines ADARs et APOBECs dans la cellule (diversification du protéome, défense contre les virus, maintien de l'intégrité du génome) sont également proches. Nous allons voir à présent qu'effectivement ces deux familles de protéine ont une histoire évolutive commune.

## **III.3 Histoire évolutive des enzymes d'édition des ARN**

Nous avons vu que chez les métazoaires l'édition des par conversion de bases consiste majoritairement en des désaminations C-en-U catalysées par la protéine APOBEC1, et A-en-I catalysées par les enzymes ADATs, et ADARs. D'autres exemples, rares, ont été rapportés comme des désaminations U-en-C ou G-en-A qui pourraient être liés à l'action de transaminases agissant sur des polynucléotides mais à ce jour de telles enzymes n'ont pas été caractérisées.

### *III.3.1 La superfamille des désaminases zinc-dépendantes*

La protéine APOBEC1 appartient à une famille cytidine-désaminase zinc-dépendante agissant sur des polynucléotides (ADN double-brin ou ARN simple brin). Les protéines ADATs et ADARs forment également une famille d'adénine désaminases zinc-dépendantes agissant sur



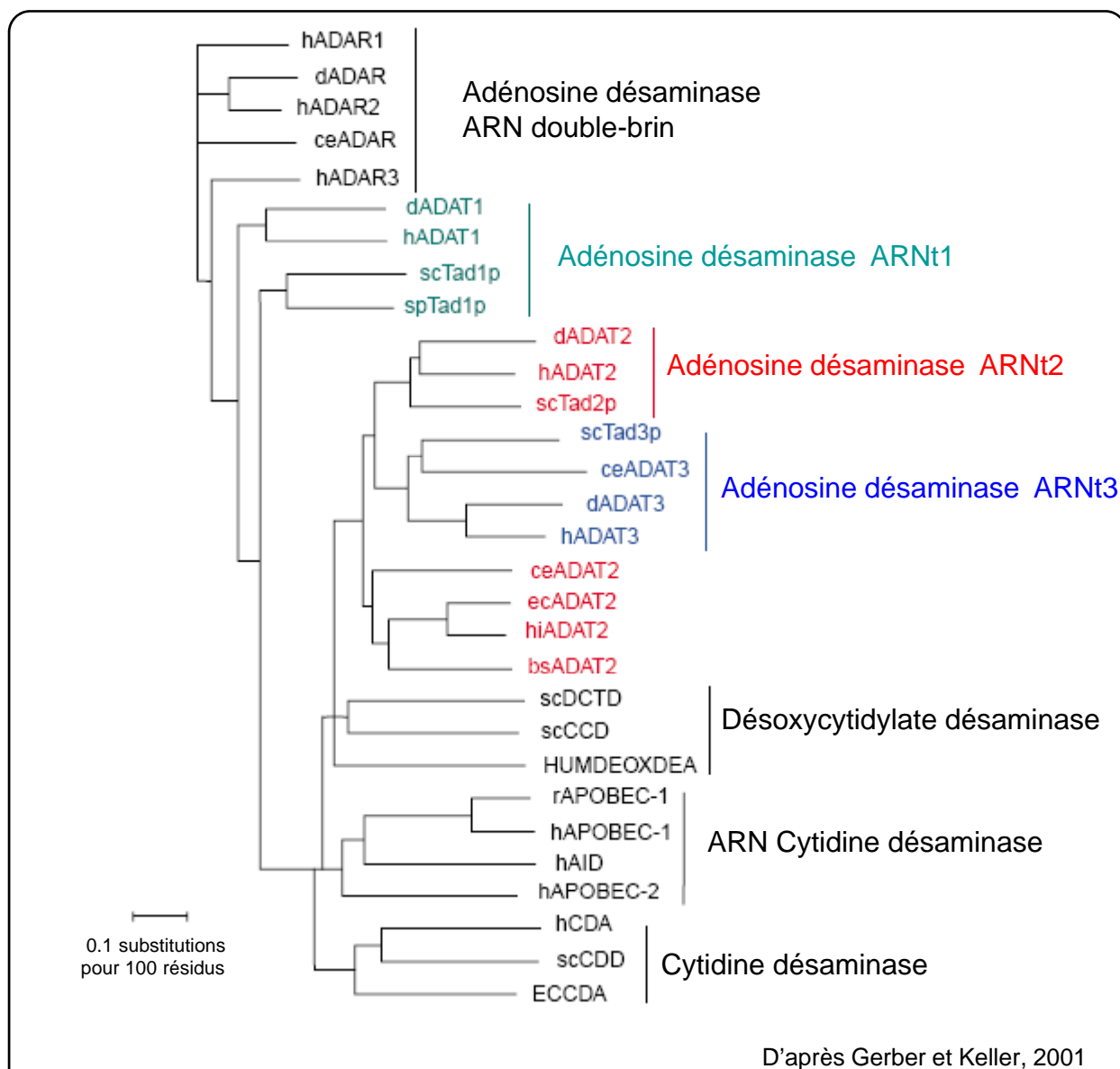


Figure n°16 : Arbre phylogénétique établi à partir des séquences du site catalytique (36 acides aminés) de nombreuses enzymes de la superfamille de désaminases zinc-dépendantes, présentant en particulier les désaminases catalysant les réactions d'édition des ARN.

<b>enzyme</b>	<b>substrats</b>	<b>organismes</b>
Cytosine désaminase	Cytosine/méthylcytosine	bactérie, archée, levure
Cytosine désaminase	Cytidine/deoxycytidine	bactérie, archée, levure, plante métazoaire
dCMP désaminase	dCMP	bactérie, archée, levure, plante métazoaire, virus
TadA/Tad2p-3p/ADAT2-3	Adénosine-34 dans ARNt	bactérie, levure, métazoaire
Tad1p/ADAT1	Adénosine-37 dans ARNtAla	levure, métazoaire
ADAR1-2-3	Adénine dans ARNm	métazoaire
AID/APOBECs	Cytosine dans ADN/ARNm/méthylcytosine	vertébré
Riboflavine désaminase	Synthèse des riboflavines	bactérie, archée, champignons, plantes
Guanine désaminase	Guanine	quelques bactéries, archée

D'après Conticello et al., 2007

Table n°1 : Membres de la superfamille de désaminases zinc-dépendante.

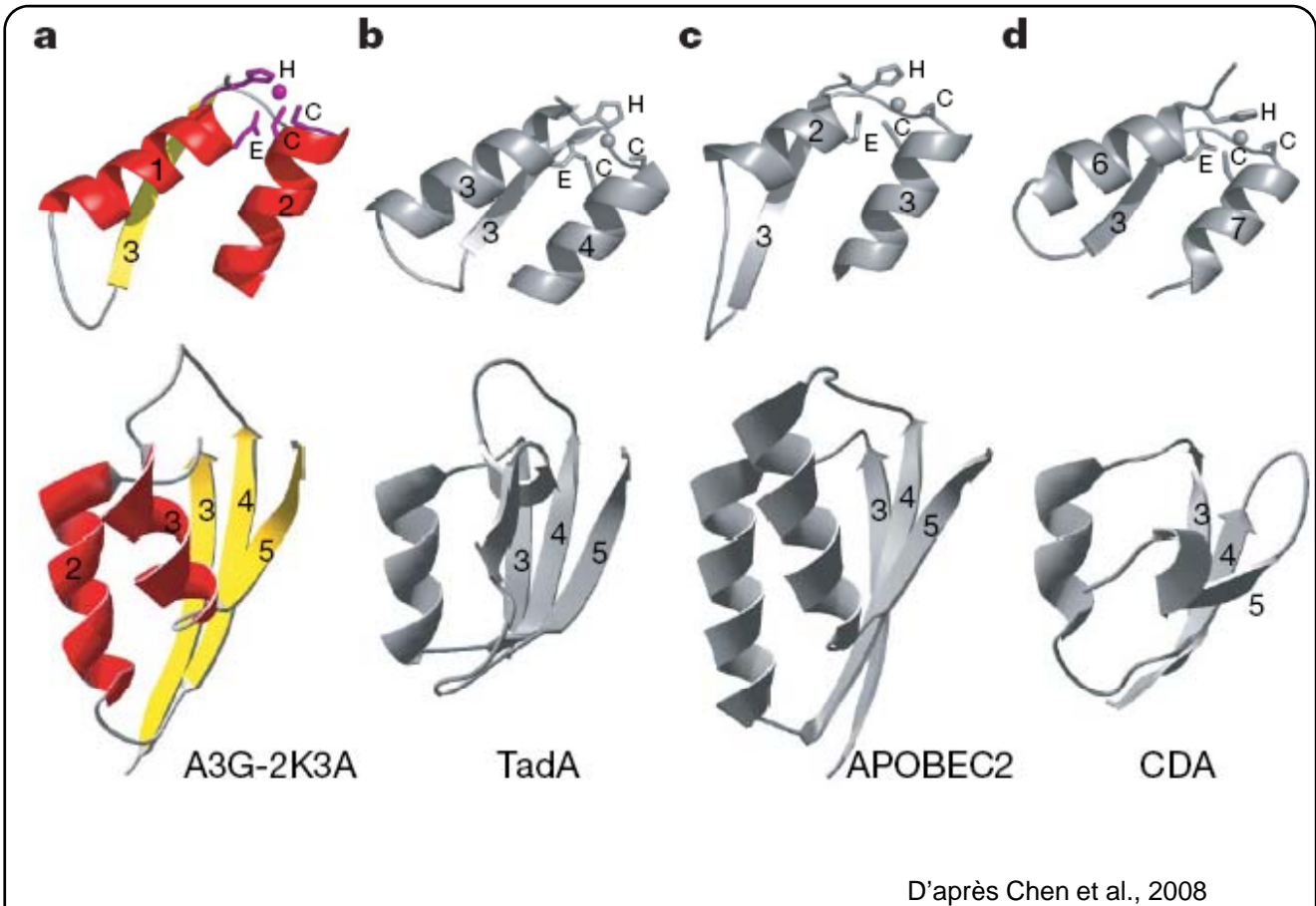
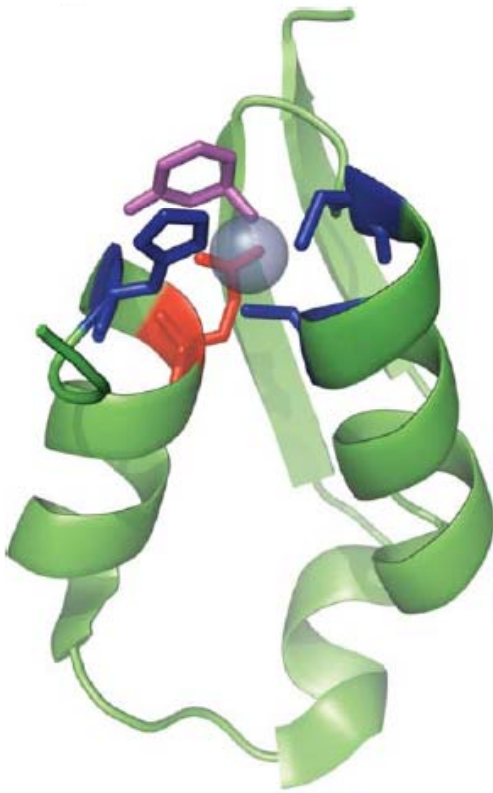


Figure n°17 : Structures du domaine catalytique de différents membres de la super famille de désaminases zinc-dépendante.

a) APOBEC3G humaine; b) saTadA de *S. aureus*; c) APOBEC2 humaine; d) ECCDA d'*E. coli*

Ligne du haut: le domaine de fixation du zinc (partie N-terminale du site catalytique)

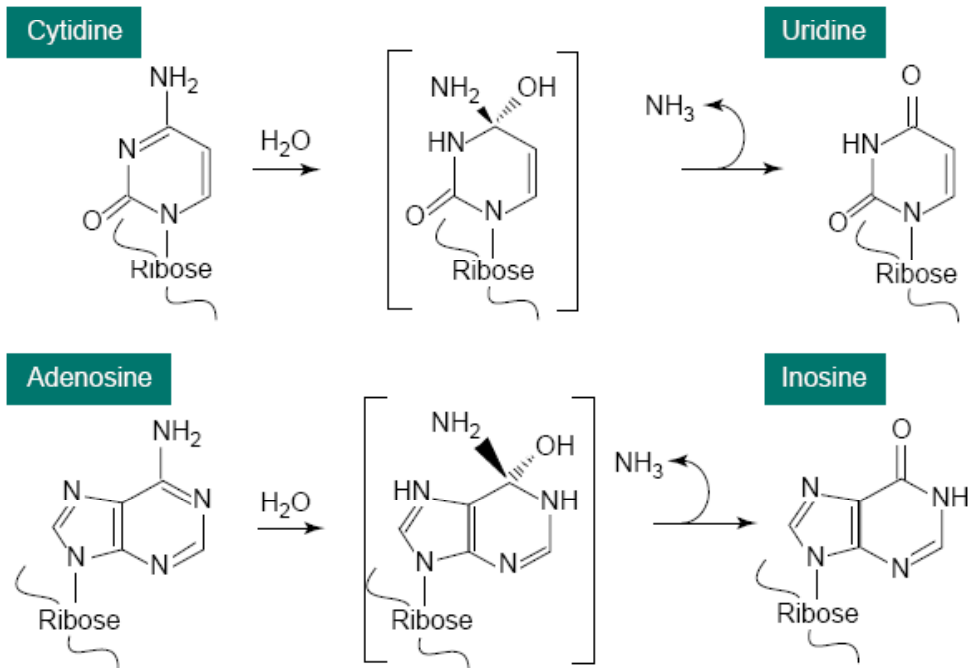
Ligne du bas: domaine d'interaction avec le substrat (partie C-terminale du site catalytique)



D'après Conticello et al., 2007

Figure n°18: Structure du domaine catalytique d'ECCDA d'*E.coli*

Ce domaine consiste en une poche où le zinc va être fixé grâce à deux cystéines (en bleu) et une histidine (en bleu) et est formé par deux hélices alpha d'orientation opposée. Un quatrième ligand participe à la fixation du zinc, il s'agit d'une molécule d'eau qui sert de donneur pour l'attaque hydroxyde de la réaction de désamination. L'hélice alpha qui contient le résidu histidine, contient également le glutamate (en rouge) qui sert de donneur de proton au cours de la réaction. La désamination est une hydrolyse initiée par l'attaque nucléophile de l'hydroxyde de zinc à la position C4 du résidu à désaminer, et est facilitée par la présence du résidu glutamate.



D'après Gerber et Keller, 2001

Figure n°19: Mécanisme des réactions de désamination de cytosine et adénine. Il s'agit d'une désamination hydrolytique, passant par un intermédiaire tétravalent, conduisant à la libération d'une molécule d'ammonium et d'un résidu uracile ou inosine, respectivement.

des ARNs simple brin ou double-brin. Une étude phylogénétique a permis de mettre évidence que toutes ces protéines appartiennent à la superfamille des désaminases zinc-dépendantes qui sont largement distribuées dans les différents groupes taxonomiques et agissent sur des purines et des pyrimidines (figure n° 16) (Gerber and Keller, 2001; Conticello et al., 2005). Cette super famille se compose, en outre des désaminases agissant sur des ADN et ARN (qui ne présentent pas d'activité sur des résidus monomériques), d'une grande majorité de désaminases agissant sur des bases libres, des nucléosides et des nucléotides monomériques (dCMP/cytosine/cytidine). Ces protéines permettent la formation d'uracile via l'hydrolyse de groupement amine du carbone C4 et participent à la voie métabolique des pyrimidines (table n°1).

### III.3.2 Corrélations structure-fonction au sein des désaminases

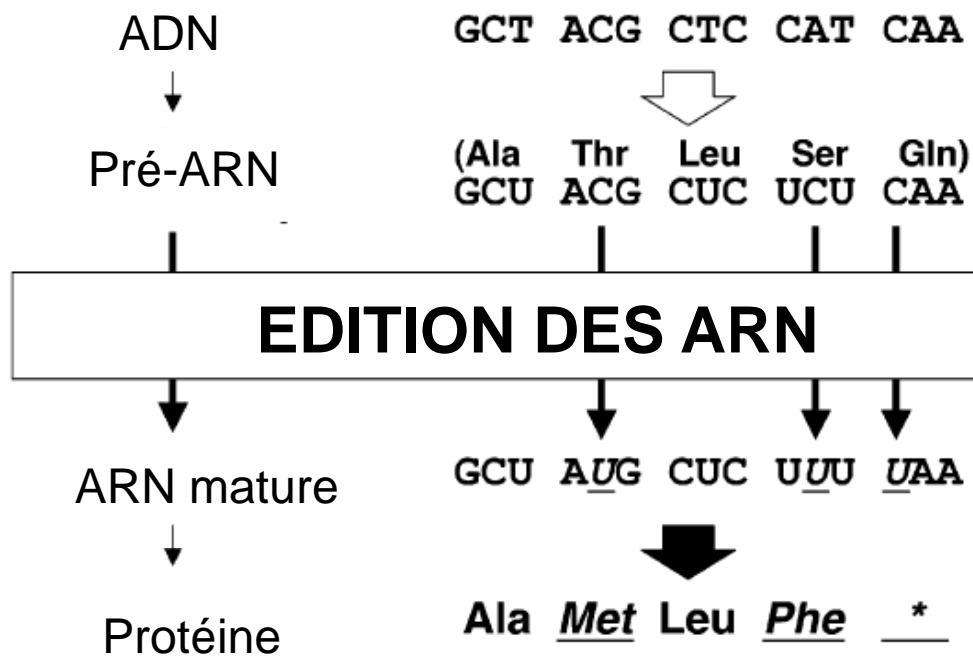
#### Comparaison de structures déterminées de certaines désaminases

Dans un premier temps, les structures cristallographiques de plusieurs cytidines et cytosines désaminases agissant sur des résidus libres ont été déterminées (Betts et al., 1994; Ireton et al., 2002; Johansson et al., 2002; Ko et al., 2003; Xie et al., 2004). Puis la structure d'une ADAT bactérienne, la protéine saTadA de *Staphylococcus aureus* a été obtenue (Losey et al., 2006). Enfin, récemment, les structures de deux membres de la famille APOBEC humaine (APOBEC2 et APOBEC3G) ont été publiées (Prochnow et al., 2007; Chen et al., 2008).

La comparaison de ces structures a permis d'identifier une organisation commune de la partie N-terminale du site catalytique (figure n°17, ligne du haut). Cette région correspond au motif de fixation du zinc (figures n°18 et n°19).

En revanche, l'organisation structurale de la partie C-terminale du site catalytique de ces protéines permet de distinguer les cytidines désaminases agissant sur des résidus monomériques des autres enzymes de cette super-famille (figure n°17, ligne du bas). En effet, dans les protéines APOBEC3G, APOBEC2 et saTadA, le feuillet beta du motif catalytique s'organise de façon parallèle à deux autres feuillets beta, alors que pour les cytidines désaminases agissant sur des résidus libres d'*E. coli*, de *Bacillus subtilis*, de *Saccharomyces cerevisiae* et humaine, les feuilles beta 4 et 5 sont antiparallèles et séparés par une petite boucle. Cette différence structurale est probablement « clé » pour fixer spécifiquement des substrats différents. En ce sens les études structurales, supportées par les expériences de mutagenèse dirigée, ont montré qu'en effet les résidus localisés juste après le feuillet beta 4 sont indispensables à la fixation à l'ADN pour les protéines APOBEC 3 (Langlois et al., 2005; Chen et al., 2008). De même la région équivalente de la protéine saTadA est impliquée dans la reconnaissance de l'adénosine34 dans l'ARNt substrat (Langlois et al., 2005; Losey et al., 2006).

L'étude de la structure quaternaire de ces protéines a révélé à nouveau une distinction entre les cytidines désaminases agissant sur des résidus monomériques des autres enzymes de cette super-famille. En effet, les cytidines désaminases agissant sur des résidus libres d'un grand nombre d'espèces forment des homotétramères contenant 4 sites catalytiques (Johansson et al., 2002). En revanche les cytosines désaminases, saTadA et APOBEC2 forment des homodimères à deux sites catalytiques (Ko et al., 2003; Losey et al., 2006; Prochnow et al., 2007).



D'après Takenaka et al., 2008

Figure n°20: L'édition des transcrits des organites chez les plantes.

Ce processus de régulation post-transcriptionnelle convertit des cytidines en uraciles dans les ARNm. Ceci conduit à une protéine ayant une séquence d'acides aminés différente de la séquence codée par l'ADN génomique. Plusieurs exemples de changements des acides aminés sont présentés.

Enfin, les études par RMN (résonance magnétique nucléaire) ont confirmé l'existence d'un mécanisme de « flipping out » pour les désaminases saTadA et APOBEC3G agissant sur des polynucléotides simple ou double-brin. Ce mécanisme permet de rendre la base à désaminer accessible au site catalytique (Losey et al., 2006; Chen et al., 2008).

#### Les prédictions similaires pour les autres désaminases agissant sur des polynucléotides

Bien que nous manquions d'informations structurales concernant les autres désaminases agissant sur des polynucléotides, on sait que toutes ces protéines ont une séquence conservée (H/C)(A/V)E(X<sub>24-30</sub>)(PCXXC) correspondant à la partie C-terminale du site catalytique et elles présentent toutes un long domaine C-terminal dont les prédictions de structures secondaires sont similaires à celles des cytosines désaminases, ou à la protéine saTadA. Par ailleurs, comme nous l'avons déjà évoqué, toutes ces protéines doivent former des homodimères ou hétérodimères pour être actives. Enfin, il a été proposé que les protéines ADARs ou la protéine AID utilisent un mécanisme de « base-flipping » rendant ainsi accessible la base à désaminer (Polson and Bass, 1994; Yu et al., 2004). De même, l'étude de la structure du *apoB* lors de sa désamination par APOBEC1 a également relevé l'existence d'un mécanisme de « base-flipping » (Maris et al., 2005).

Pour résumer, il apparaît que les enzymes catalysant les réactions d'édition des ARNs chez les métazoaires appartiennent à une superfamille de désaminases zinc-dépendantes. Parce que la cytidine-désaminase ECCDA d'*E.coli* fut la première enzyme de ce groupe dont la structure a été déterminée, il a été longtemps considéré que les enzymes de la famille APOBECs, ADARs et ADATs auraient une structure identique. En réalité, le site catalytique est conservé, mais ces enzymes sont plus proches des cytosines désaminases et leur partie C-terminale a évolué afin de s'adapter à un substrat polynucléotidique.

### **IV Le processus d'édition chez les plantes**

#### **IV.1 Généralités (Takenaka et al., 2008) (figure n°20)**

Chez les plantes, l'édition est un processus qui altère la séquence d'ARNs des organites, majoritairement par des conversions de résidu cytidine spécifique en uracile (C-en-U) et moins fréquemment d'uracile en cytidine (U-en C). Il a été mis en évidence, en 1989, dans les transcrits mitochondriaux (Covello and Gray, 1989; Gualberto et al., 1989; Hiesel et al., 1989) et en 1991 dans les transcrits chloroplastiques (Hoch et al., 1991).

L'édition des ARN des organites chez les plantes est un processus fréquent. En effet, dans les chloroplastes, approximativement 35 sites d'édition ont été répertoriés (Shikanai, 2006), en particulier 34 sites sont actuellement décrits chez *Arabidopsis thaliana* (Chateigner-Boutin and Small, 2007). Ce processus est encore plus abondant dans les mitochondries puisque 441, 427 et 491 sites d'édition ont été décrits chez *Arabidopsis thaliana*, *Brassica napus* et *Oryza sativa*, respectivement (Giege and Brennicke, 1999; Notsu et al., 2002; Handa, 2003). Les sites actuellement décrits se retrouvent majoritairement dans les ARNm codant des protéines, mais des sites édités dans régions transcrites mais non traduites, tels que les ARN de transferts et les introns sont également répertoriés. En revanche, les ARN ribosomiaux mitochondriaux et chloroplastiques ne semblent pas être édités. Actuellement, pour aucune espèce, les régions non codantes n'ont été analysées dans leur intégralité vis-à-vis de l'édition, si bien que les

chiffres actuels sous-estiment, très certainement, l'abondance réelle du phénomène d'édition dans les organites. Cependant, par extrapolation, on peut penser que dans le transcriptome mitochondrial des plantes à fleurs, il y aurait 20 à 50 sites d'édition supplémentaires (Takenaka et al., 2008).

La distribution asymétrique des sites d'édition entre les régions codantes et non codantes des transcrits suggère que ce phénomène est soumis à une certaine pression évolutive. Sans une nécessité fonctionnelle, le processus d'édition ne serait sans doute pas maintenu. Ce processus est une étape essentielle à l'expression des gènes mitochondriaux et chloroplastiques, en particulier pour permettre la synthèse de protéines fonctionnelles. En effet, l'édition peut conduire à des modifications de la séquence en acides aminés d'une protéine permettant ainsi la restauration d'acides aminés qui sont conservés chez des protéines homologues et qui sont essentiels à sa fonction (Bock et al., 1994; Sasaki et al., 2001). Ces modifications post-transcriptionnelles peuvent également permettre l'introduction d'un codon AUG initiateur de traduction (Hoch et al., 1991; Kotera et al., 2005), ou la génération d'un codon STOP (Wintz and Hanson, 1991). Par ailleurs, il a été montré que ce processus semble être indispensable à un repliement correct des ARN de transferts mitochondriaux afin qu'ils puissent assurer leur fonction (Marchfelder and Binder, 2004). De même, dans les introns, l'édition peut être le prérequis nécessaire à un repliement structural correct et un épissage efficace (Carrillo and Bonen, 1997). Ainsi, que ce soit dans les mitochondries ou dans les chloroplastes, l'édition est un processus essentiel pour assurer la maturation des ARN sans lequel les complexes de la chaîne respiratoire et photosynthétique ne pourraient être assemblés, maintenus et fonctionnels dans la cellule végétale.

L'édition des ARN a été décrit chez toutes les plantes terrestres, à l'exception d'un sous-groupe d'hépatiques, les Marchantiidae (Oda et al., 1992; Malek et al., 1996). Les hépatiques des autres groupes éditent avec une haute fréquence leurs ARN (Malek et al., 1996), suggérant que ce processus aurait été perdu secondairement chez les Marchantiidae (voir aussi discussion, partie 1 des résultats). De façon intéressante, les hépatiques constituent un groupe de plantes terrestres relativement proche des algues vertes, telles que les charales, pour lesquelles l'édition des ARN ne semble pas exister (Turmel et al., 2003). Ainsi, il est admis que l'édition serait apparue à l'émergence des plantes terrestres, il y a 450 millions d'années (Knoop, 2004).

#### **IV.2 Quel avantage évolutif à l'édition des ARN des organites chez les plantes ?**

Devant l'importance, tant quantitative que qualitative, du processus d'édition dans les transcrits des organites chez les plantes, très vite s'est posée la question « philosophique » du pourquoi : pourquoi ce processus si coûteux en énergie est apparu, et quels avantages offre-t-il aux plantes expliquant qu'il ait été maintenu ?

Si le rôle de l'édition est de rétablir des séquences qui ont évolutivement divergé, alors comme dans le modèle de la bureaucratie proposé par Muligan (Muligan, 2004), ce processus ressemble un vaste gaspillage d'énergie. En effet, il serait plus efficace de corriger l'information génétique directement au niveau de l'ADN génomique. Si l'édition est apparue à l'émergence des plantes terrestres et a été conservée pour pratiquement toutes les espèces, c'est que ce processus doit conférer un avantage sélectif.



Il apparaît que, en particulier dans les mitochondries, les transcrits peuvent être édités avec différents degrés selon les sites. De plus, bien que cela ne soit pas caractérisé en détails, il apparaît que l'édition varie en fonction des tissus végétaux, le stade de développement de la plante, et les conditions environnementales.

Dans un premier temps, il a été proposé que l'édition, par introduction d'un codon AUG initiateur de traduction ou d'un codon stop, pourrait permettre une régulation rapide de la traduction d'un transcrit et de ce fait de la quantité de protéine correspondante. Si on prend l'exemple du transcrit *ndhD* (codant pour une sous-unité du complexe NADH de la chaîne photosynthétique) chez *Arabidopsis thaliana*, l'édition introduit un codon AUG essentiel pour initier la synthèse de la protéine correspondante. L'édition est régulée par le développement de la plante, si bien que le transcrit *ndhD* n'est pas édité dans les racines (où le complexe NADH n'est pas actif), et est partiellement édité dans les feuilles (où le complexe est actif) (Hirose and Sugiura, 1997). Ainsi l'édition du transcrit *ndhD* paraît bénéfique pour la plante puisqu'elle permet une expression tissu-spécifique de la protéine correspondante. Cependant, il existe 4 autres sites d'édition dans ce transcrit qui doivent être également édités afin de produire une protéine fonctionnelle. Quel peut être le rôle de ces sites additionnels ? Par ailleurs l'expression du transcrit *ndhD* est régulée par un autre mécanisme post transcriptionnel : un clivage de l'ARNm entre le transcrit *psaC* et *ndhD* (Del Campo et al., 2002). Enfin, chez les monocotylédones, le codon AUG de ce transcrit est spécifié directement par la séquence d'ADN génomique suggérant que l'édition de ce site ne soit pas une étape essentielle au développement de toutes les plantes. Pris ensemble, tous ces éléments interrogent sur un réel effet bénéfique de l'édition pour réguler la synthèse de protéines dans les organites.

Par analogie à ce qui est décrit chez les animaux pour les récepteurs de neurotransmetteurs ou le transcrit *apoB*, il a été proposé que l'édition chez les plantes pourrait permettre de générer des variants protéiques. En effet, le transcrit *nad3* d'*Oenothera* contient plusieurs sites différenciellement édités, générant ainsi une population hétérogène de transcrits (Schuster et al., 1990). D'autres études expérimentales ont confirmé que l'édition conduit à une population hétérogène de transcrits qui peuvent permettre la synthèse de protéines ayant des différences mineures de séquence (Muligan, 2004). Cependant, l'étude des gènes mitochondriaux *nad9*, *atp6*, et *rps12*, dont l'édition permet de générer plusieurs transcrits, indique qu'une seule protéine est produite à partir de l'ARNm correctement édité (Grohmann et al., 1994; Lu and Hanson, 1994; Phreaner et al., 1996). Les molécules d'ARN complètement éditées permettent la synthèse de protéines plus conservées d'un point de vue évolutif, et doivent être sélectionnées (ou à l'inverse les variants protéiques sont contre sélectionnés) dans la cellule. Par ailleurs, l'intégration du gène mitochondrial *atp9* dans le génome nucléaire du tabac permet la production de cette protéine à partir d'un transcrit non édité. Cette dernière est correctement adressée aux mitochondries où sa présence (probablement par compétition avec la protéine endogène synthétisée à partir d'un transcrit édité) perturbe l'activité mitochondriale résultant en un dysfonctionnement de la production du pollen et une stérilité mâle (Hernould et al., 1993; Zabaleta et al., 1996). Ainsi, non seulement, l'édition chez les plantes ne conduit pas à une diversification protéique, mais de plus l'expression de tels variants apparaît délétère.

Une troisième alternative a été proposée quant au rôle de l'édition chez les plantes : ce processus pourrait prévenir le transfert de gènes des organites vers le génome nucléaire. En effet, l'édition dans le noyau n'a jamais été décrite chez les plantes. Ainsi pour qu'un gène des organites soit transféré fonctionnellement il faut que l'ADN à intégrer dans le génome

nucléaire provienne d'une molécule d'ARN totalement éditée et retrotranscrite. En outre, de façon intéressante, il se trouve que chez un grand nombre d'organismes, en particulier chez les animaux, l'usage de codons est différent dans le noyau et la mitochondrie, et ceci prévient le transfert fonctionnel de certains gènes mitochondriaux. Chez les plantes, sans doute suite à l'acquisition fortuite par le génome mitochondrial et chloroplastique de gènes nucléaires codant pour des ARNt, l'usage de codons est commun dans ces compartiments cellulaires, ainsi l'édition pourrait exercer une pression afin de maintenir le génome des organites.

### **IV.3 Comment une cytidine spécifique est reconnue dans un transcrit pour être modifiée ?**

Le fait qu'il soit possible d'observer des molécules d'ARN partiellement éditées indique que l'édition des molécules d'ARN est un processus post-transcriptionnel distinct de la transcription ; et qu'il n'existe pas un signal dans l'ADN génomique pour spécifier la base à modifier. De plus, l'existence de ces molécules d'ARN partiellement éditées suggère que chaque site pourrait être spécifié individuellement et que certains sites sont édités plus vite que d'autres (modèle dit de « hit and run » nous y reviendrons par la suite).

#### *IV.3.1 Une région d'ARN indispensable (dit « cis-élément ») pour la reconnaissance du site d'édition*

Si chaque site est spécifié individuellement, il va de soit que le contexte nucléotidique dans lequel se trouve la base à éditer est l'élément déterminant sa reconnaissance par la machinerie d'édition. Un calcul a permis de montrer que théoriquement une séquence de 9 nucléotides doit suffire à l'identification d'une base unique dans le transcriptome mitochondrial ou chloroplastique.

L'une des premières investigations menées pour caractériser ces « séquences » spécifiant la base à éditer a consisté en l'analyse de chimères naturelles (Lippok et al., 1994; Kubo and Kadowaki, 1997). En effet, lors de duplications internes du génome du riz (une monocotylédone) et d'*Oenothera* (une dicotylédone) des fragments chimériques naturels ont été générés pour certains gènes. Leur analyse a révélé que si la séquence en amont du site à éditer avait été conservée au cours de la duplication, alors l'édition avait également été conservée ; en revanche, lorsque c'est la séquence en aval du site d'édition qui a été dupliquée, l'édition est perdue. Ainsi, les éléments spécifiant les bases à éditer doivent être localisés en amont de celles-ci.

*In silico*, la comparaison des séquences entourant les 441 sites d'édition décrits dans le transcriptome mitochondrial d'*Arabidopsis thaliana* n'a pas permis d'établir un « consensus » ou un « motif commun » pouvant spécifier l'édition (Giege and Brennicke, 1999). L'unique tendance commune qui se dégage est la présence de pyrimidines en amont de la base à modifier plutôt que de rares résidus guanines qui pourraient rendre l'accès plus difficile à la machinerie d'édition.

Des expériences de transformation de chloroplastes de tabac, technique originellement développée chez l'algue verte *Chlamydomonas reinhardtii*, ont permis d'apporter des réponses à la question de la reconnaissance du site d'édition. Cette technique permet en effet,

l'introduction de gènes « étrangers », *in vivo*, dans les chloroplastes de tabac (plante dicotylédone) (Chaudhuri et al., 1995; Bock et al., 1996; Chaudhuri and Maliga, 1996).

Ainsi, chez le tabac, une région de 22 nucléotides est essentielle pour l'édition du site contenu dans le transcrit *psbL*. Cette région comprend 16 nucléotides en amont et 5 nucléotides après la cytidine à modifier (Chaudhuri and Maliga, 1996). De même, sur le transcrit *ndhB* qui contient deux sites d'édition *ndhB-6* et *ndhB-7*, il a été montré qu'un fragment d'ARN de -12 à -2 était essentiel pour l'édition (Bock et al., 1996).

Malheureusement, en raison de l'absence de marqueurs de sélection, la transformation de mitochondries, *in vivo*, n'est pas possible à l'heure actuelle. Cependant, un système d'édition *in organello* a été établi en transformant des mitochondries purifiées à partir du blé et du maïs (plantes monocotylédones) (Farre and Araya, 2001; Staudinger and Kempken, 2003).

Ces études ont permis de mettre en évidence que les 16 nucléotides en amont et les 6 nucléotides en aval de la cytidine à modifier dans le transcrit *coxII* suffisent à spécifier le site d'édition pour ces deux espèces (Farre et al., 2001; Choury et al., 2004). Si la taille de cette séquence minimale est identique pour les deux espèces, l'identité des nucléotides qui la compose diffère.

Les systèmes d'édition *in vitro* ont d'abord été mis au point en utilisant un extrait enrichi en protéines solubles de chloroplastes de tabac (Hirose and Sugiura, 2001; Miyamoto et al., 2002). Ils ont ensuite été établis de façon analogue à partir d'extrait mitochondrial de chou-fleur et de pois (plantes dicotylédones) (Takenaka and Brennicke, 2003; Takenaka et al., 2004). Ces approches ont permis de confirmer qu'une séquence « cis-élément » d'une vingtaine de nucléotides en amont de la base à éditer, et un seul nucléotide ou un nombre très restreint, en aval, sont nécessaires à l'édition (Hirose and Sugiura, 2001; Takenaka et al., 2004).

En conclusion, toutes ces approches ont permis de souligner l'existence d'une séquence « cis-élément » d'une vingtaine de nucléotides (essentiellement en amont de la base à éditer) nécessaire pour spécifier le site d'édition. L'obtention de résultats similaires chez différentes espèces de plantes (monocotylédones et dicotylédones) et pour les deux organites pourrait suggérer une origine commune de la machinerie d'édition à l'émergence des plantes terrestres et dont le fonctionnement a été conservé au moins chez toutes les plantes à fleurs.

#### IV.3.2 Existence de signaux plus éloignés (séquences « enhancer ») et effets longue distance

Dans le système d'édition *in vitro*, l'utilisation d'un extrait protéique mitochondrial de pois a permis de montrer que si l'édition du transcrit *atp9* est possible avec une séquence minimale de 20 nucléotides en amont de site d'édition, elle est encore plus efficace avec une séquence de 40 nucléotides (Takenaka et al., 2004). Par des expériences de mutagenèse dirigée, il a en effet été démontré que la séquence -5 à -20 nucléotides est essentielle à l'édition et que la séquence -20 à -40 influence l'efficacité d'édition. Son absence conduit à un rendement diminué de moitié, il s'agit donc d'une séquence « enhancer » ou « activatrice » d'édition (Takenaka et al., 2004; Neuwirt et al., 2005).

Le remplacement de la séquence -20 à -40 du pois par la séquence de chou fleur inhibe l'édition, ce qui a permis de montrer que ces séquences « enhancer » sont spécifiques à chaque espèce (Takenaka et al., 2004). Enfin, de façon inattendue, il a été montré qu'*in vitro*, cette séquence « enhancer » favorise également l'édition d'un deuxième site localisé à plus de

70 nucléotides en aval de cette séquence (van der Merwe et al., 2006). Cette dernière observation suggère que les facteurs de la machinerie d'édition progressent linéairement le long de la molécule d'ARN lors de ce processus (modèle du « scanning »). Cependant, ceci est contradictoire avec les observations faites précédemment indiquant que la machinerie d'édition sélectionne individuellement, et sans ordre, chaque site d'édition (modèle du « hit-and-run »). Ce modèle a été également soutenu par des expériences d'édition *in vitro*, au cours desquelles, il a été montré que d'une part, l'existence de séquences partiellement éditées est réelle et non artéfactuelle, et que d'autre part lorsque deux sites d'édition sont constitutifs le fait que le premier soit édité n'influe pas la probabilité d'édition du second (Verbitskiy et al., 2006). Ce résultat confirme que la machinerie d'édition reconnaît individuellement, et généralement sans ordre de préférence, chaque site.

Une explication alternative au modèle du « hit and run » et à celui de la progression linéaire serait d'imaginer que des facteurs, propres à chaque espèce (telle qu'une hélicase, par exemple) pourraient « préparer » la molécule d'ARN afin de la rendre plus accessible à la machinerie d'édition (Fairman et al., 2004).

#### IV.3.3 Un facteur agissant en trans dans le processus d'édition

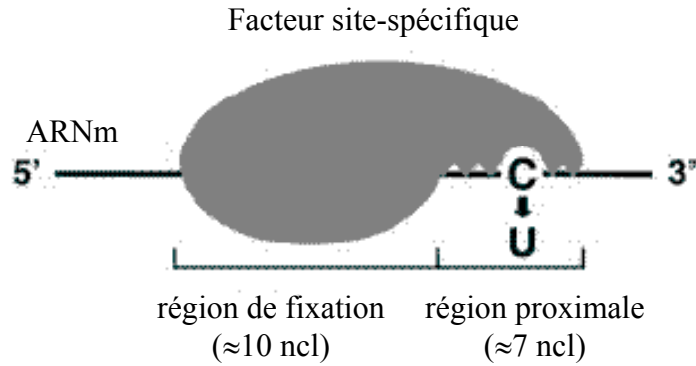
L'existence d'un facteur agissant en trans a d'abord été suggéré par les expériences d'édition réalisées *in vivo* en chloroplastes de tabac (Chaudhuri et al., 1995). En effet, l'introduction d'un transcrit chimérique contenant le site d'édition *psbL* entraîne une diminution significative de l'édition du site *psbL* endogène. Cet effet de compétition entre le site porté par le transgène et le site porté par le gène endogène souligne l'existence d'un possible trans-facteur en quantité limitante.

Cette hypothèse a été largement soutenue par l'approche *in vitro* utilisant un extrait chloroplastique de tabac (Hirose and Sugiura, 2001). En effet, l'édition des transcrits *psbL* et *ndhB* est bloquée par l'ajout d'un excès d'ARN contenant la région « cis-élément ».

Enfin, ces phénomènes de compétition ont été également mis en évidence dans des expériences réalisées *in organello* dans la mitochondrie de blé ou *in vitro* à partir d'un extrait protéique mitochondrial (Choury and Araya, 2006; Verbitskiy et al., 2006). Il apparaît qu'une séquence minimale d'édition (« cis-élément ») homologue à la séquence naturelle peut « compétitionner » pour l'édition, alors qu'une séquence mutagenisée ne le peut pas.

Il a été longtemps considéré que ces effets de compétition sont spécifiques de chaque site suggérant qu'il existe autant de trans-facteurs que de sites d'édition chez les plantes. Par ailleurs, pour un même site d'édition, l'effet de compétition est conservé au sein d'espèces différentes, indiquant que ces facteurs sont communs et conservés au moins chez certaines espèces de plantes à fleurs.

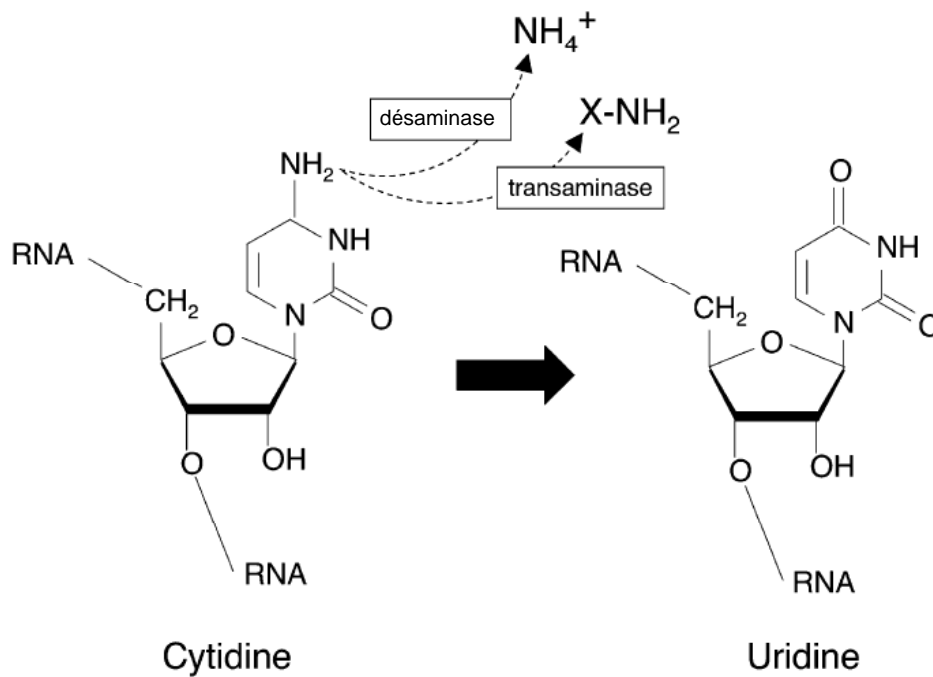
Cependant, la rigidité du modèle selon lequel chaque site d'édition est reconnu par un facteur différent est actuellement remise en cause. En effet, il a été montré qu'une forte surexpression, en chloroplastes de tabac, d'un transcrit *rpoB* ou *ndhF* entraîne une diminution de l'efficacité d'édition au niveau des sites endogènes correspondants mais aussi au niveau d'autres sites d'édition (Chateigner-Boutin and Hanson, 2002). De faibles similarités dans les séquences « cis-éléments » de tous ces sites existent suggérant qu'un même trans-facteur pourrait éventuellement reconnaître un « motif commun » dans ces séquences (Chateigner-Boutin and Hanson, 2002). Soutenant cette observation, l'existence d'un phénomène de compétition entre les sites d'édition *rpoB* C467 et *rps14* C80 a été montré récemment *in vitro*,



d'après Miyamoto *et al.*, 2004

Figure n°21 : Modèle du fonctionnement de la machinerie d'édition dans les chloroplastes.

Un facteur d'édition (ou facteur site-spécifique) se lie, dans un premier temps, avec une haute affinité à une séquence en amont du site d'édition (région de fixation) dont il détermine ainsi la spécificité. Puis ce facteur assure le changement C-en-U, directement ou via le recrutement d'autres facteurs inconnus.



D'après Takenaka *et al.*, 2008

Figure n°22: Réaction biochimique de l'édition chez les plantes

Le groupement amine du carbone C4 de la cytidine est retiré par désamination ou transféré sur une molécule accepteuse de groupement amine par transamination.

et il apparaît que 5 nucléotides de la séquence « cis élément » détermine cet effet (Heller et al., 2008).

La question de l'identité du facteur d'édition agissant *en trans* a alors été soulevée.

Dans un premier temps, l'utilisation d'un ARN guide pour spécifier un site d'édition, telle qu'elle a été caractérisée chez le trypanosome, a séduit les végétalistes. Cependant, lorsque les extraits chloroplastiques utilisés dans le système d'édition *in vitro* sont traités à la nucléase micrococciale (dégradant les ARNs), ces extraits sont toujours capables d'activité (Hirose and Sugiura, 2001). Aussi, par des essais d'« UV cross-linking », la présence d'un ARN associé au transcrit *psbL* n'a jamais été identifiée. En revanche, ces expériences d'« UV-cross linking » ont suggéré l'implication de facteurs protéiques (Hirose and Sugiura, 2001; Miyamoto et al., 2002, , 2004). Des protéines de 25, 56 et 70 kDa, se liant spécifiquement à la région dite « cis-élément » des transcrits *psbL*, *psbE*, et *petB* respectivement ont été isolées. Ceci montre que très certainement le trans-facteur spécifiant le site d'édition chez les plantes est une protéine, plutôt qu'un ARN guide (figure n°21). Récemment, soutenant l'idée qu'une même protéine pourrait spécifier deux sites d'édition distincts, il a été montré qu'une protéine de 95kDa se fixait sur la séquence « cis-élément » des transcrits chloroplastiques *ndhB9* et *ndhF1* de tabac (Kobayashi et al., 2008).

Ainsi, il existe plus de 400 sites d'édition chez les plantes à fleurs et ces sites doivent être spécifiés par un grand nombre de protéines conservées au cours de l'évolution. Nous verrons, dans la deuxième partie de l'introduction, que la récente découverte de la famille de protéines PPR (pour pentatricopeptides repeat) constituée de plus de 450 protéines conservées chez les plantes à fleurs, en font d'excellents candidats pour être ces trans-facteurs impliqués dans le processus d'édition (Small and Peeters, 2000; Lurin et al., 2004). En effet, ces protéines sont adressées, pour une grande majorité si ce n'est pas toutes, aux organites où elles participent à des processus de régulation post-transcriptionnelle des ARNm dont l'édition pour certaines d'entre-elles (Lurin et al., 2004; Kotera et al., 2005; Okuda et al., 2007; Chateigner-Boutin et al., soumis).

#### **IV.4 Quelle(s) enzyme(s) sont impliquées dans l'édition chez les plantes**

##### *IV.4.1 Est-ce qu'une cytidine-désaminase est impliquée dans l'édition ?*

Chez les plantes à fleurs, l'édition des transcrits chloroplastiques et mitochondriaux procède par des conversions C-en-U uniquement. L'édition inverse (conversion U-en-C) n'existe pas, ou à de très rares exceptions ; et n'a jamais été mise en évidence chez *Arabidopsis thaliana*, *Brassica napus* et *Oryza sativa* (Giege and Brennicke, 1999; Notsu et al., 2002; Handa, 2003). Lors des tests d'édition *in vitro*, il a été montré que ce processus procède par désamination ou transamination de cytidines en uraciles dans les transcrits chloroplastiques et mitochondriaux, et non pas par remplacement de nucléotides comme décrit chez le trypanosome ou *Physarum polycephalum* (Yu and Schuster, 1995; Hirose and Sugiura, 2001) (figure n°22). Ainsi, l'enzyme responsable de la réaction d'édition chez les plantes pourrait être une désaminase analogue à la protéine APOBEC1 qui catalyse la désamination C-en-U du transcrit *apoB* chez les mammifères.

Dans le génome d'*Arabidopsis thaliana*, il existe 9 gènes nucléaires codant pour des cytidines désaminases (Favre-Nitschke et al., 1999). Une ou plusieurs de ces protéines pourraient être

adressées aux mitochondries et/ou aux chloroplastes et catalyser la réaction d'édition. Pour évaluer cette hypothèse, des essais ont été développés lors des tests d'édition *in vitro* à partir d'extraits chloroplastiques ou mitochondriaux.

Toutes les cytidines désaminases fixent un atome de zinc dans leur site catalytique lors de la désamination, et l'utilisation de chélateurs de zinc, telle que l'*o*-phénanthroline, inhibe leur activité (Carter, 1998). Il apparaît que l'utilisation de chélateurs de zinc, inhibe l'édition *in vitro* à partir des extraits chloroplastiques, mais est sans effet à partir d'extraits mitochondriaux (Takenaka and Brennicke, 2003; Hegeman et al., 2005). De plus, lorsque le zinc est retiré des extraits mitochondriaux, cela n'a pas d'effet sur l'efficacité de la réaction d'édition *in vitro* (Takenaka and Brennicke, 2003). Ainsi, il est tout à fait possible qu'une cytidine-désaminase catalyse la réaction d'édition dans les chloroplastes. Cette hypothèse reste envisageable dans les mitochondries car d'une part, il a été montré que certaines enzymes fixent le zinc de façon « protectrice » dans leur structure de sorte qu'il est difficile à extraire et que les chélateurs sont sans effet sur leur activité (Spath et al., 2007) ; d'autre part, on peut imaginer qu'une cytidine-désaminase ait pu évoluer de sorte que le zinc ne soit plus essentiel à son activité.

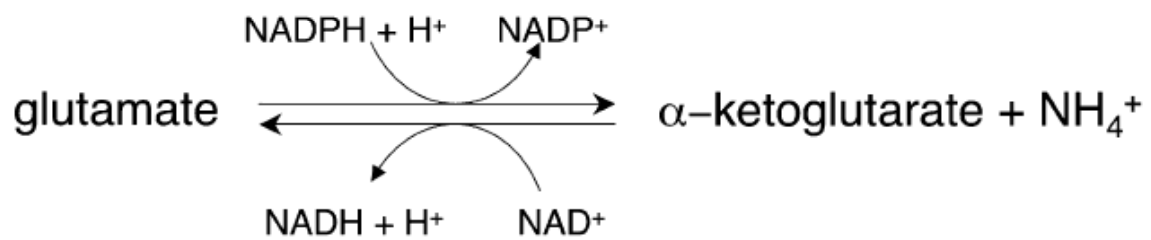
Cependant, une cytidine-désaminase « classique » ne peut pas, d'un point de vue énergétique, catalyser les réactions d'édition inverse U-en-C fréquente chez les mousses, les fougères, les lycopodiées et, en particulier chez les « Hornworts » (Malek et al., 1996; Knoop, 2004; Groth-Malonek et al., 2005). Ceci suppose que, soit deux enzymes distinctes catalysent ces deux types d'édition, soit avec une vue plus parcimonieuse, on peut supposer qu'une transaminase intervienne. D'un point de vue évolutif, on peut penser qu'à l'origine, la machinerie d'édition faisait intervenir une transaminase catalysant à la fois des désaminations C-en-U et des aminations U-en-C. Cette enzyme aurait été conservée chez les plantes « inférieures » pour lesquelles les deux types d'éditions co-existent. Elle aurait évolué chez les plantes à fleur en une désaminase capable uniquement de conversions de cytidines en uraciles.

#### IV.4.2 Est-ce qu'une transaminase est impliquée dans l'édition ?

Les réactions de transamination, qui permettent le transfert d'un groupement amine d'une molécule à l'autre, sont fréquentes dans différentes voies de biosynthèse des acides aminés et requièrent généralement un composé intermédiaire accepteur de groupement amine.

Dans les tests d'édition *in vitro* réalisés à partir d'extraits mitochondriaux, l'ajout de différentes molécules accepteuses de groupement amines, telle que l'oxaloacétate ou l' $\alpha$ -kétoglutarate, s'est révélé sans effet sur l'activité catalytique (Takenaka et al., 2007). De plus, ces molécules de petite taille (inférieure à 10kDa) sont certainement absentes des extraits mitochondriaux utilisés. Cependant, il est envisageable qu'une transaminase ayant évolué afin de présenter une activité catalytique sur un substrat polynucléotidique plutôt qu'un résidu libre utilise un autre type de molécules accepteuses de groupements amines, tels que des ARN par exemple. D'autre part, ces essais ont été mis au point à partir d'extraits mitochondriaux de pois, une espèce ne réalisant pas d'édition U-en-C, et de ce fait ne possédant peut être pas l'enzyme catalysant ce type de réaction. Ainsi, l'implication potentielle d'une transaminase dans les réactions d'édition ne peut pas être écartée.

Par ailleurs, les glutamates déshydrogénases (GDH), sont des enzymes capables de catalyser des désaminations ou des aminations en fonction de la concentration de différents régulateurs



D'après Takenaka et al., 2007

Figure n°23: Les glutamates déshydrogénases sont des enzymes qui catalysent des réactions de désamination ou d'amination en utilisant le NAD(P) ou NAD(P)H comme co-facteurs.

L'équilibre entre ces deux réactions est de nature allostérique et influencé par la concentration en ATP.



allostériques ou isostériques (figure n°23). De façon intéressante, il a été montré que de telles enzymes sont présentes dans les mitochondries de nombreuses plantes et peuvent se lier à l'ARN, ce qui en font de bons candidats pour catalyser les réactions d'édition (Takenaka et al., 2007). Ces enzymes utilisent le NADH (lors de désamination) et le NADP (lors d'amination) comme cofacteurs essentiels à leur activité. Il a été montré que l'ajout de NADH ou de NADPH, dans les tests d'édition *in vitro* réalisés à partir d'extraits mitochondriaux, active la réaction d'édition, alors que l'ajout de NAD ou NADP est sans effet (Takenaka et al., 2007). De plus, l'utilisation d'un inhibiteur des protéines GDH, la phosphinothricine n'affecte pas la réaction d'édition (Takenaka et al., 2007). Pris ensemble, ces éléments indiquent que les glutamates déshydrogénases ne sont sans doute pas les transaminases catalysant les réactions d'édition chez les plantes.

#### *IV.4.3 Est-ce qu'une CTP synthase est impliquée dans l'édition ?*

Les CTP synthases sont des enzymes capables d'amincer un résidu uracile libre afin de produire une cytidine. Une telle enzyme aurait pu évoluer afin de catalyser les réactions d'édition U-en-C sur un substrat ARN. Cependant, il est moins probable que ce type d'enzymes puisse catalyser les réactions d'édition C-en-U, car une activité de désamination n'a jamais été mise en évidence pour toutes les CTP synthases caractérisées à ce jour.

Ces enzymes requièrent pour leur activité des molécules donneuses de groupement amines de faible poids moléculaire, le plus souvent de l'ammonium ou de la glutamine. Comme évoqué précédemment, ces molécules de taille inférieure à 10kDa, sont certainement absentes des extraits mitochondriaux utilisés lors des tests d'édition *in vitro*. Cependant, pour les mêmes raisons que celles évoquées précédemment (possibilité qu'une enzyme ayant évolué à partir d'une CTP synthase utilise un autre type de molécule donneuse de groupement amine, et le fait les extraits mitochondriaux utilisés proviennent du pois, une espèce ne réalisant pas d'édition U-en-C), l'implication d'une telle enzyme dans le processus d'édition reste une hypothèse envisageable.

#### *IV.4.4 Quelle(s) autre(s) enzyme(s) pourraient être impliquée(s) dans l'édition ?*

Alternativement, il a été proposé que les enzymes de modification des ARNt ou moins probablement les enzymes impliquées dans les mécanismes de réparation de l'ADN, puissent également avoir évolué pour catalyser la réaction d'édition chez les plantes. Cependant, à l'heure actuelle, aucune expérimentation n'a été développée afin d'évaluer de telles hypothèses.

L'identification du ou des facteur(s) catalysant les réactions d'édition chez les plantes est une étape fondamentale à la compréhension de ce processus. Nous verrons dans la deuxième partie de cette introduction que certaines protéines PPR pourraient porter un domaine catalytique et qu'une partie de mon travail de thèse a été d'évaluer si ce dernier pouvait être le facteur enzymatique du processus d'édition chez les plantes.

## **V Bilan sur l'édition des ARNs**

### **V.1 Avantages et conséquences de ce mécanisme**

En résumé, l'édition des ARNs est un mécanisme qui altère l'information génétique codé par le génome.

Pour les ARN mitochondriaux des trypanosomes (et peut être pour les ARN mitochondriaux et chloroplastiques des plantes), ce mécanisme pourrait permettre de contrôler l'expression des gènes de façon post-transcriptionnelle permettant ainsi une rapide adaptation aux conditions environnementales et développementales. De plus, chez les plantes, l'édition pourrait également exercer une pression afin de maintenir l'intégrité du génome des organites. Si, chez les plantes, il est très clair que l'édition répare ou corrige l'information codée par le génome, chez les animaux, l'édition a pour but de diversifier cette information offrant à l'organisme une plus grande complexité. Récemment, chez les animaux, l'édition de séquences transcrites non codantes a révélé d'autres rôles possibles à ce processus dans la régulation de l'expression de gènes via la voie de l'ARN interférent, la défense contre les virus ou le maintien de l'intégrité du génome nucléaire.

### **V.2 Origine et évolution de ce processus**

Considérant les divers types et caractéristiques de l'édition des ARNs au sein des différents groupes taxonomiques, ce processus doit probablement être apparu plusieurs fois indépendamment au cours de l'évolution.

Chez les trypanosomes, la machinerie d'édition a évolué à partir d'enzymes dont l'activité existait déjà chez leur ancêtre commun, telles que des endonucléases, ligases... Il pourrait s'agir de protéines impliquées dans la réparation de l'ADN. Ces enzymes ont évolué et se sont associées pour former un nouveau complexe : l'éditosome. Chez les animaux, l'apparition de l'édition est liée à l'émergence d'une nouvelle classe d'enzyme, des désaminases capables d'agir sur des polynucléotides (dérivant de désaminases agissant sur des résidus libres).

Par ailleurs, trois mécanismes différents, permettant de spécifier le site d'édition, ont été caractérisés : l'utilisation de petits ARN guide, la reconnaissance d'une séquence nucléotidique autour du site d'édition, ou encore la reconnaissance de la structure secondaire de l'ARN cible.

Chez les plantes, bien que ce processus ait été mis en évidence il y a presque vingt ans, les différents facteurs de la machinerie d'édition ne sont pas encore clairement caractérisés à ce jour. Cependant, comme nous allons le décrire à présent, certains membres de la famille de protéines PPR semblent impliqués dans ce processus.

## **Partie 2 :**

### **LA FAMILLE DE PROTEINES PPR (pour pentatricopeptide repeat)**

#### **I Caractéristiques des protéines PPR**

##### **I.1 Mise en évidence d'une grande famille de protéines adressées aux organites chez Arabidopsis**

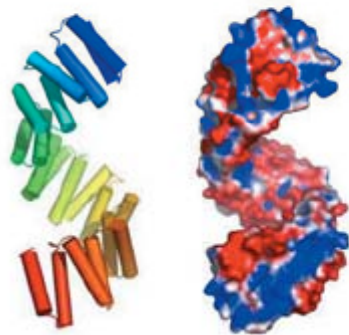
Lorsqu'en décembre 2000, la séquence complète du génome d'*Arabidopsis thaliana* a été publiée, l'existence de milliers de gènes, jusque là insoupçonnés, a été dévoilée. Les premières analyses de la séquence de ce génome ont permis d'estimer que 31% de ces gènes ne présentaient aucune homologie avec des gènes de fonction connue (Arabidopsis Genome Initiative, 2000). Par l'utilisation du logiciel Predotar, permettant la prédiction de la localisation subcellulaire des protéines, l'existence d'une grande famille de 450 protéines chez Arabidopsis, la famille PPR, a été mise en évidence (Small and Peeters, 2000).

En effet, au moins deux tiers des protéines de cette famille portent un signal d'adressage vers les organites (chloroplastes ou mitochondries). L'étude de localisation subcellulaire de 48 protéines PPR (ou de leur peptide d'adressage) fusionnées à une protéine fluorescente (GFP ou Ds-RedII) a été réalisée au laboratoire d'Evry. Ces travaux ont permis de confirmer les prédictions indiquant dans tous les cas une localisation chloroplastique ou mitochondriale (Lurin et al., 2004).

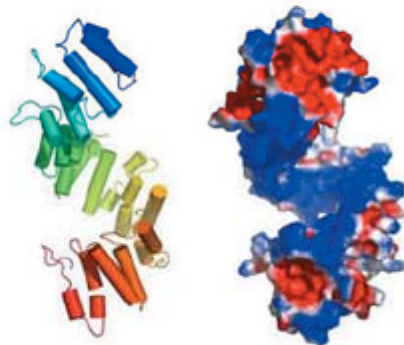
Cependant, d'après les logiciels de prédiction, près de 30% des protéines PPR ne portent pas de signal d'adressage vers ces deux organites. Ces protéines peuvent-elles avoir une autre localisation dans la cellule ?

D'après les données expérimentales actuelles, quatre protéines PPR apparaissent localisées hors des organites : la protéine BSF de drosophile (pour Bicoïde mRNA stabilisation Factor) (Mancebo et al., 2001) ; la protéine humaine LRP130 (Mili and Pinol-Roma, 2003) ; les protéines PPR d'Arabidopsis At1g10270 et GRP23 (Brown et al., 2005; Ding et al., 2006). Cependant, ces résultats restent à confirmer par des analyses complémentaires. Par ailleurs, au laboratoire d'Evry, l'étude de localisation subcellulaire de près de 100 protéines PPR pour lesquelles aucun adressage vers les organites n'a été prédit, est en cours. Les résultats préliminaires indiquent que la plupart d'entre elles sont localisées dans les mitochondries et plus rarement dans les chloroplastes. Cependant un petit nombre d'entre elles présentent une localisation nucléaire ou cytosolique et par extrapolation, il apparaît qu'un maximum de 20 protéines PPR (5% de la famille) pourrait être localisé en dehors des organites chez Arabidopsis.

GlcNac transférase  
10 motifs TPR



At1g05750  
10 motifs PPR



D'après Delannoy et al., 2007

Figure n°24: Comparaison des structures super hélicoïdales formées par la répétition de 10 motifs TPR de la GlcNac transférase et la répétition de 10 motifs PPR de la protéine At1g05750. Les chaînes exposées dans le sillon central de la GlcNac transférase sont acides alors que celles de la protéine PPR At1g05750 sont basiques (caractéristique commune à toutes les protéines PPR). Cette différence doit permettre aux protéines PPR d'interagir avec une molécule d'ARN plutôt qu'une protéine.

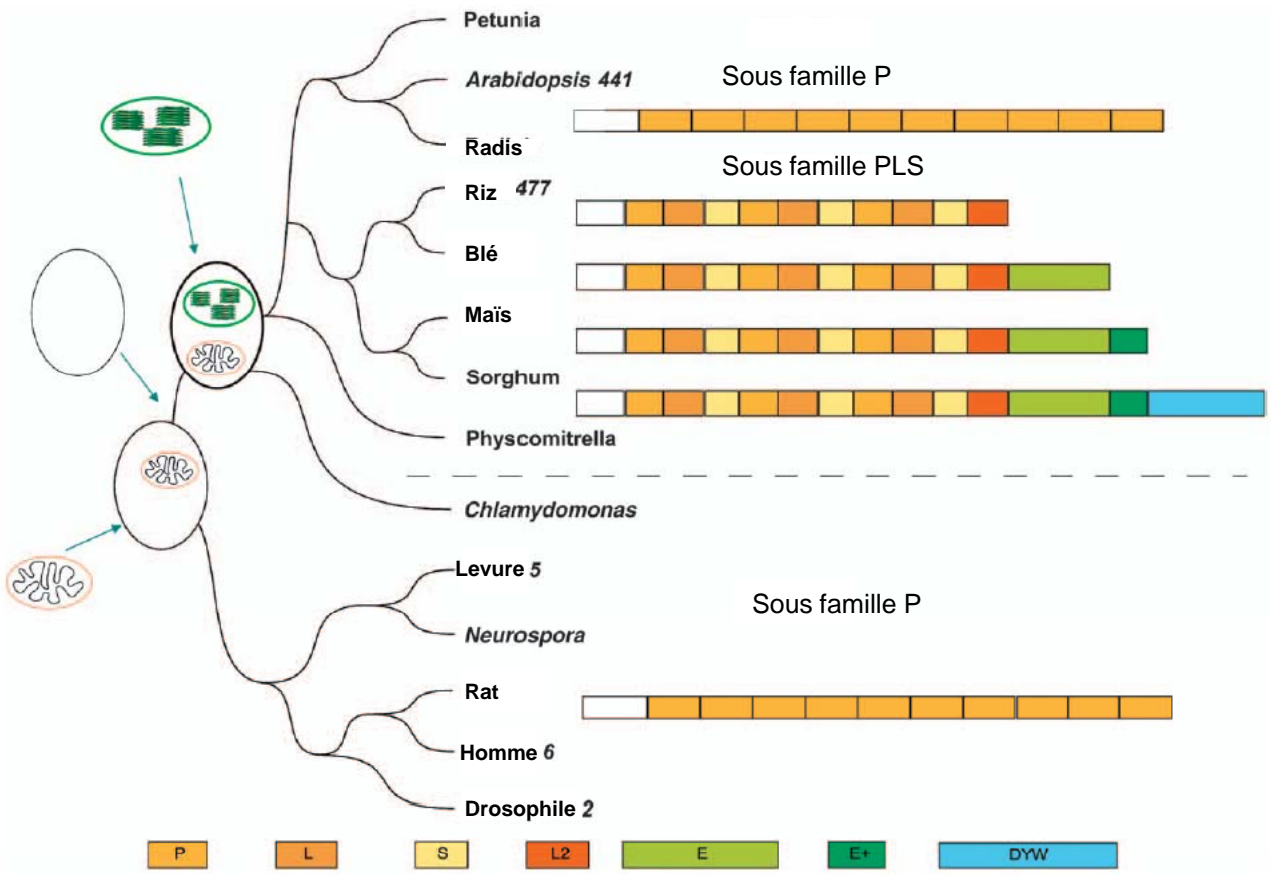
## **I.2 Le motif PPR : une séquence de fixation (spécifique) aux ARN**

La principale caractéristique des protéines PPR est la présence de répétitions en tandem d'un motif dit « PPR » (pour pentatricopeptide repeat) qui a donné son nom à la famille. Ce motif est constitué de 35 acides aminés, et le nombre de répétitions de ce dernier peut varier de 2 à 26 par molécule, avec une moyenne de 12 chez les plantes (Small and Peeters, 2000; Lurin et al., 2004). Le motif PPR est proche du motif TPR (pour tetratricopeptide repeat), impliqué dans des interactions protéine-protéine (Blatch and Lassle, 1999; Small and Peeters, 2000). Le motif TPR consiste en une paire d'hélices alpha antiparallèles, qui en répétitions en tandem dans les protéines, conduit à la formation d'une structure super hélicoïdale englobant un sillon central correspondant au site de fixation à un ligand protéique (Das et al., 1998). L'analyse des motifs PPR a également permis de proposer pour les enchaînements de motifs PPR une structure super hélicoïdale (Small and Peeters, 2000) (figure n°24). Cependant, à la différence de ce qui est observé pour les protéines TPR, les chaînes exposées dans le sillon central des protéines PPR seraient dans la majorité des cas hydrophiles et basiques ce qui suggère que ces protéines seraient dédiées à des interactions moléculaires avec des cibles chargées négativement (protéines acides ou acides nucléiques) (Small and Peeters, 2000).

En accord avec cela, des expériences *in vitro*, telles que des expériences de retard sur gel, d'« UV-cross-linking » ou d'affinité, ont démontré que des protéines PPR semblent interagir avec des acides nucléiques, notamment des ARNm (Meierhoff et al., 2003; Mili and Pinol-Roma, 2003; Lurin et al., 2004; Kobayashi et al., 2007). D'ailleurs, quelques protéines PPR, comme la protéine PPR4 du maïs, contiennent un motif RRM de fixation aux ARN, reconnaissable (Schmitz-Linneweber et al., 2006). Cependant, il a été montré que la protéine PPR-P63 de blé pourrait se fixer à l'ADN suggérant ainsi que certaines protéines PPR pourraient agir sur un ADN plutôt qu'un ARN (Ikeda and Gray, 1999).

Par ailleurs, il apparaît qu'en présence des 12 répétitions de motif PPR qui la composent, la protéine HCF152 se fixe à l'ARN avec une haute affinité, alors que 2 répétitions de motifs PPR uniquement ne suffisent pas à conférer une haute affinité de fixation à l'ARN (Nakamura et al., 2003). De façon contradictoire, pour la protéine humaine LRP130, 2 des 11 motifs PPR suffisent à une fixation affine sur un ARN (Mili and Pinol-Roma, 2003). Ainsi, l'ensemble des données publiées actuellement sur les protéines PPR permet d'affirmer qu'une vaste majorité d'entre elles peut se fixer aux ARN.

En outre, plusieurs études suggèrent que les protéines PPR semblent fixer une molécule d'ARN avec une haute spécificité. Ainsi, par des expériences de RIP-CHIP (RNA immunoprecipitation-chip) il a été montré que la protéine PPR-CRP1 de maïs se fixe *in vivo* à une région de 69 nucléotides dans la région 5'-UTR des transcrits *psaC* et *petA* (Schmitz-Linneweber et al., 2005). De même, il a été montré que les protéines CRR4 et CRR21 pourraient spécifier des sites d'édition du transcrit *ndhD* en se fixant sur une région de 35 nucléotides autour des sites d'édition (Okuda et al., 2006). Cependant, il faudra attendre d'autres confirmations expérimentales pour vérifier la nature spécifique de l'interaction entre une protéine PPR et un ARN. En effet, à l'heure actuelle, on ne peut pas totalement exclure que d'autres facteurs soient impliqués et nécessaires dans ce type d'interactions. De même, nous ignorons beaucoup, voir tout, du mécanisme moléculaire sous-jacent à ces interactions.



D'après Andrès et al., 2006

Figure n°25: Organisation de la famille PPR en fonction des motifs présents dans les protéines

La sous-famille de protéines PPR-P est présente chez tous les eucaryotes, en revanche la sous-famille de PPR-PLS est spécifique des plantes terrestres. Quand il est connu, le nombre de protéines PPR présentes chez les différentes espèces est indiqué.

### **I.3 Distribution phylogénétique des protéines PPR et caractérisation de différents motifs**

La recherche du motif PPR dans de nombreux protéomes a révélé que ces protéines sont absentes du règne procaryote, en revanche on retrouve des PPR chez tous les eucaryotes mais avec une extraordinaire diversité en nombre entre les plantes et les autres organismes (Lurin et al., 2004). Par exemple, alors que dans le génome d'Arabidopsis et du riz on trouve plusieurs centaines de protéines PPR, il n'y en a que 6 chez l'Homme. Une étude récente a montré que cette expansion de la famille PPR chez les plantes supérieures s'est principalement effectuée par des événements massifs de rétrotransposition qui ont eu lieu principalement avant la séparation des monocotylédones et des dicotylédones (O'Toole et al., 2008).

Une étude bioinformatique des motifs PPR a révélé l'existence de motifs PPR « dérivés » de taille un peu différente du motif de référence de 35 acides aminés (appelé P pour « pure »). Ces motifs dérivés du motif PPR ont été appelés L pour « Long » et S pour « Short ». Cette distinction a permis de mettre en évidence l'existence de deux sous-familles de protéines PPR : les PPR-P (qui ne comportent que des motifs de type « Pure ») ; et les PPR-PLS et qui portent également des motifs L et S (figure n°25) (Lurin et al., 2004). De façon intéressante, il apparaît que les protéines PPR-P sont présentes chez tous les eucaryotes, alors que les protéines PPR-PLS sont spécifiques des plantes. Chez Arabidopsis sur les 442 protéines PPR décrites, 242 sont de types PPR-P, tandis que les autres sont de types PPR-PLS (Lurin et al., 2004).

La sous-famille de protéines PPR-PLS est elle-même divisée en sous-groupes suivant la nature de la région C-terminale. En effet, ces protéines peuvent comporter un motif additionnel E (PPR-PLS de type « E », 191 membres chez Arabidopsis), auquel peut s'ajouter un motif E+ (PPR-PLS de type « E+ », 145 membres chez Arabidopsis), auquel peut encore éventuellement s'ajouter un domaine DYW (PPR-PLS de type « DYW », 87 membres chez Arabidopsis) (Lurin et al., 2004) (figure n°25). Si les motifs E et E+ sont hautement dégénérés et parfois difficiles à identifier, le domaine DYW (~100 acides aminés) est très conservé et composé notamment de résidus histidines, cystéines et d'un triplet final aspartate (D)-tyrosine (Y) tryptophane (W) quasiment invariants. Il a été proposé que ce domaine pourrait porter une activité catalytique (Aubourg et al., 2000; Lurin et al., 2004).

### **I.4 Quelques données sur l'expression des protéines PPR chez Arabidopsis**

Une des raisons de la découverte tardive de la famille de protéines PPR chez Arabidopsis est sans doute lié au fait qu'il existe une relativement faible quantité d'EST correspondant à ces gènes dans les banques de données suggérant qu'ils sont peu ou non exprimés dans cette plante.

L'analyse des données de puces CATMA (Complete Arabidopsis Transcriptome MicroArray) a révélé qu'en réalité une vaste majorité des gènes PPR d'Arabidopsis sont exprimés dans plusieurs tissus de la plante, en particulier les feuilles et les fleurs ; et a permis de confirmer que l'expression de ces gènes est extrêmement faible, notamment pour la sous-famille de protéines de type PPR-PLS (Lurin et al., 2004).

En ce sens, une étude récente a montré que les transcrits codants pour des protéines PPR ont une demi-vie bien plus courte que la plupart des autres transcrits chez Arabidopsis (Narsai et al., 2007).

Table n°2: Table récapitulative des différents mutants de protéines PPR présentés en introduction

gène	organisme	processus moléculaires	cible	type	localisation	références
<i>Rfk1</i>	<i>Raphanus sativus</i>	CMS	<i>orf125</i>	P	mt	Koizuka et al., 2003
<i>Rfo</i>	<i>Raphanus sativus</i>	CMS	<i>orf138</i>	P	mt	Brown et al., 2003; Desloire et al., 2003
<i>Rf1a-Rf1b</i>	<i>Oryza sativa</i>	CMS	<i>atp6-orf79</i>	P	mt	Kazama et Toriyama, 2003 ; Akagi et al., 2004 Komori et al., 2004 ; Wang et al., 2006
<i>CRR2</i>	<i>Arabidopsis thaliana</i>	Clivage	<i>rps7-ndhB</i>	DYW	cp	Hashimoto et al., 2003
<i>CRR4</i>	<i>Arabidopsis thaliana</i>	Edition	<i>ndhD</i>	E/E+	cp	Kotera et al., 2005
<i>CRR21</i>	<i>Arabidopsis thaliana</i>	Edition	<i>ndhD</i>	E/E+	cp	Okuda et al., 2006
<i>PGR3</i>	<i>Arabidopsis thaliana</i>	Clivage	<i>petL-petG-psaJ</i>	P	cp	Yamazaki et al., 2004
<i>HCF152</i>	<i>Arabidopsis thaliana</i>	Clivage	<i>psbT-psbH-petB-petD</i>	P	cp	Meterhoff et al., 2003
<i>MCA1</i>	<i>Chlamydomonas reinhardtii</i>	Stabilité	<i>petA</i>	P	cp	Lown et al., 2001
<i>CRP1</i>	<i>Zea mays</i>	Traduction	<i>petA, psaC</i>	P	cp	Fisk et al., 1999; Schmitz-Linneweber et al., 2005
<i>PPR4</i>	<i>Zea mays</i>	Epissage	<i>rps12</i>	P	cp	Schmitz-Linneweber et al., 2006
<i>hLRP130</i>	<i>Zea mays</i>	Stabilité	<i>coxI, coxIII</i>	P	mt	Mili et Pinol-Roma, 2003
<i>Cya5</i>	<i>Neurospora crassa</i>	Traduction	<i>cox1</i>	P	mt	Cofin et al., 1997
<i>Pet309</i>	<i>Saccharomyces cerevisiae</i>	Traduction	<i>cox1</i>	P	mt	Manthey et McEwen, 1995 ; Manthey et al., 1998
<i>Aep3</i>	<i>Saccharomyces cerevisiae</i>	Clivage	<i>atp6-atp8</i>	P	mt	Ellis et al., 2004
<i>BSF</i>	<i>Drosophila melanogaster</i>	Stabilité	<i>bicoid</i>	P	-	Mancebo et al., 2001
<i>P63</i>	<i>Triticum aestivum</i>	Transcription	<i>coxII</i>	P	mt	Ikeda et Gray, 1999
<i>OTP43</i>	<i>Arabidopsis thaliana</i>	Epissage	<i>nad1</i>	P	mt	Falcon de Longevialle et al., 2007
<i>CLB19</i>	<i>Arabidopsis thaliana</i>	Edition	<i>rpoA, clpP</i>	E/E+	cp	Chateigner-Boutin et al., soumis
<i>PPRlagladg</i>	<i>Arabidopsis thaliana</i>	Epissage	<i>ycf3</i>	P	cp	Falcon de Longevialle et al., en préparation
<i>FLV</i>	<i>Arabidopsis thaliana</i>	Edition	<i>rpoC1</i>	DYW	cp	Chateigner-Boutin et al., résultats non publiés
<i>PPR336</i>	<i>Arabidopsis thaliana</i>	Traduction	-	P	mt	Uytewaal et al., 2008
<i>pTAC2</i>	<i>Arabidopsis thaliana</i>	Transcription	-	P	cp	Pfalz et al., 2005
<i>GRP23</i>	<i>Arabidopsis thaliana</i>	Transcription	-	P	-	Ding et al., 2006
<i>Rfk1</i>	<i>Petunia parodii</i>	CMS	<i>pcf</i>	P	mt	Bentolila et al., 2002
<i>rfl</i>	<i>Shorgum bicolor</i>	CMS	-	E/E+	mt	Klein et al., 2005
<i>PPR2</i>	<i>Zea mays</i>	Traduction	-	P	cp	Williams et Barkan, 2003



## **II Rôle des protéines PPR**

Dès 2000, bien que leur fonction ait été encore peu connue, un certain nombre d'éléments permettaient de penser que les protéines PPR étaient impliquées dans un certain nombre de processus de régulation post-transcriptionnelle de l'expression des gènes dans les organites. Depuis, de plus en plus de données sur des mutants affectés dans les gènes PPR de différents organismes ont été publiées, confirmant l'idée que les protéines PPR participent au métabolisme des ARN à différents niveaux tels que : la transcription, la traduction, la maturation, la dégradation, l'épissage, la stabilité, la dégradation ou encore l'édition (table n°2).

### **II.1 PPR et initiation de la transcription des ARNm**

La protéine mitochondriale PPR-p63 du blé est le premier exemple de protéine PPR décrite pour participer au mécanisme de transcription des ARN (Ikeda and Gray, 1999). En effet, cette protéine a été purifiée au sein d'un extrait protéique mitochondrial capable d'une activité de transcription *in vitro*, dans lequel elle apparaît essentielle à l'activation de la transcription du transcrit *coxII* et se fixe sur la région promotrice du gène correspondant (Ikeda and Gray, 1999). Par ailleurs, une protéine très similaire chez *Arabidopsis*, la protéine pTAC2 a été identifiée au sein d'un extrait protéique permettant la transcription de gènes chloroplastiques (Pfalz et al., 2006). Enfin, il apparaît que la protéine PPR-GRP23 d'*Arabidopsis*, impliquée dans le développement précoce de l'embryon, pourrait interagir avec la sous-unité III de l'ARN polymérase de type II dans le noyau (Ding et al., 2006). Ces trois protéines PPR de plantes sont de type PPR-P. Ainsi, pris ensemble, ces éléments suggèrent que certaines protéines PPR, en se fixant sur des régions de l'ADN impliquées dans la régulation de la transcription, pourraient intervenir dans ce processus.

Par ailleurs, il a été montré que les ARN polymérases mitochondriales de mammifères contiennent des motifs PPR (Shadel, 2004). De plus, chez la levure, des protéines PPR associées à l'ARN polymérase mitochondriale, se lient aux chaînes d'ARN néosynthétisées afin de les stabiliser (Shadel, 2004). Ainsi, il existe un lien entre les ARN polymérases (et de ce fait la transcription) et les protéines PPR dont le rôle pourrait constituer en la coordination de la transcription des ARN et des étapes de maturation ultérieures jusqu'à la traduction.

### **II.2 PPR et maturation des ARNm**

L'expression des gènes, en particulier dans les organites, peut être contrôlée par des mécanismes de régulations post-transcriptionnelles visant à maintenir la stabilité des ARNm ou au contraire conduire à leur dégradation. Plusieurs protéines PPR ont été décrites pour être impliquées dans ces phénomènes. Par exemple, chez la drosophile, la protéine PPR-BSF (pour Bicoïd mRNA Stability Factor) se fixe à la région 3'UTR du transcrit *bicoïd* afin de le stabiliser (Mancebo et al., 2001). L'homologue humain de cette protéine, la protéine PPR-LRP130 fixe des ARNm nucléaires et mitochondriaux (Mili and Pinol-Roma, 2003). Des mutations dans le gène codant pour cette protéine conduisent à une maladie génétique : la LSFC (pour Leigh syndrome French Canadian variant) (Mootha et al., 2003). Chez les patients atteints de cette maladie, une diminution de la quantité des sous-unités COXI et COXIII de la chaîne respiratoire et des transcrits mitochondriaux correspondants a été caractérisée (Xu et al., 2004). De même, chez l'algue verte unicellulaire, *Chlamydomonas*

*reinhardtii*, il a été montré que la protéine PPR-MCA1 est essentielle à l'accumulation correcte du transcrit *petA* (Lown et al., 2001). Il a été proposé que toutes ces protéines PPR de type PPR-P sont directement impliquées dans le maintien de la stabilité de ces transcrits, bien qu'un défaut dans l'initiation de la traduction conduirait sans doute aux mêmes observations moléculaires.

La maturation des ARNm peut également consister en un ou plusieurs clivages d'ARN précurseurs polycistroniques, et là encore, nous disposons de plusieurs exemples de protéines PPR impliquées dans ce processus. Ainsi, chez la levure *Saccharomyces cerevisiae*, un mutant n'exprimant plus la protéine Aep3 présente une réduction dramatique de la quantité de transcrits *atp6* et *atp8* associée à un défaut de clivage de l'ARN polycistronique précurseur correspondant (Ellis et al., 2004). Une observation similaire a été faite pour plusieurs mutants de trois protéines PPR chloroplastiques chez *Arabidopsis thaliana*. En effet, la protéine PPR-CRR2 (pour chlororespiratory reduction) participe au clivage du transcrit *rps7-ndhB* (Hashimoto et al., 2003). De même, les protéines PPR PGR3 (pour proton gradient regulation 3) et HCF152 (pour high chlorophyll fluorescence) sont essentielles à la maturation des transcrits polycistronique *petL-petG-psaJ* et *psbT-psbH-petB-petD*, respectivement (Meierhoff et al., 2003; Yamazaki et al., 2004). De façon intéressante, toutes ces protéines (à l'exception de la protéine CRR2) appartiennent à la sous-famille de protéines PPR-P.

Une autre étape de maturation des ARNm, pour laquelle le rôle des protéines PPR a été plus récemment mis en évidence, est l'épissage des introns. En effet, chez *Arabidopsis*, l'étude de mutants de la protéine PPR OTP43 (protéine PPR-P) a révélé l'absence totale de transcrit mitochondrial *nad1* mature, dont le gène correspondant est morcelé en 5 exons répartis sur trois transcrits indépendants (de Longevialle et al., 2007). De ce fait, l'assemblage du transcrit *nad1* mature nécessite 3 épissages en « trans » et un en « cis » et il a été démontré que la protéine PPR OTP43 participe au trans-épissage réunissant les exons 1 et 2 (de Longevialle et al., 2007). Par ailleurs, également chez *Arabidopsis*, au laboratoire de Perth, l'étude du mutant d'un gène codant pour une PPR atypique, la protéine PPR-laglidag a été réalisée. Cette protéine est constituée de 6 motifs PPR (4P et 2S) et de deux motifs Laglidag caractéristiques des endonucléases de « homing » trouvées chez les intéines, les introns mobiles de groupe I et II et les introns archéens (Stoddard, 2005). L'analyse du phénotype moléculaire de ce mutant a révélé l'existence d'un défaut de « cis » épissage de l'intron 2 du transcrit chloroplastique *ycf3* (Falcon de Longevialle et al., sous presse). Enfin, chez le maïs, la protéine PPR4 participe à l'épissage en « trans » de l'intron 1 du transcrit chloroplastique *rps12* (Schmitz-Linneweber et al., 2006). A nouveau, toutes ces protéines PPR appartiennent à la sous-famille de PPR-P.

### **II.3 PPR et traduction**

La dernière étape de maturation d'un ARNm est l'initiation de la traduction, et là encore les protéines PPR semblent jouer un rôle.

En effet, chez la levure et chez *Neurospora crassa*, les mutants *pet309* (Manthey and McEwen, 1995; Manthey et al., 1998) et *cya-5* (Coffin et al., 1997) présentent un défaut de maintien de la stabilité du transcrit mitochondrial *cox1* et de sa traduction. Chez la levure, la protéine PET309 se fixe à la région 5'UTR du transcrit *cox1* et fait partie d'un complexe de la membrane mitochondriale interne dans lequel se trouve des facteurs de traduction (Naithani et al., 2003; Krause et al., 2004). De plus, les protéines PET309 de levure et de mammifères peuvent s'associer à la petite sous-unité mitochondriale du ribosome (Gavin et al., 2002; Koc

and Spremulli, 2003). Enfin, il a également été souligné que la protéine PET309 de levure peut s'associer à la protéine Nam1, qui se fixe elle-même sur l'ARN polymérase (Naithani et al., 2003). Pris ensemble, tous ces éléments suggèrent que la protéine PPR PET309 pourrait participer à toutes les étapes de maturation du transcrit *cox1* depuis sa transcription jusqu'à sa traduction et l'assemblage du complexe correspondant dans la membrane interne mitochondriale (Krause et al., 2004; Shadel, 2004).

Par ailleurs, chez le maïs, l'étude d'un mutant d'un gène codant pour la protéine PPR2 a révélé un défaut global de traduction dans les chloroplastes (Williams and Barkan, 2003). De plus, chez cette même espèce, il a été montré que la protéine PPR CRP1 est essentielle à la traduction des transcrits *petA* et *psaC*, et de ce fait au fonctionnement du cytochrome *b6f* et du photosystème I de la chaîne photosynthétique (Fisk et al., 1999). La protéine CRP1 peut fixer la région 5'UTR de ces transcrits (Schmitz-Linneweber et al., 2005). Enfin, tout récemment, la protéine PPR336 d'*Arabidopsis* a été isolée au sein d'une fraction polysomale mitochondriale, et cette protéine peut interagir avec les protéines de ribosomes mitochondriaux (Uyttewaal et al., 2008). Ainsi ces éléments suggèrent que les protéines PPR ne sont pas en elle-même des composants de la machinerie de traduction mais plutôt des facteurs permettant de recruter de cette machinerie afin d'initier la traduction.

#### **II.4 Rôle de protéines PPR dans des mécanismes de régulation de l'expression des gènes propres aux plantes**

##### *Cas des PPR impliquées dans la stérilité mâle cytoplasmique (CMS) et la restauration de la fertilité*

La stérilité mâle cytoplasmique (CMS) est un caractère à hérédité maternelle qui conduit à une perturbation de la production des grains de pollen et transforme ainsi des plantes hermaphrodites en plantes femelles (Budar et al., 2003). Dans les populations naturelles, ce processus permet de réguler le nombre de plantes mâle-stériles et femelle-stériles et favorise les croisements entre individus. Chez de nombreuses espèces, la stérilité mâle cytoplasmique est induite par l'expression d'une protéine mitochondriale hydrophobique (nommée ORF CMS) et la fertilité peut être restaurée sous l'action d'un facteur nucléaire (nommé restaurateur de fertilité ou gène *Rf*). Les couples ORF CMS/ gène *Rf* sont propres à chaque espèce. Cependant, au cours des dernières années, à l'exception du gène *Rf2* du maïs codant pour une aldéhyde déshydrogénase (Cui et al., 1996), il s'est avéré que tous les gènes *Rf* identifiés codent pour des protéines PPR.

C'est chez le pétunia que le premier gène PPR restaurateur de fertilité a été isolé (Bentolila et al., 2002). L'expression de la protéine PPR correspondante conduit à diminution de la quantité de protéine mitochondriale PCF associée à une réduction de la quantité du transcrit correspondant ainsi qu'à des défauts de maturation. Ainsi il apparaît que la protéine PPR-*Rf* régule l'expression de l'ORF CMS *pcf* à un niveau post transcriptionnel. Une observation similaire a été faite pour deux gènes *PPR* de radis. En effet, les gènes *PPR Rfo* et *Rfk1* réduisent l'expression des ORF CMS mitochondriales 138 et 125 (Brown et al., 2003; Desloire et al., 2003; Koizuka et al., 2003). Chez le riz, la CMS est induite par l'expression de l'ORF mitochondriale 79 localisée juste en aval d'une copie du gène *atp6* (Wang et al., 2006). La quantité de transcrits matures *atp6* et *orf79* est extrêmement réduite chez les plantes dont la fertilité mâle est restaurée. Quatre équipes de recherche ont montré l'implication du locus *Rf-1* dans la restauration de la CMS chez cet espèce (Kazama and Toriyama, 2003; Akagi et

al., 2004; Komori et al., 2004; Wang et al., 2006). Ce locus contient deux gènes codant pour les protéines PPR Rf-1a et Rf-1b. Il apparaît que la protéine Rf-1a participe au clivage de l'ARN précurseur *atp6-orf79*, alors que la protéine Rf-1b est impliquée dans sa dégradation. De plus, l'expression du gène *Rf-1a* est épistatique sur celle du gène *Rf-1b*. Enfin, chez le mil (*Sorghum bicolor*), il apparaît que le gène codant pour la PPR Rf1 est également impliqué dans la restauration de la stérilité mâle cytoplasmique (Klein et al., 2005).

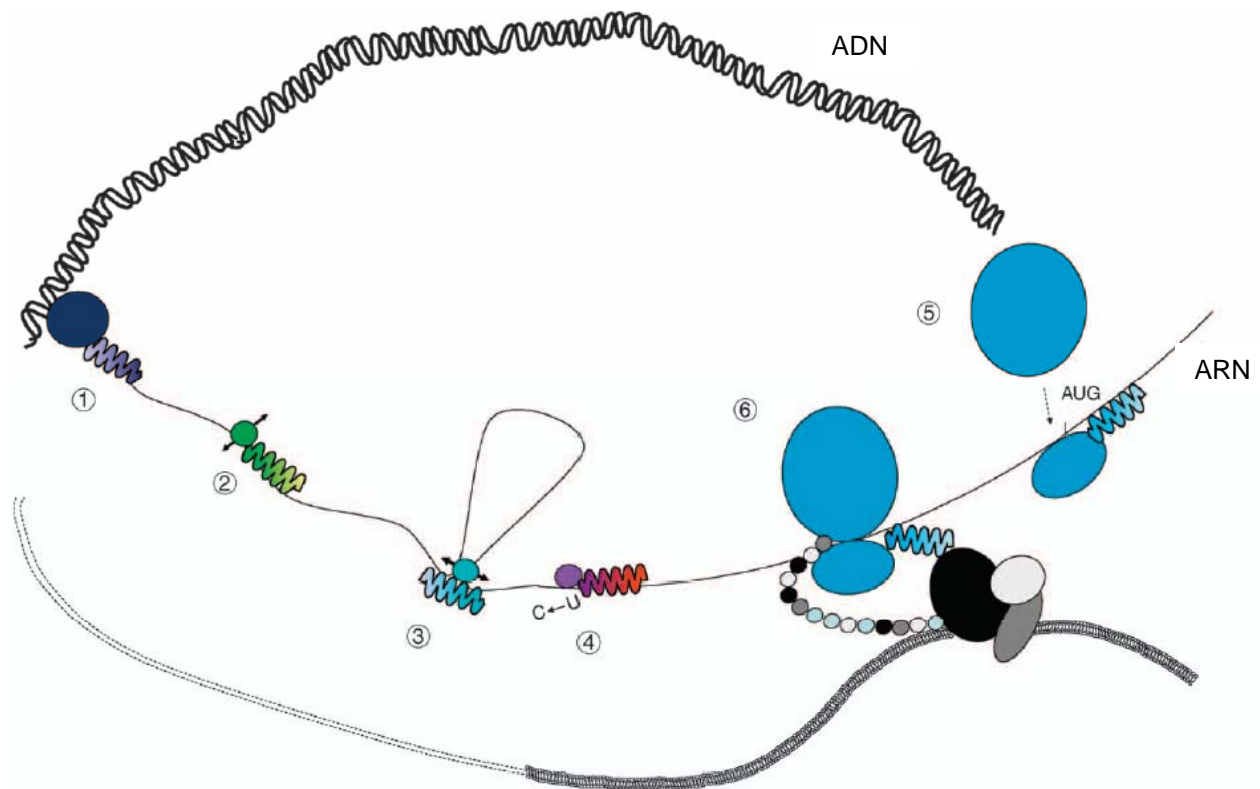
De façon intéressante, à l'exception de la protéine PPR de mil (de type PLS-E), toutes les autres protéines PPR impliquées dans la restauration de la CMS chez les plantes appartiennent à la sous-famille de PPR-P. Bien que le mécanisme moléculaire sous-jacent la restauration de la CMS ne soit pas encore finement caractérisé, il apparaît ces protéines PPR régulent des étapes de maturation (ou éventuellement d'initiation de la traduction) des ARNm des ORF CMS mitochondriales.

### *Cas des PPR impliquées dans le processus d'édition*

Nous l'avons évoqué précédemment, chez les plantes, au cours de la régulation post-transcriptionnelle des gènes des organites, l'édition C-en-U des ARN (et plus rarement U-en-C) est une étape essentielle à la synthèse de protéines fonctionnelles. Récemment, l'étude de différents mutants d'*Arabidopsis* a permis de mettre en évidence que plusieurs protéines PPR sont impliquées dans l'édition C-en-U des transcrits chloroplastiques.

La première à avoir été caractérisée est la protéine PPR CRR4, essentielle à l'édition d'un site unique dans le transcrit chloroplastique *ndhD* (Kotera et al., 2005). L'édition de ce site permet l'introduction d'un codon AUG initiateur de traduction et il a été montré que la protéine CRR4 se fixe sur le transcrit *ndhD* dans une région de quelques nucléotides en amont de ce site (Okuda et al., 2006). De la même façon, il a été montré que la protéine PPR CRR21 est également essentielle à l'édition d'un second site du transcrit *ndhD* (Okuda et al., 2007). Ces deux protéines PPR sont de type PPR E/E+ ; et des expériences d'échange de domaines entre ces protéines, ont montré que la spécificité d'interaction avec l'ARN cible était déterminée par les motifs PPR (Okuda et al., 2007). L'étude de mutants d'un gène codant pour une troisième protéine PPR de type E/E+, la protéine CLB19, a révélé que cette dernière était essentielle à l'édition de deux sites dans les transcrits chloroplastiques *rpoA* et *clpP* (Chateigner-Boutin et al., soumis). Chez ces lignées mutantes, des défauts de maturation dans d'autres transcrits chloroplastiques ont été mis en évidence. Il a été proposé que l'abolition de l'édition du site *rpoA* explique ce phénotype moléculaire pléiotropique, car le gène *rpoA* code pour la sous-unité alpha de la PEP (pour plastid-encoded RNA polymerase), enzyme responsable de la transcription de nombreux gènes chloroplastiques (Chateigner-Boutin et al., soumis). Ces trois protéines PPR de type E/E+ ne présentent aucun domaine permettant de proposer une activité catalytique, suggérant qu'elles assurent le rôle de facteur de spécificité dans la réaction d'édition et qu'elles peuvent éventuellement participer au recrutement du ou des facteur(s) catalytique(s) dont l'identité est encore inconnue (Kotera et al., 2005; Okuda et al., 2007; Chateigner-Boutin et al., soumis).

Par ailleurs, une analyse de QTL (quantitative locus trait) a permis d'isoler une région d'ADN, contenant deux gènes PPR de type DYW, qui est essentielle à l'édition dans la mitochondrie chez *Arabidopsis thaliana* (Bentolila et al., 2005). Et récemment, au laboratoire de Perth, l'étude d'un mutant dans un gène codant une protéine PPR chloroplastique de type DYW, la protéine FLV, a révélé un défaut d'édition dans le transcrit *rpoC1* (Chateigner-Boutin et al., résultats non publiés).



D'après Andrés et al., 2006

Figure n°26: Les différentes fonctions des protéines PPR dans le contrôle de l'expression des gènes. Des protéines PPR sont impliquées à chaque étape de l'expression des gènes entre la transcription et la traduction. 1: Des protéines PPR ont été copurifiées avec l'ARN polymérase, 2: Des protéines PPR sont impliquées dans le clivage des ARN, 3: Des protéines PPR sont impliquées dans l'épissage des introns, 4: Des protéines PPR sont nécessaires à l'édition des ARNm, 5 et 6: Des protéines PPR sont impliquées dans l'initiation et l'élongation de la traduction.

Ainsi à l'heure actuelle, plusieurs protéines PPR, toutes de type PLS, sont impliquées dans l'édition C-en-U des transcrits chloroplastiques (et peut-être mitochondriaux) chez *Arabidopsis*. De façon intéressante, la sous-famille de protéines PPR de type PLS, comme l'édition C-en-U des ARN des organites, sont propres aux plantes terrestres. Ceci pourrait expliquer l'explosion de cette sous famille de protéines chez les plantes supérieures (Lurin et al., 2004; Shikanai, 2006).

## **II.5 Principales conclusions quant aux rôles des protéines PPR (figure n°26)**

A la différence des gènes nucléaires, généralement régulés au niveau transcriptionnel, l'expression des gènes des organites est majoritairement contrôlée de façon post-transcriptionnelle. A l'heure actuelle, toutes les études indiquent que les protéines PPR sont importées dans les organites afin d'assurer cette régulation depuis la transcription jusqu'à la traduction (figure n°26). Toutes les étapes de maturation des ARN ne sont pas distinctes et indépendantes les unes des autres, et une même protéine PPR peut intervenir aux niveaux de plusieurs mécanismes. Ainsi il apparaît que les protéines PPR sont sans doute des facteurs de spécificité ciblant un ou un petit nombre de transcrits plutôt que les effecteurs de ces différents processus de maturation. Cependant, l'un des enjeux de l'étude de la famille PPR reste de caractériser le mécanisme moléculaire permettant d'une part, l'interaction (présumée directe) entre une protéine PPR et son ARN cible, et d'autre part le recrutement éventuel de facteurs enzymatiques intervenant dans ces différents processus de maturation.

Par ailleurs, les effets physiologiques liés à l'absence de chaque protéine PPR dépendent du ou des gènes encodés par son (ses) ARN cible(s) et de ce fait ne sont pas prévisibles. Ainsi chez les plantes, un défaut d'expression d'un gène codant une protéine PPR peut conduire à une large variété de phénotypes sévères tels que : la létalité embryonnaire, la stérilité mâle cytoplasmique, des défauts de croissance, des défauts de développement des graines ou encore des défauts de fonctionnement de la photosynthèse (Delannoy et al., 2007). Ces résultats indiquent que les membres de la famille PPR chez les plantes, malgré leur grand nombre et leur similarité, ne sont pas fonctionnellement redondants (Lurin et al., 2004).

Un autre enjeu de l'étude de la famille PPR est de comprendre pourquoi cette petite famille de quelques gènes chez la plupart des eucaryotes a « explosé » chez les plantes supérieures. En effet, hormis l'édition (et la sous famille de PPR de type PLS) et éventuellement la restauration de la stérilité mâle cytoplasmique, il ne semble pas que l'explosion de cette famille soit liée à l'apparition de nouvelles fonctions chez les plantes supérieures. Ainsi, la plupart des fonctions remplies par les protéines PPR sont assurées par d'autres protéines chez les autres organismes, et de ce fait la raison ayant favorisé cette expansion reste actuellement inconnue.

## **III Présentation du sujet de thèse : « Identification du facteur catalytique impliqué dans le processus d'édition des ARN des organites chez les plantes »**

Depuis sa découverte il y a presque 20 ans, les différents facteurs de la machinerie d'édition chez les plantes sont restés inconnus. Il est admis qu'un ou plusieurs facteurs de spécificité nucléaires se fixent à un ARN en amont de la base à modifier et catalysent la réaction d'édition directement ou *via* le recrutement d'un facteur catalytique.

Dès la découverte de la famille PPR chez *Arabidopsis*, en 2000, plusieurs éléments ont permis de suggérer qu'un certain nombre d'entre elles pouvaient être impliquées dans le processus d'édition. Tout d'abord, d'un point de vue quantitatif, il s'agit de la seule famille de protéines de liaison à l'ARN adressées aux organites et dont le nombre est du même ordre de grandeur que celui des sites édités. Mais aussi, d'un point de vue moléculaire, la modélisation de la structure d'un enchaînement de motifs PPR suggère que les protéines PPR pourraient reconnaître avec un haut niveau de spécificité une séquence ARN d'une dizaine de bases. D'un point de vue phylogénique, la sous-famille de protéines PPR de type PLS est spécifique des plantes terrestres, tout comme le phénomène d'édition. Enfin, notre attention s'est tout particulièrement portée sur la sous-famille de PPR à domaine « DYW ». En effet, d'un point de vue biochimique, l'édition nécessite une activité enzymatique, et le domaine DYW (d'environ 100 acides aminés) possède certains résidus conservés qui suggèrent une activité catalytique. **Nous avons rapidement proposé que ce domaine pouvait être le facteur enzymatique nécessaire à l'édition.** Par ailleurs, nous avons identifié une protéine particulière, **la protéine AtDYW1**, qui se distingue des autres protéines PPR par le fait qu'elle est uniquement composée du domaine DYW. De ce fait, elle **pouvait être un transfacteur recruté par les protéines PPR pour lesquelles aucun domaine catalytique n'a pu être prédit.** L'identification récente de plusieurs protéines PPR de type E/E+ ou de type DYW (appartenant donc à la sous-famille de protéine PPR de type PLS), qui sont impliquées dans le processus d'édition de transcrits chloroplastiques, supporte en partie ces hypothèses.

L'enjeu principal de mon travail de doctorat a consisté d'une part à évaluer si le domaine DYW pouvait effectivement être le facteur catalytique de la réaction d'édition chez les plantes et d'autre part à caractériser la fonction de la protéine AtDYW1 chez *Arabidopsis*. Pour se faire, j'ai mené une étude globale du domaine DYW visant à identifier son origine (recherche de motifs comparables dans les banques de données) et j'ai examiné sa distribution phylogénétique en particulier parmi les plantes primitives afin d'établir une corrélation éventuelle entre la présence du domaine DYW et le phénomène d'édition (Partie 1 des résultats). Par ailleurs, j'ai analysé l'expression et la localisation subcellulaire de la protéine AtDYW1 chez *Arabidopsis* (Partie 2 des résultats), et par l'analyse de plantes mutantes affectées dans l'expression de ce gène, j'ai étudié sa fonction en particulier son implication dans le processus d'édition dans les organites (Partie 3 des résultats). J'ai également recherché à caractériser l'activité enzymatique éventuelle de cette protéine (Partie 4 des résultats).

# **Présentation des résultats**



## **Partie 1:**

### **DES DONNEES EN FAVEUR DE L'IMPLICATION DU DOMAINE DYW DANS LE MECANISME D'EDITION DES ARN CHEZ LES PLANTES**

#### **I Une hypothèse à tester**

Nous avons souligné en introduction que les protéines PPR sont impliquées dans de nombreux mécanismes de régulations post-transcriptionnelles au sein des organites, en particulier l'édition C-en-U des ARN. De plus, via leur motif PPR, ces protéines pourraient être des facteurs capables de lier spécifiquement un ARN.

A l'heure actuelle, les protéines PPR décrites pour être impliquées dans l'édition appartiennent à la sous-famille PPR de type PLS, spécifiques des plantes terrestres. Au sein de cette sous-famille de PPR, nous savons qu'il existe, en plus des répétitions de motifs PPR, des domaines C-terminaux additionnels (E/E+ et DYW). Seul le domaine DYW des protéines PPR semble pouvoir « porter » une éventuelle activité catalytique.

Tous ces éléments pris ensemble, nous ont menés à poser l'hypothèse suivante : le domaine DYW des protéines PPR pouvait être le facteur enzymatique central impliqué dans le processus d'édition.

Dans ce premier chapitre de ma thèse, je présente comment, grâce à une étude du domaine DYW dans son ensemble, nous avons cherché à tester cette hypothèse.

Nous avons, dans premier temps, souligné l'existence de similarités entre le site actif de cytidines désaminases et le domaine DYW. Puis, par une étude de la distribution du domaine DYW dans le règne végétal, nous avons montré l'existence d'une corrélation phylogénétique stricte entre la présence du domaine DYW et le phénomène d'édition. Les résultats obtenus dans ces deux premières études étayaient fortement une implication du domaine DYW dans l'édition des ARN chez les plantes, suggérant même que celui-lui pourrait être le facteur enzymatique. Ces résultats ont fait l'objet d'une publication (ci-jointe, paragraphe II.3) qui détaille l'ensemble du travail (méthodes, résultats, et interprétations). Dans un souci de clarté et de cohérence de ce manuscrit, je détaille ici les étapes essentielles à la mise en place de ces approches, ainsi qu'un bref résumé des résultats présentés dans l'article faisant suite.

Enfin, je détaille les enjeux et résultats d'une étude complémentaire dans laquelle nous avons recherché la présence éventuelle du domaine DYW chez les dinoflagellées.

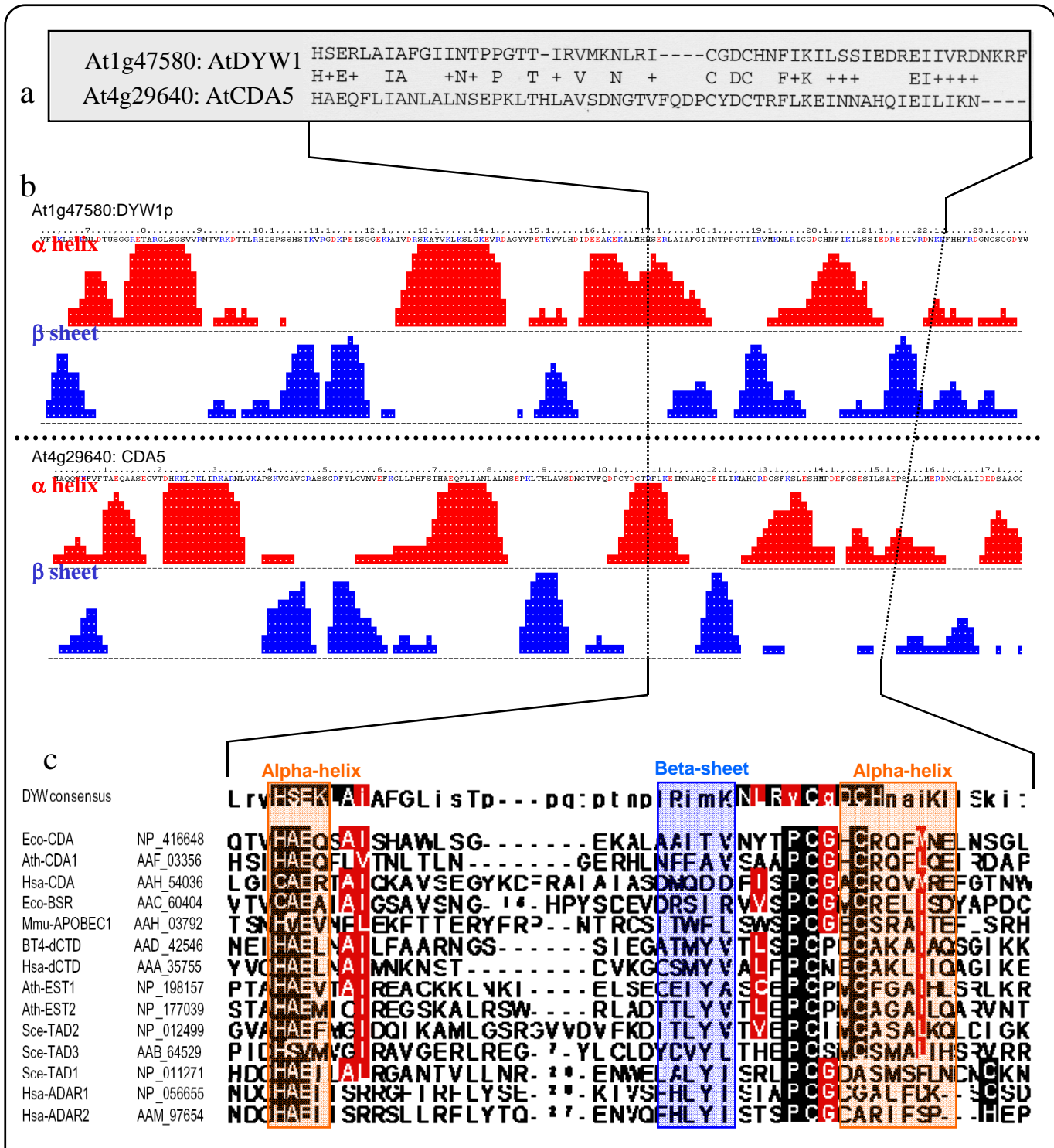


Figure n°27: Homologie en le domaine DYW et le site catalytique de désaminases

a) Alignement des séquences protéiques AtDYW1 et AtCDA5 (résultat de BLAST)

b) Prédictions structurales du logiciel PredictProt pour les protéines AtDYW1 et AtCDA5

c) Similarités entre les séquences primaires et secondaires du consensus du domaine DYW et de 14 cytidines désaminases (CDA) ou enzymes impliquées dans l'édition des ARN chez différentes espèces.

Abréviations: Eco, *Escherichia coli*, Ath, *Arabidopsis thaliana*, Hsa, *Homo sapiens*, BT4, bacteriophage T4, Mmu, *Mus musculus*, Sce, *Saccharomyces cerevisiae* (les numéros d'accessions gene bank sont indiqués)

## II Deux approches parallèles pour tester notre hypothèse

### II.1 Mise en évidence et caractérisation de similarités entre le site actif de cytidines désaminases et le domaine DYW

Dès sa caractérisation, la présence d'histidines et de cystéines conservées suggérait une potentielle activité catalytique pour le domaine DYW (Aubourg et al., 2000; Lurin et al., 2004). Afin d'obtenir plus d'informations quant à cette activité « potentielle », j'ai mené au début de ma thèse une série d'analyses *in silico*.

#### *II.1.1 Des conservations de séquences primaires entre le domaine DYW et les désaminases*

Un premier résultat a suscité mon attention ; l'alignement, par BLAST, de la séquence de la protéine AtDYW1 (At1g47580) avec la séquence d'une cytidine-désaminase d'*Arabidopsis thaliana* (At4g29640) (figure n° 27a). Si le degré de conservation est faible, ce résultat nous a paru néanmoins convainquant dans la mesure où les séquences s'alignant correspondent au site actif de cette cytidine-désaminase et à la partie C-terminale du motif DYW (partie la plus conservée).

J'ai ensuite montré que ce résultat était généralisable, en alignant la partie C-terminale du consensus du domaine DYW et le site actif de plusieurs désaminases, dont certaines, telle qu'APOBEC1 ou les protéines ADAR, sont impliquées dans des processus d'édition chez les mammifères (figure n° 27b).

#### *II.1.2 Le motif catalytique des CDA est comparable à un motif conservé du domaine DYW*

Les cytidines désaminases contiennent un motif caractéristique de leur site actif **C/HxEx<sub>n</sub>PCxxC** (PFAM 00383). Dans leur étude cristallographique, Betts *et al.* (Betts et al., 1994) montraient que les 2 cystéines et l'histidine conservées dans le site actif des cytidines désaminases permettent la fixation du zinc, et que le résidu glutamate est essentiel à la catalyse de la désamination.

Nous avons déterminé que les domaines DYW d'*Arabidopsis thaliana* contiennent dans leur partie C-terminale, un motif conservé **HxEx<sub>n</sub>CxxC**. A l'exception du résidu proline, la similarité entre le motif catalytique des cytidines désaminases et ce motif conservé du domaine DYW est grande. De plus, les 4 résidus décrits comme essentiels à la fixation du zinc et à la catalyse de la désamination pour les cytidines désaminases sont présents et quasiment invariants dans toutes les séquences de protéines à domaine DYW isolées à l'heure actuelle.

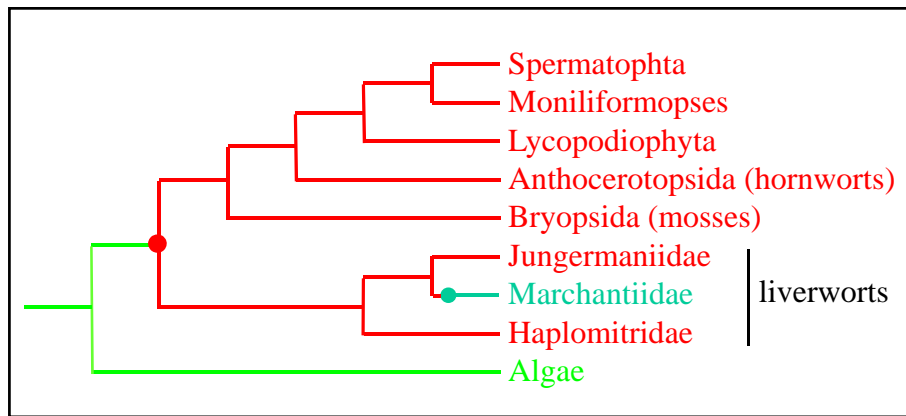


Figure n°28: Distribution phylogénique de l'édition chez les plantes vertes

En rouge: acquisition du processus d'édition par les plantes terrestres.

En vert-bleu: perte secondaire du processus d'édition par les Marchantiidae.

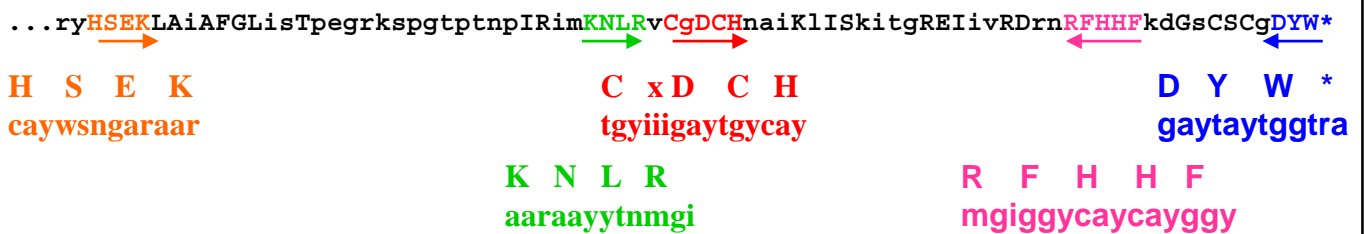


Figure n°29 : Représentation schématique des différentes amorces dégénérées dessinées pour amplifier le domaine DYW

En lettres capitales: consensus en acides aminés des amorces.

En lettres minuscules: séquence des amorces.

Leur localisation sur le consensus (partie C-terminale) du domaine DYW est indiquée.

### *II.1.3 Des conservations de structures secondaires entre le domaine DYW et les cytidines désaminases*

En utilisant le logiciel predict prot (Rost and Liu, 2003), j'ai pu mettre en évidence, qu'en plus d'une conservation au niveau de la séquence primaire, il existe de grandes similarités dans les prédictions de structures secondaires entre les protéines AtDYW1 (At1g47580) et CDA5 (At4g29640) (figure n°27c).

Là encore, j'ai pu généraliser ce résultat à l'ensemble des protéines à domaine DYW ainsi que plusieurs cytidines désaminases présentes chez Arabidopsis.

### *II.1.4 Bilan*

Cette approche a permis de mettre en évidence des similarités de séquences entre le consensus du domaine DYW et un certain nombre de désaminases (dont certaines catalysent des réactions d'édition chez les mammifères et les champignons). De plus, les résidus décrits comme essentiels à l'activité des cytidine-désaminases sont présents et extrêmement conservés dans le domaine DYW. Enfin, le domaine DYW et les cytidines désaminases d'Arabidopsis présentent une structure comparable. Ainsi, le domaine DYW pourrait avoir une activité catalytique proche de celle des cytidines désaminases.

## **II.2 Une étude de la distribution du domaine DYW dans le règne végétal**

Le phénomène d'édition C-en-U des transcrits des organites est décrit chez toutes les plantes terrestres à l'exception d'un sous-groupe d'hépatiques, les Marchantiidae. Ce phénomène est absent chez les algues vertes. Ainsi, il semble que l'édition des ARN soit apparue au moment de l'émergence des plantes terrestres, il y a plus de 450 millions d'années, puis aurait été secondairement perdu chez les hépatiques à thalles complexes (Marchantiidae) uniquement (figure n°28).

Nous avons cherché à connaître la distribution du domaine DYW dans le règne végétal, en particulier chez les hépatiques dont certains éditent leur ARN (les Jungermanniidae) et d'autres non (les Marchantiidae). J'ai débuté cette étude à l'URGV, puis je l'ai poursuivie, dans le cadre d'une collaboration, avec l'équipe de Volker Knoop (Université de Bonn, Allemagne) spécialisée dans l'étude de l'évolution de l'édition dans la lignée verte.

### *II.2.1 Une analyse bioinformatique*

Nous avons recherché parmi l'ensemble des génomes séquencés et des collections d'ESTs du NCBI, la séquence de toutes les protéines comportant un domaine DYW. Pour cela nous avons réalisé des analyses de TBLASTN à partir de la séquence consensus du domaine DYW.

Les résultats de cette étude sont détaillés dans l'article (paragraphe II.3). En résumé, cette analyse bioinformatique, m'a permis de montrer que **le domaine DYW est spécifique des plantes**. En effet, on ne trouve aucune séquence correspondant à ce domaine chez les algues, les animaux, les champignons et les bactéries, alors que plus d'un millier de séquences a pu

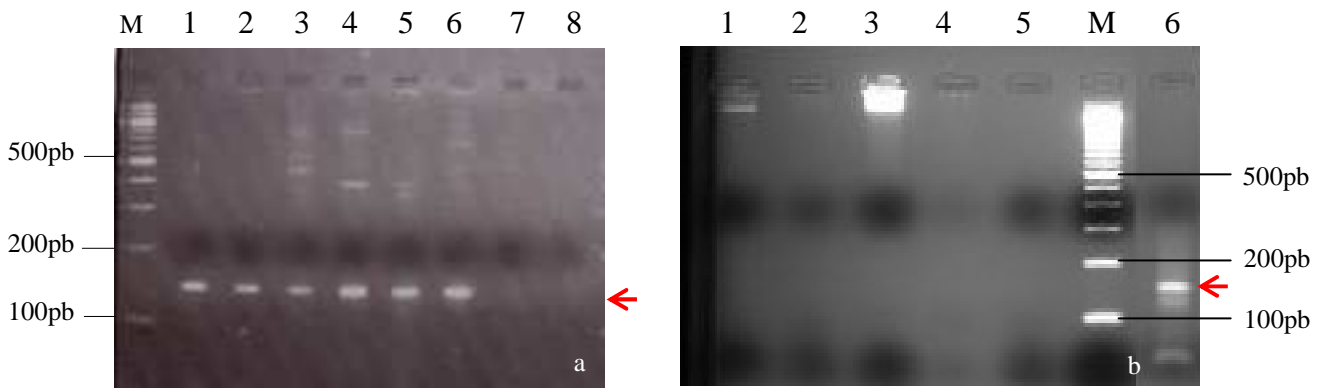


Figure n° 30 : Résultats des réactions PCR réalisée avec le couple d’amorces dégénérées CxDCH/DYW\* : mise au point sur des espèces témoins

a: témoins positifs

- Piste M : Marqueurs de taille 100pb BioLabs
- Piste 1: ADN génomique d’*Arabidopsis thaliana*
- Piste 2: ADN de plasmide (pAtDYW1)
- Piste 3: ADN génomique de pois
- Piste 4: ADN génomique de melon
- Piste 5: ADN génomique de tomate
- Piste 6: ADN génomique de colza
- Piste 7 et 8 : Contrôle négatif (eau)

b: témoins négatifs

- Piste 1: ADN génomique *Chlamydomonas reinardhii* (1)
- Piste 2: ADN génomique *Chlamydomonas reinardhii* (2)
- Piste 3: ADN génomique *Chlamydomonas reinardhii* (3)
- Piste 4: piste vide
- Piste 5: ADN génomique *Mus musculus*
- Piste M : Marqueurs de taille 100pb BioLabs
- Piste 6: Contrôle positif (ADN génomique d’*Arabidopsis thaliana*)

➔ Produit d’amplification d’intérêt (~150pb)

Table n° 3: Liste des espèces d’intérêts

Groupe	Espèce	puits
<b>Marchantiidae</b> (pas d’édition)	<i>Lunularia cruciata</i>	1
<b>Jungermanniidae</b> (édition)	<i>Fossombriona alskana</i>	2
	<i>Haplomitrium mnioides</i>	3
	<i>Metzgeria furcata</i>	4
	<i>Radula complanata</i>	5
	<i>Frullania tamarisci</i>	6
	<i>Scapania nemorea</i>	7
<b>Anthocerotopsida</b> (édition)	<i>Anthoceros agrestis</i>	8
<b>Lycopodopsida</b> (édition)	<i>Huperzia hilderbrandtii</i>	9
<b>Spermatophyta</b> (édition)	<i>Dicksonia antarctica</i>	10

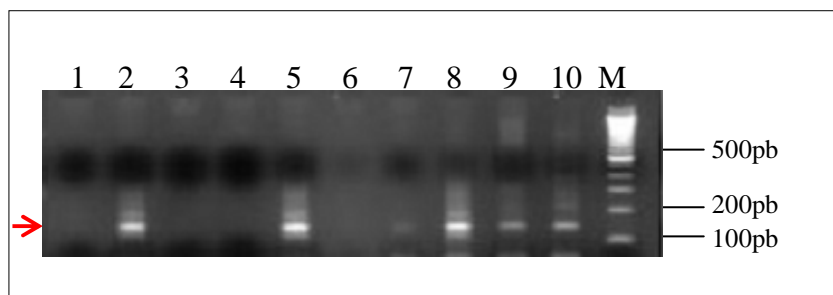


Figure n° 31 : Résultats des réactions PCR réalisée avec le couple d’amorces dégénérées CxDCH/DYW\* sur l’ADN génomique d’espèces d’intérêt (*Espèces listées dans la table n° ci-dessus*)

être identifié chez 135 espèces de plantes « supérieures ». De plus, l'alignement multiple de 1066 séquences ainsi isolées, m'a permis de souligner **l'extrême conservation du domaine DYW** parmi toutes ces espèces, suggérant une forte pression de sélection sur celui-ci.

### *II.2.2 Mise au point d'une approche par PCR à l'aide d'amorces dégénérées*

#### Génération d'amorces dégénérées

La partie C-terminale du domaine DYW présente une conservation telle que certains acides aminés sont invariants. Grâce à cette dernière caractéristique, plusieurs amorces dégénérées ont été dessinées (figure n° 29).

#### Test PCR préliminaires sur des espèces témoins

L'utilisation d'amorces dégénérées en PCR nécessite d'adapter les conditions de réaction. Nous avons établi, au cours de cette phase de mise au point, les conditions optimales de réaction PCR (ajout de magnésium dans le tampon de réaction, concentrations des amorces ...) ainsi qu'un programme de PCR « par palier » ou « TOUCHDOWN » (voir article, Matériel et Méthodes).

L'ensemble de ces mises au point a été réalisé sur une série d'extractions d'ADN génomique d'Angiospermes, connues pour posséder des gènes codant des protéines PPR de la sous classe DYW (témoins positifs). Les espèces utilisées sont : *Arabidopsis thaliana*, le pois, le melon, la tomate et le colza. Nous avons également utilisé une préparation d'ADN plasmidique contenant le gène *DYWI* d'*Arabidopsis thaliana* (pDYW1). Des extractions d'ADN génomique de l'algue unicellulaire *Chlamydomonas reinardhii*, ainsi que de *Mus musculus* (souris) ont été utilisées comme témoins négatifs ; puisque nous savons que le domaine DYW est absent du génome de ces deux espèces.

Toutes les combinaisons d'oligonucléotides dégénérés possibles ont été testées et il s'est rapidement avéré que la combinaison d'amorce CxDCH/DYW\* donnait une meilleure amplification sur les témoins positifs et aucune amplification sur les témoins négatifs (figure n°30a et b).

### *II.2.3 PCR à l'aide d'amorces dégénérées sur l'ADN génomique d'espèces d'intérêts*

#### Les premiers résultats

Dans un premier temps, j'ai travaillé sur une série d'extractions d'ADN génomique provenant de 10 espèces différentes parmi lesquelles des bryophytes, des lycopodiées et des spermatophytes (table n°3). J'ai réalisé un ensemble de PCR avec le couple d'amorces dégénérées CxDCH/DYW\*. J'ai obtenu des produits d'amplification de 150pb pour 6 espèces (figure n°31).

Les produits d'amplifications obtenus ont été extraits du gel, purifiés avant d'être clonés dans le plasmide pGEMT-easy. Ainsi, par séquençage, j'ai vérifié que ces amplifications correspondaient à des domaines DYW (résultats non présentés).

Cependant, à ce stade de notre approche, j'étais confrontée à des difficultés techniques : peu de reproductibilité des amplifications en PCR, détection difficile de produits d'amplifications pour certaines espèces... De ce fait, l'interprétation des résultats était complexe, et il était peu envisageable de réaliser une analyse plus systématique sur un grand nombre d'espèces afin d'étudier la distribution phylogénétique du domaine DYW dans la lignée verte.

### De nouvelles amorces dégénérées

Nous avons étudié les séquences de domaines DYW de différentes espèces de Bryopsida, Anthocerotopsida, Jungermanniidae et Haplomitriidae, obtenues par ces premières approches de PCR ou disponibles dans les bases de données.

Nous avons souligné la présence, en amont des cystéines conservées de l'amorce CxDCH, de résidus Leucine, Arginine et Valine quasiment invariants pour ces espèces d'intérêts. Ceci nous a permis de dessiner une extension à l'amorce CxDCH préexistante, ainsi devenue LRVCxDCH.

De la même façon, il existe un doublet invariant de cystéines en amont du triplet final DYW dans les domaines DYW de toutes les espèces de la lignée verte étudiée. Chez les mousses et les hépatiques, nous avons observé que ce doublet invariant de cystéines était majoritairement encadré par soit un résidu alanine et glycine, soit un résidu sérine et argine. Ainsi, nous avons proposé deux extensions à l'amorce DYW\* préexistante : CSCRDYW\* d'une part et CACGDYW\* d'autre part.

### Une « systématisation » de l'approche

J'ai testé ces nouvelles amorces sur l'ADNg de différentes espèces d'intérêt et il s'est avéré que les modifications apportées ont permis d'améliorer significativement l'efficacité des PCR. Ceci est particulièrement vrai lorsque l'on utilise la combinaison LRVCxDCH/ CSCRDYW\*.

Ainsi nous avons pu rechercher la présence du domaine DYW par PCR à l'aide de ce couple d'amorces dégénérées sur près de 40 espèces, en particulier des hépatiques dont certaines éditent leur ARN (les Jungermanniidae) et d'autres non (les Marchantiidae). Une amplification du domaine DYW a été obtenue pour toutes les espèces testées, excepté pour les espèces d'algues vertes et de Marchantiidae. Ces résultats sont détaillés dans l'article (Voir Article : table n°2 et figure n° 3A, paragraphe II-3).

Afin de vérifier que l'absence d'amplification n'est pas liée à un problème trivial de qualité ou concentration de la préparation d'ADNg, nous avons pris soin de tester les ADNg en réalisant des PCR avec des amorces spécifiques de l'ARNr 18s de chacune de ces espèces (Voir Article : figure n° 3B, paragraphe II-3).

En conclusion, cette étude nous a permis de démontrer l'existence d'**une corrélation phylogénétique stricte entre la présence du domaine DYW et le phénomène d'édition** dans la lignée verte.

### II.3 “A hypothesis on the identification of the editing enzyme in plant organelles”



## Hypothesis

## A hypothesis on the identification of the editing enzyme in plant organelles

Véronique Salone<sup>a,b</sup>, Mareike Rüdinger<sup>c</sup>, Monika Polsakiewicz<sup>c</sup>, Beate Hoffmann<sup>a</sup>, Milena Groth-Malonek<sup>c</sup>, Boris Szurek<sup>a</sup>, Ian Small<sup>a,b</sup>, Volker Knoop<sup>c</sup>, Claire Lurin<sup>a,\*</sup><sup>a</sup> URGV, 2 Rue Gaston Crémieux, F-91057 Evry Cedex, France<sup>b</sup> ARC Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley 6009, Australia<sup>c</sup> IZMB, Abt. Molekulare Evolution, Kirschallee 1, D-53115 Bonn, Germany

Received 31 May 2007; revised 20 July 2007; accepted 30 July 2007

Available online 10 August 2007

Edited by Micheal R. Sussman

**Abstract** RNA editing in plant organelles is an enigmatic process leading to conversion of cytidines into uridines. Editing specificity is determined by proteins; both those known so far are pentatricopeptide repeat (PPR) proteins. The enzyme catalysing RNA editing in plants is still totally unknown. We propose that the DYW domain found in many higher plant PPR proteins is the missing catalytic domain. This hypothesis is based on two compelling observations: (i) the DYW domain contains invariant residues that match the active site of cytidine deaminases; (ii) the phylogenetic distribution of the DYW domain is strictly correlated with RNA editing.

© 2007 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

**Keywords:** RNA editing; PPR protein; DYW domain; Plant evolution

## 1. Introduction

In plants, organellar transcripts are subject to RNA editing which is a specific post-transcriptional pyrimidine exchange (mainly C-to-U conversions, frequently accompanied by U-to-C conversions in hornworts and ferns). By restoring conserved amino acid codons, most editing events are obligatory steps in the biosynthesis of functional proteins [1,2]. RNA editing has been detected, both in chloroplast and mitochondrial transcripts, among all vascular plants (seed plants, ferns and fern allies) and also in the three classes of bryophytes (hornworts, liverworts and mosses) [3–5]. In contrast, RNA editing has not been reported in algae [6,7]. Plant organellar editing probably appeared with the emergence of land plants and nearly simultaneously in mitochondria and chloroplasts, suggesting a common origin for the editing machinery in both organelles [1–3]. RNA editing has been described in all groups of bryophytes with the exception of the subclass of complex thalloid liverworts (Marchantiidae) [8]. The presence of editing in all other land plant lineages, especially in the Jungermanniidae, the sister group of the Marchantiidae, suggests either two

independent gains or a secondary loss of the editing mechanism in the latter subclass [7].

Although discovered 15 years ago, RNA editing remains an enigmatic process. It is now accepted that sequence-specific RNA-binding proteins must be involved in editing in plants [1,9]. Recently, the *Arabidopsis thaliana* PPR proteins CRR4 and CRR21 were described as essential for editing single sites in *ndhD* and *ndhB* transcripts in plastids [10,11]. CRR4 binds specifically to sequences just upstream of the editing site [12]. PPR proteins are characterised by the presence of tandem arrays of a 35 amino acid (pentatricopeptide) motif [13]. The PPR family is subdivided in two major subfamilies: the P subfamily, whose members (~250 in the *Arabidopsis* genome) contain only canonical PPR motifs and the PLS subfamily, whose members (~200 in *Arabidopsis*) also contain non-canonical shorter and longer PPR-like motifs [14]. Canonical PPR motifs are found in the genomes of all eukaryotes that have a mitochondrial genome, but the non-canonical PPR-like motifs (PLS) seem to be plant-specific [14]. It has thus been proposed that PLS proteins could be involved in plant-specific post-transcriptional processes and especially in RNA editing [1,14]. Neither CRR4 or CRR21 (both PLS subfamily proteins) contain any domain likely to have catalytic activity [10,11], suggesting that they are only specificity factors. The enzyme(s) catalysing RNA editing in plant organelles is (are) totally unknown. In its usual direction, the reaction has been shown to proceed by deamination [15,16]. In mammals, the APOBEC editing enzyme, responsible for C-to-U editing of mammalian apolipoprotein B mRNA, is a zinc-dependent enzyme clearly related to cytidine deaminase [17,18]. Editing assays using plastidial plant extracts were shown to be zinc-dependent [19]. Cytidine deaminase-like enzymes have thus been popularly supposed to be involved in RNA editing in plants. However, to date, none of the nine putative cytidine deaminase proteins encoded by the *A. thaliana* genome appear to be involved in editing [20].

About 90 *Arabidopsis* proteins from the PLS subfamily have an additional and highly conserved C-terminal domain (c. 100 amino acids) named the DYW domain because of its characteristic terminating tripeptide [14,21]. This domain has not been found in any other protein or in any organism apart from land plants (embryophytes). The presence of invariant cysteine and histidine residues in the DYW motif prompted the proposal of a catalytic function for this domain [14,21]. Taken

\*Corresponding author. Fax: +33 1 60 87 45 10.  
E-mail address: lurin@evry.inra.fr (C. Lurin).

together, the RNA-binding properties of PPR proteins, their subcellular localisation, the specificity to land plants of the PLS subfamily, the putative enzymatic function of DYW domains and the description of two editing mutants lacking specific PLS proteins [10,11], lead us to the hypothesis that the DYW domain may provide the catalytic activity necessary for RNA editing. To examine this hypothesis, we compared the DYW consensus sequence to the active site of cytidine deaminases and found some sequence and structural similarities. Moreover, we performed a study of the phylogenetic distribution of the DYW domain in green plants, in particular in liverworts, and demonstrated a striking correlation between the presence of the DYW domain and RNA editing. These data provide strong arguments in favour of the role of the DYW domain in RNA editing in plant organelles.

**2. The DYW domain has similarities to the active site of cytidine deaminases**

Cytidine deaminases are zinc-dependent enzymes containing a typical motif (PFAM 00383) that corresponds to the active site of the protein. In particular, the degenerate motif C/HxEx<sub>r</sub>PCxxC is found in cytidine deaminases and mammalian editing enzymes [22]. When doing a BLASTP [23] analysis against the *Arabidopsis* proteome using AtCDA3 and AtCDA5 *Arabidopsis* cytidine deaminase sequences, we found a weak similarity with several DYW-containing PPR proteins (data not shown). A comparison of the DYW domain consensus and the cytidine deaminase zinc-binding domain was performed and the most highly conserved residues of the DYW domain were shown to align with the active site of cytidine deaminases (Fig. 1), both in terms of primary sequence and secondary structure prediction. In particular, the residues HxEx<sub>r</sub>CxxC are conserved in DYW proteins, although the proline preceding the first cysteine in this motif is not. The conserved histidine and cysteines were shown to bind the zinc atom at the active site of cytidine deaminases, whereas the conserved glutamate is required for catalysis of the deamination reaction [24]. As these four residues are

identical and invariant in all known DYW proteins, we propose a putative activity of the DYW domain in nucleotide deamination.

**3. The DYW domain is restricted to plant sequence databases**

TBLASTN [23] analyses were performed on EST collections and complete genome sequences available at the NCBI website using the DYW and PPR consensus sequences developed previously [14]. As previously shown [14], ESTs encoding PPR proteins were found in all large eukaryote data sets. In contrast, no animal, fungal or bacterial proteins showed a domain convincingly similar to the DYW domain, confirming its plant specificity. In plants, 1550 sequences encoding DYW domain-like polypeptides were found. Most of them were identified in seed plants (~125 species), but also in two mosses (*Ceratodon*, *Physcomitrella*), two ferns (*Adiantum*, *Ceratopteris*) and the lycophodiophyte *Selaginella* (see Table 1). Notably, no evidence for DYW-like domains was found in algae, despite the presence of almost 500 000 algal ESTs in the databases searched, among which about one hundred were shown to encode PPR proteins. In addition, in the three unicellular algae *Chlamydomonas reinhardtii*, *Ostreococcus tauri* and *Thalassiosira pseudonana* whose genomes have been completely or almost completely sequenced, no indication of DYW domains was found whereas 8, 17 and 27 PPR genes were found in these genomes, respectively. No DYW domain sequences were found in ESTs from *Marchantia* either, whereas 20 PPR-encoding ESTs were identified from this genus (Table 1). Although only 33 000 ESTs are available from *Marchantia*, all other land plant groups with an equivalent number of EST and/or PPR sequences have significant matches to the DYW consensus.

**4. The DYW domain is highly correlated with RNA editing**

A multiple alignment of the 1066 DYW domain sequences available at the start of this work from 97 species (including



Fig. 1. Similarity of the DYW domain to the active site of cytidine deaminases and editing enzymes. Eighty-seven DYW domain sequences from *Arabidopsis* were aligned using CLUSTALW to derive the consensus sequence shown on top (C-terminal half of the domain is shown). Capital letters in the consensus sequence indicate the most highly conserved amino acids (found in more than 60% of the *Arabidopsis* DYW domains). The DYW consensus is aligned with a multiple alignment of 14 typical cytidine deaminases (CDAs) or RNA editing enzymes. Species abbreviations are Eco, *Escherichia coli*; Ath, *Arabidopsis thaliana*; Hsa, *Homo sapiens*; BT4, bacteriophage T4; Mmu, *Mus musculus*; Sce, *Saccharomyces cerevisiae*. Genbank accession numbers are indicated. The amino acids involved in the active site (and the corresponding amino acids in the DYW consensus) are shaded in black. The DYW and CDA alignments were also used independently for secondary structure predictions using the PredictProtein service (<http://cubic.bioc.columbia.edu/predictprotein/>). CDA structure predictions agreed with the known crystal structures for CDAs. The  $\alpha\beta$  secondary structure, the three quasi-invariant amino acids bound to the active site zinc ion and the catalytic glutamate characteristic of the CDAs are all present in the DYW domain sequences.

Table 1  
Search for DYW and PPR domain sequences in plant ESTs

Taxonomic group	# of ESTs	PPR domains		DYW domains	
		# of sequences	# of species	# of sequences	# of species
<i>Spermatophyta</i>					
Coniferophyta	638911	471	10	184	9
Cycadophyta	17240	30	2	11	3
Ginkgophyta	6248	13	1	10	1
Gnetophyta	14363	21	3	2	2
<i>Magnoliophyta</i>					
Basal Magnoliophyta	19147	57	2	15	2
Eudicotyledons	6221190	4416	127	692	85
Liliopsida	4474911	2305	37	596	26
Magnoliids	30036	86	4	12	3
<i>Moniliformopses (ferns)</i>					
<i>Lycopodiophyta</i>	15616	388	2	18	2
<i>Lycopodiophyta</i>	5873	25	2	1	1
<i>Bryophyta</i>					
Bryopsida (mosses)	205805	64	3	10	2
Marchantiopsida (liverworts)	33692	20	1	0	0
<i>Algae</i>					
<i>Algae</i>	473244	104	13	0	0
Total		8000	207	1550	135

The non-redundant EST database at NCBI was searched by TBLASTN using the Arabidopsis DYW domain consensus sequences and a typical triplet PPR coding sequence. '# sequences' indicates the number of sequences with significant matches in each phylogenetic group (containing '# species'). '# of ESTs' indicates the total number of non-redundant ESTs corresponding to each phylogenetic group in the database.

sequences from *Physcomitrella*, *Selaginella* and *Ceratopteris*) was performed (Fig. 2 and data not shown). Interestingly, the DYW domain showed high sequence conservation among even distantly related species, suggesting a strong selection

pressure linked to an essential function in these plants. The presence of invariant amino acids allowed us to design different degenerate oligonucleotides (Fig. 2) to search by PCR amplification for the presence of DYW coding sequences in

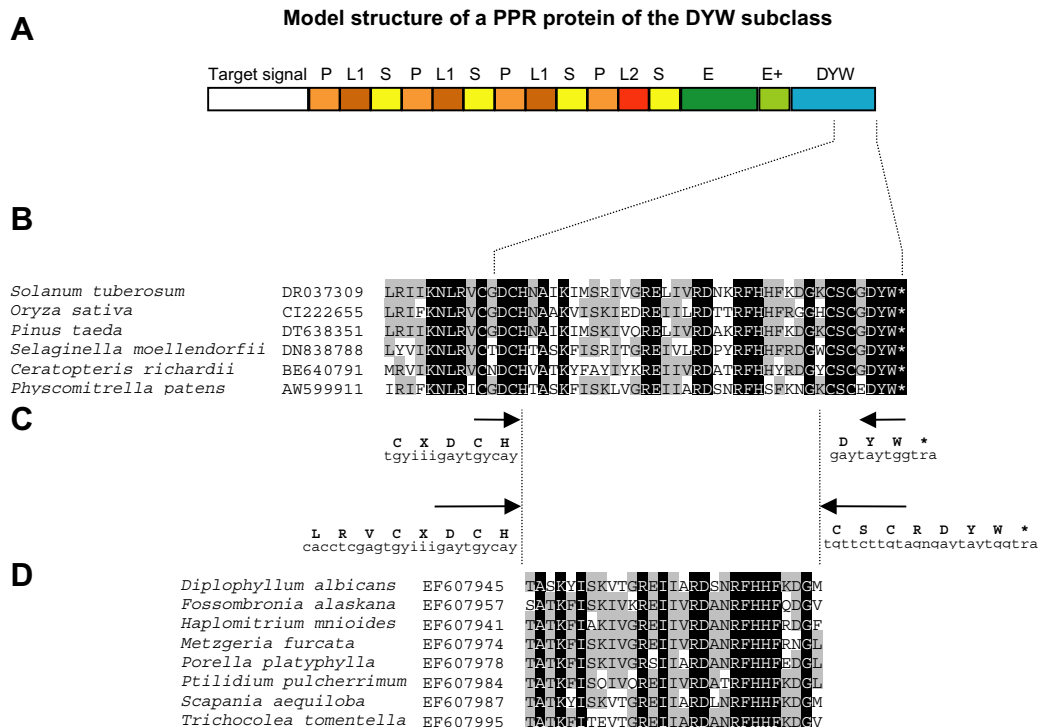


Fig. 2. Primer design for amplifying sequences encoding DYW domains. (A) Model structure of a PPR protein from the DYW subclass. (B) CLUSTALW multiple alignment of a selection of typical DYW domain sequences from amongst 1066 land plant sequences. (C) First and second generation degenerate oligonucleotide sequences used to amplify sequences encoding DYW domains from various plant species. The second generation of longer degenerate oligonucleotide primers was designed using the bryophyte sequences as guide; these gave better PCR amplification from bryophyte samples. (D) Typical DCH–DYW amplicon sequences obtained from selected jungermanniid liverworts show the same conservation of amino acid positions as higher plants. In A and D accession numbers are indicated.

a wide range of non-seed land plants (liverworts, mosses, hornworts, ferns and fern allies, Table 2). Using a first generation of oligonucleotides encoding very short conserved peptides (see Fig. 2C), we were able to amplify DYW coding sequences from many bryophytes, but not from *Marchantia* or its close relatives, nor from algae. Knowing these *Bryophyta* sequences, we designed a second generation of longer oligonucleotides which were especially targeted towards lower plant DYW sequences. After optimization of the PCR conditions, amplification of the expected c. 135 bp fragments were obtained for all taxa except for the green algae and marchantiid liverworts included in our sampling (Fig. 3A). As low concentrations or poor quality DNA could be a trivial explanation for PCR failure, genomic DNA samples that did not show any amplification using DYW primers were tested by PCR using oligonucleotides designed to amplify the gene for the mitochondrial 18S ribosomal RNA. Amplification products were obtained in all samples using these primers as positive controls (Fig. 3B). To verify the identity

of the DYW domain amplification products, these were isolated, cloned and sequenced (Table 2). To check for diversity in the amplicons, several clones were examined for each species and different sequences were indeed identified as expected for a diversified gene family (Table 2). All 83 novel sequences were deposited in the database under Genbank accession numbers (EF607933–EF608015). The sequences showed conservation of the same amino acid positions as previously observed for the (mostly angiosperm) sequences in the database (Fig. 2). Special attention in our taxon sampling was given to *Haplomitrium*, an isolated liverwort genus with extremely high editing frequency [25,26]. Indeed, 12 different amplicon sequences were rapidly identified in *Haplomitrium mnioides*. Hence, the presence and apparent absence of the DYW domain seems to fully correlate with the known occurrence of pyrimidine exchange type of RNA editing in the land plant lineage which is present in all clades of land plants except the marchantiid liverworts and is also absent in algae.

Table 2  
A summary of DYW domain amplification and sequencing results

Taxonomic group	Species	PCR amplification	# of diff. seqs retr.	# of sequenced clones	Accessions numbers
Green algae	<i>Chaetosphaeridium globosum</i>	No			
	<i>Chlamydomonas reinhardtii</i>	No			
	<i>Chlorokybus atmophyticus</i>	No			
	<i>Closterium ehrenbergii</i>	No			
	<i>Coleochaete scutata</i>	No			
	<i>Klebsormidium flaccidum</i>	No			
Liverworts (Marchantiopsida)	Marchantiidae				
	<i>Asterella blumeana</i>	No			
	<i>Bucegia romanica</i>	No			
	<i>Conocephalum conicum</i>	No			
	<i>Corsinia coriandrina</i>	No			
	<i>Lunularia cruciata</i>	No			
	<i>Marchantia polymorpha</i>	No			
	<i>Monoclea gottschei</i>	No			
	<i>Monosolenium tenerum</i>	No			
	<i>Reboulia hemisphaerica</i>	No			
	<i>Riccia breidleri</i>	No			
	<i>Riccia fluitans</i>	No			
	<i>Ricciocarpos natans</i>	No			
	Haplomitriidae				
	<i>Haplomitrium mnioides</i>	Yes	12	13	EF607933–44
	Jungermanniidae				
	<i>Diplophyllum albicans</i>	Yes	8	9	EF607945–52
	<i>Fossombronia alaskana</i>	Yes	8	13	EF607953–60
	<i>Frullania tamarisci</i>	Yes	3	3	EF607961–63
<i>Lepidogyna hodgsoniae</i>	Yes	5	5	EF607964–68	
<i>Metzgeria furcata</i>	Yes	6	8	EF607969–74	
<i>Plagiochila asplenioides</i>	Yes	2	2	EF607975–76	
<i>Porella platyphylla</i>	Yes	3	3	EF607977–79	
<i>Ptilidium pulcherrimum</i>	Yes	6	6	EF607980–85	
<i>Scapania aequiloba</i>	Yes	5	5	EF607986–90	
<i>Scapania nemorea</i>	Yes	4	4	EF607991–94	
<i>Trichocolea tomentella</i>	Yes	2	2	EF607995–96	
Mosses (Bryopsida)					
<i>Mnium hornum</i>	Yes	2	2	EF607997–98	
<i>Physcomitrella patens</i>	Yes	1	1	EF607999	
<i>Sphagnum fallax</i>	Yes	1	1	EF608000	
Hornworts (Anthocerotopsida)					
<i>Anthoceros agrestis</i>	Yes	2	2	EF608001–02	
<i>Dicksonia antarctica</i>	Yes	2	2	EF608003–04	
<i>Isoetes velata</i>	Yes	2	4	EF608005–06	
<i>Selaginella tamariscina</i>	Yes	5	5	EF608007–11	
<i>Selaginella uncinata</i>	Yes	4	4	EF608012–15	
Total			83	94	

The '# of diff. seqs retr.' column indicates the number of distinguishable sequences obtained amongst the clones that were sequenced (# of sequenced clones). Sequences were submitted to Genbank and accessions numbers are indicated.

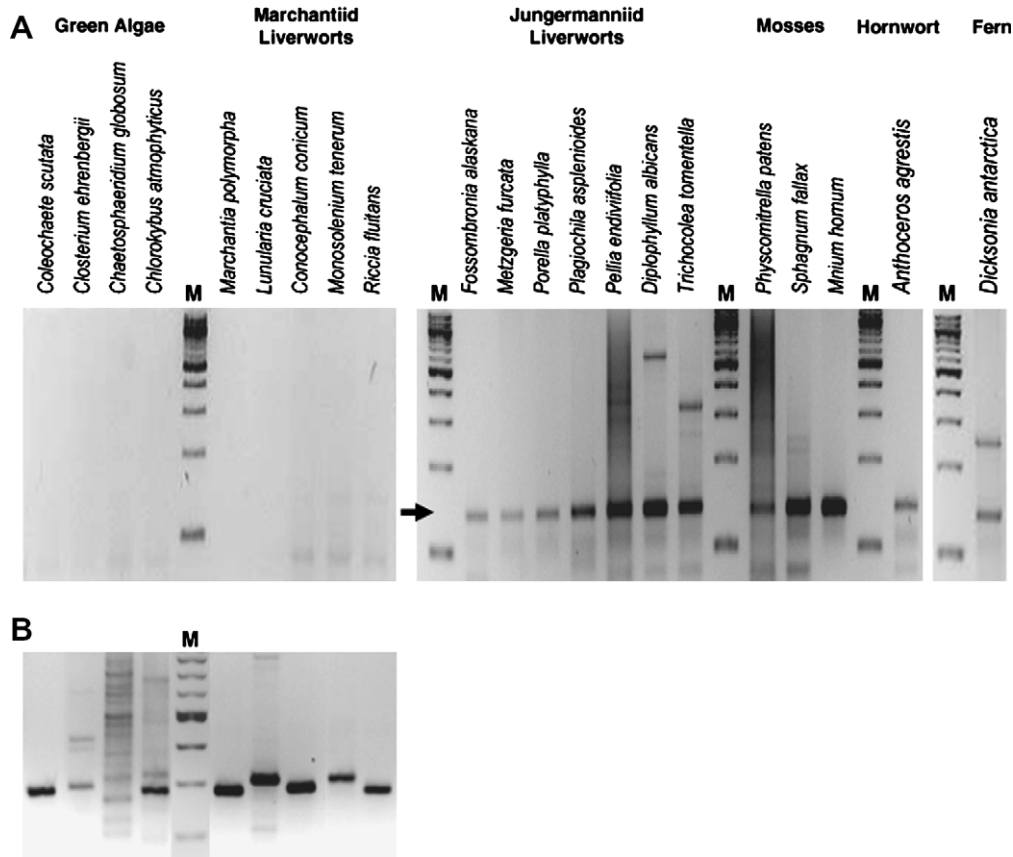


Fig. 3. Exemplar PCR amplification assays of the DYW domain in plants. (A) Successful PCR amplification of the DCH–DYW amplicon was generally observed for jungermanniid liverworts, mosses, hornworts and ferns giving rise to a 135 bp PCR product (arrow, right panels) but not for marchantiid liverworts or green algae (left panel) in our taxon sampling. M designates a 100 bp ladder used as size marker. (B) 18S rDNA amplification as a positive control for algal and marchantiid liverwort DNA qualities (in the same order from left to right as under A). Genomic DNA was extracted using Nucleobond kit to extract plant nucleic acids (Macherey & Nagel). Five  $\mu\text{g}$  of genomic DNA were digested by EcoRV or SmaI independently and then mixed together to make the template DNA solution. The PCR amplification assays contained 10 ng to 1  $\mu\text{g}$  of genomic DNA, 2.5  $\mu\text{l}$  of 10 $\times$  PCR buffer (200 mM Tris–HCl, 100 mM  $(\text{NH}_4)_2\text{SO}_4$ , 100 mM KCl, 20 mM  $\text{MgSO}_4$ , 1% Triton X-100, pH 8.8), 0.2 mM of each dNTP, 1  $\mu\text{M}$  of each primer, 2 mM  $\text{MgCl}_2$ , 0.5 U of *Taq* DNA polymerase (Biolabs). We used a touchdown program including an initial denaturation step (5 min, 94  $^\circ\text{C}$ ), followed by 20 cycles (each comprising 30 s at 94  $^\circ\text{C}$ , 1 min at 45  $^\circ\text{C}$  decreasing by 0.5  $^\circ\text{C}/\text{cycle}$  and 2 min at 72  $^\circ\text{C}$ ) and by 30 cycles (30 s at 94  $^\circ\text{C}$ , 1 min at 40  $^\circ\text{C}$ , 2 min at 72  $^\circ\text{C}$ ) and a final step of 7 min at 72  $^\circ\text{C}$ . PCR products were run on a 4% agarose gel.

## 5. Concluding remarks

To date, no plant mutants affected in the enzymatic step of RNA editing have been identified. Biochemical data are also inconclusive, as so far the *in vitro* editing assays that have been developed have not led to the identification of any of the specific molecular factors involved [9,19,27–29]. We have here addressed the question of the identification of the editing catalytic domain from an evolutionary perspective by studying the phylogenetic correlation between the DYW domain and RNA editing in a wide range of different plant groups. We found no evidence for the DYW domain in green algae, but the domain is unequivocally present in all vascular plants tested and all bryophytes apart from the subclade of complex-thalloid liverworts, the Marchantiidae. The phylogenetic distribution of the DYW domain hence exactly matches the distribution of RNA editing in plants demonstrating a strong correlation between the two (Fig. 4). Liverworts are thought to be monophyletic and to form a sister group to all other extant land plants, as e.g. most recently shown with a multi-taxa-multi-loci study [30]. Within liverworts, the Jungermanniidae and the Marchantiidae are identified as sister groups by all re-

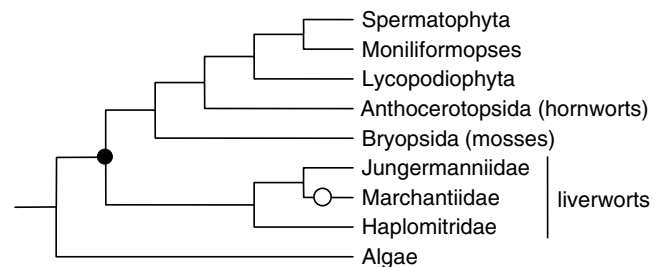


Fig. 4. Phylogenetic distribution of RNA editing in green plants. RNA editing events have been described in all land plant lineages with the exception of the marchantiid liverworts and no RNA editing has been identified in green algae. In the light of current models of land plant phylogeny this implies a unique gain of RNA editing in the earliest land plants (filled circle) and a secondary loss in the complex-thalloid, marchantiid, liverworts (open circle).

cent molecular studies independent of taxon or locus sampling [31–33]. Of particular importance is the phylogenetic position of the isolated genus *Haplomitrium*, formerly included among the Jungermanniidae but now strongly supported as branching basal to a jungermanniid–marchantiid dichotomy [26] (and ref-



erences therein). The strong phylogenetic evidence argues for the appearance of editing prior to the divergence of liverworts from the remaining land plants and for a secondary loss of editing in the marchantiid liverworts (Fig. 4). DYW domains are also absent from algae and appear to be absent from marchantiid liverworts correlating well with the presence or absence of RNA editing. It is important to note that this correlation holds over both the appearance of editing (*Haplomitrium*, the earliest branching lineage to carry out RNA editing clearly contains many DYW domain genes) and the loss of editing in the marchantiid lineage (we found no DYW domain sequences in the 11 marchantiid genera tested). Clearly, the presence of DYW domain sequences is a strong marker for RNA editing. In addition to this phylogenetic correlation, the DYW domain shows some structural and sequence similarities to cytidine deaminases and cytidine-deaminase-like proteins capable of C-to-U RNA editing in animals and fungi. In particular, the amino acid residues known to be essential for the deaminase activity are present in DYW domains and are extremely highly conserved amongst all DYW-containing proteins.

To date, the only mutant affected in expression of a DYW protein for which a molecular defect has been proposed is *err2*, an *Arabidopsis* mutant that lacks the plastid NAD(P)H dehydrogenase (NDH) complex [34]. CRR2 is needed for processing and possibly translation of *ndhB* transcripts, but mutant plants have not been shown to be defective in editing [34]. However, a role in editing for CRR2 cannot be ruled out as not all the editing sites in *Arabidopsis* plastids have been catalogued [35], much less tested in the mutant. Supporting genetic evidence in favour of a link between DYW proteins and editing comes from a study that mapped a locus quantitatively affecting editing of mitochondrial *cb206* transcripts to the vicinity of two genes encoding DYW proteins [36].

In conclusion, our data provide a strong phylogenetic argument in favour of a role of the DYW domain in RNA editing. Its sequence similarities with cytidine deaminases are consistent with a proposed role in catalyzing editing reactions. Together, these observations make the implication of DYW proteins in RNA editing a highly attractive hypothesis.

**Acknowledgements:** The work in the authors' laboratories is funded by the French National Institute for Agricultural Research (C.L.), French National Centre for Scientific Research (V.S.), the Australian Research Council (Grant CE0561495), the State Government of Western Australia (I.S.) and the Deutsche Forschungsgemeinschaft (V.K.).

## References

- Shikanai, T. (2006) RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol. Life Sci.* 63, 698–708.
- Tillich, M., Lehwark, P., Morton, B.R. and Maier, U.G. (2006) The evolution of chloroplast RNA editing. *Mol. Biol. Evol.* 23, 1912–1921.
- Bock, R. (2000) Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing. *Biochimie* 82, 549–557.
- Freyer, R., Kiefer-Meyer, M.C. and Kossel, H. (1997) Occurrence of plastid RNA editing in all major lineages of land plants. *Proc. Natl. Acad. Sci. USA* 94, 6285–6290.
- Malek, O., Lattig, K., Hiesel, R., Brennicke, A. and Knoop, V. (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. *Embo J.* 15, 1403–1411.
- Hiesel, R., Combettes, B. and Brennicke, A. (1994) Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. *Proc. Natl. Acad. Sci. USA* 91, 629–633.
- Steinhauser, S., Beckert, S., Capesius, I., Malek, O. and Knoop, V. (1999) Plant mitochondrial RNA editing. *J. Mol. Evol.* 48, 303–312.
- Ohyama, K. (1996) Chloroplast and mitochondrial genomes from a liverwort, *Marchantia polymorpha*—gene organization and molecular evolution. *Biosci. Biotechnol. Biochem.* 60, 16–24.
- Hirose, T. and Sugiura, M. (2001) Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast in vitro RNA editing system. *Embo J.* 20, 1144–1152.
- Kotera, E., Tasaka, M. and Shikanai, T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433, 326–330.
- Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K. and Shikanai, T. (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc. Natl. Acad. Sci. USA* 104, 8178–8183.
- Okuda, K., Nakamura, T., Sugita, M., Shimizu, T. and Shikanai, T. (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J. Biol. Chem.* 281, 37661–37667.
- Small, I.D. and Peeters, N. (2000) The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* 25, 46–47.
- Lurin, C. et al. (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16, 2089–2103.
- Blanc, V., Litvak, S. and Araya, A. (1995) RNA editing in wheat mitochondria proceeds by a deamination mechanism. *FEBS Lett.* 373, 56–60.
- Yu, W. and Schuster, W. (1995) Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria. *J. Biol. Chem.* 270, 18227–18233.
- Chan, L., Chang, B.H., Nakamura, M., Li, W.H. and Smith, L.C. (1997) Apobec-1 and apolipoprotein B mRNA editing. *Biochim. Biophys. Acta* 1345, 11–26.
- Wedekind, J.E., Dance, G.S., Sowden, M.P. and Smith, H.C. (2003) Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet.* 19, 207–216.
- Hegeman, C.E., Hayes, M.L. and Hanson, M.R. (2005) Substrate and cofactor requirements for RNA editing of chloroplast transcripts in *Arabidopsis* in vitro. *Plant J.* 42, 124–132.
- Faivre-Nitschke, S.E., Grienenberger, J.M. and Gualberto, J.M. (1999) A prokaryotic-type cytidine deaminase from *Arabidopsis thaliana* gene expression and functional characterization. *Eur. J. Biochem.* 263, 896–903.
- Aubourg, S., Boudet, N., Kreis, M. and Lecharny, A. (2000) In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol. Biol.* 42, 603–613.
- Bhattacharya, S., Navaratnam, N., Morrison, J.R., Scott, J. and Taylor, W.R. (1994) Cytosine nucleoside/nucleotide deaminases and apolipoprotein B mRNA editing. *Trends Biochem. Sci.* 19, 105–106.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Betts, L., Xiang, S., Short, S.A., Wolfenden, R. and Carter Jr., C.W. (1994) Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. *J. Mol. Biol.* 235, 635–656.
- Groth-Malonek, M., Pruchner, D., Grewe, F. and Knoop, V. (2005) Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol. Biol. Evol.* 22, 117–125.
- Groth-Malonek, M., Wahrmund, U., Polsakiewicz, M. and Knoop, V. (2007) Evolution of a pseudogene: exclusive survival of a functional mitochondrial *nad7* gene supports *Haplomitrium* as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. *Mol. Biol. Evol.* 24, 1068–1074.

- [27] Nakajima, Y. and Mulligan, R.M. (2005) Nucleotide specificity of the RNA editing reaction in pea chloroplasts. *J. Plant Physiol.* 162, 1347–1354.
- [28] Neuwirt, J., Takenaka, M., van der Merwe, J.A. and Brennicke, A. (2005) An in vitro RNA editing system from cauliflower mitochondria: editing site recognition parameters can vary in different plant species. *RNA* 11, 1563–1570.
- [29] Sasaki, T., Yukawa, Y., Wakasugi, T., Yamada, K. and Sugiura, M. (2006) A simple in vitro RNA editing assay for chloroplast transcripts using fluorescent dideoxynucleotides: distinct types of sequence elements required for editing of *ndh* transcripts. *Plant J.* 47, 802–810.
- [30] Qiu, Y.L. et al. (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. USA* 103, 15511–15516.
- [31] Davis, E.C. (2004) A molecular phylogeny of leafy liverworts (Jungermanniidae: Marchantiophyta) (Goffinet, B., Hollowell, V. and Magill, R., Eds.), *Molecular Systematics of Bryophytes. Monographs in Systematic Botany*, Vol. 98, pp. 61–86, Missouri Botanical Garden Press, St. Louis.
- [32] Forrest, L.L. and Crandall-Stotler, B.J. (2005) Progress towards a robust phylogeny for the liverworts, with particular focus on the simple thalloids. *J. Hattori Bot. Lab.* 97, 127–159.
- [33] Heinrichs, J., Gradstein, S.R., Wilson, R. and Schneider, H. (2005) Towards a natural classification of liverworts (Marchantiophyta) based on the chloroplast gene *rbcL*. *Cryptogamie Bryologie* 26, 215–233.
- [34] Hashimoto, M., Endo, T., Peltier, G., Tasaka, M. and Shikanai, T. (2003) A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast *ndhB* in *Arabidopsis*. *Plant J.* 36, 541–549.
- [35] Tillich, M., Funk, H.T., Schmitz-Linneweber, C., Poltnigg, P., Sabater, B., Martin, M. and Maier, R.M. (2005) Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J.* 43, 708–715.
- [36] Bentolila, S., Chateigner-Boutin, A.L. and Hanson, M.R. (2005) Ecotype allelic variation in C-to-U editing extent of a mitochondrial transcript identifies RNA-editing quantitative trait loci in *Arabidopsis*. *Plant Physiol.* 139, 2006–2016.

DYW1 At1g	121	GEKKAIVDRSKAYVKLSLKGKVRDAGYVPETKYV--LHDIDEEAKEKALMHHSERLAI	177
est_k_brevis3144	539	GE+ + + YV++ + + + Y P+ + V +D + + + L HH+E+ A+	700
DYW1 At1g	178	AFGIINTPPGTTIRVMKNLRCGDCHNFIKILSSI	212
est_k_brevis3144	701	A + N ++ V N R C DCH+F + S +	799

Figure n° 32 : Alignement de la partie C-terminale de la séquence protéique de DYW1 (At1g47580) et de la partie C-terminale de l'EST (gi48705712) de *Karenia brevis*.

Résultat de TBLASTN: capture d'écran du site NCBI

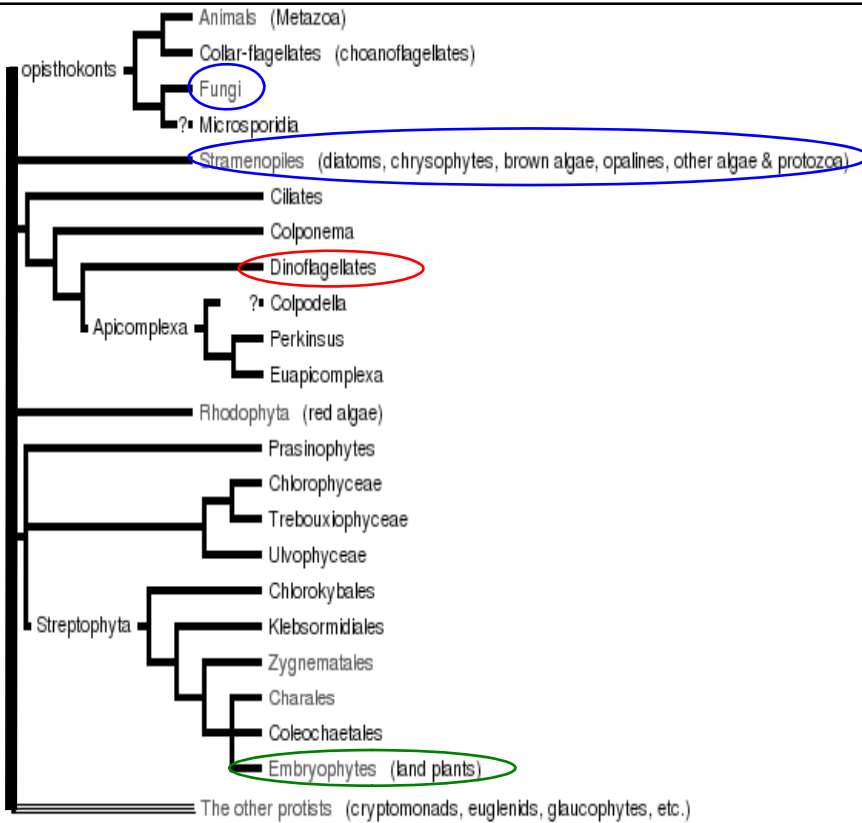


Figure n°33: Arbre phylogénétique représentant l'évolution des eucaryotes

Encerclé en rouge les dinoflagellées (ou Dinoflagellates)

Encerclé en vert les plantes terrestres (ou Embryophytes)

Encerclé en bleu les groupes taxonomiques des protozoaires et des champignons (auxquels appartient les trypanosomes et *Physarum*, respectivement)

Encadré n°1: **Edition chez les dinoflagellées**

<i>Pfiesteria piscicida</i> <i>Prorocentrum minimum</i> <i>Cryptocodinium cohnii</i>	}	A→G	47%
		U→C	23%
		C→U	17%
		<b>Lin et al., 2002</b>	

<i>Ceratium horridum</i>	}	A→G
		A→C
		<b>Zauner et al., 2004</b>



### **III Etude complémentaire : recherche de la présence éventuelle du domaine DYW chez les dinoflagellées.**

#### **III.1 Contexte de l'étude**

Au cours de mes analyses bioinformatiques (TBLASTN), j'ai découvert une EST (gi48705712) contenant un motif PPR à domaine DYW chez une dinoflagellée (*Karenia brevis*) (voir alignement, figure n°32). Ce fait a retenu mon attention : Comment expliquer évolutivement cette observation ?

Les dinoflagellées sont un groupe taxonomique phylogénétiquement distant des plantes terrestres (figure n°33). Il s'agit d'eucaryotes unicellulaires capables ou non d'activité photosynthétique, vivant le plus souvent dans les courants marins. Pour 4 espèces de ce groupe : *Pfiesteria piscicida*, *Prorocentrum minimum*, *Cryptocodinium cohnii*, et *Ceratium horridum*, un processus d'édition a été décrit dans des transcrits chloroplastiques (Lin et al., 2002; Zauner et al., 2004). A l'heure actuelle, seules ces deux publications font référence à un phénomène d'édition dans ce groupe taxonomique, si bien que nous n'avons pas une idée très précise quant à son étendue. De surcroît, il apparaît que le mécanisme d'édition chez ces 4 espèces est différent de celui décrit chez les plantes terrestres. En effet, il s'agit de plusieurs types de modification de bases (A-en-G, U-en-C, C-en-U, et A-en-C), en proportions différentes selon les espèces (voir encadré n°1).

La présence de l'EST contenant un motif PPR à domaine DYW chez *Karenia brevis* soulève plusieurs questions. *Karenia brevis* est-elle capable de réaliser le processus d'édition ? Existe-t-il un domaine DYW chez d'autres espèces de ce groupe taxonomique, notamment chez les espèces capable de réaliser l'édition ? Cette EST peut-elle avoir été attribuée à cette espèce suite à une erreur d'annotation dans la banque de données NCBI (présence d'une séquence contaminante)?

Afin de répondre à ces questions, j'ai développé deux approches parallèles : la recherche de la présence du domaine DYW dans ce groupe taxonomique, notamment chez les espèces réalisant l'édition, et la validation de la présence de la PPR de type DYW chez *Karenia brevis*.

#### **III.2 Recherche de la présence du domaine DYW chez les dinoflagellées**

Pour cette étude, j'ai utilisé l'outil technique développé pour étudier la distribution phylogénétique du domaine DYW dans la lignée verte, à savoir, la PCR avec des amorces dégénérées dessinées sur ce domaine. Ce choix peut paraître risqué étant donné la grande distance évolutive entre les plantes terrestres et les dinoflagellées. Cependant, l'EST trouvée chez *Karenia brevis*, suggère que, s'il existe dans ce groupe taxonomique, le domaine DYW est conservé (voir alignement, figure n°32).

Des PCRs avec 2 couples différents d'amorces dégénérées (DCH/CACGDYW et DCH/CSCRDYW) ont été réalisées sur l'ADNg de 9 espèces, y compris *Karenia brevis* (voir table n°4). Aucune amplification n'a pu être obtenue. Afin, de vérifier que l'absence d'amplification n'est pas liée à un problème trivial de qualité ou concentration de la

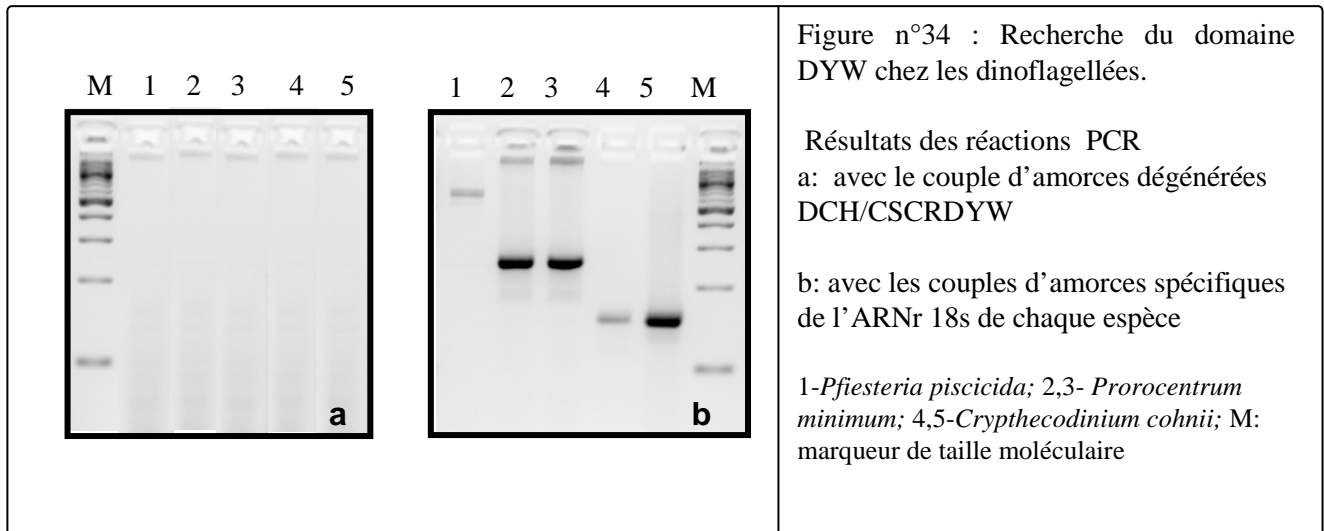
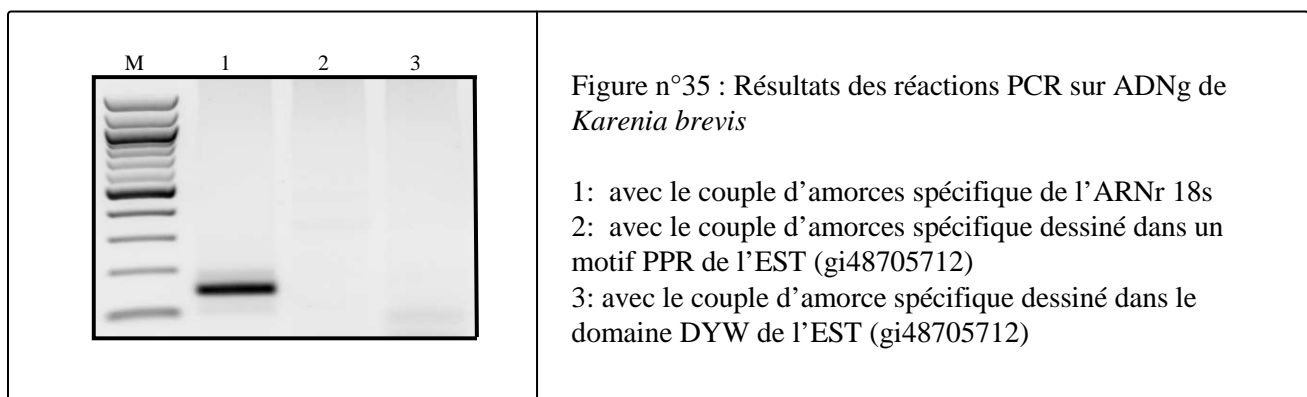


Table n°4: Recherche du domaine DYW chez les dinoflagellées: bilan des résultats PCR

Espèce	Résultat PCR avec les amorces dégénérées DCH/CSCRDYW	Résultat PCR avec les amorces dégénérées DCH/CACGDYW	Résultat PCR avec les amorces spécifiques de ARNr 18s
<i>Karenia brevis</i>	Pas d'amplification	Pas d'amplification	Amplification
<i>Cryptocodinium cohnii</i>	Pas d'amplification	Pas d'amplification	Amplification
<i>Prorocentrum minimum</i>	Pas d'amplification	Pas d'amplification	Amplification
<i>Pfiesteria piscicida</i>	Pas d'amplification	Pas d'amplification	Amplification
<i>Alexandrium tamarense</i>	Pas d'amplification	Pas d'amplification	Amplification
<i>Amphidinium carterae</i>	Pas d'amplification	Pas d'amplification	Amplification
<i>Ceratium horridum</i>	Pas d'amplification	Pas d'amplification	Amplification
<i>Pyrocystis lunula</i>	Pas d'amplification	Pas d'amplification	Amplification
<i>Guillardia theta</i>	Pas d'amplification	Pas d'amplification	Amplification



préparation d'ADNg, nous avons pris soin de tester les ADNg en réalisant des PCR avec des amorces spécifiques de l'ARNr 18s de chacune de ces espèces.

La figure n°34 présente les résultats de PCR pour 3 espèces : *Prorocentrum minimum*, *Cryptocodinium cohnii* et *Pfiesteria piscicida* (dinoflagellées capables d'édition). L'ensemble des résultats obtenus est résumé dans la table n°4.

**En résumé, par cette approche de PCR avec des amorces dégénérées, la présence d'un éventuel domaine DYW n'a pas pu être mise en évidence chez les dinoflagellées.**

### **III.3 Validation de la présence d'une PPR de type DYW chez *Karenia brevis***

Un couple d'amorces spécifique de l'ARNr 18s de *K. brevis* a permis de contrôler la qualité de l'ADN génomique de cette espèce. Des PCR ont également été réalisées en utilisant deux couples d'amorces spécifiques de l'EST d'intérêt : l'un dans le domaine DYW, et l'autre dans la région PPR. Une amplification avec le couple d'amorces spécifiques de l'ARN 18s a été obtenue, tandis qu'aucune amplification avec les deux couples d'amorces spécifiques de l'EST recherchée (figure n° 35) n'a été observée.

**Pour conclure, la séquence de PPR à domaine DYW trouvée dans la banque de données du NCBI semble avoir été attribuée à *K. brevis* par erreur.**

Les deux couples d'amorces dessinés spécifiquement sur l'EST (gi48705712) n'ont pas, faute de matrice appropriée, pu être validés comme fonctionnels en PCR. Il nous est donc impossible d'exclure totalement la possibilité d'une absence d'amplification due à un problème technique lié à ces amorces (mauvaise synthèse, problème de quantité...). Cependant le fait d'avoir utilisé deux couples différents réduit cette probabilité. D'autre part, comme nous l'avons déjà souligné, le niveau de conservation du domaine DYW de cet EST est grand (voir alignement figure n°32). Si cet EST existait vraiment chez *K. brevis*, les PCR utilisant les couples d'amorces dégénérées dessinées sur le domaine DYW auraient, très certainement, donné un produit d'amplification, or j'ai obtenu le résultat contraire (voir table n°4).

### **III.4 Bilan**

Les résultats de cette approche complémentaire indiquent qu'il n'existe probablement pas de domaine DYW chez les dinoflagellées, et que l'EST (gi48705712) qui correspond à une PPR de type DYW a été attribuée à *Karenia brevis* sans doute par erreur. Cela conforte, d'une certaine façon, le résultat obtenu précédemment, à savoir que le domaine DYW est spécifique de la lignée verte, et que sa distribution corrèle avec celle du processus d'édition C-en-U des transcrits des organites.

Ce bilan n'est pas vraiment surprenant, car, bien que des hypothèses d'acquisitions indépendantes ou de transfert horizontal entre espèces auraient pu être discutées, il semblait évolutivement difficile d'expliquer la présence d'un domaine DYW dans un groupe taxonomique aussi distant des plantes terrestres.

D'autre part, comme le suggère Lin et al. (Lin et al., 2002), l'édition telle qu'elle a été mise en évidence chez les dinoflagellées (modification de plusieurs types de nucléotides) semble « comparable » au processus d'édition par insertion/délétion de nucléotides existant chez les trypanosomes et les champignons myxomycètes. Qui plus est, les dinoflagellées sont évolutivement plus proches de ces groupes taxonomiques, que des plantes terrestres (figure n°33).

#### **IV Discussion**

Lorsque nous avons débuté l'approche présentée dans ce chapitre, en 2004, nous savions que les protéines PPR semblaient impliquées dans des mécanismes de régulation post-transcriptionnelle aux seins des organites. Bien que leur implication dans le processus d'édition n'était pas encore établie, nous avons remarqué que l'explosion de la sous-famille de PPR de type PLS était spécifique des plantes terrestres, tout comme ce processus (Lurin et al., 2004). Par ailleurs, le domaine DYW des protéines PPR de type PLS semblait posséder une activité catalytique potentielle (Aubourg et al., 2000; Lurin et al., 2004). Pris ensemble, ces éléments suggéraient que les protéines PPR de type PLS pouvaient être impliquées dans le processus d'édition C-en-U dans les organites.

L'enjeu était de développer une étude globale sur le domaine DYW afin de préciser son rôle dans la biologie de la plante.

Pour cela, à l'aide d'une approche *in silico*, j'ai mis en évidence des similarités entre la région C-terminale très conservée de ce domaine et le site actif des cytidines désaminases, en particulier, les protéines APOBEC impliquées dans des mécanismes d'édition C-en-U chez les mammifères (Chan et al., 1997). Bien que ces similarités aient été observées au niveau de la séquence primaire et secondaire, on ne peut pas parler d'homologie forte entre le domaine DYW et les désaminases. Il est alors légitime de s'interroger sur la signification biologique que l'on peut attribuer à de telles ressemblances. Nous avons proposé que le domaine DYW pourrait « dériver » de désaminases agissant sur des ARN, en particulier les cytidines désaminases. En aucun cas, nos résultats ne démontrent une telle hypothèse.

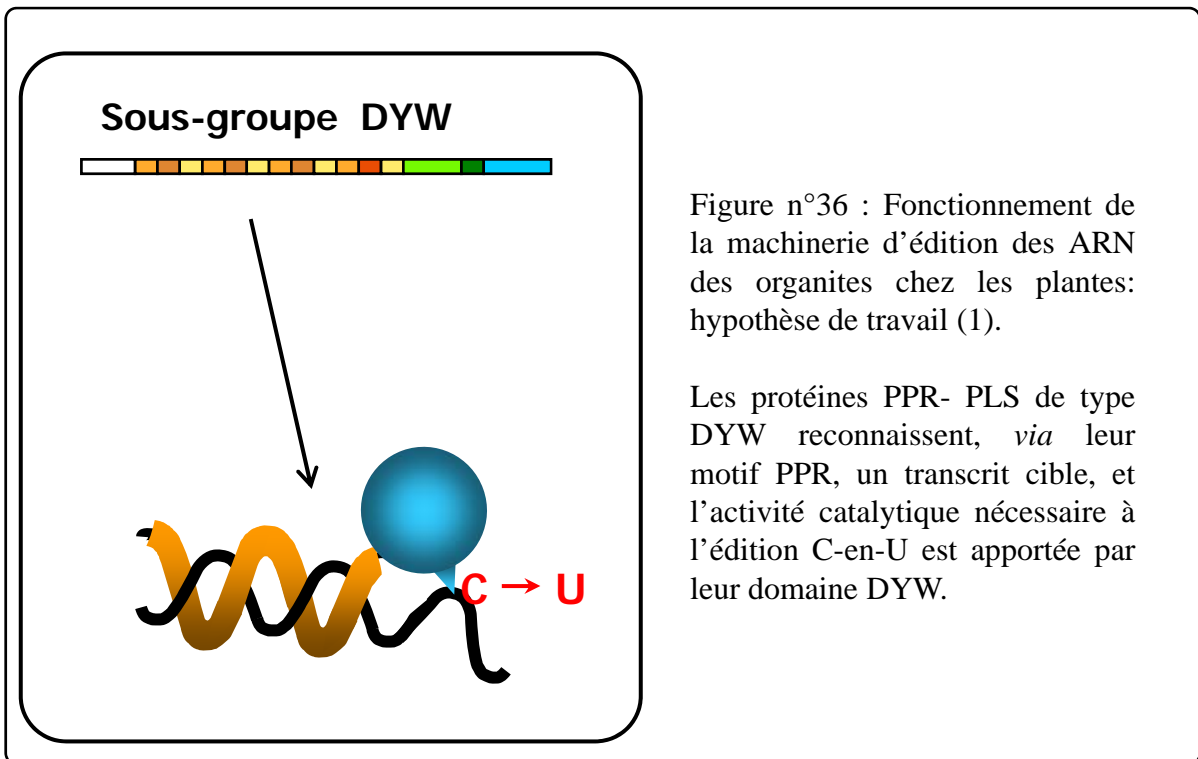
Par ailleurs, dans cette approche bioinformatique, nous avons observé qu'il existait des centaines de séquences de protéines PPR à domaine DYW chez les angiospermes, alors qu'un nombre très restreint avait été isolé pour d'autres groupes taxonomiques comme les mousses. Ceci pouvait suggérer que l'explosion de la sous-famille de PPR à domaine DYW au sein de la lignée verte peut corrélérer avec l'étendue du phénomène d'édition. Une comparaison récente de la famille de protéine PPR d'*Arabidopsis thaliana*, *Oryza sativa* (le riz) et *Physcomitrella patens* (une mousse) a permis de montrer qu'il existe effectivement une corrélation entre le nombre de protéine PPR de type PLS (dont celles de type DYW) et le nombre de sites édités chez les plantes (O'Toole et al., 2008). En effet, il existe près de 250 de protéines PPR de la sous-famille PLS chez le riz et *Arabidopsis*, où plus de 400 sites d'édition ont été répertoriés dans les transcrits mitochondriaux et au moins une trentaine dans les transcrits chloroplastiques (Giege and Brennicke, 1999; Tsudzuki et al., 2001; Notsu et al., 2002; Chateigner-Boutin and Small, 2007). En revanche, chez la mousse *Physcomitrella*, la sous famille de protéines PPR de type PLS se restreint à 10 protéines PPR de type DYW et de façon intéressante, le phénomène d'édition des transcrits des organites est moins fréquent puisque seuls 9 sites sont actuellement décrits chez cet organisme (Miyata et al., 2002; Miyata and Sugita, 2004; O'Toole et al., 2008). Ainsi, non seulement la sous-famille de protéines

PPR de type PLS et le processus d'édition sont spécifiques des plantes, mais de surcroît l'extension rapide de cette sous-famille de protéines et du phénomène d'édition corrélent.

Afin de déterminer la distribution phylogénétique du domaine DYW dans la lignée verte, nous avons choisi le pari d'une approche par PCR en utilisant des amorces dégénérées. Après quelques mises au point techniques, cette tentative s'est avérée fructueuse. J'ai ainsi pu mettre en évidence que le domaine DYW, tel qu'il est décrit chez les plantes à fleurs, est absent des algues vertes, mais présent chez toutes les plantes vasculaires, en particulier chez toutes les bryophytes, à l'exception d'un sous-groupe d'hépatiques : les Marchantiidae. Cette distribution phylogénétique corréle strictement avec la distribution du phénomène d'édition dans la lignée verte. En effet, dans la phylogénie de la lignée verte admise par l'ensemble de la communauté scientifique à l'heure actuelle, les hépatiques sont un groupe monophylogénétique (dérivé d'un même ancêtre commun). Au sein de ce groupe, les Marchantiidae et les Jungermanniidae sont deux groupes sœurs, et les Haplomitridae aurait divergé avant leur séparation (Davis and Wurdack, 2004; Forrest and Crandall-Stotler, 2005; Groth-Malonek et al., 2007). Ainsi, la vision la plus parcimonieuse quant à l'émergence du phénomène d'édition propose que ce phénomène soit apparu au moment de l'émergence des plantes terrestres, il y a plus de 450 millions d'années. Ce mécanisme aurait donc été présent chez l'ancêtre commun des hépatiques et secondairement perdu chez les Marchantiidae. Il paraît important de souligner ici, que nous avons en réalité une double corrélation phylogénétique entre la présence du domaine DYW et le processus d'édition dans la lignée verte, au niveau de leur émergence d'une part et de leur perte secondaire chez les Marchantiidae, d'autre part.

Cette approche, n'a été possible que sur la base d'une propriété du domaine DYW : sa grande conservation au sein de tous les groupes taxonomiques. Le couple d'oligonucléotides dégénérés utilisé, a été dessiné dans la région C-terminale la plus conservée du domaine, et est de ce fait de taille très restreinte (~150pb). Une limitation majeure de cette approche réside dans le fait qu'il est possible que chez les algues vertes et les Marchantiidae, le domaine DYW soit présent mais suffisamment divergent au moins dans cette région pour que les amorces n'aient pas pu s'hybrider. Afin de palier à ce problème, nous avons tenté, sans succès, l'utilisation d'autres couples d'amorces dégénérées dans des régions moins conservées du domaine DYW. Aussi, il semble difficile par ce type d'approche d'amplifier l'ensemble du domaine DYW, et strictement, cela nous permet de démontrer l'existence d'une corrélation entre la présence d'une région conservée du domaine DYW et le processus d'édition au sein de la lignée verte. Cependant, les amorces dégénérées utilisées amplifient la partie C-terminale du domaine DYW qui contient les résidus histidines et cystéines conservés suggérant une potentielle activité catalytique pour le domaine DYW (Aubourg et al., 2000; Lurin et al., 2004), et pour laquelle j'ai montré qu'il existe des homologies avec le site actif de cytidines désaminases. Dans l'hypothèse où il existerait, chez les algues vertes et/ou les Marchantiidae, un domaine DYW plus divergent et ne possédant pas cette conservation dans la partie C-terminale, il est alors légitime de se demander si un tel domaine pourrait avoir conservé une éventuelle activité catalytique.

Il faut également noter que chez les anthocérotes (Anthocerotopsida), bien que l'on observe des cas d'édition C-en-U, c'est l'édition « inverse » (de type U-en-C) qui prédomine (Yoshinaga et al., 1996). J'ai montré que le domaine DYW est présent dans ce groupe taxonomique, et nous pensons que le domaine DYW peut catalyser la désamination C-en-U des réactions d'édition dans la lignée verte. Peut-on envisager que le domaine DYW trouvé chez les anthocérotes soit également capable de catalyser la réaction d'édition « inverse » U-



en-C, ou bien est-ce une enzyme différente qui porte cette activité catalytique ? A l'heure actuelle, il est difficile d'apporter une réponse définitive à cette question car aucune transaminase agissant sur des ARN (capable de catalyser des désaminations C-en-U et U-en-C) n'a été décrite dans la littérature. D'ailleurs une telle enzyme existe-t-elle ou bien est-ce deux enzymes différentes qui catalysent ces deux types de réactions ? Notons cependant que les domaines DYW isolés chez les anthocérotes sont très similaires à ceux trouvés chez les autres espèces. S'ils devaient porter une activité catalytique de transaminase et non de simple cytidine-désaminases, n'aurait-on pas observé certaines différences ? Cependant, comme nous venons d'en discuter, nous savons que les oligonucléotides dégénérés utilisés ne permettent l'amplification que de séquences très conservées. Si chez ces espèces, certaines protéines portent un domaine DYW plus divergent et porteur d'une nouvelle activité catalytique, transaminase par exemple, notre approche nous aurait-elle permis de les amplifier ?

### **En conclusion**

Cette étude du domaine DYW dans son ensemble nous a permis de mettre en évidence de nouveaux éléments en faveur de son implication dans le processus d'édition des ARN chez les organites des plantes ; et ces indications soutiennent solidement qu'il pourrait catalyser les réactions d'édition.

Ainsi, nous avons proposé un modèle de travail quant au rôle des protéines PPR à domaine DYW dans le processus d'édition (figure n°36). En effet, ces protéines, *via* leur motif PPR, pourraient jouer le rôle de facteurs capables de spécifier la cytidine à modifier dans la molécule d'ARN ; et leur domaine DYW pourrait être le facteur enzymatique catalysant la désamination C-en-U de la réaction d'édition. L'identification récente d'un mutant d'une PPR à domaine DYW (*FLV*) impliquée dans l'édition chez *Arabidopsis*, soutient fortement ce modèle (Chateigner-Boutin *et al.*, résultats non publiés).

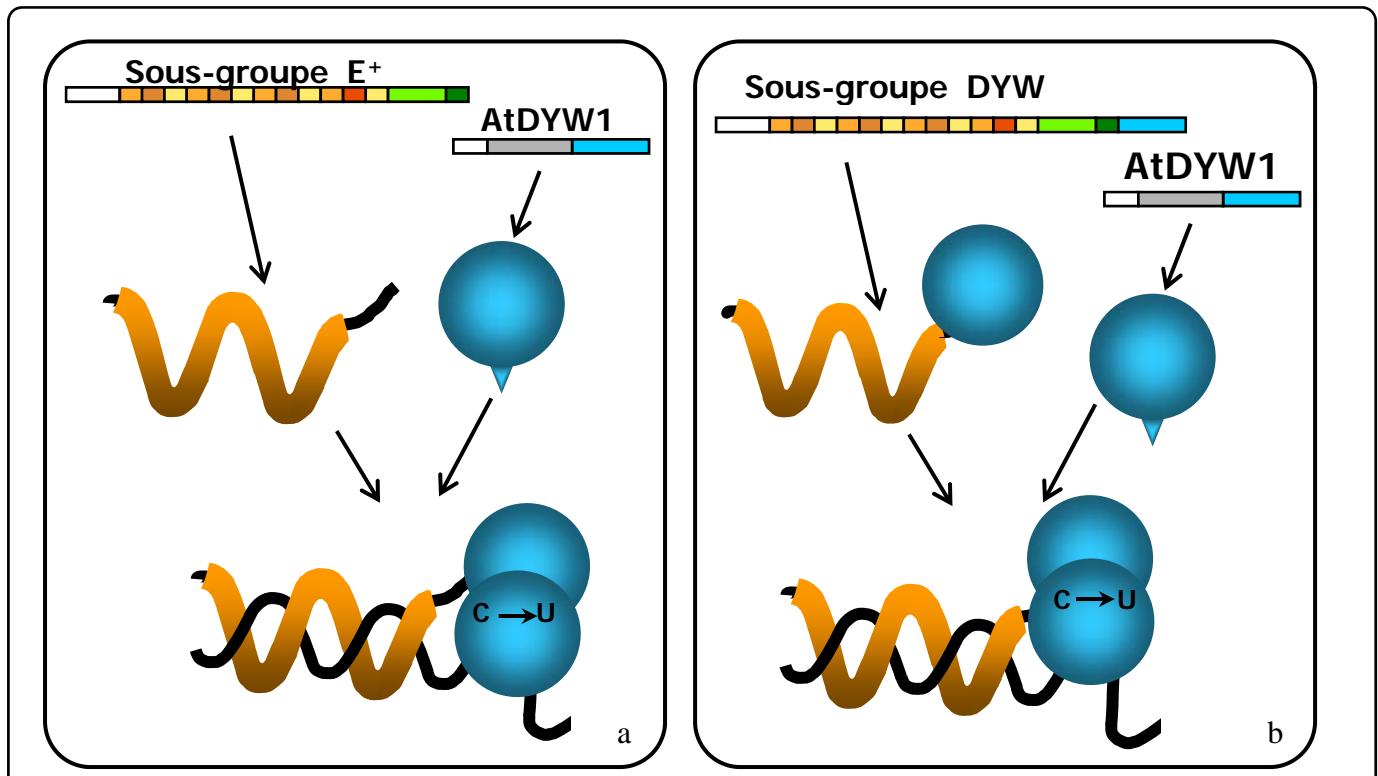
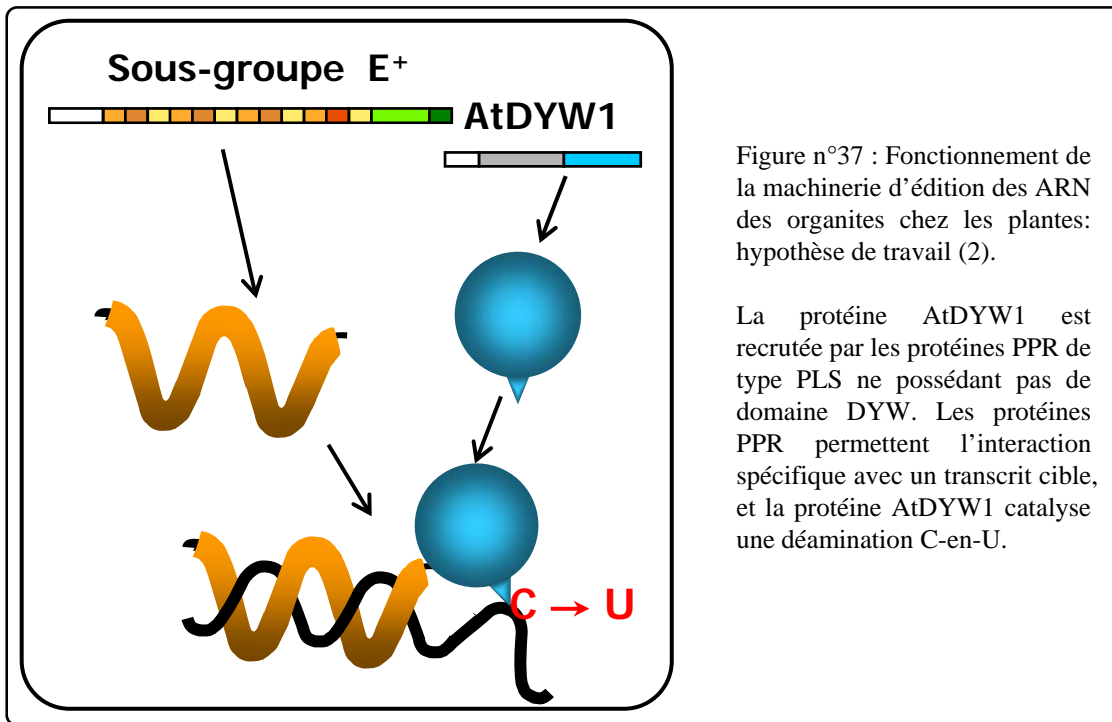


Figure n°38 : Fonctionnement de la machinerie d'édition des ARN des organites chez les plantes: précision du modèle dans l'hypothèse où (a) la protéine AtDYW1 interagit avec les protéines PPR de type PLS ne possédant pas de domaine DYW, sous forme de dimère, afin de catalyser la réaction d'édition, (b) les protéines PPR- PLS de type DYW recrutent la protéine AtDYW1 afin de catalyser la réaction d'édition.



## **Partie 2:**

### **LE CANDIDAT *AtDYW1***

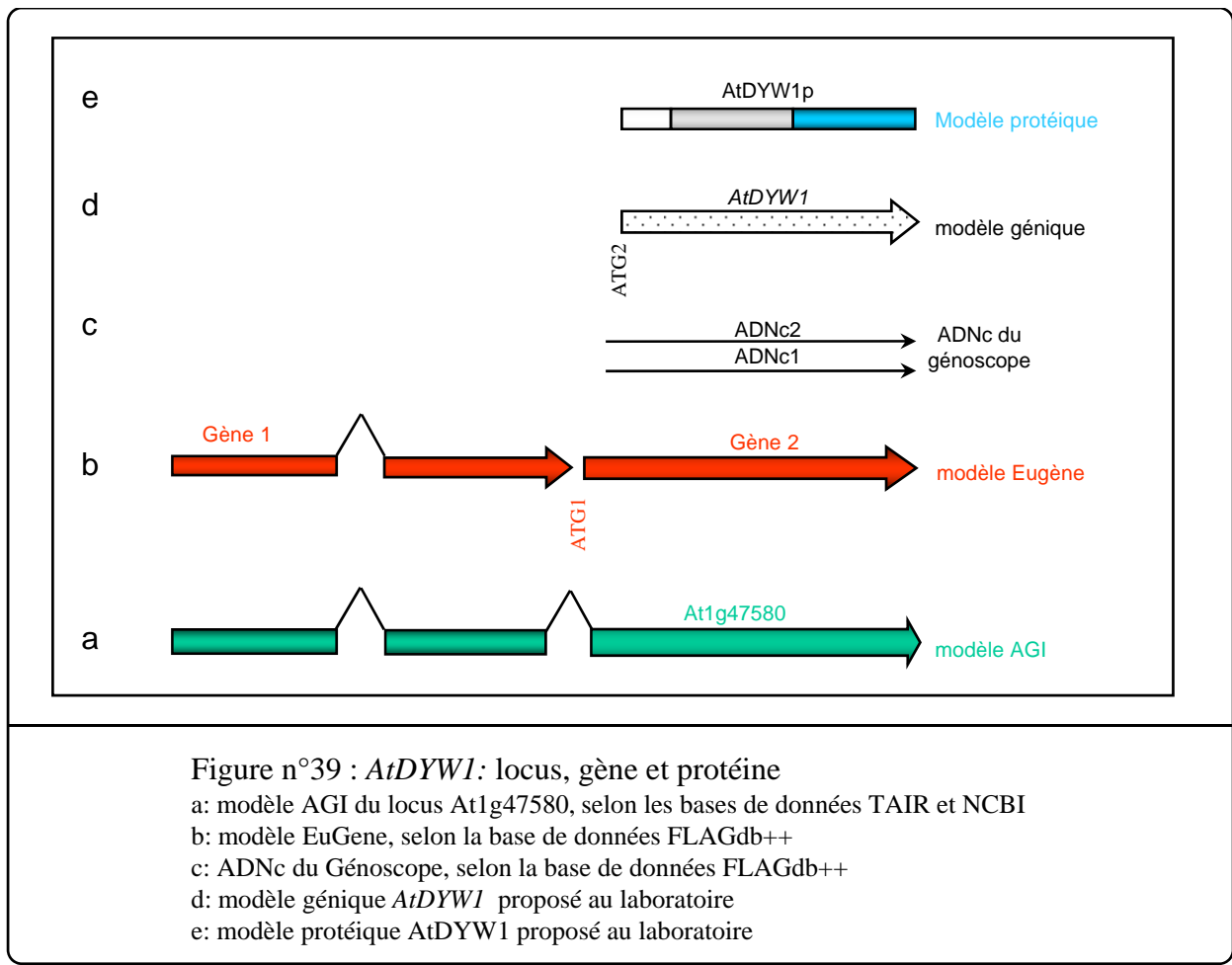
#### **I Contexte de l'étude**

Après avoir étudié le domaine DYW dans son ensemble et proposé un modèle quant au rôle des protéines PPR-PLS de type DYW dans le processus d'édition C-en-U des transcrits des organites chez les plantes, j'ai consacré la suite de mon travail de thèse à la caractérisation d'un candidat particulier : l'ORF *At1g47580*. Cette ORF *At1g47580* a retenu tout particulièrement notre attention car elle code pour une protéine unique chez *Arabidopsis* : la protéine *AtDYW1* composée d'un peptide d'adressage pour les organites et d'un domaine DYW, sans réel motif PPR.

#### **II Rôle putatif de *AtDYW1* : une hypothèse de travail**

A l'heure actuelle, nous savons que plusieurs protéines PPR, toutes de type PLS, sont impliquées dans l'édition C-en-U des transcrits chloroplastiques chez *Arabidopsis*. A l'exception de la PPR-FLV (de type DYW), les autres PPR décrites (CRR4, CRR21 et CLB19) sont de type E/E+ et ne portent ni domaine DYW, ni aucun autre domaine permettant éventuellement de proposer une activité catalytique (Kotera et al., 2005; Okuda et al., 2007; Chateigner-Boutin et al., soumis). Par conséquent, ces protéines pourraient être des facteurs de spécificité dans le processus d'édition, et elles doivent recruter en *trans* un facteur catalytique afin de catalyser la désamination C-en-U. Nous proposons que la protéine *AtDYW1*, de part sa structure uniquement composée du domaine DYW, pourrait être ce facteur catalysant la réaction d'édition C-en-U, recruté par les protéines PPR de type PLS ne possédant pas de domaine DYW (figure n°37).

Sachant que de nombreuses cytidines désaminases agissent sous forme de dimères ou de multimères, nous avons proposé une alternative quant à ce modèle de travail. En effet, afin d'être dans une conformation active pour catalyser la réaction d'édition, la protéine *AtDYW1* formerait un homodimère (ou éventuellement multimère) lorsqu'elle est recrutée par des protéines PPR de type E/E+ (figure n° 38a). De même, nous avons proposé à la fin du chapitre précédent, un modèle quant au rôle des protéines PPR-PLS de type DYW dans le processus d'édition C-en-U. Nous envisageons également que la protéine *AtDYW1* pourrait s'associer au domaine DYW d'une protéine PPR afin de former un hétérodimère actif catalysant la réaction d'édition (figure n° 38b).



---KKGVVVYVDPTRF---VLHDIIDQEA---KE-QALLYHSERLAIAYGIITCP---PRKT---	LTIIKKNLRCVCGDCHNF IKIMSKIIGRVLIVRDNKRPFHHFK---DGKCSCGDYW
---YVPETKY---VLHDIIDQEA---KE-KALMHHSERLAIAPGIINTP---PGTT---	IRVMKNLRCVCGDCHNF IKILLSIEDREIIVRDNKRPFHHFK---DGNCSGCDYW
LAGYRPIT--TOVLLNLEBED---KE-NVLHYHSEKLATAPGLLISTS---PGKP---	IRVKNLRCVCGDCHNSIKLISKVYKRKITVDRDKRPFHHFK---DGCSCMDYW
EIGYAPVS--EVMHVEEED---KE-FALRYHSEKLAVAFGLLKT---VQVA---	ITVKNLRCVCGDCHSAFKYISIVSNROIIVRDKLRFHHFK---NGSCSCLDYW
SEGYVQI--SNVVDVEEED---KE-NAVYHSEKLAIAPMLISTP---ERSP---	ITVKNLRCVCGDCHLAIKLVSKVYNRIVVDRSRPFHHFK---NGSCSQDYW
LAGYVAEI--TSELHNLESEED---KG-YALRYHSEKLAIAPGLIVTP---PGTT---	IRVTKNLRICVCGDCHNFAKVFSDVYDRIVVDRSRPFHHFK---GSCSCNDFW
LAGYKPED--SNRTAPTLESERC---LG-SADTSHSEKLAIAYGLMRTT---AGTT---	LRITKKNLRCVCRDCHSFTKYYSKAFNREIIVRDRVRPFHHFK---DGICSCNDYW
SNGYPEI--SOVLIIEEED---KE-OSLNLHSEKLAIAYGLISTE---APKV---	IRVKNLRCVCGDCHSVAKLISQLYDREIIVRDRVRPFHHFK---NGQCSNDFW
LAGYKADT--TPVMFDIEEED---KE-DALCLHSEKAAIAPGIMSLK---EDVP---	IRVKNLRCVCGDCHQVSMMSKIFNREIIVRDRNRPFHHFK---DGHCSGDFW
EDGYVAQT--GLVLDHIGEEED---KE-NALCYHSEKLAVAYGLMMDGA---DEESP---	VRVINLRCVCGDCHVFKHISKYKREIIVRDRVRPFHHFK---DGCSCSRDFW
EAGYVTD--KEVLLDLNEKD---KE-IALSYHSEKLAIAPCLMKTR---PGTP---	VRINKNLRICGDCHLVMMKMSKLSFIREIIVRDCNRPFHHFK---DGCSCSRDFW
LIGYKNT--GDAFFDVDEED---KE-SSIHMHSEKLAIAYGMKTK---PGTT---	IRVKNLRCVCGDCHTIVTKLISEVYGREIIVRDRNRPFHHFK---NGVCSRDYW
FEGYSPDL--TOILLNVDEEED---KE-RLK--GHSOKVAIAPGLLYTP---PGSI---	IKIARNLRMCSDCHTYTKLISMIYEREIIVRDRNRPFHHFK---GGTCSCKDYW
KEGYIPDP--REVSONLEEED---KE-RLICGHSERLAIAPGLLNTE---PGTP---	LOVMKNLRCVCGDCHVTKLISKIVGREILVDRNRPFHHFK---DGTCSCKDRW
KEGYVPT--SCVLHVEEED---KE-ILLCGHSEKLAIAPGLINTS---PGTI---	IRVAKNLRVCGDCHLAKTPIISKIVDREIILVDRVRPFHHFK---NGTCSGCDYW
KEGYVADT--SGVLHDVDEEH---KR-ELLFGHSEKLAVAFGIINTE---PGTT---	IRVTKNLRICVCGDCHVAIKPTISKITEREIVRDRSRPFHHFK---RGNCSGCDYW
EAGYVPT--NLVLG--VENDV---KE-QLLSTHSEKLAIAPGLLNTT---AGTT---	IRVKNLRCVCGDCHNATKYLISLVGTREIIVRDRMORPFHHFK---NGACSCGDYW
LSGYVPT--SMVHANMDOE---KE-ITLRYHSEKLAIATFGLLNTT---PGTT---	IRVKNLRCVCRDCHNAAKLISLIFGRKIVLRDVRPFHHFK---DGCSCGDFW
EAGYVNT--SSVCH-DVDEED---KE-MTLRVHSEKLAIAPGIMNTV---PGST---	VNVKNLRCVCGDCHNVIKLISKIVDREFVVRDAKRFHHFK---DGCSCGDFW
ELGYAPDY--RSALH-DVEDEQ---KE-EMLWYHSEKLAIAPGLINTV---EGSA---	VTVMKNLRCVCGDCHTIVKLSGVVGREIIVLRDIPRFHHFK---NGTCSGCDYW
ELGYVPT--SHVIV-HVDOQE---RE-VNLQYHSEKLAIAPALLNTT---PGST---	ITHKKNLRVCGDCHTAIKLASKVVGREIIVRDRNRPFHHFK---DGNCSCKDYW
SIGYLPDRS--QAPIVDATNDGS---KE-YSLRLHSEKLAIAPGLINLP---PQTP---	IRIFKNLRCVCGDCHVEVTKLISKVFNTEIIVRDRVRPFHHFK---DGCSCLDYW
MIGYVDPH--RFVLQDVEDD---E-KE-HILMHSERLAIAPALLIATP---AKTT---	IRIFKNLRCVCGDCHSVTKLISKITEREIVRDRSRPFHHFK---NGVCSGCDYW
ENGYVEL--EMLLDLVDDDE---RE-AIVHSHSEKLAIATYGLIKTK---PGTI---	IRIMKNLRCVCGDCHKVTKLISKIYKRDIVMRDRTRPFHHFK---DGCSCGDFW
EAGYQET--ELALHD-VYEEED---RE-LMVKYHSEKLAIAPGLIATE---PGTE---	IRIKNLRCVCGDCHVTKLISKITEREIVRDRNRPFHHFK---DGYCSGCDYW
SHGYVPT--NTVLQD-LEETE---KE-OSLQVHSEKLAIAYGLISTK---PGSP---	LKIFKNLRCVCGDCHTIVTKLISKITGRKIVMRDRNRPFHHFK---DGCSCGDFW
DLGYEPT--SYVLQD-IDDEH---KE-AVLAQHSEKLAIAPGLIATP---KQSP---	LLIKNLRCVCGDCHLVKILAKIETEREIVRDRSRPFHHFKSDGVCSCGDFW
KAGYVPT--SEVLQE-MEEEW---KE-GALRHHSEKLAIAPGLISTK---PGTK---	LRIVKNLRCVCRNCHETAIKLISKIYKREIIVRDRTRPFHHFK---DGCSCNDYW
KMGYVPT--ASVLHD-LEEEV---KE-QILRHHSEKLAIAPGLISTP---DKTT---	LRIMKNLRCVCGDCHTAIKFISKIVGREIIVRDRTRPFHHFK---DGFCSRDYW
SAGYVDPVS--GLVSHSEGDLE---EKE-QALCCHSEKLAVCPGLLETK---PSTP---	LVVFNLRVCRDCHSVMKIVSKVYDREIIVRDRNRPFHHFK---GGSCSCDYW
PLGYSPDVAG--AFHDVDEE---QRE-DNLSYHSEKLAVAYGLIKTP---EKSP---	LVVTKNLRVCRDCHSAIKLISKISKRYIVRDRSRPFHHFK---DGCSCCDYW
KSGHRPNLDF--ALHDVEEQ---EQE-QMLWGHSEKLAVVFGLLNTP---DGTP---	LOVKNLRCVCGDCHAVIKFISYAGREIFIRDRNRPFHHFK---DGLCSGDFW
LACYKPELEF--ALHNVEEED---OKE-KLLWHSEKLAVAFGLIKLP---QGSQ---	IOVFNLRVCRDCHKATKFISETKREIIVRDRTRPFHHFK---DGCSCGDFW
KVGYVMDKKW--AMVKEDEACEGDEE---MRLGHSEKLAVAFGLMSTR---DQSE---	LVVKNLRCVCGDCHNAIKFMSAVAGREIIVRDRNRPFHHFK---DGCSCGDFW
SEMGYKSKTEF--VLHDVGED---VKE-EMINOHSEKLAIAPGLIVVS---PME---	PITIKNLRVCGDCHETAIKLISKIYKREIIVRDRSRPFHHFK---DGCSCGDFW
KFGYKPSVEH--ELHDVEIG---KKE-ELLMOHSEKLALVFGMLCLP---ESS---	PVRIMKNLRVCRDCHSVMKIAATRFRIIVRDRNRPFHHFK---DGHCSGDFW
ELGYVDPSES--ALHDVEED---DKE-THLAVHSEKLAIAPFALMNTK---EEEDSN---	NTIRITKNLRVCRDCHVAAKLISQITSRREIIVRDRNRPFHHFK---FGVCSGDFW
EMGYVPEMRS--ASQDIEED---EQEEDNLHSEKLALAFVMEGR---KSSDCGV---	NLQICMKNLRVCRDCHNFMKLSKLLGKELVDRDRNRPFHHFK---DGCSCNDYW
DEAGYTPVAVS--ALYDVEEED---EKE-DQVSKHSEKLALAFALARLS---KGAN---	IRVKNLRCVCRDCHAVMKTLSKYVYGREIIVRDRNRPFHHFK---DGCSCRDYW
GIGYVPTNF--VLQDLEGE---QME-DSLRHHSEKLALAFGLMTRLP---IKEV---	IRITKNLRCVCGDCHVFKLASKLEIIVRDRTRPFHHFK---DGCSCRDYW

Figure n°40 : Alignement multiple des 40 protéines à domaine DYW les plus proches du consensus établi pour ce domaine.  
(de haut en bas signifiant de la plus proche à la plus éloignée).  
En rouge, la séquence correspondant à la protéine *AtDYW1* (partie C-terminale)

### **III Une annotation corrigée pour le locus At1g47580**

Dans les bases de données (TAIR et NCBI), le locus At1g47580 est décrit comme codant potentiellement une lipoyltransférase (ORF de 1251pb) (figure n°39a). Cependant, les prédictions d'annotation du logiciel EuGène (Foissac et al., 2003) présentées dans la base de données FLAGdb++ (Samson et al., 2004), indiquent à ce locus l'existence de deux gènes potentiels. Le premier, incluant la région d'ADN qui présente une homologie avec une lipoyltransférase (LIP2p), et un second gène en aval portant un motif DYW (notre gène d'intérêt, comme nous l'expliquerons, ci-après) (figure n° 39b).

Par ailleurs, l'étude des ADNc d'*Arabidopsis thaliana* séquencés par le Génoscope®, nous a permis de montrer qu'il existe 2 ADNc exprimés à partir du locus At1g47580 qui correspondent (à quelques bases près) au second gène prédit par EuGène (figure n°39c).

Forts de ces éléments, nous proposons que l'annotation AGI du locus At1g47580, en une ORF unique de 1251pb, est incorrecte. Il existe un fragment génomique de 555pb, présentant une homologie avec la LIP2, et pour lequel aucun ADNc n'a pu être mis en évidence. Nous proposons qu'il s'agisse d'un pseudogène résultant d'une duplication d'un exon de la LIP2p, existant ailleurs dans le génome. Notre ORF d'intérêt, de 696pb, se situe juste en aval de ce pseudogène. L'examen des ADNc exprimés à partir du locus At1g47580 a permis de déterminer que le codon ATG initiateur de traduction (que nous appelons ATG2) est localisé 24pb en aval de celui prédit par EuGène (arbitrairement appelé ATG1) (figure n°39d). Notons que, comme pour les protéines PPR, l'examen des données des puces CATMA indique que le gène *AtDYWI* est faiblement exprimé dans plusieurs tissus de la plante, tels que les feuilles et les fleurs (Jean-Pierre Renou, communication personnelle).

Comme nous l'avons déjà évoqué, l'ORF At1g47580 code pour une protéine de 239 acides aminés, composée en N-terminal, d'un peptide d'adressage pour les organites, et en C-terminal, d'un domaine DYW (figure n°39e). Une analyse de la séquence en acides aminés de cette protéine, *via* le logiciel predict protein (Rost and Liu, 2003), permet de souligner l'existence d'un motif de fixation au zinc (résultats non présentés). De plus, afin de préciser la nature de la séquence peptidique se trouvant dans la protéine *AtDYWI* entre la partie N-terminale correspondant au peptide d'adressage et la partie C-terminale correspondant au domaine DYW, nous avons mené des analyses de blast de cette séquence sur la banque de données de NCBI. Bien que cette région ne corresponde pas à de « véritables » motifs PPR, elle présente, cependant, des homologies avec le motif PPR de protéines d'*Arabidopsis* suggérant qu'il pourrait s'agir d'un motif PPR ancestral (voir discussion générale, paragraphe V-1).

Il existe près de 90 protéines PPR ayant un domaine DYW chez *Arabidopsis thaliana*. Cependant, le gène At1g47580 est le seul gène de cette sous-famille ne codant pas pour un (ou plusieurs) motifs PPR réel(s). Ainsi, ce gène est unique dans le génome d'*Arabidopsis*, et a été baptisé *AtDYWI*. Par ailleurs, le domaine DYW du gène *AtDYWI* est l'un des plus proches du consensus de ce domaine établi à partir des séquences de toutes les PPR à domaine DYW chez *Arabidopsis* (figure n°40). Cette grande conservation suggère une forte pression de sélection sur ce gène, et par conséquent un rôle important dans la « biologie » de la plante (survie cellulaire, développement...).

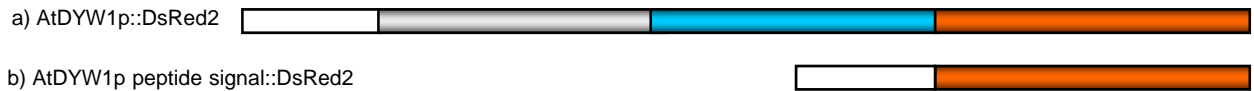


Figure 41: Constructions utilisées pour les expériences de localisation subcellulaire de la protéine AtDYW1

- a) ORF pleine longueur *AtDYW1* fusionnée à la DsRed2  
 b) peptide d'adressage du gène *AtDYW1* fusionné à la DsRed2

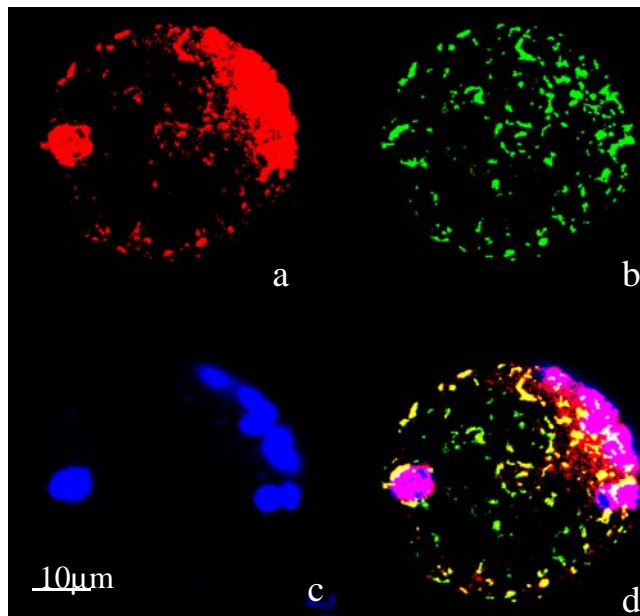


Figure 42: Observation au microscope confocal de protoplastes co-exprimant le peptide d'adressage du gène *AtDYW1* fusionné à la DsRed2 et le peptide d'adressage du gène de la prolyl-tRNA synthétase fusionné à la GFP.

- a) Signaux émis dans le spectre de la RFP (DsRed2: Excitation/Emission maximum: 558nm/583nm), correspondant à la fluorescence du peptide d'adressage du gène *DYW1* fusionné à la RFP.
- b) Signaux émis dans le spectre de la GFP (Excitation/Emission maximum: 479nm/509nm), correspondant à la fluorescence du peptide d'adressage du gène de la prolyl-tRNA synthétase fusionné à la GFP. La prolyl-tRNA synthétase étant une protéine mitochondriale, ceci facilite la visualisation de ces organites.
- c) Signaux d'«autofluorescence» correspondant à la fluorescence des pigments chlorophylliens des thylacoïdes à l'intérieur des chloroplastes (Emission >600nm).
- d) Superposition de trois signaux mettant en évidence une colocalisation des signaux RFP avec les signaux d'autofluorescence des chloroplastes (rose) et avec les signaux GFP (jaune/orangé)

## **IV Localisation subcellulaire de la protéine AtDYW1**

### **IV.1 Analyses bioinformatiques**

Les logiciels de prédiction de localisation subcellulaire TargetP et Predotar prédisent uniquement, avec une probabilité de 46 et 84% respectivement, une localisation chloroplastique pour la protéine AtDYW1. En revanche, l'utilisation du logiciel pSort (Nakai and Horton, 1999) m'a permis de noter qu'un adressage mitochondrial (62%) et chloroplastique (52%) étaient tout autant envisageable pour la protéine AtDYW1. D'après le logiciel TargetP, un site de clivage du peptide signal est prédit après le vingt-septième acide aminé, et ainsi la protéine AtDYW1 mature aurait une taille de ~25 kDA.

### **IV.2 Etude de la localisation subcellulaire de protéines de fusion**

Pour cette étude, deux constructions ont été utilisées : une fusion de la protéine AtDYW1 pleine longueur (du codon ATG2 au dernier codon avant le stop) avec la protéine DsRed2 et une fusion de la préséquence de la protéine AtDYW1 (arbitrairement fixée au 100 premiers acides aminés) avec la protéine DsRed2 (figure n°41).

J'ai étudié la localisation subcellulaire de ces protéines de fusion par expression transitoire dans des feuilles de *N. Benthamiana*. Pour ces deux constructions, une localisation chloroplastique évidente et reproductible a été observée (figure n°42). De plus, pour la protéine chimérique composée du peptide d'adressage de la protéine AtDYW1 fusionné à la RFP, un signal mitochondrial a pu être observé (figure n°42). Cependant ce signal est faible, comparé au signal chloroplastique, et il n'est pas reproductible dans tous les protoplastes transformés.

### **IV.3 Expériences de western blot utilisant l'anticorps anti-AtDYW1**

J'ai également étudié la localisation de la protéine AtDYW1 chez *Arabidopsis*, par des expériences de western blot utilisant un anticorps dirigé contre cette protéine. Je détaille dans la quatrième partie des résultats la production de l'anticorps dirigé contre la protéine AtDYW1 et la mise au point de son utilisation au laboratoire (paragraphe II.3, partie 4 des résultats). Toutefois, on peut noter, dès à présent, que successivement trois anticorps polyclonaux ont été produits (notés respectivement AC1, AC2 et AC3), et que seul le dernier (AC3), nous a permis de détecter la protéine AtDYW1 dans cette étude.

En utilisant cet anticorps (AC3), j'ai montré que la protéine AtDYW1 est exprimée chez *Arabidopsis* et enrichie dans une fraction chloroplastique (comparée à une fraction de protéines totales). En revanche, dans des fractions enrichies en mitochondries préparées à partir de feuilles d'*Arabidopsis*, je n'ai pas pu mettre en évidence la présence de cette dernière (figure n° 43). La qualité de ces fractions enrichies en protéines mitochondriales a été testée, grâce à un anticorps anti-FDH (protéine mitochondriale d'*Arabidopsis*), par Catherine Colas des Francs-Small (résultats non présentés).

En revanche, la présence d'une bande d'apparence spécifique a été détectée sur des fractions enrichies en mitochondries préparées à partir de pomme de terre. Ce résultat a été obtenu par

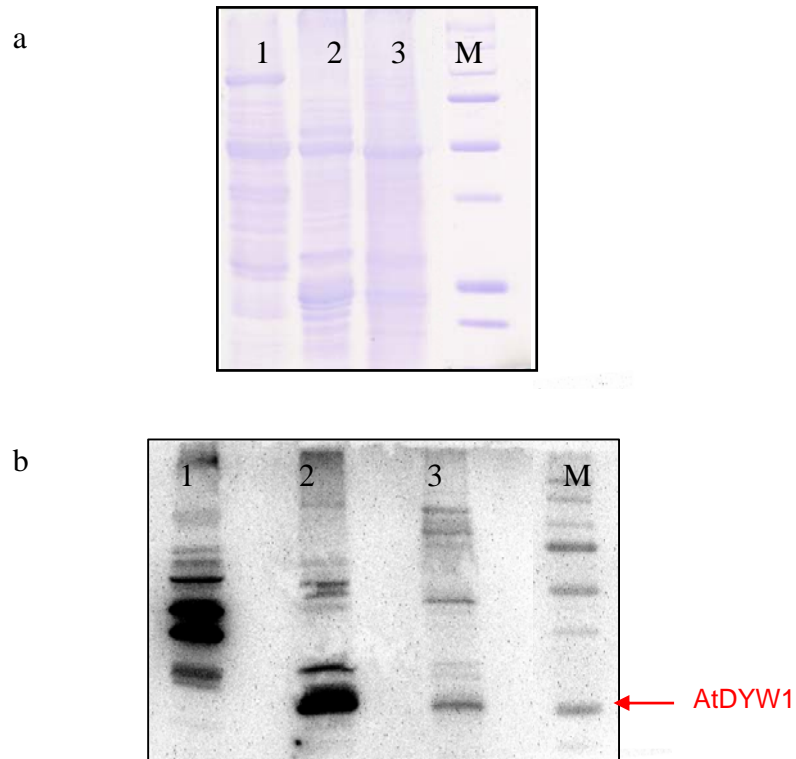


Figure 43: Etude de la localisation de la protéine AtDYW1 par des expériences de western blot utilisant un anticorps dirigé contre cette protéine.

- a) Gel coloré en bleu de Coomassie permettant de vérifier la quantité de protéines déposées.
- b) Membrane de western blot après révélation  
(hybridation avec l'anticorps primaire AC dirigé contre la protéine AtDYW1 et l'anticorps secondaire anti-lapin)

Échantillons:

- (1) Fraction protéique d'Arabidopsis enrichie en mitochondries
- (2) Fraction protéique d'Arabidopsis enrichie en chloroplastes
- (3) Fraction de protéines totales d'Arabidopsis

*Nota bene:* La présence de bandes non spécifiques est discutée dans la 4ème partie des résultats.

western blot en utilisant l'anticorps (AC2) dirigé contre la protéine *AtDYW1* (Catherine Colas des Francs-Small, communication personnelle).

## **V Interprétations et Bilan**

L'apport majeur de cette étude est d'avoir démontré une expression et une localisation chloroplastique, sans équivoque, de la protéine *AtDYW1*. Ces résultats sont repris dans le manuscrit proposé au paragraphe III.5 (troisième partie des résultats).

En revanche, il est plus difficile de conclure quant à une éventuelle localisation mitochondriale.

Tout d'abord, nous savons que, dans le cas des protéines adressées aux deux organites, il existe un biais de prédiction, de sorte que ces logiciels prévoient préférentiellement une localisation chloroplastique (le score mitochondrial est généralement assez faible) (Ian Small, communication personnelle).

De plus, le signal mitochondrial observé en protoplastes de *N. benthamiana* pour la protéine de fusion préséquence-*AtDYW1*::RFP est faible et peu reproductible. Ceci ne nous permet pas de trancher entre une faible quantité de protéine réellement localisée dans les mitochondries ou un artefact expérimental. Suite à l'examen des banques d'ADNc du Génoscope, nous utilisons l'ATG2 comme codon ATG initiateur de traduction, bien qu'EuGène prédise un ATG différent, ajoutant 8 acides aminés en N-terminal du peptide signal. Que l'on utilise l'un ou l'autre ATG, les prédictions de localisation obtenues avec TargetP et Predotar restent très similaires indiquant préférentiellement une localisation chloroplastique. Cependant nous pouvons nous demander s'il en est de même dans la réalité et nous interroger sur les conséquences d'une éventuelle erreur dans « notre modèle génique » pour *AtDYW1* (voir aussi discussion générale, paragraphe IV.1).

Par ailleurs, à notre seuil de détection, nous n'avons pas pu mettre en évidence la présence de la protéine *AtDYW1* dans les mitochondries d'*Arabidopsis*. Toutefois la production et l'utilisation d'un anticorps dirigée contre *AtDYW1* au laboratoire ont révélé certaines difficultés techniques (voir partie 4 des résultats, paragraphe II.3). Par conséquent, il est envisageable que notre anticorps ne soit pas assez sensible pour détecter une faible quantité de protéine *AtDYW1* dans les mitochondries d'*Arabidopsis*, si elle existe.

Enfin, il est également légitime de s'interroger sur les informations que l'on peut réellement tirer des expériences de western blot réalisées sur une préparation des mitochondries de pomme de terre. En effet, l'anticorps utilisé (AC2) pourrait « cross-interagir » avec une protéine homologue d'*AtDYW1* présente dans les mitochondries de pommes de terre. Mais quelle est la réelle identité d'une telle protéine : une protéine de structure identique à *AtDYW1* ou une protéine PPR à motif DYW ? *A priori*, la petite taille de la bande obtenue par Cathie Colas des Francs dans ces expériences suggère qu'il s'agisse bien d'une protéine homologue à *AtDYW1*. Cependant, nous savons qu'il existe quelques protéines PPR à domaine DYW ayant un nombre très réduit de motif PPR, dont la taille pourrait être assez similaire.

En conclusion, dans cette partie, nous avons mis en évidence l'existence de la protéine *AtDYW1*, qui de part ses caractéristiques, pourrait être un facteur enzymatique essentiel pour catalyser les réactions d'édition chez *Arabidopsis*. Nous avons également démontré que cette protéine est exprimée et localisée dans les chloroplastes; et qu'une éventuelle localisation mitochondriale ne peut pas être totalement exclue.



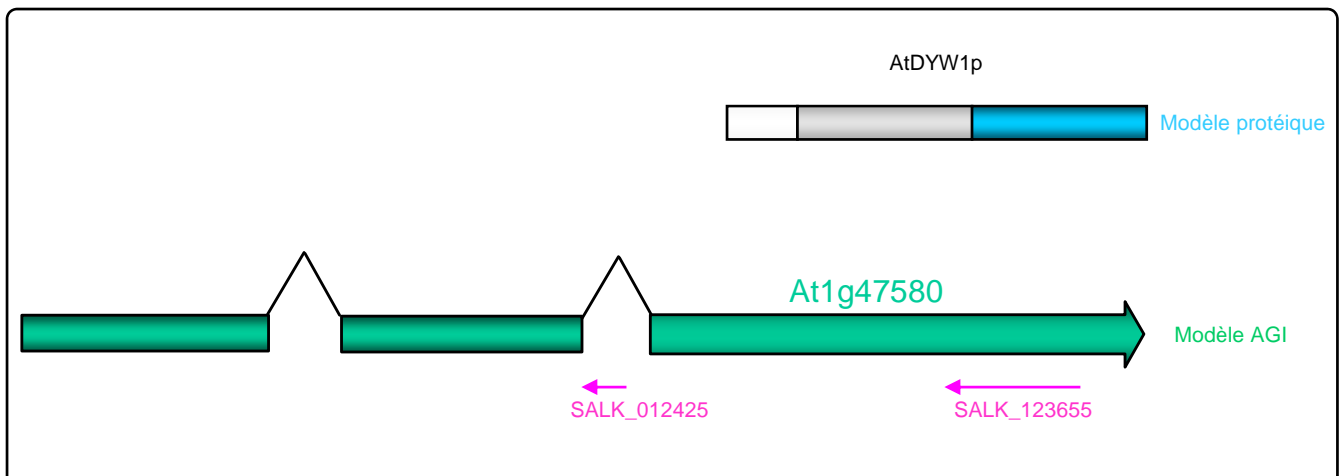


Figure n°44 : Localisations probables des insertions dans le locus At1g47580 pour les lignées SALK-123655 et SALK-012425.

Les modèles AGI et la protéine AtDYW1 correspondante sont représentés

## **Partie 3:**

### **UNE ETUDE GENETIQUE DU GENE *AtDYWI***

#### **I L'approche génétique : les grandes lignes**

Nous avons proposé que la protéine *AtDYWI*, de part ses caractéristiques, pourrait être une enzyme essentielle aux réactions d'édition chez *Arabidopsis*. Afin, d'évaluer cette hypothèse, j'ai consacré une grande partie de mon travail de thèse à développer une étude génétique sur ce gène candidat.

J'ai mené plusieurs approches en parallèle : la recherche de lignées pour lesquelles l'extinction de l'expression du gène *AtDYWI* est totale (mutants d'insertions), l'élaboration et la caractérisation de lignées pour lesquelles l'extinction de l'expression du gène *AtDYWI* est partielle (mutants « ARNi » pour ARN interférent : constitutifs ou inductibles), et enfin l'élaboration et la caractérisation de lignées surexprimant des copies du gène *AtDYWI* sauvage ou porteur de mutations (lignées de surexpression ou lignées « dominant-négatif »).

#### **II Mutants d'insertion**

##### **II.1 Contexte de l'étude**

Afin d'étudier la fonction du gène *AtDYWI* chez *Arabidopsis thaliana*, nous avons cherché l'existence de mutants d'insertion parmi les collections internationales (SALK Institute, INRA Versailles, GABI).

Il s'avère que deux lignées du SALK institute (SALK-123655, SALK-012425) touchent potentiellement le gène *AtDYWI* (figure n°44). Les mutants d'insertion sont obtenus après infiltration d'*Arabidopsis* avec une souche d'*Agrobacterium tumefaciens* comportant un vecteur binaire permettant l'intégration d'un ADN-T (ADN dit de transfert) dans le génome de la plante (génération T0). Notons qu'il y a en moyenne deux insertions par plante et que la stratégie de génétique inverse mise en place au SALK Institute permet de localiser, de façon approximative, une seule insertion sur la carte du génome d'*Arabidopsis thaliana*. C'est pourquoi, il est nécessaire de vérifier que l'insertion de l'ADN-T est bien localisée précisément dans le gène d'intérêt.

Ces deux lignées ont été reçues sous la forme d'un mélange de graines issues des plantes de la génération T3. Il s'agit donc d'un mélange de graines de génotypes potentiellement sauvage, hétérozygote ou homozygote pour l'insertion. C'est pourquoi nous devons analyser le phénotype macroscopique et le génotype de chaque plante individuellement.

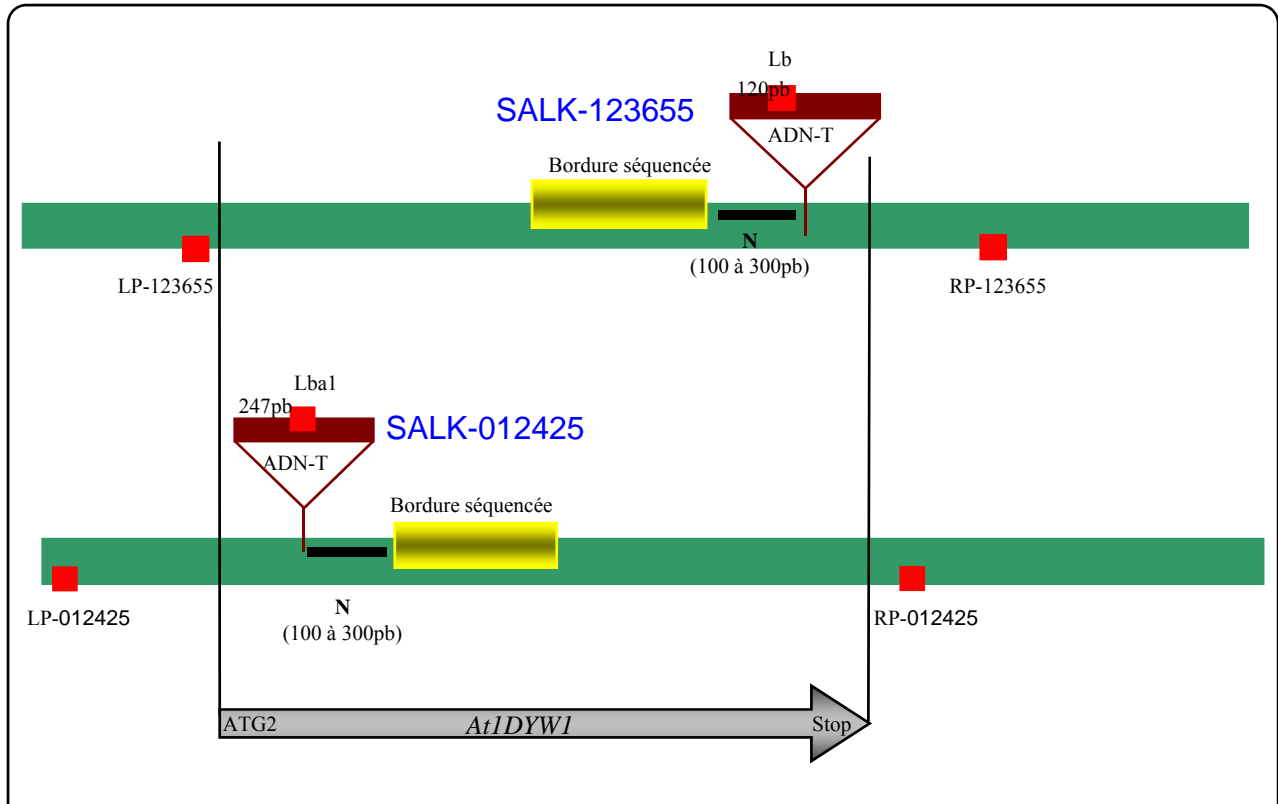


Figure n°45 : Principe du génotypage des lignées SALK-123655 et SALK-012425

#### Cas de la lignée SALK-123655

LP-123655/ RP-123655: amorces spécifiques de l'allèle sauvage (~900pb)

Lb: amorce spécifique de l'ADN-T et localisée à 120pb de son extrémité gauche

*Remarque: L'ADN-T est localisé dans le gène AtDYWI, si la taille du produit d'amplification correspondant à l'allèle mutant (additionnée de 120 pb) est inférieure à la distance LP123655-stop (distance de 480 pb).*

#### Cas de la lignée SALK-012425

LP-012425/ RP-012425: amorces spécifiques de l'allèle sauvage (~1000pb)

Lba1: amorce spécifique de l'ADN-T et localisée à 120pb de son extrémité gauche

*Remarque: L'ADN-T est localisé dans le gène AtDYWI, si la taille du produit d'amplification correspondant à l'allèle mutant (retranchée de 247 pb) est supérieure à la distance LP-ATG2 (distance de 545 pb).*

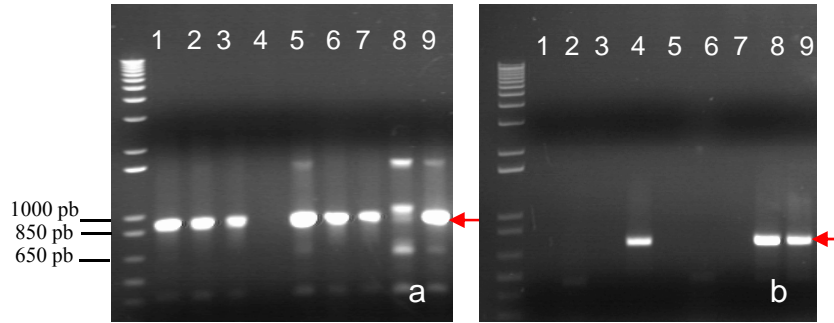


Figure n°46 : Résultats des PCR de génotypage des plantes de la lignée SALK-12655

a : amplification de l'allèle sauvage avec le couple LP123655/ RP123655 (~900pb)  
 b : amplification de l'allèle mutant avec le couple LP-123655/ Lb (~750pb)

1-3, 5-7: plantes homozygotes sauvages pour l'insertion  
 4 : plante homozygote mutante pour l'insertion  
 8,9 : plantes hétérozygotes pour l'insertion

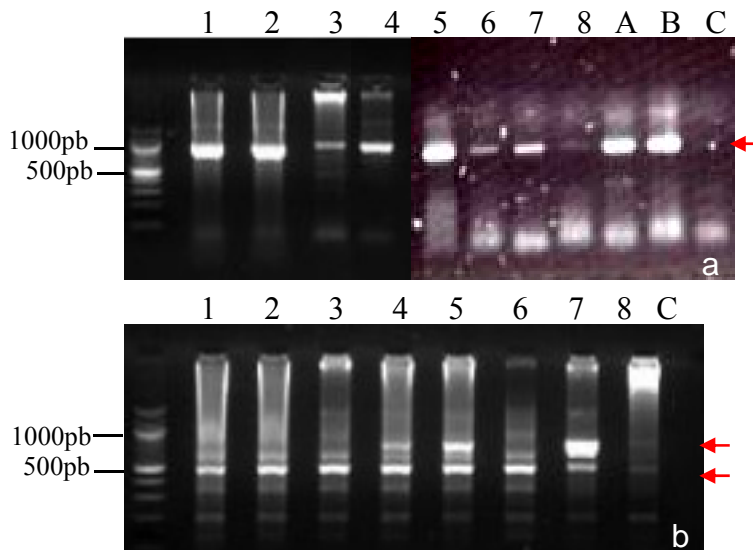


Figure n°47 : Résultats des PCR de génotypage des plantes de la lignée SALK-012425

a : amplification de l'allèle sauvage avec le couple LP012425/ RP012425 (~1000pb)  
 b : amplification de l'allèle mutant avec le couple LP012425/ Lba1 (~500pb et ~700pb)

1-8: plantes d'intérêts  
 A-B: témoin positif (ADN génomique d'une plante sauvage)  
 C: témoin négatif : eau

## **II.2 Phénotype macroscopique des plantes de génération T4 des lignées SALK-123655 et SALK-012425**

L'analyse phénotypique des mutants consiste en l'étude de plusieurs caractères comme le pourcentage de germination, la taille et l'aspect des germinations et des plantes adultes. Sur ces critères, pour les 2 lignées SALK, aucun phénotype macroscopique particulier lié à l'insertion de l'ADN-T n'a pu être observé pour les plantes de génération T4 étudiées.

## **II.3 Génotypage des plantes de génération T4 des lignées SALK-123655 et SALK-012425**

### *II.3.1 Principe du génotypage*

Pour chaque lignée SALK, nous utilisons un couple d'amorces LP/RP spécifiques de l'allèle sauvage et recouvrant la zone supposée d'insertion de l'ADN-T. L'ADN-T faisant plusieurs milliers de kilobases, aucune amplification n'est obtenue avec ces amorces à partir d'un allèle portant une insertion. Les amorces LB et LBa1 sont spécifiques de l'ADN-T. Combinées à l'amorce génomique LP, ces dernières donnent un produit d'amplification correspondant à l'allèle muté (figure n°45).

### *II.3.2 Résultats des réactions de PCR*

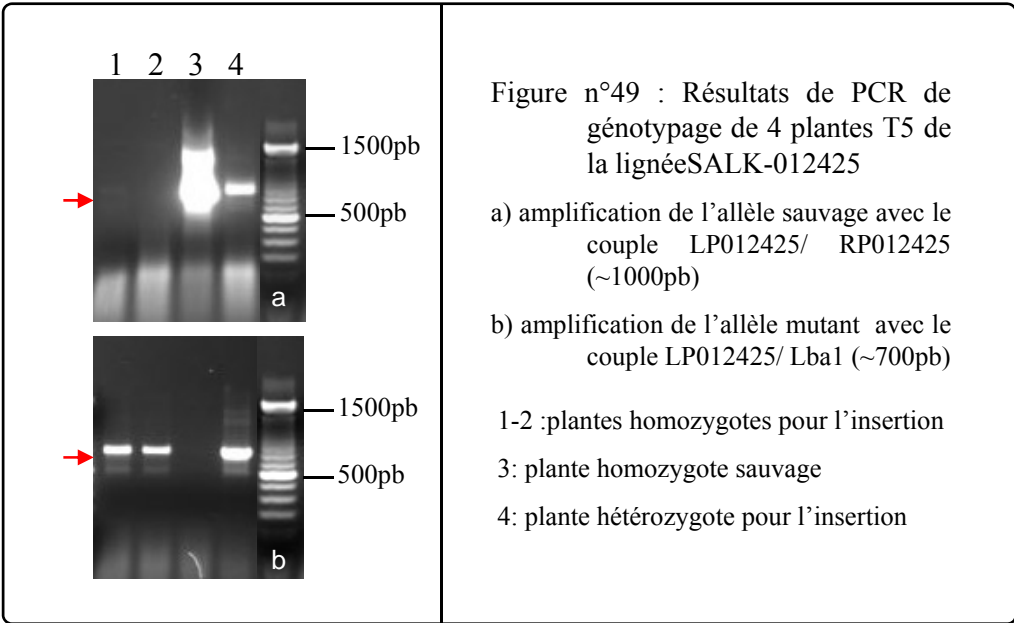
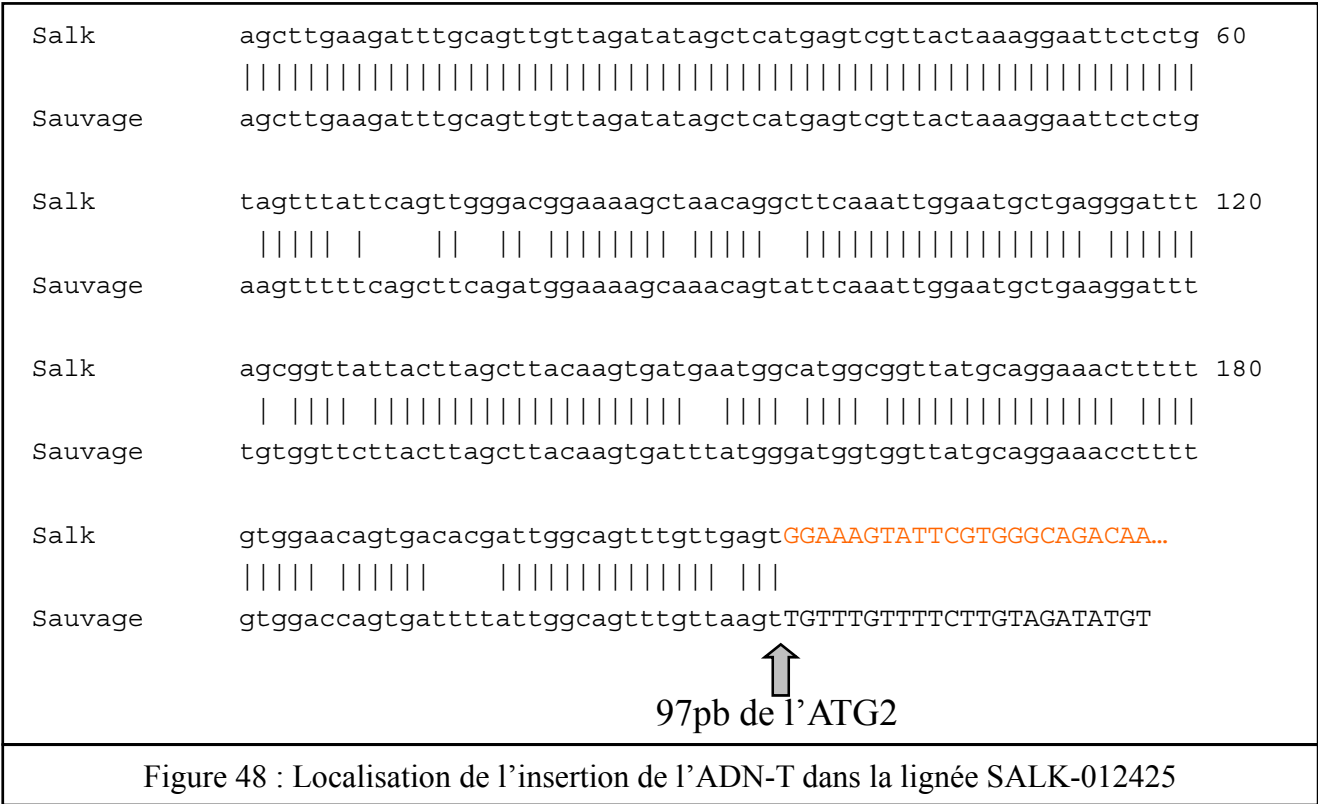
#### **Cas de la lignée SALK-123655**

La première réaction PCR utilisant le couple d'oligonucléotides LP-123655 et RP-123655, a permis de déterminer la présence ou non de l'allèle sauvage (figure n° 46a). La seconde réaction PCR, utilisant l'amorce Lb (spécifique de l'ADN-T) combinée à l'amorce LP-123655, a permis de déterminer la présence ou non de l'allèle muté (figure n°46b).

Ainsi, sur les 9 plantes génotypées, 6 étaient sauvages, 2 étaient hétérozygotes pour l'insertion et une plante était homozygote pour l'insertion. De plus, notons que le produit d'amplification correspondant à l'allèle muté a une taille estimée à 750 pb. De ce fait, nous pouvons conclure que l'ADN-T n'est pas inséré dans le gène *AtDYWI* mais à environ 150 pb en aval de celui-ci.

#### **Cas de la lignée SALK-012425**

L'emploi des oligonucléotides LP-012425/RP-012425 a permis de mettre en évidence pour les 8 plantes de génération T4, la présence de l'allèle sauvage (figure n°47a). Il n'y a donc pas de plantes homozygotes pour l'insertion. La seconde PCR, réalisée avec le couple d'amorces LP-012425/LBa1, a donné lieu à deux produits d'amplification (figure n° 47b). L'un, d'une taille estimée à 500 pb, a été amplifié à partir des extractions d'ADN génomique de toutes les plantes. Le second, amplifié à partir des extractions d'ADN génomique des plantes 4, 5 et 7 a une taille estimée à 700 pb.



### Séquençage des produits d'amplification obtenus pour l'allèle mutant

Nous avons séquencé ces deux produits d'amplification afin de savoir lequel permet de déterminer la présence de l'allèle mutant. Les séquences obtenues ont été utilisées pour réaliser un BLAST sur l'ensemble des données du NCBI. Les résultats indiquent que le produit de 500pb correspond à une amplification d'ADN plasmidique (contamination) et que le produit de 700 pb correspond à l'allèle muté. Ainsi, parmi les 8 plantes de la lignée SALK-012425 étudiées, 5 sont de génotype sauvage et 3 sont de génotype hétérozygote pour l'insertion. De plus, l'analyse de la séquence de l'allèle mutant nous a permis de localiser le point exact de l'insertion à 97 pb en amont de l'ATG du gène *DYWI* (figure n° 48).

### **II.4 Que signifie l'absence de plantes homozygotes en génération T4, dans la lignée SALK-012425 ?**

L'absence de plantes homozygotes pour l'insertion dans la lignée SALK-012425 à la génération T4 peut avoir deux explications.

- L'une, d'ordre statistique, suppose que l'échantillon de 8 plantes que nous avons étudié, n'est pas représentatif de l'ensemble de la population T4. Ainsi, la même étude réalisée sur un nombre plus important de plantes devrait permettre d'observer des plantes homozygotes pour l'insertion.
- L'autre explication, d'ordre biologique, est liée à l'insertion de l'ADN-T à 97 pb en amont de l'ATG du gène *AtDYWI*. Cette insertion pourrait affecter la région promotrice de ce gène, conduisant à une altération de son expression. Dans ce cas, l'absence de plantes homozygotes pour l'insertion pourrait être liée à une létalité à des stades précoces du développement de telles plantes.

### Analyse des plantes de cinquième génération (T5), dans la lignée SALK-012425

Nous avons voulu savoir s'il était possible d'obtenir de plantes homozygotes pour l'insertion caractérisée dans la lignée SALK-012425. Pour cela nous avons récolté les graines des trois plantes de la génération T4, hétérozygotes pour l'insertion. Trente graines issues de chacun de ces parents ont été semées, et nous avons, comme précédemment, analysé le phénotype macroscopique et le génotype de chaque plante individuellement.

#### Etude du phénotype macroscopique

Sur les mêmes critères que précédemment, aucun phénotype macroscopique particulier lié à l'insertion de l'ADN-T n'a pu être observé pour l'ensemble des plantes de génération T5.

#### Génotypage des plantes de cinquième génération

Le génotypage de ces plantes de cinquième génération a été réalisé par PCR, selon le même principe que pour la génération précédente. La présence de plantes homozygotes a été mise en évidence (figure n° 49).

## **II.5 Discussion sur l'absence de mutants d'insertion**

Cette étude nous a permis de montrer que les plantes de la lignée SALK-123655 portent un ADN-T inséré à 150pb en aval du codon stop du gène *AtDYWI*. Les plantes de la lignée SALK-012425 97pb ont, quant à elles, un ADN-T inséré à 97pb en amont de l'ATG du gène *DYWI*.

Ce résultat est lié au fait que la stratégie de génétique inverse mise en place au SALK Institute ne permet malheureusement pas de localiser précisément le point d'insertion de l'ADN-T dans le génome d'*Arabidopsis thaliana*. On estime qu'il est à une distance inférieure à 300 pb de la bordure séquencée mais il peut, comme dans notre cas, être situé plus loin et donc hors du gène d'intérêt.

Nous avons vérifié que ces deux insertions, à l'état hétérozygote et homozygote, ne conduisaient pas à un phénotype macroscopique. Nous aurions pu vérifier également si ces insertions affectées l'expression du gène *AtDYWI* à un niveau transcriptionnel (expérience de qRT-PCR) ou à un niveau traductionnel (expérience de western blot utilisant un anticorps anti-*AtDYWI*). Cependant, au moment de cette étude, je n'avais pas encore mis au point l'utilisation de tels outils au laboratoire. D'autre part, nous avons choisi de consacrer nos efforts à la recherche et à la caractérisation de mutants présentant de plus forts phénotypes macroscopiques.

Il est intéressant de noter que toutes collections confondues, il existe chez *Arabidopsis thaliana*, plus de 250 000 mutants d'insertion déjà caractérisés, et que parmi ceux-ci on trouve des insertions dans près 70% des ORF décrites. L'absence d'insertion dans les ORF restantes peut être liée soit à la petite taille de celles-ci, soit au fait qu'elles codent pour des protéines essentielles à la biologie de la plante. Dans notre cas, nous savons que le gène *AtDYWI* est de petite taille, ce qui pourrait expliquer l'absence d'insertion dans ce dernier. Compte-tenu de notre hypothèse de travail, l'éventualité d'une létalité au stade gamétophytique d'une insertion dans l'ORF *AtDYWI* est également envisageable.

## **III Lignées « ARNi »**

### **III.1 Le projet AGRIKOLA**

Si le gène *AtDYWI* est essentiel à la biologie de la plante, alors l'obtention de lignées mutantes présentant une extinction totale de son expression pourrait être impossible. Ainsi, j'ai consacré une partie de mes recherches de doctorat à établir et caractériser des lignées pour lesquelles l'expression du gène *AtDYWI* ne serait que partiellement diminuée, par la méthode dite de l'ARN « interférent » (« ARNi »). La caractérisation de lignées « ARNi » constitutives et inductibles a été menée en parallèle.

Débutée à l'URGV, cette étude a été grandement facilitée par le fait que notre équipe de recherche était une équipe partenaire du projet AGRIKOLA (pour *Arabidopsis* Genomic RNAi Knock-out Line Analysis : [www.agrikola.org](http://www.agrikola.org)). L'objectif de ce projet européen était d'établir et mettre à disposition de la communauté scientifique des lignées « ARNi » pour des milliers de gènes d'*Arabidopsis*. Notre équipe, ayant eu en charge l'aspect « clonage à haut débit » de ce projet, j'ai bénéficié d'une aide technique, méthodologique et matérielle



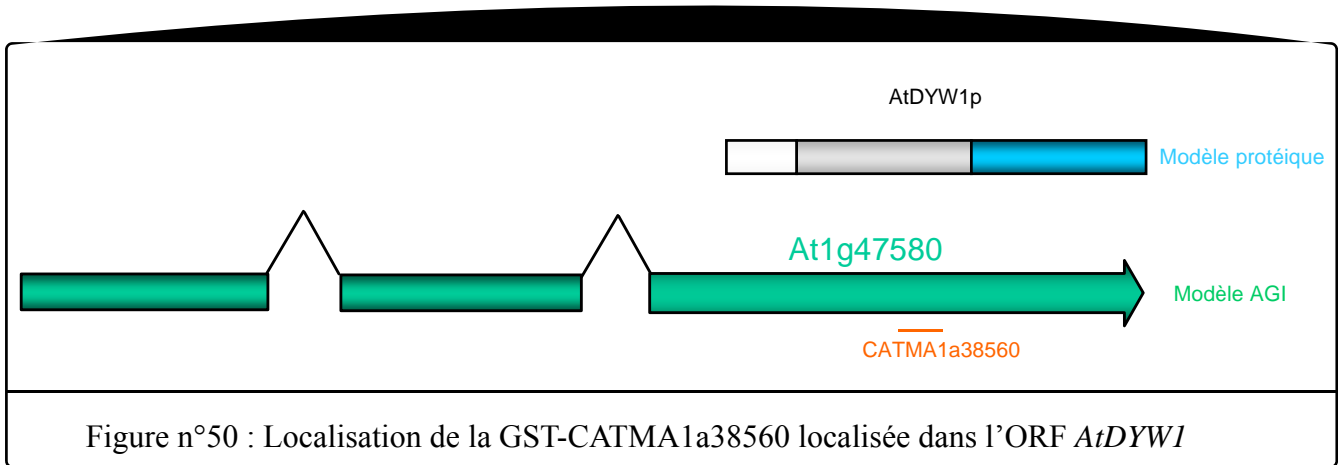


Figure n°50 : Localisation de la GST-CATMA1a38560 localisée dans l'ORF *AtDYW1*

précieuse pour l'établissement des différents types de lignées « ARNi » étudiées dans ma thèse.

Chez *Arabidopsis*, il a été montré que, l'expression d'un ARN en épingle à cheveu (ou hairpin RNA, hpRNA) composé d'un fragment double-brin spécifique du transcrite cible, et d'une boucle simple brin composée d'introns qui seront épissés de la molécule d'ARN mature, permet d'éteindre efficacement l'expression d'un gène de façon post-transcriptionnelle dans une plante transformée (Smith et al., 2000).

La banque de fragments CATMA (Crowe et al., 2003) regroupe plus de 20000 GSTs (pour gene-specific sequence tag) d'*Arabidopsis thaliana*. Ces petites séquences (de 100 à 500pb) ont été sélectionnées spécifiquement pour éviter des « cross-hybridation » avec d'autres gènes de cette plante lors d'expériences d'analyse du transcriptome. De ce fait, les GST du projet CATMA présentent des propriétés idéales pour générer un hpRNA spécifique d'un transcrite cible.

Ainsi, la GST-CATMA1a38560 est une séquence de 158pb, spécifique de notre gène d'intérêt *AtDYWI* (figure n°50). Cette GST a été clonée dans le vecteur d'entrée pDONR207 du système Gateway® et ce clone a servi de point de départ pour établir toutes les constructions utilisées dans les différents types de lignées « ARNi » établies (constitutives et inductibles) au cours de mon doctorat. En effet, les différents vecteurs d'expression *in planta*, mis à notre disposition dans le cadre du projet AGRIKOLA, ont la particularité de contenir deux cassettes de recombinaison Gateway. Ainsi, via une double réaction de LR clonase, la GST d'intérêt clonée dans le pDONR207 se retrouve clonée en « tête-bêche » dans le vecteur d'expression (Hilson et al., 2004).

### **III.2 Lignées « ARNi » constitutives**

#### *III.2.1 Etablissement des lignées « ARNi » constitutives*

##### Construction utilisée : le vecteur pAGRIKOLA

Le vecteur pAGRIKOLA permet l'expression, *in planta*, d'un hpRNA sous contrôle d'un promoteur constitutif fort (P35s CaMW) et du terminateur de l'octopine synthase (ocs). Les 2 cassettes de recombinaisons Gateway clonées « en tête-bêche » dans ce plasmide sont séparées par l'intron-2 du gène *Pdk* de *Flaveria*, et l'intron du gène *cat* (Helliwell and Waterhouse, 2003). Après une double recombinaison de LR clonase, la GST d'intérêt est insérée en lieu et place des cassettes Gateway du vecteur pAGRIKOLA.

##### Génotypage de la construction utilisée

Afin de vérifier que la GST d'intérêt a été intégrée dans le plasmide pAGRIKOLA en deux copies « tête-bêche », lors de la double recombinaison LR, on réalise un génotypage des colonies d'*E.coli*, puis *A. tumefaciens* transformées par notre construction.

Le principe de ce génotypage est décrit sur la figure n°51. Notons seulement ici qu'un mouvement dit de « flip-flop » des introns pouvant se produire au moment de la double recombinaison LR, on attend deux types de constructions correctes (sens 1 ou sens 2).

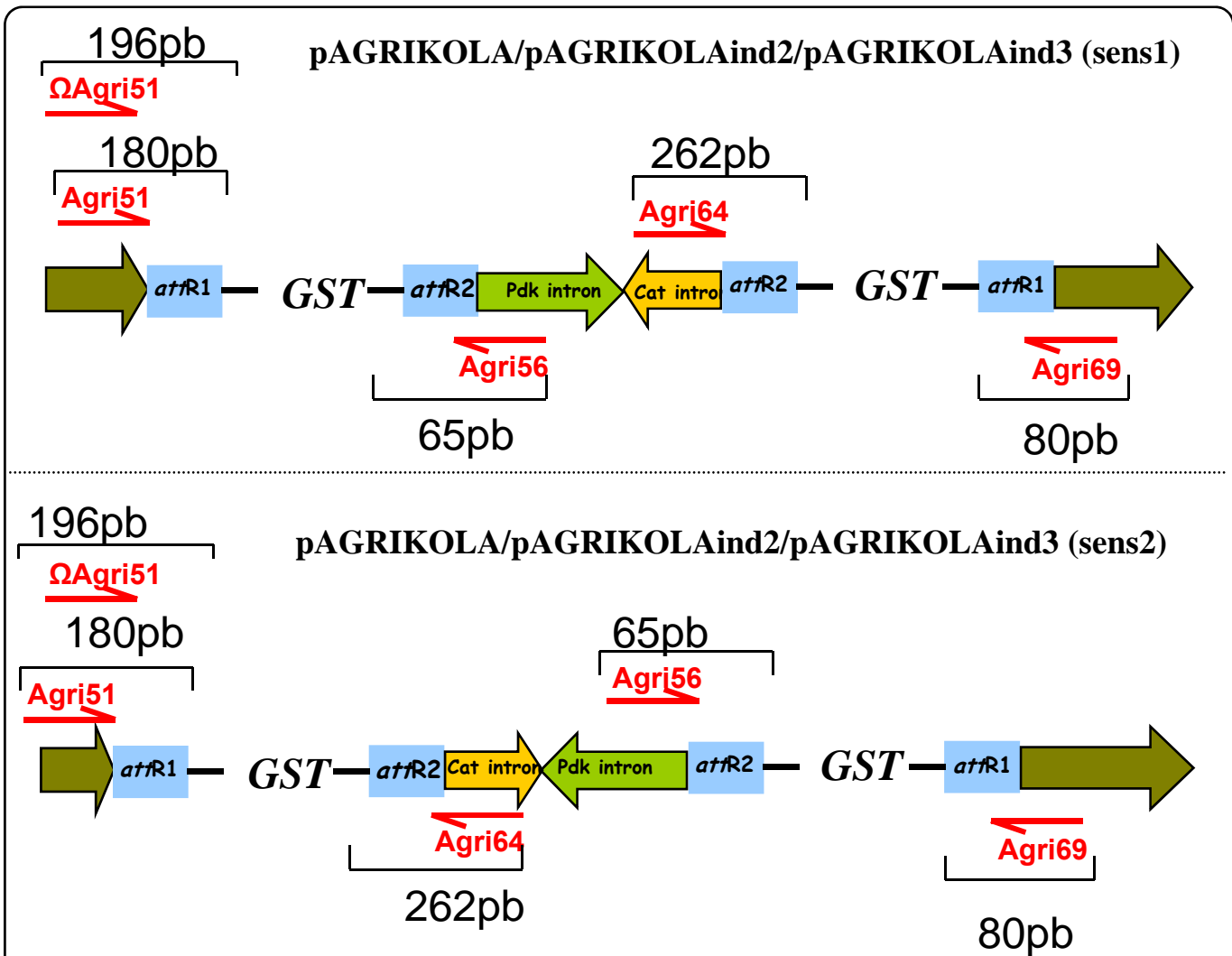


Figure n°51 : Principe du génotypage des constructions du projet AGRIKOLA

Les vecteurs pAGRIKOLA/ pAGRIKOLAind2/ pAGRIKOLAind3 ont été conçus afin qu'après une double recombinaison de LR clonase, la GST d'intérêt se retrouve clonée en deux copies « tête bêche » séparées par les introns *Cat* et *Pdk*. Un mouvement dit de « flip-flop » des introns peut se produire au cours de la double recombinaison LR, si bien qu'on attend deux types de constructions correctes (sens 1 ou sens 2). Un couple d'amorces de part et d'autre de chacune des 2 copies de la GST d'intérêt permet de vérifier l'intégralité de la construction.

Pour le vecteur pAGRIKOLA (constitutif):

dans une configuration de sens 1:

Le couple Agri51/Agri56 permet d'amplifier la première GST (produit PCR de 403pb)

Le couple Agri64/Agri69 permet d'amplifier la seconde GST (produit PCR de 500pb)

Dans une configuration de sens2:

Le couple Agri51/Agri64 permet d'amplifier la première GST (produit PCR de 600pb)

Le couple Agri56/Agri69 permet d'amplifier la seconde GST (produit PCR de 303pb)

Pour les vecteurs pAGRIKOLAind2/pAGRIKOLAind3 (inductibles):

dans une configuration de sens 1:

Le couple Agri51/Agri56 permet d'amplifier la première GST (produit PCR de 419pb)

Le couple Agri64/Agri69 permet d'amplifier la seconde GST (produit PCR de 500pb)

Dans une configuration de sens2:

Le couple Agri51/Agri64 permet d'amplifier la première GST (produit PCR de 616pb)

Le couple Agri56/Agri69 permet d'amplifier la seconde GST (produit PCR de 303pb)

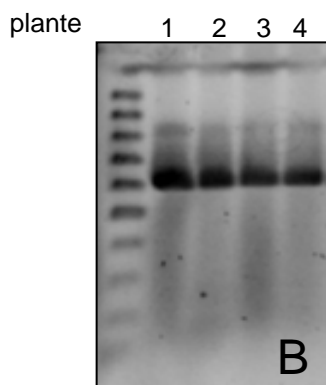
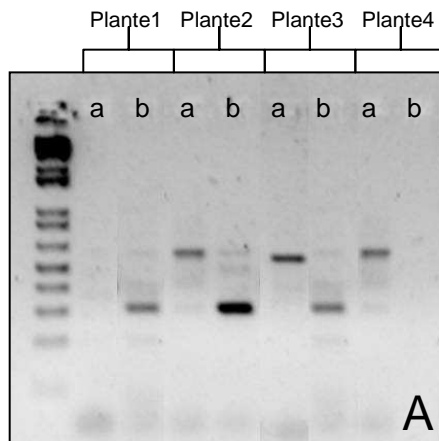


Figure n°52 : Résultats des réactions de PCR de génotypage de plantes T1 des lignées *AtDYWI* « ARNi » constitutif

A/ Détection de la présence des 2 GST

a : amplification de la première GST avec le couple Agri51/Agri64(600pb)

b : amplification de la deuxième GST avec le couple Agri56/Agri69 (303pb)

La plante 1 ne comporte que la deuxième GST

La plante 4 ne comporte que la première GST

Les plantes 2 et 3 comportent les 2 GST.

B/ Contrôle de la qualité des extractions d'ADNg

La qualité des extractions d'ADNg a été vérifiée par PCR à l'aide d'un couple d'amorces localisé dans le gène *AtDYWI*

Dans notre cas, une colonie d'*A. tumefaciens*, contenant une construction de sens 2, a été ainsi isolée et utilisée pour réaliser une transformation d'*Arabidopsis thaliana* (écotype WS).

### III.2.2 Sélection et génotypage des lignées

Après un traitement des cultures à l'herbicide « basta », 49 transformants (T1) indépendants ont été sélectionnés. Les graines de ces 49 plantes ont été récoltées séparément de façon à obtenir 49 lignées (T2) indépendantes.

Parallèlement, une feuille de chaque plante T1 a été prélevée et placée dans en plaque 96 puits, pour réaliser une extraction d'ADN génomique et les PCR de génotypage. La démarche utilisée pour le génotypage est identique à celle décrite sur figure n°51.

Après analyse des produits PCR obtenus, pour seulement 12 lignées, nous avons pu détecter la présence des 2 GST dans la configuration attendue (produits d'amplification à 600 et 303pb). Pour les autres lignées, nous avons : soit une amplification correspondant à l'une ou l'autre des 2 GST, soit une absence d'amplification pour les 2 GSTs (figure n°52).

### III.2.3 Caractérisation phénotypique des 12 lignées d'intérêt

Ainsi, je disposais de 12 lignées indépendantes, de génotype correct, et pouvant potentiellement présenter une diminution de l'expression du gène *AtDYWI*. Pour poursuivre mon analyse, j'ai mené une caractérisation phénotypique de ces lignées.

30 graines (T2) de chaque lignée ont été semées. Suivant les critères d'analyse phénotypique habituels (pourcentage de germination, la taille et l'aspect des germinations et des plantes adultes), aucun phénotype macroscopique particulier, n'a pu être observé pour ces 12 lignées.

### III.2.4 Bilan

Ainsi, la recherche de lignées *AtDYWI* « ARNi » constitutives, m'a permis d'isoler 12 lignées, pour lesquelles potentiellement l'expression du gène *AtDYWI* serait affectée. Aucun phénotype macroscopique n'a été observé pour l'ensemble de ces lignées.

Cela peut avoir deux significations biologiques :

- Soit, dans ces lignées, l'expression du gène *AtDYWI* est effectivement affectée post-transcriptionnellement, et cela n'a aucune incidence dramatique sur la vie de la plante (pas de phénotype macroscopique « visible »).
- Soit, les lignées sélectionnées possèdent un ADN-T inséré dans une région génomique transcriptionnellement peu ou non active. Ainsi il n'y aurait pas d'expression ou une expression insuffisante du hpRNA. La cascade de « réactions ARNi », conduisant à l'extinction du gène cible, ne serait donc pas « déclenchée ».

Une méthode efficace pour choisir entre ces deux hypothèses, consiste à analyser dans ces 12 lignées, l'expression de gène *AtDYWI*, à un niveau transcriptionnel ou à un niveau traductionnel. Cependant, nous avons choisi de ne pas développer ce type d'analyses sur les

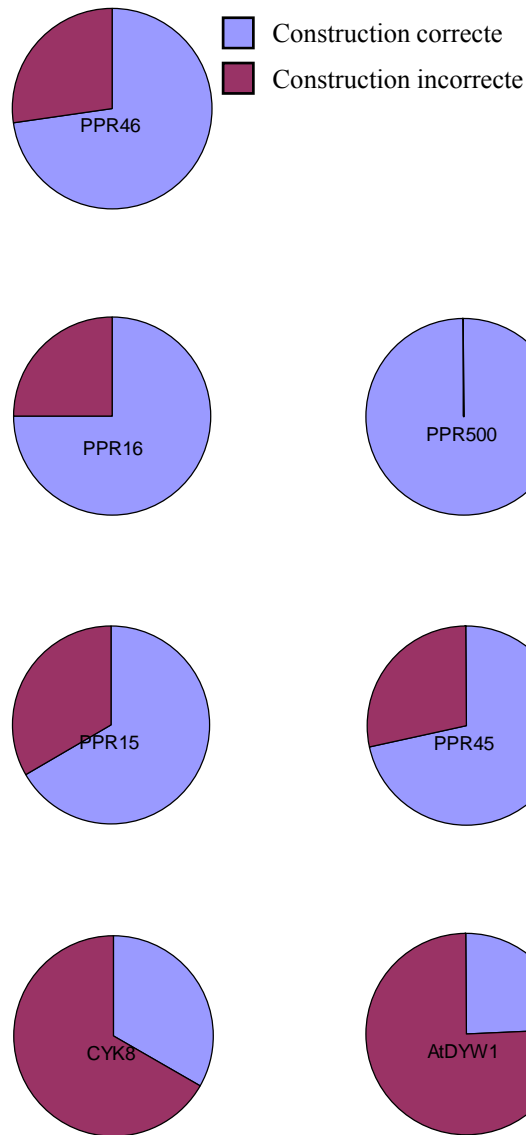


Figure n°53: Comparaison du nombre de plantes T1 portant des constructions « correctes » (présence des 2 GST) et « incorrectes » des différentes lignées « ARNi » constitutif produites dans le cadre du projet AGRIKOLA

12 lignées *AtDYWI* « ARNi » constitutives, car nous disposons d'éléments en faveur de la seconde hypothèse, suggérant qu'un « silencing » constitutif du gène *AtDYWI* ne pouvait pas (ou difficilement) être obtenu.

(1) Tout d'abord, dans une approche menée en parallèle, nous avons obtenu des lignées *AtDYWI* « ARNi » inductibles présentant des plantes au développement très affecté, pour lesquelles l'expression du gène *AtDYWI* est diminuée. Nous avons choisi de concentrer nos efforts sur ces lignées prometteuses (je détaille l'ensemble de ces résultats, ci-après, paragraphe III.3).

(2) De plus, remarquons que lors du génotypage des lignées *AtDYWI* « ARNi » constitutives, nous obtenons une très forte proportion (37/49) de lignées de génotype « incorrect » (absence de l'une ou l'autre des 2 GSTs, voir des deux). Ceci suggère une forte contre-sélection des « transgènes corrects » au moment de leur insertion dans le génome de la plante. Cette observation a-t-elle un sens biologique ?

Afin de répondre à cette interrogation, j'ai comparé mes résultats de génotypage à ceux obtenus pour d'autres lignées « ARNi » constitutives, construites et sélectionnées parallèlement aux lignées *AtDYWI*. Ces lignées, visant des gènes codant pour des protéines PPR d'intérêt pour le laboratoire, ont été génotypées et caractérisées, principalement par Andéol Falcon De Longevialle.

J'ai comparé, pour chaque gène d'intérêt, le nombre de transformants obtenus ayant une construction « pAGRIKOLA correcte » par rapport au nombre de transformants ayant une construction « pAGRIKOLA incorrecte » (figure n°53). Ce rapport est toujours supérieur à 1/2, excepté pour deux lignées *CYK8* et *AtDYWI*, pour lesquelles il est de l'ordre de 1/3 et 1/4, respectivement. De façon intéressante, pour le gène *CYK8*, nous possédons au laboratoire un mutant d'insertion létal à l'état homozygote. Par ailleurs, nous savons que, pour certaines autres lignées (telle que PPR46), une insertion à l'état homozygote est viable.

Ainsi, il semble difficile d'obtenir des lignées « ARNi » constitutives pour des gènes essentiels, tel que *CYK8*. Une forte contre sélection des cassettes de « silencing » intègres semble avoir lieu, probablement lors de leur insertion dans le génome de la plante afin de permettre sa survie. D'un point de vue du génotypage des transformants primaires de lignées « ARNi » constitutives, nous obtenons des résultats similaires pour le gène *CYK8* et *AtDYWI*. Cette similarité statistique n'a pas forcément de sens biologique. Cependant, comme pouvait déjà le suggérer l'absence de lignées d'insertion, le gène *AtDYWI* pourrait se comporter comme un gène dont l'extinction conduit à une létalité.

En conclusion, il apparaît qu'en essayant de produire des lignées *AtDYWI* « ARNi » constitutives, nous avons obtenu un grand nombre de transformants primaires de génotype « incorrect », présentant des remaniements dans la cassette de « silencing pAGRIKOLA ». D'autre part, une approche parallèle, nous a permis d'obtenir des plantes pour lesquelles l'expression du gène *AtDYWI* est diminuée, et qui présentent un phénotype macroscopique fort. Pris ensemble, ces deux observations, nous poussent à penser, bien que nous ne l'ayons pas vérifié, qu'il n'y a pas de « silencing » actif dans les 12 lignées *AtDYWI* « ARNi » constitutives obtenues.

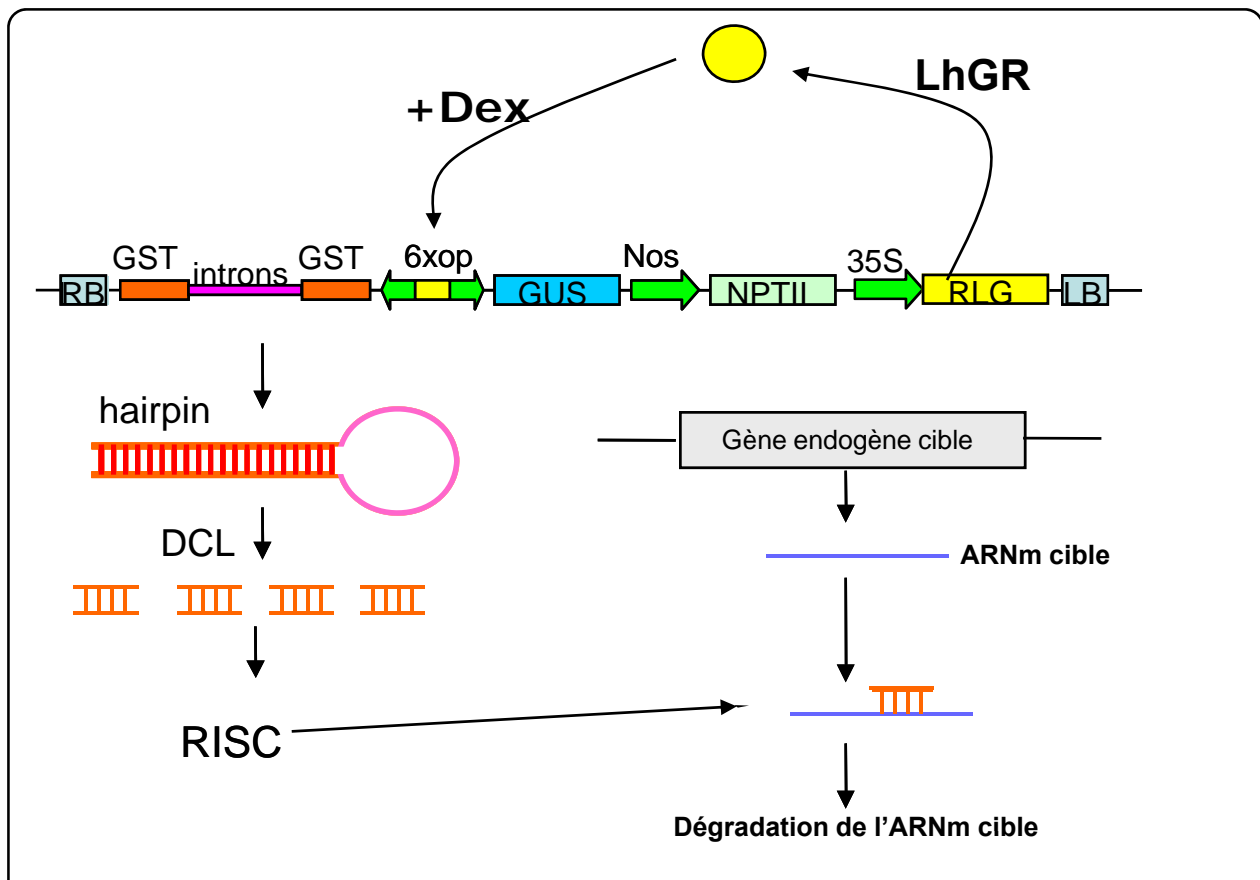


Figure n°54 : Représentation schématique de la construction utilisée pour produire les lignées *AtDYW1* « ARNi » inducibles de première génération.

La GST-CATMA1a38560 a été clonée, en utilisant la technologie Gateway, en deux copies « tête-bêche » séparées par des séquences introniques qui seront épissées du transcrite afin de produire un ARN en « épingle à cheveux » (hpRNA). Cet hpRNA déclenche une cascade de réaction conduisant à l'extinction de l'expression du gène *AtDYW1*.

Cette construction est sous le contrôle du promoteur inducible 6xop, bidirectionnel, il conduit aussi l'expression du gène rapporteur *GUS*. Les plantes transformées par cette construction expriment, sous contrôle du promoteur constitutif P35s-CaMW, le facteur de transcription LhG4 fusionné au domaine récepteur de la glucocorticoïde de rat. Cette protéine chimérique, après application de dexaméthasone, va reconnaître et activer la transcription du promoteur 6xop. Enfin, les plantes transgéniques portent, sous contrôle du promoteur constitutif pNos, le gène *NPTII* conférant une résistance à la kanamycine.



### **III.3 Lignées « ARNi » inductibles de première génération**

#### *III.3.1 Contexte de l'étude*

Comme évoqué précédemment, nous avons, parallèlement à l'obtention de lignées *AtDYWI* « ARNi » constitutives, débuté la construction et l'analyse de lignées *AtDYWI* « ARNi » inductibles. L'étude de ces lignées a été très approfondie et les résultats obtenus nous ont permis d'écrire un article que nous espérons soumettre prochainement. Dans ce souci de clarté et cohérence quant à la lecture de ce manuscrit, l'ensemble des étapes intermédiaires et résultats obtenus dans ce travail est présenté ici, et l'article correspondant fait suite (paragraphe III.5).

#### *III.3.2 Etablissement des lignées « ARNi » inductibles de première génération*

##### Construction utilisée

Nous avons utilisé le vecteur pAGRII, dérivé du vecteur pAGRIKOLA, et qui porte le système 6xpOp/LhGR (Craft et al., 2005). Ce système a 2 composants essentiels : un promoteur bidirectionnel (6xpOp) contenant 6 répétitions du domaine de fixation du facteur de transcription LhGR et une cassette permettant, sous contrôle du promoteur constitutif P35s CaMW, l'expression du facteur de transcription LhGR fusionné au récepteur de glucocorticoïdes du rat.

Ainsi, après une induction chimique par application de glucocorticoïde exogène (la dexaméthasone), le promoteur 6xpOp va conduire à l'expression, *in planta*, d'un hpRNA et du gène rapporteur *GUS*. De la même façon que pour le vecteur pAGRIKOLA, une double recombinaison de LR clonase va permettre le clonage de la GST d'intérêt « en tête-bêche » dans le vecteur pAGRII (figure n°54).

##### Génotypage

Le génotypage des bactéries, agrobactéries et plantes porteuses de la cassette pAGRII s'effectue selon une méthode très similaire à celle utilisée pour pAGRIKOLA, si ce n'est que l'amorce Agri51 est remplacée par l'amorce AgriQ51 se fixant sur le promoteur 6xpOp (figure n°51).

#### *III.3.3 Sélection, génotypage et phénotypage des transformants T1*

Après une sélection sur un milieu de culture *in vitro* (1/2MS) contenant de la kanamycine, 5 transformants (T1) indépendants ont été isolés.

Ces plantes ont été repiquées en terre et après croissance, une feuille de chaque plante T1 a été prélevée afin de réaliser une extraction d'ADN génomique ainsi que les PCR de génotypage. Après analyse des produits PCR obtenus, les cinq transformants T1 possèdent une construction pAGRII correcte, c'est-à-dire porteuse des 2 GST dans la configuration attendue.



T1 Plante 1 Plante 2 Plante3 Plante4 Plante5

Figure n°55 : Lignées *AtDYWI* « ARNi » inductibles : Morphologie des 5 transformants T1 obtenus

Plante 1 et 2: phénotype comparable à une plante sauvage

Plante 3: Croissance correcte avec une fertilité très réduite

Plante 4 et 5: Fort retard de croissance et fertilité très réduite

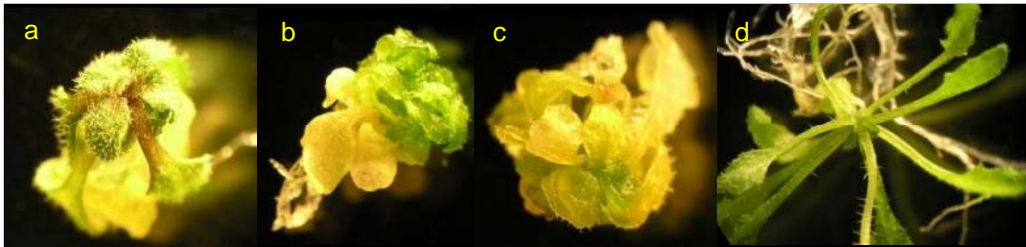


Figure n°56 : Lignées *AtDYWI* « ARNi » inductibles : Observation de la morphologie des «petits-choux » qui apparaissent dans les lignées T2 (sous loupe binoculaire) après 21 jours de culture *in vitro*

- a) Plantule affectée de la lignée 1
- b) Plantule affectée de la lignée 4
- c) Plantule affectée de la lignée 5
- d) Plantule sauvage Col0

**Lignée 1**

**Plantes sauvages**



Figure n°57 : Lignées *AtDYWI* « ARNi » inductibles: Morphologie des plantules « petits choux » qui survivent en terre après transfert.

Cette observation n'a été possible que pour quelques très rares plantules affectées de la lignée. Généralement, les plantules de phénotype « petits-choux » ne survivent pas , même après transfert.

Le phénotype de ces 5 plantes T1 a été examiné, révélant une apparence comparable à une plante sauvage pour les plantes 1 et 2, alors que les plantes 3 et 5 présentent une fertilité réduite, et que les plantes 4 et 5 présentent un fort retard de croissance (figure n°55).

Les graines de ces 5 plantes T1 ont été récoltées séparément de façon à obtenir 5 lignées indépendantes. Notons que très peu de graines ont pu être obtenues pour les lignées 3 et 5, et que certaines graines étaient de très petite taille, en particulier pour la lignée 3.

### III.3.4 Analyse phénotypique des lignées *AtDYW1* « ARNi » inductible à la génération T2

L'étude phénotypique de plantes issues de ces 5 lignées *AtDYW1* « ARNi » inductible a été réalisée après semis de graines sur milieu de culture *in vitro* (1/2MS) contenant ou non 10 $\mu$ M de dexaméthasone.

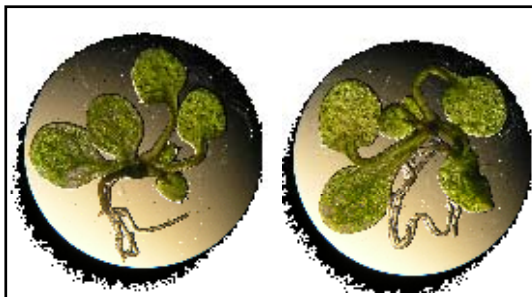
#### - Détection d'un phénotype dit « petits choux »

Pour la lignée 2, aucun phénotype n'a pu être mis en évidence, quelles que soient les conditions de croissance. En revanche, l'analyse phénotypique des lignées 1, 4 et 5 a révélé l'existence de plantules ayant un phénotype très particulier que nous avons surnommées « petits choux » ou « twisty plants » (figure n°56). Ce phénotype est complexe et se caractérise par plusieurs défauts morphologiques tels qu'une croissance très affectée, le recourbement des feuilles, et un jaunissement des plantules.

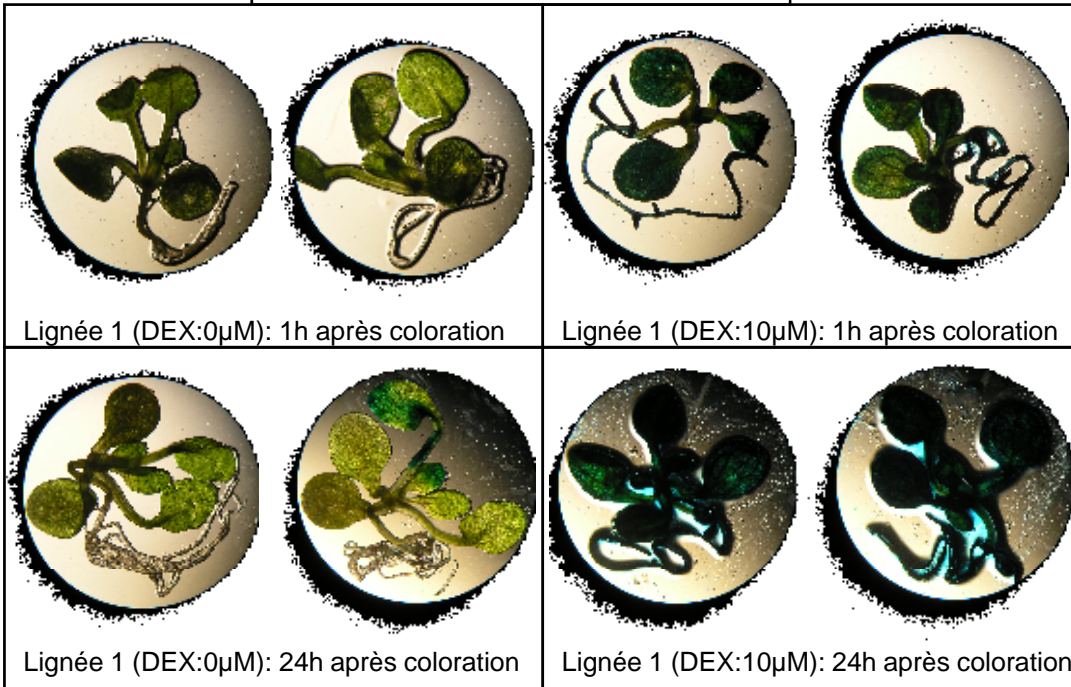
Ce phénotype apparaît progressivement après germination des graines et il est très net entre 20 et 28 jours après germination. Les plantules « petits choux » ne survivent généralement pas, même après un repiquage en sol. Cependant, j'ai obtenu un nombre très restreint de plantules de la lignée 1 ayant survécu après transfert et pour lesquelles un phénotype très fort persiste tout au long de la vie de telles plantes (figure n°57). Cette survie des plantules affectées après transfert en terre est très exceptionnelle et n'a pu être obtenue que pour une dizaine de plantes en 2 ans d'analyse. Par ailleurs, l'obtention de graines à partir de ces plantes s'avère difficile (très peu de graines sont obtenues, et par la suite, elles ne germent pas sur milieu de culture *in vitro*). Ainsi, peu d'analyses ont été menées sur les plantes transférées.

Pour les lignées 1, 4 et 5, le phénotype dit « petits-choux » est apparu sur des plantules ayant poussé sur un milieu contenant ou non de l'inducteur. Environ 10% des plantules sont ainsi affectées dans la lignée 1 et 30% dans la lignée 4. Pour ces deux lignées, ces observations ont pu être reproduites sur des graines de même génération, mais aussi sur des graines de génération T3 et T4. En revanche, la proportion de plantules présentant un phénotype « petits choux » dans la lignée 5 n'a pas pu être déterminée car nous disposions de peu de graines F2 et beaucoup d'entre elles n'ont pas germé que ce soit sur un milieu contenant ou non de la dexaméthasone. Ce phénomène (forte proportion de graines qui ne germent pas) s'est accentué en T3, limitant la caractérisation de cette lignée.

Pour la lignée 3, aucune des graines récoltées en T2, n'a germé sur milieu de culture *in vitro* contenant ou non de la dexaméthasone, rendant impossible l'étude de cette lignée.



Col0 (DEX:10μM): 24h après coloration



Lignée 1 (DEX:0μM): 1h après coloration

Lignée 1 (DEX:10μM): 1h après coloration

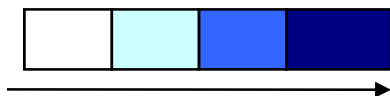
Lignée 1 (DEX:0μM): 24h après coloration

Lignée 1 (DEX:10μM): 24h après coloration

A

	Lignée 1		Lignée 2		Lignée 4		Col0	
	0μM	10μM	0μM	10μM	0μM	10μM	0μM	10μM
1h après coloration		■		■		■		
24h après coloration	■	■		■	■	■		

Intensité de coloration « GUS »



B

Figure n°58: Lignées *AtDYWI* « ARNi » inductibles : Détection qualitative de l'activité du gène *GUS* par coloration directe de plantules.

A/ Exemples de résultats obtenus pour des plantules de la lignée 1 ayant poussé sur un milieu (1/2 MS) contenant ou non 10μM de dexaméthasone.

B/ Table récapitulative des résultats obtenus pour l'ensemble des lignées *AtDYWI* « ARNi » inductibles

### - Bilan

Ainsi, la caractérisation des lignées *AtDYWI* « ARNi » inductible a révélé des plantules présentant un phénotype macroscopique marqué dit « petits choux » pour trois lignées. L'apparition d'un tel phénotype ne semble pas induite par la dexaméthasone, puisque nous obtenons une proportion similaire de plantes affectées quelles que soient les conditions de croissance (avec ou sans inducteur).

J'ai donc consacré une partie de mon travail à vérifier, d'une part, l'efficacité de l'induction dans nos expériences, et d'autre part, à déterminer si le phénotype « petits choux » pouvait être corrélé à une diminution d'expression du gène *AtDYWI*.

#### *III.3.5 Efficacité de l'induction via la dexaméthasone*

Comme nous l'avons déjà évoqué, le promoteur 6xpOp est bidirectionnel, ainsi après une induction à la dexaméthasone, il permet l'expression d'un hpRNA et du gène rapporteur *GUS* (codant pour la  $\beta$ -glucuronidase d'*E. coli*), dans les lignées *AtDYWI* « ARNi » inductibles. Afin de vérifier l'efficacité de l'induction dans nos expérimentations, nous avons étudié l'activité de ce gène rapporteur par des essais qualitatifs ou quantitatifs réalisés sur des plantules des lignées 1, 2 et 4. Nous avons également étudié l'expression du transcrite générant le hpRNA par des expériences de RT-PCR.

#### \* Détection qualitative de l'expression du gène rapporteur *GUS* par coloration directe de plantules

Nous avons prélevé une dizaine de plantules des lignées 1, 2 et 4, ainsi que des plantules sauvages (Col-0), ayant poussé 7 jours sur un milieu de culture *in vitro* (1/2MS) contenant ou non 10 $\mu$ M de dexaméthasone. Les plantules ont été mises en contact direct avec le substrat X-Gluc, et nous avons observé l'apparition d'une éventuelle coloration bleue après 1 et 24 heures. Les résultats obtenus sont présentés en figure n°58.

Une heure après coloration, nous observons que les plantules des 3 lignées « ARNi » inductibles ayant poussé sur un milieu de culture contenant de la dexaméthasone, ont une couleur bleue intense. Les plantules de ces lignées, provenant de boîtes ne contenant pas cet inducteur, ne présentent aucune coloration après une heure d'observation. Pour quelques plantules des lignées « ARNi » inductibles 1 et 4, une légère teinte bleue a pu être observée 24 heures après coloration. Enfin, dans toutes les conditions de cette étude, aucune coloration bleue n'a pu être détectée pour les plantules Col-0.

#### \* Quantification de l'activité du gène *GUS* par tests fluorimétriques

Ce test fluorimétrique est basé sur l'hydrolyse, par la  $\beta$ -glucuronidase, du substrat 4-MUG en un composé fluorescent à une longueur d'onde de 460nm. Nous avons ainsi mesuré les cinétiques de l'activité de la  $\beta$ -glucuronidase dans des extraits protéiques provenant de plantules des lignées « ARNi » inductibles 1 et 2, et des plantules sauvages (Col-0), ayant poussé dans différentes conditions.

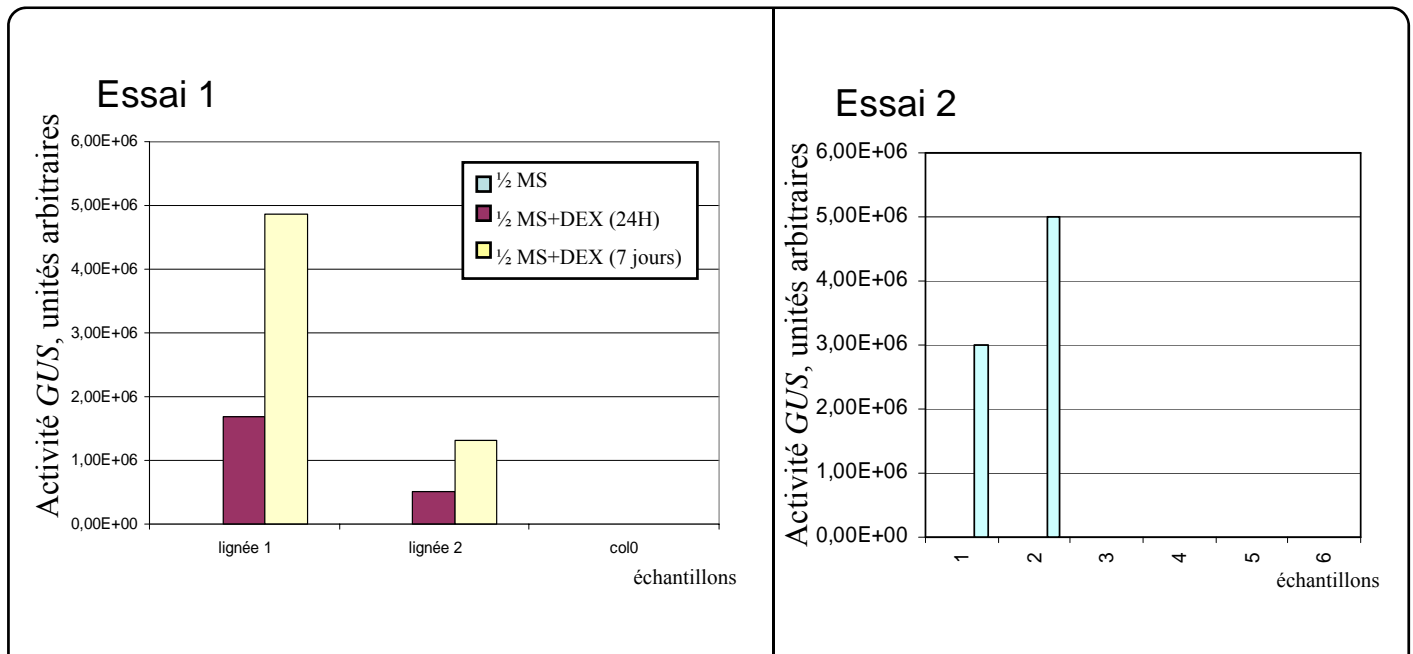


Figure n°59: Lignées *AtDYWI* « ARNi » inductibles: Quantification de l'activité du gène *GUS* par tests fluorimétriques

Les valeurs indiquées correspondent à la moyenne des valeurs obtenues pour 5 plantules dans chacune des conditions.

Essai 1: Comparaison pour des plantules des lignées 1, 2 et Col0 poussées 7 jours sur un milieu sans inducteur (1/2 MS) puis transférées ou non, 24 heures ou 7 jours sur un milieu (1/2 MS) contenant de 10µM de dexaméthasone.

Essai 2: Comparaison pour des plantules de la lignée 1 et Col0 poussées 21 jours sur un milieu de culture

- (1) avec inducteur (1/2 MS+DEX) et présentant un phénotype « petits choux » (lignée 1)
- (2) avec inducteur (1/2 MS +DEX) et ne présentant pas de phénotype (lignée 1)
- (3) sans inducteur (1/2 MS) et présentant un phénotype « petits choux » (lignée 1)
- (4) sans inducteur (1/2 MS) et ne présentant pas de phénotype (lignée 1)
- (5) avec inducteur (1/2 MS +DEX) et ne présentant pas de phénotype (Col0)
- (6) sans inducteur (1/2 MS) et ne présentant pas de phénotype (Col0)

Essai n°1 (figure n°59)

Dans un premier temps, nous avons utilisé des plantules des lignées « ARNi » inductibles 1 et 2, et sauvages (Col-0), ayant poussé 7 jours sur un milieu sans inducteur (1/2 MS) puis transférées ou non, 24 heures ou 7 jours sur un milieu (1/2 MS) contenant de 10µM de dexaméthasone.

Par cette méthode, nous n'avons détecté aucune activité du gène *GUS* dans les plantules sauvages (témoin négatif), et pour toutes les plantules des lignées 1 et 2, ayant poussé sur un milieu de culture ne contenant pas d'inducteur. En revanche, les plantules de ces lignées, ayant poussé 1 et 7 jours sur un milieu contenant de 10µM de dexaméthasone présentent une activité β-glucuronidase. Pour ces 2 lignées, cette activité est approximativement deux fois supérieure après 7 jours d'induction, qu'après 24 heures. Enfin, il apparaît que l'activité de la β-glucuronidase est 2 à 4 fois plus importante dans la lignée 1 que dans la lignée 2.

Essai n°2 (figure n°59)

Nous avons ensuite conduit le même type d'analyses sur des plantules de la lignée 1, ayant poussé 21 jours sur un milieu de culture avec ou sans inducteur, et présentant ou non un phénotype « petits-choux ». Comme précédemment, des plantules sauvages Col-0, ayant poussé dans les mêmes conditions, ont servi de témoin négatif dans ces expériences.

Aucune activité β-glucuronidase n'est détectée pour ces dernières, tout comme pour les plantules de la lignée 1 ayant poussé sur un milieu sans inducteur, quel que soit leur phénotype. En revanche, une activité β-glucuronidase a pu être mesurée pour toutes les plantules de la lignée 1 ayant poussé sur un milieu contenant de la dexaméthasone, indépendamment de leur phénotype. Notons que cette activité est comparable ( $\sim 5 \cdot 10^6$ ) à celle observée dans le test précédent réalisé sur des plantules de la lignée 1 à 7 jours d'induction sur un milieu contenant de la dexaméthasone.

\* Détection du transcrite correspondant au gène *GUS* et du transcrite générant le hpRNA

Via des expériences de RT-PCR, nous avons finalement évalué l'expression du gène rapporteur *GUS* au niveau transcriptionnel, et recherché la présence du transcrite générant le hpRNA. Ces analyses ont été réalisées à partir de plantules de la lignée 1, ayant poussé 21 jours sur un milieu de culture avec ou sans inducteur, et présentant ou non un phénotype « petits-choux ». Des plantules sauvages Col-0, ayant poussé dans les mêmes conditions, ont à nouveau servi de témoin négatif dans ces expériences.

Le transcrite correspondant au gène *GUS* ainsi que le transcrite générant le hpRNA, sont amplifiés chez toutes les plantes de la lignée 1 ayant poussé sur un milieu contenant de la dexaméthasone, indépendamment de leur phénotype (figures n°60 et n°61). En revanche, ils n'ont pu être détectés chez toutes les plantes de cette lignée ayant poussé sur un milieu sans inducteur et les plantes sauvages Col-0.



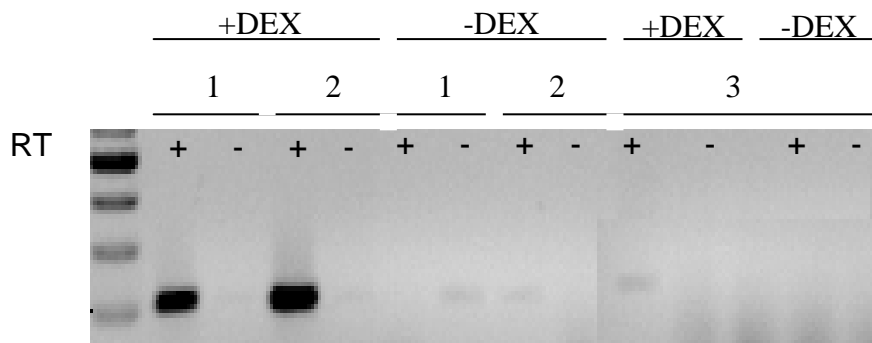


Figure n°60: Lignées *AtDYWI* « ARNi » inductibles : Détection du transcrit *GUS*, résultats de RT-PCR

- (1) ADNc provenant d'un pool de 5 plantules de phénotype « petits-choux » de la lignée 1, ayant poussé 21 jours sur un milieu contenant ou non de l'inducteur (DEX)
- (2) ADNc provenant d'un pool de 5 plantules sans phénotype de la lignée 1, ayant poussé 21 jours sur un milieu contenant ou non de l'inducteur (DEX)
- (3) ADNc provenant d'un pool de 5 plantules Col0 ayant poussé dans les mêmes conditions

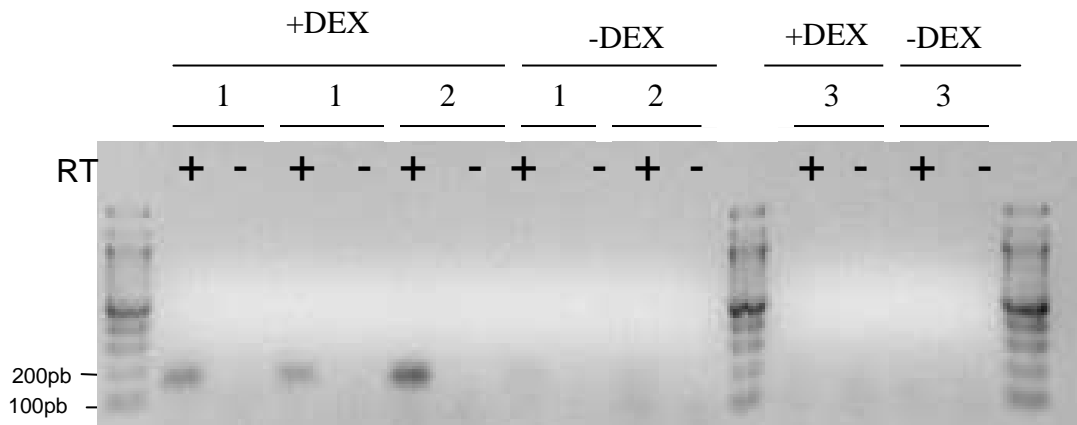


Figure n°61: Lignées *AtDYWI* « ARNi » inductibles : Détection du transcrit générant le hprRNA, résultats de RT-PCR

- (1) ADNc provenant d'une plantule de phénotype « petits-choux » de la lignée 1, ayant poussé 21 jours sur un milieu contenant ou non de l'inducteur (DEX)
- (2) ADNc provenant d'une plantule sans phénotype de la lignée 1, ayant poussé 21 jours sur un milieu contenant ou non de l'inducteur (DEX)
- (3) ADNc provenant d'un pool de 5 plantules Col0 ayant poussé dans les mêmes conditions

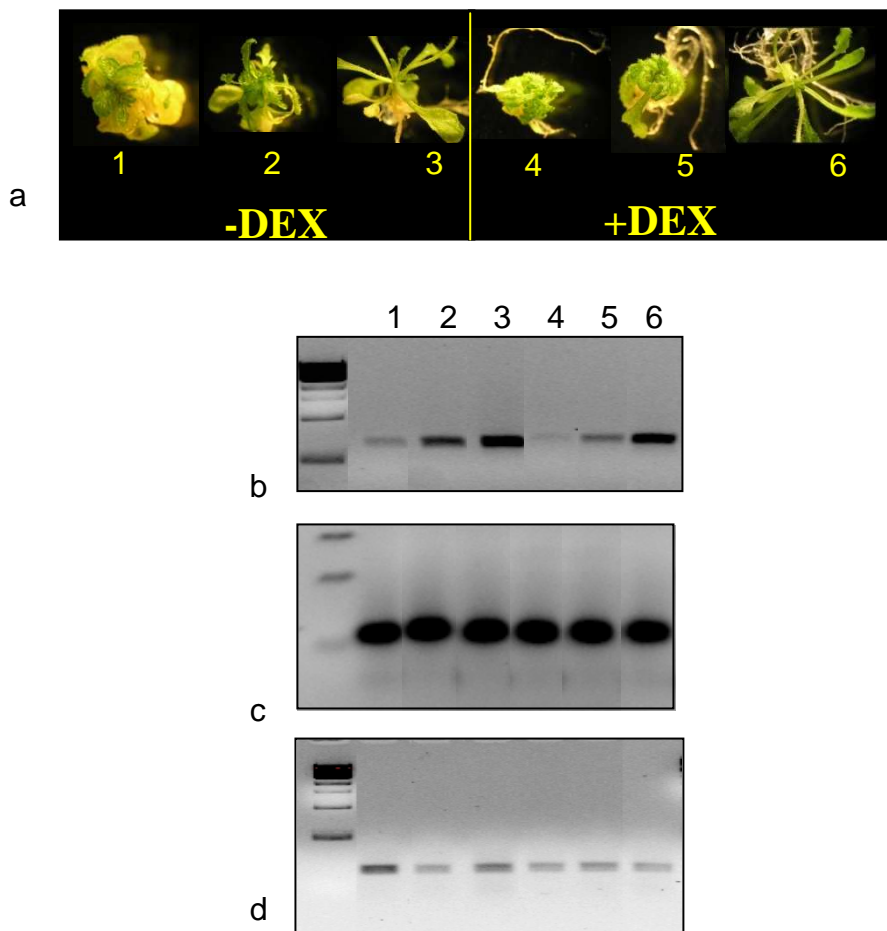


Figure n°62 : Lignées *AtDYWI* « ARNi » inductibles: résultats de RT-PCR

(a) Morphologie des plantules étudiées

- (1) Plantule très affectée ayant poussé 21 jours sur un milieu ne contenant pas d'inducteur (-DEX)
- (2) Plantule de phénotype intermédiaire ayant poussé 21 jours sur un milieu ne contenant pas d'inducteur (-DEX)
- (3) Plantule sans phénotype ayant poussé 21 jours sur un milieu ne contenant pas d'inducteur (-DEX)
- (4) Plantule très affectée ayant poussé 21 jours sur un milieu contenant de l'inducteur (+DEX)
- (5) Plantule de phénotype intermédiaire ayant poussé 21 jours sur un milieu contenant de l'inducteur (+DEX)
- (6) Plantule sans phénotype ayant poussé 21 jours sur un milieu contenant de l'inducteur (+DEX)

(b) Détection du transcrit *AtDYWI*

(c) Détection du transcrit *Actine*

(d) Détection du transcrit *PPR43*

### \* Bilan

Que ce soit par l'étude de l'expression du gène rapporteur *GUS* ou la détection du transcrit générant le hpRNA, nos résultats concordent et indiquent que l'activité transcriptionnelle du promoteur 6xpOp est bien induite par la dexaméthasone dans les lignées *AtDYWI* « ARNi » inductibles. De plus, par ces études, nous avons montré que l'efficacité de l'induction via la dexaméthasone varie en fonction de la lignée étudiée et semble dépendante du temps d'induction jusqu'à l'obtention d'un seuil maximal (atteint avant 7 jours et maintenu au moins pendant 21 jours d'induction).

Pour les lignées 1 et 4, dans nos expériences de coloration directe de plantules, nous avons noté, après 24h d'observation, l'apparition de plantules légèrement bleues provenant d'un milieu sans dexaméthasone. Ce résultat pourrait suggérer l'existence d'une expression basale (sans induction) du gène *GUS* dans ces lignées.

Enfin, les résultats obtenus nous permettent de confirmer que l'apparition des plantules avec un phénotype « petits-choux » dans ces lignées *AtDYWI* « ARNi » inductible, est indépendante de l'induction du promoteur 6xpOp via l'ajout de dexaméthasone dans le milieu de culture.

### *III.3.6 Niveau d'expression du gène AtDYWI dans les lignées « ARNi » inductibles*

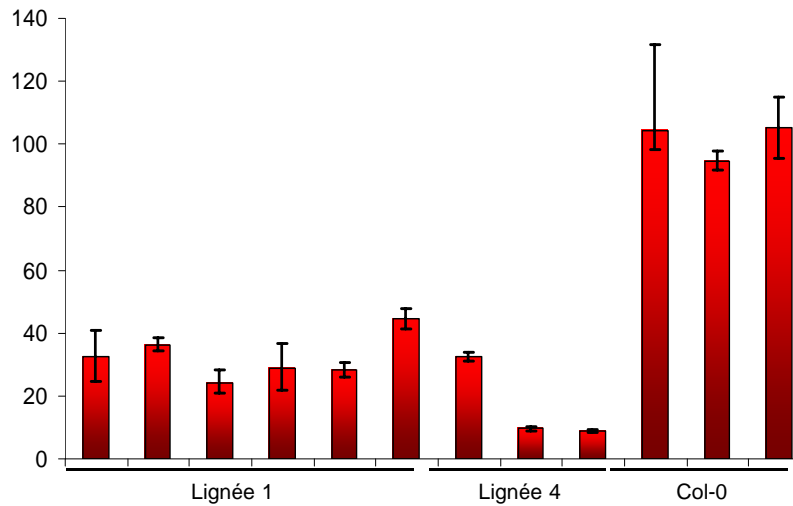
Dans les lignées *AtDYWI* « ARNi » inductibles, nous observons, pour au moins trois lignées, l'apparition de plantules au développement très affecté. Ce phénotype dit « petits-choux » n'est pas inductible chimiquement. Est-il lié à une modification d'expression du gène *AtDYWI* ?

### \* Détection du transcrit correspondant au gène *AtDYWI* par RT-PCR

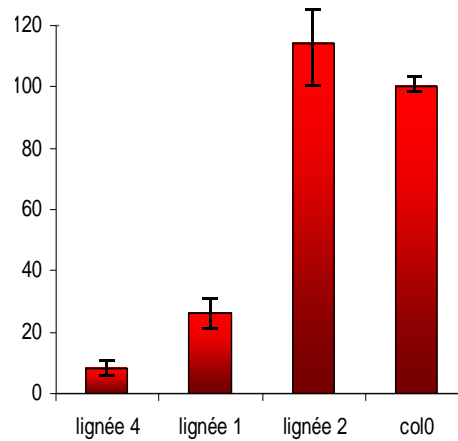
Via des expériences de RT-PCR, nous avons évalué l'expression du gène *AtDYWI* au niveau transcriptionnel, dans des plantules de la lignée 1, 2 et 4, ayant poussé 21 jours sur un milieu de culture avec ou sans inducteur, et présentant ou non un phénotype « petits-choux ». Rappelons qu'il n'existe pas de plantules de la lignée 2 au développement affecté.

En utilisant un couple d'amorces localisé dans le gène *AtDYWI*, nous avons obtenu, quelque soit les conditions de culture, une amplification plus faible de ce transcrit dans les plantules affectées de la lignée 1 et 4 (figure n°62). Par ailleurs, les plantules non affectées des lignées 1, 2 et 4, ont une quantité de transcrit *AtDYWI* comparable à celle de plantules sauvages poussées dans les mêmes conditions.

Une mauvaise qualité des préparations d'ARN ou d'ADNc provenant des plantules affectées pourrait expliquer une telle observation. Aussi, nous avons étudié, via la même méthode, l'expression des gènes constitutifs de l'actine (*At3g18780* et *At1g49240*) et l'expression d'un autre gène codant pour une protéine PPR (*PPR43*) étudiée à l'URGV. Une amplification similaire de ces transcrits a été observée quelque soit l'origine des échantillons (plantes affectées ou non) (figure n°62).



A



B

Figure n°63: Lignées *AtDYW1* « ARNi » inductibles : résultats de PCR quantitative

(A) Quantification de l'abondance du transcrite *AtDYW1* dans des plantules affectées des lignées 1, 4 et dans des plantules sauvages

(B) Quantification de l'abondance du transcrite *AtDYW1* dans des plantules affectées des lignées 1, 4 et dans des plantules sans phénotype de la lignée 2 et Col0. (Les résultats ici correspondent à la moyenne des résultats obtenus pour deux plantules de chaque lignée).

Les données brutes ont été normalisées en utilisant l'expression des gènes constitutifs de l'actine (AT3G18780 et AT1G49240) et les ratios sont calculés en prenant une plante sauvage Col0 comme référence

Ainsi, il semble que la quantité de transcrit correspondant au gène *AtDYWI* soit réduite spécifiquement dans les plantules au phénotype « petits-choux » des lignées *AtDYWI* « ARNi » inductibles.

#### \* Expériences de PCR quantitatives

Afin de confirmer et évaluer la diminution de l'expression du gène *AtDYWI* dans les plantules affectées des lignées « ARNi » inductibles 1 et 4, une approche par PCR quantitative a été développée.

L'abondance des transcrits correspondants au gène *AtDYWI* et de deux gènes constitutifs de l'actine (*At3g18780* et *At1g49240*) a été mesurée, dans des plantules affectées des lignées 1 et 4 et des plantules sauvages Col-0 poussées dans les mêmes conditions (1/2GB). Une déviation standard a été calculée à partir de 3 répétitions techniques et une normalisation est appliquée en utilisant les données d'expression des gènes de l'actine (expression constitutive). Les résultats indiquent une très forte diminution de la quantité de transcrit *AtDYWI* (de 50% à 90%) dans les plantules affectées comparées aux plantules sauvages (figure n°63A). De façon intéressante, notons que cette diminution semble plus importante pour certaines plantules de la lignée 4 (présentant moins de 10% de la quantité « normale » de transcrit *AtDYWI*), que pour les plantules de la lignée 1 (présentant toujours plus de 24% de la quantité « normale » de ce même transcrit). Nous avons également mesuré, par PCR quantitative, l'abondance du transcrit *AtDYWI* dans des plantules de la lignée 2. Ces plantules présentent une quantité de transcrit *AtDYWI* comparable à celles de plantules sauvages Col-0 poussées dans les mêmes conditions (1/2GB) (figure n°63B).

#### \* Expériences de western blot utilisant un anticorps dirigé contre la protéine AtDYWI

Enfin, nous avons évalué la quantité de protéine AtDYWI présente dans les plantules affectées des lignées 1 et 4 par des expériences de western blot.

J'ai réalisé une extraction de protéines totales à partir de plantules de phénotype « petits-choux » des lignées 1 et 4 et de plantules sauvage Col-0 poussées dans les mêmes conditions (1/2 MS). La quantité de protéines de ces extractions a été évaluée sur un mini-gel dénaturant de protéines coloré en bleu de Coomassie (figure n°4, dans le manuscrit présenté au paragraphe III.5). Un western blot a été réalisé sur ces échantillons en utilisant l'anticorps (AC3) dirigé contre la protéine AtDYWI. Une faible bande de taille attendue (~25kDa) est obtenue pour les plantes sauvages alors qu'aucune bande ne peut être observée, à notre seuil de détection, pour les plantes mutantes au phénotype « petits-choux » (voir figure n°4, dans le manuscrit présenté au paragraphe III.5).

#### \* Bilan et interprétation des résultats

Cette étude nous a permis d'isoler 5 lignées *AtDYWI* « ARNi » inductibles. Deux d'entre-elles ont montré un phénotype très affecté dès la première génération et n'ont pu être étudiée de façon approfondie (lignée 3 et 5). L'une d'entre-elle n'a pas montré de phénotype particulier, ni de défaut d'expression du gène *AtDYWI* (lignée 2). En revanche, l'étude des lignées 1 et 4, a révélé l'existence de plantules de phénotype dit « petits choux », et pour lesquelles l'expression du gène *AtDYWI* est affectée. En effet, des analyses de PCR quantitative ont permis de montrer une diminution (de 50 à 90%) de la quantité de transcrit *AtDYWI* dans les plantules au développement affecté. De même, par des expériences de western blot, nous n'avons pas pu mettre en évidence la présence de la protéine AtDYWI dans ces plantules. Cependant, nous savons que notre anticorps est peu sensible (voir aussi

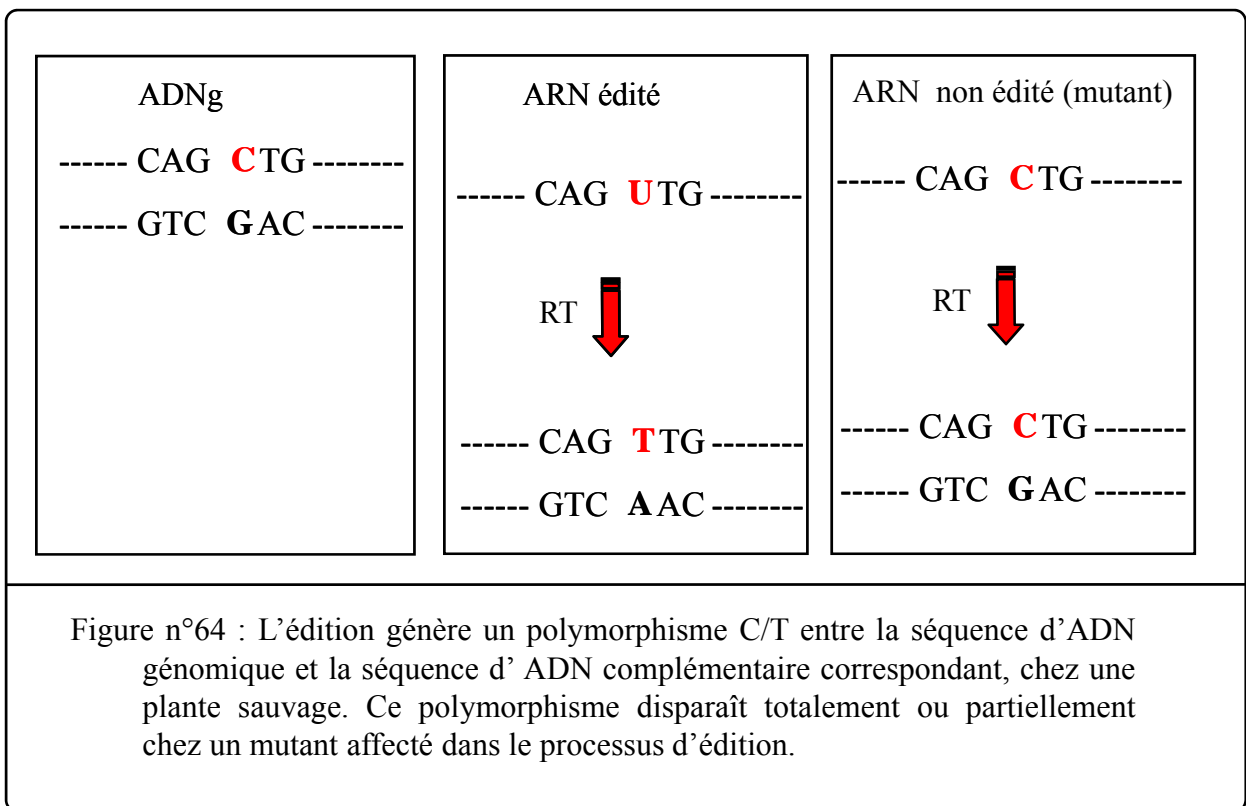
partie 4 paragraphe II.3), et par conséquent, il est envisageable qu'à notre seuil de détection, nous ne puissions pas observer l'éventuelle présence de la protéine *AtDYWI* dans ces plantules, car sa quantité pourrait être plus faible que la quantité présente dans les plantules sauvages (déjà tout juste détectable).

De façon intéressante, nous avons observé une plus grande proportion de plantules affectées dans la lignée 4 que la lignée 1, et d'après nos expériences de PCR quantitative, il semble que les plantules de la lignée 4 expriment généralement moins de transcrite *AtDYWI* que les plantules de la lignée 1. Pris ensemble ces résultats suggèrent l'existence d'une corrélation entre l'expression du gène d'intérêt et le phénotype dit « petit-choux » observé dans ces lignées.

De façon surprenante, nous avons vu que dans ces lignées, l'extinction du gène *AtDYWI* et le phénotype « petit-choux » apparaissaient dans 10 à 30% des plantes, de façon indépendante à une induction à la dexaméthasone. Nous avons cependant vérifié l'efficacité d'induction du promoteur 6xpOp *via* l'ajout de dexaméthasone dans le milieu de culture, lors de nos expérimentations. Ainsi, il s'avère que ni l'expression du gène rapporteur *GUS*, ni la présence du transcrite générant le hpRNA, ne témoignent d'une éventuelle extinction post-transcriptionnelle de l'expression du gène *AtDYWI* dans nos lignées. On se demande alors comment expliquer l'apparition de ces plantules au phénotype marqué et affectées dans l'expression du gène *AtDYWI* ?

Une légère expression basale du promoteur 6xpOp en absence d'inducteur (phénomène de fuite) pourrait expliquer l'apparition de plantes aux phénotypes affectés sur un milieu ne contenant pas de dexaméthasone. Ce phénomène de fuite pour le promoteur inductible 6xpOp a déjà été rapporté (Ian Moore, communication personnelle). De plus, dans les lignées 1 et 4, nos expériences de colorations directes de plantules afin d'examiner l'expression du gène *GUS*, avaient déjà suggéré l'existence d'une expression basale (sans induction) de ce gène. Il est possible que nos autres méthodes d'analyses (test fluorimétrique et RT-PCR pour détecter le transcrite *GUS* et le transcrite générant le hpRNA) ne soit pas assez sensibles pour mettre en évidence un tel phénomène. Toutefois, même si l'existence d'une fuite pourrait expliquer l'apparition de plantes aux phénotypes affectés sur un milieu ne contenant pas de dexaméthasone, elle ne nous permet pas de comprendre pourquoi l'ajout de dexaméthasone et la production importante de transcrite générant le hpRNA n'augmente pas la proportion de plantes de phénotype « petits-choux » observés.

L'absence de corrélation entre la présence du transcrite générant le hpRNA et l'extinction du gène *AtDYWI* dans les plantules ayant poussé sur un milieu contenant de la dexaméthasone suggère qu'une des étapes de la cascade de réactions conduisant au « silencing » du gène *AtDYWI* soit « bloquée » dans les plantes sans phénotype. Dans les plantules de phénotype « petit-choux », l'extinction de l'expression du gène *AtDYWI* se déclenche par une « voie inconnue » liée à la présence de la cassette transgénique mais qui semble indépendante de la présence du hpRNA. Le fait que nous ne puissions pas expliquer moléculairement comment l'extinction du gène *AtDYWI* est activée dans ces lignées « ARNi » inductibles est sans doute la critique majeure qui peut être faite quant à la caractérisation de ces lignées. C'est pourquoi, j'ai continué tout au long de ma thèse (comme je le décrirai dans les paragraphes suivants) à chercher à obtenir d'autres types de lignées mutantes dont la caractérisation pourrait apporter des résultats complémentaires.



Une autre objection pourrait être formulée quant au choix de la méthode « ARNi » pour éteindre l'expression du gène *AtDYW1*, membre d'une famille multigénique. En effet, nous avons vu qu'il existe, chez *Arabidopsis thaliana*, près de 90 gènes codant des protéines PPR de type DYW et que la séquence de ce domaine est relativement bien conservée d'un membre à l'autre. Comment s'assurer que l'expression du hpRNA dans les lignées *AtDYW1* « ARNi » inductibles n'affecte pas l'expression d'autres membres de la famille PPR ? Comme nous l'avons déjà évoqué, la GST-CATMA1a38560, générant le hpRNA, est une séquence de 158pb spécifique de notre gène d'intérêt *AtDYW1*. Nous avons vérifié par une analyse de blast que cette séquence ne présente pas d'homologie (même sur une dizaine de paire de bases) avec d'autre membre de la famille PPR. Cependant, une homologie parfaite de 16 paires de bases avec la séquence de l'ORF At4g19990 codant pour une protéine ayant un domaine de liaison au zinc et similaire à la protéine FHY3 (pour FAR-RED elongated hypocotyls 3) a été ainsi mise en évidence.

A ce stade de mon doctorat, je disposais donc de 2 lignées « ARNi » inductibles indépendantes pour lesquelles des plantules, présentant une expression du gène *AtDYW1* réduite et un phénotype très affecté, apparaissent. Etant donné notre hypothèse de travail suggérant que la protéine *AtDYW1* pourrait être essentielle au processus d'édition des ARN des organites ; j'ai consacré une partie de cette étude à l'analyse du niveau d'édition dans les plantules affectées de ces deux lignées *AtDYW1* « ARNi » inductibles.

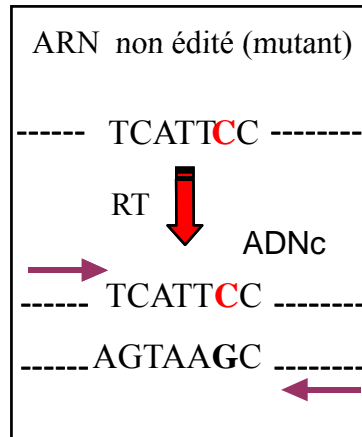
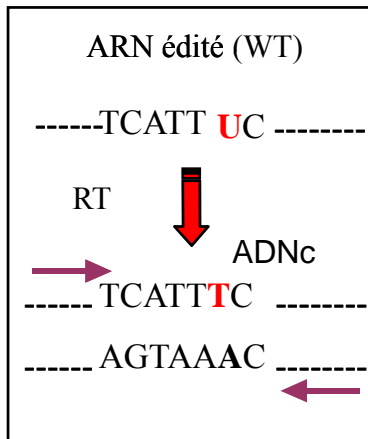
**Nota bene :** Comme nous venons d'en discuter, le « silencing » du gène *AtDYW1* n'est pas inductible à la dexaméthasone dans ces lignées. Cependant, dans un souci de clarté, ces lignées restent nommées *AtDYW1* « ARNi » inductibles dans la suite du manuscrit.

### III.3.7 Recherche de défauts d'édition dans les lignées « ARNi » inductibles 1 et 4

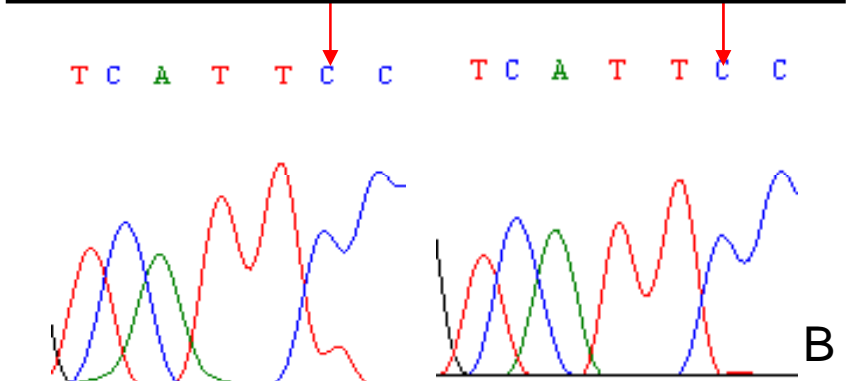
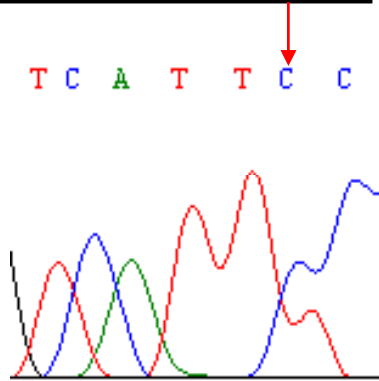
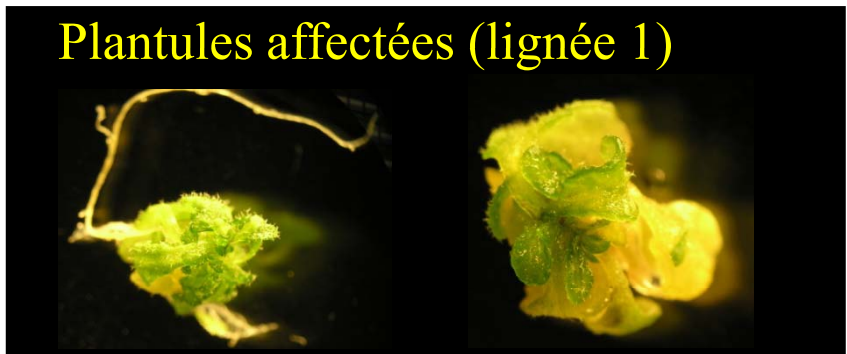
Selon notre modèle de travail, la protéine *AtDYW1* pourrait être impliquée dans l'édition d'un certain nombre de sites d'édition chloroplastiques et mitochondriaux (si elle est recrutée uniquement par des protéines PPR de type E/E+), voir tous les sites dans l'hypothèse où elle agirait sous forme d'un dimère ou multimère. Par ailleurs, l'extinction du gène *AtDYW1* dans les plantules affectées n'étant que partielle, les défauts d'édition observés pourraient également n'être que partiels. Pour ces deux raisons, il est impossible de prédire *a priori*, les sites d'édition qui pourraient être affectés dans nos plantes d'intérêts. C'est pourquoi, afin de rechercher d'éventuels défauts d'édition, nous avons choisi, dans un premier temps, d'utiliser des approches systématiques sur un grand nombre de sites.

Cependant, nous avons vu en introduction qu'il existe chez *A. thaliana*, respectivement 34 et 441 sites d'édition décrits dans les chloroplastes (Chateigner-Boutin and Small, 2007) et les mitochondries (Giege and Brennicke, 1999). De plus, ces chiffres sous-estiment sans doute la réalité car de nouveaux sites peuvent être découverts notamment dans les régions non traduites. La répartition de ces sites, en particulier dans les transcrits mitochondriaux, est très inégale, si bien qu'il existe des régions à forte densité en sites édités (où ils peuvent être distants d'à peine quelques paires de bases, voir être directement consécutifs). Nous avons également évoqué que si pour certains sites l'édition est un processus efficace (100% des transcrits synthétisés sont édités), pour d'autres sites, en particulier dans les mitochondries, l'édition peut être partielle. Enfin, le niveau d'édition est variable selon les tissus étudiés, le stade de développement de la plante, et les conditions environnementales.





A



B

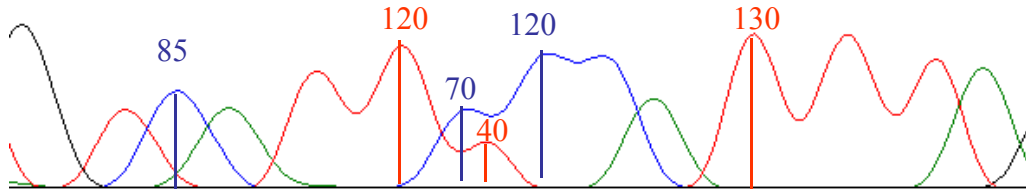
Figure n°65 : Crible par séquençage direct de produits PCR amplifiés à partir d'ADN complémentaires

A/ Représentation schématique du principe de ce crible

B/ Exemples de résultats obtenus, pour une plantule sauvage et deux plantules affectées de la lignée 1, au site d'édition C(30599) dans le transcrit mitochondrial *ccb206*

Plantule sauvage : transcrit mitochondrial *ccb206* site d'édition (position 30599)

G T C A T T C C C A T T T A  
 C-1 C C+1  
 T-1 T T+1



Plantule affectée: transcrit mitochondrial *ccb206* site d'édition (position 30599)

G T C A T T C C C A T T T A  
 C-1 C C+1  
 T-1 T T+1

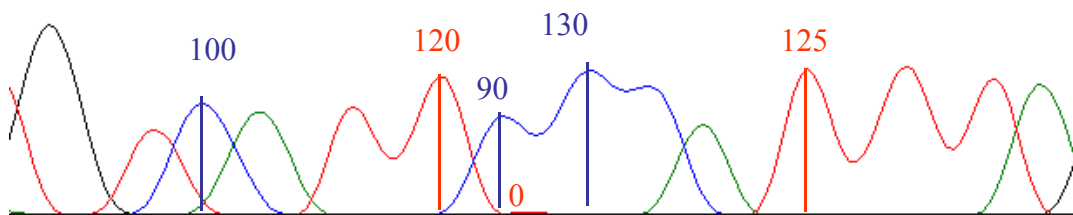


Figure n°66 : Approximation du niveau d'édition par une méthode de mesure de « maximum » sur chromatographe

Le maximum (pic) du site édité (cytidine: notée C) est mesuré, ainsi que le maximum de la cytidine précédente (notée C-1) et suivante (notée C+1)

Les rapports  $R1 = C-1 / C+1$  ( en théorie égal à 1) et  $R2 = C / C+1$  (en théorie inf. à 1) sont déterminés. Le rapport R1 doit être identique chez un mutant et un sauvage, alors qu'en cas de défaut d'édition le rapport R2 doit augmenter chez le mutant.

Les mesures et rapport « en miroir » sont également effectué en mesurant les maximums de Thymidines correspondante (notée T, T-1 et T+1).

Table n°5: Résultats du crible de recherche de défauts d'édition par séquençage direct de produits PCR amplifiés à partir d'ADNc de plantules affectées des lignées *AtDYWI* « ARNi » inductibles

Locus (Transcrit)	amorce	position génomique	Edition
AtCg00040 (matK)	Chloro 260	2931	NA
AtCg00130 (atpF)	Chloro 27	12707	NA
AtCg00170 (rpoB)	23581F-24093R	23898	A
AtCg00170 (rpoB)	Chloro 45	25779	PA
AtCg00170 (rpoB)	Chloro 45	25992	PA
AtCg00500 (accD)	Chloro 112	57868	PA
AtCg00570 (psbF)	Chloro 125	63985	NA
AtCg00580 (psbE)	Chloro 126	64109	NA
AtCg00590 (petL)	Chloro 130	65716	A
AtCg00670 (clpP)	Chloro 141	69942	PA
AtCg00890 (ndhB)	Chloro 185	949999	A
AtCg00890 (ndhB)	Chloro 189	96698	PA
AtCg00890 (ndhB)	Chloro 190	97016	PA
AtCg01010 (ndhF)	Chloro 204	112349	NA
AtCg01050 (ndhD)	Chloro 212	116281	NA
AtCg01050 (ndhD)	Chloro 212	116290	PA
AtCg01050 (ndhD)	Chloro 212	116494	PA
AtCg01050 (ndhD)	Chloro 214	117166	A
AtM30498( ccb206)	Mito 28	30616	A
AtM30498( ccb206)	Mito 28	30621	PA
AtM30498( ccb206)	Mito 28	30622	PA
AtM30498( ccb206)	Mito 28	30626	PA
AtM30498( ccb206)	Mito 28	30634	PA
AtM30498( ccb206)	Mito 28	30641	PA
AtM30498( ccb206)	Mito 28	30643	PA
AtM30498( ccb206)	Mito 28	30650	PA
AtM30498( ccb206)	Mito 28	30655	PA
AtM30498( ccb206)	Mito 28	30656	PA
AtM30498( ccb206)	Mito 28	30748	PA
AtM30498( ccb206)	Mito 28	30766	PA
AtM30498( ccb206)	Mito 28	30800	PA
AtM134288( nad7)	Mito 128	134300	PA
AtM134358( nad7)	Mito 128	134309	PA
AtM134358( nad7)	Mito 128	134340	PA
AtM134358( nad7)	Mito 128	134347	PA
AtM134358( nad7)	Mito 128	134412	PA
AtM134358( nad7)	Mito 128	134431	PA
AtM134358( nad7)	Mito 128	134440	PA
AtM134358( nad7)	Mito 128	134674	PA
AtM157500( orfX)	Mito 171	157634	NA
AtM157500( orfX)	Mito 171	157635	NA
AtM157500( orfX)	Mito 171	157651	NA
AtM157500( orfX)	Mito 171	157654	NA
AtM157500( orfX)	Mito 171	157663	NA
AtM157500( orfX)	Mito 171	157690	NA
AtM157500( orfX)	Mito 171	157851	NA
AtM157500( orfX)	Mito 171	157854	NA
AtM157500( orfX)	Mito 171	157867	NA
AtM8198(rps4)	Mito 95	82060	A
AtM8198(rps4)	Mito 95	82065	PA
AtM8198(rps4)	Mito 95	82074	PA
AtM8198(rps4)	Mito 95	82075	PA
AtM8198(rps4)	Mito 95	82125	A
AtM8198(rps4)	Mito 95	82150	PA
AtM8198(rps4)	Mito 94	82158	PA
AtM8198(rps4)	Mito 95	82161	PA
AtM27775(nad9)	Mito 11	23797	NA
AtM27775(nad9)	Mito 11	23838	NA
AtM27775(nad9)	Mito 11	23908	NA
AtM27775(nad9)	Mito 11	23938	NA
AtM27775(nad9)	Mito 11	24046	NA
AtM27775(nad9)	Mito 11	24069	NA
AtM27775(nad9)	Mito 11	24074	NA

Pris ensemble ces éléments constituent des limitations techniques importantes au développement d'un crible « haut-débit » efficace, sensible, rapide et peu coûteux, qui permettrait de connaître l'« éditeur » (mesure de la fréquence d'édition pour l'ensemble des sites décrits) d'un mutant. Si un tel crible n'existe pas encore actuellement, nous avons toutefois tenté plusieurs approches, en parallèle. Elles sont toutes basées sur la détection et éventuellement la quantification du polymorphisme C/T existant, au niveau du site édité, entre la séquence de l'ADN génomique et l'ADN complémentaire correspondant, chez une plante sauvage, polymorphisme qui doit disparaître totalement ou partiellement chez un mutant affecté dans le processus d'édition (figure n°64).

#### \* Crible utilisant des enzymes de restrictions

Le principe de ce crible repose sur le fait que le polymorphisme généré par l'édition peut, pour certains sites, faire apparaître ou disparaître un site reconnu par une enzyme de restriction. Lors de la mise au point de l'utilisation de ce crible au laboratoire d'Evry, j'ai mis en évidence qu'une telle méthode n'est pas appropriée à l'étude de sites partiellement édités (résultats non présentés). Par conséquent, il semble peu adapté à la mise en évidence de baisses partielles d'édition dans les lignées *AtDYWI* « ARNi » inductibles, et nous avons choisi d'utiliser d'autres approches.

#### \* Crible par séquençage direct de produits PCR amplifiés à partir d'ADN complémentaires :

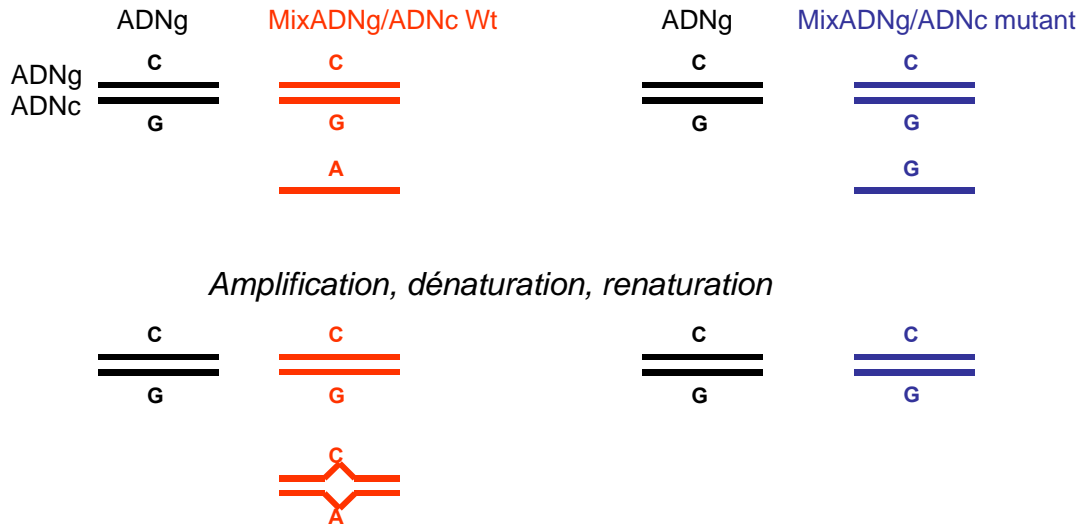
Pour réaliser ce crible, les ARN totaux de plantules de 21 jours présentant un phénotype « petits-choux » (lignées *AtDYWI* « ARNi » inductible 1 et 4) ou de plantules sauvages (Col-0) poussées dans les mêmes conditions ont été extraits et des ADNc ont été obtenus par transcription inverse. Des séquences chloroplastiques et mitochondriales, renfermant des sites édités, ont été amplifiées par PCR, et les produits PCR ainsi obtenus ont été analysés par séquençage direct (figure n°65A). Afin de limiter le coût de cette approche, nous avons pris soin d'amplifier des séquences ayant plusieurs sites d'éditeurs rapprochés, en particulier pour les transcrits mitochondriaux.

Ainsi pour 5 à 10 plantes affectées de chaque lignée, 63 sites d'édition ont été analysés (dont 45 sites mitochondriaux répartis sur 5 transcrits et 18 sites chloroplastiques répartis sur 11 transcrits). De nombreux défauts d'édition, tant dans les transcrits mitochondriaux que dans les transcrits chloroplastiques, ont été révélés par cette approche (figure n°65B). Nous avons développé une méthode permettant d'évaluer le niveau d'édition pour chacun des 63 sites étudiés (figure n°66), et répartir nos observations en 3 catégories : les sites ne semblant pas affectés (NA), les sites semblant partiellement affectés (PA) et les sites semblent très affectés (A). Au total, 47 des sites étudiés présentent des défauts d'édition : soit 29 sites mitochondriaux répartis dans 3 transcrits et 12 sites chloroplastiques répartis dans 6 transcrits (table n°5).

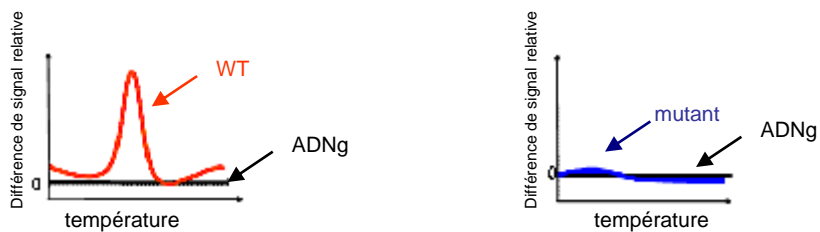
La distribution de ces défauts d'édition ne semble pas se faire au hasard puisque tous les sites étudiés d'un même transcrit sont affectés ou partiellement affectés. Ceci laisse penser que, comme nous l'avons évoqué en introduction, la machinerie d'édition pourrait progresser linéairement le long de la molécule d'« ARN », au moins au niveau de ces transcrits. Alternativement, l'existence d'un défaut d'édition dans un transcrit pourrait le déstabiliser entraînant une diminution de l'efficacité d'édition des autres sites.

En revanche, l'efficacité d'édition varie aléatoirement d'un site à l'autre et semble totalement abolie pour 7 de ces sites. Par ailleurs, nous avons également analysé par cette méthode le niveau de l'édition de ces sites pour deux plantules de la lignée 2 qui ne présentent ni

## 1. PCR en temps réel (produits PCR contenant un site d'édition)



## 2. High resolution melting, acquisition du signal et analyse



D'après Chateigner-Boutin et Small, 2007

Figure n°67 : Principe du crible par HRM (pour High-Resolution Melting) non quantitatif

La comparaison des profils de «melting», obtenus à partir d'un mélange équimolaire de produits PCR amplifiés à partir d'un ADNg et d'ADNc provenant d'une plante mutante ou sauvage poussée dans les mêmes conditions, permet de détecter des défauts d'édition.

phénotype, ni diminution de l'expression du gène *AtDYWI*. Aucun défaut d'édition n'a pu être détecté (résultats non présentés).

L'une des limitations majeures de cette approche est qu'elle ne nous donne qu'une « mesure très imprécise » du taux d'édition pour chaque site étudié. En effet, plusieurs biais ne sont pas pris en compte dans notre méthode d'analyse des chromatogrammes de séquençage. Par exemple, lors de la réaction de séquençage, le mélange de dNTPs utilisés peut ne pas être équimolaire. Par ailleurs, dans les chromatogrammes, les « maximums » peuvent être incertains (« épaulement » dans les pics) gênant leur mesure. Par conséquent, on peut se demander si une mesure d'aire serait plus appropriée. Ainsi, il faut rester très prudent quant à l'interprétation des résultats obtenus. Cette approche nous permet d'établir une liste de sites d'édition potentiellement affectés dans les lignées *AtDYWI* « ARNi » inductibles. Certains sites, semblant totalement affectés, sont sans doute ceux à étudier prioritairement par une approche plus sensible afin de confirmer et quantifier les défauts observés. En effet, cela permettra de déterminer s'il reste ou non un faible quantité de transcrit édité dans les plantules affectées.

#### \* Crible par HRM (pour High-Resolution Melting), non quantitatif

L'analyse par HRM (pour High-Resolution Melting) repose sur une amplification par PCR d'une courte séquence d'ADN en présence d'un composé fluorescent capable de s'intercaler entre les paires de bases d'un ADN double-brin. Les produits PCR obtenus sont alors soumis à une dénaturation thermique au cours de laquelle le changement de fluorescence (provoqué par la libération de l'agent intercalant fluorescent lors de l'ouverture des brins d'ADN) est précisément mesuré. On obtient ainsi le profil de dénaturation-renaturation (« melting »), des produits PCR analysés, en fonction de la température. Un produit d'amplification contenant un ou plusieurs mésappariements (dits « hétéroduplex ») va être dénaturé à une température plus basse qu'un produit d'amplification ne contenant aucun mésappariement (dits « homoduplex »). Cette différence peut être analysée par comparaison des profils de « melting » obtenus.

Au laboratoire de Perth, la machine de PCR en temps réel LC480 (Roche) permet de réaliser des analyses HRM sur 384 échantillons en parallèle, et de façon suffisamment résolutive pour permettre la détection d'un mésappariement d'une seule base dans un produit d'amplification d'une taille inférieure ou égale à 700pb. Ceci a permis le développement d'un crible « haut-débit » pour étudier l'édition des 34 sites chloroplastiques d'un mutant chez *A. thaliana* (Chateigner-Boutin and Small, 2007). J'ai utilisé ce crible et comparé les profils de dénaturation obtenus à partir d'un mélange équimolaire de produits PCR amplifiés à partir d'un ADN<sub>g</sub> et d'ADN<sub>c</sub> provenant d'une plantule affectée de chaque lignée *AtDYWI* « ARNi » inductible ou d'une plantule sauvage poussée dans les mêmes conditions (figure n° 67). Un profil de dénaturation différent est obtenu entre la plante sauvage et les plantes affectées pour 5 mélanges de produits PCR. Ceci suggère que les sites d'édition *psbZ* (35800), *accD* (57868), *rps12* (69553), *rpoA* (78691), et *ndhF* (112349) sont potentiellement affectés dans les lignées *AtDYWI* « ARNi » inductibles.

Le crible utilisé est uniquement qualitatif et peut permettre la détection un défaut d'édition au niveau d'un ou plusieurs sites mais ne permet pas de conclure à l'absence d'un défaut pour les autres sites, notamment s'ils s'avèrent partiellement édités chez le mutant. A nouveau, il faut donc rester prudent quant à l'interprétation des résultats. Les 5 sites ainsi identifiés constituent sans doute des candidats à étudier en priorité par une approche plus sensible afin de confirmer

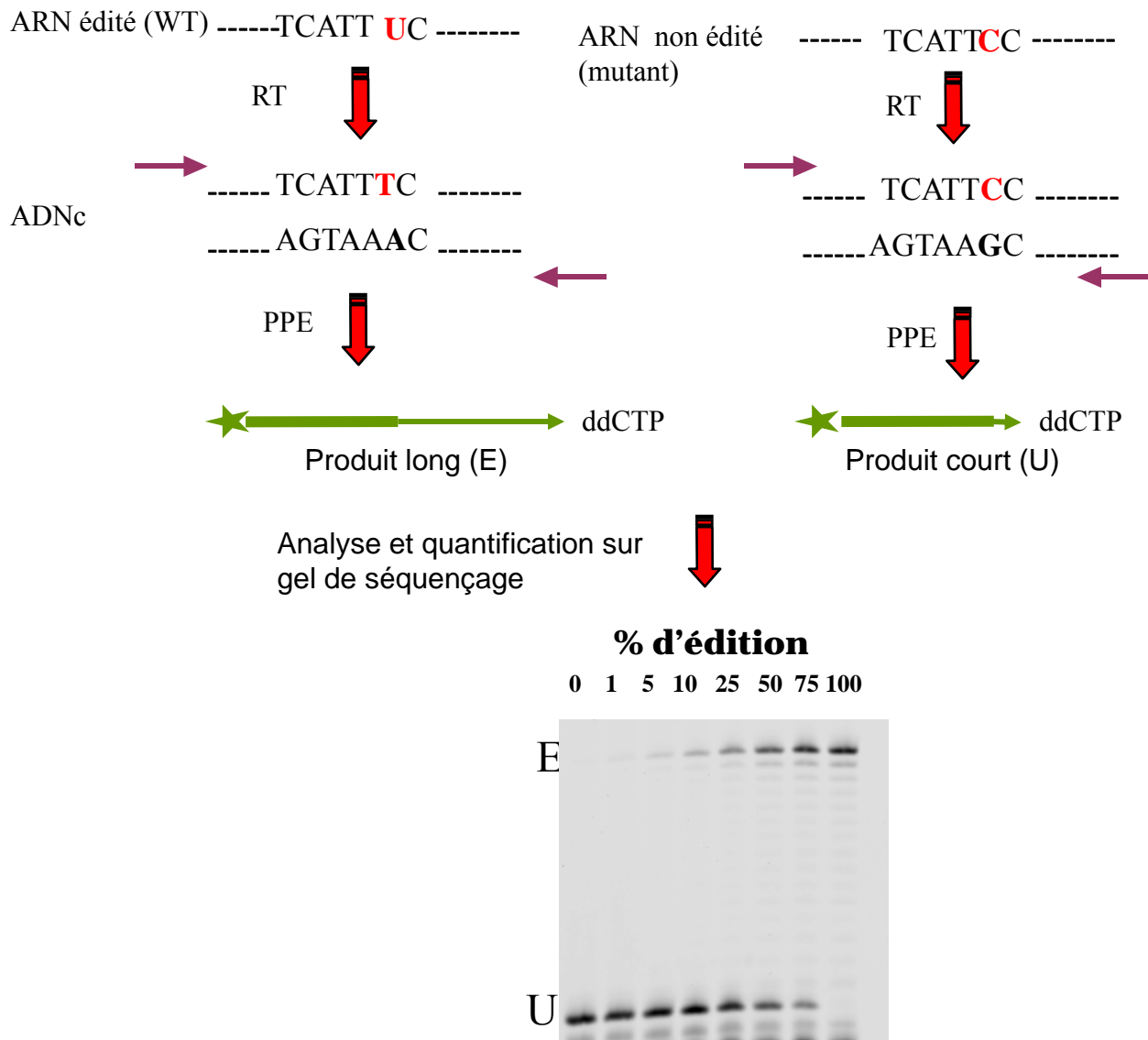


Figure n°68 : Principe de l'analyse quantitative de l'édition par PPE (poisoned primer extension)

Des produits PCR amplifiés à partir d'ADN complémentaires de plantes sauvages ou mutantes sont soumis à une réaction de PPE.

Un oligonucléotide marqué en 5' (marquage fluorescent FAM) est allongé par une polymérase en présence d'un mélange de nucléotides contenant un ddNTP (ddCTP, dans cet exemple).

Si l'ADNc a été obtenu à partir d'un ARN contenant un site édité (WT), l'extension est longue (elle s'arrête à la première cytosine après le site d'édition).

Si l'ADNc a été obtenu à partir d'un ARN ne contenant pas de site édité (mutant), l'extension est courte (elle s'arrête à la cytosine non éditée).

Les produits sont séparés sur gel de séquençage et la fluorescence révélée est quantifiée afin de déterminer le niveau d'édition. Le gel présente l'exemple des résultats que l'on peut obtenir à partir d'une gamme de produits édités de 0 à 100%

et quantifier les défauts observés, mais il faut garder à l'esprit que des défauts d'édition au niveau d'autres sites peuvent exister dans les plantules affectées des lignées *AtDYWI* « ARNi » inductibles.

\* Analyse par PPE (pour Poisoned Primer Extension)

L'extension « empoisonnée » d'un oligonucléotide (ou PPE) par l'utilisation de ddXTP permet de quantifier très précisément la fréquence d'édition à un site particulier (figure n°68). Cette méthode est couramment utilisée car elle permet de détecter une activité résiduelle d'édition (seuil de sensibilité de 1 % environ) (Hayes and Hanson, 2007), cependant sa lourde mise place technique ne permet pas de l'utiliser pour « cribler » de nombreux sites. Ainsi, nous avons fait le choix d'utiliser cette technologie pour étudier, dans un premier temps, l'édition de 11 sites chloroplastiques dans des plantules affectées des lignées *AtDYWI* « ARNi » inductibles.

9 des sites choisis correspondent aux meilleurs candidats à tester d'après les résultats obtenus dans les cribles par séquençage direct et HRM (respectivement, 4 et 5 sites chloroplastiques). Par ailleurs, le site *clpP* (69942) a été ajouté à cette liste pour les raisons suivantes : (1) nous savons que la protéine PPR CLB19 est impliquée dans l'édition des sites *rpoA* (78691) et *clpP* (69942) (2) d'après les résultats du crible HRM, *rpoA* est un site affecté dans l'édition dans nos lignées mutantes (3) de plus, d'après les résultats du crible par séquençage, *clpP*, bien qu'il ne soit pas parmi les 4 meilleurs candidats chloroplastiques, semble également potentiellement affectés dans l'édition dans ces lignées (4) enfin l'amorce PPE permettant l'analyse du site *clpP* était disponible au laboratoire à Perth facilitant la mise en place de cette expérimentation. Enfin, le site *rpoC1* (21806) a également été choisi pour être analysé par PPE dans nos lignées mutantes. Dans aucun crible préliminaire, ce site n'est apparu affecté, et l'amorce PPE permettant son analyse était également disponible au laboratoire de Perth. Ainsi, ce site apparaît comme un bon candidat pour servir de « témoin négatif » pour les analyses PPE (on entend par là, un site dont la fréquence d'édition serait inchangée dans nos lignées mutantes par rapport à celle des plantules sauvages). Par ailleurs, rappelons que nous savons qu'une PPR de la sous-classe DYW (FLV) est impliquée dans l'édition de ce site (Chateigner-Boutin, résultats non publiés).

Pour mener ces analyses de PPE, afin de faciliter l'interprétation des résultats, nous avons pris soin d'étudier des échantillons biologiques comparables, c'est-à-dire, des plantules de phénotype « petits-choux » très similaires, cultivées dans une seule condition et provenant toutes de la même génération (T3) pour les lignées « ARNi » inductibles 1 et 4. De plus, la diminution de la quantité de transcrit *AtDYWI* a été précisément évaluée par PCR quantitative pour chaque plantule individuelle avant de mener ces analyses de PPE (comme décrit dans le manuscrit, présenté au paragraphe III.5, figure n°4).

Pour 9 sites ainsi étudiés, des défauts d'édition ont pu être confirmés par cette analyse en PPE (table n°6). En effet, pour les 3 sites *ndhD* (gp :117166), *accD* (gp :57868) et *rpoB* (gp :23898), l'édition est totalement abolie dans toutes les plantules affectées des lignées mutantes 1 et 4 (figure n° 5 dans le manuscrit présenté au paragraphe III.5). Pour 4 autres sites (*psbZ* (gp :35800), *petL* (gp :65716), *rpoA* (gp :78691) et *ndhF* (gp :112349)), l'édition est partiellement affectée pour certaines plantules de phénotype « petits-choux » dans ces lignées (figure n° 5 dans le manuscrit présenté au paragraphe III.5). De plus, les sites d'édition *clpP* (gp : 69942) et *rps12* (gp : 69553) sont partiellement affectés uniquement dans des plantules de phénotype « petits-choux » de la lignée 4 (figure n°5 dans le manuscrit



Locus (Transcrit)	Amorces	position génomique	AA changement	PPE ( % d'édition)		
				<i>AtDYW1</i> "ARNi" lignée1	<i>AtDYW1</i> "ARNi" lignée4	WT
AtCg00170 (rpoB)	23581F-24093R	23898	S>L	0	0	94-99
AtCg00300 (psbZ)	Chloro 68	35800	S>L	80-95	59-94	93-95
AtCg00500 (accD)	Chloro 112	57868	S>L	0	0	88-92
AtCg00590 (petL)	Chloro 130	65716	P>L	40-72	0	69-73
AtCg00650 (rps12)	Chloro 140	69553		78-84	16-20	77-84
AtCg00670 (clpP)	Chloro 141	69942	H>Y	86-92	38-92	81-92
AtCg00740 (rpoA)	Chloro 159-286	78691	S>F	63-78	42-67	78-82
AtCg01010 (ndhF)	Chloro 204	112349	S>L	73-94	55-95	90-95
AtCg01050 (ndhD)	Chloro 214	117166	>M	0	0	66-67

Table n°6: Résultats, pour 9 sites chloroplastiques, de la quantification du niveau d'édition (obtenue par analyse de PPE) dans des plantules affectées des lignées *AtDYW1* « ARNi » inductibles

présenté au paragraphe III.5). Pour le site *rpoC1*, aucun défaut d'édition n'a pu être mis en évidence dans toutes les plantules affectées des lignées mutantes 1 et 4. Enfin, un problème technique (non incorporation de l'amorce PPE) ne nous a pas permis d'analyser le niveau d'édition du site *ndhB* (gp : 949999).

Remarque : Nous avons également analysé par cette méthode le niveau de l'édition des sites *ndhD* (gp : 117166), *rpoA* (gp : 78691) et *clpP* (gp : 69942) pour des plantules de la lignée 2 qui ne présentent ni phénotype, ni diminution de l'expression du gène *AtDYW1*. Aucun défaut d'édition n'a pu être détecté (résultats non présentés).

### III.3.8 Bilan et interprétations des résultats

La recherche de défauts d'édition dans les lignées « ARNi » affectées dans l'expression du gène *AtDYW1* s'est déroulée en deux étapes principales. Dans un premier temps, nous avons mené une recherche sans *a priori* sur de nombreux sites d'édition chloroplastiques et mitochondriaux par différentes méthodes de cribles. Ces études nous ont permis d'identifier de nombreux défauts d'édition dans les transcrits des deux organites pour les plantules affectées de ces lignées mutantes. Cependant, ces cribles sont qualitatifs et peu sensibles, ils présentent donc des limites pour la caractérisation notamment des défauts partiels d'édition et les sites ainsi identifiés doivent être confirmés. Dans un second temps, une analyse par PPE a été menée afin de réaliser une quantification très sensible du niveau d'édition pour une dizaine de sites chloroplastiques. Cette étude a permis de confirmer de défauts d'édition pour 9 sites chloroplastiques dans les plantules de phénotype « petits-choux » des lignées 1 et 4.

*Vers l'établissement de l' « éditotype » chloroplastique dans les lignées « ARNi » affectées dans l'expression du gène AtDYW1, par PPE*

Une des premières interrogations quant au modèle consiste à se demander si la protéine AtDYW1 intervient au niveau de tous les sites.

Dans les plantules de phénotype « petits-choux » des lignées 1 et 4, la quantification de l'édition au niveau d'une dizaine de sites chloroplastiques a indiqué que pour 3 sites *ndhD* (117166), *accD* (57868) et *rpoB* (23898), l'édition est totalement abolie dans toutes les plantules de phénotype « petits-choux », suggérant ainsi que la protéine AtDYW1 doit certainement être un facteur essentiel pour l'édition de ces sites.

Le gène chloroplastique *ndhD* code pour une sous-unité du complexe NDH (pour NAD(P)H déhydrogenase), et l'absence totale d'édition au site *ndhD* (gp : 117166) a déjà été observée dans des lignées mutantes d'*A. thaliana* affectées dans l'expression du gène *CRR4* (Kotera et al., 2005). L'édition à ce site permettant la création d'un codon AUG initiateur de traduction, dans ces mutants il a été montré que la sous-unité *ndhD* correspondante n'est pas synthétisée ce qui conduit à un défaut d'assemblage et d'activité du complexe NDH dans les chloroplastes, néanmoins le développement et la croissance de ces mutants ne sont que peu affectés (Kotera et al., 2005). Ainsi, le fort phénotype « petits-choux » observés dans les lignées « ARNi » affectées dans l'expression du gène *AtDYW1* n'est sans doute pas lié à l'absence d'édition du site *ndhD* (gp : 117166).

En revanche, le gène *accD* code pour la beta-carboxytransférase, une sous-unité du complexe chloroplastique Acétyl-CoA carboxylase (ACCCase). Chez *Arabidopsis*, l'étude d'un mutant d'une autre sous-unité de ce complexe (CAC1) a révélé des plantes morphologiquement très affectée (Hui-Rong Qian, 1998). Quant au gène *rpoB*, il code pour la sous-unité bêta de l'ARN polymérase chloroplastique (PEP). L'analyse de mutants affectés dans l'expression de

ce gène chez l'orge ou le tabac a révélé des plantes albinos déficientes dans le fonctionnement chloroplastique (Inada et al., 1997; De Santis-MacLossek G, 1999). L'édition de ces deux sites permet la restauration d'un acide aminé conservé chez des protéines homologues d'autres espèces. Ainsi, l'absence totale d'édition à ces deux sites, dans nos plantules « affectées » des lignées « ARNi » inductibles, peut résulter en l'absence de ces deux protéines fonctionnelles dans les chloroplastes, et expliquer le fort phénotype observé.

En revanche, 6 autres sites ne semblent que partiellement affectés, et enfin il existe un site pour lequel aucun défaut n'a été mis en évidence. Pour expliquer ces observations, on peut proposer que la protéine *AtDYW1* est un facteur essentiel intervenant dans l'édition de tous les sites, mais comme son extinction n'est que partielle dans nos lignées, nous observons une « gamme d'effet » sur l'édition de différents sites. Alternativement, on peut imaginer qu'en absence d'une quantité « normale » de protéine *AtDYW1*, d'autres protéines (par exemple des partenaires du complexe d'édition) puissent « assurer » plus ou moins efficacement le rôle de cette protéine.

Supportant l'une ou l'autre de ces hypothèses, l'analyse des défauts d'édition partielle suggère l'existence d'une corrélation, entre le niveau d'expression du gène *AtDYW1*, la survie des plantes et le niveau d'édition dans ces plantes. En effet, il apparaît que dans la lignée 4, la plus affectée dans l'expression du gène *AtDYW1*, on observe également des défauts d'édition partiels plus sévères que dans la lignée 1. De plus, deux sites d'édition, dans les transcrits *clpP* et *rps12*, ne semblent affectés que dans des plantules de la lignée 4. Néanmoins, comme nous l'avons vu en introduction, l'édition partielle est un phénomène courant dans les transcrits des organites chez les plantes. De plus, la fréquence d'édition à un site donné peut varier en fonction du tissu végétal étudié, du stade de développement de la plante mais aussi des conditions environnementales extérieures. De ce fait, établir une corrélation stricte entre le niveau d'expression du gène *AtDYW1* et le niveau d'édition aux différents sites, dans nos lignées mutantes, s'avère un exercice délicat.

Par ailleurs, aux vues des éléments dont nous disposons, on ne peut pas exclure que la protéine *AtDYW1* soit essentielle ou nécessaire pour l'édition de certains sites, et non impliquée dans l'édition d'autres sites.

Une autre façon de s'interroger quant au modèle consiste à se demander si la protéine *AtDYW1* intervient au niveau de tous les sites pour lesquels l'édition dépend d'une protéine PPR de type E/E+.

Grâce à nos connaissances actuelles sur l'analyse de mutant d'*Arabidopsis*, nous savons que 3 protéines PPR de type E/E+ (*CCR4*, *CRR21* et *CLB19*) sont essentielles à l'édition de 4 sites chloroplastiques (*ndhD* : gp : 117166 ; *ndhD* : gp : 116785 ; *clpP* : gp : 69942 ; et *rpoA* : gp : 78691). Dans les lignées « ARNi » affectées dans l'expression du gène *AtDYW1*, il s'avère que l'édition du site *ndhD* (gp : 117166) est totalement abolie et que ce processus est partiellement affecté pour deux autres sites *clpP* (gp : 69942) et *rpoA* (gp : 78691). En revanche, d'après le crible HRM, l'édition du site *ndhD* (gp : 116785) n'apparaît pas affectée. Cependant, pour ce site, le niveau d'édition n'a pas été précisément quantifié par PPE, et il est tout à fait envisageable que l'édition de ce site soit partiellement affectée. De plus, comme évoqué précédemment, l'extinction du gène *AtDYW1* n'est que partielle dans ces lignées, et il pourrait rester une quantité suffisante de protéine *AtDYW1* pour permettre l'édition ce site. Cependant, le fait que la protéine *AtDYW1* n'intervienne pas dans l'édition de ce site, reste une éventualité que l'on ne peut pas exclure à l'heure actuelle. En conclusion, il apparaît que la protéine *AtDYW1* est impliquée dans l'édition d'au moins trois des 4 sites pour lesquels une protéine PPR de type E/E+ est essentielle. Ceci supporte le modèle selon lequel la protéine *AtDYW1* est recrutée par les protéines PPR de type E/E+.

Par ailleurs, afin d'évaluer si la protéine AtDYW1 peut s'associer au domaine DYW des protéines PPR et former un hétérodimère au cours de la réaction d'édition, on peut s'interroger sur le niveau d'édition du site *rpoC1* (gp : 21806).

En effet, il s'agit de l'unique site pour lequel nous savons qu'une protéine PPR de type DYW est essentielle pour la réaction d'édition. Dans les lignées « ARNi » affectées dans l'expression du gène *AtDYW1*, après quantification par PPE, il s'avère qu'aucun défaut d'édition n'a pu être mis en évidence pour ce site. Ceci pourrait suggérer que la protéine AtDYW1 ne s'associe pas au domaine DYW de la protéine PPR-FLV pour permettre l'édition du site *rpoC1*. Cependant, comme évoqué précédemment, le fait que l'extinction du gène *AtDYW1* ne soit que partielle dans les lignées étudiées, ne nous permet pas d'affirmer que la protéine AtDYW1 ne participe à l'édition de ce site.

A l'heure actuelle, en raison de limitations de temps, l'« éditotype » chloroplastique complet des lignées « ARNi » affectées dans l'expression du gène *AtDYW1* n'a pas été réalisé. Cependant, la mise au point d'un crible HRM quantitatif au laboratoire de Perth en plaque de 384 puits, et l'adaptation de la technique de PPE à « moyen » débit au laboratoire d'Evry devrait permettre dans un futur proche de compléter cette étude.

Par ailleurs, les sites étudiés ont été choisis d'après les résultats de crible permettant la recherche systématique de défauts d'édition. Cependant ces cribles ont révélés qu'un plus grand nombre de sites semble être affecté dans ces lignées mutantes dans les transcrits chloroplastiques, mais aussi dans les transcrits mitochondriaux comme nous allons le discuter à présent.

*Existe-t-il des défauts d'édition dans les transcrits mitochondriaux dans les lignées « ARNi » affectées dans l'expression du gène AtDYW1 ?*

Lors du crible par séquençage direct de produits PCR amplifiés à partir d'ADN complémentaires, des défauts d'édition au niveau de 29 sites répartis sur 3 transcrits mitochondriaux ont été mis en évidence dans les lignées « ARNi » affectées dans l'expression du gène *AtDYW1*. Dans ce crible, nous n'avons évalué l'édition de moins d'un quart des 441 sites décrits dans le transcriptome mitochondrial d'*Arabidopsis*, suggérant que dans ces lignées un plus grand nombre de sites pourrait être affecté.

Dues à des limitations techniques et de temps, au cours de ma thèse, je n'ai pas pu confirmer, par une analyse quantitative par PPE, l'existence de tels défauts d'édition dans le transcriptome mitochondrial. En effet, lors du crible par séquençage, afin de limiter son coût, j'ai choisi d'analyser l'édition au niveau de transcrits à forte densité en nombre de sites. Les sites sont donc proches les uns des autres et parfois même consécutifs. Ces caractéristiques rendent leur étude peu ou non envisageable dans le cadre d'une analyse PPE, car d'une part, le dessin d'une amorce PPE dans une telle région s'avère délicat et, d'autre part, la présence de défauts multiples d'édition dans un même produit PCR conduirait à un profil complexe de produits PPE dont l'interprétation est difficile. De ce fait, réinitier la recherche de sites potentiellement affectés dans des transcrits mitochondriaux ayant un ou un petit nombre de sites d'édition semble une étape préalable indispensable à la quantification de l'édition au niveau de sites mitochondriaux dans nos lignées d'intérêt.

L'enjeu de cette étude est important car, si l'existence de défauts d'édition mitochondriaux dans les lignées « ARNi » affectées dans l'expression du gène *AtDYW1* est confirmée, alors ceci soutient l'hypothèse selon laquelle la protéine AtDYW1 doit être présente à la fois dans les chloroplastes et les mitochondries. Ceci suggère également que le fonctionnement de la machinerie d'édition dans ces deux organites serait identique.

Cependant, s'il s'avère que la protéine AtDYW1 est chloroplastique uniquement, et que dans les lignées « ARNi » affectées dans l'expression du gène *AtDYW1* il existe réellement des défauts d'édition dans les transcrits mitochondriaux, quelle (s) interprétation (s) pourrait-on avancer pour expliquer ces observations?

Dans un tel cas, il est possible que les défauts d'édition observés ne soient pas directement liés à une diminution de l'expression du gène *AtDYW1*, mais apparaîtraient secondairement. En effet, dans une plante au phénotype très affecté, le fonctionnement des centres producteurs d'énergie (chloroplastes et mitochondries) peut être perturbé et l'édition de différents transcrits indirectement dérégulé. Une telle éventualité remettrait en cause l'interprétation de nos résultats quant à l'implication de la protéine AtDYW1 dans l'édition des transcrits chloroplastiques et mitochondriaux. Alternativement, comme évoqué précédemment, on peut s'interroger sur la spécificité du « silencing » dans ces plantes. En effet, il est peut être envisageable que l'expression d'autres gènes, en particulier des gènes codant pour des protéines PPR à domaine DYW, puissent être affectée dans ces lignées.

En conclusion : L'étude des lignées « ARNi » inductibles de première génération a révélé des plantes présentant des défauts de développement (phénotype « petits choux ») et pour lesquelles l'expression du gène *AtDYW1* et le processus d'édition sont affectés. L'un des points forts de cette étude est qu'elle souligne l'implication de la protéine AtDYW1 au niveau de plusieurs sites d'édition suggérant qu'elle ne joue sans doute pas le rôle de facteur de spécificité, mais plutôt qu'elle est impliquée au niveau d'une étape commune à tous ces sites, par exemple la catalyse de la réaction de désamination comme le suggère notre modèle. Cependant, comme nous l'avons évoqué, deux critiques majeures peuvent être apportées quant à cette étude. D'une part, le matériel génétique utilisé est problématique dans la mesure où l'on ne peut pas expliquer moléculairement comment l'extinction du gène *AtDYW1* se produit dans ces lignées. J'ai donc consacré une partie de mon travail de doctorat à établir et caractériser d'autres lignées mutantes pour ce gène, comme je vais le décrire à présent. D'autre part, on ne peut pas affirmer que les défauts d'édition observés dans ces lignées soient directement liés à l'absence de protéine AtDYW1, ni que cette dernière soit effectivement l'enzyme centrale du processus d'édition. C'est pourquoi, j'ai également une caractérisation « fonctionnelle » de cette dernière, comme nous le verrons dans la quatrième partie des résultats.

### **III.4 Lignées « ARNi » inductibles de deuxième génération**

#### *III.4.1 Etablissement des lignées « ARNi » inductibles de deuxième génération*

L'objectif de cette étude est d'obtenir d'autres lignées *AtDYW1* « ARNi » présentant idéalement un phénotype lié à une diminution de l'expression du gène *AtDYW1* et dont l'apparition serait inductible et « contrôlée ».

#### Constructions utilisées

Pour se faire nous avons utilisé le vecteur pAGRIKOLAind3 dérivé du vecteur pAGRII utilisé ci-avant, et qui permet l'expression, *in planta*, d'un hpRNA sous contrôle du promoteur 6xpOp, inductible à la dexaméthasone. La différence majeure réside dans le fait que la cassette de production de l'hormone LhGR, toujours sous contrôle du promoteur constitutif (P35s CaMW), n'est plus apportée dans le génome des plantes transformées, par ce vecteur.

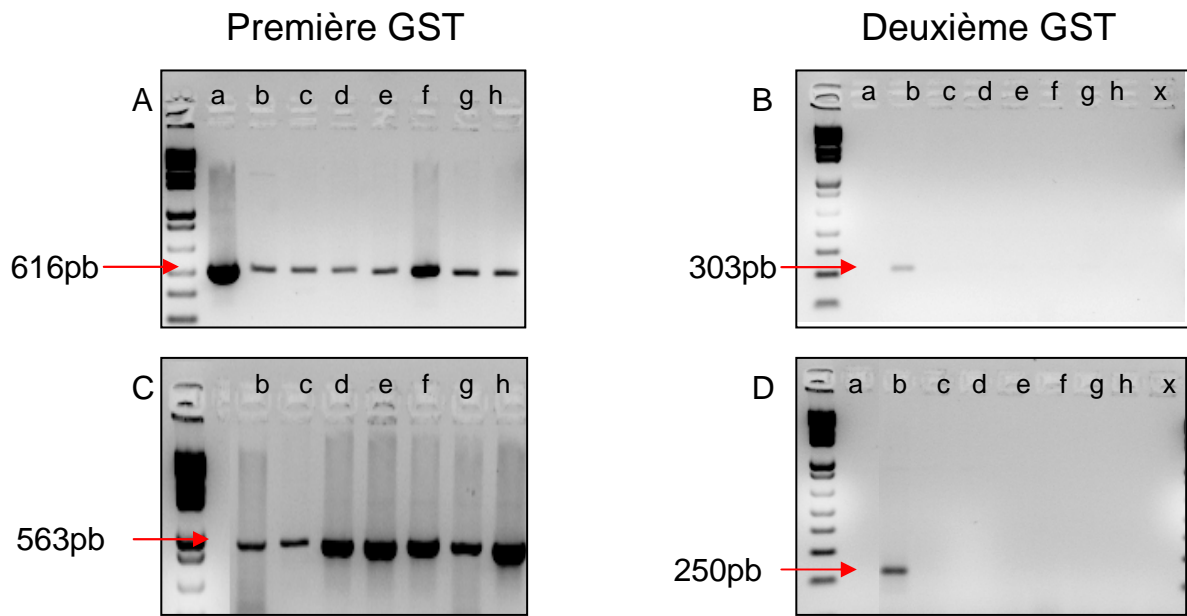


Figure n°69 : Résultats des PCR de génotypage de 8 plantes T1 des lignées *AtDYWI* « ARNi » inducible de deuxième génération

- A/ amplification de la première GST avec le couple AgriΩ51/Agri64(616pb)
- B/ amplification de la deuxième GST avec le couple Agri56/Agri69 (303pb)
- C/ amplification de la première GST avec le couple 5'GST/Agri64 (563pb)
- D/ amplification de la deuxième GST avec le couple 5'GST/Agri69(250pb)

Seule la plante b porte les deux GST, toutes les autres plantes ne portent que la première GST

Cette cassette a été intégrée de façon stable dans le génome d'une lignée dite « pilote », appelée 4C (lignée homozygote pour cette construction).

### Principe d'utilisation

Comme précédemment, une double recombinaison de LR clonase, permet de cloner la GST d'intérêt en lieu et place des cassettes Gateway dans le vecteur pAGRIKOLAind3. Les plantes transformées par cette construction donnent naissance à des lignées qui peuvent être alors croisées avec la lignée 4C. Ainsi, pour la moitié des plantes de la descendance d'un tel croisement, le système 6xpOp/LhGR est reconstitué et l'induction à la dexaméthasone peut être réalisée.

Ce système a été développé par l'équipe de Ian Moore (Université d'Oxford) afin d'améliorer l'efficacité d'induction et d'éviter les problèmes de « fuite » du promoteur 6xpOp dont nous avons discuté précédemment. En effet, notons qu'avant croisement avec la lignée 4C, les plantes des lignées porteuses de la cassette pAGRIKOLAind3 ne produisent pas la LhGR ce qui réduit considérablement tout risque d'expression « basale » du promoteur 6xpOp.

### Génotypage

Le génotypage des bactéries, agrobactéries et plantes porteuses de la cassette pAGRIKOLAind3 s'effectue selon la même méthode que celle utilisée pour pAGRII (décrite figure n°51).

### Sélection et génotypage des lignées

Après un traitement des cultures à l'herbicide « basta », 14 transformants (T1) indépendants ont été sélectionnés. Les graines de ces plantes ont été récoltées séparément de façon à obtenir 14 lignées (T2) indépendantes.

Parallèlement, une feuille de chaque plante T1 a été prélevée afin de réaliser une extraction d'ADN génomique ainsi que les PCR de génotypage.

Après analyse des produits PCR obtenus, pour seulement 2 lignées, nous avons pu détecter la présence des 2 GST dans la configuration attendue pour le sens 2. Pour les autres lignées, nous obtenons toujours l'amplification d'une seule GST (figure n°69). Afin de confirmer le génotypage de nos plantes (T1), nous avons dessiné une amorce dans la région 5' de la GST-CATMA1a38560. Lors d'un second génotypage utilisant cette nouvelle amorce, les résultats obtenus précédemment ont été confirmés soit : 2 transformants possédant une construction « correcte » avec 2 GST, et 12 transformants possédant une construction « incomplète » avec seulement l'une des 2 GST (figure n°69).

### Croisement avec la lignée 4C

Les graines obtenues à partir de ces deux transformants T1 ont été semées sur milieu de culture *in vitro* (1/2 GB) contenant de l'herbicide « basta ». Les plantules F2 résistantes au « basta » ont été repiquées en terre.

Pour 10 plantes de chacune des 2 lignées, 3 à 4 croisements sont réalisés avec la lignée 4C. Les croisements sont réalisés dans le sens [ $\text{♀mutant} \times \text{♂4C}$ ], et réciproquement

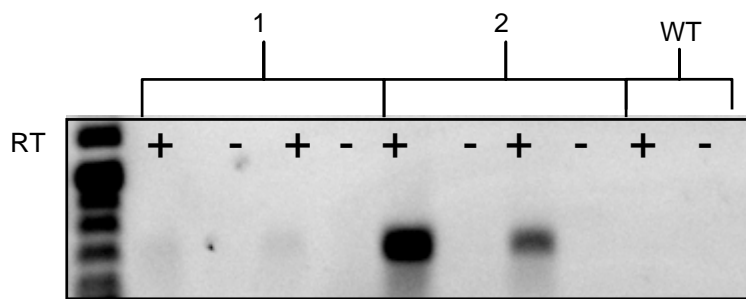


Figure n°70 : Lignées *AtDYWI* « ARNi » inductibles, deuxième génération:  
 Détection du transcrit *GUS*, résultats de RT-PCR

(1) plantes d'intérêt sans phénotype ayant poussé sur un milieu ne contenant pas d'inducteur

(2 et WT) plantes d'intérêt et sauvage sans phénotype ayant poussé sur un milieu contenant 10 $\mu$ M d'inducteur (dexaméthasone)



[♀4CX♂mutant]. Les graines F3 des siliques issues de ces croisements sont récoltées de façon indépendante (soit ~10 à 15 graines par lot).

*Nota Bene* : Nous avons fait le choix de ne pas mélanger les graines de siliques différentes (même si elles proviennent d'une même plante mère), car des erreurs techniques dues à des contaminations polliniques (c'est-à-dire que ce n'est pas le pollen souhaité qui féconde le pistil) sont fréquentes dans ce type d'approches. En cas de mélange de graines, ces erreurs peuvent biaiser la ségrégation en cas d'apparition d'un éventuel phénotype lors de l'induction. Ceci est particulièrement vrai dans notre cas, car nous travaillons sur un échantillon relativement restreint en graines.

#### III.4.2 Génotypage et phénotypage de plantes issues des croisements

Les graines T3 ont germé sur un milieu de culture *in vitro* (1/2 GB) pendant 10 jours. Stérilement, une feuille de chaque plantule a été prélevée et placée dans une plaque 96 puits, pour réaliser une extraction d'ADN génomique et les PCR de génotypage. Ces PCR servent à vérifier que le pollen fécondant est bien celui d'intérêt. Ainsi, pour les croisements [♀4CX♂mutant], nous utilisons les couples d'amorces Agri69/56 et AgriΩ51/64, et pour les croisements réciproques [♀mutantX♂4C], nous utilisons un couple d'amorces dans le gène de résistance à la kanamycine.

En théorie, les graines T3 d'intérêt peuvent être directement sélectionnées sur un milieu de culture *in vitro* contenant à la fois l'herbicide « basta » et de l'antibiotique kanamycine, afin d'éviter ces expériences de génotypage. Cependant, nous avons pensé que de telles conditions de croissance pourraient « stresser » les plantes et conduire à l'observation de phénotypes non spécifiques.

Après 10 jours de culture et vérification du génotype des plantules, pour chaque croisement, la moitié des plantules est transférée sur un milieu de culture *in vitro* contenant 10µM de dexaméthasone, et l'autre moitié sur un milieu de culture *in vitro* ne contenant pas cet inducteur. Après 21 jours de culture, aucun phénotype particulier n'a pu être observé pour des plantules ayant poussé sur un milieu contenant ou non l'inducteur.

Des graines de plantes sauvages (écotype Col0), ainsi que des graines de la lignée 4C et des deux lignées mutantes, ont germé dans les mêmes conditions. Elles nous ont servi de témoins tout au long de cette analyse phénotypique.

#### III.4.3 Validation de l'efficacité d'induction à la dexaméthasone

Un problème technique lors de l'induction pourrait expliquer l'absence d'apparition d'un phénotype dans ces deux lignées *AtDYW1* « ARNi » inductible. Afin de tester cette hypothèse, j'ai étudié, par des expériences de RT-PCR, l'expression du gène rapporteur *GUS* sur des plantes ayant poussé en présence ou non d'inducteur (figure n°70). Les résultats montrent une amplification du transcrite *GUS* uniquement pour les plantules ayant poussé en présence de l'inducteur. Ainsi, dans ces expériences, l'induction via la dexaméthasone semble efficace.

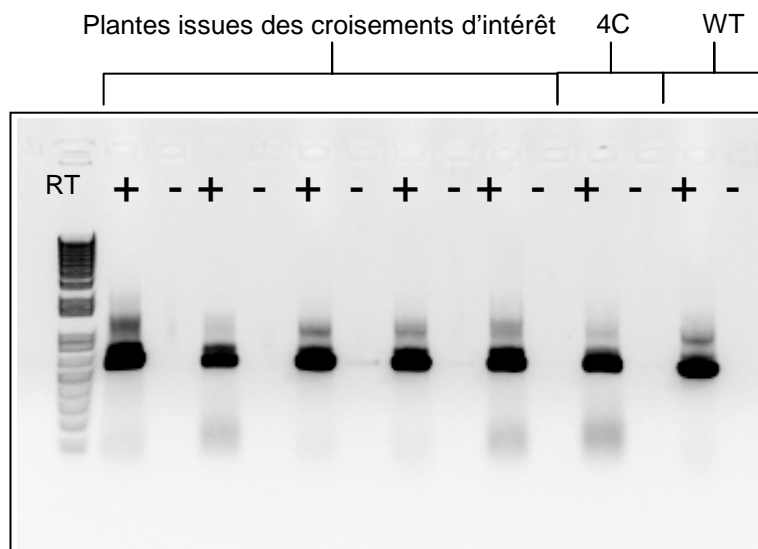


Figure n°71 : Lignées *AtDYWI* « ARNi » inductibles, deuxième génération:

Détection du transcrit *AtDW1* par RT-PCR sur des plantules sans phénotype ayant poussé sur un milieu contenant 10 $\mu$ M d'inducteur (dexaméthasone), les plantes 4C (lignée « pilote ») et WT(sauvage) servant de témoin.

#### III.4.4 Niveau d'expression du gène *AtDYWI* après 21 jours d'induction

Bien que je ne l'ai pas réellement quantifié, j'ai toutefois évalué le niveau d'expression du gène *AtDYWI* par RT-PCR dans les plantules ayant poussé en présence de l'inducteur (figure n°71). Par comparaison avec son expression dans une plante sauvage poussée dans les mêmes conditions, nos résultats n'indiquent aucune variation notable de la quantité de transcrite *AtDYWI* dans ces plantules (à notre seuil de détection).

#### III.4.5 Bilan

Ainsi, dans cette étude, nous avons établi 2 nouvelles lignées *AtDYWI* « ARNi » inductibles. Après croisement de ces lignées avec la lignée « pilote » 4C et induction à la dexaméthasone, aucun phénotype macroscopique n'a été observé. Nous avons vérifié que, dans nos expérimentations, l'induction via la dexaméthasone était efficace, en étudiant l'expression du gène rapporteur *GUS*. Par ailleurs, malgré cette induction chimique, l'expression du gène *AtDYWI* ne semble pas affectée dans ces 2 lignées.

Ainsi, comme nous l'avons déjà constaté lors de l'étude des lignées *AtDYWI* « ARNi » inductibles de première génération, une induction efficace, c'est-à-dire capable d'activer l'expression du gène rapporteur *GUS* (et probablement celle du hpRNA), ne traduit pas une extinction post-transcriptionnelle de l'expression du gène cible. Nous imaginons qu'une des étapes de la cascade de réactions conduisant au « silencing » du gène *AtDYWI* est bloquée, peut être afin d'assurer la survie de la plante.

Par ailleurs, lors du génotypage de ces lignées « ARNi » inductibles de deuxième génération, nous avons obtenu une incroyable proportion (6/7) de transformants primaires de génotype « incorrect », n'ayant qu'une seule des deux GST. Ceci suggère qu'une forte contre-sélection des plantes portant la cassette de « silencing » pAGRIKOLAind3 intègre semble avoir lieu. Une remarque similaire avait été émise lors de l'étude des lignées *AtDYWI* « ARNi » constitutives (paragraphe III.2). L'hypothèse sous-jacente à ces observations est que la seule présence de ces cassettes de « silencing » dans le génome des plantes transformées pourrait être néfaste à leur survie. Cependant, une telle observation n'est pas vraie pour les cinq transformants T1 obtenus lors de la production des lignées *AtDYWI* « ARNi » inductibles de première génération. En effet, leur génotypage a révélé cinq plantes portant une cassette de « silencing pAGRII » intègre. Comment expliquer ces différents résultats ?

La sélection des transformants portant la cassette de « silencing pAGRII » a été réalisée sur milieu de culture *in vitro* contenant de la kanamycine, alors que la sélection des transformants portant les autres type de cassettes a été réalisé en terre en utilisant l'herbicide « basta ». Il est généralement observé que des plantes, présentant un développement affecté, peuvent se développer sur milieu de culture *in vitro*, alors qu'elles ne germent pas directement en terre. Ainsi, il est raisonnable de penser que les transformants portant les cassettes de « silencing pAGRIKOLA ou pAGRIKOLAind3 » intègres pouvaient présenter un développement affecté et n'ont pas poussé en terre. En revanche, les transformants portant la cassette de « silencing pAGRII » d'intérêt ont poussé en milieu de culture *in vitro* bien que

quatre d'entre eux présentaient des défauts de développement que nous avons pu observer après leur transfert en terre.

**III.5 The Arabidopsis AtDYW1 protein: a key component of RNA editing machinery in chloroplasts (manuscript en préparation)**

## Manuscript in preparation

# The Arabidopsis DYW1 protein: a key component of the RNA editing machinery in chloroplasts

### *Introduction:*

In land plant chloroplasts and mitochondria, RNA molecules are altered by RNA editing, a specific post-transcriptional pyrimidine transition (mainly C-to-U conversions). In flowering plants, about 400 and 35 RNA editing sites have been described in mitochondrial and plastid transcripts, respectively (Giege and Brennicke, 1999; Notsu et al., 2002; Handa, 2003; Shikanai, 2006; Chateigner-Boutin and Small, 2007). By restoring conserved amino acid codons, most editing events are obligatory steps in the biosynthesis of functional proteins. Despite years of investigation, RNA editing remains an enigmatic process.

Recently, in *Arabidopsis thaliana*, 3 PPR proteins CRR4, CRR21, and CLB19 were described as essential for editing of *ndh-D*, *rpoA* and *clpP* chloroplast transcripts (Kotera et al., 2005; Okuda et al., 2007) (Chateigner-Boutin *et al.*, submitted). PPR proteins are defined by the presence of a tandem array of a 35 amino acid (pentatricopeptide) motif (Small and Peeters, 2000). A growing mass of data supports the idea that PPR proteins are involved in post-transcriptional regulation of gene expression in both plastids and mitochondria. In *Arabidopsis*, the PPR protein family is extraordinarily large, containing 450 members, and roughly half of the PPR proteins are in the PLS subfamily, which is specific to land plants (Aubourg et al., 2000; Lurin et al., 2004). It has been proposed that PLS proteins could be involved in plant-specific post-transcriptional processes, such as RNA editing which is unique to land plants (Lurin et al., 2004; Shikanai, 2006). CRR4, CRR21 and CLB19 belong to the E/E+ subgroup of the PLS-PPR subfamily, and do not contain domains likely to have catalytic activity, suggesting that these PPR proteins may be only specificity factors in RNA editing processes. Supporting this idea, CRR4 binds specifically to sequences just upstream of the editing site (Okuda et al., 2007). However, about 90 PLS-PPR proteins in *Arabidopsis* have an additional and highly conserved C-terminal domain (~100 amino acids), termed the DYW domain because of its characteristic terminating tripeptide (Aubourg et al., 2000; Lurin et al., 2004). We recently demonstrated a striking correlation between the presence of the DYW domain and RNA editing in land plants and showed similarities between the DYW domain consensus and the active site of cytidine deaminases (Salone et al., 2007). Together, these observations lead us to propose that the DYW domain of PPR proteins of the PLS subfamily may provide the catalytic activity for editing reactions.

In this study, we investigated the role of the DYW1 protein which, because of its particular structure consisting only of a plastid targeting sequence and a DYW domain without any PPR motifs, is unique in *Arabidopsis*. Studying two RNAi mutant lines, we report that this protein is essential for RNA editing of numerous chloroplast transcripts. These findings suggest that the mutant plants may be affected in the enzymatic step of RNA editing, and lead us to hypothesise that the DYW1 protein is a catalytic factor of the editing machinery in chloroplasts recruited by RNA-binding specificity proteins, such as those PPR proteins that lack the DYW domain.

## **Material and Methods:**

### **Bioinformatics:**

A modified version of the FLAGdb++ database and interface (Samson et al., 2004) was used for much of the visualization needed for manual annotation of *At1g47580* locus, including display of PPR motifs, AGI gene models, cloned cDNA sequences and EuGene predictions (Foissac et al., 2003). Predotar version 1.03 (Small et al., 2004) and TargetP version 1.01 (Emanuelsson et al., 2000) were used for the prediction of the subcellular localisation of *DYW1*.

### **Plant material and growth conditions:**

*Arabidopsis thaliana* Columbia ecotype (Col-0) was used in this study. Plants grown in sterile conditions were plated on germination medium (GM) containing 0.5X Murashige and Skoog basal salts (Sigma) 1% (w/v) sucrose and when indicated, 10 $\mu$ M of dexamethasone or 50mg/L of kanamycin antibiotic. Seedlings and plants in soil were grown under controlled conditions (16h light:8h dark cycle, light intensity of 100 $\mu$ Mm<sup>-2</sup> sec<sup>-1</sup> at 22°C day/16°C night, humidity 80%).

*Nicotiana benthamiana* plants were grown in soil under controlled conditions: 16h light:8h dark cycle, light intensity of 100 $\mu$ Mm<sup>-2</sup> sec<sup>-1</sup> at 16°C, humidity 80%.

### **Subcellular localization of *DYW1* protein:**

The full length ORF (from the start to the last codon before the stop codon) or just the first 100 codons of the *DYW1* gene were PCR amplified from *Arabidopsis* genomic DNA using *Pfx* DNA polymerase (Invitrogen) and primers flanked with *attB1* and *attB2* sites (Supplementary Table 1). PCR products were cloned into the entry vector pDONR207 by Gateway® recombination (Invitrogen). The p0229-RFP binary destination vector was created (as described by (Lurin et al., 2004)) by modifying pGreenII 0229 (Hellens et al., 2000); and the sequence coding for *DYW1* protein or putative targeting sequence were shuttled into this vector by Gateway® recombination (Invitrogen) from the corresponding pDONR207 entry clone. RFP fusion proteins were transiently expressed in *Nicotiana benthamiana* leaves via agroinfiltration, as described by (Lurin et al., 2004). Protoplasts were prepared from leaf material (harvested 48h after infiltration) cut into small pieces and incubated in enzyme solution (Chupeau, 1974) at 30°C for 2 or 3 hours. Fluorescent proteins were visualised using a confocal microscope (Leica TCS-NT), allowing RFP detection (DsRed2 filter BP580/80, maximum Exciter/Emitter: 558 nm/583nm; GFP filter BP530/30, maximum Exciter/Emitter: 479nm/509nm) and natural fluorescence detection (LP590 filter, Emitter > 600nm). Signal were collected in independent photomultipliers.

### **Antibodies against *DYW1* protein:**

The last 570 base pairs of *DYW1* coding sequence were amplified and cloned into the entry vector pDONR207 by Gateway® recombination (Invitrogen), as described before. From the resulting plasmid, the *DYW1* fragment was introduced into the pDEST17 bacterial expression vector (Invitrogen). A 6xHis-tagged-*DYW1* fusion protein was produced in *E. coli* BL21-SI strain (Invitrogen) induced 10 hours at 28°C by adding NaCl to a final concentration of 0.3M. This protein was purified from *E. coli* extracts by affinity chromatography using Ni-NTA Agarose (Qiagen) under denaturing conditions. 1 mg of this purified protein was used to generate a polyclonal antibody by the Institute of Medical and Veterinary Science, Adelaide, following their usual procedure.

### **Protein gel blot analysis:**

Protein analyses were from total protein extracts, where indicated, from chloroplast-enriched fractions prepared from Col-0 leaves as described by Hegeman et al. (2005). Protein concentration was determined with Bradford reagent (BioRad). Proteins were separated by SDS-PAGE and transferred onto Immun-Blot PVDF Membrane (BioRad). To verify equal protein loading, samples were run on a second gel and stained with Coomassie Brilliant Blue R-250. Immunodetection was performed using the following dilutions of the polyclonal antibodies: 1:1000 for *DYW1* and 1:3000 for formate

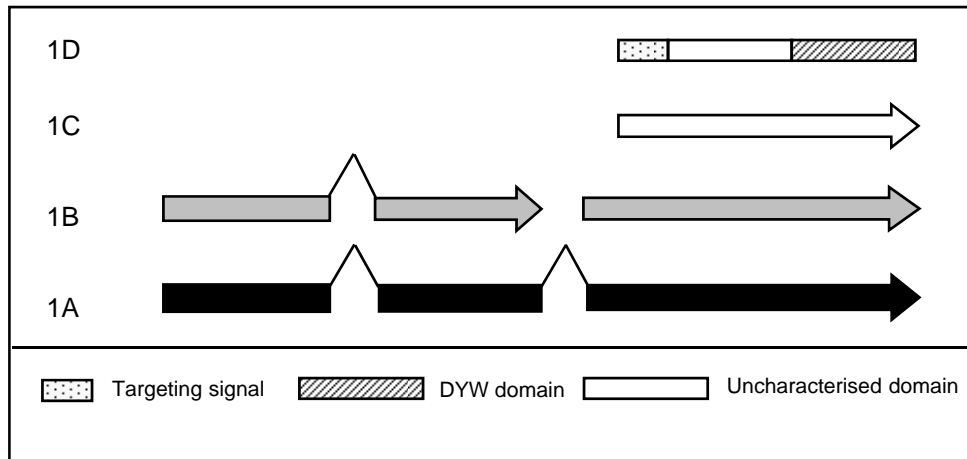


Figure 1: **DYW1 gene and protein structures.** Exons are represented as arrows, and introns by a line. The At1g47580 locus as described by its AGI model available in TAIR database (<http://www.arabidopsis.org>) (1A). Two genes predicted by EuGene for this locus. (1B) *DYW1* gene model proposed after examining EuGene predictions and cDNAs expressed from the At1g47580 locus: an intronless gene of 696 base pairs, from 17488133 to 17488829 position on chromosome 1 (1C). The corresponding DYW1 protein consists of a putative N-terminal plastid targeting signal and a C-terminal DYW domain (1D).

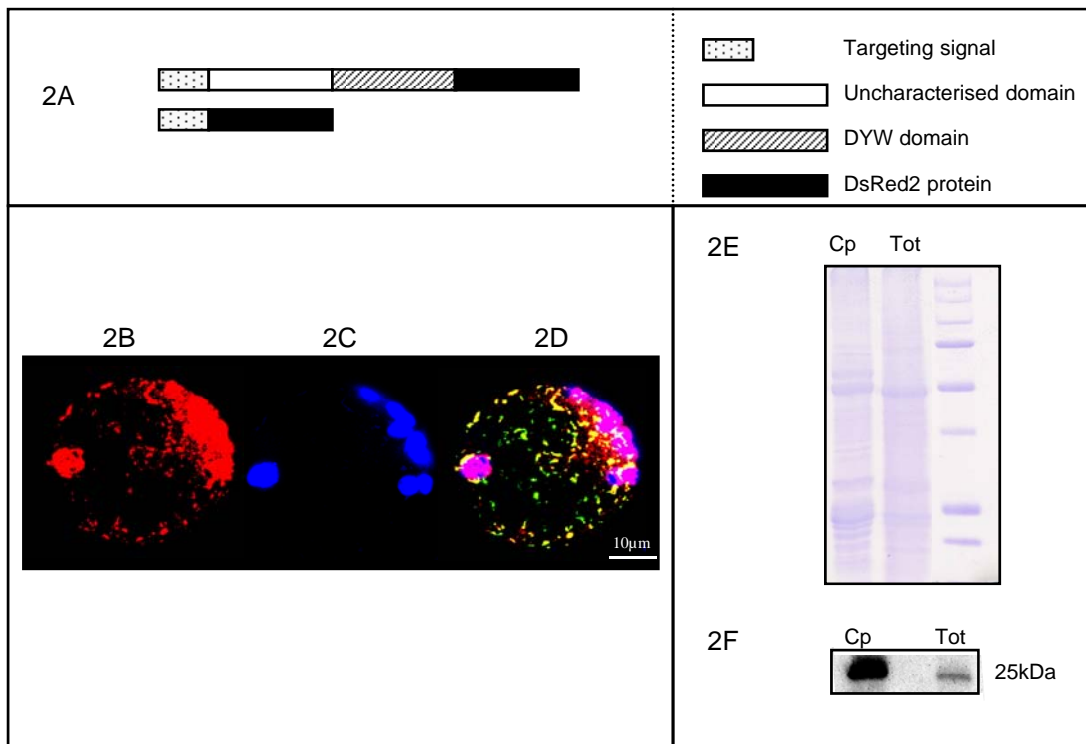


Figure 2: **Subcellular localisation of DYW1.**

Schematic representation of the two RFP fusion constructs used in subcellular localization assays (2A). Localisation of DsRed2 fluorescence in *N. benthamiana* protoplasts, observed under a confocal microscope, 48 hours after agroinfiltration of leaves by these two constructs (2B). The chlorophyll autofluorescence signal acquired separately (2C), allowed us to visualise the chloroplast localisation of the DsRed2 signal by superimposition of both signals (2D). Western blot analysis was done by loading 15µg of total protein extract or chloroplast-enriched fractions from 3 week-old *Arabidopsis* leaves. After transfer, membranes were hybridised with an antibody raised against recombinant DYW1 (2F). The coomassie blue-stained gel is shown as a loading control (2E).

dehydrogenase (FDH). Detection of the primary antibodies was by goat anti-rabbit IgG-alkaline phosphatase using the ECL Western Blotting Analysis System (GE Healthcare).

### ***DYWI RNAi lines***

The gene-specific sequence tag (GST) CATMA1a38560 (Crowe et al., 2003) was PCR amplified from *Arabidopsis* and cloned into pDONR207 Gateway® entry vector as described before. Then, using a double LR clonase reaction, this entry clone was recombined with both pAGRII *attR1-ccdB-attR2* cassettes (Figure 3A). The pAGRII expression vector is modified from pAGRIKOLA vector and carry the 6xpOp/LHGR system allowing inducible expression of the hpRNA in transgenic plant after dexamethasone application (Hilson et al., 2004; Craft et al., 2005; Wielopolska et al., 2005). Following AGRKOLA project procedure (Hilson et al., 2004), this construction was introduced into *A. tumefaciens* strain (C58C1). *Arabidopsis* transformation was performed by the conventional floral dip method (Clough and Bent, 1998). Seeds from transformed plants were germinated on germination medium under kanamycin selection. The integrity of the GST hairpin transgene, in selected plants, was monitored via multiplex PCR using AgriQ51, Agri56, Agri64 and Agri69 primers as described (Hilson et al., 2004).

### ***RNA preparation and DYWI transcript abundance:***

Total RNA was isolated from seedlings using the RNeasy plant minikit (Qiagen), and genomic DNA was removed using DNA-free DNase kit (Ambion). DNA-free RNA (3µg) was reverse transcribed with oligo(dT) primers. Quantitative PCR was performed with the LightCycler 480 Real-time PCR system (Roche) using the LightCycler 480 SYBR green 1 master mix and the primers 5' TGGAAAGTTGCATAAATGGCTTA 3', 5' AGATGGTAACTGCTCTTGTGGAG 3'. The data was normalized using constitutively expressed actin genes (AT3G18780 and AT1G49240) and corresponding primers: PQactin28\_F GGTAACATTGTGCTCAGTGGTGG and PQactin28\_R AACGACCTTAATCTTCATGCTGC.

### ***RNA preparation and analysis of RNA editing:***

Total RNA was isolated from seedlings, as described before, and DNA-free RNA (3µg) was reverse transcribed with random hexamers.

-Sequences including the editing sites were amplified by PCR with a primer set described in Supplementary Table 2. The RT-PCR products were sequenced directly.

-For high resolution melting (HRM) of PCR amplicons and analysis, we used the procedure and primer set described by Chateigner-Boutin and Small, 2007.

-Poisoned primer extension (PPE) of RT-PCR products and determination of editing efficiency were performed as previously described by Chateigner-Boutin and Small, 2007. PCR templates were generated using the primer set described in Supplementary Table 3. The PPE primers and corresponding ddNTPs used were detailed on the same Supplementary Table 3.

## ***Results:***

### ***DYWI: a unique ORF in Arabidopsis genome***

The At1g47580 locus, located on the first chromosome of *Arabidopsis thaliana* genome, is annotated as a putative lipoyltransferase, with an estimated full-length cDNA of 1251 bp (Figure 1A). The first exon of At1g47580 is almost identical to an exon of the lipoyltransferase *LIP2*. The second exon encodes a DYW domain, a highly conserved C-terminal domain found in about 90 PPR proteins (Aubourg et al., 2000; Lurin et al., 2004). Two cDNA clones covering this second exon (Salanoubat et al., 2000), ~680 bp in length, indicate it can be expressed without being spliced to the upstream exon, and indeed there is no evidence that these exons are ever spliced together. EuGene predictions (Foissac et al., 2003) indicate 2 separate genes at this locus. We propose that the *LIP2*-homologous exons



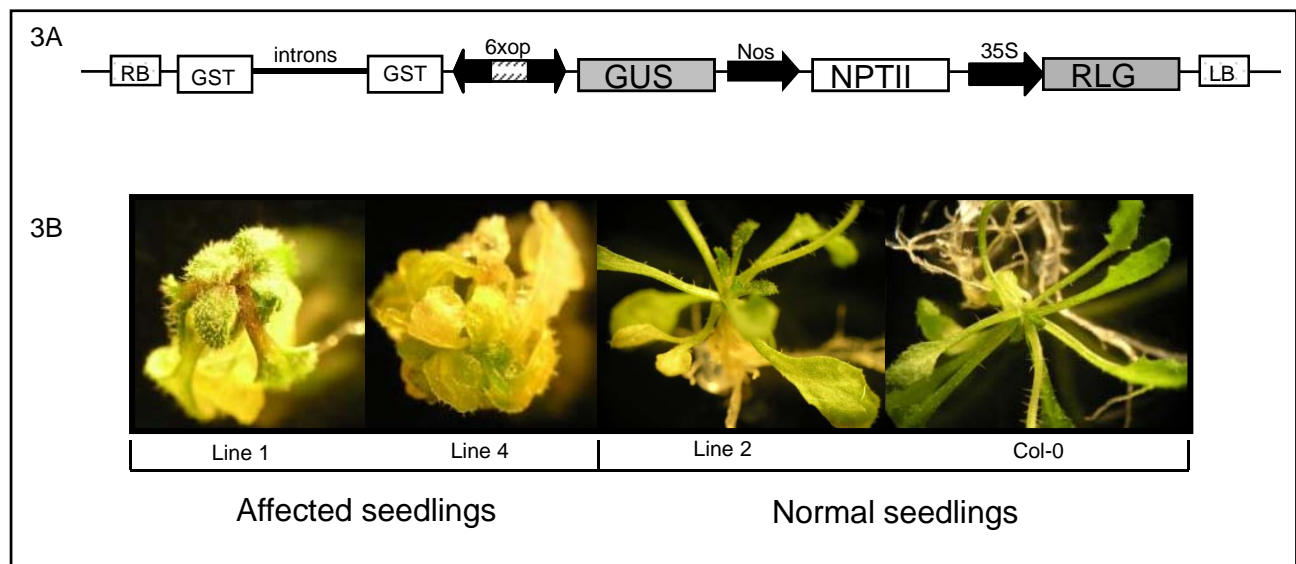


Figure 3: ***DYW1* RNAi lines**

(3A) Schematic representation of the construct used to produce *DYW1* RNAi lines. The CATMA1a38560 gene-specific sequence tag (GST) is cloned in an inverse repeated conformation, and separated by intron sequences that are spliced out from the RNA molecule, allowing expression of a hairpin RNA (hpRNA). This construct is under control of the inducible 6xpOp promoter, made of 6 repeats of the LhGR transcription factor fixation domain. The 6xpOp promoter is bidirectional and also controls the expression of the *GUS* reporter gene. Transgenic plants also carry, under control of the constitutive 35S CaMV promoter, a chimeric gene allowing expression of LhGR transcription factor. This chimeric protein, upon application of external glucocorticoid such as dexamethasone, will activate the 6xpOp promoter. (3B) Morphology of 21 day-old WT and T2 *DYW1* RNAi lines seedlings. Seedlings from line 1 and 4 showing an the severe phenotype seen in 10-30% of the lines, and seedlings from line 2 and WT showing normal growth.

corresponds to a pseudogene, but just downstream of it (from 17488133 to 17488829 on chromosome 1) an intronless gene is expressed encoding a DYW domain (Figure 1C). The putative protein encoded by this gene contains 239 amino acids (~25kDa), with a predicted 35 amino acid N-terminal plastid targeting sequence and a DYW domain (Figure 1D). This putative protein would not contain any PPR motifs, which makes this gene unique amongst the 90 DYW-containing genes in the *Arabidopsis* genome. For this reason, we named it *DYWI*.

### **A plastid subcellular localization for DYWI protein**

According to the Target P (Emanuelsson et al., 2000) and Predotar (Small et al., 2004) prediction programs, the N-terminal coding sequence of *DYWI* is predicted to target *DYWI* to plastids, with scores of 0.46 or 0.84 respectively.

The full-length protein (from the start to the last codon before the stop) and the first hundred amino acids of *DYWI* were fused to DsRed2 fluorescent protein and the subcellular localisation of the fusion proteins were experimentally analysed. Both fusion proteins consistently labelled plastids in transient expression experiments (Figure 2A, 2B).

An antibody was raised to a recombinant *DYWI* and used to probe western blots of total leaf extract and a chloroplast-enriched fraction of *Arabidopsis* leaves. A protein of the expected size (~25kDa) is strongly enriched in chloroplast extract, with respect to its concentration in total leaf extract (Figure 2C and 2D).

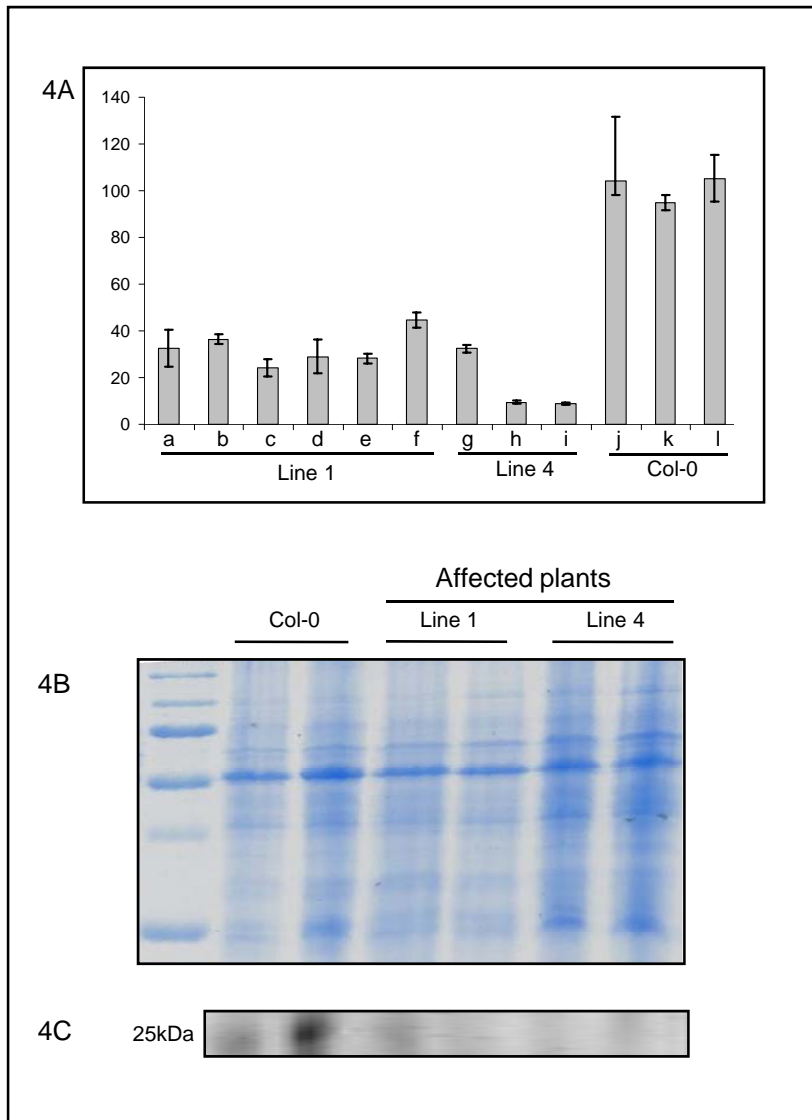
### **Phenotypic characterization of DYWI RNAi lines**

Two potential T-DNA insertion mutants in *DYWI* (SALK-123655, SALK-012425) were obtained, but after genotyping, we demonstrated that for both lines the T-DNA insertion is outside of the *DYWI* ORF (data not shown). Therefore we employed an RNAi approach using the inducible pAGRII expression vector developed for the AGRICOLA project (Hilson et al., 2004) (Karimi and Hilson, unpublished) (Figure 3A).

Five individual T1 transformants were isolated, and four of them showed an altered phenotype compared to wild type plants, especially a growth delay and reduced fertility (data not shown). For 3 of these 5 primary transformants viable T2 RNAi lines were obtained. Phenotypic characterization was performed on plants grown on synthetic media with or without 10 $\mu$ M dexamethasone inducer. For the line 1 and 4, respectively, about 10% and 30% of seedlings showed an extremely severe phenotype (Figure 3B). Most of the affected plants died within 3 weeks. The rate of appearance of this phenotype was unaffected by the presence or absence of dexamethasone in the medium and uncorrelated with either hairpin RNA levels or GUS expression (data not shown), and thus probably corresponds to transcriptional silencing rather than post-transcriptional silencing.

### **A decrease in DYWI transcript and protein accumulation in severely affected plants from DYWI RNAi lines:**

The transcript abundance of *DYWI* from affected and wild-type seedlings was measured by qRT-PCR. Results shown a strong decrease in *DYWI* transcripts in affected seedlings from both RNAi lines compared to wild type seedlings (Figure 4A). Interestingly, in line 4, that exhibited the highest frequency of affected plants, we obtained seedlings with less than 10% of the wild-type *DYWI* transcript level, whereas in line 1, we never observed decreases to below 24% of wild-type. In the third line with no visible phenotype, in the same analysis, we



**Figure 4: *DYW1* transcript and protein accumulation.**

*DYW1* transcript abundance from 21-day-old affected seedlings (Line 1 and 4) and WT seedlings was measured by quantitative PCR in individual plants. The average abundance of *DYW1* transcript in the 3 Col0 seedlings was set to 100% after normalization using actin transcripts as a reference. The standard deviation was calculated from 3 technical replicates (4A).

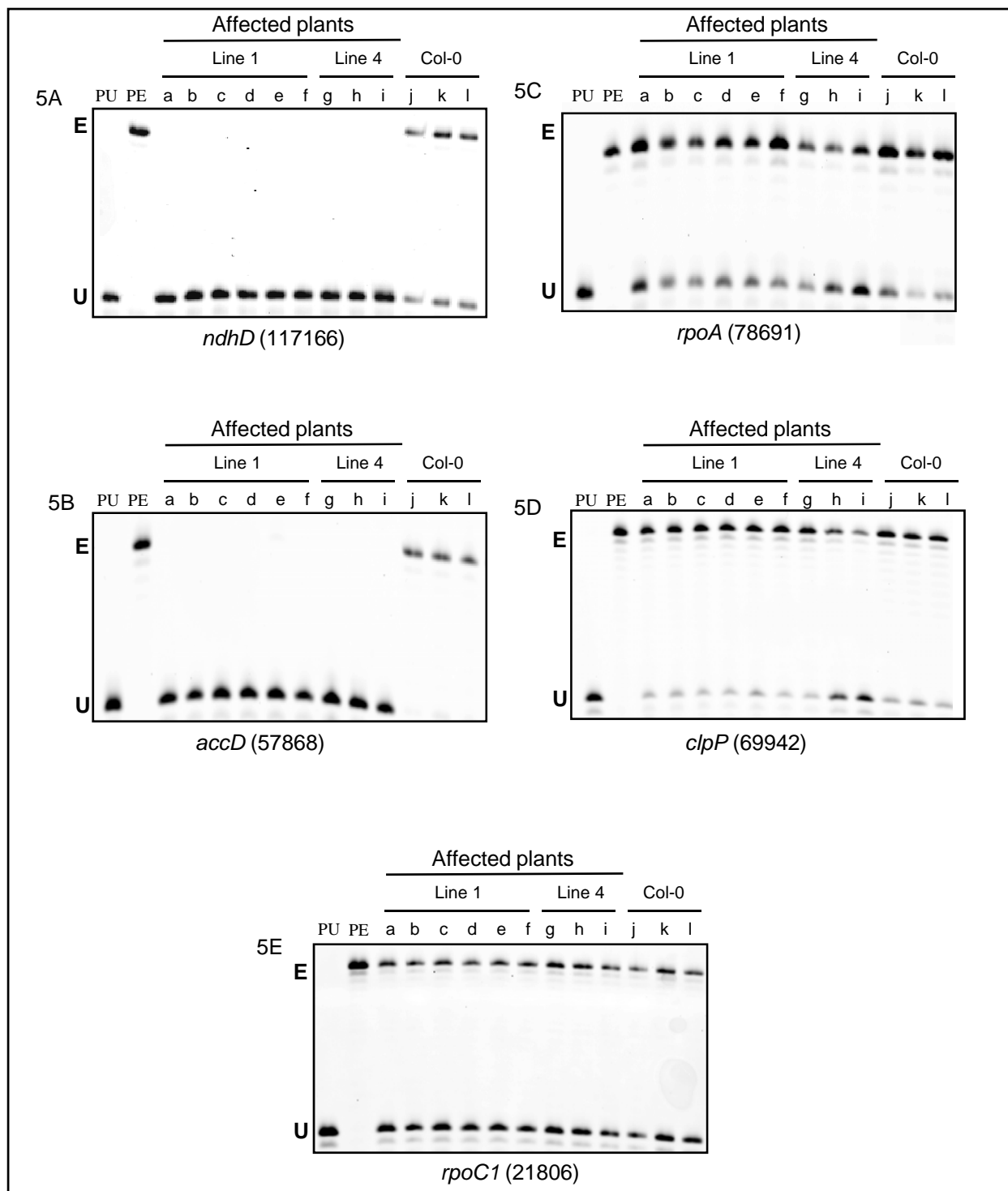
Western blot analyses were done by loading 20µg of total protein extracts from 21-day-old affected seedlings (Line 1 and 4) and WT seedlings. Membranes were hybridised with anti-*DYW1* antibody, revealing a 25kDa signal in WT seedlings (4C). The coomassie blue-stained gel is shown as a loading control (4B).

Locus (Transcript)	Primer pair	genome position	AA change (CDS)	Direct cDNA sequencing	HRM	PPE (range value in %)		
						<i>AtDYW1</i> RNAi line1	<i>AtDYW1</i> RNAi line4	WT
AtCg00040 (matK)	Chloro 260	2931	H>Y	NI				
AtCg00130 (atpF)	Chloro 27	12707	P>L	NI				
AtCg00180 (rpoC1)	Chloro 39F-21861R	21806	S>L					
AtCg00170 (rpoB)	23581F-24093R	23898	S>L	I		0	0	94-99
AtCg00170 (rpoB)	Chloro 45	25779	S>L	PI				
AtCg00170 (rpoB)	Chloro 45	25992	S>L	PI				
AtCg00300 (psbZ)	Chloro 68	35800	S>L		yes	80-95	59-94	93-95
AtCg00330 (rps14)	Rps14.AT.rev-chloro71R	37092	P>L					
AtCg00330 (rps14)	Chloro 72	37161	S>L					
AtCg00500 (accD)	Chloro 112	57868	S>L	PI	yes	0	0	88-92
AtCg00500 (accD)	Chloro 113	58642						
AtCg00570 (psbF)	Chloro 125	63985	S>F	NI				
AtCg00580 (psbE)	Chloro 126	64109	P>S	NI				
AtCg00590 (petL)	Chloro 130	65716	P>L	I		40-72	0	69-73
AtCg00650 (rps12)	Chloro 140	69553			yes	78-84	16-20	77-84
AtCg00670 (clpP)	Chloro 141	69942	H>Y	PI		86-92	38-92	81-92
AtCg00740 (rpoA)	Chloro 159-286	78691	S>F		yes	63-78	42-67	78-82
AtCg00840 (rpl23)	Chloro 178	86056	S>L					
AtCg00890 (ndhB)	Chloro 185	949999	P>L	I				
AtCg00890 (ndhB)	Chloro 186	95225	H>Y					
AtCg00890 (ndhB)	Chloro 187	95608	S>L					
AtCg00890 (ndhB)	Chloro 187	95644	S>L					
AtCg00890 (ndhB)	Chloro 187	95650	S>L					
AtCg00890 (ndhB)	Chloro 189	96419	S>F					
AtCg00890 (ndhB)	Chloro 189	96579	H>Y					
AtCg00890 (ndhB)	Chloro 189	96698	P>L	PI				
AtCg00890 (ndhB)	Chloro 190	97016	S>L	PI				
AtCg01010 (ndhF)	Chloro 204	112349	S>L	NI	yes	73-94	55-95	90-95
AtCg01050 (ndhD)	Chloro 212	116281	P>L	NI				
AtCg01050 (ndhD)	Chloro 212	116290	S>L	PI				
AtCg01050 (ndhD)	Chloro 212	116494	S>L	PI				
AtCg01050 (ndhD)	Chloro 213	116785	S>L					
AtCg01050 (ndhD)	Chloro 214	117166	>M	I		0	0	66-67
AtCg01080 (ndhG)	Chloro 218	118858	S>F					

**Table 1: List of Chloroplasmic Editing Defects in “twisty” plants**

The 34 currently known editing sites in *A. thaliana* plastid genes are listed in this table. The transcript affected is given in the first column, and for each site the primers used to amplify the cDNA sequence containing the editing site are given. The position on the chloroplast genome is given in column 3 for each editing site, and the resulting alteration in the coding sequence.

The results for 18 RNA editing sites, analyzed by direct cDNA sequencing, from individual affected seedlings from both *DYW1* RNAi lines are indicated (column 5) and divided into 3 classes considering a rough measure of editing level: impaired editing sites (I), putative impaired editing sites (PI) or unimpaired editing sites (NI). An HRM screen on all known editing sites was performed and to look for differences in melting curves for a mix of gDNA and wild type cDNA versus a mix of gDNA and cDNA from affected plants. This screen indicated 5 putative impaired editing sites (indicated by yes in column 6). Quantification of editing level was conducted, by PPE analysis, on nine editing sites for individual affected seedlings from line 1 (column 7) and line 4 (column 8) and WT (column 9). The results are indicated as percentage range.



### Figure 5: Examples of Chloroplast Editing Defects in "twisty" plants

Poisoned primer extension assays were performed on known editing sites. Data for individual 21-day-old affected seedlings and WT seedlings (the same samples as those presented in the qPCR analysis shown in figure 4A) are shown. The editing sites shown are: (5A) *ndhD* (117166), (5B) *accD* (57868), (5C) *rpoA* (78691), (5D) *clpP* (69942) and (5E) *rpoC1* (21806). RT-PCR products were obtained with primers surrounding the editing sites and served as templates for an extension reaction from a 5' FAM-PPE primer which anneals near to the target editing site. The extension was stopped by the incorporation of ddCTP/ddGTP at the location of the editing site (producing a short product, if the template is unedited, indicated by a "U" on the gel) or at the subsequent C/G (producing a longer product, if the template is edited, indicated by an "E" on the gel). As technical controls, PPE was performed on plasmids containing an edited (PE) or unedited (PU) clone of interest. We observed editing defects for several sites including *ndhD* (117166), *accD* (57868), *rpoA* (78691), *clpP* (69942) as shown in this figure. *rpoC1* (21806) is an example of a unimpaired editing site characterised in affected seedlings.

were unable to detect any significant variation in *DYW1* transcript abundance compared to wild type plants (data not shown).

The accumulation of DYW1 protein was also investigated by western blotting on total protein extract from affected seedlings from both *DYW1* RNAi lines and wild type seedlings using an antibody raised against recombinant DYW1. No signal was detected in affected seedlings, while a protein of the expected size (~25kDa) is detectable in wild type seedlings (Figure 4D).

### **RNA editing of several chloroplast transcripts are impaired in severely affected plants from *DYW1* RNAi lines**

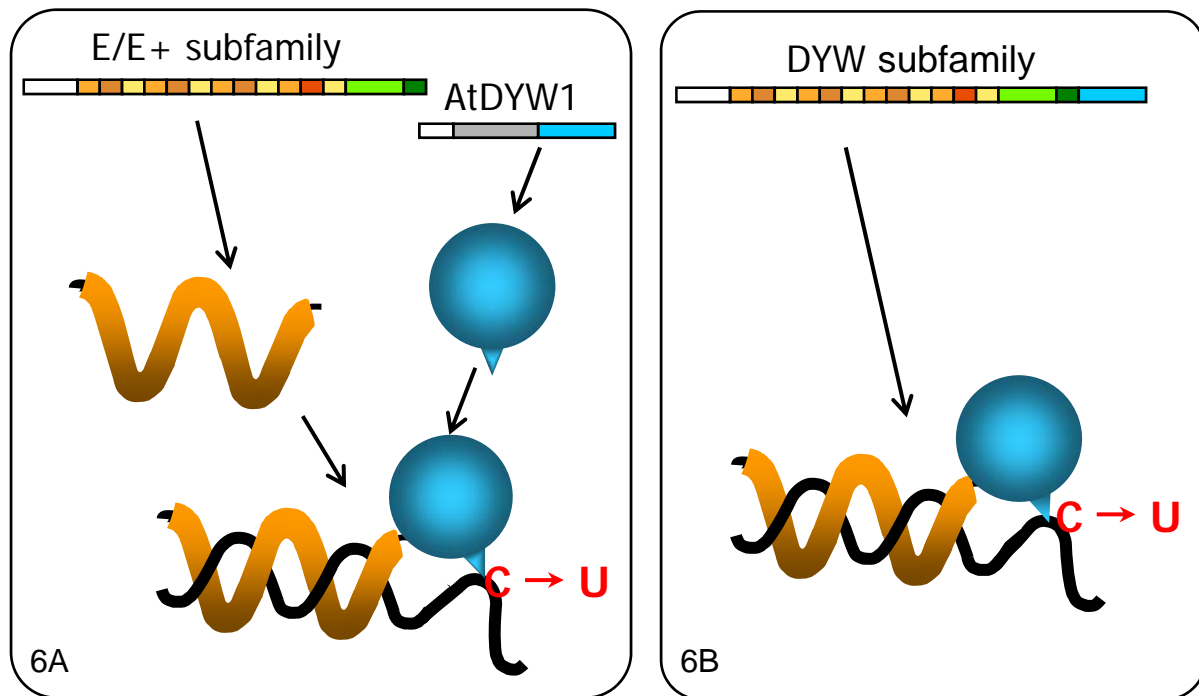
We previously proposed that the DYW domain of PPR proteins in plants may have a role in catalyzing editing reactions (Salone et al., 2007), and this led us to consider that RNA editing could be impaired in *DYW1* RNAi lines 1 and 4.

To assess this possibility, editing was evaluated using two systematic methods: direct cDNA sequencing and high resolution melting (HRM). A decrease in RNA editing was suggested for numerous sites (Table 1). Nine candidate sites were verified by poisoned primer extension assays (PPE). For 3 editing sites - *ndhD* (genomic position 117166), *accD* (57868) and *rpoB* (23898) transcript - PPE results demonstrated a total loss of RNA editing for all affected seedlings from both *DYW1* RNAi lines (Figure 5A, 5B and Table 1). For the 6 other editing sites, we detected a partial decrease in RNA editing for several, but not all, affected seedlings from *DYW1* RNAi lines 1 and 4 (Figure 5C, 5D and Table 1).

### ***Discussion:***

Our data demonstrated that DYW1 protein is a key component in RNA editing machinery in *Arabidopsis* chloroplasts. Systematic screens of the known editing sites in chloroplast transcripts revealed that RNA editing is impaired when DYW1 expression is drastically decreased. RNA editing is totally abolished for 3 of the tested RNA editing sites (in *ndhD*, *accD* and *rpoB* transcripts), and partially impaired for 6 other sites. These observations indicate that DYW1 is required for RNA editing at at least 9 sites in *Arabidopsis* chloroplasts. These widespread effects suggest that *DYW1* RNAi mutants may be affected in a step common for many sites. To date, all PPR proteins described as essential for RNA editing in chloroplasts belong to the E/E+ subgroup of the PLS-PPR subfamily (Kotera et al., 2005; Okuda et al., 2007)(Chateigner-Boutin *et al.*, submitted), and do not contain any domain likely to have catalytic activity (Aubourg et al., 2000; Lurin et al., 2004). Based on phylogenetic distribution and sequence similarity with known cytidine deaminases, we proposed that the DYW domain may catalyse the editing reaction (Salone et al., 2007; O'Toole et al., 2008). These observations lead us to propose that DYW1 protein could be recruited by E/E+ PLS-PPR proteins and provide the catalytic activity necessary for RNA editing in *Arabidopsis* chloroplasts (Figure 6A). Interestingly the DYW domain of DYW1 protein is one of the closest to the *Arabidopsis* DYW domain consensus suggesting high selection pressure on this gene sequence (data not shown).

Supporting this model, in *DYW1* RNAi lines, we found that RNA editing is totally abolished for *ndhD* (gp: 117166) site and partially impaired for *clpP* (gp: 69942 ) and *rpoA* (gp: 78691) sites (Figure 5 and table 1). The PPR protein CRR4 is essential for RNA editing of this *ndhD* site (Kotera et al., 2005) and the PPR protein CLB19 is essential for RNA editing of both the *clpP* and *rpoA* sites (Chateigner-Boutin et al., submitted). The incomplete inhibition of *DYW1* expression in the RNAi mutants may explain the fact that we were unable to detect any defect in editing for the *ndhD* (116785) site that requires CRR21 (Okuda et al., 2007).



**Figure 6: A model for editing machinery in chloroplasts**

(6A) E/E+ PLS-PPR proteins, such as CRR4, CRR21 or CLB19, may act as specificity factors to bind RNA molecules and recruit the DYW1 protein as a catalytic partner to edit RNA molecules. (6B) DYW PLS-PPR proteins may act as specificity factors to bind RNA molecules and use their DYW domain to edit RNA molecules without a requirement for DYW1.

The *ndhD* gene encodes a subunit of the chloroplast NAD(P)H dehydrogenase (NDH) complex, and the lack of RNA editing in the *ndhD* (117166) site was previously described for *Arabidopsis crr4* mutants that are defective in NDH activity, but with normal growth and development (Kotera et al., 2005). Consequently, this molecular defect is unlikely to be responsible for the strong phenotype observed in *DYW1* RNAi lines. However, the *accD* gene encodes the carboxytransferase beta subunit of the acetyl-CoA carboxylase (ACCase) complex in plastids, and an *Arabidopsis* mutant in the CAC1 subunit of the ACCase complex is morphologically affected (Hui-Rong Qian, 1998). Moreover, the *rpoB* gene encodes the beta subunit of the plastid encoded RNA polymerase (PEP). Studies of tobacco and barley mutants for the corresponding gene revealed albino plants with defective plastids (Inada et al., 1997; De Santis-Maclossek G, 1999). For both editing sites, RNA editing restores conserved amino acids in protein sequences. It is reasonable to propose that a lack of RNA editing at these sites could result in an absence of two functional proteins and lead to the dramatic effects seen in the affected plants.

The large number of E and DYW-class PPR genes in the rice and *Arabidopsis* genomes is correlated with a similarly large number of editing sites in these two species. A conserved putative orthologue of *DYW1* gene is found in a single copy in the rice genome (O'Toole et al., 2008)(and unpublished data). Consequently, *DYW1*-like proteins may be a key component in RNA editing machinery conserved in all angiosperm plastids. However, the genome of the moss *Physcomitrella patens* contains no *DYW1*-like protein, no E/E+ PPR proteins, and only 10 DYW PPR proteins, in correlation with the very few editing sites described for this species (O'Toole et al., 2008).

It is currently thought that RNA editing appeared at the emergence of land plants and almost simultaneously in both organelles, suggesting a common origin of the RNA editing machinery in mitochondria and chloroplasts. E/E+ and DYW PPR proteins are present in both organelles of many angiosperms, therefore a dual localisation, in mitochondria and chloroplasts, for *DYW1* and *DYW1*-like proteins needs to be investigated. The rice and *Arabidopsis* genomes encode several short mitochondrial DYW proteins that contain only one or two repetitions of the PPR motif (Lurin et al., 2004; O'Toole et al., 2008). These proteins might act as *DYW1*-like proteins in mitochondria.



## References:

- Aubourg, S., Boudet, N., Kreis, M., and Lecharny, A.** (2000). In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol Biol* **42**, 603-613.
- Chateigner-Boutin, A.L., and Small, I.** (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res* **35**, e114.
- Chupeau, Y., Bourgin, J.P., Missonier, C., Dorion, N., and Morel, G.** (1974). Préparation et culture de protoplastes de divers *Nicotiana*. *C.R. Acad. Sci. Paris* **278**, 1565-1568.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**, 735-743.
- Craft, J., Samalova, M., Baroux, C., Townley, H., Martinez, A., Jepson, I., Tsiantis, M., and Moore, I.** (2005). New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in *Arabidopsis*. *Plant J* **41**, 899-918.
- Crowe, M.L., Serizet, C., Thareau, V., Aubourg, S., Rouze, P., Hilson, P., Beynon, J., Weisbeek, P., van Hummelen, P., Reymond, P., Paz-Ares, J., Nietfeld, W., and Trick, M.** (2003). CATMA: a complete *Arabidopsis* GST database. *Nucleic Acids Res* **31**, 156-158.
- De Santis-Maclossek G, K.W., Bock A, Schoch S, Maier RM, Wanner G, Rüdiger W, Koop HU, Herrmann RG.** (1999). Targeted disruption of the plastid RNA polymerase genes *rpoA*, *B* and *C1*: molecular biology, biochemistry and ultrastructure. *PLANT J*. **18**, 477-489.
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G.** (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**, 1005-1016.
- Foissac, S., Bardou, P., Moisan, A., Cros, M.J., and Schiex, T.** (2003). EUGENE'HOM: A generic similarity-based gene finder using multiple homologous sequences. *Nucleic Acids Res* **31**, 3742-3745.
- Giege, P., and Brennicke, A.** (1999). RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A* **96**, 15324-15329.
- Handa, H.** (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res* **31**, 5907-5916.
- Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M.** (2000). pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol Biol* **42**, 819-832.
- Hilson, P., Allemeersch, J., Altmann, T., Aubourg, S., Avon, A., Beynon, J., Bhalerao, R.P., Bitton, F., Caboche, M., Cannoot, B., Chardakov, V., Cognet-Holliger, C., Colot, V., Crowe, M., Darimont, C., Durinck, S., Eickhoff, H., de Longevialle, A.F., Farmer, E.E., Grant, M., Kuiper, M.T., Lehrach, H., Leon, C., Leyva, A., Lundeberg, J., Lurin, C., Moreau, Y., Nietfeld, W., Paz-Ares, J., Reymond, P., Rouze, P., Sandberg, G., Segura, M.D., Serizet, C., Tabrett, A., Taconnat, L., Thareau, V., Van Hummelen, P., Vercruyssen, S., Vuylsteke, M., Weingartner, M., Weisbeek, P.J., Wirta, V., Wittink, F.R., Zabeau, M., and Small, I.** (2004). Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications. *Genome Res* **14**, 2176-2189.

- Hui-Rong Qian, B.J.N., Eve Syrkin Wurtele** (1998). Molecular genetics of acetyl-CoA carboxylase. In 9TH INTERNATIONAL CONFERENCE ON ARABIDOPSIS RESEARCH.
- Inada, H., Seki, M., Morikawa, H., Nishimura, M., and Iba, K.** (1997). Existence of three regulatory regions each containing a highly conserved motif in the promoter of plastid-encoded RNA polymerase gene (*rpoB*). *Plant J* **11**, 883-890.
- Kotera, E., Tasaka, M., and Shikanai, T.** (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**, 326-330.
- Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I.** (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089-2103.
- Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K.** (2002). The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics* **268**, 434-445.
- O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and Small, I.** (2008). On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* **25**, 1120-1128.
- Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T.** (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A* **104**, 8178-8183.
- Salanoubat, M., Lemcke, K., Rieger, M., Ansoerge, W., Unseld, M., Fartmann, B., Valle, G., Blocker, H., Perez-Alonso, M., Obermaier, B., Delseny, M., Boutry, M., Grivell, L.A., Mache, R., Puigdomenech, P., De Simone, V., Choisne, N., Artiguenave, F., Robert, C., Brottier, P., Wincker, P., Cattolico, L., Weissenbach, J., Saurin, W., Quetier, F., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Benes, V., Wurmbach, E., Drzonek, H., Erfle, H., Jordan, N., Bangert, S., Wiedelmann, R., Kranz, H., Voss, H., Holland, R., Brandt, P., Nyakatura, G., Vezzi, A., D'Angelo, M., Pallavicini, A., Toppo, S., Simionati, B., Conrad, A., Hornischer, K., Kauer, G., Lohnert, T.H., Nordsiek, G., Reichelt, J., Scharfe, M., Schon, O., Barges, M., Terol, J., Climent, J., Navarro, P., Collado, C., Perez-Perez, A., Ottenwalder, B., Duchemin, D., Cooke, R., Laudie, M., Berger-Llauro, C., Purnelle, B., Masuy, D., de Haan, M., Maarse, A.C., Alcaraz, J.P., Cottet, A., Casacuberta, E., Monfort, A., Argiriou, A., flores, M., Liguori, R., Vitale, D., Mannhaupt, G., Haase, D., Schoof, H., Rudd, S., Zaccaria, P., Mewes, H.W., Mayer, K.F., Kaul, S., Town, C.D., Koo, H.L., Tallon, L.J., Jenkins, J., Rooney, T., Rizzo, M., Walts, A., Utterback, T., Fujii, C.Y., Shea, T.P., Creasy, T.H., Haas, B., Maiti, R., Wu, D., Peterson, J., Van Aken, S., Pai, G., Militscher, J., Sellers, P., Gill, J.E., Feldblyum, T.V., Preuss, D., Lin, X., Nierman, W.C., Salzberg, S.L., White, O., Venter, J.C., Fraser, C.M., Kaneko, T., Nakamura, Y., Sato, S., Kato, T., Asamizu, E., Sasamoto, S., Kimura, T., Idesawa, K., Kawashima, K., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakayama, S., Nakazaki, N., Shinpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S.** (2000). Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* **408**, 820-822.

- Salone, V., Rudinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C.** (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* **581**, 4132-4138.
- Samson, F., Brunaud, V., Duchene, S., De Oliveira, Y., Caboche, M., Lecharny, A., and Aubourg, S.** (2004). FLAGdb++: a database for the functional analysis of the Arabidopsis genome. *Nucleic Acids Res* **32**, D347-350.
- Shikanai, T.** (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* **63**, 698-708.
- Small, I., Peeters, N., Legeai, F., and Lurin, C.** (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**, 1581-1590.
- Small, I.D., and Peeters, N.** (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* **25**, 46-47.
- Wielopolska, A., Townley, H., Moore, I., Waterhouse, P., and Helliwell, C.** (2005). A high-throughput inducible RNAi vector for plants. *Plant Biotechnol J* **3**, 583-590.

## **IV Mutants surexprimant la protéine AtDYW1**

### **IV.1 Contexte de l'étude**

Face à l'enjeu que représente l'étude de la protéine AtDYW1, et les difficultés d'obtention de lignées pour lesquelles l'expression de ce gène est totalement ou partiellement abolie, j'ai initié au cours de ma thèse une autre stratégie de « mutagenèse ». L'objectif était d'établir des lignées surexprimant la protéine AtDYW1 sous sa forme « sauvage » ou présentant des mutations ponctuelles de résidus conservés du motif DYW (lignées dites « dominant négatif »).

Comme nous l'avons déjà évoqué, le domaine DYW présente des homologies avec le site catalytique de désaminases fixant le zinc. Ainsi le changement d'un ou d'un petit nombre d'acides aminés conservés dans le site catalytique supposé de la protéine AtDYW1 doit pouvoir abolir sa fonction. De plus, à l'image des autres désaminases, AtDYW1 pourrait agir sous forme de dimère ou multimère dans la cellule végétale afin catalyser la réaction d'édition. Enfin, d'après notre modèle, cette protéine pourrait interagir avec des protéines PPR (et d'autres facteurs) lors des réactions d'éditations. Ces arguments rendent les stratégies de surexpression attrayantes pour étudier le rôle de la protéine AtDYW1. En effet, un grand nombre de copies surnuméraires de la protéine AtDYW1 fonctionnelle ou supposée non fonctionnelle (forme mutée) vont s'exprimer dans les cellules, et éventuellement en se fixant au(x) partenaire(s) sauvages(s) déréguler, bloquer ou diminuer l'activité du complexe d'édition.

Les approches génétiques par des stratégies de surexpression ou dites « dominant négatif » sont couramment utilisées chez un grand nombre d'organismes modèles, notamment pour étudier le rôle de protéines impliquées dans des « cascades de régulation ». Par exemple récemment, le rôle de la protéine MAP kinase MPK6 a été ainsi étudié chez *Arabidopsis* (Bush and Krysan, 2007).

### **IV.2 Mutant de surexpression de première génération**

#### *IV.2.1 Etablissement des lignées de surexpression*

La forme pleine longueur de l'ORF *AtDYW1* (ATG2/stop) a été clonée dans un vecteur donneur (pDONR207) du système Gateway, puis une réaction de recombinaison de LR clonase a permis le clonage directionnel de cette ORF dans un vecteur de destination pGWB2. Ce vecteur permet l'expression du gène d'intérêt dans la plante, sous contrôle du promoteur fort et constitutif (p35s-CaMW). Ce vecteur a été introduit dans une souche *d'A. tumefaciens* afin de transformer des plantes d'*Arabidopsis* (écotype WS).

#### *IV.2.2 Sélection des transformants obtenus*

Les graines ont été semées sur un milieu de culture *in vitro* (½ MS) contenant soit de l'antibiotique kanamycine, soit de l'antibiotique hygromycine. 42 transformants (T1) indépendants ont été ainsi sélectionnés. Après cette sélection, les plantules d'intérêts ont été

Plantes de la lignée H4



Plante de la lignée C1



Figure n°72 : Lignées de surexpression du gène *AtDYW1* : Observation phénotypique des plantes de la génération T2 (4 semaines après transfert en sol)

Le phénotype est complexe: retard de croissance, anormalité des tiges et des feuilles, nombre de feuilles de rosettes surnuméraires, perte de la dominance apicale...

repiquées en terre et les graines de ces plantes ont été récoltées séparément de façon à obtenir 42 lignées (T2) indépendantes.

Afin de réduire les risques de sélectionner des lignées ayant intégré une cassette non complète, nous avons tiré parti du fait que les gènes marqueurs de sélection (gène de résistance aux antibiotiques kanamycine et hygromycine) sont positionnés de part et d'autre de la construction d'intérêt. Ainsi les 42 lignées de génération T2 ont été soumises à une nouvelle sélection sur un milieu de culture *in vitro* (½ MS) contenant l'antibiotique opposé à celui utilisé lors de la sélection à la génération précédente. Une centaine de graines de chaque lignée ont été semées et la ségrégation a été examinée. On attend théoriquement ¾ de plantes résistantes et ¼ de plantes sensibles (en cas d'une insertion unique).

A la suite de cette deuxième sélection, nous avons obtenu :

- 6 lignées (T2) ne présentant aucune plante résistante, ce qui suggère un remaniement de la cassette d'intérêt au moment de son intégration dans le génome de la plante de sorte qu'elle ne soit pas complète. L'étude de ces lignées n'a pas été plus approfondie.

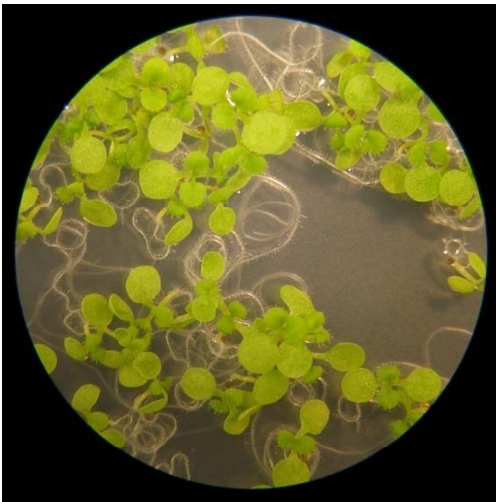
- 7 lignées (T2) ne présentant que des plantes résistantes (ou quasiment), cette ségrégation pourrait être le reflet d'insertion multiple de la cassette d'intérêt dans le génome de la plante. L'étude de ces lignées n'a pas été plus approfondie.

- 25 lignées (T2) présentant des plantes sensibles et résistantes dans la proportion théorique attendue. Ces lignées ont été retenues pour mener des analyses plus approfondies. Une dizaine de plantules résistantes de chacune de ces lignées a été alors repiquée en terre et montée en graine (récoltées plante à plante) afin de constituer la génération T3.

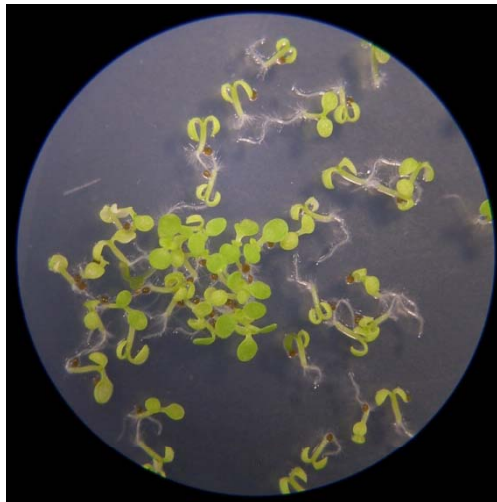
- et enfin, 4 lignées (T2) présentant des plantes sensibles et résistantes, avec une proportion de plantes sensibles supérieures à ¼. Ce fait a attiré notre attention et nous avons proposé trois hypothèses pour expliquer cette observation. (1) Il est souvent décrit dans ce genre d'approche génétique, qu'essayant de surexprimer un gène d'intérêt dans une plante, on déclenche en réalité une cascade de réactions conduisant à l'extinction de l'expression non seulement de la cassette transgénique mais aussi de la copie endogène du gène d'intérêt (phénomène dits de « co-suppression »). Par ailleurs, il a été rapporté que la présence des cassettes gateway de part et d'autre de l'ORF d'intérêt dans le vecteur pGWB2 pourrait favoriser l'apparition d'une structure en ARN double-brin et favoriser le déclenchement d'une telle cascade de réactions chez les plantes transformées par cette construction. Forts de ces connaissances, nous pouvons alors nous demander si la présence d'un grand nombre de plantules sensibles, dans ces 4 lignées, pourrait être liée au fait qu'un phénomène de co-suppression soit activé dans ces plantules lignées. (2) Nous savons que la protéine *AtDYWI* est localisée dans les chloroplastes. Chez *Arabidopsis*, un défaut de fonctionnement des chloroplastes peut conduire, en culture *in vitro*, à l'apparition de plantes albinos au développement très affecté. Or, lors de la sélection sur un milieu de culture *in vitro* contenant les antibiotiques kanamycine ou hygromycine, les plantules sensibles qui sont blanches et présentent un développement très affecté. A notre niveau d'observation, ces deux phénotypes ne sont pas distinguables, et ceci pourrait expliquer la présence surnuméraire de plantules d'apparence « sensible ». (3) Une autre explication consiste à penser que la fertilité des gamètes mâles ou femelles dans ces lignées soit affectée. De ce fait, une mauvaise transmission de l'insertion d'intérêt au cours des générations est envisageable.

Aussi, nous avons fait le choix de poursuivre l'analyse de ces 4 lignées (notée A10, H4, C1, et A7). Une dizaine de plantules résistantes de chacune de ces lignées a donc été alors repiquée en terre et montée en graine (récolte plante à plante) afin de constituer la génération T3.

A



B



C



Figure n°73: Lignées de surexpression du gène *AtDYW1* : Observation phénotypique des plantes de la génération T2, germées sur un milieu de culture (1/2 MS)

- a) Lignée dont les plantules ne présentent pas de phénotype
- b) Lignée pour laquelle deux populations de plantes apparaissent : des plantules sans phénotype et des plantules plus petites et vert-pâle.
- c) Agrandissement d'une plantule vert-pâle

#### IV.2.3 Phénotypage et génotypage des 29 lignées en génération T2

Le phénotype de plantes (T2) repiquées en terre est alors examiné. 4 semaines après transfert, pour 14 lignées, plusieurs plantes présentent un phénotype complexe : retard de croissance, anomalies des feuilles, perte de la dominance apicale... (figure n°72). De façon, intéressante, les 4 lignées ne présentant pas une ségrégation  $\frac{3}{4}$ ,  $\frac{1}{4}$ , semblent être les plus affectées.

Parallèlement à cette étude phénotypique, une feuille de deux plantes de chacune de ces 29 lignées T2 a été prélevée et placée dans une plaque 96 puits, pour réaliser une extraction d'ADN génomique et une PCR de génotypage utilisant une amorce dans le promoteur p35s et une amorce dans la région terminale de l'ORF *AtDYW1*. Une amplification de taille attendue (1065pb) a été obtenue pour toutes les lignées testées alors qu'aucune amplification n'est observée avec l'ADNg de WS (témoin négatif).

#### IV.2.4 Analyse des lignées de surexpression en T3

Afin d'affiner les observations faites à la génération T2, nous avons poursuivi nos analyses sur la génération T3, en concentrant nos efforts sur les 14 lignées ayant présenté un phénotype à la génération précédente.

##### \* Analyse de ségrégation

Nous avons débuté par une analyse de ségrégation avec deux objectifs en tête, d'une part, savoir s'il était possible d'isoler des mutants homozygotes pour l'insertion dans ces lignées, et d'autre part, savoir s'il existe à nouveau un nombre surnuméraire de plantes sensibles, en particulier pour les 4 lignées qui avaient montré ce phénomène en génération T2.

Pour se faire, pour les 14 lignées d'intérêt, nous avons semé une centaine de graines T3 provenant de 6 plantes T2 individuelles, sur milieu de culture *in vitro* ( $\frac{1}{2}$  MS) contenant de l'antibiotique kanamycine.

Pour les 10 lignées présentant une ségrégation  $\frac{3}{4}$ ,  $\frac{1}{4}$  de la résistance à l'antibiotique en T2, nous observons en T3 une ségrégation également conforme à celle attendue théoriquement. C'est-à-dire des graines issues d'une plante T2 homozygote pour l'insertion, donnant en T3 100% de plantes résistantes à la kanamycine, et des graines issues d'une plante T2 hétérozygote pour l'insertion, donnant en T3  $\frac{3}{4}$  de plantes résistantes à la kanamycine et  $\frac{1}{4}$  de plantes sensibles.

En revanche, pour les 4 lignées qui présentaient en T2 un excès de plantes sensibles, nous retrouvons cet excès de plantes sensibles à la kanamycine à la génération T3. Pour les lignées A10 et H4, il apparaît qu'aucune plante homozygote pour l'insertion n'a pu être isolée. Pour les lignées C1 et A7, des plantes potentiellement homozygotes pour l'insertion semblent avoir été isolées, car la ségrégation en génération T3 donne quasiment 100% de plantes résistantes à la kanamycine, avec toutefois quelques plantes sensibles (moins de 10%).

*Nota bene* : Pour ces 4 lignées, une analyse de ségrégation en génération T4 a été également menée et a montré qu'à cette génération l'excès de plantes sensibles à la kanamycine était encore plus net. En effet, il reste moins de 10 à 20% de plantes résistantes à cet antibiotique.



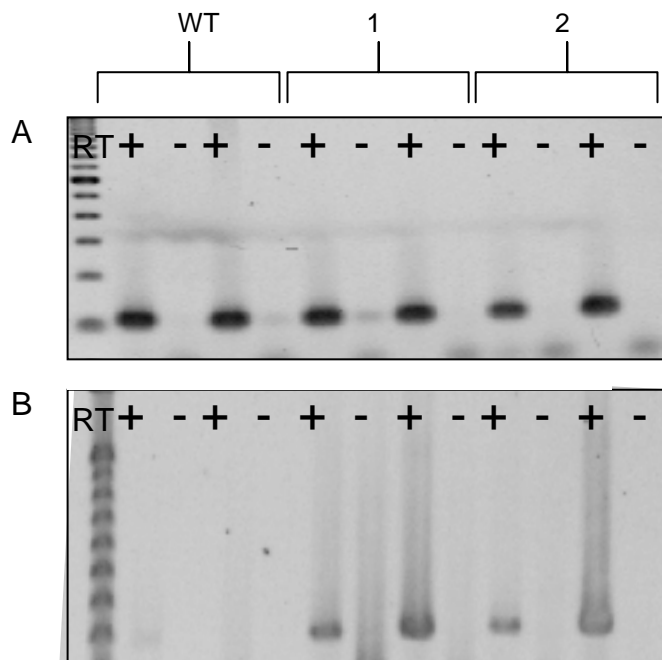


Figure n°74 : Lignées de surexpression du gène *AtDYWI* :

A/ Détection du niveau global de transcrit *AtDYWI* (couple d'amorces interne au gène *AtDYWI*)

B/ Détection du transcrit *AtDYWI* correspondant au transgène (couple d'amorces U3-U5)

WT: ADNc de plantules sauvages ayant poussé sur 1/2MS

1: ADNc de plantules sans phénotype de la lignée H4 ayant poussé sur 1/2MS

2: ADNc de plantules petites présentant une coloration « vert-pâle » de la lignée H4 ayant poussé sur 1/2MS

### \* Approfondissement de l'étude phénotypique

Nous avons dans un premier temps voulu tester la reproductibilité du phénotype observé en T2. Pour chaque lignée, une vingtaine de plantules résistantes à la kanamycine, obtenues lors de l'analyse de ségrégation en T3, est repiquée en terre. Comme précédemment, 4 à 6 semaines après transfert, certaines plantes présentent un phénotype complexe : retard de croissance, anomalies des feuilles et surtout des tiges, perte de la dominance apicale... Ces phénotypes sont d'autant plus marqués dans les plantes provenant des lignées H4 et C1.

Ces plantes proviennent de boîtes de culture *in vitro* contenant un antibiotique et, bien que résistantes à cet antibiotique, cette étape de sélection pourrait constituer un « stress » pour les plantes et être à l'origine du phénotype que nous observons. Afin de tester que ce phénotype apparaît de façon indépendante à cette sélection, nous avons fait germer des graines de ces lignées sur un milieu de culture *in vitro* ne contenant pas d'antibiotique. Pour 8 de ces lignées (en particulier les lignées C1, A7, H4 et A10), nous avons observé l'apparition de deux populations de plantules au sein d'une même boîte de culture : des plantules comparables à des plantules sauvages, et des plantules plus petites avec une coloration vert pâle (figure n°73). Notons que pour les lignées C1, A7, H4 et A10 aucune plantule de type albinos n'est apparue.

Des petites plantules de toutes ces lignées ont été transférées en terre. A nouveau, 4 à 6 semaines après transfert, plusieurs plantes présentent un phénotype complexe, comparable à celui déjà observé.

### \* Etude de l'expression du gène *AtDYWI* dans les deux populations de plantules apparues en culture *in vitro*

Via des expériences de RT-PCR, nous avons évalué l'expression du gène *AtDYWI* dans ces deux populations de plantules apparues sur milieu de culture *in vitro*. Les ARN totaux de petites et grosses plantules ainsi que de plantules sauvages (écotype WS) poussées dans les mêmes conditions ont été extraits et des ADNc ont été synthétisés.

L'utilisation d'un couple d'amorces localisé dans le gène *AtDYWI*, nous a permis d'obtenir une amplification correspondant à la fois à l'expression du gène *AtDYWI* endogène et de la copie transgénique présente dans ces plantules. On observe un produit d'amplification comparable chez toutes les plantules mutantes et sauvages (figure n°74A). L'utilisation du couple d'amorces U3-U5, localisé dans la cassette transgénique, nous a permis de détecter l'expression du transgène d'intérêt spécifiquement. Un produit d'amplification comparable chez toutes les plantes mutantes a été obtenu (figure n°74B).

Sous contrôle du promoteur P35s-CaMW on attend, théoriquement, une très forte expression de la copie transgénique du gène *AtDYWI* dans les lignées mutantes. Or de façon surprenante, bien que son expression soit détectable, elle semble modérée. En effet, le niveau global d'expression du gène *AtDYWI* dans ces lignées (expression de la copie transgénique et de la copie endogène) semble assez comparable au niveau d'expression du gène *AtDYWI* dans les plantes sauvages. Ceci suggère, bien que nous ne l'ayons pas vérifié par des expériences de western blot, que la quantité de protéine *AtDYWI* dans ces lignées doit être comparable à celle d'une plante sauvage. Par ailleurs, à notre seuil de détection, nos résultats indiquent que l'apparition de petites plantules de couleur vert-pâle n'est pas liée à une variation de l'expression du gène *AtDYWI* dans nos lignées.

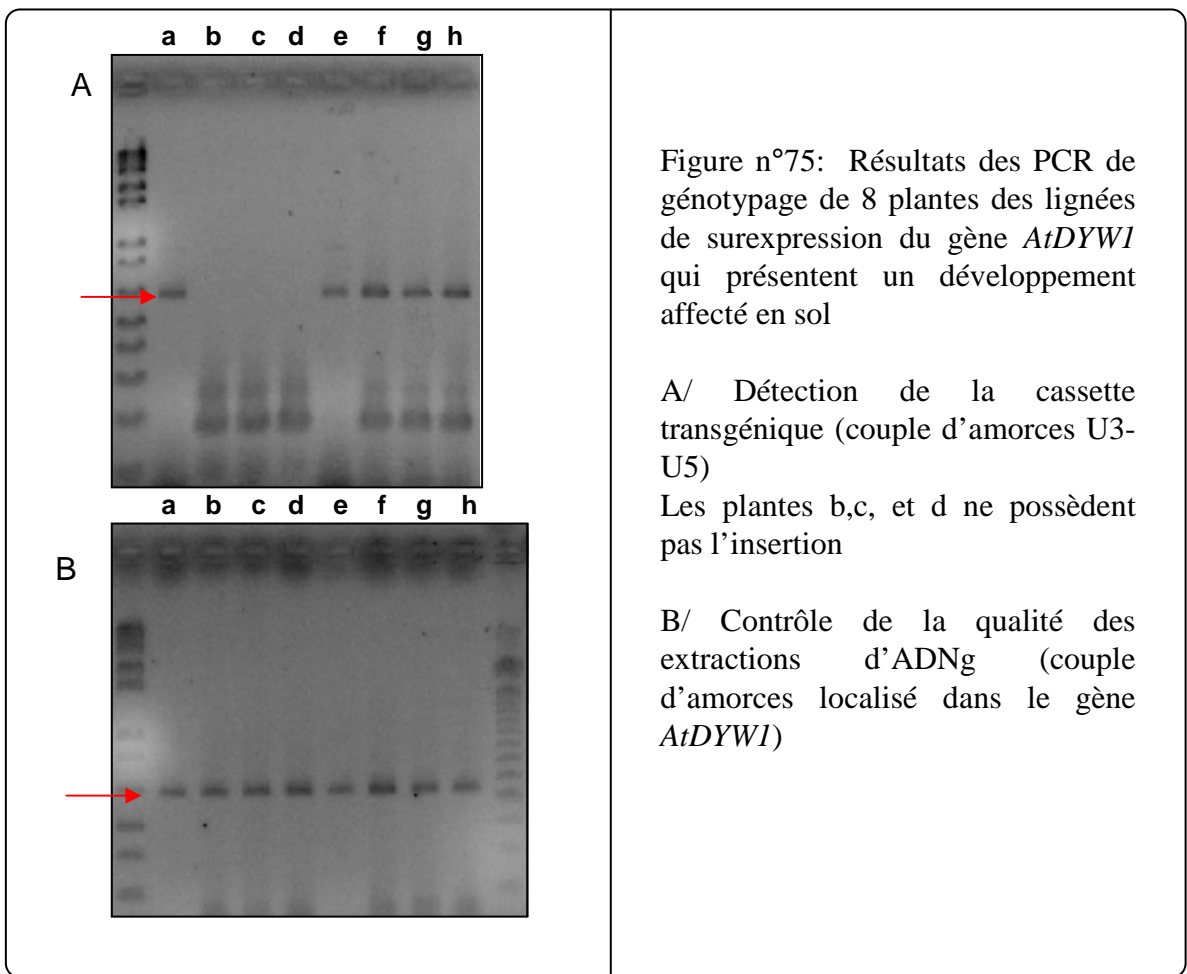


Figure n°75: Résultats des PCR de génotypage de 8 plantes des lignées de surexpression du gène *AtDYWI* qui présentent un développement affecté en sol

A/ Détection de la cassette transgénique (couple d'amorces U3-U5)

Les plantes b,c, et d ne possèdent pas l'insertion

B/ Contrôle de la qualité des extractions d'ADNg (couple d'amorces localisé dans le gène *AtDYWI*)

#### \* Génotypage des plantes présentant un phénotype en terre

Les petites plantules vert-pâle repiquées en terre et donnant des plantes avec un phénotype macroscopique complexe proviennent d'un milieu de culture *in vitro* sans sélection. Ainsi, nous avons voulu vérifier dans cette population que le phénotype observé co-ségrège avec notre construction d'intérêt. Pour se faire, une feuille de chaque plantes affectées a été prélevée et placée dans en plaque 96 puits afin de réaliser une extraction d'ADN génomique et une PCR de génotypage.

L'utilisation du couple d'amorces U3-U5 a permis de vérifier l'éventuelle présence de la cassette transgénique dans ces plantes. La qualité des extractions d'ADNg a été vérifiée en utilisant un couple d'amorces localisé dans le gène *AtDYW1* (permettant l'amplification du gène endogène). De façon tout à fait surprenante, les résultats de PCR indiquent que certaines plantes de phénotype affecté ne possèdent pas l'insertion d'intérêt dans leur génome, alors que d'autres possèdent cette insertion (figure n°75).

#### IV.2.5 Bilan et interprétations des résultats

Ces résultats peuvent être interprétés de deux façons. Tout d'abord, on peut proposer que le phénotype macroscopique complexe observé dans ces lignées ne co-ségrège pas avec la cassette de surexpression d'intérêt. Son apparition serait alors liée à la présence d'une mutation non caractérisée dans le génome de ces lignées qui vraisemblablement proviendrait d'une erreur technique (les plantes T0 transformées n'étaient pas sauvages). Ainsi, l'étude de ces lignées ne peut pas nous apporter une aide précieuse quant à la compréhension du rôle de la protéine AtDYW1 chez *Arabidopsis thaliana*.

Cependant, une hypothèse alternative peut expliquer ces observations. En effet, nous avons constaté, d'une part, que l'expression de la cassette transgénique, bien que sous contrôle du promoteur fort P35S-CaMW, apparaît modérée dans ces lignées. D'autre part, 4 lignées présentent un excès de plantes sensibles à la kanamycine dès la génération T2 et ce phénomène s'amplifie au cours des générations. Pris ensemble, ces résultats indiquent l'existence d'un processus de « co-suppression » ou tout au moins de répression de la cassette transgénique dans ces lignées. Un tel processus est souvent lié à une hyperméthylation de la cassette transgénique entraînant un blocage de la transcription de la région correspondante (phénomène de TGS pour transcriptional gene silencing). Il est raisonnable d'envisager que l'hyperméthylation de cette région d'ADNg diminue la qualité de l'ADNg extrait et/ou l'efficacité des PCR de génotypage. Ceci pourrait expliquer l'absence de détection de la construction d'intérêt dans les plantes présentant un phénotype macroscopique complexe en terre.

Une analyse du niveau de méthylation de la cassette transgénique dans ces lignées nous aurait permis de choisir entre ces deux hypothèses, mais par manque de temps, ces expérimentations n'ont pas été entreprises au cours de mon doctorat.

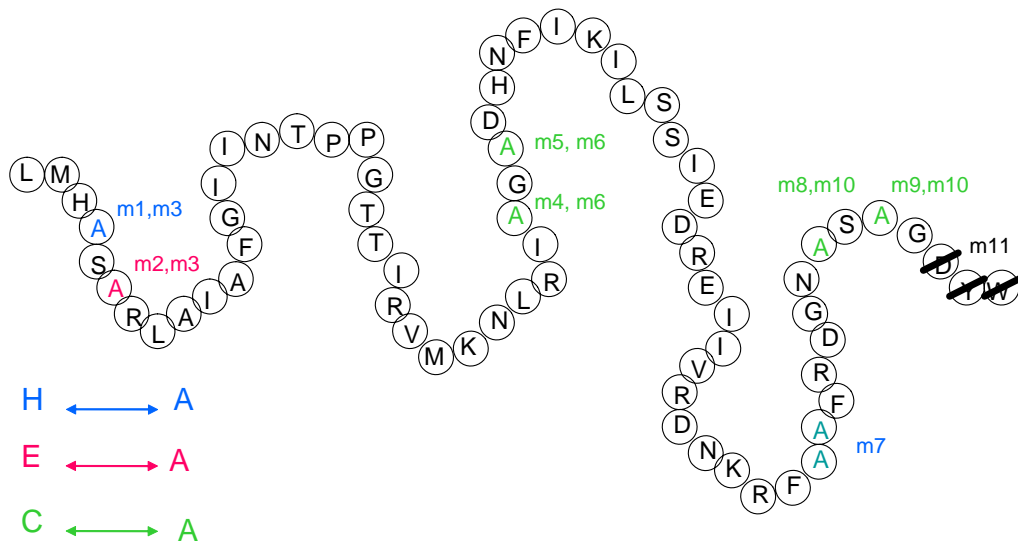


Figure n°76 : Représentation schématique des mutations ponctuelles envisagées dans le site supposé catalytique de la protéine AtDYW1 afin d'établir des lignées dites « dominant-négatif »

Les acides aminés modifiés sont spécifiés par un code couleur.  
 Le bleu détermine un changement d'une histidine en alanine  
 Le rose détermine un changement d'un acide glutamique en alanine  
 Le vert détermine un d'une cystéine en alanine

Les mutants m3, m6, m7 et m10 ont deux acides aminés modifiés.  
 Le mutant m11 correspond à une délétion du triplet final DYW.  
 Les autres mutants ont un acide aminé modifié.

### **IV.3 Mutant de surexpression de deuxième génération**

Parallèlement à l'approche que nous venons de décrire, j'ai également cherché à l'établir un second type de lignées de surexpression du gène *AtDYWI* sous sa forme « sauvage » ou présentant des mutations ponctuelles de résidus conservés du motif DYW.

#### *IV.3.1 Etablissement des lignées de surexpression de deuxième génération*

##### Choix des mutations ponctuelles et mutagenèse dirigée

Dans la littérature, plusieurs études montrent que le changement d'un acide aminé conservé dans le site catalytique d'une désaminase peut totalement abolir sa fonction; par exemple, le remplacement d'une cystéine en alanine de la CDA humaine APOBEC3G (Opi et al., 2006), ou celui d'une histidine en alanine d'une adénosine désaminase de souris (Sideraki et al., 1996), ou encore celui d'un acide glutamique en alanine d'une cytidine-désaminase d'*E. coli* (Carlow et al., 1996). Ainsi, une mutagenèse dirigée dans la partie C-terminale de la protéine *AtDYWI* (site catalytique supposé) a été effectuée afin d'affecter un ou plusieurs résidus cystéine, histidine, ou acide glutamique conservés chez toutes les protéines ayant un domaine DYW. Ces mutations sont notées m1 à m10 (figure n°76). De plus, le triplet final DYW est quasiment invariant dans l'alignement de milliers de séquence du domaine DYW de nombreuses espèces, ce qui peut supposer un rôle fondamental pour ces 3 derniers acides aminés. Une onzième mutation a consisté en la délétion du triplet final DYW de la protéine *AtDYWI*.

La mutagenèse dirigée permettant l'obtention des clones d'entrée du système gateway (pDNR221) porteur de l'ORF modifiée *AtDYWI* s'est faite à l'aide de deux PCR successives et d'une réaction de BP clonase (voir matériel et méthode, paragraphe II.1.6.3, figure n°100). Suite à cette étape, la séquence des clones d'entrée a été vérifiée par séquençage à l'aide des amorces seqLB et M13F. Toutes les formes mutantes de l'ORF *AtDYWI* pleine longueur (ATG2/STOP) ont été ainsi obtenues à l'exception des mutations m7 et m10.

##### Choix du promoteur et vecteur d'expression

Pour cette approche, j'ai choisi d'utiliser le promoteur *AmasPmas* qui permet une expression modérée d'un transgène dans la plante (Etienne Delannoy, communication personnelle). Ce promoteur a été cloné dans un vecteur donneur (pDONR207) du système Gateway, et une double réaction de recombinaison de LR clonase a permis son clonage directionnel en phase avec nos ORF d'intérêts dans le vecteur d'expression pKm24GW. Dans cette approche, nous avons utilisé soit la forme pleine longueur de l'ORF *AtDYWI* (ATG2/stop), soit une forme pleine longueur de cette ORF porteuse des mutations ponctuelles décrites ci-avant.

L'utilisation de cette réaction de double LR a révélé des difficultés techniques si bien que nous avons obtenu, dans un premier temps, uniquement les constructions porteuses de la forme sauvage du gène *AtDYWI*, ou des formes mutantes m2, m6 et m8. Ces constructions ont été introduites dans une souche d'*A. tumefaciens* afin de transformer de plantes d'*Arabidopsis* (écotype Col0).

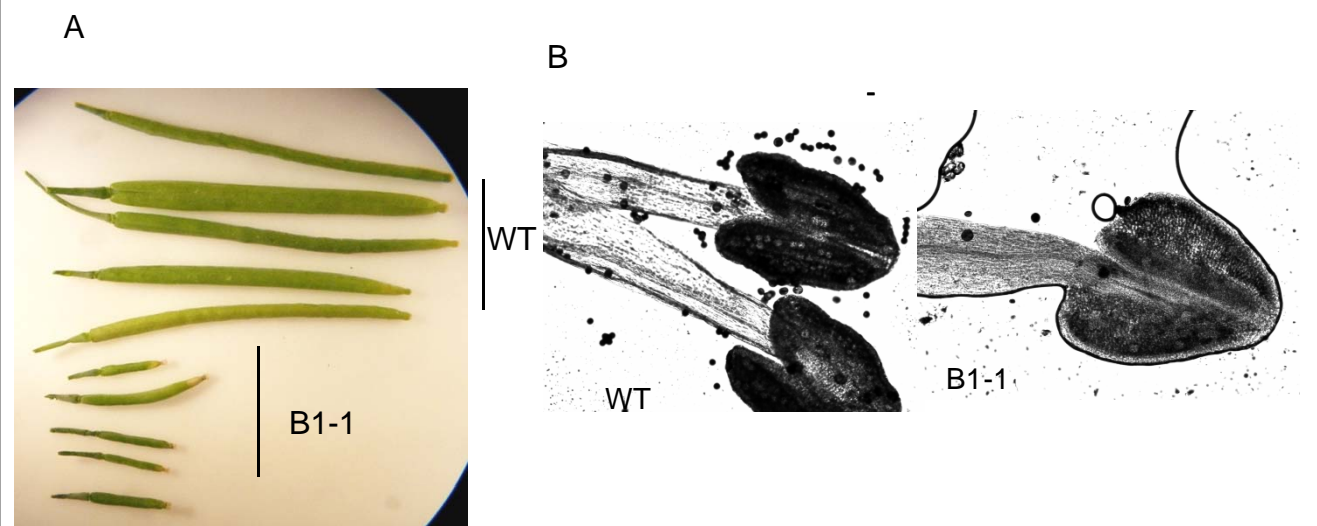


Figure n°77 : Le transformant T1 nommée B1-1 (portant la mutation m2: H171-> A) semble présenter une fertilité réduite

A/ Les siliques sont de taille inférieure à celle d'une plante sauvage, et comportent moins de graines

B/ Les étamines possèdent peu de pollen. Nous avons, selon la méthode décrite par Alexander (1969), mis en évidence que les grains de pollen de la plante B1-1 et d'une plante sauvage se colorent en bleu intense (coloration détectée sous microscope à lumière blanche), suggérant qu'ils sont non avortés.

#### IV.3.2 Sélection et génotypage des transformants obtenus

Les graines ont été semées sur un milieu de culture *in vitro* ( $\frac{1}{2}$  MS) contenant de l'antibiotique kanamycine. Pour chaque lignée, 2 à 4 transformants (T1) indépendants ont été ainsi sélectionnés. Les plantules résistantes à la kanamycine ont été repiquées en terre et les graines de ces plantes ont été récoltées séparément de façon à obtenir des lignées (T2) indépendantes.

Parallèlement, une feuille de chaque plante a été prélevée pour réaliser une extraction d'ADN génomique et une PCR de génotypage utilisant une amorce dans le promoteur AmasPmas et une amorce dans la région terminale de l'ORF *AtDYWI*. Une amplification de taille attendue a été obtenue pour toutes ces plantes. Les produits PCR obtenus ont été séquencés à l'aide d'une amorce localisée dans le gène *AtDYWI* (49int) en amont des éventuelles mutations d'intérêts afin de vérifier la présence de celles-ci.

#### IV.3.3 Caractérisation phénotypique des transformants obtenus

Le phénotype des plantules résistantes à la kanamycine et repiquées en terre a été examiné. Toutes les plantes ont un aspect morphologique et un développement comparable à une plante sauvage, à l'exception de la plante B1-1 (porteuse de la mutation m2 dans la copie transgénique du gène *AtDYWI*).

En effet, les siliques de cette dernière sont de taille extrêmement réduites comparées à celles d'une plante sauvage, et ne contiennent qu'un petit nombre de graines (figure n°77A). De même, l'examen des étamines de cette plante a relevé une morphologie correcte mais très peu de grains de pollen. Par la méthode de coloration histochimique décrite par Alexander (1969), la fertilité des grains de pollen de cette plante B1-1 a été examinée (figure n°77B), révélant que les grains de pollen ne sont pas avortés. Cependant, la plante B1-1 portant la construction d'intérêt à l'état hétérozygote, ces grains de pollens pourraient être tous sauvages.

Afin de s'assurer que nous aurons des graines viables pour cette lignée des croisements [ $\text{♀}$ B1-1 X  $\text{♂}$ sauvage] et croisement réciproque [ $\text{♂}$ B1-1 X  $\text{♀}$ sauvage] ont été réalisés. Ces croisements, quel que soit le sens considéré, ont permis l'obtention de graines donnant naissance à des plantes résistantes à la kanamycine. Ceci confirme la fertilité non seulement des grains de pollen de la plante B1-1, mais aussi des pistils. Par ailleurs, les graines issues de l'autofécondation de la plante B1-1 ont été récoltées et ont un aspect normal. Ces graines ont germé sur un milieu de culture *in vitro* contenant de la kanamycine, et les plantes de deuxième génération (T2) résistantes ne présentent pas de phénotype particulier après transfert en terre. Ainsi le phénotype observé sur la plante B1-1 n'est pas transmissible à la génération suivante.

#### IV.3.4 Etude de l'expression du gène *AtDYWI* par des expériences de PCR quantitative sur les plantes de génération T1

Afin d'étudier l'expression du gène *AtDYWI* dans ces lignées, une feuille de chaque plante T1 a été prélevée pour réaliser une extraction d'ARN totaux et une transcription « inverse » afin d'obtenir des ADNc. Des analyses de PCR quantitative ont été menées à l'aide de 4 couples d'amorces :

- un couple (endoF/R), spécifique de la région 3'UTR de la copie endogène du gène *AtDYWI* et permettant une mesure de l'abondance du transcrite correspondant ;



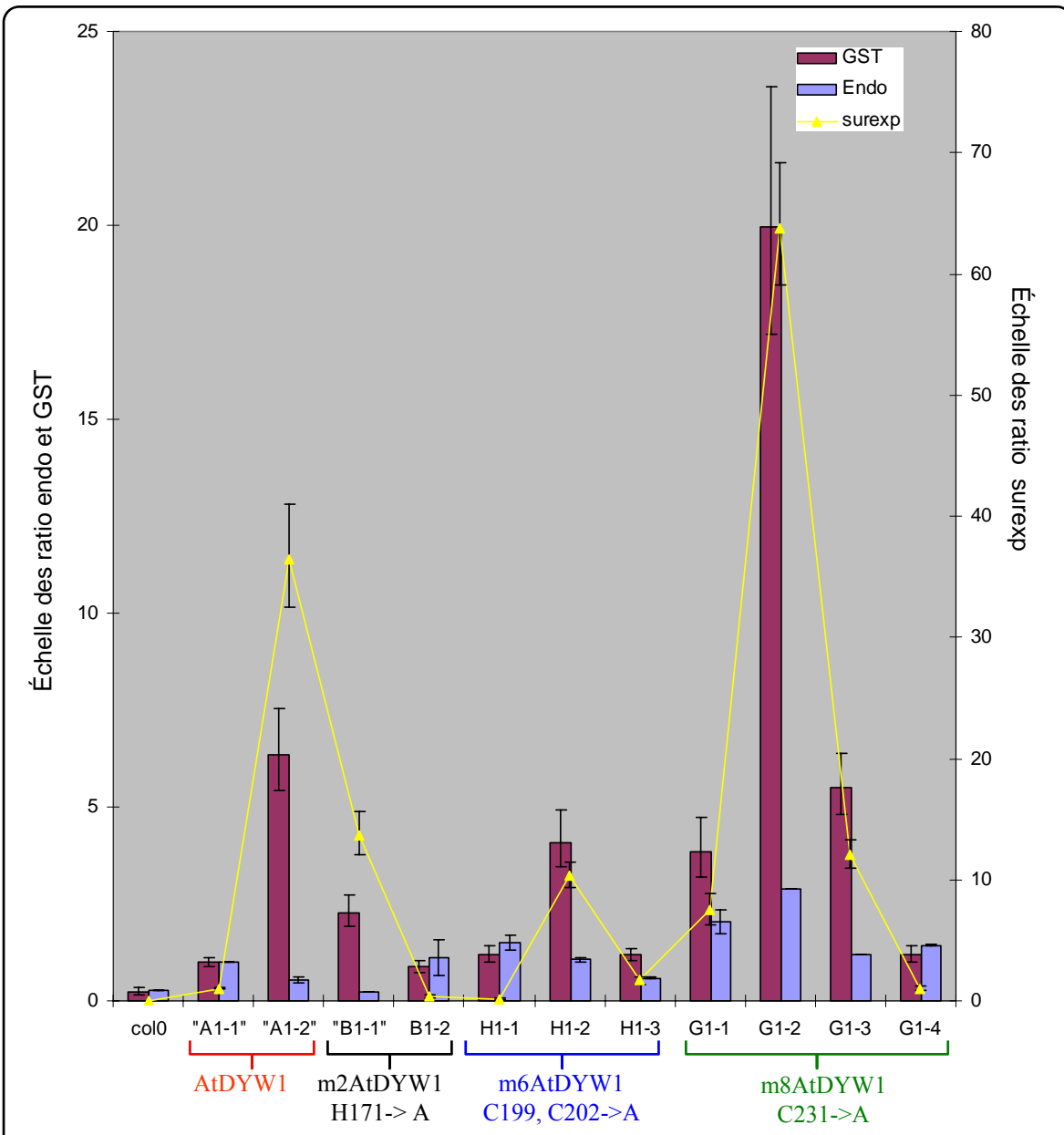


Figure n°78 : Lignées de surexpression ou dite « dominant-négatif » du gène *AtDYW1* : résultats de PCR quantitative

Les données brutes ont été normalisées en utilisant l'expression des gènes constitutifs de l'actine et les ratios sont calculés en prenant la plante sauvage A1-1 comme référence

- un couple (surexpF/R), spécifique de la région 3'UTR du transgène *AtDYW1* et permettant une mesure de l'abondance du transcrite correspondant ;
- un couple (GSTR/F), localisé dans la région codante du gène *AtDYW1*, permettant une mesure de l'abondance des transcrits correspondant à l'expression du gène endogène et du transgène ;
- un couple (ActineF/R), spécifique de deux gènes constitutifs de l'actine (*At3g18780* et *At1g49240*), permettant une normalisation des données.

L'abondance de ces différents transcrits a été mesurée pour chaque plante T1 de ces 4 lignées. Une normalisation est appliquée en utilisant les données d'expression des gènes *actine*, et une déviation standard a été calculée à partir de 3 répétitions techniques. Les résultats indiquent que toutes les plantes T1 expriment le transgène d'intérêt mais avec une extrême variabilité. Si bien que d'une plante à l'autre, le niveau global d'expression du gène *AtDYW1* peut varier d'un niveau comparable à celui d'une plante sauvage (comme dans les plantes B1-2 ou H1-1) à un niveau très augmenté (comme dans les plantes A1-2 ou G1-2)(figure n°78).

#### IV.3.5 Recherche de défauts d'édition sur les plantes de génération T1 : Analyse par PPE

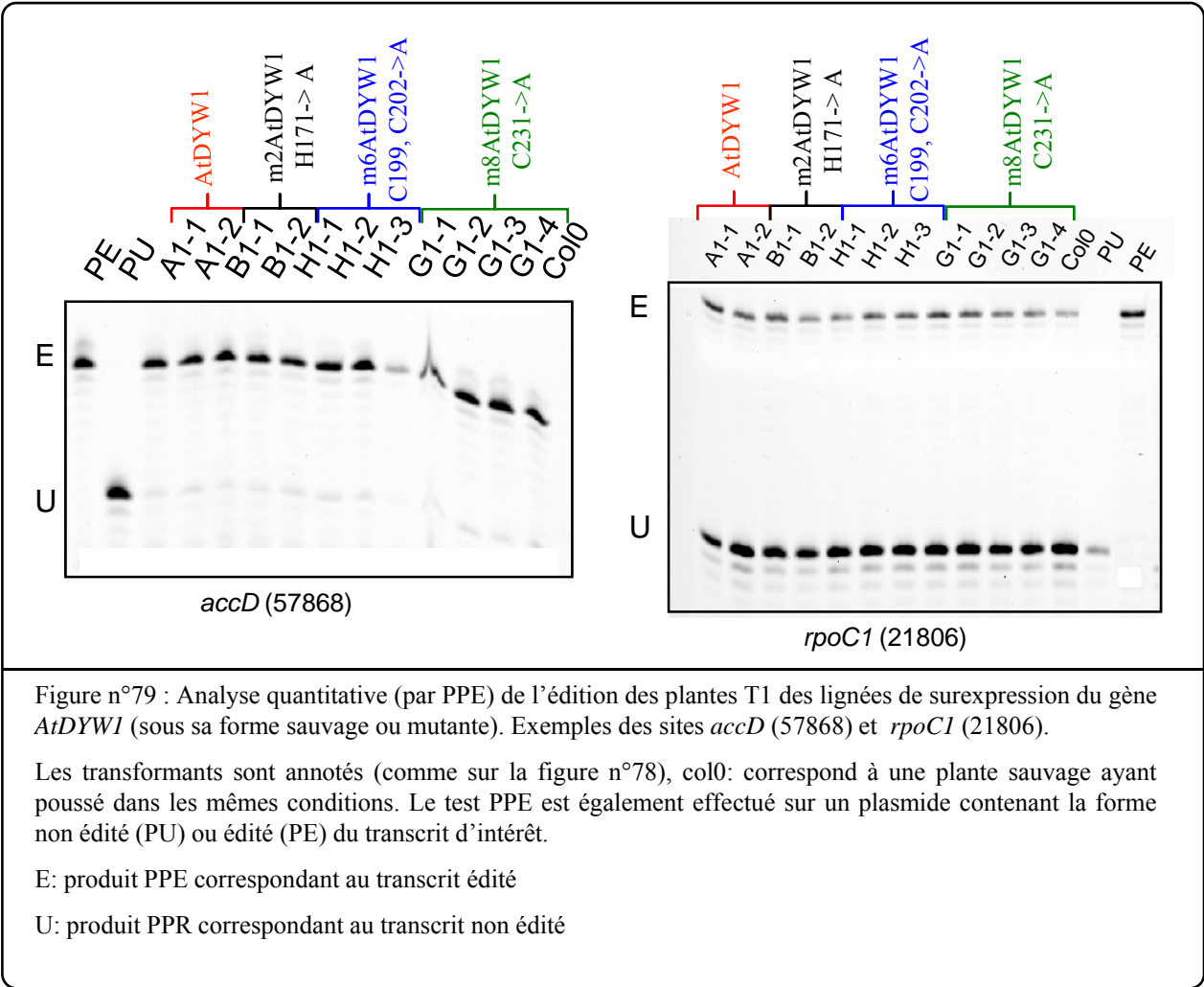
J'ai profité des expériences de mise au point des analyses PPE (décrites au paragraphe III.3.7), pour tester le niveau d'édition de 4 sites chloroplastiques *accD* (gp : 57868), *rpoA* (gp : 78691), *clpP* (gp : 69942) et *rpoC1* (gp : 21806) des plantes T1 de ces 4 lignées. Pour toutes les plantes examinées, aucun défaut d'édition n'a été mis en évidence à ces 4 sites (figure n°79).

#### IV.3.6 Etude de l'expression du gène *AtDYW1* : expériences de western blot sur des plantes de génération T2

Afin, de confirmer la surexpression du gène *AtDYW1* à un niveau protéique dans nos 4 lignées d'intérêt, j'ai entrepris des expériences de western blot.

Pour se faire, des graines de plantes T2 de ces lignées ont été semées sur un milieu de culture *in vitro* (½ MS) contenant de l'antibiotique kanamycine, et 3 plantules résistantes pour chaque lignée ont été repiquées en terre. Le développement et la morphologie de ces plantes (y compris, comme nous l'avons évoqué précédemment, pour les graines issue de la plante T1, B1-1) sont comparables à ceux de plantes sauvages, confirmant l'absence de phénotype macroscopique dans ces lignées d'intérêt.

Ensuite, j'ai réalisé une extraction de protéines totales à partir d'une feuille de chaque plante de nos lignées d'intérêt ainsi qu'une feuille de plantes sauvages Col-0 poussées dans les mêmes conditions. La quantité de protéines de ces extractions a été évaluée sur un mini-gel dénaturant de protéines coloré en bleu de Coomassie (figure n°80A). Un western blot a été réalisé sur ces échantillons en utilisant l'anticorps (AC3) dirigé contre la protéine *AtDYW1*. Une faible bande de taille attendue (~25kDa) est obtenue pour toutes les plantes étudiées. Cette bande est de même intensité que celles obtenues pour les plantes sauvages suggérant qu'il n'y a pas de surexpression de la protéine *AtDYW1* dans nos lignées d'intérêt (figure n°80B).



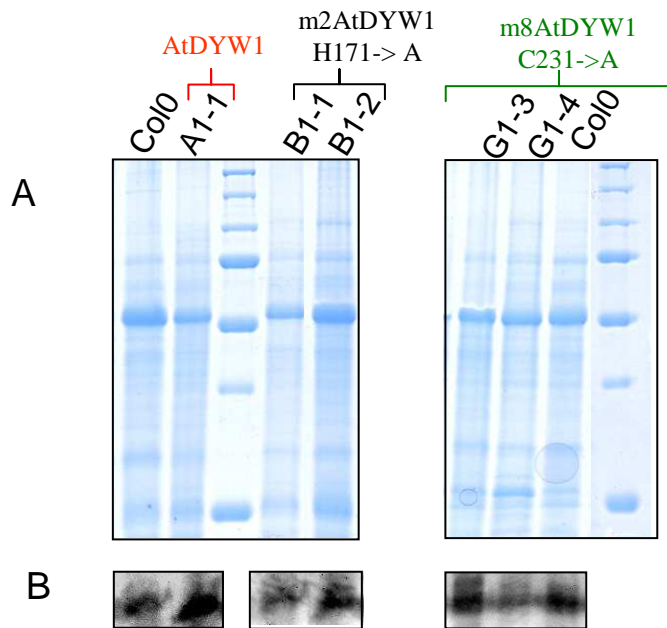


Figure n°80 : Détection de la quantité de protéines AtDYW1 dans des plantes T2 des lignées de surexpression du gène *AtDYW1* (sous sa forme sauvage ou mutante)

- Gel coloré en bleu de Coomassie permettant de vérifier la quantité de protéines déposées
- Membrane du western blot réalisé en utilisant l'anticorps (AC3) dirigé contre la protéine AtDYW1

Résultats présentés pour certains transformants (annotés comme précédemment), col0: correspond à une plante sauvage ayant poussé dans les mêmes conditions.

#### IV.3.7 Bilan et interprétations des résultats

Ainsi, par cette étude, j'ai obtenu des lignées surexprimant « théoriquement » le gène *AtDYW1* sous sa forme sauvage ou portant des mutations ponctuelles dans le site supposé catalytique. Ces lignées ne présentent pas de phénotype macroscopique particulier, et les analyses PPE menées indiquent que dans ces lignées, il n'y a pas de défaut d'édition aux sites *accD* (57868), *rpoA* (78691), *clpP* (gp : 69942) et *rpoC1* (21806).

Par des analyses de PCR quantitative, j'ai mesuré l'expression du transgène dans les différentes plantes (T1) et mis en évidence que son expression varie dans chacune des lignées. Puis, par des analyses de western blot, j'ai montré que, dans ces lignées, la quantité de protéine *AtDYW1* présente était faible et comparable (à notre seuil de détection) à la quantité détectée dans une plante sauvage. Ceci est vrai pour toutes les lignées étudiées, y compris celles pour lesquelles le transcrit correspondant au transgène est très abondant. Ainsi nos résultats suggèrent l'existence d'une régulation post-transcriptionnelle, traductionnelle ou post-traductionnelle importante de l'expression du gène *AtDYW1* dans ces lignées, de sorte qu'une surexpression de la protéine correspondante (porteuse ou non de mutations) soit réprimée. Ainsi, l'étude de telles lignées ne peut pas nous aider à comprendre le rôle de la protéine *AtDYW1* chez *Arabidopsis thaliana*.

Si l'on dresse un bilan, il apparaît que, quelle que soit la stratégie utilisée, l'obtention de lignées surexprimant la protéine *AtDYW1* s'avère délicate. Les résultats obtenus suggèrent l'existence de mécanismes de régulation extrêmement fins de l'expression du gène *AtDYW1* chez *Arabidopsis*. A la fin de ce manuscrit, je propose une réflexion quant à la raison fonctionnelle de cette régulation dans la biologie de la plante. De même, les perspectives envisagées pour poursuivre une caractérisation génétique de ce gène candidat seront évoquées (discussion générale, paragraphes IV et V.4).

Vecteur	Promoteur	Sélection	Tag	mode de clonage	Site de coupure par une protéase	Origine
pDEST 17	T7/lac	AmpR	N-6his	Gateway	non	Invitrogen
pETG30K	T7/lac	KanR	N-6his; N-GST	Gateway	non	EMBL
pETG30A	T7/lac	AmpR	N-6his; N-GST	Gateway	non	EMBL
pETM-11	T7/lac	KanR	N-6his	multisite	TEV	EMBL
pETM-30	T7/lac	KanR	N-6his; N-GST	multisite	TEV	EMBL
pET-15	T7/lac	AmpR	N-6his	multisite	TEV	Pr C. Bond

Table n°7 : Description des vecteurs d'expression utilisés pour produire la protéine rDYW1 en système bactérien

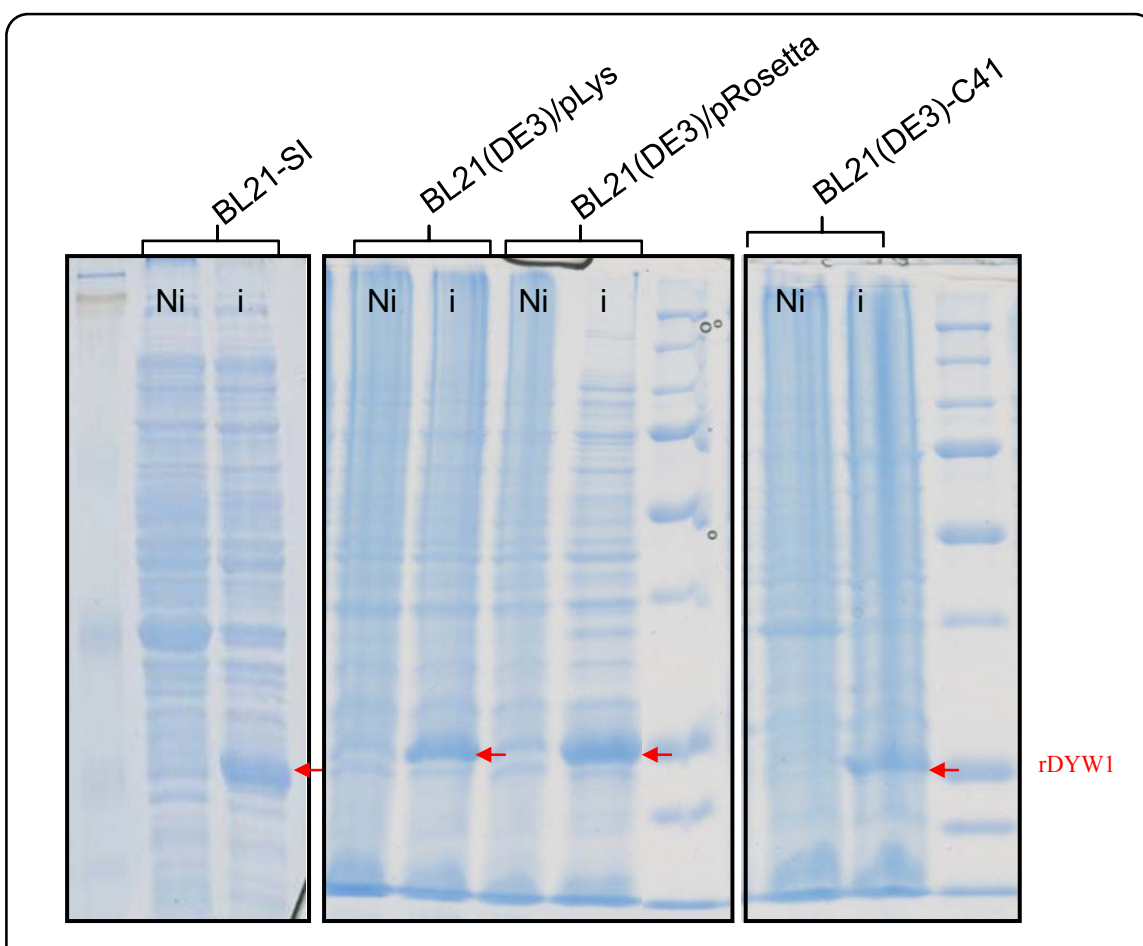


Figure n°81 : Vérification de l'efficacité d'induction sur gel SDS-Page coloré en bleu de Coomassie

Conditions inductions: 37°C, 3h, NaCl (0.3M) pour la souche BL21-SI, 37°C, 3h, IPTG (1mM) pour les autres souches

La flèche rouge indique l'apparition de la protéine rDYW1 recombinante après induction (~25KDa)

## **Partie 4:**

### **UNE ETUDE « FONCTIONNELLE » DE LA PROTEINE AtDYW1**

#### **I Les objectifs de l'approche**

Selon notre modèle de travail, la protéine AtDYW1 pourrait être le facteur enzymatique central catalysant les désaminations C-en-U des réactions d'édition dans les chloroplastes et les mitochondries d'*Arabidopsis*. Nous avons choisi de développer, parallèlement à l'étude génétique que nous venons de décrire, une approche biochimique afin d'obtenir des informations complémentaires d'ordre fonctionnel (test d'activité) et éventuellement structural (cristallographie) quant à cette protéine.

#### **II Production de la protéine AtDYW1 en système bactérien**

##### **II.1 Construction utilisée**

Dans un premier temps, la forme dite « interne/stop » de l'ORF *AtDYW1*, c'est-à-dire sans peptide signal (délétion arbitraire des 126 premières paires de bases) a été clonée dans le vecteur d'entrée du système Gateway® pDONR207. Une réaction LR clonase a permis le clonage directionnel de ce fragment d'ORF *AtDYW1* dans le vecteur de destination pDEST17 (Invitrogen). Ce vecteur permet l'expression, sous contrôle du promoteur de l'ARN T7 polymérase, du fragment protéique d'intérêt fusionné dans sa partie N-terminale à une étiquette composée de 6 résidus histidine (Tag 6-his). La protéine recombinante ainsi produite est notée rDYW1 afin de la distinguer de la protéine sauvage AtDYW1.

Par la suite, au laboratoire, nous avons également tenté le clonage de ce fragment « interne/stop » de l'ORF *AtDYW1*, ou de l'ORF pleine longueur (ATG2/stop) ou encore d'un fragment de cette ORF plus court ne comprenant que le domaine DYW, dans différents vecteurs de destination (liste table n°7). Les étapes d'amplification par PCR, de purification des différents produits PCR, et de clonage dans le vecteur d'entrée du système Gateway® pDONR207 n'ont pas posé de problème technique. En revanche, le clonage dans ces différents vecteurs d'expression (que ce soit en utilisant la technologie Gateway, ou des méthodes plus classiques *via* des enzymes de restrictions) a échoué et ce malgré plusieurs tentatives. Ceci soulève la question d'une éventuelle toxicité de nos constructions pour les bactéries.

Souche d' <i>E.coli</i>	description de la souche	utilisation et particularités	inducteur	Origine	références
B121(DE3)	déficiente pour les protéases <i>lon</i> et <i>ompT</i>	expression inductible de l'ARN polymérase T7, augmente la stabilité des protéines produites	IPTG	Novagen	Studier <i>et al.</i> , 1990
B121-SI	déficiente pour les protéases <i>lon</i> et <i>ompT</i>	expression inductible de l'ARN polymérase T7, augmente la stabilité des protéines produites, augmente la solubilité des protéines produites	NaCl	Invitrogen	Donahue and Bebee, 1999
B121(DE3)/pRosetta	déficiente pour les protéases <i>lon</i> et <i>ompT</i> , contient sur un plasmide les gènes d'ARNt: <i>argU</i> , <i>argW</i> , <i>glyT</i> , <i>lleX</i> , <i>leuW</i> , <i>metF</i> , <i>proL</i> , <i>thrT</i> , <i>thrU</i> , et <i>tyrU</i>	expression inductible de l'ARN polymérase T7, augmente la stabilité des protéines produites, expression de protéines eucaryotes contenant des codons rares	IPTG	Novagen	-
B121(DE3)/pLys	déficiente pour les protéases <i>lon</i> et <i>ompT</i> , contient sur un plasmide les gènes répresseurs <i>lac</i> et T7 lysozyme	expression inductible de l'ARN polymérase T7, augmente la stabilité des protéines produites, contrôle plus strict de l'induction (réduit l'expression basale), facilite la lyse des cellules	IPTG	Novagen	-
B121(DE3)-C41	déficiente pour les protéases <i>lon</i> et <i>ompT</i> , contient deux mutations non caractérisées	expression inductible de l'ARN polymérase T7, augmente la stabilité des protéines produites, facilite l'expression de nombreuses protéines toxiques telles que les protéines membranaires et globulaires	IPTG	Beys Biolog	Miroux and Walker, 1996
B121(DE3)-Turner	déficiente pour les protéases <i>lon</i> et <i>ompT</i> , mutant de la lactose perméase ( <i>lacY</i> )	expression inductible de l'ARN polymérase T7, augmente la stabilité des protéines produites, permet une induction plus homogène de cellules	IPTG	Novagen	-
B121(DE3)-Origami	déficiente pour les protéases <i>lon</i> et <i>ompT</i> , porteuse de mutations dans les gènes de la thioredoxine réductase ( <i>trxB</i> ) et la glutathione reductase ( <i>gor</i> )	expression inductible de l'ARN polymérase T7, augmente la stabilité des protéines produites, réduit la formation de ponts disulfures entre les molécules	IPTG	Novagen	Lobel <i>et al.</i> , 2001
B121(DE3)-STAR	déficiente pour les protéases <i>lon</i> et <i>ompT</i> , porteuse de mutations dans le gène de la RNaseE ( <i>rneI31</i> )	expression inductible de l'ARN polymérase T7, augmente la stabilité des protéines produites, réduit la dégradation des ARNm	IPTG	Invitrogen	Lopez <i>et al.</i> , 1999

Table n°8 : Description des souches d'*E. coli* utilisées pour produire la protéine rDYW1



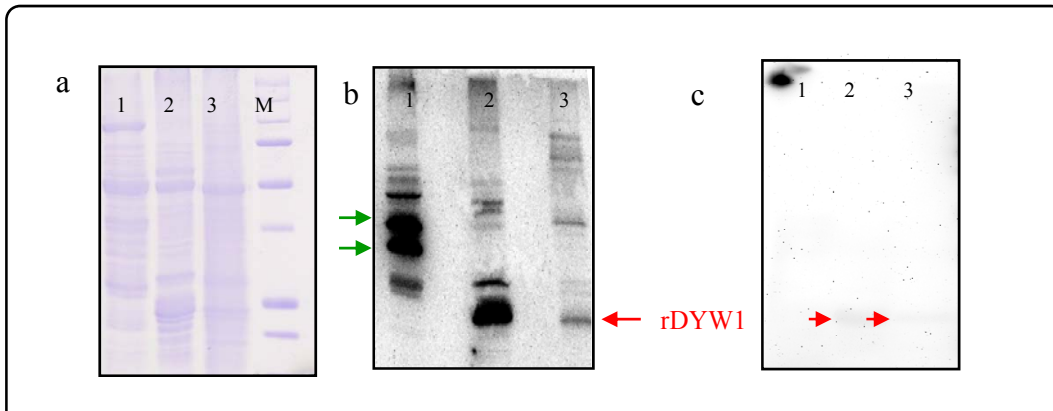
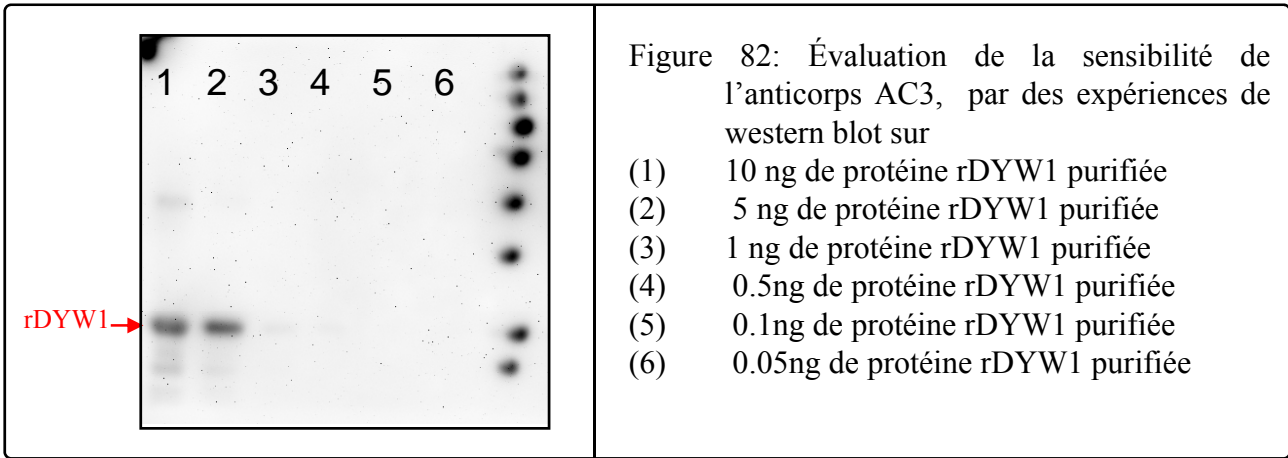


Figure 83 : Évaluation de la spécificité de l'anticorps AC3, par des expériences de western blot sur

- (1) Fraction protéique d'Arabidopsis enrichie en mitochondries
- (2) Fraction protéique d'Arabidopsis enrichie en chloroplastes
- (3) Fraction de protéines totales d'Arabidopsis

a) Gel colorée en bleu de Coomassie permettant de vérifier la quantité de protéines déposées.

b) Membrane de western blot après révélation. L'anticorps AC3 a été dilué 1/500

c) Membrane de western blot après révélation. L'anticorps AC3 a été dilué 1/1000

Les flèches vertes indiquent la présence de bandes aspécifiques correspondant à des protéines enrichies dans la fraction mitochondriale, et dont nous pensons qu'il pourrait s'agir de protéines PPR à domaine DYW .

## **II.2 Expression de la protéine rDYW1 en système bactérien**

Ainsi, afin d'exprimer notre protéine d'intérêt en système bactérien, nous avons utilisé le vecteur d'expression pDEST17 porteur du fragment « interne/stop » de l'ORF *AtDYW1*. Ce vecteur a été utilisé pour transformer différentes souches d'*E. coli* (liste table n°8), toutes dérivées de la souche BL21 et permettant une expression inductible de l'ARN T7 polymérase et conséquemment de notre protéine recombinante d'intérêt.

Comme nous nous interrogeons sur l'éventuelle toxicité de la protéine AtDYW1 exprimée dans *E. coli*, nous avons pris soin de vérifier que les colonies sélectionnées et utilisées dans les tests d'expression portent une construction intègre. Pour cela une minipréparation d'ADN plasmidique et un séquençage a été réalisé, ne révélant aucune mutation dans le fragment « interne/stop » de l'ORF *AtDYW1* d'intérêt.

Toutes les souches ont été induites 3h à 37°C en ajoutant l'inducteur approprié (IPTG ou NaCl) dans le milieu de culture. L'efficacité de l'induction a été vérifiée par dépôt sur mini-gel dénaturant (SDS-page) et coloration en bleu de Coomassie d'une petite fraction de protéines totales provenant de cellules induites et non induites. Nous avons obtenues une quantité correcte de protéine rDYW1 pour 4 souches : BL21-SI, BL21-pLys, BL21-C41 et BL21-Rosetta (figure n°81).

## **II.3 Production d'un anticorps dirigé contre la protéine AtDYW1**

Afin d'obtenir un anticorps polyclonal dirigé contre la protéine AtDYW1, la souche BL21-SI a été induite et les cellules exprimant la protéine rDYW1 ont été lysées en conditions dénaturantes. Une grande quantité de protéine rDYW1 ainsi produite a été purifiée sur colonne Ni-NTA (voir aussi paragraphe II.5.1) avant d'être injectée à des lapins. Ainsi, nous avons pu obtenir 3 anticorps polyclonaux notés (AC1, AC2, et AC3) dirigés contre notre protéine d'intérêt.

Par des expériences de western blot, l'utilisation de ces 3 anticorps a permis la détection de la protéine rDYW1 produite en cellules bactériennes. La sensibilité de ces anticorps a été évaluée en utilisant une gamme de concentration de protéine rDYW1 (produite en système bactérien et purifiée sur colonne Ni-NTA). Les résultats indiquent que nous pouvons détecter une quantité de l'ordre de 1ng, 5ng, et 10 ng de protéine avec l'anticorps AC2, AC3 et AC1, respectivement (figure n°82, exemple pour l'anticorps AC3).

Par ailleurs, toujours par des expériences de western blot, ces anticorps permettent la détection d'un signal de taille attendu (~25kDa) pour la protéine AtDYW1 à partir d'extraits protéiques d'*Arabidopsis*. Ce signal est obtenu avec une relativement faible dilution des anticorps (1/500), et il n'est plus ou pratiquement plus détectable avec une dilution plus grande (1/1000) (figure n°83). Un fort bruit de fond a été obtenu avec les anticorps AC1 et AC2, et de ce fait nous avons préférentiellement utilisé l'anticorps AC3 dans les expérimentations menées au cours de mon doctorat.

Cet anticorps est relativement sensible puisqu'il permet de détecter quelques nanogrammes de protéine rDYW1 purifiée. Cependant, le gène *AtDYW1* étant faiblement exprimé chez *Arabidopsis*, il est difficile, malgré cette sensibilité, de mettre en évidence un signal correspondant à la protéine AtDYW1 à partir d'un extrait de protéines totales.

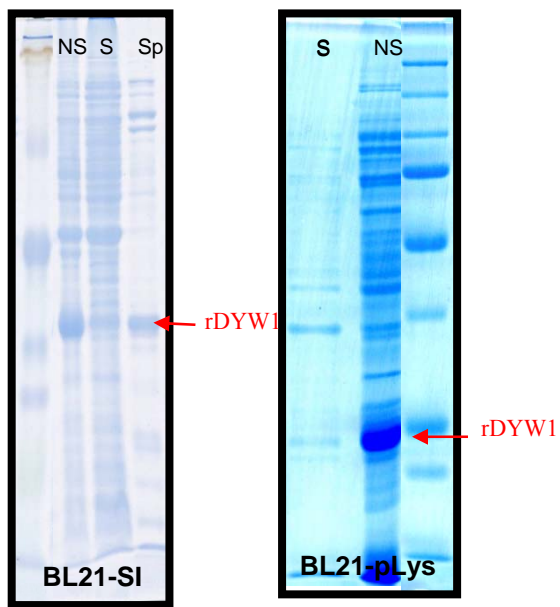


Figure n°84 : Comparaison de la solubilité de la protéine rDYW1 produite dans la souche BL21-SI et BL21-pLys

NS: correspond à la fraction protéique non soluble

S: correspond à la fraction protéique soluble

Sp: correspond à la fraction protéique soluble après purification sur colonne Ni-NTA, s'il y a lieu.

Tampon	composition
1	50mM Tris-HCl pH:8,0 150mM NaCl 10 mM imimidazole pH:8,0
2	50mM Tris-HCl pH:8,0 150mM NaCl 10 mM imimidazole pH:8,0 5mM $\beta$ -mercaptoethanol
3	50mM Tris-HCl pH:8,0 150mM NaCl 10 mM imimidazole pH:8,0 5mM $\beta$ -mercaptoethanol 10% glycérol
4	50mM Tris-HCl pH:8,0 150mM NaCl 10 mM imimidazole pH:8,0 5mM $\beta$ -mercaptoethanol 10% glycérol 1M urea
5	50mM Tris-HCl pH:8,0 10 mM imimidazole pH:8,0
6	50mM Tris-HCl pH:8,0 1M NaCl 10 mM imimidazole pH:8,0
7	50mM Tris-HCl pH:8,0 0.5M ammonium sulphate 10 mM imimidazole pH:8,0
8	50mM Tris-HCl pH:8,0 150mM KBr 10 mM imimidazole pH:8,0
9	100mM K.phos. pH:8,0 10 mM imimidazole pH:8,0 5mM $\beta$ -mercaptoethanol
10	50mM Tris-HCl pH:8,0 150mM NaCl 10 mM imimidazole pH:8,0 0.05% Triton X-100

Table n°9 : Composition des différents tampons de lyse.

*Nota Bene:* Tous ces tampons sont complétés par du lysozyme additionné fraîchement au moment de la lyse des cellules

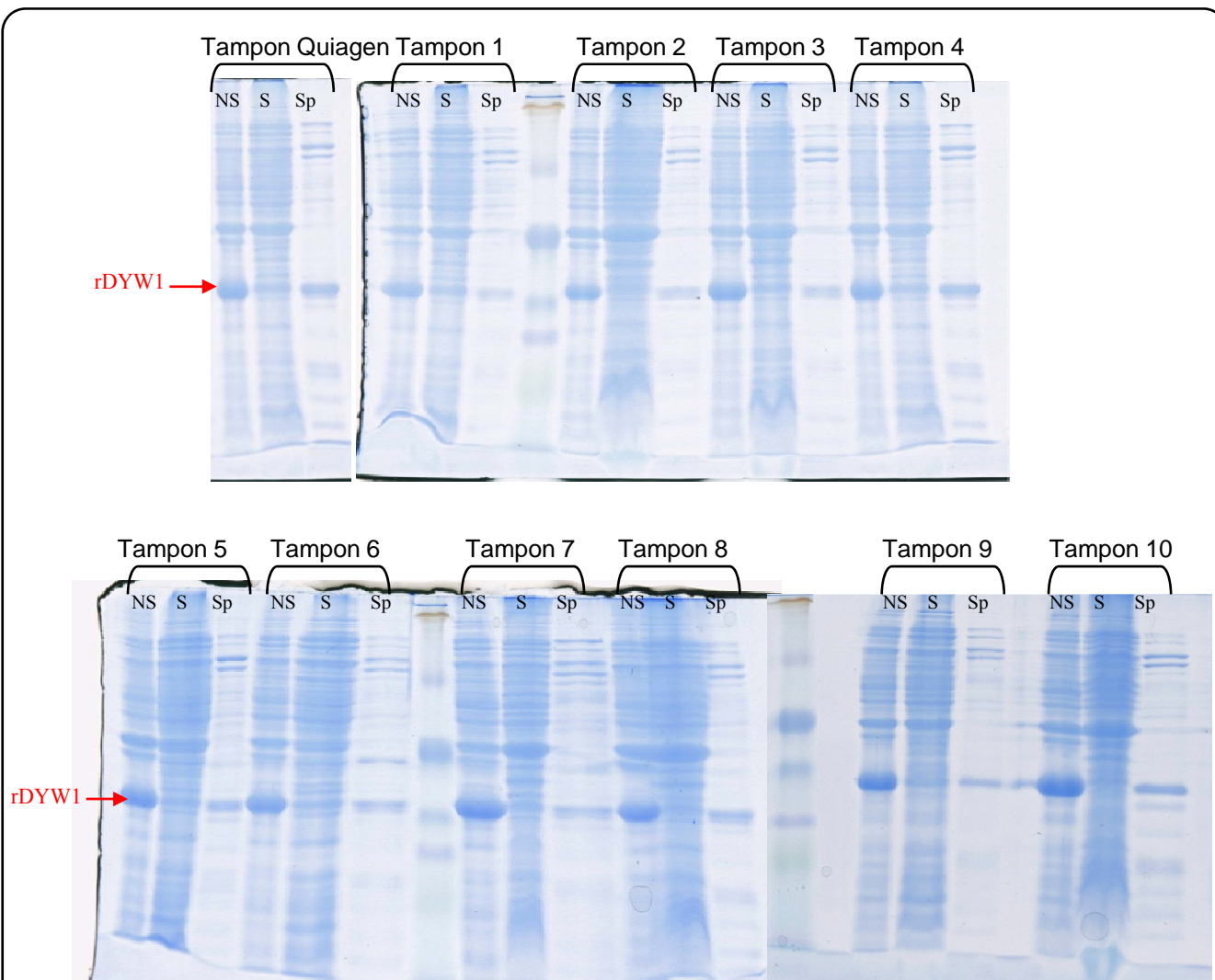


Figure n°85 : Optimisation des conditions de lyse

Différents tampons 1 à 10 (voir composition dans la table n°9) et un tampon commercial (Quiagen) ont été utilisés pour lyser les cellules induites BL21-SI.

NS: correspond à la fraction protéique non soluble

S: correspond à la fraction protéique soluble

Sp: correspond à la fraction protéique soluble après purification sur colonne Ni-NTA.

Par ailleurs, on peut s'interroger quant à la spécificité de cet anticorps. En effet, de façon intéressante, lors de l'utilisation de l'anticorps (AC3) sur une fraction de protéines d'Arabidopsis enrichie en protéines mitochondriales, nous avons observé un fort enrichissement de deux bandes de taille supérieure à celle attendue pour la protéine AtDYW1 (figure n°83). Nous nous sommes interrogés sur l'identité des protéines qui pourraient être reconnues par notre anticorps et avoir une localisation mitochondriale dans les cellules d'Arabidopsis. Sachant que le domaine DYW est relativement bien conservé, pourrait-il s'agir de petites protéines PPR à domaine DYW ? Notons que nous avons déjà évoqué cette éventualité, pour interpréter les résultats obtenus avec l'anticorps (AC2) lors de l'étude de la localisation de la protéine AtDYW1 par des expériences de western blot sur fraction de protéines mitochondriales de pommes de terre (résultats partie II, paragraphe V). Ainsi il est envisageable que ces anticorps reconnaissent spécifiquement la protéine AtDYW1, mais aussi d'autres protéines PPR de type DYW (en particulier dans les mitochondries).

#### **II.4 Optimisation de la production de protéine rDYW1 soluble**

Afin de pouvoir réaliser des études structurales ou fonctionnelles sur la protéine rDYW1, nous avons cherché à évaluer et optimiser la quantité de protéine soluble que nous pouvons produire.

Pour se faire, dans un premier temps, j'ai réalisé une mise au point de l'induction (différentes conditions de température et de temps d'induction) afin de produire de la protéine rDYW1 dans la souche BL21-SI (inductible aux sels). Ainsi, j'ai constaté qu'une induction de 12h à 28°C me donnait un rendement optimum en termes de quantité de protéine soluble rDYW1 produite. En revanche, par manque de temps, pour les trois autres souches inductibles à l'IPTG (BL21-pLys, BL21-C41, BL21-Rosetta) une seule condition d'induction a été testé (3h à 37°C).

Après avoir lysé les cellules en conditions natives (voir aussi paragraphe ci-après), la fraction protéique soluble a été séparé de la fraction insoluble par centrifugation. Nous avons comparé la quantité de protéine rDYW1 soluble obtenue après induction des différentes souches (BL21-SI, BL21-pLys, BL21-C41 et BL21-Rosetta) (figure n°84). Il apparaît très nettement que nous obtenons une bien meilleure quantité de protéine rDYW1 soluble avec la souche BL21-SI comparée aux 3 autres souches. Ceci peut avoir deux explications. D'une part, l'utilisation de la souche BL21-SI est effectivement conseillée pour améliorer la solubilité des protéines produites (Donahue and Bebee, 1999). D'autre part, en raison d'une limitation de temps, j'ai bien mieux optimisé les conditions d'induction et de lyse pour cette souche, que je ne l'ai fait pour les autres souches.

Parallèlement, j'ai cherché à optimiser la lyse en conditions natives des cellules induites. J'ai utilisé soit des tampons de lyse commerciaux (QuIA express Ni-NTA fast start kit lysis buffer commercialisé par Quiagen, ou BugBuster® commercialisé par Novagen) combiné à une sonication des cellules, soit différentes compositions de tampon de lyse préparé au laboratoire (table n°9) toujours associé à une sonication.

Nous obtenons une quantité de protéine rDYW1 soluble assez comparable quelque soit la composition des tampons, avec toutefois peut-être un meilleur rendement pour les tampons contenant de l'imimidazole et du glycérol (figure n°85).

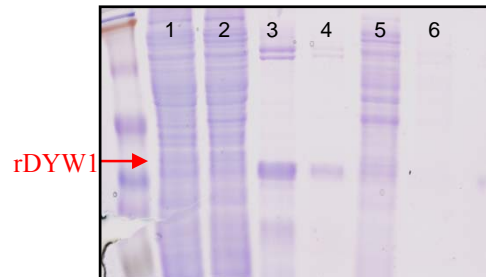


Figure n°86 : Purification de la protéine rDYW1 recombinante sur colonne Ni-NTA produite en système bactérien à partir de la souche BL21-SI

- 1: fraction protéique soluble après lyse en conditions natives
- 2: fraction protéique collectée après passage sur la colonne Ni-NTA
- 5: fraction protéique collectée après un premier lavage de la colonne
- 6: fraction protéique collectée après un premier lavage de la colonne
- 3: fraction protéique collectée après une première élution
- 4: fraction protéique collectée après une seconde élution

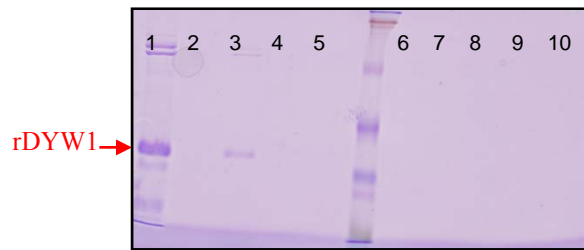


Figure n°87: Purification de la protéine rDYW1 recombinante sur colonne d'échange de cation

- 1: fraction protéique chargée sur la colonne (Protéine rDYW1 après purification sur colonne Ni-NTA, produite en système bactérien, à partir de la souche BL21-SI)
- 2: fraction protéique collectée après passage sur la colonne
- 3: fraction protéique collectée après un premier lavage de la colonne
- 4: fraction protéique collectée après un second lavage de la colonne
- 5: fraction protéique collectée après un lavage final de la colonne avec 3M NaCL
- 6 à 10: différentes fractions protéiques collectées lors de l'élution sur gradient NaCL

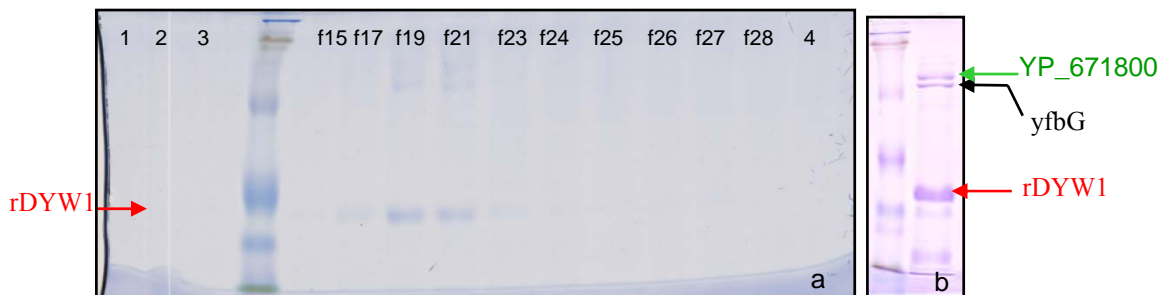


Figure n°88 : Purification de la protéine rDYW1 recombinante sur colonne d'échange d'anion

- a) 1: fraction protéique collectée après passage sur la colonne
- 2 : fraction protéique collectée après un premier lavage de la colonne
- 3: fraction protéique collectée après un second lavage de la colonne
- 4: fraction protéique collectée après un lavage final de la colonne avec 3M NaCL
- f15 à f28: différentes fractions protéiques collectées lors de l'élution sur gradient NaCL
- b) Fractions protéiques F17 à F23 après concentration sur microcon et identification des bandes contaminantes par spectrométrie de masse (indiquée en noir et vert)



## **II.5 Purification de la protéine rDYW1**

### *II.5.1 Chromatographie d'affinité : utilisation du système Ni-NTA (Quiagen)*

Afin de purifier la protéine rDYW1 soluble ainsi produite, nous avons utilisé des colonnes de Nickel-Nitrilotriacetic acid (Ni-NTA) ou des billes magnétiques Ni-NTA. En effet, le tag 6his, présent en N-terminal de notre protéine d'intérêt, a une grande affinité pour le métal et se fixe à ces colonnes avec une bonne affinité. Tirant partie de ce système, j'ai pu obtenir environ 1mg de protéine rDYW1 purifiée à partir de 400mL de culture de cellules BL21-SI (figure n°86, la concentration a été évaluée par dosage au Bradford). Toutefois, même après cette étape de purification, nous pouvons observer la présence des bandes « contaminantes » (notamment d'un doublet de haut poids moléculaire) sur gel.

### *II.5.2 Chromatographie par échange d'ions*

Nous avons ensuite réalisé une deuxième étape de purification sur des colonnes échangeuses d'ions constituées par un support insoluble et chimiquement inerte, chargé d'ions liés électrostatiquement qui vont être remplacés, de manière réversible, par des ions en solution. Ainsi, les colonnes d'échange d'anions sont chargées positivement et vont retenir des protéines chargées négativement et inversement, les colonnes d'échange de cations sont chargées négativement et retiennent les protéines chargées positivement. La charge net d'une protéine varie en fonction du pH et de la concentration saline de la solution dans laquelle elle se trouve (voir matériel et méthodes, figure n°103).

Dans un premier temps, à l'aide du programme de prédiction ProtParam (disponible sur le site ExPasy), j'ai déterminé le point isoélectrique théorique (PI) de la protéine rDYW1 recombinante produite. Le PI correspond à la valeur de pH pour laquelle la protéine ne porte pas de charge, il est remarquablement élevée pour la protéine rDYW1 (PI=9.26).

Ainsi la protéine rDYW1 dans une solution de NaP (50mM), NaCL (100mM) à pH 7.2 sera, théoriquement, immobilisée sur une colonne d'échange de cations. Nous avons donc réalisé une chromatographie par échange de cations dans ces conditions et une élution fractionnée sur un gradient de sels (NaCL, 100mM à 1M). De façon surprenante, nos résultats (figure n°87) indiquent que la protéine rDYW1 s'est fixée avec une très faible affinité sur la colonne échangeuse de cations, puisqu'on la retrouve éluée dans le premier lavage de colonne.

Nous avons alors repris notre protéine rDYW1 purifiée dans une solution de Tris (20mM), NaCL (50mM) à pH 8.5 afin de réaliser une chromatographie par échange d'anions et, comme précédemment, une élution sur un gradient de sels. Cette fois-ci, nous avons élué la protéine rDYW1 dans les fractions 17 à 23 collectées (figure n°88a). Cependant, après cette deuxième étape de purification et concentration de la protéine purifiée obtenue, on remarque que les deux bandes contaminantes (doublet de haut poids moléculaire) sont toujours présentes (figure n°88b).

Lorsque que l'on utilise un tampon de lyse dénaturant avant de purifier la protéine rDYW1 recombinante, on constate que ce doublet de bandes contaminantes disparaît, ce qui suggère qu'il pourrait s'agir de protéines d'*E. coli* interagissant avec la protéine rDYW1. Ainsi, lors de la lyse en conditions natives, ces protéines sont certainement complexées à notre protéine rDYW1 d'intérêt. Ce qui a pour conséquence de modifier la valeur de PI théorique calculée

tampon	composition
1	eau
2	50mM tris (pH7,5)
3	50mM tris (pH7,5) 5mM β-mercaptoéthanol
4	50mM tris (pH7,5) 150mM NaCL
5	50mM tris (pH7,5) 100mM Arginine
6	50mM tris (pH7,5) 1M KBr
7	50mM tris (pH7,5) 3% TMAO
8	50mM tris (pH7,5) 10% glycérol
9	50mM tris (pH7,5) 1mM ZnSO4
10	50mM tris (pH7,5) 100mM Sucrose
11	8M Urée
12	100mM Kpi 5mM β-mercaptoéthanol
13	50mM tris (pH7,5) 5% glycérol 50mM Sucrose 5mM β-mercaptoéthanol
14	identique au tampon n°13 + 10μM ZnSO4
15	identique au tampon n°14, dilué 2X dans de l'eau
16	identique au tampon n°14, dilué 4X dans de l'eau

Table n°10 : Composition des différents tampon utilisés dans nos essais de « renaturation » de la protéine rDYW1.

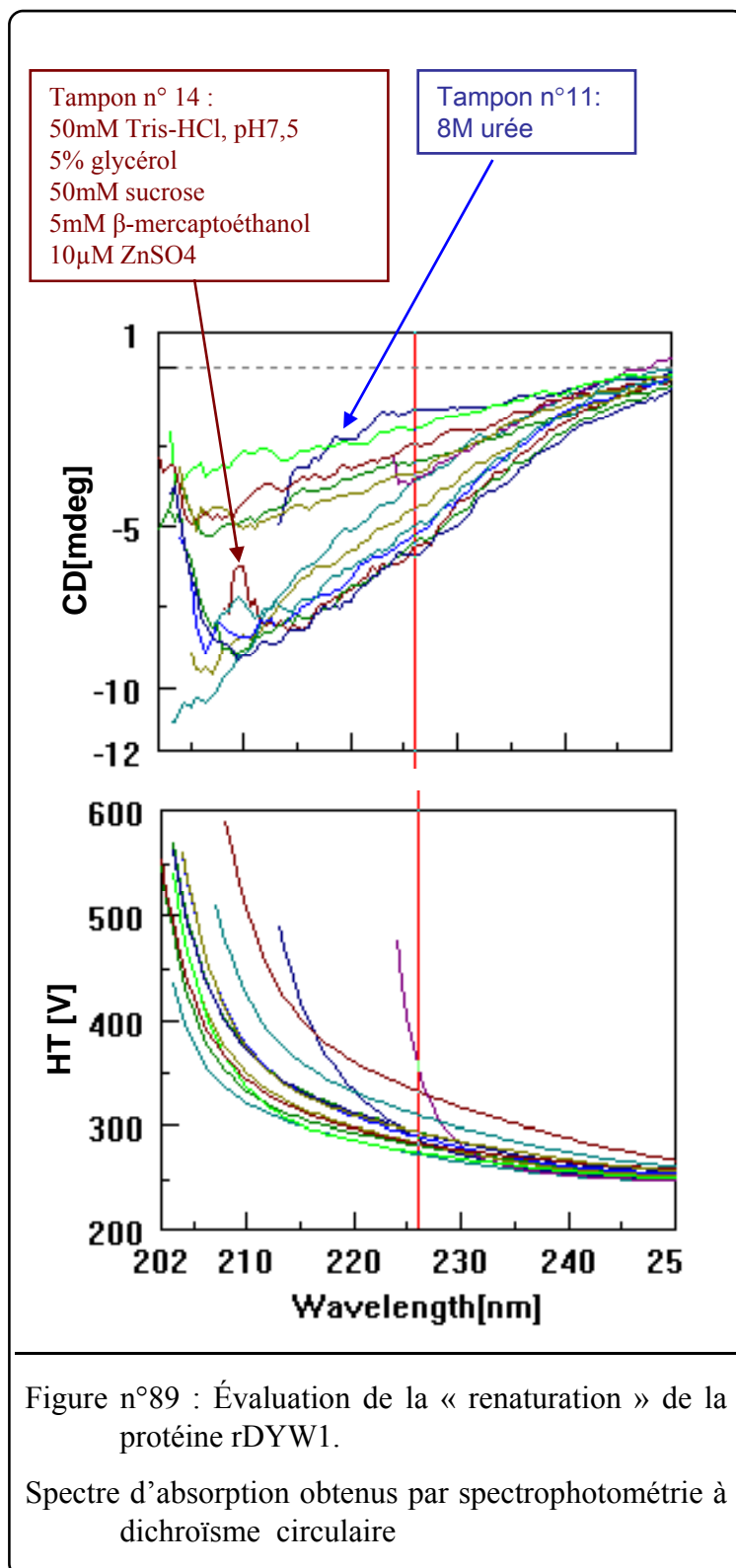


Figure n°89 : Évaluation de la « renaturation » de la protéine rDYW1.

Spectre d'absorption obtenus par spectrophotométrie à dichroïsme circulaire

précédemment et pourrait expliquer le fait que la protéine rDYW1 se soit fixée avec une très faible affinité sur la colonne échangeuse de cations.

### *II.5.3 Identification des « contaminants » par spectrométrie de masse*

Afin d'identifier la nature des protéines qui semblent interagir avec la protéine rDYW1 lors de sa production en système bactérien, nous avons prélevé les bandes correspondantes sur le gel et les peptides extraits ont été analysés par spectrométrie de masse (réalisé en collaboration par Nic Taylor, Plant Energy Biology). Les résultats (figure n°88b) indiquent que la bande de plus haut poids moléculaire correspond à la glucosamine-fructose-6-phosphate aminotransférase (YP\_671800) dont l'homologue le plus proche chez Arabidopsis est une transaminase (At3g24090). Quant à la deuxième bande, elle correspond à une potentielle ribonucléotide transformylase (yfbG) dont l'homologue le plus proche chez Arabidopsis est une UDP-glucuronate décarboxylase (At2g27860). Nous avons également vérifié, par ce biais, que notre bande d'intérêt correspond bien à un fragment peptidique de la protéine AtDYW1.

### **II.6 Tests préliminaires afin d'évaluer, par spectrophotométrie à dichroïsme circulaire, la « renaturation » de la protéine rDYW1**

Nous avons constaté qu'après induction de la souche BL21-pLys et lyse en conditions dénaturantes des cellules exprimant la protéine rDYW1, nous pouvons obtenir une grande quantité de protéine rDYW1 relativement pure. La protéine rDYW1 ainsi produite a été purifiée sur colonne Ni-NTA et resuspendue dans des tampons contenant différents solvants, pH, concentration de sels... (table n°10).

Afin d'étudier la « renaturation » (ou « repliement ») éventuelle de la protéine rDYW1 dans ces différentes solutions, nous avons utilisé un spectrophotomètre à dichroïsme circulaire. En effet, connaissant le spectre d'absorption de différentes structures (telle que les hélices alpha, les feuilletts beta...), on peut estimer mathématiquement la composition de la structure secondaire d'une protéine par l'analyse de son spectre d'absorption de dichroïsme circulaire. Cette méthode ne permet pas de décrire ou de localiser précisément ces différents éléments de structures secondaires, mais permet de détecter relativement facilement des changements de conformation d'une protéine dans différentes solutions et ainsi d'évaluer sa « renaturation » (voir aussi matériel et méthodes, figure n° 104).

De façon préliminaire, nous avons constaté que si la protéine rDYW1 semble totalement « dénaturée » dans une solution concentrée en urée (8M), elle pourrait prendre une certaine conformation (détection de structure en hélice alpha) dans des tampons contenant des solvants organiques, en particulier dans la solution suivante : 50mM TrisHCl pH 7.5 ; 5% glycérol ; 50mM sucrose ; 10µM ZnSO<sub>4</sub>, ~800µM urée (figure n°89).

### **III Production de la protéine rDYW1 en système *in vitro***

Comme nous venons de l'évoquer, au cours de nos essais de production de la protéine AtDYW1 en système bactérien, nous nous sommes interrogés sur son éventuelle toxicité pour les cellules. D'autre part, nous avons été confortés à la présence de protéines bactériennes

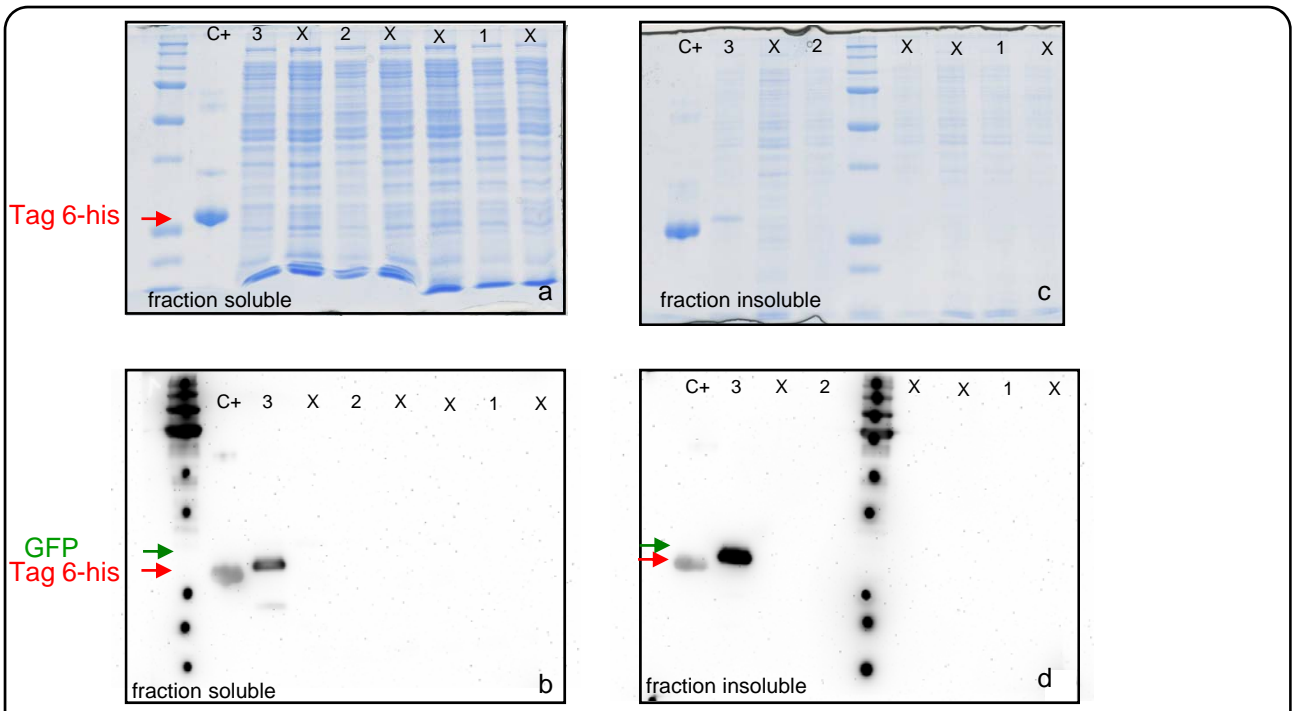


Figure n°90: Production de la protéine rDYW1 recombinante en RTS, par utilisation d'un lysat d'*E. coli*

a et b: Fractions protéiques soluble et insoluble, respectivement, séparées sur gel SDS-Page coloré en bleu de Coomassie

c et d: membrane de western blot réalisé sur les fractions protéiques soluble et insoluble, respectivement, avec l'anticorps dirigé contre le tag 6his

Échantillons:

(1) Protéine d'intérêt produite à partir de 0.2 µg de matrice ADN pDEST17/int-stop*AtDYW1*

(2) Protéine d'intérêt produite à partir d'1 µg de matrice ADN pDEST17/int-stop*AtDYW1*

(3) Protéine GFP produite à partir de 0,5µg de matrice ADN (plasmide contrôle fourni dans le kit), indiqué par une flèche verte

C+: Tag 6-his purifié, indiqué par une flèche rouge

X :échantillons non relatifs à l'expérience décrite

interagissant avec notre protéine recombinante d'intérêt lors de sa purification. Ainsi, pour palier à ces éventuels problèmes techniques, nous avons envisagé l'utilisation d'un système de production *in vitro* pour la protéine rDYW1.

### **III.1 Généralités sur la technologie RTS (Rapid Translation System)**

Le système RTS, aussi surnommé « production acellulaire de protéines », utilise un lysat procaryotique ou eucaryotique « optimisé » afin de réaliser simultanément des réactions de transcription et traduction *in vitro* (Hoffmann et al., 2004). On utilise une matrice d'ADN (soit un plasmide ou bien directement un produit PCR chimérique comprenant le promoteur de l'ARN T7 polymérase) et un lysat qui apporte l'ARN T7 polymérase nécessaire pour la transcription, ainsi que les ribosomes et les composants additionnels nécessaires à la traduction (voir matériel et méthode : figure n°102A). Les substrats énergétiques, les nucléotides et les acides aminés sont ajoutés au milieu réactionnel directement (Batch) ou de façon continue grâce à la technologie CECF (pour continuous-exchange cell-free, voir matériel et méthode : figure n°102B), ce qui augmente significativement le taux de protéine produite.

### **III.2 Production à partir d'un extrait d'*E. coli* (ou *E. coli* system) ou de germe de blé (*Wheat Germ system*)**

Afin de produire la protéine rDYW1, j'ai testé deux « kits » RTS commercialisés par Roche, l'un utilisant un lysat d'*E. coli*, et l'autre utilisant un lysat de germe de blé couplé à la technologie CECF. Ces deux kits sont développés pour un petit volume réactionnel (50 à 100µL) afin d'évaluer dans un premier temps l'efficacité de production (de l'ordre de plusieurs nanogrammes, théoriquement) et la solubilité de la protéine d'intérêt.

J'ai utilisé le vecteur d'expression pDEST17 porteur du fragment « interne/stop » de l'ORF *AtDYW1* comme matrice pour les réactions. L'utilisation du lysat d'*E. coli* n'a pas permis l'obtention de protéine rDYW1 recombinante ou en quantité insuffisante pour permettre sa détection en western blot en utilisant un anticorps dirigé contre le tag 6-his (figure n°90). En revanche, nous avons pu produire cette protéine d'intérêt en utilisant le lysat de germe de blé combiné à la technologie CECF. La quantité de protéine produite est relativement faible (de l'ordre de ~20ng) si bien qu'elle n'est pas détectable sur gel dénaturant coloré en bleu de Coomassie (figure n°91a) mais l'est par western blot en utilisant soit l'anticorps (AC3) dirigé contre la protéine AtDYW1, soit l'anticorps dirigé contre le tag 6-his (figure n°91b et 91c). La protéine rDYW1 ainsi produite est pour moitié dans la fraction protéique soluble et pour moitié dans la fraction protéique insoluble.

J'ai pu obtenir une petite quantité de protéine rDYW1 soluble après purification sur billes magnétiques Ni-NTA, cependant la révélation par western blot ne permet pas d'évaluer la pureté de la protéine ainsi obtenue. Théoriquement, l'utilisation d'un kit RTS adapté pour un grand volume réactionnel permettrait d'obtenir plusieurs milligrammes de protéine d'intérêt et tester sa pureté, sur gel dénaturant coloré en bleu de Coomassie, après purification sur colonne Ni-NTA, malheureusement, par manque de temps, je n'ai pu réaliser un tel essai.

Enfin, notons qu'il est probable qu'un système RTS basé sur l'utilisation d'un lysat d'origine végétal est plus approprié pour produire la protéine rDYW1. Cependant, nous avons vu qu'il

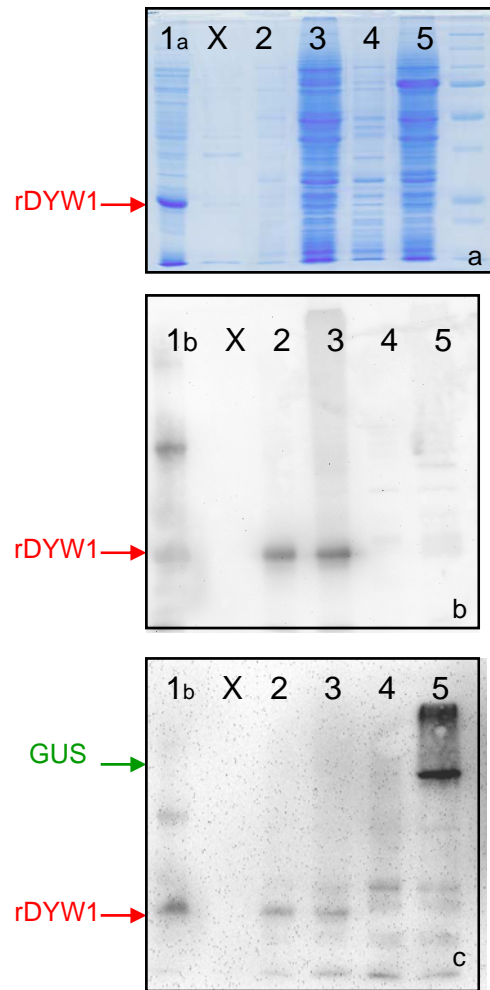


Figure n°91 : Production de la protéine rDYW1 recombinante en RTS, par utilisation d'un lysat de germe de blé

a : Fractions protéiques soluble et insoluble, séparée sur gel SDS-Page coloré en bleu de Coomassie

b: membrane de western blot réalisé sur les fractions protéiques soluble et insoluble, respectivement, avec l'anticorps (AC3) dirigé contre la protéine AtDYW1

c: membrane de western blot réalisé sur les fractions protéiques soluble et insoluble, respectivement, avec l'anticorps dirigé contre le tag 6-his.

Échantillons:

(1a) Protéine d'intérêt produite en système bactérien (souche BL21-SI).

(1b) Protéine d'intérêt produite en système bactérien (souche BL21-SI), et purifiée sur colonne Ni-NTA (~5ng ont été déposés)

(2) Protéine d'intérêt produite à partir de 3 µg de matrice ADN pDEST17/int-stopAtDYW1, fraction protéique insoluble

(3) Protéine d'intérêt produite à partir de 3 µg de matrice ADN pDEST17/int-stopAtDYW1, fraction protéique soluble

(4) Protéine GUS produite à partir d'1µL de matrice ADN (plasmide contrôle fourni dans le kit), fraction protéique insoluble

(5):Protéine GUS produite à partir d'1µL de matrice ADN (plasmide contrôle fourni dans le kit), fraction protéique insoluble

X :échantillons non relatifs à l'expérience décrite

En rouge: rDYW1 indique notre protéine d'intérêt

En vert: GUS indique le témoin GUS

était tout à fait possible de produire cette protéine recombinante dans une souche *E. coli*. De plus, parce que le kit RTS utilisant le lysat d' *E. coli* n'est pas combiné à la technologie CECF, il produit les protéines d'intérêt avec une moins grande efficacité que le kit utilisant le lysat de germe de blé. Pris ensemble, ces éléments suggèrent que la quantité de protéine rDYW1 éventuellement produite avec le kit RTS utilisant le lysat d'*E. coli* serait trop faible pour être mise en évidence à notre seuil de détection.

#### **IV Bilan de la production de la protéine AtDYW1**

Ainsi, la production de la protéine AtDYW1, en système bactérien ou en système *in vitro*, a permis d'obtenir une forme unique de protéine recombinante (rDYW1) correspondant au fragment peptidique dit interne-stop de l'ORF *AtDYW1* fusionné en N-terminal à une étiquette 6-his. Le fait que nous ayons échoué pour produire d'autres formes de la protéine AtDYW1 en système bactérien peut s'expliquer par un problème technique de clonage, mais soulève également la question d'une éventuelle toxicité de la protéine AtDYW1 pour les cellules bactériennes. Il est courant que l'expression des protéines portant un peptide signal vers les organites soit toxique pour une bactérie. On peut alors penser que le peptide signal de la protéine AtDYW1 confère une certaine toxicité à cette dernière en cellules bactériennes. Supportant cette idée, la forme protéique rDYW1 que nous avons réussi à produire ne porte pas de peptide d'adressage. Alternativement, selon notre modèle, la protéine AtDYW1 est une enzyme de modification des ARN catalysant des réactions de désamination C-en-U. Exprimée sous une forme active dans les cellules bactériennes, cette protéine pourrait éditer des ARNm de façon aberrante entraînant une létalité pour les bactéries. Dans le cadre d'une telle hypothèse, il est alors légitime de se demander si la protéine rDYW1 que nous avons réussi à produire est fonctionnelle.

Par ailleurs, au cours de ces études, nous avons remarqué que le point isoélectrique théorique de la protéine rDYW1 est remarquablement élevé (>9), il en est de même pour la forme « sauvage » de la protéine AtDYW1. Ceci suppose qu'à un pH physiologique, dans une cellule végétale, la protéine AtDYW1 pourrait porter une charge nette de surface positive, et ainsi interagir avec des molécules ou peptides chargés négativement, comme par exemple, des acides nucléiques.

Lors de la production de la protéine rDYW1 en système bactérien, nous avons été confrontés à deux problèmes majeurs : tout d'abord, la protéine produite est majoritairement insoluble; par ailleurs, sa purification s'est avérée difficile car cette protéine semble interagir avec d'autres protéines bactériennes dans le cytoplasme d'*E. coli*. Nous avons identifié, par spectrométrie de masse, l'identité des deux protéines bactériennes majoritairement complexées à notre protéine d'intérêt. Il est remarquable qu'il s'agisse, non pas de protéines de type « chaperonnes », mais de protéines de modification des nucléotides, telle qu'une aminotransférase et une ribonucléotide transformylase. Etant donné notre hypothèse quant à la fonction de la protéine AtDYW1, (enzyme catalysant les réactions de désamination C-en-U du processus d'édition pouvant éventuellement agir sous la forme d'un dimère ou multimère) nous pouvons nous interroger sur le sens « biologique » de cette observation. En effet, il est possible que la protéine rDYW1 ne s'associe pas par hasard à ces enzymes de modification des nucléotides, mais parce qu'elles pourraient présenter des homologies structurales. J'ai d'ailleurs, en utilisant un logiciel de prédiction de la structure (PredictProt), vérifié que l'enchaînement des hélices alpha et des feuillettes bêta la protéine rDYW1 et de l'aminotransférase étaient comparables ; en revanche la ribonucléotide transformylase semble

être structurellement différente. Par ailleurs, on peut envisager que la protéine AtDYW1 recrute des protéines de ce type dans la cellule végétale afin de catalyser les réactions d'édition (voir aussi discussion générale, paragraphe III.1). Cependant, comme nous travaillons en système hétérologue, il faut rester prudent quant à l'interprétation que l'on peut tirer de ces observations. Par ailleurs, nous pouvons envisager que la protéine rDYW1, ainsi complexée à d'autres protéines lors de sa production dans les cellules d'*E. coli*, pourrait se trouver dans une conformation fonctionnellement inactive (voir aussi paragraphe V.4).

En utilisant un lysat de germe de blé, j'ai pu produire *in vitro* une petite quantité de protéine rDYW1. D'après les résultats obtenus, le système RTS semble une alternative intéressante pour produire la protéine AtDYW1 car il augmente la solubilité de la protéine rDYW1 ainsi produite. En effet, si on compare la quantité de protéine soluble obtenue rapportée à la quantité totale produite, le système RTS semble plus efficace que la production dans *E. coli*. De plus, bien que je ne l'ai pas réellement testé, il est raisonnable de penser que la pureté de la protéine rDYW1 soit meilleure quand elle est produite de façon « acellulaire ». En effet, le système RTS a été conçu afin de réduire la probabilité de formation de complexe avec des protéines d'origine bactérienne ou végétale, car le lysat utilisé contient théoriquement un nombre « réduit » de protéines nécessaires aux étapes de transcription et de traduction *in vitro*. Enfin, il s'agit d'un système de choix pour produire des protéines présentant une toxicité dans les cellules bactériennes, ce qui pourrait être le cas de la protéine AtDYW1.

Par des analyses de spectrométrie à dichroïsme circulaire, nous avons également envisagé la production d'une protéine rDYW1 dénaturée puis « renaturée ». Nos résultats préliminaires pourraient indiquer que cette alternative est envisageable, notamment par l'utilisation de solvants organiques, mais il faut rester très prudent quant à l'interprétation des résultats que l'on peut obtenir dans ce type d'approches. En effet, la seule « réelle » démonstration d'un repliement dans une conformation fonctionnelle de la protéine rDYW1 serait de montrer qu'elle retrouve ainsi une activité (voir paragraphe V.2).

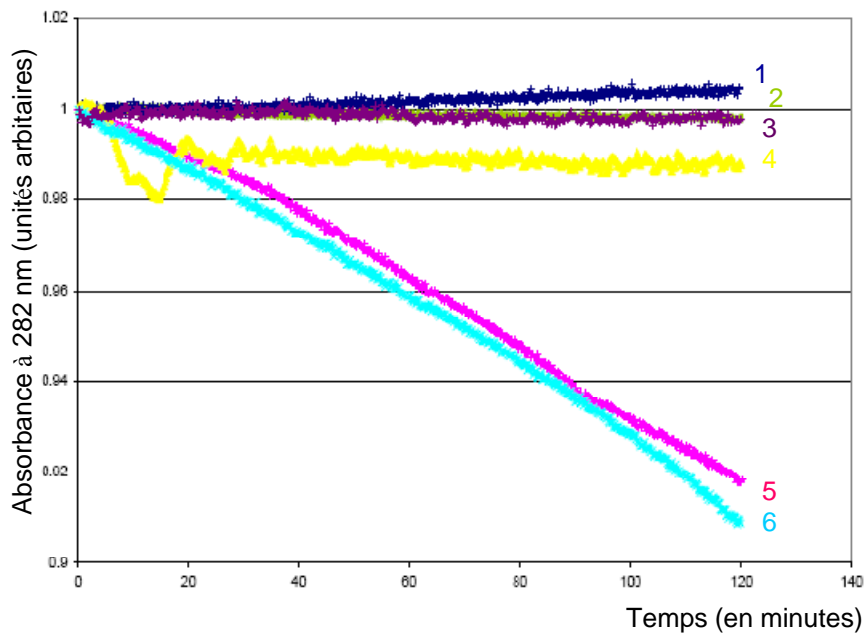
En conclusion, au cours de ma thèse, j'ai testé différentes approches afin de produire la protéine AtDYW1. J'ai pu obtenir une quantité raisonnable de protéine rDYW1 soluble et suffisamment pure pour envisager des approches fonctionnelles (test d'activité). En revanche, la pureté (dans le cas d'une production en système bactérien) ou la quantité (via système RTS) de la protéine ainsi produite sont insuffisantes pour envisager des approches structurales telle que l'obtention d'un cristal de cette protéine.

## **V Une étude « fonctionnelle » de la protéine AtDYW1**

### **V.1 Mesure d'une activité cytidine-désaminase par spectrophotométrie**

Initialement décrite en 1955 (Wang, 1955), cette méthode consiste mesurer l'activité enzymatique d'une cytidine-désaminase par spectrophotométrie. En effet, à 282nm, on observe une décroissance d'absorption du substrat (résidu cytidine ou 2'-deoxycytidine) au fur et à mesure de sa transformation en uracile. L'activité enzymatique de plusieurs cytidine-désaminases agissant sur des résidus cytosine « libre », comme la protéine AtCDA1 d'*Arabidopsis* (Faivre-Nitschke et al., 1999), a été ainsi montrée. Cette méthode a également permis de démontrer l'activité cytidine-désaminase de la protéine CDD1 de levure, homologue de la protéine humaine APOBEC1 (Dance et al., 2001).





#### Légende

- 1: Témoin négatif
- 2: rDYW1 purifiée
- 3: Cellules BL21-SI bouillies
- 4: Cellules BL21-SI bouillies surexprimant la protéine rDYW1
- 5: Cellules BL21-SI surexprimant la protéine rDYW1
- 6: Cellules BL21-SI

Figure n°92 : Cinétiques d'activité pour le substrat cytidine

- en présence d'un lysat bactérien obtenu à partir de cellules BL21-SI surexprimant ou non la protéine rDYW1 : une décroissance similaire et très faible a pu être observée (courbes 5 et 6).
- Cette décroissance disparaît lorsque les lysats sont bouillies (courbes 3 et 4).
- en présence de la protéine rDYW1 purifiée, aucun changement d'absorption de la cytidine n'a pu être observé (courbe 2).
- l'absorbance de la cytidine à 282nm, est stable à 25°C, durant au moins 2 heures (courbe 1).

Afin de déterminer si la protéine AtDYW1 est capable de catalyser une désamination hydrolytique d'un résidu cytidine ou 2'-déoxycytidine, j'ai testé cette approche en utilisant la protéine rDYW1 soluble, produite en système bactérien BL21-SI et purifiée en conditions natives (comme décrit ci avant). J'ai mesuré durant 2 heures, à 25°C, l'absorbance à 282nm de la cytidine ou de la 2'-déoxycytidine en présence de différentes concentrations de protéine rDYW1 purifiée. J'ai également effectué les mêmes mesures en présence de lysat bactérien obtenu à partir de cellules surexprimant ou non la protéine DYW1.

Des observations similaires ont été obtenues en utilisant le substrat cytidine (figure n°92) ou 2'-déoxycytidine (résultats non présentés). Nous observons une faible décroissance de l'absorbance de ces deux substrats en présence de lysats bactériens. L'absorbance de ces substrats est stable en absence d'enzyme, ainsi qu'en présence des lysats bouillis. Il est donc raisonnable de penser que, bien que faible, nous mesurons une réelle activité cytidine-désaminase. Cependant, cette activité est identique pour des lysats contenant ou non une quantité importante de protéine rDYW1. De plus, aucune variation d'absorption n'est mesurée en présence de protéine rDYW1 purifiée. Ceci suggère que l'activité désaminase mesurée n'est pas liée à la protéine rDYW1 elle-même, mais probablement à des cytidines désaminases bactériennes.

De nombreuses hypothèses peuvent être avancées afin d'expliquer l'absence d'activité cytidine-désaminase pour la protéine rDYW1 dans ce test, comme nous le discuterons ci-après (paragraphe V-4 et discussion générale paragraphe III.1).

Cependant, s'il est relativement facile à mettre en place, ce test, parce qu'il utilise des résidus cytidine ou 2'-déoxycytidine « libres » comme substrat, pourrait ne pas être approprié pour mettre en évidence l'activité d'enzymes agissant sur des ARN dans une cellule. D'ailleurs, il a été utilisé sans succès pour la protéine APOBEC1 humaine, catalysant la réaction d'édition du transcrit de l'apolipoprotéine B (MacGinnitie et al., 1995). Ainsi, au cours de ma thèse, j'ai développé d'autres approches afin de tester l'éventuelle activité cytidine-désaminase de la protéine AtDYW1.

## **V.2 Mise en évidence d'une activité cytidine-désaminase sur sondes d'ARN radio-marquées**

### *V.2.1 Description générale du test utilisé*

Ce test permet de mettre en évidence une activité de désamination de résidus radiomarqués dans une sonde d'acide nucléique (ADN ou ARN). En effet, la sonde portant des résidus marqués radioactivement (cytidines par exemple) est incubée en présence de l'enzyme d'intérêt. Puis une digestion nucléotidique à la nucléase P1 est réalisée, et les nucléotides sont séparés par chromatographie sur couche mince de cellulose (TLC). La radioactivité est révélée, et, en cas d'activité désaminase, en plus d'un signal correspondant aux résidus radiomarqués (cytidines, par exemple), un signal correspondant à des résidus désaminés (uridines, dans notre exemple) peut être détecté.

Il s'agit d'une méthode de choix pour mettre en évidence l'activité d'enzymes impliquées dans l'édition des ARN car le substrat est un ARN et non pas une base « libre ». D'autre part,

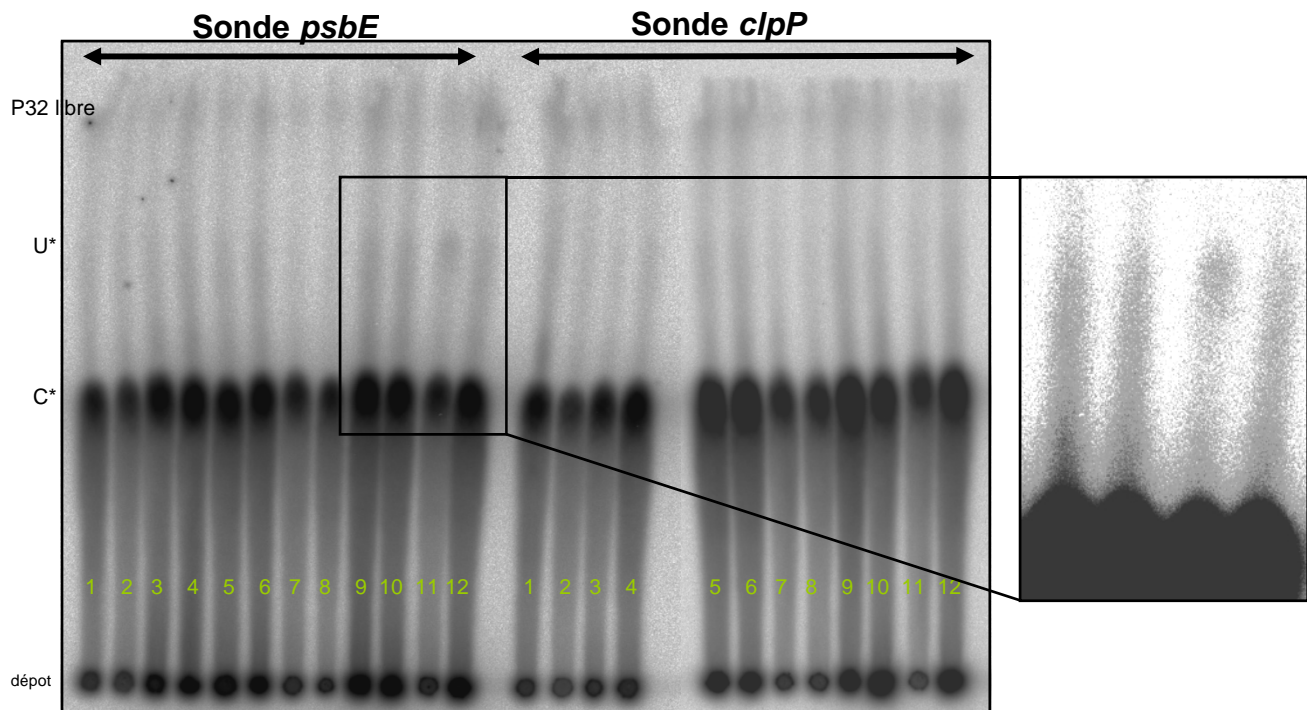


Figure n°93 : Détection d'une activité d'édition sur sondes d'ARN radiomarquées

Echantillons

- 1: lysat bactérien obtenu à partir de cellules BL21-pLys
- 2: lysat bactérien obtenu à partir de cellules BL21-pLys surexprimant la protéine rDYW1
- 3: protéine rDYW1 soluble et purifiée sur billes magnétiques Ni-NTA, produite en cellule BL21-SI,
- 4: protéine rDYW1 soluble et purifiée sur billes magnétiques Ni-NTA, produite en cellule BL21-pLys
- 5: protéine rDYW1 soluble et purifiée sur billes magnétiques Ni-NTA, produite en système RTS à partir d'un lysat de germe de blé.
- 6: protéine GUS soluble et purifiée sur billes magnétiques Ni-NTA, produite en système RTS à partir d'un lysat de germe de blé.
- 7: fraction protéique soluble contenant la protéine rDYW1 non purifiée, produite en système RTS à partir d'un lysat de germe de blé.
- 8: fraction protéique soluble contenant la protéine GUS non purifiée, produite en système RTS à partir d'un lysat de germe de blé.
- 9: protéine rDYW1 potentiellement « renaturée »
- 10: protéine rDYW1 « dénaturée »
- 11: extrait chloroplastique d'Arabidopsis enrichi en protéines solubles (témoin « positif »)
- 12: sonde ARN non incubée en présence de protéine (témoin négatif)

si l'enzyme testée est capable de désaminer plusieurs résidus dans un transcrit, alors cette méthode très sensible car elle permet la détection de l'édition non pas sur un résidu donné du transcrit mais potentiellement pour tous les résidus marqués du transcrit. Cette méthode a été utilisée avec succès pour mettre en évidence l'activité cytidine-désaminase de l'enzyme d'édition humaine APOBEC1 (Petersen-Mahrt and Neuberger, 2003), ou encore l'activité adénosine désaminase de l'enzyme d'édition d'*E. coli* TadA (Kim et al., 2006).

### *V.2.2 Mise en place de ce test pour évaluer une éventuelle activité cytidine-désaminase de la protéine AtDYW1*

#### Choix de la sonde

Etant donné notre hypothèse de travail, il est apparu évident de tester l'éventuelle activité cytidine-désaminase de la protéine AtDYW1 sur une sonde d'ARN portant des résidus cytidines marqués radioactivement. Nous aurions pu tout à fait mener une approche sans *a priori* sur l'ARN « cible » et utiliser un ARN synthétique (tel qu'un ARN polyC) comme sonde dans notre test. Cependant, nous avons choisi une stratégie différente, et synthétisé deux sondes d'ARN correspondant à des fragments des transcrits chloroplastiques AtCg00580 (*PsbE*) et AtCg00670 (*ClpP*).

Les raisons de ces choix ont été multiples. D'une part, les produits PCR chimériques permettant de générer ces 2 sondes, par des réactions de transcription *in vitro*, étaient disponibles au laboratoire de Perth où cette expérience a été mise en place. De plus, Anne-Laure Chateigner-Boutin a pu détecter une activité spécifique d'édition, dans un test *in vitro* utilisant un extrait chloroplastique d'*Arabidopsis* enrichi en protéines solubles, pour la sonde correspondant à un fragment du transcrit chloroplastique *PsbE*. Enfin, d'après nos analyses sur les lignées *AtDYW1* « ARNi » inductibles, la protéine AtDYW1 pourrait être impliquée dans l'édition du site *clpP* (69942).

#### Echantillons protéiques utilisés dans ce test

Lors de notre expérimentation, afin de produire la protéine AtDYW1, nous avons tiré parti des méthodes décrites précédemment. Ainsi, nous avons utilisé 1µg de protéine rDYW1 soluble, produite en système bactérien (souche BL21-SI et BL21-pLys) et purifiée sur billes magnétiques de Ni-NTA. Afin de s'affranchir d'une éventuelle inactivation de la protéine rDYW1 ainsi produite au cours de sa purification, nous avons également testé la potentielle activité cytidine-désaminase d'un lysat total de cellules bactériennes BL21-pLys exprimant ou non la protéine rDYW1.

Par ailleurs, lors de nos expériences de « renaturation » de la protéine rDYW1, nous avons constaté que cette protéine pourrait avoir une certaine conformation dans un tampon contenant des solvants organiques. Ainsi, nous avons utilisé un « aliquote » de protéine rDYW1, produite dans les cellules BL21-pLys, et resuspendue soit dans une solution à forte concentration en urée (forme « débobinée » de la protéine rDYW1, témoin négatif) ou dans un tampon contenant : 50mM TrisHCl pH 7.5 ; 5% glycérol ; 50mM sucrose ; 10µM ZnSO<sub>4</sub>, ~800µM urée (forme « potentiellement repliée » de la protéine rDYW1).

De plus, nous avons utilisé 1µg de protéine rDYW1 soluble et purifiée sur billes magnétiques Ni-NTA, produite *in vitro* par RTS en utilisant un extrait de germe de blé. L'extrait de germe

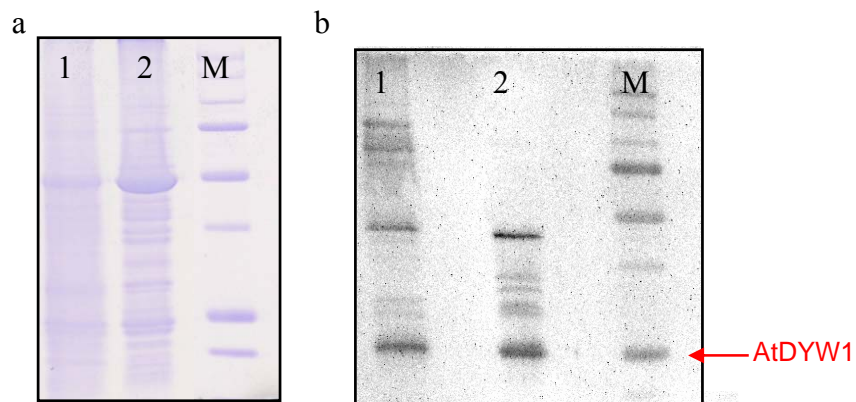


Figure 94 : Détection de la protéine AtDYW1 par des expériences de western blot utilisant un anticorps dirigé contre cette protéine (AC3).

a) Gel colorée en bleu de Coomassie permettant de vérifier la quantité de protéines déposées.

b) Membrane de western blot après révélation

Échantillons:

(1) Fraction de protéines totales d'Arabidopsis

(2) Extrait chloroplastique d'Arabidopsis enrichi en protéines solubles pour lequel une activité d'édition a été démontrée.

de blé peut éventuellement contenir une ou plusieurs cytidines désaminases capables de catalyser des réactions d'édition, et par conséquent, afin d'éviter une éventuelle erreur d'interprétation de nos résultats, nous avons également utilisé 1µg de la protéine GUS produite dans les mêmes conditions (témoin négatif).

Enfin, un extrait chloroplastique d'*Arabidopsis* enrichi en protéines solubles a été préparé au laboratoire et a permis d'obtenir, dans certaines conditions, une édition *in vitro* au site C (64109) du transcrite chloroplastique *psbE* (Anne-Laure Chateigner Boutin, communication personnelle). Ainsi, cet extrait a servi de « témoin positif » dans notre expérience.

### V.2.3 Résultats et interprétations

Les résultats de cette expérience indiquent qu'en utilisant un extrait chloroplastique d'*Arabidopsis* enrichi en protéines solubles sur la sonde *psbE*, nous pouvons détecter un faible signal correspondant à l'apparition d'uridines radiomarquées (témoin positif de l'expérience) (figure n°93). De façon intéressante, par des expériences de western blot utilisant l'anticorps AC3 dirigé contre la protéine AtDYW1, j'ai pu mettre en évidence la présence de cette protéine d'intérêt dans cet extrait chloroplastique d'*Arabidopsis* enrichi en protéines solubles et présentant une activité d'édition des ARN (figure n°94).

Au laboratoire de Perth, Anne-Laure Chateigner-Boutin développe actuellement un autre test d'édition *in vitro* dans lequel elle peut mesurer par PPE, l'apparition de l'édition en un site particulier sur des sondes d'ARN non marquées. Par cette méthode et en utilisant le même extrait chloroplastique d'*Arabidopsis* enrichi en protéines solubles, elle a obtenu des résultats similaires à ceux obtenus dans ce test, à savoir : la mise en évidence d'une activité d'édition au site C (64109) du transcrite chloroplastique *psbE* et pas de détection d'une telle activité au site C (69942) du transcrite chloroplastique *clpP*.

D'après nos analyses sur les lignées *AtDYW1* « ARNi » inductibles, si la protéine AtDYW1 semble impliquée dans l'édition du site *clpP*(69942), elle pourrait être essentielle à l'édition de 3 sites dans les transcrits *ndhD* (117166), *accD* (57868) et *rpoB* (23898). Pour des raisons de commodité nous avons utilisé les sondes disponibles au laboratoire de Perth à savoir *psbE* et *clpP*, mais il est légitime de se demander si nos observations auraient été les mêmes sur d'autres sondes ARN telles que *ndhD*, *accD* et *rpoB*.

Par ailleurs, l'activité d'édition mesurée pour le transcrite *psbE*, en utilisant l'extrait chloroplastique d'*Arabidopsis* enrichi en protéines solubles est très faible dans notre essai (de l'ordre de 1%) et très supérieure (de l'ordre 80%) lorsqu'elle est analysé par PPE (Anne-Laure Chateigner-Boutin, communication personnelle).

Ceci pourrait suggérer que nos conditions de réaction doivent pouvoir être optimisées, mais ce n'est sans doute pas la raison principale expliquant ces différences. En effet, l'extrait chloroplastique utilisé permet une désamination spécifique du résidu C (64109) du transcrite chloroplastique *psbE*. Par PPE, l'efficacité d'édition n'est mesurée qu'à ce site, alors que dans notre essai tous les résidus cytidines de la sonde (~45) sont marqués et donc le signal d'édition spécifique de cette cytidine est « dilué » et doit être de l'ordre de 2% (ce qui correspond grossièrement à ce que nous observons).

Comme nous l'avons déjà souligné, cette méthode est très sensible pour tester l'activité d'une enzyme impliquée dans l'édition des ARN C-en-U, si cette dernière désamine, sans spécifié de reconnaissance, plusieurs voir tous les résidus cytidines des sondes. Dans l'hypothèse où la protéine AtDYW1 agirait sans spécificité, son activité devrait donc être détectée avec

beaucoup plus d'aisance dans les échantillons concentrés en cette protéine purifiée qu'avec le témoin positif de cette expérience. Cependant, il est possible qu'à la différence de ce qui est observé pour de nombreuses désaminases agissant sur des polynucléotides, la protéine AtDYW1 ne désamine qu'avec une très haute spécificité un résidu unique dans une sonde. Dans ce cas, une analyse de l'efficacité d'édition par PPE pourrait être un test plus approprié pour mettre en évidence sa potentielle activité enzymatique.

Enfin, des hypothèses autres que le choix de la sonde ou de la méthode d'analyse de l'efficacité d'édition, peuvent être avancées pour expliquer que nous n'avons pas pu mettre en évidence une activité cytidine-désaminase pour la protéine rDYW1. Nous y reviendrons à la fin de ce chapitre (paragraphe V.4) et en discussion générale (paragraphe III.1).

### **V.3 Reconstitution d'un complexe « minimal » d'édition**

#### *V.3.1 Objectifs de l'approche*

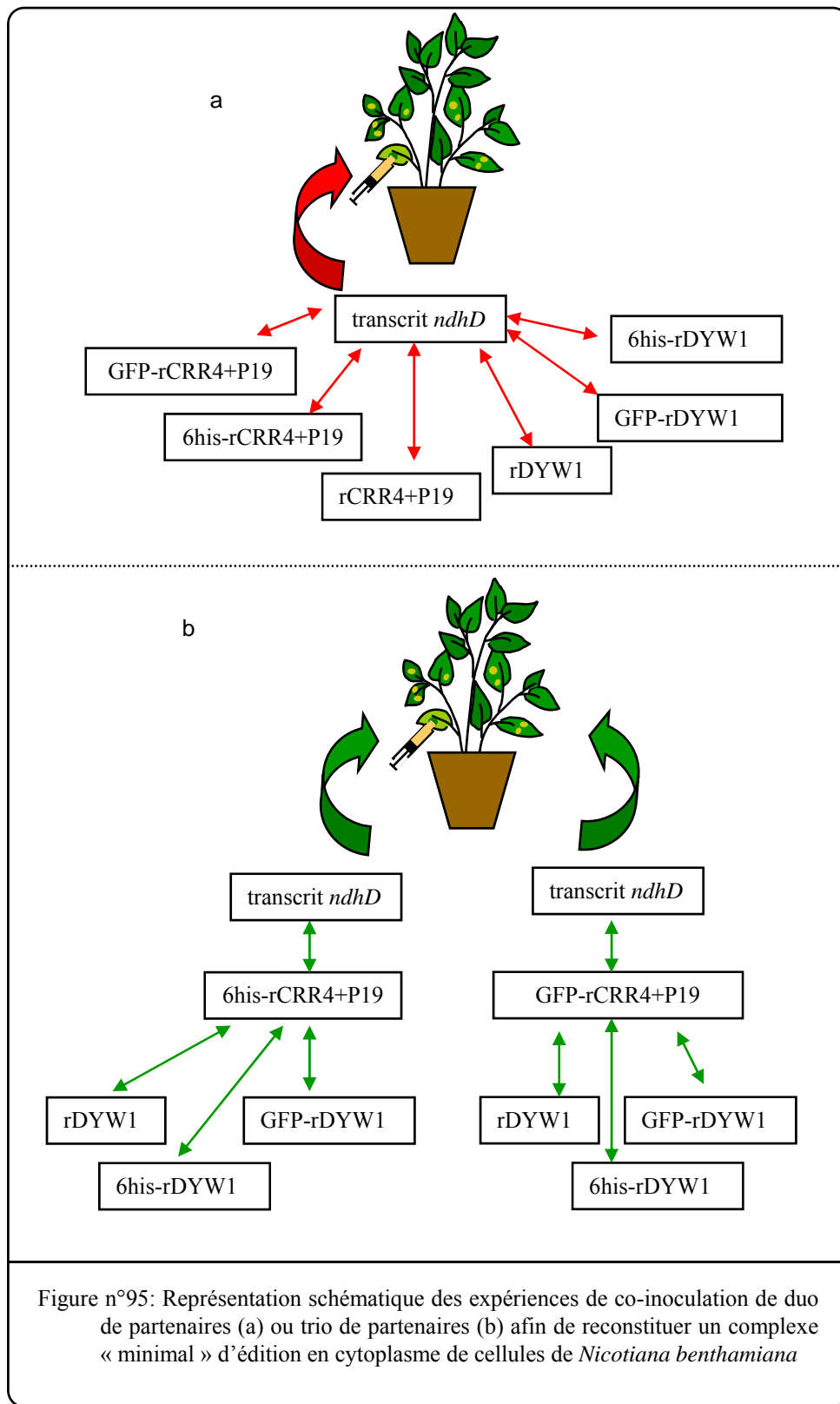
Afin de tester, par une nouvelle méthode, notre hypothèse selon laquelle la protéine AtDYW1 pourrait être le facteur enzymatique catalysant les réactions d'édition C-en-U, j'ai envisagé des approches dont l'objectif était de reconstituer un complexe « minimal » d'édition dans un compartiment cellulaire où ce processus n'existe pas tel que dans la mitochondrie de levure ou le cytoplasme de cellules de tabac. Avec une vue simplifiée, nous pouvons dire que la machinerie impliquée dans les réactions d'édition dans les organites des plantes se compose de trois partenaires essentiels: un transcrit contenant une cytidine à modifier, un facteur de reconnaissance spécifique de la séquence d'ARN encadrant cette base à modifier et une enzyme catalysant la désamination C-en-U.

Selon notre modèle de travail, une protéine PPR pourrait être le facteur de reconnaissance spécifique d'un transcrit et recruterait un facteur enzymatique AtDYW1 afin de catalyser la réaction d'édition. Si cette hypothèse de travail est juste, nous faisons le pari que l'expression de ces trois candidats dans un compartiment cellulaire où ce processus d'édition n'existe pas devrait suffire à induire la réaction d'édition, et que l'absence de l'un d'entre eux devrait abolir ce mécanisme.

Une approche similaire (Lellek et al., 2002) a été décrite chez la levure, où les protéines APOBEC1 et ACF (APOBEC1 complementation factor) ont été exprimées en présence d'un gène chimérique comprenant un fragment du transcrit de l'apolipoprotéine B (apoB) fusionné à un marqueur de sélection. Ainsi, un complexe d'édition a été reconstitué avec succès dans une cellule eucaryote où ce processus n'existe normalement pas.

#### *V.3.2 Reconstitution d'un complexe « minimal » d'édition dans les mitochondries de levure :*

Comme nous l'avons vu en introduction, l'un des premiers mutants affectés dans l'expression d'un gène codant pour une protéine PPR a été décrit chez la levure. En effet, chez cet organisme, la protéine PPR-PET309 participe à la maturation du transcrit mitochondrial *cox1* (Manthey and McEwen, 1995; Manthey et al., 1998). Au cours de mon doctorat, nous avons envisagé une approche de reconstitution d'un complexe « minimal » d'édition dans les mitochondries de levure. Pour se faire, nous avons le projet d'intégrer le domaine DYW de la





protéine AtDYW1 dans la région C-terminale du gène *Pet309* par recombinaison homologue. Ensuite, nous aurions souhaité examiner, si dans la souche ainsi obtenue, la protéine de fusion PET309::DYW pouvait éditer de façon « aberrante » dans le transcrit *cox1*. Malheureusement, par manque de temps, je n'ai pas pu finir la production de la construction à intégrer par recombinaison homologue dans le génome de la levure.

### V.3.3 Reconstitution d'un complexe « minimal » d'édition en cytoplasme de cellules de *N. benthamiana*

#### 1) Mise en place du test

Comme nous l'avons souligné, la protéine PPR-CRR4 d'Arabidopsis est essentielle à l'édition du transcrit chloroplastique *ndhD* (Kotera et al., 2005). Nous avons alors souhaité surexprimer transitoirement les protéines CRR4, AtDYW1 (sans peptide d'adressage pour les organites) et le transcrit *ndhD* en cytoplasme de cellules de *Nicotiana benthamiana* par agroinfiltration afin de reconstituer un complexe « minimal » d'édition.

#### Constructions utilisées

Dans un premier temps, la forme dite « interne/stop » de l'ORF *CRR4*, c'est-à-dire sans peptide signal (délétion arbitraire des 135 premières paires de bases) a été clonée dans le vecteur d'entrée du système Gateway® pDONR207. De la même façon, un fragment du transcrit *ndhD* comprenant 44 pb de part et d'autre du site d'édition C (117166) a été introduit dans le vecteur d'entrée pDONR207. Enfin, pour la protéine AtDYW1, nous avons utilisé le vecteur d'entrée du système Gateway® pDONR207 contenant la forme dite « interne/stop » de l'ORF *AtDYW1*.

Par une réaction de recombinaison de «LR clonase», nous avons cloné les fragments d'ORF *CRR4* et *AtDYW1* dans différents vecteurs de destination permettant l'expression de ces ORF d'intérêt en cellules végétales sous contrôle du promoteur P35s CaMW. Trois vecteurs d'expression différents ont été utilisés : le vecteur pGWB2 permettant l'expression d'une protéine sans étiquette, le vecteur pGWB6 permettant l'expression d'une protéine avec une étiquette fluorescente (GFP) fusionnée en N-terminal du peptide d'intérêt, et enfin le vecteur pGWB9 permettant l'expression d'une protéine avec une étiquette 6-his fusionnée en N-terminal du peptide d'intérêt. Pour le troisième partenaire, nous souhaitons obtenir une expression d'un fragment de transcrit *ndhD*. Ainsi, ce fragment de transcrit *ndhD*, a été introduit uniquement dans le vecteur de destination pGWB2 par une réaction de recombinaison de «LR clonase».

Une souche d'*A. tumefaciens* a été transformée par ces différentes constructions afin de pouvoir réaliser des expériences d'agroinfiltration de feuilles de *N. benthamiana*.

#### Stratégie d'agroinfiltration

Des inoculations et co-inoculation (mélange de souches) ont été réalisées en parallèle afin d'obtenir une surexpression transitoire d'un ou plusieurs partenaires dans le cytoplasme de cellule de *N. benthamiana*. Toutes les combinaisons décrites dans sur la figure n°95 ont été testées.

Ainsi, l'inoculation d'un partenaire seul et la co-inoculation d'un partenaire seul et du vecteur (P19) exprimant une protéine suppresseur de « silencing » ont permis de tester l'expression de chacune de nos constructions à un niveau transcriptionnel ou traductionnel (résultats décrits au paragraphe suivant).

Par ailleurs, l'inoculation des trois partenaires permet de tester si en présence des protéines CRR4 et AtDYW1, le transcript *ndhD* est édité. Enfin, l'inoculation de duos de partenaires permet de vérifier si la présence des 2 partenaires protéiques est nécessaire et suffisante à l'édition du transcript *ndhD*. En effet, en présence de CRR4 ou AtDYW1 seule, on ne s'attend pas à observer l'édition du transcrit cible d'après notre hypothèse de travail.

### Vérification du succès de l'agroinfiltration

Lors de chaque expérience d'agroinfiltration, une souche d'agrobactéries contenant un vecteur permettant la surexpression de la GFP a été inoculée dans des feuilles *N. benthamiana*. Ceci sert de témoin d'agroinfiltration car les feuilles ainsi infiltrées sont soumises à un rayonnement UV. Ces feuilles, due à la surexpression de la protéine GFP, fluorescent et nous permettent ainsi de vérifier que l'agroinfiltration s'est bien déroulée. Cette fluorescence a pu être observée à 2, 4, 6 et 8 jours après inoculation.

#### 2) Vérification de l'expression des différents partenaires

##### A un niveau transcriptionnel:

Via des expériences de RT-PCR, nous avons évalué l'expression des 3 partenaires (infiltré seul ou co-infiltré avec le suppresseur de « silencing » P19) à 2, 4, 6 et 8 jours après inoculation dans les feuilles de *N. benthamiana*. Les ARN totaux des feuilles agroinfiltrées ont été extraits et des ADNc ont été synthétisés par transcription inverse.

En utilisant le couple d'amorces U3-U5, nous avons montré par PCR que le transcrit correspondant au fragment d'ARN cible *ndhD* était exprimé dans toutes ces conditions. De même, en utilisant un couple d'amorces localisé dans le gène *AtDYWI*, nous avons montré que les transcrits correspondants aux différentes formes de protéines recombinantes rDYW1 étaient exprimés dans toutes ces conditions. Enfin, l'utilisation d'un couple d'amorces dans le gène *CRR4*, a permis de détecter les transcrits correspondants aux différentes formes de protéines recombinantes rCRR4 uniquement lors des co-infiltrations avec le suppresseur de « silencing » P19.

##### A un niveau traductionnel, par détection de la fluorescence de l'étiquette GFP

Théoriquement, la vérification de l'expression des protéines d'intérêt au niveau traductionnel est facile à réaliser lorsque ces dernières portent une étiquette GFP. En effet, comme pour le témoin d'agroinfiltration, il suffit de soumettre les feuilles de tabac agroinfiltrées à un rayonnement UV et d'observer l'apparition éventuelle d'un signal fluorescent. Cependant, par cette approche, aucune fluorescence n'a pu être décelée, avec les protéines recombinantes GFP-rCRR4 et GFP-rDYW1, à 2, 4, 6 et 8 jours après inoculation dans les feuilles de *N. benthamiana*.

De ce fait, nous avons également recherché l'apparition d'un signal fluorescent par observation sous microscope des cellules épidermiques des feuilles agroinfiltrées. Cette méthode, plus sensible, nous a permis de mettre en évidence une fluorescence spécifique de la

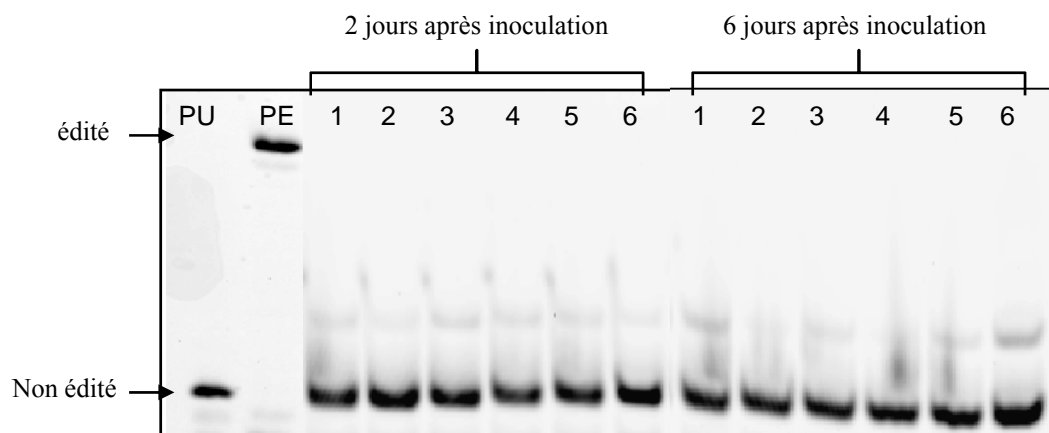


Figure n°96 : Mesure PPE de l'édition éventuelle au site C (117166) du transcrit *ndhD*, 2 et 6 jours après agroinfiltration de trio de partenaires en feuilles de *N. benthamiana*

- (1) rDYW1+ 6his-rCRR4(+P19)+ transcrit *ndhD*
- (2) 6his-rDYW1 + 6his-rCRR4(+P19)+ transcrit *ndhD*
- (3) GFP-rDYW1 + 6his-rCRR4(+P19)+ transcrit *ndhD*
- (4) rDYW1+ GFP-rCRR4(+P19)+ transcrit *ndhD*
- (5) 6his-rDYW1 + GFP-rCRR4(+P19)+ transcrit *ndhD*
- (6) GFP-rDYW1 + 6his-rCRR4(+P19)+ transcrit *ndhD*

PE: plasmide contrôle contenant un produit PCR correspondant à la forme éditée du transcrit *ndhD*

PU: plasmide contrôle contenant un produit PCR correspondant à la forme non-éditée du transcrit *ndhD*

protéine de fusion GFP-rDYW1 à 2, 4, 6 et 8 jours après inoculation dans les feuilles de *N. benthamiana*. En revanche, pour la protéine de fusion GFP-rCRR4, un signal fluorescent spécifique n'a pu être observé à 2, 4, 6 et 8 jours après inoculation, uniquement lors des co-infiltrations avec le suppresseur de « silencing » P19.

#### A un niveau traductionnel, par expériences de western blot

J'ai également vérifié, à 4 jours après inoculation, l'expression de la protéine rDYW1 ne portant pas d'étiquette par des expériences de western blot utilisant l'anticorps (AC3) dirigé contre la protéine AtDYW1. Aucun signal n'a été observé à notre seuil de détection.

De plus, nous ne disposons pas, au laboratoire, d'anticorps dirigé contre la protéine CRR4 et de ce fait je n'ai pas pu vérifier l'expression de la protéine rCRR4 ne portant pas d'étiquette, en cellules de tabac. Enfin, il est tout à fait possible de tester l'expression des protéines de fusion 6his-rCRR4 et 6his-rDYW1, par des expériences de western blot utilisant un anticorps dirigé contre le tag 6his, mais par manque de temps, je n'ai malheureusement pas mené ces analyses.

#### *3) Analyse du niveau d'édition du transcrit cible *ndhD**

Parallèlement à la vérification de l'expression des différents partenaires, j'ai également étudié le niveau d'édition du transcrit *ndhD* lors de l'inoculation de duo ou trio de partenaires. Les ARN totaux des feuilles agroinfiltrées ont été extraits et des ADNc ont été synthétisés par transcription inverse. Pour chaque condition d'agroinfiltration (duo ou trio de partenaires), j'ai amplifié à l'aide du couple d'amorces U3-U5, un produit PCR correspondant au transcrit cible *ndhD*.

#### Séquençage direct de produits PCR

Les produits PCR ainsi obtenus ont été analysés par séquençage direct afin d'étudier l'apparition éventuelle d'édition au site C (117166). Par cette méthode, aucune trace d'édition n'a pu être mise en évidence quelle que soit la combinaison de partenaires testée.

#### Analyse par PPE

Ces mêmes produits PCR ont alors été utilisés pour mesurer, par une analyse PPE, l'édition éventuelle au site C (117166) avec une plus grande sensibilité. A nouveau, aucune trace d'édition n'a pu être mise en évidence quelle que soit la combinaison de partenaires testée (figure n°96).

#### *4) Bilan*

En conclusion, l'édition du transcrit *ndhD* dans le cytoplasme de cellules tabac ne semble pas apparaître lorsque l'on co-exprime la protéine PPR-CRR4 pouvant être le facteur de spécificité, et la protéine AtDYW1 pouvant être le facteur enzymatique de la réaction d'édition.

Dans un premier temps, on peut s'interroger sur la mise au point technique de cette expérience. A-t-on réellement surexprimé les différents partenaires, en particulier les protéines rCRR4 et rDYW1 ? En effet, via des expériences RT-PCR, j'ai pu, relativement

facilement, mettre en évidence l'expression des différents transcrits correspondant aux trois partenaires du complexe « minimal » d'édition. Cependant, la détection au niveau protéique de rCRR4 et rDYW1 s'est révélée plus délicate. En effet, nous avons observé un faible signal fluorescent pour les protéines recombinantes GFP-rCRR4 et GFP-rDYW1, ce qui pourrait suggérer une expression modérée de ces protéines dans les cellules de *N. Benthamiana*. De même, à notre seuil de détection, nous n'avons pas pu mettre en évidence l'expression de la protéine rDYW1 par expérience de western blot en utilisant l'anticorps (AC3). Comme nous l'avons déjà évoqué, il est envisageable que notre anticorps ne soit pas assez sensible pour détecter une faible quantité de protéine rDYW1 dans les cellules de tabac, toutefois ce résultat indique qu'il n'y a pas de « surexpression » de protéine rDYW1 dans cette expérience. Ainsi, il n'existe pas de corrélation entre la quantité de transcrit *rDYW1* et la quantité de protéine correspondante. Comme nous l'avons déjà souligné lors de l'étude des lignées de surexpression de la protéine AtDYW1 chez *Arabidopsis* (chapitre 3, paragraphe IV.3.7), ceci suggère l'existence de mécanisme de régulation traductionnelle ou post-traductionnelle de l'expression du gène *AtDYW1* dans une cellule végétale (voir aussi discussion générale).

Outre ces considérations techniques, par cette approche, nous n'avons pas pu mettre en évidence une activité cytidine-désaminase pour la protéine rDYW1. De nombreuses hypothèses peuvent être avancées pour expliquer ce résultat, comme nous le discuterons au paragraphe suivant. Ceci soulève également des interrogations quant à notre modèle de travail et nous y reviendrons en discussion générale (paragraphe III.1).

#### **V.4 Bilan sur la recherche d'une activité cytidine-désaminase pour la protéine AtDYW1**

Ainsi, quel que soit le test utilisé, nous n'avons pas pu mettre en évidence une activité cytidine-désaminase pour la protéine rDYW1.

Notons, tout d'abord que dans tous ces tests, nous utilisons un fragment peptidique rDYW1 correspondant à la protéine AtDYW1 sans peptide signal (forme dite interne/stop). Pour se faire, nous avons arbitrairement retiré les 42 premiers acides aminés de la protéine AtDYW1. Cependant, selon le logiciel de prédiction TargetP, le site de clivage pour le peptide signal se situe après le vingt-septième acide aminé dans la protéine AtDYW1. Ainsi, les 15 premiers acides aminés de la région N-terminale de la protéine AtDYW1 mature pourraient être manquants dans la forme protéique rDYW1 que nous avons utilisé dans l'ensemble de nos tests. Précédemment, nous avons mis en évidence que la partie C-terminale du domaine DYW était la plus conservée et présentait des similarités avec le site actif de cytidines désaminases (voir présentation des résultats, partie 1). Ainsi, le site catalytique potentiel de la protéine AtDYW1 se localiserait sans doute dans sa partie C-terminale. Cependant il est légitime de s'interroger sur l'impact éventuel de l'absence de ces 15 acides aminés dans la partie N-terminale notamment sur la conformation de la protéine rDYW1 ; et de ce fait sur sa fonction et ses interactions avec ses partenaires potentiels (ARN ou protéines).

Dans la majorité de nos essais, la protéine rDYW1 recombinante utilisée porte une étiquette (6-his ou GFP) dans sa partie N-terminale. Nous avons fait le choix de placer l'étiquette du côté N-terminal, toujours parce que nos études précédentes suggèrent une localisation C-terminale du site catalytique potentiel de la protéine AtDYW1. Malgré tout, on peut se demander si la présence de telles étiquettes ne pourrait pas avoir un impact sur la fonction de la protéine rDYW1, notamment en modifiant son repliement structural. *Nota bene* : Afin de palier à cet éventuel problème, nous avons produit une version protéique rDYW1 sans étiquette pour nos expériences de reconstitution d'un complexe « minimal »

d'édition en cytoplasme de tabac. Cependant, nous n'avons pas pu mettre en évidence que cette protéine était correctement exprimée.

En résumé, dans ces approches, nous ne testons pas l'activité cytidine-désaminase de la protéine AtDYW1 native mais celle d'une protéine rDYW1 recombinante qui pourrait ne pas présenter les mêmes caractéristiques et la même activité que la protéine sauvage.

Par ailleurs, lorsqu'elle est produite en cellules bactériennes, nous avons mis en évidence que la protéine rDYW1 se retrouve complexée à d'autres protéines ce qui pourrait l'inactiver. Bien que nous n'ayons pas évalué cette hypothèse, il est également envisageable que dans le cytoplasme de cellules de tabac, la protéine rDYW1 interagisse avec des protéines absentes dans les organites. Ceci pourrait éventuellement inactiver sa fonction expliquant nos résultats négatifs.

De plus, il est tout à fait possible d'envisager que la protéine AtDYW1 dans les organites recrute des cofacteurs ou d'autres partenaires afin de catalyser la réaction de désamination C-en-U. Ces éléments pourraient être absents dans nos tests d'activité qu'ils soient réalisés *in vitro* ou en cytoplasme de cellule de tabac.

Enfin, même si la protéine AtDYW1 peut catalyser la réaction d'édition, il est envisageable, en particulier parce qu'elle ne possède pas de motif PPR, qu'elle ne puisse pas se fixer d'elle-même sur un ARN. Ceci pourrait expliquer pourquoi nous n'avons pas pu détecter une activité cytidine-désaminase pour la protéine rDYW1 dans notre test sur sonde d'ARN radiomarquée. Par ailleurs, dans notre essai de reconstitution d'un complexe « minimal » d'édition en cytoplasme de tabac, nous avons essayé de palier à ce problème en co-exprimant rDYW1 avec la protéine PPR CRR4. Cependant, pour les mêmes raisons que celles évoquées pour la protéine rDYW1 (délétion de 45 acides aminés dans la partie N-terminale, présence d'une étiquette 6his ou GFP, possibilité d'interaction avec des protéines « contaminantes » dans le cytoplasme des cellules de tabac) il est possible que la protéine rCRR4 ne puissent pas interagir avec son transcrite cible ou la protéine rDYW1. De plus, ces interactions pourraient nécessiter la présence de cofacteurs ou d'autres partenaires absents dans le cytoplasme de cellules de tabac.

En résumé, nos tests d'activité se déroulent *in vitro* ou en système hétérologue, conditions qui pourraient différer de celles existantes dans les organites où se déroulent normalement les réactions d'édition dans la cellule végétale. Ainsi certains éléments nécessaires au processus d'édition, en particulier à l'activité de la protéine rDYW1, pourraient manquer dans nos conditions de test. De même la présence d'éléments normalement absents dans les organites pourrait inhiber l'activité potentielle de la protéine rDYW1.

Enfin, le fait que nous n'avons pas pu mettre en évidence une activité cytidine-désaminase pour la protéine rDYW1 soulève également des interrogations quant à notre modèle de travail : la protéine AtDYW1 est-elle réellement le facteur catalytique des réactions d'édition ? Peut-on envisager un autre rôle pour cette protéine ? Notre modèle de travail est-il correct ? Nous proposons une large réflexion quant à ces questions dans la discussion générale (ci après).

# **Discussion générale**

## **I Le modèle d'étude : un bref rappel**

Depuis leur identification en 2000, de nombreuses études, essentiellement d'ordre génétique, indiquent que les protéines de la famille PPR participent à différentes étapes de maturation des ARN depuis la transcription jusqu'à la traduction dans les organites.

Les protéines PPR sont présentes chez tous les eucaryotes, mais leur nombre a « explosé » chez les plantes, en particulier les angiospermes (Lurin et al., 2004). De même, l'émergence de la sous-famille de protéines PPR de type PLS est spécifique des plantes terrestres (Lurin et al., 2004). Cependant, le génome des organites des plantes ne contient pas un nombre plus important de gènes que ceux des autres espèces. De plus, à l'exception de l'édition des ARN uniquement décrite chez les plantes terrestres, les mécanismes de régulation post-transcriptionnelle de ces gènes ne sont pas plus complexes. Ainsi, afin d'expliquer l'apparition et l'expansion de la sous famille de protéines PPR de type PLS, nous avons rapidement proposé qu'elles puissent être impliquées dans l'édition.

Par ailleurs, le domaine DYW (~100 acides aminés), localisé dans la région C-terminale de certaines protéines PPR de la sous famille PLS, est très conservé et notamment composé de résidus histidines et cystéines quasiment invariants, suggérant qu'il pourrait porter une activité catalytique (Aubourg et al., 2000; Lurin et al., 2004). Rapidement, nous avons proposé que ce domaine puisse être le facteur catalytique dans les réactions d'édition. Nous avons également émis l'hypothèse selon laquelle les protéines PPR de type E/E+ qui ne portent pas ce domaine DYW, recrutent la protéine AtDYW1 en tant que facteur catalytique au cours de la réaction d'édition. Enfin, sachant que de nombreuses cytidines-désaminases agissent sous forme de dimères, nous avons proposé que, pour être active, la protéine AtDYW1 pourrait éventuellement former un homodimère lorsqu'elle est recrutée par des protéines PPR de type E/E+, ou un hétérodimère s'associant au domaine DYW d'une protéine PPR.

Ce modèle de travail soulève deux grandes interrogations. D'une part, est-ce que les protéines PPR de type PLS et la protéine AtDYW1 sont réellement impliquées dans l'édition ? D'autre part, le domaine DYW et la protéine AtDYW1 sont-ils des facteurs enzymatiques dans la réaction d'édition ?

## **II Résultats obtenus en faveur du modèle**

### **II.1 Des arguments phylogéniques en faveur du modèle**

Supportant l'hypothèse selon laquelle le domaine DYW pourrait être impliqué dans l'édition, au cours de mon travail de doctorat, j'ai pu montrer que sa distribution phylogénétique dans lignée verte corrèle strictement avec le phénomène d'édition.

De plus, une comparaison de la famille de protéine PPR d'*Arabidopsis thaliana*, *Oryza sativa* (le riz) et *Physcomitrella patens* (une mousse) a permis de montrer qu'il existe une corrélation entre l'expansion de la sous-famille de protéines PPR-PLS et celle du processus d'édition chez les plantes terrestres (O'Toole et al., 2008).



Par ailleurs, notre modèle de travail implique qu'une protéine semblable à AtDYW1 existe dans le génome de toutes les plantes ayant des protéines PPR de type E/E+. Qu'en est-il dans la réalité ? Nous savons que chez le riz, une ORF (noté OsDYW1), également en copie unique, présente des caractéristiques très similaires à AtDYW1. C'est-à-dire qu'elle pourrait coder pour une protéine composée uniquement d'un peptide d'adressage pour les organites et d'un domaine DYW, sans pour autant comporter de répétitions de motifs PPR. Il est donc envisageable que l'ancêtre commun des protéines DYW1 soit apparu il y a au moins 200 millions d'années, avant la divergence des monocotylédones et des dicotylédones, et que des protéines « DYW1-like » existent chez toutes les angiospermes. D'autre part, il est frappant de remarquer que chez la mousse *Physcomitrella patens*, il n'existe ni de protéine PPR de type E/E+, ni de protéine similaire à AtDYW1 (O'Toole et al., 2008). Cette observation corréle et supporte fortement notre hypothèse de travail.

## **II.2 Des arguments génétiques en faveur du modèle**

Le travail de plusieurs équipes de recherche a permis de mettre en évidence que 4 protéines PPR sont impliquées dans l'édition C-en-U de transcrits chloroplastiques chez Arabidopsis. Trois d'entre elles, CRR4, CRR21 et CLB19 sont de type E/E+ et la quatrième, la protéine FLV est de type DYW (Kotera et al., 2005; Okuda et al., 2007; Chateigner-Boutin et al., soumis; Chateigner-Boutin *et al.*, résultats non publiés). Par conséquent, ces résultats soutiennent fortement que les protéines PPR de la famille PLS participent au processus d'édition dans les chloroplastes chez Arabidopsis.

Au cours de mon doctorat, j'ai pu étudier deux lignées « ARNi » affectées dans l'expression du gène *AtDYW1*. Leur analyse a permis de révéler des plantes présentant des défauts de développement (phénotype « petits choux ») et pour lesquelles l'édition est affectée au niveau de plusieurs sites chloroplastiques mais aussi, bien que cela reste à confirmer, mitochondriaux. Ceci soutient fortement l'hypothèse selon laquelle la protéine AtDYW1 est impliquée dans le processus d'édition au moins dans les chloroplastes chez Arabidopsis. Par ailleurs, le fait que plusieurs sites d'édition soient affectés suggère qu'elle ne joue sans doute pas le rôle de facteur de spécificité, mais plutôt qu'elle soit impliquée au niveau d'une étape commune à tous ces sites, par exemple la catalyse de la réaction de désamination comme le suggère notre modèle. Une des interrogations quant au modèle consiste à se demander si la protéine AtDYW1 intervient au niveau de tous les sites, ou bien uniquement au niveau des sites dont l'édition pourrait être spécifiée par des protéines PPR de type E/E+. Comme nous en avons déjà discuté (troisième partie des résultats, paragraphe III.3.7), l'expression du gène *AtDYW1* n'est que partiellement diminuée dans ces lignées ce qui limite nos interprétations. Cependant, à l'heure actuelle, nos résultats soutiennent que la protéine AtDYW1 pourrait être recrutée au moins par les protéines PPR de type PLS ne possédant pas de domaine DYW.

## **III Des interrogations quant au modèle**

### **III.1 Le domaine DYW et la protéine AtDYW1 sont-ils des facteurs enzymatiques dans la réaction d'édition ?**

Comme évoqué dans la première partie des résultats, les résidus conservés dans le consensus du domaine DYW s'alignent sur le site catalytique de désaminases fixant du zinc, en

particulier les cytidines désaminases impliquées dans les réactions d'édition chez les mammifères, telle que la protéine APOBEC1. Ceci pourrait suggérer que la protéine AtDYW1 et le domaine DYW pourraient agir telle des désaminases fixant du zinc afin de catalyser les réactions d'édition chez les plantes. Cependant, durant mon travail de thèse, j'ai cherché à évaluer l'éventuelle activité de cytidine-désaminase à partir de la protéine AtDYW1 purifiée ou par reconstitution d'un complexe « minimal » d'édition en cytoplasme de tabac. Ces tentatives ont été infructueuses et à l'heure actuelle, il est impossible d'exclure un autre rôle essentiel pour la protéine AtDYW1 ou le domaine DYW dans le complexe d'édition, comme par exemple le recrutement du facteur catalytique.

### **III.2 Le modèle est-il valable pour les deux organites ?**

Il est communément admis que l'édition des ARN est un processus ancien qui est probablement apparu au moment de l'émergence des plantes terrestres il y a plus de 450 millions d'années, et de surcroît simultanément dans les chloroplastes et les mitochondries, suggérant une origine commune de la machinerie d'édition dans ces deux organites. Chez *Arabidopsis*, il existe des protéines PPR de type PLS dans les deux organites. Par ailleurs, une analyse de QTL (quantitative locus trait) a permis d'isoler une région d'ADN, contenant deux gènes PPR de type DYW, qui est essentielle à l'édition dans la mitochondrie chez *Arabidopsis thaliana* (Bentolila et al., 2005). Pris ensemble, ces éléments suggèrent que les protéines PPR de type PLS pourraient être impliquées dans l'édition dans des transcrits mitochondriaux et chloroplastiques, et il est envisageable que le domaine DYW puisse catalyser les réactions d'édition à la fois dans les chloroplastes et les mitochondries.

En revanche, ce modèle suggère également une double localisation chloroplastique et mitochondriale pour la protéine AtDYW1. Or, si nos résultats indiquent une localisation chloroplastique évidente pour cette protéine, sa localisation mitochondriale est incertaine. Si la protéine AtDYW1 est uniquement chloroplastique, cela remet-il en cause notre modèle de travail ? Dans ce cas, quelle hypothèse peut-on avancer quant au fonctionnement de la « machinerie » du processus d'édition dans les mitochondries ?

Nous avons observé la présence de quelques protéines PPR à domaine DYW pour lesquelles le nombre de motif PPR est très restreint (2 à 3). D'après les prédictions de Predotar et TargetP, certaines pourraient être adressées vers les mitochondries. Chez de telles protéines, il est difficile d'envisager que ce petit nombre de répétitions du motif PPR résulte en une structure super-hélicoïdale permettant une reconnaissance spécifique avec un ARN cible. Aussi, nous proposons qu'elles pourraient éventuellement, à l'image de la protéine AtDYW1, être recrutées par des protéines PPR de type E/E+ afin d'apporter l'activité catalytique nécessaire dans le processus d'édition, en particulier dans les mitochondries.

Alternativement, nous pouvons envisager que les enzymes catalysant les réactions d'édition dans les chloroplastes et les mitochondries soient différentes. En ce sens, il apparaît que les réactions d'édition *in vitro* réalisées à partir d'un extrait chloroplastique nécessitent la présence de zinc, alors que ce co-facteur ne semble pas indispensable aux réactions d'édition *in vitro* réalisées à partir d'un extrait mitochondrial (Carter, 1998; Takenaka and Brennicke, 2003; Hegeman et al., 2005). Afin d'expliquer cette observation, il a été proposé que l'enzyme catalysant l'édition dans les mitochondries et les chloroplastes a divergé à partir d'un ancêtre commun au cours de l'évolution (Takenaka et al., 2008). Cependant, il est important de garder à l'esprit que, bien que communément admise, la vision parcimonieuse

d'une origine commune de la machinerie d'édition dans les deux organites reste une hypothèse non démontrée à ce jour.

### **III.3 Combien de protéines PPR sont impliquées dans l'édition ?**

Est-ce que toutes les protéines PPR de type PLS sont impliquées dans l'édition?

Nous l'avons vu en introduction, certaines protéines PPR de type PLS pourraient être impliquées dans d'autres processus que l'édition. Par exemple, chez *Arabidopsis*, la protéine PPR-CRR2 de type DYW participe au clivage du transcrit *rps7-ndhB* (Hashimoto et al., 2003). Cependant ceci n'exclut pas un rôle de cette protéine dans le processus d'édition notamment au niveau d'un site non caractérisé à l'heure actuelle. De plus, les étapes de maturation des ARN ne sont pas distinctes et indépendantes les une des autres. Il faut donc garder à l'esprit qu'une même protéine PPR de type PLS pourrait intervenir aux niveaux de plusieurs mécanismes.

En ce sens, chez *Arabidopsis*, 34 sites d'édition sont actuellement décrits dans les transcrits chloroplastiques, alors qu'un plus grand nombre de protéines PPR de type PLS semblent adressées à ces organites (Lurin et al., 2004; Chateigner-Boutin and Small, 2007). De ce fait, il est possible d'envisager que d'autres sites d'édition chloroplastiques, notamment dans les régions transcrites mais non traduites, restent à identifier.

Alternativement, il se peut qu'au cours de l'expansion de la sous-famille de PPR de type PLS, certains membres aient évolués et acquis de nouvelles fonctions, notamment dans d'autres mécanismes de régulation post-transcriptionnelle, comme pourrait le suggérer la caractérisation de la protéine PPR-CRR2 d'*Arabidopsis*.

Est-ce que les protéines PPR de type P sont également impliquées dans l'édition ?

Les protéines PPR « pures » sont présentes chez tous les eucaryotes ayant un génome mitochondrial. De ce fait il n'existe pas de corrélation phylogénétique entre l'apparition du phénomène d'édition C-en-U dans les organites de plantes et cette sous-famille de protéines PPR. Par ailleurs, l'étude de plusieurs mutants dans des gènes codant pour des protéines PPR-P a permis de montrer que ces protéines intervenaient dans de nombreux mécanismes de maturation des ARN, jamais dans l'édition (au moins à ce jour). Ainsi, bien qu'on ne puisse pas totalement l'exclure, les protéines PPR de la sous famille « pure » ne semblent pas impliquées dans l'édition, et la raison fonctionnelle de l'expansion de cette sous famille de protéines chez les plantes supérieures demeure inconnue à ce jour.

### **III.4 Les protéines PPR de type PLS sont –elles des facteurs de spécificité dans le processus d'édition?**

Plusieurs éléments suggèrent que les protéines PPR de la sous-famille PLS pourraient assurer la reconnaissance d'une séquence « cis-élément » afin de spécifier le site d'édition dans les transcrits chloroplastiques et mitochondriaux de nombreuses espèces de plantes.

Tout d'abord, la modélisation de la structure d'un enchaînement de motifs PPR suggère que les protéines PPR pourraient reconnaître avec un haut niveau de spécificité une séquence ARN d'une dizaine de bases et l'affinité des protéines PPR pour les ARN a été démontrée (Delannoy et al., 2007). Par ailleurs, trois des quatre protéines PPR décrites pour être impliquées dans l'édition sont de type E/E+ et ne présentent aucun domaine permettant de proposer une activité catalytique suggérant qu'elles assurent le rôle de facteur de spécificité dans la réaction d'édition (Kotera et al., 2005; Okuda et al., 2007; Chateigner-Boutin et al., soumis). De plus, il a été montré que la protéine CRR4 peut spécifier le site d'édition dans le transcrit *ndhD* en se fixant sur une région de 35 nucléotides principalement en amont de ce dernier (Okuda et al., 2006). Enfin, la comparaison de la famille de protéines PPR du riz et d'*Arabidopsis*, a révélé que non seulement le nombre de gènes *PPR* est très similaire chez ces 2 espèces (respectivement de 450 et 477 membres), mais aussi de façon toute à fait inattendue, que plus de 92% de ces gènes constituent des paires d'orthologues conservés au cours de l'évolution (O'Toole et al., 2008). Cette observation implique que les protéines orthologues du riz, et d'*Arabidopsis* (et certainement d'autres plantes à fleurs) pourraient assumer les mêmes fonctions dans ces organismes et donc reconnaître les mêmes cibles. Dans le cadre des réactions d'édition, ceci concorde avec l'existence de phénomènes de compétition conservés entre différentes espèces qui suggèrait que les facteurs de spécificité puissent être communs pour au moins certaines espèces de plantes à fleurs (Chaudhuri et al., 1995; Hirose and Sugiura, 2001; Choury and Araya, 2006; Verbitskiy et al., 2006).

Cependant, une évolution dynamique des sites d'édition dans les transcrits des organites chez les plantes a été soulignée, si bien que deux espèces relativement proches présentent différents « *patterns* » d'édition (Freyer et al., 1995; Schmitz-Linneweber et al., 2002; Sasaki et al., 2003). De ce fait, avec une vue parcimonieuse, on s'attend à l'existence d'une co-évolution entre les sites d'édition et leur facteur de spécificité, de sorte que lorsqu'un site d'édition n'est pas conservé entre deux espèces, le facteur correspondant ne le soit également pas.

Cette observation apparaît contradictoire avec l'existence d'une remarquable conservation entre les protéines PPR orthologues du riz et d'*Arabidopsis*. De même, il a été montré que chez le tabac (*Nicotiana tabacum*), le transcrit chloroplastique *ndhA* endogène n'est pas édité. Néanmoins, cette espèce est capable d'éditer la séquence d'un transcrit *ndhA* hétérologue provenant d'autres espèces, en particulier les espèces « ancêtres » *Nicotiana tomentosiformis* et *Nicotiana glauca* (Schmitz-Linneweber et al., 2001; Tillich et al., 2006). Ainsi, de façon surprenante, il semble que le facteur de spécificité du processus d'édition du transcrit *ndhA* a été conservé chez le tabac, bien que le site d'édition correspondant ait disparu.

Une explication possible quant à cette observation consiste à penser que la perte de l'édition du site *ndhA* est récente et que le(s) facteur(s) de spécificité correspondant finira (ont) par être également perdu(s) au cours de l'évolution. Néanmoins, si l'on considère, la relativement grande distance évolutive entre le riz et *Arabidopsis* et la grande conservation des paires de protéines PPR orthologues, cette hypothèse semble moins probable. Alternativement, il a été proposé que les facteurs de spécificité du processus d'édition sont soumis à une pression évolutive afin d'être conservés indépendamment du maintien de l'édition du transcrit cible correspondant. En effet, ces facteurs pourraient avoir évolué et être impliqués dans un nouveau rôle, notamment au niveau d'un nouveau site d'édition (Tillich et al., 2006). En ce sens, nous allons évoquer au paragraphe suivant, que les protéines PPR de type PLS pourraient effectivement être impliquées au niveau de plusieurs sites d'édition. Par ailleurs, nous avons également évoqué au paragraphe précédent que certaines protéines PPR de type PLS, telles que la protéine CRR2 d'*Arabidopsis*, pourraient être impliquées dans d'autres mécanismes de régulation post transcriptionnelle.

### **III.5 Combien de séquences « cis-élément » une protéine PPR de type PLS peut-elle reconnaître ?**

Il existe plus de 400 sites d'édition dans les transcrits mitochondriaux et au moins 30 dans les transcrits chloroplastiques du riz et d'*Arabidopsis* (Giege and Brennicke, 1999; Tsudzuki et al., 2001; Notsu et al., 2002; Chateigner-Boutin and Small, 2007). Ce nombre corrèle approximativement avec le nombre de gènes *PPR* contenu dans le génome de ces organismes (450 pour *Arabidopsis* et 477 pour le riz).

Le modèle jusqu'alors proposé quant au fonctionnement de la machinerie d'édition chez les plantes stipulait que chaque site d'édition est spécifié par un facteur protéique différent. Transposé à la famille de protéine PPR, un tel modèle implique qu'afin de couvrir l'ensemble des sites d'édition, toutes les protéines PPR participent à ce processus et spécifient un site unique dans le transcriptome mitochondrial ou chloroplastique. Or, nous savons qu'un tel modèle est certainement erroné pour deux raisons majeures.

D'une part, il apparaît que toutes les protéines PPR ne sont pas impliquées dans l'édition, en particulier comme nous venons d'en discuter, la sous-famille de protéines PPR « pure », soit environ 250 membres chez le riz et *Arabidopsis* qui semble impliquée dans d'autres mécanismes de régulation post-transcriptionnelle.

D'autre part, un tel modèle impliquerait que seule une trentaine de protéine de la famille PPR soient adressées aux chloroplastes et les quelques 400 autres auraient une localisation mitochondriale. Dans la réalité, les prédictions quant à la localisation des différents membres de la famille PPR pour *Arabidopsis* et le riz ne corrèlent pas avec ces proportions. En effet, bien que la proportion de protéines PPR prédites mitochondriales soit supérieure à celle des protéines PPR chloroplastiques, il apparaît qu'au moins 100 protéines PPR portent un signal d'adressage pour les chloroplastes (Lurin et al., 2004; O'Toole et al., 2008).

Ainsi, un modèle alternatif peut être proposé selon lequel une protéine PPR de type PLS pourrait reconnaître une ou un petit nombre de séquences « cis-élément » et de ce fait être impliquée dans l'édition non pas d'un site unique, mais d'un petit nombre de sites. Ce modèle apparaît contradictoire avec la haute spécificité d'interaction que l'on attend entre une protéine PPR et sa cible ARN (Delannoy et al., 2007). De plus, chez les mutants n'exprimant plus les gènes codant pour les protéines PPR *CRR4*, *CRR21* ou encore *FLV*, un seul site d'édition apparaît affecté. Cependant, supportant ce modèle alternatif, il apparaît que la protéine PPR-*CLB19* est essentielle à l'édition de deux sites dans les transcrits chloroplastiques *rpoA* et *clpP* (Chateigner-Boutin et al., soumis). De même, comme nous l'avons vu en introduction plusieurs études récentes supportent l'hypothèse selon laquelle les séquences « cis-élément » de sites d'édition distincts pourraient être reconnues par un même facteur de spécificité (Chateigner-Boutin and Hanson, 2002; Heller et al., 2008; Kobayashi et al., 2008).

## **IV Perspectives**

### **IV.1 Des données qui pourraient compléter le manuscrit en préparation**

Une grande partie de mon travail de doctorat a consisté en l'étude de la protéine AtDYW1, et les résultats obtenus sont une première étape pour démontrer l'importance de cette dernière dans le processus d'édition chez les plantes supérieures.

Cependant, il nous paraît essentiel de valider expérimentalement le modèle du gène *AtDYW1* par des expériences de RACE-PCR. Ainsi, la longueur exacte du peptide signal de cette protéine pourra être vérifiée, avant de tester à nouveau sa localisation mitochondriale. De même, il faut compléter l'éditotype chloroplastique dans les lignées « ARNi » affectées dans l'expression du gène *AtDYW1*, et confirmer l'existence d'éventuels défauts d'édition dans les transcrits mitochondriaux.

Néanmoins, comme nous l'avons déjà évoqué (troisième partie des résultats, paragraphe III.3.7), le fait que, nous ne puissions ni expliquer comment l'extinction du gène *AtDYW1* est activée dans ces lignées, ni être certains de la spécificité de cette dernière, reste problématique. Si au cours de mon doctorat l'obtention d'un matériel génétique complémentaire s'est avérée infructueuse, récemment, grâce à l'approche « TILLING », de nouvelles lignées portant différentes mutations ponctuelles dans le gène *AtDYW1* sont disponibles au laboratoire d'Evry. Leur analyse est en cours, et de façon préliminaire et encourageante, dans certaines lignées des plantes au développement très affecté apparaissent et pour l'une d'entre elles, l'existence d'un défaut d'édition au niveau du site chloroplastique *rpoB* a déjà été soulignée. Ainsi, la caractérisation de ces lignées constitue une approche complémentaire à l'analyse des lignées « ARNi » affectées dans l'expression du gène *AtDYW1* menée au cours de mon doctorat. Pris ensemble, les résultats de ces études offriront sans doute la possibilité de démontrer plus solidement l'implication de la protéine AtDYW1 dans l'édition.

Par ailleurs, dans les laboratoires de Perth et d'Evry, des approches sont actuellement menées afin de tester l'interaction entre la protéine AtDYW1 et différentes protéines PPR de type E/E+ ou de type DYW. Ces approches sont développées soit en système hétérologue (notamment par des expériences de double-hybride dans la levure), ou *in planta* (via les systèmes dit « split-DHFR ou « split-YFP » ; et des expériences d'échange ou délétion de domaine dans les protéines PPR). L'éventuelle dimérisation de la protéine AtDYW1 pourra également ainsi être évaluée. Les résultats de ces expériences permettront certainement de préciser le modèle d'étude.

Par ailleurs, à la suite de mon travail de doctorat, il paraît également essentiel de poursuivre les investigations afin d'évaluer l'activité cytidine-désaminases de la protéine AtDYW1 purifiée (ou éventuellement d'un domaine DYW d'une protéine PPR) sur des nucléotides libres ou des sondes d'ARN marquées. Dans ce but, une amélioration de la production et de la purification de la protéine AtDYW1 soluble est actuellement poursuivie au laboratoire de Perth car cela peut déterminer le succès des futurs essais. De plus, il s'agit d'une étape préalable à la mise en place d'approches structurales, en particulier l'obtention d'un cristal de la protéine AtDYW1 (ou alternativement d'un domaine DYW d'une protéine PPR), qui offriront peut être la possibilité de confirmer sa ressemblance avec des désaminases fixant du zinc et/ou apporteront de nouveaux éléments quant à sa fonction.

De même, la reconstitution d'un complexe « minimal » d'édition en cellules de tabac ou d'*Arabidopsis*, associée à des expériences de délétions de domaines PPR ou de mutagenèse ponctuelle du domaine DYW (ou de la protéine AtDYW1) sont également des approches poursuivies ou envisagées au laboratoire de Perth et d'Evry. Enfin, des essais de « piégeage » de la protéine AtDYW1 (en utilisant l'anticorps AC3) afin de la soustraire de l'extrait chloroplastique utilisé dans les réactions d'édition *in vitro* sont également prévus.

Néanmoins, nous avons vu en introduction, que chez de nombreux organismes, les réactions d'édition des ARN se font au sein d'un large complexe protéique nommé « éditosome », et

nous pensons qu'il en est peut être de même chez les plantes. Deux réflexions importantes peuvent être faites à partir d'un tel constat.

D'une part, tous les essais de détection de l'activité enzymatique sur la protéine AtDYW1 purifiée (ou alternativement le domaine DYW d'une protéine PPR) restent très incertains car l'activité pourrait dépendre de la présence d'autres éléments du complexe. Dans un tel cas, il est alors légitime de s'interroger de la sorte : « Existe-t-il un test qui pourrait démontrer formellement que la protéine AtDYW1 ou le domaine DYW sont les facteurs enzymatiques centraux de la réaction d'édition ? »

D'autre part, l'isolement de l'« éditosome » chez les plantes et l'identification des différents partenaires de ce complexe permettraient sans doute des avancées considérables dans la compréhension du fonctionnement de la machinerie d'édition. En ce sens, l'étude des complexes ribonucléiques dans lesquels se trouve la protéine AtDYW1 ou les protéines PPR de type PLS, constituerait une première étape dans la caractérisation moléculaire et fonctionnelle du complexe d'« éditosome » chez les plantes.

#### **IV.2 Autre perspective : vers une meilleure compréhension de l'interaction PPR/ARN substrat**

D'une façon générale, l'un des enjeux majeur de l'étude de la famille PPR reste la compréhension moléculaire de l'interaction (supposée directe) entre une protéine PPR et une molécule d'ARN. En effet, élucider le mécanisme moléculaire permettant l'interaction entre une protéine PPR et son (ses) transcrit(s) cible(s) permettra sans doute de mieux comprendre comment se fait la spécificité de reconnaissance de la base à éditer au sein d'un transcrit.

Ainsi, comme évoqué précédemment, l'étude des complexes ribonucléiques dans lesquels se trouvent les protéines PPR peut être envisagée par des approches de purification par affinité et analyse des transcrits (par RIP-chip) et des protéines (en spectrométrie de masse).

D'autre part, la confrontation de la séquence des motifs PPR à celles de leurs cibles ARN (en particulier les séquences « cis-éléments ») offrira sans doute la possibilité de réaliser une analyse fine de la relation entre la structure et la fonction du motif PPR et éventuellement de « décoder » la relation entre successions de motifs PPR et séquence d'ARN. Réciproquement, des éléments déterminants pour cette interaction dans les séquences cibles (résidus essentiels et éléments structuraux) seront peut être ainsi révélés. Par la suite, ces prédictions pourront être confirmées par des expériences de mutagenèse dirigée des motifs PPR et des séquences « cis-éléments ».

Enfin, nous avons vu en introduction que chez les autres organismes lors des réactions d'édition, la structure des ARN substrats évolue afin de rendre la base à éditer accessible au site catalytique des désaminases (mouvement dit de « base-flipping »). Ainsi, il serait intéressant d'étudier, par résonance magnétique nucléaire, l'évolution de la structure d'un ARN substrat, en présence ou non d'une protéine PPR de type PLS, lors d'une réaction d'édition *in vitro*. En effet, ceci permettrait d'évaluer si les protéines PPR peuvent faciliter un tel mouvement lors de la réaction d'édition chez les plantes.

## **V D'autres pistes de travail**

### **V.1 Vers une compréhension de l'origine et de l'évolution des protéines PPR, du domaine DYW et de la protéine AtDYW1**

L'origine de la famille PPR demeure actuellement obscure. En effet, ces protéines sont absentes du génome des procaryotes (y compris dans les espèces pouvant être proches des ancêtres endosymbiontes des mitochondries). En revanche, elles sont présentes chez tous les eucaryotes ayant un génome mitochondrial et l'expansion de cette famille est spectaculaire chez les plantes terrestres. Le mécanisme d'expansion des protéines PPR chez les plantes terrestres, *via* des événements massifs de rétrotransposition, a été récemment élucidé (O'Toole et al., 2008). Néanmoins, hormis le cas de la sous-famille de type PLS éventuellement liée au processus d'édition, la raison fonctionnelle de cette expansion reste inconnue à ce jour.

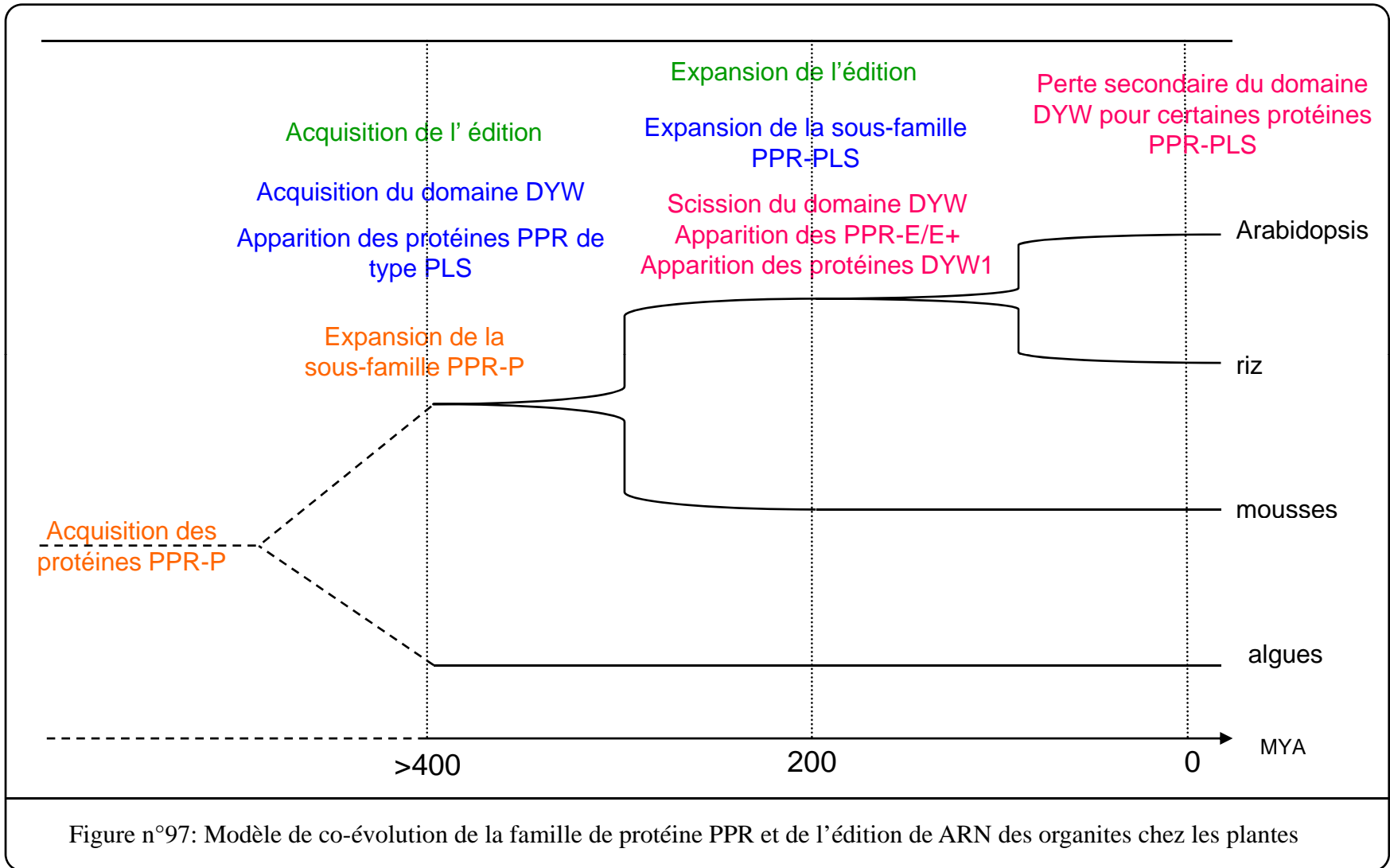
L'origine du domaine DYW est restée également longtemps mystérieuse. En effet, ce domaine est spécifique des plantes terrestres. Il a été montré que de nombreux gènes spécifiques des plantes ont été acquis lors du transfert de gènes des ancêtres endosymbiontes des chloroplastes vers le génome nucléaire de leur cellule hôte. Cependant, le domaine DYW semble absent du règne procaryote (y compris dans les espèces pouvant être proches des ancêtres des organites de la cellule végétale), et donc cette hypothèse ne semble pas expliquer son origine. Par ailleurs, au cours de mon doctorat, j'ai pu souligner l'existence d'homologies entre le domaine DYW et le site actif de cytidines désaminases fixant le zinc. Ainsi, à l'image de ce qui est décrit pour les protéines de la famille ADAR ou AID/APOBEC impliquées dans l'édition chez les mammifères, l'apparition du domaine DYW chez les plantes pourrait être liée à l'évolution d'une désaminase ancestrale agissant sur des mononucléotides qui aurait acquis la possibilité d'agir sur un substrat polynucléotidique (un ARN en particulier).

Une question sous-jacente à l'origine du domaine DYW concerne celle de la protéine AtDYW1. En effet, on peut s'interroger de la sorte : est-ce que les protéines PPR de type DYW résultent d'une fusion entre des protéines PPR de type E/E+ et d'une protéine ancestrale uniquement composée du domaine DYW, ou bien, à l'inverse est-ce que la protéine AtDYW1 est apparue relativement plus récemment au cours de l'évolution par scission du domaine DYW des protéines PPR ?

Comme nous l'avons déjà souligné, bien que la protéine AtDYW1 ne possède pas de « vrai » motif PPR, elle comporte une région présentant des homologies avec le motif PPR et qui pourrait correspondre à un motif PPR dégénéré. Ceci supporte l'hypothèse d'une scission d'une protéine PPR à motif DYW pour expliquer l'origine de la protéine AtDYW1. L'existence d'une ORF similaire chez le riz (OsDYW1) suggère que la scission du domaine DYW des PPR permettant l'apparition des protéines de « DYW1-like » a eu lieu, il y a au moins 200 millions d'années, avant la divergence des monocotylédones et des dicotylédones. De plus, l'absence de protéine de type DYW1 et de protéines PPR de type E/E+ chez *Physcomitrella patens* suggère que cette scission est postérieure à l'apparition des mousses, il y a 400 millions d'années.

Nous avons remarqué précédemment que l'expansion de la sous-famille PPR de type PLS et celle du processus d'édition corrélaient chez les plantes. Selon le modèle évolutif du processus d'édition chez les plantes, proposé par Covello et Gray (Covello et Gray, 1993), à l'émergence des plantes terrestres une protéine ancestrale aurait acquis la capacité d'éditer un ou un petit nombre de sites. Au cours de l'évolution, l'accumulation de mutations dans le





génomique des organites aurait entraîné une multiplication des sites d'édition et rendu ce processus indispensable à la survie des plantes. L'ensemble des données quant à l'évolution de la sous-famille de protéine PPR de type PLS, nous permet aujourd'hui de proposer un modèle cohérent de « co-évolution » entre l'édition et l'évolution de la famille PPR (figure n°97).

Selon ce modèle, chez l'ancêtre commun de tous les eucaryotes une protéine PPR « pure » est apparue, et s'est modérément dupliquée au cours de l'évolution dans toutes les branches évolutives (<10 membres chez la majorité des eucaryotes). Lors de l'émergence des plantes terrestres, les protéines PPR de type P se sont dupliquées massivement par des événements de rétrotransposition et les protéines de type PLS sont apparues.

Une protéine PPR ancestrale a acquis le domaine DYW et de ce fait la capacité de catalyser les réactions d'édition des ARN dans les organites. Actuellement, nous ne disposons pas de suffisamment d'informations pour déterminer si la première protéine PPR ayant acquis le domaine DYW était de type PPR-P et a ensuite évolué en type PLS ; ou si le scénario évolutif inverse, selon lequel l'apparition des protéines PPR de type PLS est antérieure à l'acquisition du domaine DYW, s'est déroulé.

Au cours de l'évolution, avant la divergence entre les monocotylédones et les dicotylédones, d'une part la scission du domaine DYW aurait permis de donner naissance aux protéines de type DYW1 et aux protéines PPR de type E/E+, et d'autre part, une multiplication massive des protéines PPR de type PLS aurait eu lieu.

Au cours de ces événements de multiplication, les protéines PPR auraient pu accumuler des modifications mineures leur permettant de reconnaître de nouveaux transcrits cibles, expliquant l'expansion du processus d'édition chez les angiospermes. Par ailleurs, ces modifications des protéines PPR de type PLS peuvent résulter également en une diversification de leur fonction, suggérant ainsi l'existence éventuelle de protéines PPR de type PLS impliquées dans d'autres mécanismes de régulation post-transcriptionnelle que l'édition.

De façon intéressante, la comparaison de la sous famille de protéines PPR de type PLS chez le riz et *Arabidopsis* a révélé l'existence de 13 paires de protéines orthologues pour lesquelles l'un des membres ne porte plus le domaine DYW. Ceci pourrait suggérer une « récente » perte secondaire de ce dernier. D'un point de vue fonctionnel, nous envisageons que suite à l'apparition des protéines de type DYW1 dans le génome de certaines espèces d'angiospermes, le domaine DYW des protéines PPR pourrait ne plus être indispensable pour catalyser les réactions d'édition. De ce fait, ce domaine pourrait être amené à disparaître au cours de l'évolution.

## **V.2 Edition des ARNt dans les organites des plantes**

Chez les plantes, les sites d'édition actuellement décrits se retrouvent majoritairement dans les ARNm codant des protéines, mais des sites édités dans les ARN de transferts ont également été répertoriés. Par exemple, dans les chloroplastes, il existe un site d'édition C-en-U dans un ARNt décrit chez une espèce de mousse et un site d'édition A-en-I en position 34 de l'ARNt<sup>Arg</sup> de nombreuses espèces (Sprinzl et al., 1998; Miyata et al., 2008). Dans les mitochondries, plusieurs sites d'édition C-en-U ont également été décrits dans des ARNt, en revanche aucune inosine n'a été identifiée (Sprinzl et al., 1998; Verbitskiy et al., 2006).

Nous avons vu, en introduction, que le site d'édition A-en-I en position 34 de l'ARNt<sup>Arg</sup> était commun à plusieurs organismes procaryotes et eucaryotes, si ce n'est pas tous, et que chez de nombreuses espèces les protéines de la famille ADAT catalysent ces réactions d'édition

(Grosjean et al., 1995; Sprinzl et al., 1998; Gerber and Keller, 1999). Par ailleurs, chez le trypanosome, il a été récemment montré que deux protéines ADATs forment un hétérodimère capable de catalyser à la fois des désaminations C-en-U et A-en-I dans un ARNt (Rubio et al., 2007).

Ainsi, il est légitime d'imaginer que, chez les plantes, des protéines ADATs peuvent catalyser à la fois les réactions d'édition C-en-U et A-en-I des ARNt ; et que certaines d'entre elles pourraient avoir évolué afin de catalyser la réaction d'édition C-en-U des ARNm. Alternativement, sachant que les protéines ADATs appartiennent à la super famille des désaminases fixant le zinc, et que le domaine DYW des protéines PPR présente des homologies avec le site actif des enzymes de cette famille, on peut émettre l'hypothèse selon laquelle le domaine DYW des protéines PPR assure à la fois les réactions d'édition C-en-U et A-en-I des ARNt, et C-en-U des ARNm dans les organites.

### **V.3 Edition des ARNm nucléaires?**

A l'heure actuelle, il n'existe aucune démonstration quant à l'existence d'édition dans les ARNm nucléaires chez les plantes. Cependant, il n'est pas exclu que des conversions de bases (C-en-U, A-en-I ou U-en-C) puissent survenir dans un certain nombre de messagers (Sébastien Aubourg et Jean Colcombet, communication personnelle). Le développement de nouvelles technologies d'analyse du transcriptome, basées sur le séquençage massif d'ADNc provenant d'un extrait cellulaire, nous permettra, dans un avenir proche, d'accumuler de nombreuses données sur la séquence des transcrits et d'évaluer ainsi l'existence éventuelle d'édition dans les transcrits nucléaires chez les plantes. Dans un tel cas, au moins deux pistes de travail pourraient être explorées.

D'une part, comme nous l'avons évoqué en introduction, quelques protéines PPR pourraient avoir une localisation nucléaire ou cytosolique. Si de telles protéines existent réellement alors il est raisonnable de proposer que, comme dans les organites, elles participeraient à des mécanismes de régulation post transcriptionnelle de l'expression de gènes. De ce fait, il serait intéressant d'évaluer leur implication potentielle dans des mécanismes d'édition des ARNm nucléaires.

D'autre part, on peut s'interroger sur le rôle de l'édition des ARNm nucléaires dans la biologie de la plante. Par analogie à ce qui a été décrit chez les métazoaires, on peut envisager que l'édition permette une diversification de l'information codée par le génome des plantes. De plus, d'autres rôles peuvent être aussi envisagés telle qu'une implication dans la régulation de l'expression de gènes *via* la voie de l'ARN interférent, ou encore dans la défense contre les virus ou le maintien de l'intégrité du génome nucléaire

### **V.4 Une expression des protéines PPR et de la protéine AtDYW1 extrêmement régulée dans la cellule végétale**

Nous avons vu que l'expression des gènes PPR (notamment ceux de la sous-famille PLS) et celle du gène *AtDYWI* est faible dans les différents tissus d'*Arabidopsis* (Lurin et al., 2004)(Jean-Pierre Renou, communication personnelle). De même, la demi-vie des ARNm codant pour des protéines de la famille PPR est remarquablement plus courte que celle d'une vaste majorité de gènes chez cette espèce (Narsai et al., 2007). Par ailleurs, une étude a révélé que de nombreux gènes *PPR* pourraient être «ciblés » par des tasiARN (Howell et al., 2007). Pris ensemble, tous ces éléments suggèrent que l'expression des gènes *PPR*, et

éventuellement celle du gène *AtDYW1*, est extrêmement régulée chez *Arabidopsis*, et que cette régulation pourrait se faire par des mécanismes d'interférence à l'ARN.

Au cours de mon doctorat, au moins deux éléments sont venus supporter cette hypothèse. D'une part, lors des essais de reconstitution d'un complexe « minimal » d'édition en cytoplasme de tabac, l'expression du transcrite correspondant à la protéine PPR-CRR4 n'a été possible qu'en co-exprimant un gène suppresseur de « silencing ». D'autre part, l'expression d'une cassette portant le gène *AtDYW1* sous contrôle d'un promoteur fort (P35S) dans des lignées transgéniques d'*Arabidopsis* semble induire un mécanisme « silencing » de cette cassette. Nous avons également noté que la présence de transcrits correspondant à une copie transgénique du gène *AtDYW1* chez le tabac et chez *Arabidopsis* ne conduit pas à une expression de la protéine correspondante, suggérant qu'il existe, au moins pour ce gène, un autre niveau de régulation de son expression.

Face à ces observations, on peut se demander quels sont les mécanismes moléculaires permettant de réguler le « silencing » des gènes codant pour des protéines PPR et éventuellement la protéine *AtDYW1*. Nous disposons de peu d'éléments pour répondre à cette question actuellement. Cependant, comme nous l'avons évoqué en introduction, il a été montré que la protéine ADAR2, catalysant les réaction d'édition A-en-I des ARNm chez les animaux, « auto régule » son expression par édition de son propre transcrite. Serait-il possible d'envisager que les protéines PPR ou la protéine *AtDYW1* pourraient « auto réguler » leur propre expression ? Autrement dit, existe-t-il un lien entre la voie de « silencing » et l'édition chez les plantes ? Si tel est le cas, ceci expliquerait sans doute les difficultés expérimentales rencontrées quant à l'établissement de lignées affectées dans l'expression du gène *AtDYW1*. Cependant, il faut garder à l'esprit que la protéine *AtDYW1* et la majorité des protéines PPR sont adressées aux organites, alors que les voies de « silencing » sont nucléaires et cytosoliques. Ainsi, si un lien existe entre ces mécanismes, il pourrait ne pas être direct et de ce fait difficile à mettre en évidence.

On peut également s'interroger sur le sens biologique d'une régulation aussi fine de l'expression du gène *AtDYW1* et des gènes codant pour des protéines PPR chez *Arabidopsis*. Une première hypothèse consiste à proposer que la régulation de l'expression du gène *AtDYW1* permet de « contrôler » le niveau d'édition des transcrits des organites. En ce sens, il apparaît que la surexpression du facteur de spécificité PPR-CRR4 n'entraîne pas une sur-édition du transcrite cible *ndhD* (Shikanai, 2006). Ainsi, l'enzyme catalysant la réaction d'édition (ou un autre composant de la machinerie d'édition) pourrait être l'élément limitant dans la réaction d'édition.

Une autre hypothèse, selon laquelle la régulation de l'expression des gènes PPR participe à la régulation de la production de l'énergie dans la cellule végétale, est attrayante à deux niveaux. D'une part, les protéines produites par les gènes des chloroplastes et des mitochondries sont indispensables aux processus de photosynthèse et de respiration et l'expression de ces gènes dépend majoritairement d'une régulation post-transcriptionnelle à laquelle les protéines PPR participent. D'autre part, selon la théorie de l'origine endosymbiotique des organites, ces derniers dérivent d'organismes procaryotes ancestraux dont le génome aurait subi d'importants réarrangements au cours de l'évolution, en particulier le transfert de nombreux gènes essentiels à leur fonctionnement vers le génome nucléaire. De ce fait, dans la cellule végétale, l'expression des trois génomes doit être coordonnée. Chez *Arabidopsis thaliana*, l'existence de signaux multiples provenant des chloroplastes et permettant de réguler la transcription de gènes nucléaires codant pour des protéines adressées à ces organites a été révélée par l'étude de mutants GUN (pour genomes uncoupled) (Gray et al., 2003). Les mutants GUN sont affectés dans l'expression de gènes nucléaires, et déficients dans la

perception des signaux provenant des chloroplastes. Récemment, il a été montré que l'un d'entre eux correspondait à un gène codant pour une protéine PPR, suggérant une implication des protéines PPR dans ces signaux de coordination de l'expression des gènes des trois génomes dans la cellule végétale (Koussevitzky et al., 2007).

Ainsi, si les mécanismes moléculaires de maturation des ARN des organites, et l'implication des protéines PPR, sont aujourd'hui de mieux en mieux caractérisés, comprendre leur régulation et leur rôle au niveau cellulaire et au cours de la vie de la plante reste l'un des enjeux des prochaines années de recherche (biologie intégrative).

## **VI Une « critique » de l'approche menée au cours de mon doctorat**

Comme son titre l'indique, mon travail de doctorat a consisté en l'« identification du facteur catalytique impliqué dans le processus d'édition des ARN des organites chez les plantes ».

### **VI.1 Le danger du modèle**

Pour ce faire, nous avons proposé un modèle de travail selon lequel le domaine DYW des protéines PPR de type PLS ou la protéine AtDYW1 pourraient être le(s) facteur(s) catalytique(s) du processus d'édition chez les plantes. Ainsi, au cours de mon doctorat, j'ai multiplié le nombre d'approches afin de tester cette hypothèse.

A l'heure actuelle, nous disposons d'un certain nombre d'éléments étayant l'implication des protéines PPR de type PLS et de la protéine AtDYW1 dans les réactions d'édition chez *Arabidopsis*. Cependant, l'hypothèse selon laquelle les domaines DYW de certaines protéines PPR et la protéine AtDYW1 assurent l'activité catalytique des réactions reste actuellement bien moins démontrée. Il est alors légitime de se demander si notre modèle de travail est correct, et dans le cas contraire, combien d'échecs et de temps investi sont nécessaires avant de remettre en cause ce modèle ou alternativement l'ajuster.

Nous avons déjà proposé une alternative quant au rôle du domaine DYW ou la protéine AtDYW1 dans le processus d'édition. En effet, ils pourraient participer au recrutement du facteur catalytique de l'édition. Ainsi, le développement d'approches visant à identifier les partenaires des protéines PPR à domaine DYW ou de la protéine AtDYW1 paraît judicieux.

Par ailleurs, notre modèle de travail ne propose qu'une vue très simplifiée du fonctionnement de la machinerie d'édition. En ce sens, il est forcément inexact et ne reflète pas la réalité de la complexité du processus d'édition chez les plantes ; notamment il ne tient pas compte des mécanismes de régulation de l'édition et de ses conséquences sur la biologie de la plante. Ceci pourrait expliquer certaines difficultés expérimentales rencontrées au cours de mon doctorat, en particulier pour éteindre l'expression du gène *AtDYW1* et tester l'éventuelle activité catalytique de la protéine correspondante.

Enfin, les éléments dont nous disposons ne constituent en aucun cas une preuve absolue d'une implication directe du domaine DYW des protéines PPR ou de la protéine AtDYW1 dans le processus d'édition. Il faut donc rester « attentif » à toute nouvelle piste de travail qui pourrait se présenter.

## **VI.2 Et si c'était par la fin que tout commençait...**

A ce jour, il est légitime de se demander si l'identification du facteur catalytique d'édition, sans *a priori*, est envisageable. En ce sens, je crois qu'effectivement une approche visant à purifier le complexe d'« éditosome » chez les plantes et identifier les différents facteurs qui le compose serait moins restrictive que les études menées au cours de mon doctorat. Cependant, des investigations menées en sens par plusieurs équipes de recherche n'ont pas encore portées leur fruit.

### ***Le mot de la fin***

Le « tâtonnement » n'est-il pas l'un des « moteurs » de la recherche ? Autrement dit, les expériences infructueuses ne permettent-elles pas d'avancer ? Je termine ce manuscrit sur une anecdote illustrant cette idée.

Lors d'une interview, un grand chercheur avouait avoir « échoué » deux cents fois avant de trouver l'explication d'un phénomène scientifique. La journaliste lui demande alors : « Et cela ne vous a pas découragé de vous tromper aussi souvent ? » Sa réponse fût : « Je ne me suis jamais trompé ! J'ai découvert deux cents mauvaises méthodes »

Propos rapportés par le Docteur Susan Jeffers

# **Matériel et méthodes**

## **I. Matériel biologique**

### **I.1 Les espèces végétales**

#### **-*Arabidopsis thaliana***

L'essentiel de mon travail de doctorat a été développé sur la plante modèle *Arabidopsis thaliana*. Les écotypes *Columbia* (noté Col0) et *Wassilevskija* (noté WS) ont été utilisés.

#### **-*Nicotiana benthamiana***

Les expériences d'expression transitoire de protéines en cellules végétales ont été réalisées sur une plante modèle proche du tabac, *Nicotiana benthamiana*.

### **I.2 Les souches des bactéries**

#### **- La souche DH5 $\alpha$**

La souche d'*Escherichia coli* DH5 $\alpha$  est couramment utilisée dans de nombreux laboratoires de biologie, notamment pour les clonages et la multiplication plasmidique, car il est possible de la rendre très compétente (de l'ordre de  $10^7$ - $10^8$  transformants par  $\mu\text{g}$  d'ADN) pour une transformation par choc thermique. Son génotype est : *supE44*,  $\Delta$ *lacU169*( $\phi$ 80*lacZ* $\Delta$ M15), *recA1*, *endA1*, *hsdR17*, *thi-1*, *gyrA96*, *relA1*

#### **- La souche DB3.1**

DB3.1 est une souche bactérienne d'*Escherichia coli* modifiée génétiquement. Sa principale caractéristique est de posséder la mutation *gyrA462* qui protège de l'effet toxique du gène *ccdB*. Nous utilisons cette souche pour multiplier les vecteurs donneurs et destination du système Gateway (voir paragraphe II.1.6.3). Le génotype de cette souche est : F<sup>-</sup> *gyrA462* *endA1*  $\Delta$  (*sr1-recA*) *mcrB* *mrr* *hsdS20*(<sub>rB-</sub>, <sub>mB-</sub>) *supE44* *ara-14* *galK2* *lacY1* *proA2* *rpsL20*(Sm<sup>R</sup>) *xyl-5*  $\lambda$ - *leu* *mtl1*

#### **-Les souches dérivées de BL21**

La souche BL21 d'*Escherichia coli* est une souche couramment utilisée afin de produire d'une protéine d'intérêt en cellules bactériennes. La principale caractéristique de cette souche est de contenir dans son génome le gène codant pour l'ARN polymérase T7 sous contrôle d'un promoteur inductible. Au cours de mon doctorat, j'ai utilisé différentes souches dérivées de cette dernière, dont la liste et les caractéristiques sont présentées dans la table n°8 (Partie 4 des résultats, paragraphe II).

#### **-La souche C58C1**

La souche d'agrobactérie (*Agrobacterium tumefaciens*) C58C1 est utilisée au laboratoire pour les études d'expressions transitoire en cellules de *Nicotiana Benthamiana* de



protéine dans les plantes et les transformations stables d'*Arabidopsis thaliana*. Elle peut contenir différents plasmides « helper » tels que pMP90, pSoup ou le plasmide d'« hypervirulence » pCH32. Le génotype de cette souche est : pTi<sup>-</sup>, Ery<sup>R</sup>, Chl<sup>R</sup>.

### **I.3 Conditions de croissance et milieux de culture**

#### *I.3.1 Culture des plantes Arabidopsis thaliana et Nicotiana benthamiana sur terreau*

Les plantes d'*Arabidopsis thaliana* sont cultivées sur terreau en atmosphère contrôlée, 16h de lumière à 9600 Lux à 22°C, 8 h d'obscurité à 16°C, hygrométrie 65-75%. Les graines sont réhydratées dans de l'eau et placées 3 jours à 4°C avant semis, pour lever la dormance.

Pour les sélections de plantes d'*Arabidopsis* transformées et résistantes à l'herbicide « basta », les graines sont mises à germer sur du sable humide, et arrosées à plusieurs reprises avec une solution d'herbicide (240-250 mg/L de basta, additionnée de Silwet-77(500µL/L).

Les plants de *Nicotiana benthamiana* sont cultivés sur terreau en atmosphère contrôlée, 16h de lumière à 9600 Lux à 26°, 8 h d'obscurité à 16°C, hygrométrie 65-75%.

#### *I.3.2 Culture des plantes Arabidopsis thaliana in vitro*

Les plantes d'*Arabidopsis thaliana* sont cultivées stérilement *in vitro* sur milieu Murashige et Skoog, nommé ½ MS (2,2 g/L mix Murashige et Skoog (Sigma #M5519), 1% sucrose, 0,5 g/L MES, pH ajusté à 5,7 avec du KOH, 8 g/L Agar type E). Les graines sont stérilisées avant réhydratation selon un protocole classique basé sur l'utilisation d'éthanol (3 min dans un bain d'ETOH 70%+0.05% SDS, 1 min dans un bain d'ETOH 95%, séchage ~1 h00 sous hotte stérile). Les conditions de culture de la chambre sont 16h de lumière à 9600 Lux à 22°C, 8 h d'obscurité à 16°C, hygrométrie 65-75%.

Selon les besoins expérimentaux le milieu ½MS peut être additionné de dexaméthasone (concentration finale : 10 µM) ou de différents antibiotiques tels que la kanamycine (50mg/L) et l'hygromycine (25mg/L).

#### *I.3.3 Culture des souches de bactéries*

Les souches bactériennes d'*E. coli* sont cultivées à 37°C, sur milieu Luria-Bertani, nommé LB (10g/L Bacto-tryptone, 5g/L Bacto-yeast extract, 10g/L NaCl, pH ajusté à 7 avec du NaOH, contenant 15g/L de Bacto-Agar pour les cultures en boîte de Pétri), et contenant les antibiotiques nécessaires à la sélection des plasmides utilisés.

Les bactéries de la souche d'*E. coli* BL21-SI sont cultivées sur milieu Luria-Bertani dépourvu de sel nommé LB-ON (10g/L Bacto-tryptone, 5g/L Bacto-yeast extract, pH ajusté à 7 avec du NaOH, contenant 15g/L de Bacto-Agar pour les cultures en boîte de Pétri), et contenant les antibiotiques nécessaires à la sélection des plasmides utilisés. La température utilisée pour leur croissance varie de 28°C à 37°C selon les besoins expérimentaux (voir aussi paragraphe III.5).

Les agrobactéries sont cultivées à 30°C, sur milieu Luria-Bertani appauvri en sel, nommé LB-Agro (10g/L Bacto-tryptone, 5g/L Bacto-yeast extract, 5g/L NaCl, pH ajusté à 7 avec du NaOH, contenant 15g/L de Bacto-Agar pour les cultures en boîte de Pétri), et contenant les antibiotiques nécessaires à la sélection des plasmides utilisés.

Les antibiotiques couramment utilisés sont : carbénicilline (50mg/L) ; chloramphénicol (20mg/L) ; kanamycine (50mg/L) ; gentamicine (15mg/L) ; tétracycline (5mg/L) ; rifampicine (37,5mg/L).

#### **I.4 Autre matériel**

Annexe N°1 : Liste des principaux plasmides et vecteurs d'expression utilisés

Annexe N°2 : Liste des principales constructions réalisées au cours de mon doctorat

Annexe N°3 : Liste des principaux oligonucléotides utilisés

## **II Méthodes de biologie moléculaire**

### **II.1 Méthodes relatives à l'ADN**

#### *II.1.1 Extraction d'ADN génomique « rapide » d'Arabidopsis*

Ce protocole a été utilisé pour préparer rapidement de l'ADN génomique d'Arabidopsis à partir d'un échantillon végétal de quelques milligrammes (jeune feuille, bouton ou fleur). Ce protocole peut être réalisé en tube ou en plaque 96 puits.

#### Protocole d'extraction ADN génomique « rapide » d'Arabidopsis

1. Collecter et broyer les tissus dans l'azote liquide
2. Ajouter 400µl de Tampon de lyse
3. Vortexer
4. Incuber 15 minutes à 50°C
5. Centrifuger 15 minutes à 1500 G et 4°C
6. Récupérer 150 µl de surnageant propre
7. Ajouter 150µl d'isopropanol
8. Vortexer
9. Incuber 20 minutes à température ambiante
10. Centrifuger 30 minutes à 1500 G à 4°C
11. Eliminer le surnageant, rincer les culots avec 125µl d'éthanol à 70%
12. Centrifuger 15 minutes à 1500 G à 4°C
13. Eliminer le surnageant et sécher les culots
14. Reprendre les culots dans 125µl de TE

Tampon de Lyse (concentration finale) : SDS 0,5 % ; NaCl 250 M ; EDTA 25 mM; Tris-HCl (pH :7,5) 200 mM

La qualité de l'extraction est généralement suffisante pour les PCR de génotypage. Toutefois, une étape de purification supplémentaire de l'ADNg ainsi extrait peut être réalisée par précipitation au phénol/chloroforme.

### Protocole de purification de l'ADNg par précipitation au phénol/chloroforme

1. Ajouter un volume d'un mélange phénol/chloroforme (1 : 1)
2. Vortexer
3. Centrifuger 10 minutes à vitesse maximale à température ambiante
4. Reprendre la phase supérieure
5. Ajouter un volume chloroforme
6. Vortexer la phase supérieure
7. Centrifuger 10 minutes à vitesse maximale à température ambiante
8. Reprendre la phase supérieure
9. Ajouter 1/10ème de volume de NaAc (5M) et 2 volumes  $\mu$ L ETOH 95°C froid
10. Centrifuger 20 minutes à vitesse maximale à 4°C
11. Eliminer le surnageant, rincer les culots avec 125 $\mu$ l d'éthanol à 70%
12. Centrifuger 15 minutes à vitesse maximale à 4°C
13. Eliminer le surnageant et sécher les culots
14. Reprendre les culots dans 125 $\mu$ l de TE

#### II.1.2 Echantillons d'ADN génomique provenant d'autres espèces

Au cours de l'étude de la distribution phylogénétique du domaine DYW, l'ADNg de nombreuses espèces a été utilisé.

##### Espèces « témoins »

- ADNg d'*Arabidopsis thaliana*, extrait à partir de culture de cellules, fourni par Claire Lurin, URGV, Evry
- ADNg de pois, colza, tomate, melon, fourni par Bénédicte Sturbois, URGV, Evry
- ADNg de *Mus musculus*, fourni par Théodora Nialt, Université de Vienne, Autriche
- ADNg de *Chlamydomonas reinhardtii* fourni par le laboratoire du Pr F.A. Wollman, Université Pierre et Marie Curie, Paris

##### Espèces d' « intérêt »

- ADNg de *Prorocentrum minimum* et de *Pfiesteria piscicidia* fourni par le laboratoire du Pr Senjie Lin, Université de Connecticut, Canada.
- ADNg total de *Cryptocodinium cohnii* fourni par laboratoire du Pr Dave Spencer, Université de Dalhousie, Canada
- ADNg de *Karenia brevis*, extrait à partir de cellules de la souche CCMP2228 commercialisée par le laboratoire Provasoli-Guillard National Center for Culture of Marine Phytoplankton- Floride-USA, en utilisant le Genomic DNA from plant Kit- Nucleospin® Plant (Macherey-Nagel)
- ADNg d'*Alexandrium tamarense*, *Amphidinium carterae*, *Ceratium horridum*, *Pyrocystis lunula*, *Guillardia theta* fourni par le laboratoire du Pr U.G. Maier, Université de Marburg, Allemagne.
- ADNg de toutes les autres espèces d'intérêt, fourni par le laboratoire du Pr Volker Knoop, Université de Bonn, Allemagne.

### *II.1.3 Amplification d'ADN par PCR*

#### II.1.3.1 Méthode de PCR « classique »

Les réactions de PCR de type « classique » sont utilisées dans le cadre des réactions de génotypage, contrôle de la qualité des extractions d'ADNg, test des colonies, RT-PCR...

Les réactions sont réalisées en utilisant une TaqPolymérase préparée au laboratoire d'Evry ou de Perth, dans un tampon adapté (Composition du tampon 10X : Tris-HCl 100mM pH8, MgCl<sub>2</sub> 25mM, KCl 500mM, Nonidet P40 0.5%).

Les réactions PCR ont été réalisées à l'aide d'un programme dont la température d'hybridation est adaptée aux différents couples d'oligonucléotides, et le temps d'élongation qui dépend de la taille du fragment à amplifier.

#### Protocole de PCR « classique »

Dans volume final de 10 à 50 µl  
Matrice ADN 25-75ng  
Rouge de crésol 1X  
Tampon 1X  
dNTP 0,2 mM chacun  
Taq polymérase 1 à 5 unités  
Oligonucléotides 1µM

#### Programmation du thermocycler

5 min à 94°C  
30 sec à 94°C  
45 sec à X°C (généralement entre 45°C et 60°C)  
X min à 72°C (généralement entre 30sec et 1min 30)  
5 min à 72°C

} 35 cycles

Rouge de crésol (solution 10 X : Rouge de crésol 2 mg/ml ; Saccharose 600 mg/ml)

#### II.1.3.2 Méthode de PCR « clonage » ou « séquençage »

Lorsque nous amplifions un fragment d'ADN par PCR afin de le cloner ou le séquencer, les réactions de PCR sont réalisées comme décrit précédemment (paragraphe II.1.3.1) à la différence près que la TaqPolymérase utilisée est une enzyme dite « fidèle » (avec une activité de relecture).

Différentes enzymes ont été utilisées telles que : la platinum®Pfx (Invitrogen) au laboratoire d'Evry, et la DNA polymerase (Takara) au laboratoire de Perth (voir aussi PCR de clonage Gateway paragraphe II.1.6.3)

#### II.1.3.3 Méthode de PCR utilisant des amorces dégénérées

Pour les PCR utilisant des amorces dégénérées, nous avons utilisé des conditions de PCR adaptées, en particulier un programme utilisant un gradient de palier de température d'hybridation descendant pour les 20 premiers cycles. La Taq polymérase utilisée est la Taq DNA polymérase commercialisée par new England BioLabs, utilisée dans un tampon adapté

(Tampon 10X : 200 mM Tris-HCl, 100 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100 mM KCl, 20 mM MgSO<sub>4</sub>, 1% Triton X-100, pH 8.8).

Afin de faciliter la réaction PCR, 5µg d'ADNg total est digéré soit par l'enzyme EcoRV, soit par l'enzyme SmaI, indépendamment, puis un mélange équimolaire est réalisé.

Protocole de PCR utilisant les amorces dégénérées :

Dans volume final de 10 à 50 µl  
Matrice ADN 10-75ng d'ADN génomique  
Tampon 1X  
dNTP 0,2 mM chacun  
Taq polymérase 0.5 unité  
Oligonucléotides 1mM chacun  
MgCl<sub>2</sub> 2mM

Programmation du thermocycler

5 min à 94°C

94°C (30 sec)  
45°C (-0.5°C/cycle) (1min) } 20 cycles  
72°C (2min)

94°C (30 sec) } 30 cycles  
40°C (1min)  
72°C (2min)

72°C (7min)

II.1.3.4 Méthode de PCR quantitative

Les réactions de PCR quantitatives ont été réalisées au laboratoire de Perth, dans la machine LightCycler 480 Real-time PCR system (Roche) à l'aide du master mix LightCycler 480 SYBR green 1 (Roche). Les oligonucléotides utilisés figurent en Annexe n°3.

II.1.4 Visualisation des produits PCR par électrophorèse

Les molécules d'ADN sont séparées sur gel d'agarose (concentration 0,5 à 4%) contenant du BET (Bromure d'Ethidium) dans un tampon de migration (TBE 1X : pH8,0, 89 mM TRIS, 89 mM Acide borique, 2 mM EDTA) par application d'un champ électrique d'une tension variant entre 50 à 200 V suivant la taille du gel (7 à 30 cm de longueur). La visualisation s'effectue sous UV.

### *II.1.5 Purification de produits PCR en vue d'un clonage ou d'un séquençage*

#### *II.1.5.1 Utilisation de Microcon®*

Les produits de PCR destinés à être clonés ont été concentrés et dessalés sur membrane à très faible adsorption en cellulose régénérée (Microcon® YM-50, Millipore). Les produits de PCR sont déposés sur la membrane puis centrifugés à 14000G, la membrane est ensuite transférée et disposée à l'envers dans un nouveau tube après ajout de 10µl d'eau ultrapure, les produits de PCR purifiés sont récupérés par centrifugation à 14000G.

#### *II.1.5.2 Utilisation du kit QIAquick (Qiagen)*

Alternativement, les produits de PCR, destinés à être clonés ou séquencés, peuvent être directement extraits du gel d'électrophorèse purifiés et concentrés en utilisant le kit QIAquick et en suivant les indications du fournisseur.

#### *II.1.5.3 Purification au PEG*

Enfin, les produits PCR destinés à être clonés peuvent être précipités, purifiés et concentrés, par ajout de 5 volumes d'une solution de PEG/MgCl<sub>2</sub>/TE (solution PEG/MgCl<sub>2</sub>: 12mM MgCl<sub>2</sub>, 12% PEG 8000, 6mM Tris, 0,6mM EDTA pH 7,5). Après centrifugation pendant 1 heure à 4500xg à température ambiante, le culot est repris dans 20µL de TE (10mM Tris, 1mM EDTA pH 7,5).

### *II.1.6 Clonage de produit PCR*

#### *II.1.6.1 Clonage « classique » (restriction, ligation)*

Les produits PCR sont amplifiés à l'aide d'oligonucléotides portant des sites de restriction à leurs extrémités 5' (comme décrit au paragraphe II.1.3.1). Après digestion du fragment PCR et du vecteur dans lequel l'insert doit être inséré, le vecteur est éventuellement déphosphorylé (cas d'une digestion simple) (utilisation de la BAP : Bacterial Alkaline Phosphatase, Fermentas, selon les instructions du fournisseur). Le vecteur et le fragment PCR ainsi préparés sont purifiés sur gel (comme décrit au paragraphe II.1.5.2), et quantifiés. La réaction de ligation est conduite généralement sur la nuit à 4°C, en utilisant la T4 DNA ligase (New England Biolabs). Le rapport entre la quantité d'ADN correspondant à l'insert et au plasmide a été ajusté en fonction des expérimentations. Après ligation, une transformation de bactéries compétentes est réalisée (comme décrit au paragraphe II.1.7).

#### *II.1.6.2 Clonage dans le vecteur pGEMT-Easy (Promega)*

Les produits PCR purifiés peuvent être directement clonés dans le vecteur pGEMT-Easy en suivant les instructions du fournisseur (Promega), avant transformation de bactéries compétentes (comme décrit au paragraphe II.1.7).

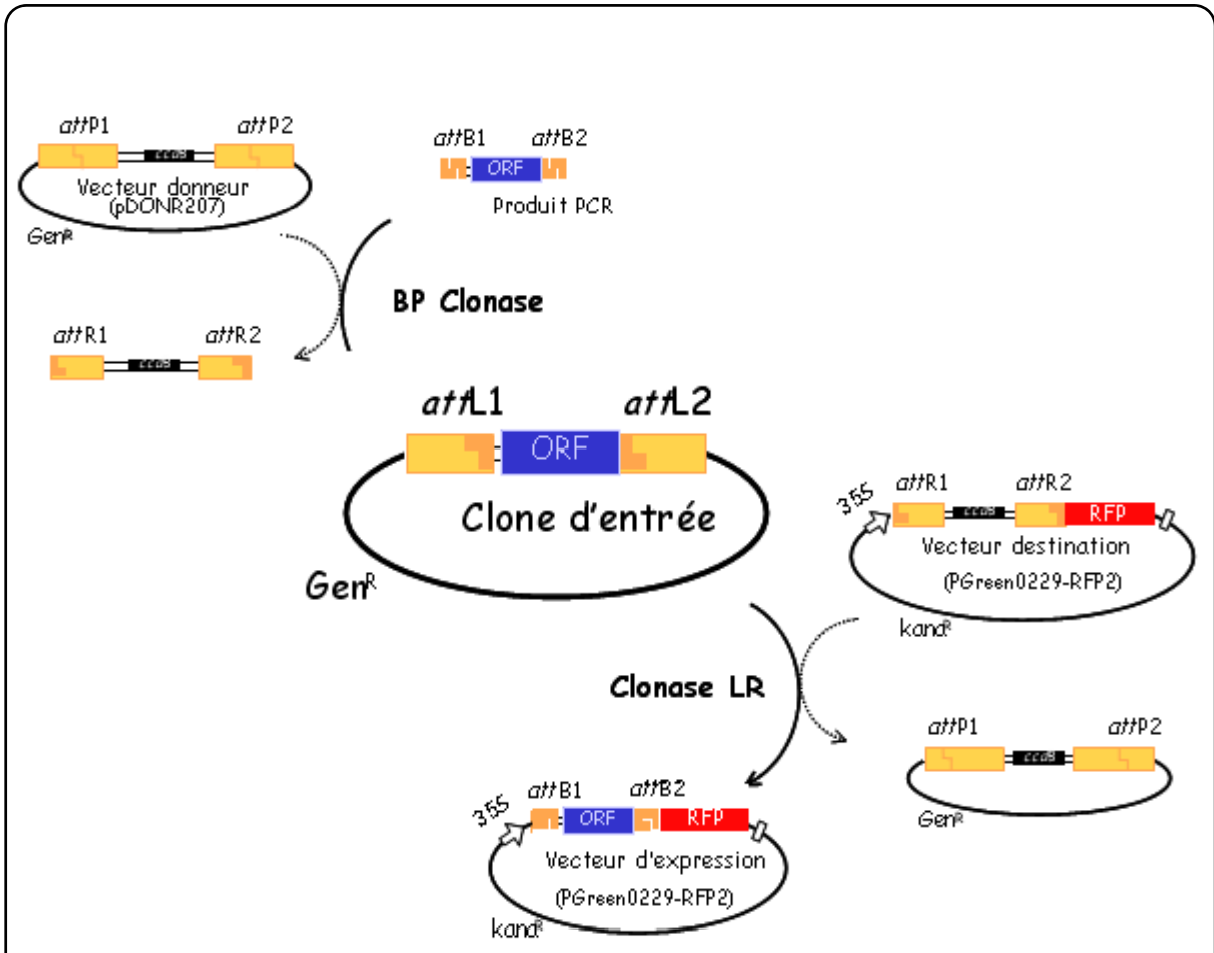


Figure n°98 :système de clonage par recombinaison Gateway® (Invitrogen)

### II.1.6.3 Technologie de clonage Gateway<sup>TM</sup> (Invitrogen)

#### Principe du clonage Gateway

Au cours de mon doctorat, l'essentiel des constructions ont été établies en utilisant la technologie Gateway<sup>TM</sup> (Invitrogen). Son principe repose sur les réactions de recombinaison permettant l'intégration et l'excision du phage lambda dans le génome d'*E. coli*. Les clonages sont réalisés en deux étapes. La première (réaction BP) consiste en la recombinaison directionnelle d'un gène d'intérêt avec un vecteur dit « donneur », dont le produit donne le vecteur dit « d'entrée ». Ce dernier sert de matrice à une seconde réaction (réaction LR) de recombinaison avec un vecteur de destination, aboutissant à un vecteur dit « d'expression » (figure n°98).

L'ensemble des vecteurs donneurs et de destination utilisés au cours de mon doctorat figure dans la table en Annexe n°1.

#### PCR de clonage

A l'aide deux PCR successives, le fragment d'ADN à insérer est amplifié et les sites de recombinaisons *attB1* et *attB2* sont ajoutés. Les produits PCR obtenus sont purifiés par précipitation au PEG (comme décrit au paragraphe II.1.5.3).



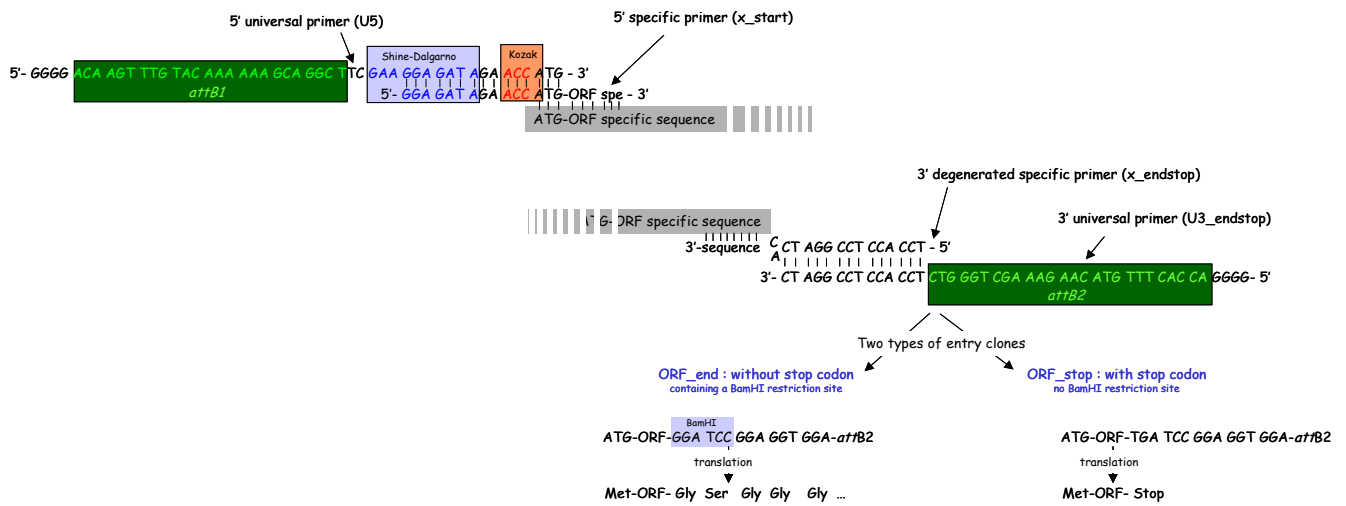


Figure n°99 : Structure des régions 5' et 3' entourant le gène d'intérêt dans le cadre d'une amplification avec les amorces U5-5'PTO/Amorce U3end-stop-5'PTO)

Protocole de PCR1 en vue d'un clonage Gateway

1  $\mu$ l de chaque primer (0.5  $\mu$ M)  
1  $\mu$ l d'AND matrice (~50 ng/ $\mu$ l)  
0,2  $\mu$ l de dnTP (100 nM)  
0,2  $\mu$ l de MgSO<sub>4</sub> (50 nM)  
1  $\mu$ l de tampon platinum®Pfx (Invitrogen)  
0,08  $\mu$ l de platinum®Pfx (Invitrogen)  
4,52  $\mu$ l H<sub>2</sub>O

Programmation du thermocycler

5 min à 94°C  
15 sec à 94°C  
30 sec à 55°C  
x sec à 68°C  
5 min à 68°C

} 10 cycles

Protocole de PCR2 en vue d'un clonage Gateway

1  $\mu$ l de amorce U5\*(10 $\mu$ M)  
1 $\mu$ l de amorce U3\*(10 $\mu$ M)  
0,8  $\mu$ l de dnTP (10mM)  
0,8  $\mu$ l de MgSO<sub>4</sub> (50mM)  
1 $\mu$ l de Tampon platinum®Pfx  
0,32  $\mu$ l de platinum®Pfx  
32,08  $\mu$ l d'H<sub>2</sub>O

Programmation du thermocycler

5 min à 94°C  
15 sec à 94°C  
30 sec à 55°C  
x sec à 68°C  
5 min à 68°C

} 25 cycles

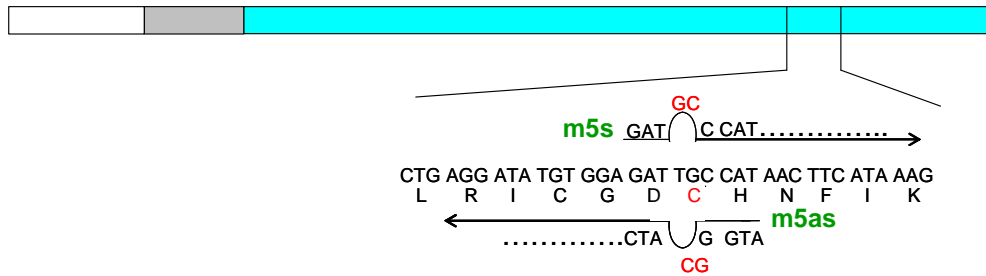
\*Note relative aux amorces U3 et U5

Il existe plusieurs versions des amorces U3 et U5 (les séquences sont disponibles en Annexe n°3). En effet originellement, dans le cadre du projet ATOME, développé au laboratoire d'Evry, les amorces U5 et U3 ont été développées.

Puis ces amorces ont été modifiées de sorte qu'elles portent une liaison « pto » (pour phosphothioate) entre les deux nucléotides de l'extrémité 5' des amorces, ce qui rend les fragments générés au cours d'une PCR insensibles à l'activité exonucléase 3'→5' de la Pfx.

De même, l'amorce U3end/stop porte une base dégénérée permettant de générer, en une même réaction de PCR et de BP clonase, des vecteurs d'entrée contenant une ORF pleine longueur soit avec un codon STOP, soit sans le codon STOP. En l'absence de ce codon stop, un site BamHI est présent, ainsi par simple digestion par l'enzyme de restriction BamHI, nous distinguons les clones avec ou sans codon stop (figure n°99).

## Génération des amorces sens et antisens



## PCR successives en vue d'un clonage par réaction de recombinaison BP

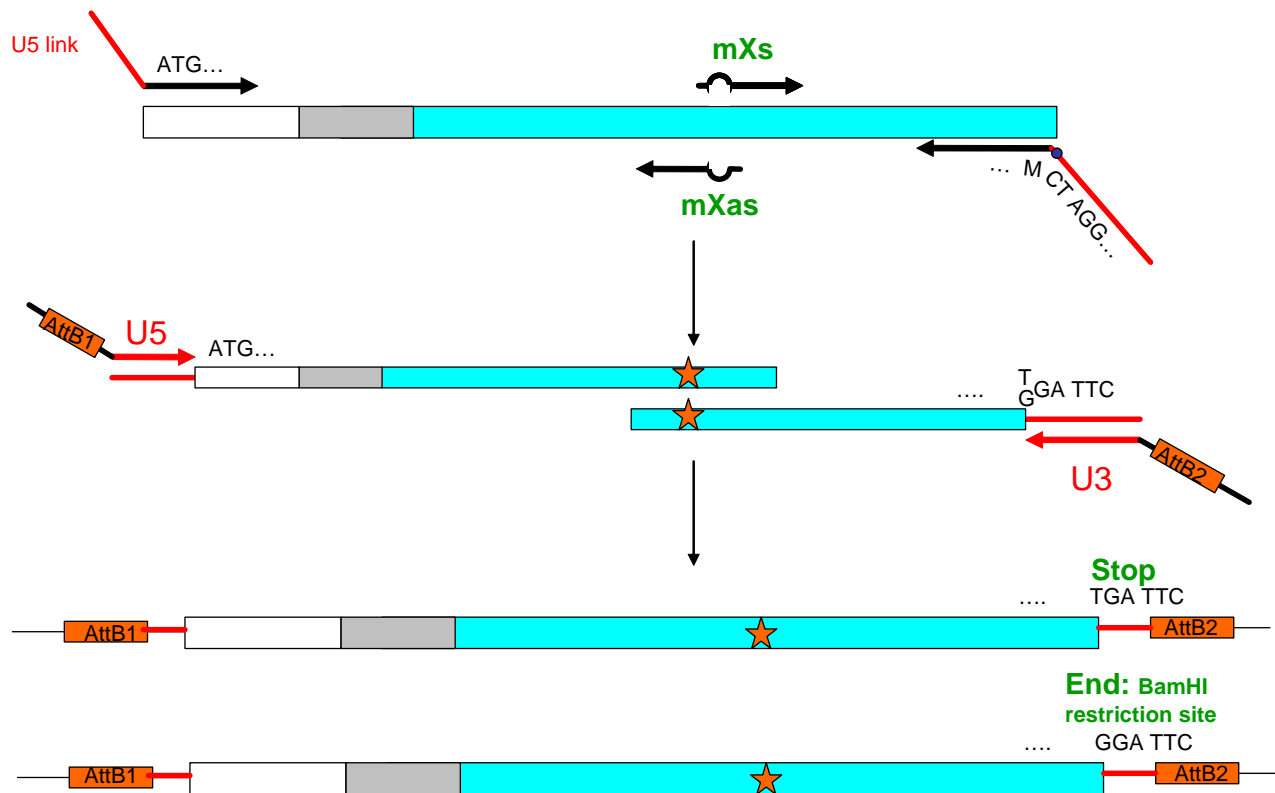


Figure n°100:

Mutagenèse ponctuelle par PCR en vue d'un clonage dans un vecteur d'entrée du système Gateway

Enfin l'amorce U5 contient une séquence ATG et Shine Dalgarno Kozak. Ces séquences pouvant être présentes dans certains vecteurs d'expression, une version modifiée de l'amorce U5, ne contenant pas ces séquences, a été générée.

#### Mutagenèse ponctuelle par PCR en vue d'un clonage dans un vecteur d'entrée du système Gateway

Au cours de mon doctorat, j'ai voulu surexprimer le gène *AtDYWI* porteur ou non de mutations ponctuelles dans *Arabidopsis* (Partie 3 des résultats, paragraphe IV.3.1). Ainsi une mutagenèse ponctuelle a été réalisée par PCR, ainsi que l'ajout des sites de recombinaison Gateway.

Pour chaque mutation, deux amorces chevauchant le site ou les sites à muter ont été générées : l'une est appelée sens, et l'autre antisens (figure n°100). La dégénérescence du code génétique a permis de dessiner des amorces ne contenant qu'une à deux substitutions par site muté. Ces substitutions ont été positionnées en 5' des amorces afin de laisser un minimum de 10 bases d'homologie parfaite en 3'.

Tout d'abord, on réalise, en parallèle, une PCR avec l'amorce antisens et l'amorce DYWIATG2- attB1 et une PCR avec l'amorce sens et l'amorce DYWIend/stop. Pour ces PCR, une TAQ polymérase sans activité de relecture est utilisée afin que les substitutions portées par les amorces ne soient pas corrigées au cours de la réaction PCR, et le nombre de cycles est réduit pour éviter d'accumuler d'autres mutations non souhaitées. Puis, les deux produits d'amplifications obtenus sont mélangés de façon équimolaire et servent de matrice dans une deuxième réaction PCR réalisée avec les amorces 5'PtoU5/ 5'PtoU3 (comme décrit précédemment).

#### Réaction de BP clonase

La réaction de BP permet de cloner la séquence encadrée par les sites *attB1* et *attB2* dans un vecteur d'entrée contenant les sites *attL1* et *attL2*

##### Protocole de réaction de BP clonase

1. Réaliser le mélange réactionnel suivant :
  - 2µL produit PCR purifié
  - 50-100ng vecteur pDONR207
  - 1µL tampon de BP 5X
  - 0,4µL Clonase BP (Invitrogen)
  - Compléter à 5µL avec du TE pH8
2. incubation sur la nuit à 25°C
3. ajout de protéinase K (2 µg)
4. incubation 15 minutes à 37°C
5. transformation de bactéries compétentes (paragraphe II.1.7)
6. test de colonies par PCR (paragraphe II.1.3.1)
7. minipréparation d'ADN (paragraphe II.1.9)

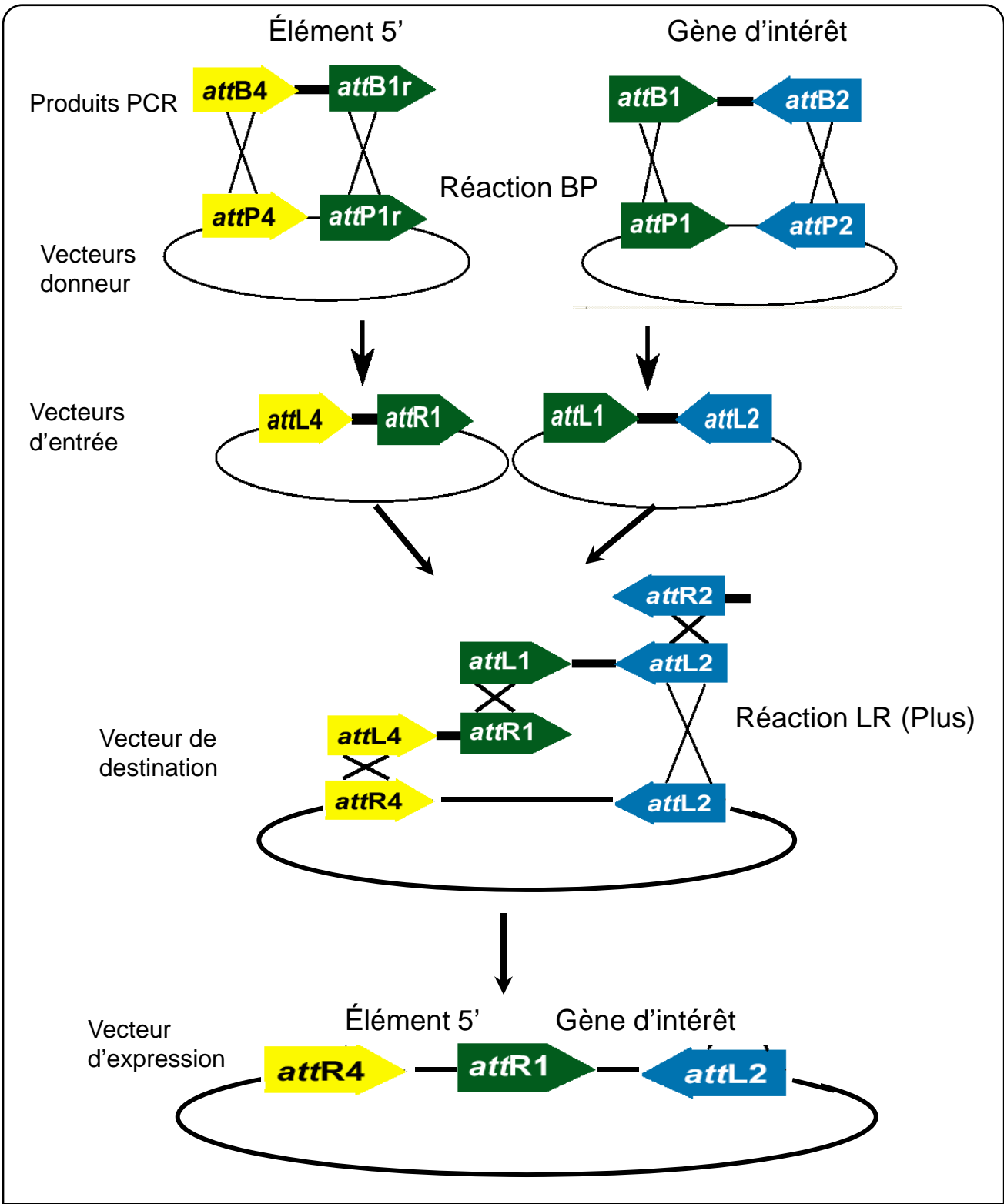


Figure n°101 : Système de clonage multisites par recombinaison Gateway® (Invitrogen)

Ici, deux vecteurs donneurs sont recombines (réaction de LR plus clonase) afin de permettre la fusion d'un élément 5' (un promoteur par exemple) en amont du gène d'intérêt.

### Réaction de LR clonase

La réaction de LR clonase va permettre de transférer les inserts clonés dans les vecteurs d' « entrée » dans un vecteur destination entre les sites de recombinaison *attR1* et *attR2*

#### Protocole de réaction de LR clonase

1. Réaliser le mélange réactionnel suivant :
  - 50-100ng de clone d'entrée
  - 100ng de vecteur destination
  - 1µL du tampon LR 5X
  - 0,4µL Clonase LR (Invitrogen)
  - Compléter à 5µL final avec du TE pH8
2. incubation sur la nuit à 25°C
3. ajout de protéinase K (2 µg)
4. incubation 15 minutes à 37°C
5. transformation de bactéries compétentes (paragraphe II.1.7)
6. test de colonies par PCR (paragraphe II.1.3.1)

### Cas du clonage Gateway multisite

Une évolution de la technologie Gateway permet également de cloner jusqu'à trois fragments PCR en phase dans un vecteur d'expression. On parle alors de clonage Gateway multisite. Au cours de mon doctorat, j'ai utilisé cette technologie afin de réaliser des fusions entre le promoteur *AmasPmas* et le gène *AtDYWI* porteur ou non de mutations ponctuelles (Partie 3 des résultats, paragraphe IV.3.1).

La démarche expérimentale reste très similaire à celle décrite précédemment, si ce n'est que l'on utilise des sites de recombinaison « modifiés » et une réaction de LR Plus Clonase (figure n°101).

#### Protocole de réaction de LR clonase Plus

1. Réaliser le mélange réactionnel suivant :
  - clones d'entrées (20-25 fmol chacun)
  - 60ng de vecteur destination
  - 4 µL Clonase LR Plus (Invitrogen)
  - Compléter à 16 µL final avec du TE pH8
2. incubation sur la nuit à 25°C
3. ajout de protéinase K (2 µg)
4. incubation 15 minutes à 37°C
5. transformation de bactéries compétentes (paragraphe II.1.7)
6. test de colonies par PCR (paragraphe II.1.3.1)

### II.1.7 Transformation bactérienne, sélection et test des colonies transformantes

Des cellules bactériennes de la souche DH5 $\alpha$  compétentes (de l'ordre de  $10^7$ - $10^8$  transformants par  $\mu\text{g}$  d'ADN) pour la transformation par choc sont utilisées.

#### Protocole de préparation de cellules DH5 $\alpha$ compétentes

1. Préculture de 10mL dans du LB sur la nuit à 37°C, 200rpm
2. Ensemencement de la préculture dans 500mL de milieu SOB
3. Culture à 19°C, 200rpm jusqu'à l'obtention d'une DO à 600nm comprise entre 0,5 et 0,6
4. Incubation 10 minutes sur la glace
5. Centrifugation 5000G, 10 minutes à 4°C (en 2 flasques de 1L)
6. Chaque culot est repris dans 80mL de TB froid
7. Incubation 10 minutes sur la glace
8. Centrifugation 5000G, 10 minutes à 4°C
9. Chaque culot est repris dans 20mL de TB froid
10. Ajout de 2,8mL de DMSO
11. Incubation 10 minutes sur la glace
12. Préparation d' « aliquote » (généralement 2mL)
13. Congélation dans l'azote liquide et conservation à -80°C

Milieu SOB : 20g/L Bacto-tryptone, 5g/L Bacto-yeast extract, 10mM NaCl, 2,5mM KCl, 10mM MgCl<sub>2</sub>, 10mM MgSO<sub>4</sub>, pH ajusté à 7 avec du NaOH  
TB : 10mM Pipes, 15mM CaCl<sub>2</sub>, 250mM KCl, ajusté à pH 6,7 avec du KOH, 55mM MnCl<sub>2</sub>, stérilisé par filtration (pas d'autoclave)

#### Protocole de transformation par choc thermique de cellules DH5 $\alpha$ compétentes

1. Mélanger 1 $\mu\text{g}$  de plasmide (ou le mélange réactionnel de réaction BP ou LR) à 100 $\mu\text{l}$  de bactéries compétentes
2. incubation 30 min à 4°C
3. incubation 75 secondes à 42°C (en tube eppendorf)
4. incubation 3 min à 4°C
5. ajout de 900 $\mu\text{L}$  de milieu SOC
6. Culture liquide, 1h à 37°C
7. Etalement sur boite de milieu LB additionné de l'antibiotique approprié
8. test de colonies par PCR (paragraphe II.1.3.1)

Milieu SOC : SOB + 20mM glucose

### II.1.8 Transformation d'Agrobactéries, sélection et test des colonies transformantes

Des cellules bactériennes de la souche C58C1 compétentes (de l'ordre de  $10^2$ - $10^3$  transformants par  $\mu\text{g}$  d'ADN) pour la transformation par choc sont utilisées.

Protocole de préparation de cellules C58C1 compétentes

1. Préculture de 1,6mL dans du LB-Agro sur la nuit à 28°C, 200rpm
2. Ensemencement de la préculture dans 40mL de LB-Agro
3. Culture à 29°C, 200rpm jusqu'à l'obtention d'une DO à 600nm comprise entre 0,4 et 0,8
4. Centrifugation 5000G, 15 minutes à 4°C
5. Le culot est repris dans 10mL 10 mM Tris-HCl pH 7,5 froid
6. Centrifugation 5000G, 15 minutes à 4°C
7. Le culot est repris dans 10mL 10 mM Tris-HCl pH 7,5 froid
8. Préparation d'« aliquote » (généralement 1mL)
9. Congélation dans l'azote liquide et conservation à -80°C

Protocole de transformation par choc thermique de cellules C58C1 compétentes

1. Mélanger 1µg de plasmide à 200µl de bactéries compétentes
2. incubation 5 min à 0°C
3. incubation 5 min dans l'azote liquide
4. incubation 5 min à 37°C
5. ajout de 800µL de milieu LBAgro
6. Culture liquide, 2h à 28°C
7. Etalement sur boîte de milieu LBAgro additionné de l'antibiotique approprié
8. test de colonies par PCR (paragraphe II.1.3.1)

*II.1.9 Extraction et purification de plasmide (minipréparation d'ADN)*

Selon les laboratoires, différents kits commerciaux ont été utilisés tels que :

- le kit SV Miniprep (Promega) au laboratoire d'Evry
- les kits : QIAprep Spin Miniprep Kit (Qiagen) et PureLink HiPure Plasmid Miniprep Kit (Invitrogen) au laboratoire de Perth.

La quantité de plasmide obtenue est systématiquement vérifiée par spectrométrie (mesure de la DO à 680nm). On obtient généralement des rendements allant de 50ng-1µg/mL pour 5mL de cultures.

*II.1.10 Séquençage d'ADN*

Le séquençage des produits de PCR ou des ARN plasmidique a été effectué par une entreprise externe, GenomeExpress puis Macrogen à partir de janvier 2006.



## **II-2 Méthodes relatives à l'ARN**

### *II.2.1 Extraction d'ARN total de tissus végétaux*

L'ARN total est extrait à partir d'un échantillon végétal de quelques milligrammes (une feuille ou une plantule pour *Arabidopsis*, un fragment de feuille pour *Nicotiana benthamiana*) suivant une méthode dite « au TRIzol® » au laboratoire d'Evry, ou par l'utilisation du kit RNeasy plant minikit (Qiagen) en suivant les instructions du fournisseur.

#### Protocole d'extraction d'ARN au « TRIzol® »

1. Récolter les tissus, les peser et les plonger dans l'azote liquide
2. Réduire les tissus en fine poudre au mortier et dans l'azote liquide
3. Reprendre dans 1ml de TRIzol® par 100mg de tissus
4. Incuber 5 minutes à température ambiante
5. Ajouter 0,2 volume de chloroforme et vortexer
8. Incuber 5 minutes à température ambiante
9. Centrifuger 15 minutes à un maximum de 12000 G à température ambiante
10. Récupérer la phase aqueuse
11. Ajouter un volume d'isopropanol (0,5 volume de TRIzol® initial)
12. Incuber 10 à 30 minutes à température ambiante
13. Centrifuger 15 minutes à maximum 12000 G à 4°C
14. Récupérer le culot, puis le laver dans l'éthanol 70%, eau DEPC 30%
15. Centrifuger 15 minutes à maximum 12000 G
16. Eliminer l'éthanol
17. Sécher le culot
18. Reprendre le culot dans de l'eau DEPC à 65°C, généralement 1/10ème à 1/20ème du volume de TRIzol® utilisé
19. Ajouter 0,1 volume de phénol/chloroforme/IAA (25 :24 :1) et vortexer
20. Incuber 5 minutes à température ambiante
21. Centrifuger 15 minutes à un maximum de 12000 G à température ambiante
22. Récupérer la phase aqueuse
23. Ajout 0,1 volume NaAC (3M) traité au DEPC et 2 volumes ETOH 95% eau DEPC 5%
24. Placer 30 min (minimum) à -80°C
25. Centrifuger 15 minutes à maximum 12000 G à 4°C
26. Récupérer le culot, puis le laver dans l'éthanol 70%, eau DEPC 30%
27. Centrifuger 15 minutes à maximum 12000 G
28. Eliminer l'éthanol
29. Sécher le culot
30. Reprendre le culot dans de l'eau DEPC à 65°C, généralement 1/10ème à 1/20ème du volume de TRIzol® utilisé

### *II.2.2 Elimination de l'ADN contaminant des extraits d'ARN totaux*

Les ARN totaux sont traités à la DNaseI (DNA-free™ DNase Treatment and Removal Reagents AM1906) suivant les instructions du fournisseur (Ambion®). Les ARN totaux sont quantifiés au spectrophotomètre par mesure de l'absorption à 260nm, et visualisés sur gel d'agarose 2 %, tampon TBE.

### II.2.3 Réaction de transcription « inverse »

La synthèse du premier brin d'ADN complémentaire d'ARN d'organites se fait à l'aide d'hexanucléotides aléatoires. En revanche, la synthèse du premier brin d'ADN complémentaire d'ARN nucléaires se fait à l'aide d'un oligonucléotide dT. La synthèse d'ADN complémentaire est effectuée en utilisant la reverse transcriptase SuperscriptII™ (11904-018, Invitrogen™).

#### Protocole de transcription « inverse »

1. Réaliser le mélange suivant
  - 3 µg d'ARN
  - 1 µl de dNTP (10mM)
  - 100ng d'hexanucléotides ou 2 µL oligonucléotide dT (10 µM)
  - Eau DEPC, qsp 12 µl
2. incuber 5 minutes à 65°C
3. incuber 5 à 10 minutes dans la glace
4. Ajouter
  - 4 µl de tampon de réaction 5X
  - 2 µl de DTT (0,1 M)
  - 1 µl de RNaseOUT™ (10777-019 Invitrogen™)
  - 1 µl de SuperscriptII™ (200 unités) (sauf dans le contrôle RT-)
5. Incubation 10 minutes à 25°C (dans le cas de l'utilisation d'hexanucléotides)  
Incubation 10 minutes à 37°C (dans le cas de l'utilisation oligonucléotide dT)
6. Incuber 60 minutes à 42°C
7. Incuber 15 minutes à 75°C
8. Incuber dans la glace

## **II.3 Mesure du niveau d'édition**

### II.3.1 Crible HRM (pour High Resolution Melting)

Ce crible non quantitatif, développé au laboratoire de Perth par Anne-Laure Chateigner-Boutin, permet de détecter éventuellement un défaut d'édition au niveau d'un des 34 sites chloroplastiques décrits chez Arabidopsis (Chateigner-Boutin and Small, 2007). Son principe est décrit dans la troisième partie des résultats, paragraphe III.3.7 et figure n°67).

### Protocole pour l'analyse HRM

1. Matrices utilisées : 1ng d'ADNg d'Arabidopsis (préparé comme décrit au paragraphe II.1.1), 3 $\mu$ L d'ADNc d'une plante sauvage ou mutante dilué 1/35 (préparé comme décrit au paragraphe II.2.3), ou un mélange de ces deux matrices.
2. La réaction de PCR s'effectue dans la machine LightCycler 480 Real-time PCR system (Roche) dans un volume de 10 $\mu$ l contenant : le master mix LightCycler 480 SYBR green 1 (Roche), l'agent intercalant : le LCGreen Plus (Idaho Technology) et 200nM de chaque amorce.
3. Programmation du LightCycler :
  - 10 min à 95°C
  - |               |   |           |
|---------------|---|-----------|
| 20 sec à 95°C | } | 40 cycles |
| 20 sec à 60°C |   |           |
| 20 sec à 72°C |   |           |
  - 30sec à 95°C
  - 30sec à 70°C
  - Fusion de 70°C à 90°C (acquisition de 30 signaux/degré)
4. Analyse par utilisation du logiciel du LightCycler 480 permettant de déterminer le « Crossing point » (CP), et la cinétique de la température de fusion «  $T_m$  » (représentant la fluorescence émise entre 450-500nm, en fonction de l'augmentation de température). Le « pic » de fusion correspondant à la dérivée négative ( $-d/dT$ ) du signal de fluorescent et la valeur de  $T_m$  sont déterminés.
5. Analyse par utilisation du logiciel Gene Scanning Software des courbes de « fusion » ainsi obtenues.

### II.3.2 Analyse par PPE (*pour Poisoned Primer Extension*)

Cette technique permet de mesurer l'efficacité de l'édition de l'ARN au niveau d'un site particulier dans un transcrit. Son principe est décrit dans la troisième partie des résultats, paragraphe III.3.7).

### Protocole pour l'analyse PPE

1. Les ARN d'une plante sauvage et mutante sont extraits (comme décrit paragraphe II.2) et desADNc sont synthétisés (comme décrit paragraphe II.2.1). Alternativement, pour les témoins, 25pg d'ADN plasmidique (contenant la séquence du transcrit d'intérêt édité ou non édité) servent de matrice.
2. Une PCR est effectuée en utilisant un couple d'amorces flanquant le site d'intérêt dans le transcrit.
3. 2,5µl de PCR sont purifiés par 1µl d'enzyme ExoSAP-IT (USB 78200) et incubés à 37°C pendant 20 minutes pour éliminer les nucléotides et les primers non incorporés. L'enzyme est inactivée à 80°C pendant 15 minutes.
4. La réaction de PPE s'effectue par le biais d'un kit ThermoSequenase Cycle Sequencing kit (USB 78500).

Mélange réactionnel pour l'extension PPE :

100ng de produit PCR purifié  
2µl de tampon 10X  
0,5pmol d'amorce marquée en 5' à la 6-carboxyfluorescein (FAM)  
1 unité ThermoSequenase DNA polymerase  
100nm de chaque nucléotides (mélange contenant trois dNTP+ un ddNTP)  
qsp 20µl d'eau

#### Programmation du thermocycler

1 min à 94°C  
30 sec à 94°C  
30 sec à 55°C  
10 sec à 72°C  
3 min à 68°C

} 35 cycles

5. Migration sur gel de séquençage à 12% acrylamide (Sequagel) à 55W.
6. Révélation par utilisation d'un Typhoon Trio Imager (GE Healthcare, ex-Amersham), avec un filtre d'émission 520BP40. L'acquisition du signal fluorescent se fait à 488nm.
7. Quantification par utilisation du logiciel ImageQuant TL 1D gel analysis (Amersham Bioscience). Les bandes correspondant aux produits édités et non édités sont manuellement déterminées, et le logiciel calcule un volume correspondant à l'intensité de fluorescence dans cette bande. Le bruit de fond est déterminé par la mesure d'une bande « absente » correspondant au produit édité et non édité, obtenue à partir des plasmides contrôle non édité et édité, respectivement. L'efficacité (ou %) d'édition est alors calculée comme suit :

$\%E = 100 \times (\text{volume de la bande édité} - \text{volume du bruit de fond de la bande édité}) / (\text{volume de la bande édité} - \text{le volume du bruit de fond de la bande édité}) + (\text{volume de la bande non édité} - \text{le volume du bruit de fond de la bande non édité})$ .

### **III Méthodes relatives aux protéines**

#### **III.1 Extraction de protéines de tissus végétaux**

Les extractions de protéines totales sont réalisées à partir d'un échantillon végétal de quelques milligrammes (une feuille ou une plantule pour *Arabidopsis*, un fragment de feuille pour *Nicotiana benthamiana*).

Protocole d'extraction de protéines totales

1. Collecter et broyer les tissus dans l'azote liquide
2. Ajouter 400µl de Tampon de lyse
3. Incuber 30 minutes (vortexer de temps en temps)
4. Centrifuger 15 minutes à 12000 G
5. Récupérer le surnageant

Tampon de Lyse (concentration finale) :  $\beta$ -mercaptoéthanol 40mM; EDTA 5 mM; Tris-HCl (pH :7,5) 100 mM

Ajouter « fraîchement » une pastille d'inhibiteur de protéase (Roche)/ 10mL de Tampon avant utilisation

Au cours de mon doctorat, j'ai également utilisé des fractions enrichies en protéines mitochondriales et chloroplastiques solubles d'*Arabidopsis*, préparées par Cathie Small-Colas des Francs et Anne-Laure Chateigner-Boutin (Plant Energy Biology, Perth).

#### **III.2 Dosage de protéines**

La quantité de protéines extraites est mesurée par Test de Bradford (Bradford 1976), en utilisant le réactif Coomassie Plus (Pierce Biotechnology™).

#### **III.3 Migration des protéines sur gel dénaturant**

Les protéines sont séparées sur gel 7cm, de type SDS-PAGE 12,5 % :

Composition d'un gel SDS-PAGE

	Gel de migration (Running)	Gel de concentration (Stacking)
40% Acrylamide	3,3ml	335µl
2% bis- Acrylamide	1,2mL	177µl
Tris-HCl	2,5mL (1,5M ; pH 8,8)	850µl (0,5M ; pH 6,8)
SDS 10%	100µL	66,7µL
Eau	2,5mL	2mL
Au dernier moment :		
APS 10% (frais)	50µl	16.6µl
TEMED	10µl	6,75µl

### Préparation des échantillons

Les échantillons sont mélangés à un tampon de dénaturation/charge (tampon concentré 5X: 15%  $\beta$ -mercaptoéthanol, 15% SDS, 1,5% bleu de bromophénol et 50% glycérol) puis dénaturés à 95°C pendant 5 minutes.

### Migration des échantillons

La migration se fait à voltage constant à 80V dans le gel de concentration et 160V dans le gel de migration. Le tampon de migration est un tampon Laemmli 1X (Tampon 10X : Glycine 1,92M, Tris base 0,25M, pH : 8,3) additionné de 0,1% SDS.

### Visualisation des protéines par coloration directe de gel dénaturant

Les gels sont directement colorés en utilisant le réactif Coomassie Brilliant Blue R-250 (Invitrogen) en suivant les instructions du fournisseur.

## **III.4 Expériences de western blot**

### Transfert sur membrane

Après séparation sur gel SDS-PAGE, les protéines sont transférées sur une membrane PVDF (Immun-Blot PVDF Membrane BioRad) par électro-transfert avec le système Mini Trans-Blot Cell (Bio-Rad), dans un tampon (Glycine 0,1 M, Tris base 768 mM, 20% ETOH). Le transfert peut s'effectuer durant 1h à voltage constant (100V) ou sur la nuit à ampérage constant (100mA).

### Saturation de la membrane et incubation avec les anticorps

1. Saturation de la membrane pendant 1h, à 37°C dans une solution de TBS-T contenant 5% de lait lyophilisé
2. 3 fois 5 à 10 min de lavage par incubation dans du TBS-T
3. Incubation de 2h dans l'anticorps primaire dilué dans du TBS-T (Généralement AC3 dirigé contre la protéine AtDYW1, dilution 1/500 ou un anticorps dirigé contre le tag 6-his, dilution 1/1000)
4. 3 fois 5 à 10 min de lavage par incubation dans du TBS-T
5. Incubation de 45min dans l'anticorps secondaire dilué dans du TBS-T (Généralement Anti-lapin dilution 1/10000 si l'anticorps primaire est AC3 dirigé contre la protéine AtDYW1 et Anti-souris dilution 1/10000 si l'anticorps primaire est dirigé contre le tag 6-his)
6. 3 fois 5 à 10 min de lavage par incubation dans du TBS-T

Tampon TBS-T : 1X TBS, 0,1% de tween 20

Tampon TBS 10X : 200mM Tris ; 1,37M NaCl ; pH :7,4

### Révélation de western blot

Le western blot est révélé en utilisant le Kit de réaction ECL Western Blotting Analysis System (GE Healthcare, ex-Amersham), et visualisé grâce à un LAS-1000 (Fujifilm).

### **III.5 Production de la protéine rDYW1 en système bactérien**

Au cours de mon doctorat, j'ai produit la protéine rDYW1 en système bactérien, à partir d'une version dite « interne-stop » de l'ORF AtDYW1 clonée, *via* le système Gateway, dans le vecteur d'expression pDEST17 (voir aussi paragraphe II.1.6.3 et partie 4 des résultats paragraphe II).

Cette construction a été introduite, *via* une transformation bactérienne par choc thermique (comme décrit au paragraphe II.1.7) dans différentes souches toutes dérivées de la souche BL21 (voir liste Partie 4 des résultats, table n°8). Ces souches d'expression contiennent dans leur génome le gène codant pour l'ARN polymérase T7 sous contrôle d'un promoteur inductible. (*proU* promoteur : promoteur inductible aux sels pour BL21-SI et *lac/tac* promoteur : promoteur inductible à l'IPTG pour les autres souches). Comme il n'y a pas de pression de sélection pour maintenir cette cassette dans le génome d'*E.coli*, elle peut, éventuellement, être perdue au fur et à mesure des divisions cellulaires rendant l'induction impossible. Afin de palier à ce problème, il est donc préférable de partir de cellules fraîchement transformées avant chaque induction.

Protocole d'induction « classique » des souches dérivées de la souche BL21:

**1. Préculture à 37° C, sur la nuit, 200 rpm**

La préculture est réalisée dans un milieu LB contenant 50mg/mL de carbenicilline pour toutes les souches, à l'exception de la souche BL21-SI pour laquelle, on utilise un milieu dépourvu de sels (LBONCb). La DO600 de cette préculture est mesurée au matin et une dilution à une DO600= 0.1 est réalisée pour démarrer la culture.

**2. Culture à 37°C, 200 rpm jusqu'à l'obtention d'une DO600 : 0.5**

Cette phase dure de 2 à 3h, selon la souche d'*E.coli* utilisée. Le milieu de culture choisi est identique à celui des précultures.

**3. Induction à 37°C, 200rpm, 3h**

L'induction se fait par ajout de sels NaCl (0,3M concentration finale) pour la souche BL21-SI, ou d'IPTG (1mM concentration finale) pour les autres souches dans le milieu de culture.

**4. Vérification de l'induction**

Un « aliquote » de culture avant et après induction (généralement 1mL) est prélevé, et centrifugé à 12000G pendant 10 min. Le culot obtenu est resuspendu dans 15uL de tampon de dénaturation/charge et chauffé 5minutes à 95°C (comme décrit au paragraphe III.4). Les protéines contenues dans ces « aliquotes » seront analysées sur un gel dénaturant colorées au bleu de Coomassie (comme décrit au paragraphe III.4).

Afin d'augmenter la quantité de protéine rDYW1 soluble produite, les conditions d'induction ont été optimisées pour la souche BL21-SI. La préculture et la culture sont réalisées à une température de 28°C. J'ai constaté que l'oxygénation des cultures affecte de beaucoup la quantité finale de protéines solubles obtenues. Ainsi durant ces expérimentations, j'ai utilisé un volume de culture liquide inférieur ou égale au cinquième du volume total du récipient. La phase d'induction dure de 10 à 12h. Le reste du protocole reste inchangé par rapport à ce qui est décrit dans le protocole d'induction « classique » des souches dérivées de la souche BL21.

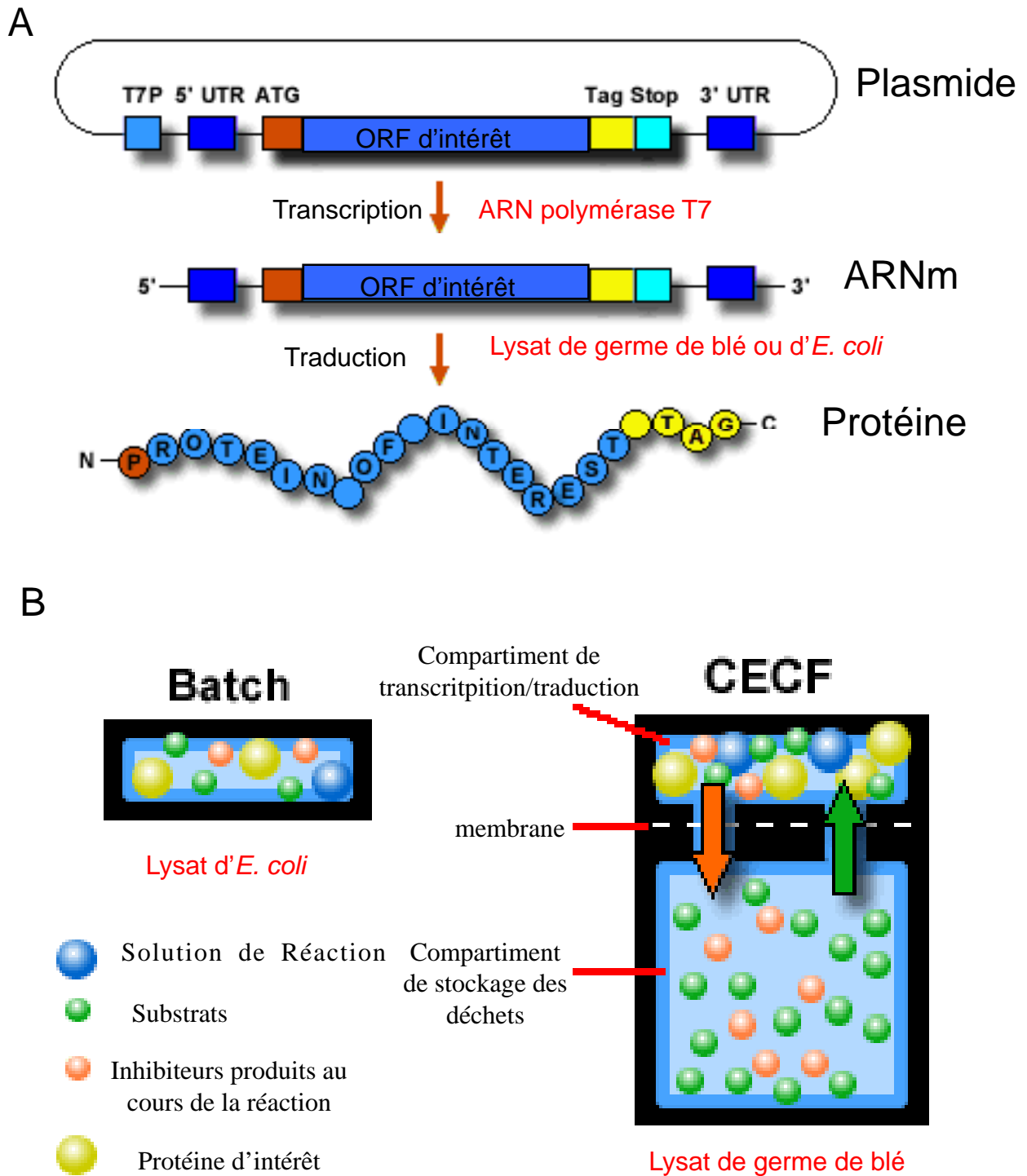


Figure n°102 : (A) Principe du système RTS (Rapid Translation System) de Roche, (B) utilisation d'un mélange réactionnel en Batch (cas du lysat d'*E. coli*) ou en utilisant la technologie CECF pour continuous-exchange cell-free (cas du lysat de germe de blé)



### Lyse des cellules en conditions natives

Différents tampons de lyse en conditions natives ont été utilisés au cours de mon doctorat (voir partie 4 des résultats paragraphe II.4 et table n°9). La lyse des cellules est effectuée pendant 30 minutes, à température ambiante, avec une agitation douce. Elle est généralement combinée à une étape de sonication (2X5min à puissance intermédiaire).

### Lyse des cellules en conditions dénaturantes

En conditions dénaturantes, les cellules sont lysées dans le tampon suivant : 50mM Tris-HCl (pH8,5), 8M urée, 10mM imidazole, 5mM  $\beta$ -mercaptoéthanol. L'urée étant très efficace pour la lyse, une étape de sonication n'est pas nécessaire.

## **III.6 Production de la protéine rDYW1 par le système RTS (Rapid Translation System, Roche)**

La production « acellulaire » de la protéine rDYW1 a été réalisée au cours de mon doctorat à partir de 3 $\mu$ g de plasmide pDEST17 contenant version dite « interne-stop » de l'ORF AtDYW1.

Le principe de la technologie RTS est présenté sur la figure n°102a. Il s'agit d'ajouter la matrice dans un mélange réactionnel contenant un lysat « amélioré » apportant l'ARN T7 polymérase, les ribosomes, et les composants additionnels nécessaires aux réactions de transcription et de traduction *in vitro*. Deux kits différents, commercialisés par Roche, ont été utilisés : RTS wheat germ lysat et le RTS *E. coli* lysat en suivant les instructions du fournisseur. Le premier utilise la technologie CECF (pour continuous-exchange cell-free) qui permet l'ajout en continu de nucléotides et acides aminés au cours de la réaction (figure n°102b). Par opposition, dans le kit *E. coli* lysat, les substrats sont apportés en une fois directement (méthode dite « Batch ») (figure n°102b).

## **III.7 Purification de la protéine rDYW1**

### *III.7.1 Par le système Ni-NTA, pour Nickel-Nitrilotriacetic acid (Qiagen)*

La protéine rDYW1 porte un tag N-terminal 6-his et de ce fait peut être purifiée sur des colonnes Ni-NTA commercialisées par Qiagen. Nous avons utilisé le kit Qiaexpress® Ni-NTA fast start (3060) en suivant les instructions du fournisseur. Alternativement, pour purifier des petites quantités de protéines des billes magnétiques Ni-NTA (Ni-NTA Magnetic Agarose, Qiagen) ont été utilisées.

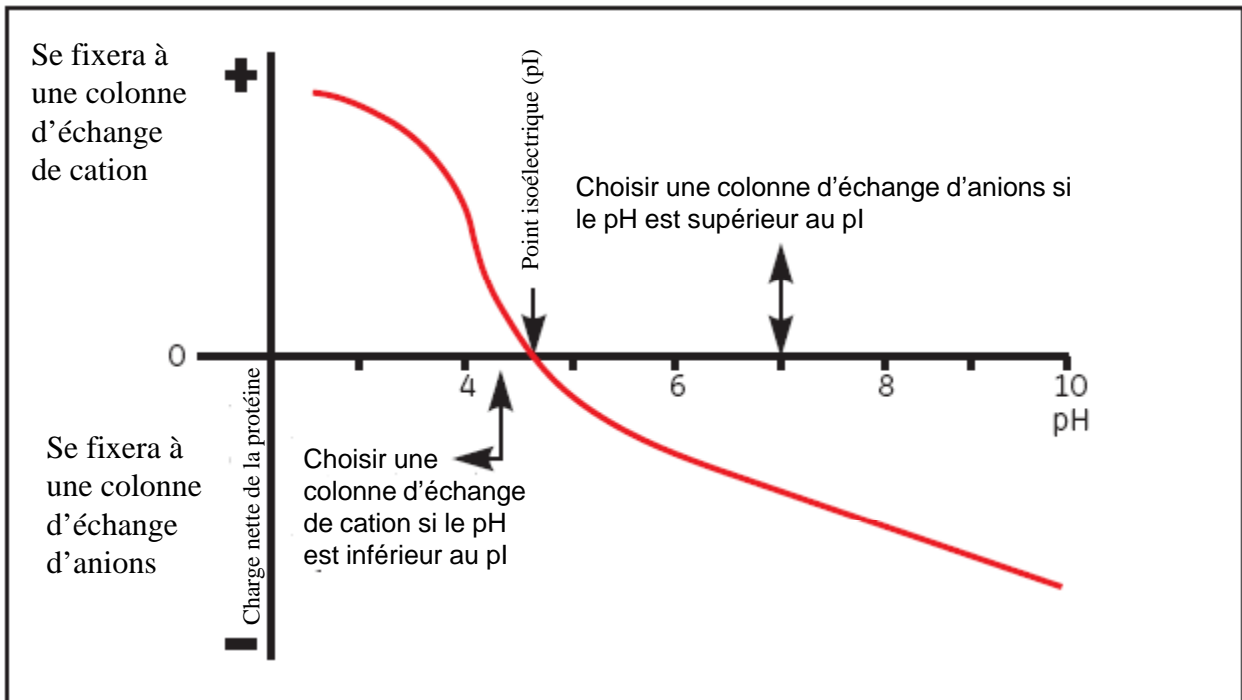


Figure n°103 : Recommandations de GE Healthcare pour déterminer le type de colonnes échangeuses d'ions à utiliser pour fixer une protéine d'intérêt en fonction de son pI.

#### Protocole d'utilisation des billes magnétites Ni-NTA

- 1- Resuspendre les billes dans leur milieu de stockage par pipetage ou vortex léger.
- 2- Prélever un volume de 50 $\mu$ L de billes, les placer dans un tube eppendorff 1,5mL.
- 3- Culotter les billes à l'aide du support aimanté. Enlever par pipetage le surnageant.
- 4- Laver les billes dans 400 $\mu$ L de tampon de lyse. Bien agiter (pipetage ou vortex léger)
- 5- Culotter les billes à l'aide du support aimanté. Enlever par pipetage le surnageant.
- 6- Ajouter 1mL de cellules fraîchement lysées. Bien agiter par retournement du tube (4 à 5 fois).
- 7- Culotter les billes à l'aide du support aimanté. Enlever par pipetage le surnageant.
- 8- Laver les billes dans 400 $\mu$ L de tampon de lyse. Bien agiter par retournement du tube (4 à 5 fois).
- 9- Culotter les billes à l'aide du support aimanté. Enlever par pipetage le surnageant.
- 10- Eluer à l'aide de 25 $\mu$ L de solution tampon de lyse additionné d'imidazole (2M). Bien agiter par retournement du tube (4 à 5 fois).
- 11- Culotter les billes à l'aide du support aimanté. Prélever le surnageant et le conserver à 4°C jusqu'à sa prochaine utilisation.

#### III.7.2 Chromatographie d'échange d'ions

Au cours de mon doctorat, la protéine rDYW1 a été purifiée sur des colonnes d'échange de cations HiTrap SP HP (GE Healthcare), ou d'anions HiTrap Q HP (GE Healthcare). La charge nette d'une protéine varie en fonction du pH et de la concentration saline de la solution dans laquelle elle se trouve. Ainsi, connaissant le pI (point isoélectrique) théorique de la protéine rDYW1, nous avons déterminé, en suivant les instructions du fournisseur, les conditions optimales pour purifier cette protéine sur ces deux types de colonnes (figure n°103). Une élution fractionnée sur gradient de sels a été réalisée afin d'améliorer la purification.

#### Conditions utilisées pour la chromatographie d'échange d'ions

Sur colonnes d'échange de cations HiTrap SP HP

Tampon de resuspension de la protéine rDYW1 (~1mg): NaP (50mM), NaCL (100mM) à pH 7,2

Gradient d'élution : NaCL, 100mM à 1M

Sur colonnes d'échange d'anions HiTrap Q HP

Tampon de resuspension de la protéine rDYW1 (~1mg) : Tris (20mM), NaCL (50mM) à pH 8,5

Gradient d'élution : NaCL, 50mM à 1M

#### **III.8 Production d'anticorps polyclonaux**

La protéine rDYW1 a été produite en cellules BL21-SI selon un protocole d'induction « classique » (comme décrit au paragraphe III.5). Les cellules ont été lysées en conditions dénaturantes (comme décrit au paragraphe III.5) et la protéine rDYW1 a été purifiée sur colonne Ni-NTA (comme décrit au paragraphe III.7.1). Une dialyse est réalisée afin de reprendre la protéine purifiée dans une solution de 50mM potassium phosphate (pH8), 5mM  $\beta$ -mercaptoéthanol. Cette solution contenant (~1mg/mL) de protéine rDYW1 est envoyée à une société afin de réaliser les injections aux lapins.

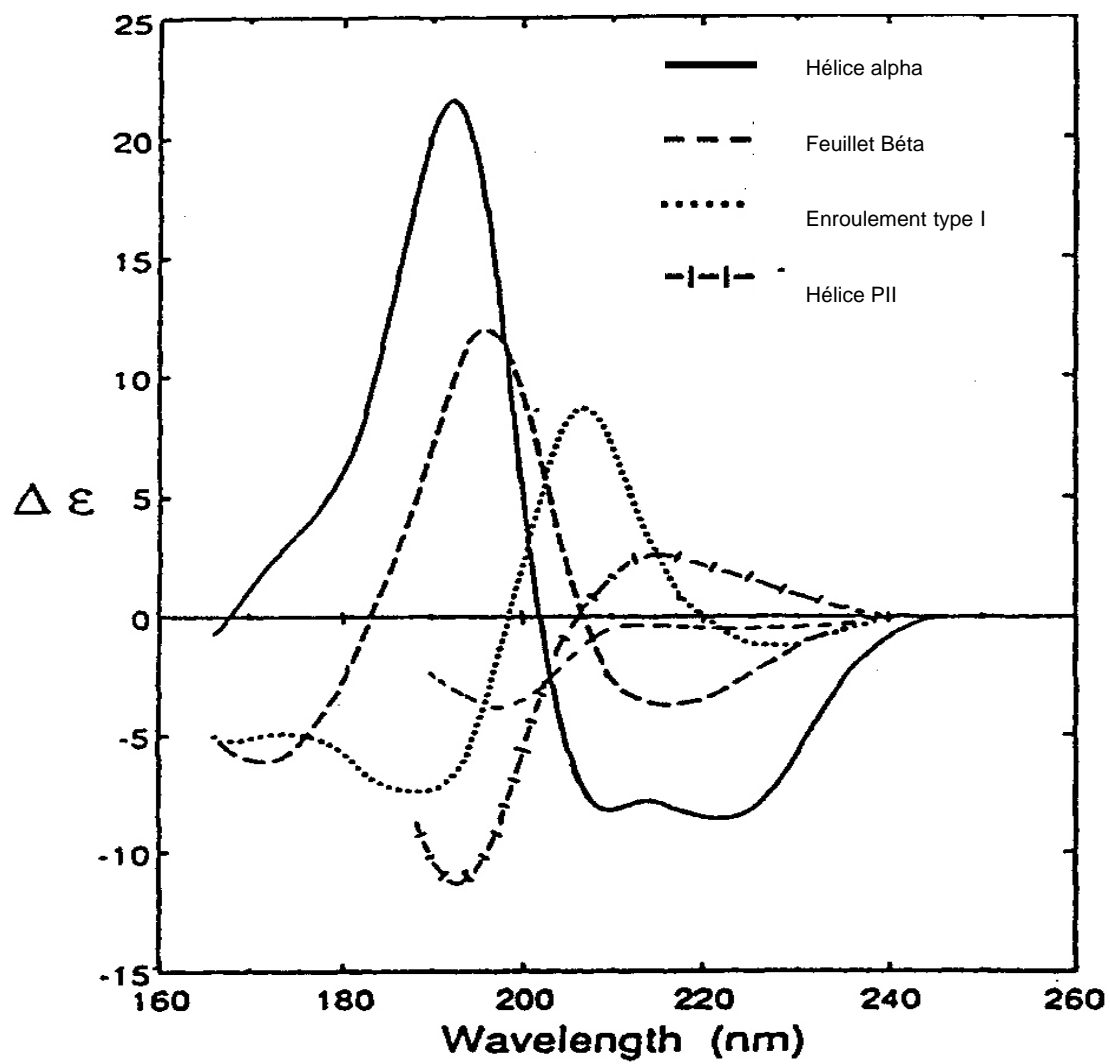


Figure n°104 : Spectres d'absorption obtenus, par spectrométrie à dichroïsme circulaire, pour différentes structures secondaires

La société Eurogentec (Parc industriel des Hauts-Sauts, 86 rue de l'Abbaye, 13-4040 Herstal, Belgique) a réalisé les injections et saignements des animaux afin de produire les anticorps AC1 et AC2.

La société Polyclonal Antibody Production (IMVS, Veterinary Services, 101 Blacks Rd, Gilles Plains, S.A 5086) a réalisé les injections et saignements des animaux afin de produire l'anticorps AC3.

### **III.9 Essais de renaturation de la protéine rDYW1 par dichroïsme circulaire**

Ces tests ont été réalisés en collaboration avec le laboratoire du Pr Charles Bond (Université de UWA, Perth, Australie) et avec l'aide du Dr William Stanley.

La protéine rDYW1 a été produite en cellules BL21-SI selon un protocole d'induction « classique » (comme décrit au paragraphe III.5). Les cellules ont été lysées en conditions dénaturantes (comme décrit au paragraphe III.5) et la protéine rDYW1 a été purifiée sur colonne Ni-NTA (comme décrit au paragraphe III.7.1). 0,1mg de protéine a été repris dans différentes solutions, comme décrit dans la partie 4 des résultats, paragraphe II.6 et table n°10.

Un spectre d'absorption par dichroïsme circulaire a été obtenu entre ~200 et 250 nm. Un logiciel d'exploitation (Jascow32) permet, par comparaison du spectre d'absorption de structures connues (figure n°104), d'estimer mathématiquement la structure secondaire de la protéine rDYW1 dans ces différentes solutions.

### **III.10 Mesure de l'activité cytidine-désaminase par spectrophotométrie**

Les mesures d'activité enzymatique de la cytidine-désaminase sont basées sur la méthode spectrophotométrique décrite par Wang (Wang, 1955). Le principe repose sur une absorption différentielle entre la cytidine (ou 2'déoxycytidine) et l'uridine à 282 nm. L'absorbance due à la cytidine (ou 2'déoxycytidine) diminue à fur et à mesure de sa conversion en uridine.

Les essais sont réalisés dans un milieu réactionnel de 1mL contenant du Tris-HCl 50 mM pH 7,5, NaCl 50mM et de la cytidine à 0,2 mM ou de la 2'déoxycytidine à 0,2 mM. Les essais sont réalisés à une température de 25°C, et les cinétiques sont développées sur 2 heures. La réaction débute avec l'addition de quantités variables d'extrait protéique.

### **III.11 Test d'édition sur sondes d'ARN radiomarquées**

Une activité possible d'édition des ARN d'une protéine est testée selon un protocole décrit par Yu et Schuster (Yu and Schuster, 1995).

## Protocole pour le test d'édition sur sondes d'ARN radiomarquées

### 1/ Préparation d'ARNt

Les ARNt sont mis en solution (Tris : 20mM pH 7,5, SDS : 0,5% et 50mM NaCl), à une concentration de 20µg/µL. Ils sont incubés en présence d'une unité de protéinase K pendant 1h, à 50°C. Les ARNt sont ensuite extraits au phénol/chloroforme (1 : 1), et précipités à l'éthanol (comme décrit au paragraphe II.1.1). La concentration finale des ARNt est alors mesurée, et ajustée à 10µg/µL.

### 2/ Marquage des ARN sondes

Les sondes d'ARN sont synthétisées par une réaction de transcription *in vitro* à l'aide du kit Maxiscript (Ambion) en présence d'une Cytidine 5'-triphosphate (alpha-32P) (Perkin Elmer, numéro de référence : BLU008X250UC).

### 3/ Réaction d'édition :

En théorie, les ARN ainsi marqués peuvent être dosés à l'aide d'un compteur à scintillation et environ 200 000cpm sont utilisés pour la réaction d'édition. Dans notre cas, faute d'équipement nécessaire, aucun dosage n'a été réalisé. Deux concentrations arbitraires d'ARN sonde ont été utilisées pour la réaction d'édition (sonde non diluée et sonde diluée 5X).

La réaction se fait en incubant, 2h30 à 30°C, l'ARN sonde avec 1 à 2µg de protéine purifiée (ou de mélange protéique), dans une solution finale :

- 30 mM HEPES-KOH pH7,7
- 3 mM acétate magnésium
- 45 mM acétate potassium
- 30 mM acétate ammonium
- 5% glycérol
- 10 mM ATP
- 5 mM DTT
- 1% PEG6000
- 30U RNase inhibitor
- 1x protéinase inhibitor complete

### 4/ Extraction des ARN au phénol/Chloroforme

A la fin de la réaction d'édition, 1 à 2 µg d'ARNt sont ajoutés au mélange réactionnel. Les ARN sont alors extraits au phénol/chloroforme, et précipités à l'éthanol (comme décrit au paragraphe II.1.1). Les nucléotides non incorporés sont nettoyés sur colonne Spin Columns (G50 Sephadex, Boehringer) en suivant les instructions du fournisseur.

### 5/ Digestion à la Nucléase P1 ( Nuclease 5'-Nucleotidehydrolase, 3'-Phosphohydrolase ; United States Biological)

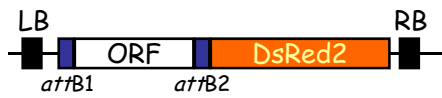
Le culot d'ARN est repris dans un volume de 10µL de concentration finale 30mM NaAc pH 5,2, et 0,5 U de nucléase P1 (enzyme qui coupe la liaison phospho-diester en 3' de chaque nucléotide). Ce mélange réactionnel est incubé 2h à 37°C.

### 6/ Séparation des produits de digestion par chromatographie sur couche mince de cellulose (TLC)

Les échantillons sont déposés à 1cm du bas d'une plaque TLC 20X20 cm, espacés de 1 à 2 cm. Le volume total de chaque échantillon est déposé par gouttes successives de 0,5µL que l'on prend soin de faire sécher entre chaque dépôt. La séparation se fait sur la nuit (12 à 15 heures), dans un mélange isopropanol :HCl :eau dans les proportions 70 :15 :15 (v :v :v), selon la méthode décrite par Nishimura (1972). La migration se fait dans une cuve que l'on peut fermer hermétiquement. La ligne de niveau de tampon dans la cuve doit être inférieure à la ligne de dépôt des échantillons (1cm).

### 7/ Révélation :

La plaque de cellulose est ensuite séchée avant révélation au phosphoimager BAS-2500 (Fujifilm).



Transformation d'*Agroacterium tumefaciens*

①

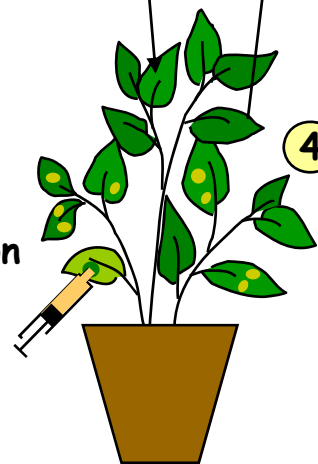
- préculture
- 24h culture dans 5 ml de milieu LB

②



- Centrifugation (5 min)
- Le culot est resuspendu dans une solution d'infiltration:
  - 10 mM MES pH5.6
  - 10 mM MgCl<sub>2</sub>
  - 150 μM acetosyringone
- DO<sub>600</sub> = 0.2-0.3
- 2h incubation at RT

③ Agrofiltration



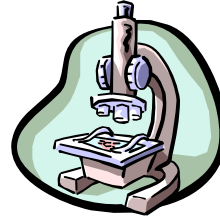
*Nicotiana benthamiana*

④

48-96h

⑥

Observation au microscope confocal



⑤

Digestion de la paroi cellulaire (2-6 h)  
dans une solution enzymatique 2X

Solution enzymatique (pour 100ml de stock 10X dans  
Protoplast medium )

- 0,2g Macerozyme R-10
- 0,5g Driselase
- 1,0g Cellulase "Onuzuka"

Protoplast medium (pour 1 litre)

- 4,3g/l Murashige and Skoog Plant Salt Mix (ICN Biomedicale)
- 2ml/l Vitamines "Morel and Wetmore (500X stock)
- 0,5g/l MES
- 20g sucrose
- 80g mannitol
- 3mg naphthaleneacetic acid
- 1mg 6-benzylaminopurine

Figure n°105 : Protocole d'agroinfiltration des feuilles de *Nicotiana benthamiana* et étude de la localisation subcellulaire de protéines fluorescentes exprimées transitoirement

## **IV Méthodes de biologie cellulaire**

### **IV.1 Expression transitoire de protéines en cellules de *Nicotiana benthamiana* par agroinfiltration**

Au cours de mon doctorat, l'expression transitoire de protéines d'intérêts en cellules de *N. benthamiana* a été réalisée pour des études de la localisation subcellulaire de la protéine AtDYW1 fusionnée à la RFP (Partie 2 des résultats, paragraphe IV.2) et dans le cadre de la reconstitution d'un « complexe minimal d'édition » (Partie 4 des résultats, paragraphe V.3).

Protocole d'agroinfiltration de feuilles de *N. benthamiana*  
(étape 1 à 4 sur la figure n°105)

1. Préculture de 5 mL de la souche d'agrobactéries d'intérêt sur la nuit, 28°C
2. Culture de 500mL de la souche d'agrobactéries d'intérêt sur la nuit, 28°C
3. Centrifuger 5 minutes à 4800G à température ambiante
4. Resuspendre le culot dans un milieu d'agroinfiltration (10mM MES/KOH pH5,6, 10mM MgCl<sub>2</sub>, 150µM d'acétylsyringone) de manière à obtenir une densité optique à 600nm comprise entre 0,2 et 0,3
5. Incubation à température ambiante pendant 2 heures
6. Infiltration à l'aide d'une seringue sans aiguille de la face inférieure d'une feuille de *N. benthamiana*

### **IV.2 Obtention de protoplastes et observations microscopiques**

Protocole d'obtention de protoplastes  
(3 à 5 jours après agroinfiltration, étape 5 sur la figure n°105)

1. Découper et lacérer les feuilles en fines lamelles
2. Incuber dans 5 mL de milieu de digestion 2 à 6 heures à 30°C à l'obscurité
3. Dépôt d'une goutte de solution entre lame et lamelle

Milieu de digestion : 4,3 g/L de Murashige and Skoog Plant Salt Mix (ICN Biomedicale), 0,5 g/L de MES, 20 g/L de sucrose, 80 g/L de mannitol, 0,4 g/L de Macerozyme R-10, 1 g/L de Driselase, 2 g/L de Cellulase "Onuzuka"

Les protoplastes ainsi obtenus sont observés à l'aide d'un microscope confocal (Leica TCS-NT) qui permet la détection de la fluorescence de la RFP (DsRed2 : Excitation/Emission maximum : 558 nm/583nm) (filtre BP580/80), de la GFP (Excitation /Emission maximum: 479nm/509nm) (filtre BP530/30), et ceux d'autofluorescence naturelle des cellules (Emission >600nm)( filtre LP590). Ces signaux sont collectés par des photomultiplicateurs indépendamment.



## **V Méthodes propres à Arabidopsis**

### **V.1 Transformation d'Arabidopsis par « floral deeping »**

Les plantes d'Arabidopsis sont transformées par trempage des boutons floraux dans une solution d'agrobactéries d'intérêt, selon la méthode décrite par Clough et Bent, (Clough and Bent, 1998).

Protocole de transformation d'Arabidopsis :

1. Culture de la souche d'agrobactéries d'intérêt dans du LB-Agro jusqu'à une DO à 600 nm comprise entre 0,8 et 1,0 (le volume dépend de la quantité de plantes que l'on souhaite transformer)
2. Centrifugation des cellules et obtention d'un culot
3. Resuspension du culot dans une solution 5% de sucrose, 0.02% de Silwet-L77 (volume identique au volume de culture)
4. Trempage des boutons floraux 2 à 5 min

### **V.2 Croisement d'Arabidopsis**

*Arabidopsis thaliana* est une plante autogame, ainsi pour croiser deux lignées entre elles, il faut manuellement « castrer » les plantes d'une lignée (notée A). Pour ce faire, 3 ou 4 boutons floraux, sont ouverts avec des pinces fines sous loupe binoculaire et débarrassés de leurs 6 étamines. Ainsi seuls les pistils (partie ♀) des plantes de la lignée A sont conservés et mis en contact avec du pollen (partie ♂) de l'autre lignée (notée B). Ce croisement est noté : [♀AX♂B]. De la même façon, le croisement réciproque peut être effectué [♀BX♂A].

### **V.3 Etude du phénotype**

L'analyse phénotypique des mutants consiste en l'étude de plusieurs caractères comme : le pourcentage de germination, la taille et l'aspect des germinations et des plantes adultes.

### **V.4 Coloration de grains de pollen, méthode d'Alexander (Alexander, 1969)**

Cette coloration histochimique permet de distinguer des grains de pollen avortés ou non avortés. En effet, cette méthode de coloration se base sur l'utilisation de deux composants essentiels : le malachite green (qui colore la paroi cellulosique du grain de pollen) et l'acide fuschique (qui colore le protoplasme des grains de pollen). Les grains de pollens non avortés prennent en quelques minutes une coloration bleue intense, détectable sous microscope à lumière blanche. En revanche, les grains de pollens avortés ne contiennent pas de protoplasme et ne se colorent pas.

#### Protocole de la coloration Alexander (1969)

1. Déposer une goutte de solution de coloration 1X sur une lame de microscope
2. Prélever une anthère et tapoter dans la goutte afin de libérer le pollen
3. Incuber 5 min à température ambiante
4. Placer une lamelle
5. Visualiser la coloration dans un microscope à lumière blanche

Solution Alexander 50X (pour 100mL : 10mL ETOH 95%, 5mL de 1% malachite vert dans de l'ETOH 95%, 5mL de 1% d'acide fuschique dans de l'eau, 0,5% d'orange G dans de l'eau, 2mL d'acide acétique glacial, 25mL de glycérol, 7mL de phénol , 47,5mL d'eau)

#### **V.5 Détection de l'activité GUS par coloration directe de plantules d'Arabidopsis**

L'expression du gène rapporteur *GUS* (codant la  $\beta$ -glucuronidase d'*E. coli*) peut être détectée par coloration de plantules d'*Arabidopsis* mises en contact direct avec le substrat X-Gluc (5-bromo-4-chloro-3-indolyl- $\beta$ -D-glucuronide).

#### Protocole de détection de l'activité GUS par coloration directe de plantules d'Arabidopsis

1. Incubation 20 à 30 minutes des plantules dans une solution d'acétone froid
2. Lavage dans le tampon de coloration (ne contenant pas de X-Gluc) sur glace
3. Transférer dans le tampon de coloration
4. Incubation à 37°C
5. Observation des plantules après 1 à 24h d'incubation

Tampon de coloration : 0.2% triton, 50mM NaPO<sub>4</sub>, 2mM Ferri-Ferro, 2mM X-Gluc (dans du DMSF)

#### **V.6 Détection de l'activité GUS par fluorimétrie**

Ce test fluorimétrique est basé sur l'hydrolyse, par la  $\beta$ -glucuronidase, du substrat 4-MUG (4-Methylumbelliferyl- $\beta$ -D-glucuronide) en un composé fluorescent à une longueur d'onde de 460nm. Une extraction de protéines totales à partir de plantules d'*Arabidopsis* est réalisée (comme décrit au paragraphe III.1), et l'extrait est dosé la méthode de Bradford (comme décrit au paragraphe III.2). La cinétique de l'activité de la  $\beta$ -glucuronidase est réalisée par mesure de la DO à 460nm toutes les 3 min pendant 2h, à l'aide de l'appareil Fluostar Galaxy (Biolise). L'activité enzymatique est mesurée en pmole de substrat hydrolysé par minute et par  $\mu$ g de protéines.

## **VI Outils bioinformatiques**

### \* Analyse de séquences peptidiques ou nucléotidiques

BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1997) j'utilise la version v2.2.9 disponible sur le site du NCBI (National Center for Biotechnology Information) : [www.ncbi.nlm.gov/BLAST/](http://www.ncbi.nlm.gov/BLAST/), avec les paramètres par défaut.

ClustalW (Thompson et al., 1994) j'utilise la version v2 disponible sur le site de l'EMBL-EBI : <http://www.ebi.ac.uk/Tools/clustalw2/index.html>, avec les paramètres par défaut.

Pour le dessin d'oligonucléotides, j'utilise Primer3 (v. 0.4.0) disponible à l'adresse suivante : <http://fokker.wi.mit.edu/primer3/input.htm>, avec les paramètres par défaut.

Pour les prédictions de motifs et domaines dans les séquences protéiques, j'utilise le logiciel ProSite (Hulo et al., 2006) disponible sur le site d'Expasy : <http://www.expasy.org/prosite/>

Pour la détermination des propriétés biochimiques d'une protéine, j'utilise le logiciel ProtParam (Gasteiger E., 2005) disponible sur le site d'Expasy : <http://www.expasy.org/tools/protparam.html>

Pour prédire la structure secondaire des protéines, j'utilise le logiciel PredictProtein (Rost and Liu, 2003) disponible à l'adresse suivante : <http://www.predictprotein.org/>

Les prédictions de structure des gènes d'Arabidopsis du logiciel Eugène (Foissac et al., 2003) sont examinés sur la base de données FLAGdb++ (Samson et al., 2004) disponible sur le réseau informatique de l'URGV.

La banque de fragments CATMA (Crowe et al., 2003) est également disponible sur la base de données FLAGdb++ (Samson et al., 2004) disponible sur le réseau informatique de l'URGV.

Les ADNc d'Arabidopsis séquencés par le Génoscope figurent également sur la base de données FLAGdb++ (Samson et al., 2004) disponible sur le réseau informatique de l'URGV.

La banque de données TAIR a été également utilisée, en particulier pour la recherche de mutants d'insertion par les collections: <http://www.arabidopsis.org/>

### \* Logiciels de prédiction de localisation cellulaire

Predotar (Small et al., 2004) : j'utilise la version 1.03 disponible à l'adresse suivante : <http://genoplante-info.infobiogen.fr/predotar>

TargetP (Emanuelsson et al., 2000): j'utilise la version v1.01 disponible à l'adresse suivante : <http://www.cbs.dtu.dk/services/targetP>

Psort (Nakai and Horton, 1999): j'utilise la version PSORT II disponible à l'adresse suivante : [www.http://psорт.nibb.ac.jp](http://www.psорт.nibb.ac.jp).

# **Annexes**

**Annexe N°1 : Liste des principaux plasmides et vecteurs d'expression utilisés**

nom	utilisation principale	origine
pDONR207	plasmide "donneur" du système Gateway	Invitrogen
pDONR207-amasPmas	plasmide "donneur" du système Gateway contenant le promoteur AmasPmas	PEB-Perth
pDONR221	plasmide "donneur" du système Gateway	Invitrogen
pGWB2	vecteur d'expression en plante-système Gateway-P35S	URGV-Evry
pGWB6	vecteur d'expression en plante-système Gateway-P35S-fusion N-6his	URGV-Evry
pGWB9	vecteur d'expression en plante-système Gateway-P35S-fusion N-GFP	URGV-Evry
pKm24GW	vecteur d'expression en plante-système Gateway-recombinaison multisite	PEB-Perth
p19	vecteur permettant l'expression d'un gène suppresseur de "silencing" en cellules végétales	URGV-Evry
pGFP	vecteur permettant l'expression de la protéine GFP en cellules végétales	URGV-Evry
pGreen0229-RFP2	vecteur d'expression en plante-système Gateway-fusion C-DsRed2	URGV-Evry
pAGRIKOLA	vecteur d'expression en plante-système Gateway-"ARNi" constitutif	projet Agrikola
pAGRII	vecteur d'expression en plante-système Gateway-"ARNi" inductible 6xpOp/LhGR	projet Agrikola
pAGRIKOLAind3	vecteur d'expression en plante-système Gateway-"ARNi" inductible 6xpOp/LhGR	projet Agrikola
vecteur d'expression en système bactérien voir table n° 7		

**Annexe N°2 : Liste des principales constructions réalisées au cours de mon doctorat**

nom	caractéristiques principales
pDONR207-AtDYW1-ATG2-STOP	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur <i>AtDYW1</i>
pDONR207-AtDYW1-sansATG2-STOP	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur sans ATG
pDONR207-AtDYW1-ATG2-préséquence	vecteur "donneur" Gateway-préséquence de l'ORF <i>AtDYW1</i>
pDONR207-AtDYW1-ATG2-end	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur sans CODON STOP
pDONR207-AtDYW1-GST-CATMA1a38560	vecteur "donneur" Gateway- GST CATMA1a38560 de l'ORF <i>AtDYW1</i>
pDNOR207-AtDYW1-int-STOP	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> sans préséquence (délétion arbitraire de 100 premiers acides aminés)
pDNOR207-CRR4-int-STOP	vecteur "donneur" Gateway-ORF CRR4 sans préséquence (délétion arbitraire de 100 premiers acides aminés)
pDONR207-ndhD	vecteur "donneur" Gateway-fragment du transcrit ndhD non édité
pDONR221-AtDYW1-ATG2-STOP-mutation1	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M1
pDONR221-AtDYW1-ATG2-STOP-mutation2	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M2
pDONR221-AtDYW1-ATG2-STOP-mutation3	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M3
pDONR221-AtDYW1-ATG2-STOP-mutation4	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M4
pDONR221-AtDYW1-ATG2-STOP-mutation5	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M5
pDONR221-AtDYW1-ATG2-STOP-mutation6	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M7
pDONR221-AtDYW1-ATG2-STOP-mutation8	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M8
pDONR221-AtDYW1-ATG2-STOP-mutation9	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M9
pDONR221-AtDYW1-ATG2-STOP-mutation11	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur portant la mutation M11
pDONR221-AtDYW1-ATG2-END-mutation2	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur sans CODON STOP portant la mutation M2
pDONR221-AtDYW1-ATG2-END-mutation4	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur sans CODON STOP portant la mutation M4
pDONR221-AtDYW1-ATG2-END-mutation5	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur sans CODON STOP portant la mutation M5
pDONR221-AtDYW1-ATG2-END-mutation11	vecteur "donneur" Gateway-ORF <i>AtDYW1</i> pleine longueur sans CODON STOP portant la mutation M11

Annexes

pGreen0229-RFP2-AtDYW1-préséquence	expression en cellules végétales d'une protéine de fusion préséquenceAtDYW1::DsRed2
pGreen0229-RFP2-AtDYW1-ATG2-end	expression en cellules végétales d'une protéine de fusion AtDYW1::DsRed3
pAGRIKOLA-GST-CATMA1a38560	expression en cellules végétales d'un hpRNA dirigé contre l'ORF <i>AtDYW1</i>
pAGRII-GST-CATMA1a38560	expression en cellules végétales d'un hpRNA dirigé contre l'ORF <i>AtDYW1</i>
pAGRIKOLAind3-GST-CATMA1a38560	expression en cellules végétales d'un hpRNA dirigé contre l'ORF <i>AtDYW1</i>
pGWB2-AtDYW1-int-STOP	expression en cellules végétales de la protéine rDYW1
pGWB6-AtDYW1-int-STOP	expression en cellules végétales de la protéine rDYW1 (tag 6his N-terminal)
pGWB9-AtDYW1-int-STOP	expression en cellules végétales de la protéine rDYW1 (tag GFP N-terminal)
pGWB2-CRR4-int-STOP	expression en cellules végétales de la protéine rCRR4
pGWB6-CRR4-int-STOP	expression en cellules végétales de la protéine rCRR4 (tag 6his N-terminal)
pGWB9-CRR4-int-STOP	expression en cellules végétales de la protéine rCRR4 (tag GFP N-terminal)
pGWB2-ndhD	expression en cellules végétales d'un fragment de transcrit <i>ndhD</i>
pDEST17-AtDYW1-int-STOP	expression en cellules bactériennes de la protéine rDYW1 (tag 6his N-terminal)
pGWB2-AtDYW1-ATG2-STOP	expression en cellules végétales de la protéine AtDYW1
pKm24GW-AmasPmas-ATG2-STOP	expression en cellules végétales de la protéine AtDYW1
pKm24GW-AmasPmas-ATG2-STOP-mutation2	expression en cellules végétales de la protéine AtDYW1portant la mutation m2
pKm24GW-AmasPmas-ATG2-STOP-mutation6	expression en cellules végétales de la protéine AtDYW1portant la mutation m6
pKm24GW-AmasPmas-ATG2-STOP-mutation8	expression en cellules végétales de la protéine AtDYW1portant la mutation m8

**Annexe N°3 : Liste des principaux oligonucléotides utilisés**

nom de l'oligonucléotide	séquence de 5' en 3'	principale utilisation
HSEK	CACCTCGAGCAYWSNGARAAR	recherche du domaine DYW
KNLR	CACCTCGAGAAARAAYTNM	recherche du domaine DYW
CxDCH	CACCTCGAGTGYIIIIAYTGICAY	recherche du domaine DYW
RFHHF	GCTCTAGARAARTGRTGRAANC	recherche du domaine DYW
DYWSTOP	GCTCTAGAYYACCARTARTC	recherche du domaine DYW
CACGDYW	TYACCARTARTCNCCACAAGCACA	recherche du domaine DYW
CSCRDYW	TYACCARTARTCNCTACAAGAACA	recherche du domaine DYW

Le code d'utilisation des bases dégénérées est présenté en annexe n°4.

kbestF	GCGCAACAGTATTGAATGA	Validation EST karenia brevis
kbestR	CCTCTGCAAGAAGGTCGAAG	Validation EST karenia brevis
kbdywF	AAATTCAGGCACGACAGCTT	Validation EST karenia brevis
kbdywR	CCAGCAGTTGTGAGGCATAC	Validation EST karenia brevis

kb18sF	CAGTTCCGACGGTAGGGTA	Contrôle qualité de l'ADNg d'espèce d'intérêt
kb18sR	ACCTCCCTGTGCAGGATTG	Contrôle qualité de l'ADNg d'espèce d'intérêt
Dinocox1F	AAAAATTGTAATCATAAACGCTTAAGG	Contrôle qualité de l'ADNg d'espèce d'intérêt
Dinocox1R	TGTTGAGCCACCTATAGTAAACATTA	Contrôle qualité de l'ADNg d'espèce d'intérêt
leftcryp	AGCTGCGAATGGCTCATTAT	Contrôle qualité de l'ADNg d'espèce d'intérêt
rightcryp	AGGCCAATACCCTACCATCC	Contrôle qualité de l'ADNg d'espèce d'intérêt
Leftpfiest	GCTCGAACAAACGATGAAGG	Contrôle qualité de l'ADNg d'espèce d'intérêt
Rightpfiest	ACTCCGGAAGAAGAGGTTGG	Contrôle qualité de l'ADNg d'espèce d'intérêt
leftproro	TTACAACCTTCAGCGACGGATG	Contrôle qualité de l'ADNg d'espèce d'intérêt
rightproro	AGCAGGCATGCTTTCAGG	Contrôle qualité de l'ADNg d'espèce d'intérêt
ColeoF18s	CGATGGAACCTTTGAGGCAAT	Contrôle qualité de l'ADNg d'espèce d'intérêt
ColeoR18s	CGAAAACCTTCACATCAT	Contrôle qualité de l'ADNg d'espèce d'intérêt
ClostF18s	CTCACCCATCGAAGCGATAC	Contrôle qualité de l'ADNg d'espèce d'intérêt
ClostR18s	CGCTTGCCGACTGTATCTTA	Contrôle qualité de l'ADNg d'espèce d'intérêt
KlebF18s	ACCTCTTCAAAGCCGTCGTA	Contrôle qualité de l'ADNg d'espèce d'intérêt



## Annexes

KlebR18s	ATGCCACGTGTCACCATCTA	Contrôle qualité de l'ADNg d'espèce d'intérêt
ChloF18S	TTCTTCGTGGCCAACTTCTT	Contrôle qualité de l'ADNg d'espèce d'intérêt
ChloR18S	GTACAAAGGGCAGGGACGTA	Contrôle qualité de l'ADNg d'espèce d'intérêt
ChaeF18s	CTAGCCGATGGAAGTTCGAG	Contrôle qualité de l'ADNg d'espèce d'intérêt
ChaeR18s	CGAAAACTACCCGGATCAT	Contrôle qualité de l'ADNg d'espèce d'intérêt
MarcF18s	ATGGCCGTTCTTAGTTGGTG	Contrôle qualité de l'ADNg d'espèce d'intérêt
MarcR18s	TCCCCATCACGATGAATTTT	Contrôle qualité de l'ADNg d'espèce d'intérêt
LunuF18s	AAACGGCTACCACATCCAAG	Contrôle qualité de l'ADNg d'espèce d'intérêt
LunuR18s	CAAAGAGAGAGGCGACCAG	Contrôle qualité de l'ADNg d'espèce d'intérêt
ConoF18s	ATGGCCGTTCTTAGTTGGTG	Contrôle qualité de l'ADNg d'espèce d'intérêt
ConoR18s	TCCCCATCACGATGAATTTT	Contrôle qualité de l'ADNg d'espèce d'intérêt
AsteF26s	ACGTTTGAGGCGCAAATAAC	Contrôle qualité de l'ADNg d'espèce d'intérêt
AsteR26s	CTCCTCGTGCTCCAGCTATC	Contrôle qualité de l'ADNg d'espèce d'intérêt
MonoF26s	CGGGTTAGTCGATCCTGAGA	Contrôle qualité de l'ADNg d'espèce d'intérêt
MonoR26s	GCGATGAGTACTCGCTCCTC	Contrôle qualité de l'ADNg d'espèce d'intérêt
RiccF18s	ATGGCCGTTCTTAGTTGGTG	Contrôle qualité de l'ADNg d'espèce d'intérêt
RiccR18s	TCCCCATCACGATGAATTTT	Contrôle qualité de l'ADNg d'espèce d'intérêt

PQpdkintron_R	TTGATAGATCTTGCGCTTTGTT	RT-PCR Agrikola
PQpdkintron_L	TGTTAGAAATTCCAATCTGCTTGT	RT-PCR Agrikola
ActineF	GCCATCCAAGCTGTTCTCTC	RT-PCR Agrikola
ActineR	CCCTCGTAGATTGGCACAGT	RT-PCR Agrikola
GusR	GGCACAGCACATCAAAGAGA	RT-PCR Agrikola
GusF	CTGATAGCGCGTGACAAAAA	RT-PCR Agrikola
PPR43F	TTTTCAGAGCTGGGCAGATT	RT-PCR Agrikola
PPR43R	AGCATTGTACGTCGCAACAG	RT-PCR Agrikola
AGRI51	CAACCACGTCTTCAAAGCAA	Génotypage lignées Agrikola
omegaAGRI51	GTTCATTTTCATTTGGAGAGGA	Génotypage lignées Agrikola
AGRI56	CTG GGG TAC CGA ATT CCT C	Génotypage lignées Agrikola
AGRI69	AGG CGT CTC GCA TAT CTC AT	Génotypage lignées Agrikola
AGRI64	CTT GCG CTG CAG TTA TCA TC	Génotypage lignées Agrikola
5'GSTDYW	GTGGCAGAGAAACAGCAAGAG	Génotypage lignées Agrikola

Annexes

GST-CATMA1a38560	ATTCAGCTTTACGTAGGCCTTGCTTCTAT	clonage Gateway AtDYW1 GST
GST-CATMA1a38561	CAAGAAAGCTGGGTCCCTACCAATAATCTCCACAAGAG	clonage Gateway AtDYW1 GST
DYW1-atg-sdk	AAAAAAGCAGGCTTCGTTGAGACTCAAGTCAAGGAAAC	clonage Gateway AtDYW1 ORF pleine longueur sens
49ATG2	GGAGATAGAACCATGAGTTTTTGCTCCTTCTCTGT	clonage Gateway AtDYW1 ORF pleine longueur sens (-ATG-SDK)
49end	CAAGAAAGCTGGGTCCCAATAATCTCCACAAGAGCAG	clonage Gateway AtDYW1 ORF pleine longueur antisens (sans STOP)
49stop	CAAGAAAGCTGGGTCCCTACCAATAATCTCCACAAGAG	clonage Gateway AtDYW1 ORF pleine longueur antisens (avec STOP)
49interne	GGAGATAGAACCATGTTGAGACTCAAGTCAAGGAAAC	clonage Gateway AtDYW1 ORF sans peptide signal sens
49préséquence	CAAGAAAGCTGGGTCTCGTAAAGTAGTATCTTTTCGTACAG	clonage Gateway AtDYW1 peptide signal antisens
DYW1m1s	CATGCTAGTGAACGGTTA	clonage Gateway AtDYW1 forme mutante1
DYW1m1as	ACTAGCATGCATCAAGGC	clonage Gateway AtDYW1 forme mutante1
DYW1m2s	AGTGACAGGTTAGCTATT	clonage Gateway AtDYW1 forme mutante2
DYW1m2as	CCGTGCACTATGATGCATC	clonage Gateway AtDYW1 forme mutante2
DYW1m3s	CATGCTAGTGCACGGTTAGCT	clonage Gateway AtDYW1 forme mutante3
DYW1m3as	CCGTGCACTAGCATGCATCAAG	clonage Gateway AtDYW1 forme mutante3
DYW1m4s	ATAGCTGGAGATTGC	clonage Gateway AtDYW1 forme mutante4
DYW1m4as	TCCAGCTATCCTCAG	clonage Gateway AtDYW1 forme mutante4
DYW1m5s	GATGCCATAACTTCA	clonage Gateway AtDYW1 forme mutante5
DYW1m5as	ATGGGCATCTCCACATAT	clonage Gateway AtDYW1 forme mutante5
DYW1m6s	ATAGCTGGAGATGCCATAACTT	clonage Gateway AtDYW1 forme mutante6
DYW1m6as	ATGGGCATCTCCAGCTATCCTCAG	clonage Gateway AtDYW1 forme mutante6
DYW1m7s	TTCGCTGCTTTTAGAGAT	clonage Gateway AtDYW1 forme mutante7
DYW1m7as	AAAAGCAGCGAACCTCTT	clonage Gateway AtDYW1 forme mutante7
DYW1m8s	AACGCCTCTTGTGG	clonage Gateway AtDYW1 forme mutante8
DYW1m8as	AGAGGCGTTACCATC	clonage Gateway AtDYW1 forme mutante8
DYW1m9s	TCTGCTGGAGATTAT	clonage Gateway AtDYW1 forme mutante9
DYW1m9as	TCCAGCAGAGCAGTT	clonage Gateway AtDYW1 forme mutante9
DYW1m10s	AACGCCTCTGCTGGAGATTATTG	clonage Gateway AtDYW1 forme mutante10
DYW1m10as	TCCAGCAGAGGCGTTACCATC	clonage Gateway AtDYW1 forme mutante10
DYW1m11as	TCCACCTCCGGATCMTCCACAAGAGCAGTTACCATC	clonage Gateway AtDYW1 forme mutante11
DYW1end/stop	TCCACCTCCGGATCMCCAATAATCTCCACAAGAGCAG	clonage Gateway AtDYW1 antisens (end/stop)
M13R	CAGGAAACAGCTATGAC	clonage/séquencage
M13F	GTTTTCCCAGTCACGAC	clonage/séquencage

## Annexes

AmasPmas	TCTAGATTTTTCAAATCAGTG	Génotypage lignées de surexpression
P35S	CAACCACGTCTTCAAAGCAA	Génotypage lignées de surexpression
CRR4rtF	TGTGATGCCGAGAAGAGATG	RT-PCR gène CRR4
CRR4rtR	CAACACCGAGTTTTCCACCT	RT-PCR gène CRR4
CRR4int	GGAGATAGAACCATGGCTTTTGCCTCTTCTCGACGTCCGTATC	clonage Gateway CRR4
CRR4end/stop	TCCACCTCCGGATCMCAATGTA CTG GAAACTTCAATGCTATCTAC	clonage Gateway CRR4
ndhDR	TCCACCTCCGGATCTATTGAAAAACAACAATTA	clonage Gateway ndhD
ndhDF	GAAGGAGATAGAACCTTTGAGTACGCGTTCTTGG	clonage Gateway ndhD
PQactin28_F	GGTAACATTGTGCTCAGTGGTGG	qPCR
PQactin28_R	AACGACCTTAATCTTCATGCTGC	qPCR
DYW1endoF	AGATGGTAACTGCTCTTGTTGGAG	qPCR
DYW1endoR	TGGAAGTTGCATAAAATGGCTTA	qPCR
DYW1surexpF	TGGTGATTTTTGCGGACTCTA	qPCR
DYW1surexpR	AGAGATGGTAACTGCTCTTGTTGG	qPCR
GSTDYW1R	ATTCAGCTTTACGTAGGCCTTGCTTCTAT	RT-PCR / qPCR gène AtDYW1
GSTDYW1F	GCCTGAAAGAAATTTAGATACTTGGAGC	RT-PCR / qPCR gène AtDYW1
Mito11F	TTGGGTCATCTCAATGGGTT	editotype mitochondrial
Mito11R	CGATCGATATTTGCGGAGTT	editotype mitochondrial
Mito28F	CGTATATCGTCGTAACGCC	editotype mitochondrial
Mito28R	AAGAGCCGAACGAGAATGAA	editotype mitochondrial
Mito128F	AGTGGGAGAGCCGTGTTATG	editotype mitochondrial
Mito128R	CTTTGCTTGCTGTGCAGTGA	editotype mitochondrial
Mito171F	TTAATTTGCGATCGGAAACG	editotype mitochondrial
Mito171R	CAGGAACAAGAAGAAGCGAGA	editotype mitochondrial
Mito95F	GCCTTGATTATGCCGGG	editotype mitochondrial
Mito95R	TGAGACAAGAGAATGAAGAGG	editotype mitochondrial

Chloro 260F	AAGATGTCGATCGTAAATGCG	editotype chloroplastique
Chloro 260R	TCGTTTGATTATCCCCTAAGG	editotype chloroplastique
Chloro 27F	CAGCCTATTCTTGAATGAACAAC	editotype chloroplastique
Chloro 27R	GATTGCTTTCTTCCTTCCC	editotype chloroplastique
23581F	GAGGTGGGTTCCAGAAAAAGG	editotype chloroplastique
24093R	TATCTGTCCTACATTCATGCG	editotype chloroplastique
Chloro 45F	AAATTTCCAGGGTAGCAAACATTC	editotype chloroplastique
Chloro 45R	AAATTGAAGATATAGATCACGAAATTGA	editotype chloroplastique
Chloro 112F	AAATTACAAGATAAGAATAATAGGAATCG	editotype chloroplastique
Chloro 112R	AGAGGTAAACATTGATTGGTAGCAT	editotype chloroplastique
Chloro 125F	ACGTCTATAACTCAGCTCATCCAA	editotype chloroplastique
Chloro 125R	CATGGACTAGCTGTACCTACCGT	editotype chloroplastique
Chloro 126F	GGTTTATCGTTGGATGAACTGC	editotype chloroplastique
Chloro 126R	AAATACCCTTGGTACTTTATTGACG	editotype chloroplastique
Chloro 130F	GCCAAAGACGATGAATTGAGTT	editotype chloroplastique
Chloro 130R	GAGAGGTACTATCTATTCTAGCTCTTAATC	editotype chloroplastique
Chloro 141F	ACATGTTCTCGTCGCTGA	editotype chloroplastique
Chloro 141R	TTTATGAGGCACAAACGGGA	editotype chloroplastique
Chloro 185F	AGGTCAGATTTGCGCTATTCC	editotype chloroplastique
Chloro 185R	TGACTGGACGAAACCAAGAA	editotype chloroplastique
Chloro 189F	TCTCAGAGATAGATAGAGAGGTAGGAATTT	editotype chloroplastique
Chloro 189R	CATTGAATGTACAGAAATGGCTATAA	editotype chloroplastique
Chloro 190F	AAACATTCCTCCTAGAGTAGCTGTTAAT	editotype chloroplastique
Chloro 190R	TTCCGTAATATTCCATTCCAAA	editotype chloroplastique
Chloro 204F	GGTATCCTTGATCATGCGACATA	editotype chloroplastique
Chloro 204R	TCAATATTCATGGATCATCCCTTT	editotype chloroplastique
Chloro 212F	AGCGCCAATAAATCCATGAG	editotype chloroplastique
Chloro 212R	GTTCGAACGAACCAACATTAATA	editotype chloroplastique
Chloro 214F	TCGATCCATTTATAATCTTCGGA	editotype chloroplastique
Chloro 214R	GACTGTGTTGGTTGTAAGAGATGTG	editotype chloroplastique

## Annexes

LP-SALK-123655	CAGCAAGAGGACTTTCAGGTTCA	Génotypage lignées SALK
RP-SALK-123655	TGGCGTTTCGGTTTATGGATG	Génotypage lignées SALK
RP-SALK-012425	CACCCCTGAAATCTCTGGCT	Génotypage lignées SALK
LP-SALK-012425	TGTGGTTTGATCTGATTCTGTC	Génotypage lignées SALK
LBa1	TGGTTCACGTAGTGGCCATCG	Génotypage lignées SALK
Lb	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC	Génotypage lignées SALK

U3 standard	GGGGACCACTTTGTACAAGAAAGCTGGGTC	Clonage Gateway
U5 standard	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG	Clonage Gateway
U5 PTO	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG	Clonage Gateway
U3 PTO end/stop	CTAGGCCTCCACCTCTGGGTCGAAAGAACATGTTTCACCAGGGG	Clonage Gateway
U5-ATG-SDK	GGGGACAAGTTTGTACAAAAAAGCAGGCTTC	Clonage Gateway
seqLB	GTAACATCAGAGATTTTGAGACAC	Clonage Gateway

Amorce PPE PsbZ	TTGGCAGTTTTGCATTAATTACTTTC	PPE
Amorce PPE ndhF	TTCGGTTACTTTATCTTTATCGATCCACTTACTTC	PPE
Amorce PPE Rps12	CATTCTATTTTATTCATAGGTAAGTACTGATCCG	PPE
Amorce PPE NdhB12	TGATTGTATGTGTGATAGCATCTACTATAC	PPE
Amorce PPE accD	TGAAAATGAGTAGTTCAGAAAGAATCGAGC	PPE
Amorce PPE clpP	ATGTTTTTATGTCAGCAACAGAAGCCCAAGC	PPE
Amorce PPE rpoA	CGAAATAGAAGGAACATGTATTACACGTGC	PPE
Amorce PPE ndhD	TCTTTGGACCTGGTGTATCTTGCTTTACC	PPE
Amorce PPE ndhG	CACTCCCAGACCCCTACTAGAAGACCG	PPE
Amorce PPE rpoC1	TAATAAAAAACCTACTTTCTTACgATTAC	PPE
Amorce PPE petL	AGTAGAAAACCGAAATAACTAGTTATAGTA	PPE
Amorce PPE rps12	CATTCTATTTTATTCATAGGTAAGTACTGATCCG	PPE

Amorces de crible HRM	voir Chateigner-Boutin et Small, 2007 (table supplémentaire, disponible sur le site NAR online)	
--------------------------	---	--

#### **Annexe N°4 : Code d'utilisation des bases dégénérées**

**Les oligonucléotides "dégénérés" : un mélange de molécules d'ADN**

La nomenclature internationale inclut des abréviations pour les bases "normales" d'ADN : les purines G (Guanine) et A (Adénine), les pyrimidines C (Cytosine) et T (Thymine) ; en plus, il existe 11 lettres supplémentaires pour les bases "dégénérées". En fait, les bases dans un oligonucléotide ne sont pas dégénérées. Ce qui est dégénéré est la position, et non la base elle-même. Par exemple, si l'amorce contient un R à une position donnée, votre amorce correspondra à un mélange équimolaire d'oligonucléotides dont environ 50 % des brins d'ADN auront un G dans cette position et les autres 50 % posséderont un A.

Nomenclature Internationale de bases dégénérées

R = G ou A

K = G ou T

S = G ou C

W = A ou T

M = A ou C

Y = T ou C

D = G ou A ou T

V = G ou A ou C

B = G ou T ou C

H = A ou T ou C

N = G ou A ou T ou C

## **Références Bibliographiques**

- Akagi, H., Nakamura, A., Yokozeki-Misono, Y., Inagaki, A., Takahashi, H., Mori, K., and Fujimura, T.** (2004). Positional cloning of the rice Rf-1 gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. *Theor Appl Genet* **108**, 1449-1457.
- Akbarian, S., Smith, M.A., and Jones, E.G.** (1995). Editing for an AMPA receptor subunit RNA in prefrontal cortex and striatum in Alzheimer's disease, Huntington's disease and schizophrenia. *Brain Res* **699**, 297-304.
- Alexander, M.P.** (1969). Differential staining of aborted and nonaborted pollen. *Stain Technol* **44**, 117-122.
- Alfonzo, J.D., Thiemann, O.H., and Simpson, L.** (1998). Purification and characterization of MAR1. A mitochondrial associated ribonuclease from *Leishmania tarentolae*. *J Biol Chem* **273**, 30003-30011.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.
- Amaro, R.E., Swift, R.V., and McCammon, J.A.** (2007). Functional and Structural Insights Revealed by Molecular Dynamics Simulations of an Essential RNA Editing Ligase in *Trypanosoma brucei*. *PLoS Negl Trop Dis* **1**, e68.
- Ammerman, M.L., Fisk, J.C., and Read, L.K.** (2008). gRNA/pre-mRNA annealing and RNA chaperone activities of RBP16. *Rna* **14**, 1069-1080.
- Anant, S., and Davidson, N.O.** (2000). An AU-rich sequence element (UUUN[A/U]U) downstream of the edited C in apolipoprotein B mRNA is a high-affinity binding site for Apobec-1: binding of Apobec-1 to this motif in the 3' untranslated region of c-myc increases mRNA stability. *Mol Cell Biol* **20**, 1982-1992.
- Anant, S., MacGinnitie, A.J., and Davidson, N.O.** (1995). apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, is a novel RNA-binding protein. *J Biol Chem* **270**, 14762-14767.
- Aphasizhev, R., and Simpson, L.** (2001). Isolation and characterization of a U-specific 3'-5'-exonuclease from mitochondria of *Leishmania tarentolae*. *J Biol Chem* **276**, 21280-21284.
- Aphasizhev, R., and Aphasizheva, I.** (2008). Terminal RNA uridylyltransferases of trypanosomes. *Biochim Biophys Acta* **1779**, 270-280.
- Aphasizhev, R., Aphasizheva, I., and Simpson, L.** (2003a). A tale of two TUTases. *Proc Natl Acad Sci U S A* **100**, 10617-10622.
- Aphasizhev, R., Aphasizheva, I., Nelson, R.E., and Simpson, L.** (2003b). A 100-kD complex of two RNA-binding proteins from mitochondria of *Leishmania tarentolae* catalyzes RNA annealing and interacts with several RNA editing components. *Rna* **9**, 62-76.
- Aphasizhev, R., Sbicego, S., Peris, M., Jang, S.H., Aphasizheva, I., Simpson, A.M., Rivlin, A., and Simpson, L.** (2002). Trypanosome mitochondrial 3' terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing. *Cell* **108**, 637-648.

- Aphasizhev, R., Aphasizheva, I., Nelson, R.E., Gao, G., Simpson, A.M., Kang, X., Falick, A.M., Sbicego, S., and Simpson, L.** (2003c). Isolation of a U-insertion/deletion editing complex from *Leishmania tarentolae* mitochondria. *Embo J* **22**, 913-924.
- Aruscavage, P.J., and Bass, B.L.** (2000). A phylogenetic analysis reveals an unusual sequence conservation within introns involved in RNA editing. *Rna* **6**, 257-269.
- Athanasiadis, A., Rich, A., and Maas, S.** (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* **2**, e391.
- Aubourg, S., Boudet, N., Kreis, M., and Lecharny, A.** (2000). In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol Biol* **42**, 603-613.
- Auxilien, S., Crain, P.F., Trewyn, R.W., and Grosjean, H.** (1996). Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *J Mol Biol* **262**, 437-458.
- Backus, J.W., and Smith, H.C.** (1992). Three distinct RNA sequence elements are required for efficient apolipoprotein B (apoB) RNA editing in vitro. *Nucleic Acids Res* **20**, 6007-6014.
- Bass, B.L.** (2002). RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* **71**, 817-846.
- Batzer, M.A., and Deininger, P.L.** (2002). Alu repeats and human genomic diversity. *Nat Rev Genet* **3**, 370-379.
- Beghini, A., Ripamonti, C.B., Peterlongo, P., Roversi, G., Cairoli, R., Morra, E., and Larizza, L.** (2000). RNA hyperediting and alternative splicing of hematopoietic cell phosphatase (PTPN6) gene in acute myeloid leukemia. *Hum Mol Genet* **9**, 2297-2304.
- Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., and Tromp, M.C.** (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**, 819-826.
- Bentolila, S., Alfonso, A.A., and Hanson, M.R.** (2002). A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci U S A* **99**, 10887-10892.
- Bentolila, S., Chateigner-Boutin, A.L., and Hanson, M.R.** (2005). Ecotype allelic variation in C-to-U editing extent of a mitochondrial transcript identifies RNA-editing quantitative trait loci in *Arabidopsis*. *Plant Physiol* **139**, 2006-2016.
- Betts, L., Xiang, S., Short, S.A., Wolfenden, R., and Carter, C.W., Jr.** (1994). Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. *J Mol Biol* **235**, 635-656.
- Bhalla, T., Rosenthal, J.J., Holmgren, M., and Reenan, R.** (2004). Control of human potassium channel inactivation by editing of a small mRNA hairpin. *Nat Struct Mol Biol* **11**, 950-956.
- Billeter, M.A., Cattaneo, R., Spielhofer, P., Kaelin, K., Huber, M., Schmid, A., Bacsko, K., and ter Meulen, V.** (1994). Generation and properties of measles virus mutations typically associated with subacute sclerosing panencephalitis. *Ann N Y Acad Sci* **724**, 367-377.
- Bjork, G.R., Jacobsson, K., Nilsson, K., Johansson, M.J., Bystrom, A.S., and Persson, O.P.** (2001). A primordial tRNA modification required for the evolution of life? *Embo J* **20**, 231-239.
- Blanc, V., Alfonzo, J.D., Aphasizhev, R., and Simpson, L.** (1999). The mitochondrial RNA ligase from *Leishmania tarentolae* can join RNA molecules bridged by a complementary RNA. *J Biol Chem* **274**, 24289-24296.



- Blanc, V., Henderson, J.O., Kennedy, S., and Davidson, N.O.** (2001a). Mutagenesis of apobec-1 complementation factor reveals distinct domains that modulate RNA binding, protein-protein interaction with apobec-1, and complementation of C to U RNA-editing activity. *J Biol Chem* **276**, 46386-46393.
- Blanc, V., Henderson, J.O., Newberry, E.P., Kennedy, S., Luo, J., and Davidson, N.O.** (2005). Targeted deletion of the murine apobec-1 complementation factor (acf) gene results in embryonic lethality. *Mol Cell Biol* **25**, 7260-7269.
- Blanc, V., Navaratnam, N., Henderson, J.O., Anant, S., Kennedy, S., Jarmuz, A., Scott, J., and Davidson, N.O.** (2001b). Identification of GRY-RBP as an apolipoprotein B RNA-binding protein that interacts with both apobec-1 and apobec-1 complementation factor to modulate C to U editing. *J Biol Chem* **276**, 10272-10283.
- Blatch, G.L., and Lassle, M.** (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**, 932-939.
- Blom, D., Burg, J., Breek, C.K., Speijer, D., Muijsers, A.O., and Benne, R.** (2001). Cloning and characterization of two guide RNA-binding proteins from mitochondria of *Crithidia fasciculata*: gBP27, a novel protein, and gBP29, the orthologue of *Trypanosoma brucei* gBP21. *Nucleic Acids Res* **29**, 2950-2962.
- Blow, M., Futreal, P.A., Wooster, R., and Stratton, M.R.** (2004). A survey of RNA editing in human brain. *Genome Res* **14**, 2379-2387.
- Bock, R., Kossel, H., and Maliga, P.** (1994). Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. *Embo J* **13**, 4623-4628.
- Bock, R., Hermann, M., and Kossel, H.** (1996). In vivo dissection of cis-acting determinants for plastid RNA editing. *Embo J* **15**, 5052-5059.
- Bratt, E., and Ohman, M.** (2003). Coordination of editing and splicing of glutamate receptor pre-mRNA. *Rna* **9**, 309-318.
- Brown, G.G., Formanova, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., Lafortest, M., Zhang, J., Cheung, W.Y., and Landry, B.S.** (2003). The radish Rfo restorer gene of *Ogura* cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J* **35**, 262-272.
- Brown, J.W., Shaw, P.J., Shaw, P., and Marshall, D.F.** (2005). Arabidopsis nucleolar protein database (AtNoPDB). *Nucleic Acids Res* **33**, D633-636.
- Budar, F., Touzet, P., and De Paepe, R.** (2003). The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica* **117**, 3-16.
- Burgess, M.L., and Stuart, K.** (2000). Sequence bias in edited kinetoplastid RNAs. *Rna* **6**, 1492-1497.
- Burgess, M.L., Heidmann, S., and Stuart, K.** (1999). Kinetoplastid RNA editing does not require the terminal 3' hydroxyl of guide RNA, but modifications to the guide RNA terminus can inhibit in vitro U insertion. *Rna* **5**, 883-892.
- Burns, C.M., Chu, H., Rueter, S.M., Hutchinson, L.K., Canton, H., Sanders-Bush, E., and Emeson, R.B.** (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**, 303-308.
- Bush, S.M., and Krysan, P.J.** (2007). Mutational evidence that the Arabidopsis MAP kinase MPK6 is involved in anther, inflorescence, and embryo development. *J Exp Bot* **58**, 2181-2191.
- Byrne, E.M., Visomirski-Robic, L., Cheng, Y.W., Rhee, A.C., and Gott, J.M.** (2007). RNA editing in *Physarum* mitochondria: assays and biochemical approaches. *Methods Enzymol* **424**, 143-172.
- Cappione, A.J., French, B.L., and Skuse, G.R.** (1997). A potential role for NF1 mRNA editing in the pathogenesis of NF1 tumors. *Am J Hum Genet* **60**, 305-312.

- Carlow, D.C., Short, S.A., and Wolfenden, R.** (1996). Role of glutamate-104 in generating a transition state analogue inhibitor at the active site of cytidine deaminase. *Biochemistry* **35**, 948-954.
- Carlson, C.B., Stephens, O.M., and Beal, P.A.** (2003). Recognition of double-stranded RNA by proteins and small molecules. *Biopolymers* **70**, 86-102.
- Carrillo, C., and Bonen, L.** (1997). RNA editing status of nad7 intron domains in wheat mitochondria. *Nucleic Acids Res* **25**, 403-409.
- Carter, C.** (1998). Modification and editing of RNA. In *Modification and editing of RNA*, H. Grosjean and R. Benne, eds (ASM Press), pp. 363-375.
- Carter, C.W., Jr.** (1995). The nucleoside deaminases for cytidine and adenosine: structure, transition state stabilization, mechanism, and evolution. *Biochimie* **77**, 92-98.
- Cattaneo, R.** (1994). Biased (A→I) hypermutation of animal RNA virus genomes. *Curr Opin Genet Dev* **4**, 895-900.
- Caudy, A.A., Ketting, R.F., Hammond, S.M., Denli, A.M., Bathoorn, A.M., Tops, B.B., Silva, J.M., Myers, M.M., Hannon, G.J., and Plasterk, R.H.** (2003). A micrococcal nuclease homologue in RNAi effector complexes. *Nature* **425**, 411-414.
- Chan, L., Chang, B.H., Nakamuta, M., Li, W.H., and Smith, L.C.** (1997). Apobec-1 and apolipoprotein B mRNA editing. *Biochim Biophys Acta* **1345**, 11-26.
- Chateigner-Boutin, A.L., and Hanson, M.R.** (2002). Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared cis elements. *Mol Cell Biol* **22**, 8448-8456.
- Chateigner-Boutin, A.L., and Small, I.** (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res* **35**, e114.
- Chaudhuri, S., and Maliga, P.** (1996). Sequences directing C to U editing of the plastid psbL mRNA are located within a 22 nucleotide segment spanning the editing site. *Embo J* **15**, 5958-5964.
- Chaudhuri, S., Carrer, H., and Maliga, P.** (1995). Site-specific factor involved in the editing of the psbL mRNA in tobacco plastids. *Embo J* **14**, 2951-2957.
- Chen, C.X., Cho, D.S., Wang, Q., Lai, F., Carter, K.C., and Nishikura, K.** (2000). A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *Rna* **6**, 755-767.
- Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P.** (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83-86.
- Chen, K.M., Harjes, E., Gross, P.J., Fahmy, A., Lu, Y., Shindo, K., Harris, R.S., and Matsuo, H.** (2008). Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116-119.
- Chen, S.H., Habib, G., Yang, C.Y., Gu, Z.W., Lee, B.R., Weng, S.A., Silberman, S.R., Cai, S.J., Deslypere, J.P., Rosseneu, M., and et al.** (1987). Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* **238**, 363-366.
- Chester, A., Weinreb, V., Carter, C.W., Jr., and Navaratnam, N.** (2004). Optimization of apolipoprotein B mRNA editing by APOBEC1 apoenzyme and the role of its auxiliary factor, ACF. *Rna* **10**, 1399-1411.
- Chester, A., Somasekaram, A., Tzimina, M., Jarmuz, A., Gisbourne, J., O'Keefe, R., Scott, J., and Navaratnam, N.** (2003). The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *Embo J* **22**, 3971-3982.

- Cho, D.S., Yang, W., Lee, J.T., Shiekhattar, R., Murray, J.M., and Nishikura, K.** (2003). Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J Biol Chem* **278**, 17093-17102.
- Choury, D., and Araya, A.** (2006). RNA editing site recognition in heterologous plant mitochondria. *Curr Genet* **50**, 405-416.
- Choury, D., Farre, J.C., Jordana, X., and Araya, A.** (2004). Different patterns in the recognition of editing sites in plant mitochondria. *Nucleic Acids Res* **32**, 6397-6406.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**, 735-743.
- Coffin, J.W., Dhillon, R., Ritzel, R.G., and Nargang, F.E.** (1997). The *Neurospora crassa* *cya-5* nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein. *Curr Genet* **32**, 273-280.
- Coticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., and Neuberger, M.S.** (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* **22**, 367-377.
- Coticello, S.G., Langlois, M.A., Yang, Z., and Neuberger, M.S.** (2007). DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv Immunol* **94**, 37-73.
- Covello, P.S., and Gray, M.W.** (1989). RNA editing in plant mitochondria. *Nature* **341**, 662-666.
- Craft, J., Samalova, M., Baroux, C., Townley, H., Martinez, A., Jepson, I., Tsiantis, M., and Moore, I.** (2005). New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in *Arabidopsis*. *Plant J* **41**, 899-918.
- Crowe, M.L., Serizet, C., Thareau, V., Aubourg, S., Rouze, P., Hilson, P., Beynon, J., Weisbeek, P., van Hummelen, P., Reymond, P., Paz-Ares, J., Nietfeld, W., and Trick, M.** (2003). CATMA: a complete *Arabidopsis* GST database. *Nucleic Acids Res* **31**, 156-158.
- Cruz-Reyes, J.** (2007). RNA-protein interactions in assembled editing complexes in trypanosomes. *Methods Enzymol* **424**, 107-125.
- Cruz-Reyes, J., and Sollner-Webb, B.** (1996). Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. *Proc Natl Acad Sci U S A* **93**, 8901-8906.
- Cruz-Reyes, J., Rusche, L.N., Piller, K.J., and Sollner-Webb, B.** (1998). *T. brucei* RNA editing: adenosine nucleotides inversely affect U-deletion and U-insertion reactions at mRNA cleavage. *Mol Cell* **1**, 401-409.
- Cruz-Reyes, J., Zhelonkina, A.G., Huang, C.E., and Sollner-Webb, B.** (2002). Distinct functions of two RNA ligases in active *Trypanosoma brucei* RNA editing complexes. *Mol Cell Biol* **22**, 4652-4660.
- Cui, X., Wise, R.P., and Schnable, P.S.** (1996). The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* **272**, 1334-1336.
- Dance, G.S., Beemiller, P., Yang, Y., Mater, D.V., Mian, I.S., and Smith, H.C.** (2001). Identification of the yeast cytidine deaminase CDD1 as an orphan C $\rightarrow$ U RNA editase. *Nucleic Acids Res* **29**, 1772-1780.
- Das, A.K., Cohen, P.W., and Barford, D.** (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *Embo J* **17**, 1192-1199.
- Davidson, N.O., Innerarity, T.L., Scott, J., Smith, H., Driscoll, D.M., Teng, B., and Chan, L.** (1995). Proposed nomenclature for the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme: APOBEC-1. *Rna* **1**, 3.

- Davies, M.S., Wallis, S.C., Driscoll, D.M., Wynne, J.K., Williams, G.W., Powell, L.M., and Scott, J.** (1989). Sequence requirements for apolipoprotein B RNA editing in transfected rat hepatoma cells. *J Biol Chem* **264**, 13395-13398.
- Davis, C.C., and Wurdack, K.J.** (2004). Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* **305**, 676-678.
- Dawson, T.R., Sansam, C.L., and Emeson, R.B.** (2004). Structure and sequence determinants required for the RNA editing of ADAR2 substrates. *J Biol Chem* **279**, 4941-4951.
- de Longevialle, A.F., Meyer, E.H., Andres, C., Taylor, N.L., Lurin, C., Millar, A.H., and Small, I.D.** (2007). The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in *Arabidopsis thaliana*. *Plant Cell* **19**, 3256-3265.
- De Santis-Maclossek G, K.W., Bock A, Schoch S, Maier RM, Wanner G, Rüdiger W, Koop HU, Herrmann RG.** (1999). Targeted disruption of the plastid RNA polymerase genes rpoA, B and C1: molecular biology, biochemistry and ultrastructure. *PLANT J.* **18**, 477-489.
- Del Campo, E.M., Sabater, B., and Martin, M.** (2002). Post-transcriptional control of chloroplast gene expression. Accumulation of stable psaC mRNA is due to downstream RNA cleavages in the ndhD gene. *J Biol Chem* **277**, 36457-36464.
- Delannoy, E., Stanley, W.A., Bond, C.S., and Small, I.D.** (2007). Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochem Soc Trans* **35**, 1643-1647.
- Deng, J., Schnauffer, A., Salavati, R., Stuart, K.D., and Hol, W.G.** (2004). High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1. *J Mol Biol* **343**, 601-613.
- Desloire, S., Gherbi, H., Laloui, W., Marhadour, S., Clouet, V., Cattolico, L., Falentin, C., Giancola, S., Renard, M., Budar, F., Small, I., Caboche, M., Delourme, R., and Bendahmane, A.** (2003). Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep* **4**, 588-594.
- Ding, Y.H., Liu, N.Y., Tang, Z.S., Liu, J., and Yang, W.C.** (2006). *Arabidopsis* GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. *Plant Cell* **18**, 815-830.
- Donahue, R.A., and Bebee, R.L.** (1999). BL21-SI Competent cells for protein expression in *E.coli*. *Life Technologies* **21**, 49-52.
- Driscoll, D.M., and Casanova, E.** (1990). Characterization of the apolipoprotein B mRNA editing activity in enterocyte extracts. *J Biol Chem* **265**, 21401-21403.
- Driscoll, D.M., and Zhang, Q.** (1994). Expression and characterization of p27, the catalytic subunit of the apolipoprotein B mRNA editing enzyme. *J Biol Chem* **269**, 19843-19847.
- Driscoll, D.M., Wynne, J.K., Wallis, S.C., and Scott, J.** (1989). An in vitro system for the editing of apolipoprotein B mRNA. *Cell* **58**, 519-525.
- Driscoll, D.M., Lakhe-Reddy, S., Oleksa, L.M., and Martinez, D.** (1993). Induction of RNA editing at heterologous sites by sequences in apolipoprotein B mRNA. *Mol Cell Biol* **13**, 7288-7294.
- Drozd, M., Palazzo, S.S., Salavati, R., O'Rear, J., Clayton, C., and Stuart, K.** (2002). TbMP81 is required for RNA editing in *Trypanosoma brucei*. *Embo J* **21**, 1791-1799.
- Eckmann, C.R., Neunteufl, A., Pfaffstetter, L., and Jantsch, M.F.** (2001). The human but not the *Xenopus* RNA-editing enzyme ADAR1 has an atypical nuclear localization

- signal and displays the characteristics of a shuttling protein. *Mol Biol Cell* **12**, 1911-1924.
- Egebjerg, J., and Heinemann, S.F.** (1993). Ca<sup>2+</sup> permeability of unedited and edited versions of the kainate selective glutamate receptor GluR6. *Proc Natl Acad Sci U S A* **90**, 755-759.
- Ellis, T.P., Helfenbein, K.G., Tzagoloff, A., and Dieckmann, C.L.** (2004). Aep3p stabilizes the mitochondrial bicistronic mRNA encoding subunits 6 and 8 of the H<sup>+</sup>-translocating ATP synthase of *Saccharomyces cerevisiae*. *J Biol Chem* **279**, 15728-15733.
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G.** (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**, 1005-1016.
- Ernst, N.L., Panicucci, B., Igo, R.P., Jr., Panigrahi, A.K., Salavati, R., and Stuart, K.** (2003). TbMP57 is a 3' terminal uridylyl transferase (TUTase) of the *Trypanosoma brucei* editosome. *Mol Cell* **11**, 1525-1536.
- Fairman, M.E., Maroney, P.A., Wang, W., Bowers, H.A., Gollnick, P., Nilsen, T.W., and Jankowsky, E.** (2004). Protein displacement by DExH/D "RNA helicases" without duplex unwinding. *Science* **304**, 730-734.
- Faivre-Nitschke, S.E., Grienenberger, J.M., and Gualberto, J.M.** (1999). A prokaryotic-type cytidine deaminase from *Arabidopsis thaliana* gene expression and functional characterization. *Eur J Biochem* **263**, 896-903.
- Farre, J.C., and Araya, A.** (2001). Gene expression in isolated plant mitochondria: high fidelity of transcription, splicing and editing of a transgene product in electroporated organelles. *Nucleic Acids Res* **29**, 2484-2491.
- Farre, J.C., Leon, G., Jordana, X., and Araya, A.** (2001). cis Recognition elements in plant mitochondrion RNA editing. *Mol Cell Biol* **21**, 6731-6737.
- Fisk, D.G., Walker, M.B., and Barkan, A.** (1999). Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. *Embo J* **18**, 2621-2630.
- Fitzgerald, L.W., Iyer, G., Conklin, D.S., Krause, C.M., Marshall, A., Patterson, J.P., Tran, D.P., Jonak, G.J., and Hartig, P.R.** (1999). Messenger RNA editing of the human serotonin 5-HT<sub>2C</sub> receptor. *Neuropsychopharmacology* **21**, 82S-90S.
- Flomen, R., Knight, J., Sham, P., Kerwin, R., and Makoff, A.** (2004). Evidence that RNA editing modulates splice site selection in the 5-HT<sub>2C</sub> receptor gene. *Nucleic Acids Res* **32**, 2113-2122.
- Foissac, S., Bardou, P., Moisan, A., Cros, M.J., and Schiex, T.** (2003). EUGENE'HOM: A generic similarity-based gene finder using multiple homologous sequences. *Nucleic Acids Res* **31**, 3742-3745.
- Forrest, L.L., and Crandall-Stotler, B.J.** (2005). Progress towards a robust phylogeny for the liverworts, with particular focus on the simple thalloid. *J. Hattori Bot. Lab.* **97**, 127-159.
- Freyer, R., Lopez, C., Maier, R.M., Martin, M., Sabater, B., and Kossel, H.** (1995). Editing of the chloroplast *ndhB* encoded transcript shows divergence between closely related members of the grass family (Poaceae). *Plant Mol Biol* **29**, 679-684.
- Gallo, A., Keegan, L.P., Ring, G.M., and O'Connell, M.A.** (2003). An ADAR that edits transcripts encoding ion channel subunits functions as a dimer. *Embo J* **22**, 3421-3430.
- Gao, G., and Simpson, L.** (2003). Is the *Trypanosoma brucei* REL1 RNA ligase specific for U-deletion RNA editing, and is the REL2 RNA ligase specific for U-insertion editing? *J Biol Chem* **278**, 27570-27574.

- Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.** (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, J.M. Walker, ed (Humana Press), pp. 571-607
- Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G.** (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147.
- George, C.X., and Samuel, C.E.** (1999a). Characterization of the 5'-flanking region of the human RNA-specific adenosine deaminase ADAR1 gene and identification of an interferon-inducible ADAR1 promoter. *Gene* **229**, 203-213.
- George, C.X., and Samuel, C.E.** (1999b). Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible. *Proc Natl Acad Sci U S A* **96**, 4621-4626.
- Gerber, A., Grosjean, H., Melcher, T., and Keller, W.** (1998). Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *Embo J* **17**, 4780-4789.
- Gerber, A.P., and Keller, W.** (1999). An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **286**, 1146-1149.
- Gerber, A.P., and Keller, W.** (2001). RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem Sci* **26**, 376-384.
- Giege, P., and Brennicke, A.** (1999). RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A* **96**, 15324-15329.
- Golden, D.E., and Hajduk, S.L.** (2006). The importance of RNA structure in RNA editing and a potential proofreading mechanism for correct guide RNA:pre-mRNA binary complex formation. *J Mol Biol* **359**, 585-596.
- Gray, J.C., Sullivan, J.A., Wang, J.H., Jerome, C.A., and MacLean, D.** (2003). Coordination of plastid and nuclear gene expression. *Philos Trans R Soc Lond B Biol Sci* **358**, 135-144; discussion 144-135.
- Greeve, J., Navaratnam, N., and Scott, J.** (1991). Characterization of the apolipoprotein B mRNA editing enzyme: no similarity to the proposed mechanism of RNA editing in kinetoplastid protozoa. *Nucleic Acids Res* **19**, 3569-3576.
- Greeve, J., Altkemper, I., Dieterich, J.H., Greten, H., and Windler, E.** (1993). Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins. *J Lipid Res* **34**, 1367-1383.
- Greger, I.H., Khatri, L., and Ziff, E.B.** (2002). RNA editing at arg607 controls AMPA receptor exit from the endoplasmic reticulum. *Neuron* **34**, 759-772.
- Greger, I.H., Khatri, L., Kong, X., and Ziff, E.B.** (2003). AMPA receptor tetramerization is mediated by Q/R editing. *Neuron* **40**, 763-774.
- Grohmann, L., Thieck, O., Herz, U., Schroder, W., and Brennicke, A.** (1994). Translation of nad9 mRNAs in mitochondria from *Solanum tuberosum* is restricted to completely edited transcripts. *Nucleic Acids Res* **22**, 3304-3311.
- Grosjean, H., Sprinzl, M., and Steinberg, S.** (1995). Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie* **77**, 139-141.

- Grosjean, H., Auxilien, S., Constantinesco, F., Simon, C., Corda, Y., Becker, H.F., Foiret, D., Morin, A., Jin, Y.X., Fournier, M., and Fourrey, J.L.** (1996). Enzymatic conversion of adenosine to inosine and to N1-methylinosine in transfer RNAs: a review. *Biochimie* **78**, 488-501.
- Groth-Malonek, M., Pruchner, D., Grewe, F., and Knoop, V.** (2005). Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol Biol Evol* **22**, 117-125.
- Groth-Malonek, M., Wahrmond, U., Polsakiewicz, M., and Knoop, V.** (2007). Evolution of a pseudogene: exclusive survival of a functional mitochondrial nad7 gene supports Haplomitrium as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. *Mol Biol Evol* **24**, 1068-1074.
- Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H., and Grienenberger, J.M.** (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* **341**, 660-662.
- Gurevich, I., Tamir, H., Arango, V., Dwork, A.J., Mann, J.J., and Schmauss, C.** (2002). Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. *Neuron* **34**, 349-356.
- Hajjar, A.M., and Linial, M.L.** (1995). Modification of retroviral RNA by double-stranded RNA adenosine deaminase. *J Virol* **69**, 5878-5882.
- Handa, H.** (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res* **31**, 5907-5916.
- Harris, S.G., Sabio, I., Mayer, E., Steinberg, M.F., Backus, J.W., Sparks, J.D., Sparks, C.E., and Smith, H.C.** (1993). Extract-specific heterogeneity in high-order complexes containing apolipoprotein B mRNA editing activity and RNA-binding proteins. *J Biol Chem* **268**, 7382-7392.
- Hashimoto, M., Endo, T., Peltier, G., Tasaka, M., and Shikanai, T.** (2003). A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in *Arabidopsis*. *Plant J* **36**, 541-549.
- Hayes, M.L., and Hanson, M.R.** (2007). Assay of editing of exogenous RNAs in chloroplast extracts of *Arabidopsis*, maize, pea, and tobacco. *Methods Enzymol* **424**, 459-482.
- Hayman, M.L., and Read, L.K.** (1999). Trypanosoma brucei RBP16 is a mitochondrial Y-box family protein with guide RNA binding activity. *J Biol Chem* **274**, 12067-12074.
- Hegeman, C.E., Hayes, M.L., and Hanson, M.R.** (2005). Substrate and cofactor requirements for RNA editing of chloroplast transcripts in *Arabidopsis* in vitro. *Plant J* **42**, 124-132.
- Heller, W.P., Hayes, M.L., and Hanson, M.R.** (2008). Cross-competition in editing of chloroplast RNA transcripts in vitro implicates sharing of trans-factors between different C targets. *J Biol Chem* **283**, 7314-7319.
- Helliwell, C., and Waterhouse, P.** (2003). Constructs and methods for high-throughput gene silencing in plants. *Methods* **30**, 289-295.
- Hernould, M., Suharsono, S., Litvak, S., Araya, A., and Mouras, A.** (1993). Male-sterility induction in transgenic tobacco plants with an unedited atp9 mitochondrial gene from wheat. *Proc Natl Acad Sci U S A* **90**, 2370-2374.
- Herrick-Davis, K., Grinde, E., and Niswender, C.M.** (1999). Serotonin 5-HT<sub>2C</sub> receptor RNA editing alters receptor basal activity: implications for serotonergic signal transduction. *J Neurochem* **73**, 1711-1717.

- Hersberger, M., and Innerarity, T.L.** (1998). Two efficiency elements flanking the editing site of cytidine 6666 in the apolipoprotein B mRNA support mooring-dependent editing. *J Biol Chem* **273**, 9435-9442.
- Hersberger, M., Patarroyo-White, S., Arnold, K.S., and Innerarity, T.L.** (1999). Phylogenetic analysis of the apolipoprotein B mRNA-editing region. Evidence for a secondary structure between the mooring sequence and the 3' efficiency element. *J Biol Chem* **274**, 34590-34597.
- Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A.** (1989). RNA editing in plant mitochondria. *Science* **246**, 1632-1634.
- Higuchi, M., Single, F.N., Kohler, M., Sommer, B., Sprengel, R., and Seeburg, P.H.** (1993). RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* **75**, 1361-1370.
- Higuchi, M., Maas, S., Single, F.N., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R., and Seeburg, P.H.** (2000). Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406**, 78-81.
- Hilson, P., Allemeersch, J., Altmann, T., Aubourg, S., Avon, A., Beynon, J., Bhalerao, R.P., Bitton, F., Caboche, M., Cannoot, B., Chardakov, V., Cognet-Holliger, C., Colot, V., Crowe, M., Darimont, C., Durinck, S., Eickhoff, H., de Longevialle, A.F., Farmer, E.E., Grant, M., Kuiper, M.T., Lehrach, H., Leon, C., Leyva, A., Lundeberg, J., Lurin, C., Moreau, Y., Nietfeld, W., Paz-Ares, J., Reymond, P., Rouze, P., Sandberg, G., Segura, M.D., Serizet, C., Tabrett, A., Tacconnat, L., Thareau, V., Van Hummelen, P., Vercruyse, S., Vuylsteke, M., Weingartner, M., Weisbeek, P.J., Wirta, V., Wittink, F.R., Zabeau, M., and Small, I.** (2004). Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. *Genome Res* **14**, 2176-2189.
- Hirose, T., and Sugiura, M.** (1997). Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast *ndhD* mRNA: a possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes. *Embo J* **16**, 6804-6811.
- Hirose, T., and Sugiura, M.** (2001). Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast in vitro RNA editing system. *Embo J* **20**, 1144-1152.
- Ho, C.K., Wang, L.K., Lima, C.D., and Shuman, S.** (2004). Structure and mechanism of RNA ligase. *Structure* **12**, 327-339.
- Hoch, B., Maier, R.M., Appel, K., Igloi, G.L., and Kossel, H.** (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* **353**, 178-180.
- Hoffmann, M., Nemetz, C., Madin, K., and Buchberger, B.** (2004). Rapid translation system: a novel cell-free way from gene to protein. *Biotechnol Annu Rev* **10**, 1-30.
- Holley, R.W., Everett, G.A., Madison, J.T., and Zamir, A.** (1965). Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid. *J Biol Chem* **240**, 2122-2128.
- Hoopengardner, B., Bhalla, T., Staber, C., and Reenan, R.** (2003). Nervous system targets of RNA editing identified by comparative genomics. *Science* **301**, 832-836.
- Horton, T.L., and Landweber, L.F.** (2000). Evolution of four types of RNA editing in myxomycetes. *Rna* **6**, 1339-1346.
- Horton, T.L., and Landweber, L.F.** (2002). Rewriting the information in DNA: RNA editing in kinetoplasts and myxomycetes. *Curr Opin Microbiol* **5**, 620-626.
- Hough, R.F., Lingam, A.T., and Bass, B.L.** (1999). *Caenorhabditis elegans* mRNAs that encode a protein similar to ADARs derive from an operon containing six genes. *Nucleic Acids Res* **27**, 3424-3432.



- Howell, M.D., Fahlgren, N., Chapman, E.J., Cumbie, J.S., Sullivan, C.M., Givan, S.A., Kasschau, K.D., and Carrington, J.C.** (2007). Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* **19**, 926-942.
- Huang, C.E., O'Hearn, S.F., and Sollner-Webb, B.** (2002). Assembly and function of the RNA editing complex in *Trypanosoma brucei* requires band III protein. *Mol Cell Biol* **22**, 3194-3203.
- Huang, C.E., Cruz-Reyes, J., Zhelonkina, A.G., O'Hearn, S., Wirtz, E., and Sollner-Webb, B.** (2001). Roles for ligases in the RNA editing complex of *Trypanosoma brucei*: band IV is needed for U-deletion and RNA repair. *Embo J* **20**, 4694-4703.
- Hughes, S.D., Rouy, D., Navaratnam, N., Scott, J., and Rubin, E.M.** (1996). Gene transfer of cytidine deaminase apoBEC-1 lowers lipoprotein(a) in transgenic mice and induces apolipoprotein B editing in rabbits. *Hum Gene Ther* **7**, 39-49.
- Hui-Rong Qian, B.J.N., Eve Syrkin Wurtele** (1998). Molecular genetics of acetyl-CoA carboxylase. In 9TH INTERNATIONAL CONFERENCE ON ARABIDOPSIS RESEARCH.
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J.** (2006). The PROSITE database. *Nucleic Acids Res* **34**, D227-230.
- Igo, R.P., Jr., Palazzo, S.S., Burgess, M.L., Panigrahi, A.K., and Stuart, K.** (2000). Uridylate addition and RNA ligation contribute to the specificity of kinetoplastid insertion RNA editing. *Mol Cell Biol* **20**, 8447-8457.
- Igo, R.P., Jr., Weston, D.S., Ernst, N.L., Panigrahi, A.K., Salavati, R., and Stuart, K.** (2002). Role of uridylate-specific exoribonuclease activity in *Trypanosoma brucei* RNA editing. *Eukaryot Cell* **1**, 112-118.
- Ikeda, T.M., and Gray, M.W.** (1999). Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. *Mol Cell Biol* **19**, 8113-8122.
- Inada, H., Seki, M., Morikawa, H., Nishimura, M., and Iba, K.** (1997). Existence of three regulatory regions each containing a highly conserved motif in the promoter of plastid-encoded RNA polymerase gene (*rpoB*). *Plant J* **11**, 883-890.
- Ireton, G.C., McDermott, G., Black, M.E., and Stoddard, B.L.** (2002). The structure of *Escherichia coli* cytosine deaminase. *J Mol Biol* **315**, 687-697.
- Jasinska, A., and Krzyzosiak, W.J.** (2004). Repetitive sequences that shape the human transcriptome. *FEBS Lett* **567**, 136-141.
- Johansson, E., Mejlhede, N., Neuhard, J., and Larsen, S.** (2002). Crystal structure of the tetrameric cytidine deaminase from *Bacillus subtilis* at 2.0 Å resolution. *Biochemistry* **41**, 2563-2570.
- Kable, M.L., Seiwert, S.D., Heidmann, S., and Stuart, K.** (1996). RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA. *Science* **273**, 1189-1195.
- Kallman, A.M., Sahlin, M., and Ohman, M.** (2003). ADAR2 A-->I editing: site selectivity and editing efficiency are separate events. *Nucleic Acids Res* **31**, 4874-4881.
- Kao, C.Y., and Read, L.K.** (2007). Targeted depletion of a mitochondrial nucleotidyltransferase suggests the presence of multiple enzymes that polymerize mRNA 3' tails in *Trypanosoma brucei* mitochondria. *Mol Biochem Parasitol* **154**, 158-169.
- Kawahara, Y., Ito, K., Sun, H., Ito, M., Kanazawa, I., and Kwak, S.** (2004). Regulation of glutamate receptor RNA editing and ADAR mRNA expression in developing human normal and Down's syndrome brains. *Brain Res Dev Brain Res* **148**, 151-155.

- Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., and Nishikura, K.** (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* **315**, 1137-1140.
- Kawahara, Y., Kwak, S., Sun, H., Ito, K., Hashida, H., Aizawa, H., Jeong, S.Y., and Kanazawa, I.** (2003). Human spinal motoneurons express low relative abundance of GluR2 mRNA: an implication for excitotoxicity in ALS. *J Neurochem* **85**, 680-689.
- Kawakubo, K., and Samuel, C.E.** (2000). Human RNA-specific adenosine deaminase (ADAR1) gene specifies transcripts that initiate from a constitutively active alternative promoter. *Gene* **258**, 165-172.
- Kazama, T., and Toriyama, K.** (2003). A pentatricopeptide repeat-containing gene that promotes the processing of aberrant atp6 RNA of cytoplasmic male-sterile rice. *FEBS Lett* **544**, 99-102.
- Kikuno, R., Nagase, T., Waki, M., and Ohara, O.** (2002). HUGE: a database for human large proteins identified in the Kazusa cDNA sequencing project. *Nucleic Acids Res* **30**, 166-168.
- Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matisse, T.C., Buyske, S., and Gabriel, A.** (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res* **14**, 1719-1725.
- Kim, J., Malashkevich, V., Roday, S., Lisbin, M., Schramm, V.L., and Almo, S.C.** (2006). Structural and kinetic characterization of Escherichia coli TadA, the wobble-specific tRNA deaminase. *Biochemistry* **45**, 6407-6416.
- Kim, U., Wang, Y., Sanford, T., Zeng, Y., and Nishikura, K.** (1994). Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc Natl Acad Sci U S A* **91**, 11457-11461.
- Kinnaird, J.H., Maitland, K., Walker, G.A., Wheatley, I., Thompson, F.J., and Devaney, E.** (2004). HRP-2, a heterogeneous nuclear ribonucleoprotein, is essential for embryogenesis and oogenesis in *Caenorhabditis elegans*. *Exp Cell Res* **298**, 418-430.
- Klein, R.R., Klein, P.E., Mullet, J.E., Minx, P., Rooney, W.L., and Schertz, K.F.** (2005). Fertility restorer locus Rf1 [corrected] of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor Appl Genet* **111**, 994-1012.
- Knoop, V.** (2004). The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet* **46**, 123-139.
- Ko, T.P., Lin, J.J., Hu, C.Y., Hsu, Y.H., Wang, A.H., and Liaw, S.H.** (2003). Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. *J Biol Chem* **278**, 19111-19117.
- Kobayashi, K., Suzuki, M., Tang, J., Nagata, N., Ohyama, K., Seki, H., Kiuchi, R., Kaneko, Y., Nakazawa, M., Matsui, M., Matsumoto, S., Yoshida, S., and Muranaka, T.** (2007). Lovastatin insensitive 1, a Novel pentatricopeptide repeat protein, is a potential regulatory factor of isoprenoid biosynthesis in *Arabidopsis*. *Plant Cell Physiol* **48**, 322-331.
- Kobayashi, Y., Matsuo, M., Sakamoto, K., Wakasugi, T., Yamada, K., and Obokata, J.** (2008). Two RNA editing sites with cis-acting elements of moderate sequence identity are recognized by an identical site-recognition protein in tobacco chloroplasts. *Nucleic Acids Res* **36**, 311-318.
- Koc, E.C., and Spremulli, L.L.** (2003). RNA-binding proteins of mammalian mitochondria. *Mitochondrion* **2**, 277-291.
- Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T., Kawasaki, S., and Imamura, J.** (2003). Genetic characterization of a

- pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. *Plant J* **34**, 407-415.
- Kolakofsky, D., Roux, L., Garcin, D., and Ruigrok, R.W.** (2005). Paramyxovirus mRNA editing, the "rule of six" and error catastrophe: a hypothesis. *J Gen Virol* **86**, 1869-1877.
- Koller, J., Muller, U.F., Schmid, B., Missel, A., Krufft, V., Stuart, K., and Goring, H.U.** (1997). Trypanosoma brucei gBP21. An arginine-rich mitochondrial protein that binds to guide RNA with high affinity. *J Biol Chem* **272**, 3749-3757.
- Komori, T., Ohta, S., Murai, N., Takakura, Y., Kuraya, Y., Suzuki, S., Hiei, Y., Imaseki, H., and Nitta, N.** (2004). Map-based cloning of a fertility restorer gene, Rf-1, in rice (*Oryza sativa* L.). *Plant J* **37**, 315-325.
- Kos, A., Dijkema, R., Arnberg, A.C., van der Meide, P.H., and Schellekens, H.** (1986). The hepatitis delta (delta) virus possesses a circular RNA. *Nature* **323**, 558-560.
- Koslowsky, D.J., Bhat, G.J., Read, L.K., and Stuart, K.** (1991). Cycles of progressive realignment of gRNA with mRNA in RNA editing. *Cell* **67**, 537-546.
- Koslowsky, D.J., Reifur, L., Yu, L.E., and Chen, W.** (2004). Evidence for U-tail stabilization of gRNA/mRNA interactions in kinetoplastid RNA editing. *RNA Biol* **1**, 28-34.
- Kotera, E., Tasaka, M., and Shikanai, T.** (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**, 326-330.
- Koussevitzky, S., Nott, A., Mockler, T.C., Hong, F., Sachetto-Martins, G., Surpin, M., Lim, J., Mittler, R., and Chory, J.** (2007). Signals from chloroplasts converge to regulate nuclear gene expression. *Science* **316**, 715-719.
- Krause, K., Lopes de Souza, R., Roberts, D.G., and Dieckmann, C.L.** (2004). The mitochondrial message-specific mRNA protectors Cbp1 and Pet309 are associated in a high-molecular weight complex. *Mol Biol Cell* **15**, 2674-2683.
- Kubo, N., and Kadowaki, K.** (1997). Involvement of 5' flanking sequence for specifying RNA editing sites in plant mitochondria. *FEBS Lett* **413**, 40-44.
- Kumar, M., and Carmichael, G.G.** (1997). Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proc Natl Acad Sci U S A* **94**, 3542-3547.
- Kuryatov, A., Gerzanich, V., Nelson, M., Olale, F., and Lindstrom, J.** (1997). Mutation causing autosomal dominant nocturnal frontal lobe epilepsy alters Ca<sup>2+</sup> permeability, conductance, and gating of human alpha4beta2 nicotinic acetylcholine receptors. *J Neurosci* **17**, 9035-9047.
- Lai, F., Drakas, R., and Nishikura, K.** (1995). Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts. *J Biol Chem* **270**, 17098-17105.
- Lambert, L., Muller, U.F., Souza, A.E., and Goring, H.U.** (1999). The involvement of gRNA-binding protein gBP21 in RNA editing-an in vitro and in vivo analysis. *Nucleic Acids Res* **27**, 1429-1436.
- Langlois, M.A., Beale, R.C., Conticello, S.G., and Neuberger, M.S.** (2005). Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res* **33**, 1913-1923.
- Lau, P.P., Zhu, H.J., Nakamuta, M., and Chan, L.** (1997). Cloning of an Apobec-1-binding protein that also interacts with apolipoprotein B mRNA and evidence for its involvement in RNA editing. *J Biol Chem* **272**, 1452-1455.

- Lau, P.P., Zhu, H.J., Baldini, A., Charnsangavej, C., and Chan, L.** (1994). Dimeric structure of a human apolipoprotein B mRNA editing protein and cloning and chromosomal localization of its gene. *Proc Natl Acad Sci U S A* **91**, 8522-8526.
- Law, J.A., O'Hearn, S.F., and Sollner-Webb, B.** (2008). Trypanosoma brucei RNA editing protein TbMP42 (band VI) is crucial for the endonucleolytic cleavages but not the subsequent steps of U-deletion and U-insertion. *Rna* **14**, 1187-1200.
- Lawson, S.D., Igo, R.P., Jr., Salavati, R., and Stuart, K.D.** (2001). The specificity of nucleotide removal during RNA editing in Trypanosoma brucei. *Rna* **7**, 1793-1802.
- Lehmann, K.A., and Bass, B.L.** (1999). The importance of internal loops within RNA substrates of ADAR1. *J Mol Biol* **291**, 1-13.
- Lehmann, K.A., and Bass, B.L.** (2000). Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875-12884.
- Lellek, H., Welker, S., Diehl, I., Kirsten, R., and Greeve, J.** (2002). Reconstitution of mRNA editing in yeast using a Gal4-apoB-Gal80 fusion transcript as the selectable marker. *J Biol Chem* **277**, 23638-23644.
- Lellek, H., Kirsten, R., Diehl, I., Apostel, F., Buck, F., and Greeve, J.** (2000). Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex. *J Biol Chem* **275**, 19848-19856.
- Leung, S.S., and Koslowsky, D.J.** (2001). Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of an mRNA bound to its cognate gRNA. *Rna* **7**, 1803-1816.
- Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D., Olshansky, M., Rechavi, G., and Jantsch, M.F.** (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* **22**, 1001-1005.
- Liang, H., and Landweber, L.F.** (2007). Hypothesis: RNA editing of microRNA target sites in humans? *Rna* **13**, 463-467.
- Lin, S., Zhang, H., Spencer, D.F., Norman, J.E., and Gray, M.W.** (2002). Widespread and extensive editing of mitochondrial mRNAs in dinoflagellates. *J Mol Biol* **320**, 727-739.
- Lippok, B., Brennicke, A., and Wissinger, B.** (1994). Differential RNA editing in closely related introns in Oenothera mitochondria. *Mol Gen Genet* **243**, 39-46.
- Liu, Y., and Samuel, C.E.** (1996). Mechanism of interferon action: functionally distinct RNA-binding and catalytic domains in the interferon-inducible, double-stranded RNA-specific adenosine deaminase. *J Virol* **70**, 1961-1968.
- Lobel, L.I., Pollak, S., Klein, J., and Lustbader, J.W.** (2001). High-level bacterial expression of a natively folded, soluble extracellular domain fusion protein of the human luteinizing hormone/chorionic gonadotropin receptor in the cytoplasm of Escherichia coli. *Endocrine* **14**, 205-212.
- Lomeli, H., Mosbacher, J., Melcher, T., Hoger, T., Geiger, J.R., Kuner, T., Monyer, H., Higuchi, M., Bach, A., and Seeburg, P.H.** (1994). Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **266**, 1709-1713.
- Lopez, P.J., Marchand, I., Joyce, S.A., and Dreyfus, M.** (1999). The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Mol Microbiol* **33**, 188-199.
- Losey, H.C., Ruthenburg, A.J., and Verdine, G.L.** (2006). Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* **13**, 153-159.

- Lown, F.J., Watson, A.T., and Purton, S.** (2001). Chlamydomonas nuclear mutants that fail to assemble respiratory or photosynthetic electron transfer complexes. *Biochem Soc Trans* **29**, 452-455.
- Lu, B., and Hanson, M.R.** (1994). A single homogeneous form of ATP6 protein accumulates in petunia mitochondria despite the presence of differentially edited atp6 transcripts. *Plant Cell* **6**, 1955-1968.
- Luciano, D.J., Mirsky, H., Vendetti, N.J., and Maas, S.** (2004). RNA editing of a miRNA precursor. *Rna* **10**, 1174-1177.
- Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L., and Small, I.** (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089-2103.
- Maas, S., Gerber, A.P., and Rich, A.** (1999). Identification and characterization of a human tRNA-specific adenosine deaminase related to the ADAR family of pre-mRNA editing enzymes. *Proc Natl Acad Sci U S A* **96**, 8895-8900.
- Maas, S., Kim, Y.G., and Rich, A.** (2000). Sequence, genomic organization and functional expression of the murine tRNA-specific adenosine deaminase ADAT1. *Gene* **243**, 59-66.
- Maas, S., Rich, A., and Nishikura, K.** (2003). A-to-I RNA editing: recent news and residual mysteries. *J Biol Chem* **278**, 1391-1394.
- Maas, S., Patt, S., Schrey, M., and Rich, A.** (2001). Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. *Proc Natl Acad Sci U S A* **98**, 14687-14692.
- MacGinnitie, A.J., Anant, S., and Davidson, N.O.** (1995). Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. *J Biol Chem* **270**, 14768-14775.
- Madej, M.J., Niemann, M., Huttenhofer, A., and Goringer, H.U.** (2008). Identification of novel guide RNAs from the mitochondria of *Trypanosoma brucei*. *RNA Biol* **5**.
- Madison-Antenucci, S., Grams, J., and Hajduk, S.L.** (2002). Editing machines: the complexities of trypanosome RNA editing. *Cell* **108**, 435-438.
- Madison-Antenucci, S., Sabatini, R.S., Pollard, V.W., and Hajduk, S.L.** (1998). Kinetoplastid RNA-editing-associated protein 1 (REAP-1): a novel editing complex protein with repetitive domains. *Embo J* **17**, 6368-6376.
- Malek, O., Lattig, K., Hiesel, R., Brennicke, A., and Knoop, V.** (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. *Embo J* **15**, 1403-1411.
- Mancebo, R., Zhou, X., Shillinglaw, W., Henzel, W., and Macdonald, P.M.** (2001). BSF binds specifically to the bicoid mRNA 3' untranslated region and contributes to stabilization of bicoid mRNA. *Mol Cell Biol* **21**, 3462-3471.
- Manthey, G.M., and McEwen, J.E.** (1995). The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COX1 locus of *Saccharomyces cerevisiae*. *Embo J* **14**, 4031-4043.
- Manthey, G.M., Przybyla-Zawislak, B.D., and McEwen, J.E.** (1998). The *Saccharomyces cerevisiae* Pet309 protein is embedded in the mitochondrial inner membrane. *Eur J Biochem* **255**, 156-161.
- Marchfelder, A., and Binder, S.** (2004). Plastid and plant mitochondrial RNA processing and RNA stability. In *Molecular biology and biotechnology of plants organelles*, H. Daniell and C.D. Chase, eds (Springer), pp. 261-294.

- Maris, C., Masse, J., Chester, A., Navaratnam, N., and Allain, F.H.** (2005). NMR structure of the apoB mRNA stem-loop and its interaction with the C to U editing APOBEC1 complementary factor. *Rna* **11**, 173-186.
- Maylie, B., Bissonnette, E., Virk, M., Adelman, J.P., and Maylie, J.G.** (2002). Episodic ataxia type 1 mutations in the human Kv1.1 potassium channel alter hKvbeta 1-induced N-type inactivation. *J Neurosci* **22**, 4786-4793.
- Mehta, A., and Driscoll, D.M.** (1998). A sequence-specific RNA-binding protein complements apobec-1 To edit apolipoprotein B mRNA. *Mol Cell Biol* **18**, 4426-4432.
- Mehta, A., and Driscoll, D.M.** (2002). Identification of domains in apobec-1 complementation factor required for RNA binding and apolipoprotein-B mRNA editing. *Rna* **8**, 69-82.
- Mehta, A., Kinter, M.T., Sherman, N.E., and Driscoll, D.M.** (2000). Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Mol Cell Biol* **20**, 1846-1854.
- Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G.** (2003). HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* **15**, 1480-1495.
- Melcher, T., Maas, S., Herb, A., Sprengel, R., Higuchi, M., and Seeburg, P.H.** (1996). RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J Biol Chem* **271**, 31795-31798.
- Mili, S., and Pinol-Roma, S.** (2003). LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. *Mol Cell Biol* **23**, 4972-4982.
- Miller, M.L., Antes, T.J., Qian, F., and Miller, D.L.** (2006a). Identification of a putative mitochondrial RNA polymerase from *Physarum polycephalum*: characterization, expression, purification, and transcription in vitro. *Curr Genet* **49**, 259-271.
- Miller, M.M., Halbig, K., Cruz-Reyes, J., and Read, L.K.** (2006b). RBP16 stimulates trypanosome RNA editing in vitro at an early step in the editing reaction. *Rna* **12**, 1292-1303.
- Miroux, B., and Walker, J.E.** (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289-298.
- Missel, A., Souza, A.E., Norskau, G., and Goringer, H.U.** (1997). Disruption of a gene encoding a novel mitochondrial DEAD-box protein in *Trypanosoma brucei* affects edited mRNAs. *Mol Cell Biol* **17**, 4895-4903.
- Miyamoto, T., Obokata, J., and Sugiura, M.** (2002). Recognition of RNA editing sites is directed by unique proteins in chloroplasts: biochemical identification of cis-acting elements and trans-acting factors involved in RNA editing in tobacco and pea chloroplasts. *Mol Cell Biol* **22**, 6726-6734.
- Miyamoto, T., Obokata, J., and Sugiura, M.** (2004). A site-specific factor interacts directly with its cognate RNA editing site in chloroplast transcripts. *Proc Natl Acad Sci U S A* **101**, 48-52.
- Miyamura, Y., Suzuki, T., Kono, M., Inagaki, K., Ito, S., Suzuki, N., and Tomita, Y.** (2003). Mutations of the RNA-specific adenosine deaminase gene (DSRAD) are involved in dyschromatosis symmetrica hereditaria. *Am J Hum Genet* **73**, 693-699.
- Miyata, Y., and Sugita, M.** (2004). Tissue- and stage-specific RNA editing of rps 14 transcripts in moss (*Physcomitrella patens*) chloroplasts. *J Plant Physiol* **161**, 113-115.

- Miyata, Y., Sugita, C., Maruyama, K., and Sugita, M.** (2008). RNA editing in the anticodon of tRNA Leu (CAA) occurs before group I intron splicing in plastids of a moss *Takakia lepidozioides* S. Hatt. & Inoue. *Plant Biol (Stuttg)* **10**, 250-255.
- Miyata, Y., Sugiura, C., Kobayashi, Y., Hagiwara, M., and Sugita, M.** (2002). Chloroplast ribosomal S14 protein transcript is edited to create a translation initiation codon in the moss *Physcomitrella patens*. *Biochim Biophys Acta* **1576**, 346-349.
- Mootha, V.K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F., Mitchell, G.A., Morin, C., Mann, M., Hudson, T.J., Robinson, B., Rioux, J.D., and Lander, E.S.** (2003). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A* **100**, 605-610.
- Morrison, J.R., Paszty, C., Stevens, M.E., Hughes, S.D., Forte, T., Scott, J., and Rubin, E.M.** (1996). Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. *Proc Natl Acad Sci U S A* **93**, 7154-7159.
- Morse, D.P., and Bass, B.L.** (1999). Long RNA hairpins that contain inosine are present in *Caenorhabditis elegans* poly(A)<sup>+</sup> RNA. *Proc Natl Acad Sci U S A* **96**, 6048-6053.
- Mukopadhyay, S., Anant, S., Lee, R., and Davidson, N.O.** (2000). In Digestive disease week (San Diego).
- Muligan, M.** (2004). RNA editing in plant organelles In *Molecular biology and biotechnology of plant organelles*, H. Daniell and C.D. Chase, eds (Springer), pp. 239-260.
- Naithani, S., Saracco, S.A., Butler, C.A., and Fox, T.D.** (2003). Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of *Saccharomyces cerevisiae*. *Mol Biol Cell* **14**, 324-333.
- Nakai, K., and Horton, P.** (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**, 34-36.
- Nakamura, T., Meierhoff, K., Westhoff, P., and Schuster, G.** (2003). RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA. *Eur J Biochem* **270**, 4070-4081.
- Nakamuta, M., Tsai, A., Chan, L., Davidson, N.O., and Teng, B.B.** (1999). Sequence elements required for apolipoprotein B mRNA editing enhancement activity from chicken enterocytes. *Biochem Biophys Res Commun* **254**, 744-750.
- Nakamuta, M., Chang, B.H., Zsigmond, E., Kobayashi, K., Lei, H., Ishida, B.Y., Oka, K., Li, E., and Chan, L.** (1996). Complete phenotypic characterization of apobec-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of Apobec-1. *J Biol Chem* **271**, 25981-25988.
- Narsai, R., Howell, K.A., Millar, A.H., O'Toole, N., Small, I., and Whelan, J.** (2007). Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell* **19**, 3418-3436.
- Navaratnam, N., Bhattacharya, S., Fujino, T., Patel, D., Jarmuz, A.L., and Scott, J.** (1995). Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site. *Cell* **81**, 187-195.
- Navaratnam, N., Fujino, T., Bayliss, J., Jarmuz, A., How, A., Richardson, N., Somasekaram, A., Bhattacharya, S., Carter, C., and Scott, J.** (1998). *Escherichia coli* cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition. *J Mol Biol* **275**, 695-714.

- Neuwirt, J., Takenaka, M., van der Merwe, J.A., and Brennicke, A.** (2005). An in vitro RNA editing system from cauliflower mitochondria: editing site recognition parameters can vary in different plant species. *Rna* **11**, 1563-1570.
- Nishikura, K., Yoo, C., Kim, U., Murray, J.M., Estes, P.A., Cash, F.E., and Liebhaber, S.A.** (1991). Substrate specificity of the dsRNA unwinding/modifying activity. *Embo J* **10**, 3523-3532.
- Niswender, C.M., Sanders-Bush, E., and Emeson, R.B.** (1998). Identification and characterization of RNA editing events within the 5-HT<sub>2C</sub> receptor. *Ann N Y Acad Sci* **861**, 38-48.
- Niswender, C.M., Copeland, S.C., Herrick-Davis, K., Emeson, R.B., and Sanders-Bush, E.** (1999). RNA editing of the human serotonin 5-hydroxytryptamine 2C receptor silences constitutive activity. *J Biol Chem* **274**, 9472-9478.
- Niswender, C.M., Herrick-Davis, K., Dilley, G.E., Meltzer, H.Y., Overholser, J.C., Stockmeier, C.A., Emeson, R.B., and Sanders-Bush, E.** (2001). RNA editing of the human serotonin 5-HT<sub>2C</sub> receptor. alterations in suicide and implications for serotonergic pharmacotherapy. *Neuropsychopharmacology* **24**, 478-491.
- Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K.** (2002). The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics* **268**, 434-445.
- O'Connell, M.A., Krause, S., Higuchi, M., Hsuan, J.J., Totty, N.F., Jenny, A., and Keller, W.** (1995). Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase. *Mol Cell Biol* **15**, 1389-1397.
- O'Hara, P.J., Horodyski, F.M., Nichol, S.T., and Holland, J.J.** (1984). Vesicular stomatitis virus mutants resistant to defective-interfering particles accumulate stable 5'-terminal and fewer 3'-terminal mutations in a stepwise manner. *J Virol* **49**, 793-798.
- O'Hearn, S.F., Huang, C.E., Hemann, M., Zhelonkina, A., and Sollner-Webb, B.** (2003). *Trypanosoma brucei* RNA editing complex: band II is structurally critical and maintains band V ligase, which is nonessential. *Mol Cell Biol* **23**, 7909-7919.
- O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., and Small, I.** (2008). On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* **25**, 1120-1128.
- Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., and et al.** (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol* **223**, 1-7.
- Ohman, M.** (2007). A-to-I editing challenger or ally to the microRNA process. *Biochimie* **89**, 1171-1176.
- Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T.** (2006). A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* **281**, 37661-37667.
- Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T.** (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A* **104**, 8178-8183.
- Opi, S., Takeuchi, H., Kao, S., Khan, M.A., Miyagi, E., Goila-Gaur, R., Iwatani, Y., Levin, J.G., and Strebel, K.** (2006). Monomeric APOBEC3G is catalytically active and has antiviral activity. *J Virol* **80**, 4673-4682.
- Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A.** (2000a). dADAR, a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *Rna* **6**, 1004-1018.



- Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A.** (2000b). A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity. *Cell* **102**, 437-449.
- Panigrahi, A.K., Schnauffer, A., Carmean, N., Igo, R.P., Jr., Gygi, S.P., Ernst, N.L., Palazzo, S.S., Weston, D.S., Aebersold, R., Salavati, R., and Stuart, K.D.** (2001). Four related proteins of the *Trypanosoma brucei* RNA editing complex. *Mol Cell Biol* **21**, 6833-6840.
- Patterson, J.B., and Samuel, C.E.** (1995). Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol Cell Biol* **15**, 5376-5388.
- Patton, D.E., Silva, T., and Bezanilla, F.** (1997). RNA editing generates a diverse array of transcripts encoding squid Kv2 K<sup>+</sup> channels with altered functional properties. *Neuron* **19**, 711-722.
- Paul, M.S., and Bass, B.L.** (1998). Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *Embo J* **17**, 1120-1127.
- Pelletier, M., and Read, L.K.** (2003). RBP16 is a multifunctional gene regulatory protein involved in editing and stabilization of specific mitochondrial mRNAs in *Trypanosoma brucei*. *Rna* **9**, 457-468.
- Pelletier, M., Miller, M.M., and Read, L.K.** (2000). RNA-binding properties of the mitochondrial Y-box protein RBP16. *Nucleic Acids Res* **28**, 1266-1275.
- Petersen-Mahrt, S.K., and Neuberger, M.S.** (2003). In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *J Biol Chem* **278**, 19583-19586.
- Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.J., and Oelmuller, R.** (2006). pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* **18**, 176-197.
- Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., Russo, J.J., Ju, J., Randall, G., Lindenbach, B.D., Rice, C.M., Simon, V., Ho, D.D., Zavolan, M., and Tuschl, T.** (2005). Identification of microRNAs of the herpesvirus family. *Nat Methods* **2**, 269-276.
- Phreaner, C.G., Williams, M.A., and Mulligan, R.M.** (1996). Incomplete editing of rps12 transcripts results in the synthesis of polymorphic polypeptides in plant mitochondria. *Plant Cell* **8**, 107-117.
- Piller, K.J., Rusche, L.N., Cruz-Reyes, J., and Sollner-Webb, B.** (1997). Resolution of the RNA editing gRNA-directed endonuclease from two other endonucleases of *Trypanosoma brucei* mitochondria. *Rna* **3**, 279-290.
- Polson, A.G., and Bass, B.L.** (1994). Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *Embo J* **13**, 5701-5711.
- Polson, A.G., Bass, B.L., and Casey, J.L.** (1996). RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature* **380**, 454-456.
- Polson, A.G., Ley, H.L., 3rd, Bass, B.L., and Casey, J.L.** (1998). Hepatitis delta virus RNA editing is highly specific for the amber/W site and is suppressed by hepatitis delta antigen. *Mol Cell Biol* **18**, 1919-1926.
- Polson, A.G., Crain, P.F., Pomerantz, S.C., McCloskey, J.A., and Bass, B.L.** (1991). The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. *Biochemistry* **30**, 11507-11514.

- Poulsen, H., Nilsson, J., Damgaard, C.K., Egebjerg, J., and Kjems, J.** (2001). CRM1 mediates the export of ADAR1 through a nuclear export signal within the Z-DNA binding domain. *Mol Cell Biol* **21**, 7862-7871.
- Powell, L.M., Wallis, S.C., Pease, R.J., Edwards, Y.H., Knott, T.J., and Scott, J.** (1987). A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50**, 831-840.
- Price, D.H., and Gray, M.W.** (1998). Editing of tRNA. In *Modification and Editing of RNA*, H. Grosjean and R. Benne, eds (AMS Press, Washington D.C.), pp. 289-306.
- Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F., and Chen, X.S.** (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* **445**, 447-451.
- Raitskin, O., Cho, D.S., Sperling, J., Nishikura, K., and Sperling, R.** (2001). RNA editing activity is associated with splicing factors in InRNP particles: The nuclear pre-mRNA processing machinery. *Proc Natl Acad Sci U S A* **98**, 6571-6576.
- Reenan, R.A., Hanrahan, C.J., and Barry, G.** (2000). The mle(naps) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na<sup>+</sup> channel transcript in a region of RNA editing. *Neuron* **25**, 139-149.
- Richardson, N., Navaratnam, N., and Scott, J.** (1998). Secondary structure for the apolipoprotein B mRNA editing site. Au-binding proteins interact with a stem loop. *J Biol Chem* **273**, 31707-31717.
- Rosenthal, J.J., and Bezanilla, F.** (2002). Extensive editing of mRNAs for the squid delayed rectifier K<sup>+</sup> channel regulates subunit tetramerization. *Neuron* **34**, 743-757.
- Rost, B., and Liu, J.** (2003). The PredictProtein server. *Nucleic Acids Res* **31**, 3300-3304.
- Rubio, M.A., Pastar, I., Gaston, K.W., Ragone, F.L., Janzen, C.J., Cross, G.A., Papavasiliou, F.N., and Alfonzo, J.D.** (2007). An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. *Proc Natl Acad Sci U S A* **104**, 7821-7826.
- Rueter, S.M., Dawson, T.R., and Emeson, R.B.** (1999). Regulation of alternative splicing by RNA editing. *Nature* **399**, 75-80.
- Rusche, L.N., Huang, C.E., Piller, K.J., Hemann, M., Wirtz, E., and Sollner-Webb, B.** (2001). The two RNA ligases of the *Trypanosoma brucei* RNA editing complex: cloning the essential band IV gene and identifying the band V gene. *Mol Cell Biol* **21**, 979-989.
- Saccomanno, L., and Bass, B.L.** (1999). A minor fraction of basic fibroblast growth factor mRNA is deaminated in *Xenopus* stage VI and matured oocytes. *Rna* **5**, 39-48.
- Salavati, R., Panigrahi, A.K., and Stuart, K.D.** (2001). Mitochondrial ribonuclease P activity of *Trypanosoma brucei*. *Mol Biochem Parasitol* **115**, 109-117.
- Salavati, R., Panigrahi, A.K., Morach, B.A., Palazzo, S.S., Igo, R.P., and Stuart, K.** (2002). Endoribonuclease activities of *Trypanosoma brucei* mitochondria. *Mol Biochem Parasitol* **120**, 23-31.
- Samson, F., Brunaud, V., Duchene, S., De Oliveira, Y., Caboche, M., Lecharny, A., and Aubourg, S.** (2004). FLAGdb<sup>++</sup>: a database for the functional analysis of the Arabidopsis genome. *Nucleic Acids Res* **32**, D347-350.
- Sansam, C.L., Wells, K.S., and Emeson, R.B.** (2003). Modulation of RNA editing by functional nucleolar sequestration of ADAR2. *Proc Natl Acad Sci U S A* **100**, 14018-14023.
- Sasaki, T., Yukawa, Y., Miyamoto, T., Obokata, J., and Sugiura, M.** (2003). Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for ndhB editing. *Mol Biol Evol* **20**, 1028-1035.

- Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H., and Nagano, Y.** (2001). Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *J Biol Chem* **276**, 3937-3940.
- Saunders, L.R., and Barber, G.N.** (2003). The dsRNA binding protein family: critical roles, diverse cellular functions. *Faseb J* **17**, 961-983.
- Scadden, A.D.** (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat Struct Mol Biol* **12**, 489-496.
- Scadden, A.D., and Smith, C.W.** (1997). A ribonuclease specific for inosine-containing RNA: a potential role in antiviral defence? *Embo J* **16**, 2140-2149.
- Schaub, M., and Keller, W.** (2002). RNA editing by adenosine deaminases generates RNA and protein diversity. *Biochimie* **84**, 791-803.
- Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A.** (2005). RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. *Plant Cell* **17**, 2791-2804.
- Schmitz-Linneweber, C., Tillich, M., Herrmann, R.G., and Maier, R.M.** (2001). Heterologous, splicing-dependent RNA editing in chloroplasts: allotetraploidy provides trans-factors. *Embo J* **20**, 4874-4883.
- Schmitz-Linneweber, C., Regel, R., Du, T.G., Hupfer, H., Herrmann, R.G., and Maier, R.M.** (2002). The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. *Mol Biol Evol* **19**, 1602-1612.
- Schmitz-Linneweber, C., Williams-Carrier, R.E., Williams-Voelker, P.M., Kroeger, T.S., Vichas, A., and Barkan, A.** (2006). A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell* **18**, 2650-2663.
- Schnauffer, A., Ernst, N.L., Palazzo, S.S., O'Rear, J., Salavati, R., and Stuart, K.** (2003). Separate insertion and deletion subcomplexes of the *Trypanosoma brucei* RNA editing complex. *Mol Cell* **12**, 307-319.
- Schnauffer, A., Panigrahi, A.K., Panicucci, B., Igo, R.P., Jr., Wirtz, E., Salavati, R., and Stuart, K.** (2001). An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science* **291**, 2159-2162.
- Schock, D., Kuo, S.R., Steinburg, M.F., Bolognino, M., Sparks, J.D., Sparks, C.E., and Smith, H.C.** (1996). An auxiliary factor containing a 240-kDa protein complex is involved in apolipoprotein B RNA editing. *Proc Natl Acad Sci U S A* **93**, 1097-1102.
- Schramke, V., and Allshire, R.** (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* **301**, 1069-1074.
- Schramke, V., and Allshire, R.** (2004). Those interfering little RNAs! Silencing and eliminating chromatin. *Curr Opin Genet Dev* **14**, 174-180.
- Schuster, W., Wissinger, B., Unsel, M., and Brennicke, A.** (1990). Transcripts of the NADH-dehydrogenase subunit 3 gene are differentially edited in *Oenothera* mitochondria. *Embo J* **9**, 263-269.
- Scott, J.** (1995). A place in the world for RNA editing. *Cell* **81**, 833-836.
- Seeburg, P.H., and Hartner, J.** (2003). Regulation of ion channel/neurotransmitter receptor function by RNA editing. *Curr Opin Neurobiol* **13**, 279-283.
- Seeburg, P.H., Single, F., Kuner, T., Higuchi, M., and Sprengel, R.** (2001). Genetic manipulation of key determinants of ion flow in glutamate receptor channels in the mouse. *Brain Res* **907**, 233-243.
- Seiwert, S.D., Heidmann, S., and Stuart, K.** (1996). Direct visualization of uridylyate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. *Cell* **84**, 831-841.

- Shadel, G.S.** (2004). Coupling the mitochondrial transcription machinery to human disease. *Trends Genet* **20**, 513-519.
- Shah, R.R., Knott, T.J., Legros, J.E., Navaratnam, N., Greeve, J.C., and Scott, J.** (1991). Sequence requirements for the editing of apolipoprotein B mRNA. *J Biol Chem* **266**, 16301-16304.
- Shikanai, T.** (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* **63**, 698-708.
- Sideraki, V., Wilson, D.K., Kurz, L.C., Quioco, F.A., and Rudolph, F.B.** (1996). Site-directed mutagenesis of histidine 238 in mouse adenosine deaminase: substitution of histidine 238 does not impede hydroxylate formation. *Biochemistry* **35**, 15019-15028.
- Simpson, L., Aphasizhev, R., Gao, G., and Kang, X.** (2004). Mitochondrial proteins and complexes in *Leishmania* and *Trypanosoma* involved in U-insertion/deletion RNA editing. *Rna* **10**, 159-170.
- Skuse, G.R., Cappione, A.J., Sowden, M., Metheny, L.J., and Smith, H.C.** (1996). The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. *Nucleic Acids Res* **24**, 478-485.
- Slavov, D., Clark, M., and Gardiner, K.** (2000a). Comparative analysis of the RED1 and RED2 A-to-I RNA editing genes from mammals, pufferfish and zebrafish. *Gene* **250**, 41-51.
- Slavov, D., Crnogorac-Jurcevic, T., Clark, M., and Gardiner, K.** (2000b). Comparative analysis of the DRADA A-to-I RNA editing gene from mammals, pufferfish and zebrafish. *Gene* **250**, 53-60.
- Small, I., Peeters, N., Legeai, F., and Lurin, C.** (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**, 1581-1590.
- Small, I.D., and Peeters, N.** (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* **25**, 46-47.
- Smith, N.A., Singh, S.P., Wang, M.B., Stoutjesdijk, P.A., Green, A.G., and Waterhouse, P.M.** (2000). Total silencing by intron-spliced hairpin RNAs. *Nature* **407**, 319-320.
- Sodhi, M.S., Burnet, P.W., Makoff, A.J., Kerwin, R.W., and Harrison, P.J.** (2001). RNA editing of the 5-HT(2C) receptor is reduced in schizophrenia. *Mol Psychiatry* **6**, 373-379.
- Sommer, B., Kohler, M., Sprengel, R., and Seeburg, P.H.** (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11-19.
- Sowden, M.P., Lehmann, D.M., Lin, X., Smith, C.O., and Smith, H.C.** (2004). Identification of novel alternative splice variants of APOBEC-1 complementation factor with different capacities to support apolipoprotein B mRNA editing. *J Biol Chem* **279**, 197-206.
- Spath, B., Settele, F., Schilling, O., D'Angelo, I., Vogel, A., Feldmann, I., Meyer-Klaucke, W., and Marchfelder, A.** (2007). Metal requirements and phosphodiesterase activity of tRNase Z enzymes. *Biochemistry* **46**, 14742-14750.
- Sprinzi, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S.** (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26**, 148-153.
- Stagno, J., Aphasizheva, I., Rosengarth, A., Luecke, H., and Aphasizhev, R.** (2007). UTP-bound and Apo structures of a minimal RNA uridylyltransferase. *J Mol Biol* **366**, 882-899.
- Staudinger, M., and Kempken, F.** (2003). Electroporation of isolated higher-plant mitochondria: transcripts of an introduced *cox2* gene, but not an *atp6* gene, are edited in organello. *Mol Genet Genomics* **269**, 553-561.

- Stephens, O.M., Yi-Brunozzi, H.Y., and Beal, P.A.** (2000). Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site. *Biochemistry* **39**, 12243-12251.
- Stoddard, B.L.** (2005). Homing endonuclease structure and function. *Q Rev Biophys* **38**, 49-95.
- Stuart, K., Allen, T.E., Heidmann, S., and Seiwert, S.D.** (1997). RNA editing in kinetoplastid protozoa. *Microbiol Mol Biol Rev* **61**, 105-120.
- Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W.** (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**, 60-89.
- Takenaka, M., and Brennicke, A.** (2003). In vitro RNA editing in pea mitochondria requires NTP or dNTP, suggesting involvement of an RNA helicase. *J Biol Chem* **278**, 47526-47533.
- Takenaka, M., Neuwirt, J., and Brennicke, A.** (2004). Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res* **32**, 4137-4144.
- Takenaka, M., Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., and Brennicke, A.** (2008). The process of RNA editing in plant mitochondria. *Mitochondrion* **8**, 35-46.
- Takenaka, M., Verbitskiy, D., van der Merwe, J.A., Zehrmann, A., Plessmann, U., Urlaub, H., and Brennicke, A.** (2007). In vitro RNA editing in plant mitochondria does not require added energy. *FEBS Lett* **581**, 2743-2747.
- Taylor, J.M.** (1999). Replication of human hepatitis delta virus: influence of studies on subviral plant pathogens. *Adv Virus Res* **54**, 45-60.
- Teng, B., Burant, C.F., and Davidson, N.O.** (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816-1819.
- Teng, B., Blumenthal, S., Forte, T., Navaratnam, N., Scott, J., Gotto, A.M., Jr., and Chan, L.** (1994). Adenovirus-mediated gene transfer of rat apolipoprotein B mRNA-editing protein in mice virtually eliminates apolipoprotein B-100 and normal low density lipoprotein production. *J Biol Chem* **269**, 29395-29404.
- Teng, B., Ishida, B., Forte, T.M., Blumenthal, S., Song, L.Z., Gotto, A.M., Jr., and Chan, L.** (1997). Effective lowering of plasma, LDL, and esterified cholesterol in LDL receptor-knockout mice by adenovirus-mediated gene delivery of ApoB mRNA editing enzyme (ApoBec1). *Arterioscler Thromb Vasc Biol* **17**, 889-897.
- 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.
- Tillich, M., Poltnigg, P., Kushnir, S., and Schmitz-Linneweber, C.** (2006). Maintenance of plastid RNA editing activities independently of their target sites. *EMBO Rep* **7**, 308-313.
- Tonkin, L.A., and Bass, B.L.** (2003). Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science* **302**, 1725.
- Tonkin, L.A., Saccomanno, L., Morse, D.P., Brodigan, T., Krause, M., and Bass, B.L.** (2002). RNA editing by ADARs is important for normal behavior in *Caenorhabditis elegans*. *Embo J* **21**, 6025-6035.
- Tsudzuki, T., Wakasugi, T., and Sugiura, M.** (2001). Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol* **53**, 327-332.
- Turmel, M., Otis, C., and Lemieux, C.** (2003). The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *Plant Cell* **15**, 1888-1903.

- Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giege, P.** (2008). PPR336 is associated with polysomes in plant mitochondria. *J Mol Biol* **375**, 626-636.
- Valente, L., and Nishikura, K.** (2005). ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. *Prog Nucleic Acid Res Mol Biol* **79**, 299-338.
- van der Merwe, J.A., Takenaka, M., Neuwirt, J., Verbitskiy, D., and Brennicke, A.** (2006). RNA editing sites in plant mitochondria can share cis-elements. *FEBS Lett* **580**, 268-272.
- Vanhamme, L., Perez-Morga, D., Marchal, C., Speijer, D., Lambert, L., Geuskens, M., Alexandre, S., Ismaili, N., Goringe, U., Benne, R., and Pays, E.** (1998). Trypanosoma brucei TBRGG1, a mitochondrial oligo(U)-binding protein that co-localizes with an in vitro RNA editing activity. *J Biol Chem* **273**, 21825-21833.
- Verbitskiy, D., Takenaka, M., Neuwirt, J., van der Merwe, J.A., and Brennicke, A.** (2006). Partially edited RNAs are intermediates of RNA editing in plant mitochondria. *Plant J* **47**, 408-416.
- Vollmar, W., Gloger, J., Berger, E., Kortenbruck, G., Kohling, R., Speckmann, E.J., and Musshoff, U.** (2004). RNA editing (R/G site) and flip-flop splicing of the AMPA receptor subunit GluR2 in nervous tissue of epilepsy patients. *Neurobiol Dis* **15**, 371-379.
- Vondruskova, E., van den Burg, J., Zikova, A., Ernst, N.L., Stuart, K., Benne, R., and Lukes, J.** (2005). RNA interference analyses suggest a transcript-specific regulatory role for mitochondrial RNA-binding proteins MRP1 and MRP2 in RNA editing and other RNA processing in Trypanosoma brucei. *J Biol Chem* **280**, 2429-2438.
- Wang, Q., Khillan, J., Gadue, P., and Nishikura, K.** (2000a). Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science* **290**, 1765-1768.
- Wang, Q., O'Brien, P.J., Chen, C.X., Cho, D.S., Murray, J.M., and Nishikura, K.** (2000b). Altered G protein-coupling functions of RNA editing isoform and splicing variant serotonin<sub>2C</sub> receptors. *J Neurochem* **74**, 1290-1300.
- Wang, Q., Miyakoda, M., Yang, W., Khillan, J., Stachura, D.L., Weiss, M.J., and Nishikura, K.** (2004). Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. *J Biol Chem* **279**, 4952-4961.
- Wang, T.P.** (1955). Cytosine nucleoside deaminase from Escherichia coli. *Methods Enzymol* **2**, 478-481.
- Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., Luo, D., Long, Y., Zhong, Y., and Liu, Y.G.** (2006). Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* **18**, 676-687.
- Williams, P.M., and Barkan, A.** (2003). A chloroplast-localized PPR protein required for plastid ribosome accumulation. *Plant J* **36**, 675-686.
- Wilson, D.K., Rudolph, F.B., and Quijcho, F.A.** (1991). Atomic structure of adenosine deaminase complexed with a transition-state analog: understanding catalysis and immunodeficiency mutations. *Science* **252**, 1278-1284.
- Wintz, H., and Hanson, M.R.** (1991). A termination codon is created by RNA editing in the petunia atp9 transcript. *Curr Genet* **19**, 61-64.
- Wolf, J., Gerber, A.P., and Keller, W.** (2002). tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. *Embo J* **21**, 3841-3851.
- Wong, S.K., Sato, S., and Lazinski, D.W.** (2001). Substrate recognition by ADAR1 and ADAR2. *Rna* **7**, 846-858.

- Worthey, E.A., Schnauffer, A., Mian, I.S., Stuart, K., and Salavati, R.** (2003). Comparative analysis of editosome proteins in trypanosomatids. *Nucleic Acids Res* **31**, 6392-6408.
- Xie, K., Sowden, M.P., Dance, G.S., Torelli, A.T., Smith, H.C., and Wedekind, J.E.** (2004). The structure of a yeast RNA-editing deaminase provides insight into the fold and function of activation-induced deaminase and APOBEC-1. *Proc Natl Acad Sci U S A* **101**, 8114-8119.
- Xu, F., Morin, C., Mitchell, G., Ackerley, C., and Robinson, B.H.** (2004). The role of the LRPPRC (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome c oxidase) I and COX III mRNA. *Biochem J* **382**, 331-336.
- Yamanaka, S., Poksay, K.S., Arnold, K.S., and Innerarity, T.L.** (1997). A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme. *Genes Dev* **11**, 321-333.
- Yamanaka, S., Poksay, K.S., Balestra, M.E., Zeng, G.Q., and Innerarity, T.L.** (1994). Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A zinc motif is essential for catalytic activity, and noncatalytic auxiliary factor(s) of the editing complex are widely distributed. *J Biol Chem* **269**, 21725-21734.
- Yamanaka, S., Balestra, M.E., Ferrell, L.D., Fan, J., Arnold, K.S., Taylor, S., Taylor, J.M., and Innerarity, T.L.** (1995). Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci U S A* **92**, 8483-8487.
- Yamazaki, H., Tasaka, M., and Shikanai, T.** (2004). PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in Arabidopsis. *Plant J* **38**, 152-163.
- Yang, J.H., Sklar, P., Axel, R., and Maniatis, T.** (1995). Editing of glutamate receptor subunit B pre-mRNA in vitro by site-specific deamination of adenosine. *Nature* **374**, 77-81.
- Yang, W., Wang, Q., Howell, K.L., Lee, J.T., Cho, D.S., Murray, J.M., and Nishikura, K.** (2005). ADAR1 RNA deaminase limits short interfering RNA efficacy in mammalian cells. *J Biol Chem* **280**, 3946-3953.
- Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., and Nishikura, K.** (2006). Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* **13**, 13-21.
- Yi-Brunozzi, H.Y., Stephens, O.M., and Beal, P.A.** (2001). Conformational changes that occur during an RNA-editing adenosine deamination reaction. *J Biol Chem* **276**, 37827-37833.
- Yoshinaga, K., Iinuma, H., Masuzawa, T., and Uedal, K.** (1996). Extensive RNA editing of U to C in addition to C to U substitution in the rbcL transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Res* **24**, 1008-1014.
- Yu, K., Huang, F.T., and Lieber, M.R.** (2004). DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J Biol Chem* **279**, 6496-6500.
- Yu, W., and Schuster, W.** (1995). Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria. *J Biol Chem* **270**, 18227-18233.

- Zabaleta, E., Mouras, A., Herould, M., Suharsono, S., and Araya, A.** (1996). Transgenic male-sterile plant induced by an unedited *atp9* gene is restored to fertility by inhibiting its expression with antisense RNA. *Proc Natl Acad Sci U S A* **93**, 11259-11263.
- Zauner, S., Greilinger, D., Laatsch, T., Kowallik, K.V., and Maier, U.G.** (2004). Substitutional editing of transcripts from genes of cyanobacterial origin in the dinoflagellate *Ceratium horridum*. *FEBS Lett* **577**, 535-538.
- Zhang, X.J., He, P.P., Li, M., He, C.D., Yan, K.L., Cui, Y., Yang, S., Zhang, K.Y., Gao, M., Chen, J.J., Li, C.R., Jin, L., Chen, H.D., Xu, S.J., and Huang, W.** (2004). Seven novel mutations of the ADAR gene in Chinese families and sporadic patients with dyschromatosis symmetrica hereditaria (DSH). *Hum Mutat* **23**, 629-630.
- Zhang, Z., and Carmichael, G.G.** (2001). The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* **106**, 465-475.



## **Abstract**

RNA editing in plants organelle transcripts is a process leading to specific post-transcriptional pyrimidine interconversions (mainly C-to-U). Editing specificity is determined by proteins, and recently a few **pentatricopeptide repeat (PPR)** proteins were described to be essential in this process and could act as specificity factors. However, the enzyme(s) catalysing RNA editing in plants is still totally unknown. We propose that the DYW domain found in a plant-specific PPR subfamily may be a catalytic domain necessary for the RNA editing process. During my PhD, I found that the phylogenetic distribution of the DYW domain is strictly correlated with RNA editing in green plants. Moreover, invariant residues of the DYW domain show similarities with the active site of cytidine deaminases. Taken together, these data strongly support the hypothesis that the DYW domain is implicated in RNA editing.

In addition, in *Arabidopsis thaliana*, a particular protein, named AtDYW1, consists of a targeting signal to the organelles and a DYW domain without any intervening PPR motifs. Inducible RNAi lines in which this gene has been targeted for knock-down exhibit strongly altered development. In the affected seedlings, *AtDYW1* gene expression is specifically decreased, and several editing defects in plastids transcripts were characterized. These results suggest that this protein may play a central role in RNA editing by acting as a trans-factor recruited by PPR proteins that lack the DYW domain. Demonstrating, that the AtDYW1 protein is the central enzyme in the RNA editing process will be an important finding in plant biology.

Key words : RNA editing, PPR proteins, DYW domain, mitochondria, chloroplast

## **Résumé**

Chez les plantes, l'édition des ARN dans les organites conduit principalement à des conversions de cytidines en uraciles. La spécificité de ce processus est déterminée par des protéines nucléaires. Récemment, il a été montré que certaines protéines de la famille PPR (**pentatricopeptide repeat**) sont essentielles à l'édition agissant probablement en tant que facteurs de spécificité. Cependant, l'identité de l'enzyme(s) catalysant la réaction d'édition reste inconnue. Nous avons proposé que le domaine « DYW » présent chez certaines protéines PPR puisse être le facteur catalytique de ce processus. En ce sens, mon travail de doctorat a permis de souligner l'existence d'une corrélation de la distribution phylogénétique du domaine DYW et de l'édition dans la lignée verte. De plus, les résidus invariants de ce domaine s'alignent avec le site catalytique de cytidines désaminases.

Par ailleurs, chez *Arabidopsis thaliana*, la caractérisation d'une protéine particulière de la famille PPR, la protéine AtDYW1 constituée uniquement du domaine DYW et d'un peptide signal pour les organites, a été entreprise. Cette protéine pourrait être un trans-facteur recruté par des protéines PPR, n'ayant pas de domaine DYW, pour venir éditer une base donnée au sein des transcrits. L'analyse de mutants affectés dans l'expression du gène *AtDYW1* a révélé l'existence de plantules au développement très affecté et pour lesquelles la protéine d'intérêt n'est plus exprimée. Plusieurs défauts d'édition dans des ARNm chloroplastiques ont pu être caractérisés dans ces mutants suggérant que la protéine AtDYW1 joue un rôle essentiel dans l'édition. L'enjeu principal reste de démontrer que cette protéine est l'enzyme centrale catalysant les réactions d'édition chez les plantes.

Mots clés : Edition des ARN, protéine PPR, domaine DYW, mitochondrie, chloroplaste